Unsupervised Domain Adaptation for Remote Sensing Images Using Metric Learning and Correlation Alignment

Aniruddha Mahapatra[1(✉)] and Biplab Banerjee[2]

[1] Indian Institute of Technology, Roorkee, India
amahapatra@cs.iitr.ac.in
[2] Indian Institute of Technology Bombay, Mumbai, India
bbanerjee@iitb.ac.in

Abstract. We address the problem of domain adaptation (DA) in the context of remote sensing (RS) image classification in this paper. By definition, the problem of unsupervised DA aims at classifying samples from a *target* domain which is strictly devoid of any label information while assuming that enough training data are available from a related yet non-identical (in terms of data distributions) *source* domain. A number of existing approaches in this regard are focused towards matching the underlying distributions of the data from both the domains in a shared latent space without explicitly considering: i) the discriminativeness of the embedding space, ii) the usefulness of a manifold distance is pulling the domains towards each other over the standard Euclidean measures. However, we argue the importance of both the aspects in learning the latent space, particularly for fine-grained classes. Our model jointly optimizes both the terms in an end-to-end fashion and the learned latent space is found to properly align the classes with high precision. Experimental results obtained on a hyper-spectral and a multi-spectral dataset confirm the superior performance of the approach over a number of techniques from the literature.

Keywords: Domain adaptation · Metric learning · Representation learning · Remote sensing.

1 Introduction

Remote sensing (RS) image analysis [1] is currently considered an active field of research, thanks to the ample amount of data acquired by a wide range of satellite-onboard sensors periodically. However many of the machine learning algorithms (predominantly supervised methods), that work on RS images inherently assume that the training and test samples are drawn from similar underlying distributions, which is often violated in analyzing multi-temporal RS images where the land-cover properties change due to seasonal effects, presence of cloud

R. V. Babu et al. (Eds.): NCVPRIPG 2019, CCIS 1249, pp. 133–142, 2020.
https://doi.org/10.1007/978-981-15-8697-2_12

covers during the imaging process (considering passive sensors for data acquisition), man-made changes on ground, from one image to another.

Generation of training samples, in general, is a costly and time-consuming process and the challenges proportionately grow for the case of multi-temporal RS image sequences with many images. However considering that the labeled samples are available for some of the images in the sequence, the problem can now be formulated to classifying the images with no prior label information by judicious knowledge transfer from the images with available training data. Domain Adaptation (DA) [3] is a popular inductive transfer learning approach to handle such a critical classification framework.

DA techniques are used to build a classifier taking into consideration the mismatch in the distributions governing the training and test data. The training domain is generally termed as the *source* domain which is accumulated with an ample amount of labeled data (images with training samples) while the test domain is termed as the *target* domain. Note that the target domain may or may not contain any label information, based on which DA can further be classified as semi-supervised or unsupervised DA, respectively. Our focus is on the challenging unsupervised DA setup which tries to compensate for the degradation in classification performance by transferring knowledge from the labeled source domain to the unlabeled target domain.

Historically, machine learning literature is rich in DA techniques. There have been several endeavors towards feature adaptation (making the source and target features overlapping in some latent space) and classifier adaptation (making the classifier trained on the source domain samples to gradually adapt to the target domain properties). However, the classifier adaptation based techniques seldom suffer from the problem of *source forgetting* since the classifier iteratively finds its bias towards the target domain. The feature adaptation based approaches are resilient to such bottlenecks and have shown better performance in diverse scenarios including image, speech, text, to name a few.

Amongst the feature adaptation based approaches, the current trend to overcome the problem of domain mismatch by mapping the cross-domain samples on a shared subspace using deep neural network-based modules which make the entire process data-driven. This is generally achieved by minimizing some measure of domain variance, like the Maximum Mean Discrepancy (MMD) [7]. Likewise, a recently proposed Deep CORAL method [5] aligns second-order statistics of the source and the target distribution by constructing a differentiable loss function that minimizes the difference between the source and target correlations which the authors termed as the CORAL loss. However, to the best of our knowledge, the majority of such methods do not consider the discriminativeness of the learned embedding space, which nonetheless causes misclassification for fine-grained classes. Hence, the notion of learning a discriminative subspace is of prime interest for improved cross-domain classification.

