



Synthetic Aperture Radar Image Ship Detection Based on YOLO-SARshipNet

Bing Chen¹, Yuting Zhu², Chen Wang¹, and Xiaoqing Wang²(✉)

¹ Transport Planning and Research Institute Ministry of Transport, Beijing, China

² Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou, China

xqwang@mail.sysu.edu.cn

Abstract. Automatic ship detection from synthetic aperture radar (SAR) imagery plays a significant role in many urban applications. Recently, owing to the impressive performance of deep learning, various SAR ship detection methods based on the convolution neural network (CNN) have been proposed for optical SAR images. However, existing CNN-based methods and spatial-domain-based methods exhibit certain limitations. Some algorithms do not consider the detection speed and model scale when improving the detection accuracy, which limits the real-time application and deployment of SAR. To solve this problem, a lightweight, high-speed and high-accurate SAR ship detection method based on yolov3 has been proposed. First, the backbone part of the model is improved, the pure integer quantization network is applied as the core to reduce a small amount of accuracy while reducing the model scale by more than half; Second, modify the feature pyramid network to improve the detection performance of small-scale ships by enhancing feature receptive fields; third, introduce the IoU loss branch to further improve the detection and positioning accuracy; finally, the feature distillation is applied to handle the problem of accuracy decrease caused by model integer quantization. The experimental results on the two public SAR ship datasets show that this algorithm has certain practical significance in the real-time SAR application, and its lightweight parameters are helpful for future FPGA or DSP hardware transplantation.

Keywords: Ship detection · Synthetic aperture radar · Pure integer quantization · Feature pyramid network · Feature distillation

1 Introduction

Synthetic aperture radar (SAR) has important applications in various fields, such as military, agriculture, and oceanography, owing to its high resolution and all-weather features among other advantage. [1] Ship target detection is an

Bing Chen and Yuting Zhu contributed to the work equally and should be regarded as co-first authors.

important link in the implementation of marine surveillance, and has important applications in military reconnaissance, marine transportation management, and maritime crime combat. In this way, SAR images become an effective remote sensing data source for ship targets, and SAR ship target detection has been widely studied. [2]

The traditional ship detection methods are mainly based on the statistical distribution of sea clutter, such as the constant false alarm rate (CFAR) method, and the method based on machine learning to extract features [3]. However, these traditional methods are highly dependent on the pre-defined distribution of ship geometric features or semantic features, which degrades the ship detection performance in SAR images. Therefore, it is difficult for these methods to perform ship detection accurately and robustly. In addition, some super-pixel-based ship detection methods have been proposed, which are also difficult to accurately detect ships in nearshore and offshore scenes [4]. Following the advancement of deep learning technology, many methods have been designed to use deep learning to solve the problem of SAR ship detection. The target detection algorithm based on deep learning is not limited by the scene, does not need sea-land-separation, and can learn the characteristics of ships spontaneously, which improves the shortcomings of traditional methods to a certain extent [5].

Supervised deep learning methods often require many training samples as support. In the field of SAR image ship detection, the Naval Engineering University released the first public SAR Ship Detection Dataset (SSDD) in China [6]. Since then, there are mainly high-resolution SAR ship target datasets (SAR-Ship-Dataset) constructed by the team of the Institute of Aeronautical and Astronautical Information, Chinese Academy of Sciences [7]. On this basis, SAR ship detection methods based on deep learning have developed rapidly. On the SSDD dataset, [8] uses the optimized Faster R-CNN model, with feature fusion and transfer training to achieve a detection accuracy of 78.8%. [9] uses the improved YOLOv3 [10] algorithm, combined with the feature pyramid structure, to improve the detection performance of small-sized ships, the accuracy is improved but the detection speed is reduced to a certain extent. [11] proposed a detection method SARShipNet-20 based on the depth separation convolutional neural network, and adopted the attention mechanism, which achieved good results in detection accuracy and detection speed. [7] introduces the SAR-Ship-Dataset dataset in detail and uses different object detection algorithms to conduct experiments on this dataset. [12] uses Retinal-Net as the basis, with the pyramid network and class imbalance loss function to improve the accuracy, reaching 96% accuracy. The above methods have achieved good results in SAR ship detection, but they all have the characteristics of large number of parameters, which improves the accuracy but reduces the detection speed. Therefore, there are certain defects in the hardware transplantation of the model and the application of SAR ship detection with high real-time requirements.

