SS-HIDA: Semi-Supervised Heterogeneous Image Domain Adaptation

Mihailo Obrenovic, Thomas Lampert, François Monde-Kossi, Pierre Gançarski, Miloš Ivanović

SS-HIDA: Semi-Supervised Heterogeneous Image Domain Adaptation*

Mihailo Obrenović[1,2][0000−0003−1272−7437], Thomas Lampert[1][0000−0002−4911−3941], François Monde-Kossi[1][0000−0002−1580−5704], Miloš Ivanović[2][0000−0002−8974−2267], and Pierre Gançarski[1][0000−0003−1230−6560]

[1] ICube, University of Strasbourg, Strasbourg, France
[2] University of Kragujevac, Faculty of Science, Radoja Domanovica 12, 34000 Kragujevac, Serbia
mobrenovic@unistra.fr

Abstract. Heterogenous domain adaptation is a very challenging area, where labelled data from one dataset should help in learning on another unlabelled or scarcely labelled dataset, coming from a different input space. This problem is especially interesting in remote sensing, where a variety of sensors are used, producing images of different modalities and having a different number of bands. However, surprisingly, not much work has been done to address this problem. In this paper, we propose a novel approach for semi-supervised heterogeneous image domain adaptation named SS-HIDA. We evaluate on two heterogeneous remote sensing datasets, one being RGB, and the other multispectral, and show that SS-HIDA successfully outperforms the baseline method for the task of land-cover classification.

Keywords: Remote Sensing · Domain Adaptation · Deep Learning · Representation Learning

1 Introduction

In recent years, deep learning (DL) techniques have made a huge improvement in the field of computer vision (CV). Supervised DL methods rely heavily on the existence of large-scale labelled datasets. However, reference data is often difficult to obtain. This is especially true in the field of remote sensing (RS). Satellites generate a huge amount of data on a daily basis, and since labelling is a manual process, it is slow and expensive. Not to mention that the Earth's surface is constantly evolving, meaning that reference data may not be reusable for images taken at a later time. Different satellites have different sensors and capture images at different seasons or different places, all of which lead to very

* This work was granted access to the HPC resources of IDRIS under the allocation 2020-A0091011872 made by GENCI. We thank Nvidia Corporation for donating a Quadro P6000 GPU and the Centre de Calcul de l'Université de Strasbourg for access to the GPUs used for this research. Supported by the French Government through co-tutelle PhD funding.

different data distributions. Since DL models (and machine learning methods in general) often generalise poorly, we cannot apply existing trained models to other datasets. To overcome this problem, focus has turned to domain adaptation (DA) techniques.

Domain adaptation involves learning a model on one data distribution (named source - typically labelled), and applying it to another, different but related data distribution (called target - typically with little or no reference data) by reducing the shift between domains. This approach proves to be very successful, however, most of the DA methods assume RGB images in both domains (homogeneous DA), while in remote sensing, different sensors can capture images having a different number of channels (multispectral, hyperspectral, etc), or of different resolutions. As such, existing DL domain adaptation approaches cannot be applied in such heterogeneous situations because their structure is fixed, preventing images of different characteristics being used within the pipeline.

In this paper we propose a novel semi-supervised heterogeneous domain adaptation (HDA) approach for images called SS-HIDA with the task of correctly classifying land cover from satellite and aerial images. Although HDA methods exist, they focus on adapting from image to text data [20, 5] or adapting between SURF - DeCAF [13, 25] and DeCAF - ImageNet features [22]. As such, the development of such models would be very beneficial for the RS community where a variety of different sensors are used, some of them being RGB, multispectral, hyperspectral, SAR, LiDAR, panchromatic etc. To the best of our knowledge this is the first work on extracting domain invariant features from two heterogeneous unpaired image-data domains with a different number of bands.

Existing work on different modalities in the RS community have focused on data fusion [16] where different domains have corresponding paired images. However, in the proposed work, such a constraint does not exist, and therefore datasets with completely independent images can be used, possibly taken from different parts of the world.