Contributions: Inspired from the aforementioned discussions, we prose an end-to-end trainable neural network-based unsupervised DA module that learns a

shared embedding space for the samples of both the domains which are also deemed to be discriminative. In particular, while a contrastive loss measure [6] is considered on the latent source domain labeled samples which is responsible to make the latent representations compact class-wise, simultaneously the domain difference is reduced in terms of minimizing the difference of the between-domain higher-order statistics. Experimental results obtained on two benchmark RS datasets showcase the superiority of the proposed DA module over several standard approaches.

2 Related Work

Unsupervised DA techniques are extensively applied in handling multi-temporal RS images on the ground captured at different time instances. Ad-hoc approaches for unsupervised DA usually consist of matching the feature distribution between the source and target domain [10,11] by exploring standard distance measures between distributions. These methods can generally be divided into two categories: (i) sample re-weighting [12,13] and (ii) feature space transformation [15,16], respectively. Besides, techniques based on metric learning [19], subspace alignment [20] and nonlinear transformation based on graph node matching [21] are explored in conjunction with different cross-domain RS images.

On the other hand, adaptive deep neural networks have recently been explored for unsupervised DA for image classification. For example, DLID [22] trains a joint source and target CNN architecture with two adaptation layers. Similarly, DDC [23] applies a single linear kernel to one layer to minimize the Maximum Mean Discrepancy (MMD). Deep CORAL applies CORAL loss to minimize the difference in learned feature covariances across source and target domain. Deep LogCORAL similar to the Deep CORAL method, proposes to use the Riemann distance, approximated by Log-Euclidean distance to replace the naive Euclidean distance in Deep CORAL.

We propose a metric learning-based approach that minimizes the feature covariances of the source and the target domain (using CORAL and LogCORAL losses separately) and at the same time form better clustering of similar images by Contrastive loss on labeled source data for better fine-grained classification of multi-temporal RS images.

3 Method

We consider the unsupervised DA situation where there are no labels associated with the target domain data distribution. Since the RS image data can be very challenging, like the Botswana dataset, with overlapping classes and very small feature vector for each image, we first try to learn a similarity-based classifier to obtain good clustering of data based on their respective classes to avoid overlap. For the second goal of reducing covariance between the source and target domain, we propose to minimize the difference in second-order statistics between the source and target feature activations, i.e. the CORAL/logCORAL loss. Figure 1

shows a sample architecture of our proposed model. The two losses are trained end-to-end.

This approach is based on the assumption that minimizing the difference between the second-order statics would bring the target domain to overlap with the source domain. Since cross-domain data point is inherently closer to their respective classes than of different classes of different domains, similarity-based learning on the source domain data would also make nice clusterings for target domain data points in the domain invariant feature space. Joint training with both the losses is likely to learn features that work well on the target domain:

$$L_{Total} = KL_{Source} + \lambda(1 - K)L_{CrossDomain} \tag{1}$$

where covariance loss denotes CORAL or logCORAL loss, $K = 1$ if pairs are from the same domain, $K = 0$ if pairs are from a different domain. λ is a trade off between domain adaptation and clustering accuracy on the source domain.

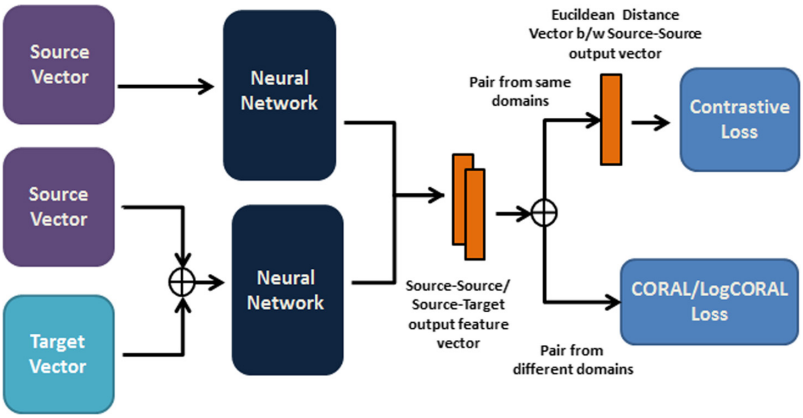

Fig. 1. Sample structure of our model. We determine whether the inputs are from same domain or not using the similarity labels of pair. If the input pairs are both from source domain, then we compute the Source Loss based on their class labels. Otherwise if one vector from source and other from target domain, we compute the Cross-Domain Loss.

