Journal of Japan Association for Earthquake Engineering, Vol. 21, No. 3, 2021

DEVELOPMENT OF THE DEEP LEARNING BASED DAMAGE DETECTION MODEL FOR BUILDINGS UTILIZING AERIAL PHOTOGRAPHS OF MULTIPLE EARTHQUAKES

Shohei NAITO[1], Hiromitsu TOMOZAWA[2], Yuji MORI[3],
Naokazu MONMA[4], Hiromitsu NAKAMURA[5] and Hiroyuki FUJIWARA[6]

[1] Member, Research Fellow, National Research Institute for Earth Science and Disaster Resilience, Ibaraki, Japan, naito@bosai.go.jp
[2] Mizuho Information & Research Institute, Inc., Tokyo, Japan, hiromitsu.tomozawa@mizuho-ir.co.jp
[3] Mizuho Information & Research Institute, Inc., Tokyo, Japan, yuji.mori@mizuho-ir.co.jp
[4] Member, PASCO Corporation, Tokyo, Japan, naamon8759@pasco.co.jp
[5] Member, Deputy Manager, National Research Institute for Earth Science and Disaster Resilience, Ibaraki, Japan, manta@bosai.go.jp
[6] Member, Manager, National Research Institute for Earth Science and Disaster Resilience, Ibaraki, Japan, fujiwara@bosai.go.jp

ABSTRACT: In order to support disaster response activities, we developed an automatic damage classification model using aerial photographs obtained from several earthquakes in Japan. First, we visually classified all buildings into one of four damage levels, then constructed a training and test data set covering four damage levels. By using this training data set, we were able to develop a CNN-based damage detection model with higher performance than previous models. As a result, an average recall value of 70% was obtained, and we confirmed that it is sufficiently accurate to assess the state of disaster damage to wooden buildings in Japan.

Keywords: Remote sensing, Building damage, Damage detection, Deep learning

1. INTRODUCTION

Gaining a rapid understanding of the state of damage immediately following the onset of a disaster is of paramount importance so that various public organizations can make decisions concerning how they should respond. As a result, real-time estimation systems based on space interpolation of earthquake observation records[1)–3)] have been developed to help understand the state of seismic motion damage to various structures just after the onset of a disaster. Although these systems are effective in providing information for determining the degree of damage in the initial stages of a disaster response, the estimated damage is sometimes inconsistent with the actual damage observed because of the influence of localized amplifications of seismic motions, space resolutions for the ground, buildings and other data used for the estimation, uncertainty with the estimation model's damage function, and other factors. It is, therefore, necessary to compile damage information verified by surveys, inspections and patrols, etc., over time[4)] when responding to a disaster from the initial first-aid stages through to the restoration period. However, because a longer time is needed to clarify the true extent of damage by means of field surveys as the scale of damage increases, it is appropriate that remote sensing involving the use of artificial satellites, aircraft, and other means be used in a mutually supplementary manner[5)].

One remote sensing technique that can be adopted to make observations is the use of fixed-wing aircraft. The use of fixed-wing aircraft makes it possible to take images at flight altitudes of about 1 to 2 km under daytime conditions, even with relatively small amounts of cloud cover, thus producing higher resolution images than are possible with an artificial satellite[6)]. Aircraft are also superior to helicopters and unmanned aerial vehicles (UAVs) because an extensive survey can be carried out in a single flight. Furthermore, the overall time taken, from take-off through to imaging, data transfer, orthographic processing and other steps in the preparation of readable images usually only takes a few days, although this depends on when the disaster occurs, weather, and other conditions[7)]. For these reasons, aerial photography is better in terms of resolution, coverage and speed, and can be said to be an effective remote sensing tool for determining the state of damage.

When determining the amount of building damage due to an earthquake, for example, the Guideline for Operating the Assessment Criteria for Disaster Damage to Dwelling Houses produced by the Cabinet Office (announced in March 2020)[8)] states that "a dwelling house can be judged to be completely destroyed if its collapse is confirmed by reading an aerial photograph"; aerial photography is expected to be further utilized in support for disaster responses in the future. Previous studies have shown that the normal resolution aerial photographs allow effective visual estimation of building damage, and that such damage can be classified into up to four levels, approximately, when using vertical images taken from fixed-wing aircraft, or up to six levels when using oblique images taken from a helicopter or the like[9)–13)].

In the case of large disasters, however, damage detection by visual means can be problematic—requiring much time and many personnel. Research has, therefore, been conducted into the automatic assessment of building damage using aerial photographs and machine learning techniques, including deep learning[14)–21)]. This has shown that machine learning, using post-quake images, can facilitate the automatic identification of collapsed or swept-away buildings[14)–20)] and blue tarp-covered buildings[21)] with a high degree of precision.

Our group, on the other hand, have also developed two machine learning models that classify building damage into four levels, to enable damage assessment to be carried out even when earthquakes cause many buildings to be half or even completely destroyed, although not collapsed. These models were developed using vertical images taken by a fixed-wing aircraft just after the main shock of the 2016 Kumamoto Earthquake[22), 23)]. One of the models we have developed is for SVM classification based on image characteristic amounts, and the other is a deep learning model using a convolutional neural network. When these two techniques were compared in terms of damage detection accuracy, the deep learning model was found to be the more accurate. However, since this deep learning model has been optimized for use with the images acquired after the main shock of the 2016 Kumamoto Earthquake, there is a problem with decreased accuracy due to its inability to accurately distinguish between buildings and non-buildings and to accurately determine the degree of damage when applied to the aftermath of other earthquakes, although it will still exhibit relatively high detection accuracy when applied to other aerial photographs taken in the 2016 Kumamoto Earthquake around the same time. This issue is described in further detail in Section 4.1, based on actual verification results.

Since the present study aims to make the best use of a building damage detection model to support disaster responses, it is necessary to be able to quickly and accurately identify damage not just from past earthquakes but also in the event of any future disaster. Therefore, in order to improve the generalization performance of the existing model[22)], deep learning was performed using six kinds of aerial photograph taken under different imaging conditions and for more than one earthquake. First, using vertical aerial photographs taken just after the onset of the 1995 Southern Hyogo Prefecture Earthquake, the 2011 off the Pacific Coast of Tohoku Earthquake, the foreshock and main shock of the 2016 Kumamoto Earthquake, and the 2018 Hokkaido Iburi Eastern Earthquake, damage levels sustained by buildings shown in the images obtained were classified into four levels. Comparing these degrees of damage with those obtained by field surveys, we confirmed that damage could be identified at practically feasible levels of accuracy with up to three degrees of damage.

Next, we developed an automatic classification model for building damage that was capable of determining damage ratings for a wider-than-normal range of aerial photographs by performing deep learning—using these ratings and building images extracted from various different aerial photographs as training data. Employing this model, we then developed a program for generating colored images from aerial photographs where the color shows the degree of building damage. We also emphasized speed of operation, enabling users to quickly identify regions with a high concentration of damage and to determine the overall degree of damage by superposing output images and aerial photographs, even if aerial photographs were the only data source available. Furthermore, we tried to make it possible to compile data on the number of buildings by degree of three damage levels, if polygon data for buildings were available, by superposing the data with such images and classifying the degrees of damage by setting threshold levels.

As stated above, the present study was aimed at helping various governmental and private organizations in Japan to speed up and streamline their actions to ensure a quick recovery from earthquake disasters involving major damage to wooden buildings with varying degrees of destruction by developing a higher generalization performance and effective damage detection model that would allow them to immediately assess the overall state of damage.

2. PREPARATION OF TRAINING DATA

2.1 Procedures for visually checking aerial photographs

The training data used in the present study were derived from several target earthquakes that were selected from among the various disasters that had recently occurred in Japan. Specifically, we used data on building damage that were classified by visually checking vertical aerial photographs taken from fixed-wing aircraft for more than one earthquake that damaged wooden buildings over a wide area due to seismic motions (at an intensity of 6-lower or more) and compared them with aerial photographs obtained soon after the onset of the disaster. As shown in Table 1, imaging conditions were classified into six types. The data used for the 1995 Southern Hyogo Prefecture Earthquake were obtained by scanner-digitizing and orthographic processing of 23 aerial photographs of Kobe and Ashiya that had been taken by Nakanihon Air Co., Ltd., using an analogue camera on the day of the disaster (Fig. 1). The data used for the 2011 off the Pacific Coast of Tohoku Earthquake were obtained from 20 orthographic images that had been taken by PASCO Corporation above Sendai City, using a digital area sensor (UCX), within one month of the disaster (Fig. 2). The data used for the 2016 Kumamoto Earthquake were obtained from 13 and 20 orthographic images that had been taken by PASCO Corporation over Mashiki Town and other affected regions, using a digital multiline sensor (ADS), several days after the foreshock of April 14 and the main shock of April 16, respectively (Figs. 3 and 4). For generalization performance verification purposes, locations of training data and test data were exchanged after the foreshock (Fig. 3) and main shock (Fig. 4). The data used for the main shock of the 2016 Kumamoto Earthquake were obtained from 17 orthographic images that had been taken above Minami-Aso Village and Nishihara Village using a UCX (Fig. 5). The data used for the 2018 Hokkaido Iburi Eastern Earthquake were obtained from 10 orthographic images in regions where the intensity was estimated to be 5-upper or greater using the J-

RISQ real-time damage estimation system[3], and where a relatively large number of buildings were present, using aerial photographs that had been taken by PASCO Corporation above Chitose City, Abira Town, Atsuma Town, and Mukawa Town using a UCX, several days after the onset of the earthquake (Fig. 6). Referring to the left panels in Figs. 1–6, the blue/gray-framed images are aerial photographs for training; the red-framed images are aerial photographs for testing (A and B show how to acquire the training data described in Section 2.3 below); the background map indicates estimated quake intensities based on a J-RISQ earthquake report[3]. The purple-framed images in the right panel are aerial photographs following the frame-to-frame cross validation described in Section 4.5 below.

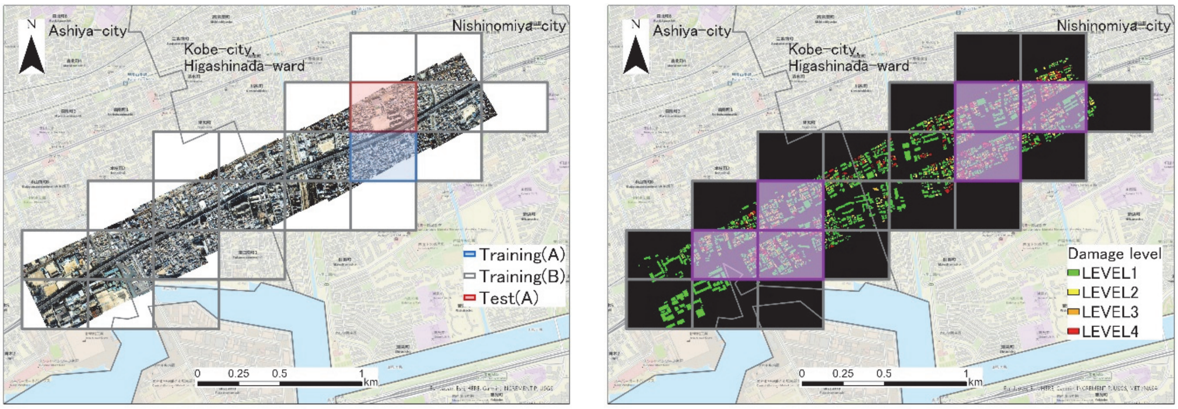


Fig. 1 Training data for the 1995 Southern Hyogo Prefecture Earthquake (left panel, aerial photographs; right panel, masked images)

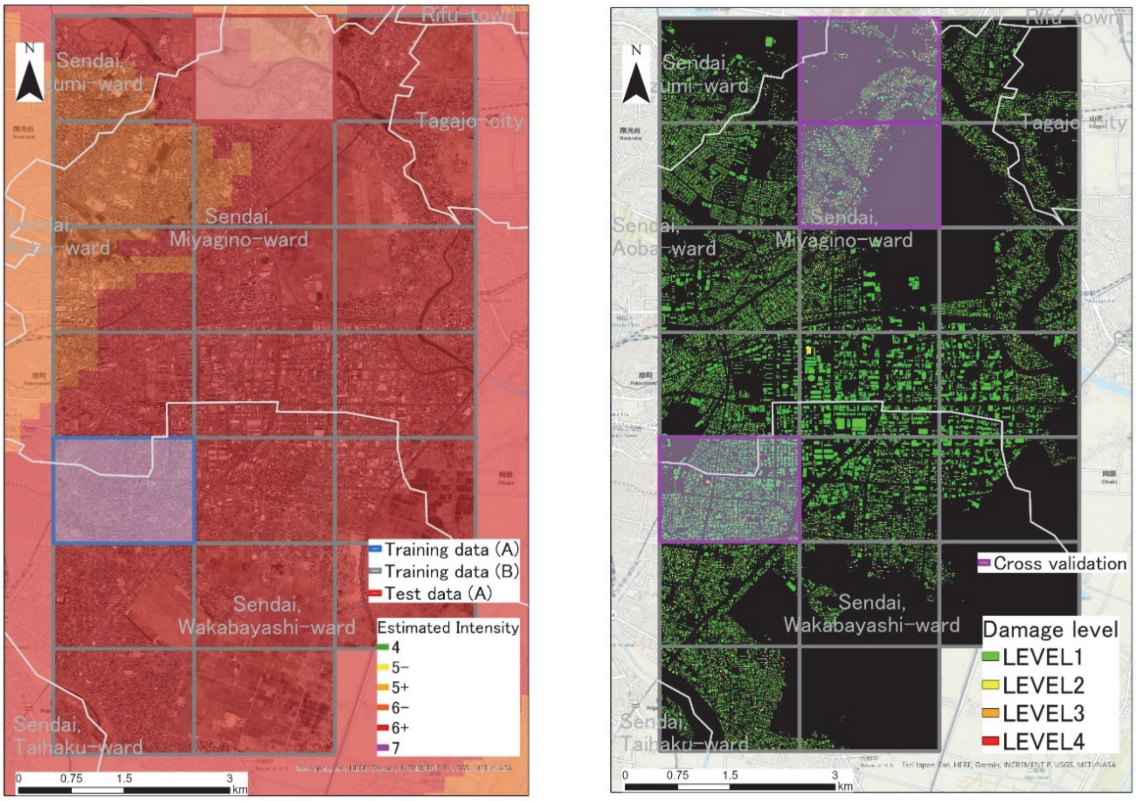


Fig. 2 Training data for the 2011 off the Pacific Coast of Tohoku Earthquake

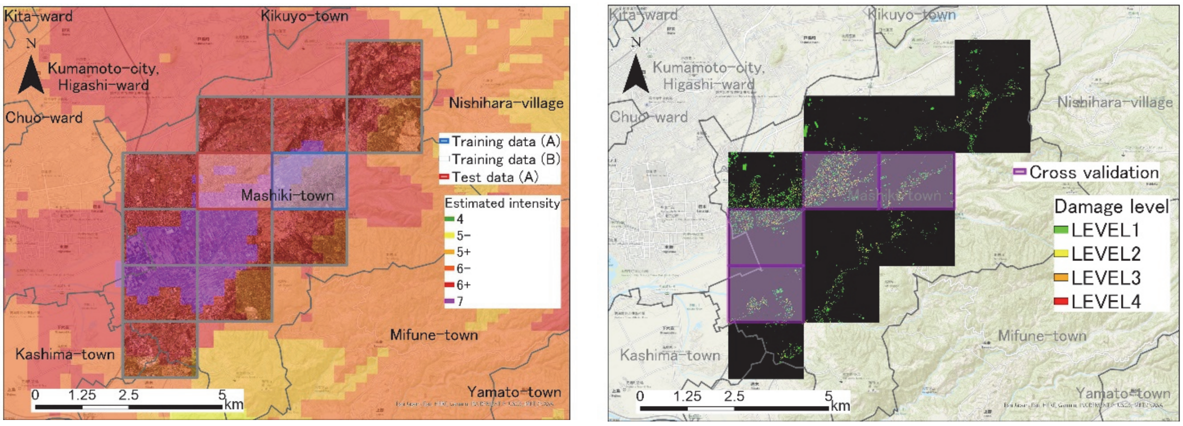


Fig. 3 Training data for the 2016 Kumamoto Earthquake (foreshock)

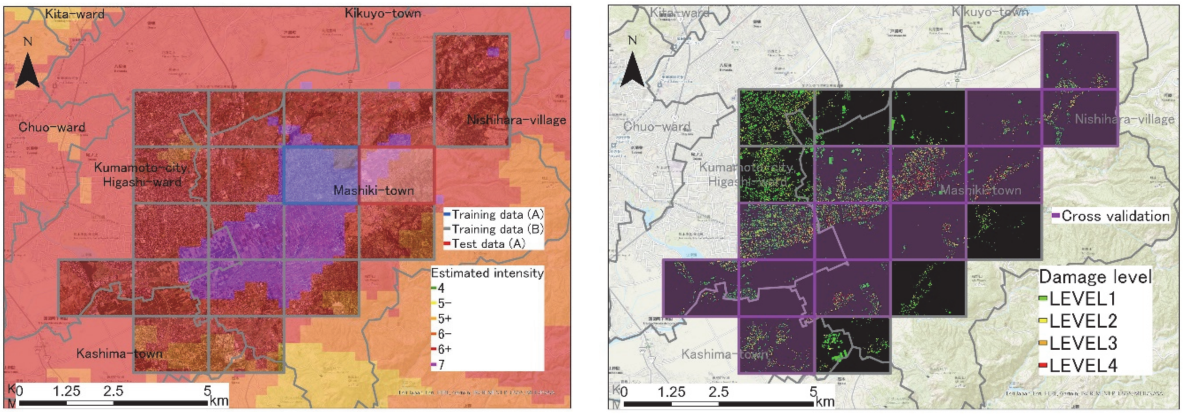


Fig. 4 Training data for the 2016 Kumamoto Earthquake (main shock, around Mashiki Town)

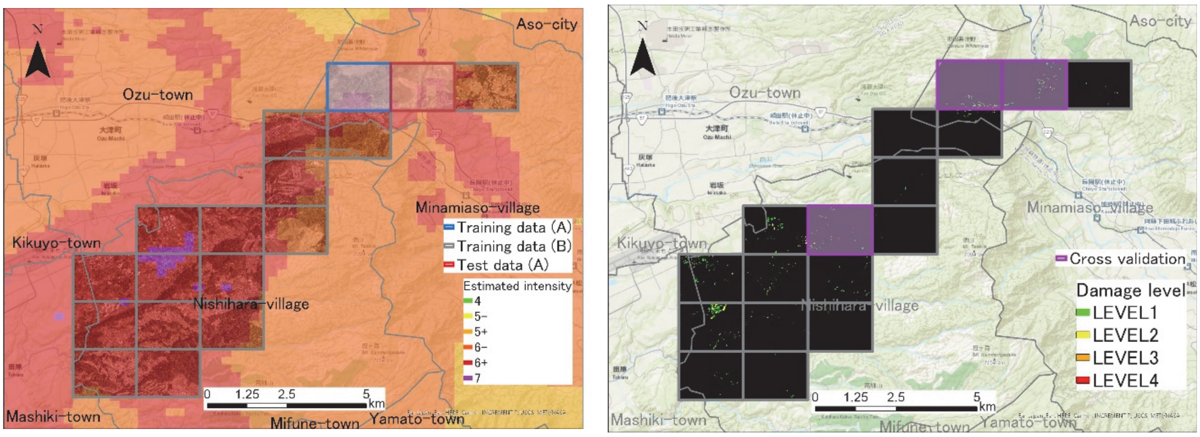


Fig. 5 Training data for the 2016 Kumamoto Earthquake (main shock, around Minami-Aso Village)

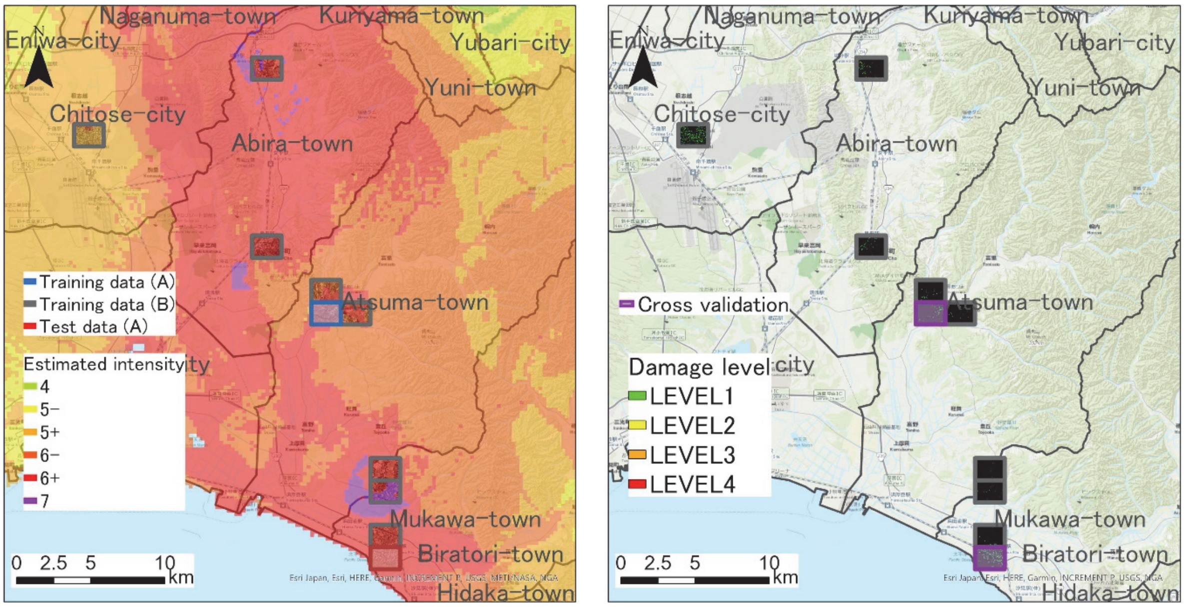


Fig. 6 Training data for the 2018 Hokkaido Iburi Eastern Earthquake

Table 1 Imaging conditions for visually checked aerial photographs

	Year of imaging	Date of imaging	Imaging organization	Imaging equipment	Resolution	Number of images
Southern Hyogo Prefecture Earthquake	1995	January 17	Nakanihon Air Co., Ltd.	Analogue	4 cm	23
off the Pacific Coast of Tohoku Earthquake	2011	March-April	PASCO Corporation	Area sensor	20 cm	20
Kumamoto Earthquake (foreshock)	2016	April 15	PASCO Corporation	Line sensor	20 cm	13
Kumamoto Earthquake (main shock) Mashiki Town	2016	April 19	PASCO Corporation	Line sensor	20 cm	20
Kumamoto Earthquake (main shock) Minami-Aso Village	2016	April 19	PASCO Corporation	Area sensor	20 cm	17
Hokkaido Iburi Eastern Earthquake	2018	September 11	PASCO Corporation	Area sensor	20 cm	10

The actual spatial region included in each aerial photograph covered 0.4 km × 0.3 km for the 1995 Southern Hyogo Prefecture Earthquake and 2 km × 1.5 km for the other earthquakes. Since the 1995 Southern Hyogo Prefecture Earthquake had a different image resolution, we unified the sizes of the buildings in the patch images by reducing the image size to a resolution of 20 cm when identifying those patch images to be used as deep learning training data (described in Section 2.3 below).

