remote sensing

MDPI

Article

Aircraft Detection in High Spatial Resolution Remote Sensing Images Combining Multi-Angle Features Driven and Majority Voting CNN

Fengcheng Ji [1], Dongping Ming [1,2,*], Beichen Zeng [1], Jiawei Yu [1], Yuanzhao Qing [1], Tongyao Du [1] and Xinyi Zhang [1]

[1] School of Information Engineering, China University of Geosciences (Beijing), 29 Xueyuan Road, Beijing 100083, China; jifc@cugb.edu.cn (F.J.); 2004190028@cugb.edu.cn (B.Z.); 2004190032@cugb.edu.cn (J.Y.); tsingyz@cugb.edu.cn (Y.Q.); 2004200038@cugb.edu.cn (T.D.); 2004200043@cugb.edu.cn (X.Z.)

[2] Polytechnic Center for Natural Resources Big-Data, Ministry of Natural Resources of China, Beijing 100036, China

* Correspondence: mingdp@cugb.edu.cn; Tel.: +86-10-13520907831

Abstract: Aircraft is a means of transportation and weaponry, which is crucial for civil and military fields to detect from remote sensing images. However, detecting aircraft effectively is still a problem due to the diversity of the pose, size, and position of the aircraft and the variety of objects in the image. At present, the target detection methods based on convolutional neural networks (CNNs) lack the sufficient extraction of remote sensing image information and the post-processing of detection results, which results in a high missed detection rate and false alarm rate when facing complex and dense targets. Aiming at the above questions, we proposed a target detection model based on Faster R-CNN, which combines multi-angle features driven and majority voting strategy. Specifically, we designed a multi-angle transformation module to transform the input image to realize the multi-angle feature extraction of the targets in the image. In addition, we added a majority voting mechanism at the end of the model to deal with the results of the multi-angle feature extraction. The average precision (AP) of this method reaches 94.82% and 95.25% on the public and private datasets, respectively, which are 6.81% and 8.98% higher than that of the Faster R-CNN. The experimental results show that the method can detect aircraft effectively, obtaining better performance than mature target detection networks.

Keywords: aircraft detection; remote sensing image; multi-angle; majority voting; convolutional neural network

Citation: Ji, F.; Ming, D.; Zeng, B.; Yu, J.; Qing, Y.; Du, T.; Zhang, X. Aircraft Detection in High Spatial Resolution Remote Sensing Images Combining Multi-Angle Features Driven and Majority Voting CNN. *Remote Sens.* **2021**, *13*, 2207. https://doi.org/10.3390/rs13112207

Academic Editors: Edoardo Pasolli, Zhou Zhang, Zhengxia Zou and Zhiyong Lv

Received: 23 April 2021
Accepted: 3 June 2021
Published: 4 June 2021

1. Introduction

Remote sensing image is a crucial and indispensable resource widely used in civil and military fields [1,2]. As one of the most important tasks and research hotspots of remote sensing image interpretation, object detection has attracted the attention of academia and industry with the higher spatial resolution of remote sensing images and the richer information contained in the images. Aircraft, as an important means of transportation and weaponry, is one of the most important targets in the field of object detection. The accurate detection of aircraft has crucial practical significance and military value [3]. Therefore, aircraft detection from remote sensing images has become the focus of attention.

With the increasing capacity of computer data processing, deep learning methods based on convolutional neural networks have made remarkable achievements in speech recognition, computer vision, autonomous driving, and other fields [4–7]. Compared with the traditional methods, the deep learning methods based on convolutional neural network can extract features with richer semantic information, higher level, stronger robustness and generalization ability from samples in a data-driven way. In recent years, the convolutional neural networks with excellent feature expression abilities have been widely used in

image classification [8,9], object detection [10,11], and semantic segmentation [12–14]. Also, the object detection networks based on deep convolutional neural network have good performance in natural image detection.

In view of the success of convolutional neural networks in natural image detection, some researchers tried to apply convolutional neural network to aircraft detection in remote sensing images. In order to improve the detection accuracy of aircrafts in remote sensing images, Hu et al. [15] used the saliency detection algorithm to reduce the number of proposal boxes, and obtained the target position information by using the saliency algorithm based on the background priori, and finally, a deep convolutional neural network was used to determine the category and fine-tune bounding boxes of the objects. Shi et al. [16] proposed a model of aircraft detection called DPANet based on deconvolution and position attention to extract the external and internal structure features of the aircraft. Besides, aiming at the problem of multi-angle distribution of aircrafts in remote sensing images, the rotation invariant detection network based on convolutional neural networks was proposed and widely used in the detection of aircraft targets in remote sensing images [17,18]. Although the accuracy of the detection can be improved to a certain extent by applying convolutional neural networks to the detection of aircrafts in remote sensing images, there are still two challenges in the target detection of remote sensing images. Firstly, unlike natural images, remote sensing images have more complex objects and special imaging angle and capture mostly the top information of objects, resulting in the similarity (e.g., spectrum similarity or geometry similarity) of the targets and other objects. In addition, the imaging results of remote sensing images are susceptible to interference from the atmosphere, electromagnetic waves and so on, which makes it difficult for object detection in remote sensing images [19,20]. Secondly, the network cannot avoid the loss of information in the process of feature extraction, so there is often a high rate of missed detection and false alarm when detecting complex and dense small objects and weak objects [21]. In general, all the above reasons will cause difficulties in extracting information from remote sensing images, resulting in a decrease in the accuracy of target detection. At present, mature object detection networks, such as R-CNN, SSD, YOLO, etc., have greatly improved the accuracy of detection, but in the face of the above difficulties, they are still unable to avoid the problems of missed detection rate and false alarm rate [22–24].

To solve the above problems, we proposed a simple, effective, and more universal multi-angle features driven method. Through adding the multi-angle transformation module, the features of object from multiple angles can be extracted to reduce the missed detection rate of the model. In addition, aiming at the common problem of false detections in the existing object detection networks, we proposed a box detection post-processing method based on majority voting strategy. Through post-processing of the detection results, we can further judge whether the detection boxes contain the target objects, thereby reducing the possibility of misjudgment and improving the overall performance of the model. The experimental results showed that the proposed method achieved better performance than the existing object detection networks on both the public dataset and the private dataset.

2. Materials and Methods

2.1. Related Work

2.1.1. One-Stage Target Detection Algorithm Based on Convolution Neural Network

The one-stage object detection algorithm based on convolutional neural network has been paid more attention because of its simple structure, high computational efficiency, and high detection accuracy [25]. It regards the object detection as a regression analysis problem, and uses mature convolutional neural networks, such as VGG [26], Resnet [8], etc., as the backbone to determine the object category and location. The representative one-stage object detection networks include SSD [27], YOLO [11], and improved methods based on them, such as DSSD [28], YOLO-V2 [29], YOLO-V3 [30], and so on. Although the one-stage object detection method has advantages in calculation speed, its accuracy

is usually lower than that of the two-stage object detection method. There will be serious missed detections when facing remote sensing images with a large range. Therefore, few scholars apply it to large-scale remote sensing image target detection directly.

