applied sciences

MDPI

Article

A Road Crack Detection Method Based on Residual and Attention Mechanism

Jianwu Xie [1], Weiwei Li [2], Wenwen Liu [3] and Hang Chen [4,*]

[1] E-School-Enterprise Cooperation Management Center, Tianjin Transportation Technical College, Tianjin 300393, China
[2] School of Computer Science and Technology, Tiangong University, Tianjin 300387, China
[3] Beijing Institute of Control and Electronic Technology, Beijing 102308, China
[4] School of Space Information, Space Engineering University, Beijing 101416, China
* Correspondence: hangchen@alu.hit.edu.cn

Abstract: This paper proposes a road crack detection method based on residual and attention mechanisms to address the issues of difficult detection of small cracks on road surfaces in complex backgrounds and inaccurate crack detection edges. This method introduces residual modules in the encoder stage to better extract crack detail features and introduces attention mechanism modules in the skip connection structure of the network to better locate crack positions. Training and testing on public datasets have shown that compared with existing partial detection methods, our method has improved segmentation accuracy and generalization, and is more precise in segmenting small cracks, thus verifying the superiority of the proposed method in this paper.

Keywords: image segmentation; crack detection; encoder-decoder; attention mechanism

check for updates

Citation: Xie, J.; Li, W.; Liu, W.; Chen, H. A Road Crack Detection Method Based on Residual and Attention Mechanism. *Appl. Sci.* **2024**, *14*, 5749. https://doi.org/10.3390/app14135749

Academic Editor: Luís Picado Santos

Received: 25 March 2024
Revised: 20 June 2024
Accepted: 27 June 2024
Published: 1 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The traditional road crack detection method records the location and degree of damage of cracks by investigators driving along the road. This method has the disadvantages of low efficiency, high human resource consumption, and significant subjective factors, and can no longer meet the increasing demand for road maintenance [1,2]. Therefore, it is very necessary to achieve automated detection of road cracks. Early machine vision-based image processing algorithms were sensitive to interference factors such as noise, lighting changes, and complex backgrounds in images, which could easily lead to inaccurate processing results, such as median filtering [3], threshold segmentation [4], edge detection [5,6], etc. In recent years, with the development of artificial intelligence [7,8], deep learning methods have been applied to road crack detection tasks.

The early deep learning-based pavement crack detection methods mainly adopted the classic fully convolutional network (FCN) and convolutional neural network (CNN) architectures. FCN is an image semantic segmentation model derived from CNN, which achieves pixel-level semantic segmentation by replacing traditional fully connected layers with fully convolutional layers. However, due to the pooling and upsampling operations of FCN, the network may experience a certain degree of spatial information loss during processing. In order to reduce information loss, Ronneberger et al. created a U-Net network model based on CNN architecture and applied it to biomedical segmentation [9]. U-Net adopts a skip connection structure to retain global image information and extract better features. U-Net and its encoding and decoding structure are widely used, and Liu et al. [10] first used the U-Net network model in crack detection, achieving higher accuracy than FCN. Gou et al. [11] proposed a crack detection method based on the Faster RCNN model and combined it with an improved feature extraction network, which has good crack detection accuracy. Ren et al. [12] proposed a network consisting of extended convolution, spatial pyramid pooling, and skip connection modules. Fan et al. [13] improved the accuracy of

detection by designing multi-layer extension modules and hierarchical feature learning modules. The above methods utilize the idea of multi-scale feature fusion to effectively improve the performance of crack detection, but are susceptible to noise and have the problem of missed detections. Lau et al. [14] used residual networks to extract crack features. Thitirat Siriborvornratanakul [15] proposed a convolutional neural network using severely imbalanced data for crack detection on road surfaces. Zou et al. [16] proposed a road crack detection network called DeepCrack based on SegNet, which combines the same scale features output by the encoding and decoding layers in pairs, and then concatenates and fuses all the fused images as the output of the network. Zhang et al. [17] proposed the CrackNet crack detection network, which was designed without the use of pooling layers, resulting in detection errors. Although the above methods have improved segmentation accuracy compared to manual methods, there are still some shortcomings. Current crack detection methods are difficult to ensure accurate detection of small cracks, and the detection results are easily affected by noise, resulting in missed detections.

Taking inspiration from the above literature, this paper proposes a road crack detection model based on residual and attention modules in an encoder decoder structure to address the problem of traditional road segmentation algorithms being difficult to identify narrow cracks and inaccurate segmentation edges. The rest of this article is organized as follows. Section 2 briefly introduces the relevant work. Section 3 introduces the details of the experimental Implementation and the corresponding experimental results. Section 4 is the conclusion.

