Article

Boosting SAR Aircraft Detection Performance with Multi-Stage Domain Adaptation Training

Wenbo Yu [1,2] , **Jiamu Li** [1,2], **Zijian Wang** [1,2] **and Zhongjun Yu** [1,2,*]

1 Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China; yuwenbo19@mails.ucas.ac.cn (W.Y.); lijiamu19@mails.ucas.ac.cn (J.L.); wangzijian191@mails.ucas.ac.cn (Z.W.)
2 School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 101408, China
* Correspondence: yuzj@ucas.ac.cn; Tel.: +86-010-56535329

Abstract: Deep learning has achieved significant success in various synthetic aperture radar (SAR) imagery interpretation tasks. However, automatic aircraft detection is still challenging due to the high labeling cost and limited data quantity. To address this issue, we propose a multi-stage domain adaptation training framework to efficiently transfer the knowledge from optical imagery and boost SAR aircraft detection performance. To overcome the significant domain discrepancy between optical and SAR images, the training process can be divided into three stages: image translation, domain adaptive pretraining, and domain adaptive finetuning. First, CycleGAN is used to translate optical images into SAR-style images and reduce global-level image divergence. Next, we propose multilayer feature alignment to further reduce the local-level feature distribution distance. By applying domain adversarial learning in both the pretrain and finetune stages, the detector can learn to extract domain-invariant features that are beneficial to the learning of generic aircraft characteristics. To evaluate the proposed method, extensive experiments were conducted on a self-built SAR aircraft detection dataset. The results indicate that by using the proposed training framework, the average precision of Faster RCNN gained an increase of 2.4, and that of YOLOv3 was improved by 2.6, which outperformed other domain adaptation methods. By reducing the domain discrepancy between optical and SAR in three progressive stages, the proposed method can effectively mitigate the domain shift, thereby enhancing the efficiency of knowledge transfer. It greatly improves the detection performance of aircraft and offers an effective approach to address the limited training data problem of SAR aircraft detection.

Keywords: convolutional neural network (CNN); domain adaptation; aircraft detection; synthetic aperture radar (SAR)

1. Introduction

Synthetic aperture radar (SAR) can provide all-day all-weather ground observation and has drawn considerable attention in various fields, such as maritime surveillance, agricultural survey, and disaster monitoring. With the rapid development of SAR technology and the substantial increase in radar platform numbers, more and more high-quality SAR data are generated, leading to an increasing demand for SAR image interpretation algorithms. Between the applications, aircraft detection in high-resolution SAR images aims to precisely locate all the airplanes in the image automatically and plays a critical role in airport management. However, due to the complex airport background and low intuitiveness of aircraft in SAR images, precisely detecting aircraft in SAR images is still a challenging task.

Represented by the constant false alarm rate (CFAR) [1], traditional object detection methods [2–4] mainly focus on the modeling of clutter distribution and threshold extraction, which highly rely on the scattering intensity of objects and perform poorly under complex backgrounds. In contrast, neural networks can automatically learn to extract useful semantic features from data without designing handcrafted features. With the accumulation

of image data and the development of high-speed computation hardware, deep learning methods have shown great potential in computer vision, and various object detection methods using convolutional neural networks (CNN) have been designed, which can be divided into two categories: two-stage detectors [5–8] and one-stage detectors [9–14]. To extract precise object features, two-stage detectors first generate region proposals that tend to contain objects, which are further analyzed to obtain final classification and localization results. On the contrary, one-stage detectors directly predict object bounding boxes based on grid regions. Compared to two-stage methods, one-stage detectors are faster and can be trained end-to-end, drawing more and more research attention.

On this basis, CNNs have been applied to various SAR image object detection tasks, including vehicles [15,16], ships [17–19], bridges [20], etc. Among these tasks, the detection of SAR aircraft is complex and challenging due to the complex background and scale heterogeneity of aircraft. Therefore, numerous studies have been put forward to improve aircraft detection performance. For instance, to achieve better aircraft description ability and detect complete targets, He et al. [21] used two parallel networks and a constraint layer to utilize the depth characteristics and component structure of airplane targets. Diao et al. [22] introduced a CFAR-based aircraft pre-locating algorithm to generate high-quantity region proposals. Wang et al. [23] employed airport runway masks to remove false alarms and designed a weighted feature fusion module, achieving higher detection accuracy. Guo et al. [24] proposed a scattering enhancement strategy and an attention pyramid network to detect SAR aircraft precisely. Zhao et al. [25] designed a multibranch dilated convolution module to extract discrete backscattering features of aircraft and achieved better detection performance. Kang et al. [26] put forward an innovative scattering feature relation network (SFR-Net) to enhance the relationships among the scattering points of aircraft and guarantee the completeness of aircraft detection results. To extract multiscale features of the aircraft and suppress background noise, Chen et al. [27] designed an efficient pyramid convolution attention fusion module and a parallel residual spatial attention module, which achieved improved detection accuracy. From the research above, it can be drawn that most studies are focused on designing specialized network structures and complex detection procedures. Through customized design, the discrete and weak aircraft features can be effectively extracted, and the detection accuracy can be significantly improved. However, besides network structure, the training process, which helps the network to learn to extract representative target features and to distinguish targets from the background, also plays a critical role in the final detection performance. To train a reliable CNN detector, a large amount of labeled data is always required. Nonetheless, different from optical images with high data accessibility, SAR images are usually more difficult to obtain and interpret. For aircraft detection tasks, due to the discrete scattering point form of aircraft and the background interference of airport buildings, target annotation has a higher difficulty, and the construction of a large-scale SAR aircraft detection dataset is more arduous and time-consuming, which impedes the detector's training process and decreases detection performance. Therefore, it is necessary to focus on how to effectively learn generic aircraft characteristics with limited training samples.

One effective solution is to use transfer learning, i.e., utilizing knowledge from other tasks to improve SAR aircraft detection performance. For instance, optical images can indicate ground objects using the visible and near-infrared portions of the electromagnetic spectrum. Compared to SAR images, optical images can depict fine-grained object information at a high resolution. As optical images capture similar visual appearances to human eyes, target annotating in optical images is more intuitive and low-cost, making the process of dataset construction more simplified and efficient. In remote sensing scenarios, SAR and optical images are two different renderings of the same ground object, and there exists an inner connection between them. Therefore, transferring knowledge from optical images for SAR image interpretation is feasible and practical. Li et al. [28] proposed to use the pretrained weights from optical images to boost ship detection performance under limited data. Bao et al. [29] designed a complementary pretraining method to transfer

the characteristics of optical ships to SAR images. However, while optical images mainly contain target contour and texture details, SAR images reflect targets as discrete strong scattering points and inevitably generate speckle noise. Transferring these domain-specific characteristics from one domain to the other may decrease the performance and cause negative transfer, making knowledge transfer difficult and inefficient.