This article is organised as follows: in Section 2, a review of related existing work is given, followed by a description of the proposed SS-HIDA architecture in Section 3. Section 4 describes the experimental setup and results. Finally, the conclusions are given and future work is discussed in Section 5.

2 Literature Review

The emergence of Generative Adversarial Networks (GANs) [9, 1] inspired numerous domain adaptation techniques for computer vision [7, 19]. The idea of making real and fake data indistinguishable is naturally extended to DA where two domains should be brought to the same space.

A large majority of DA methods for computer vision are concerned with RGB domains [7, 19, 4]. Some work on heterogeneous DA has been studied in domains having different features, e.g. SURF and DeCAF features [13, 25], DeCAF and GoogleNet features [22] etc. And others tackle the problem of image-text DA [20, 5].

One of the first methods applied between modalities (RGB and depth images) was Adversarial Discriminative Domain Adaptation (ADDA) [23], but it is worth noting that it is primarily used on RGB images.

Perhaps the most promising methods for image based HDA are Image-to-Image translation GANs [27, 26], which can translate images from one domain to another. One of the most famous architectures, CycleGAN [27], has been successfully applied to DA in RGB images [11].

Transfer learning is much more difficult task in RS when compared to CV. For many CV object classification datasets, models pre-trained on ImageNet give transferable features. However, the equivalent large-scale curated datasets in RS are only starting to exist [12]. An interesting study on transfer learning across multiple remote sensing datasets was given in [15]. Nevertheless, the fact that heterogeneous data (i.e. data with > 3 and/or non-RGB bands) exists would prevent them from being applied in the same manner as in CV.

Tasar et al. propose an approach that uses image-to-image translation for DA in RS [21], nevertheless it is applied to RGB domains only. CycleGAN has been used for translating between optical and SAR images [17] in change detection. There have been works on DA for semantic segmentation of land cover maps using data from different sensors in different domains [3, 2], but in one case, though the bands may be different, their number still has to be the same [3], while in the other case, labelled segmentation masks are needed in the target domain, and these (segmentation masks) are used as an intermediate space during the translation from the target domain to the source domain [2], this approach therefore does not extend to classification. Voreiter et al. propose the most similar method to that presented herein [24], the authors use variant of CycleGAN and apply it to two remote sensing datasets of different resolutions.

In the literature, unsupervised domain adaptation (UDA) is addressed more often than semi-supervised domain adaptation (SSDA). However it was shown that existing UDA methods do not scale well to the semi-supervised setting [18] and that there is a need for methods specifically tailored for SSDA. One such deep learning method is successfully applied to RS [14] using the Bayesian paradigm.

3 Semi-Supervised Heterogeneous Image Domain Adaptation (SS-HIDA)

Most of the existing HDA methods are based on the idea of translating data from one domain to the other, either in pixel space using image-to-image methods [27, 26], or in feature space, e.g. ADDA [23]. When trained in this manner, however, the resulting models are only applicable to the target domain. They are therefore bound to either simplify or invent the difference between domains during the translation, since the target data distribution must be made to match the source's distribution. Instead, we propose a method that extracts domain invariant features. The extracted features are neither in the source, nor target data space, but in a learnt common latent space. The hypothesis being that

this will allow the model to enhance the latent representation using information from both domains. Our method is inspired by homogeneous DA methods such as DANN [7], WDGRL [19], and DSN [4] which also extract domain invariant features, but are limited to working with homogeneous domains only.

We extend the homogeneous, unsupervised domain adaptation approach Wasserstein Distance Guided Representation Learning (WDGRL) [19] to the case of heterogeneous image data.

