Detection of Landslide Candidate Interference Fringes in DInSAR Imagery Using Deep Learning

Joko KAMIYAMA[1]*, Tomoyuki NORO[1], Masayuki SAKAGAMI[1], Yamato SUZUKI[1]
Kazuo YOSHIKAWA[2], Shuhei HIKOSAKA[2] and Ikushi HIRATA[2]

1 National Institute for Land and Infrastructure Management, MLIT, Japan
2 PASCO Corporation, Japan
*Corresponding author. E-mail: kamiyama-j253@mlit.go.jp

Interferometric synthetic aperture radar (InSAR) is an effective technique for monitoring the risk of large-scale sediment movements because it can broadly and routinely observe the extent of landslides. To detect interference fringes with the possibility of landslide movements from differential interferograms, it is common for experts to interpret these fringes considering the effects of water vapor as well as the topography and other factors. Increasing the accuracy of detecting landslides is an important issue in the usage of InSAR. Convolutional Neural Networks (CNNs) that enable image recognition with sufficient accuracy have recently been developed. To efficiently detect landslide candidate interference fringes, this study evaluated the effectiveness of introducing a CNN model to detect the interference fringes representing landslide movements using similar processes as experts techniques. As a result, the CNN model was able to detect landslide candidate interference fringes that had been detected by experts with a recall of approximately 90%.

Key words: Landslide detection, DInSAR, Deep Learning, Convolutional Neural Networks

1. INTRODUCTION

In Japan, serious damage is often caused by large-scale erosion and sedimentation processes, such as the heavy rain caused by the 2011 Typhoon No. 12 (Talas) and the deep-seated landslides and landslide dams caused by the 2008 Iwate-Miyagi Inland Earthquake. In the prevention/mitigation of sediment disasters, the monitoring of potential landslide slopes based on their detection is essential.

An interference analysis using synthetic aperture radar mounted on a satellite, i.e., interferometric synthetic aperture radar (InSAR), is a monitoring technique that can broadly and routinely observe minor landslide movements. Differential Interferometric SAR (DInSAR), which provides a measurement accuracy of several centimeters, is an effective technique for detecting landslides. Because Japan's mountainous areas are widely covered with forests, analyses of ground surface movements with a few SAR images can be performed to meet requirements such as the number of persistent scatterers or the need for multiple analytic pairs with good coherence.

In recent years, multiple artificial satellites equipped with SAR with resolutions of several meters have been in operation worldwide. Increases in the observation frequency due to the operation of multiple satellites and the expansion of the observation width will enable more frequent landslide monitoring in the same location and DInSAR for landslide monitoring will be used more than ever before.

Changes in the interference fringes in differential interferograms obtained from DInSAR may appear due to factors other than ground surface movements. There are various factors that produce interference fringes, such as satellite orbit differences or the delay of radio waves due to water vapor [*Shimada*, 1999]. However, it is difficult to thoroughly remove local effects that appear in slopes from global effects on a large spatial scale. Therefore, the detection of interference fringes that may indicate landslide movements depends on expert interpretations. However, it would require enormous amounts of time and effort for experts to interpret all the differential interferograms created from observational data obtained over wide areas at high frequency. In addition, variations in the judgment criteria are likely to occur among experts. To

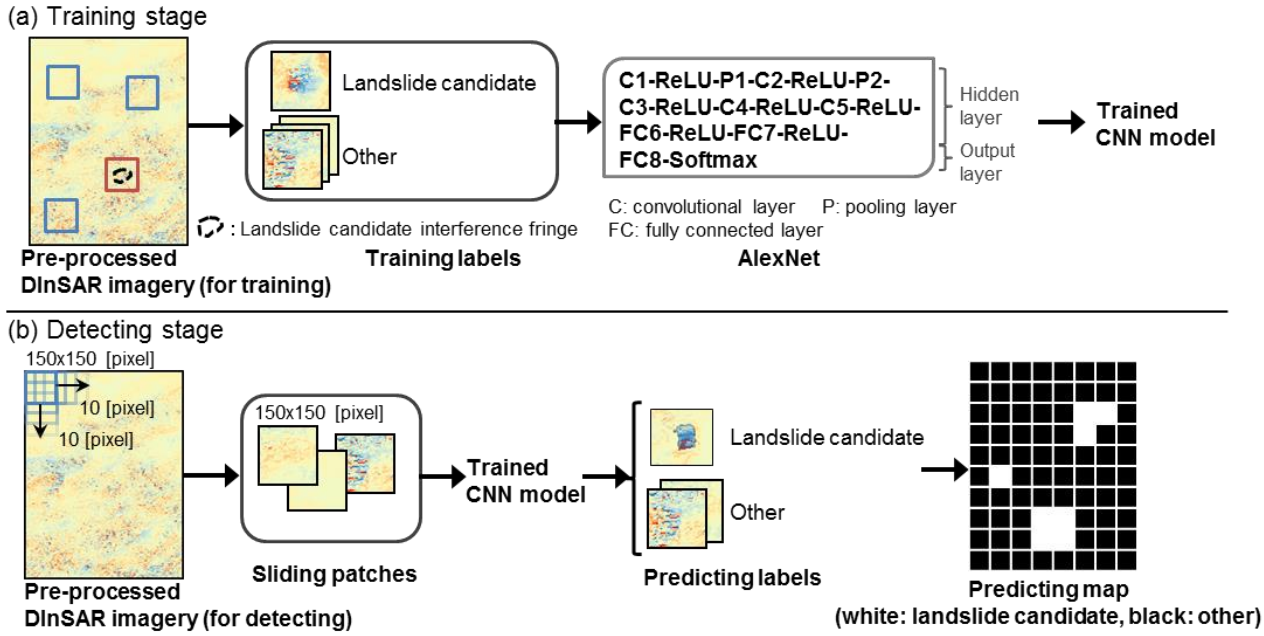

Fig. 1 Architecture of the CNN model and image of (a) the training stage and (b) the detection stage

increase the accuracy of landslide detection is an important issue in the usage of DInSAR.

