Method and Research of Concrete Pavement Crack Detection Based on Convolution Neural Network

Fanli Wu

School of Civil Engineering, Zhejiang University of Technology, Hangzhou, China.

13250812688@163.com

Abstract. Being essential to civil engineering, concrete pavement is prone to cracks in the process of use, which affects the bearing capacity and driving safety of concrete roads. In this paper, a convolution neural network (CNN) is used to identify concrete pavement cracks, with the Concrete Crack Images for Classification of Mendeley Data utilized to test the CNN performance. Mendeley Data is used to compare the performance of concrete pavement crack detection between CNN and other three classical network models. The experimental results show that Accuracy, Precision, Recall, and F1-Score are 0.97, 0.98, 0.96, and 0.97 respectively. Compared with image recognition methods such as RNN-LSTM, Contour Threshold, and ResMLP, the Accuracy of CNN is the highest in this paper, which proves that CNN has high recognition performance in concrete pavement crack recognition. Besides, it is of high reference value for the behavior and decision-making of concrete pavement crack recognition.

Keywords: Concrete Pavement; Mendeley Data; Convolution Neural Network; Crack Detection; Civil Engineering.

1. Introduction

Functions of infrastructures, including buildings, bridges, dams, etc., will be easily deteriorated over time in use [1]. To ensure their safety and usability, the physical and functional conditions of infrastructures must be inspected and evaluated [2]. For example, the damage caused by concrete cracks undercuts infrastructure such as roads and pavements, but it is unrealistic to analyze a large amount of data and identify patterns by human or traditional computer programs using rule-based methods [3].

To accurately identify the cracks in concrete pavement, many related studies have been put forward. For example, Wu et al. [4] proposed a new multi-scale deep learning method for concrete crack detection, and the model they designed had good portability. Abdelkader [5] et al. Combined an artificial neural network with a frog leaping algorithm and improved the prediction performance of the artificial neural network by optimizing the connection weights between neurons. Gupta [6] et al. used an ERT conductivity map of electrical resistance tomography to identify and characterize the severity, location, and mode of cracks. Zhong [7] et al. put forward an algorithm for concrete pavement crack detection based on an attention mechanism and multi-feature fusion to realize fast focusing of cracks. Yu [8] et al. designed an improved Dempster-Shafer algorithm for decision-level image fusion, which can accurately identify crack contours in the case of limited area prediction errors. Although there is constant research on concrete pavement crack detection, its performance is still not desirable enough.

With the development of deep learning and computing devices, CNN are widely used in computer vision, such as image recognition. Russakovsky [9] et al. analyzed in detail the current situation in large-scale image classification and target detection as well as the key breakthrough in classified target recognition. Zhen [10] et al. proposed a new semi-supervised CNN method, which optimized the classification of thresholds and improved the reliability of unlabeled recognition samples. Wang [11] et al. Used the automatic search strategy based on deep neural networks to find the network architecture suitable for HRS image recognition. Mujahid [12] et al. proposed a lightweight model based on YOLOv3 and DarkNet-53 CNN for gesture feature extraction and recognition. According to the results, the algorithm has high recognition accuracy and efficiency for concrete pavement crack recognition. Wu [13] et al. designed a graph neural network model based on multi-view image feature similarity, which has a strong generalization ability and the anti-interference ability for forest fire recognition. Although the method of concrete pavement crack recognition based on artificial intelligence has certain achievements in recognition, low recognition performance still exists.

In this paper, the performance of concrete pavement crack detection by CNN is studied. Concrete Crack Images for Classification of Mendeley Data [14] [15] is used as the dataset. The method of this paper is based on a CNN, which is a simple and effective detection method, providing guidance and suggestions for decision-making of related industries.

2. Related Technology

2.1 Artificial Neural Networks

Artificial Neural Network (ANN) assists in architecture design to improve the performance by finding the best parameter values, which can extract scene detection patterns and trends. Generally speaking, the more hidden layers, neurons, and the activation function in ANN, the better the result. ANN structural diagram is shown in Figure 1.

