Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Unsupervised Domain Adaptation for Cross-domain Remote Sensing Object Detection Via Joint Input and Feature Space

Deliang Chen
Nanjing University of Posts and Telecommunications
Taotao Cheng
Nanjing University of Posts and Telecommunications
Siyu Hong
Nanjing University of Posts and Telecommunications
Wen Gao
Nanjing University of Posts and Telecommunications
Zixuan Lu
Nanjing University of Posts and Telecommunications
Liu Yang
Nanjing University of Posts and Telecommunications
Chen Ji
gisjc@nju.edu.cn

Nanjing University
Liang Cheng
Nanjing University

Article

Keywords: Unsupervised domain adaptation, Joint Color Space Transformation(JCST), High-resolution remote sensing images, Cross-domain object detection, Multi-model federation, Multi-scale object

Posted Date: May 30th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6642304/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

The rapid advancement of deep learning has led to significant achievements in remote sensing object detection. However, domain shift often causes notable performance drops when models trained on one domain are applied to real-world scenarios. Unsupervised domain adaptation (UDA) offers a solution by narrowing domain gaps. Generative adversarial networks (GANs) are commonly used for this purpose, but they can degrade key textures and details in source images. To address this, we propose a method that integrates transformations in both input and feature spaces. First, we standardize image dimensions across source and target domains. Then, a Joint Color Space Transformation (JCST) module operates in the feature space to decouple and recombine color channels, preserving crucial image details while aligning data distributions. We validated our approach on a dataset containing large-, medium-, and small-scale objects, using multiple object detection models. Results show that our method boosts average detection accuracy by 2–4% on source domain images, demonstrating improved generalization and robustness in cross-domain tasks.

1. Introduction

Progress in earth observation technology has led to greater accessibility of high-resolution remote sensing imagery, now obtainable more easily and at a lower cost [1–3]. We have now entered the era of big data in remote sensing, when automated picture interpretation is crucial for realizing its full potential [4, 5].

In recent years, there have been substantial breakthroughs in the application of deep learning approaches for the detection of objects using remote sensing [6, 7]. When suitable labeled image examples are available, it is possible to train object detection models based on deep learning in an integrated fashion, which has distinct advantages over conventional methods that rely on manually produced characteristics. Nonetheless, these deep learning methodologies are significantly reliant on substantial supervision (i.e., big, well labeled datasets) and necessitate analogous data distributions between training and testing sets. In contrast to natural photographs, remote sensing photos frequently display considerable diversity among several datasets due to changing imaging conditions, including time, location, and sensor specifications. Consequently, utilizing a model trained on source-domain images (i.e., the training set) directly on target-domain images (i.e., the test set) frequently results in a significant decline in performance [8]. Therefore, it is an immediate need to fix the disparity between the datasets for the source and target domains.

In order to lessen the impact of domain bias, unsupervised domain adaptation models aim to bring the distributions of features from the source and target domains into agreement [9, 10]. By projecting the characteristics of the two domains into a large-scale space and then reducing the difference between their distributions, approaches that use the maximum divergence of averages facilitate the transfer of characteristics [11]. On the other hand, adversarial learning-based methods use adversarial training to help the source and target domains' feature distributions become more comparable [12]. These algorithms encounter difficulties when used on high-resolution imagery from satellites, even if they have shown notable success in cross-domain object detection for natural images. Remote sensing images markedly differ from natural images owing to discrepancies in capture angles, temporal conditions, spatial resolutions, and display considerable intra- and inter-class variability. This variability sometimes causes the target-domain images generated by adversarial learning to lose important texture and feature information, resulting in differences from the source images. These differences create additional challenges for detecting tasks.

Prior research has primarily focused on cross-domain recognition of tiny targets, but large targets such as airports and runways in remote sensing imagery have received scant attention. As the target size enlarges, the detection frame

encompasses additional background information. A significant domain disparity between the source and target leads to diminished detection accuracy for larger targets.

This paper investigates the issues related to cross-domain object detection in remote sensing. To mitigate performance deterioration caused by uneven distributions across domains, we propose a joint domain adaptation strategy designed for cross-domain detection. To prevent the loss of detail and texture when transferring images from the source to the destination domain, we provide the JCST module for effective image information transfer. Moreover, extensively utilized remote sensing datasets, including DOTA[13], DIOR [7], and LEVIR [14], encompass varied data sources and comparable target scales. Nevertheless, datasets that provide consistent data sources and diverse target scales are uncommon. To rectify this deficiency, we created a novel dataset from uniform data sources while incorporating varied goal scales to enhance the assessment of our method's efficacy. A multitude of studies validates the efficacy of our methodology across various detection models.

The primary contributions of this paper include:

1. We present the first use case of the JCST module, which is for object detection in remote sensing using unsupervised domain adaptation. The JCST is incorporated into a generative adversarial network to facilitate picture transfer between domains, hence improving detection efficacy.
2. We present the DSOD dataset, which includes uniform data sources and multi-scale targets (large, medium, small) to evaluate the framework's resilience. The dataset is available at https://github.com/ChiYuIII/DSOD.

The subsequent sections of the paper are structured as follows: Section II addresses related work, Section III outlines the proposed method, Section IV provides experimental results and discussions, and Section V concludes the paper.

2. Related work

2.1. Unsupervised deep domain adaptation

The goal of unsupervised domain adaptation is to find a solution to the distributional inconsistencies that exist between the data from the source domain and the data from the target domain. This is a problem that is common across a variety of datasets. The main aim is to efficiently leverage shared features across these domains and implement them in tasks within the source domain [8]. Prior approaches in this field extracted domain-invariant features using the Maximum Mean Discrepancy (MMD) technique. With the growing integration of deep learning in domain adaptation research, unsupervised deep domain adaptation has become a prominent methodology. The methods can be classified into four distinct categories.

One significant method is domain adaptation using Generative Adversarial Networks (GANs), which aim to balance source and target domains through an adversarial formation that consists of a generator and a discriminator. Notable techniques, including CycleGAN [15], StarGAN [16], and GatedGAN [17], help the target domain's image styles to be transferred from the source domain.

Methods based on adversarial networks integrate adversarial strategies into the domain adaptation problem, enabling the trained network to differentiate between categories while also capturing domain-invariant features. Class-level adversarial networks, as introduced by [18], maintain semantic consistency locally. [19] Enhance adversarial learning by incorporating both strong and weak data enhancement techniques, as well as mutual learning within a teacher-student framework, thereby enabling the student model to acquire unique domain characteristics through self-learning.

The reconstruction-based approach utilizes a self-encoder to reduce information loss in the reconstruction of input data, thereby preserving the essential features of the original data. [20] To improve the extraction and refinement of shared characteristics, a domain separation network has been implemented to isolate specific information in source and target domains. [21] suggested a collaborative learning technique to find shared characteristics between synthetic and real images in the target domain at the pixel, region, and image levels.

Domain distribution differences are used to bridge the gap between domains to reduce generalization error in the target domain. [22] implemented an optimal multiple kernel selection strategy for mean embedding alignment, which significantly minimized domain discrepancies and facilitated the effective adaptation of deep convolutional neural networks in domain transfer contexts. These approaches have notably advanced research in domain adaptation, enhancing the development of the field.

