

PAPER • OPEN ACCESS

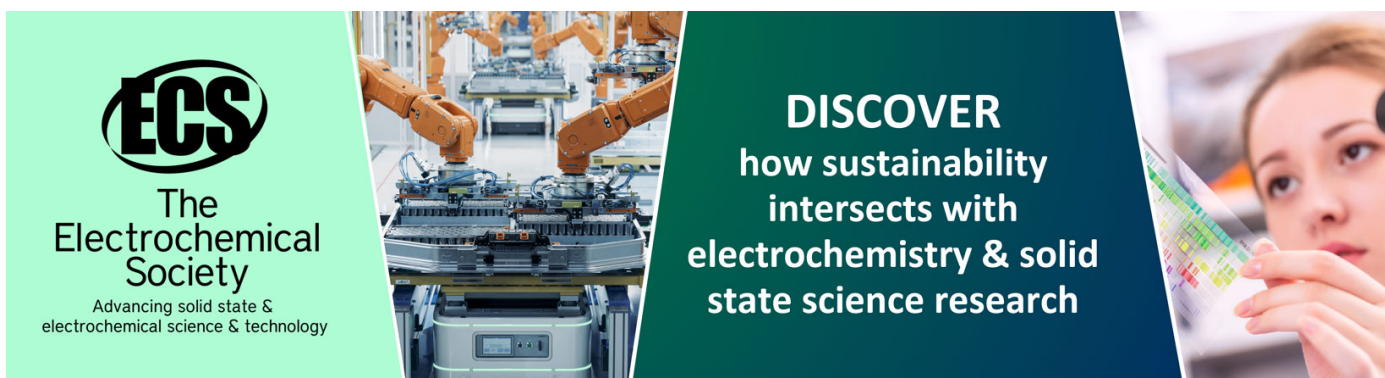
Automatic Crack Detection for Concrete Infrastructures Using Image Processing and Deep Learning

To cite this article: Cuong Nguyen Kim *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **829** 012027

View the [article online](#) for updates and enhancements.

You may also like

- [Lightweight pixel-wise segmentation for efficient concrete crack detection using hierarchical convolutional neural network](#)
Jin Kim, Seungbo Shim, Yohan Cha et al.
- [Development and field application of a nonlinear ultrasonic modulation technique for fatigue crack detection without reference data from an intact condition](#)
Hyung Jin Lim, Yongtak Kim, Gunhee Koo et al.
- [A Gabor filter based weak crack detection method for glassivation passivation parts wafer](#)
Sizhe Lang, Fei Zheng, Panyu Li et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Automatic Crack Detection for Concrete Infrastructures Using Image Processing and Deep Learning

Cuong Nguyen Kim^{1a}, Kei Kawamura^{2b}, Hideaki Nakamura^{3c}, and Amir Tarighat^{4d}

¹Faculty of highway& bridge, Mien Trung of civil engineering, Vietnam

²Graduate School of Science & Technology for Innovation, Yamaguchi University, Japan

³Graduate School of Science & Technology for Innovation, Yamaguchi University, Japan

⁴Dept. of Civil Engineering, Shahid Rajae Teacher Training University, Iran

^a nguyenkimcuong@muce.edu.vn, ^b kay@yamaguchi-u.ac.jp,

^c Nakahide@yamaguchi-u.ac.jp, ^d tarighat@srttu.edu

Abstract. Automatic crack detection is a main task in a crack map generation of the existing concrete infrastructure inspection. This paper presents an automatic crack detection and classification method based on genetic algorithm (GA) to optimize the parameters of image processing techniques (IPTs). The crack detection results of concrete infrastructure surface images under various complex photometric conditions still remain noise pixels. Next, a deep convolution neural network (CNN) method is applied to classify crack candidates and non-crack candidates automatically. Moreover, the proposed method is compared with the state-of-the-art methods for crack detection. The experimental results validate the reasonable accuracy in practical application.

