



Contents lists available at ScienceDirect

Journal of Rock Mechanics and Geotechnical Engineering

journal homepage: www.jrmge.cn

Full Length Article

A hybrid attention deep learning network for refined segmentation of cracks from shield tunnel lining images

Shuai Zhao^a, Guokai Zhang^b, Dongming Zhang^c, Daoyuan Tan^{a,*}, Hongwei Huang^c^a Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, 999077, Hong Kong, China^b School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, 200093, China^c Key Laboratory of Geotechnical and Underground Engineering of Ministry of Education, Tongji University, Shanghai, 200092, China

ARTICLE INFO

Article history:

Received 14 October 2022

Received in revised form

7 January 2023

Accepted 16 February 2023

Available online 14 April 2023

Keywords:

Crack segmentation

Crack disjoint problem

U-net

Channel attention

Position attention

ABSTRACT

This research developed a hybrid position-channel network (named PCNet) through incorporating newly designed channel and position attention modules into U-Net to alleviate the crack discontinuity problem in channel and spatial dimensions. In PCNet, the U-Net is used as a baseline to extract informative spatial and channel-wise features from shield tunnel lining crack images. A channel and a position attention module are designed and embedded after each convolution layer of U-Net to model the feature interdependencies in channel and spatial dimensions. These attention modules can make the U-Net adaptively integrate local crack features with their global dependencies. Experiments were conducted utilizing the dataset based on the images from Shanghai metro shield tunnels. The results validate the effectiveness of the designed channel and position attention modules, since they can individually increase balanced accuracy (BA) by 11.25% and 12.95%, intersection over union (IoU) by 10.79% and 11.83%, and F₁ score by 9.96% and 10.63%, respectively. In comparison with the state-of-the-art models (i.e. LinkNet, PSPNet, U-Net, PANet, and Mask R-CNN) on the testing dataset, the proposed PCNet outperforms others with an improvement of BA, IoU, and F₁ score owing to the implementation of the channel and position attention modules. These evaluation metrics indicate that the proposed PCNet presents refined crack segmentation with improved performance and is a practicable approach to segment shield tunnel lining cracks in field practice.

© 2023 Institute of Rock and Soil Mechanics, Chinese Academy of Sciences. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Cracks usually occur on the concrete linings of shield tunnels during in-service or construction stages owing to structure degradation or environmental changes. The continued propagation of cracks will accelerate the structure deterioration, bringing about failure or unreliability to the structural systems (Abdel-Qader et al., 2003; Wang et al., 2022). Cracks are one of the earliest indicators in estimating structure degradation. Therefore, accurate and timely detection of cracks will have high practical application values in preventing greater damage or potential accidents to infrastructures, such as shield tunnels.

Powered by the computer vision techniques, inspection method of shield tunnel lining defects has changed from human-based method to computer vision-based approach. The computer vision-based method can produce numerous tunnel lining images, and image processing techniques are used to identify defects, such as concrete spalling, leakage area, and cracks in the obtained images (Dawood et al., 2018). Currently, identification of defects in images has changed from conventional image processing techniques to deep learning techniques because of the poor performance of conventional image processing methods for processing images that have complicated backgrounds (Zhao et al., 2021; Huang et al., 2022). The emergence of deep learning techniques has greatly promoted the revolution of geotechnical engineering (Zhang et al., 2021a, 2021b; Chen et al., 2022, 2023). Zhang et al. (2021c) and Phoon and Zhang (2022) have conducted detailed review studies to identify achievements and to share prospects for deep learning techniques in addressing geotechnical problems.

One main advantage is that deep learning techniques can automatically learn the features needed for classification and

* Corresponding author.

E-mail address: dytan@polyu.edu.hk (D. Tan).

Peer review under responsibility of Institute of Rock and Soil Mechanics, Chinese Academy of Sciences.

prediction in large datasets, which make deep learning methods have been used in some fields for rock slopes' key-block classification (Zhu et al., 2022), lithology identification (Xu et al., 2022), rock mass quality classification (Sheng et al., 2022), working face ground identification (Liu et al., 2021), braced excavation-induced wall deflection prediction (Wu et al., 2022), and leakage detection (Xue et al., 2020, 2022). In addition, deep learning approaches have proven to be robust for crack detection (Cha et al., 2017), and numerous researches have shown the advantages of U-Net (Ji et al., 2018), fully convolutional networks (FCN) (Huang et al., 2018; Yang et al., 2018; Dung and Anh, 2019; Zhai et al., 2022), convolutional neural networks (CNNs) (Alipour and Harris, 2020; Hsieh and Tsai, 2020; Hsieh et al., 2020; Lee et al., 2020; Pan et al., 2020; Adam and Sathesh, 2021), SegNet (Song et al., 2019), CrackNet (Zhang et al., 2017), DeepCrack (Zou et al., 2018; Liu et al., 2019), CrackU-net (Huyan et al., 2020), YOLOv4 (Yu et al., 2021), Faster R-CNN (Deng et al., 2020), CAL-ISM (Zheng et al., 2022), Mask R-CNN (Kim and Cho, 2019; Kalfarisi et al., 2020), and compressive sensing with generative adversarial network (CS-GAN) (Huang et al., 2021) in dealing with crack prediction tasks.

Among these aforementioned studies, it is pretty remarkable that the CS-GAN proposed by Huang et al. (2021) and the synthetic data augmentation method proposed by Zhai et al. (2022) are effective approaches to improve crack segmentation performance from the perspective of optimizing image datasets; and, most notably, U-Net and FCN are widely applied to modern semantic segmentation models. For example, Yang et al. (2018) used FCN for segmentation of cracks, and skeletonizing operations were then performed on the cracks segmented by FCN to compute the width and length of cracks. Their proposed approach achieved a low measurement error for crack width calculation. However, the error rate for crack length calculation was slightly high because of the 'disjoint problem' in cracks segmented by FCN. The 'disjoint problem' is that a continuous tunnel lining crack (Fig. 1a) is misidentified as several disconnected crack parts during the segmentation process (Fig. 1b), and it significantly influences the quantification precision when computing the whole length of a crack. Similar 'disjoint problem' also exists in the authors' previous work (Huang et al., 2022), where a mask-region-based convolutional neural network (Mask R-CNN) combining a morphological closing operation was proposed to segment shield tunnel lining cracks. To address this 'disjoint problems', the authors subsequently tried to modify a PANet combining a morphological closing

operation to segment cracks (Zhao et al., 2021). The authors' proposed approach further mitigated the 'disjoint problem'. Nevertheless, the crack segmentation performance improved slightly from the Mask R-CNN to the modified PANet. Actually, the crack width is on the millimeter level, and therefore even a very small improvement of crack segmentation performance is of great significance to accurately calculate the length and width of cracks. Thus, it is essential to design a refined approach to further mitigate the 'disjoint problem' in order to precisely compute the whole length of a crack.

The 'disjoint problem' of segmented cracks (Fig. 1b) is attributed to the multiple convolution and pooling operation during the feature extraction process, in which the features of a crack covering only a few pixels are frequently lost (Chaurasia and Culurciello, 2017; Zhao et al., 2021). The loss of crack features results in that some crack parts not being identified. Therefore, it is necessary to select or develop a suitable baseline (or backbone architecture) model to finely extract crack features so as to prevent feature loss. U-Net, a symmetrical encoder-decoder structure, is frequently used as baseline and has demonstrated great performance for the semantic segmentation of crack images owing to its skip connections (Jenkins et al., 2018; Ji et al., 2018; Zou et al., 2018; Huyan et al., 2020; Hsieh and Tsai, 2020). These skip connections have also acted a vital role in the success of PANet (Liu et al., 2018) and Mask R-CNN (He et al., 2018). Moreover, most of the state-of-the-art models for crack image segmentation are variants of U-Net (e.g. CrackU-net (Huyan et al., 2020) and DeepCrack (Zou et al., 2018)). U-Net-, FCN-, Mask R-CNN-, and PANet-based approaches usually predict the category for each pixel of crack images; however, the relationship between pixels is ignored, making a segmented crack not continuous or obviously making the size and shape of a segmented crack quite different from that of ground truth. Therefore, the spatial correlation between pixels has an effect on the 'disjoint problem'. To address the spatial inconsistency between pixels, the attention mechanism has been incorporated into neural networks to solve segmentation tasks (Hu et al., 2018; Fu et al., 2019; Pan et al., 2020). Motivated by the successful application of U-Net and attention mechanism, this study attempts to develop a hybrid network through incorporating designed attention layers into U-Net to mitigate the 'disjoint problem'.

This study developed a hybrid network termed the position-channel network (PCNet) to reduce the 'disjoint problem' of segmented cracks in channel and spatial dimensions. For this purpose, a position attention (PosAtt) module and a channel attention (ChaAtt) module are designed and embedded after each convolution layer of U-Net to form a hybrid network. These two attention modules can make the U-Net adaptively integrate local crack features with their global dependencies. As a consequence, feature interdependencies in both the channel and spatial dimensions can be modeled and used to strengthen the quality of obtained features, contributing to a more precise crack segmentation process to alleviate the 'disjoint problem'.

2. Methodology

2.1. The proposed approach

The U-Net composes of an encoder sub-network and a decoder sub-network, and has a great many of feature channels in the decoder sub-network, which enables the network to propagate context information to higher layers (Ronneberger et al., 2015). Moreover, the semantic feature maps from decoder sub-network are merged with the low-level feature maps from encoder sub-network through skip connections, and thus the U-Net can yield more precise segmentations of target objects. U-Net has proven to

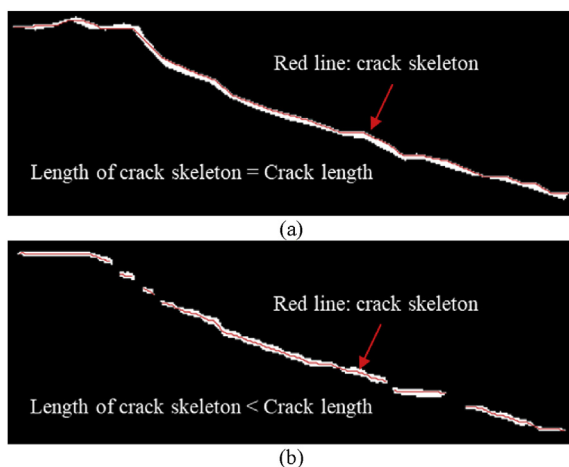


Fig. 1. Identified cracks. The length of crack skeleton is frequently computed as crack length. (a) An identified continuous crack, and (b) A continuous crack segmented as disconnected parts.

be effective for crack image segmentation (Zou et al., 2018; Huyan et al., 2020). Therefore, this study uses the U-Net as baseline/backbone to extract features from crack images.