To overcome the discrepancy between the two domains and to achieve a better transfer effect, domain adaptation (DA) is widely studied by researchers. As a specific type of transfer learning, domain adaptation aims to improve the effectiveness of source domain knowledge on a related target domain with the same task. Specifically, domain adaptation focuses on learning a type of mapping that can map the source and target domains into a common space, such that the knowledge learned on the source domain can also be applied to the target domain [30]. For instance, Ganin et al. [31] combined domain adaptation and deep feature learning within one training process by using adversarial learning. Chen et al. [32] focused on improving the cross-domain robustness of object detection through feature-level and instance-level distribution alignment. Saito et al. [33] proposed a novel method for detector adaptation based on strong local alignment and weak global alignment.
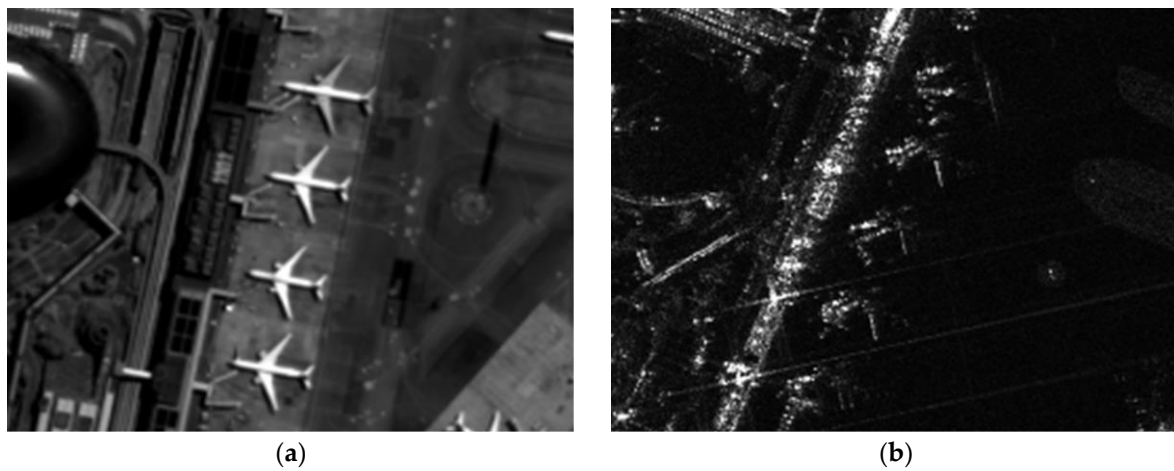
The above domain adaptation methods provide a practical solution for overcoming domain discrepancy and achieving efficient knowledge transfer. Given this perspective, some researchers have put forward domain adaptation methods to transfer knowledge in optical images to SAR images. For instance, to eliminate the need for a large labeled dataset in SAR image classification, Rostami et al. [34] proposed to transfer knowledge from the related easy-to-label Electro-Optical domain by minimizing the feature distribution distance between SAR and optical domains. Chen et al. [35] proposed a pixel-level and feature-level domain adaptation approach to achieve heterogeneous SAR target recognition. To take advantage of optical labeled data and bridge the gap between optical and SAR images, Song et al. [36] designed a two-stage transfer learning framework for SAR ship recognition. For SAR ship detection, Shi et al. [37] put forward an unsupervised domain adaptation framework based on progressive transfer by transferring knowledge from the optical domain and achieved competitive performance. These works demonstrate the effectiveness of transfer learning from optical images for SAR image interpretation tasks. However, due to high target complexity and low data quantity, domain adaptation methods for SAR aircraft detection remain a complete void.

When transferring learning from optical aircraft detection to SAR aircraft detection, due to the significant domain discrepancy, it is difficult for the detector to directly learn generic aircraft characteristics without specific optimization. Specifically, the divergence can be divided into two levels: global-level and local-level. Global-level divergence describes the overall image style, e.g., image brightness and contrast. For example, compared to optical images, SAR images usually have a dark background and high contrast, as shown in Figure 1. This is caused by different scattering characteristics of different ground objects. Additionally, there exist local-level divergences such as different object shapes and textures. One typical instance is that aircraft in SAR images show a discrete scattering point form and vary dramatically with different incident angles. As depicted in Figure 2, the outline of aircraft in SAR images is not clear, and the appearance of aircraft with different incident angles can be completely different, while aircraft in optical images are complete and stable with different incident angles. Such divergences could affect the transfer effect when the detector learns these domain-specific characteristics. Therefore, to address this issue and improve SAR aircraft detection performance, we propose a multi-stage domain adaptation training framework to reduce the global-level and local-level divergences between optical and SAR domains. Detailed contributions are summarized as follows:

1. A multi-stage domain adaptation training (MDAT) framework is proposed in this paper. The training procedure includes three stages, i.e., image translation (IT), domain adaptive pretraining (DA-P), and domain adaptive finetuning (DA-F), to gradually reduce the discrepancy between optical and SAR domains. To the best of

our knowledge, it is the first work that focuses on improving SAR aircraft detection performance by efficiently transferring knowledge from optical images.

2. To reduce the global-level image divergence between optical and SAR domains, CycleGAN [38] is adopted in the first stage to employ image-level domain adaptation. By translating optical images with aircraft targets into corresponding SAR-style ones, the overall image divergences can be effectively eliminated.

3. Additionally, multilayer feature alignment was designed to further reduce local-level divergences. By using domain adversarial learning in both the pretrain and finetune stages, the detector can extract domain-invariant features and learn generic aircraft characteristics, which improves the transfer effect and increases SAR aircraft detection accuracy.

Figure 1. Comparison of optical and SAR images: (**a**) optical image; (**b**) SAR image.

Figure 2. SAR aircraft samples with different incident angles.

The rest of the paper is arranged as follows. In Section 2, the overall structure and detailed improvements of the proposed method are described. Then, Section 3 provides the experiment dataset and evaluation metrics. Section 4 shows the results as well as the corresponding analysis. After that, a comprehensive discussion regarding the results is presented in Section 5. Lastly, the conclusion is drawn in Section 6.

2. Materials and Methods

In this section, the proposed multi-stage domain adaptation training framework is introduced in detail. As a note, the source domain is denoted as $\{X_{Opt}, Y_{Opt}\}$, where X_{Opt} denotes optical images, and Y_{Opt} is the corresponding aircraft labels. The target domain is denoted as $\{X_{SAR}, Y_{SAR}\}$, where X_{SAR} denotes SAR images and Y_{SAR} is the corresponding aircraft labels. Our purpose is to boost aircraft detection performance in SAR images by transferring knowledge from optical images.

2.1. Structure of Detection Networks

This paper proposes a multi-stage domain adaptation training framework for effectively training detection networks to achieve better SAR aircraft detection performance, which is straightforward and can be easily employed on various detection networks. To prove the effectiveness of the proposed training framework, two representative detectors, i.e., Faster RCNN [7] and YOLOv3 [10], were adopted for evaluation. In this section, the structure of the two detectors is introduced in detail.

2.1.1. Faster RCNN

Faster RCNN is the most commonly used two-stage detection network. It mainly consists of three components: a backbone network for feature extraction, a region proposal network (RPN) for predicting regions of interest (RoI), and a fully connected network (FCN) for RoI feature classification and regression. First, the backbone network takes the original image in and generates high-level feature maps that contain global semantic information regarding the contents of the image. Next, the features are sent into the RPN, which is a fully convolutional network, and it predicts a huge amount of high-quality region proposals that may contain the target to be detected. The corresponding feature of the region proposal in the feature maps is extracted with a pooling operation. Finally, the region proposal feature is used in the FCN for final classification and regression. While training the network, both the region proposals generated by the RPN and the final detection results generated by the FCN need to be optimized. Therefore, the training loss of Faster RCNN can be formulated as:

$$L_{Det} = L_{RPN\_reg} + L_{RPN\_cls} + L_{FCN\_reg} + L_{FCN\_cls} \tag{1}$$

where L_{RPN_reg} and L_{RPN_cls} are the bounding box regression loss and the objective classification loss for RPN, respectively, and L_{FCN_reg} and L_{FCN_cls} are the regression and classification loss for final detection results, respectively. To achieve better performance, we used ResNet-50 as our backbone and adopted a feature pyramid network (FPN) behind the backbone to extract multi-scale feature maps with strong semantics. More implementation details can be found in [39].

2.1.2. YOLOv3

YOLOv3 is a one-stage detector that is a fully convolutional network and directly predicts final object detection results. To obtain object bounding boxes, the input image is divided into grids with a fixed size, and each grid generates three bounding boxes that are responsible for potential objects within the grid. Multi-scale prediction is adopted to deal with objects with various scales. The network structure can also be divided into three parts: a Darknet-53 backbone, an FPN neck, and three detection heads. Darknet-53 utilizes successive convolutional blocks with residual skip connections to extract multi-level features. After that, FPN is introduced to fuse low-level feature maps that contain strong localization information with high-level feature maps that have strong semantic information, building a feature pyramid with strong semantics throughout. Finally, three detection heads are used to generate object proposals in three scales, and the training loss can be expressed as:
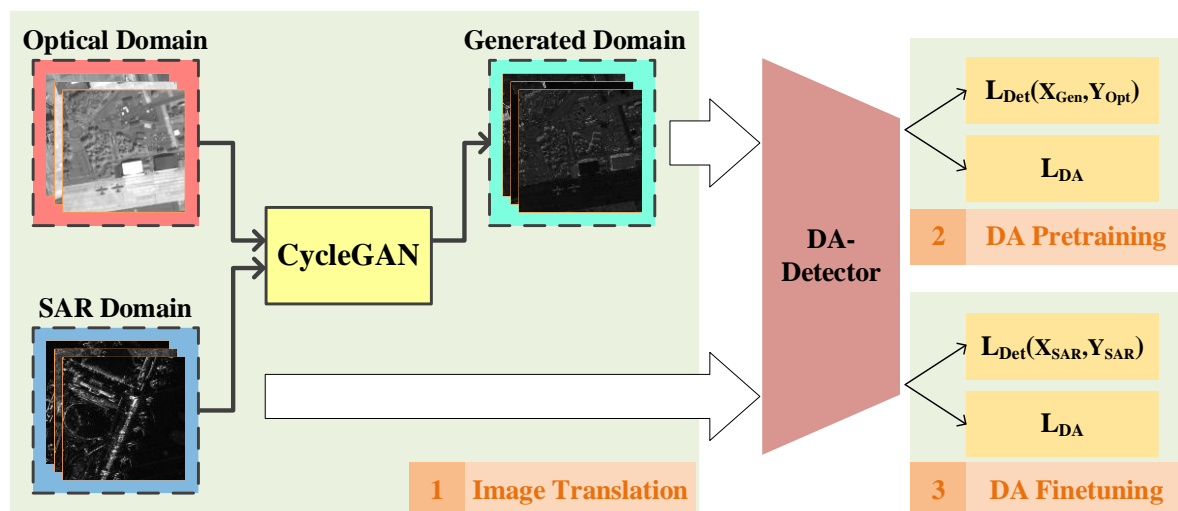
$$L_{Det} = L_{cls} + L_{conf} + L_{xy} + L_{wh} \tag{2}$$

where L_{cls}, L_{conf}, L_{xy}, and L_{wh} are the classification loss, confidence score loss, bounding box offset loss, and bounding box size loss, respectively.

