MGSFA-Net: Multiscale Global Scattering Feature Association Network for SAR Ship Target Recognition

Xianghui Zhang, Sijia Feng, Chenxi Zhao, Zhongzhen Sun, Siqian Zhang, and Kefeng Ji, *Member, IEEE*

***Abstract*—Deep learning has offered new ideas in SAR ship target recognition. Although many methods improve the recognition performance through the improvement of loss function and migration of deep networks, scattering features as the important intrinsic features of SAR targets, need to be considered in the SAR ship recognition tasks. To introduce the scattering features into the deep network and characterize the features of ship targets more comprehensively, a multiscale global scattering feature association network (MGSFA-Net) for SAR ship target recognition is proposed in this article. In the network, the SAR ship target is first separated from the background by fine target segmentation. Then, the scattering centers (SCs) of ship targets are extracted and converted to local graph structures based on the *k*-nearest neighbors algorithm. The local graph structures are associated by the scattering center feature association module and enhanced by the multiscale feature enhancement module to produce the multiscale global scattering features. Moreover, the deep features of the targets are extracted by the multikernel deep feature extraction module to characterize the high-dimensional information. Finally, the scattering features and deep features are fused by weighted integration to enrich the diversity of features. The experimental results on the FUSAR-Ship and OpenSARShip dataset show that the MGSFA-Net can significantly improve the recognition performance, even on a few-shot condition with the accuracy increasing over 2%–3%. The feature distribution and visualization show the effectiveness of the MGSFA-Net to characterize the multiscale global scattering association features.**

***Index Terms*—Deep learning, multiscale, scattering feature association, ship recognition, synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) is an active microwave sensor with all-day, all-weather, and high-resolution characteristics. SAR plays an important role in the military defense and civil economy as an indispensable means of marine information acquisition and surveillance. SAR ship target recognition is an important development prospect, especially in the aspects of marine monitoring, target surveillance, ecological protection, and disaster forecasting [1], [2], [3], [4], [5], [6].

In the past few years, the traditional SAR ship target recognition methods were mainly centered on manual features and classifier design. The common manual features for SAR ship targets include texture structure, geometry, scattering intensity, histogram of orientation gradients [7], HOG [8], SIFT [9], etc. Many classifiers of machine learning are applied to the classification of features, such as KNN [10], SVM [11], and random forest [12]. Zhao et al. [13] used the hierarchical analysis method to select the most appropriate features from the feature set consisting of geometric structural features, transformed features, and local invariant features. The KNN method was applied to make the final classification decision, which can achieve the classification of TerraSAR-X images. Jiang et al. [14] divided the ship into bow, mid, and stern sections, extracted the ultrastructural electromagnetic scattering features of each section, and used SVM for the final classification. Lang et al. [15] designed a set of more efficient geometric feature vectors, used multikernel learning to learn the combination weights of the geometric features, and made the final decision by the SVM method. Xu et al. [16] used the distance metric learning (DML) approach to improve the discriminative power of feature representation by retaining supervised information about ship samples to maximize interclass differences and minimize intraclass differences. Although the traditional handcrafted features have explicit meaning and interpretable mathematical formulas, these types of features are overly dependent on the accumulation of expert knowledge and cannot lead to high recognition performance and excellent generalization ability.

Currently, with the rapid development of deep learning, various excellent algorithms have emerged, such as ResNet, VGGNet, and Inception [17], [18], [19], which can achieve great performance. Since Chen et al. [20] proposed an A-ConvNet network that demonstrated more than 99% recognition rate on moving and stationary target acquisition and recognition (MSTAR) dataset, deep learning methods have been gradually applied to the SAR field, and have become the mainstream method for SAR target recognition. Dechesne et al. [21] designed a multitask-driven deep neural network for joint detection and geometric

Manuscript received 3 November 2023; revised 25 December 2023; accepted 18 January 2024. Date of publication 23 January 2024; date of current version 15 February 2024. This work was supported by the National Natural Science Foundation of China under Grant 62001480. *(Corresponding author: Siqian Zhang.)*

The authors are with the State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System, the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: zxhzxh0122@163.com; fengsijia12@nudt.edu.cn; zcx_nudt2021@163.com; sunzhongzhen14@163.com; zhangsiqian@nudt.edu.cn; jikefeng@nudt.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3357171

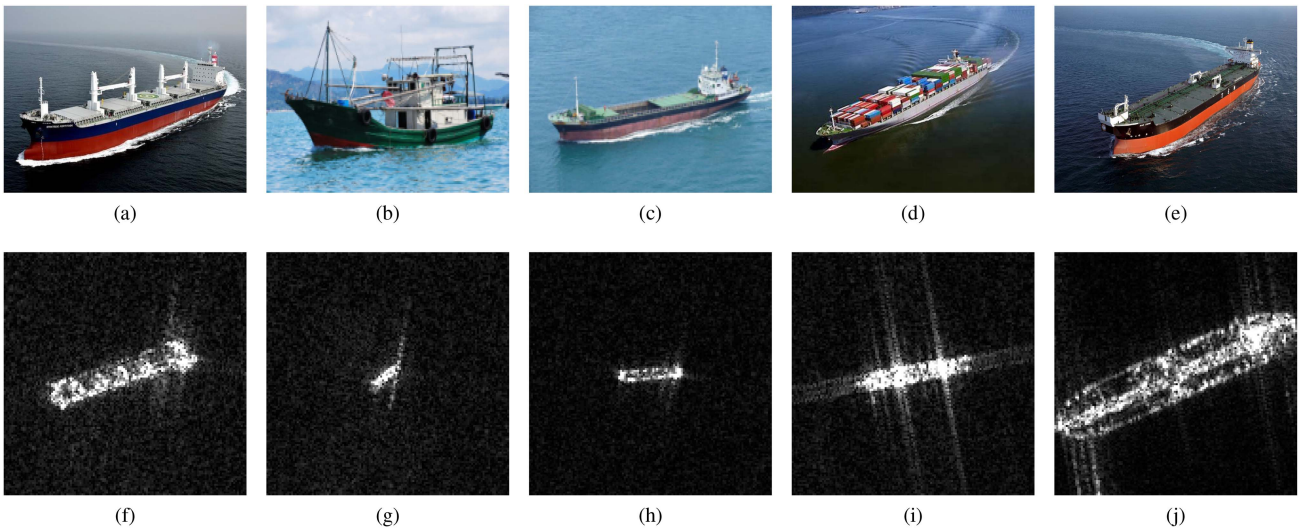


Fig. 1. SAR ship samples and the corresponding optical image. (a)–(e) are the optical images of ship targets. (f)–(j) are the SAR images of ship targets.

size estimation of ships in Sentinel-1 SAR images, which can successfully distinguish between oil tankers, cargo ships, fishing vessels, passenger ships, and trawlers. Zhang et al. [22] proposed a lightweight deep learning method combining the DML method and the gradually balanced sampling to solve the problem of the imbalance distribution of sample categories. Wang et al. [23] used a large optical dataset model as the pretraining weights and finetuned the training model to obtain high classification accuracy on the SAR ship. Lu et al. [24] proposed a new data augmentation method and combined it with transfer learning to achieve high recognition accuracy. He et al. [25] proposed a densely connected ternary CNN, and combined it with Fisher discriminant vectors to expand the feature distance between different categories, which improved the ship recognition accuracy in medium-resolution SAR images. Shang et al. [51] proposed a novel ship augmented attention network. It can utilize the kernels of different types combined with an attention mechanism to extract the multiscale features and can improve the performance of ship classification.

However, SAR ship targets have complex textures, shapes, and structures, as shown in Fig. 1. The phenomenon of defocusing and cross-highlighting spots also exist, which may lead to the poor quality of sample images. In addition, the differences in resolutions and imaging angles can also produce large feature differences in SAR ship target slices. The above deep learning methods mainly focus on the improvement of network structure, the modification of training techniques, transfer learning, and the expansion of training samples. To characterize the intrinsic features of SAR ship targets more accurately and reach better performance, researchers have focused on target recognition methods combining deep networks with traditional features. Zhang et al. [26] combined HOG features with deep features and produced better recognition results. Zeng et al. [27] fused polarization information and deep features to achieve SAR ship recognition. Electromagnetic scattering features, as the intrinsic feature of SAR targets, can better explain and characterize the electromagnetic scattering mechanism of the target by introducing them into deep networks, which has received wide attention in recent years. Ding et al. [31] combined the Attributed Scattering Centers (ASC) model and random projection features, which was the first trial of integrating the ASC feature with other features. Feng et al. [28] used the K-means clustering method to extract the features of different components of vehicle targets based on the ASC model and introduced these features into the deep network. Liu et al. [29] extracted different components of the target based on the ASC model and fused the different components with multiple deep network layers. Zhang et al. [30] constructed a visual bag-of-words from extracted the ASCs, transformed a set of the extracted parameters into features, and fused them with deep features. Zhang et al. [52] focused on the scattering point topology based on the scattering key points. They introduced the position information of the scattering key points and the distance information between the scattering points into the deep network, which can lead to an excellent recognition performance.

The ships have large sizes of targets and are mainly characterized by the combination of individual scattering structures with complex feature information. It is essential to consider the global structure of the target and the associations between different scattering features, which can help to characterize the recognition features of ship targets more accurately. In recent years, graph convolutional network (GCN) [32] has made significant progress in feature association and point cloud classification and has been gradually applied to the field of SAR image interpretation. Li et al. [33] proposed a graph learning method for SAR target recognition to acquire metaknowledge with correlation. Yang et al. [34] integrated GCN into deep networks to measure the similarity of samples, thus realizing the few-shot learning on SAR images. In addition, the graph convolution network has achieved excellent recognition results on SAR vehicle and aircraft targets. Li et al. [35] converted individual SCs into graph nodes and fully utilized the scattering centers through GCN to achieve local association of the scattering features. The extracted

SCs can better characterize the structural features of the SAR vehicle target by applying GCN to associate the scattering features. Zhao et al. [36] applied GCN to SAR aircraft targets. They converted the strong scattering points of the target into graph structures by GCN and combined the graph structure features of the scattering points with the features of the image domain extracted by improved VGGNet, which achieved effective classification of aircraft targets. However, unlike aircraft and vehicle targets, most ship targets have larger sizes with a relatively large aspect ratio of the target. The pixel distributions of different samples on the slices also vary considerably. In addition, ship targets are more compact in structure. The structure of a ship is characterized by a combination of different local scattering structures. Moreover, affected by the observation angle and imaging resolution, similar ships usually exhibit different sizes on different slices. Therefore, GCN cannot be directly used in SAR ship target recognition and necessary improvements are needed.

