Road crack detection based on faster R-CNN

Liangliang Zhu, Juan Qin, Zhiling Guo, Xiaobing Zhang, Shuo Feng, et al.

Liangliang Zhu, Juan Qin, Zhiling Guo, Xiaobing Zhang, Shuo Feng, Honglin Liu, "Road crack detection based on faster R-CNN," Proc. SPIE 12058, Fifth International Conference on Traffic Engineering and Transportation System (ICTETS 2021), 1205844 (22 December 2021); doi: 10.1117/12.2620322

Event: 5th International Conference on Traffic Engineering and Transportation System (ICTETS 2021), 2021, Chongqing, China

Road Crack Detection based on Faster R-CNN

Liangliang ZHU[a], Juan QIN* [a], Zhiling GUO[a], Xiaobing ZHANG[a], Shuo FENG[a], Honglin LIU[a]
[a]Tianjin University of Technology, School of Integrated Circuit Science and Engineering,
Tianjin, China
*jane.qin@tjut.edu.cn; phone 022-60214222

ABSTRACT

Road crack is an important factor that causes highway damage. There are three main types of road cracks: longitudinal crack, transverse crack and chapped crack. Road crack detection has developed from manual detection to semi-automatic detection based on image processing and full-automatic detection based on the combination of image processing and deep learning technology. This project is dedicating to detect road cracks by using Faster R-CNN (Faster Region-based Convolution Network). In the project, VGG16 is chosen as the backbone network architecture. The fully-connected layer and the pooling operation of the last layer are not used, only the convolution layer, pooling layer and ReLU layer of VGG16 are used. According to the characteristics of cracks, it generates the proportion size of the bounding box. The accuracy and recall rates are 90%, 95% and 98%, 97% respectively, on the self-made data set and on the open data set. The problem that the road crack identification algorithm is affected by the environment is improved.

Keywords: Object detection, Deep learning, Road crack, Faster R-CNN

1. INTRODUCTION

As an important part of traffic engineering, highway is an important infrastructure to promote economic development. With the increase of service life, different forms of road damage have appeared, among which road cracks such as longitudinal cracks, transverse cracks and crocodile cracks have become the main diseases of roads. If the road cracks can be repaired in time at the early stage, it will save a lot of manpower, material resources and financial resources for the later road maintenance work. The detection of road cracks can be regarded as an object detection problem.

The crack detection algorithm based on neural network is a nonlinear classification algorithm, which requires a high selection of network structure and training scheme [1]. In addition, the diversity of crack shape and the uncertainty of crack width lead to the complexity of training process. Cha et al. used the convolutional neural network architecture to segment a large size road surface image into 256*256 pixel sub-images for Convolutional Neural Network (CNN) training, and then applied the trained network model to test the road surface crack images [2]. Wang et al. reported a CNN architecture with three convolutional and two fully connected layers for asphalt pavement crack recognition [3]. In order to study the influence of model depth and image position on the performance of the model, reported a CNN model for pavement structure crack detection [4]. Tong et al. designed CNN architecture for recognition, location and feature extraction, which is used for crack detection, location, length measurement and 3D reconstruction of hidden cracks in ground penetrating radar (GPR) images [5]. Zhang et al. designed a six-layer CNN architecture for crack detection in pavement surfaces, and trained, validated, and tested the network with 640 k, 160 k, and 200 k images, respectively [6]. Fan et al. reported an efficient automatic pavement crack detection and measurement model integrated with CNN models [7]. The proposed AlexNet network introduces the object detection algorithm into the field of computer vision based on deep learning [8]. Zhang et al. proposed CrackNet for pixel-level crack detection on 3D asphalt surfaces [9]. CrackNet can detect cracks efficiently at the pixel level, but the architecture takes a long time to process.

