Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 12, No. 2, 2024

A new method of ship detection in complex background of SAR images based on YOLO

Xinyu Zhang, Hua Huo *

College of Information Engineering, Henan University of Science and Technology, Kaiyuan Avenue 263, Luoyang, 471023, China
* **Corresponding author**: Hua Huo

Abstract: Synthetic Aperture Radar (SAR) is a remote sensing technology capable of high-resolution imaging of the ground surface under any weather and light conditions. SAR image ship detection is a technique that uses the ship target features in SAR images to identify and locate ships, which is of great value for applications in the fields of maritime traffic management, marine environment monitoring, and maritime security and defence. However, SAR image ship detection also faces some challenges, such as complex sea surface background, low signal-to-noise ratio, and the diversity and density of ship targets, which lead to the shortcomings of traditional feature extraction and classifier-based detection methods in terms of accuracy and efficiency. To address these challenges, this paper proposes a new method for ship detection in SAR images based on YOLOv7. When dealing with SAR images containing clutter interference and complex backgrounds, many models face the dual challenge of insufficient real-time and accuracy. To address this problem, this study proposes a new PCAN aggregation network, which firstly introduces the SimAM attention mechanism, which significantly enhances the model's ability to identify the differences between ships and background clutter, further improves the ability of focusing on target features, and effectively It further improves the ability to focus on target features, effectively reduces the interference of background noise, and improves the detection accuracy in complex scenes. At the same time, the PConv strategy is added to further lighten the model structure, thus accelerating the computation process and improving the inference efficiency. We trained and tested the model on the SAR-Ship-Dataset dataset, and achieved an accuracy of 91.1%, an F1 score of 92.9, and an FPS of 46, demonstrating its excellent detection capability. By comparing and analysing the model with several classical and cutting-edge methods in the current field, the proposed model in this study demonstrates obvious performance advantages, and in addition, through ablation experiments, we further confirm the effectiveness and contribution of PCAN.

Keywords: Synthetic Aperture Radar (SAR); YOLOv7; Ship detection; Small target.

1. Introduction

With the development of water transport and ship marine technology, marine vessel inspection plays an indispensable role in both civil and military fields [1-4], examples include maritime traffic control, rescue patrols, marine environmental protection, and so on. In particular, the rapid economic development in recent years has become increasingly important for improving maritime defence and the ability to respond effectively to complex and changing situations at sea, and therefore there is a growing demand for the use of artificial intelligence methods and information technology to improve regulatory and countermeasure capabilities [8-10]. However, due to the complex and changing environment of the sea surface, the traditional optical images are highly susceptible to be affected, thus failing to capture clear pictures. Synthetic aperture radar (SAR) is a kind of active microwave imaging sensor that is all-weather, all-day, and is not affected by conditions such as weather and light, which is very suitable for the field of maritime vessel monitoring[12-14], and SAR images have the following advantages: (1) Capable of high-resolution imaging of the earth's surface under any weather and light conditions, unaffected by clouds, haze, night and other factors;(2) Being able to provide geometric, physical, electromagnetic and other multi-faceted information of the ground surface, which enriches the content of remote sensing data;(3) It can realise large-scale, long-time and continuous observation, which improves the coverage and timeliness of remote sensing data. Detection of ship targets in SAR imagery has a wide range of applications, and in military applications, SAR imagery ship detection is used for maritime surveillance, target tracking and reconnaissance [15, 16]. In civilian applications, SAR image ship detection is used for search and rescue operations, maritime security and environmental monitoring [17-20]. In maritime surveillance, SAR image ship detection is an important tool for monitoring shipping activities in coastal waters, identifying potential threats, and detecting illegal activities such as smuggling and piracy. SAR image ship detection can achieve the acquisition of information such as the number, type, position, and speed of ships at sea, thus providing a basis for maritime traffic planning, maritime rescue, and maritime law enforcement; it can also achieve the discovery and monitoring of maritime pollution sources, illegal fishing, and smuggling and drug trafficking, thus providing support for marine environmental protection and maintenance of maritime law and order; even in poor weather conditions and low light Even in bad weather conditions and low-light periods, SAR images can also provide high-resolution ship images, making it a valuable asset for surveillance operations. In search and rescue operations, SAR imagery ship detection is used to locate ships in distress and assist in the rescue of people in distress. SAR imagery can provide accurate information about the target location and size of a ship, helping search and rescue teams to co-ordinate their efforts effectively. In terms of maritime safety, SAR imagery ship detection is used to monitor ship activities and prevent accidents such as collisions and groundings. SAR imagery can provide information about the location and movement of ships, enabling the authorities to take appropriate measures to ensure the safety of maritime traffic. Detection and identification of enemy or suspicious targets can also be achieved through SAR image ship detection, thus providing reference for strategic defence at sea, tactical attack at sea, and so on. In environmental monitoring, SAR imagery ship detection is used to monitor and track oil spills, monitor marine ecosystems, and detect changes in the coastline. SAR imagery can provide high-resolution images of the affected area, enabling the authorities to take appropriate measures to minimise the impact on the environment and prevent further damage. Therefore, SAR imagery maritime vessel detection is of critical research value.

Traditional ship target recognition methods for SAR images can be summarised into three main categories: first, ship target detection based on the statistical distribution of background clutter. Constant False Alarm Rate (CFAR) [21-24] is one of the most widely used ship target detection algorithms, which determines the probability density function to obtain the false alarm rate by statistical modelling of the background clutter. Secondly, ship target detection based on polarisation decomposition. Taking advantage of the different scattering mechanisms of the ship and the sea, the secondary scattering of the ship target is further improved by de-rotation angle processing, while the polarised scattering mechanism of the sea surface is mainly based on single scattering, thus obtaining the ship detector. Third, ship target detection based on polarisation features [26-28]. In addition to the common clutter false alarm, the azimuthal direction blur caused by the ship target itself is also a difficult problem in ship target detection. In order to eliminate the effect of orientation blurring, Wang et al used the polarisation feature approach for ship detection based on the fact that ship targets possess complex metal structures and all ships are a mixture of scattering mechanisms such as single scattering, secondary scattering, and depolarisation scattering.

