APPLICATION OF DEEP LEARNING IN GLOBELAND30-2010 PRODUCT REFINEMENT

Tianfu Liu, Xuehong Chen*

State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing, China-
chenxuehong@bnu.edu.cn

Commission ICWG IV/III

KEY WORDS: GlobeLand30-2010, Deep learning, Google Earth, Training sample set, High resolution images

ABSTRACT:

GlobeLand30, as one of the best Global Land Cover (GLC) product at 30-m resolution, has been widely used in many research fields. Due to the significant spectral confusion among different land cover types and limited textual information of Landsat data, the overall accuracy of GlobeLand30 is about 80%. Although such accuracy is much higher than most other global land cover products, it cannot satisfy various applications. There is still a great need of an effective method to improve the quality of GlobeLand30. The explosive high-resolution satellite images and remarkable performance of Deep Learning on image classification provide a new opportunity to refine GlobeLand30. However, the performance of deep leaning depends on quality and quantity of training samples as well as model training strategy. Therefore, this paper 1) proposed an automatic training sample generation method via Google earth to build a large training sample set; and 2) explore the best training strategy for land cover classification using GoogleNet (Inception V3), one of the most widely used deep learning network. The result shows that the fine-tuning from first layer of Inception V3 using rough large sample set is the best strategy. The retrained network was then applied in one selected area from Xi'an city as a case study of GlobeLand30 refinement. The experiment results indicate that the proposed approach with Deep Learning and google earth imagery is a promising solution for further improving accuracy of GlobeLand30.

1. INTRODUCTION

Global land cover(GLC) product plays an important role in environmental change studies, earth system studies, land resource management, sustainable development, and other societal needs. Among existing GLC products, GlobeLand30 is the world's first 30-meter resolution GLC dataset developed by China (Chen et al. 2017). GlobeLand30 was produced based on Landsat datasets via Pixel-Object-Knowledge based (POK) classification approach (Chen et al. 2015). The traditional classification techniques usually suffer from the issue of synonyms spectrum, the foreign matter with same spectrum (Chen et al. 2015); meanwhile Landsat data cannot provide enough textual information for distinguishing land cover types. Therefore, a knowledge-based verification process was employed to further improve classification accuracy of Globeland30. However, the knowledge-based verification relies on visual interpretation of online high resolution images, such as Map World, Google Earth, and Bing Map, which is laborious and subjective. Although GlobeLand30 achieved an overall accuracy of 80.3% by POK approach, there exist misclassification between certain classes, in particular between grasslands and shrubs, artificial surfaces and barren lands (Yang et al. 2017). On the other hand, large disagreement between GlobeLand30 and other GLC datasets like Urban Atlas (UA) and OpenStreetMap for certain classes are evident, especially wetlands and shrubs (Arsanjani et al. 2016, Huang et al. 2016). GlobeLand30 lacks details necessary for certain applications. For example, it lacks detailed information of waterbodies and vegetation, which therefore limits its employment in urban planning and landscape design.