In recent years, to deal with the large amount of image data, several models have been developed [e.g., *Krizhevsky et al.*, 2012] that recognize images with high accuracy using Convolutional Neural Networks (CNNs), one of the deep learning approaches. Case studies on CNNs for remote sensing images have been increasing. Deep learning using SAR data include studies on automatic target recognition from reflection intensity images [e.g., *Wilmanski et al.*, 2016] and on land cover classification using polarimetric SAR data [e.g., *Zhou et al.*, 2016]. However, there are no cases where the locations of ground surface movements are detected from differential interferograms.

Therefore, with the aim of efficiently detecting interference fringes that may represent landslides from differential interferograms, this study attempted to evaluate the effectiveness of introducing a CNN model to detect interference fringes representing landslide movements, using similar processes to those of experts.

2. METHOD

2.1 Learning model

We used CNNs for learning. CNNs are a type of forward-propagation neural network having two-dimensional convolution layers, and as a specific feature, high-level characteristics can be obtained by synthesizing low-level characteristics. [*LeCun et al.*, 2015].

Various architectures have been proposed for CNNs [*Okatani*, 2016]. In this study, we adopted AlexNet [*Krizhevsky et al.*, 2012] because the amount of training data is small, the learning takes less time than the deep architecture, and there are many research results (**Fig. 1**). In the AlexNet architecture, the hidden layer consists of five convolutional layers, two pooling layers, and two fully connected layers and ReLU was adopted as an activation function. The output layer consists of one

Table 1 The learning conditions for the CNN model

Parameter	Meaning	Setting value
Optimization method	Optimization calculation method for searching for optimal weight parameter in learning of neural network	SGD* (Stochastic gradient descent)
Initial learning rate	Parameter of SGD. The initial value of the coefficient that determines the size of update amount to the parameters (weight, bias) learned in the CNN.	0.01
Learning rate schedule method	Parameter of SGD. A method of changing the learning rate from the initial value as the learning time passes.	1/10 each time 5,000 times of iteration
Weight decay	Parameter of SGD. For the coefficient of reducing the weight of learning to be set for preventing overfitting to the learning data of the neural network.	0.0005
Momentum	Parameter of SGD. The techniques to enhance the convergence of learning.	0.9
Mini-batch size	Number of learning data units for calculating error with one iteration of learning.	50
Number of iterations	Number of times to learning iterations	30,000

*[*David et al.*, 1986]

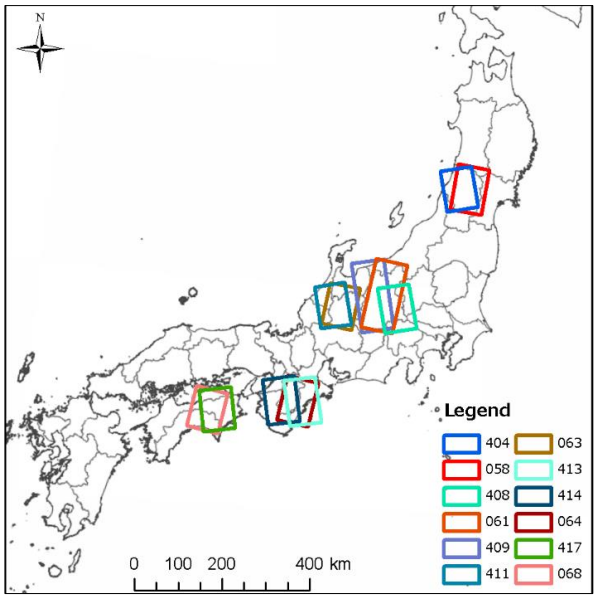

Fig. 2 Map of observed area of differential interferogram used for training data

fully connected layer, and Softmax was adopted as an activation function.

The CNN model conducted learning with back propagation using mini-batch stochastic gradient descent. For the learning conditions, the same conditions as that of AlexNet were adopted (**Table 1**), except for the parameters shown below. The parameters that depend on the number of training data (the learning rate schedule method, the mini-batch size, and the number of iterations) were determined considering the number of used training data after confirming that the learning converged.

2.2 Differential interferograms used for learning

For learning, we used the differential interferograms of ALOS/PALSAR from 2006 to 2011. We used images of Mt. Gassan in Yamagata Prefecture, the northern part of Nagano Prefecture, Mt. Hakusan in Ishikawa Prefecture, the Nara and Wakayama prefectures (the Kii Peninsula), and the eastern part of Kochi Prefecture, including locations previously detected as having the possibility of landslides [*Kamiyama et al.*, 2016, 2017]. **Fig. 2** shows the observed area and **Table 2** shows a list of the observed differential interferogram data used for the training data. The data are observed in high-resolution mode and the image processing level is L1.0.

