

Scattering Point Topology for Few-Shot Ship Classification in SAR Images

Yipeng Zhang, Dongdong Lu, Xiaolan Qiu and Fei Li

Abstract—Synthetic aperture radar (SAR) has emerged as a critical technology for detecting and classifying objects such as ships in challenging environments. However, few-shot learning remains challenging due to the limited availability of labeled SAR data, complex radar backscatter, and variations in imaging parameters. In this paper, we propose a novel network, scattering point topology for few-shot ship classification (SPT-FSC), which addresses these challenges by incorporating scattering characteristics into the network learning process through a scattering point topology (SPT) based on scattering key points. We design a topology encoding branch (TEB) through a series of operations to encode the topological information of scattering points, resulting in a SPT embedding that improves the network's adaptability to the imaging mechanism and reduces imaging variability in SAR images. To effectively fuse the SPT embedding and image features extracted from a convolutional neural network (CNN), we introduce a novel mechanism named reciprocal feature fusion attention (RFFA). Additionally, to address the limited diversity in the training data, we apply fine-tuning based methodologies and construct a fine-grained ship classification dataset by combining the OpenSARShip and FUSAR-Ship datasets. Our comprehensive experiments on these datasets demonstrate the effectiveness of our proposed SPT-FSC method, achieving high accuracy and robustness in few-shot ship classification tasks for SAR images, outperforming existing methods in this domain.

Keywords: Ship classification; few-shot learning (FSL); scattering topology; scattering point; synthetic aperture radar (SAR)

This paragraph of the first footnote will contain the date on which you submitted your paper for review, which is populated by IEEE. It is IEEE style to display support information, including sponsor and financial support acknowledgment, here and not in an acknowledgment section at the end of the article. This work was supported by the National Natural Science Foundation of China (No. 61991421 and No. 62022082. (*Corresponding author: Dongdong Lu.*))

Yipeng Zhang, Xiaolan Qiu and Fei Li are with the Suzhou Key Laboratory of Microwave Imaging, Processing and Application Technology, Suzhou 215128, China, with the Suzhou Aerospace Information Research Institute, Chinese Academy of Sciences, Suzhou 215128, China, with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, with the University of Chinese Academy of Sciences, Beijing 100190, China, and also with School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100190, China (e-mail: zhangyipeng21@mails.ucas.ac.cn; xlqiu@mail.ie.ac.cn; lifei@mail.ie.ac.cn).

Dongdong Lu is with the Suzhou Key Laboratory of Microwave Imaging, Processing and Application Technology, Suzhou 215128, China, with the Suzhou Aerospace Information Research Institute, Chinese Academy of Sciences, Suzhou 215128, China (e-mail: ludongdong@tju.edu.cn).

Color versions of one or more of the Fig.s in this article are available online at <http://ieeexplore.ieee.org>

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) is a powerful remote sensing technology that uses radar signals to create high-resolution images of the earth's surface, with the ability to penetrate through clouds and vegetation for detection and classification of objects such as ships, vehicles, and buildings [1]. SAR has become an important tool for applications such as maritime surveillance, border control, and disaster response, where its ability to detect and classify targets in challenging environments is crucial for effective decision-making [2].

Ship classification of SAR images is a challenging task due to the complexity of radar backscatter from ships and the presence of environmental clutter. Various techniques have been proposed for ship classification, including statistical analysis, texture analysis, and machine learning algorithms. At the same time, different features such as shape, size, texture, and polarimetric characteristics have been used to discriminate between different types of ships, including cargo ships, oil tankers, and other vessels [2]. Statistical methods such as principal component analysis (PCA) and independent component analysis (ICA) have been applied to extract discriminative features from SAR images for ship classification [3]. Texture analysis methods such as gray level cooccurrence matrix (GLCM) and local binary pattern (LBP) have also been used to capture the spatial variation of backscattering signals from ships [4]. However, these conventional methods are limited in their ability to handle the complex nature of SAR backscattering signals and the variability of environmental conditions, leading to reduced accuracy and robustness for ship classification in SAR images [5]. Machine learning algorithms such as support vector machines (SVM), k-nearest neighbors (KNN), and random forests (RF) have been applied to automatically classify ships in SAR images with high accuracy [6]-[8]. Some studies have also explored the use of hybrid approaches that combine multiple machine learning algorithms for ship classification, such as feature selection followed by SVM classification [9].

In recent years, deep learning methods such as convolutional neural networks (CNNs) have shown promising results for ship classification in SAR images, achieving high accuracy and robustness to variations in ship size, shape, and orientation [10]. However, the limited availability of labeled SAR data remains a bottleneck for further progress in this field. Traditionally, deep learning models for SAR image classification require a large amount of labeled data to achieve high performance. However, acquiring labeled data for SAR

images is time-consuming and expensive, which limits the applicability of deep learning models in practical scenarios. In SAR image classification, where the datasets are often limited and imbalanced, few-shot learning offers an effective solution to enhance classification model accuracy and robustness. Therefore, few-shot learning has emerged as a promising solution to address this issue by enabling deep learning models to classify SAR images accurately with a limited amount of labeled data [11], [12].

Few-shot learning is a type of machine learning that aims to train models with the ability to learn from a small number of examples [13]. By leveraging prior knowledge or meta-learning, few-shot learning algorithms can effectively generalize to new tasks with limited labeled data [14]. In the context of SAR image classification, few-shot learning approaches have been proposed to extract and leverage the underlying features of SAR images, which can be used to classify images with a limited number of labeled examples [15]. Recently, various few-shot learning methods have been developed, including data augmentation, meta-learning, and transfer learning. Data augmentation is a technique that generates additional training data by applying various transformations to the existing data [16]. In few-shot learning, data augmentation can be used to generate more labeled examples from the limited training data. This approach has been shown to improve the performance of few-shot learning methods in various applications [17]. Meta-learning is another popular few-shot learning approach that aims to learn how to learn from a few examples [18]. It involves training a model on multiple tasks and using the learned knowledge to quickly adapt to new tasks with only a few labeled examples. Meta-learning has been successfully applied in various computer vision applications, such as object recognition and segmentation [19]. Transfer learning is a technique that involves transferring knowledge from a pre-trained model to a new task [20]. In few-shot learning, transfer learning can be used to leverage the knowledge learned from a large dataset to improve the performance on a few-shot learning task with limited labeled data. Transfer learning has been widely used in various computer vision applications, including few-shot learning [21].

Few-shot learning has gained significant attention in recent years as it offers a promising approach to tackle challenges in various computer vision tasks. In the realm of image classification, few-shot learning techniques have shown remarkable adaptability to recognize objects or categories with limited labeled data. This capability is particularly valuable in domains where acquiring extensive labeled data is costly or impractical. Furthermore, few-shot learning is not limited to image classification but also extends its potential to image detection and segmentation tasks. In the field of image detection, few-shot learning empowers models to identify objects or regions of interest in images with only a few annotated instances, which is crucial for applications such as object tracking and surveillance. And few-shot object detection based on fine-tuning has received widespread attention in recent years, such as TFA [22], FSCE [23] and DeFRCN [24]. Additionally, in image segmentation, where the goal is to

segment objects or regions within an image, few-shot learning techniques have demonstrated their ability to adapt and generalize with limited labeled data [25].

Despite the success of existing few-shot learning methods, there are still several problems and challenges that need to be addressed for SAR images. Firstly, the imaging mechanism of SAR images is different from optical images, resulting in different feature representations and classification challenges [26]. Unlike optical images, SAR images contain multiple scattering effects and speckle noise, which can obscure the underlying features of the image [27]. The difference in the imaging mechanism and the characteristics of SAR images require specific adaptations and modifications to existing few-shot classification methods [28]. Secondly, SAR images exhibit significant variability due to changes in imaging parameters [29]. SAR images are affected by various factors, such as the imaging mode, incidence angle, polarization, wavelength and so on, which can result in significant variations in the image features. This variability makes it difficult to learn a robust representation that can generalize to new classes with limited labeled data [30]. Furthermore, the imaging variability of SAR images can also result in high intra-class variance and low inter-class distinction, making it difficult to distinguish between similar classes [31]. This variability can result in a high degree of ambiguity in few-shot classification, leading to poor performance. Thirdly, the training data of few-shot classification in SAR images lacks diversity [27]. SAR images are typically acquired over specific regions and at specific times, which can limit the diversity of the labeled data. When training a few-shot classifier on a small number of classes, the classifier may not be able to generalize well to new classes that are not similar to the training classes. This lack of diversity can lead to overfitting and poor generalization performance of few-shot classification models.

To address these challenges, a new network SPT-FSC is proposed in this article for few-shot ship classification in SAR images. Different from previous methods of directly applying the deep network model to the SAR images, SPT-FSC integrates the scattering characteristic into the network learning process. We propose scattering point topology (SPT) based on scattering key points [32], and a SPT embedding which introduces information on additional features to the network. Meanwhile, SPT features and the image features extracted by the CNN belong to different modalities, so we need to better integrate the two features. Consequently, we introduce the reciprocal feature fusion attention (RFFA) to comprehensively integrate two distinct sets of features. Furthermore, to alleviate the issue of limited diversity in the training data of SAR images, we apply fine-tuning based methodologies to exploit the knowledge gained from a larger dataset. In this regard, we amalgamate the OpenSARShip [33] dataset and the FUSAR-Ship [34] dataset to generate a ship classification dataset, encompassing an expanded array of categories and an augmented sample size.

The main contributions of our work can be summarized as follows:

- 1) A SPT method is proposed based on the scattering key

points, which incorporates not only the position information of the scattering key points but also the distance information between the scattering points and the inherent information of each individual scattering point;

- 2) SPT-FSC employs an effective topology encoding branch (TEB) to encode the topological information of scattering points to derive the SPT embedding. This approach was employed in few-shot classification to enhance the network's adaptability to the imaging mechanism and more effectively alleviate the imaging variability inherent in SAR images;
- 3) To achieve effective fusion of the SPT embedding and image features obtained from the CNN, a mechanism named RFFA was designed. RFFA is a mechanism that amplifies the representational capacity of neural networks by enabling them to learn from both directions of the input sequence;

Extensive studies have been conducted on OpenSARShip and FUSAR-Ship to validate the effectiveness of the proposed method. In comparison to the baseline, the accuracy of 4-way 1-shot and 4-way 5-shot tasks is improved by 11.90% and 25.15% on OpenSARShip dataset, respectively, and the accuracy of 5-way 1-shot and 5-way 5-shot tasks is increased by 24.08% and 30.72% on FUSAR-Ship dataset, respectively.

II. RELATED WORKS

A. Few-Shot Classification Methods

Few-shot classification is the task of learning to classify new categories with only a few labeled examples per category. There are several methods that have been proposed to tackle this problem. Meta-learning approaches include MAML [18], Reptile [19], and Prototypical Networks [14]. MAML learns a set of initial weights that can be quickly adapted to new tasks with few examples. Reptile performs a similar task but uses a simpler optimization method that is faster to compute. Prototypical Networks learn a prototype for each class and classify new examples based on their distance to the prototypes. Matching Networks [35], Relation Networks [36], and Siamese Networks [37] are examples of metric learning approaches. These approaches learn a distance metric between examples and use this metric to classify new examples. Matching Networks use a memory-augmented architecture to learn a similarity function between examples. Relation Networks learn a relation module that can compare pairs of examples and classify new examples based on their relation to the labeled examples. Siamese Networks learn a shared representation between pairs of examples and classify new examples based on their distance in this shared representation. DynaGAN [38], D2C [39], and ProtoGANs [40] are examples of techniques that use generative models. These methods learn to generate new examples for unseen categories based on the labeled examples. DynaGAN presents a few-shot domain adaptation method for multiple target domains, which utilizes an adaptation module to dynamically adapt a pretrained GAN model to these domains. D2C introduces a paradigm for training unconditional VAEs for

few-shot conditional image generation, which leverages diffusion-based priors and contrastive self-supervised learning to adapt to novel tasks with minimal labeled examples. ProtoGANs combine the prototype learning approach of Prototypical Networks with a generative model to generate new examples for unseen categories. Transfer learning-based methods have shown promising results in addressing this problem. Baseline++ [41] is a state-of-the-art transfer learning-based approach that aims to improve upon the baseline model by fine-tuning pre-trained models on a small labeled support set. It also utilizes the cosine similarity metric to enhance the distance metric learning process. Meta-Baseline [42] is another transfer learning-based method that utilizes meta-learning to improve few-shot classification performance. It learns to adapt to new tasks by learning a set of parameters that can be fine-tuned quickly on new tasks. It also incorporates an auxiliary loss function to encourage the model to learn more discriminative features.

Given the outstanding performance observed in few-shot learning methods relying on transfer learning, we opted for a fine-tuning-based approach in our few-shot ship classification method. Our model's baseline aligns with the Baseline++ framework, excluding the utilization of a cosine classifier.

