IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 11, 2020, accepted August 31, 2020, date of publication September 4, 2020, date of current version October 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3021895

Effective Complex Airport Object Detection in Remote Sensing Images Based on Improved End-to-End Convolutional Neural Network

YONGSAI HAN[1], SHIPING MA[2], YUELEI XU[3], LINYUAN HE[2], SHUAI LI[1], AND MINGMING ZHU[1]

[1]Graduate School, Air Force Engineering University, Xi'an 710038, China
[2]Aeronautics Engineering College, Air Force Engineering University, Xi'an 710038, China
[3]Unmanned Systems Technology Institute, Northwestern Polytechnical University, Xi'an 710038, China

Corresponding author: Shiping Ma (masp_kgd@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61701524.

ABSTRACT Airport objects are hotspots in the field of image object detection because of their specific features and value for applications. In this study, we developed a complex object detection method based on improved Faster R-CNN to achieve higher detection precision to detect seven types of remote sensing image objects in airport areas under complex conditions such as different scales, different visual angles, and different backgrounds. When building the network, we used deeper basic networks and feature fusion components to extract more robust features. At the same time, we had also modified the selection of positive and negative samples to improve sample imbalance. The main improvements in the algorithm concern the anchor size generation rule, and the addition of an a priori judgment network for the network. The effectiveness of the improved algorithm was verified in experiments. Compared with the original Faster R-CNN, the improved network brings a 12.7% increase in mAP, at the detection time of 0.307s. Finally, the model with trained weights was used to test the detection of the seven types of objects in airport areas on different data sets, and comparisons were conducted with other algorithms. The experimental results showed that the method improved the average detection accuracy and had a good performance in remote sensing airport object detection tasks.

INDEX TERMS Airport object, image processing, multi-class object detection, pattern recognition, remote sensing.

I. INTRODUCTION

Due to rapid developments in remote sensing imaging technology, the imaging resolution has increased greatly and the ability to acquire information has gradually improved. Therefore, the full exploitation and utilization of remote sensing images has become a hotspot in computer vision and remote sensing research [1]. Airport areas are key building for military and civilian usage, and thus, they have become a major focus in the field of object detection [2].

Many studies have investigated the detection of airport areas. In particular, before using CNN, many scholars mostly used artificially designed features combined with the method of locating regions of interest to detect objects in images. In this way, Zhao *et al.* [3] combined the Hough transform and graph-based visual saliency (GBVS) method for airport detection. Tao *et al.* [4] combined the segmentation region and scale-invariant feature transform (SIFT) feature statistical method for airport detection. At that time, this method accelerated the detection accuracy of airport objects to a certain extent, but due to the lack of robustness of artificial design features, it was often disturbed by similar objects. Since the 21st century, convolutional neural network (CNN) had been successfully used in a large number of fields such as detection, segmentation, object recognition, etc. Zhu *et al.* [5] obtained regions of interest (ROI) from the saliency maps of geometric saliency and local entropy, and then used AlexNet to conduct network migration learning for airport identification. Zhang *et al.* [6] proposed a type of prior knowledge to develop a regional suggestion method

The associate editor coordinating the review of this manuscript and approving it for publication was Shiqi Wang.

with linear characteristics for airports and a CNN for airport detection based on the strong classification capacity of the CNN. Due to the use of CNN, the problem of insufficient robustness of artificial design features had been overcome to a certain extent, and the detection accuracy had been improved. However, it still had the same limitations as the traditional method when selecting proposed regions. So its detection efficiency was not very high. In response to this problem, many methods that gave consideration to detection accuracy and detection efficiency have been proposed one after another. Tan *et al.* [7] proposed a novel method for aircraft detection in high-resolution SAR images based on a gradient textural saliency map. Wang *et al.* [8] proposed a novel method for aircraft detection based on high-resolution panchromatic optical remote sensing images. Dai *et al.* [9] used Faster R-CNN and a multi-component combination method for airport and aircraft detection. Cai *et al.* [10] proposed the end-to-end convolutional neural network with hard example mining method for airport detection, where the detection performance was enhanced and the time cost was reduced. Chen *et al.* [11] proposed the transfer learning method for airplane detection in remote sensing images. Their methods improved the detection rate. Xu *et al.* [12] proposed the multi-layer feature fusion in fully convolutional neural networks method for airplane detection in remote sensing images. Xu *et al.* [13] proposed the end-to-end method combining cascade region proposal networks and multi-threshold detection networks for airport detection in remote sensing images. They all built a detection framework based on a deep neural network adapted to the task for a certain type of research object (airport or airplane), and achieved good results. However, many experiments researched on airport object detection were conducted based on remote sensing images with a single object, and simple background under similar shooting distances and fixed vertical visual angle. The method was not sufficiently robustness for objects with different shooting distances, different visual angles, and more complex backgrounds. Very different results were produced when the same object image was obtained with different heights, visual angles, etc.

In this study, we constructed a data set including 7 typical types of airport objects and containing many complex conditions for the problems described above, and referred to some ideas in the above-mentioned literature to use to build a network framework, and made some improvements to the Faster R-CNN [14] algorithm to improve the detection accuracy of objects. The details of the construction of the network and improvements are as follows.

1. In this study, we constructed a multi-class, multi-scale, multi-view, and simple/complex background data set for this problem. This data set is more complete and the features are more distinct compared with other remote sensing data sets (like DOTA) with related objects (like plane). A data enhancement operation for rapid error correction in the annotation information is proposed, which greatly reduces the inspection time.

2. In the construction of the network, we establishing a feature fusion between different convolution features and the full convolutional network in order to reduce the feature dimensions and increase the nonlinear characteristics of the network, and a deeper basic network is used to extract more robust features. At the same time, in order to improve the imbalance between positive and negative training samples caused by the excessive number of negative samples during the training process, a screening network is employed to extract complex negative samples and reduce the number of optional negative samples.

3. In terms of improvement, an artificial anchor is unsuitable for the multi-class and multi-scale objects considered in this study, so a compatibility loss clustering method (CLCM) is proposed based on the compatibility loss function to autonomously generate the anchor scale. At the same time, the problem where the regional convolutional neural network judges the rationality of inter-class coexistence is solved by establishing a priori judgment network to assess the rationality of the final output.

The remainder of this paper is organized as follows. In Section 2, we describe the basic components of the Faster R-CNN, network construction, and improvements. In Section 3, we analyzed the experimental result, compared it with other methods, and tested on other commonly used databases. In Section 4, we give our overall conclusions.

