A HED-optimized Automatic Detection and Tracking Algorithm for Marine Moving Targets based on YOLO V3

Zhengye Wang[1,3] and Yongbin Zhou[2]

[1]College of Intelligence Science and Technology, National University of Defense Technology, Changsha, Hunan, 410000, P.R. China

[2]College of Intelligence Science and Technology, National University of Defense Technology, Changsha, Hunan, 410000, P.R. China

[3]Corresponding author's e-mail: 2507857020@qq.com

Abstract. Target detection technology is one of the core technologies of modern military information technology research. It has important theoretical research and application value in the fields of military such as intelligence collection, key area monitoring and weapon guidance. At present, the traditional maritime target recognition method has poor timeliness, low precision, and poor self-processing capability. This paper proposes an autonomous detection and tracking algorithm for marine moving targets based on YOLO V3 algorithm optimized by HED. The method realizes the autonomous detection and tracking of marine moving targets through the optimization of YOLO V3 deep learning network and HED edge detection algorithm. This method has two advantages: (1) fast and accurate detection of maritime targets through the YOLO V3 deep learning network; (2) detection of target edges by HED edge detection algorithm to achieve the target and background segmentation, improving the accuracy of detection.

1. Introduction
Ships are the main tool for human to explore the ocean, and are also important targets of marine target detection. In the modern war environment, the detection of marine ship targets has broad application prospects in military fields such as investigation and warning, command of the battlefield and precision guidance of weapons. With the improvement of commercial small satellite constellation and imaging technology, traditional artificial visual processing methods have been difficult to deal with massive amounts of image timely and precisely. Therefore, the real-time and precise information processing has become the focus of the technology of marine target detection.

The technology of detection of marine targets has gradually become the key direction of government and military research in various countries, and also has become one of the important indicators for evaluating the development of country's marine equipment. Many scholars have conducted a lot of research on the detection and tracking technology of marine moving targets. In these studies, there are two main types of techniques for detecting marine targets: (1) The technology of marine target detection based on non-imaging technology mainly relies on active or passive means such as radar, sonar and other devices to acquire the radiation or echo signals of the targets, and then judge and analyze to realize the detection of the targets. The implementation of marine targets detection based on non-imaging technology is relatively simple, but it can only determine the orientation and distance of the targets. It cannot accurately identify the type of the targets or obtain the environmental information around the targets. (2) The technology of marine targets detection based on imaging technology is to process and analyze the image data of targets collected by various sensors, and extract feature information of the targets to identify the type and location of the targets. It can accurately identify the type and the surrounding environment of the targets. But it's structure is complex, and it is susceptible by obstruction such as clouds. With the great development of technologies such as pattern recognition technology, imaging technology and communication technology and the global construction of small satellite constellation, the detection of marine moving targets based on imaging technology becomes more and more important. Although many methods have been proposed for the detection of maritime targets, because of the marine environment is relatively complicated, there exists a great number of false alarms.

In view of this application background, this paper proposes a technology of marine moving target detection and tracking, which is based on YOLO V3 and optimized by HED. First, we combine the original data set and the data extracted by the HED edge detection algorithm into a new dataset. Secondly, the data set obtained is calibrated for the application requirements of this paper. Lastly, YOLO V3 algorithm is trained on the prepared dataset to obtain a detection and tracking algorithm for the marine moving targets.

The basic structure of this paper is divided into the following five parts. The first part is a brief introduction to the technology of marine targets detection and tracking. The second part mainly introduces the current research status of the marine targets detection and tracking algorithm. The third part includes preparation of dataset, the basic principles of the YOLO V3 algorithm and the introduction of the basic principles of HED. The fourth part is the experimental results and analysis of the algorithm. The last part is a summary of the entire article.

2. Related work

Due to the importance of the target detection technology for marine targets, many scholars have made a lot of research on the detection and tracking technology of marine targets, and have obtained certain achievements.

2.1. Marine target detection technology based on non-imaging technology

Before the rapid development of pattern recognition technology, marine target detection technologies are more based on non-imaging technology. At present, marine targets detection technology based on non-imaging technology still occupies an important position. Marine targets detection based on non-imaging technology is generally divided into three steps: signal acquisition, signal analysis and feature extraction, target detection and classification, as shown in Figure 1. The key to the detection of the targets is the extraction of target features and the classification of targets.

