

DAMAGE DETECTION AND LOCALIZATION IN MASONRY STRUCTURE USING FASTER REGION CONVOLUTIONAL NETWORKS

LuqmanAli¹, Wasif Khan¹ and *Krisada Chaiyasarn²

¹Department of Electrical and Computer Engineering, Thammasat University, Thailand

²Department of Civil Engineering, Thammasat University, Thailand

*Corresponding Author, Received: 27 Oct. 2018, Revised: 13 Dec. 2018, Accepted: 20 Jan. 2019

ABSTRACT: With current modern technology, manual on-site inspection can be assisted by automatic inspection, which is cost-effective, efficient and not subjective. In previous work, various image-based techniques have been applied to detect damages in heritage structures based on hand-designed feature extraction and classifiers. A heritage structure is composed of masonry walls, which are the components that are typically subjected to severe damages. This paper proposed a damage detection algorithm for a masonry structure based on Faster Region Convolutional Neural Networks (FRCNN). A labeled dataset for training the damage detection system in heritage masonry structure is created in this study, which is our first contribution as, currently, there is no public dataset available for masonry structures. The second contribution is the creation of a state of the art object detection system based on FRCNN for the detection and localization of damage in masonry structures. The results show that the proposed system performs well and can be used to detect damage in masonry structures with promising computational speed.

Keywords: Object Detection, Faster Region Convolutional Networks, Masonry Structure, Automatic Inspection.

1. INTRODUCTION

Thailand has many heritage sites which require extensive maintenance work. The preservation of cultural heritage is of the utmost importance because cultural heritage maintains the nation's history and transfers valuable knowledge and skills of their ancestors. Many historical structures in Thailand have been deteriorating due to aging and man-made activities. Due to various natural and human activities, many of the cultural heritage sites have been in seriously damaged conditions. Also due to the threat of a sudden changing environment (e.g. flooding) and unforeseen natural calamities (e.g. earthquake), many sites have been damaged. Figure. 1 shows Wat Chai Wattanaram temple dated back to 16th century, located in a historic province of Ayutthaya, the former capital of Thailand, various damages can be found in this temple which is caused due to aging, natural causes or manmade activities. The majority of conservation work has been carried out by the Fine Arts department since 1969. The country requires expertise and research to establish a methodology for historical structures maintenance.

Visual inspection is commonly used to examine the present conditions of historic structures. This method is usually carried out by trained inspectors to detect various defects in historical structures. Visual

inspection cannot be carried out regularly as the process is laborious, costly and easily prone to human error. Most of the sites are inaccessible to humans especially on the top of the stupa, which is very high. This structure requires regular inspection because failure to detect a defect on time may lead to structural failure. In order to overcome this problem, this paper proposes a state of the art object detection system based on Faster Region Convolutional Neural Network (FRCNN) for the detection and localization of damages in masonry structures. Figure. 2 shows example images of various defects found around the temple.

Current defect detection algorithm depends on handcrafted features which are not comprehensive and accurate. Handcrafted features are limited to researcher experience and design features can be a complex task. As images become more complex, it is hard to distinguish between the target and background regions. As image complexity increases, the detection performance decreases. Defect detection system based on handcrafted features mainly consists of two major steps. The first step is to extract relevant features by various feature extraction algorithms such as gray level co-occurrence matrix (GLCM) [1], edges [2], multi-features [3] and principal component analysis [4].



Fig. 1 Wat Chai Wattanaram, a temple located in the historic province of Ayutthaya.

In the second step, pattern recognition is done by classifiers such as Support Vector Machines (SVM) [8] and Neural Networks [5]. To avoid the feature extraction step, Convolutional Neural Network (CNN) is used in this paper for automatic feature extraction. CNN has shown an outstanding capability in image classification as exemplified in many previous works.

In this paper, Faster R-CNN is used for automatic defect detection in masonry structure in real time. The data has been collected using unmanned aerial vehicles (UAVs) from heritage masonry structures in Thailand. The proposed architecture can provide quasi-real time damage detection in videos. In this study, only one type of damage is considered, however, other defect types can be added to the same architecture. The rest of the paper is organized as follows. In Section 2, related work is discussed. In Section 3, the overview of the proposed method is discussed. In Section 4, the database generation and the implementation of the proposed system are shown. In Section 5, experimental results are discussed and the paper ends with a conclusion, future work, acknowledgment, and references in the last sections.

