

SCAN: Scattering Characteristics Analysis Network for Few-Shot Aircraft Classification in High-Resolution SAR Images

Xian Sun, *Senior Member, IEEE*, Yixuan Lv, Zhirui Wang, *Member, IEEE*, and Kun Fu, *Member, IEEE*

Abstract—Recently, deep learning in synthetic aperture radar (SAR) automatic target recognition (ATR) has made significant progress, but the sample limitation problem in the SAR field is still obvious. Compared with the optical remote sensing images, the SAR images are insufficient, especially those containing the geospatial targets with certain target attitude angles (TAAs). To solve these problems, a novel few-shot learning framework named scattering characteristics analysis network (SCAN) is proposed in this article. First, a scattering extraction module (SEM) is designed to combine the target imaging mechanism with the network, which learns the number and distribution of the scattering points for each target type via explicit supervision. Besides, considering the imaging variability of SAR targets, a TAA-guided meta-learning network consisting of an angle self-adaption classifier (ASC) and a frequency embedded module (FEM) is designed. ASC guides the network to focus on the positive sample pairs with different TAAs. FEM combines PCT with the network training process effectively to enrich frequency domain information. Additionally, a new dataset named SAR aircraft category dataset (SAR-ACD) is constructed for the experiments. Compared with other few-shot SAR target classification approaches, our model efficiently integrates the scattering characteristics with the learning process and the test accuracy for 5-way 1-shot has been improved by 4.74%. Finally, the experimental results are provided to demonstrate the validity of the proposed method.

Index Terms—synthetic aperture radar (SAR), Automatic target recognition (ATR), few-shot learning (FSL), aircraft, scattering characteristics, classification, deep learning.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) is an active imaging system with all-time and all-weather observing ability. Recently, with the rapid development of radar

satellites, the demand for SAR image interpretation is increasing in disaster monitoring, topographic mapping, target detection and classification, etc [1]. Particularly, SAR target classification has attracted more and more attention among researchers.

Over the years, the research on SAR target classification has made significant progress. There are roughly three ways: template-based, model-based, and machine learning-based. The template-based methods compare the test samples with the sample template of each class to classify the targets [2]. The model-based approaches describe the target characteristics with the help of physical models or conceptual models [3], [4]. For machine learning, several classifiers such as AdaBoost and SVM are designed for more accurate classification [5]–[7]. However, these approaches need a high cost in manual experience for both feature selection and classifier construction. Therefore, it is necessary to apply the convolutional neural network (CNN) to extract the target features comprehensively and automatically.

Benefitted from the success of deep learning in computer vision, deep learning has been introduced into the SAR domain [8]–[10] recently. However, deep learning models are data-driven, and the application of general deep learning methods in the SAR target classification task faces the challenge of insufficient samples for several reasons. First, limited by the way of data acquisition, the dataset in the remote sensing field is generally smaller than that in the natural scene. Compared with the optical remote sensing images, the SAR images are more insufficient [11]–[15], especially the images containing geospatial targets. Second, according to the SAR mechanism, the SAR target characteristics vary significantly with different TAAs, as shown in Fig. 1. Generally, the geospatial targets in the SAR images are presented with certain TAA, which cannot meet the requirement of dataset diversity for deep learning. Furthermore, the rotation operation for data augmentation commonly used in optical images is not reasonable for SAR images. Third, the SAR target interpretation is a laborious task, which requires both the expert experience and the prior knowledge. Hence, the problem of lacking SAR samples with certain TAAs becomes more seriously. In addition, deep learning algorithms are prone to overfitting in low data regimes, which decline the classification perfor-

This work was supported by the National Nature Science Foundation of China under Grant 61725105 and 62076241. (Corresponding author: Xian Sun.)

Xian Sun, Yixuan Lv, and Kun Fu are with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the Key Laboratory of Network Information System Technology (NIST), Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China, and also with the School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: sunxian@mail.ie.ac.cn; lvyixuan19@mails.ucas.ac.cn; fukun@mail.ie.ac.cn).

Zhirui Wang is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the Key Laboratory of Network Information System Technology (NIST), Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China (e-mail: zhirui1990@126.com).

mance. Consequently, it is essential to apply the few-shot learning to the SAR target classification tasks.