Fig. 3 shows the processing flow of DInSAR. Interference fringes caused by changes in the vapor distribution in the atmospheric air between each observation often appear in DInSAR. When interference fringes resulting from atmospheric air are found in differential interferograms, it may be

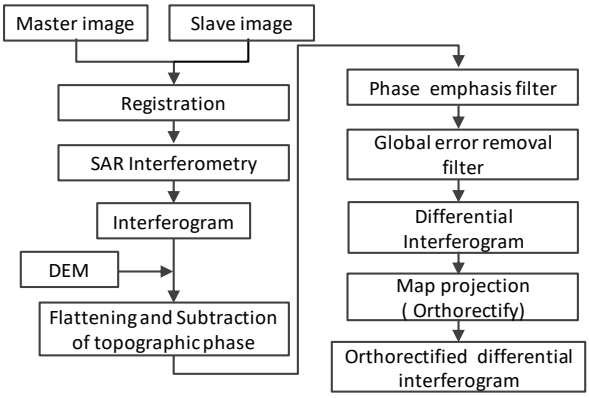

Fig. 3 Processing flow of DInSAR

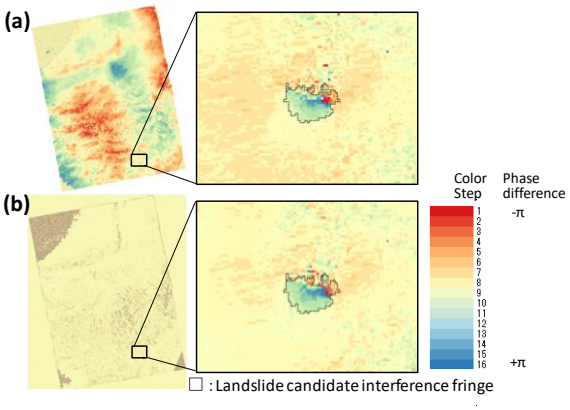

Fig. 4 Examples of differential interferograms (a) without global error removal filters and (b) with global error removal filters

unable to learn the interference fringes caused by landslides properly and may misdetect them. Therefore, we used the global error removal filter proposed by *Kusano et al.* [2015] to remove interference fringes resulting from atmospheric air, which have a larger spatial scale than landslides and appear at scales of several hundreds of meters or kilometers. The sizes of the filters we tried were 500 m, 1000 m, 2000 m, and 3000 m. We chose a filter size of 2000 m because it can detect the area of the interference fringes with the possibility of landslides, as in the case of no filters, and can remove interference fringes caused by atmospheric air. **Fig. 4** shows examples of differential interferograms for a case where the global error removal filters are applied and a case where they are not applied. In **Fig. 4**(b), the global interference fringes have been removed from the image in **Fig. 4**(a). There is nearly no change in the color gradation of the interference fringes with the possibility of landslides, as seen in the enlarged views of **Fig. 4**.

In addition, we used differential interferograms for learning by classifying the phase differences from $-\pi$ to $+\pi$ into 16 color gradation steps (**Fig. 4**).

Table 2 List of observation data for the differential interferograms used for training data

Area	Observed area		Observation date	
	Orbit number	Center frame number	Master	Slave
Mt. Gassan	404	760	25/10/2007	17/06/2010
	58	2840	12/09/2006	17/09/2008
			05/05/2009	05/08/2009
			05/05/2009	05/11/2009
			23/03/2010	08/05/2010
Northern Nagano	408	710	01/07/2007	16/08/2007
			01/07/2007	01/10/2007
			01/07/2007	03/07/2008
			01/07/2007	06/10/2009
			03/07/2008	21/08/2009
			03/07/2008	06/10/2009
			18/08/2008	06/07/2009
			03/10/2008	06/07/2009
			24/05/2010	24/08/2010
			24/05/2010	09/10/2010
			09/07/2010	24/08/2010
			09/07/2010	09/10/2010
	61	2880	05/11/2007	21/12/2007
			21/12/2007	22/03/2008
			07/08/2008	22/09/2008
	61	2890	05/11/2007	07/05/2008
			21/12/2007	22/03/2008
	409	710	20/07/2008	07/09/2009
	409	720	18/07/2007	20/07/2008
			18/07/2007	18/10/2007
			02/09/2007	18/10/2007
			20/07/2008	07/09/2009
			20/07/2008	23/10/2009
			20/10/2008	07/03/2009
			23/07/2009	07/09/2009
			07/09/2009	23/10/2009
			25/04/2010	10/09/2010
			10/06/2010	26/07/2010
			10/06/2010	10/09/2010
			10/09/2010	26/10/2010
Mt. Hakusan	411	710	06/07/2007	21/08/2007
			21/08/2007	06/10/2007
			26/08/2009	11/10/2009
			26/08/2009	26/11/2009
			29/05/2010	14/07/2010
			14/07/2010	29/08/2010
			29/08/2010	14/10/2010
			14/10/2010	29/11/2010
			06/07/2007	11/10/2009
			21/08/2007	11/10/2009
			23/05/2008	29/08/2010
			23/05/2008	14/10/2010
	63	2890	08/06/2007	24/07/2007
			08/06/2007	24/10/2007
			08/06/2007	29/10/2009
			24/07/2007	13/06/2009
			24/07/2007	29/10/2009
			24/10/2007	10/06/2008
			24/10/2007	29/10/2009
			13/06/2009	29/10/2009