Prior to preparing training data, building damage ratings were first classified into four levels by visually checking each aerial photograph, and building damage data were constructed with the use of polygons in the Japan Basic Map Information (Buildings) package[24] on ArcGIS. Even when an overlay showed a shape discrepancy between the buildings in the aerial photographs and the polygons, the most severely damaged building in the polygon was targeted for labeling with the appropriate level of damage. When no building was present in a polygon, the polygon was removed from the building training data. In contrast, when a building was found in the aerial photograph but there was no building polygon, no training data, as a rule, were newly constructed. For the 1995 Southern Hyogo Prefecture Earthquake, however, polygons were prepared by tracing building portions on aerial photographs using ArcGIS, since there was a major discrepancy between the building locations at the time of the onset of the earthquake and the polygons in the Japan Basic Map Information. The criteria used for the level of damage were the same as those in previous reports[22), 23), 25)]. When referring to a damage pattern chart for wooden buildings[26)], the criteria were defined to allow building damage to be rated using aerial photographs alone (Table 2).

Next, the building regions in the aerial photographs were classified by level of damage into the four colors shown in Table 2, using the building damage data, and a masked image was prepared with the non-building portions masked black. The above procedures yielded building damage GIS data for a total of 111,686 buildings (Table 3) and 103 masked images from 103 visually checked aerial photographs. These data and images were then used as the training data for the subsequent deep learning stage (training and testing data are described in Section 2.3 below). The coverage of the visually checked aerial photographs and the masked images are shown in Figs. 1–6.

Table 2 Definitions of building damage levels derived from aerial photographs[22), 23), 25)]

Degree of damage	Photographic features (any one is met)	Correspondence to Okada and Takai[26)]	Color (RGB)
LEVEL 1 (no damage)	Damage unidentifiable from aerial photographs	D0	Green (0,255,0)
LEVEL 2 (minor damage)	Some roof tiles were dislodged, with the roof partially covered with blue tarp	D1	Yellow (255,255,0)
LEVEL 3 (moderately damaged)	Most roof tiles were dislodged, with the roof mostly covered with blue tarp and some walls detached	D2, D3	Orange (255,127,0)
LEVEL 4 (severely damaged)	Marked inclinations, shifts or deformation of the entire building Story destruction and collapse	D4, D5	Red (255,0,0)

Table 3 Number of damaged buildings within the scope of the training data constructed

	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4	Total
1995 Southern Hyogo Prefecture Earthquake	1,880	548	396	534	3,358
2011 off the Pacific Coast of Tohoku Earthquake	47,338	5,416	1,894	47	54,695
2016 Kumamoto Earthquake (foreshock)	9,734	2,706	512	209	13,161
2016 Kumamoto Earthquake (main shock, Mashiki Town)	20,201	6,161	2,674	1,570	30,606
2016 Kumamoto Earthquake (main shock, Minami-Aso Village)	789	295	119	54	1,257
2018 Hokkaido Iburi Eastern Earthquake	8,440	113	28	28	8,609
Total	88,382	15,239	5,623	2,442	111,686

2.2 Verification of accuracy after visually checking aerial photographs

In the present study, data on the four building damage levels determined by visually checking vertical aerial photographs were used as training data to build a damage detection model. Since the objective of the study was to use this model to support disaster responses, it was necessary to verify the correspondence between the damage levels determined by visual checking of aerial photographs and actual levels of building damage. This was determined by comparing visual check results with damage level data from an actual field survey.

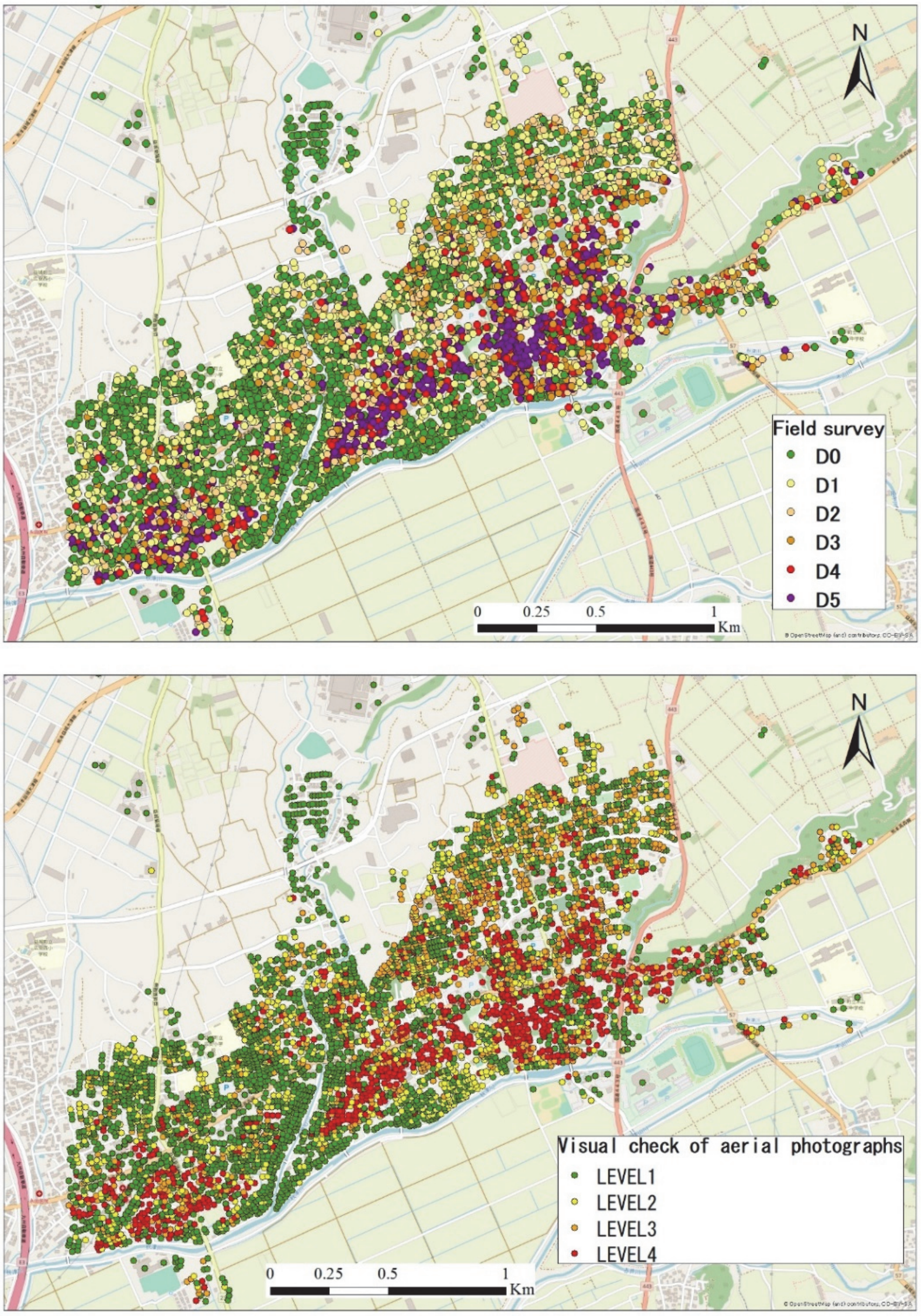


Fig. 7 Comparison of levels of building damage in a field survey (upper panel) and after checking aerial photographs (lower panel)[28), 29)]

We used the results from a visual appearance survey of building damage conducted around the Miyazono area in Mashiki Town from May to October 2016, just after the onset of the 2016 Kumamoto Earthquake[27)] as field survey data for comparison purposes. In the survey, the level of damage to each building was classified into six levels (D0 to D5) based on a damage pattern chart[26)]. For the 5,520 buildings in the same area, shown in Fig. 7, field survey results[27)] and damage levels based on visual checking of aerial photographs[25)] were compared using confusion matrices. The correspondence between the six damage levels in the field survey and the four damage levels in the visually checked aerial photographs was defined as shown in Table 2. In this correspondence of the various damage levels, the combination showing the highest correlation coefficient was selected[28)].

The term confusion matrix (Table 4) refers to a matrix showing label concordance between predicted data and actual data, in which the terms “true positive (TP)”, “false positive (FP)”, “false negative (FN)”, and “true negative (TN)” were defined in terms of label concordance/non-concordance. Usually, “recall,”, “precision,”, “F-measure,”, and “accuracy” are also used as performance indicators in a confusion matrix[22)] (Table 5).

Table 4 Confusion matrix

		Prediction (aerial photograph ratings)	
		Positive	Negative
Right answer (field surveys)	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

Table 5 Performance indicators

Index	Method of calculation
Recall	$\frac{TP}{TP+FN}$
Precision	$\frac{TP}{TP+FP}$
F-measure	$\frac{2}{\frac{1}{Precision}+\frac{1}{Recall}}$
Accuracy	$\frac{TP+TN}{TP+FP+FN+TN}$

Table 6 shows confusion matrices for the building quantity relationships based on a comparison between the levels of damage derived from the visual checking of aerial photographs and field surveys. It can be seen that both the precision and recall were no lower than 70% in the relationship between LEVEL 1 in the visually checked aerial photographs and D0 in the field survey, and the relationship between LEVEL 4 in the visually checked aerial photographs and D4+D5 in the field survey, whereas in the relationships between the intermediate damage levels LEVEL 2 and D1 and between LEVEL 3 and D2+D3, the recall was approximately 31% to 43% and the precision was approximately 44%. Overall, the average recall for all the levels was approximately 58.2%. This finding is attributable to the fact that the focus differed between the different study approaches, and that wall cracks, detachments, etc., are sometimes overlooked when using vertical aerial photographs only[29)]. As shown in Fig. 8, therefore, the proportion of undamaged buildings (LEVEL 1) in all ratings tended to be slightly higher with aerial photographs, while the proportion of largely damaged buildings (LEVEL 4) was generally similar. In damage detection using vertical aerial photographs only, the detection accuracy tends to decrease with decreasing degree of damage, and this tendency was also noted in previous studies using damage survey data from the 1995 Southern Hyogo Prefecture Earthquake[9)] and the 2011 off the Pacific Coast of Tohoku Earthquake[28)].

In the present study, therefore, the relatively low damage levels (LEVEL 2 and LEVEL 3) were merged to produce only three levels in total (LEVEL 1, LEVEL 2+3, LEVEL 4) in the visual checking of aerial photographs. Similarly, the data from the field survey were also reviewed using three levels of damage: D0 (no damage), D1–D3 (partial to half collapse), and D4 and D5 (complete collapse). The resulting data were compared using confusion matrices (Table 7). In this case, the recall and precision for LEVEL 2+3 both improved to approximately 62.1% and 75.7%, respectively, and the average recall and average precision for the three levels, overall, were approximately 73.9% and 73.6%, respectively. Therefore, the system of classification involving three damage levels was found to produce higher accuracy for training data.

Table 6 Confusion matrices comparing a field survey and visual check of aerial photographs[28), 29)] (four damage levels)

		Visual check of aerial photographs					Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4	Total		
Field survey	D0	1,878	237	69	21	2,205	85.2	77.7
	D1	470	409	414	39	1,332	30.7	36.1
	D2+D3	223	213	458	180	1,074	42.6	43.3
	D4+D5	59	72	102	676	909	74.4	74.1
	Total	2,630	931	1,043	916	5,520	-	-
Precision (%)		71.4	43.9	43.9	73.8	-	Accuracy (%)	62.0

Table 7 Confusion matrices comparing a field survey and visual check of aerial photographs (three damage levels)

		Visual check of aerial photographs				Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4	Total		
Field survey	D0	1,878	306	21	2,205	85.2	77.7
	D1+D2+D3	693	1,494	219	2,406	62.1	68.2
	D4+D5	59	174	676	909	74.4	74.1
	Total	2,630	1,974	916	5,520	-	-
Precision (%)		71.4	75.7	73.8	-	Accuracy (%)	73.3

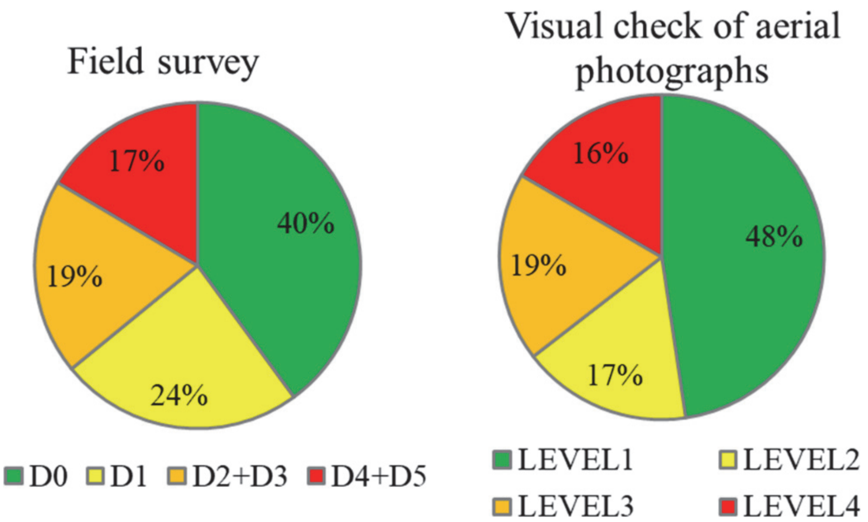


Fig. 8 Proportion of data from a field survey (left panel) and visual check of aerial photographs (right panel) (5,520 buildings in total)[28)]

These findings show that, when using only post-quake vertical aerial photographs to assess the state of building damage, it is appropriate, from a practical viewpoint, to classify damage into three levels: "no damage,", "minor to moderate damage", and "severe damage". In the present study, we also developed a deep learning model in which the extent of damage was automatically classified into four levels using visual check results based on the more extensive 4-level classification used as training data. With this in mind, the model output imaging results shown below are displayed using the 4-level classification described in the present article. In the discussion of detection accuracy, however, LEVEL 2 and LEVEL 3 were merged to obtain only three levels of damage: LEVEL 1 (no damage), LEVEL 2+3 (minor to moderate damage), and LEVEL 4 (severe damage). Likewise, the 3-level classification was used for the number of damaged buildings, as described in detail in Section 4.6 below.

2.3 Automatic identification of patch images

In image recognition by machine learning, the usual approach is to use small images prepared by cutting out the portions around the target object from the entire image, which contains various objects (patch images)[30]. In the present study, 80-pixel square patch images, which contained the majority of ordinary houses, were automatically extracted from the entire area shown on each aerial photograph to yield deep learning training data. The identified patch images were then classified into five categories: LEVEL 1, LEVEL 2, LEVEL 3, LEVEL 4, and non-buildings. The non-buildings category included random non-building items such as trees, mountainous areas, grass fields, agricultural land, parking lots, railways and roads. Examples of patch images taken from aerial photographs of the various earthquakes examined are shown in Fig. 9.

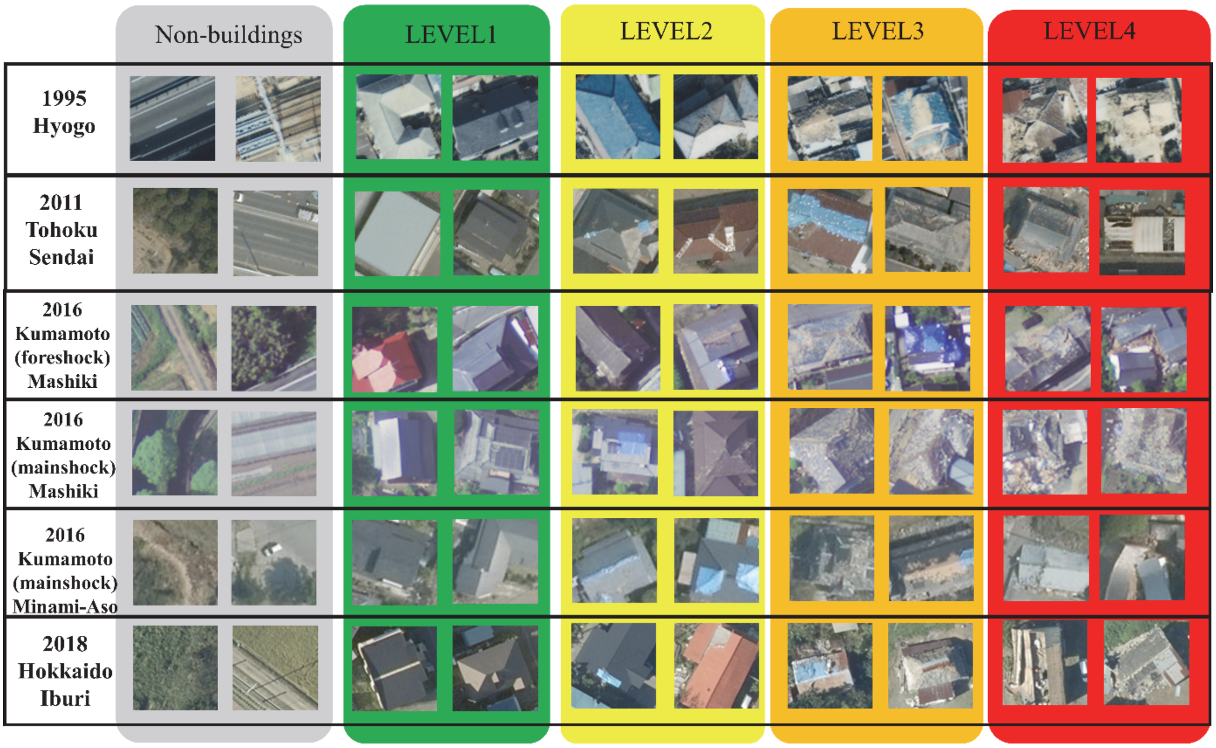


Fig. 9 Examples of patch images from the various earthquakes

Patch images were automatically acquired using two methods: (A) Data were acquired exclusively from aerial photographs showing relatively high damage levels, in order to obtain training data for both highly damaged buildings and for non-buildings. (B) Data were uniformly acquired preferentially for buildings shown on aerial photographs in order to increase the varieties of training data available for various damage levels (Fig. 10). The scan span was set at 20 pixels for aerial photographs from the 1995 Southern Hyogo Prefecture Earthquake, which covered a small area in each photo, and at 40 pixels for other aerial photographs (Table 8).