To further improve GLC datasets quality, some data fusion methods based on multi-source datum were proposed, these methods integrate advantages of spatial temporal resolution and accuracy of differing datum, providing quality refinement for GlobeLand30 product. Jung et.al (2006) exploited synergies of global land cover products for carbon cycle modeling. They merged various GLC products (GLCC, GLC2000 and MODIS LC product) into a desired classification legend and developed a new joint 1-km global land cover product (SYNMAP). This product improves characteristics for land cover parameterization of the carbon cycle models that reduce land cover uncertainties in carbon budget calculations. Yu et al (2014) generated FROM-GLC-agg(Aggregation), an improved version of FROM-GLC product, by blending NL-ISA and MODIS-urban. Huang et al (2016) assessed and improved the accuracy of GlobeLand30 data for urban area delineation by combining National Land Cover Database(NLCD), the Land Use Interpretation Map(LUIM) and Landsat images. This method firstly overlapped GlobeLand30 and other land use/land cover products (e.g., NLCD or LUIM), and then the study area was separated into reliable and unreliable areas with a majority voting rule. Finally, the unreliable areas are confirmed by use of the Landsat data with a multi-classifier system. Those construction methods based on data fusion perform well for specific land cover types at regional scale. However, these methods are impossible to be employed at global scale because of a number of factors, including the availability of good-quality imagery covering the land surface of the entire Earth and the complex spectral and textual characterization of global landscapes.
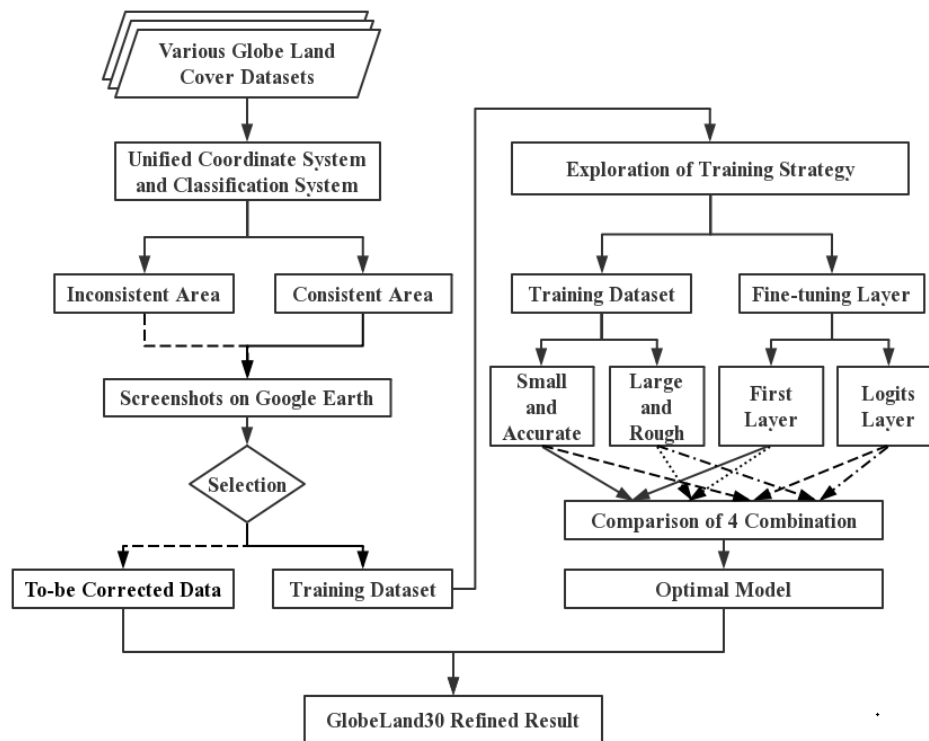
---

* Corresponding author

Figure 1. Flowchart of GlobeLand30 refinement

With the booming of big data.in last decade, deep learning has attracted more and more attentions because of its outstanding performance on image classification. A satisfactory result of image classification using deep learning requires large volume of training sample set, such amount of samples could be provided by remote sensing data. Consequently, deep learning and massive remote sensing data bring an opportunity for GlobeLand30 data refinement. Neural network model and training sample set are two main factors affecting classification accuracy of remote sensing images using deep learning. Li et al (2016) identified and counted oil palm trees using LeNet model composed of two convolutional layers, two pooling layers and a fully connected layer. They firstly selected training samples manually, and then optimize model parameters. The result suggests a higher accuracy achieved by optimized model. Hu et al (2015) produced WHU-RS dataset with 5000 samples in total, covering 20 semantic classes. This is the first time to provide a public benchmark dataset at this size on the problem of scene classification in high-resolution remote sensing imagery. Most widely used remote sensing image datasets now include UC Merced Land-Use, AID, RSSCN7, RSC11, NWPU-RESISC45. Among these datasets, NWPU-RESISC45 generated by Cheng et al (2017) is the largest one. This sample set includes 45classes and 31500 samples totally, 700 samples for per class. However, most of these datasets are produced manually, consuming large amount of time and costs as well as leading to great uncertainty in products.

This paper aims to explore application of deep learning in GlobeLand30 data refinement. First, an automatic training sample generation method via Google earth was proposed to build a large training sample set. Second, the impacts of CNN model as well as quality and quantity of training sample set on classification accuracy was analysed. Then, the optimal strategy for refining GlobeLand30 product was selected to be tested in one selected study area from Xi'an city.

