INSHORE SHIP DETECTION BASED ON MASK R-CNN

Shanlan Nie, Zhiguo Jiang, Haopeng Zhang , Bowen Cai, Yuan Yao*

Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China
Beijing Key Laboratory of Digital Media, Beijing 100191, China

ABSTRACT

Inshore ship detection is a popular research domain for optical remote sensing image understanding with many applications in harbor management. However, recent approaches on inshore ship detection depend heavily on hand-crafted features, which need a complicated procedure. In this paper, we propose a new method to achieve inshore ship detection based on Mask R-CNN. We introduce Soft-Non-Maximum Suppression (Soft-NMS) into our framework to improve the robustness to nearby inshore ships. Both battleships and merchantships can be detected in our framework. Furthermore, our framework can also obtain the binary masks of inshore ships. Experimental results on a dataset collected from Google Earth have quantitatively and qualitatively demonstrated the effectiveness of our approach.

Index Terms— object detection, inshore ship, remote sensing images, Mask R-CNN

1. INTRODUCTION

With the development of remote sensing technology, objects detection in remote sensing images becomes more prevailing. Among all the tasks from remote sensing images, ship detection is a hot researching topic because of its important role in harbor management [1]. Accurate inshore ship detection is a challenging problem, since it possesses three distinctive difficulties. Firstly, the ship targets have extremely long and thin shapes and arbitrary rotation. Secondly, the gray level and texture for both inshore ships and harbors are similar, which usually lead to many false alarms. Finally, inshore ships are often surrounded by docks and nearby ships.

In order to tackle the problems, many useful approaches have been proposed. Previous researches have emphasized the shape of ship targets, where features extracted are hand-crafted. [3] adopts invariant generalized hough transform (IGHT) to detect inshore ships. The essence of IGHT is a method of template matching; therefore, IGTH relies heavily on the consequence of target contour extraction. [4] improves the detection performance based on weighted pose voting, which makes use of radial gradient angle (RGA). Those approaches require complex procedures and cannot generalize well under different conditions. Recently, features extracted by deep convolutional neural networks (DCNNs) outperform hand-crafted features in a great many applications, thus making DCNNs achieve great success in natural image classification, object detection, image segmentation and some other related fields. With the expansion of remote sensing data, many researchers utilize DCNNs to tackle problems in remote sensing applications, such as marine ship detection [5], airplane detection [6], sea-land segmentation [7], etc. However, there are only few researches on detecting inshore ships using DCNNs. [8] introduces the rotated region based CNN to detect ships, which contains a rotated region of interest pooling layer and a rotated bounding box regression model. [9] utilizes DCNNs to sparse a remote sensing harbor image into three typical objects including sea, land and ship; therefore, it converts the detection task into image segmentation. [10] leverages a fully convolutional network (FCN) to solve the problem of inshore ship detection and classifies pixels into four classes, i.e. sea, land, ship body and foredeck/stern.

In this paper, we utilize instance segmentation model for inshore ship detection. Mask R-CNN [2] is regarded as our baseline model for its high performance in instance segmentation. Meanwhile we introduce Soft-NMS [11] into the Mask

(a) result of Mask R-CNN (b) result of our method

Fig. 1. Results of inshore ship detection. (a) is the result of Mask R-CNN [2], and (b) is our result. Both approaches can detect the inshore ship accurately. Meanwhile, our method is capable of seperating inshore ships which are close to each other.

*Corresponding to zhanghaopeng@buaa.edu.cn.