2.2. Overall Framework of MDAT

In the prominent transfer learning method for deep neural networks, i.e., the pretrain–finetune (PF) framework, the model weights trained with the source domain are utilized as the parameter initialization of the target domain. Specifically, the pretrain stage focuses on learning generic target characteristics from the source domain, and the finetune stage customizes it according to the target domain. However, the PF framework is not capable

of overcoming the huge discrepancy between optical and SAR domains, as the network also learns domain-specific features. Transferring these domain-specific features could impede the transfer learning from the optical to the SAR domain and decrease the detector's performance. Therefore, to effectively utilize the knowledge in optical images to boost SAR aircraft detection performance, it is critical to reduce the divergences and focus on the common features that are shared between both domains, e.g., aircraft spatial structure and airport background. To this end, we propose a multi-stage domain adaptation training (MDAT) framework to boost aircraft detection performance in SAR images by transferring knowledge from optical images. As depicted in Figure 3, the proposed method can be divided into three stages: image translation, domain adaptive pretraining, and domain adaptive finetuning, which are introduced as follows:

Figure 3. The overall framework of the proposed MDAT.

Stage 1: Image Translation: In the first stage, to reduce the global-level image difference between the two domains, we utilize a generative adversarial network (GAN) to translate optical images into SAR-style images (the generated domain, denoted as X_{Gen}) that have similar characteristics to SAR images. The generated images can retain the original aircraft spatial location of optical images and have a smaller divergence from the target SAR domain.

Stage 2: Domain Adaptive Pretraining: The detection model is trained on the generated domain. As there still exist local-level divergences between X_{Gen} and X_{SAR}, we apply multilayer feature alignment to reduce the feature distance between the two domains. Specifically, we employ domain classifiers on every feature level of the detection network to discriminate the features and train the detector to learn generic aircraft features on multiple scales.

Stage 3: Domain Adaptive Finetuning: The pretrained detector is finetuned on the target domain to detect SAR aircraft. To avoid the overfitting problem caused by small training samples, the domain classifiers of the pretrained model are retained and continue to align feature distributions of X_{Gen} and X_{SAR}. By initializing with the pretrained weights and constraining the feature distributions to be domain-invariant, the target information of the optical domain can be efficiently transferred to SAR aircraft detection and boost the detection performance.

The proposed MDAT training method uses three stages to gradually reduce the domain discrepancy between optical and SAR domains. The first stage reduces the global-level domain discrepancy in the original image space, which is effective and straightforward. However, though GAN is capable of learning overall image distributions and translating image style, it is difficult for the network to completely overcome the local-level divergence

caused by different imaging mechanisms. Therefore, multilayer feature alignment was proposed to further reduce the local-level divergence in the semantic feature space. With an adversarial learning strategy, the detector can learn domain-variant features that are beneficial to the learning of SAR aircraft detection. Therefore, by using the multi-stage divergence reduction framework, the considerable gaps between the optical and SAR domains can be effectively bridged, and the detector can achieve a better transfer learning effect.

2.3. Global-Level Domain Adaptation with Image Translation

Due to different imaging mechanisms, the same ground object may appear completely different in optical and SAR images. For aircraft detection, such a difference indicates that far-away points in the image space could correspond to similar detection results. As CNN aims to learn the mapping function from the image space to the detection results, to obtain better transfer learning performance, it is critical to reduce the overall distance of the two domains in the image space. On this basis, we propose to use generative adversarial networks (GAN) to translate optical images into SAR-style images and achieve global-level domain adaptation.

Since matched optical-SAR image pairs are hard to obtain, we chose CycleGAN to perform unpaired image translation, which is shown in Figure 4. Specifically, there were two generators: G_{O-S} for translating optical images into SAR-style images and G_{S-O} for translating SAR images into optical-style images. The training loss for these two generators can be expressed as:

$$L_{O-S} = \mathbb{E}_{x \sim X_{Opt}}[\log(1 - D_{SAR}(G_{O-S}(x)))] + \mathbb{E}_{x \sim X_{SAR}}[\log D_{SAR}(x)] \tag{3}$$

$$L_{S-O} = \mathbb{E}_{x \sim X_{SAR}}[\log(1 - D_{Opt}(G_{S-O}(x)))] + \mathbb{E}_{x \sim X_{Opt}}[\log D_{Opt}(x)] \tag{4}$$

where D_{SAR} and D_{Opt} are two domain discriminators. Through the adversarial training strategy, the generators learn to translate source images into fake images that are indistinguishable from target domain images. However, the original target information may be lost in the generated images due to mode collapse [38]. To address this problem, a cycle consistency loss was added to constrain the output of the generator, which is described as:

$$L_{cycle} = \mathbb{E}_{x \sim X_{Opt}}[\|G_{S-O}(G_{O-S}(x)) - x\|_1] + \mathbb{E}_{x \sim X_{SAR}}[\|G_{O-S}(G_{S-O}(x)) - x\|_1] \tag{5}$$
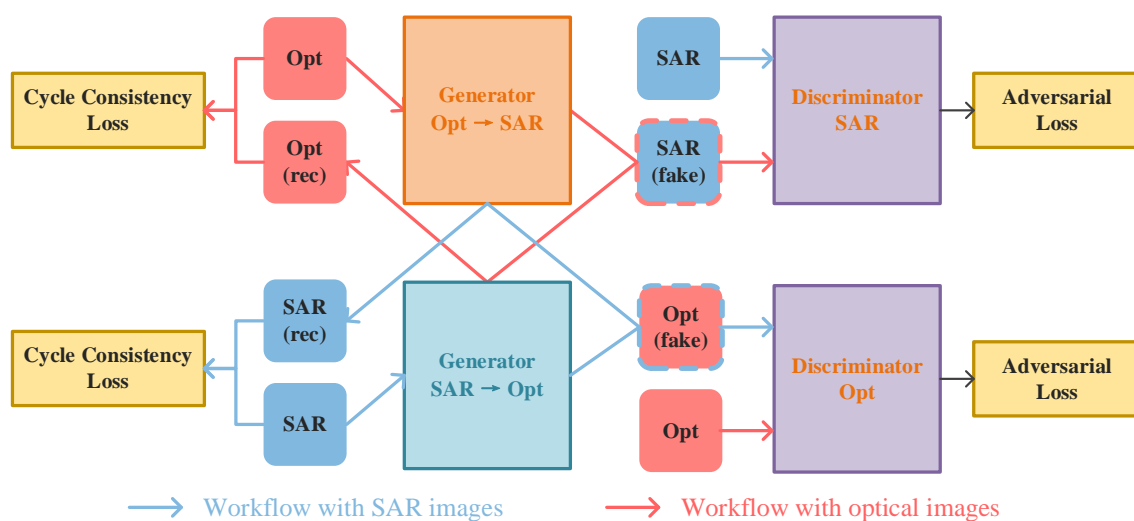

Figure 4. The training framework of the image translation stage [40].

The cycle consistency requires that the original image can be reconstructed through a generated image and thus forces the generators to maintain the original content of the input image. The overall loss for image translation is:

$$L_{S1} = L_{O-S} + L_{S-O} + \lambda \cdot L_{cycle} \tag{6}$$

where λ is a weight parameter set to 10.0. After the training of CycleGAN, the generators can capture the overall image divergence between optical and SAR images and achieve image-style transfer. Therefore, G_{O-S} is used to translate all optical images into SAR-style images. By combining these generated images and the target labels of corresponding optical images, a middle domain $\{X_{Gen}, Y_{Opt}\}$, which has a closer distance to X_{SAR}, can be obtained and used for learning generic aircraft characteristics.

