Image-Level and Feature-Level Semantic-Aware Architecture for Cross Domain Semantic Segmentation of High-Resolution Remote Sensing Imagery

Jianhao Miao [1], Xinghua Li [2 *], Xuechen Bai [1]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Hubei, China
[2] School of Remote Sensing and Information Engineering, Wuhan University, Hubei, China

Email: <jianhao_miao, lixinghua5540, baixuechen>@whu.edu.cn

Keywords: Semantic segmentation, Domain adaptation, Self-supervised learning, Optical remote sensing.

Abstract

Semantic segmentation of remote sensing images has attracted considerable attentions. For cross domain semantic segmentation, the images captured at different times inevitably exhibit significant domain gaps, which limits the segmentation performance of unlabelled domain. There are numerous methods to cope with these problems, while style transfer and domain adaptation are effective for domain gaps, the outcomes are still not ideal. Nearly all methods ignore the combination of image-level alignment and feature-level alignment, while few methods consider class-wise constraint to boost the performance. Towards this end, IFSDA, an image-level and feature-level semantic-aware architecture for cross domain semantic segmentation is put forward. In order to acquire sound outcomes, two branches of alignment strategies are realized by self-supervised learning and generative adversarial learning. Besides, a novel semantic discriminator is utilized in image translation process to optimize class-related information, thereby helping to eliminate the intra-class domain gaps between bi-temporal images and optimize the segmentation results effectively. Experiments on ISPRS 2D Semantic Labeling Contest Dataset have shown the superiority of proposed method over other models.

1. Introduction

Semantic segmentation refers to assigning a specific label to images at pixel level, and it is a crucial way for image understanding. As for remote sensing, semantic segmentation is vital to agricultural production, resource management and urban monitoring. With the acquisition of high-resolution remote sensing images (HRSIs), it is possible for us to explore abundant applications. However, the lack of annotation and the complicated radiometric features of remote sensing imagery are challenging for semantic segmentation. Due to varying atmospheric condition, sensor parameters and light condition, remote sensing imagery captured at different conditions inevitably exhibit significant domain gaps. Therefore, when applying a well-trained segmentation model on another dataset, the performance is not ideal. Transfer learning including style transfer (Gatys et al., 2016; Li et al., 2017; Luan et al., 2017; Yoo et al., 2019), domain adaptation (DA) (Chen et al., 2022a; Chen et al., 2022b; Johnson et al., 2016) and domain generalization (DG) (Iizuka et al., 2024; Liang et al., 2024; Niemeijer et al., 2024) is an effective way to deal with this issue.

As for domain adaptive semantic segmentation, the methods can be broadly categorized to feature-level alignment methods and image-level alignment methods. The former realizes implicit transfer learning while the latter conducts image translation. Although there have been many researches to cope with the domain gap of different images and combine feature alignment with specific task successfully, there are some problems.

Firstly, as for the performance of feature-level alignment, neither feature-level self-training nor extracting domain invariant features can get ideal outcomes. For example, DAFormer utilizes a teacher-student model with ClassMix strategy to distillate information from source domain to the target domain (Hoyer et al., 2022). PFST enhances the performance by a local similarity calculation (Zhang et al., 2023). With regard to domain-invariant feature extraction, ST-DASegNet uses two domain disentangled module to extract the domain invariant features of source and target domain (Zhao et al., 2024). These methods is effective for sparse implicit information, and the class-wise information is not well integrated in the network.

Besides, image-level translation for semantic segmentation also faces some challenges. It can be categorized to several aspects. The first one comes sequential adaptation-application model (Li et al., 2020; Li et al., 2023; Luo and Ji, 2022), where the semantic segmentation task is applied after image-level alignment. If the image-level alignment is ideal enough, segmentation task can benefit from it a lot, however, many image-level alignment methods ignore the class-specific information. Although some improvements (Li et al., 2023) take semantic information into consideration, they still struggle to provide ideal radiometric normalization result and bring degradation to the imagery. The other one is progressive translation-based method, they take the intermediate outcomes of image translation or pseudo label as the input, For instance, FPL+ proposes a pseudo label filtering mechanism based on both image-level and pixel-level weighting to obtain robust segmentation result for image translation (Wu et al., 2024). Cai et al. (2022) propose IterDANet for domain adaptive semantic segmentation of remote sensing imagery, it utilizes low-entropy pseudo label and generates reliable supervision. CDTCL establishes a progressive optimization strategy for cross-domain semantic segmentation and image translation, it aims to solve the inefficiency of manual labeling and retraining (Li et al., 2024). However, these methods also suffer from image degradation, the implicit information is inevitably ignored.