2. Proposed Method

2.1. Unet Model

The Unet model is a widely used segmentation network in the field of medical images, which uses multi-scale features for semantic segmentation. Figure 1 shows its basic network structure. It can be seen that the Unet network is based on encoder and decoder structures. The Unet network generates feature maps with low resolution and high-dimensional semantics through convolution and pooling operations during the encoder stage. The decoder stage restores the feature map to its original resolution through continuous convolution and upsampling operations, ultimately obtaining the segmentation result. However, the model ignores local features of key edge information, resulting in lower universality of the network. Therefore, this article improves the feature extraction part of the network by adding residual structures with better feature extraction capabilities and introduces the CBAM attention module to allow the model to focus more on crack details.

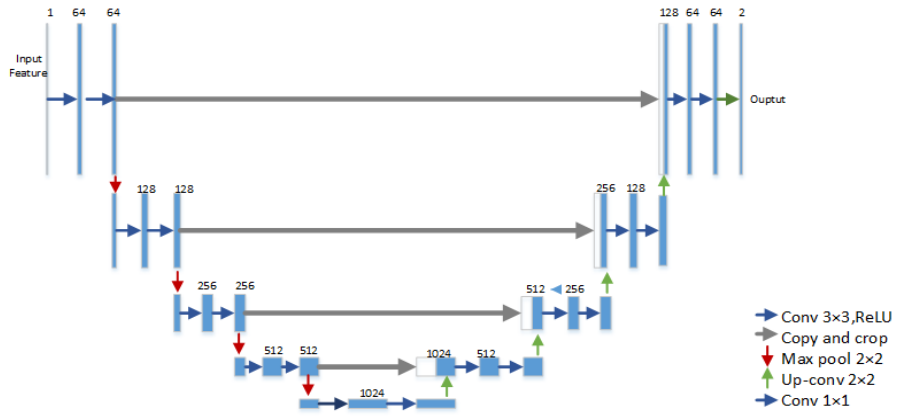


Figure 1. The network structure of Unet.

2.2. Feature Extraction Module

The number of network layers in the encoding part of the Unet model is relatively small. During the convolution process, the receptive field of each convolution kernel only covers a small local area of the input image. After convolution and downsampling

operations, the resolution of the feature map gradually decreases, leading to global information loss, which means that the feature learning ability of the model is limited [18]. For traditional U-Net, simply deepening the network can actually cause the network to face degradation problems, mainly manifested as gradient vanishing [19]. In order to extract more information from road crack images and improve the accuracy of the model in identifying road cracks, this paper chooses to introduce a deep residual structure into the backbone network, as shown in Figure 2. The cross-connection structure of the residual module adds input x to output F(x), where x is the input, and F(x) is the residual mapping function to be learned. This addition method not only does not increase the number of parameters and model complexity, but also effectively solves the problems of gradient disappearance, explosion, and network degradation in the backpropagation process of neural networks.

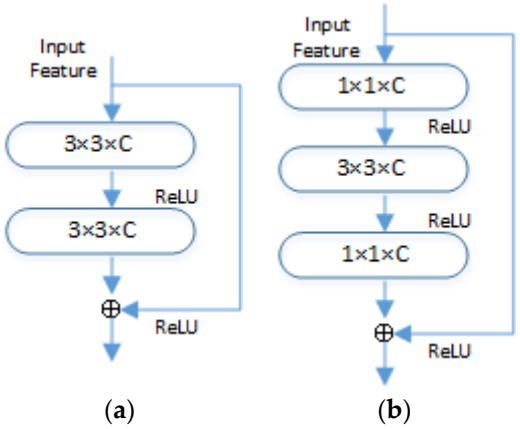


Figure 2. Residual structure. (**a**) BasicBlock module. (**b**) BottleNeck module.

In this article, the traditional convolutional layers in the upsampling and downsampling parts of the Unet network model are replaced with BasicBlock modules, and BottleNeck is embedded in the downsampling process. The introduction of these two residual structures can effectively solve the problem of model degradation and gradient disappearance as the number of network layers increases. Figure 2a shows the structure of the BasicBlock module, and Figure 2b shows the structure of the BottleNeck module.

