A New Deep Learning-based Approach for Concrete Crack Identification and Damage Assessment

Fuyan Guo[1*], Qi Cui[1], Hongwei Zhang[1], Yue Wang[1], Huidong Zhang[2*], Xinqun Zhu[3] and Jiao Chen[1]

[1]School of Control and Mechanical Engineering, Tianjin Chengjian University, Tianjin 300384, China
[2]School of Civil Engineering, Tianjin Chengjian University, Tianjin 300384, China
[3]School of Civil and Environmental Engineering, University of Technology Sydney, NSW 2007, Australia
Corresponding authors: Fuyan Guo, happytime76@163.com; Huidong Zhang, zhhuidong@126.com.

Abstract: Concrete building structures are prone to cracking as they are subjected to environmental temperatures, freeze-thaw cycles, and other operational environmental factors. Failure to detect cracks in the key building structure at the early stage can result in serious accidents and associated economic losses. A new method using the squeeze and excitation (SE)-U-Net model based on a conditional generative adversarial network (GAN) has been developed to identify small cracks in concrete structures in this paper. This proposed method was a pixel-level U-Net model based on a generative network, that was integrated the original convolutional layer with an attention mechanism, and an SE module in the jump connection section was added to improve the identifiability of the model. The discriminative network compared the generated images with real images using the PatchGAN model. Thus, the crack-recognition images generated by the generative network were infinitely close to the genuine crack images through the repeated adversarial training for the generative and discriminant networks. In damage assessments, the crack skeleton was represented by the individual pixel width and recognized using the binary morphological crack skeleton method, in which the final length, area, and average width of the crack could be determined through the geometric correction index. The results showed that compared with other methods, the proposed method could better identify subtle pixel-level cracks, and the identification accuracy is 98.48%. These methods are of great significance for the identification of cracks and the damage assessment of concrete structures in practice.

1. Introduction

The concrete building structure is affected by operational loads and external environments—such as environmental temperatures and freeze-thaw cycles, and it prone to crack. The crack could be spread gradually to form a wider or longer crack. When the crack width in the concrete structure, especially in the key building structural member, is greater than 0.3 mm, the performance of the building can be compromised seriously (ACI, 1999). Consequently, the regular crack detection for concrete structures is very important for assuring the performance, service life, and aesthetics of buildings.

Local non-destructive techniques have been developed for concrete crack detection, including the pulsed laser scanning and infrared thermographic techniques (Sabeenian et al., 2021). Pulsed laser scanning is to generate ultrasonic waves and the propagation waves are received by a fixed transducer to detect the defects in concrete members. Infrared thermographic technology is to measure the infrared waves to construct the digital image for defect detection (Hao et al., 2022; Moghnieh et al., 2022). Both techniques have a lack of clarity in its results and its high cost. The computer vision technique has been used to detect the surface crack of concrete structures and it has been attracted the attention of researchers due to its low cost. Early image processing based crack detection methods mainly use threshold recognition algorithms (Hoang, 2018a; Chen et al., 2019) and backpropagation (BP) neural networks (Li et al., 2014). Typical threshold recognition algorithms include Prewitt's algorithm (Lei et al., 2018) and Canny's algorithm (Hoang and Nguyen, 2018b) and these algorithms are susceptible to background clutter that induces a large number of burrs around cracks, resulting in incomplete crack-recognition edge contours. A BP neural network has been used to update the weights between different layer nodes through multiple iterations until the desired crack output effect could be obtained. Although this method improved accuracy, the difficulty of parameter adjustment could result in over- and under-recognition (Liu et al., 2017). The probabilistic analysis framework, which combines dynamic Bayesian network and fracture mechanics, is utilized for predicting the fatigue crack propagation of multiple fatigue details in long-span bridges and tracking the predicted component-level fatigue reliability (Xu et al., 2022). Automatic crack recognition using the Mask-Rcnn network (Li et al., 2021; Wu et al., 2021b) is also a commonly used crack-recognition algorithm. As the above, for fine cracks, the continuity between crack pixels and background pixels is ignored, and the detail edge information is missing.

In recent years, with the rapid development of computer hardware, many deep learning algorithms for crack detection have been proposed. The convolutional neural network is used to extracted features of images for automatic detection of pavement cracks by smart phones (Zhang et al., 2016). A constrained Boltzmann machine is used to detect cracks (Li et al., 2018) and a fused

convolutional neural network to detect distributed cracks in steel girders of real bridges with great success (Li et al., 2019). Using Dual Attention Network can adaptively integrate local features with their global dependencies to obtain accurate segmentation results (Fu et al., 2019). K-means is applied in the field of image recognition as a clustering-division method (Chen et al., 2017), and it is used for crack recognition by iterative approximation of the defined objective function. It needs to set and determine the parameter values manually, such as the initial number of clusters, and the recognition accuracy is affected by noise in the image (Cui et al., 2014). As above, there is still a big challenge for fine crack detection in concrete structures, such as incomplete edge contours, over- and under-recognition phenomena, loss of edge details, and low recognition accuracy.

The crack detection methods based on deep learning and computer vision needs a large amount of training data. In many cases, the data in training set is lacking or uneven and it makes insufficient training for an accurate detection model. The generative adversarial networks (GANs) are widely used to solve the problem (Sabeenian et al., 2021). To address the problem of interdependence between cracks and background pixels, and enhance the continuity of recognition results and ensure edge detail information, GANs are used in the field of target recognition image generation (Liu and Wu, 2021; Chen et al., 2021; Zhang et al., 2019). GANs have been used to recover blurred images to solve the problem of image blurring during camera imaging (Sang et al., 2019) and image recognition (He et al., 2019). However, the conventional GAN models are very unstable for training because of the inherent adversarial relationship between the generative and discriminative networks. To better balance these two networks, a conditional GAN (CGAN) is constructed by adding conditional information to both the generative and discriminative networks (Mirza and Osindero, 2019). As a method to learn data based on generative and discriminative networks, the CGAN uses games to continuously optimize the generative and discriminative networks so that the generated images are increasingly similar to the real images in terms of distribution, and achieving enhanced pixel-to-pixel continuity, thus improving recognition accuracy. Adversarial training methods are proposed for training semantic recognition models to improve the recognition accuracy using iterative game training between generative and discriminative networks with the addition of conditions (Yildiz et al., 2021). Traditional error functions, such as the mean squared error (MSE), are much sensitive to outliers. If there are outliers in the sample, the MSE assigns higher weights to the outliers, which sacrifices the prediction effect of other normal point data, eventually reducing the overall model performance. PatchGAN models perform binary discrimination on the truth or falsity of the input data (Gao et al., 2019; Huo et al., 2018). Unlike other discriminant functions that output a single value, PatchGAN outputs an $N \times N$ matrix X. Gatys et al. (2016) used PatchGAN models to enhance the consistency of a generated image using a target image in terms of texture.

Zhang et al. (2021) used PatchGAN discriminant networks in pix2pix networks, not only improving the speed of image computation but also ensuring the model focused on image details to meet the requirements of high-resolution and high-definition details.

Compared to object detection, pixel segmentation can identify and locate cracks more accurately as cracks occupy a relatively small proportion of pixels. In addition, pixel segmentation can provide the distribution path and shape of cracks, which is crucial for evaluating the health of concrete structures. To obtain quantitative information such as crack length and area, the semantic segmentation algorithm is used to detect cracks (Xu et al., 2023a) and the proposed method has achieved significant improvement in the damage identification of complex structures (Xu et al., 2023b). The U-Net network is an improved fully convolutional neural network model (Shelhamer et al., 2015), and it has been applied to improve the image recognition performance in the field, such as water body recognition for remote sensing images (Feng et al., 2019) and road detection (Zhang et al., 2018). The method projects the features of the coding layer to that of the decoding layer and a large volume of contextual information in the downsampled part is propagated to the higher-resolution network layer to avoid the information loss during the coding process, which helps to identify crack-edge contours more accurately and completely. A new self-attentive adaptive (SAA) neuron to the conventional U-Net network is added to enable the microcrack segmentation of real steel box girder bridges (Zhao et al., 2022). However, the edge detail information for fine cracks is not detected. To address the above issues such as the loss of crack-edge details during crack recognition, the squeeze and excitation (SE) module is used to identify image features by incorporating an attention mechanism in the U-Net architecture (Liu et al., 2021; Zhu et al., 2021; Chen et al., 2020). It can add the weight for each feature channel based on the value of the feature image to enhance the crack features and reduce the irrelevant features, such as background features. The effect of crack-feature recognition is enhanced, the edge detail information of cracks is retained, and then the recognition accuracy is improved. The CGAN can satisfy the requirements of pixel continuity, and the attention mechanism based on the SE module introduced in the U-Net can ensure the maintenance of crack-edge detail information in crack image recognition.

