Fishing Vessel Classification in SAR Images Using a Novel Deep Learning Model

Yanan Guan, Xi Zhang, Siwei Chen, *Senior Member, IEEE*, Genwang Liu, Yongjun Jia, Yi Zhang, Gui Gao, Jie Zhang, Zhongwei Li, and Chenghui Cao

***Abstract*— With the development of deep learning (DL), research on ship classification in synthetic aperture radar (SAR) images has made remarkable progress. However, such research has primarily focused on classifying large ships with distinct features, such as cargo ships, containers, and tankers. The classification of SAR fishing vessels is extremely challenging because of two main reasons: 1) the small size and minor interclass differences of fishing vessels make learning fine-grained features difficult and 2) determining fishing vessel types is difficult, resulting in a lack of labeled data. Hence, after designing a process framework for vessel tagging, we construct a high-resolution fine-grained fishing vessel classification dataset (FishingVesselSAR), which contains 116 gillnetters, 72 seiners, and 181 trawlers. We then propose a novel DL model (FishNet) that aims to strengthen feature extraction and utilization. In FishNet, we introduce four innovative modules to ensure superior performance in SAR fishing vessel classification: a multipath feature extraction (MUL) module, a feature fusion (FF) module, a multilevel feature aggregation (MFA) module, and a parallel channel and spatial attention (PCSA) module. Furthermore, we design an adaptive loss function to achieve better classification performance by mitigating the effects of class imbalance. In this article, we report extensive ablation studies conducted to confirm the efficacy of the five improvements listed above. Sufficient comparisons with 33 advanced methods from the DL and SAR target classification communities demonstrate that FishNet achieves an SAR fishing vessel classification accuracy of 89.79%, which is 6.77% higher than that of the second-best method.**

***Index Terms*— Fishing vessel classification, FishNet, high accuracy, synthetic aperture radar (SAR).**

I. Introduction

FISHERIES contribute to livelihoods, food security, and human health worldwide, and their sustainable development is of great significance to all humankind. However, overfishing due to illegal, unreported, and unregulated (IUU) fishing has resulted in a precipitous decline in key fish stocks, threatening biodiversity, and disrupting ecosystems out of balance [1], [2], [3]. Despite numerous initiatives to combat IUU fishing, the UN's Sustainable Development Goal (SDG) Indicator 14.6.1, which aims to eliminate IUU fishing by 2020, has yet to be achieved [4]. To realize the goal of sustainable fisheries, effective and reliable monitoring of fishing vessels in a fishing area is essential [5]. It should be noted that, due to the different species being caught, different types of fishing vessels have varying impacts on bycatch, habitat, and the rate at which overfishing occurs. Therefore, identifying the types of fishing vessels is a priority in fishery monitoring.

At present, the main technical tool for monitoring fishing vessels is the automatic identification system (AIS), which can transmit vessel-related information autonomously and continuously. AIS messages can provide vessel identification data, such as the name and Maritime Mobile Service Identity (MMSI), as well as vessel type codes, such as "70" for cargo ships, "80" for tankers, and "30" for fishing vessels. However, the type code in AIS data only indicates whether a vessel is a fishing vessel and not the type of fishing it performs. Therefore, it is not possible to effectively monitor fishing behavior based solely on the AIS information reported by fishing vessels in real-time, let alone combat IUU fishing.

Researchers have discovered that different types of fishing vessels exhibit significant variations in their fishing activities and in their supportive equipment for the deployed fishing gear on board. Variations in fishing behavior are mainly evidenced by speed, heading, and trajectory, which can be obtained through AIS data. Hence, the processing, analysis, and deep mining of AIS data with high spatiotemporal multigranularity represent the primary research direction for accurately identifying fishing vessel types [6], [7], [8], [9]. However,

Manuscript received 9 June 2023; revised 28 August 2023; accepted 3 September 2023. Date of publication 7 September 2023; date of current version 20 September 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61971455, Grant 42106177, Grant U2006207, Grant 62122091, Grant 62131019, and Grant 42206184; in part by the Foundation of Equipment Pre-Research Area under Grant 80915020107; in part by the National Defense Science and Technology Basic Enhancement Program under Grant 2022JCJQZD1470003; and in part by the Shandong Provincial Natural Science Foundation of China under Grant ZR2021QF093. *(Corresponding author: Xi Zhang.)*

Yanan Guan, Xi Zhang, Genwang Liu, Jie Zhang, and Chenghui Cao are with the First Institute of Oceanography and the Technology Innovation Center for Ocean Telemetry, Ministry of Natural Resources of China, Qingdao 266061, China (e-mail: guanyanan@fio.org.cn; xi.zhang@fio.org.cn; liu_genwang@126.com; zhangjie@fio.org.cn; caochenghui@fio.org.cn).

Siwei Chen is with the State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information Systems, National University of Defense Technology, Changsha 410073, China (e-mail: chenswnudt@163.com).

Yongjun Jia and Yi Zhang are with the National Satellite Ocean Application Service, Ministry of Natural Resources, Beijing 100081, China (e-mail: jiayongjun@mail.nsoas.org.cn; zhangyi@mail.nsoas.gov.cn).

Gui Gao is with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China (e-mail: dellar@126.com).

Zhongwei Li is with the College of Oceanography and Space Informatics, China University of Petroleum, Qingdao 266580, China (e-mail: li.zhongwei@vip.163.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/TGRS.2023.3312766, provided by the authors.

Digital Object Identifier 10.1109/TGRS.2023.3312766

the reported AIS data may be lost or erroneous for various reasons, such as operators deliberately switching off transponders or falsifying data when entering fishery management areas [10], [11], [12]. This poses a significant threat to fishing vessel management, particularly in combating IUU fishing. Therefore, there is an urgent need to develop new monitoring methods.

Unlike AIS, which relies on the active cooperation of ships, synthetic aperture radar (SAR) can capture targets under all-day and all-weather conditions by actively sending and receiving microwave pulses [13], [14]. Based on the synergistic exploitation of SAR and AIS data, scholars have conducted studies on the classification of SAR detections as fishing or nonfishing vessels [15], the detection of dark fishing fleets [16], [17], large-scale open ocean monitoring to track human activity across the oceans [18], and so on. To the best of our knowledge, however, there have been no previous studies on the identification of fishing vessel types in SAR images.

Currently, research on ship classification in SAR images mainly focuses on large ships with distinct features, such as cargo ships, container ships, and tankers. Traditional SAR ship classification methods are mainly implemented by extracting handcrafted features and applying classifiers such as sparse representation classifiers [19], [20] and support vector machines (SVMs) [21], [22]. Some common handcrafted features include geometric [23], [24], backscatter [25], histogram of oriented gradients (HOG) [26], [27], texture [28], [29], and scale-invariant feature transform (SIFT) features [30], [31]. In summary, traditional methods can achieve high performance in some cases, but they require time-consuming and laborious feature design and are weak in generalizability. However, the above shortcomings can be satisfactorily resolved with modern convolutional neural network (CNN)-based methods that have become available in recent years.

CNN-based methods can automatically extract rich features of targets although these features tend to be too abstract to be understood by humans [32]. However, these abstractions may be more consistent with the inherent logic of nature [33], which greatly improves classification performance. In 2017, Bentes et al. [34] used four classifiers to classify maritime targets, including cargo ships, tankers, windmills, platforms, and port structures. The results indicated that CNN-based models performed much better than traditional classifiers. Wang et al. [35] built a ship classification model based on very deep convolutional networks. The network model was pretrained using transfer learning and fine-tuning to achieve an average classification accuracy of 95% for bulk carriers, containers, and tankers. However, long pretraining times result in poor efficiency. OpenSARShip, a high-quality dataset dedicated to Sentinel-1 ship interpretation, was released by Huang et al. [36] as a basis for developing deep learning (DL) models for SAR ship recognition. Based on this dataset, many research results have emerged. For example, Wu et al. [37] proposed a joint convolution network with a generator and a classifier block to classify three kinds of ships. Using a deep metric learning scheme, He et al. [38] proposed a densely connected triadic CNN to expand the interclass distance and achieved satisfactory classification results for ships in medium-resolution SAR images. Arguing that the robustness of individual CNN models is weak for SAR ship classification, Zheng et al. [39] proposed an automated method for heterogeneous deep CNNs based on two-stage filtering. In addition, Hou et al. [40] released the high-resolution FUSAR-Ship dataset collected from the Gaofen-3 (GF-3) satellite. Due to the availability of large-scale datasets, the design of CNN-based SAR ship classifiers has recently attracted increasing attention from scholars [41], [42], [43].

While vigorously developing new CNN models, researchers have also attempted to develop better SAR ship classification methods by fusing traditional features with the abstract features extracted by CNNs. Wang et al. [44] proposed a deep framework for target classification utilizing multiple CNNs by incorporating the intensity and edge information of SAR images. Zhang et al. [45] designed a novel CNN that incorporates traditional HOG features. Afterward, these authors integrated dual polarization features and geometric features into CNNs [46]. Experimental results have shown that such CNN models with traditional features embedded can achieve preferable results in large ship classification.

Overall, the above studies on the classification of SAR ships with distinguishable characteristics and significant interclass differences have achieved satisfactory results. However, it remains challenging to classify different types of fishing vessels. First, different types of fishing vessels seem relatively similar in appearance. There are minor differences caused by the fishing gear and supporting facilities deployed on board, but these differences are far less significant than those between merchant ships, such as containers and tankers. Furthermore, compared to merchant ships, fishing vessels are smaller, especially offshore fishing vessels, which are typically less than 40 m in length. Due to this, obtaining detailed features and internal structures of fishing vessels from SAR images is more challenging. In addition, the SAR imaging mechanism makes such images more vulnerable to speckle noise [47], [48], and the defocusing and motion blur phenomena are more serious. Thus, the key to achieving high-precision fishing vessel classification in SAR images is to determine how to extract more effective information despite the interference of these unfavorable factors and, thereby, capture the discriminative features among different types of fishing vessels.

In addition, annotated and high-quality datasets are the basis for SAR ship classification. However, as discussed earlier, fishing vessel types cannot be directly determined from AIS information. Consequently, constructing a fine-grained fishing vessel type dataset is more complex, difficult, and time-consuming than constructing other ship datasets.

To address the abovementioned problems, we first propose a process framework for labeling fishing vessel targets in SAR images. After cropping the SAR chips corresponding to fishing vessels, we use an AIS-based fishing vessel type identification model to determine the type of each fishing vessel. This process helps us construct a high-resolution SAR fishing vessel dataset named FishingVesselSAR. Second, different from previous SAR ship classification methods based on

TABLE I
DETAILS OF THE GF-3 SAR DATA PROCESSED IN THIS ARTICLE

Imaging mode	Incidence angle (°)	Resolution (m)	Imaging swath (km)	Polarization	Number
Ultrafine strip (UFS)	20–50	3	30	Single	158
Fine strip I (FSI)	19–50	5	50	Dual	96
Full polarized strip I (QPSI)	20–41	8	20–35	Dual	88
Fine strip II (FSII)	19–50	10	95–110	Quad	44

image preprocessing or network layer stacking, we propose a novel DL model named FishNet from the perspective of strengthening feature extraction and utilization. In FishNet, four innovative modules are designed to ensure excellent classification performance for SAR fishing vessels. These modules are: 1) a multipath feature extraction (MUL) module; 2) a feature fusion (FF) module; 3) a multilevel feature aggregation (MFA) module; and 4) a parallel channel and spatial attention (PCSA) module. In addition, we propose an adaptive loss function to address the class imbalance problem in SAR fishing vessel classification, further improving the accuracy. To confirm the effectiveness of these five improvements, we conducted extensive ablation studies. Compared to 33 other advanced models, FishNet achieves the highest accuracy of 89.67% in SAR fishing vessel classification, which is 6.77% higher than the second-best model.

The main contributions of this article are given as follows.

1) The identification of fishing vessel types in SAR images is addressed for the first time.
2) A high-resolution SAR fishing vessel dataset (FishingVesselSAR) is constructed, which lays the foundation for SAR fishing vessel classification.
3) A novel DL model called FishNet is proposed for SAR fishing vessel classification.
4) In FishNet, four modules and a proprietary loss function are designed to ensure superior classification performance for SAR fishing vessels.
5) Compared with 27 advanced DL models and six advanced SAR target classification models, FishNet achieves a considerable advantage in classification accuracy.

The rest of this article is organized as follows. Section II describes the creation of the FishingVesselSAR dataset. Section III introduces the principles and implementation of FishNet in detail. Section IV introduces the details of the experiments and reports the results. Ablation studies are presented in Section V. Finally, Section VI concludes this article.

II. SAR FISHING VESSEL DATASET

A. GF-3 SAR Data

GF-3 is China's first civilian C-band SAR satellite [49]. With a 29-day revisit period, it circles the Earth in a Sun-synchronous orbit at an altitude of 755 km. The GF-3 satellite is designed with 12 imaging modes, offering the advantages of high resolution and a large imaging width.

In this article, GF-3 SAR images acquired in multiple modes in the offshore waters of China (17.53°N–40.89°N, 108.37°E–125.40°E), with ground resolutions ranging from 3 to 10 m, are utilized to undertake a study of fishing vessel type identification. A total of 386 GF-3 images were obtained from April 2019 to April 2022. The details of these scenes are summarized in Table I, and their geographical distribution is shown in Fig. S1 in the Supplementary Material. For the preprocessing of GF-3 SAR images, we employ SNAP 3.0 [50] to perform geometric correction and radiometric calibration.

B. AIS Data

The AIS data sent by a shipboard communication device can provide near-real-time messages containing static information (e.g., name, MMSI, and call sign), dynamic navigation information (e.g., position, speed, and heading), and voyage-related information (e.g., draft and destination) [51]. The AIS data used in this study were purchased from commercial companies (http://www.boloomo.com/, accessed on May 25, 2023), covering the period from 2019 to 2022.

For each SAR image, the corresponding AIS messages from the same zone were retrieved in a 10-min window centered on the SAR acquisition time. Thus, we obtained the original AIS records covering all space–time regions of the 386 GF-3 SAR scenes.