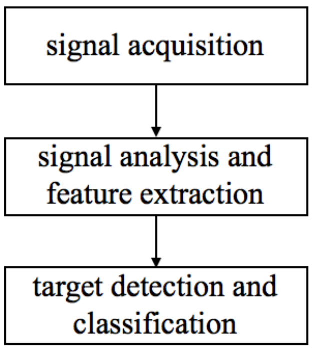


Figure 1. Basic steps for marine targets detection based on non-imaging technology

Typical methods for feature signal extraction include spectral analysis, time-domain autocorrelation, and time-frequency analysis. Spectral analysis is the most widely used method. The characteristic signal is modeled as a periodic signal. The spectral correlation coefficient in the acquired signal is used to analyze the parameters and identify the targets[1]. But its versatility is not strong. The time domain autocorrelation method is mainly applied to the analysis of radar signals, and its implementation is simple, but it cannot analyze non-linear FM signals[2]. The time-frequency analysis method mainly performs the Fourier transform on the acquired signal to establish the correspondence between the time domain and the frequency domain. It can effectively analyze the non-linear FM signal, which can deal with the target signal in more complicated cases[3].

The current target detection classifiers mainly include expert system classifier, support vector machine classifier and maximum likelihood classifier. The expert system classifier is a rule-based classifier whose implementation mechanism is to identify and classify target types through a rich professional knowledge base and logical reasoning. The support vector machine classifier is a statistical classification[4]. The implementation mechanism is to find an optimal classification plane in the complex sample model for target detection and classification. The implementation is simple and the speed of response is fast, and it is widely used in various types of target detection systems. The Maximum Likelihood Classifier is a classifier based on Bayesian theory. It is a supervised classification method that uses sufficient prior knowledge to classify objects.

2.2. Marine target detection technology based on imaging technology

With the great development of sensors, imaging technology continues to break through. Now, target detection based on imaging technology can not only determine the type and orientation of the target, but also obtain detailed environmental information of the target. Compared with non-imaging technology for marine targets detection, it processes and analyzes images of targets collected by various sensors. The type and position of the targets are identified based on the feature information obtained by analysing the images. The main steps of target detection based on imaging technology can be divided into three steps: image acquisition and processing, extraction of target feature, target detection and classification, as shown in Figure 2.

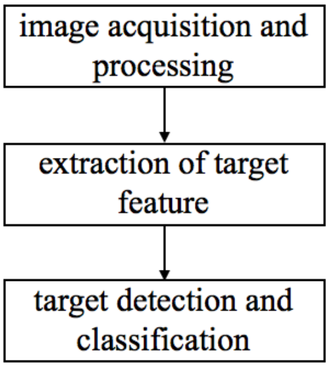


Figure 2. Basic steps for marine targets detection based on imaging technology

The development of imaging technology has also enriched imaging methods, mainly including optical imaging and radar imaging. The methods used in current optical imaging are mainly divided into visible light based imaging technology and infrared based imaging technology. The visible light imaging technology is simple to implement, and the target feature information is easier to extract and analyze. However, visible light imaging technology is susceptible to obstacles such as ocean weather and clouds. Compared with visible light imaging technology, infrared-based imaging technology has strong anti-interference ability and is not affected by cloud, waves and other obstructions, but its structure is complex, and it is greatly affected by temperature. Different from optical imaging, radar imaging technology realizes the estimation and judgment of the targets by analyzing and processing the amplitude, waveform, phase and other information of the echo signal received by the radar. Radar imaging technology is immune to weather and light, allowing real-time, sustainable, and uninterrupted monitor of the targets. The most widely used radar technology is Synthetic Aperture Radar (SAR).

For the identification of marine moving targets, the main identifying features include the following two aspects: geometric features and motion features. The geometric characteristics of the target mainly refer to the size, shape, color, etc. The motion characteristics of the target are mainly to identify the heading, speed, predicted route, etc. It helps make estimation and judgment on the subsequent actions of the target. The feature extraction algorithms for target geometric features generally include feature extraction algorithm based on image morphology, feature extraction algorithm based on active appearance model and feature extraction algorithm based on structural features. The feature extraction algorithms for motion features generally include Radon-changing feature extraction algorithm, Hough-based feature extraction algorithm and hybrid Gaussian model feature extraction algorithm[5].

After the image feature is extracted, the targets are classified according to the features. The current main classification methods are matching based on models and matching based on feature elements. Matching according to the feature model has been widely applied in the fields of fingerprint recognition, face recognition, and unmanned driving, and its real-time performance is good, but the robustness and anti-interference ability are poor. Compared with the feature recognition based on the feature model, the matching is more robust and adaptable according to the feature elements. These feature elements include the point, line or area of the target, which is suitable for a noisy detection environment.

