scientific reports

Check for updates

OPEN

# Unsupervised domain adaptation for remote sensing semantic segmentation with the 2D discrete wavelet transform

Junying Zeng[1], Yajin Gu[2], Chuanbo Qin[1]✉, Xudong Jia[1,3], Senyao Deng[1], Jiahua Xu[1] & Huiming Tian[1]

**There would be the differences in spectra, scale and resolution between the Remote Sensing datasets of the source and target domains, which would lead to the degradation of the cross-domain segmentation performance of the model. Image transfer faced two problems in the process of domain-adaptive learning: overly focusing on style features while ignoring semantic information, leading to biased transformation results, and easily overlooking the true transfer characteristics of remote sensing images, resulting in unstable model training. To address these issues, we proposes a novel dual-space generative adversarial domain adaptation segmentation framework, DS-DWTGAN, to minimize the differences between the source domain and the target domain. DS-DWTGAN aims to mitigate the distinctions between the source and target domains, thereby rectifying the imbalances in style and semantic representation.The framework introduces a network branch leveraging wavelet transform to capture comprehensive frequency domain and semantic information. It aims to preserve semantic details within the frequency domain space, mitigating image conversion deviations. Furthermore, our proposed method integrates output adaptation and data enhancement training strategies to reinforce the acquisition of domain-invariant features. This approach effectively diminishes noise interference during the migration process, bolsters model stability, and elevates the model's adaptability to remote sensing images within different domains. Experimental validation was conducted on the publicly available Potsdam and Vaihingen datasets. The findings reveal that in the PotsdamIRRG to Vaihingen task, the proposed method attains outstanding performance with mIoU and mF1 values reaching 56.04% and 67.28%, respectively. Notably, these metrics surpass the corresponding values achieved by state-of-the-art (SOTA) methods, registering an increase of 2.81% and 2.08%. In comparison to alternative approaches, our proposed framework exhibits superior efficacy in the domain of unsupervised semantic segmentation for UAV remote sensing images.**

Recently, semantic segmentation of remote sensing images using fully supervised learning has attained high accuracy and robustness, however, it necessitates a substantial amount of labeled data[1–3]. Pixel-level annotation in remote sensing is both time-consuming and costly[4]. Remote sensing images exhibit inconsistencies in landscapes across various regions and variations in image acquisition due to different sensors or weather conditions. Consequently, significant disparities arise in data styles across regions or within the same region at different times or under different sensor setups. This disparity results in a notable degradation in segmentation performance of fully supervised models in practical cross-domain segmentation tasks due to the absence of semantic annotation information in the target domain[5,6]. Particularly when dealing with Earth observation data from multiple platforms, the disparities between datasets escalate the intricacy of image semantic annotation[5,6,17].

Unsupervised semantic segmentation methods aim to minimize differences in feature distribution between the source and target domains by leveraging shared information. This enables the model to better adapt to the feature distribution of the target dataset, thus bolstering its both the generalization capabilities of the model and the precision of image semantic segmentation. This method can mitigate issues of insufficient or unlabeled annotations in the target dataset while enhancing the performance of semantic segmentation models. By

[1]School of Electronics and Information Engineering, Wuyi University, Guangdong 529020, China. [2]College of Intelligent Systems Science and Engineering, Guangzhou Huali College, Guangdong 511325, China. [3]College of Engineering and Computer Science, California State University, Northridge, Northridge, CA 91330, USA. ✉email: tenround@163.com

leveraging the unsupervised domain adaptation semantic segmentation approach, the segmentation performance of remote sensing images can be significantly improved, offering superior support for applications such as the automatic interpretation and target detection of drone remote sensing images. Unsupervised domain adaptation methods are currently mainly divided into several categories, including self supervised training[8–10], adversarial learning[6,7,11–13], and image to image conversion[14–16], while there are also some emerging methods being explored and applied. Self-supervision, while capable of diminishing reliance on annotated data, encounters difficulties in acquiring high-quality feature representation and demonstrating sufficient generalization ability. Although adversarial training can effectively capture the mapping relationship between these two domains, he training process exhibits instability, rendering convergence a formidable task. Image style transfer migrates semantic segmentation knowledge from the source domain to the target domain, aiming to preserve the stylistic attributes of the latter. This process empowers the adapted model to more effectively accommodate novel data[17,18]. In simple terms, convolutional neural networks must acquire a mapping technique capable of converting source domain images into a novel feature space, ensuring a high degree of visual coherence with the target domain data.The similarity between samples in the source domain and the target domain has a significantly positive impact on the performance of image segmentation. Tasar et al.[14] proposed a ColorMapGAN, a color mapping generative network capable of transforming the colors of training images into those of target images without any structural changes to objects in the training images. Similarly, Zhao et al.[19] designed ResiDualGAN based on residual networks and explored the adaptive potential of Generative Adversarial Networks $(GAN)$ in cross-domain semantic segmentation tasks for remote sensing images. Zhang et al.[20] proposed a local-to-global remote sensing image segmentation framework, which completes the domain adaptation process in two stages. Li et al.[21] introduced a stepwise domain adaptation remote sensing image segmentation network with mitigated covariate shift to narrow the gap between the source domain and the target domain.

Generally, unsupervised domain adaptation methods offer an effective solution for semantically segmenting remote sensing images. By employing techniques like transfer learning and feature transformation, models can more effectively align with the target domain's distribution, consequently enhancing semantic segmentation accuracy. Although existing methods have achieved some success, they are not yet perfect in handling cross-domain segmentation from real remote sensing images to real scenes. At the same time, most existing methods learn from a single space, neglecting the importance of simultaneously extracting features in the frequency domain and spatial domain. Thus, we introduce DS-DWTGAN, a dual-branch generative network based on wavelet transform. This network aims to mitigate the potential loss of semantic information and diminish disparities in data distribution by integrating insights from both frequency and spatial domains. Such an approach offers novel perspectives and avenues for tackling the challenge of cross-domain semantic segmentation in remote sensing imagery.The research's primary contributions are given below:

1. We propose a novel dual space generative network DS-DWTGAN to address the issue of excessive emphasis on style and neglect of semantic information, in order to achieve visual transformation from the source domain to the target domain and reduce the distribution differences between datasets. By applying discrete wavelet transform, a wider range of image features can be captured, while also enhancing the ability to map and model features between source and target domain images.
2. To address the instability inherent in model training, an adaptive strategy for output features has been implemented to facilitate the concurrent training of the segmentation model and output discriminator. This strategy meticulously aligns the distribution of output features, minimizing disparities across feature distributions. Consequently, the model's capacity to discern intricate image features is significantly augmented, leading to a notable enhancement in convergence speed and overall model stability.
3. In order to effectively address the characteristics of remote sensing images, this study introduces a data augmentation training strategy. This strategy enables the model to better learn the rich color and texture information in remote sensing images, while reducing the influence of noise on the model during the transfer process, enhancing the robustness and generalization of the model. We conducted cross domain semantic segmentation experiments on open-source remote sensing datasets Potsdam and Vaihingen, verifying the superiority of the proposed method in handling cross domain semantic segmentation tasks.

## Methods

### Overview

To provide a more specific description of the problem of unsupervised domain adaptation, we denote the labeled source domain as $I_S = \{(X_S, Y_S)\}^{n_S}$ and unlabeled target domain datasets as and $I_T = \{(X_T)\}^{n_T}$. Where $X_S$ represents the source domain samples, and $Y_S$ its corresponding labels. $X_T$ denote the target domain samples. $n_S$ and $n_T$ respectively denote the sample sizes of the source and target domains.

Our proposed methodology comprises two stages, depicted in Fig. 1. In the first stage, we utilize the proposed generation network to establish the mapping between the source domain and target domain image data distributions, and generate target-stylized source domain data to achieve image transformation between the source domain and target domain. In the second stage, by utilizing pseudo-target images with source domain labels obtained in the first stage to train the semantic segmentation network model in a supervised manner. Output adaptation modules and data augmentation functions were subsequently introduced in the subsequent training process to adjust and improve the segmentation results, thereby bolstering the robustness and generalization of the cross-domain segmentation model.