3.1 Source Loss

For the purpose of finding a function that maps input patterns to the domain invariant space based on neighborhood relationship between samples, we use Contrastive Loss. The basis of the creation of clusters are the class labels of source domain data inputs. This loss function works on pairs of samples. Let $\boldsymbol{X_1}, \boldsymbol{X_2} \in \mathbb{I}$ be a pair of input source domain vectors. Let Y be a binary label assigned to this pair. $Y = 0$ if $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ are deemed similar, and $Y = 1$ if they are deemed dissimilar. Define the parameterized distance function to be learned

D_W between $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ as the Euclidean distance between the outputs of G_W, where G_W is the output values of $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ from the network. That is,

$$D_W(\boldsymbol{X_1}, \boldsymbol{X_2}) = ||G_W(\boldsymbol{X_1}) - G_W(\boldsymbol{X_2})|| \tag{2}$$

To shorten notation, $D_W(\boldsymbol{X_1}, \boldsymbol{X_2})$ is written D_W. Then the loss function in its most general form is:

$$\kappa(W) = \sum_{i=1}^{P} L(W, (Y, \boldsymbol{X_1}, \boldsymbol{X_2})^i) \tag{3}$$

$$L(W, (Y, \boldsymbol{X_1}, \boldsymbol{X_2})^i = (1 - Y)L_S(D_W^i) + Y \max(0, m - L_D(D_W^i)) \tag{4}$$

where $(Y, \boldsymbol{X_1}, \boldsymbol{X_2})^i$ is the ith labeled sample pair, L_S is the partial loss function for a pair of similar points, L_D is the partial loss function for a pair of dissimilar points, and P the number of training pairs (which may be as large as the square of the number of samples). m is the margin for dissimilar sample pair. The value of m for this experiment was taken as 1.0.

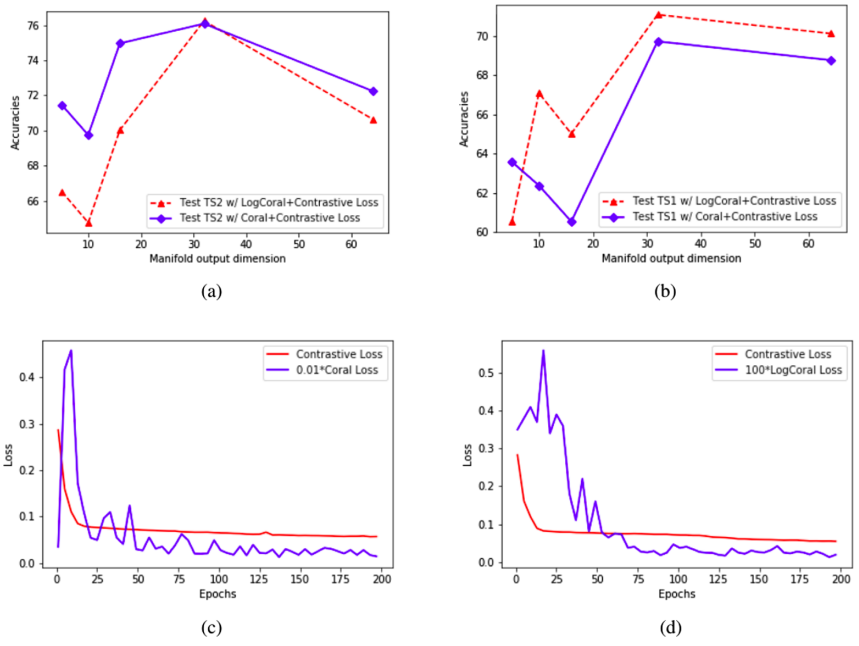

Fig. 2. Accuracy of model (a) Trained on TR1 (source) and TR2 (target) and Test on TS2, (b) Trained on TR2 (source) and TR1 (target) and Test on TS1 for different manifold output dimensions. (c) The loss curve for Contrastive and CORAL Loss, (d) Contrastive and LogCORAL Loss trained on TR2 (source) and TR1 (target) and test on TS1.

