CAN A DEEP NETWORK UNDERSTAND THE LAND COVER ACROSS SENSORS?

Zhongling Huang[1,2,3,4], Corneliu Octavian Dumitru[1], Zongxu Pan[3,4], Bin Lei[2,3,4] and Mihai Datcu[1]

[1]Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR)
Münchener Straße 20, 82234 Weßling, Germany
[2]University of Chinese Academy of Sciences, Beijing 101408, China
[3]Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China
[4]Key Laboratory of Technology in Geo-spatial Information Processing and Application System
Chinese Academy of Sciences, Beijing 100190, China

ABSTRACT

Deep learning algorithms are widely used in remote sensing image scene understanding. Generally, a large-scale annotated dataset is essential to train a deep neural network for classification. In practical terms, however, a large amount of unknown remote sensing images obtained from different sensors need to be understood which may vary from resolution, geolocation and imaging conditions compared with annotated datasets. In this paper, an unsupervised domain adaptation framework based on ResNet-18 is presented to transfer the knowledge of an existing annotated land cover dataset to other remote sensing data, decreasing the discrepancy among images across sensors. The results show a significant improvement in scene understanding of new remote sensing images.

Index Terms— land use classification, remote sensing images, transfer learning, domain adaptation

1. INTRODUCTION

Land cover classification is an important issue in remote sensing image understanding. Recently, in remote sensing scene understanding field, most researches have focused on learning hierarchical feature representations. Both unsupervised feature learning based methods, like sparse coding [1], and supervised deep learning methods, like convolutional neural networks (CNNs) [2] are conducted to find a good feature representation of land covers. The supervised deep learning methods all outperform the state-of-the-art methods with various public land cover annotated datasets [3]. Transfer learning is also applied to remote sensing data to overcome the difficulty of limited training samples [4]. However, those methods usually use the same dataset for training and testing. In practical applications, we are facing a large amount of unknown remote sensing data to be understood, collected from various sensors either in satellite or aircraft, with different resolution, imaging conditions and geolocations.

According to most existed approaches, once coming new remote sensing data from other sensors or observation methods, a new task-specific system would have to be designed. Having accumulated such remote sensing land cover datasets [5, 6, 7], it would be great if a deep network can learn from the existing remote sensing data and transfer the knowledge to understand the land covers of newly coming images which will be much more efficient. In this paper, we propose a deep residual network based framework, trying to quickly understand new unknown remote sensing data with deep transfer learning and domain adaptation approach.

2. DATASETS DESCRIPTION

Three remote sensing land cover datasets are explored in this paper, UC Merced Land Use dataset (UCM)[5], AID [6] and NWPU-RESISC45 (NWPU) [7]. In this section, we will give a brief introduction to these datasets and compare them in scale, resolution, obtained sensors, and geolocations.

Fig. 1. Some examples of *harbor/port* and *dense residential* in UCM, AID and NWPU datasets, with different resolution, imaging conditions and geolocations.

UCM is the smallest land cover dataset among them, with 21 land use classes and 100 images per class. It was collected

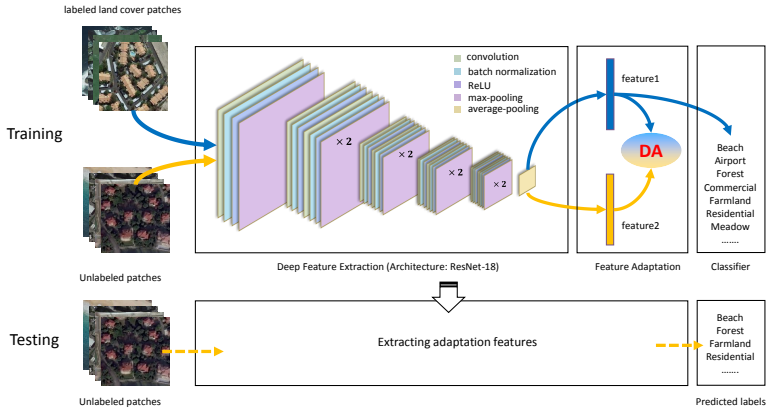


Fig. 2. The proposed framework based on ResNet-18.

Table 1. The Comparison of UCM, AID and NWPU Datasets

Dataset	UCM	AID	NWPU
Resolution	0.3048 m	0.5∼8 m	0.2∼30m
Classes	21	30	45
Total Number	2100	10000	31500
Obtained	Aerial imagery	Google Earth	Google Earth
Geolocations	USA	around the world	more than 100 countries

from aerial orthoimagery in United States urban areas, with a high resolution of 0.3048 m and a patch size of 256 × 256 pixels.

AID is an aerial image dataset collected from Google Earth, but with multi-resolutions from 0.5 m to 8 m as Google Earth images are from different remote sensing sensors. Compared with UCM, AID has a larger scale, with 30 land cover classes and a total number of 10,000, a larger patch size of 600 × 600 pixels, covering more regions around the world which makes the dataset more diversity.