2. METHODOLOGY

Globeland30-2010 dataset is divided into two types of areas (figure 1): (1) The Consistent areas, which are identified as the same land cover type by Globeland30-2010 and other selected GLC dataset, will not be refined. They are used to automatic extract training samples with a Google Earth extension plug-in. And then the training samples were used to train network model of deep learning. (2) The Inconsistent areas, in which Globeland30-2010 and other GLC dataset have disagreement on classification types, are considered as to-be corrected data. And then the inconsistent areas are put into the trained deep learning model for reclassification.

2.1 Automatic generation of a large training sample set

Datasets	Images per class	Scene class	Total images
UC Merced Land-Use	100	21	2100
WHU-RS19	~200	19	500
AID	220~420	30	10000
RSSCN7	400	7	2800
RSC11	~100	11	1232
NWPU-RESISC45	700	45	**31500**
PatternNet	**800**	38	30400

Table 1. Information of publicly available datasets

Several public and widely used existing remote sensing image datasets are listed in Table 1. Among those datasets, the largest size for total and single type is 31500 and 800 images respectively. Such datasets sizes are far from adequate for the deep learning model training. As a result, this paper proposes a fully automated method to generate a new large sample set. Specifically, we develop a screening program based on Google Earth API using C# language. In this program, users input
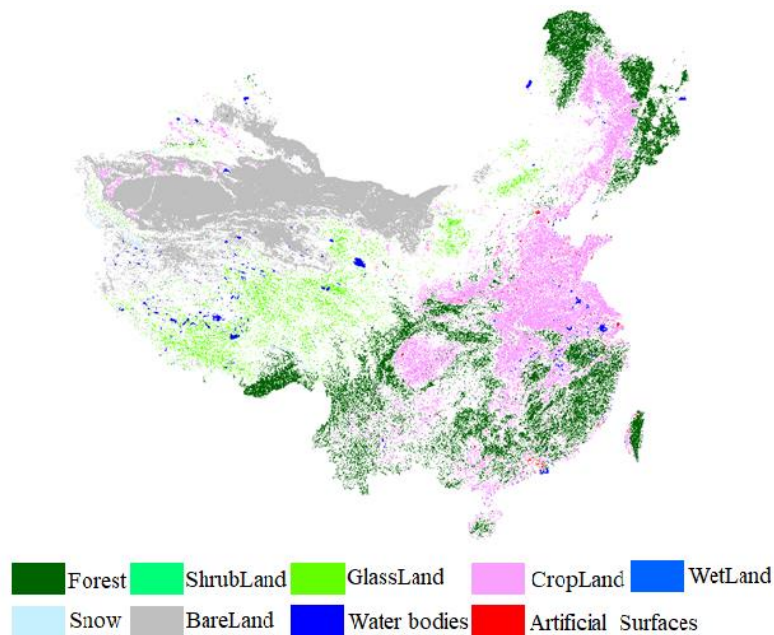
Figure 2. Union Map (GLC dataset except for GlobeLand30)

graphical coordinates of points and a specific date and get the outputs including input points-centred 90m╳90m screenshot of Google Earth images acquired around the settled date. Through this program, we generated a large sample set of seven land cover classes, with over 90,000 images in total and over 10,000 images for each class. In terms of detailed procedure in the sample set generating procedure, there are mainly 3 steps.

Firstly, we analysed the differences between Globeland30-2010 and other existing GLC dataset on the acquisition time, classification system and resolution. we selected images from other GLC datasets (MCD12Q12010, GLCNMO2008, CCI-LC2010, GlobCover2009 and GlobeLand30-2010) according to rules below:1) resolution of no lower than 500m and 2) acquisition time between 2007 and 2013. And then these datasets are converted to a unified coordinate and classification system, and the resampling methods are used to uniform the resolution of all selected images at 300m. Secondly, extract the consistency regions from images selected above (Figure 2). The consistency regions are those areas which are classified into same classes in GlobeLand30-2010 and other GLC datasets. and obtain the corresponding geographical coordinates of those consistency regions. And then using the developed screenshot program to automatically get Google Earth image screenshots. according to the inputted geographical coordinates, and then an initial large sample dataset was generated.

