Deep Learning-based Defect Detection and Assessment for Engineering Structures

Zheng Yi WU [1] and Rony KALFARISI [1]

[1]Bentley Systems, Incorporated, Watertown, USA
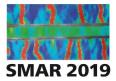
Contact e-mail: zheng.wu@bentley.com

ABSTRACT: Structure inspection is essential to ensure the safety and integrity of critical civil infrastructures. The current practice is human-based visual inspection, which is time consuming and subjective. To improve the inspection of civil infrastructure and complement human-oriented inspection, computer vision-based techniques have been employed to detect structural defects, such as cracks and corrosions. In this paper, to provide an effective and useful defect (cracks and corrosions) identification (detection and segmentation), deep learning-based approaches are developed as an integrated framework for defect detection, segmentation, assessment and visualization using 3D reality mesh modeling technology. The robustness of these techniques is evaluated and demonstrated using real cases of the bridge structures, road pavement, and tunnel inspection. The results obtained show that the proposed framework is effective and readily applicable to various engineering structures.

1   INSTRODUCTION

Civil infrastructures, such as bridges, dams, roads, and highways, are becoming more susceptible to lose their designed functions as they deteriorate over time. These structures, are often subjected to fatigue stress and periodic loading which leads to cracks on the structures' surface. The cracks on the structure reduce the local stiffness and cause material discontinuities (Budiansky & O'connel, 1976; Aboudi, 1987). In the meantime, the study conducted by National Association of Corrosion Engineering shows that the total direct cost of corrosion in five major sectors of US economy is more than $140 billion. Corrosions in infrastructure alone cost $22.6 billion, which is 16.4% of the total annual cost. Among them, $8.3 billion is the estimated cost for highways and bridges. This indicates the importance of detecting defects such as cracks and corrosions. To do so, regular inspection is essential. It is usually conducted by taking photos and videos. Due to large number of images collected during the inspection, it is imperative to effectively and efficiently process the images for defect evaluation.

This inevitable process has prompted an early detections as preventive measures to avoid further damage and possible failures. Manual inspection is still the standard practice for any damage or crack detection. In manual inspection, the process is performed by human inspector who prepare the sketch of the crack damage by recording any irregular conditions on the structures. Since this approach completely depends on the inspector's expertise, it lacks objectivity in the quantitative analysis. Therefore, many researches have been conducted to replace the process with automatic vision-based crack detection.

Due to its simplicity, several vision-based methods for detecting crack damages, primarily using image processing techniques (IPTs), have been proposed. An early comparative study was done

by Abdul-Qader et. al (Abdul-Qader, Abudayyeh, & Kelly, 2003) in which they used four edge detection methods to detect crack damages on bridges. Unfortunately, using edge detection to detect crack damages is an ill-posed problem since the results are significantly affected by the image noises such as irregular illumination and shading. IPT methods have also been applied to corrosion detection (Ortiz et al. 2016). Recently, many researchers have implemented a combination of IPT-based feature extractions and machine learning based classifications (Moon & Kim, 2011; O'Byrne, Schoefs, Ghosh, & Pakrashi, 2013; Jahanshahi, Masri, Padgett, & Sukhatme, 2013; Wu, Mokhtari, Nazef, Nam, & Yun, 2014). However, the results of these approaches still suffer from the false feature extraction of IPTs since the extracted features using IPTs are still considered hand-crafted and does not necessarily represent the true characteristics of crack damages.
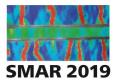
To overcome this issue, convolutional neural networks (CNN) have been adopted and implemented for an effective crack detection (Cha, Choi, & Buyukozturk, 2017; Zhang, Cheng, & Zhang, 2018) as well as corrosion detection (Hoshere et al. 2017; Atha and Jahanshahi 2018). Although the previous applications of CNN prove to be effective for crack detection with bounding box, the detected crack damages are neither segmented nor statistically evaluated for condition assessment. Mask region-based CNN (Mask R-CNN) (He, Gkioxari, Dollar, Girshick, & Ross, 2017), a state-of-the-art image segmentation with relatively fast detection speed, is found to be an effective method for this task. Mask R-CNN can generate mask image for every detected object. In this paper, Mask R-CNN is applied to detect and segment crack damages for civil infrastructures, together with 3D mesh and texture model for intuitive visualization and crack evaluation.