NWPU is a newly released large-scale annotated remote sensing dataset in 2016. It contains 45 scene classes, including land-use and land-cover classes, man-made object classes, as well as landscape nature object classes, and 700 patches in each class with a size of 256 × 256 pixels. Similarly, those images are collected from Google Earth but with a wider range of resolutions, varying from 0.2 m to 30 m in most cases. For classes like lake, mountain and island, the resolution can be lower to cover the efficient semantic areas. Those 31,500 images are collected from more than 100 countries and regions over the world, having rich image variations, high within-class diversity and between-class similarity.

Some typical examples of *port (harbor)* and *dense residential* in different datasets are shown in Fig. 1 and the comparison among three remote sensing datasets is summerized in Table 1.

3. THE PROPOSED FRAMEWORK

The deep residual network (ResNet) [8] is successful for the good performance in training a very deep network of more than 100 layers. As shown in Fig. 2, the ResNet-18 architecture which contains 4 residual blocks with 18 convolution layers is adopted as the deep feature extraction part, followed by the adaptation part and the classifier. Given an image x, a high-level feature vector $\Phi(x) \in \mathbb{R}^n$ is obtained from feature extraction stage by a stack of convolution and down-sampling layers and a global average pooling layer in the end.

The network combining the deep feature extractor together with the classifier of N classes is trained on a labeled land cover dataset $D^s = \{x_i^s, y_i^s\}$. Given a set of new remote sensing images $D^t = \{x_j^t\}$ obtained from other sensors with different resolution, imaging conditions and geolocations, classifying directly with the existed model may cause problems. In deep CNNs, the generality of features drops when going deeper [9] so that the high-level features are more specific to the dataset. When understanding new remote sensing images with a different distribution from existing data, the high-level features to be classified may lead to bias.

The adaptation layer is added in this framework to decrease the discrepancy between D^s and D^t in high-level features. Maximum mean discrepancy (MMD) [10] can be regarded as a discrepancy metric to compare the distributions based on two sets of data. Similar to the deep adaptation network (DAN) [11], by mapping the features into a Reproducing Kernel Hilbert Space (RKHS), the MMD is going to be minimized to narrow the gap between source and target do-

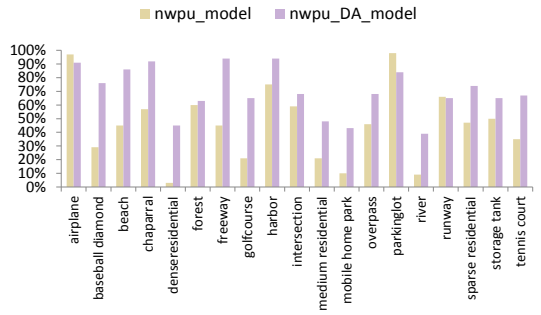
main. Given the input of x^s and x^t, denote the high-level feature discrepancy in RKHS $\mathcal{H}$ as

$$mmd(x^s, x^t) = L_{da}(\Phi(x^s), \Phi(x^t))_{\mathcal{H}}. \quad (1)$$

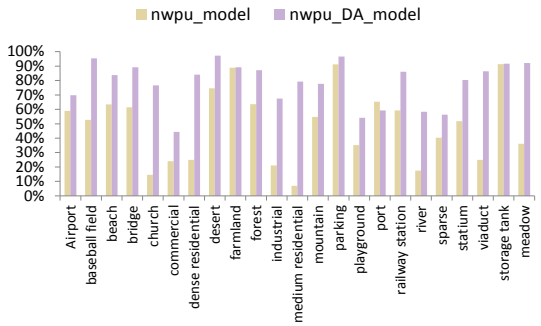
In the training stage, the classifier should be reinitialized to fit the adapted features. As a result, by inputting $\{x_i^s, y_i^s\}$ and $\{x_j^t\}$ into the network, $mmd(x^s, x^t)$ is added to the classification loss with a tradeoff λ to optimize the network, making the high-level features adapted to unknown images and the retrained classifier function well on training data simultaneously. The cost function can be given by

$$L_c(\Phi(x^s), y^s) + \lambda L_{da}(\Phi(x^s), \Phi(x^t))_{\mathcal{H}}. \quad (2)$$

In the testing stage, with the extracted adaptation features of $\{\Phi_{da}(x_j^t)\}$, unknown images $\{x_j^t\}$ can be assigned with new labels of $\{\hat{y}_j^t\}$ by the classifier.



(a)



(b)

Fig. 3. The land cover classification test accuracy on NWPU trained model before adaptation (nwpu_model) and after adaptation (nwpu_DA_model) for two different datasets, UCM (a) and AID (b).

