Wiley
Structural Control and Health Monitoring
Volume 2024, Article ID 1299095, 14 pages
https://doi.org/10.1155/2024/1299095

WILEY

Research Article

Pixel-Level Crack Identification for Bridge Concrete Structures Using Unmanned Aerial Vehicle Photography and Deep Learning

Fei Song,[1] **Bo Liu**,[2] **and Guixia Yuan**[1]

[1]*School of Information Technology, Jiangsu Open University, Nanjing 210017, China*
[2]*School of Computer and Software, Nanjing Vocational University of Industry Technology, Nanjing 210023, China*

Correspondence should be addressed to Fei Song; songf@jsou.edu.cn

Received 24 August 2023; Revised 14 December 2023; Accepted 4 May 2024; Published 24 May 2024

Academic Editor: Suparno Mukhopadhyay

Traditional manual inspection technology has the problems of high risk, low efficiency, and being time-consuming in bridge safety management. The unmanned aerial vehicle (UAV)-based detection technology is widely used in bridge structure safety monitoring. However, the existing deep learning-based concrete crack identification method has great limitations in dealing with complex background and tiny cracks in bridge structures. To address these problems, this study designs a crack pixel-level high-performance segmentation model for bridge concrete cracks that is suitable for UAV detection scenarios using machine vision (MV) and deep learning (DL) algorithms. First, considering the high requirements for the computing performance of the MV-based model for UAV-based detection, the ResNet-18-based lightweight convolutional neural network is used to represent the traditional large-scale backbone network of the pyramid scene parsing network (PSPNet) to develop a high-performance crack automatic identification model. Then, considering that bridge concrete cracks have the characteristics of subtle shapes and complex backgrounds, the spatial position self-attention module is inserted into the PSPNet to improve its detection accuracy. A concrete bridge is used for the case study, and a dataset of cracks in bridge concrete structures collected by UAVs is constructed and used for model training. The experimental results show that the loss function of the developed method in the training process results in a smooth decline, and the developed algorithm achieves the evaluation indicators of 0.9008 precision, 0.8750 recall, 0.8820 accuracy, and 0.9012 IOU on the bridge concrete crack dataset, which are significantly higher than other state-of-the-art baseline methods. In addition, four common UAV bridge detection scenarios, including low light, complex crack forms, high background roughness, and complex background scenes, are used to further test the crack detection ability of the developed crack identification model. The experimental results show that the proposed crack identification method can effectively overcome interference and real-size pixel-level segmentation of crack morphology. In addition, it also achieved a detection efficiency of 35.04 FPS, which shows that the real-time detection ability of the method has good applicability in the UAV detection scene.

1. Introduction

In recent years, with the acceleration of urbanization, China has built a large number of civil infrastructures, which are huge in size and expensive in construction and operating costs, and their safety is related to the national economy and people's livelihood [1–4]. As an important channel for transportation and the economic lifeline of China, the construction and maintenance of bridges have always been an important part of infrastructure construction. With the increase in the service life of bridges, the safety of the service performance of the engineering structure has been paid more and more attention by the relevant departments.

The traditional bridge detection technology is to evaluate the safety status of the bridge structure through manual visual inspection or the information obtained using portable instrument measurement [5]. At present, the manual detection of long-span bridges has shortcomings such as strong subjectivity, poor integrity, poor timeliness, and affecting normal operations.

To maintain the structural safety of the bridge in the long-term use process and the driving safety in the service state, the structural health monitoring (SHM) technology during operation is an effective method [6, 7]. Benefitting from its simple operation, flexible maneuverability, and excellent performance, the unmanned aerial vehicle (UAV)-based photography technology is suitable for performing tasks in complex scenes [8–10]. Moreover, in the past few years, autonomous UAV inspection technology and its related obstacle-avoidance method have been proposed and gradually applied to the field of SHM [11, 12]. Because of this, UAV-based aerial photography technology plays an important role in the daily inspection of civil infrastructure and the identification of potential diseases [10]. However, due to the difference in flight height and viewing angle when shooting, compared with natural images taken at horizontal angles, the images captured by UAVs contain more abundant small targets, and the objects in the image are arranged in disorder, random direction, and complex background [13, 14]. Therefore, it is necessary to study digital image processing methods suitable for UAV-based detection scenarios to improve detection efficiency and accuracy performance.

In the past few decades, with the continuous emergence and rapid development of emerging technologies, artificial intelligence (AI) and deep learning (DL) techniques have been gradually applied to bridge construction, operation, and maintenance [15–19]. For instance, Bae et al. [20] developed an automated crack detection method by combining deep super-resolution crack networks and deep learning models. However, the existing AI-based defect detection methods for bridge concrete structures lack the pertinence for specific UAV detection scenarios. In addition, the existing small-size crack images based on public datasets are not enough to train a high-performance neural network crack identification model to solve the detection problem of superlarge slender crack images. The bridge concrete structure cracks suffer from problems of noise, irregular distribution, and complex backgrounds. This is to say that the traditional DL-based semantic segmentation model is prone to problems such as incomplete detection of small cracks and complex cracks.

The deep convolution neural network using spatial pyramid pooling and attention mechanisms has been proven to be an effective means of building efficient and accurate detection models [21, 22]. In addition, according to the related studies, it can be seen that the attention mechanism is also an important strategy to improve the recognition effect of deep convolutional networks and increase identification accuracy. To solve the above problems, this study first uses UAV detection technology to develop a dataset of bridge concrete structure cracks. On this basis, considering the image features of concrete cracks captured by UAV aerial photography, an attention mechanism is introduced to improve the pyramid scene parsing network (PSPNet), thereby improving the feature extraction and identification capabilities for tiny cracks. A bridge that has been in operation for many years is taken as an example, combined with UAV detection technology to collect structural images and create related datasets. A series of qualitative and quantitative evaluations are used to verify the effectiveness and practical performance of the method in bridge crack detection and inference.

The main contributions of this study can be attributed as follows:

(1) The combination of the spatial position self-attention module and the ResNet-18-based lightweight convolutional neural network can improve the perception of contextual information in the spatial dimension, thereby improving the identification performance of bridge cracks

(2) The identification accuracy and inference efficiency of the constructed method have been validated in a variety of complex UAV bridge detection scenarios, including low-light conditions, diverse concrete cracks, high roughness, and high-complexity scenes

(3) Quantitative evaluation results show that the constructed bridge concrete crack segmentation model achieves the evaluation indicators of 0.9008 precision, 0.8750 recall, 0.8820 accuracy, and 0.9012 IOU on the bridge concrete crack dataset, which are significantly higher than other state-of-the-art baseline methods

The rest of this paper is described as follows. Firstly, the developed network and the mathematical principles of each part are included in Section 2. Then, a bridge after long-term service is used as the case study, and the dataset construction process is described in Section 3. Details of the model training environment, training process, and evaluation versus validation are included in Section 4. Lastly, the conclusions and limitations of this study are provided in the final part of this paper.

2. Methodology

In this section, the overall architecture of the developed network is first presented to provide a workflow. Then, the theory about the components of the model composition is further elaborated. The specific content is as follows.

2.1. The Overview of the Network: PSPNet. With the application of fully convolutional networks in semantic segmentation, a series of deep learning-based semantic segmentation networks, such as U-Net and DeepLab v3+, have developed rapidly. However, these traditional networks undergo multiple downsampling operations, such as convolution and pooling, and the detailed information of the image. This indicates that the information will be gradually lost, which will affect the accuracy of small concrete cracks. To adapt to the image characteristics of UAV high-resolution acquisition, this study chose the PSPNet model with lightweight and small parameters as the basic model.

