Article

Unsupervised Domain-Adaptive SAR Ship Detection Based on Cross-Domain Feature Interaction and Data Contribution Balance

Yanrui Yang [1,2], Jie Chen [1,2,*], Long Sun [1], Zheng Zhou [3], Zhixiang Huang [2] and Bocai Wu [1]

[1] 38th Research Institute of China Electronics Technology Group Corporation, Hefei 230088, China; 15979690595@163.com (Y.Y.); sl99goal@163.com (L.S.); 18110995593@189.cn (B.W.)
[2] Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, School of Electronics and Information Engineering, Anhui University, Hefei 230601, China; zxhuang@ahu.edu.cn
[3] Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System, College of Electronic Science and Technology, National University of Defense Technology (NUDT), Changsha 410073, China; zhouzheng19961123@163.com
[*] Correspondence: author: jiechen@ustc.edu

Abstract: Due to the complex imaging mechanism of SAR images and the lack of multi-angle and multi-parameter real scene SAR target data, the generalization performance of existing deep-learning-based synthetic aperture radar (SAR) image target detection methods are extremely limited. In this paper, we propose an unsupervised domain-adaptive SAR ship detection method based on cross-domain feature interaction and data contribution balance. First, we designed a new cross-domain image generation module called CycleGAN-SCA to narrow the gap between the source domain and the target domain. Second, to alleviate the influence of complex backgrounds on ship detection, a new backbone using a self-attention mechanism to tap the potential of feature representation was designed. Furthermore, aiming at the problems of low resolution, few features and easy information loss of small ships, a new lightweight feature fusion and feature enhancement neck was designed. Finally, to balance the influence of different quality samples on the model, a simple and efficient $E^{\frac{1}{2}}IoU$ Loss was constructed. Experimental results based on a self-built large-scale optical-SAR cross-domain target detection dataset show that compared with existing cross-domain methods, our method achieved optimal performance, with the mAP reaching 68.54%. Furthermore, our method achieved a 6.27% improvement compared to the baseline, even with only 5% of the target domain labeled data.

Keywords: deep learning; SAR; object detection; unsupervised domain adaptation

1. Introduction

Synthetic aperture radar (SAR) is an active microwave imaging radar that can acquire high-resolution radar images comparable to optical images, even in extremely harsh weather conditions. It offers the advantage of all-day and all-weather earth observation, as it is not affected by light or climatic conditions. In recent years, SAR imaging technology has advanced rapidly, leading to improved SAR image resolution and a greater amount of data. Consequently, object detection technology for SAR images has emerged as a prominent area of research. Its application potential extends beyond civil sectors such as agriculture, forestry, water management, geology, and natural disaster assessment, as it also holds significant research value in the military domain. Notably, ship detection in SAR images has garnered growing interest from both domestic and international researchers.

Target detection is a basic task in SAR image interpretation. Due to the complex imaging mechanism of SAR images, the target information in SAR images is mainly reflected in the scattering point composition, which leads to the challenge of SAR image target detection. Before the rise of deep learning algorithms, the target detection of SAR images

by researchers could be roughly divided into structure-based methods [1–3], grayscale feature-based methods [4–6] and texture feature-based methods [7,8]. These traditional detection methods can achieve good results in specific scenes, but when faced with complex scenes and multiscale targets, the performance of traditional methods is greatly affected.

With the emergence of deep learning algorithms, convolutional neural networks (CNNs) have gained significant attention and have been extensively utilized in various domains, including remote sensing. In particular, CNNs are increasingly being applied to ship detection tasks based on SAR images. Among the numerous approaches in this field, YOLO [9] and Faster R-CNN [10] are highly regarded and widely adopted for their superior accuracy, efficiency, and robustness. They are the representatives of the single-stage detection algorithm and two-stage detection algorithm.

However, CNN-based detection methods heavily rely on a large quantity of labeled SAR ship data, and in real application scenarios, due to the limitations of target orientation, imaging geometry, ship shape and construction materials, multi-angle and multi-parameter SAR ship data are seriously lacking. In recent years, although some organizations have successively released AIR-SARShip [11], HRSID [12], SSDD [13] and other SAR ship target detection datasets through the processing of SAR image data collected by SAR satellites and aircraft, these datasets have a single scenario and a limited number of samples, which cannot meet the needs of practical applications. Moreover, due to the unique imaging mechanism and complex electromagnetic wave scattering mechanism of SAR images, the ship targets in SAR images can only be interpreted by trained experts, making the creation of large-scale SAR ship image datasets a costly and time-consuming endeavor. As a result, SAR ship detection in SAR images still faces significant challenges.

In recent years, domain-adaptive approaches have evolved rapidly, which is a method of migrating models from the source domain to the target domain by learning the differences between the two. It solves the problems of incomplete, unbalanced and inaccurate labeling of datasets due to high data acquisition and labeling costs in practical applications and improves the generalization ability and performance of the model. The proposal of domain adaptation provides an idea for ship detection in SAR images. In real-world scenarios, it is difficult to acquire SAR ship data from multiple angles and parameters, and the collection of SAR ship image data is expensive. To address this issue, incorporating optical image data to assist SAR image target detection is beneficial. This not only leverages the advantages of rich image information in optical data but also solves the problem of limited SAR image data availability. In practical applications, the domain-adaptive method [14–17] has achieved good results in the object detection task in the natural scene. Inspired by the research of domain adaptation in natural scenes, research on domain-adaptive methods [18–23] has emerged in the field of remote sensing, but these methods are mainly for classification and segmentation tasks, not object detection. In recent years, there has also been initial progress in domain-adaptive object detection methods [24–29] for remote sensing images, which are usually improved on the classical domain-adaptive methods and do not explore the information characteristics of remote sensing images themselves. Moreover, there is a significant difference in information between different remote sensing images, especially between optical and SAR images. These large domain differences may impact the performance of domain adaptation methods in SAR image object detection. Furthermore, the special imaging mechanism of SAR images poses challenges for commonly used object detection algorithms, as they struggle to effectively extract features from SAR targets. Consequently, it is difficult to meet the performance requirements for ship detection in practical applications. Lastly, most existing domain-adaptive methods are based on two-stage detectors. However, for resource-limited and time-critical SAR image ship inspection tasks, we prefer to use single-stage detectors with high real-time performance and accuracy.

In summary, it is valuable and meaningful to design a ship detection algorithm for SAR images with high accuracy and high real-time performance from the optical field to the SAR field. In this paper, a new domain-adaptive SAR ship detection method based on

cross-domain feature interaction and data contribution balance is proposed, which realizes the target detection task of SAR ship images by deeply mining and fusing global and local feature information and combining new loss functions. We take the optical ship image containing rich information as the source domain and the SAR ship image as the target domain, where the source domain data are fully labeled and the target domain data are not labeled. Then, we introduce a new cross-domain image generation module, CycleGAN-SCA, to enable the generation of pseudo-images for both optical and SAR domains. Based on the knowledge distillation framework, the mean teacher (MT) model [30] is used to guide the teacher network to carry out the object detection task, realize the detection of unlabeled target domain images and generate pseudo-labels, and achieve a relatively unbiased update of the student network. Specifically, the contributions of our proposed method in the field of remote sensing are as follows:

(1) A cross-domain image generation module from the optical to SAR domain (referred to as CycleGAN-SCA) is designed. By making full use of the explicit information of the target in the optical image, the module realizes unbiased feature generation from the optical to SAR domain, which solves the problem of low performance caused by insufficient annotation data in SAR image ship detection and significantly improves the performance of SAR ship detection without increasing the annotation cost.

(2) To mitigate the influence of the complex background of SAR images on ship detection and to ensure that the model can still perform well under the condition of high-density targets, a new self-attention feature extraction backbone network (abbreviated as SFE) is designed. It can effectively capture global information and rich context information, which improves the performance of the model for SAR image ship detection tasks.

(3) Aiming at the problems of low resolution, few features, and easy information loss of small-sized ships, a new lightweight feature fusion and feature enhancement neck (referred to as LFE) is designed, and more efficient upsampling modules and context aggregation modules are introduced to learn the global spatial content and improve the detection effect of small-sized ship targets.

(4) A new simple and efficient $E^{\frac{1}{2}}IoU$ loss is constructed. $E^{\frac{1}{2}}IoU$ loss has the characteristics of balancing the contribution of high-quality samples and low-quality samples to loss, which strengthens the contribution of low-quality samples to loss, and is more suitable for ship target detection in SAR images based on unsupervised domain-adaptive detection.

2. Related Work

The existing SAR image ship detection methods are divided into traditional methods and deep learning methods. The traditional method is mainly represented by the CFAR proposed by Leng [31], and by combining the intensity and spatial distribution of SAR images, a bilateral CFAR algorithm is proposed for ship detection in SAR images, which effectively reduces the influence of SAR ambiguity and sea clutter. Although traditional methods can play a good role in some specific mission scenarios, they still cannot meet the accuracy and real-time requirements of ship target detection in a wider range of scenarios. In order to solve this problem, deep-learning-based algorithms have become the mainstream method in SAR image ship detection tasks due to their end-to-end advantages. Among them, the two-stage detection algorithm is represented by Faster R-CNN, and Li [32] proposed an improved ship detection method based on the Faster R-CNN for SAR images. Liu [33] constructed a scale-independent SAR image proposal box generator to effectively alleviate the multiscale problem. Fu [34] semantically balances multiple features at different levels to help detectors learn more about small ships in complex scenarios. However, the two-stage detection algorithm has a deep network structure, multiple parameters, and a slow detection speed, which cannot meet the actual needs of SAR image ship detection. Therefore, researchers began to turn their attention to single-stage detection algorithms with faster detection speeds, represented by YOLO. Among them, Chang [35] proposed a lightweight network model based on the improved YOLOv2 network. Zhou [36] improved

the structure of the YOLOv3 [37] network and proposed a low-sample target detection method for SAR images based on lightweight meta-learning. Feng [38] added position-enhanced attention to the latest object detector YOLOX [39], redesigning the lightweight multiscale backbone. However, these deep learning algorithms are supervised algorithms, which means that a large amount of labeled training data needs to be prepared in advance, but SAR image labeling is an expensive task, which greatly limits the development of SAR image ship detection.

