BFEA: A SAR Ship Detection Model Based on Attention Mechanism and Multiscale Feature Fusion

Liming Zhou, *Member, IEEE*, Ziye Wan, Shuai Zhao, Hongyu Han, and Yang Liu, *Member, IEEE*

***Abstract*—With the advancements in deep learning and synthetic aperture radar (SAR) technology, an increasing number of individuals are utilizing deep-learning techniques to detect ships in SAR images. However, the efficiency of SAR ship detection is affected by complex background interference and various ship sizes. Addressing these challenges, this article proposes a balanced feature enhanced attention model. First, we introduce a novel attention feature fusion network (WEF-Net) tailored for SAR multiscale ship detection. WEF-Net effectively balances the information across different backbone layers and harmonizes semantic information from various levels of the feature pyramid through aggregation and averaging. Next, we embed the receiving field extension module in WEF-Net to learn the context information and generate the global characteristics of the receiving field balance. In addition, it can extract features from multiple scales to enhance the detection capability of the model for ships of different scales. At the same time, acknowledging the impact of surrounding complex background interference on the detector, we redesigned the ELAN module by combining convolution and attention. This enhancement enables the model to better attend to target position information during feature fusion, suppress the surrounding complex background interference, and highlight the ship's feature information. Finally, owing to the prevalence of small targets in SAR images, we employ an optimized loss function to bolster the model's performance in detecting small targets. This approach accelerates training convergence, reduces instances of missed detection on small targets, and enhances overall detection performance across multiple scales. Experimental results demonstrate that our model achieves detection accuracies of 98%, 93.1%, and 76.9% on the SAR ship detection dataset, high-resolution SAR image dataset, and large-scale SAR ship detection dataset, respectively, effectively discerning ship targets amid complex backgrounds in SAR images.**

***Index Terms*—Deep learning, feature fusion network, multiscale ship detection, synthetic aperture radar (SAR).**

Manuscript received 3 April 2024; revised 15 May 2024; accepted 28 May 2024. Date of publication 3 June 2024; date of current version 17 June 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62176087, in part by the Natural Science Foundation of Henan under Grant 242300421218, in part by the Scientific and Technological Innovation Team of Universities in Henan Province under Grant 24IRTSTHN021, and in part by the Key Research and Promotion Projects of Henan Province under Grant 232102210009 and Grant 242300421218. *(Corresponding author: Hongyu Han.)*

The authors are with the Key Laboratory of Big Data Analysis and Processing, Henan University, Kaifeng 475000, China and also with the School of Computer and Information Engineering, Henan University, Kaifeng 475000, China (e-mail: lmzhou@henu.edu.cn; ziye@henu.edu.cn; hedazs@henu.edu.cn; hanhongyu@henu.edu.cn; sea@vip.henu.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/JSTARS.2024.3408339, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2024.3408339

I. Introduction

SYNTHETIC aperture radar (SAR) is an active microwave imaging radar, which can achieve all-weather, all-day observation of the ground, and can obtain high-resolution images under harsh conditions, such as night and cloud cover. It is widely used in various application fields, such as target classification [1], [2], target detection and recognition [3], [4], [5], target segmentation [6], [7], etc. Among these applications, automatic ship detection in SAR images plays an important role in both military and civilian fields, such as defense and security, fishing vessel monitoring, and maritime traffic supervision and rescue.

Due to the influence of SAR imaging characteristics, ship targets in SAR images have very low proportion and small size. Moreover, because of the scattering effect and the resolution limitation of wavelength and antenna size, it is difficult to obtain the details and boundary information of the small target, which affects the detection accuracy. At present, the deep learning related research work on SAR image small target detection is less and the effect is not good. In addition, SAR images obtained from satellites and airborne platforms generally have lower resolution and are more susceptible to background clutter and noise compared to optical images. This is because of the following conditions.

1) SAR image background is complex. Due to the influence of SAR clutter, imaging principle, structure of ground objects, and other factors, the existing detection methods are difficult to distinguish the target from the background.
2) The proportion of ships in SAR images, especially small ships, varies greatly. Due to the diversity of ship scales in SAR images, it is difficult for existing methods to extract multiscale ship features effectively.

Therefore, how to detect multiscale ship targets in complex background clutter is still a major challenge.

In recent years, the advent of deep learning, notably convolutional neural networks (CNNs), has developed rapidly, achieving remarkable results not only in optical remote sensing image processing but also in SAR image processing. Unlike traditional methods reliant on manual feature extraction, CNNs automate this process, reducing human intervention and enabling end-to-end processing.

Deep-learning-based object detection methods fall into two categories: two-stage and single-stage models. While the former generates candidate regions for subsequent classification and localization, exemplified by faster R-CNN [8], mask R-CNN [9], and libra R-CNN [10], the latter integrates region generation

with target classification, albeit at the expense of computational complexity and slower detection speeds. The single-stage algorithm treats both the target type and location as regression problems at one time, simplifying the calculation process, such as SSD [11], YOLO [12] series, FCOS [13], RetinaNet [14], and so on. In 2018, the YOLOv3 [15] algorithm was proposed, which introduced the residual structure and deepen the depth of the network, and improved the detection capability of different scales. Later, the slicing operation was introduced into the YOLOv5 [16] network model. In 2022, the YOLOv7 [17] algorithm was proposed, which introduced strategies such as multibranch stacked modules, innovative transition modules, and heavy parametric structures to further optimize the effect of target detection. YOLOv7 surpasses the previous YOLO family, with 120% FPS improvement over YOLOv5 and 180% FPS improvement over YOLOX [18].

In recent years, with the development of CNNs, deep-learning techniques have been applied in the field of SAR ship target detection, and researchers have successfully utilized them to address specific challenges. To solve the problem of large-scale variation of ship targets in SAR images, Bai et al. [19] proposed a new anchor-based SAR ship detector, which balances semantic information at different levels of the feature pyramid through aggregation and averaging, and learns global context information. Zhang et al. [20] proposed a multiscale global scattering feature association network for SAR ship target recognition, which improves the detection performance of multiscale SAR image ship detection. To solve the problem of detecting targets with complex background interference, Yu et al. [21] introduced a ship detector (MHASD), which integrates local and global attention by using mixed attention residual model (HARM), thereby enhancing ship target features at channel and spatial levels and improving detection accuracy. Sun et al. [22] proposed an anchor-free method for ship target detection in high-resolution SAR images, which improves ship target location performance in complex scenes. Sun et al. [23] proposed a complex background SAR ship recognition method based on two-branch transformer fusion network, and designed a two-branch feature extraction and fusion architecture to effectively extract local fine-grained features of ships. To solve the problem of missing small targets in SAR images, Hu and Miao [24] proposed a new small SAR ship detection network, which adopted a transformer-based dynamic sparse attention module to improve the focusing and extraction of small ship features. Kang et al. [25] proposed a deep layered network for SAR ship detection, a multilayer fusion CNN based on context area, which incorporates deep semantics and shallow high-resolution features to improve detection performance for small ships. Although the above detection methods have achieved good results, it is still a major challenge to detect multiscale ships in complex background clutter due to the multiscale and weak significance of target features and the complexity of background noise.

Therefore, in order to solve the problems of large-scale change of ship targets, many small targets and background clutter in SAR images. We construct a multiscale SAR ship detection model based on attention mechanism, called balanced feature enhanced attention model (BFEA). Because of the special imaging mechanism of SAR, it is still a big challenge to detect small ships from SAR images. BFEA can effectively solve the problem of missing small ships in SAR images with weak semantic and false alarm interference. We design a novel attention feature fusion network (WEF-Net), which can solve the problem that the ship target is not obvious and the background clutter is strong in SAR images. WEF-Net balances the semantic information of different levels of features through aggregation, and imparts an attention mechanism to suppress the interference of complex background, focusing more on the ship target. In addition, in order to solve the problem of the diversity of SAR ship scale changes and the large number of small size targets, we also embedded a multireceptive field extension module (MRFFM) in WEF-Net to learn the context features, increase the receptive field, and improve the acquisition ability of small targets. Then, in order to adaptively learn the importance of feature maps of different scales in the fusion process and improve the ability to represent multiscale features, we use the weighted feature fusion module (WFF) to introduce learnable weights to learn the importance of different input features, so as to better integrate multiscale features and retain feature information with richer context. Therefore, the model can improve the perception ability of different scale targets. In addition, the model uses the combination of convolution and attention to optimize the ELAN module. After optimization, the ELAN module can better fully integrate the feature information of different scales, enhance and improve the fusion of multiscale feature maps with complex background, so that the model can learn more global feature information, so that it can detect the ship target under complex background. Finally, for the large number of small targets in SAR images, we apply a loss function combining the complete cross-linked (CIoU) measurement scheme and Wasserstein distance. This optimization method enables the detector to perform well in capturing small ship targets, and our model achieves excellent performance on three challenging multiscale SAR ship datasets. The main contributions of this article are as follows.

1) This article constructs a new attention-mechanism-based multiscale SAR ship detection model BFEA, which has a novel WEF-Net, MRFFM, and optimized loss function. In the complex background detection of multiscale SAR ship detection, more excellent detection performance is obtained.
2) Aiming at the characteristics of inconspicuous ship targets and strong background clutter in SAR images, we propose a WEF-Net. It includes two parts: balanced feature module (BFM) and feature enhancement fusion module (FEF). WEF-Net balances the semantic information of different levels of features through aggregation, and has an embedded attention mechanism to suppress the interference of complex background and pay more attention to the ship target. The WFF and the improved ELAN module are used to better and fully integrate the feature information of different scales, enhance and improve the fusion of multiscale feature maps with complex backgrounds, and learn more global feature information, so that it can detect ship targets under complex background.

3) To solve the problem of the diversity of SAR ship scale variation and the large number of small-size targets, we embedded the MRFFM in WEF-Net to learn the context information, to build more expressive multiscale features, and to improve the detection performance of small targets.
4) To solve the problem of missing small- and medium-sized targets in SAR images, this article uses the advantages of CIoU and NWD loss function to optimize the loss function as a balance between bounding box regression accuracy and training convergence. The optimized loss function is insensitive to the movement of small targets, thus improving the detection accuracy of small ships.
5) To verify the performance of the proposed model, we performed tests on three challenging real SAR datasets: SAR ship detection dataset (SSDD), high-resolution SAR image dataset (HRSID), and large-scale SAR ship detection dataset (LS-SSDD-v1.0). Compared with the other object detection model, the average accuracy (mAP) of our model on the three datasets reaches 98%, 93.1%, and 76.9%, respectively, showing excellent performance in the task of detecting multiscale SAR ship targets in complex background.

