Noname manuscript No.
(will be inserted by the editor)

Unsupervised Satellite Image Classification based on Partial Transfer Learning

Jian Hu · Hongya Tuo · Chao Wang · Haowen Zhong · Han Pan · Zhongliang Jing

Received: date / Accepted: date

Abstract Satellite image classification plays an important role in many fields. It can be divided into two groups: supervised and unsupervised classification. The former requires a large number of labeled data, while in practice satellite images usually lack of sufficient ones. How to achieve high accuracy on unsupervised satellite image classification is a key problem. For tackling this problem, based on the idea of partial transfer learning, we propose an end-to-end unsupervised classification method novel coordinate partial adversarial domain adaptation (CPADA) for satellite images classification. Under the aid of a novel coordinate loss, our framework transfers relevant examples in the shared classes to promote performance, and ignore irrelevant ones in the specific classes to mitigate negative transfer. Experiments show that our CPADA exceeds state-of-the-art results for unsupervised satellite image classification task.

Keywords Unsupervised satellite image classification · Partial domain adaptation · Transfer learning

Jian Hu
E-mail: jianhu18@sjtu.edu.cn

Hongya Tuo
E-mail: tuohy@sjtu.edu.cn

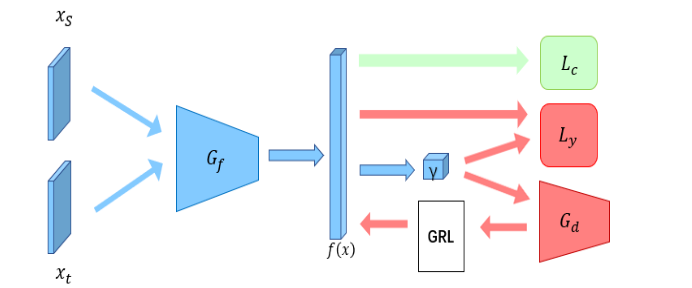
Chao Wang
E-mail: wang_chao@sjtu.edu.cn

Haowen Zhong
E-mail: haowenzhong@sjtu.edu.cn

Han Pan
E-mail: panhan@sjtu.edu.cn

Zhongliang Jing
E-mail: zljing@sjtu.edu.cn

School of Aeronautics and Astronautics
Shanghai Jiaotong University



$f(x)$:output of the feature extractor G_f :feature extractor L_y :classifier loss
x_s :sample from source domain γ :weights of the sample GRL :gradient reverse layer
x_t :sample from target domain G_d :domain discriminator L_c :coordinate loss

Figure 1: Architecture of our method

1 Introduction

Satellite image classification plays an important role in many fields, such as characteristics of the drawing, land cover classification, 3D modeling and change detection. Image classification is also a prerequisite for understanding and analyzing relevant information.

Inspired by rapid development of deep learning methods, image classification accuracy has got a huge promotion. However, deep learning approaches rely on large-scale labelled data, which is also a limitation for deep learning ones applying to many real-world applications. For satellite image classification issues, this problem is particularly essential and changeable. Thus, how to effectively use the limited image information has become a research hot spot.

Supervised or unsupervised classification techniques exist in the literature to deal with satellite image classification issues [1]. For former approaches, under the aid of deep learning methods, we can promote accuracy to a satisfactory level. However, in satellite image classification issue, we do not have plenty of labeled data, making it hard for supervised learning to practical applications. A much more realistic approach is using unsupervised image classification approaches. The generally used unsupervised methods.Unsupervised classification methods need to determine the probability distribution of data compliance, then analyze and judge preliminary results, so as to obtain better classification accuracy. However, these kinds of classification methods have a high computational cost, what's more, they only indicate distribution of the data while the labels are not clear.

Transfer learning is a new risen of unsupervised image classification methods. It focuses on utilizing the model learned from labelled source domain to tasks on unlabeled target domain. We take advantage of transfer learning to tackle the problem of label-lacking for unsupervised satellite image classification.The commonly used satellite image datasets usually include over dozens of categories while in actual situation, we do not need to classify all the categories in dataset. If we apply the standard domain adaptation to unsupervised satellite image classification issues,the source label space does not equal to target label space, so it is too hard to satisfy the basic assumption.these classes belonging to source domain but not target domain may cause negative transfer.

