MITIGATING DISTRIBUTION SHIFT FOR MULTI-SENSOR CLASSIFICATION

Sudipan Saha[1]*, Shan Zhao* [1]*, Muhammad Shahzad*[1]*, Xiao Xiang Zhu*[1,2]

Data Science in Earth Observation, Technical University of Munich, Ottobrunn, Germany[1]
Remote Sensing Technology Institute, German Aerospace Center (DLR), Weßling, Germany[2]

IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-2792-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/IGARSS46834.2022.9883596

ABSTRACT

Distribution shift may pose significant challenges in Earth observation, especially when dealing with significantly different sensors like multispectral optical and Synthetic Aperture Radar (SAR). Deep learning models trained for optical image classification generally do not generalize well for SAR images. This is due to very marked differences between them. Though there is a considerable amount of works on domain adaptation, only few deal with such strong differences. Towards this, we propose a co-teaching based domain adaptation method using dual classifier head, a Multi-layer Perceptron (MLP) classifier and a Graph Neural Network (GNN) classifier. The two classifier heads teach each other in an iterative manner, thus gradually adapting both of them for target classification. We experimentally demonstrate the efficacy of the proposed approach on Sentinel 2 (optical) as source and Sentinel 1 (SAR) images as target - both product of Copernicus program of European Space Agency.

Index Terms— Multi-sensor, Optical, Synthetic Aperture Radar, Domain adaptation, Graph Neural Network, Co-teaching.

1. INTRODUCTION

Earth observation acquisitions can be done by a wide array of active and passive sensors. Most popular among them are the passive optical multispectral and active synthetic aperture radar (SAR) sensors. Deep learning has improved the state-of-the-art in both optical and SAR image analysis. However, the success of deep learning methods largely depends upon the availability of annotated data of high quality and quantity. SAR has the advantage to provide measurements that are independent from the weather and light conditions. However, SAR images are contaminated by multiplicative noise known as speckle and lack visual saliency [1], which make it difficult to interpret and challenging to annotate. Optical images, on the other hand, are much more visually salient and thus annotations can be more easily obtained. However, significant variation between the optical and SAR data distributions exists. The traditional supervised deep models trained using the optical data without any adaptation are likely to fail on the SAR domain. However, circumnavigating this challenge may be helpful in many remote sensing problems, e.g., multi-sensor change detection [2].

Domain adaptation (DA) is a popular machine learning technique, where the system aims to adapt the knowledge learned from source domain, and applies it to a target domain without using any target annotation. In this regard, DA techniques are able to align the data distributions either explicitly using standard divergence measures [3] or implicitly using adversarial approaches [4]. However, there are few works in the domain adaptation literature towards dealing with such strong inter-domain difference like optical and SAR. Mitigating strong divergence between optical and SAR data distributions is challenging for existing divergence measures [3] and adversarial based techniques [4]. Though several methods have been introduced for domain adaptation in the context of Earth observation image classification [5] [6] [7], most of them focus on simpler settings, e.g., mitigating differences between optical images collected from different locations [7].

Co-teaching [8] simultaneously trains two deep neural networks and lets them teach each other. Roy *et.al.*[9] incorporated co-teaching in the context of domain adaptation to simulate new labeled images in the target domain. Saha *et. al.* [10] further showed its effectiveness in context of multi-city adaptation. Motivated by this, we postulate learning robust features in a unified optical-SAR space is important for DA between them. We propose to represent the source optical and the target SAR samples as a graph and then leverage Graph Neural Network (GNN) [11, 12] to aggregate semantic information from similar samples in a neighborhood. The key contribution of this paper is extending co-teaching based domain adaptation for optical-SAR adaptation. Furthermore, we created a novel test set-up using Sentinel 2 (optical) and Sentinel 1 (SAR) images.

The paper is organized as follows. Proposed method is presented in Section 2. The method is experimentally validated in Section 3. Finally, we conclude the paper in Section 4.

2. PROPOSED METHOD

We are provided with a source optical labeled dataset $\mathcal{S} = \{(\mathbf{x}_{s,i}, y_{s,i})\}_{i=1}^{n_s}$ and a target SAR unlabeled dataset $\mathcal{T} = \{(\mathbf{x}_{t,j})\}_{j=1}^{n_t}$. It is assumed that the label space of optical and SAR domains are the same, comprising of n_c classes. Our

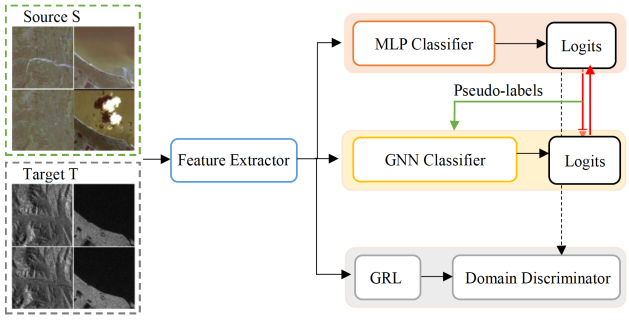

Fig. 1. Proposed joint MLP-GNN method for optical-SAR adaptation.

