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, *Member, IEEE*, and Zhenwei Shi, *Member, IEEE*

***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 was developed for image semantic segmentation. In general, these methods are trained on the source domain and then validated on the target domain to avoid labeling new data repeatedly. However, most domain-adaptation algorithms only tried to align the source domain and the target domain in the pixel level or the representation level, while ignored their cooperation. In this letter, we propose an unsupervised domain-adaptation method by Joint Pixel and Representation level Network (JPRNet) alignment. The major novelty of the JPRNet is that it achieves joint domain adaptation in an end-to-end manner, so as to avoid the multisource problem in the remote sensing images. JPRNet is composed of two branches, each of which is a generative-adversarial network (GAN). In one branch, pixel-level domain adaptation is implemented by the style transfer with the 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 the public data sets have indicated the effectiveness of the JPRNet.**

***Index Terms*— Domain adaptation, generative-adversarial network (GAN), remote sensing, semantic segmentation.**

I. Introduction

SEMANTIC segmentation that aims at assigning a 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 the remote sensing images [1]–[3]. However, most of them have to train the models on the large labeled data sets, while it is a time-consuming process to collect such pixel-level annotated data sets.

An attractive alternative is to use domain adaptation that aims to transfer the model learned on a labeled source domain to a target domain. During the past decade, researchers have proposed some domain-adaptation algorithms for the remote sensing image semantic segmentation [4]–[7]. More recently, the generative-adversarial networks (GANs) have achieved promising performance in addressing the problem. In the domain-adaptation methods for the semantic segmentation of the remote sensing images, a 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 letter, inspired by the idea of hierarchical domain adaptation, we propose an end-to-end network, which can address the Joint Pixel and Representation level Network (JPRNet) domain adaptation. JPRNet is developed based on the Cycle GAN [13], which is a popular pixel-level backbone. A representation-level domain-adaptation approach is proposed to improve the Cycle GAN.

To some extent, the 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, the JPRNet is an end-to-end model, while an FCAN directly cascades two domain-adaptation algorithms. Due to the multisource problem of the remote sensing images, the images obtained by different satellites are quite different.If not adapting the end-to-end manner for the training domain-adaptation networks, users may have to select manually 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 that helps to improve the robustness of the algorithm.

The 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 the Cycle GAN that could transfer the image style from the source-domain images to the transferred source-domain images. In another branch, the Representation level Adaptation Network (RAN) is used to realize the domain-invariant representation between the transferred source-domain images and

Manuscript received January 16, 2020; revised February 24, 2020 and May 13, 2020; accepted July 15, 2020. Date of publication August 4, 2020; date of current version October 26, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC1405605, in part by the National Natural Science Foundation of China under Grant 61671037, in part by the Natural Science Foundation of Hebei Province of China under Grant F2020202008, and in part by the National Defense Science and Technology Key Laboratory China under Grant 61420020401. *(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 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).

Color versions of one or more of the figures in this letter are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2020.3010591

