






# Class Reconstruction Driven Adversarial Domain Adaptation for Hyperspectral Image Classification

Shivam Pande<sup>1</sup>(✉) , Biplab Banerjee<sup>1</sup>(✉) , and Aleksandra Pižurica<sup>2</sup>(✉) 

<sup>1</sup> Center of Studies in Resources Engineering,  
Indian Institute of Technology Bombay, Mumbai, India  
184314002@iitb.ac.in, getbiplab@gmail.com

<sup>2</sup> Department of Telecommunications and Information Processing,  
Ghent University, Ghent, Belgium  
Aleksandra.Pizurica@ugent.be

**Abstract.** We address the problem of cross-domain classification of hyperspectral image (HSI) pairs under the notion of unsupervised domain adaptation (UDA). The UDA problem aims at classifying the test samples of a target domain by exploiting the labeled training samples from a related but different source domain. In this respect, the use of adversarial training driven domain classifiers is popular which seeks to learn a shared feature space for both the domains. However, such a formalism apparently fails to ensure the (i) discriminativeness, and (ii) non-redundancy of the learned space. In general, the feature space learned by domain classifier does not convey any meaningful insight regarding the data. On the other hand, we are interested in constraining the space which is deemed to be simultaneously discriminative and reconstructive at the class-scale. In particular, the reconstructive constraint enables the learning of category-specific meaningful feature abstractions and UDA in such a latent space is expected to better associate the domains. On the other hand, we consider an orthogonality constraint to ensure non-redundancy of the learned space. Experimental results obtained on benchmark HSI datasets (Botswana and Pavia) confirm the efficacy of the proposal approach.

**Keywords:** Domain adaptation · Adversarial training · Hyperspectral images

## 1 Introduction

The current era has witnessed the acquisition of a large volume of satellite remote sensing (RS) images of varied modalities, thanks to several national and international satellite missions. Such images showcase relevance in a range of important applications in areas including urban studies, disaster management, national security and many more. One of the major applications in this regard concerns

the analysis of (i) images of a given area on ground but acquired at different time instants, and (ii) images of different geographical areas but composed of similar land-cover types. Usually, it is non-trivial to generate training samples for all the images and hence it is a common practice to reuse the training samples obtained from images with *similar* characteristics to new images for carrying out the supervised learning tasks. To this end, the paradigm of inductive transfer learning, in particular domain adaptation, is extremely popular.

By definition, the unsupervised domain adaptation (UDA) techniques typically consider two related yet diverse data domains: a source domain  $\mathcal{S}$  equipped with ample amount of training samples, and a target domain  $\mathcal{T}$  where the test samples are accumulated. Since the data distributions are different for the two domains:  $P(\mathcal{S}) \neq P(\mathcal{T})$ , the classifier trained on  $\mathcal{S}$  fails to generalize for  $\mathcal{T}$  following the probably approximately correct (PAC) assumptions of the statistical learning theory [17, 18].

Traditional UDA techniques can broadly be classified into categories based on: (i) classifier adaptation, and (ii) domain invariant feature space learning. In particular, a common feature space is learned where the notion of domain divergence is minimized or a transformation matrix is modelled to project the samples of (source) target domain to the other counterpart [4, 13]. Some of the popular ad hoc methods in this category include transfer component analysis (TCA) [15], subspace alignment (SA) [6], geodesic flow kernel (GFK) [9] based manifold alignment etc. Likewise, UDA approaches based on the idea of maximum mean discrepancy (MMD) [20] learn the domain invariant space in a kernel induced Hilbert space. Recently, the idea of adversarial training has become extremely popular in UDA. Specifically, such approaches are based on a min-max type game between two modules: a feature generator ( $G$ ) and a discriminator ( $D$ ). While  $D$  tries to distinguish samples coming from  $\mathcal{S}$  and  $\mathcal{T}$ ,  $G$  is trained to make the target features indistinguishable from  $\mathcal{S}$  [11]. The RevGrad algorithm is of particular interest in this respect as it introduces a gradient reversal layer for maximizing the gradient of the  $D$  loss [7]. This, in turn, directs  $G$  to learn a domain-confused feature space, thus reducing the domain gap substantially. Adversarial residual transform networks (ARTN) [3] is another notable approach that uses adversarial learning in UDA. Besides, the use of generative adversarial networks (GAN) have been predominant in the recent past for varied cross-domain inference tasks: image style transfer, cross-modal image generation, to name a few. Some of the GAN based endeavors in this regard are: DAN [8], CycleGAN [5] and ADDA [19].