Apart from DA methods for semantic segmentation, DG (Iizuka et al., 2024; Liang et al., 2024; Niemeijer et al., 2024) is an effective way to solve the out-of-distribution segmentation. However, compared to DA, DG methods have relatively low accuracy, because it is designed to prioritize generalization to unseen domains and sacrifice the accuracy of specific dataset.

In order to build up a robust model for domain adaptive semantic segmentation and overcome the aforementioned shortcomings, we believe there are several key points and

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-193-2026

correspond optimization strategies to these key points. The first one is integration of semantic segmentation and DA. As many researches suggested, an accurate segmentation result is beneficial for DA, and sound domain adaptation outcomes contribute to transfer learning. Besides, feature-level alignment are believed to be insufficient to some extent, image-level DA can compensate for their shortcomings.

Therefore, a novel image and feature-level self-supervised domain adaptive (IFSDA) segmentation network for HRSIs is proposed, the main contributions can be summarized as follows.

- Two kinds of alignment strategies, including a global image-level alignment and a feature-level alignment are integrated and help for the segmentation.
- Self-supervised teacher-student model and two levels of ClassMix strategy enable robust target pseudo information to supervise the segmentation.
- Class-wise information including source label and target pseudo label are considered to further boost transfer learning.

2. Methodology

2.1 Concept and Model Formulation

Cross domain semantic segmentation utilize images and annotations from one domain to predict the labels of imagery from another domain. The former is source domain and the latter is target domain.

As shown in Figure. 1, $\{x^s, y^s\}$ represents the source domain images x^s along with their corresponding labels y^s, which are treated as the input of classifier in optimization process, $\{x^t\}$ denotes the images from target domain without manual annotation, they serve as the input of model optimization and testing set. The task of cross domain semantic segmentation is to predict labels for target domain images $\{x^t\}$ using only $\{x^s, y^s, x^t\}$.

2.2 Network Architecture

An overall workflow of proposed IFSDA is depicted in Figure. 1. The training process of IFSDA can be divided into two stages, although both stages are optimized simultaneously. As depicted in the left panel of Figure. 1, the first stage is designed to realize image-level alignment, while the second stage is the optimization of teacher-student model and the refinement of label generation where ClassMix strategy and ensemble learning are adopted. As for the test phase, the well-trained student model and image-level segmentation network can cope with the segmentation of target image.

In the Stage 1, image-level alignment is realized by generative adversarial image translation. The image translation can help for segmentation task because it narrows the domain gap. It is believed that if the translation outcomes are similar to the target image enough, the source supervision can greatly help for target segmentation. The process of image translation includes fake image generation, reconstruction image generation and cycle-consistent reconstruction image generation. Source image is fed to the S transfer Net ($\boldsymbol{G_s}$) to get fake source image which has similar radiometric feature to target image, the generated fake data is then fed to T transfer Net ($\boldsymbol{G_t}$) to get reconstruction source image, vice versa for target image. The S transfer Net and T transfer Net are both UNet-like architecture with residual blocks and a convolutional block attention module, with seven downsampling processes and seven upsampling processes.

