remote sensing

MDPI

Article

A Fast Aircraft Detection Method for SAR Images Based on Efficient Bidirectional Path Aggregated Attention Network

Ru Luo [1,2], Lifu Chen [1,2,*], Jin Xing [3], Zhihui Yuan [1,2], Siyu Tan [1,2], Xingmin Cai [1,2] and Jielan Wang [2,4]

[1] School of Electrical and Information Engineering, Changsha University of Science & Technology, Changsha 410114, China; Luoru@stu.csust.edu.cn (R.L.); yuanzhihui@csust.edu.cn (Z.Y.); tansiyu@stu.csust.edu.cn (S.T.); cc182022@stu.csust.edu.cn (X.C.)

[2] Laboratory of Radar Remote Sensing Applications, Changsha University of Science & Technology, Changsha 410014, China; ningyuan@stu.csust.edu.cn

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

[4] School of Computer and Communication Engineering, Changsha University of Science & Technology, Changsha 410114, China

* Correspondence: lifu_chen@csust.edu.cn; Tel.: +86-182-2997-5986

Abstract: In aircraft detection from synthetic aperture radar (SAR) images, there are several major challenges: the shattered features of the aircraft, the size heterogeneity and the interference of a complex background. To address these problems, an Efficient Bidirectional Path Aggregation Attention Network (EBPA2N) is proposed. In EBPA2N, YOLOv5s is used as the base network and then the Involution Enhanced Path Aggregation (IEPA) module and Effective Residual Shuffle Attention (ERSA) module are proposed and systematically integrated to improve the detection accuracy of the aircraft. The IEPA module aims to effectively extract advanced semantic and spatial information to better capture multi-scale scattering features of aircraft. Then, the lightweight ERSA module further enhances the extracted features to overcome the interference of complex background and speckle noise, so as to reduce false alarms. To verify the effectiveness of the proposed network, Gaofen-3 airports SAR data with 1 m resolution are utilized in the experiment. The detection rate and false alarm rate of our EBPA2N algorithm are 93.05% and 4.49%, respectively, which is superior to the latest networks of EfficientDet-D0 and YOLOv5s, and it also has an advantage of detection speed.

Keywords: synthetic aperture radar; aircraft detection; deep learning; involution; residual attention

check for updates

Citation: Luo, R.; Chen, L.; Xing, J.; Yuan, Z.; Tan, S.; Cai, X.; Wang, J. A Fast Aircraft Detection Method for SAR Images Based on Efficient Bidirectional Path Aggregated Attention Network. *Remote Sens.* **2021**, *13*, 2940. https://doi.org/10.3390/rs13152940

Academic Editor: Andrea Monti Guarnieri

Received: 19 June 2021
Accepted: 23 July 2021
Published: 27 July 2021

1. Introduction

Synthetic aperture radar (SAR) can provide continuous and stable observation all day and all night, which has been widely used in various fields [1]. With the fast development of SAR techniques, a large number of high-resolution spaceborne and airborne data have been acquired, which provides new opportunities for SAR targets detection. Nowadays, timely aircraft detection plays a pivotal role in airport management and military activities [2]. This is because the scheduling and placement of aircraft are sensitive spatio-temporally.

Currently, there are three major challenges in aircraft detection: the shattered image features of aircraft, their size heterogeneity, and the interference of complex background. In SAR imaging, features of the aircraft are composed of a series of discrete scattering points, unlike the obvious fuselage and wing information of the aircraft in optical images, which are not easy for visual interpretation. In addition, the size heterogeneity between aircraft cannot be ignored, which could be different significantly among small and big aircraft. Small aircraft are more likely to be missed, resulting in a lower detection rate of the algorithm. Moreover, facilities around the aircraft could be recorded as features similar to those of the aircraft due to scattering [3], which further increases the difficulty of aircraft detection in SAR images. Therefore, it is essential for detection algorithms to recognize effective features of the aircraft.

In recent years, with the rapid development of deep learning technologies, more advanced deep learning models have been widely applied in the field of object detection. Convolutional neural networks (CNNs) have strong feature extraction ability with end-to-end structures, which makes them the mainstream deep learning algorithms for target detection [4]. The You Only Look Once (YOLO) series is a typical one-stage algorithm, which is famous for its fast detection speed. In 2015, Redmon et al. proposed the YOLO [5] algorithm to predict the targets by directly performing classification regression on the input image. The average accuracy of the YOLO algorithm on Pascal VOC 2007 testing dataset was 63.4%, sacrificing about 10% of detection accuracy compared with Fast R-CNN [6], but the detection speed was much higher than Fast R-CNN. In 2017, The YOLOv2 [7] with Darknet-19 as the backbone network was proposed, in which the anchor mechanism based on k-means clustering calculation was used to improve the detection accuracy. In 2018, Redmon et al. [8] utilized multi-scale prediction to obtain detection results, which greatly improved the accuracy of small targets detection. However, with continuous improvement of YOLO network performance, the number and complexity of parameters were sky-scrapping, and the detection speed of the network was also decreasing. In 2019, Cross Stage Partial Network (CSPNet) [9] was proposed, which effectively solved the gradient information repetition problem in the optimization of large convolutional neural networks, and ensured the speed and accuracy of training. In 2020, YOLOv4 [10] was proposed to further improve the speed of detection. CSPDarknet53 was used as the backbone network of feature extraction, as a typical lightweight framework. Subsequently, the appearance of Scaled-YOLOv4 [11] and YOLOv5 [12] attracted more attention of researchers and the industry for its advantages of the fast speed and satisfying accuracy.

Inspired by these advanced methods, many remote-sensing experts tried to introduce deep learning into the field of SAR target detection. In the large-view SAR image, the aircraft belongs to the category of small target, which occupies less pixels and is very easily affected by background factors. However, due to the unique imaging mechanism of SAR, the complexity of the background region of SAR image has been much larger than that of optical images, which needed higher requirements for the robustness of the detection network. To solve the above problems, some researchers have firstly detected the airport and then detected aircraft in the airport area to reduce unnecessary background information. This method can effectively improve the efficiency of aircraft detection. Furthermore, combined with the runway-mask method, false alarms have been further reduced [13–16]. However, the mask method would possibly miss aircraft located outside the airport runway areas, and the extraction precision of the runway might undermine the detection performance of the aircraft. In addition, some experts proposed to combine saliency information to guide convolutional neural networks to highlight information region of the target, and then extract robust features. Du et al. [17] proposed a saliency-guided single shot multi-box detector (S-SSD) for target detection in SAR images. In their experiments, the performance of the S-SSD network has been significantly improved. However, compared with the original SSD, S-SSD network adds 18.7 M extra parameters, which makes the detection speed slower than SSD, and there is still room for optimization in the detection speed.

The attention mechanism is similar to the visual pattern of human brains, which adaptively extracts the region of interest and then focuses more on capturing effective information of small targets. Due to the small number of parameters carried by attention, the addition of an attention mechanism will not significantly prolong the detection efficiency of the network. Many scholars have explored and applied attention mechanisms to improve the performance of deep neural networks. Combined with the characteristics of different SAR image interpretation tasks, Chen et al. [18] proposed a multi-level and densely dual attention (MDDA) network to realize the high-precision automatic extraction of airport runway from high-resolution SAR images. Zhang et al. [19] carefully designed a network based on attention mechanism, and achieved good results in water detection from SAR image. For aircraft detection, Zhao et al. [20] designed a pyramid attention dilated network

(PADN) to enhance the learning of aircraft scattering characteristics, and the detection rate of the network was 85.68% on the data of Gaofen-3 airport. Subsequently, Guo et al. [21] have proposed scattering information enhancement module to preprocess the input data to highlight scattering characteristics of the target; then the enhanced data were input into the attention pyramid network for aircraft detection. In their experiment, the average precision (AP) was 83.25%, but the complexity of the whole network should also be noticed.