As the UDA problem is frequently encountered in RS, the aforementioned ad hoc techniques have already been explored in the RS domain [18]. A recent example [1] proposes a hierarchical subspace learning strategy which considers the semantic similarity among the land-cover classes at multiple levels and learns a series of domain-invariant subspaces. The use of a shared dictionary between the domains is also a popular practise for HSI pairs [21]. As far as the deep learning techniques are concerned, the use of GAN or domain independent convolution networks are also explored in this regard [2].

In this work, we specifically focus on the domain classifier (DC) based adversarial approach towards UDA. Precisely, the DC based UDA approaches simultaneously train the domain classifier and a source specific classifier using the feature generator-discriminator framework. While the domain classifier is entrusted with the task of making the domains overlapping, the source classifier helps in avoiding any trivial mapping. However, we find the following shortcomings of the standard DC based approaches: (i) the learned space does not encourage discriminativeness. In particular, the notion of intra-class compactness is not explicitly taken into account, which may result in overlapping of samples belonging to fine-grained categories. (ii) the learned space is ideally unbounded and does not convey any meaningful interpretation and may be redundant in nature.

In order to resolve both the aforementioned issues, we propose an advanced autoencoder based approach as an extension to the typical DC based UDA. In addition to jointly training the binary domain classifier and the source-specific multiclass classifier, we specifically add two other constraints on the learned latent space for the source specific samples. The first one is the reconstructive constraint that is directed to reconstruct one sample from another sample from  $\mathcal{S}$  both sharing the same class label. This essentially captures the classwise abstract attributes better than a typical autoencoder setup. Further, this loss helps in concentrating the samples from  $\mathcal{S}$  at the category level. The other one is the orthogonality constraint to ensure that the non-redundancy of the reconstructed features in the source domain. Optimization of all four loss measures together is experimentally found to better correspond  $\mathcal{S}$  and  $\mathcal{T}$ . The main contributions of this paper are:

- We introduce a class-level sample reconstruction loss for the samples in  $\mathcal{S}$  in a typical DC based UDA framework. This makes the learned space constrained and bounded.
- We enforce an orthogonality constraint over the source domain to keep the reconstructed features in the source domain non-redundant.
- Extensive experiments are conducted on the Botswana and Pavia HSI datasets where improved classification performance on  $\mathcal{T}$  can be observed.

The subsequent sections of the paper discuss the methodology followed by the experiments conducted and concluding remarks.

## 2 Methodology

In this section, we detail the UDA problem followed by our proposed solution.

**Preliminaries:** Let  $\mathcal{X}_S = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{N_S} \in X_S \otimes Y_S$  be the source domain training samples with  $\mathbf{x}_i^s \in \mathbb{R}^d$  and  $y_i^s \in \{1, 2, \dots, C\}$ , respectively. Likewise, let  $\mathcal{X}_T = \{(\mathbf{x}_j^t)\}_{j=1}^{N_T} \in X_T$  be the target domain samples obtained from the same categories as of  $\mathcal{X}_S$ . However,  $P_S(X_S) \neq P_T(X_T)$ . Under this setup, the UDA problem aims at learning  $f_S : X_S \rightarrow Y_S$  which is guaranteed to generalize well for  $\mathcal{X}_T$ .

In order to learn an effective  $f_S$ , we propose an end-to-end encoder-decoder based neural network architecture comprising of the following components: (i)

a feature encoder  $f_E$ , (ii) a domain classifier  $f_D$ , (iii) a source specific classifier  $f_S$ , and (iv) a reconstructive class-specific decoder  $f_{DE}$ . Note that the feature encoder is typically implemented in terms of the *fully-connected* (fc) layers with non-linearity. For notational convenience, we denote the encoded feature representation corresponding to an input  $\mathbf{x}$  by  $f_E(\mathbf{x})$ .

We elaborate the proposed training and inference stages in the following. A depiction of our model can be found in Fig. 1.

