



Classification of Concrete Crack using Deep Residual Network

Umme Hafsa Billah¹, Hung Manh La¹, Nenad Gucunski², Alireza Tavakkoli¹

¹ *Department of Computer Science and Engineering, University of Nevada – Reno, NV89557, USA. Corresponding Author's Email: hla@unr.edu.*

² *Department of Civil and Environmental Engineering, Rutgers, the State of University of New Jersey – Piscataway, NJ08854, USA.*

Abstract

Inspection of civil infrastructure is a demanding task, which also requires in-depth knowledge and experience about infrastructure assets. Manual inspection is often labor intensive, making the task more difficult and expensive. Besides, the time-consuming process of manual inspection and omission to detect critical damage may lead to serious failures. Automated systems can mitigate most of the adversities stemming from manual inspection. In this paper, a crack detection technique for civil infrastructures, like roads and bridge decks, is presented. The technique can detect cracks of different shapes and orientation. A data set of 21,996 images of various roads, highways and bridge decks was prepared at different times of the day and for different light orientation conditions. Though convolutional neural network (CNN) is very popular for image classification, it shows degradation when a deep network is used. For this reason, a 50 layered residual learning network was used to train the data. The network shows significant successes on different test cases collected from an arbitrary source, with a 94% percent crack detection accuracy. This work provides sufficient evidence that deep networks with large datasets can be used for crack classification without any degradation in accuracy of the network.

1. Introduction

Modern transportation systems rely on well performing civil infrastructure, including roads, bridges, railways, etc. It can be said that the transportation infrastructure is the backbone of people and goods movement.. Concrete, the most widely used material to build the infrastructure is susceptible to deterioration with time from various causes, such as repeated overloading, chemical processes, like corrosion, environmental effects, and many more. Internal deterioration in concrete will lead to formation of defects, most notably to cracking and delamination. Presence of such defects will affect the performance and longevity of the infrastructure asset. Therefore, regular inspection, and corresponding maintenance activities, is needed to retain the performance and prevent premature failure. Manual inspection of concrete elements, in particular detection and characterization of cracks is time consuming, labor-intensive, error prone and uneconomic. An automated inspection system, which can assess the condition of an infrastructure asset in a continuous manner, is critically needed. Automated inspection system for

civil infrastructure was addressed in many studies, including more recently by Gibb et al. (2018); La et al. (2013); La et al. (2017), Cha et al. (2017). Crack detection in concrete has many challenges. The crack shape, orientation and the background varies with structure and the environment they are in. The illumination, noise present in real time crack images also makes crack detection challenging. There are many traditional image processing techniques, which fail in cases of crack detection due to the variability of real time images, as mentioned in Cha et al. (2017). There have been studies showing results, where state-of-the-art edge detection techniques were used, like canny edge detector, sobel operator fails to classify crack in concrete smoothly, e.g. in. Prasanna et al. (2012); Sham et al. (2018). In addition, Laplacian of Gaussian transforms have been previously used for extracting image cracks, e.g. by Lim et al. (2014). However, the factors enabling a real time crack detection, as mentioned above, degrade the performance of such techniques. Some machine learning techniques, such as SVM (Prasanna et al. (2012)), Adaboost and Random forest (Prasanna et al. (2016)), CNN ((Gibb et al. (2018); Cha et al. (2017); Dinh et al. (2016))), were previously used for crack classification in concrete. As mentioned in Cha et al. (2017), deep architectures like CNN, and large data sets collected in varying conditions are needed for detection of cracks from images in real time. Various deep CNN structures are used for classification of cracks, such as AlexNet designed by Krizhevsky et al. (2012), or VGGNet represented by Simonyan et al. (2014). The deep CNN structures are subject to degradation (in accuracy) and optimization is a problem when more layers are added to the network (He et al. (2016)). Therefore, considering the factors of crack detection, using a deep CNN structure could be a challenge. To address this problem, a residual learning network designed by He et al. (2016), widely known as Resnet, is used in this study to recognize cracks in images. With residual learning approach, significant improvements are made in the crack detection, and with less complexity. The following sections provide a discussion on the residual network architecture and training process, experimental data collection and results, and the conclusions of the study.