For dealing with this problem, we introduce partial domain adaptation to unsupervised satellite image classification issues,the architecture of our method is shown in Figure 1. The framework includes two part: the adversarial part and the classifiication part.The adversarial process is a two-player minimax games. The first is feature extractor G_f ,which tries to learn domain-invariant features from source and target domain, and the second is domain discriminator G_d which tries to distinguish the domains. The source classifier is trained by minimizing the cross-entropy loss L_y.we put forward a novel coordinate loss. Its main effect is to promote weights of shared classes up to almost 1,and decrease the outlier classes weights down to 0,so as to improve classification ability. Finally, we use the actual satellite image datasets, and compare our algorithm with existing mainstream learning and partial domain adaptation methods, the experimental results shows that CPADA reaches the state-of-the-art result and suitable on satellite image classification problem.

2 Related Work

In this section, we briefly review related works of satellite image classification and domain adaptation,one of the main methods of transfer learning.

2.1 Satellite image classification methods

Most of the satellite image classification methods focus on feature extraction.There are two kinds of methods to extract features including hand-craft feature extracting methods and feature learning methods.

Formal methods aim to extract features artificially designed. [3] [4] try to utilize a Bayesian framework to reduce the gap between low-level features and high-level features. [2] first extracts object specific features by using a modified cloud basis function neural network, then a relax labeling process assisting to get classified images. Above two methods aim to model the satellite image with a bottom-up scheme while the bottom features are still designed by researchers. [5] [11] extract low-dimension features and cluster into different categories, then the features from input images are mapped to its closest label. Even if these methods perform better than previous ones, they use various hand-craft local image descriptors to represent aerial scenes, and may lack the flexibility to different scenes.

Feature learning methods try to learn a high-level representation from input data.they can be categorized into two kinds: the first method tries to learn a scene representation from the unlabeled input data. [12] and [14] utilize a unsupervised method to generate a sparse feature representation.[20] presents an improved unsupervised feature learning method, which can not only learn a good global representation, but it can also find out intrinsic structures of local image patches.

Another method is deep learning method. Deep learning has been applied to many fields and shown astounding results. This kind of ways is also applied to satellite image classification.[17] finetunes Google-Net on the target dataset and achieves good performance. [6] takes advantage of CaffeNet to obtain high-level features, then classify these features from target images with a SVM classifier.

All the methods in [12] [14] [17] [20] and [6] can learn features automatically. But the unsupervised methods spend a lot of time on training while the classify accuracy is not good. And deep learning methods need efficient labeled data, which is hard to satisfy.

2.2 Domain adaptation and partial domain adaptation

Transfer learning focuses on freeing deep learning from plenty of labelled data. Domain adaptation plays an important role in transfer learning. Domain adaptation tries to narrow the gap between source and target domain while it can learn domain-invariant features across domains to assist task on target domain perform well.

It has been applied to many fields such as computer vision, natural language processing and machine learning. A large number of domain adaptation methods have been proposed to deal with label-lacking issues. They mainly focus on closing high-order statistics features and minimax game processing. [22], [9] and [10] aim for adding adaptation layers to match high-order statistics of distribution. Once the high-order features are matched, the gap between source and target domains are eliminated. [8] [21] and [19] try to use a feature extractor to obtain domain-invariant features, discriminator tries to distinguish which domain the feature comes from. Based on this minimax game processing between feature extractor and discriminator, the framework can perform well on target domain with support from source domain.

Partial domain adaptation is a promotion of standard domain adaptation. In standard domain adaptation, a basic assumption is that source and target domain share the identical label space. But in partial domain adaptation, this assumption has been relaxed to source label space contains target one. This new assumption makes it much easier to apply partial domain adaptation to real world application compared with standard domain adaptation. There are three methods [13] [15] [16] dealing with partial domain adaptation issues by down-weighting outlier classes and up-weighting shared classes.