To address the above challenges, the MGSFA-Net for SAR ship target recognition is proposed in this article. First, the target is split from the background by target fine segmentation to avoid the background clutter. Then the SCs of ship targets are extracted and converted to graph structures. The local graph structures are associated to characterize the global structure of targets based on the scattering center feature association (SCFA) module and enhanced by the multiscale feature enhancement (MSFE) module to obtain the multiscale global scattering association features. In addition, the deep features of ship targets are extracted based on the multikernel deep feature extraction (MKDFE) module to get the features of the high-dimensional layer. Finally, the scattering features and deep features are fused to increase the diversity of features and improve the recognition accuracy of the SAR ship. The contributions of this article are summarized as follows.

1) A MGSFA-Net for SAR ship target recognition is proposed to combine the deep network with the internal scattering features of ship targets.
2) The SCFA module is designed to construct the local graph structure of SCs and associate the local structure to characterize the global topology structure feature of ship targets, which can improve the ability to characterize the global features and utilize the scattering information more fully.
3) The MSFE module is designed with a multiscale fusion method to fit the difference in the sizes of ship targets and reduce the loss of scattering information of high levels in SCFA module.
4) Due to the multiscale characteristics of ship targets, the MKDFE module is designed to better characterize the deep features of the high-dimensional layer.

The remaining structure of this article is organized as follows. In Section II, the details of the proposed MGSFA-Net are presented. In Section III, the implementation details of experiments are presented, and the results are shown. The conclusions of the experiments are shown in Section IV.

II. Proposed Methodology

A multiscale global scattering feature association network (MGSFA-Net) is proposed for SAR ship target recognition, the framework of which is shown in Fig. 2. It consists of five modules:

1) Target fine segmentation;
2) SCFA module;
3) MSFE module;
4) MKDFE module;
5) Feature fusion module.

In this section, the details of the proposed network are presented.

A. Target Fine Segmentation

SAR sensor will produce azimuth and range blur in ship target imaging. Azimuth blur is formed by the overlap of main signals and the azimuth upwards side flap signals. Range blur is formed by the overlap of echoes around the side of a ship and the echoes in the distance. In addition, the wave motion of the sea surface will make the larger ship target swing at a certain angle, resulting in the jump phenomenon of scattering centers in the range direction of the resolution unit. Also, there will be a diffusion phenomenon on the ship target in the range direction and azimuth direction, leading to outspread shadows and crossed shadows at the edge of the target imaging area. Considering the above situation, the target fine segmentation method is adopted in this article to realize the separation of the SAR target and background area.

First of all, the segmentation method based on the Radon transform is used to separate the ship target and imaging interference region from the perspective of the frequency domain. The threshold segmentation is performed on the separated ship target slices, and morphological means are used to process the segmented image to reduce the influence of side flaps and extract the target main region accurately. Also, the target region is refined based on ellipsoidal constraints to solve the regional fracture phenomenon and obtain the image segmentation region of the ship target. The specific process of refined segmentation is shown in Fig. 3.

B. SCFA Module

This module can extract the SCs, convert the SCs into the local graph structure, and associate the local structure to characterize the global scattering structure of the ship target.

1) Scattering Centers Extraction: Due to the specific characteristics of SAR target imaging, the SAR ship target region can be approximated as the sum of the responses of multiple independent scattering centers, which can be expressed as follows [35]:

$$\mathbf{E}(f,\varphi;\theta)=\sum_{i=1}^{k}\mathbf{E}_i\left(f,\varphi;\theta_i\right) \tag{1}$$

where k denotes the total number of the SCs, f denotes the frequency, and φ denotes aspect angle. The backscattering model can be expressed as follows:

$$\mathbf{E}_i\left(f,\varphi;\theta_i\right)=S_i\cdot\exp\left(\frac{-j4\pi f}{c}\left(x_i\cos\varphi+y_i\sin\varphi\right)\right) \tag{2}$$

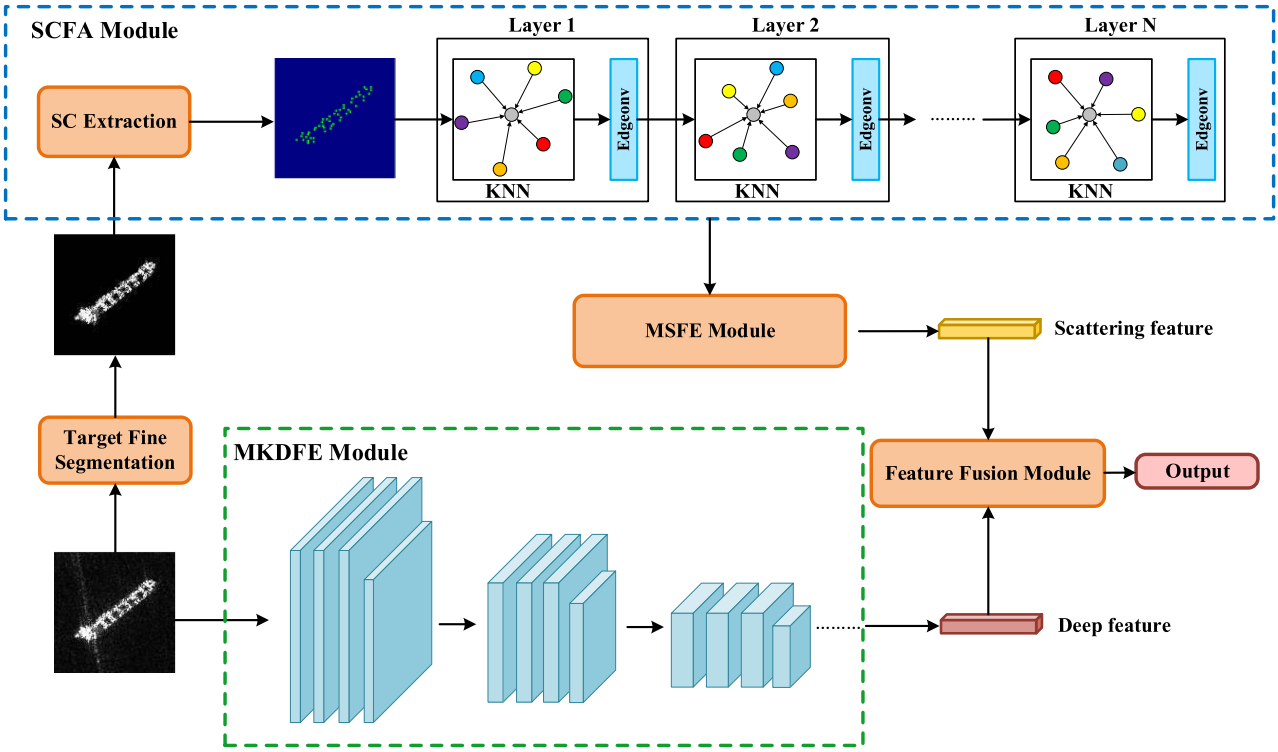


Fig. 2. Overall structure of multiscale global scattering feature association network for SAR ship target recognition.

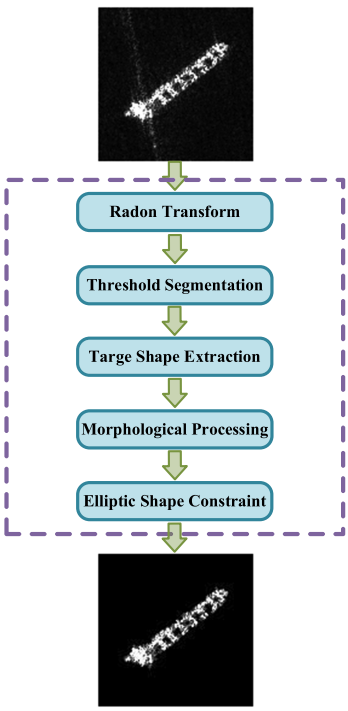


Fig. 3. Example of SAR ship target fine segmentation.

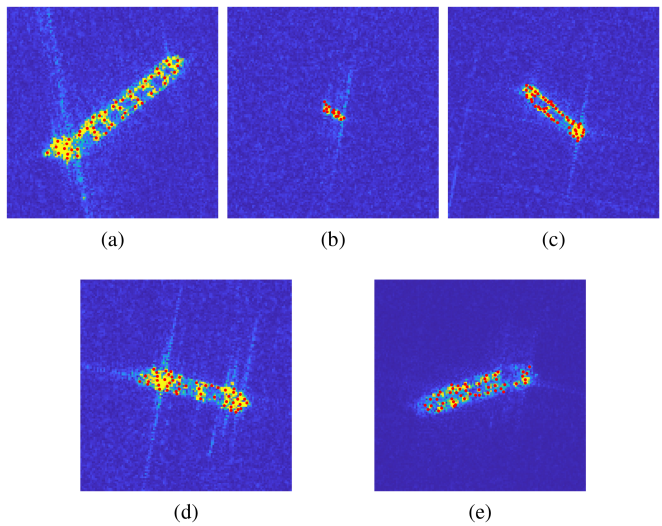


Fig. 4. Scattering centers extraction of ships with different categories. (a) BulkCarrier. (b) Fishing. (c) GeneralCargo. (d) Containership. (e) Tanker.

where S_i denotes the backscattering coefficient, x_i and y_i denote the position coordinates along the range dimension, and azimuth dimension, respectively. These scattering parameters are estimated by the GA algorithm [37], and each extracted scattering center can be characterized as $\mathbf{P}_i = [S_i, X_i, Y_i]$, which contains the scattering intensity and the position information. Then the ship target can be characterized as the sum of n scattering centers $\{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_n\}$. The extracted results are shown in Fig. 4.