PSPNet is a DL-based decode-encode model for image semantic segmentation. Figure 1 demonstrates the diagram of the overall architecture of the network. Compared with other DL-based semantic segmentation networks, one

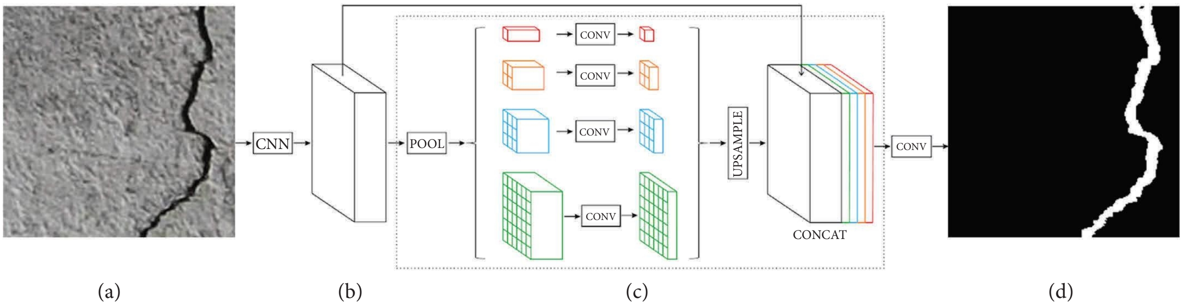


Figure 1: The PSPNet-based semantic segmentation architecture. (a) Input image. (b) Feature map. (c) Pyramid pooling model. (d) Final prediction.

significant advantage of PSPNet is its ability to effectively capture global context information from an image. Pyramid pooling operation (PPO) is the core ingredient of the PSPNet, which captures multiscale features and aggregates them into a fixed-length representation. It involves dividing an image into a set of grids and then pooling features within each grid at multiple scales. This allows for the extraction of both local and global features, which can improve the performance of tasks such as object recognition and semantic segmentation.

In addition, PSPNet has high-efficiency computing characteristics, mainly because dilated convolution is used to expand the receptive field of the convolution kernel, thereby reducing the number of parameters and calculations and improving the efficiency of model training and inference in actual scenarios. By incorporating this global context information, PSPNet can achieve state-of-the-art performance on several semantic segmentation benchmarks, while maintaining a relatively small number of parameters compared with other models. Due to the use of global context information, PSPNet can better understand the relationship between objects in the image, thus improving the accuracy of crack segmentation. Additionally, PSPNet can be trained end to end, which simplifies the training process and reduces the need for postprocessing steps.

2.2. The ResNet-Based Backbone for Feature Extraction.

The purpose of upsampling and downsampling in the PSPNet is mainly to transfer and extract image information. One of the key advantages of the PSPNet model is that the residual network (e.g., ResNet-101) is used as the feature extraction network, and the PPO module is used to collect information on different regions and different scales, which has improved spatial accuracy. Figure 2 shows the diagram of the architecture of the ResNet feature extraction backbone network. The ResNet architecture consists of several building blocks called residual blocks, each of which contains one or more residual connections. However, the limited computing power of detection equipment such as UAVs has led to higher requirements for the computing power and detection performance of the network.

Based on this, this study proposes a lightweight convolutional neural network ResNet-18 combined with a targeted improvement of the network feature extraction layer. The overall network of ResNet-18 used as the feature extraction network of the PSPNet is depicted in Figure 3.

It can be seen from figure that ResNet-18 is a variant of the ResNet architecture that has 18 layers, indicating that it requires less memory and computational resources and making it shallower and faster to train than deeper variants such as ResNet-50. Additionally, ResNet-18 is less prone to overfitting than deeper models, which can be beneficial when working with small datasets.

The ResNet architecture is a type of deep convolutional neural network that uses residual connections to enable the training of very deep networks. The formula for ResNet can be expressed as follows:

$$y = \mathcal{F}\left(x, \{w_i\}\right) + x, \qquad (1)$$

where x and y are the input and output feature maps, respectively, and $\mathcal{F}()$ is a residual function that is implemented as a series of convolutional layers with batch normalization and ReLU activation. The weights of the convolutional layers are denoted byw_i. The residual connection allows the input to bypass one or more layers of the network and be added directly to the output of the residual function. This helps to mitigate the vanishing gradient problem that can occur in very deep networks, allowing ResNet to be trained with up to hundreds of layers.

In this study, the pretrained ResNet-18 is a variant of the ResNet-18 architecture that has been trained on a large dataset (such as ImageNet), and the weights of the model have been saved. Figure 4 shows the overall diagram of the pretrained model. These pretrained weights can then be used as a starting point for other MV-based tasks, where the model is fine-tuned on a smaller dataset, like the bridge concrete crack dataset. The use of the pretrained ResNet-18 model can provide several advantages, including faster convergence during training, better generalization to new datasets, and improved accuracy on the target task. This is because the pretrained model has already learned a set of

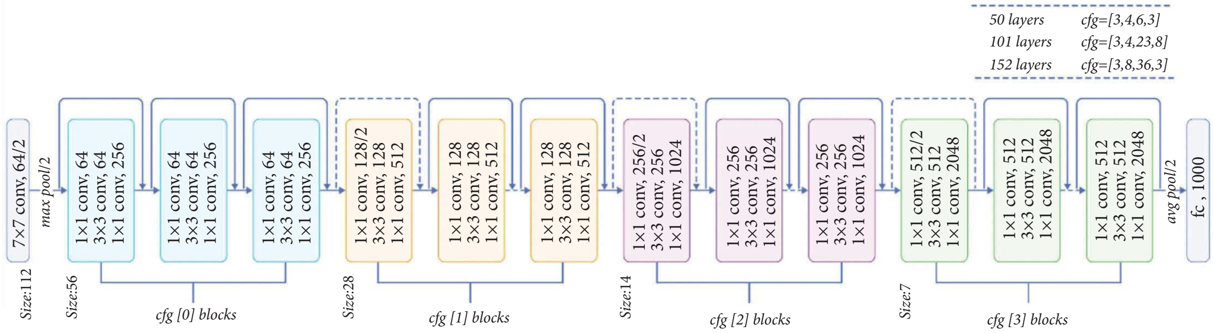


Figure 2: The overall diagram of ResNet-50 architecture as the backbone extractor.

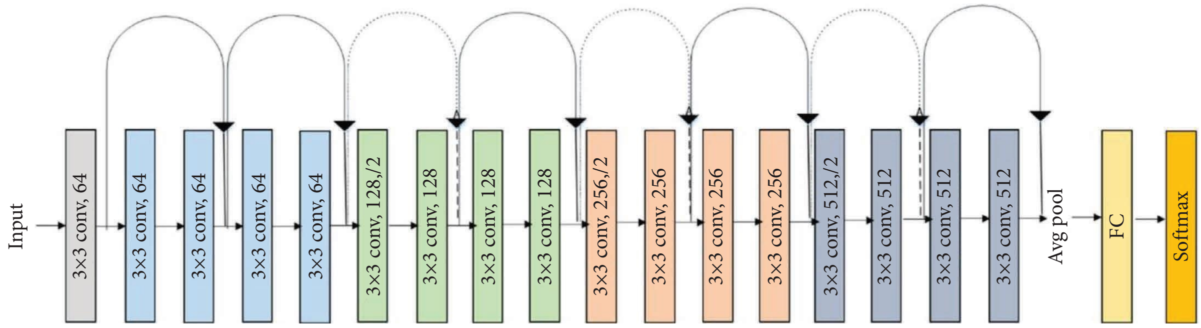


Figure 3: The intuitive illumination of the ResNet-18 architecture.

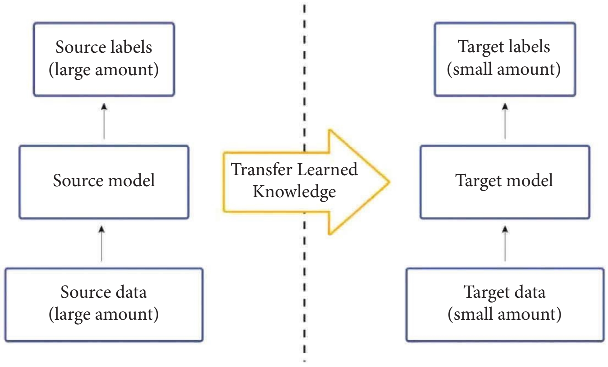


Figure 4: The overall diagram of the pretrained model.

features that are useful for many MV-based tasks, and fine-tuning allows the model to adapt these features to the specific task at hand.