3 Coordinate Partial Adversarial Domain Adaptation

Based on present partial domain adaptation methods, we present a novel coordinate partial adversarial domain adaptation (CPADA) to deal with unsupervised satellite image classification issue. It is the first time to apply partial domain adaptation methods to satellite image classification issue.

In partial domain adaptation setting, source domain is $D_s = (x_i^s, y_i^s)$representing each source sample with a label, while in target domain $D_t = (x_i^t)$representing each target sample without label. When it comes to our satellite image classification issue, the source domain is a satellite image dataset with labeled data while the target domain is a satellite image dataset without label, the target label space is a subset of source label space.

3.1 Domain adaptation neural network

For narrowing the distance between source and target domain, our work bases on Domain Adversarial Neural Network (DaNN) [8], in which this problem tackled by a two-player minimax processing. The first is feature extractor G_f to learn domain-invariant features from source and target domains, and the second is domain discriminator G_d to distinguish the domains. The source classifier G_y is trained by minimizing the cross-entropy loss L_y. The G_y and G_d can be learned as follow:

$$\begin{aligned} minL(G_d, G_f, G_y) &= \frac{1}{n_s}\sum_{i=1}^{n_s} L_y(G_y(G_f(x_i^s)), y_i^s) \\ &- \frac{1}{n_s}\sum_{i=1}^{n_s}[logG_d(G_f(x_i^s))] \\ &- \frac{1}{n_t}\sum_{i=1}^{n_t}[log(1 - G_d(G_f(x_i^t)))] \end{aligned} \tag{1}$$

3.2 Shared-class weights

the classes belonging to both source and target domains are defined as shared classes, while the classes only belonging to source domain as outlier classes. Source classifier G_y is trained on source domain. The output of this classifier is a $|C_s|$-dimension vector, C_s is the number of source classes. Each element represents the possibility of the input sample belonging to the jth class. A basic assumption is that the difference between source outlier label space and target label space should significantly more obvious than the difference between source shared label space and target label space,even if the target label space disjoints with source label space, Thus, we utilize output of G_y for target samples to evaluate the possibility of the source images belonging to target label space. It is can be defined as shard-class weight w_c, which is described as follows:

$$w_c = \frac{1}{n_t}\sum_{i=1}^{n_t} G_y(G_f(x_i^t)) \tag{2}$$

What's more,w_c is normalized by dividing its largest element to eliminate the influence of extremum. Hence, the objective of our framework is as follow:

$$\begin{aligned} minL(G_d, G_f, G_y) &= \frac{1}{n_s}\sum_{i=1}^{n_s} w_c L_y(G_y(G_f(x_i^s)), y_i^s) \\ &- \frac{1}{n_s}\sum_{i=1}^{n_s} w_c[logG_d(G_f(x_i^s))] \\ &- \frac{1}{n_t}\sum_{i=1}^{n_t}[log(1 - G_d(G_f(x_i^t)))] \end{aligned} \tag{3}$$

Compared with DaNN, this new framework adds a shared-class weight w_c on the source classifier G_y and domain discriminator G_d when it works in source domain. This new measure can reduce negative transfer caused by outlier source classes while extract domain-invariant features from shared source and target space.

3.3 Limitation of present methods

Shared-class weight w_c is presented to down-weight outlier classes while up-weight shared classes. As shown in 3.2, each element in w_c represents the possibility of the input satellite image belonging to the jth class. In ideal situation, the element subjecting to correct class should be as high as possible while the other (not matching) should be very low. The training processing aims to minimize the cross-entropy of the source satellite samples and applies this trained distribution to the target domain. However, during the training processing, the lower score values for the elements of the vector are not penalized. This phenomenon may cause relavant low score values for the correct satellite image classes during inference. In supervised learning processing, this phenomenon may not cause bad outcome because as long as the correct class get the highest scores. However, when it comes to w_c, which aims to evaluate whether a class subject to outlier classes, this behavior causes a problem that having low scores for the positive class will result in false negatives.