2) Local Graph Structure Construction: The extracted SCs can be regarded as nodes of the first layer of graph structure input $\mathbf{V}_i = [S_i, x_i, y_i]$. By using the k-nearest neighbors (KNN) algorithm, the nodes are connected with the n neighboring nodes around them to generate the local graph structure.

Algorithm 1: Scattering Center Feature Association Module.

Input: SAR ship target with segmentation
Output: Scattering feature vector of different Layers
Initialization: Layers: n; KNN points: k;

1: Extract the scattering centers by GA algorithm [37], each SC is represented as a set of three parameters:
   $\rightarrow \mathbf{P} = [S, x, y]$
2: **for** $(l = 0; l < n; l++)$ **do**
3: Select the k nearest neighboring points of the scattering centers and regard them as nodes:
   $\rightarrow \mathbf{V}^{(l)} = \{v_1^{(l)}, v_2^{(l)} \ldots\ldots, v_k^{(l)}\}$
4: Construct graph structure between scattering centers and other neighboring points:
5: **if** $l = 0$ **then**
6: $\rightarrow \mathbf{G}^0 = (\mathbf{V}^{(0)}, \mathbf{E}^{(0)})$
7: **else**
8: $\rightarrow \mathbf{G}^l = (\mathbf{V}^{(l)}, \mathbf{E}^{(l)})$
9: **end if**
10: Calculate the edge feature by $v_j - v_i$ and v_i nodes with the $edgeconv$ operation:
   $\rightarrow e_{ij} = \bar{h}_\Theta(v_i, v_j - v_i)$
11: Update and calculate the nodes of the next layer by the edge of the previous layer:
   $\rightarrow v_i^{(l+l)} = \sum_{j:(i,j)\in\mathbf{E}^{(l)}} h_\Theta^{(l)}(v_i^{(l)}, v_j^{(l)})$
12: **end for**

The set of distances between its vertices and the adjacent k points can be regarded as edges $\mathbf{E}$ of the graph structure. Therefore, the input graph structure can be regarded as $\mathbf{G}^0 = (\mathbf{V}^{(0)}, \mathbf{E}^{(0)})$. The set of nodes at layer l of graph structure is $\mathbf{V}^{(l)} = \{v_1^{(l)}, v_2^{(l)}, \ldots, v_n^{(l)}\}$, and the graph structure constructed at lth layer is $\mathbf{G}^l = (\mathbf{V}^{(l)}, \mathbf{E}^{(l)})$.

3) Global Scattering Feature Association: The algorithm mainly uses the convolution of edge (named as $edgeconv$) [44] to calculate the relationship between the local graph structure, associate the local graph structure of scattering centers, and characterize the global structure of the targets.

Define the edge feature between nodes as $e_{ij} = h_\Theta(v_i, v_j)$, where v_i and v_j are the ith and jth nodes, respectively. To combine the global structure with the local neighborhood feature, the edge features are calculated by the relative positions of the neighbor points captured by $v_j - v_i$ and the center points captured by v_i. This method of calculation can take into account both the global structure and the local scattering features, which can associate the local features to characterize the global structure. Therefore, the calculation formula of the edge features is expressed as

$$h_\Theta(v_i, v_j) = \bar{h}_\Theta(v_i, v_j - v_i). \tag{3}$$

Thus, edge features can be calculated by local features and global features as

$$e_{ijm} = f(\theta_m \cdot (v_j - v_i) + \phi_m \cdot v_i) \tag{4}$$

where e_{ijm} denotes the mth edge feature from node x_i to node x_j, $\Theta = \{\theta_1, \ldots\ldots, \theta_n, \phi_1 \ldots\ldots, \phi_n\}$ denotes the key weight parameter for the convolution operation of the network, and $f(\cdot)$ denotes the activation function.

The feature vectors of each node in the next graph structure layer are obtained by aggregating and mapping the edge features of the previous layer to achieve the update of the graph structure network layer and iteration of the feature extraction. Therefore, the next layer can be represented as

$$v_i^{(l+l)} = \sum_{j:(i,j)\in\mathbf{E}^{(l)}} h_\Theta^{(l)}\left(v_i^{(l)}, v_j^{(l)}\right) \tag{5}$$

where $v_i^{(l+l)}$ denotes the ith node of layer $(l+1)$, $v_i^{(l)}$ denotes the ith node of the layer l, and $v_j^{(l)}$ denotes the jth node of layer l. The update of the node is completed by the iteration of $edgeconv$ operation. Therefore, the SCFA module consists of several Layers, each of which is composed of a KNN layer and a $edgeconv$. The details of SCFA module can be concluded as Algorithm 1.

C. MSFE Module

In order to fit the great differences in the sizes of ship targets and reduce the loss of scattering information in the SCFA module, a MSFE module is designed to stitch the scattering features extracted from different layers of the SCFA module.

This module can characterize the multiscale scattering features and compensate for the loss of scattering information. First, the different layers of the SCFA module are fused through the concatenation operation. The fusion operation of different layers can be expressed as

$$\mathbf{F_{multilayer}} = Concat(\mathbf{F_1}, \mathbf{F_2}, \ldots, \mathbf{F_N}) \tag{6}$$

where $\mathbf{F_{multilayer}}$ represents the concatenated features of different layers and $\mathbf{F_i}$ represents the features of the ith layer. Then, the pooling operations are performed on $\mathbf{F_{multilayer}}$ by using MaxPooling and AvePooling operations in a parallel manner. Finally, the multiscale features $\mathbf{F_{multiscale}}$ are extracted by concatenating the two features that pass through the two different pooling layers. The framework of the module is shown in Fig. 5.

The multiscale features can better characterize the scattering features of targets of different sizes and can characterize the scattering structure of the targets more accurately and comprehensively. In addition, the increase in the number of layers in the SCFA module can lead to the loss of scattering information. Therefore, the features extracted from the last layer cannot characterize the scattering information of the target accurately. Multiscale features can utilize the features of the low-dimensional layer to complement the detailed features of the high-dimensional layer to reduce the loss of information in the scattering features obtained from higher layers of the network. The effects of different pooling operations on the recognition performance of the algorithm will be analyzed in a detailed comparison in the subsequent experimental section.

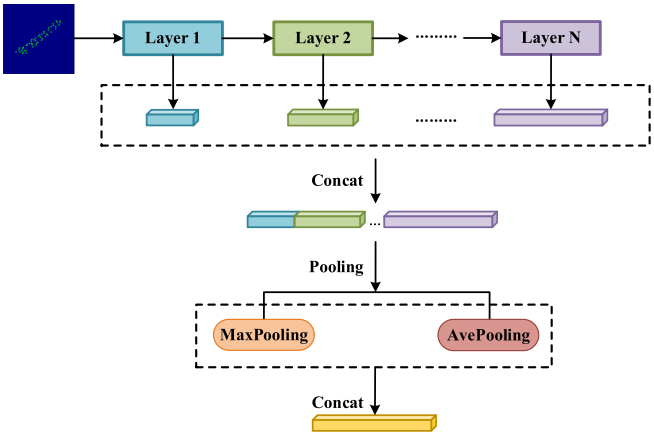


Fig. 5. Framework of MSFE module.

TABLE I
TRAINING-TEST DIVISION OF THREE-CATEGORY FUSAR-SHIP DATASET

Class	Train	Test	Total
BulkCarrier	150	82	232
Fishing	150	105	255
GeneralCargo	150	159	309

TABLE II
TRAINING-TEST DIVISION OF FIVE-CATEGORY FUSAR-SHIP DATASET

Class	Train	Test	Total
BulkCarrier	150	82	232
Fishing	150	105	255
GeneralCargo	150	159	309
ContainerShip	150	75	225
Tanker	150	150	300

D. MKDFE Module

The multikernel deep feature extraction module is designed based on the fusion of multiple convolution kernels to address the problem of large differences in the target size of ship slices and to extract deep features of ships more comprehensively. In the module, 3×3, 5×5, and 7×7 kernels are used to perform convolution operations on the image respectively with the structure of the module shown in Fig. 6. The 3×3 kernel has a smaller receptive field, which pays more attention to the local details and texture information. Therefore, it can learn the information of texture and details of the target, which is beneficial for extracting the features of the targets with small sizes. The 5×5 kernel has a larger receptive field than 3×3. It can capture the information of a wider area and can learn the global structure and shape of the image better. The 7×7 kernel has a larger receptive field, which can learn the global structure and shape of the target better. The 7×7 kernel has a larger receptive field and pays more attention to the global information in the image, which can extract the features of targets with large sizes. Then the feature maps extracted from convolutional layers of different kernel sizes are fused to extract the global structural features of the target while preserving the detailed texture features of the target and making the extracted features more complete. This allows for better adaptation to the different sizes of the target. In addition, the last pooling layer is designed with a size of 8×8, which can highlight the high-dimensional features of the input image and the global spatial framework of the target without losing too much local information. Finally, the feature vector of 128 dimensions is produced through the linear layer to obtain the deep features of the extracted ship target.

