Measurement 237 (2024) 115197

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Automated intelligent measurement of cracks on bridge piers using a ring-climbing vision scanning operation robot

Hao Du [a], Huifeng Wang [a,b,*], Xiaowei Zhang [a], Haonan Peng [a], Rong Gao [a], Xueyan Zheng [a], Yaxiong Tong [a], Yuanhe Shan [a], Zefeng Pan [a], He Huang [a]

[a] School of Electronic & Control Engineering Chang'an University, Xi'an 710064, China
[b] Road Traffic Intelligent Detection and Equipment Technology Research Center of Shaanxi, Xi'an 710064, China

ARTICLE INFO

Keywords:
Crack segmentation
Ring-climbing robot
Deep learning
Efficient self-attention

ABSTRACT

Detecting cracks on the surface of bridge piers is critical to the health condition of bridges. Aiming at the limitations of existing unmanned aerial vehicles (UAVs) and climbing robots, this research proposes a ring-climbing visual scanning operation robotic disease acquisition system, which achieves entire region, high-definition, and fast acquisition of cracks on bridge piers surfaces. In addition, a crack segmentation network based on an efficient self-attention mechanism and a refined segmentation algorithm are proposed to achieve accurate detection and quantification of cracks on bridge piers. Finally, an experimental prototype was built, and experiments were conducted. The results show that the maximum error of crack quantification is within 0.1 mm. The designed ring-climbing visual scanning operation robotic disease acquisition system can be applied to the surface of any pier with a 1.2 to 1.5 m diameter, which is of great significance for the automatic detection of the health condition of bridge piers.

1. Introduction

Bridges play a vital role in infrastructure, supporting the country's economic progress. The quality of bridges has a direct impact on both economic growth and public safety[1,2]. As traffic volume and service time increase, bridges' performance and security are put at risk. The health of the bridge piers is crucial to the bridges' overall structural safety and service life. Damage on the surface of bridge piers can significantly impact their structural integrity. Therefore, detecting cracks has always been crucial to managing bridge safety[3–5]. By detecting and quantifying surface cracks, the bridge piers' condition can be assessed promptly, establishing a scientific foundation for maintenance and administration[6,7]. Therefore, conducting comprehensive, accurate, and efficient automated crack detection on bridge piers surfaces is crucial to prolonging bridge service life and enhancing economic benefits.

Automated acquisition of images of surface cracks on bridge piers primarily depends on UAVs. Ding et al. [8] utilized a UAV system to acquire images of concrete cracks and calibrate the gimbal camera. Consequently, accurate positioning and quantitative measurement of concrete cracks were achieved. He et al. [9] employed a UAV bridge inspection program with a high-definition zoom camera, laser rangefinder, and global positioning system to capture images of bridge diseases. This method provided a novel approach to bridge inspection. Wang et al. [10] proposed a system for detecting cracks in bridges that utilize a tethered creeping UAV and an image processing method, resulting in high measurement accuracy of the crack parameters. Recently, there has been an increased use of wall-climbing robots in bridge inspections. Mahmoud et al. [11] designed an omnidirectional magnetic wheeled climbing robot that safely and smoothly traversed a steel cable-stayed bridge. Lan et al. [12] developed a vacuum adsorption wall-climbing robot that achieved attachment to the surface of a bridge piers and real-time information collection. Jang et al. [13] proposed a robotic system for ring-climbing inspection operations using multiple cameras to perform image acquisition. Although the entire region climbing the bridge piers can be realized, if the diameter of the bridge piers to be measured is large, the number of cameras required to ensure the entire region acquisition of the image will increase, causing a rise in the cost and the extension of the installation time of the equipment.

Accurately extracting the geometric features of surface cracks on a bridge pier, captured through imaging, is vital for its health assessment and maintenance. Two main methods are used for this purpose: image

[*] Corresponding author at: School of Electronic & Control Engineering Chang'an University, Xi'an 710064, China.
E-mail address: fwang@chd.edu.cn (H. Wang).

https://doi.org/10.1016/j.measurement.2024.115197
Received 21 February 2024; Received in revised form 9 June 2024; Accepted 26 June 2024
Available online 26 June 2024

processing and deep learning. Image processing involves manually designing feature extraction algorithms to capture the cracks' geometric features[14]. Ma [15]extracted the crack boundary using morpho-geometric and texture features. The cracks' maximum width was calculated via its skeleton line and normal. Liu et al. [16] introduced an algorithm for crack extraction with adaptive thresholding, automatically selecting window sizes and thresholds based on specific metrics, effectively removing noise and extracting the cracks. However, there are limitations to using image processing for crack extraction. If the background has a similar color or texture to the cracks, the image processing method may not accurately extract the crack from the complex background. Crack segmentation based on deep learning has emerged as a popular research topic in recent years, as deep learning can automatically learn features and address complex nonlinear problems. Li et al. [17] utilized a generative adversarial network to create fractures in bridges amidst intricate backgrounds and utilized an advanced DenseNet to differentiate fractures with precision surpassing 93 % precisely. Wang et al. [18] proposed a model for segmenting fractures based on the Vision Transformer and achieved the mIoU of 92.63 %, which provided good segmentation results.

Although the UAVs exhibit strong maneuverability and can comprehensively capture images of bridge piers through a predefined flight path, acquiring images of the bridge piers' surface at a close distance is not feasible due to safety concerns. When capturing images, the camera distance is adjusted, which hinders the precise assessment of micro-scale cracks in high-definition images. The navigation and obstacle avoidance systems must be supplemented to ensure consistent functioning. Even though existing wall-climbing robots can capture close-range images of surface defects, they rely mainly on manual robot motion control or path planning during inspections. Additionally, most wall-climbing robots have a slow crawling speed and prove challenging to move laterally. Collecting global images of the bridge piers surface requires resources and human resources, which can result in inadequate battery power for the wall-climbing robot and potential safety hazards.

Furthermore, although deep learning has demonstrated high accuracy in crack detection tasks, many researchers have employed convolutional neural networks (CNNs) as their network models [19]. However, when presented with large images, CNNs may face challenges in capturing global information and receptive field[20]. Finally, the current datasets of bridge piers cracks display significant inconsistency concerning the distances from the camera to the bridge piers surface during image capture, resulting in varied pixel equivalents for crack images. Therefore, most researchers have yet to perform further accurate quantification of the cracks identified by the neural network.

Aiming at the problems of existing UAVs and climbing robots, which are time-consuming, costly and challenging to achieve entire region detection, this research proposes a disease acquisition system using a novel ring-climbing visual scanning operation robot. The robot system will capture images of the bridge piers through automated climbing and a visual scanning vehicle, allowing for a comprehensive view of any damage present on the surface. The detection accuracy is 0.1 mm, and it takes only about 40 min to detect the entire region on a 10 m high bridge pier. This technology applies to columnar bridge piers with a 1.2 to 1.5 m diameter. It achieves the entire region, high-definition, fast and automated acquisition of images of the bridge piers. To address the issue of insufficient fine segmentation and quantification of cracks identified by the neural network, a dataset consisting of surface cracks of bridge piers with uniform pixel equivalence was built. A method that combines the crack segmentation network with a refined segmentation algorithm has also been introduced to quantify the cracks. This approach has resulted in accurately detecting and quantifying surface cracks in bridge piers with a quantification error of within 0.1 mm. This study is significant for automating surface crack detection and assessment on bridge piers. The main contributions of this research are outlined below:

(1) The proposed robotic system is easy to assemble and economical. Using only one camera, it can achieve the entire region, high-definition, fast and automated acquisition of surface disease images of any bridge pier with a 1.2 to 1.5 m diameter.
(2) A dataset of bridge piers surface cracks with uniform pixel equivalence in complex backgrounds is contributed.
(3) The proposed crack segmentation network and refined segmentation algorithm can make the quantization error of crack width within 0.1 mm.

2. Ring-climbing visual scanning operation robot

In this research, a ring-climbing visual scanning robot system is proposed for the automated acquisition of bridge piers surface diseases. Fig. 1 illustrates the system's schematic diagram.

The upper computer control unit establishes wireless communication using Lora with the robot motion control unit. The robot climbs and descends steadily on the bridge piers under control. Simultaneously, it wirelessly communicates with the circular vision scanning image acquisition unit to govern the image acquisition and storage. Ultimately, it achieves the automation of the bridge piers surface climbing and the acquisition of disease images. The overall structure of the ring-climbing robot system is shown in Fig. 2.

The robot's body frame adopts a six-section divided structure for simple transportation and installation. The connector links and bolts fix the adjacent spilt frame in place. The diameter of the body frame measures 2160 mm. By regulating the elasticity of the preload spring on the bridge pier s' surface, the robot performs a climbing of a 1.2 to 1.5 m diameter column-shaped bridge pier. The robot's motion control unit consists of the robot's body frame, battery, gyroscope, and other devices. After receiving the climbing command from the upper computer, the motion attitude of the robot is determined by the FPGA in the control unit through signals from six gyroscopes. The control unit then assigns different running speeds to each motor, ultimately achieving safe and steady climbing of the robot. The robot motion control unit is shown in Fig. 3.