2.3. The Improved Self-Attention Mechanism. However, it is also worth noting that the PPO module used in PSPNet may lead to the problem of context information loss between different subregions to a certain extent. This is mainly because the model only decodes the high-level feature map of cracks with a small resolution, does not fully consider the relationship between pixels, and ignores the impact of the underlying spatial detail features on crack image segmentation. This problem is more common when dealing with high-resolution crack images acquired by UAVs, so it is necessary to study targeted improvement algorithms to improve the detection effect of PSPNet in small and complex background cracks.

Figure 5 shows the overall view of the spatial self-attention mechanism architecture. In semantic segmentation, the self-attention mechanism can be used to extract semantic information of different regions in an image, thereby helping the model to more accurately classify pixels into different semantic categories. In particular, the self-attention mechanism allows the model to focus on pixels at different positions in the image and calculate the similarity between them, thereby generating a pixel-level weighted vector to represent the semantic information of each pixel. It has been proven that the self-attention mechanism has achieved good results in semantic segmentation tasks,

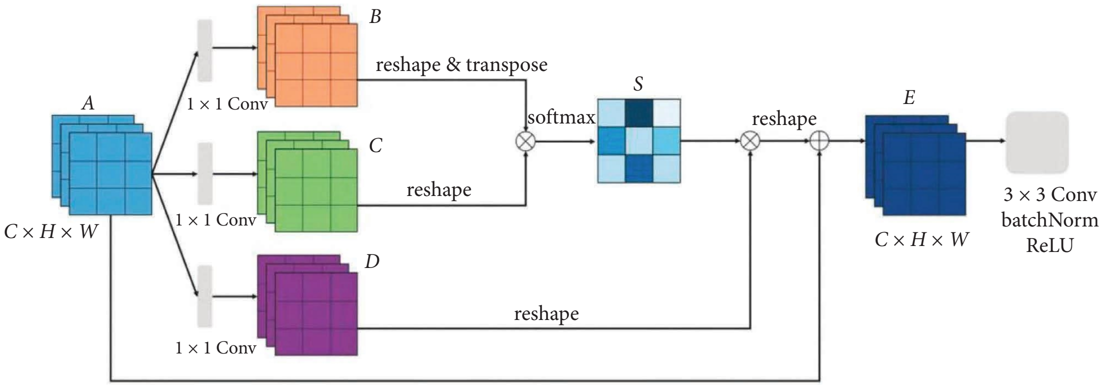


Figure 5: The spatial self-attention mechanism architecture.

especially for image data with complex structures such as bridge concrete cracks with small sizes.

The main computation mechanism of the spatial attention mapping matrix is shown as follows:

$$s_{ji} = \frac{\exp\left(B_i \cdot C_j\right)}{\sum_{i=1}^{N} \exp\left(B_i \cdot C_j\right)}, \tag{2}$$

where B_i represents the i-th element in the matrix B; C_j represents the j-th element in the matrix C; and s_{ji} represents the influence of the feature of the i-th position on the j-th position. Finally, it is added element-wise with the feature map to obtain the output and then goes through a convolutional layer to get the final output:

$$E_j = \alpha \sum_{i=1}^{N} \left(s_{ji} D_i\right) + A_j. \tag{3}$$

2.4. The Combined Loss Function for Model Training. The loss function is the object optimized by the neural network during the training process, which can guide the training direction of the model. The selection of an appropriate loss function can improve the learning performance of the model on cracks. The essence of bridge concrete structure crack segmentation is to perform binary classification on each pixel of the crack image. However, due to the low proportion of crack pixels in the bridge crack image, the ordinary binary classification cross-entropy loss function is used for gradient backpropagation. The samples of each category have the same degree of attention, which is very susceptible to the influence of sample imbalance, which leads to the inhibition of the learning of crack features, making the prediction results more biased towards the image background.

To address these problems, this study designs a hybrid loss function considering negative sample mining based on the improvement of the binary cross-entropy loss function. The hybrid loss function includes the focal loss function and the dice loss function, which can be described as follows:

$$\begin{aligned} L_{\text{hybrid}} &= L_{\text{Focal}} + L_{\text{Dice}}, \\ L_{\text{Focal}} &= -\frac{1}{N} \sum_{i=1}^{N} \left[-x_i \beta \left(1 - y_i\right)^{\gamma} \log\left(y_i\right) + \left(1 - x_i\right)\left(1 - \beta\right) y_i^{\gamma} \log\left(1 - y_i\right)\right], \\ L_{\text{Dice}} &= 1 - \frac{\sum_{i=1}^{N} x_i y_i + \varepsilon}{\sum_{i=1}^{N} x_i + y_i + \varepsilon} - \frac{\sum_{i=1}^{N} \left(1 - x_i\right)\left(1 - y_i\right) + \varepsilon}{\sum_{i=1}^{N} 2 - x_i - y_i + \varepsilon}, \end{aligned} \tag{4}$$

where N represents the total number of pixels; x_i represents the label value of the i-th pixel; and y_i represents the predicted value of the i-th pixel.

The application of the focal loss function, which is used to emphasize the importance of negative samples (cracks), can alleviate the problem of using the dice loss function alone in some extremely unbalanced data. The inability to learn the correct gradient descent direction leads to the difficulty of training. Moreover, introducing a lower proportion of the dice loss function in the focal loss function can further alleviate the sample imbalance problem in the crack image.

3. Case Study

3.1. Project Description and UAV-Based Process. The project is located in the areas south of the Yangtze River, China, and was completed and opened to traffic in August 1999. The total length of this bridge is about 1531 meters with a total width of 17 meters and a total of 56 spans. Due to the rapid development of the economy, the increasing traffic volume, including heavy vehicles, has brought greater pressure on the bridge structures. Therefore, the UAV-based inspection technique was introduced to determine the overall situation of bridge service behavior changes. The UAV used in this study is produced by the DJI company, and its model is the Mavic 3 Pro. The basic information and parameters about the UAV and its equipped sensor equipment are shown in Table 1.

Step 1: Flight test stage. The physical map of the UAV equipment and its corresponding bridge detection scene process are shown in Figure 6. The specific process of UAV bridge crack detection is as follows.

In the process of the UAV-based bridge inspection task, the first step is to use the test flight mode to determine the best detection distance. After several tests, it was finally determined that the image quality collected when the UAV lens was 40 cm away from the crack surface was the best.

Step 2: Adjust to the right position. The UAV was hovered above the crack position and adjusted the position of the gimbal, indicating that the camera lens is parallel to the crack surface.

Step 3: Capture images at a fixed location. The UAV was turned from the hovering state to the steady flight state along the crack area, and pictures are taken every 1s.

Based on the above steps, a total of 85 concrete crack images with a resolution of 5000 ∗ 3000 pixels were obtained. These images contain different types of concrete cracks, including single cracks, intersecting cracks, and network cracks.

3.2. Image Preprocessing and Dataset Production. As the size of cracks in bridge concrete structures is generally relatively small, they occupy a relatively low ratio of crack image pixels. Therefore, this study combines sliding windows and digital image processing techniques to divide superlarge-sized images into small images for easy dataset construction.

Aiming at the characteristics of the uncertain distribution of cracks in the image, the large proportion of length and thinness, and the low proportion of crack pixels in the overall pixels in actual engineering shooting, multiple steps are used to process superlarge-size images to construct a training set. The specific steps are as follows:

Step 1: Initial images of bridge concrete cracks are obtained in structures of oversize through the UAV inspection technique.

Step 2: According to the size of the image, a sliding window with a fixed size is selected and the image is divided into small images according to a certain step size. In this study, the image with a resolution of 6000 ∗ 4000 pixels is divided into small images with a size of 600 ∗ 400 and 1200 ∗ 800 pixels for model training.