E. Feature Fusion Module

In this section, the electromagnetic scattering features are combined with deep features at the output by weighted integration to combine the advantages of both features. Mathematically, the fusion method can be expressed by

$$\mathbf{F_{fusion}} = w_{SF} \cdot \mathbf{F_{SF}} + w_{DF} \cdot \mathbf{F_{DF}} \tag{7}$$

where w_{SF} and w_{DF} are the weights of the scattering features $\mathbf{F_{SF}}$ and deep features $\mathbf{F_{DF}}$, respectively. $\mathbf{F_{fusion}}$ denotes the fusion features of both features. This fusion of features can combine the scattering structure features of the target and the high-dimensional semantic features extracted by the deep network to achieve the complementary advantages of the two features. In this article, the weight of scattering features w_{SF} and the weight of deep features w_{DF} are selected as 0.6 and 0.4, respectively. The effect of different fusion weights on the recognition performance will be analyzed in detail in Section III.

III. EXPERIMENTS AND RESULTS

A. Dataset Description

The effectiveness of our algorithm is verified on two SAR ship datasets that is FUSAR-Ship dataset and OpenSARShip dataset.

1) FUSAR-Ship Dataset: The FUSAR-ship dataset is a SAR ship dataset released by Hou et al. [6] in 2020, which is a high-resolution SAR ship classification dataset acquired from the quad polarization Gaofen-3 satellite. The dataset consists of five main target categories: BulkCarrier, Fishing, General-Cargo, ContainerShip, and Tanker. The Five-Category dataset is constructed on these five categories. In addition, the categories of BulkCarrier, Fishing, GeneralCargo are more representative. Thus, the Three-Category experiment dataset is constructed based on these three categories. The samples are expanded by flipping and rotating operations. The specific compositions of Three-Category and Five-Category datasets are shown in Tables I and II, respectively. The examples of several SAR ship samples from the FUSAR-Ship dataset are presented in Fig. 7.

2) OpenSARShip Dataset: The OpenSARShip dataset is a SAR ship dataset released by Huang et al. [50] in 2018, which is acquired from the dual-polarization SAR detected by the Sentinel-1 satellite of European Space Agency, including both VH and VV polarization channels. The sizes of the slices are not the same. The dataset utilized in our experiment is ground range detected (GRD) products. In order to verify the effectiveness of our methods further, we select BulkCarrier, ContainerShip, and Tanker to compose the three-category OpenSARShip dataset. The specific compositions of the data are shown in Table III.

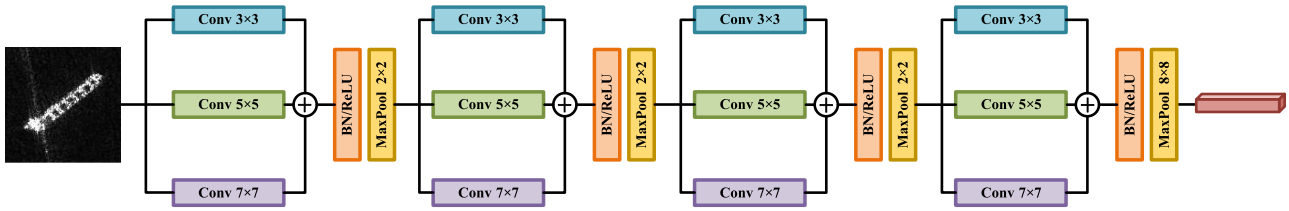


Fig. 6. Structure of MKDFE module. The module consists of several kernels of convolutional layers, which can extract complicated deep feature information of SAR ship targets.

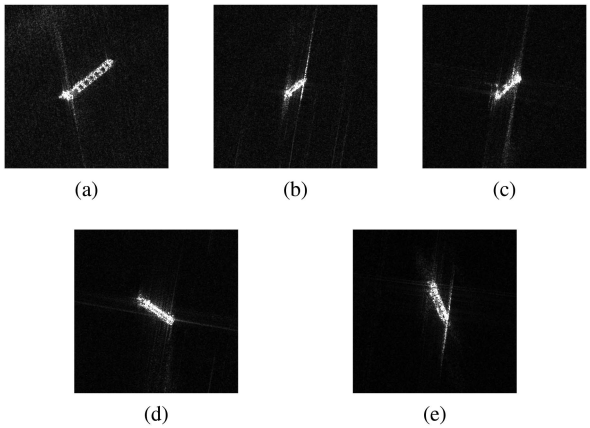


Fig. 7. SAR ship samples in FUSAR-Ship. (a) BulkCarrier. (b) Fishing. (c) GeneralCargo. (d) ContainerShip. (e) Tanker.

TABLE III
TRAINING-TEST DIVISION OF OPENSARSHIP DATASET

Class	Train	Test	Total
BulkCarrier	448	906	1354
ContainerShip	448	526	974
Tanker	448	193	641

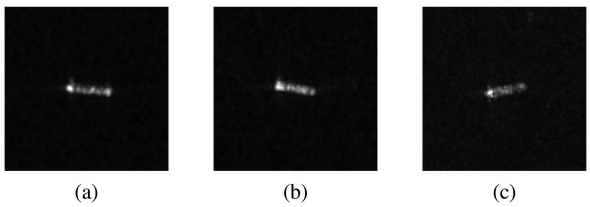


Fig. 8. SAR ship samples in OpenSARShip. (a) BulkCarrier. (b) ContainerShip. (c) Tanker.

Fig. 8 presents several SAR ship samples from the OpenSARShip dataset.

B. Experimental Setup

In order to improve the computational efficiency, all samples of FUSAR-Ship are cropped to 256 × 256 and then resized to 128 × 128 in this article. The samples of OpenSARShip are cropped and padding to 128 × 128. The experimental learning rate is set to 0.001 and the learning momentum is set to 0.9. The updating method of the parameter is stochastic gradient descent, the training epoch is set to 150 and the batch size is set to 16. All the experiments are carried out with both MATLAB platform and the Pytorch deep learning framework on a NVIDIA GeForce RTX 3090 GPU. We use the PyTorch 1.8 and Python 3.8 compilers run on Windows 10. Besides, CUDA 11.6 is employed for the algorithm acceleration of training and inference.

C. Evaluation Indicators

In this article, accuracy is used as the main indicator to evaluate the effect of experiment methods. In addition, other three types of evaluation indexes are used, that is recall, precision, and F1, to compare the experimental performance of different methods and to verify the effectiveness of the proposed method more completely and accurately.

Recall is defined by

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{8}$$

Precision is defined by

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{9}$$

F1 is defined by

$$\text{F1} = 2 \cdot \frac{\text{Recall} \cdot \text{precision}}{\text{Recall} + \text{precision}}. \tag{10}$$

Accuracy is defined by

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{11}$$

where TP denotes the true positives, FP denotes the false positives, FN denotes the false negatives, and TN denotes the true negatives. For the convenience of representation, recall is abbreviated as "R," precision is abbreviated as "P," and accuracy is abbreviated as "Acc."

D. Model Analysis

In this section, the impact of different parameters of different modules is analyzed. The ablation study of the proposed model is conducted to select the optimal parameters and architecture of the model.

1) Impact of Different Parameters in MGSFA-Net: The two most important parameters in the network are the nearest neighbor point k in KNN algorithm and the number of layers N. The recognition performance on different conditions of k and N are analyzed, with the results shown in Fig. 9.

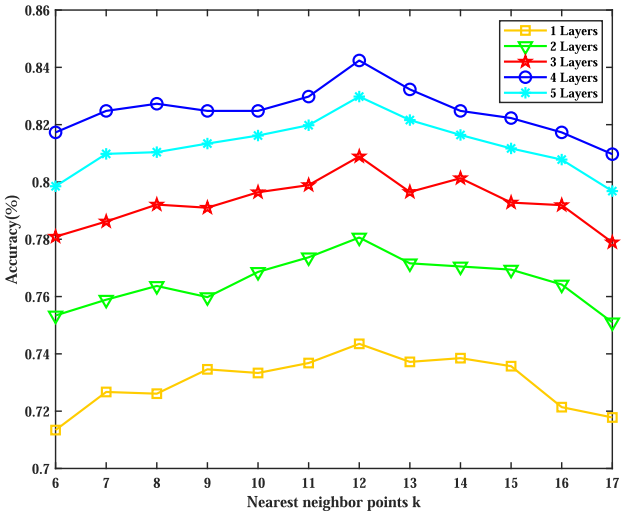


Fig. 9. Influence of nearest neighbor point k in KNN algorithm and the number of layers N.

a) Nearest neighbor point k: As shown in Fig. 9, under different k-nearest neighbor conditions, the accuracy rate shows a trend of increasing and then decreasing as the value of k keeps increasing. As the value of k increases, more nearest neighbor points can be selected, which can help construct a closer and more accurate topology association structure. However, if the value of k is too high, it will lead to an extremely complex structure association and inaccurate association of features, which can produce plenty of redundant association features, thus having a negative impact on the recognition performance. The optimal performance of the network can be achieved when the value of k is equal to 12.

b) Number of layers N: It can be seen from the curve that the recognition performance tends to get better and then worse with the number of layers increasing, reaching the best accuracy of 84.24% when the number of layers is 4. The high-dimensional layer has more dimension of the extracted features, which can characterize more intrinsic associated features. However, too many layers will cause some loss of scattering information when extracting scattering features. This will result in some feature redundancy in the fusion of features in the MSFE module, leading to a decline in the recognition performance.

Therefore, the number of layers is set to 4 and the number of KNN nearest neighbor points is set to 12 in this article to achieve the best performance in our experiment.

2) Different Pooling Methods of MSFE Module: To enhance scattering features better, we compare the effect of maximum pooling, average pooling, adding of the two poolings, and concatenation of the two poolings, as the results shown in Table IV. From the analysis of the results, it can be concluded that concatenating the maximum pooling and the average pooling achieves better results. Maximum pooling retains the most obvious and salient information locally and pays more attention to texture details. Average pooling pays more attention to the global structure and global feature distribution. Therefore concatenating the two pooling methods can extract more accurate and comprehensive scattering features, reduce the feature information loss caused by an increase in the dimensions of the network, and enhance the multiscale features better.