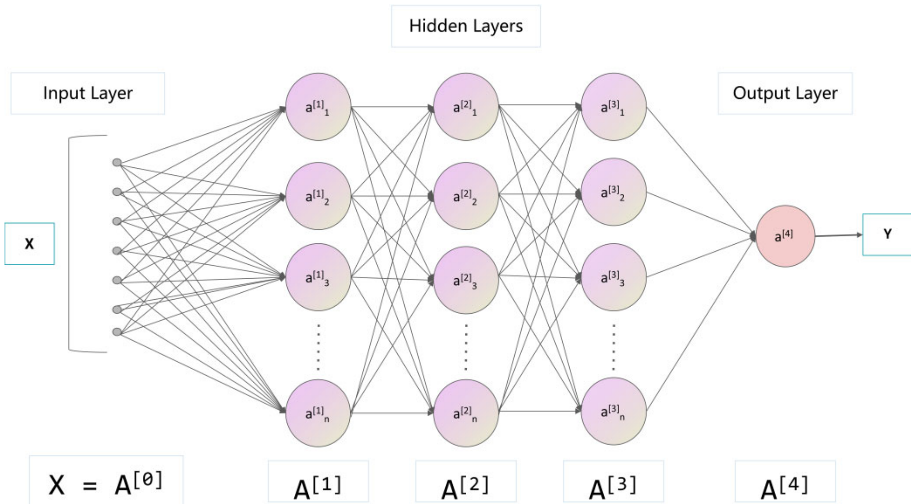


Figure 1. ANN Structural Diagram

In Figure 1, X represents the input of data, $a[1]_1$ represents the first node of the first hidden layer, arrows represent the weighted average relationship between nodes, and each arrow represents a weight. The last layer is the output layer, which represents the prediction results of the neural network. ANN has many connection modes. Figure 1 shows a fully connected neural network, which represents that every node between adjacent network layers has a connection relationship. The structural diagram of the fully connected network is shown in Figure 2.

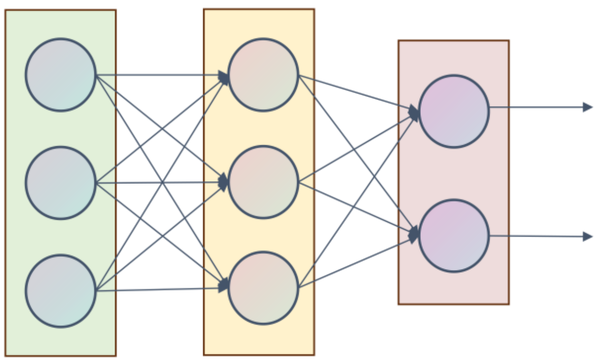


Figure 2. Full Connection Structural Diagram

2.2 Convolution Neural Network

Convolutional Neural Network (CNN) aims to train the extraction of relevant features directly from pixel images [16], which usually consists of the input layer, convolution layer, pooling layer, and output layer. The convolution layer is shown in Figure 3.