2.2. Cross Domain Object Detection

Recent advancements in unsupervised domain adaptation techniques for object detection tasks have been significant. There are three basic types of cross-domain object identification techniques that use unsupervised domain adaptation: disparity-based techniques, adversarial-based techniques, and reconstruction-based techniques. [23] to reduce the discrepancy between formation and test data at the example and image level, we developed two adaptation modules based on the H-divergence theory. [24] utilized the CycleGAN model to improve detection performance by transforming datasets from daytime to nighttime, employing the generated images for training the detection model. [25] suggested a technique for unsupervised domain adaptation that combines less strict global alignment with strong local feature alignment. In a similar vein, [26] implemented consistency regularization within the Mean Teacher framework to mitigate domain bias during the transfer from synthetic to real images. In the same vein, [27] developed an automatic annotation framework that enhances the robustness of pedestrian detection via a dual-stream region proposal network. [28] generated high-quality synthetic images from 3D CAD models to mitigate data scarcity. [29] offered a way to generate images without human supervision with the goal of reducing domain inconsistencies in thermal and visual photography. [30] With ingenuity, the inter-domain reconnaissance problem has been solved by employing a customized dorsal network to extract local and global information, thereby enabling the recognition of small and closely clustered targets. While these methods have improved object detection in natural images, cross-domain remote sensing object detection continues to pose challenges due to the complex structure of remote sensing imagery.

2.3. Remote sensing cross domain object detection

Significant progress has been made in related domains in recent years, but cross-domain remote sensing object identification is still understudied and its adaptability potential is yet untapped. [31] presented a deep learning migration-based change detection method that pre-trains and fine-tunes source and target domain models to address poor adaption. [32] suggested a progressive migration-based unsupervised SAR ship detecting system, this enables a precise detection in non-labelled SAR images by aligning the characteristics of the source and target fields at the pixel level, as well as the characteristics and predictions. [33] introduced a semi-synthetic data generator for the production of remotely sensed data and developed domain-adapted detectors to address cross-domain recognition challenges in remote sensing images. Additionally, [34] suggested a multi-source change network that is domain-adaptive to address the shortcomings of adversarial domain adaptation methods, particularly the neglect of instance features. [35] network includes a data augmentation module SDA, a noise-reducing and feature-enhancing module SFR, and a PLG module that utilizes source domain knowledge to label unlabeled target domains, thereby mitigating domain bias in remote sensing object detection. These studies offer significant theoretical and methodological foundations for the advancement of cross-domain remote sensing object detection.

3. Methodology

This study's overall framework is depicted in Fig. 1. The first step is to perform a domain-supervised adaptation by employing the JCST_GAN model to convert the images via the source into the destination. After that, the related detection models are trained on a multi-scale target detection dataset. In the concluding phase, multiple detection models are utilized on the domain-adapted images to yield the final detection outcomes.

3.1. Multi-scale cross domain remote sensing object detection dataset

Current publicly available datasets exhibit significant variability in data sources, while target sizes tend to be consistent. We developed a remote sensing target detection dataset utilizing consistent data sources while varying target sizes to assess the precision and robustness of our method. The dataset comprises 14,640 Google tiles, each possessing a resolution of 1024 × 1024 pixels. Although the image dimensions remain constant, the spatial resolutions vary, resulting in significant geographic and scene diversity. Each subset was balanced and representative of the target categories and distributions by partitioning the dataset into training, validation, and test sets in a 6:2:2 ratio.

The dataset comprises four distinct target categories: storage tanks, airplanes, wind turbines, and airstrips. The targets were chosen based on significant size variations observed in remote sensing images. Storage tanks and airplanes are categorized as small targets, wind turbines as medium-sized targets, and airstrips as large-scale targets. This variety allows the dataset to more accurately represent complex real-world conditions, providing a thorough evaluation of the detection model's capabilities. Figure 2 displays sample images for each target type, while Table 1 outlines the quantity, spatial distribution, and associated Google tile levels of the dataset's targets. This configuration enables the assessment of model performance across various target sizes and environmental conditions.

Table 1
Data labeling details.

Label	Train	Validate	Test	Label box	Total	Google level	Scale
Storage	1415	470	472	36468	2357	18	Small
Airplane	1707	569	569	33379	2845	17	Small
Wind turbine	4449	1483	1483	29892	7415	17	Medium
Runway	1214	404	405	2806	2023	16	Large
Total	8785	2926	2929	-	14640	-	-

3.2. Unsupervised Domain Adaptation Based on Generative Adversity

Through finding invariant semantic characteristics, unsupervised techniques may efficiently translate pictures from the source domain to the destination domain, hence reducing domain differences. Unsupervised domain adaptation techniques, mostly utilizing Generative Adversarial Networks (GANs), are often categorized into two types: those necessitating paired images and those that do not. The former utilizes strictly paired images for supervised learning, exemplified by conditional GANs, initially developed for image-to-image translation with paired data. This approach is ill-suited for cross-domain remote sensing object detection problems. Conversely, the latter does not depend on paired images, rendering it more suitable for generalization to cross-domain remote sensing object detection

applications. CycleGAN minimizes dependence on image pairs by employing a cyclic consistency loss and adversarial treatment between the generator and discriminator to generate highly qualitative images that accurately reflect the target domain. Nonetheless, the varied acquisition conditions of remote sensing photos, including imaging modes, object scale, color saturation, and shooting angles, can lead to substantial domain-specific discrepancies. The interrelated challenges lead to significant stylistic variations across remote sensing photos from various domains. Consequently, the accuracy of target detection may be impacted by the loss of detail and texture information that may result from the application of style transfer.

We propose an improved CycleGAN model that maintains detail and texture during style transfer to resolve this issue. Furthermore, we do quantitative comparisons between our enhanced CycleGAN and other prominent unsupervised style transfer techniques, with experimental findings illustrating the better efficacy of our method in preserving visual details and texture.

3.3. Unsupervised Domain Adaptation Based on Generative Adversity

3.3.1. Input space alignment

Fixed input size in the model results in fluctuations in image resolution, which causes discrepancies in context and layout, hence complicating adversarial learning. To resolve this, we utilize uniform scaling and fixed-window cropping, as demonstrated in Fig. 3, to synchronize the resolution between the source and target domains. We attain contextual consistency for low-resolution images in the target domain by scaling them to correspond with the resolution of the source domain. We then crop the scaled photos with a fixed-size window to match the dimensions of the source domain images, ensuring layout consistency [36, 37]. When the target domain image possesses a greater resolution, employing fixed-window cropping on the source domain photos, in conjunction with scaling the target images, yields a comparable outcome.

$$\begin{cases} \varphi\left(\delta\left(\mathrm{x_T}, \backslash{:}\mathrm{r}\right), \backslash{:}\mathrm{S}\right) \mathrm{if} \mathrm{R_T} < \mathrm{R_S} \\ \delta\left(\mathrm{x_T}, \backslash{:}\mathrm{r}\right), \varphi\left(\mathrm{x_S}, \backslash{:}\mathrm{S_{x_T \cdot r}}\right) \mathrm{if} \mathrm{R_T} > \mathrm{R_S} \end{cases}$$

1

where δ and φ denote uniform scaling and fixed-window cropping, $\mathrm{x_T}$ and $\mathrm{x_S}$ denote the target and source images, respectively, $\mathrm{R_S}$ and $\mathrm{R_T}$ are the resolutions of the source and target images, respectively, S is the input size, r is the resizing factor, i.e., the ratio between $\mathrm{R_T}$ and $\mathrm{R_S}$ ratio, and $\mathrm{S_{x_T \cdot r}}$ is the size of $\mathrm{x_T}$ after resizing.

