ASM Sc. J., 13, Special Issue 2, 2020 for ICSCC2019, 74-82

Automated Building Crack Identification using Enhanced Mask R-CNN

Wei Qin Lee[1]*, King Hann Lim[1], Chin Hong Lim[2], Wen Loong Lim[2] and Huei Ee Yap[3]

[1]*Department of Electrical and Computer Engineering, Curtin University Malaysia, Miri Sarawak*

[2]*SAFE-T5 (M) Sdn. Bhd., Selangor, Malaysia.*

[3]*LP-Research Inc., Tokyo, Japan.*

Automated building crack identification is highly required in the field of civil and construction to increase the working efficiency of building surveyors. The current image processing techniques are difficult to capture visual representation of cracks in a complex background due to the irregular crack patterns and size. Hence, deep learning neural network using Mask Region-Convolutional Neural Network (Mask R-CNN) is proposed in this paper to identify cracks in a complex background. Mask R-CNN can detect the wall cracks with precise instance segmentation and high classification accuracy. In the experiment, Mask R-CNN can achieve mAP_{50} of 50.0% for crack identification. Mask R-CNN is compared with the state-of-the-art object detection method, You Only Look Once version 3 (YOLOv3), which achieves only mAP_{50} of 28.7%. Subsequently, Sobel filters are added in the Mask R-CNN process to enhance the performance of crack identification. The effect of Sobel filter improves the mask of Mask R-CNN by archieving mAP_{50} of 63.3% with its guiding effect along the edges.

Keywords: crack identification; mask region-convolutional neural network (Mask R-CNN); you only look once version 3 (YOLOv3); sobel filter

I. INTRODUCTION

Building structures safety and reliability are important in the cities with high-rise buildings due to the gradual structural deterioration. The appearance of cracks on buildings could further increase the rate of deterioration and reduce the lifespan of structural integrity under the uncrontrolled environmental circumstances (Suresh *et al.*, 2004). Three types of common cracking distress can be categorized, i.e. longitudinal, transverse and alligator cracks (Saar and Talvik, 2010, Wang *et al.*, 2004). The location and extent of the cracks are important to be known in order to determine the damaging condition of a building. Local stiffness of building structures may be drastically reduced due to the cracks appearance that causes the material discontinuities (Aboudi, 1987, Budiansky and O'Connell, 1976). Serious cracks may contribute to the structure failure if the structures are not maintained on regular basis. Therefore, an early crack identification is required to allow precautionary measurement to avoid critical damage and failure to be occurred (Dhital and Lee, 2012).

Crack identification refers to a process to detect and locate cracks on structures using either manual visual inspection or computerized auto-detection. In conventional approach, building structural surveys are inspected manually by professional surveyors to collect data through visually observation of the cracks on site. However, manual building survey is time consuming especially in the visual inspection of mega structures (Prasanna *et al.*, 2016). During surveying, the outline of the cracks are prepared manually by specialists, and the cracks condition are noted on a paper. As this is completely dependent on the specialist's knowledge and experience, it is hard to assess the building structure deterioration objectively using visual inspection (Fujita and Hamamoto, 2011).

Due to the current advancement of technology and algorithms, automated building crack identification has recently received great attention in the field of civil and construction to replace the conventional manual surveying method (Broberg, 2013). The accuracy in automated cracks identification is a challenging task due to the irregular cracks pattern, size and complexity. The conventional image

*Corresponding author's e-mail: weiqin94@hotmail.com

processing techniques based on image pixels often faced the difficutly in thresholding value determination and filter size selection (McCann *et al.*, 2017). Hence, this issue has led to the difficulty of constructing a good internal visual representations of cracks (Jain *et al.*, 2014).

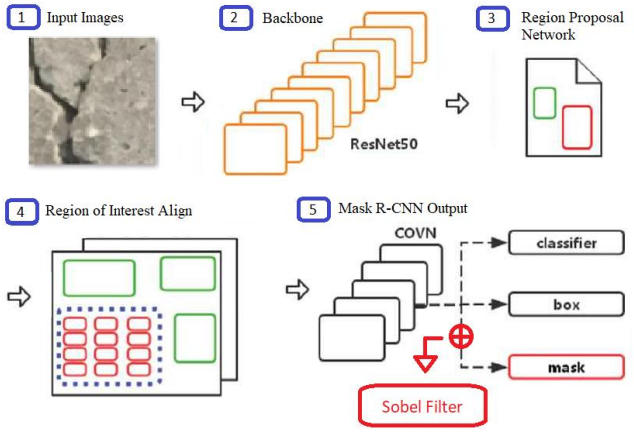


Figure 1. Framework of the enhanced Mask R-CNN with Sobel filter for building crack detection

