An End-to-End Network for Remote Sensing Imagery Semantic Segmentation via Joint Pixel and Representation Level Domain Adaptation

Lukui Shi, Ziyuan Wang, Bin Pan, and Zhenwei Shi

Abstract

It requires pixel-by-pixel annotations to obtain sufficient training data in supervised remote sensing image segmentation, which is a quite time-consuming process. In recent years, a series of domain adaptation methods were developed for image semantic segmentation. Generally, these methods are trained on the source domain and then validated on the target domain so as to avoid repeatedly labelling new data. However, most domain adaptation algorithms only tried to align the source domain and the target domain in pixel level or representation level, while ignored their cooperation. In this paper, we propose an unsupervised domain adaptation method by joint pixel and representation level align (JPRNet). The major novelty of JPRNet is that it achieves joint domain adaptation in an end-to-end manner, so as to avoid the multi-source problem in remote sensing images. JPRNet is composed of two branches, each of which is a generative adversarial network (GAN). In one branch, the pixel level domain adaptation is implemented by the style transfer with Cycle GAN, which could transfer the source domain to a target domain. In the other branch, the representation level domain adaptation is realized by adversarial learning between the transferred source domain images and the target domain images. The experimental results on public datasets have indicated the effectiveness of JPRNet.

Index Terms

Remote sensing, semantic segmentation, domain adaptation, GAN.

I. Introduction

Semantic segmentation, which aims at assigning label to each pixel in an image, is a fundamental and challenging problem in the field of aerial and satellite images. In recent years, researchers have proposed many semantic segmentation algorithms based on deep learning for remote sensing images [1]–[3]. However, most of them have to train the models on the large labeled datasets, while it is a time consuming process to collect such pixel level annotated datasets.

An attractive alternative is to use domain adaptation, which aims to transfer the model learnt on a labeled source domain to a target domain. During the past decade, researchers have proposed some domain adaptation algorithms for remote sensing image semantic segmentation [4]–[7]. More recently, the generative adversarial networks (GANs) have achieved promising performance in addressing the problem. In domain adaptation methods for the semantic segmentation of remote sensing images, GAN was used in [8]–[12].

However, the above methods only attempted to solve the domain shift problem by aligning either the pixel space or the representation space. In this paper, inspired by the idea of hierarchical domain adaptation, we propose an end-to-end network, which can address joint pixel and representation level domain adaptation (JPRNet). JPRNet is developed based on Cycle GAN [13], which is a popular pixel level backbone. A representation level domain adaptation approach is proposed to improve Cycle GAN.

To some extent, JPRNet involves the similar idea as Fully Convolutional Adaptation Networks for Semantic Segmentation (FCAN) [14]. However, they are quite different in the optimization manners. First and foremost, JPRNet is an end-to-end model while FCAN directly cascades two domain adaptation algorithms. Due to the multi-source problem of remote sensing images, the images obtained by different satellites are quite different. If not adopt the end-to-end manner for training domain adaptation networks, users may have to manually select different hyperparameters for any two remote sensing data, which will significantly increase the artificial interference. Therefore, the end-to-end structure can reduce the human intervention which helps to improve the robustness of the algorithm.

JPRNet contains the pixel level and representation level domain adaptation branches, each of which is a GAN. In the pixel level branch, domain adaptation is conducted by Cycle GAN which could transfer image style from the source domain images

The work was supported by the National Key R&D Program of China under the Grant 2017YFC1405605, the National Natural Science Foundation of China under the Grants 61671037, and the Natural Science Foundation of Hebei Province of China under the Grants F2020202008.*(Corresponding author: Bin Pan)*

Lukui Shi and Ziyuan Wang are with the School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China, and also with the Hebei Province Key Laboratory of Big Data Calculation, TianJin 300401, China (e-mail:shilukui@scse.hebut.edu.cn; wangziyuan.hebut@hotmail.com).