### Image generation stage

The architectural design with dual-branch architecture has been effectively utilized in various fully-supervised tasks of semantic segmentation[22–24]. With this structure, each branch possesses its unique approach to processing

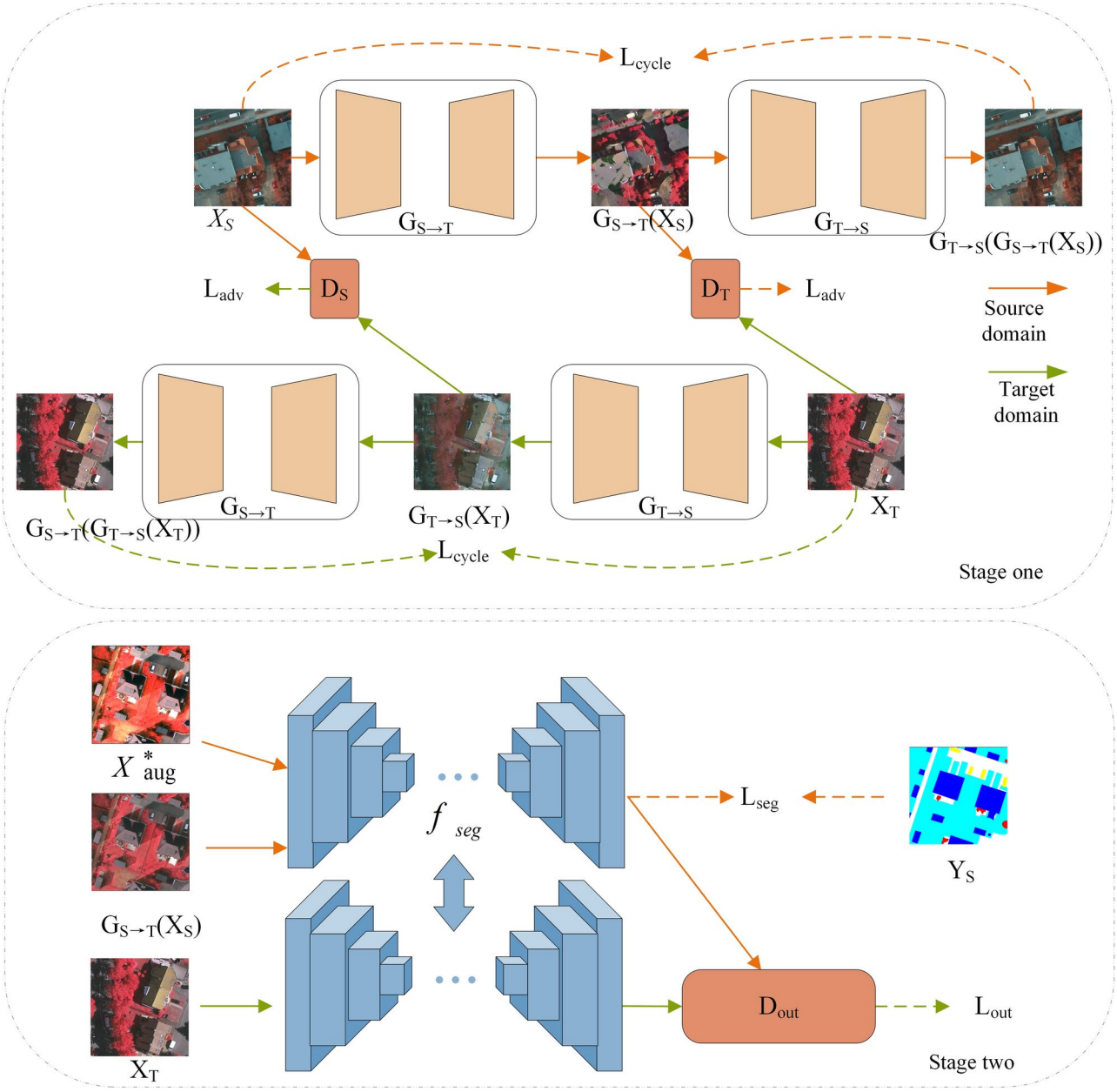


**Fig. 1**. Overall framework. Orange streamlines indicate source-domain sample transformations, green streamlines indicate target-domain sample transformations, L denotes the training loss.

information and is capable of extracting feature information of varying dimensions from the same input. Integrating feature maps from both branches directly easily loses context information around detailed features. Therefore, we chose validated FFM feature fusion modules to complement each other. Fully utilizing diverse feature information, the disparity among images across domains is reduced, thereby enhancing performance in image style transfer. Based on the aforementioned idea, we designed a dual-space GAN that simultaneously learns in the frequency and spatial domains, as illustrated in Fig. 2.

*Wavelet generation network*

This branch focuses on learning the mapping of spectral information from source domain images to target domain images. We devised a Wavelet Generation Network structured upon the U-Net architecture[25], depicted in Fig. 2 , comprising an encoder, a decoder, and interconnecting jump connections at every feature scale. Discrete wavelet transform is employed during the feature extraction stage, decomposing input features into high-frequency and low-frequency components. Low-frequency components and convolutional outputs are cascaded to form downsampled features, whereas high-frequency components are integrated into the upsampling module of the wavelet transform via skip connections. This enables our network to learn spatial information as well as rich frequency domain information. The up-sampling and down-sampling modules of the wavelet transform are

