

Evaluation of Building Damage due to Natural Disaster using CNN and GAN

Haruka Yamada^{1, a*}, Takenori Hida^{2, b}, Xin Wang^{3, c}, and Masayuki Nagano^{1, d}

¹ Tokyo University of Science, Noda, Chiba, Japan

² Ibaraki University, Hitachi, Ibaraki, Japan

³ Ashikaga University, Ashikaga, Tochigi, Japan

^a7117115@alumni.tus.ac.jp, ^btakenori.hida.mn75@vc.ibaraki.ac.jp,

^cwang.xin@g.ashikaga.ac.jp, ^dnagano-m@rs.tus.ac.jp

Keywords: Building Damage, Deep Learning, Convolutional Neural Networks, Generative Adversarial Networks, Image Recognition, Grad-CAM

Abstract. After destructive natural disasters, it is necessary to quickly grasp the damage situation for the initial response. In recent years, studies on the method of the automatic evaluation of building damages due to disasters using the convolutional neural network (CNN), which is a deep learning methodology for image recognition, were conducted. In these studies, it was clarified that a large number of images are necessary to train the CNN with sufficiently high accuracy. However, the number of images of damaged building is limited. Therefore, in the present study, we used the generative adversarial network (GAN) to automatically generate a large number of imitation images of damaged and undamaged buildings and trained the CNN using imitation images to obtain a higher accuracy rate of the CNN. Then, the validity of the CNN for judgment of “damaged” and “undamaged” using imitation images was confirmed. In addition, photographs of actual buildings were input to the trained CNN as test data.

Introduction

In order to properly allocate people and equipment for emergency activities and emergency response after natural disasters such as earthquakes, it is essential to quickly evaluate the building damages after disasters. Usually, the emergency risk judgment and damage classification of buildings are visually determined by experts after disasters. However, it tends to take a long time to grasp the damage level of all buildings affected, because of the shortage of experts in rapid risk assessment and the widespread distribution of the damaged area. For example, in the case of the 2016 Kumamoto earthquake, damage inspection by May 16 was done for 41,907 buildings, which is only approximately 30% of the total number of damaged buildings inspected by June 16 [1]. It is difficult to grasp the whole picture of building damages only by human inspection in a short period just after massive earthquakes.

To solve this problem, several studies aiming to shorten the damage assessment period using the convolutional neural network (CNN [2]), which is a methodology of deep learning for image recognition, have recently been conducted. In these studies, the degree of damage of buildings is automatically determined from photographs of the exterior [3]. For training of the CNN, a large amount of image data is essential to improve the accuracy of judgment. However, preparing a large amount of building damage images is difficult because destructive disasters rarely occur. In order to solve these problems, Fujiu et al. [4], Yamaguchi et al. [5], and the present authors [6] have used 3DCG models to generate a large number of images of buildings. The present authors pointed out that the assessment accuracy of CNNs trained by 3DCG- imitation images was low because the similarity between 3DCG building models and actual buildings was low. Time and effort are



required to replicate various types of damaged buildings in detail in 3DCG that are comparable to those in actual buildings.

Therefore, in the present study, we use the generative adversarial network (GAN [7]) to easily generate many imitation images with high similarity to actual building images and used these images to train the CNN for building damage assessment. In addition, to validate the proposed method, photographs of actual damaged and undamaged buildings, were input to the trained CNN for building damage assessment.

Generation of imitation images of damaged and undamaged buildings using the generative adversarial network (GAN)

In this study, the GAN is used to generate numerous imitation images of damaged and undamaged buildings. The network structure of the GAN used in the present study is shown in Figure 1. The GAN is a neural network consisting of two parts. One is the generator, which generates imitation image data, and the other is the discriminator, which judges whether the image is a real image.

The model used for image generation in the present study is FastGAN [8], which is one of the least computationally intensive GANs and can generate images with relatively little image data. In the process of sequential up-sampling, the generator uses the *skip-layer excitation* module to fuse the most recent feature map with those of four lower levels, improving stability without increasing the computational complexity.

