Feature Extraction and Classification of Multispectral Imagery by Using Convolutional Neural Network

Atsushi Yoshihara
Graduate School of System Informatics,
Kobe University,
Nada, Kobe 657-8501,Japan
Email: yoshihara@me.cs.scitec.kobe-u.ac.jp

Tetsuya Takiguchi
Graduate School of System Informatics,
Kobe University,
Nada, Kobe 657-8501, Japan
Email:takigu@kobe-u.ac.jp

Yasuo Ariki
Graduate School of System Informatics,
Kobe University,
Nada, Kobe 657-8501, Japan
Email:ariki@kobe-u.ac.jp

***Abstract*—Once large-scale disasters occur like earthquake and tunami, it becomes very important task to secure safe evacuation and rescue routes. In this task, land cover classification can be performed on satellite images using remote sensing technology. But, there is a problem that the accuracy of these classification is still low in order to be practical in large-scale disasters. Therefore, we decided to create a detailed situation map right after the disaster by overlaying the cover classification result on the map information before the disaster. As a first step for this purpose, we propose a method to improve the classification accuracy. In this paper, we employ a convolutional neural network(CNN) for feature extraction. Generally, in classification tasks, most of deep learning models employ the softmax activation. But, [1] shows that by simply replacing softmax with linear SVMs, it gives significant gain. So, we employ the RBF(Radial Basis Function) SVM for classification. From our experimental results, we demonstrated that classification method using SVM is about 4(%) more accurate than softmax used in general classification method using CNN.**

I. Introduction

Every year, large-scale disasters occur all over the world and cause great damage. Among them, the Japanese archipelago is located on multiple plates, so it is a region with many earthquakes. In particular, by the Great East Japan Earthquake that occurred in 2011, the Pacific coastal area of the Tohoku region suffered tremendously. Reconstruction activities continue even after five years have passed. In the event of such a large-scale disaster, securing safe evacuation and rescue routes, and considering reconstruction measures are very important tasks. For these tasks, it is necessary to collect wide area information at once. In recent years, remote sensing technology has been drawing attention to realize these tasks. Remote sensing is a technique of observing the reflection of electromagnetic waves and measuring the object remotely from sensors mounted on platforms such as artificial satellites and aircraft. This technique has advantages such as remoteness, wide area and periodicity, and it is utilized in various fields such as land use survey. In this research, we apply this technique to the situation of the area damaged by the disaster. In this task, land cover classification processing is performed in various ways, and research on this has been done in many ways [1-7]. In these studies, various classification methods have been proposed for the purpose of improving classification accuracy. But, there is a problem that the accuracy of these classification is still low in order to be practical in large-scale disasters.

Therefore, we propose a method to improve classification accuracy by using CNN which shows excellent performance in various fields such as speech recognition, image classification and natural language processing in recent years. Generally, in classification tasks, most of deep learning models employ the softmax activation. But, [1] shows that by simply replacing softmax with linear SVMs, it gives significant gain. So, in this paper, we employ a CNN for feature extraction and the SVM for classification to improve classification accuracy.

The rest of this paper is organized as follows: In Section 2, related works are described and our method is proposed in Section 3. In Section 4, the experimental data is evaluated, and the final section is devoted to our conclusions and future work.

II. Related Work

Classical method for land cover classification of multispectral satellite images includes supervised classifiers such as the support vector machine (SVM) [2], [3], the conditional random fields (CRF) [4], [5], and random forest (RF)[6], [7], [8].

The support vector machine(SVM) is supervised non-parametric statistical learning technique. Its training algorithm aims to find a hyperplane that separates the dataset into a discrete predefined number of classes in a fashion consistent with the training examples. In [2], [3], they shows that SVMs demonstrate good performance in the remote sensing field due to improvement of the classification accuracies.

The Conditional Random Fields are a probabilistic framework for labeling and contextual classification. The CRF is a form of undirected graphical model that defines a single log-linear distribution over label sequences given a particular observation sequence. In [4], [5] (L.Albert et al.), a two-layer CRF model is proposed for simultaneous classification of land cover and land use. This results shows their approach yields good accuracies for the land use classes.

