



Towards More Robust Detection for Small and Densely Arranged Ships in SAR Image

Jingpu Wang^(✉), Youquan Lin, Long Zhuang, and Jie Guo

Nanjing Research Institute of Electronics Technology, Jiangsu, China
13813085308@163.com

Abstract. Independent of sunlight and weather conditions, synthetic aperture radar (SAR) imagery is widely applied to detect ships in marine surveillance. This paper proposes a ship target detection algorithm, aiming at the missing and false detection in the situation of dense ship targets in SAR images. Firstly, we used the spatial pyramid pooling (SPP) to enhance the feature extraction capability in different scales. Then, we modified the regression loss function with three factors of center distance, overlap area and length-width ratio to reduce the error of location. Finally, we proposed double threshold soft non maximum suppression (DTSOFT-NMS) to reduce the missing detections for dense ships. The experimental results reveal that our model exhibits excellent performance on the open SAR-ship-dataset and improves average precision (AP) by 6.5% compared with the baseline YOLOv3 model.

Keywords: Synthetic Aperture Radar (SAR) · Spatial pyramid pooling · Double threshold soft non maximum suppression (DTSOFT-NMS)

1 Introduction

Synthetic Aperture Radar (SAR) is an active microwave sensor that can emit electromagnetic waves and receive echo signals for active imaging without being restricted by external conditions such as weather and light. With all-weather characteristics, SAR images have been widely applied in civil and military fields, and ship target detection based on SAR images has become one of the hot research issues.

Traditional ship detection methods were mainly based on the different backscattering characteristics between ship targets and the ocean. Common methods include constant false alarm rate (CFAR) [1], polarization decomposition, wavelet decomposition and template method. With the development of SAR, the resolution of SAR images has further improved, and traditional detection methods are difficult to meet the current actual needs in terms of accuracy and detection efficiency.

Due to the strong feature extraction capabilities of convolutional neural network (CNN), deep learning has achieved great success in object detection. Object detection algorithms based on CNN can be divided into two categories: one is a two-stage object detection algorithm, also known as a region-based algorithm, includes Fast R-CNN [2], Faster R-CNN [3], R-FCN [3], and Mask R-CNN [5]. The other one is a one-stage

object detection algorithm, also called regression-based object detection algorithm. This type of algorithm directly generates the category and position information of the target. Representative algorithms include SSD [6], RetinaNet [7] and YOLO series [8–10].

Zhang [11] proposed a ship target detection algorithm based on visual attention via CNN coupling cascade (3c2n-guided), which can improve detection performance and greatly reduce missed detection and false positives. CHEN [12] proposed a target detection network combined with attention mechanism to accurately locate the target in complex scenarios. JIAO [13] proposed a densely connected multiscale neural network based on Fasters R-CNN framework to solve the multiscale multi-scene SAR ship detection problem. Miao [14] proposed a detection method that the traditional CFAR was combined with Faster R-CNN network to improve the detection performance. Yang [15] proposed a new detection method based on SSD, which combines context information fusion, migration model learning and SSD, and conducts training and testing on the public SSDD dataset. In the past years, deep learning has made great progress in ship target detection of SAR images, but there are still many problems.

Because the SAR system has multiple imaging modes and imaging resolutions, ships have different sizes in SAR images. Current detection methods have poor detection effect on multi-size ships, especially small-size ships. Besides, in the dense areas with small adjacent spaces between ships, the detection strategy of CNN will lead to missing and false detection.

In this paper, we propose a method for multiscale and densely arranged ships detection in SAR Image. The main contributions of this paper are as follows.

1. We chose YOLOv3 as the baseline and use the SPP module to improve the feature extraction capability of the network;
2. We redesigned the regression loss function based on CIOU [17], which can reduce the scale sensitivity and improve the location accuracy;
3. We proposed DTSoft-NMS into the final processing to improve the detection effect for dense ships;
4. Based on the public SAR ship dataset [18] proposed by the Chinese Academy of Sciences, we use k-means clustering to redesign the anchor boxes, train and test the proposed network.

The organization of this paper is as follows. Section 2 relates to the related work of the baseline model. Section 3 illustrates our proposed method and network structure. Section 4 introduces the dataset used in our experiments and describes the experimental details and results. Section 5 presents the conclusion.

