International Journal of Advanced Computer Technology (IJACT)
ISSN: 2319-7900 Volume VII, Issue II, April, 2018

DETECTION METHOD OF HIGHWAY CRACK BASED ON DEEP CONVOLUTION NEURAL NETWORK

Gang zhan Liu[1], Weili[2], WeinaYue[3], [3]Zhengzhou University of Light Industry 471023, China;
[1,2]Information Engineering College, Henan University of Science and Technology, Luoyang, China;

Abstract

In order to further improve the accuracy of the crack image feature detection accuracy, a neural network model of expansion convolution based on deep learning method is proposed: the traditional convolution module is replaced by the expansion convolution module in the convolution layer the largest aggregation and mean aggregation modules are replaced by the threshold mean module in the converging layer, then classify the collected features with softmax classifier. Experimental results show that the improved network model compared to the traditional convolution neural network model the best improvement rate of convergence performance is 18.36% and the maximum decrement rate of recognition error is 26.74%, so improved neural network model of expansion convolution has better generalization ability and robustness.

Keywords: - *Deep learning Convolution neural network Expansion convolution Threshold mean Dimension reduction*

1. Introduction

With the increasing progress of the rapid development of economy and society, more and more high frequency of use of highway, there have been serious problems of highway crack, how check and repair accurately the road cracks has become a prominent problem, Li Wei et al designed the improved BP neural network algorithm based on composite error function[1], which improves the crack image detection accuracy and speed; They also designed the improved algorithm of [2] adaptive region merging based on watershed, to solve the over segmentation problem; For the problem of inaccurate about edge cracks in the asphalt pavement, they also designed the detection algorithm of image based on morphological multi-scale fracture[3], which can not only set an adaptive weights, but also make the operator to detect various types of edges and effectively restrain the noise, to realize the accurate extraction of various types of edges. With the rapid development of artificial intelligence, increasing the number of fracture characteristics, to extract image features using the neural network method has been widely used, but the shallow neural network learning has the following disadvantages: the convergence speed slowed because of the " saw tooth phenomenon" [4];the generalization ability [5] and prediction rate decreased for selecting samples artificially , therefore, this method has been unable to meet the demand, in view of the above problems, this paper will introduce deep learning method to crack image detection, proposing a convolution neural network model which by convolution and aggregation alternate ways may extract the feature information needed, and repeatedly train through many layers to accurately classify characteristics.

2. CNN network structure

Convolution neural network is a multilayer artificial neural network to deal specifically with the characteristic of two-dimensional images, each layer is composed of many feature maps, each feature map has a plurality of mutually isolated neurons there is a deep connection between the adjacent neurons in the adjacent layers, and the adjacent neurons on the same feature map are not connected to [6]. There are three types of layers in the structure of convolution neural network: convolution layer, aggregation layer and full connection layer, all layers arranged in accordance with the order of the structure, each layer can be extracted from 2D image features, using the differential function to activate the data from one layer to another layer, and using the back-propagation algorithm to optimize the weights of network parameters to improve the classification effect.

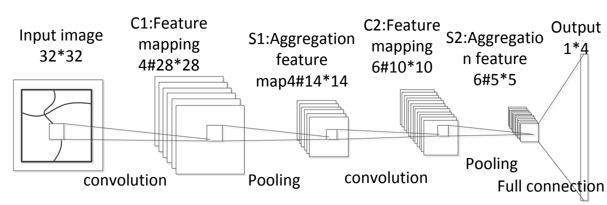

Fig.1 CNN network structure of fracture recognition

As shown in figure 1, the network consists of five layers, including two convolution layers (C1 and C3), two aggregation layers (S2 and S4), and one full connection layer. Input of the original image is the size of 32 × 32 fractures, C1 layer has 4 filters, convolution kernels is 5 ×5, boundary fills 0 element, step length is 1, 4 28×28 feature graphs is obtained through convolution operation. In s2 layer there are aggregation and sampling operation, convolution kernels is 2×2, step length is 2, boundary fills 0 element, four 14×14 feature graphs are gained, halve the redundant information, a large number of parameters are also shrinking, does not affect the image of effective information, the C3 layer increased to 6 convolution kernel, going on executing convolution operation for characteristics after polymerization process， similar to the previous C1 layer, to obtain Characteristics of different cracks. At the S4 layer, feature aggregation is used to remove useless feature information, and 6 neurons are formed at the full connection layer. The function score and feature classification are achieved by combining softmax classifier at output layer.

