Deep Learning-Based Crack Detection Using Mask R-CNN Technique

3 authors:

Chengjun Tan
Hunan University
10 PUBLICATIONS **31** CITATIONS

Nasim Uddin
University of Alabama at Birmingham
262 PUBLICATIONS **936** CITATIONS

Yahya M. Mohammed
University of Alabama at Birmingham
8 PUBLICATIONS **14** CITATIONS

Deep Learning-Based Crack Detection Using Mask R-CNN Technique

Chengjun Tan[1], Nasim Uddin [2], Yahya M. Mohammed [3]

[1, 2, 3] *Dept. of Civil, Construction, and Environmental Engineering, The University of Alabama at Birmingham – USA, Email:* tan1990@uab.edu

Abstract

Cracks of civil infrastructures, including bridges, dams, roads, and skyscrapers, potentially reduce local stiffness and cause material discontinuities, so as to lose their designed functions and threaten public safety. This inevitable process signifier urgent maintenance issues. Early detection can take preventive measures to prevent damage and possible failure. With the increasing size of image data, machine/deep learning based method have become an important branch in detecting cracks from images. This study is to build an automatic crack detector using the state-of-the-art technique referred to as Mask Regional Convolution Neural Network (R-CNN), which is kind of deep learning. Mask R-CNN technique is a recently proposed algorithm not only for object detection and object localization but also for object instance segmentation of natural images. It is found that the built crack detector is able to perform highly effective and efficient automatic segmentation of a wide range of images of cracks. In addition, this proposed automatic detector could work on videos as well; indicating that this detector based on Mask R-CNN provides a robust and feasible ability on detecting cracks exist and their shapes in real time on-site.

1. Introduction

Cracks of civil infrastructures, including bridges, dams, roads, and skyscrapers, potentially reduce local stiffness and cause material discontinuities, so as to lose their designed functions and threaten the public safety (Budiansky and O'connell 1976; Kinra et al. 2006). This inevitable process signifier urgent maintenance issues. Early inspection and detection can take preventive measures to prevent damage and possible failure (Dhital and Lee 2012). Manual inspection is the most traditional approach for crack inspection, which highly depends on the specialist's knowledge and experience, lacking objectivity in the quantitative. In the manual inspection, the sketch of the crack is prepared manually and the conditions of the irregularities are noted, leading to labor-intensive, time-consuming and even dangerous (Mohan and Poobal 2017; Shi et al. 2016).

For fast and reliable surface defect detection, automatic detection is expected to develop instead of the slower and subjective traditional human inspection. Thereby there is increasing interest in vision-based crack detection using image-processing techniques (IPTs) for Non-destructive inspection or structural health monitoring (SHM). IPTs has a significant advantage of that all superficial defects are almost likely identifiable (Cha et al. 2017). In early studies, the cracks on a concrete surface generally are considered to possess two key properties i.e. it has low luminance and their structure shape is thinner and different with the other textural patterns (Sankarasrinivasan et al. 2015). However, edge detection actually is an ill-posed problem, because the results are substantially affected by the noises with no optimal solution (Ziou and Tabbone 1998).

Recently, image-processing algorithms based on convolutional neural networks (CNNs) have led to dramatic advances in computer vision for feature extraction (He et al. 2016). This feature extraction technique based on CNN have been adopted to learn the cracks features and then detect them (Cha et al. 2017; Yokoyama and Matsumoto 2017). However, cracks in real-world situations vary extensively leading to learning hardly at the global view, meaning that both procedures of crack characteristic learning and detection are based on small tokens (small piece pixel) as mentioned above rather than a whole picture. As shown in Figure 1, the sliding window technique was applied to help detect cracks on the entire picture. Strictly speaking, these approaches only use CNN to classify crack or uncrack via every small piece pixel of a picture, rather than detect cracks location and the shape directly at a global view. Hence, this proposed methods cannot detect crack in real-time on site.

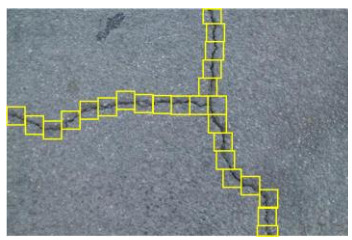

Figure 1 Schematic of CNN crack detector

In this study, we investigate to build a real-time automatic crack detector using state-of-the-art technique of Mask R-CNN to detect crack from image datasets. Mask R-CNN technique is a recently proposed algorithm for object detection, object localization, and object instance segmentation of natural images (He et al. 2017). We defined a new threshold value for decision final output mask leading to good performance of instance segmentation on irregular long-thin object. Extensively varying images were applied to training the Mask R-CNN based detector and the trained detector offers good results in extensively varying images and even video datasets.