The acquisition of high-definition surface images of the bridge piers' entire region in an automated and efficient manner is made possible by installing a circular rail and a visual scanning vehicle on the robot frame. The visual scanning vehicle primarily comprises an FPGA, CCD camera, motor, and other devices. Upon receiving the image acquisition command from the upper computer, the FPGA drives the vehicle's bottom wheel, allowing it to move along the rail. The encoder determines the vehicle's current position, triggering the camera to acquire images of diseases when it reaches the designated point. At this moment, the LED lights on the light board are activated to supplement the lighting of the recorded images, and all the images are saved directly onto the industrial computer. The visual scanning vehicle is illustrated in Fig. 4.

3. Segmentation and quantification of cracks

This research proposes an accurate and lightweight Crack-Segmentor model based on an efficient self-attention mechanism [21] to compensate for the convolutional neural network's insufficient receptive field and perform better with fewer parameters than other models.

It is worth noting that many researchers currently rely on deep learning to detect cracks in binary images before quantifying them. Since manual dataset labeling can result in errors and the cracks on the surface of the bridge piers are minuscule, multiple interfering factors can cause errors between the natural cracks and the network recognition results. Consequently, directly employing the results predicted by the network model for quantification could lead to inaccuracies. We propose a crack refined segmentation algorithm using the ROI threshold segmentation method to reduce errors, which allows for more precise binary imaging of the cracks.

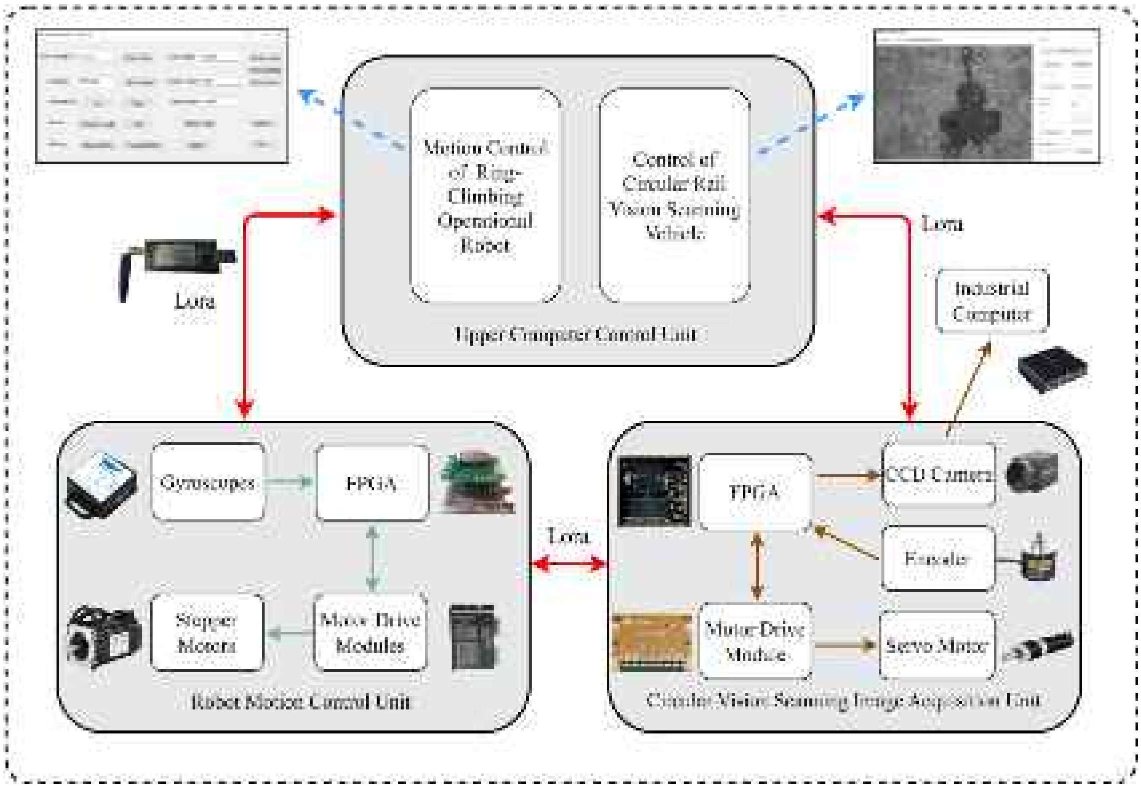

Motion Control of Ring-Climbing Operational Robot
Control of Circular Rail Vision Scanning Vehicle
Upper Computer Control Unit
Lora
Lora
Industrial Computer
Gyroscopes
FPGA
Stepper Motors
Motor Drive Modules
Robot Motion Control Unit
Lora
FPGA
CCD Camera
Encoder
Motor Drive Module
Servo Motor
Circular Vision Scanning Image Acquisition Unit

Fig. 1. Robot system schematic diagram.

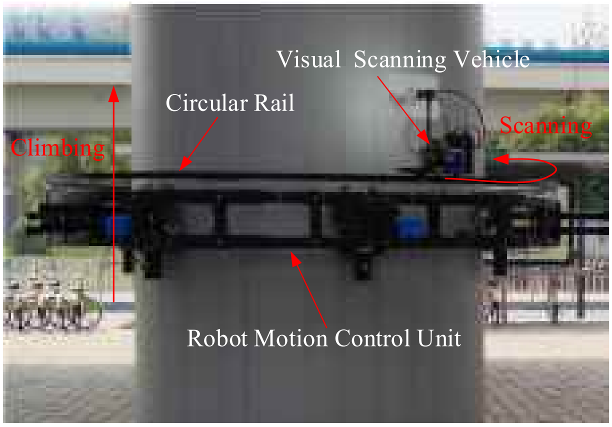


Fig. 2. Robot system overall structure.

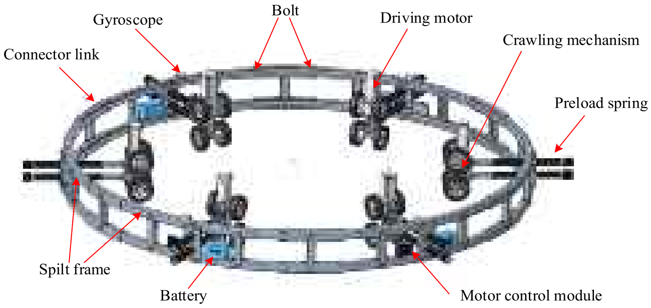


Fig. 3. Robot motion control unit.

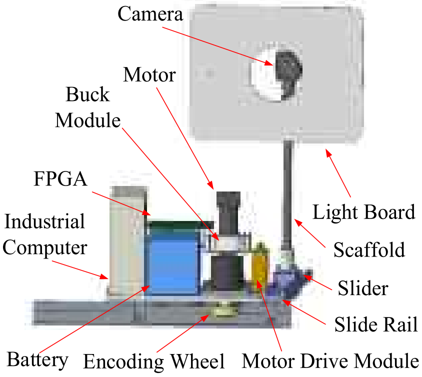


Fig. 4. Visual scanning vehicle.

3.1. Crack-Segmentor model

The Crack-Segmentor network structure is depicted in Fig. 5. The model input is a 512 × 512 RGB crack image. In the encoder, the image is first blocked by Overlap Patch Embedding (OPE) [22]and converted into 2D vectors. Then, a Transformer Block based on an efficient self-attention mechanism is used to extract the cracks feature maps. Overlap Patch merging (OPM) [23] is used to down-sample the feature maps and obtain multilayer feature maps for subsequent feature fusion. After the input image passes through the encoder, four high-resolution low-level feature maps and low-resolution high-level feature maps are obtained with resolutions of 1/4, 1/8, 1/16, and 1/32 of the original image. The feature fusion network adopts the BiFPN[24]. The BiFPN realizes bidirectional information transfer and multi-scale weighted feature fusion by down-sampling the high-level feature maps and up-sampling the low-level feature maps, improving the model's performance in the crack segmentation task. Without bells and whistles, the