3 Improved expansion convolution structures

In order to collect large scale features in a few layers, in the convolution neural network model a new expansion convolution structure is designed, replacing the original convolution process, the structure can promote receptive field in linear growth without losing the image's important information, Its core is the operator of expansion convolution, in the convolution layer, when the original convolution kernel and original image execute convolution, between the all original row elements and column elements about the original convolution kernel are filled with 0 elements, the number of rows and columns filling 0 elements is called the filling coefficient, if 1 rows 0 and 1 column 0 are filled , (filling coefficient is 1) filled convolution kernel is called one-dimensional convolution kernel of the expansion coefficient; if 2 rows 0 and 2 column 0 are filled , it is called two-dimensional convolution kernel of the expansion coefficient, at the time of convolution, The receptive field is expanding, operator is a filter through the expansion way to modify the parameters and realize convolution in different regions of the image.

As shown below :

$W1_{x1}$	$w2_{x2}$	$w3_{x3}$
$w4_{x4}$	$w5_{x5}$	$w6_{x6}$
$w7_{x7}$	$w8_{x8}$	$W9_{x9}$

Fig 2. Original convolution kernel

$w1_{x1}$	0_{x2}	$w2_{x3}$	0_{x4}	$w3_{x5}$
0_{x6}	0_{x7}	0_{x8}	0_{x9}	0_{x10}
$w4_{x11}$	0_{x12}	$w5_{x13}$	0_{x14}	$w6_{x15}$
0_{x16}	0_{x17}	0_{x18}	0_{x19}	0_{x20}
$w7_{x21}$	0_{x22}	$w8_{x23}$	0_{x24}	$w9_{x25}$

Fig 3. one-dimensional convolution kernel of the expansion coefficient

Assume that figure 2 is the original convolution kernel, which is 3 x 3 filters, corresponding image region is also 3 x 3 receptive fields, where Wi is the filter value, Xi is the value of the characteristic graph data corresponding to the filter. figure 3 is the convolution kernel expanded ,whose size is 5 ×5, that filter is called the 1 dimension convolution kernel of expansion coefficient by the original convolution kernel executing expansion, expansion elements are all 0.

$w1_{x1}$	0_{x2}	0_{x3}	$w2_{x4}$	0_{x5}	0_{x6}	$w3_{x7}$
0_{x8}	0_{x9}	0_{x10}	0_{x11}	0_{x12}	0_{x13}	0_{x14}
0_{x15}	0_{x16}	0_{x17}	0_{x18}	0_{x19}	0_{x20}	0_{x21}
$w4_{x22}$	0_{x23}	0_{x24}	$w5_{x25}$	0_{x25}	0_{x26}	$w6_{x28}$
0_{x29}	0_{x30}	0_{x31}	0_{x32}	0_{x33}	0_{x34}	0_{x35}
0_{x36}	0_{x37}	0_{x38}	0_{x39}	0_{x40}	0_{x41}	0_{x42}
$w7_{x43}$	0_{x44}	0_{x45}	$w8_{x46}$	0_{x47}	0_{x48}	$w9_{x49}$

Fig 4. two-dimensional convolution kernel of the expansion coefficient

Figure 4 is a 7×7 filter which is derived from an original filter by expanding the 2 times expansion coefficient and also known as a 2 dimensional expansion coefficient convolution kernel. The expansion and convolution process is as follows:

(1) the construction of the expansion convolution core. When the first convolution layer is executed convolution operation, an expansive convolution filter is constructed, which makes the original convolution kernel filled with a row and a column of 0 elements between all row elements and the column elements, and becomes a 5× 5 filter, Since the number of rows and columns of 0 elements are 1, it is also called the 1 dimensional coefficient expansion convolution, that is, the expansion coefficient is 1, and the region of the image receptive field is also 5 × 5. The convolution kernel corresponding to second convolution layer is obtained by increasing 2 times expansion coefficient from the original convolution kernel [7]. Filling 0 elements of 2 rows and 2 columns can get 7 x 7 filter, that filter is also known as 2 dimensional convolution kernel of expansion coefficient, and its corresponding receptive field area is 7 × 7. It is assumed that the size of the original convolution kernel is W, the expansion coefficient of the convolution core is K in the long and wide direction, and the size of convolution kernel after expansion is S, then S=W x K+W-K. (2) convolution operation. Suppose that figure 2 is an original convolution kernel with a bias of B1, the area data of the corresponding receptive field is x1, X2, X3... X9, the step length is 1. If the original image is directly implemented convolution operation, the first value of neuron function is obtained after convolution operation:

f(x)=x1×w1+x2×w2+x3×w3+x4×w4+x5×w5+x6×w6+x7×w7+x8×w8+b0. If the expansion operation is executed [8],

ISSN: 2319-7900 Volume VII, Issue II, April, 2018

after the 1 dimension coefficient expansion convolution, the corresponding receptive field area of each neuron on the same feature map will increase, and there will be 1 gap.

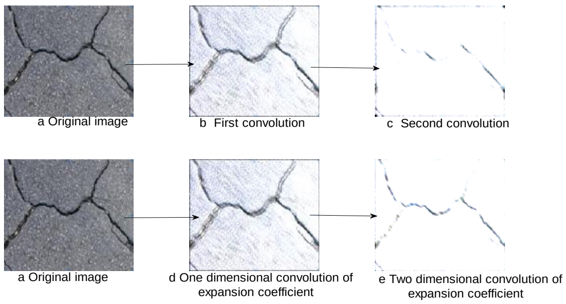

Fig 5 Process contrast diagram between traditional convolution and expansion convolution

As shown in Figure 3, the first neuron function value is f(x)=x1×w1+x3×w2+x5×w3+x11×w4+x13×w5+x15×w6+x21×w7+x23×w8+x25×w9+b1. After the two dimensional coefficient expansion convolution, as shown in Figure 4, the first neuron function value is f(x)=x1×w1+x4×w2+x7×w3+x22×w4+x25×w5+x28×w6+x43×w7+x46×w8+x49×w9+b2. Figure 5 is the result of the experiment given above. The experiment is divided into two parts: the first part is the two convolution test of the original convolution kernel a, b to c. the second part is the two expansion convolution experiment of a, d and e. Both of them used 3 * 3 convolution kernel for the first time. The characteristic graph e extracted after the convolution kernel expansion is clearer than the former figure C, and the texture information is more abundant. The former uses less convolution parameters in the extraction of feature maps, but the effect of extracting features at the same time in the network scale is poor. The latter can not only converge the large scale features of the original image faster, simplify the construction of the model, but also extract the features faster, and the accuracy of the extracted features is higher.

4. Improved aggregation models

A. Traditional aggregation model

In order to gather the effective information in the feature graph, a large number of redundant features in the feature map should be removed, and the method of aggregation operation is usually adopted. The common polymerization methods have the maximum aggregation and the mean aggregation [9]. The maximum aggregation is to take out the maximum value pixel value in the aggregated domain as the eigenvalue of the region during the polymerization process. The average aggregation is the eigenvalue of the average value of all pixel values in the aggregated domain. The size of the input feature map is P, the size of the aggregated domain is a x a, the step length is s, the bias is B, and the output feature graph is Q. The size of the output feature graph Q can be obtained by non - overlapping polymerization: Q= (P-a) ⁄s+1, The values of the neuron after the average polymerization and the maximum value polymerization are also obtained. The formula is as follows:

Average polymerization:

$$Q_{mn} = \frac{\sum_{m=1}^{a}\sum_{n=1}^{a} P_{mn}}{a^{2}} + b \quad (1)$$

Maximum aggregation:

$$Q_{mn} = \max_{m=1,n=1}^{a} p_{mn} + b \quad (2)$$

B. Improved aggregation model

If the maximum value aggregation is applied in the extraction of the polymer layer, it will reduce the error of the convolution layer to estimate the mean shift and preserve the texture characteristics of the feature map better. By means of mean aggregation, the variance of the estimated value brought by the confinement of the domain will be reduced, and the background features of the feature map are extracted better [10-12]. A new model is proposed in view of the above shortcomings: the threshold value mean aggregation model. The model can reduce the influence of the former model on image feature details, remove redundant features, reduce computation parameters, and keep image translation, expansion and rotation invariance.

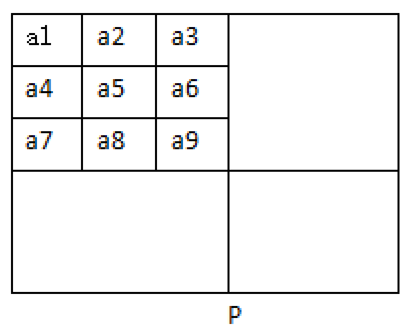

Fig 6 Convolution feature graph

The method is further optimized on the basis of the previous two, and the following is the following: First set the initial threshold to find out the average of all the elements in the aggregation domain, which usually takes the mean as a threshold more scientific. And then select all the features larger than the threshold in the aggregated domain. Finally, the average of its sum is obtained. The mean value is the representative feature of the reduced dimension sampling [13-15] in the aggregated domain.

ISSN: 2319-7900 Volume VII, Issue II, April, 2018

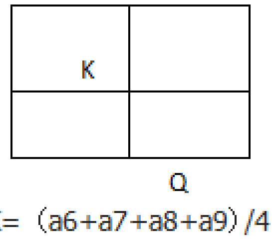

Fig 7 Aggregated feature graph

As shown in figures 6 and Figure 7, First, find the mean value I of all the elements ：I=a1+a2+a3+a4+a5+a6+a7+a8+a9/9, and make the result as a threshold, again, all the features in the aggregation larger than the threshold are screened, and get the mean of its sum. Assuming a5<I<a6<a7<a8<a9, the feature value of the ideal aggregation is K= （a6+a7+a8+a9)/4 after the ideal aggregation., this method not only has clear effect and richer content, but also avoids the influence of details weakening caused by maximum pooling and mean pooling. In addition, for the convolution network model, the prediction accuracy is better and the performance is more stable. The threshold algorithm is as follows:

1) set the threshold, usually using the average value of the element of the aggregation domain as the threshold, and the formula is as follows:

$$V_{mn} = \sum_{m=1}^{a}\sum_{n=1}^{a} P_{mn} \qquad (3)$$

2) the number of all elements larger than the mean is selected and the formula is as follows:

$$T = \underset{m=1,n=1}{\overset{a}{Count}}(P_{mn} > V_{mn}) \qquad (4)$$

3) the sum of all the elements larger than the mean in the aggregation region, the formula is as follows:

$$Y = \sum_{m=1}^{a}\sum_{n=1}^{a}(P_{mn} > V_{mn}) \qquad (5)$$

4) the mean value is calculated and used as the value of the aggregated posterior element, and the formula is as follows:

$$Q_{mn} = {}^{Y}\!/_{T} + b \qquad (6)$$

5. Experimental results and analysis

Experimental equipment Intel core i5-2400 cpu@3.10GHZ CPU, Kingston 8G memory，64 bit windows10 operating system，Matlab2014a software. A total of 3000 cracks are collected to make a data set, 2400 pictures are made into a training set, and 600 are made into a test set. And the collected color images are preprocessed, that is, gray scale transformation. Some samples are shown in Figure 8. Input 32 * 32 size picture, train the network model, record the number of iterations, periodic consumption time and error recognition rate, and draw the mean square error curve.