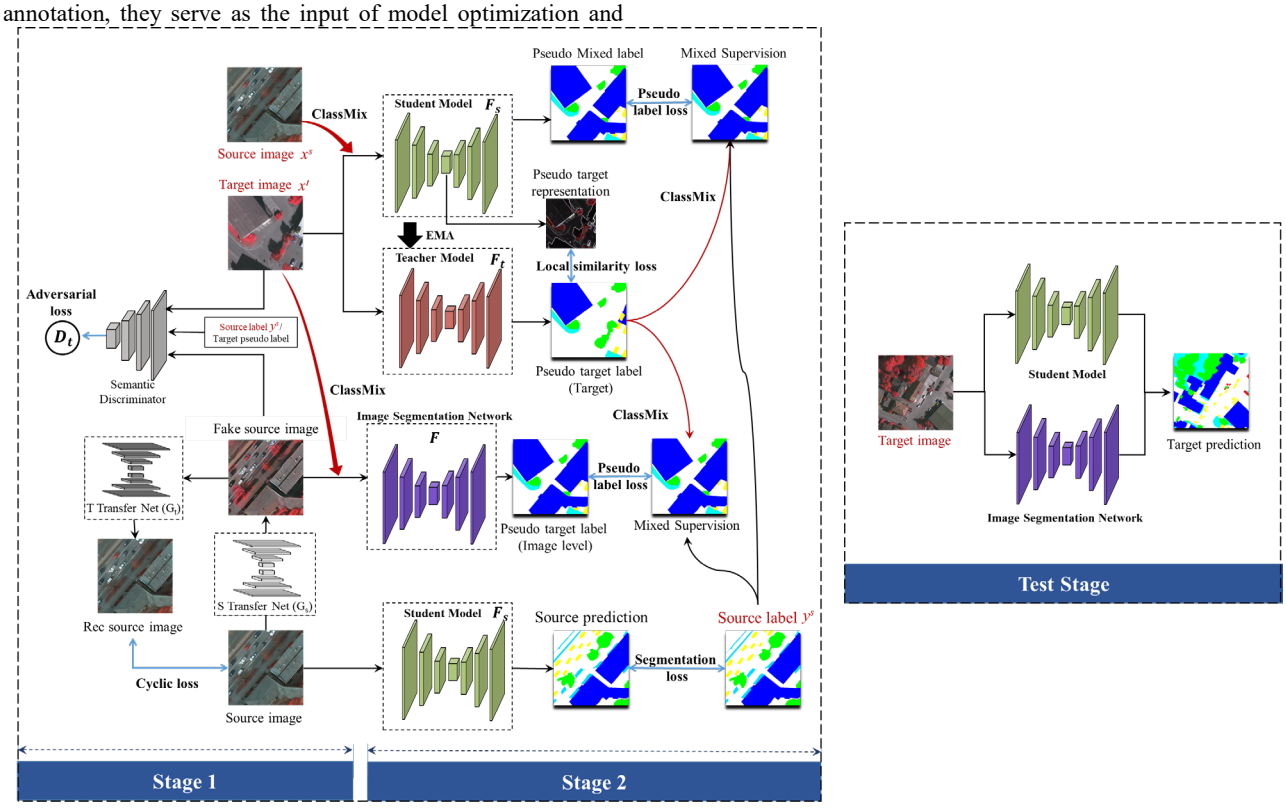

Figure 1. The architecture of IFSDA.

Fake images and corresponding real images are fed into discriminator. Similarly, taking the source domain as an example, the loss function is as follows:

$$\begin{cases} L_{adv}(\boldsymbol{D_t}) = \mathbb{E}_{x^t \sim pdata(x^t)}\left[\boldsymbol{D_t}(x^t)^2\right] + \mathbb{E}_{x^s \sim pdata(x^s)}\left[1 - \boldsymbol{D_t}(\boldsymbol{G_s}(x^s))\right)^2\Big] \\ L_{adv}(\boldsymbol{G_s}) = \mathbb{E}_{x^s \sim pdata(x^s)}\left[1 - \boldsymbol{D_t}(\boldsymbol{G_s}(x^s))\right)^2\Big] \end{cases} \quad (1)$$

where D denotes the discriminator and G denotes the generator. And the cyclic loss is calculated by translation result and original

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-193-2026 | © Author(s) 2026. CC BY 4.0 License.

input:

$$L_{cyc}(G) = \mathbb{E}_{x^s \sim pdata(x^s)} \left[\left|\left| x^s - \boldsymbol{G_t}(\boldsymbol{G_s}(x^s)) \right|\right|_1 \right] \quad (2)$$

where the $||*||_1$ denotes the L1 norm, and the above losses of target domain are similar.

Besides, in order to boost the translation performance and get robust result for the subsequent semantic segmentation, a semantic discriminator is utilized here.

