A Deep Residual Network with Transfer Learning for Pixel-level Road Crack Detection

Seongdeok Bang[a], Somin Park[a], Hongjo Kim[a], Yeo-san Yoon[a], and Hyoungkwan Kim[a]

[a]Department of Civil and Environmental Engineering, Yonsei University, Seoul, South Korea
E-mail: bangdeok@yonsei.ac.kr, somin109@yonsei.ac.kr, hongjo@yonsei.ac.kr, yeosanyoon92@yonsei.ac.kr, hyoungkwan@yonsei.ac.kr

Abstract –

Image-based crack detection methods have been extensively studied due to their cost-effectiveness in terms of data acquisition and processing. However, automated crack detection still remain a challenging task due to complexity of image background and different patterns of cracks. To address these issues, this paper proposes a deep residual network with transfer learning for pixel-level crack detection on road surface images. The network was trained on 71 images of CrackForest dataset and tested on 47 images of it. Experimental results suggest that the deep residual network is superior to the existing algorithms with recall value and precision value of 84.90% and 93.57%, respectively.

Keywords – Crack Detection; Residual Network; Deep learning; Transfer Leaning

1 Introduction

Cracks on a road pavement that are not timely repaired increase maintenance and repair costs due to the acceleration of surface deterioration. To establish a proactive maintenance plan, road inspection data should be obtained frequently. In general, a government agency responsible for road monitoring either manually inspects the status of road networks or by using special road monitoring vehicles equipped with laser scanners, infra-red cameras, or high definition video cameras. However, it is still challenging to timely monitor a significant portion of road networks.

Image processing and machine learning algorithms have been widely studied to detect cracks on road pavement. Radopoulou and Brilakis [1] proposed the pavement management system that automatically detects patches in video frames. The authors used the video acquired by the car's parking camera in patch detection. Ouma and Hahn [2] proposed a pothole detection method on asphalt road pavements using the fuzzy c-means algorithm. The consumer-grade smartphone devices were used as road data collectors for the low-cost automated system. Oliveira and Correia [3] proposed a system for the automatic detection and characterization of cracks on road pavement images captured by a digital camera. These papers shows that the road crack detection methods based on image processing and machine learning algorithms allow frequent updates of road conditions at low cost.

CNN (Convolutional Neural Networks) have been used for road crack detection since CNN demonstrated an excellent performance in ImageNet Large Scale Visual Recognition Competition (ILSVRC) 2012 [4]. Gopalakrishnan et al. [5] used the VGG16 model, a kind of CNN, to identify the presence of cracks in images. Compared with the performance by traditional machine learning methods such as SVM (Support Vector Machine), RF (Random Forest), and LR (Logistic Regression), the CNN model proposed in [5] showed the highest accuracy for crack detection in highway images. Zhang et al. [6] proposed a CNN model with six layers for detecting cracks on road images. They conducted a comparative research with existing algorithms, such as SVM and boosting method, to confirm the superior crack detection performance of the proposed CNN. Wang et al. [7] applied a CNN to detect the existence of pavement crack using two different scales of grid (32×32 and 64×64).

Although the CNNs studied so far have performed well in terms of identifying the presence of cracks in a certain image region, they had a limitation in terms of detecting cracks at pixel-level. Sufficient pixel-level accuracy that can quantify the width, shape, and length of the crack enables detailed analysis of the road surface distress.

In this paper, we propose a deep residual network with transfer learning for pixel-level crack detection. The deep residual network classifies all pixels in road images into cracks and background. To achieve high performance with a relatively small number of training images, this study uses transfer learning, which is a learning method of fine-tuning the weights of a pre-

