Landslide Classification from Synthetic Aperture Radar Images Using Convolutional Neural Network with Multichannel Information

Shingo Mabu, Yoshiaki Nakayama and Takashi Kuremoto

Graduate School of Sciences and Technology for Innovation, Yamaguchi University
2-16-1 Tokiwadai, Ube, Yamaguchi 755-8611, Japan
E-mail: mabu@yamaguchi-u.ac.jp

Abstract Detection of disaster-stricken areas using synthetic aperture radar (SAR) images is important in countries and regions with heavy rain and earthquakes. Although it is important to immediately find disaster-stricken areas when a disaster occurs, it takes time to read SAR images and also needs experience and expertise. Therefore, machine learning, especially deep learning, is expected to be applied to the classification of disaster-stricken areas. Classification using deep learning is often executed on patch images of local areas. However, patch-based classification would miss information on the surrounding areas such as topographic features. In this study, a convolutional neural network (CNN) is applied to the classification of SAR images using the following techniques. When making the images input to a CNN, two multichannel image generation methods, i.e., a zero-padding method and map-concatenation method, are used, where the target areas to be classified and their surrounding areas are combined to form multichannel images. In the experiments, the zero-padding method and map-concatenation method are evaluated by the classification performance of SAR images that cover the northern Kyushu area in Japan, where large-scale landslides due to heavy rain occurred in 2017. Through the experiments, we clarify the appropriate CNN structures with multichannel information for landslide classification.

Keywords: convolutional neural network, classification, multichannel image, SAR image

1. Introduction

When large-scale disasters simultaneously occur in multiple and wide areas, aerial photographs [1] have been used to find disaster-stricken areas. However, it takes time to take photographs of wide areas, which makes the rapid detection of disaster areas difficult. Therefore, satellite remote sensing, which can observe a wide area, has recently attracted attention. It is important for the countries and regions where natural disasters frequently occur to detect disaster areas quickly. In Japan, for example, typhoons and torrential rain occur every year, and large-scale disasters such as the metropolitan-area earthquake and the Nankai Trough earthquake are imminent; thus, detecting disaster-stricken areas by satellite remote sensing is extremely important. Earthquakes and torrential rain cause landslides, but finding the areas of landslides is difficult because they often occur in areas that are not easily accessed.

Research on landslide detection using satellite remote sensing has been actively conducted [2]–[4]. Many methods detect landslides on the basis of the amount of change in the normalized difference vegetation index (NDVI) calculated from optical satellite images before and after the disaster [5], [6]. NDVI
is calculated using the degree of reflection of visible and near-infrared rays. In [5], an automated object-oriented and multitemporal landslide detection approach was proposed to realize landslide hazard and risk assessment. Landslides are automatically identified on the basis of NDVI trajectories. In [6], the information obtained by two satellites, Landsat-8 and Advanced Land Observing Satellite No. 2 (ALOS-2), was combined to make damage maps after a disaster occurs. This method uses a modified normalized difference water index (MNDWI) and NDVI to detect the areas damaged by an earthquake that occurred in Japan. In [7], Landsat images and a predictive model of E30 [8] were used to find the soil erosion and the loss of agricultural land in the period from 2000 to 2014 in the Jiadhal basin in northeast India. This study showed the effectiveness of remote sensing in gathering spatial information to find areas needing immediate attention.

Although optical satellite images can cover wide areas, the information of the ground cannot be observed correctly at night and in bad weather. Therefore, finding disaster-stricken areas from synthetic aperture radar (SAR) images that can be taken without being affected by the time and weather has attracted attention. Disaster areas are usually detected by measuring temporal differences between SAR images. In [9], damage assessment methods that use multitemporal SAR were reviewed, where some methods use coherence and intensity correlation, and some methods use combinations of SAR and additional data such as geographic information system (GIS) data and optical imagery. However, since it is more difficult to read SAR images than optical images, considerable knowledge and experience are necessary to design SAR image processing systems. For example, the types of features useful for detecting disaster areas depends on the designers’ skills.