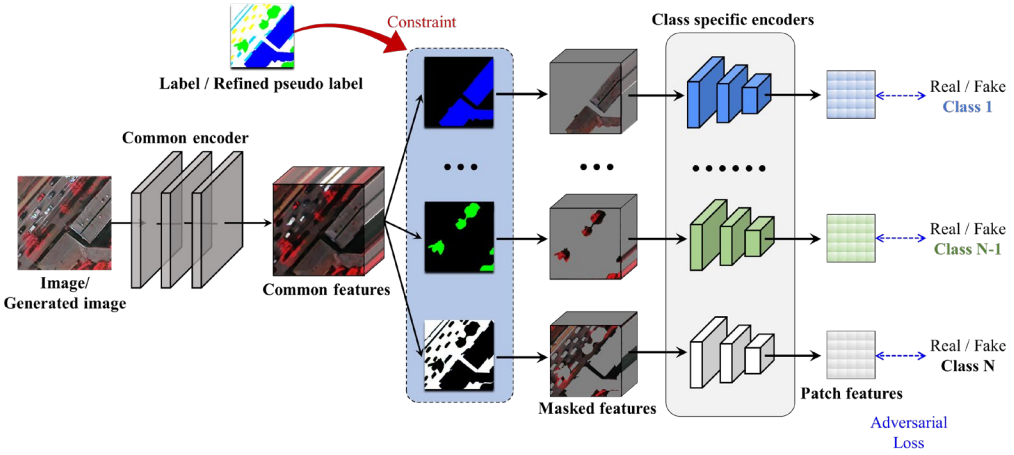

Figure 2. The architecture of semantic discriminator

Its core idea is separately considering the areas associated with different classes during the discrimination process. As depicted in Figure. 2, the semantic discriminator consists of a common encoder and class specific encoders. Specifically, the data go through the two processes and generate a metric to conduct adversarial process. Firstly, the real image and generated fake image are fed to a common encoder to extract shared features. Then, based on the corresponding label or pseudo label, features corresponding to specific class areas are activated within the intermediate feature maps. These activated features are subsequently passed into the relevant class-specific encoder to acquire deep, class-discriminative features. Finally, these features can be constrained by maximizing the discrimination score of PatchGAN for real features and minimizing it for generated fake features. By integrating the Semantic discriminator into the framework, the image-level alignment can be optimized by class-wise constraint. During the network training, the pseudo label becomes more accurate and sound translation outcomes are acquired through the generative adversarial process. The semantic discriminator is a convolutional neural network, and both the common encoder and class-specific encoder contain two convolution layers.

The second stage solves the problem of cross-domain segmentation by combining feature-level alignment and image-level alignment. Thus, this stage has three feature-level constraints and two image-level constraints. As depicted in Figure. 1, $\boldsymbol{F_s}$, $\boldsymbol{F_t}$ and $\boldsymbol{F}$ denote the student model, teacher model and image-level segmentation model. They share the same structure, the encoder is a resnet50_v1c network from OpenMMLab and the segmentation head is a depthwise separable segmentation head with atrous spatial pyramid pooling.

As for implicit feature-level segmentation, the first one is segmentation loss supervised by the label of source domain.

$$L_{seg}(\boldsymbol{F_s}) = -\frac{1}{H \times W} \sum_{i=1}^{H \times W} \log\left(\boldsymbol{F_s}(x_i^s)\right) \times onehot(y_i^s) \quad (3)$$

where H, W denote the height and width of image respectively, $onehot(*)$ denotes the onehot encoding.

Self-supervised learning is crucial when target label is not available, and ClassMix strategy can compensate this shortcoming effectively.

In brief, teacher model uses exponential moving average to take both historical parameters and the current learned parameters into consideration. It is believed that the teacher model is the combination of previous optimization, it can produce robust segmentation results for supervision on target domain. ClassMix strategy is utilized, which is realized by replacing half of the categories in target pseudo label by source label and corresponding target image or generated fake image by source image. And this mixed supervision can takes robust pseudo supervision and accurate source supervision into consideration.

For the feature-level alignment, we utilize a pseudo label loss L_{pse} to make the mixed pseudo target segmentation to learn from accurate source supervision, thereby encouraging model to perform confident predictions on the part of target domain:

$$L_{pse}(\boldsymbol{F_s}) = \frac{-q(x^t)}{H \times W} \times \sum_{i=1}^{H \times W} \log(\boldsymbol{F_s}(\mathbf{Mix}(x_i^t, x_i^s))) \times \mathbf{Mix}(onehot(\boldsymbol{F_t}(x_i^t), y_i^s)) \quad (4)$$

where **Mix** denotes the ClassMix strategy and $q(x^t)$ is a number to count classwise maximum output probability which is larger than the threshold τ:

$$q(x^t) = \frac{1}{H \times W} \sum_{i=1}^{H \times W} [\max \boldsymbol{F_t}(x_i^t)_c > \tau] \quad (5)$$

where $\max \boldsymbol{F_t}(x_i^t)_c$ is the maximum probability of the classes in the position i among all the probability after the Softmax activation. A big value of q occurs when the pseudo prediction of teacher model has high confidence on a specific class, it denotes the robustness of the prediction. It is noticeable that strong augmentation is utilized to mixed image including random color jitter and Gaussian blur.

Besides, a local similarity loss L_{sim} is utilized to enhance the performance of student model, and source feature distribution loss is utilized to maximize feature similarity of the same category in feature-level DA, which is similar to PFST (Zhang et al., 2023).

$$L_{sim}(\boldsymbol{F_s}) = \frac{-1}{H \cdot W \cdot |\Omega|} \sum_{i=1}^{H \times W} \sum_{j \in \Omega} \cos\left(\boldsymbol{et}(x_i^t), \boldsymbol{et}(x_j^t)\right) p + \left(1 - \cos\left(\boldsymbol{et}(x_i^t), \boldsymbol{et}(x_j^t)\right)\right)(1-p) \quad (6)$$

where Ω denotes the neighbor area of pixel i, $\boldsymbol{et}$ is the encoder of the teacher model, $\cos(x, y)$ denotes the cosine similarity of x and y, p is the probability that x_i^t, x_j^t share the same class, it is calculated by the dot product of $\boldsymbol{F_s}(x_i^t)$ and $\boldsymbol{F_s}(x_j^t)$.