Bin Pan (Corresponding author) is with the School of Statistics and Data Science, Nankai University, Tianjin 300071, China (e-mail:panbin@nankai.edu.cn).

Zhenwei Shi is with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China (e-mail:shizhenwei@buaa.edu.cn).

to the transferred source domain images. In another branch, the representation level adaptation network (RAN) is used to realize domain invariant representation between the transferred source domain images and the target domain images. Our contributions are summarized as follows:

- We propose a domain adaptation method (JPRNet) for remote sensing imagery semantic segmentation, which can be trained on a labeled dataset and applied its model to another unlabeled dataset;
- We construct JPRNet with two GANs, which could simultaneously train the pixel and representation level branches via an end-to-end manner.

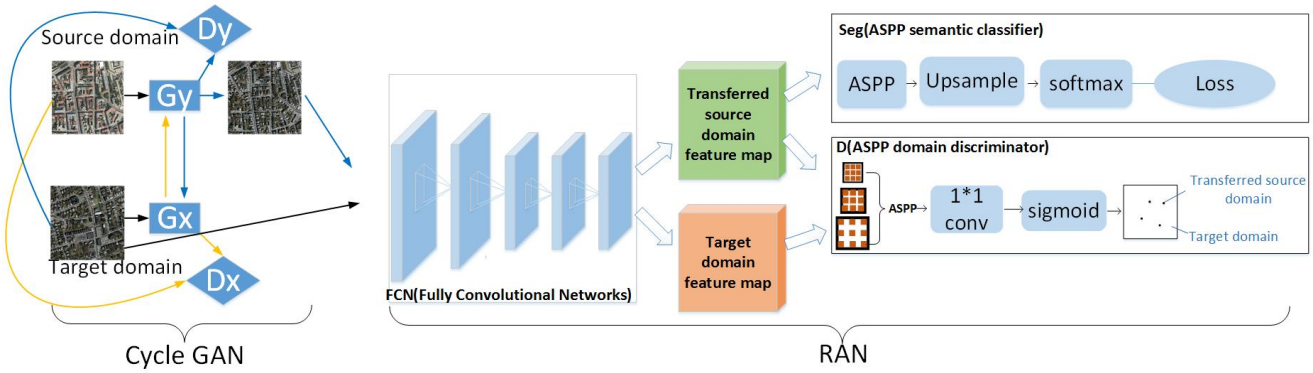

Fig. 1. Overall architecture of the proposed method. It consists of two main components: the pixel adaptation network (Cycle GAN) on the left and the representation adaptation network (RAN) on the right. Cycle GAN could transfer the image style of the source domain. RAN learns the domain-invariant representations of between the target domain images and the transferred source domain images in an adversarial manner.

II. Methods

Our proposed adaptive semantic segmentation network is illustrated in Figure 1. It consists of the pixel adaptation network (Cycle GAN) and the representation adaptation network (RAN). Given images from the source domain and the target domain, Cycle GAN transfers images from one domain to the other from the perspective of the pixel level in an adversarial manner. RAN learns the representation domain adaptation in an adversarial manner and a domain discriminator is designed to classify the image regions corresponding to the receptive field of each spatial unit in the feature map. RAN is to guide the representation learning in both domains, and makes the discriminator difficult to distinguish between the transferred source domain representations and the target domain representations. As a result, our algorithm addresses domain adaptation problem from both the pixel level and the representation level.

A. Pixel level adaptation network (PAN)

PAN is designed to transfer images from one domain to the other under as possible as preserving appearance similarity and to segment the transferred source domain images. In PAN, this goal is achieved by using Cycle GAN and FCN. The PAN network consists of five components: G_X, G_Y, D_X, D_Y and FCN, where G_X, G_Y, D_X and D_Y are parts of Cycle GAN, and FCN is a semantic segmentation network. Suppose that X represents the source domain dataset, Y represents the target domain dataset, $x_i \in$ X and $y_i \in$ Y. PAN aims to learn two mappings $G_Y(x)$ and $G_X(y)$, and train FCN. $G_Y(x)$ maps data from X to Y and $G_X(y)$ maps data from Y to X. FCN is trained by using the transferred source domain images and the source labels. Next, we summarize the objective functions of PAN.

