

# Building-damage detection method based on machine learning utilizing aerial photographs of the Kumamoto earthquake

**Shohei Naito<sup>1</sup>, Hiromitsu Tomozawa<sup>2</sup>, Yuji Mori<sup>2</sup>, Takeshi Nagata<sup>2</sup>, Naokazu Monma<sup>3</sup>, Hiromitsu Nakamura<sup>4</sup>, Hiroyuki Fujiwara<sup>4</sup>, and Gaku Shoji<sup>5</sup>**

## Abstract

This article presents a method for detecting damaged buildings in the event of an earthquake using machine learning models and aerial photographs. We initially created training data for machine learning models using aerial photographs captured around the town of Mashiki immediately after the main shock of the 2016 Kumamoto earthquake. All buildings are classified into one of the four damage levels by visual interpretation. Subsequently, two damage discrimination models are developed: a bag-of-visual-words model and a model based on a convolutional neural network. Results are compared and validated in terms of accuracy, revealing that the latter model is preferable. Moreover, for the convolutional neural network model, the target areas are expanded and the recalls of damage classification at the four levels range approximately from 66% to 81%.

## Keywords

Damage estimation, earthquake damage, earthquake hazards mitigation, Japan earthquakes, remote sensing

Date received: 26 December 2019; accepted: 27 December 2019

<sup>1</sup>Multi-hazard Risk Assessment Research Division, National Research Institute for Earth Science and Disaster Resilience, Tsukuba, Japan

<sup>2</sup>Mizuho Information & Research Institute, Inc., Tokyo, Japan

<sup>3</sup>PASCO Corporation, Tokyo, Japan

<sup>4</sup>National Research Institute for Earth Science and Disaster Resilience, Ibaraki, Japan

<sup>5</sup>Department of Engineering Mechanics and Energy, University of Tsukuba, Ibaraki, Japan

## Corresponding author:

Shohei Naito, Multi-hazard Risk Assessment Research Division, National Research Institute for Earth Science and Disaster Resilience, 3-1 Tennodai, Tsukuba 305-0006, Ibaraki, Japan.

Email: [naito@bosai.go.jp](mailto:naito@bosai.go.jp)

## Introduction

It is important to detect serious damages immediately after the occurrence of an earthquake, in terms of both supporting the various organizations that are involved in the disaster responses and in reducing the number of casualties. To this end, we previously developed a real-time damage estimation system to estimate the building damage and human casualties within 15 min of an earthquake (Fujiwara et al., 2019). Our developed system uses numerous seismic observations and estimates damage with multiple damage functions. The derived damage estimate is based on mesh unitary data, which include site amplification factors, building attributes, and demographic movements (Nakamura et al., 2013). The real-time damage estimation system is effective as a primary information source during the period of an initial disaster response. However, damage estimations may not correspond to actual damage owing to uncertainty associated with the accuracy of data and damage estimation models. It is therefore necessary to aggregate damage information confirmed through observations for applicable decision-making in the disaster response.

The time needed to obtain the entire extent of damage from field investigations increases with the scale of disaster. It is therefore appropriate to make complementary use of remote sensing images. As examples, Matsuoka and Yamazaki (2004) used synthetic-aperture radar (SAR) satellite imagery to detect damage to buildings on the basis of differences in the backscattering coefficient; Miura et al. (2013) conducted pixel-based texture analyses using optical satellite images for the extraction of areas of damage; Wu et al. (2014) used satellite images to estimate building heights on the basis of shadow lengths in the assessment of seismic vulnerability; and Menderes et al. (2015) used the pre- and post-event digital surface models generated from aerial photographs in the detection of collapsed buildings.

An effective remote sensing method is used to make observations using a fixed-wing aircraft. Aerial photography can be conducted from an altitude of approximately 1.2–3.5 km during the day when there is little cloud and therefore captures images that are of higher resolution than the images acquired by satellite (Yamazaki and Liu, 2016). Ogawa and Yamazaki (2000), for example, classified damage levels of buildings on the basis of the visual judgment of aerial photographs, showing that the approach was effective in identifying severely damaged or collapsed wooden buildings. Furthermore, with the aim of detecting damage in the event of a widespread disaster, image capturing by fixed-wing aircraft is superior to that by helicopters or that by unmanned aerial vehicles (UAVs), because the use of a fixed-wing aircraft allows investigation over an extensive area during a single flight. Accordingly, aerial photography has superior resolution, wider coverage, and better immediacy and is therefore effective in assessing building damage resulting from a huge earthquake.

In case of a large-scale earthquake, however, many man-hours are required for the visual interpretation of large aerial photographs. It is more efficient to extract building damage automatically by adopting a machine learning method. As examples, Gerke and Kerle (2011) applied a conditional random field and AdaBoost algorithm to damage classification using oblique aerial photographs of the post-earthquake scene; Cooner et al. (2016) applied an artificial neural network and random forest to post-earthquake satellite images in the detection of damaged buildings; Vetrivel et al. (2016) applied a damage classification model based on a bag-of-visual-words, which comprises the histogram-of-oriented-gradients descriptor and classifiers, such as a support vector machine (SVM),

random forest, and Adaboost, to post-earthquake oblique aerial images; and Tu et al. (2016) applied automatic damage detection models of rooftop areas based on a bag-of-visual-words, comprising the histogram-of-oriented-gradients descriptor and SVM classifier, to post-earthquake aerial photographs. The results of the aforementioned studies clarify that machine learning methods are effective in the automatic detection of damaged and undamaged buildings in remote sensing images, with accuracy ranging approximately from 60% to 90%.

Especially in recent years, it has been clarified that the use of a model based on a deep neural network (Hinton et al., 2006) provides better performance than the use of other supervised machine learning methods. Likewise, the deep learning model is highly applicable to the classification of remote sensing images. As examples, Castelluccio et al. (2015) applied the convolutional neural network (CNN) to the semantic classification of remote sensing scenes for two types of public dataset; Hu et al. (2015) applied plural models of CNNs for the classification of remote sensing scenes using public datasets; and Maggiori et al. (2017) proposed a framework for the dense classification of satellite imagery and generation of classification maps using a CNN and a fully convolutional network. The results of the aforementioned studies clarify that deep learning methods are effective in the detailed classification of remote sensing images, having accuracy of approximately more than 90%.

Deep learning is similarly applicable to the detection of damage to buildings in the event of a natural disaster. As examples, Vetrivel et al. (2017) developed a damage detection model with which to classify buildings as damaged or undamaged, by integrating the CNN and three-dimensional point cloud features and using post-earthquake images obtained by airplanes and UAVs; Fujita et al. (2017) proposed a CNN-based damage detection model with which to classify buildings as washed-away or surviving, using pre-tsunami and post-tsunami aerial images; Attari et al. (2017) proposed a CNN-based fine-grained damage classification method with which to classify building damage as mild, medium, or severe, using UAV images acquired in the aftermath of a cyclone; Duarte et al. (2018) developed a model with which to classify damaged and undamaged regions from satellite images, using a CNN trained with satellite and airborne image samples; Bai et al. (2018) proposed a CNN-based damage mapping model with which to detect washed-away or collapsed buildings, using pre- and post-tsunami satellite images; Li et al. (2019) developed a damaged building assessment method adopting post-hurricane aerial images to classify images as debris and mild, using the single-shot multibox detector model; Cao and Choe (2019) proposed a CNN-based model with which to classify images as flooded or undamaged using post-hurricane satellite imagery; and Ishii et al. (2018) developed a CNN model with which to classify buildings as damaged or other, using aerial and local photographs. The results of the aforementioned studies clarify that building-damage detection models based on deep learning are effective in the classification of individual buildings as damaged or undamaged, with accuracy approximately ranging from 70% to above 90%.

As previously noted, many researchers have proposed methods of automatically detecting building damage using machine learning techniques and remote sensing images. The adoption of these methods allows the extraction of damaged areas and classification of the damage to buildings into two or three classes, such as undamaged, moderately damaged, and severely damaged.

