Image-based crack recognition of tunnel lining using residual U-Net convolutional neural network

S K Hou\(^1\), Z G Ou\(^1,2\), P X Qin\(^3\), Y L Wang\(^4\), Y R Liu\(^1\)*

\(^1\)State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing 100084, China
\(^2\)State Key Laboratory of Stimulation and Regulation of Water Cycles in River Basins, China Institute of Water Resources and Hydropower Research, Beijing 100038, China
\(^3\)Huaneng Yarlung Zangbo River Hydropower and Investment Co., Ltd, Chengdu 610093, China
\(^4\)Hebei Institute of Water Science, Shijiazhuang 050051, China

Corresponding author’s e-mail: liuyaoru@tsinghua.edu.cn

Abstract. Concrete crack is a common disease of lining segments in tunnel projects, which will cause adverse risks without being repaired in time. Currently, various convolutional neural networks (CNNs) have been successfully applied in the computer vision field. An image-based crack recognition model of tunnel lining using Residual U-Net (ResU-Net) network is proposed. The residual learning units are added to the encoding path of the U-Net network to solve the problem of model degradation. Based on a highway tunnel project in western China, a dataset of lining crack is built. And through the size adjustment, annotation and binary processing, 2880 image samples with the same size of 448×448 is obtained. The dataset is divided into training set, validation set and test set with the ratio of 6:1:1. The quantitative results show that three evaluation metrics of pixel accuracy (PA), intersection over union (IoU) and Dice coefficient (Dice) are 98.67%, 56.45% and 68.09%, respectively, which is better than that of typical U-Net. It indicates that the ResU-Net model has good performance and robustness in the pixel level segmentation of tunnel lining cracks.

1. Introduction
In recent years, many projects such as tunnels, bridges and roads have gradually shifted from the construction stage to the operation and maintenance stage in China [1]. With the increase of the operation time, and affected by the deterioration of concrete material, design and construction defects, untimely maintenance etc., many concrete structure projects have appear various diseases, which limit the use performance and safety [2]. A large number of statistical data show that lining cracking is one of the most common disease in tunnel engineering [3]. The occurrence of cracks will affect the integrity and reliability of the structure to a certain, and even cause serious safety accidents without limiting its expansion. Therefore, it is necessary to carry out regular structural health monitoring (SHM) to ensure the long-term safety and stability of the tunnel structure.

The tradition lining crack detection of tunnels is mainly rely on the manual detection, which requires qualified inspectors to enter the tunnel and inspect the lining crack by their naked eye. The efficiency of this method is low, and the accuracy of crack identification depends on the experience and judgment of the inspector to a large extent [4]. Currently, tunnel disease diagnosis based on the images and computer vision (CV) technology has become a research hotspot. The tradition CV methods used for crack recognition include the fuzzy C-means clustering [5] and edge detection [6] etc., which have...
been proved to be effective. However, the traditional CV methods have strong dependence on images quality and their generalization capability and adaptability are relatively poor [7]. Moreover, due to the complex environment of the tunnel lining surfaces, the shooting conditions are poor with the problems of low contrast, uneven illumination and serious noise, which makes it difficult to construct an effective detection model using the traditional CV methods. The development of deep learning techniques brought a breakthrough for tunnel crack recognition. The convolutional neural network (CNN) can directly extract appropriate features from the two-dimensional image automatically and shows the outstanding performance in the field of images analysis, such as object defection, semantic segmentation and instance segmentation [8].

In this study, an image-based crack recognition model of tunnel lining using Residual U-Net (ResU-Net) network is proposed. By adding several residual learning units to the convolutional layers of the encoding path in the U-Net network, the training degradation problem are improved, and the semantic segmentation of crack defects is realized. Based on a highway tunnel in western China, the images of lining segments with crack were collected. Through the image size adjustment, the image is processed to the same size of 448×448. Then, the cracks of the tunnel lining segments are annotated and the crack binary segmentation image is obtained through the image conversion code. The pixel value of the crack area is 255 (i.e., white), and the pixel value of the non-crack area is 0 (i.e., black). Finally, a dataset containing 2880 lining crack images are obtained. The training set, verification set and test set are divided by random sampling with the ratio of 6:1:1, respectively. The mask prediction performance of the ResU-Net model and the comparison with the U-Net model are analyzed.