2.3. Convolutional Attention Module

CBAM (convolutional block attention module) [20] is an attention module used to enhance the performance of convolutional neural networks. Figure 3 shows the structure of CBAM. CBAM is a lightweight general-purpose module consisting of two parts: channel attention module and spatial attention module. This structure not only saves parameters and computational power, but also ensures its successful integration into existing network architectures. The channel attention module can adaptively select important feature channels, while spatial attention can adaptively focus on important image regions. The focus of road crack segmentation is on precise segmentation of crack boundaries. The CBAM module combines channel and spatial attention to help the model capture the feature information of road cracks more comprehensively, further achieving the goal of improving segmentation accuracy. In this article, the feature maps extracted by the residual module are used as inputs to the CBAM module, which can enable the optimized network model to focus more on the key features of road crack formation and ignore irrelevant information, thereby improving the performance of the network.

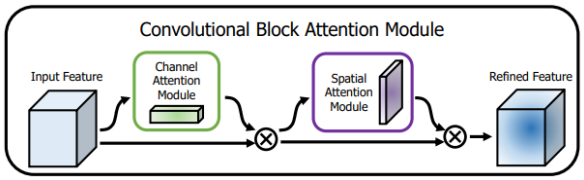


Figure 3. The structure of the CBAM module.

1. Channel attention module

The channel attention module aggregates spatial features through two methods: global average pooling and maximum pooling, as shown in Formula (1). The input features are represented by F, with a size of C × H × W. After performing global average pooling and maximum pooling operations on input feature, represented by F, two C × 1 × 1 channels are obtained. Then, these two channels are, respectively, fed into a multi-layer perceptron (MLP) with hidden layers. Next, the two obtained features are added together and passed through a sigmoid activation function to obtain the weight coefficient represented by Mc. Finally, Mc and F are multiplied to obtain the new feature, represented by F', as shown in Formula (2). The structure of the channel attention module is shown in Figure 4.

$$Mc(F) = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MPL}(\mathrm{MaxPool}(F))) \quad (1)$$

$$F' = Mc(F) \otimes F \quad (2)$$

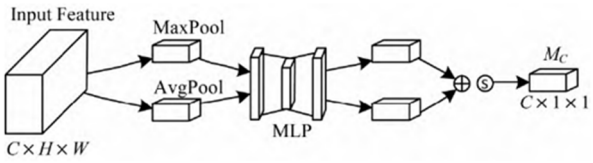


Figure 4. Structure of channel attention mechanism module.

Among them, F represents the input feature, Mc represents the weight coefficient, and F' represents the new feature obtained using the channel attention module. σ represents the sigmoid function, while AvgPool and MaxPool represent the average pooling and maximum pooling functions, respectively.

- Spatial attention module

The spatial attention module mainly focuses on the positions of meaningful features, which effectively supplement the information that the channel attention module focuses on. The structure of the spatial attention module is shown in Figure 5. The input of the spatial attention module is the feature obtained by the channel attention module represented by F'. Similar to the channel attention module, the spatial attention module also performs global average pooling and maximum pooling operations on the input features in the channel dimension, resulting in two 1 × H × W channels. Then, these two channels are concatenated together and compressed through a 7 × 7 convolutional layer. Next, the weight coefficients represented by Ms are obtained through a Sigmoid activation function operation, as shown in Formula (3). Finally, the weight coefficients represented by Ms and the input features represented by F' are multiplied to obtain the new features processed by the spatial processing module, represented by F'', as shown in Formula (4). Where σ represents the sigmoid function and $f^{7\times7}$ represents a convolution operation with a size of 7 × 7.

$$Ms(F) = \sigma(f^{7\times7}([\mathrm{AvgPool}(F), \mathrm{MaxPool}(F)])) \quad (3)$$

$$\mathrm{F}'' = Ms(F') \otimes F' \quad (4)$$

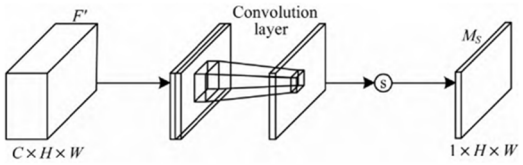


Figure 5. Structure of the spatial attention mechanism module.

2.4. Our Method

Existing road crack detection algorithms often have the problem of low accuracy in detecting small cracks. In order to solve these problems, this article introduces the ResNet model's excellent residual structure in the encoding stage of UNet's model to improve the network's ability to extract crack detail features. In addition, we have improved the UNet structure's skip connections by introducing attention mechanisms in the skip connections section, aiming to enhance the network's crack localization and anti-interference capabilities. Figure 6 shows the network structure of our model. We evaluated the network model proposed in this article using publicly available datasets. The experimental results on public datasets showed that our model outperforms state-of-the-art crack segmentation models in terms of overall performance in road crack segmentation.

