Using 3-D Convolution and Multimodal Architecture for Earthquake Damage Detection Based on Satellite Imagery and Digital Urban Data

Takashi Miyamoto and Yudai Yamamoto

Abstract—**When a large earthquake occurs, it is quite important to quickly figure out the damage distribution of housing structures for disaster prevention measures. Currently, the information is confirmed manually by local public organizations, which takes a lot of time. Therefore, a method is required for gathering the information more swiftly and objectively. In this work, a novel method for detecting damage to single buildings from a set of multitemporal satellite images is developed by applying a recent machine learning approach. The damage detection system is designed as a deep learning model that uses multimodal data, consisting of optical satellite images and structural attributes. The proposed method achieved over 90% detection accuracy on damaged housing in the affected area of 2016 Kumamoto earthquake, Japan from satellite images taken by Pleiades as well as digital urban data.**

Index Terms—**3-D convolution, earthquake damage detection, multimodal learning, satellite imagery, spatiotemporal data.**

I. Introduction

PREPAREDNESS for and mitigation of natural disasters that frequently occur throughout the world is one of society's most pressing problems. Owing to the difficulty of earthquake predictions and the scope of damage incurred, earthquake mitigation continues to be the focus of disaster planning in Japan and many other countries, despite the infrequency of earthquakes. Earthquake disaster planning is broadly composed of four chronological phases: prevention, preparation, response, and reconstruction. Among these, the response phase refers to activities conducted immediately after the disaster, such as search and rescue operations and the provisioning of necessary supplies. These activities are particularly important for earthquakes where the complete prevention of damage is difficult. In the response phase, promptly identifying the state of damage—the earthquake location and damage degree—is crucial as it enables the appropriate allocation of resources and sound decision-making by individuals and disaster support organizations. Moreover, such a prompt assessment can optimize subsequent activities.

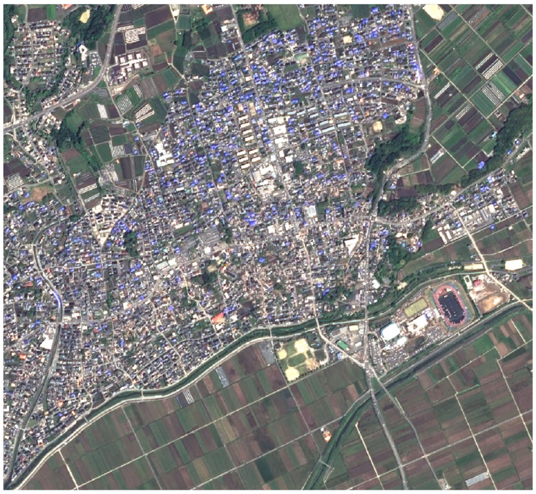
Manuscript received May 30, 2021; revised July 19, 2021; accepted July 28, 2021. Date of publication August 5, 2021; date of current version September 9, 2021. This work was supported in part by JSPS Kakenhi under Grant JP18K13966 and in part by ACT-I, JST. *(Corresponding author: Takashi Miyamoto.)*

The authors are with the Department of Civil and Environmental Engineering, University of Yamanashi, Yamanashi 4008511, Japan (e-mail: tmiyamoto@yamanashi.ac.jp; g20tca02@yamanashi.ac.jp).

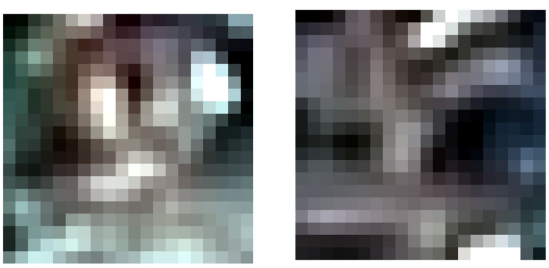
Digital Object Identifier 10.1109/JSTARS.2021.3102701

Information related to earthquake damage is generally gathered through human efforts, such as on-site verification by public organizations. However, research is ongoing for a faster method, which specifically involves the use of remote sensing data, whereby information is gathered by sensors mounted on airplanes, as a means of gathering such information. An effective means of obtaining information on the overall postdisaster situation of a wide area involves the analysis of data collected from satellites equipped with optical sensors capable of high-definition imagery or synthetic aperture radar (SAR) capable of photographing the Earth's surface regardless of weather or time of day [1], [2]. Recent research and development in satellite technology include the development of next-generation optical satellites and the establishment of on-demand launching systems for compact SAR satellites during disasters. These developments have been implemented practically in various disaster-preparedness initiatives. Implementation examples include the Ministry of Land, Infrastructure, Transport and Tourism's satellite data usage guidelines for flooding and sediment disasters, and the construction of specific operational systems.