In recent years, domain-adaptive methods have developed rapidly, and object detection methods based on unsupervised domain adaptation have been widely used in many fields where it is difficult to obtain a large amount of labeled data, but they are less applied in SAR image object detection. In contrast, in natural scenes, there are already many object detection methods based on unsupervised domain adaptation that have achieved good results. For example, Chen [14] designed two domain-adaptive components at the image level and instance level based on the Faster R-CNN model to reduce domain differences. Chen [15] considered the contradiction between transferability and discriminability in adversarial adaptation and proposed a hierarchical transferability calibration network. Deng [16] proposed a cross-domain mean teacher model distillation method to maximize the knowledge of the teacher model and solve the problem of model bias. In addition, Zhou [17] integrated the single-stage detector YOLOv5 [40] with domain adaptation, which improved the detection performance and generalization of the model.

The successful application of an unsupervised domain-adaptive target detection method in natural scenes provides direction for SAR image ship detection. Recently, there has been work on the application of domain-adaptive methods in the field of remote sensing. Among them, Shi [24] reduced the domain difference from the image level and the instance level and realized unsupervised domain adaptation between different SAR datasets. Li [25] and Chen [26] realized unsupervised domain adaptation from optical images to SAR images through adversarial training methods. Guo [27] used a small number of labeled SAR image samples to learn transferable features by using instance-level adaptation. Zhang [28] extracted and aligned the global structure and local instance information between SAR and optical images, and a hierarchical domain-adaptive network for SAR ship detection was constructed. Li [29] transferred knowledge from the optical domain to the SAR domain for SAR ship detection from the three levels of pixels, features and predictions. However, these methods do not make full use of the image information of the SAR image itself, but only improve on the classical domain-adaptive method.

Therefore, according to the actual task requirements, we designed a domain-adaptive SAR ship detection method based on cross-domain feature interaction and data contribution balance. In this method, the labeled optical ship image is used as the source domain and the unlabeled SAR ship image is used as the target domain, so as to realize the SAR image ship detection without labeling of SAR data.

3. Methods

In this paper, a new domain-adaptive SAR ship detection method called CFD based on cross-domain feature interaction and data contribution balance is proposed. CFD adopts the MT model structure. The MT model was originally designed for semi-supervised learning tasks; it has the ability to correct cross-domain differences and perceive target-related features. We apply the MT model to SAR ship target detection with unsupervised domain adaptation. In this section, we will detail the CFD algorithm flow. In the unsupervised domain adaptation SAR ship detection task, we use the optical dataset as the source domain.

For N_s source domain images I^s, N target bounding box $B = \{B_j|_{j=1}^{N}, B_j = (x_j, y_j, w_j, h_j)\}$ and their corresponding class labels $C = \{C_j|_{j=1}^{N}, C_j = 0\}$; define the source domain as $D_s = \{(I_i^s, B_i^s, C_i^s)|_{i=1}^{N_s}\}$. The SAR dataset is used as the target domain; likewise, the target

domain of N_t unlabeled images I^t is defined as $D_t = \{I^t_i|_{i=1}^{N_t}\}$. CycleGAN-SCA is used to learn with invariant features through weak alignment in the global scene and generate the class source image I^s_f and class target image I^t_f. During the training process, $\left(I^s, I^s_f\right)$ and $\left(I^t, I^t_f\right)$ are required to appear in pairs, enter into the MT model, and input (I^s, I^s_f, I^t) into the student model and I^t_f into the teacher model, where the loss of source domain data training is expressed as:

$$L_{\text{det}}(I^s) = L_{\text{box}}(B^s; I^s) + L_{\text{cls,obj}}(C^s; I^s) \tag{1}$$

Similarly, the loss at training time for class source domain data is expressed as:

$$L_{\text{det}}\left(I^s_f\right) = L_{\text{box}}\left(B^s; I^s_f\right) + L_{\text{cls,obj}}\left(C^s; I^s_f\right) \tag{2}$$

In the MT model, both the student model and the teacher model have the same structures, and we use SFE as the backbone network, which combines global information and rich context information to explore the potential of feature representation through the self-attention mechanism and extract ship target features more effectively. LFE includes an upsampling module and a context aggregation module with larger receptive fields, which improves the resolution of ship targets, enhances the semantic features of different scales, and effectively enhances the detection performance of the model, especially for small ship targets. Finally, the enhanced features are predicted by three detection layers at different scales, and the parameters of the loss-updating model are calculated. The exponential moving average (EMA) of the weight parameters of the student model is used to guide the update of the weight parameters of the teacher model, assuming that the weight parameter of the student model is P^s and the weight parameter of the teacher model is P^t. P^t_n are updated after the Nth training batch:

$$P^t_n = \alpha P^t_{n-1} + (1-\alpha)P^s_n \tag{3}$$

where α is exponential decay, usually 0.99, 0.999, etc. In the distillation process, the teacher model selects a high-probability bounding box as a pseudolabel by predicting the I^t_f of the class target image, which guides the student model training, reduces the detection loss of the student model on the target domain, and enhances the robustness of the model. Distillation losses are expressed as:

$$L_{dis}\left(I^t, I^t_f\right) = L_{\text{det}}\left(I^t; P\left(I^t_f\right)\right) \tag{4}$$

where $P\left(I^t_f\right)$ is the pseudolabel generated by the teacher model for I^t_f prediction of the class target image. Therefore, the total loss of the model is expressed as:

$$L = L_{\text{det}}(I^s) + L_{\text{det}}\left(I^s_f\right) + L_{\text{dis}}\left(I^t, I^t_f\right) \tag{5}$$

In the calculation of the regression box loss function, to accelerate the convergence of the function, balance the contribution of different quality samples to the loss, and optimize the impact of low-quality samples, a new regression box loss named $E^{\frac{1}{2}}IoU$ loss is designed. Finally, we use the trained student model to reason about the target domain image.

The cross-domain image generation module CycleGAN-SCA, self-attention feature extraction backbone network SFE, lightweight feature fusion and feature enhancement neck LFE, and simple and efficient $E^{\frac{1}{2}}IoU$ loss are the main modules and strategies of the CFD model, and Figure 1 shows the overall network structure of the model. In the following sections, we will cover these four aspects in detail.
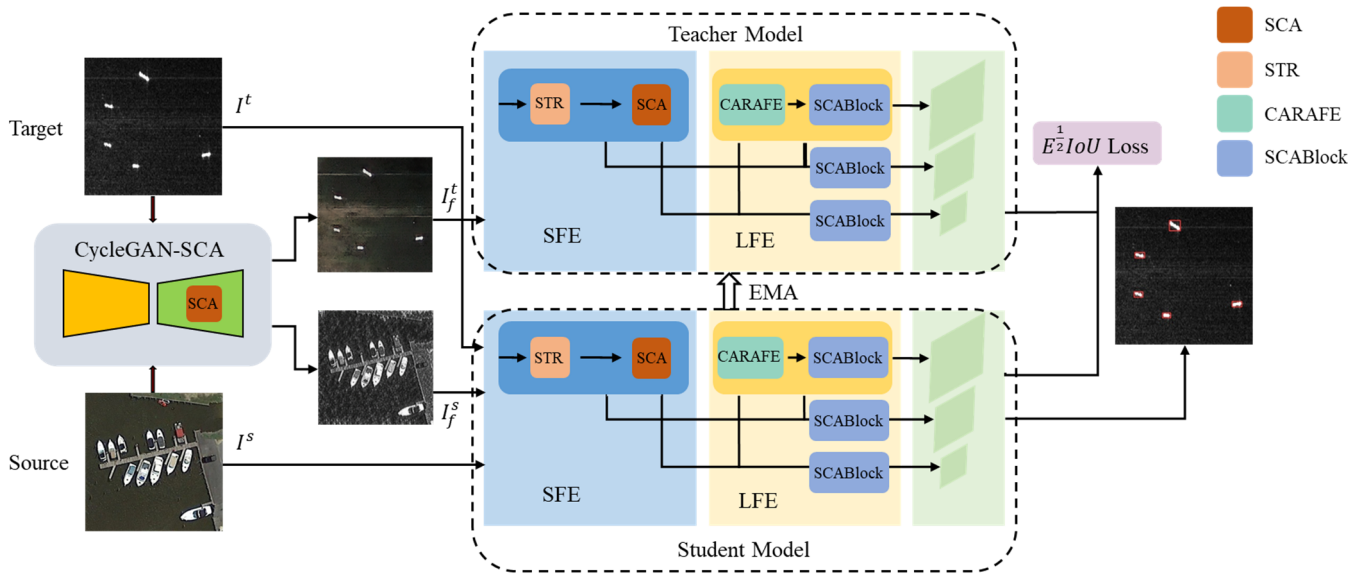
Figure 1. Network structure of CFD.

3.1. Cross-Domain Image Generation Module: CycleGAN-SCA

In SAR ship detection based on unsupervised domain adaptation, there are significant differences in texture, color, and appearance between the optical images in the source domain and the SAR images in the target domain due to the disparity in imaging principles. These domain differences greatly affect the model's performance in domain-adaptive tasks. To address this issue, we propose a cross-domain image generation module called CycleGAN-SCA, based on the CycleGAN framework [41], to generate pseudo-images that resemble both the source and target domains. CycleGAN is a cyclic image generation module that learns the features between different domains and generates pseudo-images with consistent content but different styles. In our approach, the style of the pseudo-image from the source domain resembles that of the target domain, and vice versa.