Despite the advantage of the U-Net for crack image segmentation, the U-Net ignores the relationships between pixels, and thus the spatial consistency in the final segmentation map of U-Net cannot be guaranteed (Lin et al., 2016). As we know, CNNs extract informative features by fusing channel-wise and spatial information together within local receptive fields. Therefore, this study attempts to improve the representational power of the U-Net in both channel and spatial dimensions so as to strength the relationships between pixels and thus reduce the ‘disjoint problem’. To achieve this, a PosAtt module and a ChaAtt module are designed and incorporated into the U-Net to form a hybrid PCNet. The main structure of the designed PCNet is illustrated in Fig. 2.

The designed ChaAtt layers and PosAtt layers are placed after each convolution layer, as presented in Fig. 2. These two attention layers are placed in parallel to capture the interdependencies in channel and spatial dimensions separately, and then the features are fused, as displayed in Fig. 3. The ChaAtt layer and the PosAtt layer are complementary to each other, i.e. the ChaAtt layers guarantee the information is learned channel-wisely in a global view, while the PosAtt layers guarantee the pixel-level information is strengthened in a local view. Specifically, the features from a convolution layer are first fed into the PosAtt layer to capture the spatial dependencies between any two pixels through assigning different weights to pixels, which makes the pixels with similar or same properties build associations. As a result, the continuity of crack pixels is strengthened, and thus the ‘disjoint problem’ is mitigated at some extent. Meanwhile, the ChaAtt layers capture the channel relationship features through assigning different weights to different channels, which is equivalent to enlarging the receptive field through nonlinear interaction between channels. Therefore, the global context information can be leveraged to help prevent the crack parts from being misidentified as backgrounds, and as a result the ‘disjoint problem’ is mitigated at some extent. Finally, the strengthened features from the two designed attention layers are aggregated through element-wise summation to achieve better crack segmentation performance. More details about the structure of the ChaAtt layer and the PosAtt layer are introduced in Sections 2.2 and 2.3.

The PCNet consists of the encoder and decoder stages (Fig. 2). At the encoder stage, the network extracts hierarchy features of input crack images through convolution operations. At the decoder stage, deconvolution operations restore the high-level feature map produced at the encoder stage to the resolution of original crack images. The encoder and decoder stages are connected through four skip connections (Fig. 2), which ensures the finally restored feature map integrates features from different scales and more low-level features.

2.2. Channel attention module

During the convolution operations, each of convolution layers works with a local receptive field and consequently contextual information outside of this local receptive field cannot be utilized by each unit of output features (Hu et al., 2018). This problem becomes more serious at lower convolution layers whose receptive field sizes are small. Inspired by the successful use of ‘inception module’ (Szegedy et al., 2015) and ‘squeeze-and-excitation block’ (Hu et al., 2018), a new ChaAtt module is designed in this study to mitigate the above problem, as displayed in Fig. 4.

The intermediate feature is denoted as follows:

$$\mathbf{F} = [F_1, F_2, \dots, F_i, \dots, F_C] \quad (1)$$

where F_i is the feature map from i th channel of \mathbf{F} , $\mathbf{F} \in \mathbb{R}^{C \times W \times H}$ (C , H , and W are the total channel number of \mathbf{F} , the height, and the width, respectively) and $i \in \{1, 2, \dots, C\}$. For each F_i , a 5×5 , a 3×3 , and a 1×1 convolutional layer are applied to F_i to obtain multi-scale context features, and these features are concatenated (i.e. these multi-scale context features are combined through merging the channels). The combined features are subsequently put into a 3×3 convolutional layer to reduce channels (Fig. 4). Then, a global average pooling (GAP) is performed on the obtained features to produce channel-wise global averaged vector as $\mathbf{z} \in \mathbb{R}^C$. The z_i of the i th channel is calculated as

$$z_i = \frac{1}{HW} \sum_x \sum_y F_i(x, y) \quad (2)$$

Moreover, a sigmoid function (σ) is utilized to accomplish

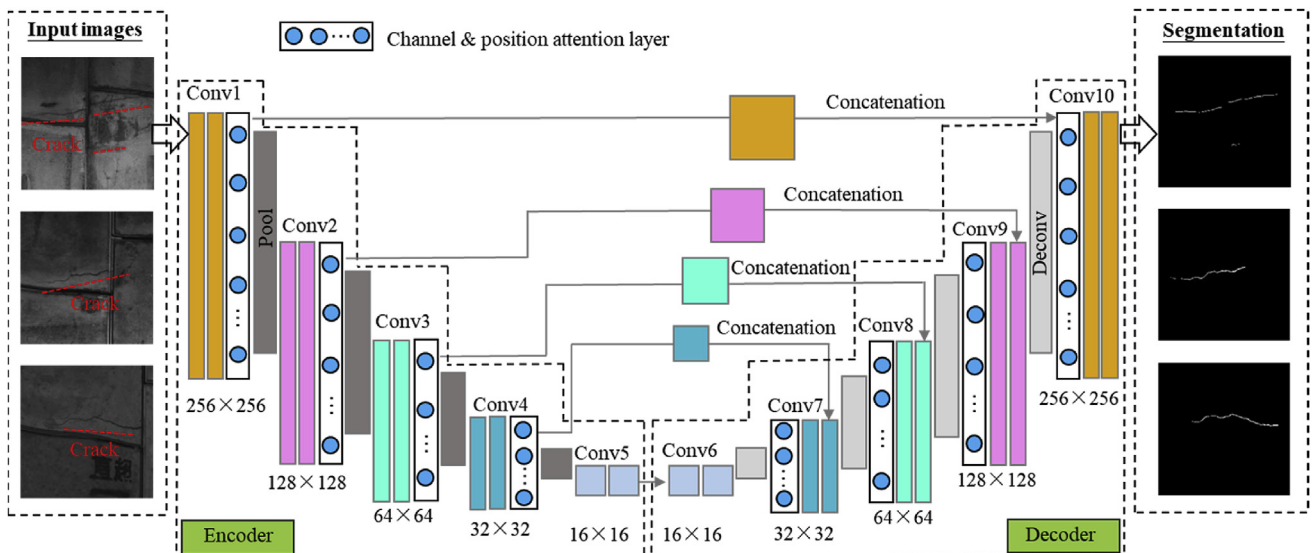


Fig. 2. Schematic diagram of the proposed approach.

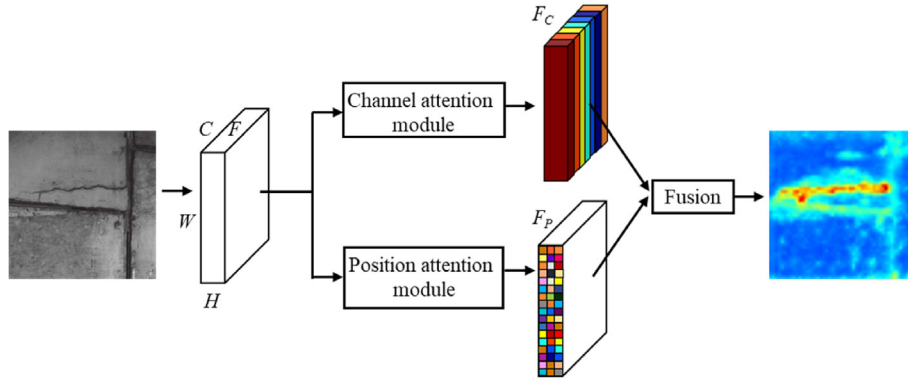


Fig. 3. The connection between ChaAtt and PosAtt layers.

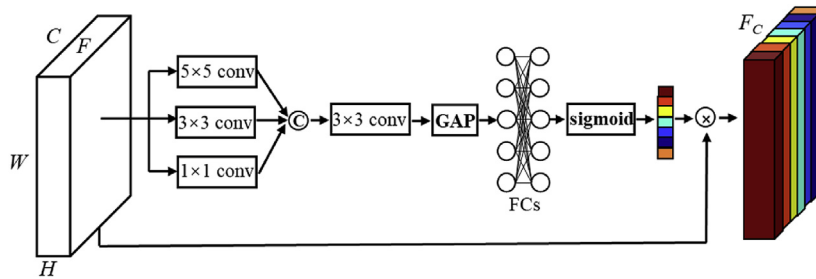


Fig. 4. Structure of ChaAtt module.

information aggregation in order to fully capture channel-wise dependencies as follows:

$$\mathbf{z}' = \sigma(W_2 \delta(W_1 \mathbf{z})) \quad (3)$$

where δ represents the ReLU function, which is applied to guaranteeing that multiple channels are able to be strengthened when building the nonlinear relationships among different channels through sigmoid activation. $W_1 \in \mathbb{R}^{F \times C}$ and $W_2 \in \mathbb{R}^{C \times F}$ are the weights of the fully connected layers (FCs), respectively (Fig. 4). Here, two FCs are used to improve the nonlinearity, however, which also increases computational cost because of the increased parameters (Hu et al., 2018). Therefore, a reduction ratio, r , is used to reduce the dimension so as to decrease the number of parameters and computational cost. Herein, the r is determined and set to 4, since the value of 4 can achieve a good trade-off between complexity and accuracy after conducting experiments with different values (i.e. 4, 8, 12, and 16).