The rest of this article is organized as follows. Section II reviews related methods. Section III describes the proposed method in detail. Section IV shows experiment and results. Finally, Section V concludes this article.

II. Related Work

A. SAR Target Detection in Complex Background

In SAR images, complex background target detection is a difficult problem because the imaging mechanisms of SAR images and optical images are very different, it is easy to be affected by the texture information is not significant, the edge is difficult to detect, the coherent point noise is serious, lack of geometric structure, and other factors.

Many scholars have studied deep-learning-based approaches to solve these problems. Yang et al. [26] introduced the coordinate attention module to extract more discerning semantic features, facilitating precise localization and discrimination of ship targets. Wang et al. [27] proposed an improved YOLOv5 method to address the low accuracy of multiship target detection tasks in complex scenes, and introduced asymmetric cone nonlocal block and sim attention mechanism to reduce nearshore background interference. Chen et al. [28] proposed an anchor-free method SAD-Det based on transformer and adaptive features, which can detect rotating invariant ship targets with high average accuracy in SAR images, and realize high-precision ship detection in any direction in complex scenes. Additionally, Huang et al. [29] devised the horizontal synchronous attention mechanism (LSAM), which dynamically adjusts weights for different regions to bolster information capture regarding target entities. Through these improved methods, the performance of SAR image complex background target detection has been further improved, which brings new possibilities for more accurate target detection and recognition.

B. Small Target Detection

In SAR images, small target detection is a challenging problem. According to the characteristics of SAR imaging, SAR image targets are sparse and small in size. In object detection, there are cases of missing detection because the SAR image is blurred or the target object itself is small in size, which leads to the target appearing in the SAR image occupying fewer pixels.

This section delves into various fundamental network-based optimization strategies. Hu and Miao [24] proposed an SAR detection model based on transformer architecture to improve the detection accuracy. Fu et al. [30] presented the feature balancing and refinement network to harmonize semantic features across diverse levels, facilitating comprehensive information acquisition for small ship detection. Additionally, Lu et al. [31] introduced the information enhancement feature pyramid network (FPN). This architecture enriches the spatial context information flow and refines the representation of small and multiscale ships. These network optimization methodologies hold promise for achieving more precise small target detection, thereby enabling real-time detection capabilities.

C. Multiscale Target Detection

Given the SAR imaging characteristics, ship targets in SAR images exhibit multiscale attributes. The utilization of an FPN [32] to construct multilevel features stands as a primary technique in addressing multiscale target detection challenges. However, a noticeable semantic disparity exists among FPNs at distinct levels, potentially diminishing the representational capacity of multiscale features.

Therefore, researchers have proposed new solutions to the above problems. Yang et al. [26] devised a receptive field enhancement module, thereby enhancing ship detection performance across diverse scales. Tang et al. [33] proposed an attention mechanism for multiscale receptive fields convolution block (AMMRF). AMMRF not only makes full use of the location information of the feature map to accurately capture the regions in the feature map that are useful for detection results, but also effectively captures the relationship between the feature map channels, so as to better learn the relationship between the ship and the background. Tan et al. [34] proposed bidirectional feature pyramid (BiFPN), augmenting horizontal skip connections. thereby facilitating superior feature fusion. These methodologies empower the model to glean richer contextual information, thereby enhancing its capacity to detect multiscale features.

III. Methodology

A. Network Structure

To solve the problem of large-scale change and background clutter of SAR image, we propose a BFEA, as shown in Fig. 1. First, we extract features based on the CSPDarknet53 architecture. Then, we propose a new WEF-Net to balance features at different levels, which consists of a BFM and an FEF. BFM can balance the information of each backbone layer and generate the global features of the sensitivity field balance, while FEF

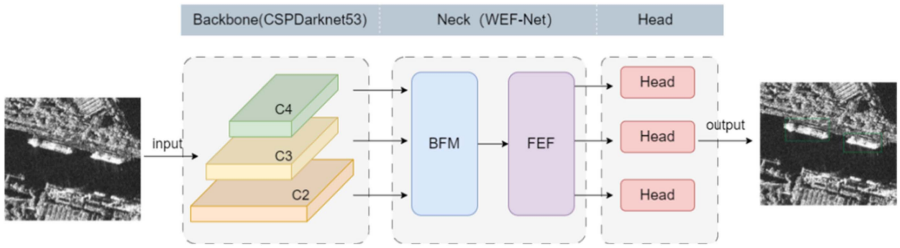

Fig. 1. Overall structure of our proposed model BFEA. The neck network WEF-Net is divided into balanced feature module BFM and feature enhancement fusion module FEF.

can balance the semantic information of different levels of the feature pyramid by averaging the global features. Finally, the output from WEF-Net is sent to the header.

B. Multifield Extension Module (MRFFM)

In the proposed WEF-Net neck network, the MRFFM is embedded to balance the features of different receptive fields and extract more context features. First, the input feature graph F is subjected to the expansion convolution with an expansion rate of 2, then the CA mechanism is used to extract features to obtain the $R1$ feature graph, and the channel cascade operation is performed with the input original feature graph F. The fused feature graph is first subjected to 1×1 convolution to change the number of channels, and then the expanded convolution with an expansion rate of 4 and the CA mechanism to extract features to obtain the $R2$ feature graph. Moreover, the original feature map and $R1$ feature map are continued to be superimposed, and the fused feature map is first subjected to 1×1 convolution to change the number of channels, and then into the expansion convolution with an expansion rate of 8 and the CA mechanism to extract features to obtain the $R3$ feature map. Finally, the channels of $R1$, $R2$, $R3$, and the original input feature graph F are stacked and transformed into the feature graph P with the same number of channels as the input feature graph by 1×1 convolution.

Since small targets lack semantic information, it is urgent to extract global context information with large receiving fields as an important compensation. Therefore, in the BFM part of WEF-Net, we embed MRFFM into the C4 layer with the most abundant semantic information to extract more important feature information conducive to the detection of small ships. In addition, we embed MRFFM into the FEF part of WEF-Net to extract global feature information for balanced features. The MRFFM structure is shown in Fig. 2, and its formula is shown as follows:

$$R1 = \mathrm{CA}\left[D\mathrm{Conv}^{[1]}_{3\times3}(F)\right] \tag{1}$$

$$R2 = \mathrm{CA}\left(D\mathrm{Conv}^{[2]}_{3\times3}(\mathrm{Conv}_{1\times1}\left[R1, F\right])\right) \tag{2}$$

$$R3 = \mathrm{CA}\left(D\mathrm{Conv}^{[3]}_{3\times3}(\mathrm{Conv}_{1\times1}\left[R1, R2, F\right])\right) \tag{3}$$

$$P = \mathrm{Conv}_{1\times1}\left[R1, R2, R3, F\right] \tag{4}$$

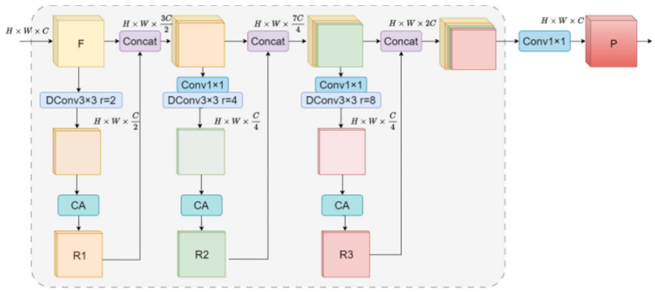

Fig. 2. Structure of MRFFM, F represents the input original feature map, $Ri(i = 1, 2, 3)$ represents the intermediate feature map of the expansion convolution and coordination attention mechanism through different expansion rates, and P is the output feature map. Conv 1×1 is 1×1 standard convolution, DConv 3×3 $r = 2, 4, 8$ denote expansive convolutions of 3×3 with expansion rates of 2, 4, and 8, respectively. CA represents the coordination attention mechanisms. Concat represents channel cascade operation.

where F denotes the input feature map, $Ri(i = 1, 2, 3)$ denotes the intermediate variable, while $D\mathrm{Conv}^{[1]}_{3\times3}$ denote $3 \times 3 \times C \times \frac{C}{2}$ expansion convolution with expansion rate of 2. Similarly, $D\mathrm{Conv}^{[2]}_{3\times3}$ and $D\mathrm{Conv}^{[3]}_{3\times3}$ denote $3 \times 3 \times \frac{C}{4} \times \frac{C}{4}$ expansive convolutions with expansion rates of 4 and 8, respectively. $\mathrm{Conv}_{1\times1}$ is 1×1 standard convolution, $[\cdot]$ represents the operation of channel concatenation, and P denotes the output feature map. CA represents the coordination attention mechanism.

Compared to other popular modules based on expansive convolution, MRFFM has four innovative features. First, MRFFM uses the combination of dilatation convolution and coordinated attention mechanism with different dilatation rates to extract global context features, enlarge receptive field, and suppress complex background interference. Second, unlike certain modules based on dilated convolution, such as the spatial pyramid pooling structure [35], which overlooks the meshing of dilated convolution, thereby risking loss of local information, MRFFM tackles this issue by overlaying and reutilizing original receptive fields and diverse extended receptive fields to alleviate gridding challenges. Third, the channel configuration in MRFFM is meticulously devised. While numerous model architectures based on expansive convolution overlook the primacy of the original feature map in models with distinct receptive fields, resulting in the loss of original features, our MRFFM module preserves the distinct advantages of the original feature information. Last, many other modules use only one of the

Algorithm 1: Feature Fusion Method of MRFFM.