In this study, an SE-U-Net recognition model based on CGANs has been developed to identify pixel-level fine cracks in concrete structures. The issues of incomplete edge contours, loss of edge details, and low recognition accuracy in the fine crack detection are studied. Different with the traditional crack detection method based on image features, the proposed method utilises the CGAN as the basic architecture combined with an improved U-Net pixel-level image recognition method. The U-Net model for pixel-level recognition is used as the generative network and the PatchGAN

model is used as the discriminative network. The output of the generative network is fed into the discriminant network in the form of a convolution to recognize the image features and discriminate the image source. The discriminant result is then fed back to the generative network to reduce the error between the recognition result and the true value. The generative network is trained against the discriminant network to continuously improve the image recognition accuracy and solve the problems of incomplete edge contours and loss of edge details in crack-recognition images. To achieve a quantitative calculation of the cracks, the morphological characteristics of cracks, including crack area, length, width and other attributes, is identified and calculated using the diametrical morphological skeleton recognition method. The results show that the proposed method could locate crack widths down to 0.1 mm and provide valuable insights for the health assessment of concrete and wooden structures.

2. Theory: Crack Identification based on CGAN SE-U-Net Model

2.1 Conditional generative adversarial networks

The pixel classification using the baseline model can improve the accuracy of pixel recognition. However, the continuity between crack pixels and background pixels can be easily ignored, causing the loss of crack-edge details or the cracks to differ considerably in size and shape from the real standard in the recognition results. A GAN can better solve such problems. However, an inherent adversarial relationship exists between the generative network (G) and discriminative network (D) in a GAN, making the generated results unstable. To better balance the two networks, a CGAN can be constructed by adding conditional information to both the generative and discriminative networks. This allows the generative network to generate the final result in a way that is not completely free and unsupervised, but rather generates the appropriate result according to the specified conditions, avoiding the problem that the generative network would incorrectly identify the cracks. CGAN can effectively solve the problem that GAN is difficult to converge, and does not significantly exacerbate the training burden of GAN. The objective function of the CGAN can be expressed as follows:

$$L_{cGAN}(D,G)=E_{x,y}[logD(x,y)]+E_{x,z}[log\left(1-D\left(x,G(x,z)\right)\right)] \tag{1}$$

where x denotes the input-assembled prefabricated component crack image, y denotes the manually segmented labeled image, function $D\,(x,\,y)$ denotes the probability that the input image is from manual segmentation, $D\,(x,\,G\,(x,\,z))$ denotes the probability that the input image is from the generator, G attempts to minimize the objective function and D attempts to maximize it. Mathematically, this can be expressed as follows:

$$L = arg\ min_G max_D L_{cGAN}(D,G) \qquad (2)$$

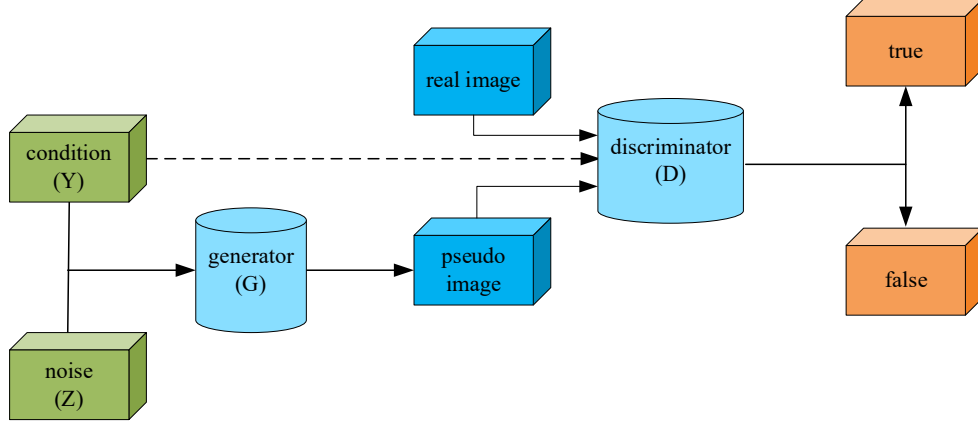where L denotes the outcome generator when **Eq. (2)** reaches the Nash equilibrium.

Fig. 1. The CGAN structure.

To enhance the generative power of the generator, stabilize the training of the generator, and ensure the integrity of the crack-edge segmentation, a distance function *L1* is added to the objective function as follows:

$$L_{L1}(G) = E_{x,y,z}[\|y\text{-}G(x,z)\|] \qquad (3)$$

The new objective function can be obtained by combining the original objective function with the distance function *L1* as follows:

$$L^{*} = L_{cGAN}(D,G) + \lambda L_{L1}(G) \qquad (4)$$

where λ denotes the penalty coefficient of the *L1* loss.

2.2 Generative networks: improving the U-net model

The U-Net network is a semantic segmentation convolutional neural network, comprising three parts—that is, an encoder, decoder, and skip connection (Olaf et al., 2015), as shown in **Fig. 2**. In contrast to a general convolutional neural network, the U-Net network adds upsampling while performing deconvolution to obtain a pixel-level segmented image. It requires less training data and is suitable for accurate segmentation of small targets.
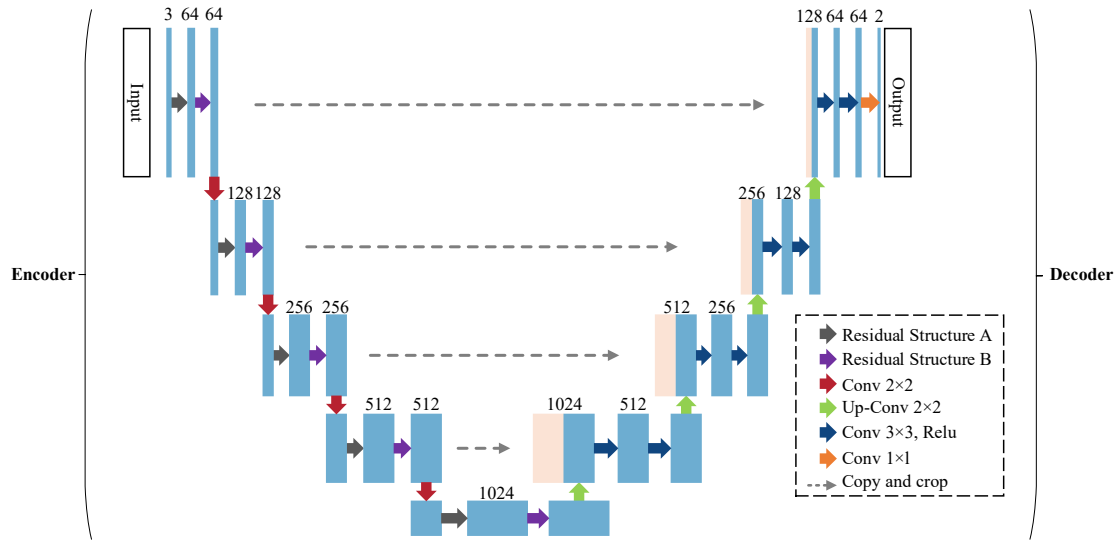
Fig. 2. The U-Net network model.

In this study, a pixel-level U-Net segmentation network was used as the generator, and the convolutional layer was replaced with an SE module to improve the model segmentation effect. The SE module was also added to the jump connection to enhance the model's ability to extract crack details, and a dropout layer was added to prevent the network from overfitting.