Transfer learning models promote the applicability of CNN without high computational costs and without the need to understand how CNN operates. Transfer learning models with image data as input include Google models, i.e., the Visual GeometryGroup's VGGNet, Inception-V3 and Microsoft's ResNet. The VGGNet model with 2000 labeled images (4:1 training to test data) was used to detect various types of structural damage [10]. Feng and Zhang improved the architecture of the Inception V3 model, and trained it with transfer learning to detect structural damage in concrete water pipes [11]. Zhang and Cheng used an ImageNet-based pre-trained model to identify cracks and seal cracks in pavement images [12]. Qu et al. presented the crack detection algorithm for concrete pavement with convolutional neural network and modified the output dimension of LeNet-5 model [13]. Li et al. improved the accuracy and robustness through fully convolutional neural network based on densely connected and deeply supervised [14]. The deep learning

Fifth International Conference on Traffic Engineering and Transportation System (ICTETS 2021),
edited by Yongkang Xing, Proc. of SPIE Vol. 12058, 1205844
© 2021 SPIE · 0277-786X · doi: 10.1117/12.2620322

method greatly expands the universality and robustness of traditional methods and shows good performance in solving the problem of crack detection [15]. Faster R-CNN has good general characteristics and robustness over multiple datasets [16].

The advantage of Faster R-CNN algorithm is that the RPN (Regional Proposal Networks) module improves the efficiency of objection detection. In this paper, VGG16 network is used to extract the image features of pavement cracks. The RPN module is used to generate high quality anchors. And the rules for generating anchors have been modified, the ratio of anchor frame is changed from 0.5, 1, 2 to 0.3, 1, 3 based on the original algorithm. The size of the anchor frame has been changed from 8, 16, 32 to 4, 8, and 16. Then, the feature map and the high-quality anchors generated by the RPN module should be jointly transmitted to the ROI Pooling layer to classify and locate the road cracks. This paper will realize the detection of road cracks based on Faster R-CNN. The innovation of this project lies in the application of the Faster R-CNN algorithm to the road crack identification, and the improved results have been achieved.

2. PROPOSED ALGORTHM

In this paper, the object detection and recognition algorithm based on Faster R-CNN is proposed, and its algorithm flow is shown in Figure 1. As a CNN network object detection algorithm, Faster R-CNN uses a set of Conv Layer + ReLU Layer + Pooling Layer to extract the features of the input images, and then applies it to the subsequent RPN network and the fully connected layer. The first correction process of the object detection box is also carried out. ROI Pooling makes use of the proposals by RPN and the feature maps by the last layer of VGG16 to generate the fixed-size proposals feature map, and access the subsequent network to realize object identification and positioning through fully-connected Layer.

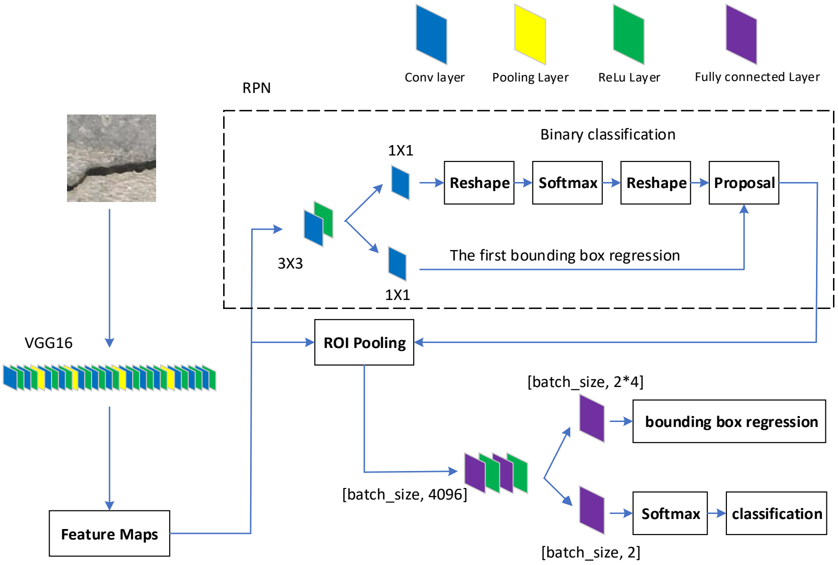


Figure 1. Overall network architecture

Classifier forms a fixed-size Feature map of the ROI Pooling layer for full connection operation, and Softmax is used for classification of specific categories. At the same time, L1 Loss is used to complete the regression operation of the bounding box to obtain the precise position of the crack.

Most road cracks are long and narrow. According to the morphological characteristics of cracks, the ratio of anchor frame is changed from 0.5, 1, 2 to 0.3, 1, 3 based on the original algorithm. The size of the anchor frame has been changed from 8, 16, 32 to 4, 8, and 16 to improve the recognition rate of the trained model on the smaller cracks.