1. Introduction

Many concrete components of existing infrastructure systems such as bridges, and tunnels have suffered from various geologic, loading and environmental conditions cause to cracks which make influent to quality of operations. Therefore, the condition assessment of the existing infrastructures is an important task not only for warning against deterioration but also for guaranteeing soon maintenance. Concrete cracks are important indicators reflecting the safety of infrastructure. The automatic crack detection based on image data has been considered significantly due to accuracy, objectivity and timing inspection. This technique can be implemented using some of different image data captured from ultrasonic device, infrared and thermal device, laser scanning, and commonly digital cameras.

Image processing techniques consist of three approaches: edge detection, threshold technique (Fujita et al.2006) and mathematical morphology (Nguyen et al 2016). Machine learning algorithms (MLAs) are commonly used to decide the parameter value of IPTs. Therefore, they are applied to detect and classify concrete infrastructure surface cracks. In recently years, many automatic crack detection and classification methods based on a combination of IPTs and MLAs are implemented as decision tree (DT) (Kei et al.2013), support vector machine (SVM), k-clustering nearest neighbour (K-NN), and artificial neural network (ANN) (Zhang et al.2014).



Moreover, the parameter-optimization algorithms of IPTs such as genetic algorithm (GA), particle swarm optimization (PSO), artificial bee colony (ABC), and differential evolution (DE) are typically utilized.

In this article, the image processing parameters (IPPs) are adjusted to the optimized value in order to increase the accuracy of crack detection before use of convolution neural network (CNN) to eliminate non-crack candidates. As a result, the crack map will be improved significantly.