Fine crack mis-detection can be relatively easy in the difficult task of image feature recognition in concrete structures with complex backgrounds. In the construction of the GAN generative and discriminative networks, this study adopted a pixel-level U-Net structure as the generative network for the pixel-level recognition of crack edges.
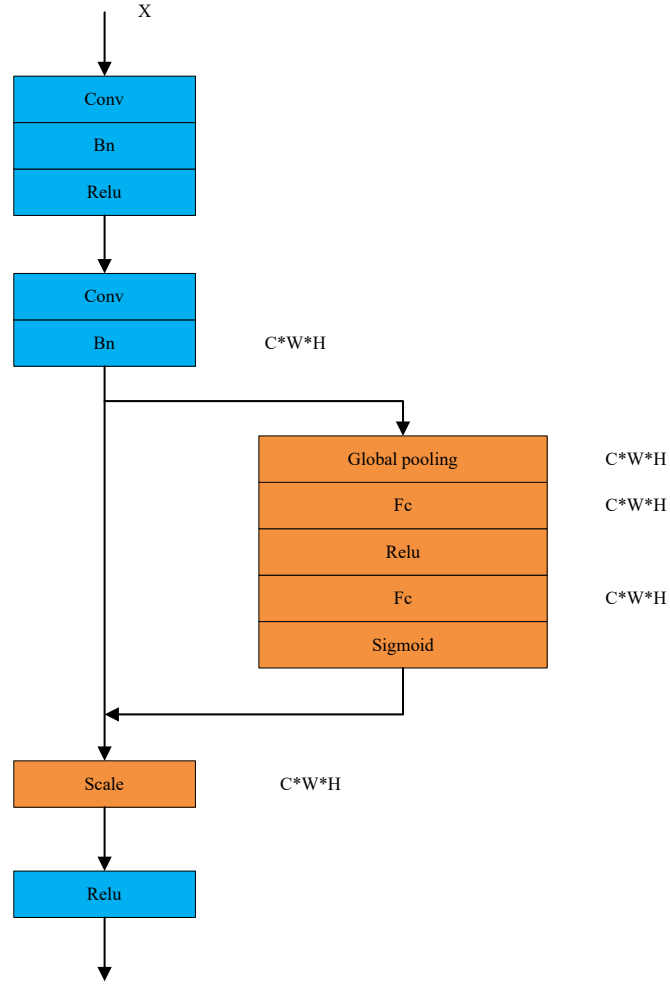
Fig. 3. The SE module.

To further solve problems such as the loss of crack-edge details during crack recognition, this study used the SE structure to recognize the image features. It weighs each feature channel based on the value of the feature image, enhances the weight of crack features, and reduces the weight of irrelevant features—such as the background—to improve the effect of crack feature recognition. Consequently, in this study, the original convolutional structure was replaced by a convolution and SE structure. The combined SE module is shown in **Fig. 3**.

The SE module uses two convolutions—that is, the first adjusts the number of channels and the resolution of the input feature image to the desired output size and the second is used to perform the SE operation.

First, after global pooling of the feature images, a fully connected network with a bottleneck structure is passed through, and a sigmoid function is used for the output of each neuron to obtain weights with a value range of (0,1), where the number of neurons is determined by the hyperparameter r (r being the compression multiplier of the neurons). Replacing the convolutional layer with an SE module—which contains a convolutional layer—and a convolutional layer with an SE structure—compared with a single convolutional layer—further improves the model's

recognition ability.

Second, to satisfy the need for multilevel feature learning in recognition tasks, the attention mechanism SE module is added to the jump-connected part of the U-Net structure to learn the multiscale features of the encoder and obtain the importance of each feature channel. Crack features are boosted based on the importance of each feature channel, and the weights of feature recognition are distributed to all stages of the contraction path to solve the problem of the U-Net relying on fixed weights in the process of image feature recognition to enhance the effect of feature recognition and improve the recognition accuracy of the model.

After adding the SE module, the U-Net model is as shown in **Fig. 4**, the U-Net encoder comprising the SE module with downsampling—the main function of which is to recognize features and expand the perceptual field—and the decoder consisting of the SE module with upsampling—the main function of which is to recognize features and expand the resolution. A jump connection exists between the encoder and the decoder, which splices the shallow features in the encoder with the deep features in the decoder, providing detailed information to the decoder and enabling it to obtain finer recognition boundaries.
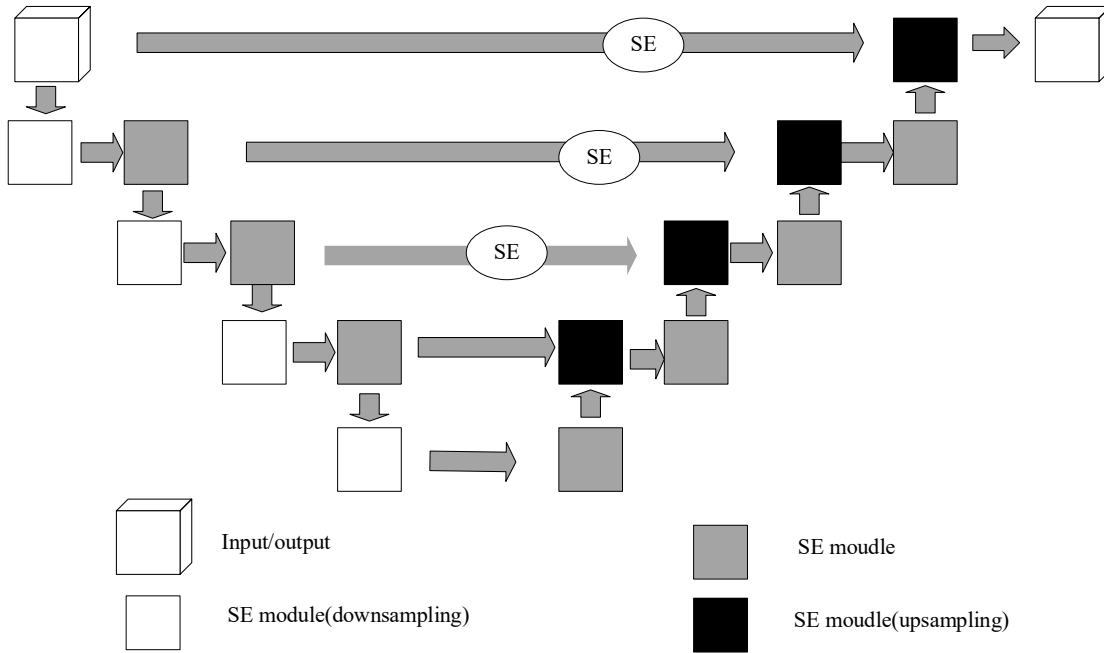
Fig. 4. The U-Net model added to the SE module.

2.3 Discriminative networks: PatchGAN

When selecting a discriminative network, different construction forms can be selected based on different data types. In the field of imaging, convolutional networks are often used to discriminate

between input data sources. To meet the demands of adversarial networks, focus on image details, and ensure high-resolution and high-definition details, the PatchGAN model was used as the discriminative network. The output image of the generative network and labeled image are used as inputs separately, and the image is convolved using a convolution kernel adapted to the input image to reduce the image area and identify the image features. A fully connected convolution was used to find the discriminant network values corresponding to the generative network sub-pest image and recognition label image. The discriminative network loss values can then be calculated using cross-entropy, the Ranger method being used for network learning optimization.

The U-Net was used as the generative network. After obtaining the generative network image recognition data, the PatchGAN multilayer convolutional network was used as the discriminative network model to maintain the high resolution and detail of the image. The output recognition image and the real labeled image of the generative network are used as the input, and the image is convolved using the convolution kernel adapted to the input image to reduce the image area and identify the image features. Each pixel of the real image and the generated image is then scored. Finally, all the pixel scores are averaged to determine the authenticity of the image. Fully-connected convolution is used to determine the discriminative network values corresponding to the generated network-recognized image and the recognized labeled image. In this study, a convolutional kernel with a step size of 2 and a length and width of 3 × 3 was selected for the four-layer convolutional network, the specific discriminative network structure of which is as shown in **Fig. 5**.

Although the number of parameters can be reduced by decreasing the number of convolutional kernels in the network, the number of eigenmaps generated per layer also decreases, leading to a decrease in the expressiveness of the network. In this study, we attempted to replace the stochastic gradient descent (SGD) algorithm with the Ranger optimizer. Stochastic gradient descent is one of the most widely used optimization algorithms in current neural networks, but the training speed of this algorithm can be slow, and the model can fall into local optimal solutions. For concrete crack-recognition applications, it is desirable for the model to exhibit a better generalization ability and faster training speed.
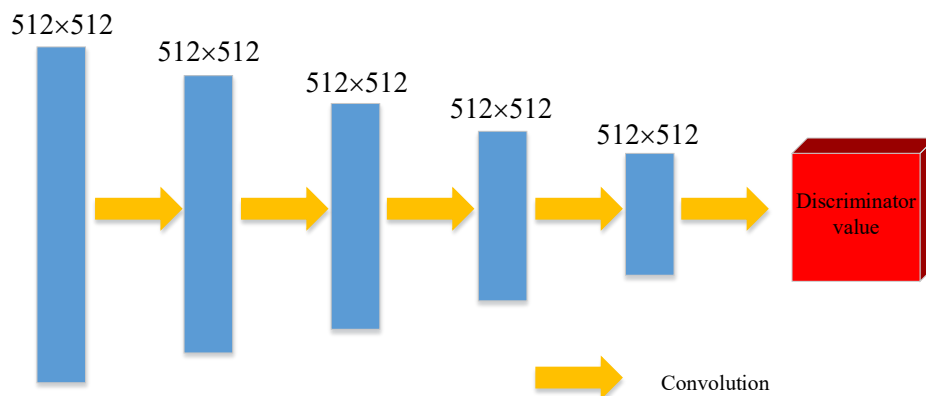
Fig. 5. The discriminator structure model.