Meanwhile, as long as the correct class produces the highest value, the cross-entropy loss does not penalize values of unrelated classes. As a result, even if weights of shared classes are obviously higher than those of outlier classes, few of them can achieve desired values. Inaccurate cross-class relationships are encouraged during training. The deviation between predicted and ideal weights may misalign the features of outlier source classes and target classes. Hence, it is necessary to force the weights in shared classes close to 1, while the weights in outlier classes almost to 0 to wipe out negative transfer.

3.4 Coordinate loss

In our approach, we present a novel coordinate loss to deal with the above limitation we mentioned. The basic assumption of our method is that if we identify the sample belongs to a class, the element of this corresponding class in w_c should be 1 otherwise should be 0.

Because w_c is gained from the source classifier, the more reasonable the output of the source classifier is, the more reasonable the w_c is. Hence, we plan to utilize the coordinate loss L_c to force only the most possible class get an obvious high weight while others are close to 0.

We interpret each transformed activation score as the probability of the input image belonging to each individual class. In source domain, we have a ground truth label. And the output of the source classifier is a $|C_s|$-dimension vector, C_s is the number of source classes. Each element represents the possibility of the input sample belonging to the jth class. The network learns possibilities for the membership of each class as follow:

$$P(y=i) = G_y(G_f(x_i^t))_(y=i) \quad i\epsilon(1,2,\cdots c) \tag{4}$$

We define the possibility of the correct class as R_c and the possibility of the wrong class as R_w. In this way, coordinate loss as the risk of classification is defined as

$$L_c = R_c + \alpha * R_w \tag{5}$$

α is a trade-off parameter.

In our formulation, we hope the possibility of the correct class should be as high as possible, Thus, we define R_c as follow:

$$\begin{aligned} R_c &= \sum_{i=1}^{n}\sum_{s=0}^{c}(1-P(y=s))^2 \\ &= \sum_{i=1}^{n}\sum_{s=0}^{c}(1-G_y(G_f(x_i^t))_(y=s))^2 \end{aligned} \tag{6}$$

By minimizing R_c, the possibility of the correct class can be forced to be high. Meanwhile, we hope the weights of the incorrect classes should be as low as possible. Hence, similar to R_c, R_w can be defined as:

$$\begin{aligned} R_w =& \frac{1}{(C-1)}(P(y\neq s))^2 \\ =& \frac{1}{(C-1)}\sum_{i=1}^{n}\sum_{s=0}^{c}((G_y(G_f(x_i^t))_(y\neq s))^2 \end{aligned} \tag{7}$$

By substitution, coordinate loss is presented as follow:

$$\begin{aligned} L_c =& \sum_{i=1}^{n}\sum_{s=0}^{c}(1-G_y(G_f(x_i^t))_(y=s))^2 \\ &+\alpha * \frac{1}{(C-1)}\sum_{i=1}^{n}\sum_{s=0}^{c}((G_y(G_f(x_i^t))_(y\neq s))^2 \end{aligned} \tag{8}$$

Then, the final objective of CPADA is defined as:

$$
\begin{aligned}
& minL(G_d, G_f, G_y) \\
&\quad = \frac{1}{n_s}\sum_{i=1}^{n_s} w_c L_y(G_y(G_f(x_i^s)), y_i^s) \\
&\quad - \frac{1}{n_s}\sum_{i=1}^{n_s} w_c[logG_d(G_f(x_i^s))] \\
&\quad - \frac{1}{n_t}\sum_{i=1}^{n_t} [log(1 - G_d(G_f(x_i^t)))] \\
&\quad + \sum_{i=1}^{n}\sum_{s=0}^{c} (1 - G_y(G_f(x_i^t))_(y = s))^2 \\
&\quad + \alpha * \frac{1}{(C-1)}\sum_{i=1}^{n}\sum_{s=0}^{c} ((G_y(G_f(x_i^t))_(y \neq s))^2
\end{aligned}
\tag{9}
$$

After adding the coordinate loss, only one element of the source classifier close to 1 while others are close to 0. When this classifier is utilized to evaluate w_c, target samples that produce small weights will also be penalized regardless of the weights of the other classes. Moreover,when the coordinate loss is used together with cross-entropy loss, as the network is being trained, the most possible target class gets a relatively higher weight while others are close to 0.

