Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 2, No. 1, 2025

Ship Target Detection in SAR Images Based on Multi-Scale Dynamic Fusion Model of YOLOv8

Weihe Huang [a], Yinhong Song [b]

School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102616, China
[a] 2108570022077@stu.bucea.edu.cn, [b] 2108570022123@stu.bucea.edu.cn

Abstract: Target detection is an important part of radar image interpretation. The use of SAR technology combined with deep learning network models has become an important means of ship target detection. Compared with traditional methods, deep learning has powerful data processing and feature extraction capabilities. Therefore, this paper proposes a multi-scale dynamic fusion (DFS) detection model for target detection based on YOLOv8. The model includes three parts: multi-scale feature extraction and fusion, dynamic alignment, and attention mechanism feature enhancement. Finally, experiments were conducted on the SAR-ship dataset. The experimental results show that the detection results of the model reach 96.4%.

Keywords: SAR, deep learning, YOLO, Multi-scale dynamic fusion, Ship target detection.

1. Introduction

Synthetic Aperture Radar is an active microwave remote sensing imaging radar that emits coherent electromagnetic waves to illuminate the surface and then receives scattered echoes from surface targets to obtain images [1]. It is usually installed on flying platforms such as satellites, drones and spacecraft. Compared with traditional optical sensors, SAR systems have strong penetration capabilities and can penetrate clouds, haze and rainfall, and are not restricted by lighting conditions. As one of the dominant forces in maritime combat, the real-time and accurate information obtained by ships provides effective support for the military field [2].

Different from traditional methods that rely on artificially designed features. Convolutional Neural Networks have shown great application potential in the field of SAR image target detection due to their powerful end-to-end feature extraction capabilities. Through multi-layer convolution operations, CNN can automatically learn and extract target features, significantly reducing dependence on manual intervention, and performs particularly well in complex and changeable target detection tasks. Current target detection algorithms are mainly divided into two categories: Anchor-Free and Anchor-Based. The former directly completes the positioning and classification of the target through key point detection, center point regression or bounding box parameter prediction. Typical representatives include CenterNet [3] and FCOS [4]. The latter generates a large number of candidate boxes at various positions in the input image by predefining a set of anchor boxes of different scales and aspect ratios. These anchor boxes are subsequently optimized through classification and regression to predict the target category and accurately locate the target position. Classic models include YOLOv3 [5] and SSD [6].

In recent years, remarkable progress has been made in the research of SAR image ship detection, but due to the unique imaging mechanism of SAR, it still faces many challenges in practical application, such as unclear ship target contour information, complex background, strong scattering, and diverse environmental interference. In response, many scholars have proposed innovative ways to improve detection performance and efficiency. Jiarui Zhang [7] proposed a Lightweight SAR Ship Detection Algorithm , which aims to improve the robustness of the detection model through a position enhancement strategy to solve the problem of fuzzy contours of ship targets. Aiming at the problem of large number of parameters in traditional deep learning models, Zhou Long [8] proposed a lightweight convolutional neural network LiraNet by combining dense joining, residual joining and group convolution technologies, which provides a feasible scheme for the deployment of SAR ship target detection on radar equipment.

These studies not only highlight the flexibility and strong adaptability of deep learning models in SAR image object detection, but also demonstrate the important role of advanced technologies such as attention mechanism, Transformer, and dilated convolution in improving the detection accuracy and efficiency, making deep learning methods more suitable for complex and diverse scenarios and real-time applications. Despite the remarkable progress made in the above research, there are still important challenges in the field of SAR ship detection. On the one hand, the sea clutter background and target characteristics of SAR images make the model susceptible to noise interference during the detection process. On the other hand, the resource-constrained nature of mobile devices places higher demands on the parameters of the model. Therefore, in order to solve the above problems, YOLOv8 proposes a multi-scale dynamic fusion SAR ship target detection algorithm, which can effectively improve the detection accuracy while reducing the number of parameters and calculations.