As with other natural disasters, research attempting to analyze the state of earthquake damage from the satellite sensor data has proceeded both in Japan and abroad. As reviewed in detail by Dong *et al.* [3], data obtained from photo-optical, SAR, as well as light detection and ranging sensors have been used to detect structural damage from earthquakes via remote sensing. Gong *et al.* [4] proposed to use a combination of SAR imagery and footprint map for building damage detection. Tong *et al.* [5] proposed a method using the difference in elevation in each structure before and after an earthquake as determined via DEMs, while Matsuoka *et al.* [6] proposed a method for determining structural damage via threshold values after calculating individual differences in the SAR backscattering coefficient. Additionally, a considerable amount of research has been conducted on methods that employ machine learning models to determine damage. Mansouri *et al.* [7] used machine learning models, such as support vector machine (SVM), based on feature values in differential images identified from optical images before and after a disaster. Bai *et al.* [8] distinguished damaged structures from undamaged structures by using the *K*-nearest neighbor method on feature values from differential images obtained before and after an earthquake using SAR. However, owing to the limited spatial resolution of satellite observation data,



(a)



(b)

Fig. 1. Example of satellite imagery of earthquake-suffered area: taken by Pleiades (0.5 m/pixel resolution) at the time of the 2016 Kumamoto earthquake, Japan.

these approaches are currently not sufficiently accurate to detect damage in micro-phenomena, such as the collapse of residential buildings and highway blockages, which is the major type of damage in earthquakes. Hence, increasing the accuracy of these approaches is a major challenge.

Given the above background, in this study, we investigated technology that can assess the state of damage in a single residential building using wide-area disaster photography on a scale that captures an entire city block. As illustrated in Fig. 1, while satellite images can capture the entire city, they do not have sufficient resolution to show the appearance of individual buildings clearly. Our proposed approach leverages satellite sensing through recent developments in machine learning, and it consists of two processes. First, by using the location and external information on structures stored in a GIS database, the range of photographs for each residence is identified from wide-area photography, and image fragments on the scale of a single residence are extracted. Second, the extent of damage in groups of residential structures in disaster areas is ascertained through classifiers that determine whether the structures in the extracted image fragments have been destroyed.

In this study, we examine the effectiveness of reflecting the following two ideas in the classifier in order to improve the performance of determining collapsed and noncollapsed buildings from satellite images. The first concept is to improve classification performance by inputting to the classifier the photographic images of the disaster area immediately before and immediately after an earthquake. This study examined the effectiveness of further generalizing the application of image difference values (an approach that has been implemented widely in the past) to the use of spatiotemporal convolutional layers. The second concept is to use two types of information other than satellite imagery, the age and materials of the structures, that could be employed in damage detection. They are inputted to multimodal frameworks and the results are assessed.

II. Related Works

The detection of earthquake damage from satellite images can be regarded as a problem of image recognition by machine learning from the viewpoint of information science. Machine learning technology has significantly advanced in recent years through the development of deep learning [9]. Convolutional neural networks (CNNs) were proposed by Lecun *et al.* [10] and demonstrated for their effectiveness recently by Krizhevsky *et al.* [11]. Notably for image recognition, CNNs use backpropagation to create automatic filters to identify feature values from images for improving their effectiveness in sorting tasks. In recent years, numerous examples of CNN filters have been expanded to the third dimension to identify feature values in spatiotemporal data. Tran *et al.* [12] expounded on the differences between this approach and the stacking of groups of images in the feature dimension direction. They investigated the application of a two-dimensional (2-D) filter as well as the advantages and disadvantages of each method. Ji *et al.* [13] demonstrated the effectiveness of using 3-D CNNs in behavior recognition based on time-series images. Given this context, in the present study, a method was designed to improve earthquake damage-identification task-processing performance by applying these recent advances in machine learning to satellite sensing images.

Further, one of the recent developments in deep learning models is to utilize multimodal data [14], which improves the performance of the models for feature learning and task processing. In the detection of earthquake damage, it is expected that the detection performance can be improved by using not only satellite images but also various other data that can be obtained in advance. Therefore, in this study, we designed a neural network that can handle such multimodal data.

III. Proposed Scheme

A. Overview of the Earthquake Damage Detection System

The system developed in this study distinguishes damaged structures from undamaged structures in two stages, as discussed below. First, a geospatial information database providing structure locations and their footprints is used to identify the image fragments of each structure from wide-area satellite imagery. Next, to discern the state of damage on individual residences appearing in images obtained by satellites, a deep learning model