Although great progress has been made in the detection and tracking methods for marine moving targets, these algorithms have a large number of missed and false detected situations under actual conditions. Therefore, this paper proposes a YOLO V3 algorithm optimized by HED to realize the detection and tracking of marine moving targets.

3. Method

3.1. Dataset preparation

This paper is mainly aimed at detecting and tracking marine moving targets. There is currently no specific dataset designed for this aspect. This paper combines online search with some suitable data in existing datasets. The basic examples of image are shown in Figure 3, which constitute the basic dataset of the algorithm. This paper mainly prepares the dataset through the space-based marine target image. Since the dataset is not rich enough, the dataset is enriched by rotation, cutting, etc. Finally, the number of images in the dataset is 1000, and 80% of the dataset was used as the training dataset and 20% was used as the testing dataset.

Figure 3. Some image data examples in the dataset.

3.2. Design of marine moving targets detection and tracking algorithm

The marine targets detection and tracking algorithm in this paper is mainly based on YOLO V3 algorithm and HED. Among them, YOLO V3 algorithm is the main algorithm to realize detection and tracking. HED is aimed to optimize the input image quality and improve the accuracy of algorithm.

3.2.1. The basic principle of YOLO V3 algorithm. Since the born of YOLO[6], it has been tagged with two labels: fast and not good at detecting small objects. But YOLO V3[7] has a significant improvement in detecting small objects and has a good ability to generalize the model. Compared with the predecessors, YOLO is very fast, 1000 times faster than R-CNN and 100 times faster than Fast R-CNN. YOLO V3 detection accuracy is higher than DSSD, close to FPN, but the detection time is only less than one-third of the latter two. YOLO V3 adds the multi-level prediction of top-down, which solves the problem that YOLO are weak at detecting small targets. V2[8] has only one detection, V3 becomes three, which are a subsampled whose feature map is 13*13 and two upsampling whose feature maps are 26*26, 52*52. Which means v3 has used 52 feature maps, while V2 uses only 13 feature maps, which should be the most influential on small targets. What's more, V3 replaces V2's softmax loss into logistic loss. When the predicted target categories are complex, it is more efficient to use logistic regression for classification, such as classification in the Open Images Dataset dataset. In this dataset, there are a lot of overlapping tags, such as women and people. If you use softmax, it means that each candidate box only corresponds to one category, but it is not always like this. The composite label approach enables better modeling of the data.

3.2.2. Basic principle of HED algorithm. The HED network is modified based on the VGG (Visual Geometry Group) network. With the powerful feature extraction capability of the VGG network, the multi-level and multi-scale features in the image are extracted. The HED network is designed with five side output layers that are output from the VGG16 network's conv1-2, conv2-2, conv3-3, conv4-3, and conv5-3, and remove all fully connected layers of the VGG network. The receptive fields of these output layers are 5 × 5, 14 × 14, 40 × 40, 92 × 92 and 196 × 196, respectively. The HED network deconvolutes the high-level feature maps, expands the size to the same as the original image, and fuses the feature maps through the convolutional network. HED utilizes the accurate location information of the low-level network and the rich semantic information of the high-level network to get more accurate test results. Compared with the VGG network, the HED network removes the fully connected layer of VGG and implements training and detection for input pictures of any size, while reducing the number of parameters and network complexity of the entire network. At the same time, during the training process, the HED network proposed a different training method than the previous network. The HED network improves the cost function of training, and the training steps are streamlined by the simultaneous training of the side output layer and the fusion layer. Meanwhile, due to the huge difference between the number of contour lines and non-contour lines in edge extraction, using the normal loss function for training will cause model errors and affect the training results. Therefore, the HED network amplifies the coefficient of the edge position loss function. Finally, the HED network iteratively optimizes the parameters and fusion weights in the convolutional network kernel according to the sum of the fusion cost and the side cost. The HED network has proven to be a good performing deep learning network for edge detection, which can provide better target feature information for target detection in this paper.

3.2.3. Implementation of sea moving target detection and tracking algorithm. The main tasks of the algorithm in this paper are detection and tracking. The detection task is mainly to detect potential moving targets at the picture, solving the problem of the presence or absence of the marine moving target in the surveillance image. The tracking task is mainly to predict the heading of the target. In the study of this paper, the heading prediction is weakened. Accurate heading predictions need to predict the heading of the target in 360 degrees, since the calibration of the heading cannot be too refined. The amount of labels required for 360 degrees calibration is too large and because the amount of data is too small, the amount of training data per label is too small. Therefore, in this paper, the heading prediction is weakened into heading prediction in eight directions. The heading is mainly divided into up, down, left, right, left-up, and left-down, right-up and right-down. The data set is calibrated for these two requirements. Then the YOLO V3 algorithm and HED algorithm are trained and tested on the data set to obtain an algorithm suitable for marine moving targets detection and tracking. The basic flow of the algorithm is shown in Fig 4. The algorithm combines the high speed and high precision of YOLO V3 algorithm with the advantages of HED algorithm, which can suppress the background and highlight the target, and realize the accurate and rapid detection and tracking of the marine targets.