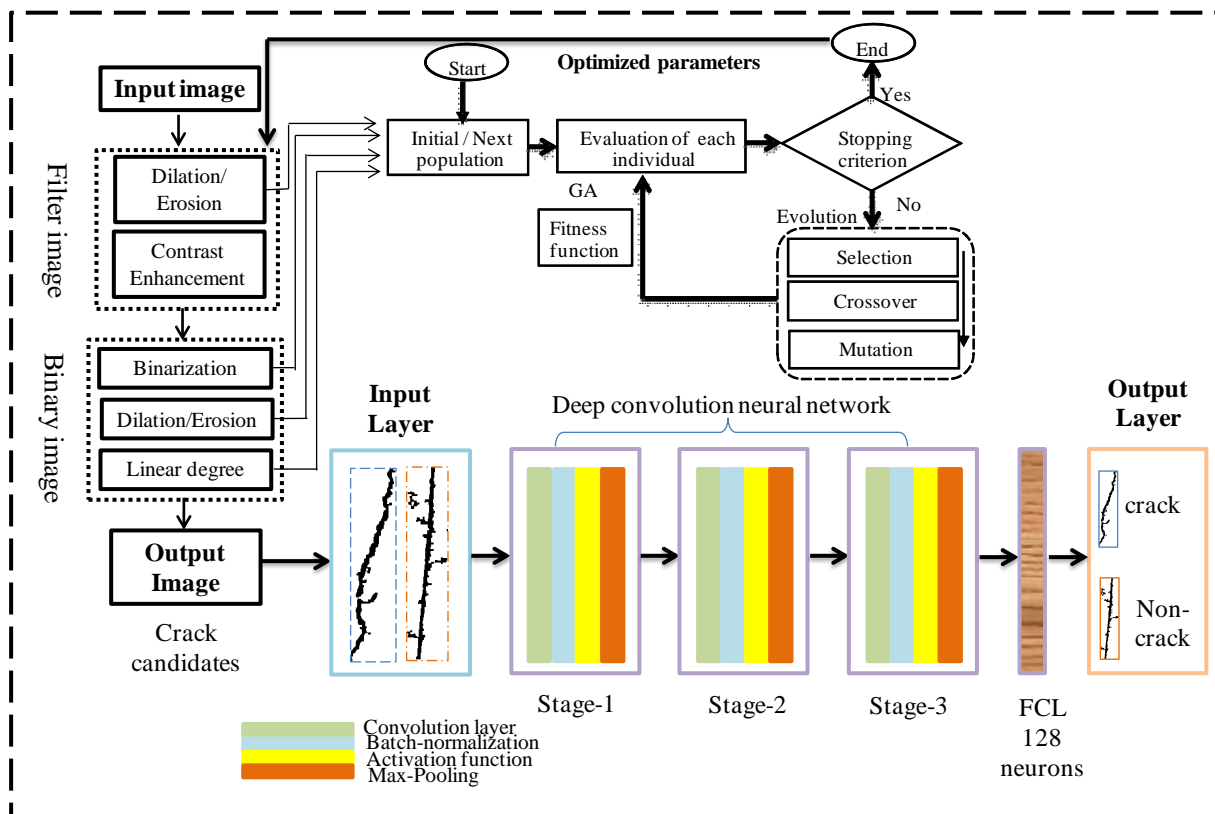


Figure 1. Pipeline of the automatic crack detection and classification using GA-CNN.

2. Proposed method

Figure 1 shows a pipeline of automatic crack detection and classification based on IPTs combined with image processing and deep learning.

2.1 Image Processing Techniques (IPTs)

The image processing techniques compose of three main parts. Namely, there are filtering image part, binary image part, and feature extraction part. Therein, filter image part comprises of morphological transform and contrast enhancement of the gray-scale image. The binary image part consists of binarization, and dilation-erosion transform of the binary image. To the end, geometric transform composes of labelling and linear degree.

2.1.1 Morphological transform. The morphology filter consists of opening transform of closing transform on gray-scale image with a predefined structuring element as the following equation:

$$M = \max([(G \oplus S) \ominus S] \ominus S], G) - G \quad (1)$$

Where M means the smooth image after morphological image processing, S is a structuring element, G is a gray-scale image converting from the corresponded original image. Additionally opening transform

includes dilation of erosion. Closing is an inverse operator of the opening as the following Eqs. (2) and (3):

$$(G \oplus S)_{(x,y)} = \max_{u,v} (G_{(x-u,y-v)} + S_{(u,v)}) \quad (2)$$

$$(G \ominus S)_{(x,y)} = \min_{u,v} (G_{(x-u,y-v)} - S_{(u,v)}) \quad (3)$$

The purpose of this part is firstly to smooth image as well as eliminate noises and shading and then make the sharpness of crack edge pixels. For an example, figure 2 (b) shows the result of smooth image M .

2.1.2. Structuring element design. Because the shape of crack is irregular elongation, the shape of the structuring element of line-type is adopted. To retrieve fully crack information in the various directions, combination line types of four directions into the structuring element such as $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ or six directions $\{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ\}$. In this study, the former type is used to reduce computational volume.

2.2. Contrast enhancement

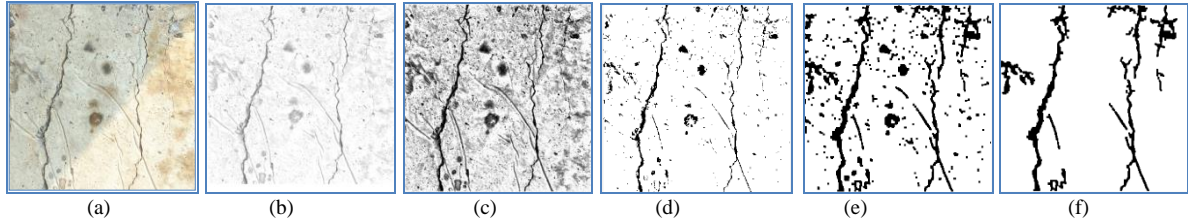


Figure 2. The illustrative results of image processing procedure. (a): Original image; (b): Morphological filter; (c): Contrast enhancement; (d):Binarization; (e): Dilation-thinning transform; (f): Final crack image processing.