Let $X^s = \{(x_i^s, y_i^s)\}_{i=1}^{n^s}$ be a labelled source dataset of n^s samples from the domain $\mathcal{D}_s$ following the data distribution $\mathbb{P}_{x^s}$. SS-HIDA uses a small amount of target labels, so let us define two separate sets of target data, one being labelled $X^{tl} = \left\{(x_j^{tl}, y_j^t)\right\}_{j=1}^{n^{tl}}$, and the other being unlabelled $X^{tu} = \{x_k^{tu}\}_{k=1}^{n^{tu}}$, $n^{tl} << n^{tu}$, where target samples $x^t \in \left\{x_j^{tl}\right\}_{j=1}^{n^{tl}} \cup \{x_k^{tu}\}_{k=1}^{n^{tu}}$ come from the domain $\mathcal{D}_t$ and follow the data distribution $\mathbb{P}_{x^t}$. Unlike WDGRL, SS-HIDA is able to work with heterogeneous domains, i.e. $x^s \in \mathcal{X}^s$, $x^t \in \mathcal{X}^t$, $\mathcal{X}^s \neq \mathcal{X}^t$ where the dimensions d^s and d^t of spaces $\mathcal{X}^s$ and $\mathcal{X}^t$ may or may not differ.

SS-HIDA's architecture is presented in Figure 1, and consists of 5 neural network components: 3 feature extractors, a domain critic, and a class discriminator. To be able to work with the data coming from two different spaces, possibly of different input sizes, two different input branches are needed. Therefore, SS-HIDA has two separate feature extractors — $FE_s : \mathcal{X}^s \to \mathbb{R}^c$ and $FE_t : \mathcal{X}^t \to \mathbb{R}^c$ — these have the task to bring the data to a feature space of the same size — $g^s = FE_s(x^s)$, $g^t = FE_t(x^t)$. Furthermore, another shared feature extractor $FE_{sh} : \mathbb{R}^c \to \mathbb{R}^d$ is employed to model the similarity of the data domains, and to extract domain invariant features — $h^s = FE_{sh}(g^s)$, $h^t = FE_{sh}(g^t)$. Note that in Figure 1, the specific architecture presented is for use on RESISC45 and EuroSAT datasets, which can be adapted to other datasets.

Wasserstein distance is used to measure the distance between domains. This metric comes from the theory of optimal transport. Since calculating Wasserstein distance is computationally expensive, the domain critic component $DC : \mathbb{R}^d \to \mathbb{R}$ is trained to approximate it [1, 19], which makes the training process much faster. The domain critic utilises the whole target dataset x^t including the unlabelled part, i.e. a total of $n^t = n^{tl} + n^{tu}$ samples. The loss of this component is defined as:

$$\mathcal{L}_{wd}(h^s, h^t) = \frac{1}{n^s} \sum_{i=1}^{n^s} DC(h_i^s) - \frac{1}{n^t} \sum_{j=1}^{n^t} DC(h_j^t). \tag{1}$$

In order to calculate empirical Wasserstein distance, Equation (1) needs to be maximised, therefore the domain critic component is trained by solving:

$$\max_{\theta_{dc}} (\mathcal{L}_{wd} - \gamma \mathcal{L}_{grad}), \tag{2}$$

where θ_{dc} are the domain critic's weights and $\gamma \mathcal{L}_{grad}$ is a regularisation term enforcing the Lipschitz constraint (Eq. (5) from [19]), which improves upon simple weight clipping [1].