With the goal of constructing SAR fishing vessel samples, we retained the AIS data of all fishing vessels (that is, data with the unique type code "30") and deleted all the rest to facilitate subsequent SAR-AIS coregistration. Furthermore, data cleaning operations were performed to eliminate duplicate and incorrect records.

C. Establishment of the FishingVesselSAR Dataset

Despite the large number of GF3 images available for use, the small number and size of the fishing vessels make it necessary to complete the entire process of constructing SAR fishing vessel samples by hand to ensure the accurate extraction of each fishing vessel from each SAR image. Therefore, we develop a process framework for labeling SAR fishing vessel targets, which mainly consists of two phases: cropping SAR fishing vessels and identifying fishing vessel types.

1) Cropping SAR Fishing Vessels: Establishing reliable links between AIS messages and the fishing vessels in SAR images is a prerequisite for cropping SAR fishing vessels. In fact, the acquisition time of an SAR image may not completely coincide with the reporting time of the most closely corresponding AIS data, and Doppler shifts may lead to positional differences between SAR vessels and the associated AIS messages. To overcome the former challenge, we interpolate the AIS data to derive the vessel positions at the

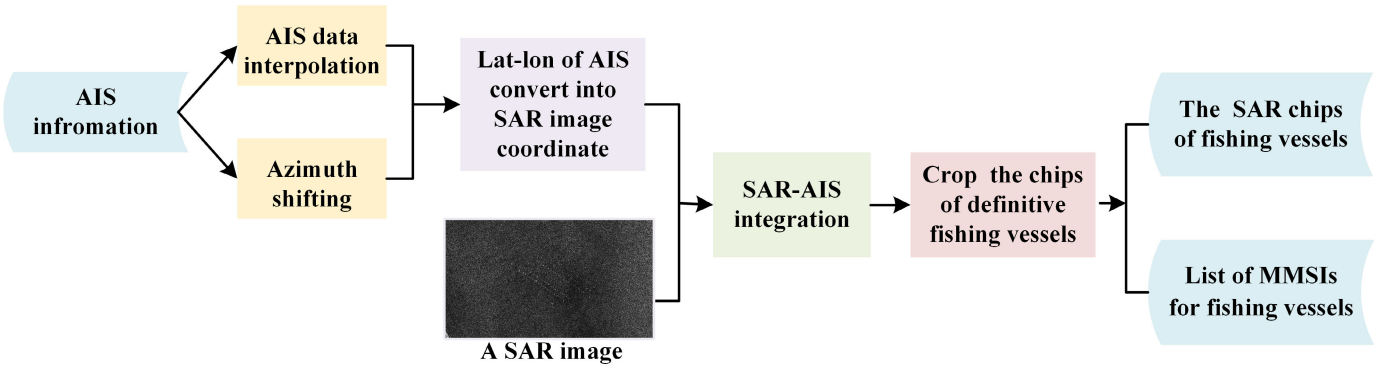


Fig. 1. Flowchart of the process of cropping SAR fishing vessel chips.

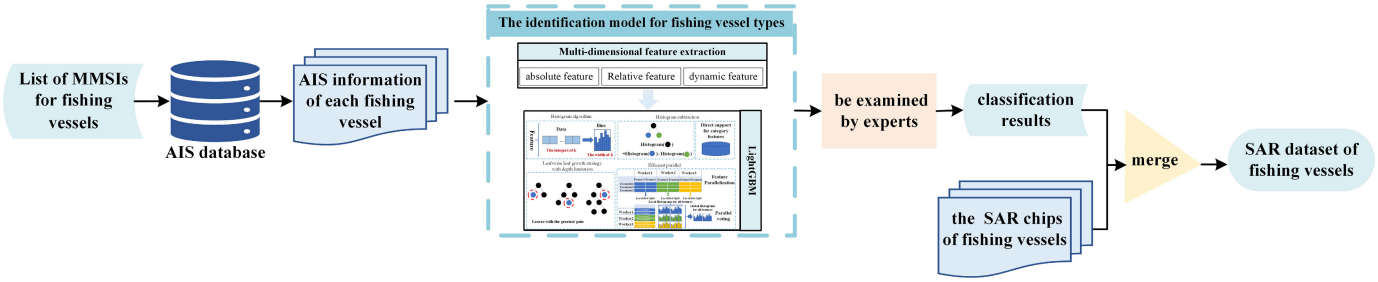


Fig. 2. Flowchart of the process of identifying fishing vessel types.

time of SAR acquisition. To address the latter, we calculate the Doppler shifts [52] based on the velocity vectors in the AIS messages to correct the dead reckoning positions of the vessels. Fig. 1 illustrates the detailed operational flow. After achieving the optimal spatiotemporal matches, we obtained SAR chips corresponding to 511 fishing vessels subject to strict quality control. All chips were saved as single-precision floating-point amplitude images.

2) Identifying Fishing Vessel Types: Gillnetters, seiners, and trawlers are the main types of fishing vessels in Chinese waters, and their combined fishing production accounted for 78.30% of the total fishing in 2019. The employment of different fishing methods results in distinct differences in the trajectories of these fishing vessels plotted by AIS data (see Fig. S2 in the Supplementary Material), and a detailed description of these differences is provided in the Supplementary Material. Therefore, we can identify the type of fishing vessel by tracking its AIS trajectory 15 days before and after the corresponding SAR acquisition time. Fig. 2 shows the type identification model used in the above process. Based on AIS data, the identification model can reliably identify the fishing type of each vessel by comprehensively mining the differences that characterize different fishing vessel types. For detailed principles and the implementation process, please refer to [8].

For the construction of the FishingVesselSAR dataset, the identification model identifies 511 fishing vessels as 116 gillnetters, 72 seiners, 181 trawlers, and 142 fishing vessels of uncertain type. Thus, the final FishingVesselSAR dataset contains SAR chips corresponding to 369 fishing vessels of known types. Among them, the SAR chips of 267 fishing vessels have resolutions better than 8 m, accounting for approximately 91.32% of the total samples. The detailed statistics of the FishingVesselSAR dataset are shown in Fig. S3 in the Supplementary Material.

III. Method and Model

In this section, we first introduce the overall structure of the proposed FishNet and describe the feature maps involved. After that, four important modules and the proprietary loss function in FishNet are introduced, namely, the MUL module, the FF module, the MFA module, the PCSA module, and the adaptive loss function.

A. Overview of the Model

Due to the small size of the targets and the noise of SAR images, capturing the interclass differences among fishing vessels is challenging. To this end, we have developed a novel DL model named FishNet that maximizes feature extraction and utilization to fully obtain SAR fishing vessel discriminative features. The overall architecture of FishNet is shown in Fig. 3.

Four innovative modules and a novel loss function are designed in FishNet for solving fishing vessel types of identification in SAR images. FishNet's construction process is given as follows.

First, the MUL module is designed to fully extract features from input images so that subtle differences between fishing vessels can be captured to the greatest extent. FishNet's basic backbone for feature extraction is formed by connecting multiple MUL modules with Transition modules. As shown in Fig. 3, the Transition module uses average pooling layers with a stride of 2 (marked @2) to achieve a size reduction of feature

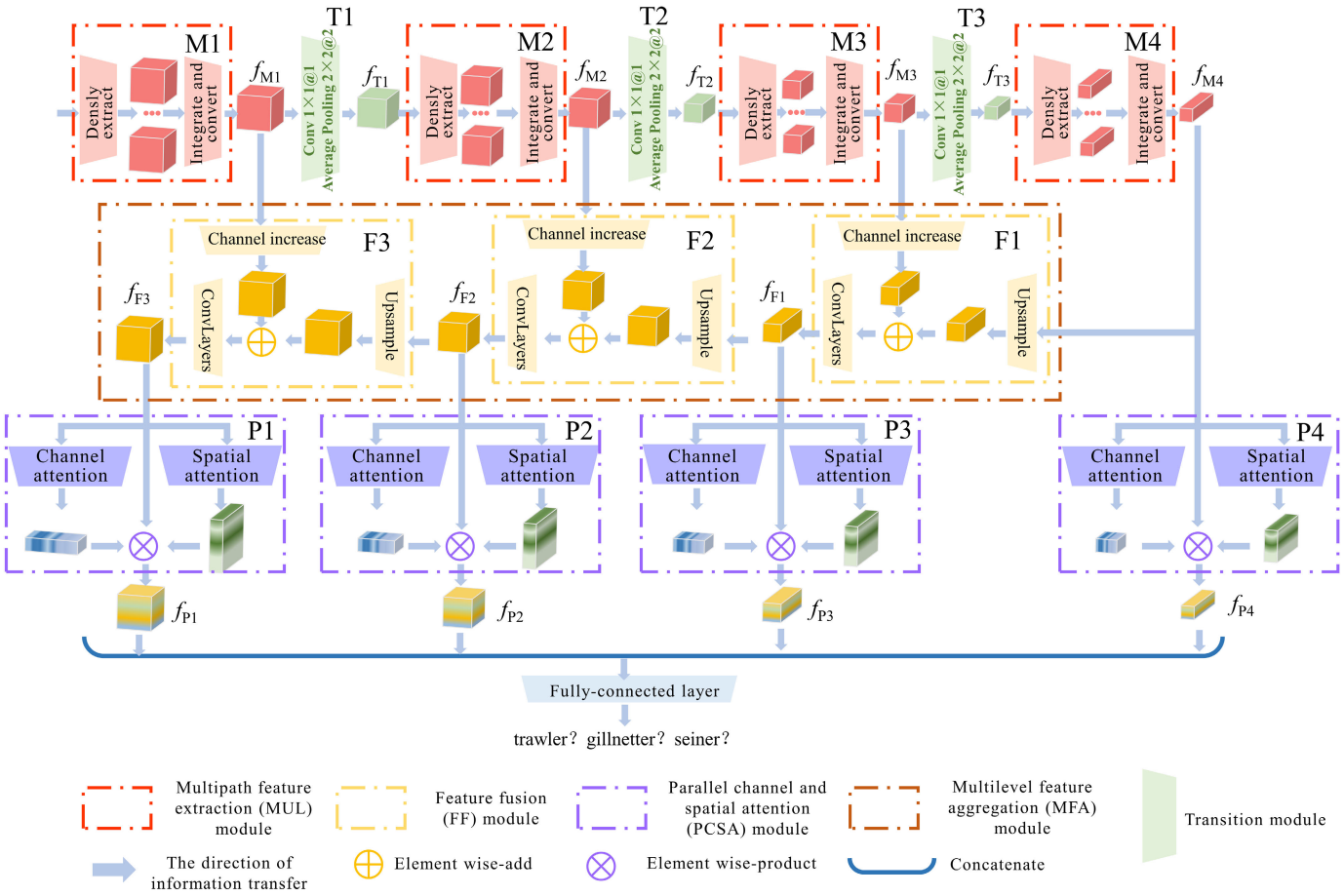


Fig. 3. Network architecture of FishNet.

maps extracted by the previous MUL module. As a result, the network parameters of FishNet can be compressed, while the feature receptive field of the subsequent MUL module can be increased.

Second, fishing vessels are relatively small, and to improve classification accuracy, it is crucial to emphasize details and local information while fully extracting deep features. We have designed the FF module to fuse the features at different levels extracted by the backbone since high- and low-level features capture sufficient semantic and detailed information, respectively. Furthermore, we have designed the MFA module based on the FF module to achieve multilevel FF. By jointly applying the FF module and the MFA module, the high-efficiency fusion of the features in different receptive fields contributes to improving the spatial information contained in the features, enabling FishNet to obtain more comprehensive feature information from fishing vessels.

To suppress background features caused by interference phenomena such as speckle noise while highlighting interclass differences, we design the PCSA module to perform channel and spatial attention (SA) processing on the fused features in parallel.

Then, after attention-based feature recalibration by the PCSA module, the fused feature maps are integrated as the elements to be input to the fully connected layer for better classification results.

Finally, an adaptive loss function is proposed to address the data imbalance problem for SAR fishing vessels and is used to guide the training process.

Considering the statistical distribution of the pixel sizes (see Fig. S3 in the Supplementary Material), SAR images must first be scaled to a uniform size of 56 × 56 pixels before being input into FishNet. Table II displays the main modules in FishNet and the names and sizes of their outputs.

B. MUL Module

Due to the small interclass differences and the limited number of SAR fishing vessel samples available for model training, a prerequisite for improving classification accuracy is to extract as much effective feature information as possible. Taking this into account, we design the MUL module that consists of a MUL-DenseBlock for enhanced feature extraction and a ConversionBlock for feature compression, as shown in Fig. 4.

As shown in Fig. 4(a), to extract sufficient features, the input feature maps that enter the MUL module are first processed by the MUL-DenseBlock, which generates more than twice the feature maps. Inspired by DenseNet's [53] dense connection mechanism, we introduce the DenseBlock module with a feature reuse function in each branch of the MUL-DenseBlock to extract features from the input feature maps.

TABLE II
DESCRIPTION OF THE COMPONENTS OF FISHNET

Layer	Output	Output size
M1	f_{M1}	56×56
T1	f_{T1}	28×28
M2	f_{M2}	28×28
T2	f_{T2}	14×14
M3	f_{M3}	14×14
T3	f_{T3}	7×7
M4	f_{M4}	7×7
F1	f_{F1}	14×14
F2	f_{F2}	28×28
F3	f_{F3}	56×56
P1	f_{P1}	56×56
P2	f_{P2}	28×28
P3	f_{P3}	14×14
P4	f_{P4}	7×7
Class	-	1×1

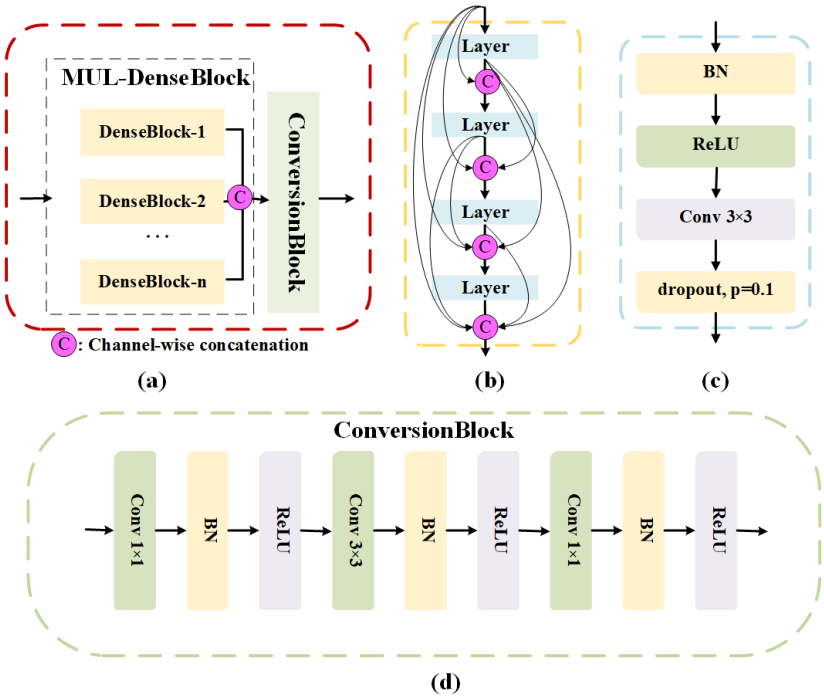


Fig. 4. MUL module: (a) overall architecture, (b) four-layer DenseBlock, and (c) ConversionBlock.

