Improved High Precision Aircraft Target Detection Method of YOLT

Jiaxing Mao\textsuperscript{*}, Xinyu Zhang\textsuperscript{2}, Yang Ji\textsuperscript{3}, Zhen Zhang\textsuperscript{4}, Zihao Guo\textsuperscript{5}

\textsuperscript{1}Systems Engineering Research Institute of China State Shipbuilding Corporation, Beijing, 100000, China
\textsuperscript{2}Systems Engineering Research Institute of China State Shipbuilding Corporation, Beijing, 100000, China
\textsuperscript{3}Systems Engineering Research Institute of China State Shipbuilding Corporation, Beijing, 100000, China
\textsuperscript{4}Systems Engineering Research Institute of China State Shipbuilding Corporation, Beijing, 100000, China
\textsuperscript{5}Systems Engineering Research Institute of China State Shipbuilding Corporation, Beijing, 100000, China
\textsuperscript{*}Corresponding author’s e-mail: zhangfan61@meituan.com

Abstract. The first remote sensing dataset that can be used for aircraft target detection is created and a 3D model simulation method for data augmentation is proposed. After statistical analysis of aircraft sizes, a high-precision improved YOLT method for target detection is proposed. YOLT is the first light algorithm focusing on target detection in remote sensing images. We improve its network structure, design a creative method based on receptive field improvement and adopt an optimized non-maximum suppression strategy. The results show that our method has better performance than other main target detection algorithms, especially for small targets and cross-scale samples.

1. Introduction
High resolution remote sensing image contains sufficient target information, which is of great significance in many fields. The research on target detection and recognition of high-resolution image is an important research work at present. Among the numerous remote sensing image targets, aircraft has become a research hotspot because of its strategic significance in the military field and air transportation field.

In recent years, with the rapid development of deep learning technology, more and more research work has been done on aircraft target detection in remote sensing image background [1-5]. Many researchers have improved on the general and mature target detection algorithms, but lack of solutions for the special problems of remote sensing images. Xingyue Li [8] proposed a method of adding context information to supplement the feature expression of the target. The target candidate region was fused with the background features, and the fused features were used to locate the target. Although the accuracy was improved to a certain extent, the proposed method was only embedded in Faster R-CNN, which lacked experimental universality. YOLT [3], on the basis of Yolov2 [6], specifically proposes several improved methods for target detection of remote sensing images, which solves the problems of...
large size and dense target arrangement of remote sensing images. However, in order to realize the end-to-end target detection as fast as possible, this method adopts a relatively simple and lightweight network structure design. The experiment results show that the performance is poor when the target size span in the image is large and there are many small targets. In order to achieve high precision aircraft target detection in the background of remote sensing images, this paper proposes an improved YOLT aircraft target detection method to solve the problem of large target size variation and small target.

2. Related Work
Tong et al. [8], aiming at the complex remote sensing background, proposed a detection method combining background semantic information, designed sub-networks to extract and predict the background features in blocks, and combined background prediction information and target information to output the final result. Wang et al. [9] took RPN as the basic method and designed a sub-network to generate multi-angle candidate regions, which improved the detection accuracy of vehicles and aircraft, but this method increased the complexity of the model and the performance improvement effect was limited. Pang et al. [10] proposed a new deep learning detection method R2-CNN based on residual module as the main network for detecting small and medium-sized targets in remote sensing scenes. The method is composed of four modules: feature extraction network, global attention, classifier and re-detector. The detection results of small targets are modified step by step, and a higher detection accuracy is obtained.

3. Our Method
Firstly, aiming at the shortage of aircraft target dataset of remote sensing images, we construct an aircraft detection dataset RSAD (Remote Sensing Aircraft Detection), and innovatively adopt a variety of data enhancement methods. Then, we deeply analyze the image of aircraft targets under the same resolution, size distribution and the range of small target pixel information in network design, use the targeted context connection module, design the layer number and the location of the pooling layer, which make aircrafts of different sizes have a better adaptability, and complete the receptive field in the form of dilated convolution of ascension. Finally, in the post-processing stage, the traditional non-maximum suppression method is improved to improve the detection accuracy again.