Table 1: Main parameters of UAV aerial photography equipment.

Main indicator	Parameter value
Weight	703 g
Size	7.2 × 10 × 3 inch
Flight time	31 min
Maximum ascent speed	5 m/s
Maximum descent speed	4 m/s
Maximum horizontal flight speed	21 m/s
Maximum flight altitude	100 m
Maximum wind resistance rating	4 wind
Maximum transmission distance	7.5 km
Optical camera pixels	12 million

Step 3: All the split images with cracks are kept, and images that only contain the background are also randomly selected and entered into the final training set.

The developed bridge concrete crack dataset contains a total of 2500 crack and background images. These images are divided into training sets, validation sets, and test sets according to the ratio of 6 : 2:2 for model construction and performance evaluation. Figure 7 shows images of cracks in bridge sections captured by UAVs and the corresponding annotation files. Table 2 shows the description of the developed image dataset. It can be seen from the figure and the table that the dataset contains images of cracks with different widths and different background roughnesses, which fully reflects the diversity of cracks in bridge concrete structures in real-world scenarios.

4. Experimental Result and Discussion

4.1. The Model Training Process. All calculation processes of the algorithms involved in this article are carried out on the same image workstation. The main performance parameters of this workstation are shown in Table 3. The proposed and other comparison algorithms are coded using Python and the PyTorch platform.

From Figure 8, it can be seen that as the iteration process increases, the model training loss function gradually and steadily decreases and approaches convergence. Such changes indicate that the application of transfer learning technology can improve the stability and speed of the model convergence process, which can be verified by gradually improving the accuracy of the verification set. After each iteration, the evaluation indicator accuracy is calculated on the validation set, and the model weight coefficient is saved. The selection criterion for the optimal model is the one with the highest crack segmentation effect on the verification set. Finally, the 84th iteration results and the corresponding weight files were retained for model inference testing.

4.2. Ablation Experiment and State-of-the-Art Vision-Based Model Evaluation. A series of ablation comparison experiments are used to evaluate the impact of the attention mechanism and the ResNet backbone network module on model performance. Table 4 shows the ablation experimental results of different modules. It can be inferred that the

(a)

(b)

Figure 6: Carrying out bridge inspection based on UAV detection technology. (a) UAV-based inspection equipment equipped with cameras. (b) UAV-based bridge damage detection.

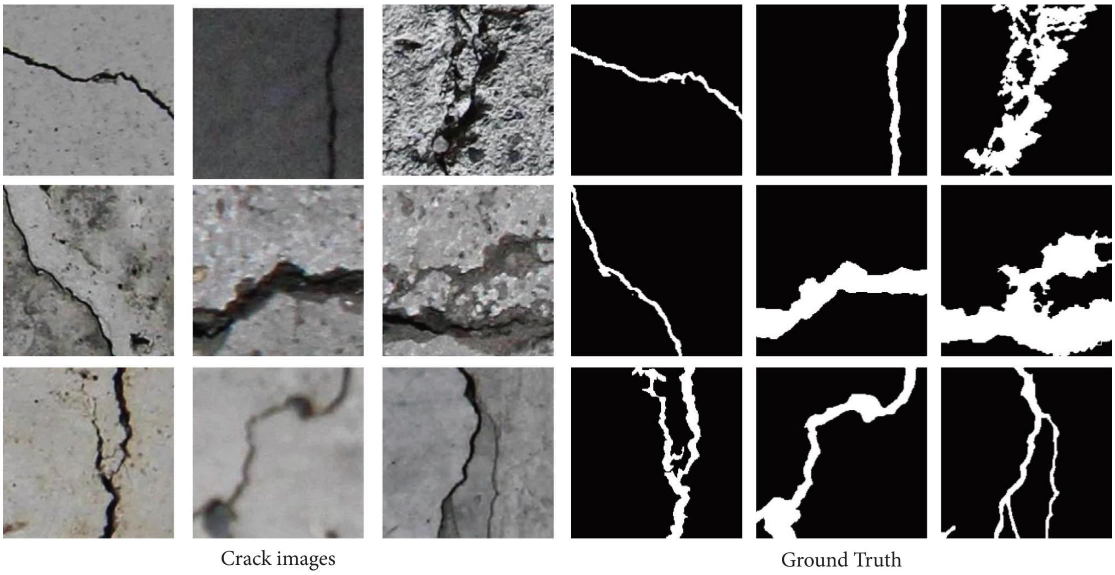


Figure 7: The bridge concrete crack images and corresponding labeling results.

Table 2: Image dataset description of the building model.

Parameter	Training set	Validation set	Test set
Number of images	1500	500	500
Image resolution	600 * 400, 1200 * 800	600 * 400	600 * 400, 6000 * 4000

Table 3: Main hardware parameters of the computing environment.

Parameter	Training set
Central processing unit	2 × Xeon® Gold 5118
Memory	256 GB
Graphics processing unit	1 × Tesla T4
Storage	1 TB SSD

proposed improved self-attention mechanism has an obvious effect on improving the crack segmentation performance. As can be seen from the table, compared with the MobileNet backbone network, which is also designed for lightweight models, the introduction of the residual network can significantly improve the model detection accuracy.

To further verify the applicability of the proposed vision-based crack detection model in UAV detection scenarios, multiple state-of-the-art DL-based concrete crack segmentation networks, including the U-Net [23], the SegNet [24], the DeepLab V3 [25], and the FCN [26], are used as benchmark methods. These detection models are

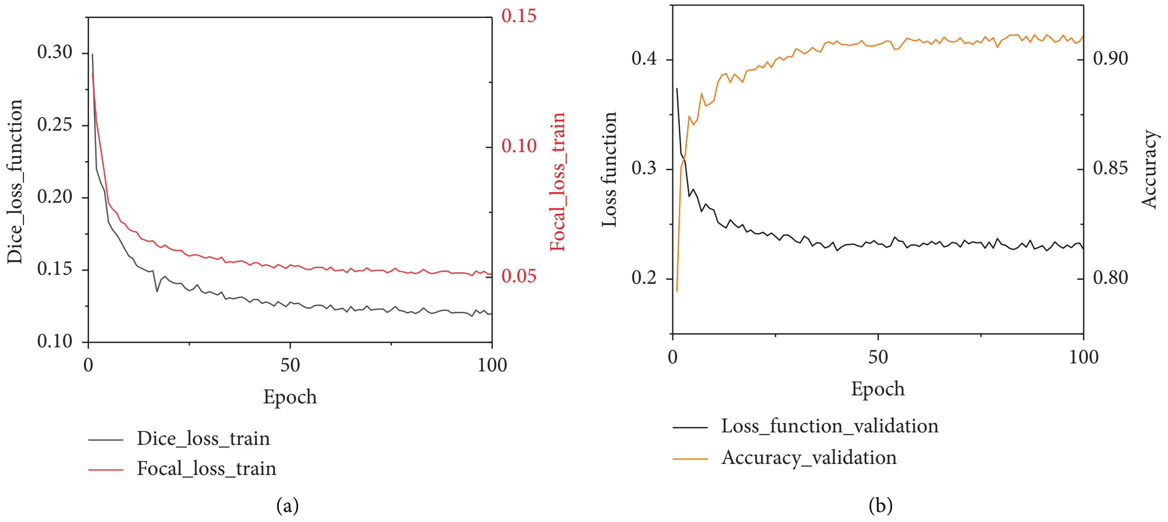


Figure 8: Process line changes during the model training process. (a) Loss function change in the training stage. (b) Loss function change in the validation stage.

Table 4: Analysis of ablation experimental results of different components.

Backbone	With/without the self-attention mechanism	Intersection over union (IOU)	Accuracy
ResNet-18	w/	0.9012	0.8820
ResNet-18	w/o	0.8830	0.8626
MobileNet	w/	0.8771	0.8532
MobileNet	w/o	0.8636	0.8411

trained with the same input data and the same number of iterations as the construction method.