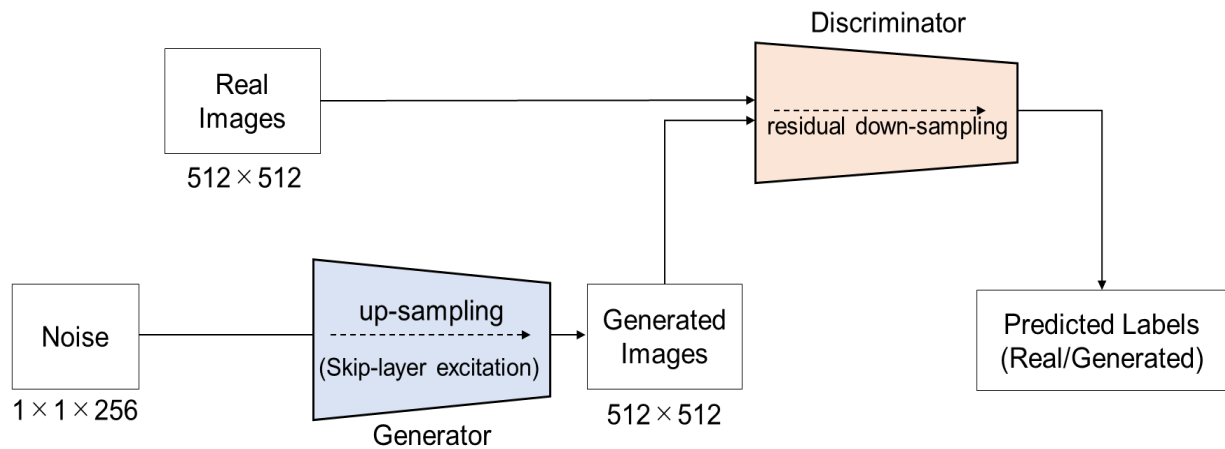


Figure 1. Network structure of the generative adversarial network (GAN) based on FastGAN [8].

The mutual learning process of GAN is generally expressed by the following equation:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{x \sim p_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

where V is the objective function, x and z are input data, and noise variables, $D(x)$ and $G(z)$, are the output of the discriminator and generator, respectively, p_{data} and p_z are distributions of training data and the noise variables, respectively, and \mathbb{E} represents the expected value. The generator learns to minimize the objective function so that it generates data that the discriminator cannot distinguish from the training data. On the other hand, the discriminator maximizes the objective function so that the discriminator can classify the training data and the generated data with high accuracy.

In the present study, two types of GANs are constructed to generate images of damaged and undamaged buildings by preparing two types of photographs. The photographs of damaged

buildings for training were taken after the 1995 Kobe earthquake, the 2011 Tohoku earthquake (including tsunami damage), and the 2016 Kumamoto earthquake, in Japan. Some of the photographs of undamaged buildings used in the training are of actual undamaged buildings taken in urban areas and some were obtained from the Internet. The sizes of images used in the present study were modified to 512×512 pixels. A total of 4,040 photographs, consisting of 1,995 and 2,045 photographs of damaged and undamaged buildings respectively, were given to the GAN. The photographs in Figure 2 show examples of input data. The batch size was set to 10. The gradient accumulation was set to 4. The learning rate was set to 0.0002, and the number of epochs (number of trials) was set to 150,000.

The photographs in Figure 3 are examples of imitation images. In the images with 1,000 epochs, trees and debris are reproduced fairly well, but the building images are completely unacceptable. In the image with 100,000 epochs, trees, roads, and buildings are created fairly well, and the images are relatively similar to photographs of the actual items. However, the overall images are not excellent, suggesting that learning is still insufficient. When the number of epochs reaches 150,000, buildings, debris, and trees are clearly represented, and the imitation images are approximately equivalent to photographs of actual items. Therefore, we use the generated images by the GAN with 150,000 epochs to train the CNN, which is described below.



Figure 2. Actual building photographs input for the generative adversarial network (GAN).