This study proposes a method of detecting building damage based on machine learning, using aerial photographs acquired by a fixed-wing airplane in the aftermath of the main

shock of the 2016 Kumamoto earthquake. The machine learning method has two models, namely a bag-of-visual-words model and a CNN model. Adopting these models, each building is classified into one of the four levels of damage that are equivalent to levels in a detailed validation result obtained from the visual interpretation of aerial photographs.

No previously developed machine learning model has classified the damage levels of individual buildings into four different classes. For the purpose of supporting municipalities in disaster assessment, it is important to determine the distributions and amounts of minor and moderate damages in cases where there are not many severely damaged buildings. Therefore, for the first time, we developed detailed damage classification models in this article. Detailed damage classification results obtained using the proposed method will be provided to municipalities, and we consider that such detailed information will be available for prioritizing manpower and other resources during disaster responses.

Furthermore, this article aims to update global damage information using the partial detection result obtained from aerial photography that can be used in a disaster response. Matsuoka and Nojima (2010), for example, developed an integration method using seismic intensity information and pre- and post-event satellite images. In this article, we developed a building-based damage detection method using building polygons released by the Geospatial Information Authority of Japan (GSI). Adopting this method, we can estimate the number of buildings for each damage level and then aggregate the damaged buildings in certain regions. By this procedure, the building-damage estimation result generated by the real-time damage estimation system (Nakamura et al., 2017) can be updated using the Bayesian updating method developed by Kusaka et al. (2017). Subsequently, the updated building-damage estimation results will be made available to government and municipalities to be used in the emergency and recovery stages of disaster responses (Figure 1).

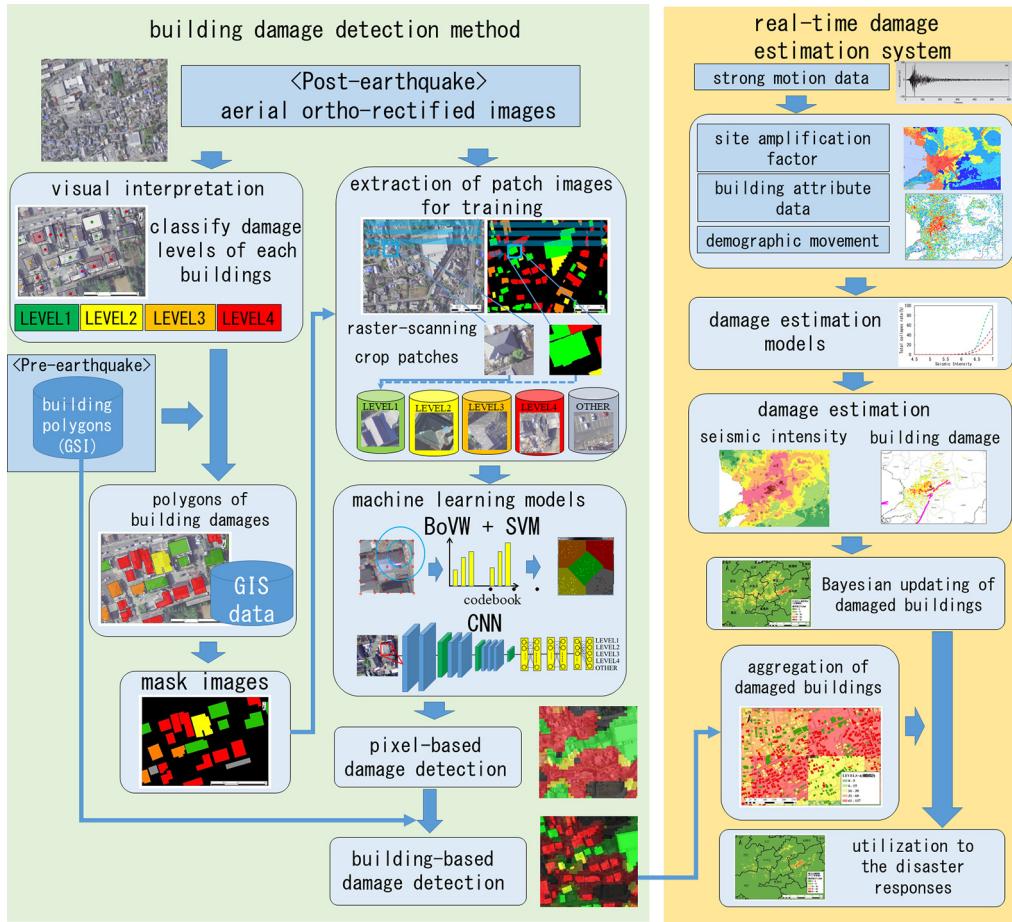
## **Method of creating training data for machine learning**

The main shock of the 2016 Kumamoto earthquake struck with a magnitude of 7.3 at 1:25 a.m. on 16 April. Considerable damage to buildings resulted in the deaths of 269 people and a total collapse of 8643 residential buildings (Kumamoto Prefecture, 2018).

This study classifies the damage to each building into one of the four categories through visual interpretation, using aerial orthorectified images with a definition of 20 cm/pixel. These images were taken for a devastated area in the town of Mashiki on 19 April 2016 by PASCO Corporation using a multi-line digital aerial sensor system. We defined the criteria for the damage classification by referencing the damage pattern chart of wooden buildings defined by Okada and Takai (2000), such that building damage is interpreted using only aerial photographs (Table 1). Figure 2 shows an example of building images for different damage classifications.

The visual interpretation results are subsequently constructed as data of a geographical information system, using basic geographic data of Japan released by the GSI (2018). Moreover, building polygons are coded with four colors for damage classification, as listed in Table 1, and mask images in which parts other than buildings are blacked out are generated using ArcGIS software, as illustrated in Figure 3.

The aerial photographs and mask images of all areas are scanned for each patch of  $80 \times 80$  pixels that included a major section of the roof of a residence with a scan width of 20 pixels (Figure 4). If buildings account for 30% of a patch, the patch is cropped as

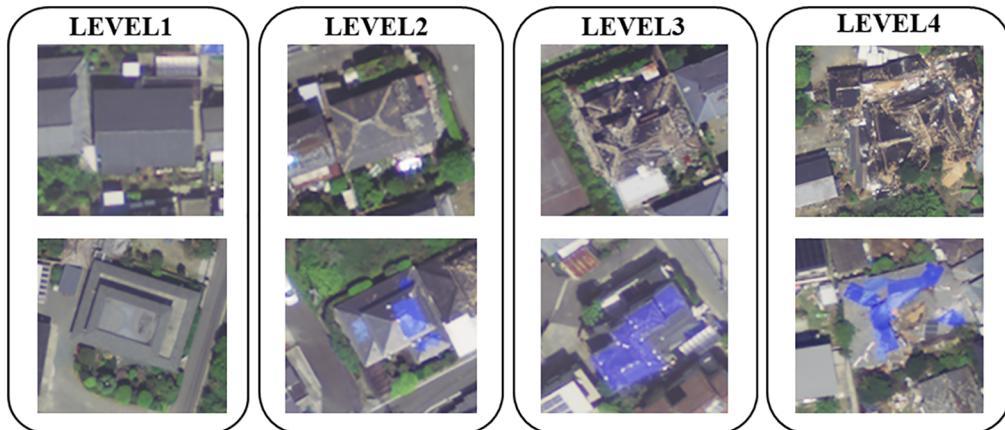


**Figure 1.** Overall illustration of the building-damage detection method and the real-time damage estimation system.

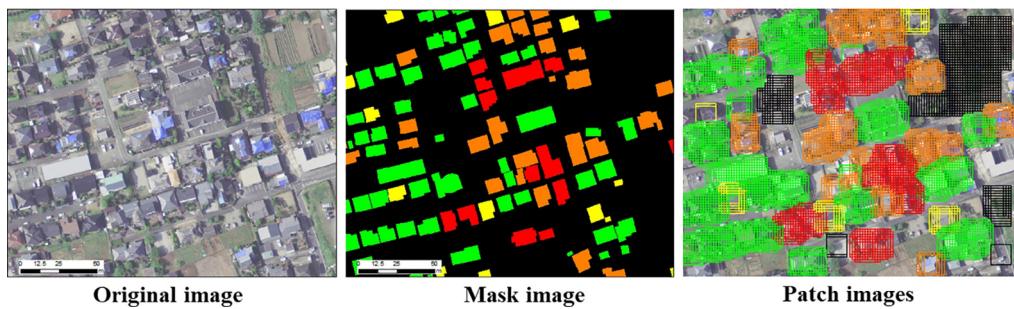
**Table I.** Building damage classifications.