Area	Observed area		Observation date	
	Orbit number	Center frame number	Master	Slave
Kii Peninsula	413	660	24/06/2007	09/08/2007
			24/06/2007	26/06/2008
			09/11/2007	11/05/2008
			26/03/2008	17/08/2010
			26/09/2008	11/02/2009
			29/09/2009	14/11/2009
	414	660	08/10/2006	13/07/2008
			08/01/2007	13/10/2008
			11/10/2007	03/03/2010
			11/10/2007	18/04/2010
			11/01/2008	18/04/2010
			11/01/2008	03/06/2010
			13/07/2008	16/07/2009
			16/07/2009	16/10/2009
			03/03/2010	19/07/2010
			18/04/2010	19/10/2010
			03/06/2010	03/09/2010
			03/06/2010	19/10/2010
			19/07/2010	19/10/2010
	414	670	08/10/2006	08/01/2007
			08/10/2006	13/07/2008
			11/07/2007	26/08/2007
			11/07/2007	11/10/2007
			11/07/2007	11/01/2008
			11/07/2007	01/12/2009
			26/08/2007	11/01/2008
			26/08/2007	16/01/2010
			26/08/2007	03/03/2010
			11/10/2007	11/01/2008
			11/10/2007	26/02/2008
			11/10/2007	03/03/2010
			11/10/2007	18/04/2010
			11/10/2007	03/06/2010
			11/01/2008	03/03/2010
			11/01/2008	18/04/2010
			11/01/2008	03/06/2010
			13/07/2008	28/02/2009
			13/07/2008	16/07/2009
			13/07/2008	16/10/2009
			28/08/2008	13/10/2008
			16/10/2009	18/04/2010
			16/10/2009	03/06/2010
			01/12/2009	18/04/2010
			01/12/2009	03/06/2010
			16/01/2010	03/06/2010
			03/03/2010	03/09/2010
			18/04/2010	03/09/2010
			18/04/2010	19/10/2010
			03/06/2010	19/10/2010
			19/07/2010	19/10/2010
	64	660	23/12/2006	30/09/2009
			26/12/2007	18/08/2010
	64	2940	23/12/2006	30/09/2009
			10/08/2007	30/09/2009
			26/12/2007	18/05/2010
			26/12/2007	18/08/2010
Eastern Kochi	417	660	28/08/2006	31/08/2007
			13/10/2006	18/07/2008
			13/01/2007	03/12/2008
			16/10/2007	08/03/2010
			16/10/2007	23/04/2010
			16/10/2007	08/06/2010
			01/12/2007	08/03/2010
			01/12/2007	23/04/2010
			01/12/2007	08/06/2010
			16/01/2008	23/04/2010
			16/01/2008	24/07/2010
	68	2940	14/01/2007	17/01/2008
			09/06/2010	25/07/2010
			09/06/2010	09/09/2010
			25/07/2010	09/09/2010
			25/07/2010	25/10/2010

2.3 Classes of training data

For the CNN model, we used two classes of training data, i.e., "landslide candidates", meaning interference fringes where landslides are likely to occur, and "others", areas detected in a random manner from areas other than the landslide candidates in the differential interferograms. Out of the 127 differential interferograms, we used 54 samples as landslide candidates (included in 51 differential interferograms). These samples were detected as possible candidates for locations where landslides could occur based on interference fringes detected on the slopes at multiple different times and on the existence of microtopography formed by landslides, as in the expert technique. Interference fringes representing landslide candidates are characterized by successive changes in color gradation as well as step-wise differences in the color gradation around the interference fringe [*Geospatial Information Authority of Japan*, 2011].

For others we sampled several locations at random from areas other than the landslide candidates from the 127 differential interferograms. Differential interferograms are expected to include interference fringes caused by various factors on ground surfaces with different gradients. Therefore, it is necessary to evenly prepare the model learning the features of the interference fringes of the slopes and flatlands. When the slopes and flatlands are defined to have an average gradient of 15 degrees or more or less than 15 degrees, respectively, slopes accounted for 37% and flatland accounted for 63% of the differential interferograms classified as others. The division between slope and flatland was set to 15 degrees (on a 1 km mesh) considering the lower limit value of the average relief energy of 1 km^2 in the mountainous area is 300 m [*Suzuki*, 2000a] and the value of the sedimentary topography due to debris flow e.g. the maximum inclination of sedimentary terrains such as alluvial fans belonging to flatlands is 15 degrees [*Suzuki*, 2000b]. We confirmed that the ratios of the slopes and the flatlands sampled at random in the others class were equivalent. Note that the elevation/gradient data of the 1-km mesh of the National Land Numerical Information archived by the *National Spatial Planning and Regional Policy Bureau, MLIT* [2011] was used for the average gradient.

For these training data, images were cut into small areas (150 × 150 pixels = 17.64 km^2 each) to be used for learning. The size of the small areas was set so that the landslide candidate interference fringes used for learning included those areas as well as the margins to include the information near the fringes.

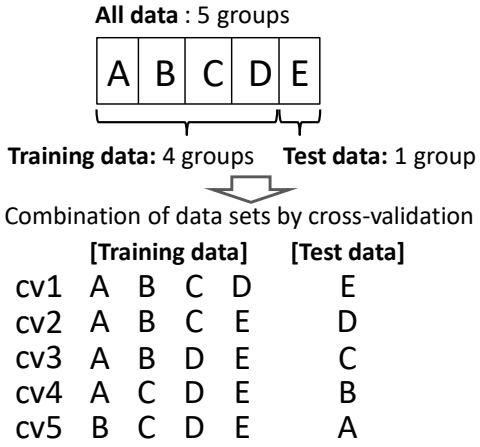

Fig. 5 The datasets used for cross-validation

2.4 Learning method and dataset

To evaluate the classification performance of the trained CNN model, it is necessary to divide the data for training and testing. However, the number of data in the landslide candidate class is small and the data for testing are limited; therefore, there was a concern that the validity of the validation result would be low. Accordingly, we used a k-fold cross-validation to maximize the use of the limited data and evaluate the generalization of the model. In this method, part of the data, which are divided into k pieces, is analyzed and the analysis result is verified using the remaining data. Then, the validity of the analysis is verified by averaging the k validation results. In this study, dividing the all data into five groups, four of which are used for training while the remaining group is used for testing, we conducted the training and validation five times in total by changing the combination of groups each time (cv1–cv5), as shown in **Fig. 5**. The number of divisions of data was determined taking into consideration the time required for validation.