Parallel DenseBlocks with similar depths greatly enhance the network architecture, enabling MUL-DenseBlock to learn richer information about SAR fishing vessels.

As illustrated in Fig. 4(b), a DenseBlock comprises L layers, each producing k feature maps, where k represents the growth rate of the network. Furthermore, each layer implements a nonlinear transformation $H_l(.)$, where l is the index of the layer. $H_l(.)$ denotes a composite function consisting of batch normalization (BN), a rectified linear unit (ReLU) activation function, a convolution (Conv), and a dropout layer with a rate of 0.1, as shown in Fig. 4(c). To strengthen feature propagation, encourage feature reuse, and, thus, extract more global and meaningful features of the fishing vessels, the output of each layer of the DenseBlock is combined with the outputs of all preceding layers in the channel dimension to serve as the input to the next layer. However, the distribution of data obtained by the model after training at each layer is different. Therefore, it is necessary to first normalize the input elements in the next layer using BN, which is implemented as follows:

$$\mathrm{BN}(z) = \lambda \frac{z - \mu(z)}{\sqrt{\sigma^2(z)}} + \gamma \tag{1}$$

where μ denotes the mean of the input feature maps, σ^2 denotes the variance, and λ and γ are network parameters, which are updated during the training process. The BN layer

increases network stability while speeding up the model's convergence to some extent.

A nonlinear transformation is then implemented using the ReLU activation function, and its output is computed as follows:

$$\mathrm{Act}(\mathrm{BN}(z)) = \max(0, \mathrm{BN}(z)). \tag{2}$$

Afterward, the convolution layer extracts feature maps by introducing filters with a size of 3×3 as follows:

$$o[m,h] = \sum_{g=0}^{3}\sum_{f=0}^{3} z(m-g, h-f) * w(g,f) \tag{3}$$

where z denotes the convolutional input, w denotes the filter, and o indicates the output feature map.

Finally, the dropout layer is employed to prevent overfitting in the model. Dropout is a regularization strategy that reduces the reliance of each neuron on others by randomly switching off some neurons. This improves the robustness and generalization ability of the network model.

Therefore, the output of the lth layer $(1 \le l)$ is

$$x_l = H_l\left(\left[x_0, x_1, x_2, \ldots, x_{(l-1)}\right]\right) \tag{4}$$

where $[x_0, x_1, x_2, \ldots, x_{(l-1)}]$ represents the concatenation of the features produced in layers $0, \ldots, l-1$.

The output of the DenseBlock for a single path is

$$y_i = x_0 + H_1(x_0) + H_2(x_1) + \cdots + H_l\left(x_{(l-1)}\right). \tag{5}$$

Then, the final output of the MUL-DenseBlock in the jth MUL module of FishNet can be expressed as

$$Y_j = y_{j1} + y_{j2} + \cdots + y_{jn} \tag{6}$$

where n represents the number of parallel DenseBlocks, that is, the number of paths in the MUL-DenseBlock, and "+" denotes the concatenation operation.

With the above formulas, various feature maps from different channels can be assembled, which is helpful for modeling the feature correlations of the channels and extracting strong features in the following steps.

Here, we set the parameter L of the DenseBlocks in the MUL-DenseBlock of each of the four MUL modules to 6, 12, 24, and 16, respectively, and we set the growth rate k to 16. The above parameters are determined based on the successful experience of DenseNet; the details can be found in [53]. Moreover, for the most important parameter in the MUL-DenseBlock, namely, n, we set it to 3 in accordance with the experimental results (see Section V-A).

The application of Mul-DenseBlock enables adequate extraction of features for the fine-grained classification of fishing vessels while also resulting in a substantial increase in the number of model parameters. Therefore, the ConversionBlock is designed to process the output of the MUL-DenseBlock before it is forwarded to the TransitionBlock. As shown in Fig. 4(c), the ConversionBlock consists of three composite layers containing 1×1, 3×3, and 1×1 convolution filters, respectively. The ConversionBlock not only decreases the number of feature maps (while preserving the spatial dimensions) but also filters information and learns intermediate features.

C. FF Module

CNN models extract features of the target by layer-by-layer abstraction. Most existing CNN-based classification models utilize only the last layer of the network, which in FishNet is only f_{M4} for the SAR fishing vessel classification task. However, higher level features with larger receptive fields can represent richer semantic information but lack finer details due to their coarser resolution. In contrast, lower level features capture sufficient geometric details but offer weaker semantic representations.

SAR fishing vessel classification requires high-level features that retain strong semantic information and low-level features with high resolution that can be mined for more detailed information to differentiate the nuances of different fishing vessel types. For this reason, the FF module fuses the feature maps extracted at different levels to realize feature-level fusion with the purpose of exploring the intrinsic connections between features to produce feature maps with more powerful semantic representations.

Before fusing the high-level feature maps F_{high} $(H \times W \times C_1)$ and the low-level feature maps $F_{\mathrm{low}}(2H \times 2W \times C_2, C_2 < C_1)$, the size and dimensions of both need to be consistent. To minimize the loss of discriminative information on the fishing vessels and ensure that the fused feature maps possess more comprehensive information and greater classification power, the size of the fused feature maps output by the FF module is set to $2H \times 2W \times C_1$. Specifically, the FF module has two branches, as illustrated in Fig. 5.

In one branch, 2-D nearest-neighbor upsampling is applied to F_{high} to retain as much semantic information as possible, resulting in fused feature maps with both robust spatial information and strong semantic information. In addition, a 3×3 convolution is implemented on each upsampled map to reduce the aliasing effect of upsampling. The above process can be expressed as follows:

$$F'_{\mathrm{high}} = \mathrm{BN}\left(\mathrm{Conv}_{3\times3\times c_1}\left(f_{\mathrm{upsample}_2}\left(F_{\mathrm{high}}\right)\right)\right) \tag{7}$$

where f_{upsample_2} represents the operation of 2-D nearest-neighbor upsampling.

In another branch, the first 3×3 convolutional layer increases the dimension of the feature map F_{low} to C_1, and the second convolutional layer further processes the feature map after the number of channels is increased to obtain more feature information. Thus, the series of operations performed on F_{low} can be expressed as follows:

$$F'_{\mathrm{low}} = \mathrm{BN}\left(\mathrm{Conv}_{3\times3\times c_1}\left(\mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{Conv}_{3\times3\times c_1}(F_{\mathrm{low}})\right)\right)\right)\right). \tag{8}$$

In short, the function of one branch is to obtain the characteristic information of the fishing vessel under a larger receptive field from the higher level features, and the function of the other is to increase the dimension of the lower level features while preserving details.

Finally, we consider different methods of combining F'_{low} and F'_{high}: the elementwise product and the elementwise sum. Experimental results show that the elementwise sum provides

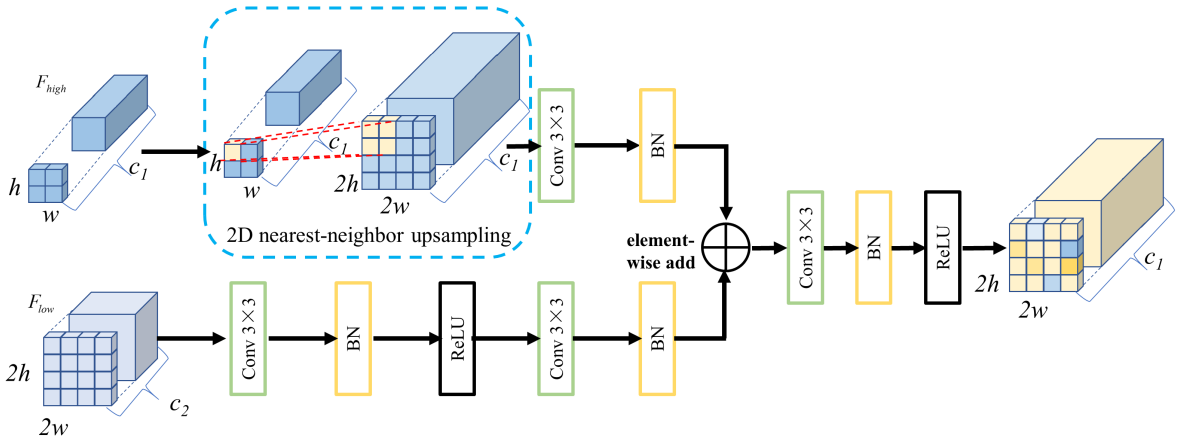


Fig. 5. FF module.

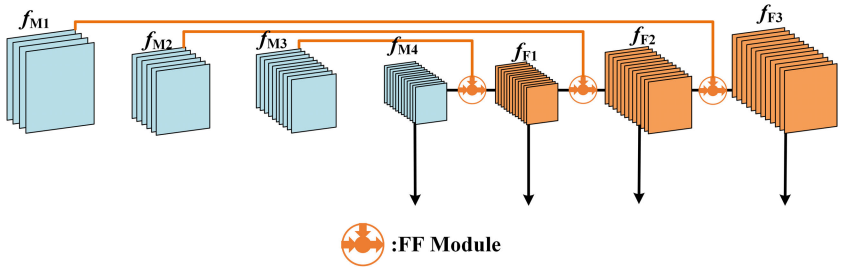


Fig. 6. MFA module.

better accuracy (see Section V-B). Thus, the process of obtaining the final feature maps F' can be expressed as follows:

$$F' = \mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{Conv}_{3\times3\times c_1}\left(F'_{\mathrm{low}} \oplus F'_{\mathrm{high}}\right)\right)\right) \tag{9}$$

where $\oplus$ denotes the elementwise sum.

D. MFA Module

Based on the FF module, a new top-down MFA module has been implemented in the FishNet model. By inserting FF modules between layers in FishNet's basic backbone, the MFA module reduces the loss of feature information during feature extraction due to resolution reduction. The architecture of the MFA module is shown in Fig. 6.

As shown in Fig. 6, the MFA module uses the feature maps $\{f_{M4}, f_{M3}, f_{M2}, f_{M1}\}$ from the top-down pathway that are output from each MUL module at different levels. Unlike in most previous methods [54], the output of each FF module in the MFA module is used as the input to the next FF module in order to extract sufficient feature information to efficiently differentiate SAR fishing vessels. The specific implementation processes are given as follows. The top-level features f_{M4} and the third-level features f_{M3} are simultaneously input into the MFA module, and f_{M4} is upsampled and merged with f_{M3} to obtain the fused features f_{F1}. Then, the second-level features f_{M2} are input into the MFA module and fused with the upsampled f_{F1} to obtain the fused features f_{F2}, and a similar process is repeated to obtain f_{F3}.

By fusing multilevel information from low- and high-level feature maps, the MFA module aggregates both detailed spatial information and contextual semantic information for small fishing vessels, thereby improving the feature description capabilities of FishNet.

E. PCSA Module

Compared to large ships, fishing vessel classification recognition requires addressing the difficulty of feature extraction caused by minor differences while attenuating interference caused by unfavorable factors such as scattered noise in SAR images. By designing the MUL module, FishNet can extract feature maps containing a great deal of information about SAR fishing vessels, and the FF and MFA modules further enrich the feature maps. Therefore, highlighting the more relevant and representative interclass differences of fishing vessels from so much feature information is a crucial task to improve SAR fishing vessel type identification.

Existing studies have proven that feature discrimination can be improved by integrating attention mechanisms into the network model to help capture spatial and channel correlations between features. Therefore, to minimize the interference of redundant information and increase the differentiation of the features used for final classification, we propose the PCSA module. In this module, attention mechanisms are applied to the spatial and channel dimensions simultaneously, guiding FishNet to pay more attention to the subtle interclass differences among fishing vessels while limiting the influence of unnecessary information to achieve a better classification effect for SAR fishing vessels. The architecture of the PCSA module is shown in Fig. 7.

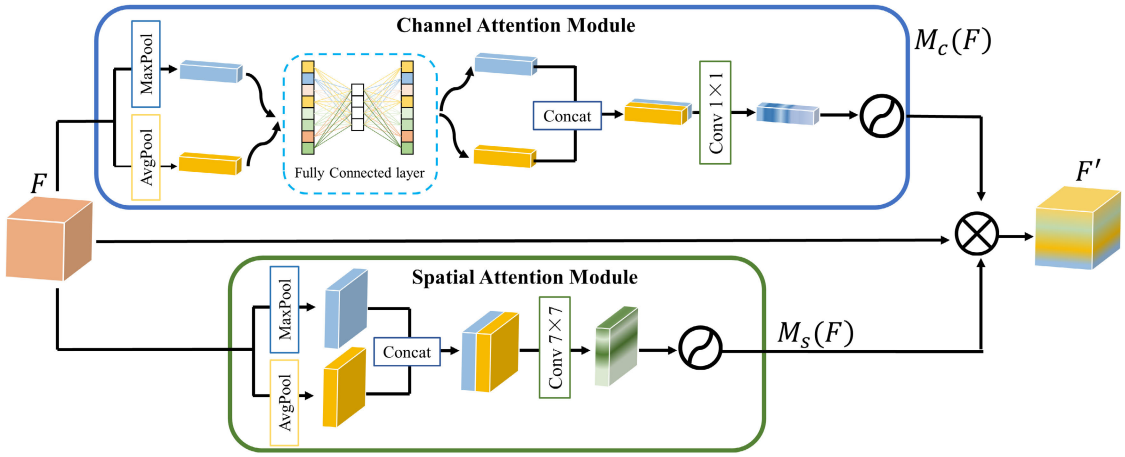


Fig. 7. PCSA module.