2 Related Work

In 2016, You Only Look Once (YOLO) has been introduced which unlocks the potential of real-time performance. YOLO uses regression method to predict the coordinates of the bounding boxes and achieve the classification of the targets. YOLO applies a grid approach for bounding box detection and adds a SoftMax layer to directly predict the object class. The detection speed reaches 30 FPS, but YOLO has serious location error,

causing poor detection accuracy. YOLOv2 modified the feature extraction network with Darknet-19, uses Anchor Box to predict targets and applies Batch Normalization (BN) for training. Compared with YOLO, the running speed and detection accuracy of YOLOv2 have been significantly improved. YOLOv3 uses the deep residual network Darknet-53 to extract features and achieve multi-scale prediction, which make the accuracy and speed meet the actual engineering needs.

The network structure of YOLOv3 is composed of feature extraction network darknet-53 and feature pyramid network (FPN).

Inspired by residual network, darknet-53 has 53 convolution layers, which can be divided into 5 residual convolution blocks. Each convolution block is composed of multiple residual units. The residual operation is carried out by the input and two digital cumulative modeling (DBL) units to construct the residual unit. Every DBL cell contains convolution layer, BN, and leaky ReLU activation functions. By introducing residual units, the depth of the network can be increased to reduce the gradient vanishing problem.

The FPN of YOLOv3 is shown in Fig. 1. In the process of detection, the input image will be down-sampled for 5 times, and the last three feature maps are fused with larger feature maps by up-sampling. When the input image size is 416×416 , the detection is performed on the scales of 13×13 , 26×26 and 52×52 .

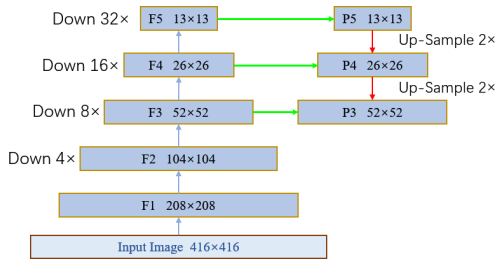


Fig. 1. The feature pyramid network of YOLOv3

Feature maps of different scales contain different feature information. The small maps can usually provide deep semantic information, while large maps contain more targets position information, especially small targets information. Thus, YOLOv3 fuses feature maps of different scales to improve multiscale detection ability. It can not only detect large size targets, but also optimize the detection of small size targets.

3 Methods

3.1 Spatial Pyramid Pooling

In this paper, we use SPP module to improve the extraction ability of local and global features between multiscale ships in SAR images.

The SPP module is based on the SPP-net [16], which is used to solve the problem of different image input to fixed size output to the full connection layer. The application of SPP in our method can alleviate the over-sensitivity of the convolutional layer for

position information. It can realize the fusion of local feature and global feature at the feature maps level.

The structure of SPP is shown in Fig. 2, which is all pooling layers. SPP does not change the size of feature maps before and after processing. There are three sizes of pooling kernel, with 3×3 , 5×5 , 7×7 , and then superimpose the initial feature map to obtain the new feature map. We add three SPP modules before the detectors of the three scale feature maps in the baseline model. The new structure of our models is shown as Fig. 3.

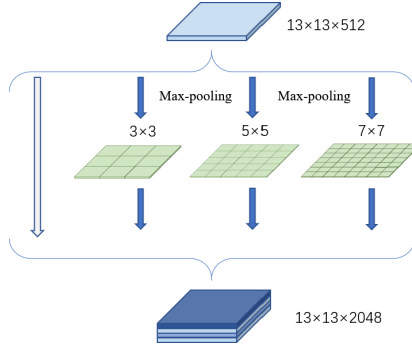


Fig. 2. The structure of SPP

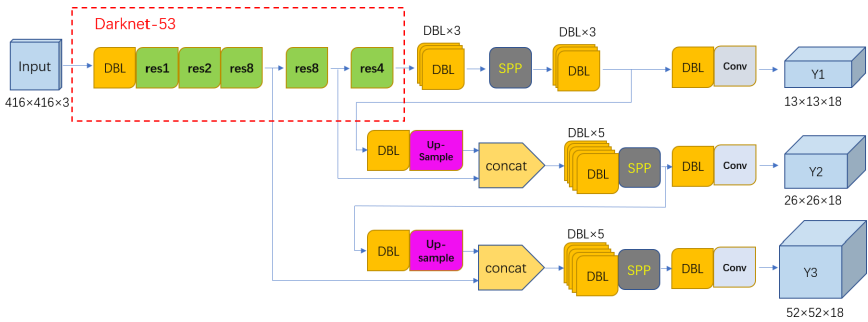


Fig. 3. The structure we proposed

In SAR images, the shapes of ships are different, so ships in the same image may have multiple sizes. Besides, the same ship in various resolution images has different sizes. Therefore, multiscale ship is a difficult problem for SAR ship detection. YOLOv3 performed object detection on three scale feature maps and each of which is responsible for different ships.