3. THE NETWORK ARCHITECTURE

3.1 VGG16 Model

In this paper, VGG16 is used to extract the cracks features. The VGG16 network includes 13 Conv layers, 13 ReLU layers and 5 MaxPooling layers. But the algorithm doesn't use the last pooling layer. The algorithm supports the input of images of any size, but before entering the network, the input image needs to be set to the scale of normalization. That is, the input image should be clipped or zeroed. In the convolution layer, the size of the output feature graph is calculated as shown as (1). Where, $\mathrm{output}_{\mathrm{size}}$ represents the size of the output feature layer, $\mathrm{input}_{\mathrm{size}}$ represents the input size, $\mathrm{ker\,nel}_{\mathrm{size}}$ represents the size of the convolution kernel, pad represents the filling size, and stride represents the convolution step.

$$\mathrm{output}_{\mathrm{size}} = \frac{\left(\mathrm{input}_{\mathrm{size}} - \mathrm{ker\,nel}_{\mathrm{size}} + 2\mathrm{pad}\right)}{\mathrm{stride}} + 1 \qquad (1)$$

Besides, the parameters of network structure are shown in table 1. Therefore, after 4-layer pooling, the image size becomes 1/16 of the original. The Feature maps obtained from the VGG16 network are used for RPN and fully connected layers.

Table 1. VGG16 related parameters

Layer	**Conv1**	**Pool1**	**Conv2**	**Pool2**	**Conv3**	**Pool3**	**Conv4**	**Pool4**	**Conv5**
Kernel_size	3x3	2x2	3x3	2x2	3x3	2x2	3x3	2x2	3x3
Feature_map	224*224		112*112		56*56		28*28		14*14
Input_size	3	64	64	128	128	256	256	512	512
Output_size	64	64	128	128	256	256	512	512	512

3.2 RPN

RPN is mainly used to extract proposals, and its processing is shown in Figure 2. The introduction of RPN can be said to truly integrate the objection detection process into the neural network. In RCNN and Fast-RCNN, the algorithm used to extract proposals is usually Selective Search [17], but it is time-consuming. The introduction of RPN not only consumes less time, but also is easy to combine with Fast -R-CNN. After the Feature maps enter the RPN network, a 3*3 convolution operation is carried out, followed by two branches. One is the classification operation, which is used to determine whether the Anchor belongs to the foreground or the background. It is a binary classification problem. The other branch goes into the first adjustment of the bounding box, thus discarding a large number of inaccurate proposals after NMS (Non-maximum suppression).

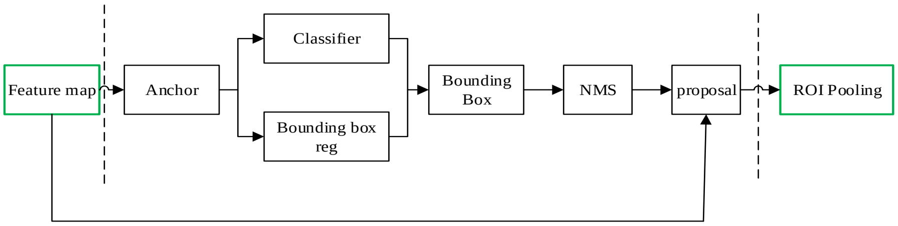


Figure 2. Regional Proposal Networks

1) CLASSIFIER AND SOFTMAX

The Softmax function converts the multi-classification output into probability in (2) [18]. Where, x represents the input, W_{y} represents the weight of the layer, and f_{y} represents the output result of the layer. Besides, N_c represents the number of categories. The numerator maps the real number output to zero to infinity by exponential function, and the denominator adds all the results of the categories to normalize.

In this paper, the softmax function will be used for two classification tasks. One is used to distinguish between foreground and background, and the other is used for specific classification, namely whether it is a crack or not.

$$\begin{cases} f_{\mathrm{y}} = \mathrm{W}_{\mathrm{y}} \times x \\ \mathrm{soft\,max}(f)_y = \dfrac{\exp(f_{\mathrm{y}})}{\sum_{c=1}^{\mathrm{N}_c} \exp(f_c)} \end{cases} \qquad (2)$$

2) BOUNDING BOX REGRESSION

The proposal can be represented by a set of four-dimensional vectors (m,n,w,h), with m,n representing the central coordinates of the proposal and w,h representing the width and height. **P** represents the original Proposal and **G** represents the Ground Truth. The bounding box regression is used to find a mapping relationship $\Re$, so that the original input proposal **P** can get a proposal **F** closer to **G** through the mapping $\Re$. The mapping $\Re$ can be described in (3).