Note that the optimization of teacher model considers the

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-193-2026 | © Author(s) 2026. CC BY 4.0 License.

historical teacher parameter s and current student parameters, it is updated by exponential moving average:

$$\theta_t^{Teacher} = \alpha \cdot \theta_{t-1}^{Teacher} + (1-\alpha) \cdot \theta_t^{Student} \quad (7)$$

where θ represents the model parameters, $Teacher$, $Student$ denote teacher model and student model respectively, α denotes the EMA momentum, and t is the number of iteration.

It is noticeable that in IFSDA, image-level translation result is adopted to further refine the segmentation result. The ClassMix strategy of generated fake source image is depicted in Figure. 3.

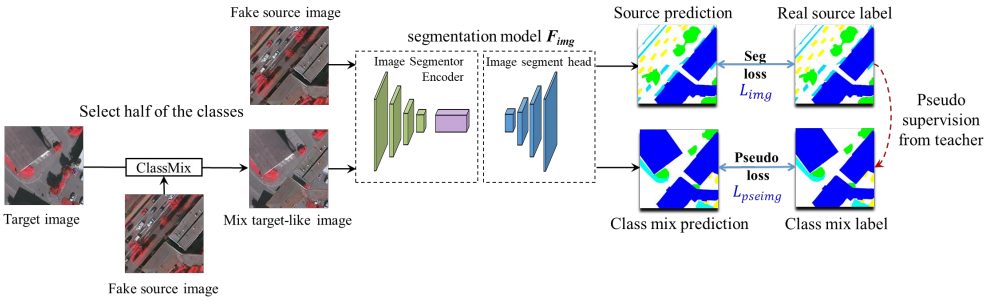

Figure 3. The architecture of self-supervised learning and the ClassMix Strategy.

As depicted in Figure. 3, $\boldsymbol{F_{img}}$ is designed for the segmentation of target stylized images (Fake source image) and it shares the same structure with the feature-level segmentation model. The $\boldsymbol{F_{img}}$ is constrained by two manners, the translation image supervision and the supervision by the mixture of target-like images.

Given the image translation of source image has the same class distribution as source image, source label is utilized to constrain the segmentation of translation image, and L_{img} is acquired as follows:

$$L_{img}(x^t, x^{s'}) = \sum_{i=1}^{H\times W} \log(\boldsymbol{F_{img}}(\boldsymbol{G_s}(x^s))) \times onehot(y_i^s) \quad (8)$$

Similar to feature-level ClassMix, another kind of ClassMix is adopted in the translation-guided segmentation process, by selecting half of the source label and replacing it by pseudo target label, simultaneously, the corresponding area of translation image is replaced by target image. The pseudo information loss of translation L_{pseimg} is then acquired:

$$L_{pseimg}(\boldsymbol{F_{img}}) = \frac{-q(x^t)}{H\times W} \sum_{i=1}^{H\times W} \log(\boldsymbol{F_{img}}(\mathbf{Mix}(x_i^t, \boldsymbol{G_s}(x^s)))) \times \mathbf{Mix}(onehot(\boldsymbol{F_t}(x_i^t), y_i^s)) \quad (9)$$

where q is the robustness of prediction calculated by Formula (5).

As for the overall optimization goal, when optimizing the adversarial network of image generation process, the objective function is

$$L_{adv}(\boldsymbol{D_s}, \boldsymbol{D_t}) = -L_{adv}(\boldsymbol{D_s}) - L_{adv}(\boldsymbol{D_t}) \quad (10)$$

and when training the segmentation network, the objective function is

$$L_G(\boldsymbol{G_s}, \boldsymbol{G_t}, \boldsymbol{F_s}, \boldsymbol{F_{img}}) = L_{adv}(\boldsymbol{G_s}, \boldsymbol{G_t}) + \lambda_1 L_{cyc}(\boldsymbol{G_s}, \boldsymbol{G_t}) + \lambda_2 L_{seg}(\boldsymbol{F_s}) + \lambda_3 L_{pse}(\boldsymbol{F_s}) + \lambda_4 L_{sim}(\boldsymbol{F_s}) + \lambda_5 L_{pseimg}(\boldsymbol{F_{img}}) \quad (10)$$

where λ_1 to λ_5 are different weights assigned to each loss.

3. Experiments

3.1 Datasets and Experimental Setting

ISPRS Potsdam and Vaihingen datasets are chosen as the experimental datasets. Potsdam is a German city with large architectural complexes, narrow streets, and dense settlement structures, while Vaihingen is a small German village. Both datasets have six consistent categories including impervious surface, building, low vegetation, tree, car and clutter. As many researches adopted, the Potsdam dataset has four bands including infrared (IR), red (R), green (G) and blue (B) bands while Vaihingen dataset has three bands including R, G and B. 24 patches of Potsdam dataset and 16 patches of Vaihingen dataset are set as training set, which is the official setting.