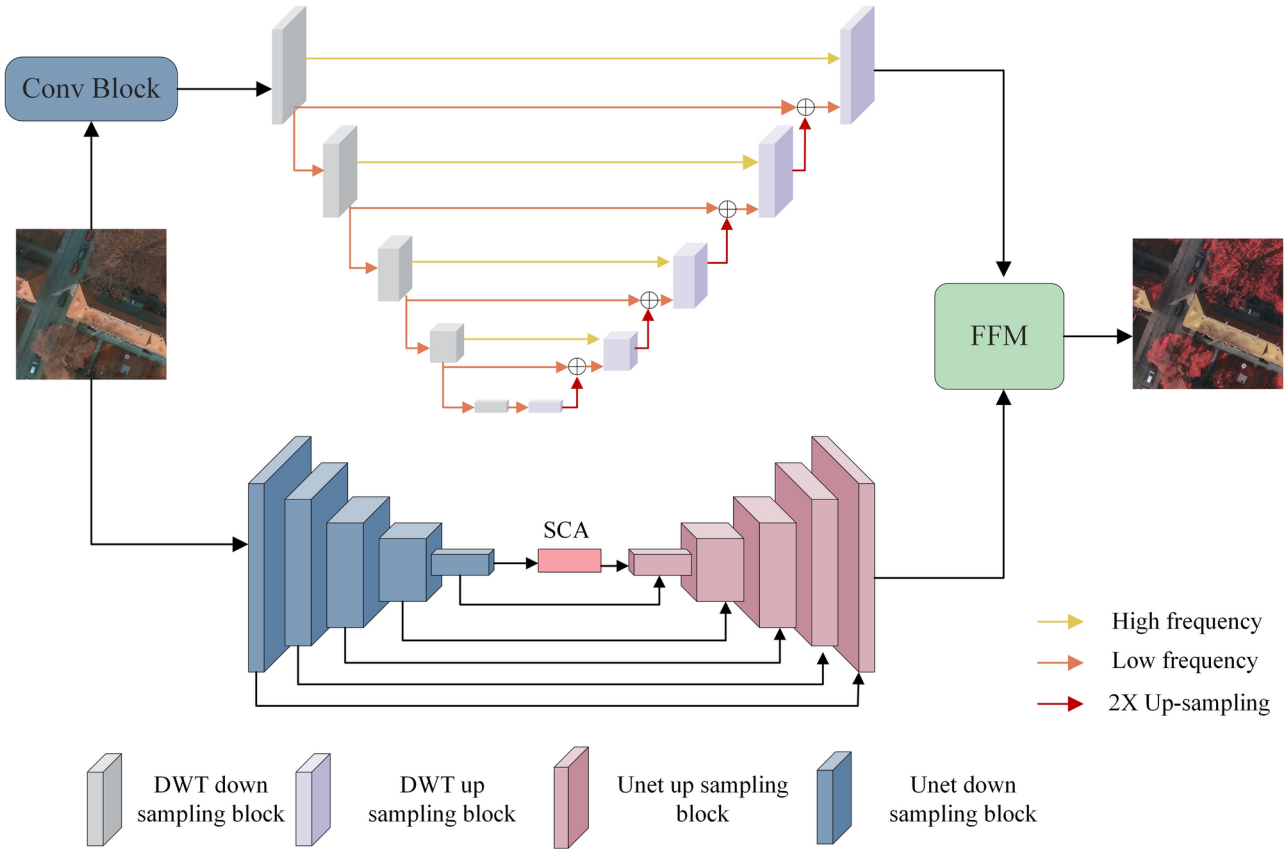


**Fig. 2**. Dual space adversarial generative network.

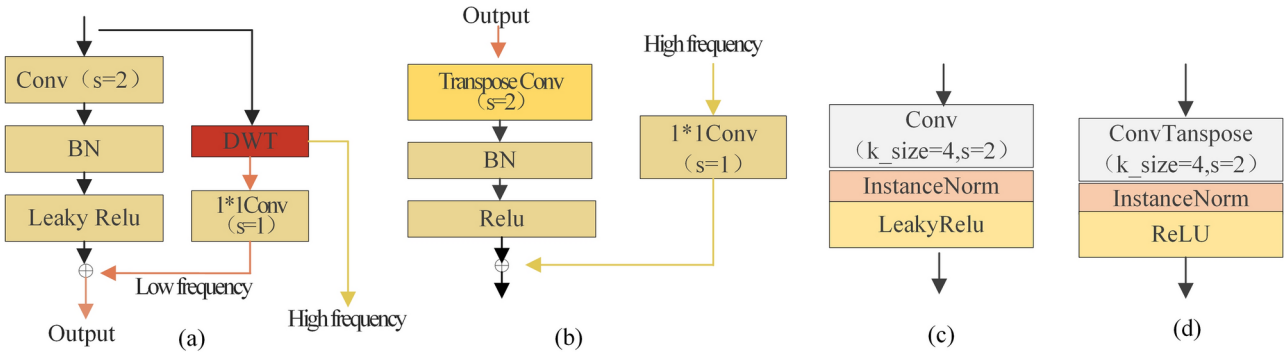


**Fig. 3**. Up and down sampling module. Where (**a**) and (**b**) denote the modules for downsampling and upsampling of wavelet transform, and (**c**) and (**d**) denote the modules for downsampling and upsampling of convolutional branching, respectively.

illustrated in Fig. 3a,b, respectively.Wavelet transform is a fundamentally time-frequency analysis method[26–28], which decomposes the input signal into images of different frequencies through high-pass $(\mathrm{F_{LH}}, \mathrm{F_{HL}}, \mathrm{F_{HH}})$ and low-pass $F_{LL}$ filters. DWT stands out for its ability to facilitate reversible downsampling. It achieves this by decomposing two-dimensional data into four discrete wavelet components: a low-frequency component $I_{LL}$ and three high-frequency components $(\mathrm{I_{LH}}, \mathrm{I_{HL}}, \mathrm{I_{HH}})$ through filter convolution.

The low-frequency component, denoted as $I_{LL} = F_{LL} * X$ ,$*$ is expressed through a convolution operation, while the high-frequency component shares a similar expression to the low-frequency one. Leveraging Discrete Wavelet Transform, we can capture detailed information in the wavelet domain of images across various scales, particularly from the $I_{LH}$ , $I_{HL}$ and $I_{HH}$ components. However, due to the limited size of the remote sensing dataset, achieving optimal performance solely through the DWT branch proves challenging. Consequently, we introduce a secondary branch to augment the learning process with additional information features, thereby enhancing the overall performance on the dataset.

*Convolutional manipulation module*

The convolutional operation branches and the generation network, akin to prevalent models[17,18,29], comprise the downsampling and upsampling of U-Net alongside skip connections. To accommodate the significant scale transformation of remote sensing images and the prevalence of small targets, we introduce a Spatial Channel Attention module. During the feature extraction stage, greater emphasis can be placed on small targets of interest to alleviate the issue of their neglect during the image learning process. Within an image, a correlation exists between the geographical positions of objects, like buildings and urban roads, where the pixels of cars and roads exhibit spatial connectivity. Convolution can extract long-distance contextual information to enhance model performance. Additionally, the relationship between feature mappings at various channel levels within the image is crucial for semantic segmentation. Consequently, we propose a Spatial Channel Attention module to augment image generation by leveraging spatial positions and channel relationships, as illustrated in Fig. 4.

*Generative adversarial learning*

In this paper, we first perform the image generation process to preserve the semantic information of the source domain image and learn the stylized representation of the target domain image. GAN based structure this paper uses two generators $G_{S\to T}$ and $G_{T\to S}$ and two discriminators $D_S$ and $D_T$ . With $X_{S\to T}$ representing the transformation from source domain image to target domain image, $G_{S\to T}$ denotes the target domain generator and $G_{T\to S}$ denotes the source domain generator. The source domain discriminator $D_S$ distinguishes between the source domain image and the generated pseudo-target image, while the target domain discriminator $D_T$ distinguishes between the target domain image and the generated pseudo-source image. The generator contains a wavelet generation network $G_{DWT}$ a convolutional generation network $G_{Conv}$ (Fig. 2 shows), and the image generation process is illustrated in Eqs. (1) and 2.

$$X_{S\to T} = G_{S\to T}(X_S) = G_{DWT}(X_S) + G_{Conv}(X_S), \tag{1}$$

$$X_{T\to S} = G_{T\to S}(X_T) = G_{DWT}(X_T) + G_{Conv}(X_T), \tag{2}$$

Through iterative processes, this study trains the generator to produce images that deceive the discriminator, which concurrently endeavors to discern whether an image is authentic or generated. The adversarial dynamic between the generator and discriminator is encapsulated in Eqs. (3) and (4).

$$\mathscr{L}_{adv}^{S\to T}(D_T, G_{S\to T}) = \mathbb{E}_{x_T\sim l_T}[(D_T(x_T))] + \mathbb{E}_{x_s\sim l_S}\left[(D_T(G_{S\to T}(x_S)))\right], \tag{3}$$

$$\mathscr{L}_{adv}^{T\to S}(D_S, G_{T\to S}) = \mathbb{E}_{x_s\sim I_S}\left[\left(D_S(x_S)\right)\right] + \mathbb{E}_{x_T\sim I_T}\left[\left(D_S\left(G_{T\to S}(x_T)\right)\right)\right], \tag{4}$$

To promote content preservation from the source domain image throughout the image transformation process, we integrated an image cycle consistency constraint . This constraint aims to minimize the error between the reconstructed and original images. The introduction of the L1 norm is utilized to regulate image consistency, as illustrated in Eq. (5).

$$\mathscr{L}_{cyc} = \mathbb{E}_{x_s\sim I_s}(\|G_{T\to S}(G_{S\to T}(x_S)) - x_S\|_1) + \mathbb{E}_{x_T\sim I_T}(\|G_{S\to T}(G_{T\to S}(x_T)) - x_T\|_1), \tag{5}$$