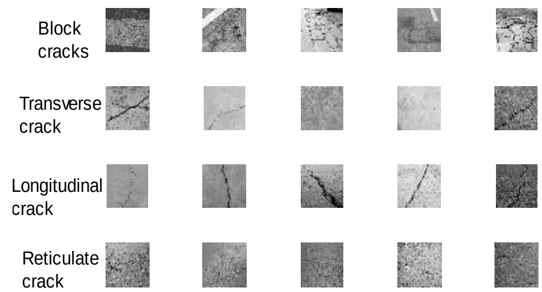

Fig 8 Partial crack image sample

A. Mean aggregation model of convolution network

The traditional convolution neural network of the average aggregation model was iterated for 4 times. The changes of training batches, mean square error, and cycle consumption time and error recognition rate were recorded at each iteration. The mean square error curve is shown in Fig .9.

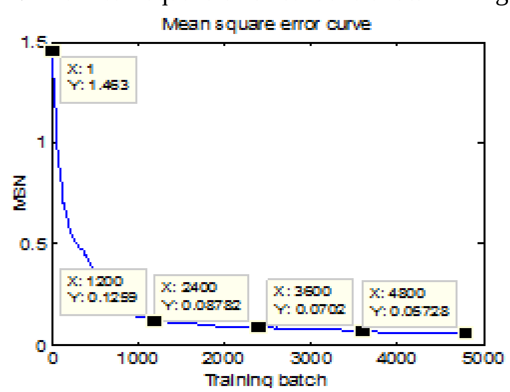

Fig 9 Mean square error curve of average polymerization model

In Figure 9，From (1,1.463) to (1200,0.1259) is a variation of the mean square error curve in the first iteration of the first period, from (1200,0.1259) to (2400,0.08782) as the variation curve of the second iteration of the second cycle, from (2400,0.08782) to (3600,0.0702) as the variation curve of the third iteration of third cycles, from (3600,0.0702) to (4800,0.05728) as the curve of the fourth iteration of fourth cycles, For the first iteration, the downward trend of the curve is obvious, the speed of convergence is faster, and the latter three stages tend to be stable. As shown in Table 1, the number of iterations is proportional to the period consumption time, and is inversely proportional to the error recognition rate. With the increase of the number of iterations, the time difference of the time consumed in a single iteration is less than that of the single iteration. The total time consuming time of the number of iterations is increasing, and the error recognition rate is decreasing, and the recognition rate of the image prediction is improved.

Table 1 Average aggregation model parameters of traditional convolution neural network

Number of iterations	First cycle /s	Second cycle /s	Third cycle /s	Fourth cycle /s	Error Recognition rate /%
1	210.24				11.53
2	228.35	226.07			7.91
3	209.29	209.35	209.22		6.29
4	209.14	211.19	219.14	218.87	5.35

B. Maximum aggregated model of convolution network

According to the number of iterations, the changes of mean square error, cycle and error recognition rate of the maximum aggregation model convolution network are recorded, as shown in Figure 10.

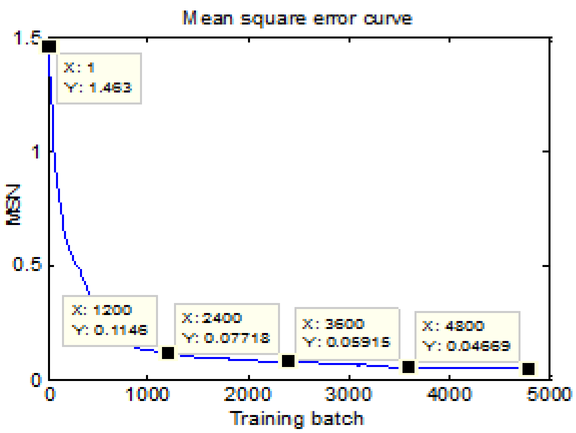

Fig 10 Maximum mean square error curve

Table 2 Parameters of the maximum aggregation model of the traditional convolution network