Damage classifications in this study	Features in photographs	Damage grade (Okada and Takai, 2000)	Color (R, G, B)
LEVEL 1 (no damage)	No damage can be confirmed	D0	Green (0, 255, 0)
LEVEL 2 (minor damage)	Some roof tiles have collapsed	D1	Yellow (255, 255, 0)
LEVEL 3 (moderate damage)	Most of the roof tiles have collapsed, or part of a wall has fallen	D2, D3	Orange (255, 127, 0)
LEVEL 4 (major damage)	Distortion of the entire building, or destruction or collapse	D4, D5	Red (255, 0, 0)

training data for buildings in each damage classification; furthermore, if no building is included in a patch and the surrounding 40-pixel area, the patch is taken as training data for objects other than buildings. The patch images of  $80 \times 80$  pixels are thus used for the



**Figure 2.** Examples of building images for different damage classifications.



**Figure 3.** Examples of original, masked, and patched building images for different damage classifications.

machine learning of five classifications and automatically extracted using OpenCV, which is an open-source image analysis library (OpenCV team, 2018).

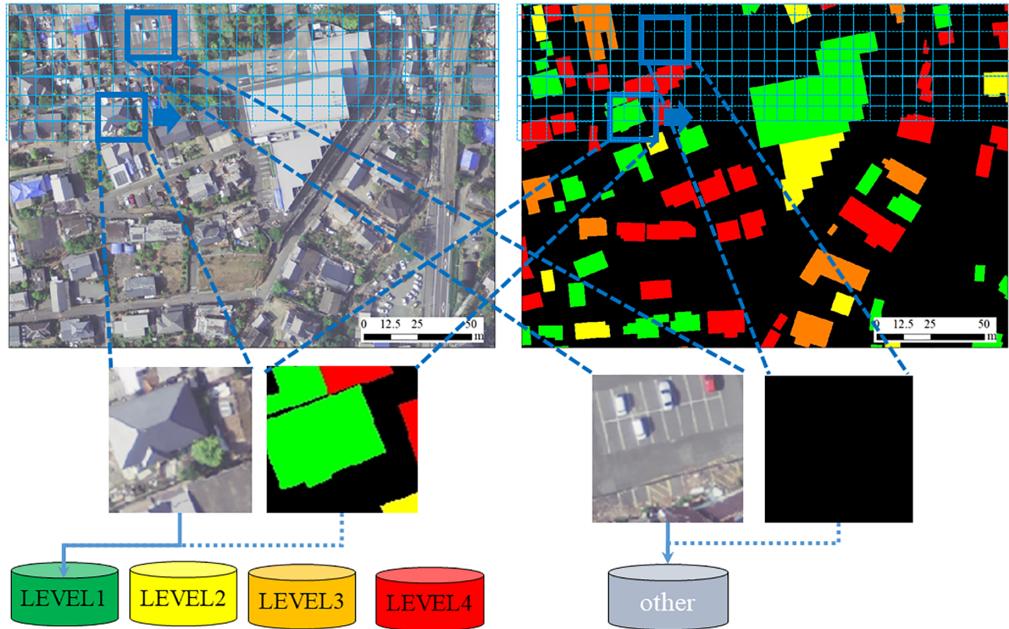
Following the above procedure, 132,574 patch images are automatically extracted using an image frame ( $2 \text{ km} \times 1.5 \text{ km}$ ) of the aerial photographs captured in the Miyazono area of Mashiki, which is one of the most devastated areas by the Kumamoto earthquake (Figure 5). Thereafter, patch images are screened such that the numbers of data for different classifications are equal, and 2500 patch images per classification are used for machine learning, as listed in Table 2.

## Development of machine learning models based on a bag-of-visual-words and SVM

In this article, we developed the building-damage detection method using two machine learning models known for their high discrimination performance, and compare the classification accuracies of the two models.

The first model is based on feature extraction and discrimination using a classifier. The scale-invariant feature transform (SIFT) developed by Lowe (2004) is applied as the

- (1) Raster-scan overall the aerial photos and mask images for each patch of  $80 \times 80$  pixels with a scan width of 20 pixels



(2) If buildings accounted for 30% of a patch, the patch was cropped as training data for buildings in each damage classification.

(3) If no building was included, the patch was determined as training data for objects other than buildings.

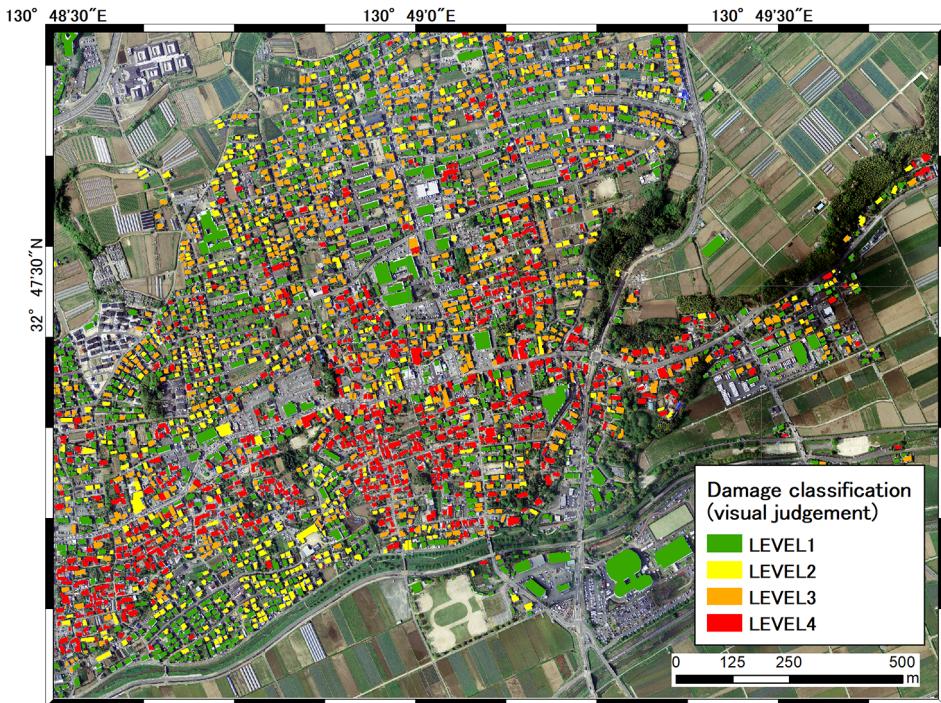
**Figure 4.** Procedure of extracting patch images from the Miyazono area.

**Table 2.** Training data extracted from an image frame of the Miyazono area.

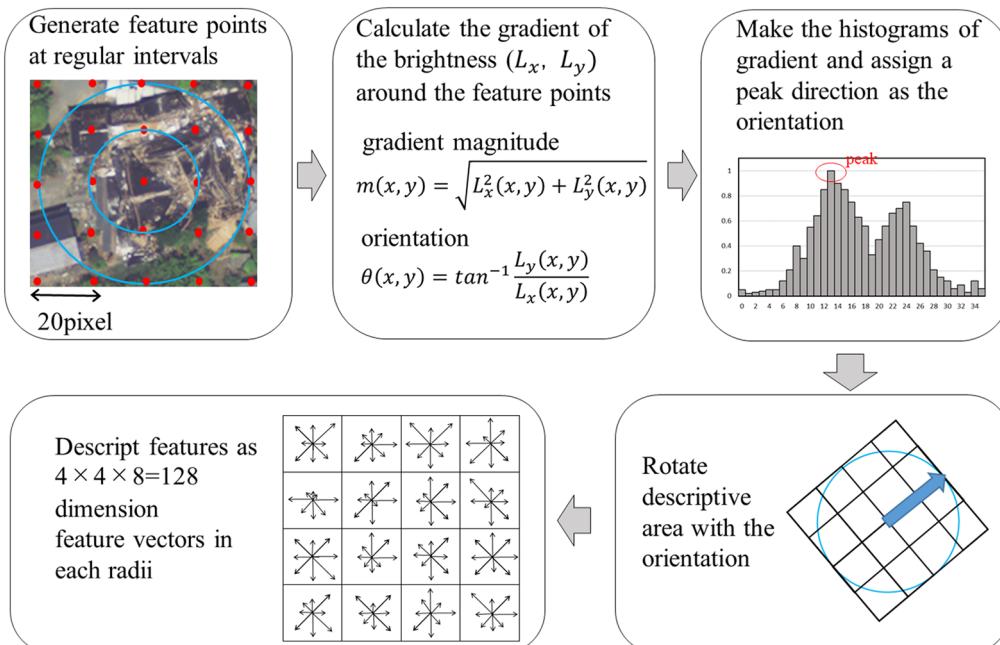
	Except buildings	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4	Total
Number of buildings	–	1257	523	745	691	3216
Number of patch images	100,087	13,008	4537	7237	7705	132,574
Number of training data	2500	2500	2500	2500	2500	12,500