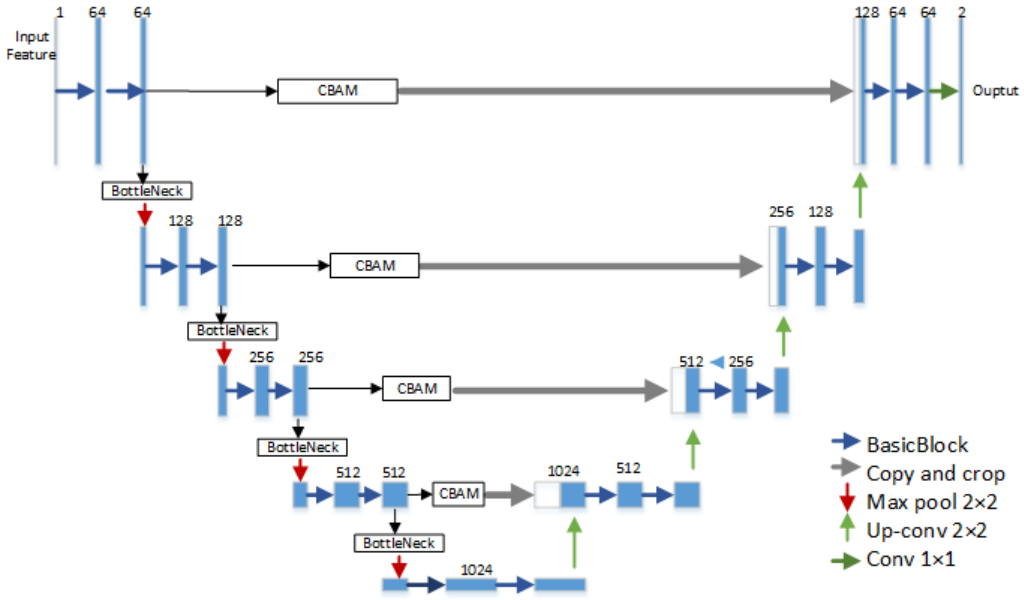


Figure 6. The network structure of our model.

3. Experimental Results and Analysis

In this section, we first introduce the experimental setup, and then present the crack detection comparison results between the proposed algorithm and other cash algorithms and conduct an in-depth analysis of the comparison results. Finally, we conduct ablation experiments on the proposed network structure and analyze the experimental results.

3.1. Experimental Setup

2. Experimental platform: All network models in this article were implemented based on a deep learning architecture called PyTorch. In our proposed network structure, the SGD optimization method was used to update parameters, with a learning rate initialized to 0.01 and momentum optimization algorithm set to 0.9. All experiments in this paper were carried out using GeForce RTX 3060 GPU, which is from NVIDIA, Santa Clara, CA, USA.
3. Datasets: This study used three publicly available crack datasets, namely CrackTree260, CrackLS315, and DeepCrack [21]. The CrackTree260 dataset is a dataset

of asphalt pavement images, which includes 260 pavement images with a size of 800 × 600 pixels. The CrackLS315 dataset includes 315 crack images with a size of 512 × 512 pixels. The DeepCrack dataset is a dataset of concrete pavement images, which includes 537 pavement images with dimensions of 544 × 384 pixels. The labeled images in these datasets were manually labeled. To address the issue of an insufficient number of images in the dataset, we employed four data augmentation methods: vertical reflection, mirror reflection, translation, and random rotation. Through the above operations, the number of training datasets can be expanded by 8 times. To validate the established neural network models, we selected 75% of each dataset as training data and 25% as test data.

Factors such as lighting and exposure can lead to unclear crack boundaries in road surface images, which increases the difficulty of road surface crack detection. To address this issue, this article adopted the CLAHE (contrast limited adaptive histogram equalization) method to enhance the crack areas in road crack images by adjusting the contrast of local regions. Compared with ordinary histogram equalization algorithms, CLAHE divides the entire image into blocks, and then counts the frequency of each pixel value appearing in each block to obtain the histogram of each block. Contrast limiting is applied to each block, also known as cropping limiting. Usually, we evenly distribute the trimmed parts to other parts of the histogram. An image processed by the CLAHE algorithm can greatly enhance the difference between crack information and road background, which greatly improves the quality of the sample image. The images before and after CLAHE enhancement processing are shown in Figure 7.

