CETRA2022 11-13 May 2022, Pula, Croatia
7th International Conferenceon Road and Rail Infrastructure

EVALUATION OF TRAINING DATA QUALITY FOR DEEP LEARNING-BASED DAMAGE DETECTION

Tomotaka Fukuoka[1], Mai Yoshikura[2], Makoto Fujiu[1]
[1] *Kanazawa University, Geosciences and Civil Engineering, Japan*
[2] *Kanazawa University, Department of Environmental Design, Japan*

Abstract

A bridge inspection needs a lot of costs. It causes a lack of engineers and budget. So, some local governments couldn't complete the bridge's aggressive preventive maintenance in Japan. Recently, deep learning-based damage detection methods have been studied by many researchers to reduce the cost of the bridge's aggressive preventive maintenance. This kind of method could detect the damage to the bridge by photo image with a detection model which has been trained with a large training dataset. In contrast to the increase in dataset size, the effectiveness of the quality of training data is not discussed enough. In this paper, the ratio of negative samples in the training data is regarded as the quality of training data. In this study, we targeted the automatic detection process of the rebar exposure and the peeling on the surface of bridges. The effects of the ratio of the negative example data in a training dataset have been evaluated. In this study, the negative sample means the annotated data with no target damage. The negative sample image is the part of a bridge but does not include target damages. Many previous studies generate this kind of image when generating annotation data from real bridge photo images, but do not utilize it. A three-fold cross-validation method has been adopted to keep the robustness of the detection. The seven different detection models have trained with seven different training dataset which has different negative sample ratios. As a result of comparing the detection results of each model, the influence of the negative example data on the training data was confirmed. There was a peak of the recall curve in the middle of increasing the negative sample ratios. The correct negative sample ratio could improve the accuracy of damage detection.

Keywords: bridge inspection, deep learning, image processing, training data

1 Introduction

In Japan, there are about 730,000 bridges with a length of 2 m or longer, and many of these were built during a period of high economic growth. It is expected that the number of such old bridges would increase in the future. Therefore, their rebuilding and the extension of their service life must be considered [1]. An owner of a bridge is required to monitor a bridge with close visual inspection per 5 years according to the national criteria in 2014 in Japan [2]. However, it is expected difficult to continue a close visual inspection method in the future for some problems. For example, this method needs high cost, a lack of an engineer for the bridge inspection, and a lack of budget for the maintenance. There are many studies for solving these problems by proposing a low-cost maintenance method, an efficient maintenance method, and so on.

DOI: https://doi.org/10.5592/CO/CETRA.2022.1393

An image processing-based automatic damage detection is one of the solutions. There are 26 kinds of damage that would be considered for bridge inspection in Japan [2]. Most of these damages are confirmed with visual inspection by an inspection engineer. Therefore, an image processing-based automatic damage detection method could be an alternative method. For example, a crack is detected by a visual inspection of an inspection engineer. They measure appearance points, shape, wide, and length of cracks. An inspection report stores this information. Some previous studies could detect a part of cracks [3, 4]. Recently, deep learning-based has been proposed and could detect cracks with high accuracy [5, 6, 7]. A deep learning-based image processing method is adapted for the detection of other damages. An automatic detection method has been proposed for erosion, palling, rebar exposure, leakage, and so on [8, 9]. Some studies use UAVs to take a photo for an input image of the deep learning-based automatic damage detection method [10, 11].

In general, a deep learning-based method needs a huge size dataset to train the model to detect the target well. The content of the dataset depends on the task. One of the major damage detection methods is the semantic segmentation method. This method receives image data as input data and outputs image data making the target damage area with pixel unit. In this case, the dataset is pair of data: raw image data and the mask image cover the target damage area on the raw image. This method can accumulate detection, but the cost of dataset generation is high.