II. METHODS

A. CONSTRUCTION OF THE OBJECT DETECTION NETWORK IN THE AIRPORT AREA OBJECT DETECTION TASK

1) DETECTION FRAMEWORK BASED ON FASTER R-CNN

Machine self-identification is mainly based on the model's decision regarding the input image, where the model is learned based on a large training data set using machine learning methods. At present, the neural network model obtains the best recognition performance and a deep neural network model trained with the deep learning method obtains the highest object recognition accuracy [15], [16]. Therefore, a deep region convolutional neural network (CNN) is employed to train the constructed data set in our proposed method. The main reference is the Faster R-CNN algorithm. The schematic illustration of the process is shown in Figure 1.

A regional CNN mainly comprises a detection network and a region proposal network (RPN). The detection network mainly determines the positioning and classification of the object, and it comprises a convolution network, ROI pooling layer, and fully connected layer. The convolutional network extracts the discriminative deep features of the object in a large number of convolution operations. The convolution formula is as follows:

$$y' = \frac{y + 2p - k}{s} + 1 \tag{1}$$

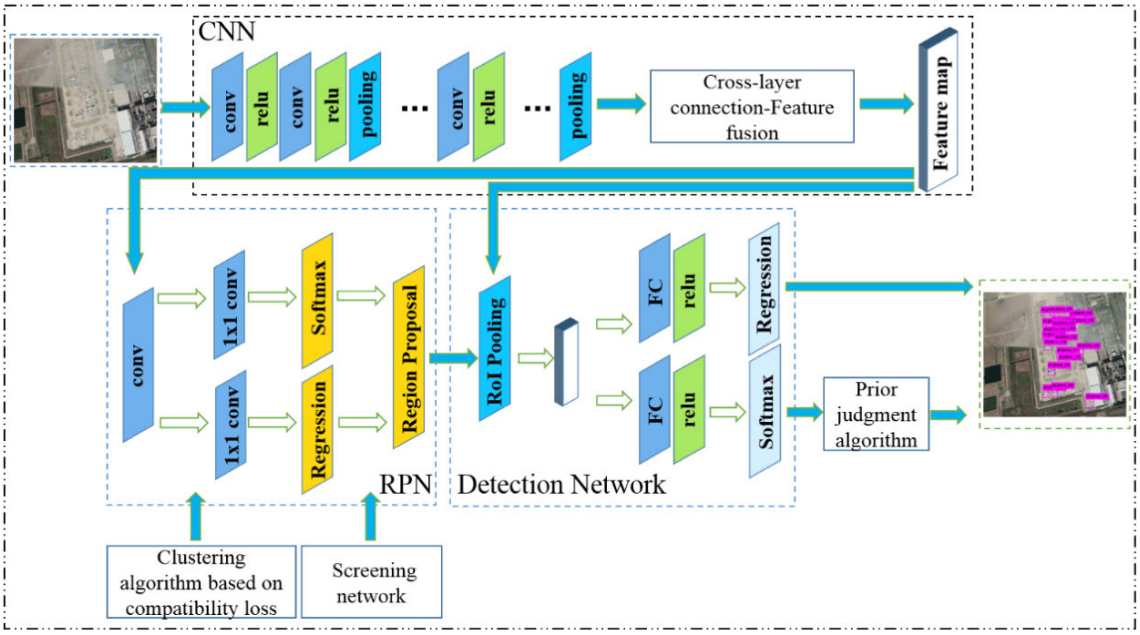

FIGURE 1. Simplified illustration of the network framework.

where y' is the size of the feature map, y is the size of the input matrix, p is the number of zero padding layers, k is the size of the convolution kernel, and s is the stride.

The ROI pooling layer mainly performs max-pooling based on the feature map generated by the RPN and the feature map generated by the model to generate a proper feature map, and then enters the full connection layer for subsequent training. The Pooling formula is as follows:

$$\max\_pooling_{out} = \max(a_{ij})_{i\in w, j\in h} \tag{2}$$

where $pooling_{out}$ is the output of the pooling layer and the $\max\_pooling_{out}$ represents that the pooling layer adopts the maximum pooling method, a_{ij} is an element of a max-pooling matrix, w is the width of a max-pooling matrix, h is the height of a max-pooling matrix, $\max(a_{ij})_{i\in w, j\in h}$ means to take the maximum value of the elements in the matrix.

The fully connected layer is mainly responsible for the specific classification of the object and further precise positioning of the position, but it requires a fixed size feature map, so the appropriate feature map output from the ROI pooling layer has a fixed size.

The RPN mainly makes regional recommendations where the extraction of the proposed boxes is based on the feature map. Clearly, a lower resolution feature map requires a smaller amount of calculations, thereby greatly reducing the time and storage space consumption requirements for the network's filtering windows based on the original image and because the overlapping portions of the suggested frame are extracted repeatedly multiple times, and thus the detection speed is also greatly improved. The specific details of this method were described in a previous study [14].

2) DEEPER BASIC NETWORK - RESNET

The depth of the neural network plays a crucial role in the improvement of model performance. Most of the networks that won on the early ImageNet dataset [17] adopted a deeper network structure, such as VGG-16 [18], GoogLeNet [19] And so on, so it also reveals the importance of network depth. However, as the number of network layers increases, it will inevitably bring the problems of gradient disappearance or explosion and decrease of accuracy after saturation due to simply increasing the network depth. The ResNet[20] proposed the concept of ''layer jump connection'' based on this problem, that is, the input is directly superimposed on the output. During the back propagation, the gradient of this path is transmitted back intact, and the correlation is very strong, as shown in Formula 3. Using it as the basic network of the detection network will be more conducive to feature extraction.

$$H_m(x) = F_m(x) + \omega x_n \tag{3}$$

where $H_m(x)$ is the output of the m-layer through ''shortcut connections'', $F_m(x)$ is the output of the original m-layer, x_n is the output of the n-layer ($m > n$), and ω is the convolution operation, which is used to change the dimension of x_n to make it consistent with $F_m(x)$.

3) SETTING OF FEATURE FUSION COMPONENTS AND SCREENING NETWORK

a: FEATURE FUSION COMPONENTS

For the small object detection problem, just like fighter, helicopter and oil tank are shown in Figure 2, if the last layer feature is detected separately according to the original algorithm due to the small object occupying few pixels in

(a) (b) (c)

FIGURE 2. The schematic diagram of fighter(a), helicopter(b) and oil tank object(c).

the high-level feature map, then the information contained is highly abstract. Studies have shown that high-level semantic information is more suitable for the classification of objects whereas lower-level feature information is better for the positioning of objects. Therefore, the detection of small objects based on multi-layer feature fusion is more suitable for the classification and localization of small objects.