2.1.2. Two-Stage Target Detection Algorithm Based on Convolution Neural Network

Compared with one-stage object detection methods, two-stage methods based on convolutional neural network have higher accuracy and better performance in locating and recognizing targets, so they are widely used in the object detection of remote sensing images. The typical two-stage object detection networks are the R-CNN [31] series, such as R-CNN, Fast R-CNN [10], and Faster R-CNN [32]. Besides, some scholars were inspired by the idea of them, putting forward their own methods to detect objects in remote sensing images. Wu et al. [33] used Edgeboxes algorithm to generate region proposals, and then used convolutional neural networks to perform feature extraction and classification. Similarly, Yang et al. [34] proposed a "region proposal–classification–accurate object localization" framework for detecting objects in remote sensing images. However, all the above methods have the problem of redundancy in candidate regions, in order to solve those problems, Liu et al. [35] proposed an aircraft detection method based on corner clustering and convolutional neural networks, which used the mean-shift clustering algorithm to generate candidate regions for the corner points detected in the binary image, and then utilized CNN to determine the target category in the candidate region. Although the two-stage method can improve the accuracy of object detection, the existing two-stage target detection methods ignore the differences of features extracted after image rotation, resulting in insufficient features extracted, which easily leads to false detections or missed detections. Therefore, how to use the object detection network to extract image information more comprehensively has become one of the future research directions.

2.1.3. Improvement on Mature Target Detection Networks

At present, in addition to directly applying mature target detection networks to remote sensing images, another idea is to improve the performance of the existing mature target detection networks as summarized in Table 1.

Table 1. Related works on the improvement of mature target detection network.

Baseline	Paper	Characteristic	Advantage	Disadvantage
Faster R-CNN	Chen et al. [22]	Add a constraint to sieve low quality positive samples	High precision	Low utilization of image information
	Feng et al. [36]	Optimize the generation method of foreground samples, reduce the ineffective foreground samples		Relatively high time consumption
	Li et al. [37]	Modify the anchor box size, scale and loss function of the network for specific target	High precision and less time-consuming	Only for small-scale images
	Fu et al. [38]	Add a rotation-aware object detector to solve the problem of inconsistent target orientation in remote sensing images	High precision	Complex structure and large amount of calculation
SSD	Bao et al. [39]	Determine the detection boxes and target category through two consecutive regressions	Real-time and high precision	Low utilization of image information
	Guo et al. [23]	The DepthFire module is added, which reduces the amount of calculation and improves processing efficiency	Real-time	Compared with two-stage target detection, the overall accuracy is low
	Qu et al. [40]	Combination of dilated convolution and feature fusion		
	Yin et al. [41]	Design encoding-decoding module to detect small objects	Real-time and high precision	Low utilization of image information

Table 1. *Cont.*

Baseline	Paper	Characteristic	Advantage	Disadvantage
YOLO-V1	Xie et al. [24]	Designed a Locally–Constrained module to improve the detection performance for cluster small targets	High accuracy of small target detection	Only for small objects
YOLO-V3	Pham et al. [42]	Replace the large-scale factors in YOLO-V3 with (very) small-scale factors for small target detection	Real-time and high accuracy of small target detection	
	Zhou et al. [43]	Combine the idea of dense connections, residual connections and group convolution	Lightweight	Universality to be investigated

By improving the mature target detection networks, the existing achievements can be fully utilized. However, although the improvement methods such as fine-tuning for network parameters and structure have improved the performance of target detection, most of them are optimized for the feature extraction process or certain types of targets. In the face of small, weak, and dense targets, there are still missed detections and false detections caused by the low utilization rate of image information and insufficient extraction of target potential feature [21]. Therefore, based on the method of making full use of remote sensing image information and extracting potential target characteristics as much as possible, we proposed a target detection model based on multi-angle features driven and majority voting strategy.

2.2. Methods

The target detection model proposed in this paper consists of two modules: multi-angle features driven and majority voting. Also, the multi-angle features driven module includes three parts: multi-angle transformation, feature pyramid network (FPN) [21] and Faster R-CNN. The part of multi-angle transformation is used for transformation of images. The feature pyramid network is embedded in the backbone of Faster R-CNN to extract multi-scale features of targets. The majority voting strategy is used to process the detection results of multi-angle features driven strategy. In addition, the accuracy of Resnet is higher, and it can effectively solve the problem of optimization and training of networks as the number of layers deepen. Therefore, resnet50 was utilized as the backbone network. We describe these two modules in detail below.

2.2.1. Aircraft Detection Based on Multi-Angle Features Driven Strategy

For a remote sensing image rotated at different angles, it becomes a new image compared with the original one. Therefore, the extracted features are different when the image is input to the network at different angles. The reasons of feature differences caused by rotation are analyzed as follows:

(1) In the process of forward propagation, the neural network will lose information due to various operations, such as convolution, pooling, etc. [8,44]. Moreover, the loss will become more serious as the depth of forward propagation increases. Therefore, the difference of final features will be caused due to the loss of their respective information when the image is input to the network at different angles.

(2) For the convolutional neural networks, the extracted features are closely related to the stride and the convolution kernel of the network. Moreover, for the image input to the network at different angles, although the stride and the convolution kernel of the network are same, the extracted features are different due to the different angle in the convolution direction.

As shown in Figure 1, for a 15×15 image, the stride and padding of the network are set to 2 and 3, respectively, and the convolution kernel size is set to 7. The feature map extracted by the convolution process is different from that extracted after 270° rotation of the image. Therefore, multi-angle features driven strategy can effectively compensate for the difference of the extracted features, to make better use of the information in remote

sensing images and reduce the missed detections and false detections caused by the loss of information.