2. Multi-scale Dynamic Fusion Attention Mechanism

This module jointly realizes multi-scale feature extraction and fusion, dynamic feature alignment and feature enhancement of attention mechanism. In the multi-scale feature extraction stage, adaptive maximum pooling and average pooling are performed on larger feature maps, and neighbor interpolation is used to upsample smaller feature maps. Finally, the following three feature maps are spliced along the channel dimension. Figure 1 shows the network structure.

$$l^{'} = AdaptiveMaxPool(l,size)+AdaptiveAvgPool(l,size) \quad (1)$$

$$s^{'} = Interpolate(s, size, \mathrm{mod}\, e = 'nearest') \quad (2)$$

$$output=Concat(l^{'}, m, s^{'}, \dim = 1) \quad (3)$$

In the feature alignment stage, dynamic convolution is used to replace the original interpolation for medium-scale features (P4) and large-scale features (P5). This process can adjust the resolution of the feature map to achieve the goal of accurate alignment. The aligned feature map uses convolution to perform three-dimensional feature fusion, and finally the fused feature map is dimensional compressed to obtain a multi-scale feature map.

$$P_4^{'} = \mathrm{Conv}(P_4), P_4^{''} = DySample(P_5, scale = 4) \quad (4)$$

$$P_5^{'} = Conv(P_5), P_5^{'} = DySample(P_5, scale = 4) \quad (5)$$

$$combine = Concat(P_3^{3d}, P_4^{''3d}, P_5^{''3d}) \quad (5)$$

$$output = Sequeeze(\max pool(output_{3d}, \dim = 2)) \quad (7)$$

In the attention mechanism stage, global pooling and 1×1 convolution used to build channel attention, and then the weighted feature map is generated channel by channel. Then, local spatial attention is achieved through average pooling, convolution and Sigmoid activation in the height and width directions, thereby refining the spatial features.

$$x_1^{'} = channel(x_1), \quad x_{fusion} = x_1^{'} + x_2 \quad (8)$$

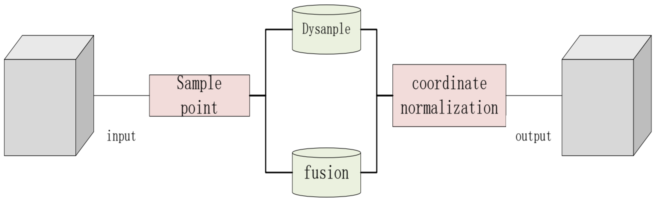

Figure 1. DFS model network structure.

3. Experiment

Based on the SAR data of GF-3 and Sentinel-1 in China, the SAR-Ship-Dataset includes 102 GF-3 images and 108 Sentinel-1 images, constructing a total of 1693 high-resolution ship images and annotating 43819 ship instances [9].

The evaluation indicators used in this experiment included average precision, recall, precision, and model parameters.

Mean Average Precision is a key indicator to measure the overall performance of the object detection model, and the higher the mAP, the better the comprehensive performance of the model when detecting different target classes.

$$mAP = \int_0^1 P(R)dR \quad (9)$$

Recall indicates the proportion of positive samples that can be correctly detected by the model to all actual positive samples, and a higher recall rate indicates that the lower the missed detection rate of the model.

$$\mathrm{Re}\,call = \frac{TP}{TP + FP} \quad (10)$$

Precision indicates the proportion of positive samples in the detection results predicted by the model to be positive, and the higher the precision, the lower the false detection rate of the model in the prediction.

$$\mathrm{Pr}\,ecision = \frac{TP}{TP + FN} \quad (11)$$

Table 1. Experiment comparison

v8n	DFS	P (%)	R (%)	mAP (%)
√		91.7	92.7	94.6
√	√	95.8	91.3	96.4

As shown in Table 1, we can clearly see that on the SAR-ship dataset, the detection accuracy of ship targets has been improved, and the number of parameters has been reduced on the original basis. This result shows the effectiveness of the model in target detection.