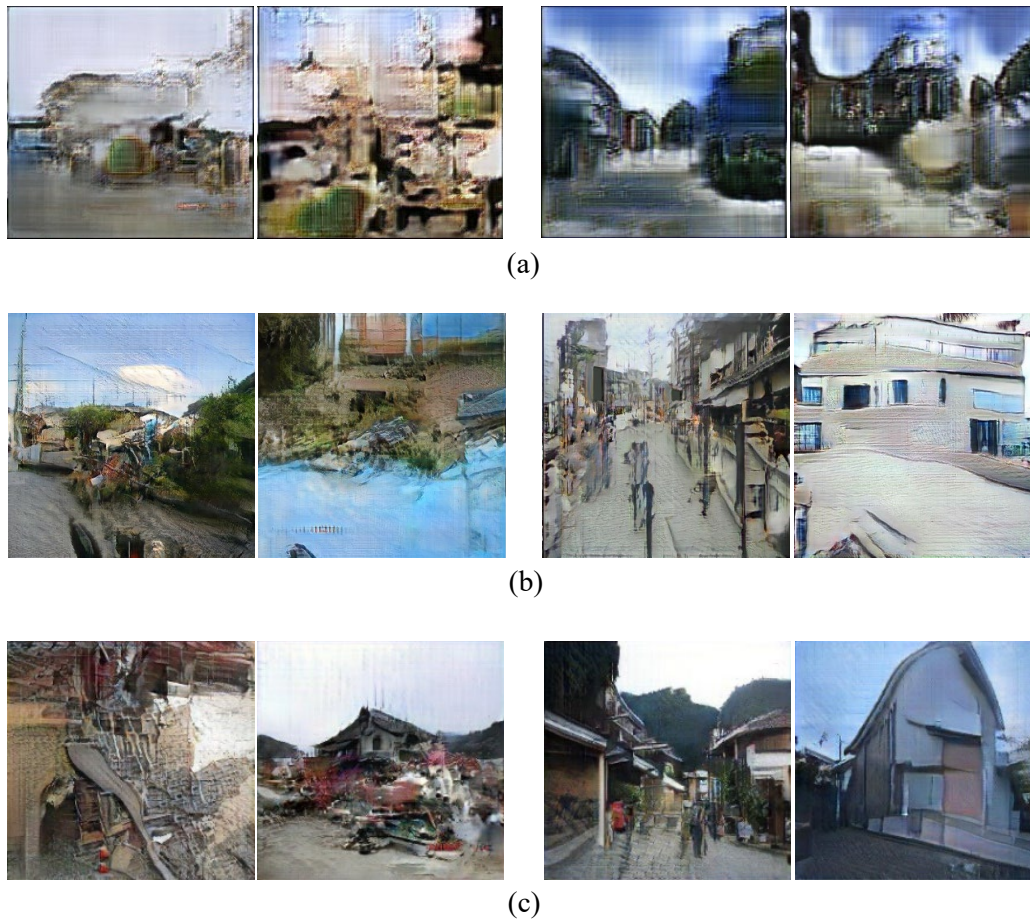


Figure 3. Examples of damaged and undamaged building images generated by the GAN: (a) 1,000 epochs, (b) 100,000 epochs, and (c) 150,000 epochs.

Training and validation of the convolutional neural network (CNN) for damage evaluation using imitation images of buildings

In this section, the CNN model used in the present study is introduced. Then, the results of CNN training and validation for damage evaluation using imitation images of buildings will be presented.

Overview of the CNN

We construct the CNN, which takes advantage of the fact that many data have local features, as proposed by LeCun et al. [9], to determine whether a building is damaged.

In the present study, we use Caffe [10], a deep learning framework, and AlexNet [11] as the CNN architecture. The structure of AlexNet in Figure 4 consists of five convolutional layers, three pooling layers, and three fully connected layers. Image data were previously resized to 256×256 pixels, converted to normalized RGB values, and input to the CNN. The training data were classified into two categories, “damaged” and “undamaged,” with “damaged” set to 0 and “undamaged” set to 1, and the learning rate was set to 0.01.