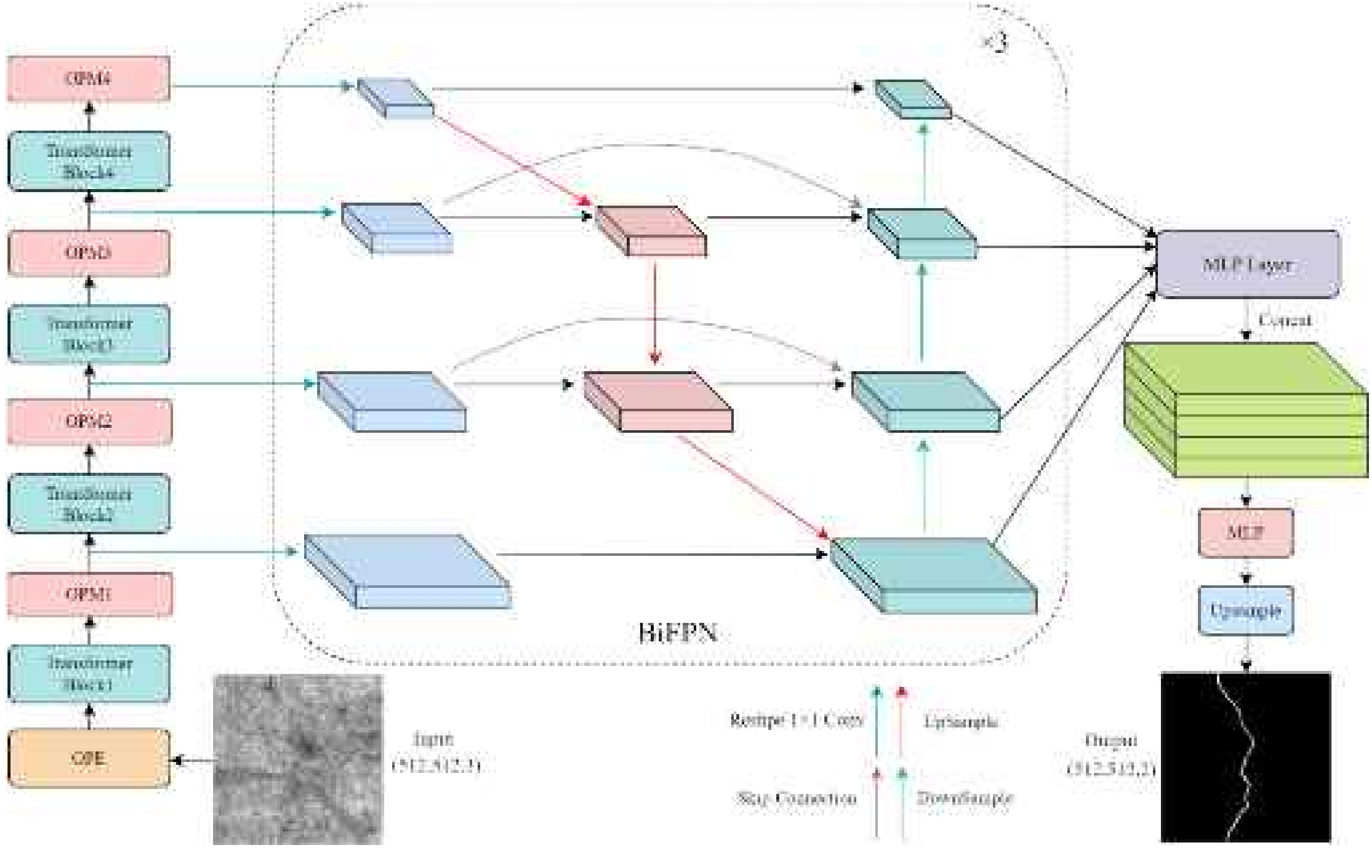
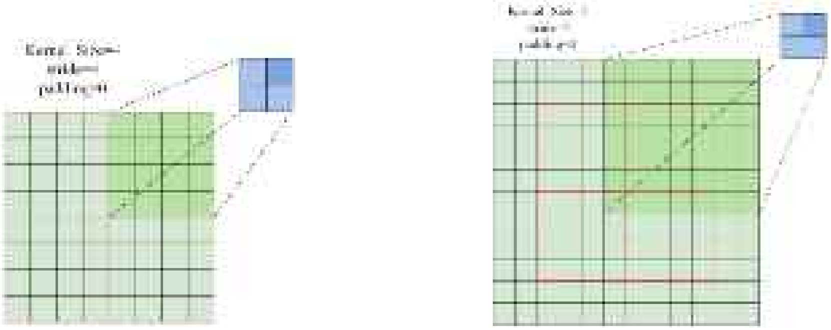


Fig. 5. Crack-Segmentor network structure.

decoder adopts a lightweight network, the fused feature maps are first linearly mapped by MLP Layer to unify the feature maps' shape, then concatenated in the channel direction. The further fused feature maps go through the MLP for its channel number adjustment. Finally, they go through the up-sample for image resolution adjustment to output a binary image the same size as the original image.

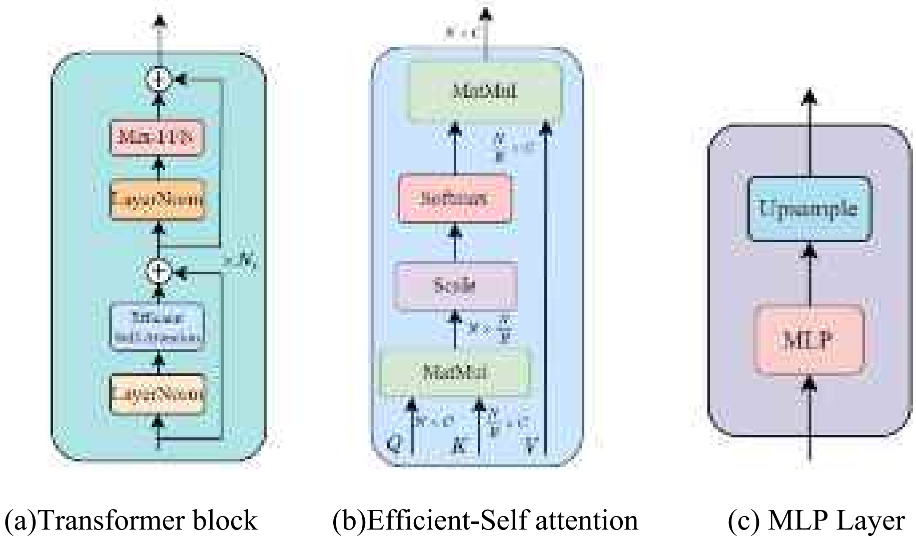
For RGB images, the data format is a three-dimensional matrix, while the Transformer Block requires the data format to be a two-dimensional vector. Therefore, the data is transformed from 3D to 2D by OPE module. Firstly, an image is divided into patches overlapping to increase the smoothness and continuity of the image after division, especially in the edge region of the image. Then each patch is mapped to a one-dimensional vector by linear mapping. Finally, each image is converted to a two-dimensional vector. For program implementation, 2D convolution is used to achieve the blocking operation of the image. Fig. 6 shows the principle of performance of two Patch Embedding for image blocking. After Overlap Patch Embedding, the image is converted to a $(128 \times 128, 32)$ 2D vector.

The structure of the Transformer Block is shown in Fig. 7(a). The most important in the Transformer Block is the efficient self-attention mechanism module, which can extract the features of the image more efficiently and is an improvement of the traditional self-attention mechanism [25].LayerNorm [26] performs normalization between each layer, which can improve the gradient propagation and increase the model generalization. Mix-FFN [27]improves the traditional feed-forward neural network structure, which can adapt the model to image inputs of different scales and enhance the model's generalization ability. From Stage 1 to Stage 4, the Transformer Block is stacked N_i times, and N_i is set to [2,2,2,2] respectively during model design. Eq. (1) is the formula for the traditional self-attention mechanism.



(a) Non-overlap patch embedding (b) Overlap patch embedding

Fig. 6. Patch Embedding. (a) Non-overlap patch embedding (b) Overlap patch embedding.



(a)Transformer block (b)Efficient-Self attention (c) MLP Layer

Fig. 7. Structures in the network. (a)Transformer block (b)Efficient-Self attention (c) MLP Layer.

$$Attention(Q, K, V) = softmax\left(\frac{Q \times K^T}{\sqrt{d_k}}\right)V \tag{1}$$

Where d_k is the number of heads of the multi-head self-attention mechanism, and d_k is set to [1,2,5,8] from stage 1 to stage 4. Q, K, and V are $N \times C$ vectors whose computational complexity is $o(N^2)$.

When the input image is increased by N times, the computation is increased by the square of N. Therefore, an efficient self-attention mechanism is used to reduce the computation. The Efficient-Self Attention mechanism reshapes K and V into $\frac{N}{R \times C}$, reducing the computational complexity to $o(\frac{N^2}{R})$. Fig. 7(b) shows the structure of Efficient-Self Attention, where R is set to [64,6,4,1] from stage 1 to stage 4 in the model.

After the encoder, it is necessary to get the high-resolution low-level feature maps and low-resolution high-level feature maps. OPM is used to down-sample to get the feature maps of different scales. The program implementation of OPM is the same as OPE, down-sampling of feature maps is achieved by 2D convolution with different strides and kernel sizes. After the four stages of feature extraction OPM, the output four feature maps are $(128 \times 128, 32)$,$(64 \times 64, 64)$, $(32 \times 32, 160)$ and $(16 \times 16, 256)$,which are then input to the feature fusion network for fusion.

The feature pyramid network is most common in feature fusion networks [28], where up-sampling by top-down can pass high semantic information to shallow features while retaining the high-resolution properties of shallow features. Although the up-sampling of FPN passes the semantic information of the deep features to the low-level features, the detailed information of the deep features still needs to be improved. In contrast, BiFPN uses bidirectional connections to fuse features, which not only passes the semantic information of the deep features to the shallow features but also passes the detailed information of the shallow features directly to the deep features to enhance the detailed information of the deep features. Traditional feature fusion is often just a simple summation or superposition between feature maps, such as using concat or shortcut connection, without differentiating between feature maps fed in simultaneously. However, different feature maps have different resolutions, and their contributions to the fused feature map are also other, so simply performing a summation or superposition of them is not the best operation. So, we use a simple but efficient mechanism of weighted feature fusion. Such bidirectional information transfer and weighted feature fusion can better handle the feature representation of multi-scale targets and enhance the model's perception of cracks, thus improving the model's performance. After the BiFPN module, the output four feature layers are $(128, 128, 64)$,$(64, 64, 64)$,$(32, 32, 64)$ and $(16, 16, 64)$.