YOLOv5s is the lightest model among all YOLOv5 networks in depth and width, which presents excellent speed and precision in automatic object detection. Therefore, we choose it as the backbone network and further combined it with our own Efficient Bidirectional Path Aggregation Attention Network (EBPA2N) for efficient aircraft detection. The main contributions of this paper are summarized as follows:

(1) An effective and efficient aircraft detection network EBPA2N is proposed for SAR image analytics. Combined with the sliding window detection method, an end-to-end aircraft detection framework based on EBPA2N was established, which offers accurate and real-time aircraft detection from large-scale SAR images.

(2) As far as we know, we are the first to apply involution in SAR image analytics. We invented the Involution Enhanced Path Aggregation (IEPA) and Effective Residual Shuffle Attention (ERSA) module in an independent efficient Bidirectional Path Aggregation Attention Module (BPA2M). The IEPA module is proposed to capture the relationship among aircraft's backscattering features to better encode multi-scale geospatial information. As the basic module of the IEPA module, involution redefines the design method of feature extraction. By contrast with the traditional standard convolution, it uses different involution kernels in different spatial positions (i.e., spatial specificity) to integrate pixel spatial information, which is more conducive to establishing the correlation between aircraft scattering features in SAR images. On the other hand, the ERSA module mainly focuses on the scattering features information of the target and suppresses the influence of background clutter, then the influence of speckle noise in SAR images can be reduced.

(3) Our experiment has proved the outstanding performance of EBPA2N, which indicates the success of implementing multi-scale SAR image analytics as geospatial attention within deep neural networks. This paper has paved the path for further integration of SAR domain knowledge and advanced deep learning algorithms.

The rest of this paper is organized as follows. In Section 2, the aircraft detection framework of SAR images proposed in this paper is introduced in detail. Section 3 presents the experimental results and corresponding analysis of aircraft detection on three airport SAR data with different networks. In Section 4, the experimental results are discussed and the future research direction is proposed. Section 5 briefly summarizes the results of this paper.

2. Methodology

2.1. Overall Detection Framework

In this paper, an Efficient Bidirectional Path Aggregation and Attention Network (EBPA2N) is proposed for aircraft automatic detection in SAR images as shown in Figure 1. Based on the trade-off between the speed and the precision, the YOLOv5s backbone network was selected for feature extraction to achieve substantial representation. Then, the last three output feature maps C_3, C_4, and C_5 of the backbone network were chosen to inject the bidirectional path aggregation and attention module (BPA2M) to enrich the expression of features. BPA2M is a new object detection module proposed in this paper. It consists of IEPAM and three parallel ERSA. The IEPA module was used to fuse three feature maps of different sizes output from backbone network to learn multi-scale geospatial information. Then, three parallel attention mechanisms were used to refine the multi-scale features of the aircraft. Furthermore, a classification and box prediction network was used to obtain preliminary prediction results. The Non-Maximum Suppression (NMS) [22] method was employed to remove overlapping prediction boxes to obtain detection results. Last but not the least, a new sliding window algorithm is proposed to improve the detection efficiency of large-scale SAR images.

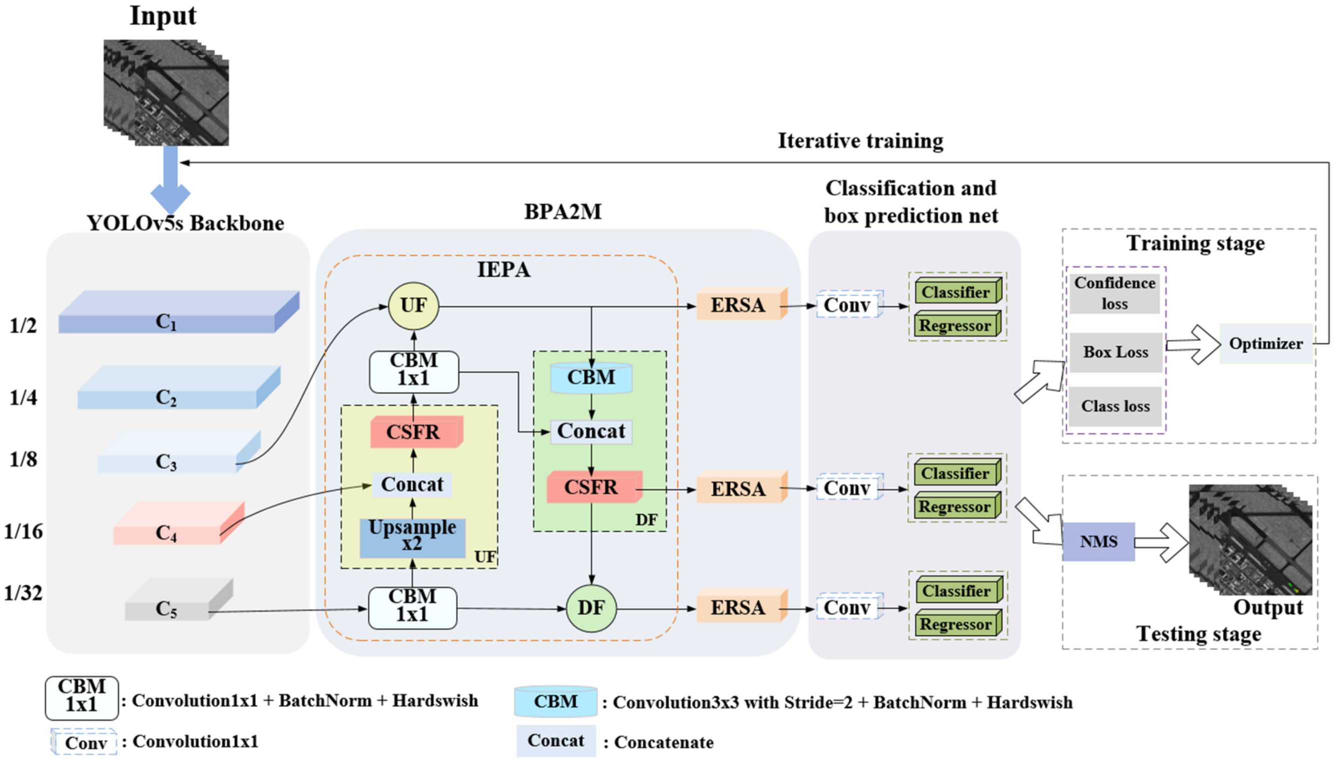

Figure 1. The architecture of the Efficient Bidirectional Path Aggregation and Attention Network (EBPA2N).

2.2. *YOLOv5s Backbone*

The input data with the size of 512 × 512 were sliced to generate feature map $C_1 \epsilon R^{32\times256\times256}$. In this paper, the three-dimensional tensor is expressed as $X \epsilon R^{C\times H\times W}$, where C, H and W represent the channel dimension, height and width of the feature map. Then, through four feature extraction modules with downsampling, rich image features are extracted to form feature maps as $C_2 \epsilon R^{64\times128\times128}$, $C_3 \epsilon R^{128\times64\times64}$, $C_4 \epsilon R^{256\times32\times32}$, and $C_5 \epsilon R^{512\times16\times16}$. In particular, the spatial pyramid pooling [23] module is embedded in the top feature extraction block. By using multi-scale pooling to construct features on multiple receptive fields to learn the multi-scale characteristics of aircraft. The detailed internal structure of the YOLOv5s backbone network can be referred to the introduction in [24,25].

