Dual-domain Joint Dense Multiple Small Ship Target Detection Algorithm for Spaceborne SAR Images

JIA Peng[1], *DONG Tiancheng*[2], *WANG Taoyang*[3], *ZHANG Guo*[2], *SHENG Qinghong*[1], *LI Jun*[1*]

1. College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P. R. China; 2. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, P. R. China; 3. School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, P. R. China

(Received 10 April 2024; revised 9 November 2024; accepted 30 November 2024)

Abstract: Ship detection via spaceborne synthetic aperture radar (SAR) has become a research hotspot. However, existing small ship detection methods based on the radar signal domain and SAR image features cannot obtain highly accurate results because of the obvious coherent speckle noise at sea and strong reflection interference from near-shore objects. To resolve the above problems, this study proposes a dual-domain joint dense multiple small ship target detection method for spaceborne SAR image that simultaneously detects objects in the image and frequency domains. This method uses an attention mechanism module and algorithm structure adjustments to improve the small ship target feature mining ability. In the frequency-based image generation, extreme signal strength values are detected in the azimuth and range directions, with the results of the two complementing each other to realize dual-domain joint small ship target detection. The comprehensive qualitative and quantitative evaluation results show that the proposed method can attain a final precision rate of 92.25% and achieve accurate results for SAR ship detection in open-sea, coastal, and port area ships. The test results for the self-built SAR small-ship dataset demonstrate the effectiveness and universality of the method.

Key words: synthetic aperture radar (SAR); small ship detection; deep learning; attention module; YOLO; dual-domain joint

CLC number: V19 **Document code:** A **Article ID:** 1005-1120(2024)06-0725-14

0 Introduction

Ship detection methods based on synthetic aperture radar (SAR)[1] images primarily leverage the statistical distribution of vessels and the amplitude, phase, and texture characteristics of the images. Numerous traditional detection methods, such as the constant false alarm rate (CFAR) algorithm[2-3] and automatic sea and land segmentation[4] algorithms, emerged during the early developmental stages[5] and demonstrated high detection accuracy in experiments conducted on typical SAR images. However, these methods often rely on statistical distribution[6] of ships and the range information of the images, limiting their effectiveness in complex scenes and failing to meet the requirements for high-precision detection[7]. In contrast, recent advancements in ship detection[8-10] have been driven by deep learning techniques applied to SAR images, yielding satisfactory results. Deep learning-based object detection methods are typically categorized into two-stage and single-stage approaches[11]. Despite the prevalence of small targets in SAR data, object detection methods encounter challenges when targeting small ships. Consequently, achieving accurate detection of small ships in SAR images remains a significant and ongoing research endeavor[12].

*Corresponding author, E-mail address: jun.li@nuaa.edu.cn.

How to cite this article: JIA Peng, DONG Tiancheng, WANG Taoyang, et al. Dual-domain joint dense multiple small ship target detection algorithm for spaceborne SAR images[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2024,41(6):725-738.

http://dx.doi.org/10.16356/j.1005-1120.2024.06.005

In cases where the SAR image resolution is low or the ship's actual size is small, the ship may manifest as a bright spot within the image. Leveraging these distinctive characteristics, scholars have successfully detected small ship targets in SAR images by incorporating attention mechanism modules, conducting multi-scale feature parallel mining, and deploying enhanced multi-directional algorithms[13-14]. The attention mechanism stands at the forefront of deep learning network optimization, enhancing the feature extraction capability of backbone networks. Consequently, it has emerged as a preferred technique for refining small target detection methods in SAR images of ships. For instance, Lin et al.[15] introduced a squeeze excitation (SE)[16] mechanism to enhance network feature extraction. However, the SE module focuses solely on channel correlations, overlooking the significance of spatial information. Addressing this limitation, Zhao et al.[17] proposed an attention receptive pyramid network (ARPN) that integrates the convolutional block attention module (CBAM) introduced by Woo et al[18]. This integration aims to mitigate the adverse impact of surrounding environments. Nonetheless, the CBAM module is limited to capturing local correlations and fails to establish long-range feature dependencies.

In addition to the attention mechanism, scholars have explored multiscale parallel methods for optimization. For instance, Li et al.[19] introduced the lightweight faster region-convolutional neural network (Faster R-CNN), which leverages parallel multiscale convolution operations to extract feature information effectively. Similarly, Zhang et al.[20] proposed a lightweight SAR ship detector that shares features across different detection scales via upsampling and downsampling. This approach has shown promising results in detecting multiscale targets amidst complex backgrounds. However, the above methods do not focus on losing position information for high-level feature of small objects in the networks. Therefore, when the information used for position refinement is mapped to the final feature map, a significant amount of data is lost.

However, the aforementioned methods do not address the issue of losing information about small targets in deep feature maps. Consequently, during the mapping of position refinement information onto the final feature map, a notable amount of data is lost. Moreover, due to the similarity in scattering mechanisms between certain local areas of ship targets in SAR images and their surrounding regions, as well as the resemblance of image features between small ship targets and shoreline features, the aforementioned algorithms tend to produce numerous false alarms in near-shore areas. Furthermore, as the data pass through the convolutional neural network model, the deep convolutional layer may inadvertently discard valuable feature information about the boat, thereby increasing the likelihood of errors in small-target detection[21].

In addition to the small-target detection methods based on SAR images mentioned earlier, Ding et al.[22-23] proposed an innovative dual-domain joint detection method for identifying small moving targets in video SAR data. This approach combines high-resolution SAR images with low-resolution range-Doppler (RD) spectrum data, supplementing missing RD spectrum target energy in the image and synchronously extracting target image features and energy. However, the method primarily focuses on the RD spectrum during frequency spectrum processing, specifically employing fast Fourier transform solely in the azimuthal direction, neglecting the distance-range to the target energy. Consequently, when maritime targets are close to land or other objects with similar reflectivity, similar object interference may arise, as observed in the aforementioned video SAR scenario. Furthermore, while the concept of dual-domain joint detection shows promise, its application in various scenarios warrants further exploration and development.