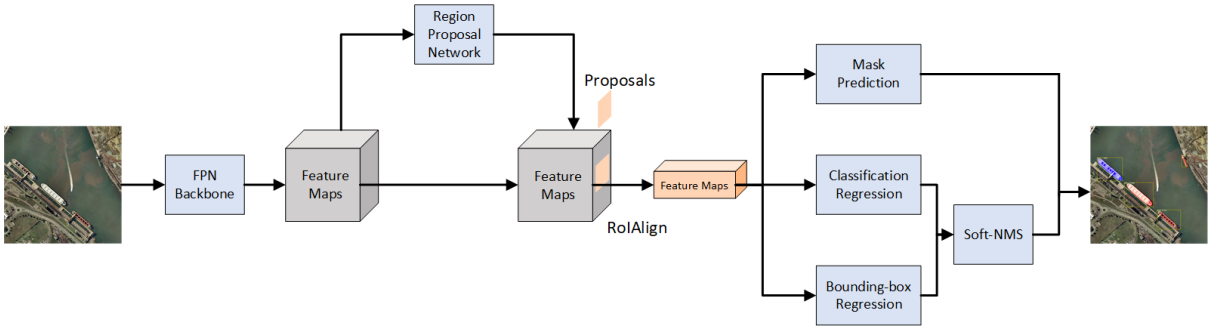

Fig. 2. The framework of our proposed method.

R-CNN to improve object detection performance. There are three advantages of our proposed method. Firstly, our approach can detect both battleships and merchantships in one framework. Secondly, our approach is more robust to those inshore ships, which are surrounded by each other. Thirdly, our approach is capable of achieving ship detection and ship segmentation in a single framework. Therefore, our method is able to obtain the masks of ships, which is useful for predicting the extra information of ships, like area, circumference, direction, etc.

The rest of this paper is organized as follows. In Section 2, we describe Mask R-CNN and the process of our method. Experimental results on a remote sensing dataset are introduced in Section 3. Section 4 concludes the paper.

2. METHODOLOGY

As shown in Fig. 2, our framework is based on Mask R-CNN, which includes two stage procedure. The first one is a Region Proposal Network. The second stage is made up of a Fast R-CNN classifier and a binary mask prediction branch. We replace Non-Maximum Suppression (NMS) by Soft-NMS, which performs as a post-processing step to obtain the final set of detections. In this section, we introduce our proposed detection framework in details.

2.1. Mask R-CNN Based Architecture

Mask R-CNN is a general framework for object instance segmentation, which can detect objects in an image accurately while generating a segmentation mask for each instance. Our method leverages Mask R-CNN as our baseline. Mask R-CNN adopts two stage procedure. The first stage is a Region Proposal Network (RPN) [12], which is used to propose candidate object bounding boxes. The second stage consists of a Fast R-CNN classifier [13] and a binary mask prediction branch. The second stage extracts features by using RoIAlign from each candidate box. Then the Fast R-CNN classifier performs classification and bounding box regression; meanwhile the mask prediction branch can output a binary mask for each candidate box. More technical details can be referred to [2].

Feature Pyramid Networks (FPN) [14] architecture plays an important role of feature extractor in Mask R-CNN. In our approach, we adopt ResNet-50-FPN to extract remote sensing image features. FPN utilizes a top-down architecture with lateral connections to build high-level semantic feature maps at all scales, which is suitable for detecting inshore ships with different sizes. FPN takes a remote sensing image as input and outputs a feature pyramid, which consists of five different scale features. Then, according to the anchors, RPN generates a batch of the region of interests (RoIs) in those different scale features. Subsequently, features of these RoIs are extracted by using RoIAlign. Then the features are fed into the Fast R-CNN classifier, which can finally produce softmax probability estimations about inshore ship and refine the bounding box positions for the ship targets. Meanwhile, the features are also fed into the mask prediction branch, which is made up of four convolution layers and one deconvolution layer, to predict the ship target mask.

2.2. Soft-NMS

NMS is an essential part of the object detection network, which performs as a post-processing step to obtain final detections. After Fast R-CNN classifier refines the region proposals, network could lead to cluttered detections since multiple region proposals often get regressed to the same region of interest (RoI). Therefore, NMS functions to obtain the final detections in most of state-of-the-art detectors [11].