2.3. *Bidirectional Path Aggregation and Attention Module (BPA2M)*

2.3.1. Involution Enhanced Path Aggregation (IEPA) Module

For the multi-level feature map formed by the backbone network, the low-level feature maps can provide rich spatial details of the target which is helpful for target localization. High-level feature maps with abundant semantic information play an important role in distinguishing foreground from background. Therefore, the effective fusion of feature maps with different scales is helpful to improve the detection accuracy of targets.

In this paper, IEPA module is proposed to adequately fuse the feature maps of C_3, C_4 and C_5 output from the backbone network. As shown in Figure 2a. In IEPA module, the 1 × 1 convolution module is used to adjust the number of channels. The stacking of up fusion (UF) and down fusion (DF) modules forms a bidirectional path, which effectively fuses shallow detail features and deep semantic information to achieve complementary advantages, which is conducive to improving the detection rate of targets with different scales. The Cross Stage Feature Refinement (CSFR) module is used in both UF and DF modules. The CSFR module is a new module proposed in this paper, which is inspired by involution [26]. It aims to learn features in a large receptive field to establish the relationship between discrete features of the aircraft, and then improve the performance of the network for aircraft detection.

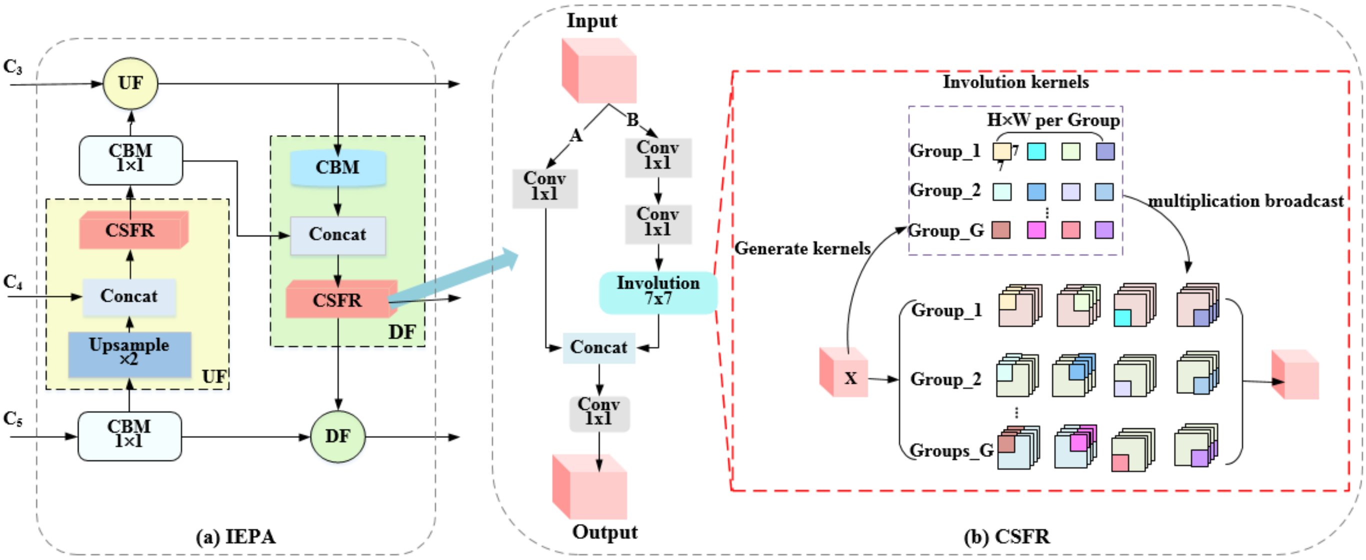

Figure 2. The structure of Involution Enhanced Path Aggregation (IEPA) and Cross Stage Feature Refinement (CSFR) module.

As shown in Figure 2b, in the CSFR module, the input feature map is processed by two respective branches. In these two branches, the input feature map adopts 1 × 1 convolution reducing the dimension of features by half to reduce the amount of computation. In branch B, the input feature map after dimension reduction is input into the 1 × 1 convolution module to learn cross channel information interaction. Then, 7 × 7 involution is used to capture the relationship between the scattering features of aircraft in a large range to obtain rich feature representation ability. In the involution, in order to speed up the operation of the network, the input feature map $X \epsilon R^{C \times H \times W}$ are divided into several groups of sub-feature maps by feature grouping, and then they are processed in parallel. Meanwhile, two 1 × 1 convolutions are used as the involution kernels generation function, and then the corresponding involution kernels are adaptively generated according to each pixel of the feature map. Therefore, each group of sub-feature maps with the size of $H \times W$ will generate a corresponding involution group, which contains $H \times W$ involution kernels. Different involution kernels are represented by different colors in Figure 2b. It is easy to know that the whole involution kernel and the input feature map are automatically aligned in the spatial dimension. Furthermore, each group of feature maps is multiplied with the involution kernels of the corresponding group to learn features. That is, each group of feature maps share the same involution kernels in the channel dimension, and different involution kernels are used in different spatial locations to learn the spatial visual patterns of different spatial regions. Subsequently, all the group feature maps are aggregated to obtain the output of involution module. Finally, branch A and branch B are concatenated, and then the channel information is fused by 1 × 1 convolution to get the output of CSFR module. Importantly, the number of parameters in the CSFR module is less than the standard 3 × 3 convolution, which also facilitates network lightweighting.

2.3.2. Effective Residual Shuffle Attention (ERSA) Module

The feature map processed by the IEPA module contains abundant image information. The scattering characteristics of aircraft should be highlighted timely and effectively, which is beneficial to improve the detection performance of aircraft. The attentional mechanism is a signal processing mechanism similar to the human brain, which can select high-value information relevant to the task at hand from a large amount of information. Chen et al. [27] revealed the effectiveness of using the attention mechanism in SAR image detection.

In this paper, the ERSA module is proposed to filter the features so that the network pays more attention to the channel features and the spatial regions containing the target information. It is an ultra-lightweight dual attention mechanism module that is inspired by the ideas of residual [28] and shuffle attention [29]. In order to lighten the weight as much as possible, the $F_c(\cdot)$ function and sigmoid function are used to form gating mechanism in channel attention and spatial attention branches to adaptively learn the correlation of the channel features and the importance of the spatial features.

The channel attention module is defined as follows:

$$F_{GAP} = \frac{1}{H \times W} \sum_{i,j=1}^{W,H} X_{i,j} \tag{1}$$

$$E_1 = F_{GAP}(X_1) \tag{2}$$

$$F_c(E_1) = W_1 \cdot E_1 + b_1 \tag{3}$$

$$X_{11} = \delta(F_c(E_1)) \otimes X_1 \tag{4}$$

where $X_1 \; \epsilon \; R^{C \times H \times W}$ is the input feature map, F_{GAP} is the global average pooling, and $E_1 \; \epsilon \; R^{C \times 1 \times 1}$ is the channel vector with global receptive field. W_1 and b_1 are a pair of learnable parameters, which are used to scale and translate the channel vector respectively, and learn different importance of channel dimensions. δ is the sigmoid function for normalization. Finally, the normalized channel attention weight vector $\delta(F_c(E_1))\epsilon R^{C \times 1 \times 1}$ is multiplied by the input feature map $X_1 \epsilon R^{C \times H \times W}$ to realize the feature recalibration.

Similarly, in spatial attention, the group norm (GN) performed on the feature maps X_2 to capture the spatial information, and then the similarity gating mechanism is used to adaptively adjust the input feature map X_2 to highlight the essential spatial information.