The number and scale of model parameters and the time required to train the effective model are important links to evaluate the applicability of the model in the actual UAV detection scene. Table 5 demonstrates the parameter scale comparison of the constructed model and other state-of-the-art visual crack detection models. It can be seen from the table that the number of parameters of the model is much smaller than that of other advanced algorithms, indicating that it is a lightweight detection algorithm, which is suitable for UAV-based edge computing scenarios. In addition, the training time of the model is also significantly faster than that of other advanced comparison algorithms, indicating that it has efficient modeling ability. At the same time, the training time of the construction method is also significantly lower than that of other benchmark methods, which is mainly due to the application of the transfer learning strategy, which reduces the cost of model training. At the same time, the training time of the developed method is also significantly lower than that of other benchmark methods, which is mainly due to the application of the transfer learning strategy, which reduces the cost of model training.

The evaluation comparison of identification results of different crack identification methods is shown in Table 6 and Figure 9. It can be inferred from the figure and table that the proposed method achieves better recognition results than other methods in terms of intuitive visual comparison and various evaluation indicators. In addition, it can also be seen from the table that the proposed method has achieved an inference efficiency of more than 30 FPS, indicating that it has real-time detection capabilities. This is mainly due to the introduction of lightweight convolutional networks, which effectively reduce the number of model calculation parameters and improve model calculation efficiency.

4.3. Complicated Environment Validation. To further verify the applicability of the constructed crack detection algorithm in different scenes, four different scenarios including dark light, diversity of cracks, high roughness, and complex background are selected as the test environment. The evaluation benchmark is the comparison of the results of manual judgment based on the experience of experts with rich experience in bridge inspection.

4.3.1. Low-Light Conditions. Figure 10 demonstrates the application of the constructed method to the crack identification results under UAV-based dark and light conditions. It can be seen from the figure that the identification effect of building the vision-based crack detection model is similar to the result of manual annotation. Even for those scenes with extremely dark lighting conditions and severely insufficient light, the proposed method can still accurately identify and segment the geometry of concrete cracks, which shows that the method has strong generalization and adaptability. This

Table 5: Model training and model building parameter comparison.

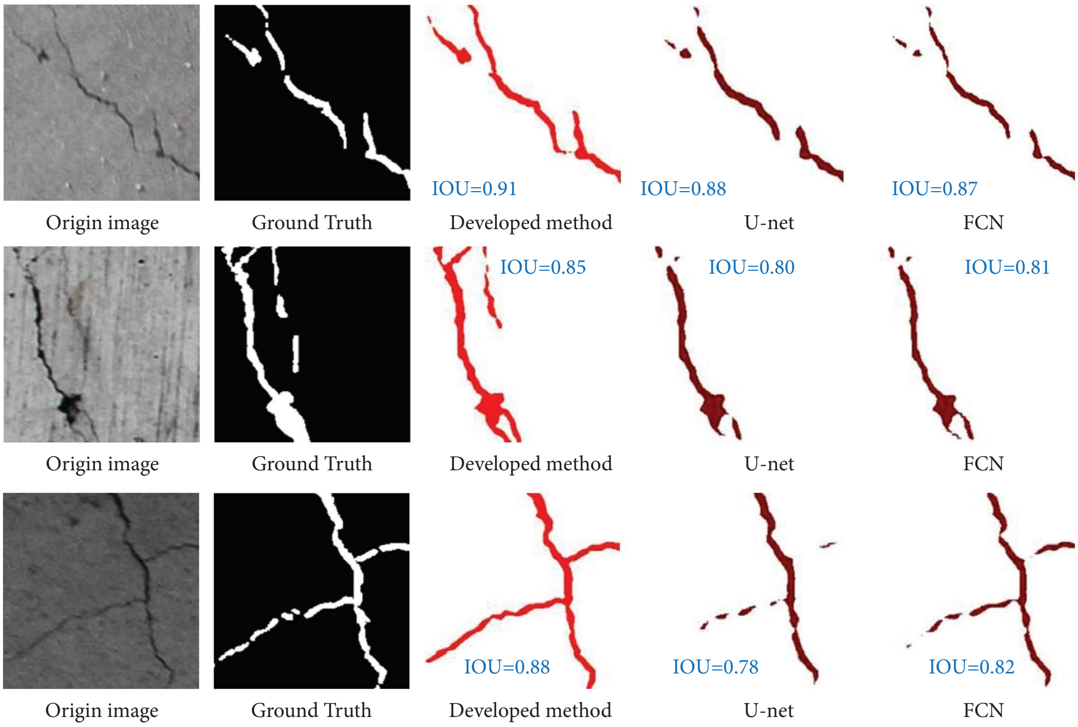
Models	Total model size (MB)	Training speed (h)
The developed method	54.221	1.2
Model 1	97.248	2.1
Model 2	105.31	2.3
Model 3	101.052	2.2
Model 4	75.252	1.85
Model 5	112.250	2.52
Model 6	120.454	2.72

Note. Model 1–Model 6 denote U-Net, U-Net++, FCN, SegNet, DeepLab v3, and DeepLab v3+, respectively.

Table 6: Evaluation of crack identification effects.

Models	Precision (%)	Recall (%)	Accuracy (%)	IOU (%)	Speed (FPS)
Proposed	0.9008	0.8750	0.8820	0.9012	35.04
Model 1	0.8046	0.8485	0.8680	0.8709	27.32
Model 2	0.8124	0.8520	0.8863	0.8821	26.36
Model 3	0.8105	0.8472	0.8735	0.8624	26.25
Model 4	0.8216	0.8025	0.8132	0.8327	28.62
Model 5	0.8536	0.8432	0.8238	0.8526	20.26
Model 6	0.8631	0.8573	0.8386	0.8635	21.31

Note. Model 1–Model 6 denote U-Net, U-Net++, FCN, SegNet, DeepLab v3, and DeepLab v3+, respectively.



Origin image Ground Truth Developed method U-net FCN

Origin image Ground Truth Developed method U-net FCN

Origin image Ground Truth Developed method U-net FCN

Figure 9: Comparison of different test results.

shows that the machine vision-based crack identification method has better environmental adaptability than manual detection.

4.3.2. *Diverse Concrete Cracks.* Figure 11 demonstrates the application of the constructed method to the crack identification results under the diverse concrete crack scene. It can be seen from the picture that there are significant differences in cracks in bridge concrete structures, including both wide cracks and small cracks with narrow widths. The difference in cracks in bridge concrete images will bring some challenges and difficulties to the vision-based detection algorithm. In fact, from the results, the construction of the vision-based crack detection model can effectively identify