2. Methodology

Deep residual learning network designed by He et al. (2016) was used in this study for the classification purpose. The performance of a very deep network degrades, as well as optimization issue of hyperparameters occurs, when more layers are added to the network. These problems can be resolved by using a residual mapping instead of adding new layers into a CNN to the underlying mapping. Therefore, a 50 layer deep network with a residual mapping function is used in this study. The complete workflow of the system is represented in Figure 1.

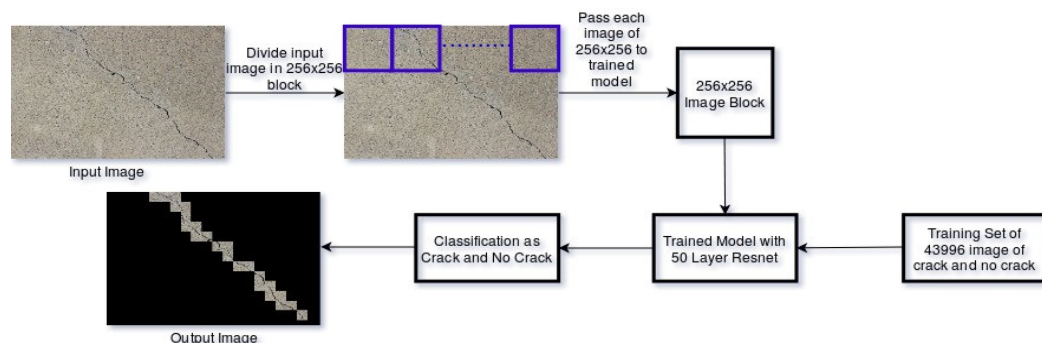


Figure 1. The work flow of the proposed system.

2.1 Network Architecture

The CNN architecture consists of a fifty layer deep network with residual learning function. The input layer takes an image of a size of 256x256x3. Those represent the height, width, and the number of channels of the image. It is first fed to the network. The input image is passed through a 7x7 convolutional layer, which outputs a 112x112 image. A 3x3 *maxpooling* is performed on the image after that results with an output of 52x52 dimensions. The residual function is performed on each three layer stack consisting of 1x1, 3x3, and 1x1 filters. Later, a global average pooling is performed on the network, followed by a 1x1 *soft max* activation in the fully connected layer. Table 1 represents the architecture of a fifty layer resnet network.

Table 1. Architecture of fifty layer residual network.

Name of the Layer	Output Size	50 Layer Organization
Conv1	112x112	7x7, 64, stride 2
Conv2_x	56x56	3x3 max pool, stride 2
		1x1, 64 3x3, 64 x 3 1x1, 256
Conv3_x	28x28	1x1, 128 3x3, 128 x 4 1x1, 512
Conv4_x	14x14	1x1, 256 3x3, 256 x 6 1x1, 1024
Conv5_x	7x7	1x1, 512 3x3, 512 x 3 1x1, 2048
	1x1	Fully connected layer

When extra layers are added to construct a very deep network, it is more efficient to map these layers into a residual function, rather than mapping the added layers to underlying mapping of the network. Let us assume that the underlying mapping of an added layer is $H(x)$, where x is the input of the first layer. The residual mapping of the added layers would be $F(x) : H(x)-x$. Using the equation $F(x)+x$, a shortcut connection can be added to after a number of stacked layers, consisting of existing layers and added new layers. This residual learning function is applied to every few stacked layer. Initially, Resnet by He et al. (2016) uses a 32 layer architecture and is applying residual learning in every two layers with 3x3 filters. However, in Resnet with 50 layer a building block is designed with applying a residual function in every three layers. The stacked layers consist of filters with 1x1, 3x3, 1x1 dimensions, respectively.