Augmentation methods such as rotation, flipping, and scaling are often used to generate huge scale datasets when training data is small scale in many previous studies. However, augmented images by such methods have a similar feature. In general, a well-work deep learning model needs not only the amount of data but also needs the variation and the quality of images. Noise data is regarded as a part of the quality of training data in this paper. Noise data is important to train a robust model in some cases. But a discussion about the noise ratio in the training data is not enough in the damage detection field.

In view of the features, all bridges are different from each other in Japan. This means that bridges have different surfaces, scales, structures, the circumstance surrounding the bridge, and so on. Some bridges are located difficult to take a picture of. It is difficult to take a picture with stable quality. These pictures include noise such as the background. These facts cause low detection accuracy. One of the solutions to this problem is adopting negative sample data as training data for the detection model could distinguish damage and noise with more accuracy.

In this paper, the effects of the ratio of a negative sample in training data have been evaluated. A semantic segmentation method has been adopted to detect the rebar exposure and the peeling. The evaluation adapts the cross-validation method to keep the robustness of the evaluation and the effect of the ratio of negative sample data.

The training data use real photo images of a bridge. In general, these picture image size is larger than the input size of a deep-learning model. In ordinary, these picture images are clopped or scaled into the input size of a deep-learning model. When splitting an image, there are two kinds of images: target damage image and not including target damage image. A not including target damage image is regarded as noise data. But, these noise data are part of a bridge image. These images may be useful to distinguish between an image of damage and an image of no damage. These data have been set as a negative sample for training the deep-learning model in this paper.

2 Related works

A semantic segmentation method could detect the target object in an input image and output the masked image with the pixel unit [12-14]. The pixel unit output enables us to not only know a damaged position but also calculate real damage size using shooting range, focus range, and so on. Li et al. evaluated the detection of some damage: crack, peeling, and so on. They adapted FCN [13] and SegNet [14] as a semantic segmentation model [15]. The evaluation set real constructure as a target and take test photo images twenty centimeters from the target damage area. A pre-trained model have used when training the deep-learning model in many cases. These models have trained with large-scale datasets for a different task at first. The fine-tuning method or the transfer learning method trains these models with another dataset optimized for a current task. Nakajima et al. [15] have evaluated the rebar exposure detection method using a model trained with transfer learning. They adapt famous networks such as FCN, SegNet, and so on. They used a pre-trained model which had been trained for an image classification task and compared the result of each model. They used 208 pictures taken while close visual inspection as training and test data. This evaluation shows that a semantic segmentation method does not need training image data taken by the same filming condition.

3 Evaluation method

In this paper, the SegNet model has been adopted as a detection model for rebar exposure and peeling. This model needs more small cost to train model than the FCN model. As an evaluation method, seven models were trained with seven datasets that have a different ratio of the negative samples and the results of the detection of each model have compared.

3.1 Deep learning model

The SegNet model has trained with transfer learning. A classification model VGG16 trained by the ImageNet dataset. The ImageNet dataset has more than one hundred million images classified into one thousand classes. This model has been adopted as a pre-trained model for transfer learning for training the SegNet semantic classification model. As a result of pre-examination, the multi-class semantic segmentation method frequently detects the rebar exposure damage as the peeling, the binary semantic segmentation models have been trained: detect only rebar exposure and detect only peeling.

3.2 Dataset

As evaluation data, 179 photo images of the bridge including the rebar exposure and the peeling in a bridge inspection record has used. To train the models, the rebar exposure only annotated datasets and peeling only annotated datasets has generated from the evaluation data. The annotated data generated by masking the rebar exposure area is red and the peeling black color at first. The rebar exposure annotated data was made by converting it from a black-colored area to white and converting from a red-colored area to black. The peeling annotated data was made by converting it from a red-colored area to white (Fig. 1). As each photo images are a different size and not suitable for SegNet model input, all images have been split into small size images (height is 224 pixels, width is 224 pixels). If a fraction exists, crop 224 pixels from the right or 224 pixels from the bottom of the image. As an example, Fig. 2 shows five images including peeling, one image including rebar exposure and one negative sample image from one photo of the bridge inspection record. To improve robustness, a three-fold cross-validation method had adopted. Table 1 shows the description of the dataset of each case.

