The Seventh Asian-Pacific Symposium on Structural Reliability and Its Applications (APSSRA2020)
October 4–7 2020, Tokyo, Japan
T. Takada, I. Yoshida & T. Itoi (editors)

Image-Based Building Damage Evaluation Based on Semantic Segmentation and Convolutional Neural Network

M. Yamaguchi[1], T. Hida[2], T. Itoi[3] and T. Takada[4]

[1]*Former Undergraduate Student, The University of Tokyo. Email: yamaguchi@load.arch.t.u-tokyo.ac.jp*
[2]*Assistant Professor, School of Engineering, The University of Tokyo. Email: hida@load.arch.t.u-tokyo.ac.jp*
[3]*Associate Professor, School of Engineering, The University of Tokyo. Email: tatsuya.itoi@arch.t.u-tokyo.ac.jp*
[4]*Director, Nuclear Safety and Disaster Prevention Support Section, Japan Atomic Energy Agency, Email: takada.tsuyoshi@jaea.go.jp*

Abstract: In this study, we introduced the semantic segmentation based on U-net methodology to image-based building damage evaluation with CNN. First, the 3D CG models were constructed to generate the data sets for training the U-net. Second, the building damage was evaluated through CNN based on the original image and segmented image by U-net. Then, the validity of damage evaluation result by using segmented image was shown based on Grad-CAM technique. Finally, the geographical distribution of building damage was visualized by comparing actual damage distribution and the results of building damage evaluated by CNN.

Keywords: Structural Health Monitoring, Damage Evaluation, Artificial Intelligence, CNN, Semantic Segmentation.

1. Introduction

It is important to grasp the damage situation of structures due to a natural disaster promptly because it is related to the strategic judgment of the input of human resources and goods for emergency response. However, the initial response in the event of a disaster is delayed due to the delay in understanding the damage situation. For instance, Fig. 1 shows the increase of the number of identified building damages in case of Typhoon No. 15 in 2019 by date and time (Fire Department 2019). It took a week or more to grasp the overall disaster scale.

In recent years, the development of deep learning in the field of image recognition has been remarkably advanced. Image recognition using a computer is characterized in that it can process a large number of images at higher speed than humans. In other words, the use of artificial intelligence is appropriate for prompt damage identification of structures. Therefore, it is expected to be able to quickly determine the damage status by mechanically analysing the images of damaged structures trough deep learning.

Previous studies (Hida et al. 2018) investigated the capability of structural damage identification using a convolutional neural network (CNN, LeCun et al. 1998), which is one of the deep learning methods. The study pointed that there was a problem that the CNN can react to parts other than buildings in the photo, such as the ground, plants, and clouds. In such cases, the accuracy of damage evaluation tends to be low. Therefore, in order to increase the accuracy, it is necessary to let CNN recognize where the building is in the image.

To solve this problem, we use semantic segmentation, which is a pixel-by-pixel labelling. It is an issue for semantic segmentation to provide image segmentation data as training data. However, damaged buildings have a wide variety of factors such as structural types and building heights. Furthermore, a large amount of image data is needed for training. To deal with it, we propose a method that uses a 3D CG model as training data in order to perform segmentation on a wide variety of buildings and a number of images. Then, segmented building images are used for training the CNN, and the accuracy of damage evaluation of buildings is validated by using the photos took in the disaster area of the 2016 Kumamoto earthquake.

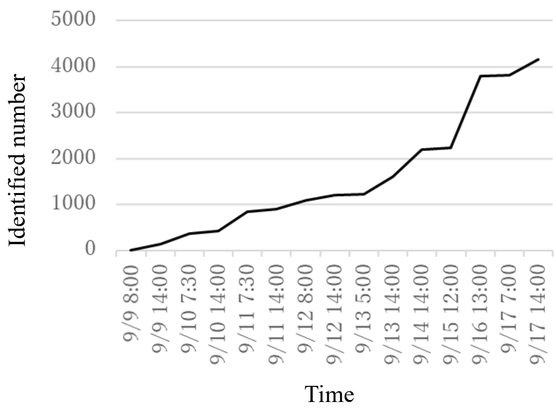

Figure 1. Increase of the number of identified damaged buildings in case of the Typhoon No. 15 occurred in 2019 in Japan