2.4. Local-Level Domain Adaptation with Multilayer Feature Alignment

Though image translation can reduce overall image divergence and achieve global-level domain adaptation, the local-level divergence caused by different aircraft characteristics still impedes the detector training process in both the pretrain and finetune stages. In the context of limited data, detectors are more prone to overfitting on the few training samples rather than learning generic aircraft characteristics, leading to limited aircraft-detection performance.

As the detector tends to extract discriminative object features during the training process, the difference of feature distribution between the two domains can reflect the local-level object divergences learned by the detector. Therefore, learning similar feature distributions of the two domains is critical for reducing the local-level divergences. On this basis, we propose multilayer feature alignment to further reduce the distribution distance between X_{Gen} and X_{SAR} in the feature space. Specifically, adversarial learning was adopted on multiple feature maps of the detector to force the detector to learn generic features in all scales. As shown in Figure 5, one domain classifier (DC) was added to each feature layer after the feature pyramid network (FPN) to align the feature distributions between the optical and SAR domains. While training the detector, the detection head generates prediction results based on the extracted features and helps the network to learn discriminative features for aircraft detection. By applying DC to all feature scales, aircraft features of various scales can be properly aligned and learned by the detector. Specially, in Faster RCNN, we denote the region proposal network and the FCN together as the detection head. DC is used to generate domain classification results and help the network to learn domain-invariant features between X_{Gen} and X_{SAR}.

As depicted in Figure 6, a DC consists of a gradient reverse layer (GRL), two convolution layers, and a sigmoid function. The GRL keeps the input feature unchanged and multiplies the gradient by a negative scalar during the back propagation. By reversing the gradient between the DC and the detector, the detector is more likely to learn domain-invariant features that are indistinguishable by the DC. In each DC, two 1×1 convolution layers are used to transform the feature map into domain classification results. Subsequently, a sigmoid function is employed to normalize the output. The domain classifiers are trained simultaneously with the detector, and their loss function adopts the cross entropy (CE) loss, which can be expressed as:

$$L_{DC} = \mathbb{E}_{x \sim X_{SAR}}[\log F(x)] + \mathbb{E}_{x \sim X_{Gen}}[\log(1 - F(x))] \tag{7}$$

where $F(\cdot)$ is the mapping from the input image to the DC output.

While training the detector, images from two domains, i.e., a training domain and an auxiliary domain, are sent into the network simultaneously. As shown in Figure 7, in the DA-P stage, the detector is pretrained on X_{Gen}, and X_{SAR} is the auxiliary domain. The training loss for the second stage is:

$$L_{S2} = L_{Det}(X_{Gen}, Y_{Opt}) + \alpha \cdot L_{DC} \tag{8}$$

where $L_{Det}(\cdot)$ denotes the original detection loss, and α is a weight parameter. In our practice, α was set to 1.0. The detector can learn to discriminate aircraft in X_{Gen} while focusing on the common features that are shared within both domains.
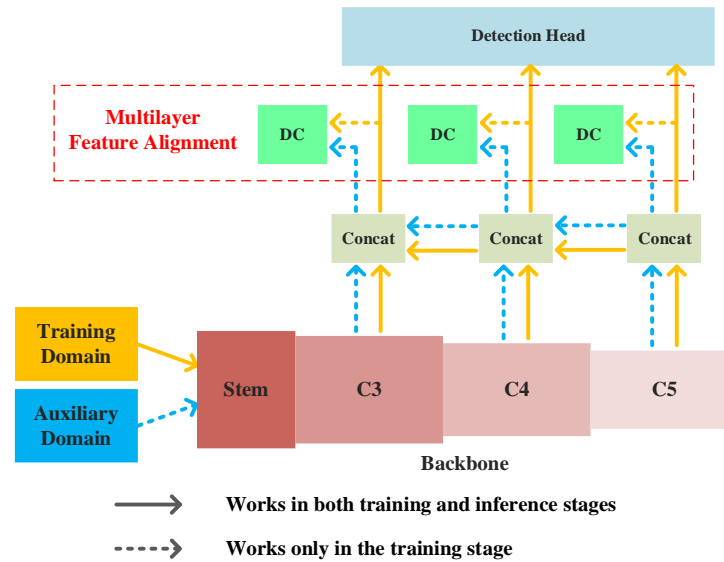

Figure 5. The structure of the detection network with the proposed multilayer feature alignment.
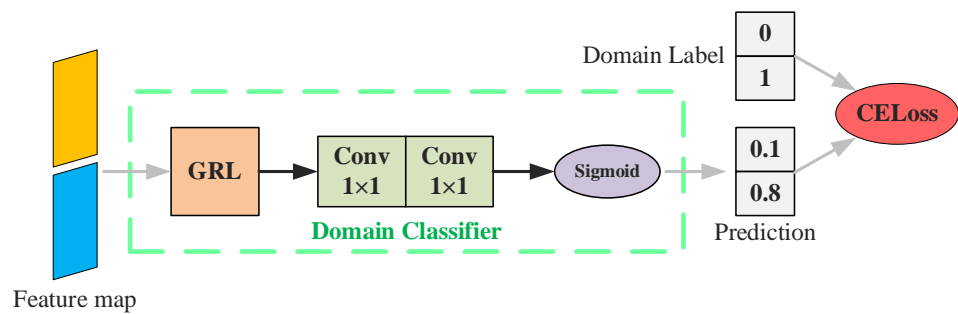

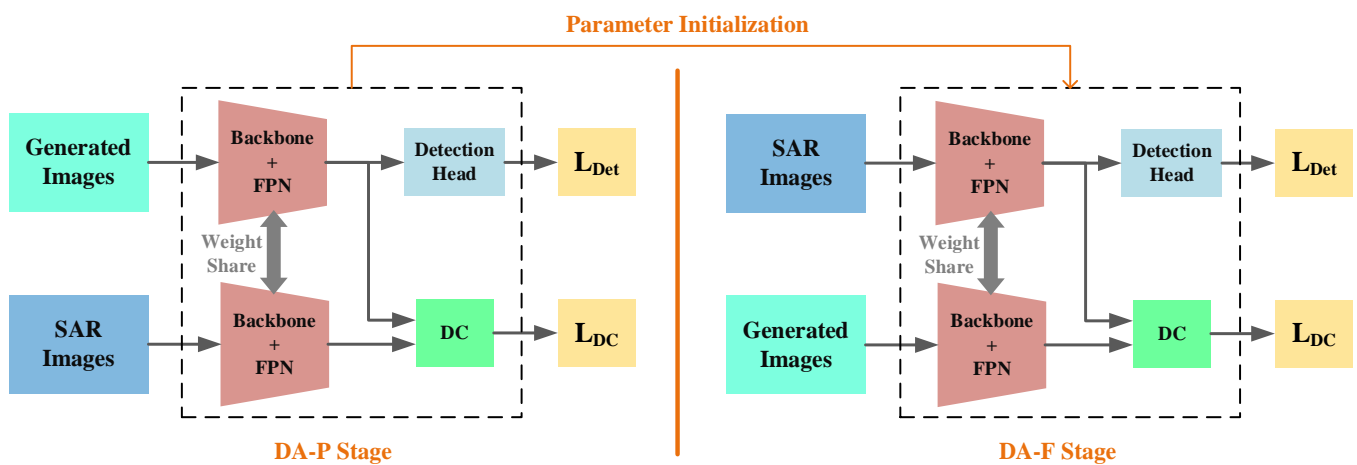**Figure 6.** The workflow of the domain classifier.

Figure 7. The training framework of the detection network in DA-P and DA-F stages.

In the DA-F stage, the detector is finetuned on X_{SAR}, and X_{Gen} is the auxiliary domain. The domain classifiers of the pretrained model are retained and continue to align feature distributions. The training loss for the last stage is:

$$L_{S3} = L_{Det}(X_{SAR}, Y_{SAR}) + \alpha \cdot L_{DC} \tag{9}$$

With the learned weights as model initialization and the domain adaptation constraint, the detector can effectively utilize the transferred knowledge and achieve better SAR aircraft detection performance. In the reference stage, the detector can detect SAR aircraft without DC and the auxiliary domain.

3. Experiment and Parameters Evaluation

3.1. Datasets

Since public SAR aircraft detection datasets are rare, we carefully selected 13 large-scale SAR images acquired by the GaoFen-3 system in spotlight mode, representing the target SAR domain. Additionally, we chose seven large-scale panchromatic images acquired by the GaoFen-2 system, serving as the source optical domain. These large-scale images cover different airports and have image sizes of several thousand pixels. The aircraft targets in these images are manually annotated with the reference of optical remote sensing images that cover the same area and are confirmed by SAR image interpretation experts.