3.3.2. JCST Module

Remote sensing object detection typically utilizes high-resolution images that encompass abundant detailed information; however, CycleGAN is primarily engineered for style transfer of low- and medium-resolution images. The current network's depth and complexity are insufficient to fully capture all detailed information in high-resolution images. The generator is incapable of accurately capturing and reproducing intricate textures and edge details while producing high-resolution photos. This results in the loss of much detail and texture information in the moved image. To address the aforementioned issues and guarantee the complete preservation of detail and texture information from the source image during style migration with CycleGAN, we developed a Joint Color Space Transformation (JCST) module, as illustrated in Fig. 4.

The CycleGAN network uses style migration to convert the source domain image into the target domain image. Following the accomplishment of getting the migrated target domain picture, a color space conversion is carried out in order to change the image from the RGB color space to the YCbCr color space, which delineates the luminance information (Y channel) from the chrominance information (Cb and Cr channels). The Y, Cb, and Cr channels of the transferred target domain image are subsequently isolated. Luminance data (Y channel) and chrominance data (Cb and Cr channels) are extracted, and the Y channel from the original source domain image is merged with the Cb and Cr channels from the migrated target domain image. This method integrates luminance data from the source domain image to preserve the structure and details of the original image, while imparting the target domain image with the chromatic style derived from its chromaticity information. The resultant image preserves the details and structure of the source domain while integrating the stylistic attributes of the target domain, so producing a target domain image characterized by visual coherence and superior quality.

3.4. Experimental design

3.4.1. Task setup

In this research, we conducted remote sensing cross domain object detection ex-periments using the constructed multi-scale remote sensing object detection dataset (DSOD). Specifically, we take the DSOD dataset as the target domain data and select the Gaofen-2 image of a region as the source domain data. The Gaofen-2 image under-goes preprocessing to provide RGB three-channel data, which is subsequently cropped to match the dimensions of the DSOD dataset. We conducted three experimental sets for object detection employing a multi-model detection strategy:

1. 1. Remote sensing object detection directly on images in the source domain;
2. 2. Remote sensing object detection on images post-style migration via CycleGAN;
3. 3. Using our suggested technique for distant sensing object detection on images after style migration.PyTorch is used for each and every experiment to be carried out in this investigation. The model training is expedited utilizing the NVIDIA RTX 4080, which possesses 24GB of graphics RAM.

3.4.2. Evaluation of the indexes

We propose to use Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) to evaluate the quality of the images acquired using the proposed method.

$$\mathrm{PSNR}=10\bullet\log_{10}(\frac{\mathrm{MAX}^2}{\mathrm{MSE}})$$

2

$$\mathrm{SSIM}(\mathrm{x},\backslash:\mathrm{y})=\frac{(2\mu_{\mathrm{x}}\mu_{\mathrm{y}}+\mathrm{C}_1)(2\sigma_{\mathrm{xy}}+\mathrm{C}_2)}{(\mu_{\mathrm{x}}^2+\mu_{\mathrm{y}}^2+\mathrm{C}_1)(\sigma_{\mathrm{x}}^2+\sigma_{\mathrm{y}}^2+\mathrm{C}_2)}$$

3

Where MAX is the maximum possible pixel value of the image, MSE is the mean square error, μ_{x} and μ_{y} are the mean values of the image x and y, σ_{x}^2 and σ_{y}^2 are the variances of the image x and y, σ_{xy} is the covariance of the image x and y, and C_1 and C_2 are constants used for stabilization calculation.

We assessed the detection model's performance in our tests using five traditional evaluation metrics obtained from the confusion matrix: accuracy, recall, F1 score, and mAP value. Accuracy assesses the model's capacity to forecast positive instances, recall evaluates the model's proficiency in properly identifying positive instances, and the F1 score is the harmonic mean of accuracy and recall, offering a balanced evaluation of the model's performance. mAP, conversely, indicates the comprehensive performance of the model across all areas.

$$\mathrm{Precision}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$$

4

$$\mathrm{Precision}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$$

5

$$\mathrm{F}_1=\frac{2\times\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}}$$

6

$$\mathrm{mAP}=\frac{1}{\mathrm{n}}\sum{}_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{AP}_{\mathrm{i}}$$

7

TP , FP, and FN are the number of true cases, false positive cases, and false negative cases respectively, n denotes the number of categories, and AP_{i}denotes the AP value of the ith category.

3.4.3. Multi-modal detection

A multi-model detection strategy is proposed, aiming to synthesize high-quality images and multiple types of object detection models generated by the improved CycleGAN through an integrated learning approach, for the purpose of verifying the proposed method's accuracy and robustness in the detection system.

First, we normalize and resize the input images to ensure that they can adapt to the input requirements of each target detection model and enhance the model's flexibility in adjusting to various settings and illumination. Subsequently, the optimized CycleGAN network is utilized for image style migration to generate higher quality and more detailed target images. The JCST module, which is incorporated into the upgraded CycleGAN, efficiently maintains the target objects' features and edge information while also greatly improving the output images' clarity and realism.

Next, the generated high-quality images are input into the following four different types of object detection models for detection, respectively. Single-stage detection model: directly predicts the locations and classes of all targets in the image through a single forward propagation[38]. Models based on deformed convolutional and Transformer architectures: combining deformed convolutional networks and Transformer mechanisms, they can effectively capture shape changes and contextual information of targets, and are suitable for target detection and pose estimation in complex scenes[39]. A two-stage model based on region proposition network (RPN): creates potential target regions first, after which these regions are subjected to categorization and bounding box regression.[40]. Fully convolutional one-stage model: the location and class of targets are directly predicted by a fully convolutional network on feature maps at different scales[41]. These techniques are employed to confirm that the enhanced CycleGAN's high-quality pictures may be applied to multi-model detection tasks. This strategy not only brings substantial

performance improvement for target detection tasks in complex scenes, but also demonstrates the potential and practical application value of the improved CycleGAN in remote sensing image processing applications.

4. Experimental results and Discussion

4.1. Unsupervised domain adaptation results

The aim of this research is to improve target detection through style migration with the advanced CycleGAN. The source domain comprises the preprocessed Gaofen-2 images, whereas the target domain consists of the image from the DOSD dataset. Table 2 presents the quantitative experimental data, whereas Fig. 5 illustrates the visualization outcomes. The CycleGAN generator employs a U-Net-like architecture, which enhances its ability to preserve global information; however, it exhibits shortcomings in processing high-frequency image details, particularly during the downsampling and upsampling phases, leading to a loss of detail and texture information. This negatively impacts the future target detection. To reduce the loss of texture and detail in produced images, we integrate the JCST module into the CycleGAN network, aiming to modify just the image style while retaining enhanced detail information.