Therefore, deep learning has recently been applied to remote sensing [10] including disaster area classification to realize end-to-end learning including feature extraction and classification. In [11], a deep belief network [12] was used as a feature extraction method, and an iterative algorithm to select effective features for classification was proposed. This method was applied to a classification problem of seven classes with 2800 remote sensing images. In [13], a deep autoencoder was used to learn high-level features and the representation of remote sensing images. A preprocessing method using wavelet transformation and a corrupting and denoising method were designed to enhance the robustness for recognizing landslide features.

In this study, we investigate the ability of a convolutional neural network (CNN) enhanced by multichannel images to classify normal and abnormal (disaster) areas. In detail, the proposed method aims to find landslide areas in SAR images only taken after a disaster, where the SAR images are divided into small patch images to carry out patch-based classification. Patch-based classification is useful when we aim to identify the points of landslide areas. However, there is a problem in the patch-based classification, that is, the CNN only uses the local information of SAR images and does not have information on the surrounding areas, e.g., topographic features. Therefore, we apply some methods that combine the patch image to be classified and the images of its surrounding areas to form multichannel images. When combining patch images and surrounding area images, two newly designed techniques are used: one is a zero-padding method and the other is a map-concatenation method. Through the experimental results, we focus on the effectiveness of the two methods by comparing some single-channel and multichannel methods.

The contributions of this paper are as follows. 1) The proposed method uses not only the target area images to be classified but also their surrounding area images as context, and combines them to generate multichannel images for accurate classification. 2) Several kinds of multichannel CNN structures are evaluated and the appropriate structures for landslide classification of SAR images are clarified.

A CNN has already been applied to SAR image analysis. In [14], a new type of CNN with sparsely connected layers was proposed to adapt to a limited number of training images, where all the layers consist of convolution layers and fully connected layers are not used. In [15], a transfer-learning-based CNN was proposed, where the CNN consists of a reconstruction pathway for pretraining the network by unsupervised learning and a classification pathway for adapting to the target task by supervised learning. These meth-
ods were applied to the classification of patch images of SAR provided by the Moving and Stationary Target Acquisition and Recognition (MSTAR) public dataset [16]. However, there is a limitation in these methods, that is, both methods are designed to use patch images only and do not consider information on the surrounding area, making them unsuitable for disaster area detection. When detecting disaster areas, human experts interpret SAR images focusing on not only local areas but also on the whole images; therefore, here we design architectures that consider local areas and their surrounding wide areas for classification.

This paper is organized as follows. Section 2 describes the mechanism and features of SAR. Section 3 reviews the basic structure of the CNN and the criteria for evaluating classifiers. Section 4 describes the classification methods using the CNN with multichannel images. Section 5 gives the experimental results and Sect. 6 provides a discussion. Section 7 is devoted to conclusions.

2. Synthetic Aperture Radar (SAR)

SAR is an active image radar that synthesizes small antennas mounted on a platform, such as an aircraft or satellite, to realize large virtual antennas and generates high-resolution radar images [17], [18]. Because SAR is an active sensor that emits microwaves, it is possible to observe the surface of the earth regardless of the presence or absence of sunlight and clouds. SAR images are applied to the research fields of agriculture, disaster, oceans, earth science, and so on [19]–[21].

SAR emits microwaves and receives the reflected microwaves from the surface of the earth (Fig. 1). When a microwave emitted from the SAR antenna enters a conductor or dielectric, a current is induced and the microwave is reemitted from the induced current. This is called scattering, and scattering in the opposite direction of the incident wave is called backscattering. Because backscattering is the diffuse reflection caused by scattering, backscattering is different from specular reflection. SAR receives the backscattering and executes image reproduction. The scattering intensity of the microwaves strongly depends on the frequency, wavelength and electric characteristics (dielectric constant, etc.) of the scatterer. Therefore, for example, seawater or cars made of metal strongly reflect microwaves because a current is easily induced. On the other hand, sand and trees have low reflectivity because hardly any current is induced. In each pixel of SAR images, the intensity of the received microwave is recorded.

3. Convolutional Neural Network (CNN)

In this section, we firstly review the basic structure of a CNN and mini-batch learning, whose concepts are used to build classifiers in Sect. 4. Then, the SAR images used for the experiments and some evaluation criteria are introduced.