3.2 Cross-Domain Loss

Let us denote $\mathbf{D_S} = [\mathbf{x_1}, ..., \mathbf{x_{n_S}}]$ as the source domain features extracted from the last layer of the model as shown in the Fig. 1, where $\mathbf{x_i}$ is the *ith* source sample, and $\mathbf{D_T} = [\mathbf{u_1}, ..., \mathbf{u_{n_T}}]$ as the unlabeled target domain feature extracted from the final layer, where $\mathbf{u_i}$ is the *ith* training sample. To overlap the target source domain features with their corresponding source domain features, we try to reduce the covariance between them. For this purpose, we found that the most appropriate loss metrics to be CORAL and logCORAL loss and trained our model on them separately to compare performance.

CORAL loss [5] calculates the distance of second-order statics between the two domains. The covariance matrix of features is calculated from the output of the final layer for each domain which then minimizes the Euclidean distance of the covariance matrices of two domains. The CORAL loss is defined as follows:

$$L_{CORAL} = \frac{1}{4d^2}||\mathbf{C_S} - \mathbf{C_T}||^2 \tag{5}$$

in which the covariance matrices $\mathbf{C_S}$ and $\mathbf{C_T}$ are defined as follows:

$$\mathbf{C_S} = \frac{1}{n_S - 1}(\mathbf{D_S^T D_S} - \frac{1}{n_S}(\mathbf{1^T D_S})^\mathbf{T}(\mathbf{1^T D_S})) \tag{6}$$

$$\mathbf{C_T} = \frac{1}{n_T - 1}(\mathbf{D_T^T D_T} - \frac{1}{n_T}(\mathbf{1^T D_T})^\mathbf{T}(\mathbf{1^T D_T})) \tag{7}$$

where n_S, n_T is the batch size of the source domain and target domain respectively. d is the feature dimension, and 1^T is a vector that all elements equals to 1.

[28] shows that measuring matrix distance on Riemannian manifold may be more precise that Euclidean manifold and give better results in problems related to Domain Adaptation. According to the above assumption, the logCORAL loss is a distance measure in the Riemannian manifold. The logCORAL distance is defined as the Euclidean distance between the logarithm of covariance matrices:

$$L_{CORAL} = \frac{1}{4d^2}||log(\mathbf{C_S}) - log(\mathbf{C_T})||^2 \tag{8}$$

where the $log()$ operation is the logarithm of the covariance matrix and $\mathbf{C_S}$ and $\mathbf{C_T}$ are the covariance matrix for source and target domain respectively as defined in CORAL loss.

Minimizing the Source loss itself is going to over-fit for the source domain and give poor clusterings for the target domain. On the other hand, reducing the Cross-Domain loss would lead to degenerate features as the network will project all the source and target data to a single point leading to zero Cross-Domain loss. Jointly training the network on Cross-Domain loss with Source loss will make the network learn the desirable domain invariant mapping into feature space. We show that these two losses play counterparts and reach an equilibrium at the end of the training, where the final features are expected to work well on the target domain.

Table 1. Performance comparison with baseline models in domain adaptation. The first column gives test accuracy on TS2 with training on TR1 (source) + TR2 (target) while, second column gives test accuracy on TS1 with training on TR2 (source) + TR1 (target). The second last model uses CORAL w/ CONTRASTIVE Loss. The last model achieves state of the art result LogCORAL w/ CONTRASTIVE Loss.

Model	TR1 → TS2	TR2 → TS1
TCA	69.88	61.00
GFK	72.89	65.50
CORAL	54.54	47.36
SA	72.88	68.52
STK	75.28	70.20
BDA	62.52	50.72
Auto-encoder	67.46	62.69
Our (CORAL)	76.08	69.73
Our (LogCORAL)	**76.24**	**71.09**

4 Experiment

4.1 Dataset

We validate the proposed framework on commonly available but challenging Botswana hyper-spectral dataset acquired by the Hyperion sensor of the EO-1 satellite over a 1476 256 pixel study area located in the Okavango Delta, Botswana. 10 bands among the set of 145 bands are selected based on their discrimination capability using the method mentioned in [29]. Here, 14 land-cover classes are identified for two different spatially disjoint areas. There are a total of 4 sets in this data (2 for each domain). TR1 (train) and TS1 (test) for one domain and TR2 (train) and TS2 (test) for the second domain. We use all the labeled source data and all the target unlabeled data.