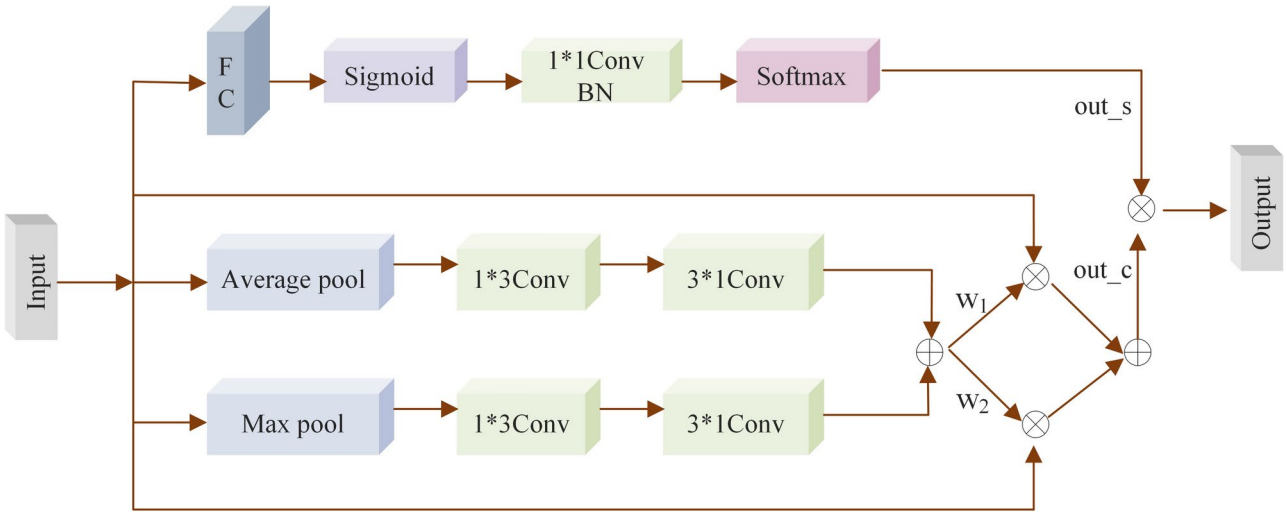


**Fig. 4**. Space channel attention module. Spatial attention and channel attention work together to enhance the expression ability of small target features.

### Segmentation training

Our primary aim during the segmentation phase is to develop a semantic segmentation model denoted as $f_{seg}$, capable of achieving optimal performance on unlabeled target domains . The semantic segmentation model undergoes training on labeled data from the source domain, stylized to resemble the target domain, and assimilates features transferred from the latter. Despite this, the features learned during training prove inadequate for direct application to authentic target data. Hence, to enhance the generalization performance across remote sensing images, we introduce an output adaptive module and a data enhancement function. These augmentations aim to refine the model's capability to adapt to the intricacies of the target domain, thereby improving overall segmentation quality.

For our semantic segmentation model, we opted for DeepLabV3+[3], and for expedited model inference, we selected ResNet34[30] as the backbone of the DeepLabV3+ network. The coding structure employs null convolution for multi-scale feature extraction, while the decoding structure incorporates Dropout at the final layer to mitigate overfitting issues during training.

*Output space adaptive (OSA)*

Throughout the training phase of semantic segmentation, the feature encoder grapples with high-dimensional structural and textural data, rendering the inference process intricate and challenging for accomplishing domain adaptation tasks. Hence, this paper focuses on addressing domain adaptation in the output space.In the output space, specifically within the segmentation network's softmax output, we propose utilizing a Generative Adversarial Network to align the distributions of both the Potsdam and Vaihingen datasets. While the images from both datasets manifest noteworthy dissimilarities in spectral and visual characteristics, they also manifest numerous congruences in their outputs. These include spatial layout, characterized by a prevalence of buildings in urban locales and a profusion of vegetation in rural settings, as well as local contextual features like the proximity of vehicles to buildings. Consequently, we argue that regardless of the dataset origin, its segmentation outcomes ought to exhibit specific resemblances.

When executing the output adaptive module, this paper treats the segmentation model $f_{seg}$ as a traditional GAN generator, which produces softmax predictive output probability maps for two inputs $X_{S\to T}$ and $X_T$. The outputs of the segmentation model $f_{seg}$ are the same as those of the output adaptive module. At the same time, a discriminator $D_{out}$ is used to distinguish the output of $f_{seg}$ from either $X_{S\to T}$ or $X_T$. As in the traditional GAN approach, the discriminator is trained to distinguish the true from the false, and then the generator is trained to produce images that can deceive the discriminator. The discriminator training process is shown in Eq. (6) .

$$\mathscr{L}_{out} = \mathbb{E}\left(\log_2\left(1 - D_{out}(f_{seg}(X_T))\right)\right) - \mathbb{E}\left(\log_2\left(D_{out}(f_{seg}(X_{S\to T}))\right)\right), \tag{6}$$

*Data augmentation*

Remote sensing images often possess abundant color and texture features. To mitigate noise interference during image feature transfer learning, we employ color jittering techniques to improve image quality and stability, facilitating segmentation algorithms in better identifying features and enhancing segmentation accuracy. Remote sensing images typically contain abundant color and texture information.

Color jitter is a method that boosts image contrast by leveraging alterations in color. Introducing variations in color within the original image's color space enhances contrast. Such alterations may include adjusting pixel brightness, saturation, or hue to produce a visual effect that highlights the targets in the image. In practical applications, Color jitter introduces an offset to the grayscale value of the current pixel, determined by the error values of neighboring pixels within the color space distribution. The implementation of the algorithm involves the following steps: (1) randomly selecting a pixel from the original image; (2) randomly adjusting the color of the selected pixel, including altering brightness, hue, or saturation; (3) reintegrating the modified pixel into the original image; (4) iterating through these steps until either all pixels have undergone modification or a predetermined number of iterations has been achieved.

## Experiments

### Dataset

The Potsdam and Vaihingen Remote Sensing datasets represent prominent 2D semantic segmentation benchmarks within the International Society for Photogrammetry and Remote Sensing (http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-potsdam.html, http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-vaihingen.html), originating from aerial photography. The distinct geographic locations and spectral characteristics inherent to these datasets offer diverse experimental scenarios for cross-domain adaptation. Therefore, this study assesses the efficacy of our modeling framework across both datasets to comprehensively evaluate its performance.

These two datasets are widely utilized in remote sensing research, featuring consistent semantic annotation categories. These categories include impervious surfaces, buildings, low vegetation, trees, cars, and clutter. The Potsdam remote sensing dataset comprises aerial imagery captured over the city of Potsdam, Germany. It encompasses 38 high-resolution remote sensing images, each with dimensions of $6000 \times 6000$ pixels and a ground sampling distance (GSD) of 5 cm. This dataset provides information across four bands: infrared, red, green, and blue (IRRGB), utilized for both IRRG and RGB band analyses, denoted as PotsdamIRRG and RGB, respectively. Conversely, the Vaihingen dataset was acquired over various regions of the city of Vaihingen, Germany. It comprises 33 high-resolution remote sensing images, each with dimensions of approximately $2000 \times 2000$ pixels and a GSD of 9 cm. This dataset includes information from three bands: infrared, red, and

green . To address computational constraints, this study preprocesses both datasets using an image cropping method to optimize memory usage.

## Experimental detail

The entire model was implemented using the PyTorch framework. All experiments were conducted on a machine featuring an Intel Core i9-12900K CPU, 32 GB of RAM, and an NVIDIA GeForce RTX A4000 GPU with 16 GB of graphics memory.

In the experimental section, we evaluate the segmentation performance of cross-domain Very High Resolution remote sensing images using two key metrics: the mean Intersection over Union (mIoU) and the mean F1 score (mF1). These widely accepted statistical measures facilitate a comprehensive comparison between the performance of our proposed GAN architecture and existing methodologies. Specifically, we compute mF1 and mIoU for five foreground classes :buildings, trees, low vegetation, car, and impervious surfaces, according to Eqs. (7) and (8), respectively.

$$mIoU = \frac{1}{N}\sum_{n=1}^{N}\frac{TP_n}{TP_n + FP_n + FN_n}, \tag{7}$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{8}$$

where $precision = TP/(TP + FP)$ and $recall = TP/(TP + FN)$. $TP_n$, $FP_n$, $TN_n$ and $FN_n$ represent true positives, false positives, true negatives, and false negatives, respectively, for feature information indexed to category n.

## Ablation experiment

We performed ablation experiments on the PotsdamIRRG, PotsdamRGB, and VaihingenIRRG datasets to ascertain the significance and impact of various modules on the task at hand, as delineated in Table 1. The PotsdamIRRG and VaihingenIRRG datasets differ in terms of category distribution, geographical coverage, and resolution. In comparison with PotsdamRGB and VaihingenIRRG, other than variations in category distribution, geographical coverage, and resolution, there are also differing factors related to image spectra.