Surface images of civil infrastructure are the typical dark environmental condition, narrow contrast. The intensity of crack pixels and background pixels is not much difference. Therefore, a histogram of the raw image is equalized before binarization to sharpen crack pixel as the following equation:

$$q_k = \sum_{i=0}^k \frac{n_i}{n} \quad k = 0, 1, 2, \dots, L-1$$

$$I_k(\text{out}) = (I_{\max} - I_{\min}) \times q_k(\text{in}) + I_{\min} \quad (4)$$

Where, n_i : is number of pixels that have i^{th} gray-scale value. n : the number of pixels; k : the input gray-scale level. L : the maximum gray-scale level (255). q_k : acquired normalized histogram. I_{\max} and I_{\min} are respective to the maximum and minimum intensities of the original image. $I_k(\text{out})$ is the new gray-scale intensity of the input pixel having $q_k(\text{in})$. As a result, figure 2.(c) shows crack enhancement after morphology filter is applied to smooth the original image.

2.2.1. Binarization. As shown in Eq. (5), the purpose of binarization is to segment gray-scale image into binary image. The binary image only has two values 255 (white pixel is background) and 0 (black pixel is crack or noise).

$$P(i) = \begin{cases} 0 & \text{if } I(x_i) < T \\ 255 & \text{otherwise} \end{cases} \quad (5)$$

Where T is a threshold value of the binarization (binary), it is necessary to find out an optimal value. $P(i)$ is the i^{th} pixel value after binarization step. figure 2(d) shows the results of binarization transform.

2.2.2. Dilation. The aim of a dilation operation is to connect crack fragments meanwhile noises are separated from the cracks. Therefore it results in the reduction of loss pixel. However, widths of crack pixels increase along with the crack shape. The size of structuring element of the dilation operation is considered as an adjusted parameter to optimized value by GA. In this paper, the shape of structuring

element is predefined in “square” type. The rule of the dilation operation is if any pixel in the input neighborhood is “1”, the output pixel is “1”. Otherwise, the output pixel is “0”.

2.2.3. Thinning. Thinning to a binary image is a morphology operator used to remove noise surrounding crack pixels. The purpose of this step is to prune branches from the crack shape as well as reduce noises after dilation operator is performed. Figure 2 (e) shows the image processing result of dilation and thinning transform.

2.2.4. Labeling and length threshold. After thinning, the pixel values of the image are presented by the 0s or 255s. The automatic labeling step is to combine the 8th-connection components as a single object with the same number. Hence, a single object can be a crack object or a noise object as shown in figure 2(f). Therefore, the length threshold decides as if the object is removed following equation:

$$P^{(i)} = \begin{cases} 255 \frac{Max(R_{xi}, R_{yi})^2}{S_i} < T' \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Where T' is a threshold value of the i^{th} label length to distinguish between a crack object and a noise object. S_i is the total number of pixels of the i^{th} label. R_{xi} , and R_{yi} are the number of pixels of the i^{th} label in the horizontal and vertical directions, respectively.

The purpose of the first part is to make blurred images as well as eliminate noises and shading. The main aim of the binarization is to segment grayscale image into binary image depending on a threshold value. The purpose of dilation/erosion that is processed in the binary image part is to connect fragment images of crack meanwhile noises are separated from the cracks. In the end, the threshold defined in linear degree is to decide whether the single objects in the binary image are removed.

The productivity of the filtered image depends on the parameter value of structuring element size in dilation and erosion transform. In binarization, if the threshold value is too high, many crack pixels are lost. If threshold value is too low, more noise will occur. Such it is necessary to find out an optimum threshold value. Similarly, the parameter values of dilation/erosion and linear degree in the binary image part also affect to the quality of output image. Therefore, these parameters are adjusted to the optimized values based on GA.

2.3. Application of GA to the image processing parameters optimization

2.3.1 Represented chromosome Design for solution candidates. The IPPs are combined together for creating an individual in a population. Next each individual is represented by a chromosome encoded to a binary string, as shown in figure 3.

Namely, the size of structuring element (s) is assigned by 6 bits, the threshold value of binarization (t) is expressed by 8 bits, dilation transform parameter (d) is expressed by 4 bits, and the linear degree is expressed by 6 bits (l). Table 1 shows the parameter value range which design based on the preliminary experiments.