The second dataset contains two scenes acquired by the ROSIS sensor during a flight campaign over Pavia, northern Italy. The number of spectral bands is 102 for Pavia Centre and 103 for Pavia University. To make both of them of equal dimension, we use PCA to reduce the number of dimensions to 50. The geometric resolution is 1.3 m. Both image ground-truths differentiate 9 classes each out of which we use 7 classes common to both.

4.2 Model Specifications

In this experiment, we apply the Source loss and Cross-Domain loss from the features extracted from the last layer ($fc3$) of the model. In training for both losses, we set the batch size to 32, the learning rate to 10^{-3}, weight decay to 0. We train a small 3-layer fully-connected neural network on the features extracted from $fc3$ for source domain for classification purpose (Table 2).

Table 2. Performance comparison with our models with Auto-Encoder in domain adaptation. The first column gives test accuracy on PaviaU with training on PaviaCentre while, second column gives test accuracy on PaviaCentre with training on PaviaU. The second last model uses CORAL and CONTRASTIVE Loss. The last model achieves state of the art result LogCORAL w/ CONTRASTIVE Loss.

Model	PaviaC → PaviaU	PaviaU → PaviaC
Auto-Encoder	46.14	46.64
D-CORAL	47.72	45.86
D-LogCORAL	**47.81**	**56.32**

4.3 Choosing $fc3$ (final) Layer Dimension

To decide what is the optimal dimension of the $fc3$ layer for the Botswana and Pavia dataset, we train different networks with output dimension sizes of 5, 10, 16, 32 and 64. Based on the accuracies from the classifier applied to the output of $fc3$ for the domain shifts on both datasets, we get the maximum accuracy for the $fc3$ layer with dimension 32. In Fig. 2(a) and 2(b) we have plotted the accuracies from different output layer dimensions for the Botswana dataset.

4.4 Performance Between CORAL and LogCORAL Loss

We try using CORAL and logCORAL loss separately for reducing covariance across domains. From accuracy results, we find that logCORAL loss outperforms CORAL loss by a narrow margin of around 0.2%−0.3% (source TR2, target TR1, test TS1: CORAL accuracy-69.73%, logCORAL accuracy-71.09% and source TR1, target TR2, test TS2: CORAL accuracy-76.08%, logCORAL accuracy-76.24%) for Botswana Dataset and relatively a large margin of 9% on Pavia Dataset (source PaviaU, target PaviaC: CORAL accuracy-47.81%, logCORAL accuracy-56.32%).

4.5 Comparison with Popular da Techniques

We compare the generalization performance of the proposed framework with that of six popular and diverse unsupervised DA techniques from the literature as follows:

a) TCA [9]
b) Subspace alignment (SA) based DA [14]
c) GFK-based subspace projection [35]
d) CORAL with SVM
e) STL
f) BDA

GFK, SA, and TCA are manifold based methods that project the source and target distributions into a lower-dimensional manifold and are not end-to-end deep methods. In all the cases, we first project the data in the embedding space and further design a multiclass SVM classifier (with RBF kernel) in the new space exploiting the projected source domain training samples. The classifier is further evaluated on the projected target domain test samples. From Table 1 we see that our model achieves better performance for both domain shifts than the 6 baseline methods with a relative margin of 1–2%. We can see that even though Cross-Domain loss is not always decreasing, it gets to a relatively stable state after a few epochs.

5 Conclusion

In this paper, we propose a novel neural network architecture for unsupervised domain adaptation. We show that semantic similarity learning with the reduction in covariance between the source and target dataset can outperform the 'standard' convolution neural network in the domain adaptation problem. The method is feasible and simple. We demonstrate that this model achieves state-of-the-art performance on the Botswana dataset. Future works include using a similarity learning that adaptively assesses similarity based on distribution on representation space rather than penalizing individual pairs based on the notion of labels and with different domain discrepancy reduction algorithms (such as MMD and its variants). It will also be of interest to determine how this model performs on very high-resolution satellite images.