PAN is designed by adding FCN on the basis of Cycle GAN. Therefore, we firstly introduce the objective function of Cycle GAN, which consists of four components. The adversarial term of the loss function for training G_Y and D_X can be written as follows:

$$\begin{aligned}\mathcal{L}_{X\to Y} =& E_{y\sim p_y(y)}[\log D_Y(y)]+ \\ & E_{x\sim p_x(x)}[\log(1-D_Y(G_Y(x)))]\end{aligned} \tag{1}$$

The most significant difference between Cycle GAN and other GAN networks is that Cycle GAN introduces the cycle consistency loss. The loss requires that the transferred image can be mapped back to itself in the original domain, namely: x$\to G_Y$(x)$\to G_X(G_Y(x))\approx$x. It is defined as follows:

$$\begin{aligned}\mathcal{L}_{cyc}(G_X, G_Y) =& E_{x\sim p_x(x)}[\|G_X(G_Y(x))-x\|_1]+ \\ & E_{y\sim p_y(y)}[\|G_Y(G_X(y))-y\|_1]\end{aligned} \tag{2}$$

According to the structure of Cycle GAN, the objective function of Cycle GAN can be written as follows:

$$\begin{aligned}\mathcal{L}_{cyclegan}(\tilde{G},\tilde{D}) = \mathcal{L}_{X\to Y}(G_Y, D_Y)+\\ \mathcal{L}_{Y\to X}(G_X, D_X) + \mathcal{L}_{cyc}(G_X, G_Y)\end{aligned} \tag{3}$$

where $\tilde{G}$ represents G_X and G_Y, $\tilde{D}$ represents D_X and D_Y.

Compared to the pixel level domain adaptation network AAN in FCAN, Cycle GAN implements pixel level domain adaptation in a generative-adversarial manner, while AAN implements pixel level domain adaptation in a reconstructed manner. AAN would use too many artificially set hyperparameters during the reconstruction process, which may lead to excessive human intervention. Therefore, we selected Cycle GAN with less human intervention as our pixel level domain adaptation network.

Then we introduce the objective function of FCN. Suppose that $c \in \{0,1\}$ represents the pixel-wise binary label of the image x, the loss function for the segmentation task can be written as:

$$\begin{aligned}\mathcal{L}_{seg}(FCN, G_Y) = -E_{(x,c)\sim p(x,c)}[c\log(FCN(G_Y(x)))\\ +(1-c)\log(1-FCN(G_Y(x)))]\end{aligned} \tag{4}$$

So, the objective function of PAN can be defined as follows:

$$\begin{aligned}\mathcal{L}_{PAN}(\tilde{G},\tilde{D},FCN) =&\mathcal{L}_{cyclegan}(\tilde{G},\tilde{D})\\ &+\mathcal{L}_{seg}(FCN, G_Y)\end{aligned} \tag{5}$$

B. Representation level adaptation network (RAN)

The purpose of RAN is to learn domain-invariant representations by an adversarial manner. In RAN, the feature representations of two domains are learnt by fooling a domain discriminator. It consists of FCN, ASPP semantic classifier Seg, and ASPP discriminators D. FCN is part of the segmentation network as well as the generator of GAN to generate domain-invariant representations.

Atrous Spatial Pyramid Pooling (ASPP) [15] uses multi-rate dilated convolution to extract multi-scale features in the form of spatial pyramid, which has proven to be effective in extracting multi-scale information. The ASPP semantic classifier Seg could promote segmentation results by fusing multi-scale features from different convolutional layers. In the semantic classifier, the settings of ASPP are the same as those of DeepLab V3.