With the rapid development of artificial intelligence in recent years, deep convolutional neural networks (CNNs) have achieved great success in many computer vision tasks such as image classification, target detection, semantic segmentation, and so on. The current mainstream target detection algorithms can be divided into two categories: one category is single-stage target detection algorithms, which only need a convolutional neural network to directly obtain the class probability and position coordinates of different targets, and representative algorithms include YOLO, SSD, and so on. The other class is two-stage target detection algorithms, which firstly need to use Selective search or Region Proposal Network (RPN) to generate proposed regions, and then regression classification is performed in these proposed regions, representative algorithms are R-CNN, Faster R- CNN, etc. Ship detection algorithms for SAR images have also been researched by a large number of scholars. Sun et al proposed a SAR ship detector based on bi-directional feature fusion and angular classification of YOLO (BiFA-YOLO), which employs a novel bi-directional feature module (Bi-DFFM) to efficiently aggregate multi-scale features for ship detection. Li et al proposed an algorithm for ship detection based on the Faster R-CNN, the lightweight Lite Faster R-CNN is proposed by fully considering the local and global multi-scale feature information through the hopping connection, which makes the network more suitable for the feature extraction of SAR images. Cui et al proposed embedding the attention mechanism into the FPN while using the dense connection, which achieves a good detection. Zhu et al proposed a bilateral repeating feature pyramid (DB-YOLO) SAR ship detector, which is improved for small-scale ship targets and improves the detection real-time. Sun et al proposed a lightweight densely-connected sparse-activation detector (DSDet) for ship target detection, which utilises a densely-connected sparse-activation network to constructing a lightweight backbone network, thus achieving a balance between performance and computational complexity, however, the accuracy is not particularly satisfactory. Tang et al investigated a noise level classifier (NLC) module and a SAR target potential area extraction (STPAE) module. The NLC module implements the extraction of the SAR image noise level for classification, and the STPAE is used to extract the potential target's complete region, combined with YOLOv5 achieved good detection results, but in real-time is not quite able to meet the needs of practical application scenarios.

Despite the better performance of deep learning approaches compared to traditional methods, SAR image ship detection still faces a number of challenges [43, 44], including the following:

(1) Complex sea surface background: The sea surface background in SAR images often contains a variety of factors such as waves, sea ice, islands, coastlines, etc. These factors will produce echo signals that are similar or coherent with the ship target, leading to a reduction in the contrast between the ship target and the background, which increases the difficulty of detection;

(2) Low signal-to-noise ratio. The noise in SAR images mainly includes speckle noise, scattering noise, quantisation noise, etc. These noises reduce the clarity and quality of SAR images, affect the extraction of edge and detail features of ship targets, and reduce the accuracy of detection;

(3) Diversity and density of ship targets. ship targets in SAR images have different types, sizes, shapes, orientations, attitudes and other features, which cause ship targets to present different radar scattering characteristics in SAR images, increasing the complexity of detection;

(4) Ship targets in SAR images are often densely distributed in harbours, wharves, waterways and other areas, and ship targets in these areas will obscure or overlap each other, which leads to difficulties in positioning and segmentation of the detection frame, and increases the detection error.

In summary, how to maintain the detection accuracy while meeting the needs of real-time detection, so that the speed and accuracy to achieve a good balance is one of the important problems that need to be solved. Therefore, to address the problems and challenges of ship detection in SAR images, this paper proposes a high-speed and lightweight detection network based on YOLOv7, which improves and optimises the YOLOv7 model for the characteristics of the ship targets in SAR images, and improves the detection speed with the same high accuracy, and achieves quite good results. The main contributions of our work are as follows:

(1) Aiming at the problem of SAR image ship detection in complex sea background, this paper proposes an efficient aggregation network of PCAN, which further reduces the number of parameters on the lightweight model to speed up the detection speed, and at the same time fuses the three-dimensional attentional weights to improve the accuracy of detecting SAR image ships.

(2) Reduced redundant computation and frequent memory access requirements by introducing partial convolution, accelerating the overall computation process and extracting

spatial features more effectively.

(3) The 3D attention weights are inferred without adding parameters to the original network by introducing the SimAM attention mechanism, which further enhances the model's ability to recognise and learn target features without increasing the computational burden.

(4) In this paper, the proposed model is experimented on the SAR-Ship-Dataset dataset, which possesses better robustness compared with other state-of-the-art methods, and achieves better results in terms of combined speed and accuracy performance.

The rest of the paper is organised as follows. Section II describes our proposed module and model in detail. In Section III, we present the experimental details and conduct experiments using two publicly available datasets, compare them with other classical and advanced detection models to validate the effectiveness of our model, and show the visualisation results. Section IV conducts an ablation study to verify the usefulness of each module. Finally, conclusions are drawn and summarised in Section V.

2. Proposed Method

In the face of the actual scenario of ship detection at sea, the ability to make full use of the limited computational resources and processing power, the use of higher detection accuracy, lower model complexity, faster detection of the lightweight detector is the key to the ship detection.YOLOv7 is one of the most advanced target detection algorithms at present, with a very high speed and accuracy, the main network structure of the YOLOv7 can be roughly divided into YOLOv7-tiny, YOLOv7, YOLOv7X, YOLOv7-D6, YOLOv7-W6, and YOLOv7-E6. Among them, YOLOv7-tiny has the smallest network model and the fastest speed, which is the most suitable for ship detection from the practical scene. However, since the YOLOv7 series is designed for optical image detection, it is not fully competent in the field of SAR image ship detection, so we have designed and improved the YOLOv7-tiny network based on the YOLOv7-tiny network, which is able to specialise in the characteristics of this field, and achieve better detection results.

