A STUDY TO EVALUATE THE EFFECTIVENESS OF THE SIZE OF IMAGE FOR THE TRANSFORMER MODEL-BASED BRIDGE DAMAGE DETECTION METHOD

T. Fukuoka[1] **and M. Fujiu**[2]

[1] *Assistant Professor, College of Transdisciplinary Sciences for Innovation, Kanazawa University, Kanazawa, Japan,*
[2] *Associate Professor, College of Transdisciplinary Sciences for Innovation, Kanazawa University, Kanazawa, Japan,*
Correspond to Assistant Professor, T. Fukuoka (tfukuoka@se.kanazawa-u.ac.jp)

Keywords*: bridge inspection, damage detection, image processing, image size, transformer model*

1. GENERAL INSTRUCTIONS

Recently, the study of extending the service life of bridges has gained attention. In Japan, periodic inspections of bridges by the close visual inspection method are conducted once every five years. This bridge inspection method needs much cost. Because of the lack of engineers and budget, some local governments couldn't complete the bridge's aggressive preventive maintenance in Japan. To solve those problems, studies of automation have been made to reduce the cost of the inspection task which depends on human power.

Deep learning-based damage detection methods are one of the methods to reduce human power. An image processing method could detect damage from a photo image of a bridge. We focus on the state-of-the-art technology for semantic segmentation method which uses the Transformer model[1]. This kind of method has high accuracy to detect the target from an input image. On the other hand, there is not enough discussion about the effectiveness of the size of an image for this new method. In many cases, the size of an input image is different from the assumption of the input layer of a detection model. It is necessary for preprocessing: rescale, cut out, split, and so on. Such preprocessing changes the information of an input image but there is not enough discussion about its effectiveness when using the Transformer model-based method.

2. EXAMINATION

We set the task of detecting the peeling and the rebar exposure on the surface image of a bridge as an evaluation. As a semantic segmentation method, SegFormer[2] model has been trained to detect the peeling and the rebar exposure.

As an evaluation dataset, 179 annotated images have been used. The average value of image height is 595 pixels. The average value of image width is 811 pixels. The dataset has been split into training data and test data. 119 images have been used as training data and 60 images have been used as test data. In this examination, three training datasets have been generated from the original training dataset for training detection models by splitting images with different split-size (Fig 1). Dataset A only includes images that height is 224 pixels and width is 224 pixels. Dataset B only includes images that height is 448 pixels and width is 448 pixels. Dataset C includes raw images that have not been split. Each dataset has 17180 images by augmenting its split or raw images.

In this evaluation, the result of the precision value and

Figure 2. The example of a different split-size image

the recall value of each detection model has been compared. Detection model SF_A has trained with dataset A. Detection model SF_B has trained with model B. Detection model SF_C has trained with dataset C. As a preprocessing of the detection with each model, input images have been split with the size of images each dataset has.

Table 1. show the detection results of the peeling and the rebar exposure. These results show that the split size of the input image affects detection accuracy.

Table 1. Results of the peeling (upper case) and the rebar exposure (lower case) detection

	SF_A	SF_B	SF_C
Precision	0.771	0.769	0.808
Recall	0.658	0.674	0.741
Precision	0.747	0.765	0.708
Recall	0.486	0.563	0.685

3. CONCLUSIONS

The state-of-the-art technology for semantic segmentation method which uses the Transformer model has not been enough discussion about its effectiveness of split-size of image yet. In this paper, the effectiveness of the size of images for detection of the peeling and the rebar exposure have been evaluated. We would evaluate the effectiveness of the size larger than the assumption size of the detection model in future works.

REFERENCES

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin.: Attention is All You Need. Proceedings of the 31st International Conference on Neural Information Processing Systems, pp.6000-6010, 2017.

[2] Dosovitskiy, A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houlsby N.: An Image is Worth 16x16 Words: Transformers for Image Recogni-tion at Scale, Proceedings of International Conference on Learning Representations, 2021.