The overall ERSA module structure is shown in Figure 3. In order to improve the efficiency of the network, the input feature maps are divided into 32 groups of independent sub-features along the channel dimension for parallel processing. In each group of sub-features, the input feature maps are divided into two parts according to the channel dimension. They are input into the channel attention branch and the spatial attention branch respectively to learn the important features of the target. Then the two branches are fused to obtain the sub-features after attention enhancement. All the sub-features of groups after attention enhancement are aggregated, and then channel shuffle operator is used to enhance the information flow between feature channels to get the fine features after attention enhancement. Finally, with the addition of skip connection, the coarse-grained features of the initial input feature map are retained effectively and the training process becomes more robust.

2.4. Classification and Box Prediction Network

After extracting the features of the aircraft, the BPA2M module outputs effective prediction feature maps at three scales, in which the grid area is divided into 64 × 64, 32 × 32 and 16 × 16 separately. The schematic diagram of classification and box prediction is shown in Figure 4. For each grid cell, the classification and box prediction network generates three anchor boxes with different sizes and aspect ratios (represented by orange box in Figure 4). The size and shape of anchor boxes are obtained by clustering algorithm based on the size of the target box in the dataset. Then, a 1 × 1 convolution is directly used to predict the location, confidence and category of each bounding box through classification regression.

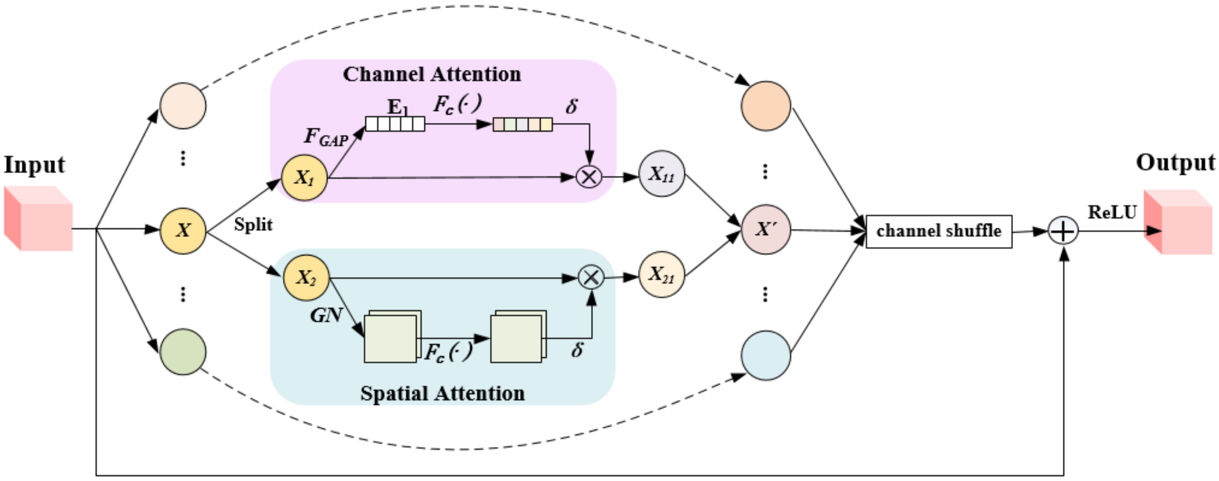

Figure 3. The structure of the Effective Residual Shuffle Attention (ERSA) module.

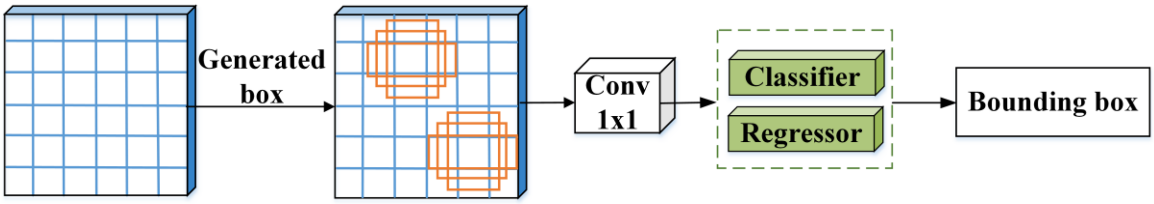

Figure 4. The schematic diagram of classification and box prediction, the anchor box is represented by orange.

After obtaining the bounding box, the NMS is used to remove the overlap box, and the final output detection result is obtained, as show in Figure 1. In the training stage, the loss of the network is calculated. The network is optimized by adjusting network parameters to minimize network loss.

2.5. Detection by Sliding

In this paper, the detection method by sliding window is proposed to perform aircraft detection from large-scale SAR images, which can improve the efficiency of the network [30], as shown in Figure 5. First, the input large-scale SAR image is clipped by sliding window with the window size of 512 × 512 and the step size of 450. In this way, the size of the aircraft is relatively larger in the test samples which is conducive for aircraft detection [31]. Then, the test samples are fed into EBPA2N to detect the aircraft. After the detection results are obtained, coordinate mapping and NMS are used as post-processing modules to achieve coordinate aggregation. The final detection results of aircraft from large-scale SAR images are then produced.

3. Experimental Results and Analyzer

3.1. Data and Evaluation Metrics

The data used in this experiment were more than 10 large-scale SAR images containing airports from the Gaofen-3 system with 1 m resolution. For the insufficiency of manually annotated aircraft data, we used rotation, translation (data augmentation in width and height directions), flipping and mirror to expand the data. Finally, 4396 samples of aircraft data with the size of 512 × 512 pixels were obtained, and the training set and verification set were divided by a ratio of 8:2. In order to evaluate the performance of the network more objectively and efficiently, four evaluation indexes were used in this paper: detection rate (*DR*) [32], false alarm rate (*FAR*) [32], training time and testing time. The calculation formulas of *DR* and *FAR* are as follows:

$$DR = \frac{N_{DT}}{N_{GT}} \tag{5}$$

$$FAR = \frac{N_{DF}}{N_{DT} + N_{DF}} \tag{6}$$

where N_{DT}, N_{DF} and N_{GT} are the number of correctly detected aircraft, falsely detected aircraft, and the real aircraft targets. We use *N* to represent numbers, and the subscripts *DT* stands for Detected True, *DF* for Detected False, and *GT* means Ground Truth, respectively.

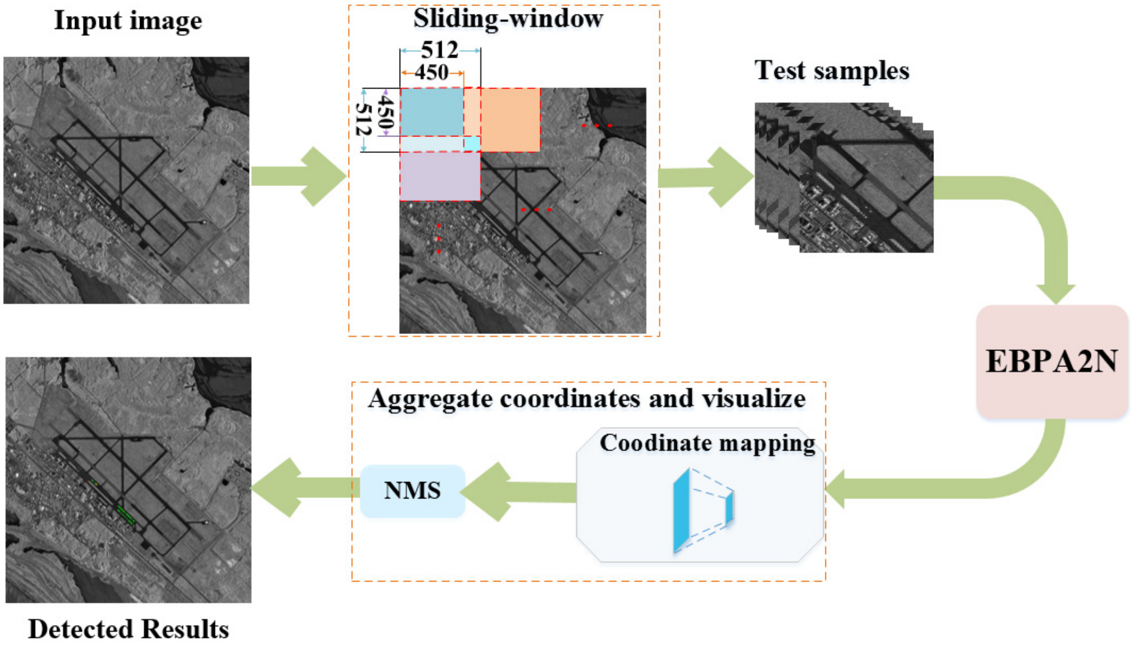

Figure 5. The architecture of sliding window detection.