In the attention mechanism applied by the previously proposed convolutional block attention module (CBAM) [55], the feature maps are processed by a channel attention (CA) module before being fed to an SA module. However, this can lead to some level of bias in the feature information learned by the later SA module. The deviation caused by the serial connection of the CA and SA modules can lead to instability of the effect of the overall attention module, compromising the classification performance for fine-grained fishing vessels. Therefore, in the proposed PCSA module, as shown in Fig. 7, the CA and SA modules are applied in parallel instead. Consequently, each attention module can learn directly from the original feature maps, ensuring superior processing results for SAR fishing vessel feature maps. Therefore, the PCSA module consists of three parallel channels, with the CA module and SA submodule at its core.

Given an intermediate feature map F as input, a 1-D CA map $M_c(F)$ and a 2-D SA map $M_s(F)$ are inferred by the CA module and the SA module, respectively. Finally, the two attention maps are directly multiplied in an elementwise manner with F to obtain the final refined output F'. The refined feature map F' is computed as follows:

$$F' = M_c(F) \otimes F \otimes M_s(F) \tag{10}$$

where $\otimes$ denotes the elementwise multiplication.

1) CA Module: When the image is processed by a multilayer convolution layer, the feature matrix of multiple channels can be obtained. However, not all channels contain important information. For SAR fishing vessel identification, some channels are extremely valuable, while others may be of little value or even detrimental. In view of this, the PCSA module is designed with a CA module that can enhance the weight of useful information and suppress the weight of useless information in the feature channel.

The CA module exploits the interchannel relationships among the features to generate a CA map, which directs the model's attention to the meaningful content in the input data. In the CA module, the input features are processed separately with average and maximum pooling to calculate the statistical characteristics of the feature map for each channel, which are then sent to a shared fully connected layer. The output feature vectors are then merged through concatenation, and a 1×1 convolutional layer is applied. After the sigmoid activation function is applied, the global weight descriptor is generated. In short, the CA map is computed as follows:

$$M_c(F) = \sigma(\text{Conv}_{1\times1}(\text{concat}[\xi(\text{GAvgPool}(F)), \xi(\text{GMaxPool}(F))])) \tag{11}$$

where $\sigma(.)$ denotes the sigmoid function, MLP(.) is the multilayer perceptron (MLP), $\text{Conv}_{1\times1}(.)$ represents convolution with a kernel size of 1×1, concat(.) represents concatenation in the channel dimension, and GAvgPool(.) and GMaxPool(.) represent the average pooling and the maximum pooling, respectively.

2) SA Module: As previously mentioned, SAR images of fishing vessels are prone to background features caused by interference phenomena such as speckles. Therefore, it is necessary to design the SA module to make the model focus on the areas in the SAR image that is meaningful for identifying fishing vessels. Unlike the CA module, the SA module generates an SA map by applying average and maximum pooling processes along the channel axis using a convolutional layer. In short, the SA map is computed as follows:

$$M_c(F) = \sigma(\text{Conv}_{7\times7}(\text{concat}[\text{GAvgPool}(F), \text{GMaxPool}(F)])). \tag{12}$$

As shown in Fig. 3, the output of PCSA modules, expressed as $\{f_{P1}, f_{P2}, f_{P3}, f_{P4}\}$, are used to obtain the final SAR fishing vessel classification decision.

F. Adaptive Loss Function

The cross-entropy loss function calculates a value between 0 and 1 to indicate whether the model is learning to make accurate predictions. The cross-entropy loss for multiclass classification is defined as follows:

$$\text{Loss} = -\sum_{i=0}^{C-1} y_i \log(p_i) \tag{13}$$

where $p = [p_0, \ldots, p_{C-1}]$ is the predicted distribution, with p_i representing the probability that the sample belongs to

class i; $y = [y_0, \ldots, y_{C-1}]$ is the one-hot encoding of the true labels, with $y_i = 1$ when the sample belongs to class i and $y_i = 0$ otherwise; and C is the total number of samples.

The cross-entropy loss is a measure of the difference between the probability distributions of the real values y and the predicted values p.d In other words, the smaller the loss, the better the model performs. As shown in (13), this loss function is evaluated with the same weight for each class. As a result, training a model on an imbalanced dataset can bias the learning process toward the majority of classes. As a consequence, the model may perform poorly for minority classes, even when the cross-entropy loss reaches a small value during training. Therefore, employing the cross-entropy loss as the loss function for FishNet would be negatively impacted by the class imbalance of the SAR fishing vessel samples, resulting in an adverse effect on the model's learning.

Based on our preliminary experiments, we discovered that the numbers of samples of various types of fishing vessels in each training batch were unbalanced. Furthermore, due to the differences in quantity among the different classes, it is challenging to determine the proportions of the various classes appearing in each minibatch, even after random disruption of all samples. As a solution to these issues, we propose an adaptive loss function defined as follows:

$$\text{Loss} = -\sum_{i=0}^{C-1} y_i e^{-\frac{N_k}{N}} \log(p_i) \tag{14}$$

where $e^{-(N_k/N)}$ is the weight of a sample of class k in the minibatch, N denotes the total number of samples in the minibatch, and N_k denotes the number of samples of class k in the minibatch.

The proposed adaptive loss function can adjust the weight of each sample in each minibatch based on the class to which it belongs, thus balancing the contribution of each class to the overall loss. Consequently, the adaptive loss of each sample is not only dependent on the sample itself but also related to the total number of samples of the same class in the batch. Implementing the adaptive loss function balances the parameter offset in the learning process and mitigates the adverse effect of class imbalance on the training of FishNet, ultimately resulting in better SAR fishing vessel classification performance.

IV. Experiments and Analysis

Our experiments are run on a personal computer with an Intel E5-2678 v3 CPU and an NVIDIA RTX A2000 GPU. Our codes are written based on the PyTorch framework, and CUDA 11.4 and cuDNN 8.2 are used to call the GPU to accelerate training.

To prevent overfitting due to the limited amount of labeled data and improve the generalizability of the model, various data augmentation methods, such as flipping, panning, and rotation at varying angles, are utilized to expand the dataset. In addition, there are fewer seiners in the FishingVesselSAR dataset than trawlers and gillnetters, creating a class imbalance problem. Such imbalance will lead to poor performance for the minority classes in the model. If the same data augmentation methods are to be used for each type of fishing vessel, the imbalance in the dataset would be exacerbated. For this reason, we increase the number of gillnetters and seiners by using geometric transformation methods. After data augmentation, there are 1160 gillnetters, 1080 seiners, and 1268 trawlers in the SAR fishing vessel dataset.

As described in Section III, the SAR fishing vessel images are resized to 56 × 56 pixels for input to FishNet, which is trained from scratch without pretraining. Stochastic gradient descent (SGD) [56] is used to train FishNet for 200 epochs with a batch size of 16. The Nesterov momentum and weight decay parameters of SGD are set to 0.9 and 0.0001, respectively. The learning rate is initially set to 0.0001 and is dynamically adjusted using the cosine annealing method. All the following experiments are conducted with the same parameter settings to ensure fair comparisons. In addition, the experimental results are obtained through fivefold cross-validation, and the mean and standard deviation of each evaluation index are calculated by taking the optimal results of each experiment.

A. Evaluation Indices

In this article, accuracy, precision, recall, and F-score are used to evaluate model performance.

As the most important evaluation metric, accuracy represents the percentage of correct samples relative to the total. Precision is the percentage of true positives among the samples judged to be positive by the classifier. Recall is defined as the ratio of correctly predicted positive cases to total positive cases. The F-score is the harmonic mean of the precision and recall values, with a maximum value of 1 and a minimum value of 0. These metrics are defined as follows [57]:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{15}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{16}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{17}$$

$$\text{F} - \text{score} = \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{18}$$

where TP, FP, TN, and FN represent the numbers of true positives, false positives, true negatives, and false negatives, respectively. Moreover, the confusion matrix [58], the receiver operating characteristic (ROC) curve [59], [60], and the area under the ROC curve (AUC) [61] are used to further evaluate the classification performance.

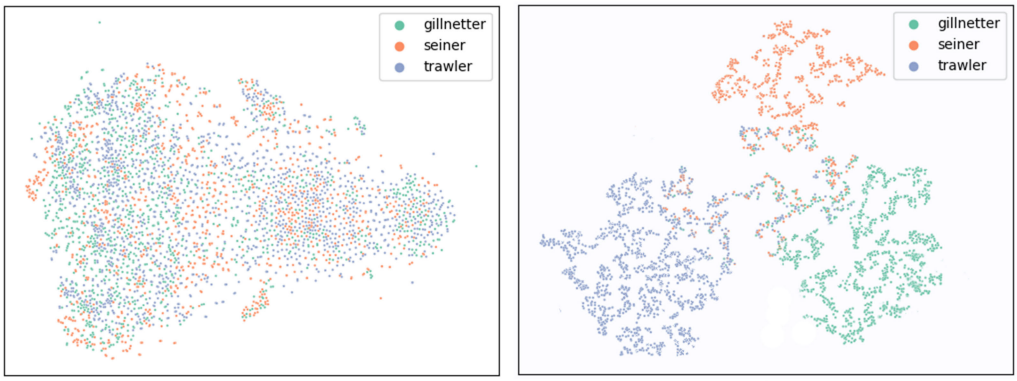
B. Classification Results

Table III shows the performance evaluation results of FishNet. The average accuracy is 89.79%, and the other evaluation indices reach approximately 90%, indicating the superior performance of FishNet for SAR fishing vessel classification. Moreover, FishNet exhibits relatively robust stability of its classification performance, as evidenced by its standard deviation of less than 1.00% for each evaluation index.

To intuitively demonstrate the classification ability of FishNet, we visualize the SAR fishing vessel samples before and

TABLE III
PERFORMANCE EVALUATION OF FISHNET USING THE FIVEFOLD CROSS-VALIDATION PROCEDURE

Fold no.	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
1	90.32	91.13	90.33	90.07
2	88.75	89.67	88.67	88.71
3	89.53	89.36	89.33	89.33
4	89.35	89.34	89.72	89.33
5	91.02	91.33	91.01	91.12
AVG	89.79±0.89	90.17±0.98	89.81±0.90	89.71±0.92



(a) (b)

Fig. 8. Visualizations of (a) input fishing vessel samples and (b) corresponding output of the FishNet model using t-SNE, where samples represented in the same color belong to the same class.

after processing by FishNet using the t-distributed stochastic neighbor embedding (t-SNE) algorithm [62], as shown in Fig. 8.

Fig. 8(a) shows that the visualization results for the input sample set are disordered and chaotic, and the classes are almost impossible to distinguish. In comparison, after being processed by FishNet, samples with the same category labels are closer together, while samples with different category labels tend to be separated from one another [see Fig. 8(b)]. The proposed FishNet model shows superior classification performance in Fig. 8(b).

1) Visualization of Feature Maps: To observe the processing of SAR images by FishNet, we visualize some of the output feature maps of each key innovative module (see Fig. 9). The blue arrows in Fig. 9 indicate the direction of information transfer within the model.

Fig. 9(a) shows an original SAR fishing vessel sample inputted into the model, and Fig. 9(b)–(e) shows some of the feature maps outputted by the MUL modules at different levels, with dimensions of 56 × 56, 28 × 28, 14 × 14, and 7 × 7, respectively. The extracted features in the different feature maps are clearly distinct, but they are all valid in that they focus on the hull rather than the sea surface. In addition, as the network depth increases, the extracted fishing vessel features gradually transition from local texture features to higher levels of abstraction. Although these features are too abstract to be interpreted explicitly, the high accuracy achieved by FishNet reflects its powerful ability to extract features without any human participation.

Fig. 9(f)–(h) shows some of the output feature maps from the FF modules, with dimensions of 56 × 56, 28 × 28, and 14 × 14, respectively. Compared with Fig. 9(d) and (e), the feature maps outputted from FF module F1 retain not only semantic information that is useful for classifying fishing vessels but also discriminative detailed features such as contours and textures. With this combination of high-level semantic and low-level physical information, these features are more representative. Therefore, the model can acquire subtle features from SAR images of fishing vessels that are more representative of the vessel type to achieve better classification results. This is particularly evident in the feature maps outputted by the F3 module [see Fig. 9(f)], which integrates all levels of feature information.

The feature maps in Fig. 9(i)–(l) are obtained by processing the feature maps of Fig. 9(e)–(h), respectively, with the PCSA module. They show that the proposed PCSA module can accurately focus on the target and produce precise attention results.

2) Interclass Euclidean Distances: Inspired by the work of Zeiler and Fergus [63], we describe the classification performance in terms of the Euclidean distances between the feature vectors extracted from different classes of fishing vessels. This way, the improvement in classification performance caused by each key innovative module can be quantified.