The four feature maps after BiFPN module have inconsistent dimensions and must to be dimensionally unified. Input to MLP Layer for MLP linear mapping, then resize feature maps to unity by Upsample for channel direction concatenation. Fig. 7(c) shows the structure of the MLP Layer.

3.2. ROI threshold segmentation algorithm

The process of ROI threshold segmentation algorithm is shown in Fig. 8. Firstly, an ROI is selected on Image 1, which is the crack region in the neural network prediction result Mask 1. Secondly, only the ROI is operated with contrast enhancement to make the crack in the ROI more apparent, and the result is in Image 2. Then the pixel value of Mask 1 is reversed to get Mask 2. The pixel value of the crack in Mask 2 is 0, and the pixel value of the background is 255. Then the pixel value of Mask 2 and the contrast-enhanced image of ROI are added together to get Image 3. In Image 3, the pixel values of ROI remain the same as in Image 2, while the pixel values in the non-ROI (background region) are all changed to 255 (8-bit binary image with a maximum pixel value of 255). Because the pixel values in the non-ROI are all changed to 255, the influence of complex background on crack threshold segmentation is eliminated, and the drawback of misrecognizing the complex background as cracks in traditional threshold segmentation is solved. However, traditional threshold segmentation requires manual selection of the threshold, and when the image changes, the threshold may need to be re-selected. Therefore, Otsu's algorithm [29] is used for the ROI region to automatically select the threshold and segmentation, further removing the background in the ROI and obtaining finer crack edges. Image 4 shows the image after threshold segmentation. Finally, small voids in the connectivity domain caused by threshold segmentation are removed by morphological closure operation, and the contours of the outer part of the crack are smoothed. Eventually, the pixel values are reversed to obtain the final binary image Mask 3.

Threshold segmentation may lead to localized breaks in the segmented cracks, which can adversely affect their quantification. Therefore, this research uses a crack connection algorithm based on Dijkstra's algorithm[30]for the fracture connection of cracks. The steps of the algorithm are as follows.

(1) Firstly, the binary image of the cracks after ROI threshold segmentation processing is sequentially divided into connectivity domains. Each connectivity domain is an independent crack

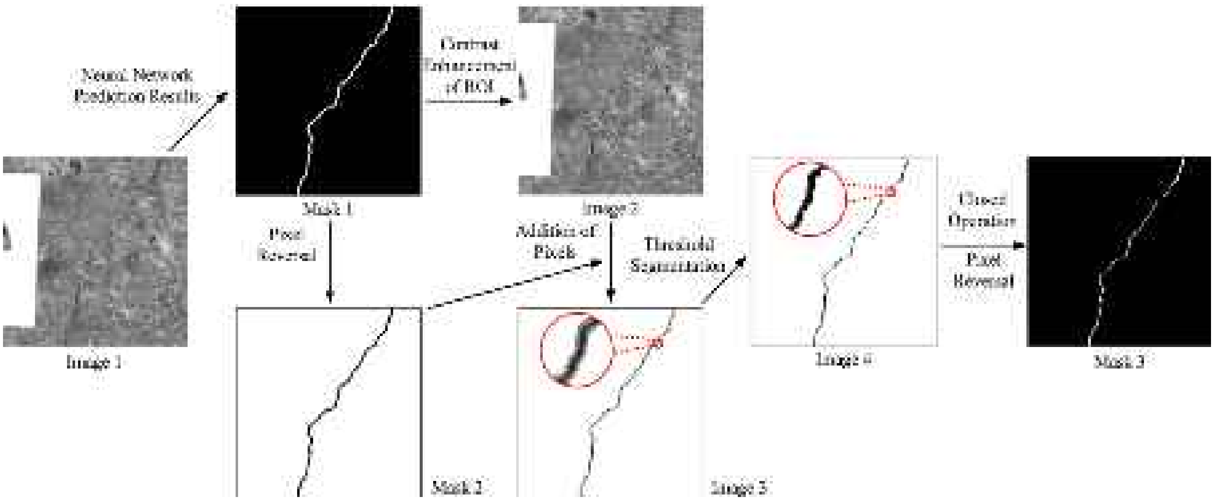


Fig. 8. ROI threshold segmentation.

segment, and the topmost connectivity domain is selected as the starting unit.

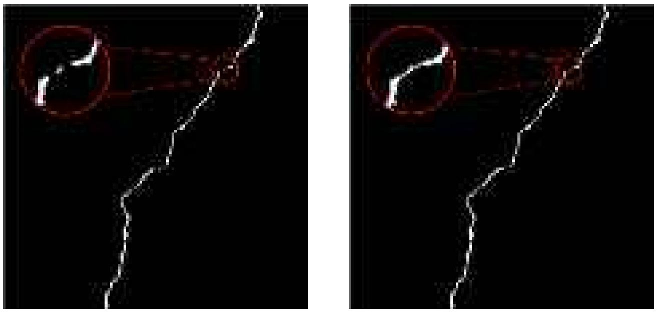
(2) Calculate the Euclidean distance from the starting unit to each remaining unit and connect the starting unit with its closest distance to the endpoints to realize the connecting domains' merging.

(3) Perform the repetition of step 2). The merged unit is again used as the new starting unit to connect and merge the connected domains according to the shortest Euclidean distance until all the connected domains are merged.

A distance threshold is designed when cracks are connected. When the Euclidean distance between two connected domains of a crack is bigger than the distance threshold, it is considered two independent cracks, and no connection of connected domains is performed. The crack-break connection using the crack connection algorithm is shown in Fig. 9.

3.3. Crack quantification algorithm

In this research, the length and width are quantified based on the skeleton of the crack and its edges. The skeleton of the crack is extracted using the Zhang-Suen algorithm [31], and the edges of the crack are extracted using the Sobel algorithm [32]. The Zhang-Suen algorithm is a relatively simple and effective refinement algorithm for skeletonizing target objects in binary images. The algorithm can retain the main structural features of the target object and remove unnecessary pixels to obtain a more streamlined skeletonization result. The Sobel algorithm is one of the standard operators used for image edge detection. It defines two convolution kernels that perform convolution operations in the horizontal and vertical directions of the image and then combine the gradients from both directions into a single gradient. The Sobel algorithm detects the edges by obtaining the gradient value at each pixel in the image. The length of the crack is calculated from the length of the skeleton since the crack is roughly the same length as its skeleton.



(a) Before connection (b) After connection

Fig. 9. Connection of cracks. (a) Before connection (b) After connection.

The width of cracks is an essential basis for evaluating the crack condition. In this research, the skeleton and eight-direction search methods are used to obtain the width of cracks. The principle of crack width measurement is shown in Fig. 10. The specific calculation steps are as follows:

(1) Identify the topmost pixel point of the skeleton as the starting seed point.

(2) Within the segmented crack region, the distances from the seed point to the edge line of the crack were calculated for each of the eight directions 0°, 45°, 90°, 135°, 180°, 225°, 270°, 315° and recorded as $w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8$ respectively.

(3) Combine the distances in the four directions of 0° and 180°, 45° and 315°, 90° and 270°, 135° and 225° by combining $d_1 = w_1 + w_5$, $d_2 = w_2 + w_6$, $d_3 = w_3 + w_7$, $d_4 = w_4 + w_8$, let $D = \min(d_1, d_2, d_3, d_4)$, D is the width of the crack at the seed point.

(4) Calculate $W = W_0 \times D$. Where W_0 is the pixel equivalent, and W is the actual width at the seed point.

(5) Update to the next seed point and repeat steps 2) ~ 4), until the width at all seed points is measured.

4. Experimental results and analysis

4.1. Performance experiments comparing the UAV and the ring-climbing robot

To verify the effectiveness of the proposed ring-climbing visual

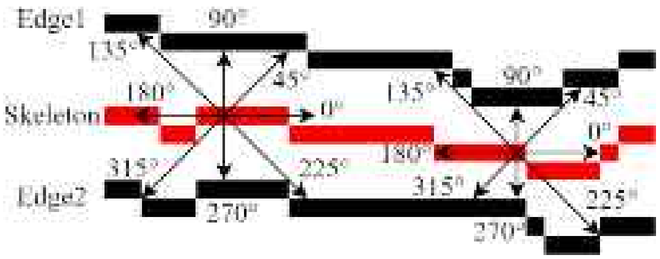


Fig. 10. Crack width quantification.

scanning operation robot system, we performed a comparative experiment of bridge piers disease detection using a DJI UAV and a ring-climbing robot at Shouchun Bridge in Huainan, respectively. Figs. 11 and 12 show the results of our disease acquisition using the UAV and the ring operation robot at the same location of a bridge pier, and it can be seen that the images collected by our proposed ring-climbing visual scanning robot acquire images with higher quality and accuracy compared to the UAV.