However, there are some unsatisfactory images in this dataset. Those images have the too low spatial resolution or their acquisition time is too far from 2010. So a method has been proposed to solve this problem, which scans the dataset according to the photographing time and sources of the images, excluded unsatisfactory images (NASA and Landsat images and the time before 2007 or after 2013), finally we obtain the new large-scale datasets. The new datasets have 7 categories (artificial land, cultivated land, bare land, grassland, forest, glacier and permanent snow cover and water bodies), excluding tundra, shrub lands and wetlands in GlobeLand30-2010 (Figure 3). Shrubs are merged into grassland type since they are truly confusable. Tundra is excluded from consideration for there is no this type defined in other GLC datasets. And wetlands are

also neglected for its ambiguous definitions. And the sample size of each category is no less than 10,000, it's far more than that of other datasets.

2.2 Comparison of different training strategies

In the large-scale training dataset obtained in this paper, there are a few incorrect label images, which means they are identified as the same class among GLC datasets above but actually are misclassified. For example, artificial lands are labelled as cultivated barren land samples. That means our new dataset cannot provide both high quality and large size samples. We have to trade quality and quantity. This impact on deep learning model is not so sure Therefore, two kinds of training sample sets were generated based on the sample size (large and small) and data quality (rough and accurate), which are rough large sample set (large number of samples with a few incorrect labels for each type) and accurate small sample set (small number of total samples but with no wrong labels). This accurate small sample set is similar to the existing remote sensing sample sets mentioned above.

The total number of rough large samples obtained in this paper is 91,000, with a size of 13,000 for each land cover type. And the proportion of correctly classified samples is approximately 90%. During the process of generating the rough large sample set, it was found that the grassland looks very similar to the bare land, and the water body in winter is similar to ice. So high-resolution images of grasslands and water bodies acquired only between May and September were included in this sample set. However, the number of selected grassland and water body samples is far less than 13,000. Then a sample augmentation method through rotation angles was adopted to enlarge the grassland and water body samples to 13000. The total number of accurate small sample set manually selected is 5,600, and the single sample is 800.

For a convolutional neural network(CNN) model, it performs better with more parameters. However, a large number of parameters means that the model has to be trained on very large data set from scratch and there will be a very intensive
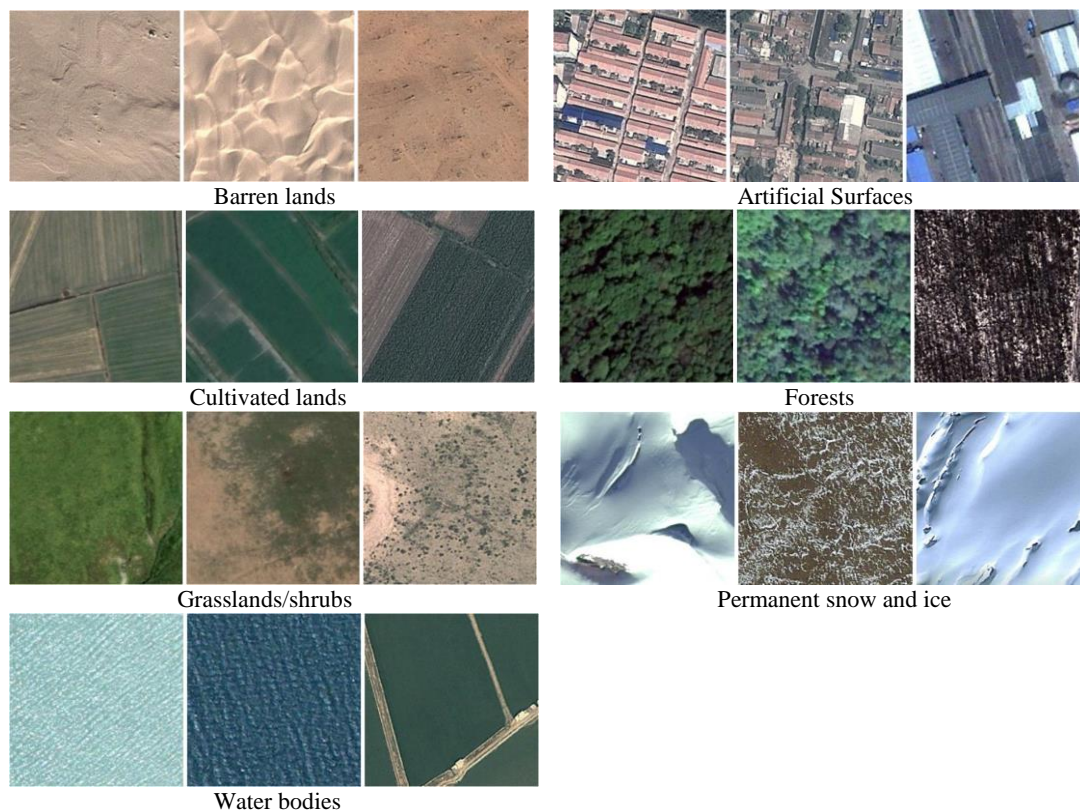
Figure 3. Sample set obtained in this paper: three examples of each class. There are more than 10, 000 images in each category