During training, the Ranger optimizer updates the parameters during each iteration round based on the BP algorithm. Owing to the slender contour characteristics of cracks, the initial learning rate was ten in this study. Moreover, the smaller the loss, the smaller the deviation between the identification result and the true value. If the loss value is large or changes considerably, the parameters—that is, the initial learning rate, iteration rounds, and network depth—must be finely adjusted. The Ranger optimizer was used to dynamically adjust the learning rate during training so that the model could achieve a higher convergence rate.

2.4 The overall composition of the crack image segmentation network

In the adversarial learning process between the generator and discriminator, not only can the loss functions of the discriminator and generator be calculated, but the discriminator's ability to discriminate the segmentation result of the generated image as false can also be improved, while the labeled image can be fitted by the continuous optimization of the generator, which can make the discriminator discriminate it as true. The two are continuously played off against each other during the training, the adversarial update based on the discriminator and generator models being used as an update of the entire network after training to improve the generator image segmentation accuracy. At the end of the training process, accurately segmented images are obtained in the generator. The overall framework of the model is as shown in **Fig. 6**.
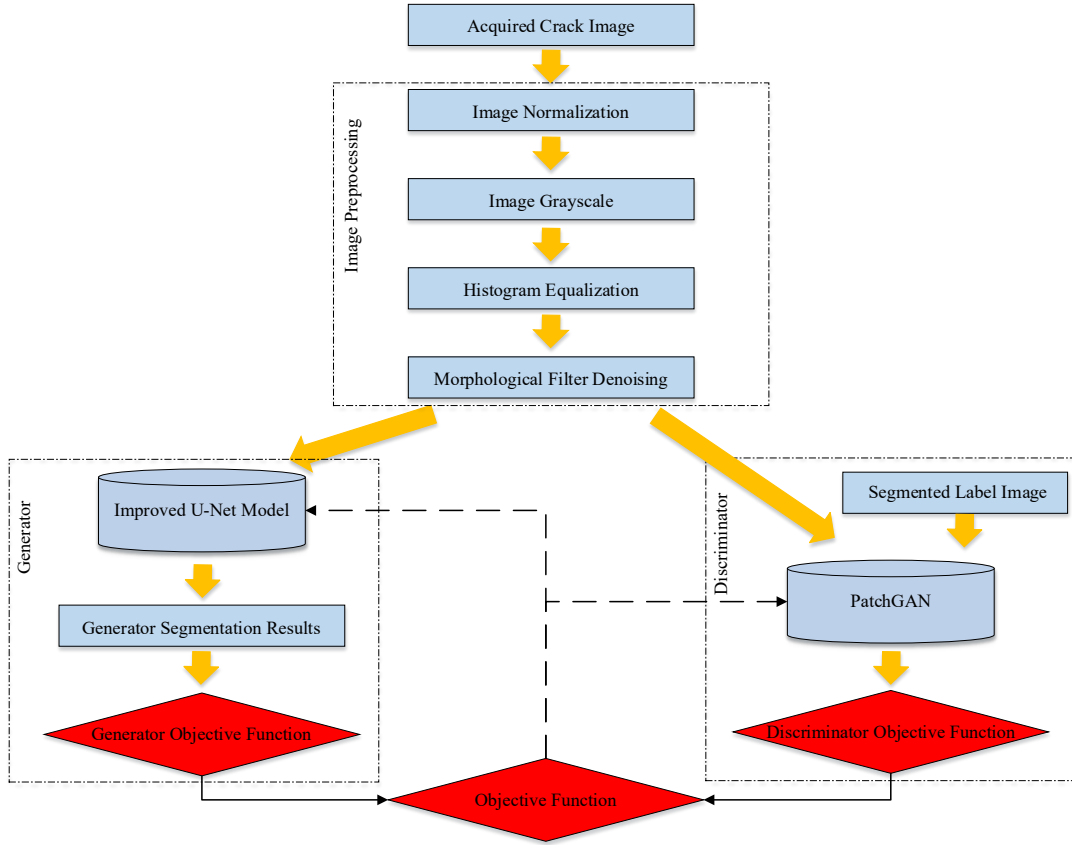
Fig. 6. The overall model framework.

3. Methods: Project examples

3.1 Crack database establishment and testing environment

In the field of image processing, it is critical to acquire and create high-quality images. CCD technology, a semiconductor imaging sensor, occupies an important position in solid-state arrays with a light-sensitive substrate comprising discrete silicon elements that can effectively capture subtle changes in the environment by injecting, transmitting, and detecting charges to provide accurate image data. CCD technology offers better performance, higher sensitivity, less image distortion, and faster response time, offering considerable advantages in crack detection. In this study, we considered this to be the best image-acquisition device.

To effectively suppress external interference, we used a light source as an important part of image acquisition and carefully studied its indicators, including brightness, distribution of the light source, distribution of the spectrum, age of use, and contrast. For example, if the light is too dim it can reduce the image contrast and be more likely to cause noise. Consequently, it is necessary to use light continuously to ensure image accuracy, selecting the light to ensure that it echoes the surroundings and better captures and displays all information. Based on the selection indicators mentioned earlier, as well as the characteristics of the concrete and precast components, we chose a light-emitting diode (LED) as the light source for image acquisition. LED light sources exhibit

excellent brightness, efficient luminescence, easily adjustable brightness, long life, and rapid response. Moreover, the light source comprises many LEDs, thus meeting the needs of a variety of illumination angles.

Before capturing images of cracks in precast concrete, it is important to ensure that the distance between the camera and member is reasonable, depending on the accuracy of the camera and crack detection. When cracks of size 0.2–0.3 mm appear in precast concrete members, the accuracy of the measurement must be lower than 0.2 mm to ensure its proper use. For this purpose, the pixel resolution was 1600 × 1200; therefore, we could obtain images that included the actual dimensions of the precast members—that is, 160 × 120 mm.
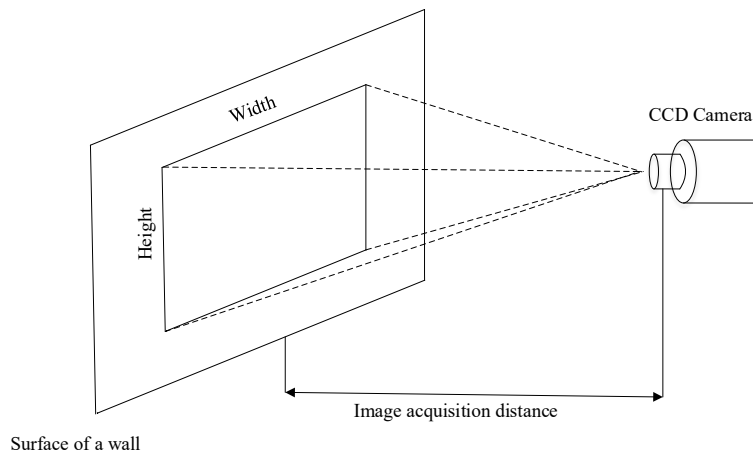
(a) Image-acquisition distance diagram.

(b) Precast concrete members.
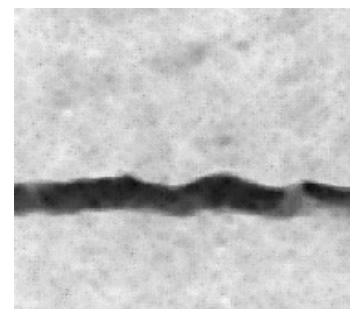
| (1) | (2) | (3) |
| Original image | Grayscale image | Multi-structure morphology filtering denoising |

(c)Equalization transformation of the crack image histogram.

Fig. 7. Crack image acquisition and processing

As shown in **Fig. 7(a)**, the following steps can be performed to determine the position of the camera:

1) Place a 160 or 120 mm rectangle on a blank sheet of paper and attach it to the wall.

2) The image-acquisition device is mounted in a horizontal position and the head is adjusted to ensure that it remains perpendicular to the blank sheet of paper with a rectangle drawn on it.

3) By adjusting the position of the image-acquisition device, the rectangle on the sheet of paper can be fully captured.