## 2.1 Training

Given the encoded feature representations, the proposed loss measure is composed of the losses from the following components in the decoder:

**Source Classifier  $f_S$ :** The mapping,  $f_S$  is a multiclass softmax classifier trained solely on  $\mathcal{X}_S$ . We express the corresponding loss in terms of the cross-entropy that is defined as the log-likelihood between the training data and the model distribution [10]. Specifically, we deploy an empirical categorical cross-entropy based loss,

$$\mathcal{L}_S = -\mathbb{E}_{(\mathbf{x}_i^s, y_i^s) \in \mathcal{X}_S} [y_i^s \log f_S(f_E(\mathbf{x}_i^s))] \quad (1)$$

where  $\mathbb{E}_{\mathcal{D}}$  denotes the empirical expectation over domain  $\mathcal{D}$ .

**The Class-Specific Source Reconstruction  $f_{DE}$ :** Note that  $f_S$  ensures better inter-class separation of the source domain samples in the learned space. However, it does not consider the notion of intraclass compactness which is essential for demarcating highly overlapping categories. In addition, we simultaneously require the learned space to be meaningful and to capture the inherent class-level abstract features of both  $\mathcal{S}$  and  $\mathcal{T}$ .

To this end, let us define two data matrices  $X_S \in \mathbb{R}^{N_S \times d}$  and  $\hat{X}_S \in \mathbb{R}^{N_S \times d}$  from  $X_S$  in such a way that the  $i^{th}$  row of both the matrices refers to a pair of distinct data points obtained from a given category. Under this setup,  $f_{DE}$  aims to reconstruct  $\hat{X}$  in the decoder branch given  $f_E(X_S)$ . We formulate the corresponding loss as:

$$\mathcal{L}_R = \sum_{i=1}^{N_S} \|\tilde{X}_S - \hat{X}_S\|_F^2 \quad (2)$$

Note that  $\tilde{X}_S$  denotes the projected  $f_E(X_S)$  onto the decoder. Since we perform cross-sample reconstruction in this encoder decoder branch ( $f_E$  and  $f_{DE}$ ),  $f_E$  essentially captures abstract class-level features of  $\mathcal{X}_S$ . Besides,  $\mathcal{L}_R$  further ensures within-class compactness. As a whole, the joint minimization of  $\mathcal{L}_S$  and  $\mathcal{L}_R$  ensures that  $f_E$  essentially learns a space which is simultaneously discriminative and meaningful.

**Domain Classifier  $f_D$ :** The role of  $f_D$  is to project the samples from  $\mathcal{S}$  and  $\mathcal{T}$  onto the shared space modelled by  $\Delta f_E$ . Let us assign the domain label 0 to all the source samples  $X_i^s$  and label 1 to all the target samples  $X_i^t$ . We define  $X_D = [X_S, X_T]$  and  $Y_D = [\hat{Y}_S, \hat{Y}_T]$  where  $\hat{Y}_S = \mathbf{0}$  is an all zero vector of size  $N_S$

and  $\hat{Y}_T = \mathbf{1}$  of size  $N_T$ . Given that,  $f_D$  maximizes a typical binary cross-entropy based classification error through a min-max game between  $f_E$  and  $f_D$  in such a way that the learned space becomes highly domain invariant. Formally we define the loss measure for  $f_D$  as:

$$\mathcal{L}_D = -\mathbb{E}_{(\mathbf{x}_k^D, y_k^D) \in (X_D, Y_D)} [y_k^D \log f_D(f_E(\mathbf{x}_k^D))] \tag{3}$$

**Orthogonality Constraint:** An additional orthogonality constraint over the reconstructed source domain features is added to the total loss to ensure their non-redundancy. The constraint is given as:

$$f_{DE}(X_S)^T f_{DE}(X_S) = I \tag{4}$$

However, Eq. (4) imposes a hard constraint over the optimization problem, so instead of incorporating it in the total loss, we minimize a softer version given as:

$$\mathcal{L}_O = f_{DE}(X_S)^T f_{DE}(X_S) - I \tag{5}$$

where  $I$  denotes identity matrix.