Input: $x \in R^{W\times H\times C}$refers to the input feature map
Output: $y \in R^{W\times H\times C}$refers to the output feature map
Step: $Ri(i = 1, 2, 3)$is the intermediate feature map.
$DConv^{[1]}_{3\times3}$ represents $3 \times 3 \times C \times \frac{C}{2}$ expansive convolution with expansion rate of 2. $DConv^{[2]}_{3\times3}$ and $DConv^{[3]}_{3\times3}$ represents $3 \times 3 \times \frac{C}{4} \times \frac{C}{4}$ expansive convolution with expansion rate of 4 and 8, respectively. $Conv_{1\times1}$ is a 1×1 standard convolution, $[\cdot]$ representing the concatenation operation. CA represents the coordination attention mechanisms.

```
for i = 1 to 3 do
  if i = 1 then
    Ri = CA(DConv^[1]_3×3(x))
  if i = 2 then
    Ri = CA(DConv^[2]_3×3(Conv_1×1[Ri − 1, x]))
  else
    Ri =
    CA(DConv^[3]_3×3(Conv_1×1[Ri − 1, Ri − 2, x]))
    y = Conv_1×1[Ri, Ri − 1, Ri − 2, x]
    return y
  end if
end for
```

parallel or cascading structures. In parallel structures, achieving equivalent expansion sizes necessitates higher expansion rates, significantly reducing effective convolution parameters compared to the recommended MRFFM. Conversely, in cascade structures, the original feature map's information gets diluted as the convolution layers deepen. MRFFM amalgamates the strengths of parallel and cascading architectures to mitigate these challenges.

In addition, the implementation process is summarized in Algorithm 1.

C. Efficient Layer Aggregation Module Based on Attention Mechanism (ELAN-A)

In order to solve the problem of complex background clutter affecting SAR image ship detection, we improve the ELAN-W [17] fusion module by mixing convolution and attention module, and named it ELAN-A. Global context information plays a key role in ELAN modules. To achieve this goal, we introduce CA [36] mechanism to help models better focus on global information. In contrast to other prevalent attention mechanisms, CA's computational scale is small, endowing the network with the capability to attend to a larger range of targets. By adding it to ELAN-W, ELAN can learn more and better characteristic information, suppress the surrounding complex background information, and highlight the characteristic information of ship targets. The structure of ELAN-A is shown in Fig. 3.

D. WEF-Net

In order to solve the difficulty of detecting ship targets in SAR images under complex background, this article designs

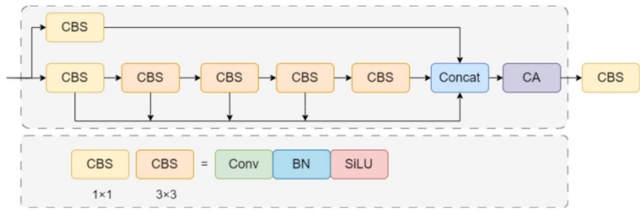

Fig. 3. Structure of ELAN-A. ELAN-A can help extract effective features that focus more on the target. CBS is composed of convolution, batch normalization, and SiLU activation function. CA represents the coordination attention mechanisms. Concat represents channel cascade operation.

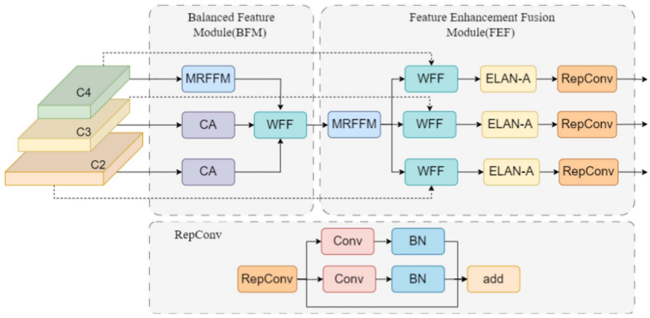

Fig. 4. Structure of WEF-Net. The WEF-Net is divided into balanced feature module (BFM) and feature enhancement fusion module (FEF). BFM can balance the information of each backbone layer and generate the global features of the sensitivity field balance, while FEF can balance the semantic information of different levels of the feature pyramid by averaging the global features. CA represents the coordination attention mechanisms. WFF represents weighted feature fusion module. RepConv represents reparameterized convolution.

a neck network WEF-Net, which balances the feature maps of deep layer and shallow layer with different resolutions. It includes two parts: a BFM and an FEF. BFM can balance the information of each backbone layer and generate the global features of the sensitivity field balance, while FEF can balance the semantic information of different levels of the feature pyramid by averaging the global features. Because different feature maps have different contributions, to balance the semantic information and resolution of different layers, WEF-Net balances the proportion of deep and shallow position information, improves the effectiveness of information, and reduces the redundancy of information. The structure of WEF-Net is shown in Fig. 4.

First, in the BFM, the last three layers of feature diagrams $C2$, $C3$ and $C4$ output from the backbone network are sent into the BFM module. First, we conduct an MRFFM for $C4$, and then use upsampling and downsampling, respectively, for $C4$ and $C2$ layers to adjust the shape to the same as $C3$. Subsequently, the CA mechanism is introduced to layers $C2$ and $C3$, aiding the network in learning location information across different layers and channels. This facilitates the generation of adaptively balanced features and the adjustment of weights through coordinated attention. It is worth noting that $C4$ is not followed by a coordinated attention mechanisms, which is because the $C4$ layer, as the deepest layer, has the richest semantic information and keeps its original feature information from being lost. Then, the feature maps of $C2$, $C3$, and $C4$ branches are transformed into

feature maps with the same resolution. Finally, using the idea of weighted BiFPN, a WFF is designed to fuse the three branches' feature graphs. WFF is not a simple cascade of channels; instead, it sets a set of learnable parameters for each feature map at different scales to carry out weighted cooperative operations. This approach is able to identify the importance of different feature maps during training, ensuring that key information about small ship targets is retained. The fusion process of BFM can be expressed as follows:

$$F = \frac{\begin{array}{c} w_1 \left(\text{UpSample}\left(\text{MRFFM}\left(C4\right)\right)\right) + w_2 \left(\text{CA}\left(C3\right)\right) \\ +w_3 \left(\text{MP}\left(\text{CA}\left(C2\right)\right)\right) \end{array}}{w_1 + w_2 + w_3 + \Delta} \tag{5}$$

where w_1, w_2, and w_3 are learnable parameters, but Δ is deliberately set as a small value to prevent numerical instability. $Ci(i= 2, 3, 4)$ is the initial feature map of the corresponding layer, CA represents the coordination attention mechanisms, MP represents the maximum pooling operation, it can achieve the function of downsampling, and F is the fusion output feature map. UpSample indicates the upsampling operation.

In the FEF, the BFM module inputs the balanced fusion features to the multifield extension module MRFFM for context feature extraction, and then decomposed into three feature maps with the same resolution as the original three-layer feature map. The feature maps output from MRFFM are upsampled and downsampled, respectively, WFF is performed with the initial features of *C*2 layer and *C*4 layer with the same resolution, respectively, and then input to ELAN-A module for feature learning, and finally transmitted to reparameterized convolutional RepConv. In addition, there is a branch from the MRFFM output directly WFF with the C3 layer initial feature map with the same resolution, which is then transmitted to ELAN-A and RepConv for feature extraction and reparameterized convolution. The process of FEF is shown as follows:

$$P4 = \text{Rep}\left(E\left(\frac{w_1 \left(\text{MP}(\text{MRFFM}(F))\right) + w_2 \left(C4\right)}{w_1 + w_2 + \Delta}\right)\right) \tag{6}$$

$$P3 = \text{Rep}\left(E\left(\frac{w_1 \left(\text{MRFFM}(F)\right) + w_2 \left(C3\right)}{w_1 + w_2 + \Delta}\right)\right) \tag{7}$$

$$P2 = \text{Rep}\left(E\left(\frac{w_1 \left(\text{UpSample}(\text{MRFFM}(F))\right) + w_2 \left(C2\right)}{w_1 + w_2 + \Delta}\right)\right) \tag{8}$$

where w_1 and w_2 are learning parameters, but Δ is a small number to avoid numerical instability, $Ci(i = 2, 3, 4)$ is the initial feature map of the corresponding layer, $Pi(i = 2, 3, 4)$ is the neck output feature map of the corresponding layer $Ci(i = 2, 3, 4)$, MP indicating the maximum pooling operation, it can realize the function of downsampling. F is the fusion feature map, E is the ELAN-A, and Rep is the RepConv operation.

In addition, we introduce the coordinated attention mechanisms in BFM, the primary steps of CA entail embedding coordinate information and generating coordinate attention, enabling the encoding of channel relationships. The input feature map X is first pooled globally in horizontal and vertical directions and compressed into $1 \times H \times C$ and $W \times 1 \times C$, respectively. Subsequently, these two compressed feature maps

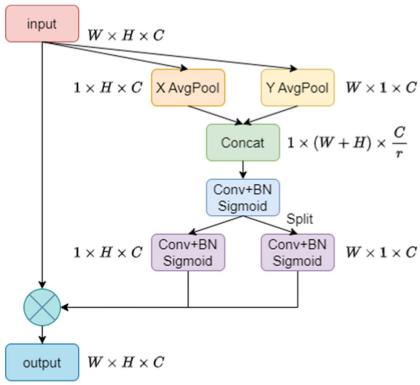

Fig. 5. Structure of CA. BN represents the batch normalization operation. *X* AvgPool and *Y* AvgPool represent average pooling operations in the horizontal and vertical directions, respectively. Sigmoid represents nonlinear activation function. Concat represents channel cascade operation.

are fused together, resulting in concatenated outputs consolidated to $1 \times (W + H) \times C/r$ by 1×1 convolution block. The amalgamated result is then split into two different feature maps, with their dimensions transformed into $1 \times H \times C$ and $W \times 1 \times C$ by two additional 1×1 convolution and sigmoid functions, respectively, generating weighted matrices. Finally, the input feature map X is subjected to multiplication with the two weighting matrices. The operational procedure of the CA module can be succinctly summarized as follows:

$$Z_C^h(h) = \frac{1}{W} \sum_{0 \le i \le W} X_c(h, i) \tag{9}$$

$$Z_C^w(w) = \frac{1}{H} \sum_{0 \le j \le H} X_c(j, w) \tag{10}$$

$$F = \sigma(\text{BN}(\text{Conv}_{1\times1}\left[{Z_C}^h, {Z_C}^w\right])) \tag{11}$$

$$F_c(i, j) = X_c(i, j) \times \sigma\left(\text{Conv}_{1\times1}\left(F^h\right)\right) \times \sigma\left(\text{Conv}_{1\times1}\left(F^w\right)\right) \tag{12}$$

where W and H are the width and height of the input feature map, respectively, and $Z_C^h(h)$ and $Z_C^w(w)$ represent the result of X average pooling and Y average pooling, respectively. $F \in R^{1\times(W+H)\times C/r}$ represents the intermediate vector, F^h and F^w of the attention weights are obtained by the sigmoid activation function. $\text{Conv}_{1\times1}$ represents the convolution of the kernel size of 1×1, σ represents the sigmoid activation function, BN represents the batch normalization operation, and $[\cdot]$ represents the channel cascade operation. The CA structure is shown in Fig. 5.

