









# SA<sup>2</sup>Net: Ship Augmented Attention Network for Ship Recognition in SAR Images

Yuanzhe Shang , Graduate Student Member, IEEE, Wei Pu , Member, IEEE, Danling Liao , Ji Yang, Congwen Wu , Graduate Student Member, IEEE, Yulin Huang , Senior Member, IEEE, Yin Zhang , Member, IEEE, Junjie Wu , Member, IEEE, Jianyu Yang , Member, IEEE, and Jianqi Wu

**Abstract**—Maritime surveillance is extensively concerned by worldwide authorities, in which ship recognition in synthetic aperture radar (SAR) images is a significant and fundamental component. Though some development has been achieved in the SAR ship recognition task, two areas remain inadequately explored, which are the comprehensive utilization of multiscale features and the deployment of the prior knowledge of the ship shape. In this article, a novel ship augmented attention network (SA<sup>2</sup>Net) for ship recognition is proposed, which comprehensively utilizes the multiscale features and integrates the ship shape prior to the end-to-end network. On one hand, due to the unequal effects of different scales, a scale attention module is proposed to adaptively select and assign weights to desired feature scales while disregarding irrelevant scales. Moreover, a feature weaving module (FWM) is constructed to merge semantic and detailed features produced by the high-to-low backbone, enriching representations across all scales of ship targets. On the other hand, in order to incorporate the prior knowledge of the ship shape into the network, we develop a feature augmentation module (FAM) to further boost the ship recognition accuracy. This module can provide rectangular receptive fields that align with the shape of ships, wherein a limitation encountered with traditional square convolutions. Comprehensive experiments on representative three- and six-category OpenSAR-Ship tasks and seven-category FUSAR-Ship tasks show that our SA<sup>2</sup>Net demonstrates superior performance when compared to the current state-of-the-art methods.

**Index Terms**—Synthetic aperture radar (SAR), ship recognition, convolutional neural networks (CNNs), shape prior knowledge, feature augmented module, scale attention module (SAM).

## I. INTRODUCTION

NOWADAYS, maritime surveillance is extensively concerned by worldwide authorities. As the basis of

maritime surveillance tasks, ship monitoring plays a key role in both military and civil activities, such as trade management, marine traffic, transportation monitoring, and national maritime safeguarding [1], etc. Automatic identification system (AIS) and vessel traffic service are conventional techniques for ship monitoring. However, neither of these techniques is enough to achieve general purpose vessel monitoring with the demanded independence, temporal coverage, and spatial coverage [2]. Synthetic aperture radar (SAR), with all-day monitoring capability, can monitor large areas independently of meteorological conditions [3], which stands out as an effective substitution and has been extensively studied for ship recognition in recent decades.

In the past few decades, many hand-crafted feature methods have been introduced for SAR ship recognition, such as scattering statistics features, texture features, geometric features, moment features, scale-invariant features [4], and HOG features [5]. To increase the recognition accuracy, some machine learning methods are jointly used, including  $K$ -nearest neighbor (KNN) [6], support vector machine (SVM) [7], and random forest (RF) [8].

Although these traditional recognition algorithms have produced good results, they are always based on hand-crafted features. These features may be suitable for specific data but lack adaptability. In contrast, deep learning diverges from conventional approaches as it leverages neural networks to autonomously extract features through end-to-end learning. This data-driven paradigm has achieved remarkable success across various domains, particularly in the realm of image recognition [9]. In recent years, the convolutional neural network (CNN)-based methods tend to be the mainstream for SAR ship recognition [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25]. To improve recognition accuracy, many studies have been put forward and achieved good SAR ship performance in various aspects. To solve the issue of class imbalance, Li et al. [12] proposed a dense residual network (DRNet) combining upsampling data augmentation and ratio batching. Shao et al. [13] proposed a balanced batch-based sampling method to avoid learning imbalance during training. Zhang et al. [14] presented a method for training CNN that integrates deep metric learning (DML) with progressively balanced sampling. Raj et al. [15] proposed a one-shot learning-based deep learning model. To address the problem of small training dataset due to few available data, Lu et al. [16] established a CNN with data augmentation. Yuanyuan et al. [17]

Manuscript received 30 June 2023; revised 17 August 2023; accepted 6 September 2023. Date of publication 20 September 2023; date of current version 14 November 2023. This work was supported by the National Natural Science Foundation of China under Grant 61901091 and Grant 61901090. (Corresponding author: Wei Pu.)

Yuanzhe Shang, Wei Pu, Danling Liao, Congwen Wu, Yulin Huang, Yin Zhang, Junjie Wu, and Jianyu Yang are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 610056, China (e-mail: shangyuanzhe@std.uestc.edu.cn; pwuestc@163.com; liaodanling77@gmail.com; 202011012311@std.uestc.edu.cn; yulinhuang@uestc.edu.cn; yinzhang@uestc.edu.cn; junjie\_wu@uestc.edu.cn; yangjianyu@uestc.edu.cn).

Ji Yang is with the Unit 31308 of the PLA, China (e-mail: yangji@163.com).

Jianqi Wu is with the East China Research Institute of Electronic Engineering, Hefei 230031, China, and also with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 610056, China (e-mail: wujianqi38@163.com).

Digital Object Identifier 10.1109/JSTARS.2023.3317489

conducted some small sample SAR ship recognition research based on transfer learning. To tackle with the challenge of large intraclass variation and small interclass separation of ships, Xu and Lang [18] and He et al. [19] used a DML scheme to expand the distance between different classes. To resolve the weak robustness of individual models in high risk scenarios, Zheng et al. [10] introduced an automated approach for ensemble modeling of heterogeneous deep convolutional neural networks (DCNNs), employing a two-stage filtration process. This self-configuring algorithm dynamically determines the optimal combination of base classifiers by automatically identifying the suitable types and quantities.

Although these approaches have achieved notable success, the majority of the aforementioned works tend to focus on iterative modifications of network structures, training trick optimizations, loss function adjustments, and so on, rather than design a task-specific network from the characteristics of SAR image and empirical knowledge of the ship. In recent years, an expanding body of scholars has taken notice of this phenomenon. Huang introduced a new Deep SAR-Net [20] that considers complex-valued SAR images to learn the spatial texture information and backscattering patterns of ships. Zhang et al. [21] fused HOG features into CNNs and proposed four mechanisms to ensure superior recognition accuracy. Zeng et al. [22] designed a hybrid channel feature loss that jointly utilizes the information contained in the polarized channels (VV and VH). He et al. [23] established a group bilinear pooling and a MPFL loss to fully exploit the dual-polarized SAR images for promising fine-grained ship recognition. Xiong et al. [24] developed a miniature hourglass region extraction network dedicated to dual-channel feature fusion. Zhang and Zhang [25] designed a SE-LPN-DPFF to perform dual-polarization feature fusion and balance each polarization feature's contribution. Although the above methods achieve good results, leaving room for further improvement in the performance of the network. First of all, the comprehensive utilization of multiscale features holds paramount importance in enhancing SAR ship recognition, an area that remains inadequately investigated. Ships usually appear with diverse sizes, which is challenging to achieve state-of-the-art (SOTA) recognition result by using a single scale features of CNN [26]. This is why maximizing the use of multiscale features is crucial. Xu [27] and Zhang [21] have made some preliminary explorations to deal with this issue. However, they simply flattened the multiscale features without fusing them, which resulted in a failure to provide sufficient features at all scales. In addition, these previous works aggregated multiscale features of CNN to recognize ships using unified weights, e.g., a simple summation, which ignores the different importance of different scales. Second, the ship class in SAR imagery has special shape prior characteristics. To the best of our knowledge, no work has yet integrated the ship shape prior into an end-to-end network to perform SAR ship recognition.

Based on the analysis above, we propose a task-specific ship augmented attention network (SA<sup>2</sup>Net) for comprehensively utilizing the multiscale features and integrating the ship shape prior into an end-to-end network. Among SA<sup>2</sup>Net, the feature weaving module (FWM) is designed to generate rich and reliable representations at all scales. The scale attention module (SAM)

has been constructed to select and assign weights to relevant feature scales while disregarding irrelevant scales. The feature augmentation module (FAM) has been designed to enhance ship features, which incorporate the priory knowledge of the ship shape. Comprehensive experiments demonstrate the superiority of our SA<sup>2</sup>Net compared with several SOTA methods. In contrast to previous works, the novelties and contributions can be summarized as follows:

- 1) We proposed a SA<sup>2</sup>Net that jointly applies SAM and FWM to fully exploit the multiscale features of ship targets. The shallow scale features contain more detailed information while deep scale features contained more semantic information, which is unequally effective for recognition. Instead of simply combining different-scale features, the proposed SAM is developed to control information flow of different scales using a leaned weight vector, and then adaptively selects and assigns weights to desired feature scales while disregarding irrelevant scales. The proposed FWM aggregates semantic and detailed features of different scales by integrating high-level semantic information and low-level detailed information through a similar weaving process, resulting in rich representations at all scales.
- 2) FAM is first proposed in this article to leverage empirical knowledge regarding ships, which commonly exhibit elongated and narrow characteristics. In contrast to prior approaches that employ square convolution kernels for ship feature extraction, the FAM introduces rectangular convolutions. This design can provide rectangular receptive fields that align with the shape of ships, a limitation encountered with traditional square convolutions.
- 3) We conduct extensive experiments on benchmark OpenSARShip [28] and FUSAR-Ship [29]. The results show that SA<sup>2</sup>Net exceeds existing methods, including traditional feature-based methods, classic object recognition CNNs, and novel task-specific SAR ship recognition CNNs. The experimental results demonstrate the effectiveness of our method.