Through ReLU function and sigmoid function, the importance of each feature channel can be determined and represented through \mathbf{z}' . The ChaAtt features, \mathbf{F}_C , will be gained by multiplying \mathbf{z}' with \mathbf{F} as

$$\mathbf{F}_C = \mathbf{F} \otimes \mathbf{z}' = [F_1 z'_1, \dots, F_i z'_i, \dots, F_C z'_C] \quad (4)$$

where \otimes is element-wise multiplication.

The ChaAtt layers make the convolution layers learn more global information through suppressing less useful features and highlighting informative ones. Thus, more effective global features of an input crack image are gained, and consequently the segmentation performance of the network is supposed to be improved.

2.3. Position attention module

Considering that local information acts an important part in obtaining a precise segmentation result for image segmentation task, a PosAtt module is designed to parallel with the ChaAtt module to extract the pixel-level and subtle feature information (Fig. 3). The PosAtt module helps the up-sampled feature map contain more subtle features. The structure of the proposed PosAtt module is illustrated in Fig. 5.

In this study, the input feature is denoted as follows:

$$\mathbf{F} = [F_{1,1}, F_{1,2}, \dots, F_{i,j}, \dots, F_{H,W}] \quad (5)$$

where $F_{i,j} \in \mathbb{R}^{C \times 1 \times 1}$ represents the feature at the position (i,j) on \mathbf{F} .

The PosAtt module focuses on the selection of feature position, and therefore the features that are relevant to the segmentation performance are strengthened spatially. For this purpose, a 5×5 , a 3×3 , and a 1×1 convolutional layer are applied to \mathbf{F} to obtain multi-scale context features, and then these features are added. Next, a 1×1 convolution is applied to fusing features from all the channels so as to obtain the fused feature $\mathbf{F}_f \in \mathbb{R}^{H \times W}$. A sigmoid function (σ) is then implemented to the fused features \mathbf{F}_f to generate the weight matrix (\mathbf{W}_p for all pixel positions). The \mathbf{W}_p is calculated as follows:

$$\mathbf{W}_p = \frac{e^{\mathbf{F}_f}}{1 + e^{\mathbf{F}_f}} \quad (6)$$

Finally, an element-wise multiplication is performed between \mathbf{W}_p and \mathbf{F} . Moreover, a residual connection is used to obtain the final PosAtt features \mathbf{F}_p by

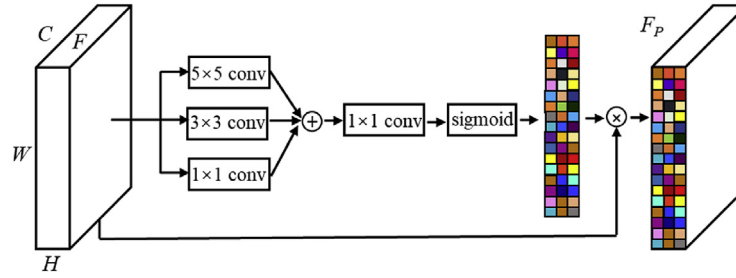


Fig. 5. Structure of PosAtt module.

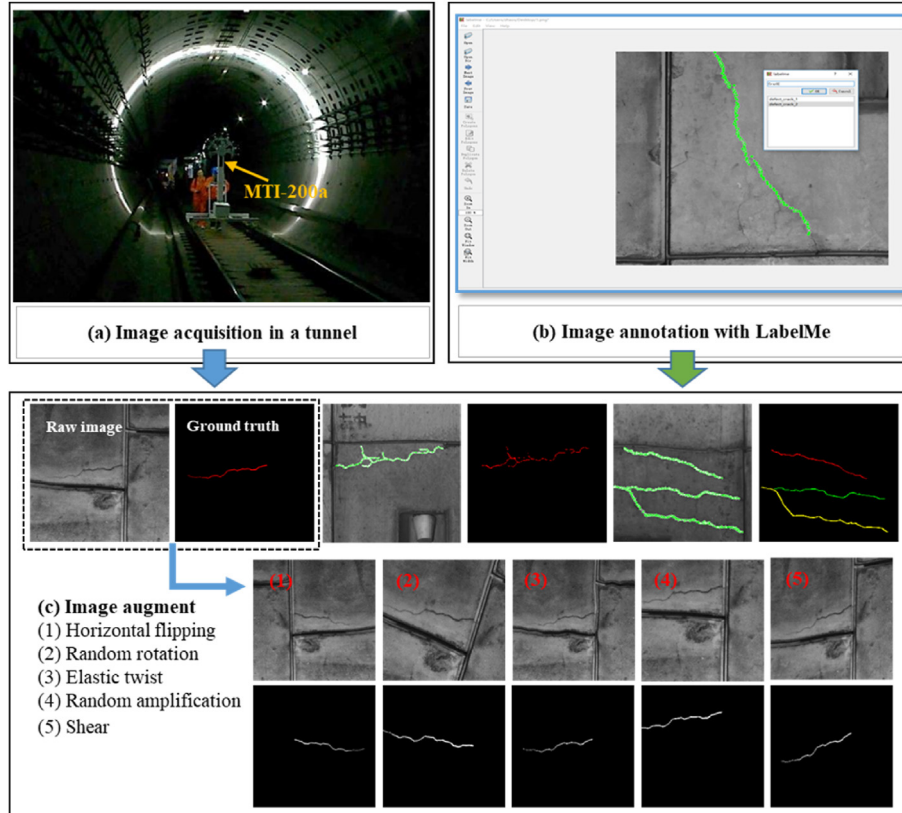


Fig. 6. Schematic diagram of data acquisition and processing.

$$\mathbf{F}_p = \mathbf{F} \otimes \mathbf{W}_p = [F_{1,1}w_p(1,1), F_{1,2}w_p(1,2), \dots, F_{H,W}w_p(H,W)] \quad (7)$$

where $w_p(H, W)$ is the weight at the position (H, W) of feature \mathbf{F} .

The designed PosAtt module considers the relationships between pixels, which strengthens the spatial consistency of pixels. Therefore, the designed PosAtt module improves the continuity of segmentation results in term of pixels at some extent.

2.4. Loss function used for the proposed model

In this study, the pixels of an image are classified into two categories: crack pixels and background pixels. A binary cross-entropy loss function is much suitable for binary segmentation because of its high efficiency in updating weights. The weights can be updated faster when the gradient calculated by a binary cross-entropy loss function is higher. Thus, a binary cross-entropy loss function is

selected as the loss function (denoted by L) of the proposed PCNet, and it is defined as follows:

$$L = -\frac{1}{N} \sum_{i=1}^N [p_i^* \log_{10} p_i^u + (1 - p_i^*) \log_{10} (1 - p_i^u)] \quad (8)$$

where $p_i^* = 1$ or 0 indicates the pixel i classified as class u or background, p_i^u represents the predicted probability of the pixel i being the class u , and N is the number of different classes.

3. Experimental results using the proposed model

3.1. Dataset of crack images

The crack dataset previously established by the authors (Zhao et al., 2021) was applied for the training, validation, and testing of the proposed PCNet. The crack dataset was established through three steps: image acquisition, image cropping, and image

annotation (Zhao et al., 2021; Huang et al., 2022). As shown in Fig. 6a, a moving tunnel inspection machine (MTI-200a) was used to collect images at the 5.5-m diameter shield tunnels in Shanghai, China, and six line-scan cameras on MTI-200a machine achieved a pixel precision of 0.3 mm/pixel for the obtained images through the field calibration test (Huang et al., 2017). The acquired images were then extracted from the machine and cropped to the size from 800×800 pixels to 3000×3000 pixels, and the image size was selected according to the size of cracks in order to make each image contained at least a whole crack. After image cropping, the cracks in the obtained images were labeled along their boundary through the LabelMe tool (Fig. 6b). Cracks were labeled using per-instance label to aid in accurate crack localization. When image label is finished, a JavaScript Object Notation (json) file corresponding to images was generated. As shown in Fig. 7, the json file contains coordinates of the crack edge points, the length and width of the minimum enclosing rectangle of each crack, the area of a crack, the width and height of the images, and the serial number for the images and categories. Then the ground truth images were obtained using the Python language to process the json file.

The detailed establishment steps for the dataset can refer to Zhao et al. (2021) and Huang et al. (2022). The crack dataset consists of 572 training images, 143 validation images, and 79 testing images, all of which covered the aforementioned image size (i.e. ranging from 800×800 pixels to 3000×3000 pixels). When using the training images to train the PCNet, the PCNet automatically performed the horizontal flip, random rotation, elastic twist, random amplification, and shear operations to each training image for data augmentation (Fig. 6c). These augmentation techniques

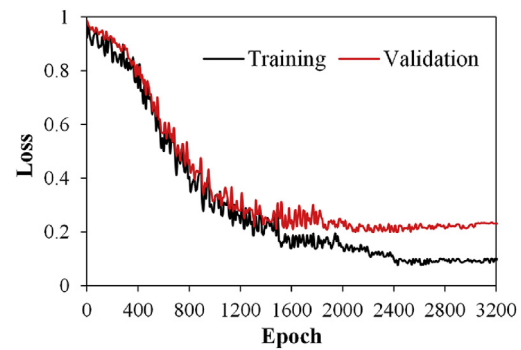


Fig. 8. Loss variation during the process of training and validation.

were only used for the training dataset to improve the segmentation performance of the PCNet during the training process.

3.2. Training settings and results

The proposed model was executed on a desktop configured with one NVIDIA GTX 1080 GPU. The calculation software environment was configured with TensorFlow framework. Through trial and error, the initial learning rate was set to 0.001, and it decreased by 0.0005 after 400 epochs. The Adam optimizer (Kingma and Ba, 2014) was used to update the weights in the proposed PCNet, and the parameters of the model were initialized by Gaussian distribution. All input images were re-sized (i.e. shrunk) to 256×256 pixels using area-based interpolation (Joshi, 2015) after being put into the PCNet. During the training process, the horizontal flip, random rotation, elastic twist, random amplification, and shear were used for data augmentation.