2 DEEP LEARNING

Deep learning is one of machine learning techniques. It is multi-layered artificial neural networks which can learn and represent data using multiple levels of abstraction. Deep learning models are different from traditional machine learning techniques in that they can learn the representations of the data without introducing any hand-crafted rules or knowledge. This means that deep learning technique is highly flexible and effective for solving many challenging problems, including but not limited to image processing, medicine, biometrics, and engineering (Vargas, Mosavi, & Ruiz, 2017). One of deep learning method is Mask R-CNN which achieves state-of-the-art performance for object detection and instance segmentation. This method is directly applicable and appropriate for crack damage detection and segmentation.

2.1 Mask R-CNN Architecture

Mask R-CNN is built and improved upon Faster R-CNN. The original Faster R-CNN has two outputs for each candidate object, a class label and a bounding box offset. Mask R-CNN adds a third branch that outputs the object mask giving a much finer spatial layout of an object. As a result, Mask R-CNN can perform object detection and segmentation with better efficiency and accuracy. Mask R-CNN has the same two-stage detection process as Faster R-CNN, as seen in Fig. 1. For the first stage, region proposal network (RPN), outputs a set of potential regions of interest (RoI) with certain aspect ratios (anchors) and evaluate how good each of these anchors are. These RoIs will become the input for the second stage. In the second stage, the algorithm

will output bounding box offset, the class name, and binary mask for each object in parallel. The loss function is defined for the three tasks on each sampled RoI, class prediction, bounding box refinement, and mask generation. Class prediction and bounding box regression loss are collected from both RPN and mask generation stages, whereas mask loss is taken only from mask generation stage.
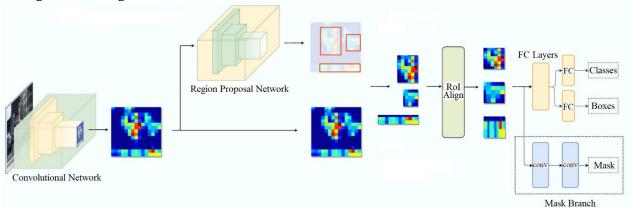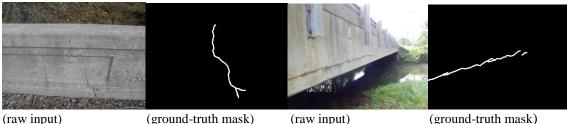

Fig. 1: Mask R-CNN architecture for instance segmentation (Hui, 2018)

3 TRAINING MASK R-CNN

The image dataset was acquired from a provider of field inspection applications and asset management services for bridges. In this dataset, all the images were captured by the field inspector using regular camera and taken without any prior conditioning. 1250 images of bridges and structures were selected where crack damages can be seen visually. The dataset was split into two categories, 1000 images are used for training and 250 images are for testing. Training and testing samples were annotated manually using image labelling tool for creating bounding box and Microsoft Paint for generating crack mask labels. Fig 2 shows some examples for the image and the mask label in the datasets.
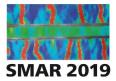

(raw input)   (ground-truth mask)   (raw input)   (ground-truth mask)
Fig. 2: Some examples of the image in the crack damage datasets.

3.1 Transfer Learning

Deep convolutional neural network usually contains millions of trainable parameters to train; therefore, a relatively large dataset is needed. Directly training such a huge number of parameters is problematic, especially when the training dataset is small. Transfer learning provides a way to improve one model by transferring the knowledge of another model that has been trained previously. In transfer learning, a base network is first trained on a base dataset and
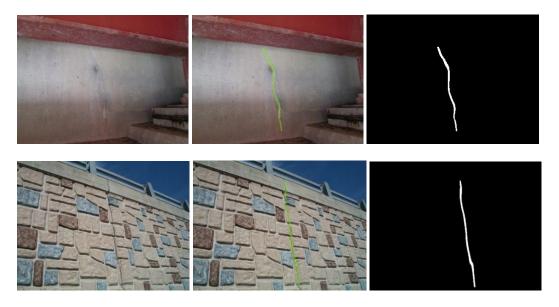
task, and then the learned features were transferred to a second target network to be re-trained on the target dataset and task. The improvement is extremely significant and that's the reason the technique is adopted in this paper.

