Bridge Crack Identification Based on Feature Fusion of Convolutional Neural Networks

Qiuyue Wang, Baolin Li and Xiucai Nie

ABSTRACT

Aiming at the problem of low image quality of bridge cracks and poor identification of bridge cracks by single convolutional neural network method, this paper proposes a bridge crack identification algorithm based on feature fusion of convolutional neural networks. The algorithm first augments the image to obtain enough training samples to simulate the real scene. Then, using super-resolution technology to reconstruct the image size to enhance the image details; then, using AlexNet and VGG11 to construct and train the bridge crack feature fusion model. Finally, the SoftMax classifier is used to classify and identify the merged features to obtain the crack detection results. The experimental results show that the proposed algorithm is better than the single AlexNet and VGG11 feature extraction, and the image recognition accuracy is increased by 0.32% and 0.41% respectively. The loss value is smaller under the same iteration number, and the improved algorithm is better.

1. INTRODUCTION

With the continuous development of the economy and society, China's transportation industry plays an extremely important role in life. Bridges are an integral part of the transportation system and an important infrastructure for China's modernization. Under the influence of various factors, the service life and quality of bridges will have serious impacts, threatening the safety of people's lives and property. Therefore, bridge testing has become a topic of increasing concern.
In 2011, Guan et al. proposed the FoSA algorithm, which uses F* seed growth to obtain crack target points and then links to cracks. The method performs well on smooth road surface, and it is easy to miss and misdetect the crack image with complex background. In 2012, Q Zhou et al. proposed a method of seed point detection and tensor voting for crack detection, which opened up a new situation. This method achieves a high detection accuracy, but there is a disconnection when the crack is linked. In 2015, Guan et al. proposed the ITV-crack algorithm, which uses iterative tensor voting to detect cracks. Although this method can improve the detection accuracy, the calculation is complicated and the detection speed is not high.

Based on the above problems, this paper proposes a bridge crack identification algorithm based on convolutional neural network feature fusion. The algorithm first uses image augmentation technology to obtain training samples, then uses super-resolution technology to reconstruct images, and then uses two different convolutional neural network models. The crack characteristics of the bridge are merged, and finally the classification results are obtained by using the classifier to obtain the detection result.

2. RELATED TECHNOLOGY

Bridge crack identification is generally divided into two main links: feature extraction and feature classification. For feature extraction, there are currently two main methods: based on traditional methods and methods based on deep learning. Traditional feature extraction methods mainly include histogram of Oriented Gradient (HOG), Scale-invariant feature transform (SIFT), Local Binary Pattern (LBP), and affine scale. A single feature extraction method such as affine scale invariant feature transform (ASIFT) and difference of Gaussian (DOG), and a combination of various feature extraction methods developed later. With the continuous development of information technology, deep learning technology has also developed rapidly. Various methods for extracting image features using convolutional neural networks have emerged. At the same time, various models and combinations of various model combinations have emerged. In this paper, two deep convolutional networks, AlexNet and VGG11, are used to implement feature fusion to extract features. For classification algorithms, there are Support Vector Machine (SVM), linear projection based on sparse representation, minimum distance classification, SoftMax classification, random forest and Bayesian classification. In the past, a large number of classification algorithms based on deep learning have also appeared in the past, and this classification algorithm is related to a specific task, and the full-connection layer is added to classify after the feature extraction part of the convolutional neural network.

In this paper, we use two kinds of deep convolutional neural networks, namely AlexNet and VGG11, to classify bridge cracks based on feature fusion of
convolutional neural networks. The features extracted by AlexNet and VGG11 are merged and entered into the fully connected layer, the dropout layer, and the SoftMax classifier for identification.

The innovations and main work of this paper have the following points: 1). Using data augmentation to solve the classification model training problem of small batch data; 2). Using super-resolution algorithm to reconstruct the details of bridge crack image; 3). Building AlexNet and VGG11 The characteristics of the fusion network, and then use the SoftMax classifier to achieve image classification, improve recognition rate.