B. Few-Shot Classification in SAR images

Few-shot classification in SAR images has been the subject of considerable research in recent years. Various approaches have been proposed to tackle this problem. One of the methods for few-shot SAR image classification is the instance-aware transformer (IAT) model [43]. The IAT model incorporates instance-level information into the classification process, which can enhance the performance of few-shot classification in SAR images. Another approach involves utilizing a weighted distance and feature fusion strategy [44]. The similarity between a query image and support images is evaluated using a distance metric through this method, and the accuracy of few-shot classification is improved by combining the features extracted from both the query and support images. The technique proposed for few-shot SAR image classification, known as the spatial transformed prototypical network (ST-PN), has been introduced in the network [45], which incorporates spatial transformations in the prototypical network to enhance its ability to recognize new classes. A few-shot learning framework is based on the prototypical network with a limited number of training samples [46], allowing it to learn a classifier capable of recognizing new classes with only a limited number of examples. A technique utilizing a meta-learning approach with amortized variational inference has been proposed [47], which involves a meta-learning framework to train a model capable of quickly adapting to new classes with only a few examples. The proposed method for few-shot classification in SAR images employs a metric learning approach that utilizes a siamese network to train a similarity metric [48]. Deep transfer learning has also been applied to few-shot SAR image classification [49]. For the identification of few-shot SAR targets, Lu et al. presented a deep convolutional neural network based on transfer learning [50]. They utilized

the refined model to categorize new cases after fine-tuning a pre-trained network model on a few samples from new classes. Disentangled attention modules were used by Tai et al. to selectively transfer characteristics from electro-optical (EO) samples to SAR samples, eliminating the need for additional SAR samples [51]. Additionally, in the current landscape, a number of deep learning-based methods integrate object detection and classification. Several few-shot object detection methods for SAR images also include classification modules. Chen et al. [52] presented a novel few-shot SAR object detection framework based on meta-learning, utilizing an attention mechanism to emphasize class-specific features and a dynamic relationship learning paradigm involving graph convolutional networks and orthogonality constraints to enhance feature similarity. Zhou et al. [53] proposed a novel FSODM method for optical remote sensing that includes a lightweight meta-feature extractor (DarknetS) and an aggregation module (AggregationS) to improve SAR image feature representation and generalization for new classes.

Nevertheless, the majority of prior model optimization research has primarily utilized general models designed for natural scenes, without taking into account the distinct imaging characteristics inherent to SAR and optical images. Our method innovatively designs a SPT, thereby uniquely integrating the scattering characteristics of SAR in the learning process of the network.

C. Scattering Key Points for Deep Neural Network

The concept of key points in deep neural networks has emerged as a pivotal aspect of modern computer vision and machine learning research. Key points, often referred to as keypoints or landmarks, represent specific spatial locations in an image or feature map. They serve as salient reference points, capturing essential information about the underlying structure and content of an image. While the idea of key points has been present in computer vision for decades, the advent of deep learning has revitalized its relevance and utility in transformative ways. One of the primary roles of key points in deep neural networks is feature extraction. CNNs use learned filters to identify these key points, enabling the networks to recognize patterns, objects, or structures within the data. This has led to remarkable progress in image classification, object detection, and segmentation tasks. Key points also facilitate spatial understanding, enabling models to infer the relative positions and orientations of objects in an image.

The integration of scattering key points into deep neural networks for the analysis of SAR images represents a significant breakthrough in remote sensing and computer vision. SAR images, characterized by their complex and texture-rich content, have long posed unique challenges for traditional computer vision approaches. The introduction of scattering key points marks a promising shift towards harnessing the full potential of deep neural networks in SAR image analysis. These key points serve as salient landmarks, allowing deep neural networks to effectively understand, classify, and interpret the content of SAR images. The use of scattering key points has led to remarkable strides in object

detection and recognition within SAR images.

Recent studies [32], [54]-[57] have proposed the use of a set of key points to model local scattering regions for guiding networks in SAR object detection and classification. An anchor-free network was proposed by Fu et al. [54] that uses scattering key points to guide network training and employs predicted scattering key point positions as offsets for deformable convolution modules. It has been found that the scattering key points obtained using the Harris corner detector roughly capture the structural characteristics and capture the distinguishing features of ships. The proposed SFR-Net [55] aims to address the completeness issue of aircraft detection results in SAR images by utilizing the scattering point relation module (SPRM) to analyze and correlate discrete scattering points. The SPRM enables the extraction of characteristics and establishment of connections among scattering points, thereby overcoming the discreteness issue inherent to individual aircraft in SAR imaging. Sun et al. [56] proposes a novel unified framework named SPAN for accurately locating and classifying ships in large-scale SAR images by capturing the distribution characteristics of strong scattering points in the ship area. SPAN utilizes a ship classification encoder module to extract the correlation and distribution characteristics between each scattering point, and combines the features and distribution information of strong scattering points to recognize the ship category. A novel integrated framework named SCAN [57] was proposed for few-shot SAR aircraft classification, which includes a classification path and a scattering extraction branch that utilizes a scattering extraction module (SEM) to guide the network to learn the number and distribution of strong scattering points for each target type. The scattering extraction branch employs a SEM to improve the SAR target feature representation and optimize the main classification task, which leads to better performance on novel category few-shot recognition tasks. The proposed SPG-OSD [32] method for ship detection in SAR images incorporates the scattering characteristics of SAR images to guide the network through an oriented two-stage detection module and a scattering-point-guided RPN. The innovative use of key scattering points in the RPN effectively reduces false alarms and missing ships, while the RPN predicts the position of these points and utilizes location information to extract features near them.

Most of the aforementioned research endeavors have leveraged scattering key points to facilitate network learning of pertinent features and enhance precise localization. However, it is imperative to acknowledge that, in practice, the distribution and spatial interrelationships among scattering key points can divulge pivotal insights into the fundamental shape, boundaries, and internal composition of the target. Consequently, we introduce a topological structure founded upon scattering key points, aiming to elevate the representation of ships' structural characteristics. This enriched feature representation is subsequently applied within the domain of few-shot ship classification.

III. METHODOLOGY

A. Problem Setup

In standard few-shot classification, the objective is to acquire knowledge of new concepts within a set of novel classes C_{novel} , with only a limited number of samples, while having access to a labeled dataset of base classes C_{base} containing a substantial number of images, where $C_{base} \cap C_{novel} = \emptyset$. The test set is divided into two distinct subsets, namely the support set and the query set. The support set comprises of a limited number of annotated samples belonging to the newly introduced classes, while the query set contains unlabeled samples of the same label space. The main goal of few-shot classification is to precisely classify unlabeled query samples based on the information provided in the support set. Specifically, when the support set contains N classes and each class is represented by K labeled samples, the problem is referred to as N -way K -shot few-shot classification. At the same time, the query set contains Q samples per class drawn from the same N classes, and the objective is to classify the $N \times Q$ query images accurately into their respective N classes.

In our setup, we use three separate datasets: a training set, a support set, and a testing set. The label space for the support set and testing set is identical, while the training set has its own unique label space that is completely separate from both the support and testing sets. During the training phase, our training method is exactly the same as normal image classification, as shown in Fig. 1. It refers to training a classifier f_{θ} with standard cross-entropy loss on all base classes. To accommodate the recognition of novel classes in the test stage, we replace the

classification head of the model. Specifically,

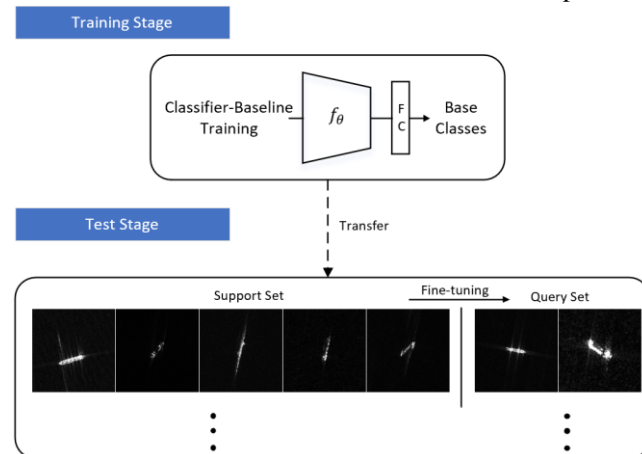


Fig. 1. Training and testing of few-shot classification in our setup

the original classification head used during training on base classes is replaced with a new one designed to handle unseen categories. This replacement ensures that the model's output aligns with the novel class labels during testing, facilitating accurate classification. To make the model capable of recognizing novel classes during testing, we fine-tune it using the $N \times K$ images from the support set. This fine-tuning process involves updating the model's parameters to minimize the loss, allowing it to become more attuned to the novel classes' characteristics. Notably, both the backbone part and the TEB remain frozen during the fine-tuning process. This means that only the fully connected layer is updated using the support set images from the novel classes, allowing it to adapt to the unique characteristics of these classes. Subsequently, we evaluate the

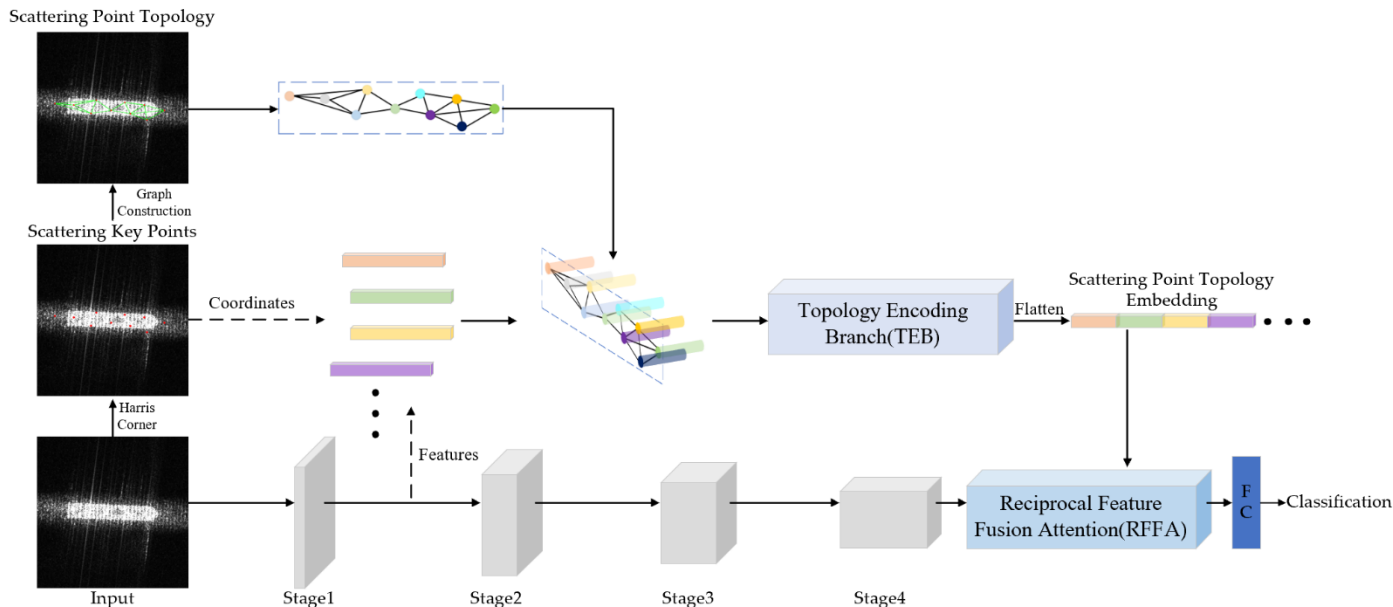


Fig. 2. Overall architecture of SPT-FSL. First, we extract the scattering key points from the objects and establish the scattering point topology. Next, we use the positions of these scattering key points to obtain feature vectors for corresponding nodes from Stage 1 of the backbone, and feed them into the GCN and MLP-mixer. Finally, the RFFA fuses the SPT embedding with features extracted by the CNN to generate the final prediction.

model's performance on the $N \times Q$ images from the query set using the fine-tuned model. This assessment phase ensures that the model can effectively recognize and classify novel classes based on the adapted knowledge acquired during fine-tuning.