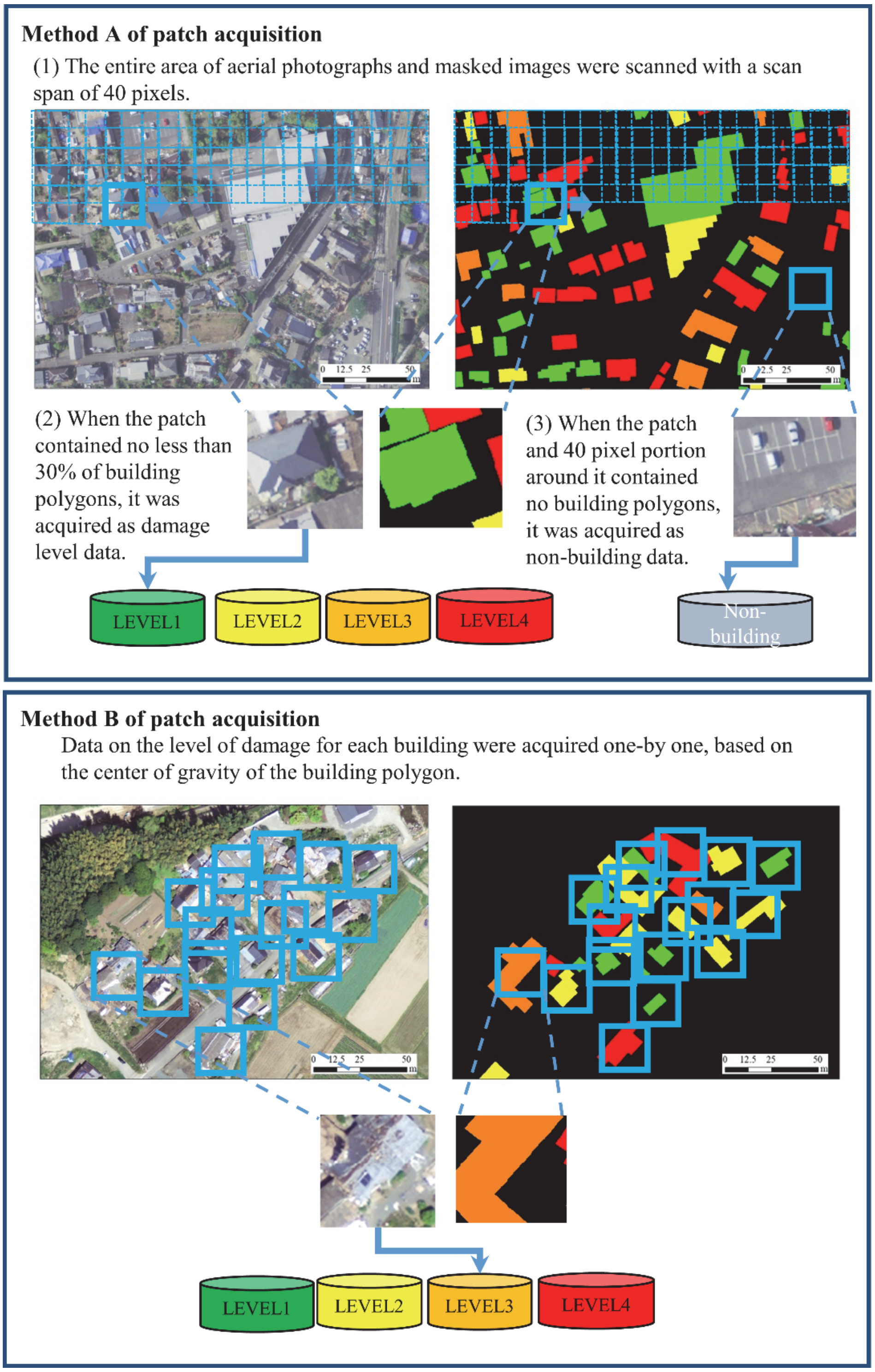


Fig. 10 Outline of various methods used for automatic acquisition of patch images[28)]

Table 8 Definitions of methods used for automatic acquisition of patch images

Method	Target aerial photograph	Target label	Method of patch acquisition
A	One image for training One image for testing	Buildings	The entire area was scanned with a scan span of 40 pixels. When the patch contained no less than 30% of building polygons it was acquired as damage level data (scan span set at 20 pixels for the 1995 Southern Hyogo Prefecture Earthquake only)
		Non-buildings	The entire area was scanned with a scan span of 40 pixels. When the patch and 40-pixel portion around it contained no building polygons it was acquired as non-building data (scan span set at 20 pixels for the 1995 Southern Hyogo Prefecture Earthquake only)
B	Other images for training	Buildings	Data on the level of damage for each building were acquired one-by-one, based on the center of gravity of the building polygon.

Specific procedures are described below. First, one training photo and one testing photo were selected from among the range of aerial photographs from each earthquake, and respective patch images were automatically identified using method A. Next, using the remaining aerial photographs, patch images were automatically identified using method B. The training patch images acquired using methods A and B were merged for use as training data. The locations of the aerial photographs used as training data (methods A and B used in combination) and test data (method A only), as described above, are shown in Figs. 1–6, respectively. A total of 431,974 items of training and testing data were automatically acquired using the above procedures (Table 9). For generalization performance verification, after the foreshock and main shock in the 2016 Kumamoto Earthquake, locations were exchanged to prevent the training data and testing data acquired using method A being duplicated in the same place (Figs. 3 and 4).

Next, to improve the quality of training data, an image checker (other than the visual checker used for the aerial photographs) visually checked each patch image and removed data that seemed to be difficult to classify from patch images alone for any of the following reasons (Fig. 11). This screening yielded a total of 409,818 patch images for use as training and testing data (Table 10). The reasons of data removal are as follows, and the impacts of this data screening on the detection accuracy are described in Section 4.3 below.

- The items in the image are difficult to perceive because of the large unwanted images of the outer frames of the aerial photographs and reflected light.
- Buildings are difficult to identify because the images were too dark or hidden by vegetation etc.
- Because there are marked differences in building location between building polygons and aerial photographs, non-buildings were cut out as buildings, or buildings were cut out as non-buildings.
- There are markedly different buildings around the target building, which have different levels of damage.
- The checker's assessments differed markedly between aerial photograph ratings and the visual check of images.

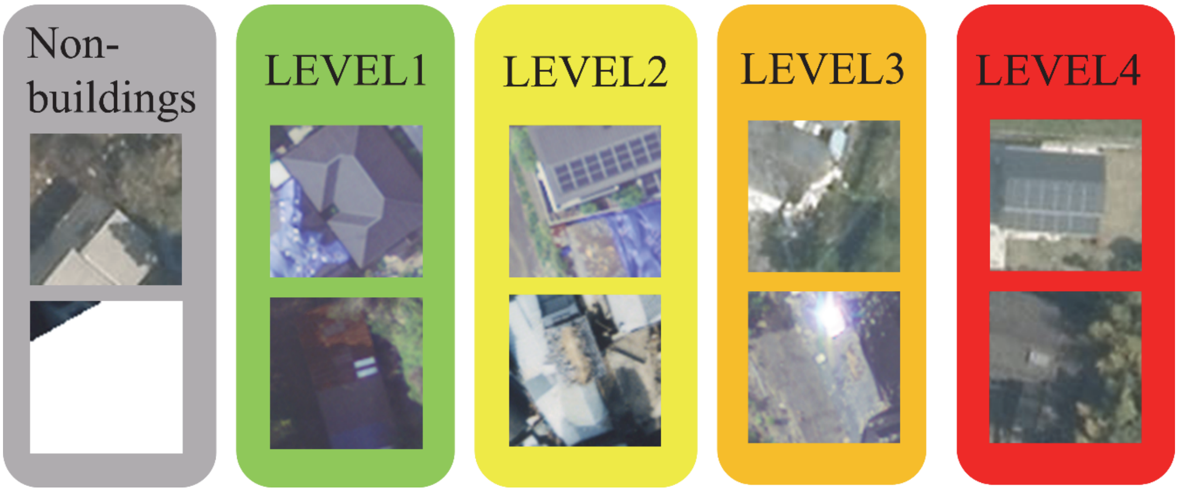


Fig. 11 Examples of patch images excluded from the training data

Table 9 Number of automatically identified patch images

Target earthquake	Type (method of determination)	Non-buildings	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4	Total
1995 Southern Hyogo Prefecture Earthquake	For training (A)	1,484	1,354	125	150	591	3,704
	For training (B)	0	909	177	99	216	1,401
	For testing (A)	1,462	1,450	219	115	275	3,521
2011 off the Pacific Coast of Tohoku Earthquake	For training (A)	5,187	19,492	675	145	5	25,504
	For training (B)	0	36,281	4,066	1,504	30	41,881
	For testing (A)	27,795	4,208	849	399	24	33,275
2016 Kumamoto Earthquake foreshock	For training (A)	36,119	1,499	692	105	39	38,454
	For training (B)	0	5,059	1,178	190	61	6,488
	For testing (A)	19,635	5,207	2,604	788	241	28,475
2016 Kumamoto Earthquake (Mashiki Town)	For training (A)	19,565	3,424	1,180	1,922	2,064	28,155
	For training (B)	0	14,532	4,665	1,520	533	21,250
	For testing (A)	35,894	889	437	458	490	38,168
2016 Kumamoto Earthquake (Minami-Aso Village)	For training (A)	42,560	243	224	84	35	43,146
	For training (B)	0	592	214	80	31	917
	For testing (A)	42,168	547	107	70	70	42,962
2018 Hokkaido Iburi Eastern Earthquake	For training (A)	37,247	1,988	23	5	17	39,280
	For training (B)	0	5,336	58	16	9	5,419
	For testing (A)	24,402	5,502	37	2	31	29,974
Grand total		293,518	108,512	17,530	7,652	4,762	431,974

Table 10 Number of patch images following visual screening

Target earthquake	Type (method of determination)	Non-buildings	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4	Total
1995 Southern Hyogo Prefecture Earthquake	For training (A)	509	1,071	80	127	333	2,120
	For training (B)	0	754	89	85	133	1,061
	For testing (A)	778	1,280	170	87	140	2,455
2011 off the Pacific Coast of Tohoku Earthquake	For training (A)	5,176	18,898	378	75	2	24,529
	For training (B)	0	35,571	3,359	1,261	15	40,206
	For testing (A)	27,788	3,638	562	337	11	32,336
Foreshock of the 2016 Kumamoto Earthquake	For training (A)	36,106	1,221	165	36	9	37,537
	For training (B)	0	4,075	881	123	45	5,124
	For testing (A)	19,158	4,859	1,644	683	137	26,481
2016 Kumamoto Earthquake (Mashiki Town)	For training (A)	19,467	2,785	306	1,251	1,187	24,996
	For training (B)	0	12,936	3,378	1,254	336	17,904
	For testing (A)	35,887	707	148	153	184	37,079
2016 Kumamoto Earthquake (Minami-Aso Village)	For training (A)	41,409	179	116	49	20	41,773
	For training (B)	0	463	112	75	13	663
	For testing (A)	41,486	508	34	35	44	42,107
2018 Hokkaido Iburi Eastern Earthquake	For training (A)	36,462	1,936	11	1	2	38,412
	For training (B)	0	5,296	18	10	5	5,329
	For testing (A)	24,228	5,449	8	2	19	29,706
Grand total		288,454	101,626	11,459	5,644	2,635	409,818

3. DEVELOPING A DAMAGE DETECTION MODEL USING DEEP LEARNING

In the present study, a damage detection model was developed using a convolutional neural network (CNN) —a deep learning technique with confirmed high performance for image classification[31)]. In the CNN, features were extracted by combining, in multiple stages, a convolution layer in which convolution computing was performed for each pixel z_{ij} of the image u_{ij} with the filter h_{pq} of the size $(H \times H)$ (Eq. (1)), and a pooling layer in which the image was compressed by extracting a representative value (maximum in the present study) from the pixel z_{pq} in the layer region P_{ij} (Eq. (2)). In the final layer, i.e., the fully connected layer, the probability y_k that the input x belongs to the class C_k was output using the Softmax function (Eq. (3)). To minimize the cross entropy E_n (Eq. (4)) indicating the error between this output value y_k and the training dataset $t_{n,k}$, the weighting parameter h_{pq} was updated using the error back propagation method (a method in which the value δ obtained by differentiating the E_n by the input value u_{ij} (Eq. (5)) is calculated retrogradely toward the input layer (l-1 layer) side), and repeated calculations were made until the error became a minimum value[32)] (Fig. 12).

$$u_{ij}{}^{(l)} = \sum_{p=0}^{H-1}\sum_{q=0}^{H-1} z_{i+q,j+q}{}^{(l-1)} h_{pq} \tag{1}$$

$$u_{ij}{}^{(l)} = \max_{(p,q)\in P_{ij}} z_{pq}{}^{(l-1)} \tag{2}$$

$$y_k = p(C_k|x) = \frac{exp(u_k)}{\sum_{j=1}^{K} exp(u_j)} \tag{3}$$

$$E_n = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{n,k} \log y_k \tag{4}$$

$$\delta_{ij}{}^{(l)} \equiv \frac{\partial E_n}{\partial u_{ij}{}^{(l)}} \tag{5}$$

In the present study, improvements were made to enable simultaneous learning using multiple datasets with increased generalization performance, on the basis of a CNN model developed in a prior study[22)]. The CNN model, like those described in the literature[22)], was based on VGG-16[33)] However, VGG-16 originally assumed 224-pixel square input images, losing image texture in the downstream pooling layer, thus reducing the detection accuracy[28)]. For this reason, the model was rebuilt with the convolution layer shallowed according to the 80-pixel square patch image (Table 11).

In addition, the model described in the literature[22)] showed learning instability due to increased variation in precision and loss values for each epoch. Therefore, the stochastic gradient descent (SGD) method was used for optimization to linearly reduce the learning coefficient for each epoch and to stabilize the loss function value. In addition, efforts were made to reduce the variation in training data and improve distinguishing performance by adding batch normalization layers and dropout layers. In building this model, the deep learning framework Keras[34)] and the Python programming language were used. Table 12 summarizes a comparison of the models described in the literature[22)] and the present study. The selected CNN parameters used are shown in Table 13.

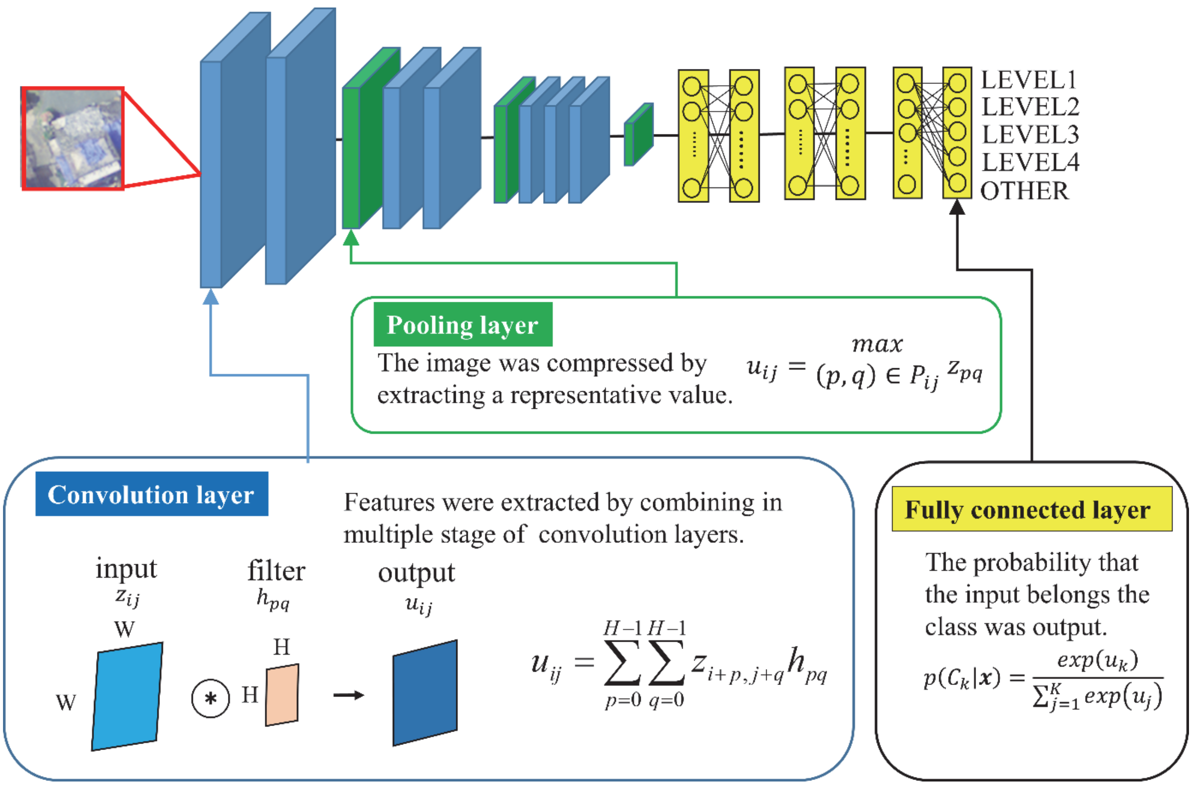


Fig. 12 Outline of CNN-based machine learning model[28)]

Table 11 Rebuilt CNN model[28)]

Layer	Data size	Filter size	Stride	Activation function
Input	80×80×3	-	-	-
Convolution layer	80×80×64	3×3	1	ReLU
Batch normalization layer	80×80×64	-	-	-
Convolution layer	80×80×64	3×3	1	ReLU
Batch normalization layer	80×80×64	-	-	-
Pooling layers	40×40×64	2×2	2	-
Convolution layer	40×40×128	3×3	1	ReLU
Batch normalization layer	40×40×128	-	-	-
Convolution layer	40×40×128	3×3	1	ReLU
Batch normalization layer	40×40×128	-	-	-
Pooling layers	20×20×128	2×2	2	-
Convolution layer	20×20×256	3×3	1	ReLU
Batch normalization layer	20×20×256	-	-	-
Convolution layer	20×20×256	3×3	1	ReLU
Batch normalization layer	20×20×256	-	-	-
Convolution layer	20×20×256	3×3	1	ReLU
Batch normalization layer	20×20×256	-	-	-
Pooling layers	10×10×256	2×2	2	-
Fully connected layer	1×1×1024	-	-	ReLU
Dropout layer	1×1×512	-	-	-
Fully connected layer	1×1×1024	-	-	ReLU
Dropout layer	1×1×512	-	-	-
Fully connected layer	1×1×5	-	-	Softmax
Output	-	-	-	-

Table 12 Comparison with a damage detection model reported in a prior study

	Model described in the literature[22)]	Model in the present study
CNN composition	Convolution layers: 13 layers Pooling layers: 5 layers Fully connected layers: 2 layers	Convolution layers: 7 layers Pooling layers: 3 layers Fully connected layers: 3 layers
Acquisition of learning data	Learning with use of some data (5,000 data items for each class) extracted from all data	Learning with 1000 randomly acquired data items per 2 epochs for each class with data exchange
Data augmentation	No	Yes
Learning coefficient	Fixed value	Linearly attenuated for each epoch
Method of optimization	Adam	SGD
Model adoption criteria	The model of epochs showing the highest accuracy was adopted.	The model of epochs showing the largest minimum value of recall was adopted.
Framework	MXNet	Keras

The training algorithm used in the present study is described in detail below. First, the training data were divided into those for training and those of verification. In the first epoch, CNN-based learning was performed on the basis of 100 mini-batches using 1000 data items randomly extracted from among the training data for each of the five classes (four damage levels + non-buildings). In this case, data augmentation (lateral reversal and RGB brilliancy value ± five random shifts) was performed to yield 1000 data items even when there were less than 1000 originally present. In the next epoch, learning was performed again using 1000 data items acquired in the previous epoch. Thereafter, 1000 new data items were randomly extracted for an even-numbered epoch, followed by data augmentation, whereas for an odd-numbered epoch, learning was performed using the data acquired in the previous epoch. This procedure was performed repeatedly. Class rating was performed for each epoch using verification data in order to calculate the minimum recall for each class as the assessment value V. Following the above procedure, a model of the epoch for the maximum assessment value V, after the end of the last epoch, was adopted as the training model (Fig. 13). The above procedure was performed for both training and detection, and it was found that the detection accuracy of the verification data eventually stabilized[28)]. In the dataset used in the present study, the detection accuracy of verification data stabilized within 500 to 2000 epochs, as shown in Fig. 14. Here are the time courses of detection accuracy and loss function in 10-fold cross validation performed using the learning data shown in Table 10. Therefore, the final model was developed basically using 2000 epochs. In the cross validation described in Section 4.5 below, however, the learning time was greatly prolonged, so the number of attempts was set at 500 epochs.