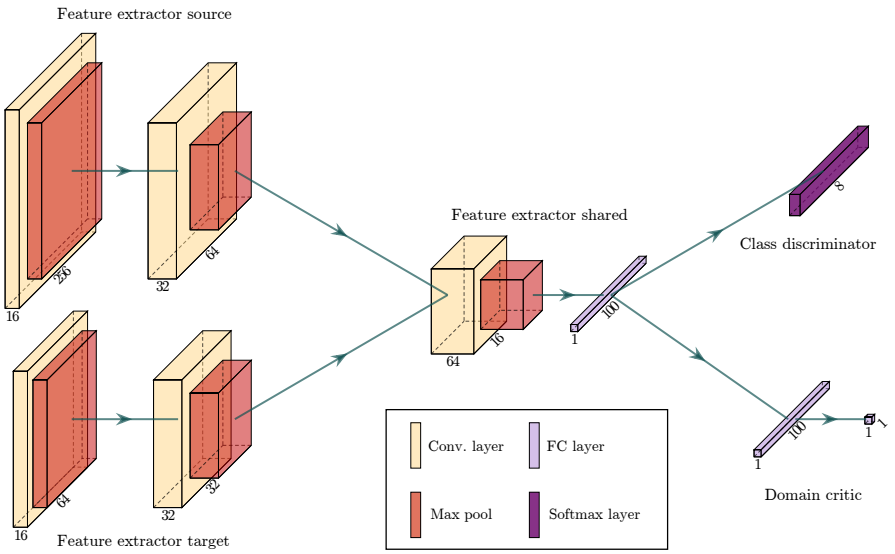


Fig. 1: The proposed heterogeneous semi-supervised domain adaptation model. The specific architecture presented is used for the case when the source dataset is RESISC45 and target dataset is EuroSAT. The kernel size of all convolutional layers is 5×5.

Finally, the class discriminator $C : \mathbb{R}^d \to \mathbb{R}^l$ (l being the number of classes) is trained on the extracted features of all the labelled samples $(h, y) = (h^s, y^s) \cup (h^{tl}, y^{tl})$. If labels y_i are one-hot encoded, cross-entropy classification loss is defined as:

$$\mathcal{L}_c(h, y) = -\frac{1}{n^s + n^{tl}} \sum_{i=1}^{n^s+n^{tl}} \sum_{k=1}^{l} y_{i,k} \log C(h_i). \tag{3}$$

If we denote the weights of the feature extractor as θ_{fe}, and the weights of class discriminator as θ_c, the final min-max adversarial optimisation problem to be solved is:

$$\min_{\theta_{fe}, \theta_c} \left\{ \mathcal{L}_c + \lambda \max_{\theta_{wd}} \left[\mathcal{L}_{wd} - \gamma \mathcal{L}_{grad} \right] \right\}. \tag{4}$$

4 Experimental results

Unlike WDGRL, whose components are fully connected neural networks, SS-HIDA is a convolutional architecture (see Figure 1 for details). The DC loss' weight λ is 0.1, the learning rate is 10^{-4}, and the Adam optimiser is used. The input data is standardised per channel so that each channel has mean 0 and standard deviation 1. The following augmentation transformations are used: flipping with a probability of 0.45, rotation from 0° to 180° with a probability of

Name	Source	Image Size	Samples	Classes	Resolution
RESISC45	Aerial	$256 \times 256 \times 3$	31,500	45	0.2 m−30 m
EuroSAT	Satellite	$64 \times 64 \times 13$	27,000	10	10 m

Table 1: Characteristics of NWPU-RESISC45 and EuroSAT datasets

0.75, changing contrast with the probability of 0.33 by multiplying the values of the pixels with the coefficient ranging between 0.5 and 1.5, changing brightness with the probability of 0.33 by adding the coefficient ranging between -0.3 and 0.3 scaled by the mean of pixel values per channel before standardisation, blurring with the probability of 0.33 with Gaussian filter with σ parameter values ranging from 1.5 to 1.8, and finally adding Gaussian noise with mean 0 and standard deviation between 10 and 15 with the probability of 0.33. In each iteration, half of the training batch comes from the source, and the other half from the target domain. The model is trained for 40 epochs, and the one with the lowest validation loss is chosen.