Two cross domain segmentation tasks are conducted, including (a) Potsdam IRRG to Vaihingen IRRG, and (b) Vaihingen IRRG to Potsdam IRRG, where the former dataset serves as the source domain.

The experiments were conducted on Super G7210 Rack-mounted workstation with Nvidia Quadro RTX A5000 GPUs and Intel Xeon 6330 CPU, The deep learning algorithms were built based on PyTorch and MMCV architecture. Intersection over Union (IoU) was used to evaluate the performance.

3.2 Comparative experiments

In order to verify the effectiveness of the proposed IFSDA, experiments have been conducted on ISPRS 2D semantic labeling contest dataset. And several state-of-the-art (SOTA) comparative methods were adopted to exhibit the superiority of IFSDA, including DAFormer (Hoyer et al., 2022), PFST (Zhang et al., 2023), and image-level methods CycleGAN (Zhu et al., 2017) is also considered with the segmentation network realized by self-training. For fair, all the comparative methods were conducted under the codebase of MMSegmentation, their dataloader, network formulation, optimizers are the same.

First, Table 1 shows the metrics of the results of Potsdam IRRG to Vaihingen IRRG segmentation. The "Source" denotes the baseline model trained on source data only.

Method / Class	Source	DAFormer	PFST	CycleGAN	IFSDA
Impv.	58.20	75.78	78.81	70.38	**78.94**
Build.	73.51	87.19	**87.91**	84.00	87.20
Low.	35.81	49.98	**55.81**	49.52	55.25
Tree	59.83	64.92	57.47	**68.12**	60.28
Car	48.76	60.51	61.98	60.86	**63.69**
Clut.	16.61	20.08	35.44	20.08	**39.50**
mIoU	48.79	59.74	62.90	58.83	**64.14**

Table 1. IOU (%) of different methods on Potsdam IRRG to Vaihingen IRRG segmentation

As shown in the results, the IFSDA has three highest IoUs over all comparative methods, including highest mIoU, which improves the performance by 15.35% and outperforms DAFormer, PFST and self-training segmentation by CycleGAN for 4.40%, 1.24% and 5.31% respectively. Besides, the IoUs of Building and Low Vegetation are just slightly lower than the optimal value (PFST) by 0.71% and 0.56%, which further exhibits it superiority. The visualization results of this experiment are shown in Figure. 4. In the first row, it is obvious that the building detection result of IFSDA inside the red rectangle is the most complete. Besides, the rectangles in the second row denote misclassification of Clutter and the ignorance

of Building, while there is significant misclassification of DAFormer and CycleGAN is significant, while the results of PFST and IFSDA are better. The third row exhibits the performance of distinguishing low vegetation and tree, the result of CycleGAN wrongly takes low vegetation as tree, although it has highest IoU of tree. Meanwhile, PFST has many noises of clutter. In comparison, IFSDA has the most accurate segmentation result.

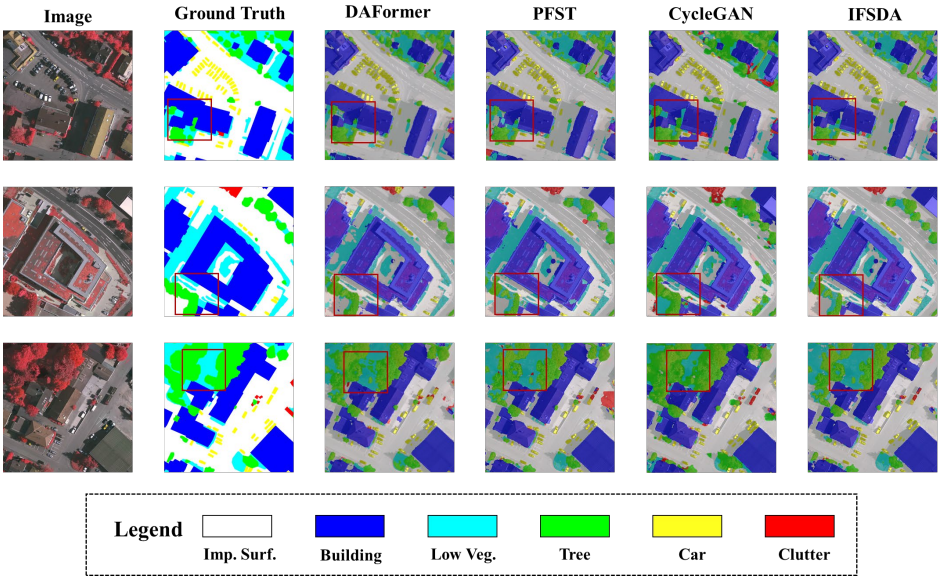

Figure 4. Visualization results on Potsdam to Vaihingen Segmentation.