The rest of this article is organized as follows. In Section II, we present the details of our proposed SA<sup>2</sup>Net. In Section III, implementation details are reported, and extensive experimental results are provided. Section IV presents the conclusion.

## II. METHODOLOGY

### A. Network Structure

The overall framework of our ship augmented attention network (SA<sup>2</sup>Net) for SAR ship recognition is illustrated in Fig. 1. To achieve ship recognition with diverse sizes, we propose SAM and FWM, which comprehensively utilize multi-scale features. Besides, FAM is designed to enhance ship features by incorporating the priory knowledge of the ship shape. In SA<sup>2</sup>Net, the pretrained ResNet-50 [30] has been leveraged as the backbone for its enormous performance in feature extraction. Given a SAR ship image, the FWM integrates high-level semantic information and low-level detailed information through repeatedly fusing the representations produced by the high-to-low backbone to obtain better representations at all scales. Besides, in view of

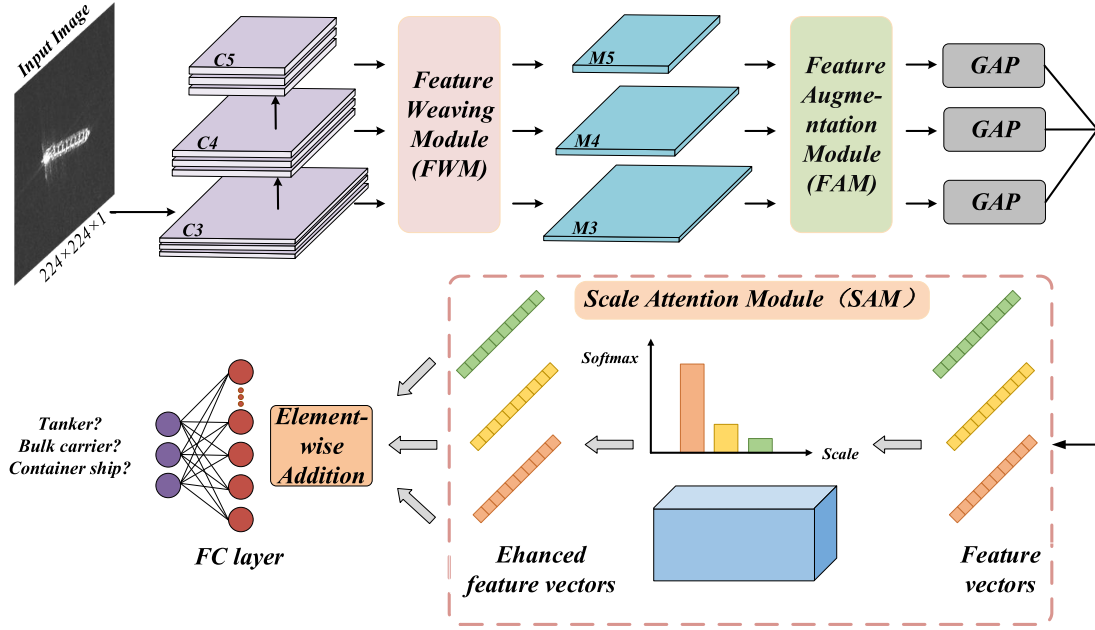


Fig. 1. Overall architecture of ship augmented attention network (SA<sup>2</sup>Net). The architecture consists of a backbone for feature extraction and three modules to refine the extracted features for final precise recognition. ResNet-50 is adopted as the backbone due to its impressive performance. The three modules are FWM, FAM, and SAM, respectively.

the distinctive prior characteristics pertaining to the shape of ship class, FAM has been devised to augment ship features by integrating the prior knowledge of ship shape. At last, to select the effective feature scales for the final recognition, SAM utilizes the relevance scores to select and assign weights to relevant feature scales while disregarding irrelevant scales. This network guarantees more accurate recognition for SAR ship to squeeze

out the benefits of the multiscale features and integrated the ship shape prior. Details are provided in the Algorithm flow below.

### B. Feature Weaving Module

Accomplishing robust SAR ship recognition across diverse sizes proves challenging when using a single-scale feature representation from CNN. To address this issue effectively, leveraging the multiscale features obtained from intermediate layers of the CNN presents a viable solution. In CNN, the receptive field of layers become larger as the layer becomes deeper. The feature maps obtained from the lower layer focus on detailed information while the feature maps obtained from the deeper layer focus on semantic information. Inspired by HRNet [31], with a focus on sufficiently making use of multiscale features to obtain rich representations at all scales, FWM is proposed.

As shown in Fig. 2, FWM fully mines and combines the feature maps of different scales through a feature fusion mechanism called feature weaving. It generates reliable rich feature representations through repeatedly fusing the representations produced by the high-to-low backbone convolutional layers. The details of FWM are presented as follows.

ResNet-50 is utilized as the backbone. The output of the last layer of different residual blocks in Conv3, Conv4, and Conv5 levels of ResNet-50 is indicated as  $C_i$  ( $i = 3, 4, 5$ ). The specific pattern of generating each  $M_i$  layer corresponding to  $C_i$  layer is shown in Fig. 2(a), with  $i \in \{3, 4, 5\}$ . For higher level, same level and lower level features, the features are processed by bilinear interpolation upsampling,  $1 \times 1$  convolution, and convolution layer downsampling, respectively. In this step, the channel dimension is uniformly adjusted to 256. Finally, different layers are consolidated with elementwise summation.

---

#### Algorithm 1: SA<sup>2</sup>Net for SAR Ship Recognition.

---

**Given:** A gray SAR ship image  $\mathbf{X} \in \mathbb{R}^{224 \times 224 \times 1}$ ;

**Output:** Recognition result  $\text{Out}_{SA^2Net} \in \mathbb{R}^{n_{class} \times 1}$ .

- 1: Image  $\mathbf{X}$  loading, do preprocessing on  $\mathbf{X}$  to get  $\mathbf{X} \in \mathbb{R}^{224 \times 224 \times 3}$ , hyperparameters setting;
  - 2: Extract hierarchical feature maps  $C_3, C_4, C_5$  with backbone ResNet-50, build enhanced feature maps  $M_3, M_4, M_5$  with  $C_3, C_4, C_5$ .
  - 3: **Stage1 :**
  - 4: Input  $M_1$  with  $l$  as  $\{3, 4, 5\}$ , obtain  $M_{1h}, M_{1v}, M_{1l}, M_{1r}, M_{1s}$  by five parallel branches with distinct convolution kernels;
  - 5: Concatenate  $M_{1h}, M_{1v}, M_{1l}, M_{1r}, M_{1s}$  to get  $M'_1$ ;
  - 6: Perform composite function  $comp(\cdot)$  on  $M'_1$  to get output  $A_1$ .
  - 7: **Stage2 :**
  - 8: Generate feature vector of each scale  $f_i$  with  $A_1$  by GAP;
  - 9: Concatenate  $f_i$  and obtain the learned scale relevance scores  $w$  with FC and Softmax function;
  - 10: Gain final enhanced features  $\hat{f}$  with  $f_i$  and  $w$ ;
  - 11: Predict the SAR ship recognition scores  
 $\text{Out}_{SA^2Net} = \text{softmax}(FC(\hat{f}))$ .
-

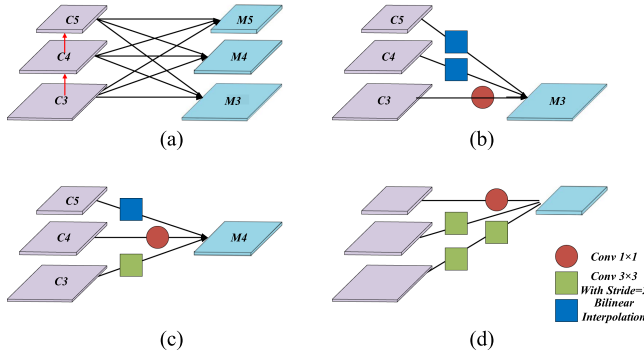


Fig. 2. (a) Overall depiction of FWM is presented. Subsequently, (b)–(d) elaborate on the specific details of generating  $M_3$ ,  $M_4$ , and  $M_5$ , respectively. Moreover, it is worth noting that the pathway represented by the blue square signifies upsampling utilizing bilinear interpolation, the pathway indicated by the green square signifies downsampling using one or two  $3 \times 3$  convolutions, and the pathway denoted by the red circle signifies aligning channel dimensions through  $1 \times 1$  convolutions.

The operations of FWM are computed as follows:

$$\begin{cases} M_3 = \text{Conv}U(C_5) + \text{Conv}U(C_4) + \text{Conv}(C_3) \\ M_4 = \text{Conv}U(C_5) + \text{Conv}(C_4) + \text{Conv}D(C_3) \\ M_5 = \text{Conv}(C_5) + \text{Conv}D(C_4) + \text{Conv}D(C_3) \end{cases} \quad (1)$$

where  $\text{Conv}(\cdot)$  denotes  $1 \times 1$  convolution to align the channel dimensions,  $\text{Conv}U(\cdot)$  is the bilinear interpolation upsampling, and  $\text{Conv}D(\cdot)$  indicates  $3 \times 3$  convolution with stride 2 down-sampling.