The attention mechanism fails to exhibit long-range dependence in detecting small ship targets in SAR images. Building upon this observation and insights derived from previous optimization strategies, this paper introduces a novel dual-domain joint SAR image small-ship target detection algorithm for addressing numerous challenges, including a high incidence of false alarms in multiscale parallelism and incomplete detection of single-directional frequency

spectrum target energy. The dual domains considered in this study encompass the original SAR image domain and the frequency domain generated by Fourier transform in two directions. Firstly, to mitigate feature loss in high-level convolutional networks encountered in prior SAR image domain processing studies, region proposals were enhanced[24] to incorporate two-scale downsampling layers within the spatial pyramid pooling (SPP) module integrated into the feature extraction modules of the YOLOv4 network. Additionally, a dilated convolutional layer was introduced to augment the receptive field, discern small-scale target features, and reduce the false alarm rate. Secondly, to mitigate similar reflections on the shore, one-dimensional intensity regional maximum value detection was independently performed, retaining and positioning the targets with regional maximum value in both range and azimuth directions. Finally, the two domains were merged to yield the final detection outcomes. Evaluation on various public datasets (SSDD and SAR-ship-dataset) demonstrated the proposed method's ability to achieve robust detection of small ship targets across diverse environments, including open seas and near-shore regions.

The subsequent sections of this paper are organized as follows: Section 1 provides an overview of relevant literature encompassing key factors examined in this study, such as existing radar signals, deep learning methods for small-target detection, and SAR small-target datasets. It also outlines the initial steps taken to address the research objectives. Section 2 delves into efforts to enhance the YOLO detection method and frequency detection algorithm. Section 3 details the experimental setup and analysis of results. Finally, Section 4 offers the concluding remarks of this paper.

1 Related Work

1.1 Existing SAR ship datasets

Existing public SAR ship datasets include SAR-ship-dataset[25], SSDD[26], and AIR-SARShip-1.0/2.0[27]. In the SAR-ship-dataset, the majority of ships exhibit relatively small and widely varying scales, and some examples of the SAR-ship-dataset are shown in Fig.1. Fig.2 illustrates the distribution of bounding box pixels for different ship sizes: 35 025 ships with 0—1 000 pixels are classified as small targets, 23 125 ships with 1 000—4 000 pixels as medium targets, and 1 385 ships as large targets. Similar data distributions are observed in the SSDD and AIR-SARShip-1.0 datasets. However, existing research has not yet reached the necessary level of detection capability for small-ship targets in SAR images, with mainstream algorithms primarily focusing on medium and large ships.

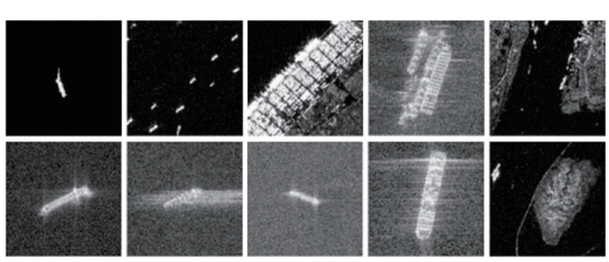


Fig.1 Examples of the SAR-ship-dataset

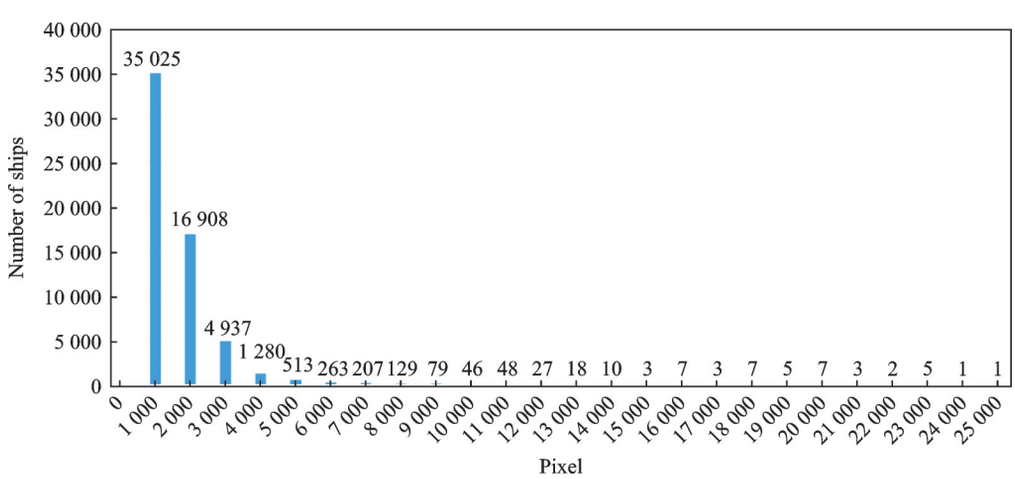


Fig.2 Distribution of bounding box pixels for ships of different sizes in SAR-ship-dataset

1.2 Existing deep learning-based methods for (small) object detection

1.2.1 Deep learning-based small object detection

The detection of small targets has always been a challenging but popular topic in the field of target detection. A small target in an image refers to one with a small area in the image. The main reasons for the difficulty in detection are that: (1) Small-target features are scarce; (2) the number of datasets is small; (3) the design of anchor boxes is difficult; and (4) deep network features can easily be lost. Mainstream solutions include the pyramid structure, backbone network, anchor frame design, optimization objective, and gain component (attention mechanism module). Existing algorithms such as the YOLO series, through the combination of the feature extraction network CSPDarknet53 and the spatial pyramid pooling layer, have better realized the detection of large-and small-scale targets and have become the improved basic network for many deep-learning small-target detection methods.

This solution is also applicable to the small-target detection of SAR images. In SAR images, small targets can be defined as targets with less than 1 000 bounding box pixels. The bounding box is indicated by the red area in Fig.3.

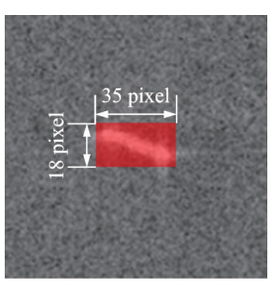


Fig.3 Diagram of the bounding box

1.2.2 YOLO algorithm

The YOLOv4[28] algorithm proposed in 2020 is based on the YOLOv3 algorithm, which integrates the ideas and training techniques of the algorithm model with the advantages obtained in deep learning neural networks in recent years. This algorithm can simultaneously compensate for many defects, thereby achieving dramatic improvements in speed and detection accuracy.

In YOLOv4, the parameters (the pixel coordinates x and y, width w, and height h of the anchor box) in YOLO-Head are not independent and require a loss function to express their interrelationships. Therefore, the complete intersection over union (CIoU)[29] was adopted, which achieved better results.

