Automatic ship detection in SAR Image based on Multi-scale Faster R-CNN

Yiqing Zhou\textsuperscript{2,\*}, Zemin Cai\textsuperscript{2}, Yuting Zhu\textsuperscript{2} and Jingwen Yan\textsuperscript{1,2,\*}

\textsuperscript{1} Guangdong Provincial Key Laboratory of Digital Signal and Image Processing, Shantou, 515063, Guangdong, China
\textsuperscript{2} College of Engineering, Shantou University, Shantou, 515063, Shantou, China
\* E-mail: jwyan@stu.edu.cn, 18yqzhou@stu.edu.cn

Abstract. Automatic ship detection of synthetic aperture radar (SAR) images has been widely used in maritime surveillance. SAR images have the characteristics of all-weather, all-day detection. Therefore, many object detection methods ranging from traditional to deep learning techniques have been proposed. However, the objects in large-scale remote sensing images are relatively small, and objects are often appeared at different scales. What's more, the current ship detection methods are insensitive to small-scale vessels. To solve these problems, a novel multi-scale ship detection method based on a Multi-scale Faster R-CNN network in SAR images is proposed in this paper. Firstly, a multi-scale network is used to decompose the SAR images into a pyramid structure and extract the features. Then, the region proposal network (RPN) is performed using the feature map for each layer to get the proposals that contains ship targets. Finally, these proposals are fed to the classification and regression network to obtain the final detection results. Multi-scale Faster R-CNN achieves the mean average precision (mAP) score 0.986 on the dataset of SAR-Ship-Dataset, which indicates that the proposed method has high detection accuracy and low missing rate.

1. Introduction

Synthetic Aperture radar (SAR) images are not affected by weather and light. And are widely used in marine monitoring, environmental monitoring and military applications, such as maritime transport safety, ship detection, etc. Germany launched the TerraSAR-X satellite in June 2007, the European Space Agency (ESA) launched Sentinel-1 in April 2014 and China launched the Gaofen-3 in August 2016. With the addition of satellite platforms, a vast amount of SAR images has been produced, SAR ship detection become an urgent technical.

The traditional SAR ship target detection mainly use Constant False Alarm Rate (CFAR) detection method, but the CFAR method cannot adapt to ship detection of all sizes. CFAR can only use pixel-based information, which is difficult to extract effective features. Under the circumstance of great complexity, there are too many false warning targets, which will affect the detection performance of CFAR detection method\cite{1}. Cell-average CFAR (CA-CFAR) and order statistic CFAR (OS-CFAR) detectors are presented in\cite{2}\cite{3}. These two detectors use the self-adaptive methods to determine the detection threshold. However, these two methods ignore the local characteristics in SAR ship images and have low detection performance for complex scenes. In\cite{4}, a multi-scale fused heterogeneity detector and a contrarious decision-based target detector are proposed. By reducing the interference of background and speckle noise, the target can be highlighted effectively, and the target and background can be separated accurately. But the traditional detectors are slow in speed and low in precision. In\cite{5},
the Single Shot Multibox Detector (SSD) is used in SAR ship images. However, the accuracy of SSD detector is relatively low. In [6], it put forward the application of Faster R-CNN to the SAR image of ship detection. The accuracy of Faster RCNN is improved to some extent, but the faster RCNN does not consider the multi-scale characteristics of SAR ship images. In [7], RetinaNet was used to detect ship targets. But there exists the problem with small targets that have not been detected.

Based on the above analysis, Multi-scale Faster R-CNN is proposed for detecting small targets of SAR ship images. It contains a multi-scale network to extract the multi-scale characteristics of ships. Through this network, multi-scale features of ships are extracted and put them into the Region Proposal network (RPN) to generate candidate regions. Finally, get the detection targets.

This paper is organized as follows. Section 2 relates to the proposed method. Section 3 reports on the experiments, including the dataset and experimental analysis. Section 4 offers the conclusion.

2. Multi-scale Faster R-CNN

The architecture of Multi-scale Faster R-CNN has three components, the backbone with the multi-scale network, RPN, classification and regression. The network structure is shown in figure 1.

2.1. Multi-scale Network

In SAR ship images, ships usually have different scales, which result in different characteristics. For convolutional neural networks, different depths correspond to different levels of semantic characteristics. The original Faster R-CNN used the top-level features, ignoring the lower-level features. However, these low-level features are very helpful for detecting small objects [8]. Therefore, based on Faster R-CNN, this paper adds a multi-scale sub-network, as shown in figure 2, which consists of a bottom-up line and a top-down line and a horizontal connection.
The bottom-up pathway uses ResNet to construct. It consists of many convolution modules each has many convolution layers. As move up, the spatial dimension is reduced by 1/2. The output of each convolution module later is used in the top-down pathway. Specifically, the feature maps from bottom-up pathway undergoes $1 \times 1$ convolution filter to reduce the channel dimensions.