The convolutional layer detects local features of the image data by convoluting the input image with filters. The pooling layers are used to achieve invariance to small movements, thereby transforming the input data into a more manageable form. AlexNet uses max pooling in the pooling layer, which is a function that outputs the value of the largest component of the input. The fully connected layers combine the image data from which the feature portions were extracted in the convolutional and pooling layers into a single node and outputs the values transformed by the

activation function. The output of the last of the three layers, the fully connected layer, contains a SoftMax function that transforms the output score of each image into the probability between 0 and 1.

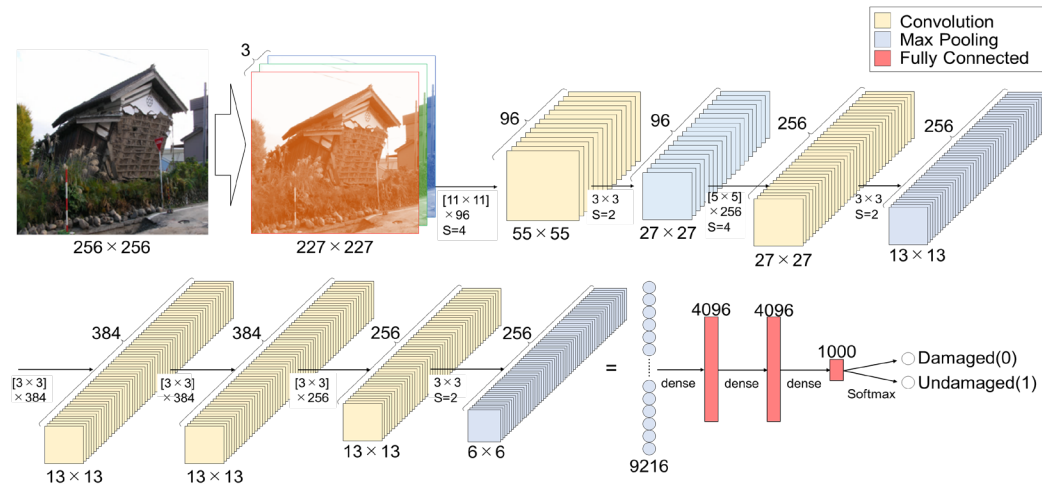


Figure 4. The structure of the CNN based on AlexNet [11].

Convolutional neural network training using imitation images generated by the GAN

The CNN was trained and validated using a total of 300,000 imitation images of buildings, 150,000 each of images of damaged and undamaged building generated by the GAN. Here, 200,000 images were used for training, and the remaining 100,000 images were used to verify the training results. The training results in Figure 5 show that the accuracy of the validation data for 10 epochs was 99.56%, and the loss function value was 0.012. This indicates that the learning process progressed well. The accuracy of the validation data for the first epoch was 98.28%, and the loss function value was 0.045. One possible reason for this is that the training data of 200,000 images were sufficient.

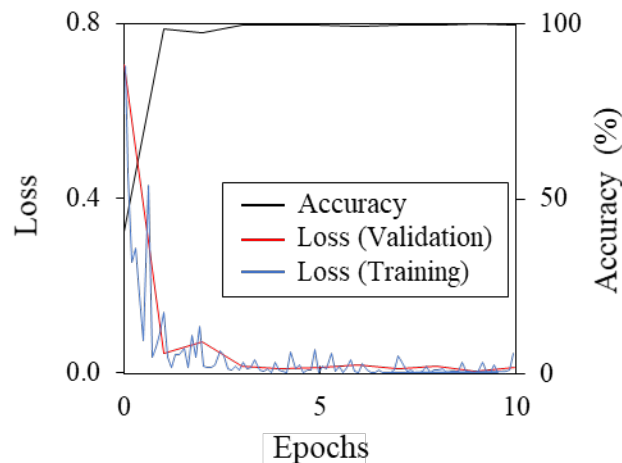


Figure 5. Convolutional neural network training and validation result with building imitation images.

Validation of CNN using photographs of actual buildings

In this section, photographs of actual damaged and undamaged buildings were input to the trained CNN as test data to examine the validity of the damage assessment of the CNN trained by images generated through the GAN.

Validation using photographs of actual buildings and evaluation indices

A total of 1,000 photographs of actual buildings, 500 each for damaged and undamaged buildings, were input as test data to verify the accuracy of the classification results.