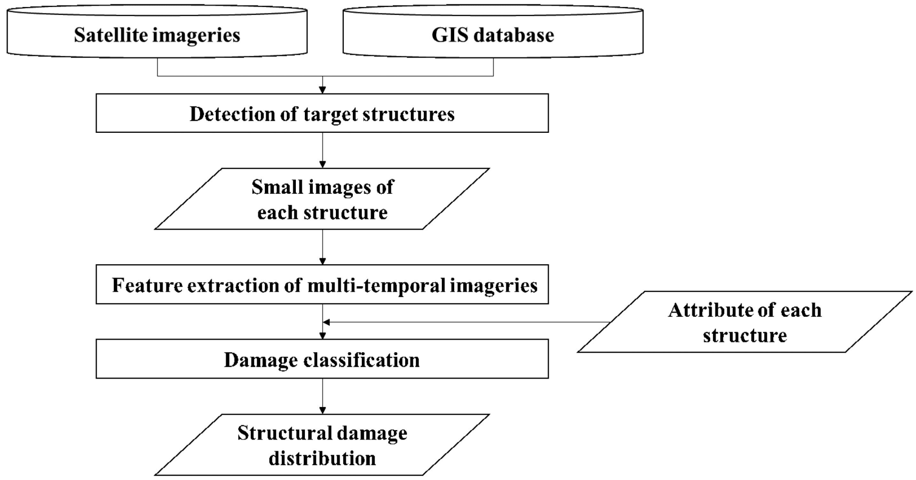


Fig. 2. Flowchart of the proposed scheme.

is used to categorize each fragment as being in one of two classes: damaged or undamaged.

In this damage detection process shown in Fig. 2, the first phase of identifying the location of buildings and extracting image fragments from satellite images can be performed with ease because geospatial information is sufficiently available in Japan, and satellite images are assigned spatial coordinates in pixel units. Therefore, the main challenge in the development of our system is to design a deep learning model that can accurately discriminate the presence or absence of damage in images of single houses, and we will discuss this in detail in the following sections.

B. Deep Learning Model

1) Feature Identification From Pairs of Photographic Images Obtained Before and After an Earthquake: Artificial satellites periodically photograph the Earth's surface; therefore, when detecting disaster damage, it is possible to compare data photographed before the disaster with that obtained immediately after it. Previous studies have reported improved detection accuracy by comparing information before and after a disaster through computing difference images. Computation of difference images is equivalent to applying a difference filter to a pair of images. However, the difference filter is only one of the various filters that can be used to extract features from image pairs. By applying more filters, more detailed features of image pairs can be extracted, and the performance of earthquake damage discrimination can be further improved. Therefore, in this study, we used a 3-D CNN for classification tasks involving time-series images.

The 3-D CNNs expand the 2-D filters of standard CNNs to the third dimension. Filters with a one-dimensional degree of freedom in the temporal direction and a 2-D degree of freedom in the spatial direction were used. In 3-D CNNs, backpropagation is used to create the shapes of the filters automatically, thereby automatically extracting the feature values of the chronological and spatial data that are useful for the given task. Unlike a 2-D CNN, which extracts features from a single image, a 3-D CNN extracts features from a time series of two or more images. A 3-D CNN extracts spatial features, such as edges and their temporal changes. Therefore, the application of 3-D CNNs is considered to be effective for detecting earthquake damage that appears as changes in the shape of structures before and after an earthquake. Target areas were classified as damaged or undamaged based on chronological image data, using a deep learning model that combines 3-D CNN feature extraction layers with the fully connected layers of a classifier.

2) Multimodal Framework Integrating Structural Information: Satellite remote sensing provides a wealth of information that can be used to estimate and understand damage after an earthquake, including both photographic imagery and ground structure and seismic intensity distribution. Structural information, such as construction age, has been hypothesized and statistically confirmed to be deeply connected to the presence of damage in individual buildings. Therefore, we used as inputs the aforementioned pairs of photographic images before and after the earthquake and the two types of structural information: construction age and structure type. A multimodal learning structure capable of identifying the presence of damage in an integrated manner was obtained from such heterogeneous information.

The structure of the proposed damage-identification deep learning model is shown in Fig. 3.

IV. Assessing the Developed System

A. Dataset

To assess the damage-detection performance of the designed deep learning model, a dataset consisting of satellite images of residences damaged by the 2016 Kumamoto earthquake was created. The images were obtained from Mashiki Town in Kumamoto prefecture. This dataset was assembled from images of 310 destroyed residences and 2030 undestroyed residences. Each residence was confirmed and labeled as "damaged" or "undamaged" based on the results of a comprehensive survey by the Architectural Institute of Japan (AIJ) [15]. For the satellite