This section utilizes the convolutional Unet generation network as the baseline model. For the cross-domain segmentation task PotsdamIRRG → VaihingenIRRG, the mIoU and mF1 scores stand at $45.85\%$ and $58.90\%$, respectively. The $Dwt\_unet$ denotes the fusion of a wavelet generation network with a convolutional generation network. While $Dwt\_unet$ initially performs slightly less effectively than the baseline model in direct segmentation, the subsequent introduction of an output adaptive module and a data augmentation strategy significantly enhances segmentation accuracy. After the introduction of output adaptation and data augmentation methods into the baseline model, respective improvements of $5.69\%$ and $9.06\%$ in mIoU were observed. Importantly, within the $Dwt\_unet$ generation network, the integration of the OSA module and the utilized data augmentation strategy yielded mIoU and mF1 values of $56.04\%$ and $67.28\%$ respectively. In

| | Unet | Dwt_unet | OSA | aug | mIoU (%) | mF1 (%) |
|---|---|---|---|---|---|---|
| PotsdamIRRG ↓ VaihingenIRRG | ✓ | | | | 45.85 | 58.9 |
| | ✓ | | ✓ | | 51.54 | 63.68 |
| | ✓ | | ✓ | ✓ | 54.91 | 66.53 |
| | | ✓ | | | 45.29 | 58.22 |
| | | ✓ | ✓ | | 53.23 | 65.20 |
| | | ✓ | ✓ | ✓ | **56.04** | **67.28** |
| VaihingenIRRG ↓ PotsdamRGB | ✓ | | | | 43.45 | 56.50 |
| | ✓ | | ✓ | | 51.16 | 62.28 |
| | ✓ | | ✓ | ✓ | 53.54 | 64.43 |
| | | ✓ | | | 43.56 | 56.27 |
| | | ✓ | ✓ | | 49.6 | 61.61 |
| | | ✓ | ✓ | ✓ | **53.94** | **64.67** |
| PotsdamRGB ↓ VaihingenIRRG | ✓ | | | | 40.81 | 52.85 |
| | ✓ | | ✓ | | 42.39 | 54.69 |
| | ✓ | | ✓ | ✓ | 49.4 | 61.29 |
| | | ✓ | | | 43.42 | 55.49 |
| | | ✓ | ✓ | | 50.15 | 62.61 |
| | | ✓ | ✓ | ✓ | **51.03** | **63.20** |

**Table 1**. Ablation studies of different modules of DU-DWTGAN on different tasks. Significant values are in bold.

comparison to the baseline model, the experimental outcomes demonstrated a rise of 10.19% in mIoU and 8.38% in mF1. This enhancement serves as a comprehensive demonstration of the efficacy of the output adaptation module and data augmentation techniques in bolstering the stability of denoising and enhancing models, consequently enhancing the model's generalization capability. In the cross-domain adaptation experiments from VaihingenIRRG to PotsdamRGB and from PotsdamRGB to VaihingenIRRG, the segmentation performance of the two domains improved after the addition of various modules. Specifically, the mIoU scores were 53.94% and 51.03%, and the mF1 scores were 64.67% and 63.20%, respectively. These significant enhancements in data substantiate the efficacy of the modules proposed by us for cross-domain segmentation tasks. Our approach effectively reduce differences in dataset distributions, thereby significantly enhancing the accuracy and reliability of model segmentation.

## Comparison with other methods

In the experimental validation phase, the efficacy of the proposed method is substantiated by employing UAV Remote Sensing datasets from diverse domains as source datasets and conducting comparative experiments on three distinct UAV remote sensing datasets. The comparison encompasses several models, namely DualGAN, CycleGAN, FADA, MemoryAdaptNet[31], MBATA-GAN[32] and ResiDualGAN. The former two methods, along with ResiDualGAN, are specialized in image-to-image style transformations.MemoryAdaptNet is an output space adversarial learning method. MBATA-GAN is a domain adaptation model based on global attention transformation. FADA, on the other hand, focuses on fine-grained adversarial learning for cross-domain semantic segmentation tasks. The experimental findings are presented in Tables 2, 3, and 4. In this study, DeepLabv3+ with ResNet34 serving as the backbone network is adopted as the baseline model to assess the segmentation model's real-world performance in the presence of domain disparities. The baseline model is trained on labeled datasets and evaluated on unlabeled datasets. As evident from the data presented in the subsequent tables, the segmentation outcomes post-domain adaptation notably surpass those of the baseline model, underscoring the efficacy of the domain-adapted segmentation approach in mitigating data distribution disparities and enhancing the segmentation of minute targets.

Tables 2 and 3 present the results of cross-domain segmentation tasks between the PotsdamIRRG and VaihingenIRRG datasets. In the cross-domain task PotsdamIRRG → VaihingenIRRG, the baseline model achieved segmentation results with an mIoU of 29.85% and an mF1 of 41.76%. In contrast, our proposed framework exhibits superior performance in remote sensing image semantic segmentation tasks. Specifically, our model achieved an mIoU of 56.04% and an mF1 of 67.28%, indicating a 26.19% increase in mIoU compared to the baseline model. Compared to the second-best model, our model showed further improvement, with an increase of 3.95% in mIoU and 3.03% in mF1 based on performance metrics.

In the cross-domain segmentation task in Table 3, VaihingenIRRG is the source domain, while PotsdamIRRG is the target domain. The baseline model exhibited the poorest segmentation performance, with mIoU and mF1 values of 29.85% and 41.76% respectively. Following domain adaptation, both existing methods and our proposed model demonstrated enhanced performance in evaluation metrics. Specially, our proposed DS-DWTGAN model outperformed others, achieving the highest mIoU and mF1 scores of 56.68% and 67.25% respectively. In comparison to the suboptimal ResidualGAN model, our model has shown enhancements of 3.95% and 3.03% in terms of mIoU and mF1, respectively. Moreover, our proposed approach exhibits elevated

| Task | Methods | | Clutter | Imp. surfaces | Car | Tree | Low vegetation | Building | Overall |
|---|---|---|---|---|---|---|---|---|---|
| PotsdamIRRG ↓ VaihingenIRRG | Source-Only | IOU% | 1.83 | 30.48 | 17.93 | 52.46 | 16.63 | 59.76 | 29.85 |
| | | F1% | 3.23 | 46.15 | 26.79 | 68.61 | 28.22 | 74.57 | 41.76 |
| | CycleGAN | IOU% | 3.11 | 43.73 | 9.85 | 58.44 | 33.28 | 49.54 | 32.99 |
| | | F1% | 4.80 | 60.3 | 17.25 | 73.58 | 49.49 | 65.9 | 45.22 |
| | DualGAN | IOU% | 3.58 | 52.67 | 19.74 | 62.24 | 39.8 | 62.83 | 40.14 |
| | | F1% | 5.66 | 68.53 | 32.19 | 76.63 | 56.53 | 76.86 | 52.73 |
| | FADA | IOU% | 10.83 | 62.26 | 39.82 | 64.13 | 43.22 | 72.22 | 48.75 |
| | | F1% | 19.54 | 76.74 | 56.96 | 78.18 | 60.35 | 83.87 | 62.6 |
| | ResidualGAN | IOU% | 7.76 | 70.09 | 50.27 | 60.53 | 46.19 | 77.71 | 52.09 |
| | | F1% | 11.42 | 82.28 | 66.44 | 75.22 | 62.78 | 87.38 | 64.25 |
| | MemoryAdaptNet | IOU% | 10.86 | 67.82 | 44.30 | 50.27 | 46.06 | 76.92 | 49.37 |
| | | F1% | 19.59 | 80.82 | 61.40 | 66.91 | 63.07 | 86.95 | 63.12 |
| | MBATA-GAN | IOU% | 3.78 | 64.92 | 38.51 | 51.33 | 40.19 | 74.67 | 45.57 |
| | | F1% | 7.29 | 78.73 | 55.61 | 67.85 | 57.34 | 85.50 | 58.72 |
| | Ours | IOU% | 7.95 | 72.64 | 56.67 | 65.63 | 49.92 | 83.42 | **56.04** |
| | | F1% | 11.15 | 84.05 | 72.05 | 79.16 | 66.33 | 90.92 | **67.28** |