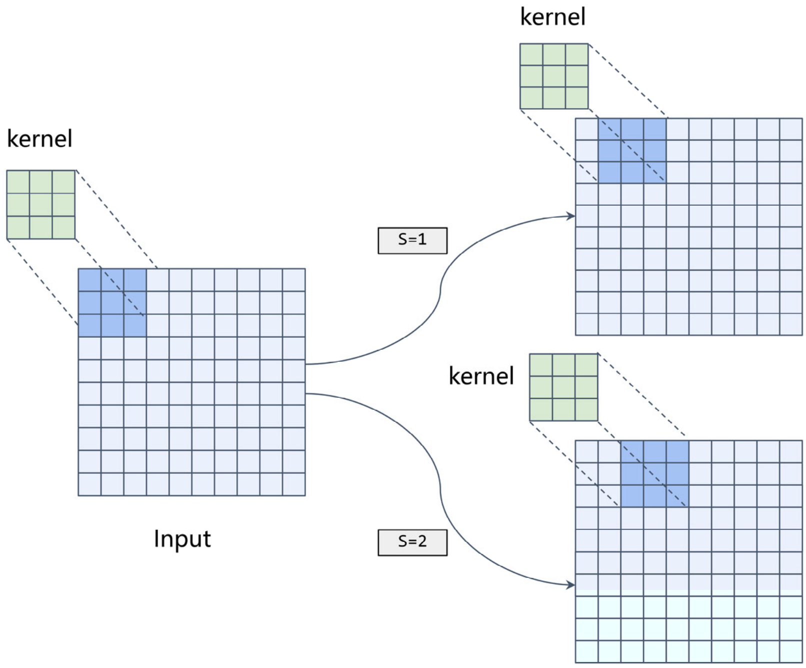
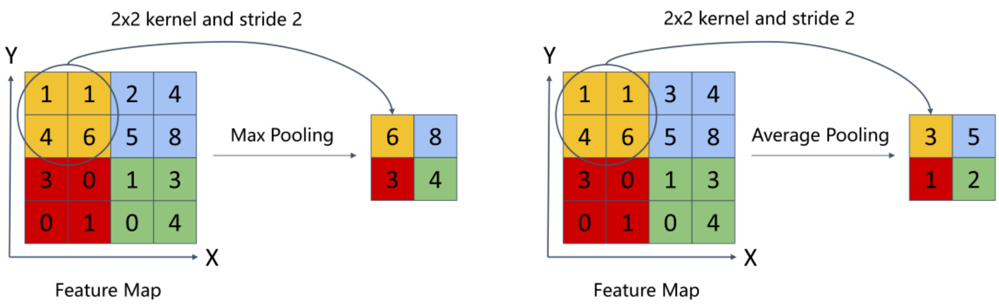


Figure 3. Structural Diagram of Convolution Layer

The convolution layer is to abstract image feature information through convolution operation, which usually contains multiple convolution kernels, with each corresponding to a feature map. The parameters of the convolution layer include the sliding step and the size of the convolution kernel. During the convolution of the feature graph, the convolution kernel slides along a transverse or longitudinal direction according to step size. The output feature map of the convolution layer also needs to undergo activation function operation to get more nonlinear features. The shallow convolution layer can extract the edge and other position information of the image. The deep convolution layer can extract more abundant and advanced semantic information. Using CNN with the characteristics of parameter sharing and sparse connection, the convolution layer can greatly reduce its parameter quantity and complexity.

Pooling divides the input feature map into several local regions, which reduces the size of the feature map by counting the data characteristics in each local region. The core of the convolution layer is the convolution core, and the core of pooling is the pooling core. When pooling the feature map, the pooling core strides according to the pooling step size. Common pooling methods are max pooling and average pooling, as shown in Figure 4.



(a) Max Pooling (b) Average Pooling

Figure 4. Structural Diagram of Pooling

In Figure 4, the max pooling uses the maximum to represent the local area, which makes the overall feature contrast of the image significant, while the average pooling uses the average to represent the local area, which makes the overall feature information of the image smoother. When training the model, the pooling can help the network focus on learning the pixel features of the image, which is helpful to provide the generalization and robustness of the network and avoid the network over-fitting.

2.3 Model Evaluation Index

In classification and recognition, 1 and 0 are usually used to represent two categories. 1 means that an event exists and 0 means the opposite. In image recognition, whether a certain feature exists or not is also a binary classification problem. The state relationship of the two categories is represented by the confusion matrix, as shown in Table 1.