In addition, to evaluate the performance of the model according to the difference in the number of training data, the 127 differential interferograms were divided into two groups, Dataset A and Dataset B. Dataset A consists of 107 differential interferograms and 42 landslide candidate samples. Dataset B consists of 127 differential interferograms (Dataset A plus 20) and 52 landslide candidate samples (Dataset A plus 12). Based on this, we prepared the training data (**Table 3**) and test data (**Table 4**) to perform the intersection validation five times. The number of landslide candidate data in the one-time training is 33 or 34 samples for Dataset A and 42–44 samples for Dataset B. Because there were few landslide candidate interference fringes for use as training data, there was a concern about the possibility of overlearning in the CNN model for the

Table 3 Number of samples for each class used for training

Class		cv1		cv2		cv3		cv4		cv5	
		Before padded	After padded	Before padded	After padded	Before padded	After padded	Before padded	After padded	Before padded	After padded
Dataset A	Landslide candidate	33	40,000	33	40,000	33	40,000	34	40,000	34	40,000
	Others	40,000	40,000	40,000	40,000	40,000	40,000	40,000	40,000	40,000	40,000
Dataset B	Landslide candidate	43	45,000	43	45,000	42	45,000	43	45,000	44	45,000
	Others	45,000	45,000	45,000	45,000	45,000	45,000	45,000	45,000	45,000	45,000

limited training data. It is generally desirable for the training data to represent the actual data space of the learning object. However, in this study, the number of landslide candidate interference fringes is small and the directions and positions in which the interference fringes appear are limited. Accordingly, we padded the numbers of training data, via rotation, reversal, and parallel movements, to 40,000 for Dataset A and to 5,000 for Dataset B so that the interference fringes in the interferograms have different directions and positions (**Table 3**). For rotation padding, the interferograms were rotated at random based on the rotation angle θ ($0° \leq \theta < 180°$). For reversal padding, they were vertically, horizontally, and vertically and horizontally reversed. For parallel movement padding, they were selected randomly within the scope of −10 to +10 pixels in the X-axial and Y-axial directions of images. The same numbers of other data and landslide candidate data were sampled.

In addition, test data were prepared from image data other than training data (Datasets A' and B' in **Table 4**).

We created CNN models that trained Dataset A (CNN-A) and another CNN model that trained Dataset B (CNN-B).

2.5 Evaluation index of the classification performance

The classification performance of the trained CNN models were evaluated using the following method. First, we evaluated the performance demonstrated when the sample data for testing (**Table 4**) was classified into each class using the trained CNN models. Next, from the differential interferograms including the samples of each class shown in **Table 4**, we cut out small area images of 150 × 150 pixels at intervals of 10 × 10 pixels and evaluated the results interpreted by the CNN models (**Fig. 1**(b)). When the classification result by the CNN models for the class of test data is correct, it is indicated as true positive (TP) and true negative (TN), and in case of error it is indicated as false positive (FP) and false negative (FN) (**Table 5**).

Table 4 Number of samples for each class used for testing

Class		cv1	cv2	cv3	cv4	cv5
Dataset A'	Landslide candidate	9	9	8	8	8
	Other	1,000	1,000	1,000	1,000	1,000
Dataset B'	Landslide candidate	11	11	11	11	11
	Other	1,500	1,500	1,500	1,500	1,500

Table 5 Evaluation index of the classification performance

		Class of test data	
		Landslide candidate	Other
Classification result	Landslide candidate	True Positive (TP)	False Positive (FP)
	Other	False Negative (FN))	True Negative (TN)

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

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

As indices for evaluating the classification performance, we used the recall, precision, and break even point (BEP). The recall value represents the ratio of the data classified by the CNN models as landside candidates to the data that are actually landside candidates (**Table 5**, Eq. (1)). The precision value represents the ratio of the data that are actually landside candidates to the data classified by the CNN models as landside candidates (**Table 5**, Eq. (2)). The recall and precision values are generally in a trade-off relationship; therefore, we used BEP to evaluate the classification performance when they are balanced. BEP is the value at which the recall equals the precision when the threshold to the probability value of the landside candidates class is raised or lowered. In this study, the recall and precision are calculated when the threshold value is varied in increments of 0.01 and the average value of the precision between the two points where the magnitude relationship between the recall and the precision is reversed is taken to be the BEP.

3. RESULT

3.1 Classification result of the test data

Table 6 shows the result of the classification of

Table 6 Evaluation result of classification performance of CNN models

Case	Model	Validation data	Evaluation Index	cv1	cv2	cv3	cv4	cv5	Average
1	CNN-A	Dataset A'	TP	448.4	504.5	504.0	441.0	504.0	–
			FN	56.1	0.0	0.0	63.0	0.0	–
			FP	2.0	1.5	1.0	0.0	1.5	–
			Recall	88.9	100	100	87.5	100	95.3
			Precision	99.6	99.7	99.8	100	99.7	99.8
			BEP	99.6	100	99.9	99.7	100	99.8
2	CNN-A	Dataset B'	TP	618.1	686.8	618.1	549.5	679.5	–
			FN	137.4	68.7	137.4	206.0	75.5	–
			FP	2.0	3.0	1.5	0.5	2.5	–
			Recall	81.8	90.9	81.8	72.7	90.0	83.5
			Precision	99.7	99.6	99.8	99.9	99.6	99.7
			BEP	95.8	98.5	91.9	99.3	96.8	96.3
3	CNN-B	Dataset B'	TP	618.1	755.5	618.1	686.8	755.0	–
			FN	137.4	0.0	137.4	68.7	0.0	–
			FP	3.5	6.0	2.5	2.0	5.0	–
			Recall	81.8	100	81.8	90.9	100	90.9
			Precision	99.4	99.2	99.6	99.7	99.3	99.5
			BEP	99.0	99.5	99.0	99.7	99.9	99.4

the test data in **Table 4** from models created using cv1–cv5 as training data in Datasets A and B in **Table 3**. Note that, because the numbers of test data between the classes are balanced in general classification evaluation, we conducted normalization (total number of test data/number of classes) for the evaluation result of the classification of the test data. Here, the validation was conducted for three cases with different combinations of the model and classification data. It is conceivable that the model CNN-B including the same training data as the model CNN-A learns features in a similar way. Validation against Dataset A' for the classification using model CNN-B was not performed because it is similar to Case 1 in **Table 6**.