Hence, when we apply the coordinate loss into our framework, the weights of shared classes are higher than these without coordinate loss, while the weights of outlier classes are mostly close to 0. This improvement can effectively eliminate negative transfer.

4 Experiment

We conduct our experiment on two datasets Office-31 dataset and NWPU-Merced-Land satellite image dataset. Office-31 dataset is utilized to compare our method with other classical unsupervised domain adaptation approaches to illustrate the effect of our method. Then, we apply our algorithm to an unsupervised satellite image classification issue.In this case,NWPU-RESISC-45 satellite image dataset with label is set as source domian, UC Merced Land-19 dataset is a without label is set as target domian.

4.1 Set up

Office-31 dataset is a standard benchmark dataset for unsupervised domain adaptation problem. This dataset includes 4652 images in 31 categories. There are three domains, named A, W, D, collected from amazon.com (A), DSLR (D) and web camera (W). we select the ten categories belonging to both office-31 and Caltech-256 dataset to build new target domain. In source domain there are 31 categories while in target domain there are only 10 categories. We compare the accuracy of several standard domain adaptation and partial domain adaptation methods with ours to evaluate our approach.

Figure 2: Sample images of UC Merced Land dataset

NWPU-Merced-Land satellite image dataset includes two part, NWPU-RESISC-45 satellite image dataset, which contains 45 common ground object categories with label, while UC Merced Land-19 dataset contain 19 categories without label. It is worthwhile to notice that the 19 categories is a subset of NWPU-RESISC45 label space. In this task, we set NWPU-RESISC45 satellite image dataset as source domain and UC Merced Land-19 dataset as target domain. sample images of UC Merced Land-19 dataset are shown in Figure 2. Our purpose is to classify the UC Merced Land satellte images,we introduce the NWPU-RESISC-45 satellite image dataset to assist us to apply our CPADA.Moreover,we compare the accuracy of our method to other traditional unsupervised machine learning methods and deep learning method to evalute our approach.

4.2 Result

The classification results based on ResNet-50 network are measured on the six classification tasks on Office-31 dataset and unsupervised satellite image classification task on NWPU-Merced-Land satellite image dataset are respectively show in Table 1 and Table 2. It is obvious that CPADA outperform than others in terms of accuracy, showing that our method performs well on different benchmark and improves the effect of algorithm in unsupervised satellite image classification issues.

Table 1.Accuracy of Partial Domain Adaptation Tasks on Office-31 Dataset

Methods	Office-31						
	A31 → W10	**D31 → W10**	**W31 → D10**	**A31 → D10**	**D31 → A10**	**W31 → A10**	Avg
ResNet[7]	75.59	96.27	98.09	83.44	83.92	84.97	87.05
DAN[10]	59.32	73.90	90.45	61.78	74.95	67.64	71.34
DaNN[8]	73.56	96.27	98.73	81.53	82.78	86.12	86.50
IWAN[13]	76.27	98.98	**100.00**	78.98	89.46	81.73	87.57
SAN[15]	81.82	98.64	**100.00**	81.28	80.58	83.09	87.27
PADA[16]	86.54	99.32	**100.00**	82.27	92.69	95.41	92.69
CPADA	**87.80**	**100.00**	**100.00**	**87.26**	**93.74**	**95.02**	**94.03**

Table 2.Accuracy of Unsupervised Satellite Image classification Tasks on NWPU-RESISC-UC-Merced-Land Satellite Image Dataset

Methods	NWPU-RESISC-UC-Merced-Land
ResNet[7]	69.78
DaNN[8]	51.36
PADA[16]	88.74
CPADA	**89.96**

From the results, we have some insightful observations. Based on Table 2, we can find out 1) DaNN performs worse than ResNet, which implies that the misalignment of domains has a negative transfer on results.2) RTN applies the entropy minimization criterion to their classifier, which is helpful to restrain negative transfer.3) Our scenario is more accurate than previous partial domain adaptation methods because of novel coordinate loss L_c.