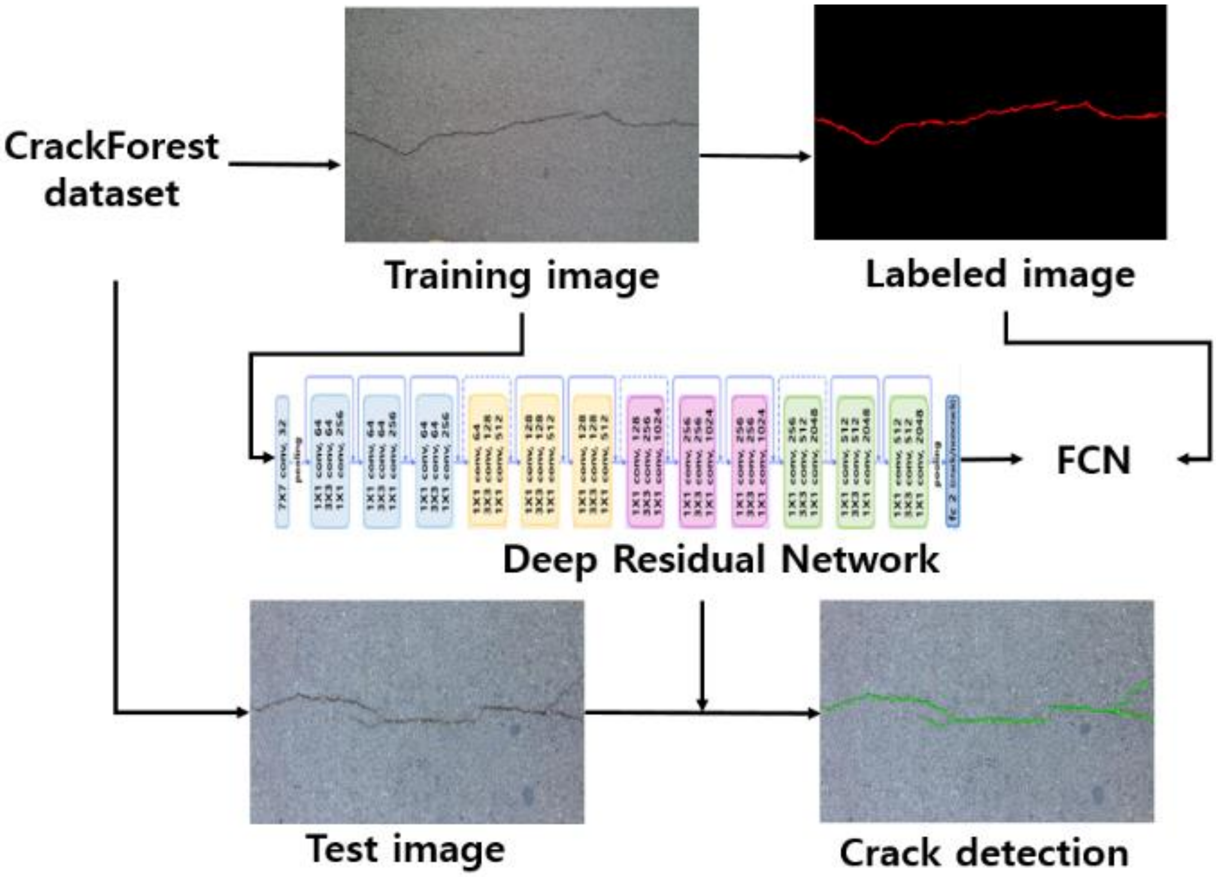

Figure 1. The deep residual network for road crack detection

trained CNN [8]. Experiments are conducted to demonstrate the performance of the proposed network.

2 Deep neural network with transfer learning for crack detection

CNN has been shown to be very effective in processing image data to perform computer vision tasks, such as semantic segmentation, image classification, object detection, and regression. A typical CNN consists of three kinds of layers: a convolution layer, a sub-sampling layer, and a fully-connected layer. The convolution layer applies convolution filters to the input image to produce output images. This layer has parameters such as number of channels, kernel size, number of filters, and stride size. The sub-sampling layer obtains feature maps with a lower dimension using average pooling or max pooling in the receptive field. In a CNN, convolutional layers and sub-sampling layers are typically repeated several times, and a fully-connected layer is placed at the end. The softmax function value of the fully-connected layer is the prediction value of the CNN.

The existing CNNs have the vanishing gradient problem that prevents the front layers of the model from learning in cases where many layers exist. The residual network [9] tried to solve this problem with a novel learning method that trains the network with the residuals–the difference between input and output of each computational block. The residual network with 152 layers shows better performance in large-scale object recognition problems than the VGG network [10] with 16 layers.

Figure 1 represents the proposed network. The network inputs are 480 × 320 × 3 images. The first layer is a 7 × 7 convolution layer with 64 depths and the second layer is a 3 × 3 max-pooling layer to perform down-sampling. The blocks, which consist of a 1 × 1 convolution layer, a 3 × 3 convolution layer, and a 1 × 1 convolution layer, are stacked behind the two layers (7 × 7 convolution layer and 3 × 3 max-pooling layer). An image passed through the 152 layers is reduced to a feature of 15 × 10 size. After that, the method of fully convolution network (FCN) [11] is applied to detect cracks using features extracted by the deep residual network. FCN enables pixel-level prediction by restoring features to original image size using the skip connection, which utilizes the intermediate layers' results and the final features of the residual network. The results of the intermediate layers compensate for losses in the up-sampling process performed by the bilinear interpolation.