TABLE IV
DIFFERENT FUSION METHODS OF MSFE MODULE

Methods	Recall(%)	Precision(%)	Accuracy(%)
Max(·)	78.64	78.97	78.63
Ave(·)	78.20	80.31	79.16
Add[Max(·), Ave(·)]	81.32	80.50	81.09
Cat[Max(·), Ave(·)]	**84.94**	**83.79**	**84.24**

The bold values represent the best result.

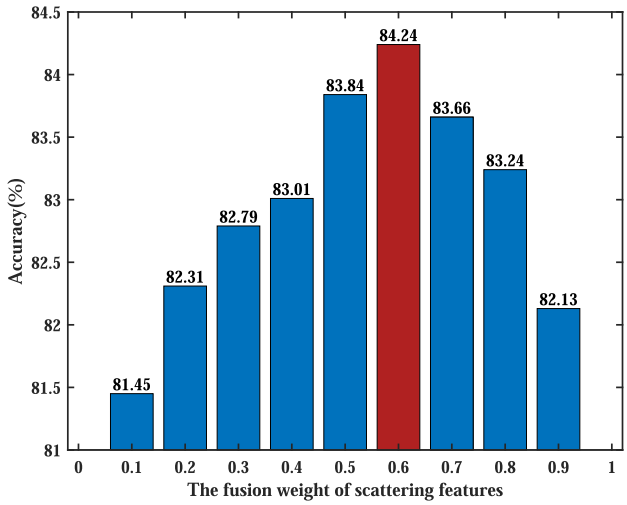


Fig. 10. Influence of fusion weight of scattering features in MGSFA-Net.

3) Weight Analysis of Feature Fusion Module: The recognition performance of different fusion weight conditions under the experimental conditions of five-category is compared and analyzed to select the optimal fusion weights for the fusion of scattering features and deep features. The bar charts of the recognition accuracy are shown in Fig. 10. The experimental results show that the recognition accuracy is not high when the weights of single features are too large or too small. For example, when the weights of scattering features are 0.1 or 0.9, the recognition accuracy is 81.45% and 82.13%, respectively, which is the worst performance among all weights. When the weights of scattering features and deep learning features are 0.6 and 0.4, respectively, the recognition accuracy is the highest, which can reach 84.24%. Therefore, the weights of scattering features and deep learning features are selected as 0.6 and 0.4, respectively, to reach the best results.

4) Ablation Experiment: The ablation experiments are conducted on the five-category FUSAR-Ship dataset to compare the effect of the MKDFE module, the SCFA module, and the MSFE module. The results are summarized in Table V.

In Table IV, the "✓" denotes that the module is involved while the "×" denotes that the module is not involved. When only the SCFA module is involved in the experiment, the extracted scattering features are the features extracted from the last layer in the SCFA module and the recognition rate can only reach

TABLE V
ABLATION STUDIES OF MGSFA-NET

MKDFE	SCFA	MSFE	Feature Fusion	Accuracy(%)
×	✓	×	×	64.57
×	✓	✓	×	71.93
✓	×	×	×	80.45
✓	✓	×	✓	82.10
✓	✓	✓	✓	**84.24**

The bold values represent the best result.

64.57%. This is because the simple scattering points cannot represent all the useful information of the target recognition network. In addition, there is scattering information loss in the high-dimensional feature layer, which can lead to a low accuracy of recognition. However, with the involvement of the MSFE module, the accuracy can increase to 71.93%. This indicates that the module fusing scattering features of different scales can effectively solve the problem of the loss of scattering information in high-dimensional layers, and can enhance the characterization ability of the features, which is more conducive to improve the target recognition performance. When only the MKDFE module is used in the experiment, the accuracy is 80.45%. The feature extraction in the image domain can acquire the high-dimensional abstract features from the deep network through different convolutional kernels. With the integration of the SCFA, MSFE, and MKDFE modules, the proposed method can combine the advantages of deep network and scattering information. The accuracy of MGSFA-Net can reach 84.24%. The results show that the recognition performance is significantly improved by fusing scattering features and deep features, and the SCFA module can assist the deep learning network in characterizing the features of the target more comprehensively.

E. Recognition Performance and Comparison

The classical target recognition methods are selected and compared with the proposed method to verify the performance of MGSFA-Net. The comparison methods include excellent deep learning networks, such as Hou [6], GoogleNet [38], MobileNet [41], DenseNet [39], Inceptionv3 [19], ShuffleNet [43], ResNet [17], A-ConvNet [20], VGGNet [18], and InceptionNet [19]. The algorithms appearing in computer vision recently, such as Swin Transformer [46], Vision Transformer [47], ConvNext [48], and Internimage [49] are also compared. The methods fusing deep features and other traditional features, such as MSDF [29], FEC [30], SFSA [36], and HOG-ShipCLSNet [26], are included as well. The experiments on FUSAR-Ship datasets and OpenSARShip datasets are conducted, respectively, with the results listed in Table IX.

1) Results of Experiments on Three-Category FUSAR-Ship: Among the three-category experiments, the highest recognition rate in the classical classification network of deep learning can reach 83.24%. In the method of fusion of traditional features and deep features, FEC can reach the highest recognition rate of 83.25%. The methods related to the Transformer show bad performance on the SAR ship target recognition. The accuracy of vision transformer can only reach 69.65% on the three-category FUSAR-Ship. Compared with these methods, the recognition accuracy of the proposed MGSFA-Net can reach 85.26%. In terms of recall and precision, the method in this article still achieves an excellent performance. The precision of the proposed method reaches 87.52% and the recall rate reaches 84.87%, which is higher than the rest methods. The confusion matrix for the three-category experiments is shown in Table VI. In the confusion matrix, BulkCarrier and GeneralCargo have the highest rates of recall, both of which are more than 90%. While the fishing target recall is only 69.52%. Since the size of the fishing target is too small, the number of extracted scattering centers is small, and the scattering characteristics are not obvious, the fishing category may result in a low recognition rate. In addition, BulkCarrier and GeneralCargo are prone to confusion since the two types of targets are of similar size and have similar features. The bar charts of accuracy under different methods on three-category FUSAR-Ship are shown in Fig. 11.

TABLE VI
CONFUSION OF EXPERIMENT OF THREE-CATEGORY FUSAR-SHIP

Predict \ True	BulkCarrier	Fishing	GeneralCargo	Recall(%)
BulkCarrier	77	0	5	93.90
Fishing	0	73	32	69.52
GeneralCargo	6	8	145	91.19
Precision(%)	92.77	90.12	79.67	**Accuracy = 85.26%**
F1(%)	93.33	78.49	85.04	

TABLE VII
CONFUSION OF EXPERIMENT OF FIVE-CATEGORY FUSAR-SHIP

Predict \ True	BulkCarrier	ContainerShip	Fishing	GeneralCargo	Tanker	Recall(%)
BulkCarrier	76	0	2	3	1	92.68
ContainerShip	0	83	19	1	2	79.04
Fishing	2	19	115	6	4	72.33
GeneralCargo	3	1	6	64	3	85.33
Tanker	1	2	5	1	143	95.33
Precision(%)	85.39	77.57	78.77	83.12	94.08	**Accuracy = 84.24%**
F1(%)	88.89	78.30	75.41	84.21	94.70	

2) Results of Experiments on Five-Category FUSAR-Ship: In the five-category of target classification experiments, the recognition rate of A-ConvNets, VGGNet, HOG-ShipCLSNet, MSDF, and FEC can exceed 80%, butit is still lower than the result of MGSFA-Net, which is 84.24%. The accuracy of swin transformer and vision transformer can only reach 62.52% and 54.12%, respectively, much lower than other methods. In addition, the ConvNext and Internimage methods can achieve 65.67% and 70.40%, respectively, which cannot classify the SAR ship targets effectively. The confusion matrix of the five-category experiments is shown in Table VII. The confusion matrix shows that the larger the ship target size is, the better the recognition performance is, reaching 92.68% and 95.33% for BulkCarrier and Tank, respectively. This is because these two categories of target sizes are large and the identifiable features are prominent. Moreover, they contain more scattering information in the target and can reflect more significant scattering characteristics. The network can learn richer features,

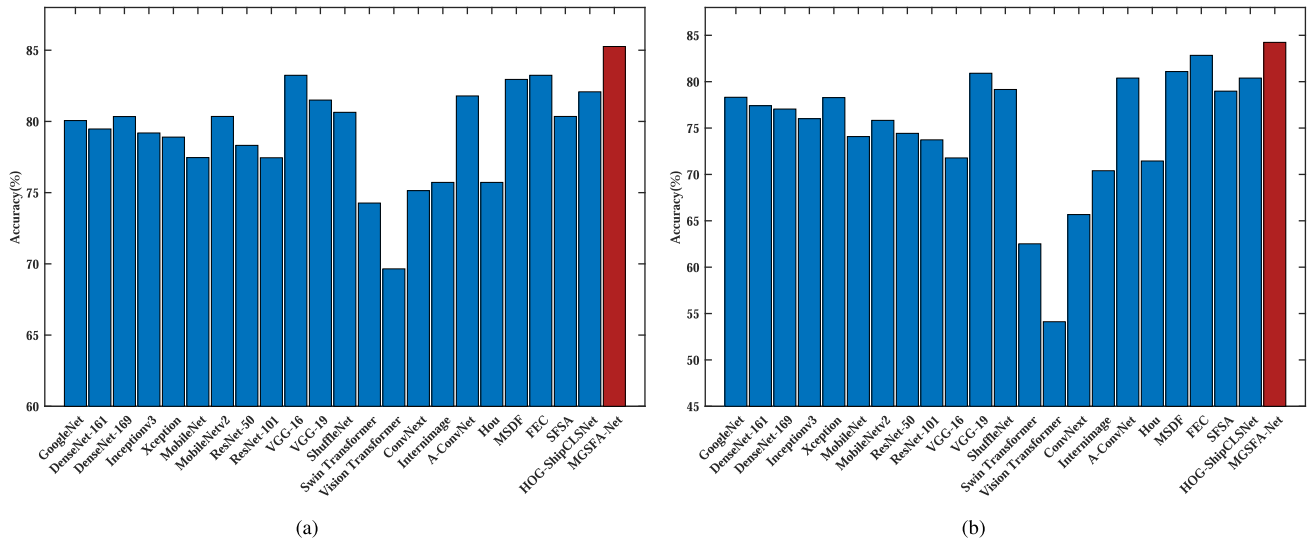


Fig. 11. Bar charts of accuracy under different methods on FUSAR-Ship dataset. (a) Accuracy of experiment on three-category FUSAR-Ship. (b) Accuracy of experiment on five-category FUSAR-Ship.