The test results in Table 1 suggest that the model developed in the present study is generally suitable for determining both damaged and undamaged buildings for photographs of actual buildings. However, 57 of 500 (11.4%) photographs of damaged buildings were determined to be of undamaged buildings, whereas 102 of 500 (20.4%) photographs of undamaged buildings were determined to be of damaged buildings. Figure 6 shows examples of photographs of actual buildings that were misclassified. The damaged buildings tended to be determined to be undamaged when the building experienced story collapse, the damage was of a small to moderate level, or there was little debris. The undamaged buildings appear to be judged to be damaged when electric wires and trees were included in the photographs.

Table 1. Convolutional neural network test results for photographs of actual buildings.

Estimation / Correct	Damaged	Undamaged
Damaged	443	102
Undamaged	57	398

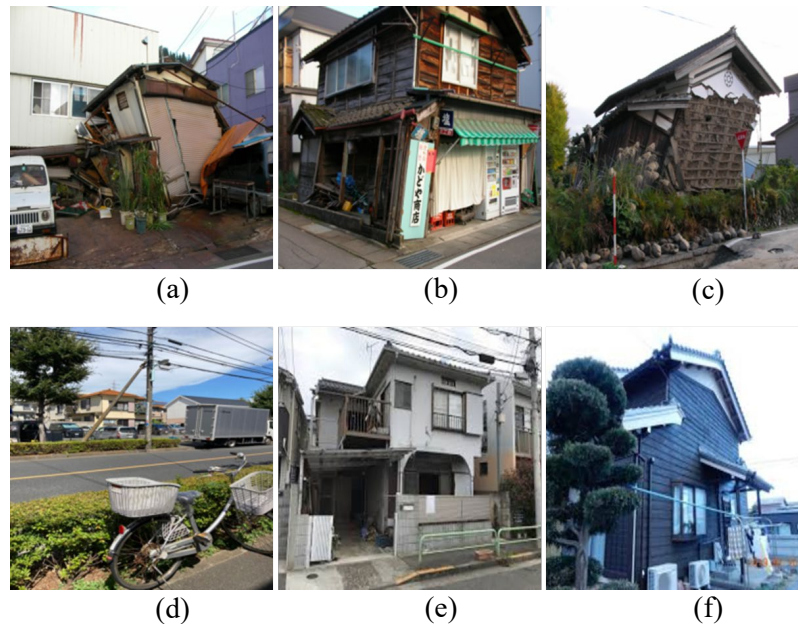


Figure 6. Examples of photographs in which the actual building were misclassified: (a) through (c) are photographs of damaged buildings judged to be undamaged buildings, and (d) through (f) are photographs of buildings judged to be damaged.

Next, the indices for evaluating the test results are examined more specifically. The evaluation indices are accuracy, precision, recall, and F-measure, which are calculated by the following equations:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

where TP (true positive) is the number of damaged buildings determined to be damaged by the CNN, TN (true negative) is the number of undamaged buildings determined to be undamaged, FP (false positive) is the number of undamaged buildings judged to be damaged, and FN (false negative) is the number of damaged buildings judged to be undamaged.

In Figure 7, all the indices are higher than 0.8. In particular, recall was higher than precision. Although the ratio of damaged buildings judged to be damaged was high, relatively many undamaged buildings were also judged to be damaged. Precision, which is the ratio of the number of photographs of correctly classified damage to the number of photographs of improperly classified damaged, was 0.81, indicating that the proportion of false positives was low. Therefore, F-measure, the harmonic mean of precision and recall, is also generally reasonable. The accuracy was 0.84, which means that all of the predicted results showed good agreement with the correct results. These results indicate that the CNN trained by the imitation images generated by the GAN was assessed validly.

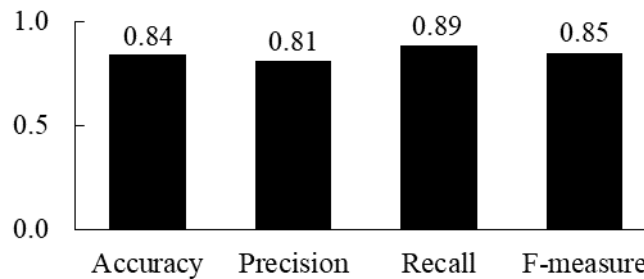


Figure 7. Indices for evaluating the test results.