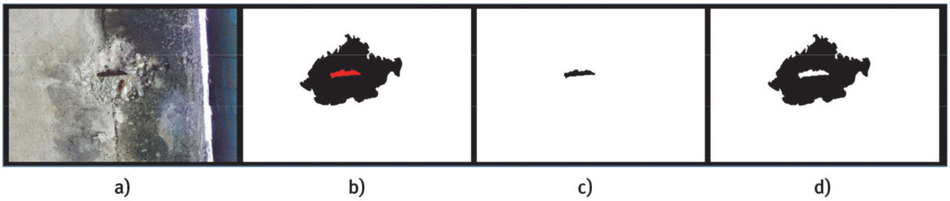

Figure 1 An example of an annotation converted to binary annotation (a: raw image, b: rebar exposure and peeling annotated image, c: only rebar exposure annotated image, d: only peeling annotated image)

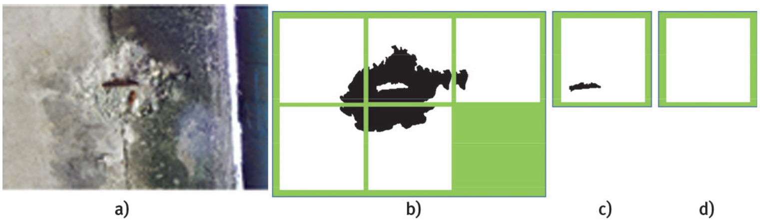

Figure 2 An example of a split annotated image (a: raw image, b: peeling annotation, c: rebar exposure annotation, d: negative sample image)

Table 1 Training data of each cross-validation case

	Case 1	Case 2	Case 3
The number of original training images	119	119	120
The number of images after splitting	1718	1751	1791
The number of annotated images in the rebar exposure image after splitting	392	405	389
The number of annotated images in detached images after splitting	596	627	607

This table shows that the average ratio of the rebar exposure including images is about 22 % (negative sample is about 78 %), and one of the peelings including images is about 35 % (negative sample is about 65 %). The number of datasets for training the model is seven: annotated image only dataset, five datasets including a random sampled negative sample with fix ratio to damage including data (from 10 % to 50 % increase by 10), the dataset using all split images.

3.3 Evaluation method

Each model has been evaluated by comparing the result of damage detection by calculating precision and recall. Precision means the ratio of correct detected damage pixels. Recall means the ratio of detected damage pixels occupies a correct damaged area. High precision means the model could detect damaged areas with little not damaged areas. High recall means the model could detect damaged areas with little miss damaged areas. eqn. (1) is a formula of precision. eqn. (2) is a formula of recall.

$$\text{Precison} = \frac{\begin{array}{c}\text{The number of pixels detected by the model as the damage to be detected}\\ \text{that were correctly detected as damage to be detected}\end{array}}{\text{The number of pixels detected by the model as the damage}} \tag{1}$$

$$\text{Recall} = \frac{\text{The number of pixels detected by the model as the damage to be detected that were correctly detected as damage to be detected}}{\text{The number of pixels of damage to be detected}} \quad (2)$$

4 Evaluation

The precision of the detection results for rebar exposure and the recall are shown in Table 2 and Table 3. The precision of the detection results for peeling and the recall are shown in Table 4 and Table 5. A red-colored value in each table means the highest value between each model in each case. Model 1, indicated by the columns in the table, is the model trained using only the annotated images; models 2 to 6 are the models trained using all the annotated images and the negative example data randomly extracted so that the negative example data account for 10 % to 50 % of the training data; model 7 is the model trained using all the data after splitting. The numbers colored red in the table represent the highest values of the models compared within each cross-validation. Fig. 3 shows the average values of the three cases for each model in Table 2-5.