B. Overall Architecture

The overall architecture of the proposed network model primarily consists of three components: the backbone, the topology encoding branch named TEB, and the reciprocal feature fusion attention module named RFFA, as shown in Fig. 2. The backbone of the network is based on ResNet-12 [42], which is composed of four stages, each containing a residual

layers with a 3×3 kernel, followed by Batch Normalization and Leaky ReLU activation with a 0.1 slope. The number of channels in the convolutional layers for each residual block are 64, 128, 320, and 640, respectively. A 2×2 max-pooling layer is applied after each residual block, and a 5×5 global average pooling is used at the end to obtain a 640-dimensional feature vector. This architecture is consistent with recent works and is kept simple without additional modifications such as Drop-Block and wider channels. The TEB includes two layers of graph convolutional networks (GCNs) [58] and an MLP-Mixer [59] layer to extract the SPT embedding. The purpose of this branch is to capture the intricate spatial relationships among the scattering key points in the SPT. The RFFA module is designed to effectively fuse the SPT embedding with the features extracted from the CNN. This fusion allows for better concatenation of the features before they are input into a fully connected layer for predication.

C. Scattering Point Topology (SPT)

The SPT method is developed based on the principle that strong scattering points possess substantial significance and stability, containing an abundance of scattering information. This approach aims to explore the stable scattering characteristics of various targets, such as ships, to better understand and represent their structural and shape features. To initiate this process, the Harris detector [60] is applied to identify applied to identify important points that reflect the structural outline of the target, as illustrated in Fig. 3(c) and (d). This process results in a set of points that capture the shape and structure of the object. Subsequently, we apply the K-means clustering algorithm to classify the extracted points into N distinct clusters with high scattering intensities, as illustrated in Fig. 3(e) and (f). This step effectively eliminates redundant points and derives a more regular structure, which is essential for further analysis. The N cluster centers are defined as the scattering key points, denoted as $P = \{p_1, p_2, \dots, p_N\}$. These points represent the local distributions of scattering intensity and structural characteristics of the targets. By capturing these local features, the method is able to enrich the discriminative features of ships, leading to improved classification and recognition performance. Following the clustering process, the scattering key point coordinate $V = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ of the scattering key points on the original images are obtained for establishing SPT.

In effect, there is geometric structure information for the distribution of key scattering points of a target. Key scattering points of ships of the same category have comparable topology structure, which can also be confirmed by the similar distribution of key scattering points of the two containers in Fig. 3(e) and (f). Therefore, we construct a topology structure based on the scattering key points, so as to fully utilize the distribution information of the scattering key points to characterize the structural characteristics of the ship. First, for a set of scattering key points P , calculate the Euclidean distance $d(x_i, x_j)$ between each pair of scattering key points. This can be represented by the following equation:

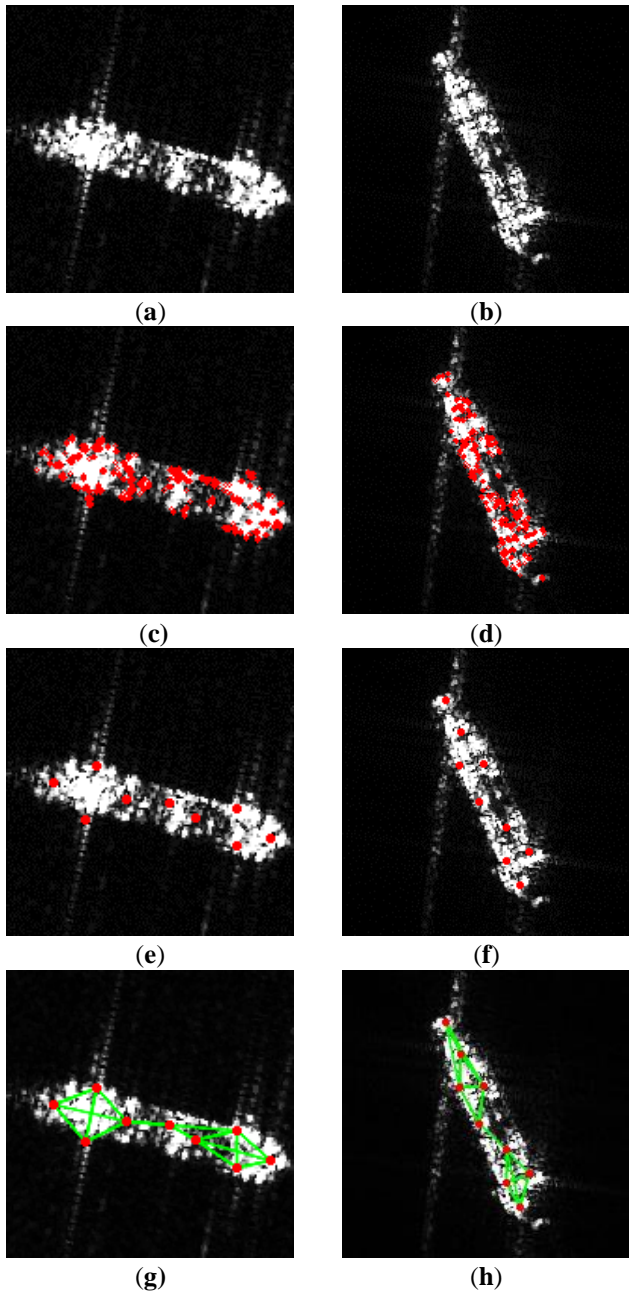


Fig. 3. Illustration of the SPT. (a) and (b) Origin SAR images of containers. (c) and (d) Extracted points by Harris corner detector. (e) and (f) Key scattering points. (g) and (h) SPT.

block. Each residual block consists of three convolutional

$$d(p_i, p_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (1)$$

For each scattering key point p_i , according to its Euclidean distance with other scattering key points, determine its K nearest neighbor data points. Determined using the following formula:

$$N_i = \operatorname{argmin}_{j \neq i} d(p_i, p_j) : j = 1, 2, \dots, N, N \neq i \quad (2)$$

where N_i represents the index of K scattering key points closest to scattering key point p_i .

And then, using the information of K nearest neighbors, construct an adjacency matrix $A \in \mathbb{R}^{N \times N}$, where $A(i, j)$ indicates whether there is a connection between scattering key points p_i and p_j . In general, $A(i, j)$ is 1 if p_j is one of the K nearest neighbors of p_i , and 0 otherwise, which can be expressed as:

$$A(i, j) = \begin{cases} 1, & \text{if } j \in N_i \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Finally, using Dijkstra's algorithm, starting from each scattering key point p_i , calculate the shortest path distance to other scattering key points to obtain the distance matrix $E \in \mathbb{R}^{N \times N}$, where $e_{ij} \in E$ represents the geodesic distance from p_i to p_j . This forms an undirected graph $T = \{P, E\}$, also named as the SPT. Visualization of the SPT is shown in Fig. 3(g) and (h). This topology structure provides valuable insights into the distribution patterns of the scattering key points, as well as their

N 32-dimensional feature vectors, denoted as $F' \in \mathbb{R}^{N \times 32}$, which is expressed as

$$f'_i = \Theta \sum_{j \in \mathcal{N}(i) \cup \{i\}} \frac{e_{j,i}}{\sqrt{d_j d_i}} f_j \quad (4)$$

with $\hat{d}_i = 1 + \sum_{j \in \mathcal{N}(i)} e_{j,i}$, where f'_i represents the i -th row of F' , f_j represents the j -th row of F , Θ represents trainable weights, $\mathcal{N}(i)$ represents all neighbor nodes of i and $e_{j,i}$ denotes the edge weight from source node j to target node i .

However, directly concatenating these vectors would introduce a dependency on the order of input feature points, resulting in a SPT embedding that could exhibit variations for different targets of the same category. In this scenario, scattering features representing the same feature area might appear in different positions, potentially compromising the effective utilization of ship scattering features. This limitation arises due to the constrained feature extraction capabilities of the subsequent fully connected layer.

To confront and mitigate this challenge while bolstering the robustness of the embedding, the final layer incorporates a dedicated MLP-Mixer layer. This MLP-Mixer layer is meticulously engineered to further process the feature vectors while maintaining the same number of output channels as the input channels. The Mixer layer comprises two types of MLP sub-layers: channel-mixing MLPs and token-mixing MLPs. Within the Mixer layer, there are two distinct categories of MLP sub-layers, namely channel-mixing MLPs and token-mixing MLPs. These two sub-layer types are designed to serve specific purposes in enhancing feature fusion and communication. The

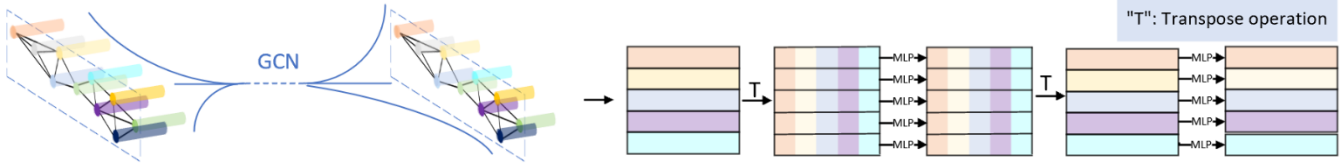


Fig. 4. Framework of the TEB.

positional relationships. By analyzing these relationships, it is possible to characterize the structural and shape features of ship targets more accurately.

D. Topology Encoding Branch (TEB)

After constructing the SPT, it is essential to access the feature map from the backbone's Stage 1 and identify the 64-dimensional feature vectors corresponding to each scattering key point based on their coordinates V . These vectors serve as the attributes of the nodes in the topology, with the corresponding vector matrix denoted as $F \in \mathbb{R}^{N \times 64}$. With the complete SPT in place, denoted as $T = \{P, F, E\}$, we can proceed to input it into the TEB, as shown in Fig. 4.

The branch's first layer comprises a graph convolutional layer with the same number of input and output channels. This layer is followed by another graph convolutional layer, which compresses the output channels down to 32 dimensions. This process effectively extracts and condenses the features, yielding

channel-mixing MLPs are responsible for facilitating communication among different channels. They operate independently on each token, treating individual rows of the data table as their inputs. In contrast, the token-mixing MLPs are geared towards enabling communication across different spatial locations or tokens. They work independently on each channel and utilize individual columns of the data table as their inputs. These two types of sub-layers are intentionally interleaved within the Mixer layer. This arrangement ensures that both channel-level and spatial-level interactions are effectively harnessed, facilitating a comprehensive and holistic fusion of information across both input dimensions.

Distinct from a conventional MLP, which primarily extracts features from individual vectors, the MLP-Mixer introduces an innovative capability: it facilitates communication between different spatial locations, specifically the features of scattering key points located at varying positions. This communication mechanism empowers the comprehensive fusion of feature

attributes from scattering key points situated in diverse areas. This fusion can be mathematically expressed as:

$$\begin{aligned} \mathbf{U}_{*,i} &= \mathbf{F}'_{*,i} + \mathbf{W}_2 \sigma(\mathbf{W}_1 \text{LayerNorm}(\mathbf{F}'_{*,i})), \quad \text{for } i = 1 \dots 32, \\ \mathbf{Y}_{j,*} &= \mathbf{U}_{j,*} + \mathbf{W}_4 \sigma(\mathbf{W}_3 \text{LayerNorm}(\mathbf{U}_{j,*})), \quad \text{for } j = 1 \dots N. \end{aligned} \quad (5)$$

Here σ is an element-wise nonlinearity (GELU [61]), \mathbf{F} is a patch composed of f_2 . Ultimately, by concatenating these N 32-dimensional feature vectors, we obtain a SPT embedding that is largely independent of the scattering point placements.

E. Reciprocal Feature Fusion Attention (RFFA)

After completing the feature extraction process within the TEB, we transition from the SPT to the SPT embedding. Nevertheless, a pivotal challenge arises when it comes to effectively amalgamating the SPT embedding with the visual information, which has been extracted by the CNN. This challenge is critical to address as it determines the model's ability to synergize both visual and topological aspects. On one hand, SPT embedding encapsulates the topological characteristics of the target, providing essential insights into its structure and organization. On the other hand, the visual information obtained through the CNN offers rich details and patterns that are crucial for accurate classification and recognition tasks. Bridging the gap between these two types of data and ensuring that they complement each other seamlessly is of paramount importance to improve the overall performance of the model.

To tackle this challenge, we introduce the innovative concept of RFFA. RFFA serves as a mechanism to facilitate the fusion of visual information and SPT embedding, thereby creating a unified representation. This integration plays a pivotal role in enabling the subsequent fully connected neural network to acquire a more comprehensive understanding of the target by simultaneously learning from both its visual and topological features. By incorporating two interactive cross-attention[62] functions within the RFFA module, we facilitate cross-modality learning, as shown in Fig. 5. These cross-attention functions are meticulously designed to discern subtle alignments and correspondences between the topological and visual information. They effectively identify and enhance the features that are corroborated and confirmed by both modalities, thus enriching the final fused representation.