To solve the above shortcomings, this paper proposes a lightweight high-speed high-precision SAR ship detection method YOLO-SARshipNet based on the improved YOLOv3 framework. First, a pure integer quantization network is applied as the backbone network. The network structure is compact and efficient,

including depth separation convolution, and the size of the model parameters is 1/6 of the original YOLOv3 framework; second, the spatial pyramid pooling module is introduced into the feature pyramid, which improves the detection ability of the model for small-sized ships by enhancing feature receptive field; third, introduce the IoU loss branch to further improve the accuracy of detection and positioning; finally, feature distillation is applied to solve the problem of accuracy decrease caused by model integer quantization. The experimental results demonstrate that our method can achieve an excellent SAR image ship detection effect. The contributions of this study are as follows:

- We propose a YOLO-SARshipNet detection model that successfully reduces the model size by one-sixth without sacrificing accuracy. The detection model can effectively detect ships in the SAR images;
- We introduce feature distillation to successfully solve the problem of model accuracy degradation;
- The experimental results on SSDD and SAR-Ship-Dataset show that YOLO-SARshipNet has achieved relatively good results in terms of accuracy, speed, and model size, and its lightweight parameters are helpful for FPGA or DSP hardware porting.

The rest of this paper is organized as follows. In Sect. 2, the principles and implementation of the proposed algorithm is introduced. Section 3 presents experimental results and discussion. Finally, a summary is provided in Sect. 4.

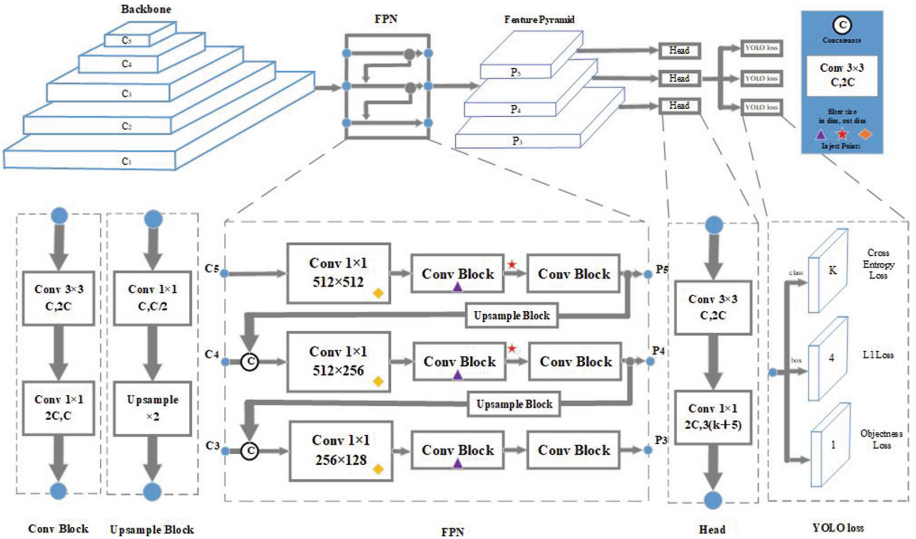


Fig. 1. Schematic of YOLO-SARshipNet algorithm flow.

2 Model Structure and Improvement Methods

Synthetic aperture radar (SAR) has important applications in various fields, such as military, agriculture, and oceanography, owing to its high resolution and all-weather features among other advantage. Ship target detection is an important link in the implementation of marine surveillance, and has important applications in military reconnaissance, marine transportation management, and maritime crime combat. In this way, SAR images become an effective remote sensing data source for ship targets, and SAR ship target detection has been widely studied.

Target detection models based on deep learning have achieved great success in natural image target detection tasks, but most of those models are not suitable for SAR image ship detection. On the one hand, there is a big difference between SAR images and ordinary optical images; on the other hand, most detectors are designed for multi-target detection. If they are directly applied to SAR detection, there must be a certain degree of redundancy. Therefore, different from the existing optical image target detection models, this paper designs a YOLO-SARshipNet network that is more suitable for SAR ship detection and deployment based on the YOLOv3 network [10].