E. Optimized Loss Function

With the application of CNNs in object detection, the conventional loss function proves inadequate for assessing the similarity between predicted and actual frames. In response, researchers have proposed alternative loss functions, such as Intersection over Union (IoU) loss [37], [38], [39], with CIoU loss emerging as a pioneering approach. In contrast to previous IoU, CIoU incorporates considerations for overlap area, center

point distance, and aspect ratio [38]. The CIoU loss function can be mathematically represented as follows:

$$R_{\mathrm{CIoU}} = \frac{D^2}{C^2} + \alpha\upsilon \tag{13}$$

$$\upsilon = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{14}$$

$$\alpha = \frac{\upsilon}{(1-\mathrm{IoU})+\upsilon} \tag{15}$$

$$L_{\mathrm{CIoU}} = 1 - \mathrm{IoU} + R_{\mathrm{CIoU}} \tag{16}$$

where D is the Euclidean distance between the center point of the prediction box and the real box. C is the diagonal length of the smallest surrounding box covering the two boxes, α is the weight factor, and υ is used to measure the similarity of the aspect ratio between the real box and the predicted box. In addition, w^{gt} is the width of the ground reality box, h^{gt} is the height of the ground reality box, w is the width of the prediction box, and h is the height of the prediction box.

The sensitivity of CIoU losses to different scale targets varies greatly, and for tiny objects with 6×6 pixels, small positional deviations can lead to significant IoU drops. For normal objects with 36×36 pixels, the IoU changes less, which will lead to defects in label allocation. So CIoU is not a good metric for small objects.

Henceforth, the article employs the NWD loss function [40], a novel metric leveraging Wasserstein distance to quantify the similarity between bounding boxes, as opposed to the conventional IoU metric. Notably, NWD demonstrates insensitivity to object displacement across varying scales, rendering it particularly adept at assessing similarity among minute objects. Moreover, the NWD loss function integrates a sample weight mechanism, enabling the model to prioritize the detection performance of small targets. However, the applicability of the NWD loss function is primarily confined to scenarios involving small targets, given its inherent scale sensitivity. While it effectively mitigates missed detections for small targets, it may inadvertently induce false detections and impede convergence rates for medium and large targets. The formula for the NWD loss function is delineated as follows:

$$W_2^2(N_a, N_b) = \left\| \left([cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2}]^T, \left[cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2}\right]^T \right) \right\|_2^2 \tag{17}$$

$$\mathrm{NWD}(N_a, N_b) = e^{\left(-\frac{\sqrt{W_2^2(N_a,N_b)}}{c}\right)} \tag{18}$$

where $N_a = (cx_a, cy_a \frac{w_a}{2}, \frac{h_a}{2})$ denotes boundary box A, $N_b = (cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2})$ denotes boundary box B, $\|\cdot\|_2$ denotes 2-norm, and c is a constant factor.

To solve this problem, we combine CIoU and NWD loss functions to leverage their respective strengths and optimize the loss function [41], [42], where, *r*, as a tradeoff factor, can effectively balance the NWD index and CIoU method by adjusting its value. When *r* is large, the NWD loss does not have much effect, making the boundary box loss similar to CIoU. Conversely, with a smaller *r*, the boundary box is conceptualized as a Gaussian distribution. Consequently, the refined boundary box loss demonstrates robust performance in detecting small-scale SAR ship targets. The calculation formula can be articulated as follows:

$$L_{\mathrm{CIoU-NWD}} = (1-r) * (1-\mathrm{NWD}) + r * L_{\mathrm{CIOU}}. \tag{19}$$

In addition, for confidence loss L_{obj} and classification loss L_{cls}, we use binary cross-entropy loss (BCEWithLogitsLoss) to calculate [43], the formula is delineated as follows:

$$L_{\mathrm{BCE}} = -\sum_{n=1}^{N} \hat{y_i} \log(\sigma(y)) + (1-\hat{y_i}) \log(\sigma(1-y)) \tag{20}$$

where $\hat{y_i}$and y are prediction vectors and real vectors, respectively. σ is the sigmoid function.

To sum up, we propose that the loss function of the model is divided into three parts: L_{box}, L_{obj}, and L_{cls}, respectively, measure the loss of the detector on boundary box regression, object confidence, and classification, P_{box} and P'_{box}, respectively, represent the prediction box and ground reality box. P_{obj} and P'_{obj} represent forecast confidence and ground reality confidence, respectively. In addition, P_{cls} and P'_{cls} represent forecast probability and ground reality probability, respectively. Hyperparametersλ_1, λ_2, and λ_3 default to 0.05, 0.7, and 0.3 as follows:

$$\begin{aligned}\mathrm{Loss} &= \lambda_1 L_{\mathrm{box}} + \lambda_2 L_{\mathrm{obj}} + \lambda_3 L_{\mathrm{cls}} \\ &= \lambda_1 L_{\mathrm{CIoU-NWD}}(P_{\mathrm{box}}, P'_{\mathrm{box}}) + \lambda_2 L_{\mathrm{BCE}}(P_{\mathrm{obj}}, P'_{\mathrm{obj}}) \\ &\quad + \lambda_3 L_{\mathrm{BCE}}(P_{\mathrm{cls}}, P'_{\mathrm{cls}}).\end{aligned} \tag{21}$$

F. Model Flowchart

Fig. 6 depicts the overall detection flow of the BFEA, which is divided into three stages: dataset preparation, training, and inference. First, we collect SAR images from a dataset and annotate the images, then use data enhancement to get a more robust dataset for training and validation. We then optimize and train the BFEA model on the developed dataset until the loss converges. In the inference stage, we use the nonmaximum suppression method to filter the redundant bounding boxes to improve the positioning accuracy and detection effect. Finally, the optimal model of SAR image target detection is obtained.

IV. Experiments

In this section, we begin by providing a succinct overview of the datasets and experimental setup employed in this article. Subsequently, we compare the detection outcomes of our model against those of several other object detection models, showcasing the merits of our proposed approach. Finally, we conduct a series of ablation experiments and optimized loss function tradeoff factor control experiments. These experiments are designed to substantiate the capability of our proposed BFEA detection model in effectively detecting multiscale ship targets amid complex backgrounds.

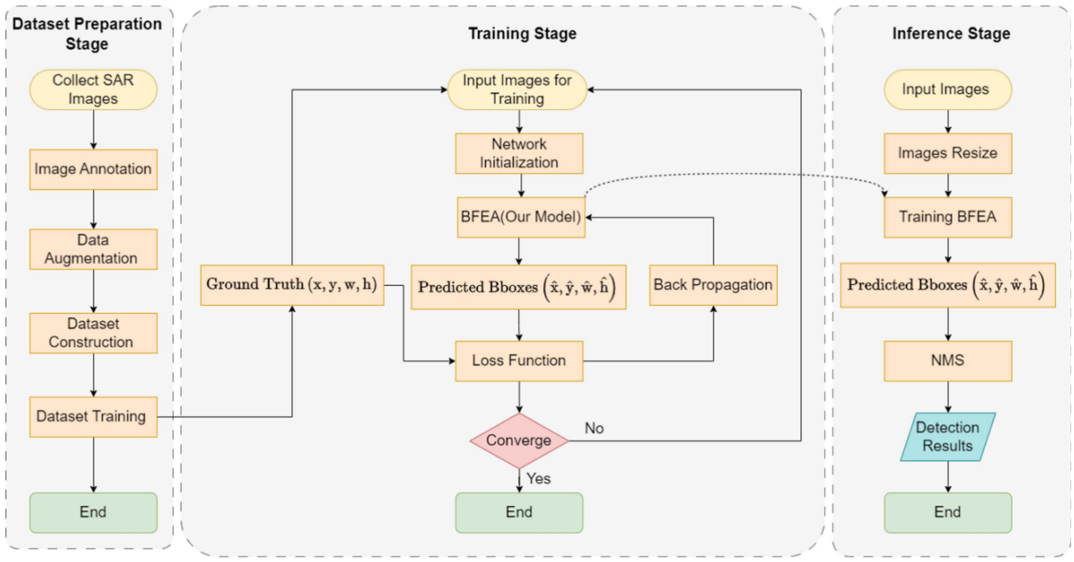

Fig. 6. Our model flowchart. It includes three stages: dataset preparation stage, training stage, and inference stage.

TABLE I
STATISTICS OF THE THREE DATASETS

Details	SSDD	HRSID	LS-SSDD-v1.0
Satellite	RadarSat-2,TerraSAR-X, Sentinel-1	Sentinel-1B,TerraSAR-X,TanDEM	Sentinel-1
Polarization	HH, HV, VV, VH	HH, HV, VV	VV, VH
Resolution (m)	1–15	0.5, 1, 3	5 × 20
Image number	1160	5604	9000
Train/test ratio	928/232	3643/1961	6000/3000
Ship number	2456	16,951	6015
Image size(pixel2)	217×214-526×646	800×800	800×800
Average image size (pixel2)	481×331	800×800	800×800

A. Experimental Setup and Hyperparameters

In the experiment, the experimental environment was configured with an NVIDIA GeForce RTX 3070 graphics card (8 GB), an Intel Core i7-10700K CPU, and an Ubuntu 20.04.4LTS operating system.

In our experiment, we compare the detection performance of our detector with other target detectors. During the training phase, we employed the SGD optimizer. For SSDD, we examined the SAR image size in this dataset and the average SAR image size on SSDD was 481 × 331. Therefore, we set the input size to 512 × 512, which does not result in a large loss of image information. HRSID and LS-SSDD-v1.0 use a standard image resolution of 800 × 800.

B. Dataset

This article evaluates the effectiveness of our proposed method using three publicly available SAR datasets: the SSDD [44], the HRSID [45], and the LS-SSDD-v1.0 [46]. Details of these datasets are summarized in Table I.

