SEFEPNet: Scale Expansion and Feature Enhancement Pyramid Network for SAR Aircraft Detection With Small Sample Dataset

Peng Zhang ⓘ, Hao Xu ⓘ, Tian Tian ⓘ, Peng Gao ⓘ, Linfeng Li, Tianming Zhao ⓘ, Nan Zhang, and Jinwen Tian

Abstract—Aircraft detection in synthetic aperture radar (SAR) images is still a challenging research task because of the insufficient public data, the difficulty of multiscale target detection, and the complexity of background interference. In this article, we construct a public SAR aircraft detection dataset (SADD) with complex background and interference objects to facilitate the research in SAR aircraft detection. Then, we propose the scale expansion and feature enhancement pyramid network as the SADD baseline. It uses a four-scale fusion method to combine the shallow position information with the deep semantic information, effectively adapting to the multiscale target detection in SAR images, significantly improving the detection effect of small targets. The feature enhancement pyramid structure is connected behind the backbone network to weaken the background texture and highlight the target to achieve feature enhancement, improving the ability to extract target features in complex backgrounds. Finally, to further improve the detection performance of the small-scale SAR aircraft dataset, we propose a domain adaptive transfer learning method. Experiments on SADD show that this method can significantly improve the recall rate and F1 score. At the same time, we find that the transfer effect of using homologous but different types of targets as the source domain is better than those of heterologous but same types of targets in SAR aircraft detection, which is instructive for future research.

Index Terms—Feature enhancement pyramid (FEP), synthetic aperture radar (SAR) aircraft detection, transfer learning.

I. INTRODUCTION

COMPARED with passive sensors, such as optics and infrared, synthetic aperture radar (SAR) has the unique advantages of all-weather and day and night. It has outstanding strategic significance in military fields, such as battlefield situational awareness [1], typical target recognition [2], and precision guidance [3]. Aircraft is an essential target in the civil field. The detection of it contributes to the effective management of airports. In the military field, the efficient and accurate acquisition of aircraft targets in the airport and airspace is of great importance, which can help to acquire battlefield military information and make battle plans in real time. Therefore, detecting aircraft targets based on SAR images is a significant research direction.

Unlike optical images, SAR imaging has a longer wavelength, a more complex imaging mechanism, and a more difficult visual interpretation of the imaging results. Therefore, SAR aircraft target detection faces some challenges. First, the target in the SAR image is discontinuous, which is composed of multiple discrete irregular scattering centre bright spots. However, the key semantic information for object recognition is hidden between these scattering centres. It is not easy to detect the complete aircraft target in this case. In addition, there are significant differences in target scales and many weak and small targets in SAR images, which makes detection difficult. Moreover, the interference of complex background also brings difficulties to SAR aircraft target detection. A large number of background highlight scattering points are distributed around the aircraft target, which will be confused with aircraft target components to a certain extent, making it challenging to locate and identify the aircraft accurately.

In the previous work, the traditional SAR aircraft target detection algorithms are roughly divided into three stages: 1) detection, 2) identification, and 3) recognition. First, suspicious targets, including targets and false alarms, are detected by detection algorithms, such as constant false alarm rate (CFAR) [4] and its derivative and improved algorithms [5], [6]. In the identification stage, targets and false alarms can be distinguished by other aircraft features, such as contour, shape, texture, and scattering centre. As Chen *et al.* [7] analyzed the scattering features of civil aircraft in high-resolution TerraSAR-X images. On this basis, the Harris–Laplace corner detector was used to extract the strong scattering points and proposed to describe them by using the convex dot vector to realize the extraction and description of the scattering features of aircraft targets. In the final recognition stage, the template matching method, model-based method, and machine learning method [8] are often used to realize the fine-grained division of aircraft types. However, traditional SAR aircraft detection algorithms often rely on hand-designed features, and the implementation process is complex. It is difficult to achieve good detection results in complex scenes, and its generalization ability is poor.

Target detection algorithm based on deep learning does not need manual design features, and has the characteristics of a simple implementation process and high detection accuracy, so

it has been widely used in the field of SAR images. Target detection algorithms based on deep learning can be divided into two categories: 1) two-stage detection algorithms, such as R-CNN series [9]–[11], which have high detection accuracy; and 2) one-stage detection algorithms, such as SSD series [12]–[14] and YOLO Series [15]–[17], which have very fast detection speed. In the task of SAR aircraft target detection, according to actual requirements, scholars usually improve the model based on the above algorithms to obtain good detection results. Just as He *et al.* [18] improved the algorithm based on YOLO [15] and combined the target component information to achieve satisfactory results in Terra SAR-X data. However, when the sizes of targets in SAR images are very different and there are many small targets, the existing algorithms are difficult to achieve satisfactory results. Moreover, in the case of complex background interference, the extracted target features are insufficient, resulting in poor detection accuracy. In addition, methods based on deep learning require a large number of samples for supervised training. Insufficient samples can easily lead to overfitting and affect the detection effect.

Most of the datasets used in SAR aircraft detection research are not publicly available, so it is impossible to train a robust model using large amounts of SAR aircraft data. In this case, using a small amount of SAR aircraft image data to improve the detection performance of the model is a significant research direction. A suitable solution to the small sample problem is to use transfer learning methods. The main idea of transfer learning is to use the existing knowledge (source domain) to learn new knowledge (target domain), find the similarity between them, and use the data in the source domain to transfer the knowledge to the target domain to establish the model. According to the learning method [19], transfer learning methods can be divided into four categories: 1) instance-based transfer learning, 2) feature-based transfer learning, 3) model-based transfer learning, and 4) relation-based transfer learning. The model-based transfer learning method has been widely used in convolutional neural networks, adopting parameter sharing. At present, a well-established paradigm is to use large-scale data to pretrain the model (e.g., ImageNet [20]) and then use the model-based transfer learning method to fine-tune the model on target tasks, which usually have less training data. In the field of natural image processing, model-based transfer learning methods have achieved state-of-the-art results on many tasks, including object detection [11], [12], [17], image segmentation [21], [22], and action recognition [23], [24]. In the SAR aircraft detection mission, due to the lack of public data, Li *et al.* [25] tried to use public SAR ship data for pretraining, which improved the detection performance of the model to a certain extent. However, due to the different categories of source domain data and target domain data, the similarity between domains is not obvious enough, so it is not easy to achieve better detection results only by using the model-based transfer method. Therefore, designing a feature-level transfer learning method for SAR aircraft detection with small sample datasets is necessary.