Here, we input the fishing vessel samples from the test set for each experiment as specified in Table III into the well-trained FishNet model for that specific experiment. Then, we calculate the Euclidean distance between the feature vectors extracted by each module for each pair of types of SAR fishing vessel samples. The means and standard deviations of the interclass Euclidean distances obtained from the five experiments are shown in Fig. 10. A red or green arrow

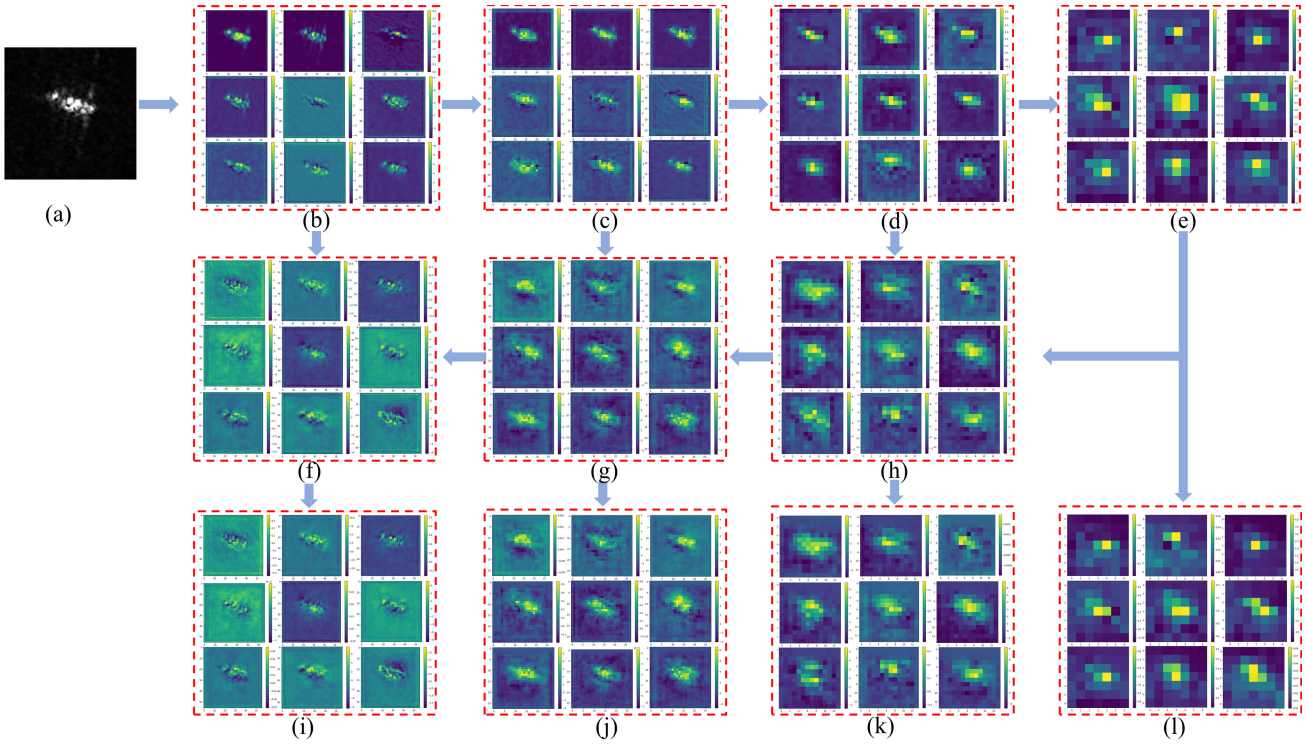


Fig. 9. Visualization of feature maps: (a) input; output feature maps from MUL modules: (b) M1, (c) M2, (d) M3, and (e) M4; output feature maps from FF modules: (f) F3, (g) F2, and (h) F1; and output feature maps from PCSA modules: (i) P1, (j) P2, (k) P3, and (l) P4.

represents an increase or decrease, respectively, in the Euclidean distance calculated in this module compared to the previous module. This is done to facilitate the evaluation of changes in the classification performance of the model. When there is more than one upstream module, as shown in Fig. 10(e), only the smallest change value is displayed for each distance.

The MUL module is proposed to extract more sufficient target features. As seen in Fig. 10(a)–(c), with the deepening of the backbone, the three interclass Euclidean distances gradually increase. However, they decrease after the deepest M4 module [see Fig. 10(d)]. This phenomenon may possibly result from the fact that, as the receptive field of the model increases with depth, detailed information such as the structures of fishing vessels may vanish entirely in the extracted feature vectors, reducing the ability to classify fishing vessels. This problem can be solved to some extent by using the FF and MFA modules, as evidenced by the gradual increases in the three interclass Euclidean distances for the feature vectors output by the F1 [see Fig. 10(g)], F2 [see Fig. 10(f)], and F3 [see Fig. 10(e)] modules. By integrating features at different levels, the FF and MFA modules enhance the information richness of the features considered for classification, thus making them more discriminative.

The PCSA module is designed to suppress background features caused by interference phenomena such as speckle noise while enhancing the representativeness of useful features. In FishNet, the PCSA modules P1, P2, P3, and P4 are used to process the feature maps output from F3, F2, F1, and M4, respectively. As seen in Fig. 10(h)–(k), the Euclidean distances are increased by 2.30%–11.94% compared to the previous module, thus proving the effectiveness of the PCSA module.

Finally, all feature vectors inputted into the classification module are integrated into one output for the final fishing vessel classification. In Fig. 10(l), the Euclidean distances between the three classes as calculated from the integrated feature vectors reach their maximum values, indicating that FishNet can achieve superior classification performance for gillnetters, seiners, and trawlers.

C. Comparisons With Advanced Models

Two experiments are conducted to comprehensively verify FishNet's superior classification performance. One comparison is made with advanced models from the DL community that has achieved remarkable success in image classification. Another comparison is made with advanced models from the SAR target classification community.

1) Comparison With Advanced Models From the DL Community: Table IV shows the performance comparison of the proposed model with 27 advanced DL models. The other DL models are implemented based on the original proposals. The classification accuracies of all the models in Table IV exceed 60%, which indicates the feasibility of identifying fishing vessels from SAR images.

Compared to other DL models with excellent classification performance in other fields, FishNet, which is specifically designed to address the difficulties encountered in SAR fishing vessel classification, such as the small size of the targets

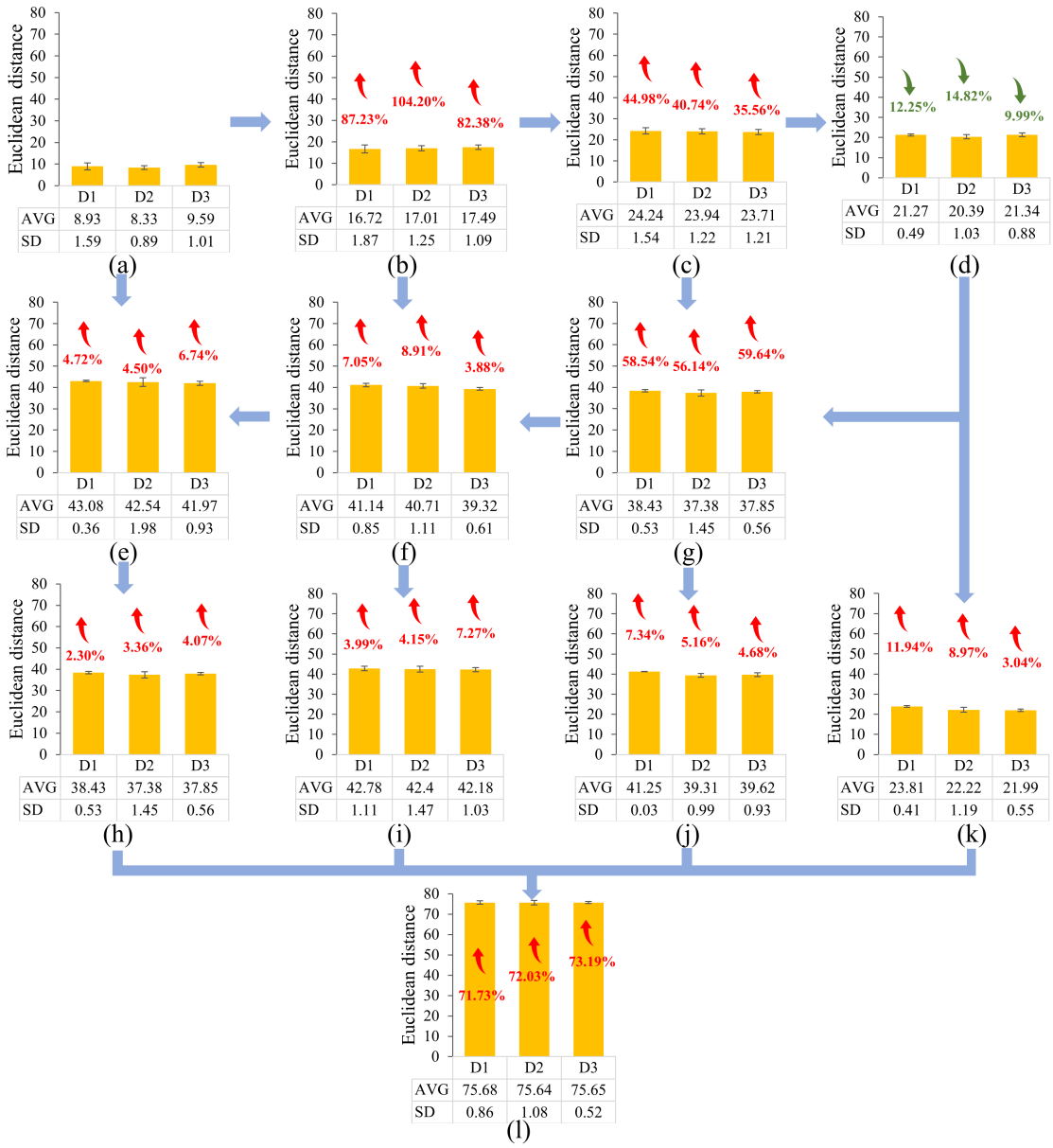


Fig. 10. Interclass Euclidean distances between the feature vectors output by MUL modules: (a) M1, (b) M2, (c) M3, and (d) M4; FF modules: (e) F3, (f) F2, and (g) F1; PCSA modules: (h) P1, (i) P2, (j) P3, and (k) P4; and (l) feature vectors used for final fishing vessel classification in the classification module. *** "D1" denotes the interclass Euclidean distance between gillnetters and seiners, "D2" denotes the interclass Euclidean distance between gillnetters and trawlers, and "D3" denotes the interclass Euclidean distance between seiners and trawlers.

and the subtle interclass differences between classes, achieves an accuracy of 89.79%, which is 6.77% higher than the second-best model and much higher than those of the others. In addition, the overall precision, recall, and F-score values outperform those of currently popular advanced DL models.

2) Comparison With Advanced Models From the SAR Target Classification Community: To the best of our knowledge, no previous research has been conducted on the type classification of fishing vessels in SAR images. Several DL models have been widely used in SAR target classification and have produced excellent classification results. Their performance for SAR fishing vessel classification is shown in Table IV. Therefore, we choose six advanced methods specially designed for SAR target classification to compare with our FishNet model in this section.

In Table V, A-ConvNets [74], capsule [75], and Zhang et al. [76] are small-sample classification methods based on the Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset; GSESCNNs [77], Hou et al. [78], and HOG-ShipCLSNet [45] are ship classification methods developed based on the OpenSARShip and FUSAR-Ship datasets, which focus on large ships such as tankers.

As shown in Table V, regardless of whether the methods are based on the large-ship or MSTAR datasets, their classification results for SAR fishing vessels are much lower than those of FishNet, with the maximum difference in accuracy reaching 42.20%. There are two main reasons why the other models may fail to perform well in classifying fishing vessels. On the one hand, these models are smaller than the DL models in

TABLE IV
PERFORMANCE COMPARISON WITH 27 ADVANCED DL MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
ShuffleNet_v2_x1_0 [64]	71.08±0.59	70.94±1.25	69.32±0.91	70.01±0.54
ShuffleNet_v2_x2_0 [64]	74.56±0.57	74.34±0.52	73.07±1.06	71.39±4.61
ShuffleNet_v2_x1_5 [64]	73.32±0.78	73.58±0.55	71.56±0.95	72.21±0.92
ShuffleNet_v2_x0_5 [64]	65.18±0.41	65.6±0.74	62.84±0.71	63.26±0.29
ResNet_34 [65]	79.48±1.26	79.26±1.21	76.73±3.78	78.73±1.32
ResNet_50 [65]	79.19±2.89	79.87±2.34	80.02±2.15	79.33±2.71
ResNet_101 [65]	81.48±1.90	81.46±2.58	80.33±1.90	80.67±2.01
RegNet [66]	45.84±0.32	47.68±0.25	42.68±0.41	42.22±2.74
AlexNet [67]	66.56±1.22	66.83±1.611	66.92±1.53	66.26±1.19
GoogLeNet [68]	68.46±1.12	69.38±2.01	68.46±0.87	68.34±1.13
MobileNetV1 [69]	65.96±1.06	67.73±2.29	63.66±1.08	63.39±1.66
MobileNetV2 [70]	66.58±1.82	68.93±1.14	64.73±2.30	64.67±2.82
MobileNetV3_small [71]	68.52±1.49	68.66±1.18	67.80±1.66	67.93±1.78
MobileNetV3_large [71]	72.58±1.17	73.06±1.23	71.93±1.19	71.85±1.03
EfficientNet_b0 [72]	74.20±0.37	74.23±0.31	71.74±0.29	72.01±0.21
EfficientNet_b1 [72]	75.03±0.21	75.34±0.41	72.38±0.18	72.60±0.30
EfficientNet_b2 [72]	76.94±1.31	77.54±1.63	76.22±1.43	76.32±1.42
EfficientNet_b3 [72]	75.36±0.34	76.54±0.59	73.86±0.54	73.72±0.44
EfficientNet_b4 [72]	76.26±0.29	76.58±0.94	74.82±0.73	74.88±0.27
EfficientNet_b5 [72]	74.36±1.14	74.88±1.19	73.18±0.61	73.00±0.94
EfficientNetV2_s [73]	76.30±1.15	76.73±0.76	73.93±1.23	73.93±1.29
EfficientNetV2_m [73]	75.92±0.73	76.68±0.41	73.73±0.96	73.46±1.22
EfficientNetV2_l [73]	76.72±0.74	77.33±0.52	74.20±0.77	74.33±0.97
DenseNet121 [53]	82.06±1.22	82.73±1.20	81.65±1.14	81.54±0.87
DenseNet161 [53]	82.14±0.91	82.13±0.98	81.39±0.75	81.66±0.79
DenseNet169 [53]	83.02±1.12	74.02±1.59	73.39±0.77	73.37±0.89
DenseNet201 [53]	82.84±0.67	82.94±0.75	82.66±1.19	81.92±0.73
FishNet	89.79±0.89	90.17±0.98	89.81±0.90	89.71±0.92

TABLE V
PERFORMANCE COMPARISON WITH ADVANCED METHODS FROM THE SAR TARGET CLASSIFICATION COMMUNITY

Method	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
A-ConvNets	61.55±1.33	62.72±1.41	62.04±1.22	61.78±0.84
capsule	62.31±1.62	62.02±1.02	62.03±1.12	61.79±0.95
Zhang et al.	68.97±1.17	71.64±1.10	68.30±1.04	67.35±0.70
GSESCNNs	62.48±0.71	62.07±0.95	62.06±1.03	60.28±1.02
Hou et al.	59.60±0.43	60.30±0.83	60.66±1.01	59.70±1.06
HOG-ShipCLSNet	47.59±1.67	47.73±0.67	47.02±1.24	47.07±0.11
FishNet	89.79±0.89	90.17±0.98	89.81±0.90	89.71±0.92

Table IV in terms of both depth and width, and as a result, they fail to extract sufficient effective discriminative features from SAR fishing vessels to capture the small interclass differences, resulting in poor classification. On the other hand, most of them do not utilize all of the features extracted in previous layers and, instead, rely solely on the features extracted in the last layer.