$$\Re\left(\mathrm{P}_m,\mathrm{P}_n,\mathrm{P}_w,\mathrm{P}_h\right)=\left(\mathrm{F}_m,\mathrm{F}_n,\mathrm{F}_w,\mathrm{F}_h\right)\approx(\mathrm{G}_m,\mathrm{G}_n,\mathrm{G}_w,\mathrm{G}_h) \tag{3}$$

3.3 NMS and ROI Pooling

NMS has very important applications in the field of computer vision, such as video object tracking, data mining, 3D reconstruction, object recognition and texture analysis. NMS is a process of finding a local maximum. For the same object, many bounding boxes will be generated in the image during object detection, and then a score will be obtained, which will be entered into the classifier. Select the bounding box with the highest score to calculate the Intersection-over-Union (IOU) of other bounding boxes. If the IOU is greater than the threshold, delete the corresponding bounding box.

ROI Pooling enables mapping from the original image area to the last convolution area in VGG16, and pooling to a fixed size. This layer samples the region proposal to a 7*7 feature map. In VGG16, CONV5_3 has 512 feature maps, so all region proposals correspond to a 7*7*512 dimension feature vector as the input of the fully connected layer.

4. EXPERIMENT

Faster R-CNN includes two loss functions: loss of the RPN network and loss of the RCNN network. Each loss includes classification loss and regression loss of adjustment bounding box. Softmax Loss is used for classification loss and Smooth L1 Loss is used for regression loss. Smooth L1 loss is defined in (4):

$$\mathrm{smooth}_{\mathrm{L1}}(r)=\begin{cases}0.5r^2\dfrac{1}{\sigma^2}, & if\ |r|<\dfrac{1}{\sigma^2}\\ |r|-0.5, & otherwise\end{cases} \tag{4}$$

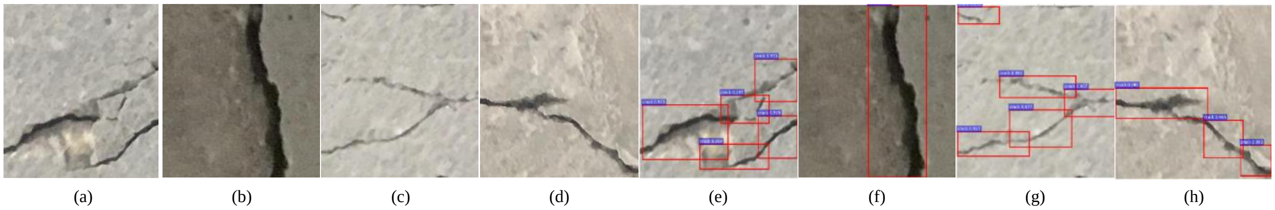
Therefore, the loss of RPN and RCNN can be calculated in (5).

$$L\left(\{p_i\},\{t_i\}\right)=\frac{1}{\mathrm{N}_{\mathrm{cls}}}\sum\nolimits_i L_{\mathrm{cls}}\left(p_i,p_i^*\right)+\mu\frac{1}{N_{reg}}\sum\nolimits_i p_i^* L_{\mathrm{reg}}(t_i,t_i^*) \tag{5}$$

Where, i is anchors index.

After 1.4 million iterations, the RPN classification loss is 0.015, and the suggestion box regression loss is 0.084. The classification loss of RCNN is 0.024, while the regression loss of Bounding Box is 1.82*1e-3. The overall loss function value is 0.069.

In the experiment, an open dataset created by researchers at Middle East Technical University is used. The data set contains 40,000 color images (pixel 227*227), of which 20,000 images have crack information and 20,000 images have no crack. As shown in Figure 3, after adjusting the ratio of the anchor frame, it can be found that the tiny cracks in the upper left corner of (g).



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 3. images in public dataset. (a), (b), (c), (d) represent the original images in the public dataset, and (e), (f), (g), (h) represent the corresponding detection image. Using the trained model, the cracks in the images can be accurately located and identified.

In addition, the precision of the model is about 98% and the recall rate is about 97%. The precision can be calculated in

(7). TP represents the number of images with cracks and detected, FP represents the number of images without cracks but incorrectly detected as having cracks.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

The recall rate can be calculated in (8). FN represents the number of images with cracks but detected as non-cracks.