Target box loss should consider three important geometric factors: Overlap area, center point range, and aspect ratio. However, the distance-IoU (DIoU)[30] in YOLOv2 and YOLOv3 does not consider the aspect ratios of the detection boxes. Detection boxes with similar aspect ratios should exhibit lower losses. Therefore, CIoU was used in this study, shown as

$$\mathrm{CIoU}=1-\mathrm{IoU}+\frac{\rho^2(b,b^{\mathrm{gt}})}{c^2}+\alpha\nu \tag{1}$$

where b and b^{gt} mean the central points of anchor box and target box, ρ means the distance between b and b^{gt}, c the diagonal length of the smallest enclosing box covering two boxes, α the positive measurement parameter, and ν the consistency of the aspect ratio. Thus, the overlapping area factor had a higher priority in the regression, especially in nonoverlapping scenarios.

YOLOv5, launched a month after YOLOv4, introduces significant improvements in its network architecture, especially with adaptive anchor boxes that are automatically learned from the training data, offering an advantage in detecting small, fixed-size objects. While YOLOv5 is more flexible and faster, YOLOv4 struggles with weak features for small targets, leading to a trade-off between accuracy and efficiency in this study, which chooses to modify YOLOv4 for better precision. The YOLO series algorithm processes target detection by dividing the image into $S\times S$ (S means the size of the grid) grids and predicting target presence in each, followed by selecting the most appropriate frame. YOLOv4 enhances YOLOv3's feature extraction by using CSPDarknet53, Mish activation and Dropblock regularization, and introduces SPP and path aggregation network (PAN) for handling varying image sizes and maintaining a fixed output value. YOLOv4 retains YOLOv3's detection head, resulting in a model with CSPDarknet53, SPP-PANet, and YOLO-Head components. The structure of the YOLOv4 network is illustrated in Table 1.

Table 1 YOLOv4 structure components

Step	Content
Backbone	Feature extractor Darknet-53 or CSPDarknet-53 composes the backbone, which is responsible for extracting rich feature representations from input images based on multiple convolutional layers and residual connections; thus, it can effectively capture features at different scales and semantic levels.
Neck	SPP and PAN modules compose the neck and fuse features from the backbone network to obtain higher-level semantic information. SPP captures context information at different scales by pooling the feature maps, while PAN achieves feature fusion via cascading and cross-layer connections, thereby enhancing the accuracy and multi-scale perception capability of object detection.
Head	A multi-scale prediction strategy is used for the head, which is responsible for generating bounding boxes and class probabilities for targets, with each scale having an independent detection head. Each detection head includes a series of convolutional layers and fully connected layers.
Output	Loss functions, such as IoU loss, class balance, and CIoU loss, compose the output. Non-maximum suppression (NMS) is applied to predicted bounding boxes and class probabilities to filter overlapping bounding boxes. This process outputs the positions (bounding box coordinates) and class probabilities of the detected objects, resulting in the final detection results.

1.3 Detection of radar one-dimensional signals and two-dimensional images

Radar data primarily consist of radar echo signal data, which can be transformed into various representations including one-dimensional high-resolution range data, two-dimensional RD data, and SAR images. Each dimension of data representation exhibits distinct characteristics, leading to variations in detection methods based on each data point.

Mainstream radar two-dimensional image detection methods can be categorized into RD image detection, micro-Doppler image (M-D) detection, and SAR image detection. The RD map is generated from the original radar echo signal through fast Fourier transform in the fast time domain to provide range dimension information. Subsequently, Fourier transform is applied to each range unit in the slow time domain to generate the RD spectrum map, revealing the energy distribution of echoes in the RD domain. The RD spectrum proves effective in distinguishing energy differences between targets and interference. Both domestic and international researchers have initiated target detection based on the RD spectrum to separate complex targets from the background, significantly reducing the false alarm detection rate, as depicted in Fig.4.

SAR image detection stands as the prevailing direction for two-dimensional target detection in radar systems. With the ongoing evolution of deep-learning algorithms, a growing array of deep-learning methods are finding applications in this domain. However, the overall performance of deep-learning algorithms in SAR lags behind that of the computer vision field. Additionally, due to inherent disparities between optical and SAR images, many detection algorithms lack necessary adjustments to accurately interpret the fundamental characteristics of SAR imagery. Moreover, research and application pertaining to target scattering mechanisms remain relatively insufficient in comparison.

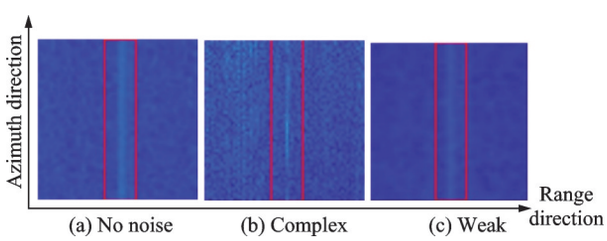


Fig.4 Detection performance of RD spectrum in various situations

2 Methodology Improvement

Inspired by the idea of two-dimensional RD spectrum detection, this study attempts to mine the characteristics of small targets using the azimuth-time distance-frequency domain data and range-time azimuth-frequency domain data derived from the SAR intensity information. This work further supplements the results of SAR image detection and realizes SAR small ship target detection in complex environments with high background interference.

2.1 Overall algorithm workflow

The algorithm flow of this study (Fig.5) comprises three main steps: data preprocessing, dual-track object detection, and collaborative processing. Compressing the SAR image in the azimuth and range directions to yield the range-time domain azimuth-frequency domain data and the range-frequency domain and azimuth-time domain data.

Dual-track object detection: In the image track, the SAR image serves as the data source to construct a dataset, and the improved YOLOv4 algorithm is utilized for training and verification. Meanwhile, the frequency track conducts one-dimensional signal detection on the compressed results in the azimuth and range directions to detect extreme energy values.