### C. Feature Augmentation Module

As shown in Fig. 4, the ship class in SAR images exhibits a prominent geometric characteristic, a large aspect ratio. In addition, in contrast to natural images captured from a horizontal view, SAR images are acquired from a top-down perspective. This leads to objects appearing at arbitrary orientations. Traditional convolution operations commonly employ square kernels such as  $3 \times 3$ ,  $5 \times 5$ , as they are well-suited for capturing block-like structures like vehicles and buildings. However, the unique shape characteristics of the ship class, which exhibits a strip-like structure and arbitrary orientations of ships pose challenges for effective extraction using traditional convolution kernels. Therefore, rectangular convolutions with different directions are introduced, which can provide rectangular receptive fields to match the shape and arbitrary orientations of ships. We develop the FAM to replace the traditional square convolution with a combination of a square convolution and four rectangular convolutions implemented through separate branches. The original features are preserved by the square convolution branch, while horizontal convolution, vertical convolution, left diagonal convolution, and right diagonal convolution refine the details by providing rectangular receptive fields.

Fig. 3 demonstrates the structure of FAM. The parallel organization of five branches with different kernel sizes are constructed. Assuming the convolutional layers take a  $C$ -channel feature map as input. As illustrated in Fig. 3, FAM incorporates rectangular convolutions in four distinct orientations: horizontal, vertical, left diagonal, and right diagonal. Concretely, for the replacement of a  $3 \times 3$  square kernel  $S \in \mathbb{R}^{3 \times 3 \times C}$ , FAM comprises five parallel branches including four rectangular convolution kernels and a square convolution kernel. The horizontal kernel  $S_1 \in \mathbb{R}^{1 \times 3 \times C}$ , vertical kernel  $S_2 \in \mathbb{R}^{3 \times 1 \times C}$ , left diagonal kernel  $S_3 \in \mathbb{R}^{[left\,diag] \times C}$ , and right diagonal kernel  $S_4 \in \mathbb{R}^{[right\,diag] \times C}$  are rectangular kernels that align with the shape of ships. Let  $M_l \in \mathbb{R}^{H \times W \times C}$  be the input of FAM, with  $l$  as  $\{3, 4, 5\}$ .  $X$  is fed into five juxtaposed paths. Then, five output feature maps  $M_{lh}, M_{lv}, M_{ll}, M_{lr}, M_{ls} \in \mathbb{R}^{H \times W \times C}$  are obtained. Then, the concatenate operator of five feature maps is performed to obtain  $M'_l \in \mathbb{R}^{H \times W \times 5C}$ . This progress can be described as

$$\begin{aligned} M'_l &= \text{cat}(M_{lh}, M_{lv}, M_{ll}, M_{lr}, M_{ls}) \\ &= \text{cat}(M_l * S_1, M_l * S_2, M_l * S_3, M_l * S_4, M_l * S) \end{aligned} \quad (2)$$

where  $*$  indicates the convolution operation, and  $\text{cat}$  is the concatenate operator.

We define  $\text{comp}(\cdot)$  as a composite function to get the final output of FAM.  $\text{comp}(\cdot)$  is consist of three consecutive operations: batch normalization (BN), a rectified linear unit (ReLU), and a  $3 \times 3$  convolution (conv). As for  $M_l$ , the corresponding output of FAM can be denoted as

$$A_l = \text{comp}(M'_l) \quad (3)$$

The reason why we employ  $\text{cat}(\cdot)$  operator and  $\text{comp}(\cdot)$  function, rather than simply summation of the output of five branches is motivated by DenseNet [32]. First of all, when  $M_{lh}, M_{lv}, M_{ll}, M_{lr}, M_{ls}$  are combined by summation, which may impede the information flow in the network [32], leading to ship feature extraction insufficiency. Second, the statistical characteristics of the five juxtaposed branches differ from each other, e.g., there may be large differences in the mean and variance of the pixels in each branch. So it is important to perform the batch normalization (BN) [33] layer after concatenation, rather than before. This is the ingenuity of  $\text{comp}(\cdot)$  function. Applying the BN layer before the concatenation operator may result in an internal covariate shift in new feature maps, reducing the generalization capability of the network. Finally, the outputs  $A_3, A_4, A_5$  are fed into SAM for the next step.

### D. Scale Attention Module

Most previous works aggregate multiscale features of CNN to recognize ships using unified weights, e.g., a simple summation, which ignore the unequal effectiveness of different scales. To address this problem, we propose SAM, as shown in Fig. 5. This module weights desired feature scales according to the relevance scores between each scale and final recognition probabilities, selecting the effective feature scales while excluding irrelevant scales.

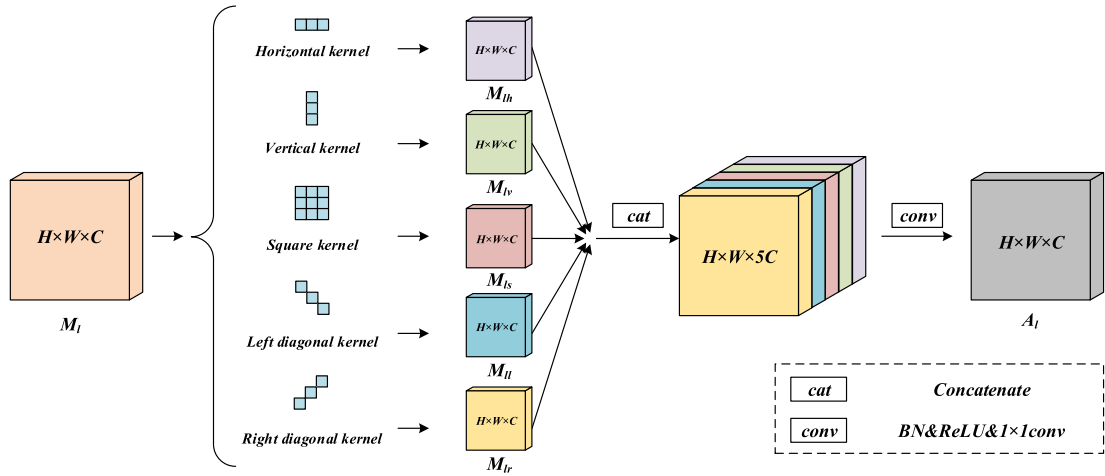


Fig. 3. Illustration of the FAM. In this figure, the FAM contains five parallel layers with kernel sizes.

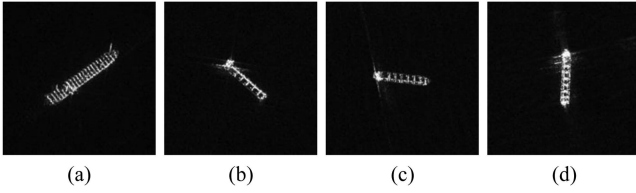


Fig. 4. SAR ships with four directions. (a) Right diagonal. (b) Left diagonal. (c) Horizontal. (d) Vertical.

Denote  $A_3, A_4, A_5$  as a set of multiscales feature maps. We embed the three feature maps of each scale into a vector  $f_i \in \mathbb{R}^d$  respectively for global information by global average pooling (GAP), where  $d = 256$ , and cascade these feature vectors by the following process:

$$\begin{aligned} f &= \text{cat}(f_1, f_2, f_3) \\ &= \text{cat}(\text{GAP}(A_3), \text{GAP}(A_4), \text{GAP}(A_5)) \end{aligned} \quad (4)$$

where  $f \in \mathbb{R}^{3d}$  is the feature vector after concatenating.

To make the module automatically select the desired feature scales to obtain preferable recognition scores, the designed SAM can generate a learned scale relevance scores. The weight  $p = [w_1, w_2, w_3] \in \mathbb{R}^3$  of selecting feature scales for each specific recognition score can be described as

$$w_i = \text{Softmax}(w_a^T f) \quad (5)$$

where  $w_a \in \mathbb{R}^{3d \times 3}$  is the attention weight, which combines features of distinct scales into a weight vector with three dimension. Based on the above weight predictor values  $w_i$ , the feature scales  $f_i$  can be weighted and summed to gain the final enhanced features  $\tilde{f} \in \mathbb{R}^d$  for preferable SAR ship recognition results:

$$\tilde{f} = (w_1 \otimes f_1) \oplus (w_2 \otimes f_2) \oplus (w_3 \otimes f_3). \quad (6)$$

Subsequently, a fully connected (FC) layer and a softmax function are needed for achieving the final recognition.

### III. EXPERIMENTS AND RESULTS

In this section, we will perform extensive experiments to verify the effectiveness of the proposed method on benchmark dataset OpenSARShip and FUSAR-Ship. First, we describe the dataset and give the dataset settings. Then, we present the implementation details, including image preprocessing, parameter settings, evaluation metrics, loss function, and backbone. Next, the experimental results are demonstrated for OpenSARShip

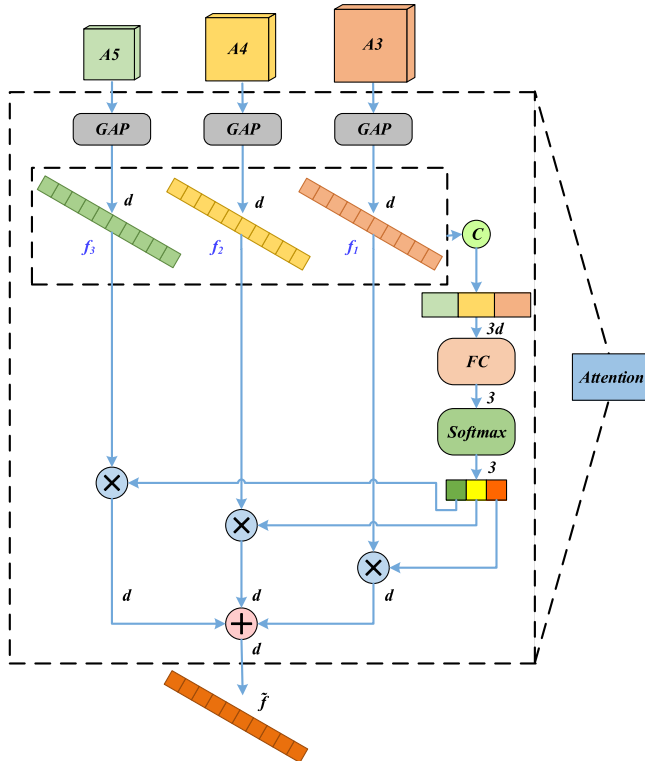


Fig. 5. Illustration of the SAM. It adaptively selects and assigns weights to desired feature scales while disregarding irrelevant scales. C denotes the concatenate operator.  $\otimes$  is channelwise product and  $\oplus$  is elementwise sum.