2. Methodology

2.1. Semantic segmentation

In the field of computer vision, there are two common methods to recognize structural surface defects by convolutional neural networks (CNNs), including object defection and semantic segmentation [9]. The purpose of object defection is to find out the boundary box and assign a category label for each defect in the image. However, the object defection is relatively rough. In order to obtain the geometric features such as the width, length and area of the target surface defect, it is necessary to locate the defect more accurately. The task of semantic segmentation is to classify the pixels of the image and extract pixels of defects form the image.

2.2. U-Net network

CNN model can automatically extract the multi-layer features of the two-dimensional image, and classify the image pixels and recognize the targets. To overcome the limitations of the CNN in fine image segmentation, Long et al. [10] proposed the full convolutional network (FCN). Compared with CNN, FCN changes the final full connection layer to convolution layer with convolution kernel size of 1×1, which improves the accuracy of semantic segmentation and enables the model to accept any size of input.

U-Net network is a classic improved fully convolutional neural network (FCN) used for semantic segmentation [11]. The basic structure of U-Net is composed of an encoder part and a decoder part. The main characteristics of U-Net is adopting the skip connection idea, which can concatenate the encoding features to the decoding features, making it have a good performance for the small training dataset.

2.3. Residual U-Net

According to the experience, the deeper the neural network is, the more complex features can be extracted, which means the better prediction performance. However, many studies have found that the deeper layers may also lead to the model degradation [12]. Therefore, He et al. [13] proposed the residual neural network to solve the problem caused by the increase of the network depth. The schematic diagram of residual learning unit is shown as Figure 1. A residual learning unit includes two 3×3 convolutional layers with a stride of 2 and two batch normalization (BN) layers. The first
convolutional layer is following a BN layer and a rectified linear unit (ReLU) activation layer. The $i$-th residual input vectors can be consider in the $(i+1)$-th residual learning unit.

**Figure 1.** Schematic diagram of residual learning unit.

In the study, by adding several residual learning units to the convolutional layers of the encoding path in the U-Net network, a residual U-Net (ResU-Net) network is established [14]. The architecture of the ResU-Net network is shown as Figure 4. The left part of the ResU-Net is the encoding path, which includes an input block, a head block, four residual blocks and a maximum pooling block. The head block includes a $7 \times 7$ convolutional layers with a stride of 1, followed by a BN layers, a ReLU activation layer and a maximum pooling layer. The next four residual blocks are composed of 3, 4, 6 and 3 residual learning units, respectively. The right part of the proposed model is the decoding path, which is composed of six up-sampling blocks, two addition blocks and an output block. Each up-sampling block is composed of a $3 \times 3$ convolutional layer with a stride of 1 and an up-sampling layer. The convolutional kernel size of each up-sampling layer is $4 \times 4$ with a stride of 2. Furthermore, the concatenation between the first four up-sampling blocks and the corresponding four residual blocks is established (see Figure 2). The first addition block includes a $3 \times 3$ convolutional layer with a stride of 1 and a ReLU activation layer. The second addition block includes a dropout layer, a $1 \times 1$ convolutional layer and a Sigmoid activation layer.
2.4. Evaluation metrics

Confusion matrix is a common visualization tool to describe the mask prediction accuracy of neural network. Table 1 shows the confusion matrix of the binary classification problem. For the prediction accuracy evaluation of image segmentation, it is necessary to compare the correctness of each pixel point in the real image and the predicted image. The columns of the confusion matrix represent that the predicted values of pixels are positive (indicating cracks) or negative (indicating non-cracks), and the rows of the confusion matrix represent that the real values of image pixels are positive or negative.

| Actual value | Predicted value |                   |
|--------------|-----------------|-------------------|
|              | Positive        | Negative          |
| Positive     | TP              | FN                |
| Negative     | FP              | TN                |

Pixel accuracy \( (PA) \) is the simplest metrics to evaluate the model performance of image segmentation, which means the proportion of correctly predicted pixels (including positive and negative) in the total pixels. The higher the value of pixel accuracy, the better the mask prediction performance of the model. Pixel accuracy can be calculated as follows:

\[
PA = \frac{TP + TN}{TP + FP + FN + TN}
\]

