AN CNN-BASED OPERATIONAL APPROACH FOR LAND COVER MAPPING IN CHINA

Xuemei Zhao¹, Lianru Gao²*, Zhengchao Chen², Bing Zhang²,³

1. School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, 541004, China
2. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100094, China
3. College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: National land cover map with 30m resolution, an important database for studying the interaction between human and the environment, is a tedious work. The rise of deep learning technique provides a new idea for the work. This paper reports a novel method based on deep convolutional neural networks for the national land cover mapping task. The proposed method has four major parts: classification system, data sources, training samples selection, and training & inferencing. The produced deep learning (DL)-based land cover map is compared with two highly accepted land cover maps (the reference land cover map and the GLC30). Overall accuracies of GLC30 and DL-land cover map are 76.45% and 82.59% when considering the reference land cover map as the ground truth. Overall accuracies of the reference land cover map and the DL-land cover map are 74.25% and 78.87% when the GLC30 is treated as the ground truth.

Index Terms— National land cover map, deep learning, Landsat image classification

1. INTRODUCTION

Medium resolution large-scale land cover map is an important database for land cover change detection, climate change, land resource management, and other applications. As a result, a lot of contributions have been made on large-scale image classification. However, the application of traditional image classification algorithms on large-scale land cover mapping is limited due to their preference for different kinds of samples. Experimental results in [1] have shown that Maximum Likelihood Classifier (MLC), decision tree (J4.8), Random Forest (RF) and Support Vector Machine (SVM) obtained 53.9%, 57.9%, 59.8% and 64.9% overall accuracies on a global scale. To produce a large-scale land cover map with high accuracy, Chen et al. acquired the classification result of each class by combining pixel- and object-based algorithms with the corresponding prior knowledge which can improve the recognition ability. Different pixel- and object-based algorithms are employed for the classification of different classes. The produced GLC30 is a combination of classification results for all the classes [2]. However, the production of the GLC30 involves a lot of human labor, including expert interpretation and cross-validation. The national land cover database for the United States produced by the U.S. Geological Survey used a decision tree-based classifier with multi-source training data [3]. It also employs several artificial interpreted auxiliary data and a hierarchical theme-based post-classification which consumes much human labor. Overall, large-scale classification based on conventional algorithms is usually complex and time-consuming.

Deep learning is efficient in dealing with big data. It is able to learn the essential features of different classes from a large training set. A deep neural network can be considered as a complex function composed of stacked convolutional layers. The stacked convolutional layers can learn semantic information of the input image and output its label. It is a multi-to-one function, which means pixels inside a patch of the
image shares the same label when they are fed to the network. The FCN introduced in [4] provides a new idea for training a semantic segmentation network in an end-to-end way. Also, a number of variant networks were proposed to increase the end-to-end prediction accuracy, such as U-Net [5] and PSPNet [6].

To fully utilize the ability of deep neural networks, this paper employs PSPNet as a classifier for Landsat image classification of China. Taking into account existing land cover classification systems and the distribution of different classes, a classification scheme containing 8 classes is proposed, namely forest, grass land, wet land, water body, cultivated land, artificial surface, bare land, and snow & ice. To improve the performance, parameters pretrained from ImageNet were used to finetune the PSPNet.

The rest of this paper is organized as follows. Section 2 describes details of the produced deep learning-based land cover map of China. Corresponding analyses are carried out in Section 3. Section 4 draws the conclusions.

2. METHODOLOGY

2.1 Overall Architecture

The CNN-based operational approach is described in the flowchart as shown in Fig. 1. Original Landsat images and the reference land cover map are used as images and ground truth after they are preprocessed, which is discussed in Sections 2.2 and 2.3. Then training samples are selected according to the proposed four selection principles as described in Section 2.4. Then the DL-land cover map can be produced by inferencing the images with the trained PSPNet (see more details in Section 2.5). This is a step by step process, in which the accuracy of the former step has a high influence on the later step. Therefore, quality control of each part is of important.