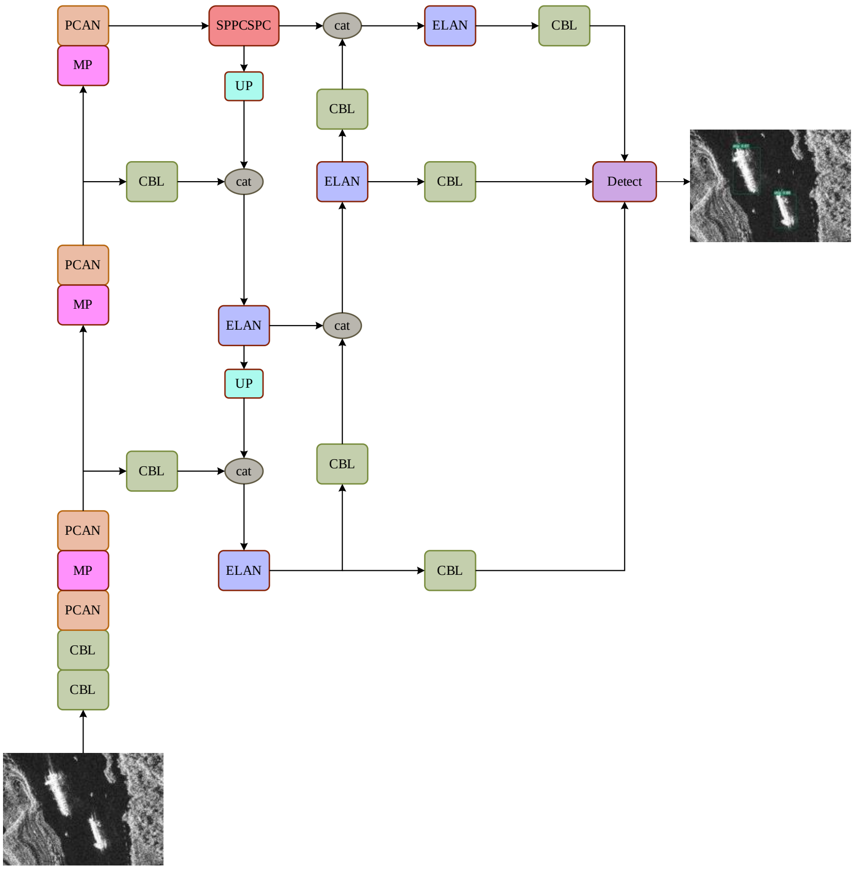

Fig. 1 Schematic diagram of the network structure.

The overall structure of our proposed model is shown in Fig. 1: it consists of Input, Backbone, Head and Output. Backbone is the part used for feature extraction, which consists of CBL module, partial convolutional and attention-free aggregation networks (PCAN) and MP, which ultimately outputs three different scales of feature maps (C3, C4, C5). Among them, CBL is a basic unit composed of Conv, BN and LeakyReLU activation functions, which are used to change the number of channels or feature extraction. The PCAN module is designed on the basis of the ELAN module, which, as an innovation in YOLOv7, demonstrates an efficient design of network structure. Its core idea is to make the network capture richer feature information during training by precisely regulating the shortest and longest gradient paths in the network. This not only improves the learning ability of the network, but also makes it more robust when facing various complex scenarios. The PCAN module cleverly employs the partial convolution PConv , which replaces the conventional convolution Conv. The introduction of partial convolution effectively reduces the redundant computations in the network and reduces the number of frequent memory accesses, which significantly improves the computational efficiency of the network. . More importantly, this design makes the network more accurate and efficient in extracting spatial features, which provides a strong guarantee for precise target location and identification. Simultaneously, the SimAM attention mechanism is added to further improve the network performance, and the specific schematic diagram is shown in Fig. 2.

Head: Head is the part used for feature fusion and up-sampling, which consists of the SPPCSPC module, the ELAN module and the CBL module, which ultimately outputs three feature maps at different scales (P3, P4, P5).The SPPCSPC module is a module consisting of the Spatial Pyramid Pooling Layer (SPP) and the Cross Stage Partial Connection Layer (CSP), which is used to increase the receptive field and reduce the computational effort. The SPP layer is used to obtain the features of different receptive fields through the maximum pooling layer for adapting to different scales of targets. While in the design of the CSP layer, the features are divided into two independent parts. On the one hand, one part of the features is computed by traditional processing, on the other hand, the other part of the features is specially processed by using the SPP structure, and finally these two parts are fused in the CSP layer. This design not only significantly reduces the amount of computation by about half and improves the computational efficiency, but also effectively improves the accuracy of the model, so that the whole system achieves a higher performance level while maintaining high efficiency.

Output: the Output is the part used for target detection, which consists of three CBL modules and finally outputs three different scales of detection results (D3, D4, D5). Compared to YOLOv7, it eliminates the use of REP-heavy parameterisation modules, making it more lightweight, and finally uses IDetect for prediction, including the location, category and confidence of the target.

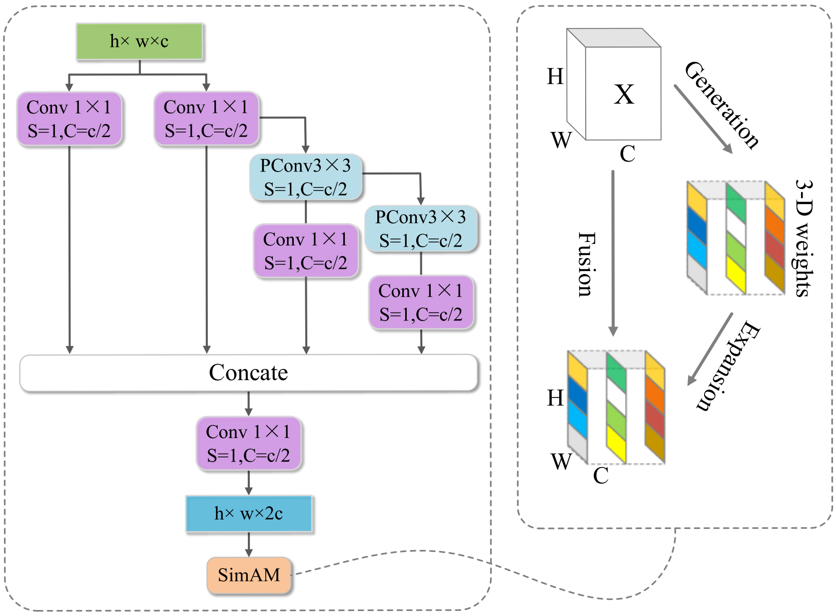

Fig. 2 PCAN module structure diagram.

2.1. Partial Convolution

Partial Convolution (PConv), as an advanced convolution operation technique, has the main goal of reducing the redundancy of traditional convolution in terms of computation and memory access, the principle of which is illustrated in Fig. 3. PConv achieves this by selectively convolving a portion of the input channel, thus exploiting the high similarity of feature maps across channels to point out the redundancy that exists in the traditional methods. redundancy that exists. By implementing partial ratios (e.g., 1/4), the need for floating point operations (FLOPs) is significantly reduced, effectively improving computational efficiency without compromising feature extraction. This strategy is crucial for optimising traditional deep learning models in resource-limited situations.