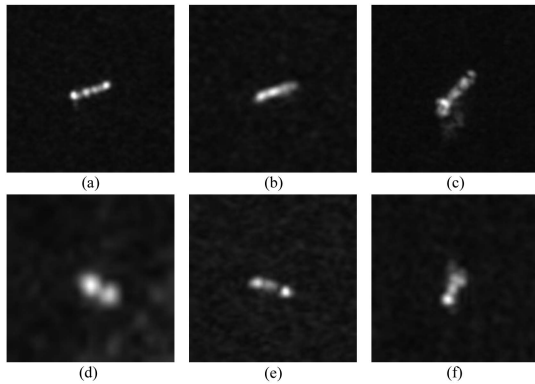


Fig. 6. SAR ship samples in OpenSARShip. (a) Bulk carrier. (b) Cargo. (c) Container ship. (d) Fishing. (e) General cargo. (f) Tanker.

TABLE I  
TRAINING-TESTING DIVISION OF THE THREE-CATEGORY DATASET IN  
OPENSARSHIP

Type	Train	Test	All
Bulk carrier	300	375	675
Container ship	300	711	1011
Tanker	300	254	554

under a three-category recognition task, a six-category recognition task, and a FUSAR-Ship under seven-category task. In addition, we perform a comprehensive comparison between the proposed method and some SOTA methods, encompassing traditional classifiers, classic CNN methods, and modern CNN methods designed for SAR ship recognition. Ablation studies and discussion are conducted at last.

#### A. Dataset

1) *OpenSARShip*: In our study, we utilize OpenSARShip, a benchmark dataset that comes from the Sentinel-1 satellite. OpenSARShip possesses five critical properties, namely specificity, large-scale coverage, diversity, reliability, and public availability, which collectively contribute to its significant value in practical applications. The ship labels in the OpenSARShip dataset are expertly assigned through a semiautomated process, with support from AIS, ensuring their accuracy. The dataset utilized in our experiment is ground range detected (GRD) products captured by the Sentinel-1 IW mode. It possesses a resolution of  $20 \text{ m} \times 22 \text{ m}$  and a pixel size of  $10.0 \text{ m} \times 10.0 \text{ m}$  in both the azimuth and range directions [28]. Based on OpenSARShip, two recognition datasets are conducted. Fig. 6 shows some SAR ship samples in OpenSARShip.

a) *Three-Category*: Container ships, tankers, and bulk carriers are chosen to establish the representative dataset. These three classes of ships are the most common and representative ships occupying 80% of the international shipping market [28]. The number of each class of ship is uneven in OpenSARShip. To avoid the effect of class imbalance, the number of training samples in each class is equal. Table I shows the training-testing sets of the three-category dataset.

TABLE II  
TRAINING-TESTING DIVISION OF THE SIX-CATEGORY DATASET IN  
OPENSARSHIP

Type	Train	Test	All
Bulk carrier	200	475	675
Container ship	200	811	1011
Tanker	200	354	554
Cargo	200	557	757
Fishing	200	121	321
General cargo	200	165	365

TABLE III  
TRAINING-TEST DIVISION OF THE FUSAR-SHIP

Dataset	Category	Train	Test	All
FUSAR-Ship	Bulk carrier	1150	494	1644
	Container ship	1219	523	1742
	Fishing	1101	473	1574
	Tanker	1215	521	1736
	General cargo	1205	517	1722
	Other cargo	1214	521	1735
	Other	1211	520	1731

b) *Six-Category*: On the basis of the three category, another three classes, cargo ship, fishing, and general cargo are selected to organize one more challenging six-category recognition experiment. Based on the detailed ship classes provided by the Maritime Traffic AIS information, six ship classes are specifically selected for analysis as their sample numbers exceed 200. Furthermore, categories with insufficient samples in the raw OpenSARShip dataset are excluded to ensure a more reasonable experimental setup. Table II shows the training-testing sets of the six-category dataset.

2) *FUSAR-Ship*: Another benchmark dataset FUSAR-Ship is introduced to further confirm the effectiveness of SA<sup>2</sup>Net. The high-resolution dataset FUSAR-Ship originates from China's Gaofen-3 (GF-3) satellite, the country's maiden civil C-band fully polarimetric spaceborne SAR. The GF-3 SAR images possess an azimuth resolution of 1.124 m and a slant range resolution ranging from 1.700 to 1.754 m. The FUSAR-Ship dataset is assembled through an automatic SAR-AIS matchup procedure encompassing over 100 GF-3 scenes, containing over 5000 ship image chips integrated with AIS information. In this article, FUSAR-Ship consists of seven main categories, namely bulk carriers, container ships, fishings, tankers, general cargo ships, other cargo ships, and others. Table III shows the ship sample numbers of each category in FUSAR-Ship. Fig. 7 presents several SAR ship samples from the FUSAR-Ship dataset.

#### B. Implementation Details

All the experiments are implemented on a personal computer (PC) with NVIDIA GeForce RTX 2060 VENTUS (12G) GPU and 24G RAM. The software development process is carried out within the Python programming language environment, utilizing

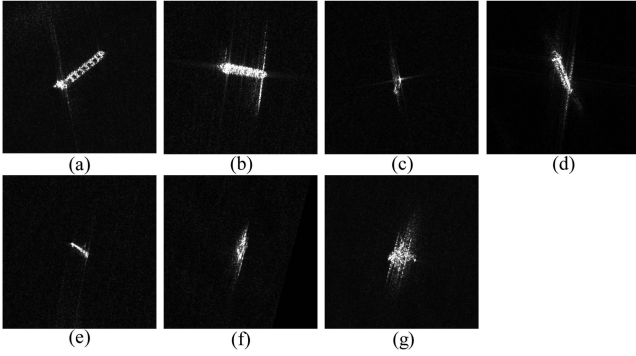


Fig. 7. SAR ship samples in FUSAR-Ship. (a) Bulk carrier. (b) Container ship. (c) Fishing. (d) Tanker. (e) General cargo. (f) Other cargo. (g) Other.

the open-source PyTorch machine learning library. For training and inference acceleration, CUDA10.1 is employed.

1) *Image Preprocessing*: The backbone we utilized is pretrained ResNet-50. The pretrained weights of ResNet-50 are based on natural images, which are three-channel images. However, the SAR images we exploit in our manuscript are single-channel. In order to utilize the pretrained ResNet-50, we replicate the grayscale value across all three channels. In other words, the values in all three channels are the same at each pixel position since our grayscale image has only one channel. Such conversion can also be found in other classic work [34].

2) *Parameter Setting*: These experiments are trained under the same parameters. The size of the input images in OpenSARShip are unified to  $224 \times 224$ . Using stochastic gradient descent (SGD) optimizer with the weight decay parameter 0.001 and the momentum parameter 0.9, the proposed network is trained by 10 000 iterations. The batch size is set to 16 due to the limited GPU memory. To alleviate the adverse impact of vanishing training gradients, we assigned a relatively low learning rate of 0.0001, which is appropriate for our method.

3) *Loss Function*: The cross entropy(CE) loss is served as the loss function

$$L = -\frac{1}{N} \sum_{m=1}^N y'_m \log(y_m) \quad (7)$$

where the  $m$ th sample recognition result is denoted as  $y_m$ , the  $m$ th sample ground truth is denoted as  $y'_m$ , and the total number of training samples is denoted as  $N$ .

4) *Evaluation Metrics*: Similar to most scholars, accuracy (%) is used as the core evaluation criteria to measure recognition performance and confirm effectiveness of the proposed modules. For comprehensive evaluations of SAR ship recognition results, four additional performance metrics are employed in the experiments, including: 1) F1; 2) precision; 3) recall; and 4) confusion matrix. The definition of these metrics are as follows.

Accuracy is defined by

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (8)$$

where TP denotes true positives, TN denotes true negatives, FP denotes false positives, and FN denotes false negatives.

TABLE IV  
RECOGNITION PERFORMANCE OF DIFFERENT BACKBONES ON THREE-CATEGORY OPENSARSHIP AND FUSAR-SHIP

Backbone	Three-Category OpenSARShip(%)	FUSAR-Ship(%)
ResNet-18	81.64	87.11
ResNet-34	82.54	87.47
ResNet-50	<b>82.91</b>	<b>88.28</b>
ResNet-101	82.39	87.98

The bold values mean the recognition performance of ResNet-50 shows the optimal accuracy not only on OpenSARShip, but FUSAR-Ship as well.