2. Outline of proposed method

Figure 2 shows the flow chart of this study. First, we examine the segmentation addition process to improve the accuracy of damage evaluation. We construct 3D CG models and generate the image datasets from the 3D models. The image data sets are used for training of U-Net, which is a neural network for semantic segmentation. After that, we will verify the applicability of U-Net to actual building images based on the photos taken in the natural disaster areas.

Next, we perform damage evaluation by Convolutional Neural Network (CNN) and compare the existing method (without semantic segmentation) with the proposed method. In the damage level judgment using CNN, the accuracy is calculated for each of the cases where segmentation processing is performed and those where it

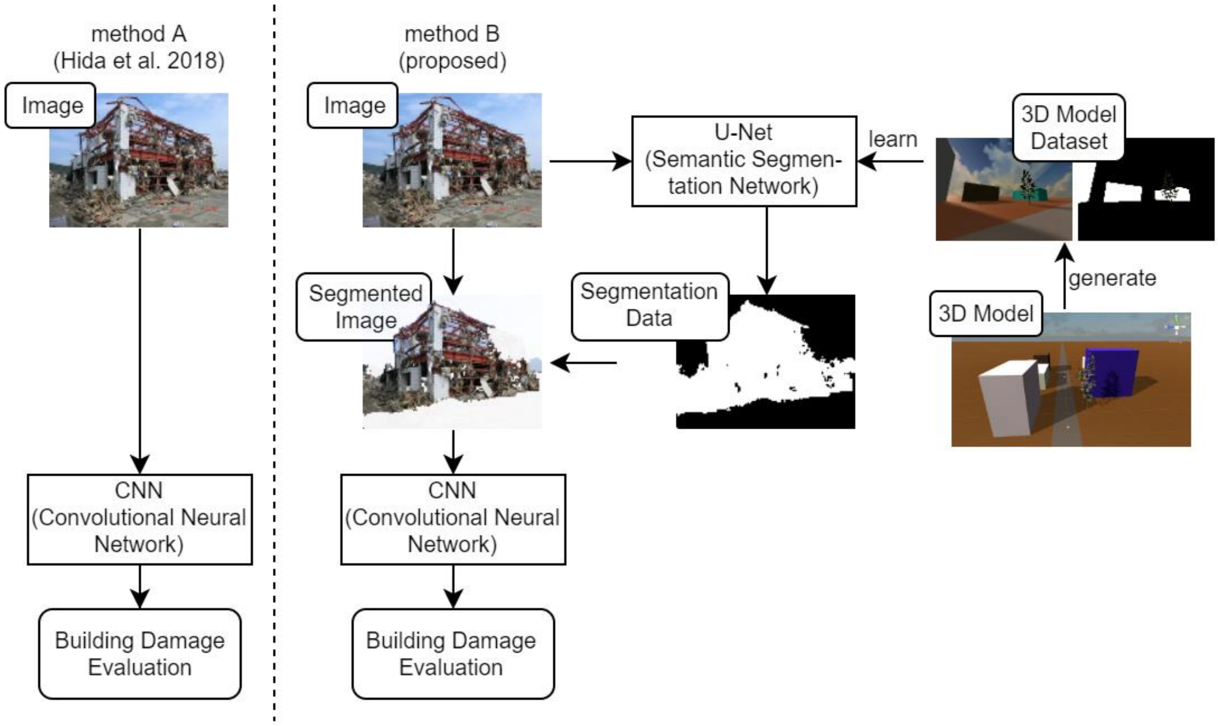

Figure 2. Flowchart of proposed methods

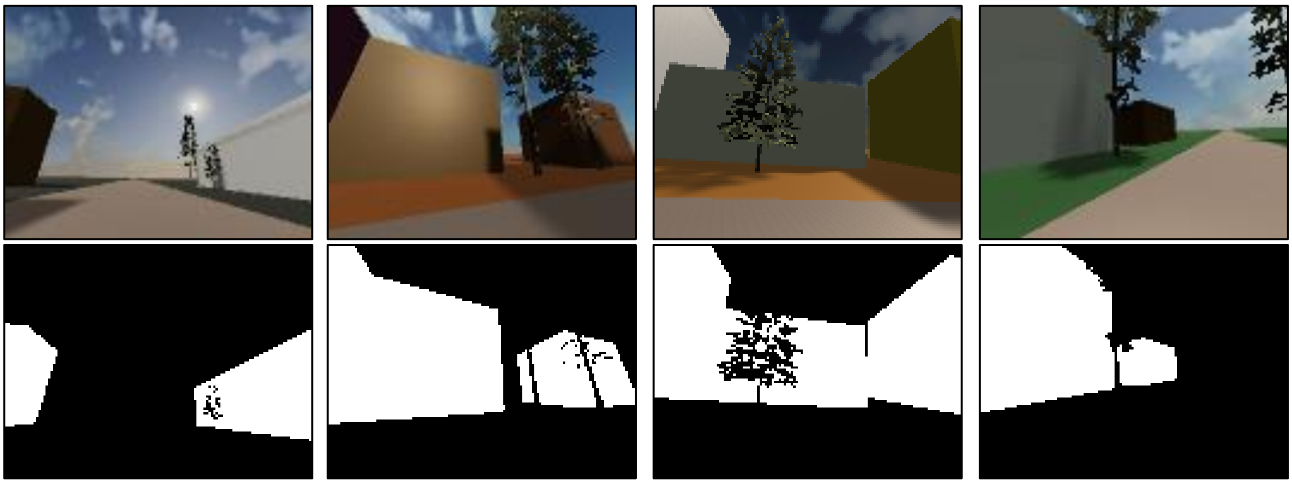

Figure 3. Examples of training data set

is not performed. In addition, the validity of proposed method is confirmed by using the CNN created by the proposed method.

3. Semantic Segmentation

Semantic segmentation is the methodology outputting pixel-based labelling. In this study, U-Net (Ronneberger et al. 2015) is used as a semantic segmentation network.

In the field of semantic segmentation problem, thousands of data sets that consists of image and segmented image as training data are required. In this study, 3D CG model is used for automatic generation of training data. Figure 3 illustrates the examples of the virtual space and data sets. A 3D CG engine, Unity, is used as a tool for making 3D model. The virtual space consists of ground, road, trees, sunlight, sky, and buildings are constructed by using the 3D model. A large amount of data sets can be obtained by setting the angle of the camera to be random and generating the images of the virtual space.

The buildings are cubes, the color, the height, the size and the location are set randomly. The trees and sky are duplications of Unity asset, the location and the number are random. The sunlight is a parallel light from random direction.

The Unity program generated 10200 data sets consist of image and segmented image. The white area of the segmented images shows where building is. Image size is