The recall values of cv1–cv5 for each model range from 72.7% to 100%, the precision values range from 99.2% to 100%, the BEP values range from 91.9% to 100%, and the performance of cv1–cv5 varies. Because it is difficult to evaluate the detection method in this study in the validations of cv1–cv5, we believe that it is appropriate to take an average value of the five validation results for each model.

The average value of the five validations for each model and the classification was 95.3% for recall, 99.8% for precision, and 99.8% for BEP in Case 1 of **Table 6**. In Case 2, the recall value was 83.5%, the precision value was 99.7%, and the BEP value was 96.2%. In Case 3, the recall value was 90.9%, the precision value was 99.5%, and the BEP value was 99.4%; each of which was lower than the comparable values in Case 1.

Table 7 Evaluation result of detection performance of CNN model

(a) Size filtering not applied

Case	Model	Validation data	Evaluation Index	cv1	cv2	cv3	cv4	cv5	Average
1	CNN-A	Dataset A'	TP	9	9	8	8	10	–
			FN	0	0	0	0	0	–
			FP	319	291	284	288	222	–
			Recall	100	100	100	100	100	100
			Precision	2.7	3.0	2.7	2.7	4.3	3.1
2	CNN-A	Dataset B'	TP	10	10	9	10	11	–
			FN	1	1	2	1	1	–
			FP	344	357	375	495	308	–
			Recall	90.9	90.9	81.8	90.9	91.7	89.2
			Precision	2.8	2.7	2.3	2.0	3.4	2.7
3	CNN-B	Dataset B'	TP	9	9	8	7	10	–
			FN	0	0	0	1	0	–
			FP	700	658	677	320	573	–
			Recall	100	100	90.9	81.8	100	94.5
			Precision	1.4	1.4	1.3	1.7	1.6	1.5

(b) Size filtering applied

Case	Model	Validation data	Evaluation Index	cv1	cv2	cv3	cv4	cv5	Average
1	CNN-A	Dataset A'	TP	9	9	8	8	10	–
			FN	0	0	0	0	0	–
			FP	236	197	207	199	158	–
			Recall	100	100	100	100	100	100
			Precision	3.7	4.4	3.7	3.9	6.0	4.3
2	CNN-A	Dataset B'	TP	10	10	9	10	11	–
			FN	1	1	2	1	1	–
			FP	251	242	282	328	231	–
			Recall	90.9	90.9	81.8	90.9	91.7	89.2
			Precision	3.8	4.0	3.1	3.0	4.5	3.7
3	CNN-B	Dataset B'	TP	9	9	8	7	10	–
			FN	0	0	0	1	0	–
			FP	489	428	457	221	386	–
			Recall	100	100	90.9	81.8	100	94.5
			Precision	2.0	2.1	1.9	2.5	2.3	2.1

3.2 Result of the interpretation of the differential interferograms

Table 7 shows the result of the classification when the differential interferograms were interpreted using the trained CNN models. In Case 1 of **Table 7**(a), the recall value was 100% and the precision value was 3.1%, both of which were the highest values compared to the other cases. In Case 3, the recall value was 94.5%, higher than in Case 2, and the precision value was 1.5%, the lowest value of the precision seen.

Fig. 6 shows an example of an interpretation result. We created an interpretation result map considering that the center of each small area image with an area of 10 × 10 pixels corresponds to the interpretation result of the small area images. Many of the fringes detected as landslide candidates from the interpretation result (**Fig. 6** (b)), as detected by the CNN models, had an area smaller than landslide candidates used for the training data. For interference fringes with small areas, it is difficult even for experts to distinguish them as landslide candidates or as others.