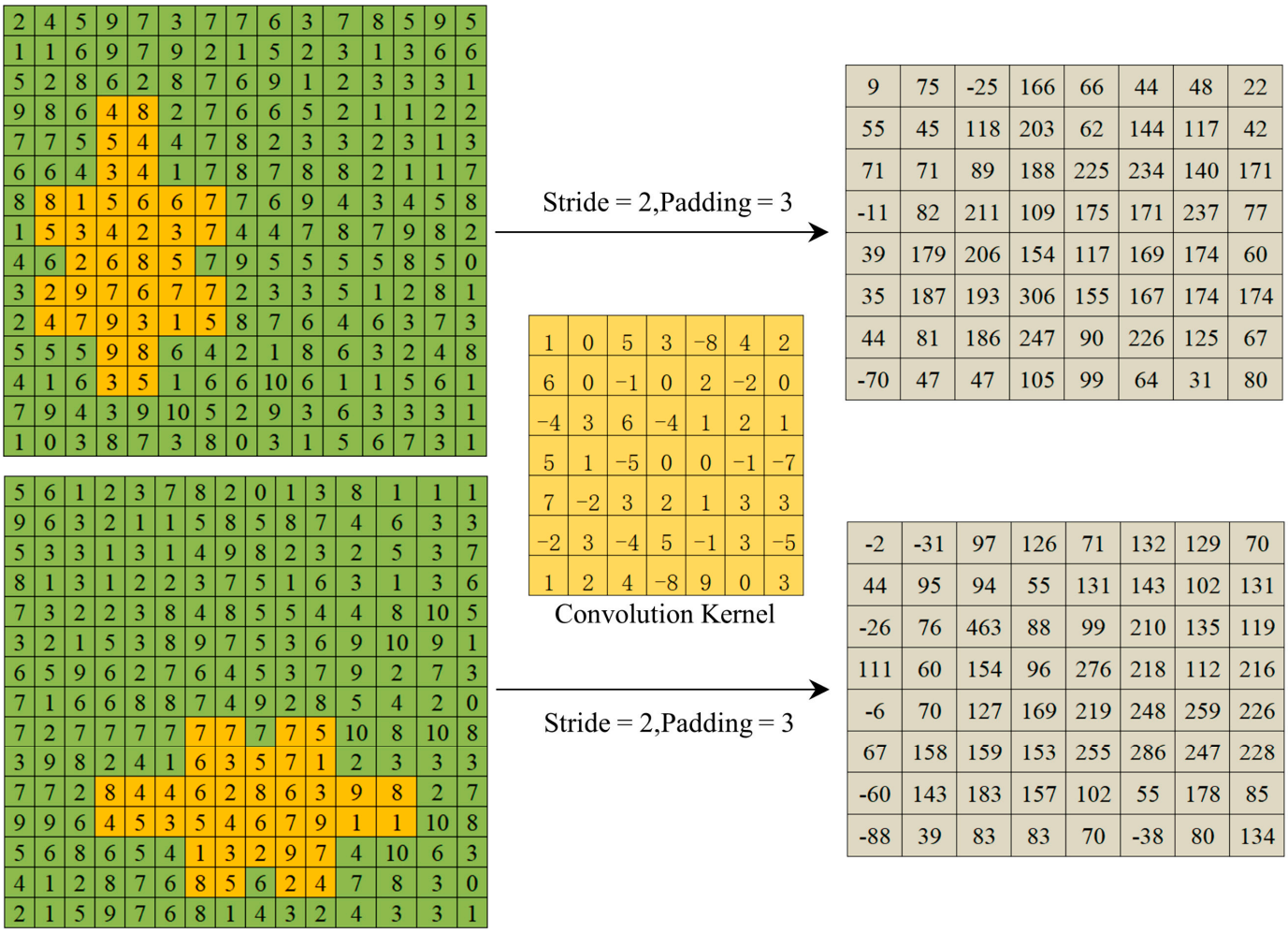


Figure 1. (**Top**) original image and feature map. (**Bottom**) 270° rotation and feature map.

A rotation invariant detection network is usually capable of identifying the targets distributed at different angles in the image by adding rotation invariant layer, using rotation invariance regularization constraint or directly using data augmentation [17,45,46]. The targets of aircrafts in the image are mostly distributed at random angles, which can make the network have rotation invariant to a certain extent by the training of dataset. Therefore, the strategy of multi-angle features driven proposed in this paper did not focus on the rotation invariant but emphasized the use of differences caused by image rotation. It performed multiple feature extractions at different angles for the input image. The purpose is to reduce the loss of target features and improve the performance of the model. Figure 2 shows the process of multi-angle features driven strategy. In order to extract multi-angle features, we designed the multi-angle transformation module. It performs multi-angle (0°, 90°, 180°, 270° and mirror image) transformation on the remote sensing image and inputs it into the trained target detection network for prediction.

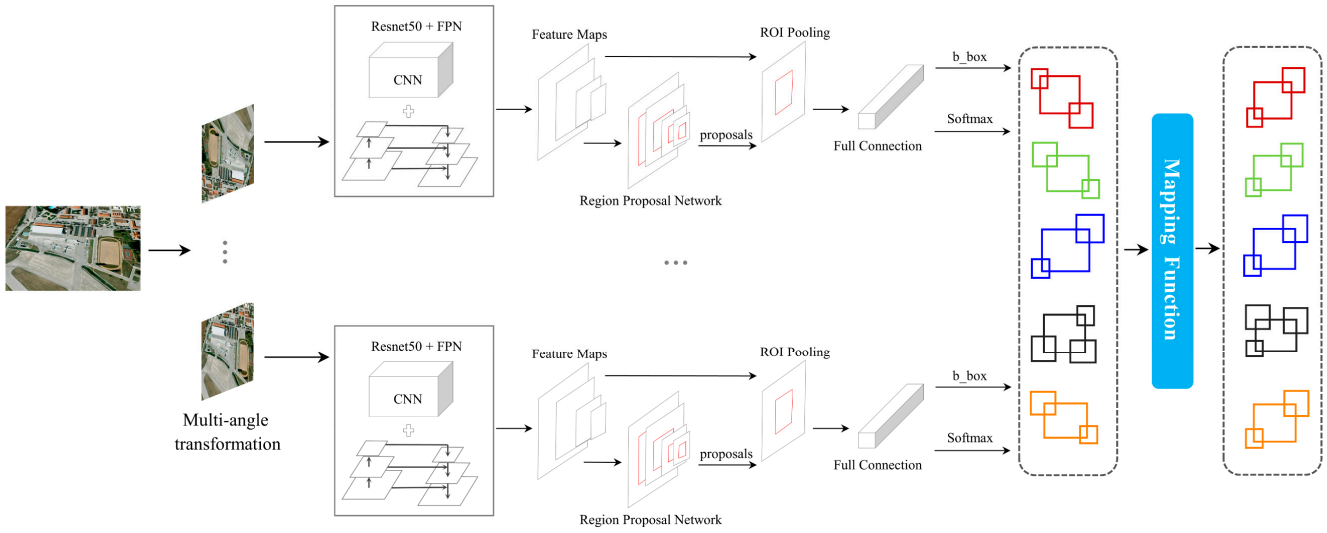


Figure 2. Aircraft detection process of remote sensing images based on multi-angle features driven strategy.

In addition, in order to reduce the loss of small target information caused by the increase of network depth and improve the overall detection performance of the network, we used a feature pyramid network to fuse the multi-scale features of remote sensing image. The combination of multi-angle features driven strategy and feature pyramid network can effectively use the information in the image. For the problem of inconsistency of direction of detection boxes after multi-angle transformation, we used the mapping function to map the detection boxes to the corresponding position in the original image, and then used them for the subsequent majority voting process.

2.2.2. Detection Boxes Processing Based on Majority Voting Strategy

For the existing target detection networks, the final detection result is similar to a "one-shot decision". In other words, the preliminary detection results of the network are the final result, lacking the process of judgement for the detection results. Besides, although the strategy of multi-angle features driven can reduce the loss of information, it also causes the accumulation of wrong information. All the above reasons will lead to the high false alarm rate of the detection results, reducing the final target detection accuracy. Thus, a post-processing method of box detection called majority voting was proposed on the basis of the result of multi-angle features driven strategy. It was achieved by stacking the detection results after multi-angle feature extraction and voting on the stacking results of the detection boxes at each position.

We believe that when the number of votes is advantaged ($n \geq 3$, n is the number of votes), then it can be determined that there is a positive sample at the position of the detection boxes. However, it is also possible that a positive sample exists in the detection box when the number of votes is disadvantaged ($1 \leq n \leq 2$, n is the number of votes) caused by the limitations of the network structure. Therefore, in order to further improve the accuracy of detection and the overall performance of the network, a simple binary classification network was designed in this paper. It was used to judge whether there is positive sample in the detection box with inferior votes. The detailed introduction of the binary classification network will be described in Section 2.2.3.