![Flowchart of the operational CNN-based approach.](image)

2.2 Classification System

Landsat images covering large scale areas have complex features. For example, water body with different components or imaged under different lighting conditions may present very different spectral characteristics. On the contrary, objects of different classes always share similar spectral features such as forest and grass land. Considering the representation ability of single Landsat image with 30m resolution and the distribution of different objects in China, this paper adopts a classification scheme with 8 classes referencing existing first level classification systems. The 8 classes are forest, grass land, wet land, water body, cultivated land, artificial surface, bare land, and snow & ice.
2.3 Data Sources

528 Level 1T Landsat 5 images in growing season around 2010 are collected. They have been geometrically corrected with ground control points. So, we just stretched and mosaiced them as shown in Fig. 2(a). Radiation correction is not carried out. Hence, the stitching lines are obvious among the scenes of images.

```
(a) Landsat images of China     (b) Reference land cover map
```

“Land Cover Map of the People’s Republic of China for 2010”, obtained from the National Earth System Science Data Sharing Infrastructure, National Science & Technology Infrastructure of China (http://www.geodata.cn), is considered as the reference land cover map, as shown in Fig. 2(b). The land cover map was produced by 143014 labeled samples and validated with more than 30,000 independent samples. It contains 6 first level classes and 38 second level classes with accuracies of 94% and 86%, respectively. 38 second level classes are merged to 8 first level classes according to the proposed classification system as discussed in Section 2.3.

2.4 Selection of Training Samples

The convolutional neural network is a self-learning algorithm. Its learning ability highly depends on the quality of the training set. Therefore, selecting images with high accuracy is significant to optimize the training accuracy of the network. During the selection, the following lines should be followed: 1) Selected images should be representative. The higher the accuracy of training samples, the more likely the network can learn the essential characteristics of different classes. 2) The balance between classes should be considered. The network will sacrifice the accuracy of rare classes to ensure the overall accuracy of the whole training set when imbalanced training samples are fed to the network. 3) The size of the samples is important. The larger the sample, the easier the network learns the global characteristics of bigger objects in the detected image. However, the larger the sample, the larger the memory required by GPU. To make a trade-off between accuracy and the requirements of GPU memory, the training samples are clipped to 512×512 pixels. 4) The distribution of samples also has a great impact on the training results. Influenced by geographical conditions, the same class may present different features in different locations. Taking in to account the spatial distributions of the training samples will have a positive impact on the classification of China.

2.5 Training and Inferencing

In this paper, PSPNet is employed as the classifier to learn features of different classes from the training set and inference the land cover map with the trained parameters. PSPNet learns semantic features by stacked convolutional layers and adjusts objects with different sizes by a multiscale pyramid pooling layer. Six bands except the thermal infrared band are used to train the network. Transfer learning theory is employed to adjust the pretrained parameters on ImageNet to the features of Landsat images. The program is recoded by pytorch with the learning rate set to 10^-4. The momentum and batch size are set to 0.1 and 20, respectively. It takes about 4 days to train 20000 samples for 300 epochs with 4 Titan XP. In the inference process, the original images are clipped to 640×640 pixels with an overlay of 320 pixels to overcome the
drawback of lower accuracy in the boundary. It consumes about 4 hours to classify all the clipped images of China, which is of 568G.

Taiwan and South China Sea Islands are not included in the original reference land cover map.

3. RESULTS AND ANALYSIS

The produced DL-land cover map is shown in Fig. 3. In this experiment, about 30% of the total area is used to train the network to avoid overfitting. Some details of the classification results are listed in Fig. 4, where (a1)-(e1) are original images; (a2)-(e2) are classification results of the reference land cover map; (a3)-(e3) are classification results from GLC30 [2]; (a4)-(e4) are classification results of the DL-land cover map. GLC30 is produced by integrating pixel- and object-based method with prior knowledge, it has an overall accuracy of over 80%. The produced DL-land cover map is able to extract artificial surface with linear features which cannot be extracted in the GLC30 as shown in Fig. 4(a1)-(a4). Forest, grass land and cultivated land have similar spectral features due to the reason that vegetation has similar reflectance characteristics. Therefore, confusions between forest and grass land is a common problem in GLC30 product. As shown in Fig. 3(b1)-(b4), the DL-land cover map can recognize the cultivated land while the reference land cover map and the GLC30 cannot. Overall, the CNN-based operational approach can learn the essential features of each class. Therefore, the produced DL-land cover map outperforms the GLC30 in detail information extraction and decreasing the confusion among different classes of vegetation. Sometimes, it is only possible to collect images in which period the vegetation is not the most prosperous, such as Fig. 4(c1) or some parts of the images presents different spectral features due to imaging conditions (Fig. 4(d1)). The learning ability of the proposed CNN-based operational approach makes it possible to distinguish different classes with their semantic information. Therefore, requirements of preprocessing on the original Landsat images can be relaxed. That is the reason why only stretch is carried out in this paper. From Fig. 4(b1)-Fig. 4(d4), it can be found that the CNN-based operational approach is able to obtain correct classification results even though misclassification exists in the reference land cover map, which is used to train the network.
Fig. 4 Details of the classification results, where (a1)-(e1) are original images; (a2)-(c2) are classification results of the reference land cover map; (a3)-(e3) are classification results from GLC30; (a4)-(e4) are classification results of the DL-land cover map.