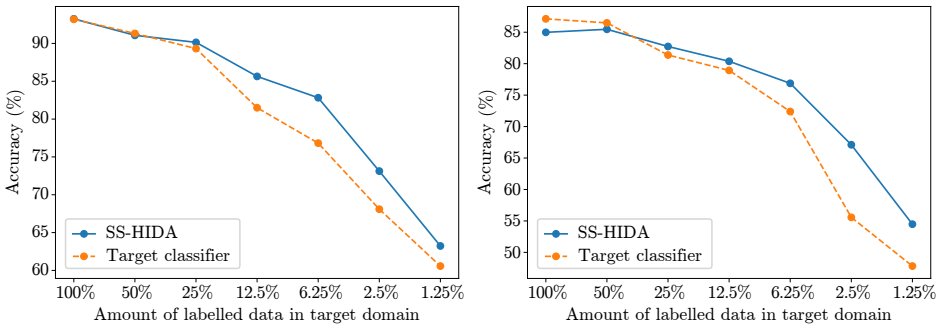
SS-HIDA is compared to the target baseline, a classifier trained on the same amount of labelled target data as our semi-supervised DA model. The same architecture is used, i.e. the same layers as the target FE, shared FE, and class discriminator. The models are evaluated on different amounts of labelled target data, ranging from fully labelled target dataset (100%) to only $n^{tl} = 5$ labelled samples per class (1.25%).

4.1 Data

The proposed approach is evaluated on the following eight corresponding classes from two heterogeneous remote sensing datasets (details given in Table 1):

- NWPU-RESISC45 [6] (high resolution aerial RGB images extracted from Google Earth) — dense residential, forest, freeway, industrial area, lake, meadow, rectangular farmland, and river.
- EuroSAT [10] (low resolution multi-spectral images from the Sentinel-2A satellite) — residential, forest, highway, industrial, sealake, pasture, annual crop and permanent crop (two classes merged into one), river.

The datasets are split into train, validation, and test sets with the proportion of 60:20:20 while keeping the classes balanced in all sets. For the cases of 12.5% labelled target data and lower, the target validation set consists of n^{tl} images, the same number of samples as in the target training set. Otherwise we would have an unrealistic situation in which the target validation set would be bigger than target training set. This evaluation protocol is similar to that used by Saito et al. [18]. As SS-HIDA has access to the source labelled data, it uses both the source and target validation set, while the target baseline can only use the target validation set.



(a) Source: RESISC45, Target: EuroSAT (b) Source: EuroSAT, Target: RESISC45

Fig. 2: Comparison of the results of SS-HIDA and the baseline classifier with varying numbers of labelled training images. The numbers are expressed in percentages of labelled images.

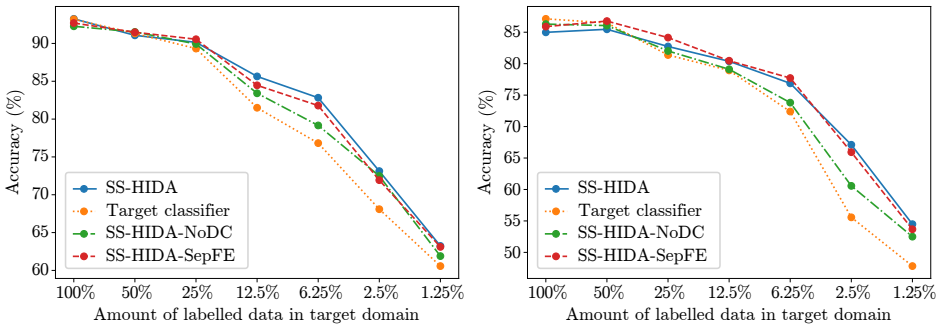
4.2 Results

The results are presented in Figure 2. They show that SS-HIDA performs very well in semi-supervised domain adaptation where a small amount of target labelled data is available. For the case when RESISC45 is taken as the source (and EuroSAT as the target), SS-HIDA is almost equal in performance to the baseline when there is a large amount of labelled data, and it strongly outperforms the baseline in the more realistic scenario in which only a small fraction of target data is labelled. When the percentage of labelled target data is 12.5%, 6.25%, or 2.5%, SS-HIDA achieves 5–6% higher accuracy than the baseline.

When EuroSAT is taken as the source and RESISC45 as the target, the advantage of SS-HIDA compared to the target baseline is even more pronounced, gaining almost 12% when only 2.5% of target data is labelled.