In practice, the RFFA process starts with layer normalization to ensure consistent scaling and to enhance training stability. Following this, the input features from both modalities are processed through a shared fully connected layer, projecting them onto a shared inner dimension. We denote the modality of the topology and visual representations with subscripts T and V, resulting in \mathbf{H}_T^l and \mathbf{H}_V^l , respectively. The overarching concept is to enhance the importance of visual features within a target when their relevance is substantiated by the semantic context provided by the topology features. To formalize this, we introduce a cross-attention function, expressed mathematically as follows:

$$f_{\text{CrossAtt}}(\mathbf{H}_T^l) = \text{softmax}\left(\frac{q(\mathbf{H}_T^l)k(\mathbf{H}_V^l)}{\sqrt{d_k}}\right)v(\mathbf{H}_T^l) \quad (6)$$

$$f_{\text{CrossAtt}}(\mathbf{H}_V^l) = \text{softmax}\left(\frac{q(\mathbf{H}_V^l)k(\mathbf{H}_T^l)}{\sqrt{d_k}}\right)v(\mathbf{H}_V^l) \quad (7)$$

In this formulation, q , k , and v represent the query, key, and value functions, respectively, while d_k is the dimension of the key vectors. After applying the cross-attention functions, we utilize fully connected layers to separately project the topology features and visual features back to their original dimensions, thereby obtaining the new topology features and visual features. To generate the final output, we concatenate these updated features, effectively combining the relevant information from both modalities to produce a more discriminative and robust representation.

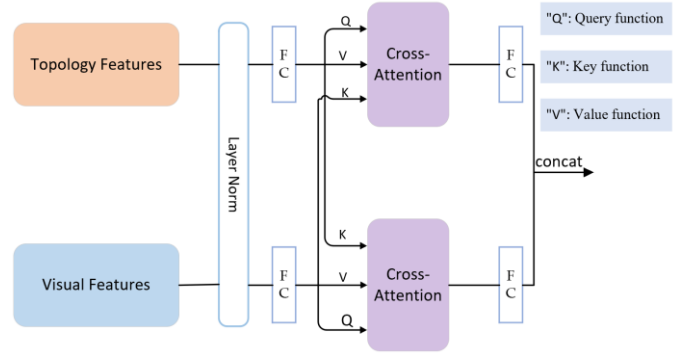


Fig. 5. Illustration of the RFFA.

IV. EXPERIMENTS

A. Datasets

1) *OpenSARShip Dataset*: The OpenSARShip dataset is a comprehensive collection of Sentinel-1 SAR images specifically designed for ship interpretation tasks. The dataset is compiled from Sentinel-1 images acquired across various sea areas, environmental conditions, and seasons, making it a valuable resource for research in maritime surveillance and related applications. This dataset comprises over 60,000 SAR image chips, with each chip measuring 256×256 pixels. The chips contain approximately 20,000 ships, with detailed annotations, including ship position, length, and width, as well as relevant environmental information, such as wind speed and wave height. The dataset also includes a wide range of ship types, such as bulk carriers, container ships, fishing vessels, oil tankers, passenger ships, and tugs, among others. Excluding a few categories with particularly small sample sizes, we utilized 8 different classes from the dataset. We randomly selected 4 classes for the training set and the remaining 4 classes for the test set, which is shown in Table I.

TABLE I
THE DETAILED DATASET SETTINGS IN OUR EXPERIMENT.

Dataset	Base Classes	Novel Classes
OpenSARShip	Cargo	Dredgin
	Fishing	Passeng
	Tanker	Pilot
	Tug	Search
FUSAR-Ship	Cargo	HighSpeedCraft
	Dredger	LawEnforce
	Fishing	Reserved
	Tanker	Passenger
	Unspecified	Tug
Open-FUSAR	BulkCarrier, CargoShip	Unspecified
	CrudeOilTanker, Fishing	ContainerShip
	Passenger, Pleasurecraft	GeneralCargo
	Tanker, Tanker-HazardB	Tanker-HazardA
	Tug, WingInGrnd	PilotVessel
	HighSpeedCraft, OffshoreSupplyShip	Oil_ChemicalTanker
	OilProductsTanker	

2) *FUSAR-Ship Dataset*: The FUSAR-Ship dataset is a high-resolution SAR-AIS matchup dataset built using Gaofen-3 SAR images and associated automatic identification system (AIS) data. The dataset is specifically designed to facilitate research in ship detection, recognition, and maritime surveillance by providing high-resolution SAR images with accurate ship position and identification information from corresponding AIS data. FUSAR-Ship contains over 10,000 SAR image chips, with each chip measuring 500×500 pixels. These images are acquired from Gaofen-3, a high-resolution SAR satellite with a spatial resolution of up to 1 meter. The dataset covers various sea areas, environmental conditions, and ship types, enhancing its versatility for algorithm development and evaluation. The AIS data included in the dataset provide ship identification, position, course, speed, dimensions, and other relevant attributes. Following the general settings in FSL, we utilized 10 different classes from the dataset. We randomly selected 5 classes for the training set and the remaining 5 classes for the test set, which is shown in Table I.

3) *Open-FUSAR Dataset*: To address the challenge of fine-grained classification in SAR images, where the subtle differences between various classes make it difficult for few-shot learning models to effectively learn discriminative features, we have integrated the aforementioned two datasets into a unified dataset called Open-FUSAR Dataset. During the integration process, we combined instances from the same categories across both datasets to increase the diversity and the number of samples for each class. Meanwhile, distinct categories were kept separate to maintain a broad range of ship types for classification tasks. After the integration, we randomly selected 13 classes to be included in the training set, ensuring a sufficient number of examples for each class to facilitate the effective training of the model. The remaining 6 classes were designated as the test set, which enabled us to

evaluate the performance of the trained model on unseen data and assess its generalization capabilities, which is shown in Table I. By combining and processing the two datasets in this manner, we aimed to enhance the model's learning ability by exposing it to a diverse set of examples across various ship types, environments, and conditions. Furthermore, this approach allows researchers to develop more robust and accurate fine-grained classification algorithms for SAR images, ultimately improving maritime surveillance and related applications.

B. Implementation Details

In our experiments, we employ a testing setup using 4way-1shot and 4way-5shot configurations on OpenSARShip Dataset, using 5way-1shot and 5way-5shot configurations on FUSAR-Ship Dataset and using 6way-1shot and 6way-5shot configurations on Open-FUSAR Dataset and, with 5 query images per class. The input images are resized to a fixed resolution of 84×84 pixels. During the training phase, we set the number of epochs to 50. For both the training and fine-tuning processes, we utilize the stochastic gradient descent (SGD) algorithm as the optimizer. The optimizer is configured with a learning rate of 0.01, momentum of 0.9, dampening of 0.9, and weight decay of 0.001. The fine-tuning process is executed with a total of 150 steps for 1-shot and 600 steps for 5-shot. In the testing phase, we set the number of test episodes to 100 and employ a batch size of 4. In the context of the SPT, each scattering point's feature dimension extracted by the TEB is set at 32. Consequently, the dimensionality of the TEB output equals $N \times 32$. In parallel, the output of the backbone network is flattened, resulting in 640-dimensional vectors. To maintain consistency, we have set the default number of scattering key points (N) to 20, ensuring that the dimensionality of the TEB

output matches the output dimension of the backbone network. As a result, the SPT embedding shares the same dimensionality as the backbone output. Upon concatenation of the SPT embedding and the backbone output, the combined feature vector attains a dimensionality of 1280. This feature vector is subsequently input into the fully connected layer for further processing. Additionally, we have set the default number of nearest neighbor nodes (K) for the SPT at 3.

C. Evaluation Metric

For each episode, the accuracy (Acc_i) is computed as the ratio of the number of correctly classified query images to the total number of query images. The computation process is presented in Equation (8), where N_q denotes the total number of query images. The variables p_i and t_i represent the predicted class and ground truth class of the i -th query image, respectively. The indicator function \mathbb{I} , is equal to 1 if $p_i = t_i$ and 0 otherwise.

$$Acc_i = \frac{1}{N_q} \sum_{j=0}^{N_q} \mathbb{I}(p_i = t_i) \quad (8)$$

To mitigate the impact of randomness, the experiment is conducted multiple times, and the final accuracy is determined as the average of all episode accuracies (Acc_i), as shown in Equation (9). The variable N in the equation refers to the number of repetitions of the experiment. In the main experiment, N is set to 100 for all models, providing a more robust evaluation.

$$Acc = \frac{1}{N} \sum_{j=0}^N Acc_i \quad (9)$$

By utilizing this evaluation metric, we can achieve a more reliable and precise estimation of the models' performance.

D. Ablation Studies

we conduct a series of ablation experiments on three datasets to analyze the effectiveness of each proposed component in SPT-FSC. All subsequent experiments maintain the same settings to make a fair comparison. Specifically, we explore the effects of different numbers of scattering key points, varying numbers of nearest neighbors, the utilization of TEB including GCN and MLP-Mixer, and the impact of RFFA. By

systematically examining these aspects, we aim to gain deeper insights into the underlying mechanisms of SPT and provide valuable guidance for optimizing its performance.

1) *Effects of different number of scattering key points:* In the analysis of the effects of different numbers of scattering key points on classification accuracy, it can be observed that an increase in the number of scattering key points generally leads to an improvement in classification accuracy for both the 1-shot and 5-shot scenarios, as shown in table II. However, the improvement tends to plateau or even slightly decline when the number of key points surpasses a certain threshold, such as $N=20$ for the OpenSARShip and FUSAR-Ship datasets. And the performance improvement is more pronounced in the 5-shot scenario compared to the 1-shot scenario. Furthermore, the impact of increasing the number of scattering key points on classification accuracy appears to be more significant for the Open-FUSAR dataset compared to the OpenSARShip and FUSAR-Ship datasets. This demonstrates the effectiveness of expanding the base class size in the dataset to address the lack of diversity in the training data for few-shot classification in SAR images.

Fig. 6 offers a visualization of the SPT with different numbers of scattering key points, showing that the topology becomes more intricate as the number of key points increases. This increased complexity allows the method to capture more discriminative features, resulting in better classification performance. Nonetheless, it is crucial to strike a balance between the complexity of the SPT and the risk of overfitting, as utilizing too many scattering key points may lead to a decline in generalization capabilities.

In conclusion, the increase in the number of scattering key points enhances the classification accuracy of the SPT method, particularly in the 5-shot scenario. However, it is essential to find an optimal balance between the complexity of the SPT and the model's generalization capabilities, as excessive complexity may lead to overfitting and reduced performance on unseen data.

Table II
EFFECTS OF DIFFERENT NUMBER OF SCATTERING KEY POINTS

N	OpenSARShip		FUSAR-Ship		Open-FUSAR	
	1shot	5shot	1shot	5shot	1shot	5shot
10	44.35 \pm 2.37	59.90 \pm 2.05	52.12 \pm 1.87	62.08 \pm 1.89	51.44 \pm 2.24	73.56 \pm 1.72
15	47.25 \pm 2.19	66.40 \pm 2.27	51.12 \pm 2.27	66.72 \pm 1.55	56.00 \pm 2.30	77.56 \pm 1.59
20	47.75 \pm 2.47	68.55 \pm 2.36	55.76 \pm 2.28	68.04 \pm 1.64	59.48 \pm 2.61	81.48 \pm 1.59
25	51.95 \pm 2.92	66.80 \pm 1.94	50.08 \pm 2.46	66.80 \pm 1.80	60.60 \pm 2.60	79.76 \pm 1.58
30	51.15 \pm 2.49	72.00 \pm 2.03	52.68 \pm 2.51	69.04 \pm 1.67	62.68 \pm 2.51	80.40 \pm 1.53