Table 1. Two-head network architecture.

Network component	Architecture
Feature extractor(F)	Resnet-18 excluding FC layer
MLP classifier(G_{mlp})	FC layer
Edge network(f_{edge})	Conv(256,256,1), Conv(256,128,1), Conv(128,1,1)
Node classifier(f_{node})	Conv(512, $2 \times n_c$,1), Conv($2 \times n_c$,n_c,1)

goal is to learn a predictor for SAR domain using the data in $\mathcal{S} \cup \mathcal{T}$. We process the images using a feature extractor F, the output of which is fed to a MLP classifier $\mathcal{G}_{mlp}$ and a GNN classifier $\mathcal{G}_{gnn}$. While $\mathcal{G}_{mlp}$ is aimed towards instance-level independent prediction, $\mathcal{G}_{gnn}$ optimizes the prediction assuming that a minibatch of samples are available. The proposed framework is shown in Figure 1.

2.1. Feature extractor

Let F be a feature extractor network. For a given sample $\mathbf{x}$, it outputs $\mathbf{f} = F(\mathbf{x})$. We feed the network with a minibatch of B samples drawn from the source (optical) domain and B samples drawn from the target (SAR) domain. Minibatch is fed to the feature extractor to obtain $\mathcal{F} = \{\mathbf{f_{s,i}}, \mathbf{f_{t,i}}\}_{i=1}^{B}$ that is subsequently fed to both MLP network and GNN network. The feature extractor network can be realized using residual networks like ResNet-18 and ResNet-50 [13].

2.2. MLP classifier

Ingesting $\mathcal{F}$ as input, the MLP classifier $\mathcal{G}_{mlp}$ produces logits $\hat{\mathcal{G}} = \{\hat{\mathbf{g}}_{\mathbf{s,i}}, \hat{\mathbf{g}}_{\mathbf{t,i}}\}_{i=1}^{B}$. The classwise prediction $p(\hat{y} = c; c \in n_c)$ is obtained by passing $\hat{\mathcal{G}}$ through softmax function. The samplewise prediction of MLP classifier is not affected by the other samples in the minibatch, i.e., it makes instance-level independent predictions. The MLP classifier is trained with cross-entropy loss using only the samples from the source $\mathcal{S}$, i.e., excluding the target samples in minibatch.

2.3. GNN classifier

GNN classifier $\mathcal{G}_{gnn}$ consists of an edge network f_{edge} and a node classifier f_{node}. Each node represents a feature vector of an image and edges encode the relationship between every two nodes. In each iteration, a graph whose structure is implied in a $2B \times 2B$ affinity matrix is created.

2.3.1. *Edge Network*

The edge network is directly fed with the output of feature extractor, i.e., $\mathcal{F}$ and it produces an affinity matrix $\hat{\mathcal{A}}$. The entries $\hat{a}_{i,j}$ in $\hat{\mathcal{A}}$ indicate the similarity between sample i and j in the minibatch. When the sample i and j belong to the same semantic category, the entries $\hat{a}_{i,j}$ is set as 1 and 0 otherwise. To train f_{edge}, we need a target affinity matrix $\hat{\mathcal{A}}_{tar}$, with entries $\hat{a}_{i,j}^{tar}$.

The ground truth $\hat{\mathcal{A}}_{tar}$ needs to connect the objects of the same class as neighboring nodes. To do so we require class labels of all samples in the minibatch. While class labels of samples belonging to source domain are known, labels for samples belonging to the target domain are unknown. Towards training, it is aided by $\mathcal{G}_{mlp}$ that generates pseudo-label for the target samples. Nevertheless, not all pseudo-labels returned by $\mathcal{G}_{mlp}$ are trustworthy. The detailed filtering strategy is described Section 2.3.3 to select the high quality pseudo labels. The f_{edge} is trained using a binary cross entropy loss.

$$\mathcal{L}_{edge} = \hat{a}_{i,j}^{tar} \log p(\hat{a}_{i,j}) + (1 - \hat{a}_{i,j}^{tar}) \log (1 - p(\hat{a}_{i,j})) \quad (1)$$