In order to strengthen the network's attention to local information, we add SCA attention mechanism to the discriminator of CycleGAN to improve the quality of the generated pseudo-images. SCA is an improved version of the coordinate attention (CA) mechanism [42]. Since the Sigmoid activation function in CA is prone to gradient disappearance and gradient explosion, it is not suitable for deep neural networks. The ReLU activation function has the problem of neuronal death, and the problem of gradient disappearance may occur in the case of negative input. Especially in the small target detection task of SAR ships, the gradient disappearance problem will cause that the gradient cannot be effectively transmitted to the shallow network layer in the process of backpropagation, so the shallow network cannot learn effective feature representation. Therefore, the Sigmoid activation function in CA is replaced by the SiLU activation function. SiLU introduces nonlinearity on the basis of Sigmoid function and has the advantages of both sigmoid and ReLU. Approaching ReLU on the positive interval, while approaching sigmoid on the negative interval, can better retain the information of the SAR ship images. The activation function expressions for Sigmoid and SiLU are as follows:

$$Sigmoid(x) = \frac{1}{1 + e^{-x}} \qquad (6)$$

$$SiLU(x) = x * Sigmoid(x) = x * \frac{1}{1 + e^{-x}} \qquad (7)$$

The two activation functions are shown in Figure 2. We refer to the improved attention mechanism as SCA, as shown in Figure 3. SCA not only accurately captures the location of the region of interest but also effectively learns the relationships between channels. The

capabilities of this model can enhance the discrimination of different regions in the image and improve the quality of the resulting pseudo-images. At the same time, the high-quality target domain pseudoimages share common labels with source domain samples, which improves the diversity of trainable samples and enhances the generalization of the model.
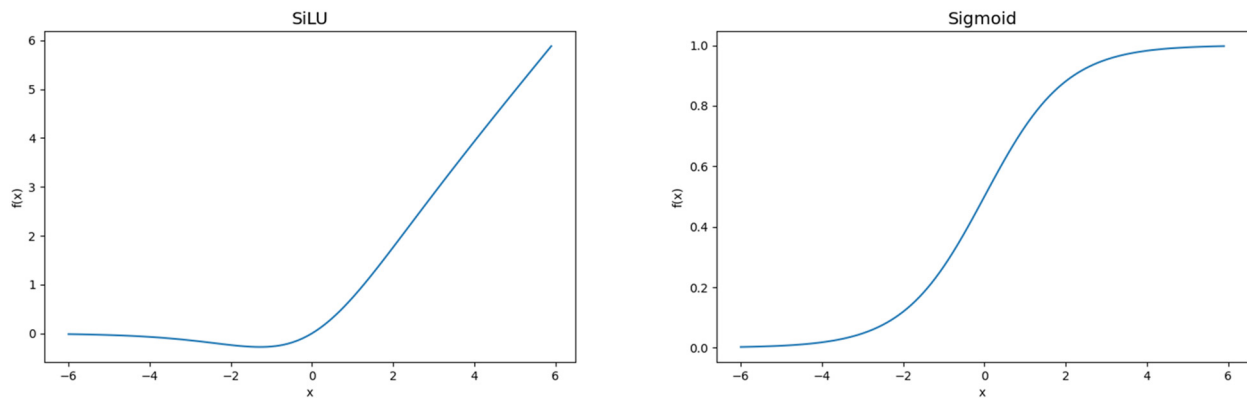

Figure 2. Comparison of the SiLU activation function with the Sigmoid activation function.
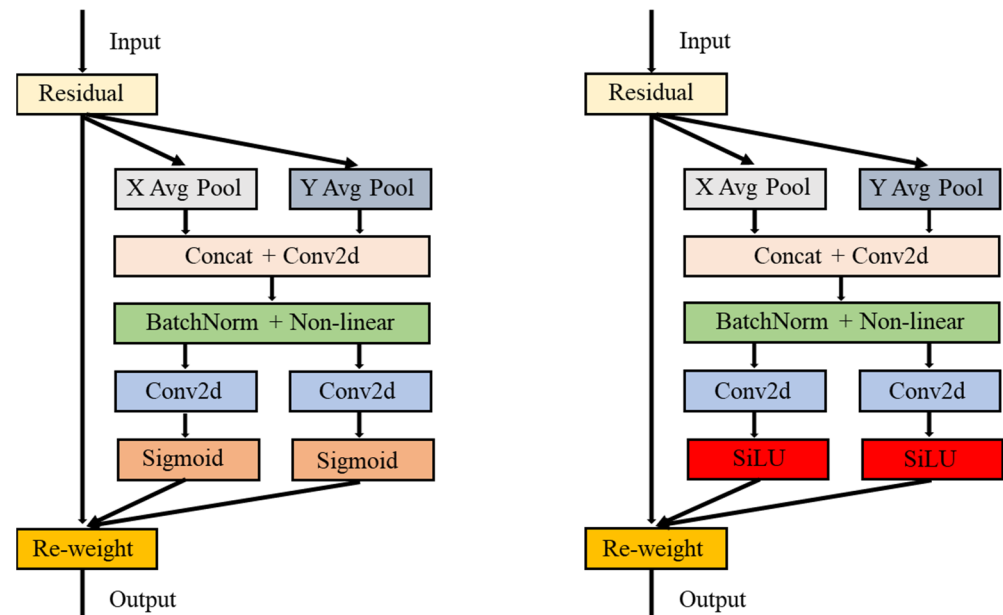

Figure 3. Comparison of the structure of the CA module with the improved SCA module.

3.2. Self-Attention Feature Extraction Backbone Network: SFE

In the SAR image ship detection task, considering the special imaging mechanism of SAR images, it is difficult for the traditional backbone network to extract discrete scattered feature information in SAR images, and a new self-attention feature extraction backbone network SFE is designed. SFE combines the advantages of local feature extraction of the CNN network and transformer [43] global relationship modeling and deeply mines the potential of image feature representation through global information and rich context information, achieving more efficient extraction of ship target features.

In image processing tasks, by introducing position coding into the image, the transformer is able to capture contextual information at different locations of the image, enabling more accurate modeling and prediction. In deep networks, the transformer can perform deeper feature extraction while consuming less memory and computing resources. Compared to traditional CNNs, the Transformer model has obvious advantages. However, completely discarding CNNs will make the model lose the ability to capture local features,

and combining the advantages of both CNNs and transformers will bring better results. Therefore, we added the Swim-Transformer block module in Swim-Transformer [44] and the newly designed SCA attention mechanism module to the backbone network, and through a large number of experimental verifications and analyses, a new backbone network SFE is finally designed, and Figure 4 shows the network structure of the SFE.
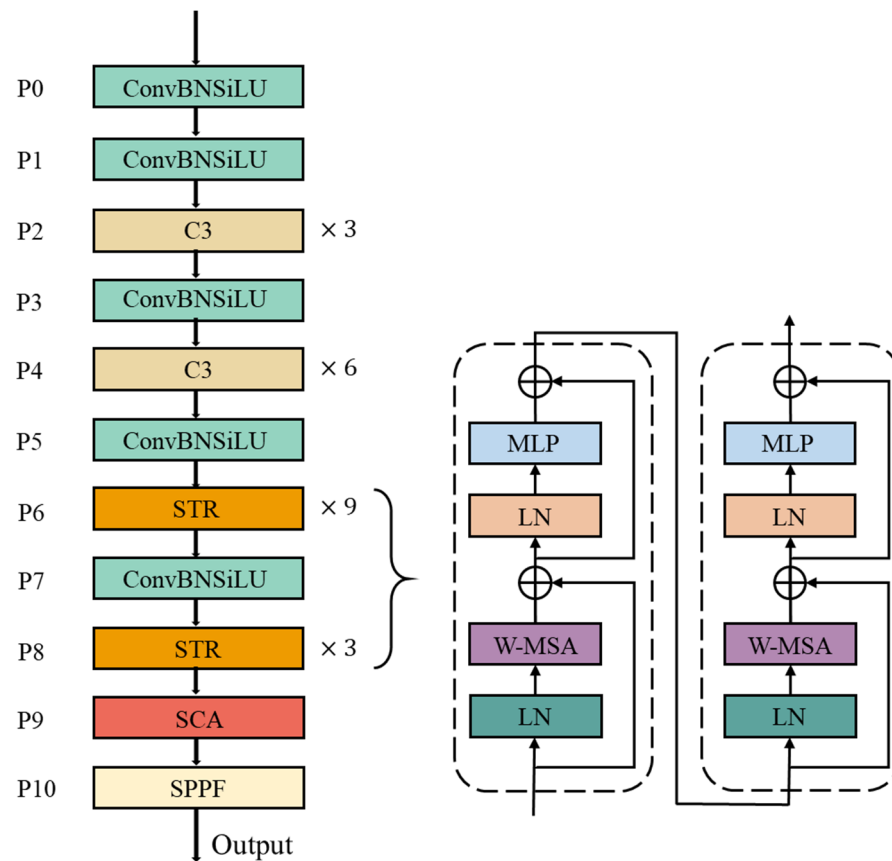

Figure 4. Network structure of SFE.