C. Evaluation Index

In object detection tasks, in order to calculate mAP, we first need to determine the precision and recall ratio. Accuracy (P) and recall (R) are defined as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{22}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{23}$$

In the formula, true positive (TP) represents the instances where the model correctly predicts a positive label that aligns with the ground truth. False positives (FP) denote instances where the model incorrectly predicts a positive label for a negative ground truth. False negatives (FN) indicate instances where the model incorrectly predicts a negative label for a positive ground truth. A test result is deemed TP only if its true IoU exceeds the specified IoU threshold (typically 0.5).

By varying the IoU thresholds, we can construct a precision–recall (PR) curve. The area under the PR curve, denoted as mAP, is calculated to quantify the model's performance. The mAP calculation formula is shown as follows:

$$\mathrm{mAP} = \int_0^1 P(R)dR. \tag{24}$$

The $F1$ score is a metric by which the performance of the model is analyzed, and is defined as follows:

$$F1 = \frac{2 \times P \times R}{P + R}. \tag{25}$$

D. Results of SSDD

As shown in Table II, we compared BFEA with some popular and universal detectors in the SSDD dataset, and our $F1$, mAP, recall rate, and accuracy rate all achieved the best results. They were 95.1%, 98.0%, 95.2%, and 95.1%, respectively.

TABLE II
COMPARISON RESULTS WITH OTHER OBJECT DETECTION MODELS ON SSDD

Method	mAP	*P*	*R*	*F*1
Faster R-CNN[8]	91.3	87.3	89.2	88.3
Libra R-CNN[10]	93.9	86.1	94.6	90.1
YOLOv4[47]	92.1	92.8	84.3	88.3
YOLOv5-l[16]	92.1	93.6	89.1	91.2
YOLOv7[17]	94.6	91.0	90.3	90.6
SAR-ShipNet[48]	89.1	95.1	76.3	84.6
GFB-Net[49]	93.0	85.6	94.0	89.6
CR2A-Net[50]	89.8	94.0	87.8	90.7
PPA-Net[51]	95.2	95.2	91.2	93.1
YOLO-FA[52]	96.8	95.2	95.0	95.0
BiFF-FESA[53]	97.8	**96.1**	94.0	95.0
BFEA(ours)	**98.0**	95.2	**95.1**	**95.1**

The best results are in bold.

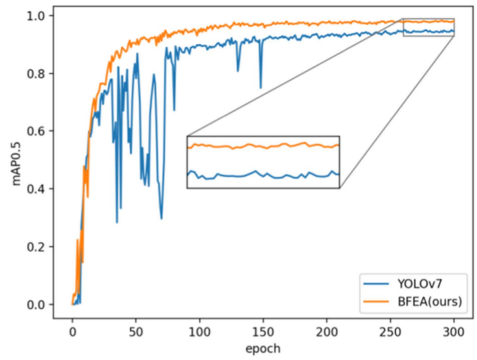

Fig. 7. Average accuracy curve of baseline and our proposed BFEA on the SSDD dataset.

In comparison to other SAR ship detectors, our model exhibits superior detection performance. While BiFF–FESA achieves a commendable precision rate of 96.1%, our model outperforms in *F*1 score and mAP. Notably, exceed the BiFF–FESA in the table for the best result. The experimental findings unequivocally demonstrate the superior detection capabilities of our model compared to other target detectors.

In Fig. 7, the average accuracy curve of our model and the baseline on the SSDD dataset are drawn to directly compare the learning ability and detection effect of the model. At the beginning, the average accuracy curves of the two models increase sharply, but the oscillation was obvious. After training with about 150 epochs, the mean accuracy curve of the baseline begins to level off. That is to say, these modules fully learn the characteristics of SAR ships through training. The baseline curve is significantly lower than the BFEA curve. In other words, our model learned better than the baseline.

In addition, the visualization of the BFEA versus the baseline is shown in Fig. 8. It can be seen from Fig. 8(a) and (b) that a misdetection occurred at the baseline, misdetecting the ship's tail waves and buildings as ships. Fig. 8(c) and (d) show simultaneous missed and false detection of the baseline, mistaken detection of shore structures as ships, and missed detection of overlapping ships. It shows that the BFEA can correctly identify ship information in complex scenes with surrounding background interference. And it has a good extraction effect on different scale ship features. Our proposed model can detect multiple small ships offshore and large ships near shore.

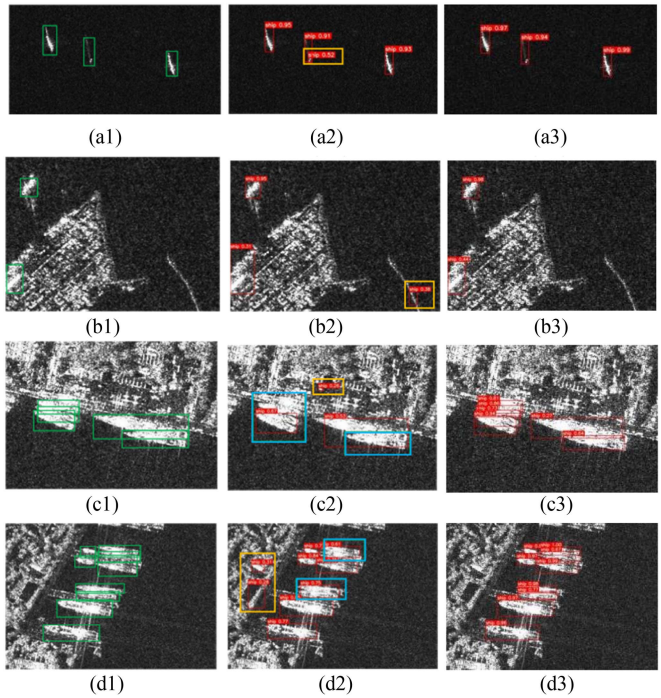

Fig. 8. Visualization result of our designed BFEA on SSDD. The first column is the original SAR image and green box is ground truth box, the second column is the corresponding baseline test results, and the third column is the test results of our model. The yellow box indicates false detection, and the blue box indicates missed detection. (a) Sparse offshore scenes. (b) Near shore scenes with complex background interference. (c) and (d) Dense nearshore scenes.

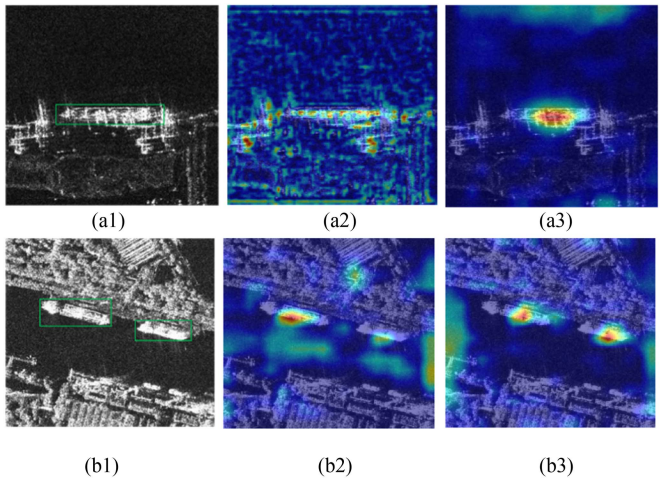

Fig. 9. Thermal map results of baseline and our proposed BFEA on the SSDD dataset. The first column is the original SAR image and green box is ground truth box, the second column is the corresponding baseline heat map results, and the third column is the heatmap results of our model.

To further validate the efficiency of feature extraction on the SSDD dataset by our proposed method, we conducted a Grad-CAM visual test. This visual test involved comparing the results between the baseline and our proposed method, as illustrated in Fig. 9. Regions highlighted in red suggest heightened attention

TABLE III
COMPARISON RESULTS WITH OTHER OBJECT DETECTION MODELS ON HRSID

Method	mAP	P	R	$F1$
Faster R-CNN[8]	79.4	80.2	82.1	81.1
Cascade R-CNN[54]	87.4	89.0	85.7	87.3
YOLOv4[47]	81.7	85.9	76.3	80.8
YOLOv5[16]	85.9	85.6	85.5	85.5
YOLOv7[17]	90.0	91.9	80.1	85.5
DetNet[55]	83.2	84.3	80.6	82.4
PPA-Net[51]	89.2	90.3	88.2	89.2
ATSD[56]	88.1	90.2	86.5	88.3
DRPSM[57]	86.5	-	**90.1**	-
SIFNet[58]	79.7	79.9	82.7	81.2
RFD[59]	92.4	-	84.3	-
BFEA(ours)	**93.1**	**93.6**	85.9	**89.5**

The best results are in bold.

TABLE IV
COMPARISON RESULTS WITH OTHER OBJECT DETECTION MODELS ON LS-SSDD-V1.0

Method	mAP	P	R	$F1$
Faster R-CNN[8]	57.7	58.0	61.6	59.7
Cascade R-CNN[54]	59.0	54.1	66.2	59.5
YOLOv5[16]	73.3	84.0	63.6	72.3
YOLOv7[17]	71.6	**86.1**	63.3	72.9
YOLOv8-s[60]	74.4	82.4	67.0	73.9
SII-Net[61]	76.1	68.2	**79.3**	73.3
MHASD[21]	75.5	83.4	67.9	74.8
LRTransDet[41]	76.2	82.6	68.2	74.7
BFEA(ours)	**76.9**	83.5	68.8	**75.4**

The best results are in bold.

TABLE V
COMPARISON RESULTS WITH OTHER OBJECT DETECTION MODELS ON HRSID DATASET

Method	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
FCOS[13]	87.0	65.9	60.7	63.4	25.2
RetinaNet[14]	80.1	59.4	54.6	60.0	22.1
Mask R-CNN[9]	81.1	67.7	59.9	63.7	13.9
YOLO-SD[62]	83.7	67.6	62.3	51.9	1.3
LPEDet[63]	89.7	71.8	65.8	63.4	24.2
FEPS-Net[64]	90.7	74.3	66.8	65.2	31.6
SDNet[65]	90.6	75.1	66.1	72.1	56.9
ESD-Net[66]	91.0	77.7	67.6	65.6	40.8
NAS-YOLOX[67]	91.1	71.9	65.2	68.6	34.1
SFPN[68]	91.2	77.3	**69.0**	69.6	29.7
BFEA	**93.1**	**77.9**	57.4	**80.0**	**64.2**

The best results are in bold.