**Table 2**. PotsdamIRRG → VaihingenIRRG quantitative results for cross-domain segmentation. Significant values are in bold.

| Task | Methods | | Clutter | Imp.surfaces | Car | Tree | Low vegetation | building | overall |
|---|---|---|---|---|---|---|---|---|---|
| VaihingenIRRG ↓ PotsdamIRRG | Source-Only | IOU % | 5.82 | 48.44 | 37.12 | 12.03 | 37.71 | 49.48 | 31.77 |
| | | F1% | 7.73 | 64.48 | 52.63 | 20.63 | 53.63 | 65.43 | 44.09 |
| | CycleGAN | IOU% | 7.53 | 52.65 | 38.89 | 39.27 | 39.25 | 48.98 | 37.76 |
| | | F1% | 10.42 | 68.37 | 54.94 | 55.68 | 55.51 | 64.87 | 51.63 |
| | DualGAN | IOU% | 7.28 | 47.95 | 45.73 | 34.77 | 45.02 | 49.17 | 38.32 |
| | | F1% | 10.03 | 64.02 | 61.73 | 50.67 | 61.14 | 64.92 | 52.09 |
| | FADA | IOU% | 17.40 | 57.53 | 66.55 | 35.85 | 45.22 | 61.25 | 47.30 |
| | | F1% | 29.64 | 73.04 | 79.92 | 52.78 | 62.28 | 75.97 | 62.27 |
| | ResidualGAN | IOU% | 2.52 | 72.41 | 68.29 | 47.46 | 49.47 | 82.27 | 53.74 |
| | | F1% | 3.87 | 83.83 | 81.04 | 64.00 | 65.58 | 90.11 | 64.74 |
| | MemoryAdaptNet | IOU% | 10.60 | 59.59 | 62.62 | 44.93 | 44.75 | 64.15 | 47.78 |
| | | F1% | 19.16 | 74.68 | 77.01 | 62.00 | 61.83 | 78.16 | 62.14 |
| | MBATA-GAN | IOU% | 0.56 | 60.90 | 47.58 | 36.13 | 26.90 | 68.06 | 40.02 |
| | | F1% | 1.11 | 75.70 | 64.48 | 53.08 | 42.39 | 81.00 | 52.96 |
| | Ours | IOU% | 3.44 | 72.64 | 71.69 | 53.57 | 55.61 | 83.13 | **56.68** |
| | | F1% | 5.12 | 84.00 | 83.38 | 69.44 | 70.89 | 90.66 | **67.25** |

**Table 3**. VaihingenIRRG→PotsdamIRRG quantitative results for cross-domain segmentation. Significant values are in bold.

| Task | Methods | | Clutter | Imp.surfaces | Car | Tree | Low vegetation | building | overall |
|---|---|---|---|---|---|---|---|---|---|
| PotsdamRGB ↓ VaihingenIRRG | Source-Only | IOU% | 1.81 | 33.41 | 13.59 | 55.48 | 12.51 | 55.40 | 28.70 |
| | | F1% | 3.11 | 49.38 | 23.39 | 71.2 | 22.08 | 71.06 | 40.04 |
| | CycleGAN | IOU% | 2.22 | 43.44 | 14.69 | 54.22 | 13.72 | 52.53 | 30.14 |
| | | F1% | 3.75 | 60.03 | 24.67 | 70.12 | 23.77 | 68.54 | 41.82 |
| | DualGAN | IOU% | 2.97 | 39.76 | 13.67 | 57.66 | 15.78 | 57.50 | 31.22 |
| | | F1% | 4.92 | 56.40 | 23.12 | 72.99 | 26.99 | 72.74 | 42.86 |
| | FADA | IOU% | 12.01 | 49.83 | 35.22 | 46.91 | 29.54 | 73.64 | 41.19 |
| | | F1% | 21.45 | 66.51 | 52.10 | 63.86 | 45.61 | 84.81 | 55.73 |
| | ResidualGAN | IOU% | 6.83 | 65.27 | 57.19 | 57.41 | 38.61 | 79.63 | 50.82 |
| | | F1% | 10.20 | 78.80 | 72.48 | 72.76 | 55.37 | 88.59 | 63.03 |
| | MemoryAdaptNet | IOU% | 18.40 | 66.14 | 49.41 | 35.48 | 36.76 | 77.90 | 47.34 |
| | | F1% | 31.70 | 79.62 | 66.14 | 52.38 | 53.76 | 87.58 | 61.76 |
| | MBATA-GAN | IOU% | 0.85 | 48.88 | 34.37 | 42.58 | 27.56 | 73.21 | 37.91 |
| | | F1% | 1.70 | 65.67 | 51.16 | 59.73 | 43.21 | 84.54 | 51.00 |
| | Ours | IOU% | 7.44 | 64.07 | 55.59 | 57.66 | 40.19 | 81.20 | **51.03** |
| | | F1% | 10.81 | 77.88 | 71.00 | 73.00 | 56.96 | 89.57 | **63.20** |

**Table 4**. PotsdamRGB→VaihingenIRRG quantitative results for cross-domain segmentation. Significant values are in bold.

precision and reliability in identifying and handling small target objects, such as those classified as 'car' and 'tree' pixels, with IOU enhancements reaching 71.69% and 53.57%, respectively. This unequivocally showcases the superiority of our model in fine object recognition. Furthermore, regarding the segmentation performance of other categories, our method has additionally achieved effective enhancements, thereby further augmenting the overall effectiveness of the segmentation task.

In conclusion, our method demonstrates superior semantic segmentation performance compared to other methods, while retaining a greater amount of semantic information, as evidenced by the experimental results presented in Tables 2 and 3. These findings robustly validate the efficacy of our proposed model in effectively addressing cross-domain segmentation tasks using the PotsdamIRRG and VaihingenIRRG datasets. It is worth noting that our model not only preserves a richer set of semantic details but also consistently delivers improved segmentation outcomes.