Experiments show that the upgraded CycleGAN outperforms the original in structural similarity (SSIM) and peak signal-to-noise ratio (PSNR), while the JCST module improves images. In the visualization findings, our method excels in retaining high-frequency information, and the generated images closely resemble the original images, offering more dependable inputs for the target recognition model. While our enhanced method alleviates the issue of high-frequency information loss to a degree, it remains incapable of completely restoring all details in very intricate texture areas. Furthermore, despite attaining notable enhancements in SSIM and PSNR, inconsistencies may persist between the generated image and the target domain image in other quantitative measures (e.g., color and spatial consistency). This may influence the efficacy of target detection across various application contexts; hence, subsequent study should investigate in greater depth how to regulate the extent of style migration to maintain semantic coherence within the produced image and the target domain image.

Table 2
Comparison of image of quality generated using CycleGAN and JCST_GAN.

	CycleGAN	JCST_GAN
SSIM	0.85	0.99
PSNR	23.99	30.82

In addition, the introduction of the JCST module increases the complexity and training time of the model, especially when processing high-resolution images, and the demand for computational resources increases significantly, which may limit its application in large-scale datasets or real-time applications. Future exploration of model trimming or parameter optimization may further decrease computing expenses. And different detection tasks and datasets may on the effect of generated images, so validating the generalization of the method in a wider range of future research will continue to focus on application scenarios. CycleGAN is chosen as the base model for style migration mainly based on its wide application and excellent performance in image translation tasks. Compared with other mainstream models, CycleGAN achieves high-quality image style migration without the need for pairs of training data, which makes it outstanding in diverse datasets and unsupervised environments. From the results in Table 3, compared to other popular models, CycleGAN produces images of noticeably higher quality., which further validates its advantages in the style migration task. In addition, the architecture of CycleGAN is highly flexible and easy to combine with other

models. For example, the JCST module introduced in this study is an example, which realizes a significant improvement in the detail fidelity of the generated images by seamlessly combining with CycleGAN.

Table 3
Comparison of image quality generated by different models.

	DualGAN	DiscoGAN	CycleGAN
SSIM	0.74	0.69	**0.82**
PSNR	17.63	14.83	**21.36**

4.2. Multi-model object detection results

The performance of many object identification models is compared and analyzed in this part, with an emphasis on the outcomes before and after the suggested approach was used. Figure 6 presents the results, while Table 4 illustrates the detection outcomes for targets including wind turbines, storage tanks, airplanes, and airport runways. The comparison indicates that the style-migrated images outperform the original images in the object detection task, with an improvement in the mAP of 2%-4%. The findings show that the suggested CycleGAN augmentation technique effectively migrates picture styles while simultaneously improving the object identification model's performance, particularly in situations involving many targets, showing significant advantages.

In terms of specific model performance, YOLOv8 and FCOS show a more comprehensive improvement in all object detection tasks, with F1 scores improved by 2%-6%. This suggests that style migration enhances the detection capability of these models, especially in complex multi-target scenarios.The RE-DETR model also exhibits F1 score enhancement in the detection of wind turbines, airplanes, and airport runways, but the F1 scores are unchanged for the detection of storage tank targets due to the fact that while the accuracy improves, the decrease in the recall rate offsets this advantage. On the contrary, the F1 score of Faster R-CNN decreases in the detection of wind motors, mainly due to the decrease in recall after style migration, and the model fails to recognize some of the wind motor targets, but its accuracy improves, indicating that the style-migrated photos more closely resemble the model training set data distribution.

Despite its strong overall performance, the suggested approach has several drawbacks. First, the style migration is not effective enough to detect certain targets (e.g., storage tanks), and there is also a decrease in recall on some models (e.g., Faster R-CNN), suggesting that different models are differently adapted to style-migrated images. Future research can perform personalized optimization for different detection models to enhance the applicability of style migration. Second, in extremely complex scenes, style migration may still not be able to sufficiently improve the contrast between target and background, It implies that in order to further boost the detection performance in intricate scenarios, we can include further picture enhancing techniques or intricate multi-scale feature extraction procedures.

Table 4
Proposed methodology and baseline methodology object detection results.

	Wind turbine			Storage			Airplane			Runway			
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	mAP
YOLO V8	0.86	0.88	0.87	0.57	1.00	0.72	0.60	1.00	0.75	0.46	0.81	0.59	0.92
JCST_YOLO V8	0.81	0.97	**0.89**	0.62	1.00	**0.77**	0.67	1.00	**0.81**	0.54	0.83	**0.65**	**0.95**
RE-DETR	0.81	0.95	0.88	0.63	0.99	0.77	0.36	0.92	0.52	0.86	0.77	0.81	0.91
JCST_RE-DETR	0.86	1.00	**0.93**	0.62	1.00	0.77	0.44	0.95	**0.60**	0.86	0.80	**0.83**	**0.94**
Faster R-CNN	0.86	0.97	0.91	0.94	0.42	0.58	0.95	0.96	0.95	0.64	0.90	0.75	0.81
JCST_Faster R-CNN	0.72	1.00	0.84	0.96	0.42	**0.59**	0.95	0.98	**0.96**	0.64	1.00	**0.78**	**0.85**
FCOS	0.84	0.88	0.86	0.90	0.93	0.91	0.94	0.96	0.95	0.57	0.70	0.63	0.87
JCST_FCOS	0.84	0.90	**0.87**	0.92	0.95	**0.94**	0.95	0.99	**0.97**	0.57	0.73	**0.64**	**0.89**

4.3. Ablation experiment

We created and carried out a number of comparison tests to evaluate the efficacy of the suggested approach. First, the images designated for detection were input directly into the target detection model without any style migration. Second, the images underwent style migration using the traditional CycleGAN, and the resulting images were then fed into the object detection model for recognition. Finally, we employed the improved CycleGAN, which includes the JCST module, to perform style migration on the images prior to target detection. Table 5 provides a summary of these trials' findings.

The data presented in the table indicates that the F1 scores and mAP values for object detection using images generated by the traditional CycleGAN are lower than those obtained from the original images. This can primarily be attributed to the significant loss of detail and texture information in the CycleGAN-generated images compared to the originals, which hampers the detection model’s ability to effectively extract key features of the targets, thereby affecting class discrimination and leading to decreased detection performance. However, the implementation of our proposed improved CycleGAN, enhanced by the JCST module, leads to a noticeable enhancement in the transferred images' object detection ability. Notably, for the Faster R-CNN model, the mAP value improved by 4%., with similar improvements observed across all other models. This suggests that the JCST module effectively mitigates the loss of image detail and texture, enhancing the applicability of the generated images to the target detection models and consequently improving detection accuracy.

Table 5
Quantitative evaluation of ablation experiments.