2. LITERATURE REVIEW

From previous work, it can be observed that the majority of damage detection systems are based on handcrafted features. Shahid Kabir et al. (2010) [6] used Gray Level Co-Occurrence Matrix (GLCM) features and Artificial Neural Network (ANN) to detect damages caused by alkali-aggregate reaction (AAR). Abdul- Qadir (2003) [7] worked on crack identification bridges using edge detection techniques.

The limitation of edge detection algorithms is due to noise. The edge detection techniques can perform better in noise-free images but do not perform well for real scene images contaminated with noise [26]. Nishikawa (2012) [9] used multiple sequential image filtering for crack detection in concrete structures. German et al. (2012) [10] implemented machine vision for the detection and properties measurement of concrete spalling for post-earthquake safety assessment. Yuen (2015) [11] proposed a vision-based automated crack detection system for the inspection of bridges. Cha (2016) [12] proposed a vision-based system for the detection of loosened bolts using Hough transform features and Support Vector Machine classifier. Zalama (2014) [13] used visual features extracted by Gabor filters for the damage detection in road pavement. Wu (2016) [14], Liao (2016) [15], Chen (2012) [16], and Mirzaei (2016) [17] have improved the accuracy of image-based techniques for damage detection in structural components but these methods still required post-processing and pre-processing steps which can be computationally expensive for the damage detection task.

To overcome the problems associated with feature extraction, automatic feature extraction based on learning techniques such as deep learning can perform efficiently compared to the techniques based on handcrafted features. Cha (2017) [18] proposed a CNN based crack detection system for cracks in concrete structures. Zhang (2016) [19] proposed a deep convolutional neural network system for road crack detection using images collected by a low-cost smartphone. The CNN has the capability to classify multiple classes. For damage localization, a sliding window technique can be applied. In the sliding

window technique, finding the optimal size of the window can be difficult as the size of images varies. To address this problem, Girshick (2014) [20] proposed an R-CNN system, in which object proposals and images are given as an input to the system. The R-CNN system uses CNN for automatic feature extraction and SVM classifier for localization and classification. The R-CNN can increase the accuracy of various recognition and detection tasks, but it was computationally slow. To overcome the issue, He (2014) [21] proposed a spatial pyramid pooling network (SPP-net). The speed of the proposed network was greater than R-CNN, but the implementation and training process of this network is as hard as training R-CNN. To address the issue associated with both SPP-net and R-CNN, Girshick 2015 [22] proposed Fast R-CNN. The proposed Fast R-CNN has better performance and speed than SPP-net and R-CNN, but its speed is slow because of the generation of object proposals from the method like a selective search. Ren (2016) [23] proposed faster R-CNN by combining Fast R-CNN with a region proposal network. Faster R-CNN provides a low computational cost with high accuracy.

3. PROPOSED METHODOLOGY

The outline of the damage detection and localization system is shown in Figure. 2 below. The proposed system consists of three main modules explained in details below; (1) image acquisition via a drone and DSLR camera, (2) the training of Fast R-CNN, (3) the testing of the proposed system. The output from the system images with localized damaged regions. The modification and implementation of Faster R-CNN architecture are discussed in this section.

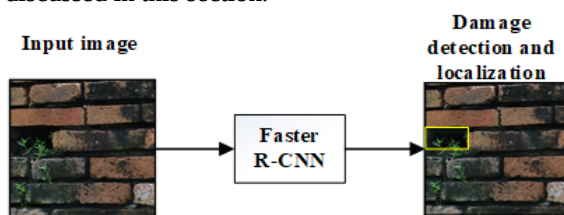


Fig. 2 The Outline of the proposed system

3.1 Image acquisition

The use of UAVs for civil inspection is an increasing trend and has replaced conventional surveying methods since they are cheaper, faster and simpler. In the proposed research work, a DSLR camera and a drone DJI phantom 4 were used for data acquisition. The full coverage of a masonry structure is obtained by flying a drone in a preplanned flight path strategy known as the Point of Interest (POI) strategy.