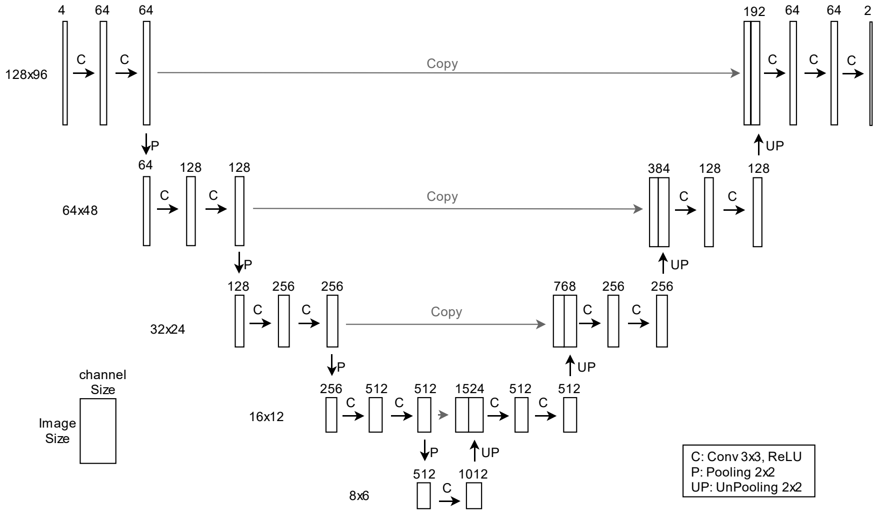

Figure 4. U-Net configuration

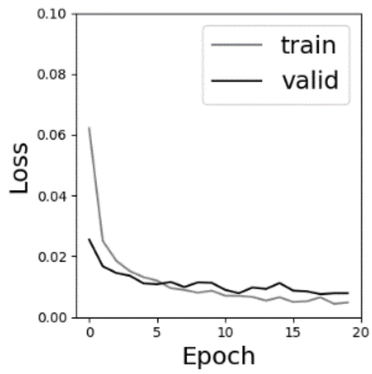

Figure 5. U-Net learning curve

Figure 6. Examples of (segmented) images in datasets

128x96 pixels. The generated 10200 data sets include 10000 for train, 200 for test.

Figure 4 shows U-Net, which is the neural network for semantic segmentation. U-Net realizes the function of extracting higher-order features by narrowing down the amount of information in the deepest layer of the U-shape, while creating several skip layers to create a structure that does not lose detailed information. The input data is a 4ch (ARGB) image. In an ARGB image, in addition to 3 colour layers (Red, Green, Blue), there is an alpha layer that represents transparency. Each pixel takes an integer value from 0 to 255. All alpha layers of normal images are 255 (opaque). In the figure 4, “C” represents a 3x3 convolutional layer, “P” represents a 2x2 max pooling layer, and “UP” represents a 2x2 unpooling layer. The output is a binary value indicating whether it is a building or not.

Figure 5 shows the loss curve. Loss converged well after 20 epochs of training.

Figure 6 shows examples of segmented images. Original real images and U-Net outputs are merged to segmented images. In this study, the accuracy of damage evaluation based on three methods are compared. First, the original image data are used (method A).Then two cases of segmented image are generated through U-net. The one is “overlay” image (method B), which is made by copying segmentation result to alpha channel of original image. The other is “blackout”, made by multiplying original images’ RGB data with segmentation data at each pixel (method C). The dataset includes 628 pictures of damaged buildings and 792 pictures of undamaged buildings. The photos of damaged buildings are taken in the areas that suffer severe damage due to the Niigata Prefecture Chuetsu Earthquake in 2004, the Niigata prefecture Chuetsu-oki Earthquake in 2007, the 2010 Chile Earthquake, the 2011 off the Pacific coast of Tohoku Earthquake, the Koshigaya city tornado disaster in 2013, and the 2016 Kumamoto Earthquake.

Although there are some cases in which the sky and trees have been extracted from the image, the evaluation results of the building position by U-net are generally of good accuracy.

