Research on YOLOv3 target detection model in the field of remote sensing

Yao Yang¹, Cong Huang¹, Huajun Wang*, Jun Wan¹, ZhenHeng Wang¹ and Yu Ma¹
¹ Chengdu University of Technology, SiChuan Province, 610059, China
*Corresponding author’s e-mail: 2019020390@stu.cdut.edu.cn

Abstract: The target recognition problem of high-resolution remote sensing images is marine traffic monitoring, disaster reduction emergency search and rescue, unmanned autonomous systems (UAS) (such as unmanned aerial vehicles, unmanned vehicles, unmanned subsmeribles, unmanned surface craft and other autonomous robots) and other civilian applications. The core technology of the system is also the key technology of military automatic target recognition (ATR) systems such as military reconnaissance, precision guidance, and maritime monitoring. With the development of high-resolution earth observation systems, more and more industrial applications require the extraction of more valuable target details from high-resolution remote sensing images. Then, due to the complex background of high-resolution remote sensing images and the difference in image quality (factors such as size, resolution, noise, etc.), low target detection accuracy and slow speed are the main bottlenecks that affect the combination of artificial intelligence and remote sensing. In order to solve the technical bottleneck of target detection, this paper focuses on exploring the detection performance of the channel attention module SElayer and YOLOv3 model in the field of high-resolution remote sensing images, providing basic technical support for related research.

1. Introduction
In recent years, my country has made rapid progress in high-resolution remote sensing satellites, and it has become the world's most advanced. In January 2018, the number of satellites in orbit reached ten. The Gaojing-1 series satellites, which have been put into use in 2016, are the first 0.5-meter high-resolution commercial remote sensing satellites in space in my country. They can provide a wide range of more than 60km and a resolution better than full-color 0.5m or a multi-spectral 2m height. Resolution image. In December 2018, China Aerospace Science and Technology Corporation developed a high-resolution commercial remote sensing satellite with a resolution of 0.3 meters, comparable in performance to the WorldView-4 (best commercial remote sensing satellite in the United States).

High-resolution optical remote sensing images have the advantages of high accuracy, intuitiveness, less noise, less distortion, and immunity to electromagnetic interference, and have broad application prospects in many fields. Target detection is one of the important researches and application directions of various remote sensing images, including optical remote sensing images. The purpose is to automatically search for objects of interest in optical remote sensing images, calibrate their positions, and identify their specific categories. Higher resolution optical remote sensing images bring richer and more intuitive ground object information, making it possible for more demanding target detection applications, such as the detection of multiple targets and small targets in complex backgrounds. The target detection technology based on optical remote sensing images has great potential in many fields.
such as resources and environmental protection, emergency and disaster relief, shipping scheduling, urban management, agricultural and forestry production, national defense and military. For example, in urban planning and management, it can help to build roads, bridges, ports and other infrastructure more rationally by detecting traffic jams in the city, and to reasonably distribute industrial and living land; to detect civil ships on rivers and oceans, and to monitor ports and waterways. Fishery management, rescue and disaster relief bring more and more convenient possibilities; in the national defense and military, target detection technology based on optical remote sensing images has been put into use to detect enemy parties such as warships, combat vehicles, airport runways, military aircraft, and nuclear facilities. Important military targets are indispensable for dynamic monitoring and precise strikes against military operations.

The increasing number of remote sensing satellites launched by humans and the improvement of the resolution of optical remote sensing images have brought new challenges to remote sensing image processing technology. First of all, the amount of image data has increased sharply, which puts forward higher requirements for the processing speed of the target detection algorithm. At the same time, high-resolution optical remote sensing images have richer target color and texture information, more complex backgrounds, and interference from lighting, shadows, and more similar objects, which have an adverse effect on target detection. Optical remote sensing technology is developing rapidly, but the traditional optical remote sensing image processing technology is relatively insufficient in accuracy, efficiency and generalization ability, which also limits the mining and use of high-resolution optical remote sensing image information.

In recent years, deep learning technology has made natural image image processing technology an unprecedented leap. Affected by this, some researchers have begun to apply deep learning-based target detection methods to remote sensing image target detection tasks, and have made some progress, demonstrating its applicability and great potential. However, the current target detection technology has practical requirements: high accuracy, strong adaptability, good real-time performance, and general multi-target detection technology suitable for optical remote sensing images. There is still a significant gap. This paper mainly studies the application of deep learning technology in optical remote sensing images, and studies the high-precision multi-type target detection method that adapts to remote sensing images, which is of great significance to better realize the automation, intelligence, and rapid processing of optical remote sensing images.

2. Related work

In this article, we use the YOLOv3 algorithm as the benchmark experimental network. The YOLOv3 backbone network is Darknet-53, with a residual network module added, and three detectors for multi-scale detection of targets[2-3]. This article tests the experiment on the RSOD data set. The RSOD data set is used for target detection in the field of remote sensing images. The target types in the data set include airplanes, oil tanks, track and field fields, and overpasses. There are 446 aircraft images, a total of 4993 aircraft, 189 track and field images, a total of 191 track and field examples, 176 overpass images, 180 overpass examples, 165 oil tank images, and 1586 oil tank examples. On the basis of the basic model, use the YOLOv3 model to train and test on the RSOD data set.

2.1 Difficulties

At this stage, the excellent detection algorithms that can be realized on natural images are disrespectful in realizing remote sensing images, mainly because of the characteristics of remote sensing images themselves. Remote sensing image data sets usually have the following characteristics[1]:

1. Large changes in image size: Due to the many types of remote sensing image sources (Google Earth, etc.), the image size is extremely uneven, with a resolution as large as 10000*10000 and as small as 400*400.