We added two Swim-Transformer modules to the backbone's deep network and renamed them STR modules. Compared with the extraction of simple features of SAR ship targets in shallow networks, we expect to extract the semantic features of the entire SAR image in deep networks so that ship targets can be detected from complex backgrounds more effectively. Therefore, we use the STR module in the deep network and keep the original C3 module in the shallow network. In addition, we introduce the SCA mechanism module after establishing the STR module to enhance the network's feature extraction ability for local information in the images. The SCA module considers both the position relationship and channel attention in the SFE backbone. This allows us to accurately capture the region of interest by utilizing the captured position information effectively, while also capturing the relationship between channels. Compared with the SE module and the CBAM module, the SE module only considers spatial attention, the CBAM module separates spatial attention and channel attention, and the design of SCA is more suitable for SAR image ship detection tasks.

Through extensive experimental comparisons, the results demonstrate that the design of the SFE module significantly improves the model's feature extraction capability.

3.3. Lightweight Feature Fusion and Feature Enhancement Neck: LFE

In the task of ship detection with SAR images, we are faced with the problems of multiscale ship targets, few features of small-sized targets, and easy loss of information. In order to solve these problems, we expect to design a module that can fully integrate deep

and shallow semantic features at different scales. To this end, we designed a lightweight neck LFE with feature fusion and feature enhancement. LFE consists of the feature map upsampling module CARAFE [45] and the context aggregation module SCABlock, and the network structure of the LFE is shown in Figure 5.
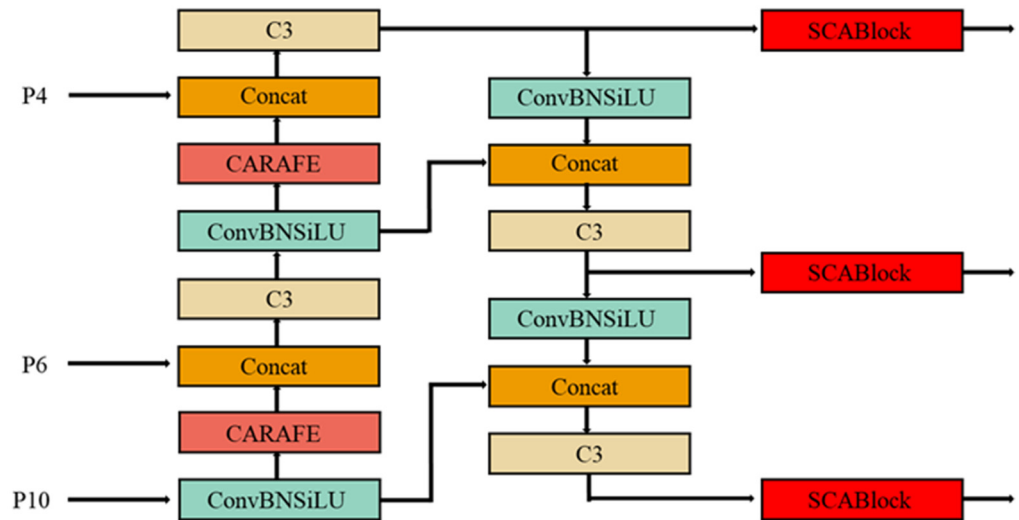

Figure 5. Network structure of LFE.

To address the limitations of conventional upsampling methods, such as limited receptive fields, difficulty in capturing semantic information from feature maps, and high parameter and computational requirements, we employ the CARAFE upsampling module proposed by Wang [45]. CARAFE offers a larger receptive field and effectively utilizes surrounding information, which is crucial for dealing with complex and irregular feature maps, such as ship targets in SAR images. At the same time, the CARAFE module can dynamically generate an adaptive kernel based on the semantic information of the input content, so as to better adapt to the characteristics of different targets. In addition, CARAFE introduces little computational overhead, exhibits fast calculation speed and is a lightweight and efficient upsampling module. Using CARAFE as the upsampling module can alleviate the problem of loss of small-size target information in SAR image ship detection and make more efficient use of the feature information of ship targets.

Finally, before object detection of feature maps of three different scales, the context aggregation module SCABlock is added to further enhance the features by learning the global spatial context of each scale. SCABlock is an improved version of CABlock used by Liu [46] in optical remote sensing image instance segmentation tasks, as shown in Figure 6. As shown in the figure, there is a clear difference in the structure of the two modules. SCABlock has an extra layer of shortcut connection than CABlock, and the Sigmoid activation function in the original position is replaced by the SiLU activation function. The shortcut connection is added to effectively fuse local and global features before SAR ship detection, reduce information confusion, enhance feature extraction of feature maps of different scales, alleviate the problems of multiscale and insufficient features of ship targets, and improve the accuracy of ship target detection in SAR images.
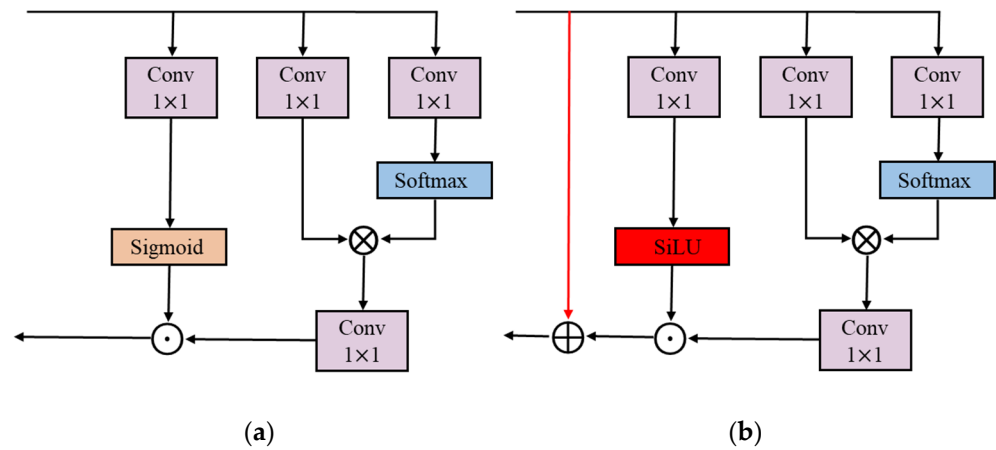
Figure 6. CABlock (**a**) compared to SCABlock (**b**).

3.4. Simple and Efficient Regression Box Loss: $E^{\frac{1}{2}}IoU$ Loss

In the object detection task, the loss function used for bounding box regression is based on the intersection over union (IoU) [47], which calculates the overlap between the predicted bounding box and the real bounding box. Although IoU Loss solves the two major problems of smooth L1 series variables being independent of each other and not having scale invariance, it also has two problems. First, when there is no overlap between the prediction box and the real bounding box, the loss does not reflect the distance between the two boxes. In these cases, the loss value is 0 and the gradient cannot be used for further training. Second, when the size and shape of the prediction box and the real bounding box are fixed, different intersection modes can obtain the same IoU value, which cannot reflect how the two intersect. In response to these problems, the variants of IoU Loss, GIoU Loss [48], DIoU Loss [49], CIoU Loss [49], and EIoU Loss [50] have been proposed. Among them, EIoU Loss proposed by Zhang [50] adds factors such as the center point distance, aspect ratio, and overlapping area between the predicted bounding box and the real bounding box to the calculation of IoU Loss. This modification results in faster convergence speed and improved positioning ability of the model. Furthermore, Focal-EIoU Loss is introduced by combining Focal Loss [51] to address the problem of sample imbalance in bounding box regression. This loss function reduces the contribution of prediction boxes with low overlap with the target bounding box to the regression box loss, thus prioritizing high-quality prediction boxes during the regression process. The formula for Focal-EIoU Loss is as follows:

$$L_{\text{Focal-EIoU}} = \text{IoU}^{\gamma} L_{EIoU} \tag{8}$$

γ is a parameter that ranges from 0 to 1, and in the original paper, the authors found through ablation experiments that the best trade-off can be achieved when $\gamma = 0.5$.

In SAR image ship detection missions, because the target feature information is difficult to extract, the number of low-quality samples is far more than the number of high-quality samples, and the training samples are extremely unbalanced. Therefore, based on the idea of EIoU Loss, we design a simpler and more efficient regression box $E^{\frac{1}{2}}IoU$ Loss, which is in the following form:

$$L_{E^{\frac{1}{2}}IoU} = (L_{EIoU})^{\frac{1}{2}} \tag{9}$$

Since the function expression of the Focal-EIoU Loss is more complex, to compare the differences between the EIoU Loss, Focal-EIoU Loss, and $E^{\frac{1}{2}}IoU$ Loss, the IoU in the Focal-EIoU Loss function expression is approximately replaced with the EIoU, and the changes in the three IoU Losses with the EIoU are shown in Figure 7. From the figure, it is evident that the Focal-EIoU Loss and $E^{\frac{1}{2}}IoU$ Loss have different gradients when dealing with high quality and low-quality samples. The Focal-EIoU Loss strongly

suppresses the loss for low-quality samples, while enhancing the loss for high-quality samples. This design allows the network model to focus more on detecting high-quality samples. However, in the SAR image ship detection task dominated by small targets, low-quality prediction boxes and low-confidence predictions may be ignored or assigned to small targets, which will severely impact the performance of the model. $E^{\frac{1}{2}}IoU$ Loss is improved on this basis, which balances the contribution of high-quality samples and low-quality samples to loss, strengthens the attention to low-quality samples, and improves the model's detection performance for small-sized target ships. Furthermore, $E^{\frac{1}{2}}IoU$ IoU Loss has lower computational complexity, and compared with Focal-EIoU Loss, there are fewer calculation parameters, which accelerates the convergence speed of the model. Experimental results show that the design of $E^{\frac{1}{2}}IoU$ Loss is effective.
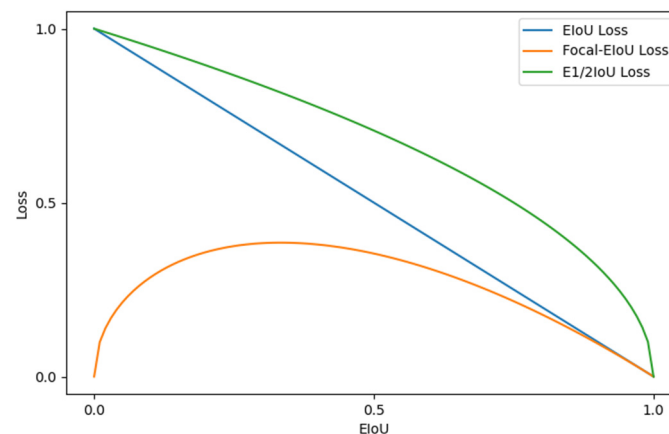

Figure 7. Function graph comparison of EIoU Loss, Focal-EIoU Loss, and $E^{\frac{1}{2}}IoU$ IoU Loss.