Furthermore, the HOG-ShipCLSNet model fuses traditional manual features with CNN-based abstract features, which is a popular approach in SAR ship classification research, but its accuracy reaches only 47.59% for the classification of fishing vessels. To explore possible reasons for this, we select a cargo ship and a gillnetter from the OpenSARShip dataset and the FishingVesselSAR dataset, respectively, and extract and visualize their HOG features, as shown in Fig. 11. Compared to the gillnetter [see Fig. 11(c)], the cargo ship [see Fig. 11(a)] is larger and has a more pronounced superstructure. As shown in Fig. 11(b), the HOG features extracted from the cargo ship can describe the shapes and edges of local targets, which is beneficial in identifying ship types. Therefore, the addition of HOG features to the HOG-ShipCLSNet model results in more effective classification of SAR ships in Zhang et al.'s work [45]. However, the HOG features extracted from the gillnetter are unable to represent its geometric features [see Fig. 11(d)]. A CNN-based model incorporating such HOG features might be less effective. As a result, a DL model that

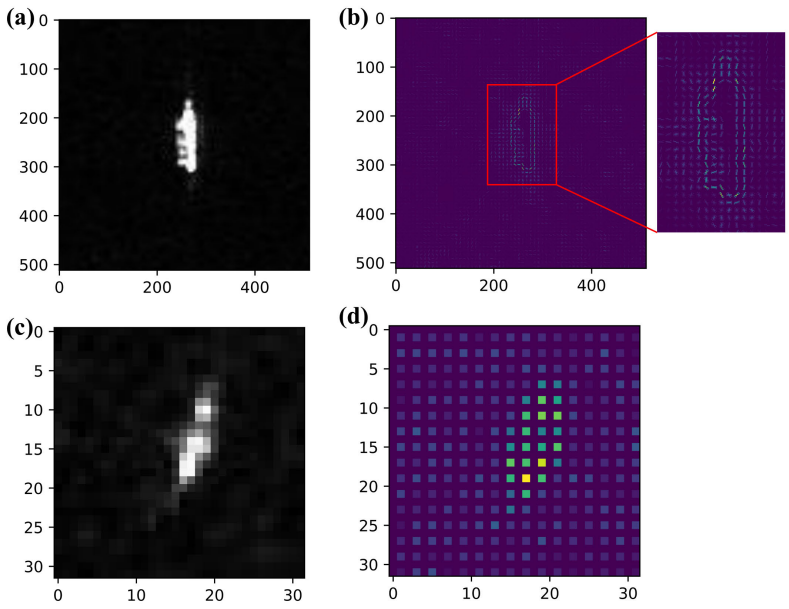


Fig. 11. Visualizations of a cargo ship in (a) OpenSARShip and (b) its HOG features and of a gillnetter in (c) FishingVesselSAR and (d) its HOG features.

TABLE VI
PERFORMANCE COMPARISON WITH DIFFERENT VALUES OF n

n	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
1	86.47±0.69	87.85±0.87	85.55±0.95	85.38±1.29
2	87.57±1.66	88.92±1.03	88.89±1.39	88.88±1.75
3	89.79±0.89	90.17±0.98	89.81±0.90	89.71±0.92
4	89.44±1.06	88.45±1.28	89.01±1.02	88.92±0.77
5	88.08±1.35	87.58±1.74	88.33±1.88	88.09±2.11

can freely extract features from targets may be a more effective method of achieving high-precision classification for fishing vessels with unclear features.

3) ROC and AUC Metrics: An objective evaluation of the different models' classification performance is presented in this section in terms of the ROC curve and AUC. Fig. 12 shows the ROC curves and the corresponding AUCs of the various classification methods. The performance of a model is better when the ROC curve is steeper. The AUC is defined as the area under the ROC curve and above the coordinate axis. In general, the larger the AUC is, the better the classification ability. Compared with the others, the proposed FishNet model achieves a higher true positive rate (TPR) while maintaining a lower false positive rate (FPR) [see Fig. 12(a)], resulting in the highest AUC value [see Fig. 12(b)]. This finding confirms its excellent classification performance for fishing vessels.

V. ABLATION STUDY

A. Ablation Study on the MUL Module

We perform a series of experiments to verify the effectiveness of the MUL module and to determine the influence of the hyperparameter n. The results are listed in Table VI. Note that, when $n = 1$, the MUL module is replaced with a primitive DenseBlock. This means that the backbone network of FishNet consists of DenseBlocks and TransitionBlocks instead of MUL modules and TransitionBlocks.

As seen in Table VI, FishNet with the MUL module ($n > 1$) achieves better classification performance for SAR fishing vessels, with the classification accuracy being improved by at least 1.10%. When $n = 3$, the accuracy is improved by 3.32% to a maximum value of 89.79%, thus proving the effectiveness of the MUL module.

Theoretically, the feature extraction ability of the MUL module should be further improved as n increases. However, the classification accuracy of FishNet for SAR fishing vessels decreases instead when $n > 3$, as shown in Table VI. As mentioned above, due to the small size of SAR fishing vessels and the subtle interclass differences, the effective features that the model can extract are limited. It appears that the degradation in classification accuracy when n exceeds 3 indicates that, at this point, the feature extraction capacity of the model is already excessive for SAR fishing vessel samples. In this case, numerous irrelevant and redundant features may be extracted by the model, which can adversely affect the final classification performance. Therefore, for the SAR fishing vessel classification task, the optimal value of the parameter n in the FishNet model is 3.

B. Ablation Study on the FF Module and the MFA Module

To verify the effectiveness of the FF module and the MFA module and investigate the influence of some detailed network settings, we perform two groups of experiments. Experiment 1

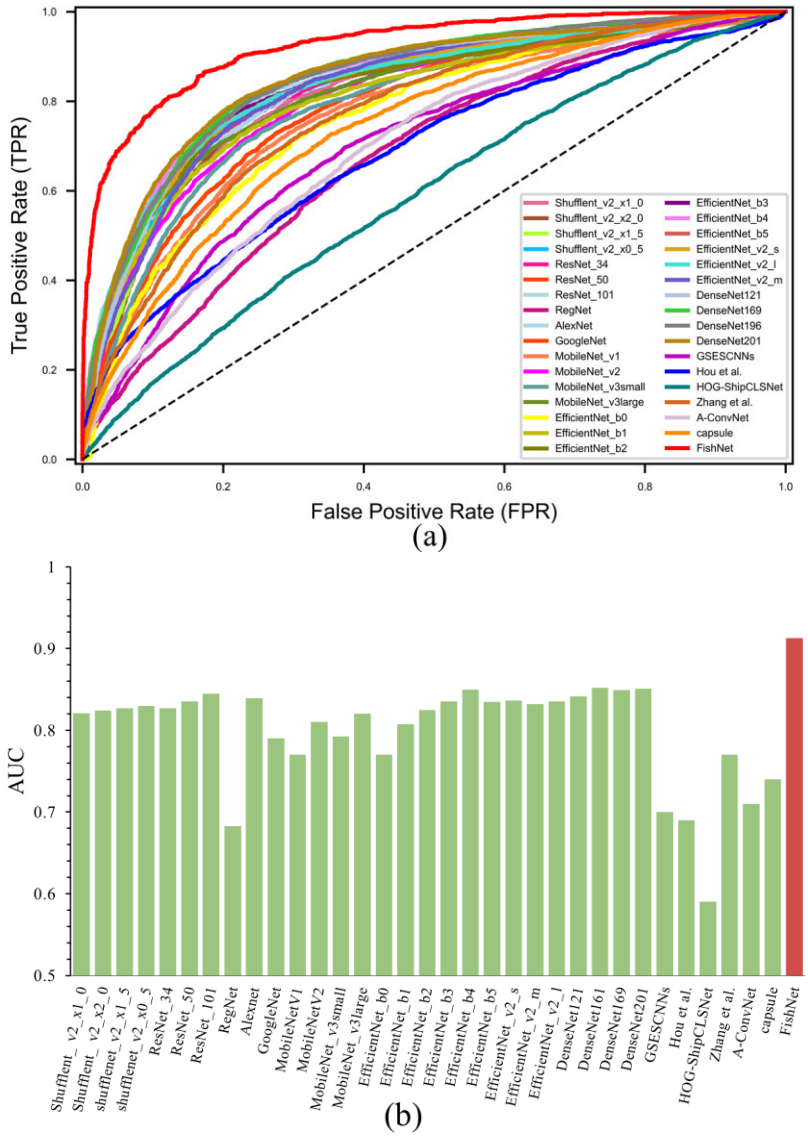


Fig. 12. (a) ROC curves and (b) corresponding AUC values of the different methods.

TABLE VII
PERFORMANCE COMPARISON OF DIFFERENT COMBINATION METHODS

Combination method	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
Elementwise product	87.54±0.76	87.74±2.08	87.67±0.41	87.60±0.87
Elementwise sum	89.79±0.89	90.17±0.98	89.81±0.90	89.71±0.92

is conducted to verify the effects of using different feature combination methods in the FF module. Experiment 2 is conducted to study the effects of different levels of features on classification accuracy.

1) Experiment 1: Effectiveness of Different Combination Methods: Table VII compares the performance achieved when using different combination methods in the FF module. The classification accuracy of the model is higher when the elementwise sum method is used to combine features of different levels. This result suggests that the elementwise sum combination method is more beneficial for fishing vessel classification in SAR images.

2) Experiment 2: Effectiveness of the FF Module and the MFA Module: Table VIII shows the classification performance results obtained using different levels of features in the FishNet model. In case 1, only the f_{M4} feature maps are used, which is equivalent to not using the MFA module and the FF module. As seen in Table VIII, the accuracy increases with an increasing number of features at different levels.

In case 4, FishNet adopts the full MFA module, using the features from all levels for the final classification. In this case, the model achieves the highest classification accuracy, which is approximately 15.67% higher than that in case 1. Therefore, the FF module and the MFA module have a significant

TABLE VIII
PERFORMANCE COMPARISON OF DIFFERENT NUMBERS OF FEATURE LEVELS

	f_{M1}	f_{M2}	f_{M3}	f_{M4}	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
Case 1	×	×	×	√	83.31±0.88	83.50±1.37	82.60±0.89	82.44±1.09
Case 2	×	×	√	√	84.53±1.23	83.62±1.59	84.82±1.06	84.52±1.19
Case 3	×	√	√	√	87.31±0.68	87.42±0.88	87.14±1.02	86.84±0.92
Case 4	√	√	√	√	89.79±0.89	90.17±0.98	89.81±0.90	89.71±0.92

TABLE IX
PERFORMANCE COMPARISON OF DIFFERENT ATTENTION MECHANISMS

Attention type	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
×	85.85±0.17	85.83±0.80	85.28±0.41	85.31±0.30
SENet	87.63±0.74	87.70±0.92	87.60±0.79	87.69±0.85
ECA	86.11±0.69	86.51±0.77	85.61±0.92	85.71±0.90
GAM	86.25±1.22	86.30±1.49	85.54±1.19	85.71±1.22
CBAM	88.19±0.65	89.00±0.91	88.01±0.71	88.13±0.61
PCSA	89.79±0.89	90.17±0.98	89.81±0.90	89.71±0.92

TABLE X
PERFORMANCE COMPARISON OF DIFFERENT LOSS FUNCTIONS

Loss function	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
Cross-entropy loss	88.71±0.44	89.33±0.16	88.21±0.18	88.36±0.18
Focal loss	84.71±0.81	84.31±0.51	84.08±0.03	84.17±0.41
Adaptive loss	89.79±0.89	90.17±0.98	89.81±0.90	89.71±0.92

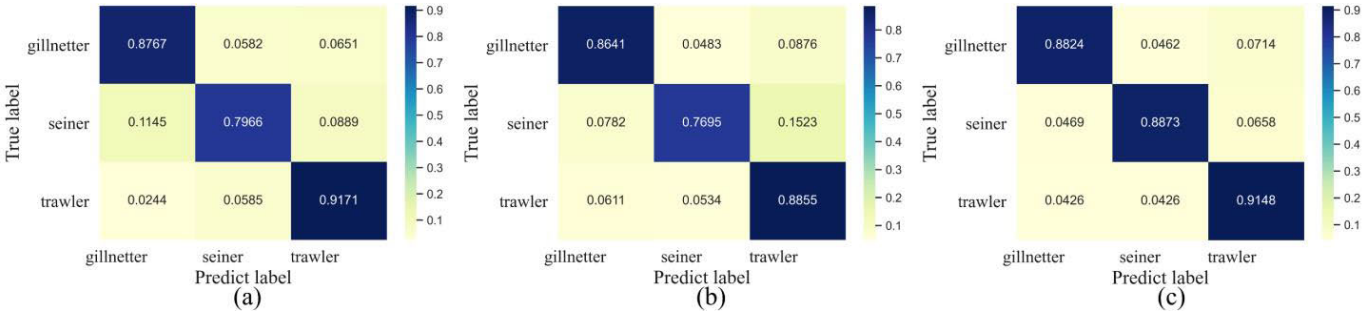


Fig. 13. Normalized confusion matrices of SAR fishing vessel classification with FishNet using (a) cross-entropy loss, (b) focal loss, and (c) adaptive loss.

impact on improving the performance of SAR fishing vessel classification.