A CNN typically require a large amount of labeled images to achieve high prediction accuracy. However, it is difficult to collect hundreds of thousands or even millions of road images and to label them manually. To address this problem, transfer learning was proposed to fine-tune pre-trained weights with big data, such as ImageNet. Unlike existing learning methods that randomly initialize the weights of CNN, transfer learning uses the weights of a pre-trained CNN as initial weights before the training process. Transfer learning has been shown to be useful in solving image classification problems of various domains, such as medical image

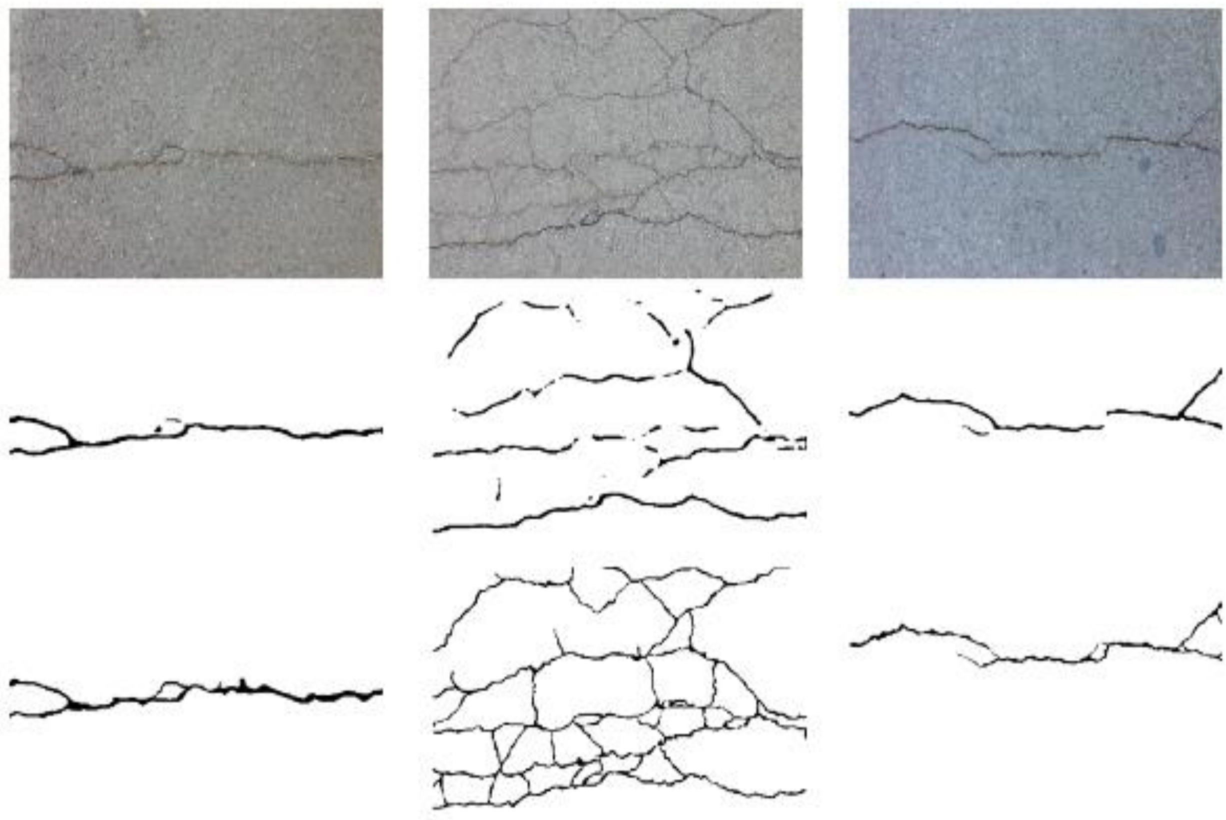

Figure 2. Results of residual network on CrackForest dataset (from top to bottom: original image, network prediction, ground truth)

processing [12]. In this context, we propose a deep residual network with transfer learning. Figure 1 illustrates the training and testing process of the proposed network. In this paper, the residual network with 152 layers, pre-trained by ImageNet dataset were used to detect cracks in images.