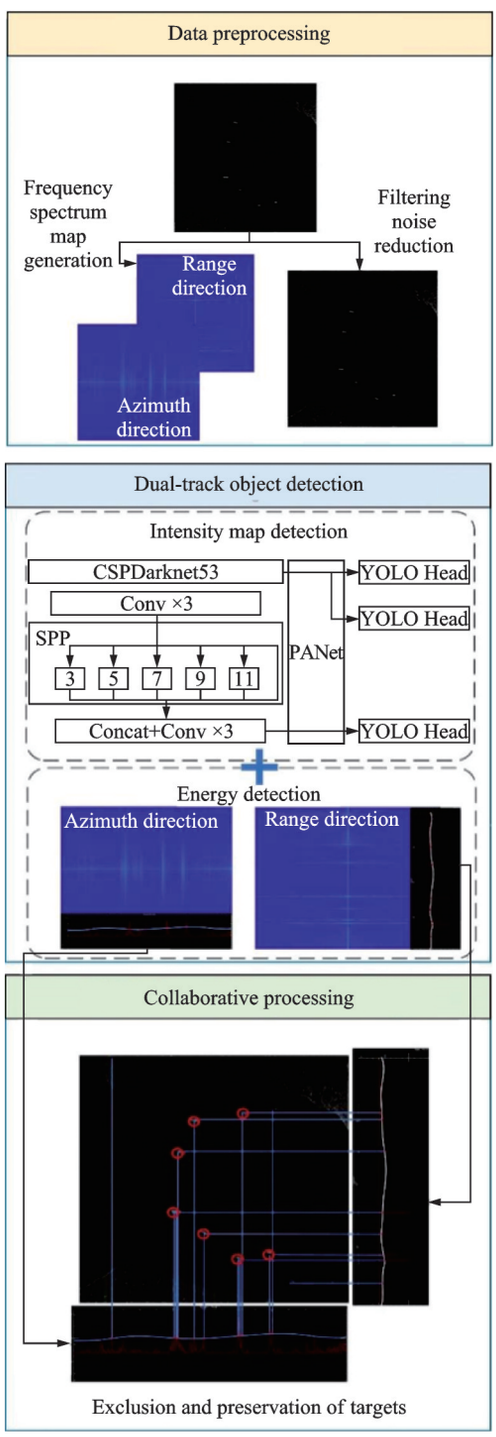


Fig.5 Overall algorithm flowchart

Collaborative processing: Targets identified in the image track and those at the energy extremes in the frequency track are synthesized. Weak signal targets are eliminated, while targets missed by image detection are supplemented to enable dual-domain collaborative detection.

In contrast to prior work, this study emphasizes feature mining in the frequency domain rather than relying solely on image features. This approach reduces the miss rate associated with single-image feature mining. The method proposed in this study aims to offer a more dependable detection scheme for measured data.

2.2 Improvement of the YOLO algorithm

2.2.1 Attention mechanism

During deep learning image analysis, all features and regions in an image are typically considered equally important. However, attention models, such as the one described in Ref.[31], autonomously learn the importance of each feature and region. By assigning higher weights to important features and regions, attention models enhance the feature extraction capability of deep learning algorithms without significantly increasing computational complexity or parameters[32].

One of the leading attention mechanism modules is CBAM[33], which builds upon the concepts of Senet while integrating spatial and channel features. CBAM comprises two key modules: the spatial attention mechanism module and the channel attention mechanism module.

In the channel attention mechanism module, global maximum pooling and global average pooling layers are employed to condense the spatial dimensions of the input feature map, yielding two one-dimensional feature vectors. These vectors are then fed into a shared neural network with ReLU ac-

tivation function. After combining the output values and applying a sigmoid activation function, a new feature map is generated.

The specific formula is as follows

$$M_c(F)=\sigma(\mathrm{MLP}(\mathrm{AvgPool}(F))+\mathrm{MLP}(\mathrm{MaxPool}(F)))=\sigma\left(\boldsymbol{W}_1(\boldsymbol{W}_0(F^c_{\mathrm{avg}}))+\boldsymbol{W}_1(\boldsymbol{W}_0(F^c_{\mathrm{max}}))\right) \tag{2}$$

where F means the input feature; F^c_{avg} and F^c_{max} represent the output features of global average pooling and global maximum pooling along channel, respectively; σ denotes the sigmoid function; $\boldsymbol{W}_0\in\mathbf{R}^{C/r\times C}$, and $\boldsymbol{W}_1\in\mathbf{R}^{C\times C/r}$. Note that the MLP weights, $\boldsymbol{W}_0$ and $\boldsymbol{W}_1$, are shared for both inputs and the ReLU activation function is followed by $\boldsymbol{W}_0$.

The spatial attention mechanism module compresses the feature image processed by the channel attention mechanism module using the global maximum pooling layer and the global average pooling layer to obtain two one-dimensional feature vectors, and it then directly stitches them together according to the channel dimension to obtain a feature map with a channel dimension of two. Finally, through a fixed convolution layer, the activation function remains sigmoid to obtain the final feature result map.

The specific formula is as follows

$$M_s(F)=\sigma\left(f^{7\times7}([\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)])\right)=\sigma(f^{7\times7}([F^s_{\mathrm{aug}};F^s_{\mathrm{max}}])) \tag{3}$$

where $F^s_{\mathrm{aug}}\in\mathbf{R}^{1\times H\times W}$ and $F^s_{\mathrm{max}}\in\mathbf{R}^{1\times H\times W}$, represent the output features of global average pooling and global maximum pooling along spatial, respecctirely; while $f^{7\times7}$ represents a convolution operation with the filter size of 7×7.

The convolutional layer network structure of the feature extraction after adding the attention mechanism is shown in Fig.6.

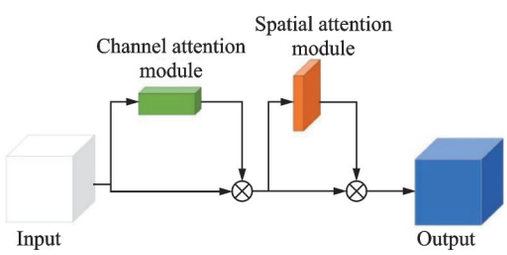


Fig.6 Convolutional layer structure that incorporates CBAM

In order to improve the perception ability of the detection model, the CBAM module was embedded in the backbone of YOLOv4 (i.e., CSPDarknet).

2.2.2 Optimization of the network architecture

(1) CSPDarknet53 network optimization

In optimizing the CSPDarknet53 network, the channel dimensions of all convolutional layers within the backbone network were reduced to one-third of their original size. This adjustment results in an exceptionally lightweight model. As a consequence, the lightweight backbone network does not require reliance on pre-trained models for classification tasks and can directly utilize small datasets, such as SSDD, to train the detection model.

(2) Spatial pyramid pooling structure optimization

Drawing inspiration from advancements in region proposals, the feature extraction module of the YOLOv4 network was enhanced with the addition of two scale-down sampling layers within the SPP module. Additionally, a dilated convolution layer was introduced to expand the receptive field, isolate small-scale target features, and diminish the false alarm rate of detection.