Despite the rapid development of SAR aircraft detection methods, there are still some obstacles to further development of the field, especially the lack of a common benchmark dataset for algorithm evaluation. Therefore, the motivation of this article
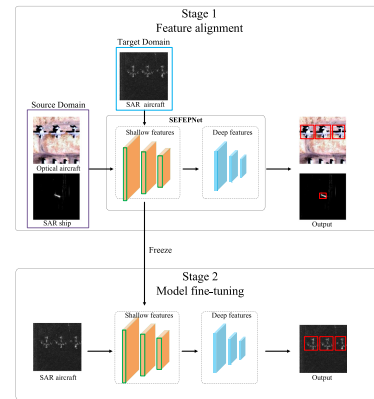

Fig. 1. Schematic diagram of the feature adaptive transfer learning method. It first aligns the shallow features of the source domain data with the target domain data, and then uses the target domain data to fine-tune the model.

can be summarized as follows. First, we aim to establish a public SAR aircraft detection dataset (SADD) to provide a benchmark for evaluating different algorithms and facilitate other scholars' research. Second, for the challenges of complex background interference and target size diversity in SADD, we plan to design a scale expansion and feature enhancement pyramid network (SEFEPNet) as the baseline of SADD. Third, although we provide the first public dataset in the related field, our SADD scale is relatively small due to the limitations of data sources and difficulties in labeling. Therefore, we plan to use the transfer learning method to improve detection performance. We believe that the selection of source domain data is as important as the design of the method, so we also intend to explore the impact of source domain data types on SAR aircraft detection performance, which can guide scholars on how to select source domain data to effectively improve the performance of SAR aircraft detection and has practical application value in real scenes.

In this article, we propose the SEFEPNet architecture. Our SEFEPNet adopts a four-scale feature fusion method to combine semantic features of different depths, effectively alleviating the problem of difficult detection of small and weak targets and adapting to multiscale target detection. In addition, our SEFEPNet adopts the feature enhanced spatial pyramid pooling structure, which can strengthen the target feature and weaken the background texture information, helping to improve the ability to extract target features in complex backgrounds. Moreover, the structure can fully extract local features through the pooling layer of various sizes and greatly enrich the information of the receiver field in combination with global features, which is conducive to the integrity detection of SAR aircraft targets and can also adapt to the problem of the large difference in target size. Experimental results show that our SEFEPNet is superior to the general algorithms in SAR target detection.

To solve the problem of insufficient samples of SAR aircraft, we propose a feature-adaptive transfer learning method, as shown in Fig. 1. This method consists of two stages: first, train the network with the source domain data and align the distribution of shallow features with the target domain; then, freeze the shallow network and fine-tune the model using only the target

domain data to get the final output. In this method, an adaptive layer is added to the network, and the distance between the extracted source domain features and the target domain features is measured by maximum mean discrepancy (MMD), so that the knowledge learned from the source domain data is more consistent with the target domain feature distribution. Since there is no open-source SAR aircraft dataset, we, respectively, use open optical aircraft images [26] and SAR ship images [27] as a source domain for transfer learning, which has some intuitive similarities with SAR aircraft targets. Experiments show that this domain-adaptive transfer learning method can significantly improve detection performance. Moreover, we found that in the field of SAR aircraft detection, using homologous but different types of targets as the source domain is better than those of heterologous but same types of targets, which is instructive for future research.

The main contributions of this article are as follows.

1) First, for the lack of SAR aircraft detection data, we construct a public dataset (SADD), which provides a benchmark for comparing different algorithms to facilitate the research of other scholars.

2) Second, for the complex background interference and target multiscale problems in SAR aircraft detection, we propose SEFEPNet as the baseline of SADD, which can effectively improve the detection accuracy and is superior to the mainstream object detection algorithms.

3) Third, to further improve the detection performance of the small-scale SADD, we propose a domain adaptive transfer learning method. Moreover, for SAR aircraft target detection, we find that using homologous but different types of targets as the source domain is better than those of heterologous but same types of targets, which has guiding significance and practical application value for future research.

The rest of this article is organized as follows. Section II introduces the related work of SAR aircraft target detection and the related work of transfer learning in SAR. Section III introduces the SAR aircraft dataset and production details. Section IV explains the SEFEPNet and the feature-adaptive transfer learning method. Section V provides ablation experiments of SEFEPNet structure, as well as rich comparison experiments with other general detection algorithms. Section VI further discusses and analyzes the results of transfer learning. Finally, Section VII concludes this article.

II. Related Work

A. SAR Aircraft Detection With Deep Learning

The general process of aircraft target detection and recognition system in SAR image can be roughly divided into five steps: 1) image preprocessing, 2) airport detection, 3) aircraft target rough detection, 4) false alarm elimination, and 5) aircraft target recognition. Since there are many methods to eliminate false alarm targets, rough detection of aircraft targets is the most important part that affects the performance of SAR aircraft target detection performance. In the rough detection of SAR aircraft targets, the aim is to detect the real aircraft targets as much as possible. Readers can see recent surveys [28], [29], and [30] for

more details in the SAR aircraft detection. Here, we only review several milestone works.

The traditional SAR target detection algorithms need to design the features manually, the steps are complex, and the generalization performance is poor. To solve the shortcomings of traditional algorithms, Wang Siyu *et al.* [31] proposed an aircraft target detection algorithm in SAR images based on data enhancement and the convolutional neural network. First, candidate slices are selected by a sliding window, and then the candidate slices are identified by a convolutional neural network. However, the network structure of this algorithm is relatively simple, and there are many missing detections. To improve the detection accuracy of the algorithm, Zhao Danxin *et al.* [32] proposed a new method of aircraft target detection based on RESNET [33]. The method first constructs the graph and template pyramid for multiscale detection. It then uses the fully convolutional network to extract the context information of different layers to achieve high-precision detection of aircraft targets. Later, Guo Qian *et al.* [34] proposed an algorithm based on FPN to detect aircraft in high-resolution images. This structure combines the high-resolution information of basic features with the high semantic information of depth features, contributing to the accuracy of detection results. However, there are many small targets in SAR images, and the detection effect of the existing algorithms is poor. To solve the problem that small targets are challenging to detect, An *et al.* [35] adopted a feature pyramid network based on rotating minimum adjacency rectangular frame to improve the detection effect of small targets to a certain extent.

However, for the small and weak aircraft targets in SAR images, the existing algorithms are still challenging to achieve satisfactory results. In addition, an SAR image is prone to a significant difference in the size distribution of aircraft targets, which increases the difficulty of SAR aircraft detection. Therefore, it is an urgent problem to improve the detection accuracy of small and weak targets and fully extract target features of different sizes in SAR images, which is also one of the motivations of this article.

B. Transfer Learning of SAR Images

Transfer learning is a method of transferring knowledge from one domain (source domain) to another domain (target domain) to obtain better learning results. The model-based transfer method is commonly used in deep convolutional neural networks. It initializes the model by using the parameters learned from the large source domain dataset, and then fine-tunes [36] the initialized model with a small number of labeled target domain samples, which can prevent model overfitting and effectively improve the test performance of the target domain dataset. The following are the milestone works of transfer learning in SAR image target recognition and detection.