3.1. RSAD dataset construction
In order to solve the problem of remote sensing image dataset shortage, this paper constructs a special dataset for remote sensing aircraft target detection and recognition. The images of different time, different background complexity and different light intensity are collected to ensure the diversity of the dataset.

In order to analyze the multi-scale problem and small-scale target problem of aircraft target detection in remote sensing background, this section makes a quantitative analysis on the target scale of RSAD and counts 1435 aircraft targets in 400 images of RSAD dataset. In this section, according to the convention of target detection, targets with both length and width of 20 pixels or less are called small aircraft targets, while targets with width of 400 pixels or more are called large aircraft targets. These two types of targets, large aircraft and small aircraft, are the key to improve the performance of the current detection methods.

The basic data enhancement methods include rotation scaling, contrast change, etc. Although such basic enhancement method can increase the number of samples, the samples generated are still images of the original dataset in essence, which fails to achieve the purpose of generating new samples, and the data enhancement effect is limited. Based on the commonly used data enhancement methods, this section innovatively proposes a method to enhance samples through 3D model simulation from the perspective of 3D model. This method can generate new sample images for the original dataset, enhance the diversity of the samples, and ensure the clarity and authenticity of the generated samples. The flow chart of this method is shown in Figure 1.
Firstly, we use the prior knowledge of the target characteristics to build a three-dimensional model of the aircraft. The characteristic of the three-dimensional model is that it can show various characteristics of the target in all directions. Secondly, Autodesk 3dsMax software is used to render and simulate the model to generate two-dimensional images. In the simulation process, the angle and attitude of the target can be changed without adding redundant information or losing information. Then, the segmentation algorithm is used to separate the target and the background in the simulation image, and the aircraft target is obtained. Finally, Google Earth mapping software is used in this section to intercept different backgrounds such as airfields and airport runways, and the segmented aircraft targets are fused with different backgrounds to obtain the generated new samples. Some sample images generated by this method are shown in Figure 2.

Figure 1. Method for data enhancement

3.2. Network structure

YOLT adopts four times of pooling operation, which makes the feature graph decrease continuously, and the size of the feature graph in the last layer is 26×26. The network design of YOLT restricts its detection performance of multi-scale targets such as large and small aircraft. Aiming at this problem, we propose an improved high-precision aircraft target detection method.

The network structure proposed in this paper is shown in Figure 3. The network includes 18 convolution layers and 3 pooling layers, and dilated convolution is used to lift receptive field and fusion features. Three pooling layers make the resolution of the final feature map 1/8 of the input image, which not only ensures a large receptive field, but also avoids too small resolution of the deep feature map, which affects the detection effect.

Figure 2. Sample images of RSAD from 3D simulation

Figure 3. The network structure of improved YOLT

3.3. Reception field promotion and feature fusion

For larger aircraft targets, the receptive field of the shallow layer network is relatively small, so it is difficult to locate and detect the aircraft as a whole in some cases. Compared with smaller aircraft targets,
with the deepening of the network, especially after the pooling operation, the resolution of the feature map gradually decreases and the pixel size of the target gradually decreases. If the pixel size of an aircraft target in the input image is 16×16, it will change to 2×2 after four times of pooling operation, which brings difficulties to the detection of small aircraft. In order to solve the detection problem of multi-scale aircraft and improve the detection performance of large and small aircraft, the dilated convolution method is introduced in this section. On the one hand, the receptive field of the second layer is enlarged and combined with the output of the 16th layer, so that the receptive field of the large aircraft target is improved while the features of the deep and shallow layers are fused. On the other hand, replacing ordinary convolution with dilated convolution in deep network can avoid reducing the image resolution, reduce the loss of effective information, and further improve the detection ability of small aircraft targets. The pyramid structure design often makes the detection network complex. The network design in this paper avoids this defect, and the network is more efficient and simple, and the performance is also significantly improved.