Figure 3. A represented chromosome for solution candidates.

Table 1. Properties of parameter.

Variable	Range	Steps	Bits
s	[1 127]	2	6
t	[0 255]	1	8
d	[1 31]	0.5	4
l	[0 32.5]	0.5	6

2.3.2. Genetic algorithm. Figure 4 indicates a sequence of GA including into the crucial three stages. Namely, they consist of the initial population generation, fitness evaluation of each individual in the current population, and evolution operation to create the next generation. Namely, the detailed steps are presented as the following three steps:

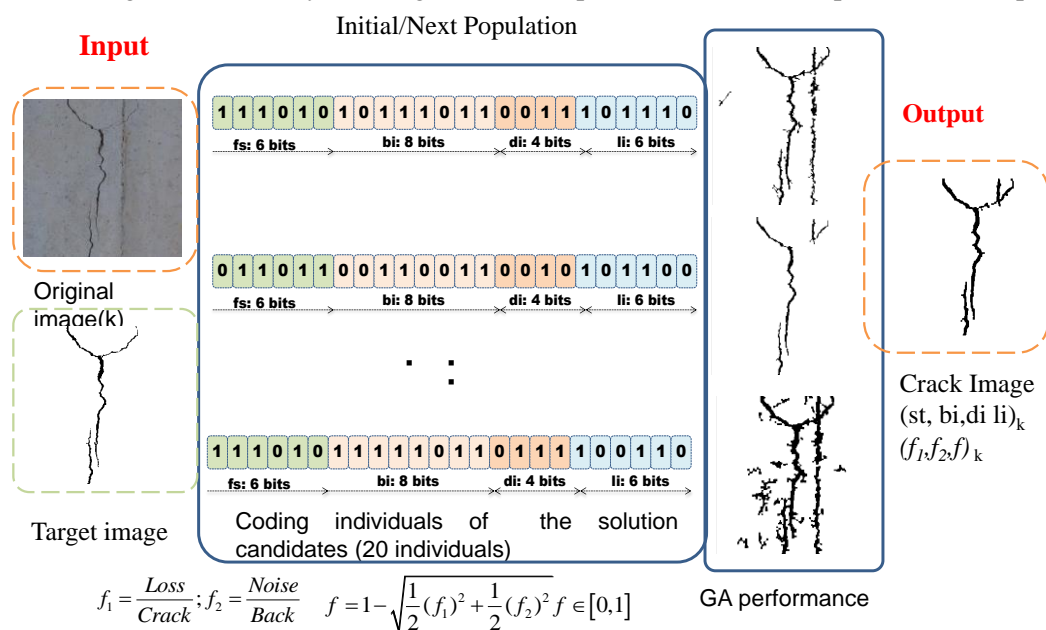
Step1. Generate initial population randomly

An initial population including 20 individuals was generated randomly with respect to 20 phenotypes to start fitness evaluation. To assess the fitness of the individual in the current population, an objective function to assess crack detection accuracy is defined as the Eq.(7). Loss and noise are computed based on comparison between the processed image and the target image shown in figure 5. As a result, the objective function (f) has to ensure the accuracy of extracted crack information with the minimum noises and losses as much as possible. The accuracy and the processing time can be considered as evaluation costs.

Step 2. Evolution operation

The evolution operation comprised of selection, crossover, and mutation is repeated until finding best solution. Each binary string encoded from the searching range of the parameter values has a corresponded fitness value. The probability of each string to be selected is proportional to its fitness value based on the Roulette wheel rotation randomly. The process is repeated for the second parent. Two elite members are kept forward to the next generation.