References

1. J. A. Richards and J. Richards, Remote Sensing Digital Image Analysis. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-30062-2_9
2. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. ICML, 2014
3. Patel, V.M., Gopalan, R., Li, R., Chellappa, R.: Visual domain adaptation: A survey of recent advances. IEEE Signal Process. Mag. **32**(3), 53–69 (2015)
4. A. Torralba and A. A. Efros. Unbiased look at dataset bias. In CVPR, 2011
5. Sun, Baochen, Saenko, Kate: Deep CORAL: Correlation Alignment for Deep Domain Adaptation. In: Hua, Gang, Jégou, Hervé (eds.) ECCV 2016. LNCS, vol. 9915, pp. 443–450. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49409-8_35
6. R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In CVPR, 2006
7. A. Gretton, K. Borgwardt, M. Rasch, B. Scholkopf, and A. Smola. A kernel two-sample test. JMLR, 2012
8. M. Long, H. Zhu, J. Wang, and M. I. Jordan. Deep transfer learning with joint adaptation networks. In ICML, 2017
9. Wang, Yifei, Li, Wen, Dai, Dengxin, Van Gool, Luc: Deep Domain Adaptation by Geodesic Distance Minimization. ICCV 2651–2657 (2017)

10. A. Gretton, A. Smola, J. Huang, M. Schmittfull, K. Borgwardt, and B. Scholkopf. Covariate shift and local learning by distribution matching. 2009
11. G. Csurka. A comprehensive survey on domain adaptation for visual applications. In Domain Adaptation in Computer Vision Applications. 2017
12. J. Jiang and C. Zhai. Instance weighting for domain adaptation in nlp. In ACL, 2007
13. J. Huang, A. Gretton, K. M. Borgwardt, P. B. Scholkopf, and A. J. Smola. Correcting sample selection bias by unlabelled data. In NIPS, 2007
14. B. Gong, K. Grauman, and F. Sha. Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation. In ICML, 2013
15. S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang. Domain adaptation via transfer component analysis. Transactions on Neural Networks. 2017
16. R. Gopalan, R. Li, and R. Chellappa. Domain adaptation for object recognition: An unsupervised approach. In ICCV, 2011
17. M. Baktashmotlagh, M. T. Harandi, B. C. Lovell, and M. Salzmann. Unsupervised domain adaptation by domain invariant projection. In ICCV, 2013
18. B. Sun, J. Feng, and K. Saenko. Return of frustratingly easy domain adaptation. In AAAI, 2016
19. Geng, B., Tao, D., Xu, C.: DAML: Domain adaptation metric learning. IEEE Trans. Image Process. **20**(10), 2980–2989 (2011)
20. B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. in Proc. IEEE Int. Conf. Comput. Vis., 2013
21. D. Tuia, J. Munoz-Mari, L. Gomez-Chova, and J. Malo. Graph matching for adaptation in remote sensing. IEEE Trans. Geosci. Remote Sens., vol. 51, no. 1, 2013
22. Chopra, S., Balakrishnan, S., Gopalan, R.: Dlid: Deep learning for domain adaptation by interpolating between domains. In: ICML Workshop, 2013
23. Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T.: Deep domain confusion: Maximizing for domain invariance. 2014
24. O. Rippel, M. Paluri, P. Dollr, and L. D. Bourdev. Metric learning with adaptive density discrimination. In ICLR, 2015
25. R. Salakhutdinov and G. E. Hinton. Learning a nonlinear embedding by preserving class neighborhood structure. In AISTSTS, 2017
26. F. Siyahjani, R. Almohsen, S. Sabri, and G. Doretto. A supervised low-rank method for learning invariant subspaces. In ICCV, 2015
27. Pedro, O.: Pinheiro. Unsupervised Domain Adaptation with Similarity Learning, In CVPR (2017)
28. Z. Huang and L. Van Gool. A riemannian network for spd matrix learning. 2016
29. Bruzzone, L., Persello, C.: A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability. IEEE Trans. Geosci. Remote Sens. **47**(9), 3180–3191 (2009)
30. Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q.: Domain adaptation via transfer component analysis. IEEE Trans. Neural Netw. **22**(2), 199–210 (2011)
31. B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 2960–2967
32. B. Gong, Y. Shi, F. Sha, and K. Grauman. Geodesic flow kernel for unsupervised domain adaptation. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2012, pp. 2066–2073