feature descriptor. The SIFT is a feature extraction algorithm that is highly robust against changes in scale, rotation, and illumination. Furthermore, we use SIFT descriptors of two different radii (20 and 40 pixels) for each of the 25 feature points generated by dividing a patch image at regular intervals as shown in Figure 6.

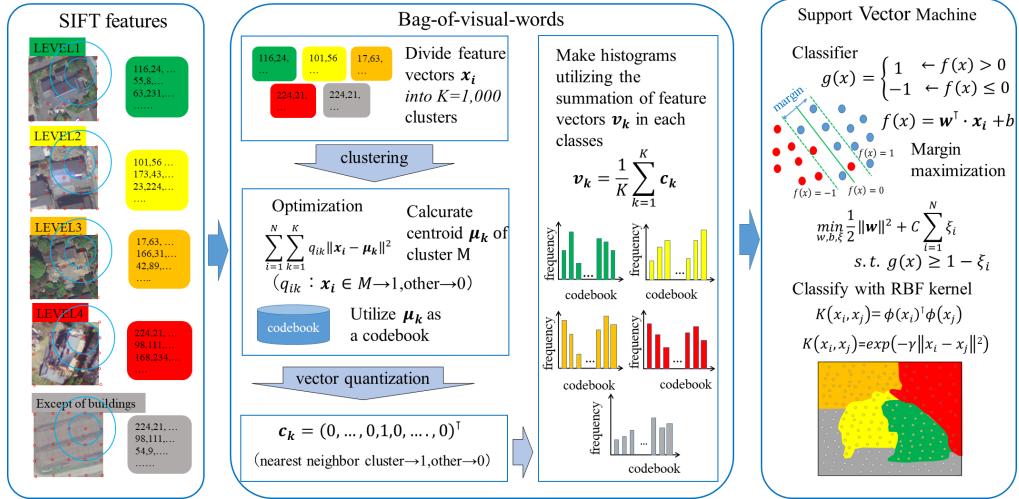
The entire SIFT feature vectors are divided into 1000 centroids by adopting k-means clustering, and these centroids are used as the codebook. Subsequently, histograms of each class are created using the frequencies of vector-quantized features based on the closest neighbor clusters (Figure 7). This method, using what is known as a bag-of-visual-words model, was developed by Csurka et al. (2004). It applies a bag-of-words model (Manning and Schütze, 1999), which classifies text according to the frequency of words while ignoring the word order, to image recognition. Similarly, the method is robustness against geometric



**Figure 5.** Image frame of the Miyazono area and the visually classified damage to buildings, with a background photograph taken by PASCO Corporation.



**Figure 6.** Overview of the feature extraction method using the SIFT descriptor.



**Figure 7.** Overview of the machine learning model based on a bag-of-visual-words and SVM.

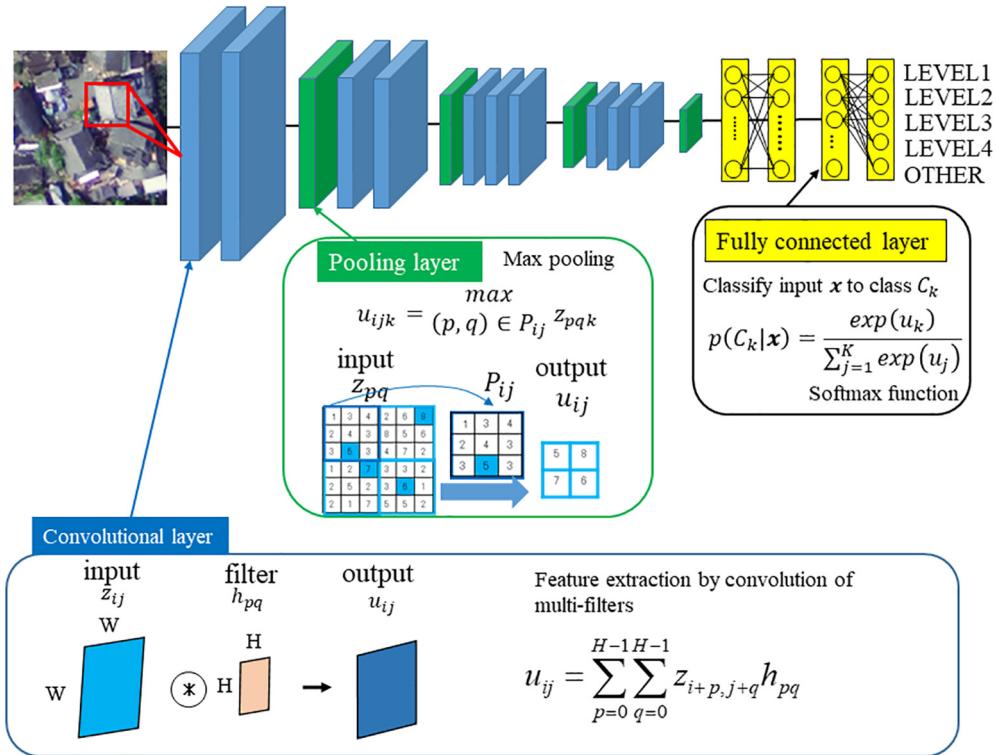
transformation in object recognition in that it ignores the relative positional relationship of local features.

Furthermore, we adopt the SVM developed by Vapnik (1998) using the radial basis function kernel as a classifier to set decision boundaries for each damage classification. In this study, the models based on a bag-of-visual-words and the SVM are constructed using OpenCV.

## Development of a machine learning model based on the CNN

Another model is the CNN, which is a deep learning model used widely in the field of image recognition (LeCun et al., 1989). The CNN comprises several convolutional layers, pooling layers, and fully connected layers (Figure 8). In a convolutional layer, a feature of an image is extracted by convolution using a plurality of filters. In a pooling layer, an image is compressed by extracting the maximum value between adjacent pixels. In a fully connected layer, the discrimination result is output as a probability of each class label. Furthermore, the discrimination accuracy is improved over a number of iterations, using error back propagation to minimize the difference between the discrimination results and correct value.

In this study, we developed a CNN model based on the visual geometry group (VGG) model (Simonyan and Zisserman, 2014) using MXNet (Apache Software foundation, 2018), a deep learning library (Table 3). The parameters of the CNN are given in Table 4. In addition, we add batch normalization layers (Ioffe and Szegedy, 2015) behind each convolutional layer; these layers prevent an increase in dispersion between mini-batches. Furthermore, we add a dropout layer between fully connected layers to prevent overfitting and we shuffle training data and rearrange mini-batches in each epoch to eliminate biased data associated with learning orders.



**Figure 8.** Overview of the CNN model.

## Accuracy comparison of machine learning models

We compared the classification performances of the machine learning models using a 10-fold cross-validation method. For both models, 12,500 patch images are used as data in cross validation (Table 2); these images are extracted from the Miyazono area (Figure 5). Initially, 2500 patch images of each class are divided into 10 randomly selected groups. Among these 10 groups, data of nine groups are used for training and data of the one remaining group are used for validation. The combination of groups is then switched for a total of 10 repetitions such that each group is used in validation. All patch images are thus used for training and validation.