In other words, the numerator denotes the number of correctly recognized ship samples, the denominator denotes the number of all test ship samples.

Recall [21] is defined by

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (9)$$

Precision [21] is defined by

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (10)$$

F1 [21] is defined by

$$\text{F1} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (11)$$

Furthermore, in order to evaluate the ship recognition performance in a more specific manner, a confusion matrix is adopted as a classwise measure to evaluate the recognition ability of each category. This evaluation method has been commonly utilized in previous studies on SAR ship recognition as well.

5) *Backbone*: Generally speaking, the backbone will directly influence the recognition performance. In the context of SAR ship detection task, ResNet serves as the most favored backbone in some popular and substantial works [35], [36], [37]. Thus, we also apply it to SAR ship recognition task empirically. To choose the most suitable backbone for our task, we conduct the experiments of ResNet-18, ResNet-34, ResNet-50, and ResNet-101 as the backbone networks of SA<sup>2</sup>Net. The experimental results are presented in Table IV. From the experimental results, we found that the recognition performance of ResNet-50 shows the optimal accuracy not only on OpenSARShip, but FUSAR-Ship as well. The primary reason for this observation is that the features learned by ResNet-18 and ResNet-34 are insufficient, and ResNet-101 is prone overfitting due to its deep network. Therefore, we choose the pretrained ResNet-50 as the backbone in the subsequent experiments.

### C. Recognition Results

1) *Quantitative Evaluation*: Table V shows the evaluation metrics of SA<sup>2</sup>Net on the three-category OpenSARShip task, six-category in classical algorithms task, and seven-category FUSAR-Ship task. From Table V, the SAR ship recognition

TABLE V  
EVALUATION METRICS OF SA<sup>2</sup>NET ON OPENSARSHIP AND FUSAR-SHIP TASKS

Task	$r(\%)$	$p(\%)$	$f1(\%)$	$Acc(\%)$
Three-Category OpenSARShip	78.79	83.23	80.95	82.91
Six-Category OpenSARShip	59.54	59.28	59.41	61.10
Seven-Category FUSAR-Ship	88.24	88.52	88.38	88.28

accuracy on three-category OpenSARShip is 82.91%, on six-category OpenSARShip is 61.10%, and that on the seven-category FUSAR-Ship task is 88.28%. As for OpenSARShip, the latter performance is significantly lower compared to the former, primarily due to the inherently higher complexity of the six-category recognition task as compared to the three-category task. In addition, the number of training samples available for the six-category task is smaller than that of the three-category task, further amplifying the recognition challenge associated with the six-category task. Due to FUSAR-Ship has a better resolution with more ship detailed representations can be learned, the recognition accuracy can reach 88.28%.

2) *Confusion Matrix*: The recognition performance under three-category and six-category tasks for each ship class in confusion matrix forms are offered by Tables VI, VII, and VIII, respectively. Most diagonal values are higher than others in the same line from both tables, which indicate that most ships can be recognized correctly. A notable observation from the three tables is that most diagonal values predominantly surpass the corresponding values in the same row, implying a high rate of correct recognition for most ships, but there are still some classes which are easily confused. From Table VI, the bulk carrier is recognized as a container ship mistakenly. This phenomenon may arise due to the outline of the ship is too vague, which acts as a strong scattering point, thereby limiting its capacity to facilitate effective recognition. From Table VII, the general cargo is recognized as a cargo mistakenly. In fact, their class differences are rather small, and the general cargo can be regarded as a special cargo. From Table VIII, the primary source of class prediction confusion lies among the categories of fishing, other, and other cargo. This phenomenon can be attributed to the analogous geometries shared by these three ship classes.

#### D. Comparison With Traditional Methods and Modern CNN-Based Methods

To thoroughly evaluate the efficiency of the proposed method, we compare the experimental results with the state-of-art methods, including the traditional feature-based methods [6], [7], [8], classic object recognition CNNs [9], [38], [39], [30], [40], and novel task-specific SAR ship recognition CNNs [21], [22], [25], [27], [28], [29]. The comparison methods are our reappearance and our experiments are as consistent as possible with their original reports. It should be noted that the inputs of Zeng et al. [22] and SE-LPN-DPFF [25] are paired VV-VH SAR amplitude images. More specific, the input of training sample number is 150 VV-VH SAR amplitude images for three-category task

and 100 VV-VH SAR amplitude images for six-category task. For other approaches, the inputs consist of unpaired VV and VH SAR amplitude images, wherein single-channel VV and single-channel VH SAR images are sequentially fed directly into the networks. Please note that the FUSAR-Ship dataset solely offers single-channel SAR images, thereby preventing the reappearance of Zeng [22] and SE-LPN-DPFF [25]. Table IX shows the quantitative SAR ship recognition performance with traditional methods and modern CNN-Based methods. From Table IX, the following conclusions can be drawn:

- 1) Among all traditional methods, on the three-category OpenSARShip dataset, the optimal recognition accuracy is 61.72% from KNN, but is still greatly lower than our SA<sup>2</sup>Net (61.72% << 82.91%). On the six-category OpenSARShip dataset, among all traditional methods, the optimal recognition accuracy is 43.54% achieved by SVM. However, this accuracy remains significantly lower compared to our proposed SA<sup>2</sup>Net (43.54% << 60.10%). On the FUSAR-Ship dataset, among all traditional methods, the optimal recognition accuracy is 77.19% achieved by RF. However, this accuracy remains significantly lower compared to our proposed SA<sup>2</sup>Net (77.19% << 88.28%). Modern CNN-based models typically exhibit superior recognition accuracies compared to traditional method, aligning with expectations. This observation suggests that the features extracted by modern CNNs may possess enhanced characterization capabilities.
- 2) On the three-category OpenSARShip dataset, SA<sup>2</sup>Net offer the highest recognition than other modern CNN-based methods. Among all of them, the suboptimal recognition methods is 80.82% from SE-LPN-DPFF [25]. However, it is still lower than our SA<sup>2</sup>Net by 2.09%, which shows the SOTA SAR ship recognition performance of our proposed SA<sup>2</sup>Net.
- 3) On the six-category OpenSARShip dataset, SA<sup>2</sup>Net also offer the highest recognition accuracy than others. Among all of them, the suboptimal recognition method is 59.73% from SE-LPN-DPFF [25]. Nevertheless, our SA<sup>2</sup>Net achieves a 1.37% higher accuracy, showcasing its superior performance as the state-of-the-art SAR ship recognition model.
- 4) On the seven-category FUSAR-Ship dataset, SA<sup>2</sup>Net also offer the highest recognition accuracy than others. Among all of them, the suboptimal recognition methods is 86.69% from HOG-ShipCLSNet [21]. Nevertheless, our SA<sup>2</sup>Net achieves a 1.59% higher accuracy, indicating its superior performance.
- 5) Although SE-LPN-DPFF use the dual-polarization coherence features to characterize ship feature relationships in different polarization channels to improve recognition accuracy, the method neither comprehensively utilize the multiscale features nor leveraged empirical knowledge regarding ships. Thus, its recognition performances are inferior to SA<sup>2</sup>Net's. In addition, although HOG-ShipCLSNet [21] utilized the multiscale features, it simply flattened them and use each feature scale equally, which reduce the ability of the network to extract and choose effective features for precise recognition.



TABLE VI  
CONFUSION MATRIX OF SA<sup>2</sup>NET RECOGNITION RESULTS ON THREE-CATEGORY OPENSARSHIP

True \ Predicted	Bulk carrier	Container ship	Tanker	Recall(%)
	Bulk carrier	255	116	4
Container ship	53	653	5	91.84
Tanker	24	27	203	79.92
Precision(%)	76.81	82.04	95.75	Accuracy=82.91(%)
F1(%)	72.14	86.66	87.12	

TABLE VII  
CONFUSION MATRIX OF SA<sup>2</sup>NET RECOGNITION RESULTS ON SIX-CATEGORY OPENSARSHIP

True \ Predicted	Bulk carrier	Container ship	Tanker	Cargo	Fishing	General cargo	Recall(%)
	Bulk carrier	314	75	17	52	3	14
Container ship	124	580	7	37	12	51	71.52
Tanker	18	18	170	75	13	60	48.02
Cargo	126	20	36	293	0	82	52.60
Fishing	0	0	14	3	100	4	82.64
General cargo	16	8	13	65	3	60	36.36
Precision(%)	52.51	82.74	66.14	55.81	76.34	22.14	Accuracy=61.10%
F1(%)	58.53	76.72	55.65	54.16	79.37	27.52	

TABLE VIII  
CONFUSION MATRIX OF SA<sup>2</sup>NET RECOGNITION RESULTS ON FUSAR-SHIP

Dataset	True \ Predicted	Bulk carrier	Container ship	Fishing	General cargo	Other	Other cargo	Tanker	Recall(%)
		FUSAR-Ship	Bulk carrier	465	6	0	7	5	6
	Container ship	0	516	2	0	2	2	1	98.66
	Fishing	4	0	383	0	40	45	1	80.97
	General cargo	3	2	0	508	0	2	2	98.26
	Other	0	2	38	0	408	64	8	78.46
	Other cargo	5	0	31	2	57	418	8	80.23
	Tanker	6	0	18	2	22	20	453	86.95
	Precision(%)	96.27	98.09	81.14	97.88	76.41	75.04	94.77	Accuracy=88.28%
	F1(%)	95.19	98.38	81.06	98.07	77.42	77.55	90.69	