The ASPP discriminator D attempts to distinguish the representation of the source domain and the target domain. It outputs the domain prediction of each image region that corresponds to the spatial unit in the final feature map. In the discriminator, specifically, *k* dilated convolutions with different sampling rates are exploited in parallel to produce *k* feature representations after the outputs of FCN are input into the discriminator. Here, each feature map has *c* feature channels. Then all feature channels are combined into *c*k* channels. These channels pass a 1×1 convolutional layer plus a sigmoid layer to generate the final score map. Each spatial unit in the final score map represents the probability of belonging to the target domain.

Because buildings in remote sensing images have different sizes, we attempt to use multi-scale representations to enhance adversarial learning and building segmentation. It is the traditional way for solving multi-scale problems to adjust the resolution of the input image and use parallel weight sharing network, which will consume a lot of memory and training time. In our network, ASPP is used not only to solve the multi-scale problem of segmentation, but also to solve the multi-scale discrimination of adversarial network.

C. Joint Pixel and Representation level Network (JPRNet)

JPRNet adds a representation level domain adaptation on the basis of Cycle GAN. As shown in Figure 1, Cycle GAN can achieve the pixel level domain adaptation. Its generator can output the target-like images, which have the common labels with images in the source domain. Then it is to learn domain-invariant representations between the transferred source domain images and Massachusetts Buildings dataset (the target domain). RAN is used to produce representations across domains and segment the transferred source domain images. Suppose that Y_{fake} represents the transferred source domain dataset, Y represents the target domain dataset, $y_{fake} \in Y_{fake}$ and y ∈ Y, the adversarial objective function and the objective function of RAN can be respectively written as:

$$\begin{aligned}\mathcal{L}_{adv}(FCN, D) = E_{y\sim Y}[\frac{1}{Z}\sum_{i=1}^{Z}\log(D_i(FCN(y)))]+\\ E_{y_{fake}\sim Y_{fake}}[\frac{1}{Z}\sum_{i=1}^{Z}\log(1-D_i(FCN(y_{fake})))]\end{aligned} \tag{6}$$

$$\mathcal{L}_{RAN}(FCN, D, Seg) = \mathcal{L}_{seg}(FCN, Seg) + \mu\mathcal{L}_{adv}(FCN, D) \tag{7}$$

where Z is the number of the spatial units in the output of D, and μ is the tradeoff parameter and the loss $\mathcal{L}_{seg}$ is the same as equation (4).

Besides, similar to literature [16], we also add the loss of semantic consistency as follow:

$$\begin{aligned}\mathcal{L}_{sem}(G_X,F) = \lambda E_{x\sim p_x(x)}[\|F(x) - F(G_Y(x))\|_1]+\\ \lambda E_{x\sim p_x(x)}[\|F(G_X(G_Y(x))) - F(x)\|_1]\end{aligned} \tag{8}$$

where F is a pre-trained segmentation network in the source domain, and F is frozen during the training process.

Through fooling the domain discriminator with the transferred source and target representations, RAN is able to produce domain-invariant representations. Therefore, JPRNet firstly performs the pixel level domain transfer from the source domain to the target domain, and the transferred images are then input into RAN for representation level domain adaption.

Algorithm 1 JPRNet training details and process

Input:

Data: source domain downsampling Inria images X, target domain Massachusetts Buildings images Y, source domain labels C, suppose x $\in$ X, y $\in$ Y, $c \in C$.

Output:

Predicted labels of the target domain: C_y

while iteration is effective **do**
  $y_{fake} \leftarrow G_Y(x)$ {forward pass}
  $D_{Ymap} \leftarrow D_Y(\{y_{fake}, y\})$ {forward pass}
  $x_{fake} \leftarrow G_x(y_{fake})$ {forward pass}
  Compare(x, x_{fake}) {Consistency comparison}
  The above process is a cycle from the source domain to the target domain. The process from the target domain to the source domain is similar to this.
  $\{R_{fakey}, R_y\} \leftarrow F(\{y_{fake}, y\})$ {forward pass}
  $Seg_{map} \leftarrow$ Seg($\{R_{fakey}, c\}$) {forward pass}
  $D_{map} \leftarrow D(\{R_{fakey}, R_y\})$ {forward pass}
  G_Y, D_Y, G_X, D_X can be optimized according to equation (3).
  FCN, Classifier, and Discriminator can be optimized according to equation (7).
end while