In the proposed strategy, a drone flew around an object of interest at a pre-determined radius and took pictures every 2-3 seconds interval to ensure a 50% overlap between consecutive images. The drone flight was programmed by Auto Flight Logic, which requires radius, velocity, altitude, and camera viewing angles as input. The drone path was planned carefully to ensure that it can perform the task in specified battery time. At the ground level, some close-up images were collected by researchers similar to drone motion around the structure. The close-up images were collected so that the features are more prevalent in images as shown in Figure. 3.



Fig. 3 Sample images of the structure taken by handheld DSLR camera.

3.2 Faster R-CNN

The schematic architecture of Faster R-CNN is shown in Figure. 4. In Faster R-CNN, the feature extraction computation of CNN is shared between region proposal network (RPN) and Fast R-CNN. The main advantage of Faster R-CNN is that a selective search used in Fast R-CNN is replaced by RPN, which makes Faster R-CNN fast. Secondly, convolutional maps are shared between RPN, a detection network which makes the RPN almost cost-free.

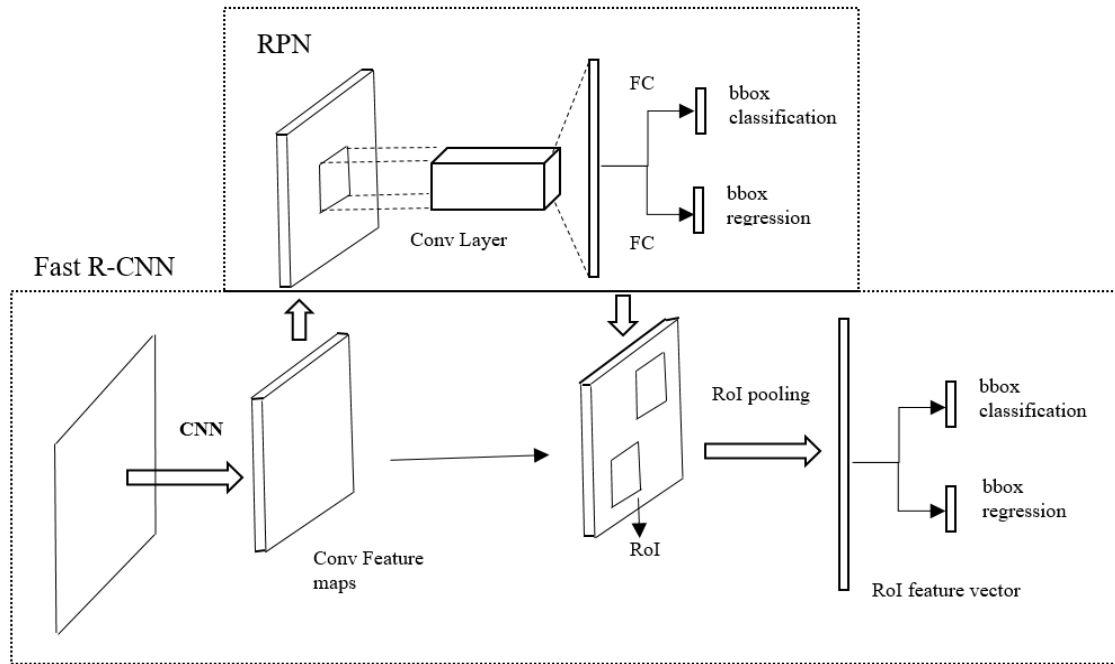


Fig. 4: The schematic architecture of Faster R-CNN (The top portion of the figure represents region proposal network while the down portion represent Fast R-CNN).

3.2.1 Convolutional Neural Network

Damage detection system using Faster R-CNN is divided into two main networks, Fast R-CNN and RPN. Both networks perform damage detection by sharing the same CNN architecture. A Convolutional Neural Network is a type of multilayer feedforward Artificial Neural Network which can solve many real-world problems. In the proposed work, ZF-net is used for training and testing as it is fast and has proven to be efficient in many real-time detection problems (Li 2016 [24]; Ren 2016 [23]). The ZF-net architecture is composed of 13 layers, Convolution layer, max pooling layer, and a softmax layer. In the proposed Faster R-CNN damage detection system, the architecture of ZF-net is modified for both Fast R-CNN and RPN. In the case of RPN, two layers are modified, the last max pooling layer and fully connected (FC) layer. These layers are replaced by sliding CONV and FC layer which has feature vectors with the depth of 256. Secondly, the softmax layer is modified with a softmax and regressors. For the ZF-net Fast R-CNN architecture modification, the last max-pooling layer is modified by Region of interest (RoI) pooling layer. The dropout layers with the threshold of 0.5 are sandwiched between the first and second FC layers in order to prevent the overfitting problem during the training process and similarly, the

dropout layers are inserted between the second and third FC layer in ZF-net. The softmax layer is modified to a softmax and regressors. The modification details of both the RPN and Fast R-CNN layer are shown in Table 1 and 2 below.