3. IMPLEMENTATION PROCESS

The framework of bridge cracks based on feature fusion of convolutional neural networks proposed in this paper is shown in Figure 1. It mainly consists of three parts: image preprocessing, feature extraction and fusion, and feature classification. After inputting the image of the bridge crack, firstly, the image is subjected to data augmentation, super-resolution and other pre-processing; then, the pre-processed bridge crack image is input into the feature extraction module, and after AlexNet and VGG11 extract the image features, Feature fusion is performed by series connection of feature vectors; finally, the merged feature is passed to the recognition module, and the recognition module uses the SoftMax classifier to identify the type of the image according to the input feature.
3.1 Image Preprocessing

In order to eliminate irrelevant information in the image and restore the real useful information, it is necessary to pre-process the image, enhance the detectability
of the information and minimize the data, thereby improving the reliability of feature extraction, image segmentation, matching and recognition.

In view of the different sizes of the collected images, if the unified normalization operation is directly performed, the picture will be excessively blurred, and a large number of mosaics or artifacts appear. Therefore, in order to improve the resolution of the image without degrading its quality, the image in the data set needs to be super-resolution preprocessed, and the image details are reconstructed while the image is enlarged.

In addition to this, for the training of the data set, it is augmented to obtain a sufficient set of training samples. Since super-resolution uses the a priori information of the image to enhance the image resolution, the augmentation operation of the image changes the a priori information of the image, and then the superization is performed. Therefore, the data of the image set is firstly analyzed. Augmentation operation, the difference of prior information will make the image resolution after super-resolution more complicated and complex, and then simulate various uncertain factors in the actual imaging and detection process, and realize the availability of the data set.

3.1.1 IMAGE AUGMENTATION

Although the current depth recognition-based image recognition algorithms can achieve very high accuracy, such as K-Nearest Neighbor (KNN) algorithm, SVM, BP neural network, Convolutional Neural Networks (CNN) and migration Learning, etc., but all are high-accuracy results obtained after training through a large number of data samples. When the training data samples are small, the effect of the deep learning based algorithm is significantly reduced. However, a small amount of sample learning based on data augmentation can solve this problem.

The training data set is augmented to compensate for the shortcomings of data shortage. There are many ways to increase the data, such as flipping, translation, rotation, cutting, scaling, Gaussian noise and Gaussian blur. This article mainly uses the following data augmentation methods, and the specific effect diagram of some data is shown in Figure 2.

1). Image rotation: Rotate the image 90° clockwise and 90° counterclockwise.

2). Image blur: Gaussian blur with standard deviation of 0.5 and 1.0 is performed on the original image.

3). Image noise addition: Gaussian noise with standard deviations of 0.01 and 0.02 is added to the original image.

4). Image noise and image blur: After the image is added with Gaussian noise with a standard deviation of 0.01, the Gaussian kernel with a standard deviation of 0.1 is used for blurring.

5). Image noise and image rotation: After adding Gaussian noise with a standard deviation of 0.1 to the image, rotate 90° clockwise.
6). Image blurring and image rotation: After the Gaussian blur with a standard deviation of 0.1 on the original image, rotate 90° counterclockwise.

![Image blurring and image rotation](image)

Figure 2. The specific effect diagram of some data.

### 3.1.2 SUPER RESOLUTION

Due to the different image sizes of the collected bridge cracks, especially when photographed at a long angle, the target area of the bridge crack in the image is too small. After detection and cutting, the resolution of the bridge crack image is too small, not conducive to the recognition of images. Therefore, it is necessary to perform super-resolution operation on the low-resolution image, and to enlarge the size of the image while making the edge details of the image clearer.

### 3.2 Feature extraction and fusion based on AlexNet and VGG11

This paper chooses AlexNet and VGG11 as the network for feature extraction and feature fusion. Compared with other network models such as GoogleNet, ResNet, DensNet, etc., AlexNet and VGG11 have the advantages of simple structure, few network layers and few parameters, which are suitable for the processing of bridge crack images. Since the recognition accuracy of any of these single networks needs to be improved, the feature fusion method is used to merge the extracted features as the input of the classifier, which can effectively improve the recognition accuracy of the algorithm. The following will be for AlexNet, VGGnet, VGG11 and feature fusion algorithms are introduced separately.