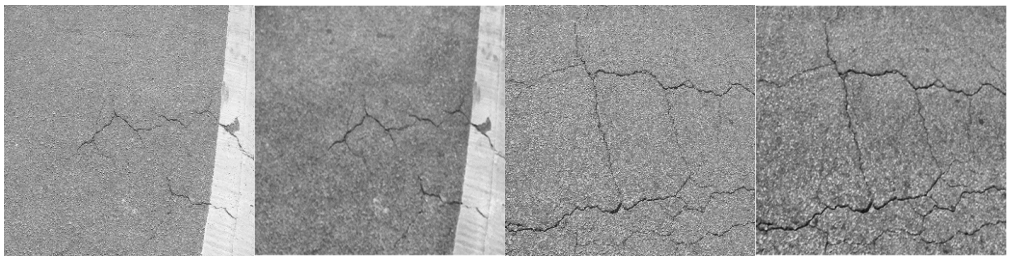


Figure 7. Contrast images before and after CLAHE enhancement. (**a**) Original image. (**b**) Enhanced image corresponding to (**a**). (**c**) Original image. (**d**) Enhanced image corresponding to (**c**).

4. Evaluation indicators: In this article, classic evaluation indicators in the field of semantic segmentation, such as precision (P), recall (R), and F1 score, were selected for evaluation. Formula (5) is a calculation method for accuracy, which can express the ratio of detection results to ground-truth. Formula (6) is the calculation method for recall rate, which represents the percentage of correctly detected crack pixels to detected crack pixels. Formula (7) is the calculation method for F1 value, which can measure both accuracy and recall. The higher the F1 value, the better the detection effect.

Among them, TP represents the number of pixels correctly classified as cracks, FP represents the number of pixels incorrectly classified as cracks, and FN represents the number of pixels incorrectly classified as background, i.e., false negatives

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

$$\mathrm{F1} = \frac{2 \times P \times R}{P + R} \tag{7}$$

5. Comparison method: This article compared the advanced crack segmentation methods; all comparison methods were based on deep learning as follows:
   - U-Net. Its network structure mainly consists of three parts: encoder, decoder, and skip connection, and is widely used in the field of image segmentation;
   - Jing et al. [22] proposed a deep convolutional neural network based on attention mechanism and residual structure;
   - Junzhou Chen et al. [23] proposed a refined crack detection method via LECS-Former for autonomous road inspection vehicles.

3.2. Visualization Analysis of Experimental Results

Table 1 presents the test results of different network models on the CrackTree260, CrackLS315, and DeepCrack datasets under the same experimental setup. The optimal values for each evaluation indicator have been bolded. From Table 1, it can be seen that the proposed network model had the best F1 score on all three test datasets. Compared with the methods mentioned in UNet and Refs. [21,22], our model achieved an accuracy improvement of 5.7%, 1.3%, and 1.8% on the CrackTree260 dataset, respectively. Our model had improved accuracy by 6.7%, 3.5%, and 1.4% on the CrackLS315 dataset, respectively. Our model had improved accuracy by 0.5%, 0.2%, and 3.5% on the DeepCrack dataset, respectively.

Table 1. Performance of different methods on CrackTree260, CrackLS315, and DeepCrack datasets.

Methods	CrackTree260			CrackLS315			DeepCrack		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
UNet	0.729	0.757	0.743	0.677	0.712	0.694	0.746	0.606	0.669
Ref. [22]	0.773	0.791	0.782	0.709	0.694	0.702	0.749	**0.766**	0.757
Ref. [23]	0.768	**0.795**	0.781	0.730	**0.742**	0.736	0.716	0.706	0.711
Our method	**0.786**	0.784	**0.785**	**0.744**	0.732	**0.738**	**0.751**	0.765	**0.758**

Figure 8 shows the detection results of different methods on the CrackTree260 dataset. Compared with the results of the crack detection methods shown in Figure 8c–e, the edge information of the cracks detected by our model was richer and had better continuity. Although the crack detection methods in Figure 8 can obtain relatively complete crack shapes, the crack detection methods in Figure 8c–e lost a lot of crack detail information and were more susceptible to noise compared to the model proposed in this paper. For example, there are relatively small cracks in the red and blue rectangles of the first image. Other methods can also detect cracks here, but there may be varying degrees of fracture, and the method proposed in this article can also detect small cracks more completely and continuously. In the third image, the noise is quite severe. Except for the algorithm in this article, all other algorithms are affected to varying degrees by the noise and background information in the image, resulting in missed detections and false positives of small crack details in the image.