Visualization of the evidence of predictions using Grad-CAM

In deep learning, the judgment of constructed models is often regarded as a black box, and it is often difficult to explain the evidence for the judgment. Therefore, in order to examine the cause for the judgments of misclassification of photographs of actual buildings by the trained CNN, we applied gradient-weighted class activation mapping (Grad-CAM [12]). This is a technique to visualize the important regions in the image, based on the idea that the areas that contribute most to output values of the predicted class are important for classification. In general, the gradient of the output value in the predicted class of the final convolutional layer is used.

Figure 8 shows visualization examples of misclassified by applying Grad-CAM to photographs of damaged and undamaged buildings. First, in visualization of discriminative regions in photographs of actual damaged buildings, there were cases in which the gradient of the concepts

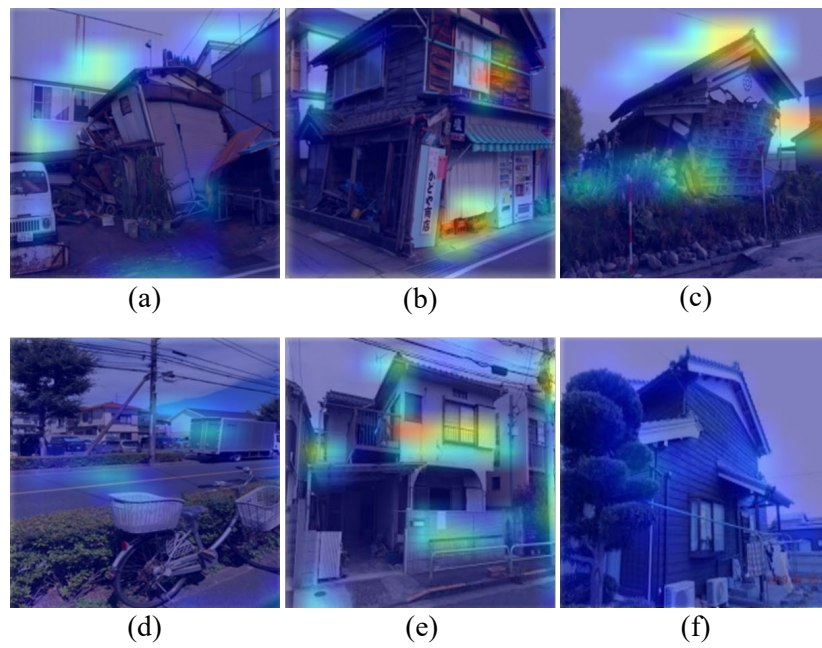


Figure 8. Examples of visualization of the evidence of predictions using Grad-CAM to photographs of actual buildings that were misclassified: (a) through (c) are photographs of damaged visualized areas of interest, and (d) through (f) are photographs of undamaged visualized areas of interest.

was large not only for the target building in the image, as in Figure 8(b), but also for buildings other than the target building and the sky, as in Figures 8(a) and 8(c). Next, in visualization of discriminative regions in photographs of actual undamaged buildings, in Figures 8(d) and 8(e), the gradient in the center of the photograph is larger, indicating that this region corresponds to the building, respectively. On the other hand, in some cases, such as Figure 8(f), the overall gradient was small and the evidence for the predictions was unclear. Therefore, for more accurate judgments, it is considered effective to apply a method such as semantic segmentation [13] to both the training and test images to mask the areas other than buildings and focus only on buildings.

Conclusions

In the present study, we firstly generated imitation images of buildings using the GAN, which is an image generation method based on deep learning, to easily generate a large amount of exterior image data of the damaged and undamaged buildings. Then the generated images are used as the input data for training a CNN that determines damage to buildings caused by a natural disaster. The accuracy of the assessment based on indices such as accuracy, precision, recall, and F-measure are investigated. Finally, the evidence of predictions of the CNN was investigated through Grad-CAM.