In the final step, in order to get the final results, the Intersection Over Union (IOU) [47] index was utilized to remove the redundant detection boxes. IOU is usually used to measure the coincidence degree between the detection box and the ground true box, and here it represents the coincidence degree between multiple detection boxes in the same position. The calculation of *IOU* is shown in Formula (1).

$$IOU = \frac{Area \cap}{Area \cup} \tag{1}$$

where *Area*∩ represents the intersection between the detection boxes, and *Area*∪ represents the union between the detection boxes.

For multiple detection boxes in the same position, only one detection box will be saved, and the redundant detection boxes will be deleted when IOU is higher than 0.5. The processing flow is shown in Figure 3.

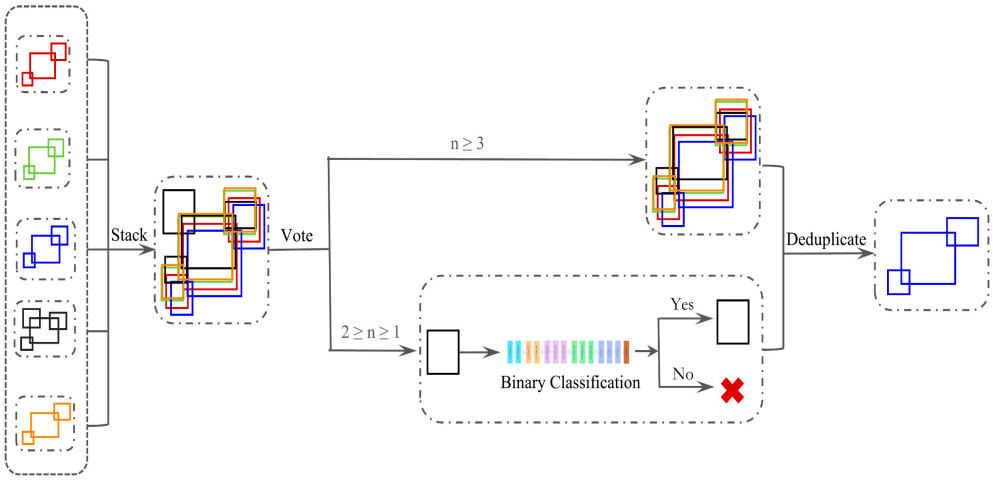


Figure 3. The processing flow of the detection boxes based on majority voting strategy.

2.2.3. The Binary Classification Network

Considering the hardware and time cost, the binary classification network used in this paper adopted the simplified VGG-16 network. The standard VGG-16 network requires the size of the input image to be 224 × 224. However, according to the size of the detection boxes, the size of the input image was adjusted from 224 × 224 to 64 × 64. Meanwhile, in order to make the network more lightweight, we deleted the last three convolution layers and pooling layer in the original model and adjusted the input dimension of the full connection layer from 7 × 7 × 512 to 4 × 4 × 512. By modifying the structure of VGG-16 network, the amount of training parameters of the whole network is greatly reduced, which makes it more suitable for the binary classification task in this study. Finally, for the selection of the loss function and optimizer of the binary classification network, we used the CrossEntropy Loss and Adam optimizer of the original network. The specific structure of binary classification network used in this paper is shown in Figure 4.

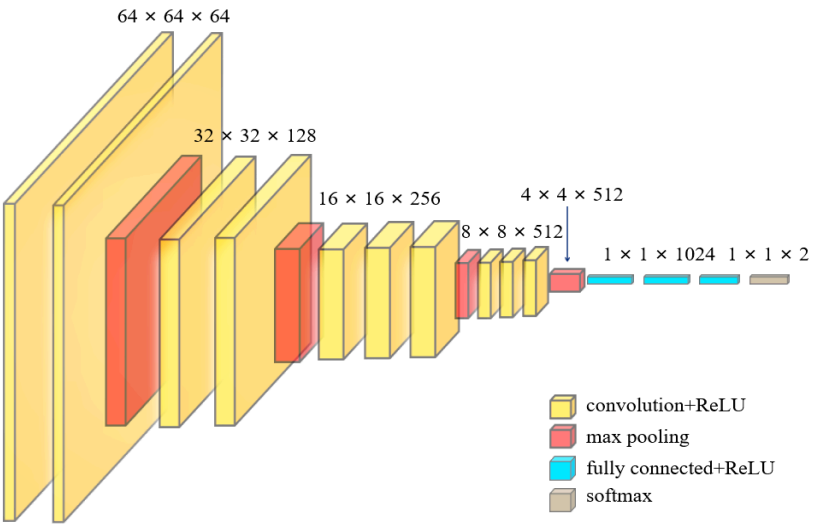


Figure 4. The structure of Binary Classification Network.

2.2.4. Comprehensive Accuracy Evaluation Method

In order to evaluate the effectiveness of the proposed method, the AP value used by Pascal VOC 2012 was introduced into our experiments as the performance evaluation index of object detection, which can effectively evaluate the performance of the network in a certain category. In general, the higher the AP value, the better the performance of the network in a certain category [48]. The AP value can be obtained by calculating the area under the smooth curve formed by the combination of a series of *Recall* and *Precision* values. The calculation formulas of Precision, Recall and *AP* are as follows:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$AP = \int_0^1 P(R)dR \tag{4}$$

where *TP* denotes the number of true positives identified, *FP* denotes the number of false positives identified, *FN* denotes the number of false negatives identified, and *P*(*R*) denotes the *Precision* value corresponding to the *Recall* value.

At the same time, in order to study the time consumption of the model, the Average Time index was employed to evaluate the detection speed of the model. The Average Time refers to how long it would take from the input of the image to the output of the final result, including the time spent on pre-processing, network transmission and post-processing. The *Average Time* is calculated as follows:

$$Average\ Time = \frac{\sum_{k=0}^{n}(ET_K - ST_K)}{n} \tag{5}$$

where n is the number of images in test dataset, ET_k is the output time of the k-th result, and ST_k is the input time of the k-th image.

3. Experiments

All the experiments in this article were performed on the Windows10 system, and PyTorch was employed as a deep learning framework. The hardware configuration consisted of IntelCore i5-9400F CPU, RAM (16 GB) and GPU (Nvidia GeForce RTX 2080Ti 11 GB).