Regarding the results of Vaihingen IRRG to Potsdam IRRG segmentation, the metrics and visualization results are shown in Table 2 and Figure. 5. Given the amount of Clutter in the Vaihingen dataset is insufficient to train a robust prediction, the results of the Clutter are omitted. IFSDA improves the mIoU by 6.87% and exceeds DAFormer, PFST and DAFormer for 3.66%, 2.43% and 1.08%. As for the visualization results, the highlighted road in the first row and the highlighted building in the second row show the excellent ability of IFSDA in distinguishing details. Besides, in the third row, the circled area shows the best performance of IFSDA in distinguishing individual trees, while the building segmentation of IFSDA is the optimal.

Method / Class	Source	DAFormer	PFST	CycleGAN	IFSDA
Impv.	65.87	68.46	70.09	68.01	**72.13**
Build.	70.42	73.20	70.35	80.36	**81.54**
Low.	50.86	54.86	49.30	48.46	**56.15**
Tree	19.46	32.57	23.25	**54.47**	40.84
Car	**72.78**	71.50	69.40	58.76	68.38
mIoU	55.89	60.12	56.49	62.01	**63.81**

Table 2. IOU (%) of different methods on Vaihingen IRRG to Potsdam IRRG segmentation

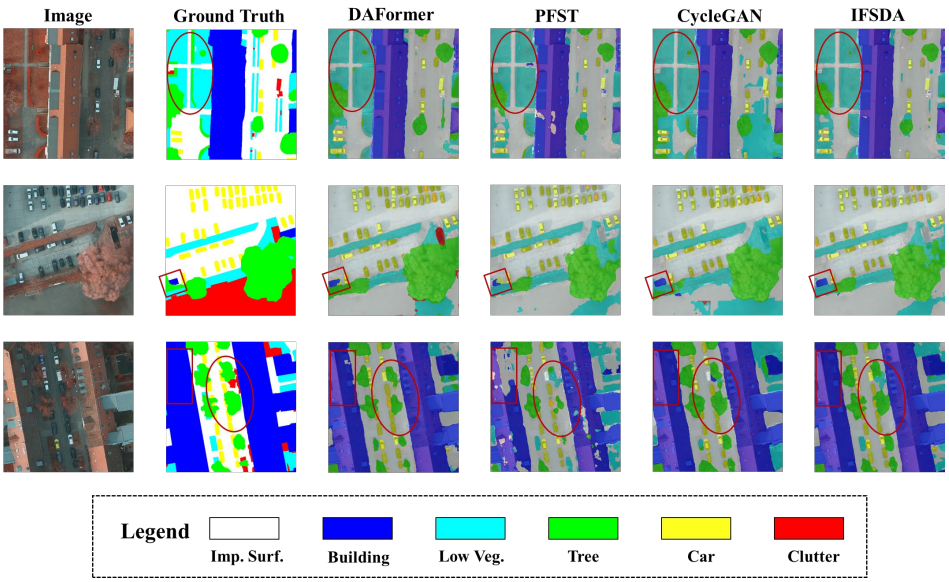

Figure 5. Visualization results on Vaihingen IRRG to Potsdam IRRG Segmentation.

4. Conclusion and Future Works

In this paper, a novel image-level and feature-level semantic-aware domain adaptation method for bi-temporal segmentation named IFSDA is proposed. Two levels of alignment strategies and two kinds of ClassMix are utilized to effectively combine the explicit and implicit features. Besides, class-wise discriminator strategy further refines the results by integrating source annotation and target pseudo information. Experiments on Potsdam and Vaihingen datasets have demonstrated its superiority over feature-alignment methods and a deep translation-based method, which shows great potential in cross domain semantic segmentation of remote sensing imagery. In the future, we will dedicate ourselves to exploring advanced generative structures to achieve superior image-level translation results, realizing strong integration of feature extraction network, discovering the mutual reinforcement of transfer learning and semantic segmentation and improving the efficiency by network light-weighting.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No. 42571368.

References

Cai, Y., Yang, Y., Shang, Y., Chen, Z., Shen, Z., & Yin, J., 2022. IterDANet: Iterative Intra-Domain Adaptation for Semantic Segmentation of Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-17. https://doi.org/10.1109/tgrs.2022.3203040.

Chen, H., Zhang, H., Yang, G., Li, S., & Zhang, L., 2022a. A Mutual Information Domain Adaptation Network for Remotely Sensed Semantic Segmentation. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-16. https://doi.org/10.1109/tgrs.2022.3203910.

Chen, X., Pan, S., & Chong, Y., 2022b. Unsupervised Domain Adaptation for Remote Sensing Image Semantic Segmentation Using Region and Category Adaptive Domain Discriminator. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-13. https://doi.org/10.1109/tgrs.2022.3200246.