	Wind turbine	Storage	Airplane	Runway	
	F_1	F_1	F_1	F_1	mAP
YOLO V8	0.87	0.72	0.75	0.59	0.92
CycleGAN + YOLO V8	0.72	0.56	0.61	0.45	0.81
JCST_YOLO V8	0.89	0.77	0.81	0.65	0.95
RE-DETR	0.88	0.77	0.52	0.81	0.91
CycleGAN + RE-DETR	0.74	0.68	0.46	0.69	0.79
JCST_RE-DETR	0.93	0.77	0.60	0.83	0.94
Faster R-CNN	0.91	0.58	0.95	0.75	0.81
CycleGAN + Faster R-CNN	0.81	0.47	0.84	0.69	0.71
JCST_Faster R-CNN	0.84	0.59	0.96	0.78	0.85
FCOS	0.86	0.91	0.95	0.63	0.87
CycleGAN + FCOS	0.79	0.83	0.89	0.55	0.82
JCST_FCOS	0.87	0.94	0.97	0.64	0.89

5. Conclusions

Due to the rapid advancements in artificial intelligence and object detection technology, many object detection tasks now employ deep learning techniques. However, remote sensing images are sourced from various origins, and even images taken by the same sensor may exhibit inconsistent data distributions due to differences in acquisition times and imaging angles. This disparity between the training dataset and the detection model's accuracy is greatly impacted by the images that need to be detected. We propose a method to harmonize the data distributions between the model's training set and the target images in order to resolve this issue. This technique is then applied to real-world detection tasks.

In particular, we incorporate a JCST module into the CycleGAN framework to effectively preserve the details and texture of the images during the style transfer process, thereby improving the quality of the resultant images. The experimental results suggest that our method can enhance the accuracy of target detection by 2–4%.

Moreover, to evaluate the effectiveness of target detection across various scales, we have developed a multi-scale target detection dataset that considers differences in target sizes. This dataset encompasses targets of three scales: large, medium, and small, with all images originating from the same source. The final experimental results indicate that our proposed method not only substantially improves the performance of various detection models but also improves the accuracy of detection for targets of varying sizes. The integration of the domain adaptation model with the target detection model will be the primary focus of future research in order to enhance the model's efficacy and achieve a higher level of automation.

Declarations

Data availability

All data generated or analyzed during this study are included in this published article.

Author Contributions: Conceptualization, D.C. and T.C.; methodology, D.C., T.C. and S.H.; validation, W.G., Z.L. and L.Y.; data curation, S.H.; writing—original draft preparation, D.C. and T.C.; writing—review and editing, S.H.; project administration, L.C.; funding acquisition, C.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number No. 42401548.

Competing interests

The authors declare no competing interests.

Consent to participate

All authors voluntarily agree to participate in this research study.

Consent to publish

All authors voluntarily approved the publication of this research study.

References

1. Chen, Z. et al. Joint alignment of the distribution in input and feature space for cross-domain aerial image semantic segmentation. *International J. Appl. Earth Observation Geoinformation* **115** (2022).
2. Zhang, L., Zhang, L. & Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience remote Sens. magazine*. **4**, 22–40 (2016).
3. Chi, M. et al. Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE* 104, 2207–2219 (2016).
4. Sun, X. et al. From single-to multi-modal remote sensing imagery interpretation: A survey and taxonomy. *Sci. China Inform. Sci.* **66**, 140301 (2023).
5. Li, Y. et al. Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation. *ISPRS J. Photogrammetry Remote Sens.* **175**, 20–33 (2021).
6. Zhang, X. et al. Remote sensing object detection meets deep learning: A metareview of challenges and advances. *IEEE Geoscience Remote Sens. Magazine* (2023).
7. Li, K., Wan, G., Cheng, G., Meng, L. & Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. photogrammetry remote Sens.* **159**, 296–307 (2020).
8. Oza, P., Sindagi, V. A., Sharmini, V. V. & Patel, V. M. Unsupervised domain adaptation of object detectors: A survey. *IEEE Trans. Pattern Anal. Mach. Intelligence* (2023).
9. Shi, Y., Du, L., Li, C., Guo, Y. & Du, Y. Unsupervised domain adaptation for SAR target classification based on domain-and class-level alignment: From simulated to real data. *ISPRS J. Photogrammetry Remote Sens.* **207**, 1–13 (2024).
10. Wilson, G. & Cook, D. J. A survey of unsupervised deep domain adaptation. *ACM Trans. Intell. Syst. Technol. (TIST)*. **11**, 1–46 (2020).

11. Yan, H. et al. Weighted and class-specific maximum mean discrepancy for unsupervised domain adaptation. *IEEE Trans. Multimedia*. **22**, 2420–2433 (2019).
12. Benjdira, B., Bazi, Y., Koubaa, A. & Ouni, K. Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images. *Remote Sens.* **11**, 1369 (2019).
13. Xia, G. S. et al. in *Proceedings of the IEEE conference on computer vision and pattern recognition.* 3974–3983.
14. Zou, Z. & Shi, Z. Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images. *IEEE Trans. Image Process.* **27**, 1100–1111 (2017).
15. Chu, C., Zhmoginov, A. & Sandler, M. Cyclegan, a master of steganography. *arXiv preprint arXiv:1712.02950* (2017).
16. Choi, Y. et al. in *Proceedings of the IEEE conference on computer vision and pattern recognition.* 8789–8797.
17. Chen, X., Xu, C., Yang, X., Song, L. & Tao, D. Gated-gan: Adversarial gated networks for multi-collection style transfer. *IEEE Trans. Image Process.* **28**, 546–560 (2018).
18. Luo, Y., Zheng, L., Guan, T., Yu, J. & Yang, Y. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 2507–2516.
19. Li, Y. J. et al. Cross-domain object detection via adaptive self-training. *CoRR* (2021).
20. Tsai, J. C. & Chien, J. T. in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP).* 1–6 (IEEE).
21. Sun, R. et al. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 4360–4369.
22. Long, M., Cao, Y., Wang, J. & Jordan, M. in *International conference on machine learning.* 97–105 (PMLR).
23. Chen, Y., Li, W., Sakaridis, C., Dai, D. & Van Gool, L. in *Proceedings of the IEEE conference on computer vision and pattern recognition.* 3339–3348.
24. Arruda, V. F. et al. in *2019 International Joint Conference on Neural Networks (IJCNN).* 1–8 (IEEE).
25. Saito, K., Ushiku, Y., Harada, T. & Saenko, K. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 6956–6965.
26. Cai, Q. et al. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 11457–11466.
27. Cao, Y. et al. Pedestrian detection with unsupervised multispectral feature learning using deep neural networks. *Inform. fusion*. **46**, 206–217 (2019).
28. Liu, W., Luo, B. & Liu, J. Synthetic data augmentation using multiscale attention CycleGAN for aircraft detection in remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **19**, 1–5 (2021).
29. Liu, P., Li, F., Yuan, S. & Li, W. Unsupervised Image-Generation Enhanced Adaptation for Object Detection in Thermal Images. *Mobile information systems* 1837894 (2021). (2021).
30. Biswas, D. & Tešić, J. Domain adaptation with contrastive learning for object detection in satellite imagery. *IEEE Trans. Geoscience Remote Sensing* (2024).
31. Yang, M., Jiao, L., Liu, F., Hou, B. & Yang, S. Transferred deep learning-based change detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **57**, 6960–6973 (2019).
32. Shi, Y., Du, L., Guo, Y. & Du, Y. Unsupervised domain adaptation based on progressive transfer for ship detection: From optical to SAR images. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–17 (2022).
33. Xu, T. et al. Feature aligned domain adaptive object detection in remote sensing imagery. *IEEE Trans. Geoscience Remote Sens. 60*. **FADA**, 1–16 (2022).