TABLE VI
COMPARISON RESULTS WITH OTHER OBJECT DETECTION MODELS ON LS-SSDD-V1.0

Method	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Faster R-CNN[8]	66.6	10.8	24.6	37.4	—
Libra faster R-CNN[10]	65.9	10.0	24.0	38.5	—
RetinaNet[14]	60.7	7.2	20.2	36.7	—
YOLO-SD[62]	72.0	12.2	27.1	39.9	—
ATSS[69]	66.5	8.9	23.3	35.2	—
LssDet[70]	74.3	9.4	25.8	36.9	—
BFEA	**76.9**	**15.0**	**28.7**	**42.3**	—

The best results are in bold.

by the network, whereas regions in blue indicate lower attention by the network.

Fig. 9 depicts the thermal map visualization results of our proposed model on the SSDD dataset. To provide a clearer illustration of the detection performance, we selected two near-shore SAR images with intricate backgrounds for comparison in our experiment. As evident from the figure, YOLOv7 struggles to accurately detect ships in the near-shore scene, resulting in both missed and false detections. This deficiency can be attributed to the complexity of the near-shore background and the consequent interference, which adversely affects the performance of YOLOv7. In contrast, our proposed BFEA exhibits a superior capability in capturing ships amid complex background interference near the shore. This is mainly due to the ELAN-A module and WEF-Net, which can suppress background interference while highlighting the ship's main features.

E. Results of Large-Scale Datasets

In the model verification phase, we applied the model to two large-scale SAR ship detection platforms HRSID and LS-SSDD-v1.0 for multiscale ship detection.

As shown in Table III, we compared the BFEA model to other target detectors in the HRSID dataset, and our model achieved an mAP of 93.1%, obtaining better mAP and $F1$ than other SAR ship detectors. On the HRSID dataset, its mAP is 0.7% higher than the best detector RFD in the table, and although DRPSM gets a higher indicator in recall, our model gets the best results on mAP, accuracy, and $F1$, with a 6.6% improvement in mAP. Our model achieves a great advantage on this dataset.

We also compared our model to other target detectors in the LS-SSDD-v1.0 dataset, and the results are shown in Table IV. The experimental results show that BFEA achieves 76.9% mAP on the LS-SSDD-v1.0 dataset and has better mAP and $F1$ than other target detectors in large-scale SAR image detection tasks. In LS-SSDD-v1.0, we achieved the best $F1$ by improving mAP by 0.7% compared to the best detector, LRTransDet. These results prove that our model performs well on large-scale datasets with a large number of small ships.

To comprehensively assess the efficacy of our methods, we conducted comparative evaluations with other detectors on the HRSID and LS-SSDD-v1.0 datasets, employing a range of evaluation metrics including AP_{50}, AP_{75}, AP_S, AP_M, and AP_L. These metrics adhere to the standardized definitions outlined in the COCO dataset. Specifically, AP_{50} denotes performance at IoU $= 0.5$, while AP_{75} offers a more rigorous evaluation reflecting positional accuracy at IoU $= 0.75$. Additionally, AP_S, AP_M, and AP_L metrics gauge our model's ability in detecting objects across various scales, with AP scores computed for small (area $< 32^2$ pixels), medium ($32^2 <$ area $< 64^2$ pixels), and large (area $> 64^2$ pixels) targets, as per the defined area occupancy thresholds. In our analysis, AP_{50} serves as the primary performance indicator, with other metrics serving as supplementary references. Detailed results are tabulated in Tables V and VI for comprehensive evaluation.

We compared the proposed BFEA with 15 other target detectors in the HRSID and LS-SSDD-v1.0 datasets. As shown in

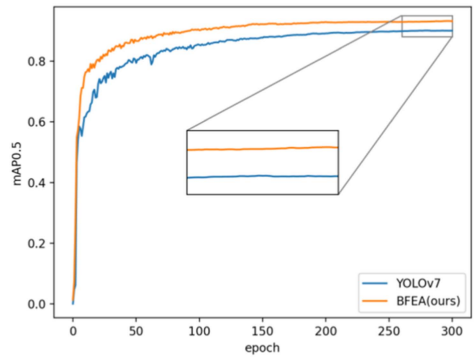

Fig. 10. Average accuracy curve of baseline and our proposed BFEA on the HRSID dataset.

the table, BFEA performs AP_{50}, AP_{75}, AP_M, and AP_L on the HRSID dataset achieves the best results (93.1%, 77.9%, 80.0%, and 64.2%, respectively). Although our model is lower than other models in small size ship detection, our model is 1.9%, 0.6%, 10.4%, and 34.5% higher than the best model SFPN in the table on AP_{50}, AP_{75}, AP_M, andAP_L, respectively. In addition, the results of BFEA on the LS-SSDD-v1.0 dataset were superior to those of other methods (76.9%, 15.0%, 28.7%, and 42.3%, respectively). This may be because WEF-Net can balance different levels of semantic information. Meanwhile, ELAN-A further improves the detector's performance by suppressing background interference and highlighting ship targets.

Fig. 10 illustrates the average accuracy curves of both our model and the baseline on the HRSID dataset, providing a direct comparison of their learning capabilities and detection efficacy. Initially, during the early stages of training, both models exhibited sharp increases in detection performance, albeit with noticeable oscillations. However, after approximately 100 epochs, the mean accuracy curves of both the baseline and our proposed method began to stabilize. Last, with the curve of our proposed BFEA model notably surpassing that of the baseline. This observation suggests that our model demonstrates superior learning performance on the HRSID dataset compared to the baseline.

In addition, the visualization results of BFEA and baseline on the HRSID dataset are shown in Fig. 11. In Fig. 11(a), it can be seen that YOLOv7 misdetected the sea clutter and shore as ships. In Fig. 11(b), YOLOv7 missed detection. In Fig. 11(c) and (d), YOLOv7 has both missed and false detections, but our model can correctly detect all ships. Compared with SSDD, there are more complex scenes and small targets in HRSID nearshore SAR images, but the BFEA proposed in this article can correctly identify these small ships even in complex scenes with surrounding background interference.

Fig. 12 depicts the average accuracy curve comparing our model with the baseline on the LS-SSDD-v1.0 dataset, facilitating a direct assessment of their learning capacity and detection efficacy. During the initial stages of training, both models exhibited sharp increases in detection performance, although notable fluctuations were observed, particularly in the baseline model. Following approximately 150 epochs of training, the mean accuracy curves of both the baseline and our proposed

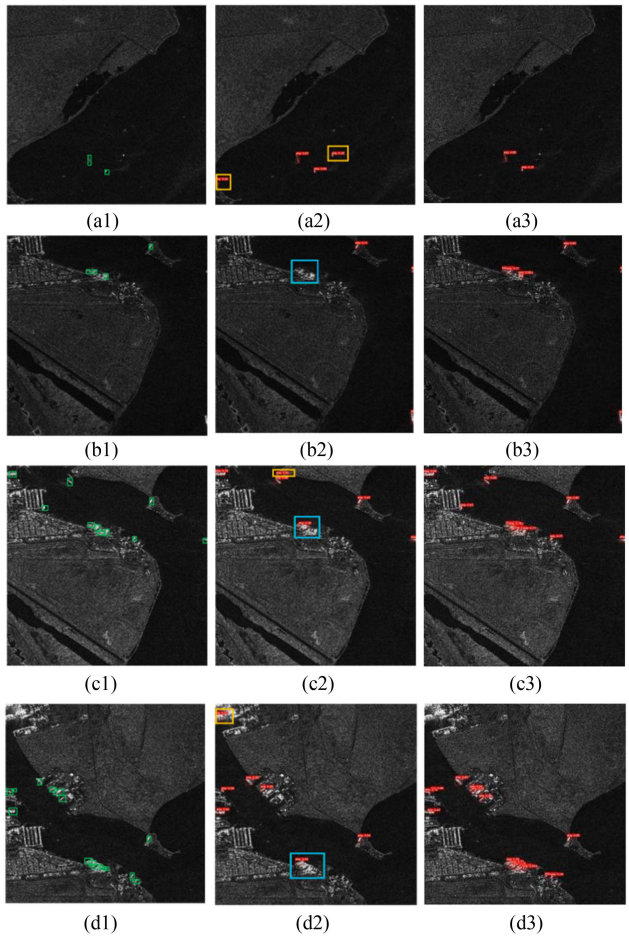

Fig. 11. Visualization of our designed BFEA on HRSID. The first column is the original SAR image and green box is ground truth box, the second column is the corresponding baseline test results, and the third column is the test results of our model. The yellow box indicates false detection, and the blue box indicates missed detection. (a) Sparse offshore scenes. (b) Near shore scenes with complex background interference. (c) (d) Dense nearshore scenes.

method began to stabilize. Last, with the curve of our proposed BFEA model markedly outperforming that of the baseline. This indicates that our model achieves superior learning outcomes on the LS-SSDD-v1.0 dataset compared to the baseline.

In addition, the visualization results of BFEA and baseline on the LS-SSDD-v1.0 dataset are shown in Fig. 13. In Fig. 13(a), (b), (c), and (d), the baseline was missed, and our model correctly detected the missed vessel.

F. Ablation Experiment

To verify the role of each BFEA module in the SAR ship target detection, the ablation experiments were carried out on the SSDD dataset. The experimental results are shown in Table VII.

Upon scrutinizing the findings presented in Table VII, our experimental results align closely with the underlying theoretical principles. Within the WEF-Net architecture, the MRFFM module ensures a balanced receptive field, thereby capturing essential global context information and compensating for the limited semantic features inherent to SAR ships. As indicated in the first row of the table, the baseline achieved only 94.5%

TABLE VII
ABLATION EXPERIMENTS ON SSDD DATASETS

Baseline	WEF-Net	NWD Loss	ELAN-A	P	R	$F1$	mAP_{50}	mAP_{50-95}
√				89.0	90.0	90.0	94.5	63.9
√	√			94.7	89.0	91.7	96.1	66.0
√		√		95.9	91.2	93.4	96.0	65.9
√			√	94.9	91.8	93.3	96.2	66.1
√	√	√		93.1	94.5	94.0	97.0	69.3
√	√	√	√	95.2	95.1	95.1	98.0	69.3

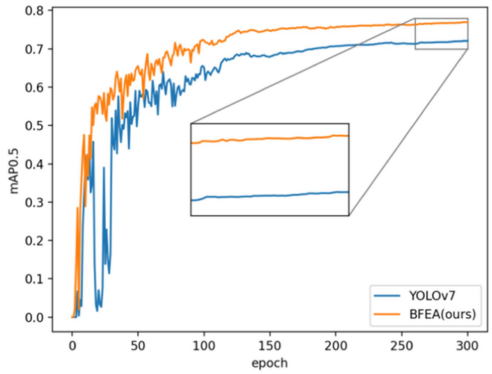

Fig. 12. Average accuracy curve of baseline and our proposed BFEA on the LS-SSDD-v1.0 dataset.