To fit into GPU memory, all images were resized to a maximum dimension of 1024 pixel (either width or height) while retaining the original aspect ratio. Stochastic gradient descent (SGD) was performed for 30,000 iterations with initial learning rate of 0.003 and momentum of 0.9. Additional augmentation was also done to each individual image. This includes image rotations by 90°, flipping (vertical and horizontal), scaling, and shifting. The Mask R-CNN code is based on the open-source implementation of TensorFlow object detection API and is available online (Huang, et al., 2017). Training was done using pre-trained model on COCO dataset and then the model was fine-tuned on crack damage dataset. Two base networks, ResNet101 and Inception-Resnet, were used to train the model. The calculation was performed on Amazon EC2 instances.

4   RESULT

Some examples of the crack damage segmentation results on bridges and structures can be seen in Fig. 3. The experimental results demonstrate the effectiveness and advantage of Mask R-CNN in detecting crack damage in a complex scenery.

(input image)   (inferenced image)   (inferenced mask)
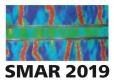Fig. 3: Sample results of crack detection and segmentation.

Figure 4. Sample results of corrosion detection and segmentation

4.1 Evaluation Metrics

Discrete scores were used as the criteria for quantitative comparison for the benchmark and one common approach to evaluate instance segmentation performance is to use average precision (AP). Each detected mask is considered a true positive if its intersection-over-union (IoU) ratio with its ground-truth annotation is greater than some IoU threshold. The value of IoU threshold is referenced from MS COCO dataset evaluation metrics with fixed IoU threshold at 0.5 ($AP_{0.50}$) and averaging AP over [0.50, 0.95] with an increment of 0.05 (AP). There is also AP_{medium} for objects that has $32^2 <$ area $< 96^2$ and AP_{large} for objects that has area $> 96^2$.
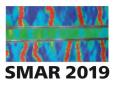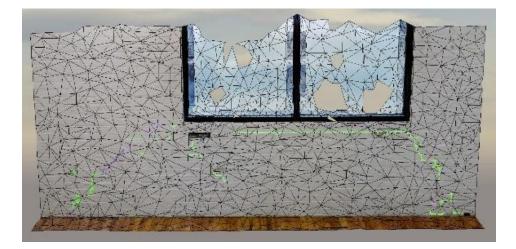
Table 1: Evaluation result of the trained models.

	AP	$AP_{0.50}$	AP_{medium}	AP_{large}
Resnet101	0.0065	0.048	0.0010	0.025
Inception-Resnet	0.0070	0.050	0.0013	0.050

The result shows high confidence in predicting bounding boxes for crack damage objects. However, the value of average precision for mask image seems rather low. This can be attributed to multiple detections of the same crack damage objects in an image and, unfortunately, this happens quite frequently. The reason for this is because there is no definite shape and orientation for a crack damage object making it impossible to get a consistent label in creating the ground truth for the masking crack damage. As a result, the network seems to confuse itself on how to determine the number of crack object in the image. This issue may be overcome if there are more images in the training dataset.
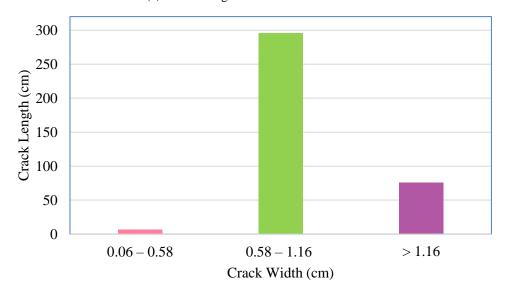
In evaluating the crack damage statistics, reality modeling technology (Bentley 2018) was used to construct the 3D model with crack-detected and segmented images. For every connected crack in the inferenced mask images, the area of the crack is first measured by calculating the number of pixels in the inferenced mask image. The length of the crack was obtained by applying thinning algorithm iteratively until the shape of the crack only shows one-pixel width. Once this is achieved, the average crack width is estimated with the area divided by the length. Finally, all the quantified cracks are classified in different categories or levels according to the estimated width, and the crack statistics can be generated for the assessment of infrastructure condition.

(a) Cracks segmented with 3D mesh model

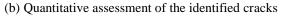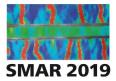(b) Quantitative assessment of the identified cracks

Fig. 4: Visualization of detected cracks with 3D models and crack evaluation.