Table 1. Evaluation Index of Confusion Matrix

Confusion Matrix		Predicted	
		0	1
Actual	0	TN	FP
	1	FN	TP

In Table 1, the case where the predicted is 0 and the actual is 0 is TN (True Negative). The case where the predicted is 1 and the actual is 0 is FP (False Positive). The case where the predicted is 0 and the actual is 1 is FN (False Negative). The case where the predicted is 1 and the actual is 1 is TP (True Positive).

Based on the confusion matrix, the common evaluation indexes of the classification model include Accuracy, Loss, Precision, Recall, and F1-Score.

Accuracy represents the proportion of all samples that are correctly classified, as in Formula 1:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Precision represents the proportion of samples that are actually 1 in the sample predicted as 1 by the model, as in Formula 2:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recall represents the proportion of samples that are actually 1 that are correctly predicted to be 1, as in Formula 3:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

F1-Score is the harmonic average of recall rate and accuracy rate, which represents the comprehensive evaluation of the prediction performance of the model, as shown in Formula 4:

$$F1-Score = \frac{2 \bullet Recall \bullet Precision}{Recall + Precision} \tag{4}$$

3. CNN-Based Concrete Pavement Cracks Detection

In this paper, CNN is used to identify cracks in concrete pavement, with its structural diagram shown in Figure 5.

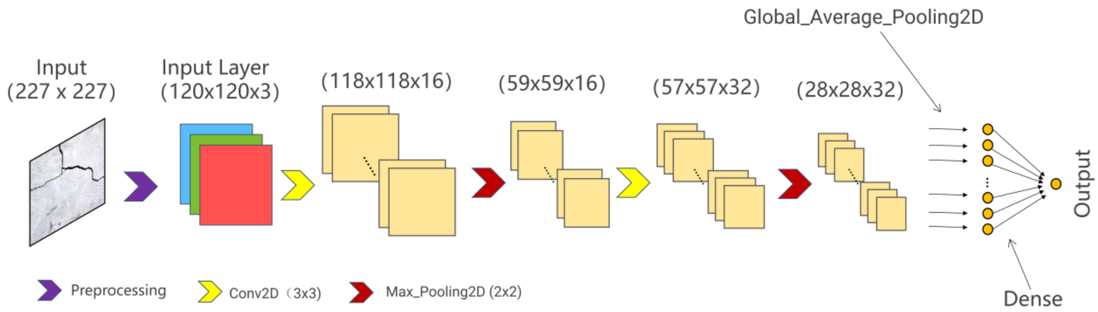


Figure 5. CNN Structural Diagram

In Figure 5, 227×227 pixel image samples are uniformly used for data input, and 3 as the depth represents RGB value, thus obtaining three feature maps. After preprocessing, the image is scaled to 120×120 as a convolution input layer. In the first convolution layer, 16 3×3×3 Filters are used to convolve the image, and 16 118× 118 feature maps are obtained, which extract 16 different features of the image. Max-Pooling adopts a 2×2 max pooling kernel, which continuously samples 16 feature maps and gets 16 feature maps sizing 59×59. Thus, non-maximum values in local areas are eliminated, parameters and redundant information are reduced, and the robustness of extracted features is improved. The second convolution layer has 32 above Filters, each of which convolves 16 feature maps after the previous down-sampling together, resulting in a new 57×57 feature map. Then, the second Max-Pooling continues to sample 32 feature maps, which obtains 32 28×28 feature maps. Using global average pooling, all pixel values of 32 feature maps are added and averaged, and the corresponding feature maps are represented by the obtained node values, thus reducing the parameters and over-fitting. Finally, there is a full connection layer, each neuron is connected with 32 nodes in the previous layer, and the output of the whole network is obtained. Based on Figure 5, the training parameters of each convolution and pooling layer of CNN are shown in Table 2.

