Advances in Computer Science Research, volume 88

International Conference on Computer, Network, Communication and Information Systems (CNCI 2019)

A Waterway Monitoring Method of Unmanned Surface Vehicle Based on Deep Learning

Yiheng Wu[1], Xing Li[1], Zhangjie Yin[1], Jian Li[1] and Yan Zhou[1]

[1]School of Computer Science, Hubei University of Technology, Hubei, China

*514773500@qq.com

Abstract. A waterway monitoring method of unmanned surface vehicle based on deep learning is proposed to help identify ship accidents in real time. Compared with the traditional waterway monitoring method, the use of unmanned surface vehicle is more flexible and the recognition method of deep learning is more real-time. First, we build a dataset of ship accidents, and then use this dataset to train the Faster R-CNN network, we optimize the RPN in the fast R-CNN framework for the problem of missed detection at the same time, and finally obtain the target detection model. Experiments show that the method can effectively improve the efficiency of the waterway monitoring, greatly reduce the labour cost, and facilitate the management personnel to grasp the waterway situation in real time.

Keywords: Waterway Monitoring Method, RPN network.

1. Introduction

The waterway is the basis of water transport and the key factor for shipping development. Waterway monitoring has a great influence in maintenance of water safety, water traffic order, and handling water accidents in time.

With the development of unmanned technology in recent years, unmanned surface vehicle has become more and more widely used in various fields. Compared with high-altitude remote sensing technology, unmanned surface vehicle is an unmanned device with flexible operation, free shooting time. Therefore, the application of unmanned surface vehicle to monitor waterway has become a new research direction.

At present, many research institutes and universities have conducted in-depth research on navigation channel monitoring and put forward many effective solutions [1, 2, 3]. The paper [2] proposes an acoustic signal collected from a passing ship through a fiber optic hydrophone, combined with energy threshold and characteristic strong line spectrum, to achieve monitoring of the number of passing ships in the channel.

Traditional monitoring of the waterway is mainly relying on passive monitoring system. Only when the waterway management personnel get more accurate information about ship accidents they can retrieve the required parts from the numerous monitoring videos. This manpower-dependent monitoring program does not detect water traffic accidents in a timely manner which causes the abnormality of the channel cannot be processed in the fastest way. More importantly, traditional monitoring equipment is difficult to use on the waterway due to power and installation restrictions.

In recent years, deep learning has developed rapidly [4, 5]. Its advantage is that it does not need to artificially specify the feature values that need to be extracted. Instead, it uses big data to train the model to learn the characteristics of the target itself [6].In this paper, the passive monitoring system is replaced by active acquisition of channel information, using real-time information to classify incidents, which has good application value.

2. Date Set Establishment

Establishing a data set plays an important role in deep learning model, the data set must accurately reflect the specific classification. There are many reasons for ship accidents on the water. It is not only the collision between ships, besides there are also reef accidents at sea. The most common scene when a ship accident occurs is the overturning of the ship. Therefore, when we set up a data set that can reflect the abnormal events of the waterway, we collected ship capsizing photos for deep learning.

Fig. 1 Ship capsizing photos

3. Ship Identification of the Faster R-CNN Process

In the field of target detection, the CNN (Convolutional Neural Network) has excellent characteristics, and its weight-sharing network structure can reduce the complexity and weight of the network model. The advantage of CNN is that the recognition accuracy is high. However, a large number of convolution operations make its operation efficiency low and the real-time performance poor. As an improvement, Faster R-CNN introduces the concept of RPN (Region Proposal Network). This paper uses Faster R-CNN to train the object detection model.