In the field of target detection and recognition in SAR images, to solve the problem of poor model generalization performance caused by insufficient target domain data, Shengna Wu *et al.* [37] proposed a convolutional neural network algorithm based on model transfer. In this method, source domain data were obtained from ESA ENVISAT/ASAR, and target domain data were obtained from ESA ERS-2/SAR. With the adoption of
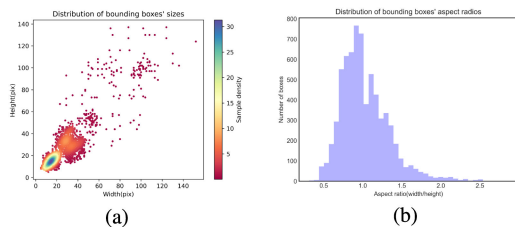
Fig. 2. Distributions of bounding boxes' sizes and aspect ratios of the SADD.

model transfer, the test performance of the target field has been greatly improved. Just as Shengna Wu *et al.* [37] did, in the field of SAR image transfer learning, researchers usually use SAR image datasets as the source domain data, which are homologous to the target domain and of the same type. However, in some specific SAR target detection and recognition tasks, it is difficult to find enough homology and the same category data as the source domain. Pan and Yang [19] showed that even if the training and future data are not in the same feature space or even the distribution is different, knowledge transfer can also improve the performance of machine learning. Inspired by [19], Yong Li *et al.* [38] adopted the optical dataset Pascal VOC 2007 as source domain data for model transfer and achieved a certain performance improvement in the SAR ship detection task.

In the field of SAR image detection and recognition, previous scholars have proved that using homologous or heterogeneous data as the source domain can improve the performance of the model. However, there is almost no research to guide us on how to choose source domain data. We need to know whether to use homologous data or heterologous data as the source domain to improve model performance, which will be of great inspiration for future research. This is one of the motivations of this article.

III. DATASET AND PRODUCTION DETAILS

Since there is no publicly available aircraft detection dataset in SAR images, we collect and construct an aircraft slice dataset to investigate the detection performance of our method. We name the SAR aircraft detection dataset SADD and make it public to facilitate the research of other scholars. The SADD is available at[1]

The SADD is collected from the German TerraSAR-X satellite, which works in x-band and HH polarization mode with image resolutions ranging from 0.5 to 3 m. The ground truth of aircraft are manually annotated by SAR ATR experts according to the prior knowledge and corresponding optical images. After cropping the large images, 2966 nonoverlapped 224×224 slices are collected with 7835 aircraft targets, of which structures, outlines, and main components are clear. The distributions of bounding boxes' sizes in pixels and aspect ratios are shown in Fig. 2(a) and (b), respectively. Aircraft targets in SADD have various sizes, and there are a large number of small-size targets.

The target background of SADD is relatively complex, including various scenes, such as airport runway, airport apron, and civil aviation airport. The negative samples are mainly around the airport, including open space and forest. Fig. 3 shows the sample images in SADD.
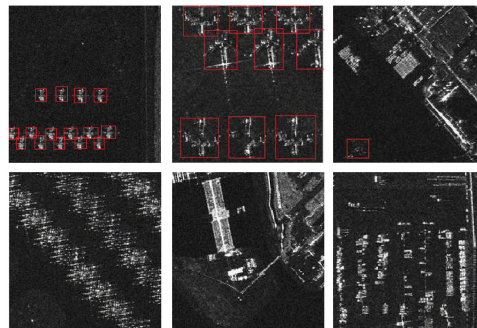
Fig. 3. Examples in SADD. Above are real SAR aircrafts, below are jamming targets.

TABLE I
DIVISION OF THE DATASET

Dataset	Positive samples	Negative samples	Ground truth
Train	796	1666	6948
Test	88	416	887

TABLE II
STATISTICS OF AIRCRAFT TARGET SIZE IN TRAINING SET AND TEST SET

number ⟍ Size (pixel) Dataset	< 20	$20 - 40$	> 40
Train	3425	2319	1204
Test	495	251	141

In this article, to verify our method, we randomly divide the images in SADD into the training set and test set according to the ratio of 5:1. The training set includes 796 positive samples and 1666 negative samples, and the number of aircraft boxes is 6948. The test set includes 88 positive samples and 416 negative samples, and the number of aircraft boxes is 887, as shown in Table I. The aircraft target size distribution in the training set and test set is shown in Table II, where < 20 means that both the width and height of the bounding box are less than 20.

IV. METHODS

Our SAR aircraft detection algorithm is mainly composed of a transfer learning module and detection module. First, the detection network is used to train the source domain data (such as SAR ships), and the feature transfer learning method is used to align the source domain data with the target domain data (SAR aircraft) in the shallow position of the network. Then a small sample of the SAR aircraft dataset (SADD) is used to fine-tune the detection network. The idea diagram of our algorithm is shown in Fig 1. The following is a detailed introduction of the detection network and the feature transfer learning method.

A. Detection Network

Our SEFEPNet consists of feature extraction, feature enhancement pyramid, feature fusion, and feature decoding modules. Inspired by YOLOV3 [17], our feature extraction module uses Darknet-53 as the backbone, which can achieve better results in terms of speed and accuracy. An overview of our network is shown in Fig. 4.
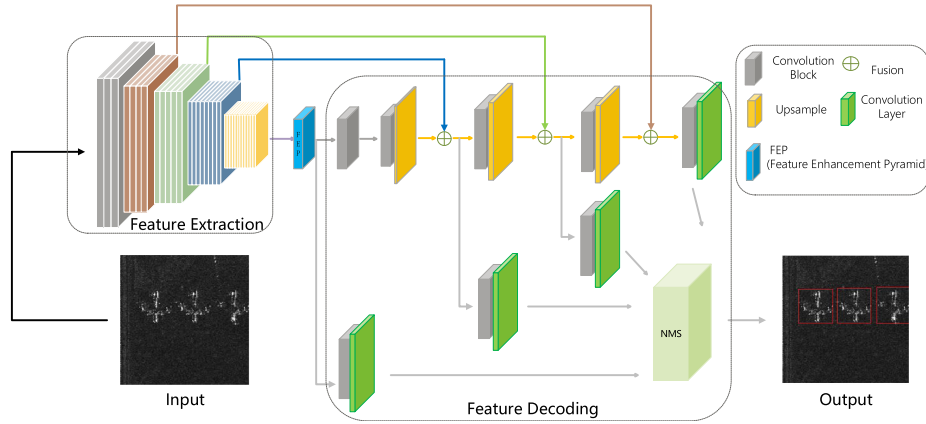
Fig. 4. Overview of the SEFEPNet.

Algorithm 1: Improved K Clustering Algorithm for Generating Anchor Boxes.