The training and validation processes continued for a maximum of 3200 epochs. As presented in Fig. 8, the loss values of the proposed model started to converge after 2800 epochs, and the training stopped with best validation performance registered as 0.231091 at Epoch 3200. Thus, the established models trained for 3200 epochs were applied to conducting experiments on the established testing dataset.

3.3. Performance and results on the testing dataset

3.3.1. Segmentation results of the proposed approach

To demonstrate the performance of the proposed PCNet for real-life crack images, two true positive sample images that were output by the proposed PCNet are presented in Fig. 9. The original images and the corresponding ground truths are also presented in Fig. 9. The crack in each image is correctly segmented as a continuous crack, and the boundary and shape of the crack region are relative accurate. Moreover, the crack masks predicted by the PCNet are much closer to the cracks in ground truth images. It should be noted that the pixel precision of the images in the established dataset is 0.3 mm/pixel, as introduced in Section 3.1. Therefore, the width of the cracks that are segmented by the PCNet is above 0.3 mm, which means that the proposed PCNet can identify the cracks with width above 0.3 mm.

The proposed PCNet are fed with images containing multiple cracks to evaluate the robustness of the model in identifying multiple cracks. If multiple cracks are separate in an image, it takes little effort for the proposed model to correctly segment all the cracks, as presented in Fig. 10a–c. All the cracks are correctly segmented by the proposed model, and the shape of segmented cracks is close to the actual shape of the corresponding ground truth. However, the proposed model is not effective for the map

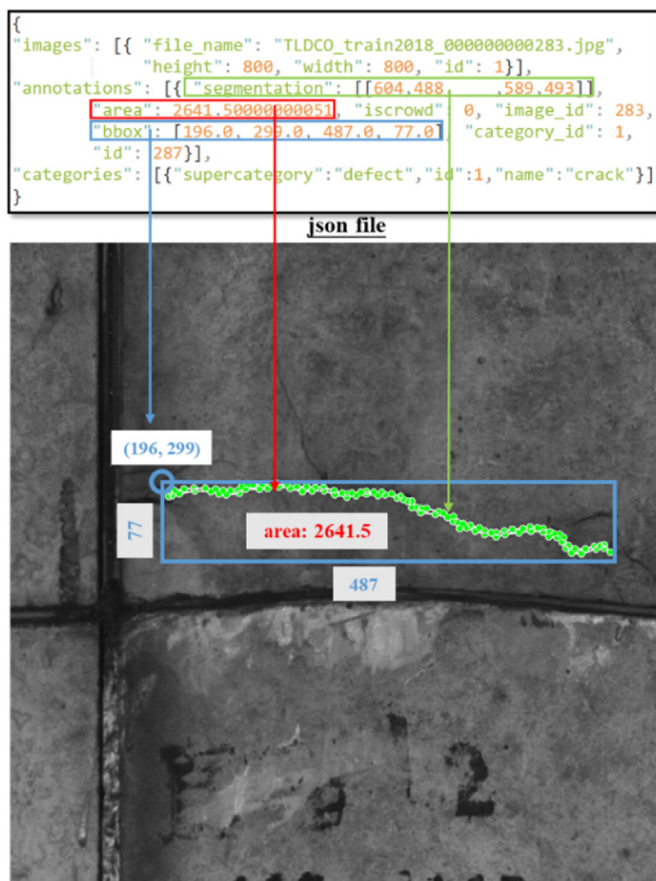


Fig. 7. The information of a crack image in the json file.

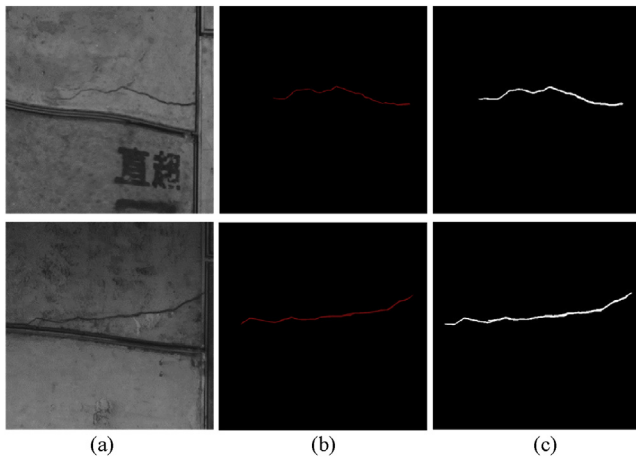


Fig. 9. Two true positive results of PCNet: (a) Original image, (b) Ground truth, and (c) Image output by PCNet.

cracks (Fig. 10d). The basic shape of the map cracks can be roughly segmented, but the crack shape at the fork points is difficult to be finely captured (Fig. 10d). That is because the number of map crack samples in the training dataset is very small, since only a very small number of map crack samples are collected during shield tunnel inspection. The less trained samples of map cracks cause that the proposed model cannot fully learn the features of map cracks during training process, which, in turn, leads to the poor

segmentation effect of the proposed model on map cracks among testing dataset. The segmentation results should be better if more map crack samples are collected and used to train the proposed model.

In summary, the proposed model is robust in identifying multiple separate cracks in an image. Nevertheless, cracks segmented by the proposed approach are still facing the ‘disjoint problem’, i.e. a continuous crack is segmented into multiple disconnected parts (Fig. 10a and b). Among the testing dataset, the proposed PCNet has an average of 0.4 discontinuities for each segmented crack compared to 1.8 discontinuities for the PANet and 2.4 discontinuities for the Mask R-CNN (Zhao et al., 2021). Thus, the proposed PCNet alleviates the ‘disjoint problem’ to some extent.

3.3.2. Effectiveness of the designed attention modules

To validate the effectiveness of ChaAtt and PosAtt modules of the proposed PCNet, the performance of the proposed PCNet is compared with that of the model without the PosAtt and ChaAtt layer (i.e. U-Net), the model only with the PosAtt layer (i.e. PNet), and the model only with the ChaAtt layer (i.e. CNet). The balanced accuracy (BA), IoU, and F₁ score are used as indicators (Huang et al., 2022) to evaluate the crack segmentation performance of these models among the testing dataset, and the results are summarized in Table 1. It can be known the BA, IoU, and F₁ score of the model without ChaAtt and PosAtt (70.31%, 40.21%, and 56.37%) are poorer than that of PNet (83.26%, 52.04%, and 67%) and CNet (81.56%, 51%, and 66.33%). Among these models, the proposed PCNet using the two attention modules achieves the best performance, which is because the PosAtt module and the ChaAtt module allow the

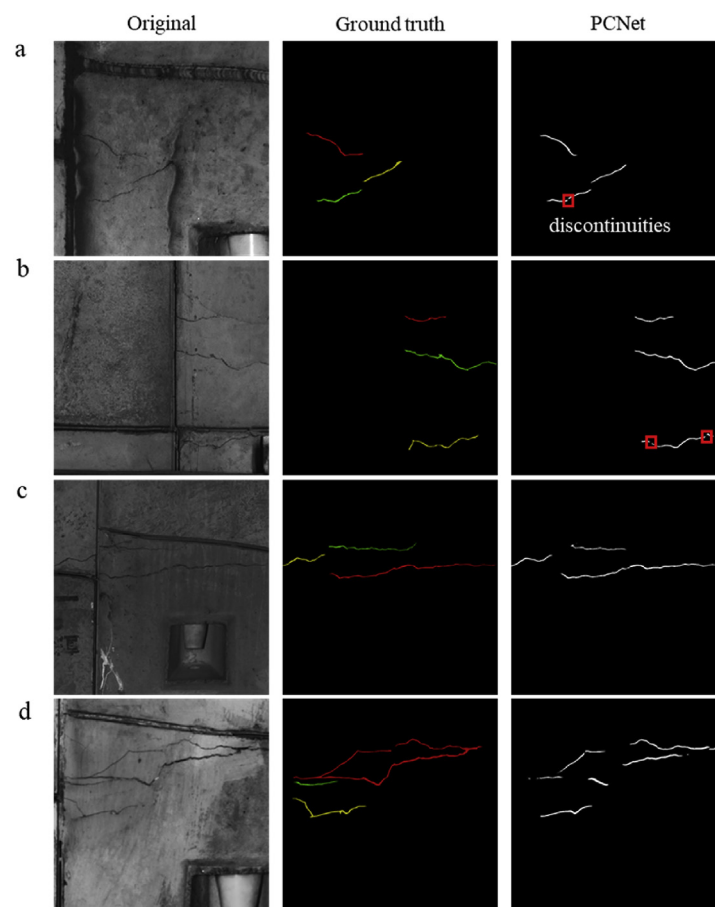


Fig. 10. The crack samples segmented by PCNet with ‘disjoint problem’: (a) Example a, (b) Example b, (c) Example c, and (d) Example d.

Table 1
Performance of the model with and without attention modules.

| Model | Backbone | PosAtt | ChaAtt | BA (%) | IoU (%) | F ₁ (%) |
|-------|----------|--------|--------|--------|---------|--------------------|
| U-Net | — | | | 70.31 | 40.21 | 56.37 |
| PNet | U-Net | ✓ | | 83.26 | 52.04 | 67 |
| CNet | U-Net | | ✓ | 81.56 | 51 | 66.33 |
| PCNet | U-Net | ✓ | ✓ | 84.59 | 53.14 | 68.07 |

convolutional layers to learn more local and global representations that can further improve the crack segmentation performance of the proposed PCNet.

Some examples are visualized using models with different attention modules and are presented in Fig. 11. It can be known from Fig. 11 that images output by the PCNet are closer to the ground truth compared to U-Net model. Comparing images output by CNet model and PNet model with images output by the proposed PCNet, it can be seen that ChaAtt layers make the convolutional layers focus on the global structure of crack region, while the PosAtt layers tend to detect subtle crack boundaries.