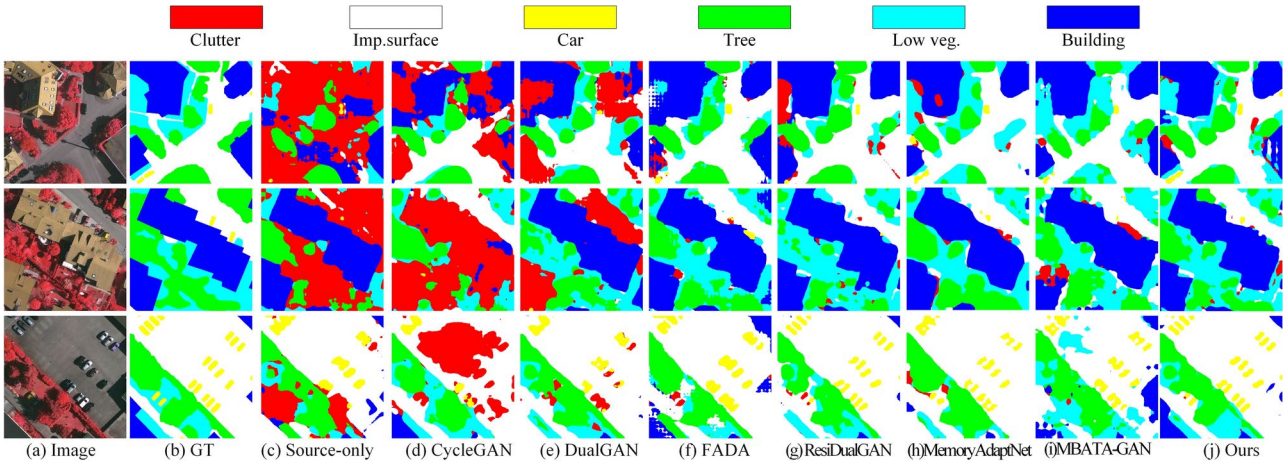


**Fig. 5**. Quantitative visualization of the PotsdamIRRG →VaihingenIRRG task.

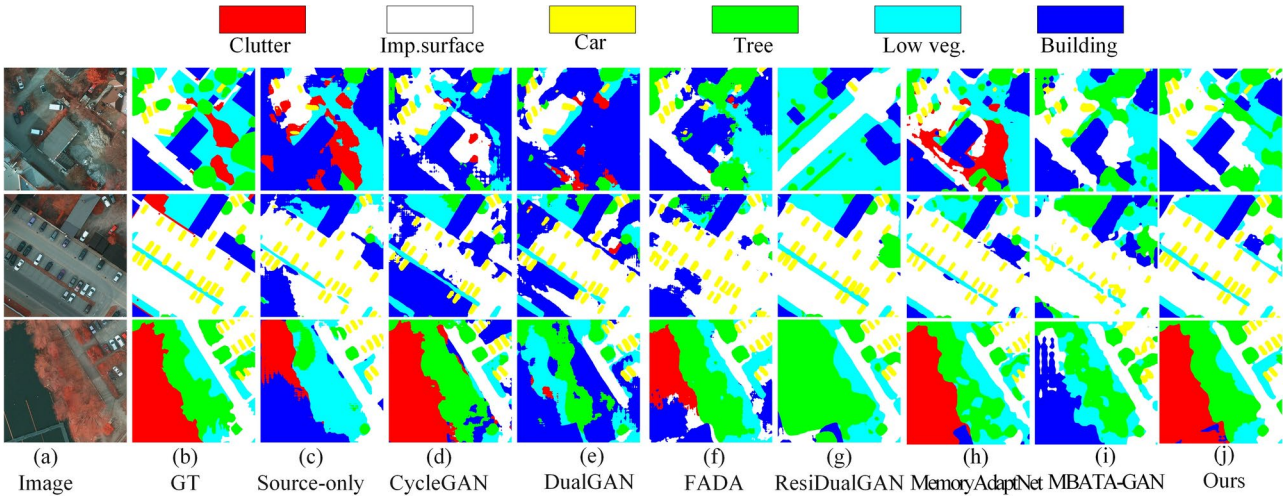


**Fig. 6**. Quantitative visualization of the VaihingenIRRG → PotsdamIRRG task.

Figure 5 displays the visualization results of the model's inference on the cross-domain task PotsdamIRRG → VaihingenIRRG, while Fig. 6 presents the corresponding outcomes for the VaihingenIRRG → PotsdamIRRG task. A clear observation from Figs. 5 and 6 reveals that the visualization effect of our proposed semantic segmentation method closely resembles the real labels. This outstanding performance stems from the model's comprehensive acquisition of frequency domain information during the image generation phase, coupled with effective model optimization during the segmentation phase. Notably, our method demonstrates proficient performance even for smaller object categories, such as cars and low vegetation. The visualization outcomes underscore our proficiency in handling cross-domain segmentation tasks and affirm our method's capability to accurately segment small objects in complex scenes, thereby validating the model's effectiveness in both image generation and segmentation phases.

The cross-domain segmentation results for the PotsdamRGB→VaihingenIRRG task is shown in Table 4 , which shows that the domain offsets increase the band factor of the imaging compared to the cross-domain tasks between the PotsdamIRRG and VaihingenIRRG datasets.

In the cross-domain segmentation task involving PotsdamRGB as the source domain and VaihingenIRRG as the target domain, the baseline model demonstrates modest segmentation performance, yielding 28.70% and 40.04% for mIoU and mF1, respectively. Notably, individual category predictions exhibit notable deficiencies, particularly in "clutter," "car," and "low vegetation," with IOUs of only 1.81%, 13.59%, and 12.51%, respectively. Introducing the proposed DS-DWTGAN model results in improved performance, achieving 51.03% mIoU and 63.20% mF1. Although the enhancement over the baseline method is relatively marginal, the DS-DWTGAN model exhibits significant progress with 22.33% increase in mIoU and 23.16% increase in mF1. In the PotsdamRGB→VaihingenIRRG task, the proposed method substantially improves segmentation across various categories, with "clutter," "impervious surfaces," "car," "tree," "low vegetation," and "building" categories reaching IOU of 7.44%, 64.07%, 55.59%, 57.66%, 40.19%, and 81.20%, respectively. Comparing the results from

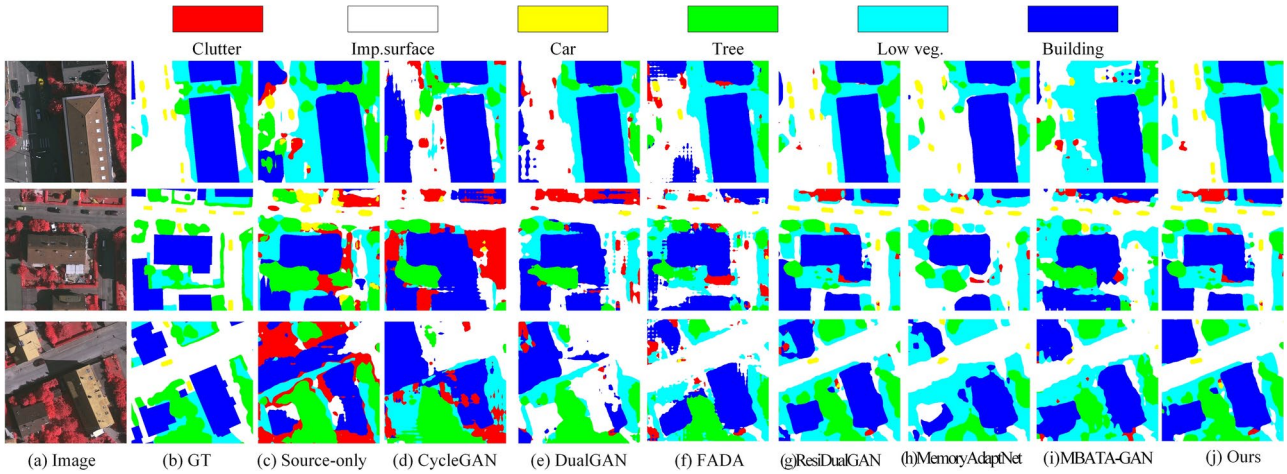


**Fig. 7**. Quantitative visualization of the PotsdamRGB→VaihingenIRRG task.

| Method | CycleGAN | DualGAN | FADA | ResiDualGAN | MemeoryAdaptNet | MBATA-GAN | Ours |
|---|---|---|---|---|---|---|---|
| Samples/second | 2.11 | 2.11 | 3.38 | 2.11 | 2.82 | 1.24 | 2.34 |
| Params (M) | 22.44 | 22.44 | 42.94 | 22.44 | 58.63 | 143.91 | 1.81 |
| Flops (G) | 63.36 | 63.36 | 359.96 | 63.36 | 484.22 | 1393.13 | 11.10 |

**Table 5**. Model efficiency of different methods.

Tables 2 and 4 reveals a noteworthy impact on the cross-domain segmentation task despite PotsdamIRRG and PotsdamRGB datasets capturing images within the same geographical region but employing different bands.

Figure 7 depicts the visualization outcomes of several model inferences for the cross-domain tasks PotsdamRGB →VaihingenIRRG . The cross-domain tasks involving the PotsdamRGB and VaihingenIRRG datasets are relatively intricate, with the overall segmentation effect appearing slightly inferior when compared to the visualization results in Figs. 5 and 6. Figure 7 exhibits the predicted images obtained through testing methods, showcasing relatively high accuracy in the predicted pixel categories of our semantic segmentation. Despite slightly inferior performance in complex cross-domain tasks, our proposed DS-DWTGAN method exhibits significant potential and relatively high prediction accuracy in semantic segmentation.

In order to comprehensively demonstrate the advantages of our method, we have conducted experiments on the efficiency of the model algorithm. Table 5 compares the efficiency of different domain adaptation models. We compared the inference time, Params and FLOPS of each sample for the 7 different methods in the table. Due to the use of the same network and strategy in the segmentation stages of CycleGAN, DualGAN, and ResiDualGAN models, their performance in inference time, parameter count, and floating-point operations for each sample is consistent. In terms of inference time, MBATA-GAN is the fastest, reaching 1.24 seconds per sample, while our model's inference time is in the middle position. In the comparison of Params, our model is the lightest, only 1.81M, while MBATA-GAN has the largest parameter quantity, reaching 143.91M. As for FLOPS, our model has the lowest complexity, only 11.10G, while MBATA-GAN's FLOPS reaches 1393.13G, with the highest complexity, followed by the MemoryAdaptNet model. Based on the previous analysis of segmentation performance, it has been further confirmed that the model not only performs well in segmentation performance, but also has advantages in execution efficiency.