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

The value of F_1 can be calculated in (9).

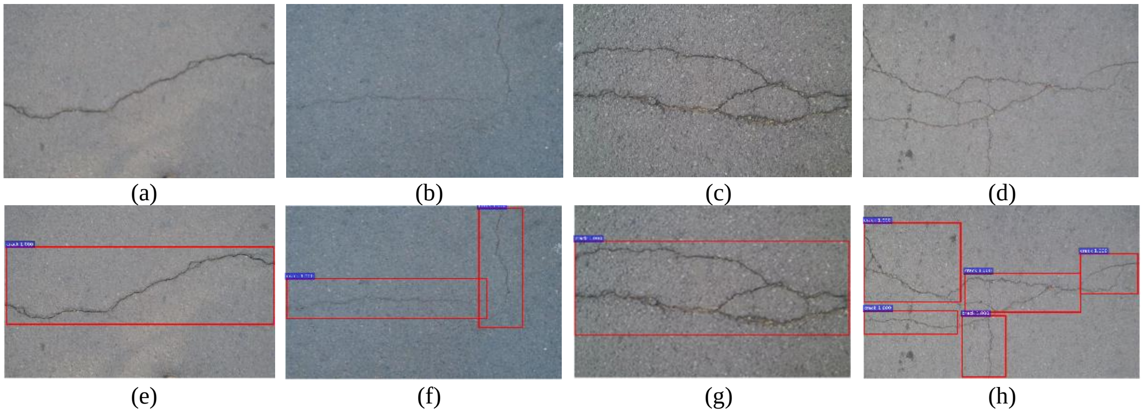
$$F_1 = \frac{2\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

In addition, we compared the recognition results with the neural network recognition algorithm by Li Yongxu *et al.* using CLAHE algorithm [1]. This is shown in Table 2.

Table 2. Results of different methods are compared

	the evaluation index		
	Precision	Recall	F_1
CLAHE-CNN-B	0.835	0.899	0.866
CLAHE-CNN-C	0.927	0.973	0.949
Ours	**0.980**	**0.970**	**0.974**

In the homemade data set, a total of 151 pictures of road cracks are collected. The pixel size of each image in the dataset is 480*320. As shown in Figure 4, the dataset contains information about longitudinal cracks, transverse cracks, and cracks to ensure that the network model can learn the experience of identifying various cracks.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4. images in homemade dataset. (a), (b), (c), (d) represent the original images in the homemade dataset, and (e), (f), (g), (h)

The performance evaluation index results of the training model on the homemade dataset and the public dataset are shown in Table 3. According to the table, all evaluation indexes on the public dataset are better than the homemade dataset. The reason is that the images in the public dataset have the feature of high similarity, so the network model can better recognize the images with high similarity after learning the experience. However, its disadvantage is that the crack images in real environment are often asphalt roads, which are quite different from the images in the dataset. Therefore, the network model obtained can get better results on the test set of the public dataset, but its generalization ability is relatively weak.

Table 3. Comparison between homemade and public datasets

	the evaluation index		
	Precision	Recall	F1
Homemade dataset	0.900	0.950	0.924
Public dataset	0.980	0.970	0.974

5. CONCLUSION AND PROSPECT

In the paper，this recognition algorithm is applied to the actual road detection engineering field based on Faster R-CNN, instead of using the traditional road recognition algorithm based on image processing and morphology operation. Thus, the problem that the road crack identification algorithm is affected by the environment is improved. According to the characteristics of cracks, the proportion size of the bounding box generated is modified. The accuracy and recall rates are 90%, 95% and 98%, 97% respectively, on the self-made data set and on the open data set created by researchers at Middle East Technical University. Compared with traditional algorithms, the road crack recognition algorithm based on deep learning has better recognition rate and generalization ability. Besides, the model can be trained on a PC and transplanted to Python productivity for Zynq (PYNQ) to form a tiny detection system. Therefore, a small cost, convenient detection device can be realized. In addition, the embedded function of PYNQ can be used to give detection personnel corresponding prompts (such as flashing LED and buzzer sounding) when cracks are detected.

This work is supported by the National college students' innovation and entrepreneurship training program 202010060003.