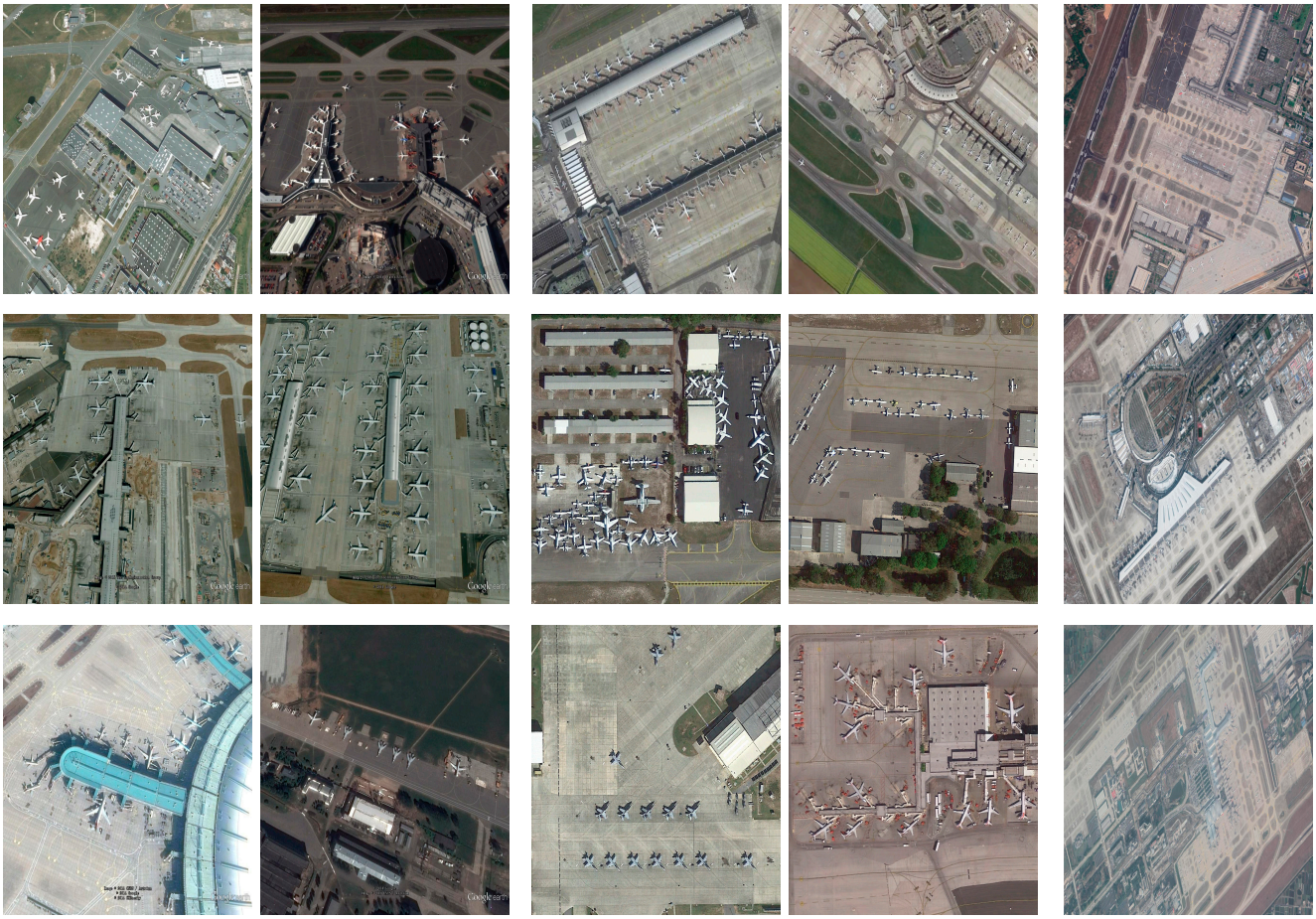
3.1. Datasets Description

3.1.1. Object Detection Datasets

In order to verify the effectiveness of the model, we evaluated the performance of our model on three datasets of remote sensing images, namely RSOD [49], DIOR [50], and private dataset, named I–III, respectively. Some samples of the datasets are shown in Figure 5.

The RSOD (https://github.com/RSIA-LIESMARS-WHU/RSOD-Dataset-, accessed on 23 April 2021) dataset contains four typical target categories of remote sensing image. It has 976 images captured by Google Earth from some airports around the world and obtained by manually marking the locations and attributes. We only used images of aircraft as our training dataset. DIOR (http://www.escience.cn/people/gongcheng/DIOR.html, accessed on 23 April 2021) dataset is a large public dataset proposed by Northwestern Polytechnical University in 2018, which contains 23,463 images and 20 land species categories, including about 1200 images of aircraft, we randomly selected 300 images as our test dataset. The object instances of DIOR dataset have a wide range of spatial resolution, inter-class and intra-class variation, and have high inter-class similarity and intra-class diversity. At the same time, images were obtained under different weather, season, and imaging conditions, so it can be used as a test set to better verify the robustness of the network. In addition, we collected 25 images with different size and spatial resolution

from Google Earth, including a total of 1064 aircrafts. The range of images in this dataset is relatively large and the aircrafts are small and weak targets in the image. Therefore, it can be used to verify the generalization and practicability of the model, the basic information of different image datasets is shown in Table 2:



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

Figure 5. Object detection dataset subsets. (**a**) Dataset I. (**b**) Dataset II. (**c**) Dataset III.

Table 2. Information of image datasets.

Number	Dataset	Resolution/m	Image Size/Pixel	Number of Images	Number of Aircrafts	Purpose
I	RSOD	0.5–2.0	1000 × 900	446	4993	train
II	DIOR [part]	0.5–30	800 × 800	300	3943	test
III	Private Dataset	0.6–1.2	1400 × 900–3000 × 2500	25	1064	test

3.1.2. Binary Classification Network Datasets for "Inferiority Box" Discrimination

The binary classification network dataset adopted in this paper collected a total of 1200 image blocks with different size and resolution arbitrarily intercepted from the RSOD dataset, including 800 positive samples and 400 negative samples. The research shows that data augmentation can make the network fully learn the change of objects and enhance the ability to recognize the complex change of objects. Therefore, we expanded the number of samples in dataset by means of rotation and cropping from 1200 to 4700, including 3000 positive samples and 1700 negative samples. We randomly selected 80% as the training set and 20% as the test set. Partial samples of the dataset are shown in Figure 6. Sample sizes in the dataset range from 30 × 30 pixels to 100 × 100 pixels.

Figure 6. Subsets of binary classification network dataset. (**a**) Positive samples. (**b**) Negative samples.

3.2. Training and Parameters Setting for Network

3.2.1. Training of the Network

In this paper, network training was divided into two parts. The first part is the training of the target detection network. Based on the pre-training weights provided by the official, we embedded the feature pyramid network into the backbone of Faster R-CNN and used the training dataset for transfer learning. The second part is the training of binary classification network. The dataset was trained based on the simplified VGG-16 network, and the binary classification network was saved as a ".pkl" file for parsing and calling by the model.

3.2.2. Parameters Setting

Effective training of the network has an extremely important influence on the performance of network, and some parameters for training need to be clearly stated. In order to train the target detection network, the learning rate controlling the learning progress was 0.005. The number of epoch and batch size used in this paper were 100 and 5, respectively. In addition, as a crucial part of detection boxes post-processing, the parameters setting and the training result of the binary classification network directly affect the final accuracy. Therefore, for training the binary classification network, the learning rate was set as 0.0001. Meanwhile, the epoch and batch_size were 50 and 32, respectively.

3.3. Experimental Results

In order to evaluate the effectiveness of the model proposed in this paper, we demonstrated the detection results on different datasets. We used the trained model to detect the test datasets, including a total of 325 images and 5007 aircrafts. Figure 7 shows part of the detection results on the test datasets using the model proposed in this paper. It can be seen from the figures that the model can detect targets effectively when facing large-scale images or the small and dense targets in the image.

(**a**) (**b**)

Figure 7. Results on different datasets. (**a**) Results on Dataset II. (**b**) Results on Dataset III.

In addition, we also used the comprehensive accuracy evaluation method proposed in Section 2.2.4 to quantitatively evaluate the detection results, and the results are shown in Table 3.

Table 3. Performance of the model on test datasets.