Cross-layer connection is a classic method for the multi-layer feature fusion problem based on a full convolution network [21], HyperNet [22], and other approaches. In addition, the dimensionality problem after cross-layer connection can be addressed effectively with the 1×1 convolution kernel in GoogLeNet and ResNet. This method can not only change the dimensionality under the premise of a constant feature scale, but also the subsequent nonlinear activation function is used to increase the nonlinear characteristics of the network. We employ the convolution method to achieve the feature fusion and 1×1 convolutions to decrease feature dimensions and increase the nonlinear characteristics of the network. The original and improved network structures are shown in Figure 3.

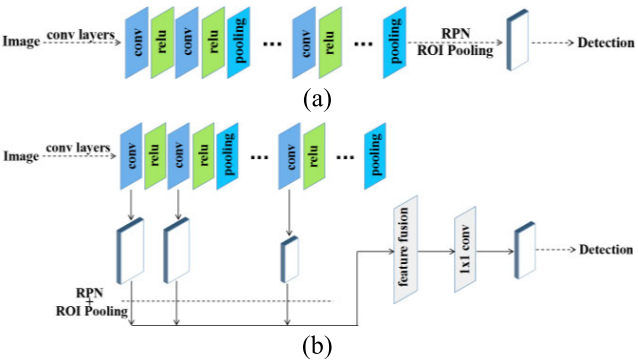

FIGURE 3. The network structure. (a) original network; (b) improved network.

Here, we first use RPN and ROI Pooling to extract the feature vectors of the region proposal that need to be fused. And then, the feature is fused by convolution operations. The relevant formula is as follows.

$$z_{ff}(x, y, w) = \sum_{i=1}^{k} x_i * w_i + \sum_{i=1}^{k} y_i * w_{i+k} \tag{4}$$

where z_{ff} is the fused output, x_i is the one of the input channel (x), y_i is another one of the input channel (y), k is the number of the input channel, w_i is the convolution parameters (i), and this formula also works for more than two inputs.

After that, there is possible to increase the output feature map dimension. Because if the feature maps of the corresponding channels have a large semantic difference, we should add new convolution kernel by formula 4. Finally, the 1×1 convolution is used to decrease feature dimensions and increase the nonlinear characteristics of the network.

b: SCREENING NETWORK

Establish a screening network by defining a sample with an IoU value between (0, 0.1) as a hard negative sample and filter it out as a negative sample for training, while strictly controlling the number of samples so they equal the number of positive samples. If this is not the case, add the remaining negative samples. Improve the balance among positive and negative samples during the training process.

The traditional method for defining a positive sample involves determining the box with the highest score among a set of boxes and a box with a ground truth IoU value greater than 0.7. The method for defining a negative sample involves determining a box with a ground truth IoU value less than 0.3 among a set of boxes. The remaining boxes do not contribute to the training process. In the airport object detection task. Most of the objects considered in the proposed method are small in size, so there will be many negative samples, which can readily lead to an imbalance between positive and negative samples during training, thereby making the training process sensitive to negative samples (background) and insensitive to positive samples (objects). Thus, a screening network is established to select hard negative samples with values between (0, 0.1) based on the IoU values calculated using the non-maximum suppression (NMS) algorithm for training the RPN in order to reduce the amount of optional negative samples, and thus, the balance between positive and negative samples is improved.

B. IMPROVEMENT TECHNIQUES

Here is our related improvement techniques on the limitations of the original algorithm on the performance of the object detected in this paper.

1) AN IMPROVEMENT ANCHOR METHOD– CLCM

The compatibility loss clustering method (CLCM) is used to generate the size and proportion of the anchor instead of the original anchor generation rule so the anchor can be a more suitable fit to the size of most ground truth data, as well as improving the quality of the positive sample and eliminating many of the redundant calculations due to the use of a very large anchor, and avoiding not extracting global features of the object due to the use of a very small anchor.

The traditional method for determining the size of an anchor involves manually designing the size of an appropriate anchor based on the size of the object in the data set. However, due to knowledge limitations and the large volume of data, it is difficult to meet the requirements for detecting a

large number of objects by using many artificially designed anchors with suitable sizes, and problems will also be caused by using an anchor with an unreasonable size. Thus, the data set constructed in the present study contains 7 types of objects, where each object has many different shapes and sizes, especially the airport objects. It is very difficult to manually formulate the sizes of anchors to satisfy all of these objects. Therefore, we employ the CLCM to automatically calculate the sizes of the best 9 anchors (consistent with the number in Faster R-CNN) based on our data set. The specific steps in the algorithm are shown in Table 1.

TABLE 1. The steps of CLCM.

CLCM to generate the anchor
Step 1: Extract the ground truth area and proportional information for some objects from each type of object category from the RPN network as samples.
Step 2: The information extracted from various types of objects is expanded into a two-dimensional Euclidean space.
Step 3: Randomly initialize nine anchor boxes (the number selected is modeled based on the Faster R-CNN detection algorithm, where a very high number increases the computational complexity and a very low number might not represent the full size of the object) and compare the nine anchor boxes with samples selected from the ground truth information, before calculating the loss value for each box.
Step 4: The ground truth cluster with a small loss value becomes a class around the corresponding anchor.
Step 5: Calculate the average size of the boxes in each class as the new anchor.
Step 6: Repeat the steps described above until the loss value changes little after each iteration to obtain the optimal nine anchors.

We define the compatibility loss value for the anchor and ground truth in the algorithm as *Compati Loss*. When designing the *Compati Loss* function, we consider the IoU of the machine-generated anchor and the ground truth as the independent variable. The *Compati Loss* should be smaller when IoU is larger. The range of *Compati Loss* value is defined from zero to infinity. A value of zero indicates that the anchor matches completely with the ground truth. The compatibility of the anchor and ground truth is lower when the *Compati Loss* value is larger. In addition, when the compatibility increases from zero, the value of *Compati Loss* should become less sensitive to it and the decay rate will slow, so the value of *Compati Loss* will not change greatly when the value of IoU is close to one. The value of *Compati Loss* should become very sensitive when the IoU increases from zero to quickly react to the closeness of the predicted box to the ground truth box. In summary, the function is defined as follows:

$$Compati\ Loss = -\ln[IoU(w_{gt}^{i}, h_{gt}^{i}, w_{a}^{k}, h_{a}^{k})] \quad (5)$$

$$IoU_{(a,gt)} = \frac{S(A \cap G)}{S(A \cup G)} \quad (6)$$

where w_{gt}^{i} is the width of the ground truth (i), h_{gt}^{i} is the height of the ground truth (i), w_{a}^{k} is the width of the anchor (k), h_{a}^{k} is the height of the anchor (k), $IoU_{(a,gt)}$ is the intersection-over-union of the anchor and ground truth, $S(A \cap G)$ is the intersection area of the anchor and ground truth, and $S(A \cup G)$ is the union area of the anchor and ground truth.