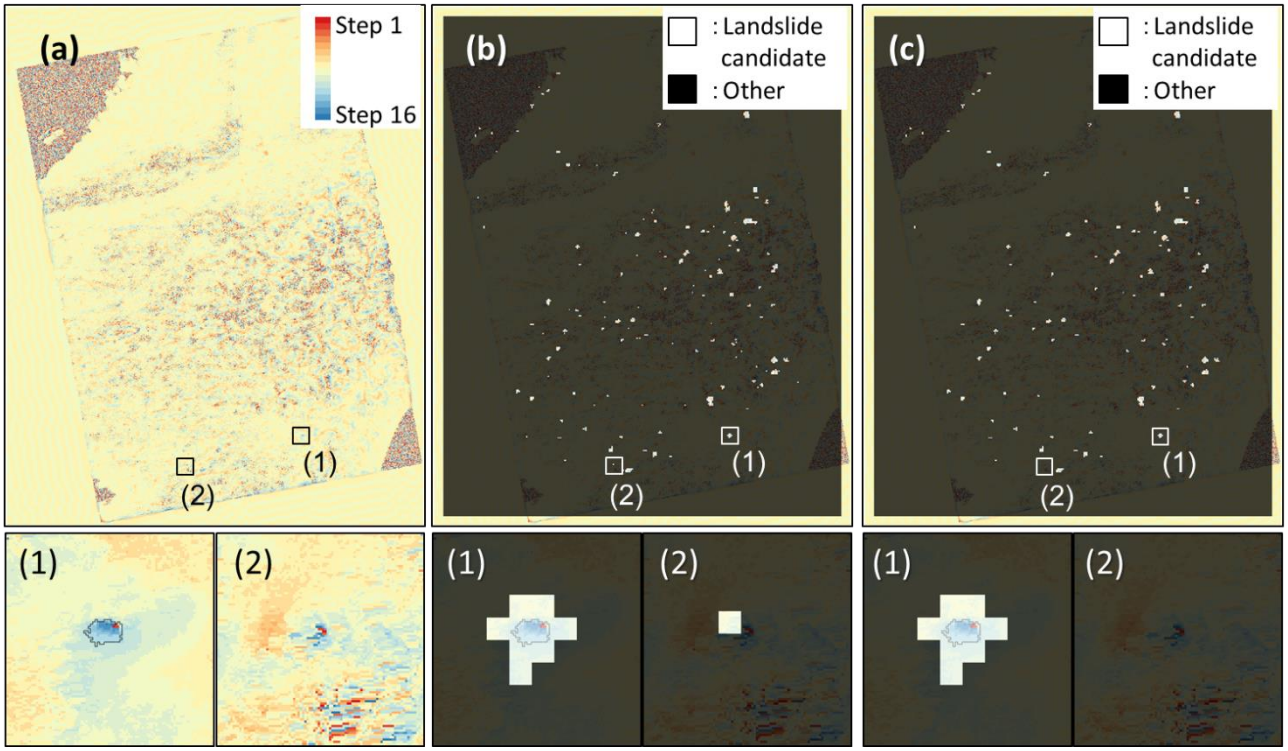

Fig. 6 An example of predicting map
(a) DInSAR imagery, (b) Predicting map (Not applied size filtering), (c) Predicting map (Applied size filtering)

Therefore, we provided size filtering to remove the landslide candidates of less than 10 × 10 pixels (approx. 0.078 km^2), which is less than the minimum area (0.16 km^2) of landslide candidates in the training data, from the interpretation result of the CNN model. The index value and predicting map in this case is shown in **Table 7**(b) and **Fig. 6** (c), respectively. As a result, in comparison with the case where no size filtering is applied (**Table 7(a)**), the recall value was the same but the precision value slightly improved for all cases. Because the filter size is smaller than the minimum area of the landslide candidates used for the training data, the number of true positives in Eq. (2) (**Table 5**) does not change. Meanwhile, interference fringes with a small area included in others were rejected by the filter and the number of false positives was reduced, so the precision improved.

4. DISCUSSION

4.1 Comparison of the validation result for each case

As a result of the classification of the test data in **Table 6**, others and landslide candidates were classified accurately in each case.

On the other hand, the interpretation result of **Table 7** is a high recall but low precision. Comparing the results of each case in **Table 7**; the value of the recall is lower in Case 2 than in Cases 1 and 3. **Fig. 7** shows samples of landslide candidates included in the dataset used for training the CNN models. Panels (a) and (b) in **Fig. 7** are included in Datasets A and B but panels (c) and (d) are included only in Dataset B. In Case 2, Dataset B' includes samples of landslide candidates having characteristics different from the training data (Dataset A) used for modeling (**Figs.7** (c) and (d)). Therefore, these samples could not be detected. In Case 3, the value of the precision was the lowest. Samples of landslide candidates in Dataset B include some patterns of interference fringes that are difficult to distinguish from those in others (**Figs.7** (c) and (d)). This likely caused misdetections of interference fringes in the others class. Even though in Case 3 a model with more training data than in

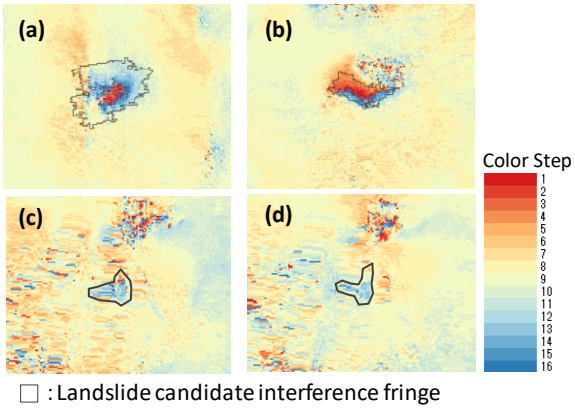

Fig. 7 Sample of Landslide candidate: (a) and (b) Dataset A, (c) and (d) Dataset B

Case 1 was created, the characteristics of the added landslide candidate interference fringes, rather than the increased amount of training data, might have affected the image classification performance.

4.2 Detection trends of models

Landslide candidate samples in the test data of Case 3 in **Table 7** have one or two candidate sites in a single differential interferogram. Conversely, in the interpretation result of Case 3, landslide candidate interference fringes detected by the CNN-B model averaged 46 sites in a single differential interferogram. In these areas, the ratios of the slopes, which were defined as 15 degrees or more in gradient, and the flatlands, which were defined as less than 15 degrees were 43% (approximately 20 sites) and 56% (approximately 26 sites), respectively. Because the ratios of the slopes and flatlands in the test data were 33% and 67%, respectively, the detection ratios of the interference fringes per unit area are higher on slopes.

For the sites misdetected on the flatlands and slopes, it is difficult to visually detect clear differences in the patterns of their respective interference fringes. However, compared to the flatlands, differences are more likely to appear on slopes where continuous interference fringes recognized as landslide candidates are characterized by step differences from neighboring areas. In addition, another cause could be the misdetection of noise as landslide candidate interference fringes; noise occurs because the coherence is lower on the slopes than on the flatlands.