3.3.3. Comparative experiments

To further show the superiority of the proposed PCNet for crack segmentation task, the results of implementing LinkNet (Chaurasia and Culurciello, 2017), PSPNet (Zhao et al., 2017), and U-Net (Ronneberger et al., 2015) on the testing dataset are displayed in Fig. 12 in comparison with the results of implementing the proposed PCNet.

Compared with the other three methods, the proposed PCNet shows better performances for crack segmentation. Such as in the third row of Fig. 12, distractors (or noise) are detected as cracks by PSPNet and U-Net, while the proposed PCNet successfully filters out these distractors. Cracks segmented by the PSPNet, LinkNet, and U-Net are discontinuous (Fig. 12), while the proposed PCNet can obtain a more precise shape and boundary of the crack region. Moreover, the crack region segmented by the PCNet is close to the ground truth. These findings indicate that the designed PosAtt and ChaAtt modules help the network obtain more local subtle and global features, which boosts the segmentation of cracks.

As presented in Table 2, the proposed PCNet demonstrates remarkable improvement in the BA, IoU and F₁ score in comparison with other models. The values of BA, IoU, and F₁ score of the proposed PCNet (84.59%, 53.14%, and 68.07%) are much higher than those for the LinkNet (74.64%, 43.24%, and 60.04%), PSPNet (72.59%, 41.56%, and 58.04%), and U-Net (70.31%, 40.21%, and 56.37%). Among all the models in Table 2, U-Net performs poorly when it is used for shield tunnel lining images whose backgrounds are complicated. Compared with the performance of U-Net, the performance of the proposed model demonstrates noticeable improvement, i.e. the values of BA, IoU, and F₁ score increase by 14.28%, 12.93%, and 11.7%, respectively. The difference between U-Net and the proposed model is that the ChaAtt layers and PosAtt layer are added to the network to improve the segmentation performance. The improvement of the segmentation performance

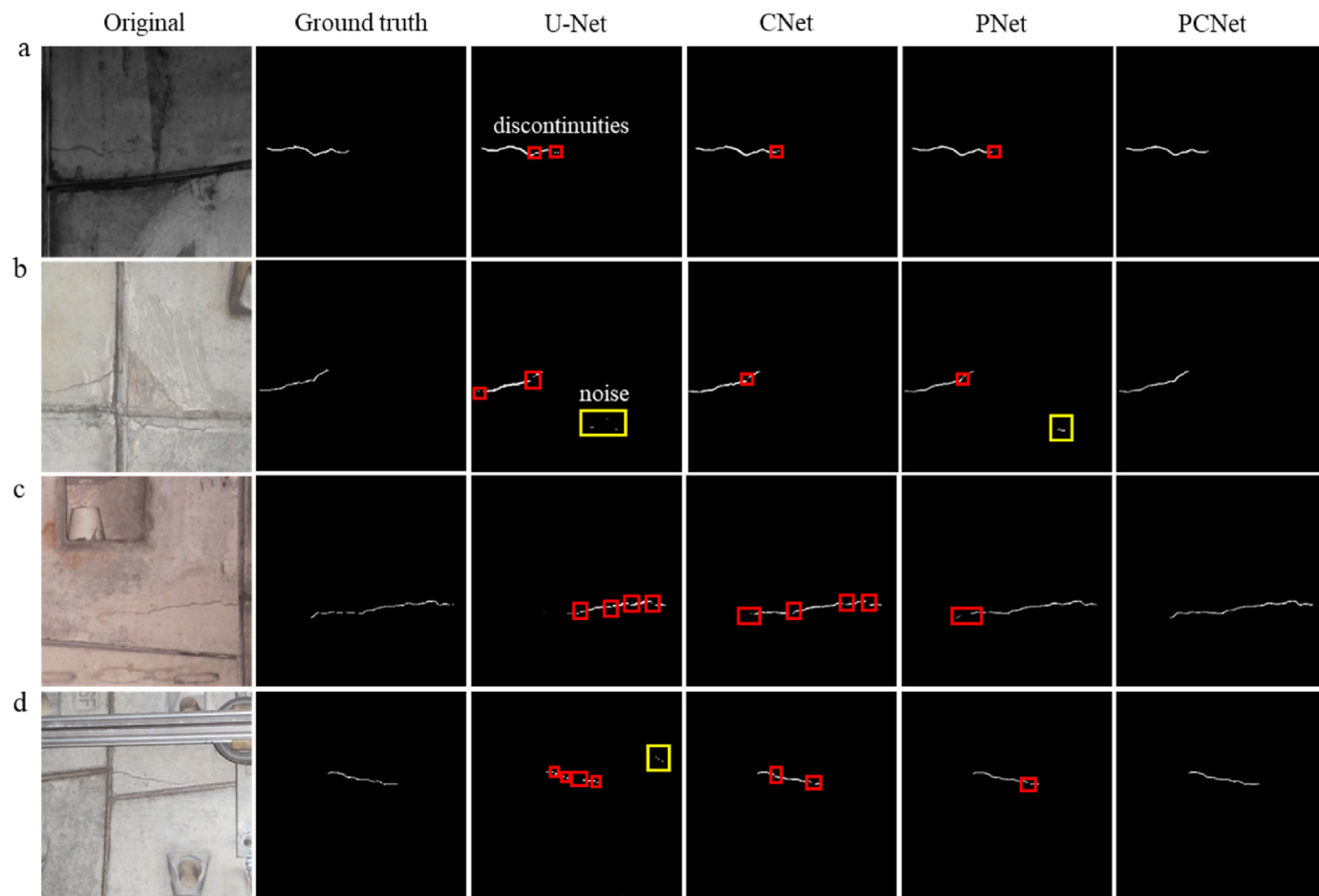


Fig. 11. Segmentation results with different attention modules: (a) Example a, (b) Example b, (c) Example c, and (d) Example d.

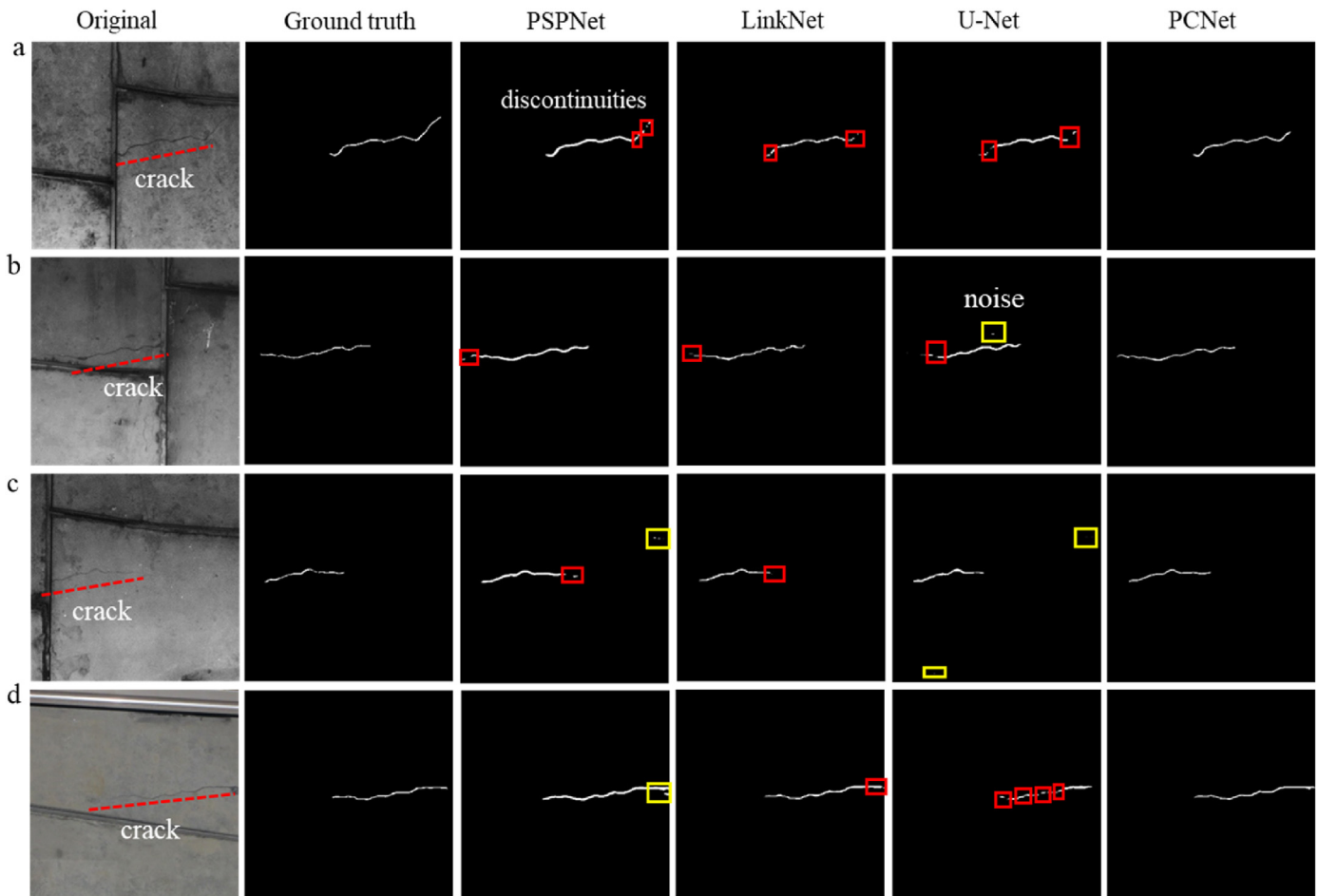


Fig. 12. Four crack samples segmented by different models: (a) Example a, (b) Example b, (c) Example c, and (d) Example d.

Table 2
Performance of the four models on the testing dataset.

| Model | BA (%) | IoU (%) | F ₁ (%) |
|---------|--------|---------|--------------------|
| LinkNet | 74.64 | 43.24 | 60.04 |
| PSPNet | 72.59 | 41.56 | 58.04 |
| U-Net | 70.31 | 40.21 | 56.37 |
| PCNet | 84.59 | 53.14 | 68.07 |

indicates that the above two attention modules contribute to the crack segmentation.