The overall network structure of the designed YOLO-SARshipNet is shown in Fig. 1. First, a deep convolutional backbone network is used to extract features of ship targets in SAR images; second, a feature pyramid network is constructed using the extracted feature layers to fuse features of different scales; finally, on the fused feature map, the position information of the target, the confidence information of the category and the intersection ratio information of the position are applied for regression prediction.

2.1 Backbone

The traditional YOLOv3 [10] uses Darknet-53 as the feature extraction skeleton network, which consists of a series of 1 and 3 convolutional layers (each convolutional layer is followed by a normalization layer and a pooling layer), the network contains a total of 53 convolutional layers, and a residual structure is introduced [13], so it is called Darknet-53. However, the YOLOv3 is aimed at optical images containing multi-scale and multi-target, while the SAR image ship target is small, the size is relatively single, and the gradient between the target area and the background area is not obvious. In addition, the size of the YOLOv3 model using Darknet-53 as the backbone network that reaches 139.25 MB, which is not suitable for hardware transplantation of FPGA or DSP. To handle these problems, this paper constructs a pure integer quantization network as a backbone network.

The pure integer quantization network is based on the Efficient-Net network proposed in [14]. Applying EfficientNet-B4 as the skeleton network of the model can reduce the number of parameters of the model by half without reducing the detection accuracy. However, there are still two problems in the deployment of

this model on embedded devices. First, many ship detection devices have limited floating-point operation support; second, each hardware supports different operation acceleration operations, and several operations in the network cannot be supported by hardware. In response to the above problems, the tools provided by Google [15] are used to quantize the model, that is, the floating-point operation of the model is converted into an integer operation. This operation not only helps the device-side operation, but also successfully reduces the size of the model. Second, the attention mechanism in Efficient-Net is removed, which facilitates device deployment and reduces the ship detection time of the model. The constructed pure integer quantization network is based on EfficientNet-B4, which contains 19 network layers and 5 down-sampling layers.

2.2 Feature Pyramid Networks

The structure of Feature Pyramid Networks (FPN) [16] is shown in Fig. 1. When the image passes through the backbone network, feature blocks with larger receptive fields and richer semantic information are finally obtained through continuous convolution and down-sampling operations. However, larger down-sampling will lose part of the semantic information, which cause missed detection of small-sized objects. In order to solve this problem, this paper uses FPN to perform convolution and up-sampling operations on the feature blocks obtained by the backbone network and adopts a feature fusion strategy to fuse the shallow semantic information and deep semantic information, thereby improving the detection ability of small-sized ships.

In addition, this paper introduces the Spatial Pyramid Pooling Layer (SPP) [18] in FPN. The SPP module applied in this paper concatenates the output of max pooling with kernel size $k = 1, 5, 9, 13$, the stride is 1, and the insertion position is the “red star” position in the Fig. 1. Introducing the SPP module to the feature blocks containing rich semantic information has the following advantages:

- The SPP module is multi-scale, which improves the receptive field of the feature blocks;
- The SPP module does not lead additional parameters, effectively ensuring the model;
- The detection accuracy of the model is effectively improved after the introduction of the SPP module.

2.3 Loss Function

After passing through the FPN, three feature layers are selected for two convolutions to obtain the following feature parameters of the SAR image: (1) The center coordinates (x, y) of the detection anchor; (2) The width and height coordinates (w, h) of the detection anchor; (3) Ship confidence C of detection anchor. We utilize these feature parameters to construct a new well-designed loss function:

$$\text{loss} = \text{loss}_{box} + \text{loss}_{obj} + \text{loss}_{IoU} \quad (1)$$

where loss_{box} is the coordinate loss of the detection anchor, defined as:

$$\text{loss}_{box} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (2 - w_i \times h_i) \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2 \right] \quad (2)$$

where x_i^j, y_i^j, w_i and h_i are the real coordinates of the j -th frame of the i -th grid, and $\hat{x}_i^j, \hat{y}_i^j, \hat{w}_i$ and \hat{h}_i are the corresponding predicted coordinates; $I_{ij}^{obj} = 1$ when the grid contains a ship or part of a ship, otherwise $I_{ij}^{obj} = 0$. And λ_{coord} is the weight coefficient of coordinate loss, B is the number of detection frames, and S is the number of divided grids. loss_{obj} is the confidence loss function, which is defined as:

$$\text{loss}_{obj} = \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} (c_i - \hat{c}_i)^2 + \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (c_i - \hat{c}_i)^2 \quad (3)$$

where λ_{noobj} and λ_{obj} are the weight coefficients of the j -th box of the i -th grid with and without the target, respectively, and the value is 0.5.