By performing learning using 1000 data items for each class, as described above, the numbers for each class became constant, preventing over-learning for any particular class. In addition, efforts were made to prevent over-learning by exchanging the data for every two epochs, and to shorten the data reading time. Furthermore, the use of a model for the maximum assessment value made it possible to improve the data accuracy for every class even when the number of data items was not uniform among the various classes.

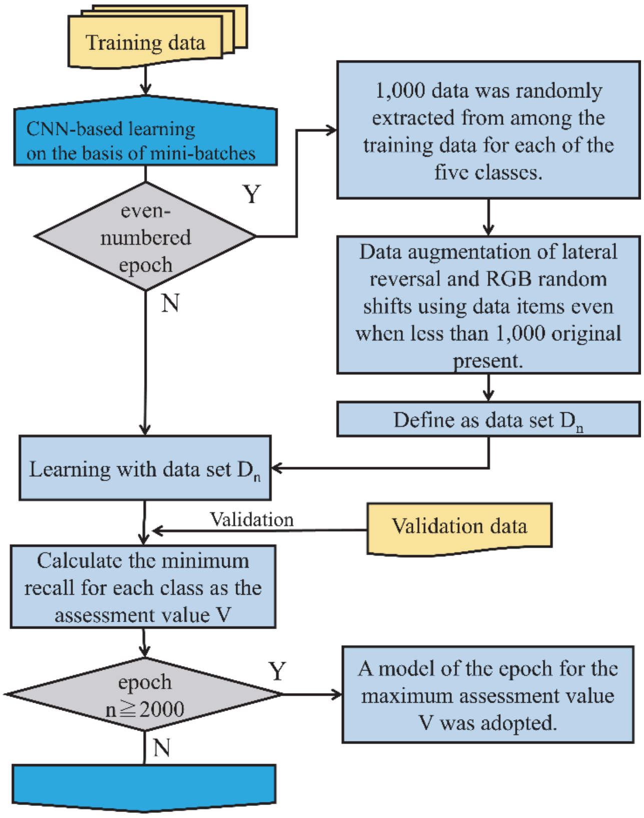


Fig. 13 Training model flow chart[28)]

Table 13 CNN parameters

Parameter	Set value
Activation function	ReLU
Mini-batch size	100
Method of optimization	SGD
Learning coefficient	10^{-3} – 10^{-10}
Weight attenuation	0
Number of attempts	500 – 2000 epoch

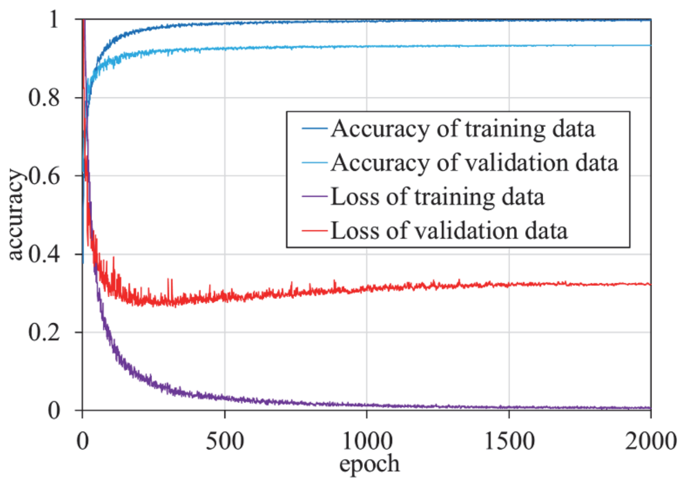


Fig. 14 Rating accuracy and loss function for each epoch

4. VERIFICATION OF THE ACCURACY OF DAMAGE DETECTION MODELS

4.1 Generalization performance issues in models reported in previous studies

As noted earlier, some issues regarding the generalization performance of the model have been reported in the literature[22)]. In this study, a total of 42,900 training data items obtained after the main shock of the 2016 Kumamoto Earthquake (Mashiki Town) (Table 10) were learned using the model described in the literature[22]. Thereafter, damage detection results were determined for all areas from the various aerial photographs taken following the main shock of the 2016 Kumamoto Earthquake, the 1995 Southern Hyogo Prefecture Earthquake, and the 2011 off the Pacific Coast of Tohoku Earthquake (at the locations shown in Fig. 4, Fig. 1, and Fig. 2, respectively) that were used for testing. The results are shown in Fig. 15. Each image was examined using a program in which damage detection results for an 80-pixel square patch were determined by scanning the entire area of the aerial photograph. Details of the procedures of image preparation used have already been published elsewhere in the literature[22), 28)] and have, therefore, been omitted here.

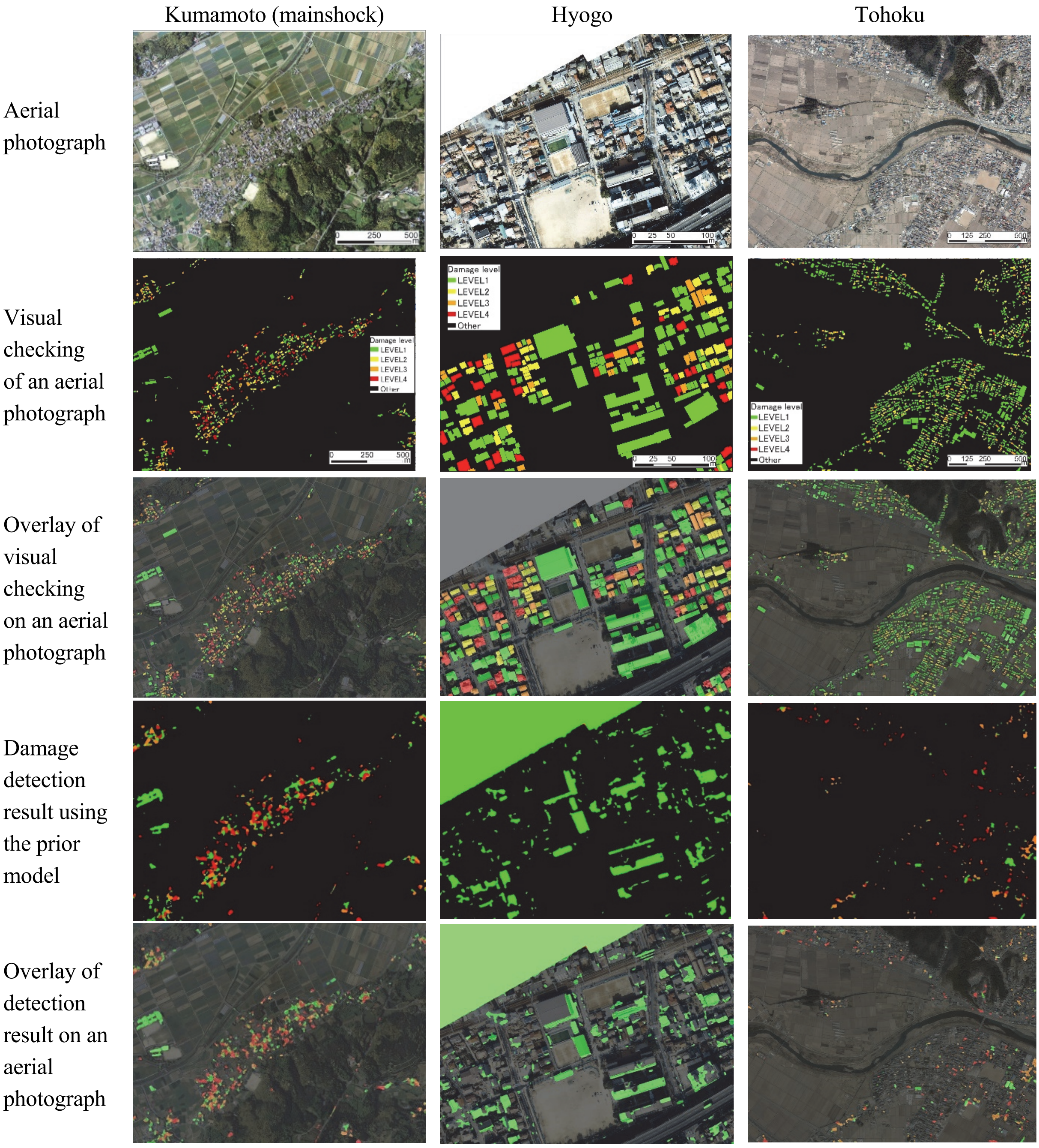


Fig. 15 Detection results from unlearned aerial photographs using a model reported in a prior study[22)]

The first row in Fig. 15 shows the results from the main shock of the 2016 Kumamoto Earthquake (Mashiki Town); the second row shows the 1995 Southern Hyogo Prefecture Earthquake; and the third row shows the 2011 off the Pacific Coast of Tohoku Earthquake. The first line shows one of the aerial photographs used with a scale added. The second line shows an image colored for various building polygons on the basis of the results of visual checking of aerial photographs, with a scale and legend added. The third line shows an overlay of a visual checking image on an aerial photograph at a transmittance of 50%. The fourth line shows a damage detection image in a prior-study model of learning data for the main shock of the 2016 Kumamoto Earthquake only. The fifth line shows an overlay of an image of damage detection results on an aerial photograph at a transmittance of 50%. When comparing the visual check results on the second and third lines and the damage detection results on the fourth and fifth lines in each earthquake, it can be seen that the images from the 2016 Kumamoto Earthquake, which were of the same kind as the learning data, generally performed well in extracting the distributions and total volumes of LEVEL 1 and LEVEL 4 buildings. However, in the images from the 1995 Southern Hyogo Prefecture Earthquake, the upper left zone, which was painted white at the time of digitization since it was outside the range of aerial photographs, was misidentified as LEVEL 1 and many damaged buildings were misidentified as non-buildings meaning that, overall, the levels of damage seem to have been underrated. For the 2011 off the Pacific Coast of Tohoku Earthquake, many buildings were misidentified as non-buildings and LEVEL 4 was over-identified.

Figure 16 shows plots of detection accuracy (recall, precision, F-measures) for testing data in various different cases (Table 10). When assessing the main shock of the 2016 Kumamoto Earthquake, the average recall for the various levels of damage was approximately 60.2%, the average precision was approximately 49.1%, and the average F-measure was approximately 52.6%. In the 1995 Southern Hyogo Prefecture Earthquake, the average recall was approximately 30.9%, the average precision was approximately 46.1%, and the average F-measure was approximately 33.2%. In the 2011 off the Pacific Coast of Tohoku Earthquake, the average recall was approximately 15.1%, the average precision was approximately 43.2%, and the average F-measure was approximately 14.4%. All these values were low. The above findings show that the prior-study model is unsatisfactory in terms of generalization performance for images taken under different imaging conditions.

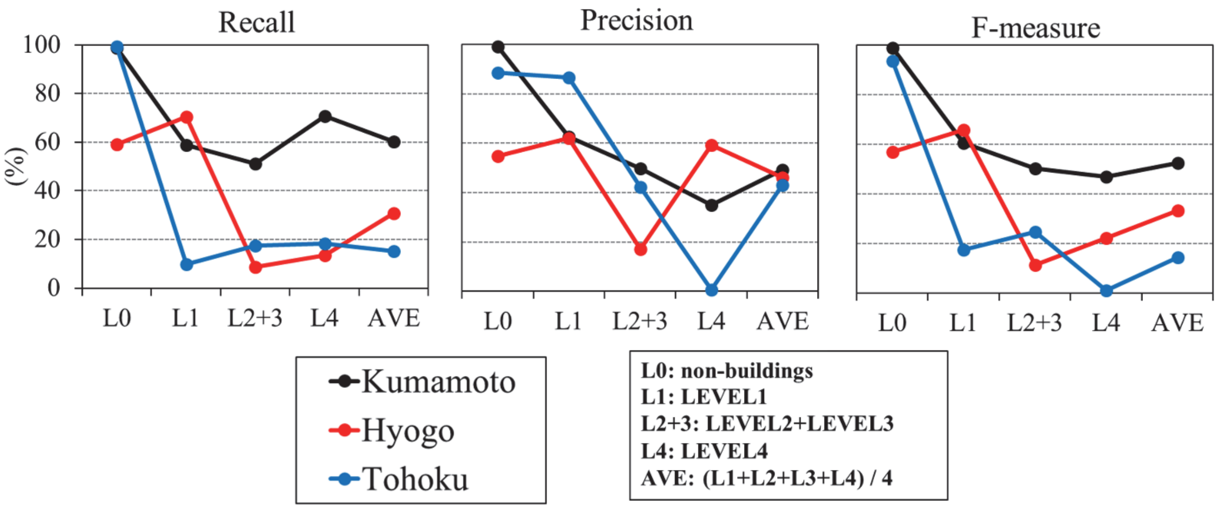


Fig. 16 Recall, precision, and F-measure values for test data with learning using a model described in a prior study[22)]

4.2 Comparison of damage detection accuracies between the present model and a prior-study model

The failure to improve the generalization performance of the prior-study model can be attributed to two factors: the training model and the training data. To determine the influence of model improvement on its generalization performance, training was performed for shared learning data using models from the present study and those reported in the literature[22], after which the detection accuracy of each model was verified by automatically assessing damage using shared testing data. A total of 239,654 patch images for training, extracted from six kinds of aerial photograph, using the method described in Chapter 2, were used as the training data (Table 10).

Results of damage detection using each model for the entire area of the aerial photographs tested are shown in Fig. 17. The first row in Fig. 17 shows the results from the main shock of the 2016 Kumamoto Earthquake (Mashiki Town); the second row shows the 1995 Southern Hyogo Prefecture Earthquake; and the third row shows the 2011 off the Pacific Coast of Tohoku Earthquake. The first line shows one of the aerial photographs used with a scale added. The second line shows an image colored for various building polygons based on the results of visual checking of the aerial photographs. The third line shows a damage detection image obtained by learning using one of the models reported in the literature[22]. The fourth line shows an overlay of a damage detection image obtained using a prior-study model on an aerial photograph at a transmittance of 50%. The fifth line shows a damage detection image obtained by learning using the present model. The sixth line shows an overlay of a damage detection result from the present model and an aerial photograph at a transmittance of 50%.

When comparing damage detection results in the prior-study model shown in Fig. 17 (third and fourth lines) with visual check results (second line), it can be seen that the prior-study model rated more buildings as LEVEL 4 than did visual checking of the 2016 Kumamoto Earthquake. In the 1995 Southern Hyogo Prefecture Earthquake, the outlined portion outside the scope of aerial photographs was rated as LEVEL 1, and identification of damage at LEVEL 2 or higher failed. Furthermore, in the 2011 off the Pacific Coast of Tohoku Earthquake, non-buildings were rated as LEVEL 4 in some portions—all of which were found to have been erroneously identified. In the damage detection results obtained using the present model (lines 5 and 6), on the other hand, the detection results for all aerial photographs were in general agreement with the overall degrees of damage and areas with concentrated damage identified in the visual check results. Therefore, the damage detection accuracy for testing aerial photographs was successfully improved, compared with the prior-study model.

As described above, the prior-study model often underrated or overrated damage levels for different test images. This is attributable to the major impact of brilliancy changes due to variation in the imaging conditions for different images. In a previous study[23], image histogram normalization produced a range of improvements. On the other hand, the present model also involved a random shift of brilliancy values in learning data expansion, as described in Chapter 3, and efforts were made to make the model more robust to brilliancy changes even without histogram normalization.

Table 14 shows the results of damage detection for 170,164 testing patch images taken from six kinds of aerial photograph (Table 10) using the present model. Table 15 shows the results of damage detection using a model reported in the literature[22]. Both sets of data are displayed in the form of a confusion matrix, along with the performance indicators shown in Table 5. These results were compared on the basis of the average values of the various indicators used for LEVEL 1 to LEVEL 4. In the model developed in the present study (CASE 1), the average recall was approximately 75.0%, the average precision was approximately 54.7%, and the average F-measure was approximately 62.8%, whereas in the model used for the prior study (CASE 2), the average recall was approximately 62.8%, the average precision was approximately 30.7%, and the average F-measure was approximately 35.8%. When using any of the indicators, the model developed using the method described in the present study was found to be superior in terms of damage detection accuracy (Fig. 18).

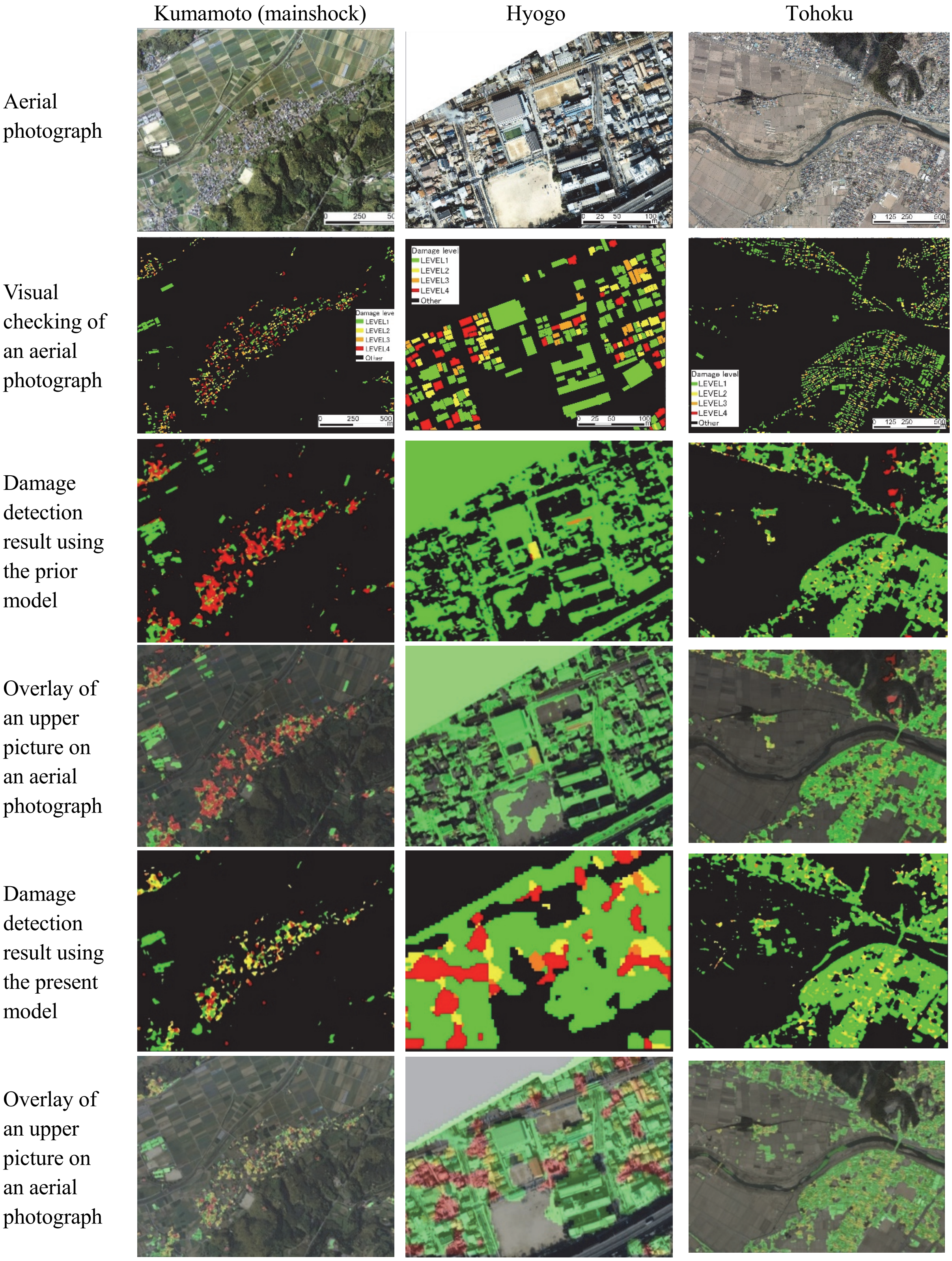


Fig. 17 Damage detection results from unlearned aerial photographs using the present model and a model from a prior study[22)]

Table 14 Ratings in the model developed using the method described in the present study (CASE 1)

		Automatic classification				Total	Recall (%)	F-measure (%)
		Non-buildings	LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	Non-buildings	143,942	4,305	792	286	149,325	96.4	97.9
	LEVEL 1	686	13,649	2,049	57	16,441	83.0	78.4
	LEVEL 2+3	37	407	3,253	166	3,863	84.2	64.3
	LEVEL 4	31	33	162	309	535	57.8	45.7
Total		144,696	18,394	6,256	818	170,164	-	-
Precision (%)		99.5	74.2	52.0	37.8	-	Accuracy (%)	94.7

Table 15 Ratings in a model developed using a method described in the literature[22] (CASE 2)

		Automatic classification				Total	Recall (%)	F-measure (%)
		Non-buildings	LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	Non-buildings	137,414	5,749	2,829	3,333	149,325	92.0	95.5
	LEVEL 1	1,028	10,751	3,190	1472	16,441	65.4	63.6
	LEVEL 2+3	98	832	1,921	1012	3,863	49.7	32.3
	LEVEL 4	20	47	76	392	535	73.3	11.6
Total		138,560	17,379	8,016	6,209	170,164	-	-
Precision (%)		99.2	61.9	24.0	6.3	-	Accuracy (%)	88.4

In all these results, the precision decreased for LEVEL 2+3 and LEVEL 4 compared with non-buildings and LEVEL 1. This could be because the numbers of testing data for non-buildings and LEVEL 1 were larger than those for LEVEL 2+3 and LEVEL 4, and the proportion of detection errors might have been higher in the former, even when relatively low overall. Likewise, the absolute value of precision for LEVEL 2+3 and LEVEL 4 often decreased because of the use of different amounts of testing data in each class. The accuracy was affected to a greater extent by non-buildings and LEVEL 1, which contained larger amounts of data. For these reasons, the present study emphasized recall as a performance indicator. However, precision and F-measure values were also used to help gain an understanding of the relative differences in detection performance among the different models' training data.