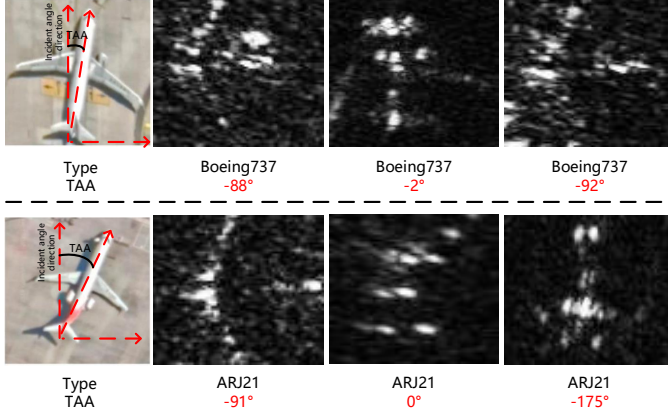


Fig. 1: Optical remote sensing images and the SAR images of the same category. The Top is Boeing737, and the bottom is ARJ21. The TAA in this article is denoted as follows: The direction of the incident angle is denoted as 0° . It goes from 0° to -180° counterclockwise and from 0° to 180° clockwise.

In the SAR domain, few-shot learning methods roughly start from two directions: data expansion and model optimization [16]. Data expansion is a direct method to solve the sample limitation problem, including traditional data expansion methods like brightness and contrast conversion, and synthesis of new samples through simulation software or generative models [17]–[19]. However, data-expansion methods for SAR images are complicated and tend to affect the classification efficiency. For model optimization, transfer learning and meta-learning are two representative methods. Transfer learning aims to establish a connection between source domain and target domain, and maximizes the use of source domain knowledge to solve the problem of insufficient samples in the target domain [19]–[28]. Meta-learning is upgraded from sample-level learning to task-level learning, with the aim of achieving rapid learning of new tasks with only a few samples [29]–[36].

However, most previous model optimization studies directly apply the general models in natural scenes and do not consider the different imaging properties between SAR and optical images. Meanwhile, compared with the SAR ships, the discrete structure of the aircraft makes it more meaningful to take advantage of scattering characteristics. Generally, different from the optical imaging mechanism, the imaging result of SAR aircraft is composed of many discrete strong scattering points. The strong scattering points represent the maximum pixel value of the local scattering region in SAR images and have superior saliency and stability properties [37], [38]. Therefore, combining with the strong scattering points is a feasible way to optimize the performance of SAR aircraft classification. Through the above analysis, a novel few-shot SAR target classification network named

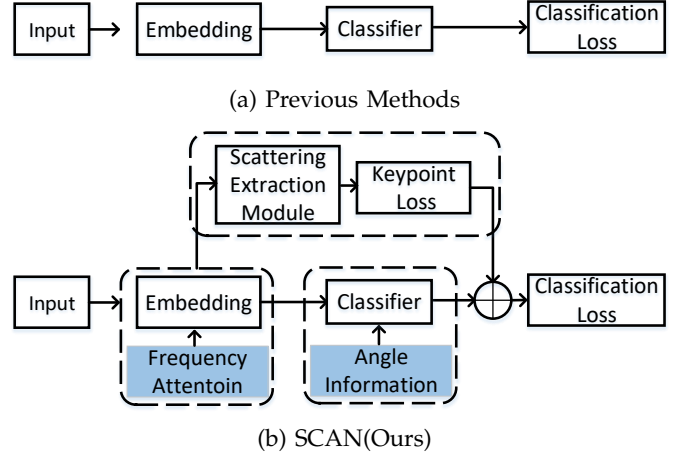


Fig. 2: Illustration of the few-shot learning SAR target classification methods. (a) Previous methods are the general few-shot learning methods in the natural scene. (b) SCAN (Ours) that combines scattering characteristics.

scattering characteristics analysis network (SCAN) is proposed. Different from previous methods, SCAN integrates the scattering characteristic and TAAs with the network learning process as exhibited in Fig. 2. A scattering extraction module (SEM) serves as an auxiliary task to introduce additional supervision information to the network. Meanwhile, the target imaging result is sensitive to the TAAs, which makes the angular distribution of the aircraft is sparse. Hence, a TAA-guided meta-learning network is designed with the guidance of the TAAs. Compared with the few-shot SAR target classification approach [39] involving new categories, our method adopts the metric learning structure where fine-tuning is not required during the testing procedure.

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

- 1) We present a novel integrated framework named SCAN for few-shot SAR aircraft classification, which consists of a classification path and a scattering extraction branch. Our method shows the best performance on the novel category few-shot recognition tasks without fine-tuning.
- 2) To enhance SAR target feature representation, SEM is designed to assist the optimization of the main classification task. It can guide the network to learn the number of strong scattering points and their spatial location distribution of different categories.
- 3) To alleviate the imaging variability of SAR targets caused by TAAs, a TAA-guided meta-learning network consisting of a frequency embedded module (FEM) and an angle self-adaption classifier (ASC) is proposed. FEM combines PCT with the training process effectively to enrich frequency domain information. ASC guides the network to focus on the positive sample pairs with different TAAs during training and learn more representative and distinguishing features.

