



Article

Integrating Weighted Feature Fusion and the Spatial Attention Module with Convolutional Neural Networks for Automatic Aircraft Detection from SAR Images

Jielan Wang^{1,2}, Hongguang Xiao^{1,*}, Lifu Chen^{2,3} , Jin Xing⁴ , Zhouhao Pan⁵, Ru Luo^{2,3} and Xingmin Cai^{2,3}

¹ School of Computer and Communication Engineering, Changsha University of Science & Technology, Changsha 410114, China; ningyuan@stu.csust.edu.cn

² Laboratory of Radar Remote Sensing Applications, Changsha University of Science & Technology, Changsha 410014, China; lifu_chen@csust.edu.cn (L.C.); Luoru@stu.csust.edu.cn (R.L.); cc182022@stu.csust.edu.cn (X.C.)

³ School of Electrical and Information Engineering, Changsha University of Science & Technology, Changsha 410114, China

⁴ School of Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, UK; Jin.Xing@newcastle.ac.uk

⁵ China Academy of Electronics and Information Technology, Beijing 100041, China; panzhouhao09@mailsucas.ac.cn

* Correspondence: lgx2020@csust.edu.cn; Tel.: +86-135-7416-8208



Citation: 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. <https://doi.org/10.3390/rs13050910>

Academic Editor: Mercedes E. Paoletti

Received: 14 January 2021

Accepted: 23 February 2021

Published: 28 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The automatic detection of aircrafts from SAR images is widely applied in both military and civil fields, but there are still considerable challenges. To address the high variety of aircraft sizes and complex background information in SAR images, a new fast detection framework based on convolution neural networks is proposed, which achieves automatic and rapid detection of aircraft with high accuracy. First, the airport runway areas are detected to generate the airport runway mask and rectangular contour of the whole airport are generated. Then, a new deep neural network proposed in this paper, named Efficient Weighted Feature Fusion and Attention Network (EWFAN), is used to detect aircrafts. EWFAN integrates the weighted feature fusion module, the spatial attention mechanism, and the CIF loss function. EWFAN can effectively reduce the interference of negative samples and enhance feature extraction, thereby significantly improving the detection accuracy. Finally, the airport runway mask is applied to the detected results to reduce false alarms and produce the final aircraft detection results. To evaluate the performance of the proposed framework, large-scale Gaofen-3 SAR images with 1 m resolution are utilized in the experiment. The detection rate and false alarm rate of our EWFAN algorithm are 95.4% and 3.3%, respectively, which outperforms Efficientdet and YOLOv4. In addition, the average test time with the proposed framework is only 15.40 s, indicating satisfying efficiency of automatic aircraft detection.

Keywords: SAR image; aircraft detection; airport detection; deep learning; feature fusion; spatial attention module

1. Introduction

Synthetic Aperture Radar (SAR) is an advanced active microwave earth observation approach, which is insensitive to clouds and fogs, and images the earth surface all day and all night. SAR turns out to be fascinating for reconnaissance missions under various weather conditions. At present, the resolution of SAR platforms can reach the centimeter-level, which offers opportunities to identify detailed targets in various application domains. Since 1978, SAR has attracted considerable attention of the radar scientific community because of its unique imaging mechanism; and it has been widely used in both military and civilian fields.

The detection and identification of aircraft are essential to the effective management of the airport. Acquiring the number, type, location, and status information of aircrafts is

of great military value. Therefore, research on automatic aircraft detection algorithm with SAR imagery is very necessary.

A novel framework is proposed in this paper to perform efficient aircraft detection from SAR images automatically and efficiently. The main contributions of this article are summarized as follows:

- (1) For the efficient management of the airport (in the civil field) and real-time acquisition of battlefield military intelligence and formulation of combat plans (in the military field). An efficient and automatic aircrafts detection framework is proposed. First, the airport runway areas are extracted, then the aircraft is detected using the Efficient Weighted Feature Fusion and Attention Network (EWFAN), and finally the runway area is employed to filter false alarms. This framework can provide a generalizable workflow for aircraft detection and achieve high-precision and rapid detection.
- (2) EWFAN is proposed to perform aircrafts detection by integrating SAR image analytics and deep neural networks. It effectively integrates the weighted feature fusion module and the spatial attention mechanism, with the CIF loss function. This network is a lightweight network with the advantages of high detection accuracy and fast detection speed; it provides an important reference to other scholars and can also be extended for the detection of other dense targets in SAR image analytics, such as vehicles and ships.

For the efficient management of the airport (in the civil field) and real-time acquisition of battlefield military intelligence and formulation of combat plans (in the military field). An Efficient Weighted Feature Fusion and Attention Network (EWFAN) is proposed to perform aircrafts detection. It effectively integrates the weighted feature fusion module and the spatial attention mechanism, with the CIF loss function. This network is a lightweight network with the advantages of high detection accuracy and fast detection speed; it provides an important reference for other scholars and can also be used for the detection of other dense targets in SAR image analytics, such as vehicles and ships.

The rest of this paper is arranged as follows. Section 2 introduces the state-of-the-art of the deep learning on objects detection and the development of the bridge detection from SAR images. In Section 3, the aircraft detection framework of SAR images proposed in this paper is introduced in detail. Section 4 indicates specific experimental results and analysis to verify the efficiency of the algorithm proposed in this paper. Finally, the summary of the paper and the discussion of the future research are given.

2. State-of-the-Art

Traditional target detection methods of SAR image are mainly divided into two categories: (1) single feature-based method and (2) multi-feature-based method. The single feature-based approach usually uses Radar Cross Section (RCS) information to extract brighter areas as candidate targets. The most common method is the Constant False Alarm Rate (CFAR) algorithm, which is based on clutter statistics and threshold extraction [1]. In 1968, on the basis of CFAR, Finn et al. [2] proposed the CA-CFAR algorithm, which used all pixels in the clutter region of the sliding window to estimate parameters of the corresponding clutter statistical model, and it is optimal for the detection in homogenous regions. In 1973, Goldstein [3] further extended the CFAR method to address background noise with lognormal distribution and Weibull distribution, to forge the famous Goldstein detector. However, the CFAR algorithm did not include the structural information of the target, which would lead to inaccurate targets positioning. The target detection of the multi-feature-based method is based on the fusion of multiple image features. In 2015, Tan et al. [4] combined gradient texture saliency map with CFAR algorithm to detect aircrafts. However, the manual feature craft process was complex and time-consuming, and the feature fusion algorithm was scenario-dependent, which usually incurred additional errors. In general, manual feature-based target detection in SAR image analytics is facing various challenges, such as poor robustness and low automation [5].

Recently, many scholars began to investigate machine learning algorithms for SAR image target detection, such as Support Vector Machine (SVM) [6] and Adaptive Boosting (AdaBoost) [7]. Although machine learning algorithms have improved detection accuracy to a certain extent compared with traditional target detection methods, they are only suitable for small samples [8], and the manually crafted features are still poor in generalization.

Nowadays, deep learning algorithms have developed rapidly and achieved good results in many fields. Object detection within deep learning can be grouped into two types: “two-stage detection” and “one-stage detection” [9]. The former can be depicted as a “coarse to-fine” process while the latter is described as “complete detection in one step.” Among two-stage detection algorithms, Girshick et al. [10] proposed Region-Convolution Neural Network (R-CNN), which first extracted candidate boxes from the prior boxes, and then filtered the candidate boxes to obtain the final prediction results. But this algorithm suffered from the slow detection speed. In 2015, Girshick [11] further improved the network on the basis of R-CNN [10] and SPPNet [12], as Fast R-CNN, with enhanced detection accuracy and speed. In 2017, He et al. [13] proposed Mask R-CNN, which not only improved the accuracy of the target detection, but also satisfied the accuracy requirements of semantic segmentation tasks. In 2015, R. Joseph et al. [14] proposed the “You Only Look Once” (YOLO) algorithm, which was the first one-stage deep learning target detection algorithm. Compared with the two-stage detection algorithm, its detection speed was impressive, but the detection accuracy was reduced meanwhile. In 2019, Tan et al. [15] proposed Efficientdet, which introduced the weighted BI-directional Feature Pyramid Network (BIFPN) and the compound zoom method, to improve the detection efficiency of the network. In 2020, Bochkovskiy et al. [16] proposed the YOLOv4 algorithm. Compared with previous versions, YOLOv4 further improved the detection speed and accuracy considerably.

With the rapid development of deep learning and SAR imaging technology, many scholars use deep learning algorithms to perform targets detection in SAR image analytics. Compared with traditional methods for targets detection of SAR images, deep learning can achieve higher detection accuracy and faster detection speed, especially with end-to-end detection.