TABLE I: Spectral bands used in the multispectral imagery

Band	Bandwidth [nm]
Red	$655 - 690$
Green	$510 - 580$
Blue	$450 - 510$
Near-infrared	$780 - 920$
Panchromatic	$450 - 800$

Random Forest (RF) is proposed by Breman in 2001 for classification and clustering. RF grows many decision tree in the forest. Each tree gives a classification, and the output of the classifier is determined by a majority vote of the trees. In [7] (Ozlem Aker et al.), the classification results of RF classifier are compared with the results obtained from other classification algorithms to evaluate RF performance. And, this experimental results indicates that RF algorithm gives higher classification accuracies than other methods.

While, in recent years, a Convolutional Neural Network (CNN) has shown excellent performance in various fields, such as speech recognition, image recognition and natural language processing[9], [10]. CNN consists of various combination of the convolutional layers, pooling layers and fully connected layers. They tightly couple feature extraction, model construction and classification. In [9] (Wei Hu et al.), they employed deep convolutional neural networks to classify hyperspectral images. Their experiment demonstrates that the proposed method can achieve better performance than some traditional methods, such as SVM, and the conventional deep learning methods. So, we employ CNN to extract features of pixels in satellite images. Generally, in CNN, the softmax activation function is often employed for classification.

But in [1] (Yichuan Tang et al.), they show that SVM works better than softmax on 2 standard datasets (MNIST,CIFAR-10) and a recent dataset. So, instead of the softmax function, we employ the SVM for classification, and compare these classification accuracies.

III. Feature extraction and classification method

A. Data and study area

The data that are used in this work are very high spatial resolution Geoeye-1 satellite images data obtained from Geoeye-1 sensor. Table. I shows the bands and their respective bandwidths. The size of the orthographic images is 10314×10312 pixels with a spatial resolution of 0.5 meters per pixel.

And, the study area is located in Ishinomaki city. This city was damaged by the tsunami by the Great East Japan Earthquake that occurred in 2011. We extracted sample region (256×256 pixels) that includes all five classes (building, water, vegetation, asphalt, ground) to be classified, and verified classification accuracy as a preliminary experiment. Figure. 1 shows the study area in our work.

Fig. 1: Study Area

B. Method

The process of classifying a single pixel in a satellite image using a CNN can be seen in Figure. 2. The input to the first CNN layer consists of c number of spectral bands of contextual size m $\times$ m, where the pixel to be classified is located at the center. The full architecture consists of standard CNN layers, followed by a fully-connected (FC) layer. Finally, Method 1 followed a softmax classifier, while Method 2 followed SVM. The convolutional and pooling layer for the first CNN consists of k number of feature and pooling maps, one for each filter. Rectified linear units (ReLU) are used as the non-linear activation function after the convolutional step. A normalization step with local contrast normalization (LCN) is performed after the non-linear activation function step in order to normalize the non-saturated output caused by the ReLU activation function.

The hidden layer of the fully-connected layer is then used as input to classifier for the final classification. Training of the filters, the fully-connected layer and the softmax classifier of Method 1 are learned with k-means by extracting 10,000 randomly-extracted patches of size m $\times$ m and using them as input to the k-means algorithm. The parameters of the fully-connected layer and the softmax classifier are trained from random initialization and then trained with backpropagation using stochastic gradient descent (SGD).

In Method 2, we employed a radial basis function (RBF) SVM for classification. In this process, scaling, greedy research for tuning parameters, and cross validation are done. In SVM using RBF kernel, two hyperparameters are adjusted using greedy research method. One is a cost parameter C that determines how much misclassification is acceptable. And, the other is γ that used in the RBF kernel.