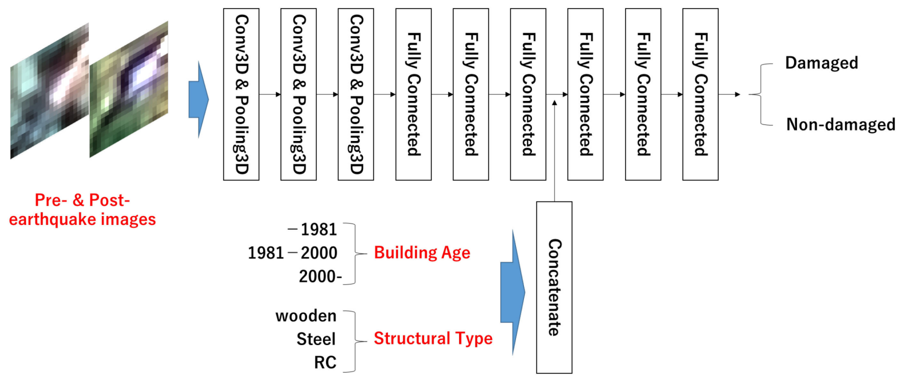


Fig. 3. Architecture of damage-identification neural network with 3-D convolutional layers and multimodal learning structure.

TABLE I
DAMAGE GRADE CHART OF THE BUILDINGS GIVEN BY THE FIELD SURVEY

Grade	Description
0	Undamaged
1	Cracks on the side walls or peeling off of exterior materials
2	Significant spalling of roof or wall mortar
3	Some columns, beams, or walls have been destroyed, but no interior space is missing
4	Some columns, beams, or walls have been destroyed, but no interior space is missing
5	Partially collapsed
6	Totally collapsed

TABLE II
NUMBER OF BUILDING DATA FOR THE CROSS-VALIDATION TEST: DATA AUGMENTATION IS APPLIED TO THE TRAINING DATA

	Non-damaged	Damaged
Training data	8120(=1015*8)	1240(=155*8)
Validation data	1015	155

images, the pairs of images used were captured before and after the earthquake, on January 16, 2016 and April 20, 2016, respectively, at optical satellite Pleiades. The resolution of the images was 0.5 m/pixel. Using the process described above, small images of each residence were extracted. Each fragment was then resized to 20–20 pixels to unify the image sizes. In addition, the labeling of the damaged structures was based on the damage grade scale (0–6), which was used by the AIJ's comprehensive survey (Table I). Structures assessed from 0 to 4 were labeled as "non-damaged," whereas structures assessed as 5 or above were labeled as "damaged."

Data on the construction age and structure type are typically managed by the local authorities. In this study, we used data collected and organized through a comprehensive survey. Construction age was classified into three categories based on the years of revision of Japan's Building Standards Act, as shown in Fig. 3. For the structure type, each structure was classified into one of four categories: "wood," "steel," "reinforced concrete," or "other."

To verify the accuracy of the deep learning model, the dataset was divided into training data and assessment data, as shown in Table II. It was then cross-validated. The training data were magnified 8× via mirror reflection and rotation.

B. Performance Assessment of 3-D Convolution

To verify the effectiveness of using 3-D CNNs as a feature extraction method from predisaster and postdisaster image pairs, we constructed a deep learning model with multiple input data formats and corresponding feature extraction layers as shown in Table III, and we evaluated the classification performance of the model.

No. 1-1 was a CNN learning and inference case for which an input tensor was created using only postearthquake images. Here, one input tensor had three dimensions—image width w, height h, and RGB 3-ch. The CNN model was equivalent to those used for standard image recognition.

No. 1-2 was a case in which an input tensor was constructed by calculating difference images from pre-earthquake and postearthquake images. Here, the number of tensor dimensions and the CNN model format were the same as those in No. 1-1; however, they differed from those in No. 1-1 in that pre-earthquake image information was used.

No. 1-3 was a case where a $w \times h \times 6$ ch input tensor was created by stacking pre-earthquake images to use as 4–6-ch data for postearthquake image RGB-dimension 3-ch data. Here, the calculation process was similar to that of a standard CNN, except that the number of feature values in the input layer was 6 ch. Compared with the No. 1-2 case, however, data were input as-is, without calculating the differences from before and after images; therefore, a richer quantity of information was expected to be obtained.

No. 1-4 was a case in which an input tensor was created as $w \times h \times 3$-ch spatiotemporal data through stacking before and after images in the temporal dimension. Here, the CNN calculation process was applied to three-dimensional data, and the created filters were expected to be able to capture through learning data features, such as spatial and temporal features, in more detail than in No. 1-3.