With the consideration of pose, scale, conformation, clutter and illumination, these image processing techniques encounter massive challenges to correctly identify cracks (LeCun *et al.*, 2010). Deep learning neural network is recently adopted to overcome the limitation of the conventional image processing techniques due to its non-linearity input-output mapping relationship and self-adaption learning characteristics (Haykin and Network, 2004). However, the effectiveness of deep learning neural network is highly dependent on the amount of quantity or quality of the training data that is fed in to the network (Ding *et al.*, 2017). In most of real world applications, training data is often costly due to the difficulty of data collection (Pan and Yang, 2010). In addition, data labeling is another great labor efforts for deep learning neural networks training (Zitnick and Dollár, 2014). Thus, it is common to deal with limited resources condition to train an effective deep learning model.

In this paper, Mask Regional-Convolutional Neural Network (Mask R-CNN) is applied to perform crack identification and instance segmentation. To further improve the performance of Mask R-CNN, the Edge Agreement Head (Zimmermann and Siems, 2019) is participated with Sobel filter in modifying the output of the mask branch to guide the training process. Sobel filter could detect the significant edges and cracks direction, and subsequently offer an auxiliary task for Mask R-CNN to segment the cracks instances.

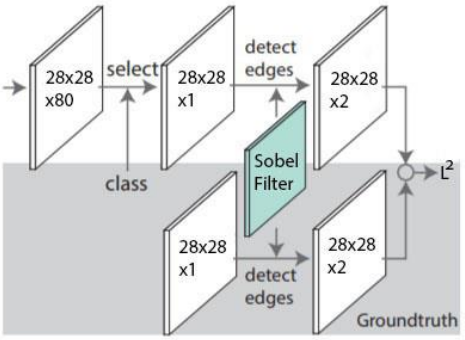
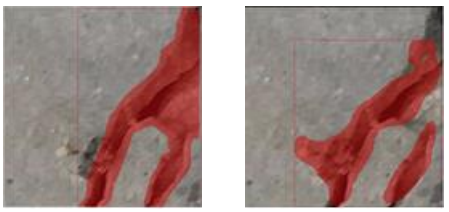


Figure 3. Edge Agreement Head: Extended mask branch architecture of Mask R-CNN. After the mask output with 28×28×80 dimensional, mask will assigned to the correct class. Then, it will now pass through a 3×3×2 Sobel filter and the same applied to the ground truth mask. Both output was then compare and calculate with L^2 norm loss function which is then results to edge loss L_{Edge} (Zimmermann and Siems, 2019)



(a) (b)

Figure 2. (a) Ground-truth Mask and (b)Predicted Mask with oversegmented region

With the modified Edge Agreement Head, the proposed Mask R-CNN is able to learn faster and hence decrease the desired amount of training data to achieve better results provided by the guiding effect of Sobel filter. The segmented output mask of the enhanced Mask R-CNN is compared with the ground truth mask to measure the accuracy.

II. ENHANCED MASK R-CNN WITH SOBEL FILTERING HEAD

A. Mask R-CNN

Mask R-CNN (He *et al.*, 2017) has the same general architecture with the Faster R-CNN (Ren *et al.*, 2017) except that the region of interest pooling module of Faster R-CNN is replaced with the region of interest align module. An additional branch known as Regional Proposal Network (RPN) is connected toward region of interest (RoI) align for generating mask as illustrated in Figure 1. Mask R-CNN is a two stages framework for instance segmentation. In first stage, the region that contains object with high probability

will be selected as RoI align. Subsequently, the region of interest will filtered by RPN nd fed into three parallel branches of the network, i.e. edge agreement head along the mask predictor branch with sobel filtering, bounding box offset regressor, and softmax classifier.

B. Edge Agreement Head with Sobel Filtering

In the observation towards instance segmentation of Mask R-CNN, the masking does not produce accurate predicted output as compared with the real object boundary. The output of predicted mask may be over segmented or missing segmentation or both existed as the worst case scenario as illustrated in Figure 2 as compared to the ground truth mask. As observed from the predicted output, it may have some part missing of segmentation and some other part was over segmented which indicated the worst case scenario. To boost the performance to avoid the mistakes such over segmented or missing of segmentation, Sobel operator was then applied for edge detection and it is effective for crack detection (Talab *et al.*, 2016, Jain *et al.*, 2014).