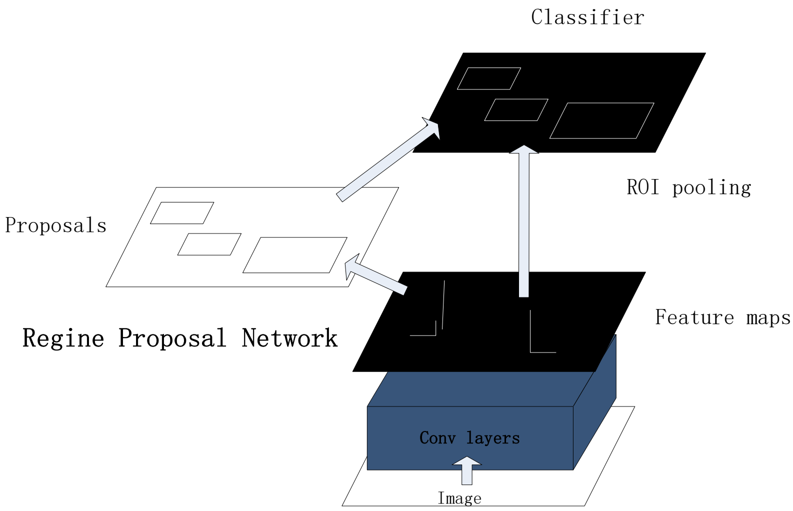

Fig. 3 Network model of RPN

After inputting the sample image and the ground truth into the Faster R-CNN network, conv layers extracts the target features in the marked frame to obtain the feature map, The RPN generates the candidate region and its internal feature information on the feature map, and then uses the classifier to identify the target class in the candidate frame with the regression device to correct the candidate frame position and finally outputs the detection result.

3.1 Region Proposal Network

The RPN takes an arbitrary-scale picture as input and output is series of rectangular object proposals, each with an objectless score. The RPN network first generates a sliding window on the convolution feature of the conv layers output, maps the window which is fully connected to the cls layer and reg layer a low-dimensional vector. For each window position, there are 9 initial anchor boxes, each anchor box containing four parameters (x, y, w, h) that determine the position and size of the suggestion box.

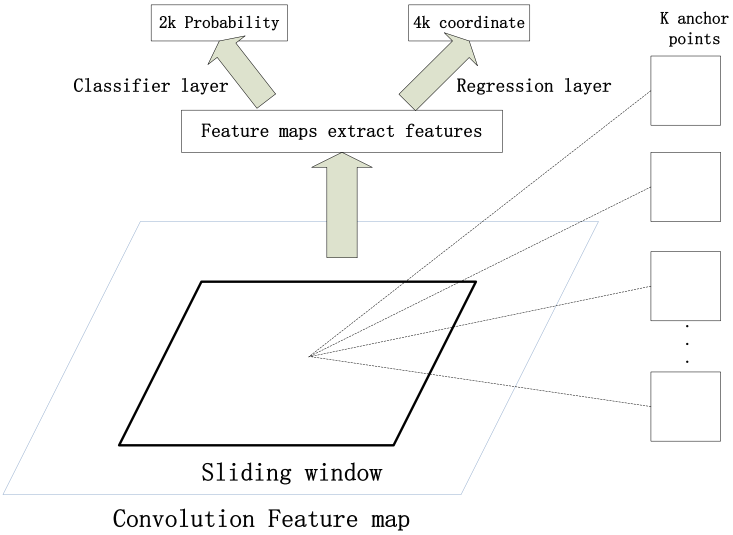

Fig. 2 Framework of faster R-CNN model.