computation process requiring days or even weeks. Fortunately, it is proved that fine-tuning based on pre-trained CNN models is effective and more efficient (Hu et al. 2015). So this paper suggests using pre-trained models that have been trained using other image classification dataset. Among existing pre-trained CNN models, GoogleNet model series acquired a better result with relatively fewer parameters than others and inceptionV3 is the most robust and effective version of GoogleNet. Therefore, inceptionV3 model is adapted in this paper, but it's necessary to do fine-tuning on this model for better suitability for our study. We adopted two fine-tuning strategies—fine-tuning from the first layer and from logits respectively, and then we analysed and compared their impacts on classification accuracy.

Given the two aspects discussed above, we constructed four training strategies, including the fine-tuning model from first layer using rough large sample set, the fine-tuning model from logits using rough large sample set, the fine-tuning model from first layer using accurate small sample set and fine-tuning model from logits using accurate small sample set. Then we analysed and compared the effects of four strategies on classification accuracy, and decided the best training strategy. It is worth noting that we conducted our experiment on the platform of TensorFlow in Windows operating system.

2.3 GlobeLand30-2010 Refinement

The inconsistent areas in GlobeLand30-2010 dataset are put into the trained inception V3 model for reclassification to acquire the refinement results. Actually, we selected a region with various land cover types from Xi'an city as a case application.
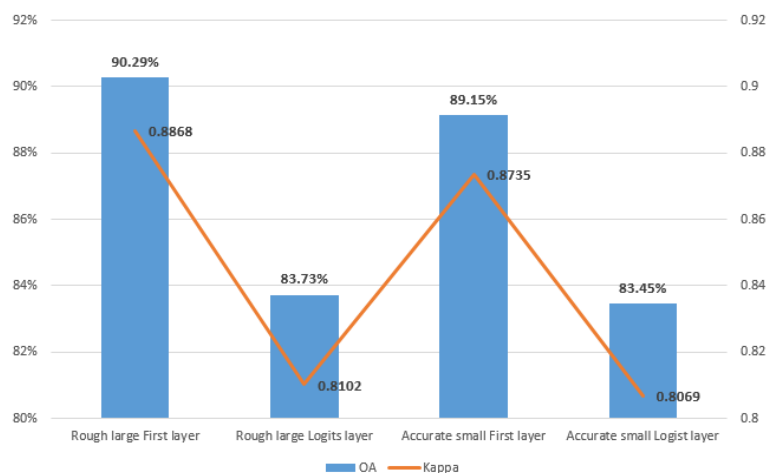

Figure 4. Result of four training strategies (OA: Overall Accuracy)

Number of Pixels		Reference							
		Artificial surface	Barren lands	Cultivated land	Forest	Grassland	Permanent Snow & ice	Water	Total
Classification	Artificial Surface	100	0	1	0	0	0	0	101
	Barren lands	0	92	0	0	8	5	0	99
	Cultivated land	1	1	92	8	7	0	2	111
	Forest	0	0	1	91	4	2	0	96
	Grassland	0	13	5	1	81	0	2	104
	Permanent Snow & ice	0	0	1	0	0	91	4	96
	Water	0	0	0	0	0	2	92	94
	Total	101	100	100	100	100	100	0	701