TABLE IX  
COMPARISON OF SA<sup>2</sup>NET ON THE THREE-CATEGORY AND SIX-CATEGORY UNDER OPENSARSHIP

Model	Three-Category OpenSARShip				Six-Category OpenSARShip				FUSAR-Ship			
	r(%)	p(%)	f1(%)	Acc(%)	r(%)	p(%)	f1(%)	Acc(%)	r(%)	p(%)	f1(%)	Acc(%)
KNN [6]	71.19	71.11	71.15	61.72	32.09	29.38	30.92	40.52	47.03	48.37	47.69	46.54
SVM [7]	54.40	70.29	61.33	57.39	42.06	56.35	48.17	43.54	60.43	59.81	60.12	60.04
RF [8]	39.11	51.89	44.60	56.34	40.39	51.06	45.11	38.90	77.33	77.65	77.49	77.19
AlexNet [9]	53.87	62.22	57.75	63.96	43.05	46.95	44.92	47.1	77.11	77.37	77.24	77.03
VGG-16 [38]	69.15	80.8	74.52	73.67	44.97	53.13	48.71	51.67	80.53	80.67	80.60	80.45
GoogLeNet [39]	65.03	80.23	71.83	73.06	44.04	55.06	48.94	50.06	80.53	80.67	80.60	80.4
ResNet-50 [30]	70.11	81.74	75.48	77.01	46.56	57.58	51.49	56.79	81.40	81.36	81.38	81.31
MobileNet-v2 [40]	69.90	81.46	75.24	76.19	47.21	54.29	50.5	52.19	74.10	74.10	74.10	74.63
Hou [29]	66.30	70.74	68.45	70.60	47.21	54.29	50.5	51.15	71.96	72.19	72.08	71.89
Huang [28]	69.75	81.02	74.96	76.49	<u>58.56</u>	58.71	<u>58.64</u>	56.50	83.13	83.34	83.27	83.19
Zeng [22]	72.84	81.82	77.07	78.13	53.15	59.68	56.23	56.99	-	-	-	-
MS-CNN [27]	70.97	78.06	74.35	77.31	48.13	54.49	51.11	53.52	78.89	80.00	79.50	79.01
HOG-ShipCLSNet [21]	<u>77.30</u>	82.42	79.78	79.55	53.36	55.66	56.23	54.49	<u>86.62</u>	<u>86.54</u>	<u>86.58</u>	<u>86.69</u>
SE-LPN-DPPF [25]	76.17	<u>83.99</u>	<u>79.89</u>	<u>80.82</u>	56.14	<b>59.70</b>	57.87	<u>59.73</u>	-	-	-	-
<b>SA<sup>2</sup>Net(Ours)</b>	<b>79.92</b>	<b>84.87</b>	<b>82.32</b>	<b>82.91</b>	<b>59.54</b>	<u>59.28</u>	<b>59.41</b>	<b>61.10</b>	<b>88.24</b>	<b>88.52</b>	<b>88.38</b>	<b>88.28</b>

TABLE X  
ABLATION STUDIES OF EACH MODULE IN SA<sup>2</sup>NET

Model			Accuracy(%)		
FWM	FAM	SAM	Three-Category	Six-Category	Seven-Category
×	×	×	77.01	56.79	81.31
✓	×	×	80.22	59.52	86.10
×	✓	×	79.70	58.32	84.67
✓	✓	×	81.87	60.29	87.06
✓	×	✓	81.26	60.38	86.69
✓	✓	✓	<b>82.91</b>	<b>61.10</b>	<b>88.28</b>

### E. Ablation Study

In this part, a series of ablation studies on OpenSARShip and FUSAR-Ship are performed to verify the effectiveness of FWM, FAM, and SAM. For a fair comparison, all subsequent studies are performed with the same settings. The overall comparisons are displayed in Table X. Most specifically, adding any of FWM, FAM, i.e., the first three rows of Table X, could boost the recognition accuracy of our model, resulting from the powerful feature supplementation and refinement donated by our task-specific modules. Besides, as can be seen from the fourth and fifth columns of Table X, the accuracy gains further improvements when enabling two modules. Eventually, as can be seen from the sixth column of Table X, compared with the baseline, when applying FWM, FAM, and SAM together, the accuracy of our method achieved the highest on both OpenSARShip and FUSAR-Ship datasets. Next, we will analyze the effectiveness of FWM, FAM, and SAM in detail.

1) *Effect of FWM*: Most existing methods simply extract multiscale features of the network, which limits the performance of SAR ship recognition. To get rich representations at all scales, we leverage semantic and detailed features of different scales extracted by the backbone to construct FWM. Through feature weaving, FWM combines high-level and low-level features to obtain enriched representations. This approach enhances feature discrimination in comparison to the direct utilization of multiscale features extracted solely by the backbone.

Table X provides the results of the FWM in the ablation experiments for both datasets. It should be noted that “×” means that only the last layer features of the backbone network are utilized, ignoring the multiscale features of middle layers from CNN. From Table X, compared with the baseline, FWM gains 3.21% accuracy boost on three-category OpenSARShip task, 2.73% accuracy boost on six-category OpenSARShip task, and 4.79% accuracy boost on FUSAR-Ship task, which is an impressive improvement. To get a comprehensive understanding of FWM, another experiment is conducted to validate the effectiveness of feature weaving, which is named as “ablation study intra FWM.” Table XI provides the results. The “×” means that our SA<sup>2</sup>Net does not perform feature weaving. In other words, the multiscale features are not fused in SA<sup>2</sup>Net. The results in Table XI show that feature weaving achieves 0.52% and 0.89% improvements in accuracy under three-category OpenSARShip and seven-category FUSAR-Ship tasks. Although the improvements in feature weaving are not impressive as other modules, it still demonstrates that integrating high-level and

TABLE XI  
ABLATION STUDY INTRA FWM

Three-Category OpenSARShip		FUSAR-Ship	
Feature Weaving	Accuracy(%)	Feature Weaving	Accuracy(%)
×	82.39	×	87.39
✓	<b>82.91</b>	✓	<b>88.28</b>

low-level information is an effective way to improve SAR ship recognition accuracy.

The improvements of two groups of ablation studies indicate the necessity and effectiveness of combining different scales of CNN to achieve SAR ship recognition of various sizes.

2) *Effect of FAM*: The proposed FAM is introduced to incorporate the priory knowledge of the ship shape into the network by providing rectangular receptive fields that align with the shape of ships. In addition, the directional rectangular kernels can deal with the challenges of arbitrary orientations of ships pose. Table X provides the results of the FAM in the ablation experiments for both datasets. The “×” means that only the square kernel is employed for feature extraction. The results in Table X show that FAM module achieves reasonable 2.69%, 1.53%, and 3.36% improvements in accuracy under the three-category OpenSARShip, six-category OpenSARShip, and FUSAR-Ship tasks compared with the baseline. The improvements show that introducing the rectangular kernels breaks through the limitation of traditional fixed kernel, making the feature extraction more powerful. So FAM is rational for the recognition task of SAR ship target with large aspect ratio and arbitrary orientations.

3) *Effect of SAM*: Although FWM can provide rich representation at all scales, different scale features are not equally effective for recognition. Compared to deep features, shallow features are often not discriminative enough. To adaptively select and assign weights to desired feature scales while disregarding irrelevant scales, we propose SAM to control the information flow of different scales. Table X shows the recognition results with and without SAM. The “×” means that the multiscale features are fused by a simple summation. From Table X, the results show that SAM module achieves reasonable 1.04%, 0.81%, and 1.22% improvements in accuracy under the three-category OpenSARShip, six-category OpenSARShip, and FUSAR-Ship tasks compared with SA<sup>2</sup>Net without SAM. This is in line with our knowledge of CNNs. The shallow features contains more detailed information, which is less in discriminative. So

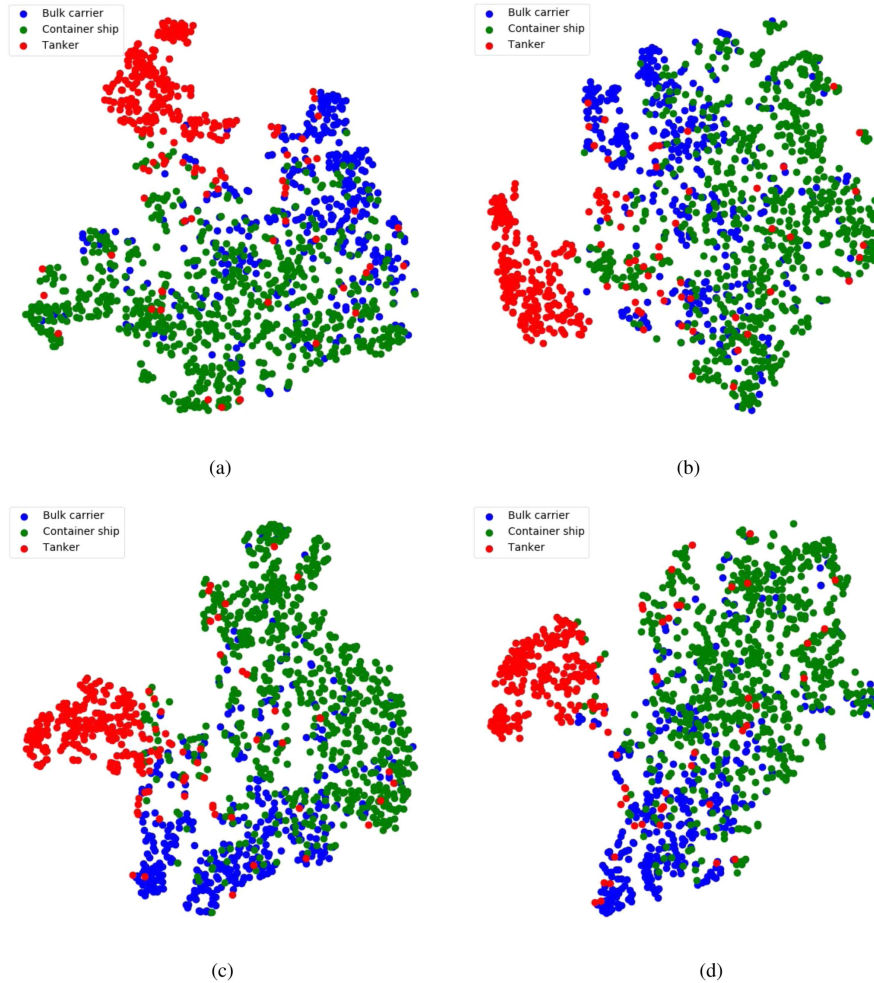


Fig. 8. Three-category task t-SNE feature visualization of the embedding vector distribution. (a) Our network without FWM. (b) Our network without FAM. (c) Our network without SAM. (d) Our network SA<sup>2</sup>Net.