Dataset	Number of Image	Number of Aircraft	AP (%)	Average Time (s)
II	300	3943	94.82	0.49
III	25	1064	95.25	0.63

4. Discussion

4.1. Comparison with the Advanced Models

In Tables 4 and 5 and Figure 8, we show the results of our methods in Dataset II and Dataset III compared with the performance of the state-of-the-art target detection networks, including SSD300, YOLOV4 and Faster R-CNN. The results of the comparison are as follow.

Table 4. Performance of different model on Dataset II.

Model	Backbone	AP (%)	Average Time (s)
SSD300	VGG16	87.20	0.07
YOLOV4	Darknet	93.91	0.09
Faster R-CNN(with FPN)	Resnet50	88.01	0.26
Ours	Resnet50	94.82	0.49

Table 5. Performance of different model on Dataset III.

Model	Backbone	AP (%)	Average Time (s)
SSD300	VGG16	40.92	0.30
YOLOV4	Darknet	69.33	0.28
Faster R-CNN(with FPN)	Resnet50	86.27	0.38
Ours	Resnet50	95.25	0.63

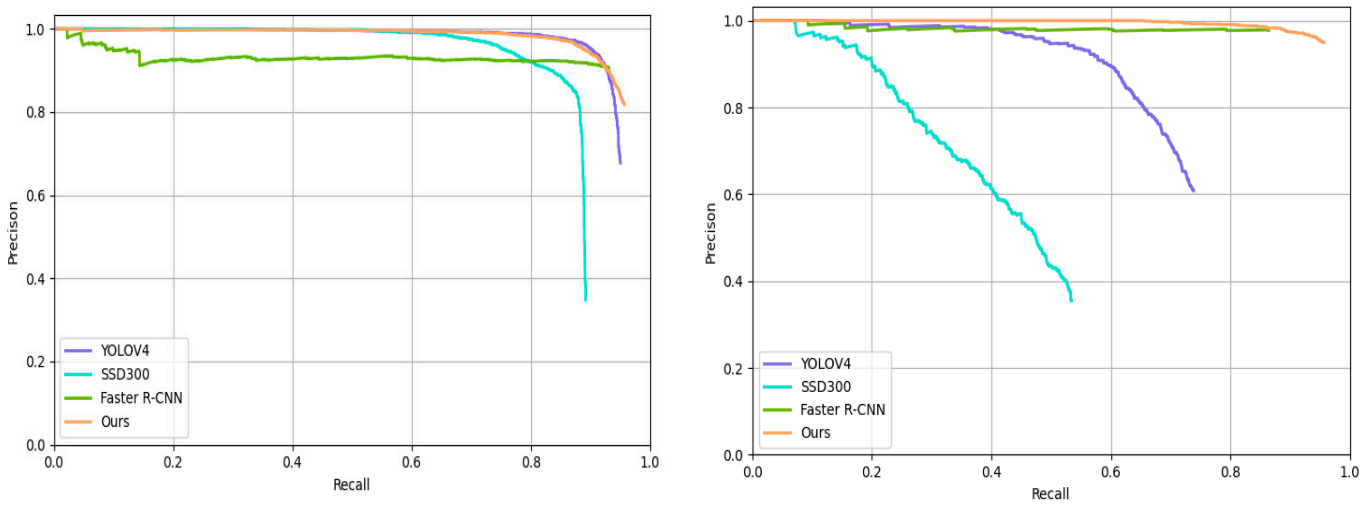


Figure 8. P-R curve of different models on dataset. (**a**) Dataset II. (**b**) Dataset III.

As we can see from Tables 4 and 5 and Figure 8, the method of combing multi-angle features driven strategy and majority voting strategy proposed in this paper has the best performance in terms of AP. The AP of the model proposed in this paper is 94.82% on Dataset II, which achieves better accuracy than the most advanced target detection networks at present. On Dataset III, the AP also achieved 95.25%. In addition, on Dataset II with a small image range and clear targets, all models achieved good performance. However, on Dataset III with a large image range and small and weak targets, the performance of one-stage target detection networks changed a lot, the change of two-stage target detection networks was relatively little. The performance of our proposed methods combining multi-angle features driven strategy and majority voting strategy is stable, and the AP still reached 95.25%.

4.2. Ablation Experiment

4.2.1. The Effectiveness of Multi-Angle Features Driven Strategy

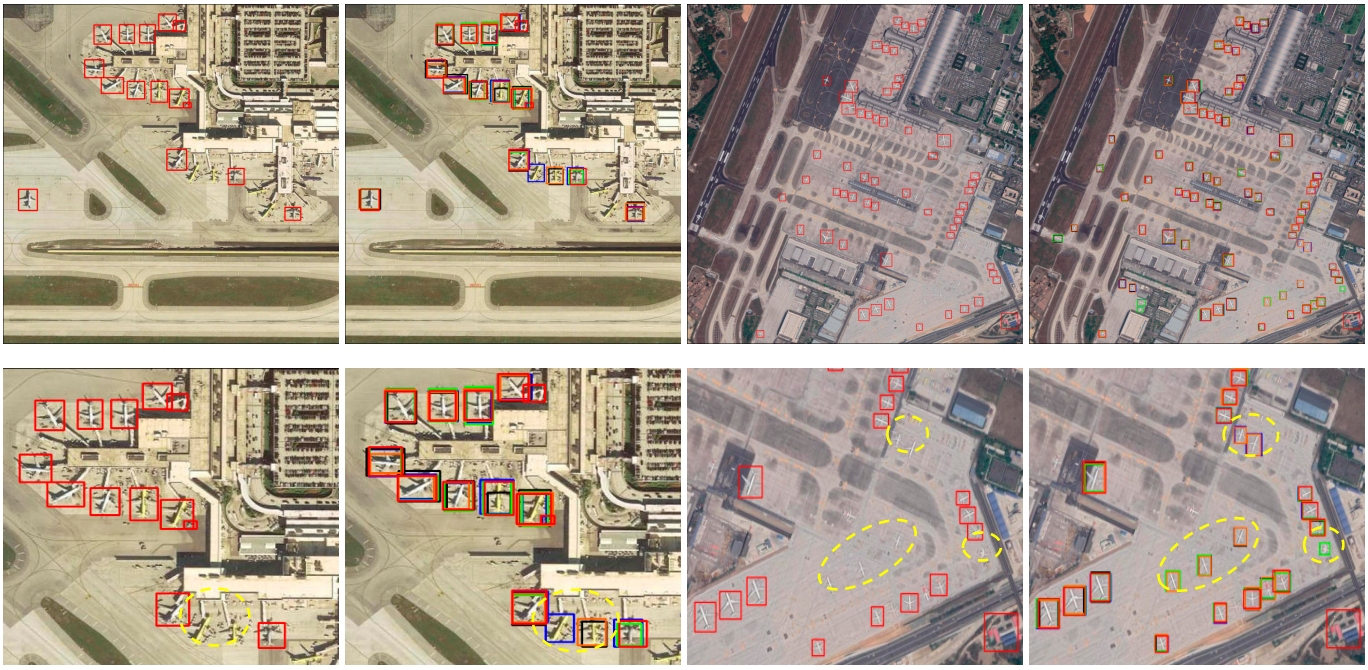
To verify the effectiveness of the multi-angle features driven strategy, we tested it on Dataset II and Dataset III without using multi-angle transformation module to verify the performance of the model. As we can see from Tables 6 and 7, the AP of the model with multi-angle features driven strategy is significantly higher than that without it, and the AP reaches 93.09% and 94.51%, respectively. Compared with the model that lacks multi-angle features driven strategy, the AP is improved by 5.08% and 8.24%, respectively. Figure 9 shows some samples that compare the results of model with multi-angle features driven strategy with that without multi-angle features driven strategy. It can be seen from the figures, although the targets are missed due to the large image range or the small and dense aircrafts in the image, the missed aircrafts are detected again through the multi-angle features driven strategy, which makes the missed detection rate of the network reduce.