Table 2 shows the promotion of our method in unsupervised satellite image classification. 1) ResNet-50 utilizing finetune performs not too bad and the accuracy cannot satisfy our need. 2) The standard domain adaptation method DaNN performs worse than finetuned ResNet-50 because of negative transfer. 3) On account of negative transfer having been reduced, the accuracy of PADA has achieved over 88%. 4) CPADA presents coordinate loss to reduce the influence of outlier classes. The accuracy of our method has achieved almost 90%, which is the state-of-the-art on unsupervised satellite image classification task. 5) if we remove the coordinate loss, CPADA degrades into PADA, the comparison between these two methods also shows the role of coordinate loss.

4.3 Analysis

Classes Weights: Fig. 3(a)-3(c) are weights histograms of source classes on the unsupervised satellite image classification task using ResNet-50, DaNN and CPADA. The blue bins represent shared classes, while the red ones are outlier classes.

ResNet-50 can give most shared-classes higher weights owing to the effect of finetune, even if the distinction between shared classes and outlier classes are not obvious. DaNN barely classifies outlier weights because of the negative transfer. Meanwhile, the gap between outlier and shared is more obvious which is conducive to avoid negative transfer. Our method can assign much larger weights to the shared classes while much lower weights to the outlier ones compared with previous methods.

Feature Visualization: We visualize the t-SNE embeddings of the bottleneck layer learned by ResNet-50, DaNN and CPADA on the unsupervised satellite image classification task in Fig. 4(a)-4(c). From Fig.4(a),the ResNet can classify some outlier samples but performs far away from perfect. Fig.4(b) illustrates DaNN can distinguish source domain samples well while target samples are mixed to every cluster that they cannot be discriminated. Compared with above scenarios, CPADA classifier in Fig.4(c) can classify target samples better, most samples are aggregated to correct classes. The result of t-SNE shows that our model can better discrim-