To improve quality of individual fitness, the crossover operation is used to create two new children from two selected parents with predefined probability. Crossover point is point laid on between 0 to the end of chromosome length. In this study, the single crossover point is selected. The part of the first parent

**Figure 4.** Procedure of genetic algorithm.

chromosome that runs until the crossover point is spliced with the part of the second parent chromosome that includes, and runs after, the crossover point shown in figure 3. The whole new generation is selected

in this manner. The mutation of bit strings ensue through bit flips at random positions. The purpose of the mutation operation is to create genotype diversification in the population in order to avoid local optimization leading to finding the best solution. Mutation point is chosen randomly. However, mutation rate is very small under 1% to avoid collapsing the genetic structure of the current population.

Step3. Stopping criterion

Evaluation of each individual meets the predefined maximum generation

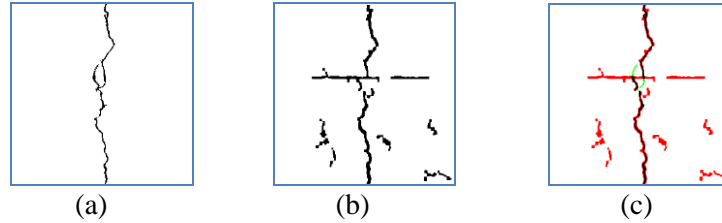


Figure 5. Result of supervised learning algorithm with one-pixel accuracy. (a): Ground-truth image; (b): Processed image; (c): Compared image. A number of red pixels in the compared image are noise pixels annotated Noise. A number of green pixels in the compared image are loss pixels annotated Loss.

$$f = \frac{1}{\sqrt{w_1 f_1^2 + w_2 f_2^2}}, f \in [0, 1]$$

$$f_1 = \frac{\text{Loss}}{\text{Crack}}; f_2 = \frac{\text{Noise}}{\text{Back}} \quad (7)$$

Where Crack and Back are the number of black pixels and white pixels in the ground-truth image, respectively. In this paper, the weight parameters of the objective function $w_1 = w_2 = 0.5$.

f_1, f_2 are loss rate and noise rate, respectively. f measures the accuracy of crack detection. f is larger value, the accuracy is higher.

3. Deep neural network

The output results of the image processing technique are crack images including crack pixels and crack-like non-crack pixels. The brightness of crack-like non-crack pixels is similar to the brightness of the crack pixels. The brightness is intensity value of pixel of binary image in 0 (black) or 255 (white). Figure 6 shows a crack image result of fully automated crack detection method using IPTs combined GA. As a result, the crack image contains many noise and crack like non-crack. The major challenge is how to classify them automatically in order to only keep true crack pixels.

3.1 Convolution layer

The DCNN composed of three stages. Each stage comprises of convolution layer, batch-normalization layer, max-pooling layer with drop out, and activate function (Cha et al.2017). The last layer is fully-connected layer so as to map noise candidates or crack candidates.

Table 2 indicates dimension of layers of DCNN shown in figure 1 as well as parameters of stride and padding. Where C1, C2, and C3 are the convolution layers; M-P1, M-P2, and M-P3 are the max pooling. FCL1 and FCL2 are fully-connected layers. Dr1 and Dr2 are drop out with a predefined probability ratio (Tong et al.2018). In this paper, these probability ratios are 0.4.

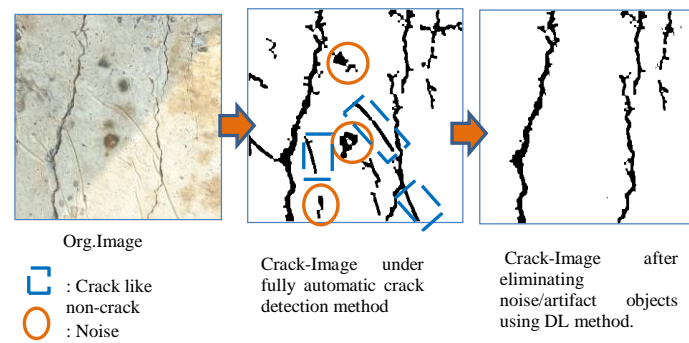


Figure 6. Application of deep learning method for true-crack pixel detection.

Table 2. Dimension of layers and parameters.