Gatys, L.A., Ecker, A.S., & Bethge, M., 2016. Image Style Transfer Using Convolutional Neural Networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2414-2423. https://doi.org/10.1109/cvpr.2016.265.

Hoyer, L., Dai, D., & Van Gool, L., 2022. DAFormer: Improving Network Architectures and Training Strategies for Domain-Adaptive Semantic Segmentation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 9914-9925. https://doi.org/10.1109/cvpr52688.2022.00969.

Iizuka, R., Xia, J., & Yokoya, N., 2024. Frequency-Based Optimal Style Mix for Domain Generalization in Semantic Segmentation of Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing, 62, 1-14. https://doi.org/10.1109/tgrs.2023.3344670.

Johnson, J., Alahi, A., & Li, F., 2016. Perceptual Losses for Real-time Style Transfer and Super-resolution. European Conference on Computer Vision (ECCV), Springer, 694-711.

Li, X., Zhang, L., Wang, Q., & Ai, H., 2020. Multi-temporal Remote Sensing Imagery Semantic Segmentation Color Consistency Adversarial Network. Acta Geodaetica et Cartographica Sinica, 49, 1473-1484. https://doi.org/10.11947/j.AGCS.2020.20190439.

Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., & Yang, M.-H., 2017. Universal Style Transfer via Feature Transforms. Advances in Neural Information Processing Systems (NIPS), 386-396.

Li, Y., Shi, T., Zhang, Y., & Ma, J., 2023. SPGAN-DA: Semantic-Preserved Generative Adversarial Network for Domain Adaptive Remote Sensing Image Semantic Segmentation. IEEE Transactions on Geoscience and Remote Sensing, 61, 1-17. https://doi.org/10.1109/tgrs.2023.3313883.

Li, Z., Lei, Z., Xie, M., Ji, H., Li, Y., Zhu, J., & Gao, Z., 2024. CDTCL: Cross-Domain Remote Sensing Image Translation for Semantic Segmentation Leveraging Contrastive Learning. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 7354-7358. https://doi.org/10.1109/igarss53475.2024.10641195.

Liang, C., Li, W., Dong, Y., & Fu, W., 2024. Single Domain Generalization Method for Remote Sensing Image Segmentation via Category Consistency on Domain Randomization. IEEE Transactions on Geoscience and Remote Sensing, 62, 1-16. https://doi.org/10.1109/tgrs.2024.3379669.

Luan, F., Paris, S., Shechtman, E., & Bala, K., 2017. Deep Photo Style Transfer. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6997-7005. https://doi.org/10.1109/cvpr.2017.740.

Luo, M., & Ji, S., 2022. Cross-spatiotemporal land-cover classification from VHR remote sensing images with deep learning based domain adaptation. ISPRS Journal of Photogrammetry and Remote Sensing, 191, 105-128. https://doi.org/10.1016/j.isprsjprs.2022.07.011.

Niemeijer, J., Schwonberg, M., Termöhlen, J.-A., Schmidt, N.M., & Fingscheidt, T., 2024. Generalization by Adaptation: Diffusion-Based Domain Extension for Domain-Generalized Semantic Segmentation. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2818-2828. https://doi.org/10.1109/wacv57701.2024.00281.

Wu, J., Guo, D., Wang, G., Yue, Q., Yu, H., Li, K., & Zhang, S., 2024. FPL+: Filtered Pseudo Label-Based Unsupervised Cross-Modality Adaptation for 3D Medical Image Segmentation. IEEE Transactions on Medical Imaging, 43, 3098-3109. https://doi.org/10.1109/TMI.2024.3387415.

Yoo, J., Uh, Y., Chun, S., Kang, B., & Ha, J.-W., 2019. Photorealistic Style Transfer via Wavelet Transforms. IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 9035-9044. https://doi.org/10.1109/ICCV.2019.00913.

Zhang, F., Shi, Y., Xiong, Z., Huang, W., & Zhu, X.X., 2023. Pseudo Features-Guided Self-Training for Domain Adaptive Semantic Segmentation of Satellite Images. IEEE Transactions on Geoscience and Remote Sensing, 61, 1-14. https://doi.org/10.1109/tgrs.2023.3281503.

Zhao, Q., Lyu, S., Zhao, H., Liu, B., Chen, L., & Cheng, G., 2024. Self-training guided disentangled adaptation for cross-domain remote sensing image semantic segmentation. International Journal of Applied Earth Observation and Geoinformation, 127, 1-18. https://doi.org/10.1016/j.jag.2023.103646.

Zhu, J.-Y., Park, T., Isola, P., & Efros, A.A., 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. IEEE International Conference on Computer Vision (ICCV), IEEE, 2242-2251. https://doi.org/10.1109/ICCV.2017.244.