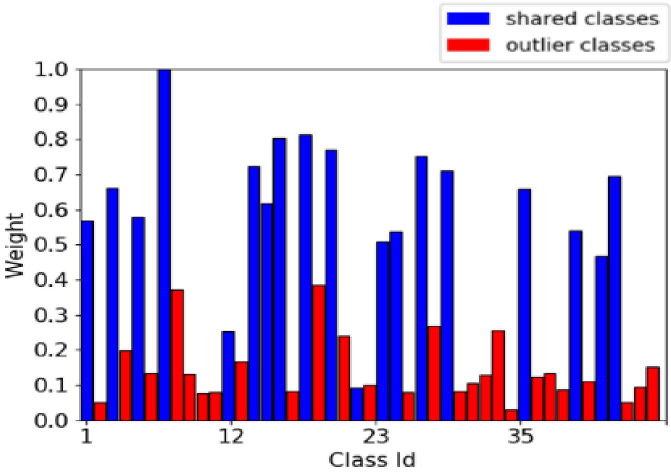


Figure 3 (a): ResNet-50

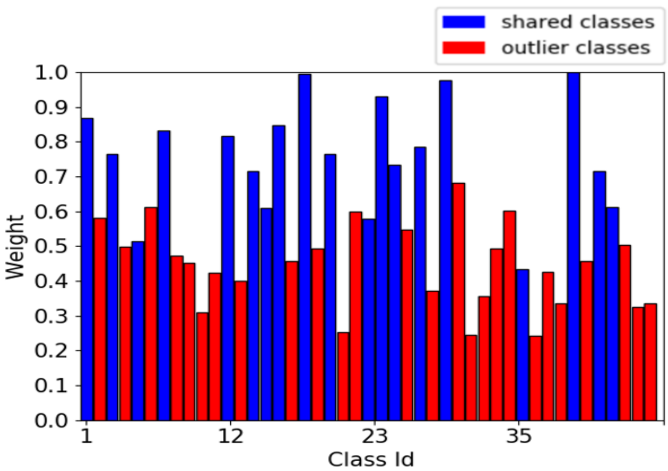


Figure 3 (b): DaNN

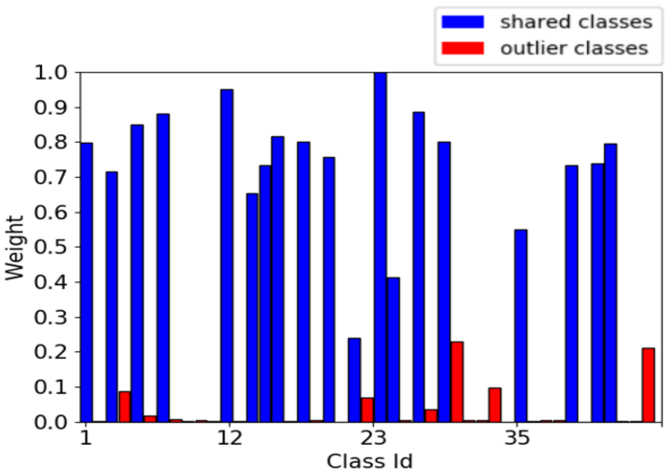


Figure 3 (c): CPADA

Figure 3:Histograms of class weights learned by ResNet-50, DaNN and CPADA on unsupervised satellite image classification tasks

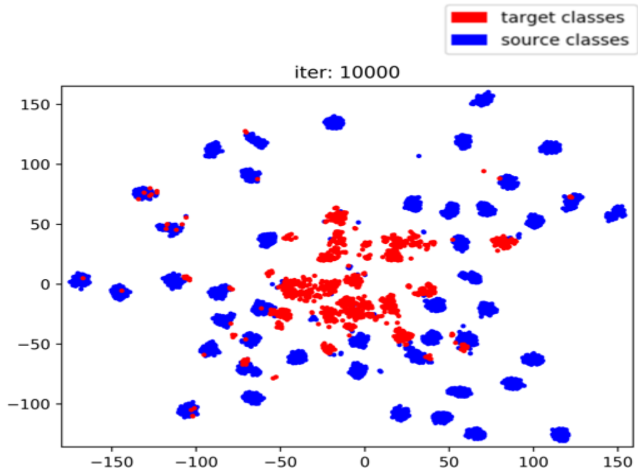


Figure 4 (a): ResNet-50

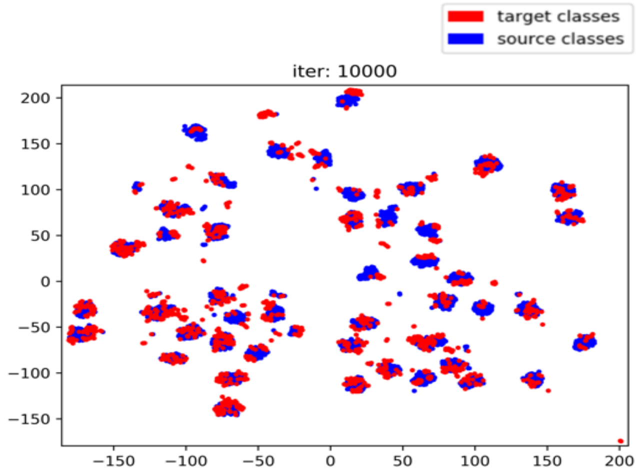


Figure 4 (b): DaNN

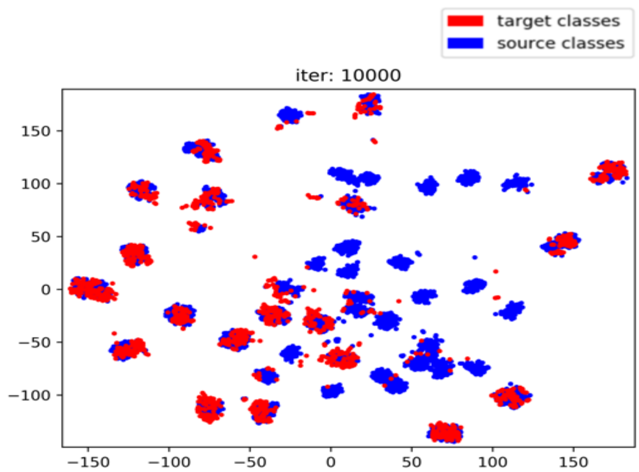


Figure 4 (c): CPADA

Figure 4:The t-SNE visualization of ResNet-50, DaNN and CPADA with domain information

inate both source and target examples than compared methods.