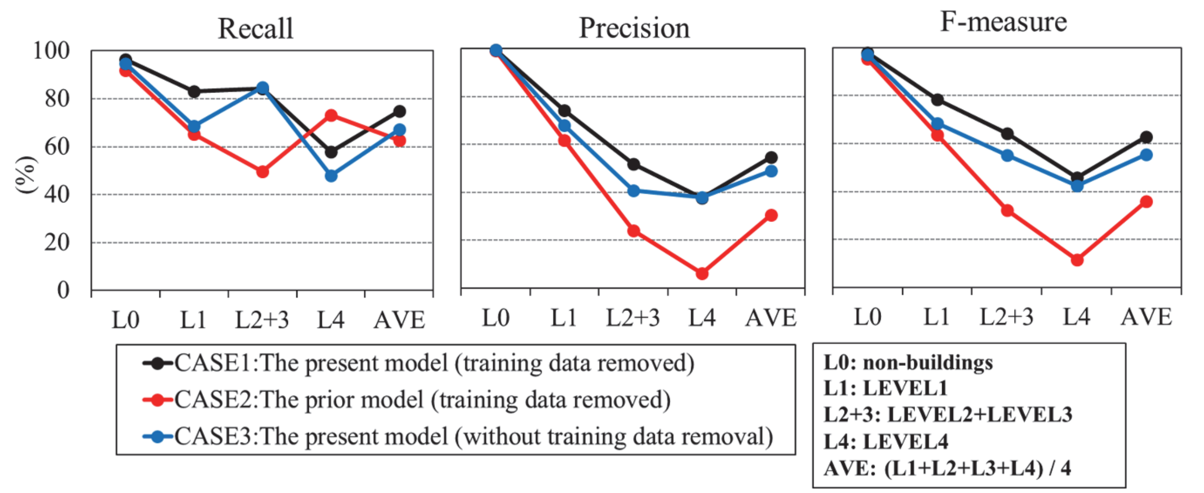


Fig. 18 Variation in detection accuracy among the various classes used for different models and different conditions of training data preparation

4.3 Influence of visual removal of training data on detection accuracy

Here, the influence of the training data used on the detection performance is discussed. First, the influence of training data screening on the detection accuracy is described.

Section 2.3 described how data were visually removed when preparing the training data. To determine the influence of this step on the detection accuracy, training was performed using training data prior to data removal (Table 9) and training data following data removal (Table 10), and then levels of damage were assessed using the resulting models on the same testing data.

Damage detection results for all areas of the aerial photographs used for testing are shown in Fig. 19. The first row in Fig. 19 shows the results obtained after the main shock of the 2016 Kumamoto Earthquake (Mashiki Town); the second row shows the 1995 Southern Hyogo Prefecture Earthquake; and the third row shows the 2011 off the Pacific Coast of Tohoku Earthquake. The first line shows one of the aerial photographs used with a scale added. The second line shows an image of visual check results. The third line shows an image of detection results with learning of data prior to data removal. The fourth line shows an overlay of this image on an aerial photograph at a transmittance of 50%. The fifth line shows an image of detection results with data training following data removal. The sixth line shows an overlay of this image on an aerial photograph at a transmittance of 50%. It can be seen that a comparison of the images on the third and fifth lines shows no major difference in the degree of damage identified before and after data removal.

On the other hand, 176,375 items of training data derived from six kinds of aerial photograph, prior to data removal, were used as test data. The resulting confusion matrix for detection results is shown in Table 16. When comparing these results and the detection results with training data following data removal (Table 9), based on the mean values for each indicator from LEVEL 1 to LEVEL 4, the average recall was approximately 67.1%, the average precision was approximately 49.0%, and the average F-measure value was approximately 55.3% for the detection results prior to data removal (CASE 3). When using any of the indicators, the detection accuracy decreased slightly compared with the detection results following data removal (CASE 1) (Fig. 18). This finding shows that visual screening of training data is effective, at least to some extent, in building a highly generalization performance model.

Table 16 Ratings obtained in a model without training data removal (CASE 3)

		Automatic classification				Total	Recall (%)	F-measure (%)
		Non-buildings	LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	Non-buildings	143,495	5,172	2,301	388	151,356	94.8	97.1
	LEVEL 1	692	12,235	4,706	170	17,803	68.7	68.4
	LEVEL 2+3	100	502	5,159	324	6,085	84.8	55.2
	LEVEL 4	66	81	442	542	1,131	47.9	42.4
Total		144,353	17,990	12,608	1,424	176,375	-	-
Precision (%)		99.4	68.0	40.9	38.1	-	Accuracy (%)	91.5

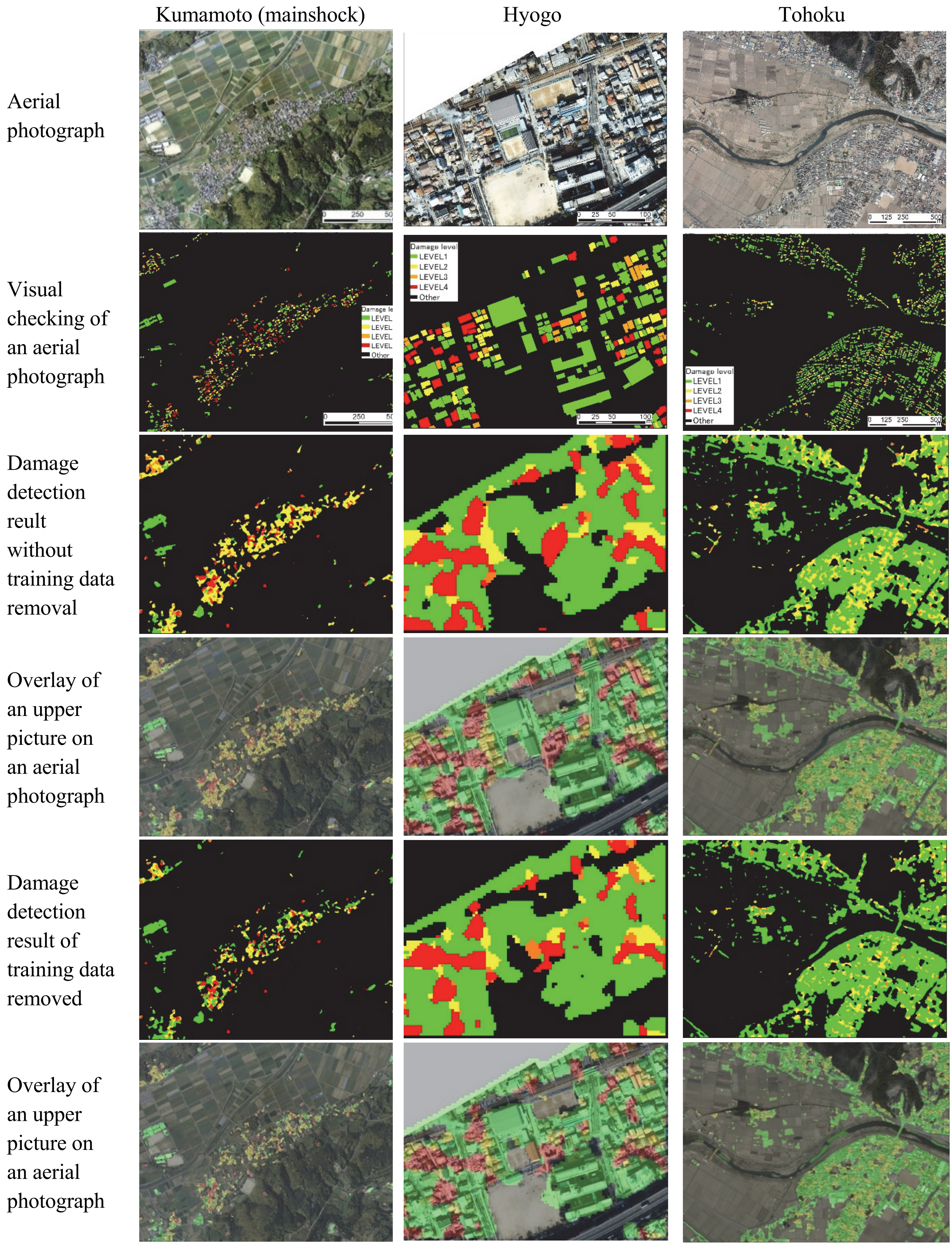


Fig. 19 Damage detection result differences derived from unlearned aerial photographs in the presence and absence of visual removal

4.4 Influence of multiplicity of training data on detection accuracy

Next, the influence of the multiplicity of training data types on the detection accuracy is described—comparing simultaneous learning of multiple training datasets acquired under different imaging conditions. Figure 20 shows detection results obtained by acquiring training data from one type, two types, three types, four types, five types, and six types of aerial photograph used as training data following visual removal of data (Table 10) (denoted DATA 1, DATA 2, DATA 3, DATA 4, DATA 5, and DATA 6, respectively). This was used to build models, and for assessing damage when testing aerial photographs using the models. The first row of Fig. 20 shows the results obtained after the main shock of the 2016 Kumamoto Earthquake (Mashiki Town). The second row shows the results for the 1995 Southern Hyogo Prefecture Earthquake. The third row shows the results for the 2011 off the Pacific Coast of Tohoku Earthquake. The first line shows an overlay of an image of detection results with DATA 1 learning on an aerial photograph at a transmittance of 50%. The second line shows an overlay of an image of detection results with DATA 2 learning on an aerial photograph at a transmittance of 50%. The subsequent lines show an overlay of an image of detection results with DATA 3, DATA 4, DATA 5, or DATA 6 learning and an aerial photograph at a transmittance of 50%.

When comparing the various results with the visual check results shown in Fig. 19, it can be seen that the detection accuracy was low with a few kinds of training data. For example, damage levels were underrated for DATA 1 in the 2016 Kumamoto Earthquake and the 1995 Southern Hyogo Prefecture Earthquake, but were overrated in the 2011 off the Pacific Coast of Tohoku Earthquake. Overall, however, the degree of consistency with the visual check results increased with increasing types of data, so that the best results were obtained when using DATA 6.

Next, the detection accuracies for 170,164 items of testing data acquired from the six kinds of aerial photograph were compared using various different models. The results are shown in Fig. 21. The average recall at each damage level was higher for DATA 3, 4, 5 and 6 than for DATA 1 and 2. When comparing the various confusion matrices for each case, the rate of visual misidentification of LEVEL 2+3 buildings as LEVEL 1 buildings was higher with the use of DATA 1 and 2 (Tables 17 and 18), and the recall for LEVEL 2+3 was lower. However, as the number of types of data increased, these classification errors decreased and the average recall value increased. With regard to precision, the average precision value decreased because the automatic classification results for LEVEL 2+3 using DATA 3, 4 and 5 included many visual check results for non-buildings and LEVEL 1 (Tables 19 to 21). The proportions of these classification errors decreased with DATA 6, and this also produced the highest results for average precision at each damage level (Table 22). As a result, the average recall, average precision, and average F-measure for each damage level were all maximized with DATA 6.

Therefore, when considering damage classification for testing data collected under various different conditions, classification errors, etc., can sometimes occur depending on the combination of data used when there are fewer kinds of training data than test data. However, it can be seen that both the average recall and the average precision improved in the model when a sufficient number of different types of training data suitable for the test data were available.

While the number of datasets used for learning increased from DATA 1 to DATA 6, the number of data items used for epoch-by-epoch learning was uniformly set at 1000 for each class, as stated in Chapter 3. Therefore, the improved accuracy can be attributed to the abundance of different kinds of training data, rather than to the increase in the amount of training data.

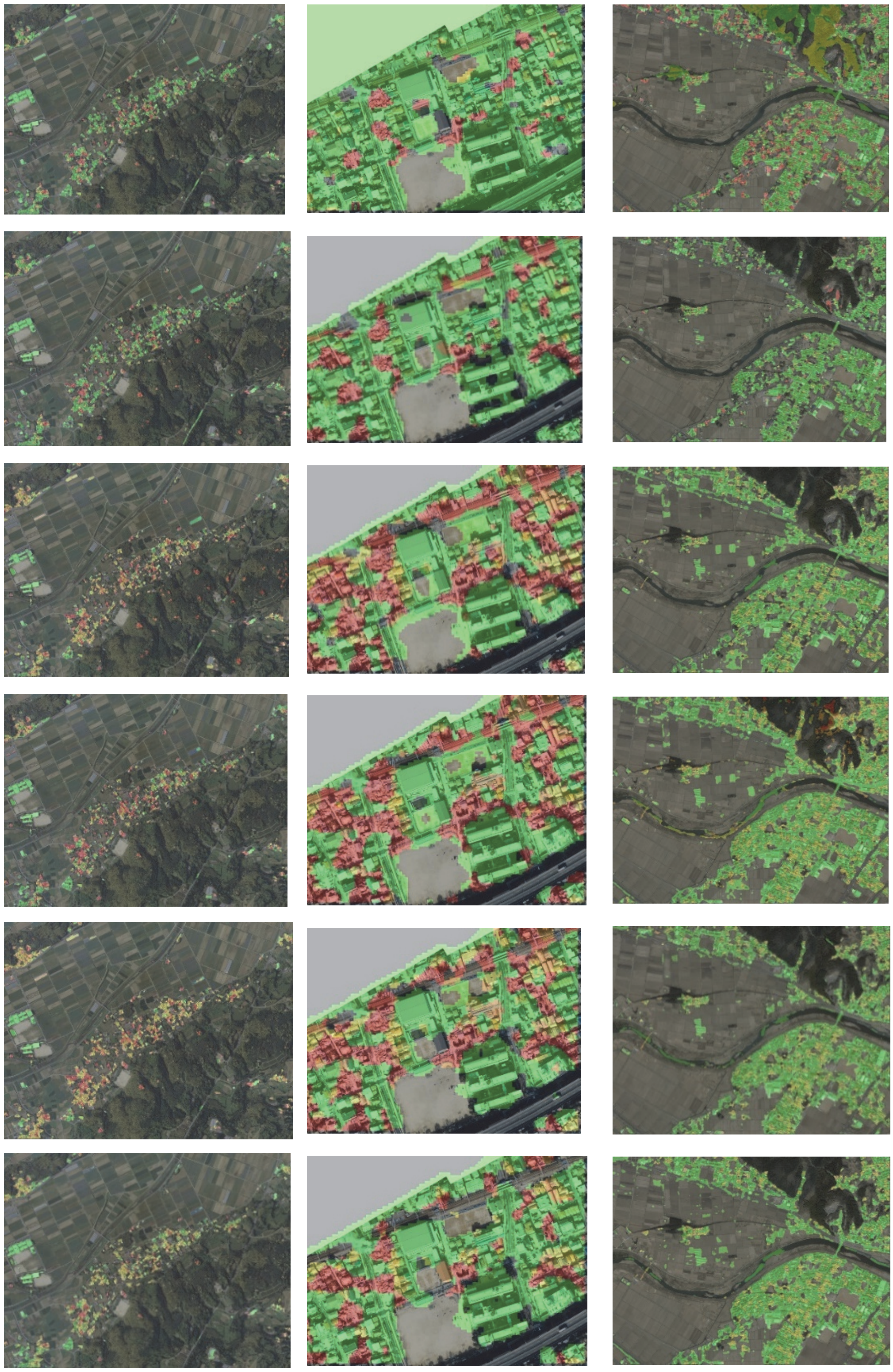


Fig. 20 Superposition of damage detection results and aerial photographs with different kinds of training data (The color classes of the various detection results are the same as those shown in Fig. 19)

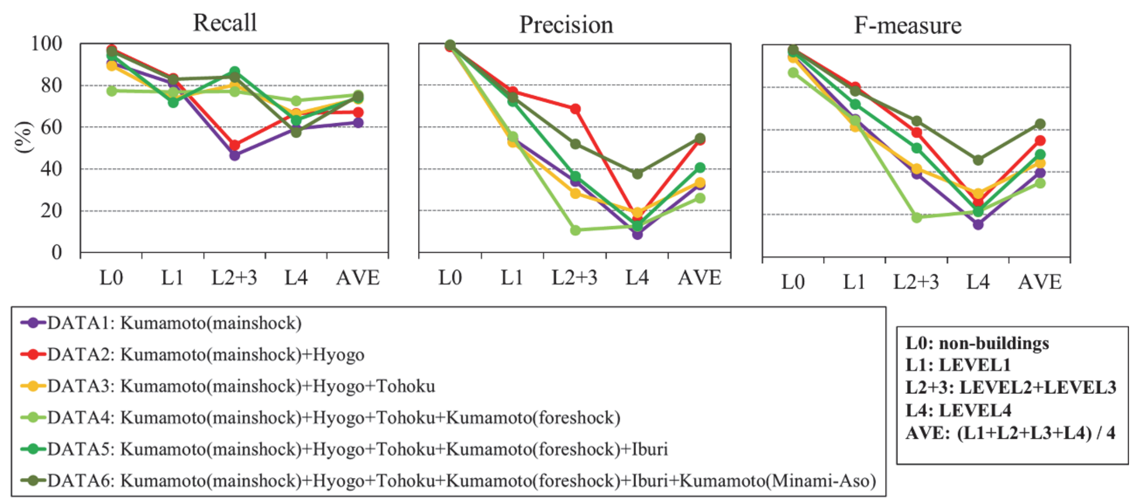


Fig. 21 Damage detection accuracy differences, by class, with different kinds of training data

Table 17 Damage detection results for test data in the model trained by DATA 1

		Automatic classification				Total	Recall (%)	F-measure (%)
		Non-buildings	LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	Non-buildings	135,289	9,611	2,967	1,458	149,325	90.6	94.6
	LEVEL 1	1,384	13,348	434	1275	16,441	81.2	65.2
	LEVEL 2+3	75	1437	1,798	553	3,863	46.5	39.4
	LEVEL 4	38	104	76	317	535	59.3	15.3
Total		136,786	24,500	5,275	3,603	170,164	-	-
Precision (%)		98.9	54.5	34.1	8.8	-	Accuracy (%)	88.6

Table 18 Damage detection results for test data in the model trained by DATA 2

		Automatic classification				Total	Recall (%)	F-measure (%)
		Non-buildings	LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	Non-buildings	145,170	2,667	246	1,242	149,325	97.2	97.9
	LEVEL 1	1,812	13,747	574	308	16,441	83.6	80.2
	LEVEL 2+3	186	1,367	1,989	321	3,863	51.5	58.9
	LEVEL 4	35	65	78	357	535	66.7	25.8
Total		147,203	17,846	2,887	2,228	170,164	-	-
Precision (%)		98.6	77.0	68.9	16.0	-	Accuracy (%)	94.8

Table 19 Damage detection results for test data in the model trained by DATA 3

		Automatic classification				Total	Recall (%)	F-measure (%)
		Non-buildings	LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	Non-buildings	133,661	10,495	4,345	824	149,325	89.5	94.2
	LEVEL 1	643	12,144	3,399	255	16,441	73.9	61.6
	LEVEL 2+3	61	299	3,101	402	3,863	80.3	41.8
	LEVEL 4	35	25	119	356	535	66.5	30.0
Total		134,400	22,963	10,964	1,837	170,164	-	-
Precision (%)		99.5	52.9	28.3	19.4	-	Accuracy (%)	87.7

Table 20 Damage detection results for test data in the model trained by DATA 4

		Automatic classification				Total	Recall (%)	F-measure (%)
		Non-buildings	LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	Non-buildings	115,645	9,671	21,962	2,047	149,325	77.4	87.1
	LEVEL 1	525	12,635	3,051	230	16,441	76.9	64.5
	LEVEL 2+3	43	394	2,987	439	3,863	77.3	18.7
	LEVEL 4	15	35	95	390	535	72.9	21.4
Total		116,228	22,735	28,095	3,106	170,164	-	-
Precision (%)		99.5	55.6	10.6	12.6	-	Accuracy (%)	77.4

Table 21 Damage detection results for test data in the model trained by DATA 5

		Automatic classification				Total	Recall (%)	F-measure (%)
		Non-buildings	LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	Non-buildings	141,185	4,248	2,025	1,867	149,325	94.5	96.9
	LEVEL 1	863	11,820	3,607	151	16,441	71.9	72.2
	LEVEL 2+3	41	229	3,354	239	3,863	86.8	51.6
	LEVEL 4	23	19	153	340	535	63.6	21.7
Total		142,112	16,316	9,139	2,597	170,164	-	-
Precision (%)		99.3	72.4	36.7	13.1	-	Accuracy (%)	92.1

Table 22 Damage detection results for test data in the model trained by DATA 6

		Automatic classification				Total	Recall (%)	F-measure (%)
		Non-buildings	LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	Non-buildings	143,942	4,305	792	286	149,325	96.4	97.9
	LEVEL 1	686	13,649	2,049	57	16,441	83.0	78.4
	LEVEL 2+3	37	407	3,253	166	3,863	84.2	64.3
	LEVEL 4	31	33	162	309	535	57.8	45.7
Total		144,696	18,394	6,256	818	170,164	-	-
Precision (%)		99.5	74.2	52.0	37.8	-	Accuracy (%)	94.7

4.5 Influence of training data space distribution on damage detection accuracy

In the previous sections, accuracy was checked using one optional choice from the six kinds of aerial photograph used as training data, along with other choices from those used as testing data. The possibility of detection accuracy differences due to spatial changes in the combination of these training data cannot be ruled out. Therefore, a cross validation was performed by exchanging space distribution of training data and testing data based on the aerial photograph frames.