loss_{IoU} is the intersection ratio loss function, defined as:

$$\text{loss}_{IoU} = -\ln IoU \quad (4)$$

where IoU is the intersection ratio of the predicted anchor and the label anchor.

$$IoU = \frac{\text{area}(B_G \cap B_P)}{\text{area}(B_G \cup B_P)} \quad (5)$$

where B_P is the predicted box, B_G is the label box, that is, IoU is the area ratio of the intersection and union of the predicted frame and the real frame area.

2.4 Feature Distillation

In this paper, the integer quantization technology is applied to reduce the number of parameters and calculation of the model, but it has a bad impact on the detection accuracy of the model. To solve this issue, we use a feature distillation-based approach.

The main idea of feature distillation is to train a small network model to learn a large pre-train network model, the large network is called the ‘‘teacher network’’, and the small network is the ‘‘student network’’. Feature distillation expects the student network to achieve accuracy like or even better than the teacher network with fewer parameters and a smaller scale. The FSP matrix is used to characterize the data association between different layers of the ‘‘teacher

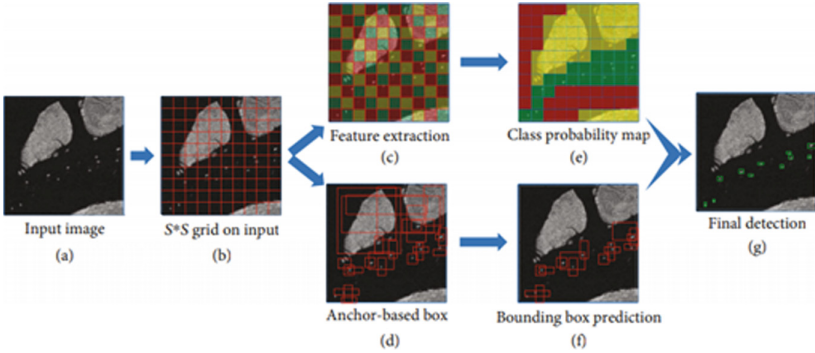


Fig. 2. The proposed YOLO-SARshipNet with feature distillation flow.

network” and the “student network”, and then the $L2$ loss is used to fit the FSP matrix of the corresponding layer of the small model and the FSP matrix of the corresponding layer of the large model, as shown in Fig. 2. The complete Efficient-Net is applied as the teacher model and YOLO-SARshipNet as the student model to construct the FSP matrix of the corresponding layers. The advantage of this algorithm is that it allows the small model to learn the intermediate process and obtain more prior knowledge. We verified the effectiveness of this method through experiments, and the specific analysis is shown in Sect. 3.2.

3 Experimental Results and Discussion

In this section, we verify the effect of YOLO-SARshipNet on SSDD and SAR-Ship-Database. All experiments were done using a single graphics card (NVIDIA GeForce GTX TITAN XP).

3.1 Dataset and Experiment Setup

In the experiment, we tested the effectiveness of the proposed algorithm using SSDD and SAR-Ship-Dataset. The SSDD dataset contains 1160 SAR images with a total of 2358 ships, and the smallest ships marked are 7×7 pixels and the largest are 211×298 pixels. The SAR-Ship-Dataset contains 43,819 SAR image slices of size 256×256 , whose data sources are the Gaofen-3 satellite and the European Sentinel-1 satellite. The SAR images in the two datasets have multiple polarization modes, multiple scenes, and multiple resolutions, which can effectively verify the effect of the ship detection algorithm.

The two datasets are randomly divided into training set, validation set, and test set according to the ratio of 7 : 2 : 1. The SGD optimizer is used to iteratively update the network parameters. The initial learning rate is set to 0.0013 and the decay coefficient is 0.1. A total of 10000 iterations are performed, and the decay is once at the 8000st and 9000st times. The momentum parameter of SGD is set to 0.9. Batch-size for each training is set to 8.