3.2. Implementation Details

The experimental environment was based on a single 12G memory NVIDIA RTX 2080ti GPU and Unbuntu18.04 system. All networks were based on Pytorch framework, training with 100 epochs by the same data set and recording the training time. The batch size was 16. The learning rates for EfficientDet-D0 [33], YOLOv5s [12] and our network (EBPA2N) were 3e-4, 1e-3 and 1e-3, respectively. In particular, none of these networks was loaded with pretraining models. In addition, multi-scale training and additional enhancement testing techniques (e.g., Test Time Augmentation, TTA) were not used.

3.3. Analysis of the Experimental Results

In order to evaluate the performance of the framework proposed in this paper, three large-scale SAR images from the Gaofen-3 system with 1m resolution are used as independent tests, which are airport I (Hongqiao Airport) with 12,000 × 14,400 pixels, airport II (Capital Airport) with 14,400 × 16,800 pixels and airport III (Military Airport) with 9600 × 9600 pixels.

3.3.1. Analysis of Aircraft Detection for Airport I

Figure 6 shows the detection results of aircraft by different networks in Airport I. correct detection, false alarms and missed detection are indicated by green, red and yellow boxes respectively. This airport is a large civil airport (Hongqiao airport in Shanghai, China) with heavy transportation. As shown in Figure 6a, the distribution of aircraft is relatively dense, and the distance between adjacent aircraft is small. The backscattering characteristics of the aircraft at the bend of the airport are complex and diverse, which increases the difficulty of the detection. According to Figure 6b, EfficientDet-D0 has many false alarms (red color boxes) outside the airport. However, there are no obvious false alarms in YOLOv5s (Figure 6c) and EBPA2N (as shown in Figure 6d). This shows that the robustness of YOLOv5s and EBPA2N is much better. In the enlarged detail of the bend area of the airport, we can find that EfficientDet-D0 is better than YOLOv5s (as shown in Figure 6c) in missed detection, but it has much more false alarms. However, in the detection results of YOLOv5s, many aircraft are missed. Because IEPA and ERSA modules in EBPA2N can effectively detect aircraft and eliminate false alarms, the detection performance is much better than that of YOLOv5s and EfficientDet-D0. Furthermore, it shows that EBPA2N has good ability of feature learning and can fit the multi-scale and multi-directional features of aircraft well.

3.3.2. Analysis of Aircraft Detection for Airport II

Airport II is another large-scale civil airport (the capital airport in Beijing, China). As shown in Figure 7a, the background area covers large commercial and residential areas. There are also lots of strong scattering highlights, which are likely to cause false alarms. In addition, the mechanical and metal equipment around the aircraft appears to have similar texture as the aircraft, which is also very prone to generate false alarms. The detection results of aircraft for Airport II by each network are shown in Figure 7b–d. According to the results, there is a serious problem of false alarms in EfficientDet-D0 (as shown in Figure 7b), especially at the edge of the airport. In contrast, the detection results of YOLOv5s and EBPA2N are more satisfactory (as shown in Figure 7c,d). According to the magnified view of the local details, the EfficientDet-D0 has only two missed aircraft, but it has five false alarms. YOLOv5s has three missed detection and one false alarm. The result of EBPA2N (as shown in Figure 7d) is the closest to that of the ground truth (as shown in Figure 7a), in which, only one aircraft with weak brightness information has been missed. Remarkably, this aircraft is missed in both networks. This also demonstrates the advantage of the parallel ERSA module used in our network, which can focus more on learning the characteristics of the aircraft, thus the detection rate of the aircraft is improved.

3.3.3. Analysis of Aircraft Detection for Airport III

Airport III is a military airport with 33 aircraft, as shown in Figure 8a, it can be seen that the overall background area is relatively clean. Only a few scattered buildings with strong scattering points in the left area of the airport are prone to interference. At the same time, although the aircraft at the airport are smaller than civil aircraft, the scattering characteristics are obvious, so the detection of the three networks have high detection integrity. Figure 8a–d show the ground truth and the detection results by different networks. It can be seen that only one aircraft has been missed by EfficientDet-D0, which is marked with a yellow box in the Figure 8b, but there are 10 false alarms. This shows that the EfficientDet-D0 has a weak ability to distinguish the effective features of the aircraft. There are only two false alarms in YOLOv5s (as shown in Figure 8c), one of which is distributed in the airport runway and the other is outside the airport area. However, it missed three aircraft. In the proposed EBPA2N network, there is only one missed detection and one false alarm located at the edge of the equipment inside the airport (as shown in Figure 8d). Compared with the original YOLOv5s, the number of missed detection by EBPA2N is reduced, which indicates that the addition of IEPA module enhances the ability of the network to capture scattering characteristics of aircraft.