4. Experiments

4.1. Datasets Introduction and Experimental Setup

To ensure the accuracy and effectiveness of our CFD model, we perform experimental verification on the self-built optical-SAR cross-domain target detection dataset. The dataset is based on the commonly used optical image dataset DIOR [52] and SAR image ship target detection dataset SSDD.

(1)　DIOR: The DIOR dataset is a large-scale, publicly available optical remote sensing image target detection benchmark dataset proposed by Northwestern Polytechnical University. The dataset adopts a horizontal bounding box (HBB) annotation format, containing 23,463 images and 190,288 instances, covering 20 types of remote sensing ground objects, with an image size of $800 \times 800 \times 3$ and a resolution of 0.3–30 m.

In this paper, we only utilize images from the DIOR dataset that contain ship targets. After strict screening, 928 ship images were selected as the source domain, which was consistent with the number of images in the target domain, to avoid the problem of model performance degradation caused by the imbalance between the two domains. The source domain data include ship targets near the shore and far out at sea, and target sizes are widely distributed.

(2)　SSDD: The SSDD dataset is the earliest dataset published at home and abroad specifically for SAR image ship target detection, and the dataset draws on the production process of the PASCAL VOC dataset. The data are sourced from multiple satellite sensors, including four polarization methods, with resolutions ranging from 1 m to 15 m, covering a large area of ship targets near the shore and far out at sea.

In the SSDD dataset, there are 1160 images of a total of 2456 ships. The ratio of target length or width to image size ranges from 0.04 to 0.24, and most of them are very small targets. The aspect ratio has a wide distribution, from 0.4 to 3. We divide the dataset into

a target domain training set and a test set according to the ratio of 8:2, where the target domain training set has 928 images, which is the same number as the source domain images. The remaining 232 images serve as the target domain test set.

In our experiment, first, the CycleGAN-SCA model generates high-quality class target domain data and class source domain data from the input source domain data and target domain training data. Then, these four types of data are uniformly cut to a size of 512×512 and input into the CFD model. Finally, the model's performance is tested on the target domain test data. To balance computing power and real-time requirements, we use partial training parameters of YOLOv5s and set the batch size to 4 for a total of 500 epochs. The parameter α of the EMA in the MT model is set to 0.99. The experiment uses an NVIDIA Quadro RTX 6000 GPU and the Ubuntu 16.04 Linux operating system, based on the PyTorch deep learning framework. Figure 8 shows instances selected from the DIOR and SSDD datasets.
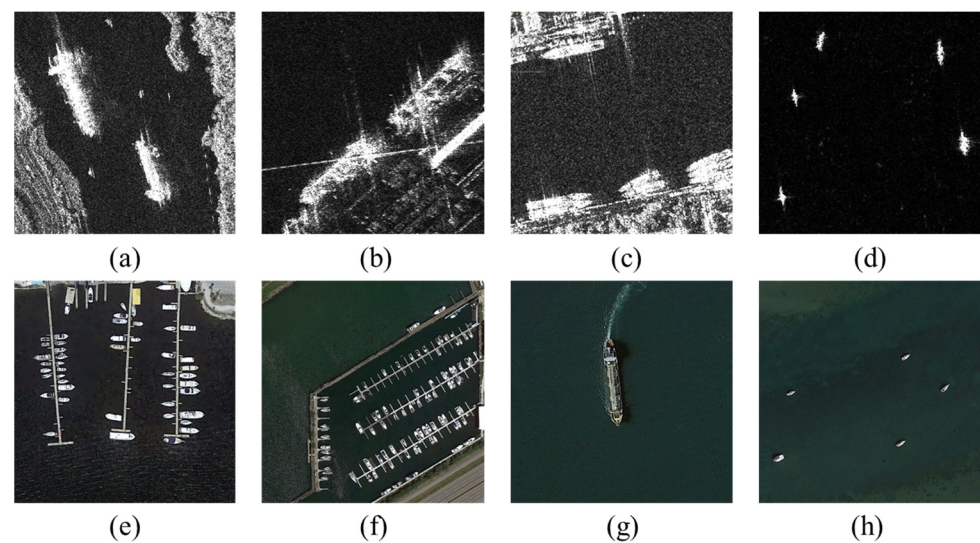

(a)	(b)	(c)	(d)
(e)	(f)	(g)	(h)

Figure 8. Some images selected from (**a**–**d**) SSDD datasets and (**e**–**h**) DIOR datasets.

4.2. Experimental Indices

In this experiment, mAP.5 was used as the evaluation index. The mean average precision (mAP) is a commonly used evaluation index in object detection models, calculated according to precision P and recall R, and the formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{10}$$

$$R = \frac{TP}{TP + FN} \tag{11}$$

$$mAP = \int_0^1 P(R)dR \tag{12}$$

The precision rate reflects the false alarm rate in the overall detection results, and a higher precision indicates a lower false alarm rate. The recall rate reflects the missed detection rate in the overall detection results, with higher recall rates and lower missed detection rates. The mAP is calculated from the area under the precision–recall (PR) curve, and it provides a comprehensive evaluation of the model's performance based on accuracy and recall. A higher mAP indicates better model performance.

4.3. Experimental Results

To validate the excellent performance of our proposed CFD method, we conduct a series of experiments on the preprocessed self-built large-scale optical-SAR cross-domain

target detection dataset. Firstly, we compare CFD with existing methods to demonstrate its superiority. Moreover, we perform numerous ablation experiments to verify the effectiveness of each component in the CFD model. Furthermore, we compare the newly designed $E^{\frac{1}{2}}IoU$ Loss with the existing IoU Loss, and the comparative experiments show that the $E^{\frac{1}{2}}IoU$ Loss significantly improves the performance of the model, which demonstrates the rationality of the $E^{\frac{1}{2}}IoU$ Loss design. To further reflect the application prospect of CFD in SAR image ship detection, we add different proportions of SAR annotated data to the training set, and the experimental results show that the combination of CFD and a small amount of annotated data can achieve detection accuracy comparable to supervised learning. We also discuss the influence of image generation on the detection performance and justify CycleGAN-SCA in SAR image ship detection tasks. Lastly, we compare the performance of CFD with the existing unsupervised domain adaptation methods under three background noises to verify the robustness of CFD under background noise.

4.3.1. Comparative Experiment with the Latest Unsupervised Domain Adaptation Methods

To verify the high performance of our proposed CFD, we compare CFD with the latest unsupervised domain adaptation methods on the self-built large-scale optical-SAR cross-domain target detection dataset. In the comparative experiment, we ensure that the division of the data set is consistent in each experiment to reflect the fairness of the experiment. The results of each experiment are shown in Table 1.

Table 1. Comparison with the latest unsupervised domain adaptation methods.

Method	Precision	Recall	mAP
Source_only	0.4831	0.3595	0.3330
DA [14]	0.6328	0.5413	0.5342
HTCN [15]	0.6801	0.6154	0.6145
Baseline	0.6561	0.5879	0.5934
CFD	**0.7237**	**0.6548**	**0.6854**
Upper bound	0.9662	0.9469	0.9836

The bold format is to highlight the results of our experiments.