3.3.4. Comparison of the proposed PCNet with the author's previous model

The PANet model and Mask R-CNN model combining morphological closing operation have been utilized to mitigate the 'disjoint problem' by the authors (Zhao et al., 2021). Partial results of implementing the proposed PCNet, the PANet, and the Mask R-CNN on the same crack images are displayed in Fig. 13. In Fig. 13a–c and e, the cracks segmented by the PANet and the Mask R-CNN are discontinuous, while the cracks segmented by the proposed PCNet are without 'disjoint problem'. Among the testing dataset, the proposed PCNet has an average of 0.4 discontinuities for each segmented crack compared to 1.8 discontinuities for the PANet and 2.4 discontinuities for the Mask R-CNN. The three metrics for the proposed PCNet and for the PANet and the Mask R-CNN in Zhao et al. (2021) are summarized in Table 3. It can be seen that the

improvement of crack segmentation performance is relatively large from the PANet to the proposed PCNet, i.e. the BA, IoU, and F₁ score increase by 3.64%, 2.88%, and 1.95% (Table 3). Although the crack segmentation performance improves slightly, it is of great significance to accurately calculate the length and width of cracks since the crack width is on the millimeter level.

These findings indicate the crack segmentation performance for the simple U-Net combining ChaAtt and PosAtt module is superior to that for PANet and Mask R-CNN. The reasons are as follows. Some crack-similar distractors in shield tunnel lining images have semantic information similar to cracks, which results in complex semantic information for a crack image. Therefore, among the complex semantic information, accurately identifying the semantic information of cracks and strengthening the spatial correlation between pixels are essential to solve the 'disjoint problem' of identified cracks. The use of ChaAtt layers enables the proposed PCNet to adaptively recalibrate channel-wise feature responses, which allows the network itself to use global information to selectively suppress non-crack features and highlight crack features. Furthermore, the ChaAtt layers strengthen the interdependencies between the channels of its convolutional features. Whereas, the use of PosAtt layers allows the PCNet to learn the spatial interdependencies of features, which enables the PCNet to extract more pixel-level and subtle crack features. In addition, some crack boundaries identified by Mask R-CNN or PANet may be expanded or eroded during the process of morphological closing operation (Huang et al., 2022; Zhao et al., 2021), which affects the

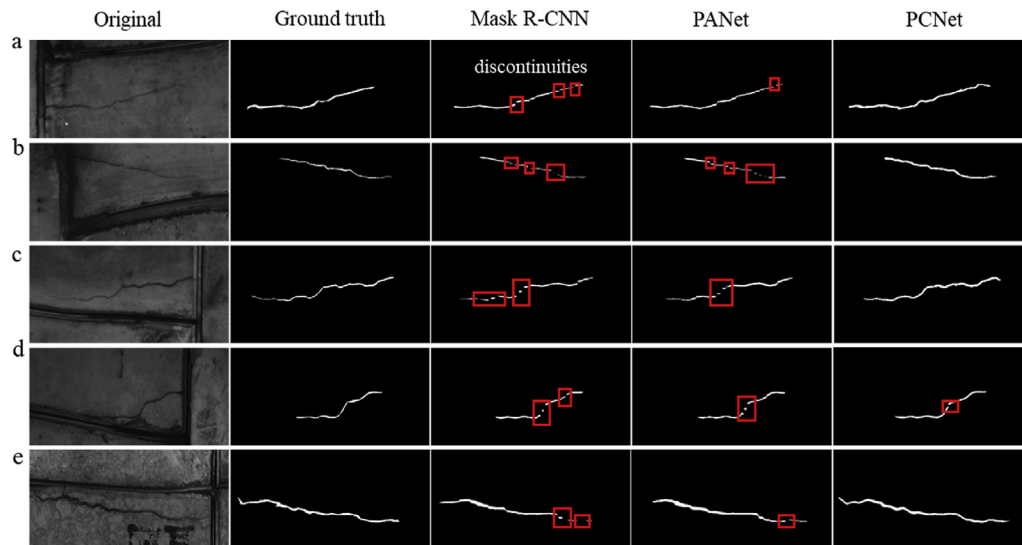


Fig. 13. Segmentation results using Mask R-CNN, PANet, and PCNet: (a) Example a, (b) Example b, (c) Example c, (d) Example d, and (e) Example e.

accuracy of identifying boundaries despite the morphological closing operation can connect the disjoint crack parts. Therefore, the proposed PCNet shows better performance compared to the Mask R-CNN and PANet.

However, for some cracks that are segmented without 'disjoint problem' by the proposed PCNet, some distractors near them are identified as cracks, as presented in Fig. 14. The reasons are that mask prediction competes between classes with respect to the proposed PCNet since pixel classification and mask prediction are coupled. Thus, crack-similar distractors can be identified as cracks by the proposed PCNet.

4. Discussion

4.1. Effect of 'disjoint problem' on quantification error of crack size

Overcoming the 'disjoint problem' for identified cracks is important for the accurate calculation of crack length and width in engineering applications. Similar to the authors' previous study (Zhao et al., 2021), A* algorithm (Lester, 2005; Zhao et al., 2021) is also applied to calculating the width and length of a crack in order to analyze the effect of the severity of 'disjoint problem' on the crack quantification error. The error rate (E_r) is used to evaluate the quantification error, and its calculation process is illustrated in Fig. 15. In the error rate calculation equation (Fig. 15), the size (i.e. length and width) of crack ground truths is used as the actual size, while the size of the crack segmented by the PCNet is used as the predicted size.

The mean error of crack length and width calculation for the proposed PCNet are 9.57% and 12.82%, as presented in Fig. 16. The mean error of crack size calculation for the Mask R-CNN and PANet in Zhao et al. (2021) is also summarized in Fig. 16. Compared with the PANet, the PCNet reduces the number of discontinuities of the segmented crack by three times, and the mean errors of crack length and width calculation respectively decrease by 3.77% and 2.16%, which indicates that the 'disjoint problem' largely influences the mean error of crack length calculation. The mean error of crack width calculation also has a large decrease from the PANet to the PCNet, and the reduced value (2.16%) is higher than that (0.68%) from the Mask R-CNN to the PANet. Although the reduced quantification error rate is relatively small, it is worth reducing the quantification error since the crack width is on the millimeter level.

It should be noted that the crack boundaries identified by the PCNet also affect the mean error of width calculation. Moreover, because the width calculation relies too much on the identified crack boundaries (a series of points A_i and B_i in Fig. 17) and the length calculation just relies two points of identified crack boundaries (points L and R in Fig. 17) (Zhao et al., 2021), the effect of identified crack boundaries on the mean width calculation error is much greater than that on the mean length calculation error. In the previous Mask R-CNN and PANet models, the morphological closing operation is applied to connecting the disconnected crack parts through expanding and eroding the identified boundaries, which affects the accuracy of the identified boundaries and consequently affects the mean error of width calculation. The morphological closing operation is not applied to the proposed PCNet; however, the PCNet also achieves a low number of discontinuities and higher values of BA, IoU and F_1 score (Table 3), which means the PCNet identifies the crack boundaries more accurately than the PANet and the Mask R-CNN. That is why the reduced value of the mean width calculation error from the PANet to the PCNet is much higher than that from the Mask R-CNN to the PANet. The above analysis further validates the contribution of the designed two attention modules in accurate crack boundary segmentation and consequently the low crack quantification error.

4.2. Explanation of the proposed PCNet through visualization

Although three metrics have been used to evaluate the effectiveness of the obtained models in Sections 3.3.2 and 3.3.3, the internal mechanism should be considered to better explain the results, particularly to check whether the proposed PCNet focuses on the target parts and whether the proposed attention modules contribute to the improved performance. As such, in recent years, more works have focused on understanding how deep learning works, such as the visualization of active areas of CNN using class activation maps and heat maps (Selvaraju et al., 2017; Gao and Mosalam, 2018), and visual explanations for deep CNN using Grad-CAM++ (Chattopadhyay et al., 2018). In this study, the heat maps are used to visualize the weighted combination of the resulting feature maps at different convolution stages of the PCNet and the pure U-Net. The heat maps can suggest which parts of an input image were inspected by the PCNet and the pure U-Net.

Table 3

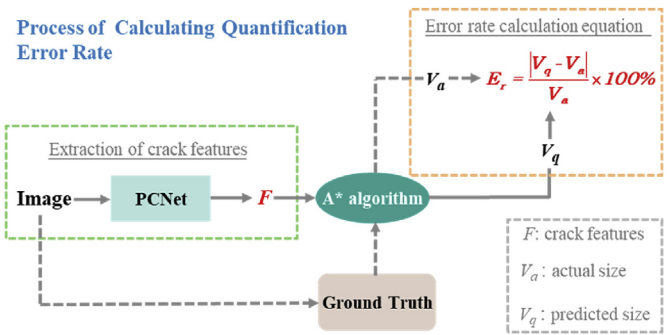
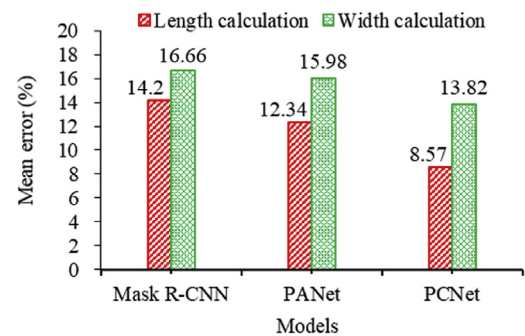
Crack segmentation performance for the three models.

| Model | BA (%) | IoU (%) | F ₁ (%) |
|--------------------------------|--------|---------|--------------------|
| Mask R-CNN (Zhao et al., 2021) | 78.91 | 48.82 | 64.88 |
| PANet (Zhao et al., 2021) | 80.95 | 50.26 | 66.12 |
| Proposed PCNet | 84.59 | 53.14 | 68.07 |