3.1 Basic structure of CNN

A neural network was proposed as an algorithm that imitates complicated brain functions and has been applied to solve various real-world problems. The CNN was proposed on the basis of the visual cortex of living organisms obtained from neuroscience as a means of recognizing objects in images. To recognize objects, the CNN transforms pixel values of images into feature values using multiple processing layers. The first research on applying the functions of the visual cortex to pattern recognition was the neocognitron [22], and the direct origin of the CNN is LeNet[23], [24]. An example of a CNN structure is shown in Fig. 2. The CNN consists mainly of three types of layers: a convolution layer, a pooling layer.
and a fully connected layer. A convolution layer executes the filtering process on the input image. The filtering process extracts features that are useful for recognizing objects in images. A pooling layer is basically placed immediately after the convolution layer, which decreases the position sensitivity of the features obtained by the convolution layer. The features extracted by multiple convolution and pooling layers are integrated and recognized by the fully connected layer(s) [25].

### 3.2 Mini-batch learning

There are several methods to prepare a set of input samples for the learning of neural networks. One of them is sequential learning (stochastic gradient descent) [26], which calculates an error for one sample and updates the internal parameters using the backpropagation algorithm [27]. Since sequential learning updates for one sample, the magnitude of parameter changes strongly depends on the samples. If there is a sample containing noise, the magnitude of parameter changes will be large and the error may not converge. On the other hand, batch learning is a method of using all the samples to calculate the error. Although it is robust against noise, it takes a longer time to update parameters once; and moreover, the parameters will converge to the local optimum.

Mini-batch learning is an intermediate learning method between sequential and batch learning, which makes a set of a small number of samples and updates the parameters using the error for the small set. The number of samples contained in a mini-batch is predetermined, but this value affects the learning results. Although there is no systematic method to determine the optimum number of samples in a mini-batch, after considering parallel computing resources, convergence and errors, etc., we set the mini-batch size at 30 in this paper.

### 3.3 SAR image and evaluation criteria

Our aim is to classify landslide areas by using a CNN with some techniques. A SAR image for classification is shown in Fig. 3. It shows the northern Kyushu area in Japan on July 7, 2017, taken by ALOS-2. Large-scale landslides due to the torrential rain from July 5 to 6, 2017, occurred in this area. The red marks in Fig. 3 show the locations of the landslide areas that were annotated by the Geospatial Information Authority of Japan [28]. The size of the image is $6648 \times 4360$ [pixels] and the resolution per pixel is about $3 \text{ m}^2$.

The input image for the CNN is the original SAR image without the red marks, and the image is divided into patch images ($8 \times 8$ or $32 \times 32$ [pixels] in this paper). Figure 4 shows sample $32 \times 32$ patch images of normal and abnormal areas. Since it is difficult to distinguish normal and abnormal images by just observing the images, machine learning methods are important for quickly finding the abnormal areas. The CNN learns to classify each patch image as a normal (negative) or abnormal (landslide, positive) area. The CNN is trained by using training samples and the classification performance is evaluated by using testing samples that are not contained in the training samples. The confusion matrix is a typical way of evaluating classification performance, which is a table that counts the number of correct and incorrect classifications for each class. Table 1 shows the general form of the confusion matrix, where, for example, when the classifier predicts a sample as a positive class but the correct answer is negative, the number of false positives increases by one. This paper uses accuracy, precision, recall and F-measure based on the confusion matrix for the evaluation.

- **Accuracy**

  Accuracy (ACC) is the proportion of the correct classifications of the total number of testing
Table 1  Confusion matrix

| Correct class | Predicted Class | Positive | Negative |
|---------------|-----------------|----------|----------|
| Positive      | True Positive (TP) | False Negative (FN) |
| Negative      | False Positive (FP) | True Negative (TN) |

Fig. 3  SAR image (Northern Kyushu area, Japan on July 7, 2017)

Fig. 4  Sample 32 $\times$ 32 [pixels] patch images of normal and abnormal areas

Recall

Recall is the proportion of true positives of the number of positive samples (Eq. (3)).

$$Recall = \frac{TP}{TP + FN}$$

Precision

Precision is the proportion of true positives of the number of samples classified as positive (Eq. (2)).

$$\text{Precision} = \frac{TP}{TP + FP}$$

F-measure

F-measure evaluates the balance between precision and recall, which are in a trade-off relationship (Eq. (4)).

$$F-measure = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision}$$

4. Landslide Classification Using CNN with Multichannel Information

First, let us consider the balance between the numbers of patch images containing normal and abnormal areas. In fact, most of the patch images are usually normal even when a disaster occurs. Therefore, if the CNN is trained using the patch images, the
classification results tend to be normal and the classification of abnormal images cannot be done appropriately. Therefore, the following mini-batch generation method is used to maintain the balance between the numbers of normal and abnormal images used for training the CNN.

