ISPRS Journal of Photogrammetry and Remote Sensing 175 (2021) 20–33

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation

Yansheng Li [a,*], Te Shi [a], Yongjun Zhang [a], Wei Chen [a], Zhibin Wang [b], Hao Li [b]

[a] School of Remote Sensing and Information Engineering, Wuhan University, China
[b] Alibaba Group, China

ARTICLE INFO

Keywords:
Cross-domain remote sensing (RS) image semantic segmentation
Weakly-supervised transfer invariant constraint (WTIC)
Weakly-supervised pseudo-label constraint (WPLC)
Weakly-supervised rotation consistency constraint (WRCC)
DualGAN
Dynamic optimization strategy

ABSTRACT

Due to its wide applications, remote sensing (RS) image semantic segmentation has attracted increasing research interest in recent years. Benefiting from its hierarchical abstract ability, the deep semantic segmentation network (DSSN) has achieved tremendous success on RS image semantic segmentation and has gradually become the mainstream technology. However, the superior performance of DSSN highly depends on two conditions: (I) massive quantities of labeled training data exist; (II) the testing data seriously resemble the training data. In actual RS applications, it is difficult to fully meet these conditions due to the RS sensor variation and the distinct landscape variation in different geographic locations. To make DSSN fit the actual RS scenario, this paper exploits the cross-domain RS image semantic segmentation task, which means that DSSN is trained on one labeled dataset (i.e., the source domain) but is tested on another varied dataset (i.e., the target domain). In this setting, the performance of DSSN is inevitably very limited due to the data shift between the source and target domains. To reduce the disadvantageous influence of data shift, this paper proposes a novel objective function with multiple weakly-supervised constraints to learn DSSN for cross-domain RS image semantic segmentation. Through carefully examining the characteristics of cross-domain RS image semantic segmentation, multiple weakly-supervised constraints include the weakly-supervised transfer invariant constraint (WTIC), weakly-supervised pseudo-label constraint (WPLC) and weakly-supervised rotation consistency constraint (WRCC). Specifically, DualGAN is recommended to conduct unsupervised style transfer between the source and target domains to carry out WTIC. To make full use of the merits of multiple constraints, this paper presents a dynamic optimization strategy that dynamically adjusts the constraint weights of the objective function during the training process. With full consideration of the characteristics of the cross-domain RS image semantic segmentation task, this paper gives two cross-domain RS image semantic segmentation settings: (I) variation in geographic location and (II) variation in both geographic location and imaging mode. Extensive experiments demonstrate that our proposed method remarkably outperforms the state-of-the-art methods under both of these settings. The collected datasets and evaluation benchmarks have been made publicly available online (https://github.com/te-shi/MUCSS).

1. Introduction

Along with the rapid development of multiple fields, such as remote sensing (RS), computer science and communication engineering, RS images have been growing explosively, which makes large-scale earth surface monitoring possible. As a consequence, we have entered an age of RS big data (Ma et al., 2015; Li et al., 2016; Chi et al., 2016). In this age, automatic interpretation of RS images plays an important role in effectively mining the value of RS big data. Due to its wide usage in urban planning (Shi et al., 2015; Kampffmeyer et al., 2016), crop assessment (Kussul et al., 2017; Ozdarici-Ok et al., 2015), environment monitoring (Yu et al., 2018a) and intelligent traffic (Zhang et al., 2018), RS image semantic segmentation has attracted increasing research interest in recent years. Specifically, RS image semantic segmentation

* Corresponding author.
E-mail addresses: yansheng.li@whu.edu.cn (Y. Li), te.shi@whu.edu.cn (T. Shi), zhangyj@whu.edu.cn (Y. Zhang), weichenrs@whu.edu.cn (W. Chen), zhibin.waz@alibaba-inc.com (Z. Wang), lihao.lh@alibaba-inc.com (H. Li).

https://doi.org/10.1016/j.isprsjprs.2021.02.009
Received 18 September 2020; Received in revised form 15 January 2021; Accepted 5 February 2021
Available online 6 March 2021

aims at assigning one land-cover type (e.g., building, tree) to each pixel in the RS image. Although semantic segmentation has also been exploited in the computer vision field (Tao and Liu, 2017), RS image semantic segmentation generally suffers from some additional challenges, such as the complex structures of RS images and flexible imaging orientation of remote sensors.

In the past ten years, convolutional neural networks (CN-Ns) have achieved great success in content-based RS image retrieval (Li et al., 2017), RS image object detection (Li et al., 2018, 2020a) and scene-level RS image classification (Huang et al., 2018; Li et al., 2020b). As an extension of CNN, a deep semantic segmentation network (DSSN) was first proposed by the pioneers in the computer vision community (Long et al., 2015a) and further introduced to address RS image semantic segmentation (Hu et al., 2015; Yue et al., 2015; Makantasis et al., 2015; Volpi and Tuia, 2018; Mi and Chen, 2020). When massive RS images with well-annotated pixel-level labels exist, DSSN can be trained effectively in an end-to-end manner and obviously outperforms traditional methods (e.g., hand-crafted feature-based methods) (Zhang et al., 2016; Lyu et al., 2020). However, the superior performance of DSSN highly depends on the strong supervision (i.e., there is a large quantity of labeled training data) and resembled data distribution (i.e., the testing data and training data have similar appearance characteristics).

As reported in (Cordts et al., 2016), pixel-level annotation of one natural Cityscapes image takes almost 90 min on average. Compared with natural images, RS images generally present more complex structures (Yue et al., 2019). Due to the inter-class confusion of RS images, the annotation process often requires massive domain expert knowledge. Overall, both the complex structures and inter-class confusion of RS images further make pixel-level annotation of RS images more time-consuming and costlier. In the age of RS big data, it becomes increasingly easier to collect RS images, but constructing the pixel-level labels for the RS images becomes the actual challenge. With this consideration, RS image semantic segmentation needs much more exploration around how to decrease the supervision dependency of labeled data. One potential solution is to train the DSSN with the labeled RS images from the source domain and then utilize the trained DSSN to interpret the RS images from the target domain. Apparently, this highly resembles the classic cross-domain semantic segmentation task in the computer vision field. As shown in (Chen et al., 2017; Hoffman et al., 2016, 2018), the data shift between the source and target domains often makes the performance of the DSSN degenerate seriously.

Compared with cross-domain semantic segmentation in the computer vision field, cross-domain RS image semantic segmentation suffers from more challenges. Due to the diversity of RS image acquisition conditions including imaging sensors, varied geospatial regions, ground sampling distances (GSDs) and arbitrary shooting angles (Bruzzone and Carlin, 2006; Tuia et al., 2016), RS images often present many distinct characteristics such as variety of imaging mode, multi-scale of objects and variety of color saturation. In reality, these RS characteristics are often intertwined, which dramatically enlarges the cross-domain RS image variation. For example, the variety of imaging mode in the RS field is hardly involved in the computer vision field. Compared with the cross-domain semantic segmentation task in the computer vision field (Hoffman et al., 2018; Tsai et al., 2018; Zou et al., 2018; Xu et al., 2019), cross-domain RS image semantic segmentation seems to be more challenging. So, if the cross-domain adaption methods proposed in the computer vision field are directly used to do cross-domain RS image semantic segmentation, they often do not perform as well as expected. Thus, cross-domain RS image semantic segmentation needs much more special exploration by considering the RS characteristics.

With the aforementioned consideration, this paper focuses on cross-domain RS image semantic segmentation. To minimize the disadvantageous influence of the data shift between the source and target domains, this paper proposes a new objective function with multiple weakly-supervised constraints to learn DSSN where multiple weakly-supervised constraints are composed of weakly-supervised transfer invariant constraint (WTIC), weakly-supervised pseudo-label constraint (WPLC) and weakly-supervised rotation consistency constraint (WRCC). Specifically, WTIC aims to construct the image relationship between the source and target domains with the aid of one carefully examined unsupervised style transfer model (i.e., DualGAN (Yi et al., 2017)). By the classification confidence filter, the anchor points with pseudo-labels in the images from the target domain are adaptively selected to carry out WPLC. Based on the primary fact that the inverse transformation of the segmentation result of one rotated image should equal the result of the original image, WRCC depicts the generalized rotation consistency property of the images from the target domain. All three constraints follow a weakly-supervised manner, and the image labels from the source domain are fully considered in WTIC. To make full use of the merits of multiple constraints, this paper presents a dynamic optimization strategy that dynamically adjusts the constraint weights of the objective function during the training process, which helps to alleviate the degeneration of the DSSN. Considering the special characteristics of the cross-domain RS image semantic segmentation task, this paper constructs two experimental scenarios: (I) variation in geographic location and (II) variation in both geographic location and imaging mode. Extensive experiments show that the proposed method can obviously outperform the state-of-the-art methods under these two kinds of experimental settings. The main contributions of this paper can be summarized as follows:

1) This paper proposes a novel objective function with multiple weakly-supervised constraints to learn DSSN for cross-domain RS image semantic segmentation where multiple weakly-supervised constraints include WTIC, WPLC and WRCC. Specifically, WTIC intends to bridge the images from the source and target domains, WPLC is built by adaptively mining the anchor points with pseudo-labels and WRCC is built by leveraging the rotation consistency characteristic.
2) To make full use of the well-designed constraints, this paper proposes a dynamic optimization strategy, which dynamically adjusts the constraint weights of the objective function during the training process. Extensive experimental results show the effectiveness of the presented dynamic optimization strategy in avoiding the degradation of DSSN.
3) To the best of our knowledge, this paper, for the first time, introduces DualGAN to address cross-domain RS image semantic segmentation. Compared to other unsupervised style transfer methods, this paper also demonstrates the adaptation of DualGAN for RS image style transfer from both intuitive analysis and experimental verification perspectives under our proposed framework.
4) To fully reflect the special characteristics of the RS field, this paper gives two representative cross-domain experimental settings: variation in geographic location and variation in both geographic location and imaging mode. Accordingly, we construct two evaluation datasets. Based on these two datasets, this paper releases a new benchmark for cross-domain RS image semantic segmentation.

The rest of this paper is organized as follows. Section 2 discusses the related work. Section 3 introduces the proposed method in detail. Section 4 reports the experiments and provides a discussion of the experimental results. Finally, the conclusion and potential future research directions are outlined in Section 5.

2. Related work

In this section, we briefly review the most relevant works in the literature that include cross-domain semantic segmentation in both the computer vision and remote sensing fields.

2.1. Cross-domain semantic segmentation in the computer vision field

In the field of computer vision, cross-domain semantic segmentation

has been widely exploited because of its importance in real-world missions. Until now, kinds of cross-domain semantic segmentation methods have been proposed. Generally, traditional cross-domain semantic segmentation approaches try to learn domain invariant embedding by reducing the distribution difference between the source and target domains. For example, the maximum mean deviation (MMD) (Long et al., 2015b; Tzeng et al., 2014) and its kernel variables are the most common targets for minimizing the cross-domain differences of feature distributions. With the idea of narrowing the differences between the source and target domains, a large number of approaches have been proposed. The existing methods along this avenue can be coarsely divided into two major categories: adversarial learning methods and self-learning methods. The former adopts adversarial learning to discover domain invariant representations in feature spaces. Hoffman et al. (2018) proposed a novel discriminatively trained cycle-consistent adversarial domain adaptation model that seeks to reduce the domain shift by transferring source images to the target style with a cycle consistency loss and then aligning the cross-domain feature distributions of the task network through adversarial training. Tsai et al. (2018) adopted adversarial learning in the output space considering semantic segmentations as structured outputs that contain spatial similarities between the source and target domains. By contrast, the latter takes advantage of what is learned in the source domain and then modifies it to apply to the target domain. Zou et al. (2018) proposed a novel framework based on an iterative self-training procedure, where the problem is formulated as latent variable loss minimization and can be solved by alternatively generating pseudo-labels on target data and re-training the model with these labels. Xu et al. (2019) utilized the self-ensembling attention network to extract attention-aware features for domain adaptation. Although these methods have achieved a certain extent of success on natural images, they still cannot adequately address cross-domain RS image semantic segmentation, as RS imagery often shows a more complex structure.