2. Mask R-CNN

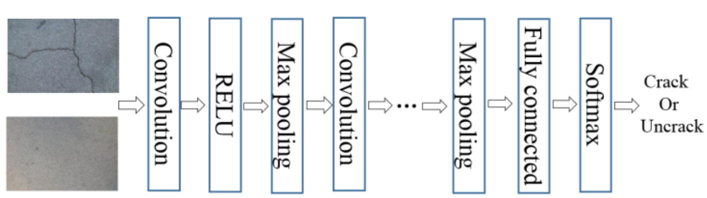

Figure 2. Convolution neural network

A convolution neural network (CNN) is a typical machine learning method especially for image recognition. By convolution, it is possible to extract every feature of images. Figure 2 shows the key architectures of CNN, which mainly includes convolution layer, activate layer, pooling layer, fully connected layer and softmax layer. The convolution process performs the convolution of features. Patten in images is to be detected by the convolution of the feature, where is automatically acquired by training/learning. Output of each convolution process is called feature map. For more mechanisms of abovementioned layers, here refer readers to references (Wu 2017; Yokoyama and Matsumoto 2017). Recently, "very deep" CNNs have significantly improved image feature extraction and image classification (Krizhevsky et al. 2012).

In Comparison to image classification, object detection from images is a more challenging task that requires more complex methods to solve. R-CNN is one of the most basic models for implementing object detection (Girshick et al. 2014). A region proposal method such as selective search (SS) and edge boxes(EB) (Girshick et al. 2014; Uijlings et al. 2013) will be applied on input images to extract around 2000 bottom-up region proposals, and then each region of interest (ROI) is computed feature with CNN. Later, Girshick (2015) and Ren et al. (2015) added the way of generating ROI into network to improve the calculation efficiency and accuracy, which are referred to as fast R-CNN and faster R-CNN respectively.

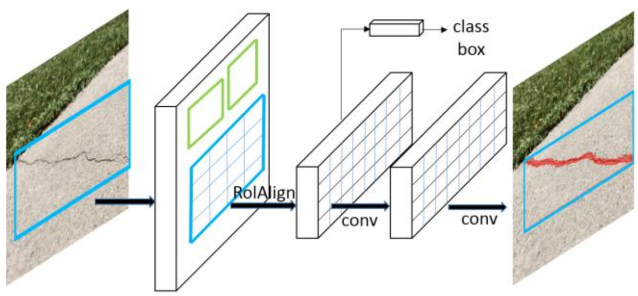

Figure 3. The Mask R-CNN framework for instance segmentation

The Mask R-CNN model was developed by Facebook AI team in 2017 and significantly extends the Faster R-CNN model for object localization, semantic segmentation, and object instance segmentation of natural images (He et al. 2017). It is described as providing a 'simple, flexible and general framework for object instance segmentation' as shown in Figure 3 (Johnson 2018). Mask R-CNN follows the Faster R-CNN model of a feature extractor followed by an operation known as ROI-Pooling to produce standard-sized outputs suitable for input to a classifier, with three important modifications (Johnson 2018). (1) Mask R-CNN replaces the imprecise ROI-Pooling operation applied in Faster R-CNN to an ROIAlign operaion that allows very precise instance segmentation masks to be created. (2) Mask R-CNN adds a network head that is also a small fully convolutional neural network to produce the desired instance segmentations. (3) Mask and class predictions are decoupled, meaning that the mask network head predicts the mask independently from the network head predicting the class (Figure 3).

3. Application

Dataset: The data used for these experiments consist of 352 crack images and corresponding annotations for crack in each image; 118 crack images and their crack ground truths are from the open resource in reference (Shi et al. 2016). The remaining crack images are randomly downloaded online and their crack ground truths are created manually with Matlab tools (2018). Of these 352 images in the dataset, 286 images were used for training; 36 images were used for validating the model and 30 images were held out for testing. Figure 4 illustrates the examples of training images with crack ground truth.