3 Experimental study

In this section, the performance of the deep residual network is analyzed. The training and test process of network was implemented in Python code. All experiments were performed on desktops with GTX 1080 Ti GPU and an Intel Core i7-7700 processor. The network was trained with a dataset used for the CrackForest [13], which is the existing crack detection method. The dataset consists of 118 images of the road surface in Beijing city acquired by iPhone5 with focus of 4mm, aperture of f/2.4, and exposure time of 1/134s. The images contain various types of noise, such as oil stains, road-specific textures, and shadows. The resolution of images is 480 × 320 × 3. Each image has a pixel-level ground truth labeled manually. All pixels except for the crack pixels are classified as background.

For direct comparison with the CrackForest algorithm [13], the experimental environment was set in the same manner. The precision, recall, and f1-measure, by examining the correspondence between the predicted pixel and ground truth, were calculated to evaluate the proposed network. We followed the assumption of the CrackForest algorithm that the predicted pixels which are within 5 pixels from the ground truth pixel are considered to be true positive pixels. 71 images of the dataset were used as training data, and the remaining 47 images were used as test data. The results are shown in Fig. 2, and the performance comparison between the proposed network and CrackForest algorithms are summarized in Table 1. Experimental results show that crack detection using the deep residual network is superior to the existing CrackForest algorithms with k-nearest neighbors algorithm (KNN) and support vector machine (SVM).

4 Conclusion

This paper proposes a deep residual network with transfer learning for pixel-level crack detection. The proposed deep residual network demonstrates better accuracy than traditional image processing and machine learning algorithms in pixel-level crack detection. The proposed network is able to quantify the geometry of cracks on various roads by performing pixel-based crack detection with high accuracy. If the network can be utilized in a variety of road image dataset with future studies, it is expected to help the asset managers establish proactive road maintenance strategies.

Table 1. Result evaluation on CrackForest dataset

Method	Precision	Recall	F1
CrackForest (KNN)	80.77%	78.15%	79.44%
CrackForest (SVM)	82.28%	**89.44%**	85.71%
CrackForest (One-Class SVM)	81.25%	86.45%	83.77%
Deep residual Network	**93.57%**	84.90%	**89.03%**

Acknowledgement

This work was supported by a grant (18CTAP-C133290-02) from Infrastructure and transportation technology promotion research Program funded by Ministry of Land, Infrastructure and Transport of Korean government.

References

[1] S.C. Radopoulou, I. Brilakis, Patch detection for pavement assessment, Automation in Construction 53 (2015) 95-104.

[2] Y.O. Ouma, M. Hahn, Pothole detection on asphalt pavements from 2D-colour pothole images using fuzzy c-means clustering and morphological reconstruction, Automation in Construction 83 (2017) 196-211.

[3] H. Oliveira, P.L. Correia, Automatic road crack detection and characterization, IEEE Transactions on Intelligent Transportation Systems 14 (1) (2013) 155-168.

[4] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems (2012) 1097-1105.

[5] K. Gopalakrishnan, S.K. Khaitan, A. Choudhary, A. Agrawal, Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection, Construction and Building Materials 157 (2017) 322-330.

[6] L. Zhang, F. Yang, Y.D. Zhang, Y.J. Zhu, Road crack detection using deep convolutional neural network, Image Processing (ICIP), 2016 IEEE International Conference on, IEEE (2016) 3708-3712.

[7] K.C. Wang, A. Zhang, J.Q. Li, Y. Fei, C. Chen, B. Li, Deep Learning for Asphalt Pavement Cracking Recognition Using Convolutional Neural Network, Airfield and Highway Pavements 2017 (2017) 166-177.

[8] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE transactions on medical imaging 35 (5) (2016) 1285-1298.

[9] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition (2016) 770-778.

[10] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).

[11] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, Proceedings of the IEEE conference on computer vision and pattern recognition (2015) 3431-3440.

[12] N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, M.B. Gotway, J. Liang, Convolutional neural networks for medical image analysis: Full training or fine tuning?, IEEE transactions on medical imaging 35 (5) (2016) 1299-1312.

[13] Y. Shi, L. Cui, Z. Qi, F. Meng, Z. Chen, Automatic road crack detection using random structured forests, IEEE Transactions on Intelligent Transportation Systems 17 (12) (2016) 3434-3445.