2.2. Cross-domain semantic segmentation in the remote sensing field

In the RS community, most of the cross-domain methods are designed for scene-level classification tasks. For example, Song et al. (2019) designed a subspace alignment (SA) and CNN-based framework to solve the cross-domain remote sensing image scene-wise classification. Othman et al. (2017) proposed a domain adaptation network (DAN) method, which aims to project the source and target data into a common space to reduce the discrepancy between source and target distributions while using graph regularization to maintain the geometrical structure of the target data. Yan et al. (2019) proposed a cross-domain distance metric learning (CDDML) framework to address cross-domain classification. Zhu et al. (2019) proposed a semi-supervised center-based discriminative adversarial learning (SCDAL) framework for cross-domain classification. In contrast, the pixel-level cross-domain RS image classification (i.e., the argued cross-domain RS image semantic segmentation in this paper) task is rarely exploited. Zhao et al. (2017) proposed a method that relied on deep neural networks for presenting the contextual information contained in different types of land covers and use a pseudo-labeling and sample selection scheme to improve the transferability of deep models to achieve cross-domain pixel-wise classification. The pioneers in (Bilel et al., 2019) were the first to propose a domain adaptation method to address cross-domain aerial image classification where CycleGAN (Zhu et al., 2017) is utilized in the domain adaptation process. The performance of the existing methods is still limited because they do not thoroughly consider the characteristics of cross-domain semantic segmentation. Therefore, cross-domain RS image semantic segmentation needs to be further studied.

3. Methodology

To facilitate clarifying the methodology, we first formulate the involved data presentation in Table 1. Let S denote the source dataset and T denote the target dataset. $S = \{(I_1^S, L_1^S), (I_2^S, L_2^S), \cdots, (I_N^S, L_N^S)\}$ contains N images, and the manually constructed label of the n-th image I_n^S is depicted by L_n^S. Let $I_i^S \in \mathbb{R}^{H\times W\times C}$ denote the i-th image, $L_i^S \in \{0,1\}^{H\times W\times C}$ represent the corresponding label, where H and W represent the height and width of the image, and C denotes the number of classes. Specifically, $L_i^S(h,w)$ is a one-hot label vector for pixel(h,w) of image I_i^S. In addition, we use $T = \{I_1^T, I_2^T, \cdots, I_M^T\}$ to represent the M images without labels from the target dataset.

Furthermore, Fig. 1 gives an overview of our proposed framework. As shown in Fig. 1, this paper proposes a novel objective function with multiple weakly-supervised constraints to learn DSSN for cross-domain RS image semantic segmentation where multiple weakly-supervised constraints include the WTIC, WPLC and WRCC. More specifically, DualGAN is recommended to conduct unsupervised style transfer between the source and target domains to carry out WTIC. By the classification confidence filter, the anchor points with pseudo-labels in the images from the target domain are adaptively selected to carry out WPLC. Based on the primary fact that the inverse transformation of the segmentation result of one rotated image should be consistent with the result of the original image, WRCC depicts the generalized rotation consistency property of the images from the target domain. To balance these multiple constraints, the optimization procedure is dynamic, which can efficiently avoid DSSN falling into a degeneration case. After training under these multiple weakly-supervised constraints, the DSSN can work well on the target dataset.

In the following, Section 3.1 introduces DualGAN to conduct unsupervised style transfer between source and target domains to carry out WTIC. We also argue the merits of DualGAN. In Section 3.2, a novel objective function with multiple weakly-supervised constraints is described, where multiple weakly-supervised constraints include WTIC, WPLC and WRCC. Finally, the dynamic optimization strategy is explained in Section 3.3.

3.1. Unsupervised style transfer

For the RS image cross-domain semantic segmentation task, the source domain has a large amount of labeled data, but the images from the target domain do not have any labels. Obviously, how to fully mine the invariant semantic features of RS image data from different domains and efficiently perform cross-domain RS image semantic segmentation is still an open problem. In the literature, by mining the invariant semantic features between the source and target domains, unsupervised style transfer can map the source images to the style of the target domain. Then, the transferred images with original labels are used to train the DSSN. Hence, this strategy naturally reduces the effect of domain shift. The existing style transfer algorithms are mainly constructed based on

Table 1
Notations.

Notation	Meaning
S	Source domain dataset
T	Target domain dataset
S'	Transferred dataset
T'	Target dataset with pseudo labels
$I_i^S (i = 1,2,\cdots,N)$	Image of source dataset
$I_i^G (i = 1,2,\cdots,N)$	Image of transferred dataset
$I_i^T (i = 1,2,\cdots,M)$	Image of target dataset
$L_i^S (i = 1,2,\cdots,N)$	Label of source dataset
$L_i^E (i = 1,2,\cdots,M)$	Pseudo label of target dataset

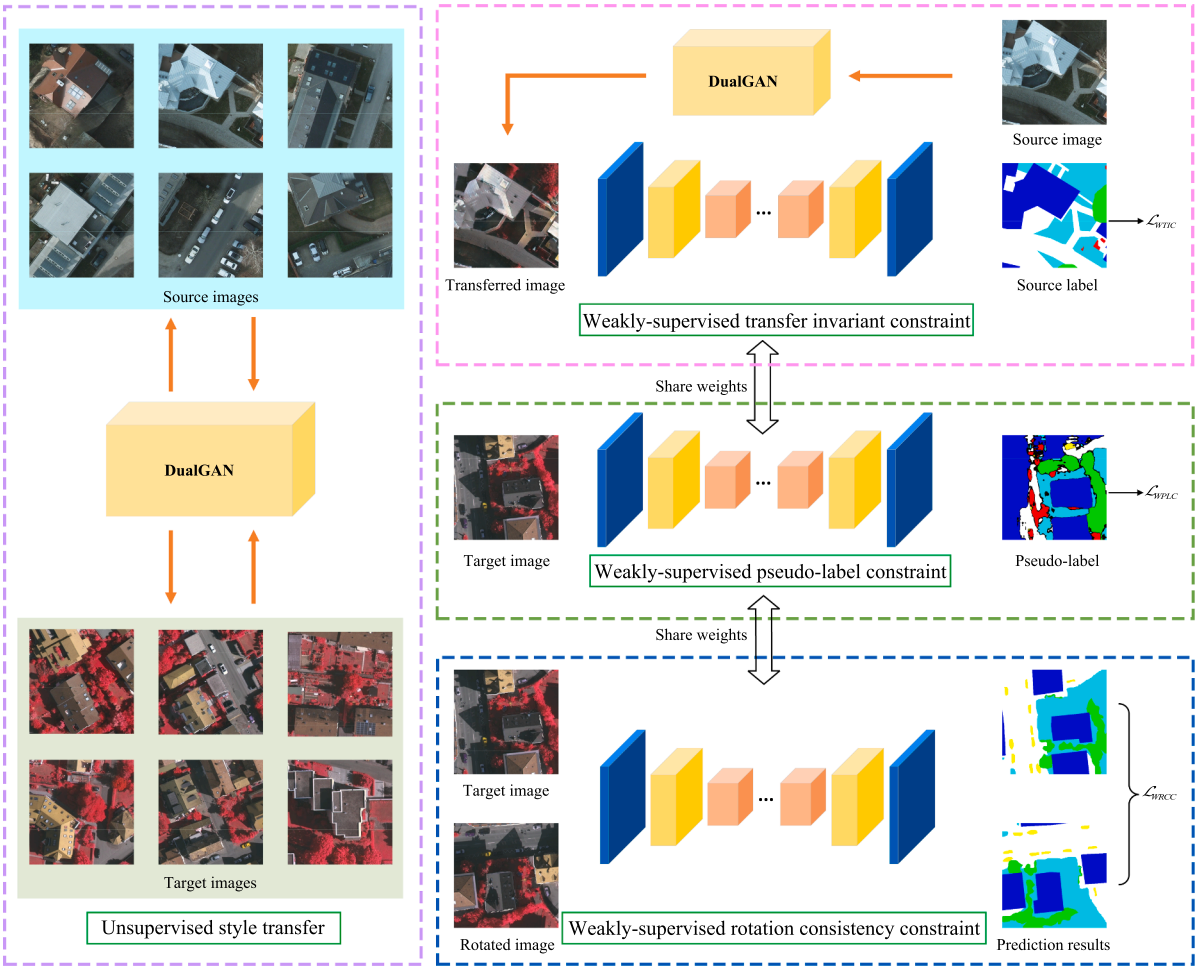

Fig. 1. Flowchart of the proposed method.

generative adversarial networks (GANs) (Goodfellow et al., 2014) and can be roughly divided into two categories: methods that require paired images (Hertzmann et al., 2001; Isola et al., 2017) and methods that do not require paired images (Zhu et al., 2017; Yi et al., 2017; Kim et al., 2017). For the former, the strictly paired images are taken as one kind of supervision constraint. For example, the conditional GAN (Isola et al., 2017), which depends on the pairs of corresponding images, is first proposed for image-to-image translation. However, this kind of method does not harmonize with the goal of generalizing cross-domain RS image semantic segmentation. The latter methods are not conditioned on the paired images, which is conducive to promoting the cross-domain RS image semantic segmentation task. In this direction, CycleGAN (Zhu et al., 2017) with the cycle consistency loss is developed to alleviate the paired information dependence. Afterwards, DiscoGAN (Kim et al., 2017) and DualGAN (Yi et al., 2017) were proposed to conduct unsupervised image-to-image translation. As stated in (Zhu et al., 2017), such unsupervised image-to-image translation methods perform well on style transfer tasks, especially involving color and texture changes. Compared with CycleGAN and DiscoGAN, DualGAN adopts one effective loss function advocated by Wasserstein GAN (WGAN), which makes the optimization procedure thorough and benefits generating high-quality images whose style is much closer to the target domain. As mentioned before, RS images present many domain characteristics such as variety of imaging mode, multi-scale of objects, variety of color saturation and arbitrary shooting angles due to the diversity of remote sensing image acquisition conditions. In practical cases, these issues are intertwined, resulting in very large style differences between remote sensing imagery from different domains. These aspects undoubtedly bring many challenges to carry out style transfer of RS images. With this consideration, DualGAN benefits thoroughly bridging the source and target domains in the RS task under our proposed framework. Hence, we recommend DualGAN to perform unsupervised image-to-image translation (i.e., the argued unsupervised style in this paper) in our proposed framework.

Intuitively, the workflow of DualGAN is visually shown in Fig. 2. As illustrated in Fig. 2, image $I^S \in S$ is converted to the target domain by G_A. Then, D_A is used to measure how well the translation $G_A(s,z)$ fits in the target domain, where z and z' is random noise to perform data augmentation. $G_A(s,z)$ is then converted back to the source domain by G_B, which outputs $G_B(G_A(s,z),z')$ as the reconstruction of I^S. Similarly, $I^T \in T$is translated to the source domain as $G_B(t,z')$ and then reconstructed as $G_A(G_B(t,z'),z)$. The discriminator D_A is trained with I^T as positive samples and $G_A(s,z)$ as negative examples, which means giving samples from T a high score and samples from $G_A(s,z)$ a low score. D_B is trained in the same way. Generators G_A and G_B are optimized to emulate "fake" outputs to confuse the corresponding discriminators D_A and D_B, as well as to minimize the reconstruction losses $\|I^S - G_A(G_B(t,z'),z)\|$ and $\|I^T - G_B(G_A(s,z),z')\|$.

With the trained DualGAN model, we transfer the source dataset S to the style of the target domain. In detail, for image $I^S \in S, I^G = G_A(I^S,z)$ is obtained. We combine the transferred images with original labels to obtain a new dataset $S' = \{(I_1^G, L_1^S), (I_2^G, L_2^S), \cdots, (I_N^G, L_N^S)\}$ whose style is similar to the target domain.

In the following, we specifically explain the adaption of DualGAN in the RS image style transfer task from the intuitive analysis perspective.