each scale feature is not equally effective for recognition. However, HOG-ShipCLSNet gave the opposite conclusion. They found that the average weighting type achieves a slightly better accuracy than the adaptive type. We analyze their network carefully to find the underlying reasons. One possible reason is that HOG-ShipCLSNet applied too much FC layer in their network. Many of them have more than 2000 neurons, a few even as high as 32 768. When the adaptive type is used, the network might fail to search the suitable weight parameter due to the heavy computational burden, which also lead to ship feature extraction insufficiency.

From the ablation study, FWM, FAM, and SAM have different effects on the recognition of SAR ship with scale variance, large aspect ratio, and arbitrary orientations. Each component of SA<sup>2</sup>Net helps each other to achieve the optimal recognition performance and tackle the problems of SAR ship recognition.

4) *t-SNE*: To provide a comprehensive understanding of the impact of FWM, FAM, and SAM, we visually present the qualitative results using t-distributed stochastic neighbor embedding (t-SNE) [41] of three-category task in Fig. 8. In the t-SNE visualization, the greater the distance between different categories, the higher the recognition accuracy achieved by the model. Fig. 8(a)–(d) illustrate the visualization based on SA<sup>2</sup>Net

without FWM, SA<sup>2</sup>Net without FAM, SA<sup>2</sup>Net without SAM, and SA<sup>2</sup>Net, respectively. It can be found that after supplementing the three modules, the recognition error is alleviated and the feature embeddings from the same class are more aggregated, which is shown in Fig. 8(d). The combination of the three modules separates the features between different categories and the intraclass features of the same category are closer together. These results indicate that the three modules help each other to achieve the optimal recognition performance and tackle with the challenges of SAR ship recognition.

## F. Discussion

In this section, we will further discuss and explain FWM. A discussion about detection and recognition integrated network is also included.

1) *FWM*: Why FWM shows impressive improvement is benefit from two aspects. One is leveraging the multiscale features obtained from intermediate layers. The other is fully mining and combining the feature maps of different scales through feature weaving. We first conduct a comprehensive ablation study to analyze how much the different scale features are related to the final recognition probability on three-category OpenSARShip

TABLE XII  
COMPARISON OF QUANTITATIVE EVALUATION INDICES WITH DIFFERENT  
NUMBER SCALES IN FWM

Dataset	C3	C4	C5	Accuracy(%)
OpenSARShip	×	×	✓	79.70
	×	✓	✓	81.57
	✓	✓	✓	<b>82.91</b>
FUSAR-Ship	×	×	✓	84.67
	×	✓	✓	87.08
	✓	✓	✓	<b>88.28</b>

The bold values mean the optimal recognition results are obtained when three layers are all utilized.

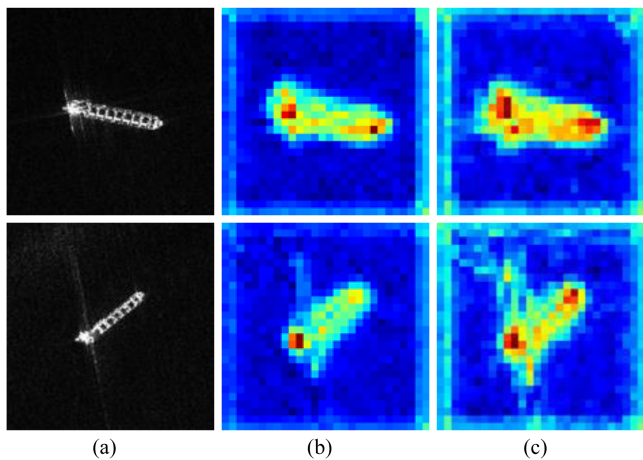


Fig. 9. Visualization of the features of different ships. (a) Original images. Visualization of feature map C3 (b) and M3 (c) with FWM.

and FUSAR-Ship. Then, the feature visualization results of FWM are given.

Table XII shows the results of how much the different scale features are related to the final recognition probability on three-category OpenSARShip and seven-category-FUSAR-Ship. From Table XII, when using a single scale, the recognition results only achieve 79.70% on three category OpenSARShip and 84.67% on seven category FUSAR-Ship. When two scales are employed, SA<sup>2</sup>Net improves results by 1.87% on three-category OpenSARShip and by 2.41% on FUSAR-Ship. The optimal recognition results are obtained when three layers are all utilized, showing the necessity of leveraging multiscale features to recognize ships of various sizes.

To validate the effectiveness of feature weaving, we present some qualitative visualization results of feature maps C3 and M3 in Fig. 9. The sizes of C3 and M3 are  $28 \times 28$  pixels. The activation heatmap of the extracted feature is the summation of the values in each row along the channel dimension. Fig. 9(a) is the original SAR ship images. As illustrated in Fig. 9(b), although features extracted by C3 focus on ship targets, the features are not sufficient enough. To deal with the problem and capture more information, feature weaving can integrate high-level and low-level information through a weaving process, resulting in rich representations. As shown in Fig. 9(c), the

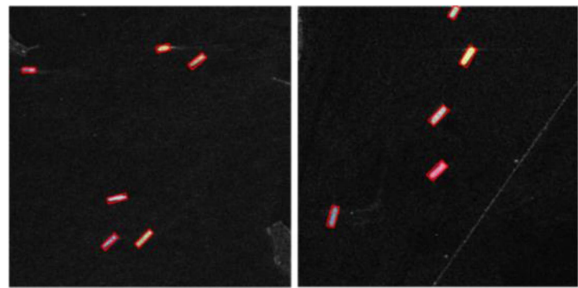


Fig. 10. Ship detection results with OBox.

network with feature weaving has more information and pays more attention to the distinguishable regions, so these important parts have higher activation scores. It proves the feature weaving effectively enriches the representations of SAR ship targets.

2) *Detection and Recognition Integrated Network*: Nowadays, an increasing number of scholars have paid more attention to establishing a unified detection and recognition SAR ship network [42], [43], [44]. However, the detection and recognition parts are independent and irrelevant in classical algorithms. On the contrary, in practical applications, it is often necessary to perform detection and recognition tasks in the SAR images simultaneously. To achieve satisfied unified detection and recognition performance, one necessary way is to inject more discriminative features extracted by SAR ship recognition methods to SAR ship detection methods. As, shown in Fig. 10, the most recent detection algorithms [36], [45] utilize oriented bounding box (OBox) to tackle with the challenge of arbitrarily oriented ships. In SA<sup>2</sup>Net, the proposed FAM utilizes directional rectangular convolution kernels to solve the same problem. In the future study, I believe the joint utilization of FAM and OBox may boost the performance of detection and recognition integrated network.

#### IV. CONCLUSION

In this article, we propose a SA<sup>2</sup>Net to further improve the performance of ship recognition in SAR image. ResNet-50 is adopted as the backbone to extract SAR ship features. Taking into account the special shape prior characteristics of the ship class, the FAM in SA<sup>2</sup>Net is designed to enhance the semantic features of ships, which incorporate the priority knowledge of the ship shape. The proposed FAM breaks through the limitation of traditional square kernels. In addition, to achieve ship recognition with diverse sizes, the comprehensive utilization of multiscale features holds paramount importance. Different from aggregate multiscale features with unified weights, SAM in SA<sup>2</sup>Net adaptively weights the desired feature scales and disregards the irrelevant scales. The proposed FWM in SA<sup>2</sup>Net generates rich and reliable representations through repeatedly fusing the representations produced by the backbone to obtain better representations at all scales. The experimental results, comparisons, and ablation studies on representative three- and six-category OpenSARShip tasks show that SA<sup>2</sup>Net greatly improves the recognition performance.