Table 1 shows the comparative experimental results of using the UAV and the ring climbing robot to detect bridge piers disease. Although the UAV has strong maneuverability and low operating difficulty, it cannot detect bridge piers disease near the acquisition of high-definition images of the surface of the bridge piers to ensure safety. It is affected by the airflow near the bridge piers, which has the risk of blowing up the aircraft. Our proposed ring-climbing visual scanning robot, although its detection efficiency is slightly lower than that of a UAV, can suspend steadily on the surface of bridge piers during the detection process and can collect high-definition, full-area surface images of bridge piers. Meanwhile, the distance between the camera and the surface of bridge piers is always constant, and the pixel equivalents between different images are always identical, and thus it is more convenient to perform accurate quantification of the captured images compared to UAVs. In addition, the ring robot carries six 24 V 8400mAh lithium batteries, which can work continuously for about 3 h, greatly increasing the endurance time relative to the UAV.

4.2. Training and prediction of Crack-Segmentor model

To build the dataset for training the Crack-Segmentor model, we obtained images of bridge piers cracks by using the ring-climbing visual scanning robotic system at Shouchun Bridge in Huainan. The image acquisition unit employs a Hikvision industrial camera, model MV-CA050-20UM, with a resolution of 2592 × 2048 for image acquisition. The software was designed using Microsoft Visual Studio 2019, the camera captures images displayed and saved in the industrial computer utilizing image acquisition software. Fig. 13 illustrates the image acquisition software.

1912 images of cracks with a resolution of 512 × 512 were cropped from the collected disease images and labeled manually using Labelme software. It was mirrored, flipped to augment the dataset to 7648, which were then divided into a 9:1 ratio of train set and test set. The original images and labels can be seen in Fig. 14.

The Crack-Segmentor model was developed within the Pytorch framework and trained on a Windows system using a NVIDIA GeForce RTX2080Ti GPU. The initial learning rate was set to 5e-5, the optimizer was the AdamW optimizer [33], the exponential decay rate for the 1st moment estimates was set to 0.9, and the exponential decay rate for the 2nd moment estimates was set to 0.999. The loss function is a weighted loss function that combines the Dice Loss[34] with the Focal Loss[35]. Dice Loss uses the intersection and concatenation between the predicted results and the actual labels. This enables the model to capture object boundaries and details better, thus improving the accuracy of segmenting cracks. A crack in a disease image may occupy only a tiny percentage of the pixels, and most of the remaining pixels are background, so crack segmentation suffers from category imbalance. Focal Loss makes the model focus more on hard-to-distinguish cracks by decreasing the weight of the background pixels and increasing the weight of the crack pixels. Combining the two loss functions can synthesize the category imbalance problem with the segmentation accuracy, thus improving the model’s performance in the crack segmentation task. The model is validated every 5 epochs of training and approaches convergence after 1300 epochs of training.

The evaluation parameters were selected as Pixel Accuracy (*PA*), Precision (*Pr*), *Recall*, and mean Intersection over Union (*mIoU*). The formula is as follows:

$$PA = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{2}$$

$$P_r = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$mIoU = mean(\frac{TP}{TP + FP + FN}) \tag{5}$$

Where *TP* is the number of pixels labeled as cracks and recognized as cracks; *TN* is the number of pixels labeled as backgrounds and recognized as backgrounds; *FP* is the number of pixels labeled as backgrounds but recognized as cracks; and *FN* is the number of pixels labeled as cracks but recognized as backgrounds.

A performance comparison is made with two Transformer networks based on efficient self-attention mechanisms to validate the Crack-Segmentor ’s performance. Table 2 shows the performance comparison results. Crack-Segmentor is higher than SegFormer and Segcrack in *PA*, *Pr*, *Recall*, and *mIoU*.

When using deep learning for disease identification of bridge piers, the inference speed of the network determines its speed and efficiency of disease identification. Therefore, the efficiency of the model is also crucial. Based on this, the Crack-Segmentor proposed in this study is compared with two semantic segmentation networks based on convolutional neural networks and two Transformer networks with efficient self-attention mechanisms for model efficiency. Table 3 shows the results of the comparison of model efficiency in a NVIDIA GeForce RTX2080Ti GPU. Crack-Segmentor is nearly the same as SegFormer and Segcrack; its *FPS* is around 38 img/s, and it has excellent efficiency. Although slightly lower than PSPNet, Crack-Segmentor ’s *mIoU* achieves 88.09 %, the highest among all models. Fig. 15 shows the results predicted by the different models.

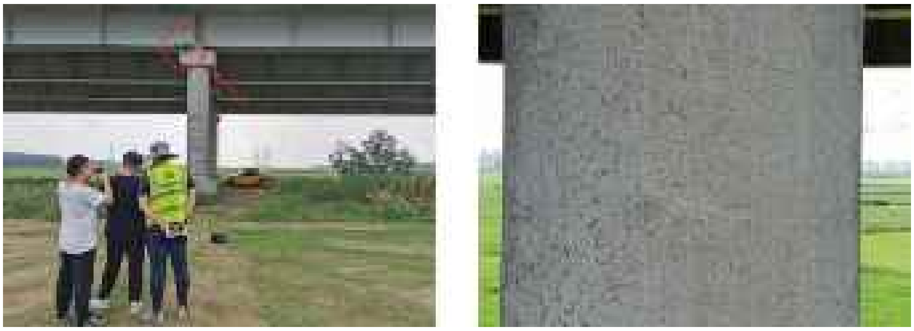


Fig. 11. UAV image acquisition.

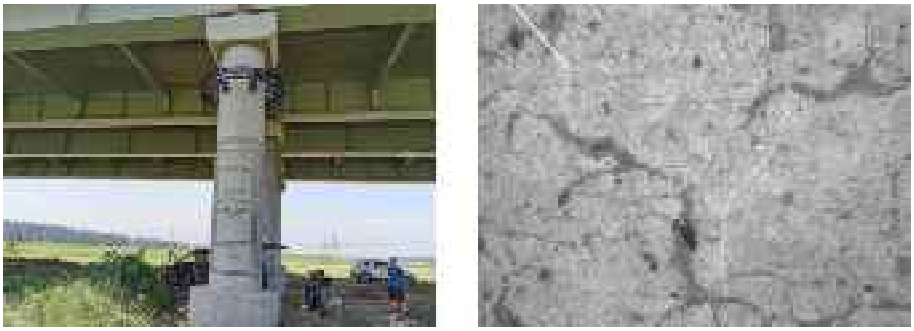


Fig. 12. Ring-climbing robot image acquisition.

Table 1
Experiments results comparing the UAV and the ring-climbing robot.

	Precision	Efficiency (10 m)	Safety	Region	Battery life
UAV	1 mm	10 min	Ordinary	Entire	30 min
Ring-climbing robot	0.1 mm	30 min	Excellent	Entire	3 h

Table 2
Model performance comparison.

	PA(%)	Pr(%)	Recall(%)	mIoU(%)
SegFormer[27]	99.61	85.53	83.91	86.55
Segcrack[18]	99.63	86.03	84.55	86.98
Crack-Segmentor	99.66	87.19	86.20	88.09

Fig. 13. Image acquisition software.

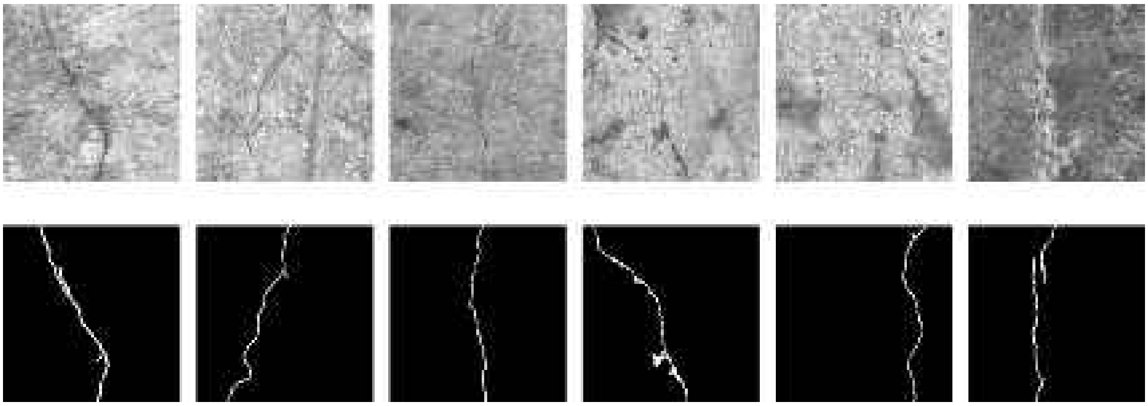


Fig. 14. Crack images and labels.

Table 3
Model efficiency comparison.

	Backbone	Image size	mIoU (%)	Params (M)	FLOPs (G)	FPS (img/s)
PSPNet[36]	MobileNet	512^2	72.81	2.376	5.870	43.78
Deeplabv3+ [37]	MobileNet	512^2	86.32	5.813	52.867	33.21
SegFormer	MiT-B0	512^2	86.55	3.707	13.489	39.19
Segcrack	MiT-B0	512^2	86.98	3.823	15.428	38.40
Crack-Segmentor	MiT-B0	512^2	88.09	3.764	14.612	37.74

4.3. *Refined segmentation and quantification of cracks*

The initial resolution of the acquired bridge piers disease image (2592 × 2048) directly input into the Crack-Segmentor network for recognition leads to a slower inference speed and higher graphics memory requirement. Down-sampling is typically used to reduce the input image, significantly decreasing the detailed information about the image, resulting in the network's inability to precisely segment the cracks. Thus, this study utilizes the sliding window technique to split the captured images of the original images into equal parts. Subsequently, these parts are fed individually into the network for recognition. Then, they are seamlessly stitched to achieve the ultimate segmentation outputs.