The cross validation employed a total of 31 frames: 12 frames of aerial photographs with patch images identified by method A (Table 8), and 19 images containing five or more training datasets for LEVEL 1 to LEVEL 4 obtained from aerial photographs with patch images identified by method B (right panels of Figs. 1–6). In these aerial photographs, damage was detected using a 31-hold cross validation process in which building patch images from one frame of an aerial photograph were used as testing data, and building patch images from 30 other aerial photograph frames were used as training data. Overall, a total of 67,893 data items were used, including exchanges of combinations of testing data and training data. The overall confusion matrix for these detection results is shown in Table 23. The overall average recall was approximately 77.2%, and the average precision was approximately 63.8%. These results show that a reasonable level of detection accuracy was obtained.

Confusion matrices for six kinds of aerial photograph are shown in Tables 24 to 29. Figure 22 shows the performance indicators used for these detection results and plots the mean values of all detection results. The recall exceeded 60% for every damage level. However, the recall for LEVEL 4 in the 2011 off the Pacific Coast of Tohoku Earthquake was relatively low at approximately 22.2%. This can be attributed to the fact that there were relatively few test data, and that much of the damage was indistinct, as described in Section 4.7 below. The precision was low at less than 5% for LEVEL 4 in the 2011 off the Pacific Coast of Tohoku Earthquake, and for LEVEL 2+3 and LEVEL 4 in the 2018 Hokkaido Iburi Eastern Earthquake. This can be attributed to the fact that there were only 18 to 22 items of test data available (Tables 26 and 29), which was several orders of magnitude less than the data for other damage levels, and that the precision also decreased as the influence of classification errors in LEVEL 1, which included a lot of data, increased compared with the former data.

The recalls, precisions, and F-measures calculated for the various levels of damage in the other earthquakes generally fell within the overall range of the mean value for all earthquakes ±1σ (standard deviation). The above findings show that various kinds of aerial photograph can be used to build up a damage detection model with an average recall for the various levels of damage of approximately 60% to 70%, and that the space distribution of training data does not significantly influence the detection accuracy.

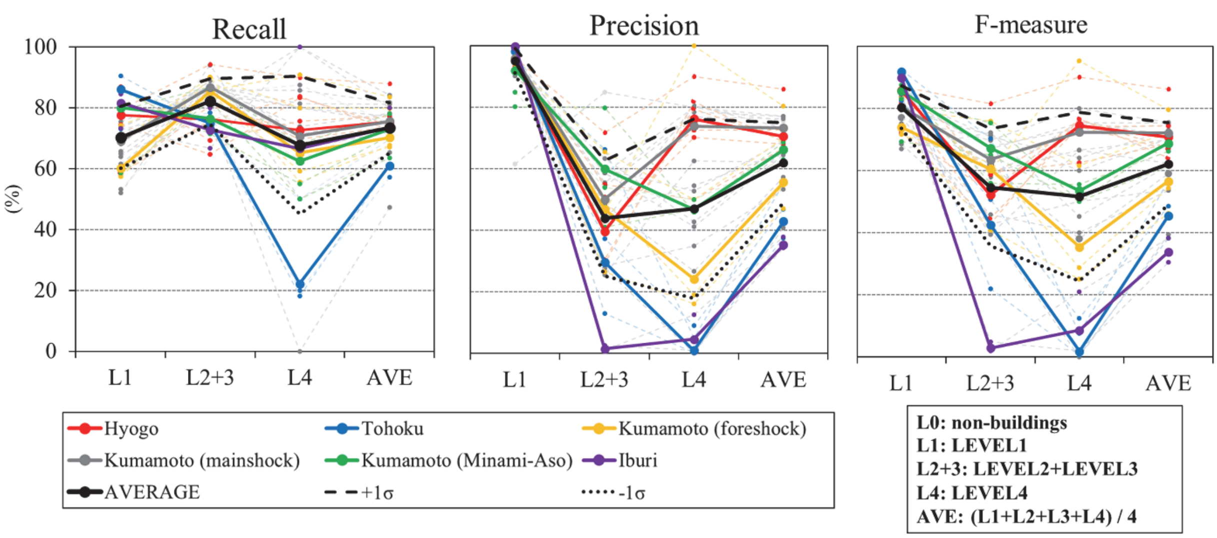


Fig. 22 Comparison of damage detection accuracies for various earthquakes in cross validation of aerial photograph frames

Table 23 Overall ratings in cross validation of aerial photograph frames

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	42,695	11,253	1,016	54,964	77.7	86.5
	LEVEL 2+3	1,042	8,663	659	10,364	83.6	55.9
	LEVEL 4	57	707	1,801	2,565	70.2	59.6
Total		43,794	20,623	3,476	67,893	-	-
Precision (%)		97.5	42.0	51.8	-	Accuracy (%)	78.3

Table 24 Damage detection results in cross validation of aerial photograph frames for the 2016 Kumamoto Earthquake (after main shock, Mashiki Town)

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	8,699	3,738	137	12,574	69.2	80.4
	LEVEL 2+3	350	4,236	293	4,879	86.8	63.5
	LEVEL 4	17	485	1,217	1,719	70.8	72.3
Total		9,066	8,459	1,647	19,172	-	-
Precision (%)		96.0	50.1	73.9	-	Accuracy (%)	73.8

Table 25 Damage detection results in cross validation of aerial photograph frames for the 1995 Southern Hyogo Prefecture Earthquake

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	2,078	535	64	2,677	77.6	85.7
	LEVEL 2+3	76	438	61	575	76.2	52.1
	LEVEL 4	18	132	401	551	72.8	74.5
Total		2,172	1,105	526	3,803	-	-
Precision (%)		95.7	39.6	76.2	-	Accuracy (%)	76.7

Table 26 Damage detection results in cross validation of aerial photograph frames for the 2011 off the Pacific Coast of Tohoku Earthquake

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	20,927	3,019	361	24,307	86.1	91.7
	LEVEL 2+3	385	1,281	44	1,710	74.9	42.6
	LEVEL 4	4	10	4	18	22.2	1.9
Total		21,316	4,310	409	26,035	-	-
Precision (%)		98.2	29.7	1.0	-	Accuracy (%)	85.3

Table 27 Damage detection results in cross validation of aerial photograph frames for the 2016 Kumamoto Earthquake (after foreshock, Mashiki Town)

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	4,391	2,780	134	7,305	60.1	73.9
	LEVEL 2+3	186	2,500	241	2,927	85.4	60.5
	LEVEL 4	3	61	120	184	65.2	35.3
Total		4,580	5,341	495	10,416	-	-
Precision (%)		95.9	46.8	24.2	-	Accuracy (%)	67.3

Table 28 Damage detection results in cross validation of aerial photograph frames for the 2016 Kumamoto Earthquake (after main shock, Minami-Aso Village)

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	573	111	32	716	80.0	85.6
	LEVEL 2+3	40	192	19	251	76.5	67.3
	LEVEL 4	10	17	45	72	62.5	53.6
Total		623	320	96	1,039	-	-
Precision (%)		92.0	60.0	46.9	-	Accuracy (%)	78.0

Table 29 Damage detection results in cross validation of aerial photograph frames for the 2018 Hokkaido Iburi Eastern Earthquake

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	6,027	1,070	288	7,385	81.6	89.8
	LEVEL 2+3	5	16	1	22	72.7	2.9
	LEVEL 4	5	2	14	21	66.7	8.6
Total		6,037	1,088	303	7,428	-	-
Precision (%)		99.8	1.5	4.6	-	Accuracy (%)	81.5

4.6 Damage detection results for buildings obtained using the present model

Building damage detection results obtained using the present model and applied to all unlearned areas of the aerial photographs used are described below. In the present model, damage is assessed for each 80-pixel square patch image, which corresponds roughly to the size needed to cover one building. In addition, we have developed a program for scanning the entire image in order to repeat damage detection for each patch, to automatically determine damage detection results for all areas of aerial photograph images, and to output colored images that clearly distinguish different damage levels. The program may also be used to superpose images with building polygons in order to automatically determine damage levels for individual buildings using the same procedures as those reported in the literature[22), 28)]. Levels of damage were rated on an area basis, using a threshold value based on the proportion of the area of each patch showing damage in the building polygon.

This area threshold was set as follows. If any patch showing a high level of damage was contained within a building polygon, it was considered a damaged portion and that building was identified as severely damaged, with priority given to identifying the higher levels of damage, overall. Based on these criteria, the area threshold for each damage level was set between 1% and 80%, and then used to maximize the average recall for the detection results with the six kinds of aerial photograph used for testing. Specifically, when LEVEL 4 patches were present in 1% or more of the building polygon, the rating given was LEVEL 4. When LEVEL 3 patches were present in 10% or more, the rating was LEVEL 3. When LEVEL 2 patches were present in 40% or more, the rating was LEVEL 2. In all other cases, the rating was set at LEVEL 1.

In this study, training data from six kinds of aerial photograph (Table 10) were used for training with the present model, after which this approach was also applied to the six aerial photographs used for testing (Figs. 1–6). In the case of the 1995 Southern Hyogo Prefecture Earthquake, we used building polygons that we had prepared ourselves during the preparation of training data based on aerial photographs. In other cases, building polygons taken from the Japan Basic Map Information database[24)] were used.

Figures. 23 and 24 show detection results for the entire image and automatic damage detection results for buildings, obtained from the training data derived from six kinds of aerial photograph (Table 10) using the present model, and then applied to the six aerial photographs used for testing (Figs. 1–6). The first row in Fig. 23 shows the results for the main shock of the 2016 Kumamoto Earthquake (Mashiki Town); the second row shows the results for the 1995 Southern Hyogo Prefecture Earthquake; and the third row shows the results for the 2011 off the Pacific Coast of Tohoku Earthquake. The first line shows one of the aerial photographs used with a scale added. The second line shows an image with the various building polygons distinguished by color on the basis of the results of a visual check of the aerial photographs. The third line shows damage detection results for the entire image area. The fourth line shows an overlay of damage detection results for the entire image area on an aerial photograph at a transmittance of 50%. The fifth line shows damage detection results for buildings based on an overlay on a building polygon. The sixth line shows an overlay of damage detection results for buildings on an aerial photograph at a transmittance of 50%. In Fig. 24, the first row shows damage detection results obtained after the foreshock of the 2016 Kumamoto Earthquake (Mashiki Town); the second row shows results obtained after the main shock of the 2016 Kumamoto Earthquake (Minami-Aso Village); and the third row shows the results obtained for the 2018 Hokkaido Iburi Eastern Earthquake. The order in which the images have been arranged in line is the same as in Fig. 23.

When comparing damage detection results for buildings in Figs. 23 and 24 (fifth line) with visual check results (second line), it can be seen in terms of qualitative nature that the overall damage levels and the distributions of severely affected areas derived from each aerial photograph could be easily understood and did not differ markedly from the visual check results.

Next, damage detection results for buildings in each aerial photograph used for testing were compared with visual check results. The resulting overall confusion matrix is shown in Table 30. It can be seen that the overall average recall was approximately 70.2%, and average precision was approximately 64.3%. These results represent a reasonable level of detection accuracy.

Figure. 25 shows recall, precision and F-measure results for various aerial photographs used for testing, and the mean and standard deviation data for various rating indicators used for the six earthquakes examined. Recall results for LEVEL 2+3 in the 2018 Hokkaido Iburi Eastern Earthquake and LEVEL 4 in the 2011 off the Pacific Coast of Tohoku Earthquake were each under one standard deviation of -1σ. In terms of precision, LEVEL 2+3 in the 2018 Hokkaido Iburi Eastern Earthquake and LEVEL 4 in the 2011 off the Pacific Coast of Tohoku Earthquake were each under one standard deviation of -1σ. The possible causes of these reductions in damage detection accuracy are discussed further in Section 4.7.

Regarding processing time, the learning stage for building each damage detection model took approximately 9 hours. However, when using a fully developed model, all the images shown in Figs. 23 and 24 could be finalized within 10 minutes for each individual aerial photograph. In contrast, performing visual checks could take several days, depending on the number of buildings in the image[25]. Therefore, in terms of processing time, this model can be considered effective in the simultaneous detection of damage states.

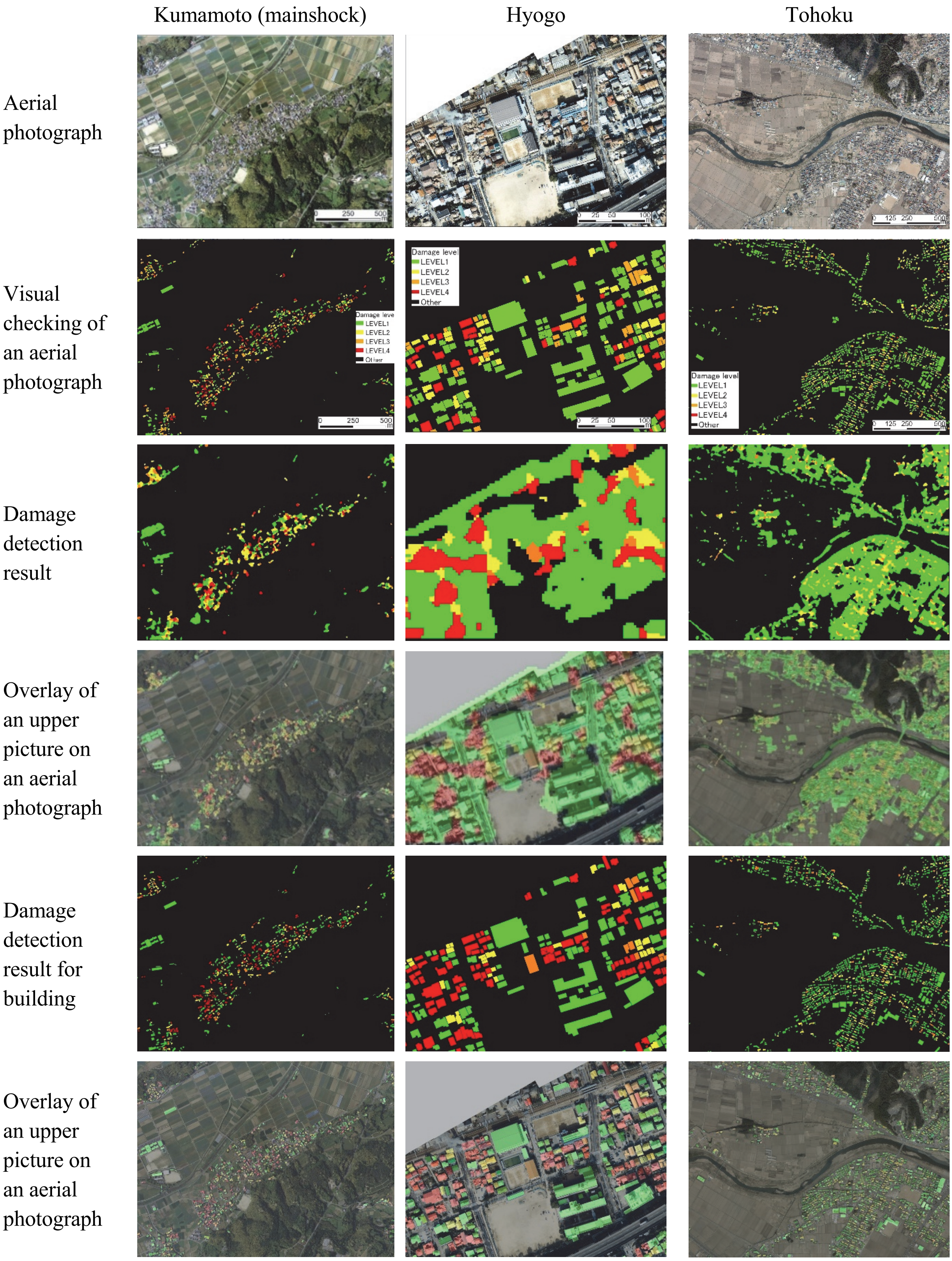


Fig. 23 Automatic damage detection results (1) obtained by applying the present model to unlearned aerial photographs