REFERENCES

[1] Y.X. Li, Zh.Zh. Xie, W.J. Tang, "Crack identification method of asphalt pavement based on convolutional neural network," *West China Transportation Science and Technology*, no. 06, pp.19-22, 2020.

[2] Y. J. Cha, W. Choi, and O. Buyukozturk, "Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks," *Comput-Aided Civ Inf,* vol. 32, no. 5, pp.361-378, May 2017.

[3] K.C.P. Wang, *et al.*, "Deep Learning for Asphalt Pavement Cracking Recognition Using Convolutional Neural Network," *International Conference on Highway Pavements and Airfield Technology 2017*, Philadelphia, PA, USA, 2017, pp.166–177.

[4] L. Pauly, D. Hogg, R. Fuentes, H. Peel, "Deeper networks for pavement crack detection," *In Proceedings of the 34th ISARC*, Taipei, Taiwan, 2017, pp.479–485.

[5] Z. Tong, J. Gao, and H. T. Zhang, "Recognition, location, measurement, and 3D reconstruction of concealed cracks using convolutional neural networks," *Constr Build Mater,* vol. 146, pp. 775-787, Aug 15 2017.

[6] L. Zhang, F. Yang, Y.D. Zhang, Y.J. Zhu, "Road crack detection using deep convolutional neural network," *In Pro. ICIP 2016*, Phoenix, AZ, USA, 25–28 September 2016, pp.3708–3712.

[7] Z. Fan *et al.*, "Ensemble of Deep Convolutional Neural Networks for Automatic Pavement Crack Detection and Measurement," *Coatings,* vol.10, no. 2, Feb 2020.

[8] M. R. Jahanshahi, S. F. Masri, C. W. Padgett, and G. S. Sukhatme, "An innovative methodology for detection and quantification of cracks through incorporation of depth perception," *Mach Vision Appl,* vol. 24, no. 2, pp. 227-241, Feb 2013.

[9] Zhang *et al.*, "Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces Using a Deep-Learning Network," *Comput-Aided Civ Inf,* vol. 32, no. 10, pp. 805-819, Oct 2017.

[10] Y. Q. Gao and K. M. Mosalam, "Deep Transfer Learning for Image-Based Structural Damage Recognition," *Comput-Aided Civ Inf,* vol. 33, no. 9, pp. 748-768, Sep 2018.

[11] C. C. Feng, H. Zhang, S. Wang, Y. L. Li, H. R. Wang, and F. Yan, "Structural Damage Detection using Deep Convolutional Neural Network and Transfer Learning," *Ksce J Civ Eng,* vol. 23, no. 10, pp. 4493-4502, Oct 2019.

[12] K. G. Zhang, H. D. Cheng, and B. Y. Zhang, "Unified Approach to Pavement Crack and Sealed Crack Detection Using Preclassification Based on Transfer Learning," *J Comput Civil Eng,* vol. 32, no. 2, Mar 2018.

[13] Z. Qu, J. Mei, L. Liu, and D. Y. Zhou, "Crack Detection of Concrete Pavement With Cross-Entropy Loss Function and Improved VGG16 Network Model," *IEEE Access,* vol. 8, pp. 54564-54573, 2020.

[14] H. F. Li, J. P. Zong, J. J. Nie, Z. L. Wu, and H. Y. Han, "Pavement Crack Detection Algorithm Based on Densely Connected and Deeply Supervised Network," *IEEE Access,* vol. 9, pp. 11835-11842, 2021.

[15] R. Kalfarisi, Z. Y. Wu, and K. Soh, "Crack Detection and Segmentation Using Deep Learning with 3D Reality Mesh Model for Quantitative Assessment and Integrated Visualization," *J Comput Civil Eng,* vol. 34, no. 3, May 1 2020.

[16] S. Q. Ren, K. M. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE T Pattern Anal,* vol. 39, no. 6, pp. 1137-1149, Jun 2017.

[17] Z. Fang，Z. Cao，Y. Xiao，K. Gong，J. Yuan, "MAT: Multianchor Visual Tracking With Selective Search Region," *IEEE Trans Cybern*. no. 99, pp.1-15, Feb. 2020.

[18] W. F. Ou *et al.*, "LinCos-Softmax: Learning Angle-Discriminative Face Representations With Linearity-Enhanced Cosine Logits," *IEEE Access*, vol. 8, pp. 109758-109769, 2020.