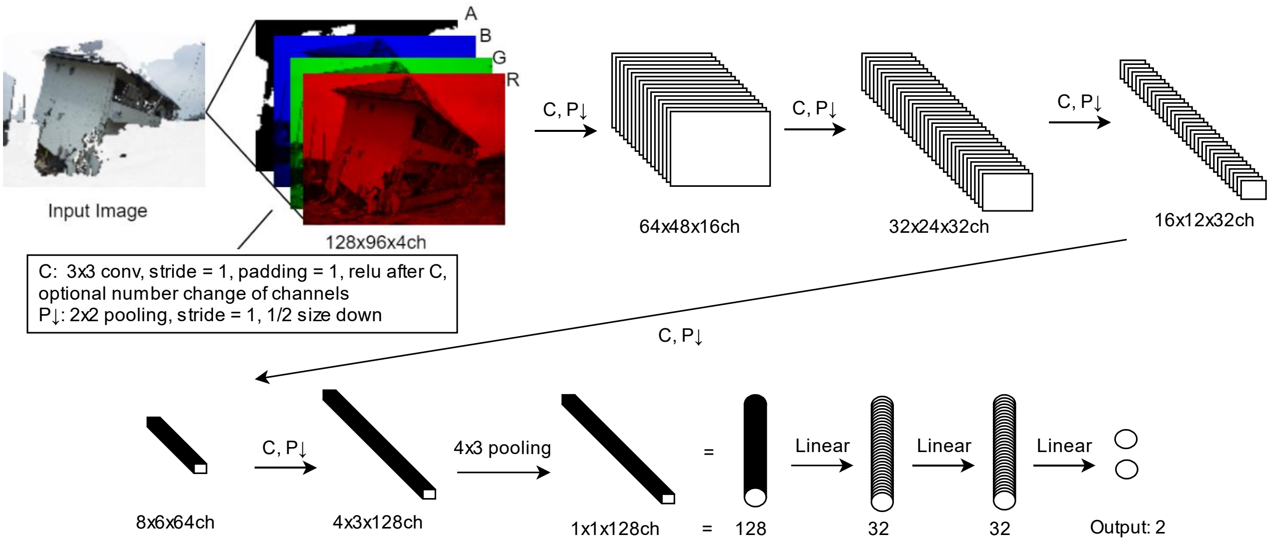

Figure 7. CNN configuration

4. Building damage evaluation with CNN

In this chapter, the damages of buildings are judged by using CNN. The accuracy of judgement of three cases shown in figure 6 are compared.

4.1 Cross Validation

Figure 7 illustrates the CNN model used in this paper. The input is an image (size of 128x96, 4 layers RGBA), the output data are 2 values whether the building is damaged or undamaged.

We use cross validation in order to evaluate the performance of CNNs. The advantage of using cross validation is that all data are eventually treated as test data. This means that all data can be analyzed.

Figure 8 shows the concept of cross validation. All the data, which are data sets consist of image and answer (damaged or undamaged,) are divided into 10 data groups. Then, the data sets are input to 100 CNN. The initial values of weights of each CNN was set to random numbers. Therefore, cross validation is conducted for 100 times in order to get damaged rate of image in steps of 1%. The CNNs use 9 groups as train data and 1 group as test data. Finally, all the images are treated as test data. Through this process, the randomity of CNN result are reduced.

Cross validation is conducted 100 times for each dataset. all CNNs learn 30 epochs. As a result, each image is judged by 100 CNN. Then, the accuracy of damage evaluation is calculated as number of correct judgements divided by 100. Finally, the average accuracy is calculated as an average of accuracy evaluated from the 10 accuracies.

4.2 Comparison of accuracy of Methods

The average accuracies are shown in Table 1.

Table 1. Average accuracies of methods

method A	method B	method C
88%	88%	75%

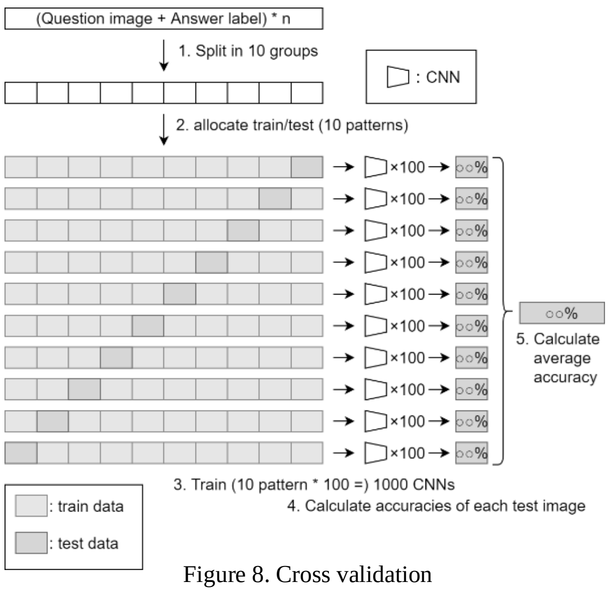

Figure 8. Cross validation

The accuracies of methods A and B are 88%, while the accuracy of method C is 75%. This is probably because method C lose a lot of information of images.