We found that images generated by GAN with 1,000 and 100,000 epochs were generally distorted and poorly learned, whereas the imitation images at 150,000 epochs had a quality almost equivalent to photographs of actual items.

The accuracy rate when the validation data were input to the CNN trained by the imitation images generated by the GAN was 99.56% for 10 epochs. This result indicates that the training was successful. Furthermore, the accuracy rate of the validation data for the first epoch was 98.28%, presumably because the training data of 200,000 images was sufficient.

Furthermore, when the CNN was tested using photographs of actual buildings, the model developed in the present study is generally suitable for determining both damaged and undamaged buildings. The CNN tended to classify undamaged buildings with electric wires and trees as

damaged ones. Whereas the damaged buildings with story collapse, small to moderate damage, and little debris were classified as undamaged.

In addition, when Grad-CAM was applied to the CNN, we found that some images had a large gradient for the target building, but in other cases, the gradient of the predictions for other buildings was larger or the overall amount of the gradient was smaller. In the future study, to improve the accuracy of the CNN, it is considered to be effective to mask the areas other than buildings and have the CNN focus only on buildings.

References

- [1] Cabinet Office, Government of Japan: Damage from the earthquake centered in Kumamoto region of Kumamoto Prefecture, (as of 8:00 a.m. on April 20) to (as of 5:15 p.m. on June 16), 2016. (in Japanese)
- [2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner: Gradient-based learning applied to document recognition, *Proc. of the IEEE*, pp.2278-2324, 1998. <https://doi.org/10.1109/5.726791>
- [3] T. Hida, T. Yaoyama and T. Takada: Building's Damage Evaluation and its Validation through CNN and Grad-CAM, *JCOSSAR 2019*, Tokyo, Japan, 2019. (in Japanese)
- [4] M. Fujiu, M. Ohara, S. Nakayama, and J. Takayama: Development of Building Damage Assessment Lessening Application Using 3d Modeled House, *JSCE Collection of papers (A1) Vol.71, No.4*, pp.I_865-I_872, 2015. (in Japanese). https://doi.org/10.2208/jscejsee.71.I_865
- [5] M. Yamaguchi, T. Hida, T. Itoi, and T. Takada: Image-Based Building Damage Evaluation Based on Semantic Segmentation and Convolutional Neural Network, *The Seventh Asian-Pacific Symposium on Structural Reliability and Its Applications, APSSRA2020*, 2020.
- [6] H. Yamada, T. Hida, X. Wang, T. Yamashita, M. Nagano: Damage evaluation of buildings after natural disaster based on 3DCG and deep learning technique, *Summaries of Technical Papers of Annual Meeting AIJ*, pp.13-14, 2021. (in Japanese)
- [7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio: Generative adversarial nets, *Proc. NIPS 2014*, pp.2672–2680, 2014.
- [8] B. Liu, Y. Zhu, K. Song, A. Elgammal: Towards Faster and Stabilized GAN Training for High-fidelity Few-shot Image Synthesis, *ICLR 2021*, 2021.
- [9] Y. LeCun, Y. Bengio: Convolutional networks for images, speech, and time series, *The Handbook of Brain Theory and Neural Networks*, Vol.3361, pp.255-258, 1995.
- [10] Y. Jia, E. Shelhamer, J. Donahue, and S. Karayev: Caffe: Convolutional architecture for fast feature embedding, In *Proc. ACM Int. Conf. on Multimedia*, pp.675-678, 2014. <https://doi.org/10.1145/2647868.2654889>
- [11] A. Krizhevsky, I. Sutskever, G. E. Hinton: ImageNet Classification with Deep Convolutional Neural Networks, In *Proc. of NIPS*, 2012.
- [12] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, Devi Parikh, and D. Batra: Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization, 2017 *IEEE International Conference on Computer Vision*, 2017. <https://doi.org/10.1109/ICCV.2017.74>
- [13] O. Ronneberger, P. Fischer, and T. Brox: U net: Convolutional networks for biomedical image segmentation, *arXiv:1505.04597*, 2015. https://doi.org/10.1007/978-3-319-24574-4_28