Input: X: The set of the length and width of ground truth in the dataset;

Output: M: Anchor box set obtained after clustering of sample set X;

1: Initial $t = 0$, and randomly select k boxes from X as the initial clustering centers;

2: **repeat:**

3: Clustering the sample X. For a fixed class center $m^t_{(w,h)} = (m^t_1, \ldots, m^t_l, \ldots, m^t_k)$, calculate the distance from each sample in X to the class center, and assign each sample to the class closest to it to form the clustering result $C^t_{(w,h)}$;

4: Calculate new class centers. For the clustering result $C^t_{(w,h)}$, calculate the median value of the length and width of the current samples in each class as the new class center $m^{t+1}(w,h) = (m^{t+1}_1, \ldots, m^{t+1}_l, \ldots, m^{t+1}_k)$;

5: Set $t = t + 1$, M $= m^t_{(w,h)}$;

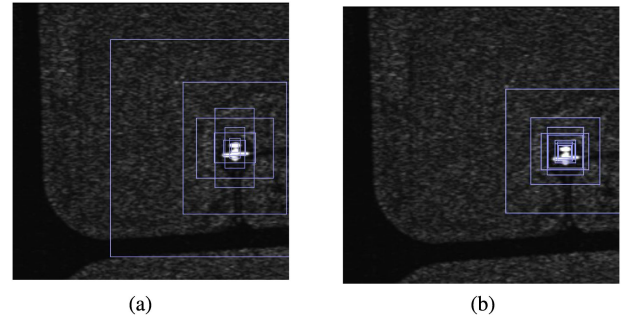6: **until:** The class center does not change and the iteration converges.

Fig. 5. Visualization of the initial anchor boxes obtained by clustering on the SADD. (a) Hand-designed anchors. (b) Improved anchors.
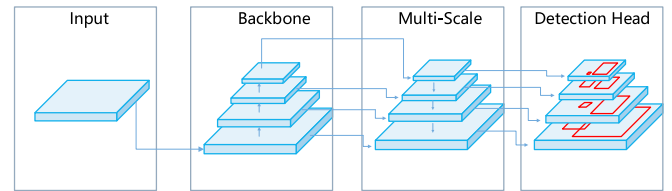

Fig. 6. Structure diagram of scale expansion.

1) Anchor Box: For anchor-based algorithms, the setting of anchors will affect the final detection performance. This article uses an improved k-clustering algorithm to allocate reasonably sized anchors for different scale detection feature maps. Our SEFEPNet adopts detection feature maps at four scales ($4\downarrow$, $8\downarrow$, $16\downarrow$, $32\downarrow$, where $\downarrow$ means downsampling), each containing three anchor boxes with different aspect ratios.

The distance measurement formula of the improved K-cluster is as follows:

$$D(b, c) = 1 - \text{IOU}(b, c) \quad (1)$$

$$\text{IOU}(b_p, b_g) = \frac{b_p \cap b_g}{b_p \cup b_g} \quad (2)$$

where $D(b, c)$ is the distance between predict box b and cluster center c, b_p is the predict box, and b_g is the ground truth. The steps of the improved k-means algorithm are as Algorithm 1.

According to the shape characteristics of aircraft targets in SAR images, the clustering center is set as 12. After 10 times of clustering analysis, the average clustering results are as follows: $(14, 14)$, $(14, 21)$, $(20, 18)$, $(21, 25)$, $(26, 21)$, $(29, 31)$, $(49, 36)$, $(39, 47)$, $(59, 45)$, $(45, 60)$, $(80, 78)$, $(137, 143)$. Compared with the hand-designed anchor boxes, such as faster R-CNN [11], anchor boxes obtained by this advanced clustering algorithm can better adapt to the size of the aircraft target. Fig. 5 is a visualization of the anchor boxes.

2) Scale Expansion: From Fig. 2, we can find that the scales of aircraft targets in SAR images are very different, and there are weak targets less than 10×10 pixels. Our SEFEPNet adopts a four-scale feature fusion detection method, adapting to multiscale target detection. By fusing four times down-sampled shallow position information with deep semantic features, the detection performance of weak targets can be effectively improved. The implementation of scale expansion is shown in Fig. 6.
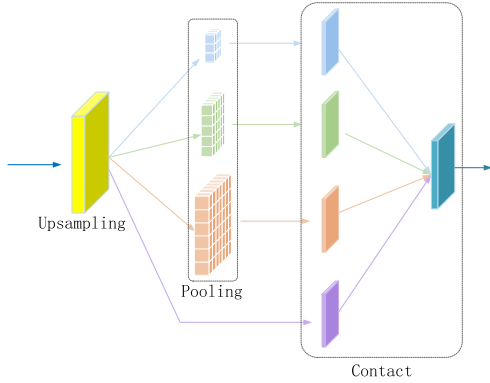
Fig. 7. Structure diagram of FEP.

3) Enhancement Pyramid (FEP): Since it is challenging to extract target features in a complex background, we propose the feature-enhanced pyramid structure (FEP), which can effectively improve the detection performance, as shown in Fig. 7. First, the FEP module up-samples the feature map output by the feature extraction backbone, strengthening the target feature and preventing the loss of the feature information of the weak target due to the subsequent spatial pyramid pool structure. Second, it uses three maximum pooling layers of different sizes ($3 \times 3, 5 \times 5, 7 \times 7$) for subsampling, which can weaken the background texture information and fully extract the local features of the target. Third, it joins the pooled feature maps and the original feature map along the channel direction to obtain an enhanced feature map. FEP realizes the fusion of local features and global features, thus enriching the expressive ability of feature maps and improving the extraction ability of target features in complex backgrounds. In addition, the pooling pyramid improves the receptive field of the feature map and solves the problem of excessive target scale differences.

4) Loss Function: The loss function in one-stage target detection comprises three parts: 1) box position loss, 2) target loss, and 3) classification loss. Since the single-class aircraft target detection task of SAR image does not need to be classified, the loss function should be composed of the first two. The final detection loss function can be expressed as

$$L = w \times (L_{xy} + L_{wh}) + L_{obj} \tag{3}$$

where L_{xy} represents the loss of the coordinate position of the prediction box, L_{wh} represents the regression loss of the width and height of the prediction box, and L_{obj} represents the loss caused by the target confidence. w is the weighting coefficient. The specific implementation details of the loss function are shown in [17].

B. Method of Transfer Learning

With the popularity of deep learning methods, more and more researchers use deep neural networks for transfer learning. Compared with traditional nondeep transfer learning methods, deep transfer learning directly improves the learning effect on different tasks. In optical image target detection, Fine-tune is a simple and popular deep network transfer method. Finetune uses source domain data to train the model and adjusts the model according to the specific task of the target domain, which can achieve a good detection effect. Yosinski *et al.* [39] proved that

Finetune could better overcome the differences between domain data, and the effect would be significantly improved if fine-tune was added to the deep transfer network.

Therefore, for the SAR aircraft target detection task, we can consider using the Finetune transfer method. Due to the shortage of SAR aircraft datasets, we can use optical aircraft images or SAR ship images as the source domain. However, the imaging mechanism between optical aircraft images and SAR aircraft images is very different, and SAR ships and SAR aircraft belong to different types, so no matter which type of dataset we choose as the source domain, there will be a large data distribution difference and feature difference with SAR aircraft targets, so only use finetune transfer method cannot have a satisfying effect. Inspired by Tzeng *et al.* [40], we designed an adaptive transfer network, as shown in Fig. 8.