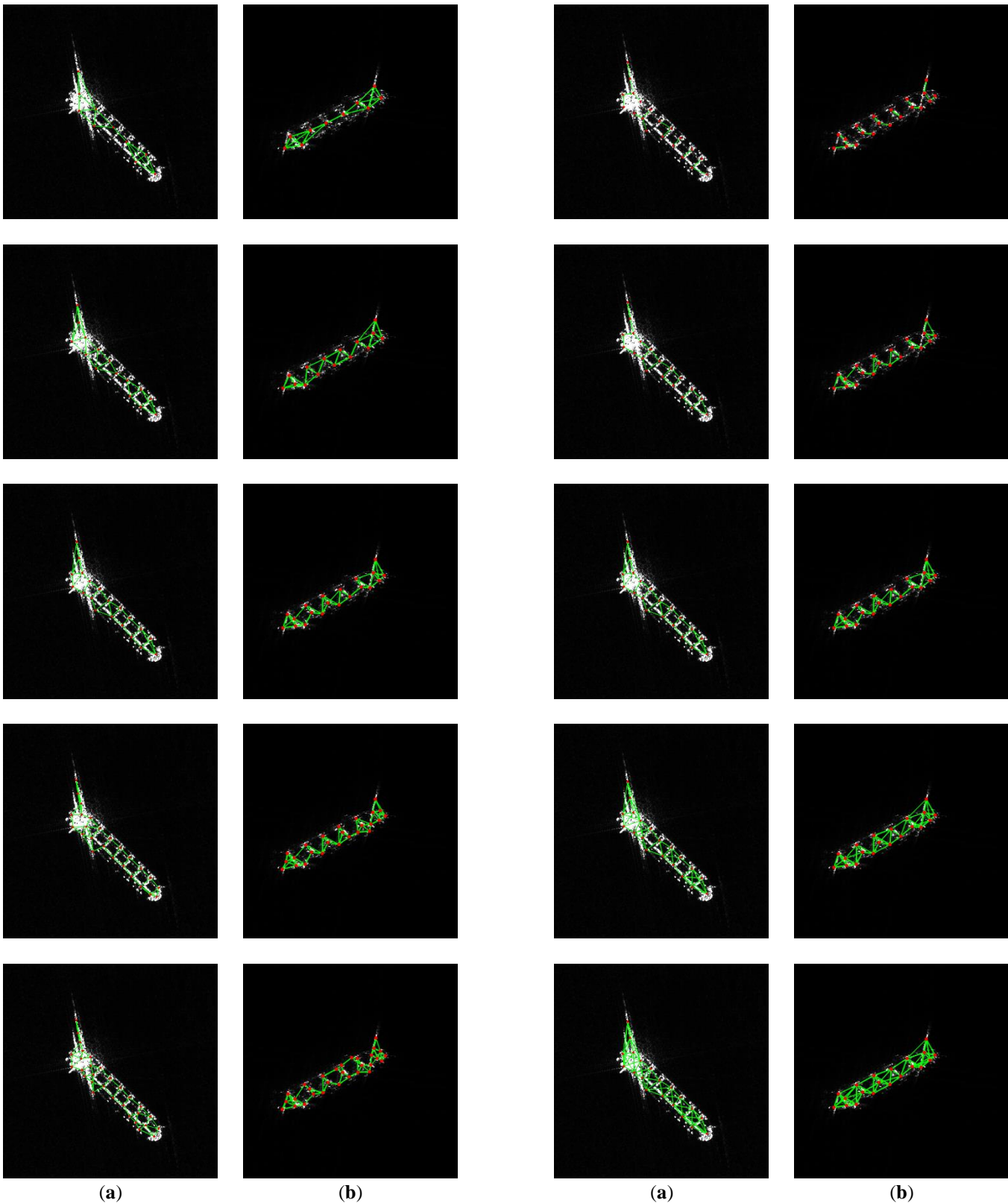


Fig. 6. Visualization of the SPT with different number of scattering key points. (a) Container of 10, 15, 20, 25, and 30 scattering key points, respectively. (b) Tanker of 10, 15, 20, 25, and 30 scattering key points, respectively.

Fig. 7. Visualization of the SPT of with different number of nearest neighbors. (a) Container of 1-5 nearest neighbors, respectively. (b) Tanker of 1-5 nearest neighbors, respectively.

2) *Effects of different number of nearest neighbors:* In this section, we discuss the effects of using different numbers of nearest neighbors K on the performance of the SPT method.

Table III presents the experimental results for different values of K , and Fig. 7 provides a visualization of the SPT with varying numbers of nearest neighbors. From Table III, we can observe that the performance of the SPT method varies with the number of nearest neighbors, which indicates that the selection of K plays a significant role in the performance of the method. These observations suggest that an optimal number of nearest neighbors exists for each dataset. When the number of nearest neighbors is too low or too high, the performance of the model deteriorates. The optimal number of nearest neighbors seems to be dataset-specific, as the best value varies between the three datasets. The visualization of the SPT with different numbers of nearest neighbors in Fig. 7 further supports these observations. The underlying reason for these patterns may be related to the trade-off between local and global information. A smaller number of nearest neighbors emphasizes local features, while a larger number of nearest neighbors captures more global information. It is crucial to find the right balance between local and global information to achieve the best performance in recognizing and classifying ships in SAR images.

The visualizations demonstrate how the SPT changes as the number of nearest neighbors increases. We can see that with a smaller K value, the topology appears sparser, while with a larger K value, the topology becomes denser. This observation suggests that the selection of K influences the complexity of the SPT, which in turn affects the representation of the structural and shape features of the ship targets.

In conclusion, the results indicate that selecting an appropriate number of nearest neighbors is crucial for the SPT method to effectively capture the structural and shape features of ship targets. By choosing an optimal K value, the method can achieve better classification and recognition performance. The potential pattern underlying these results may be that the optimal K value is dataset-dependent, meaning that it could be influenced by the specific characteristics of the dataset being used.

3) *Effects of TEB*: In this section, we delve into the analysis of the effects of employing GCN and the MLP-Mixer in the TEB of our proposed model. Our primary focus is on understanding how these components enhance the

performance of the model, especially in scenarios where the SPT plays a crucial role.

To provide some context, we began by introducing RFFA to the baseline model and conducted experiments. RFFA, in this context, operates by taking image features extracted by CNN as its inputs. It effectively acts as a self-attention mechanism, reinforcing the output features of the network. Our results, as shown in the second row of Table IV, indicated a noticeable improvement in model performance after adding RFFA. This improvement can be attributed to the ability of RFFA to enhance the discriminability of the model's output features.

Building upon this, we further enhanced the model by introducing TEB based on the baseline with RFFA, where GCN was employed. As reflected in the third row of Table IV, this addition resulted in a significant boost in performance compared to when both inputs of RFFA were image features. The test accuracy increased by 4.95% (1-shot) and 10.6% (5-shot) on OpenSARship, 6.28% (1-shot) and 12.88% (5-shot) on FUSAR-ship, and 1.56% (1-shot) and 7.48% (5-shot) on Open-FUSAR, respectively. This demonstrated the effectiveness of introducing topological features into the model, highlighting the importance of SPT in the task.

Visualizations using t-SNE on the Open-FUSAR dataset, as shown in Fig. 8, further corroborate our experimental findings. Baseline feature embeddings exhibit aliasing issues (Fig. 8(a) and (b)), which are alleviated to some extent with the addition of RFFA (Fig. 8(c) and (d)). However, it's only when we incorporate both RFFA and TEB (including GCN and MLP-Mixer) that clear boundaries between different categories emerge in the feature embeddings (Fig. 8(e) and (f)). This suggests that MLP-Mixer significantly enhances the discriminative capabilities of the feature embeddings, making it easier to distinguish between different classes. This also proves that the role of MLP-Mixer in addressing the issue of input feature point order in the SPT embedding is noteworthy. By facilitating communication between different spatial locations, MLP-Mixer enables the comprehensive fusion of feature attributes from scattering key points located in different areas. This ultimately results in a more robust SPT embedding that is less dependent on the specific positions of scattering points. Consequently, the

Table III
EFFECTS OF DIFFERENT NUMBER OF NEAREST NEIGHBORS.

K	OpenSARShip		FUSAR-Ship		Open-FUSAR	
	1shot	5shot	1shot	5shot	1shot	5shot
1	48.80 \pm 2.70	69.70 \pm 2.46	50.64 \pm 2.37	70.68 \pm 1.62	57.92 \pm 2.29	79.72 \pm 1.67
2	47.80 \pm 2.22	70.95 \pm 2.30	51.72 \pm 2.44	69.44 \pm 1.46	57.28 \pm 2.28	77.08 \pm 1.57
3	47.75 \pm 2.47	68.55 \pm 2.36	55.76 \pm 2.28	68.04 \pm 1.64	59.48 \pm 2.61	81.48 \pm 1.59
4	47.20 \pm 2.12	63.10 \pm 2.43	54.08 \pm 2.41	67.80 \pm 1.54	59.84 \pm 2.79	75.28 \pm 1.69
5	48.05 \pm 2.17	62.60 \pm 2.23	49.76 \pm 2.46	70.96 \pm 1.42	59.28 \pm 2.34	78.48 \pm 1.67

Table IV
ABLATION FOR KEY COMPONENTS PROPOSED IN SPT-FSC.

RFFA	TEB		OpenSARShip		FUSAR-Ship		Open-FUSAR	
	GCN	MLP-Mixer	1shot	5shot	1shot	5shot	1shot	5shot
×	×	×	35.85 \pm 2.00	43.40 \pm 2.04	31.68 \pm 1.56	37.32 \pm 2.00	48.78 \pm 2.24	60.52 \pm 1.84
√	×	×	39.55 \pm 1.95	53.50 \pm 2.31	39.40 \pm 1.87	50.08 \pm 1.86	50.56 \pm 2.31	68.00 \pm 1.76
√	√	×	44.50 \pm 2.38	64.10 \pm 2.53	45.68 \pm 2.27	62.96 \pm 1.82	52.12 \pm 2.14	74.12 \pm 1.93
×	√	√	35.90 \pm 2.18	37.70 \pm 1.89	30.32 \pm 1.65	35.28 \pm 1.81	49.12 \pm 2.22	62.40 \pm 1.67
√	√	√	47.75 \pm 2.47	68.55 \pm 2.36	55.76 \pm 2.28	68.04 \pm 1.64	59.48 \pm 2.61	81.48 \pm 1.59

inclusion of MLP-Mixer in the TEB contributes significantly to the overall enhancement of the model's performance across all datasets and scenarios.

Visualizations using t-SNE on the Open-FUSAR dataset, as shown in Fig. 8, further corroborate our experimental findings. Baseline feature embeddings exhibit aliasing issues (Fig. 8(a) and (b)), which are alleviated to some extent with the addition of RFFA (Fig. 8(c) and (d)). However, it's only when we incorporate both RFFA and TEB (including GCN and MLP-Mixer) that clear boundaries between different categories emerge in the feature embeddings (Fig. 8(e) and (f)). This suggests that MLP-Mixer significantly enhances the discriminative capabilities of the feature embeddings, making it easier to distinguish between different classes. This also proves that the role of MLP-Mixer in addressing the issue of input feature point order in the SPT embedding is noteworthy. By facilitating communication between different spatial locations, MLP-Mixer enables the comprehensive fusion of feature attributes from scattering key points located in different areas. This ultimately results in a more robust SPT embedding that is less dependent on the specific positions of scattering points. Consequently, the inclusion of MLP-Mixer in the TEB contributes significantly to the overall enhancement of the model's performance across all datasets and scenarios.

4) *Effects of RFFA*: The main goal of RFFA is to enhance the discriminative power of the fused features by considering the reciprocal relationships and contributions of the features from different sources, allowing the model to focus on the most relevant and informative parts of each feature set.

The incorporation of RFFA in our model is aimed at elevating the discriminative power of fused features by considering the reciprocal relationships and contributions of features from diverse sources. This strategic approach enables the model to prioritize the most relevant and informative aspects of each feature set. The impact of RFFA on the network's output features is clearly evident from the second row of the table, where RFFA is applied to the Baseline model. In this scenario, RFFA operates on two inputs, which are the image features initially extracted by the CNN. The test accuracy experiences substantial improvements, with gains of 3.7% (1-shot) and 10.1% (5-shot) on OpenSARShip, 7.72% (1-shot) and 12.76% (5-shot) on FUSAR-ship, and 1.78% (1-shot) and 7.48% (5-shot) on Open-FUSAR, respectively. These results

underscore the significant enhancement effect that RFFA has on the model's output features.

Fig. 8 (g) and (h) present t-SNE visualizations of the output features fused by simple concatenation on the Open-FUSAR dataset. In this context, the feature embedding of Baseline with TEB, including GCN and MLP-Mixer, exhibits substantial aliasing, similar to the Baseline. This observation underscores that while SPT embedding contains a wealth of feature information, it primarily pertains to topological features, which are inherently different from the visual features extracted by the CNN. Directly concatenating these visual and topological features may not efficiently leverage the topological information and could potentially hinder the extraction of visual features. In contrast, the t-SNE visualizations for both 1-shot and 5-shot scenarios (Fig. 8 (i) and (j)) demonstrate that features fused by RFFA are more discriminative and better separated compared to those fused by simple concatenation. Clear boundaries between different categories emerge, resulting in significantly improved classification performance. Consequently, the integration of RFFA for feature fusion plays a pivotal role in effectively amalgamating these distinct modalities and substantially enhancing the model's overall performance.

E. Comparison with State-of-the-Art Methods

To validate the effectiveness of SPT-FSC, we compared our method with eight state-of-the-art few-shot classification methods in the three datasets, as shown in Table V. The backbone for all the compared methods utilizes ResNet-12, and the training and testing parameters align with those employed in this paper. To ensure a fair comparison, all hyperparameters for our method and other methods are set to their default values. In comparison with state-of-the-art methods, the number of scattering key points (N) is fixedly set to 20, and the number of nearest neighbors (K) is fixedly set to 3. The parameter temperature of the Cosine classifier in Baseline++ is set to 10. The parameter margin in NegMargin is set to -0.01. The relation module of RelationNet consists of two convolutional blocks and two fully connected layers. Each of convolutional block is a 3 × 3 convolution followed by batch normalisation, ReLU non-linearity and 2 × 2 max-pooling. MAML use Adam as the meta-optimizer with a fixed step size $\alpha = 0.01$. These settings are essential for consistency in our evaluations.