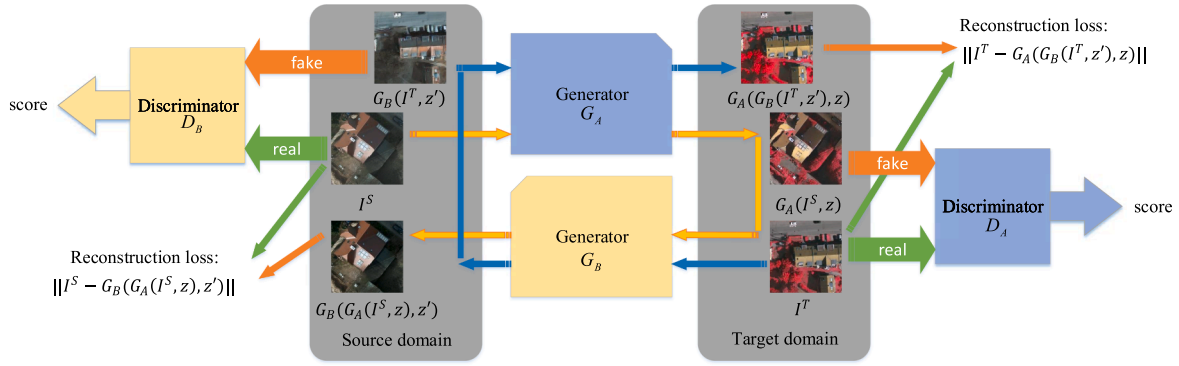

Fig. 2. Architecture of DualGAN for unsupervised style transfer.

As previously mentioned, RS images present many distinct characteristics such as variety of imaging mode, multi-scale of objects, variety of color saturation and arbitrary shooting angles. These factors undoubtedly make it more difficult to carry out unsupervised style transfer of RS images. Due to the advanced loss function (i.e., the Wasserstein loss) and network architecture (i.e., the U-shaped network architecture), DualGAN is more suitable to address the challenges in RS image style transfer. As far as the loss function, the adopted Wasserstein loss (Martin Arjovsky and Bottou, 2017) in DualGAN benefits pursuing a more sufficient update of the GAN generator compared with the traditional cross-entropy loss. More specifically, the Wasserstein loss employs the Wasserstein distance to sensitively measure the distance between the generated data and real data distributions, which helps to avoid getting into the dilemma of vanishing gradient. These characteristics of the adopted Wasserstein loss make DualGAN perform better in terms of the generator convergence speed and the stability of the optimization process. As a result, DualGAN has the ability to cope with the above variation characteristics of RS images. As for the network architecture, the U-shaped network backbone is employed in DualGAN which benefits eliminating the checkerboard artefacts and helps to generate high-quality and realistic images (Odena et al., 2016). Besides, the U-shaped network architecture can effectively fuse the low-level and high-level information, which is beneficial to style transfer of images containing multi-scale objects. Benefiting from the Wasserstein loss and U-shaped network architecture, DualGAN has a powerful ability to do unsupervised style transfer of RS images. With the aforementioned consideration, DualGAN is recommended to carry out unsupervised style transfer of RS images in this paper. By quantitatively comparing with the other unsupervised style transfer methods (e.g., DiscoGAN and CycleGAN), the advantage of the recommended DualGAN is further verified in the experimental section.

3.2. The objective function with multiple weakly-supervised constraints for learning a deep fully convolutional network

The proposed method aims to use the paired source domain images to train a model and then apply it to predict the label for the target dataset. According to the previous description, we first use DualGAN to transfer the source domain image to the style of the target domain to carry our WTIC. Then, we convert the source dataset to the style of the target domain by DualGAN. The output is a new dataset denoted as $S' = \{(I_1^G, L_1^S), (I_2^G, L_2^S), \cdots, (I_N^G, L_N^S)\}$, which is used to conduct WTIC during the training process. However, the transferred images cannot be completely consistent with the real target domain image, and there will be some differences. To this end, WPLC and WRCC are introduced in our proposed method, where WPLC is carried out by mining the anchor points with pseudo-labels and WRCC depicts the generalized rotational consistency property of the images from the target domain. Overall, all three constraints are weakly-supervised, and the image labels from the source domain are fully considered in WTIC. The total loss function can be formulated in Eq. (1).

$$\mathcal{L}_{total} = \left(1 - \frac{1}{2}\left(\alpha + \beta\right)\right)\mathcal{L}_{WTIC} + \alpha\mathcal{L}_{WPLC} + \beta\mathcal{L}_{WRCC} \tag{1}$$

where $\mathcal{L}_{WTIC}$, $\mathcal{L}_{WPLC}$ and $\mathcal{L}_{WRCC}$ represent the weakly-supervised transfer invariant constraint, weakly-supervised pseudo-label constraint and weakly-supervised rotation consistency constraint, respectively. α and β are two vital hyper-parameters that represent the weight of the weakly-supervised pseudo-label constraint and weakly-supervised rotation consistency constraint, respectively. In the initial phase of training, the network model is not stable enough. So, if α and β are too high, it disturbs training even for labeled data, and the network becomes easily stuck in a degenerate solution where no meaningful classification of the data is obtained. However, if α and β are too small, we cannot benefit from unlabeled data. Considering all these factors, we adopt the Gaussian ramp-up curve, which is also used in (Laine and Aila, 2016), to dynamically adjust the contributions of different constraints. More precisely, the α and β ramp up, starting from zero along a Gaussian curve $exp[-5(1-t)^2]$, where t equals zero first and then advances linearly on each *iter* and eventually increases to one. Thus, the weight of the first term $\mathcal{L}_{WTIC}$ is set to $(1 - 1/2(\alpha + \beta))$, decreasing from one to zero along with the training process going on, where $1/2$ is a normalization constant and aims at avoiding the weight to become a meaningless negative value.

3.2.1. weakly-supervised transfer invariant constraint

After training the DualGAN, the images in the source dataset are automatically transferred to approximate the style of the target aerial image dataset denoted as $S' = \{(I_1^G, L_1^S), (I_2^G, L_2^S), \cdots, (I_N^G, L_N^S)\}$, which benefits minimizing the influence of data shift between different domains. We use the transferred dataset S' to train a DSSN, and the cross-entropy loss function is shown as Eq. (2). In our implementation, DSSN is implemented by DeepLab v3 + (Chen et al., 2018) as it is the state-of-the-art semantic segmentation network and achieves excellent performance in the natural image semantic segmentation field.

$$\begin{aligned} &\mathcal{L}_{WTIC}(S'; \theta) \\ = \; &-\sum_{i=1}^{N} \ell_{CE}\left(P_i^G, L_i\right) \\ = \; &-\sum_{i=1}^{N}\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C}\left(L_i(h,w,c)\cdot\log\left(P_i^G(h,w,c)\right)\right) \end{aligned} \tag{2}$$

where ℓ_{CE} represents the cross-entropy loss function, $P_i^G = f_\theta(I_i^G)$ denotes the probability map, and $f_\theta(\cdot)$ represents a network with weight θ. H and W represent the height and width of the image, respectively. C denotes the number of land-cover categories.

3.2.2. weakly-supervised pseudo-label constraint

Since the transferred images cannot be completely consistent with the real target domain images, the transferred images only play a guiding role. Therefore, we consider using DSSN trained on $S^{'}$ to generate pseudo-labels with high confidence in the preparatory stage, which aims at improving the performance of DSSN for the target domain in the dynamic optimization stage. Pioneers in (Lee, 2013) proposed pseudo-label learning in semi-supervised learning and proved its effectiveness through a large number of experiments and analyzes. Therefore, we introduce the weakly-supervised pseudo-label constraint module into our framework.

Based on the aforementioned description, the DSSN can generate pseudo-labels with a high confidence level marked as $T^{'} = \{(I_1^T, L_1^E), (I_2^T, L_2^E), \cdots, (I_M^T, L_M^E)\}$ for the target dataset. Here, we use the $w \times h$ image with c classes to illusrate how to measure confidence. As we know, when the image passes through the final softmax function, we will obtain a matrix with the size of $w \times h \times c$. Each pixel has a vector with a size of $1 \times c$, and the index of the top1 value of the vector indicates the category. We calculate the difference between the top1 and top2 values. If the difference is larger than the threshold τ_{pse}, it is preserved. Otherwise, it is ignored and does not participate in the calculation of loss. The pseudo-label is available after this step. By applying this process to each image in T, we obtain $T^{'} = \{(I_1^T, L_1^E), (I_2^T, L_2^E), \cdots, (I_M^T, L_M^E)\}$.

In the dynamic optimization steps, the image from the target domain and its pseudo-label participate in training the network, and the cross-entropy loss function of this part is shown as Eq. (3).

$$\begin{aligned} & \mathcal{L}_{WPLC}(T^{'}; \theta) \\ = & -\sum_{i=1}^{M} \ell_{CE}\left(P_i^T, L_i^E\right) \\ = & -\sum_{i=1}^{M}\sum_{h=1}^{H}\sum_{W=1}^{W}\sum_{c=1}^{C}\left(L_i^E(h,w,c)\cdot\log\left(P_i^T(h,w,c)\right)\right) \end{aligned} \tag{3}$$

where ℓ_{CE} represents the cross-entropy loss function, $P_i^T = f_\theta(I_i^T)$ denotes the probability map, and $f_\theta(\cdot)$ represents the network with weight θ. H and W represent the height and width of the image, respectively. C stands for the number of land-cover categories.

3.2.3. weakly-supervised rotation consistency constraint

To use the unlabeled data, the weakly-supervised rotation consistency constraint is introduced into our framework. What we do is the pixel-wise classification (i.e., semantic segmentation), which is slightly different from scene-wise classification. Only when the transformation is completely reversible can the consistency loss be calculated. Therefore, we adopt the rotation transformation for the unlabeled image $I_i^T \in T$. Specifically, rotation transformation φ (random rotation of 90 degrees, 180 degrees, 270 degrees) is performed on the image I_i^T, and then we obtain $\widetilde{I_i^T} = \varphi(I_i^T)$. I_i^T and $\widetilde{I_i^T}$ are fed into the DSSN at the same time and two outputs $P_i^T = f_\theta(I_i^T), \widetilde{P_i^T} = f_\theta(\widetilde{I_i^T})$ are obtained. Different from the classification task, to compute the pixel-level consistency of two outputs, we have to perform the inverse transform to put every pixel to the original location. We denote inverse transforms of the random rotation as φ^{-1}. Thus, we can obtain the inverse transformed outputs $\overline{P_i^T} = \varphi^{-1}(\widetilde{P_i^T})$, and the weakly-supervised consistency loss can be computed. The consistency loss term often uses the mean squared error, which encourages the pixel-level consistency of the output under different random rotation transforms. The loss function can be described as Eq. (4).

$$\begin{aligned} & \mathcal{L}_{WRCC}(T; \theta) \\ = & \sum_{i=1}^{M} \ell_{MSE}\left(P_i^T, \overline{P_i^T}\right) \\ = & \frac{1}{H \times W}\sum_{i=1}^{M}\sum_{h=1}^{H}\sum_{W=1}^{W}\sum_{c=1}^{C}\left(\left\|P_i^T(h,w,c) - \overline{P_i^T}(h,w,c)\right\|_2\right) \end{aligned} \tag{4}$$

where H and W represent the height and width of the image, respectively, and C stands for the number of land-cover categories.

3.3. The dynamic optimization strategy for learning a deep semantic segmentation network

Before the joint dynamic optimization, some preparatory work needs to be performed. The first step is to train a DualGAN network and then transfer the images in the source domain to the style of the target domain, which bridges the images from the source and target domains. By this step, the influence of domain shift is eliminated to some degree, and initial training data are provided for DSSN. The next step is to let $\alpha = 0, \beta = 0$ and to train the network with the translated dataset $S^{'}$. This helps the model learn the patterns of the target dataset and converge to a better generalization ability of the image structure on the target dataset. After training, DSSN is used to generate pseudo-labels for the target dataset marked as $T^{'} = \{(I_1^T, L_1^E), (I_2^T, L_2^E), \cdots, (I_M^T, L_M^E)\}$.