Some of the acquired images of diseased bridge piers are selected for rough segmentation of cracks according to the process described above. Then, fined segmentation is performed using the crack refined segmentation algorithm proposed in this research, and the experimental results are shown in Fig. 16.

Some of the cracks on the bridge piers' surface are too long, the acquired original images may not contain the cracks completely, and the length of the cracks cannot be accurately obtained. Therefore, the original images of bridge piers cracks are selected to quantify the width. Nine binary images of cracks obtained by neural network rough segmentation and refined segmentation algorithm fine segmentation were used to quantify the width of the cracks. The chessboard measured the average pixel equivalent in the images affixed to the bridge piers, which was 0.069 mm/pixel in this research. The comparison results are shown in Table 4. The actual width of the cracks is obtained by the bridge piers inspectors using a crack width gauge. Rough width represents the width of the cracks obtained by rough segmentation using deep learning, and fine width represents the width of the cracks obtained by fine segmentation using the refined segmentation algorithm.

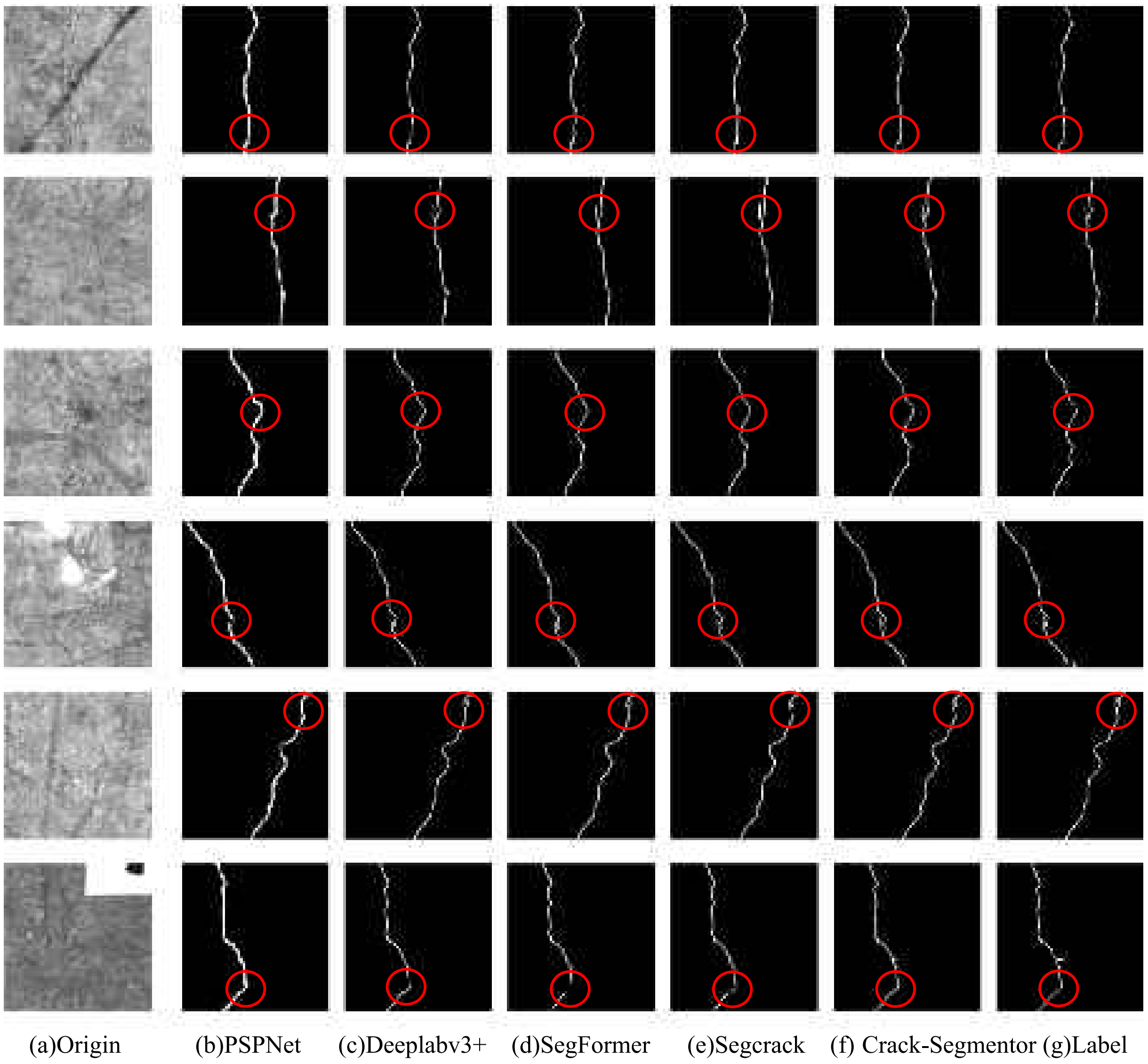
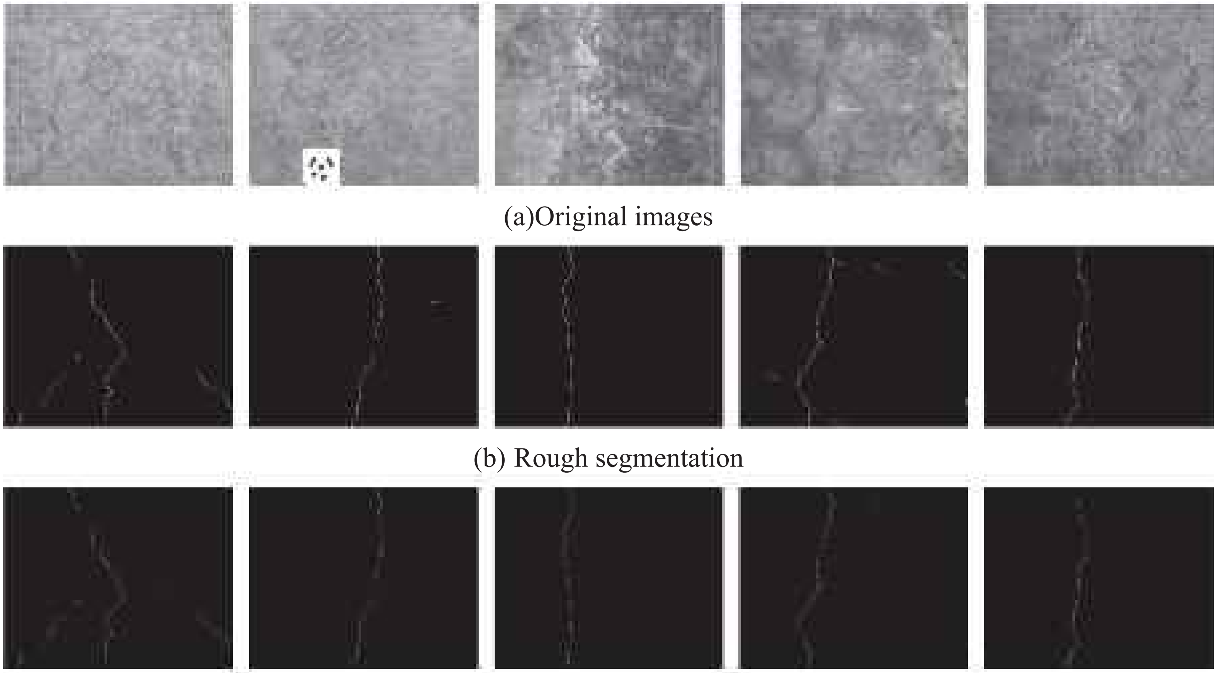


Fig. 15. Results predicted by the different models. (a)Origin (b)PSPNet (c)Deeplabv3+ (d)SegFormer (e)Segcrack (f) Crack-Segmentor (g)Label.


(a)Original images

(b) Rough segmentation

(c) Fine segmentation

Fig. 16. Original Images segmentation results. (a)Original images. (b) Rough segmentation. (c) Fine segmentation.

Table 4
Quantization of original images width.

	Rough width (mm)	Fine width (mm)	Actualwidth (mm)
Crack1	0.35	0.17	0.12
Crack2	0.31	0.16	0.13
Crack3	0.35	0.19	0.10
Crack4	0.39	0.19	0.13
Crack5	0.32	0.17	0.20
Crack6	0.38	0.21	0.20
Crack7	0.31	0.19	0.10
Crack8	0.30	0.20	0.10
Crack9	0.34	0.13	0.10

Fig. 17 shows the quantization error of crack width in the original images. The result shows that the crack refined segmentation algorithm proposed in this research dramatically reduces the quantization error of the direct use of deep learning segmentation results, and the maximum error is within 0.1 mm, which realizes the accurate quantization of cracks.