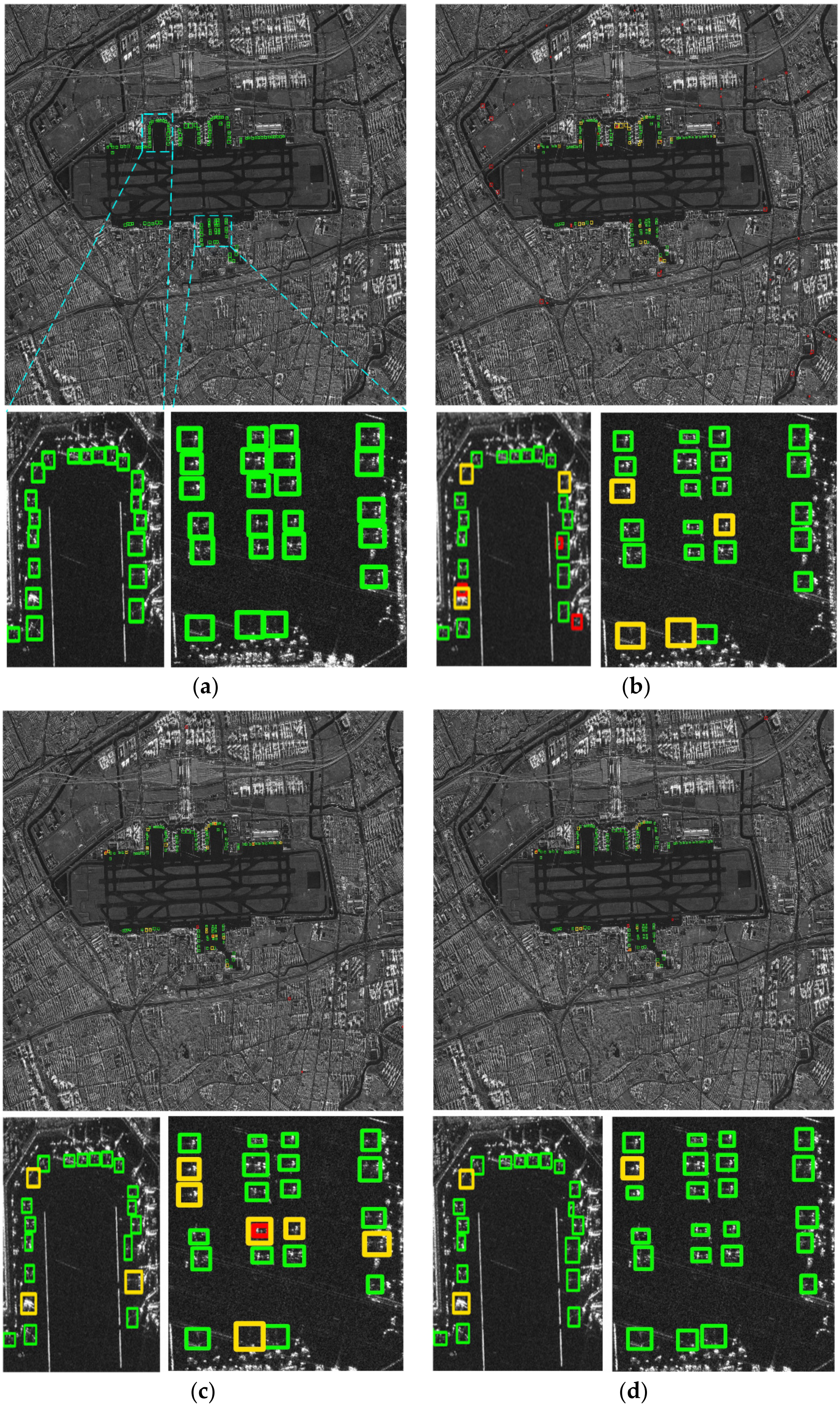

Figure 6. The experiment result for Airport I. (**a**) the ground truth of Airport I from Gaofen-3. (**b**–**d**) are the detection results of aircraft by EfficientDet-D0, YOLOv5s, and EBPA2N respectively. The corrected detection, false alarms and missed detection are indicated by green, red and yellow boxes respectively.

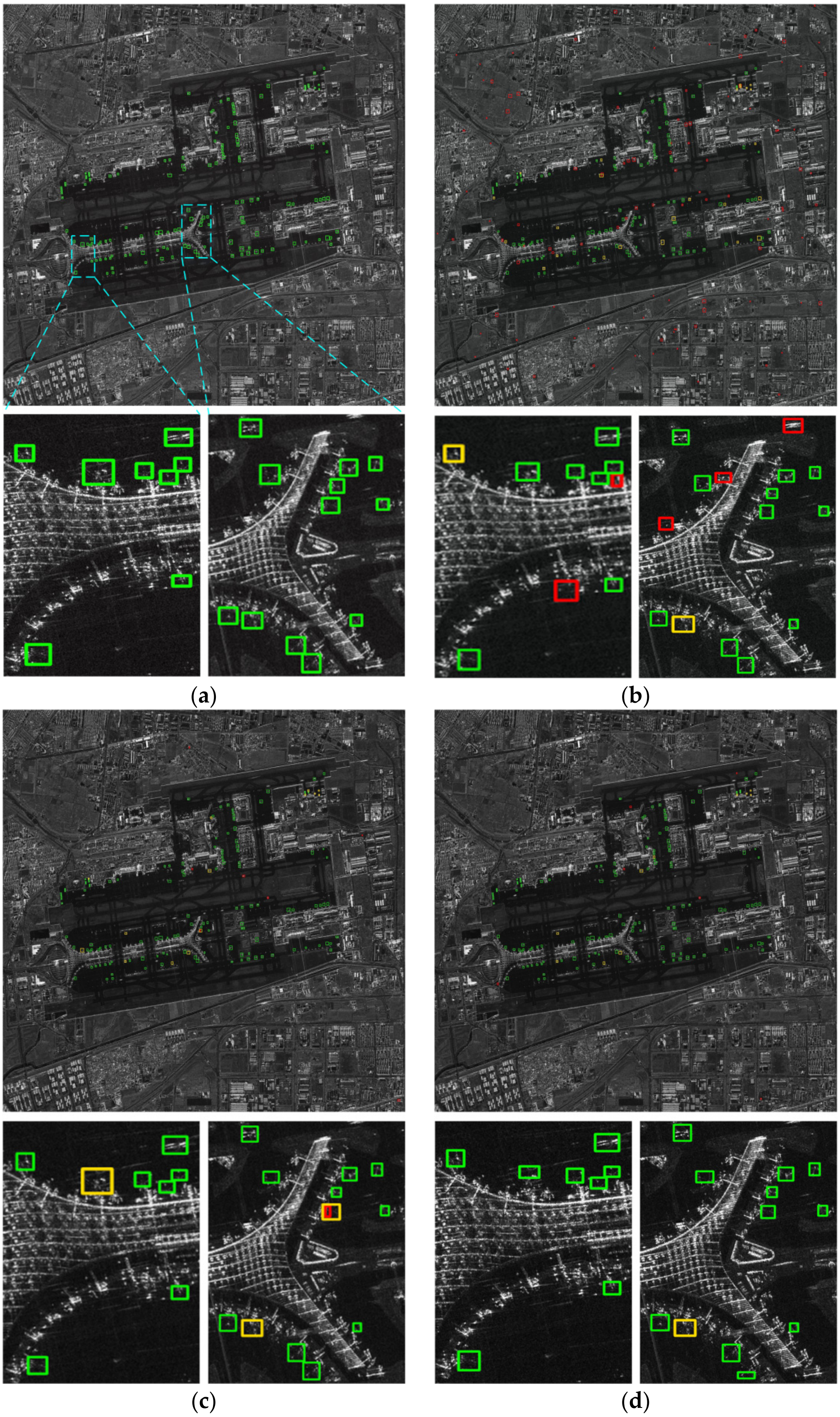

Figure 7. The detection result for Airport II. (**a**) The ground truth of Airport II from Gaofen-3 system. (**b**–**d**) are the detection results of aircraft by EfficientDet-D0, YOLOv5s, and EBPA2N, respectively. The corrected detection, false alarms and missed detection are indicated by green, red and yellow boxes, respectively.