Fig. 5 Classification accuracies of selected images.

Wet land is composed of water and vegetation and it possesses the smallest area all over China compared with other classes. It is easily confused with grass land and water body. The reference land cover map and the GLC30 classify the wet land in the left of Fig. 4(e1) as water body, whereas the CNN-based operational approach is able to recognize it correctly. Due to the learning ability of the CNN-based operational approach, the classification results of the DL-land cover map are more accurate compared with the reference land cover map and the GCL30.
25 images with 10240×10240 pixels are selected to quantitatively evaluate the three mentioned land cover maps. As most of the 8 first level classes are concentrated in eastern China, 18 typical images in this area and 7 other typical images in western China are manually selected as test images. The accuracies of all the images are shown in Fig. 5. It is easy to find that accuracies, obtained when the reference land cover map is considered as the ground truth, is more stable and higher than accuracies obtained when the GLC30 is employed as the ground truth. If the reference land cover map is treated as the ground truth, overall accuracies of the DL-land cover map and the GLC30 are 82.30% and 76.02%, respectively. If the GCL30 is considered as the ground truth, overall accuracies of the DL-land cover map and the reference land cover map are 78.47% and 73.92%. The DL-land cover map outperforms both the reference land cover map and the GLC30 when the other product is regarded as the ground truth. Actually, the official overall accuracy of the reference land cover map is higher than the GLC30. The experiments demonstrate that the overall accuracy of the DL-land cover map is higher when compared with higher accuracy product. Hence, there is a great chance that the real accuracy of DL-land cover map is higher than 82.59%.

4. CONCLUSIONS
This paper reports a fully automatic CNN-based operational approach for land cover mapping in China. With this attempt, we provide a new idea for efficient large-scale land cover mapping. The DL-land cover map was produced by inaccurate training samples selected from the reference land cover map with the training samples account for 30% of China. The CNN-based operational approach is able to learn the essential features from inconsistency training samples. This makes it possible to produce products without enough accurate samples.

ACKNOWLEDGMENT
This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant No. XDA1908302, the National Natural Science Foundation of China under Grant No. 41801233, and the 62-class General Financial Grant from the China Postdoctoral Science Foundation under Grant No. 2017M620947

REFERENCES
[1] P. Gong, J. Wang, L. Yu, Y. et al, “Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data,” Inter. J. Remote Sens., 2013, 34(7): 2607-2654.
[2] J. Chen, J. Chen, A. Liao, X, et al, “Global land cover mapping at 30m resolution: a POK-based operational approach,” ISPRS J. Photogrammetry Remote Sens., vol. 103, pp. 7-27, 2015.
[3] L. Yang, S. Jin, P. Danielson, et al, “A new generation of the United States national land cover database: requirements, research priorities, design, and implementation strategies,” ISPRS J. Photogrammetry Remote Sens., 146, pp. 108-123, 2018.
[4] E. Shelhamer, J. Long, and T. Darrell, “Fully convolutional networks for semantic segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 640-651, 2017.
[5] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: convolutional networks for biomedical image segmentation,” Inter. Conf. Medical Image Comput. Computer-Assisted Intervention, Springer, Cham, pp. 234-241, 2015.
[6] H. Zhao, J. Shi, X. Qi, et al, “Pyramid Scene Parsing Network,” IEEE Conf. Comp. Vision Pattern Recog., 2017, pp. 2881-2890.