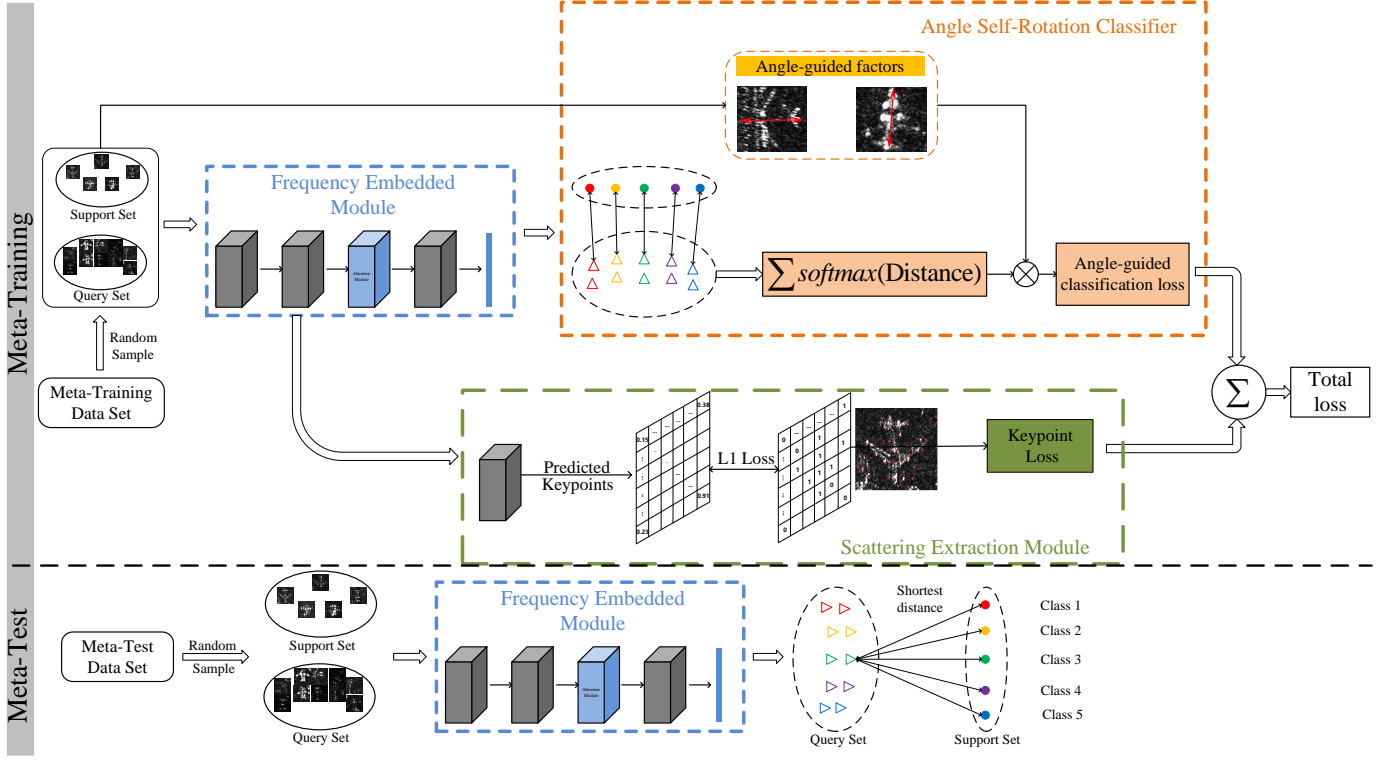


Fig. 3: The overall structure of SCAN. For the 5-way 5-shot case, angle-guided factors are not used, and other modules are reserved. Both the SEM and the angle-guided factors are not required in the testing phase.

A new SAR aircraft classification dataset (SAR-ACD) is built for research. Experimental results on SAR-ACD show the superiority of SCAN. Compared with the baseline, the accuracy on the 5-way 1-shot and 5-way 5-shot tasks are improved by 4.74% and 3.18%, respectively.

II. RELATED WORK

A. Data Expansion-Based Methods

For data expansion methods [40]–[44], they attempt to enlarge the data set by generating new samples through adversarial learning or electromagnetic simulation. In [3], a new SAR target recognition method combining Computer-aided design (CAD) model with CNN is proposed. CAD models can flexibly generate images and features in different frequency bands and angles. The RaySAR method uses ray tracing technology to build simulation samples based on the three-dimensional model of the target [45]. Specially, to solve the problem of lack of samples with specific TAAs, Song et al. [18] propose an adversarial autoencoder to generate SAR images with aspect angular diversity. Cui et al. [46] use WGAN based on gradient penalty to generate new samples. They also propose a sample selection principle to select good samples with specific TAAs from the generated samples. Ding et al. [47] propose a sample synthesis method, which generates new samples with a specific azimuth angle by performing a weighted summation of the two samples with the closest angle.