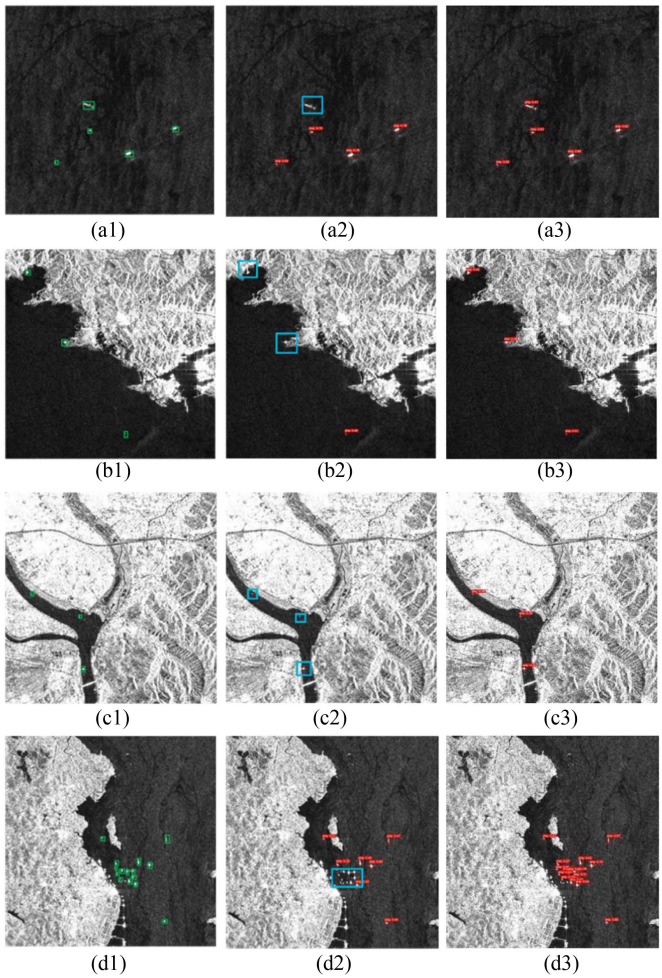

Fig. 13. Visualization result of our designed BFEA on LS-SSDD-v1.0. The first column is the original SAR image and green box is ground truth box, the second column is the corresponding baseline test results, and the third column is the test results of our model. The yellow box indicates false detection, and the blue box indicates missed detection. (a) Sparse offshore scenes. (b) Near shore scenes with complex background interference. (c) Nearshore scenery in a narrow river. (d) Dense nearshore scene.

mAP_{50}, 89.0% accuracy, and 90.0% recall. After the introduction of WEF-Net, mAP_{50} increased to 96.1%, recall rate and accuracy increased by 1.0% and 5.7%, respectively. Considering the prevalence of small targets in SAR images, we employ an optimized loss function tailored for small target detection. This loss function adeptly balances CIoU and NWD losses without imposing additional computational burden. By leveraging the optimized loss function, we preserve the antifalse detection capability inherent to CIoU while enhancing the model's ability to mitigate false detections of small targets through NWD loss integration. Our optimized loss function improved accuracy by 6.9%, recall by 1.2%, and $F1$ scores by 3.4% compared to the baseline model. The effectiveness of the optimized loss function is proven. In addition, the collaboration between WEF-Net and the optimized loss function resulted in a 2.5% improvement in mAP_{50} (from 94.5% to 97.0%) and mAP_{50-95} increased by 5.4% (from 63.9% to 69.3%). The role of the ELAN-A module is to highlight the information of the ship's target and suppress clutter in complex scenes. Based on experimental data, mAP_{50} improved by 1.7% over baseline after the introduction of ELAN-A, and in synergy with WEF-Net and the optimized loss function, accuracy and recall improved by 6.2% and 5.1% over the baseline model, mAP_{50} improved to 98.0%.

G. Optimized Loss Function Tradeoff Factor Control Experiment

To further investigate the effectiveness of our optimized loss function, we conducted control experiments on the HRSID dataset based on different r values of BFEA. Due to the large proportion of small targets in the SAR image dataset, we should reduce this value, that is, increase the NWD loss measure. Therefore, we define the r value as a threshold value between 0.1 and 0.5, and adjust the r value to find the optimal setting result. The experimental results are shown in Table VIII.

The analysis of the data provided in Table VIII reveals a trend: the recall rate tends to be higher when the value of r

TABLE VIII
RESULTS OF EACH INDEX OF DIFFERENT TRADEOFF FACTORS *r*

r	*P*	*R*	*F*1	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
0.1	92.9	86.1	89.0	92.1	74.4	54.2	78.3	66.3
0.2	92.5	87.3	90.0	93.5	76.0	56.1	79.3	62.6
0.3	93.6	85.9	90.0	93.1	77.9	57.4	80.0	64.2
0.5	93.3	84.9	89.0	92.8	77.6	56.5	79.9	64.2

is small. At this point, NWD losses dominate the loss function, aligning with our hypothesis that NWD contributes to reducing false negatives in the detection of small ship targets. Conversely, as the value of *r* increases, CIoU loss dominates, and the accuracy *P* is improved. Finally, after considering multiple indicators, we choose the tradeoff factor $r = 0.3$ that achieves the optimal result in AP_S, AP_M, and $F1$ scores as the result of loss calculation. The tradeoff factor control experiments indicating that our optimized loss function performs exceptionally well in detecting small-scale SAR ship targets. This balance is reflected in the respective measures *P* and *R*, with the $F1$ score serving as a comprehensive metric to evaluate the efficacy of the improved loss function. Notably, the data presented in Table VIII clearly demonstrate that $F1$ scores exhibit superior performance when the tradeoff factor *r* is set to 0.3.

V. CONCLUSION

This article presents a BFEA model for detecting SAR multiscale ships in complex environments. An WEF-Net is designed to aggregate and balance semantic information from different levels of features. A MRFFM is embedded in the network to learn the context information, generate the global characteristics of the receptive field balance, and enhance the detection performance of the model against ships of different scales. Then, we redesigned the ELAN module by combining convolution and attention, so that it can suppress the surrounding complex background interference while highlighting the ship's features. Finally, for small ship detection, we combine NWD measurement with CIoU measurement to optimize the loss function, which ensures strong small ship detection performance. The experimental results on SSDD, HRSID, and LS-SSDD-v1.0 show that our detector has better performance than other target detection models. Through a series of ablation experiments and tradeoff factor control experiments, we showcase our proposed feature fusion network WEF-Net, ELAN-A module, and optimized loss function can enhancing the network's performance in the detection task.

Although our model shows excellent detection performance on three challenging SAR datasets. However, the experimental results show that our model performs better on the detection of medium- and large-scale ship targets in SAR images, and there is still some room for improvement in the detection of small-scale ship targets. Therefore, in the future article, we will further design a loss function that is more suitable for SAR image small target detection, optimize its module structure, solve its limitations on small target detection, and further improve the performance of multiscale ship detection model.

REFERENCES

[1] F. Sharifzadeh, G. Akbarizadeh, and Y. Seifi Kavian, "Ship classification in SAR images using a new hybrid CNN–MLP classifier," *J. Indian Soc. Remote Sens.*, vol. 47, pp. 551–562, 2019.

[2] N. Davari, G. Akbarizadeh, and E. Mashhour, "Corona detection and power equipment classification based on GoogleNet-AlexNet: An accurate and intelligent defect detection model based on deep learning for power distribution lines," *IEEE Trans. Power Del.*, vol. 37, no. 4, pp. 2766–2774, Aug. 2022.

[3] F. Samadi, G. Akbarizadeh, and H. Kaabi, "Change detection in SAR images using deep belief network: A new training approach based on morphological images," *IET Image Process.*, vol. 13, pp. 2255–2264, 2019.

[4] N. Aghaei, G. Akbarizadeh, and A. Kosarian, "Osdes_Net: Oil spill detection based on efficient_shuffle network using synthetic aperture radar imagery," *Geocarto Int.*, vol. 37, no. 26, pp. 13539–13560, 2022.

[5] N. Aghaei, G. Akbarizadeh, and A. Kosarian, "GreyWolfLSM: An accurate oil spill detection method based on level set method from synthetic aperture radar imagery," *Eur. J. Remote Sens.*, vol. 55, no. 1, pp. 181–198, 2022.

[6] Z. Tirandaz, G. Akbarizadeh, and H. Kaabi, "PolSAR image segmentation based on feature extraction and data compression using weighted neighborhood filter bank and hidden Markov random field-expectation maximization," *Measurement*, vol. 153, 2020, Art. no. 107432.

[7] F. M. Ghara, S. B. Shokouhi, and G. Akbarizadeh, "A new technique for segmentation of the oil spills from synthetic-aperture radar images using convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8834–8844, Oct. 2022.

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[9] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, Feb. 2020.

[10] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin, "Libra R-CNN: Towards balanced learning for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 821–830.

[11] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.

[13] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9626–9635.

[14] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2999–3007.

[15] Z. Hong et al., "Multi-scale ship detection from SAR and optical imagery via A more accurate YOLOv3," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 6083–6101, 2021.

[16] G. Jocher, "YOLOv5," 2020. [Online]. Available: https://github.com/ultralytics/yolov5

[17] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 7464–7475.

[18] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

[19] L. Bai, C. Yao, Z. Ye, D. Xue, X. Lin, and M. Hui, "A novel anchor-free detector using global context-guide feature balance pyramid and united attention for SAR ship detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, Mar. 2023, Art. no. 4003005.