Table 2. Training Parameters of Convolution and Pooling Layers

Operation (Layer)	Output Size	Channels	Parameters
Data Input	227×227	3	/
Input Layer	120×120	3	0
Conv2d	118×118	16	448
Max_Pooling2d	59×59	16	0
Conv2d	57×57	32	4640
Max_Pooling2d	28×28	32	0
Global_Average_Polling2d	/	32	0
Dense	/	1	33

According to Table 2, the CNN used in this paper has a simple structure, such as fewer layers and parameters, which makes the training of the network model convenient.

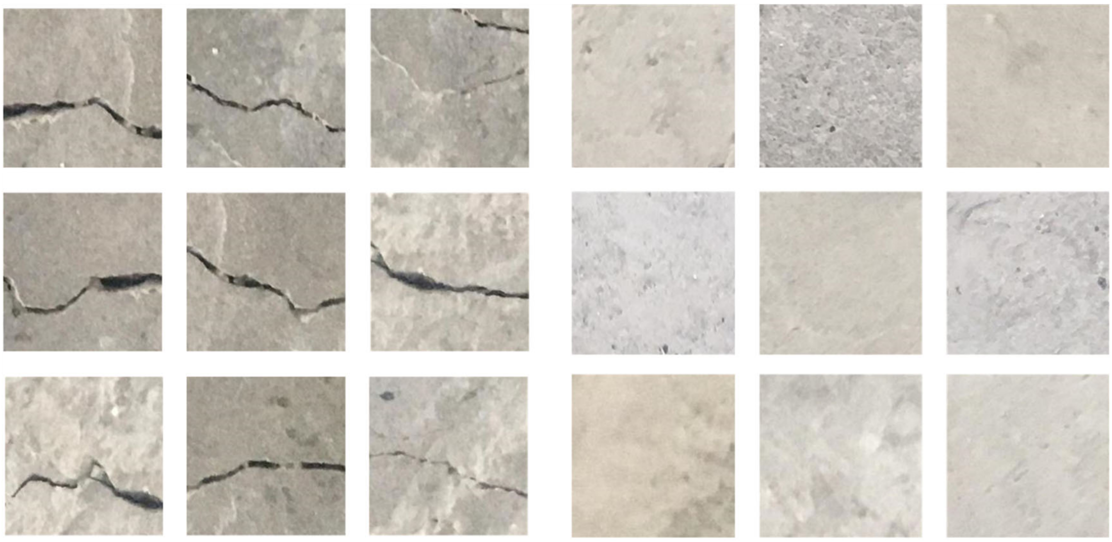
4. Experimental Design and Analysis

4.1 Experimental Environment

Using Tensorflow as the deep learning framework, this paper deploys with Baidu AI STUDIO and adopts CNN to use standardized operation after each convolution layer, thus speeding up the convergence speed of training. After standardization, ReLU is used as the activation function to increase the nonlinear relationship between each layer of the network to overcome the over-fitting and gradient disappearance. Meanwhile, Adam is used to optimizing the model. The parameters of the model are set as follows: the size of the input image is 227×227, the number of Epoch in the training is 100, and each Epoch saves the model once. All experiments were implemented in Windows 11 and Tensorflow using AMD Ryzen 7 5800H CPU and NVIDIA GeForce RTX 3050 Laptop GPU.

4.2 Experimental Dataset

The concrete pavement crack dataset used in this paper is the Concrete Crack Images for detection of Mendeley Data. It includes 40,000 concrete pavement crack images, which are composed of 500 high-resolution (4032x3024 pixels) concrete pavement images cropped or enhanced by high-resolution shooting equipment. Mendeley Data is divided into two categories with or without cracks and each category has 20,000 images with each sample based on Channel 3 of the RGB. The Mendeley Data has 28,000 crack images used to train the CNN and 12,000 images are used as test sets. As a large publicly accessible pavement crack dataset with corresponding labels, Mendeley Data is shown in Figure 6.



(a) Positive (b) Negative

Figure 6. Original Image of Concrete

Figure 6 (a) is a sample with cracks on the pavement, while Figure 6 (b) is one without cracks on the pavement.