The core idea of improvement is based on auxiliary task added to existing mask branch architecture as illustrated in Figure 3. input for the architecture is from the predicted mask and ground-truth mask. To note that the predicted mask was selected with the mask corresponding to correct class. Both input are then perform convolution with 3×3×2 Sobel kernel. The filtered output is then compared and calculate with L^2 norm loss function which is then results to edge loss L_{Edge}. L^2 norm loss function also defined as loss between the edges of ground-truth mask and predicted mask in our experiment. Therefore, the total loss function of Mask R-CNN is updated as follows:

$$L_{MRCNN} = L_{Class}+L_{BBox}+L_{Mask}+L_{Edge} \quad (1)$$

III. EXPERIMENTAL SETUP

Crack data was recorded from the surrounding of Middle East Technical University Campus Buildings (Özgenel and Sorguc, 2018) and the dataset was published at the Mendeley Library, which was named as "Concrete Crack Images for Classification". The dataset contains multiple images with and without cracks. The sample images are illustrated in Figure 4. In the experiments, three classes, i.e. longitudinal, transverse and crocodile crack contains 250 images respectively. Within these 250 images per class, 200 images was used for training and 50 images was used for validation. In total, there were 600 training images and 150 validation images for three classes. There will also have 10 images chosen from the dataset additionaly to each class for testing purpose.

Mean average precision (mAP) is the measurement metric in the experiment, which is similar to PASCAL Visual Object Classes (VOC) 2007 challenge (Everingham *et al.*, 2010) . This measurement with one intersection over union threshold will be considered and have mean value over three object classes which longitudinal crack, transverse crack, and crocodile crack

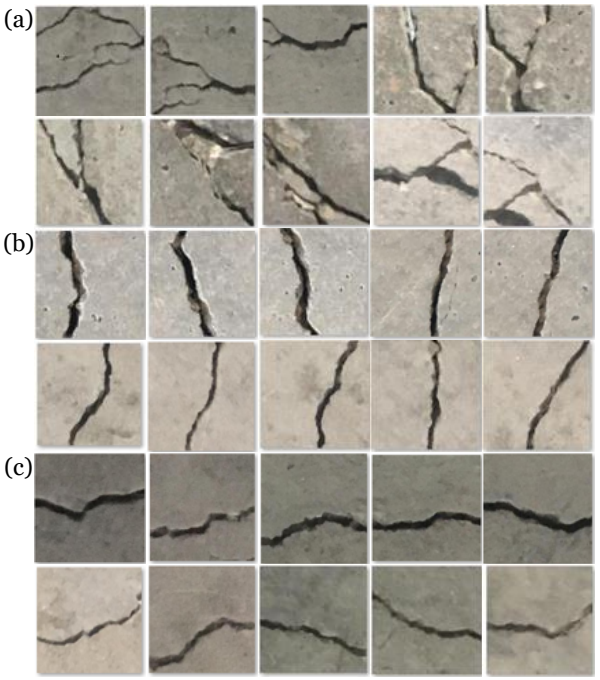


Figure 4: Mendeley Library Dataset for (a) Crocodile cracks, (b) Longitudinal cracks, (c) Transverse cracks

with an additional background dummy class. Mean average precision is calculated using the Eq. (2) as follows:

$$\text{Mean Average Precision} = \frac{\sum_{q=1}^{Q} P_{avg}(q)}{Q} \quad (2)$$

where Q is the number of queries in the set and $P_{avg}(q)$ is the average precision (AP) for a given query, q in percentage. The intersection over union is defined as follows:

$$\text{Intersection over union} = \frac{\text{Overlapping Region}}{\text{Combined Region}} \quad (3)$$

Moreover, there are mAP defined by Common Objects in Context (COCO) challenge used in this experiment which calculated across different object scales such as AP_s, AP_M, AP_L. AP_s is AP for small objects that coves area less than 32^2; AP_M is AP for medium objects that covers area greater than 32^2 but less than 96^2 and AP_L is AP for large objects that covers area greater than 96^2 . Pre-trained Common Objects in Context (COCO) (Lin *et al.*, 2014) weights were used in this experiment to initialize the weights on ResNet feature extractor. For others network such as region proposal network, the weights were initialized using Xavier uniform initializer (Glorot and Bengio, 2010).