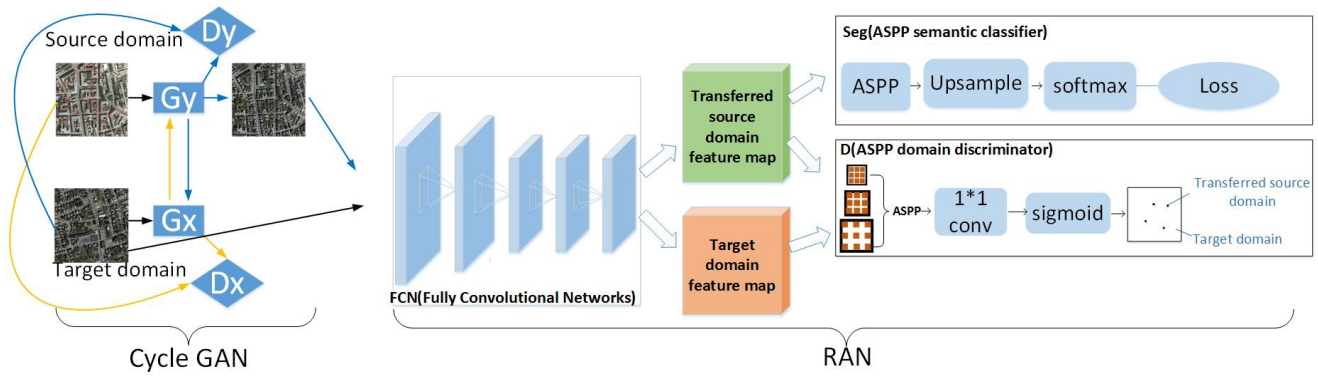

Fig. 1. Overall architecture of the proposed method. It consists of two main components: the pixel-adaptation network (Cycle GAN) on the left and the representation-adaptation network (RAN) on the right. The Cycle GAN could transfer the image style of the source domain. The RAN learns the domain-invariant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

the target-domain images. Our contributions are summarized as follows.

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

II. Methods

Our proposed adaptive semantic-segmentation network is illustrated in Fig. 1. It consists of the pixel-adaptation network (Cycle GAN) and the RAN. Given the images from the source domain and the target domain, the Cycle GAN transfers images from one domain to the other from the perspective of the pixel level in an adversarial manner. The 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. The 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 the domain-adaptation problem from both the pixel level and the representation level.

A. Pixel-Level-Adaptation Network (PAN)

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

The PAN is designed by adding an FCN on the basis of Cycle GAN. Therefore, we first 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:

$$\mathcal{L}_{X\to Y} = E_{y\sim p_y(y)}[\log D_Y(y)] + E_{x\sim p_x(x)}[\log(1 - D_Y(G_Y(x)))]. \quad (1)$$

The most significant difference between the 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:

$$\mathcal{L}_{\text{cyc}}(G_X, G_Y) = E_{x\sim p_x(x)}[\|G_X(G_Y(x)) - x\|_1] + E_{y\sim p_y(y)}[\|G_Y(G_X(y)) - y\|_1]. \quad (2)$$

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

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

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

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

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

$$\mathcal{L}_{\text{seg}}(\text{FCN}, G_Y) = -E_{(x,c)\sim p(x,c)}[c\log(\text{FCN}(G_Y(x))) + (1-c)\log(1 - \text{FCN}(G_Y(x)))]. \quad (4)$$

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

$$\mathcal{L}_{\text{PAN}}(\tilde{G}, \tilde{D}, \text{FCN}) = \mathcal{L}_{\text{cyclegan}}(\tilde{G}, \tilde{D}) + \mathcal{L}_{\text{seg}}(\text{FCN}, G_Y). \quad (5)$$

B. Representation level Adaptation Network (RAN)

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

ASPP [15] uses multirate dilated convolution to extract multiscale features in the form of spatial pyramid, which has proven to be effective in extracting multiscale information. The ASPP semantic classifier *Seg* could promote segmentation results by fusing multiscale 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 the 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 the buildings in the remote sensing images have different sizes, we attempt to use multiscale representations to enhance adversarial learning and building segmentation. It is the traditional way for solving multiscale problems to adjust the resolution of the input image and use the 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 multiscale problem of segmentation but also to solve the multiscale discrimination of the adversarial network.

C. Joint Pixel and Representation level Network (JPRNet)

The JPRNet adds a representation-level domain adaptation based on The Cycle GAN. As shown in Fig. 1, the Cycle GAN can achieve 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 the domain-invariant representations between the transferred source-domain images and the Massachusetts Buildings data set (the target domain). The RAN is used to produce representations across the domains and segment the transferred source-domain images. Suppose that Y_{fake} represents the transferred source-domain data set, Y represents the target domain data set, $y_{\text{fake}} \in Y_{\text{fake}}$, and $y \in Y$, the adversarial objective function and the objective function of the RAN can be, respectively, written as