JPRNet is an end-to-end network that combines the pixel level domain adaptation with the representation level domain adaptation. The loss function of JPRNet can be written as follow:

$$\begin{aligned}\mathcal{L}_{JPRNet}(\tilde{G}, \tilde{D}, FCN, D, Seg) = \mathcal{L}_{cyclegan}(\tilde{G}, \tilde{D})+\\ \mathcal{L}_{RAN}(F, D, Seg) + \mathcal{L}_{sem}(G_X, F) + \mathcal{L}_{sem}(G_Y, F)\end{aligned} \tag{9}$$

The major difference between JPRNet and FCAN is that JPRNet proposes an end-to-end training method for remote sensing image domain adaptation. Due to the "multi-source" problem of remote sensing images, the images captured by different sensors can be considered to come from different domains. In natural scenes, there is basically no influence of different cameras on the domain. It is impossible to set a specific domain adaptation network for any two remote sensing data sets. Therefore, we propose an end-to-end domain adaptive semantic segmentation network that can reduce human intervention.

The pseudocode of our algorithm is shown in Algorithm 1.

III. Experiments

A. Dataset and evaluation Metrics

To verify the performance of JPRNet, it is tested on the downsampled Inria dataset and Massachusetts Buildings dataset.

TABLE I
Results of baselines and our domain adaptation(%)

Methods	baseline-1	baseline-2	baseline-3	PAN	FCAN	JPRNet
+FCN	✓					
+PSPNet		✓				
+DeepLab V3			✓	✓	✓	✓
+AAN					✓	
+Cycle GAN				✓		✓
+RAN					✓	✓
IoU	56.2	57.0	58.8	60.5	61.6	62.5

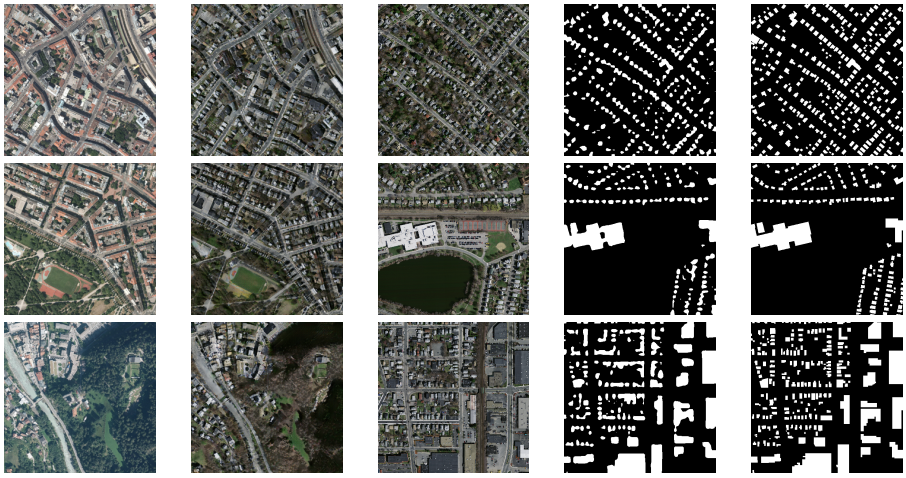

Fig. 2. The images from the left column to the right column are the source domain images(downsampled Inria dataset), the transferred source domain images, the target domain images(Massachusetts Buildings testing dataset), the predicted labels and the ground truth.