In order to investigate the validity of method A and B, we use Grad-CAM (Ramprasaath et al. 2016) technique to consider which part of the image the CNN focused to evaluate the damage. Figure 9 shows the examples of the average Grad-CAM. The average of 100 Grad-CAM is the most reliable at the point of where all CNNs are generally looking at.

As average Grad-CAM shows, CNN occasionally looks more likely at buildings with method B than A. Though the accuracy of method B is the same as that of method A, method B is more proper to judge the damage of the buildings due to an earthquake.

5. Geographical distribution of building damage

Kumamoto Earthquake occurred in April 2016 in Japan. A number of buildings suffered severe damages due to the

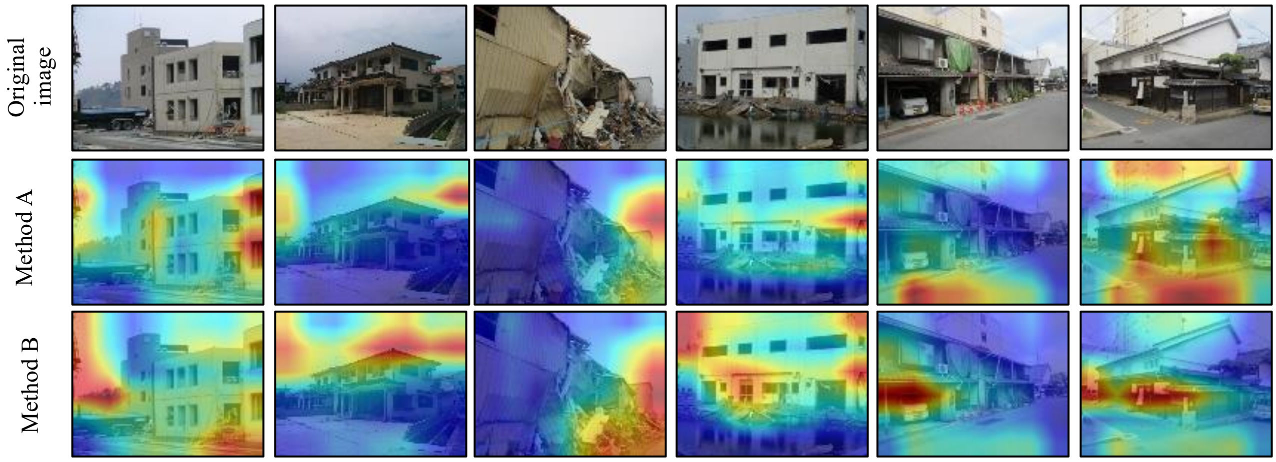

Table 9. average Grad-CAM

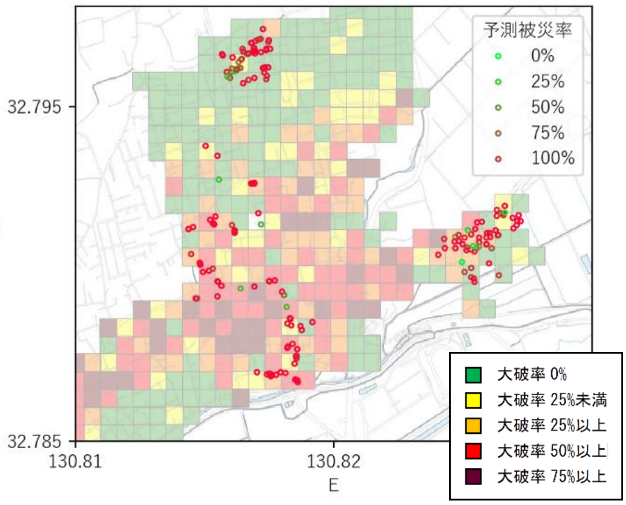

Figure 10. Building damage evaluation at Mashiki town with method A

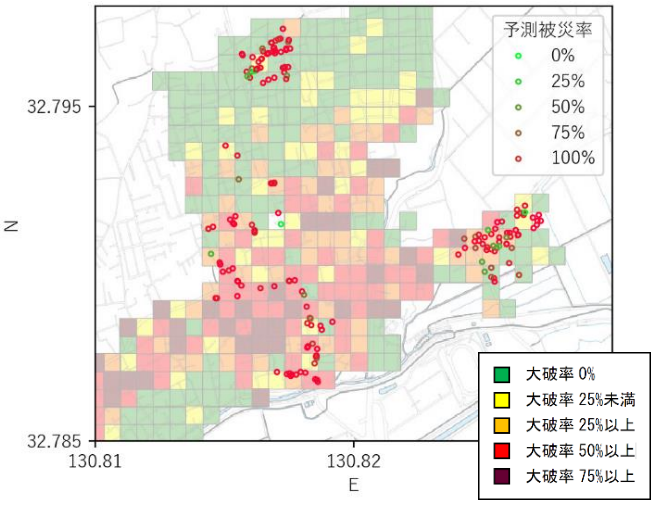

Figure 11. Building damage evaluation at Mashiki town with method B

earthquake shaking, especially in Mashiki area. Figure 10 and 11 show geographical distribution of building damage evaluated by CNNs. The collapse rate was surveyed by the National Institute for Land and Infrastructure Management (NILIM 2016). The collapse rate means the proportion of the number of collapsed buildings in the areas separated by mesh. NILIM survey defines “collapse” as damaged more than damage grade D4 that means the building is unrepairable. The size of the meshes are about 57m square.

Although there is almost no difference between the two figures, as shown earlier, Method B is more appropriate because it focuses on the building and evaluates the damage. In the north area, the collapse rate by NILIM does not match the CNNs evaluation. The damage grade is defined by Architectural Institute of Japan. D0 means not damaged, and D6 means completely collapsed. D4 means that the building is unrepairable. The CNNs were trained to evaluate whether the building is D0 or D1-D6. On the