2. The direction changes greatly: the remote sensing image is the aerial equipment overhead view, so the arbitrary orientation of the target in the image will bring more background interference to the algorithm.
3: Small targets are arranged densely and neatly: Most small targets will make it difficult to match the anchors of the anchor-based algorithm, and most of the prediction frames will be filtered out by NMS after the dense, neat and directional arrangement is predicted.

4: Large changes in target scale: factors such as the height of aerial photography and large changes in the target's own scale will cause this problem.

5: Noise interference is large: It is greatly affected by weather factors (wind, rain, clouds, etc.), and the target that needs to be carefully observed by the naked eye can greatly increase the difficulty of the computer.

6: Complex background: the scene is complex and difficult to distinguish.

2.2 Target detection based on deep learning
There are two classification criteria for target detection methods based on deep learning. They are divided into two-stage detection and single-stage detection according to whether the region of interest needs to be extracted. According to whether the anchor point frame needs to be preset, it is divided into anchor point frame-based detection and Detection based on key points, detection based on key points is also called Anchor Free detection.

In recent years, methods based on deep learning have become a new hot spot and trend in the field of remote sensing image target detection, and have achieved good results. In addition, with the rapid development of remote sensing technology and Internet technology, the available high-resolution remote sensing image data has rapidly increased in recent years, and the original lack of samples for deep learning model training has gradually alleviated[4-5].

2.3 Improvement
The convolution operation is the core calculation of the target detection network. It combines the features of the local area to improve the receptive field, including the feature fusion of space and channel. For the part with more channels, this article introduces the SELayer module to weigh the importance of each channel. The structure of SELayer is shown in Figure 1 below:

![Figure 1. SELayer structure diagram](image)

The first fully connected layer (FC) will reduce the number of input channels to one-sixteenth of the original number. When passing through the second fully connected layer, it will be expanded to the number of input channels. This has two effects: 1) Increase non-linear components, fully explore the correlation between channels; 2) Reduce a large number of parameters, and the module only needs a small amount of computing resources.
3. Experiment

3.1 Lab environment
The hardware and software environment of this experiment is Ubuntu20.04, python=3.6, RTX 2080TI, CUDA=11.2.

3.2 Hyperparameter selection
After experimental testing, the following hyperparameters are selected:

| Hyperparameter    | Parameter value | Description               |
|-------------------|-----------------|---------------------------|
| epochs            | 100             | Training epochs.          |
| batch-size        | 8               | Training batch samples.   |
| img-size          | 640*640         | Image input size.         |
| optimizer         | Adam            | Optimizer.                |
| learning-rate     | 0.01            | Learning rate.            |
| conf_thres        | 0.5             | Object confidence threshold. |
| nms_thres         | 0.5             | iou threshold.            |

3.3 Experiment and analysis

| model     | categories (ap, %) |         |         |         |         |
|-----------|--------------------|---------|---------|---------|---------|
|           | aircraft           | oiltank | overpass| playground| map     |
| YOLOv3    | 57.80              | 89.20   | 34.40   | 91.44    | 68.21   |
| YOLOv3-SE | 52.16              | 90.27   | 45.09   | 96.16    | 70.92   |
Figure 2. YOLOv3-SE detection diagram

The map is a measure of the detection accuracy. As shown in Table 2, the detection accuracy of the oil tank, overpass, and playground categories of the YOLOv3 model integrated with the Selayer module has been improved to a certain extent, and the maps of the four categories have been improved by 2.71%. The reasons for the low accuracy of overpass detection are mostly due to the diverse forms of overpasses, serious background interference such as road flower beds, and the difficulty of distinguishing the boundaries of overpasses, and it is difficult for the detector to learn such targets. The aircraft fuel tank target is clearly different from the surrounding environment, the background is relatively clean, the number of samples is large, and the detector learning speed is fast.

The YOLOv3-SE model detection diagram can be seen in Figure 2. Figure 2 lists representative test results pictures of four categories. In the airport pictures, it can be seen that most of the aircrafts of different scales have been tested, and the detection effect of densely arranged oil tanks is also very good. Most of the targets in the figure are detected. The only drawback is that the detection frame of the playground and the overpass is slightly deviated from the actual situation.

4. Conclusion

This paper integrates the SELayer module into the YOLOv3 model, which effectively improves the accuracy of model detection and improves the performance of the model. However, during the detection, it was also found that some class detection frames were slightly deviated from the actual frames (the prediction frame was larger), and further optimization was needed in future work.
References

[1] Nie Guang-Tao, Huang Hua. A survey of object detection in optical remote sensing images. Acta Automatica Sinica, 2021, 47(x): 1-19 doi: 10.16383/j.aas.c200596.

[2] J.-H. Won, D.-H. Lee, K.-M. Lee and C.-H. Lin, "An Improved YOLOv3-based Neural Network for De-identification Technology," 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), 2019, pp. 1-2, doi: 10.1109/ITC-CSCC.2019.8793382.

[3] J. Wang, W. Xiao and T. Ni, "Efficient Object Detection Method Based on Improved YOLOv3 Network for Remote Sensing Images," 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD), 2020, pp. 242-246, doi: 10.1109/ICAIBD49809.2020.9137452.

[4] Z. Deng, L. Lei, H. Sun, H. Zou, S. Zhou and J. Zhao, "An enhanced deep convolutional neural network for densely packed objects detection in remote sensing images," 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), 2017, pp. 1-4, doi: 10.1109/RSIP.2017.7958800.

[5] P. Lv, Y. Zhong, J. Zhao, A. Ma and L. Zhang, "Change detection based on structural conditional random field framework for high spatial resolution remote sensing imagery," 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017, pp. 1059-1062, doi: 10.1109/IGARSS.2017.8127138.