4.3 Data Preprocessing and Loading of Image

The preprocessing of the dataset used in this paper can be seen in Figure 7.

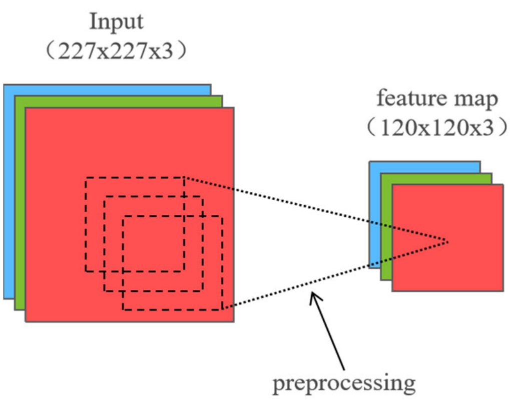


Figure 7. Data Preprocessing Model

In Figure 7, when the RGB data in images of the training set and test set are loaded, inputting the pixel values of the original image may cause the over-fitting of the algorithm, so the node data of each network is normalized. After normalization, the loading model is built. The scaling factor of the picture input is 1/255, and scaling the pixel value to between 0 and 1 is beneficial to the model convergence.

4.4 Experimental Results and Analysis

The training set and test set are used to train and test the CNN respectively in this paper. The confusion matrix obtained from the test is shown in Table 3.

Table 3. Confusion Matrix

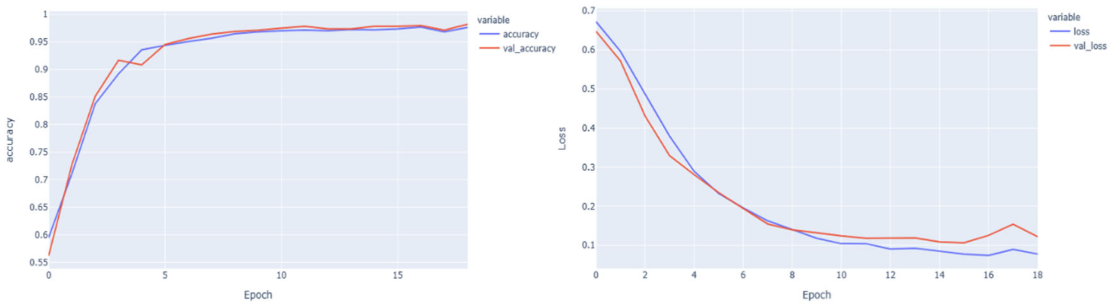
Confusion Matrix		Predicted	
		Negative	Positive
Actual	Negative	899	18
	Positive	34	849

In Table 3, TN is 899, FP is 18, FN is 34, and TP is 849. This shows that the CNN in this paper has a high prediction performance. Based on the confusion matrix in Table 3, the Accuracy, Precision, Recall, and F1-Score of the CNN in this paper are shown in Table 4.

Table 4. Evaluation Index

Classification Report	Precision	Recall	F1-score	Support
Negative	0.96	0.98	0.97	917
Positive	0.98	0.96	0.97	883
Macro avg	0.97	0.97	0.97	1800
Accuracy	/	/	0.97	1800

A
ccording to Table 4, the CNN in this paper can predict and identify cracks with an accuracy rate of 97%; predict images with cracks with an accuracy rate of 98%; accurately find images with cracks with a recall rate of 96%; and meet the threshold of higher accuracy rate and recall rate of 0.97 simultaneously. It shows that CNN has better performance for crack detection. The acquisition and loss function values of CNN for the training set and test set in this paper can be seen in Figure 8.



(a) (b)

Figure 8. (a) Diagram of Model Prediction Accuracy Rate, (b) Diagram of Model Prediction Loss Rate

As seen in Figure 8 (a), the CNN in this paper has high accuracy in both the training set and testing set under the condition of fewer epochs. When the epoch is 5, the CNN accuracy in this paper is about 95% without big fluctuation in the following epochs. According to Figure 8 (b), the Loss function value of the CNN drops very fast and there is no big fluctuation when it drops to 10. This shows that the convergence speed of CNN in this paper is fast with high stability.