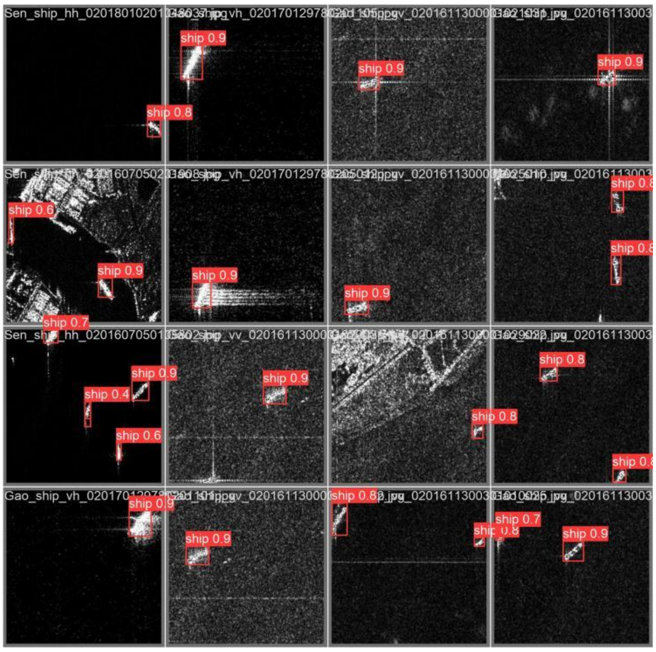

Figure 2. Prediction results of SAR-ship datasets

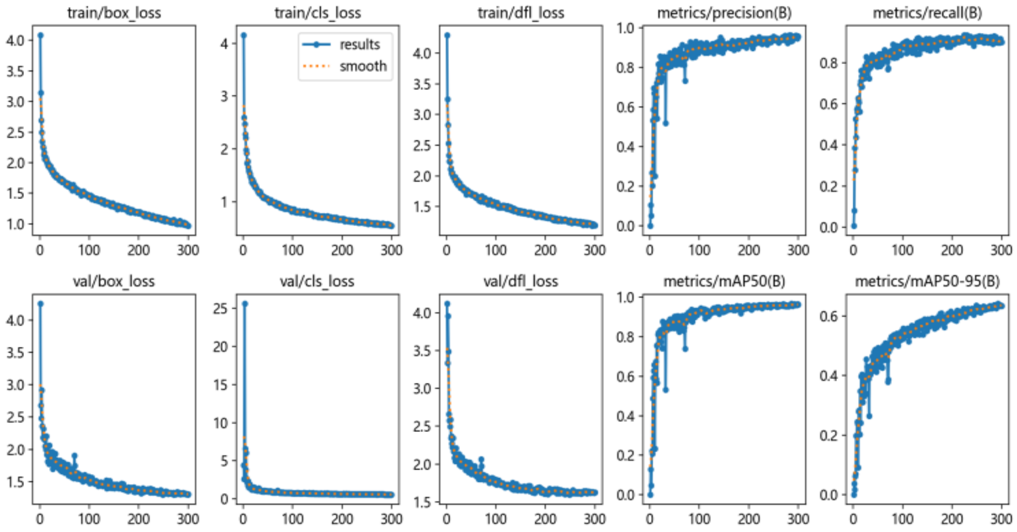

Figure 3. loss function of SAR-ship datasets

After the model training was completed, as shown in Figure 2-3, the model can stably detect ship targets regardless of the strong clutter background, showing strong anti-interference ability; for smaller ship targets, the model can also accurately locate and identify them, avoiding the common problem of small object detection difficulties.

4. Conclusion

In summary, the dynamic fusion of scales detection model proposed in this paper significantly reduces the number of model parameters and computational complexity while ensuring detection accuracy. Experimental results show that the model has achieved excellent detection performance on the SAR-ship dataset, the detection accuracy is improved to 96.4%.

References

[1] Gao, W.; Liu, Y.; Zeng, Y.; Liu, Q.; Li, Q. SAR Image Ship Target Detection Adversarial Attack and Defence Generalization Research. Sensors 2023, 23, 2266, doi:10.3390/s23042266.

[2] Song, C.; Deng, J.; Liu, Z.; Wang, B.; Wu, Y.; Bi, H. Complex-Valued Sparse SAR-Image-Based Target Detection and Classification. Remote Sensing 2022, 14, 4366, doi:10.3390/rs14174366.

[3] Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. CenterNet: Keypoint Triplets for Object Detection 2019.

[4] Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection 2019.

[5] Terven, J.; Cordova-Esparza, D. A Comprehensive Review of YOLO: From YOLOv1 and Beyond 2023.

[6] Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In; 2016; Vol. 9905, pp. 21–37.

[7] Zhang, J.; Chen, Z.; Yan, G.; Wang, Y.; Hu, B. Faster and Lightweight: An Improved YOLOv5 Object Detector for Remote Sensing Images. Remote Sensing 2023, 15, 4974, doi:10.3390/rs15204974.

[8] Long, Z.; Suyuan, W.; Zhongma, C.; Jiaqi, F.; Xiaoting, Y.; Wei, D. Lira-YOLO: A Lightweight Model for Ship Detection in Radar Images. Journal of Systems Engineering and Electronics 2020, 31, 950–956, doi:10.23919/JSEE.2020.000063.

[9] Wang, Y.; Wang, C.; Zhang, H.; Dong, Y.; Wei, S. A SAR Dataset of Ship Detection for Deep Learning under Complex Backgrounds. Remote Sensing 2019, 11, 765, doi:10.3390/rs11070765.