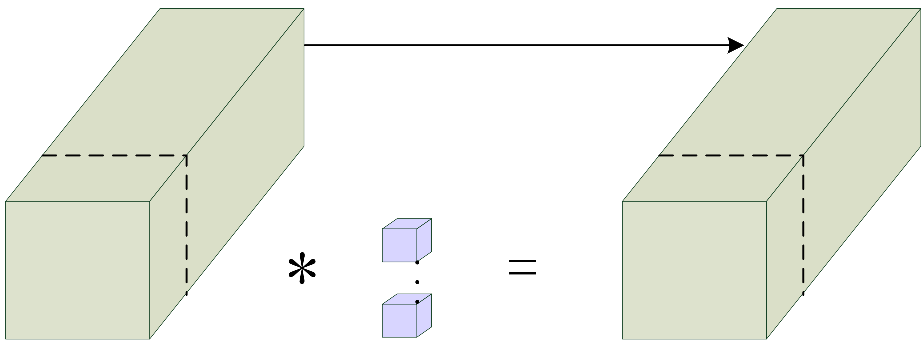

Fig. 3 PConv module structure diagram.

The implementation of PConv has not only achieved significant gains in computational efficiency, but has demonstrated its effectiveness in a variety of ways. It significantly reduces the required computational resources while maintaining or enhancing the model performance. Particularly noteworthy is that it demonstrates extreme efficacy in extracting spatial features, which is crucial for tasks requiring complex processing such as image classification, detection, and segmentation. In addition, PConv, as a core part of the FasterNet architecture, supports a series of networks designed to provide efficient and fast processing capabilities that significantly reduce latency while maintaining accuracy. These networks have demonstrated superior performance and broad applicability in numerous vision tasks, especially on challenging datasets. In SAR images, due to the specificity of radar imaging, there are often noisy, interfering, or missing data regions, which can be disruptive during target detection. By implementing partial ratios, PConv is able to automatically ignore these unreliable regions and perform convolutional operations on only valid pixels, enhancing the representation of features and thus improving target detection accuracy. PConv also has a faster computational speed, PConv reduces unnecessary computation by optimising the convolution operation and improves the operational efficiency of the model.

2.2. Attention mechanism

SimAM (Simplified Attention Module) is a simple and extra-parameter-free attention module for enhancing the feature representation of convolutional neural networks. Unlike traditional attention modules, it is able to infer 3-D attention weights for each neuron of the feature map that are variable in both channel and space without adding any parameters to the original network. This approach takes into account both spatial and channel dimensions so that each neuron is assigned a unique weight. This module is based on neuroscience theory and determines the importance of each neuron by optimising an energy function and deriving a fast closed-form solution of this energy function. The advantage of SimAM is that most operators are chosen based on the solution of the energy function, avoiding the complexity of structural adjustments. Firstly we define the following energy function for each neuron:

$$h_t(p_t, b_t, y, x_i) = (y_t - \hat{t})^2 + \frac{1}{N-1}\sum_{i=1}^{N-1}(y_m - \hat{x}_i)^2 \quad (1)$$

where $\hat{t} = p_t t + b_t$ and $\hat{x}_i = p_t x_i + b_t$ are linear transformations of t and x_i, where t and x_i are the target neuron and other neurons in a single channel of input feature $X \in \mathbb{R}^{C\times H\times W}$. $N = H \times W$ is the number of neurons in the channel. When $\hat{t} = y_t$, minimize the above formula, adopt binary labels for y_t and y_m and add regularization terms, to obtain the final energy function:

$$h_t(p_t, b_t, y, x_i) = \frac{1}{N-1}\sum_{i=1}^{N-1}(-1-(p_t x_i + b_t))^2 + (1-(p_t t + b_t))^2 + \lambda p_t^2 \quad (2)$$

There are N energy functions in each channel, and the above formula can be solved as follows:

$$p_t = -\frac{2(t-\mu_t)}{(t-\mu_t)^2 + 2\varphi_t^2 + 2\lambda} \quad (3)$$

$$b_t = -\frac{1}{2}(t+\mu_t)p_t \quad (4)$$

The μ_t and φ_t included in this formula are the mean and variance of all neurons except the target neuron, respectively. Thus the minimum energy can be calculated by the following equation:

$$h_t^* = \frac{4(\hat{\varphi}^2 + \lambda)}{(t-\hat{\mu})^2 + 2\hat{\varphi}^2 + 2\lambda} \quad (5)$$

The above equation shows that the lower the energy, the more the neuron t is distinguished from the surrounding neurons and the more important it is for visual processing, and the importance of each neuron can be obtained from $\frac{1}{h_t^*}$.

Finally, the features are refined by a scaling operation based on the attentional conditioning, as shown in Eqs. (6). In summary, the proposed attention mechanism can be implemented by adding after the convolutional layer and can be used with existing networks.

$$\tilde{X} = sigmoid(\frac{1}{H}) \odot X \quad (6)$$

2.3. Loss function

The IoU loss function is used to measure the degree of overlap between the predicted box (bounding box) and the actual box (ground truth box), and its main role is to improve the accuracy of the model in predicting the bounding box (position and size) of the object, and it is one of the commonly used loss functions in the target detection task. In the YOLOv7 series of models, CIoU is used as the loss function. The CIoU loss function, based on the IoU loss function, takes more factors into account when dealing with the bounding box regression problem by adding the centroid distances and aspect ratios, so as to be able to measure the similarity between the predicted box and the real box in a more comprehensive way.

The CIoU loss consists of three main components: the IoU loss, the centroid distance and the consistency of the aspect ratio. The formula is as follows:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b_{gt})}{c^2} + \alpha v \quad (7)$$

where b and b_{gt} represent the coordinates of the centroids of the predicted and real frames, respectively, $\rho(\cdot)$ denotes the Euclidean distance, c is the length of the diagonal of the smallest closed frame containing the two frames, and αv is the aspect ratio consistency conditioning term, which is designed to reduce the difference between the predicted frame and the real frame in terms of aspect ratio. Here, v measures the aspect ratio consistency and is defined as:

$$v = \frac{4}{\pi^2}(\arctan\frac{w_{gt}}{h_{gt}} - \arctan\frac{w}{h})^2 \quad (8)$$

where w and h represent the width and height of the predicted box, respectively, and w_{gt} and h_{gt} are the width and height of the real box. The parameter α is a weight coefficient to adjust the importance of the aspect ratio consistency term, defined as:

$$\alpha = \frac{v}{(1-IoU)+v} \quad (9)$$

The CIoU loss provides a more comprehensive approach to evaluating the difference between the predicted frame and the real frame by taking into account the IoU, the relative position of the centroids, and the consistency of the aspect ratio. This approach has a number of advantages over IoU loss, firstly by penalising the distance between the predicted frame and the centroid of the real frame, CIoU loss drives the predicted frame generated by the model to be closer to the position of the real frame, achieving a more accurate positioning. Secondly, by adding an aspect ratio consistency term, the CIoU loss helps the model to predict frames with similar aspect ratios to the real frames, which is especially important for objects with diverse shapes. Finally, by dynamically adjusting the weight of the aspect ratio consistency term, the CIoU loss is able to balance the effects of IoU, centroid distance and aspect ratio in different scenarios, so as to maintain high positioning accuracy in various scenarios.