34. Zhang, C. et al. A domain adaptation neural network for change detection with heterogeneous optical and SAR remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **109**, 102769 (2022).
35. Zhu, Y., Sun, X., Diao, W., Li, H. & Fu, K. RFA-Net: Reconstructed feature alignment network for domain adaptation object detection in remote sensing imagery. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* **15**, 5689–5703 (2022).
36. Hashemzadeh, M., Asheghi, B. & Farajzadeh, N. Content-aware image resizing: an improved and shadow-preserving seam carving method. *Sig. Process.* **155**, 233–246 (2019).
37. Asheghi, B., Salehpour, P., Khiavi, A. M. & Hashemzadeh, M. A comprehensive review on content-aware image retargeting: From classical to state-of-the-art methods. *Sig. Process.* **195**, 108496 (2022).
38. Hussain, M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. *Machines* **11**, 677 (2023).
39. Zhao, Z. et al. RT-DETR-Tomato: Tomato Target Detection Algorithm Based on Improved RT-DETR for Agricultural Safety Production. *Appl. Sci.* **14**, 6287 (2024).
40. Ren, S., He, K., Girshick, R., Sun, J. & Faster, R-C-N-N. Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2016).
41. Tian, Z., Shen, C., Chen, H. & He, T. F. C. O. S. A simple and strong anchor-free object detector. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 1922–1933 (2020).

Figures

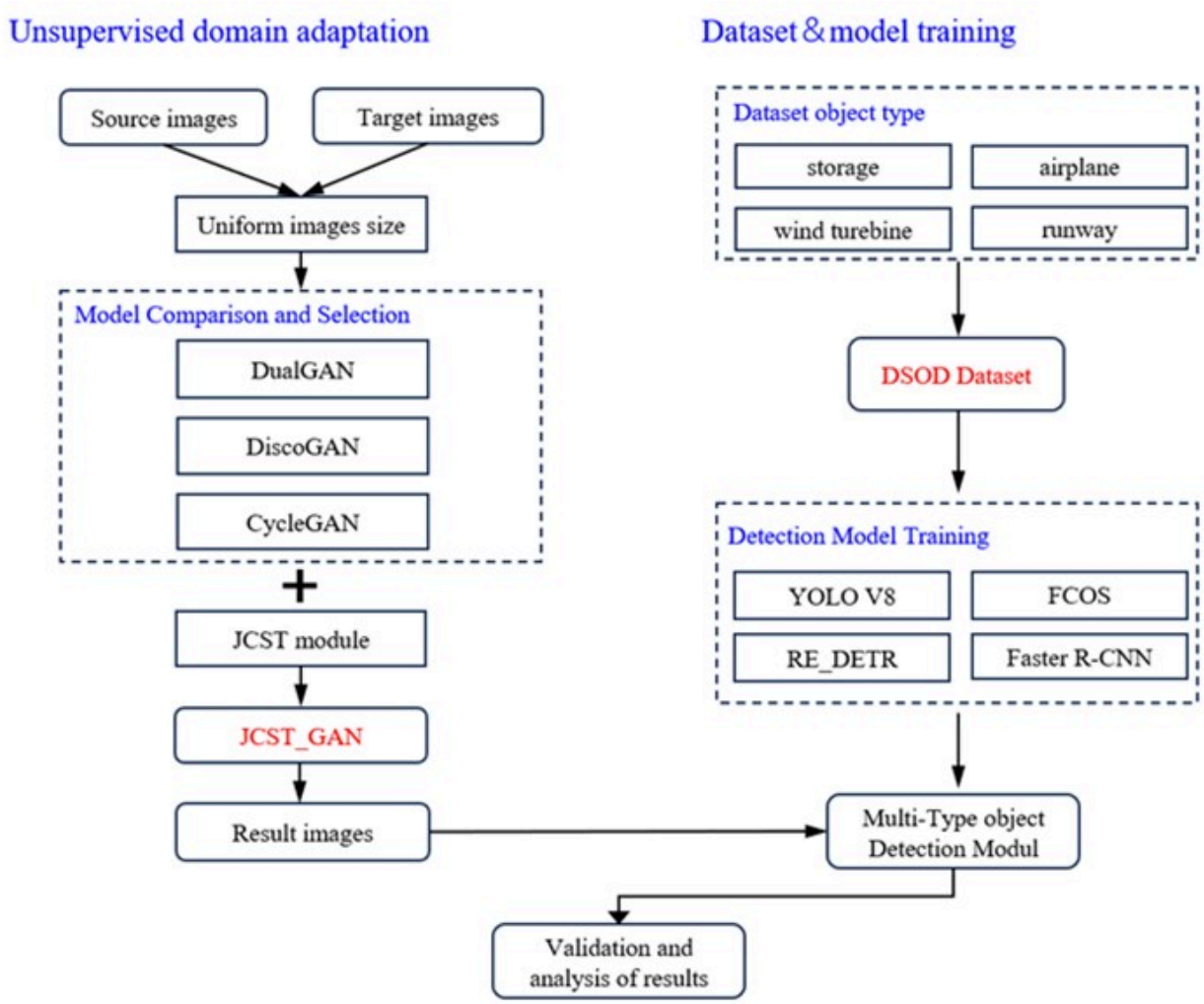


Figure 1

Overall framework diagram.