Based on the aforementioned preparatory work, the final step is to jointly learn the DSSN from scratch in an end-to-end manner. Specifically, let $\alpha \neq 0, \beta \neq 0$ and update α and β based on the iteration, which means WTIC, WPLC and WRCC modules work together, where WTIC aims to construct the image relationship between the source and target domains, WPLC is carried out by mining the anchor points with pseudo-labels and WRCC depicts the generalized rotational consistency property of the target domain images. In addition, the optimization procedure is dynamic and can efficiently balance these multiple constraints and avoid the DSSN falling into a degeneration situation. By introducing these multiple constraints, the DSSN can achieve higher performance, especially the boundary of the object, which will be clearer. Finally, the semantic segmentation network is applicable to work on the target dataset.

To benefit understanding, the whole optimization algorithm for learning DSSN is summarized in Algorithm 1.

Algorithm 1. The presented dynamic optimization strategy

Input: $S = \{(I_1^S, L_1^S), (I_2^S, L_2^S), \cdots, (I_N^S, L_N^S)\}$, $T = \{I_1^T, I_2^T, \cdots, I_M^T\}$. **Output:** weights of DSSN.

The preparatory stage

- Using images from S and images from T to train the DualGAN;
- Transferring S to $S^{'}$ by using the trained DualGAN;
- Using the transferred dataset $S^{'}$ to learn DSSN based on the objective function in Eq. (1) with $\alpha = 0, \beta = 0$;
- Generating pseudo-labels of the images from the target domain where the images with pseudo-labels are marked as $T^{'} = \{(I_1^T, L_1^E), (I_2^T, L_2^E), \cdots, (I_M^T, L_M^E)\}$;

The dynamic optimization stage

for $iter = 1 : epochs$ **do**

- Calculating the dynamic weights α, β based on the current *iter*;
- Learning DSSN based on the objective function in (1) with the updated weights α, β;

end for

4. Experimental results and discussion

In this section, we first describe the details of experiments, including datasets and evaluation metrics. Then, we analyze the confidence threshold of the pseudo-label. In Section 4.3, we conduct experiments to verify the adaptation of the DualGAN model. In Section 4.4, we conduct ablation experiments to verify the effectiveness of each constraint

module. Finally, we perform comparison experiments with existing cross-domain semantic segmentation algorithms to demonstrate the effectiveness of the proposed framework.

4.1. Experimental settings and evaluation metrics

In this subsection, we first introduce the two cross-domain RS image semantic segmentation task settings and then describe the metrics used to measure the performance of the algorithms.

4.1.1. Task settings

To fully verify the effectiveness of cross-domain RS image semantic segmentation, we conduct experiments by using Potsdam and Vaihingen datasets, which belong to the ISPRS 2D semantic segmentation benchmark dataset (Gerke, 2014). All images in both datasets are provided with their semantic labels, including six classes of ground objects: clutter/background, impervious surfaces, car, tree, low vegetation and building. For the target domain dataset, we will not use their labels in the training process. The Potsdam dataset contains 3 different imaging modes: IR-R-G: 3 channels (IR-R-G), R-G-B: 3 channels (R-G-B), RGBIR: 4 channels (R-G-B-IR). We use the first two kinds. The Vaihingen dataset contains only one imaging mode: IR-R-G: 3 channels (IR-R-G). The Potsdam dataset contains 38 very high resolution True Orthophotos (TOP) with a fixed size of 6000 × 6000. The Vaihingen dataset includes 33 very high resolution True Orthophotos (TOP) with 2000 × 2000 pixels.

In detail, we provide two cross-domain experimental settings: (I) variation in geographic location, shown as Fig. 3(a). The Potsdam IR-R-G dataset serves as he source domain, and the Vaihingen IR-R-G dataset serves as the target domain. To increase the computational efficiency, we crop the Potsdam IR-R-G dataset and their corresponding labels into the size of 512 × 512 with both horizontal and vertical strides of 512 pixels, and we obtain nearly 4000 images. All these images participate in the WTIC module. For the Vaihingen IR-R-G dataset, we crop the images to a size of 512 × 512 with both horizontal and vertical strides of 256 pixels and obtain nearly 1700 images. All the images are used in the WPLC and WRCC modules in the training phase. In addition, among the 1700 images, 500 images (cropped by original images numbered 2, 5, 7, 8, 13, 20, 22, 24) are used for validation to select our optimal model, and nearly 1200 images (cropped by the remaining images) are used for testing to evaluate the performance of the algorithms. (II) Variation in both geographic location and imaging mode, shown as Fig. 3(b). More precisely, the Potsdam R-G-B dataset serves as the source domain, and the Vaihingen IR-R-G dataset serves as the target domain. The rest of the settings are similar to the experiment (I).

This work is implemented by Pytorch and trained on a single Nvidia TITAN RTX GPU with 24 GB RAM. As an optimizer for the training, we used stochastic gradient descent (SGD) optimizer with the initial learning rate set to 0.0005 and momentum is set to 0.9 and weight decay is 5×10^{-4}. In our implementation, the batch size and epoch are set to 4 and 10, respectively.

4.1.2. Evaluation metrics

In this paper, we use the *F*1_*Score* and *IoU* to evaluate the performance of cross-domain RS image semantic segmentation.

Specifically, *F*1_*Score* can be defined by:

$$F1\_Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{5}$$

Moreover, we also used the intersection over union (*IoU*) to measure the efficiency of the segmentation. Since we have 6 different classes, *IoU* is calculated for every class separately. Then, the mean *IoU* of all classes is calculated. Eq. (6) represents how to calculate *IoU* for two different data samples.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{6}$$

where A is the set of ground-truth pixels and B is the set of predicted pixels. ∩ and ∪ denote intersection and union, respectively. $|\cdot|$ denotes calculating the number of pixels in the set.

4.2. Sensitivity analysis of the confidence threshold

In WPLC, the threshold τ_{pse} guides the generation of pseudo-labels. To analyze the sensitivity of τ_{pse}, we evaluate the performance of the proposed framework under different τ_{pse} on both the cross-domain semantic segmentation task from Potsdam IR-R-G to Vaihingen IR-R-G and the task from Potsdam R-G-B to Vaihingen IR-R-G. To reduce the number of calculations, one single experiment is used to conduct the parameter analysis.

For the pseudo-label threshold τ_{pse}, generally, a high threshold τ_{pse} fails to yield good performance since it will ignore more regions in the image than a low threshold. Thus, fewer pixels can participate in the WPLC procedure, resulting in poor performance. To obtain the best threshold τ_{pse}, we carried out a complete experimental process with various thresholds on the validation set. The results of different thresholds τ_{pse} on the two tasks are shown in Table 2 and Table 3. For the cross-domain semantic segmentation task from Potsdam IR-R-G to Vaihingen IR-R-G, it can be observed that the performance is best with a mean *F*1_*Score* of 60.76% and an m*IoU* of 44.53% on the validation set under the thresholds τ_{pse} set as 0.4. For the other task, we can deduce that the best performance is with a mean *F*1_*Score* of 52.74% and an m*IoU* of 39.19% on the validation set under the thresholds τ_{pse} set as 0.5.

4.3. Adaptation verification of the recommended DualGAN model

This section discusses the adaptation of the recommended DualGAN module. As is well known, all DiscoGAN, CycleGAN and DualGAN are qualified to conduct general-purpose image-to-image translations without requiring a joint representation to bridge the two image domains. Thus, all three algorithms are adopted to perform the image transfer on both of the argued two tasks, including from Potsdam IR-R-G to Vaihingen IR-R-G and from Potsdam R-G-B to Vaihingen IR-R-G.

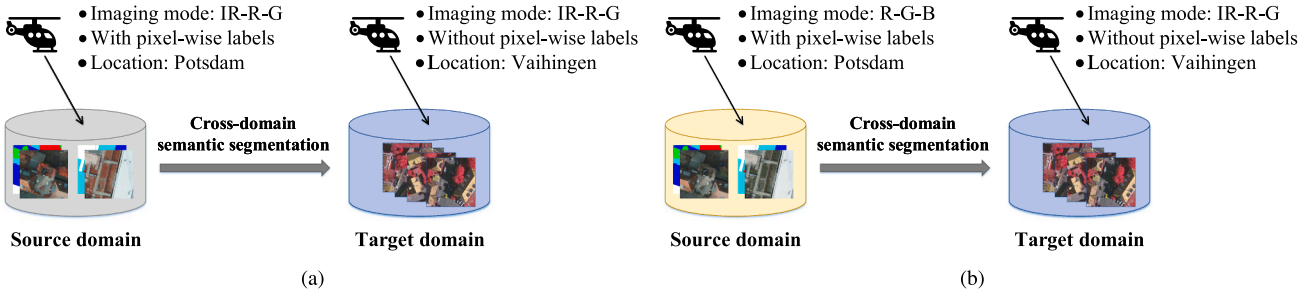

Fig. 3. Two different cross-domain semantic segmentation tasks. (a) The cross-domain task from Potsdam IR-R-G to Vaihingen IR-R-G. (b) The cross-domain task from Potsdam R-G-B to Vaihingen IR-R-G.

Table 2
Parameter analysis of the pseudo-label threshold τ_{pse} on the cross-domain semantic segmentation task from Potsdam IR-R-G to Vaihingen IR-R-G.

Threshold		Clutter/background	Impervious surfaces	Car	Tree	Low vegetation	Building	Overall
$\tau_{pse} = 0.1$	*F1_Score*	32.36	62.83	50.19	70.01	56.51	75.61	57.92
	IoU	19.30	45.80	33.51	53.85	39.39	60.78	42.11
$\tau_{pse} = 0.2$	*F1_Score*	34.75	58.60	**50.51**	73.83	59.53	74.80	58.67
	IoU	21.03	41.44	**33.78**	58.06	42.38	59.75	42.74
$\tau_{pse} = 0.3$	*F1_Score*	35.66	**65.83**	49.67	70.71	**60.92**	**78.01**	60.13
	IoU	21.70	**49.07**	33.04	54.69	43.08	**63.95**	44.26
$\tau_{pse} = 0.4$	*F1_Score*	**51.65**	59.49	44.54	**73.87**	60.29	74.81	**60.78**
	IoU	34.82	42.34	28.65	**58.45**	**43.15**	59.75	**44.53**
$\tau_{pse} = 0.5$	*F1_Score*	50.55	54.64	49.10	72.48	57.78	73.14	59.62
	IoU	**37.50**	37.59	32.54	56.84	40.62	57.66	43.79

Table 3
Parameter analysis of the pseudo-label threshold τ_{pse} on the cross-domain semantic segmentation task from Potsdam R-G-B to Vaihingen IR-R-G.

Threshold		Clutter/background	Impervious surfaces	Car	Tree	Low vegetation	Building	Overall
$\tau_{pse} = 0.4$	*F1_Score*	**1.49**	55.54	57.00	69.87	45.40	75.77	50.85
	IoU	**0.75**	38.45	39.86	53.69	29.36	61.00	37.19
$\tau_{pse} = 0.5$	*F1_Score*	1.24	59.97	**57.86**	72.73	**47.61**	**77.02**	**52.74**
	IoU	0.62	42.83	40.70	57.15	**31.24**	**62.63**	**39.20**
$\tau_{pse} = 0.6$	*F1_Score*	1.25	56.57	56.17	68.42	47.44	76.50	51.06
	IoU	0.63	39.44	39.05	52.00	31.10	31.94	32.36
$\tau_{pse} = 0.7$	*F1_Score*	1.04	**62.69**	53.30	**73.18**	46.83	75.46	52.08
	IoU	0.52	**45.66**	36.34	**57.70**	30.57	60.59	38.56
$\tau_{pse} = 0.8$	*F1_Score*	1.39	56.64	57.31	70.99	43.55	76.56	51.07
	IoU	0.70	39.51	**41.15**	55.03	27.84	62.02	37.71

For the Potsdam IR-R-G to Vaihingen IR-R-G task, the transfer results are shown in Fig. 4, where Fig. 4(e) shows the target images that are used for comparison. The transferred images generated by the DualGAN are sharper and more realistic. To quantitatively verify the adaptation of DualGAN under our proposed framework, we evaluate the performance of the proposed framework with different GAN models (i.e., DualGAN, CycleGAN and DiscoGAN) where τ_{pse} is set as 0.4. To reduce computing consumption, one single experiment is adopted for the module analysis. The quantitative results are summarized in Table 4.