We add the MMD adaptation layer behind the third residual group of the Darknet-53 feature extraction network to calculate the distance between the features of the source domain and the target domain and added it to the loss of SEFEPNet for training, which can effectively solve the problem of the large difference between the data of the source domain and the target domain. Krizhevsky *et al.* [41] revealed that the shallow layer of the neural network learns the general features, while the deep layer learns the specific features that are more focused on the learning task. For the transfer learning of SAR aircraft images, the common features of the source domain data and the target domain data are beneficial. Finally, through experiments, we chose the 8x downsampling output of the Darknet-53 network as the position of the adaptation layer. The loss function of the adaptive transfer network is expressed as

$$\ell = \ell_d(D_s, y_s) + \lambda \ell_A(D_s, D_t) \tag{4}$$

where D_S represents the source domain, D_t represents the target domain, and y_s represents the detection result of the network. ℓ represents the final loss of the network, $\ell_d(D_s, y_s)$ represents the regular detection loss of the network on labeled source domain data (which is the same as the ordinary deep network), $\ell_A(D_S, D_t)$ represents the adaptive loss of the network and λ is the weight parameter weighing the two parts of the loss function. In this article, we adopt the MMD measurement criterion as adaptive loss, which is expressed as

$$\ell_A(D_s, D_t) = \mathrm{MMD}^2(D_s, D_t) \tag{5}$$

$$\mathrm{MMD}^2(D_s, D_t) = \left\| \sum_{i=1}^{n1} \phi(x_i) - \sum_{j=1}^{n2} \phi(y_j) \right\|_{\mathcal{H}}^2 \tag{6}$$

where x_i represents the *i*th sample in the source domain, y_j represents the *j*th sample in the target domain, and $\phi(\cdot)$ is the mapping that maps the original variable to the reproducing kernel Hilbert space (RKHS) [42].

The specific steps of the transfer learning method in this article are as follows.

1) Feature alignment: Use the adaptive transfer network to train the source domain data and make the trained model more in line with the target feature distribution of the SAR aircraft through the adaptive layer.
2) Model fine-tuning: Freeze the network layer before the adaptive layer and fine-tune the trained network using a small sample SAR aircraft image dataset (target domain).
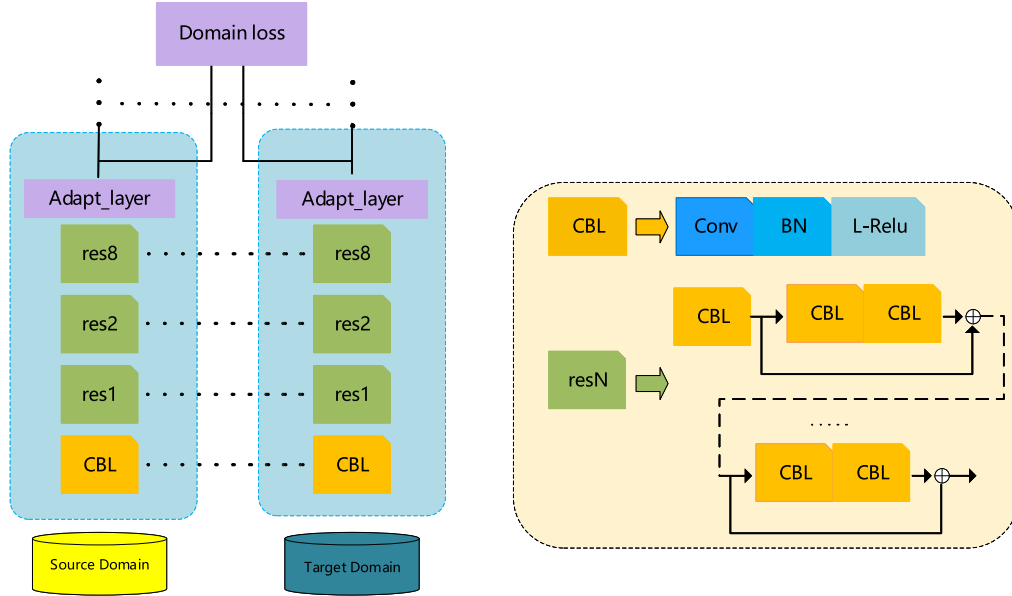
Fig. 8. Domain adaptive transfer learning method in SEFEPNet.

TABLE III
HYPERPARAMETERS SETTINGS

Hyperparameters	Value
Optimizer	SGD
Learning rate	0.01
Momentum	0.9
Decay	0.0005
Batch size	16
Epochs	100

V. EXPERIMENTS AND RESULTS

In this section, we first introduce the experimental parameter settings and evaluation metrics. Then, we conduct ablation studies to investigate our network. Finally, we compare SEFEPNet with several state-of-the-art object detection methods.

A. Experimental Setup

For our SEFEPNet, we use SAR images with a resolution of 224×224 as input and resize the images to a resolution of 416×416. Models are trained on 2080Ti with a batch size of 16 scenes per each. We use stochastic gradient descent with an initial 0.01 and drop the learning rate at the 80% and 90% of the total number of iterations by the factors of [0.1,0.01]. We set the weight decay to 5×10^{-4} and stop training when the loss plateaus. Hyperparameter settings are shown in Table III.

B. Evaluation Metrics

Precision and recall are often used as evaluation criteria. However, the precision rate and recall rate are usually contradictory, that is, when the precision rate is high, the recall rate is usually low, and when the recall rate is high, the precision rate is usually low. Therefore, we added F1-score, which is a comprehensive indicator between precision and recall. The calculation method is shown as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

$$F1 - \text{Score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{9}$$

where true positive (TP) means that the aircraft is correctly predicted, false positive (FP) means that the actual category is a false alarm but the predicted class is the aircraft, and false negative (FN) means that the actual category is the aircraft, but the predicted category is a false alarm. Since there are many methods to eliminate false alarm targets, rough detection of aircraft targets is the most important part that affects the performance of SAR aircraft target detection. Therefore, the recall rate is more important than the precision rate in the evaluation criteria of this article.

C. Ablation Study

In this subsection, we compare the performance of our SEFEPNet with different architectures and types of transfer learning to investigate the potential benefits of different design choices.

1) Multiscale Information: We investigate the benefit of the number of layers to be fused by removing the FEP module in our SEFEPNet. That is, we only use the method of scale expansion and explore the impact of different scale feature fusion layers on the model performance. As shown in Table IV, increasing the number of fusion layers can effectively improve detection performance of small-size targets. NoFEP-3L means that SEFEPNet removes the FEP module and uses three scales for feature fusion. Fig 9 is a visualization of the detection results using different scale features for fusion.