Table 2 Precision value of detection of rebar exposure

Case	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
1	0.279	0.274	0.274	0.272	0.343	0.347	0.351
2	0.320	0.307	0.338	0.344	0.304	0.362	0.384
3	0.197	0.326	0.296	0.265	0.278	0.305	0.321

Table 3 Recall value of detection of rebar exposure

Case	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
1	0.250	0.257	0.280	0.392	0.197	0.178	0.215
2	0.195	0.219	0.191	0.299	0.247	0.249	0.161
3	0.353	0.197	0.267	0.284	0.266	0.201	0.197

Table 4 Precision value of detection of peeling

Case	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
1	0.403	0.328	0.398	0.402	0.409	0.428	0.465
2	0.301	0.384	0.335	0.372	0.339	0.389	0.413
3	0.313	0.340	0.329	0.374	0.327	0.373	0.407

Table 5 Recall value of detection of peeling

Case	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
1	0.506	0.620	0.612	0.541	0.563	0.461	0.426
2	0.637	0.545	0.686	0.567	0.606	0.524	0.485
3	0.621	0.676	0.618	0.565	0.638	0.561	0.498

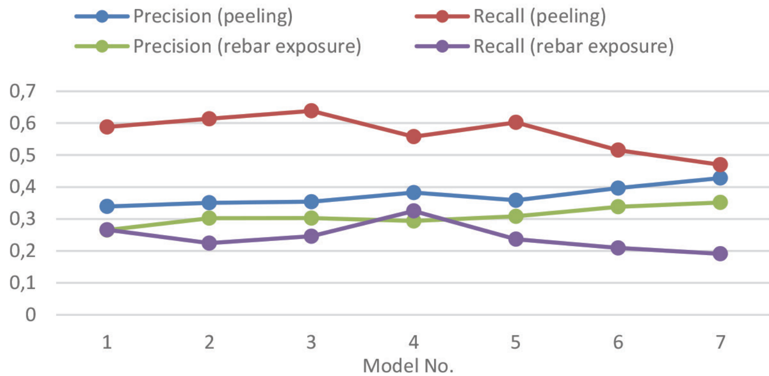

Figure 3 An average value of precision and recall of the three cases for each model of rebar exposure and peeling detection

The results shown in Table 2, Table 3, and Fig. 3, increasing the ratio of negative examples in the dataset during the training of the detection model improves the precision of the detection of rebar exposure by up to 12.9 % and the recall by up to 14.2 % compared to the results of the model without negative examples. The performance of both precision and recall tends to improve as the ratio of negative examples is increased. However, while the precision improves as the ratio of negative examples is increased, the recall is highest for model 4, and decreases as the ratio of negative examples is increased further. This result can be attributed to the fact that increasing the number of negative examples allows the detection model to more accurately distinguish between rebar exposure and rebar exposure-like areas, but the conditions for determining rebar exposure become more restrictive, making it more difficult to detect rebar exposure. In the detection model with a small ratio of negative examples, it is easy to detect false positives such as concrete joints and rusted metal parts of pipes, etc. In the model with a large ratio of negative examples, detection omissions are more likely to occur due to the color of the exposed steel bars.

From the results shown in Table 4, Table 5, and Fig. 3, the precision and recall are improved by up to 11.2 % and 11.4 %, respectively. In the same way, increasing the number of negative examples improved the precision of peeling detection, and it was confirmed that there was a tendency for a peak in the ratio of negative examples to improve the recall. A detection model has trained with a small ratio of negative examples tends to detect discolored areas incorrectly, while a model has trained with a large ratio of negative examples can only partially detect delaminated areas, resulting in detection omissions. These results indicate that the ratio of negative examples in the training data of the detection model affects the detection performance even for pixel-by-pixel damage detection. It was also shown that it is effective to set the ratio of negative cases at 20 % to 30 %, rather than simply achieving equilibrium.