Usually, the feature maps of 52×52 detect small ships and 26×26 , 13×13 feature maps detect big ships. Three SPP modules before three detectors can improve the detection effect of multiscale ship.

3.2 Loss Function

The loss function of the original YOLOv3 network is composed of four parts, including center coordinate error, width-height error, confidence error and classification error. Among them, center coordinate error and width and height coordinate error are calculated by mean square error (MSE). However, the various target sizes will lead to the large fluctuation of the mean square error loss, which will eventually affect the positioning accuracy of the target and reduce the detection accuracy.

Inspired by the CIOU [17], we modify the loss function with overlapping area, distance penalty and length-width penalty to reduce scale sensitivity enhance positioning accuracy. The new loss function is defined as follows:

$$L_{reg} = 1 - IoU + \rho^2(b, b^{gt}) + \alpha v \tag{1}$$

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \tag{2}$$

where B is the target prediction box and B^{gt} is the ground true box; b and b^{gt} represent the center points of the target prediction box and the true box; ρ is the Euclidean distance between the two points.

In the length-wide penalty, α is the coefficient used to balance the proportion and v is used to measure the proportion consistency between the anchor frame and the target frame. Their formulas are as follows:

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \tag{3}$$

$$\alpha = \frac{1}{(1 - IoU) + 1} \tag{4}$$

where w, h are the width and height of prediction box and w^{gt}, h^{gt} are the width and height of the ground true box.

It can be seen that the length-wide penalty is invariant with respect to the scale. When calculating the loss between the prediction box and the real box, our loss function not only considers the distance between the overlapping area and the center point, but also adds the influencing factor of the aspect ratio of length and width. Therefore, the new loss function can achieve better regression for different distances, directions, areas and achieve faster convergence speed and convergence effect.

3.3 DTSoft-NMS

Most detection algorithms use Non maximum suppression (NMS) to suppress overlapping boxes. The idea of NMS is to regard the box with the highest score as the best detection result and then remove all other prediction boxes with high overlap degree. The mathematical expression of this algorithm is as follows:

$$S_f = \begin{cases} S_i, & U_{IoU}(Max, B_i) < N_t \\ 0, & U_{IoU}(Max, B_i) \geq N_t \end{cases} \tag{5}$$

where S_i is the original score of the prediction box; S_f is the final score of the prediction box; Max is the prediction box with the highest score, B_i is the other boxes to be checked except Max ; $U_{IOU}(Max, B_i)$ is the IOU of B_i and Max and N_t is the threshold of IOU.

When the overlap of the two boxes is small, IOU is less than the threshold value. According to the NMS algorithm, the two prediction boxes belong to different target detection results, so two boxes are both reserved. When the overlap of two boxes is very high and IOU is greater than the threshold, as shown in Fig. 4(a), the NMS algorithm will determine that these two boxes are the detection results of the same target, so it will clear the confidence score of the box with a low score, so as to complete the filtering of the repeated prediction box.

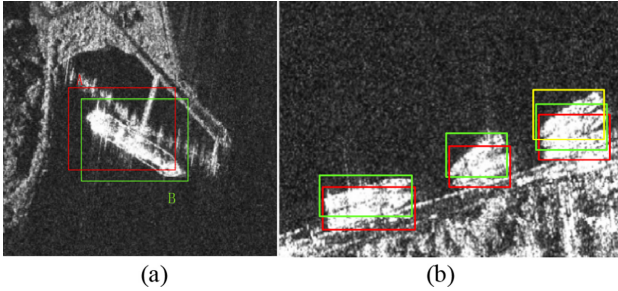


Fig. 4. An illustration of dense ships

However, when the ships are very closed, the prediction box of two different targets have large IOU and the prediction box with lower score will be cleared, resulting in a missed detection. As shown in Fig. 4 (b), the distance between ships is very close in port area. The prediction boxes of different colors represent the detection results of different ship targets. However, according to the NMS algorithm, only the prediction box of the target with the highest score will be kept, and multiple ships will be regarded as the same target, resulting in missed detection.