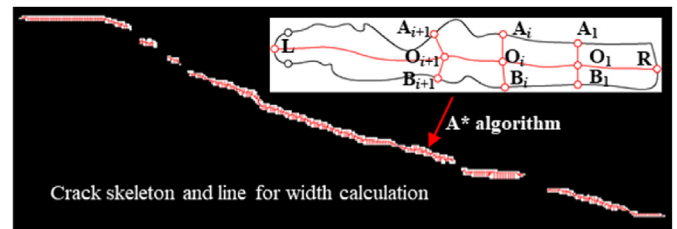
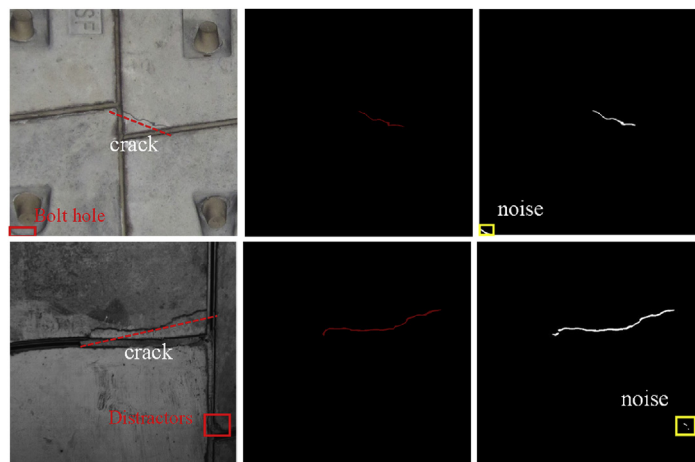
Fig. 18 presents a raw image and the corresponding heat maps from different convolutional layers of the PCNet and the pure U-Net. Fig. 18b and c demonstrates the active low-level features of the crack. The low-level features mainly refer to the crack edge and crack texture information, and both the PCNet and the pure U-Net can capture this information. In the deep layer, the proposed PCNet focuses on the entire target crack regions consistent with crack identified by human experts (Fig. 18d), while the model without PosAtt and ChaAtt module (i.e U-Net) only focuses on partial target crack regions and the longitudinal joint regions. The most active areas in Fig. 18d of the second row correspond to wrong places (i.e. longitudinal joint regions). These visualization results indicate that the two proposed attention modules help the PCNet extract more effective features and make the PCNet pay attention to the target crack regions. Thus, the proposed PCNet can be used as a baseline model, which can be improved in the future by increasing the number and type of data and utilizing more techniques that are frequently used in deep learning, such as atrous spatial pyramid pooling (Chen et al., 2018) and guided filtering (He et al., 2013).

5. Conclusions

Reducing the ‘disjoint problem’ of identified cracks is an important aspect to consider, since the ‘disjoint problem’ notably influences the accurate calculation of crack length and width. This study proposes a deep learning-based approach termed PCNet to mitigate the ‘disjoint problem’ of identified cracks. The PCNet couples the baseline U-Net with channel and PosAtt modules to form a hybrid network to reduce the ‘disjoint problem’ of segmented cracks in channel and spatial dimensions. A ChaAtt module and a position module are paralleled after each convolution layer of U-Net to model the feature interdependencies in channel and spatial dimensions. These attention modules enable the PCNet adaptively integrate local crack features with their global crack dependencies. Therefore, the PCNet can capture more subtle and

**Fig. 15.** Schematic diagram of the process for calculating crack quantification error.**Fig. 16.** Error results of different models.

pixel-level crack features, which contribute to a more precise crack segmentation process with less ‘disjoint problem’.

**Fig. 17.** Schematic diagram of the points used for length and width calculation.**Fig. 14.** False positive results of the PCNet.

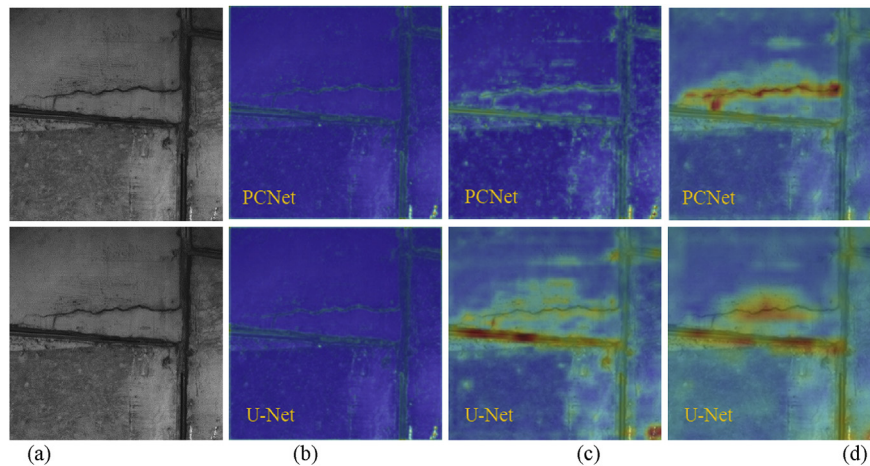


Fig. 18. Heat maps from different convolutional layers of two models: (a) Original images, (b) Heat maps of low layer, (c) Heat maps of middle layer, and (d) Heat maps of deep layer.

The performance for crack segmentation is evaluated using three metrics (i.e. BA, IoU, and F_1 score) in the established dataset containing 794 crack images. For each segmented crack, the proposed PCNet produces an average of 0.4 discontinuities, which is fewer than those produced for the PANet (1.8 discontinuities) and the Mask R-CNN (2.4 discontinuities). U-Net with individual ChaAtt module yields a BA of 81.56%, IoU of 51.00%, and F_1 score of 66.33%, each of which is 11.25%, 9.96%, and 10.79% higher than that of the baseline U-Net. Similarly, U-Net with individual PosAtt module exceeds the baseline by a BA of 12.95%, IoU of 11.83%, and F_1 score of 10.63%. It indicates both the ChaAtt module and PosAtt module can improve the crack segmentation performance. The proposed PCNet can obtain the most accurate results of crack segmentation under the highest BA of 84.59%, IoU 53.14%, and F_1 score of 68.07%, all of which are superior to the LinkNet, PSPNet, U-Net, Mask R-CNN, and PANet. The number of discontinuities of the segmented crack reduces by three times and the mean errors of crack length and width calculation respectively decrease by 3.77% and 2.16% from PANet to the proposed PCNet. In addition, the two designed attention modules can help the PCNet pay attention to the target crack regions.

This research highlights the mitigation of the ‘disjoint problem’ of identified cracks. Although the proposed PCNet improves crack segmentation performance with less ‘disjoint problem’, the PCNet misidentify some crack-similar distractors as crack because the mask prediction competes between classes. To address these limitations of the PCNet, in the future, generative adversarial network (GAN) (Goodfellow et al., 2020) can be incorporated into PCNet to reduce noises in segmented images with adversarial learning.

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.

Acknowledgments

The financial support from the Ministry of Science and Technology of the People's Republic of China (Grant No. 2021YFB2600804), the Open Research Project Programme of the State Key Laboratory of Internet of Things for Smart City (University of Macau) (Grant No. SKL-IoTSC(UM)-2021-2023/ORPF/A19/2022) and the General Research Fund (GRF) project (Grant No. 15214722)

from Research Grants Council (RGC) of Hong Kong Special Administrative Region Government of China are gratefully acknowledged.

References

- Abdel-Qader, I., Abudayyeh, O., Kelly, M.E., 2003. Analysis of edge-detection techniques for crack identification in bridges. *J. Comput. Civ. Eng.* 17 (4), 255–263.
- Adam, E.E.B., Sathesh, A., 2021. Construction of accurate crack identification on concrete structure using hybrid deep learning approach. *J. Innov. Image Process.* 3 (2), 85–99.
- Alipour, M., Harris, D.K., 2020. Increasing the robustness of material-specific deep learning models for crack detection across different materials. *Eng. Struct.* 206, 110157.
- Cha, Y.J., Choi, W., Buyukozturk, O., 2017. Deep learning-based crack damage detection using convolutional neural networks. *Comput. Aided Civ. Infrastruct. Eng.* 32 (5), 361–378.
- Chattopadhyay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N., 2018. Grad-cam++: generalized gradient-based visual explanations for deep convolutional networks. In: *Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, Lake Tahoe, NV, USA, pp. 839–847.
- Chaurasia, A., Culurciello, E., 2017. Linknet: exploiting encoder representations for efficient semantic segmentation. In: *Proceedings 2017 IEEE Visual Communications and Image Processing (VCIP)*. IEEE, St. Petersburg, FL, USA, pp. 1–4.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, Munich, Germany, pp. 801–818.
- Chen, B., Yang, Y., Yang, L., Li, W., 2022. Double-machine-learning-based data-driven stochastic flow stress model for aluminium alloys at elevated temperatures. *Mater. Today Commun.* 33, 104506.
- Chen, B., Shen, L., Zhang, H., 2023. A hybrid proper orthogonal decomposition-heteroscedastic sparse Gaussian process regression model for evaluating structural reliability with correlated stochastic material properties. *Struct. Saf.* 100, 102289.
- Dawood, T., Zhu, Z.H., Zayed, T., 2018. Computer vision-based model for moisture marks detection and recognition in subway networks. *J. Comput. Civ. Eng.* 32 (2), 04017079.
- Deng, J., Lu, Y., Lee, V.C.S., 2020. Concrete crack detection with handwriting script interferences using faster region-based convolutional neural network. *Comput. Aided Civ. Infrastruct. Eng.* 35 (4), 373–388.
- Dung, C.V., Anh, L.D., 2019. Autonomous concrete crack detection using deep fully convolutional neural network. *Autom. Construct.* 99, 52–58.
- 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*. IEEE, Long Beach, CA, USA, pp. 3146–3154.
- Gao, Y., Mosalam, K.M., 2018. Deep transfer learning for image-based structural damage recognition. *Comput. Aided Civ. Infrastruct. Eng.* 33 (9), 748–768.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2020. Generative adversarial networks. *Commun. ACM* 63 (11), 139–144.
- He, K., Sun, J., Tang, X., 2013. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (6), 1397–1409.
- He, K., Gkioxari, G., Dollar, P., Girshick, R., 2018. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (2), 386–397.
- Hsieh, Y.-A., Tsai, Y.J., 2020. Machine learning for crack detection: review and model performance comparison. *J. Comput. Civ. Eng.* 34 (5), 04020038.