As shown in Fig. 5, let the number of normal samples in the original training set $C_1$ be larger than the number of abnormal ones $C_2$. Then, $C_2$ normal samples are randomly selected by sampling without replacement. Each mini-batch is generated by selecting $C_2/M$ samples each from the normal and abnormal samples. Finally, $M$ mini-batches are obtained. The above procedure is executed every epoch and repeated until the predefined number of epochs is reached.

Next, two multichannel image generation methods are designed for the CNN to improve classification accuracy. Figure 6 shows the procedure of multichannel image generation where the target area to be classified and its surrounding areas are cropped and combined to make multichannel images. However, the sizes of the cropped images have to be the same to combine them; therefore, we propose two methods: a zero-padding method and a map-concatenation method, and evaluate their performance in Sect. 5.

In the zero-padding method, zero-padding (adding pixel values of zero) is applied to the smaller images to make their size the same as the largest one. Figure 7 shows an example of the zero-padding method where there is one target area to be classified and two surrounding areas. These three images have different sizes; therefore, zero-padding is applied to the smaller images to enlarge them and the three images are combined to obtain a three-channel image. The generated multichannel images become the input to the CNN.

In the map-concatenation method, convolution and pooling functions are applied to the images to transform them into feature maps with the same size. Figure 8 shows an example of the map-concatenation method where one target area and two surrounding areas are used. This method realizes both multichannel image generation and classification in one CNN. Each image of the target and surrounding areas is transformed to feature maps by repeating the convolution and pooling functions of the CNN. After the sizes of the feature maps become the same, all the feature maps are concatenated to obtain a multichannel image. The generated multichannel image is classified by the subsequent convolution, pooling and fully connected layers. Different from the zero-padding method, the map-concatenation method not only adjusts the size of the images but also extracts feature maps, which are important for classifying normal and abnormal images.

The reason why we propose the two methods is to compare simple and complex mechanisms. If the zero-padding method is better than the map-concatenation
method, it shows that the simple combination of the patch images is enough for disaster area classification, whereas, if the map-concatenation method is better, it shows that the feature extraction from each patch image and their combination contribute to better classification. The detailed CNN structures used in the experiments are explained in Sect. 5.2.

5. Experimental Results

The performance of landslide classification is evaluated in this section. In Sect. 5.1, CNNs with and without mini-batch generation method are compared. In Sect. 5.2, CNNs with single-channel, multichannel (zero-padding) and multichannel (map-concatenation) information are compared.

5.1 Experiment 1

The SAR image in Fig. 3 was divided into two halves; the right side was used as the training area and the left side was used as the testing area. The training and testing areas were further divided into patch images with $32 \times 32$ [pixels]. When 40% or more of the pixels of a patch image cover disaster areas, the image is regarded as an abnormal sample; otherwise, it is regarded as a normal sample. Table 2 shows the numbers of patch images belonging to each class extracted from the training and testing areas. Note that the numbers of normal and abnormal samples in the testing data are set to be the same (= 183) by selecting normal samples randomly. In this experiment, methods both with and without mini-batch generation used exactly the same number of training samples including positive and negative samples with the only difference being the way of making mini-batch samples for training the CNN.