TABLE VIII
CONFUSION OF EXPERIMENT OF OPENSARSHIP

True / Predict	BulkCarrier	Containership	Tanker	**Recall**(%)
BulkCarrier	780	106	20	86.09
ContainerShip	65	443	18	84.22
Tanker	12	12	169	87.56
Precision(%)	91.02	78.97	81.64	**Accuracy = 85.66%**
F1(%)	88.49	81.51	84.50	

and thus the recognition performance improvement is more obvious. The target of the Tanker is chosen as an example to analyze. The decks of the target have alternating horizontal and vertical bulkheads with a very obvious structure. Therefore, the extracted scattering features are more accurate and significant. As a comparison, there is no prominent structure on the surface of the fishing vessel, which means that it is difficult to extract the distinctive scattering features of the fishing target. Through the above experiments, the superior performance of MGSFA-Net on the SAR ship target recognition task is verified. The bar charts of accuracy under different methods on Five-Category FUSAR-Ship are shown in Fig. 11.

3) Results of Experiments on OpenSARShip: In the experiments of OpenSARShip, the confusion matrix is shown in Table VIII. Compared with classical deep learning methods and Transformer methods, the methods fusing traditional features and deep features show excellent performance on SAR ship target recognition. FEC and MSDF methods, the accuracy of which achieve 81.48% and 83.36%, respectively, can combine the scattering features and deep features to characterize the features comprehensively. However, the methods related to the transformer reach low accuracy, with the swin transformer and vision transformer only reaching 72.80% and 72.92%, respectively. The ConvNext and Internimage can lead to low accuracy as well. The recognition of MGSFA-Net can reach 85.66%, exceeding other methods significantly. The recall and precision can reach more than 80% for all three types of targets. Due to the low resolution of OpenSARShip dataset, BulkCarrier, and ContainerShip targets exhibit similar characterization on SAR images with similar texture structure, making these two types of targets easy to misclassify. The proposed method can better characterize the global features of the ship target while extracting the texture features. Combined with the scattering information, MGSFA-Net can better demonstrate the intrinsic features of the target and, thus, show better recognition performance on SAR ship images.

4) Comparison of Running Time: In order to compare the running time of the different network models, the training time and inference time per epoch are calculated and plotted as curves. The results of the running time are shown in Fig. 12. Since the sample size of OpenSARship is larger in this experiment, it shows longer training and inference time per epoch on OpenSARship. The curves for training and inference time show that the DenseNet and Inceptionv3 have the longest training and inference time per epoch. In our training operation environment, the training time per epoch of DenseNet and Inceptionv3 is more than 20 s. In addition, the methods related to the transformer are more complex, and most of them also show longer training and inference time. Compared to these methods, the MGSFA-Net method proposed in this article shows better performance in terms of time efficiency of training and inference. The results show that the training time per epoch of the MGSFA-Net is about 2 s and the inference time is about 1 s. Also, the methods fusing traditional features and deep features may not have a burden on running time. In conclusion, the proposed method can improve the accuracy of recognition without reducing the time efficiency too much.

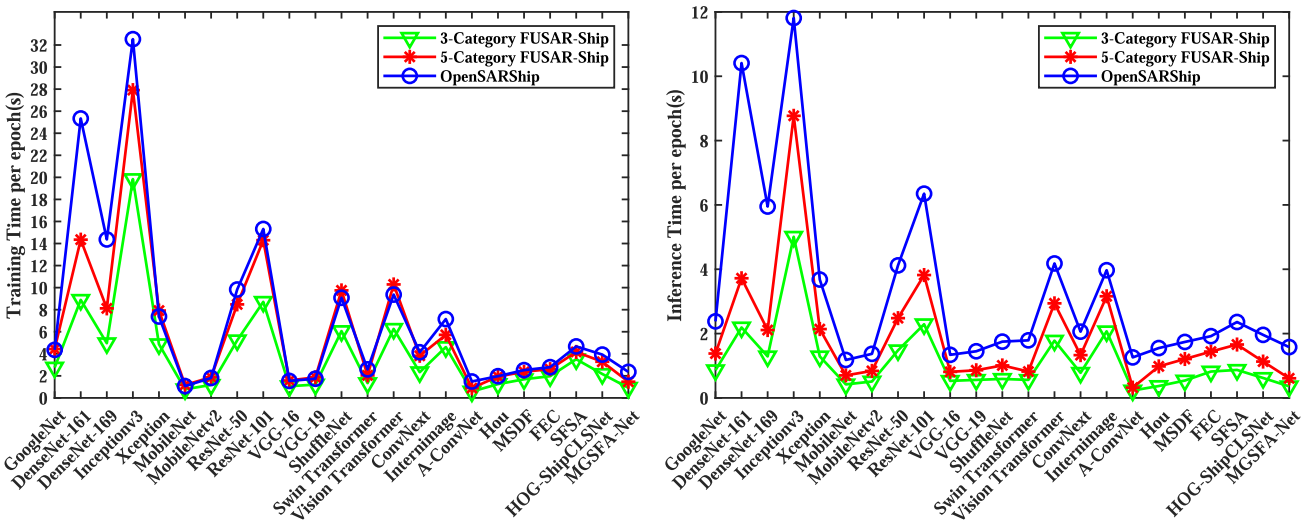
F. Recognition Performance on Few-Shot Condition

Due to the uneven distribution of the number of different categories in the SAR ship target dataset and the frequent occurrence of an insufficient amount of sample data, the few-shot conditions

TABLE IX
EXPERIMENTAL RESULTS OF DIFFERENT METHODOLOGY ON FUSAR-SHIP AND OPENSARSHIP DATASETS

Methods	Three-Category FUSAR-Ship				Five-Category FUSAR-Ship				OpenSARShip			
	P(%)	R(%)	F1(%)	Acc(%)	P(%)	R(%)	F1(%)	Acc(%)	P(%)	R(%)	F1(%)	Acc(%)
GoogleNet [38]	82.57	80.41	81.48	80.06	79.42	77.90	78.65	78.32	77.83	77.17	77.50	80.12
DenseNet-161 [39]	80.45	80.25	80.35	79.47	78.58	78.07	78.32	77.41	77.50	82.70	80.02	80.01
DenseNet-169 [39]	82.61	80.49	81.54	80.34	79.54	77.29	78.40	77.05	80.73	82.67	81.68	82.46
Inceptionv3 [19]	80.78	80.53	80.64	79.19	76.28	75.70	75.99	76.01	71.90	75.90	73.85	74.58
Xception [40]	79.95	80.15	80.05	78.90	78.37	79.80	79.08	78.28	77.13	81.53	79.27	80.62
MobileNet [41]	78.37	78.30	78.34	77.46	74.15	71.70	72.91	74.08	72.80	77.99	75.31	76.61
MobileNetv2 [42]	82.79	80.02	81.38	80.35	75.22	75.96	75.59	75.83	80.10	76.67	78.35	78.83
ResNet-50 [17]	83.34	76.17	79.60	78.32	73.34	75.90	74.60	74.43	76.47	77.26	76.86	78.03
ResNet-101 [17]	79.44	78.77	79.10	77.45	72.86	74.45	73.65	73.73	78.93	76.70	77.80	78.58
VGG-16 [18]	84.57	83.79	84.17	83.24	81.07	82.23	81.65	81.78	75.97	79.77	77.82	79.63
VGG-19 [18]	82.69	81.99	82.34	81.50	81.24	81.57	81.41	80.91	75.43	79.63	77.48	78.76
ShuffleNet [43]	81.90	81.92	81.91	80.64	79.24	78.11	78.67	79.16	72.37	78.10	75.12	76.55
Swin Transformer [46]	76.80	74.27	75.51	74.27	62.22	61.94	62.08	62.52	70.23	72.67	71.14	72.80
Vision Transformer [47]	69.65	70.60	70.20	69.65	55.30	57.54	56.39	54.12	70.53	73.77	72.11	72.92
ConvNext [48]	77.43	75.70	76.56	75.14	65.28	64.70	64.99	65.67	67.43	69.83	68.61	70.89
Internimage[49]	76.07	77.73	76.89	75.72	69.16	69.98	69.57	70.40	76.77	73.80	75.25	75.63
A-ConvNet [20]	82.97	82.60	82.78	81.79	80.56	81.28	80.92	80.39	73.80	78.10	75.89	76.98
Hou [6]	78.08	76.59	77.32	75.72	74.47	69.09	71.67	71.45	72.39	70.50	71.43	71.07
MSDF [29]	84.27	83.70	83.98	82.95	80.06	80.88	80.47	81.09	77.70	82.03	79.81	81.48
FEC [30]	84.70	84.23	84.47	83.25	81.60	84.60	82.81	82.84	81.30	83.53	82.40	83.36
SFSA [36]	82.38	81.42	81.89	80.35	78.77	80.23	79.49	78.98	78.30	83.06	80.61	81.78
HOG-ShipCLSNet [26]	83.63	83.36	83.49	82.08	80.42	79.95	80.18	80.39	78.90	83.26	81.02	82.27
MGSFA-Net	**87.52**	**84.87**	**86.17**	**85.26**	**83.79**	**84.94**	**84.36**	**84.24**	**83.86**	**85.97**	**84.90**	**85.66**

The bold values represent the best result.