The results of 10-fold cross validation are evaluated by comparing with the visual interpretation results of aerial photography, using a confusion matrix (Table 5). Adopting the confusion matrix, the discrimination accuracy is evaluated in terms of four metrics

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{F-measure} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} \quad (3)$$

**Table 3.** Structure of the CNN model.

No.	Layer	Data size	Number of filters	Size of filters	Stride width	Activation function
0	Input (RGB image)	$80 \times 80 \times 3$	—	—	—	—
1	Convolution layer	$80 \times 80$	64	$3 \times 3$	1	ReLU
	Batch normalization layer	$80 \times 80$	—	—	—	—
	Convolution layer	$80 \times 80$	64	$3 \times 3$	1	ReLU
	Batch normalization layer	$80 \times 80$	—	—	—	—
2	Pooling layer	$40 \times 40$	64	$2 \times 2$	2	—
	Convolution layer	$40 \times 40$	128	$3 \times 3$	1	ReLU
	Batch normalization layer	$40 \times 40$	—	—	—	—
	Convolution layer	$40 \times 40$	128	$3 \times 3$	1	ReLU
	Batch normalization layer	$40 \times 40$	—	—	—	—
3	Pooling layer	$20 \times 20$	128	$2 \times 2$	2	—
	Convolution layer	$20 \times 20$	256	$3 \times 3$	1	ReLU
	Batch normalization layer	$20 \times 20$	—	—	—	—
	Convolution layer	$20 \times 20$	256	$3 \times 3$	1	ReLU
	Batch normalization layer	$20 \times 20$	—	—	—	—
	Convolution layer	$20 \times 20$	256	$3 \times 3$	1	ReLU
	Batch normalization layer	$20 \times 20$	—	—	—	—
4	Pooling layer	$10 \times 10$	256	$2 \times 2$	2	—
	Convolution layer	$10 \times 10$	512	$3 \times 3$	1	ReLU
	Batch normalization layer	$10 \times 10$	—	—	—	—
	Convolution layer	$10 \times 10$	512	$3 \times 3$	1	ReLU
	Batch normalization layer	$10 \times 10$	—	—	—	—
	Convolution layer	$10 \times 10$	512	$3 \times 3$	1	ReLU
	Batch normalization layer	$10 \times 10$	—	—	—	—
5	Pooling layer	$5 \times 5$	512	$2 \times 2$	2	—
	Convolution layer	$5 \times 5$	512	$3 \times 3$	1	ReLU
	Batch normalization layer	$5 \times 5$	—	—	—	—
	Convolution layer	$5 \times 5$	512	$3 \times 3$	1	ReLU
	Batch normalization layer	$5 \times 5$	—	—	—	—
	Convolution layer	$5 \times 5$	512	$3 \times 3$	1	ReLU
	Batch normalization layer	$5 \times 5$	—	—	—	—
6	Pooling layer	$2 \times 2$	512	$2 \times 2$	2	—
7	Fully connected layer	$1 \times 1 \times 4096$	—	—	—	ReLU
	Dropout layer	$1 \times 1 \times 2048$	—	—	—	—
8	Fully connected layer	$1 \times 1 \times 5$	—	—	—	Softmax
	Output (discrimination result)	—	—	—	—	—

CNN: convolutional neural network.

**Table 4.** Parameters of the CNN model.

Item	Parameter
Optimization method	Adam
Learning rate	$10^{-3}$
Weight decay	$10^{-9}$
Batch size	100
Max epoch	500

CNN: convolutional neural network.

**Table 5.** Confusion matrix.

		Predicted labels (machine learning)	
		Positive	Negative
Actual labels (visual interpretation)	Positive	TP FP	FN TN
	Negative		

TP: true positive; FN: false negative; FP: false positive; TN: true negative.

$$\text{Overall accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (4)$$

For the model based on a bag-of-visual-words, the recalls of different damage levels range approximately from 39% to 60%, the precisions range approximately from 45% to 67%, and the overall accuracy is approximately 57% (Table 6). Meanwhile, for the model based on the CNN, the recalls of different damage levels range approximately from 81% to 92%, the precisions range approximately from 78% to 92%, and the overall accuracy is approximately 88% (Table 7). Consequently, the model based on the CNN outperforms the bag-of-visual-words model. Furthermore, in terms of the processing time, the CNN discrimination model requires less computational time for the same training data and execution environment (Table 8). Comparing with visual interpretation, the CNN accelerates the processing for discrimination (Table 9). The model based on the CNN is thus adopted to build the damage detection method as described below.

## Damage classification model for individual buildings based on building polygons

Using the damage classification model with a unit of patches, we developed the damage classification model for individual buildings according to the following procedures (Figure 9).

We raster-scan the whole area in aerial photography with a scan width of 20 pixels, crop patches of  $80 \times 80$  pixels for each scan width, and classify the damage of each patch using the CNN model. In this way, areas with building damage are visualized using a mode value of 16 discrimination results in each 20-pixel square grid. We subsequently overlay building polygons (GSI, 2018) and discriminate damage according to the area ratio of the damage level within each polygon. The damage is classified as LEVEL 4 when the area ratio of the damage level is more than 10% in a building polygon, LEVEL 3 when the area ratio is more than 10%, LEVEL 2 when the area ratio is more than 30%, and LEVEL 1 in all other cases. The given thresholds are determined by comparing the overall accuracies of test results and adopting the best combination of thresholds for each damage level from 5%, 10%, and 30%.

Adopting the described method, we classified damage to buildings in the Miyazono area using the CNN model and compared the results with the results of the visual interpretation of aerial photographs (Figure 10). The recalls of different damage levels range approximately from 61% to 93%, the precisions range approximately from 50% to 89%, and the overall accuracy is approximately 69% (Table 10). The results suggest that the discrimination accuracy is sufficient at each damage level in estimating the amount of building damage. In addition, the proposed model extracts most LEVEL 4 buildings determined by visual judgment. The result may be attributed to the thresholds of area ratios of each

**Table 6.** Comparison of results of the visual interpretation and the bag-of-visual-words model.

	Visual interpretation	Discrimination result with bag-of-visual-words				Total	Recall (%)	F-measure (%)
		Except buildings	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4		
Visual interpretation	Except buildings	1860	163	92	58	327	2500	74.4
	LEVEL 1	248	1289	496	259	208	2500	51.6
	LEVEL 2	139	262	975	661	463	2500	39.0
	LEVEL 3	51	93	303	1509	544	2500	60.4
	LEVEL 4	140	112	284	518	1446	2500	57.8
Total		2438	1919	2150	3005	2988	12,500	52.7
Precision (%)		76.3	67.2	45.3	50.2	48.4	—	Overall accuracy (%) 56.6

**Table 7.** Comparison of results of the visual interpretation and the CNN model.

	Visual interpretation	Discrimination result with CNN				Total	Recall (%)	F-measure (%)
		Except buildings	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4		
Visual interpretation	Except buildings	2409	44	15	—	31	2500	96.4
	LEVEL 1	32	2023	383	—	61	2500	80.9
	LEVEL 2	9	195	2137	62	97	2500	85.5
	LEVEL 3	2	—	95	2310	92	2500	92.4
	LEVEL 4	28	53	117	132	2170	2500	86.8
Total		2480	2316	2747	2506	2451	12,500	87.7
Precision (%)		97.1	87.3	77.8	92.2	88.5	—	Overall accuracy (%) 88.4