4) Calculate the average value of the distance between the paper and camera by measuring them several times.

Crack images of precast concrete were obtained from an engineering project in Tianjin, China. An industrial light source was used for illumination to maintain a uniform light intensity in the camera's field of view, and a CCD camera with 20 million pixels was used to capture 5000 raw images of cracks in the precast elements at a resolution of 4000 × 3000. Prefabricated elements included columns, beams, slabs, stairs, and balconies. **Fig. 7(b)** shows the precast concrete members.

To obtain a large number of crack images that met the input requirements of the model, the original images were cropped to obtain 5,000 images of resolution 512 × 512. Each image contained a sample of 262,144 pixels, the entire database containing a sample of 1,310,720,000 pixels. The crack images were edge-labeled with a 2-pixel width using the LabelMe software, black indicating the background and red indicating the crack—that is, (0 0 0) indicating a background pixel, and (255 0 0) indicating a crack pixel.

To train and validate the performance of the model, 80% of the images were used as the training dataset, 10% as the validation dataset, and 10% as the test dataset. Five thousand images were randomly disordered; 4000 images were classified as the training dataset, 500 images were classified as the validation dataset, and the remaining 500 images were used as the test dataset.

Because cracks in concrete structures can be easily disturbed by external light or electromagnetic waves, the quality of the crack images was degraded, using unclear edges and insufficient contrast. Consequently, we must first solve these problems by preprocessing to improve the accuracy of the subsequent recognition. Crack image preprocessing operations include image grayscale, image contrast enhancement, and image multi-structure morphology filtering and

denoising, as shown in **Fig. 7(c)**.

The test environment for this experiment is built using TensorFlow on windows with Intel(R) Core (TM) i7-3770 CPU @ 3.40GHz, training, validation, and testing will be executed on HP workstations configured with 8GB GPUs as Quadro P4000/PCle/SSE2. Anaconda is an environment creation software; it can be used to create different python environments for different scenarios. A virtual environment for the SE-U-Net network was created using Anaconda. CUDA 10.0 and CUDNN 7.4.1.5 are used to accelerate the training of the network.

3.2 Evaluation indicators

To quantitatively analyze the segmentation results of the model and evaluate the goodness of the crack-segmentation model for precast concrete components, the most commonly used evaluation metrics (Wang et al., 2021; Wu et al., 2021a), such as the pixel accuracy (PA), dice similarity coefficient (Dice), intersection over Union (IoU), and recall rate (Recall), for assessing the semantic segmentation effect were selected in this study.

$$PA = \frac{TP+TN}{TP+FP+TN+FN} \tag{5}$$

$$Dice = \frac{2TP}{FP+2TP+FN} \tag{6}$$

$$IoU = \frac{TP}{TP+FP+FN} \tag{7}$$

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

where PA is used to measure the accuracy of the model predictions; Dice is used to measure the degree of similarity between the model predicted values and the true values; IoU is used to measure the degree of overlap between the predicted results and the true annotations; and Recall is used to measure the completeness of the model predictions.

Description of the four parameters:

- True positive (TP): Predicted result is a positive class, the actual class is positive.

- False positive (FP): Actual negative class; the predicted result is a positive class.

- True negative (TN): Predicted result is a negative class, the actual class is negative.

- False negative (FN): Actual positive class; the predicted result is negative.

4. Results and Discussion

4.1 Model training

Instead of being classified into an entire network, each pixel was detected. The training was performed using one image as the batch size with a maximum of 50 epochs, the entire network

using a learning rate of 0.0005. A discard rate of 0.4 was used, and the network was optimally trained using Ranger gradient descent method. Network training was initiated after completing the hyperparameter configuration. There were 4000 images in the training dataset, and one image was used as the input for training in each epoch. Therefore, 4000 epochs were required to traverse the entire training dataset. During each iteration, the training accuracy and loss rate of the network were calculated. After each epoch, 30 images were randomly selected from the validation dataset to verify the performance of the network, and the validation accuracy and validation loss rate of the network were calculated. **Fig. 8(a)** and **Fig. 8(b)** show the accuracy and loss curves for the training and validation of the SE-U-Net, respectively. The red solid line represents the loss curve for the training dataset, and the blue dotted line represents the loss curve for the verification dataset. As the training period increases, the errors in the training and validation sets decrease and the segmentation accuracy improves, the training loss curves are identical to the validation loss curves. When the number of the epochs increases, both the accuracy and loss of the model no longer change significantly, and the proposed model reaches convergence.
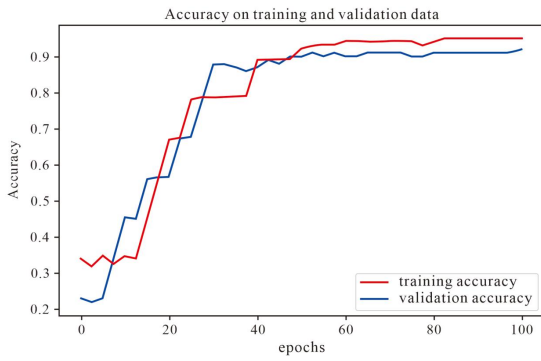
(a) Case of accuracy with epoch during training   (b) Case of loss rate with epoch during training
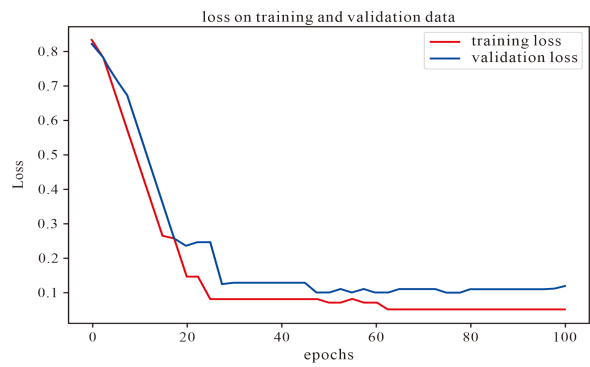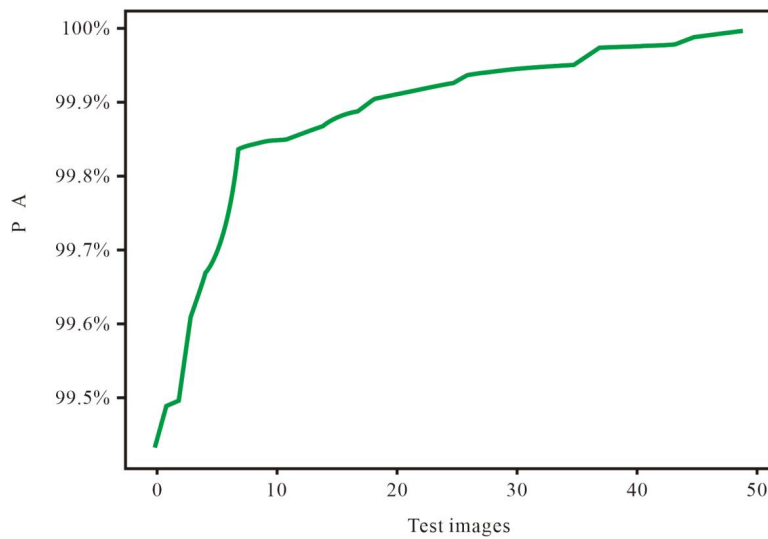
(c) The accuracy of the SE-U-Net network for the test images.
Fig. 8. The experimental results of the SE-U-Net network

After network training and verification, we viewed the model segmentation PA on the test dataset to evaluate the network performance. **Fig. 8(c)** shows the evaluation results of the SE-U-Net network on the test dataset with a learning rate of 10^{-4}. There were 500 images used in the test dataset, the PA being calculated for each image, reaching 98.48% after 50 epochs. These variations are shown in **Fig. 8(c)**.

4.2 Crack quantification calculation

To identify cracks more accurately, new methods needed to be developed to estimate their lengths and widths. These methods could be approximated by calculating the diagonal length of the outer rectangle of the crack, to identify regular or straight cracks more precisely. However, these methods could also lead to errors owing to the complexity of shape bending, preventing accurate results from being obtained. Consequently, a new method was proposed in this study to estimate the length L of the cracks as follows: After the crack skeleton was extracted, we replaced it with the width of each bit of data and determined its final length by calculating the geometric correction index, expressed as follows:

$$L = \int_c^C f(x,y)dl = \sum f(x,y)dl \tag{9}$$

$$W = \frac{\int_c^C f^2(x,y)ds}{L} = \frac{\sum f^2(x,y)ds}{\sum f(x,y)dl} \tag{10}$$

By calculating the length (L) of the crack skeleton—used for geometric correction—L can be used to determine the average width (W) of the crack based on an estimate of the crack pixel area and its ratio to the length (**Eq. (10)**).