(a) (b)

Fig. 12. Curves of running time under different methods on FUSAR-Ship dataset and OpenSARShip dataset. (a) Training time of different methods. (b) Testing time of different methods.

on five-category FUSAR-ship are designed to analyze the performance of the algorithm on condition of few-shot. The number of training samples is set to 10, 20, 30, 40, 50, 80, and 100, and the remaining samples are used as the test set. The results are shown in Table X.

When the number of target samples is 10 and 20, the recognition rates obtained by MGSFA-Net can reach 62.05% and 69.78%, respectively, which are significantly higher than other methods. As the number of samples gradually increases, the recognition accuracy also increases, but the growth rate gradually becomes slower. When the number of samples is 100, the recognition accuracy of the proposed method can reach 81.15%, while most of the other methods are less than 80%. Compared to pure deep learning methods, introducing scattering features into deep learning networks can better characterize the intrinsic essential features of the target and extract more comprehensive features. The experimental results show that under the condition of fewer samples, the recognition performance of the method proposed in this article has a significant advantage over classical deep learning methods and other feature fusion methods. Scattering features contain richer feature information. In the case of the few-shot tasks, the feature representation can be enhanced

TABLE X
ACCURACY OF DIFFERENT TRAINING SAMPLES ON FIVE-CATEGORY FUSAR-SHIP

Methods	Accuracy of different training samples(%)						
	10	20	30	40	50	80	100
GoogleNet [38]	57.91	66.17	67.29	68.86	73.72	74.97	77.95
DenseNet-161 [39]	60.97	65.27	67.86	68.82	73.02	75.87	76.31
DenseNet-169 [39]	59.12	64.84	67.25	69.14	73.26	75.34	76.79
Inceptionv3 [19]	58.44	62.15	66.42	68.51	71.86	74.23	74.87
Xception [40]	57.98	65.43	68.83	71.07	75.63	76.65	77.96
MobileNet [41]	56.25	61.34	65.83	67.38	66.19	71.01	72.47
MobileNetv2 [42]	55.78	63.72	69.17	70.38	69.75	72.86	74.54
ResNet-50 [17]	52.55	61.59	66.18	62.44	67.31	72.34	72.63
ResNet-101 [17]	51.25	61.47	68.53	69.01	70.33	72.36	73.83
VGG-16 [18]	61.68	66.83	71.05	72.79	75.87	78.37	80.51
VGG-19 [18]	57.74	64.21	70.37	73.42	76.19	78.17	79.84
ShuffleNet [43]	59.32	67.22	72.14	74.22	75.91	76.55	77.34
A-ConvNet [20]	55.61	62.57	67.29	68.51	72.27	76.33	78.56
Hou [6]	52.18	55.62	59.37	62.53	64.28	67.46	69.32
MSDF [29]	60.26	68.92	69.34	72.44	75.99	77.62	79.27
FEC [30]	61.77	68.14	71.68	**74.91**	76.01	78.01	79.11
SFSA [36]	57.29	64.72	68.01	71.38	72.27	75.64	76.28
HOG-ShipCLSNet [26]	59.81	66.25	70.18	72.05	73.26	76.04	77.16
MGSFA-Net	**62.05**	**69.78**	**72.35**	74.30	**76.37**	**78.94**	**81.15**

The bold values represent the best result.

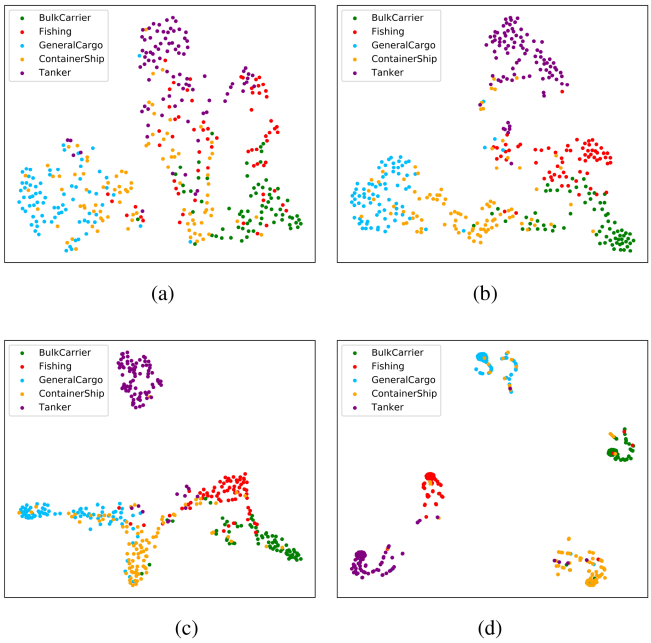


Fig. 13. Visualization of the features extracted by different methods. (a) A-ConvNet. (b) VGG-16. (c) SFSA. (d) MGSFA-Net.

to complement the features in the deep networks, preventing the overfitting phenomenon caused by the small size of training data.

G. T-SNE Feature Visualization

The t-distributed stochastic neighborhood embedding (t-SNE) visualization method is used to further prove the effectiveness of our proposed method. The method can narrow down the features to two dimensions, and the results of the feature distribution on five-category FUSAR-Ship are shown in Fig. 13. From the figure, it can be seen that the distances of features extracted by classical deep networks such as A-ConvNet and VGGNet are similar, which means that the features are hard to be separated and are easily confused. The results show that although the features extracted by the deep network can achieve a high recognition rate, the 2-D feature distribution of the deep features is similar and cannot be well distinguished. Compared with the two classical methods, the SFSA method can expand the difference of features between the categories, indicating that the separability of features is expanded with the introduction of scattering features. However, some categories of features extracted by SFSA can be confused to some extent. Compared with these methods, the features extracted by MGSFA-Net have the best feature separability with more obvious interclass differences and smaller intraclass differences. Therefore, the features extracted by the proposed method can reflect the intrinsic characteristics of the target better and are more conducive to target recognition.

H. Grad-CAM Map Visualization

To analyze and validate the specific regions of interest of the network in this article further, Grad-CAM maps [45] are used to visualize the regions of interest of different methods, as shown in Fig. 14. The region of interest of the network is analyzed by comparing the methods of VGGNet, A-ConvNet, SFSA, and MGSFA-Net. From the visualization of Grad-CAM maps, it can be seen that the compared methods, such as VGGNet, have a larger region of heat map and a wider area of interest, which cannot learn the internal features of the target region in a targeted way. These methods cannot focus on the target region and will be affected by the background region. In addition, these methods may increase the redundant features in the network, which is not conducive to the feature expression and will seriously affect the feature generalization ability. In contrast, the Grad-CAM map of MGSFA-Net in this article has a smaller area of interest and the area of interest of MGSFA-Net has a better fit to the contour of the target. Moreover, the network can pay more attention to the global structure of the target, focus more on the target area,

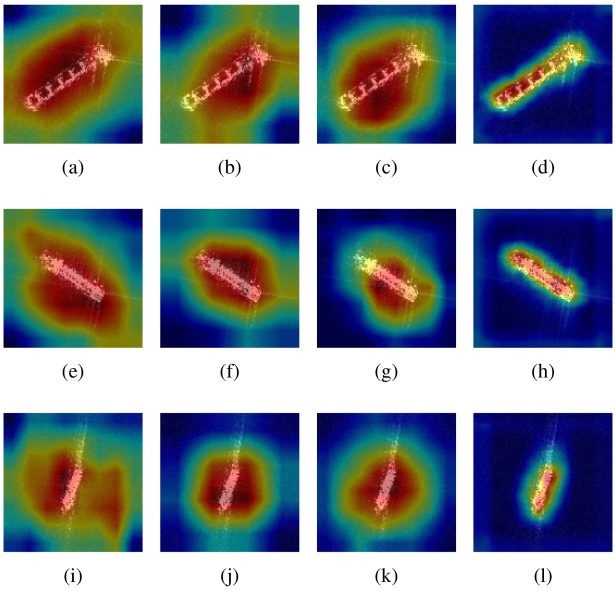


Fig. 14. Visualization of interest region of different categories under different methods. (a)–(d) BulkCarrier. (e)–(h) ContainerShip. (i)–(l) GeneralCargo. (a), (e), (i) visualization of A-ConvNets. (b), (f), (j) visualization of VGGNet. (c), (g), (k) visualization of SFSA. (d), (h), (l) visualization of MGSFA-Net.

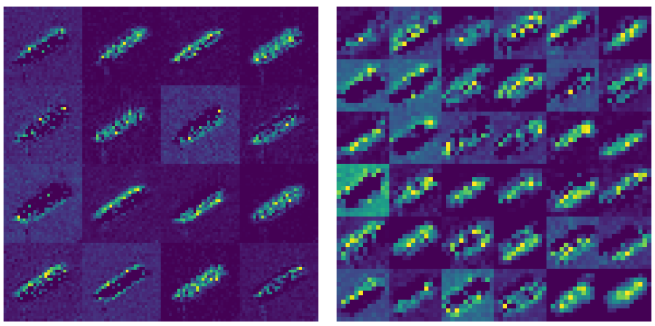


Fig. 15. Feature map of MGSFA-Net of different layers. (a) Convolution layer3. (b) Convolution layer4.

and characterize the target structural features, which can reduce the influence of the large difference in the size of the target. For targets with different sizes, such as BulkCarrier, ContainerShip, and GeneralCargo, the networks proposed in this article can efficiently and accurately focus on the target area, demonstrating better multiscale characteristics. Furthermore, it also shows that the method in this article can associate the extracted SCs to produce more efficient recognition features to characterize the global scattering structure, which can combine the scattering features effectively and verify the effectiveness of the proposed method.