Source_only represents the performance of the YOLOv5s model trained in the source domain without domain adaptation in the target domain. Baseline is a network before CFD design, DA and DTCN represent unsupervised domain adaptation methods as mentioned in [14,15], respectively, and the upper bound refers to the performance of the target domain itself on the YOLOv5s model under full supervision. The data in the analysis table show that the performance of baseline and DTCN is very similar, the performance of DA is not good, and CFD is significantly better than all comparison methods. The mAP obtained by CFD is 15.12% and 7.09% higher than that of DA and DTCN, respectively, and 35.24% higher than that of source_only, but there is still a certain gap compared with the upper bound. The prediction visualization of the method in the table is shown in Figure 9, and CFD detects small targets more accurately and has a lower false and missed detection rate than DA and DTCN. It shows that the application of CFD in the SAR image ship detection task is effective, and compared with other unsupervised domain adaptation methods, CFD is specially designed to enhance the performance of the model by considering the unique characteristics of SAR ship targets.
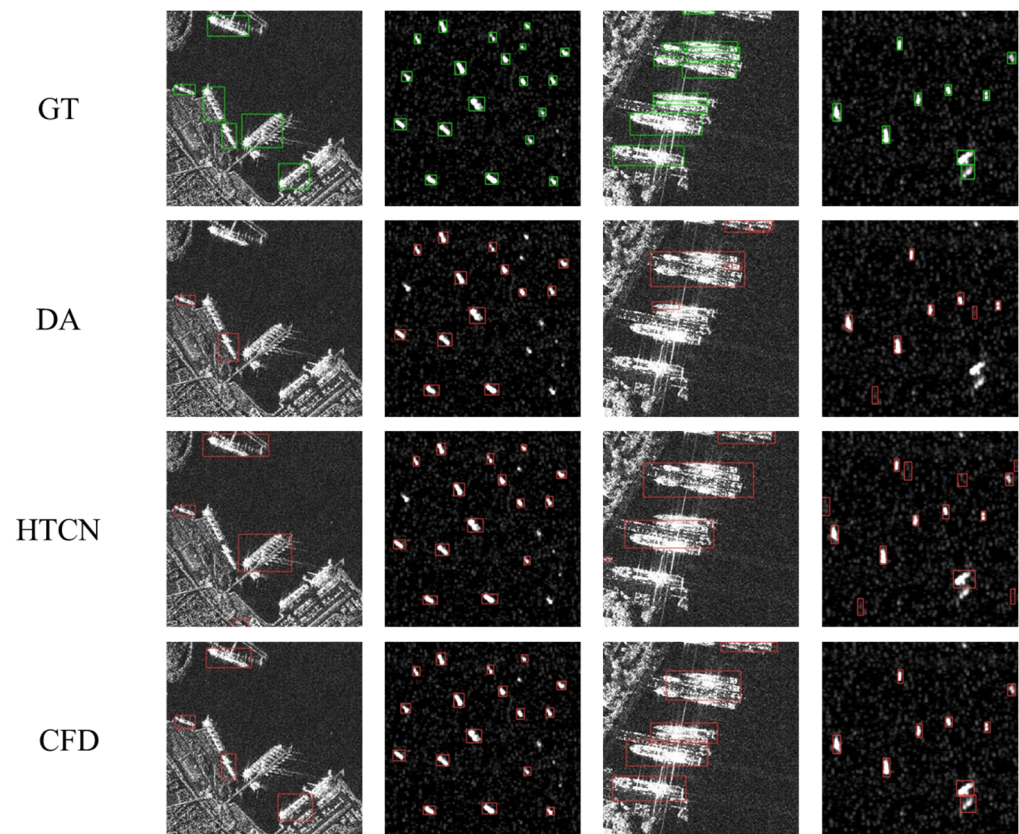
Figure 9. Compared with the detection results of the latest unsupervised domain-adaptive methods.

4.3.2. Ablation Experiments

We conducted ablation experiments on our self-built large-scale optical-SAR cross-domain target detection dataset to showcase the effectiveness of our CycleGAN-SCA, SFE, LFE, and $E^{\frac{1}{2}}IoU$ Loss designs. We combine these designs differently to better validate the impact of different designs on model detection performance.

The ablation experiment results are presented in Table 2, which shows that the cross-domain image generation module CycleGAN-SCA, the self-attention feature extraction backbone network SFE, the lightweight feature recombination and enhancement neck LFE, and the simple and efficient $E^{\frac{1}{2}}IoU$ Loss have improved the detection performance of the model.

Table 2. Ablation experimental results.

Model	SFE	LFE	$E^{\frac{1}{2}}IoU\ Loss$	Cyclegan-SCA	Precision	Recall	mAP
Baseline	×	×	×	×	0.6561	0.5879	0.5934
CFD	√	×	×	×	0.7390	0.6225	0.6489
	×	√	×	×	0.6858	0.6557	0.6385
	×	×	√	×	0.7249	0.6244	0.629
	√	√	×	×	0.7162	0.6354	0.6581
	√	×	√	×	0.7234	0.6370	0.6648
	×	√	√	×	0.7366	0.6350	0.6495
	√	√	√	×	0.7311	0.6325	0.6790
	√	√	√	√	**0.7237**	**0.6548**	**0.6854**

The bold format is to highlight the results of our experiments.

Specifically, when SFE, LFE, and $E^{\frac{1}{2}}IoU$ Loss were used alone, mAP increased by 5.55%, 4.51% and 3.56%, respectively. Among them, SFE had the greatest performance

improvement on the model, which proves that SFE has powerful feature extraction capabilities. In addition, we find that LFE alone can obtain the highest recall rate, proving that LFE can predict targets more accurately by learning deep and shallow semantic features at different scales. When these three designs are combined, the model performance is improved more than if used alone. The combination with $E^{\frac{1}{2}}IoU$ Loss increased mAP by 1.59% and 1.1%, respectively, compared with SFE and LFE alone. It is proven that the design of $E^{\frac{1}{2}}IoU$ Loss is effective in improving model performance. When all three designs were applied simultaneously, the mAP improved by 8.56%, and the performance of the model was greatly improved. Finally, we applied the newly designed cross-domain image generation module CycleGAN-SCA to the model, and the mAP reached the highest value of 68.54%, an increase of 9.2%. It is proven that CycleGAN-SCA effectively reduces the domain difference between the source domain and the target domain and enhances the performance of the model. The visual comparison results of the ablation experiment are shown in Figure 10.
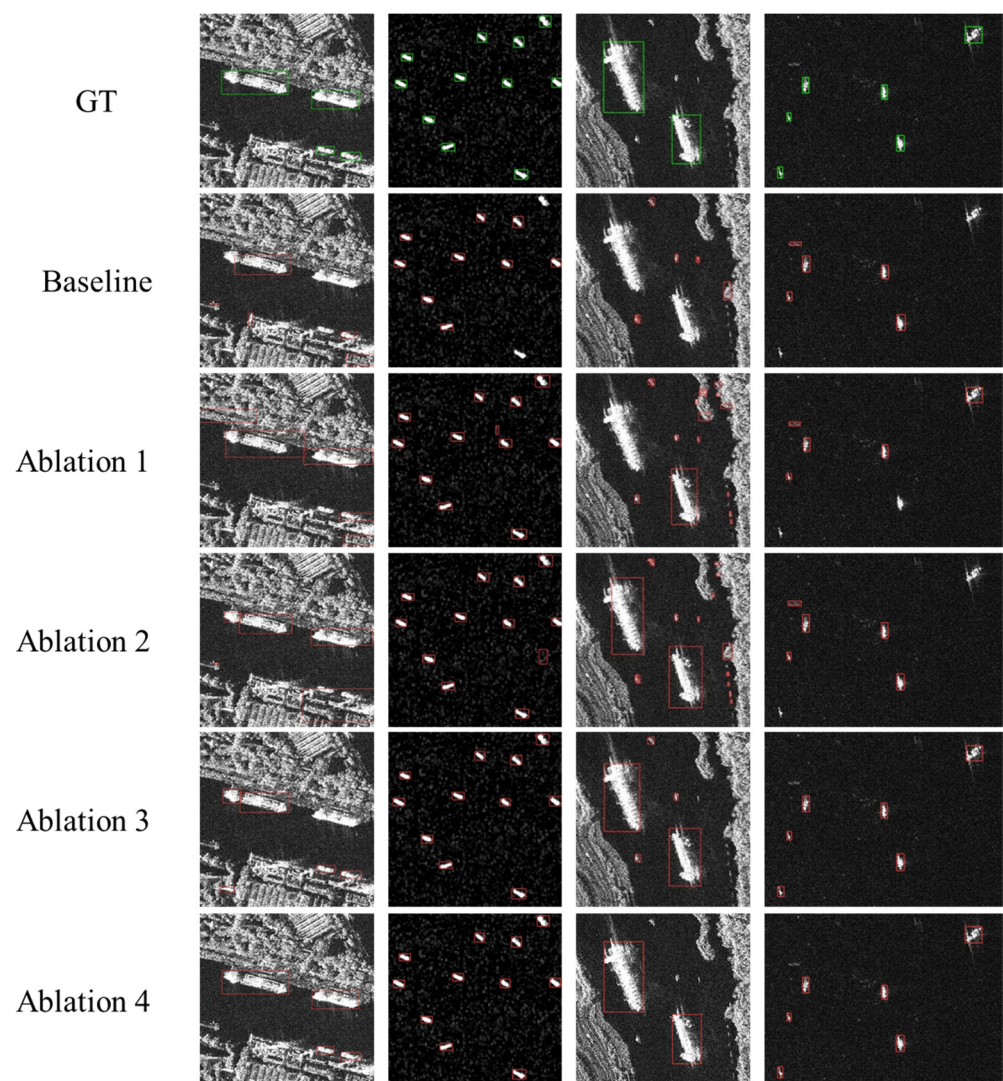

Figure 10. Visual comparison results of ablation experiments. Ablation 1 stands for SFE only; Ablation 2 represents SFE and LFE; Ablation 3 represents SFE, LFE, and $E^{\frac{1}{2}}IoU$ Loss; Ablation 4 stands for adding four designs at the same time.

In summary, the CFD proposed by us exerts excellent performance in SAR image ship detection, and the four designs in CFD contribute to the improvement of model performance.

4.3.3. IoU Loss Comparison Experiment

To further verify the applicability of our designed $E^{\frac{1}{2}} IoU$ Loss in SAR image ship detection, we keep the CFD model structure and parameters unchanged, only change the IoU Loss, and compare the $E^{\frac{1}{2}} IoU$ Loss with several classic IoU Losses. The experimental results are shown in Table 3, where CIoU Loss is the IoU Loss in baseline. EIoU Loss and Focal-EIoU Loss are the inspiration basis for our design of $E^{\frac{1}{2}} IoU$ Loss.

Table 3. IoU Loss comparison experiment.

IoU Loss	Precision	Recall	mAP
GIoU Loss	0.7296	0.6429	0.6538
DIoU Loss	0.6929	0.6410	0.6674
CIoU Loss	0.7162	0.6354	0.6581
EIoU Loss	0.7545	0.6301	0.6710
Focal − EIoU Loss	0.6438	0.6239	0.6303
$E^{\frac{1}{2}} IoU Loss$	**0.7237**	**0.6548**	**0.6854**

The bold format is to highlight the results of our experiments.