The evaluation index is an important method for effectively measuring the detection effect of the algorithm. Generally, the detection quality of an image is comprehensively evaluated based on the following five aspects: precision, recall, inference time, model size, and mean Average Precision (mAP). We used the following three objective indicators to evaluate the effectiveness of the proposed algorithm:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

$$OA = \frac{TP + FP}{TP + FP + TF + TN} \quad (8)$$

where TP , FP , TN , and FN represent the number of true positives, false positives, true negatives and false negatives, respectively.

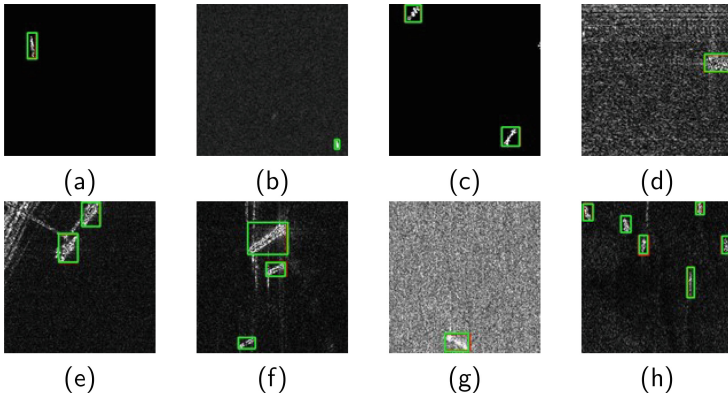


Fig. 3. Visualization of detection results on SSDD dataset.

3.2 Experiment Results

Table 1 shows the ship detection performance of different deep learning-based target detection methods on the SSDD dataset, and compares each method from Recall, Precision, and mAP. In order to ensure fairness, the compared methods all use the uniformly divided data in the experiments in this paper, and the

Table 1. Comparison of detection performance of different methods on SSDD dataset.

Method	Recall (%)	Precision (%)	mAP (%)
Faster-RCNN [10]	75.16	81.15	78.86
RetinaNet [12]	86.70	83.11	85.68
YOLOv3 [10]	93.62	88.54	90.33
YOLOv4 [17]	93.51	87.55	92.38
YOLO-SARshipNet	93.23	90.51	92.44

parameter settings, training methods and training environments are the same as those of YOLO-SARshipNet. It can be found that YOLO-SARshipNet achieves good results compared to other advanced models.

3.3 Ablation Experiment

YOLO-SARshipNet based on YOLOv3. The experiments in this section are to verify the effectiveness of the proposed improved module. This experiment used the SAR-Ship-Dataset dataset, the results are shown in Table 2, and “✓” means that this module is used.

As can be seen from Table 2, YOLO-SARshipNet is about 1/6 of the YOLOv3 model, but the inference time and accuracy are better than YOLOv3. Although the accuracy of the model is reduced by integer quantization, the prediction time is shortened again, and such a lightweight integer network is more convenient for FPGA or DSP porting. And the model overcome the above problems after feature distillation.

Figures 3 and 4 shows the ship detection results of some samples of YOLO-SARshipNet on the SAR-Ship-Dataset dataset, where the red box is the real ship, and the green box is the detected ship. It shows that YOLO-SARshipNet has a good detection effect whether it is a ship target in a blurred background, a near-shore ship, or a small target ship.

Table 2. Ablation experiments on the SAR-Ship-Dataset dataset.

Method	Backbone	FCN	IoU-Loss	Integer quantization	Feature distillation	mAP	Scale (MB)	Reference (ms)
YOLOv3						93.32	139.25	28.32
	✓					95.54	62.44	16.13
	✓	✓				96.23	62.44	17.78
	✓	✓	✓			97.19	62.44	17.78
	✓	✓	✓	✓		96.13	21.38	13.23
	✓	✓	✓	✓	✓	97.34	21.38	13.23
YOLO-SARshipNet	✓	✓	✓	✓	✓	97.34	21.38	13.23