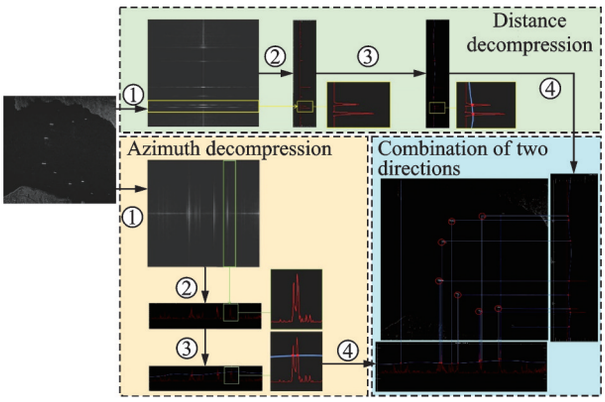
These structural optimizations to the YOLO detection algorithm heighten sensitivity to small targets, diminish the rate of misses, and ensure accurate image detection by the SAR small target ship detection algorithm.

2.2.3 Frequency domain detection method

The frequency-domain detection method used in this study is divided into three steps: two-dimensional frequency-domain generation, two-dimensional frequency-domain detection, and two-way cooperative target positioning. The overall process is illustrated in Fig.7.

Step 1 Two-dimensional frequency-domain generation

Because the original, acquired SAR image is in the space domain, it must be decompressed in the azimuth and range directions to realize the conversion from the space domain to the frequency domain. Specifically, decompression in all directions was realized through a fast Fourier transform. The decompression in azimuth direction can obtain the



① Fourier transform in the range and azimuth directions; ② Conversion of the transformed image to an intensity profile; ③ Detection of regional maximum value in the intensity profile; ④ Collection of regional maximum value in the range and azimuth directions

Fig.7 Flowchart of target detection in the range and azimuth directions

azimuth-frequency domain data, and the decompression in range direction can obtain the range frequency domain data. The azimuth-frequency domain data can be used to detect the position of the target in the upward range, whereas the frequency domain data can be used to detect the position of the target in the upward direction. The matching of the two can realize accurate positioning of the target in the azimuth and upward directions.

$$F(u,v)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\mathrm{e}^{-\mathrm{j}2\pi(ux+vy)}\mathrm{d}x\mathrm{d}y \quad (4)$$

where $F(u,v)$ is the Fourier transform of function $f(x,y)$ in the frequency domain; $f(x,y)$ a function in the spatial domain; j an imaginary unit; U and v are frequency variables on the horizontal and vertical axes, respectively; x and y the actual variables of the function $f(x,y)$.

Step 2 Two-dimensional frequency-domain detection

The azimuth-frequency domain data and range frequency domain data generated in Step 1 are converted from two-dimensional to one-dimensional maps, and the signal strength map of the center position is obtained. After the specific implementation, the results of Step 1 are shown in ② and ③ processes of Fig.7, and then the local regional maximum value is detected. The regional maximum value position corresponds to the position of the ship target in this direction.

Step 3 Two-way cooperative target positioning

Following the completion of target regional maximum value detection in Step 2, the target positions in both azimuth and range directions are obtained. These two positions are then reconstructed into the image to achieve precise target positioning. In this study, the presence of a target is determined based on extreme intensity values occurring in both azimuth and range directions. Consequently, unidirectional extreme values are disregarded. This approach allows intensity regional maximum value to better indicate the target's position at sea, minimizing interference from highly reflective objects on land and facilitating accurate target positioning in maritime environments.

3 Experimental Design

3.1 Experimental data

The SAR image small-ship detection dataset primarily relies on the SSDD dataset[34], supplemented with data from our Gaofen 3 and TerraSAR-X sources for augmentation and expansion. The finalized dataset consists of 1 680 images, each with dimensions of approximately 500 pixel×500 pixel, with some exemplar data showcased in Fig.8. All targets within the bounding box have a pixel count of less than 1 000, with the smallest target measuring 23 pixels, as depicted in Fig.9.

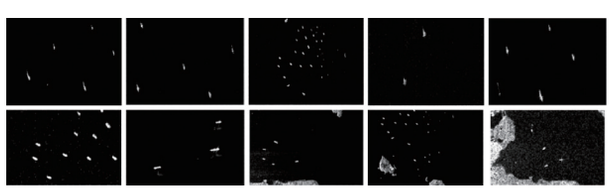


Fig.8 Partial dataset for ship detection in SAR images

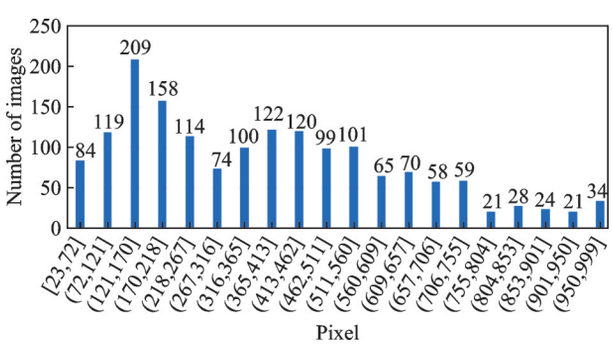


Fig.9 Distribution of pixel numbers for target surrounding boxes in the dataset

For dataset organization, 80% of the samples were randomly chosen as the training set, while the remaining 20% were evenly split between the validation and test sets. These samples were structured according to the file arrangement of the Pascal VOC 2012 Public Dataset. The distribution of specific number of ships (NoS) and number of images (NoI) is presented in Table 2.

Table 2 Correspondence between datasets NoS and NoI in small ship detection for SAR images

NoS	1	2	3	4	5	6	7	8	9	10	10+
NoI	1 056	266	130	68	65	23	22	12	6	16	16

3.2 Experimental environment

The SAR image detection algorithm was experimentally implemented using the Win11 operating system. The PyTorch environment was conFig.d, a Pychram editor was used, and the target detection algorithm was implemented, trained, and predicted using an NVIDIA GeForce RTX 3070 8G graphics card.

In the SAR image detection part of the experiment, an end-to-end training method was adopted. The hyperparameters were set as follows: initial learning rate, 0.001 3; the maximum number of iterations, 50 500; batch, 64; momentum, 0.949; and regular weight decay term, 0.000 5.

3.3 Experimental results and analysis

To verify the effectiveness of the dual-domain SAR small-ship detection method proposed in this paper for real SAR images, ablation experiments were conducted under the conditions of the experimental design arrangement in Section 2 to support the quantitative analysis, and three typical areas (far sea area, near-shore area, and port area) were selected for qualitative analysis and evaluation.

3.3.1 Experimental design

(1) Ablation experiment design

To verify the effectiveness of the method used in this study, we designed a total of four layers of ablation tests based on the basic YOLOv4 method. The first layer represented the basic YOLOv4 algorithm, the second layer represented the improved YOLOv4 algorithm, the third layer represented the improved YOLOv4 algorithm embedded in the attention mechanism module, and the fourth layer was based on the third layer. Subsequently, the frequency spectrum detection part was added. The ablation experiment demonstrated the effectiveness of each module using the improved method developed in this study. The specific design components of the ablation test structure are listed in Table 3.