Both Massachusetts Buildings dataset and the Inria dataset only contain two categories: buildings and background. Inria dataset contains 180 images of size 5000×5000. The resolution is 0.3m. Massachusetts Buildings dataset contains 151 images from aerial images of Massachusetts. The size of images is 1500×1500, and the resolution is 1m. Since Inria dataset has a higher resolution than Massachusetts Buildings dataset, we downsample the images and labels in Inria dataset from 0.3m resolution to 1m resolution with the way of average downsampling. Considering the capacity of the GPU, we cut each training sample to several 500×500 sub-images, and totally obtain 1000 pieces for training. The code of JPRNet was published in our homepage.[1]

In experiments, Intersection Over Union (IoU) is used as the evaluation metrics. It is defined as follows:

$$IoU = N_{TP}/(N_{FP} + N_{TP} + N_{FN}) \tag{10}$$

where N_{TP}, N_{FP} and N_{FN} respectively represent the number of the true positive pixels, the false positive pixels and the false negative pixels in segmentation results.

B. Implementation Details

In the pixel level domain adaptation part, the generator $\tilde{G}$ and the discriminator $\tilde{D}$ use the same configuration as [13]. In the representation level domain adaptation part, we take FCN as the segmentation network and the generator of representations. FCN is built based on ResNet-50 by removing its fully connected layers and adding a 1×1 convolution layer. Besides, to increase the output resolution, we change the stride from 2 to 1 at Conv_3 and Conv_4 to enlarge its output size from 1/32 to 1/8 of its input size. ASPP, which is the classifier of DeepLab V3, is also used as the classifier of RAN. In the adversarial branch, we use k dilated convolutions in parallel to produce multiple feature maps, each with c channels. The sampling rate of different dilated convolution kernels is respectively 1, 2, 3 and 4. Finally, after the discriminator of ASPP, a sigmoid layer is utilized to output the prediction, which is in the range of [0, 1]. In Cycle GAN part, we train Cycle GAN from a pre-trained model. After JPRNet was trained for 100 epochs, Cycle GAN was fixed, the batch size was set to 8, and another 3 epochs were trained to converge the network. We set μ = 0.01, k = 4, c = 128 and λ = 10.

C. Comparison and Ablation Study

To validate the performance, JPRNet is compared with the existing methods. These methods include FCN, PSPNet, DeepLab V3 and FCAN. FCN, PSPNet, DeepLab V3 do not adopt domain adaptation algorithms. FCAN realizes domain adaptation. They and JPRNet are respectively trained on the downsampled Inria dataset (the source domain) and then tested on Massachusetts Buildings dataset (the target domain). The experimental results are shown in TABLE I. From the table, we observe that the results from JPRNet are prior to those from these methods.

To further evaluate the effectiveness of PAN and JPRNet, we use ablation experiments to guide the analysis of the importance of each component. These components include three baselines, Cycle GAN and RAN. The results are shown in TABLE I. FCN, PSPNet and DeepLab V3 are firstly evaluated as baselines. According to the evaluated results, DeepLab V3 is chosen as the baseline in next experiments. Then we gradually integrate Cycle GAN and RAN. And they are compared with FCAN.

[1]http://levir.buaa.edu.cn/Code.htm

- **DeepLab V3**: It is firstly trained on the source domain, and then the model is evaluated on the target domain.
- **AAN**: It is the pixel level domain adaptation algorithm in FCAN.
- **Cycle GAN**: Cycle GAN is used to realize the pixel level domain adaptation in this paper.
- **RAN**: We perform the representation level adaptation on the transferred source domain images and the target domain images.

The evaluation results are given in Table I. From the results, we could observe that the integration of the pixel level domain adaptation and the representation level domain adaptation effectively improves the segmentation accuracy. Some pixel-level domain adaptation results and building segmentation results of JPRNet are shown in Fig.2.

D. Semi-Supervised Adaptation

JPRNet can also be extended to a semi-supervised version by using these labeled images. In experiments, we add a small number of labeled target domain images during training JPRNet. Results are given in Table III. Here, four cases are compared. They are respectively JPRNet, JPRNet with 100 target domain labeled images, JPRNet with 200 target domain labeled images and three baselines on the whole target domain dataset. Experimental results show that the accuracy can be improved by adding a small amount of target domain images during training. The accuracy is near to that of training and testing on the whole target domain dataset.