However, the major problem with NMS is that it sets the score for neighboring region proposals to zero. An inshore ship sometimes is surrounded by other inshore ships, so the overlap of nearby inshore ships may be present in that overlap threshold, thus making a drop in average precision. Instead of setting the score for neighboring region proposals to zero as in NMS, Soft-NMS decreases the detection scores as an increasing function of overlap. A Gaussian penalty function is used in our framework as follows:

$$s_i = s_i \times e^{-\frac{(IoU(M,b_i))^2}{\sigma}}$$

where s_i denotes the score of detections, M presents the detection box with the maximum score, and b_i denotes the detection box in the remaining detection boxes, $IoU(*)$ calculates intersection-over-union between two detection boxes. In addition, we set $\sigma = 0.6$ in our experiments.

3. EXPERIMENTS

3.1. Dataset

To demonstrate the effectiveness of our proposed approach, we build an optical remote sensing dataset collected from Google Earth, which contains 36 images of 5000×5000 pixels with a spatial resolution of 1m/pixel. The dataset includes rich harbor areas and adequate land objects, thus is qualified to test our approach. It should be mentioned that our framework detects both battleships and merchantships which enlarges the challenge of detection methods. We select 29 images for training and the rest as the test data. In our training and testing phases, 1024 × 1024 image patches are sampled covering the 5000×5000 area, due to the limitation of GPU memories.

3.2. Qualitative and Quantitative Performance

To test our method comprehensively, we augment our test data via rotation, and obtain a test data set with 1250 targets in total. Fig. 3 displays a bunch of inshore ship detection results. Detection performance of our method is shown in Table 1. We use the precision, recall [10] and average precision (AP) [15] to quantitatively evaluate our framework on the test set. AP is equal to taking the area under the PR curve. Fig. 3 indicates our framework with Soft-NMS has the ability of detecting both battleships and merchantships. However, precision and recall of our method don't reach a high level due to the detection of merchantships, as shown in Table 1.

Table 1. Detection performance of our method.

Method	Precision	Recall	AP
Method in [2]	0.8065	0.8672	0.845
Our Method	0.7990	0.8776	0.852

Compared to Mask R-CNN, our method is more robust to nearby inshore ships. As shown in the first row of Fig. 3, inshore ships, which are close to each other, fail to be detected by Mask R-CNN. Since the inshore ships are too close, the IoU of their bounding boxes has reached the overlap threshold, thus making one of inshore ships suppressed. With the help of Soft-NMS, our method is able to overcome this drawback to some extent. It can be seen from Fig. 3 and Table 1 that our framework with Soft-NMS can detect more nearby inshore ships, thus making the recall and AP of our method increasing. It should be mentioned that our test dataset doesn't contain a large number of nearby inshore ships. Therefore, the results in Table 1 is similar.

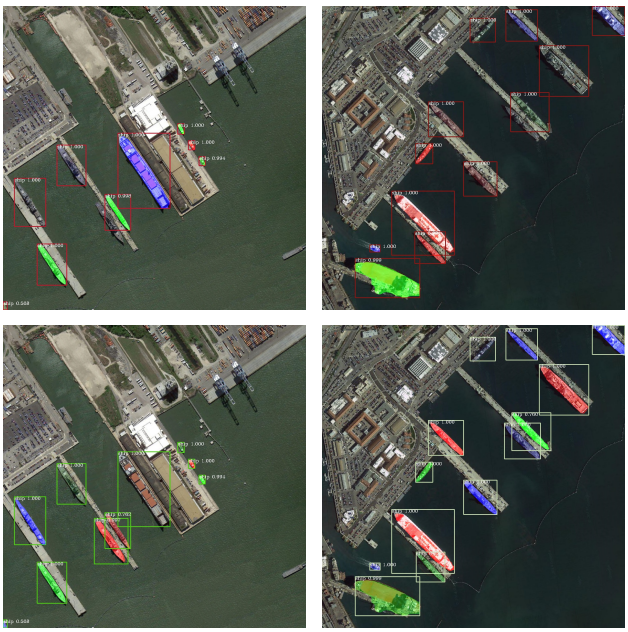

Fig. 3. Some examples of inshore ship detection results. Images in the first row are predicted with Mask R-CNN [2], and images in the second row are the result of our approach. Both approaches can obtain the bounding boxes, scores and masks for inshore ships.