[20] X. Zhang, S. Feng, C. Zhao, Z. Sun, S. Zhang, and K. Ji, "MGSFA-Net: Multiscale global scattering feature association network for SAR ship target recognition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 4611–4625, Jan. 2024.

[21] N. Yu, H. Ren, T. Deng, and X. Fan, "A lightweight radar ship detection framework with hybrid attentions," *Remote Sens.*, vol. 15, no. 11, 2023, Art. no. 2743.

[22] Z. Sun et al., "An anchor-free detection method for ship targets in high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7799–7816, Jul. 2021.

[23] Z. Sun, X. Leng, X. Zhang, B. Xiong, K. Ji, and G. Kuang, "Ship recognition for complex SAR images via dual-branch transformer fusion network," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, May 2024, Art. no. 4009905.

[24] B. Hu and H. Miao, "An improved deep neural network for small-ship detection in SAR imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 2596–2609, May 2024.

[25] M. Kang, K. Ji, X. Leng, and Z. Lin, "Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection," *Remote Sens.*, vol. 9, no. 8, 2017, Art. no. 860.

[26] X. Yang, X. Zhang, N. Wang, and X. Gao, "A robust one-stage detector for multiscale ship detection with complex background in massive SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Nov. 2021, Art. no. 5217712.

[27] Z. Wang, G. Hou, Z. Xin, G. Liao, P. Huang, and Y. Tai, "Detection of SAR image multiscale ship targets in complex inshore scenes based on improved YOLOv5," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 5804–5823, Feb. 2024.

[28] B. Chen, C. Yu, S. Zhao, and H. Song, "An anchor-free method based on transformers and adaptive features for arbitrarily oriented ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 2012–2028, Oct. 2023.

[29] M. Huang, T. Liu, and Y. Chen, "CViTF-Net: A convolutional and visual transformer fusion network for small ship target detection in synthetic aperture radar images," *Remote Sens.*, vol. 15, no. 18, 2023, Art. no. 4373.

[30] J. Fu, X. Sun, Z. Wang, and K. Fu, "An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1331–1344, Feb. 2021.

[31] Z. Lu et al., "A new deep neural network based on SwinT-FRM-ShipNet for SAR ship detection in complex near-shore and offshore environments," *Remote Sens.*, vol. 15, no. 24, 2023, Art. no. 5780.

[32] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 936–944.

[33] H. Tang et al., "A lightweight SAR image ship detection method based on improved convolution and YOLOv7," *Remote Sens.*, vol. 16, no. 3, 2024, Art. no. 486.

[34] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10778–10787.

[35] K. He et al., "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, 2015.

[36] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13708–13717.

[37] Z. Gevorgyan, "SIoU loss: More powerful learning for bounding box regression," 2022, *arXiv:2205.12740*.

[38] Z. Zheng et al., "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, Feb. 2020, vol. 34, pp. 12993–13000.

[39] Y. F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, "Focal and efficient IOU loss for accurate bounding box regression," *Neurocomputing*, vol. 506, pp. 146–157, 2022.

[40] J. Wang, C. Xu, W. Yang, and L. Yu, "A normalized Gaussian Wasserstein distance for tiny object detection," 2021, *arXiv:2110.13389*.

[41] K. Feng, L. Lun, X. Wang, and X. Cui, "LRTransDet: A real-time SAR ship-detection network with lightweight ViT and multi-scale feature fusion," *Remote Sens.*, vol. 15, no. 22, 2023, Art. no. 5309.

[42] Z. Xu et al., "DSF-Net: A dual feature shuffle guided multi-field fusion network for SAR small ship target detection," *Remote Sens.*, vol. 15, no. 18, 2023, Art. no. 4546.

[43] Z. Sun et al., "BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images," *Remote Sens.*, vol. 13, 2021, Art. no. 4209.

[44] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. SAR Big Data*, 2017, pp. 1–6.

[45] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020.

[46] T. Zhang et al., "LS-SSDD-v1.0: A deep learning dataset dedicated to small ship detection from large-scale Sentinel-1 SAR images," *Remote Sens.*, vol. 12, 2020, Art. no. 2997.

[47] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[48] Y. Deng, D. Guan, Y. Chen, W. Yuan, J. Ji, and M. Wei, "SAR-shipnet: SAR-ship detection neural network via bidirectional coordinate attention and multi-resolution feature fusion," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2022, pp. 3973–3977.

[49] C. Yao et al., "GFB-Net: A global context-guided feature balance network for arbitrary-oriented SAR ship detection," in *Proc. 7th Int. Conf. Image, Vis. Comput.*, 2022, pp. 166–171.

[50] Y. Yu, X. Yang, J. Li, and X. Gao, "A cascade rotated anchor-aided detector for ship detection in remote sensing Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2020, Art. no. 5600514.

[51] G. Tang et al., "PPA-Net: Pyramid pooling attention network for multi-scale ship detection in SAR images," *Remote Sens.*, vol. 15, no. 11, 2023, Art. no. 2855.

[52] L. Zhang, Y. Li, W. Zhao, X. Wang, G. Li, and Y. He, "Frequency-adaptive learning for SAR ship detection in clutter scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 5215514.

[53] Y. Zhou, H. Liu, F. Ma, Z. Pan, and F. Zhang, "A sidelobe-aware small ship detection network for synthetic aperture radar imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5205516.

[54] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6154–6162.

[55] Z. Li, C. Peng, G. Yu, X. Zhang, Y. Deng, and J. Sun, "Detnet: A backbone network for object detection," 2018, *arXiv:1804.06215*.

[56] C. Yao, P. Xie, L. Zhang, and Y. Fang, "ATSD: Anchor-free two-stage ship detection based on feature enhancement in SAR images," *Remote Sens.*, vol. 14, 2022, Art. no. 6058.

[57] T. Yue, Y. Zhang, J. Wang, Y. Xu, P. Liu, and C. Yu, "A precise oriented ship detector in SAR images based on dynamic rotated positive sample mining," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 10022–10035, Oct. 2023.

[58] H. Wang, S. Liu, Y. Lv, and S. Li, "Scattering information fusion network for oriented ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, Oct. 2023, Art. no. 4013105.

[59] K. Yang et al., "RFD: Detecting SAR coastal ship targets based on reducing feature decay," in *Proc. 4th Int. Conf. Big Data, Artif. Intell. Internet Things Eng.*, 2023, pp. 421–431.

[60] G. Jocher, "YOLOv8," 2023. Accessed: Feb. 14, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[61] N. Su, J. He, Y. Yan, C. Zhao, and X. Xing, "SII-net: Spatial information integration network for small target detection in SAR images," *Remote Sens.*, vol. 14, 2022, Art. no. 442.

[62] S. Wang et al., "YOLO-SD: Small ship detection in SAR images by multi-scale convolution and feature transformer module," *Remote Sens.*, vol. 14, no. 20, 2022, Art. no. 5268.

[63] Y. Feng et al., "A lightweight position-enhanced anchor-free algorithm for SAR ship detection," *Remote Sens.*, vol. 14, no. 8, 2022, Art. no. 1908.

[64] L. Bai, C. Yao, Z. Ye, D. Xue, X. Lin, and M. Hui, "Feature enhancement pyramid and shallow feature reconstruction network for SAR ship detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1042–1056, Jan. 2023.

[65] D. Quan et al., "Self-distillation feature learning network for optical and SAR image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2022, Art. no. 4706718.

[66] S. Jeong, Y. Kim, S. Kim, and K. Sohn, "Enriching SAR ship detection via multistage domain alignment," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Oct. 2022, Art. no. 4018905.

[67] H. Wang, D. Han, Z. Wu, J. Wang, Y. Fan, and Y. Zhou, "NAS-YOLOX: Ship detection based on improved YOLOX for SAR imagery," in *Proc. IEEE 10th Int. Conf. Cyber Secur. Cloud Comput. /IEEE 9th Int. Conf. Edge Comput. Scalable Cloud*, 2023, pp. 126–131.

[68] Y. Tang, S. Wang, J. Wei, Y. Zhao, and J. Lin, "Salient feature pyramid network for ship detection in SAR images," *IEEE Sensors J.*, vol. 24, no. 3, pp. 3036–3045, Feb. 2024.

[69] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9759–9768.

[70] G. Yan, Z. Chen, Y. Wang, Y. Cai, and S. Shuai, "LssDet: A lightweight deep learning detector for SAR ship detection in high-resolution SAR images," *Remote Sens.*, vol. 14, no. 20, 2022, Art. no. 5148.

Liming Zhou (Member, IEEE) received the Ph.D. degree in computer science and technology from the State Key Laboratory of Network and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China, in 2015.

He has been an Associate Professor with the School of Computer and Information Engineering, Henan University, since 2015. His research interests include deep learning, artificial intelligence, and information security.

Ziye Wan received the bachelor's degree in software engineering from the Nanyang Institute of Technology, Henan, China, in 2022. She is currently working toward the master's degree in computer technology with the School of Computer and Information Engineering, Henan University, Kaifeng, China.

Her research interests include computer vision and SAR image processing, especially target detection and recognition in SAR images.

Shuai Zhao received the B.E. degree in computer science and technology from the Henan University of Technology, Kaifeng, China, in 2021. He is currently working toward the M.S. degree in computer technology with the Key Laboratory of Big Data Analysis and Processing, Henan University, Kaifeng.

His research interests include deep learning, UAV object detection, and remote sensing.

Hongyu Han received the PhD. degree in computer science and technology from Harbin Engineering University, Harbin, China, in 2020.

He has been a Lecturer with the School of Computer and Information Engineering, Henan University, Kaifeng, China, since 2020. His research interests include sentiment analysis and deep learning.

Yang Liu (Member, IEEE) received the B.S. degree in industrial analysis and electronic measurement from the Changchun University of Science and Technology, Changchun, China, in 1996, and the M.S. degree in applied mathematics and the Ph.D. degree in remote sensing information science and technology from the Henan University, Kaifeng, China, in 2009 and 2016, respectively.

He is currently a Professor and a Ph.D. supervisor with the College of Computer Science and Information Engineering, Henan University, Kaifeng, China. He is the Director of Principle Investigator of Brain-Inspired Intelligence Science and Technology Innovative Team. His research interests include science theory and technology of brain and mind-inspired computing, and complex system engineering and application of spatiotemporal Big Data system.