TABLE II
Results of semi-supervised adaptation(%)

method \ baseline	FCN	PSPNet	DeepLab V3
JPRNet + 0 target	60.8	61.2	62.5
JPRNet + 100 target	61.5	62.7	63.3
JPRNet + 200 target	62.2	63.8	64.8
1000 target	65.3	66.3	66.5

IV. Conclusion

In this paper, we propose an end-to-end adaptive semantic segmentation architecture called JPRNet, which simultaneously conducts the pixel level and representation level domain adaptation. Pixel level and representation level domain adaptation could work together and complement each other in JPRNet. To this end, Cycle GAN is utilized to transfer an image style from the source domain to the target domain, and RAN is integrated to learn the domain invariant representation in an adversarial manner. Experimental results on the downsampled Inria dataset and Massachusetts Buildings dataset have demonstrated the effectiveness of JPRNet. Furthermore, the semi-supervised experiments indicate that JPRNet can obtain similar accuracy to baselines, which are trained and tested on the target domain.

References

[1] Y. Liu, B. Fan, L. Wang, J. Bai, S. Xiang, and C. Pan, "Semantic labeling in very high resolution images via a self-cascaded convolutional neural network," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 78–95, 2018.

[2] X. Yu, H. Zhang, C. Luo, H. Qi, and P. Ren, "Oil spill segmentation via adversarial f-divergence learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 4973–4988, 2018.

[3] B. Pan, X. Xu, Z. Shi, N. Zhang, H. Luo, and X. Lan, "DSSNet: A simple dilated semantic segmentation network for hyperspectral imagery classification," *IEEE Geoscience and Remote Sensing Letters*, 2020.

[4] Q. Wang, J. Gao, and X. Li, "Weakly supervised adversarial domain adaptation for semantic segmentation in urban scenes," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4376–4386, 2019.

[5] C. Persello and L. Bruzzone, "Kernel-based domain-invariant feature selection in hyperspectral images for transfer learning," *IEEE transactions on geoscience and remote sensing*, vol. 54, no. 5, pp. 2615–2626, 2015.

[6] S. Ghassemi, A. Fiandrotti, G. Francini, and E. Magli, "Learning and adapting robust features for satellite image segmentation on heterogeneous data sets," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[7] F. Schenkel and W. Middelmann, "Domain adaptation for semantic segmentation using convolutional neural networks," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 728–731.

[8] W. Liu, F. Su, and X. Huang, "Unsupervised adversarial domain adaptation network for semantic segmentation," *IEEE Geoscience and Remote Sensing Letters*, 2020.

[9] B. Benjdira, Y. Bazi, A. Koubaa, and K. Ouni, "Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images," *Remote Sensing*, vol. 11, no. 11, p. 1369, 2019.

[10] Q. Shi, M. Liu, X. Liu, P. Liu, P. Zhang, J. Yang, and X. Li, "Domain adaption for fine-grained urban village extraction from satellite images," *IEEE Geoscience and Remote Sensing Letters*, 2020.

[11] X. Deng, H. L. Yang, N. Makkar, and D. Lunga, "Large scale unsupervised domain adaptation of segmentation networks with adversarial learning," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 4955–4958.

[12] L. Yan, B. Fan, H. Liu, C. Huo, S. Xiang, and C. Pan, "Triplet adversarial domain adaptation for pixel-level classification of vhr remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[13] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[14] Y. Zhang, Z. Qiu, T. Yao, D. Liu, and T. Mei, "Fully convolutional adaptation networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6810–6818.

[15] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.
[16] Y. Li, L. Yuan, and N. Vasconcelos, "Bidirectional learning for domain adaptation of semantic segmentation," *arXiv: Computer Vision and Pattern Recognition*, 2019.