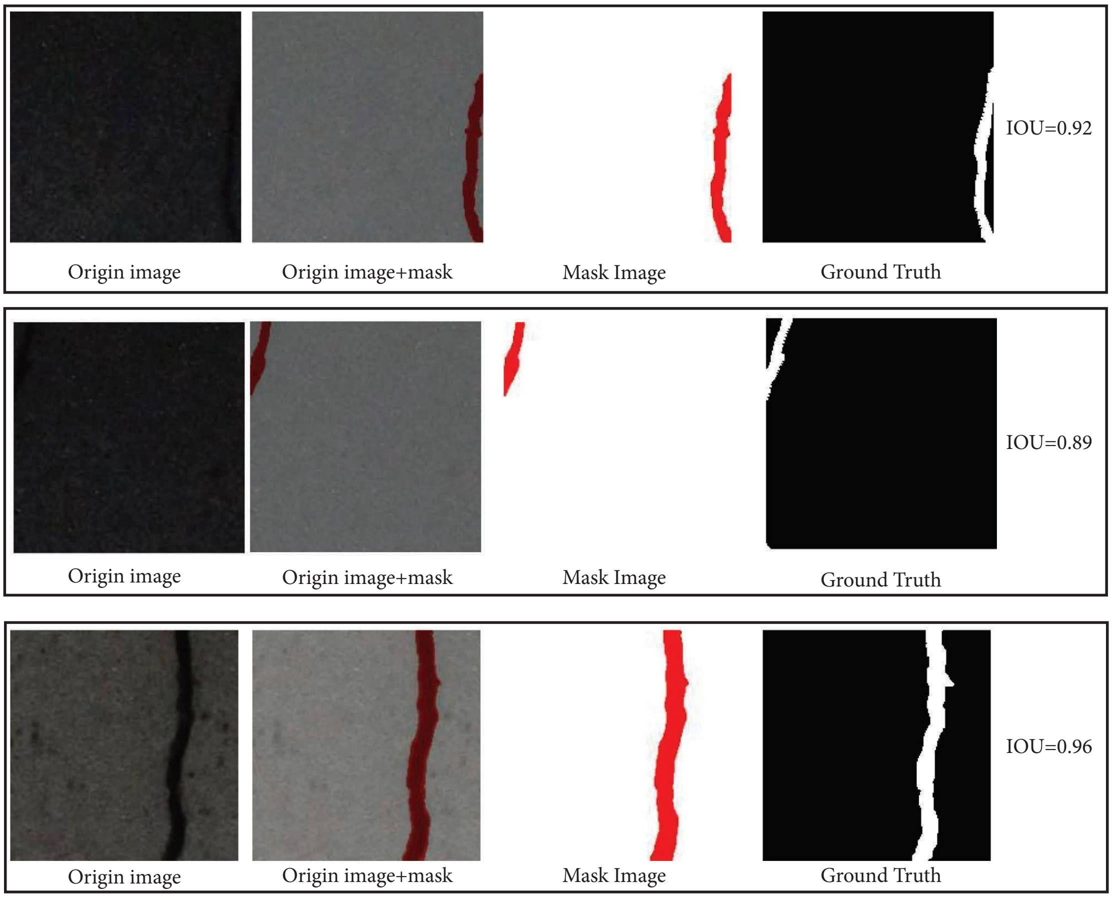


Figure 10: Evaluation of model detection results in the dark lighting scene.

and segment different types of bridge concrete cracks, and the identification results are consistent with the results of manual labeling, indicating that it has strong robustness.

4.3.3. High Roughness Background. Figure 12 demonstrates the application of the constructed method to the bridge concrete crack identification results under the high roughness background. It can be seen from the figure that the recognition effect of the vision-based crack detection model is similar to that of manual labeling. Even for scenes with different background roughness, this method can accurately identify and segment the geometric shape of concrete cracks, which shows that the method has strong generalization and adaptability.

4.3.4. High-Complexity Background. Figure 13 shows the bridge concrete crack identification effect of the developed method under a high-complexity background. It can be inferred from the results that the constructed model can accurately identify bridge concrete cracks even in the presence of significant noise contamination, including paint and stain interference. In addition, compared with the manual identification results, the crack identification results of the constructed method are similar, which can effectively replace the traditional manual detection methods.

4.4. Model Verification in Large-Scale Bridge Crack Images. To further verify the effectiveness of the proposed crack detection model in the actual UAV aerial photography inspection process, high-resolution bridge crack images were used as model input, and the sliding window method was used to verify the model recognition ability. It can be inferred from Figure 14 that the proposed crack segmentation method can achieve effective segmentation of tiny cracks in high-resolution bridge concrete structures. This proves that the proposed method has strong feature extraction performance and has broad application prospects in the fine diagnosis of cracks in bridge concrete structures in the future.

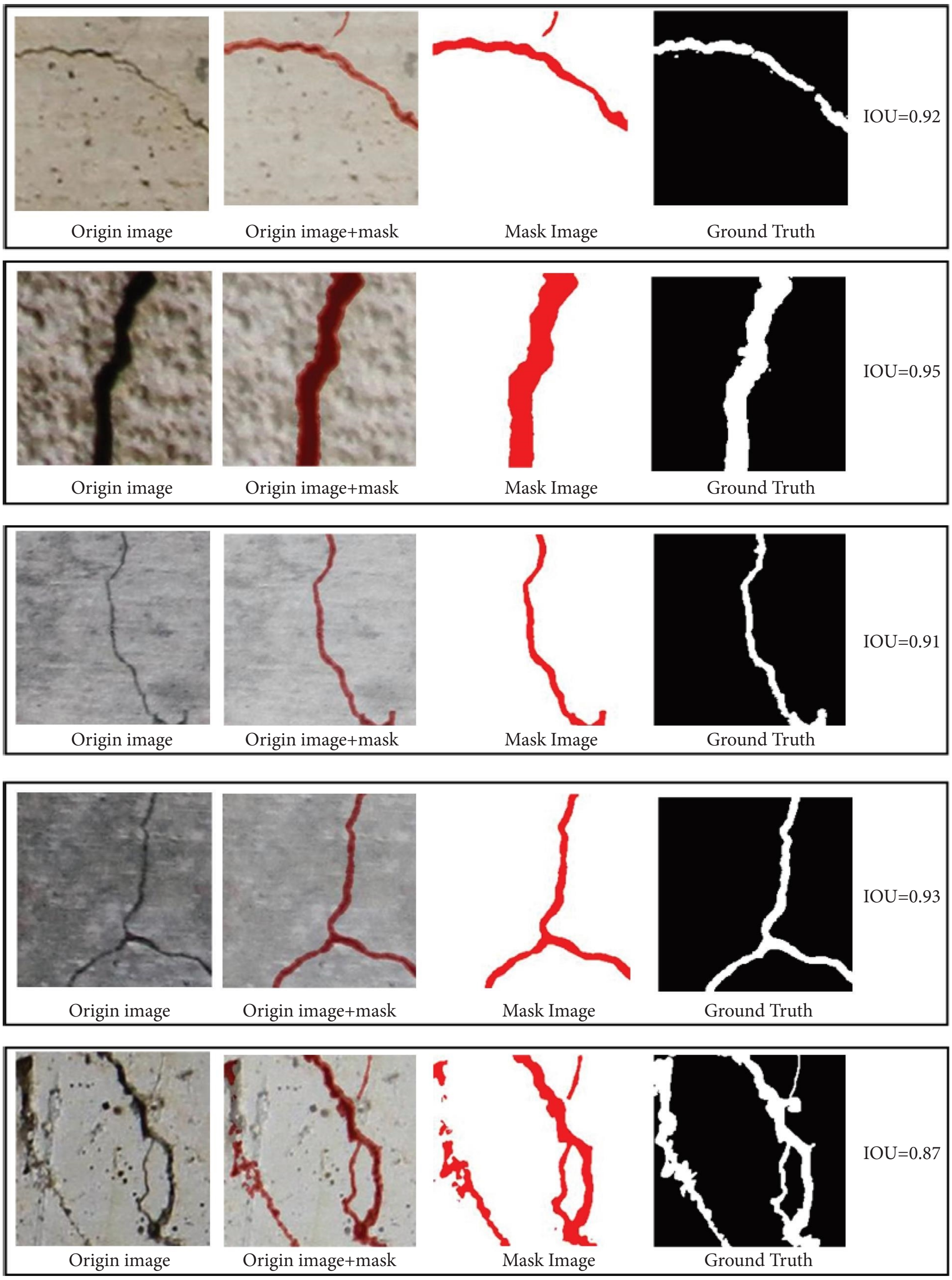


Figure 11: Evaluation of model detection results in the diverse crack scene.