Number of iterations	First cycle /s	Second cycle /s	Third cycle /s	Fourth cycle /s	Error Recognition rate /%
1	228.07				10.35
2	215.80	229.31			6.98
3	231.68	232.53	233.65		5.21
4	229.29	231.62	231.63	229.16	4.23

As shown in Figure 10, from (1, 1.463) to (1200, 0.1146) is the change curve of the first iteration of the first iteration, The curve is a little steeper than the previous average reel neural network model. From (1200, 0.1146) to (2400, 0.07718) is the change curve of the second iteration in the second cycle, the curve tends to a steady trend. From (2400, 0.07718) to (3600, 0.05915) is a graph of the third cycles of third iterations. From (3600, 0.05915) to (4800, 0.04669) is a graph of the fourth cycles of fourth iterations, among them, y dropped from 1.463 to 0.04669. Compared with the average aggregation model, the first two interval curves are steeper, the decrease is more obvious, and the latter two is parallel to the X axis. As shown in Table 2, The experiment carried out 4 iterations for the maximum aggregation model, which consumed in turn from 228.07s to 921.7s, and the total consumption time increased progressively. The sum of accumulated iteration time exceeded the average aggregation model, and the error recognition rate also decreased from 10.35% to 4.23%. The recognition rate of each iteration is lower than that of the former, and it also reflects the prediction accuracy of the model, but at the expense of the time of consumption.

C. Threshold aggregation model

The improved expansion convolution network model is shown by the mean square error curve after 4 iterations, as shown in Figure 11. The cycle consumption time and error recognition rate, as shown in Table 3, are similar to those of the previous two models. In Figure 3, from (1,1.463) to (1200,0.1099) is the change curve of the first period of the first iteration From (1,1463) to (2400,0.07354) is the total change curve of the two cycles of second iterations From (1,1463) to (3600,0.05565) is the total change curve in the three cycles of third iterations From (1,1.463) to (4800,0.04366) is the total change curve of the four cycles of fourth iterations, X increased from 1 to 4800, showing a trend of increase. Y decreased from 1.463 to 0.04366, reflecting the convergence speed of the model. Compared with the former two models, the decline is more obvious and the convergence is faster. As shown in Table 3, compared with the first two models, the improved expansion convolution network model is proportional to the number of iterations and the total consumption time, which is inversely proportional to the error recognition rate. Compared with the maximum aggregation model, the total cycle consumption time is reduced in the case of higher prediction accuracy. After 4 iterations, the accuracy is up to 3.92%, which consumes only 909.87s, while the former consumes 921.7s to reach 4.23%. The latter is based on saving time to reduce the false recognition rate and improve the accuracy of model identification. In a word, the error recognition rate tested by the model is lower than the first two, and it is more accurate for the model. The threshold aggregation model and the first two models are compared in the rate of convergence and error recognition. With the increase of the number of iterations, the mean square error ratio presents a downward trend, the convergence performance is improved and the rate of convergence can be up to 18.36%. The rate of error recognition also showed a downward trend with a decreasing rate of 26.73%. Therefore, the conclusion is that the average aggregation model is the worst, while the threshold value model predicts better.