The receptive field after the void convolution operation can be calculated by the following formula:

\[ K = (k - 1) \times r + 1 \]  

(1)

Where \( k \) is the size of the convolution kernel, \( r \) is the expansion rate, and \( K \) is the size of the equivalent convolution kernel after the "dilate" operation. It can be seen that the receptive field can increase rapidly with an exponential trend due to dilated convolution.

3.4. Improved non-maximum suppression method

Non-maximum Suppression (NMS) was applied to remove redundant candidate boxes according to certain criteria to obtain the best detection box while there were a large number of candidate boxes. In the traditional NMS algorithm, the Intersection over Union (IoU) is taken as the evaluation index of the overlapping degree of different boxes. The disadvantage of traditional thinking is that if the target to be detected appears in the overlapping area, it will be ignored directly, resulting in detection failure and reducing detection rate. In this paper, the previous non-maximum suppression method is improved by introducing DIOU (Distance IOU)[11]. The distance, overlap rate and scale between the targets and the anchor frames are included in the distance measurement to make the border regression more rapid and accurate. Its calculation formula is shown in Equation (2).

\[ DIOU = IoU - \frac{\rho^2(b, b'^{*})}{\epsilon^2} \]

(2)

Where \( b, b'^{*} \) represents the center point of the prediction box and the marker box respectively, and \( \rho \) is the Euclidean distance of the two, and \( \epsilon \) represents the diagonal distance of the smallest closure region that can contain both the prediction box and the marker box. This calculation method does not roughly remove the detection box according to the overlap area, but takes into account the distance, overlap degree and target scale, which is more reasonable and effective for aircraft detection.

4. Experiment and Evaluation

4.1. Description of experimental data

In order to evaluate the detection methods and solve the problem of uneven quality of aircraft target dataset of remote sensing image, a remote sensing image aircraft target detection dataset RSAD is constructed in this paper. The dataset comes from two software, Google Earth Map and 91 Satellite Map. In this paper, a total of 400 remote sensing images with the size between 800 and 2000 are obtained, and the spatial resolution is between 0.2m and 0.2m. Through the data enhancement methods of random cutting, random rotation and random color transformation, 4480 images were selected for network training, and the annotation information was labeled with rectangular boxes. Some images of the dataset are shown in Figure 4.
4.2. Experimental setting and evaluation indexes
The environment of all experiments in this paper is 64-bit Linux operating system, equipped with GTX1080TI graphics card. During the training, ImageNet pre-training weight was used. The experiment adopts evaluation indexes commonly used in target detection tasks - precision and recall.

4.3. Experiment results and analysis
In this section, our method is used to conduct experiments on the RSAD dataset. Figure 5– Figure 8 show four groups of detection results, which are respectively the experimental results of large-scale targets, multi-scale targets, small-scale targets and targets containing interference factors, fully proving the effectiveness and robustness of our method.
Figure 7. Display of small-scale aircraft target detection results

Figure 8. Display of anti-interference detection results

Figure 5 ~ Figure 8 show the experimental results that the largest aircraft target size is about 300×300 pixels, and the smallest target size is about 5×5 pixels. Interference factors include occlusion, complex background, and similar color between target and background, etc. The experimental results show that the algorithm in this paper has excellent adaptability in multi-scale target detection, and has a good effect on small aircraft target detection. In the case of strong interference such as complex background, similar to the target texture and color, it can achieve a stable and robust detection effect, and has a good positioning ability for the close aligned targets. It is a very effective aircraft detection algorithm.

Then, the experimental performance of the algorithm in this paper is quantitatively analyzed, and compared with the current mainstream detection algorithms Faster R-CNN [12], Yolov3 [13] and the original YOLT algorithm. The comparison results are shown in Table 1. It can be seen from the comparison results of accuracy and recall rate in the table that our method has a great improvement in accuracy and recall rate compared with the current typical target detection network. Compared with YOLT method, we improved the network, especially solved the problem of multi-scale targets from the perspective of receptive field, which could better detect small targets and significantly increase the recall rate. In addition, the improved NMS method improves the accuracy of the results to some extent.