Due to the discreteness, variability, and interference of aircraft scattering characteristics, aircraft detection in synthetic aperture radar (SAR) images is a challenging task. In 2017, Wang et al. [8] proposed an improved significance pre-detection method to achieve multi-scale fast and coarse location of aircraft candidate in SAR images. Then, Convolutional Neural Network (CNN) was used to implement accurate detection of candidate targets, which could achieve good detection accuracy with long testing time. When data enhancement methods are not used, the detection rate of their algorithm are 86.33%. In 2019, Li [17] and others combined the improved line segment detector LSD and deep learning Faster-CNN [18] to design an aircraft detection method in SAR images. The detection rate and false alarm rate of their algorithm are 95% and 5%. In 2020, Zhao et al. [19] proposed a rapid detection algorithm for aircraft targets in SAR images in complex environments and large scenes. The algorithm optimizes the overall detection process, and designs the refined extraction of airport regions based on grayscale features and the coarse detection of aircraft targets based on CNN. The detection rate and false alarm rate of their algorithm are 74.0% and 6.9%. In 2020, Guo et al. [5] proposed a method to detect aircrafts, which first adopted an adaptive discriminant operator to detect the airport, and then extract aircrafts by integrating the scattering information with deep learning. In their work, the airport detection algorithm used could reduce the computation intensity and improve the detection efficiency compared with traditional algorithms. The detection rate and false alarm rate of their algorithm are 94.5% and 8.8%. However, for high-resolution SAR images, it would take a long time to extract the airport area and be difficult to perform automatic aircraft detection. Chen et al. [20] proposed a multi-level densely connected dual-attention network to automatically detect the airport runway area, which achieved better extraction results. However, due to the use of the dual-attention mechanism, the training speed and

testing speed of this network were relatively slow. In 2019, Zhang et al. [21] proposed a new framework called Multi-Resolution Dense Encoder and Decoder (MRDED) network, which integrates Convolutional Neural Network (CNN), Residual Network (ResNet), Dense Convolutional Network (DenseNet), Global Convolutional Network (GCN), and Convolutional Long Short-Term Memory (ConvLSTM). In 2020, Chen et al. [22] proposed a new end-to-end framework based on deep learning to automatically classify water and shadow areas in SAR images. In 2020, Chen et al. [23] proposed a new deep learning-based network to identify bridges from SAR images, namely, multi-resolution attention and balance network (MABN). Chen et al. [24] propose a new scene classification framework, named Feature Recalibration Network with Multi-scale Spatial Features (FRN-MSF), to achieve high accuracy in SAR-based scene classification. Tan et al. [25] proposed a geospatial context attention mechanism (GCAM) to automatically extract airport areas. First, down-sampling were applied to the high-resolution SAR images. Then the proposed GCAM network was used to perform airport detection. Finally, coordinate mapping was utilized to obtain accurate airport detection results from high-resolution SAR images. This method not only presented high detection accuracy, but also greatly reduced the training and testing time compared to MDDA network.

Aiming at the problems of low automation of existing SAR image aircraft detection algorithms, long airport extraction time, and complex preprocessing procedures, this paper proposes an efficient and automatic aircrafts detection framework. The detection rate and false alarm rate of our aircrafts detection framework are 95.4% and 3.3%.

3. Methodology

3.1. Overall Detection Framework

For the efficient management of the airport (in the civil field) and real-time acquisition of battlefield military intelligence and formulation of combat plans (in the military field), we propose an efficient aircraft detection framework, which can automatically and quickly detect aircrafts from SAR images. The framework mainly includes three parts: airport detection, aircraft detection, and filtering. The architecture of the proposed framework is shown in Figure 1. First, the airport detection algorithm [23] is used to obtain the airport runway areas and rectangular contours of the airport, which can be employed to locate the airport area, to reduce the false alarm rate of aircraft detection. Second, after obtaining the rectangular contour of the airport, we first use EWFAN algorithm (as shown in Figure 2) to detect aircrafts via sliding windows in the rectangular box generated by airport detection. The adjacent windows will overlap by 20% (see Section 4.5), which can ensure that each area can be completely detected. Then we use coordinate mapping to convert the coordinates of the aircraft target in each window into its coordinates in the original SAR image. Since the sliding window has a certain overlap rate, it will cause the appearance of overlapping boxes. Therefore, we use the NMS algorithm to filter the overlapping boxes to get the preliminary detection result. Finally, the runway mask is incurred to remove false alarms (they are not aircrafts, but it has been tested as an aircraft), and the final aircraft detection results are generated.

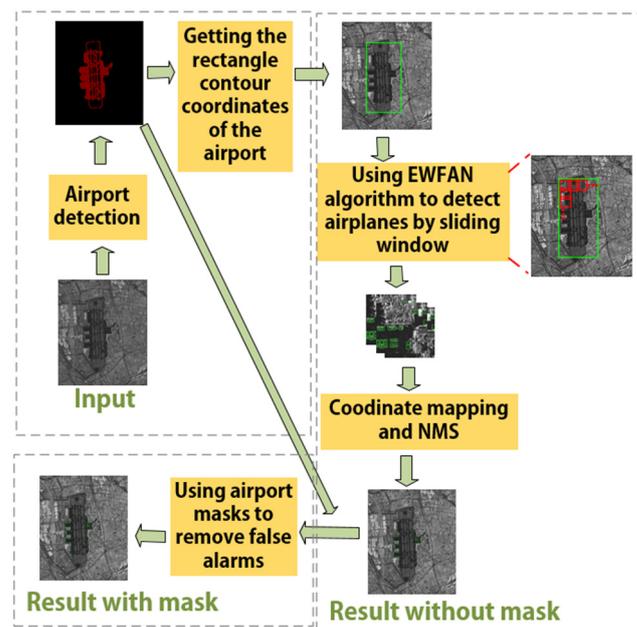


Figure 1. The efficient framework for aircraft detection.

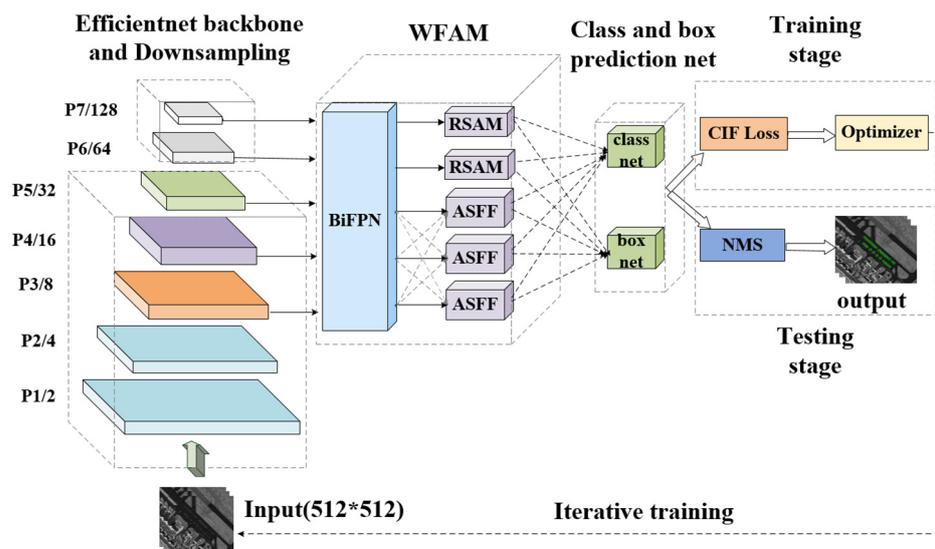


Figure 2. The framework of the Efficient Weighted Feature Fusion and Attention Network (EW-FAN) algorithm.

3.2. EfficientDet

EfficientDet is a powerful target detection network proposed by Tan et al. [15] from the Google Brain team. It combines Efficientnet [26] (also proposed by the same team) and the newly proposed Weighted Bi-directional Feature Pyramid Network (BiFPN). EfficientDet utilizes less parameter calculations to achieve a higher precision compared with other target detection algorithms. It introduces Efficientnet as the backbone network, which transforms the input image into the five-layer feature maps, which are input into BiFPN respectively, and five effective feature layers are obtained after two BiFPN layers. Furthermore, the classification regression network is used to predict the result, and the NMS algorithm is applied to generate the final detection results.

3.3. EWFAN for Aircraft Detection

The architecture of Efficient Weighted Feature Fusion and Attention Network (EWFAN) is shown in Figure 2, which is based on the EfficientDet-D0 framework. In this article, we introduce Adaptively Spatial Feature Fusion (ASFF) [27] and Residual Spatial Attention Module (RSAM) to fuse and extract features and then the Weighted Feature Fusion and Attention Module (WFAM) is proposed. In addition, CIF Loss is also presented in this paper (see Section 3.3.3).

The EWFAN algorithm proposed in this paper still adopts Efficientnet as the backbone network. First, the image is input to the backbone network and down-sampled, then five feature maps of different sizes are obtained. The sizes are 64×64 , 32×32 , 16×16 , 8×8 , and 4×4 respectively. Furthermore, five-layer feature maps are input into WFAM to obtain five effective feature maps. WFAM is composed of BiFPN, ASFF, and RSAM. BiFPN and ASFF perform weighted feature fusion on feature maps. Among them, ASFF pays more attention to spatial information fusion and effectively suppresses the interference of negative samples. RSAM is an attention mechanism proposed in this article. RSAM combines the Spatial Attention Module (SAM) with the residual connection method. Furthermore, the network generates 9 boxes with different sizes and aspect ratios on every grid of each effective feature layer. Then, the prior boxes are classified and regressed in the classification and regression network to predict the results. In the training stage, CIF Loss is proposed in this paper. CIF Loss combines CIoU Loss [28] with Focal Loss [29], which not only improves the stability and accuracy of targets regression, but also speeds up the convergence.

3.3.1. Weighted Feature Fusion and Attention Module (WFAM)

The WFAM module proposed in this paper is composed of BiFPN, ASFF, and RSAM. Among them, BiFPN can perform preliminary weighted fusion of features, and ASFF can enhance the saliency of aircraft targets and suppress the influence of background features on the detection of aircrafts [27]. In addition, this paper combines the SAM with the residual structure, and proposes the RSAM module.