TABLE IV
EFFECT OF DIFFERENT MODULES ON SEFEPNet

Method	Size	TP	FP	FN	P	R	F1
NoFEP-3L	<20	434	83	61	0.839	0.877	0.858
	20-40	233	47	18	0.832	0.928	0.877
	>40	128	29	13	0.815	0.908	0.859
NoFEP-4L	<20	445	62	50	0.877	0.899	0.888
	20-40	236	43	15	0.846	0.940	0.891
	>40	129	27	12	0.827	0.915	0.869
FEP-4L	<20	454	61	41	0.881	**0.917**	**0.899**
	20-40	243	42	8	0.853	**0.968**	**0.907**
	>40	133	27	8	0.831	**0.943**	**0.884**

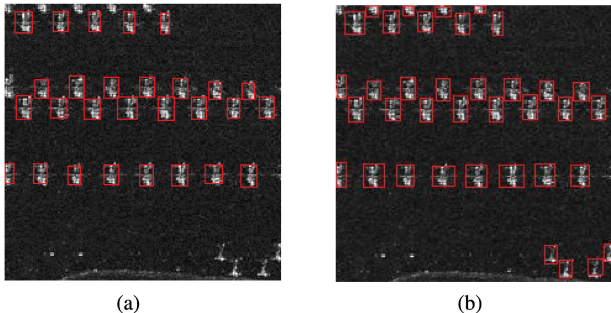The significance of bold entities indicate best values.

Fig. 9. Comparison of multiscale fusion detection results of SEFEPNet. (a) SEFEPNet-NoFEP-3L. (b) SEFEPNet-NoFEP-4L.

TABLE V
COMPARISONS OF DIFFERENT TYPES OF TRANSFER LEARNING

Method	TP	FP	FN	P	R	F1
Transfer-Free	830	130	57	0.864	0.936	0.899
Optical-Aircraft	841	133	46	0.863	0.955	0.907
SAR-Ship	871	106	16	0.891	**0.982**	**0.934**

The significance of bold entities indicate best values.

2) Feature Enhancement Pyramid: We investigate the benefits introduced by our FEP module. As shown in Table IV, it demonstrates the effectiveness of our feature-enhanced pyramid structure. FEP-4L means that SEFEPNet adopts the FEP module and uses four scales for feature fusion.

3) Types of Transfer Learning: We investigate the benefit of transfer learning on aircraft target detection in SAR images. In addition, we explore the impact of different types of transfer learning on our SEFEPNet. That is, we selected optical aircraft images from [26] and found the publicly available SAR ship images dataset [27]. Table V shows the results of transfer learning with different source domain data. We can find that when the SAR aircraft dataset is insufficient, using the same source but different object images is better than the different source but same object images for transfer learning.

D. Comparison to the State of the Arts

We compare our method to two two-stage methods, faster R-CNN [11], cascade R-CNN [43] and three one-Stage methods, SSD [12], Yolov3 [17], and YOLOX [44]. Our SEFEPNet uses

TABLE VI
COMPARISONS TO THE STATE-OF-THE-ARTS

Model	TP	FP	FN	P	R	F1
SSD	802	166	85	0.828	0.904	0.865
YOLOV3	848	148	39	0.851	0.956	0.901
Faster R-CNN	826	121	61	0.872	0.931	0.901
Cascade R-CNN	831	110	56	0.883	0.937	0.909
YOLOX	860	82	27	0.913	0.969	**0.940**
SEFEPNet	871	106	16	0.891	**0.982**	0.934

The significance of bold entities indicate best values.

the domain adaptive transfer learning method. To ensure the fairness of the comparison experiment, we use the weights obtained from SAR ship data training to initialize other comparison algorithms to improve the detection performance of SAR aircraft targets. Table VI and Fig. 10 show the comparison results of the methods. It clearly shows that our method is better than mainstream algorithms for the large difference in the target scale of SADD.

E. Experimental Analysis

Table VI shows that for the SADD dataset, our SEFEPNet has a higher recall rate and F1 value than mainstream detection algorithms. This is because there are a large number of small targets in the SADD dataset, and the target sizes vary widely, as shown in Table II. Our SEFEPNet adopts a four-scale feature fusion method to effectively alleviate the problematic detection of small targets and adapt to multiscale target detection, as shown in Table IV. In addition, our SEFEPNet uses the FEP module, which can strengthen the feature and weaken the texture, helping to extract target features in complex backgrounds. The spatial pyramid pooling structure in the FEP module can fully extract the local features of targets. It can significantly enrich the receptive field information after fusion with the global features, conducive to the integrity detection of SAR aircraft targets.

Table V shows that for the SAR aircraft detection task with a small sample dataset, when the same domain adaptive transfer learning method is adopted, the detection effect using SAR ship images as the source domain data is better than that using optical aircraft images as the source domain data. Since the imaging mechanism of SAR images and optical images are different, using optical aircraft images as source data cannot extract features similar to SAR aircraft well. Although SAR ships and SAR aircraft belong to different types of targets, their imaging mechanisms are similar, leading to more similar common features of the two targets.

Fig. 11 shows the final detection results of our algorithm. It can be seen that our algorithm can effectively adapt to multiscale targets, and it can also locate the target in the case of complex background interference.
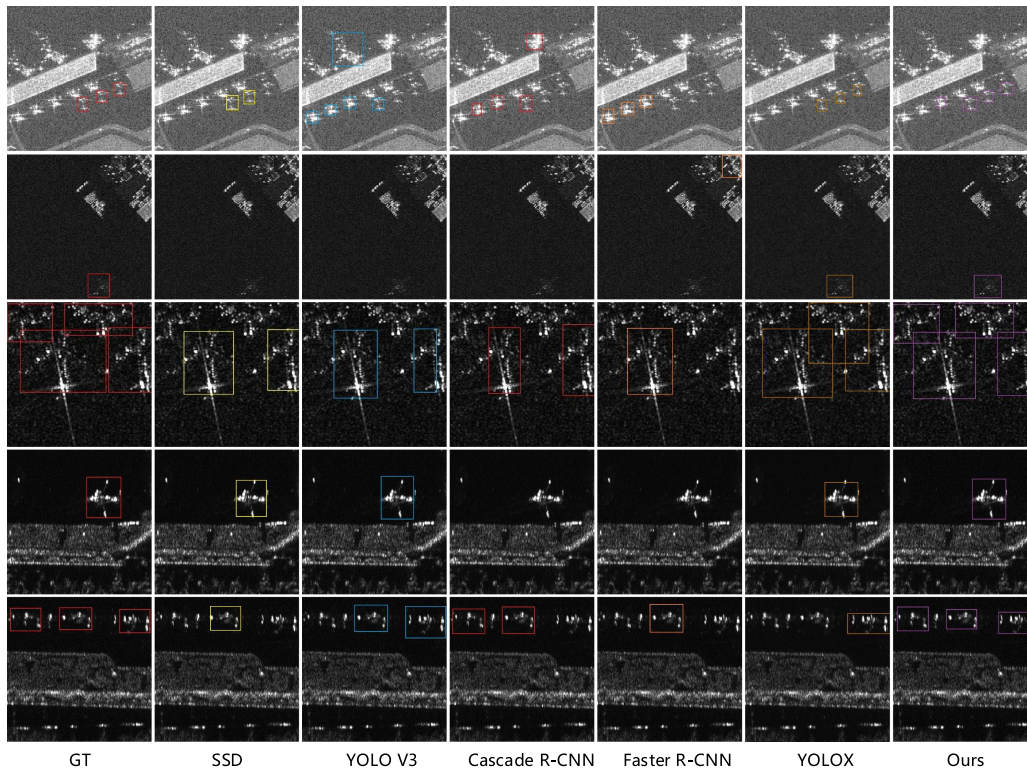
Fig. 10. Visualization of SAR aircraft detection results of different algorithms.
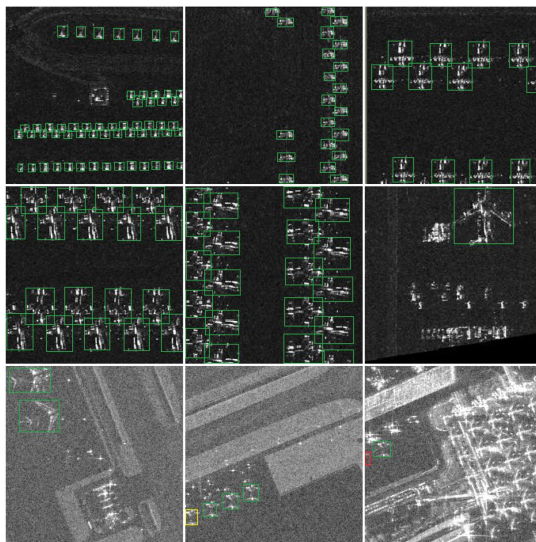