Table 2 The confusion matrix of the best strategy classification result

3. RESULT

3.1 Comparison of different training strategies

We carried out a number of experiments to explore the effect of four different training strategies on the classification results. Model training is considered completed when the total loss of training and the accuracy of validation become stable in the model training process. The test sample set was obtained by manual visual interpretation. The total number of test samples was 700, with 100 in each category. The results of the four training strategies are shown in Figure 4. As can be seen from the figure, the network fine-tuned from the first layer using rough large sample set obtained the best classification result, with overall accuracy of 90.3% and Kappa coefficient of 0.887. The confusion matrix of this strategy classification result is shown in Table 2. It can be seen from Table 2 that the classification accuracies of artificial surface, cultivated lands, forest, and permanent snow cover are particularly high, whereas the accuracies of grasslands/shrub lands and barren lands are

relatively poor. There are significant misclassifications in grasslands and barren lands, because some of the selected grassland samples are bare images actually, and there are also some images of sparse shrubs with the background of sand in barren land samples. The experimental results also show that the results of fine-tuning from the first layer are better than those from the Logits layer in Inception V3. At the same time, we find that fine-tuning from the first layer in Inception V3 using rough large sample set is the best strategy when the quality of the rough sample set is in an appropriate level.

3.2 Refinement of the GlobeLand30-2010

This paper selects one area of Xi'an city as the case of Globeland 30-2010 product refinement because there are differing types of land cover in the area. The result of the refinement is shown in Figure 5. As shown in the enlarge figures, the misclassifications in Globeland30-2010 were corrected in the experiment.
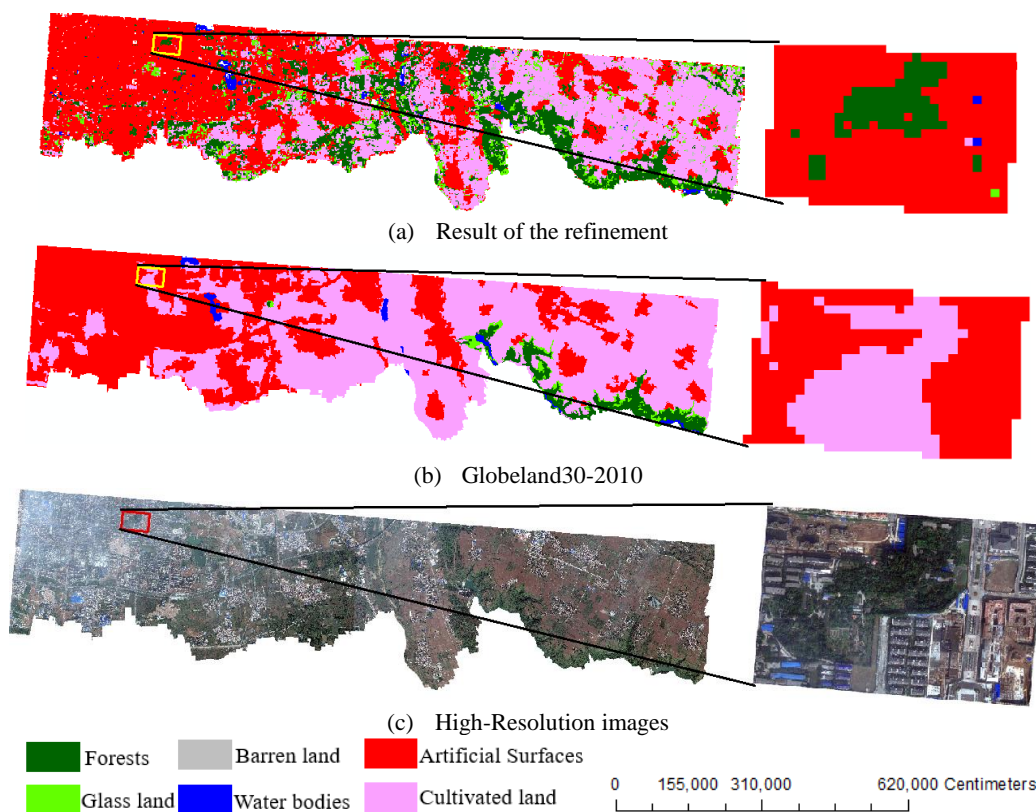

(a)　Result of the refinement

(b)　Globeland30-2010

(c)　High-Resolution images

Figure 5. Compare the result of refinement and GlobeLand30-2010

4. CONCLUSION AND DISCUSSION

This paper explores the application of deep learning in GlobeLand30 product refinement. A method was developed in this process for automatically acquiring large training samples via google earth. This method converts the existing land cover knowledge into large sample set, which greatly improves the efficiency of large sample set generation and provides us a possibility to produce large standard remote sensing sample set like ImageNet. Also, this method can automatically obtain free high-resolution images via the Google Earth to reduce the cost of sample acquisition.

