UNSUPERVISED MIXED MULTI-TARGET DOMAIN ADAPTATION FOR REMOTE SENSING IMAGES CLASSIFICATION

Juepeneg Zheng[†,§], *Wenzhao Wu*[†,§], *Haohuan Fu*[†,§] *(✉)*
Weijia Li[¶], *Runmin Dong*[†,§], *Lixian Zhang*[†,§], *Shuai Yuan*[‡]

[†]Ministry of Education Key Laboratory for Earth System Modeling,
and Department of Earth System Science, Tsinghua University, Beijing 100084, China
[§]Joint Center for Global Change Studies, Beijing 100875, China
[‡]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[¶]CUHK-SenseTime Joint Lab, The Chinese University of Hong Kong, Hong Kong, China

IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9323602

ABSTRACT

Although deep learning has been successfully applied in the field of remote sensing image classification, it still requires time-consuming and costly annotations. In recent years, domain adaptation has been witnessed to address this problem as they do not need any human interpreted in the target domain dataset. However, most of the existing works dedicate effort on the circumstance where there is only one source domain and only one target domain. In this paper, we firstly explore one source and multiple target domains issue for remote sensing application and build a challenging mixed multi-target dataset to contribute to the community. Our method constitutes three parts. Firstly, as we are blind for the multi-target domain, we adopt meta learning to divide the mixed multi-target dataset and insert sub-target domain loss as the part of the loss function. Secondly, we apply the adversarial learning to confuse the classifier to discriminate between the source domain images and the whole mixed multi-target domain images. Finally, the meta learning and the adversarial learning are dynamically iterative procedures and the labels for domain classification in mixed multi-target dataset will be updated for a particular iteration. Our method is well-performed in the four common remote sensing dataset (AID, NWPU-RESISC45, UC Merced and WHU-RS19), including five classes (agriculture, forest, river, residential and parking). Our method achieved an average accuracy of 81.59% and outperformed other domain adaptation method. The experiment results indicate our method is promising for large-scale, multi-regional and multi-temporal remote sensing applications.

***Index Terms*—** Mixed multi-target, domain adaptation, meta-learning, adversarial learning, remote sensing images classification

Correspondence: haohuan@tsinghua.edu.cn (Haohuan Fu)

1. INTRODUCTION

Although deep learning has been widely and successfully used in remote sensing images classification task [1], it requires sufficient annotations in a particular region (source domain) to extract efficient features. When it turns to a new environment (target domain), the accuracy of classifier may drop dramatically. Thanks to domain adaption (DA), we can address the scarcity of annotations for target domain by only leveraging the massively available labelled data in source domain [2, 3].

In recent years, significant performance for DA approaches indeed have been witnessed, including deep domain confusion (DDC) [4], deep adaptation network (DAN) [5], joint adaptation networks (JAN) [6] and so on. DA has been exploited in the remote sensing community to deal with multi-temporal or multi-source remote sensing images [7]. However, most of existing DA methods are preset by "single source and single-target (SSST)", that is, there is only one type source domain and one type target domain drawn from an identical distribution. Actually, SSST is an ideal situation while in reality, there are many circumstances where we only have one type source domain yet more than one type target domains. For example, in autonomous driving, we only have the training images in clear weather in most cases, however, the algorithm is supposed to be enough robust for rainy and foggy weather, or even in the dark night. It is impossible and infeasible to label images for every environment and train each model for satisfying every different condition. Although SSST DA can reduce annotations, it is also inevitable to train each DA model for every transferring circumstance, such as from clear weather to foggy weather, from clear weather to dark night condition, and etc. In addition, we should tell the model what's the exactly environment at present to help it select the appropriate DA model. In this paper, we focus on single source and multiple target (SSMT) DA problem, that is, we only utilize one source domain dataset with annotations to adapt other multiple target domain datasets.

Even in computer vision community, SSMT, especially single source and mixed multi-target is an early issue [8, 9] and to the best of our knowledge, scholars have not discussed SSMT for remote sensing image classification. Our contributions in this paper are as following two aspects:

(1) We collected four different open source remote sensing image classification dataset for five classes to validate the effectiveness of our method. This work is the first attempt for SSMT in the field of remote sensing.