#### 3.2.1 ALEXNET

Alex's AlexNet network structure model proposed in 2012 detonated the application boom of neural networks and won the 2012 ImageNet Recognition Competition, making CNN a core algorithm model for image classification. AlexNet The model is divided into 8 layers, 5 convolution layers, and 3 fully connected layers. Although the number of layers is relatively small, it has more advantages
than later GoogleNet, VGGNet, etc.: On the one hand, AlexNet is more flexible and efficient than other complex networks. On the other hand, AlexNet is the first network to use multi-GPU parallel training technology, which allows good results to be achieved with acceptable training time in complex network environments, and model training is faster. AlexNet has successfully promoted the development of supervised deep learning and laid the foundation for deep learning applications in all areas of life.

3.2.2 VGGNET

VGGNet is a deep convolutional neural network jointly developed by the Computer Visual Group of Oxford University and DeepMind. It won the second place in the classification project and the first place in the positioning project in the 2014 ILSVRC competition. It explores the relationship between the depth of the convolutional neural network and its performance. By repeatedly stacking 3×3 small convolution kernels and 2×2 maximum pooling layers, six different network structures are successfully constructed. This kind of network structure can reduce parameters on the one hand and more nonlinear mapping on the other hand, which can increase the fitting ability of the network. When training a high-level network, you can train a low-level network first, and initialize the high-level network with the weight obtained by the former, which can accelerate the convergence of the network, and has the characteristics of strong scalability and generalization ability. Therefore, VGGNet is often used to extract features and initialize the migration learning model with given training parameters to improve the training speed of the network. In this paper, in order to ensure the complexity of the model, in order to minimize the complexity of the model, VGG11 (depth 11) is selected as a network of feature fusion from various depth structures.

3.2.3 FEATURE FUSION

With the rapid development of science and technology, the performance of contemporary computers has been greatly improved. Compared with traditional algorithms, deep learning algorithms have obvious advantages. Despite this, the accuracy of deep learning algorithms still needs to be improved, making the method of multi-network feature fusion widely used. In the classification task, the performance of feature fusion after recognition is more obvious than the single feature.

Feature fusion refers to the use of multiple feature extraction methods to extract multiple features of an image, and then vector fusion according to a specific method to form a new feature vector. The way the features are fused may be series, parallel, or weighted average. Feature fusion algorithms can be divided into three categories: 1. Algorithms based on Bayesian decision theory; 2. Algorithms based on sparse representation theory; 3. Algorithms based on deep learning theory. In this paper, the
algorithm based on deep learning theory is used to classify the features extracted by two convolutional neural networks.

In this paper, the features learned by two deep convolutional neural networks, AlexNet and VGG11, are concatenated in series, as shown in Figure 1. The merged features will be sent to the SoftMax classifier for identification as new features. The deep convolutional neural network normalizes the data, so that the obtained feature structures are similar, and the fusion can be performed relatively quickly. After the feature is merged, a plurality of representative features can be retained to retain the useful information of the image to the greatest extent possible.

The image recognition classification module consists of three fully connected layers, two dropout layers and one SoftMax function, with one dropout layer between every two fully connected layers. Due to the over-fitting phenomenon of large-scale convolutional neural networks, the purpose of introducing the dropout layer is to eliminate the joint adaptability of adjacent neurons, thereby improving the generalization ability of the network model. The fully connected layer connected with the SoftMax classifier is used to solve the multi-classification problem of bridge crack identification. After identification, different neurons are mapped to different bridge crack images, and the probability of such images can be obtained.