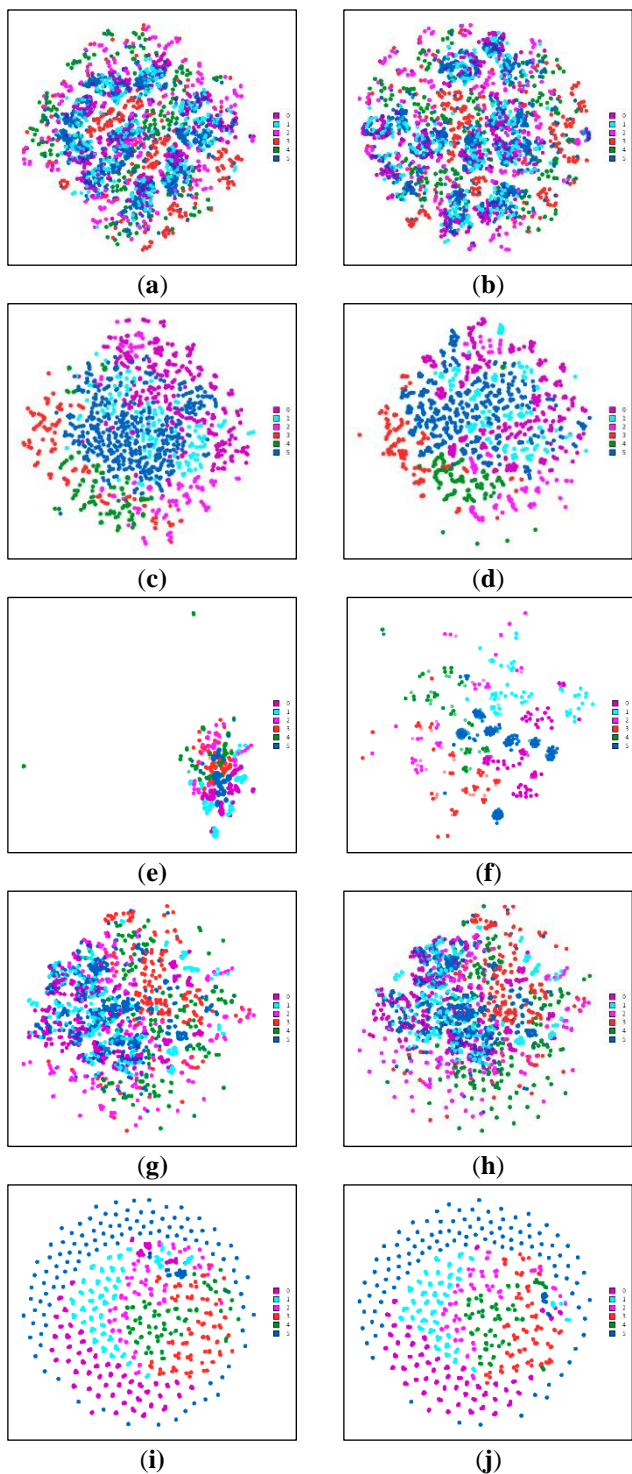


Fig. 8. t-SNE visualization of the output features on Open-FUSAR Dataset. (a) Baseline for 1-shot. (b) Baseline for 5-shot. (c) Baseline with RFFA for 1-shot. (d) Baseline with RFFA for 5-shot. (e) Baseline with RFFA and TEB including GCN for 1-shot. (f) Baseline with RFFA and TEB including GCN for 5-shot. (g) Baseline with TEB including GCN and MLP-Mixer for 1-shot. (h) Baseline with TEB including GCN and MLP-Mixer for 5-shot. (i) SPT-FSC for 1-shot. (j) SPT-FSC for 5-shot. The numbers 0-5 in the legend represent ContainerShip, GeneralCargo, oil_ChemicalTanker, PilotVessel, Tanker-HazardA and Unspecified, respectively.

Based on the experiment results presented in the table, we can provide an in-depth analysis of the performance of various few-shot classification methods, including meta-learning and transfer learning strategies, on the OpenSARShip, FUSAR-Ship, and Open-FUSAR datasets. For the OpenSARShip dataset, the proposed SPT-FSC method significantly outperforms all other methods in both the 1-shot and 5-shot scenarios, achieving 47.75% (± 2.47) and 68.55% (± 2.36) accuracy, respectively. This demonstrates the effectiveness of the SPT, TEB and RFFA in capturing detailed information from SAR images. Similarly, on the FUSAR-Ship dataset, SPT-FSC also achieves the best performance in both the 1-shot and 5-shot settings, with accuracies of 55.76% (± 2.28) and 68.04% (± 1.64), respectively. The results again showcase the strength of our proposed method in handling high-resolution SAR images with accurate ship position and identification information. In the case of the Open-FUSAR dataset, the SPT-FSC method maintains its superior performance, achieving the highest accuracy in both the 1-shot and 5-shot scenarios, with 59.48% (± 2.61) and 81.48% (± 1.59), respectively. This highlights the benefits of combining the OpenSARShip and FUSAR-Ship datasets to create a more diverse and fine-grained ship classification dataset. In general, our proposed SPT-FSC method demonstrates superior performance across all datasets and few-shot settings. This can be attributed to the effective integration of SPT information and RFFA mechanism, which enhances the network's adaptability to the imaging mechanism and alleviates the imaging variability inherent in SAR images.

The t-SNE visualization of the output features in the Open-FUSAR dataset provides valuable insights into the performance of various methods, as illustrated in Fig. 9. When the image classification model has effectively learned meaningful features from the data, the t-SNE visualization tends to reveal distinct separations between different classes within the embedding space. Notably, in Fig. 9(a), the feature embedding of ProtoNet for 1-shot learning and in Fig. 9(d), Baseline++ for 5-shot learning display highly concentrated distributions with the presence of outliers. Conversely, in Fig. 9(b), the feature embedding of ProtoNet for 5-shot learning and in Fig. 9(c), Baseline++ for 1-shot learning exhibit more dispersed distributions, but with a notable degree of overlap among classes. Moving on to Fig. 9(e), the feature embedding of SPT-FSC for 1-shot learning and Fig. 9(f), SPT-FSC for 5-shot learning demonstrates even more scattered distributions, with clear boundaries between different class embeddings, significantly mitigating the issue of overlap. It's worth highlighting that Fig. 9(f), representing SPT-FSC for 5-shot learning, exhibits fewer instances of overlap compared to Fig. 9(e), representing SPT-FSC for 1-shot learning, indicating that the number of support set samples has an impact on model performance, which aligns with the accuracy results presented. In direct comparison among ProtoNet, Baseline++, and SPT-FSC, it becomes evident that the proposed SPT-FSC method consistently leads to superior separation of different classes, whether in the 1-shot or 5-shot learning settings. These results

TABLE V
COMPARISON WITH STATE-OF-THE-ART METHODS.

Method	OpenSARShip		FUSAR-Ship		Open-FUSAR	
	1shot	5shot	1shot	5shot	1shot	5shot
MatchingNet [35]	30.20 \pm 1.66	33.30 \pm 1.70	31.88 \pm 1.96	34.00 \pm 1.44	46.41 \pm 2.26	56.86 \pm 1.98
ProtoNet [14]	32.00 \pm 1.97	37.35 \pm 2.19	30.64 \pm 2.01	35.40 \pm 1.68	40.44 \pm 2.04	52.95 \pm 1.96
RelationNet [36]	29.64 \pm 2.04	30.60 \pm 1.73	31.12 \pm 1.53	31.12 \pm 1.53	43.63 \pm 2.27	47.01 \pm 1.84
MAML [18]	29.60 \pm 2.13	33.10 \pm 1.87	31.32 \pm 1.81	32.40 \pm 1.71	42.68 \pm 2.39	52.28 \pm 1.86
Baseline [41]	35.85 \pm 2.00	43.40 \pm 2.04	31.68 \pm 1.56	37.32 \pm 2.00	48.78 \pm 2.24	60.52 \pm 1.84
Baseline++ [41]	35.35 \pm 2.09	40.90 \pm 1.84	31.96 \pm 1.79	35.64 \pm 1.83	46.36 \pm 2.35	60.32 \pm 1.80
NegMargin [63]	36.15 \pm 2.15	41.05 \pm 1.86	30.68 \pm 1.91	35.68 \pm 1.89	35.68 \pm 1.89	59.96 \pm 1.82
MetaBaseline [42]	32.51 \pm 2.03	38.60 \pm 1.86	30.32 \pm 1.88	34.52 \pm 1.83	47.59 \pm 2.21	58.68 \pm 1.87
SPT-FSC(ours)	47.50 \pm 2.47	71.05 \pm 2.48	55.76 \pm 2.28	68.04 \pm 1.64	59.48 \pm 2.61	81.48 \pm 1.59

underscore the critical role of feature in enhancing classification accuracy.

To provide a more comprehensive illustration of the experimental results, Fig. 10 presents the confusion matrices for ProtoNet, Baseline++, and SPT-FSC applied to the OpenSARship Dataset. Analyzing Fig. 10(a)-(d), it becomes evident that the classification accuracy is notably higher for the "Pilot" and "Passenger" categories, while it is relatively lower for the "Dredging" and "Search" categories. Notably, the overall accuracy for the "Search" category is particularly low. This disparity in accuracy may be attributed to the inherent challenges posed by the small size of ships within the "Search" category, making them difficult to distinguish. However, in Fig. 12(e) and (f), we observe significant improvements in the recognition accuracy across all categories. For the four categories of "Dredging," "Pilot," "Passenger," and "Search," the accuracy has increased substantially, with gains exceeding 14.8%, 9.4%, 15.4%, and 9.0% for the 1-shot scenario, respectively, and even more remarkable gains of 32.4%, 28.6%, 26.2%, and 33.4% for the 5-shot scenario. Notably, SPT-FSC outperforms in the 5-shot scenario across all four categories, achieving classification accuracies exceeding 70% in each. Furthermore, the classification accuracy for "Dredging" and "Search" categories surpasses that of the "Pilot" and "Passenger" categories in SPT-FSC. This observation highlights SPT-FSC's proficiency in capturing and leveraging fine-grained local features, enabling it to excel in distinguishing small and intricate targets. In essence, these findings underscore the effectiveness of SPT-FSC in improving classification accuracy and its potential for enhancing the model's performance in scenarios with challenging and intricate visual distinctions.

Comparing the meta-learning approaches (MatchingNet, ProtoNet, RelationNet, and MAML) with transfer learning methods (Baseline, Baseline++, and MetaBaseline), we observe that transfer learning-based methods generally perform better. This indicates that transfer learning is more suitable for handling the scarcity of labeled examples in few-shot scenarios, especially for the SAR image of ship domain. Among the meta-

learning methods, MAML shows relatively consistent performance across the three datasets. MatchingNet, ProtoNet, RelationNet, and MAML perform similarly across all datasets, with a slight variation in performance. This suggests that these meta-learning approaches have comparable abilities in learning discriminative features from few-shot examples. However, their performance is still outperformed by the transfer learning-based methods, particularly our proposed SPT-FSC approach. Baseline and Baseline++ demonstrate improved performance compared to meta-learning approaches, especially in 5-shot classification tasks. This indicates that transfer learning-based methods are more effective in few-shot classification tasks. The Baseline++ method outperforms the Baseline method, showing the importance of fine-tuning pre-trained models on the small support set and utilizing the cosine similarity metric to enhance the distance metric learning process. The NegMargin method shows a similar performance to Baseline++ in 5-shot classification tasks but has slightly worse performance in the 1-shot classification tasks. This suggests that although it is an effective approach, it might not be as robust as Baseline++ in few-shot classification tasks. MetaBaseline shows competitive performance compared to Baseline and Baseline++, but it still falls short of our proposed SPT-FSC method. This indicates that incorporating meta-learning techniques and auxiliary loss functions can improve few-shot classification performance but might not be sufficient to address the challenges posed by SAR image classification tasks. Our proposed SPT-FSC method outperforms all other methods across all datasets and both 1-shot and 5-shot classification tasks. This indicates that incorporating the SPT, SPT embedding, and RFFA significantly enhances the model's ability to learn discriminative features from few-shot examples in SAR images. Additionally, the integration of OpenSARShip and FUSAR-Ship datasets helps the model address the limited diversity in training data and further improves classification performance.

