AUTONOMOUS MULTIPLE DAMAGE DETECTION AND SEGMENTATION IN STRUCTURES USING DEEP LEARNING

Saitama University Student Member ○Sal Saad Al Deen Taher
Saitama University Regular Member Ji Dang

1. INTRODUCTION

There are nearly 730,000 road bridges (excluding railway bridges) in Japan, and more than 23% of them are over 50 years old or older as of 2020, with the number increasing to over 50% by 2030. Proper maintenance of the bridges is necessary to ensure the integrity of these structures. Thus, regular inspection of bridges is crucial to check the overall condition. However, the assessment is typically maintained manually and is dependent on a person's experience, and the task is also time-consuming, costly, and unreliable. During the last decade, numerous machine learning techniques have been utilized to identify the damages in structures. Different deep learning-based techniques, such as Convolutional Neural Network (CNN), have been used to detect the damages automatically. However, most of them are based on damage detection and are often used for a single class of damage detection. Instance segmentation is a process in which each object is detected as a separate instance. The instances can be shown separately using a Region-based CNN model such as Mask R-CNN. Instance segmentation methods are extensively used for detecting common objects around us but utilizing them to detect structural damages is very limited. The R-CNN model's training and testing for multiple damage detection are different and more challenging than detecting common objects. In this study, a Mask R-CNN based instance segmentation model is trained and tested for multiple damage detection in structures with an ambition to work under complex background.

2. METHODOLOGY

2.1 Architecture of Mask R-CNN

Mask R-CNN was developed by He, et al. (2017). It is a deep neural network designed to solve instance segmentation problems in machine learning or computer vision. In simple words, it can separate different objects from the input image or a video by generating bounding boxes, classes, and masks. Mask R-CNN consists of two stages-(i) Generation of region proposals by predicting an object based on the input image. (ii) Prediction of an object class, refinement of the bounding box, and generation of a mask in pixel level based on the first stage region proposal. Both stages are connected to a backbone structure: a Feature Pyramid Network (FPN) style CNN like ResNet or VGG.

Mask R-CNN is an improvement of previously developed Faster R-CNN. Ren, et al. (2015). Significant progress is achieved by replacing Region of Interest or ROI pooling layer with ROI Align and including a Mask head in the Fully Connected (FC) layer that can predict boxes, classes and masks simultaneously. The loss function for Mask R-CNN is as follows (Eq. 1)

$$L=L_{cls}+L_{bbox}+L_{mask} \tag{1}$$

L_{cls} is the classification loss, which tells how close the predictions are to the true class. L_{bbox} is the bounding box loss, which tells how good the model is at localization. L_{mask}, loss for mask prediction, is calculated by taking the binary cross-entropy between the predicted mask and the ground truth.

2.2 Preparation of Dataset

In this study, the Mask R-CNN model has been trained and tested to detect three types of structural damages-(i) Crack, (ii) Corrosion, and (iii) Spalling. A dataset of total 1,950 images has been used. The details is given on Table 1

Type of Damage	Crack	Corrosion	Spalling	Total
Training Images	810	450	495	1755
Testing Images	90	50	55	195
Total	**900**	**500**	**550**	**1950**

Table 1: Details of Dataset for Training and Testing

2.3 Details of Training

For the testing, we have used Pyhton 3.6, Tensorflow Version 1.15.2, Keras Version 2.1.5. The training was carried cloud-based Google Collaboratory with integrated GPU supports. Each mini-batch had one image per GPU. The network was trained for a total of 200 epochs at 100 iterations per epoch. Learning rate, weight decay, and momentum were set as 0.001, 0.0001, and 0.9, respectively. Only transfer learning was adopted, which means we only trained the network heads freezing all other layers as default. The images are annotated using the MATLAB Image Annotator tool and converted to Microsoft Common Objects in Context (MS COCO) format using Python 3.6 language. The conversion has been shown in Fig. 1.

Keywords: Damage detection, Mask R-CNN, Instance Segmentation, Deep learning
Contact Address: : Shimo-Okubo 255, Sakura-ku, Saitama-shi, 338-8570, Japan. Tel: +81-48-458-3552

(a) Original Image (b)Annotated Image (c) Ground Truth

Fig. 1: Converted Annotated Images

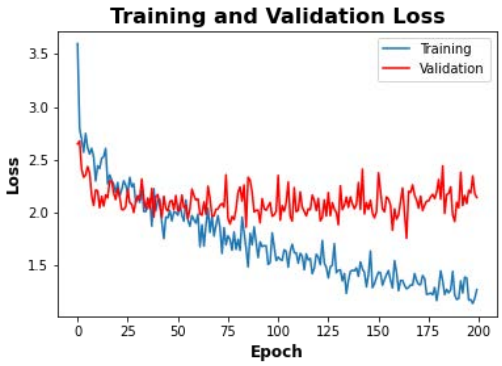


Fig. 2: Training and Validation Loss

4. RESULTS AND DISCUSSION

The model has run for 200 epochs. The training loss decreases in a regular pattern, but the validation loss pattern is not regular (Fig. 2). This means that the model has been overfitting the training data. So, we have to adopt a new method like data augmentation to reduce overfitting or may increase the number of validation data. Some validation results are shown in Table 2.

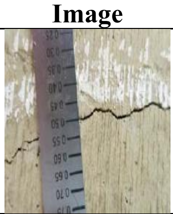
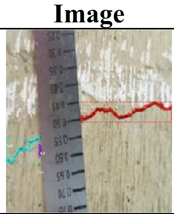
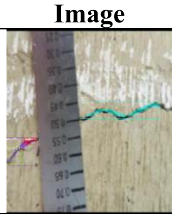
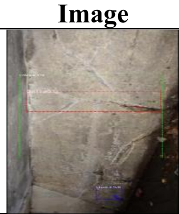
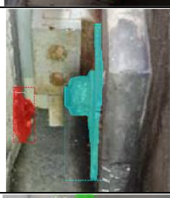
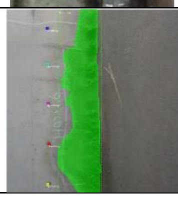
Damage Class	Image Set 1			Image Set 2		
	Original Image	Ground Truth Image	Validated Image	Original Image	Ground Truth Image	Validated Image
Crack						
Corrosion						
Spalling						

Table 2: Some Validation Result

Two instances of validation results are shown in Table-2. We can see that for the first set of images presented in Table-2, the predicted label is close to the ground truth. But for the second set of images, there is some misprediction of labels. So, improvement of the dataset and training is necessary.

5. CONCLUSIONS

In this study, the Mask R-CNN model predicts multiple structural damages and can represent them as separate instances by generating bounding boxes and masks around the predicted classes. Several improvements must be made to adopt a model that can predict multiple damages in a complex background with acceptable accuracy. For this, we can take steps like fine tuning of the model, setting best hyperparameters for the model, data augmentation, increase the number and quality of images and training all the layers of the model.

REFERENCES

He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2020. "Mask R-CNN." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42: 386–397. doi:10.1109/TPAMI.2018.2844175.

Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. 2017. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39: 1137–1149. doi:10.1109/TPAMI.2016.2577031.