CNN: convolutional neural network

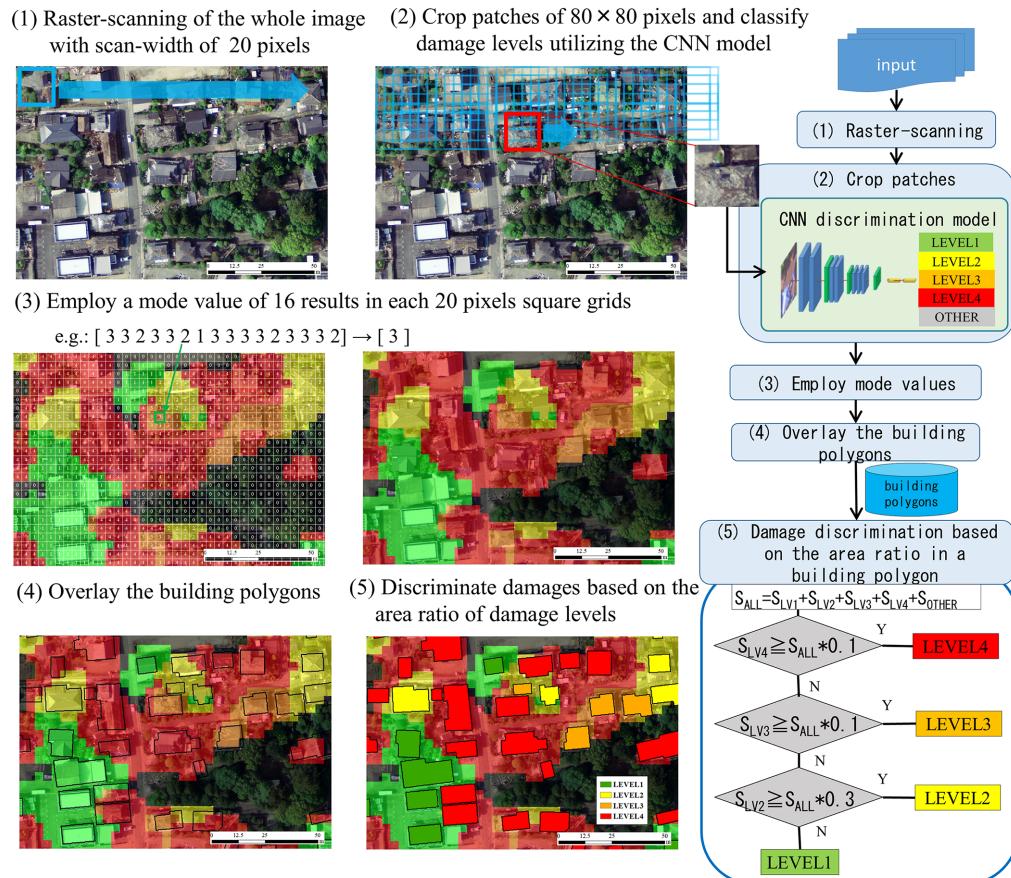
**Table 8.** Execution environment for the development of machine learning.

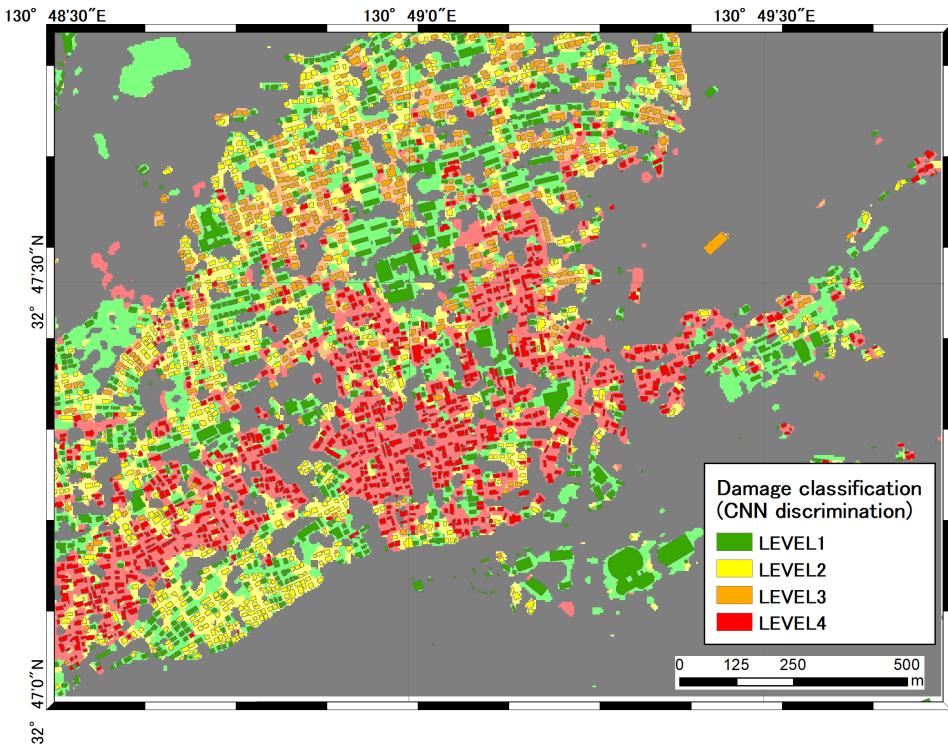
Item	Parameter
OS	Windows 10 (64 bit)
CPU	Core-i7
RAM	32 GB
GPU	GeForce GTX1070 8 GB
Platform	CUDA 8.0

**Table 9.** Comparison of the processing times of the damage detection models.

Model	Processing time
Visual interpretation	3 days (1000 buildings per day)
Bag-of-visual-words	13 h (training and discrimination)
CNN	3 h for training; 5–6 min for discrimination

CNN: convolutional neural network.

**Figure 9.** Damage classification model based on 20-pixel square grids and building polygons.



**Figure 10.** Damage classification result of buildings obtained using the CNN discrimination model and building polygons in the Miyazono area.

**Table 10.** Comparison of discrimination results of the visual interpretation and CNN model with a unit of a building.

	Discrimination result with CNN				Total	Recall (%)	F-measure (%)	
	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4				
Visual interpretation	LEVEL 1	768	251	75	163	1257	61.1	72.4
	LEVEL 2	63	375	27	58	523	71.7	59.0
	LEVEL 3	26	98	437	184	745	58.7	67.2
	LEVEL 4	9	24	17	641	691	92.8	73.8
Total		866	748	556	1046	3216	—	—
Precision (%)		88.7	50.1	78.6	61.3	—	Overall accuracy (%)	69.1

CNN: convolutional neural network.

criterion for each building polygon. We consider the thresholds are suitable for avoiding underestimation in a disaster response.

### Discrimination model covering an extended area

We developed the CNN discrimination model covering an extended area around Mashiki using 20 image frames of aerial photography and the masked images. Figure 11 illustrates

**Table 11.** Comparison of the discrimination results of the visual interpretation and CNN model in an extended area.

		Discrimination result with CNN				Total	Recall (%)	F-measure (%)	
		Except buildings	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4			
Visual interpretation	LEVEL 1	223	17,661	3603	54	187	21,728	81.3	88.0
	LEVEL 2	5	742	2454	345	104	3650	67.2	48.7
	LEVEL 3	2	10	302	741	91	1146	64.7	62.8
	LEVEL 4	7	13	60	75	296	451	65.6	52.4
Total		237	18,426	6419	1215	678	26,975	—	—
Precision (%)	—	—	95.8	38.2	61.0	43.7	—	Overall accuracy (%)	78.4

CNN: convolutional neural network.

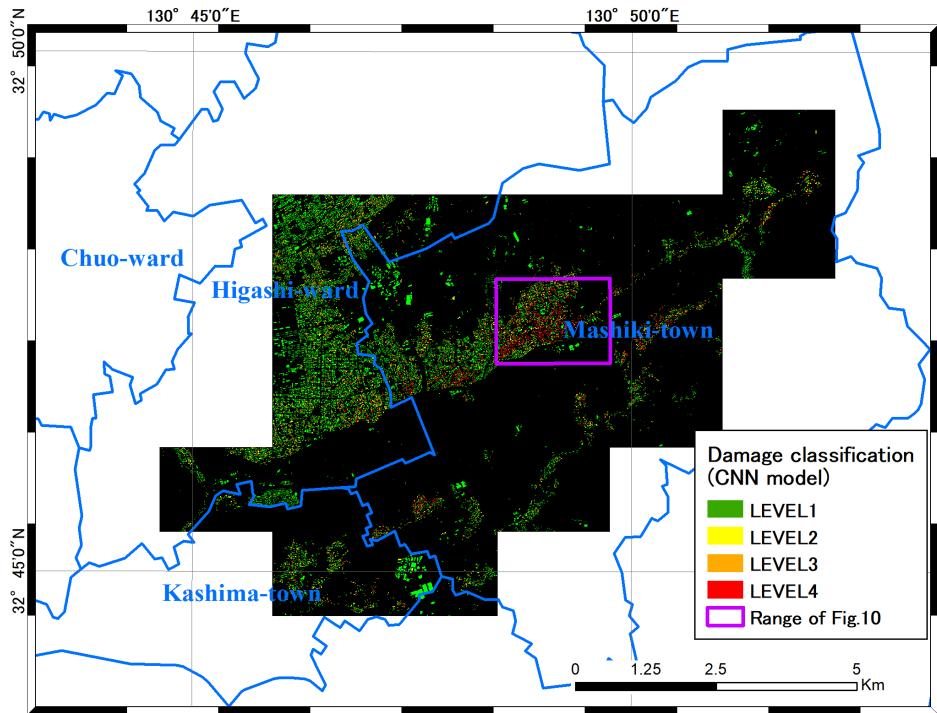
the entire area used for the training data, which includes 28,544 buildings. Patch images are initially cropped from aerial photographs using the centroids of bounding rectangles of each building polygon, and the damage level of each polygon is then labeled using the mask images. Subsequently, the data visually difficult to classify are screened, and 26,975 patch images are thus used for training data.