5 Conclusion

In this paper, we present a novel CPADA for unsupervised satellite image classification issues. Unlike previous unsupervised satellite image classification methods, the proposed approach introduces partial transfer learning method into this field,the novel coordinate loss can eliminate negative transfer by down weighting outlier satellite image classes and promote weights of shared satellite image classes. On a classical unsupervised satellite image classification task, our method outperforms 30% compared with traditional machine learning ones, and also gets the state-of-the-art result compared with partial transfer learning ones.

References

1. M. Tyagi, F. Bovolo, A. K. Mehra, S. Chaudhuri, and L. Bruzzone, A context-sensitive clustering technique based on graph-cut initialization and expectationmaximization algorithm, IEEE Geosci. Remote Sens. Lett., vol. 5, no. 1, pp. 2125, Jan. 2008.
2. I. A. Rizvi and B. K. Mohan, Object-based image analysis of high-resolution satellite images using modified cloud basis function neural network and probabilistic relaxation labeling process, IEEE Trans. Geosci. Remote Sens., vol. 49, no. 12, pp. 48154820, Dec. 2011.
3. S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, and J. C. Tilton, Learning Bayesian classifiers for scene classification with a visual grammar, IEEE Trans. Geosci. Remote Sens., vol. 43, no. 3, pp. 581589,Mar. 2005.
4. R. Bellens, S. Gautama, L. Martinez-Fonte, W. Philips, J. C. W. Chan, and F. Canters, Improved classification of VHR images of urban areas using directional morphological profiles, IEEE Trans. Geosci. Remote Sens., vol. 46, no. 10, pp. 28032813, Oct. 2008.
5. M. D. Zeiler, G. W. Taylor, and R. Fergus, Adaptive deconvolutional networks for mid and high-level feature learning, in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011, pp. 20182025.
6. O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2015, pp. 4451.
7. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
8. Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. S. Lempitsky. Domain-adversarial training of neural networks. Journal of Machine Learning Research, 17:59:159:35, 2016.
9. M. Long, H. Zhu, J. Wang, and M. I. Jordan. Unsupervised domain adaptation with residual transfer networks. In Advances in Neural Information Processing Systems, pages 136144, 2016.
10. M. Long, Y. Cao, J. Wang, and M. I. Jordan. Learning transferable features with deep adaptation networks. In International Conference on Machine Learning (ICML), 2015
11. M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, Deconvolutional networks, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 25282535.
12. Y. Bengio, A. Courville, and P. Vincent, Representation learning: A review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 17981828, Aug. 2013.
13. J. Zhang, Z. Ding, W. Li ,and P. Ogunbona. Importance weighted adversarial nets for partial domain adaptation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.
14. A. M. Cheriyadat, Unsupervised feature learning for aerial scene classification, IEEE Trans. Geosci. Remote Sens., vol. 52, no. 1, pp. 439451, Jan. 2014.
15. Z. Cao, M. Long, J. Wang, and M. I. Jordan. Partial transfer learning with selective adversarial networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.
16. Z. Cao, L. Ma, M. Long, and J. Wang. Partial adversarial domain adaptation. In European Conference on Computer Vision (ECCV), September 2018.
17. M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva. (2015). Land use classification in remote sensing images by convolutional neural networks. [Online]. Available: https://arxiv.org/abs/1508.00092
18. R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
19. S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang. Domain adaptation via transfer component analysis. IEEE Transactions on Neural Networks and Learning Systems (TNNLS), 22(2):199210, 2011
20. F. Hu, G.-S. Xia, Z. Wang, X. Huang, L. Zhang, and H. Sun, Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification, IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 5, pp. 20152030, May 2015.
21. E. Tzeng, J.Hoffman, N.Zhang, K.Saenko, and T.Darrell. Simultaneous deep transfer across domains and tasks. In IEEE International Conference on Computer Vision (ICCV), 2015.
22. E. Tzeng, J.Hoffman,N. Zhang , K.Saenko, and T.Darrell. Deep domain confusion: Maximizing for domain invariance. 2014.