Research on ship detection of optical remote sensing image based on Yolo V5

X M Sun and Y J Zhang*, H Wang, Y X Du
School of Computer Science and Software Engineering, University of Science and Technology, Liaoning, China
Corresponding author. E-mail: zhangyujun@ustl.edu.cn

Abstract. In order to achieve ship target detection based on remote sensing images in complex backgrounds, this paper proposes a ship target detection algorithm based on YOLO V5. First, the improved Kmeans is used to cluster the data set, DIUO_loss is used for the candidate frame selection strategy, and finally the channel attention mechanism is introduced. The experimental results show that under the same conditions, the improved network has a slight decrease in detection speed compared with the original YOLOv5 network, and the effect of images with densely distributed small targets is significantly improved. On the same test set, the accuracy rate is reduced by 3.84%, the recall rate is increased by 32.73%, and mAP is increased by 22.48%.

1. Introduction
With the continuous development of satellite technology, it has become very common to use satellites to monitor the conditions of various sea areas in real time. The use of remote sensing images for ship target detection has extremely broad application prospects in civil and military fields [1]. In civilian use, it can be used to find missing ships at sea; it can also be used to monitor marine fisheries; it can also monitor illegal ships, such as pirate ships, ferry boats, and so on. In military use, it can be used to detect the location of enemy ships, guide long-range missiles to accurately strike; it can also monitor the sea during wartime, illegal intrusion of warships, and so on. Due to the long-distance shooting, most of the objects to be detected in the obtained remote sensing images show the characteristics of small targets [2]. The small target in the image loses a lot of feature information after the model changes. Therefore, the recognition of small targets in remote sensing images is a hot topic with high research significance.

At present, the detection methods of ships in remote sensing images can be divided into two categories: traditional methods and deep learning-based methods. The current deep learning method is the mainstream method of optical remote sensing image detection. In the field of two-stage target detection algorithms, Gu Jiaojiao et al. used Faster RCNN to redesign the number and size of its anchor frames, which effectively improved the problem of easy repeated detection of objects to be detected in remote sensing images [3]. In the field of single-stage detection algorithms, Ma Junjie et al. used the YOLO network to establish the nonlinear observation equation of the remote sensing ship target, which was transformed into a linear equation, and then solved the specific position coordinates of the ship, which improved the positioning accuracy of the ship [4].

In summary, based on the YOLO v5s network, this article improves the YOLO V5s network to solve the problem of poor recognition of dense small targets in remote sensing images. First, use the improved Kmeans to cluster the used data set and improve the selection strategy of the target frame.
Introduce channel attention mechanism. The experimental results show that compared with the original YOLOv5 network, the improved YOLOv5s network has an improved effect in the task of remote sensing image ship target recognition.

2. Ship remote sensing detection model based on YOLO V5

2.1. Yolo V5 introduction
YOLO v5 currently has four versions: YOLO V5s, YOLO V5m, YOLO V5l, YOLO V5x. Among them, YOLO v5s is the smallest and fastest version of the model. This article uses the YOLO v5s version. YOLOv5 can be divided into 4 modules in structure: input terminal, backbone network, Neck and prediction terminal. The network structure is shown in Figure 1.

![YOLOv5 structure](image)

The image preprocessing stage of YOLOv5 introduces image preprocessing operations such as Mosaic data enhancement, adaptive image scaling, and adaptive anchor point calculation. This is conducive to the image proportion not deformed after the image is zoomed. The Backbone stage of YOLOv5 adopts the Focus and CSPDarknet53 structure, which reduces the amount of calculation for the model while ensuring accuracy. The Neck structure is mainly to generate feature pyramids. This structure generates pooled feature vectors of different fixed sizes, and also strengthens the expression ability of features, which is conducive to detecting targets with large changes in size.

2.2. Improved yolo v5 model

2.2.1. NMS improvements
Non-Maximum Suppression (NMS) is mainly used to eliminate redundant candidate frames. The target detection algorithm model will generate a lot of target frames during operation, and one target will be selected by many target frames. The role of NMS is to eliminate redundant target frames, so that a target retains only one target frame with higher confidence. However, for small targets that are too close to each other in the original NMS algorithm, the neighboring target frame may be eliminated. The original NMS algorithm will greatly reduce the detection effect of moored ships in the port terminal. This article introduces DIOU-NMS. Unlike the original NMS, DIOU-NMS not only considers the value of IoU, but also considers two boxes. The distance between the center points. A new formula is used to determine whether a box is deleted:

\[
S_i = \begin{cases} 
    s_i, & \text{IoU} - R_{DIOU}(M, B_i) < \varepsilon, \\
    0, & \text{IoU} - R_{DIOU}(M, B_i) \geq \varepsilon, 
\end{cases} \quad (1)
\]

\[
R_{DIOU} = \frac{\rho^2(b, b_{gt})}{c^2} \quad (2)
\]

Among them, \(R_{DIOU}\) is the distance between the center points of the two Boxes, Expressed by the above formula (2): Where \(p(.)\) is the distance, \(b\) and \(b_{gt}\) represent two boxes, and \(c\) is the diagonal length of the smallest box containing two boxes.
2.2.2. Channel attention mechanism

Convolutional neural networks use the idea of local receptive fields to fuse spatial information and channel information to extract image information features. The channel attention mechanism (Squeeze-and-Excitation (SE), SENet) improves its representation ability by modeling the dependence of each channel, and it can adjust the features channel by channel, so that the network can learn global information is used to selectively enhance features that contain useful information and suppress useless features[6].