(2) We employed meta learning to distinguish multiple target domains, which is an unsupervised classifier. We also adopted adversarial learning to confuse the source domain and the mixed multi-target domains. The meta learning and the adversarial learning are two dynamically iterative procedures.

2. METHODOLOGY

Our method consists of two major parts, i.e., meta learning for dividing the mixed-multi-target dataset by representative features and adversarial learning for confusing the classifier to discriminate the source dataset and the whole target dataset. Figure 1 describes the flowchart of our method.

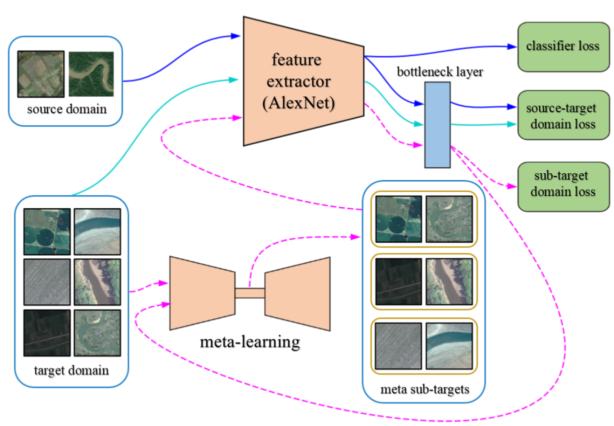

Fig. 1. The flowchart of our method.

In this paper, we concentrate on unsupervised mixed multi-target domain adaptation, which consists of an annotated source domain dataset $D_S = \{(x_i^S, y_i^S)\}_{i=1}^{n_S}$ and an unlabeled but mixed target domain dataset $D_T = \{(x_i^T)\}_{i=1}^{n_T}$, where n_S and n_T are the number of images in the source and target dataset respectively. Notably, $T = \{t_j\}_{j=1}^{k}$, where k is the number of different target datasets. For example, in our work, when we choose AID as our source domain dataset, our target datasets contains NWPU-RESIS45, UC Merced and WHU-19 and in this way, $k = 3$. It is a key message in this problem that we do not know any information about each image in target dataset belongs to which domain in advance.

2.1. Meta learning for dividing the mixed multi-target dataset

As our target datasets are compound, we do not know where each image comes from. If we put all the target images from different datasets into a union target dataset, it may occur negative transfer because of dataset gaps and category misalignment among $\{t_j\}_{j=1}^{k}$. Here we adopt meta learning to address this problem.

Meta learning [10] refers to learning to learn, which leverages all learnable technique to improve model generalization and extract more general and useful information. Our unsupervised meta learning M has an encoder and a decoder. The encoder has three fully-connected layers with the parameters of 500, 500, 1000 and the decoder has three fully-connected layers with the parameters of 1000, 500, 500. The decoder actually is a reconstruction process and the reconstruction loss is MSE loss.

After the training of reconstruction network, we divide the multi-target domain images via K-Means and each target image will have a domain class label denoted by . We insert sub-target domain loss L_{mt} to the final loss and the L_{mt} can be calculated as equation:

$$L_{mt} = \frac{1}{n_T} \sum_{x_i^T \in D_T} L_{cross-entropy}\left(M\left(x_i^T\right), \hat{z}_i\right) \quad (1)$$

2.2. Adversarial learning for confusing the discriminator between the source and the whole mixed target dataset

Let us recall the DANN, in which a two player minimax game is constructed. Similar to DANN, we can effectively confuse our network by minimizing the loss L_{st} of domain discriminator G_d between source domain images and the whole mixed multi-target domain images. L_{st} can be formulated as equation 2:

$$L_{st} = \frac{1}{n_T} \sum_{x_i \in D_S \cup D_T} L_{binary-cross-entropy}\left(G_d\left(x_i\right), d_i\right) \quad (2)$$

d_i is equal to 1 for the source domain images and 0 for the mixed multi-target domain images. For more explanation, the output of $G_d(x_i)$ is the probability of the feature map in image x_i belonging to the source domain. When $G_d(x_i)$ is larger than 0.5, it denotes that the feature map belongs to the source domain, otherwise when the probability is lower than 0.5, it represents that it belongs to the mixed multi-target domain.