3.2.2 Region Proposal Network

A regional proposal network is a fully convolutional network (FCN) that outputs a set of object proposals as rectangular bounding boxes (box) from input images [25]. Firstly, the feature maps are obtained by using CNN in RPN. After obtaining the feature maps, a CONV layer followed by the ReLU activation function is slid on each pixel to obtain feature maps as shown in Figure. 5. Each sliding window features are mapped into a vector and is given to a softmax layer and regressors, which give the prediction of bounding boxes locations and the associated probabilities. These rectangular boxes are known as anchors. The anchors are usually nine in number and are found in the combination of various aspect ratios such as 1:1, 2:1 and 1:2. The three recommended dimensions of anchors are 128×128 , 256×256 and 512×512 . The obtained region proposal described by the prediction of rectangular bounding boxes and probabilities in each box is given to FaR-CNN for further refinement in order to improve the process of classification.

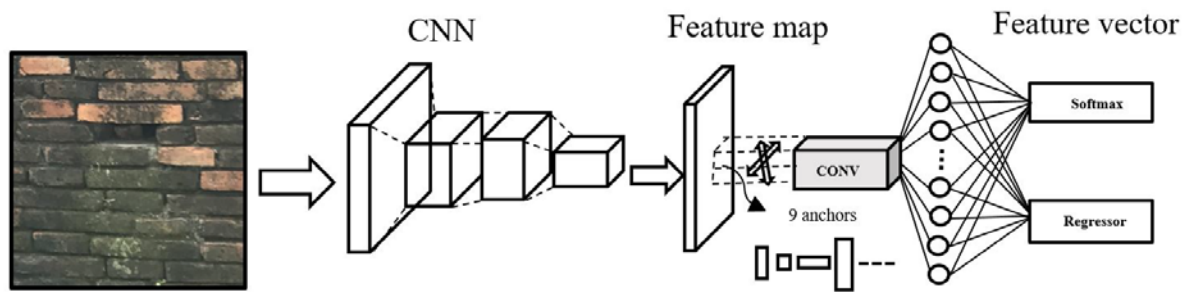


Fig. 5: The schematic architecture of a Region Proposal Network (RPN).

3.2.3 Fast R-CNN

The Fast R-CNN takes object proposal obtained from RPN as input and similar to RPN extracts feature maps by using CNN from the object proposal given as an input image. The obtained region proposal from RPN is overlaid on the obtained feature maps. The object proposal features are known as a Region of Interest (RoI). In RoI pooling, RoIs are given and the max-pooling operation is applied in order to obtain feature vector with a fixed size for each RoI. The vectors obtained are given to FC layers. The FC layers are followed by the softmax layer, and the regression softmax layer calculates the probability of damages in images while the regression layer shows the location and size of the anchors.

Table 1 RPN layer specifications

Layer	Name	Filter size	Depth	stride
1	Conv+ReLU	7×7	96	2
2	LRN	--	--	--
3	Max Pooling	3×3	96	2
4	Conv+ReLU	5×5	256	2
5	LRN	--	--	--
6	Max Pooling	3×3	256	2
7	Conv+ReLU	3×3	384	1
8	Conv+ReLU	3×3	384	1
9	Conv+ReLU	3×3	256	1
10	Sliding Conv+ReLU	3×3	256	1
11	FC	--	256	--
12	Softmax & Regressor	--	--	--

3.2.4 Network Training

The training process of the damage detection system for masonry structures is composed of four main steps. Various parameters related to the networks used are finely tuned. The process of training starts with RPN. The RPN is trained with pre-trained CNN initialized weights and object proposals are generated for Fast R-CNN. The second step in the training is the initialization of Fast R-CNN with the pre-trained weight obtained from the object proposal generation step. In step 3, RPN is again trained with the weights obtained from the previous step, object proposals are generated again. In step 4, the Fast R-CNN takes the generated object proposal and is trained with the initial parameters obtained from step 3.