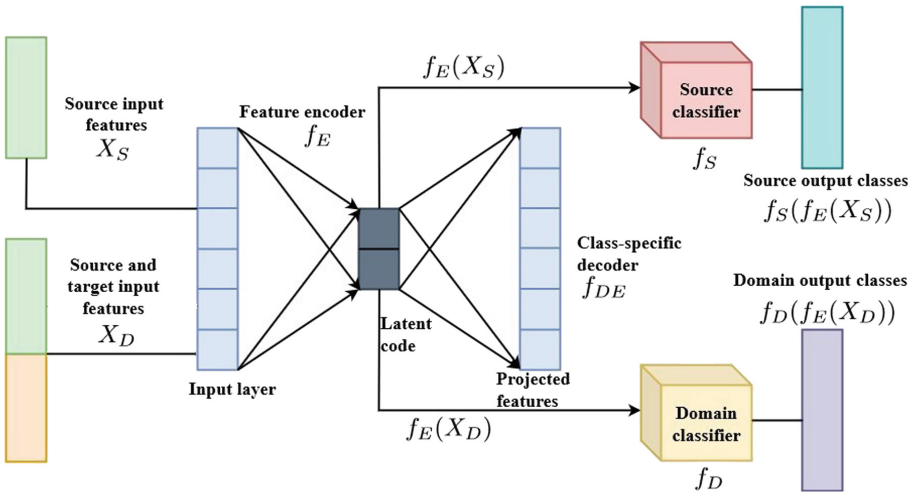


Fig. 1. Schematic flow of the proposed UDA model.

In Fig. 1, the source features  $X_S$  are encoded as  $f_E(X_S)$  and then are sent to source classifier  $f_S$ . In addition,  $X_D = [X_S, X_T]$  is encoded as  $f_E(X_D)$  and sent to domain classifier  $f_D$ . The reconstruction loss and orthogonality constraints are applied of the reconstructed source features  $f_{DE}(X_S)$ .

## 2.2 Optimization and Inference

Based on the Eqs. (1), (2), (3) and (5), the overall loss function  $\mathcal{L}$  can be represented as a two stage optimization process:

Stage 1:

$$\mathcal{L}_1 = \min_{f_S, f_E} \mathcal{L}_S + \mathcal{L}_R + \mathcal{L}_O \quad (6)$$

Stage 2:

$$\mathcal{L} = \min_{f_S, f_E} \max_{f_D} \mathcal{L}_1 - \mathcal{L}_D + \lambda \mathcal{R} \quad (7)$$

where  $\lambda$  denotes the weight of the regularizer  $\mathcal{R}$  on the learnable parameters. We follow the standard alternate stochastic mini-batch gradient descent approach to optimize  $\mathcal{L}$ . We find that the order of optimization of the individual terms does not matter in this case.

During testing, the target samples are assigned labels through  $f_S(f_E(X_T))$ .

## 3 Experiments

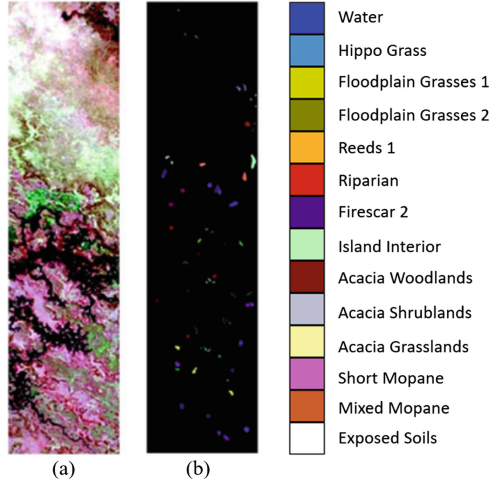
### 3.1 Datasets

Two benchmark hyperspectral datasets have been considered to validate the efficacy of our approach.

The first dataset is the Botswana hyper-spectral imagery (Fig. 2) [14]. The satellite imagery was acquired by NASA EO-1 satellite in the period 2001–2004 using the Hyperion sensor with the spatial resolution of 30 m spanning over 7.7 km strip. The imagery consists of 242 bands that covering the spectral range of 400–2500 nm. However in the current study, a preprocessed version of the dataset is used that comprises 10 bands obtained following a feature selection strategy.

Fourteen classes that correspond to land cover features on the ground are identified for the dataset. Many of the classes are fine-grained in nature with partially overlapping spectral signatures, causing the adaptation task extremely difficult. The source dataset (SD), consisting of 2621 pixels and target dataset (TD), containing 1252 pixels are created from spatially disjoint regions within the study area, leading to subtle differences in  $\mathcal{S}$  and  $\mathcal{T}$ , respectively.