5. CONCLUSIONS

In this study, we verified the possibility of detecting landslides from differential interferograms using CNN models. As a result, it was found that CNN models can detect interference fringes having the possibility of landslides with high reproducibility with recall values of approximately 90% in the validation case. However, landslide candidate interference fringes were detected in numbers larger than that of the training data, which shows that the precision is low. The image classification performance of this CNN model might be affected by the patterns of the landslide candidate interference fringes in the training data. In addition, interference fringes detected in the interpretation by the CNN model are more likely to be detected on slopes than on flatlands.

Therefore, CNN models with high recall values may be effective as a means of detecting interference fringes with the possibility of landslides from differential interferograms regularly observed over a wide area. However, the misdetection of noise is likely to occur on slopes because many landslide candidates are detected on slopes, which have lower coherences than flatlands. Therefore, it is necessary to further narrow down the interference fringes that are likely to be landslide movements from the landslide candidates detected by the CNN models based on features of the topography and geology related to landslides. Then, this model will be able to contribute to the efficient risk management of sediment disasters by prioritizing sites where on-site surveys and observations should be conducted.

However, because this study did not have sufficient training data for landslide candidates, it is difficult to say that a CNN models with high generalization have been established. Therefore, to improve the model performance, training data on interference fringes with various characteristics need to be increased. However, it is not easy to prepare a large amount of interference fringes with the possibility of landslides as training data. In this study, only differential interferograms were used as training data. However, it is sometimes difficult to detect interference fringes with the possibility of landslides only from differential interferograms. In reality, when experts detect interference fringes with the possibility of landslides, they also consider the topography and other characteristics of the site. Therefore, the use of other data such as topography for learning in addition to differential interferograms may enable more accurate detections of sites with the possibility of landslides.

Our challenge for the future is to study combinations of training data and learning methods appropriate for the detection of sites with the possibility of landslides and therefore improve the accuracy of the CNN models.

ACKNOWLEDGMENT: The authors appreciate JAXA (Japan Aerospace Exploration Agency) for provision of ALOS/PALSAR data. ALOS/PALSAR data belong to JAXA and METI (Ministry of Economy, Trade and Industry).

REFERENCES

David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Williams (1986): Learning representations by back-propagating errors, Nature, No.323, pp.533-536

Geospatial Information Authority of Japan (2011): Manual for Interpretation of SAR Interference Images for Landslide Monitoring, http://www.gsi.go.jp/chirijoho/chirijoho40089.html (accessed November 15 2017)

Ishii, T., Nakamura, R., Nakada, H., Mochizuki., Y. and Ishikawa, H. (2015): Surface object recognition with CNN and SVM in Landsat 8 images, MVA2015 IAPR International Conference on Machine Vision Applications, pp.341-344.

Japan Aerospace Exploration Agenc HP: http://www.satnavi.jaxa.jp/project/alos4/ (accessed November 15 2017)

Kamiyama, J., Kunitomo, M., Sakagami, M., Nagai, Y., Hirata, I., Niahikawa, D., Yoshikawa, K. and Sango, D. (2016): A Study of Risk Assessment Method for Landslides etc. with InSAR Analysis, Proceedings of Research Meeting of Japan Society of Erosion Control Engineering, pp.B-484-485 (in Japanese)

Kamiyama, J., Noro, T., Sakagami, M., Mantoku, M., Hirata, I., Yoshikawa, K. and Sango, D. (2017): Verification of Applicability of the Large-scale Sediment Movement Risk Assessment Method by InSAR Analysis, Proceedings of Research Meeting of Japan Society of Erosion Control Engineering, pp.198-199 (in Japanese)

Krizhevsky, A., Sutskerver, I. and Hinton, G.E. (2012): ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems 25, pp.1106-1114.

Kusano, S., Sango, D., Yamanoguchi, T. and Shimada, M. (2015): A spatial filter adaptive to slope size applied to differential SAR interferograms for landslide detection, Japan Geoscience Union Meeting 2015, STT54-P-04 (in Japanese)

LeCun, Y., Bengio, Y. and Hinton, G. (2015): Deep learning, Nature, Vol.521, pp.436-444

National Spatial Planning and Regional Policy Bureau, MLIT of Japan (2011): elevation / gradient third mesh data of the National Land Numerical Information, http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-G04-a.html (accessed November 15 2017)

Okatani, T. (2016): Recent Research Trends on Deep Learning and Application Thereof to Image Recognition, http://acsi.hpcc.jp/2016/download/ACSI2016-tutorial2.pdf (accessed November 15 2017)

Shimada, M. (1999): Correction of the Satellite's State Vector and the Atmospheric Excess Path Delay in the SAR Interferometry - An Application to Surface Deformation Detection, Journal of the Geodetic Society of Japan, Vol.45 , No.4 , pp.327-346. (in Japanese with English abstract).

Suzuki, R. (2000 a) Chapter 12 General characteristics of hills and mountains, “Introduction to topographic map interpretation for construction engineers” volume 3 Terraces/ Hills/ Mountains, p.651-684, Kokonshoin

Suzuki, R. (2000 b) Chapter 15 15.6 Debris flow topography, “Introduction to topographic map interpretation for construction engineers” volume 3 Terraces/ Hills/ Mountains, p.849-860, Kokonshoin

Wilmanski, M., Krucher, C. and Lauer, J. (2016): Modern Approaches in Deep Learning for SAR ATR, SPIE 9843, Algorithms for Synthetic Aperture Radar Imagery XXIII, 98430N

Y. Zhou, H. Wang, F. Xu, and Y. Q. Jin (2016): Polarimetric SAR image classification using deep convolutional neural networks, IEEE Geoscience and Remote Sensing Letters, Vol.13, No.12, pp.1935-1939