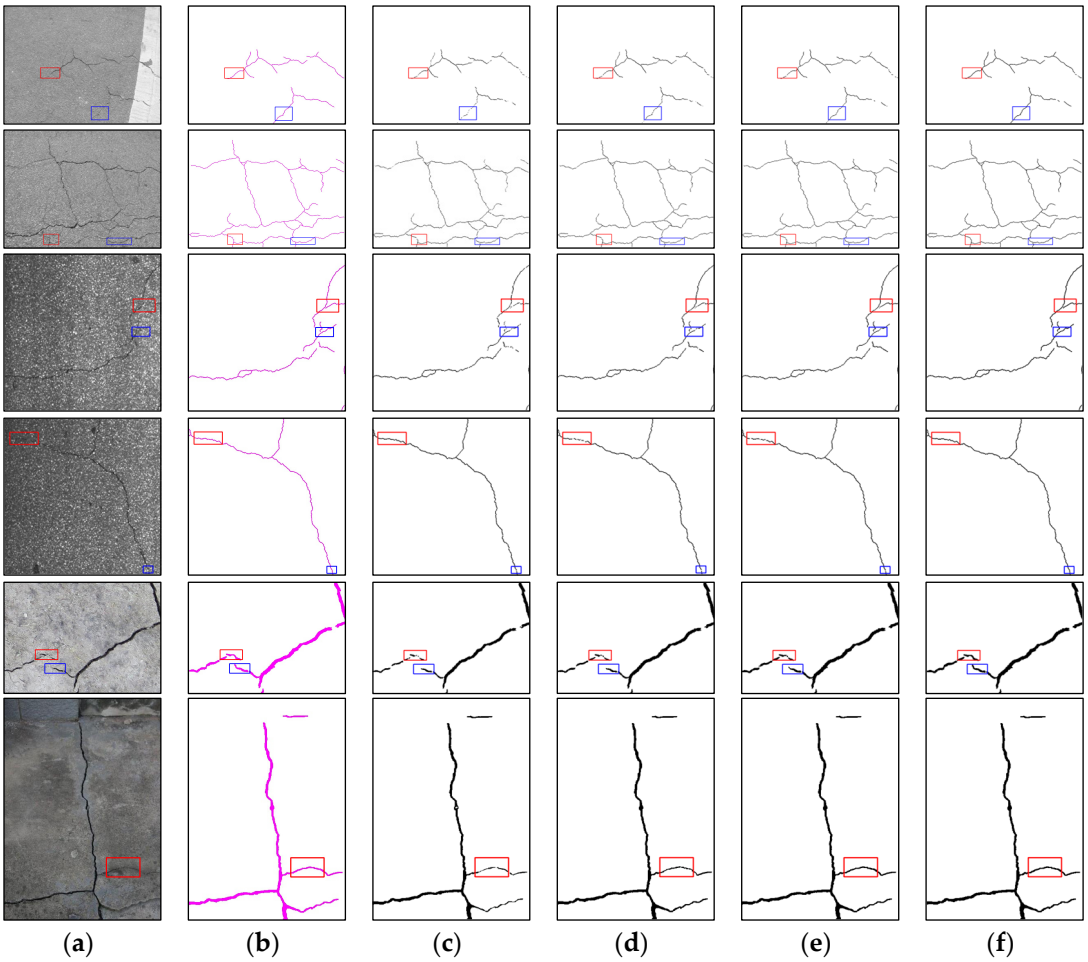


Figure 8. Detection results of four methods in CrackTree260, CrackLS315, and DeepCrack datasets, from top to bottom, correspond to 2 images selected from each dataset. (**a**) Source Image. (**b**) Ground-truth. (**c**) Detection results of UNet. (**d**) Detection results of Ref. [21]. (**e**) Detection results of Ref. [22]. (**f**) Our method.

3.3. Ablation Experiment

In order to verify the effectiveness of the BasicBlock, BottleNeck, and CBAM modules proposed in this paper, five ablation experiments were designed on the CrackTree260, CrackLS315, and DeepCrack datasets. The first group was UNet. The second group used the BasicBlock module instead of traditional convolution to extract features in the encoder structure of UNet. The third group added a BottleNeck module to the encoder structure of UNet to extract features based on the second group. The fourth group introduced the CBAM module in the skip connection part of the UNet network to make the model more focused on small cracks. The fifth group was based on UNet and included three modules: BasicBlock, BottleNeck, and CBAM.