C. Ablation Study on the PCSA Module

We perform ablation studies to verify the effectiveness of the PCSA module and to investigate the effects of other types of attention mechanisms on SAR fishing vessel classification, including the commonly used mechanisms such as SENet [78], ECA [79], GAM [80], and CBAM.

As shown in Table IX, the classification accuracy for fishing vessels improves with the addition of any attention mechanism. Among them, the model using the PCSA module has the highest accuracy, with an increase of 3.94%.

D. Ablation Study on the Adaptive Loss Function

This section compares the proposed adaptive loss with previous loss functions, namely, the cross-entropy loss and the focal loss [81]. Table X shows the accuracy results obtained with the different loss functions. Our proposed adaptive loss achieves higher accuracy than the others. Furthermore, for SAR fishing vessel classification, the model that uses focal loss has the lowest accuracy. One possible reason is that the focal loss function is used to address a serious imbalance between positive and negative samples in object detection. However, the class imbalance is less of a problem in SAR fishing vessel classification than in object detection. Therefore, using focal loss as the loss function of FishNet results in decreased model accuracy.

To further confirm the adaptive loss's ability to address the class imbalance problem in SAR fishing vessel classification, Fig. 13 shows the normalized confusion matrices of a single experimental realization for each loss function. In Fig. 13, the diagonal elements of the normalized confusion matrices are significantly greater than the other elements, indicating that most fishing vessels can be classified correctly. However, the

accuracy for seiners is generally lower than that for other types, as shown in Fig. 13(a) and (b). With adaptive loss, FishNet achieves high classification accuracy for all three types [see Fig. 13(c)]. This indicates that the adaptive loss reduces the impact of interclass imbalance on the classification performance.

VI. Conclusion

CNN-based methods have achieved satisfactory classification results for large ships with distinctive features and significant interclass differences. However, the fine-grained classification of fishing vessels in SAR images remains a challenging task.

The interclass differences among fishing vessels caused by the fishing gear and supporting facilities deployed on board are relatively minor. Moreover, due to their small size, fishing vessels in SAR images are highly susceptible to interference phenomena such as speckle noise, which makes capturing such minor interclass differences more difficult. Therefore, under the above unfavorable conditions, a powerful model capable of extracting key features that can effectively characterize the interclass differences of fishing vessels is required to achieve high classification accuracy for SAR fishing vessels. In addition, an accurate and annotated sample dataset is a prerequisite for research on SAR fishing vessel classification. However, determining fishing vessel type directly from AIS messages is not possible, which makes constructing a fine-grained dataset more complicated, difficult, and time-consuming.

In this article, we first propose a process framework for labeling SAR fishing vessels and devote considerable effort to image collection, data processing, and category labeling. An SAR fishing vessel dataset containing 116 gillnetters, 72 seiners, and 181 trawlers is constructed, providing suitable data for the realization of SAR fishing vessel classification. Due to the relatively small number of fishing vessels and the difficulty of confirming their types, the number of samples in FishingVesselSAR is relatively small compared to other vessel datasets. However, it has advantages in terms of the resolution of samples, making it more conducive to the extraction of feature information for fishing vessel classification in SAR images.

Based on FishingVesselSAR, FishNet, a novel DL model, is proposed for the classification of SAR fishing vessels. First, to solve the problem that minor interclass differences result in few useful features extracted by a generic classification model, leading to poor classification results, we propose the MUL module to achieve sufficient extraction of valuable feature information for fishing vessels. Second, to address the classification challenges presented by fishing vessels' small size, the FF module is proposed for fusing feature maps of different levels. Based on this, the MFA module for aggregating multilevel features is developed. As a result, low-level feature maps with detailed texture information and high-level feature maps with strong semantic classification information can be efficiently integrated, producing feature maps with more powerful feature representation capabilities. Third, the PCSA module is designed to adaptively recalibrate the fused features to highlight critical features and suppress invalid interference information. Fourth, the problem of class imbalance in SAR fishing vessel classification is addressed from two perspectives. Various data augmentation strategies are employed at the data level to expand the dataset, preventing model overfitting while decreasing data imbalance. At the algorithm level, an adaptive loss function that can adaptively adjust the loss of each sample during minibatch training is proposed to further solve the class imbalance problem in SAR fishing vessel classification.

Based on the results of the fivefold cross-validation experiments, FishNet's classification accuracy reaches 89.79%, which is 6.77%-47.59% higher than the accuracy of 27 DL models in the field of image classification and six advanced methods in the field of SAR target classification. Moreover, other evaluation indices, such as precision, recall, and F-score, are also superior to those of these 33 modern methods. In addition, by visualizing the feature maps and calculating the interclass Euclidean distances of the feature vectors, we evaluate and demonstrate qualitatively and quantitatively the improvement in classification performance enabled by each innovative module. Finally, for each improvement in FishNet, we present sufficient ablation studies to verify its effectiveness.

To the best of our knowledge, this study is the first to address the identification of fishing vessel types in SAR images, which is an essential step toward using SAR images to improve fishery monitoring and management and, thus, help enhance the ability to combat overfishing. For better fishing vessel classification, the FishNet model is designed with a deeper and wider network structure. This has resulted in an increased number of trainable parameters, leading to a longer training process compared to models with simpler architectures. However, once the FishNet model is trained, it enables quick and efficient prediction, making it a valuable tool for fisheries monitoring. To shorten training time without sacrificing the model's performance, we will keep improving the algorithm for the FishNet model in subsequent studies. Moreover, due to the limited data available, the applicability of our research results is presently confined to trawlers, gillnetters, and seiners in C-band SAR images. To accommodate a broader range of fishing vessel classifications, it is necessary to acquire data with different radar frequencies and conduct further studies on the identification of additional types of fishing vessels. In the future, we will continue to collect additional SAR fishing vessel samples while developing new methods for effective SAR fishing vessel type identification.

Acknowledgment

The authors would like to thank the National Satellite Ocean Application Service for providing the Gaofen-3 (GF-3) data (https://osdds.nsoas.org.cn). They thank American Journal Experts (https://china.aje.com/cn) for editing the English text of this article.

References

[1] *The State of World Fisheries and Aquaculture 2020*, Sustainability in Action, FAO, Rome, Italy, 2020, doi: 10.4060/ca9229en.

[2] E. R. Selig et al., "Revealing global risks of labor abuse and illegal, unreported, and unregulated fishing," *Nature Commun.*, vol. 13, no. 1, pp. 1–11, Apr. 2022, doi: 10.1038/s41467-022-28916-2.

[3] U. R. Sumaila et al., "Illicit trade in marine fish catch and its effects on ecosystems and people worldwide," *Sci. Adv.*, vol. 6, no. 9, Feb. 2020, Art. no. eaaz3801, doi: 10.1126/sciadv.aaz380.

[4] M. Andriamahefazafy, G. Touron-Gardic, A. March, G. Hosch, M. L. D. Palomares, and P. Failler, "Sustainable development goal 14: To what degree have we achieved the 2020 targets for our oceans?" *Ocean Coastal Manage.*, vol. 227, Aug. 2022, Art. no. 106273, doi: 10.1016/j.ocecoaman.2022.106273.

[5] R. B. Cabral et al., "Rapid and lasting gains from solving illegal fishing," *Nature Ecol. Evol.*, vol. 2, no. 4, pp. 650–658, Mar. 2018, doi: 10.1038/s41559-018-0499-1.

[6] D.A. Kroodsma et al., "Tracking the global footprint of fisheries," *Science*, vol. 359, no. 6378, pp. 904–907, Feb. 2018, doi: 10.1126/science.aao5646.

[7] H. Huang, F. Hong, J. Liu, C. Liu, Y. Feng, and Z. Guo, "FVID: Fishing vessel type identification based on VMS trajectories," *J. Ocean Univ. China*, vol. 18, no. 2, pp. 403–412, May 2018, doi: 10.1007/s11802-018-3717-1.

[8] Y. Guan et al., "Identification of fishing vessel types and analysis of seasonal activities in the Northern South China sea based on AIS data: A case study of 2018," *Remote Sens.*, vol. 13, no. 10, p. 1952, May 2021, doi: 10.3390/rs13101952.

[9] L. Pipanmekaporn, and S. Kamonsantiroj, "A deep learning approach for fishing vessel classification from VMS trajectories using recurrent neural networks," in *Proc. Int. Conf. Hum. Interact. Emerg. Technol.*, vol. 1152. Lausanne, Switzerland, Apr. 2020, pp. 5511–5533, doi: 10.1007/978-3-030-44267-5_20.

[10] D. J. McCauley et al., "Ending hide and seek at sea," *Science*, vol. 351, no. 6278, pp. 1148–1150, Mar. 2016, doi: 10.1126/science.aad5686.

[11] A. N. Tassetti, C. Ferrà, and G. Fabi, "Rating the effectiveness of fishery-regulated areas with AIS data," *Ocean Coastal Manage.*, vol. 175, pp. 90–97, Jun. 2019, doi: 10.1016/j.ocecoaman.2019.04.005.

[12] C. Ferrà, A. N. Tassetti, E. N. Armelloni, A. Galdelli, G. Scarcella, and G. Fabi, "Using AIS to attempt a quantitative evaluation of unobserved trawling activity in the Mediterranean Sea," *Frontiers Mar. Sci.*, vol. 7, Nov. 2020, Art. no. 580612, doi: 10.3389/fmars.2020.580612.

[13] G.-C. Sun, Y. Liu, J. Xiang, W. Liu, M. Xing, and J. Chen, "Spaceborne synthetic aperture radar imaging algorithms: An overview," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 161–184, Mar. 2022, doi: 10.1109/MGRS.2021.3097894.

[14] B. Snapir, T. Waine, and L. Biermann, "Maritime vessel classification to monitor fisheries with SAR: Demonstration in the North Sea," *Remote Sens.*, vol. 11, no. 3, p. 353, Feb. 2019, doi: 10.3390/rs11030353.

[15] J. Park et al., "Illuminating dark fishing fleets in North Korea," *Sci. Adv.*, vol. 6, no. 30, Jul. 2020, Art. no. eabb1197, doi: 10.1126/sciadv.abb1197.

[16] A. Galdelli, A. Mancini, C. Ferrà, and A. N. Tassetti, "A synergic integration of AIS data and SAR imagery to monitor fisheries and detect suspicious activities," *Sensors*, vol. 21, no. 8, p. 2756, Apr. 2021, doi: 10.3390/s21082756.

[17] F. Paolo et al., "XView3-SAR: Detecting dark fishing activity using synthetic aperture radar imagery," 2022, *arXiv:2206.00897*.

[18] D. A. Kroodsma, T. Hochberg, P. B. Davis, F. S. Paolo, R. Joo, and B. A. Wong, "Revealing the global longline fleet with satellite radar," *Sci. Rep.*, vol. 12, no. 1, Dec. 2022, Art. no. 21004, doi: 10.1038/s41598-022-23688-7.

[19] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010, doi: 10.1109/TIT.2010.2040894.

[20] X. Xing, K. Ji, H. Zou, W. Chen, and J. Sun, "Ship classification in TerraSAR-X images with feature space based sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1562–1566, Nov. 2013, doi: 10.1109/LGRS.2013.2262073.

[21] M. Jiang, X. Yang, Z. Dong, S. Fang, and J. Meng, "Ship classification based on superstructure scattering features in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 616–620, May 2016, doi: 10.1109/LGRS.2016.2514482.

[22] F. Wu, C. Wang, S. Jiang, H. Zhang, and B. Zhang, "Classification of vessels in single-pol COSMO-SkyMed images based on statistical and structural features," *Remote Sens.*, vol. 7, no. 5, pp. 5511–5533, May 2015, doi: 10.3390/rs70505511.

[23] G. Margarit, J. J. Mallorqui, and X. Fabregas, "Single-pass polarimetric SAR interferometry for vessel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3494–3502, Nov. 2007, doi: 10.1109/TGRS.2007.897437.

[24] H. Zhang, X. Tian, C. Wang, F. Wu, and B. Zhang, "Merchant vessel classification based on scattering component analysis for COSMO-SkyMed SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1275–1279, Nov. 2013, doi: 10.1109/LGRS.2012.2237377.

[25] E. Salerno, "Using low-resolution SAR scattering features for ship classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–4, 2022, doi: 10.1109/LGRS.2022.3183622.

[26] H. Lin, S. Song, and J. Yang, "Ship classification based on MSHOG feature and task-driven dictionary learning with structured incoherent constraints in SAR images," *Remote Sens.*, vol. 10, no. 2, p. 190, Jan. 2018, doi: 10.3390/rs10020190.

[27] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893, doi: 10.1109/CVPR.2005.177.

[28] C. Zhu, H. Zhou, R. Wang, and J. Guo, "A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3446–3456, Sep. 2010, doi: 10.1109/TGRS.2010.2046330.

[29] J. Antelo, G. Ambrosio, J. Gonzalez, and C. Galindo, "Ship detection and recognitionin high-resolution satellite images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2009, p. 514, doi: 10.1109/IGARSS.2009.5417426.

[30] E. Akagunduz, "Scale invariant sillhouette features," in *Proc. 21st Signal Process. Commun. Appl. Conf. (SIU)*, Apr. 2013, pp. 1–4, doi: 10.1109/SIU.2013.6531586.

[31] A. Agrawal, P. Mangalraj, and M. A. Bisherwal, "Target detection in SAR images using SIFT," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2015, pp. 90–94, doi: 10.1109/ISSPIT.2015.7394426.

[32] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017, doi: 10.1109/MGRS.2017.2762307.

[33] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2017, doi: 10.1038/nature14539.

[34] C. Bentes, D. Velotto, and B. Tings, "Ship classification in TerraSAR-X images with convolutional neural networks," *IEEE J. Ocean. Eng.*, vol. 43, no. 1, pp. 258–266, Jan. 2018, doi: 10.1109/JOE.2017.2767106.

[35] Y. Wang, C. Wang, and H. Zhang, "Ship classification in high-resolution SAR images using deep learning of small datasets," *Sensors*, vol. 18, no. 9, p. 2929, Sep. 2018, doi: 10.3390/s18092929.