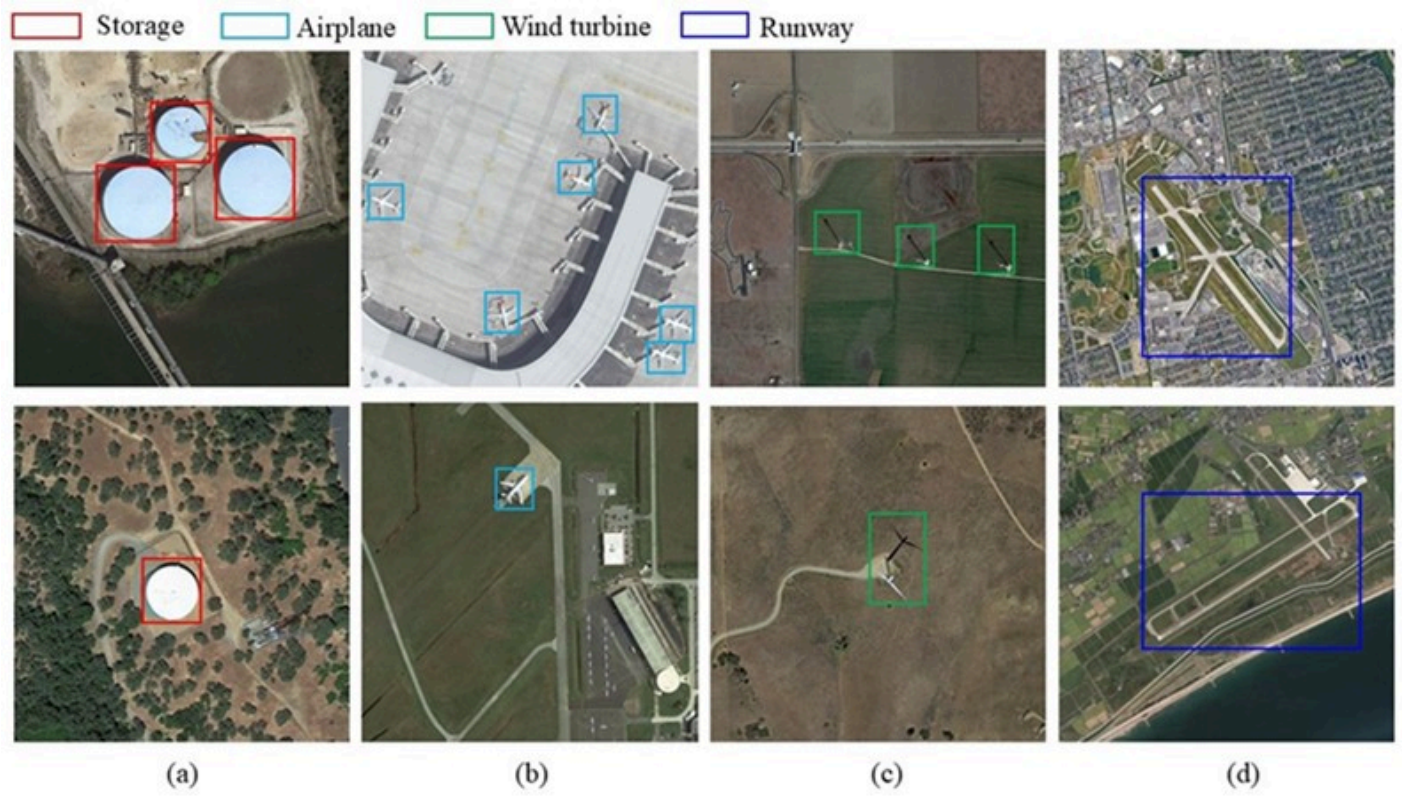


Figure 2

Sample labeling schematic.

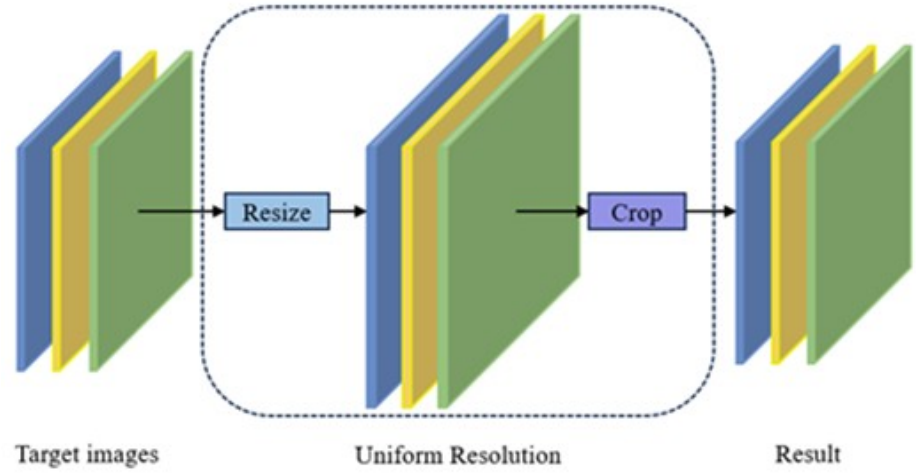


Figure 3

Uniform image resolution.

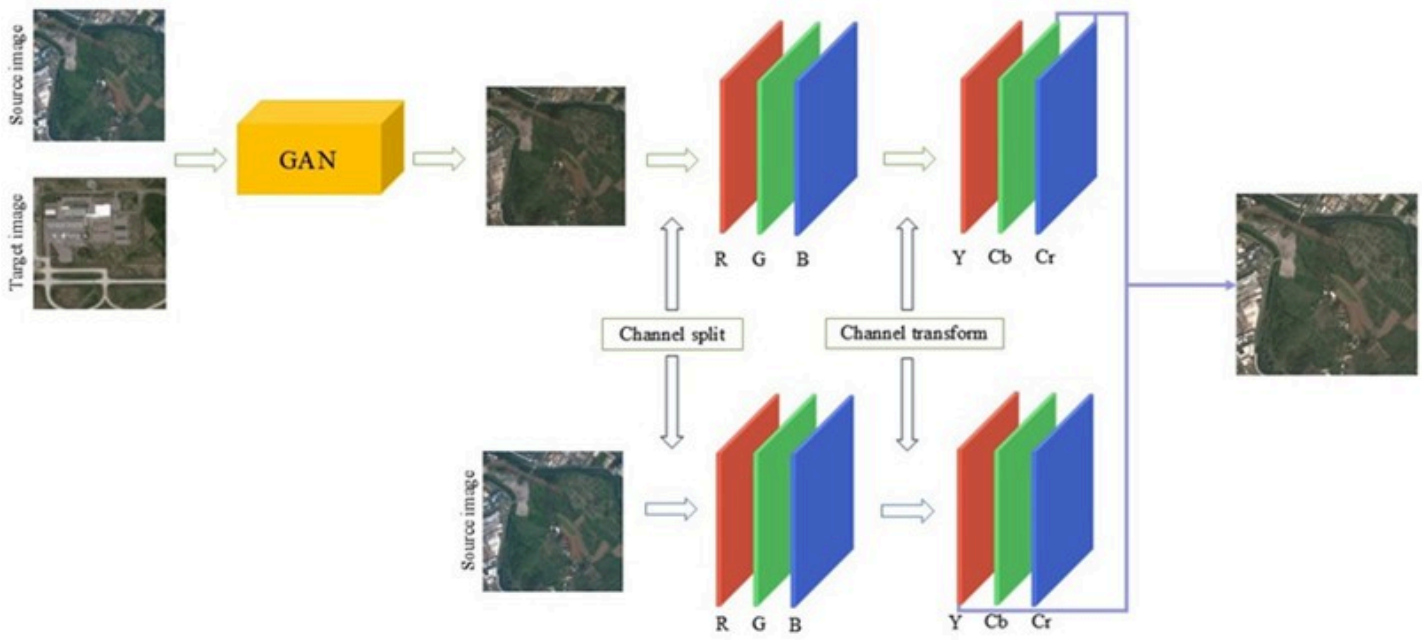


Figure 4

JCST module working diagram.