2.3.2. *Node Network*

The f_{node} aggregates the features in $\mathcal{F}$ based on the estimated affinity matrix $\hat{\mathcal{A}}$ such that for each node/sample the most similar samples in its neighbourhood contribute to its final representation. The f_{node} outputs its logits as $\bar{\mathcal{G}}^0 = \{\bar{\mathbf{g}}_{\mathbf{s,i}}, \bar{\mathbf{g}}_{\mathbf{t,i}}\}_{i=1}^{B}$. The classwise prediction $p(\bar{y} = c; c \in n_c)$ is obtained by passing $\bar{\mathcal{G}}^0$ through softmax function. Different to $p(\hat{y} = c; c \in n_c)$ that is merely an instance-level independent prediction, $p(\bar{y} = c; c \in n_c)$ accounts for the other samples in the minibatch, that helps it to obtain superior result.

$$\mathcal{L}_{node} = -\frac{1}{|\mathcal{B}_s|} \sum_{i=1}^{\mathcal{B}_s} y_{s,i} \log p(\bar{y}_{s,i}) \quad (2)$$

2.3.3. *Filtering of pseudo labels*

As mentioned in Section 2.3.1, not all pseudo labels produced by MLP are correct. The accuracy of the pseudo labels plays an important role in the adaptation process. Towards this, Roy *et.al.* [9] mask out the pseudo labels whose confidence score(entropy) is lower than a pre-chosen threshold. However, choosing threshold a priori is challenging and may need additional validation set. Luo *et.al.*[14] take the progressive

learning strategy and each time label a percentage of the target samples according to their confidence score. Again, the enlarge factor α is a hyperparameter that needs to be carefully considered against the model performance and the computational cost. To circumnavigate this problem, we propose to take the co-assistance strategy. The output of the MLP is compared with GNN, only if both of them return the same label of a given target sample, the target sample is pseudo labeled.

$$w_j = \begin{cases} 1, & \text{if } \max_{c \in n_c} p(\bar{y}_{t,j} = c) = \max_{c \in n_c} p(\hat{y}_{t,j} = c) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

when w_j is 1, pseudo label (either from GNN or MLP) of $\mathbf{x}_{t,j}$ will be used to form $\hat{\mathcal{A}}_{tar}$. Otherwise the pseudo label is discarded. In Figure 1, the double red lines are the co-assistance filtering strategy for the selection of high-quality pseudo-labels, and the filtered predictions are fed to GNN (following the green line) for the construction of affinity matrix as mentioned in Section 2.3.1.

2.4. Domain discriminator

In tandem with proposed MLP-GNN setting, similar to [15], we also train a domain discriminator, connected to the feature extractor by a gradient reversal layer (GRL), using an adversarial loss $\mathcal{L}_{adv}$.

2.5. Summarized adaptation process

The source samples $\mathcal{S}$ helps in training the MLP classifier $\mathcal{G}_{mlp}$ that in turn helps in selecting pseudo samples from target domain $\mathcal{T}$. The source and the pseudo samples from target jointly help in training the $\mathcal{G}_{gnn}$ that further helps in improving $\mathcal{G}_{mlp}$ in a co-teaching manner. This process is further aided by domain discriminator that reduces the representation gap between source and target samples.

3. EXPERIMENTS

3.1. Dataset and settings

For validation of the proposed method, we use Sen12MS dataset that is a curated dataset of georeferenced multi-Spectral Sentinel-1/2 Imagery for scene classification. Sen12MS dataset [16] consists of dual-pol synthetic aperture radar (SAR) Sentinel-1 patches and full multi-spectral Sentinel-2 image patches. The size of each image is 256×256 pixels and the images are upsampled to a ground sampling distance of 10m.

For each domain, 1000 images/class are sampled for seven different classes. The classes are forest, shrubland, savanna, grassland, cropland, urban/builtup, and water. Thus the optical-SAR dataset in our experiments consists of 7000 optical images sampled from 7 classes and similarly for the SAR domain. For optical images, we used the RGB channels.

Table 2. Quantitative comparison of the proposed method with other methods

Method	Classification accuracy
Optical trained (no adaptation)	25.63
BIN [17]	28.49
Threshold based co-teaching [9]	38.83
Proposed MLP head	44.66
Proposed GNN head	44.91

The feature extractor used is pre-trained ResNet-50. We train the model using a Stochastic Gradient Descent (SGD) optimizer having an initial learning rate of 5e-4 and decay exponentially. The λ_{node} is 0.3, and $\lambda_{edge}, \lambda_{adv}$ are set as 1. First the model is trained using mere optical images for 1000 epochs, then it is adapted on the target domain for 5000 epochs. Due to the brightness shifts between optical and SAR images, the image normalization at the preprocessing stage is done using sensor-specific own mean and standard deviation.