The CNN structure is shown in Fig. 9, where “conv” is a convolution layer, “pool” is a pooling layer and “fc” is a fully connected layer. The kernel size used in a convolution layer is set at three, the padding size is one and the stride size is one. The kernel size

Table 2 Numbers of patch images (Experiment 1): ( ) shows the number of testing samples randomly selected for evaluation.

|       | Training | Testing |
|-------|----------|---------|
| Normal| 13758    | 13825 (183) |
| Abnormal| 250     | 183     |
Table 3 Comparison between CNNs with and without mini-batch generation

|                       | Accuracy | Precision | Recall | F-measure |
|-----------------------|----------|-----------|--------|-----------|
| CNN without mini-batch generation | 0.500    | 0         | 0      | 0         |
| CNN with mini-batch generation       | 0.671    | 0.688     | 0.650  | 0.669     |

used in a pooling layer is set at two, the padding size is zero and the stride size is two. The number of units in the fully connected layer is set at 16.

Table 3 shows the accuracy, precision, recall and F-measure obtained by the methods with and without mini-batch generation. Each value is the average over five independent trials. The method without mini-batch generation makes a mini-batch by selecting training samples randomly. From Table 3, we can see that the method without mini-batch generation cannot classify abnormal samples at all owing to the large bias in the number of samples. CNN without mini-batch generation classifies all the positive (abnormal) samples as negative; thus, precision and recall in Table 3 are both zero.

There are many ways to adapt to the class imbalance of the number of data. One of the ways to prepare training samples is to randomly select samples after equalizing the numbers of positive and negative samples. In this experiment, among several methods used to adapt to the class imbalance of the number of data, we used the mini-batch generation method and confirmed its effectiveness as preparation for Experiment 2. From the experimental results, we found that the mini-batch generation method adapts to the class imbalance in SAR image classification. Hereafter, all the methods are trained by the mini-batch generation method.

5.2 Experiment 2

Next, we compare multichannel CNN (zero-padding), multichannel CNN (map-concatenation), and two types of single-channel CNNs. Hereafter, multichannel CNN (zero-padding) is referred to as Multi-CNN1, multichannel CNN (map-concatenation) is referred to as Multi-CNN2 and the two types of single-channel CNNs are referred to as single CNN1 and single CNN2. The difference between single CNN1 and single CNN2 is the input image size (8 × 8 or 128 × 128 [pixels]), which is explained later. Actually, single CNN1 and single CNN2 are based on typical CNN architectures. Both single CNN1 and single CNN2 consist of several convolution and pooling layers followed by a fully connected layer. Therefore, single CNN1 and single CNN2 are regarded as conventional methods. For the Multi-CNNs, three-channel and five-channel images are also compared to find better settings.

As in Sect. 5.1, the right half of the SAR image was used for the training and the left half was used for the testing. Each method in this experiment aimed to classify 8 × 8 [pixels] patch images. Table 4 shows the patch image sizes for each channel. For example, single CNN1 used 8 × 8 [pixels] images and single CNN2 used 128 × 128 [pixels] images as one-channel inputs. The reason why single CNN2 used images with 128 × 128 [pixels] is to receive the same amount of information as the Multi-CNNs for a fair comparison. In detail, the first channel image for the Multi-CNNs is 8 × 8 [pixels], but the images of the third and fifth channels are 128 × 128 [pixels]. Therefore, 128 × 128 [pixels] images were also given to single CNN2. Note that, although single CNN2 used 128 × 128 [pixels] images as inputs, the aim of the classification is to give normal or abnormal labels to the central 8 × 8 [pixels] areas.

Table 5 shows the numbers of patch images belonging to each class extracted from the training and testing areas. As in Sect. 5.1, the testing data contains the same numbers of normal and abnormal samples (= 5804), chosen by selecting normal samples randomly.

The structures of single CNN2, Multi-CNN1 (3ch) and Multi-CNN1 (5ch) are the same as that in Fig. 9, the structure of single CNN1 is shown in Fig. 10, that of Multi-CNN2 (3ch) is shown in Fig. 11, and that of Multi-CNN2 (5ch) is shown in Fig. 12. All the
Table 4  Patch image sizes used in the conventional and proposed methods in experiment 2: Multi-CNN1 employs the zero-padding method and Multi-CNN2 employs the map-concatenation method.