Table 3 Ablation experimental models and their architecture

Index	Model	YOLOv4	Optimization of network architecture	CBAM	Collaboration of RD spectrum maps	Precision/%	Miss rate/%	F_1 score
1	YOLOv4	√				77.46	18.82	0.792 7
2	Improved YOLOv4	√	√			89.78	23.19	0.827 9
3	Improved YOLOv4+CBAM	√	√	√		91.54	17.98	0.865 2
4	Our method	√	√	√	√	92.25	11.09	0.905 5

(2) Comparative method experimental design

After the ablation experiment was completed, different methods were compared, including the single shot MultiBox detector(SSD)[35], Faster R-CNN, YOLOv4, YOLOv5[36], and a vision-centric foundation model to explore the limits of visual representation at scale using only publicly accessible data (EVA)[37] algorithms. A comparison between advanced algorithms in related research at home and abroad and the algorithm in this study can further demonstrate the effectiveness and advancement of the algorithm proposed in this study.

3.3.2 Analysis of experimental results

(1) Evaluation of the ablation experiment results

The results of the quantitative analyses presented in Table 4 clearly show that after optimizing the network structure of YOLOv4, the overall precision of ship target detection increases from 77.46% to 89.78%. However, a 4.37% increase in the miss rate is observed. This indicates that the addition of

Table 4 Analysis of the ablation experiment results

Index	Model	Precision/%	Miss rate/%	F_1 score
1	YOLOv4	77.46	18.82	0.792 7
2	Improved YOLOv4	89.78	23.19	0.827 9
3	Improved YOLOv4+CBAM	91.54	17.98	0.865 2
4	Our method	92.25	11.09	0.905 5

scale-downsampling layers facilitates the extraction of small targets but also leads to some land targets being detected as ships. The attention mechanism described in the second section improves the feature extraction capability for ships and reduces interference from land and noise, resulting in a slight improvement in target detection accuracy. Furthermore, with the addition of azimuth and range frequency spectrogram detection, the overall precision increases by 0.71%. Additionally, because the interference from land in the frequency spectrogram is relatively low, there is a partial reduction in the miss rates. Compared with the original unimproved algorithm, our method achieves a 14.79% increase in precision and a 7.73% decrease in miss rate. From a quantitative analysis perspective, our method demonstrates good detection performance and a significant improvement in the miss rate.

(2) Qualitative analysis and evaluation of ablation results

For this qualitative evaluation, three main detection scenarios are selected: open-sea, coastal, and port area ships, as shown in Fig.10.

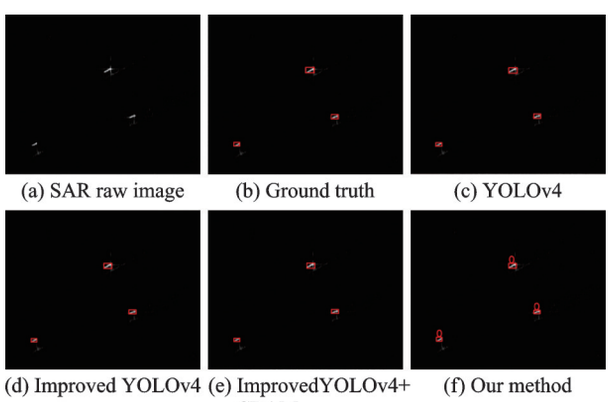


Fig.10 Detection results of various models in the far-sea area ship detection experiment

It is evident from the detection results of ships in far-sea areas that the method proposed in this paper and the original methods can achieve relatively accurate detection. There are no obvious differences in ship detection precision results between methods shown in Figs.10(c—e).

SAR small-ship detection results in the coastal ship scenario are shown in Fig.11.

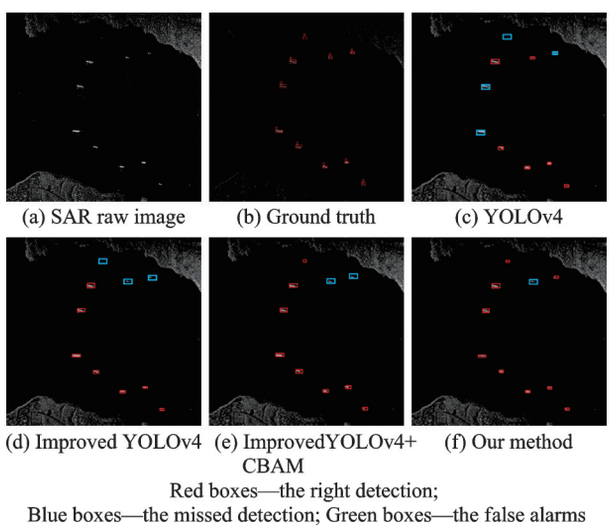


Fig.11 Detection results of various models in the coastal ship detection experiment

Fig.11 shows that the original YOLOv4 algorithm cannot achieve more accurate ship detection due to interference from shore environment noise. Four ships in the north are not detected. After adding the attention mechanism, the algorithm still misses two ships. Finally, after adding the two-way frequency spectrum detection, only one ship in the north is missed. Compared with the YOLOv4 algorithm, the detection precision of ships in this area is significantly improved. For ship detection in coastal areas, the proposed algorithm has a higher precision rate and a lower miss rate.

SAR small ship detection results in port area scenario are shown in Fig.12. Fig.12 clearly reveals that environmental noise is more severe than that in the near-shore area. Buildings in the port area and rocky formations along the shore generate strong signal reflections, causing significant interference in target detection. When only YOLOv4 is used for object detection, there are significant instances of missed targets. However, target detection is improved after network improvement and attention

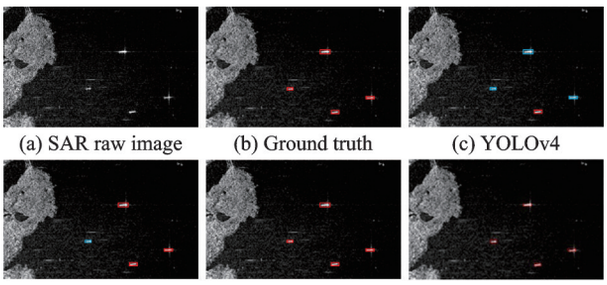


Fig.12 Detection results of various models in the port area ship detection experiment

mechanism adjustment. In this study, the proposed method further utilizes strong scattering for maritime target detection, thereby refining the position of the target. The experimental results further confirm the effectiveness of the proposed method.