Table 6. Performance of models with different module on Dataset II.

Model	AP (%)	Average Time (s)
Faster R-CNN (with FPN)	88.01	0.26
Ours (with FPN and Multi-Angle)	93.09	0.28

Table 7. Performance of models with different module on Dataset III.

Model	AP (%)	Average Time (s)
Faster R-CNN (with FPN)	86.27	0.38
Ours (with FPN and Multi-Angle)	94.51	0.40



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

Figure 9. Comparison of detection results on test set (The second line are partial enlarge views and the yellow dashed ellipse is used to emphasize the changed area). (**a**,**c**) Results of Faster R-CNN. (**b**,**d**) Results after multi-angle features driven strategy (different color boxes represent the preliminary detection results). (**a**,**b**) Dataset II. (**c**,**d**) Dataset III.

Therefore, it can be proved that multi-angle features driven strategy can effectively use the difference caused by image rotation and reduce the loss of information to a certain extent and it has practical significance for improving the detection accuracy of targets in remote sensing images.

4.2.2. The Effectiveness of Majority Voting Strategy

In this part, we compared the detection results without majority voting strategy with that using it to verify the effectiveness of majority voting strategy. The experimental results are shown in Tables 8 and 9.

Table 8. Performance of models with different module on Dataset II.

Model	AP (%)	Average Time (s)
Ours (Without Voting)	93.09%	0.28 s
Ours	94.82%	0.49 s

Table 9. Performance of models with different module on Dataset III.

Model	AP (%)	Average Time (s)
Ours (Without Voting)	94.51%	0.40 s
Ours	95.25%	0.63 s

It can be seen from Tables 8 and 9, the model with majority voting strategy has the best performance on both datasets. Compared with that without majority voting strategy, the AP of our model increased by 1.73% and 0.74% on Dataset II and Dataset III, respectively. At the same time, we can see from Figure 10 that the false detections of aircraft targets in the detection process are eliminated to a certain extent by majority voting processing, which reduces the false alarm rate of the model.

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

Figure 10. Results after majority voting strategy (The second line are partial enlarge views and the yellow dashed ellipse is used to emphasize the changed area). (**a**,**c**) Results after multi-angle features driven strategy (different color boxes represent the preliminary detection results). (**b**,**d**) Results after majority voting strategy. (**a**,**b**) Dataset II. (**c**,**d**) Dataset III.

Therefore, the combination of multi-angle features driven and majority voting strategy proposed in this paper can effectively improve the overall detection performance of the model.

4.3. The Limitation of the Model

Although the model proposed in this paper has good performance in AP, it also has certain limitations. As we can see from Tables 4 and 5, our model has increased AP, but its time consumption increased relatively. The main reasons are listed as below:

(1) Compared with single feature extraction, we performed multiple feature extraction on the image to reduce the loss of information, but this also led to an increase in the complexity of the model, resulting in excess time consumption.

(2) In the post-processing of the box detection, the voting and de-duplication of the detection boxes need to be completed through multiple cycles, which is also a crucial reason for the increased time consumption of the model.

Besides, through further exploration of the missed detections and false detections that still exist in the detection results, we summarized the reasons as follows:

(1) In the process of network training, the inability to optimize the selection of hyper-parameters, such as learning rate and batch_size, is one of the reasons for missed detections.

(2) Although the strategy of multi-angle features driven can reduce the miss of target feature to a certain extent, it still cannot fully compensate for the deficiencies of the network structure.

(3) The false alarm has been reduced to a certain extent through majority voting strategy, but the accuracy of the binary classification network for determining the targets in the inferior boxes is limited, which is also one of the reasons why false alarm still exists.

Therefore, we will pay more attention to the further simplification and optimization of the model structure and voting algorithm in future work.

5. Conclusions

Most target detection models based on convolutional neural networks are unable to fully extract features from remote sensing images and the results of target detection are mostly unprocessed, which can easily lead to missed detections and false detections of targets. In this paper, we presented a multi-angle features driven method and a majority voting strategy to adequately extract features in high resolution remote sensing images and to optimize the target detection results. Combining these methods, an aircraft target detection model for high resolution remote sensing images was proposed. Experimental results showed that the model could greatly reduce the missed detection rate and false alarm rate in target detection.

Through several groups of comparative experiments, it is verified that the performance of proposed model is obviously better than the existing networks of target detection. Although the model has an increase in time consumption compared with the existing target detection networks, it is acceptable in practical applications with the improvement of accuracy brought by the model proposed in this paper and by the development of computer hardware. In addition, it should be noted that this paper is mainly aimed at aircraft detection and the method proposed is theoretically applicable to other kinds of targets in remote sensing images. In the future, we will also conduct detections for other target types.

Author Contributions: F.J. proposed the concept and designed the model; D.M. wrote and review; B.Z. undertook the data processing and modified manuscript; J.Y. assisted in the design of model; Y.Q., T.D. and X.Z. gave comments and edits. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Key Research and Development Program (2017YFB0503600), China Geological Survey [DD20191006], the National Natural Science Foundation of China [41872253] and the Fundamental Research Funds for the Central Universities.

Data Availability Statement: Parts of related data can be found at https://github.com/RSIA-LIESMARS-WHU/RSOD-Dataset-, accessed on 23 April 2021 and http://www.escience.cn/people/gongcheng/DIOR.html, accessed on 23 April 2021.

Acknowledgments: Many thanks to Wuhan University and Northwestern Polytechnical University for providing RSOD and DIOR datasets.

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

References

1. Wu, Z.-Z.; Weise, T.; Wang, Y.; Wang, Y. Convolutional Neural Network Based Weakly Supervised Learning for Aircraft Detection From Remote Sensing Image. *IEEE Access* **2020**, *8*, 158097–158106. [CrossRef]
2. Wu, Z.-Z.; Wan, S.-H.; Wang, X.-F.; Tan, M.; Zou, L.; Li, X.-L.; Chen, Y. A benchmark data set for aircraft type recognition from remote sensing images. *Appl. Soft Comput.* **2020**, *89*, 106132. [CrossRef]
3. Zhao, A.; Fu, K.; Wang, S.; Zuo, J.; Zhang, Y.; Hu, Y.; Wang, H. Aircraft recognition based on landmark detection in remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1413–1417. [CrossRef]
4. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]