The second dataset consists of two hyperspectral imageries, one over the Pavia City Center and the other over the University of Pavia (Fig. 3) [16]. The imageries captured from Reflective Optics Spectrographic Image System (ROSIS). The Pavia City Center image consists of 1096 rows, 492 columns and 102 bands while the University of Pavia image consists of 610 rows, 340 columns and 103 bands. Seven common classes are identified in both the images out of which few share similar spectral properties thus making their classification challenging. We use Pavia University as the source dataset while Pavia City Center as the target dataset. Since Pavia City Center image consists of 102 bands, the same number of bands are used for Pavia University image as well where the last band is dropped.



**Fig. 2.** Botswana dataset with (a) colour composite of first three bands and (b) ground truth. (Color figure online)

### 3.2 Protocols

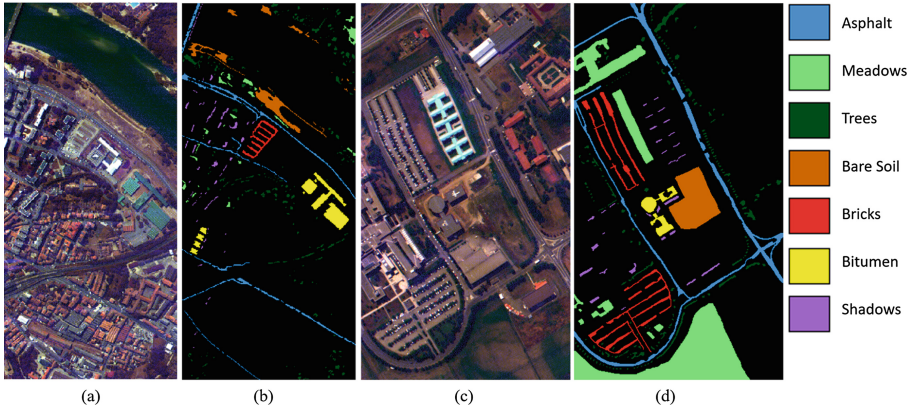
The entire network is constructed in terms of fully-connected neural network layers. In particular,  $f_E$  has two hidden layers with the dimensions of the final latent layer being 50. On the other hand, a single layer neural network is used for both the source-centric classifier and the domain classifier with the required number of output nodes.  $Relu(\cdot)$  non-linearity is used for all the layers. The weights for the loss terms are fixed through cross-validation and Adam optimizer [12] is considered with an initial learning rate of 0.001.

We report the classification accuracy at  $\mathcal{T}$  and compare the performance with the following approaches from the literature: TCA, SA, GFK, and RevGrad for Botswana dataset. However, for Pavia dataset, only GFK and RevGrad have been used for comparison since the accuracies obtained from other classifiers were quite insignificant. Note all the considered techniques aim to perform UDA in a latent space and RevGrad acts like the benchmark: it implicitly showcases the advantage of the proposed reconstructive loss term  $\mathcal{L}_R$ . In addition, we also carried out ablation study on our proposed method by eliminating reconstruction loss and orthogonality constraint one at a time.

### 3.3 Discussion

Tables 1 and 2 depict the quantitative performance evaluation and comparison to other approaches for Botswana and Pavia datasets respectively. The highest accuracy by a classifier for a given class is represented in bold.

For Botswana dataset, it can be inferred that the proposed approach outperforms the others with an overall classification accuracy of 74.5%. The RevGrad

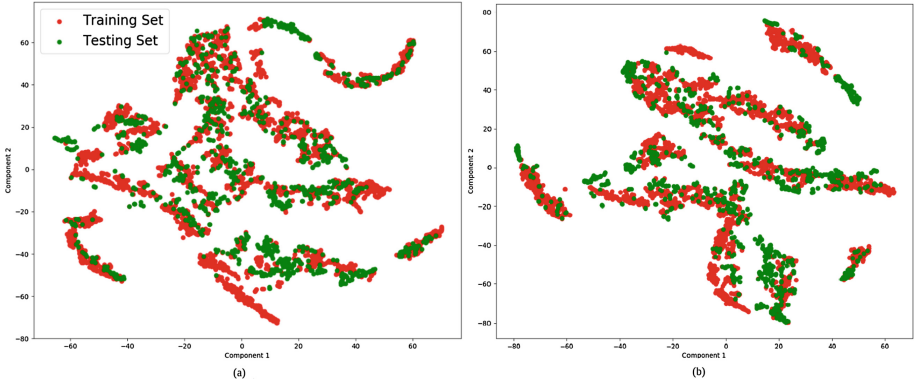


**Fig. 3.** Pavia dataset (a) colour composite and (b) the ground truth for the image of Pavia City Center(c) colour composite and (d) the ground truth for the image of the University of Pavia. (Color figure online)