Since the optical images are used as the source domain, all optical images were used for image translation and detector pretraining, and the SAR images were divided into a training set and a test set. According to their different image sizes, eight large-scale SAR images were used for model training, and the remaining five large-scale images were used for testing. To benefit the model training and testing, we adopted the sliding window method to crop these large-scale images into small image chips with a size of 512×512. Notably, image chips that contain no aircraft were used for model training in the image translation stage and were filtered in the detector training stages. The dataset information is given in Table 1. It can be seen that after removing these pure background image chips, the number of training samples for aircraft detection was quite limited. As illustrated in Figure 8, the target size of the optical dataset and that of the SAR dataset have similar ranges, where most aircraft have sizes larger than 20 pixels and smaller than 100 pixels.
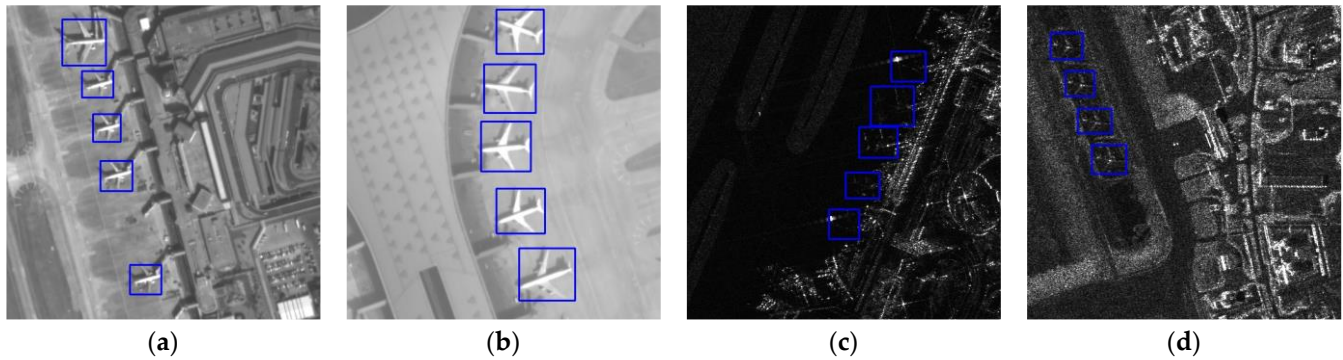
Some image chips of the dataset are shown in Figure 9. It can be seen that due to the different imaging mechanisms, the overall image style and target detail are completely different. However, the two domains share the similar airport background and aircraft structure, which could be helpful for object detection knowledge transfer.

(a)   **(b)**

Figure 8. The shape distribution of aircraft targets: (**a**) the optical domain; (**b**) the SAR domain.

Table 1. Details of the optical and SAR datasets.

Domain	Imaging Platform	Imaging Mode	Resolution	Chip Num. (All) (Train/Test)	Chip Num. (Filtered) (Train/Test)	Aircraft Num. (Train/Test)
Source	Gaofen-2	Panchromatic	0.8 m	2180/-	648/-	2083/-
Target	Gaofen-3	Spotlight	1.0 m	1120/912	300/189	1282/697

Figure 9. Image samples of the dataset: (**a**,**b**) optical images; (**c**,**d**) SAR images.

3.2. Evaluation Metrics

To assess the aircraft detection performance, precision rate (P), recall rate (R), F1-score, and average precision (AP) were used as evaluation metrics. They are defined as follows:

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

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

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{12}$$

$$AP = \int_0^1 P(R)dR \tag{13}$$

where TP, FP, and FN are the number of true positives, false positives, and false negatives, respectively. Taking into account the relatively small scale and dense arrangement of aircraft, we set the intersection over union (IoU) threshold for true positive detections as 0.45. The common metric for object detection AP_{50} is based on an IoU threshold of 0.5. And AP is calculated across the IoU thresholds from 0.5 to 0.95 with an interval of 0.05.

3.3. Implementation Setting

The proposed training framework contained two different training tasks: training a GAN for achieving global-level domain adaptation and training the detection network to detect aircraft targets. For image translation, the training of the CycleGAN followed the default settings in [38], where all the images in the optical domain and the images in the training set of the SAR domain were used for unpaired image translation. In the detector pretrain and finetune stages, the detector was trained on the source and target dataset, respectively, with the same training parameters. Specifically, Faster RCNN was trained for 72 epochs with a learning rate of 0.0025, and YOLOv3 was trained for 120 epochs with a learning rate of 0.001, where both learning rates were multiplied by 0.1 after the 2/3 and 11/12 of the total training epochs, respectively. The optimizer adopted an SGD optimizer with a momentum of 0.9 and a weight decay of 0.0001. In each training iteration, a mini-batch of four images with augmentations, including random crop and

horizontal flip, along with four randomly chosen images from the auxiliary domain were simultaneously sent into the network. In the experiments, we implemented all detection methods based on the MMDetection [39] toolkit and adopted the default parameter settings unless stated otherwise. The pytorch implementation code of the proposed method is at https://github.com/YUWEBBER/MDAT (accessed on 9 September 2023).

4. Results

4.1. Comparison with Other Methods

To evaluate the superiority of the proposed method, we employed several representative domain adaptation methods on the SAR aircraft detection dataset, including domain-adversarial neural network (DANN) [31], domain adaptation Faster RCNN (DAF) [32], and strong–weak distribution alignment (SWDA) [33]. DANN is the first work that introduced domain adaptation in deep feature learning by using adversarial learning, which is the cornerstone of many recent studies. DAF is a classic work that introduced domain adversarial adaptation into detection networks. The feature-level and instance-level design inspired other researchers and led to numerous domain adaptation studies that focused on the two-stage framework [41–44]. SWDA is also a representative domain adaptation method that is widely studied by researchers, which is also used in the study of SAR ship detection [45]. Among these methods, DANN is implemented on both Faster RCNN and YOLOv3 by directly adding a domain classifier to the feature maps of the backbone, while DAF and SWDA are implemented on Faster RCNN since they were designed based on the two-stage framework. The results in Table 2 show that the proposed method achieved the best detection performance. For the two-stage detector, the proposed method achieved an F1-score of 0.7478 and an AP_{50} of 66.8, which was the best performance among the domain adaptation methods. The result of the proposed method on YOLOv3 demonstrated an F1-score of 0.7762 and an AP_{50} of 0.69.6, which prove the effectiveness of multi-stage domain adaptation for overcoming the huge domain discrepancy between optical and SAR domains. Specifically, the proposed method achieved a recall rate of 0.7690, which was higher than other methods, indicating that the network can effectively learn aircraft structure and discover more aircraft targets.
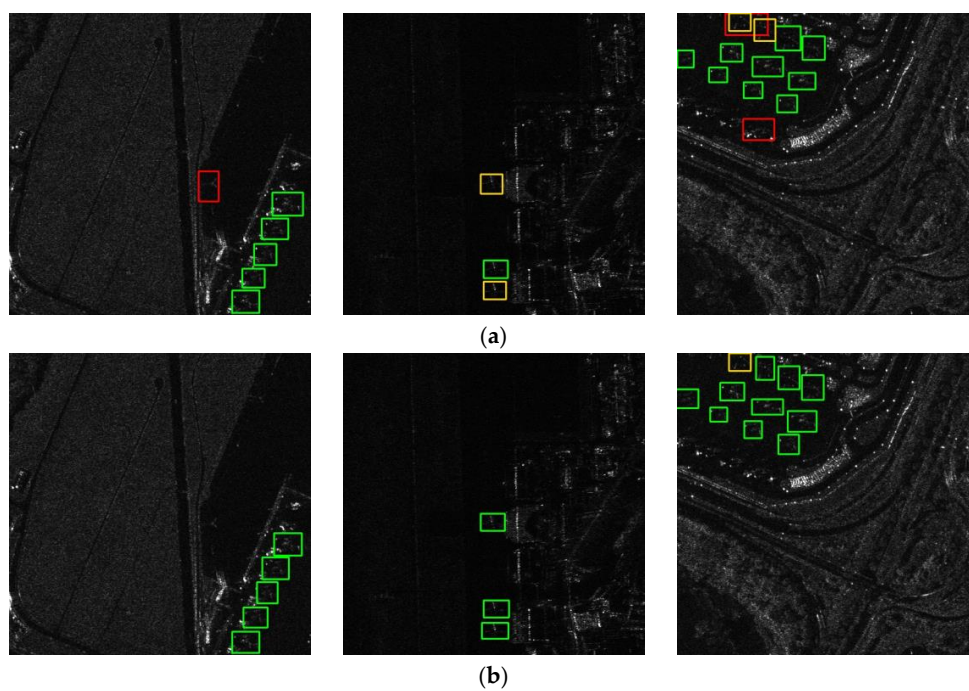
Table 2. Comparison with other methods.