Therefore, we propose DTSOFT-NMS algorithm to solve the detection problem of dense targets. The core idea of DTSOFT-NMS algorithm is to reduce the score of the prediction box with large overlap by using the penalty strategy, and to reserve the prediction box belonging to different targets to the greatest extent, so as to reduce the missed detection.

In the DTSOFT-NMS algorithm, the penalty function of the confidence score is calculated as (6). The confidence score of the prediction box with small overlap will not decay. However, the prediction box with large overlap attenuates greatly, but does not clear directly. When the IOU of the prediction box is really large, the score will be remove. Such a strategy can keep the prediction box of densely distributed ship targets and reduce the missed detection.

$$S_f = \begin{cases} 0 & , N_{t1} \leq U_{IoU}(Max, B_i) \\ S_i e^{-\frac{IoU(Max, B_i)^2}{N_{t2}}} & , N_{t2} \leq U_{IoU}(Max, B_i) < N_{t1} \\ S_i & , U_{IoU}(Max, B_i) < N_{t2} \end{cases} \quad (6)$$

4 Experiments

4.1 The Experiment Platform

All experiments were implemented on a workstation with an Intel(R) Core (TM) i5-9400F @2.90 GHz, an NVIDIA RTX 2070 GPU and the Pytorch framework. The initial learning rate of the network was set to 0.001 for the first 400 iterations and 0.0001 for the last 400 iterations. The batch size is 8. The optimization algorithm used Adam, with beta1 of 0.9, beta2 of 0.999 and epsilon of 1e-8.

4.2 Dataset

In this paper, we use the SAR-ship-Dataset 18 published by Chinese academy of sciences to train and test the deep neural network. A total of 102-scene gaofen-3 and 108-scene sentinel-1 SAR images are used to build a SAR ship target deep learning sample library. The dataset contains 43819 ship slices and their label information, while the target to be detected is only one class. Thus, it is sufficient for training our network. In our experiment, 80% of the samples were used as the training set and 20% of the sample data were testing set.

4.3 Evaluation Metrics

The evaluation metrics used in this paper are average accuracy (AP) to quantitatively evaluate the detection effect of the model. The average accuracy is defined as follows:

$$AP = \int_0^1 P(R)dR \quad (7)$$

$$R = \frac{X_{TP}}{X_{TP} + X_{FN}} \quad (8)$$

$$P = \frac{X_{TP}}{X_{TP} + X_{FP}} \quad (9)$$

where P is the detection accuracy; R is the target recall rate; X_{TP} represents the number of targets correctly detected; X_{FN} represents the number of targets not detected; X_{FP} represents the number of targets that were incorrectly checked out.

4.4 Ablution Study

To illustrate the effectiveness of the proposed approaches, we compared the effects of different versions of our method on the AP through step-by-step experiments based on the SAR-ship-Dataset. The various approaches incorporated into our model, as mentioned above are shown in Table 1.

Table 1. Ablution Study

Method	SPP	CIOU	DTSOFT -NMS	AP (%)
Proposed	✓			86.22%
		✓		85.43%
			✓	83.56%
		✓	✓	86.96%
	✓		✓	87.86%
	✓	✓		89.75%
	✓	✓	✓	89.92%

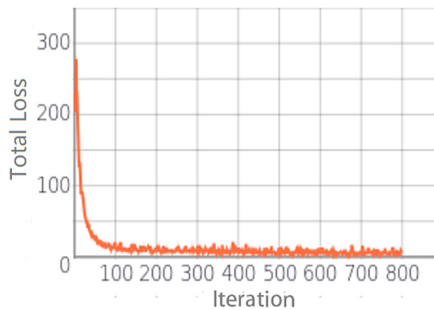
We add three SPP modules before the detectors of the three scale feature maps in the YOLOv3 network to enhance the feature extraction capability of network at different scales and the AP of the model is improved by 2.9%.

By using the CIOU loss into the loss function, the sensitivity of the network to different ship target scales is reduced, and the AP of the model is improved by 2.06%.

By contrast, using DTSOFT-NMS improves the AP of the model by only 0.17%. The reason for this lesser improvement maybe that the dataset contains relatively few densely arranged ship samples. Based on the approaches mentioned above, the final AP of the proposed model is 89.92% on the SAR-Ship-Dataset.

4.5 Results Analysis

In Fig. 5, we can see that as the number of training iterations increases, the loss value gradually decreases and converges to a low value. Finally, the loss function converges to 0.69.