When a single pixel width is used to represent the crack skeleton, the length of the crack can be obtained by summing the skeleton pixels, the crack area being the sum of all crack pixels and the average width being obtained by the ratio of area to length.

To verify the quantization performance of the model, crack quantization at the pixel level was considered. **Table 1** describes the procedure used to quantify the cracks. First, the established model is used to estimate the location of each point, morphological techniques are used to analyze the characteristics of each point to infer the length of each point, the overall area of each point is inferred from these values, and finally, the average width of each point is determined.

Table 1 Test set crack prediction results and quantitative analysis
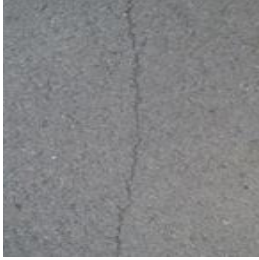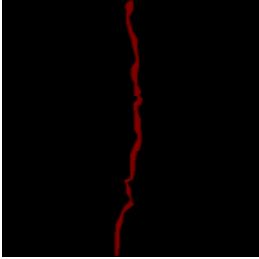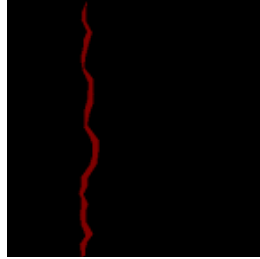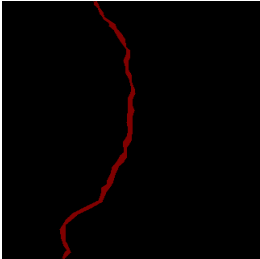
Input Image			
Predicted images			
Real Area	2310.19 *plxel*	2555.56 *plxel*	2655.56 *plxel*
Predicted area	2272.56 *plxel*	2323.25 *plxel*	2292.32 *plxel*
Real Length	665.52 *plxel*	654.23 *plxel*	563.20 *plxel*
Predicted Length	666.63 *plxel*	612.32 *plxel*	610.32 *plxel*
Real Width	3.41 *plxel*	3.51 *plxel*	3.69 *plxel*
Predicted Width	3.21 *plxel*	3.65 *plxel*	3.95 *plxel*

Table 1 lists the number of cracks predicted and quantified in the test dataset. We used a recognition model to estimate the sizes of the cracks and compute their skeletons. We also counted the number of pixels in each skeleton to derive their lengths. Finally, we summed the values of all the pixels to derive their average widths. Based on the quantitative analysis, we determined that the lengths of the cracks were closer to the predicted values of the area and average width, which also matched the true values of the labels.
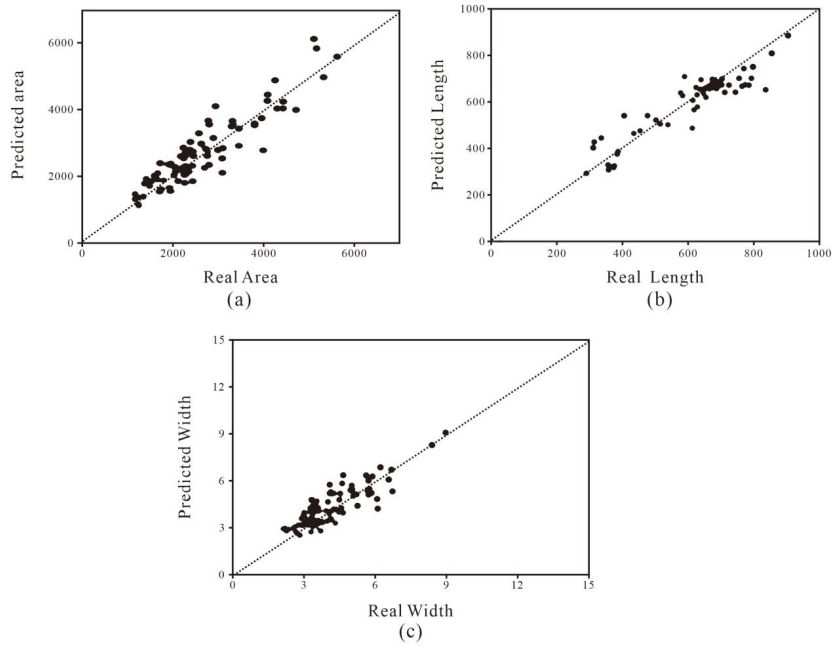
Fig. 10. Statistical results of crack geometry on the test dataset (a) area, (b) length, (c) average width.

Fig.10 shows the crack prediction metrics for a sample of 100 test dataset images, including the area, length, and width of the cracks; the statistics of these metrics matching the labeled dimensions. As is evident, except for a few anomalous prediction results, most of the data points are within the range of $y = x$, indicating that the prediction results do not differ much from the actual labeled results. Based on the statistics shown in **Fig. 10(a)** and **10(b)**, it is evident that the area of the cracks is concentrated between 3000 and 4000, whereas the length of the cracks is between 600 and 800. Additionally, **Fig. 10(c)** shows that the average width of the cracks is concentrated between 3 and 6.

5. Comparative study

Currently, with the development of convolutional neural networks, there are an increasing number of deep learning methods for image semantic segmentation, although the segmentation performance can exhibit obvious differences in the face of different tasks and methods. To further test the advantages and disadvantages of the real-time semantic segmentation methods for cracked images of assembled prefabricated components in this study, four image semantic segmentation models commonly used in current image semantic segmentation tasks—namely, DeepLabV3++, Mask-Rcnn, U-Net, and U-Net++ models—were introduced for comparison with the GAN U-Net model. The training parameters of these four models were the same as those of the GAN U-Net, and all were trained based on the strategy of automatically saving the optimal model before being tested against the test dataset. The recognition results and accuracy of different models are shown in Table 2.

Table 2 The recognition results and accuracy of different models
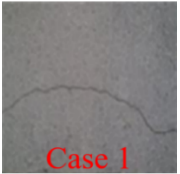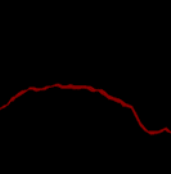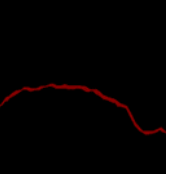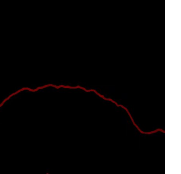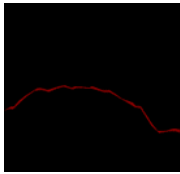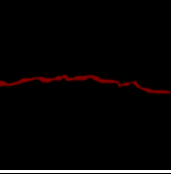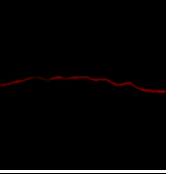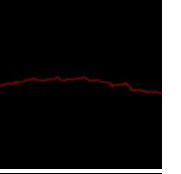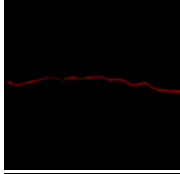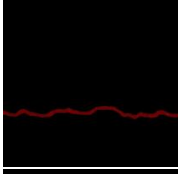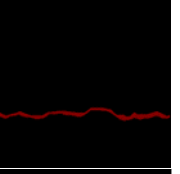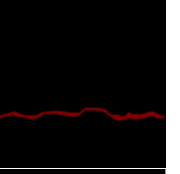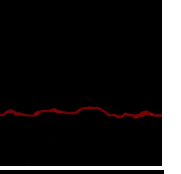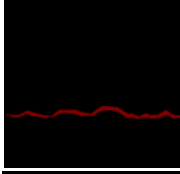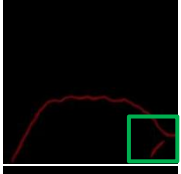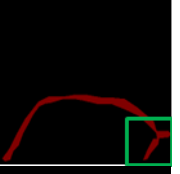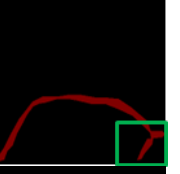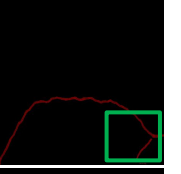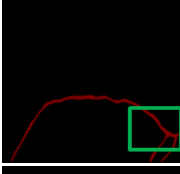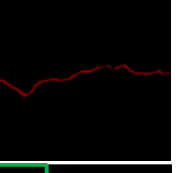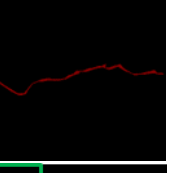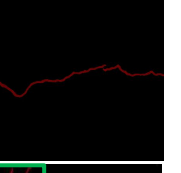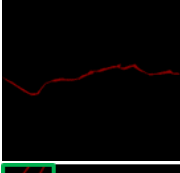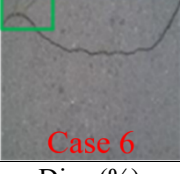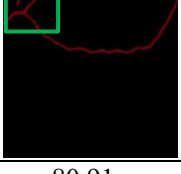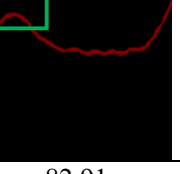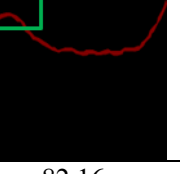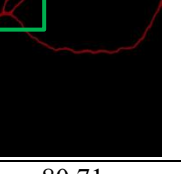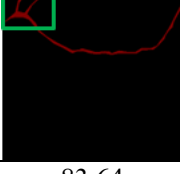
Original image	(a) DeepLabV3+	(b) Mask-Rcnn	(c) U-Net	(d) U-Net++	(e) Proposed method
Case 1					
Case 2					
Case 3					
Case 4					
Case 5					
Case 6					
Dice (%)	80.91	82.91	82.16	80.71	83.64
IoU (%)	71.92	94.73	95.68	71.71	94.80
Recall (%)	77.46	80.63	79.41	76.02	82.78
PA (%)	85.46	95.85	95.58	87.66	98.48