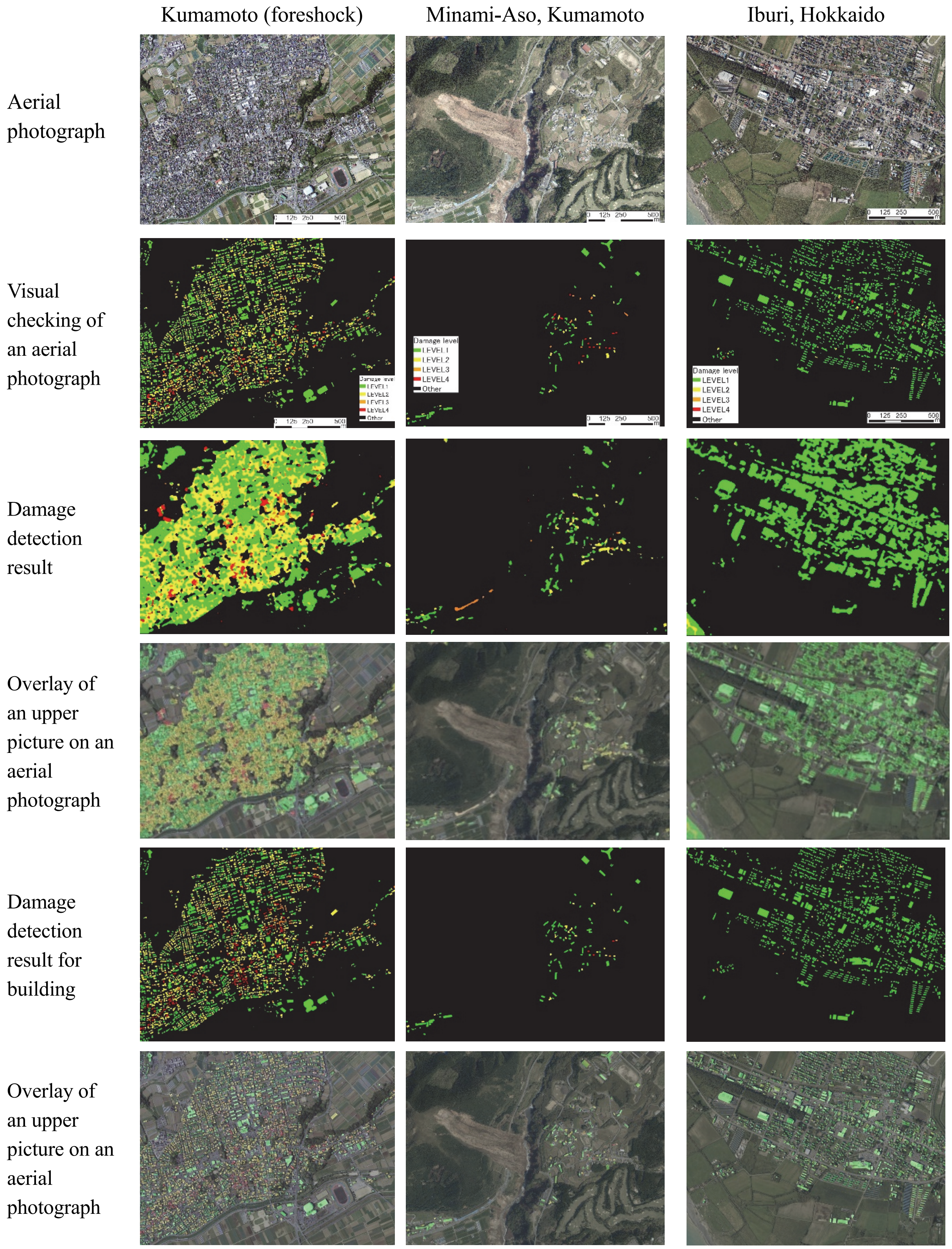


Fig. 24 Automatic damage detection results (2) obtained by applying the present model to unlearned aerial photographs

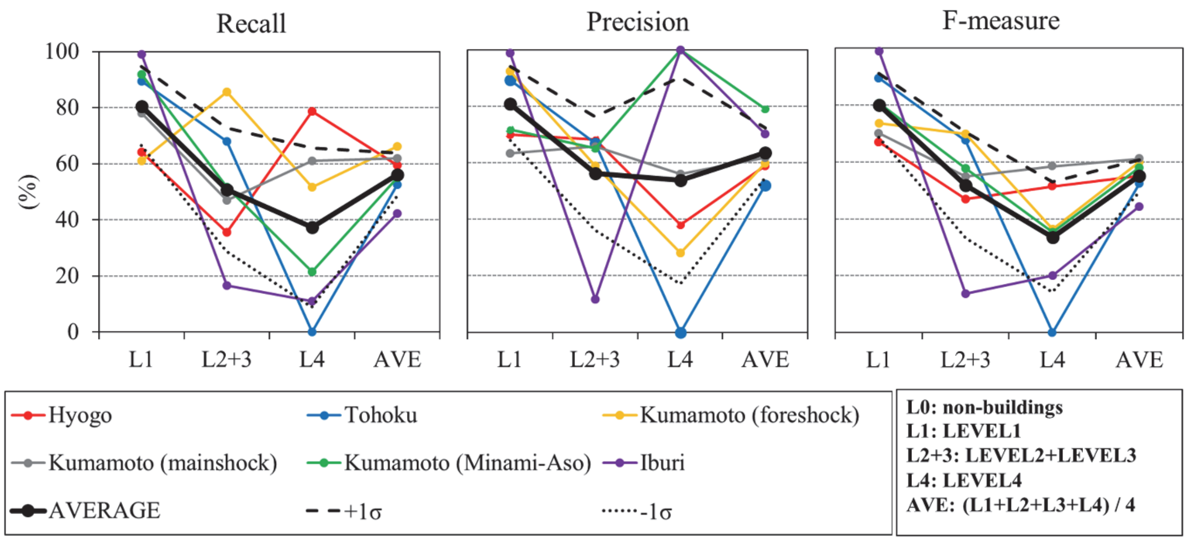


Fig. 25 Damage detection accuracy for buildings in aerial photographs used for testing

Table 30 Damage detection results for buildings (overall data)

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	4,272	938	85	5,295	80.7	85.2
	LEVEL 2+3	394	1,581	172	2,147	73.6	66.7
	LEVEL 4	65	78	185	328	56.4	48.1
Total		4,731	2,597	442	7,770	-	-
Precision (%)		90.3	60.9	41.9	-	Accuracy (%)	77.7

4.7 Characterization of damage detection results by earthquake

Damage detection results obtained when using aerial photographs for the tests mentioned in Section 4.6 were systematically analyzed, item by item, and the influences of seismic motions, building type, and other conditions on the detection accuracy of the present model were examined. In most cases, building types and the proportion of tile-roofed houses in each aerial photograph used for testing were estimated, mainly by visual checking. However, for the 2016 Kumamoto Earthquake, data obtained from a building damage information database[35)] found in a reference search were used instead.

The aerial photographs of the 2016 Kumamoto Earthquake (after main shock, Mashiki Town) that were used for testing were acquired in a region close to an inland earthquake fault where major seismic motions are known to have occurred. More than 90% of the buildings in this region were made of wood, and approximately 90% also had tiled roofs. Figure 26 shows a magnified view of such houses. Likewise, when seen from above, the damage features detected were visually distinct. Table 31 shows the confusion matrix for damage detection results in this region. The average recall was approximately 62.0% and the average precision was approximately 61.8%—reflecting a good balance in terms of results. As described above, the damage features were quite distinct and the availability of many similar features in the training data contributed to the relatively high recall and precision values obtained when assessing the various levels of damage in this region.

Table 31 Damage detection results for buildings affected by the 2016 Kumamoto Earthquake (after main shock, Mashiki Town)

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	239	47	20	306	78.1	70.0
	LEVEL 2+3	101	139	56	296	47.0	54.8
	LEVEL 4	37	25	97	159	61.0	58.4
Total		377	211	173	761	-	-
Precision (%)		63.4	65.9	56.1	-	Accuracy (%)	62.1

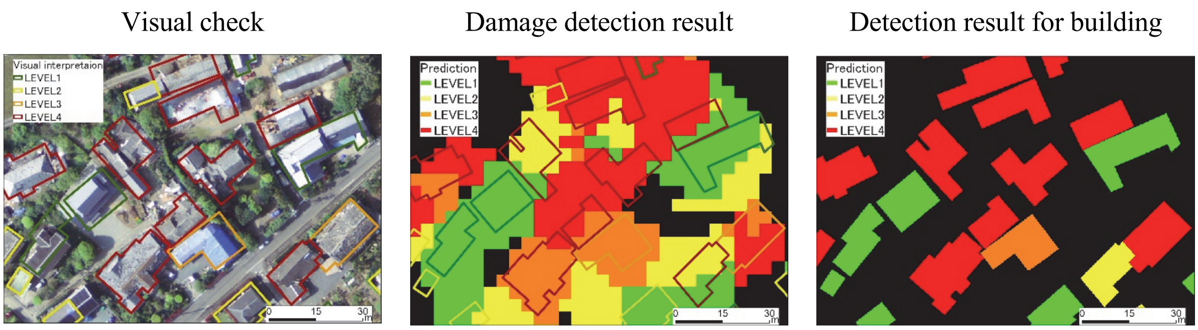


Fig. 26 Magnified view of damage detection results for the 2016 Kumamoto Earthquake (after main shock, Mashiki Town)

Aerial photographs of the 1995 Southern Hyogo Prefecture Earthquake used for testing targeted regions where strong seismic motions were caused by an inland fault. Approximately 80% of the buildings in this region were made of wood and approximately 90% had tiled roofs. The type of damage to these wooden buildings was usually quite distinct, including some cases of complete collapse. However, because the aerial photograph shown was taken on the same day as the earthquake, blue-tarp-covered buildings have yet to appear. Since the visual rating criteria used for LEVEL 2+3 included blue-tarp coverage, this might have contributed to the lower recall value observed for LEVEL 2+3. Looking at the confusion matrix, the average recall was approximately 59.5% and the average precision was approximately 58.8%. The proportion of LEVEL 2+3 was estimated to be slightly lower than the actual figure obtained from visual checking, and the proportion of LEVEL 4 was estimated to be slightly higher. The precision value for LEVEL 4 was also lower (Table 32). This can be explained as follows. In the target region, buildings were close to each other, as shown in Fig. 27. Therefore, some patches of adjoining buildings classified as LEVEL 4 were contained in the building polygons visually classified as LEVEL 1 and LEVEL 2+3, resulting in the LEVEL 4 rating. The target region also contained medium- to low-storied buildings, often constructed with reinforced concrete or steel. These structures accounted for approximately 20% of all buildings there. Even when visual checking from the air found no obvious signs of damage, structures such as pillars, beams and walls might actually have been damaged or weakened. Such damage to non-wooden buildings can be difficult to detect merely by observing from above. With this in mind, the present study focused mainly on detecting damage to wooden houses. As with all the other earthquakes examined, damage detection accuracy tended to decrease for non-wooden buildings.

Table 32 Damage detection results for buildings affected by the 1995 Southern Hyogo Prefecture Earthquake

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	75	11	31	117	64.1	67.0
	LEVEL 2+3	25	30	29	84	35.7	46.9
	LEVEL 4	7	3	37	47	78.7	51.4
Total		107	44	97	248	-	-
Precision (%)		70.1	68.2	38.1	-	Accuracy (%)	57.3

Visual check | Damage detection result | Detection result for building

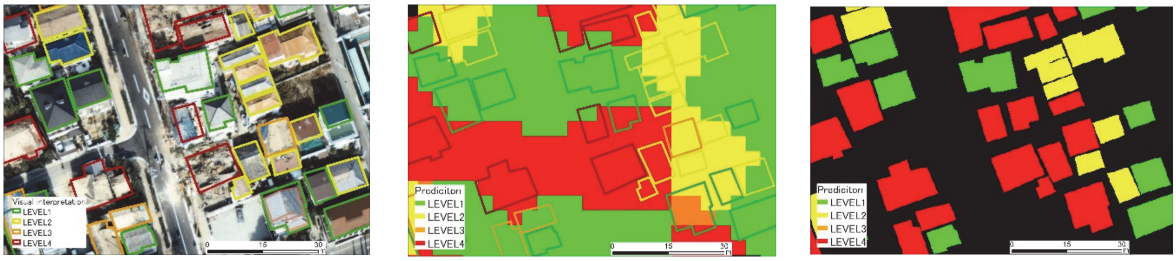


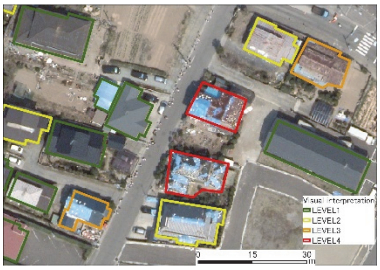
Fig. 27 Magnified view of damage detection results for the 1995 Southern Hyogo Prefecture Earthquake

The aerial photographs of the 2011 off the Pacific Coast of Tohoku Earthquake used for testing targeted regions where relatively large seismic motions were caused by a subduction-zone earthquake. Approximately 90% of the buildings in this region were made of wood, with tile-roofed buildings only accounting for approximately 20% of the total. When the proportion of tile-roofed buildings is this low, visual checking may overlook relatively low levels of damage[28]. Table 33 shows the confusion matrix for damage detection results in such a case. The average recall was approximately 52.5% and the average precision was approximately 52.1%, with no buildings identified as LEVEL 4. The fact that no LEVEL 4 damage was detected in the 2011 off the Pacific Coast of Tohoku Earthquake is probably due to the lack of any distinct features in the test data used for the LEVEL 4 rating. For example, even though two buildings near the center of the left panel in Fig. 28 were visually classified as LEVEL 4, both were automatically classified as LEVEL 3. Although these buildings were estimated to have been significantly damaged, they lacked any distinctive features indicating complete collapse, and there was probably also some variation among different assessors when classifying damaged buildings as LEVEL 3 or LEVEL 4. In the 2011 off the Pacific Coast of Tohoku Earthquake, only a few wooden houses actually collapsed around the quake intensity observation points because seismic motions in frequency zones highly correlated with building damage were absent in almost all regions[36]. Hence, there can be significant differences in seismic motion characteristics and frequency zones between inland active fault earthquakes like the 2016 Kumamoto Earthquake and subduction-zone earthquakes like the 2011 off the Pacific Coast of Tohoku Earthquake, which can in turn cause variable building damage modes and adversely affect the accuracy of any automatic damage detection systems.

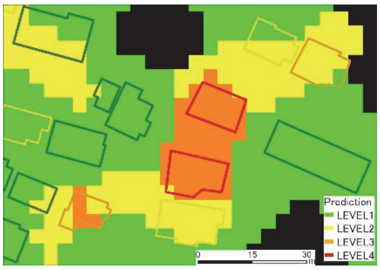
Table 33 Damage detection results for buildings affected by the 2011 off the Pacific Coast of Tohoku Earthquake

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	1,276	152	0	1,428	89.4	89.4
	LEVEL 2+3	151	322	0	473	68.1	67.5
	LEVEL 4	1	7	0	8	0.0	0.0
Total		1,428	481	0	1,909	-	-
Precision (%)		89.4	66.9	0.0	-	Accuracy (%)	83.7

Visual check



Damage detection result



Detection result for building

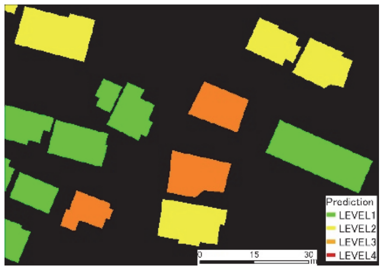


Fig. 28 Magnified view of damage detection results for the 2011 off the Pacific Coast of Tohoku Earthquake

Aerial photographs of the 2016 Kumamoto Earthquake (after foreshock) used for testing were taken in a region close to an inland earthquake fault where large seismic motions occurred. Approximately 80% of the buildings in this region were made of wood, with tile-roofed buildings also accounting for approximately 80%. Hence, the damage features in the various categories observed were quite distinct. Table 34 shows the confusion matrix for the detection results. The average recall was approximately 66.1% and the average precision was approximately 59.7%. As with the results obtained after the main shock of the 2016 Kumamoto Earthquake, the estimates of the various damage classes were quite evenly balanced. However, LEVEL 1 was slightly underestimated and LEVEL 2+3 and LEVEL 4 were slightly overestimated, compared with the visual check results (Table 34). In the magnified view of the three buildings marked with a green inverted triangle, as shown in Fig. 29, buildings for which the damage level could not be characterized by visual checking were sometimes automatically classified as LEVEL 2+3. This is attributable to the fact that some patches showing LEVEL 2+3 damage around the buildings in this region were contained in the building polygons when the buildings were close to each other, and because the same range of images obtained after the main shock of the 2016 Kumamoto Earthquake was also used as training data. The lower precision for LEVEL 4 is attributable to the fact that there were significantly fewer data items available compared with LEVEL 1 and LEVEL 2+3, so the influence of classification errors at LEVEL 4 increased accordingly.

Table 34 Damage detection results for buildings affected by the 2016 Kumamoto Earthquake (after foreshock, Mashiki Town)

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	1,167	708	34	1,909	61.1	73.6
	LEVEL 2+3	93	1,073	87	1,253	85.6	69.8
	LEVEL 4	4	40	47	91	51.6	36.3
Total		1,264	1,821	168	3,253	-	-
Precision (%)		92.3	58.9	28.0	-	Accuracy (%)	70.3

Visual check　　Damage detection result　　Detection result for building

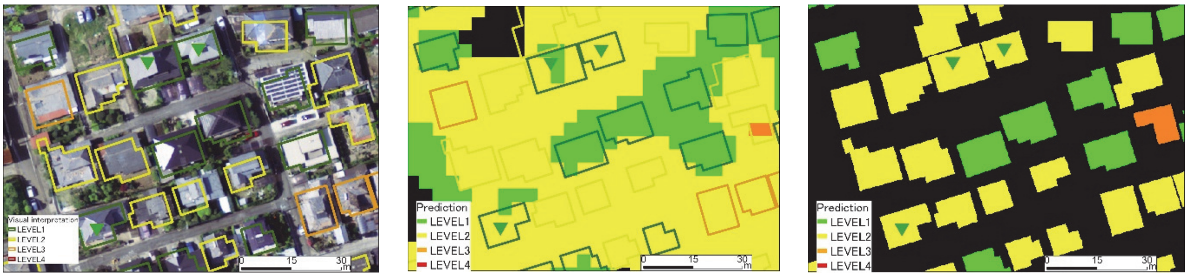


Fig. 29 Magnified view of damage detection results for the 2016 Kumamoto Earthquake (after foreshock, Mashiki Town)

Aerial photographs of the 2016 Kumamoto Earthquake (after main shock, Minami-Aso Village) used for testing were taken in a region where large seismic motions occurred due to the presence of an active fault. In addition, approximately 70% of the buildings in this region were made of wood, with tile-roofed buildings accounting for approximately 60% of the total. Hence, the damage features in the various categories were relatively distinct. As shown by the confusion matrix of ratings in Table 35, the average recall was approximately 55.0% and the average precision was approximately 79.0%. The proportion of LEVEL 4 damage was estimated to be slightly lower than the figure obtained by visual checking (Table 35). In this region, as shown in the left panel in Fig. 30, the number of building polygons was smaller than the number of actual buildings and the various locations and shapes also showed marked discrepancies. These discrepancies may have affected the detection accuracy for building damage levels. Likewise, the accuracy of building polygons in the Japan Basic Map Information dataset[24)] can be lower in non-urban regions[25)]. This polygon discrepancy also affected the automatic identification of training data, but any mis-identified data caused by this discrepancy were removed by visual screening to prevent any loss of quality in the training data used in the damage detection model (Section 2.3). With the present test data, the recall for LEVEL 4 was low, overall, at only 21.4%. The patch-based cross validation described in Section 4.6 showed that the recall value for LEVEL 4 was approximately 62.5% (Table 28), which was higher than the accuracy of building-based cross validation. Therefore, the loss of accuracy was probably due to the unsuccessful classification by threshold in cases where the extracted damaged portion was located outside the building polygon due to polygon shift, rather than to the damage detection model used.

Table 35 Damage detection results for buildings affected by the 2016 Kumamoto Earthquake (after main shock, Minami-Aso Village)

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	56	5	0	61	91.8	80.6
	LEVEL 2+3	14	15	0	29	51.7	57.7
	LEVEL 4	8	3	3	14	21.4	35.3
Total		78	23	3	104	-	-
Precision (%)		71.8	65.2	100.0	-	Accuracy (%)	71.2

Visual check   Damage detection result   Detection result for building

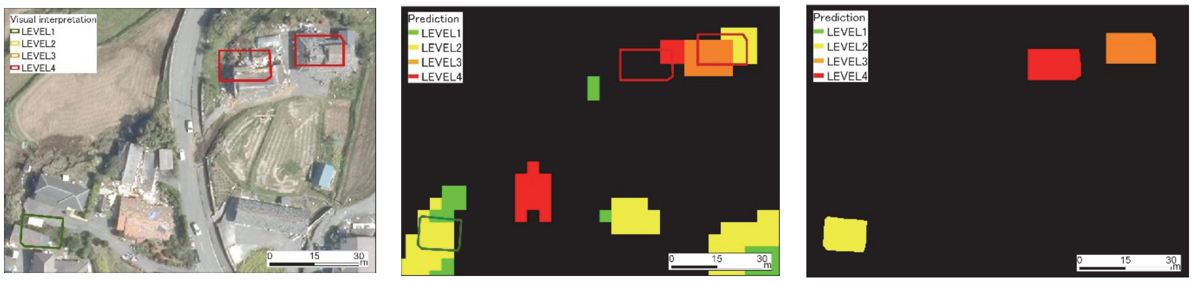


Fig. 30 Damage detection results for the 2016 Kumamoto Earthquake (after main shock, Minami-Aso Village)

Aerial photographs of the 2018 Hokkaido Iburi Eastern Earthquake used for testing were taken in a region where relatively large seismic motions were caused by an inland earthquake. In addition, approximately 90% of the buildings in this region were made of wood, with almost no tile-roofed houses. Aerial photograph ratings showed that most buildings were undamaged so, in most cases, damage features for each level were not distinct. As shown by the confusion matrix for ratings in Table 36, the average recall was approximately 42.3% and the average precision was approximately 70.2%. The proportion of LEVEL 4 damage was estimated to be slightly lower than that found by visual checking, but the proportion of LEVEL 2+3 was estimated to be slightly higher (Table 36). These low recall values for LEVEL 2+3 and LEVEL 4 can be attributed to the characteristic regional features reflected in local building shapes. The red-roofed building in the center of the left panel of Fig. 31, for example, was rated as LEVEL 4 because visual checking showed a distinct inclination, whereas the present model rated it as LEVEL 1 after automatic classification. Prior studies have shown that wooden houses in the Hokkaido region have a large amount of insulation because of their specifications, which require them to be able to withstand the snow and cold, with the proportion of houses incorporating bearing walls being higher than that in Honshu. Hence, proportion of houses with high quake resistance is higher[37]. Also, the proportion of tile-roofed houses is lower there than in other prefectures[38] and the wooden houses in Hokkaido are often markedly different, structurally, from those found in Honshu. Furthermore, images of LEVEL 2 or higher damage for the 2018 Hokkaido Iburi Eastern Earthquake accounted for only a small proportion of the training data constructed for the present study (Table 9). Hence, these low recall values could have been due to the inability of the system used to learn damage features properly. The precision for LEVEL 2+3 was also lower as a result of the relatively large influence of the mis-rating of visual check results for LEVEL 1 and other levels of damage because the number of buildings in LEVEL 2 and higher was much smaller than the number of buildings in LEVEL 1.