## REFERENCES

- [1] X. Xu, X. Zhang, and T. Zhang, "Lite-YOLOv5: A lightweight deep learning detector for on-board ship detection in large-scene Sentinel-1 SAR images," *Remote Sens.*, vol. 14, no. 4, 2022, Art. no. 1018.
- [2] G. Margarit and A. Tabasco, "Ship classification in single-pol SAR images based on fuzzy logic," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 3129–3138, Aug. 2011.
- [3] W. Pu, "Deep SAR imaging and motion compensation," *IEEE Trans. Image Process.*, vol. 30, pp. 2232–2247, 2021.
- [4] D. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE 7th Int. Conf. Comput. Vis.*, 1999, pp. 1150–1157.
- [5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 886–893.
- [6] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [7] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.
- [8] T. Kam Ho, "Random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
- [10] H. Zheng, Z. Hu, J. Liu, Y. Huang, and M. Zheng, "MetaBoost: A novel heterogeneous DCNNs ensemble network with two-stage filtration for SAR ship classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [11] Z. Shao, T. Zhang, and X. Ke, "A dual-polarization information-guided network for SAR ship classification," *Remote Sens.*, vol. 15, no. 8, 2023, Art. no. 2138.
- [12] J. Li, C. Qu, and S. Peng, "Ship classification for unbalanced SAR dataset based on convolutional neural network," *J. Appl. Remote Sens.*, vol. 12, no. 3, 2018, Art. no. 035010.
- [13] J. Q. Shao, Q. U. Chang-Wen, L. I. Jian-Wei, S. J. Peng, and N. A. University, "CNN based ship target recognition of imbalanced SAR image," *Electron. Opt. Control*, 2019.
- [14] Y. Zhang, Z. Lei, H. Yu, and L. Zhuang, "Imbalanced high-resolution SAR ship recognition method based on a lightweight CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [15] J. A. Raj, S. M. Idicula, and B. Paul, "One-shot learning-based SAR ship classification using new hybrid siamese network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [16] C. Lu and W. Li, "Ship classification in high-resolution SAR images via transfer learning with small training dataset," *Sensors*, vol. 19, no. 1, 2018, Art. no. 63.
- [17] W. Yuanyuan, W. Chao, and Z. Hong, "Ship classification in high-resolution SAR images using deep learning of small datasets," *Sensors*, vol. 18, no. 9, 2018, Art. no. 2929.
- [18] Y. Xu and H. Lang, "Distribution shift metric learning for fine-grained ship classification in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2276–2285, 2020.
- [19] J. He, Y. Wang, and H. Liu, "Ship classification in medium-resolution SAR images via densely connected triplet CNNs integrating fisher discrimination regularized metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3022–3039, Jul. 2020.
- [20] Z. Huang, M. Datcu, Z. Pan, and B. Lei, "Deep SAR-Net: Learning objects from signals," *ISPRS J. Photogrammetry Remote Sens.*, vol. 161, pp. 179–193, 2020.
- [21] T. Zhang et al., "HOG-ShipCLSNet: A novel deep learning network with HOG feature fusion for SAR ship classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–22, 2021.
- [22] L. Zeng et al., "Dual-polarized SAR ship grained classification based on CNN with hybrid channel feature loss," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.
- [23] J. He et al., "Group bilinear CNNs for dual-polarized SAR ship classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [24] G. Xiong, Y. Xi, D. Chen, and W. Yu, "Dual-polarization SAR ship target recognition based on mini hourglass region extraction and dual-channel efficient fusion network," *IEEE Access*, vol. 9, pp. 29078–29089, 2021.
- [25] T. Zhang and X. Zhang, "Squeeze-and-excitation Laplacian pyramid network with dual-polarization feature fusion for ship classification in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.
- [26] H. Qiu, H. Li, Q. Wu, F. Meng, K. N. Ngan, and H. Shi, "A2RMNet: Adaptively aspect ratio multi-scale network for object detection in remote sensing images," *Remote Sens.*, vol. 11, no. 13, 2019, Art. no. 1594.
- [27] X. Xu, X. Zhang, and T. Zhang, "Multi-scale SAR ship classification with convolutional neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 4284–4287.
- [28] L. Huang et al., "OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 195–208, Oct. 2017.
- [29] X. Hou, W. Ao, Q. Song, J. Lai, H. Wang, and F. Xu, "FUSAR-Ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition," *Sci. China-Inf. Sci.*, vol. 63, Mar. 2020, Art. no. 140303.
- [30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [31] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5686–5696.
- [32] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.
- [33] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [34] S. Wei, X. Zeng, H. Zhang, Z. Zhou, J. Shi, and X. Zhang, "LFG-Net: Low-level feature guided network for precise ship instance segmentation in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022.
- [35] 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, 2021.
- [36] J. Zhang, M. Xing, G.-C. Sun, and N. Li, "Oriented Gaussian function-based box boundary-aware vectors for oriented ship detection in multiresolution SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.
- [37] 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.
- [38] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Proc. Adv. Neural Inform. Process. Syst.*, vol. 25, 2012.
- [39] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
- [40] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: <http://arxiv.org/abs/1704.04861>
- [41] V. D. M. Laurens and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 2605, pp. 2579–2605, 2008.
- [42] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751–755, May 2019.
- [43] Z. Sun, X. Leng, Y. Lei, B. Xiong, K. Ji, and G. Kuang, "BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images," *Remote Sens.*, vol. 13, no. 21, 2021, Art. no. 4209.
- [44] 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.
- [45] Y. Sun, Z. Wang, X. Sun, and K. Fu, "SPAN: Strong scattering point aware network for ship detection and classification in large-scale SAR imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1188–1204, 2022.



**Yuanzhe Shang** (Graduate Student Member, IEEE) received the B.S. degree in electronic engineering from the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2020, where he is currently working toward the Ph.D. degree in electronic engineering.

His research interests include radar signal processing, target detection, machine learning, and automatic target recognition.



**Wei Pu** (Member, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2012 and 2018, respectively.

From 2017 to 2018, he was a Visiting Student with the Department of Electrical Engineering, Columbia University, New York, NY, USA. From 2019 to 2022, he was a Research Fellow with University College London, London, U.K. He is currently a Professor with the School of Information and Communication

Engineering, UESTC. His research interests include sparse signal processing and deep learning.

Dr. Pu was a recipient of the Newton International Fellowship from the Royal Society, U.K.



**Danling Liao** received the B.S. degree from the School of Science, Xi'an Jiaotong University, Xi'an, China, in 2020. She is currently working toward the master's degree with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China.

Her research interests include medical image analysis, medical information processing, and computer assisted interventional diagnosis.



**Ji Yang** received the B.S. degree from Logistic Engineering University, PLA, Chongqing, China.

She is currently an Engineer of Unit 31308 of the PLA. Her research interests include radar signal processing, target detection, machine learning, and automatic target recognition.



**Congwen Wu** (Graduate Student Member, IEEE) received the B.S. degree in electronic engineering from the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2020, where he is currently working toward the Ph.D. degree in electronic engineering.

His research interests include radar signal processing, machine learning, and automatic target recognition.



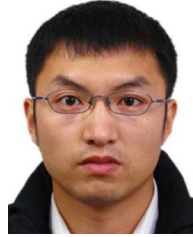
**Yulin Huang** (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from the School of Electronic Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2002 and 2008, respectively.

He is currently a Professor with the UESTC and the Dean of School of Information and Communication Engineering. His research interests include radar signal processing and SAR automatic target recognition.



**Yin Zhang** (Member, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from the School of Electronic Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2008 and 2016, respectively.

From 2014 to 2015, he had been a visiting student with the Department of Electrical and Computer Engineering, University of Delaware, Newark, USA. He is currently a research fellow at the UESTC. His research interests include signal processing and radar imaging.



**Junjie Wu** (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronic engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2004, 2007, and 2013, respectively.

He is currently a Professor with UESTC and the Vice Dean of School of Information and Communication Engineering. His research interests include signal processing and radar imaging.

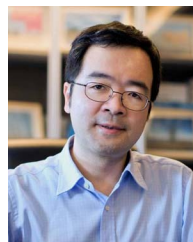


**Jianyu Yang** (Member, IEEE) received the B.S. degree from the National University of Defense Technology, Changsha, China, in 1984, and the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1987 and 1991, respectively, all in electronic engineering.

He is the Vice Chairman of the Radar Society, Chinese Institute of Electronics. From 2001 to 2005, he served as the Dean of School of Electronic Engineering of UESTC. In 2005, he was a Senior Visiting

Scholar with the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. He is currently a Professor with UESTC. He is a Senior Editor for the *Chinese Journal of Radio Science* and the *Journal of Systems Engineering and Electronics*. His research interests include synthetic aperture radar imaging and automatic target recognition.

Dr. Yang was selected as the Vice-Chairperson of the Radar Society of the Chinese Institute of Electronics (CIE) in 2016 and a Fellow of CIE in 2018. He has been awarded the titles of "Cross-Century Excellent Talent" by the Ministry of Education, "Academic and Technical Leader" by Sichuan Province, "Top 10 Excellent Scientific and Technological Workers of the Chinese Institute of Electronics," and "Fellow of the Chinese Institute of Electronics."



**Jianqi Wu** received the B.S. degree from Beijing University of Aeronautics and Astronautics, Beijing, China, in 1983 and the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1990 and 2018, respectively.

He is the Chairman of the Radar Society of Chinese Institute of Electronics. He has been working in radar for more than 30 years. He was In Charge of a Key National Defense Advanced Research Project "Sparse Array Synthetic Impulse and Aperture Radar Experimental System" and several key model projects.

Dr. Wu has received the first class award of the National Scientific and Technological Progress Award, the second class award of the National Scientific and Technological Progress Award three times, and the first class award of the National Defense Scientific and Technological Progress Prize and Outstanding Contribution Award of Science and Technology of Hefei. He is also a Fellow Member of the Chinese Academy of Engineering.