3. Experiment

3.1. Experimental condition

All experiments were run on a 64-bit Windows 10 21H2 computer with an Intel(R) Core(TM) i7-9700F CPU @ 3.00GHz, 16GB of RAM, and an NVIDIA GeForce RTX 2070 SUPER with 8GB of GPU, and were performed using the Python 3.8 computer language on the PyCharm 2022 platform, developed using the deep learning framework Pytorch and configured with the corresponding CUDA and CUDNN.

3.2. SAR datasets

In order to verify the universality of the network model, the SAR-Ship-Dataset dataset is used for testing in this experiment; SAR-Ship-Dataset uses China's domestic Gaofen-3 SAR data and Sentinel-1 SAR data as the main data source, and a total of 102 views of Gaofen-3 and 108 views of Sentinel-1 SAR images are used to A total of 102 scenes of Gaofen-3 and 108 scenes of Sentinel-1 SAR images are used to construct a deep learning sample library of high-resolution SAR ship targets. At present, the deep learning sample library contains 43819 ship slices. The imaging models of Gaofen-3 are Strip-Map (UFS), Fine Strip-Map 1 (FSI), Full Polarisation 1 (QPSI), Full Polarization 2 (QPSII) and Fine Strip-Map 2 (FSII). The resolutions of these five imaging models are 3 m, 5 m, 8 m, 25 m, and 10 m. The imaging modes of Sentinel-1 are the Strip-Map (S3 and S6) and Wide-Map modes, and the detailed information is shown in Table 1.

Table 1. Details of the SAR-Ship-Dataset.

Sensor	Imaging Mode	Resolution Rg. × Az.(m)	Swath (km)	Incident Angle(°)	Polarization	Number of Images
GF-3	UFS	3 × 3	30	20~50	Single	12
GF-3	FS1	5 × 5	50	19~50	Dual	10
GF-3	QPSI	8 × 8	30	20~41	Full	5
GF-3	FSII	10 × 10	100	19~50	Dual	15
GF-3	QPSII	25 × 25	40	20~38	Full	5
Sentinel-1	SM	1.7 × 4.3 to 3.6 × 4.9	80	20~45	Dual	49
Sentinel-1	IW	20 × 22	250	29~46	Dual	10

The dataset contains a rich variety of ship slices with diverse backgrounds, suitable for a wide range of SAR image applications. The image resolution covers 1.7 m to 25 m. The polarisation modes include HH, HV, VH and VV, and the imaging modes include ultra-fine stripe mode, fine stripe mode, fully polarised stripe mode, stripe scanning mode, and interferometric wide mode. The dataset scenarios include harbours, nearshore, islands and offshore, and the types include all kinds of common ship targets, such as oil tankers, bulk carriers, large container ships and fishing vessels.

3.3. Evaluation metrics

In order to quantitatively evaluate the detection effect of the model, the main metrics of the computer vision evaluation algorithm are P (precision) and R (recall), defined as:

$$P = \frac{\text{TP}}{\text{TP+FP}} \tag{10}$$

$$R = \frac{\text{TP}}{\text{TP+FN}} \tag{11}$$

Where TP is the number of ships correctly labelled by ship samples, FP is the number of non-ship samples labelled as ships and FN is the number of ship samples labelled as non-ship targets. In addition to this, we calculated the FPS (Frames Per Second) and F1-score, defined as:

$$FPS = 1/T \tag{12}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{13}$$

Where T is the detection time of a single image, FPS denotes the number of images detected per second on the test platform, and the F1 score is a combination of both P and R metrics, which allows for the evaluation of the comprehensive performance of different models.

3.4. Analysis and discussion of results

In order to comprehensively evaluate the effectiveness of the model proposed in this study and its practical potential, we conducted extensive experiments on the SAR-Ship-Dataset public dataset and compared the performance of the model with that of classical target detection algorithms. We comprehensively considered four key performance metrics: accuracy, recall, F1 score, and processing speed FPS, and the detailed experimental results are shown in Table 2.

Table 2. Experimental comparison results with the classical model on the SAR-Ship-Dataset.

Method	SAR-Ship-Dataset			
	P	R	F1	FPS
SSD	76.4	92.8	83.1	30.5
Faster R-CNN	82.5	84.3	83.4	13.8
Cascade R-CNN	80.2	93.5	86.3	13.7
RetinaNet	85.3	84.6	84.9	21.6
CenterNet	78.1	95.9	86.1	17.3
YOLOv5s	87.1	87.4	87.2	39
YOLOv7	87.7	85.9	86.7	37
ours	92.1	93.7	92.9	46

Based on the experimental results on the dataset, the effectiveness of our improved network model can be well illustrated. It guarantees higher detection speed in terms of FPS and significantly improves in terms of accuracy and recall, and has better detection performance in general. Compared to the classic YOLOv5s model, the accuracy improves from 87.1% to 92.1% and the recall improves from 87.4% to 93.7%, and our network model still has better results in all aspects. Compared to YOLOv7, we guarantee a higher FPS of the model while achieving improved performance. It is well known that the YOLOv7 network model is larger, so the training time is very long, the convergence speed is slow, and the FPS is lower under the same configuration, our model achieves about the same model size as the YOLOv7-tiny and the YOLOv5s, and gets the best detection results. Compared to other models, comparing Faster R-CNN, the accuracy is improved by 9.6% and the recall is improved by 9.4%. Comparing Cascade R-CNN, the accuracy is improved by 11.9% and the recall is improved by 0.2%. Among them, Faster R-CNN and Cascade R-CNN are two-stage detection models, which are slow and cannot meet the demand of real-time detection, and their FPS is only 13.8 and 13.7 on SAR-Ship-Dataset dataset, and there is no advantage in accuracy for SAR images. As for the SSD classical model, its accuracy is relatively low, with an experimental result of 76.4%, and at the same time, the FPS is not high, which makes it difficult to be applied to real-world scenarios. Compared to RetinaNet, the accuracy rate is improved by 6.8%, the recall rate is improved by 9.1%, and the FPS is also greatly improved. Compared to CenterNet, the accuracy is improved by 14%, and although our recall is slightly lower, the FPS is improved significantly, so it is obvious that our network model has a very significant advantage and better overall performance. In summary, our proposed model achieved better detection results when comparing with classical methods, proving the effectiveness of our algorithm and reflecting the superiority of our model.