Detector	Method	P	R	F1	AP	AP_{50}
Faster RCNN	DANN	0.7718	0.6987	0.7334	26.3	63.0
	DAF	0.7438	0.7374	0.7406	26.6	65.0
	SWDA	0.7646	0.7131	0.7380	27.7	65.5
	MDAT	0.7555	**0.7403**	**0.7478**	**27.8**	**66.8**
YOLOv3	DANN	0.7682	0.7274	0.7472	29.8	66.5
	MDAT	**0.7836**	**0.7690**	**0.7762**	**31.0**	**69.6**

Some visualized detection results are shown in Figures 10 and 11. As a note, green, yellow, and red rectangles represent correct detections, missed detections, and false alarms, respectively. In Figure 10, it can be seen that compared to other domain adaptation methods, the proposed MDAT can achieve better overall performance and make a better balance between precision and recall rate. In the third row of Figure 10, there is an aircraft that can be easily merged with background buildings and is missed by other methods, while the proposed method correctly detected it, indicating a better learning effect. From Figure 11, both the missed detections and false alarms of MDAT are fewer than those of DANN, proving the superiority of the proposed method for boosting SAR aircraft detection performance.

(**a**)　　　　　(**b**)　　　　　(**c**)　　　　　(**d**)

Figure 10. Visualized aircraft detection results of different domain adaptation methods with Faster RCNN: (**a**) DANN; (**b**) DAF; (**c**) SWDA; (**d**) MDAT.

(**a**)

(**b**)

Figure 11. Visualized aircraft detection results of different domain adaptation methods with YOLOv3: (**a**) DANN; (**b**) MDAT.

4.2. Ablation Study

To inspect the effectiveness of the proposed method for boosting SAR aircraft detection performance, we added the proposed components step by step, as shown in Table 3, where P-F, IT, DA-P, and DA-F mean the detector is directly pretrained on the source domain, the images of the source domain are translated, the multilayer feature alignment is used in the pretraining stage, and the multilayer feature alignment is used in the finetuning stage, respectively. It can be seen that when directly using the vanilla P-F method, the AP for both detectors was marginally increased. Higher AP indicates that the network can achieve higher precision with high IoU thresholds, which means a higher target localization ability. Therefore, pretraining on the optical domain can help the network to learn aircraft structure and scale characteristics, which can contribute to target localization. At the same time, the AP_{50} metric of Faster RCNN decreased by 3.5, and that of YOLOv3 decreased by 2.0, indicating a lower ability to discover aircraft targets. This performance drop depicts that the network can be misled by optical backgrounds and fails to discriminate aircraft from complex building interferences in SAR images. Therefore, the P-F framework is not capable of efficiently transferring knowledge from the optical domain, leading to a limited transfer learning effect.

Table 3. Ablation study results.

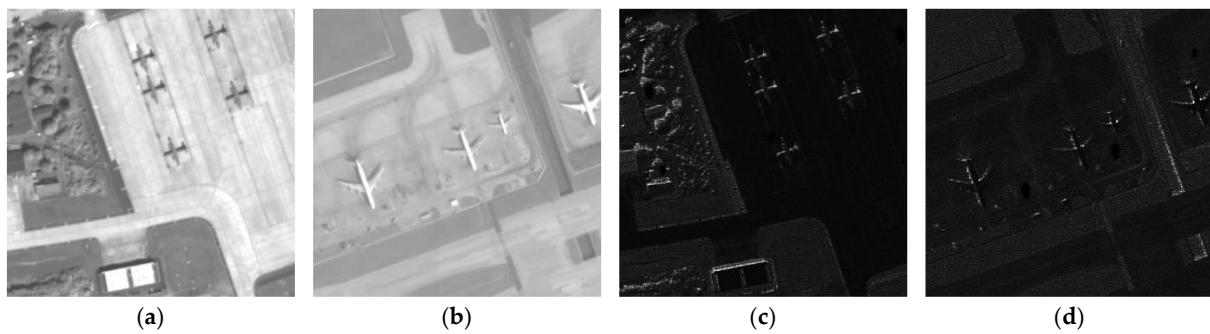| P-F	IT	DA-P	DA-F	Faster RCNN		YOLOv3	
				AP	AP_{50}	AP	AP_{50}
				25.4	65.0	28.4	67.5
✓				25.9	61.5	28.8	65.5
✓	✓			26.7	64.4	30.0	67.2
✓	✓	✓		27.2	65.8	30.6	68.0
✓	✓	✓	✓	**27.8**	**66.8**	**31.0**	**69.6**

On this basis, different from directly using the pretrained weights from the optical domain, pretraining on the translated images that are generated after image-level domain adaptation can significantly boost the detection performance, i.e., there was a 0.8 increase in AP and a 2.9 increase in AP_{50} for Faster RCNN and a 1.2 increase in AP and 1.7 increase in AP_{50} for YOLOv3. This performance increment is beneficiated from the reduced global-level image divergence between the pretrain and finetune stages. As the transferred images share more similar characteristics with SAR images, it is easier for the network to learn generic features that are beneficial to SAR aircraft detection tasks. Furthermore, adopting the multilayer feature alignment in the pretrain stage and finetune stage can both reduce the feature difference and achieve distribution alignment, improving the effect of transfer learning. Compared to training from scratch, the proposed method can boost the detection performance on both Faster RCNN and YOLOv3, proving its effectiveness in improving SAR aircraft detection accuracy.

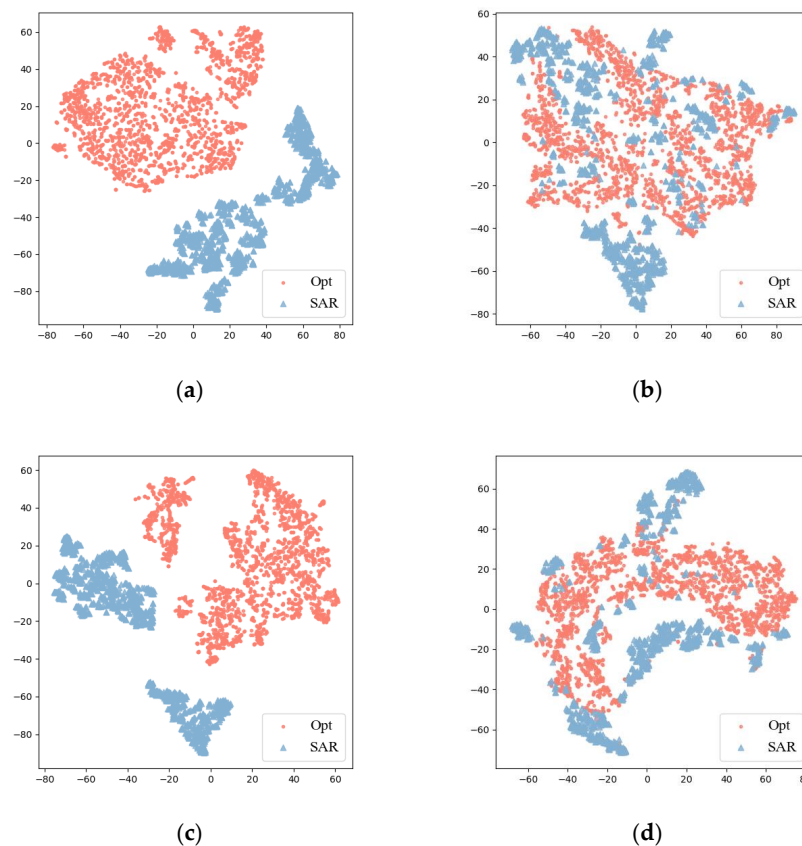### 4.3. Effect of Domain Adaptation

The proposed method uses three stages to adopt domain adaptation on image and feature levels. First, we adopted CycleGAN for transferring optical aircraft to SAR-like aircraft, and the results are given in Figure 12. It can be seen that the generated images have a similar overall visual appearance to SAR images, as the background is darkened and the aircraft show a strong intensity form, which resembles that of SAR images. The reduced image divergence can help the network learn common features that exist both in optical and SAR images, leading to a better transferring effect.