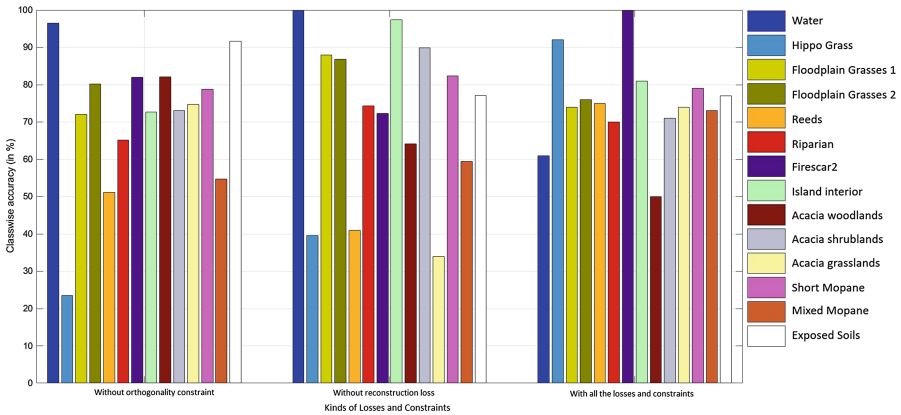
**Table 1.** Performance evaluation on the Botswana dataset (in %).

Land-cover classes	Pixel counts for SD	Pixel counts for TD	TCA [15]	SA [6]	GFK [9]	RevGrad [7]	Proposed method
Water (1)	213	57	60.0	46.0	43.0	<b>75.0</b>	61.0
Hippo grass (2)	83	81	<b>100.0</b>	<b>100.0</b>	75.0	97.0	92.0
Floodplain grasses 1 (3)	199	75	56.0	59.0	69.0	67.0	<b>74.0</b>
Floodplain grasses 2 (4)	169	91	75.0	80.0	<b>88.0</b>	79.0	76.0
Reeds (5)	219	88	78.0	<b>83.0</b>	81.0	67.0	75.0
Riparian (6)	221	109	58.0	72.0	<b>84.0</b>	65.0	70.0
Firescar 2 (7)	215	83	98.0	<b>100.0</b>	<b>100.0</b>	97.0	<b>100.0</b>
Island interior (8)	166	77	62.0	48.0	60.0	66.0	<b>81.0</b>
Acacia woodlands (9)	253	67	27.0	40.0	44.0	47.0	<b>50.0</b>
Acacia shrublands (10)	202	89	40.0	50.0	62.0	48.0	<b>71.0</b>
Acacia grasslands (11)	243	174	79.0	<b>92.0</b>	<b>92.0</b>	73.0	74.0
Short mopane (12)	154	85	89.0	<b>93.0</b>	91.0	73.0	79.0
Mixed mopane (13)	203	128	48.0	61.0	65.0	<b>77.0</b>	73.0
Exposed soil (14)	81	48	85.0	<b>100.0</b>	<b>100.0</b>	79.0	77.0
Overall accuracy (OA)	–	–	61.0	65.0	70.0	69.0	<b>74.5</b>





**Fig. 4.** Two dimensional t-SNE of Botswana’s source and target datasets (a) before domain adaptation (b) after domain adaptation.



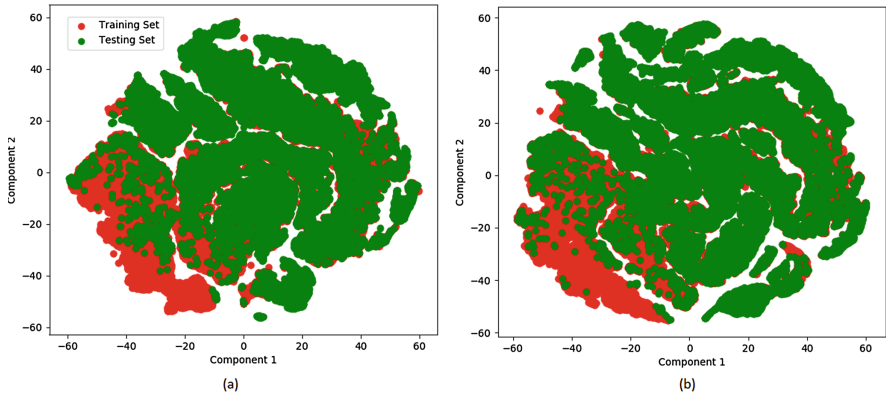
**Fig. 5.** Bar chart for ablation study comparing effects of exclusion of different losses on the classes of Botswana dataset.