In order to synthesise the performance of our algorithm, similarly we compare it with other state-of-the-art methods in recent years and the detailed results are shown in Table 3.

Table 3. Experimental comparison results with state-of-the-art methods on SAR-Ship-Dataset.

Method	SAR-Ship-Dataset			
	P	R	F1	FPS
Quad-FPN	77.6	96.1	85.9	23
EFGRNet	78.5	95.8	86.3	30.3
CenterNet+	83.5	97.6	90	30.3
CI-Net	88.1	90.3	89.2	34
DCMSNN	88.1	86.6	87.3	9
DAPN	87.3	93.4	90.2	21.5
ARPN	88.1	92	90	21.5
DCN	86.3	93.1	89.6	20.6
HR-SDNet	93.3	92.1	92.7	8.9
PANET	86.9	93.4	90	23.2
ours	92.1	93.7	92.9	49.3

Specifically, our method achieves 92.1% accuracy on the SAR-Ship-Dataset dataset, which is 14.5%, 13.6%, and 8.6% higher than the Quad-FPN, EFGRNet, and CenterNet+ methods, respectively. Our method is lower in recall, but the F1 score shows that our method is better overall, relatively more balanced, and much higher than them in FPS. As for models such as CI-Net, DCMSNN, DAPN and ARPN, we manage to take the lead across the board, with an accuracy of 93.3% compared to HR-SDNet, which is slightly higher than our method, but is far away from us in terms of detection speed. On the F1 score, a composite metric, our method achieves 92.9, which is the highest among all models. With a processing speed of 46 FPS, our method not only excels in performance metrics, but is also the fastest of all methods in terms of speed. This is in contrast to DCMSNN (9 FPS) and HR-SDNet (8.9 FPS), which perform well in some performance metrics but have slower processing speeds and may not be able to satisfy the demand for real-time processing.

3.5. Experimental visualisation of results

In order to deeply show the performance of the methods introduced in this chapter in real target detection tasks, we visualise and analyse the experimental results to be able to have an intuitive proof of performance, as shown in Fig. 4 showing the detection results in different scenarios.

By analysing the visual detection results of the dataset, we can find that: firstly, as a comparison, the first image in the dataset shows a detection scenario in a relatively calm sea environment without potential interference factors such as near-shore complex sea background, environmental factors and clutter. Under these ideal conditions, the model is able to easily identify the ship target. After that, there are detection scenarios containing various environmental disturbances and complex backgrounds, and most of the existing models often face the problem of wrong detection or inaccurate recognition when facing these scenarios. Comparatively speaking, our proposed model is able to maintain a high detection accuracy even under these challenging conditions, showing its robustness and effectiveness. In practical application scenarios, this is of great significance for application areas such as marine management and ship monitoring.

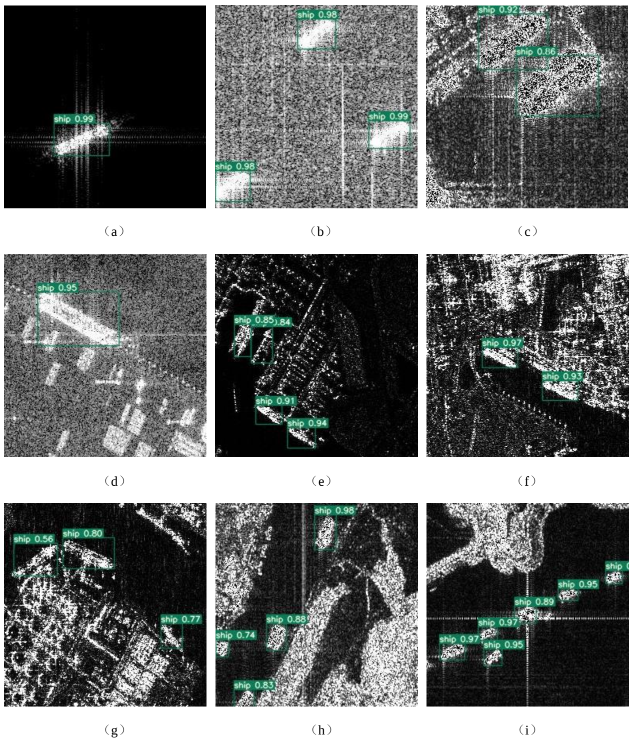

Fig. 4 Schematic diagram of the visual detection of the SAR-Ship-Dataset.

4. Experimental studies of ablation

In order to refine the importance and effectiveness of each module, for this we conducted an ablation study and analysed and discussed each module individually. We used the original YOLOv7-tiny as the base model and then added different modules for comparison to verify the actual effectiveness of the modules, and the experimental results are shown in Table 4.

Table 4. Comparison of ablation experiments on the SAR-Ship-Dataset.

Method	PConv	SimAM	P	R	F1	FPS
YOLOv7-tiny			86.3	83.2	84.7	45
ours	√		87.9	86.4	87.1	49.6
		√	90.1	91.6	90.8	45.4
	√	√	92.1	93.7	92.9	49.3

The second line "√" means adding PConv convolution to replace the regular Conv, which has the biggest advantage of reducing FLOPs, speeding up computation, and also extracting spatial features more efficiently, while again maintaining a high accuracy. It can be seen that there is a significant improvement in FPS compared to the original model after joining, and there is also a slight improvement in accuracy and recall. The third line "√" means that the SimAM attention mechanism is added, which can help the model to locate and recognise objects in the image more accurately when the background is complex or there is occlusion, and it is found that the accuracy and other indexes are significantly improved after the addition in the experiment. The addition of these new modules brings different degrees of improvement in accuracy, recall and F1 score. After incorporating all of our proposed improvement strategies, on the SAR-Ship-Dataset dataset, relative to the original model, our method improves the accuracy by 5.8%, recall by 10.5%, and FPS by 4.3, all of which are significant improvements over the original YOLOv7-tiny model, proving the proposed improvement strategies' effectiveness of the proposed improved strategy.