2) ADD PRIOR JUDGMENT ALGORITHM TO AVOID TEST RESULTS THAT DO NOT AGREE WITH PRIOR EXPERIENCE

In the training test using Faster R-CNN, some unconventional test results were obtained. For example, during aircraft detection, some areas that are similar to the shape of an airport and have similar sizes are detected as airports. Therefore, it is necessary to include a prior judgment to avoid similar errors that do not conform to the rules.

Argumentation analysis identified a few cases of misdetection, which usually had low confidence degree values (generally less than 0.7). Therefore, after the final detection network determines the category information and the confidence degree, we add a prior judgment before the final detection result to eliminate detection errors. The specific steps are shown in Table 2.

TABLE 2. The steps of prior judgment algorithm.

Prior judgment algorithm
Step 1: Read the classification result for the detection network from the log file (where each label is assigned a label value of 0, 1, 2, ..., 6 in the order shown in Table 3) and the corresponding confidence degree.
Step 2: If multiple types of label are detected and the product of the label value is 0, Step 3 is performed; otherwise, the label name is generated directly as the output.
Step 3: Compare the average value of detection confidence between the non-zero label objects and the zero label objects. Output zero label value if the average value of detection confidence with the label value of zero is large; otherwise, output all other non-zero label values.
Step 4: Read the label value in Step 3 and output the corresponding label name.

The algorithm mainly aims to eliminate the possibility of detecting other backgrounds as airports during the detection of aircraft-type objects, bridges, and oil tank objects from a lower satellite visual angle. In addition, this method may eliminate the possibility of detecting other backgrounds as aircraft-type objects, bridges, and oil tank objects during the detection of airport objects from a higher satellite visual angle.

III. RESULTS

A. DATA SET

1) PRODUCTION, CHARACTERISTICS AND LABELS OF DATA SETS

The data set used in this experiment was in the VOC2007 format and the images were uniformly set to .jpg format. In addition, the numbers of the images started from "000001" in order to facilitate the subsequent model feature extraction and training processes to optimize the path searches for various types of information in terms of their parameter weights. The specific process is shown in Figure 4.

The object detection images considered in this study comprised remote sensing images. A previous study [23]

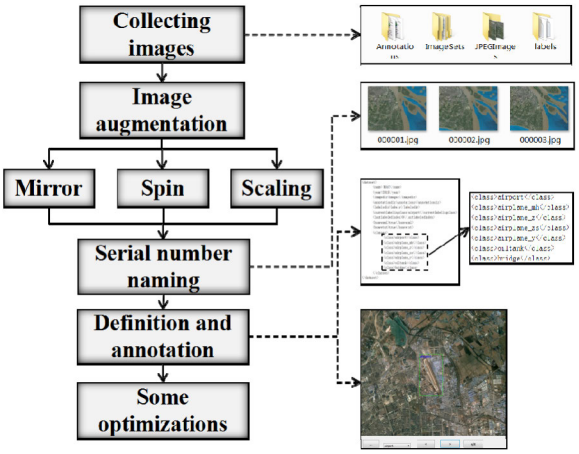

FIGURE 4. Data set construction flow chart.

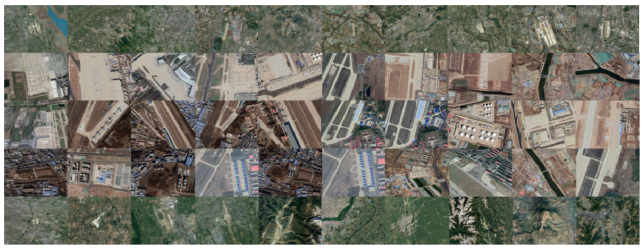

FIGURE 5. Schematic illustration of the data set.

TABLE 3. Labels and corresponding objects.

Label	Object	Label	Object
airport	airport	airplane_y	transport plane
airplane_mh	civil aircraft	bridge	bridge
airplane_z	fighter	oiltank	oil tank
airplane_zs	helicopter	-	-

showed that the remote sensing images acquired from Google map are highly generalizable for other sensors. Therefore, the remote sensing images provided by Google Earth Pro software were used as the sources for image capture and collection. Object remote sensing images with different heights and different visual angles were used by the neural network for autonomously learning multi-scale and multi-view objects. In this manner, more than 200 airports, aircraft, oil tanks, and bridges in the field were intercepted and collected, and 7264 original images were obtained without image augmentation.

A schematic illustration of the data set is shown in Figure 5.

The data set contains 1982 airport images, 1838 civil aircraft images, 565 fighter images, 715 helicopter images, 813 transport plane images, 583 bridge images, and 768 oil tank images.

The labels and the corresponding objects are shown in Table 3.

The object instances of this data set compared with the traditional data set are described in Table 4 and Figure 6.

TABLE 4. The Number of detected object instances-data set correspondence table.

	DOTA	UCAS-AOD	NWPU VHR-10	RSOD-Dataset	OURS
airport	0	0	0	0	2000
civil aircraft	15000	7482	900	4993	27500
fighter	0	0	0	0	2800
helicopter	950	0	0	0	5700
transport plane	0	0	0	0	4000
bridge	5500	0	100	0	1500
oil tank	10000	0	900	1586	24000

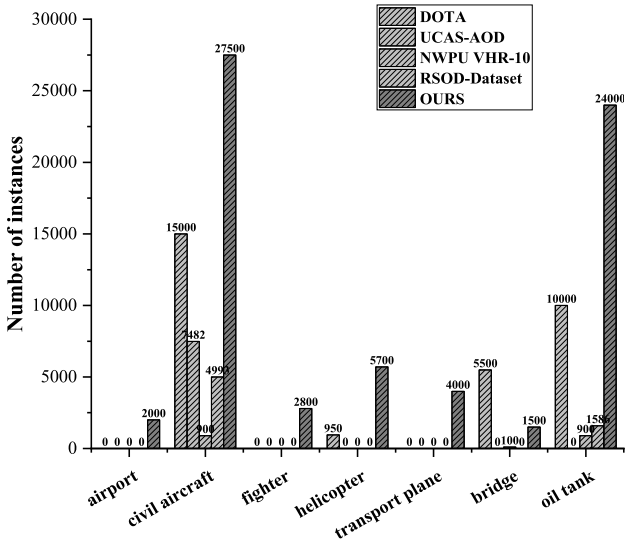

FIGURE 6. The histogram of comparison result.