The base configuration of Mask R-CNN was initiated using the setup in Table 1. There were four output classes in this context, which was crocodile crack, longitudinal crack, transverse crack, and the micessalaneous background. The backbone used in this paper is ResNet-50. The training steps per epoch is set to 100 to visuallize the total loss and mask loss. The batch size was set to 2 as the largest size of our system can handle and learning rate to 0.001 which to avoid the weights explosion that might cause the unstable learning (Goodfellow *et al.*, 2016). The optimization set as Stochastic Gradient Descent (SGD) with momentum of 0.9 and weight decay of 0.0001.

Table 1. Configurations of Mask R-CNN for Building Cracks Identification

Parameter	Configuration
BACKBONE	ResNet-50
BATCH SIZE	2
OPTIMIZER	SGD
LEARNING RATE	0.001
LEARNING MOMENTUM	0.9
WEIGHT DECAY	0.0001
TRAINING STEPS	100
RPN ANCHOR SCALES	(64, 128, 256,512, 1024)
RPN TRAIN ANCHORS PER IMAGE	320
TRAIN ROIS PER IMAGES	200

Another important parameter in Mask R-CNN is the regional proposal network anchor scales. The cracks images in the dataset contains the image dimension of 1024 as illustrated in Figure 4. The anchor scale is set from the increase of minimum value 32 to 64 to the maximum value of 512 to 1024. Figure 5 demonstrates the RoI before refinement using 32 to 512 RPN anchor scales. To avoid the slow down of training process, the RPN anchor is set to be 320 and the number of RoI is set to be 200 per image in the experiment.

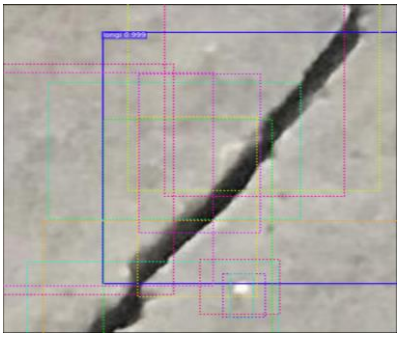


Figure 5. Example in process of detection indicated anchor scale was insufficient to provide anchor boxes for shrouding crack in image

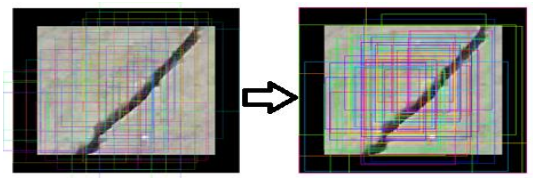
IV. RESULT AND DISCUSSION

A. Mask R-CNN vs YOLOv3

Mask R-CNN and YOLOv3 were compared in the experiemntfor for crack identification. YOLOv3 is the state-of-the-art framework of object detection which is able to detect object using single shot multiple detector. The single-shot detector requires only a single pass in the neural network and predicts all bounding boxes with one access without specifically going through two times of region proposals network as illustrated in Figure 6. However, YOLOv3 can achieve mAP_{50} of 28.7% where was much lower than Mask R-CNN, which achieved mAP_{50} of 50.0% as shown in Table 2.

Table 2. Experimental results with comparison in mAP_{50}

Method	backbone	mAP_{50}
Mask R-CNN	ResNet-50-FPN	50.0
YOLOv3	DarkNet-53	28.7



(a)

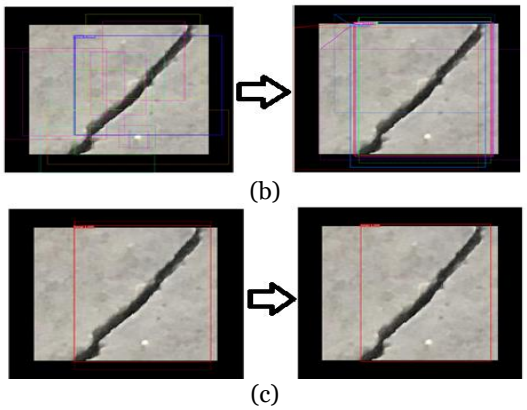


Figure 6. Process in region proposal network: (a)Bounding box generation process, (b) regions of interest refinement process, (c) detection with non-max suppression process