Comparing the performance of the models in the above four cases enabled the assessment of the suitability of various deep

TABLE III
OVERVIEW OF TEST CASES USED FOR PERFORMANCE ASSESSMENT OF 3-D CNN

No	Data used	Input tensor creation method	Input tensor dimensions	Deep learning model
1-1	Post-disaster images only	2D images	$w \times h \times 3$	2D CNN
1-2	Pre- & Post-disaster image pairs	Difference images	$w \times h \times 3$	2D CNN
1-3	Pre- & Post-disaster image pairs	Stacking in feature dimension	$w \times h \times 6$	2D CNN
1-4	Pre- & Post-disaster image pairs	Stacking in temporal dimension	$w \times h \times t \times 3$	3D CNN

TABLE IV
OVERVIEW OF TEST CASES USED FOR PERFORMANCE ASSESSMENT OF MULTIMODAL FRAMEWORK

No	Use of satellite images	Use of structural info	Classifier structure
2-1	Pre- & Post-disaster image pairs	N	3D CNN
2-2	None	Y	Support-vector machine
2-3	Pre- & Post-disaster image pairs	Y	3D CNN

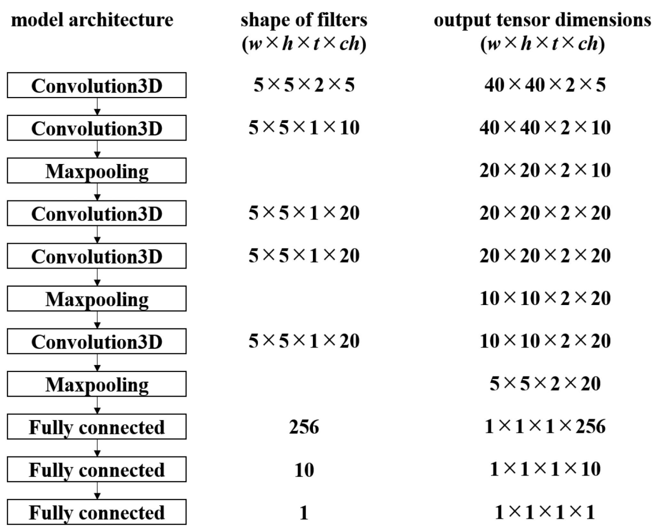


Fig. 4. 3-D CNN model structure (Case No.1-4).

learning models and data formats for identifying earthquake damage. Furthermore, the number of layers in the deep learning model in each case was uniform, and the number of internal parameters was mostly uniform for the purpose of comparison.

This comparative assessment also used random undersampling of image data in the undamaged class from the original dataset to use the same number of data items from each class. As an example of the model structure, the deep learning model used in Case No. 1-4 is shown in Fig. 4. For the other cases, the 3-D convolutional layers (3-D CNNs) in the figure were replaced with 2-D CNNs.

C. Performance Assessment of Multimodal Framework

Next, to confirm the effectiveness of the multimodal framework, a comparison of the damage-detection performances of the three machine learning models was carried out, as shown in Table IV. Model 2-1 used a 3-D CNN, similar to the proposed model; however, it did not introduce a multimodal framework and did not use structural information to distinguish between damaged and undamaged structures. Model 2-2 attempted to use only structural information via SVM to identify damage. Finally, Model 2-3 is the proposed model illustrated in Fig. 3.

TABLE V
COMPARISON OF DAMAGE DETECTION ACCURACY OF DIFFERENT METHODS FOR IDENTIFYING FEATURES FROM CHRONOLOGICAL IMAGES

No	Accuracy	Precision	Recall	ROC-AUC
1-1	0.64	**0.88**	0.60	0.72
1-2	0.69	0.56	0.75	0.79
1-3	0.70	0.67	0.72	0.78
1-4	**0.76**	0.67	**0.82**	**0.82**

Bold values indicate the highest metric values among the comparative cases.

The model uses a combination of a 3-D CNN and a multimodal framework.

For each model, learning and an assessment of damage-detection performance were conducted using the dataset presented in Table II. As the amount of data in each class in the training data used in model learning was skewed, a weighting coefficient proportional to the number of data items was introduced to the error function.

V. RESULTS AND DISCUSSION

A. Effectiveness of 3-D Convolution

The results of a comparative assessment of the accuracy of various methods of feature identification from pairs of images obtained before and after an earthquake in detecting earthquake damage are shown in Table V. First, Cases No. 1-2 to 1-4, which used pairs of before and after images, demonstrated overall superior performance compared with that of Case 1-1, which used only images captured after the earthquake. It is clear that the use of time-series satellite images is effective for detecting disaster damage. In particular, the 3-D CNN model in Case 1-4 demonstrated the best detection performance among the compared models. The 3-D convolution layer filter could extract more general features from the spatiotemporal data than image difference values. Therefore, the deep learning model is considered to have acquired the features that contribute to earthquake damage detection through the learning process.