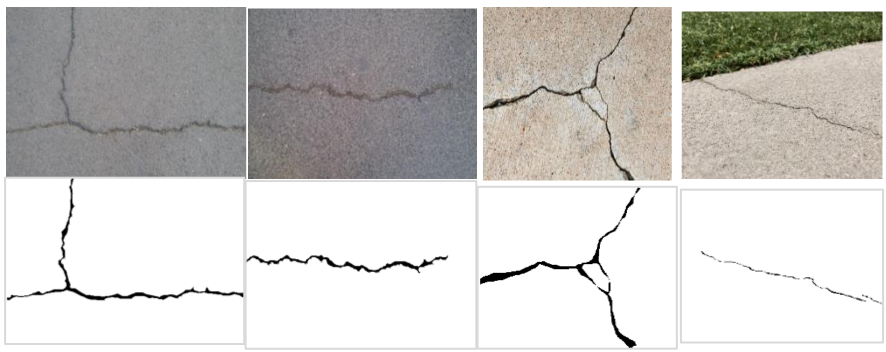

Figure 4. Examples of training images with crack ground truth

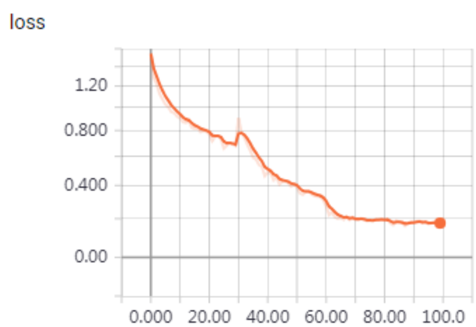

Figure 5. The total loss function versus each epoch (x axis represents epoch; y axis represents function)

Training: The Mask R-CNN backbone applied in this paper uses a "very deep' CNN of ResNet-101 (He et al. 2016). The implementation used is based on an existing implementation by Matterport Inc. released under an MIT License, and which is itself based on the open-source libraries Keras and Tensorflow (2017; 2018; 2018). For this experiment, rather than training the network end-to-end from the start, we initialize the model using the known weights obtained from pre-training on the MSCOCO dataset (Lin et al. 2014). All CNN layers was trained in three stages: (1) training only the network heads, which are randomly initialized; (2) training the upper layers of the network (from stage 4 and up in the ResNet-101 model); (3) reducing the learning rate by a factor of 10 and training end to end. The total train epochs are 100 using stochastic gradient descent with the momentum of 0.9, starting with a learning rate of 0.001 and ending with a learning rate of 0.0001. We use a batch size of two on a single NVIDIA Tesla P100 16GB GPUs. Gradients are clipped to 5.0 and weights are decayed by 0.0001 for each epoch. To help avoid over fitting, the dataset was augmented using Gaussian blurring, random rotations, and random horizontal and vertical flips. The total loss functions are illustrated in Figure 5.

Results: Figure *6* shows some excellent examples on the test images. Although the test images vary extensively, the developed crack detector is still able to find and mark the cracks, unaffected by the strong nature noises of images such as dirt, oil spills, staining, shadow and other object interferences etc. Furthermore, the trained Mask R-CNN model was applied to a video for crack detecting. As expected, it performs the excellent result as well, although the predict masks are mostly fatter than the ground truth. In addition, it was demonstrated that the Mask R-CNN based crack detector is feasible on detecting cracks exist, location and their shapes in real time on-site.