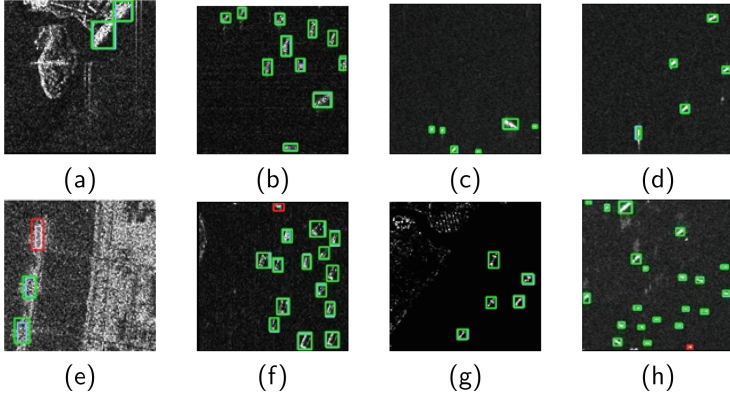


Fig. 4. Visualization of detection results on SAR-Ship-Dataset dataset.

4 Conclusion

This paper proposes a high-speed and high-precision SAR ship detection algorithm based on the YOLOv3 algorithm, namely YOLO-SARshipNet. Through the improvement of the backbone network, the accuracy is improved while reducing the size of the model; the proposed feature pyramid network effectively improves the small target detection ability of the model, and the loss function is modified for the detection of a single target SAR ship target, it also brings a certain improvement to the detection accuracy; finally, the integer quantization operation of the model allows the model to have a smaller number of parameters and calculations, and the feature distillation method has been used to solve the problem of integer quantization. The experimental results on SSDD and SAR-Ship-Dataset show the correctness and effectiveness of the method proposed in this paper, and it has certain practical significance in the real-time SAR application field. It can be transplanted to the hardware of FPGA or DSP later.

References

1. Hui, M., Xiaoqing, W., Jinsong, C., Xiangfei, W., Weiya, K.: Doppler spectrum-based NRCS estimation method for low-scattering areas in ocean SAR images. *Remote Sensing* (2017)
2. Chen, G., Li, G., Liu, Y., Zhang, X.P., Zhang, L.: Sar image despeckling based on combination of fractional-order total variation and nonlocal low rank regularization. *IEEE Trans. Geosci. Remote Sens.* **58**(3), 2056–2070 (2021)
3. Gui, G.: A parzen-window-kernel-based CFAR algorithm for ship detection in SAR images. *8*(3), 557–561 (2011)
4. Jiang, S., Chao, W., Bo, Z., Hong, Z.: Ship detection based on feature confidence for high resolution SAR images. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International* (2012)

5. Wang, C., Bi, F., Liang, C., Jing, C.: A novel threshold template algorithm for ship detection in high-resolution SAR images. In: IGARSS 2016 - 2016 IEEE International Geoscience and Remote Sensing Symposium (2016)
6. Li, J., Qu, C., Shao, J.: Ship detection in SAR images based on an improved faster R-CNN. In: Sar in Big Data Era: Models, Methods & Applications (2017)
7. Wang, Y., Wang, C., Zhang, H., Dong, Y., Wei, S.: A SAR dataset of ship detection for deep learning under complex backgrounds. *Remote Sensing*, 11(7) (2020)
8. Li Jianwei, Q., Changwen, P.S., Bing, D.: Ship target detection in SAR images based on convolutional neural network. *Syst. Eng. Electron. Technol.* **40**(9), 7 (2018)
9. Changhua, H., Chen, C., Chuan, H., Hong, P., Jianxun, Z.: Small target detection of ships. In: SAR images based on deep convolutional neural networks (3) (2022)
10. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement. *arXiv e-prints* (2018)
11. Xiaoling, Z., Tianwen, Z., Jun, S., Shunjun, W.: High-speed and high-precision SAR ship detection based on deep separation convolutional neural network. *J. Radar* **8**(6), 11 (2019)
12. Wang, Y., Wang, C., Zhang, H., Dong, Y., Wei, S.: Automatic ship detection based on retinanet using multi-resolution gaofen-3 imagery. *Remote Sens.*, 11(5) (2019)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. *IEEE* (2016)
14. Tan, M., Le, Q.V.: Efficientnet: rethinking model scaling for convolutional neural networks (2019)
15. Abadi, M., Barham, P., Chen, J., Chen, Z., Zhang, X.: A system for large-scale machine learning. *USENIX Association, Tensorflow* (2016)
16. Lin, T.Y., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. *IEEE Computer Society* (2017)
17. Bochkovskiy, A., Wang, C.Y., Liao, H.: YOLOv4: optimal speed and accuracy of object detection (2020)