It can be intuitively obtained from Figure 6 that the number of instances of typical targets such as airport, civil aircraft, fighter, helicopter, transport plane, and oil tank in our data set has exceeded the sum of the number of instances of four commonly used remote sensing data sets. For the bridge problem, we collect the bridges near the airport in a targeted manner, so the number of it is smaller than DOTA, but it is more robust for the task.

At the same time, our data set also has a larger target scale range, more shooting angles, and rich environmental characteristics and diverse structural characteristics.

Take the airport object as an example:

From the above analysis, it can be got that the data set constructed in this article may be more suitable for specific tasks-Airport Object Detection in Remote Sensing Images.

2) A SMALL OPTIMIZATION METHOD TO THE DATA SETS – RAPID MULTI-CLASS LABELS INFORMATION CORRECTION

In this paper, seven labels were used and thousands of original remote sensing images were processed before conventional data augmentation. Therefore, errors inevitably occurred during manual labeling due to factors such as misoperation and fatigue. The errors included omissions and mis-marks, and omissions could only be detected based on repeated

TABLE 5. Test results after using different basic networks and FF.

Basic network	mAP (%)	with FF	mAP(%)
ZFNet	53.3		
VGG_CNN_M_1024	58.5		
VGG-16	67.5	√	71.5
VGG-19	68.7	√	72.6
ResNet-50	69.0	√	73.1
ResNet-101	70.7	√	74.8

checks. Thus, a rapid label correction method was developed to address the problem of mis-marking. First, each type of object was placed in the same folder during image collection. Software was then applied for batch searches of the information in the folder where each object was located. Checks were performed to determine whether the ''.xml'' file contained the labels for the remaining six objects and they were located before correcting them, if necessary. After optimization, searches and correction were conducted for more than 100 mislabeled images in no more than 5 minutes, which helped the next step to achieved better training results.

B. EXPERIMENTAL PLATFORM AND PARAMETER SETTINGS

1) PLATFORM ENVIRONMENT

The processor comprised an Intel(R) Core(TM1) i9, the installed memory was 128 GB, the operating system was Ubuntu 16.04, and the graphics card was an NVIDIA GeForce GTX 2080Ti 11GB. The experimental framework was Caffe.

2) PARAMETER SETTINGS

The Caffe framework was used with open source deep learning. The pre-training network was used for initializing the parameters in the shared convolution layer. The remaining new layers were initialized randomly according to a Gaussian distribution with a mean of 0 and a standard deviation of 0.01. The initial learning rate was set to 0.001. The momentum was set to 0.9. The weight attenuation was set to 0.0005. The threshold was set to 0.7. The number of training epoch was 40,000. The sample ratio comprised training set: verification set: test set = 7:2:1.

C. ANALYSIS OF RESULTS

1) THE ANALYSIS OF COMPONENTS IN THE NETWORK CONSTRUCTION

1. Comparative analysis of detection precision after adding different basic networks and feature fusion components On the premise that the experimental environment and parameter settings are highly consistent, we have tested different basic networks on Faster R-CNN with adding feature fusion components, where ''FF'' represents the addition of feature fusion. The parts and experimental results are shown below.

Through experiments, it can be obtained that a deeper ResNet-101 is used as the basic network, and a better mAP can be obtained, with a value of 70.7%. Because it has more convolutional layers, and the ''shortcut connections'' residual structure, it can retain better feature correlation during the training process, avoiding the problem that the accuracy rate drops after saturation. In addition, through experiments, feature fusion has a better improvement effect for deeper networks. After the feature fusion component is added to ResNet-101, there is a 4.1% mAP improvement to 74.8%. Therefore, ResNet-101 is selected as the feature extraction basic network of the detection network, and feature fusion detection components are added to the network framework, and subsequent experiments are based on this.

TABLE 6. Test results obtained for data set after establishing a screening network.

Method	mAP(%)	Average detection time (s)
T1	74.8	0.306
T2	**75.3**	**0.306**

FIGURE 7. The schematic diagram of airport object.

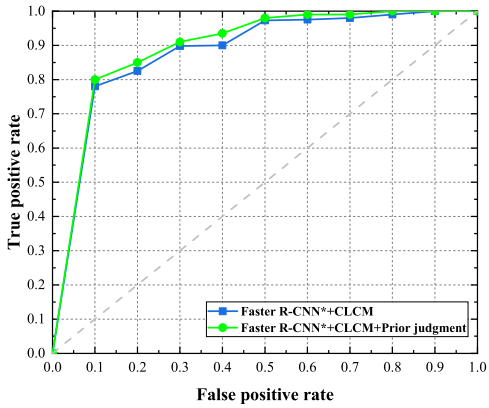

FIGURE 8. Comparison of the effects of prior judgment algorith.

2. The results after establishing a screening network.

The method for determining the training samples in the conventional manner is defined as T1, and the method for adding screening network is defined as T2. The results obtained after training and testing are shown in Table 6.

The test results showed that the model after joining the screening network improved mAP by 0.5%, and the detection time was basically unchanged, thereby demonstrating the effectiveness of the screening network.

TABLE 7. Test results after using CLCM and other methods.

Method	Average IoU	mAP(%)	Recall(%)	Precision(%)	Accuracy(%)
Faster R-CNN* + 9 anchors(a)	0.635	75.3	81.1	74.2	80.7
Faster R-CNN *+ 9 anchors(b)	0.648	75.4	81.1	74.5	81.0
Faster R-CNN *+ 12 anchors	0.663	76.6	81.4	77.6	82.5
Faster R-CNN*+K-means[27]	0.708	76.9	81.5	78.3	82.9
Faster R-CNN* + CLCM	**0.723**	**77.1**	**81.5**	**78.5**	**83.0**

FIGURE 9. The schematic diagram of airport detection.