3.2. Compared methods

The following methods are compared to the proposed method:

1. Model trained on Sentinel-2 data (no adaptation).
2. Statistical alignment based Batch-instance normalization (BIN) [17].
3. Model using co-teaching based strategy as in the proposed method, however using a threshold based pseudolabel selection as in [9].

3.3. Result

Without adaptation, the network on the optical Sentinel-2 images perform poorly on the SAR Sentinel-1 images (25.63%). In spite of poor performance, this is superior to the random guessing for balanced 7-class classification problem (14.29%).

Co-teaching [9] significantly improves the classification accuracy (38.83%). Target classification accuracy further improves by employing our proposed filtering of pseudolabels along with co-teaching. MLP head of the proposed method obtains an accuracy of 44.66%, while GNN head obtains an accuracy of 44.91%. Superior performance of the GNN head in comparison to MLP head further validates the merit of GNN in aggregating information across different domains. Thus, the proposed method obtains almost 20% improvement over the model without adaptation. Quantitative result is shown in Table 2.

4. CONCLUSIONS

This paper proposed a deep domain adaption method using a joint MLP-GNN architecture for the classification of multi-

sensor optical-SAR remote sensing data. Optical-SAR adaptation is a challenging task as the divergence between source domain (optical) and target domain (SAR) is large. Despite such strong difference between Sentinel-2 and Sentinel-1 data, the method is capable of using the former to classify the latter. The experimental results on an optical-SAR dataset showed the superiority of the GNN-based prediction in comparison to network before adaptation and simple statistical alignment based adaptation. Though there is much scope of improvement, we need to consider that SAR and optical images are considerably different and thus adapting a classifier for such scenario is significantly challenging. In future we plan to further improve the method by investigating different GNN architectures, e.g., Graph Attention Network.

Acknowledgement

The work is funded by the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" (Grant number: 01DD20001).

5. REFERENCES

[1] Sudipan Saha, Francesca Bovolo, and Lorenzo Bruzzone, "Destroyed-buildings detection from vhr sar images using deep features," in *Image and Signal Processing for Remote Sensing XXIV*. International Society for Optics and Photonics, 2018, vol. 10789, p. 107890Z.

[2] Sudipan Saha, Patrick Ebel, and Xiao Xiang Zhu, "Self-supervised multisensor change detection," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[3] Baochen Sun and Kate Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *European conference on computer vision*. Springer, 2016, pp. 443–450.

[4] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7167–7176.

[5] Shunping Ji, Dingpan Wang, and Muying Luo, "Generative adversarial network-based full-space domain adaptation for land cover classification from multiple-source remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[6] Ahmed Elshamli, Graham W Taylor, and Shawki Areibi, "Multisource domain adaptation for remote sensing using deep neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3328–3340, 2019.

[7] Wei Liu and Rongjun Qin, "A multikernel domain adaptation method for unsupervised transfer learning on cross-source and cross-region remote sensing data classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 4279–4289, 2020.

[8] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama, "Co-teaching: Robust training of deep neural networks with extremely noisy labels," *arXiv preprint arXiv:1804.06872*, 2018.

[9] Subhankar Roy, Evgeny Krivosheev, Zhun Zhong, Nicu Sebe, and Elisa Ricci, "Curriculum graph co-teaching for multi-target domain adaptation," *arXiv preprint arXiv:2104.00808*, 2021.

[10] Sudipan Saha, Shan Zhao, and Xiao Xiang Zhu, "Multitarget domain adaptation for remote sensing classification using graph neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[11] Thomas N Kipf and Max Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[12] Sudipan Saha, Lichao Mou, Xiao Xiang Zhu, Francesca Bovolo, and Lorenzo Bruzzone, "Semisupervised change detection using graph convolutional network," *IEEE Geoscience and Remote Sensing Letters*, 2020.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[14] Yadan Luo, Zijian Wang, Zi Huang, and Mahsa Baktashmotlagh, "Progressive graph learning for open-set domain adaptation," in *International Conference on Machine Learning*. PMLR, 2020, pp. 6468–6478.

[15] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan, "Conditional adversarial domain adaptation," *arXiv preprint arXiv:1705.10667*, 2017.

[16] Michael Schmitt, Lloyd Haydn Hughes, Chunping Qiu, and Xiao Xiang Zhu, "Sen12ms – a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2019, vol. IV-2/W7, pp. 153–160.

[17] Hyeonseob Nam and Hyo-Eun Kim, "Batch-instance normalization for adaptively style-invariant neural networks," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018, pp. 2563–2572.