5. Conclusions

Aiming at the problems that the existing models are not strong enough in real time and not accurate enough in SAR image ship detection, this paper proposes a new ship detector applicable to SAR images. The detector not only ensures fast and lightweight detection, easy deployment, and improved detection speed and accuracy, but also ensures excellent detection in the face of situations with clutter, complex background interference, and so on. This is mainly attributed to our proposed PCAN module, which introduces the SimAM attention mechanism to better distinguish between background clutter and the detection target and resist their interference; meanwhile, the addition of PConv takes a step further in lightweighting to achieve faster computation and inference. We perform quantification and experiments on the most representative SAR-Ship-Dataset dataset and compare it with a representative variety of classical and state-of-the-art methods, and our model achieves better performance than other mainstream detection models, while we also perform ablation experiments to verify the effectiveness of our proposed module. Our method maintains a balance of high accuracy and recall in identifying targets, and ranks first in F1 score among all compared methods, showing the superiority of the overall performance. While maintaining high performance, our method is also able to process the image data at an extremely fast speed, which makes it suitable for real-time analysis applications. Therefore, our proposed SAR image ship detection model is of great importance for future ship detection.

Despite the comprehensive performance of our new model, there are still some minor problems, for example, in near-shore scenarios, it may misclassify few scenes on land as ships, therefore, in the future, we will continue to engage in research in this area, and broadly speaking, the work is as follows:

(1) Go further in lightweighting and reduce the weight of the detection model.

(2) Further reduce the detection time without sacrificing accuracy.

(3) Optimise the detection model, mainly for the problem of false alarms that tend to occur in land scenes, to further reduce the false alarm rate, and to minimise inland interference.

References

[1] Lang H, Xi Y, Zhang X. Ship Detection in High-Resolution SAR Images by Clustering Spatially Enhanced Pixel Descriptor [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(8): 5407-5423.

[2] Chen S, Zhan R, Wang W, Zhang J. Learning Slimming SAR Ship Object Detector Through Network Pruning and Knowledge Distillation [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 1267-1282.

[3] Leng X, Ji K, Kuang G. Radio Frequency Interference Detection and Localization in Sentinel-1 Images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(11): 9270-9281.

[4] Zhang P, Luo H, Ju M, et al. Brain-Inspired Fast Saliency-Based Filtering Algorithm for Ship Detection in High-Resolution SAR Images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-9.

[5] Wang C, Jiang S, Zhang H, et al. Ship Detection for High-Resolution SAR Images Based on Feature Analysis [J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11(1): 119-123.

[6] Carmassi C, Porta I, Bertelloni C A, et al. PTSD and post-traumatic stress spectrum in the Italian Navy Operational Divers Group and corps of Coast Guard Divers employed in search and rescue activities in the Mediterranean refugees emergences and Costa Concordia shipwreck [J]. J Psychiatr Res, 2020, 129: 141-146.

[7] Zhang, Zhang, Shi, Wei. Depthwise Separable Convolution Neural Network for High-Speed SAR Ship Detection [J]. Remote Sensing, 2019, 11(21).

[8] Song S, Xu B, Yang J. SAR Target Recognition via Supervised Discriminative Dictionary Learning and Sparse Representation of the SAR-HOG Feature [J]. Remote Sensing, 2016, 8(8).

[9] Yang F, Xu Q, Li B. Ship Detection From Optical Satellite Images Based on Saliency Segmentation and Structure-LBP Feature [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(5): 602-606.

[10] Salembier P, Liesegang S, Lopez-Martinez C. Ship Detection in SAR Images Based on Maxtree Representation and Graph Signal Processing [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(5): 2709-2724.

[11] Moreira A, Prats-Iraola P, Younis M, et al. A tutorial on synthetic aperture radar [J]. IEEE Geoscience and Remote Sensing Magazine, 2013, 1(1): 6-43.

[12] Chang Y-L, Anagaw A, Chang L, et al. Ship Detection Based on YOLOv2 for SAR Imagery [J]. Remote Sensing, 2019, 11(7).

[13] Liu T, Zhang J, Gao G, et al. CFAR Ship Detection in Polarimetric Synthetic Aperture Radar Images Based on Whitening Filter [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(1): 58-81.

[14] Tang G, Liu S, Fujino I, et al. H-YOLO: A Single-Shot Ship Detection Approach Based on Region of Interest Preselected Network [J]. Remote Sensing, 2020, 12(24).

[15] Brusch S, Lehner S, Fritz T, et al. Ship Surveillance With TerraSAR-X [J]. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49(3): 1092-1103.

[16] Feng Y, Chen J, Huang Z, et al. A Lightweight Position-Enhanced Anchor-Free Algorithm for SAR Ship Detection [J]. Remote Sensing, 2022, 14(8).

[17] Solberg A H S, Storvik G O, Solberg R, Volden E. Automatic detection of oil spills in ERS SAR images [J]. IEEE Trans Geosci Remote Sens, 1999, 37: 1916-1924.

[18] Kang M, Ji K, Leng X, Lin Z. Contextual Region-Based Convolutional Neural Network with Multilayer Fusion for SAR Ship Detection [J]. Remote Sensing, 2017, 9(8).

[19] Tings B, Bentes C, Velotto D, Voinov S. Modelling ship detectability depending on TerraSAR-X-derived metocean parameters [J]. CEAS Space Journal, 2018, 11(1): 81-94.

[20] Wang Y, Wang C, Zhang H. Combining a single shot multibox detector with transfer learning for ship detection using sentinel-1 SAR images [J]. Remote Sensing Letters, 2018, 9(8): 780-788.

[21] Jiang Q, Aitnouri E, Wang S, Ziou D. Automatic Detection for Ship Target in SAR Imagery Using PNN-Model [J]. Canadian Journal of Remote Sensing, 2000, 26: 297 - 305.

[22] Seong-In H, Ouchi K. On a Novel Approach Using MLCC and CFAR for the Improvement of Ship Detection by Synthetic Aperture Radar [J]. IEEE Geoscience and Remote Sensing Letters, 2010, 7(2): 391-395.

[23] Kronauge M, Rohling H. Fast Two-Dimensional CFAR Procedure [J]. IEEE Transactions on Aerospace and Electronic Systems, 2013, 49: 1817-1823.