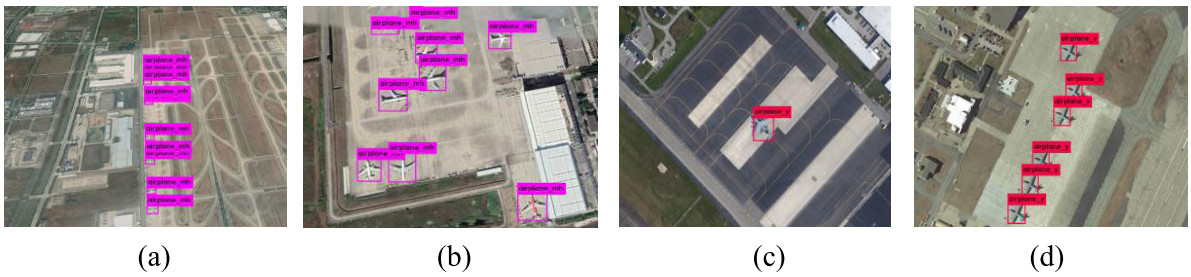

FIGURE 10. The schematic diagram of civil aviation and transport plane detection.

TABLE 8. Test results after using different basic networks.

Method	mAP	Recall	Precision	Accuracy
Faster R-CNN*+T1	77.1%	81.5%	78.5%	83.0%
Faster R-CNN*+T1+T2	**80.2%**	**81.5%**	**80.1%**	**83.8%**

2) THE ANALYSIS OF IMPROVED NETWORK RUNNING RESULTS

Here we will analyze the effects individually of the above improvement techniques through comparative experiments and related evaluation indicators.

1. The results after using CLCM.

Here, we tested the Faster R-CNN* (based on the above experimental basis) and the improved algorithm using the same experimental environment. Where Faster R-CNN*+ 9 anchors(a) represents the original algorithm. Faster R-CNN*+ 9 anchors(b) represents that the anchors are set by manual experience (the number of anchors is still 9). Faster R-CNN*+ 12 anchors represents that the anchors are set by manual experience (but the number of anchors is 12). The results are shown in Table 7.

It can be obtained from the experimental results that the addition of CLCM makes the model have improved to a certain degree in various evaluation indexes compared with the others algorithm, including commonly used k-means algorithm. This is mainly because CLCM uses a loss function that is more sensitive to IoU response. The Average IoU has the best improved effect, which has increased 8.8% compared with Faster R-CNN*, and the mAP has increased 1.8% compared with Faster R-CNN*, proving the effectiveness of the CLCM. At the same time, the process of the machine automatically fitting the 9 most suitable anchors is reflected before the detection (or cross training) process, so it does not produce time increase to the detection process.

2. The results after adding prior judgment algorithm.

The results are shown in Table 8 and the ROC curve drawn by experiment is shown in Figure 8. Here, L1 represents the CLCM algorithm, and L2 represents the prior decision algorithm.

Figure 8 shows that after adding the prior judgment algorithm, the network has a lower false positive rate at the same true positive rate, especially when the true positive rate is high. This is because the addition of a priori judgment network makes the number of false positive samples (False Positive, FP) less, while the number of true negative samples (True Negative, TN) increases. At the same time,

TABLE 9. Summary of each object test result.

Label	airport	airplane_mh	airplane_zs	airplane_z	airplane_y	oiltank	bridge
Object	airport	civil aviation	helicopter	fighter	transport plane	oil tank	bridge
AP (%)	87.3	91.4	79.3	73.2	78.2	77.4	74.5
mAP (%)				80.2			
T (s)				0.307			

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

FIGURE 11. The schematic diagram of helicopter and fighter detection.

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

FIGURE 12. The schematic diagram of oil tank and bridge detection.

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

FIGURE 13. The schematic diagram of multi-class object detection.

it is proved through experiments that the a priori judgment network has brought a 3.1% increase to 80.2% mAP, and there are also good improvements in other indicators. At the same time, as mentioned above, because the judgment network mainly brings about a decrease in FP and an increase in TN, the recall value does not change much. Thereby demonstrating the effectiveness of the prior judgment algorithm.

3) THE OVERALL TEST RESULT ANALYSIS

The results of test are shown in Table 9.

Overall, the results showed that the trained model achieved a good average precision (AP) for each type of object, and the mAP reached 80.2% with an average detection time of 0.307s. The detection of fighters and bridges (mainly near airports) achieved a lower detection rate mainly because there were less remote sensing images of fighter and bridge data compared with other objects due to some security measures, geographical factors, and other reasons. In addition, fighters

FIGURE 14. A poor detection result.

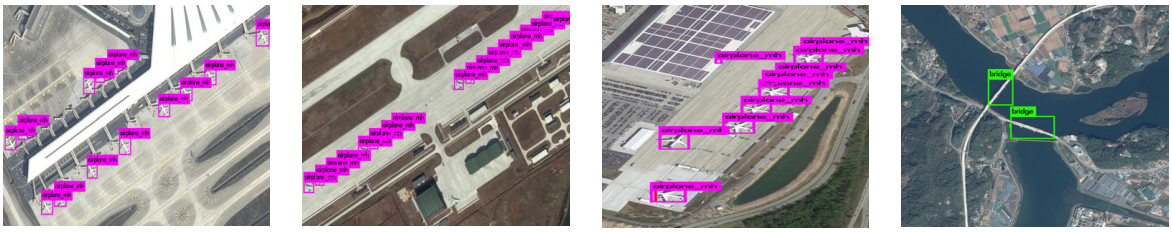

FIGURE 15. Schematic diagram of test results. The title of the images(like "P0033") is the title of the corresponding images in the DOTA data set.

FIGURE 16. Schematic diagram of test results. The title of the images(like "P0003") is the title of the corresponding images in the UCAS-AOD data set.

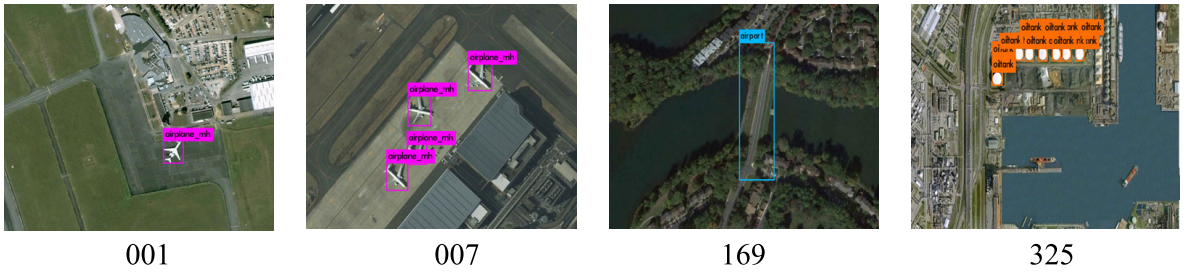

FIGURE 17. Schematic diagram of test results. The title of the images(like "001") is the title of the corresponding images in the NWPU VHR-10 data set.

are smaller than civil aircraft so fewer pixels are present in the high-range shooting angle, thereby making their detection more difficult.