- Weighted Bi-directional Feature Pyramid Network (BiFPN)

BiFPN is an efficient bidirectional feature pyramid network in EfficientDet. In this paper, the BiFPN of EfficientDet-D0 is improved, and only one layer of BiFPN is used in WFAM for preliminary feature fusion. The structure of BiFPN network is shown in Figure 3. BiFPN has made a series of improvements based on FPN.

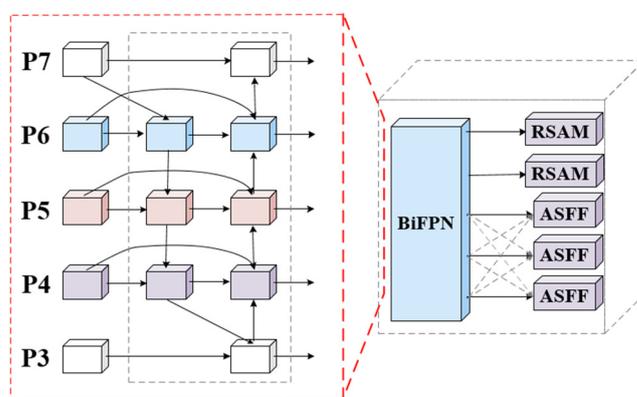


Figure 3. The structure of weighted bi-directional feature pyramid network.

- Residual Spatial Attention Module (RSAM)

We combine SAM [30] with skip connections to obtain the RSAM algorithm. The specific implementation steps of RSAM are shown in Figure 4.

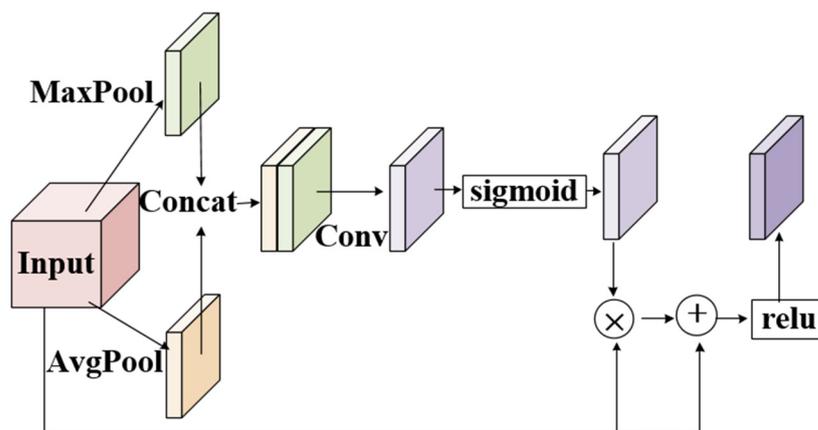


Figure 4. Schematic diagram of Residual Spatial Attention Module.

RSAM mainly focuses on spatial information. In small targets detection, because the target size is small, there is less spatial information on low-resolution feature maps. Therefore, the introduction of RSAM can effectively enhance the spatial information to improve the effect of small target detection. First, the average pooling and maximum pooling operations are performed on the feature maps in the channel dimension, so that the original feature map with a channel number of 64 becomes two intermediate feature maps with a channel number of 1, then concatenation operation is carried out on these two results. Furthermore, the sigmoid function is utilized to normalize the feature map to obtain the spatial attention feature, which is multiplied by the input, and then the skip connection is performed. Finally, the Relu activation function is used to generate the final result.

- Adaptively Spatial Feature Fusion (ASFF)

Although BiFPN performs preliminary feature fusion, it cannot address the shape and size heterogeneity of aircraft detection. There are a few missing alarms but many false alarms in the detection results, which greatly reduce the reliability of the detection results. Therefore, the Adaptive Spatial Feature Fusion (ASFF) algorithm is introduced in this paper. ASFF performs weighted fusion by setting self-learning weights for each fused feature map. This method is better than the direct connection, addition or fast normalization fusion [27]. It can effectively suppress the interference of negative samples and solve the problem of inconsistency between different feature scales in single-stage detection [27]. In addition, the ASFF algorithm has little impact on the amount of network parameters and testing speed [27]. In order to limit the interference of negative samples and enhance the saliency of features, ASFF is introduced in the P3-P5 layer. As shown in Figure 5, the principle and implementation steps of ASFF are listed as:

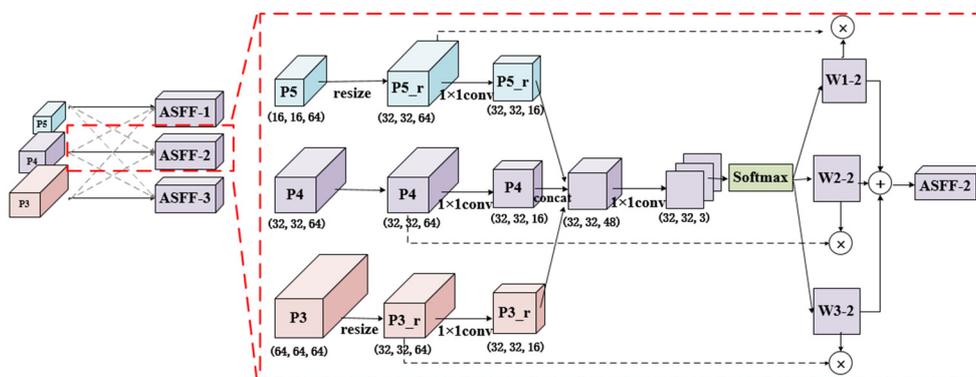


Figure 5. Schematic diagram of Adaptively Spatial Feature Fusion (Taking ASFF-2 as an example).

(1) Feature resizing. ASFF-1, ASFF-2, and ASFF-3 correspond to the P5 layer, the P4 layer, and the P3 layer, respectively. The feature resizing is demonstrated using ASFF-2. In ASFF-2, the size of the P4 layer remains unchanged, and the P3 and P5 layers are adjusted to the same size as the P4 layer. After down-sampling the feature map of the P3 layer, its size is reduced from 64 to 32 to obtain P3_resized. After up-sampling the feature map of the P5 layer, the size is increased from 16 to 32, and P5_resized is obtained. In this way, the size of the three feature maps is converged. Among these operations, up-sampling adopts interpolation method, and down-sampling employs a 3×3 convolution (stride = 2) to reduce the size to half of the original. If you want to reduce the size to 1/4 of the original size, you can first use the largest pooling layer (stride = 2), then use a 3×3 convolution (stride = 2).

(2) Adaptive fusion. We also use ASFF-2 as an example to illustrate this process. First, a 1×1 convolution operation is performed on the three adjusted feature maps to reduce the number of channels from the original 64 to 16. Then, the three feature maps are spliced together to obtain a feature layer with 48 channels. Furthermore, a 1×1 convolution is carried out to reduce the number of channels to 3. Finally, normalization is achieved through softmax, and the final weights $\alpha^3, \beta^3,$ and γ^3 are obtained. The weights $\alpha^3, \beta^3,$ and γ^3 are multiplied by P3_resized, P3_resized, and P4 respectively, and then the three results are added to obtain the new fusion feature ASFF-2. The calculation method of ASFF is shown in Equation (1) [27], and the normalization method of softmax is shown in Equation (2) [27]:

$$W_{ij}^l = \alpha_{ij}^l \cdot A_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot A_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot A_{ij}^{3 \rightarrow l} \tag{1}$$

where $A_{ij}^{n \rightarrow l}$ represents the feature vector at position (i, j) after the feature map A_{ij}^n is adjusted to the same size of A_{ij}^l . $\alpha_{ij}^l, \beta_{ij}^l,$ and γ_{ij}^l are the spatial importance weights of l at three different levels at position (i, j) respectively, and $\alpha_{ij}^l, \beta_{ij}^l, \gamma_{ij}^l \in [0, 1]$.

$$\alpha_{ij}^l = \frac{e^{a_{\alpha_{ij}}^l}}{e^{a_{\alpha_{ij}}^l} + e^{a_{\beta_{ij}}^l} + e^{a_{\gamma_{ij}}^l}} \tag{2}$$

where $a_{\alpha_{ij}}^l, a_{\beta_{ij}}^l,$ and $a_{\gamma_{ij}}^l$ are the control parameters of three weights respectively. β_{ij}^l and γ_{ij}^l use the same definition as above.

3.3.2. Classification Regression Network and Priors Boxes Generation

After the data are processed by the WFAM module, five effective feature layers are generated. The EWFAN network generates a large number of a priori boxes on each effective feature layer, and every grid point of each layer generates 9 a priori boxes. The specific structure of the classification and regression network and the generation of a priori boxes are shown in Figure 6:

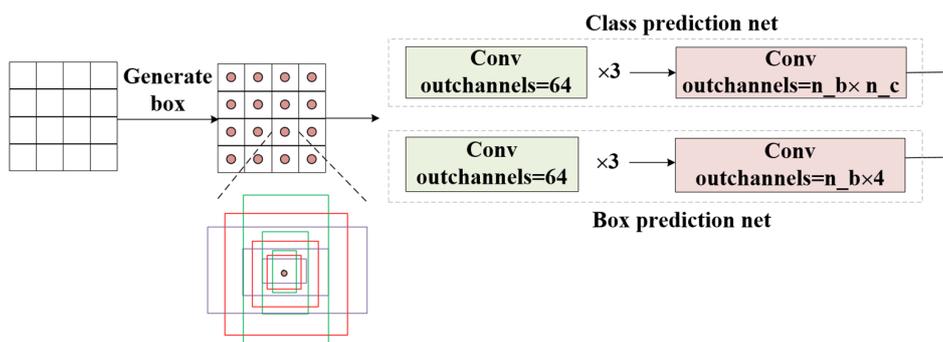


Figure 6. Classification regression network.