Intersection over union \( (IoU) \) refers to the ratio of intersection and union between predicted image pixels and the real image pixels. The larger the \( IoU \) is, the higher the coincidence degree between the predicted image and the real image is. The calculation formula of \( IoU \) is shown as follows:

\[
IoU = \frac{TP}{TP + FP + FN}
\]

Dice coefficient \( (Dice) \) refers to the similarity between the two pixel sets of the predicted image and the real image, which is positively correlated with \( IoU \). The calculation formula of Dice coefficient is as follows:

\[
Dice = \frac{2TP}{2TP + FP + FN}
\]

3. Image collection and preprocessing

In this study, an image dataset of the tunnel lining are established. The images in the dataset were collect from a tunnel projects in western of China. Figure 3 shows the scene photographs of the DXL tunnel project for image collection. DXL tunnel is a highway tunnel with the diameter of 8100 mm. The total length of the tunnel is about 4.78 km, of which 4489 m is constructed by double-shield TBM and support by lining segments. During the construction process, there appeared many cracks on the surface of the lining segments.

A representative data set is the basis of machine learning of deep learning. Generally, the larger the number of samples in the dataset and the more accurate the sample annotation is, the better the training effect and the prediction performance of the model are. In this study, an image dataset of lining crack is constructed by taking photos in the tunnel site. Firstly, each image is processed to the same size of 448×448. Then, the cracks of the tunnel lining segments are annotated by the image annotation tool LabelMe, and the crack binary segmentation image is obtained through the image conversion code. Figure 4 shows an example of the tunnel lining crack annotation process. The pixel value of the crack...
area is 255 (i.e., white), and the pixel value of the non-crack area is 0 (i.e., black). Finally, 2880 tunnel lining crack images with annotated labels are obtained.

![Figure 3. Scene photographs of the DXL tunnel project for image collection.](image)

The established tunnel lining crack dataset contains the condition of different light and different noise interference, which has strong representativeness. The dataset is divided into training set, validation set and test set with the ratio of 6:1:1 by random sampling method. The original image and the binary annotated image are taken as the input and output of the ResU-Net model during the model training process.

4. Experimental results

4.1. Mask prediction results analysis

In this section, the 2160 image samples in training set are used to training the ResU-Net model, and 360 image samples in the validation are used to tune the hyper-parameters of the model. Then, based on the optimal hyper-parameters, the established ResU-Net model is finally trained. Figure 5 shows the change of model training loss and validation loss with training epoch. In this study, cross entropy loss function is used to measure the training effect. It can been seen that the train loss and validation loss of the model decrease gradually with the increasing of epochs, and the loss is basically in a stable value when epoch is greater than 5. The value of validation loss is slightly larger than the value of training loss, which shows that the training process is reasonable.

![Figure 4. An example of the tunnel lining crack annotation process.](image)

Figure 6 shows the mask prediction results of some images in the test set. It can be seen form the figure that the predicted masks are basically consistent with the crack location. Additionally, the established ResU-Net model has strong anti-interference ability to some noise interference of the project site, such as reserved grouting hole, segment joint and segment logo etc. And the established model also has good applicability for image crack segmentation under different environment background and brightness conditions.
4.2. Model comparison

U-Net is the commonly used and effective model for semantic segmentation, and is often used as the comparison with various improved FCN models [9, 15]. In this section, a typical U-Net network model is established and compared with the ResU-Net. Then, using the same training set, validation set and test set as ResU-Net model to train and test the U-Net model, which can make the comparison results more convincing. Table 2 shows the mask prediction performance comparison of different models. It can be drawn from the table as follows:

(1) The pixel accuracy ($PA$) of the ResU-Net model and U-Net model are close and both are above 98.5%, which are 98.67% and 98.60%, respectively. It shows that the two models have good recognition effect on tunnel lining cracks, and can effectively recognize cracks, non-crack parts and various noise interferences in the image.