| Image size [pixel] | 1st channel | 2nd channel | 3rd channel | 4th channel | 5th channel |
|-------------------|-------------|-------------|-------------|-------------|-------------|
| single CNN1       | 8 × 8       | —           | —           | —           | —           |
| single CNN2       | 128 × 128   | —           | —           | —           | —           |
| Multi-CNN1 (3ch)  | 8 × 8       | 64 × 64     | 128 × 128   | —           | —           |
| Multi-CNN1 (5ch)  | 8 × 8       | 32 × 32     | 64 × 64     | 96 × 96     | 128 × 128   |
| Multi-CNN2 (3ch)  | 8 × 8       | 64 × 64     | 128 × 128   | —           | —           |
| Multi-CNN2 (5ch)  | 8 × 8       | 32 × 32     | 64 × 64     | 96 × 96     | 128 × 128   |

Table 5  Numbers of patch images (Experiment 2): ( ) shows the number of testing samples randomly selected for evaluation.

|        | Training | Testing |
|--------|----------|---------|
| Normal | 204188   | 206196  | (5804)  |
| Abnormal | 7812   | 5804    |

structures were determined by considering the balance between the input image sizes, the complexity of the networks and the learning efficiency, and showed the best results after some settings were examined. The kernel size used in each convolution layer was set at three, the padding size was one and the stride size was one. The kernel size used in each pooling layer was set at two, the padding size was zero and the stride size was two. The number of units used in the fully connected layer in each method was set at as follows: single CNN1: 16, {single CNN2, Multi-CNN1 (3ch), Multi-CNN1 (5ch), Multi-CNN2 (3ch), Multi-CNN2 (5ch)}: 256, Multi-CNN2 (5ch): 64. “concat” in Figs. 11 and 12 executes the concatenation of all the feature maps. However, since all the feature maps have to be the same size for concatenation, one “zero-padding” process is included in Fig. 12 to precisely adjust the feature map size.

Table 6 shows the accuracy, precision, recall and F-measure obtained by single-channel and multichannel CNNs at the final epoch. The bold values show the highest value for each measure among all the methods. From Table 6, Multi-CNN2 (3ch) and Multi-CNN2 (5ch) show the best F-measure, and Multi-CNN2 (5ch) also shows the best accuracy and recall.

Figure 13 shows the transition of the testing accuracy obtained by each method in the training phase. Since more information is given to single CNN2 than to single CNN1, single CNN2 shows better accuracy than single CNN1, although the target areas for the classification are the same, i.e., 8 × 8 [pixels] areas. From this result, it can be concluded that the information on the surrounding areas is important for the classification. When single CNN2 is compared with the Multi-CNNs, the Multi-CNNs show better accuracy. In particular, Multi-CNN2 (5ch) shows the best accuracy at the 50th epoch. In summary, it is clarified that the proposed methods (Multi-CNNs) are better than single CNN1 and single CNN2, i.e., the typical CNN architectures.

Figure 14 shows the transition of the training loss (error) obtained by each method. The training loss gradually decreases as the number of epochs increases, and Multi-CNN1 (5ch) and (3ch) show the lowest and second-lowest training losses at the 50th epoch, respectively. However, Fig. 15, which shows the transition of the testing loss, indicates that Multi-CNN1 (5ch) and (3ch) cause overfitting and increase the
Table 6  Comparison between the single-channel and multichannel CNNs

|                  | Accuracy | Precision | Recall | F-measure |
|------------------|----------|-----------|--------|-----------|
| single CNN1      | 0.502    | 0.302     | 0.503  | 0.377     |
| single CNN2      | 0.709    | 0.664     | 0.730  | 0.696     |
| Multi-CNN1 (3ch) | 0.721    | 0.643     | 0.761  | 0.697     |
| Multi-CNN1 (5ch) | 0.727    | 0.672     | 0.756  | 0.711     |
| Multi-CNN2 (3ch) | 0.731    | **0.689** | 0.752  | **0.719** |
| Multi-CNN2 (5ch) |          | 0.665     | **0.782** | **0.719** |

losses after around the 25th epoch. On the other hand, the testing losses of Multi-CNN2 (3ch) and (5ch) do not increase after the 25th epoch, in contrast to Multi-CNN1, which means that Multi-CNN2 reduces the effect of overfitting.