4) VISUAL OBJECT DETECTION TEST RESULTS ON OUR DATA SET AND ANALYSIS

The results obtained for various object tests are shown in Figures 9–14.

1. Airport Detection
2. Aircraft Object Detection
3. Oil Tank and Bridge Object Detection
4. Multi-Class Object Detection
5. Example of a Poor Detection Result

The experimental results demonstrated the feasibility of the improved model based on regional CNN for multi-class, multi-objective, multi-view, and complex background detection in remote sensing images. The effectiveness of detection was good for most of the objects, with highly accurate and efficient detection, but there was a problem because the effective detection of small objects and small sample objects needs to be improved. As shown in the right panel in Figure 12(a), the model was not able to fully detect the oil tank object with a small number of pixels, so this problem needs to be addressed. In addition, the results demonstrated the importance of the data set's capacity, and thus building a broader data set is one of the next steps. In addition, understanding and correcting the poor detection of the tank object similar to that shown in Figure 14 is a future research objective.

5) TESTING ON OTHER DATA SETS

1. Testing on DOTA
2. Testing on UCAS-AOD
3. Testing on NWPU VHR-10
4. Testing on RSOD-Dataset

It can be seen from the diagram that the trained network still has better detection performance for related objects in other remote sensing data sets.

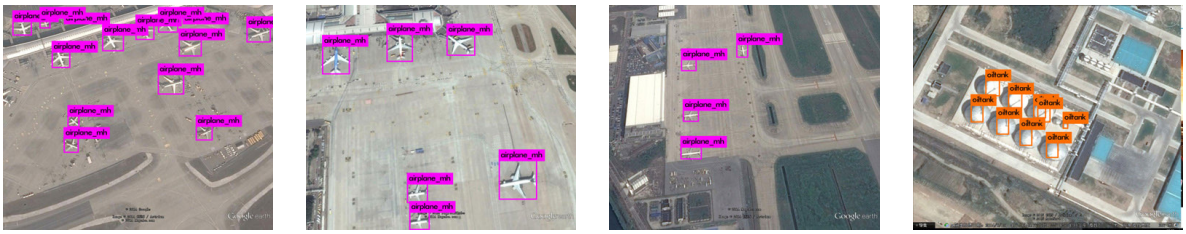

FIGURE 18. Schematic diagram of test results. The title of the images(like "aircraft_4") is the title of the corresponding images in the RSOD-Dataset data set.

TABLE 10. Comparison results of different detection methods.

Method	Faster R-CNN	R-FCN	SSD	YOLOv3	Ref.[24]	Ref.[23]	Faster R-CNN*	Proposed
mAP(%)	67.5	73.8	72.4	74.8	79.2	68.9	75.3	80.2
Detect time(s)	0.152	0.118	0.117	0.087	0.427	0.152	0.306	0.307

6) COMPARATIVE EXPERIMENT AND ANALYSIS

Using a highly consistent experimental environment and online published codes, some widely used networks were used for comparative experiments. The test results are shown in Table 10.

Through comparative experiments, it can be obtained that the network constructed in this paper has achieved higher mAP values compared with the currently widely used R-FCN [25], SSD [26], Ref [24] (Faster R-CNN+ InceptionResNetv2+ TDM), YOLOv3 [27], Ref [23] (Faster R-CNN++). The constructed network (Faster R-CNN*) brings a 7.8% improvement in mAP compared to the Faster R-CNN (with VGG-16 in source code) by using deeper ResNet-101 and feature fusion components. The improved method brought a 4.9% increase in mAP, and only increased the detection time cost of 0.001s. In summary, it proved the effectiveness and significance of the constructed network and improved methods. High detection precision can be achieved in remote sensing airport area object detection tasks.

IV. CONCLUSION

In this study, we made improvements to the two-stage detection method based on the Faster R-CNN algorithm according to the specific detection object and initial experimental results. The results showed that the improved network constructed in this article obtained better detection results with our new data set. And in other commonly used remote sensing data sets (like DOTA), it also has a good detection effect on related objects. The mAP was improved by 12.7% compared with the original Faster R-CNN algorithm. In addition, our experiments showed that the current CNN is highly dependent on the data set and the detection accuracy needs to be improved for small objects, especially those containing only a few pixels. Therefore, constructing a larger data set as well as deeper construction and optimization of the network, at the same time, on the premise of obtaining high detection accuracy, how to reduce the detection time as much as possible, are our next research priorities.

ACKNOWLEDGMENT

The author Yongsai Han thanks to Shiping Ma, Yuelei Xu, Linyuan He, Shuai Li, Mingming Zhu and all relatives and friends who participated in this work.

REFERENCES

[1] J. A. Richards, "Analysis of remotely sensed data: The formative decades and the future," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 422–432, Mar. 2005, doi: 10.1109/TGRS.2004.837326.
[2] J. Yao and Z. Zhang, "Semi-supervised learning based object detection in aerial imagery," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 1011–1016. [Online]. Available: https://ieeexplore.ieee.org/document/1467377
[3] D. Zhao, Y. Ma, Z. Jiang, and Z. Shi, "Multiresolution airport detection via hierarchical reinforcement learning saliency model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2855–2866, Jun. 2017, doi: 10.1109/JSTARS.2017.2669335.
[4] C. Tao, Y. Tan, H. Cai, and J. Tian, "Airport detection from large IKONOS images using clustered SIFT keypoints and region information," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 128–132, Jan. 2011, doi: 10.1109/LGRS.2010.2051792.
[5] T. Zhu, Y. Li, Q. Ye, H. Huo, and T. Fang, "Integrating saliency and ResNet for airport detection in large-size remote sensing images," in *Proc. 2nd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2017, pp. 20–25. [Online]. Available: https://ieeexplore.ieee.org/document/7984451.
[6] P. Zhang, X. Niu, Y. Dou, and F. Xia, "Airport detection on optical satellite images using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1183–1187, Aug. 2017, doi: 10.1109/LGRS.2017.2673118.
[7] Y. Tan, Q. Li, Y. Li, and J. Tian, "Aircraft detection in high-resolution SAR images based on a gradient textural saliency map," *Sensors*, vol. 15, no. 9, pp. 23071–23094, Sep. 2015, doi: 10.3390/s150923071.
[8] W. Wang, T. Nie, T. Fu, J. Ren, and L. Jin, "A novel method of aircraft detection based on high-resolution panchromatic optical remote sensing images," *Sensors*, vol. 17, no. 5, p. 1047, May 2017, doi: 10.3390/s17051047.
[9] C. Dai and Y. Li, "Static aircraft detection of airport scene based on Faster R-CNN and multi-component combination," *J. Comput. Appl.*, vol. 7, no. S2, pp. 85–88, Mar. 2017, doi: cnki:sun:jsjy.0.2017-S2-021.
[10] B. Cai, Z. Jiang, H. Zhang, D. Zhao, and Y. Yao, "Airport detection using End-to-End convolutional neural network with hard example mining," *Remote Sens.*, vol. 9, no. 11, p. 1198, Nov. 2017, doi: 10.3390/rs9111198.