technique on the other hand, produces an overall performance of 69%, thus implying that an overall domain alignment (without class) is not suitable for this dataset. The proposed method produces significant improvement in identifying island interior (OA = 81%), acacia woodlands (OA = 50%) and acacia shrublands (OA = 71%). These classes are difficult to handle having similar spectral properties with other classes and the ad hoc approaches considered for comparison mostly failed to identify them. For other classes, the results are comparable to the other techniques. Figure 4 shows the 2-D t-SNE comparing the source and target features (before training) with projected source and target features obtained after training.

The ablation study conducted on Botswana dataset showed that an OA of 65% is achieved when the reconstruction loss is not considered during while an

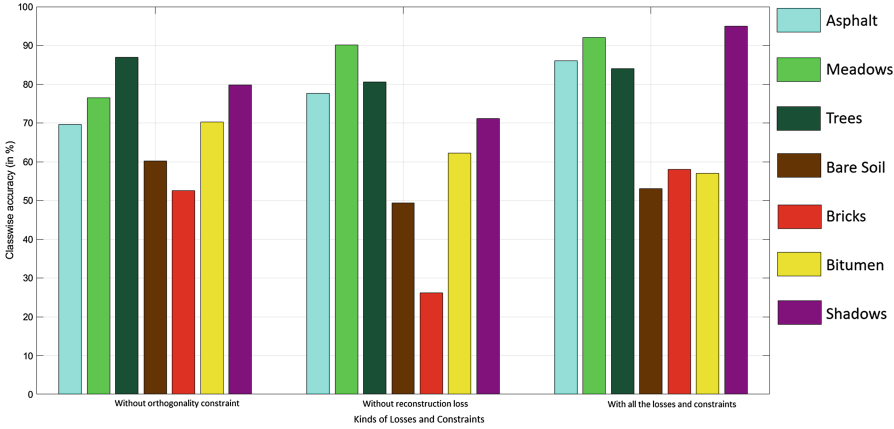
**Table 2.** Performance evaluation on the Pavia dataset (in %).

Land-cover classes	Pixel counts for SD	Pixel counts for TD	GFK [9]	RevGrad [7]	Proposed method
Asphalt (1)	6631	7585	50.0	64.0	<b>86.0</b>
Meadows (2)	18649	2905	47.0	61.5	<b>92.0</b>
Trees (3)	3064	6508	92.0	<b>94.0</b>	84.0
Baresoil (4)	5029	6549	<b>97.0</b>	72.5	53.0
Bricks (5)	3682	2140	62.0	67.0	58.0
Bitumen (6)	1330	7287	41.0	51.0	<b>57.0</b>
Shadows (7)	947	2165	<b>97.0</b>	83.5	95.0
Overall accuracy (OA)	—	—	66.0	70.5	<b>74.0</b>

**Fig. 6.** Two dimensional t-SNE of Pavia's source and target datasets (a) before domain adaptation (b) after domain adaptation.

OA of 64% is recorded in absence of orthogonality constraint. Figure 5 presents the accuracies obtained while conducting the ablation study on Botswana dataset. Significant improvement is observed in the accuracy of hippo grass, reeds and firescar 2 when all the losses and constraints are considered. It is also observed that the accuracy water class is decreased considerably for the same case. The accuracies for other classes are more or less same.

The similar trend is observed for Pavia dataset as well where our method surpasses the other classifiers with an OA of 74%, while the benchmark RevGrad classifier gives an OA of 70.5%. This affirms the inefficiency of domain alignment (without class) on the Pavia dataset as well. In addition, there is a significant improvement in classification of asphalt (OA = 86%) and meadows (OA = 92%) classes. The spectral signature of meadows class overlaps with that of that of



**Fig. 7.** Bar chart for ablation study comparing effects of exclusion of different losses on the classes of Pavia dataset.

trees (since both are a subset of vegetation), but our classifier performs well in identifying meadows much better than the other classifiers. For other classes, the classification accuracies are more or less similar to those from other classifiers. Figure 6 shows the t-SNEs of source and target features before and after training.