6. Discussion

In this section, we discuss the results obtained in Sect. 5.2. As confirmed before, the information on the surrounding areas is useful to accurately classify normal and abnormal areas. When human experts read SAR images to find the abnormal areas, they not only observe the target areas to be classified but also refer to the surrounding areas, that is, they consider the context. Therefore, the input information containing information on surrounding areas has the effect of giving context to the CNN. When comparing single CNN2 and Multi-CNNs, the given information is substantially the same, that is, all the methods use
128 × 128 [pixels] images. However, Multi-CNNs show better accuracy. This result indicates that dividing a large amount of information into multiple small pieces of information is better than giving the large amount of information as it is. This is a kind of divide-and-conquer [29], [30] where many features are extracted from various local and global viewpoints, and combined to provide a set of useful information.

It is found in Fig. 13 that Multi-CNN1 (5ch) with zero-padding shows the highest accuracy at around the 25th epoch. However, the accuracy gradually decreases at the later epochs. Multi-CNN2 (5ch) with map-concatenation shows the highest accuracy at the 50th epoch without large overfitting. When making multichannel images in Multi-CNN1, unnecessary parts are filled with zeros; thus, the simple structure of Multi-CNN1 shows better accuracy than Multi-CNN2 at early epochs. However, at the later epochs, the simple multichannel images with zero-padding are over-trained, which decreases the classification accuracy for the testing samples. Multi-CNN2 does not use zero-padding to adjust the image sizes (except for one part in Fig. 12), but uses convolution and pooling functions. That is, Multi-CNN2 makes efficient use of the given inputs to extract features and combines them to make a multichannel feature map.

There are some techniques to avoid overfitting such as dropout. However, because the main objective of this paper is to evaluate the basic classification ability of some multichannel CNN structures, the combination of multichannel CNNs and the dropout technique was not used. Of course, if a method such as dropout is applied to the proposed method, it might improve accuracy. Therefore, when the proposed method is applied to an actual system in the future, appropriate combinations of the proposed method and the conventional techniques should be employed to obtain good performance.

There are remaining problems to be solved for further improvements. 1) Classification accuracy and other measures should be further improved. In particular, in the disaster area classification, it is important to improve recall so as not to miss disaster areas. In this paper, a patch image containing 40% or more of a disaster area is regarded as abnormal; thus, some samples containing around 40% disaster area might be misclassified. Thus, we have to examine the appropriate threshold when making training
samples. 2) The proposed methods should be evaluated using other SAR images to verify the generalization ability. However, the shooting conditions of SAR images are always different; thus, it is difficult to build a classifier that can be used for all SAR images. To solve this problem, we will study a data normalization method using, for example, generative adversarial networks [31], 3) Since a CNN is based on supervised learning, which requires a large number of training samples, applying other learning mechanisms should be effective, for example, anomaly detection [32], which makes the best use of normal samples, and transfer learning [33], which uses various SAR images taken at other disaster-stricken areas in the past.

7. Conclusions

In this paper, we described landslide classification methods using a CNN with SAR images. The two multichannel image generation methods; zero-padding method and map-concatenation method, improved the classification accuracy; therefore, the effectiveness of the multichannel images was clarified. Although we applied the proposed methods to simple CNN structures to verify the basic performance, we will apply them to state-of-the-art CNN structures to evaluate the classification performance in the future.