Furthermore, to explore the effectiveness of the proposed method for reducing domain discrepancy in the feature space, we used t-SNE to visualize the features extracted by the detection network. Specifically, we took the mean value of each channel before flattening the feature map to form a feature vector. By concatenating the feature vectors of all three

layers, we could obtain a representative feature that indicates the input image, which can be used for t-SNE. As depicted in Figure 13a,c, when using the vanilla P-F method for transferring learning, the features extracted by the trained detector were decentralized, and the two domains could be easily discriminated. After adopting the proposed MDAT method, the features of the two domains were blended together and were difficult to discern from each other. This result proves that the proposed method can effectively reduce the feature distribution divergence between the optical and SAR domains. As the network tends to extract domain-invariant features that are shared between optical and SAR images, the aircraft structural characteristics in optical images can be transferred for SAR aircraft detection more effectively, boosting the detection accuracy.

Figure 12. Results of image translation: (**a**,**b**) optical images; (**c**,**d**) the corresponding translated SAR-like images.

Figure 13. Feature distribution visualization results via t-SNE: (**a**) Faster RCNN with vanilla P-F; (**b**) Faster RCNN with the proposed MDAT; (**c**) YOLOv3 with vanilla P-F; (**d**) YOLOv3 with the proposed MDAT.
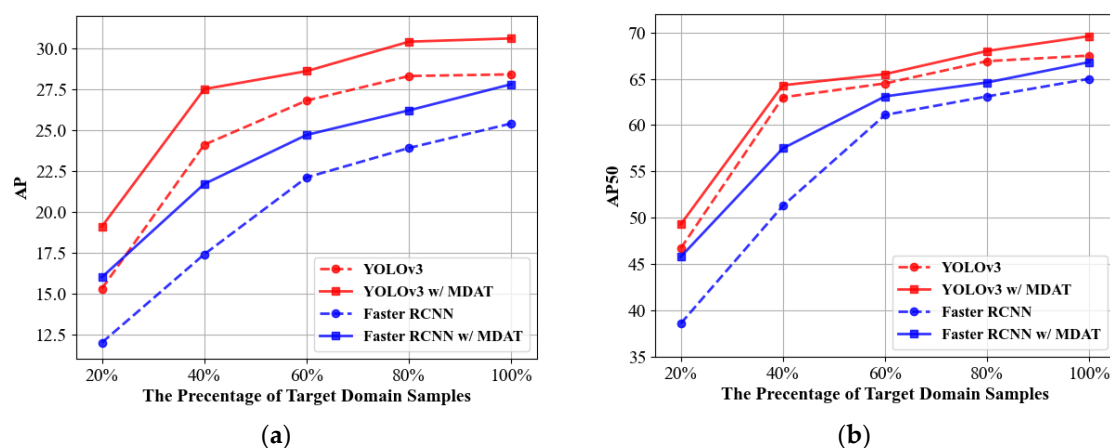
4.4. Analysis of Training Sample Scale

Due to the shortage of labeled SAR aircraft data samples, the detection accuracy when less training data were available highly reflects the robustness of a detector. Therefore, the proposed method was trained with part of the training data of the target domain. Specifically, we trained the model with all generated images and finetuned the model with 20%, 40%, 60%, and 80% of the SAR images to explore the robustness of the proposed method. The whole test set was used to evaluate detection performance. The results are depicted in Tables 4 and 5. It can be seen that the detection performance had a positive relevance with the data scale, and few training data may lead to a limited detection performance. With more SAR training data, the network can access more SAR aircraft and is more prone to learn the difference between aircraft targets and background interferences. By adopting the proposed method, the AP metric can be improved by approximately 2.0 or more. As visualized in Figure 14, after adopting the proposed MDAT, the performance of both Faster RCNN and YOLOv3 under all training data scales improved stably. This improvement indicates that with the same training data quantity, using the proposed training framework can effectively enhance the performance of aircraft detection in SAR images, which proves the robustness of the proposed method for efficiently transferring aircraft knowledge from the optical domain.

Table 4. Results of the proposed method on Faster RCNN with different training sample scales.

Training Percentage	Baseline		Proposed Method	
	AP	AP_{50}	AP	AP_{50}
20%	12.0	38.6	16.0	45.8
40%	17.4	51.3	21.7	57.5
60%	22.1	61.1	24.7	63.1
80%	23.9	63.1	26.2	64.6
100%	25.4	65.0	27.8	66.8

Table 5. Results of the proposed method on YOLOv3 with different training sample scales.

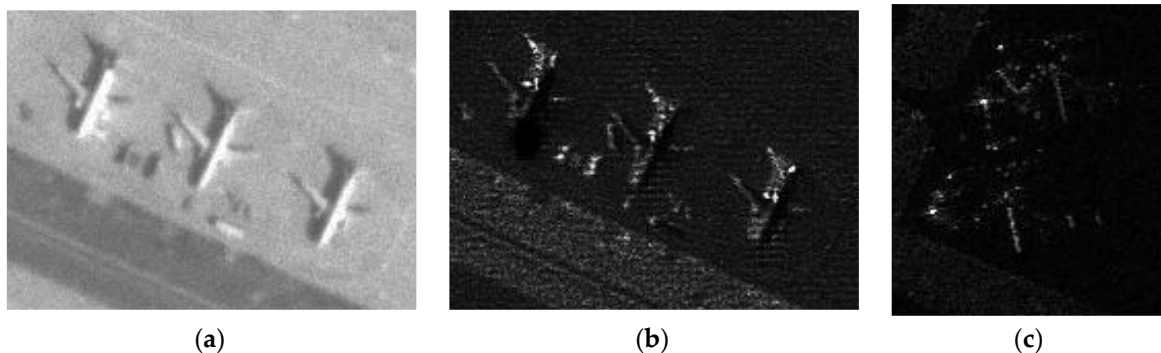| Training Percentage	Baseline		Proposed Method	
	AP	AP_{50}	AP	AP_{50}
20%	15.3	46.7	19.1	49.3
40%	24.1	63.0	27.5	64.3
60%	26.8	64.5	28.6	65.5
80%	28.3	66.9	30.4	68.0
100%	28.4	67.5	30.6	69.6

Figure 14. Detection performance with different training data rates: (**a**) AP results; (**b**) AP_{50} results.

5. Discussion

Nowadays, research on algorithms for aircraft detection in SAR images predominantly focuses on network design and feature extraction, while studies on transfer learning from other domains are scarce. In this paper, to efficiently utilize knowledge from the optical domain and improve SAR aircraft detection performance, a novel multi-stage domain adaptation training framework was proposed.

In consideration of the limited data quantity of optical and SAR images, CycleGAN was first employed to perform image translation and reduce global-level domain divergences. The generated images shown in Figure 12 exhibit a similar brightness and contrast to real SAR images. Compared to the original optical images, the flat runway area appears darker, while the uneven ground area exhibits an overall brighter characteristic, which is consistent with real SAR images. It is believed, in domain adaptation studies, that a high similarity can lead to a higher transfer effect. Therefore, the generated images can facilitate the knowledge transfer from optical to SAR, as evidenced by the results in Table 3. While reducing the overall image divergences, the generated images failed to achieve satisfactory target detail learning. As depicted in Figure 15, aircraft in SAR images exhibit incomplete structures and generally have lower scattering intensity compared to surrounding buildings, whereas the generated aircraft have complete structures. Consequently, multilayer feature alignment was further adopted to mitigate local-level domain divergences. The visualized feature distributions in Figure 13 demonstrate the effectiveness of the proposed method in eliminating the domain shift between optical and SAR images. A comparison with other domain adaptation methods indicated a higher recall rate of the proposed method, which is proof of the efficiently transferred aircraft structure knowledge. According to the results in Tables 4 and 5, the proposed MDAT can stably improve the average precision of SAR aircraft, achieving effective knowledge transfer.

| (a) | (b) | (c) |

Figure 15. Aircraft detail comparisons: (**a**) optical aircraft; (**b**) generated aircraft; (**c**) SAR aircraft.

Though the SAR aircraft detection performance was improved, the proposed MDAT framework still requires a complex training process to overcome the domain discrepancy between optical and SAR. In the future, we will continue to explore the effects of domain adaptation methods on SAR aircraft detection tasks and focus on further improving the knowledge transfer efficiency. More studies on the diversity and quantity of source domain data are needed. Additionally, introducing the prior knowledge of SAR images and aircraft scattering characteristics is also a viable approach to achieve better adaptation performance.