4.5 Comparative Experimental Analysis

To verify the CNN effectiveness studied in this paper, the CNN performance in this paper is compared with three classical ANN models based on Mendeley Data. The three network models are as follows.

(1) RNN cyclic neural network divides pictures into several sequences for training, but it has low accuracy and a long training time. End-to-end LSTM mechanism can better solve the RNN gradient disappearance, and the improved RNN model has shown good results in image recognition;

(2) It is a classical image segmentation method to use OpenCV Computer Vision Library and set Contour Threshold to classify and recognize all images;

(3) ResMLP is a simple residual network, which can be used as a multi-layer perceptron architecture for image classification. When using a large amount of data enhancement for training, it has good accuracy. However, there is instability in accuracy in sparse_categorical_accuracy;

The performance comparison diagram of the four network models is shown in Figure 9.

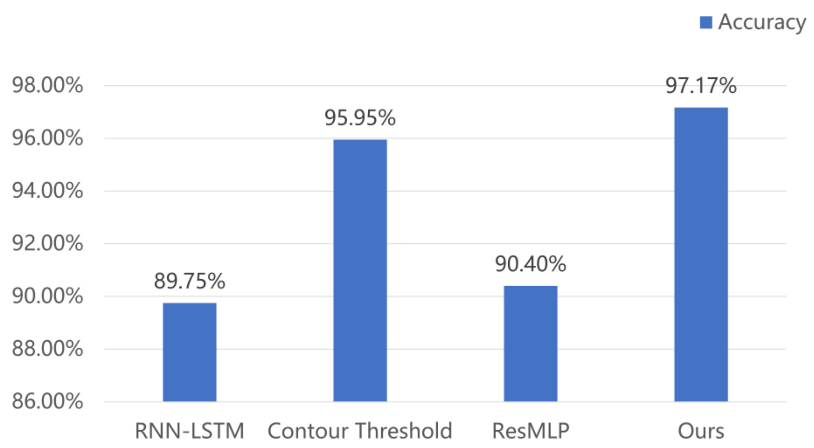


Figure 9. Comparative Experimental Results of Accuracy in Mendeley Data

As can be seen from Figure 11, the prediction accuracy of RNN-LSTM, Contour Threshold, ResMLP, and CNN for this dataset is 89.75%, 95.95%, 90.40%, and 97.17% respectively. This shows that the CNN in this paper has higher performance in identifying cracks in concrete pavement than other classical network models.

5. Summary and Prospect

The CNN used in this paper is a research and application in the field of concrete crack recognition. By enhancing the training images collected from actual scenes, the amount of training data is increased, and the generalization ability and robustness of the network model are improved in image feature recognition. According to experimental results, compared with the other three image recognition methods, the method used in this paper has higher accuracy and can be better applied to engineering practice. The main contributions of this paper are: (1) Using simple and efficient CNN to identify cracks in concrete pavement. (2) Explore the recognition performance of simple CNN on Mendeley Data.

In future research, we will further explore the crack detection of concrete pavement from the following aspects:

(1) The structure of CNN itself will be optimized and improved.

(2) Combine drones to shoot field samples, improve recognition efficiency, give play to practical application value, and lay the foundation for intelligent management and automatic monitoring;

(3) CNN can only segment and identify concrete cracks, but fails to measure the size of concrete cracks. In the follow-up, we will focus on the semantic segmentation of concrete cracks and determine the size of concrete cracks.

References

[1] Cha, Y. J., Choi, W. & Buyukozturk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. Computer-Aided Civil and Infrastructure Engineering. 32. 361-378. 10.1111/mice.12263.