Among them, classification network uses three times of the convolution with 64 channels and one time of $n_b \times n_c$ convolution to predict the category of each prediction box (n_b refers to the number of priori boxes owned by the feature layer, n_c refers to the number of network target detection categories). Regression network utilizes three times of the convolution with 64 channels and 1 time of $n_b \times 4$ convolution to predict the regression of each a priori box. In addition, according to the aspect ratio distribution of the aircraft targets in the data set (the aircraft target aspect ratio distribution is shown in Figure 7), we change the aspect ratio of the a priori boxes to 0.6, 1.12, and 1.57 in the paper, which can improve the detection accuracy of the network to a certain extent compared with the original aspect ratio.

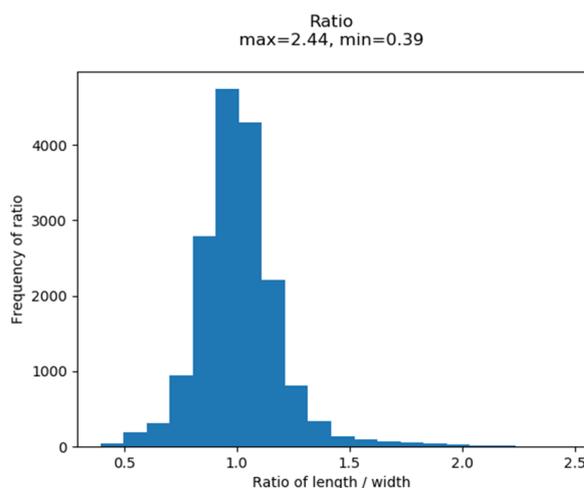


Figure 7. Aircraft target aspect ratio distribution.

3.3.3. CIF Loss Function CIF Loss Function

In the training stage, the loss function is used to calculate the difference between the network prediction result and the true value, then the optimizer is utilized to reduce the difference between the model output value and the true label. In this paper, CIF Loss is proposed, which integrated CIoU Loss and Focal Loss. CIF Loss can improve the regression accuracy and robustness of the network to a certain extent, thus further enhancing the aircraft detection effect of the network.

The traditional IoU Loss is defined as follows:

$$\text{IoU} = \frac{|B \cap B^t|}{|B \cup B^t|} \quad (3)$$

$$L_{\text{IoU}} = 1 - \text{IoU} \quad (4)$$

where B represents the prediction box, and B^t indicates the real box.

Obviously, in the traditional IoU Loss, the regression loss only considers the intersection ratio of the prediction box and the real box, which will affect the network training and testing effects. Figure 8 describes the problems of IoU Loss in different situations. (1) When two boxes have no intersection, the loss of a near non intersection box is the same as that of a faraway non intersection box, so the gradient direction is lost and the optimization cannot be implemented; (2) when the prediction boxes are contained by the real boxes, the prediction boxes with the same area in different positions are the same as the IoU of the real boxes, which will cause problems such as inaccuracy of the prediction boxes; (3) the aspect ratio between the two boxes is not considered. In the aircrafts detection, the shapes of the targets are different and the distribution is dense, and there may even be a small area overlap between the two real boxes. Therefore, if the aspect ratio is not considered, it will cause problems such as inaccurate prediction boxes and classification error.

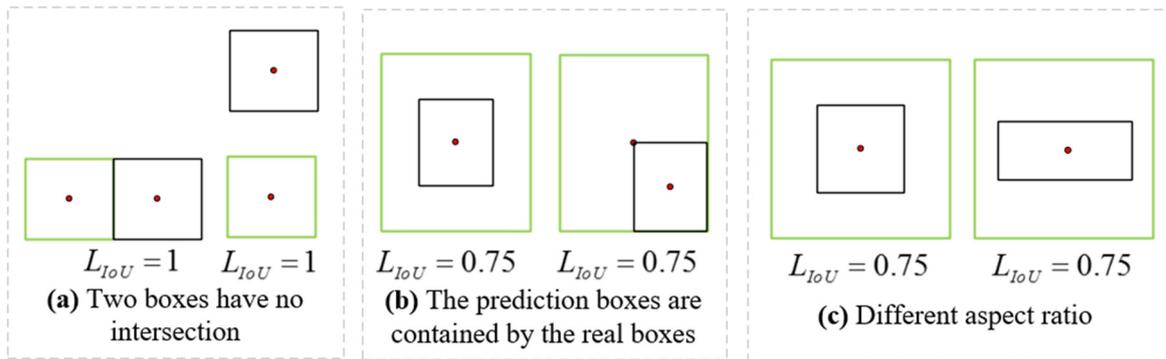


Figure 8. The problems of IoU Loss.

To solve the shortcomings of IoU Loss, CIoU Loss [28] has made improvements, which can improve the accuracy of aircrafts detection.

$$L_{CIoU} = 1 - IoU + \frac{d^2(x, x^t)}{a^2} + \mu\beta \quad (5)$$

$$\mu = \frac{\beta}{(1 - IoU) + \beta} \quad (6)$$

$$\beta = \frac{4}{\pi^2} \left(\arctan \frac{w^t}{h^t} - \arctan \frac{w}{h} \right)^2 \quad (7)$$

where x and x^t represent the center points of B and B^t respectively. $d(\cdot)$ is the Euclidean distance. a is the minimum diagonal length that can cover B and B^t . μ is the weight function of the aspect ratio, and β is used to measure the similarity of the aspect ratio. w and w^t represent the width of B and B^t respectively, and h and h^t indicate the height of B and B^t respectively.

The penalty term $\frac{d^2(x, x^t)}{a^2}$ takes into account the center distance between the prediction box and the real box. It not only solves the problem of no gradient direction when two boxes have no intersection, but also reduces the distance between the center points of the two boxes when the two boxes have intersection.

The penalty term $\mu\beta$ takes into account the difference in the aspect ratio between the predicted box and the real box, which can make the aspect ratio of the predicted box closer to that of the real box.

We combine CIoU Loss with Focal Loss and propose CIF Loss to further improve the accuracy of the network. Its definition is as follows:

$$L_{CIF} = L_{FL} + L_{CIoU} \quad (8)$$

$$L_{FL} = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (9)$$

$$p_t = \begin{cases} p, & y = 1 \\ 1 - p, & \text{otherwise,} \end{cases} \quad (10)$$

$$p_t = \begin{cases} p, & y = 1 \\ 1 - p, & \text{otherwise,} \end{cases} \quad (11)$$

where L_{FL} is the total classification loss. α_t is a balanced variant. p is the category prediction probability, and its value is between 0 and 1 [29]. y represents the label value, and y value of the positive sample is 1. L_{CIoU} indicates the total regression loss.

3.3.4. Non-Maximum Suppression (NMS)

In the testing stage, multiple prediction boxes with different confidence levels are generated on a target, which will greatly affect the detection effect. Therefore, we use

the NMS algorithm to filter the prediction boxes. The NMS algorithm retains the optimal prediction box and deletes all the remaining prediction boxes on this target, so that each target corresponds to only one prediction box, which can effectively reduce the number of false alarms. First, we sort all the prediction boxes on the detection image from high to low confidence and find the box A with the highest confidence. Then, the IoUs of A and all the other prediction boxes on the image are calculated and compared with the default IoU threshold (usually it is between 0–0.5, but it is set to 0.5 in this paper). If they are greater than the threshold, they will be deleted, otherwise they will be remained. Furthermore, it will continue to find the box with the highest confidence in the remaining prediction boxes. The above steps will be repeated until each detected target corresponds to only one prediction box. A simple schematic diagram of the NMS algorithm is shown in Figure 9.

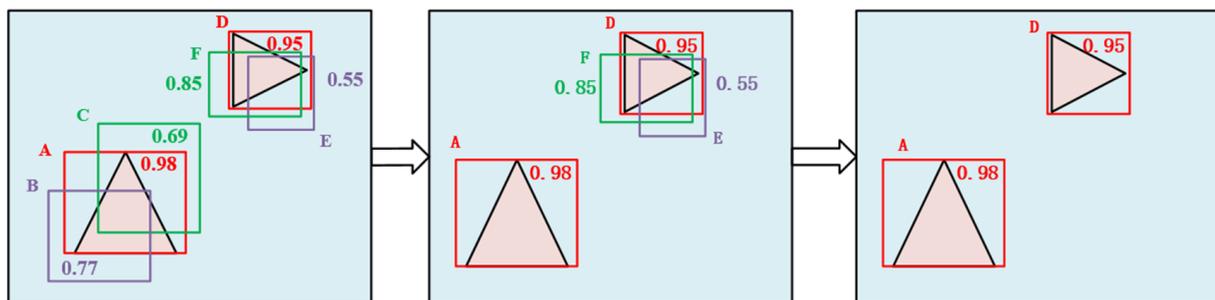


Figure 9. The schematic diagram of Non-Maximum Suppression (NMS).