6. Conclusions

In this paper, we proposed a multi-stage domain adaptation training framework for SAR aircraft detection, which can efficiently transfer knowledge from optical images by gradually reducing the domain discrepancy in three stages. In the image translation stage, CycleGAN was employed to transfer optical images into fake SAR-style images that are used for detector pretraining. With reduced global-level divergences from real SAR images, the efficiency of the model pretraining was effectively improved. Furthermore, the

proposed multilayer feature alignment was integrated into the detector in the pretrain and finetune stages to eliminate the local-level aircraft divergences and to enable the network to focus on domain-invariant features. Experiments were carried out for aircraft detection using GaoFen-2 and GaoFen-3 images. The results revealed that the detection performance of Faster RCNN and YOLOv3 was significantly improved, which verifies the excellence of the proposed framework.

Author Contributions: Conceptualization, W.Y.; methodology, W.Y. and J.L.; software, W.Y.; validation, W.Y. and J.L.; formal analysis, W.Y.; investigation, Z.W. and W.Y.; data curation, W.Y.; writing—original draft preparation, W.Y.; writing—review and editing, Z.W.; visualization, J.L.; supervision, Z.Y. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data sharing is not applicable to this article.

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

References

1. Novak, L.M.; Halversen, S.D.; Owirka, G.J.; Hiett, M. Effects of Polarization and Resolution on SAR ATR. *IEEE Trans. Aerosp. Electron. Syst.* **1997**, *33*, 102–116. [CrossRef]
2. Nunziata, F.; Migliaccio, M.; Brown, C.E. Reflection Symmetry for Polarimetric Observation of Man-Made Metallic Targets at Sea. *IEEE J. Ocean. Eng.* **2012**, *37*, 384–394. [CrossRef]
3. Dellinger, F.; Delon, J.; Gousseau, Y.; Michel, J.; Tupin, F. SAR-SIFT: A SIFT-like Algorithm for SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 453–466. [CrossRef]
4. Zhai, L.; Li, Y.; Su, Y. Inshore Ship Detection via Saliency and Context Information in High-Resolution SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1870–1874. [CrossRef]
5. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
6. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
7. 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]
8. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2961–2969.
9. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
10. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
11. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
12. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. In Proceedings of the International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9627–9636.
13. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
14. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430.
15. Bi, H.; Deng, J.; Yang, T.; Wang, J.; Wang, L. CNN-Based Target Detection and Classification When Sparse SAR Image Dataset Is Available. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6815–6826. [CrossRef]
16. Zou, B.; Qin, J.; Zhang, L. Vehicle Detection Based on Semantic-Context Enhancement for High-Resolution SAR Images in Complex Background. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4503905. [CrossRef]
17. Feng, Y.; Chen, J.; Huang, Z.; Wan, H.; Xia, R.; Wu, B. A Lightweight Position-Enhanced Anchor-Free Algorithm for SAR Ship Detection. *Remote Sens.* **2022**, *14*, 1908. [CrossRef]
18. Yu, W.; Wang, Z.; Li, J.; Luo, Y.; Yu, Z. A Lightweight Network Based on One-Level Feature for Ship Detection in SAR Images. *Remote Sens.* **2022**, *14*, 3321. [CrossRef]
19. Sun, Z.; Dai, M.; Leng, X.; Lei, Y.; Xiong, B.; Ji, K.; Kuang, G. An Anchor-Free Detection Method for Ship Targets in High-Resolution SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7799–7816. [CrossRef]

20. Chen, L.; Weng, T.; Xing, J.; Pan, Z.; Yuan, Z.; Xing, X.; Zhang, P. A New Deep Learning Network for Automatic Bridge Detection from SAR Images Based on Balanced and Attention Mechanism. *Remote Sens.* **2020**, *12*, 441. [CrossRef]
21. He, C.; Tu, M.; Xiong, D.; Tu, F.; Liao, M. A Component-Based Multi-Layer Parallel Network for Airplane Detection in SAR Imagery. *Remote Sens.* **2018**, *10*, 1016. [CrossRef]
22. Diao, W.; Dou, F.; Fu, K.; Sun, X. Aircraft Detection in SAR Images Using Saliency Based Location Regression Network. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 2334–2337. [CrossRef]
23. Wang, J.; Xiao, H.; Chen, L.; Xing, J.; Pan, Z.; Luo, R.; Cai, X. Integrating Weighted Feature Fusion and the Spatial Attention Module with Convolutional Neural Networks for Automatic Aircraft Detection from Sar Images. *Remote Sens.* **2021**, *13*, 910. [CrossRef]
24. Guo, Q.; Wang, H.; Xu, F. Scattering Enhanced Attention Pyramid Network for Aircraft Detection in SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 7570–7587. [CrossRef]
25. Zhao, Y.; Zhao, L.; Li, C.; Kuang, G. Pyramid Attention Dilated Network for Aircraft Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 662–666. [CrossRef]
26. Kang, Y.; Wang, Z.; Fu, J.; Sun, X.; Fu, K. SFR-Net: Scattering Feature Relation Network for Aircraft Detection in Complex SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5218317. [CrossRef]
27. Chen, L.; Luo, R.; Xing, J.; Li, Z.; Yuan, Z.; Cai, X. Geospatial Transformer Is What You Need for Aircraft Detection in SAR Imagery. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5225715. [CrossRef]
28. Li, J.; Qu, C.; Shao, J. Ship Detection in SAR Images Based on an Improved Faster R-CNN. In Proceedings of the SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), Beijing, China, 13–14 November 2017; pp. 1–6.
29. Bao, W.; Huang, M.; Zhang, Y.; Xu, Y.; Liu, X.; Xiang, X. Boosting Ship Detection in SAR Images with Complementary Pretraining Techniques. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8941–8954. [CrossRef]
30. Ganin, Y.; Lempitsky, V. Unsupervised Domain Adaptation by Backpropagation. In Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6–11 July 2015; Volume 2, pp. 1180–1189.
31. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn.Res.* **2016**, *17*, 2096-2030.
32. 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 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3339–3348.
33. Saito, K.; Ushiku, Y.; Harada, T.; Saenko, K. Strong-Weak Distribution Alignment for Adaptive Object Detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 6949–6958. [CrossRef]
34. Rostami, M.; Kolouri, S.; Eaton, E.; Kim, K. Deep Transfer Learning for Few-Shot SAR Image Classification. *Remote Sens.* **2019**, *11*, 1374. [CrossRef]
35. Chen, Z.; Zhao, L.; He, Q.; Kuang, G. Pixel-Level and Feature-Level Domain Adaptation for Heterogeneous SAR Target Recognition. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4515205. [CrossRef]
36. Song, Y.; Li, J.; Gao, P.; Li, L.; Tian, T.; Tian, J. Two-Stage Cross-Modality Transfer Learning Method for Military-Civilian SAR Ship Recognition. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4506405. [CrossRef]
37. 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]
38. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2242–2251. [CrossRef]
39. Chen, K.; Wang, J.; Pang, J.; Cao, Y.; Xiong, Y.; Li, X.; Sun, S.; Feng, W.; Liu, Z.; Xu, J.; et al. MMDetection: Open MMLab Detection Toolbox and Benchmark. *arXiv* **2019**, arXiv:1906.07155.
40. Yu, W.; Wang, Z.; Li, J.; Wang, Y.; Yu, Z. Unsupervised Aircraft Detection in SAR Images with Image-Level Domain Adaption from Optical Images. In Proceedings of the International Conference on Computer Vision and Pattern Analysis (ICCPA), Hangzhou, China, 31 March–2 April 2023.
41. Li, C.; Du, D.; Zhang, L.; Wen, L.; Luo, T.; Wu, Y.; Zhu, P. Spatial Attention Pyramid Network for Unsupervised Domain Adaptation. In Proceedings of the Computer Vision—ECCV, Glasgow, UK, 23–28 August 2020; Volume 12358 LNCS, pp. 481–497.
42. Zheng, Y.; Huang, D.; Liu, S.; Wang, Y. Cross-Domain Object Detection through Coarse-to-Fine Feature Adaptation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 13763–13772.
43. Fan, Q.; Zhuo, W.; Tai Tencent, Y.-W. Few-Shot Object Detection with Attention-RPN and Multi-Relation Detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 4013–4022.

44. Vibashan, V.S.; Gupta, V.; Oza, P.; Sindagi, V.A.; Patel, V.M. MEGA-CDA: Memory Guided Attention for Category-Aware Unsupervised Domain Adaptive Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 4516–4526.
45. Xu, C.; Zheng, X.; Lu, X. Multi-Level Alignment Network for Cross-Domain Ship Detection. *Remote Sens.* **2022**, *14*, 2389. [CrossRef]