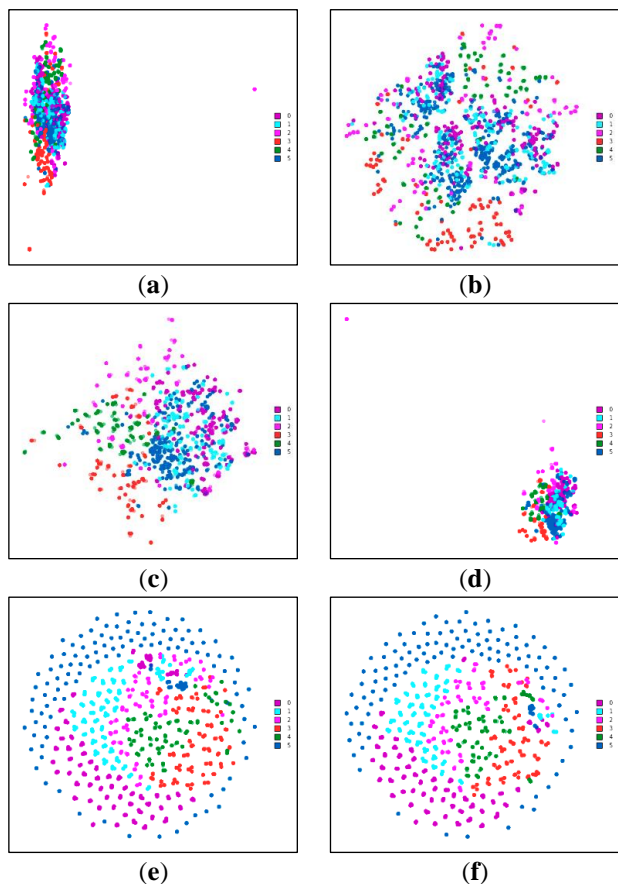


Fig. 9. t-SNE visualization of the output features on Open-FUSAR Dataset. (a) ProtoNet for 1-shot. (b) ProtoNet for 5-shot. (c) Baseline++ for 1-shot. (d) Baseline++ for 5-shot. (e) SPT-FSC for 1-shot. (f) SPT-FSC for 5-shot. The numbers 0-5 in the legend represent ContainerShip, GeneralCargo, oil_ChemicalTanker, PilotVessel, Tanker-HazardA and Unspecified, respectively.

In order to conduct a comprehensive comparison of various methods employed in the few-shot SAR image classification task, we assessed several approaches, namely, the conv-BiLSTM prototypical network (CBLPN) [64], hybrid inference network (HIN) [65], and mixed loss graph attention network (MGA-Net) [66]. These methods were evaluated using the OpenSARship dataset, with experiments consistently employing 3-way-1shot and 3-way-5shot settings for consistency and comparability. The specific configurations of base classes and novel classes are presented in Table VI. In contrast to other few-shot SAR ATR methods, SPT-FSC has made specific parameter choices by setting the number of scattering key points (N) to 30 and the number of nearest neighbors (K) to 2. These parameter selections were informed by the thorough analysis of the effects of varying N and K, which were examined in ablation experiments performed previously on the OpenSARShip dataset. Upon analyzing the results in Table VII, it becomes evident that the proposed method, denoted as SPT-FSC, achieved the highest average accuracies, reaching 70.80% in the 3-way 1-shot scenario and 88.80% in the 3-way 5-shot scenario. These results represent a significant improvement over existing few-shot SAR ATR (Automatic Target Recognition) methods, surpassing them by

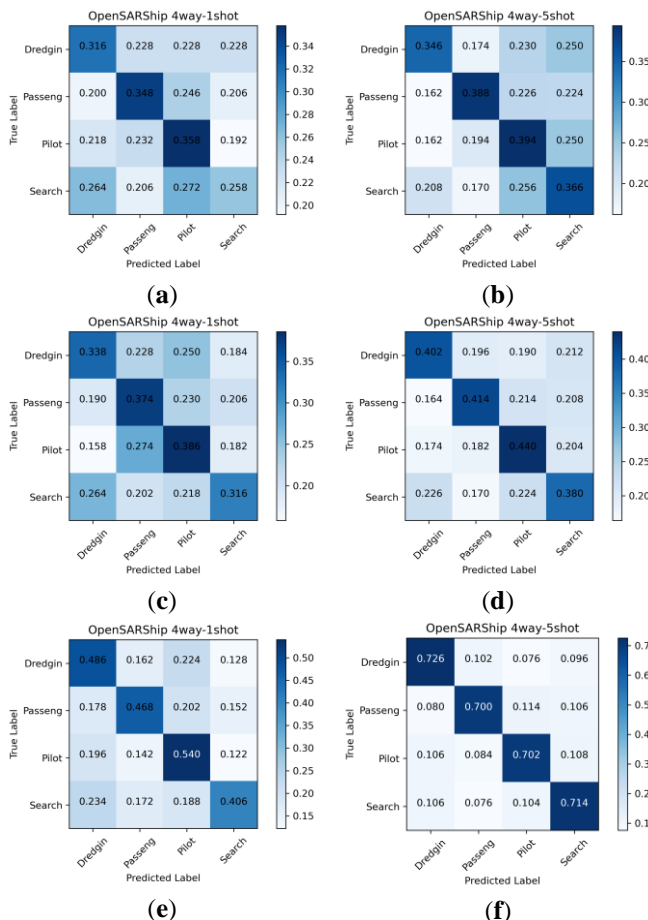


Fig. 10. Visualization of the results by confusion matrices on OpenSARship Dataset. (a) ProtoNet for 1-shot. (b) ProtoNet for 5-shot. (c) Baseline++ for 1-shot. (d) Baseline++ for 5-shot. (e) SPT-FSC for 1-shot. (f) SPT-FSC for 5-shot.

at least 0.51% and 9.11%, respectively. Notably, SPT-FSC demonstrates substantial advantages in the 3-way 5-shot scenario, indicating its robust performance. Furthermore, it is worth noting that CBLPN, HIN, and MGA-Net also exhibit competitive performance. CBLPN utilizes long short-term memory (LSTM) to integrate sequence information from SAR images after embedding, while HIN and MGA-Net employ graph networks to model relationships among samples. However, a limitation of these methods is their failure to explore fine-grained features thoroughly, resulting in biased feature distribution estimation for novel categories. This suggests that there is room for improvement in these existing approaches to further enhance their performance in few-shot SAR image classification tasks.

TABLE VI
OPENSARSHIP DATASET SETTINGS FOR FEW-SHOT SAR ATR METHODS.

Base Classes	Novel Classes
Cargo	Dredgin
Tanker	Tug
Others	fishing

TABLE VII
COMPARISON WITH FEW-SHOT SAR ATR METHODS.

Method	3-way 1-shot	3-way 5-shot
CBLPN [64]	67.51 \pm 3.14	78.89 \pm 1.15
HIN [65]	60.59 \pm 2.42	77.29 \pm 0.91
MGA-Net [66]	70.29 \pm 2.20	79.69 \pm 0.41
SPT-FSC (ours)	70.80 \pm 2.58	88.80 \pm 2.00

V. DISCUSSION

In this paper, we introduced the SPT-FSC method, which addresses the challenges of few-shot ship classification in SAR images by leveraging SPT based on scattering key points. While our approach shows promise, it's essential to acknowledge its potential limitations and suggest directions for further improvement.

One of the key limitations of our method is the parameter selection process. The choice of parameters, such as the number of scattering key points and nearest neighbors, plays a crucial role in optimizing the model's performance. However, these optimal settings may not be universally applicable and can vary depending on the specific dataset and task at hand. To mitigate this limitation, future research should focus on developing automated techniques for parameter adaptation. These techniques can reduce the reliance on manual fine-tuning and enhance the model's adaptability to different scenarios.

The introduction of SPT, TEB and RFFA has positive impacts on the model's training process, enhancing its ability to understand topological information and potentially reducing the number of training epochs required for convergence. However, these improvements come with negative impacts as well, including increased computational burden during training due to additional computations for SPT and RFFA, which can extend training times, particularly when dealing with complex SAR data or large datasets and model architectures.

Expanding the applicability of the SPT concept is another promising avenue for improvement. While the SPT-FSC method is tailored for ship classification, it's worth exploring the extension of the SPT approach to classify a broader range of objects in SAR images. Investigating the adaptability and generalization of SPT to different targets can enhance the versatility and utility of the method. One of the primary motivations for introducing SPT in the realm of few-shot classification is the inherent challenge posed by a limited number of training samples. In such scenarios, traditional deep learning networks often encounter difficulties in thoroughly learning the distinctive features and characteristics of the target class during the training phase. Consequently, these networks may exhibit suboptimal performance when tested on unseen data. The incorporation of SPT directly addresses this limitation by enriching the

network's knowledge base with additional information, thereby bolstering its capacity to learn and generalize effectively. While the potential application of SPT to general image classification is a noteworthy consideration, its impact may be more pronounced in few-shot classification settings. This distinction arises from the substantial differences in data availability between these two scenarios. General image classification typically benefits from access to extensive datasets, enabling deep learning networks to acquire comprehensive feature representations through an abundance of training samples. In such cases, the existing feature extraction mechanisms within the network are already relatively well-equipped to handle the classification task. Consequently, the introduction of SPT may not yield as substantial an improvement in standard image classification scenarios, as the network's inherent feature learning capabilities are relatively robust.

Furthermore, integrating the SPT-FSC method with SAR object detection techniques represents a significant opportunity for future research. The current trend in deep learning networks often combines object detection and classification. By incorporating SPT based classification techniques into object detection networks, we can enhance the accuracy of object recognition in SAR imagery. This integration can lead to more precise and robust target identification, particularly when dealing with SAR images captured under varying conditions and from different angles.

Moreover, leveraging the topological information encoded by SPT can improve SAR object detection itself. For example, it can facilitate more accurate target localization, which is crucial for applications like tracking and situational awareness. Additionally, SPT can assist in distinguishing genuine targets from false alarms, thereby elevating the overall accuracy of object detection in SAR images.

VI. CONCLUSIONS

In summary, this paper presents a novel framework named SPT-FSC to tackle the unique challenges associated with few-shot classification in SAR images. Our main contributions include the development of scattering point topology, the creation of a TEB to acquire SPT embedding, the design of a RFFA mechanism, and the assembly of a refined ship classification dataset. The experimental results highlight the efficacy of the proposed SPT-FSC method for ship classification in SAR images. The SPT captures the essential structural and shape features of ship targets, while the TEB enables the network to leverage the inherent information of each scattering point. The RFFA mechanism bolsters the discriminative power of the fused features by taking into account the reciprocal relationships and contributions of features from different sources. Additionally, the construction of a refined ship classification dataset allows the network to more effectively tackle the few-shot ship classification problem in SAR images. A series of analysis

provide insights into the contribution of each component to the overall performance of our proposed model. These studies confirm the importance of each component in enhancing the overall performance and offer valuable insights into the model's behavior. By addressing the imaging variability, training data diversity, and the need for specific adaptations and modifications to existing few-shot classification methods, our proposed SPT-FSC method demonstrates considerable improvement in few-shot ship classification in SAR images.