Table 2 lists the detection results as well as the accuracy of identifying six cracks with five different methods, through which it can be seen that although DeepLabV3+ can help the network to capture multi-scale information by increasing the sensory field, it still has a low recognition rate when dealing with small cracks due to the algorithmic limitations. U-Net++ captures more complex features by increasing the depth and width of the network, but it still has limitations in segmenting small areas of detail, especially when the cracks are similar to the texture of the surrounding environment. texture of the surroundings, so U-Net++ has a low recognition accuracy for small cracks. Mask-Rcnn over-segmentation and under-segmentation of the crack image processing by the Mask-Rcnn model are manifested by the fact that the crack-segmentation result is wider than the

original crack. This may be due to the low depth of the Mask-Rcnn. The sampling module in surface research is used to extract multiple branches and basic features, making it difficult to extract abstract image depth features for network teaching under a limited volume of crack data, which is unfavorable for feature localization in later stages. The crack image segmented by the U-net algorithm alone detects the cracks appearing in the image and can present the cracks completely; however, there are also cases in the detection result map where fine cracks are not segmented and small parts of the pixel points are missed or mis-segmented. Although the aforementioned methods have high accuracy, they are prone to over-segmentation, under-segmentation, discontinuous segmentation, and the loss of crack-edge details for pixel-level cracks with small boundaries. By contrast, the greater depth of the GAN-based U-Net model proposed in this study and the encoding-decoding structure make the model not only extract richer semantic features, but also join the GAN to form a confrontation through continuous gaming between the generator and discriminator during the training process, which improves the segmentation accuracy after repeated training. The proposed method improves the PA by 3%, recovers the edge information of the object, enhances the continuity between pixels, and ensures segmentation accuracy. Although in the crack detection experiments of Case 6 although U-Net++ could detect the edge fine cracks as well as the proposed method, it could not detect the edge fine cracks of Case 4. That is, the model proposed in this paper is able to accurately identify cracks up to 0.3 mm, improving the accuracy of small cracks without significantly reducing the results of normal width cracks. The Dice, IoU, Recall, and PA of the five models are shown in **Table 2**.

Based on the data summarized in **Table 2**, the proposed GAN-based U-Net segmentation method is more accurate than the other methods. The results of the algorithm proposed in this study were 83.64%, 94.80%, 82.78%, and 98.48% for the Dice, IoU, Recall, and PA, respectively. The accuracy, completeness, and similarity of the model prediction results to those of the real labeling process improved. Although the proposed network used in this study had a lower IoU index than the U-net model, it can be concluded that the overall performance was better and could achieve accurate segmentation of fine cracks by combining all the evaluation indexes.

Table 3 Ablation experiment

Method	Dice (%)	IoU (%)	Recall (%)	PA (%)
w/o SE module	82.73	95.91	82.11	96.42
Proposed method	83.64	94.80	82.78	98.48

In order to evaluate the effectiveness of the SE module in the crack segmentation task, a series of ablation experiments were designed in this study. After eliminating the SE module, the method was renamed as "w/o SE module". The experimental results are shown in **Table 3**. Compared with the

version without SE module, the addition of SE module improves the Dice coefficient of the model by 0.91 percentage points to 83.64%, although the IoU is slightly reduced by 1.17 percentage points to 94.80%. Meanwhile, the Recall and PA of the model had an increase of 0.67 percentage points and 0.02 percentage points to 82.78% and 98.48%, respectively. These results indicate that the SE module can enhance the network's ability to recognize crack features to a certain extent, especially in improving the accuracy of segmentation significantly.

6. Conclusions

A CGAN-based SE-U-Net model has been developed for identification of fine cracks in concrete structures. The proposed model used a CGAN as its architecture, a modified U-Net network to design the generative network, and an attention mechanism SE module to enhance the weights of the crack features and suppress the weights of the background features to improve the network's ability to capture crack details. PatchGAN was used as the discriminant network to score each pixel of the real and generated images and was combined with the Ranger optimization algorithm to solve the problem of too many parameters in the optimization process, making the model converge quickly and improving its recognition efficiency and accuracy. Through repeated game training of the generative and discriminative networks, the crack-recognition image generated by the generative network was very close to the real image, thus realizing the effect of the generative network in recognizing cracks with high accuracy. The results showed that the proposed method could accurately and effectively identify fine cracks in concrete structures. Some conclusions are obtained as follows,

1) Compared with traditional recognition algorithms and other models, the results showed that the proposed method could retain detailed crack-edge information for fine cracks of pixel-level width and near-fine extreme cracks and the PA rate reaching 98.48%.

2) The crack skeleton was represented by a single pixel width, identified by binary morphology. The area, length, and average width of the cracks could be calculated, providing useful information for the damage assessment of building structures.

Acknowledgements

Conflicts of Interest

The authors declare no potential conflict of interests.

References

ACI committee (1999) ACI committee institute, Building code Requirements for structural concrete and commentary[S], ACI318—99

Chen B, Zhang X, Wang R, Li Z, Deng W (2019) Detect concrete cracks based on OTSU algorithm with differential image. *The Journal of Engineering* 2019(23): 9088-9091, DOI: 10.1049/joe.2018.9191

Chen J-Y, Zheng H, Cheng L, Lai S (2017) Bridges Disease Detection Based on k-means and Probabilistic Relaxation. *Control Engineering of China* 24(S1), DOI: 10.14107/j.cnki.kzgc.160171

Chen Z, Luo W, Li L (2020) Automatic Identification of Pavement Crack Using Improved U net Model. *Journal of Data Acquisition and Processing* 35(2): 260-269, DOI: 10.16337/ j.1004 -9037.2020.02.007

Chen Z, Zhang J, Zhang Y, Huang Z (2021) Traffic Accident Data Generation Based on Improved Generative Adversarial Networks. *Sensors* 21(17): 5767, DOI: 10.3390/S21175767

Cui F, Li Z, Yao L (2014) Images crack detection technology based on improved K-means algorithm. *journal of multimedia* 9(6): 822, DOI: 10.4304/jmm.9.6.822-828

Feng W, Sui H, Huang W, Xu C, An K (2019) Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a super pixel-based conditional random field model. *IEEE Geoscience and Remote Sensing Letters* 16(4): 618-622, DOI: 10.1109/LGRS.2018.2879492

Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, Lu H (2019) Dual attention network for scene segmentation. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3146-3154).

Gao Q, Almekkawy M (2021) ASU-Net++: A nested U-Net with adaptive feature extractions for liver tumor segmentation. *Computers in Biology and Medicine* 136:104688, DOI: 10.1016/J.COMPBIOMED.2021.104688

Gao Z, Yang G, Li E, Shen T, Wang Z, Tian Y, et al. (2019) Insulator segmentation for power line inspection based on modified conditional generative adversarial network. *Journal of Sensors* 2019, DOI: 10.1155/2019/4245329

Gatys LA, Ecker AS, Bethge M (2016) Image style transfer using convolutional neural networks. Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2414-2423, DOI: 10.1109/CVPR.2016.265

Hao S, Zhang L, Qiu K, Zhang Z (2022) Conditional Generative Adversarial Network for Monocular Image Depth Map Prediction. Electronics (5). DOI: 10.3390/ELECTRONICS12051189.