Fig. 11. SEFEP-Net detection results in several test samples of SADD. The yellow boxes, red boxes, and green boxes represent false alarms, missing targets, and correct detection, repectively.
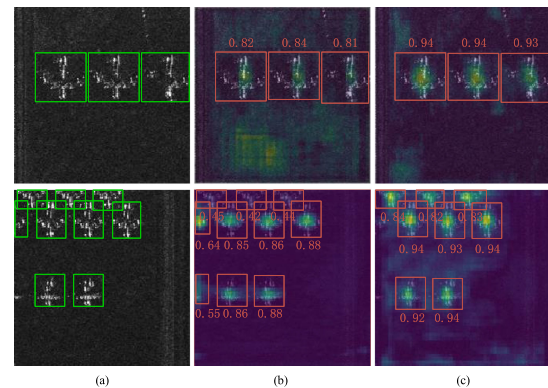

Fig. 12. Comparison visualization of the detection confidence maps. (a) Ground truth in SAR images. (b) Using optical aircraft as the source domain for transfer learning. (c) Using SAR ships as the source domain for transfer learning.

VI. DISCUSSION

This section will further discuss and analyze the work of transfer learning.

Source domain data should have similar characteristics to SAR aircraft data. Therefore, in our work, we divide source domain data into two categories: 1) heterogeneous but same type data (such as optical aircraft) and 2) homologous but different type data (such as SAR ships). We designed a domain adaptive transfer learning method for verification to compare which type of source domain data are more effective for SAR aircraft detection.

From Table V we can see that using domain adaptive transfer learning method can effectively improve the detection performance of the model, and the effect of using homologous but different type data (SAR ships) as the source domain is better than that of heterogeneous but same type data (optical aircraft). The confidence maps of the detection results of transfer learning are shown in Fig. 12. Compared with optical aircraft data, using SAR ship image as source domain data has more accurate positioning accuracy and higher confidence for SAR aircraft targets. Due to the limitation of public dataset, except for SAR
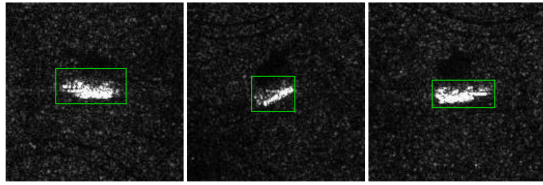
Fig. 13. Samples in MSTAR.

TABLE VII
COMPARISONS OF DIFFERENT TYPES OF TRANSFER LEARNING

Method	TP	FP	FN	P	R	F1
Optical-Aircraft	841	133	46	0.863	0.955	0.907
SAR-Vehicle	858	121	29	0.876	0.967	0.919
SAR-Ship	871	106	16	0.891	**0.982**	**0.934**

The significance of bold entities indicate best values.

ship data, we can only collect SAR vehicle data called MSTAR. Since MSTAR is a classified dataset, we manually labeled the 1000 ground truth, as shown in Fig. 13. The effect of using SAR vehicle data as the source domain is shown in Table VII. Ten thousand publicly available optical aircraft images and ten thousand SAR ship images were used in the experiment. We can see that in the field of SAR aircraft detection, the effect of using homologous but different type data (SAR vehicle) as the source domain is still better than that of heterogeneous but same type data (optical aircraft).

VII. CONCLUSION

This article constructs a public SADD with complex background and interference objects, which can provide a benchmark for the evaluation of different algorithms. Then we provide a baseline SEFEPNet for SADD, which can effectively solve the problem of detection difficulties caused by complex background interference and target size diversity in SAR images, and is superior to the mainstream object detection algorithms. To further improve the detection performance of the small-scale dataset, we design a domain adaptive transfer learning method. In the field of SAR aircraft detection, we find that the transfer effect of using homologous but different types of targets is better than that of heterologous but same types of targets. The conclusion has enlightening significance for future research and has practical application value in real scenes.

In the future, we will enrich our SADD and implement fine-grained aircraft type labeling. In addition, we will improve the model detection performance based on the idea of associative target recognition and improve our transfer learning method.

REFERENCES

[1] B. van den Broek, E. den Breejen, R. Dekker, and A. Smith, "Change detection and maritime situation awareness in the channel area—Feasibility of space borne SAR for maritime situation awareness," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 7436–7439.

[2] Y. Guo, L. Du, D. Wei, and C. Li, "Robust SAR automatic target recognition via adversarial learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 716–729, 2021.

[3] H. Qu, L. Shen, W. Guo, and J. Wang, "Ships detection in SAR images based on anchor-free model with mask guidance features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 666–675, 2022.

[4] W. Li, B. Zou, Y. Xin, L. Zhang, and Z. Wu, "An improved CFAR scheme for man-made target detection in high resolution SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 2829–2832.

[5] Y. Cui, G. Zhou, J. Yang, and Y. Yamaguchi, "On the iterative censoring for target detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 641–645, Jul. 2011.

[6] S. Brusch, S. Lehner, T. Fritz, M. Soccorsi, A. Soloviev, and B. van Schie, "Ship surveillance with TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1092–1103, Mar. 2011.