As seen from the table, $E^{\frac{1}{2}} IoU$ Loss has the largest improvement in model performance, with mAP reaching 68.54%, which is 2.73% higher than baseline, which proves the high efficiency of $E^{\frac{1}{2}} IoU$ Loss. In addition, comparing the experimental results of the first four IoU Losses in the table, it can be found that EIoU Loss obtains the largest mAP, indicating that the accurate positioning ability of EIoU Loss is more suitable for SAR image ship detection tasks, and it also proves that it is reasonable to design a new IoU Loss based on the ideas of EIoU Loss. Comparing the experimental results of EIoU Loss and Focal-EIoU Loss and $E^{\frac{1}{2}} IoU$ Loss, it can be found that the $E^{\frac{1}{2}} IoU$ Loss we designed is 1.44% higher than the mAP obtained by EIoU Loss, while Focal-EIoU Loss has the worst effect, proving that in SAR image ship inspection missions, the design of IoU Loss balances the contribution of high-quality samples and low-quality samples to loss. It is also necessary to give more attention to low-quality samples and improve the detection performance of the model for small target ships. Therefore, the $E^{\frac{1}{2}} IoU$ Loss we designed is reasonable and efficient in SAR image ship detection.

4.3.4. Additional Comparative Experiments of SAR Image Annotation Data of Different Proportions of the Target Domain

To reflect the robustness and generalization of CFD, it also provides an experimental basis for the development of SAR image ship detection to semisupervised domain adaptation. We add different scales of target domain labeling data to the trained CFD model, and a short training period is carried out; the results are shown in Table 4. Analyzing the experimental data in the table, it can be seen that only adding 5% of the target domain labeling data can increase the mAP of the CFD model by 20.09%, and when the target domain labeling data reach 30%, the performance of the model is close to that of the target domain data fully labeled. Furthermore, the experimental results of supervised training by adding target domain labeled data of the same scale to the YOLOv5s model are also recorded. Comparing the two models, it can be found that when 5% of the target domain labeled data is added, the mAP of CFD increases by 6.27%, and the advantages of the CFD model gradually decrease as the proportion of target domain labeled data gradually increases. This shows that the CFD model exhibits its own advantages in the absence of target domain labeling data. In particular, when 100% of the target domain labeled data is added, the mAP of CFD increases by 0.44%.

Table 4. Comparison experiments of added different scale target domain labeling data.

Percentage	CFD			YOLOv5s		
	Precision	Recall	mAP	Precision	Recall	mAP
0	0.7237	0.6548	0.6854	-	-	-
5	0.9083	0.8348	0.8893	0.8134	0.7906	0.8266
10	0.9112	0.846	0.9117	0.86965	0.79424	0.86653
20	0.9372	0.8707	0.9434	0.9281	0.8297	0.9046
30	0.9486	0.9121	0.9643	0.9329	0.8561	0.9228
100	0.977	0.954	0.9880	0.9662	0.9469	0.9836

Experiments show that the detection performance of the model with only a small number of target domain labeled data is comparable to that of the supervised algorithm. This shows that SAR image ship detection based on unsupervised domain adaptation has high research value and application prospects, which not only alleviates the significant lack of SAR ship image annotation data and the high cost of dataset production but also provides ideas for the application of semisupervised domain adaptation in SAR image ship detection.

4.3.5. Image Generation Comparison Experiment

To verify the rationality of CycleGAN-SCA, we compare the results of image conversion in the source domain of other image generation models. The visualization results of various image generation models are shown in Figure 11, where CUT is the image generation method [53]. It can be seen from the figure that CycleGAN and CycleGAN-SCA can correctly identify ship targets in optical images and generate higher quality pseudo-SAR images than CUT image generation models. This shows that in SAR image ship detection, the image generation module based on CycleGAN can effectively mitigate the domain difference between the source domain and the target domain. In addition, the improved CycleGAN-SCA produces ship targets that are closer to the original image in texture and shape than CycleGAN.

The experimental results of the three image generation models are shown in Table 5; CycleGAN has a great improvement in performance compared with CUT, and the mAP is improved by 4.81%. The newly designed CycleGAN-SCA also improves the performance of the model compared to CycleGAN, and the mAP is improved by 0.64%. Experiments show that it is feasible and efficient to use CycleGAN-SCA to generate images between domains to reduce domain differences in ship detection based on unsupervised domain adaptation.

Table 5. Comparative experiments of different image generation models.

Method	Precision	Recall	mAP
CUT	0.6598	0.6292	0.6309
CycleGAN [41]	0.7311	0.6325	0.6790
CycleGAN-SCA	**0.7237**	**0.6548**	**0.6854**

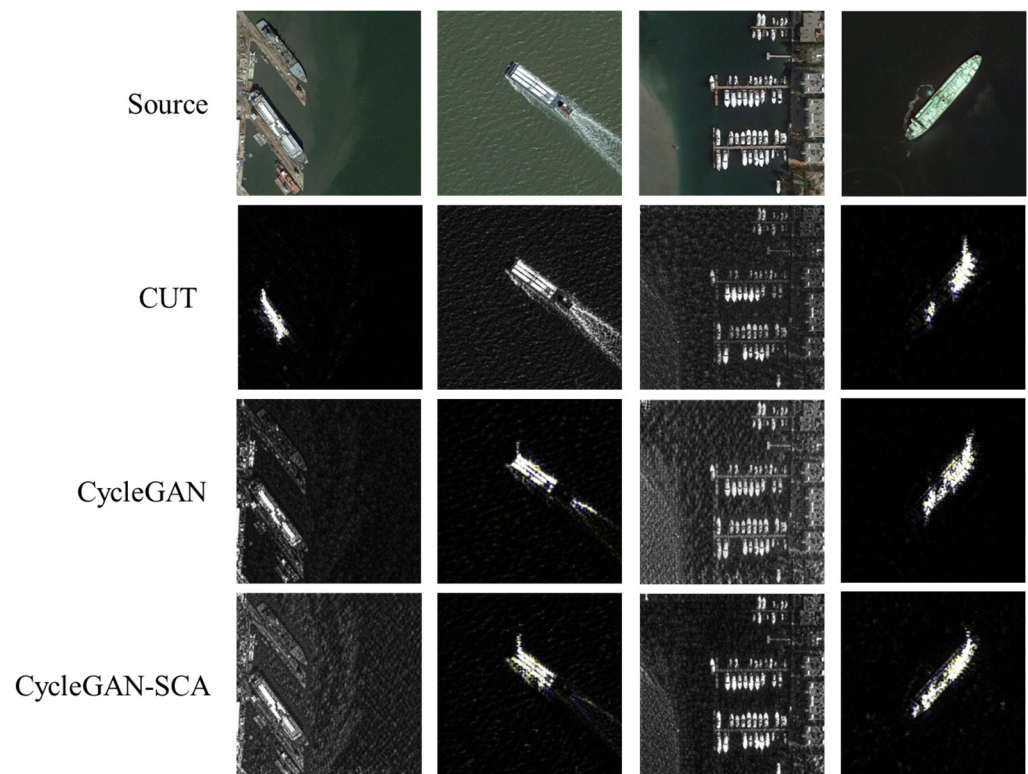The bold format is to highlight the results of our experiments.

Figure 11. Visualization of source domain image conversion by different image generation models.

4.3.6. Comparative Experiments with Background Noise

SAR image background noise is a major factor affecting SAR ship detection. In order to verify the robustness of CFD proposed by us, we added Gaussian noise, salt and pepper noise and random noise to test images, respectively, and compared the performance of CFD and existing unsupervised domain adaptation methods under these three background noises. The experimental results are shown in Table 6. The analysis of the experimental data in the table shows that, in the three background noises, the map obtained by CFD is the highest. In particular, in the case of Gaussian noise, the map obtained by CFD is 15.72% and 8.84% higher than that of DA and HTCN, respectively; in the case of salt and pepper noise, the map obtained by CFD is 15.66% and 10.05% higher than that of DA and HTCN, respectively; in the case of random noise, the map obtained by CFD is 15.35% and 11.21% higher than that of DA and HTCN, respectively. Moreover, we found that the detection performance gap between CFD and DA and HTCN is greater in the case of background noise than that of no background noise. It proves that compared with other unsupervised domain adaptation methods, CFD can better alleviate the impact of background noise on ship detection, which reflects the robustness of CFD.

Table 6. Comparative experiments with background noise.

Method	Gaussian Noise			Salt and Pepper Noise			Random Noise		
	P	R	mAP	P	R	mAP	P	R	mAP
DA [14]	0.6193	0.5264	0.5209	0.6137	0.5206	0.5123	0.6016	0.5327	0.5177
HTCN [15]	0.6621	0.5964	0.5897	0.6465	0.5742	0.5684	0.6372	0.5826	0.5591
CFD	**0.7210**	**0.6493**	**0.6781**	**0.7119**	**0.6368**	**0.6689**	**0.7145**	**0.6370**	**0.6712**

The bold format is to highlight the results of our experiments.

5. Conclusions

In this paper, CFD, a domain-adaptive SAR ship detection method based on cross-domain feature interaction and data contribution balance, is introduced. This method does not need to annotate SAR ship images and uses the existing optical dataset to efficiently complete the ship detection task of SAR images, which solves the problem of the lack of SAR image annotation data and expensive labeling cost. This method proposes a new cross-domain image generation module, CycleGAN-SCA, designs a new SFE backbone and LFE neck, and finally proposes $E^{\frac{1}{2}}IoU$ loss based on the idea of EIoU loss. CycleGAN-SCA is a cross-domain image generation module that reduces domain differences by generating high-quality pseudoimages. SFE is a self-attention feature extraction backbone network that deeply mines the potential of image feature representation through global information and rich context information and extracts ship target features more effectively. LFE is lightweight feature recombination and neck enhancement that learns semantic features at different scales to alleviate the problems of multiscale, few features and easy information loss of ship targets. $E^{\frac{1}{2}}IoU$ loss is a simple and efficient regression frame loss that balances the contribution of high-quality and low-quality samples in SAR image ship detection tasks, strengthens the focus on low-quality samples, and improves the detection performance of the model for small-size target ships.