![Diagram of SENet](image)

Figure 2. SENet is introduced to YOLO v5s structure diagram

The algorithm structure is composed of three parts: Squeeze, Excitation and Reweight, and explicitly constructs the interdependence between feature channels [7]. Specifically, it uses a method of recalibrating feature channels. In the process of learning features, the network learns the contribution value of each channel to the overall feature at the same time, and then according to this contribution value to improve useful features and suppress features that do not contribute much to the task [8].

3. Experiment

3.1. Data set

The data set of this experiment comes from the public remote sensing image data set and the image data intercepted from the Internet. Among them, 2432 pictures in the training set and 57 pictures in the validation set are selected from the public data set HRRSD. The test set selected the aerospace remote sensing target detection data set annotated by NWPU Northwestern Polytechnical University and some of the image data made by myself.

3.2. Experimental platform

The deep learning framework used in this experiment is the torch framework, and the operating platform is the ubuntu 18 system. The running memory is 16g, the video memory is 6g, the graphics card model is GTX1660ti, GPU accelerated environment 10.0, and the framework of the training model is pytorch.

3.3. Evaluation Index

Because the test set used in this article contains a part of data that only contains a single target. As a result, when the map value is calculated, the maximum accurate value corresponding to each recall stage is 1, so the effect of this experiment cannot be effectively evaluated. Therefore, the evaluation indicators of this experimental model mainly use precision (Precision, P), recall (Recal, R) and F-score, and the specific calculations are shown in equations (3), (4).

\[
\text{recall} = \frac{T_p}{T_p + F_n} \quad (3)
\]

\[
\text{precision} = \frac{T_p}{T_p + F_p} \quad (4)
\]

In the formula, TP means that the ship is correctly detected; TN means that the target is mistakenly detected as a ship; FP means that the target is not a ship, and the detection result is not a ship; FN means that the ship is mistakenly detected as another target. The mean precision AP is the area value...
of the curve enclosed by the precision rate $P$ and the recall rate $R$, and the mean precision mean mAP is the average of the mean precision values of all classes of learning.

3.4. Analysis of experimental results
After optimizing the loss function of YOLO V5 and the screening function of the prediction box, introducing a series of operations such as the channel attention mechanism, the improved detection result is shown in 11. After using CIOU_LOSS as the bounding box loss function of the predicted target and DIOU_NMS processing, it is obvious from Figure 11 that the ships in densely arranged ports and occluded locations can be effectively identified, and the confidence score has also been improved.

Figure 3: Improved YOLO V5 test results

There are a total of 30 pictures in the test set, which contain a total of 1437 boats. Using the improved YOLO V5 for detection, a total of 1 ships were detected, 1320 ships were correctly detected, 13 ships were falsely detected, and 122 ships were missed. They are in the same test set as Faster RCNN, SSD, YOLO v4, and YOLOv5s models. For experimental comparison, the detection performance of different models is shown in Table 1.

Table 1. Comparison of detection performance

| Network         | Precision rate | recall rate | mAP  |
|-----------------|----------------|-------------|------|
| Improved YOLO V5s | 89.34%         | 88.95%      | 92.6%|
| Original YOLO V5s | 93.19%         | 56.22%      | 70.12%|

It can be seen from Table 1 that the performance of the improved algorithm has been greatly improved compared with the original YOLOv5s, more small target detection samples have been detected, and the number of false detections has also increased. The recall rate increased by 32.73%, and the accuracy rate decreased by 3.84%. mAP increased by 22.48%. Compared with some other algorithms, there are similar changes. From this point of view, the performance of the improved model is significantly improved.

4. concluding remarks
Aiming at the problem of poor detection of dense small targets in remote sensing image ship target detection, this paper makes some improvements to the YOLO V5s network based on the original YOLO V5s network, and proposes a remote sensing image ship target recognition based on the improved YOLOv5 network. method. The experimental results show that the F-score of the improved target recognition network reaches 95.17%, and the accuracy and recall rate are significantly improved, and the deficiencies of the original model are improved. The next step is to improve the remote sensing image data set and improve the recognition under complex conditions. Research on accuracy and other aspects.

References
[1] Jin Wenchao. Optical remote sensing image ship target detection technology based on complex
background[D]. Harbin Engineering University, 2013.

[2] Guo Liqiao, Lin Jianyuan, Wang Wenying. The application of refined urban management based on high-scoring remote sensing data[J]. Urban Development Research, 2012, 19(11): 57-63+69.

[3] Gu Jiaojiao, Li Bingzhen, Liu Ke, et al. Infrared ship target detection algorithm based on improved Faster R-CNN [J]. Infrared Technology, 2021, 43 (2): 92-170-178.

[4] Ma Junjie, Huang Daqing, Liu Yaohui, et al. Ship target positioning based on improved ship yolo [J]. Electronic Measurement Technology, 2020, 43(5): 112-115.

[5] Lin Qingping, Zhang Qilin, Xiao Lei. Remote sensing image target recognition method using improved YOLOv5 network[J]. Journal of Air Force Early Warning Academy, 2021, 35(02): 117-120.

[6] Wang Dongfei. Application of Convolutional Neural Network Based on Channel Attention in Image Super-resolution Reconstruction[J]. Radio and Television Technology, 2018, 45(06): 63-66.

[7] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-e and excitation networks [EB/OL]. [2018-09-30]. https://ai-xiv.org/abs/1709.01507.

[8] Li Kewen, Li Xinyu. An improved Faster R-CNN pedestrian detection model based on SENet[J]. Computer Systems & Applications, 2020, 29(04): 266-271.