3.3.5. Using Airport Masks to Remove False Alarms

After using the EWFAN algorithm to obtain the preliminary aircraft detection results, the airport mask will be performed to reduce the false alarms, which will remove the prediction box outside the runway areas to obtain the final aircraft detection results. Using airport mask can further reduce the false alarm rate, thereby effectively improving the overall detection effect.

4. Experimental Results and Analyses

4.1. Data Usage

In this article, the data used in the experiment is SAR images with 1 m resolution obtained by Gaofen-3 system. We use multiple large-scale SAR images including airports and aircrafts to perform training and testing. First, we use the RSlable tool to label the aircrafts in the SAR images, which are confirmed by the SAR expert. Then, we use the generated tag files and the original SAR images to automatically generate a data set, which contains 5480 aircraft slices with a size of 500×500 and the corresponding label files. In addition, the ratio of training set to validation set is 4:1. To validate the proposed framework in the paper, three SAR images including aircrafts which are unused in the datasets are utilized to generate the aircrafts detection results.

4.2. Hyperparameter Settings

This article builds all experimental environments under the Ubuntu 18.04 system, and trains the network model based on the pytorch framework and CUDA 10.0. All models are trained using the same data set. All models are optimized using the SGD algorithm, and each model is trained using an SGD optimizer with momentum 0.9 and weight decay 4×10^{-5} . each model is trained on two 2080Ti GPUs for 100 epochs. We employed an initial learning rate of 0.0005 and the batch size of 16. We also employed the aspect ratio {0.6,1.12,1.57}. The input image size is 512×512 . It should be noted that, for fairness, none of the three models use pre-training models and automatic enhancement functions. In the testing phase, the NMS thresholds of the three models are all 0.5. For the sake of fairness,

we do not use any advanced testing techniques such as Softer-NMS or data enhancement during testing. The test time we recorded is the running time of the entire framework.

4.3. Results Evaluation Method

In order to verify the effectiveness of the method in this paper, six evaluation indexes are used to measure the effect of the network. They are detection rate (DR) [8], false alarm rate (FAR) [5], missed alarm rate (MAR) [5], false alarm number (F) [5], missed alarm number (M) [5], and mAP [31] respectively. Among them, the detection rate represents the ratio of the number of aircraft targets correctly detected by the network (C) [5] to the number of aircraft targets in the label (L), and the false alarm rate is the ratio of the number of false alarms (they are not aircrafts, but they have been detected as aircrafts) to the number of prediction boxes last output by the network (S), and the missed detection rate is the ratio of the number of missed alarms (they are aircrafts, but they have not been detected) to the number of aircraft targets in the label. Q is the number of categories, $|R(q)|$ is the number of images relevant to the category q , k is the rank in the sequence of retrieved images, n is the total number of retrieved images, $P(k)$ is the precision at cutoff k in the list, and $r(k)$ is an indicator function whose value is 1 if the image at rank k is relevant and is 0 otherwise [31]. The specific calculation equations are as follows:

$$DR = \frac{C}{L} \quad (12)$$

$$FAR = \frac{F}{S} \quad (13)$$

$$MAR = \frac{M}{L} \quad (14)$$

$$mAP = \frac{1}{Q} \sum_{q=1}^Q AP(q) \quad (15)$$

$$AP = \frac{\sum_{k=1}^n (P(k) \times r(k))}{|R(q)|} \quad (16)$$

4.4. Analysis of the Role of Airport Detection Algorithms

In this article, we compare the detection results of aircrafts using airport mask with those when they are not used, as shown in Figure 10. Then the detection results of several SAR images are counted and analyzed, as shown in Tables 1 and 2. It can be seen from the experimental results that the false alarm rate of the aircraft detection is significantly reduced after using the airport mask algorithm proposed in this paper.

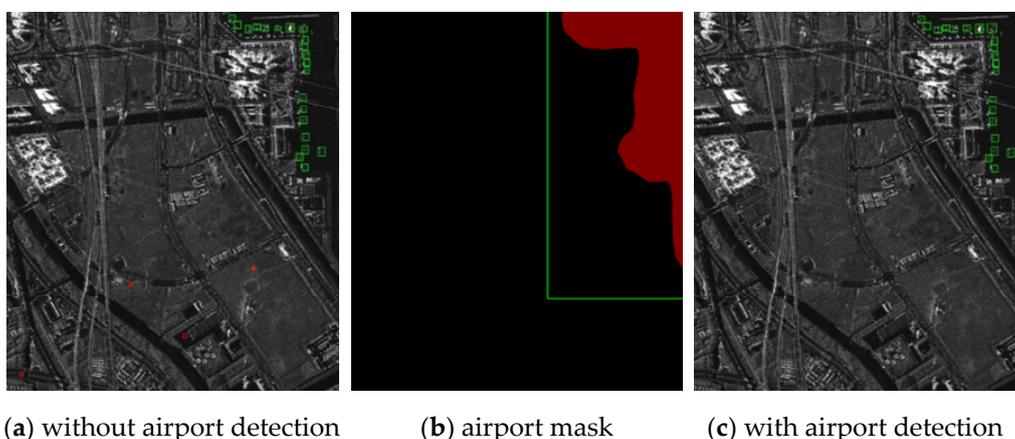


Figure 10. Effectiveness of the proposed airport detection algorithm. The green box represents the correctly detected aircraft, and the red box represents false alarms.

Table 1. Comparison of airport detection algorithms effectiveness of EfficientDet network.

Network	DR (%)	FAR (%)	MAR (%)	F	M
EfficientDet (raw) (512)	97.7	30.6	2.6	130	8
EfficientDet (mask) (512)	96.4	11.8	3.6	39	11

Table 2. Comparison of airport detection algorithms effectiveness of EWFAN network.

Network	DR (%)	FAR (%)	MAR (%)	F	M
EWFAN (raw) (512)	97.4	7.2	2.6	23	8
EWFAN (mask) (512)	95.4	3.3	4.6	10	14

In Figure 10, we show aircraft detection results with the airport mask and without airport mask. The green box represents the correctly detected aircraft, and the red box represents false alarms. It can be seen that there are 4 false alarms when the airport detection algorithm is not used, and there are no false alarms when the airport detection algorithm is used. Therefore, the false alarms of aircraft detection are significantly reduced after the airport detection algorithm is used.

Tables 1 and 2 (raw) indicate that the airport mask is not used, and (mask) indicates that the airport mask is used. (512) Represents the input size of the network. It can be seen that for the EfficientDet network, the false alarm is reduced by 18.8% when the airport mask is used, but the missed alarm rate is increased by 1.3%. The reason for the slight increase in the missed alarm rate is that the use of masks will remove aircraft targets that are not on the runway. For the EWFAN algorithm, the false alarm is reduced to only 3.3% after using the airport mask. Based on the comparison results of the two networks, it can be seen that the false alarm rate will be greatly reduced after the airport mask removal. In addition, although the use of mask to filter false alarms will lead to a slight increase in the missed alarm rate, the false alarm rate is greatly reduced, and the overall aircraft detection effect is still greatly improved.

4.5. The Influence of Different Window Overlap Rates on Aircraft Detection Performance

Table 3 shows the aircraft detection effects under different sliding window overlap ratios. The false alarm rate under 10% sliding window overlap rate is slightly lower than 20%, but the missed alarm rate under 10% sliding window overlap rate is much higher than 20%. The missed alarm rate under 30% sliding window overlap rate is slightly lower than 20%, but the false alarm rate under 30% sliding window overlap rate is 20% is much higher.

Table 3. Comparison of aircraft detection results with different window overlap rates.

Window Overlap Rates	DR (%)	FAR (%)	MAR (%)	F	M
EWFAN (10%)	93.4	2.8	6.6	8	20
EWFAN (20%)	95.4	3.3	4.6	10	14
EWFAN (30%)	95.6	5.5	4.3	17	13

Table 4 shows the aircraft detection time under different sliding window overlap ratios (the time includes airport detection time and mask time).

Table 4. Comparison of aircraft inspection time with different window overlap rates.

Window Overlap Rates	Mean
EWFAN (10%)	16.78 s
EWFAN (20%)	18.58 s
EWFAN (30%)	21.23 s

In order to trade off the missed alarm rate, false alarm rate, and detection time, this paper sets the overlap rate of the aircraft detection sliding window to 20%.

4.6. Aircrafts Detection Performance Analysis

In order to better demonstrate the performance of the algorithm proposed in this paper, three unused large airports images containing aircrafts are tested and analyzed.