CFD shows high performance when used on the self-built large-scale optical-SAR cross-domain target detection dataset, and the experimental results demonstrate that our method outperforms existing unsupervised domain adaptation methods, achieving an optimal performance of 68.54%. Furthermore, with only 5% target domain labeled data, our method improves by 6.27% compared to the baseline. These results validate the effectiveness of our proposed method. CFD has effectively solved the problem of SAR image data label dependence. In the future, we will pay more attention to the optimization of the algorithm and improve the computational efficiency of the algorithm. We will also explore how to make better use of SAR complex field data to extract more target information. In addition, we will continue to deeply explore the mechanism and principle of SAR and will further improve the ability to interpret SAR data.

Author Contributions: Writing—original draft, Y.Y.; Writing—review & editing, J.C.; Supervision, L.S., Z.Z., Z.H. and B.W. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The SSDD dataset used in this study is an open dataset. The ship data in the DIOR dataset can be downloaded in https://pan.baidu.com/s/1Z4eq-g7Crrm8PHISpJexbw?pwd=cdo7 (accessed on 11 January 2024).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Iervolino, P.; Guida, R. A novel ship detector based on the generalized-likelihood ratio test for SAR imagery. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2017**, *10*, 3616–3630. [CrossRef]
2. Gierull, C.H.; Sikaneta, I. A compound-plus-noise model for improved vessel detection in non-Gaussian SAR imagery. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 1444–1453. [CrossRef]
3. Schwegmann, C.P.; Kleynhans, W.; Salmon, B.P. Manifold adapta-tion for constant false alarm rate ship detection in South African oceans. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2015**, *8*, 3329–3337. [CrossRef]
4. Yu, W.; Wang, Y.; Liu, H.; He, J. Superpixel-based CFAR Target Detection for High-resolution SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 730–734. [CrossRef]
5. Huang, Y.; Liu, F. Detecting Cars in VHR SAR Images Via Semantic CFAR Algorithm. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 801–805. [CrossRef]
6. Li, T.; Liu, Z.; Xie, R.; Ran, L. An Improved Superpixel-level CFAR Detection Method for Ship Targets in High-resolution SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *11*, 184–194. [CrossRef]

7. Kaplan, L.M. Improved SAR target detection via extended fractal features. *IEEE Trans. Aerosp. Electron. Syst.* **2001**, *37*, 436–451. [CrossRef]

8. Tello, M.; Lopez-Martinez, C.; Mallorqui, J.J. A novel algorithm for ship detection in SAR imagery based on the wavelet transform. *IEEE Geosci. Remote Sens. Lett.* **2005**, *2*, 201–205. [CrossRef]

9. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

10. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]

11. Sun, X.; Wang, Z.; Sun, Y.; Diao, W.; Zhang, Y.; Fu, K. AIR-SARShip-1.0: High-resolution SAR ship detection dataset. *J. Radars* **2019**, *8*, 852–863.

12. Wei, S.; Zeng, X.; Qu, Q.; Wang, M.; Su, H.; Shi, J. HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation. *IEEE Access* **2020**, *8*, 120234–120254. [CrossRef]

13. Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Xu, Y.; Ke, X.; Zeng, T.; Su, H.; et al. SAR ship detection dataset (SSDD): Official release and comprehensive data analysis. *Remote Sens.* **2021**, *13*, 3690. [CrossRef]

14. Chen, Y.; Li, W.; Sakaridis, C.; Dai, D.; Van Gool, L. Domain Adaptive Faster R-CNN for Object Detection in the Wild. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3339–3348.

15. Chen, C.; Zheng, Z.; Ding, X.; Huang, Y.; Dou, Q. Harmonizing Transferability and Discriminability for Adapting Object Detectors. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 8866–8875.

16. Deng, J.; Li, W.; Chen, Y.; Duan, L. Unbiased Mean Teacher for Cross-domain Object Detection. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 4089–4099.

17. Zhou, H.; Jiang, F.; Lu, H. SSDA-YOLO: Semi-supervised domain adaptive YOLO for cross-domain object detection. *Comput. Vis. Image Underst.* **2023**, *229*, 103649. [CrossRef]

18. Zhang, X.; Yao, X.; Feng, X.; Cheng, G.; Han, J. DFENet for domain adaptation-based remote sensing scene classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5611611. [CrossRef]

19. Wu, L.; Lu, M.; Fang, L. Deep covariance alignment for domain adaptive remote sensing image segmentation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5620811. [CrossRef]

20. Zhu, Q.; Zhong, Y.; Zhang, L.; Li, D. Adaptive deep sparse semantic modeling framework for high spatial resolution image scene classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6180–6195. [CrossRef]

21. Zhong, Y.; Zhang, L.; Gong, W. Unsupervised remote sensing image classification using an artificial immune network. *Int. J. Remote Sens.* **2011**, *32*, 5461–5483. [CrossRef]

22. Zhang, J.; Liu, J.; Pan, B.; Shi, Z. Domain adaptation based on correlation subspace dynamic distribution alignment for remote sensing image scene classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7920–7930. [CrossRef]

23. Zhang, J.; Liu, J.; Pan, B.; Chen, Z.; Xu, X.; Shi, Z. An open set domain adaptation algorithm via exploring transferability and discriminability for remote sensing image scene classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5609512. [CrossRef]

24. Shi, Y.; Du, L.; Guo, Y. Unsupervised Domain Adaptation for SAR Target Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6372–6385. [CrossRef]

25. Li, L.; Zhou, Z.; Wang, B.; Miao, L.; An, Z.; Xiao, X. Domain adaptive ship detection in optical remote sensing images. *Remote Sens.* **2021**, *13*, 3168. [CrossRef]

26. Chen, S.; Zhan, R.; Wang, W.; Zhang, J. Domain adaptation for semi-supervised ship detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4507405. [CrossRef]

27. Guo, Y.; Du, L.; Lyu, G. SAR target detection based on domain adaptive faster r-cnn with small training data size. *Remote Sens.* **2021**, *13*, 4202. [CrossRef]

28. Zhang, J.; Li, S.; Dong, Y.; Pan, B.; Shi, Z. Hierarchical Similarity Alignment for Domain Adaptive Ship Detection in SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5240611. [CrossRef]

29. 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.* **2022**, *60*, 5230317. [CrossRef]

30. Tarvainen, A.; Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *arXiv* **2017**, arXiv:1703.01780.

31. Leng, X.; Ji, K.; Yang, K.; Zou, H. A Bilateral CFAR Algorithm for Ship Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1536–1540. [CrossRef]

32. Li, J.; Qu, C.; Shao, J. Ship detection in SAR images based on an improved faster R-CNN. In Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), Beijing, China, 13–14 November 2017; pp. 1–6.

33. Liu, N.; Cao, Z.; Cui, Z.; Pi, Y.; Dang, S. Multi-scale proposal generation for ship detection in SAR images. *Remote Sens.* **2019**, *11*, 526. [CrossRef]

34. Fu, J.; Sun, X.; Wang, Z.; Fu, K. An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1331–1344. [CrossRef]

35. Chang, Y.-L.; Anagaw, A.; Chang, L.; Wang, Y.; Hsiao, C.-Y.; Lee, W.-H. Ship detection based on YOLOv2 for SAR imagery. *Remote Sens.* **2019**, *11*, 786. [CrossRef]

36. Zhou, Z.; Chen, J.; Huang, Z.; Wan, H.; Chang, P.; Li, Z.; Yao, B.; Wu, B.; Sun, L.; Xing, M. FSODS: A lightweight metalearning method for few-shot object detection on SAR images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5232217. [CrossRef]

37. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

38. Feng, Y.; Chen, J.; Huang, Z.; Wan, H.; Xia, R.; Wu, B.; Sun, L.; Xing, M. A lightweight position-enhanced anchor-free algorithm for SAR ship detection. *Remote Sens.* **2022**, *14*, 1908. [CrossRef]

39. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO series in 2021. *arXiv* **2021**, arXiv:2107.08430.

40. Ultralytics. YOLOv5. Available online: https://github.com/ultralytics/yolov5 (accessed on 1 July 2023).

41. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2242–2251.

42. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 13708–13717.

43. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762.

44. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 9992–10002.

45. Wang, J.; Chen, K.; Xu, R.; Liu, Z.; Loy, C.C.; Lin, D. CARAFE: Content-Aware ReAssembly of FEatures. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3007–3016.

46. Liu, Y.; Li, H.; Hu, C.; Luo, S.; Luo, Y.; Chen, C.W. Learning to Aggregate Multi-Scale Context for Instance Segmentation in Remote Sensing Images. *arXiv* **2021**, arXiv:2111.11057.

47. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T. Unitbox: An advanced object detection network. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 1 October 2016; pp. 516–520.

48. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.

49. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34.

50. Zhang, Y.F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146–157. [CrossRef]

51. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [CrossRef]

52. 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. Photogramm. Remote Sens.* **2020**, *159*, 296–307. [CrossRef]

53. Park, T.; Efros, A.A.; Zhang, R.; Zhu, J.-Y. Contrastive Learning for Unpaired Image-to-Image Translation. *arXiv* **2020**, arXiv:2007.15651.