For the Potsdam R-G-B to Vaihingen IR-R-G task, Fig. 5 shows the transfer results. Table 5 presents the quantitative results via the proposed framework under different GAN models where τ_{pse} is set as 0.5. As shown in Table 5, the *mIoU* of DualGAN is 39.19% and the *mIoU* of CycleGAN is 37.11%. Thereby, DualGAN can outperform CycleGAN by 2%. In some small categories such as clutter/background, DualGAN performs better in task I (i.e., variation in geographic location) but CycleGAN performs better in task II (i.e., variation in both geographic location and imaging mode), that's because a little difference in the transferred images will cause a major variety in the evaluation result of small categories. Besides, GAN is a generative model which is difficult to ensure the transfer result of each class, this phenomenon is still an open problem of GAN based models. However, one thing is certain that DualGAN outperforms on the overall evaluation metrics under our proposed framework. Based on the above facts, it is not hard to draw a conclusion that DualGAN is more suitable for the cross-domain RS image semantic segmentation task under our proposed architecture.

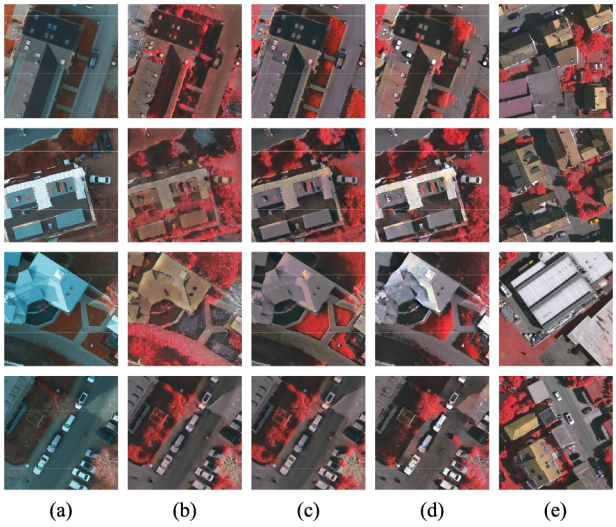

Fig. 4. The transferred images of Potsdam IR-R-G via different GAN models. (a) Images from the source domain. (b) Transferred images using DiscoGAN. (c) Transferred images using CycleGAN. (d) Transferred images using DualGAN. (e) Images from the target domain.

4.4. Ablation study

This part verifies the effectiveness of the WPLC and WRCC modules. To reduce a large amount of calculation consumption, a single experiment is adopted. In detail, four types of settings ($\alpha = 0, \beta \neq 0; \alpha \neq 0, \beta = 0; \alpha = 0, \beta = 0; \alpha \neq 0, \beta \neq 0$) are given on both cross-domain experimental settings. For example, the setting of $\alpha = 0, \beta \neq 0$ stands for the combination of weakly-supervised transfer invariant constraint and weakly-supervised rotation consistency constraint, which means only the usage of the WTIC and WRCC module to train the DSSN. The other three settings are similar to this.

As shown in Table 6 and Table 7, the WPLC and WRCC modules help improve the performance of DSSN. Especially, when these three constraints WTIC, WPLC and WRCC work together, they can benefit from each other and gain a great performance.

4.5. Comparison with the state-of-the-art methods

In this subsection, we perform comparison experiments with existing cross-domain semantic segmentation algorithms under two cross-domain experimental settings: (I) variation in geographic location and

Table 4
The quantitative results (%) of different GAN models on the cross-domain semantic segmentation task from Potsdam IR-R-G to Vaihingen IR-R-G.

		Clutter /background	Impervious surfaces	Car	Tree	Low vegetation	Building	Overall
DiscoGAN	*F1_Score*	3.55	62.63	42.06	66.27	41.26	72.81	48.10
	IoU	1.81	45.59	26.63	49.56	25.99	57.24	34.47
CycleGAN	*F1_Score*	8.84	**66.61**	**56.28**	72.97	55.22	**79.84**	56.63
	IoU	4.63	**49.94**	**39.16**	57.45	38.14	**66.44**	42.63
DualGAN	*F1_Score*	**51.65**	59.49	44.54	**73.78**	**60.29**	74.81	**60.76**
	IoU	**34.82**	42.34	28.65	**58.45**	**43.15**	59.75	**44.53**

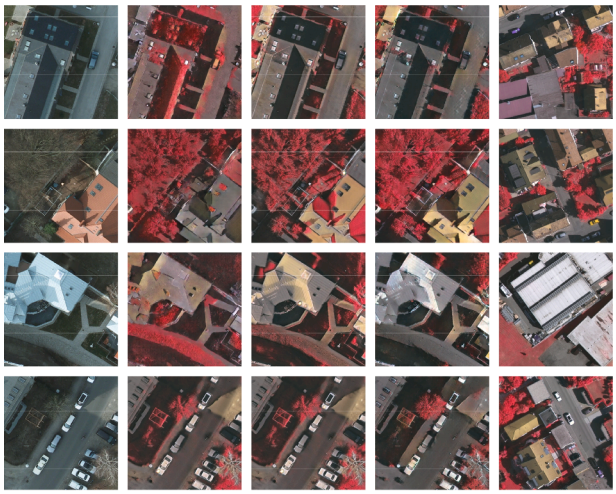

Fig. 5. The transferred images of Potsdam R-G-B via different GAN models. (a) Images from the source domain. (b) Transferred images using DiscoGAN. (c) Transferred images using CycleGAN. (d) Transferred images using DualGAN. (e) Images from the target domain.

(II) variation in both geographic location and imaging mode. For a fair comparison, all the algorithms selected the best model by the Vaihingen validation set and evaluated on the Vaihingen test set. To improve the feasibility and repeatability, we conducted three experiments for each method, and the results are presented as the *average* ± *standard deviation*.

4.5.1. *Experimental results under the variation in geographic location*

To confirm the effectiveness of our proposed method on domain shift mainly caused by region variation, we use the Potsdam IR-R-G dataset as the source domain and the Vaihingen IR-R-G dataset as the target domain. The qualitative results are displayed in Fig. 6, where the models without adaptation suffer from a serious domain shift problem. They usually appear as noisy segmentation or wrong context. After adaptation, this problem has been alleviated to a large extent. Compared with other methods, our method yields better segmentation results.

The quantitative results are listed in Table 8, where the UDA in (Bilel et al., 2019) methods are based on the BiSeNet (Yu et al., 2018b) framework and others are based on the DeepLab framework. DeepLab only is a baseline model that only utilizes the source domain for training and directly tests the obtained model on the target domain, which should clearly exhibit the problem of domain shift. The mean *F*1_*Score* and *IoU* of the baseline are 44.40% and 31.04%, respectively. After being processed by different domain adaptation methods, the segmentation performance is improved to some degree. Our best model achieves *F*1_*Score* and *IoU* as high as 61.43% and 45.38%, thereby improving the baseline by nearly 17% and 14%, respectively. Compared with the other methods, our model still has higher performance, which demonstrates that the proposed method is more beneficial to eliminate the influence of domain shift.

In addition, we use the Vaihingen IR-R-G dataset as the source domain and the Potsdam IR-R-G dataset as the target domain and then conduct the experiments with τ_{pse} set as 0.3. The visualization results and quantitative results are shown in Fig. 7 and Table 9, respectively. The results further prove that our proposed method can work well with RS image cross-domain semantic segmentation.

Table 5
The quantitative results (%) of different GAN models on the cross-domain semantic segmentation task from Potsdam R-G-B to Vaihingen IR-R-G.

		Clutter /background	Impervious surfaces	Car	Tree	Low vegetation	Building	Overall
DiscoGAN	*F1_Score*	0.67	**60.75**	46.65	55.42	37.3	72.03	45.47
	IoU	0.34	**43.63**	30.42	38.33	22.92	56.29	31.99
CycleGAN	*F1_Score*	**24.71**	52.52	53.37	65.96	38.56	**77.64**	52.13
	IoU	**14.09**	35.61	36.40	49.21	23.88	**63.45**	37.11
DualGAN	*F1_Score*	1.24	59.97	**57.86**	**72.73**	**47.61**	77.02	**52.74**
	IoU	0.62	42.83	**40.70**	**57.15**	**31.24**	62.63	**39.19**

Table 6
Quantifying the effectiveness of the WPLC and WRCC on the cross-domain semantic segmentation task from Potsdam IR-R-G to Vaihingen IR-R-G.

		Clutter /background	Impervious surfaces	Car	Tree	Low vegetation	Building	Overall
$\alpha = 0, \beta \neq 0$	*F1_Score*	1.26	**60.94**	**56.63**	72.73	49.62	**76.09**	52.88
	IoU	0.63	**43.82**	**39.50**	57.14	33.00	**61.41**	39.25
$\alpha \neq 0, \beta = 0$	*F1_Score*	1.59	59.40	39.25	69.51	52.44	73.59	49.29
	IoU	0.80	42.24	24.41	53.26	35.54	58.21	35.75
$\alpha = 0, \beta = 0$	*F1_Score*	18.94	57.45	47.27	72.57	52.95	74.34	53.92
	IoU	10.46	40.30	30.95	56.95	36.00	59.16	38.97
$\alpha \neq 0, \beta \neq 0$	*F1_Score*	**51.65**	59.49	44.54	**73.78**	**60.29**	74.81	**60.76**
	IoU	**34.82**	42.34	28.65	**58.45**	**43.15**	59.75	**44.53**

Table 7
Quantifying the effectiveness of the WPLC and WRCC on the cross-domain semantic segmentation task from Potsdam R-G-B to Vaihingen IR-R-G.

		Clutter /background	Impervious surfaces	Car	Tree	Low vegetation	Building	Overall
$\alpha = 0, \beta \neq 0$	*F1_Score*	1.39	57.47	56.31	67.31	47.46	75.72	50.94
	IoU	0.70	40.32	39.19	50.73	31.11	60.93	37.16
$\alpha \neq 0, \beta = 0$	*F1_Score*	**8.99**	56.80	48.24	69.05	40.96	75.22	49.88
	IoU	**4.71**	39.67	31.79	52.73	25.75	60.28	35.82
$\alpha = 0, \beta = 0$	*F1_Score*	1.37	52.65	53.42	67.43	42.96	74.08	48.65
	IoU	0.69	35.73	36.45	50.86	27.36	58.83	34.99
$\alpha \neq 0, \beta \neq 0$	*F1_Score*	1.24	**59.97**	**57.86**	**72.73**	**47.61**	**77.02**	**52.74**
	IoU	0.62	**42.83**	**40.70**	**57.15**	**31.24**	**62.63**	**39.19**