**Fig. 5.** Loss function curve during the training process

The application of SPP in YOLOv3 network can realize the fusion of local feature and global feature at the feature graph level, so as to enrich the expression ability. As shown in Fig. 6, the blue line represents the precision of our detection algorithm and the yellow line represents the original YOLOv3. By comparing, we can find that adding SPP structure can improve AP by 2.96%.

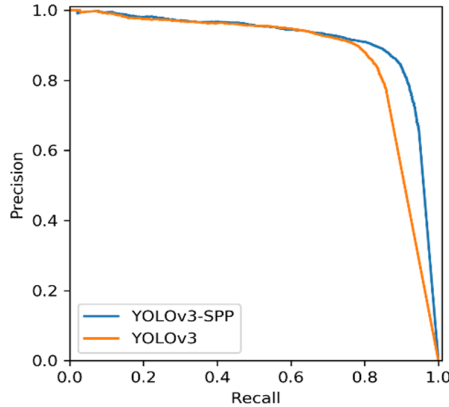


Fig. 6. The PR curves of YOLOv3 and YOLOv3-SPP

The ship detection results obtained under the different environmental conditions represented in the SAR-Ship-Dataset were analyzed, as shown in Fig. 7. We show several detection results for SAR ships.

The first row shows results for ship detection against complex backgrounds. We find that the proposed method can effectively distinguish targets from their backgrounds. The second row shows results for the detection of ships of different sizes. It is clear that the proposed algorithm achieves an improved detection effect for small ships and big ships with a lower missed detection rate. The third row shows densely arranged ships detection. It can be seen that the algorithm proposed in this paper can effectively distinguish closely spaced ships.

To better illustrate the effectiveness of the proposed DTSOFT-NMS, we select a sub-dataset from sentinel-1 SAR images with densely arranged ships. It contains 32 images and 108 ship targets. The result in Table 2 shows that DTSOFT-NMS can reduce some missed detection compare with original NMS.

4.6 Comparison with Other Methods

In this section, our proposed method is quantitatively compared with several mainstream object detection models based on deep learning in terms of the AP and detection speed. The results are shown in Table 3.

It is apparent that the proposed method achieves the best AP of 89.92% on the SAR-Ship-Dataset compared with other one-stage object detection methods including SSD and original YOLOv3. The use of SPP and DTSOFT-NMS increase the computation of

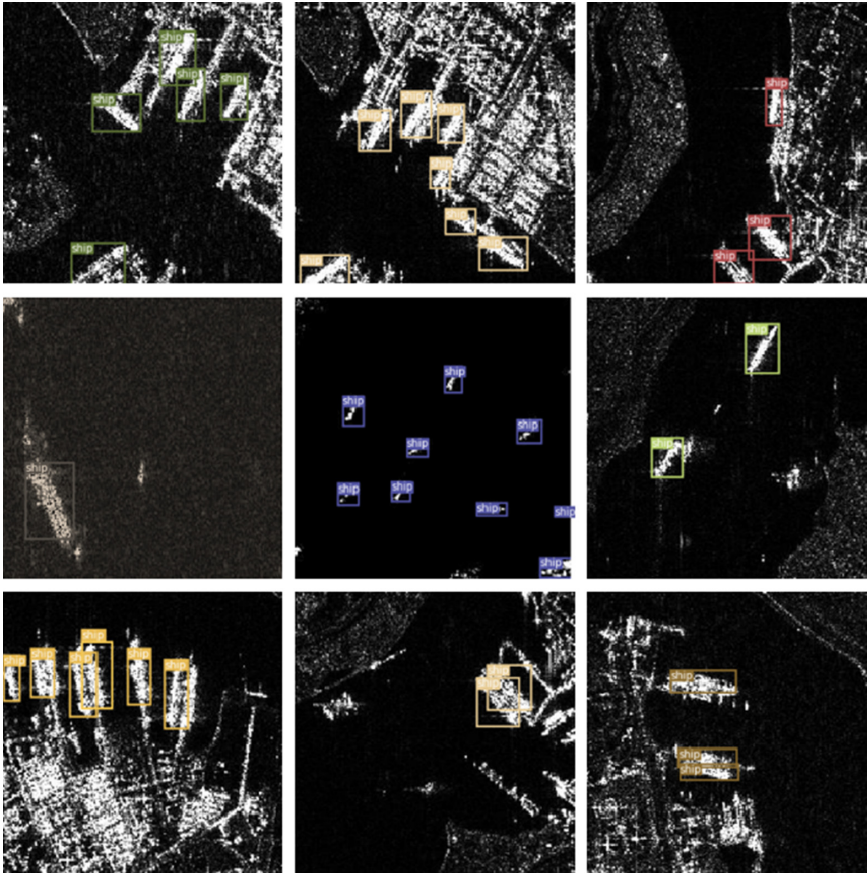


Fig. 7. Experimental results

Table 2. Densely arranged ships detection

Methods	Detected ships	True ships	False alarms	Missed ships	Precision	Recall
NMS	114	94	20	14	82.45%	87.03%
DTSOFT-NMS	122	100	22	8	81.96%	92.59%