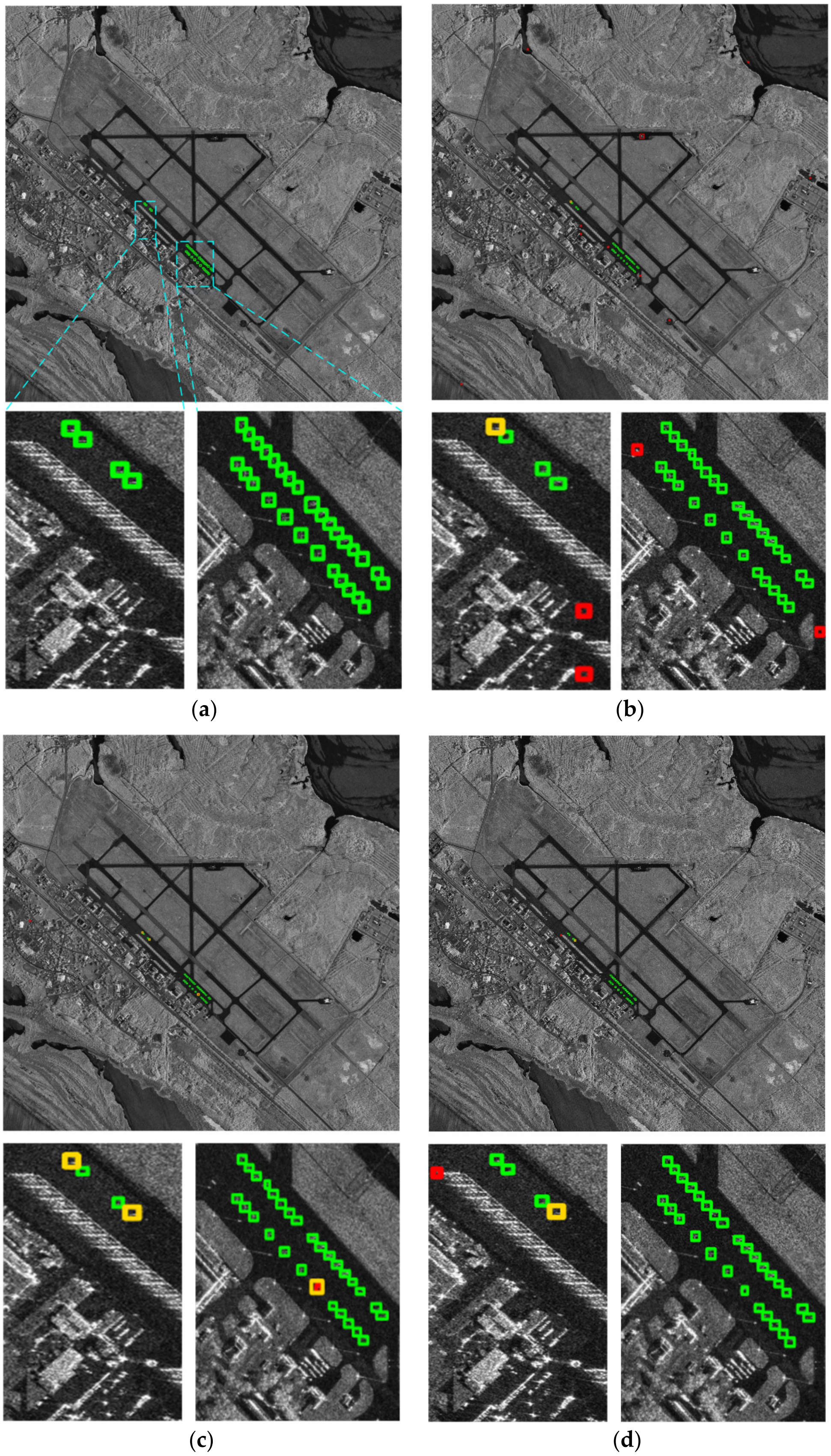

Figure 8. The detection result for Airport III. (**a**) The ground truth of Airport III from Gaofen-3 system. (**b**–**d**) are the detection results of aircraft by EfficientDet-D0, YOLOv5s, and EBPA2N respectively. The corrected detection, false alarms and missed detection are indicated by green, red and yellow boxes respectively.

3.4. Performance Evaluation

For a more intuitive performance comparison, Tables 1 and 2 show the evaluation indexes of the two airports for different networks. For Detection Rate (DR) and False Alarm Rate (FAR), EfficientDet-D0 has the worst performance. The average DR is 85.90%. The average FAR is 34.99%, which indicates that the robustness of the network for aircraft detection is poor. YOLOv5s has more balanced detection performance than EfficientDet-D0. The average FAR is 6.63%, and the average DR is 87.32%. For the proposed EBPA2N, DR is at least 5% higher than YOLOv5s and EfficientDet-D0, and FAR is only 4.49%. So, EBPA2N has much more reliable detection performance. For efficiency, combined with Tables 1 and 2 and Figure 9, EBPA2N is close to YOLOv5s in both training and testing times, which is far better than EfficientDet-D0. The above results show that the proposed network (EBPA2N) has the best detection performance and satisfying computation speed.

Table 1. Comparison of test performance and test time between different networks.

Network	Airport	Detection Rate (%)	False Alarm Rate (%)	Testing Time (s)
EfficientDet-D0	Airport I	77.50	34.51	18.05
	Airport II	83.22	46.64	28.03
	Airport III	96.97	23.81	5.98
	Mean	85.90	34.99	17.03
YOLOv5s	Airport I	80.83	8.49	8.24
	Airport II	90.21	5.15	12.11
	Airport III	90.91	6.25	4.80
	Mean	87.32	6.63	8.38
EBPA2N(Ours)	Airport I	89.17	6.14	9.68
	Airport II	93.01	4.32	13.50
	Airport III	96.97	3.03	5.01
	Mean	93.05	4.49	9.40

Table 2. Comparison of training times between different networks.

Network	Training Time (h)
EfficientDet-D0	5.10
YOLOv5s	0.69
EBPA2N(Ours)	0.882

4. Discussion

At present, most of the mainstream target detection networks are designed for optical images, and the research on SAR image is relatively immature. This paper proposes an effective aircraft detection network for large-scale SAR images, which greatly improves the accuracy of aircraft detection and provides fast detection of aircraft.

In the design and development of the detection neural network, the main consideration lies in the trade-off between the detection accuracy and speed, because complex deep neural networks that achieve high accuracy usually suffer from huge computation intensity. The YOLOv5s network is light in computation intensity, so it can be deployed at front-end devices such as mobile terminals. This is the major reason why we combine YOLOv5s with SAR image analysis as the implementation of our deep neural network. The IEPA module is designed to address the correlation between discrete features of aircraft over a larger scope, and then the lightweight ERSA attention module is proposed to adaptively choose important features of aircraft, which systematically integrates attention mechanism and residual structure. According to Tables 1 and 2, the overall FAR of EfficientDet-D0 network is 34.99%, and training and testing time of the network is the longest. This means that EfficientDet-D0 has poor ability in suppressing the interference of a complex background for the detection of small man-made targets in SAR images. The computation efficiency of YOLOv5s is very satisfying, but the detection accuracy needs to be improved. The proposed EBPA2N has achieved a good balance of accuracy and speed in aircraft detection.

Moreover, we have only conducted experiments on single-band SAR images with 1 m resolution from the Gaofen-3 system. For SAR images acquired from different systems (e.g., different band or different resolution), the performance may vary. In the future, we will employ transfer learning [34] to EBPA2N to realize rapid aircraft detection with multi-resolution and multi-band SAR images. Furthermore, the network proposed in this paper is only tested for the aircraft detection, with satisfying detection accuracy. In the following research, the network will be extended for further experimental analysis of other man-made small targets (e.g., ships, buildings and vehicles). In addition, the motion error in SAR image [35] is also worth considering, because it will indeed bring additional impact on SAR target detection.