He Z, et al (2019) Image Recognition with Conditional Wasserstein Generative Adversarial Networks with Gradient Penalty. *Computer Measurement & Control* 27(6):157-162, DOI: 10.16526/j.cnki.11-4762/tp.2019.06.034

Hoang N-D (2018a) Detection of surface crack in building structures using image processing technique with an improved Otsu method for image thresholding. *Advances in Civil Engineering* 2018, DOI: 10.1155/2018/3924120

Hoang N-D, Nguyen Q-L (2018b) Metaheuristic optimized edge detection for recognition of concrete wall cracks: a comparative study on the performances of Roberts, Prewitt, Canny, and Sobel algorithms. *Advances in Civil Engineering* 2018, DOI: 10.1155/2018/7163580

Huo Y, Xu Z, Bao S, Bermudez C, Plassard AJ, Liu J, et al. (2018) Splenomegaly segmentation using global convolutional kernels and conditional generative adversarial networks.

Cardiovascular Journal 10574:1057409, DOI: 10.1117/12.2293406

Kaur M, Sarkar RK, Dutta MK (2021) Investigation on quality enhancement of old and fragile artworks using non-linear filter and histogram equalization techniques. *Optik-International Journal for Light and ElectronOptics* 244:167564, DOI: 10.1016/J.IJLEO.2021.167564

Lei B, Wang N, Xu P, Song G (2018) New crack detection method for bridge inspection using UAV incorporating image processing. *Journal of Aerospace Engineering* 31(5):04018058, DOI: 10.1061/(ASCE)AS.1943-5525.0000879

Li L, Sun L, Ning G, Tan S (2014) Automatic pavement crack recognition based on BP neural network. *PROMET-Traffic&Transportation* 26(1):11-22.

Li W, Chen H, Zhang Y, Shi Y (2021) Track slab crack detection based on full convolutional neural network. *Annals of health law* 1848(1):012163, DOI: 10.1088/1742-6596/1848/1/012163

Lian S, Luo Z, Zhong Z, Lin X, Su S, Li S (2018) Attention guided U-Net for accurate iris segmentation. *Journal of Visual Communication and Image Representation* 56:296-304, DOI: 10.1016/j.jvcir.2018.10.001

Liu F, Wang J, Chen Z, Xu F (2021) Parallel Attention Based UNet for Crack Detection. *Journal of Computer Research and Development* 58(8):1718.

Liu J, Wu W (2021) Automatic Image Annotation Using Improved Wasserstein Generative Adversarial Networks. *IAENG International Journal of Computer Science* 48(3).

Liu SJ, Li SL, Jiang M, He D (2017) Quantitative identification of pipeline crack based on BP neural network. *Key Engineering Materials* 737:477-480, DOI:10.4028/www.scientific.net/KEM.737.477

Liu Z, Song Y-Q, Sheng VS, Wang L, Jiang R, Zhang X, et al. (2019) Liver CT sequence segmentation based with improved U-Net and graph cut. *Expert Systems with Applications* 126:54-63, DOI: 10.1016/j.eswa.2019.01.055

Mirza M, Osindero S (2019) Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Moghnieh H, Kamran S A, Hossain K F, et al. (2022) Software for segmenting and quantifying calcium signals using multi-scale generative adversarial networks. STAR Protocols (4). DOI: 10.1016/J.XPRO.2022.101852.

Olaf R, Philipp F, Thomas B (2015) U-net: Convolutional networks for biomedical image segmentation. *Cornell University*, DOI: 10.1007/978-3-319-24574-4_28

Sabeenian R S, Vijitha V (2021) A Generative Adversarial Network Based Segmentation and Classification using Deep CNN: Systematic Review. *IOP Conference Series: Materials Science and Engineering*, 1084, 012012. DOI: 10.1088/1757-899X/1084/1/012012.

Sang L，Gao S，Yin Z (2019) Motion Deblurring Based on Generative Adversarial Networks. *Computer Engineering and Applications* 55(06):173-177.

Shelhamer E, Long J, Darrell T (2015) Fully Convolutional Networks for Semantic Segmentation. *Transactions on Pattern Analysis and Machine Intelligence* 39(4): 640-651, DOI: 10.1109/CVPR.2015.7298965

Song Y, Wang J, Guo F, Lu J, Liu S (2021) Research on Supplier Selection of Prefabricated Building Elements from the Perspective of Sustainable Development. *Sustainability* 13(11): 6080, DOI: 10.3390/SU13116080

Wang Y, Bai X, Wu L, Zhang Y, Qu S (2021) Identification of maceral groups in Chinese bituminous coals based on semantic segmentation models. *Fuel* 308:121844, DOI:

10.1016/J.FUEL.2021.121844

Wu H, Liu J, Chen G, Liu W, Hao R, Liu L, et al. (2021a) Automatic Semicircular Canal Segmentation of CT Volumes Using Improved 3D U-Net with Attention Mechanism. *Computational Intelligence and Neuroscience* 2021, 9654059. DOI: 10.1155/2021/9654059

Wu Y, Yang W, Pan J, Chen P (2021b) Asphalt pavement crack detection based on multi-scale full convolutional network. *Journal of Intelligent & Fuzzy Systems* 40(1):1495-1508, DOI: 10.3233/JIFS-191105

Xu Y, Bao Y, Chen J, Zuo W, Li H (2019) Surface fatigue crack identification in steel box girder of bridges by a deep fusion convolutional neural network based on consumer-grade camera images. *Structural Health Monitoring* 18(3), 653-674.

Xu Y, Fan Y, Bao Y, et al. (2023a) Task-aware meta-learning paradigm for universal structural damage segmentation using limited images[J]. *Engineering Structures* 284: 115917.

Xu Y, Fan Y, Li H (2023b) Lightweight semantic segmentation of complex structural damage recognition for actual bridges. *Structural Health Monitoring* 22(5), 3027-3046. doi.org/10.1177/14759217221147015.

Xu Y, Li S, Zhang D, et al. (2018) Identification framework for cracks on a steel structure surface by a restricted Boltzmann machines algorithm based on consumer-grade camera images. *Structural Control and Health Monitoring* 25(2): e2075.

Xu Y, Zhu B, Zhang Z, Chen J (2022) Hierarchical dynamic Bayesian network-based fatigue crack propagation modeling considering initial defects. *Sensors* 22(18), 6777.

Yang J, Wu B, Li L, Cao P, Zaiane O (2021) MSDS-UNet: A multi-scale deeply supervised 3D U-Net for automatic segmentation of lung tumor in CT. *Computerized Medical Imaging and Graphics* 92: 101957, DOI: 10.1016/J.COMPMEDIMAG.2021.101957

Yıldız E, Arslan AT, Taş AY, Acer AF, Demir S, Şahin A, et al. (2021) Generative Adversarial Network Based Automatic Segmentation of Corneal Subbasal Nerves on In Vivo Confocal Microscopy Images. *Translational Vision Science & Technology* 10(6): 33-33, DOI: 10.1167/TVST.10.6.33

Zhao J, Hu F, Qiao W, Zhai W, Xu Y, Bao Y, Li H (2022) A modified u-net for crack segmentation by self-attention-self-adaption neuron and random elastic deformation. *Smart Structures and Systems*, 29(1), 1-16.

Zhang K, Yang X, Zhang J (2021) Deep Learning Optimization Algorithm based on High Order Exponential Smoothing Dynamic Boundary Constraint. *Information and Control*, 50(6): 685-693. DOI: 10.13976/j.cnki.xk.2021.0522

Zhang L, Yang F, Zhang Y D, Zhu Y J (2016) Road crack detection using deep convolutional neural network. *In 2016 IEEE international conference on image processing (ICIP)* (pp. 3708-3712). IEEE.

Zhang Y, Ding M, Bai Y, Ghanem B (2019) Detecting small faces in the wild based on generative adversarial network and contextual information. *Pattern Recognition* 94:74-86, DOI: 10.1016/j.patcog.2019.05.023

Zhang Z, Liu Q, Wang Y (2018) Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters* 15(5): 749-753, DOI: 10.1109/LGRS.2018.2802944

Zhu H, Zhang X, Shi Y, Li X, Sun Y (2021) Automatic segmentation of rectal tumor on diffusion‑weighted images by deep learning with U‑Net. *Journal of Applied Clinical Medical Physics* 22(9): 324-331, DOI:10.1002/ACM2.13381