REFERENCES

- [1] M. I. Skolnik, "Introduction to Radar Systems, 4th ed.," McGraw-Hill Education, New York, NY, USA, 2019, ISBN 978-1-260-12656-2.
- [2] K. Ouchi, "Current status on vessel detection and classification by synthetic aperture radar for maritime security and safety," in Proceedings of the 38th Symposium on Remote Sensing for Environmental Sciences, Gamagori, Aichi, Japan, 2016.
- [3] X. Lu, P. Han, and R. Wu, "Research on mixed PCA/ICA for SAR image feature extraction," in 2008 9th International Conference on Signal Processing, Beijing, China, 2008, pp. 2465-2468, doi: 10.1109/ICOSP.2008.4697648.
- [4] A. Tombak, E. Aptoula, and K. Kayabol, "Pixel-based classification of SAR images using features," in 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2018, pp. 1-4, doi: 10.1109/SIU.2018.8404526.
- [5] X. Zhang et al., "Multitask Learning for Ship Detection from Synthetic Aperture Radar Images," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 8048-8062, 2021, doi: 10.1109/JSTARS.2021.3102989.
- [6] W. Xiong and L. Cao, "Automatic Target Recognition Based on Rough Set-Support Vector Machine in SAR Images," in 2009 International Joint Conference on Computational Sciences and Optimization, Sanya, China, 2009, pp. 489-491, doi: 10.1109/CSO.2009.27.
- [7] Y. Ji, P. Zeng, W. Zhang, and L. Zhao, "Forest Biomass Inversion based on KNN-FIFS with Different Alos Data," in 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 2021, pp. 4540-4543, doi: 10.1109/IGARSS47720.2021.9554712.
- [8] L. Loosvelt, J. Peters, H. Skriver, B. De Baets, and N. E. C. Verhoest, "Impact of Reducing Polarimetric SAR Input on the Uncertainty of Crop Classifications Based on the Random Forests Algorithm," in IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no. 10, pp. 4185-4200, Oct. 2012, doi: 10.1109/TGRS.2012.2189012.
- [9] X. Xing et al., "Feature selection and weighted SVM classifier-based ship detection in PolSAR imagery," in International Journal of Remote Sensing, vol. 34, no. 22, pp. 7925-7944, 2013.
- [10] C. Wang, H. Zhang, F. Wu, B. Zhang, and S. Tian, "Ship classification with deep learning using COSMO-SkyMed SAR data," in 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 2017, pp. 558-561, doi: 10.1109/IGARSS.2017.8127014.
- [11] L. Zhang, L. Zhang, and B. Du, "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art," in IEEE Geoscience and Remote Sensing Magazine, vol. 4, no. 2, pp. 22-40, June 2016, doi: 10.1109/MGRS.2016.2540798.
- [12] K. Fu, T. Zhang, Y. Zhang, Z. Wang, and X. Sun, "Few-Shot SAR Target Classification via Metalearning," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-14, 2022, Art no. 2000314, doi: 10.1109/TGRS.2021.3058249.
- [13] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-Learning in Neural Networks: A Survey," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 9, pp. 5149-5169, 1 Sept. 2022, doi: 10.1109/TPAMI.2021.3079209.
- [14] J. Snell, K. Swersky, and R. Zemel, "Prototypical Networks for Few-shot Learning," in Advances in Neural Information Processing Systems, Curran Associates, Inc., Red Hook, NY, USA, 2017, pp. 4077-4087.
- [15] D. Lu, L. Cao, and H. Liu, "Few-Shot Learning Neural Network for SAR Target Recognition," in 2019 6th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), Xiamen, China, 2019, pp. 1-4, doi: 10.1109/APSAR46974.2019.9048517.
- [16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," in Nature, vol. 521, no. 7553, pp. 436-444, 2015.
- [17] C. Zhang et al., "Deepemd: Few-shot image classification with differentiable earth mover's distance and structured classifiers," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
- [18] C. Finn, P. Abbeel, and S. Levine, "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks," in Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6-11 August 2017, pp. 70-80.
- [19] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," in International Conference on Learning Representations, 2018.
- [20] Y. Bengio, "Deep learning of representations: Looking forward," in IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 4, pp. 582-598, 2014.
- [21] K. Li et al., "Adversarial feature hallucination networks for few-shot learning," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
- [22] X. Wang et al., "Frustratingly simple few-shot object detection," arXiv preprint arXiv:2003.06957, 2020.
- [23] B. Sun et al., "Fsc: Few-shot object detection via contrastive proposal encoding," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
- [24] L. Qiao et al., "Defrcn: Decoupled faster r-cnn for few-shot object detection," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.
- [25] C. Lang et al., "Global Rectification and Decoupled Registration for Few-Shot Segmentation in Remote Sensing Imagery," in IEEE Transactions on Geoscience and Remote Sensing, 2023.
- [26] H. Zhang, N. M. Nasrabadi, Y. Zhang, and T. S. Huang, "Multi-View Automatic Target Recognition using Joint Sparse Representation," in IEEE Transactions on Aerospace and Electronic Systems, vol. 48, no. 3, pp. 2481-2497, JULY 2012, doi: 10.1109/TAES.2012.6237604.
- [27] S. Deng et al., "Improving Few-Shot Image Classification with Self-supervised Learning," in Cloud Computing-CLOUD 2022: 15th International Conference, Held as Part of the Services Conference Federation, SCF 2022, Honolulu, HI, USA, December 10-14, 2022, Proceedings. Springer Nature Switzerland, 2022.
- [28] J. Tang, F. Zhang, Y. Zhou, Q. Yin, and W. Hu, "A Fast Inference Networks for SAR Target Few-Shot Learning Based on Improved Siamese Networks," in IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 2019, pp. 1212-1215, doi: 10.1109/IGARSS.2019.8898180.
- [29] R. Ran et al., "Deep transformer and few-shot learning for hyperspectral image classification," in CAAI Transactions on Intelligence Technology, 2023.
- [30] P. Clemente-Colón and X.-H. Yan, "Low-backscatter ocean features in synthetic aperture radar imagery," in Johns Hopkins APL Technical Digest, vol. 21, no. 1, pp. 116-121, 2000.
- [31] N. Inkawhich, "A Global Model Approach to Robust Few-Shot SAR Automatic Target Recognition," in IEEE Geoscience and Remote Sensing Letters, vol. 20, pp. 1-5, 2023, Art no. 4004305, doi: 10.1109/LGRS.2023.3264535.
- [32] Y. Zhang et al., "Scattering-Point-Guided RPN for Oriented Ship Detection in SAR Images," in Remote Sensing, vol. 15, p. 1411, 2023, doi: 10.3390/rs15051411.
- [33] L. Huang et al., "OpenSARShip: A Dataset Dedicated to Sentinel-1 Ship Interpretation," in IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 11, pp. 195-208, 2018, doi: 10.1109/JSTARS.2017.2755672.
- [34] 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, pp. 40-58, 2020.
- [35] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra, et al., "Matching Networks for One Shot Learning," in Advances in Neural Information Processing Systems, Red Hook, NY, USA, 2016, pp. 3630-3638.
- [36] F. Sung, Y. Yang, L. Zhang, T. Xiang, et al., "Learning to Compare: Relation Network for Few-Shot Learning," IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 8, pp. 2032-2045, 2020.
- [37] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese Neural Networks for One-shot Image Recognition," in Proceedings of the International

Conference on Machine Learning, Lille, France, 6–11 July 2015, pp. 956–964.

[38] S. Kim et al., "DynaGAN: Dynamic Few-shot Adaptation of GANs to Multiple Domains," in SIGGRAPH Asia 2022 Conference Papers, 2022.

[39] A. Sinha et al., "D2c: Diffusion-decoding models for few-shot conditional generation," in Advances in Neural Information Processing Systems 34, 2021, pp. 12533–12548.

[40] M. Yang et al., "ProtoGAN: Towards high diversity and fidelity image synthesis under limited data," *Information Sciences*, vol. 632, pp. 698–714, 2023.

[41] T. Chen et al., "A Closer Look at Few-shot Classification," in Proceedings of the International Conference on Learning Representations, 2019.

[42] Y. Chen, Z. Liu, H. Xu, et al., "Meta-Baseline: Exploring Simple Meta-Learning for Few-Shot Learning," in Proceedings of the International Conference on Computer Vision, 2021.

[43] X. Zhao et al., "Few-Shot SAR-ATR Based on Instance-Aware Transformer," *Remote Sensing*, vol. 14, no. 8, 1884, 2022.

[44] F. Gao et al., "A Few-Shot Learning Method for SAR Images Based on Weighted Distance and Feature Fusion," *Remote Sensing*, vol. 14, no. 18, 4583, 2022.

[45] J. Cai, Y. Zhang, J. Guo, X. Zhao, J. Lv, and Y. Hu, "ST-PN: A Spatial Transformed Prototypical Network for Few-Shot SAR Image Classification," *Remote Sensing*, vol. 14, no. 9, p. 2019, 2022, doi: 10.3390/rs14092019.

[46] P. Zhao, L. Huang, Y. Xin, J. Guo, and Z. Pan, "Multi-Aspect SAR Target Recognition Based on Prototypical Network with a Small Number of Training Samples," *Sensors*, vol. 21, p. 4333, 2021, doi: 10.3390/s21134333.

[47] K. Wang and G. Zhang, "SAR Target Recognition via Meta-Learning and Amortized Variational Inference," *Sensors*, vol. 20, no. 20, p. 5966, 2020, doi: 10.3390/s20205966.

[48] Z. Pan, X. Bao, Y. Zhang, B. Wang, Q. An, and B. Lei, "Siamese Network Based Metric Learning for SAR Target Classification," in IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 2019, pp. 1342–1345, doi: 10.1109/IGARSS.2019.8898210.

[49] M. Rostami, S. Kolouri, E. Eaton, and K. Kim, "Deep Transfer Learning for Few-Shot SAR Image Classification," *Remote Sensing*, vol. 11, no. 11, p. 1374, 2019, doi: 10.3390/rs11111374.

[50] D. Lu, L. Cao, and H. Liu, "Few-shot learning neural network for SAR target recognition," in 2019 6th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR). IEEE, 2019.

[51] Y. Tai et al., "Few-shot transfer learning for sar image classification without extra sar samples," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 2240–2253, 2022.

[52] S. Chen et al., "Few shot object detection for SAR images via feature enhancement and dynamic relationship modeling," *Remote Sensing*, vol. 14, no. 15, p. 3669, 2022.

[53] Z. Zhou et al., "FSODS: A Lightweight Metalearning Method for Few-Shot Object Detection on SAR Images," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–17, 2022, Art no. 5232217, doi: 10.1109/TGRS.2022.3192996.

[54] K. Fu, J. Fu, Z. Wang, and X. Sun, "Scattering-keypoint-guided network for oriented ship detection in high-resolution and large-scale SAR images," *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.*, vol. 14, pp. 11162–11178, 2021.

[55] Y. Kang, Z. Wang, J. Fu, X. Sun, and K. Fu, "SFR-Net: Scattering feature relation network for aircraft detection in complex SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2021.

[56] Y. Sun, Z. Wang, X. Sun, and K. Fu, "SPAN: Strong Scattering Point Aware Network for Ship Detection and Classification in Large-Scale SAR Imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 15, pp. 1188–1204, 2022.

[57] X. Sun, Y. Lv, Z. Wang, and K. Fu, "SCAN: Scattering Characteristics Analysis Network for Few-Shot Aircraft Classification in High-Resolution SAR Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022.

[58] S. Zhang, et al., "Graph convolutional networks: a comprehensive review," *Comput. Soc. Networks*, vol. 6, no. 1, pp. 1–23, 2019.

[59] I. O. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, and A. Dosovitskiy, "MLP-Mixer: An All-MLP Architecture for Vision," in Adv. Neural Inf. Process. Syst., vol. 34, pp. 24261–24272, 2021.

[60] C. G. Harris and M. J. Stephens, "A combined corner and edge detector," in Proceedings of the Alvey Vision Conference, 1988.

[61] D. Hendrycks and K. Gimpel, "Gaussian Error Linear Units (GELUs)," arXiv preprint arXiv:1606.08415, 2016.

[62] C.-F. R. Chen, Q. Fan, and R. Panda, "CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

[63] B. Liu, et al., "Negative margin matters: Understanding margin in few-shot classification," in Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV, Springer International Publishing, 2020.

[64] L. Wang, X. Bai, R. Xue, and F. Zhou, "Few-shot SAR automatic target recognition based on conv-BiLSTM prototypical network," *Neurocomputing*, vol. 443, pp. 235–246, Jul. 2021.

[65] L. Wang, X. Bai, C. Gong, and F. Zhou, "Hybrid inference network for few-shot SAR automatic target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9257–9269, Nov. 2021.

[66] M. Yang, X. Bai, L. Wang, and F. Zhou, "Mixed loss graph attention network for few-shot SAR target classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5216613.



Yipeng Zhang received the B.Sc. degree from China University of Mining and Technology, Xuzhou, China, in 2021. He is currently pursuing the M.Sc. degree with the Suzhou Aerospace Information Research Institute, Chinese Academy of Sciences, Suzhou, China, and the University of Chinese Academy of Sciences, Beijing, China.

His research interests include computer vision and synthetic aperture radar (SAR) image processing, especially on SAR object detection and recognition.



Dongdong Lu received the M.Sc. degrees from State Key Laboratory of Transient Physics, Nanjing University of Science and Technology. He is currently an assistant researcher at Suzhou Aerospace Information Research Institute, Chinese Academy of Sciences. And he also works at Suzhou Key Laboratory of Microwave Imaging, Processing and Application

Technology.

His research interests include intelligent interpretation of remote sensing images, target detection and recognition, target tracking and perceptual fusion.



Xiaolan Qiu received the B.Sc. degree in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 2004, and the Ph.D. degree in signal and information processing from the Graduate University of Chinese Academy of Sciences, Beijing, China,

in 2009. Since 2009, she has been with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. Since 2020, she has also been with the Suzhou Aerospace Information Research Institute, Suzhou, China.

Her research interests include synthetic aperture radar (SAR) 2-D/3-D imaging and microwave vision. She is the senior member of IEEE. She was funded by the Excellent Young Scientists Fund of NSFC. She serves as the associate editor of IEEE Geoscience and Remote Sensing Letters and Young Associate Editor of Journal of Radars.



Fei Li received the B.Sc. degree in communication engineering from Northern Jiaotong University, Beijing, China, in 1998, and the M.Sc. degree in information and communication Engineering from the University of Science and Technology of China, Hefei, China, in 2004.

He is currently a Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include remote sensing and information processing.