the network. Thus, our algorithm is not as fast as original YOLOv3. But its detection time of one image is 30.14 ms, which is sufficient for real-time detection. Compared with the two-stage object detection algorithm Faster R-CNN, the time cost of the proposed algorithm is only 30% of that of Faster R-CNN, and the AP of our method is 1% higher than Faster R-CNN.

Table 3. Compare with other methods

Method	Backbone	AP	Time
Faster R-CNN	Resnet-101	88.74%	93.29 ms
SSD	VGG16	78.62%	31.52 ms
YOLOv3	Darknet-53	83.36%	26.74 ms
Proposed	Darknet-53 with SPP	89.92%	30.14 ms

5 Conclusion

In this paper, we proposed a detection algorithm based on YOLOv3 for ship detection in SAR image and verified our method on the public SAR-Ship-Dataset. Based on the YOLOv3, this paper used the spatial pyramid pooling structure to improve the feature extraction capability of the network. Then, the regression loss function based on CIUO is proposed to make the network converge faster and obtain higher positioning accuracy during training. Finally, this paper proposed DTSOFT-NMS to reduce the missed and false detection in the process of intensive target detection. The detection time of a single image in the SAR-Ship-Dataset is 30.14 ms and the AP is 89.92%, which is significantly improved for densely arranged ships and multiscale ships.

References

1. Xing, X.W., Chen, Z.L., Zou, H.X., et al.: A fast algorithm based on two-stage CFAR for detecting ships in SAR images. In: The 2nd Asian-Pacific Conference on Synthetic Aperture Radar, pp. 506–509 (2010)
2. Girshick, R.: Fast R-CNN, IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
3. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. and Mach. Intell.* **39**(6), 1137–1149 (2017)
4. Dai, J., Li, Y., He, K., Sun, J.: R-FCN: object detection via region-based fully convolutional networks. In: *Advances in Neural Information Processing Systems*, pp. 379–387 (2016)
5. Kaiming, H., Georgia, G., Piotr, D., et al.: Mask R-CNN. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2961–2969 (2017)
6. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, Alexander C.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016*. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
7. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 99, pp. 2980–2988 (2017)
8. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788 (2016)
9. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6517–6525 (2017)

10. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement. [arXiv:1804.02767](https://arxiv.org/abs/1804.02767) (2018)
11. Zhao, J., Zhang, Z., Yu, W., et al.: A cascade coupled convolutional neural network guided visual attention method for ship detection from SAR images. *IEEE Access* **6**, 50693–50708 (2018)
12. Chen, C., He, C., Hu, C., Pei, H., Jiao, L.: A deep neural network based on an attention mechanism for SAR ship detection in multiscale and complex scenarios. *IEEE Access* **7**, 104848–104863 (2019)
13. Jiao, J., et al.: A densely connected end-to-end neural network for multiscale and multi scene SAR ship detection. *IEEE Access* **6**, 20881–20892 (2018)
14. Kang, M., Leng, X., Lin, Z., Ji, K.: A modified faster R-CNN based on CFAR algorithm for SAR ship detection. In: 2017 International Workshop on Remote Sensing with Intelligent Processing, pp. 1–4 (2017)
15. Long, Y.A.N.G., Juan, S.U., Xiang, L.I.: Ship detection in SAR images based on deep convolutional neural network. *Syst. Eng. Electr.* **41**(9), 1990–1997 (2019)
16. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8691, pp. 346–361. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10578-9_23
17. Zheng, Z., Wang, P., Liu, W.: Distance-IoU loss: faster and better learning for bounding box regression. [arXiv:1911.08287](https://arxiv.org/abs/1911.08287) (2019)
18. Wang, Y., Wang, C., Zhang, H., Dong, Y., Wei, S.: A SAR dataset of ship detection for deep learning under complex backgrounds. *Remote Sens.* **11**(7), 765–769 (2019)