Softmax classifiers can be generalized into multi-category scenarios, although other classifiers can also implement classifications, such as class determination in SVM classifiers, but the size order of the resulting class determination results only represents the sorting of categories does not have substantial physical meaning. In the SoftMax classifier, the size of the decision result value indicates the probability that the image belongs to the category, so that the implementation of the classification result is further advanced.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Experimental environment

The processor is Intel(R) Core(TM) i5-7500 CPU @ 3.40GHZ, the memory capacity is 8.0GB, and the graphics card is NVIDIA GeForce GTX 1020 2GB. The algorithm is based on the TensorFlow open source library and the Keras framework. All the data were from self-acquired images, of which 7315 bridge crack images were randomly selected as the training sample set and 2365 sheets were used as the test sample set.

The specific four types of cracks are mainly reticular cracks, longitudinal cracks, transverse cracks, and oblique cracks. And these four categories correspond to 0 to 3 category labels respectively.
4.2 Experimental parameters

For the convolution kernel size, the number of connected neurons, the weight initialization method, the activation function, the optimizer and the classifier, etc., the comparison experiment was carried out to determine the convolution kernel size of 5×5, fully connected convolutional nerve. The number of yuan is 256. The optimized Adam gradient descent algorithm is used to train, so that it can design independent adaptive learning rate for different parameters, so as to obtain faster convergence speed. At the same time, cross entropy which can control the output error learning rate is used as the loss function. The combination of the two can avoid the problem that the mean square error as the loss function leads to a gradual decrease in the learning rate. The initial learning rate is 0.001, the Dropout parameter is determined to be 0.5, and the SoftMax classifier is used to classify the fused features.

The experimental results show that the optimized Adam algorithm uses the longest time in each iteration, but the network model trained by this method has the least loss and the highest recognition accuracy, so it is used as the optimization method.

4.3 Experimental results

4.3.1 ANALYSIS OF DIFFERENT ALGORITHM RECOGNITION PERFORMANCE

As can be seen from TABLE I, since the algorithm uses the feature fusion method, the number of parameters is 23% and 33% more than that of AlexNet and VGG11, respectively, and the time of each iteration during training is 1s more, although the training time of the algorithm is It is much larger than the single AlexNet and VGG11 algorithms, but the time spent processing a similar number of images is almost the same, indicating that the generalization performance of the network model is better and the ability to express the cracks of the bridge is stronger. The recognition accuracy of this algorithm is 98.77% higher than that of AlexNet and VGG11, respectively, which is 0.32% and 0.41% higher, which indicates that the proposed algorithm has certain advantages in bridge crack identification.

| method          | Number of parameters / million | Iteration time / (s/time-1) | Loss value | Training time | Recognition accuracy / % |
|-----------------|--------------------------------|-----------------------------|------------|---------------|--------------------------|
| AlexNet         | 0.53                           | 8.15                        | 0.0921     | 734           | 98.45                    |
| VGG11           | 0.49                           | 8.03                        | 0.0945     | 682           | 98.36                    |
| This paper Algorithm | 0.65                         | 9.18                        | 0.0667     | 4122          | 98.77                    |
4.3.2 ANALYSIS OF RECOGNITION RATE OF DIFFERENT ALGORITHMS

As shown in Figure 2, the algorithm in this paper compares the black and white and color recognition rates of the input image, and the black and white recognition rate is 0.04% higher. When the black and white image is input, the recognition rate of the algorithm is 0.04 higher than AlexNet and VGG11 respectively, % and 0.16%; when the color image is input, all recognition rates are decreased, indicating that it is still to be studied and studied, but the recognition rate of the algorithm is 0.03% and 0.17% higher than AlexNet and VGG11 respectively. It can be seen that when different color images are input, the algorithm has higher recognition rate than the two single feature extraction methods.

TABLE II. COMPARISON OF THE RECOGNITION RATE BETWEEN THE ALGORITHM AND TWO SINGLE ALGORITHMS.

| Input               | method           | Recognition rate /% |
|---------------------|------------------|---------------------|
| Black and white image | AlexNet          | 98.89               |
| Black and white image | VGG11            | 98.87               |
| Black and white image | This paper       | 98.93               |
| Color image         | AlexNet          | 98.86               |
| Color image         | VGG11            | 98.82               |
| Color image         | This paper       | 98.89               |

The experimental results show that the images are preprocessed by using image augmentation and super-resolution, and then the two network models of AlexNet and VGG11 are used for feature fusion. Finally, the SoftMax classifier is used to identify and classify the merged features, which improves the training speed. The problem of slowness and low recognition accuracy can be based on the hardware conditions of common configuration, and it takes a reasonable training time to obtain a better recognition accuracy rate.