Table 1. Comparison of experiment results of different algorithms

| Method     | Precision | Recall |
|------------|-----------|--------|
| Faster R-CNN | 85.2%     | 70.3%  |
| YOLOv3     | 90.1%     | 86.8%  |
| YOLT       | 92.5%     | 85.0%  |
| Our method | 95.7%     | 91.3%  |

Finally, in order to further verify our method, experiments are carried out on other two typical remote sensing datasets and compared with YOLOV3 algorithm. NWPU-RESISC45 is a remote sensing image dataset with the largest number and category so far. UCMERCEDE_LANDUSE is one of the widely used datasets in the field of remote sensing, which is used in the classification tasks of various remote sensing images. The aircraft samples in these two datasets have sufficient diversity to be used to test the aircraft target detection algorithm. The experimental results are shown in Table 2 and Table 3. According to the table, our method has good performance on the two datasets, with high precision and recall rate, which further proves the effectiveness and completeness of our method.
Table 2. Comparison of detection results on NWPU-RESISC45

| Method   | Precision | Recall |
|----------|-----------|--------|
| YOLOv3   | 93.1%     | 86.5%  |
| Our method | 94.3%     | 89.2%  |

Table 3. Comparison of detection results on UCMerced_LandUse

| Method   | Precision | Recall |
|----------|-----------|--------|
| YOLOv3   | 92.6%     | 85.4%  |
| Our method | 95.2%     | 88.5%  |

5. Summary
Aiming at the problems and shortcomings of current aircraft detection methods in remote sensing images, this paper proposes a high-precision aircraft target detection method based on YOLT. Firstly, we construct a remote sensing dataset RSAD, which can be used for aircraft detection, and use various data enhancement techniques to enhance the dataset. Then, based on YOLT algorithm, the network structure is improved, especially for multi-scale target and small target detection. Finally, part of the experimental results is presented from four perspectives and quantitative comparative experiments are carried out from two perspectives. According to the experimental results, our method has higher precision and recall rate, and better performance.

References
[1] Yu F, Koltun V, Funkhouser T. (2017) Dilated residual networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii. pp. 472-480.
[2] Xu C, Duan H. (2010) Artificial bee colony (ABC) optimized edge potential function (EPF) approach to target recognition for low-altitude aircraft. Pattern Recognition Letters, 31(13): 1759-1772.
[3] Van Etten A. (2018) You Only Look Twice: Rapid Multi-Scale Object Detection In Satellite Imagery. https://arxiv.org/pdf/1805.09512.pdf.
[4] Zuo J, Xu G, Fu K, et al. (2018) Aircraft Type Recognition Based on Segmentation With Deep Convolutional Neural Networks. IEEE Geoscience and Remote Sensing Letters, 15(2): 282-286.
[5] Hu G, Yang Z, Han J, et al. (2018) Aircraft detection in remote sensing images based on saliency and convolution neural network. EURASIP Journal on Wireless Communications and Networking, 22(1): 1-16.
[6] Redmon J, Farhadi A. (2017)YOLO9000: Better, Faster, Stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii. pp. 7263-7271.
[7] Yu F, Koltun V, Funkhouser T. (2017) Dilated residual networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii. pp. 7263-7271.
[8] Tong X Y, Xia G S, Liu Q, et al. (2020) Land-over Classification with High-Resolution Remote Sensing Images Using Transferable Deep Models. Remote Sensing of Environment, 237-322.
[9] Wang H, Liang W, Shan G. (2019) An Efficient Method of Detection and Recognition in Remote Sensing Image Based on multi-angle Region of Interests. https://arxiv.org/ftp/arxiv/papers/1907/1907.09320.pdf.
[10] Pang J, Li C, Shi J, et al. (2019) R2-CNN: Fast Tiny Object Detection in Large-Scale Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing, 1-13.
[11] Zheng Z, Wang P, Liu W, et al. (2020) Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. In: AAAI Conference on Artificial Intelligence. Vancouver. pp: 300-315.
[12] Girshick R, Donahue J, Darrell T, et al. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus. pp. 580-587.
[13] Redmon J, Farhadi A. (2018) YOLOv3: An incremental improvement. https://arxiv.org/pdf/1804.02767.pdf.