5. Fu, Z.; Chen, Y.; Yong, H.; Jiang, R.; Zhang, L.; Hua, X.-S. Foreground gating and background refining network for surveillance object detection. *IEEE Trans. Image Process.* **2019**, *28*, 6077–6090. [CrossRef]
6. Dai, X. HybridNet: A fast vehicle detection system for autonomous driving. *Signal. Process. Image Commun.* **2019**, *70*, 79–88. [CrossRef]
7. Zhang, X.; Wang, H.; Xu, C.; Lv, Y.; Fu, C.; Xiao, H.; He, Y. A lightweight feature optimizing network for ship detection in SAR image. *IEEE Access* **2019**, *7*, 141662–141678. [CrossRef]
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
9. Chang, D.; Ding, Y.; Xie, J.; Bhunia, A.K.; Li, X.; Ma, Z.; Wu, M.; Guo, J.; Song, Y.-Z. The devil is in the channels: Mutual-channel loss for fine-grained image classification. *IEEE Trans. Image Process.* **2020**, *29*, 4683–4695. [CrossRef] [PubMed]
10. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
11. 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.
12. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
13. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Santiago, Chile, 13–16 December 2015; pp. 3431–3440.
14. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef]
15. Hu, G.; Yang, Z.; Han, J.; Huang, L.; Gong, J.; Xiong, N. Aircraft detection in remote sensing images based on saliency and convolution neural network. *Eurasip J. Wirel. Commun. Netw.* **2018**, *2018*, 1–16. [CrossRef]
16. Shi, L.; Tang, Z.; Wang, T.; Xu, X.; Liu, J.; Zhang, J. Aircraft detection in remote sensing images based on deconvolution and position attention. *Int. J. Remote Sens.* **2021**, *42*, 4241–4260. [CrossRef]
17. Cheng, G.; Zhou, P.; Han, J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7405–7415. [CrossRef]
18. Zhang, Z.; Jiang, R.; Mei, S.; Zhang, S.; Zhang, Y. Rotation-invariant feature learning for object detection in VHR optical remote sensing images by double-net. *IEEE Access* **2019**, *8*, 20818–20827. [CrossRef]
19. Wu, Y.; Ma, W.; Gong, M.; Bai, Z.; Zhao, W.; Guo, Q.; Chen, X.; Miao, Q. A coarse-to-fine network for ship detection in optical remote sensing images. *Remote Sens.* **2020**, *12*, 246. [CrossRef]
20. Zhu, M.; Xu, Y.; Ma, S.; Li, S.; Ma, H.; Han, Y. Effective airplane detection in remote sensing images based on multilayer feature fusion and improved nonmaximal suppression algorithm. *Remote Sens.* **2019**, *11*, 1062. [CrossRef]
21. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
22. Chen, F.; Ren, R.; Van de Voorde, T.; Xu, W.; Zhou, G.; Zhou, Y. Fast automatic airport detection in remote sensing images using convolutional neural networks. *Remote Sens.* **2018**, *10*, 443. [CrossRef]
23. Guo, H.; Bai, H.; Zhou, Y.; Li, W. DF-SSD: A deep convolutional neural network-based embedded lightweight object detection framework for remote sensing imagery. *J. Appl. Remote Sens.* **2020**, *14*, 014521. [CrossRef]
24. Xie, Y.; Cai, J.; Bhojwani, R.; Shekhar, S.; Knight, J. A locally-constrained yolo framework for detecting small and densely-distributed building footprints. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 777–801. [CrossRef]
25. Chen, K.; Li, J.; Lin, W.; See, J.; Wang, J.; Duan, L.; Chen, Z.; He, C.; Zou, J. Towards accurate one-stage object detection with ap-loss. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5119–5127.
26. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
27. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
28. Fu, C.-Y.; Liu, W.; Ranga, A.; Tyagi, A.; Berg, A.C. Dssd: Deconvolutional single shot detector. *arXiv* **2017**, arXiv:1701.06659.
29. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
30. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
31. 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; pp. 580–587.
32. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv* **2015**, arXiv:1506.01497. [CrossRef] [PubMed]
33. Wu, H.; Zhang, H.; Zhang, J.; Xu, F. Typical target detection in satellite images based on convolutional neural networks. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China, 9–12 October 2015; pp. 2956–2961.

34. Long, Y.; Gong, Y.; Xiao, Z.; Liu, Q. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2486–2498. [CrossRef]
35. Liu, Q.; Xiang, X.; Wang, Y.; Luo, Z.; Fang, F. Aircraft detection in remote sensing image based on corner clustering and deep learning. *Eng. Appl. Artif. Intell.* **2020**, *87*, 103333. [CrossRef]
36. Feng, Y.; Wang, L.; Zhang, M. A multi-scale target detection method for optical remote sensing images. *Multimed. Tools Appl.* **2019**, *78*, 8751–8766. [CrossRef]
37. Li, Y.; Zhang, S.; Zhao, J.; Tan, W. Aircraft Detection in Remote Sensing Images Based on Deep Convolutional Neural Network. In Proceedings of the 2017 International Conference on Computer Technology, Electronics and Communication (ICCTEC), Dalian, China, 20–21 October 2018; pp. 942–945.
38. Fu, K.; Chang, Z.; Zhang, Y.; Xu, G.; Zhang, K.; Sun, X. Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *161*, 294–308. [CrossRef]
39. Bao, S.; Zhong, X.; Zhu, R.; Zhang, X.; Li, Z.; Li, M. Single shot anchor refinement network for oriented object detection in optical remote sensing imagery. *IEEE Access* **2019**, *7*, 87150–87161. [CrossRef]
40. Qu, J.; Su, C.; Zhang, Z.; Razi, A. Dilated convolution and feature fusion SSD network for small object detection in remote sensing images. *IEEE Access* **2020**, *8*, 82832–82843. [CrossRef]
41. Yin, R.; Zhao, W.; Fan, X.; Yin, Y. AF-SSD: An Accurate and Fast Single Shot Detector for High Spatial Remote Sensing Imagery. *Sensors* **2020**, *20*, 6530. [CrossRef]
42. Pham, M.-T.; Courtrai, L.; Friguet, C.; Lefèvre, S.; Baussard, A. YOLO-Fine: One-stage detector of small objects under various backgrounds in remote sensing images. *Remote Sens.* **2020**, *12*, 2501. [CrossRef]
43. Long, Z.; Suyuan, W.; Zhongma, C.; Jiaqi, F.; Xiaoting, Y.; Wei, D. Lira-YOLO: A lightweight model for ship detection in radar images. *J. Syst. Eng. Electron.* **2020**, *31*, 950–956. [CrossRef]
44. Honari, S.; Yosinski, J.; Vincent, P.; Pal, C. Recombinator networks: Learning coarse-to-fine feature aggregation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5743–5752.
45. Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv* **2017**, arXiv:1712.04621.
46. Zhou, Y.; Liu, X.; Zhao, J.; Ma, D.; Yao, R.; Liu, B.; Zheng, Y. Remote sensing scene classification based on rotation-invariant feature learning and joint decision making. *Eurasip J. Image Video Process.* **2019**, *2019*, 1–11. [CrossRef]
47. Yan, H. Aircraft detection in remote sensing images using centre-based proposal regions and invariant features. *Remote Sens. Lett.* **2020**, *11*, 787–796. [CrossRef]
48. Li, L.; Zhang, S.; Wu, J. Efficient object detection framework and hardware architecture for remote sensing images. *Remote Sens.* **2019**, *11*, 2376. [CrossRef]
49. Xiao, Z.; Liu, Q.; Tang, G.; Zhai, X. Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images. *Int. J. Remote Sens.* **2015**, *36*, 618–644. [CrossRef]
50. Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 296–307. [CrossRef]