A qualitative analysis of the above three scenarios indicates that the improvement in the detection precision of the proposed method is not obvious compared with that of the original deep learning method when detecting ships in the open sea area. However, the improvement is more obvious when detecting ships in the near-shore and port areas based on the reduced false alarm rate caused by land and port buildings. Thus, the final precision is significantly improved.

(3) Evaluation of comparative experimental results

The quantitative results in Table 5 indicate that our method has higher precision than all other five algorithms, 1.41% higher than the second highest algorithm (EVA algorithm), and the miss rate also reaches the second lowest value. The F_1 score of our method is only little lower than the EVA algorithm, and overall, our method achieves better results compared to existing algorithms.

Table 5 Analysis of the ablation experiment results

Index	Model	Precision/%	Miss rate/%	F_1 score
1	SSD	81.59	48.95	0.628 0
2	Faster R-CNN	73.79	42.78	0.644 6
3	YOLOv4	77.46	18.82	0.792 8
4	Improved YOLOv5	86.26	14.33	0.859 6
5	EVA	90.84	6.52	0.921 4
6	Our method	92.25	11.09	0.905 5

(4) Qualitative analysis and evaluation of comparative experimental results

For the qualitative evaluation, three typical scene images are selected for comparison: far-sea areas, dense target scene, and near-shore land interference, as shown in Fig.13. The algorithm achieves good results for these scenes. In particular, when dealing with dense scenes of small targets, the image slice contains 24 small targets, 17 of which are accurately detected by the EVA algorithm and, 22 of which are accurately detected by the V5 algorithm, and there are three false alarms. However, the proposed algorithm realizes the accurate detection of 23 small targets and only produces two target false alarms. At the same time, in coastal scenarios, the effect of the EVA algorithm is far inferior to our method. Only Faster R-CNN and the method proposed in this paper have achieved good detection for several coastal ships in the southern part of the scene. The comparative experiments with the meth-

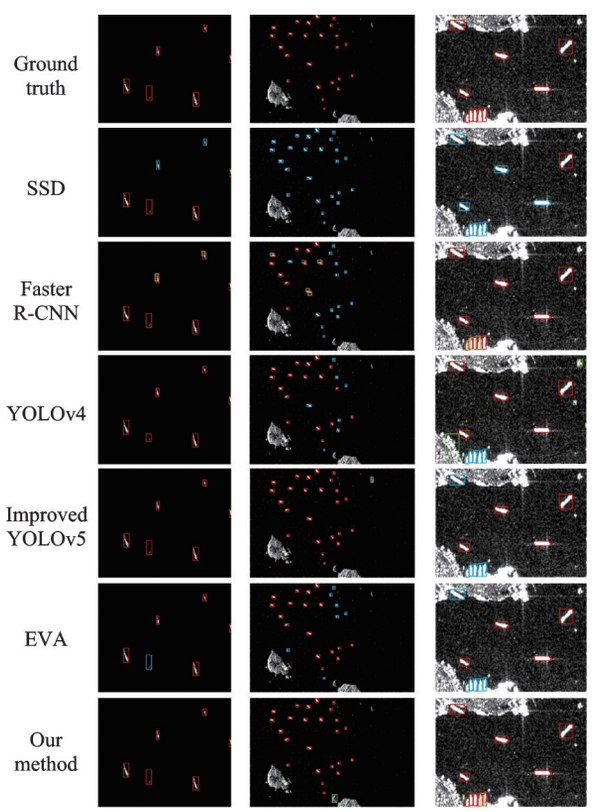


Fig.13 Results of small-ship targets for various models in comparative experiments

ods described in this section further demonstrate the high detection accuracy and low miss rate of the proposed algorithm in complex and multi-background interference scenarios.

4 Conclusions

This study introduces an innovative approach by combining the SAR small ship detection algorithm commonly utilized in the image domain with an intensity detection method in the frequency domain. To tackle the challenges inherent in SAR small-ship detection, particularly susceptibility to interference from highly reflective objects on the shore, a dual-domain joint SAR ship target detection algorithm is proposed. This algorithm integrates the lightweight attention mechanism module CBAM into the feature extraction stage of the YOLOv4 algorithm and optimizes its structure to enhance ship target features while suppressing interference images, such as environmental features. The proposed SAR small-ship detection dataset facilitates achieving higher detection precision and lower miss rates.

The combined quantitative and qualitative evaluation approach yields superior results in ablation experiments across far-sea, near-shore, and port areas, as well as in method comparison experiments covering far-sea areas, small target dense scenes, and near-shore land interference scenarios.

However, experiments reveales a high false alarm rate in the ship detection process within port areas using the proposed algorithm. Additionally, the combination of the two algorithm domains employed in this study is found to be insufficient. Therefore, further research efforts should focus on developing methods to reduce the false alarm rate of SAR small ship detection in complex environments and enhance the internal integration of the algorithm for improved performance in future studies.

References

[1] ZHOU X, CHANG N B, LI S. Applications of SAR interferometry in earth and environmental science research[J]. Sensors, 2009, 9(3): 1876-1912.

[2] AI J, CAO Z, MAO Y. An improved bilateral CFAR ship detection algorithm for SAR image in complex environment[J]. Journal of Radars, 2021, 10: 499-515.

[3] LENG X, JI K, YANG K, et al. A bilateral CFAR algorithm for ship detection in SAR images[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12: 1536-1540.

[4] LIU Y, ZHANG M, XU P, et al. SAR ship detection using sea-land segmentation-based convolutional neural network[C]//Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP). Shanghai, China: IEEE, 2017: 1-4.

[5] HWANG S I, 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: 391-395.

[6] WANG C, WANG Y, LIAO M. Removal of azimuth ambiguities and detection of a ship: Using polarimetric airborne C-cand SAR images[J]. International Journal of Remote Sensing, 2012, 33: 3197-3210.

[7] XU G A, HANG S, LONG G, et al. Feature aligned ship detection based on improved RPDet in SAR images[J]. Displays, 2022, 74: 102191.

[8] AI J, TIAN R, LUO Q, et al. Multi-scale rotation-invariant Haar-like feature integrated CNN-based ship detection algorithm of multiple-target environment in SAR imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57: 10070-10087.

[9] YU L, WU H, ZHONG Z, et al. TWC-Net: A SAR ship detection using two-way convolution and multiscale feature mapping[J]. Remote Sensing, 2021, 13: 2558.

[10] AI J, MAO Y, LUO Q, et al. SAR target classification using the multikernel-size feature fusion-based convolutional neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-13.