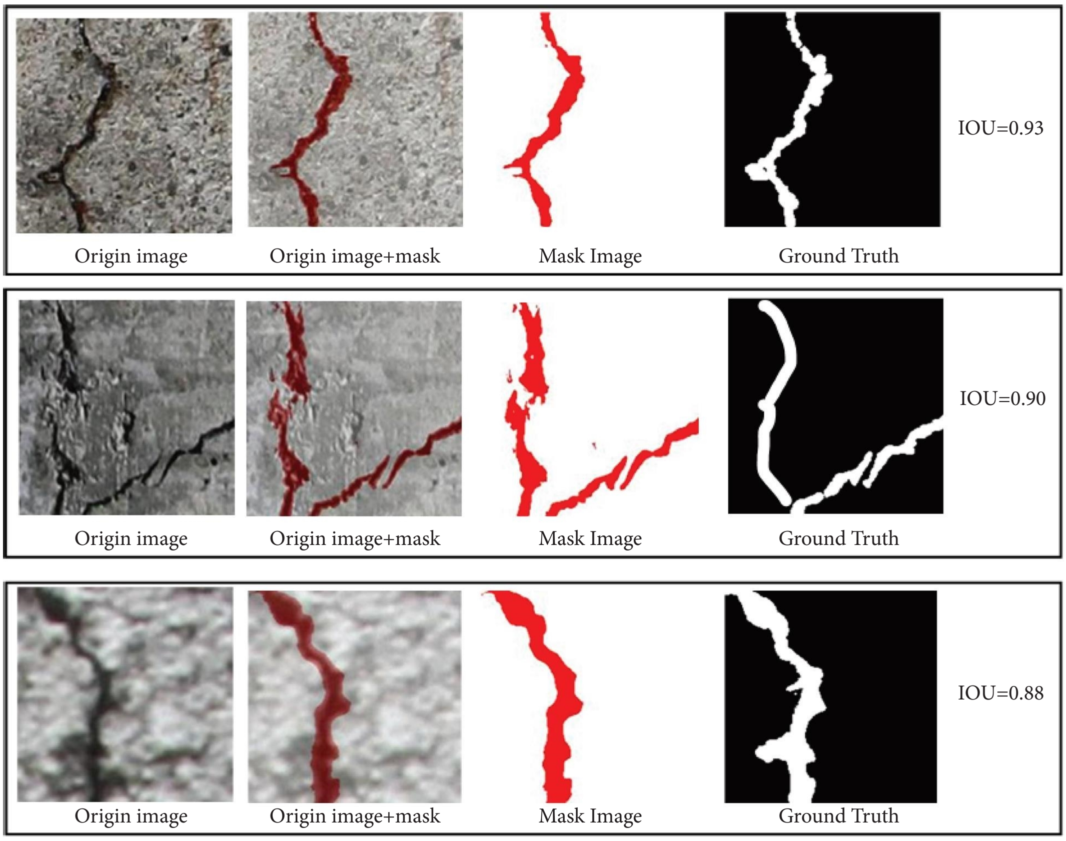


Figure 12: Evaluation of model detection results in the high robustness scene.

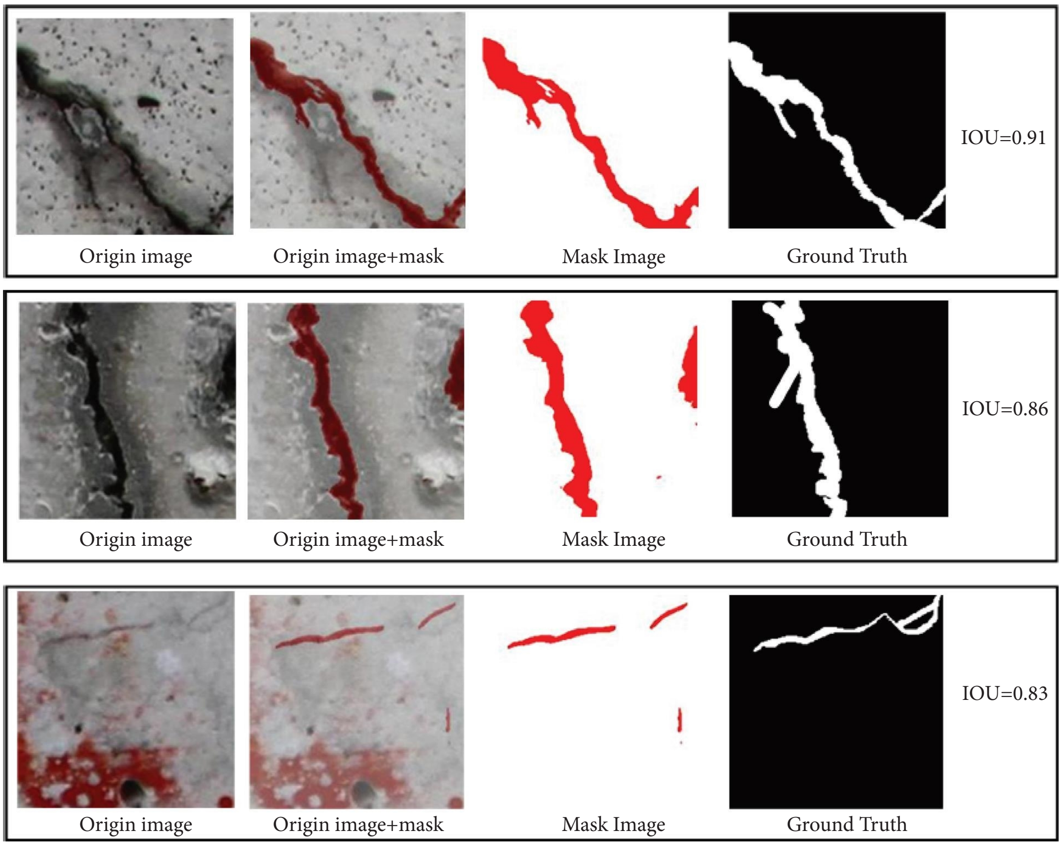


Figure 13: Evaluation of model detection results in the high complexity scene.

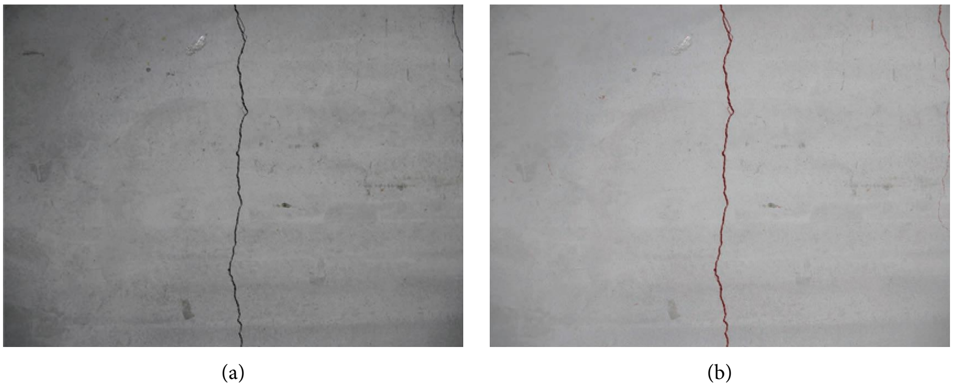


Figure 14: The constructed method applied to full-scale bridge crack image detection. (a) Original image. (b) Crack segmentation result.

5. Conclusions and Future Discussion

5.1. Conclusions. As a common kind of structural defect of bridges, cracks are a sign that the bridge structure has reached the limit of its bearing capacity, which seriously affects the safe operation of the bridge. Manual detection methods have the disadvantages of high risk, large errors, and low efficiency in the inspection of bridge cracks. Digital image processing technology has been proven as an effective tool, but it also has problems in terms of detection efficiency and complex scene generalization ability.

To solve these problems, this study combines DL and machine vision technology to propose an automatic concrete crack identification method suitable for UAV bridge detection scenarios. The significant advantage of this method is to insert the spatial position module based on the self-attention mechanism into the decoding-encoding semantic segmentation network. This method can realize the accurate identification of subtle and complex background cracks and efficiently infer concrete crack images while having real-time and high detection efficiency. The key contributions of this study are as follows.

The specific contributions of this study are as follows:

(a) The combination of the PSPNet lightweight network and spatial location module based on the self-attention mechanism can improve the ability of the model to obtain rich contextual information in the spatial dimension and then realize the effective identification of diverse cracks in bridges in complex scenes.