Furthermore, cross validation with a unit of image frames is adopted in validating the discrimination accuracy. The image frames are divided into an image frame for validation and 19 image frames for training. However, an image frame of the Miyazono area (Figure 10) is excluded from the data for validation because it has already been used in training data. The discrimination results of these data are given in Table 11 using a confusion matrix.

Table 11 shows that the total number of areas, having different damage levels, are disproportional in comparison with the visual interpretation. The precisions of LEVEL 2 and LEVEL 4 are relatively low because all data are affected by the misclassification of LEVEL 1, for which there are a large number of buildings. The overall accuracy is high owing to the strong effect of the discrimination result of LEVEL 1.

For the reasons stated above, we mainly use recall as the evaluation index of classification performance in this section. Recall is not affected by the number of data for each class, and the recalls of different damage levels thus range approximately from 66% to 81%. We thus confirm sufficient discrimination accuracy, in the context of assessing the number of damaged buildings at each damage level.

We furthermore compared the ratios of each damage level obtained in this study with governmental inspection results (Monma et al., 2018) for the same area (Figure 12 and Table 12). It can be seen that the ratio of LEVEL 1 in this study nearly corresponds to the ratio of no damage or minor damage in the governmental inspection, the ratio of LEVEL 2 nearly corresponds to the ratio of moderate damage, and the ratios of LEVEL 3 and LEVEL 4 nearly correspond to the ratio of major damage. It is difficult to simply compare with each criterion because the investigation based on aerial photography and the governmental inspection based on field surveys differ in terms of the examination method and purpose. However, using these correspondence relationships, the building-damage detection method would be helpful in a disaster response. As an example, the distribution of



**Figure 11.** Area of the training data for the CNN discrimination model.

**Table 12.** Number of buildings at each damage level according to the CNN model, the visual interpretation of aerial photographs, and a governmental inspection based on field surveys.

	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4	Total
Automatic classification using the CNN model	15,743	7104	2614	3083	28,544
Visual interpretation of aerial photograph	18,640	5862	2530	1472	28,544
		No damage or minor damage	Moderate damage	Major damage	Total
Governmental inspection based on field surveys	17,970	6485	7567		32,022

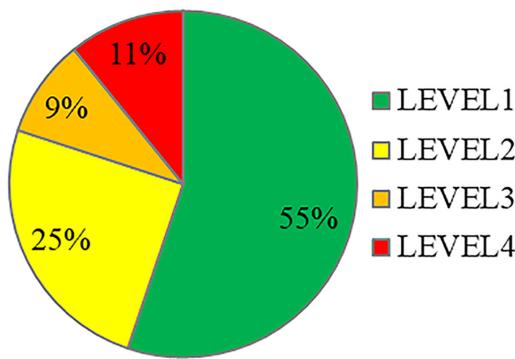
CNN: convolutional neural network.

major damage could be used by municipalities in estimating the amount of damage, assigning human resources, and arranging relief supplies.

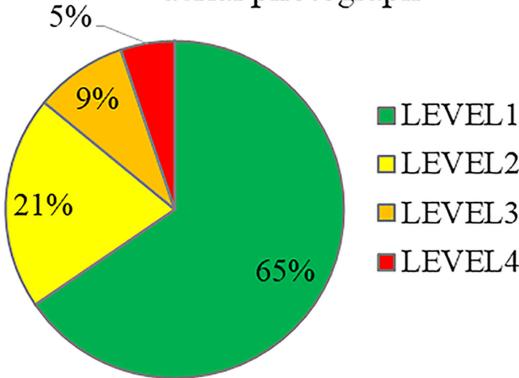
## Conclusion and future study

A method of detecting building damage using two types of machine learning model was proposed and applied to aerial photographs taken immediately after the main shock of the

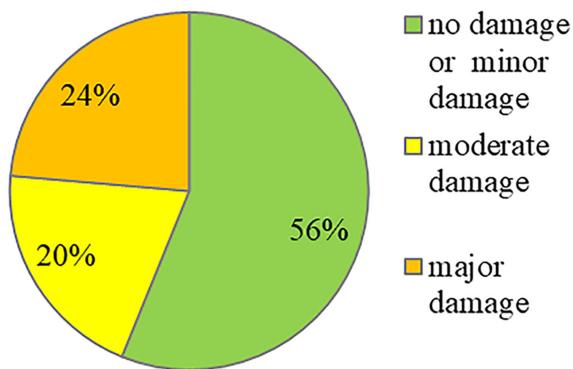
automatic classification  
using the CNN model



visual interpretation of an  
aerial photograph



governmental inspection  
based on field surveys



**Figure 12.** Comparison of the ratios of different damage levels obtained using the CNN discrimination model, the visual interpretation of aerial photographs, and a governmental inspection based on field surveys.

2016 Kumamoto earthquake. We subsequently compared the classification results of the models and confirmed that higher accuracy was achieved using the CNN model, for which the recalls of different damage levels ranged approximately from 81% to 92%.

We additionally developed a damage classification method for individual buildings using polygon data released by the GSI; the recalls of different damage levels ranged approximately from 61% to 93%.

Furthermore, applying the cross-validation method using 20 image frames of aerial photography resulted in recalls of different damage levels ranging approximately from 66% to 81%. For the results stated above, we confirmed that the proposed method performs sufficiently well in estimating the amount and distribution of residential damage, and the judgment is more immediate than the human judgment.

As described above, the presented method provided high classification performance when applied to a sequence of aerial photographs taken at around the same time. However, the discrimination accuracy is expected to diminish under other photographic conditions, such as adverse weather conditions or low light, or for other regional characteristics of buildings. We must therefore improve the generalization performance of the machine learning model.

To this end, we are developing a model that has extensive generalization performance by employing images of other earthquakes. As an example, the developed model has been trained using data captured from other aerial photographs acquired soon after the 1995 southern Hyogo prefecture earthquake and the 2011 earthquake off the Pacific coast of Tohoku.

The transfer learning method using a supervised pre-trained network with a rich dataset has been used effectively in constructing a CNN model with high generalization performance (Agrawal et al., 2014, Girshick et al., 2014). By adopting transfer learning, it seems possible to enhance the generalization performance of the CNN-based building-damage detection model. In addition, by adopting semi-supervised learning based on generative adversarial networks (Goodfellow et al., 2014), numerous training data can be generated from a few labeled data. By employing such state-of-the-art deep learning methods, it is possible to develop a model that has high generalization performance, in the case that a huge training dataset is not available. Furthermore, using the additional training data and various deep learning methods described above, our proposed method would be applicable to the detection of damage to buildings in the event of other types of disaster, such as tsunami, typhoons, and landslides.

Moreover, we are developing a method of aggregating damaged buildings with meshes using the automatic damage discrimination result. Using these data, it is possible to apply a Bayesian update to the result generated by the real-time damage estimation system on the basis of earthquake motion (Figure 1). The updated damage estimation result will be available to support the initial disaster responses of municipal administrations and insurance companies. We will therefore develop a method of detecting building damage that will have a higher generalization performance, contributing to the prompt restoration of infrastructure.