A detailed investigation was compared in Yao *et al.* (2019) using the Average Precision (AP) measurement metrice over small, medium and large object areas between the YOLOv3 (Redmon and Farhadi, 2018) and Mask R-CNN (He *et al.*, 2017). The small AP scale is set as the area size below 32^2; AP for mediurm objects is within 32^2 and 96^2; and AP for large object with area more than 96^2, as defined in the COCO evaluation metrics (Lin *et al.*, 2014). The performance of Mask R-CNN acehived $AP_S = 15.5\%$, $AP_M = 38.1\%$, $AP_L = 52.4\%$ while the YOLOv3 received the result of $AP_S = 18.3\%$, $AP_M = 35.4\%$, $AP_L = 41.9\%$. As in comparision between both performance, the YOLOv3 performed worse result in the large object segmention with $AP_L = 41.9\%$ as compared to Mask R-CNN with $AP_L = 52.4\%$. This result indicates that Mask R-CNN has consistent performance across all scales of object size. As a result, our experiment phenomena can be explained as YOLOv3 segemented worse boundary of large object as it has much lower values in AP_L as compared to Mask R-CNN. In addition, the crack dataset used in the experiment were full scale crack images. As recorded in the experiment by Murao *et al.* (2019), they concluded that YOLO framework performed badly in crack identification with the precision ratio of 27.95% in the worst case and 41.63% in the best case. In our experiment, the cracks samples detected using YOLOv3 were demonstrated in Figure 7. YOLOv3 was sensitive to tiny changes in an image, and therefore, the segmentation results was worse in the crack identification. As illustratated in Figure 8, Mask R-CNN performed better segmentation results in all three types of cracks. However, there are some misclassfication in Mask R-CNN to segment crack correctly in the experiment.

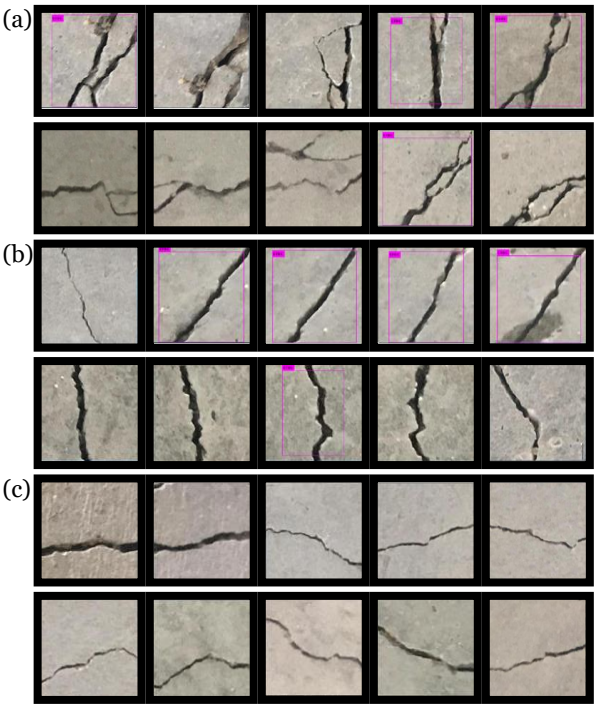


Figure 7. YOLOv3 detection results (a) Crocodile cracks, (b) Longitudinal cracks, (c) Transverse cracks

B. Mask R-CNN vs Enhanced Mask R-CNN

Fang *et al.* (2019) demonstrated the Mask R-CNN models performance comparision with FCN8s, SegNet, PSPNet, U Net, and Dilate ResNet for crack detetction. Mask R-CNN performance was however not very prominent between models with precision of 35.20%. This is due to the crack pixel and background pixel do not have clear boundary in the picture. In Zimmermann and Siems (2019), it also stated that masking provided by Mask R-CNN often do not segment the real object boundaries accordingly. However, in instance segmentation for crack detection, it is especially critical to extract the boundary between the crack and the background to achieve an accurate result. Therefore, to improve the performance of Mask R-CNN, an enchanced Edge Agreement Head was added to evaluate the performance of Mask R-CNN. Additional mask edge loss was added on to the total loss of Mask R-CNN as an indicator for the auxiliary loss for Edge Agreement Head. Figure 9 showed the total loss of Mask R-CNN before and after applying Edge Agreement Head. However, the added mask edge loss has contributed the total loss increment.