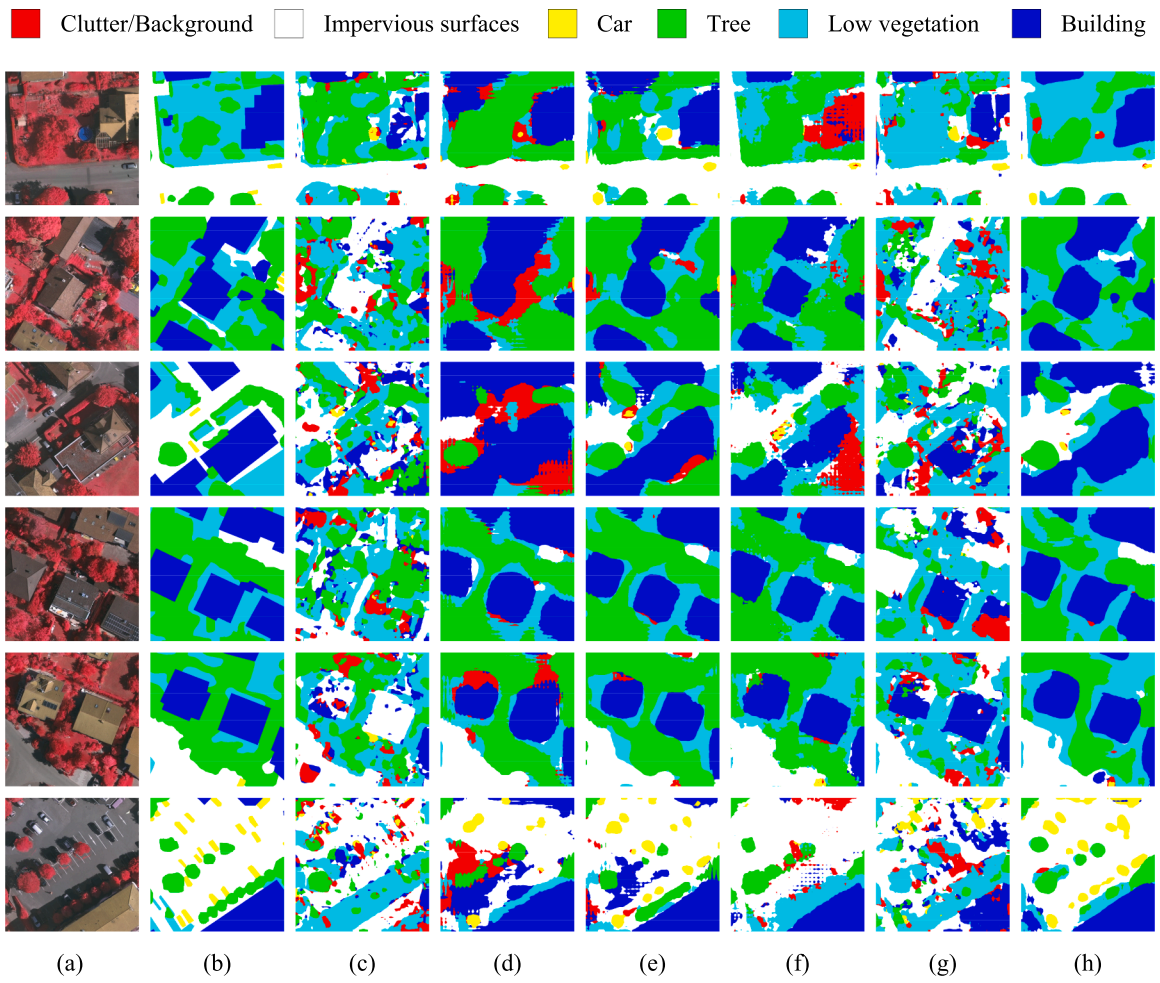

Fig. 6. The qualitative results of the cross-domain semantic segmentation task from Potsdam IR-R-G to Vaihingen IR-R-G. (a) Target images. (b) Ground truth. (c) BiseNet only. (d) DeepLab v3+ only (e) SEANet (Xu et al., 2019). (f) AdaptSegNet (Tsai et al., 2018). (g) UDA in (Bilel et al., 2019). (h) Ours.

4.5.2. Experimental results under the variation in both geographic location and imaging mode

Similar to the first experiment, the Potsdam R-G-B dataset serves as the source domain, and the Vaihingen IR-R-G dataset serves as the target domain to evaluate the effectiveness of our method on domain shift caused by variations in both the geographic location and imaging mode. The qualitative results are shown in Fig. 8. Similarly, the baseline model suffers from a severe domain shift problem. After adaptation, the segmentation performance for each category has been significantly improved. Compared with other methods, our method can preserve more details on edges, which demonstrates the superiority of the proposed method from a qualitative perspective.

Table 10 describes the performance of our method and other competitive methods. The source-only method achieves very poor performance, with a mean *IoU* and *F*1_*Score* of 23.72% and 34.50%, respectively. For the category of clutter and low vegetation, the *IoU* drops to a very low degree. By observing the original data of the two datasets, we find that the saturation between them is also very different, and the color of the tree is green in Potsdam R-G-B and red in Vaihingen IR-R-G due to the different imaging modes used in these two datasets. Unsurprisingly, those differences directly lead to a sharp drop in performance. The performance of our method is superior to other methods, and its mean *IoU* and mean *F*1_*Score* achieve 39.93% and 54.82%, respectively. The above experimental results demonstrate the superiority and effectiveness of our method.

Furthermore, we take the Vaihingen IR-R-G dataset as the source domain and the Potsdam R-G-B dataset as the target domain and then conduct the experiments with τ_{pse} set as 0.3. The visualization results

Table 8
The quantitative results (%) of the cross-domain semantic segmentation task from Potsdam IR-R-G to Vaihingen IR-R-G.

		Clutter /background	Impervious surfaces	Car	Tree	Low vegetation	Building	Overall
BiSeNet only	*F1_Score*	3.28 ± 0.23	55.55 ± 0.77	17.2 ± 2.18	62.57 ± 1.72	37.44 ± 0.17	43.93 ± 0.72	36.65 ± 0.01
	IoU	1.67 ± 0.12	38.46 ± 0.74	9.42 ± 1.31	45.54 ± 1.82	23.04 ± 0.13	28.15 ± 0.59	24.38 ± 0.15
DeepLab v3 + only	*F1_Score*	10.79 ± 2.43	52.73 ± 3.03	33.70 ± 1.36	70.92 ± 1.24	30.26 ± 5.34	68.06 ± 0.64	44.40 ± 1.13
	IoU	5.71 ± 1.36	35.84 ± 2.79	20.27 ± 0.99	54.95 ± 1.50	17.88 ± 3.71	51.59 ± 0.73	31.04 ± 0.67
SEANet (Xu et al., 2019)	*F1_Score*	20.00 ± 0.19	62.59 ± 1.01	49.29 ± 0.47	66.83 ± 3.68	37.45 ± 1.48	73.27 ± 2.91	51.60 ± 1.63
	IoU	11.11 ± 0.23	45.57 ± 2.14	32.71 ± 0.83	50.42 ± 8.31	23.06 ± 2.25	57.99 ± 7.25	36.81 ± 3.50
AdaptSegNet (Tsai et al., 2018)	*F1_Score*	8.76 ± 2.80	**70.39 ± 4.12**	11.99 ± 3.44	68.96 ± 2.12	44.91 ± 3.42	**77.40 ± 0.28**	47.05 ± 0.64
	IoU	4.60 ± 1.52	**54.39 ± 4.91**	6.40 ± 1.94	52.65 ± 2.47	28.98 ± 2.84	**63.14 ± 0.37**	35.02 ± 1.19
UDA in (Bilel et al., 2019)	*F1_Score*	4.15 ± 0.12	57.02 ± 0.66	15.15 ± 0.71	41.97 ± 0.76	41.94 ± 0.21	58.10 ± 3.10	36.40 ± 0.14
	IoU	2.12 ± 0.06	39.88 ± 0.64	8.20 ± 0.42	26.56 ± 0.61	26.53 ± 0.17	40.97 ± 3.08	24.04 ± 0.21
Ours	*F1_Score*	**45.65 ± 4.59**	66.13 ± 1.22	**51.09 ± 2.89**	**73.14 ± 0.83**	**55.97 ± 2.03**	76.77 ± 0.68	**61.43 ± 1.03**
	IoU	**29.66 ± 3.82**	49.41 ± 1.37	**34.34 ± 2.59**	**57.66 ± 1.03**	**38.87 ± 1.97**	62.30 ± 0.90	**45.38 ± 0.98**

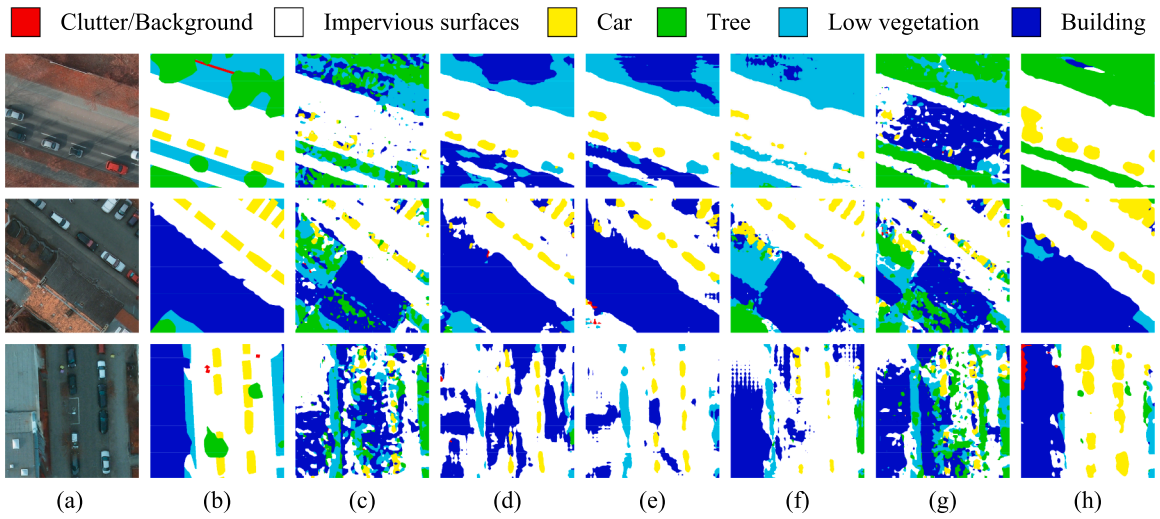

Fig. 7. The qualitative results of the cross-domain semantic segmentation task from Vaihingen IR-R-G to Potsdam IR-R-G. (a) Target images. (b) Ground truth. (c) BiseNet only. (d) DeepLab v3+ only (e) SEANet (Xu et al., 2019). (f) AdaptSegNet (Tsai et al., 2018). (g) UDA in (Bilel et al., 2019). (h) Ours.

Table 9
The quantitative results (%) of the cross-domain semantic segmentation task from Vaihingen IR-R-G to Potsdam IR-R-G.

		Clutter /background	Impervious surfaces	Car	Tree	Low vegetation	Building	Overall
BiSeNet only	*F1_Score*	**44.97 ± 0.80**	36.99 ± 2.52	1.36 ± 0.51	**58.71 ± 1.53**	41.42 ± 0.96	34.61 ± 0.90	36.34 ± 0.13
	IoU	**29.01 ± 0.67**	22.70 ± 1.88	0.69 ± 0.26	**41.56 ± 1.54**	26.12 ± 0.76	20.93 ± 0.66	23.5 ± 0.12
DeepLab v3+ only	*F1_Score*	16.86 ± 7.52	65.93 ± 0.54	55.6 ± 1.62	14.24 ± 0.52	45.34 ± 0.37	53.97 ± 0.93	41.99 ± 1.92
	IoU	9.3 ± 4.48	49.18 ± 0.60	38.51 ± 1.56	7.67 ± 0.30	29.32 ± 0.31	36.96 ± 0.88	28.49 ± 1.36
SEANet (Xu et al., 2019)	F1_Score	23.23 ± 2.90	**67.79 ± 3.20**	59.56 ± 2.56	9.79 ± 1.01	45.13 ± 1.23	56.73 ± 1.78	43.70 ± 0.95
	IoU	13.16 ± 1.84	**51.33 ± 3.62**	42.44 ± 2.61	5.15 ± 0.56	29.14 ± 1.05	39.61 ± 1.75	30.14 ± 1.07
AdaptSegNet (Tsai et al., 2018)	*F1_Score*	15.33 ± 5.95	64.64 ± 1.24	58.11 ± 0.93	36.79 ± 4.53	**61.50 ± 2.59**	63.41 ± 2.60	49.96 ± 0.52
	IoU	8.36 ± 3.49	49.55 ± 1.58	40.95 ± 0.92	22.59 ± 3.40	34.43 ± 2.69	48.01 ± 3.24	33.98 ± 0.26
UDA in (Bilel et al., 2019)	*F1_Score*	43.43 ± 1.76	20.71 ± 1.61	1.82 ± 1.13	31.93 ± 0.83	31.08 ± 1.10	24.24 ± 0.51	25.54 ± 0.12
	IoU	27.39 ± 0.37	18.66 ± 2.28	0.59 ± 0.09	32.06 ± 2.20	19.72 ± 0.71	27.40 ± 0.18	20.97 ± 0.12
Ours	*F1_Score*	20.56 ± 3.94	67.53 ± 2.59	**65.31 ± 0.67**	51.82 ± 1.02	53.48 ± 0.66	**69.59 ± 1.33**	**54.71 ± 1.26**
	IoU	11.48 ± 2.45	51.01 ± 2.96	**48.49 ± 0.74**	34.98 ± 0.93	**36.50 ± 0.62**	**53.37 ± 1.56**	**39.30 ± 1.09**

and quantitative results are shown in Fig. 9 and Table 11, respectively. The results also show the superiority of our proposed method.