B. Effectiveness of Multimodal Framework

The assessment results for the performance of each model in Table IV for categorizing the assessment data are presented in

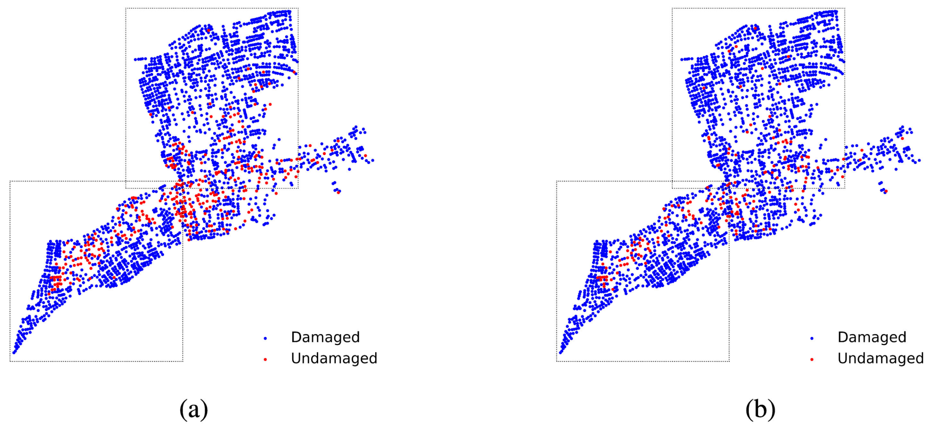


Fig. 5. Result of earthquake damage detection by the proposed scheme: The two boxes in the figures represent two 10-km square areas whose nature of damage distribution is discussed in the text. (a) Ground truth. (b) Inference.

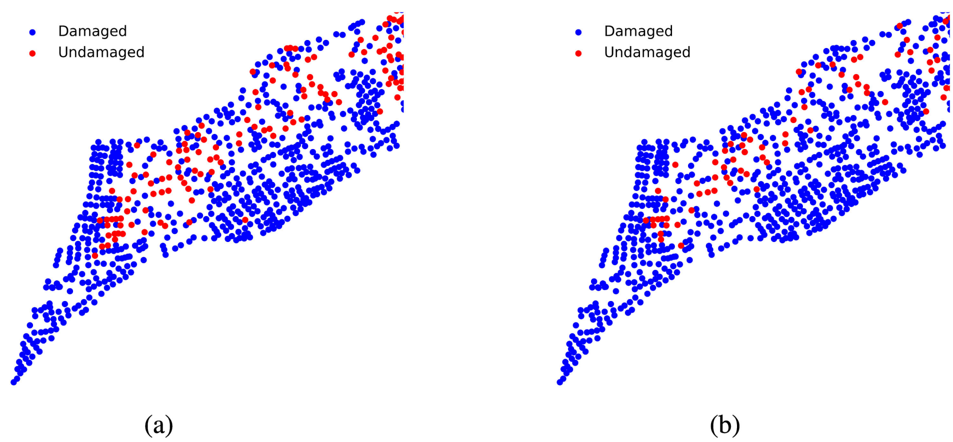


Fig. 6. Damage distributions in southwest district. (a) Ground truth. (b) Inference.

TABLE VI
COMPARISON OF DAMAGE DETECTION ACCURACY FOR PRESENCE/ABSENCE OF MULTIMODAL FRAMEWORK

No	Accuracy	Precision	Recall	ROC-AUC
2-1	0.72	0.26	**0.86**	0.82
2-2	0.70	0.27	0.74	0.74
2-3	**0.92**	**0.88**	0.47	**0.90**

Bold values indicate the highest metric values among the comparative cases.

Table VI. The performance of each model was evaluated using four metrics: accuracy, precision, recall, and ROC-AUC.

A comparison of the performance of each model reveals that the proposed model, Model 2-3, generally demonstrated the best performance in terms of the four metrics. This finding suggests that combining satellite imagery and structural information is more effective than using them independently. In particular, a high ROC-AUC metric is particularly desirable for detecting damage [16]. Fig. 5 illustrates a comparison of the inference result and actual residence damage conditions in Mashiki Town after the Kumamoto earthquake. A comparison with the on-site survey results reveals that the proposed method is capable of detecting damage trends for each district.

For a more detailed discussion, we review the distribution of damage and its prediction results in two 10-km square districts (southwest district and northward district) indicated by the boxed lines in Fig. 5. The southwest district, shown in Fig. 6, is characterized by an east-west band of damage through the district's center. The damage detection result of the proposed method captures this band-shaped damage and detects the damage trend of the district appropriately. In the northern part of the district shown in Fig. 7, the proposed method slightly underestimates the damage but reproduces the trend that almost no damage occurs in the northern part of the district compared with the southern part. Thus, it can be concluded that the proposed method has sufficient accuracy in understanding the trend of damage distribution in each area.