Layer Name	Number of Filter	Size	Stride	Padding	Layer Name	Number of Filter	Size	Stride	Padding
C1	32	16x16	2	2	C3	128	6x6	2	2
M-P1	1	3x3	2	-	M-P3	1	2x2	2	-
Dr1	1				Dr2	1			
C2	64	7x7	2	2	FCL1	1	128	-	-
M-P2	1	3x3	2	-	FCL2	1	2	-	-

4. Experiment results

The accuracy of the classification is expressed as the following equation:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

Where TP is the total of crack images detected correctly, TN is the total of non-crack images detected correctly. FP is the total of the crack images detected incorrectly. FN is the total of the non-crack images detected incorrectly.

Figure 7 shows comparison of the training accuracy between the proposed method and Cha_2017 method. The result expresses that the proposed method outperforms the other, and it gains maximum accuracy at epoch 30. Further, the experiment results of the proposed training model with various image data sizes. Training maximum accuracy (ACC) is 96.1%, the test data accuracy is 91% with 3000 images for each class. Moreover, the minimum testing accuracy belonged to 4000 image data set is 87.5%.

5. Conclusion

This paper found the optimization of parameters of IPTs using GA. Moreover, the IPTs results combined with DCNN method resulted in high automatic crack detection. The accuracy of training model is depending on the diverse of input data. The difference of accuracy between training and testing is large. It is necessary to improve more deep CNN layers to get high score.

The final purpose was to create crack map therefore requiring the pixel-level accuracy automatically. Disadvantage of DCNN method needed a large number of input data so as to gain the high fixed accuracy. More, the computation of DCNN model is heavy relied on GPU and computer configuration.

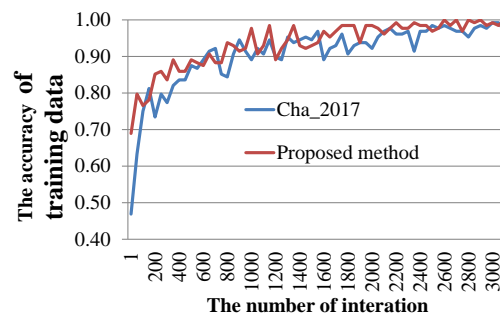


Figure 7. Comparison result between the proposed method and Cha_2017.

6. References

- [1] Cha, Y., Choi, W. (2017). Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks, *Computer-aided Civil and Infrastructure Engineering*, vol. 32, pp. 361–378.
- [2] Fujita, Y., Mitani, Y., and Hamamoto, Y. (2006). A method for crack detection on a concrete structure, *ICPR2006: IEEE 18th Int. Conf. on Pattern Recognition*, Vol. 3, New York, pp. 901–904.
- [3] Kawamura, K., Yoshino, K., Nakamura, H., Sato, Tarighat, A. (2013). A Valid Parameter Range Identification Method of a Digital Image Processing Algorithm for Concrete Surface Cracks Detection Using Genetic Algorithm and Decision Tree, *Proc. Japan Soc. Civ. Eng.* Vol. 69.
- [4] Kawamura, K., Miyamoto, Y., Nakamura, A., H., Sato, R. (2003). Proposal of a crack pattern extraction method from digital images using an interactive genetic algorithm, *Proc. Japan Soc. Civ. Eng.* Vol. 72, pp. 115–131.
- [5] Nguyen, C., Kawamura, K., Tarighat, A. (2016) A study on semi-automatic concrete cracks detection using interactive genetic algorithm, *JCI*, vol 38(1), pp. 2061–2066.
- [6] Tong, Z., Gao, J., Sha, A., and Hu, L., (2018). Convolution neural network for asphalt pavement surface texture analysis, *Computer-aided civil and infrastructure engineering*, Vol 0, pp. 1–17.
- [7] Zhang, W., Zhang, Z., Qi, D., Liu, Y. (2014). Automatic crack detection and classification method for subway tunnel safety monitoring, *Sensors*. Vol. 14, pp. 19307–19328.

Acknowledgments

This research is supported by JSPS KAKENHI grant number 15K06180.