Table 2. Segmentation performance of our method.

Method	AP	AP_{50}
Method in [2]	0.5807	0.8679
Our Method	0.5861	0.8774

Our method can not only obtain the classification score and bounding box of inshore ships, but also get the mask of inshore ships. Segmentation performance of our method is shown in Table 2. We leverage the standard COCO metrics to assess the segmentation, including AP, AP_{50} [16]. Here, AP_{50} means that the threshold of IoU is set as 0.50. AP represents that the threshold of IoU is set from 0.50 to 0.95, with a step of 0.05. Compared to battleships, which have comparatively same shapes, merchantships have various shapes and sizes. Hence the segmentation of merchantships can be more difficult. In our result, few merchantships aren't well-segmented, like the biggest merchantship in Fig. 3. However, most of battleships can get accurate binary masks, which guarantees that our method has a good performance in segmentation.

4. CONCLUSION

In this paper, we have proposed an Mask R-CNN based method for inshore ship detection. Soft-NMS is integrated in our framework for better detection. The proposed approach is evaluated on Google Earth images to show its capability of detecting both battleships and merchantships. With the help of Soft-NMS, our framework is more robust to nearby inshore ships. Meanwhile, our method has a good performance in segmentation of inshore ships.

Acknowledgment

This work was supported in part by the National Key Research and Development Program of China (2016YFB0501300, 2016YFB0501302), the National Natural Science Foundation of China (Grant Nos. 61501009, 61771031, 61471016 and 61371134) and Aerospace Science and Technology Innovation Fund of CASC.

5. REFERENCES

[1] C. Zhu, H. Zhou, R. Wang, and J. Guo, "A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 9, pp. 3446–3456, Sept 2010.

[2] K. He, G. Gkioxari, P. Dollr, and R. Girshick, "Mask r-cnn," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 2980–2988.

[3] J. Xu, X. Sun, D. Zhang, and K. Fu, "Automatic detection of inshore ships in high-resolution remote sensing images using robust invariant generalized hough transform," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 12, pp. 2070–2074, Dec 2014.

[4] H. He, Y. Lin, F. Chen, H. M. Tai, and Z. Yin, "Inshore ship detection in remote sensing images via weighted pose voting," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3091–3107, June 2017.

[5] Yuan Yao, Zhiguo Jiang, Haopeng Zhang, Danpei Zhao, and Bowen Cai, "Ship detection in optical remote sensing images based on deep convolutional neural networks," *Journal of Applied Remote Sensing*, vol. 11, pp. 11 – 11 – 12, 2017.

[6] X. Li and S. Wang, "Object detection using convolutional neural networks in a coarse-to-fine manner," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2037–2041, Nov 2017.

[7] R. Li, W. Liu, L. Yang, S. Sun, W. Hu, F. Zhang, and W. Li, "DeepUNet: A Deep Fully Convolutional Network for Pixel-level Sea-Land Segmentation," *ArXiv e-prints*, Sept. 2017.

[8] Zikun Liu, Jingao Hu, Lubin Weng, and Yiping Yang, "Rotated region based cnn for ship detection," in *2017 IEEE International Conference on Image Process ing. Piscataway, NJ: IEEE*, 2017.

[9] D. Cheng, G. Meng, S. Xiang, and C. Pan, "Fusionnet: Edge aware deep convolutional networks for semantic segmentation of remote sensing harbor images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 12, pp. 5769–5783, Dec 2017.

[10] H. Lin, Z. Shi, and Z. Zou, "Fully convolutional network with task partitioning for inshore ship detection in optical remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1665–1669, Oct 2017.

[11] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Soft-nms x2014; improving object detection with one line of code," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 5562–5570.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, June 2017.

[13] R. Girshick, "Fast r-cnn," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 1440–1448.

[14] T. Y. Lin, P. Dollr, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 936–944.

[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 580–587.

[16] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.