$$\min_{\beta} \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{N}\xi_i \tag{1}$$

$$K(x,x') = \exp(-\gamma\|x-x'\|^2) \tag{2}$$

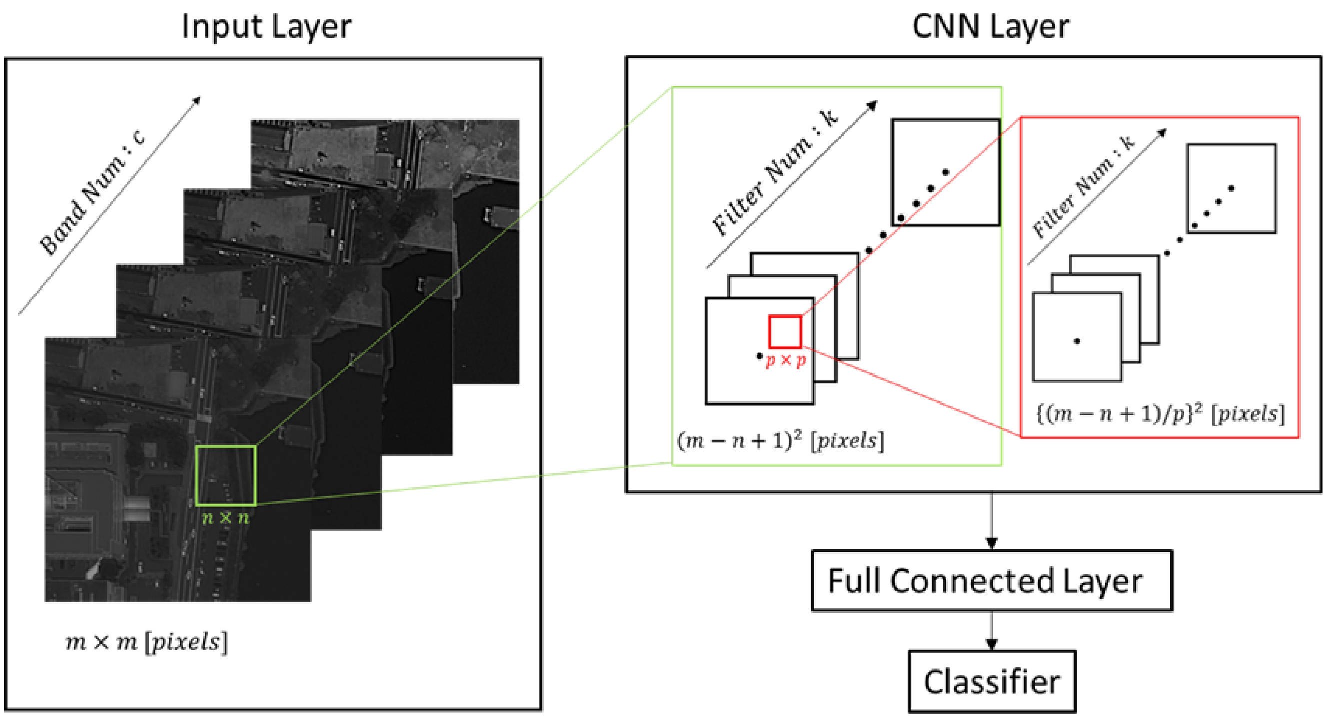

Fig. 2: Overview of the method

IV. Experiments

A. Experimental condition

The data consist of 4 multispectral images with a spatial resolution of 0.5 meters per pixel of a sample area in Ishinomaki city. The data were manually labeled into five categories (building, water, vegetation, asphalt, ground). The date sets are created by randomly selecting 100 pixels from each of the five categories and then assigning 80(%) of them as the training set, 20(%) as the testing set. In our experiments, the CNN architecture model parameters that used are the same values evaluated in [11].

B. Experimental result

In order to compare the accuracy of the classification results created by the two methods, softmax classifier and SVM classifier, the same set of ground truth was used. TABLE III, TABLE IV and TABLE V show the classification results by the softmax classifier and SVM classifier. We evaluate classification accuracy by calculating Recall, Precision and F-measure (ref. TABLE II). Recall refers to the probability that a certain land cover of an area on the ground is classified as such, while Precision refers to the probability that a pixel labeled as a certain land cover class in the map is really this class. And, F-measure is a measure of a test's accuracy and is defined as the weighted harmonic mean of the Recall and Precision of the test.