C. Characteristics of Misclassified Data

As shown in Table VI, the proposed method demonstrates an overall high percentage of correct hits; however, its

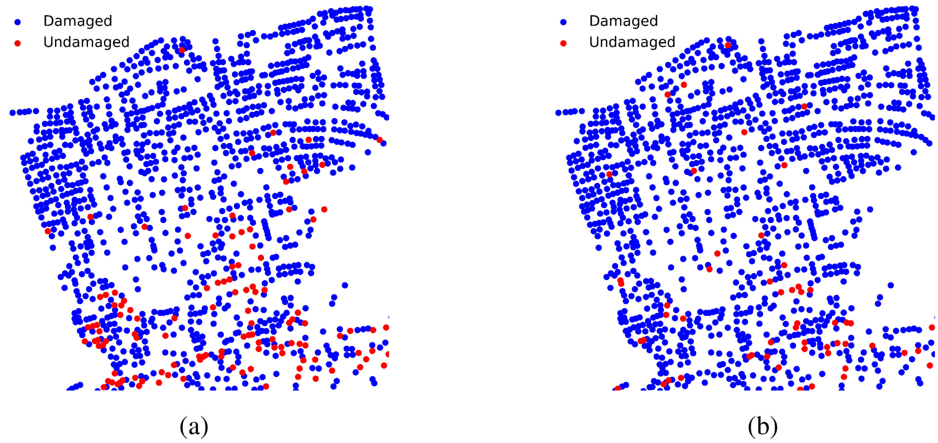


Fig. 7. Damage distributions in northward district.

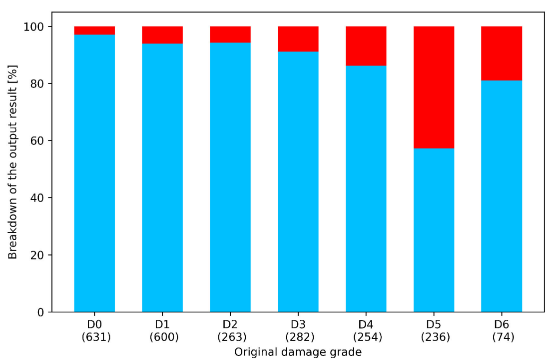


Fig. 8. Percentage correct for each damage grade: The values in () indicate the number of data per grade.

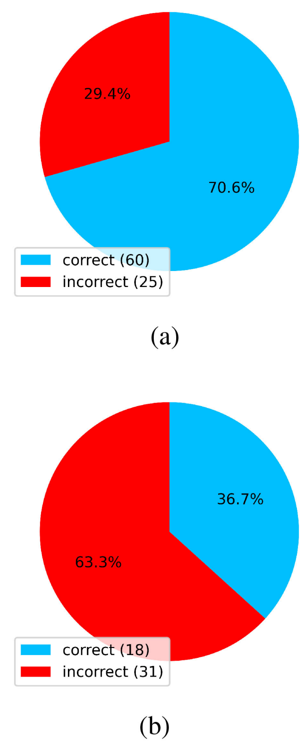


Fig. 9. Breakdown of correct/incorrect classifications by damage type category in D5: The values in () indicate the number of data.

recall—i.e., its likelihood of correctly recognizing structures labeled as "damaged"—is fairly low. To identify the reason, Fig. 8 depicts the percentage of correct judgments per original damage grade prior to the two-value labeling shown in Table I. The figure clearly shows that the percentage correct for damage grade 5 (D5) is far lower than that for the other grades. Structures classified as D5 are labeled as "damaged" under the study's two-value labeling; therefore, misidentification of structures belonging to this grade is believed to have led to the low recall value.

Compared with damage grade 6 (D6), which signifies completely collapsed structures, D5 contains a large number of structures that are difficult to classify using overhead photography because this grade contains damage patterns such as the collapse of only the first floor or the tilting of the structure. Therefore, based on structural information detailing the types and formats of damage from the results of a comprehensive survey that identified trends and points of commonality in misclassified data, D5 was reclassified into seven categories of damage types: "first-floor damage," "second-floor damage," "total collapse," "partial collapse," "partial damage," "deformity/tilting," and "other." Based on the results of this reclassification, three of the categories—second-floor damage, total collapse, and partial collapse—were posited as categories where the type of damage was discernible via overhead photography; the other four categories were posited as categories where the type of damage was difficult to discern from overhead. The proposed method's percentage of correct classifications was then rechecked for each category group.

Fig. 9 details the percentages of correct and incorrect classifications regarding damage detection for each category group for the proposed method. For the category group considered discernible, the proposed method's classifications were 71% correct, whereas for the category group considered difficult to discern, its classifications were only approximately 37% correct. From this result, it is clear that the correctness of the model is greatly influenced by whether the damage patterns can be judged from the overhead photography.

These results suggest that it is necessary to note that damage that can be discerned via satellite imagery differs from the damage grades assessed via field surveys. The results also suggest that approaching damaged/nondamaged dataset labeling from the same perspective would be effective in enhancing the performance of machine learning models.