As for top-down path, the spatial resolution is upsampled by a factor of 2 using the nearest neighbour. Then the feature maps are merged from the bottom-up pathway and the top-down pathway by element-wise addition. Finally, each layer uses a $3 \times 3$ convolution filter to generate the final feature map, which is to reduce the aliasing effect of upsampling. Then these feature maps are fed into RPN.

2.2. RPN

In this paper, in order to generate proposals, the RPN in Faster R-CNN is used[9]. RPN is a full convolutional network used to generate bounding box that contains candidate targets. The feature maps of different scales generated by the multi-scale network are input into RPN head. Anchors are generated by the RPN head are fed into two full-connection layers, the classification layers and the regression layers. The classification layers output the probability of each anchor being object and the probability that it is not object. The probability is in terms of $p_i$ , $i$ is the index of anchor in a mini-batch.

$$p_i^* = \begin{cases} 0, & \text{negativelabel} \\ 1, & \text{positivelabel} \end{cases}$$

(1)

For bounding box regression, use the parameterizations of the four coordinates following[9] :

$$t_x = \frac{(G_x - P_x)}{P_w}, \quad t_y = \frac{(G_y - P_y)}{P_h}$$

(2)

$$t_w = \log \frac{G_w}{P_w}, \quad t_h = \log \frac{G_h}{P_h}$$

(3)

$$t_{x*} = \frac{(\hat{G}_x - P_x)}{P_w}, \quad t_{y*} = \frac{(\hat{G}_y - P_y)}{P_h}$$

(4)

$$t_{w*} = \log \frac{\hat{G}_w}{P_w}, \quad t_{h*} = \log \frac{\hat{G}_h}{P_h}$$

(5)

where $x$, $y$, $w$ and $h$ denote the box’s center coordinates and its width and height, $G_i = (G_x, G_y, G_w, G_h)$, $P_i = (P_x, P_y, P_w, P_h)$, $\hat{G}_i = (\hat{G}_x, \hat{G}_y, \hat{G}_w, \hat{G}_h)$ respectively represents the predicted box, the anchor box and the ground-truth box. The above formulas can be regarded as bounding-box regression from an anchor box to a nearby ground-truth box.

The loss function of the RPN portion is as follows:

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

(6)

Where $t_i^* = \{t_{x*}^*(P), t_{y*}^*(P), t_{w*}^*(P), t_{h*}^*(P)\}$, $t_i = \{t_x(P), t_y(P), t_w(P), t_h(P)\}$. The $L_{cls}$ term in (6) is normalized by the mini-batch size and the $L_{reg}$ term is normalized by the number of anchor locations. By default, we set $\lambda = 10$ and thus both classification and regression terms are roughly equally weighted. At the end of RPN, 1000–2000 anchors are generated and will be used later.

2.3. Classification and Regression

Due to the problem of zone mismatch in RoiPooling, RoIAlign was selected in this paper[10]. RoIAlign traverses each candidate region, keeping the floating points from being quantified. Firstly, the candidate
region is divided into $k \times k$ cells, and the boundary of each cell is not quantified. Four coordinate positions are calculated in each cell, the output coordinates are calculated by bilinear interpolation. Then, the maximum pooling is used to obtain the pixel value of the small region. After RoIAlign, the generated proposals are sent into the full connection layers for classification and regression to get the results.

3. Experiments

3.1. Experiments on SAR Ship Dataset

The dataset used in this paper is the SAR Ship Dataset published in[11]. A total of 102 Chinese Gaofen-3 images and 108 Sentinel-1 images are used. Details are shown in the table 1.

| Sensor   | Imaging Mode | Resolution $\times$ Az.(m) | Swath(km) | Incident Angle ($^\circ$) | Polarization |
|----------|--------------|-----------------------------|-----------|--------------------------|--------------|
| GF-3     | UPS          | $3 \times 3$                | 30        | 20–50                    | Single       |
| GF-3     | FSI          | $5 \times 5$                | 50        | 19–50                    | Dual         |
| GF-3     | QPSI         | $8 \times 8$                | 30        | 20–41                    | Full         |
| GF-3     | FSII         | $10 \times 10$              | 100       | 19–50                    | Dual         |
| GF-3     | QPSII        | $25 \times 25$              | 40        | 20–38                    | Full         |
| Sentinel-1| SM           | $1.7 \times 4.3$ to $3.6 \times 4.9$ | 80        | 20–45                    | Dual         |
| Sentinel-11 | IW          | $20 \times 22$              | 250       | 29–46                    | Dual         |

The dataset including 43,819 ship slices containing 59,535 ship targets. There are some examples of ship slices in figure 3.