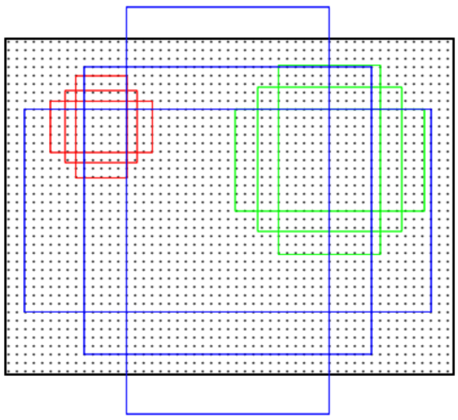

Fig. 4 Anchor boxes

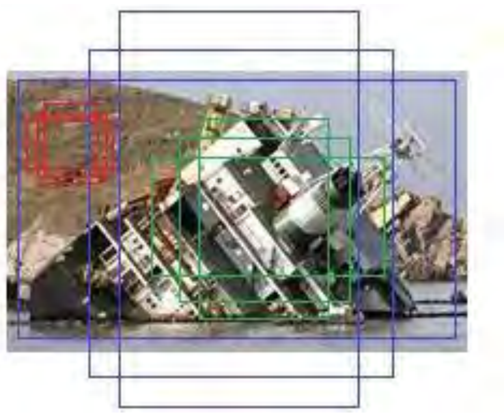

Fig. 5 Anchor boxes on an image

In RPN, the loss function is the combined loss of classification loss and regression loss.

$$L(\{p_i\}+\{t_i\})=\frac{1}{N_{cls}}\times\sum_i L_{cls}\left(p_i,p_i^*\right)$$
$$+\frac{\lambda}{N_{reg}}\times\sum_i p_i^*\times L_{reg}\left(t_i,t_i^*\right) \tag{1}$$

Among them, i represents the i-th anchor point,p_i represents the probability that the anchor point i is the target, $p_i^*=1$ indicates that the i-th anchor point is a positive sample,$t_i$represents the four parameterized coordinate vectors of the predicted bounding box,t_i^*represents deviation between the candidate area border and the real target border, L_{cls} and L_{reg} represent for classification loss, N_{cls} and N_{reg} represent the normalized parameters of the classification loss and the regression loss, respectively, indicating the balance weight

$$L_{cls}(p_i,p_i^*)=-\log\left[p_i p_i^*+(1-p_i)(1-p_i^*)\right] \tag{2}$$

The four coordinates in the regression are as follows,

$$t_x=\frac{x-x_a}{w_a} \qquad t_y=\frac{y-y_a}{h_a} \qquad t_w=\log\left(\frac{w}{w_a}\right) \qquad t_h=\log\left(\frac{h}{h_a}\right) \tag{3}$$

$$t_x^*=\frac{x^*-x_a}{w_a} \qquad t_y^*=\frac{y^*-y_a}{h_a} \qquad t_w^*=\log\left(\frac{w^*}{w_a}\right) \qquad t_h^*=\log\left(\frac{h^*}{h_a}\right) \tag{4}$$

x, y, w, h represents the bounding box center coordinates, width, height; variables x, x_a , x* represent the predicted rectangle, the rectangle of the anchor, and the ground truth rectangle.

3.2 Improvements for RPN

The ratio of positive and negative Anchors of RPN is 1:1 ideally, but in the case of a larger batch size, especially for images with a small number of targets, the number of positive anchors will be very small, the model thus trained will be biased towards the background category and objects will be easily missed. A limit has been made on this ratio to balance the positive and negative anchor ratios. The number of negative samples not exceed 1.2 times that of the positive samples.

4. Test Results and Analysis

When training the target detection network, the Fedora 28 operating system was used and the training process was accelerated using the NVidia GeForce GTX 1080 graphics card. The deep learning framework that the Faster R-CNN model relies on is Tensorflow. The training phase is optimized by a random gradient descent algorithm with a momentum of 0.85, the initial learning rate is set to 0.002, and the regularization coefficient is set to 0.000 04. The learning rate per 25,000 times dropped by 0.1 times. The data expansion methods used in the training process is random horizontal flipping and random image cropping. After the training is completed, the model is streamlined and standardized, and then deployed to the embedded development platform RK3399 for testing.

This experiment uses the ship accident rescue video data in recent years, and the frames containing the accident ship in these videos are used as test data sets. The results show that within a certain distance range, the detection effect is better, and it can be recognized by certain occlusion. The overall recognition rate is above 90%. When transmitting video, the average recognition time per frame is 0.032s, which enables real-time recognition.