## Discussion

This study introduces a novel unsupervised domain adaptation method, DS-DWTGAN, tailored for cross-domain semantic segmentation of remote sensing images. DS-DWTGAN integrates discrete wavelet transform-based image transformation to address biases stemming from geographical variations and imaging modalities across Remote Sensing datasets. By incorporating wavelet transform and a spatial channel module within the generative network, the proposed method not only preserves semantic content from the source domain, often overlooked in traditional approaches, but also captures rich frequency domain information while mitigating domain discrepancies. During the segmentation process, we mitigate noise interference and enhance the reliability of image segmentation by optimizing the output space adaptive module and employing data augmentation techniques. DS-DWTGAN has been tested and validated with Remote Sensing datasets, demonstrating remarkable robustness and generalization capabilities. This model can effectively reduce domain shift and improve the cross-domain semantic segmentation of remote sensing images. While the method proposed in this paper has attained a degree of success in addressing the issue of semantic segmentation in cross-domain remote sensing images, it still exhibits certain limitations. To accelerate the deployment of unmanned

aerial remote sensing images, unsupervised domain adaptation methods for remote sensing image semantic segmentation should maximize the utilization of multi-source data's complementary characteristics, all while maintaining computational efficiency. This approach will enhance the model's capability to comprehend and differentiate surface objects with similar features, thereby enhancing its generalization ability and segmentation accuracy across diverse domains.

## Data availability

The datasets analysed during the current study are available in the website, http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-potsdam.html.

Received: 12 May 2024; Accepted: 30 September 2024
Published online: 09 October 2024

## References

1. Xu, M., Wu, M., Chen, K., Zhang, C. & Guo, J. The eyes of the gods: A survey of unsupervised domain adaptation methods based on remote sensing data. *Remote Sens.* **14**(17), 4380 (2022).
2. Tsai, Y. H., Hung, W. C., Schulter, S., Sohn, K., Yang, M. H. & Chandraker, M. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7472–7481 (2018).
3. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, 801–818 (2018).
4. Luo, S., Li, H. & Shen, H. Deeply supervised convolutional neural network for shadow detection based on a novel aerial shadow imagery dataset. *ISPRS J. Photogram. Remote Sens.* **167**, 443–457 (2020).
5. He, Y., Wang, J., Liao, C., Shan, B. & Zhou, X. ClassHyPer: ClassMix-based hybrid perturbations for deep semi-supervised semantic segmentation of remote sensing imagery. *Remote Sens.* **14**(4), 879 (2022).
6. Li, Y., Shi, T., Zhang, Y. & Ma, J. SPGAN-DA: Semantic-preserved generative adversarial network for domain adaptive remote sensing image semantic segmentation. In *IEEE Transactions on Geoscience and Remote Sensing* (2023).
7. Wang, H., Shen, T., Zhang, W., Duan, L. Y. & Mei, T. Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation. In *European Conference on Computer Vision*, 642–659 (2020).
8. Chaitanya, K., Erdil, E., Karani, N. & Konukoglu, E. Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation. *Med. Image Anal.* **87**, 102792 (2023).
9. Caron, M., Houlsby, N. & Schmid, C. Location-aware self-supervised transformers for semantic segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 117–127 (2024).
10. Mei, K., Zhu, C., Zou, J. & Zhang, S. Instance adaptive self-training for unsupervised domain adaptation. In Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVI 16, 415–430 (2020).
11. Zou, Y., Yu, Z., Kumar, B. V. K. & Wang, J. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 289–305 (2018).
12. Vu, T. H., Jain, H., Bucher, M., Cord, M. & Pérez, P. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2517–2526 (2019).
13. Luo, Y., Zheng, L., Guan, T., Yu, J. & Yang, Y. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2507–2516 (2019).
14. Tasar, O., Happy, S. L., Tarabalka, Y. & Alliez, P. ColorMapGAN: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks. *IEEE Trans. Geosci. Remote Sens.* **58**(10), 7178–7193 (2020).
15. Cai, Y. et al. BiFDANet: Unsupervised bidirectional domain adaptation for semantic segmentation of remote sensing images. *Remote Sens.* **14**(1), 190 (2022).
16. Ismael, S. F., Kayabol, K. & Aptoula, E. Unsupervised domain adaptation for the semantic segmentation of remote sensing images via one-shot image-to-image translation. In *IEEE Geoscience and Remote Sensing Letters* (2023).
17. Yi, Z., Zhang, H., Tan, P. & Gong, M. Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE International Conference on Computer Vision*, 2849–2857 (2017).
18. Zhu, J. Y., Park, T., Isola, P. & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, 2223–2232 (2017).
19. Zhao, Y., Guo, P., Sun, Z., Chen, X. & Gao, H. ResiDualGAN: Resize-residual DualGAN for cross-domain remote sensing images semantic segmentation. *Remote Sens.* **15**(5), 1428 (2023).
20. Zhang, B., Chen, T. & Wang, B. Curriculum-style local-to-global adaptation for cross-domain remote sensing image segmentation. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–12 (2021).
21. Li, J. et al. A stepwise domain adaptive segmentation network with covariate shift alleviation for remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–15 (2022).
22. Yu, C., Wang, J., Peng, C., Gao, C., Yu, G. & Sang, N. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 325–341 (2018).
23. Poudel, R. P., Bonde, U. D., Liwicki, S. & Zach, C. ContextNet: Exploring Context and Detail for Semantic Segmentation in Real-time. Preprint at arXiv:1805.04554 (2018).
24. Fan, M., Lai, S., Huang, J., Wei, X., Chai, Z., Luo, J. & Wei, X. Rethinking bisenet for real-time semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9716–9725 (2021).
25. Ronneberger, O., Fischer, P. &Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, part III 18*, 234–241 (2015).
26. Fu, M., Liu, H., Yu, Y., Chen, J. & Wang, K. Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 203–212 (2021).
27. Xiang, S., Liang, Q. & Fang, L. Discrete wavelet transform-based Gaussian mixture model for remote sensing image compression. In *IEEE Transactions on Geoscience and Remote Sensing* (2023).
28. Li, Q. & Shen, L. Wavesnet: Wavelet integrated deep networks for image segmentation. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, 325–337 (2022).
29. Creswell, A. et al. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* **35**(1), 53–65 (2018).
30. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778 (2016).
31. Zhu, J. et al. Unsupervised domain adaptation semantic segmentation of high-resolution remote sensing imagery with invariant domain-level prototype memory. *IEEE Trans. Geosci. Remote Sens.* **61**, 1–18 (2023).
32. Ma, X., Zhang, X., Wang, Z. & Pun, M. Unsupervised domain adaptation augmented by mutually boosted attention for semantic segmentation of VHR remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **61**, 1–15 (2023).

## Author contributions

Writing-review and editing and supervision by Junying Zeng and Chuanbo Qin; Conceptualization, methodology and writing-original draft by Yajin Gu; Project administration and supervision by Xudong Jia; Data curation, validation and experiments by Senyao Deng; visualization and experiments by Jiahua Xu and Huiming Tian. All authors have read and agreed to the published vision of the manuscript.

## Funding

Guangdong Basic and Applied Basic Research Foundation (2021A1515011576), Special Project in key Areas of Artificial Intelligence in Guangdong Universities (No.2019KZDZX1017), Key Research Projects for Universities of Guangdong Provincial Education Department (No.2020ZDZX3031), Guangdong, Hong Kong, Macao and the Greater Bay Area International Science and Technology Innovation Cooperation Project (No. 2021A0 50530080), 2024 Key Research Platform Project for Ordinary Universities in Guangdong Province (2024ZDZX1008), 2022 Guangdong Provincial Education Department GraduateEducation Innovation Project (Guangdong Education and Research Letter ([2022]No.2).

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-74781-y.

**Correspondence** and requests for materials should be addressed to C.Q.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.