4.6.1. Analysis of Aircrafts Detection for Airport I

The airport is Hongqiao Airport in Shanghai, China. In this airport, the aircrafts are small. In addition, there are many objects around the aircrafts that are similar to aircrafts, which will interfere with the aircraft detection results, thereby false alarms will be increased. In addition, the distribution of aircrafts in Hongqiao Airport is very dense, which also increases the difficulty of detection. The size of Hongqiao Airport is $12,000 \times 14,400$ pixels. Figure 11 shows the aircraft detection results of each network in Airport I. Figure 11a is the SAR image of Gaofen-3 system with 1 m resolution. Figure 11b is the aircrafts label (which is generated by SAR expert by comparing SAR image with the corresponding Google optical image). Figure 11c–e shows the detection results of EfficientDet algorithm, YOLOv4 algorithm, and proposed EWFAN algorithm respectively. The green box represents the correctly detected aircrafts, the red box represents the false alarms (they are not aircrafts, but they have been detected as aircrafts), and the yellow box represents the missed alarms (they are aircrafts, but they have not been detected aircrafts) aircrafts, the false alarms, and the missed alarms respectively. For the detail images, the detection result of the EWFAN algorithm has no false alarm, but YOLOv4 has 2 false alarms. While, EfficientDet algorithm has 4 false alarms, so it is much more than EWFAN algorithm. Neither YOLOv4 nor EfficientDet can well extract the features of aircrafts and suppress interference from difficult negative samples. In the EWFAN algorithm, the WFAM module can well enhance the saliency of positive samples and suppress the interference of negative samples, thus greatly reducing the generation of false alarm targets.

4.6.2. Analysis of Aircrafts Detection for Airport II

The aircrafts at Airport II are smaller than the aircraft at Hongqiao Airport. However, because the total number of aircrafts is less and the overall background information interference is not serious, the aircrafts detection effect for Airport II is very good. The size of Airport II is 9600×9600 pixels. Figure 12 shows the aircraft detection results of each network in Airport 2, where (a)–(e) are the same types of images as Figure 11. It can be seen from Figure 12 that there are no missing alarms in the detection results of the three networks. As for the false alarm, there are one and two false alarms for the EfficientDet algorithm (as shown in Figure 12c) and the YOLOv4 algorithm (as shown in Figure 12d) respectively, but there are no false alarms for the EWFAN algorithm (as shown in Figure 12e). Because the CIF Loss and WFAM modules in the EWFAN can effectively detect aircrafts and remove false alarms, so the detection effect is better than EfficientDet and YOLOv4.

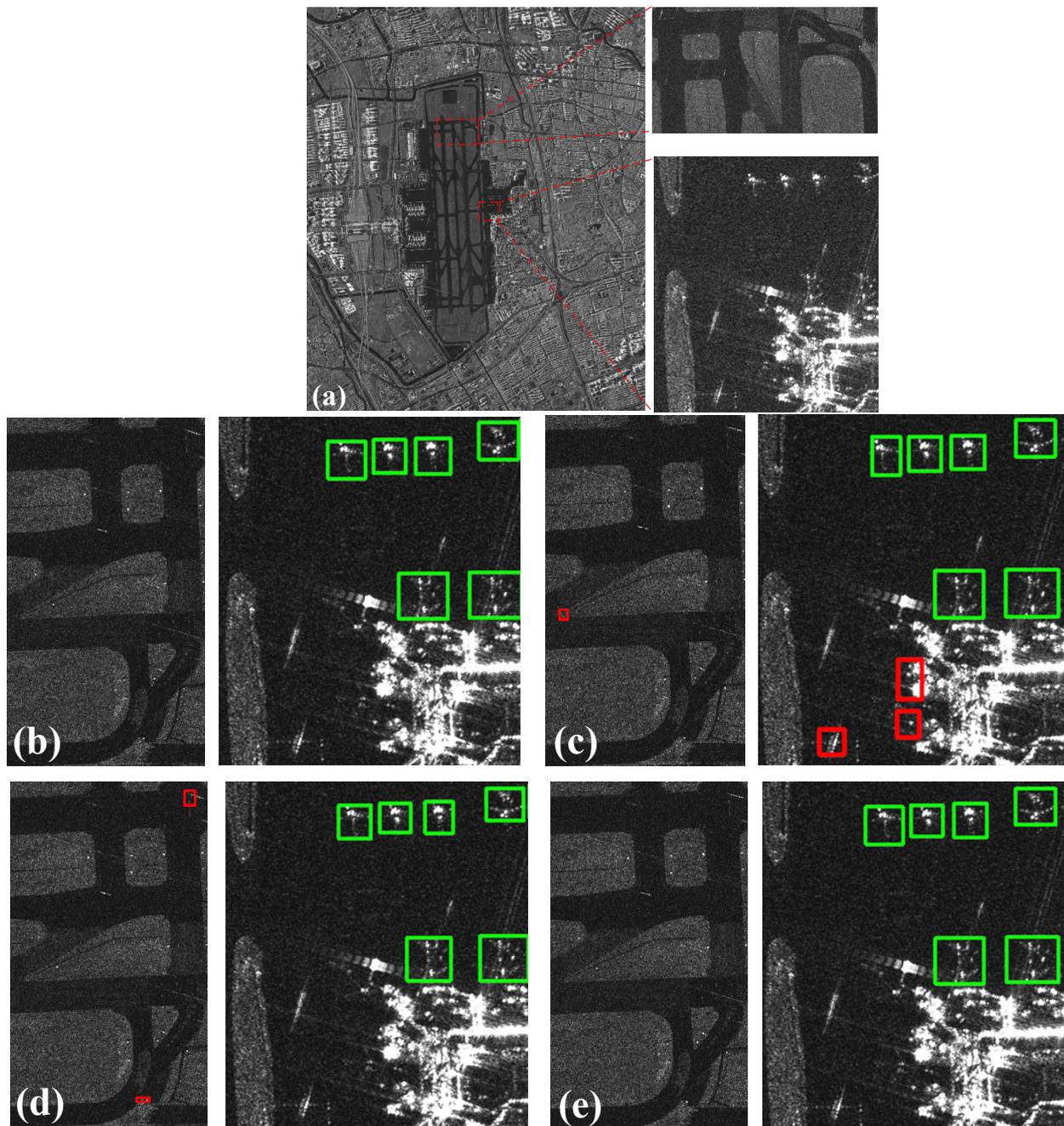


Figure 11. The experiment result for Airport I. (a) SAR image of Airport I from Gaofen-3. (b) The ground truth of the airport. (c–e) the aircraft detection results of (a) by EfficientDet, YOLOv4, and EWFAN. The green box, the red box, and the yellow box represent the correctly detected aircrafts, the false alarms, and the missed alarms respectively.

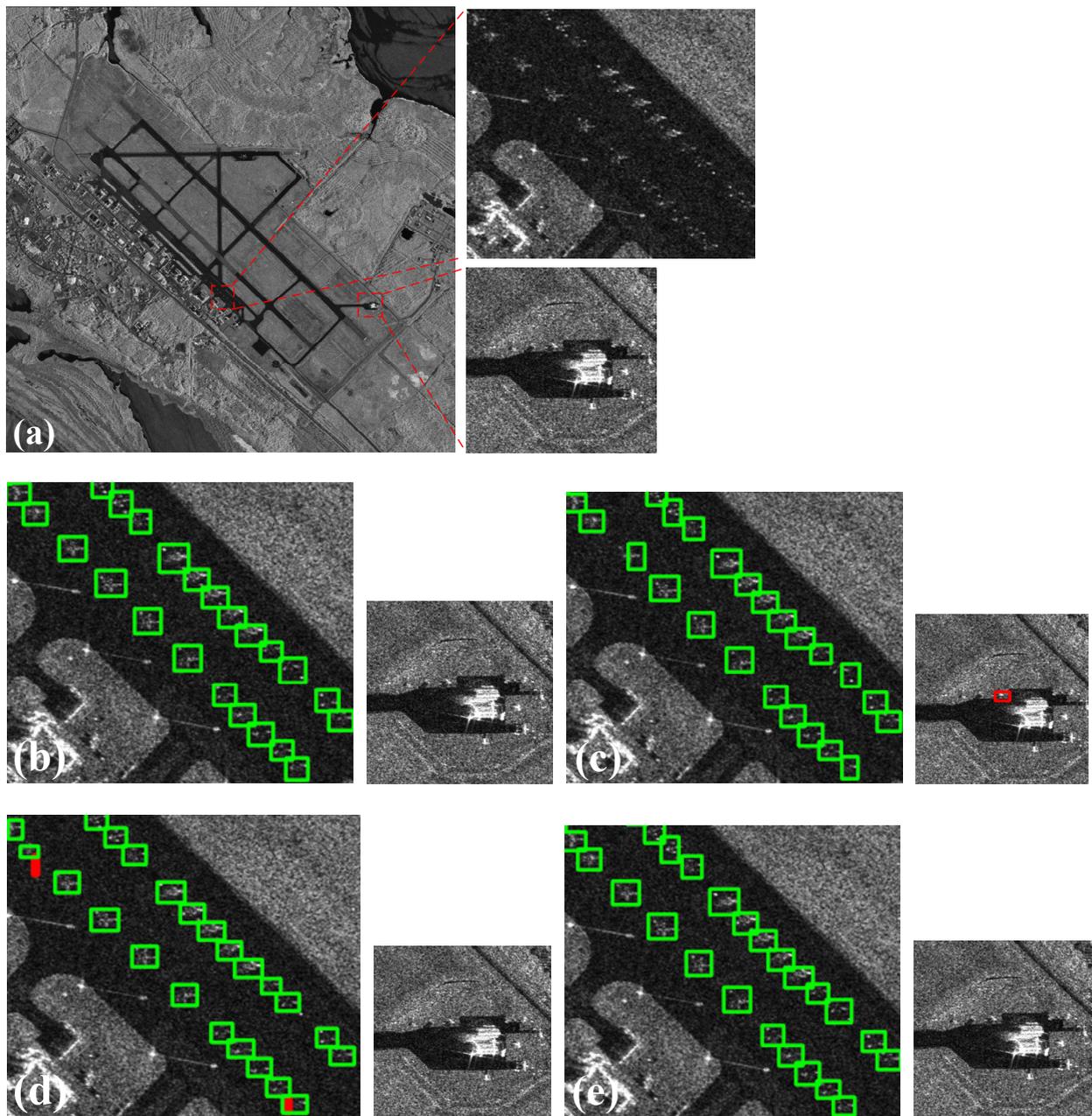


Figure 12. The experiment result for Airport I. (a) SAR image of Airport I from Gaofen-3. (b) The ground truth of the airport. (c–e) the aircraft detection results of (a) by EfficientDet, YOLOv4, and EWFAN. The green box, the red box and the yellow box represent the correctly detected aircrafts, the false alarms, and the missed alarms respectively.