Tables 2–4 show the ablation experimental results for the CrackTree260, CrackLS315, and DeepCrack datasets, respectively. By comparing the results, we can conclude that the network model with the added BasicBlock module and BottleNeck module enhanced the ability to extract crack features, thereby improving network performance. The overall crack area segmentation effect was better than UNet. The CBAM module can enhance the model's ability to obtain global features of cracks, while also improving the model's ability to extract features from crack edges and small cracks. The network model that simultaneously introduced three modules significantly exceeded the baseline network model UNet in various indicators. Among them, the accuracy was improved by 9.3% and F1 by 9.4% on the CrackTree260 dataset, while the accuracy was improved by 5.3% and F1

by 4.2% on the CrackLS315 dataset. The accuracy was improved by 3% and F1 by 4.2% on the DeepCrack dataset. The results of the ablation experiment indicate that the introduction of these three modules improves the network model's ability to detect small cracks while also enhancing the integrity of crack detection. According to the experimental results, compared to using one or two modules alone, the network model that used three modules simultaneously had better crack detection performance and richer detail.

Table 2. Results of ablation experiments on the CrackTree260 dataset.

UNet	BasicBlock	BottleNeck	CBAM	P	R	F1
●	○	○	○	0.693	0.690	0.691
●	●	○	○	0.740	0.714	0.727
●	●	●	○	0.721	0.695	0.708
●	○	○	●	0.748	0.726	0.737
●	●	●	●	0.786	0.784	0.785

Table 3. Results of ablation experiments on the CrackLS315 dataset.

UNet	BasicBlock	BottleNeck	CBAM	P	R	F1
●	○	○	○	0.691	0.702	0.696
●	●	○	○	0.702	0.715	0.708
●	●	●	○	0.732	0.709	0.720
●	○	○	●	0.723	0.708	0.715
●	●	●	●	0.744	0.732	0.738

Table 4. Results of ablation experiments on the DeepCrack dataset.

UNet	BasicBlock	BottleNeck	CBAM	P	R	F1
●	○	○	○	0.721	0.712	0.716
●	●	○	○	0.732	0.725	0.728
●	●	●	○	0.742	0.729	0.735
●	○	○	●	0.738	0.728	0.733
●	●	●	●	0.751	0.765	0.758

4. Conclusions

This article proposes a crack detection method based on residual and attention mechanisms in the structure of encoder and decoder. We added residual modules in the encoder stage to extract more detailed features of cracks and introduce attention mechanism modules in the skip connection part of the network. These operations enable our network model to pay more attention to crack features in the image and suppress the influence of irrelevant features. The experimental results showed that the proposed model had higher segmentation accuracy for small cracks compared to the compared crack detection models, and effectively suppressed the interference of background factors. In the future, we will continue to optimize the model, ensuring detection accuracy while reducing computational complexity to meet the requirements of real-time detection.