[7] J. Chen, B. Zhang, and C. Wang, "Backscattering feature analysis and recognition of civilian aircraft in TerraSAR-X images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 796–800, Apr. 2015.

[8] C. He, S. Li, Z. Liao, and M. Liao, "Texture classification of Pol-SAR data based on sparse coding of wavelet polarization textons," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 8, pp. 4576–4590, Aug. 2013.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 580–587.

[10] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 91–99, 2015.

[12] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

[13] Z. Li and F. Zhou, "FSSD: Feature fusion single shot multibox detector," 2017, *arXiv:1712.00960*.

[14] C. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: Deconvolutional single shot detector," 2017, *arXiv:1701.06659*.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.

[16] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7263–7271.

[17] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018, *arXiv:1804.02767*.

[18] C. He, M. Tu, D. Xiong, F. Tu, and M. Liao, "A component-based multilayer parallel network for airplane detection in SAR imagery," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1016.

[19] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.

[21] Z. Zhong *et al.*, "Squeeze and attention networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 13065–13074.

[22] M. Jaritz, T.-H. Vu, R. d. Charette, E. Wirbel, and P. Perez, "xMUDA: Cross-modal unsupervised domain adaptation for 3D semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 12605–12614.

[23] A. Nagrani, C. Sun, D. Ross, R. Sukthankar, C. Schmid, and A. Zisserman, "Speech2Action: Cross-modal supervision for action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10317–10326.

[24] Y. Li, B. Ji, X. Shi, J. Zhang, B. Kang, and L. Wang, "TEA: Temporal excitation and aggregation for action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 909–918.

[25] G. Li, J. Su, Y. Li, and X. Li, "SAR aircraft detection based on convolutional neural network and attention mechanism," *J. Syst. Eng. Electron.*, vol. 43, pp. 3202–3210, 2021.

[26] J. Shermeyer, T. Hossler, A. Van Etten, D. Hogan, R. Lewis, and D. Kim, "Rareplanes: Synthetic data takes flight," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2021, pp. 207–217.

[27] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7., 2019, Art. no. 765.

[28] Y. Tan, Q. Li, Y. Li, and J. Tian, "Aircraft detection in high-resolution SAR images based on a gradient textural saliency map," *Sensors*, vol. 15, no. 9, pp. 23071–23094, 2015.

[29] M. Dostovalov, R. Ermakov, and T. Moussiniants, "Detection of aircraft using Sentinel-1 SAR image series," in *Proc. Int. Radar Symp.*, 2018, pp. 1–7.

[30] W. Diao, F. Dou, K. Fu, and X. Sun, "Aircraft detection in SAR images using saliency based location regression network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 2334–2337.

[31] W. Siyu, G. Xin, S. Hao, Z. Xinwei, and S. Xian, "An aircraft detection method based on convolutional neural networks in high-resolution SAR images," *J. Radars*, vol. 6, no. 2, pp. 195–203, 2017.

[32] D. Zhao and S. Sun, "A new method for target detection of remote sensing image based on RESNET," *Electron. Des. Eng.*, vol. 26, no. 22, pp. 164–168, 2018.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[34] Q. Guo, H. Wang, and F. Xu, "Aircraft detection in high-resolution SAR images using scattering feature information," in *Proc. 6th Asia-Pacific Conf. Synthetic Aperture Radar (APSAR)*, 2019, pp. 1–5.

[35] Q. An, Z. Pan, L. Liu, and H. You, "DRBox-v2: An improved detector with rotatable boxes for target detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8333–8349, Nov. 2019.

[36] T. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1449–1457.

[37] S. Wu, K. Wang, and Y. Ouyang, "Study on small samples SAR image recognition detection method based on transfer CNN," in *Proc. 3rd Int. Conf. Electron. Inf. Technol. Comput. Eng.*, 2019, pp. 718–722.

[38] Y. Li, Z. Ding, C. Zhang, Y. Wang, and J. Chen, "SAR ship detection based on RESNET and transfer learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1188–1191.

[39] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?", *Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.

[40] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," 2014, *arXiv:1412.3474*.

[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[42] K. Borgwardt, A. Gretton, M. Rasch, H.-P. Kriegel, B. Schoelkopf, and A. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. e 49-e57, 2006.

[43] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6154–6162.

[44] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

Peng Zhang received the B.S. degree in artificial intelligence and automation from the Huazhong University of Science and Technology, Wuhan, China, in 2020. He is currently working toward the M.S. degree in control science and engineering from the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China.

His research interests include computer vision, machine learning, and remote sensing object detection.

Hao Xu received the B.S. degree in automation from the Northeastern University, Shenyang, China, in 2020. He is currently working toward the M.S. degree in control science and engineering from the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China.
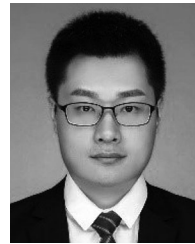
His research interests include computer vision, machine learning, and remote sensing object detection.

Tian Tian received the B.S. degree in electronic information engineering and Ph.D. degrees in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2015, respectively.

From 2012 to 2014, she was visiting Oakland University, MI, USA, as a Ph.D. candidate sponsored by China Scholarship Council. She joined the School of Computer Sciences, China University of Geosciences, Wuhan, China, in 2015, as a Postdoctoral Lecturer. She is currently an Associate Professor with School of Artificial Intelligence and Automation, Huazhong University of Science and Technology. Her research interests include remote sensing image processing, computer vision, and applications.
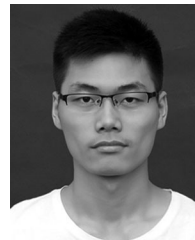
Peng Gao received the B.S. degree in optical information science and technology from the Huazhong University of Science and Technology, Wuhan, China, in 2013, and the M.S. degree in optical engineering from the Huazhong Institute of Optoelectronic Technology, Wuhan, China, in 2016. He is currently working toward the Doctor's degree in control science and engineering from the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China.

His research interests include computer vision, machine learning, and remote sensing object detection.

Linfeng Li received the B.S. degree in automation from the School of Automation, Chongqing University, Chongqing, China, in 2017. He is currently working toward the Ph.D. degree in control science and engineering from the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China.

His research interests include image processing, pattern recognition, and remote sensing.

Tianming Zhao received the B.S. degree in electronic information engineering from the School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan, China, in 2018. He is currently working toward the Ph.D. degree in control science and engineering from the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China.

His research interests include 3-D computer vision, image processing, and remote sensing object detection.

Nan Zhang received the B.S. degree in automation from Hainan University, Hainan, China, in 2019. He is currently working toward the Ph.D. degree in artificial intelligence from the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China.

His research interests include object detection, pattern recognition, and remote sensing.

Jinwen Tian received the Ph.D. degree in pattern recognition and intelligent systems from Huazhong University of Science and Technology (HUST), Wuhan, China, in 1998.

He is currently a Professor with HUST. His research interests include remote sensing image analysis, image compression, computer vision, and fractal geometry.