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

$$\begin{aligned}&\mathcal{L}_{\text{RAN}}(\text{FCN}, D, \text{Seg}) \\ &\quad = \mathcal{L}_{\text{seg}}(\text{FCN}, \text{Seg}) + \mu\mathcal{L}_{\text{adv}}(\text{FCN}, D) \quad (7)\end{aligned}$$

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

In addition, similar to the literature [16], we also add the loss of semantic consistency as follows:

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

where F is a pretrained 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, the RAN is able to produce domain-invariant representations. Therefore, the JPRNet first performs the pixel-level domain transfer from the source domain to the target domain, and the transferred images are then input into the RAN for the representation-level domain adaption.

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

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

The major difference between the JPRNet and the FCAN is that the JPRNet proposes an end-to-end training method for remote sensing image domain adaptation. Due to the "multisource" problem of the remote sensing images, the images captured by different sensors can be considered to come from different domains. In the 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. Data Set and Evaluation Metrics

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

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

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

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

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

1: **while** iteration is effective **do**
2: $y_{fake} \leftarrow G_Y(x)$ {forward pass}
3: $D_{Ymap} \leftarrow D_Y(\{y_{fake}, y\})$ {forward pass}
4: $x_{fake} \leftarrow G_x(y_{fake})$ {forward pass}
5: 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.
6: $\{R_{fakey}, R_y\} \leftarrow F(\{y_{fake}, y\})$ {forward pass}
7: $Seg_{map} \leftarrow \text{Seg}(\{R_{fakey}, c\})$ {forward pass}
$D_{map} \leftarrow D(\{R_{fakey}, R_y\})$ {forward pass}
8: 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).
9: **end while**

In the experiments, intersection over union (IoU) is used as the evaluation metrics. It is defined as follows:

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

where N_{TP}, N_{FP}, and N_{FN}, respectively, represent the number of the true-positive pixels, false-positive pixels, and false-negative pixels in the 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. The FCN is built based on ResNet-50 by removing its fully connected layers and adding a 1×1 convolution layer. Moreover, 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 the 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 discrimination of ASPP, a sigmoid layer is used to output the prediction, which is in the range of [0, 1]. In the Cycle GAN part, we train Cycle GAN from a pretrained model. After JPRNet was trained for 100 epochs, the Cycle GAN was fixed, the batch size was set to 8, and another three epochs were trained to converge the network. We set $\mu = 0.01$, $k = 4$, $c = 128$, and $\lambda = 10$.

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, and DeepLab V3 do not adapt the domain-adaptation algorithms. The FCAN realizes domain adaptation. These methods and JPRNet are, respectively, trained on the downsampled Inria data set (the source domain) and then tested on the Massachusetts Buildings data set (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 evaluate further the effectiveness of the 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. The FCN, PSPNet, and DeepLab V3 are first evaluated as baselines. According to the evaluated results, DeepLab V3 is chosen as the baseline in the next experiments. Then, we gradually integrate Cycle GAN and RAN. In addition, they are compared with the FCAN.

1) *DeepLab V3:* It is first trained on the source domain, and then the model is evaluated on the target domain.
2) *AAN:* It is the pixel-level domain-adaptation algorithm in FCAN.
3) *Cycle GAN:* Cycle GAN is used to realize the pixel-level domain adaptation in this letter.
4) *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. Semisupervised Adaptation

JPRNet can also be extended to a semisupervised version by using these labeled images. In experiments, we add a small number of labeled target-domain images during training of the JPRNet. Results are given in Table II. Here, four cases are compared. They are, respectively, JPRNet,