2.2 Training with Residual Network and Crack Detection

For training the aforementioned network architecture, a data set containing images of 256x256 dimensions have been prepared. Images of different road and bridge deck surfaces with different illumination, intensity and crack orientation are added to the training set. A fully connected layer

over the 7x7 convolutional layer of a typical Resnet architecture (He et al. (2016)) was used in this study. The network parameters were optimized using a Stochastic gradient descent (SGD) value of 0.01. The network was trained with 80 epochs, where weights were updated in each epoch. The initial weights were assigned randomly, rather than using the weight of the state-of-the-art IMAGENET data set.

A real time image of road or bridge deck surface can contain multiple crack types of different orientation. There may be sub parts of an image where non-cracks exist. To identify different patches of an image as a crack and non-crack, it is effective to subdivide an image. Therefore, an image of an arbitrary size is taken as input, and the image is then divided into sub images of 256x256x3 dimensions (height, width, channel). The images are then fed one by one into the CNN. After that, the images are stitched back together with a classification as crack or non-crack.

3. Experimental Results

The fifty layer residual network architecture classifies image in two classes as crack or non-crack. The trained model is compared with the CNN structure proposed by Gibb et al. (2018). In this section, the data set preparation and results by our experiment is discussed elaborately.

3.1 Data Set Preparation

Data were collected from two types of infrastructure assets, roads and bridge decks in particular, during different day and night times, as well as with having shadows and different backgrounds. To capture images, an autonomous robotic system with NDE sensor fusion method, represented by Gibb et al. (2018), was used to enable in-depth data collection in a limited inspection time. The Seekur Jr mobile robot from Omron Adept Technologies, Inc. was used as the base platform for the robotic system designed by Gibb et al. (2018). The Seekur mobile robot system was used previously by La et al. (2013) and La et al. (2017) for civil infrastructure inspection. However, Seekur Jr is more suitable for civil infrastructure inspection because of its mobility in adverse conditions, as reported by Gibb et al. (2018). Moreover, we have included some arbitrary images from various online sources, so that the system is not biased toward certain camera specification. The collected images have been subdivided into 256x256 pixels regions, and labelled manually into crack and non-crack classes. Images containing major cracks were included in the data set, while images with minor cracks were discarded. After that, data augmentation was performed on the crack images. For this purpose, a crack image is flipped horizontally and vertically, and rotated ninety degrees left and right. Besides, some crack images are included by decreasing brightness of images, so that the trainer can recognize cracks in very low contrast images. To decrease brightness, we have used the equation, $I = I \cdot \alpha + \beta$, where I is an image pixel, the value of α is set to 1, and β is set to -80. The data set was partitioned into two sets for training and validation. The training set contained 21996 images of 256x256x3 dimensions in the crack class, and 22000 images in the non-crack class. 300 images of crack and non-crack classes were taken for validation of the trained module. For testing, the training performance test set was created, including various images available online. The test set consisted of 326 images in the crack class, and 301 images in the non-crack class.







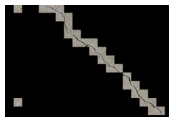
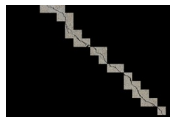
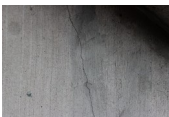



3.2 Result Analysis

The accuracy of our method is measured using the following formula:

$$Accuracy = \frac{\text{Number of correct classification}}{\text{Total number of instances}}$$

The residual learning network achieved 94% accuracy after testing on a test set of 627 images of crack and non-crack classes, while the CNN model proposed by Cha et al. (2017) and Gibb et al. (2018) achieved 89.8% and 96 % accuracy respectively (trained with same data set). Besides, six images of 3486x5184x3 dimensions were taken and fed into the system. The result showed that the Resnet model can identify cracks more accurately than Cha et al. (2018) and classifies less false positives than the CNN model by Gibb et al. (2018). Table 2 shows some examples of the crack detection result of Resnet model, CNN model by Gibb et al. (2018) and Cha et al. (2017). Though the CNN model by Gibb et al. (2018) detects more crack pixels, it detects more false positives, in comparison to the Resnet model. In some cases, the Resnet model detects more crack pixels as shown in Image1 of Table 2 than the other two CNN structures. Some cases show that the Resnet detects less crack pixels than the CNN model by Gibb et al. (2018). However, Resnet outperforms the CNN model by Cha et al. (2017) in all cases.