The ablation study on Pavia dataset showed an overall accuracy of 65% when the classifier was trained without orthogonality constraint while training without reconstruction loss gave an OA of 71%. Figure 7 compares the classwise accuracies for Pavia dataset against different losses considered in our ablation study. The results show that there is a significant improvement in the identifying of shadows (OA = 95%) and asphalt (OA = 86%) when all the losses are considered.

## 4 Conclusions

We propose a cross-domain classification algorithm for HSI based on adversarial learning. Our model incorporates an additional class-level cross-sample reconstruction loss for the samples in  $\mathcal{S}$  within the standard DC framework in order to make the learned space meaningful and classwise compact and an additional orthogonality constraint over the source domain to avoid any redundancy within the reconstructed features. Several experiments are conducted on the Botswana and Pavia datasets to assess the efficacy of the proposed technique. The results clearly establish the superiority of our approach with respect to a number of existing ad hoc and neural networks based methods. Currently, our method only relies on the spectral information. We plan to introduce the spatial aspect for improved semantic segmentation of the scene by distilling the advantages of convolution networks within the model.

## References

1. Banerjee, B., Chaudhuri, S.: Hierarchical subspace learning based unsupervised domain adaptation for cross-domain classification of remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10**(11), 5099–5109 (2017)
2. Bejiga, M.B., Melgani, F.: Gan-based domain adaptation for object classification. In: *Proceedings of the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1264–1267. IEEE (2018)
3. Cai, G., Wang, Y., Zhou, M., He, L.: Unsupervised domain adaptation with adversarial residual transform networks. arXiv preprint [arXiv:1804.09578](https://arxiv.org/abs/1804.09578) (2018)
4. Chen, H.Y., Chien, J.T.: Deep semi-supervised learning for domain adaptation. In: *Proceedings of the 25th IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6 (2015)
5. Chen, Y., Xu, W., Sundaram, H., Rikakis, T., Liu, S.M.: A dynamic decision network framework for online media adaptation in stroke rehabilitation. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **5**(1), 4 (2008)
6. Fernando, B., Habrard, A., Sebban, M., Tuytelaars, T.: Unsupervised visual domain adaptation using subspace alignment. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2960–2967 (2013)
7. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. arXiv preprint [arXiv:1409.7495](https://arxiv.org/abs/1409.7495) (2014)
8. Ganin, Y., et al.: Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **17**(1), 1–35 (2016)
9. Gong, B., Shi, Y., Sha, F., Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2066–2073 (2012)
10. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: *Deep Learning*, vol. 1. MIT Press, Cambridge (2016)
11. Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. arXiv preprint [arXiv:1412.6572](https://arxiv.org/abs/1412.6572) (2014)
12. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
13. Li, S., Song, S., Huang, G., Ding, Z., Wu, C.: Domain invariant and class discriminative feature learning for visual domain adaptation. *IEEE Trans. Image Process.* **27**(9), 4260–4273 (2018)
14. Neuenschwander, A.L., Crawford, M.M., Ringrose, S.: Results from the EO-1 experiment—a comparative study of Earth Observing-1 Advanced Land Imager (ALI) and Landsat ETM+ data for land cover mapping in the Okavango Delta, Botswana. *Int. J. Remote Sens.* **26**(19), 4321–4337 (2005)
15. 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)
16. Qin, Y., Bruzzone, L., Li, B., Ye, Y.: Tensor alignment based domain adaptation for hyperspectral image classification. arXiv preprint [arXiv:1808.09769](https://arxiv.org/abs/1808.09769) (2018)
17. Sohn, K., Shang, W., Yu, X., Chandraker, M.: Unsupervised domain adaptation for distance metric learning. In: *Proceedings of the International Conference on Learning Representations (ICLR)* (2018)
18. Tuia, D., Persello, C., Bruzzone, L.: Domain adaptation for the classification of remote sensing data: an overview of recent advances. *IEEE Geosci. Remote Sens. Mag.* **4**(2), 41–57 (2016)

19. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: *Proceedings of the Computer Vision and Pattern Recognition (CVPR)*, vol. 1, p. 4 (2017)
20. Yan, H., Ding, Y., Li, P., Wang, Q., Xu, Y., Zuo, W.: Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 3 (2017)
21. Ye, M., Qian, Y., Zhou, J., Tang, Y.Y.: Dictionary learning-based feature-level domain adaptation for cross-scene hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **55**(3), 1544–1562 (2017)