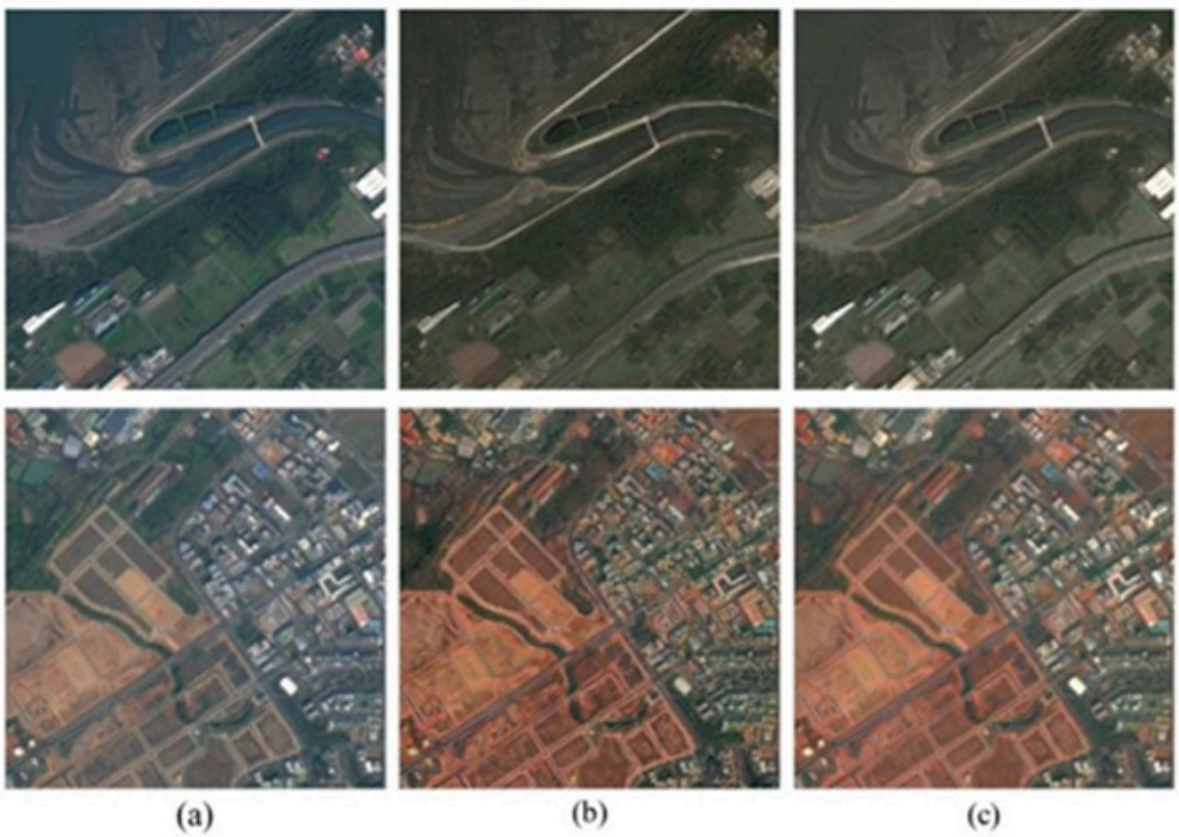


Figure 5

Comparison of image visualization before and after domain migration (a) represents the original image (b) represents the image after migration using CycleGAN (c) represents the image after migration using the proposed method.

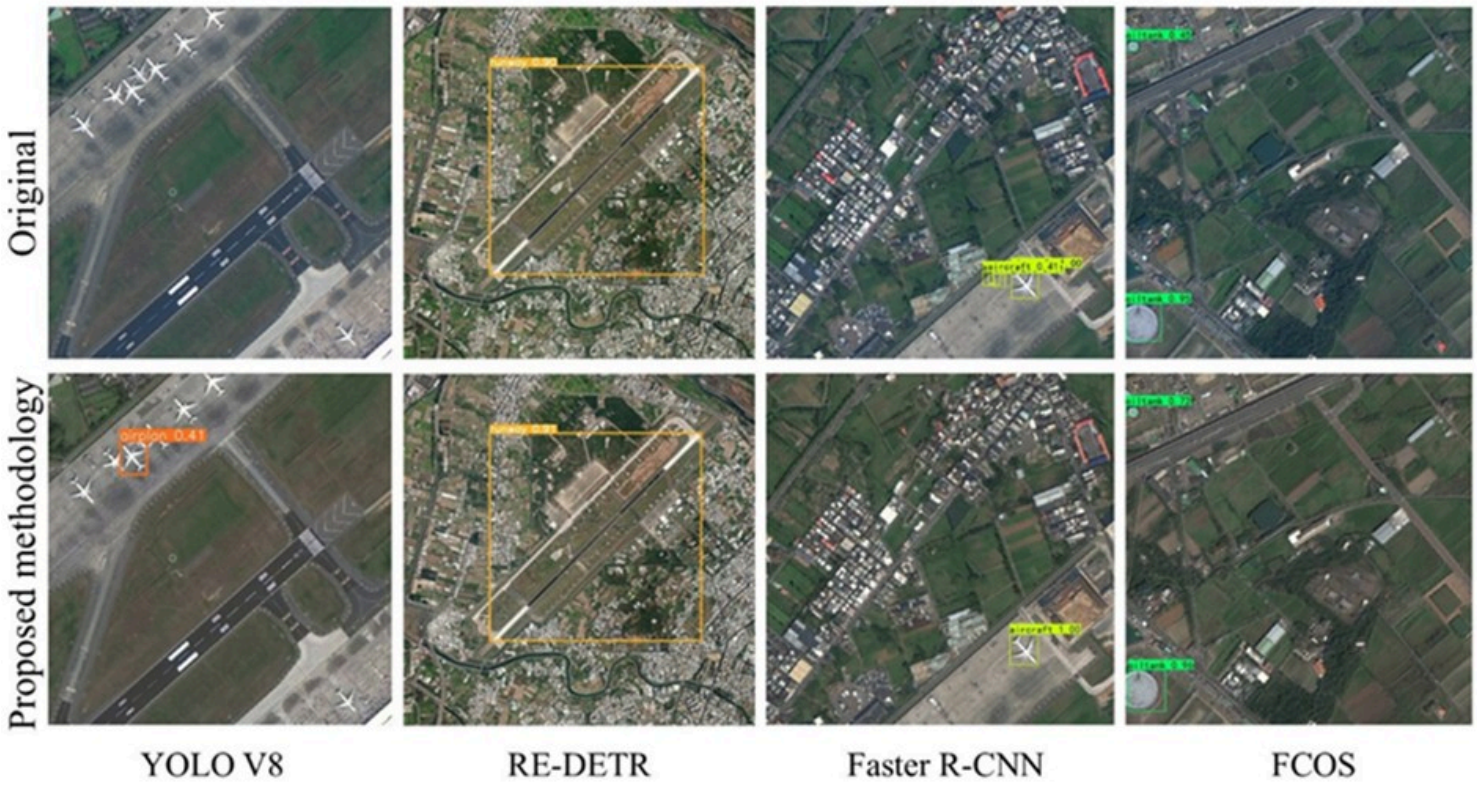


Figure 6

Detection results of different models before and after utilizing the proposed method.