In summary, our proposed method has good performance in dealing with both domain shifts mainly caused by region and imaging mode variations. The proposed method shows strong robustness and great generalization capability.

5. Conclusion

Due to its excellent feature extraction capability, DSSN has been widely used in RS image semantic segmentation and has achieved great success. However, the superiority of DSSN highly depends on the large quantity of labeled training data, and the test data and training data are

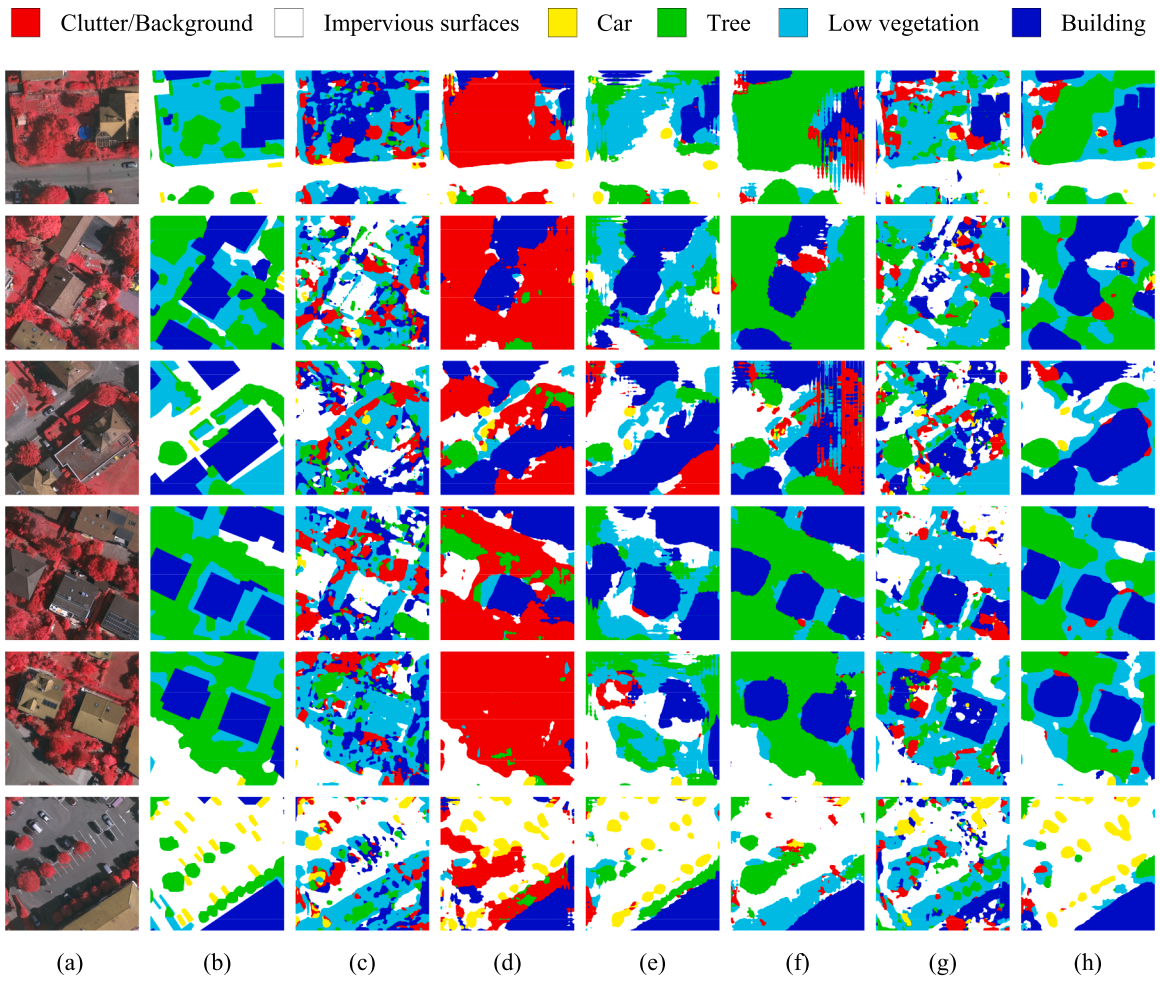

Fig. 8. The qualitative results of the cross-domain semantic segmentation task from Potsdam R-G-B to Vaihingen IR-R-G. (a) Target images. (b) Ground truth. (c) BiseNet only. (d) DeepLab v3+ only. (e) SEANet (Xu et al., 2019). (f) AdaptSegNet (Tsai et al., 2018). (g) UDA in (Bilel et al., 2019). (h) Ours.

Table 10
The quantitative results (%) of the cross-domain semantic segmentation task from Potsdam R-G-B to Vaihingen IR-R-G.

		Clutter /background	Impervious surfaces	Car	Tree	Low vegetation	Building	Overall
BiSeNet only	*F1_Score*	1.75 ± 0.26	53.57 ± 0.08	9.17 ± 0.38	14.37 ± 0.18	31.55 ± 0.90	29.49 ± 1.61	23.30 ± 0.42
	IoU	0.88 ± 0.13	36.58 ± 0.07	4.81 ± 0.21	7.74 ± 0.10	18.73 ± 0.64	17.30 ± 1.11	14.34 ± 0.28
DeepLab v3 + only	*F1_Score*	2.03 ± 0.38	63.37 ± 1.75	42.93 ± 0.14	23.73 ± 7.88	8.82 ± 0.08	66.12 ± 0.64	34.50 ± 1.70
	IoU	1.03 ± 0.19	**46.39 ± 1.87**	27.33 ± 0.11	13.58 ± 5.07	4.61 ± 0.04	49.39 ± 0.73	23.72 ± 1.23
SEANet (Xu et al., 2019)	*F1_Score*	12.92 ± 9.55	59.00 ± 0.32	48.16 ± 0.29	35.82 ± 4.06	34.48 ± 0.27	72.44 ± 1.59	43.80 ± 0.57
	IoU	**7.05 ± 2.47**	41.84 ± 0.31	31.72 ± 0.25	21.86 ± 3.02	20.83 ± 0.20	56.80 ± 1.96	30.02 ± 0.02
AdaptSegNet (Tsai et al., 2018)	*F1_Score*	5.81 ± 0.39	67.77 ± 1.02	18.54 ± 4.31	68.02 ± 0.95	22.61 ± 0.66	75.55 ± 1.37	43.05 ± 0.92
	IoU	2.99 ± 0.21	51.26 ± 1.17	10.25 ± 2.61	51.54 ± 1.10	12.75 ± 0.42	60.72 ± 1.77	31.58 ± 0.55
UDA in (Bilel et al., 2019)	*F1_Score*	4.43 ± 0.75	51.35 ± 0.20	18.00 ± 2.88	39.89 ± 3.08	36.83 ± 1.03	57.79 ± 0.68	34.70 ± 0.85
	IoU	2.27 ± 0.39	34.55 ± 0.18	9.90 ± 1.74	24.94 ± 2.40	22.58 ± 0.77	40.65 ± 0.67	22.48 ± 0.64
Ours	*F1_Score*	**13.88 ± 1.72**	**61.33 ± 1.65**	**57.88 ± 0.15**	**70.66 ± 1.90**	**42.17 ± 4.99**	**83.00 ± 6.43**	**54.82 ± 1.29**
	IoU	3.94 ± 0.44	46.19 ± 1.05	**40.31 ± 0.43**	**55.82 ± 0.62**	**27.85 ± 2.49**	**65.44 ± 1.27**	**39.93 ± 0.13**

distributed identically. In practical applications, it is difficult to satisfy these conditions. Thus, how to fully mine the invariant semantic features of RS image data from different domains and efficiently perform cross-domain RS image semantic segmentation attracts much attention. Driven by this intensive demand, this paper proposes a novel objective function with multiple weakly-supervised constraints to learn DSSN for cross-domain RS image semantic segmentation where multiple weakly-supervised constraints include the WTIC, WPLC and WRCC. Different from methods based on domain adaptation to learn the invariant features between different domains, the proposed method directly learns to map the images from the source to the target and keeps the content of the generated images similar to the original. More specifically, DualGAN is recommended for conducting unsupervised style transfer between the source and target domains to carry out WTIC. To balance these multiple constraints, the optimization procedure is dynamic, which can efficiently avoid the DSSN falling into a degeneration situation. After

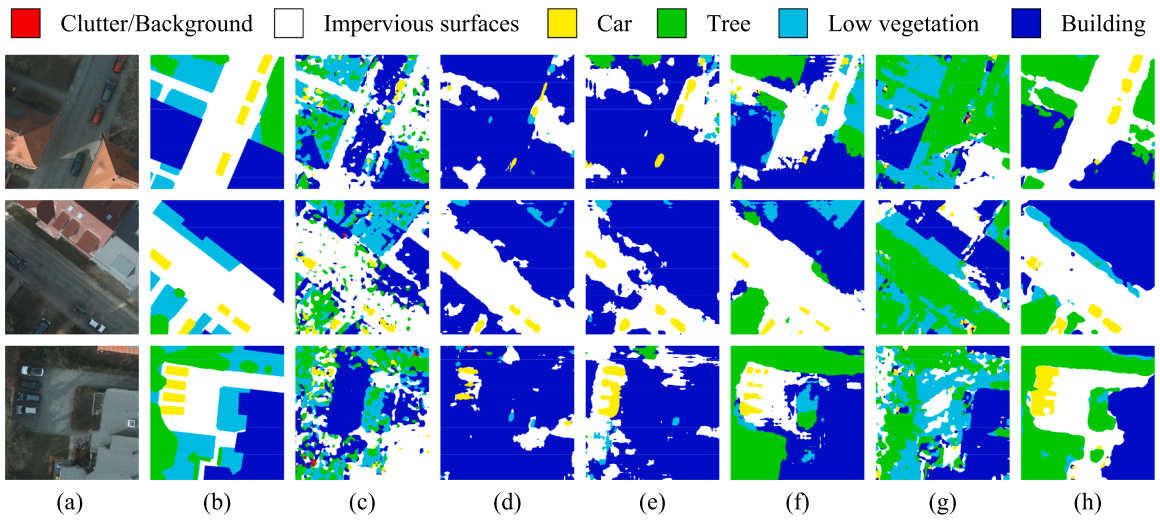

Fig. 9. The qualitative results of the cross-domain semantic segmentation task from Vaihingen IR-R-G to Potsdam R-G-B. (a) Target images. (b) Ground truth. (c) BiseNet only. (d) DeepLab v3+ only (e) SEANet (Xu et al., 2019). (f) AdaptSegNet (Tsai et al., 2018). (g) UDA in (Bilel et al., 2019). (h) Ours.

Table 11
The quantitative results (%) of the cross-domain semantic segmentation task from Vaihingen IR-R-G to Potsdam R-G-B.