Table 1 Video recognition result

Videos	Total number of frames	the number of recognized frames	Recognition rate	Average recognition time per frame
1	1450	1398	96.41%	0.045s
2	420	387	92.14%	0.038s
3	988	932	94.33%	0.042s

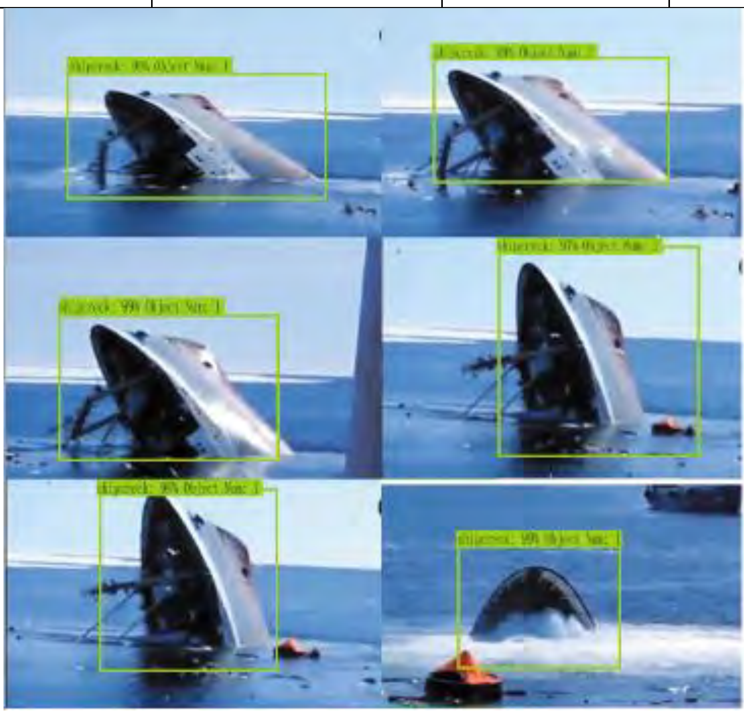

Fig. 6 Boxes of the gradually sinking ship

However, there are still some missed frames. When the target is small, the detection accuracy is greatly reduced. The main reason may be as follows, the accuracy of the model is not enough to extract higher-level features or the video has been enhanced which results in the features disappear.

5. Conclusion

Aiming at dependence on manpower and low efficiency in traditional channel monitoring, this paper proposes a real-time monitoring method for unmanned ships for navigation channels using deep learning technology. In this paper, the deep convolutional network is used to extract the characteristics of the abnormal vessel dumping vessel. The RPN network is used to generate the regional recommendations, and the deep convolutional network is used to identify the target. Finally, the target detection model based on the Faster R-CNN network is obtained. This method has the advantages of small footprint, good robustness and low cost. Simulation experiments prove the effectiveness of the method.

At present, this paper only conducts single-class detection on overturned vessels in ship accidents, and does not consider other dangerous situations on the channel. In the future, it will further explore how to comprehensively monitor the abnormal events of the channel.

Acknowledgments

This work is financially supported by National Key R&D Program of China(2017YFC1405403).

References

[1]. NIU Zuo-peng, LI Guo-jie, GUO Tao, Application of BIM technology in underwater terrain space monitoring, Port & Waterway Engineering, Oct. 2018.

[2]. Binyan Wu, Kan Gao, Jingfeng CHEN, Junping GENG, Xin LU, Dan CUI, Applications of Fiber Optic Hydrophone Array in Water-Way Monitoring of the Huangpu River, Optical Fiber & Electric Cable, 2015.No.2.

[3]. Zhijie Zhu, The Application Research on Drifting Sensor Networks Technology in Inner River Waterway Data Collection, Journal of Wuhan University of Technology (Transportation Science& Engineering),Apr.2015.

[4]. YANG Y, WANG T, CEHN L, et al. Stereo vision based obstacle avoidance strategy for quadcopter UAV[C]. 2018 Chinese Control and Decision Conference (CCDC), 2018.

[5]. HU Y, WANG Y X. Stereo vision-based fast obstacles avoidance without obstacles discrimination for indoor UAVs[C]. 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce, 2011: 4332-4337.

[6]. HINTON G E, SALAKHUTDINOV R R, Reducing the dimensionality of data with neural networks[J]. science, 2006, 313(5786): 504-507.