In both cases, when 1.25% of labelled target data is available, the benefit of SS-HIDA reduces, probably due to overfitting on the target data as there are only 5 labelled images per class available in both training and validation sets. But even though reduced, the benefit is still high when EuroSAT dataset is source and RESISC45 is target, with SS-HIDA gaining over 6%.

It is worth noting that the baseline performs surprisingly well with such few labelled images, achieving over 60% when only five labelled images per class are given. Keeping in mind that the specific architecture used is not optimised to achieve state of the art performance, this indicates that the classification problem is relatively easy (which is backed up by other findings in the literature [15, 8]) and perhaps more pronounced improvements could be found in more difficult datasets.



(a) Source: RESISC45, Target: EuroSAT (b) Source: EuroSAT, Target: RESISC45

Fig. 3: Ablation study of SS-HIDA, comparison with the model without domain critic and without shared layers with varying numbers of labelled training images. The numbers are expressed in percentages of labelled images.

4.3 Ablation Study

In order to discover the impact of each of the model's components, an ablation study is performed. One comparison model is created in which the domain critic is removed (SS-HIDA-NoDC), thus removing the domain adaptation component. A second comparison model is obtained by separating all of the layers of source and target architecture so that only the classifier is shared between them (SS-HIDA-SepFE), thus reducing the capacity of learning a general representation.

The results are shown in Figure 3. In both cases we can confirm that removing the domain critic leads to a significant drop in performance, especially in the cases where SS-HIDA has most success. In this case, there is no requirement for the model to learn overlapping distributions, therefore reducing the classifier's ability to generalise between domains. On the other hand, separating the source and target has less effect on performance. When RESISC45 is the source and EuroSAT is the target, the performance of SS-HIDA-SepFE is either the same or a little worse than SS-HIDA. It even outperforms SS-HIDA in some of the cases when EuroSAT is the source and RESISC45 is the target. But SS-HIDA remains slightly better when there is 2.5% or 1.25% labelled data in target domain. In this case, and when there are sufficient labels in the target domain, the domain critic is able to compensate and still forces the model to extract domain invariant features. It is also worth noting that none of these variations result in performance worse than the baseline.

5 Conclusions

This article has proposed a novel approach to semi-supervised heterogeneous image domain adaptation called SS-HIDA. To the best of our knowledge, this is the first approach to extract domain-invariant features. The model was evaluated

on an aerial dataset RESISC45 and a satellite dataset EuroSAT. Nevertheless, SS-HIDA is not limited to remote sensing applications only, and could be used for other cases of heterogeneous images, for example RGB and depth images. The next step of this research will be extending the usage of SS-HIDA to unsupervised DA. Initial experiments show that problems such as label-flipping need to be addressed in the unsupervised setting. Another interesting direction is to tackle the problem of semantic segmentation of images when domains are heterogeneous.