4.6.3. Analysis of Aircrafts Detection for Airport III

Airport III is the Capital Airport in Beijing, China. Its background is more complicated. There are many bright areas formed by buildings, and there are lots of interference from targets with similar shapes to aircrafts, which may cause false alarms. The size of the airport III is $14,400 \times 16,800$ pixels. Figure 13 shows the aircraft detection results of the Capital Airport, where (a)–(e) are the same types of images as Figure 11. According to the detection results, the detection rates of EWFAN, YOLOv4, and EfficientDet are all high, but the false alarm rate is quite different. We can find that the detection result of the EWFAN algorithm is the best. There is only one false alarm in the detail map and no missing alarms (as shown in Figure 13e). The detection result of the EfficientDet algorithm is the worst. It has many false alarms (as shown in Figure 13c), and its prediction box is not accurate. This

shows that the regression accuracy of the EfficientDet algorithm is low. EWFAN algorithm introduces CIF Loss, which solves this problem well.

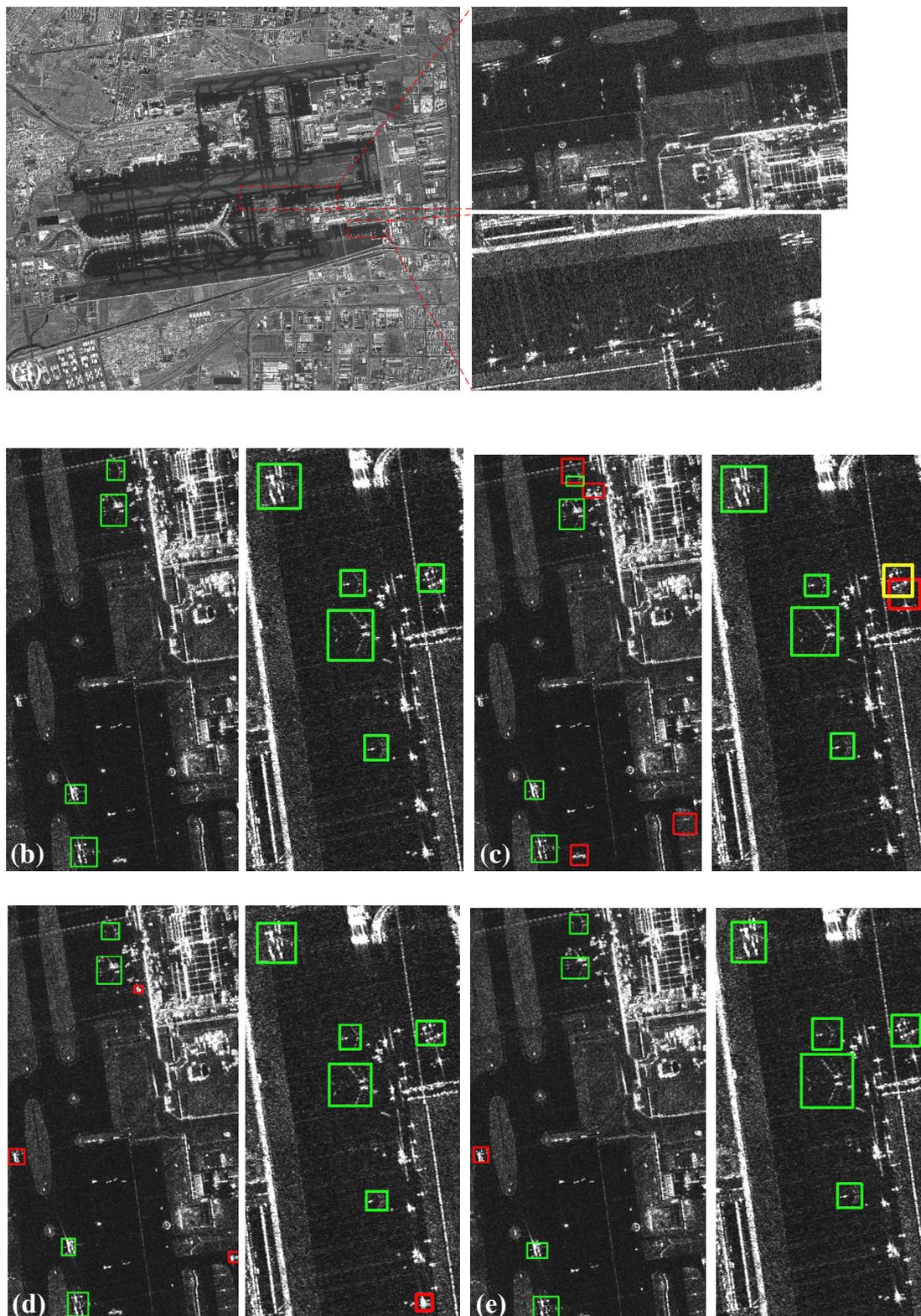


Figure 13. The experiment result for Airport I. (a) SAR image of Airport I from Gaofen-3. (b) The ground truth of the airport. (c–e) the aircraft detection results of (a) by EfficientDet, YOLOv4, and EWFAN. The green box, the red box, and the yellow box represent the correctly detected aircrafts, the false alarms, and the missed alarms respectively.

4.7. Analysis and Evaluation

Table 5 shows the aircraft detection results of SAR images with different networks when using GCAM airport detection algorithm [23] and mask to remove false alarms. Based on the analysis of the results of the three airports, the detection rates of EfficientDet, YOLOv4, and the algorithm proposed in this article are similar, and their detection rates are all over 95.0%. However, their false alarm rates vary greatly. The total false alarm rates of EfficientDet, YOLOv4, and EWFAN are 11.8%, 7.6%, and 3.3%, respectively. The false alarm rate of EfficientDet is the highest. The false alarm rate of YOLOv4 is lower compared with EfficientDet, but it is still higher than the EWFAN algorithm. The false alarm rate of EWFAN is greatly reduced compared with EfficientDet and YOLOv4, indicating that it can enhance the saliency of the target and suppress the interference of background information. Table 6 shows the results of different network on our validation dataset, the mAP of EfficientDet, YOLOv4, and EWFAN are 92.4%, 95.1%, and 97.9%, respectively. Combining Tables 1, 5 and 6, we can see that the aircraft detection framework proposed in this paper can greatly improve the overall detection effect.

Table 5. Comparison of aircraft detection results from SAR images of different network.

	Airport	DR (%)	FAR (%)	MAR (%)	F	M
EfficientDet (512)	Airport I	98.4	6.7	1.6	9	2
	Airport II	100	2.9	0	1	0
	Airport III	93.0	17.8	6.3	29	9
	Total	96.4	11.8	3.6	39	11
YOLOv4 (512)	Airport I	98.4	4.6	1.6	6	2
	Airport II	100	5.7	0	2	0
	Airport III	92.3	10.8	7.7	16	11
	Total	95.7	7.6	4.3	24	13
EWFAN (Ours) (512)	Airport I	98.4	0.8	1.6	1	2
	Airport II	100	0	0	0	0
	Airport III	91.6	6.4	8.4	9	12
	Total	95.4	3.3	4.6	10	14

Table 6. Results of different network on our validation dataset.

Network	EfficientDet	YOLOv4	EWFAN (Ours)
mAP (%)	92.4	95.1	97.9

Tables 7 and 8 respectively show the test time of different networks before and after using the airport detection algorithm. (512) Represents the input size of the network. When the airport detection algorithm is not used, the average test time of EfficientDet, YOLOv4, and EWFAN in this paper are 18.20 s, 15.91 s, and 18.58 s, respectively. Among them, the test time of the proposed network in this paper only increased by 0.38 s compared with EfficientDet, which can be almost ignored. After using the airport detection algorithm, the average test time of the EfficientDet, YOLOv4, and EWFAN algorithms are 15.25 s, 14.66 s, and 15.40 s respectively, which indicates that the time of detecting aircrafts can be shortened by introducing the airport mask. It can be seen from Table 9 that the average extraction time of the airport area is 7.89 s. For the EWFAN algorithm, the average running time of the entire framework while using the airport mask algorithm is 15.40 s, which is 3.18 s shorter than that of the unused airport extraction algorithm. This shows that the SAR image aircraft detection framework proposed in this paper can not only significantly improve the detection effect, but also shorten the detection time.

Table 7. Comparison of the test time from SAR images of different network (without airport mask algorithm).