In the exploring process of refining Globeland30 products via deep learning, this paper analyzed and compared four different training strategies based on the quality and quantity of training samples as well as model training strategy. The results of experiments show that the performances of fine-tuning from the first layer are better than those from the Logits layer in Inception V3 model. And a better result is obtained by rough large sample set than accurate small sample set, when the quality of rough sample set is within an appropriate level range. It is found that fine-tuning from the first layer in the inception V3 model using the rough large sample set to be the best strategy. Therefore, this strategy was applied to refine the GlobeLand 30-2010 product and the result shows that the refined product is better than original Globeland30-2010 before refinement.

However, the size of large sample set obtained in this paper is relatively small compared to ImageNet, and there are only 7 land cover classes included in our sample set. Therefore, it is necessary to add more samples of more land cover types to our rough large sample set. Besides, the CNN models used in this paper are all based on the pre-trained models. It may be possible to increasingly improve the accuracy of GlobeLand30 product by adopting such a strategy of training CNN model from scratch. In addition, the resolution of the land cover products in this paper is only 30 meters. It has great importance to produce a higher resolution land cover dataset with deep learning, which is our further work.

ACKNOWLEDGEMENTS

REFERENCES

Arsanjani, J. J., See, L., & Tayyebi, A. 2016. Assessing the suitability of globeland30 for mapping land cover in germany. International Journal of Digital Earth, 9(9), pp. 873-891. https://doi.org/10.1080/17538947.2016.1151956.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., & Chen, X., et al. 2015. Global land cover mapping at 30 m resolution: a pok-based operational approach. Isprs Journal of Photogrammetry & Remote Sensing, 103, pp. 7-27. https://doi.org/10.1016/j.isprsjprs.2014.09.002.

Chen, J., Liao, A., Chen, J., Peng, S., Chen, L., & Zhang, H. 2017. 30-meter global land cover data product- globe land30. Geomatics World.

Cheng, G., Han, J., & Lu, X. 2017. Remote sensing image scene classification: benchmark and state of the art. Proceedings of the IEEE, 105(10), pp. 1865-1883. https://doi.org/10.1109/JPROC.2017.2675998.

Hu, J., Jiang, T., Tong, X., Xia, G. S., & Zhang, L. (2015, July). A benchmark for scene classification of high spatial resolution remote sensing imagery. In Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International (pp. 5003-5006). IEEE. https://doi.org/10.1109/IGARSS.2015.7326956.

Huang, X., Li, Q., Liu, H., & Li, J. 2016. Assessing and improving the accuracy of globeland30 data for urban area delineation by combining multisource remote sensing data. IEEE Geoscience & Remote Sensing Letters, 13(12), pp. 1860-1864. https://doi.org/10.1109/LGRS.2016.2615318.

Jung, M., Henkel, K., Herold, M., & Churkina, G. 2006. Exploiting synergies of global land cover products for carbon cycle modeling. Remote Sensing of Environment, 101(4), pp. 534-553. https://doi.org/10.1016/j.rse.2006.01.020.

Li, W., Fu, H., Yu, L., & Cracknell, A. 2016. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. Remote Sensing, 9(1), 22. https://doi.org/10.3390/rs9010022.

Yang, Y., Xiao, P., Feng, X., & Li, H. 2017. Accuracy assessment of seven global land cover datasets over china. Isprs Journal of Photogrammetry & Remote Sensing, 125, pp. 156-173. https://doi.org/10.1016/j.isprsjprs.2017.01.016.

Yu, L., Wang, J., Li, X. C., Li, C. C., Zhao, Y. Y., & Gong, P. 2014. A multi-resolution global land cover dataset through multisource data aggregation. Science China Earth Sciences, 57(10), pp. 2317-2329. https://doi.org/10.1007/s11430-014-4919-z.

Hu, F., Xia, G. S., Hu, J., & Zhang, L. 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing, 7(11), pp. 14680-14707. doi:10.3390/rs71114680.