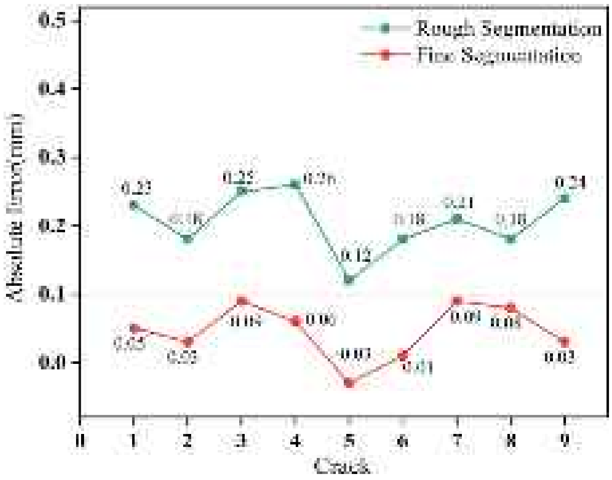


Fig. 17. Quantization error of crack width in original images.

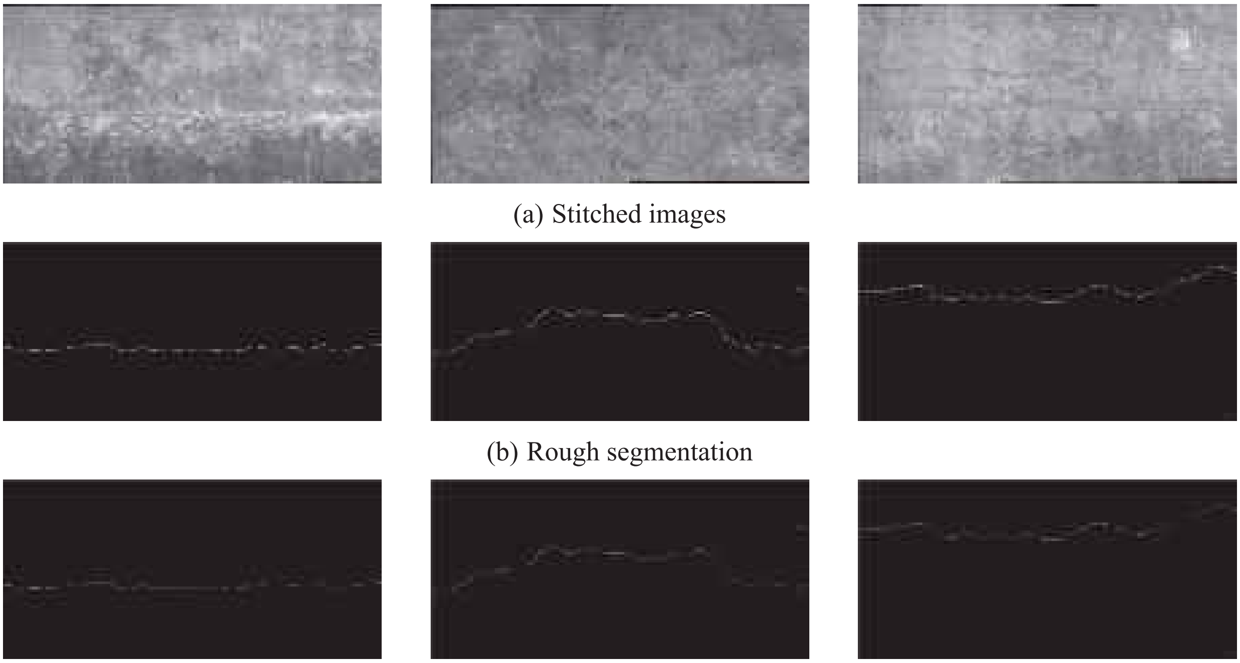
At the same time, the complete image of the bridge piers cracks obtained using feature matching stitching is selected for deep learning crack rough segmentation and refined segmentation algorithm fine segmentation, respectively. Fig. 18 shows the segmentation results.

Similarly, eight complete images of the cracks after stitching were selected, and the cracks were quantized using the crack quantization algorithm for simultaneously deep learning rough segmentation and refined segmentation algorithm after fine segmentation, and the results are shown in Table 5.

In Table 5, rough length and rough width represent the cracks' length and width for deep learning rough segmentation, and fine length and fine width represent the cracks' length and width for refined segmentation algorithm fine segmentation. From the Table 5, the maximum error in the length of the roughly segmented and finely segmented cracks is within 7 mm, indicating that the difference in the length is insignificant. Still, the difference in the width may be relatively significant. Fig. 19 shows the quantization error of crack width in stitched images. The crack width errors of deep learning rough segmentation are all over 0.15 mm, while the crack width errors of fine segmentation using the refined segmentation algorithm are maximum within 0.1 mm. The results show that the crack-refined segmentation algorithm proposed in this research can reduce the error of width quantification by directly using the deep learning segmentation results and can realize the accurate quantification of cracks.

In addition, no ring-climbing robots were used. We captured images of six bridge piers' surface cracks with a camera to verify the effectiveness of the crack detection algorithm proposed in this paper. At the same time, a hand-held crack width gauge was also used to detect the width of the cracks and the results were regarded as the actual width. Table 6 demonstrates the comparison results of the cracks' width.

As can be seen from Table 6, comparing the width obtained by the crack detection algorithm proposed in this paper with that obtained by the crack width gauge, the absolute error is only within 0.05 mm, which proves that the crack detection algorithm proposed in this paper has


(a) Stitched images

(b) Rough segmentation

(c) Fine segmentation

Fig. 18. Stitched Images segmentation results. (a) Stitched images. (b) Rough segmentation. (c) Fine segmentation.

Table 5
Quantization of stitched images.

	Rough width (mm)	Fine width (mm)	Actual width (mm)	Rough length (mm)	Fine length (mm)
Crack1	0.32	0.18	0.10	385.43	387.23
Crack2	0.40	0.17	0.13	64.24	66.31
Crack3	0.41	0.15	0.10	65.96	70.17
Crack4	0.35	0.18	0.12	188.37	192.37
Crack5	0.37	0.20	0.20	54.99	56.65
Crack6	0.45	0.25	0.20	152.42	157.94
Crack7	0.42	0.17	0.10	48.16	54.30
Crack8	0.32	0.17	0.13	83.77	82.18

Table 6
Comparison results of the cracks' width.

	Proposed method (mm)	Actual width (mm)	Absolute error (mm)
Crack1	0.23	0.21	0.02
Crack2	0.21	0.25	0.04
Crack3	0.17	0.20	0.03
Crack4	0.19	0.24	0.05
Crack5	0.32	0.28	0.04
Crack6	0.26	0.23	0.03

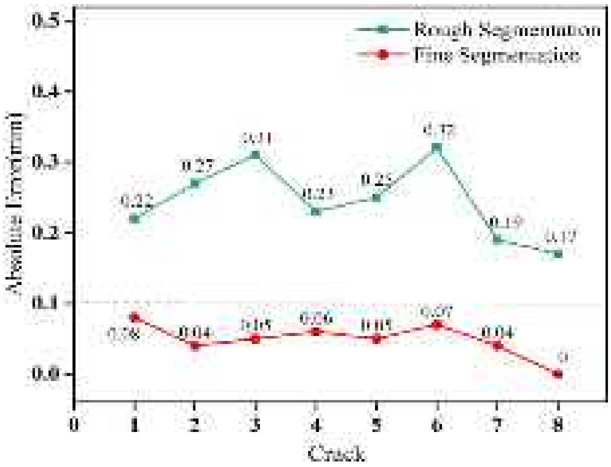


Fig. 19. Quantization error of crack width in stitched images.

excellent accuracy in quantifying the width of cracks and satisfies the requirements of engineering. However, the crack width gauge can only be manually hand-held and cannot detect the region on the bridge piers that people cannot reach. Therefore, the crack detection algorithm proposed in this paper, combined with the ring-climbing visual scanning robot, can achieve crack detection and accurate measurement of the entire region of the bridge piers.

5. Conclusion

Aiming at the problems of existing machine vision methods, which are time-consuming, costly, and challenging to achieve entire region detection, this research proposes a disease acquisition system using a novel ring-climbing visual scanning operation robot. The robot can be used on any bridge pier with a 1.2 to 1.5 m diameter, enabling entire and fast acquisition of bridge piers surface images. In the meantime, the issue of most researchers failing to conduct fine segmentation and quantification of cracks identified by the neural network has been addressed through the proposal of a crack segmentation and quantification algorithm. After experimental validation, this algorithm accurately identifies and quantifies surface cracks on bridge piers, with a quantification error of within 0.1 mm. Although the crack segmentation and quantification algorithm proposed in this study shows reasonable performance, some limitations may still need to be revised. Specifically, we would like to equip the visual scanning vehicle with an industrial computer with a GPU and deploy the crack detection algorithm proposed in this paper so that the industrial computer can perform real-time processing of the captured images. We will also research the problem of quadratic curve-based column surface inverse projection correction of bridge piers disease images and investigate the quantization error of cracks caused by image distortion.