VI. Conclusion

This study outlined an earthquake damage detection system based on satellite photographic imagery using deep learning to understand regional residential earthquake damage in an objective manner. This approach informed the design of the proposed model. As demonstrated by the research examples presented in this study, the use of multiple photographic satellite images, including images obtained during normal times, as well as the incorporation of multiple pieces of information deeply connected to disaster damage, are expected to form an effective approach for detecting general disaster damage, including and outside of the cases shown. In Japan, where disasters are becoming more intense and frequent, the development of a method for understanding and predicting the state of disaster damage that fully leverages deep learning and other current technologies will constitute in the future a vital field of research and development.

Acknowledgment

The information about structural damage distribution and structural attributes in the affected area in 2016 Kumamoto Earthquake in Japan was provided by Ministry of Land, Infrastructure, Transport and Tourism (MLIT), National Institute for Land Infrastructure Management (NILIM) and Building Research Institute. The authors would like to thank Prof. Seiichi Uchida in Kyushu University for his suggestive advice. The authors would like to express their sincere condolences to the victims of the earthquake and wish for a quick recovery and reconstruction of the affected area.

References

[1] S. Stramondo, C. Bignami, M. Chini, N. Pierdicca, and A. Tertulliani, “Satellite radar and optical remote sensing for earthquake damage detection: Results from different case studies,” *Int. J. Remote Sens.*, vol. 27, no. 20, pp. 4433–4447, 2006.

[2] T. Balz and M. Liao, “Building-damage detection using post-seismic high-resolution SAR satellite data,” *Int. J. Remote Sens.*, vol. 31, no. 13, pp. 3369–3391, 2010.

[3] L. Dong and J. Shan, “A comprehensive review of earthquake-induced building damage detection with remote sensing techniques,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 84, pp. 85–89, 2013.

[4] L. Gong, C. Wang, F. Wu, J. Zhang, H. Zhang, and Q. Li, “Earthquake-induced building damage detection with post-event sub-meter VHR TerraSAR-X staring spotlight imagery,” *Remote Sens.*, vol. 8, 2016, Art. no. 887.

[5] X. Tong *et al.*, “Building-damage detection using pre- and post-seismic high-resolution satellite stereo imagery: A case study of theMay 2008 *Wenchuan earthquake*,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 68, pp. 13–27, 2012.

[6] M. Matsuoka and F. Yamazaki, “Use of satellite SAR intensity imagery for detecting building areas damaged due to earthquake,” *Earthq. Spectra*, vol. 20, no. 3, pp. 975–994, 2004.

[7] B. Mansouri and Y. Hamednia, “A soft computing method for damage mapping using VHR optical satellite imagery,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8 no. 10, pp. 4935–4941, Oct. 2015.

[8] Y. Bai, B. Adriano, E. Mas, and S. Koshimura, “Machine learning based building damage mapping from the ALOS-2/PALSAR-2 SAR imagery: Case study of 2016 Kumamoto earthquake,” *J. Disaster Res.*, vol. 12, pp. 646–655, 2017.

[9] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, 2015.

[10] Y. LeCun *et al.*, “Backpropagation applied to handwritten zip code recognition,” *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1984.

[11] A. Krizhevsky, I. Sutskever, and G. H. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1106–1114.

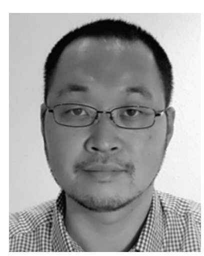
[12] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, “Learning spatiotemporal features with 3D convolutional networks,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 4489–4497.

[13] S. Ji, W. Xu, M. Yang, and K. Yu, “3D convolutional neural networks for human action recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[14] J. Ngiam, A. Khosla, M. Kim, J., Nam, H. Lee, and A. Y. Ng, “Multimodal deep learning,” in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 689–696.

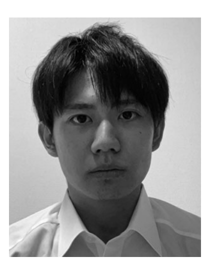
[15] Architectural Institute of Japan, “Report on the damage investigation of the 2016 Kumamoto earthquakes,” (*in Japanese*), 2018.

[16] J. Huang and X. C. X. Ling, “Using AUC and accuracy in evaluating learning algorithms,” *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 3, pp. 299–310. Mar. 2005.



Takashi Miyamoto (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in civil engineering from The University of Tokyo, Tokyo, Japan, in 2007, 2009, and 2012, respectively.

He is currently an Associate Professor with the University of Yamanashi, Yamanashi, Japan. He currently holds the post of Visiting Scholar with German Research Center for Artificial Intelligence, Kaiserslautern, Germany. His research interests include data-driven science, disaster mitigation engineering, and applied mechanics.



Yudai Yamamoto received the B.S. degree in civil and environmental engineering in 2020 from the University of Yamanashi, Yamanashi, Japan, where he is currently working toward the master’s degree in earthquake damage detection using satellite imagery with the Department of Civil and Environmental Engineering, Faculty of Engineering.

He studies machine learning application to earth observation data to detect natural disaster damage.