The result is illustrated in the Fig. 4, where a set of images of crack damages on the wall was taken and using the 3D mesh model, a fused image was reconstructed. The inferenced segmented crack damage was color coded based on the width level of the crack. To quantify the cracks (width and length) in distance unit, a conversion factor must be worked out to convert the number of pixels to metric unit. The inner side of the glass window was measured to be 108.60 cm wide on the building wall and covered by 1880 pixels in the image. Therefore, a conversion factor of 0.06 cm/pixels is obtained. Using this scale factor, each crack segment can be quantified for its average width and length. Thus, the crack statistics is conducted and the result, as illustrated in Fig. 4, is color coded based on the width level of the crack. Purple color is crack area with average width greater than 1.16 cm, green color is crack area with average width 0.58 – 1.16 cm, while pink color is crack area with average width of 0.06 – 0.58 cm.

5   CONCLUSIONS

In this study, deep learning-based approach has been developed by using Mask R-CNN to automatically detect and segment crack and corrosion for civil structures. Two base networks were trained using the bridge inspection images, which were split into training and testing datasets. The results show that defects can be detected and seghmented with a good level of precision. Quantitative evalution on seghmented cracks was propsoed by using 3D reality model technology. This research study shows that the detection and quantification of crack and corrosion using deep learning methods is not only promising but practically usefulfor civil structural inspection with satisfactory results.

6   REFERENCES

Abdul-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*(3), 255-263.

Abdul-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*(3), 255-263

Aboudi, J. (1987). Stiffness reduction of cracked solids. *Eng. Fract. Mech. , 26*(5), 637-650.

Atha, D.J. and Jahanshahi, M.R. (2018), "Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection," *Structural Health Monitoring*, DOI: 10.1177/1475921717737051

Bentley Systems, Inc (2018). ContextCapture [Computer Software]. Exton, PA

Budiansky, B., & O'connel, R. J. (1976). Elastic moduli of a cracked solid. *Int. J. Solids Struct., 12*(2), 81-87.

Cha, Y.-J., Choi, W., & Buyukozturk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 361-378.

He, K., Gkioxari, G., Dollar, P., Girshick, & Ross. (2017). Mask R-CNN. *arXiv:1703.06870*.

Hoskere, V., Narazaki, Y. Hoang, T.A., Spencer Jr., B.F. (2017). Vision-based Structural Inspection using Multiscale Deep Convolutional Neural networks. in Proc. of the 3rd Huixian International Forum on Earthquake Engineering for Young Researchers August 11-12, 2017, University of Illinois, Urbana-Champaign, United States

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., . . . Murphy, K. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. *CVPR*.

Hui, J. (2018, April). *Medium*. Retrieved 2018, from https://medium.com/@jonathan_hui/image-segmentation-with-mask-r-cnn-ebe6d793272

Jahanshahi, M. R., Masri, S. F., Padgett, C. W., & Sukhatme, G. S. (2013). An innovative methodology for detection and quantification of cracks through incorporation of depth perception. *Machine Vision and Applications, 24*(2), 227-241.

Moon, H., & Kim, J. (2011). Intelligent crack detecting algorithm on the concrete crack image using neural network. *Preceedings of the 28th ISARC*.

O'Byrne, M., Schoefs, F., Ghosh, B., & Pakrashi, V. (2013). Texture analysis based damage detection of ageing infrastructural elements. *Computer-aided Civil and Infrastructure Engineering, 28*(3), 162-177.

Ortiz, A. , Bonnin-Pascual, F., Garcia-Fidalgo, E. and Company-Corcoles, J. P. (2016). Vision-Based Corrosion Detection Assisted by a Micro-Aerial Vehicle in a Vessel Inspection Application. Sensors, **2016**, 16, 2118; doi:10.3390/s16122118.

Vargas, R., Mosavi, A., & Ruiz, L. (2017). Deep learning: a review. *Advances in Intelligent Systems and Computing*.

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 Engineering, 30*(1), 040141181-040141189.

Zhang, K., Cheng, H. D., & Zhang, B. (2018). Unified approach to pavement crack and sealed crack detection using preclassification based on transfer learning. *Journal of Computing in Civil Engineering*, 04018001-1 - 04018001-11.