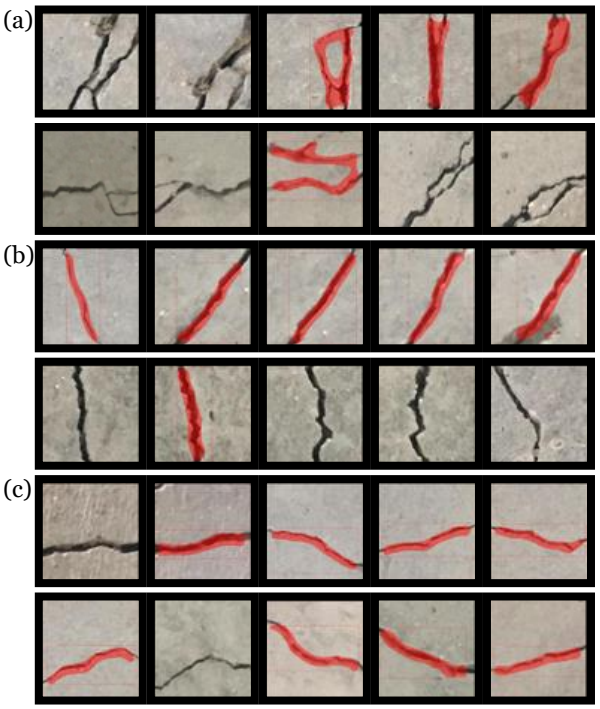


Figure 8. Mask R-CNN detection results of (a) Crocodile cracks, (b) Longitudinal cracks, (c) Transverse cracks

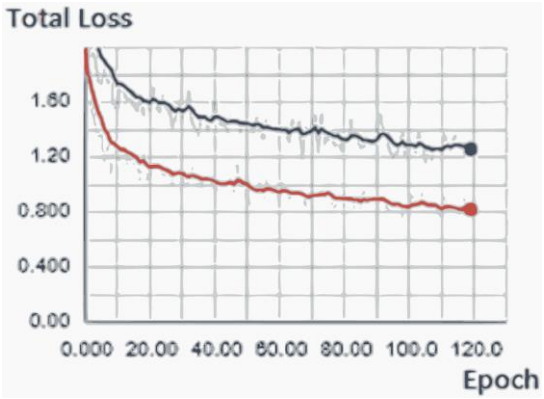


Figure 9. Total loss per epoch for Mask R-CNN before Edge Agreement Head was applied (orange) and Mask R-CNN after Edge Agreement Head was applied (blue)

Instead of investigating the total loss of network, individual loss in Mask R-CNN such as mask loss is ploted in Figure 10 for comparision. The enchanced Mask-RCNN presented the functionality to guide the learning process with the lower mask loss. The Edge Agreement Head with Sobel Filtering decreased the mask loss as compared to the original Mask R-CNN. The model with added Sobel filter as the Edge Agreement Head achieved the mAP_{50} of 63.3%, which was much higher than the original model with mAP_{50} of 50.0%. The samples of test results using the enhanced Mask R-CNN is illustrated in Figure 11.

With increasing of value of mAP_{50}, it meant that the percentage of correct positive predicted crack over the test images was increased as well as among all positive cases. According to Caruana (1993), Edge Agreement Head is able to enhance task generalization by leveraging the domain-specific information extracted in the training signals of related tasks.

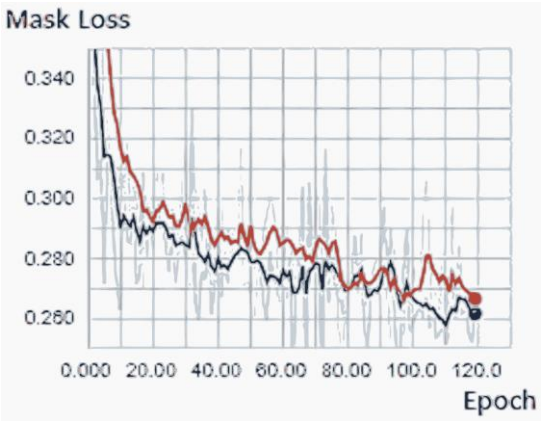


Figure 10. Mask loss per epoch at training for Mask R-CNN before (orange) and after the Edge Agreement Head was applied (blue)

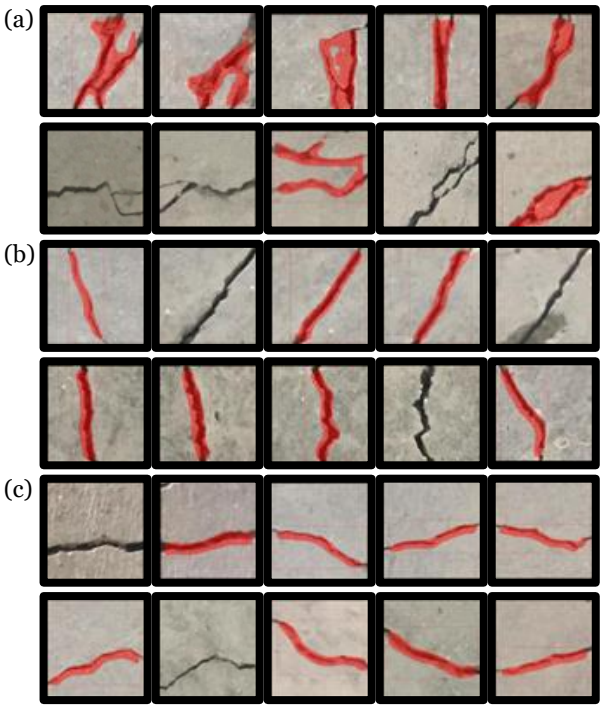


Figure 11. Mask R-CNN after the Edge Agreement Head was applied results in test data-set of crack detection of (a) Crocodile cracks, (b) Longitudinal cracks, (c) Transverse cracks