Table 2 Fast R-CNN layers specifications

Layer	Name	Filter size	Depth	stride
1	Conv+ReLU	7×7	96	2
2	LRN	--	--	--
3	Max Pooling	3×3	96	2
4	Conv+ReLU	5×5	256	2
5	LRN	--	--	--
6	Max Pooling	3×3	256	2
7	Conv+ReLU	3×3	384	1
8	Conv+ReLU	3×3	384	1
9	Conv+ReLU	3×3	256	1
10	RoI Pooling	--	256	--
11	FC+ReLU	--	4096	--
12	Dropout	--	--	--
13	FC+ReLU	--	4096	--
14	Dropout	--	--	--
15	FC+ReLU	--	6	--
16	Softmax & Regressor	--	--	--

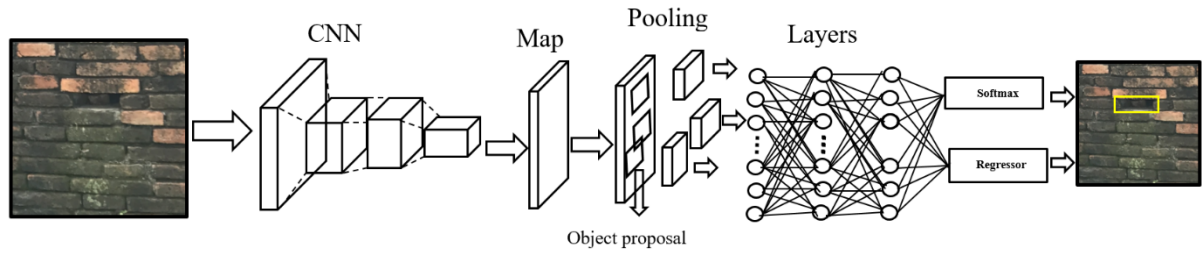


Fig. 6 The schematic architecture of Fast R-CNN

4. EXPERIMENTS AND RESULTS

4.1 Database Creation

For the creation of training and testing dataset, 466 images were taken. In these images, 163 images with dimensions of 4864×3468 are taken by a DJI Phantom 4 drone as shown in Figure 1. The remaining 303 images with a dimension of 3024×4032 are taken by an iPhone 7 camera. These images contain structural damages. The images are converted into patches of 500×500 . Out of the total patches, only 1000 patches are selected which contains damages as shown in Figure. 7 below.



Fig. 7 Sample images of the patches created for training, testing, and validation.

4.1 Implementation details

The experiment in the proposed research work is implemented by using open source library of Faster R-CNN [23], the python tensor flow, CUDA 8.0, CNN 5.1 using a core i7-7820X 3.6 GHz CPU, 64 GB DDR4 memory, and a 4GB NVIDIA GeForce GTX 1070 GPU. The RPN and Fast R-CNN learning rate is taken 0.001 while the momentum is taken 0.9. For

80,000 and 40,000 iterations, the weight decay is taken 0.0005. Average precision is used as an evaluation criterion for the damage detection system.

4.1 Training, validation and Testing Results

The network was trained by using the four-step procedure explained above. In the training process, the GPU mode took approximately 4 hours while the CPU modes took almost 1.5 days. The evaluation time for 500×500 -dimension patches in GPU mode was 0.040 sec, while in CPU mode it took almost 1 sec. For testing new images, the trained model with the mean average precision of 96.50% is selected and a new set of test images is given to the damage detection system for testing. The testing result showed some minor errors, due to lighting conditions in various patches. However, the overall performance of the proposed damage detection is promising and can be used in real time for damage detection in masonry structures. The results of the testing data are shown in Figure. 8. The training loss graph of the proposed system is shown in Figure. 9. The training loss graph shows that the loss decreases from value 5. In the graph, fluctuations occur after 8,000 iterations, and then the graph stabilizes after 70,000 iterations.