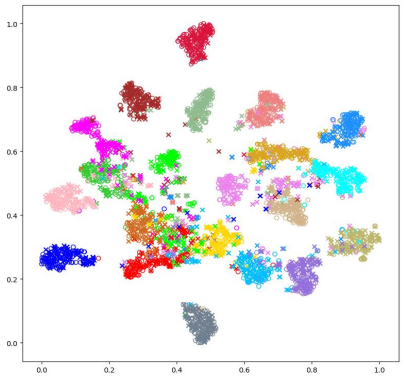
4. EXPERIMENTS

Considering the larger scale on the scene classes and the total image number, we set the NWPU as the training dataset to understand the new land cover images in AID and UCM. In our experiments, we empirically set λ as 0.5 to make a better trade-off. The Adam optimization is applied in training and the initial learning rate during domain adaptation is set to 0.0001.

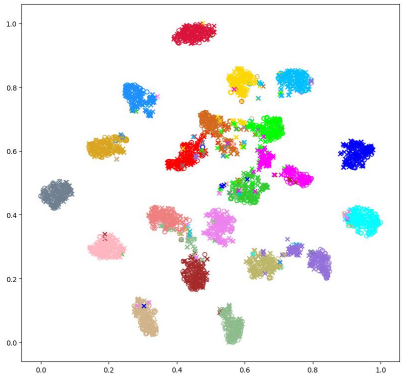
Firstly, the ResNet-18 is trained on NWPU with a training-testing split ratio of 9:1 and achieves an overall accuracy of 93.1% on the test set. Then, we use the trained model to classify the patches in AID and UCM dataset and record the predicted labels compared with the annotation. 19 land cover classes out of 21 in UCM and 23 out of 30 in AID are evaluated, according to their semantical similarity compared with NWPU. The result shows that due to the discrepancy among those land cover datasets, the NWPU trained model just achieves the accuracy of 46% and 48.82% in UCM and AID, respectively.

However, by fine-tuning the deep feature extractors to minimize the MMD in high-level features and retraining the classifier with the adapted features simultaneously, the overall accuracy has been increased to 70% and 78.37% in UCM and AID, improving 24% and 29.55%, respectively. The classification accuracy in each class of two datasets is shown in Fig. 3 and we can observe a remarkable improvement in some classes, such as dense residential areas (improving 59.27% and 42% in UCM and AID, respectively), medium residential areas (improving 72.41% and 27% in UCM and AID, respectively) and church (improving 62.09% in AID). We find it interesting that the residential areas are much various in geolocations due to the different distribution of buildings in various cities, so do the churches. Even though the model only trained on NWPU cannot predict the residential areas or churches from an unfamiliar region or resolution, domain adaptation has narrowed the gap in understanding and learnt some similar patterns across sensors.

In order to understand the domain adaptation of high-level features intuitively, we randomly select 100 images in each evaluated class of three datasets and visualize the high-level features $\Phi(x)$ and $\Phi_{da}(x)$ before and after domain adaptation, shown in Fig. 4. It can be inferred from this figure that the distribution of high-level features varies among datasets due to the decreasing generalization of features in higher layers which leads to a failure in extracting representative features of a new dataset. After adapting high-level features with unknown new dataset and retraining the classifier only with the labeled images, the discrepancy lying between two datasets is significantly decreased so that the land cover features of new images can be successfully extracted to fit the classifier.



(a)



(b)

Fig. 4. The visualization of high-level feature of different datasets before and after domain adaptation, shown in (a) and (b), respectively. Different colors represent different classes while the marker of "o" denotes NWPU dataset and "x" denotes AID.

5. CONCLUSION

The proposed framework makes a deep convolutional neural network possible to understand the land cover of new remote sensing images obtained from different sensors with various resolution, geolocations and imaging conditions. Only fine-tuning the feature extraction part to adapt the high-level features and retraining the classifier with existed labeled dataset, the unknown remote sensing land cover can be efficiently interpreted with a lower cost in time and effort.

6. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under Grant 61701478.

7. REFERENCES

[1] A. M. Cheriyadat, "Unsupervised feature learning for aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 439–451, Jan 2014.

[2] F. Zhang, B. Du, and L. Zhang, "Scene classification via a gradient boosting random convolutional network framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1793–1802, March 2016.

[3] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, June 2016.

[4] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep learning earth observation classification using imagenet pretrained networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 105–109, Jan 2016.

[5] Yi Yang and Shawn Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010, pp. 270–279.

[6] G. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, July 2017.

[7] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct 2017.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[9] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in Neural Information Processing Systems 27*, pp. 3320–3328. Curran Associates, Inc., 2014.

[10] A. Gretton, D. Sejdinovic, H. Strathmann, S. Balakrishnan, M. Pontil, K. Fukumizu, and B. K. Sriperumbudur, "Optimal kernel choice for large-scale two-sample tests," in *Advances in Neural Information Processing Systems 25*. 2012, pp. 1205–1213, Curran Associates, Inc.

[11] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proceedings of the 32nd International Conference on Machine Learning*. 2015, ICML'15, pp. 97–105, JMLR.org.