[24] Biao H, Xingzhong C, Licheng J. Multilayer CFAR Detection of Ship Targets in Very High Resolution SAR Images [J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(4): 811-815.

[25] Sugimoto M, Ouchi K, Nakamura Y. On the novel use of model-based decomposition in SAR polarimetry for target detection on the sea [J]. Remote Sensing Letters, 2013, 4(9): 843-852.

[26] Kaplan L M. Improved SAR target detection via extended fractal features [J]. IEEE Transactions on Aerospace and Electronic Systems, 2001, 37: 436-451.

[27] Huang X, Yang W, Zhang H, Xia G-S. Automatic Ship Detection in SAR Images Using Multi-Scale Heterogeneities and an A Contrario Decision [J]. Remote Sens, 2015, 7: 7695-7711.

[28] Huo W, Huang Y, Pei J, et al. Ship Detection from Ocean SAR Image Based on Local Contrast Variance Weighted Information Entropy [J]. Sensors (Basel, Switzerland), 2018, 18.

[29] Wang C, Wang Y, Liao M. Removal of azimuth ambiguities and detection of a ship: using polarimetric airborne C-band SAR images [J]. International Journal of Remote Sensing, 2011, 33(10): 3197-3210.

[30] Wang S, Sun L, Fan W-L, et al. An automated CNN recommendation system for image classification tasks [J]. 2017 IEEE International Conference on Multimedia and Expo (ICME), 2016: 283-288.

[31] Bappy J H, Roy-Chowdhury A K. CNN based region proposals for efficient object detection [J]. 2016 IEEE International Conference on Image Processing (ICIP), 2016: 3658-3662.

[32] Deng L, Yang M, Qian Y, et al. CNN based semantic segmentation for urban traffic scenes using fisheye camera [J]. 2017 IEEE Intelligent Vehicles Symposium (IV), 2017: 231-236.

[33] Redmon J, Divvala S K, Girshick R B, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection [J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 779-788.

[34] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector; proceedings of the European Conference on Computer Vision, F, 2015 [C].

[35] Girshick R B, Donahue J, Darrell T, Malik J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [J]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2013: 580-587.

[36] Ren S, He K, Girshick R B, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39: 1137-1149.

[37] Sun Z, Leng X, Lei Y, et al. BiFA-YOLO: A Novel YOLO-Based Method for Arbitrary-Oriented Ship Detection in High-Resolution SAR Images [J]. Remote Sensing, 2021, 13(21).

[38] Li Y, Zhang S, Wang W-Q. A Lightweight Faster R-CNN for Ship Detection in SAR Images [J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.

[39] Cui Z, Li Q, Cao Z, Liu N. Dense Attention Pyramid Networks for Multi-Scale Ship Detection in SAR Images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57: 8983-8997.

[40] Zhu H, Xie Y, Huang H, et al. DB-YOLO: A Duplicate Bilateral YOLO Network for Multi-Scale Ship Detection in SAR Images [J]. Sensors (Basel), 2021, 21(23).

[41] Sun K, Liang Y, Ma X, et al. DSDet: A Lightweight Densely Connected Sparsely Activated Detector for Ship Target Detection in High-Resolution SAR Images [J]. Remote Sensing, 2021, 13(14).

[42] Tang G, Zhuge Y, Claramunt C, Men S. N-YOLO: A SAR Ship Detection Using Noise-Classifying and Complete-Target Extraction [J]. Remote Sensing, 2021, 13(5).

[43] Cui J, Jia H, Wang H, Xu F. A Fast Threshold Neural Network for Ship Detection in Large-Scene SAR Images [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 6016-6032.

[44] Wang H, Zhang S, Li Y. A Robust and Lightweight Detector for Ship Target with Complex Background in SAR Image [J]. Journal of Sensors, 2022, 2022: 1-16.

[45] Wang C-Y, Bochkovskiy A, Liao H-Y M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors [J]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 7464-7475.

[46] Chen J, Kao S-H, He H, et al. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks [J]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 12021-12031.

[47] Yang L, Zhang R-Y, Li L, Xie X. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks [Z]//MARINA M, TONG Z. Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research; PMLR. 2021: 11863--11874

[48] Zheng Z, Wang P, Liu W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression; proceedings of the AAAI Conference on Artificial Intelligence, F, 2019 [C].

[49] Wang Y, Wang C, Zhang H, et al. A SAR Dataset of Ship Detection for Deep Learning under Complex Backgrounds [J]. Remote Sensing, 2019, 11(7).

[50] Cai Z, Vasconcelos N. Cascade R-CNN: Delving Into High Quality Object Detection [J]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017: 6154-6162.

[51] Lin T-Y, Goyal P, Girshick R B, et al. Focal Loss for Dense Object Detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 42: 318-327.

[52] Zhou X, Wang D, Krähenbühl P. Objects as Points [J]. ArXiv, 2019, abs/1904.07850.

[53] Zhang T, Zhang X, Ke X. Quad-FPN: A Novel Quad Feature Pyramid Network for SAR Ship Detection [J]. Remote Sens, 2021, 13: 2771.

[54] Nie J, Anwer R M, Cholakkal H, et al. Enriched Feature Guided Refinement Network for Object Detection [J]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 9536-9545.

[55] Guo H, Yang X, Wang N, Gao X. A CenterNet++ model for ship detection in SAR images [J]. Pattern Recognit, 2021, 112: 107787.

[56] Gao J-K. Research on ship frameless detection in SAR image based on multi-scale feature fusion [D]. Liaoning Technical University, 2022.

[57] Jiao J, Zhang Y, Sun H, et al. A Densely Connected End-to-End Neural Network for Multiscale and Multiscene SAR Ship Detection [J]. IEEE Access, 2018, 6: 20881-20892.

[58] Cui Z, Li Q, Cao Z, Liu N. Dense Attention Pyramid Networks for Multi-Scale Ship Detection in SAR Images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57: 8983-8997.

[59] Zhao Y, Zhao L, Xiong B, Kuang G. Attention Receptive Pyramid Network for Ship Detection in SAR Images [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 2738-2756.

[60] Dai J, Qi H, Xiong Y, et al. Deformable Convolutional Networks [J]. 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 764-773.

[61] Wei S, Su H, Ming J, et al. Precise and Robust Ship Detection for High-Resolution SAR Imagery Based on HR-SDNet [J]. Remote Sens, 2020, 12: 167.

[62] Liu S, Qi L, Qin H, et al. Path Aggregation Network for Instance Segmentation [J]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.