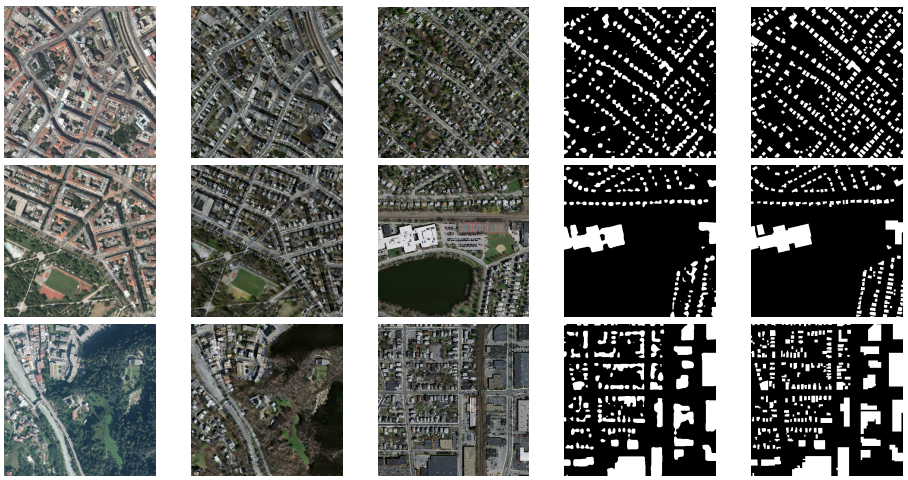

Fig. 2. Images from the left column to the right column are the source-domain images (downsampled Inria data set), the transferred source-domain images, the target-domain images (Massachusetts Buildings testing data set), the predicted labels, and the ground truth.

TABLE II
RESULTS OF SEMISUPERVISED 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

JPRNet with 100 target-domain labeled images, JPRNet with 200 target-domain labeled images, and three baselines on the whole target-domain data set. 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 data set.

IV. CONCLUSION

In this letter, we propose an end-to-end adaptive semantic-segmentation architecture called JPRNet, which simultaneously conducts the pixel-level and representation-level domain adaptations. Pixel-level and representation-level domain adaptations could work together and complement each other in the JPRNet. To this end, Cycle GAN is used 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 data set and the Massachusetts Buildings data set have demonstrated the effectiveness of JPRNet. Furthermore, the semisupervised experiments indicate that the JPRNet can obtain similar accuracy to the 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 J. Photogramm. Remote Sens.*, vol. 145, pp. 78–95, Nov. 2018.

[2] X. Yu, H. Zhang, C. Luo, H. Qi, and P. Ren, "Oil spill segmentation via adversarial f-divergence learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 4973–4988, Sep. 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 Geosci. Remote Sens. Lett.*, early access, Jan. 6, 2020, doi: 10.1109/LGRS.2019.2960528.

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

[5] C. Persello and L. Bruzzone, "Kernel-based domain-invariant feature selection in hyperspectral images for transfer learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2615–2626, May 2016.

[6] S. Ghassemi, A. Fiandrotti, G. Francini, and E. Magli, "Learning and adapting robust features for satellite image segmentation on heterogeneous data sets," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6517–6529, Sep. 2019.

[7] F. Schenkel and W. Middelmann, "Domain adaptation for semantic segmentation using convolutional neural networks," in *Proc. IGARSS-IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 728–731.

[8] W. Liu, F. Su, and X. Huang, "Unsupervised adversarial domain adaptation network for semantic segmentation," *IEEE Geosci. Remote Sens. Lett.*, early access, Dec. 13, 2019, doi: 10.1109/LGRS.2019.2956490.

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

[10] Q. Shi *et al.*, "Domain adaption for fine-grained urban village extraction from satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1430–1434, Aug. 2020.

[11] X. Deng, H. L. Yang, N. Makkar, and D. Lunga, "Large scale unsupervised domain adaptation of segmentation networks with adversarial learning," in *Proc. IGARSS-IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 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 Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3558–3573, May 2020.

[13] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.

[14] Y. Zhang, Z. Qiu, T. Yao, D. Liu, and T. Mei, "Fully convolutional adaptation networks for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6810–6818.

[15] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587

[16] Y. Li, L. Yuan, and N. Vasconcelos, "Bidirectional learning for domain adaptation of semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6936–6945.