## Acknowledgments

The authors thank Glenn Pennycook, MSc, from Edanz Group for editing the draft of this manuscript.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the Cross-ministerial Strategic Innovation Promotion Program, entitled “Enhancement of societal resiliency against natural disasters,” under the initiative of the Cabinet Office, Government of Japan.

## References

- Agrawal P, Girshick R and Malik J (2014) Analyzing the performance of multilayer neural networks for object recognition. arXiv:1407.1610v2.
- Apache Software foundation ( 2018) Apatch MXNet V1.3. Available at <https://mxnet.apache.org/> (accessed 13 November 2018)
- Attari N, Ofli F, Awad M, et al. (2017) Nazr-CNN: Fine-grained classification of UAV imagery for damage assessment. arXiv:1611.06474v2.
- Bai Y, Mas E and Koshimura S (2018) Towards operational satellite-based damage-mapping using U-Net convolutional network: A case study of 2011 Tohoku Earthquake-Tsunami. *Remote Sensing* 2018(10): 1626.
- Cao QD and Choe Y (2019) Building damage annotation on post-hurricane satellite imagery based on convolutional neural networks. arXiv:1807.01688v3.
- Castelluccio M, Poggi G, Sansone C, et al. (2015) Land use classification in remote sensing images by convolutional neural networks. arXiv:1508.00092v1.
- Cooner AJ, Shao Y and Campbell JB (2016) Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 Haiti earthquake. *Remote Sensing* 8: 868.
- Csurka G, Dance CR, Fan L, et al. (2004) Visual categorization with bags of keypoints. In: ECCV International Workshop on Statistical Learning in Computer Vision, Prague, Czech Republic, 15–16 May.
- Duarte D, Nex F, Kerle N, et al. (2018) Multi-resolution feature fusion for image classification of building damages with convolutional neural networks. *Remote Sensing* 2018(10): 1636.
- Fujita A, Sakurada K, Imaizumi T, et al. (2017) Damage detection from aerial images via Convolutional Neural Networks. In: *15th IAPR international conference on machine vision applications*, Nagoya, Japan, 8–12 May.
- Fujiwara H, Nakamura H, Senna S, et al. (2019) Development of a real-time damage estimation system. *Journal of Disaster Research* 14(2): 315–332.
- Geospatial Information Authority of Japan (GSI) ( 2018) Basic geographic data of Japan. Available at: <http://www.gsi.go.jp/kiban/index.html> (accessed 9 November 2018).
- Gerke M and Kerle N (2011) Automatic structural seismic damage assessment with airborne oblique pictometry imagery. *Photogrammetric Engineering & Remote Sensing* 77: 885–891.
- Girshick R, Donahue J, Darrell T, et al. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: *2014 IEEE conference on computer vision and pattern recognition*, Columbus, OH, USA, 23–28 June.
- Goodfellow IJ, Abadie JP, Mirza M, et al. (2014) Generative adversarial networks. arXiv:1406.2661v1.
- Hinton GE, Osindero S and Teh W (2006) A fast learning algorithm for deep belief nets. *Neural Computation* 18: 1527–1554.
- Hu F, Xia GS, Hu J, et al. (2015) Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing* 2015(7): 14680–14707.
- Ioffe S and Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167v3.

- Ishii Y, Matsuoka M, Maki N, et al. (2018) Recognition of damaged building using deep learning based on aerial and local photos taken after the 1995 Kobe Earthquake. *Journal of Structural and Construction Engineering* 83(751): 1391–1400 (in Japanese).
- Kumamoto Prefecture (2018) Damage information of Kumamoto earthquake. Available at: <http://www.pref.kumamoto.jp/> (accessed 2 November 2018).
- Kusaka A, Nakamura H, Fujiwara H, et al. (2017) Bayesian updating of damaged building distribution in post-earthquake assessment. *Earthquake Journal of Japan Association for Earthquake Engineering* 17(1): 16–29 (in Japanese).
- LeCun Y, Boser B, Denker JS, et al. (1989) Backpropagation applied to handwritten zip code recognition. *Neural Computation* 1: 541–551.
- Li Y, Hu W, Dong H, et al. (2019) Building damage detection from post-event aerial imagery using single shot multibox detector. *Applied Sciences* 9: 1128.
- Lowe DG (2004) Distinctive image features from scale-invariant key-points. *International Journal of Computer Vision* 60(2): 91–110.
- Maggiori E, Tarabalka Y, Charpiat G, et al. (2017) Convolutional Neural Networks for large-scale remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing* 55: 645–657.
- Manning CD and Schütze H (1999) *Foundation of Statistical Natural Language Processing*. Cambridge, MA: The MIT Press, 720 pp.
- Matsuoka M and Nojima N (2010) Building damage estimation by integration of seismic intensity information and satellite L-band SAR imagery. *Remote Sensing* 2010(2): 2111–2126.
- Matsuoka M and Yamazaki F (2004) Use of satellite SAR intensity imagery for detecting building areas damaged due to earthquakes. *Earthquake Spectra* 20(3): 975–994.
- Menderes A, Erener A and Sarp G (2015) Automatic detection of damaged buildings after earthquake hazard by using remote sensing and information technologies. *Procedia Earth and Planetary Science* 15: 257–262.
- Miura H, Midorikawa S and Kerle N (2013) Detection of building damage areas of the 2006 central Java, Indonesia, earthquake through digital analysis of optical satellite images. *Earthquake Spectra* 29(2): 453–473.
- Monma N, Nakamura H, Fujiwara H, et al. (2018) Construction of earthquake building damage information space database and examination of building damage curve. In: *The 15th Japan earthquake engineering symposium, PS1-01* (in Japanese), Sendai, Japan, 6–8 December.
- Nakamura H, Aoi S, Kunugi T, et al. (2013) Prototype of a real-time system for earthquake damage estimation in Japan. *Journal of Disaster Research* 8: 981–989.
- Nakamura H, Fujiwara H, Kunugi T, et al. (2017) Development of real-time earthquake damage information system in Japan. In: *Proceedings of the 16th world conference on earthquake engineering (16WCEE)*, Santiago, 9–13 January.
- Ogawa N and Yamazaki F (2000) Photo-interpretation of building damage due to earthquakes using aerial photographs. In: *Proceedings of the 12th world conference on earthquake engineering (12WCEE)*, paper no. 1906, Auckland, New Zealand, 30 January–4 February. Upper Hutt, New Zealand: New Zealand Society for Earthquake Engineering.
- Okada S and Takai N (2000) Classification of structural types and damage patterns of buildings for earthquake field investigation. In: *Proceedings of the 12th World conference on earthquake engineering (12WCEE)*, paper no. 705, Auckland, New Zealand, 30 January–4 February. Upper Hutt, New Zealand: New Zealand Society for Earthquake Engineering.
- OpenCV team (2018) OpenCV 3.2.0. Available at: <https://opencv.org/> (accessed 13 November 2018).
- Simonyan K and Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.
- Tu J, Sui H, Feng W, et al. (2016) Detection of damaged rooftop areas from high-resolution aerial images based on visual bag-of-words model. *IEEE Geoscience and Remote Sensing Letters* 13(12): 1817–1821.
- Vapnik VN (1998) *Statistical Learning Theory*. New York: Wiley-Interscience, 786pp.
- Vetrivel A, Gerke M, Kerle N, et al. (2016) Identification of structurally damaged areas in airborne oblique images using a Visual-Bag-of-Words approach. *Remote Sensing* 8: 231.

- Vetrivel A, Gerke M, Kerle N, et al. (2017) Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel learning. *ISPRS Journal of Photogrammetry and Remote Sensing* 140: 45–59.
- Wu H, Cheng Z, Shi W, et al. (2014) An object-based image analysis for building seismic vulnerability assessment using high-resolution remote sensing imagery. *Natural Hazards* 71: 151–174.
- Yamazaki F and Liu W (2016) Remote sensing technologies for post-earthquake damage assessment: A case study on the 2016 Kumamoto earthquake. In: *Proceedings of the 6th ASIA conference on earthquake engineering (6ACEE)*, Cebu City, Philippines, 22–24 September.