[36] L. Huang et al., "OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation," *IEEE J. Sel. Topics Appl. Earth Observ Remote Sens.*, vol. 11, no. 1, pp. 195–208, Jan. 2018, doi: 10.1109/JSTARS.2017.2755672.

[37] J. Wu et al., "A novel ship classification approach for high resolution SAR images based on the BDA-KELM classification model," *Int. J. Remote Sens.*, vol. 38, no. 23, pp. 6457–6476, Jul. 2017, doi: 10.1080/01431161.2017.1356487.

[38] 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, Apr. 2021, doi: 10.1109/TGRS.2020.3009284.

[39] 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, doi: 10.1109/LGRS.2022.3180793.

[40] 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, no. 4, Apr. 2020, Art. no. 140303, doi: 10.1007/s11432-019-2772-5.

[41] A. Raj J, S. M. Idicula, and B. Paul, "Lightweight SAR ship detection and 16 class classification using novel deep learning algorithm with a hybrid preprocessing technique," *Int. J. Remote Sens.*, vol. 43, nos. 15–16, pp. 5820–5847, Aug. 2022, doi: 10.1080/01431161.2021.2008544.

[42] C. Dechesne, S. Lefèvre, R. Vadaine, G. Hajduch, and R. Fablet, "Ship identification and characterization in Sentinel-1 SAR images with multi-task deep learning," *Remote Sens.*, vol. 11, no. 24, p. 2997, Dec. 2019, doi: 10.3390/rs11242997.

[43] Y. Dong, H. Zhang, C. Wang, and Y. Wang, "Fine-grained ship classification based on deep residual learning for high-resolution SAR images," *Remote Sens. Lett.*, vol. 10, no. 11, pp. 1095–1104, Nov. 2019, doi: 10.1080/2150704X.2019.1650982.

[44] N. Wang, Y. Wang, H. Liu, Q. Zuo, and J. He, "Feature-fused SAR target discrimination using multiple convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1695–1699, Oct. 2017, doi: 10.1109/LGRS.2017.2729159.

[45] 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, 2022, Art. no. 5210322, doi: 10.1109/TGRS.2021.3082759.

[46] T. Zhang and X. Zhang, "A polarization fusion network with geometric feature embedding for SAR ship classification," *Pattern Recognit.*, vol. 123, Mar. 2022, Art. no. 108365, doi: 10.1016/j.patcog.2021.108365.

[47] S.-W. Chen, X.-C. Cui, X.-S. Wang, and S.-P. Xiao, "Speckle-free SAR image ship detection," *IEEE Trans. Image Process.*, vol. 30, pp. 5969–5983, 2021, doi: 10.1109/TIP.2021.3089936.

[48] S.-W. Chen, "SAR image speckle filtering with context covariance matrix formulation and similarity test," *IEEE Trans. Image Process.*, vol. 29, pp. 6641–6654, 2020, doi: 10.1109/TIP.2020.2992883.

[49] Q. Zhang, "System design and key technologies of the GF-3 satellite," (in Chinese), *Acta Geodaetica Cartogr. Sinica*, vol. 46, no. 3, pp. 269–277, Mar. 2017, doi: 10.11947/j.AGCS.2017.20170049.

[50] (2015). *Step Science Toolbox Exploitation Platform*. [Online]. Available: http://step.esa.int/main/

[51] M. D. Ferreira, J. N. A. Campbell, and S. Matwin, "A novel machine learning approach to analyzing geospatial vessel patterns using AIS data," *GISci. Remote Sens.*, vol. 59, no. 1, pp. 1473–1490, Dec. 2022, doi: 10.1080/15481603.2022.2118437.

[52] I. I. Lin and V. Khoo, "Computer based algorithm for ship detection from ERS SAR imagery," in *Proc. 3rd ERS Symp. Space Service Our Environ.*, vol. 414, Florence, Italy, Mar. 1997, pp. 1411–1416.

[53] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.

[54] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125, doi: 10.1109/CVPR.2017.106.

[55] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.

[56] S. Song, K. Chaudhuri, and A. D. Sarwate, "Stochastic gradient descent with differentially private updates," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 245–248, doi: 10.1109/GlobalSIP.2013.6736861.

[57] G. Fu, L. Yi, and J. Pan, "Tuning model parameters in class-imbalanced learning with precision-recall curve," *Biometrical J.*, vol. 61, no. 3, pp. 652–664, May 2019, doi: 10.1002/bimj.201800148.

[58] Z. Zeng, J. Sun, Z. Han, and W. Hong, "SAR automatic target recognition method based on multi-stream complex-valued networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5228618, doi: 10.1109/TGRS.2022.3177323.

[59] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.

[60] J. Zhang, M. Xing, and Y. Xie, "FEC: A feature fusion framework for SAR target recognition based on electromagnetic scattering features and deep CNN features," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2174–2187, Mar. 2021, doi: 10.1109/TGRS.2020.3003264.

[61] J. Ai, Y. Mao, Q. Luo, L. Jia, and M. Xing, "SAR target classification using the multikernel-size feature fusion-based convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5214313, doi: 10.1109/TGRS.2021.3106915.

[62] L. van der Maaten, "Accelerating t-SNE using tree-based algorithms," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, Oct. 2014, doi: 10.5555/2627435.2697068.

[63] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, vol. 8689. Zurich, Switzerland, 2014, pp. 818–833.

[64] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, "ShuffleNet v2: Practical guidelines for efficient CNN architecture design," *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 11218. Munich, Germany, 2018, pp. 122–138, doi: 10.1007/978-3-030-01264-9_8.

[65] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[66] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollar, "Designing network design spaces," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10428–10436, doi: 10.1109/CVPR42600.2020.01044.

[67] 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, doi: 10.1145/3065386.

[68] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[69] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[70] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[71] A. Howard et al., "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324, doi: 10.1109/ICCV.2019.00140.

[72] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, Long Beach, CA, USA, 2019, pp. 6105–6114.

[73] M. Tan and Q. Le, "EfficientNetV2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 7102–7110.

[74] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806–4817, Aug. 2016, doi: 10.1109/TGRS.2016.2551720.

[75] Y. Guo, Z. Pan, M. Wang, J. Wang, and W. Yang, "Learning capsules for SAR target recognition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4663–4673, Aug. 2020, doi: 10.1109/JSTARS.2020.3015909.

[76] F. Zhang, Y. Wang, J. Ni, Y. Zhou, and W. Hu, "SAR target small sample recognition based on CNN cascaded features and AdaBoost rotation forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1008–1012, Jun. 2020, doi: 10.1109/LGRS.2019.2939156.

[77] G. Huang, X. Liu, J. Hui, Z. Wang, and Z. Zhang, "A novel group squeeze excitation sparsely connected convolutional networks for SAR target classification," *Int. J. Remote Sens.*, vol. 40, no. 11, pp. 4346–4360, Jan. 2019, doi: 10.1080/01431161.2018.1562586.

[78] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

[79] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7132–7141, doi: 10.1109/CVPR42600.2020.01155.

[80] Y. Liu, Z. Shao, and N. Hoffmann, "Global attention mechanism: Retain information to enhance channel-spatial interactions," 2021, *arXiv:2112.05561*.

[81] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007, doi: 10.1109/ICCV.2017.324.

Yanan Guan received the B.S. and Ph.D. degrees in computer technology and resource information engineering from the China University of Petroleum (East China), Qingdao, China, in 2017 and 2023, respectively.

Her research interests include investigating algorithms for ship detection, classification, and behavior analysis based on marine remote sensing data using deep learning and machine learning methods.

Xi Zhang received the B.S. degree in information systems from the Qingdao University of Science and Technology, Qingdao, China, in 2005, and the M.S. degree in signal and information processing and the Ph.D. degree in computer science from the Ocean University of China, Qingdao, China, in 2008 and 2011, respectively.

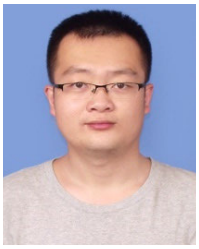
In 2011, he joined the Laboratory of Marine Physics and Remote Sensing, First Institute of Oceanography (FIO), Ministry of Natural Resources of China, Qingdao. Since 2021, he has been a Professor with FIO. Since 2020, he has been a Supervisor of Doctoral Students with the China University of Petroleum (East China), Qingdao. He has authored over 100 journal articles and conference papers, and written one book as a coauthor. His research interests include radar signal processing, synthetic aperture radar (SAR) target detection, and marine and cryosphere environment remote sensing.

Dr. Zhang also holds memberships in the Specialized Committee on Artificial Intelligence Oceanography of the Ocean Society of China and the Working Committee on Marine Spatial and Temporal Information of the China Association of Geoinformation Industry. He won four Chinese Provincial and Ministerial Awards. He is a Guest Editor (special issue) of *IEEE Geoscience and Remote Sensing Magazine* and *International Journal of Applied Earth Observation and Geoinformation*.

Siwei Chen (Senior Member, IEEE) received the Ph.D. degree (Hons.) from Tohoku University, Sendai, Japan, in 2012.

He is currently an Assistant Professor with the National University of Defense Technology, Changsha, China. He is a Coinvestigator of the European Space Agency Earth Observation Campaign and Advanced Land Observing Satellite-1 (ALOS) data. He has published more than 25 journal articles and conference papers. His research interests include radar polarimetry, polarimetric synthetic aperture radar (SAR), polarimetric SAR interferometry, and natural disaster study.

Dr. Chen was a recipient of the Outstanding Postgraduate Innovation Grant from the National University of Defense Technology from 2008 to 2009, the Young Researcher Award from the IEEE Geoscience and Remote Sensing Society (GRSS) Japan Chapter in 2011, the Dean Prize from the Graduate School of Environmental Studies, Tohoku University, in 2013, and the Best Poster Award at the IET International Radar Conference 2013. He was granted a Tohoku University President Fellowship from 2011 to 2012.



Genwang Liu received the B.S. and M.S. degrees in automation from Inner Mongolia University, Inner Mongolia, China, in 2011 and 2014, respectively, and the Ph.D. degree in computer science from the Ocean University of China, Qingdao, China, in 2017.

He is currently a Research Associate with the Laboratory of Marine Physics and Remote Sensing, First Institute of Oceanography, Ministry of Natural Resources of China, Qingdao. His research interests include ship synthetic aperture radar (SAR) detection and classification, polarimetric SAR and marine applications, SAR simulation, and image interpretation.

Yongjun Jia received the B.S. degree in information and computing science from Inner Mongolia University, Inner Mongolia, China, in 2004, and the M.A.Sc. and Ph.D. degrees in physical oceanography from the Institute of Oceanology, Chinese Academy of Sciences, Beijing, China, in 2007 and 2010, respectively.

He was a Deputy Chief Designer of the ground application system for the HY-2 satellites and new generational ocean dynamic satellites, where he organized the framework and managed the construction of the ground application system for the Chinese ocean dynamic satellites. His research interests include ocean satellite data processing, remote sensing of the ocean, and associated applications.

Yi Zhang received the Ph.D. degree from the Chinese Academy of Sciences, Beijing, China, in 2010.

In 2011, he was responsible for the development of geolocation and the normalized radar cross section (NRCS) data calculation algorithms for the HY-2A Satellite Microwave Scatterometer. He played an important role in developing the entire ground-based data processing system. From 2012 to 2016, he mainly engaged in ground system maintenance, ocean wind field applications, and so on. He is currently the Chief Designer of the HY-2B/C/D satellite microwave scatterometer ground data processing systems. His main study field includes microwave scatterometer ground-based data processing, satellite geolocation, ocean wind retrieval, and ocean wind field applications.

Gui Gao received the B.S. degree in information engineering and the M.S. and Ph.D. degrees in remote sensing information processing from the National University of Defense Technology (NUDT), Changsha, China, in 2002, 2003, and 2007, respectively.

In 2007, he joined the Faculty of Information Engineering, School of Electronic Science and Engineering, NUDT, as an Associate Professor. In 2017, he joined the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China, where he is currently a Professor. In 2018, he joined the College of Traffic Engineering, Hunan University of Technology, Zhuzhou, China, where he is a Distinguished Professor. He has authored over 100 journal articles and conference papers, and written four books and an English chapter. His research interests include radar signal processing, interferometric synthetic aperture radar (InSAR), target detection, marine environment, and synthetic aperture radar (SAR) ground moving target indication (GMTI).

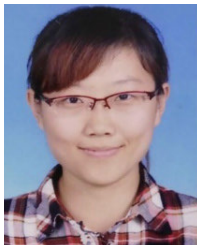
Jie Zhang received the B.S. and M.S. degrees in mathematics from Inner Mongolia University, Ürümqi, China, in 1984 and 1987, respectively, and the Ph.D. degree in applied mathematics from Tsinghua University, Beijing, China, in 1993.

He is currently the Dean of the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China. He has a broad interest in marine physics and remote sensing applications. He has been a supervisor for nearly 40 Ph.D. students and has published more than 200 papers. His research mainly focuses on the following: synthetic aperture radar (SAR) retrieval of ocean dynamics process and SAR detection of marine targets; ocean hyperspectral remote sensing; high-frequency surface-wave radar ocean detection technique; and integration of marine remote sensing application systems.

Dr. Zhang has served as a member of multiple domestic/international committees and a principal investigator/coinvestigator of many projects from the National Science Foundation of China, State High-Tech Development Plan (863), and other funding agencies.

Zhongwei Li received the Ph.D. degree from the China University of Petroleum, Qingdao, China, in 2011.

He is currently a Professor with the College of Oceanography and Space Informatics, China University of Petroleum. His research interests include remote sensing image processing and ocean numerical forecasting and cloud computing.



Chenghui Cao received the B.S. and M.S. degrees in mathematics from Qingdao Technological University, Qingdao, China, in 2014 and 2017, respectively, and the Ph.D. degree in information and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2021.

She is currently a Research Associate with the Laboratory of Marine Physics and Remote Sensing, First Institute of Oceanography, Ministry of Natural Resources of China, Qingdao. Her main research interests include radar target detection and sea clutter analysis.