Table 2: Crack detection result on several images using three different processes. The correct or false classification is identified by how many 256x256 blocks are classified as crack class

Input image	CNN model by Cha et al. (2017)	CNN model by Gibb et al. (2018)	Fifty layer Resnet model
 Image1	 Correct = 7, Incorrect = 0	 Correct = 8, Incorrect = 0	 Correct = 11, Incorrect = 0
 Image2	 Correct = 18, Incorrect = 0	 Correct = 26, Incorrect = 2	 Correct = 23, Incorrect = 0
 Image3	 Correct = 20, Incorrect = 79	 Correct = 24, Incorrect = 16	 Correct = 24, Incorrect = 14

4. Conclusion

A 50 layer deep residual network was trained using a dataset of 43996 images. The results show an improvement in crack classification over a shallow CNN structure. It was shown empirically that deep networks with large data sets can be used for classification of cracks in concrete without concerns regarding the degradation problem typical for deep networks. Though the Resnet comes with different number of layers, a fifty layer Resnet can identify cracks more

accurately than a 18 layer or 32 layer Resnet architecture. In the future work, the performance of other Resnet architectures with 101 layers and 152 layers will be analysed. Besides, the current Resnet architecture falsely classifies some of the crack blocks and some minor crack blocks as non-crack. The aim is to improve accuracy for minor cracks in concrete. Moreover, the images blocks that are fed into the system consist of 256x256 blocks. It is believed that smaller blocks for training and classification will improve the network performance. The goal is to further investigate the ideal size of image blocks for classification.

Reference

- Gibb, S., La, H. M., Le, T., Nguyen, L., Schmid, R., & Pham, H. (2018). Nondestructive evaluation sensor fusion with autonomous robotic system for civil infrastructure inspection. *Journal of Field Robotics*, 35(6), 988-1004.
- Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 32(5), 361-378.
- La, H. M., Gucunski, N., Dana, K., & Kee, S. H. (2017). Development of an autonomous bridge deck inspection robotic system. *Journal of Field Robotics*, 34(8), 1489-1504.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Prasanna, P., Dana, K. J., Gucunski, N., Basily, B. B., La, H. M., Lim, R. S., & Parvardeh, H. (2016). Automated Crack Detection on Concrete Bridges. *IEEE Trans. Automation Science and Engineering*, 13(2), 591-599.
- Dinh, T. H., Ha, Q. P., & La, H. M. (2016, November). Computer vision-based method for concrete crack detection. In *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)* (pp. 1-6). IEEE.
- Romero, A., Ballas, N., Kahou, S. E., Chassang, A., Gatta, C., & Bengio, Y. (2015). Imagenet classification with deep convolutional neural networks. In *International Conference on Learning Representations*.
- Lim, R. S., La, H. M., & Sheng, W. (2014). A robotic crack inspection and mapping system for bridge deck maintenance. *IEEE Transactions on Automation Science and Engineering*, 11(2), 367-378.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- La, H. M., Lim, R. S., Basily, B. B., Gucunski, N., Yi, J., Maher, A., ... & Parvardeh, H. (2013). Mechatronic systems design for an autonomous robotic system for high-efficiency bridge deck inspection and evaluation. *IEEE/ASME Transactions on Mechatronics*, 18(6), 1655-1664.
- Prasanna, P., Dana, K., Gucunski, N., & Basily, B. (2012, April). Computer-vision based crack detection and analysis. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2012* (Vol. 8345, p. 834542). International Society for Optics and Photonics.
- Sham, F. C., Chen, N., & Long, L. (2008). Surface crack detection by flash thermography on concrete surface. *Insight-Non-Destructive Testing and Condition Monitoring*, 50(5), 240-243.