5. CONCLUSIONS

Based on Based on the deep convolutional neural network, this paper proposes a feature fusion algorithm to realize bridge crack detection. The algorithm combines the features extracted by AlexNet and VGG11 network training models in series,
and then uses Softmax classifier to identify and classify. The experimental results show that compared with the single AlexNet and VGG11, the network achieved by this algorithm has faster training speed and higher recognition accuracy. In addition to the bridge detection algorithm, the algorithm can also be applied to antiquities identification and leakage identification.

REFERENCES

1. Yuanpan Zheng, Guangyang Li, Wei Li. 2019. “Review of the Application of Deep Learning in Image Recognition,” Computer Engineering and Applications, 55(12):20-36.
2. Wenchi Zhang, Lihui Chen, Wei Wu, Xiaomin Yang, Binyu Yan. 2019. “Traffic sign recognition based on feature fusion of convolutional neural networks,” Computer Applications, 39(1):21-25.
3. Baihua Jiang, Ya Zhang, Wenwen Zeng. 2019. “Image recognition of rail damage based on improved convolutional neural network,” Measurement & Control Technology, 38(06):19-22+27.
4. Kezheng Lin, Yuxuan Bai, Yutian Li, Wei Li. 2019. “Small sample expression recognition based on different models under deep learning,” Computer Science and Exploration, 1-13.
5. Yuxuan Li, Xiaoxia Li, Xueyuan Wang, Wei Sun. 2019. “Real-time facial expression recognition based on multi-scale kernel feature convolutional neural network,” Computer Application: 1-9.
6. Liangfu Li, Weifei Ma, Li Li, Wei Lu. 2019. “Research on bridge crack detection algorithm based on deep learning,” Acta Automatica Sinica: 1-16.
7. Na Li, Yuanfei Xu, Taotao Jia. 2019. “Fast identification method for bridge cracks with second-order moment & gray-scale difference,” Computer Applications and Software, 36(05):216-219+230.
8. Weijun Pan, Yingjie Duan, Qiang Zhang, Zhengyuan Wu, Yuchen Liu. 2019. “Research on the Recognition of Laser Radar Aircraft Wings Based on AlexNet Convolutional Neural Network,” Opto-Electronic Engineering, 46(07):121-128.
9. Qingsong Song, Chao Zhang, Zhengxin Tian, Wei Chen, Xingli Wang. 2018. “Traffic Sign Recognition Based on Multiscale Convolutional Neural Networks,” Journal of Hunan University (Natural Science), 45(08):131-137.
10. Iglovikov V, Shvets A. 2018. “TernausNet: U-Net with VGG11 encoder pre-trained on ImageNet for image segmentation,” arXiv Preprint, arXiv. 1801. 05746.
11. Xuesong Chai, Xingyong Zhu, Jianchao Li, Feng Xue, Xueshi Xin. 2018. “A tunnel lining crack identification algorithm based on deep convolutional neural network,” Railway Construction, 58(06):60-65.
12. Jingjing Zhang, Hongyu Nie, Qiang Yu. 2017. “Bridge crack detection based on multi-scale input image infiltration model,” Computer Engineering, 43(02):273-279.
13. Z W Yuan J, Zhang. 2016. “Feature extraction and image retrieval based on AlexNet,” Proceedings of the Eighth International Conference on Digital Image Processing, Bellingham: SPIE, 100330E.
14. A. Krizhevsky, I. Sutskever, and G. Hinton. 2012. “ImageNet Classification with Deep Convolutional Neural Networks,” Advances in neural information processing systems, 1097–1105.