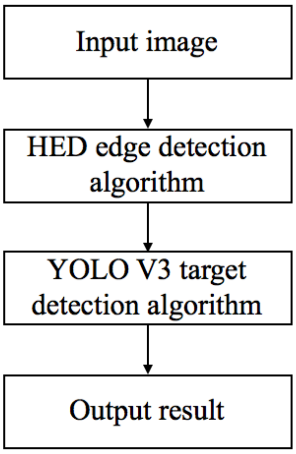


Figure 4. The basic flow of the algorithm

4. Experimental results and analysis.

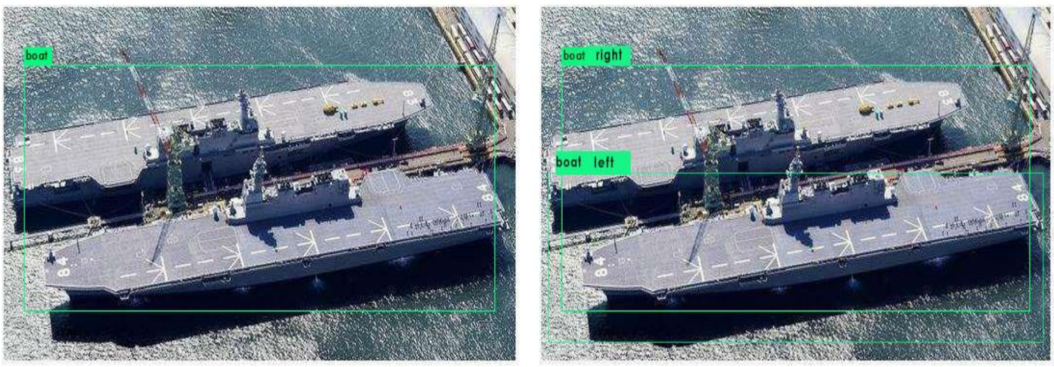
In order to verify the effectiveness of the proposed method, model training and algorithm testing are performed on the PC with GTX1060. In this paper, the effectiveness of the detection algorithm and optimization algorithm is verified by comparing the effects of without HED optimized YOLO V3 detection tracking algorithm and HED optimized algorithm.

The algorithm test was carried out on the testing dataset. Before the HED optimization, the mAP value of the marine moving target detection and tracking algorithm based on YOLO V3 is 53.48%, and the detection speed of each picture was 200 ms. After the HED algorithm optimized, the mAP value is increased by 6 percentage points to 59.58%. Due to the adding of the HED algorithm, the detection speed increases to 500ms per picture, but it still meets the real-time requirements.

When the image is clear, the background environment is simple, and the target contour is obvious, as shown in Figure 5. The experimental output is very satisfactory, and the target is accurately detected and the predicted heading is left.

Figure 5. Detection result of case one

However, when the difference between the targets in the image to be tested is not obvious and the directions of the two targets are inconsistent, as shown in Figure 6, the output of the algorithm may be a target, and the missed detection situation occurs. Moreover, the heading cannot be predicted. After the HED algorithm optimized, the output gets much better. The two ships are detected and identified separately, and the difference is obvious. No target missed detection occurs. The optimized algorithm accurately determines the heading of the two ships.