- Hsieh, Y.A., Yang, Z., James Tsai, Y.C., 2020. Convolutional neural network for automated classification of jointed plain concrete pavement conditions. *Comput. Aided Civ. Infrastruct. Eng.* 36 (11), 1382–1397.
- Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Salt Lake City, UT, USA, pp. 7132–7141.
- Huang, H., Sun, Y., Xue, Y., Wang, F., 2017. Inspection equipment study for subway tunnel defects by grey-scale image processing. *Adv. Eng. Inf.* 32, 188–201.
- Huang, H.W., Li, Q.T., Zhang, D.M., 2018. Deep learning based image recognition for crack and leakage defects of metro shield tunnel. *Tunn. Undergr. Space Technol.* 77, 166–176.
- Huang, Y., Zhang, H., Li, H., Wu, S., 2021. Recovering compressed images for automatic crack segmentation using generative models. *Mech. Syst. Signal Process.* 146, 107061.
- Huang, H., Zhao, S., Zhang, D., Chen, J., 2022. Deep learning-based instance segmentation of cracks from shield tunnel lining images. *Struct. Infrastruct. Eng.* 18 (2), 183–196.
- Huyan, J., Li, W., Tighe, S., Xu, Z., Zhai, J., 2020. CrackU-net: a novel deep convolutional neural network for pixelwise pavement crack detection. *Struct. Control Health Monit.* 27 (8), e2551.
- Jenkins, M.D., Carr, T.A., Iglesias, M.I., Buggy, T., Morison, G., 2018. A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks. In: *Proceedings of the 26th European Signal Processing Conference (EUSIPCO)*. IEEE, Rome, Italy, pp. 2120–2124.
- Ji, J., Wu, L., Chen, Z., Yu, J., Lin, P., Cheng, S., 2018. Automated pixel-level surface crack detection using U-net. In: *Proceedings of the International Conference on Multi-Disciplinary Trends in Artificial Intelligence*. Springer, Hanoi, Vietnam, pp. 69–78.
- Joshi, P., 2015. *OpenCV with Python by Example*. Packt Publishing Ltd., Birmingham, UK.
- Kalfarisi, R., Wu, Z.Y., Soh, K., 2020. Crack detection and segmentation using deep learning with 3D reality mesh model for quantitative assessment and integrated visualization. *J. Comput. Civ. Eng.* 34 (3), 04020010.
- Kim, B., Cho, S., 2019. Image-based concrete crack assessment using mask and region-based convolutional neural network. *Struct. Control Health Monit.* 26 (8), e2381.
- Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization arXiv preprint arXiv:1412.6980.
- Lee, J.S., Hwang, S.H., Choi, I.Y., Choi, Y., 2020. Estimation of crack width based on shape-sensitive kernels and semantic segmentation. *Struct. Control Health Monit.* 27 (4), e2504.
- Lester, P., 2005. A* Pathfinding for Beginners. <https://www.gamedev.net/tutorials/programming/artificial-intelligence/a-pathfinding-for-beginners-r2003/>. (Accessed 9 October 2003).
- Lin, G., Shen, C., Van Den Hengel, A., Reid, I., 2016. Efficient piecewise training of deep structured models for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Las Vegas, NV, USA, pp. 3194–3203.
- Liu, S., Qi, L., Qin, H., Shi, J., Jia, J., 2018. Path aggregation network for instance segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Salt Lake City, UT, USA, pp. 8759–8768.
- Liu, Y., Yao, J., Lu, X., Xie, R., Li, L., 2019. DeepCrack: a deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing* 338, 139–153.
- Liu, M., Liao, S., Yang, Y., Men, Y., He, J., Huang, Y., 2021. Tunnel boring machine vibration-based deep learning for the ground identification of working faces. *J. Rock Mech. Geotech. Eng.* 13 (6), 1340–1357.
- Pan, Y., Zhang, G., Zhang, L., 2020. A spatial-channel hierarchical deep learning network for pixel-level automated crack detection. *Autom. ConStruct.* 119, 103357.
- Phoon, K.-K., Zhang, W., 2022. Future of Machine Learning in Geotechnics. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, pp. 1–16.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Munich, Germany, pp. 234–241.
- Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, Venice, Italy, pp. 618–626.
- Sheng, D., Yu, J., Tan, F., Tong, D., Yan, T., Lv, J., 2022. Rock mass quality classification based on deep learning: a feasibility study for stacked autoencoders. *J. Rock Mech. Geotech. Eng.* <https://doi.org/10.1016/j.jrmge.2022.08.006>.
- Song, C., Wu, L., Chen, Z., Zhou, H., Lin, P., Cheng, S., Wu, Z., 2019. Pixel-level crack detection in images using SegNet. In: *Proceedings of the 13th International Conference on Multi-Disciplinary Trends in Artificial Intelligence*. Springer, Kuala Lumpur, Malaysia, pp. 247–254.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Boston, MA, USA, pp. 1–9.
- Wang, F., Ding, C., Pan, H., Zhang, S., 2022. A mesostructure-informed cohesion-based numerical method for fracture behavior of slate with foliation structure. *Int. J. Rock Mech. Min. Sci.* 160, 105252.
- Wu, C., Hong, L., Wang, L., Zhang, R., Pijush, S., Zhang, W., 2022. Prediction of wall deflection induced by braced excavation in spatially variable soils via convolutional neural network. *Gondwana Res.*
- Xu, Z., Ma, W., Lin, P., Hua, Y., 2022. Deep learning of rock microscopic images for intelligent lithology identification: neural network comparison and selection. *J. Rock Mech. Geotech. Eng.* 14 (4), 1140–1152.
- Xue, Y., Cai, X., Shadabfar, M., Shao, H., Zhang, S., 2020. Deep learning-based automatic recognition of water leakage area in shield tunnel lining. *Tunn. Undergr. Space Technol.* 104, 103524.
- Xue, Y., Jia, F., Cai, X., Shadabfar, M., Huang, H., 2022. An optimization strategy to improve the deep learning-based recognition model of leakage in shield tunnels. *Comput. Aided Civ. Infrastruct. Eng.* 37 (3), 386–402.
- Yang, X., Li, H., Yu, Y., Luo, X., Huang, T., Yang, X., 2018. Automatic pixel-level crack detection and measurement using fully convolutional network. *Comput. Aided Civ. Infrastruct. Eng.* 33 (12), 1090–1109.
- Yu, Z., Shen, Y., Shen, C., 2021. A real-time detection approach for bridge cracks based on YOLOv4-FPM. *Autom. ConStruct.* 122, 103514.
- Zhai, G., Narazaki, Y., Wang, S., Shajihan, S.A.V., Spencer Jr., B.F., 2022. Synthetic data augmentation for pixel-wise steel fatigue crack identification using fully convolutional networks. *Smart Struct. Syst.* 29 (1), 237–250.
- Zhang, A., Wang, K.C., Li, B., Yang, E., Dai, X., Peng, Y., Fei, Y., Liu, Y., Li, J.Q., Chen, C., 2017. Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network. *Comput. Aided Civ. Infrastruct. Eng.* 32 (10), 805–819.
- Zhang, J.-Z., Huang, H.-W., Zhang, D.-M., Phoon, K.K., Liu, Z.-Q., Tang, C., 2021a. Quantitative evaluation of geological uncertainty and its influence on tunnel structural performance using improved coupled Markov chain. *Acta Geotech* 16, 3709–3724.
- Zhang, J.-Z., Phoon, K.K., Zhang, D.-M., Huang, H.-W., Tang, C., 2021b. Novel approach to estimate vertical scale of fluctuation based on CPT data using convolutional neural networks. *Eng. Geol.* 294, 106342.
- Zhang, W., Li, H., Li, Y., Liu, H., Chen, Y., Ding, X., 2021c. Application of deep learning algorithms in geotechnical engineering: a short critical review. *Artif. Intell. Rev.* 54 (8), 5633–5673.
- Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Honolulu, HI, USA, pp. 2881–2890.
- Zhao, S., Zhang, D., Xue, Y., Zhou, M., Huang, H., 2021. A deep learning-based approach for refined crack evaluation from shield tunnel lining images. *Autom. ConStruct.* 132, 103934.
- Zheng, Y., Gao, Y., Lu, S., Mosalam, K.M., 2022. Multistage semisupervised active learning framework for crack identification, segmentation, and measurement of bridges. *Comput. Aided Civ. Infrastruct. Eng.* 37 (9), 1089–1108.
- Zhu, H., Azarafza, M., Akgün, H., 2022. Deep learning-based key-block classification framework for discontinuous rock slopes. *J. Rock Mech. Geotech. Eng.* 14 (4), 1131–1139.
- Zou, Q., Zhang, Z., Li, Q., Qi, X., Wang, Q., Wang, S., 2018. Deepcrack: learning hierarchical convolutional features for crack detection. *IEEE Trans. Image Process.* 28 (3), 1498–1512.



Dr. Shuai Zhao obtained his BSc and MSc degrees in Mining Engineering from Shandong University of Science and Technology, China, in 2014 and 2016, respectively, and his PhD in Civil Engineering from Tongji University, China, in 2022. Now, he works for Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong, China, as a Postdoctoral Fellow. He studied in Department of Civil and Environmental Engineering, Nanyang Technological University, Singapore, from 2019 to 2021 as an exchange student. His research interests mainly cover: (1) Refined detection and quantification of tunnel lining defects via deep learning approaches; (2) Mechanism of microbial grouting on shield tunnel seepage control; and (3) Resilience mechanism of horizontal deformation of shield tunnels in soft soil under additional grouting load via mechanics-informed machine learning.