I. Feature Map Visualization

This section shows the feature map produced by the MKDFE module of the MGSFA-Net in Fig. 15. The BulkCarrier category in FUSAR-Ship is taken as an example to analyze the results of the convolution layer 3 and convolution layer 4. From the results of the feature map, it can be seen that the target region is represented in the network as pixel dots with different brightnesses, where brighter colors indicate higher weights and importance in the network, and darker colors indicate lower weights in the network. The global structure of the target is composed of different pixel points. The feature map of MGSFA-Net shows that the algorithm can completely separate the target region from the background region so that the internal features of the target can be better learned. In addition, it shows that the MKDFE module can characterize the global features of the target in a better way, and the global structural features of the target can also be learned even in the high-dimensional convolution layer. As can be seen from the feature map, the method in this article can separate the ship target from the background region and highlight the global structure of the target more obviously. Moreover, the design of the module can reduce the effect of different target sizes and minimize the negative impact of a single convolution kernel when extracting targets with large sizes.

IV. Conclusion

Considering the complexity of the scattering structure of SAR ship targets and the large difference in target sizes, a multiscale global scattering feature association network for SAR ship target recognition is proposed. First, the ship target is separated from the background interference by fine segmentation. Then, the SCs are extracted and converted to graph structures, which are associated by the SCFA module and enhanced by the MSFE module to produce the multiscale global scattering features. In addition, the high-dimensional deep features of the targets are extracted by the MKDFE module. Finally, the scattering features and the deep features are fused to enhance the feature diversity. The results of recognition show that the MGSFA-Net can achieve excellent performance on the three-category experiment dataset and five-category experiment dataset, even on a few-shot condition. In addition, the results of t-SNE visualization show that the MGSFA-Net can characterize the intrinsic features of targets more accurately and expand the distance of features with different categories. Moreover, the results of the feature map and Grad-CAM visualization explain that the MGSFA-Net can associate the local structure to characterize the global features of targets and focus more on the region of targets. The research in this article can support the interpretability of deep learning studies and the analysis of target characteristics. In the future, we will be more committed to studying the scattering characteristics of various types of targets, and further investigate more robust target recognition algorithms combined with diverse features of SAR targets.

References

[1] C. Zhang et al., "Evaluation and improvement of generalization performance of SAR ship recognition algorithms," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9311–9326, 2022.

[2] Z. Sun, X. Leng, Y. Lei, B. Xiong, K. Ji, and G. Kuang, "BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images," *Remote Sens.*, vol. 13, no. 21, 2021, Art. no. 4209.

[3] S. Emanuele, "Using low-resolution SAR scattering features for ship classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–4, 2022.

[4] H. Lang, J. Zhang, X. Zhang, and J. Meng, "Ship classification in SAR image by joint feature and classifier selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 2, pp. 212–216, Feb. 2016.

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

[6] 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.

[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 886–893.

[8] 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.

[9] H. Lang, J. Zhang, T. Zhang, D. Zhao, and J. Meng, "Hierarchical ship detection and recognition with high-resolution polarimetric synthetic aperture radar imagery," *J. Appl. Remote Sens.*, vol. 8, no. 1, 2014, Art. no. 083623.

[10] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[11] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[12] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.

[13] Z. Zhao, K. Ji, X. Xing, W. Chen, and H. Zou, "Ship classification with high resolution TerraSAR-X imagery based on analytic hierarchy process," *Int. J. Antennas Propag.*, vol. 2013, 2013, Art. no. 698370.

[14] 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.

[15] H. Lang and S. Wu, "Ship classification in moderate-resolution SAR image by naive geometric features-combined multiple kernel learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1765–1769, Oct. 2017.

[16] Y. Xu and H. Lang, "Distribution shift metric learning for fine-grained ship classification in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2276–2285, 2020.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

[19] C. Szegedy, V. Vanhoucke, S. Ioffe, and J. Shlens, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.

[20] 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.

[21] C. Dechesne, S. Lefèvre, R. Vadaine, G. Hajduch, and R. Fablet, "Ship identification and characterization in Sentinel-1 SAR images with multitask deep learning," *Remote Sens.*, vol. 11, no. 24, Dec. 2019, Art. no. 2997.

[22] Y. Zhang, Z. Lei, H. Yu, and L. Zhuang, "Imbalanced high-resolution SAR ship recognition method based on a lightweight CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[23] 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, Sep. 2018, Art. no. 2929.

[24] C. Lu and W. Li, "Ship classification in high-resolution SAR images via transfer learning with small training dataset," *Sensors*, vol. 19, no. 1, p. 63, Dec. 2018.

[25] 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.

[26] T. Zhang et al., "HOG-ShipCLSNet: A novel deep learning network with HOG feature fusion for SAR ship classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–22, Jun. 2021.

[27] L. Zeng et al., "Dual-polarized SAR ship grained classification based on CNN with hybrid channel feature loss," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.

[28] S. Feng, K. Ji, F. Wang, L. Zhang, X. Ma, and G. Kuang, "Electromagnetic scattering feature (ESF) module embedded network based on ASC model for robust and interpretable SAR ATR," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.

[29] Z. Liu, L. Wang, Z. Wen, K. Li, and Q. Pan, "Multilevel scattering center and deep feature fusion learning framework for SAR target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5227914.

[30] 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.

[31] B. Ding, G. Wen, C. Ma, and X. Yang, "An efficient and robust framework for SAR target recognition by hierarchically fusing global and local features," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5983–5995, Dec. 2018.

[32] J. B. Lee, R. Rossi, and X. Kong, "Graph classification using structural attention," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov., Data Mining*, 2018, pp. 1666–1674.

[33] L. Li, J. Liu, L. Su, C. Ma, B. Li, and Y. Yu, "A novel graph metalearning method for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[34] R. Yang, X. Xu, X. Li, L. Wang, and F. Pu, "Learning relation by graph neural network for SAR image few-shot learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 1743–1746.

[35] C. Li, L. Du, Y. Li, and J. Song, "A novel SAR target recognition method combining electromagnetic scattering information and GCN," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[36] C. Zhao, S. Zhang, R. Luo, S. Feng, and G. Kuang, "Scattering features spatial-structural association network for aircraft recognition in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[37] T. Li and L. Du, "SAR automatic target recognition based on attribute scattering center model and discriminative dictionary learning," *IEEE Sensors J.*, vol. 19, no. 12, pp. 4598–4611, Jun. 2019.

[38] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[39] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.

[40] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1800–1807.

[41] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861

[42] 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.*, 2018, pp. 4510–4520.

[43] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6848–6856.

[44] Y. Wang et al., "Dynamic graph CNN for learning on point clouds," *ACM Trans. Graph.*, vol. 38, no. 5, pp. 1–12, Nov. 2019.

[45] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, pp. 336–359, 2020.

[46] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 9992–10002.

[47] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020. [Online]. Available: http://arxiv.org/abs/2010.11929

[48] Z. Liu, H. Mao, C. -Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 11976–11986.

[49] W. Wang et al., "InternImage: Exploring large-scale vision foundation models with deformable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 14408–14419.

[50] 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.

[51] Y. Shang et al., "SA2Net: Ship augmented attention network for ship recognition in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1–15, 2023.

[52] Y. Zhang, D. Lu, X. Qiu, and F. Li, "Scattering-point topology for few-shot ship classification in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 10326–10343, 2023.

Xianghui Zhang received the B.S. degree in information engineering from the National University of Defense Technology, Changsha, China, in 2022. He is currently working toward the Ph.D. degree in information and communication engineering with the State Key Laboratory of Complex Electromagnetic Environment Effects, National University of Defense Technology.

His research interests include SAR image interpretation, feature extraction, and deep learning.

Sijia Feng received the B.S. degree in information engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2016, and the M.S. and Ph.D. degrees in information and communications engineering from the NUDT in 2018 and 2023, respectively.

She joined the College of Meteorology and Oceanography, NUDT, where she is currently a Lecturer. Her research interests include remote sensing image interpretation, feature extraction, and machine learning.

Chenxi Zhao received the B.S. degree in electronic information engineering and M.S. degree in control science and technology from the Beijing University of Chemical Technology, Beijing, China, in 2018 and 2021, respectively. She is currently working toward the Ph.D. degree in information and communication engineering with the State Key Laboratory of Complex Electromagnetic Environment Effects, National University of Defense Technology, Changsha, China.

Her research interests include SAR image interpretation, feature extraction, and machine learning.

Zhongzhen Sun received the B.S. degree in communication engineering from the National University of Defense Technology, Changsha, China, in 2018, and the M.S. degree in electronic and communication engineering from National University of Defense Technology, Changsha, China, in 2021.

His current research focuses on the Intelligent electronic countermeasures and assessment, SAR image target recognition.

Siqian Zhang received the B.S. degree in electronic information engineering from Hubei University, Wuhan, China, in 2009, the M.S. degree in photogrammetry and remote sensing and the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2011 and 2015, respectively.

She is currently an Associate Professor with the College of Electronic Science and Technology, National University of Defense Technology. Her research interests include radar signal processing, SAR imaging, machine learning, compressed sensing, and sparse representation.

Kefeng Ji (Member, IEEE) received the B.S. degree in aerospace engineering from Northwestern Polytechnical University, Xi'an, China, in 1996, and the M.S. and Ph.D. degrees in information and telecommunication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 1999 and 2003, respectively.

In 2003, he joined the College of Electronic Science and Technology, NUDT, where he is currently a Professor. He has authored or coauthored over 80 papers. His research interests include signal processing, machine learning, pattern recognition, remote sensing information processing, synthetic aperture radar image interpretation, target detection, recognition, feature extraction, and marine surveillance. He is a Reviewer of several international journals and conferences, such as IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, IEEE ACCESS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, *International Journal of Remote Sensing, Remote Sensing Letters, European Journal of Remote Sensing, and Remote Sensing*.