TABLE II: Accuracy Assessment

	Relevant	Non-Relevant
Retrieved	TP	FP
Not Retrieved	FN	TN

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\mathrm{F-measure} = \frac{2\mathrm{Recall} \cdot \mathrm{Precision}}{\mathrm{Recall} + \mathrm{Precision}} \tag{5}$$

It can be seen that the SVM classifier is about 4(%) higher than the softmax classifier. From this result, it can be seen that the accuracy was improved by extracting the feature by CNN and classifying them using SVM.

TABLE III: Classification accuracy of Softmax classifier （%）

	Buildings	Water	Vegetation	Asphalt	Ground
Recall	86.36	100.0	88.89	77.78	80.00
Precision	90.48	100.0	88.89	73.68	80.00

TABLE IV: Classification accuracy of SVM classifier （%）

	Buildings	Water	Vegetation	Asphalt	Ground
Recall	81.82	100.0	88.89	100.0	84.00
Precision	94.74	100.0	94.12	72.00	95.45

TABLE V: Average F-measure (30 trials)

Softmax	SVM
86.09	90.60

V. Conclusion and Future Work

As a preliminary experiment aimed at improving land cover classification accuracy, we performed feature extraction in the sample region of the high resolution satellite imagery using CNN. In response to the report of the reference [1], we employ SVM as classifier and compare the classification accuracy using softmax and SVM for the extracted features. From this experimental results, we demonstrated that classification method using SVM is about 4(%) more accurate than softmax used in general classification method using CNN. In this way, it is considered that classification accuracy can be improved by performing feature extraction with CNN and performing classification in combination with SVM. And, it was found that classification accuracy improves by classifying by SVM even for the data we used. Therefore, in the future, we are planning to examine whether accuracy can be improved by combining with methods other than SVM. Also, the data we used this time is very small, so we should confirm how the result changes by increasing the number of data.

References

[1] Yichuan Tang. "Deep Learning using Linear Support Vector Machines." Contribution to the ICML 2013 Challenges in Representation Learning Workshop.

[2] Giorgos Mountrakis, et al. "Support vector machines in remote sensing." ISPRS Journal of Photogrammetry and Remote Sensing 66 (2011): 247-259.

[3] Yuliya Tarabalka, et al. "Spectral?Spatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques." IEEE Geoscience and Remote Sensing Society (2009): 2973-2987.

[4] L.Albert, et al. "A TWO-LAYER CONDITIONAL RANDOM FIELD MODEL FOR SIMULTANEOUS CLASSIFICATION OF LAND COVER AND LAND USE." Photogrammetry Remote Senssing and Spatial Information Sciences (2014): Volume XL-3, 17-24.

[5] L.Albert, et al. "LAND USE CLASSIFICATION USUNG CONDITIONAL RANDOM FIELDS FOR THE VERIFICAION OF GEOSPATIAL DATABASES." International Society for Photogrammetry and Remote Sensing (2014): Volume ll-4, 14-16.

[6] Jon Atli Nenediktsson, et al. "Random Forests for land cover classification." Pattern Recognition Letters 27 (2006): 294-300.

[7] Ozlem Akar, et al. "Classification of multispectral images using Random Forest algorithm." Journal of Geodesy and Geoinformation (2012): 105-112.

[8] Barrett Lowe, et al. "MULTISPECTRAL IMAGE ANALYSIS USING RANDOM FOREST." International Journal on Soft Computing (2015): Volume6 No.1

[9] Wei Hu, et al. "Deep Convolutional Neural Networks for Hyperspectral Image Classification." Journal of Sensors (2015): Article ID: 258619, 12 pages.

[10] Tomohiro Ishii, et al. "Detection by Classification of Buildings in Multispectral Satellite Imagery." International Conference on Pattern Recognition (2016): Paper ThAT2.1

[11] Martin Langkvist et al., "Classification and Segmentation of Satellite Orthoimgery Using Convolutional Neural Networks." Remote Sensing (2016): vol.8 ,issue 4, p.329.