![Figure 3. Some examples of ship slices](image)

The size of each ship slice is $256 \times 256$, these ship slices have different scales and backgrounds. Each ship slice corresponds to the label data for xml format. The label contains the name of the ship slice, the ship’s position in the ship slice, the name of the ship, etc. The 43819 ship slices were randomly divided into training set and validation set, which the training set accounted for 80% and the test set 20%.

In the target detection problem, it is necessary to classify and locate the object at the same time, and the common precision index can not be used directly, so the mean average precision (mAP) is used as the evaluation index. The calculation formula for mAP is as follows:

$$mAP = \frac{1}{C} \sum_{i=1}^{C} AP_i$$  \hspace{1cm} (7)

$C$ is the number of categories, $AP = \frac{1}{C} \sum_{i=1}^{C} precision_i$, $precision = \frac{TP}{TP + FP}$, TP stands true positive, FP stands False positive.
3.2. Experiments on SAR ship Dataset

In order to better evaluate the performance of the proposed method in this paper, using three other models for comparison, Faster R-CNN, SSD and RetinaNet, using the same dataset for training. The experimental results are shown in table 2.

| Model       | SSD  | RetinaNet | Faster R-CNN | Ours  |
|-------------|------|-----------|--------------|-------|
| mAP (%)     | 72.52| 97.56     | 85.26        | 98.59 |

It can be seen from table 2, compared with SSD, RetinaNet and Faster R-CNN, our proposed method has higher mAP. This may be the multi-scale network can reserve the features of the deep layers. There are some detection results of our method in figure 4.

![Detection results](image)

4. Conclusion

This paper designs multi-scale Faster R-CNN for ship detection in high-resolution SAR images. The experiment results show that the multi-scale Faster R-CNN detection performance is the best compared with the other target detectors selected in this paper, and the future work will focus on ship detection near the port.

Acknowledgments

This work was supported by the Department of Education of Guangdong Province (2017KCXTD015, 2016KZDXM012), the National Science Foundation of China (61876104), and the Guangdong Natural Science Foundation(2018A030307034).

References

[1] Gao, G.; Liu, L.; Zhao, L.; Shi, G.; Kuang, G. An adaptive and fast CFAR algorithm based on automatic censoring for target detection in high-resolution SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 47.6(2009), 1685–1697
[2] Finn, H.; Johnson, R. Adaptive detection mode with threshold control as a function of spatially sampled clutter-level estimates. *RCA Review*, 29, 414–464.
[3] Hansen, V.; Sawyers, J. Detectability loss due to greatest of selection in a cell averaging CFAR. *IEEE Transactions on Aerospace Electronic Systems*, 16, 115–118.
[4] Xiaojing, H., Wen, Y., Haijian, Z., & Gui-Song, X. (2015). Automatic ship detection in sar images using multi-scale heterogeneities and an a contrario decision. *Remote Sensing*, 7(6), 7695–7711.
[5] Wang, Y.; Wang, C. & Zhang, H. (2017). Combining single shot multibox detector with transfer learning for ship detection using Sentinel-1 images. *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA)*. IEEE.
[6] Kang, M.; Leng, X.; Lin, Z.; Ji, K. A modified faster R-CNN based on CFAR algorithm for SAR
ship detection. 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP). IEEE.

[7] Wang, Y., Wang, C., Zhang, H., Dong, Y., & Wei, S. (2019). Automatic ship detection based on RetinaNet using multi-resolution gaofen-3 imagery. Remote Sensing, 11(5).

[8] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, & Serge Belongie. (2017). Feature Pyramid Networks for Object Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society.

[9] Ren, Shaoqing, He, Kaiming, Girshick, Ross, & Sun, Jian. (2015). Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, 39(6), 1137-1149.

[10] He Kaiming, Gkioxari Georgia, Dollar Piotr, & Girshick Ross. Mask r-cnn. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1-1.

[11] Wang, Yuanyuan & Wang, Chao & Zhang, Hong & Yingbo, Dong & Wei, Sisi. (2019). A SAR Dataset of Ship Detection for Deep Learning under Complex Backgrounds. Remote Sensing, 11. 10.3390/rs11070765.