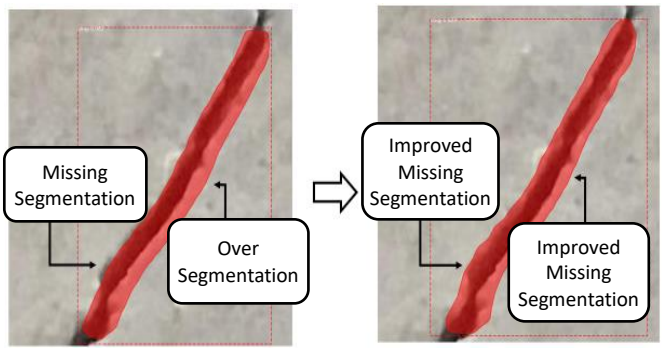
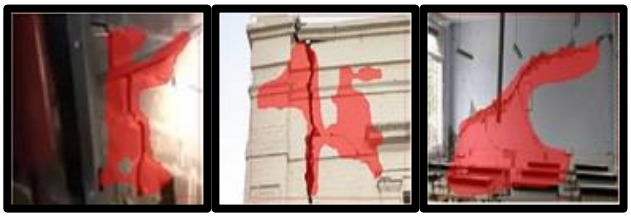


Figure 12. Sample outcome chosen from test data-set of crack detection between (a)before and (b) after Sobel Filtering Head had applied. (Arrow indicate the area affected by over segmentation and circle indiate the area affected by missing segmentation)

As sample model from Figure 12, two main improvement as expected such as reducing over segmentation and avoid missing part from the mask had gave a better results in instance segmentation. Table 3 tabulated the result of Mask R-CNN before and after Edge Agreement Head using Soble filter. With the good result toward mAP_{50}, the Mask R-CNN model was further applied on some real world situation as illustrated in Figure 13. However, the outcomes were still have a gap towards satisfied results as real world situation image consisted of some other object as a noise for crack detection but somehow the crack in images was at least detected for a big portion.

Table 3. Summary of Mask R-CNN before and after enhancement at 120 epochs

Mask R-CNN	Original	Edge Agreement Head with Sobel Filtering
mAP_{50}	50.0	63.3
Total Loss	0.7796	1.2140
Mask Loss	0.2619	0.2603
Mask Edge Loss	-	0.4369



(a) (b) (c)

Figure 13. Crack identifications in the real world scenarios such as (a) Cracks at the sewage pipes, (b) Cracks on external structures, (c) Craks on the wall

V. CONCLUSION

Enhanced Mask R-CNN with the Edge Agreement Head using Sobel Filter is proposed in this paper to identify building cracks. First, building surveying images are inserted into the Mask-RCNN for region extraction. Subsequent, an accurate instance segementation was implemented to retrieve the cracks edge. The Edge Agreement Head performed with sobel filter is to retrieve the closer cracks on the wall. The performance of enhanced Mask R-CNN can achieve up to 63.3% of mAP. As compared to the original Mask R-CNN, the enhanced version of Mask R-CNN has increased mAP of 13.3% from 50.0% to 63.3% in building crack identification. In the future work, to increase the mAP and accuracy of enhanced Mask R-CNN, more training smaples images shall be captured and inserted in the experiment training phase.

VI. REFERENCES

Aboudi, J. 1987. Stiffness reduction of cracked solids. *Engineering Fracture Mechanics,* 26**,** 637-650.

Broberg, P. 2013. Surface crack detection in welds using thermography. *NDT & E International,* 57**,** 69-73.

Budiansky, B. & O'connell, R. J. 1976. Elastic moduli of a cracked solid. *International Journal of Solids and Structures,* 12**,** 81-97.

Caruana, R. 1993. Multitask learning: a knowledge-based source of inductive bias. *Proceedings of the Tenth International Conference on International Conference on Machine Learning*. Amherst, MA, USA: Morgan Kaufmann Publishers Inc.