2.3. Iteratively collaboration between meta learning and adversarial learning

We select AlexNet as our backbone architecture. Finally, our final loss function is formulated as equation:

Table 1. The information of our dataset for five classes.

Index	AID	NWPU-RESISC45	UC Merced	WHU-RS19
Agriculture	Farmland	Circular farmland Rectangular farmland	Agriculture	Farmland
Forest	Forest	Forest	Forest	Forest
River	River	River	River	River
Residential	Dense residential Medium residential Sparse residential	Dense residential Medium residential Sparse residential	Dense residential Medium residential Sparse residential	Residential
Parking	Parking	Parking lot	Parking lot	Parking
Number	2420	5600	700	263

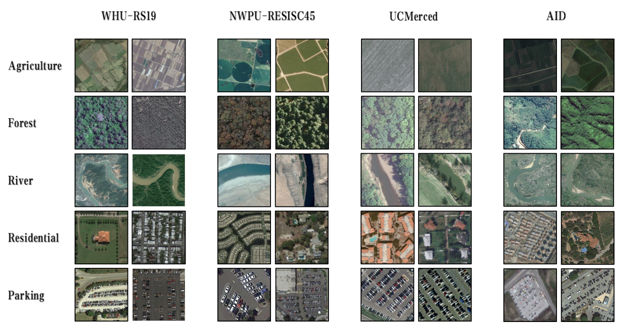

Fig. 2. Examples of our datasets from four popular remote sensing datasets.

Table 2. Comparison of the results between our method and other DA methods.

Method	A→NUW	N→AUW	U→ANW	W→ANU	Average
AlexNet	76.52	70.65	55.58	45.24	62.00
DDC	82.63	83.24	66.58	72.01	76.12
DAN	80.04	90.30	66.69	71.48	77.13
JAN	86.22	87.11	67.98	74.42	78.93
Ours	**88.42**	**92.96**	**70.55**	**74.43**	**81.59**

Table 3. Comparison of the results between without meta learning (**w/o**) and with meta learning (**w**).

Method	A→NUW	N→AUW	U→ANW	W→ANU	Average
w/o	84.16	83.68	67.40	67.53	75.69
w	**88.42**	**92.96**	**70.55**	**74.43**	**81.59**

$$L = L_C + \alpha \times L_{st} + \beta \times L_{mt} \quad (3)$$

where L_C is the classification loss of the labelled source domain dataset. and are the hyper-parameters that trade-off among L_C, L_{st} and L_{mt}. Notably, our meta learning is retrained to update the domain labels of the mixed multi-target images per N iteration. In a summary, our mete learning and adversarial learning are two collaborative and dynamically iterative procedures.

3. DATASET

As illustrated in Figure 2, we selected four different remote sensing datasets, i.e. WHU-RS19 [11], NWPU-RESISC45 [12], UC Merced [13] and AID [14]. WHU-RS19 [11] contains 12 classes of meaningful scenes downloaded from Google Earth. NWPU-RESISC45 [12] consists of 31500 images divided into 45 scene classes with the spatial resolution ranging from 30 to 0.2m per pixel. UC Merced [13] comprises 21 land use classes selected from aerial orthoimagery with a pixel resolution of one foot. AID [14] is made up of 30 aerial scene types collected from Google Earth imagery. Our dataset contains five classes including agriculture, forest, river, residential and parking. We selected some of the classes in common datasets as one class in our dataset. For example, dense residential, medium residential and sparse residential are collected as residential for AID, NWPU-RESISC45 and UC Merced. The information of our dataset is listed in Table 1 elaborately.