Table 36 Damage detection results for buildings affected by the 2018 Hokkaido Iburi Eastern Earthquake

		Automatic classification			Total	Recall (%)	F-measure (%)
		LEVEL 1	LEVEL 2+3	LEVEL 4			
Visual check	LEVEL 1	1,459	15	0	1,474	99.0	98.9
	LEVEL 2+3	10	2	0	12	16.7	13.8
	LEVEL 4	8	0	1	9	11.1	20.0
Total		1,477	17	1	1,495	-	-
Precision (%)		98.8	11.8	100.0	-	Accuracy (%)	97.8

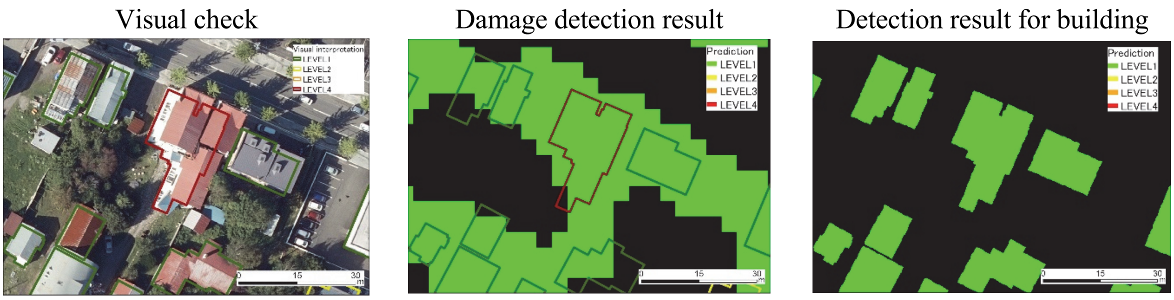


Fig. 31 Damage detection results for the 2018 Hokkaido Iburi Eastern Earthquake

5. CONCLUSION AND FUTURE ISSUES

Using six kinds of vertical aerial photographs taken under different imaging conditions just after the onset of the 1995 Southern Hyogo Prefecture Earthquake, the 2011 off the Pacific Coast of Tohoku Earthquake, the foreshock and main shock of the 2016 Kumamoto Earthquake, and the 2018 Hokkaido Iburi Eastern Earthquake, we constructed a large building damage training dataset. The accuracy of the constructed training data was then checked by comparison with field survey results from Mashiki Town, obtained after the main shock of the 2016 Kumamoto Earthquake. The average recall value with three levels of building damage (no damage, partial to half collapse, complete collapse) was determined to be approximately 74%, showing that visual checking of vertical aerial photographs is sufficiently accurate to reliably assess the damage status within these three damage levels.

We also developed a model for automatically classifying damage by applying deep learning to the same training data, and confirmed its higher rating performance compared with other models reported in previous studies. Furthermore, we found that developing a highly robust model by means of data augmentation, etc., using a large dataset comprising a wide variety of images as learning data, is a useful way of increasing generalization performance.

Using the model and dataset developed in the present study, we performed frame-by-frame cross validation using six kinds of aerial photograph to obtain training data and test data from several different places, and found that the detection performance with the three damage levels produced an average recall value of approximately 77%. We also found, using untrained aerial photographs for testing, that ratings could be determined with an average recall accuracy of approximately 70% for all three damage levels for all buildings shown in the image.

Rating results were also analyzed for characterization, and some specific issues were identified. First, the training dataset constructed did not include enough data on LEVEL 4 damage caused by subduction-zone earthquakes and LEVEL 2 or higher damage in cold regions like Hokkaido. This resulted in decreased detection accuracy for such damage. In addition, while the aerial photographs used for testing targeted regions with a high proportion of wooden houses, houses made of reinforced concrete (RC) or steel-framed (S) structures and low- and medium-storied buildings could show different forms of damage that were not always evident from their external appearance[39]. Likewise, many high-storied buildings could not be adequately assessed merely from their appearance. Therefore, the present study was short of training data covering damage to these non-wooden buildings. Furthermore, wooden but tile-roofed buildings accounted for only a small proportion of housing in the Hokkaido and Tohoku districts[38] – which could have affected the damage detection accuracy of the present model. Hence, we would like to develop improved technology for assessing earthquake damage more accurately by continuing to collect images of damage from both inside and outside Japan in order to enhance our training data. We would also like to make the best possible use of additional sources of data, other than just vertical aerial photographs.

We also identified the impacts that the resolution of the building polygons used and the building density and other local factors could have on detection accuracy. As a result of these findings, several techniques for the automatic identification of building boundaries by means of image recognition using aerial photographs[40, 41] have now been proposed, and various technologies for the automatic identification of objects of any size from images[42, 43] have been developed. In the future, we would like to further investigate any technique capable of building damage detection in those cases where highly accurate polygon data cannot be obtained using such technologies.

Moreover, in the present study, we developed a model capable of automatically detecting building damage at a reasonable level using six kinds of aerial photograph. If this model is to be applied to new disasters in the future, its accuracy is likely to be insufficient, depending on the presence of damage features related to seismic motions, architectural structures and other factors. However, the detection accuracy is expected to be increased by streamlining the training process with the use of techniques such as transfer learning[44] which would allow the weighting information constructed in the present study to be used for the next learning stage. In this study, the training time required was up to about 9 hours for the 2000 epochs of training performed in Section 4.4, and about 1 to 3 hours per training attempt for the 500 epochs of cross validation performed in Section 4.5. We found, however, that calculation time can be shortened by finishing the calculation in fewer attempts with the use of transfer learning. We are, therefore, planning to carry out a detailed verification of the effectiveness of transfer learning techniques.

In addition, while the present study exclusively targeted aerial photographs taken from fixed-wing aircraft, aerial photographs cannot always be acquired over the entire range required just after the onset of a disaster. Furthermore, aerial photographs taken solely in the vertical plane can sometimes overlook the buckling of pillars, beams, and other structural members of buildings, as well as wall damage. For this reason, it is desirable for any detailed damage classification to also include images taken using artificial satellites, helicopters, UAVs, etc., and that these be used in combination with existing aerial photographs. Using LiDAR, stereoscopic images, and other sources of data in combination, we can detect building height changes before and after an earthquake. In future, we would like to develop a building damage detection model with improved accuracy by using images taken from all these different kinds of platform sensors, three-dimensional models, and other approaches used in combination.

Finally, we would like to systematically work towards making the best possible use of the model developed in the present study to support disaster responses, in collaboration with various other organizations, by identifying any issues that may arise and continuing to develop and apply this system to help society recover from any future disasters.

ACKNOWLEDGEMENT

The present study was based on work conducted by its first author when completing his doctoral paper, with experimental results and discussion sections regarding newly constructed training data added to expand the research focus, as described in the present article. We would like to thank the mentor for my doctoral thesis, Professor Gaku Shoji at the University of Tsukuba, for providing comprehensive guidance for the present study as a whole. Our thanks should also go to my sub-mentors for their valuable opinions and contributions to streamlining our research. I would also like to thank the anonymous reviewer who provided valuable suggestions helping us to improve the content of our article. Please note that some of the operations used for training data preparation and program development were performed within the framework of the 2013–2018 Cross-ministerial Strategic Innovation Promotion Program (SIP) titled "Enhancement of Societal Resiliency against Natural Disasters."

REFERENCES

1) Yamazaki, F.: Review on the Developments of Earthquake Monitoring and Early Damage Assessment Systems in Japan, *Journal of Japan Society of Civil Engineers*, No. 577, I-41, pp. 1–16, 1997 (in Japanese).
2) Okayama, K., Kiriyama, T. and Yabuuchi, S.: Development of the Disaster Information System (DIS/Earthquakes), *Proceedings of the 30th Joint Meeting on Wind and Seismic Effects*, pp. 286–299, 1998.
3) National Research Institute for Earth Science and Disaster Resilience: *Japan Real-Time Information System for Earthquake*, https://doi.org/10.17598/nied.0017
4) Nojima, N., Sugito, M. and Kanazawa, N.: Decision Support Model for Post-Earthquake Emergency Management Based on Sequential Data Processing of Damage Information, *Journal of Japan Society of Civil Engineers*, No. 682, I-56, pp. 129–142, 2001 (in Japanese).
5) Nojima, N., Matsuoka, M., Sugito, M. and Esaki, K.: Quantitative Estimation of Building Damage Based on Data Integration of Seismic Intensities and Satellite SAR Imagery, *Journal of Japan Society of Civil Engineers A*, Vol. 62, No. 4, pp. 808–821, 2006 (in Japanese).
6) Yamazaki, F. and Liu, W.: Remote Sensing Technologies for Post-Earthquake Damage Assessment: A Case Study on the 2016 Kumamoto Earthquake, *6th ASIA Conference on Earthquake Engineering (6ACEE)*, Cebu City, Philippines, September 22nd–24th, KN-4, 13 pp., 2016.
7) Ministry of Land, Infrastructure, Transport and Tourism (MLIT): *Technical Report of the Geospatial Information Authority of Japan*, C1-400, 243 pp., 2010 (in Japanese).
https://www.gsi.go.jp/common/000226161.pdf (last accessed on January 8, 2021)
8) Cabinet Office: *Disaster Management in Japan* (in Japanese).
http://www.bousai.go.jp/taisaku/unyou.html (last accessed on May 11, 2020)
9) Ogawa, N. and Yamazaki, F.: Photo-Interpretation of Building Damage Due to Earthquakes Using Aerial Photographs, *Proceedings of the 12th World Conference on Earthquake Engineering*, 1906, 8 pp., 2000.
10) Hasegawa, H., Yamazaki, F., Matsuoka, M. and Sekimoto, I.: Determination of Building Damage Due to Earthquakes Using Aerial Television Images, *Proceedings of the 12th World Conference on Earthquake Engineering*, 1722, 8 pp., 2000.
11) Sakai, Y., Arai, K. and Akamatsu, K.: Estimation of Seismic Intensity Distribution around Omaezaki-City from Roof Tiles of Damage by the 2009 Suruga Bay Earthquake, *Journal of Japan Association for Earthquake Engineering*, Vol. 11, No. 1, pp. 48–71, 2017 (in Japanese).
12) Kojima, Y., Uchida, I., Ikeda, T. and Tsushima, K.: Evaluation of the Housing Damage at Mashiki Town Using the Aerial Photograph, *Journal of Japan Society of Civil Engineers A1*, Vol. 73, No. 4, pp. 949–958, 2017 (in Japanese).
13) Ueoka, Y., Tanaka, S. and Abe, I.: Effectiveness of Rapid Safety Assessment of Buildings after

Foreshock of the 2016 Kumamoto Earthquake Using Aerial Photographs, *Journal of Social Safety Science*, No. 32, pp. 1–9, 2018 (in Japanese).

14) Fujita, A., Sakurada, K., Imaizumi, T., Ito, R., Hikosawa, S. and Nakamura, R.: Damage Detection from Aerial Images Via Convolutional Neural Networks, *15th IAPR International Conference on Machine Vision Applications (MVA),* Nagoya, Japan, May 8–12, 4pp., 2017.
15) Tonegawa, R., Iizuka, H., Yamamoto, M., Furukawa, M. and Ohuchi, A.: Detecting Collapsed Buildings Using Convolutional Neural Network for Estimating the Disaster Debris Amount, *Information Processing Society of Japan (ISPJ) Journal*, Vol. 57, No. 6, pp. 1565–1575, 2016 (in Japanese).
16) Kamagatani, Y. and Matsuoka, M.: Damaged Building Recognition of the 2016 Kumamoto Earthquakes Using Deep Learning with Aerial Photographs, *Report of Tono Research Institute of Earthquake Science*, No. 41, pp. 49–57, 2017 (in Japanese).
17) Ueoka, Y., Tanaka, S., Abe I., Kamagatani, Y. and Matsuoka, M.: Visual Detection and Deep Learning Interpretation of Building Damages by the 2016 Kumamoto Earthquake Using Aerial Photographs, *Proceedings of the annual conference of the Institute of Social Safety Science*, No. 41, pp. 127–130, 2017 (in Japanese).
18) Ishii, Y., Matsuoka, M., Maki, N., Horie, K. and Tanaka, S.: 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*, Vol. 83, No. 751, pp. 1391–1400, 2018 (in Japanese).
19) Vetrivel, A., Gerke, M., Kerle, N., Nex, F. and Vosselman, G.: 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*, Vol. 140, pp. 45–59, 2018.
20) Ci, T., Liu, Z. and Wang, Y.: Assessment of the Degree of Building Damage Caused by Disaster Using Convolutional Neural Networks in Combination with Ordinal Regression, *Remote Sensing*, Vol. 11, No. 23, 2858, pp. 1–19, 2019, https://doi.org/10.3390/rs11232858
21) Miura, H., Aridome, T. and Matsuoka, M.: Deep Learning-Based Identification of Collapsed, Non-Collapsed and Blue Tarp-Covered Buildings from Post-Disaster Aerial Images, *Remote Sensing*, Vol. 12, No. 12, 1924, pp. 1–19, 2020, https://doi.org/10.3390/rs12121924
22) Naito, S., Tomozawa, H., Mori, Y., Nagata, T., Monma, N. Nakamura, H., Fujiwara, H. and Shoji, G.: Building Damage Detection Method Based on Machine Learning Utilizing Aerial Photographs of the Kumamoto Earthquake, *Earthquake Spectra,* Vol. 36, No.3, pp. 1166–1187, 2020, https://doi.org/10.1177/8755293019901309
23) Naito, S., Nakamura, H., Fujiwara, H., Tomozawa, H., Mori, Y., Nagata, T., Yamada, T., Shimomura, H., Monma, N. and Shoji, G.: Development of the Damage Detection Method for Buildings Utilizing Machine Learning Techniques with Aerial Photographs, *The 15th Japan Earthquake Engineering Symposium*, GO 07-01-09, pp. 1257–1266, 2018 (in Japanese).
24) Geospatial Information Authority of Japan: Fundamental Geospatial Data. https://www.gsi.go.jp/kiban/index.html (last accessed on May 13, 2020)
25) Naito, S., Monma, N., Nakamura, H., Fujiwara, H., Shimomura, H. and Yamada, T.: Investigation of Building Damages Caused by the 2016 Kumamoto Earthquake Utilizing Aerial Photographic Interpretation, *Journal of Japan Society of Civil Engineers A1*, Vol. 74, No. 4, pp. I464–I480, 2018. (in Japanese) https://doi.org/10.2208/jscejseee.74.I_464
26) Okada, N. and Takai, N.: Classification of Structural Types and Damage Patterns of Buildings for Earthquake Field Investigation, *Proceedings of the 12th World Conference on Earthquake Engineering*, 0705, 8 pp., 2000.
27) Naito, S., Hao, K. X., Senna, S., Saeki, T., Nakamura, H., Fujiwara, H. and Azuma, T.: Investigation of Damages in Immediate Vicinity of Co-Seismic Faults During the 2016 Kumamoto Earthquake, *Journal of Disaster Research*, Vol. 12, No. 5, pp. 899–915, 2017. https://doi.org/10.20965/jdr.2017.p0899
28) Naito, S.: Real-Time Seismic Damage Estimation Method for Building Utilizing Image Recognition and Machine Learning Techniques, *Doctor thesis of Graduate school of Systems and Information Engineering, University of Tsukuba*, 231 pp., September, 2019. (in Japanese)

https://doi.org/10.15068/00158013.
29) Naito, S., Nakamura, H., Fujiwara., H., Tomozawa, H., Mori, Y. and Monma, N.: CNN-Based Building Damage Detection Model for the Prompt Disaster Response Utilizing Remote Sensing Images, *17th World Conference on Earthquake Engineering (17WCEE)*, Sendai, Japan, September 13–18, 6a-0001, 12 pp., 2020.
30) Harada, T.: *Image Recognition*, Machine Learning Professional Series, Kodansha, 277 pp., 2017 (in Japanese).
31) LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. and Jackel, L. D.: Backpropagation Applied to Handwritten Zip Code Recognition, *Neural Computation*, No. 1, Vol. 4, pp. 541–551, 1989.
32) Taki, M.: *Introduction to Deep Learning*, Machine Learning Professional Series, Kodansha, 339 pp., 2017 (in Japanese).
33) Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *International Conference on Learning Representations*, San Diego, CA, May 7–9, pp. 1–14, 2015.
34) Keras Documentation. https://keras.io/ja/ (last accessed on May 14, 2020)
35) Monma, N., Nakamura, H., Fujiwara, H., Naito, S., Shimomura, H. and Yamada, T.: Construction of Earthquake Building Damage Information Space Database and Examination of Building Damage Curve, *The 15th Japan Earthquake Engineering Symposium*, PS1-01-33, pp. 2924–2933, 2018 (in Japanese).
36) Hayashi, Y., Iizuka, H., Shimomitsu, M., Kobayashi, Y. and Sakai, Y.: Damage Investigation of Surroundings of the Seismic Station at Miyagi in the 2011 off the Pacific Coast of Tohoku Earthquake and Correspondence of Damage to Buildings with Strong Ground Motions, *Journal of Japan Association for Earthquake Engineering*, Vol. 13, No. 5, pp. 62–101, 2013 (in Japanese).
37) Takeuchi, S., Okada, N. and Nakashima, T.: Advanced Method of Seismic Vulnerability Functions of Wooden Buildings in Consideration of Regional Characteristic and Construction Year—Based on the Seismic Index Distribution of Wooden Houses in Case of Hokkaido, *Journal of Structural and Construction Engineering*, Vol. 83, No. 753, pp. 1549–1559, 2018 (in Japanese).
38) Shimizu, S., Komaru, Y., Wakaura, M., Tokizane, Y., Nakamura, H. and Fujiwara, H.: Fragility Function Wooden House Considering the Roofing Type, *Journal of Social Safety Science*, No. 34, pp. 1–11, 2019 (in Japanese).
39) Takai, N. and Okada, N.: Classifications of Damage Patterns of Reinforced Concrete Buildings for Earthquake Field Investigation, *Journal of Structural and Construction Engineering*, No. 549, pp. 67–74, 2001 (in Japanese).
40) Shi, Z. and Shibasaki, R.: Automated Building Extraction from Aerial Photographs in 3D Space, *Journal of the Japan Society of Photogrammetry*, Vol. 36, No. 2, pp. 38–44, 1997 (in Japanese).
41) Susaki, J.: Automatic Extraction of Buildings in Defense Urban Areas Using Aerial Images and Airborne Lidar Data, *Journal of Japan Society of Civil Engineers F3*, Vol. 70, No. 2, pp. I123–I132, 2014 (in Japanese).
42) Redmon, J., Divvala, S., Firshick, R. and Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.
43) Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y. and Berg, A. C.: SSD: Single Shot MultiBox Detector, *European Conference on Computer Vision (ECCV)*, pp. 21–37, 2016.
44) Goodfellow, I., Bengio, Y. and Courville, A.: *Deep Learning*, The MIT Press, 800 pp., 2016.

(Original Japanese Paper Published: November, 2020)
(English Version Submitted: February 4, 2021)
(English Version Accepted: March 22, 2021)