(2) The $IoU$ and $Dice$ of ResU-Net is 56.45% and 68.09%, respectively, which is obviously higher than that of U-Net. It shows that the ResU-Net model is more sensitive to the location of cracks, and can predict the location of cracks more accurately.

| Model   | $PA$ (%) | $IoU$ (%) | $Dice$ (%) |
|---------|----------|-----------|------------|
| ResU-Net | 98.67    | 56.45     | 68.09      |
| U-Net   | 98.60    | 51.06     | 62.58      |
5. Conclusion
In this paper, an image-based crack recognition model of tunnel lining using Residual U-Net (ResU-Net) network is proposed, which is established by integrating the residual learning units to the encoder path of the U-Net. A tunnel lining crack image dataset is collected based on a highway tunnel in Western China. Through the size adjustment, annotation and binary processing, a total of 2880 lining crack image samples were obtained to training and test the established model. The mask prediction results shows that the established ResU-Net has the ability of accurate crack segmentation and strong anti-noise interference. The pixel accuracy ($PA$), intersection over union ($IoU$) and Dice coefficient ($Dice$) of ResU-Net model on test set are 98.67%, 56.45% and 68.09%, respectively, which shows the better performance than the typical U-Net. The obvious higher $IoU$ and Dice coefficient also indicates that the ResU-Net model is more sensitive to the location of cracks, and can be effectively used for the semantic segmentation of tunnel lining cracks.

Acknowledgements
This work was supported by the State Key Laboratory of Hydrosience and Engineering with Grant No. 2019-KY-03.

References
[1] Lei M F, Liu L H, Shi C H, Tan Y, Lin Y X and Wang W D 2021 A novel tunnel-lining crack recognition system based on digital image technology Tunnelling and Underground Space Technology 108 103724
[2] Yang Y, Wang L F, Zhang Y F and Han X J 2021 Multi-feature fusion based classification algorithm of surface disease image of concrete structure Journal of Chang’an University: Natural Science Edition 41 64–74 (in Chinese)
[3] Huang H W, Li Q T and Zhang D M 2018 Deep learning based image recognition for crack and leakage defects of metro shield tunnel Tunnelling and Underground Space Technology 77 166–176
[4] Koch C, Georgieva K, Kasireddy V, Akinci B and Fieguth P 2015 A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure Advanced Engineering Informatics 29 196–210.
[5] Nayak J, Naik B and Behera H 2015 Fuzzy C-means (FCM) clustering algorithm: a decade review from 2000 to 2014 Computational intelligence in data mining-volume 2 133–149
[6] Abdel-Qader I, Abudayyeh O and Kelly M E 2003 Analysis of edge-detection techniques for crack identification in bridges Journal of Computing in Civil Engineering 17 255–263
[7] Xu Y Y, Li D W, Xie Q, Wu Q Y and Wang J 2021 Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN Measurement 178 109316.
[8] Goodfellow I, Bengio Y, Courville A and Bengio Y 2016 Deep learning (Cambridge: MIT press)
[9] Zhang C B, Chang C C and Jamshidi M 2021 Simultaneous pixel-level concrete defect detection and grouping using a fully convolutional model Structural Health Monitoring 1475921720985437
[10] Long J, Shelhamer E and Darrell T 2015 Fully convolutional networks for semantic segmentation IEEE Transactions on Pattern Analysis and Machine Intelligence 39 640–651
[11] Ronneberger O, Fischer P and Brox T 2015 U-Net: Convolutional Networks for Biomedical Image Segmentation (Berlin: Springer)
[12] Qi W W, Wei M F, Yang W T, Xu C and Ma C 2020 Automatic Mapping of Landslides by the ResU-Net Remote Sensing 12 2487
[13] He M K, Zhang X Y, Ren S Q and Sun J 2016 Deep Residual Learning for Image Recognition IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp 770–778
[14] Benz C, Debus P, Ha H K and Rodehorst V 2019 Crack segmentation on UAS-based imagery using transfer learning 2019 International Conference on Image and Vision Computing New Zealand (IVCNZ) pp 1–6
[15] Ren Y P, Huang J S, Hong Z Y, Lu W, Yin J, Zou L J and Shen X H 2020 Image-based
concrete crack detection in tunnels using deep fully convolutional networks. *Construction and Building Materials* **234** 117367