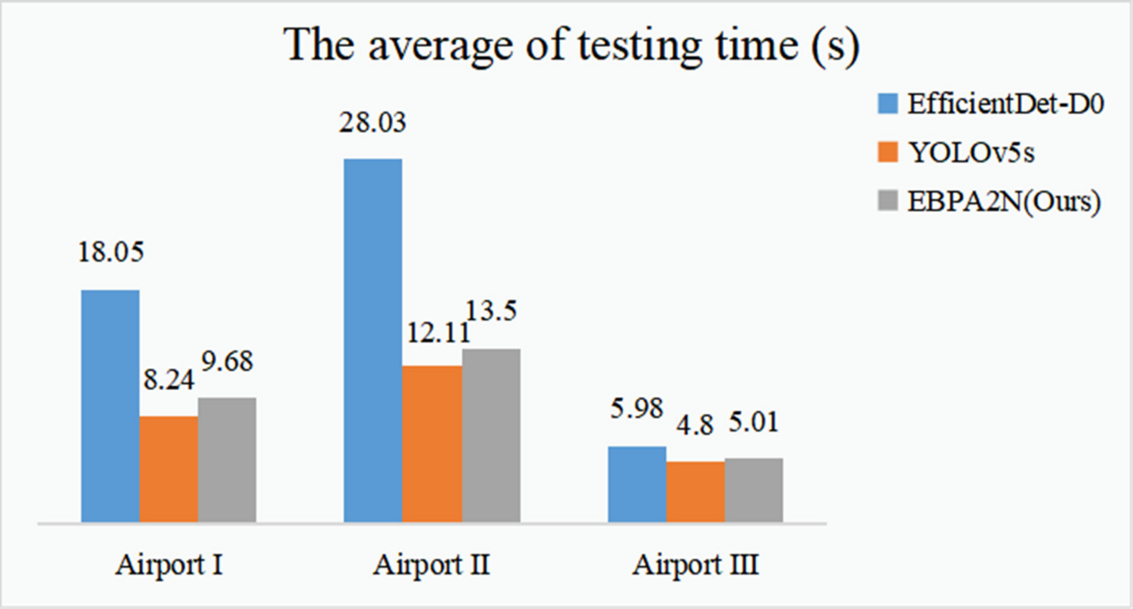

Figure 9. Visualized comparison of average testing times for different networks at three airports.

5. Conclusions

In this paper, a high-precision and efficient automatic aircraft detection network for SAR images is proposed, namely, EBPA2N. The performance improvement of EBPA2N mainly results from two innovative modules proposed in this paper. The IEPA module and ERSA module work in series, which effectively integrates the multi-scale context information of aircraft and greatly improves detection accuracy. The experimental results on three different types of airport from Gaofen-3 SAR images indicate that EBPA2N has strong and robust capabilities in extracting multi-scale and multi-direction aircraft image features, and it also can suppress the interference of the complex background very well.

By combining deep neural networks with geospatial analytics, BEPA2N can also be applied to detect other small man-made targets, such as ships, buildings and vehicles. We plan to purse these topics in our future study. BEPA2N will promote closer fusion of deep learning techniques and SAR image analytics, so as to speed up the research of intelligent target detection of SAR imagery.

Author Contributions: Methodology, R.L. and L.C.; software, R.L. and J.W.; validation, S.T., R.L. and X.C.; formal analysis, R.L.; investigation, J.X., L.C., X.C. and S.T.; data curation, J.W., Z.Y. and S.T.; writing—original draft preparation, R.L.; writing—review and editing, L.C. and J.X.; visualization, R.L.; supervision, Z.Y.; project administration, X.C., Z.Y. and J.W.; funding acquisition, L.C. and Z.Y.

All authors contributed extensively to this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by National Natural Science Foundation of China, grant number 42101468 and 61701047, partly funded by the Foundation of Hunan, Education Committee, under Grants No.20B038.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: We sincerely thank the anonymous reviewers for their critical comments and suggestions for improving the manuscript.

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

References

1. Sportouche, H.; Tupin, F.; Denise, L. Extraction and three-dimensional reconstruction of isolated buildings in urban scenes from high-resolution optical and SAR spaceborne images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3932–3945. [CrossRef]
2. Pan, B.; Tai, J.; Zheng, Q.; Zhao, S. Cascade convolutional neural network based on transfer-learning for aircraft detection on high-resolution remote sensing images. *J. Sens.* **2017**, *2017*, 1796728. [CrossRef]
3. Guo, Q.; Wang, H. Research progress on aircraft detection and recognition in SAR imagery. *J. Radars* **2020**, *9*, 497–513.
4. Zhang, T.; Zhang, X. High-Speed ship detection in SAR images based on a grid convolutional neural network. *Remote Sens.* **2019**, *11*, 1206. [CrossRef]
5. 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.
6. Ren, S.; Kaiming, H.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anl. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
7. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.
8. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767v1.
9. Wang, C.-Y.; Liao, H.; Yeh, I.-H.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 13–19 June 2020; pp. 1571–1580.
10. Bochkovskiy, A.; Wang, C.-Y. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934v1.
11. Wang, C.-Y.; Bochkovskiy, A.; Liao, H. Scaled-YOLOv4: Scaling cross stage partial network. *arXiv* **2020**, arXiv:2011.08036v2.
12. Ultralytics. yolov5. Available online: https://github.com/ultralytics/yolov5 (accessed on 18 May 2020).
13. Zhang, L.; Li, C.; Zhao, L.; Xiong, B.; Kuang, G. A cascaded three-look network for aircraft detection in SAR images. *Remote Sens. Lett.* **2019**, *11*, 57–65. [CrossRef]
14. Wang, J.; Xiao, H.; Chen, L.; Xing, J.; Pan, Z.; Luo, R.; Cai, X. Integrating weighted feature fusion and the spatial attention module with convolutional neural networks for automatic aircraft detection from SAR images. *Remote Sens.* **2021**, *13*, 910. [CrossRef]
15. 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.
16. 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.
17. Du, L.; Li, L.; Wei, D.; Mao, J. Saliency-guided single shot multibox detector for target detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3366–3376. [CrossRef]
18. 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]
19. Zhang, P.; Chen, L.; Li, Z.; Xing, J.; Xing, X.; Yuan, Z. Automatic extraction of water and shadow from SAR images based on a multi-resolution dense encoder and decoder network. *Sensors* **2019**, *19*, 3576. [CrossRef] [PubMed]
20. Zhao, Y.; Zhao, L.; Li, C.; Kuang, G. Pyramid attention dilated network for aircraft detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 662–666. [CrossRef]
21. Guo, Q.; Wang, H.; Xu, F. Scattering Enhanced attention pyramid network for aircraft detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *99*, 1–18. [CrossRef]
22. Neubeck, A.; Gool, L. Efficient non-maximum suppression. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; pp. 850–855.
23. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *Proc. IEEE Trans. Pattern Anal. Mach. Intelli.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]

24. Yan, B.; Fan, P.; Lei, X.; Yang, F. A real-time apple targets detection method for picking robot based on improved YOLOv5. *Remote Sens.* **2021**, *13*, 1619. [CrossRef]
25. Liu, Y.; Lu, B.; Peng, J.; Zhang, Z. Research on the use of YOLOv5 object detection algorithm in mask wearing recognition. *World Sci. Res. J.* **2020**, *6*, 276–284.
26. Li, D.; Hu, J.; Wang, C. Involution: Inverting the inherence of convolution for visual recognition. *arXiv* **2021**, arXiv:2103.06255.
27. 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]
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
29. Zhang, Q.-L.; Yang, Y.-B. SA-Net: Shuffle attention for deep convolutional neural networks. *arXiv* **2021**, arXiv:2102.00240v1.
30. Xing, J.; Sieber, R.; Kalacska, M. The challenges of image segmentation in big remotely sensed imagery data. *Ann. GIS* **2014**, *20*, 233–244. [CrossRef]
31. Li, S.; Gu, X.; Xu, X.; Xu, D.; Zhang, T.; Liu, Z.; Dong, Q. Detection of concealed cracks from ground penetrating radar images based on deep learning algorithm. *Constr. Build. Mater.* **2021**, *273*, 121949. [CrossRef]
32. Deng, Z.; Sun, H.; Zhou, S.; Zhao, J.; Lei, L.; Zou, H. Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 3–22. [CrossRef]
33. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and efficient object detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10778–10787.
34. Lu, J.; Vahid, B.; Hao, P.; Zuo, H.; Xue, S.; Zhang, G. Transfer learning using computational intelligence: A survey. *Knowl. Based Syst.* **2015**, *80*, 14–23. [CrossRef]
35. Wei, P. Deep SAR imaging and motion compensation. *IEEE Trans. Image Process.* **2021**, *30*, 2232–2247.