References

[1] T. Lillesand, R. W. Kiefer and J. Chipman: Remote Sensing and Image Interpretation, John Wiley & Sons, 2014.
[2] H. Shahabi and M. Hashim: Landslide susceptibility mapping using GIS-based statistical models and remote sensing data in tropical environment, Sci. Rep., Vol. 5, 9899, 2015.
[3] P. Boccardo and F. G. Tonolo: Remote sensing role in emergency mapping for disaster response, Engineering Geology for Society and Territory, Vol. 5, pp. 17-24. Springer International Publishing, 2015.
[4] O. M. Bello and Y. A. Aina: Satellite remote sensing as a tool in disaster management and sustainable development: Towards a synergistic approach, Procedia-Social Behav. Sci., Vol. 120, pp. 455-469, 2014.
[5] R. Behling, S. Roessner, D. Golovko and B. Kleinschmit: Derivation of long-term spatiotemporal landslide activity - A multi-sensor time series approach, Remote Sensing of Environment, Vol. 186, pp.88-104, 2016.
[6] N. Tamkuan and M. Nagai: Fusion of multi-temporal interferometric coherence and optical image data for the 2016 Kumamoto earthquake damage assessment, ISPRS International Journal of Geo-Information, Vol. 6, 188, 2017.
[7] A. Bormudoi and M. Nagai: A remote-sensing-based vegetative technique for flood hazard mitigation of Jiadhal basin, India, Natural Hazards, Vol. 83, No. 1, pp.411-423, 2016.
[8] M. K. Hazarika and K. Honda: Estimation of soil erosion using remote sensing and GIS: Its valuation and economic implications on agricultural production, Sustaining the Global Farm, pp. 1090-1093, 2001.
[9] S. Plank: Rapid damage assessment by means of multi-temporal SAR - A comprehensive review and outlook to Sentinel-1, Remote Sensing, Vol. 6, No. 6, pp.4870-4906, 2014.
[10] L. Zhang, L. Zhang and B. Du: Deep learning for remote sensing data: A technical tutorial on the state of the art, IEEE Geosci. and Remote Sensing Magazine, Vol. 4, No. 2, pp.22-40, 2016.
[11] Q. Zou, L. Ni, T. Zhang and Q. Wang: Deep learning based feature selection for remote sensing scene classification, IEEE Geosci. Remote Sensing Lett., Vol. 12, No. 11, pp. 2321-2325, 2015.
[12] G. E. Hinton, S. Osindero and Y.-W. Teh: A fast learning algorithm for deep belief nets, Neural Computation, Vol. 18, No. 7, pp. 1527-1554, 2006.
[13] Y. Liu and L. Wu: Geological disaster recognition on optical remote sensing images using deep learning, Procedia Computer Sci., Vol. 91, pp. 556-575, 2016.
[14] S. Chen, H. Wang, F. Xu and Y. Jin: Target classification using the deep convolutional networks for SAR Images, IEEE Trans. on Geosci. and Remote Sensing, Vol. 54, No. 8, pp. 4806-4817, 2016.
[15] Z. Huang, Z. Pan and B. Lei: Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data, Remote Sensing, Vol. 9, 907, 2017.
[16] The Air Force Moving and Stationary Target Recognition Database, Available: https://www.sdms.afrl.af.mil/
[17] J. P. Fitch: Synthetic Aperture Radar, Springer Sci. & Business Media, 2012.
[18] K. Ouchi: Principles of Synthetic Aperture Radar for Remote Sensing (in Japanese), Tokyo Denki University Press, 2009.
[19] C. Atzberger: Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs, Remote Sensing, Vol. 5, No. 2, pp.949-981, 2013.
[20] D. B. Patissier, J. F. R. Gower, A. G. Dekker, S. R. Phinn and V. E. Brando: A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans, Progress in Oceanography, Vol. 123, pp. 123-144, 2014.
[21] E. C. Barrett: Introduction to Environmental Remote Sensing, Routledge, 2013.
[22] K. Fukushima and S. Miyake: Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position, Pattern Recognition, Vol. 15, No. 6, pp. 455-469, 1982.
[23] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel: Backpropagation applied to handwritten zip code recognition, Neural Com-
Shingo Mabu received his B.E. and M.E. degrees from Kyushu University, Japan, in 2001 and 2003, respectively, and his Ph.D. degree from Waseda University, Japan, in 2006. From 2006 to 2007, he was a visiting lecturer at Waseda University, and from 2007 to 2012, he was an Assistant Professor at the same university. Since 2017, he has been an Associate Professor at the same university.

Yoshiaki Nakayama received his B.E. degree from Yamaguchi University, Japan, in 2016 and his M.E. degree from the same university in 2018. He is currently working as an engineer at Platform Development Department, AIoT Cloud Inc.

Takashi Kuremoto received his B.E. degree from the University of Shanghai for Science and Technology, China, in 1986 and his M.E. and Ph.D. degrees from Yamaguchi University, Japan, in 1996 and 2014, respectively. He worked as a system engineer at the Research Institute of Automatic Machine of Beijing from 1986 to 1992 and is currently an Assistant Professor at Yamaguchi University. He was an Academic Visitor of the School of Computer Science, The University of Manchester, U.K., in 2008.