Author Contributions: Formal analysis, Validation, J.X., W.L. (Weiwei Li), W.L. (Wenwen Liu) and H.C.; Investigation, W.L. (Weiwei Li) and H.C.; Writing—original draft, J.X., and H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: The authors are indebted to the four anonymous reviewers for their professional comments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang, D.J.; Li, Q.Q.; Chen, Y.; Cao, M.; He, L. Asphalt Pavement Crack Detection Based on Spatial Clustering Feature. *Acta Autom. Sin.* **2016**, *42*, 443–454.
2. Jiang, W.B.; Luo, Q.R.; Zhang, X.H. A Review of Concrete Roads Crack Detection Methods Based on Digital Image. *J. Xihua Univ. Nat. Sci. Ed.* **2018**, *37*, 75–84.
3. Maode, Y.; Shaobo, B.; Kun, X.; Yuyao, H. Pavement crack detection and analysis for high-grade highway. In Proceedings of the 2007 8th International Conference on Electronic Measurement and Instruments, Xi'an, China, 16–18 August 2007; pp. 4–548.
4. Oliveira, H.; Correoa, P.L. Automatic road crack segmentation using entropy and image dynamic thresholding. In Proceedings of the European Signal Processing Conference, Glasgow, UK, 24–28 August 2009; pp. 622–626.
5. Zhao, H.L.; Qin, G.F.; Wang, X.J. Improvement of canny algorithm based on pavement edge detection. In Proceedings of the 2010 3rd International Congress on Image and Signal Processing (CISP), Yantai, China, 16–18 October 2010; Volume 2, pp. 964–967.
6. Ayenu-Prah, A.; Attoh-Okine, N. Evaluating pavement cracks with bidimensional empirical mode decomposition. *EURASIP J. Adv. Signal Process.* **2008**, *2008*, 861701. [CrossRef]
7. Qu, Z.; Chen, Y.-X.; Liu, L.; Xie, Y.; Zhou, Q. The Algorithm of Concrete Surface Crack Detection Based on the Genetic Programming and Percolation Model. *IEEE Access* **2019**, *7*, 57592–57603. [CrossRef]
8. Zhang, A.; Wang, K.C.P.; Li, B.; Yang, E.; Dai, X.; Peng, Y.; Fei, Y.; Liu, Y.; Li, J.Q.; Chen, C. Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces Using a Deep-Learning Network. *Comput.-Aided Civ. Infrastruct. Eng.* **2017**, *32*, 805–819. [CrossRef]
9. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241.
10. Liu, Z.; Cao, Y.; Wang, Y.; Wang, W. Computer vision-based concrete crack detection using U-net fully convolutional networks. *Autom. Constr.* **2019**, *104*, 129–139. [CrossRef]
11. Gou, C.; Peng, B.; Li, T.; Gao, Z. Pavement Crack Detection Based on the Improved Faster-RCNN. In Proceedings of the 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Dalian, China, 14–16 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 962–967.
12. Ren, Y.; Huang, J.; Hong, Z.; Lu, W.; Yin, J.; Zou, L.; Shen, X. Image-based concrete crack detection in tunnels using deep fully convolutional networks. *Constr. Build. Mater.* **2020**, *234*, 117367. [CrossRef]
13. Fan, Z.; Li, C.; Chen, Y.; Wei, J.; Loprencipe, G.; Chen, X.; Di Mascio, P. Automatic crack detection on road pavements using encoder-decoder architecture. *Materials* **2020**, *13*, 2960. [CrossRef] [PubMed]
14. Lau, S.L.; Chong, E.K.; Yang, X.; Wang, X. Automated pavement crack segmentation using u-net-based convolutional neural network. *IEEE Access* **2020**, *8*, 114892–114899. [CrossRef]
15. Thitirat, S. Pixel-level thin crack detection on road surface using convolutional neural network for severely imbalanced data. *Comput.-Aided Civ. Infrastruct. Eng.* **2023**, *11*, 2300–2316.
16. Zou, Q.; Zhang, Z.; Li, Q.; Qi, X.; Wang, Q.; Wang, S. Deepcrack: Learning hierarchical convolutional features for crack detection. *IEEE Trans. Image Process.* **2018**, *28*, 1498–1512. [CrossRef] [PubMed]
17. Mandal, V.; Uong, L.; Adu-Gyamfi, Y. Automated Road Crack Detection Using Deep Convolutional Neural Networks. In Proceedings of the 2018 IEEE International Conference on Big Data, Seattle, WA, USA, 10–13 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 5212–5215.
18. Hui, B.; Li, Y. Pavement crack detection method based on improved U-shaped neural network. *Traffic Inf. Saf.* **2023**, *41*, 105–114.
19. Jiang, W.B.; Liu, M.; Peng, Y.N.; Wu, L.; Wang, Y. HDCB-net: A Neural Network with the Hybrid Dilated Convention for Pixel-level Crack Detection on Concrete Bridges. *IEEE Trans. Ind. Inform.* **2021**, *17*, 5485–5494. [CrossRef]
20. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
21. Liu, Y.; Yao, J.; Lu, X.; Xie, R.; Li, L. DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing* **2019**, *338*, 139–153. [CrossRef]
22. Jing, P.; Yu, H.; Hua, Z.; Xie, S.; Song, C. Road Crack Detection Using Deep Neural Network Based on Attention Mechanism and Residual Structure. *IEEE Access* **2023**, *11*, 919–929. [CrossRef]
23. Chen, J.; Zhao, N.; Zhang, R.; Chen, L.; Huang, K.; Qiu, Z. Refined Crack Detection via LECSFormer for Autonomous Road Inspection Vehicles. *IEEE Trans. Intell. Veh.* **2023**, *3*, 2049–2061. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.