References

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: Proceedings of the ICML. pp. 214–223 (2017)
2. Benjdira, B., Ammar, A., Koubaa, A., Ouni, K.: Data-efficient domain adaptation for semantic segmentation of aerial imagery using generative adversarial networks. Applied Sciences **10**(3), 1092 (2020)
3. Benjdira, B., Bazi, Y., Koubaa, A., Ouni, K.: Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images. Remote Sensing **11**(11), 1369 (2019)
4. Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D., Erhan, D.: Domain separation networks. In: Proceedings of the NIPS. pp. 343–351 (2016)
5. Chen, W.Y., Hsu, T.M.H., Tsai, Y.H.H., Wang, Y.C.F., Chen, M.S.: Transfer neural trees for heterogeneous domain adaptation. In: Proceedings of the ECCV. pp. 399–414 (2016)
6. Cheng, G., Han, J., Lu, X.: Remote sensing image scene classification: Benchmark and state of the art. Proceedings of the IEEE **105**(10), 1865–1883 (2017)
7. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. JMLR **17**(1), 2096–2030 (2016)
8. Gómez, P., Meoni, G.: MSMatch: Semi-supervised multispectral scene classification with few labels. arXiv preprint arXiv:2103.10368 (2021)
9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Proceedings of the NIPS. pp. 2672–2680 (2014)
10. Helber, P., Bischke, B., Dengel, A., Borth, D.: EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification. IEEE J Sel Top Appl Earth Obs Remote Sens **12**(7), 2217–2226 (2019)
11. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: CyCADA: Cycle-consistent adversarial domain adaptation. In: Proceedings of the ICML. pp. 1989–1998 (2018)
12. Li, H., Dou, X., Tao, C., Wu, Z., Chen, J., Peng, J., Deng, M., Zhao, L.: RSI-CB: A large-scale remote sensing image classification benchmark using crowdsourced data. Sensors **20**(6), 1594 (2020)
13. Li, J., Lu, K., Huang, Z., Zhu, L., Shen, H.T.: Heterogeneous domain adaptation through progressive alignment. IEEE Trans. Neural Netw. Learn. Syst. **30**(5), 1381–1391 (2018)
14. Lucas, B., Pelletier, C., Schmidt, D., Webb, G.I., Petitjean, F.: A Bayesian-inspired, deep learning-based, semi-supervised domain adaptation technique for land cover mapping. Machine Learning pp. 1–33 (2021)

15. Neumann, M., Pinto, A.S., Zhai, X., Houlsby, N.: Training general representations for remote sensing using in-domain knowledge. In: Proceedings of the IEEE IGARSS. pp. 6730–6733 (2020)
16. Rudner, T.G., Rußwurm, M., Fil, J., Pelich, R., Bischke, B., Kopačková, V., Biliński, P.: Multi3Net: Segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 702–709 (2019)
17. Saha, S., Bovolo, F., Bruzzone, L.: Building change detection in VHR SAR images via unsupervised deep transcoding. IEEE Trans. Geosci. Remote Sens. (2020)
18. Saito, K., Kim, D., Sclaroff, S., Darrell, T., Saenko, K.: Semi-supervised domain adaptation via minimax entropy. In: Proceedings of the IEEE ICCV. pp. 8050–8058 (2019)
19. Shen, J., Qu, Y., Zhang, W., Yu, Y.: Wasserstein distance guided representation learning for domain adaptation. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 4058–4065 (2018)
20. Shu, X., Qi, G.J., Tang, J., Wang, J.: Weakly-shared deep transfer networks for heterogeneous-domain knowledge propagation. In: Proceedings of the ACM Multimedia. pp. 35–44 (2015)
21. Tasar, O., Happy, S., Tarabalka, Y., Alliez, P.: SemI2I: Semantically consistent image-to-image translation for domain adaptation of remote sensing data. In: Proceedings of the IEEE IGARSS. pp. 1837–1840 (2020)
22. Titouan, V., Redko, I., Flamary, R., Courty, N.: CO-Optimal Transport. In: Proceedings of the NeurIPS. vol. 33, pp. 17559–17570 (2020)
23. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE CVPR. pp. 7167–7176 (2017)
24. Voreiter, C., Burnel, J.C., Lassalle, P., Spigai, M., Hugues, R., Courty, N.: A Cycle GAN approach for heterogeneous domain adaptation in land use classification. In: Proceedings of the IEEE IGARSS. pp. 1961–1964 (2020)
25. Wang, X., Ma, Y., Cheng, Y., Zou, L., Rodrigues, J.J.: Heterogeneous domain adaptation network based on autoencoder. J Parallel Distrib Comput **117**, 281–291 (2018)
26. Yi, Z., Zhang, H., Tan, P., Gong, M.: DualGAN: Unsupervised dual learning for image-to-image translation. In: Proceedings of the IEEE ICCV. pp. 2849–2857 (2017)
27. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE ICCV. pp. 2223–2232 (2017)