[11] WEI S, SU H, MING J, et al. Precise and robust ship detection for high-resolution SAR imagery based on HR-SDNet[J]. Remote Sensing, 2020, 12: 167.

[12] TANG G, ZHUGE Y, CLARAMUNT C, et al. N-YOLO: A SAR ship detection using noise-classifying and complete-target extraction[J]. Remote Sensing, 2021, 13: 871.

[13] KE X, ZHANG X, ZHANG T, et al. SAR ship detection based on an improved faster R-CNN using deformable convolution[C]//Proceedings of 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. [S.l.]: IEEE, 2021: 3565-3568.

[14] ZHANG T, ZHANG X, KE X, et al. HOG-ShipCLSNet: A novel deep learning network

with hog feature fusion for SAR ship classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-22.

[15] LIN Z, JI K, LENG X, et al. Squeeze and excitation rank faster R-CNN for ship detection in SAR images[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 16(5): 751-755.

[16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 7132-7141.

[17] ZHAO Y, ZHAO L, XIONG B, et al. 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.

[18] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: Springer, 2018: 3-19.

[19] 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, 2020, 19: 1-5.

[20] ZHANG T, ZHANG X. ShipDeNet-20: An only 20 convolution layers and < 1-MB lightweight SAR ship detector[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18(7): 1234-1238.

[21] WANG J, LIN Y, GUO J, et al. SSS-YOLO: Towards more accurate detection for small ships in SAR image[J]. Remote Sensing Letters, 2021, 12(2): 93-102.

[22] DING J, WEN L, ZHONG C, et al. Video SAR moving target indication using deep neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(10): 7194-7204.

[23] WEN L, DING J, CHENG Y, et al. Dually supervised track-before-detect processing of multichannel video SAR data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-13.

[24] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[25] 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): 765.

[26] LI J, QU C, SHAO J. Ship detection in SAR images based on an improved faster R-CNN[C]//Proceedings of 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA). [S.l.]: IEEE, 2017: 1-6.

[27] SUN X, WANG Z, SUN Y, et al. AIR-SARShip-1.0: High-resolution SAR ship detection dataset[J]. Journal of Radars, 2019, 8(6): 852-863.

[28] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[EB/OL]. (2020-04-23). https://arxiv.org/abs/2004.10934.

[29] CHEN D, MIAO D. Control distance IoU and control distance IoU loss function for better bounding box regression[J]. Pattern Recognition, 2023, 137: 109256.

[30] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2020: 12993-13000.

[31] ZHANG T, ZHANG X. A full-level context squeeze-and-excitation ROI extractor for SAR ship instance segmentation[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.

[32] WANG F, TAX D M J. Survey on the attention based RNN model and its applications in computer vision[EB/OL]. (2016-01-25). https://arxiv.org/abs/1601.06823.

[33] ZHANG M, AN J, YANG L D, et al. Convolutional neural network with attention mechanism for SAR automatic target recognition[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 19: 1-5.

[34] LI J, XU C, SU H, et al. Deep learning for SAR ship detection: Past, present and future[J]. Remote Sensing, 2022, 14(11): 2712.

[35] MIAO T, ZENG H, WANG H, et al. Inshore ship detection in SAR images via an improved SSD model with wavelet decomposition[C]//Proceedings of 2021 7th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR). [S.l.]: IEEE, 2021: 1-5.

[36] FU Q, CHEN J, YANG W, et al. Nearshore ship detection on SAR image based on YOLOv5[C]//Proceedings of 2021 2nd China International SAR Symposium (CISS). [S.l.]: IEEE, 2021: 1-4.

[37] FANG Y, WANG W, XIE B, et al. Eva: Exploring the limits of masked visual representation learning at scale[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2023: 19358-19369.

Acknowledgements This work was supported by the Foundation Strengthening Fund Project (No.2021-JCJQ-JJ-

0251); in part by the National Natural Science Foundation of China (Nos.42301384 and 42271448).

Authors Mr. **JIA Peng** received his B.S. degree from the PLA University of Information Engineering in 2002 and his M.S. degree from the PLA University of Information Engineering in 2005. Currently, he is pursuing a doctoral degree at Nanjing University of Aeronautics and Astronautics. He has been engaged in the design of spacecraft measurement and control communication systems, as well as satellite remote sensing image processing for a long time.

Dr. **LI Jun** received his B.S. degree in remote sensing science and technology, his M.S. degree in geomatics Engineering and his Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2015, 2018, and 2021, respectively. He is currently an associate research fellow at College of Astronautics, Nanjing University of Aeronautics and Astronautics. His research interests include remote sensing image processing and deep learning.

Author contributions Mr. **JIA Peng** designed the study, and complied the models. Mr. **DONG Tiancheng** conducted the analysis. Mr. **WANG Taoyang** provided the data. Mr. **ZHANG Guo** contributed to the discussion and background of the study. Prof. **SHENG Qinghong** interpreted the results. Dr. **LI Jun** wrote the manuscript. All authors commented on the manuscript draft and approved the submission.

Competing interests The authors declare no competing interests.

(Production Editor: ZHANG Huangqun)

面向星载SAR图像的双域联合密集多小舰目标检测算法

贾 鹏[1], 董天成[2], 汪韬阳[3], 张 过[2], 盛庆红[1], 李 俊[1]

(1.南京航空航天大学航天学院,南京 211106,中国; 2.武汉大学测绘遥感信息工程国家重点实验室, 武汉 430079,中国; 3.武汉大学遥感与信息工程学院,武汉 430079,中国)

摘要:通过星载合成孔径雷达(Synthetic aperture radar, SAR)进行舰船探测已成为研究热点,但由于海上相干斑点噪声明显,近岸物体反射干扰强等问题,现有的基于雷达信号域和SAR图像特征的小型舰船探测方法无法获得高精度的结果。为解决上述问题,提出了一种针对空间SAR图像的双域联合密集多重小型船舶目标检测方法,可同时在图像域和频域检测目标。该方法利用注意力机制模块和算法结构调整来提高小船目标特征挖掘能力。在基于频率的图像生成中,检测方位角和测距方向的极端信号强度值,二者结果互为补充,实现双域联合小型舰船目标检测。定性和定量的综合评价结果表明,所提出的方法最终精确率可达92.25%,在开阔海域、沿海和港区船舶的SAR船舶探测方面取得了准确的结果。自建SAR小型船舶数据集的测试结果证明了该方法的有效性和通用性。

关键词:合成孔径雷达;小型船舶探测;深度学习;注意力模块;YOLO;双域联合