5 Conclusion

In this paper, the effect of the ratio of negative examples in the training data during model training on the damage detection method for bridges using deep learning has been evaluated. In the evaluation, the pixel-by-pixel damage detection models have been created. These models have been trained with different ratios of negative examples for peeling and rebar exposure: and the detection results were compared to each other for the evaluation. The results of the three-fold cross-validation test indicated that the precision and recall of the training of the damage detection model were best improved by setting the ratio of negative cases in the training data to 20 % to 30 %, instead of simply setting the ratio to equilibrium.

However, the ratio of negative cases in the training data, which is the peak of the improvement in the reproduction rate, is different for each detection target. We will also evaluate the remaining detection targets of the 26 types of damage in bridge inspections in the future.

References

[1] White paper on land, infrastructure, transport and tourism in Japan, https://www.mlit.go.jp/hakusyo/mlit/r02/hakusho/r03/pdf/kokudo.pdf, 30 June 2021, (in Japanese)

[2] Road bureau, ministry of land, infrastructure, transport and tourism in Japan: Periodic inspection guidelines for road bridges-https://www.mlit.go.jp/road/sisaku/yobohozen/tenken/yobo4_1.pdf, 30 June 2021, (in Japanese)

[3] Yamaguchi, T., Nakamura, S., Saegusa, R., Hashimoto, S.: Image-based crack detection for real concrete surfaces. IEEJ Transactions on Electrical and Electronic Engineering, 3 (2008) 1, pp. 128-135

[4] Nguyen, H.N., Kam, T.Y., Cheng, P.Y.: An automatic approach for accurate edge detection of concrete crack utilizing 2D geometric features of crack, Journal of Signal Processing Systems, 77 (2014) 3, pp. 221-240

[5] CHUN, P.J., Igo, A.: Crack detection from image using random forest, Journal of Japan Society of Civil Engineers, F3, 71 (2015) 2, pp. I_1-I_8, (in Japanese)

[6] Yokoyama, S., Matsumoto, T.: Development of an automatic detector of cracks in concrete using machine learning, The 3rd International Conference on Sustainable Civil Engineering Structures and Construction Materials - Sustainable Structures for Future Generations, 171 (2016), pp. 1250-1255

[7] Cha, Y.J., Choi, W., Büyüköztürk, O.: Deep learning-based crack damage detection using convolutional neural networks, Computer-Aided Civil and Infrastructure Engineering, 32 (2017) 5, pp. 361-378

[8] Dong, H., Gang, T.: Damage detection of quayside crane structure based on improved Faster R-CNN, International Journal of New Developments in Engineering and Society, 3 (2019) 2, pp. 284-301

[9] Zhang, C., Chang, C., Jamshidi, M.: Concrete bridge sur-face damage detection using a single-stage detector, Computer-Aided Civil and Infrastructure Engineering, 35 (2020) 4, pp. 389-409

[10] Wang, D., Zhang, Y., Pan, Y., Peng, B., Liu, H., Ma, R.: An Automated Inspection Method for the Steel Box Gird-er Bottom of Long-Span Bridges Based on Deep Learning, IEEE Access, 8 (2020), pp. 94010-94023

[11] Dang, J., Matsuyama, T., CHUN, P.-J., Shi, J., Matsunaga, S.: Deep convolutional neural networks for bridge deterioration detection by UAV inspection, Intelligence, Informatics and Infrastructure, 1 (2020) J1, pp. 596-605, (in Japanese)

[12] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation, MICCAI, 2015.

[13] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation, CVPR, 2015.

[14] Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, ArXiv:1511.00561v3, 2016.

[15] Li, S., Zhao, X., Zhou, G.: Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network, Computer-Aided Civil and Infrastructure Engineering, 34 (2019) 7, pp. 616-634

[16] Nakajima, M., Yasuno, T., Nagatomi, D., Noda, K., Aoyanagi, K., Sekiguchi S.: Eye-inspection image-based transfer learning rebar expose using semantic segmentation, Proceedings of the Annual Conference of JSAI, 2019, (in Japanese)