CRediT authorship contribution statement

Hao Du: Writing – original draft, Resources, Methodology. **Huifeng Wang:** Writing – review & editing, Funding acquisition. **Xiaowei Zhang:** Software, Methodology. **Haonan Peng:** Validation, Supervision, Software. **Rong Gao:** Visualization, Methodology. **Xueyan Zheng:** Writing – review & editing, Formal analysis. **Yaxiong Tong:** Visualization, Resources. **Yuanhe Shan:** Validation, Project administration. **Zefeng Pan:** Visualization, Software. **He Huang:** Visualization, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This research was partly supported by The National Natural Science Foundation of China (No.52172324). Partially funded by The Science and Technology Project of Beilin District (GX2350), The Science and Technology Project of Shaanxi Provincial Transportation Department (No.21-20K) and The Fundamental Research Foundation for the Central Universities, CHD (300102323501).

References

[1] H. Zoubir, M. Rguig, M. El Aroussi, A. Chehri, R. Saadane, G. Jeon, Concrete Bridge defects identification and localization based on classification deep convolutional neural networks and transfer learning, Remote Sens. 14 (19) (2022) 4882, https://doi.org/10.3390/rs14194882.

[2] P. Hüthwohl, R. Lu, I. Brilakis, Multi-classifier for reinforced concrete bridge defects, Autom. Constr. 105 (2019) 102824, https://doi.org/10.1016/j.autcon.2019.04.019.

[3] Y.F. Liu, X. Nie, J.S. Fan, X.G. Liu, Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene reconstruction, Comput.-Aided Civil Infrastruct. Eng. 35 (5) (2020), pp. 511-529, doi: 10.1111/mice.12501.

[4] C. Song, L. Wu, Z. Chen, H. Zhou, P. Lin, S. Cheng, Z. Wu, Pixel-level crack detection in images using SegNet, in: International Conference on Multi-disciplinary Trends in Artificial Intelligence, Springer, 2019, pp. 247-254, doi: 10.1007/978-3-030-33709-4_22.

[5] J. Zhang, S. Qian, C. Tan, Automated bridge surface crack detection and segmentation using computer vision-based deep learning model, Eng. Appl. Artificial Intellig. 115 (2022) 105225, https://doi.org/10.1016/j.engappai.2022.105225.

[6] C.B. Zhang, C.C. Chang, M. Jamshidi, Simultaneous pixel-level concrete defect detection and grouping using a fully convolutional model, Struct. Health Monit. 20 (4) (2021) 2199–2215, https://doi.org/10.1177/1475921720985437.

[7] S. Dorafshan, M. Maguire, Bridge inspection: Human performance, unmanned aerial systems and automation, J. Civ. Struct. Health Monit. 8 (2018) 443–476, https://doi.org/10.1007/s13349-018-0285-4.

[8] W. Ding, K. Yu, J.P. Shu, Deep learning and UAV based crack detection method for concrete structures, China Civil Eng. J. 54 (S1) (2021) 1–12, https://doi.org/10.15951/j.tmgcxb.2021.s1.016 (in Chinese).

[9] Z.Y. He, P. Wang, Crack detection method for bridge underside cracks with drone vision, Road Mach. Construct. Mech. 36 (06) (2019) 131–137, https://doi.org/10.3969/j.issn.1000-033X.2019.06.026 (in Chinese).

[10] H.-F. Wang, L. Zhai, H. Huang, L.-M. Guan, K.-N. Mu, G.-P. Wang, Measurement for cracks at the bottom of bridges based on tethered creeping unmanned aerial vehicle, Autom. Constr. 119 (2020) 103330, https://doi.org/10.1016/j.autcon.2020.103330.

[11] M. Tavakoli, C. Viegas, L. Marques, J.N. Pires, A.T. De Almeida, OmniClimbers: Omni-directional magnetic wheeled climbing robots for inspection of ferromagnetic structures, Robot. Auton. Syst. 61 (9) (2013) 997–1007, https://doi.org/10.1016/j.robot.2013.05.005.

[12] J. Lan, W.T. Huang, G.M. Liao, H.L. Su, C.Y. Xi, Vacuum adsorption based robot for detecting diseases in piers, Technol. Innovat. Appl. 08 (2019) 44–45, https://doi.org/10.3969/j.issn.2095-2945.2019.08.015 (in Chinese).

[13] K. Jang, Y.K. An, B. Kim, S. Cho, Automated crack evaluation of a high-rise bridge pier using a ring-type climbing robot, Comput.-Aided Civil Infrastruct. Eng. 36 (1) (2021), pp. 14-29, doi: 10.1111/mice.12550.

[14] H. Ji, Development of an autonomous column-climbing robotic system for real-time detection and mapping of surface cracks on bridges, in: 2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET), Ieee, 2023, pp. 1-6, doi: 10.1109/GlobConET56651.2023.10150133.

[15] J.B. Ma, Research on bridge crack identification and measurement method based on image processing, Beijing Jiaotong University (2021), https://doi.org/10.26944/d.cnki.gbfju.2021.003153 (Thesis, in Chinese).

[16] Y. Liu, S. Cho, B.F. Spencer Jr, J. Fan, Automated assessment of cracks on concrete surfaces using adaptive digital image processing, Smart. Struct. Syst. 14 (4) (2014) 719–741, https://doi.org/10.12989/sss.2014.14.4.719.

[17] L.F. Li, M. Hu, Segmentation of fine bridge cracks based on generative adversarial network, Laser & Optoelectronics Progress. 56 (10) (2019) 102–112, https://doi.org/10.3788/LOP56.101004 (in Chinese).

[18] W. Wang, C. Su, Automatic concrete crack segmentation model based on transformer, Autom. Constr. 139 (2022) 104275, https://doi.org/10.1016/j.autcon.2022.104275.

[19] S. Sony, K. Dunphy, A. Sadhu, M. Capretz, A systematic review of convolutional neural network-based structural condition assessment techniques, Eng. Struct. 226 (2021) 111347, https://doi.org/10.1016/j.engstruct.2020.111347.

[20] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, An image is worth 16x16 words: Transformers for image recognition at scale, a. preprint, arXiv:2010.11929, doi: 10.48550/arXiv.2010.11929, 2020.

[21] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, L. Shao, Pyramid vision transformer: A versatile backbone for dense prediction without convolutions, in: In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 568–578, https://doi.org/10.48550/arXiv.2102.12122.

[22] H. Wu, B. Xiao, N. Codella, M. Liu, X. Dai, L. Yuan, L. Zhang, Cvt: Introducing convolutions to vision transformers, in: In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 22–31, https://doi.org/10.48550/arXiv.2103.15808.

[23] X. Dong, J. Bao, D. Chen, W. Zhang, N. Yu, L. Yuan, D. Chen, B. Guo, Cswin transformer: A general vision transformer backbone with cross-shaped windows, in: In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 12124–12134, https://doi.org/10.48550/arXiv.2107.00652.

[24] M. Tan, R. Pang, Q.V. Le, Efficientdet: Scalable and efficient object detection, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 10781-10790, doi: 10.48550/arXiv.1911.09070.

[25] E.A. Shamsabadi, C. Xu, A.S. Rao, T. Nguyen, T. Ngo, D. Dias-da-Costa, Vision transformer-based autonomous crack detection on asphalt and concrete surfaces, Autom. Constr. 140 (2022) 104316, https://doi.org/10.1016/j.autcon.2022.104316.

[26] J.L. Ba, J.R. Kiros, G.E. Hinton, Layer normalization, a. preprint, arXiv:1607.06450, doi: 10.48550/arXiv.1607.06450, 2016.

[27] E.Z. Xie, W.H. Wang, Z.D. Yu, A. Anandkumar, J.M. Alvarez, P. Luo, SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers, in: 35th Conference on Neural Information Processing Systems (NeurIPS), Vol. 34, Electr Network, 2021.

[28] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2117–2125, https://doi.org/10.1109/cvpr.2017.106.

[29] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1) (1979) 62–66, https://doi.org/10.1109/TSMC.1979.4310076.

[30] S.Y. Zhu, J.C. Du, Y.S. Li, X.P. Wang, Bridge crack detection using U-Net convolutional network, J. Xidian Univ. 46 (04) (2019) 35–42, https://doi.org/10.19665/j.issn1001-2400.2019.04.006 (in Chinese).

[31] T.Y. Zhang, C.Y. Suen, A fast parallel algorithm for thinning digital patterns, Commun. ACM. 27 (3) (1984) 236–239, https://doi.org/10.1145/357994.358023.

[32] S.R. Joshi, R. Koju, Study and comparison of edge detection algorithms, in: 2012 Third Asian Himalayas international conference on internet, IEEE, 2012, pp. 1-5, doi: 10.1109/AHICI.2012.6408439.

[33] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, a. preprint, arXiv:1711.05101, doi: 10.48550/arXiv.1711.05101, 2017.

[34] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 fourth international conference on 3D vision (3DV), IEEE, 2016, pp. 565-571, doi: 10.1109/3DV.2016.79.

[35] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980-2988, doi: 10.1109/iccv.2017.324.

[36] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2881–2890, https://doi.org/10.1109/CVPR.2017.660.

[37] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of

the European Conference on Computer Vision (ECCV), 2018, pp. 801–818, https://doi.org/10.48550/arXiv.1802.02611.