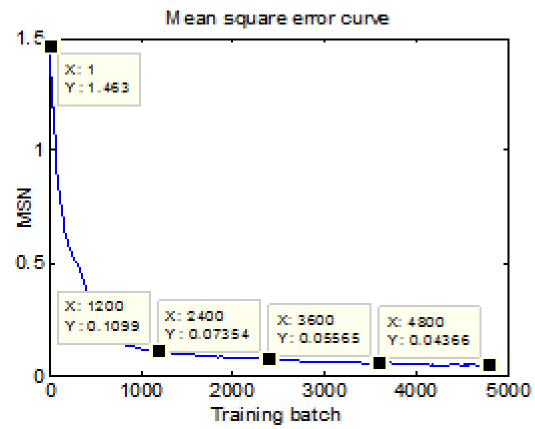

Fig 11 Threshold mean square error curve

Table 3 improved threshold model parameters of convolution neural network

Number of iterations	First cycle /s	Second cycle /s	Third cycle /s	Fourth cycle /s	Error Recognition rate /%
1	231.29				9.90
2	235.32	229.63			6.64
3	211.84	209.42	208.62		4.93
4	224.55	231.12	225.21	228.99	3.92

Table 4 comparison diagram with the performance parameters of the first two models

Comparison with average aggregation model			Comparison with the maximum aggregation model	
N Number of iterations	Percentage of speed of convergence	Reduction ratio of recognition rate	Percentage of speed of convergence	Reduction ratio of recognition rate
1	12.71	14.13	4.10	4.34
2	16.29	16.06	4.67	4.87
3	15.67	21.62	5.91	5.37
4	18.36	26.73	6.42	7.33

6. Conclusion

By analyzing the convolution process and polymerization process of convolution neural network, comparing the optimization degree of different processes in studying the characteristics of cracks and verifying the effect of feature extraction process on the number of iterations, convergence and prediction rate an improved model of deep convolution neural network is proposed, which is an expansion convolution neural network. Through experimental comparison, it is found that the fracture characteristics of the expansion convolution neural network in the convolution process are more abundant and clear. When the feature prediction is classified, the speed of the model convergence and the accuracy of the prediction are greatly improved, and the complexity of the network computing is reduced. Therefore, the model network is more suitable for accurate prediction of the image characteristics of road cracks.

Reference

[1]. Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolution neural networks J]. Advances in Neural Information Processing Systems, 2012, 25 (2): 10974105.

[2]. Grivas D A, Bhagvati C, Skolnick M M, et al. Feasibility of automating pavement distress assessment using mathematical morphology [J]. Transportation Research Record, 2004, (1435):52-58.

[3]. CHENG H D, RONG Jim. Novel approach to pavement cracking detection based on fuzzy set theory [J].Journal of Computing In Civil Engineering, 1999, 13(4):270-280.

[4]. Katakam N. Pavement Crack Detection System Through Localized Thresholding [D]. The University of Toledo, 2009.

[5]. Kalchbrenner N, Espeholt L, Simonyan K, et al. Neural Machine Translation in Linear Time [J]. 2017:16-21

[6]. Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions [J]. 2016(1):45-63

[7]. Cheng H D, Miyojim M. Automatic pavement distress detection System [J]. Information Sciences, 1998, 108(1/4):219-240.

[8]. Peggy Subirats, Jean Dumoulin, VineentLegeay. Automation of pavement surface crack detection using the continuous wavelet transform [J]. IEEE, 2009, 127:3037-3040.

[9]. Lee B J. Development of an integrated digital pavement imaging and neural network system [D]. USA: Faculty of the University of Iowa, 2001.

[10]. J, Canny. A Computational Approach to Edge Detection [J]. Pattern Analysis & Machine Intelligence IEEE Transactions on, 1986, pami-8(6):679-698.

[11]. Gary Bradski, Adrian Kaehler. 学习OpenCV（中文版） [M]. 北京：清华大学出版社, 2009.

[12]. GengHao, Luo Min, Hu Feng. Improved self-adaptive edge detection method based on canny[C]//Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China, 26-27 Aug. 2013:527 - 530.

Biographies

Gangzhan Liu, born in 1985, is a graduate student in the school of information engineering, Henan University of science and technology. The email address is 420616473@qq.com

Weili, born in 1960, is an associate professor in the school of information engineering, Henan University of science and technology.