Dhital, D. & Lee, J. R. 2012. A Fully Non-Contact Ultrasonic Propagation Imaging System for Closed Surface Crack Evaluation. *Experimental Mechanics,* 52**,** 1111-1122.

Ding, J., Li, X. & Gudivada, V. N. Augmentation and evaluation of training data for deep learning. 2017 IEEE International Conference on Big Data (Big Data), 11-14 Dec. 2017 2017. 2603-2611.

Everingham, M., Van gool, L., Williams, C. K. I., Winn, J. & Zisserman, A. 2010. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision,* 88**,** 303-338.

Fang, F., Li, L., Rice, M. & Lim, J. Towards Real-Time Crack Detection Using a Deep Neural Network With a Bayesian Fusion Algorithm. 2019 IEEE International Conference on Image Processing (ICIP), 22-25 Sept. 2019 2019. 2976-2980.

Fujita, Y. & Hamamoto, Y. 2011. A robust automatic crack detection method from noisy concrete surfaces. *Machine Vision and Applications,* 22**,** 245-254.

Glorot, X. & Bengio, Y. 2010. Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research - Proceedings Track,* 9**,** 249-256.

Goodfellow, I., Bengio, Y. & Courville, A. 2016. *Deep Learning*, MIT Press.

Haykin, S. & Network, N. 2004. A comprehensive foundation. *Neural networks,* 2**,** 41.

He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), 22-29 Oct. 2017 2017. 2980-2988.

Jain, A., Gupta, M., Tazi, S. N. & Deepika. Comparison of edge detectors. 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 7-8 Nov. 2014 2014. 289-294.

Lecun, Y., Kavukcuoglu, K. & Farabet, C. Convolutional networks and applications in vision. Proceedings of 2010 IEEE International Symposium on Circuits and Systems, 30 May-2 June 2010 2010. 253-256.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. & Zitnick, C. L. Microsoft COCO: Common Objects in Context. *In:* FLEET, D., PAJDLA, T., SCHIELE, B. & TUYTELAARS, T., eds. Computer Vision – ECCV 2014, 2014// 2014 Cham. Springer International Publishing, 740-755.

Mccann, M. T., Jin, K. H. & Unser, M. 2017. Convolutional Neural Networks for Inverse Problems in Imaging: A Review. *IEEE Signal Processing Magazine,* 34**,** 85-95.

Murao, S., Nomura, Y., Furuta, H. & Kim, C.-W. 2019. Concrete Crack Detection Using UAV and Deep Learning. *13th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP13.* Seoul, South Korea.

Özgenel, Ç. & Sorguc, A. 2018. *Performance Comparison of Pretrained Convolutional Neural Networks on Crack Detection in Buildings*.

Pan, S. J. & Yang, Q. 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering,* 22**,** 1345-1359.

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**,** 591-599.

Redmon, J. & Farhadi, A. 2018. YOLOv3: An Incremental Improvement.

Ren, S., He, K., Girshick, R. & Sun, J. 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.,* 39**,** 1137-1149.

Saar, T. & Talvik, O. Automatic Asphalt pavement crack detection and classification using Neural Networks. 2010 12th Biennial Baltic Electronics Conference, 4-6 Oct. 2010 2010. 345-348.

Suresh, S., S N, O., Ganguli, R. & Mani, V. 2004. Identification of crack location and depth in a cantilever beam using a modular neural network approach. *Smart Materials and Structures,* 13.

Talab, A. M. A., Huang, Z., XI, F. & Haiming, L. 2016. Detection crack in image using Otsu method and multiple

filtering in image processing techniques. *Optik,* 127**,** 1030-1033.

Wang, K. C., Elliott, R. P., Meadors, A. & Evans, M. Application and Validation of An Automated Cracking Survey System. Proceedings of the 6th International Conference on Managing Pavements, 2004. 1-19.

Yao, J., Yu, Z., Yu, J. and Tao, D., 2019. Single Pixel Reconstruction for One-stage Instance Segmentation. *arXiv preprint arXiv:1904.07426*.

Zimmermann, R. S. & Siems, J. N. 2019. Faster training of Mask R-CNN by focusing on instance boundaries. *Computer Vision and Image Understanding,* 188**,** 102795.

Zitnick, C. L. & Dollár, P. Edge Boxes: Locating Object Proposals from Edges. *In:* FLEET, D., PAJDLA, T., SCHIELE, B. & TUYTELAARS, T., eds. Computer Vision – ECCV 2014, 2014// 2014 Cham. Springer International Publishing, 391-405.