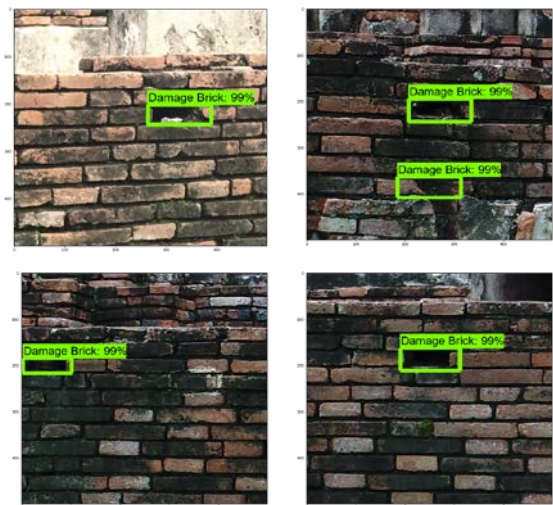


Fig. 8 The results of the testing data of the proposed system

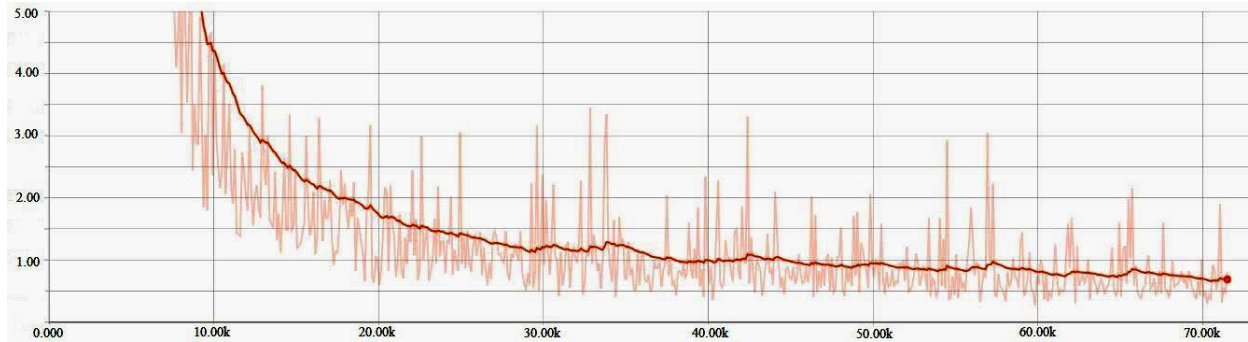


Fig. 9 The training loss graph of the proposed system

5. CONCLUSION AND DISCUSSIONS

From the experiments, it can be concluded that Faster R-CNN technique offers the possibility of automatic defect detection for masonry structures. In this work, Faster R-CNN is implemented to automatically detect and localize the defects in images of masonry structures. The system offers better resistance to noise and does not require handcrafted features as it automatically extracts features. Once a CNN based network is trained, it can accurately detect damaged regions in images, which is more efficient and cheaper than manual inspection. The accuracy of the proposed system can be further improved by using a larger training dataset. It can also be concluded that defect detection in masonry structures is difficult because of similar appearance to grout lines. Therefore, it is difficult to create good datasets as this type of scenario, i.e. the masonry surface is complex and confusing, even for human inspectors themselves. Nevertheless, good datasets are still required, however, in this work, a dataset of 1000 images have been created. The system is evaluated on the dataset and it can be concluded that a deep learning algorithm like Faster R-CNN can be trained, and with a sufficient amount of training data, and the system space can learn how to detect damages automatically.

6. FUTURE WORK

In the proposed research work, only one case of structural damages is studied. In the future, the focus is to implement damage detection system for multiple types of defects in masonry structures. Also, the research team is working on the modification of architecture of Faster R-CNN to achieve an accelerated object detection system for Civil inspection.

7. ACKNOWLEDGMENT

The authors would like to thank the grant from the Research Division, Department of Engineering,

Thammasat University to allow this research to be successfully completed.