[11] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Aircraft detection by deep belief nets," in *Proc. 2nd IAPR Asian Conf. Pattern Recognit.*, Nov. 2013, pp. 54–58. [Online]. Available: https://ieeexplore.ieee.org/document/6778281
[12] Y. Xu, M. Zhu, P. Xin, S. Li, M. Qi, and S. Ma, "Rapid airplane detection in remote sensing images based on multilayer feature fusion in fully convolutional neural networks," *Sensors*, vol. 18, no. 7, p. 2335, Jul. 2018, doi: 10.3390/s18072335.
[13] Y. Xu, M. Zhu, S. Li, H. Feng, S. Ma, and J. Che, "End-to-End airport detection in remote sensing images combining cascade region proposal networks and multi-threshold detection networks," *Remote Sens.*, vol. 10, no. 10, p. 1516, Sep. 2018, doi: 10.3390/rs10101516.
[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.
[15] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," presented at the CVPR, Jun. 2014. [Online]. Available: https://ieeexplore.ieee.org/document/6909475
[16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016, doi: 10.1109/TPAMI.2015.2437384.
[17] O. Russakovsky, J. Deng, and H. Su, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–253, Jan. 2015, doi: 10.1007/s11263-015-0816-y.
[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," presented at the CVPR, Jun. 2014. [Online]. Available: https://arxiv.org/abs/1409.1556
[19] C. Szegedy, W. Liu, Y. Jia, "Going deeper with convolutions," presented at the CVPR, Jun. 2015. [Online]. Available: https://arxiv.org/abs/1409.4842.
[20] K. He, X. Zhang, and S. Ren, "Deep residual learning for image recognition," presented at the CVPR, Jun. 2016. [Online]. Available: https://arxiv.org/pdf/1512.03385
[21] H. Liu, L. Peng, and J. Wen, "Multi-scale aware pedestrian detection algorithm based on improved full convolutional network," *Laser Optoelectron. Prog.*, vol. 55, no. 9, Apr. 2018, Art. no. 091504, doi: 10.3788/lop55.091504.
[22] T. Kong, A. Yao, Y. Chen, and F. Sun, "HyperNet: Towards accurate region proposal generation and joint object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 845–853. [Online]. Available: https://ieeexplore.ieee.org/document/7780467
[23] F. Chen, R. Ren, T. Van De Voorde, W. Xu, G. Zhou, and Y. Zhou, "Fast automatic airport detection in remote sensing images using convolutional neural networks," *Remote Sens.*, vol. 10, no. 3, p. 443, Mar. 2018, doi: 10.3390/rs10030443.
[24] A. Shrivastava, R. Sukthankar, and J. Malik, "Beyond skip connections: Top-down modulation for object detection," presented at the CVPR, Jun. 2016. [Online]. Available: https://arxiv.org/abs/1612.06851
[25] J. Dai, Y. Li, and K. He, "R-FCN: Object detection via region-based fully convolutional networks," presented at the CVPR, Jun. 2016. [Online]. Available: https://arxiv.org/pdf/1605.06409v2
[26] W. Liu, D. Anguelov, and D. Erhan, "SSD: Single Shot MultiBox Detector," presented at the ECCV, Jun. 2016. [Online]. Available: https://arxiv.org/abs/1512.02325
[27] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," presented at the CVPR, Jun. 2018. [Online]. Available: https://arxiv.org/abs/1804.02767

YONGSAI HAN received the bachelor's degree from Air Force Engineering University, in 2018, where he is currently pursuing the master's degree. He has participated in many laboratory research projects, including the National Natural Science Foundation of China and received good feedback. His research interest includes research and improvement of detection methods for multiple types of objects in airport areas under remote sensing detectors.

SHIPING MA received the bachelor's and Ph.D. degrees from Air Force Engineering University, in 1999 and 2004, respectively. He is currently working as an Associate Professor with Air Force Engineering University. He has participated in the National Natural Science Foundation and Aviation Science Foundation projects many times. His research interests include image processing, object detection, and so on. He has received the First Prize of Scientific and Technological Progress.

YUELEI XU received the Ph.D. degree from the Postdoctoral Research Mobile Station of Computer Science and Technology, Northwestern Polytechnical University, in 2012. From 2016 to 2018, he was a Visiting Scholar with the Department of Computational Neuroscience, Johns Hopkins University. He is currently working as a Professor with the Institute of Unmanned Systems Technology, Northwestern Polytechnical University. His research interests include deep intelligent perception and object detection as well as visual navigation and machine learning. He has accumulatively received six provincial and ministerial level scientific and technological progress awards.

LINYUAN HE received the bachelor's and master's degrees from Air Force Engineering University and the Ph.D. degree from Xi'an Jiaotong University. He is currently working as an Associate Professor with Air Force Engineering University. He served as the main person in charge. He has participated in the National Natural Science Foundation and Aviation Science Foundation projects many times. His research interests include image dehazing, object detection and tracking, and so on.

SHUAI LI received the bachelor's and Ph.D. degrees from Air Force Engineering University, in 2014 and 2019, respectively. He has participated in many laboratory research projects, including the National Natural Science Foundation of China and Aviation Basic Projects, and so on. His research interests include situational awareness and object detection in airport areas.

MINGMING ZHU received the bachelor's degree from Air Force Engineering University, in 2016, where he is currently pursuing the Ph.D. degree. He has participated in many laboratory research projects, including the National Natural Science Foundation of China and Aviation Basic Projects, and so on. His research interests include detection of airports and civil aviation aircraft under remote sensing detectors.