Network	Airport I	Airport II	Airport III	Mean
EfficientDet	17.97 s	8.89 s	27.75 s	18.20 s
YOLOv4	15.61 s	8.09 s	24.04 s	15.91 s
EWFAN (Ours)	18.21 s	9.03 s	28.50 s	18.58 s

Table 8. Comparison of the test time from SAR images of different network (with airport mask algorithm).

Network	Airport I	Airport II	Airport III	Mean
EfficientDet	12.79 s	11.12 s	21.85 s	15.25 s
YOLOv4	12.40 s	10.90 s	20.68 s	14.66 s
EWFAN (Ours)	12.92 s	11.19 s	22.07 s	15.40 s

Table 9. Airport extraction time from SAR image.

	Airport I	Airport II	Airport III	Mean
Time	7.60 s	6.73 s	9.33 s	7.89 s

5. Discussion

In this paper, an effective framework for aircraft detection in SAR images is proposed. This framework combines the airport detection mask with the EWFAN algorithm invented in this paper. It can actively remove false alarms, greatly improve the aircraft detection accuracy, and significantly reduce the detection time. The success of this new aircraft detection framework has proved the necessity of combining different deep neural networks in SAR image analytics.

The missed detection of aircrafts turns out to be an interesting topic for further investigation. The airport detection algorithm can quickly locate the airport area, significantly reducing the calculation amount and false alarm rate of aircraft detection. The mask removal method can further reduce the false alarm rate of aircrafts detection. Frequently, a small number of aircraft targets would be placed outside the airport runway area, which slightly increases the missed alarm. We plan to conduct in-depth analysis and address the detection of aircrafts in the flight in our future work.

We also highlight the supplementary improvement of EWFAN. This paper proposes an efficient deep learning network EWFAN for dense object detection. Compared with EfficientDet and YOLOv4, EWFAN has unique advantages in aircraft detection in SAR images. These satisfactory results are not only coming from the reasonable architecture of our EWFAN, but also the precise annotation of similar situations in our training dataset. In EWFAN, the backbone, feature fusion module, classifier, and detector are specially designed to make the training converge well. In addition, the proposed framework could be employed in near real-time aircraft detection. However, we also noticed that although the false alarm rate of EWFAN is very low, the missed alarm rate is almost the same as EfficientDet and YOLOv4. We will explore corresponding solutions in our future study.

6. Conclusions

In this paper, an efficient framework for aircrafts detection in SAR images is proposed, which includes airport detection algorithm, EWFAN algorithm, improved loss function, and mask removal method to reduce false alarms. The proposed framework achieves low false alarm rate and missed alarm rate; and the aircraft detection testing at the three airports only takes 15.40 s on average, which indicates prominent detection efficiency of the framework. Main contributions and future research directions of this paper are summarized as follows:

(1) An end-to-end aircraft detection framework is proposed in this paper. It uses the airport detection algorithm GCAM to obtain the airport runway area first. Then the airport rectangular contour is quickly obtained based on the airport runway area. In addition, aircrafts detection is performed in the rectangular box of the airport. Finally, detection results are masked by the extracted runway areas to generate the final aircrafts detection. This framework greatly reduces false alarms and improves aircraft detection efficiency, which can be utilized broadly in aircraft detection studies.

(2) An innovative aircraft detection network EWFAN is proposed in this paper. The main contributions of EWFAN are the WFAM module and the CIF loss function. In this network, BiFPN, RSAM, and ASFF are integrated to achieve effective extraction of aircraft features, which thereby can greatly reduce background interference and improve aircraft detection accuracy. It also provides a valid methodology of combining various deep neural networks for targets detection.

(3) Following the initial experiment in this paper, we will employ other SAR images. The SAR image aircraft detection framework proposed in this paper can detect aircrafts automatically and efficiently, and it is also applicable to other types of dense target detection, such as vehicles and ships.

Author Contributions: Data curation, L.C., Z.P., R.L. and X.C.; Formal analysis, J.W.; Funding acquisition, J.W. and J.X.; Investigation, H.X., L.C., R.L. and X.C.; Methodology, J.W., H.X., L.C., J.X., R.L. and X.C.; Project administration, H.X., L.C. and Z.P.; Validation, J.W.; Visualization, J.W.; Writing—Original draft, J.W.; Writing—Review & editing, H.X., L.C. and J.X. All authors contributed to the study, reviewed and approved the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by the National Natural Science Foundation (No. 41201468, 41701536, 61701047, 41674040) of China, partly funded by Natural Science Foundation of Hunan Province (No. 2017JJ3322, 2019JJ50639), partly funded by the Foundation of Hunan, Education Committee, under Grants No. 16B004, No.16C0043.

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

References

1. Steenson, B.O. Detection performance of a mean-level threshold. *IEEE Trans. Aerosp. Electron. Syst.* **1968**, *4*, 529–534. [[CrossRef](#)]
2. Finn, H.M.; Jonson, R.S. Adaptive detection mode with threshold control as a function of spatially sampled clutter-level estimates. *RCA Rev.* **1968**, *29*, 414–464.
3. Goldstein, G.B. False Alarm Regulation in Long-normal and Weibull Clutter. *IEEE Trans. Aerosp. Electron. Syst.* **1973**, *9*, 84–92. [[CrossRef](#)]
4. Tan, Y.; Li, Q.; Yan, S. Aircraft detection in high-resolution SAR images based on a gradient textural saliency map. *Sensors* **2015**, *15*, 23071–23094. [[CrossRef](#)] [[PubMed](#)]
5. Guo, Q.; Wang, H. Research progress on aircraft detection and recognition in SAR imagery. *J. Radars* **2020**, *9*, 497–513.
6. Zhao, Q.; Principe, J.C. Support vector machines for SAR automatic target recognition. *IEEE Trans. Aerosp. Electron. Syst.* **2001**, *37*, 643–654. [[CrossRef](#)]
7. Sun, Y.; Liu, Z.; Todorovic, S. Adaptive boosting for SAR automatic target recognition. *IEEE Trans. Aerosp. Electron. Syst.* **2007**, *1*, 112–125. [[CrossRef](#)]
8. Wang, S.; Gao, X.; Sun, H. An aircraft detection method based on convolutional neural networks in high-resolution SAR images. *J. Radars* **2017**, *6*, 195–203.
9. Zhengxia, Z.; Zhenwei, S. Object Detection in 20 Years: A Survey. *arXiv* **2019**, arXiv:1905.05055v2.
10. 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, Columbus, OH, USA, 23–28 June 2014; 2014; pp. 580–587.
11. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 1440–1448.
12. He, K.; Zhang, X.; Ren, S. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 1904–1916.
13. He, K.; Gkioxari, G.; Dollar, P. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.

14. 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
15. Tan, M.; Pan, R. EfficientDet: Scalable and Efficient Object Detection. *arXiv* **2019**, arXiv:1911.09070v1.
16. Bochkovskiy, A.; Wang, C. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934v1.
17. Li, C.; Zhao, L.; Kuang, G. A Two-stage Airport Detection Model for Large Scale SAR Images Based on Faster R-CNN. In Proceedings of the Eleventh International Conference on Digital Image Processing, Guangzhou, China, 10–13 May 2019; pp. 515–525.
18. Ren, S.; Kaiming, H.; 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](#)]
19. Zhao, Y.; Zhao, L.; Kuang, G. Fast detection aircrafts in complex large scene SAR images. *Chin. J. Radio Sci.* **2020**, *35*, 594–602.
20. Chen, L.; Tan, S.; Pan, Z. A New Framework for Automatic Airports Extraction from SAR Images Using Multi-Level Dual Attention Mechanism. *Remote Sens.* **2020**, *12*, 560. [[CrossRef](#)]
21. Zhang, P.; Chen, L.; Xing, J.; Li, Z.; Xing, X.; Yuan, Z. Water and shadow extraction in SAR image based on a new deep learning network. *Sensors* **2019**, *19*, 3576. [[CrossRef](#)] [[PubMed](#)]
22. Chen, L.; Zhang, P.; Xing, J.; Li, Z.; Xing, X.; Yuan, Z. A Multi-scale Deep Neural Network for Water Detection from SAR Images in the Mountainous Areas. *Remote Sens.* **2020**, *12*, 3205. [[CrossRef](#)]
23. 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](#)]
24. Chen, L.; Cui, X.; Li, Z.; Yuan, Z.; Xing, J.; Xing, X. A new Deep Learning Algorithm for SAR Scenes Classification Based on Space Statistical Modeling and Features Re-calibration. *Sensors* **2019**, *19*, 2479. [[CrossRef](#)] [[PubMed](#)]
25. Tan, S.; Chen, L.; Pan, Z. Geospatial Contextual Attention Mechanism for Automatic and Fast Airport Detection in SAR Imagery. *IEEE Access* **2020**, *8*, 173627–173640. [[CrossRef](#)]
26. Tan, M. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv* **2019**, arXiv:1905.11946v3.
27. Liu, S.; Huang, D.; Wang, Y. Learning Spatial Fusion for Single-Shot Object Detection. *arXiv* **2019**, arXiv:1911.09516v2.
28. Zheng, Z.; Wang, P. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *arXiv* **2019**, arXiv:1911.08287v1.
29. Lin, T.; Goya, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *2*, 318–327. [[CrossRef](#)] [[PubMed](#)]
30. Woo, S.; Park, J. CBAM: Convolutional Block Attention Module. *arXiv* **2018**, arXiv:1807.06521v2.
31. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The Pascal visual object classes (VOC) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [[CrossRef](#)]