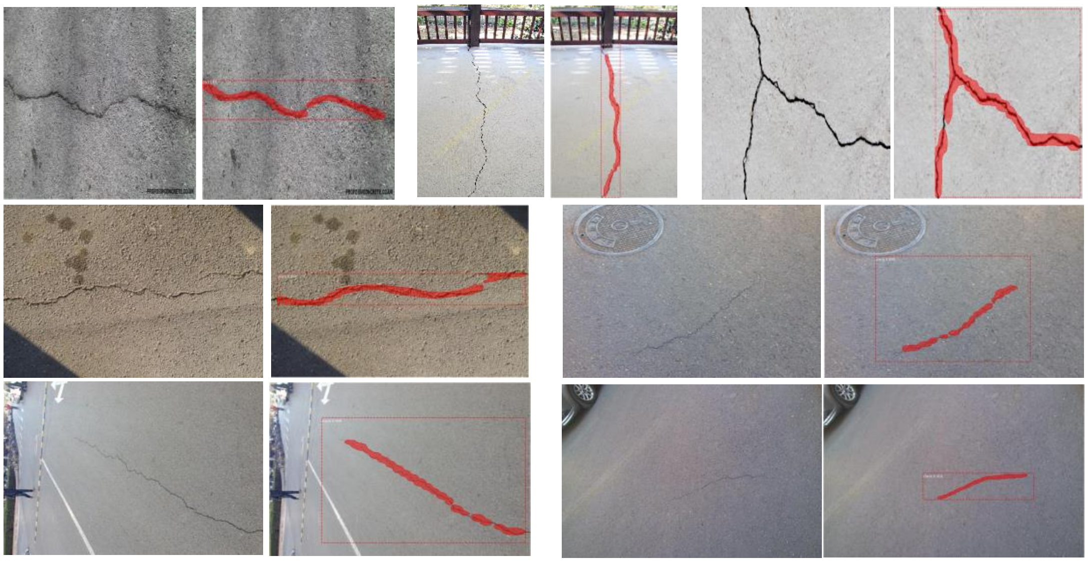

Figure 6. Examples of crack detection on test images

4 Conclusions

In this paper, we build an effective and fast automatic crack detection method based on Mask R-CNN, which can suppress noises efficiently by learning inherent feature of cracks from a 'very deep' CNN. Comparing to only CNN based method working on small batch of images, it can detect the crack of images in a global view providing a bound box for the crack exist and location. In addition, it offers mask for the predict crack showing shape of cracks simultaneously. Mask R-CNN works very fast and performs excellent on video dataset for crack detection, which can become a real-time crack detector on site equipped to the advanced UAVs.

Acknowledgment

National Science of Foundation (NSF-CNS- 1645863, and NSF-CSR- 1813949) sponsors this research. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the sponsors.

Reference

(2017). "Mask RCNN." <https://github.com/matterport/Mask_RCNN>. (15 MAR, 2018).
(2018). "Keras." <https://github.com/keras-team/keras>.
(2018). "Matlab ", <https://www.mathworks.com/products/matlab.html>.
(2018). "tensorflow ", <https://github.com/tensorflow>. (15 Mar, 2018).

Budiansky, B., and O'connell, R. J. (1976). "Elastic moduli of a cracked solid." *International journal of Solids and structures*, 12(2), 81-97.

Cha, Y.-J., Choi, W., and Büyüköztürk, O. (2017). "Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks." *Computer-Aided Civil and Infrastructure Engineering*, 32(5), 361-378.

Dhital, D., and Lee, J. (2012). "A fully non-contact ultrasonic propagation imaging system for closed surface crack evaluation." *Experimental mechanics*, 52(8), 1111-1122.

Girshick, R. (2015). "Fast r-cnn." *arXiv preprint arXiv:1504.08083*.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. "Rich feature hierarchies for accurate object detection and semantic segmentation." *Proc., Proceedings of the IEEE conference on computer vision and pattern recognition*, 580-587.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. "Mask r-cnn." *Proc., Computer Vision (ICCV), 2017 IEEE International Conference on*, IEEE, 2980-2988.

He, K., Zhang, X., Ren, S., and Sun, J. "Deep residual learning for image recognition." *Proc., Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.

Johnson, J. W. (2018). "Adapting Mask-RCNN for Automatic Nucleus Segmentation." *arXiv preprint arXiv:1805.00500*.

Kinra, V. K., Ganpatye, A. S., and Maslov, K. (2006). "Ultrasonic ply-by-ply detection of matrix cracks in laminated composites." *Journal of Nondestructive Evaluation*, 25(1), 37-49.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. "Imagenet classification with deep convolutional neural networks." *Proc., Advances in neural information processing systems*, 1097-1105.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. "Microsoft coco: Common objects in context." *Proc., European conference on computer vision*, Springer, 740-755.

Mohan, A., and Poobal, S. (2017). "Crack detection using image processing: A critical review and analysis." *Alexandria Engineering Journal*.

Ren, S., He, K., Girshick, R., and Sun, J. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Proc., Advances in neural information processing systems*, 91-99.

Sankarasrinivasan, S., Balasubramanian, E., Karthik, K., Chandrasekar, U., and Gupta, R. (2015). "Health monitoring of civil structures with integrated UAV and image processing system." *Procedia Computer Science*, 54, 508-515.

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

Uijlings, J. R., Van De Sande, K. E., Gevers, T., and Smeulders, A. W. (2013). "Selective search for object recognition." *International journal of computer vision*, 104(2), 154-171.

Wu, J. (2017). "Introduction to convolutional neural networks." *National Key Lab for Novel Software Technology. Nanjing University. China*.

Yokoyama, S., and Matsumoto, T. (2017). "Development of an Automatic Detector of Cracks in Concrete Using Machine Learning." *Procedia Engineering*, 171, 1250-1255.

Ziou, D., and Tabbone, S. (1998). "Edge detection techniques-an overview." *Pattern Recognition and Image Analysis C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenii*, 8, 537-559.