		Clutter /background	Impervious surfaces	Car	Tree	Low vegetation	Building	Overall
BiSeNet only	*F1_Score*	38.26 ± 0.54	30.12 ± 1.63	1.95 ± 1.04	**49.24 ± 0.81**	31.11 ± 1.66	22.43 ± 2.29	28.85 ± 0.17
	IoU	23.66 ± 0.41	17.74 ± 1.12	0.99 ± 0.53	**32.67 ± 0.71**	18.42 ± 1.16	12.64 ± 1.44	17.69 ± 0.10
DeepLab v3 + only	*F1_Score*	13.04 ± 2.72	60.12 ± 0.07	55.08 ± 1.00	1.06 ± 0.98	3.13 ± 0.89	45.05 ± 2.06	29.58 ± 1.29
	IoU	6.99 ± 1.56	42.98 ± 0.07	38.01 ± 0.96	0.53 ± 0.49	1.59 ± 0.47	29.09 ± 1.72	19.86 ± 0.88
SEANet (Xu et al., 2019)	*F1_Score*	8.75 ± 7.32	44.77 ± 3.85	**61.52 ± 2.87**	9.82 ± 0.78	15.88 ± 5.81	53.20 ± 0.12	32.32 ± 1.53
	IoU	4.68 ± 4.10	28.90 ± 3.24	**44.47 ± 3.02**	5.17 ± 0.44	8.70 ± 3.38	36.24 ± 0.11	21.36 ± 1.26
AdaptSegNet (Tsai et al., 2018)	*F1_Score*	11.50 ± 2.76	59.55 ± 7.01	55.95 ± 7.32	45.41 ± 8.24	25.81 ± 12.19	70.31 ± 2.60	44.75 ± 2.62
	IoU	6.11 ± 1.56	37.66 ± 0.06	42.31 ± 2.40	30.71 ± 3.20	15.10 ± 8.05	54.25 ± 3.09	31.02 ± 0.71
UDA in (Bilel et al., 2019)	*F1_Score*	**43.43 ± 1.76**	20.71 ± 1.61	1.82 ± 1.13	31.93 ± 0.83	31.08 ± 1.10	24.24 ± 0.51	25.54 ± 0.12
	IoU	**27.75 ± 1.44**	11.56 ± 1.00	0.92 ± 0.57	19.00 ± 0.59	18.40 ± 0.77	13.79 ± 0.33	15.24 ± 0.15
Ours	*F1_Score*	23.84 ± 3.92	**62.97 ± 1.24**	56.84 ± 1.03	40.97 ± 4.00	**58.87 ± 1.92**	**74.22 ± 0.66**	**52.95 ± 0.86**
	IoU	13.56 ± 2.53	**45.96 ± 1.32**	39.71 ± 1.00	25.80 ± 3.16	**41.73 ± 1.93**	**59.01 ± 0.83**	**37.63 ± 0.43**

training under multiple weakly-supervised constraints, the DSSN can perform well on the target dataset. To verify our proposed approach, we use two cross-domain experimental settings: (I) variation in geographic location and (II) variation in both geographic location and imaging mode. Extensive experiments under two typical cross-domain settings show that our proposed method can obviously outperform the state-of-the-art methods.

In the future, we plan to further improve this method from the following aspects: 1) proposing a suitable method for adaptation at multiple feature layers, 2) introducing more weakly-supervised constraints for target data, and 3) proposing a GAN based style transfer model with the stable image generation ability.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the State Key Program of the National Natural Science Foundation of China under Grant 42030102; the National Natural Science Foundation of China under Grant 41971284; the China Postdoctoral Science Foundation under Grant 2016M590716 and 2017T100581; and the Fundamental Research Funds for the Central Universities under Grant 2042020kf0218.

References

Bilel, B., Bazi, Y., Koubaa, A., Ouni, K., 2019. Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images. Remote Sens. 11, 1369.
Bruzzone, L., Carlin, L., 2006. A multilevel context-based system for classification of very high spatial resolution images. IEEE Trans. Geosci. Remote Sens. 44, 2587–2600.
Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation, in. In: Proceedings of the European conference on computer vision (ECCV), pp. 801–818.
Chen, Y.H., Chen, W.Y., Chen, Y.T., Tsai, B.C., Frank Wang, Y.C., Sun, M., 2017. No more discrimination: Cross city adaptation of road scene segmenters, in. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1992–2001.
Chi, M., Plaza, A., Benediktsson, J.A., Sun, Z., Shen, J., Zhu, Y., 2016. Big data for remote sensing: Challenges and opportunities. Proc. IEEE 104, 2207–2219.
Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding, in. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3213–3223.
Gerke, M., 2014. Use of the stair vision library within the isprs 2d semantic labeling benchmark (vaihingen).
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in neural information processing systems, pp. 2672–2680.
Hertzmann, A., Jacobs, C.E., Oliver, N., Curless, B., Salesin, D.H., 2001. Image analogies. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp. 327–340.

Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T., 2018. Cycada: Cycle-consistent adversarial domain adaptation. In: International conference on machine learning, PMLR, pp. 1989–1998.
Hoffman, J., Wang, D., Yu, F., Darrell, T., 2016. Fcns in the wild: Pixel-level adversarial and constraint-based adaptation. arXiv preprint arXiv:1612.02649.
Hu, W., Huang, Y., Wei, L., Zhang, F., Li, H., 2015. Deep convolutional neural networks for hyperspectral image classification. J. Sens..
Huang, X., Zhu, Z., Li, Y., Wu, B., Yang, M., 2018. Tea garden detection from high-resolution imagery using a scene-based framework. Photogram. Eng. Remote Sens. 84, 723–731.
Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks, in. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125–1134.
Kampffmeyer, M., Salberg, A.B., Jenssen, R., 2016. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks, in. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 1–9.
Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J., 2017. Learning to discover cross-domain relations with generative adversarial networks. arXiv preprint arXiv:1703.05192.
Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 14, 778–782.
Laine, S., Aila, T., 2016. Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242.
Lee, D.H., 2013. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML.
Li, S., Dragicevic, S., Castro, F.A., Sester, M., Winter, S., Coltekin, A., Pettit, C., Jiang, B., Haworth, J., Stein, A., et al., 2016. Geospatial big data handling theory and methods: A review and research challenges. ISPRS J. Photogram. Remote Sens. 115, 119–133.
Li, Y., Chen, W., Zhang, Y., Tao, C., Xiao, R., Tan, Y., 2020a. Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning. Remote Sens. Environ. 250, 112045.
Li, Y., Zhang, Y., Huang, X., Yuille, A.L., 2018. Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images. ISPRS J. Photogramm. Remote Sens. 146, 182–196.
Li, Y., Zhang, Y., Huang, X., Zhu, H., Ma, J., 2017. Large-scale remote sensing image retrieval by deep hashing neural networks. IEEE Trans. Geosci. Remote Sens. 56, 950–965.
Li, Y., Zhang, Y., Zhu, Z., 2020b. Error-tolerant deep learning for remote sensing image scene classification. IEEE Trans. Cybernet.
Long, J., Shelhamer, E., Darrell, T., 2015a. Fully convolutional networks for semantic segmentation, in. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440.
Long, M., Cao, Y., Wang, J., Jordan, M., 2015b. Learning transferable features with deep adaptation networks. In: International conference on machine learning, PMLR, pp. 97–105.
Lyu, Y., Vosselman, G., Xia, G.S., Yilmaz, A., Yang, M.Y., 2020. Uavid: A semantic segmentation dataset for uav imagery. ISPRS J. Photogram. Remote Sens. 165, 108–119.
Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., Jie, W., 2015. Remote sensing big data computing: Challenges and opportunities. Future Gener. Comput. Syst. 51, 47–60.
Makantasis, K., Karantzalos, K., Doulamis, A., Doulamis, N., 2015. Deep supervised learning for hyperspectral data classification through convolutional neural networks. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 4959–4962.
Martin Arjovsky, S., Bottou, L., 2017. Wasserstein generative adversarial networks. In: Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia.
Mi, L., Chen, Z., 2020. Superpixel-enhanced deep neural forest for remote sensing image semantic segmentation. ISPRS J. Photogram. Remote Sens. 159, 140–152.
Odena, A., Dumoulin, V., Olah, C., 2016. Deconvolution and checkerboard artifacts. Distill 1, e3.
Othman, E., Bazi, Y., Melgani, F., Alhichri, H., Alajlan, N., Zuair, M., 2017. Domain adaptation network for cross-scene classification. IEEE Trans. Geosci. Remote Sens. 55, 4441–4456.
Ozdarici-Ok, A., Ok, A.O., Schindler, K., 2015. Mapping of agricultural crops from single high-resolution multispectral images–data-driven smoothing vs. parcel-based smoothing. Remote Sens. 7, 5611–5638.
Shi, H., Chen, L., Bi, F.k., Chen, H., Yu, Y., 2015. Accurate urban area detection in remote sensing images. IEEE Geosci. Remote Sens. Lett. 12, 1948–1952.
Song, S., Yu, H., Miao, Z., Zhang, Q., Lin, Y., Wang, S., 2019. Domain adaptation for convolutional neural networks-based remote sensing scene classification. IEEE Geosci. Remote Sens. Lett. 16, 1324–1328.
Tao, Z., Liu, H., 2017. Image cosegmentation via saliency-guided constraint clustering with cosine similarity. In: AAAI.
Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M., 2018. Learning to adapt structured output space for semantic segmentation, in. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7472–7481.
Tuia, D., Persello, C., Bruzzone, L., 2016. Domain adaptation for the classification of remote sensing data: An overview of recent advances. IEEE Geosci. Remote Sens. Mag. 4, 41–57.
Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T., 2014. Deep domain confusion: Maximizing for domain invariance. arXiv preprint arXiv:1412.3474.
Volpi, M., Tuia, D., 2018. Deep multi-task learning for a geographically-regularized semantic segmentation of aerial images. ISPRS J. Photogram. Remote Sens. 144, 48–60.
Xu, Y., Du, B., Zhang, L., Zhang, Q., Wang, G., Zhang, L., 2019. Self-ensembling attention networks: Addressing domain shift for semantic segmentation, in. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 5581–5588.
Yan, L., Zhu, R., Mo, N., Liu, Y., 2019. Cross-domain distance metric learning framework with limited target samples for scene classification of aerial images. IEEE Trans. Geosci. Remote Sens. 57, 3840–3857.
Yi, Z., Zhang, H., Tan, P., Gong, M., 2017. Dualgan: Unsupervised dual learning for image-to-image translation, in. In: Proceedings of the IEEE international conference on computer vision, pp. 2849–2857.
Yu, B., Yang, L., Chen, F., 2018a. Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 11, 3252–3261.
Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., Sang, N., 2018b. Bisenet: Bilateral segmentation network for real-time semantic segmentation, in. In: Proceedings of the European conference on computer vision (ECCV), pp. 325–341.
Yue, J., Zhao, W., Mao, S., Liu, H., 2015. Spectral–spatial classification of hyperspectral images using deep convolutional neural networks. Remote Sens. Lett. 6, 468–477.
Yue, K., Yang, L., Li, R., Hu, W., Zhang, F., Li, W., 2019. Treeunet: Adaptive tree convolutional neural networks for subdecimeter aerial image segmentation. ISPRS J. Photogram. Remote Sens. 156, 1–13.
Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE Geosci. Remote Sens. Mag. 4, 22–40.
Zhang, Y., Lu, Y., Zhang, D., Shang, L., Wang, D., 2018. Risksens: A multi-view learning approach to identifying risky traffic locations in intelligent transportation systems using social and remote sensing. In: 2018 IEEE International Conference on Big Data (Big Data). IEEE. pp. 1544–1553.
Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881–2890.
Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, in. In: Proceedings of the IEEE international conference on computer vision, pp. 2223–2232.
Zhu, R., Yan, L., Mo, N., Liu, Y., 2019. Semi-supervised center-based discriminative adversarial learning for cross-domain scene-level land-cover classification of aerial images. ISPRS J. Photogram. Remote Sens. 155, 72–89.
Zou, Y., Yu, Z., Vijaya Kumar, B., Wang, J., 2018. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training, in. In: Proceedings of the European conference on computer vision (ECCV), pp. 289–305.