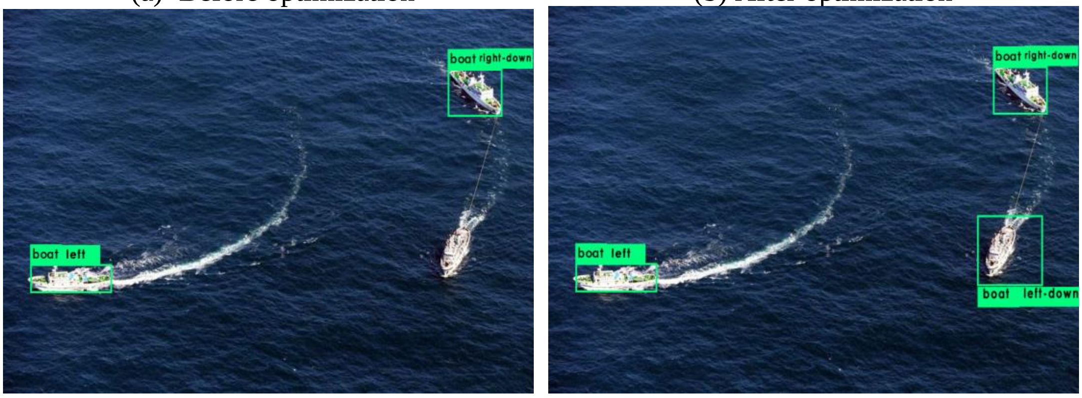


(a) Before optimization (b) After optimization

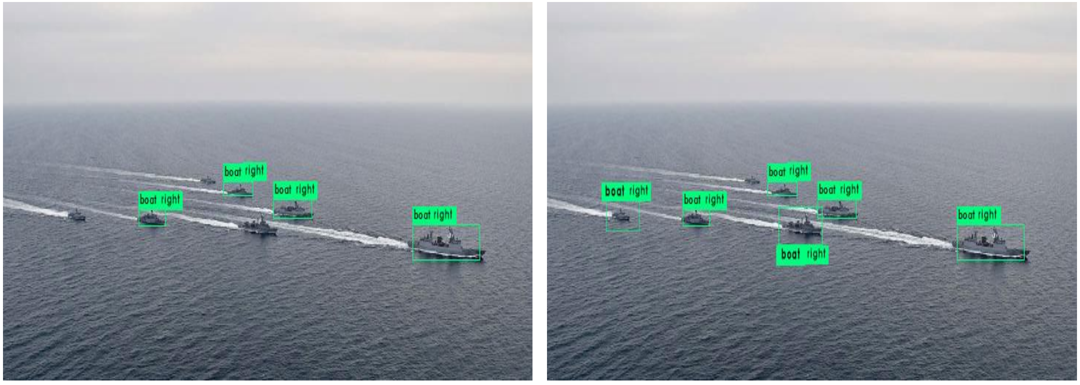
Figure 6. Detection result of case two.

Because the marine targets are in the complex environment of the sea surface, the distinction between the background and the targets is not obvious, which causes a lot of missed detection. After the optimization of the HED edge detection algorithm, the detection effect is greatly improved. In some images, some small targets can be detected after optimization, as shown in Figure 9.

(a) Before optimization (b) After optimization



(c) Before optimization (d) After optimization



(e) Before optimization (f) After optimization

Figure 7. Detection result of case three.

5. Conclusion

YOLO V3 has the advantages of high precision and fast detection speed in target detection. The HED network can output a significant abstracted edge image of the target. By combining the two methods, the feature information of the image can be fully extracted, and the accuracy and robustness of the detection are improved. The experimental results show that the proposed method can effectively suppress the background interference, obtain accurate edge images, improve the effectiveness of marine moving target detection and lay a good foundation for the construction of marine moving target detection and tracking system.

References

[1] Sheng Huiping, Lin Wangpeng, Zhou Yong. Overview of Research and Development of Signal Processing in Ship Target Recognition[J]. Ship Electronic Engineering, 2013, 12(33): 3-6.

[2] Zou Hongxing, Zhou Xiaobo, Li Yanda. Time-Frequency Analysis: Retrospect and Prospect[J]. Chinese Journal of Electronics, 2000, 28(9): 78-84.

[3] XIANG Qiang, QIN Kai-Yu. Time-Frequency Analysis Method Based on Combination of Linear Regular Transformation and Short Time Fourier Transform[J]. Chinese Journal of Electronics, 2011, 39(7): 1508-1513.

[4] Peng Wei. Research and Application of Support Vector Machine Classification Algorithm [D]. Changsha: College of Electrical and Information Engineering, Hunan University, 2007.

[5] Laxhammar,R.,Falkman,G.,*et al*.(2009)Anomaly Detection in Sea Traffic-A Comparison of the Gaussian Mixture Model and the Kernel Density Eastimator. 12*th International Conference on Information Fusion*, Seattle, 6-9 July 2009, 756-763.

[6] You only look once:unified,real-time object detection. Redmon J,Divvala S,Girshick R,et al. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition . 2016

[7] YOLOv3:An incremental improvement. J. Redmon,A. Farhadi. https://pjreddie.com/media/files/papers/YOLOv3.pdf .

[8] YOLO9000:Better,Faster,Stronger. Redmon J,Farhadi A. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . 2017

[9] XIE S, TU Z. Holistically-Nested Edge Detection[J]. International Journal of Computer Vision, 2015: 1-16.