4. RESULTS AND EXPERIMENTS

In our experiments, we set $\alpha = 0.1$ and $\beta = 0.01$. We conduct our experiments in our datasets described in Section 3. We evaluate the accuracies for four types of SSMT and their average accuracy. Although the preset of existing DA method is SSST, we compare our method with them and evaluate their classification accuracies by transferring class information from single source domain to mixed multi-target domain. Table 2 lists three state-of-the-art DA method (DDC [4], DAN [6] and JAN [5]) and baseline (AlexNet). Our method achieves the average accuracy of 81.59%. It is shown that our method improves the average accuracy by 19.59% compared with baseline and increases the average accuracy by 5.47%, 4.46% and 2.66% compared to DDC, DAN and JAN respectively.

We also compare the results between without meta learning and with meta learning (our method) in Table 3. When we do not use meta learning and drop out the sub-target domain loss, the accuracies reduce 3.15%-9.28%, indicating the positive effectiveness of meta learning for mixed multi-target domain adaptation issue.

5. CONCLUSIONS

This paper firstly discuss one source and mixed multiple target domain issue for remote sensing. Firstly, we adopt meta learning to divide the mixed multi-target dataset in an unsupervised way and insert the sub-target domain loss to the final loss function. Secondly, we applied adversarial learning to confuse the classifier to discriminate between the source domain images and the whole target domain images. Notably, the meta learning and the adversarial learning are dynamically iterative process and the labels for domain classification in mixed multi-target dataset will be updated for every iteration. We collected four different open source remote sensing image classification dataset and selected five classes to validate the effectiveness of our method, achieving an average accuracy of 81.59% and outperforming other domain adaptation methods. Experimental results indicate our method is potential for large-scale, multi-regional and multi-temporal remote sensing applications.

6. ACKNOWLEDGEMENTS

This research was supported in part by the National Key Research and Development Plan of China (Grant No.2017YFA0 604500, No.2017YFB0202204 and No.2017YFA0604401), the National Natural Science Foundation of China (Grant No.51761135015), and by Center for High Performance Computing and System Simulation, Pilot National Laboratory for Marine Science and Technology (Qingdao)

7. REFERENCES

[1] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.

[2] Sinno Jialin Pan and Qiang Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[3] Ximei Wang, Liang Li, Weirui Ye, Mingsheng Long, and Jianmin Wang, "Transferable attention for domain adaptation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, vol. 33, pp. 5345–5352.

[4] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell, "Deep domain confusion: Maximizing for domain invariance," *arXiv preprint arXiv:1412.3474*, 2014.

[5] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I Jordan, "Learning transferable features with deep adaptation networks," *arXiv preprint arXiv:1502.02791*, 2015.

[6] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan, "Deep transfer learning with joint adaptation networks," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 2208–2217.

[7] Devis Tuia, Claudio Persello, and Lorenzo Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE geoscience and remote sensing magazine*, vol. 4, no. 2, pp. 41–57, 2016.

[8] Ziliang Chen, Jingyu Zhuang, Xiaodan Liang, and Liang Lin, "Blending-target domain adaptation by adversarial meta-adaptation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2248–2257.

[9] Ziwei Liu, Zhongqi Miao, Xingang Pan, Xiaohang Zhan, Stella X Yu, Dahua Lin, and Boqing Gong, "Compound domain adaptation in an open world," *arXiv preprint arXiv:1909.03403*, 2019.

[10] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas, "Learning to learn by gradient descent by gradient descent," in *Advances in neural information processing systems*, 2016, pp. 3981–3989.

[11] Gui-Song Xia, Wen Yang, Julie Delon, Yann Gousseau, Hong Sun, and Henri Maître, "Structural High-resolution Satellite Image Indexing," in *ISPRS TC VII Symposium - 100 Years ISPRS*, Wagner W., Székely, and B., Eds., Vienna, Austria, July 2010, vol. XXXVIII, pp. 298–303.

[12] Gong Cheng, Junwei Han, and Xiaoqiang Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.

[13] 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.

[14] Gui-Song Xia, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang, and Xiaoqiang Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.