(b) Quantitative evaluation results show that the constructed bridge concrete crack segmentation model achieves the evaluation indicators of 0.9008 precision, 0.8750 recall, 0.8820 accuracy, and 0.9012 IOU on the bridge concrete crack dataset, which are significantly higher than other state-of-the-art baseline methods.

(c) The effectiveness of the constructed method has been validated in four common UAV-based bridge detection scenarios, including low light, complex crack forms, high background roughness, and complex background scenes, which are used to further test the crack detection ability of the developed crack identification model.

5.2. Limitations and Future Discussion. However, this study also has some limitations, which need to be further explained. First of all, the focus of this research is to study the bridge concrete crack identification algorithm suitable for UAV detection. Future research should focus on the physical size measurement and risk assessment of bridge concrete cracks. Moreover, apart from cracks, bridge concrete structures also have multicategory defects such as spalling, voids, depressions, and pockmarks. Therefore, the identification method of multicategory defects is studied based on visual detection algorithms. In addition, the combination of UAV and 3D reconstruction technology to construct a digital twin scene of the bridge is also the focus of future research.

Note that they are usually at a certain distance from the bridge deck, resulting in large-scale UAV images when UAVs carry out bridge inspections. However, the annotation process usually uses the original high-resolution bridge defect images as the target. However, during the manuscript presentation process, we only selected local crack images and their annotation results for visual display, which inevitably caused the manual annotation results to appear rougher and wider than the real results. The annotation of real images will affect the accuracy of the crack recognition model built based on it. In subsequent research, the authors will pay more attention to carrying out image annotation work based on local defect images to ensure the accuracy of the annotation results.

Data Availability

The data presented in this study are available from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (grant no. 62102184) and in part by the Natural Science Foundation of Jiangsu Province (grant no. BK20200784).

References

[1] H. N. Li, L. Ren, Z. G. Jia, T. H. Yi, and D. S. Li, “State-of-the-art in structural health monitoring of large and complex civil infrastructures,” *Journal of Civil Structural Health Monitoring*, vol. 6, no. 1, pp. 3–16, 2016.

[2] B. F. Spencer, V. Hoskere, and Y. Narazaki, “Advances in computer vision-based civil infrastructure inspection and monitoring,” *Engineering*, vol. 5, no. 2, pp. 199–222, 2019.

[3] Y. Deng, H. Ju, W. Zhai, A. Li, and Y. Ding, “Correlation model of deflection, vehicle load, and temperature for in-service bridge using deep learning and structural health monitoring,” *Structural Control and Health Monitoring*, vol. 29, no. 12, pp. 1–20, 2022.

[4] S. Yoon, B. F. Spencer, S. Lee, H. Jung, and I. Kim, “A novel approach to assess the seismic performance of deteriorated bridge structures by employing UAV-based damage detection,” *Structural Control and Health Monitoring*, vol. 29, no. 7, Article ID e2964, 2022.

[5] Y. Hou, H. Shi, N. Chen, Z. Liu, H. Wei, and Q. Han, “Vision image monitoring on transportation infrastructures: a lightweight transfer learning approach,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 11, pp. 12888–12899, 2023.

[6] J. Mao, H. Wang, and B. F. Spencer, “Toward data anomaly detection for automated structural health monitoring: Exploiting generative adversarial nets and autoencoders,” *Structural Health Monitoring*, vol. 20, no. 4, pp. 1609–1626, 2020.

[7] G. Hou, L. Li, Z. Xu, Q. Chen, Y. Liu, and X. Mu, “A visual management system for structural health monitoring based on web-BIM and dynamic multi-source monitoring data-driven,” *Arabian Journal for Science and Engineering*, vol. 47, no. 4, pp. 4731–4748, 2022.

[8] D. Liu, J. Chen, D. Hu, and Z. Zhang, “Dynamic BIM-augmented UAV safety inspection for water diversion project,” *Computers in Industry*, vol. 108, pp. 163–177, 2019.

[9] X. Peng, X. Zhong, C. Zhao, A. Chen, and T. Zhang, “A UAV-based machine vision method for bridge crack recognition and width quantification through hybrid feature learning,” *Construction and Building Materials*, vol. 299, Article ID 123896, 2021.

[10] M. Liu, X. Wang, A. Zhou, X. Fu, Y. Ma, and C. Piao, “Uav-yolo: small object detection on unmanned aerial vehicle perspective,” *Sensors (Switzerland)*, vol. 20, no. 8, pp. 2238–2312, 2020.

[11] K. Dunphy, A. Sadhu, and J. Wang, “Multiclass damage detection in concrete structures using a transfer learning-based generative adversarial networks,” *Structural Control and Health Monitoring*, vol. 29, no. 11, pp. 1–20, 2022.

[12] X. Pan, S. Tavasoli, and T. Y. Yang, “Autonomous 3D vision-based bolt loosening assessment using micro aerial vehicles,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 38, no. 17, pp. 2443–2454, 2023.

[13] D. C. Feng, Z. T. Liu, X. D. Wang, Z. M. Jiang, and S. X. Liang, “Failure mode classification and bearing capacity prediction for reinforced concrete columns based on ensemble machine learning algorithm,” *Advanced Engineering Informatics*, vol. 45, no. February 2019, Article ID 101126, 2020.

[14] C. Y. Liu and J. S. Chou, “Bayesian-optimized deep learning model to segment deterioration patterns underneath bridge decks photographed by unmanned aerial vehicle,” *Automation in Construction*, vol. 146, no. December 2022, Article ID 104666, 2023.

[15] H. Kim, J. Yoon, and S. H. Sim, “Automated bridge component recognition from point clouds using deep learning,” *Structural Control and Health Monitoring*, vol. 27, no. 9, pp. 1–13, 2020.

[16] Y. Zhou, Y. Pei, Z. Li, L. Fang, Y. Zhao, and W. Yi, “Vehicle weight identification system for spatiotemporal load distribution on bridges based on non-contact machine vision technology and deep learning algorithms,” *Measurement*, vol. 159, Article ID 107801, 2020.

[17] T. Jiang, G. T. Frøseth, and A. Rønnquist, “A robust bridge rivet identification method using deep learning and computer vision,” *Engineering Structures*, vol. 283, no. December 2022, Article ID 115809, 2023.

[18] M. Żarski, B. Wójcik, and K. Książek, “Finicky transfer learning-A method of pruning convolutional neural networks for cracks classification on edge devices,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 37, no. 4, pp. 500–515, 2022.

[19] F. Song, Y. Sun, and G. Yuan, “Autonomous identification of bridge concrete cracks using unmanned aircraft images and improved lightweight deep convolutional networks,” *Structural Control and Health Monitoring*, vol. 2024, pp. 1–15, 2024.

[20] H. Bae, K. Jang, and Y. K. An, “Deep super resolution crack network (SrcNet) for improving computer vision–based automated crack detectability in in situ bridges,” *Structural Health Monitoring*, vol. 20, no. 4, pp. 1428–1442, 2021.

[21] D. Zhang, L. Fu, H. Huang, H. Wu, and G. Li, “Deep learning-based automatic detection of muck types for earth pressure balance shield tunneling in soft ground,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 38, no. 7, pp. 940–955, 2022.

[22] F. Guo, Y. Qian, J. Liu, and H. Yu, “Pavement crack detection based on transformer network,” *Automation in Construction*, vol. 145, no. October 2022, Article ID 104646, 2023.

[23] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, “Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 568–576, 2020.

[24] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: a deep convolutional encoder-decoder architecture for image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[25] C. Liu, L. C. Chen, F. Schroff, H. Adam, W. Hua, and A. L. Yuille, “Auto-deeplab: hierarchical neural architecture search for semantic image segmentation,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, January 2019.

[26] L. Yan, D. Liu, Q. Xiang et al., “PSP net-based automatic segmentation network model for prostate magnetic resonance imaging,” *Computer Methods and Programs in Biomedicine*, vol. 207, Article ID 106211, 2021.