[2] Koch, C., Doycheva, K., Kasi, V., Akinci, B. & Fieguth, P. (2015). A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. Advanced Engineering Informatics. 29. 10.1016/j.aei.201 5. 01.008.

[3] Shanaka, K. B., Thilakarathna, S., Perera, J., Arashpour, M., Sharafi, P., Teodosio, B., Shringi, A. & Mendis, P. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. Automation in Construction. 141. 104440. 10.1016/j.autcon.2022.104440.

[4] Wu, X. Y., Ma, J. F., Sun, Y., Zhao, C. Q. & Basu, A. (2021). Multi-scale deep pixel distribution learning for concrete crack detection. 6577-6583. 10.1109/ICPR488 06.2021.9413312.

[5] Eslam, M. A., Abobakr, A. S. & Alfalah, G. (2021). Analyzing concrete cracks' characteristics using meta-heuristic computing. 10.1109/DASA53625.2021.9682352.

[6] Gupta, S., Lin, Y. A., Lee, H. J., Buscheck, J., Wu, R. Z., Lynch, J., Garg, N. & Loh, K. (2021). In situ crack mapping of large-scale self-sensing concrete pavements using electrical resistance tomography. Cement and Concrete Composites. 122. 104154. 10.1016/j.cemconcomp.2021.104154.

[7] Qu, Z., Chen, W., Wang, S. Y., Yi, T. M. & Liu, L. (2021). A crack detection algorithm for concrete pavement based on attention mechanism and multi-features fusion. IEEE Transactions on Intelligent Transportation Systems. pp. 1-10. 10.1109/ TITS.2021.3106647.

[8] Yu, Y., Samali, B., Rashidi, M., Mohammadi, M., Nguyen, T. & Zhang, G. (2022). Vision-based concrete crack detection using a hybrid framework considering noise effect. Journal of Building Engineering. 61. 105246. 10.1016/j.jobe.2022.105246.

[9] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z. H., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. & Li, F. F. (2014). ImageNet large scale visual recognition challenge. International Journal of Computer Vision. 115. 10.1007/s11263-015-0816-y.

[10] Yue, Z. Y., Gao, F., Xiong, X., Wang, J., Huang, T., Yang, E. F. & Zhou, H. Y. (2021). A novel semi-supervised convolutional neural network method for synthetic aperture radar image recognition. Cognitive Computation. 13. 10.1007/s12559-019-0 9639-x.

[11] Wang, J. J., Zhong, Y. F., Zheng, Z., Ma, A. L. & Zhang, L. P. (2020). RSNet: The search for remote sensing deep neural networks in recognition tasks. 10.1109/TGRS.2020.3001401.

[12] Mujahid, A., Awan, M., Yasin, A., Mohammed, M., Damaševičius, R., Maskeliunas, R. & Hameed, K. (2021). Real-time hand gesture recognition based on deep learning YOLOv3 model. Applied Sciences. 11. 4164. 10.3390/app11094164.

[13] Wu, D., Zhang, C. J., Li, J., Rong, R., Wu, H. Y. & Xu, Y. M. (2021). Forest fire recognition based on feature extraction from multi-view images. Traitement du Signal. 38. 775-783. 10.18280/ts.380324.

[14] Özgenel, Ç. & Sorguc, A. (2018). Performance comparison of pretrained convolutional neural networks on crack detection in buildings. 10.22260/ISARC2018 /0094.

[15] Zhang, L. F., Yang, Y., Zhang, D. & Zhu, Y. J. (2016). Road crack detection using deep convolutional neural network. 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, pp. 3708-3712, doi: 10.1109/ICIP.2016.75330 52.

[16] Abdolrasol, M., Hussain, S., Ustun, T. S., Sarker, M., Hannan, M. A., Mohamed, R., Ali, J., Mekhilef, S. & Milad, A. (2021). Artificial neural networks based optimization techniques: A review. Electronics. 10. 2689. 10.3390/electronics102126 89.