other hand, NILIM defines “collapse” as D4-D6. This can be the explanation for inconsistency between NILIM survey and CNN evaluation. In the southern region, where the severe damage rate is high according to the NILIM survey, there are many cases where the damage assessment by CNN also determines that there is damage. Thus, the damage assessment results by CNN are generally valid.

6. Conclusions

In this study, we introduced the semantic segmentation based on U-net methodology to image-based building damage evaluation with CNN. First, the 3D CG models were constructed to generate the data sets for training the U-net. Second, the building damage was evaluated through CNN based on the original image and segmented image by U-net. Then, the validity of damage evaluation result by using segmented image was shown based on Grad-CAM technique. Finally, the geographical distribution of building damage was visualized by comparing actual damage distribution and the results of building damage evaluated by CNN.

To improve the accuracy of semantic segmentation and to evaluate multi-grade of building damage are future task.

References

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), pp. 2278-2324.

Disaster Countermeasures Office of Fire and Disaster Management Agency: Damage caused by Typhoon No. 15 in 2019 and response status of fire department organizations (1st report) to (15th report). 2019. (in Japanese)

Ronneberger, O., Fischer, P., and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention.* pp. 234-241.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision.* pp. 618-626.

Takenori Hida, Taro Yaoyama and Tsuyoshi Takada. 2018. Damage evaluation of building via convolutional neural network. In *the 32th annual Conference of the Japanese Society for Artificial Intelligence*, Kagoshima, June 5-8, 2018. pp. 3K1OS18a03. (in Japanese)

National Institute for Land and Infrastructure Management (NILIM). 2016. Report of the committee on research of the causes of building damage in the Kumamoto earthquake.