6. REFERENCES

- [1] Hu, Y., Zhao, C. X., & Wang, H. N. (2010). Automatic pavement crack detection using texture and shape descriptors. *IETE Technical Review*, 27(5), 398-405.
- [2] Abdel-Qader, I., Abudayyeh, O., & Kelly, M. E. (2003). Analysis of edge-detection techniques for crack identification in bridges. *Journal of Computing in Civil Engineering*, 17(4), 255-263.
- [3] 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 Transactions on automation science and engineering*, 13(2), 591-599.
- [4] Abdel-Qader, I., Pashaie-Rad, S., Abudayyeh, O., & Yehia, S. (2006). PCA-based algorithm for unsupervised bridge crack detection. *Advances in Engineering Software*, 37(12), 771-778.
- [5] Kaseko, M. S., & Ritchie, S. G. (1993). A neural network-based methodology for pavement crack detection and classification. *Transportation Research Part C: Emerging Technologies*, 1(4), 275-291.
- [6] Kabir, S., Rivard, P., & Ballivy, G. (2006). Detection of AAR deterioration patterns in concrete using wavelets for multiscale texture analysis. *Special Publication*, 234, 127-146.
- [7] Abdel-Qader, I., Abudayyeh, O., & Kelly, M. E. (2003). Analysis of edge-detection techniques for crack identification in bridges. *Journal of Computing in Civil Engineering*, 17(4), 255-263.
- [8] Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, 9(3), 293-300.
- [9] Nishikawa, T., Yoshida, J., Sugiyama, T., & Fujino, Y. (2012). Concrete crack detection by multiple sequential image filtering. *Computer-Aided Civil and Infrastructure Engineering*, 27(1), 29-47.
- [10] German, S., Brilakis, I., & DesRoches, R. (2012). Rapid entropy-based detection and properties

- measurement of concrete spalling with machine vision for post-earthquake safety assessments. *Advanced Engineering Informatics*, 26(4), 846-858.
- [11] Yeum, C. M., & Dyke, S. J. (2015). Vision-based automated crack detection for bridge inspection. *Computer-Aided Civil and Infrastructure Engineering*, 30(10), 759-770.
- [12] Cha, Y.-J., You, K. & Choi, W. (2016), Vision-based detection of loosened bolts using the Hough transform and support vector machines, *Automation in Construction*, 71(2), 181– 88.
- [13] Zalama, E., Gómez-García-Bermejo, J., Medina, R., & Llamas, J. (2014). Road crack detection using visual features extracted by Gabor filters. *Computer-Aided Civil and Infrastructure Engineering*, 29(5), 342-358.
- [14] Wu, L., Mokhtari, S., Nazef, A., Nam, B., & Yun, H. B. (2014). Improvement of crack-detection accuracy using a novel crack defragmentation technique in image-based road assessment. *Journal of Computing in Civil Engineering*, 30(1), 04014118.
- [15] Liao, K. W., & Lee, Y. T. (2016). Detection of rust defects on steel bridge coatings via digital image recognition. *Automation in Construction*, 71, 294-306.
- [16] Chen, P. H., Shen, H. K., Lei, C. Y., & Chang, L. M. (2012). Support-vector-machine-based method for automated steel bridge rust assessment. *Automation in Construction*, 23, 9-19.
- [17] Mirzaei, G., Adeli, A., & Adeli, H. (2016). Imaging and machine learning techniques for diagnosis of Alzheimer's disease. *Reviews in the Neurosciences*, 27(8), 857-870.
- [18] 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.
- [19] Zhang, L., Yang, F., Zhang, Y. D., & Zhu, Y. J. (2016, September). Road crack detection using the deep convolutional neural network. In *Image Processing (ICIP), 2016 IEEE International Conference on* (pp. 3708-3712). IEEE.
- [20] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).
- [21] He, K., Zhang, X., Ren, S., & Sun, J. (2014, September). Spatial pyramid pooling in deep convolutional networks for visual recognition. In *European conference on computer vision* (pp. 346-361). Springer, Cham.
- [22] Girshick, R. (2015), Fast R-CNN, in *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, 07–13 December 2015, 1440–48.
- [23] Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6), 1137-1149.
- [24] Li, C., Kang, Q., Ge, G., Song, Q., Lu, H. & Cheng, J. (2016), Deep: learning deep binary encoding for multilabel classification, in *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Las Vegas, NV, 26 June–1 July 2016, 744–51.
- [25] Suh, G., & Cha, Y. J. (2018, March). Deep faster R-CNN-based automated detection and localization of multiple types of damage. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018* (Vol. 10598, p. 105980T). International Society for Optics and Photonics.
- [26] Bennisamoun, M. (1997, October). Edge detection: Problems and solutions. In *Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation. 1997 IEEE International Conference on* ((Vol. 4, pp. 3164-3169). IEEE.