When we can guarantee the quality of the generated samples, data expansion-based methods are significantly effective and can solve the sample limitation problem fundamentally. However, most expansion methods are a little complex, and the flow is relatively slow.

B. Methods Based on Transfer Learning

Transfer learning methods [48]–[54] utilize the prior knowledge to alleviate the sample limitation problem in the target domain. The prior knowledge is learned from the source domain where the training data is sufficient. [23] uses the optical CIFAR-10 data set to construct a pre-trained network, and the middle feature layer is used for the classification of TerraSAR-X images. In [26], [53], the source data is replaced by simulated data. The transferred knowledge is learned from simulated data by constructing pre-trained models. Specially, to solve the problem of scarce samples with specific TAAs, [55] learns the mapping relation from sufficient data in the source domain, and the mapping can be applied to the target domain to generate targets with different aspect angles.

Generally, transfer learning has high requirements for source domain data. The effect of transfer learning is closely related to the similarity between the source domain and the target domain dataset.

C. SAR Target Classification Based on Meta-learning

When there is only labelled data, meta-learning methods are more appropriate. Especially, meta-learning can

achieve more stable and better recognition results when new categories are involved. And when the number of samples is limited to a few-shot like 1 or 5, only meta-learning can recognize effectively.

Meta-learning aims to achieve rapid and effective learning when faced with a small sample of the new classes by learning on a large number of old tasks. Meta-learning methods can be divided into three categories: metric-based, model-based, and optimization-based [16]. Pan et al. [56] apply metric-based meta-learning to solve SAR target classification problems through the Siamese network. The combination of the metric-based meta-learning [57], [58] and the deep networks [59]–[61] is explored to achieve more effective deep embedding learning. Meta-training is implemented on the simulated data set of MSTAR to construct the general feature extractor. After that, the feature extractor parameters are fixed, and then the rest network parameters are fine-tuned on the basis of real MSTAR data set in [62].

Specially, to solve the problem of insufficient samples with specific TAAs, Zhao et al. [63] take multi-aspect SAR target as input based on metric-based framework. While in most of the existing approaches, the images for each category in MSTAR are partitioned into the test set and training set proportionally and do not involve new categories [34], [64]. Their purpose is to achieve higher accuracy by using fewer samples with the help of meta-training. Zhang et al. [39] propose a meta-learning framework which includes a meta-learner and a base-learner dedicated to learning a good initialization. Meanwhile, five categories in MSTAR are selected as the meta-training data set and the model's performance are tested on the remaining five new categories. More recently, Wang et al. [30] also apply meta-learning to SAR target recognition and propose a hybrid inference strategy for classification on MSTAR. The meta-training data set and meta-testing data set contain seven and three categories, respectively.

III. METHODOLOGY

The total structure of the proposed framework is shown in Fig. 3. The proposed network is mainly composed of three parts: ASC, SEM, and FEM. In this section, these three components will be described in detail. For simplicity, angle-guided classification loss is denoted as $Loss_{ASC}$, and the scattering extraction branch loss is denoted as $Loss_{SEM}$.

A. Problem Setup

Some meta-learning concepts will be introduced first. One episode in episode training is similar to a batch in the traditional training method. The meta-learning data set is partitioned into the meta-testing data set and meta-training data set. The categories of the two data set have no intersection. Assume that the meta-training data set $D_{train} = \{(x_1, y_1), \dots, (x_{N_{tr}}, y_{N_{tr}})\}$ contains a total of C_{tr} categories and N_{tr} samples. Meta-testing data set $D_{test} =$

$\{(x_1, y_1), \dots, (x_{N_{te}}, y_{N_{te}})\}$ contains C_{te} categories and the number of test samples is N_{te} .

The meta-learning process is shown in Fig. 4. In each training iteration, N categories are randomly selected from all training categories. Suppose there are a total of T training categories, the number of different category combinations during training can be calculated as :

$$C_T^N \cdot A_N^N = \left(\frac{T!}{N!(T-N)!} \right) \cdot (N(N-1) \cdots (1)) \quad (1)$$

Then the support set and the query set are formed by randomly selecting K examples and Q examples from each selected category, respectively. That is, the model need to learn how to classify these C types accurately from only $N \cdot K$ data, which is denoted as an N -way K -shot task. Similarly, in each testing iteration, the data formation is the same as the training process.

Different meta-tasks will be randomly sampled during the training process in each training episode. This learning mechanism allows the model to learn the commonality of different meta-tasks and reduce focus on the specific task-related parts. As a result, the models generalization ability is improved when faced with new meta-tasks.

B. Scattering Extraction Module(SEM)

In this section, a scattering extraction loss is proposed. A new convolution branch is added to extract the scattering keypoints of aircraft. In short, the network predicts the location and quantity information of keypoints simultaneously. The keypoints are extracted in the same way as described in Strong scattering point extraction.

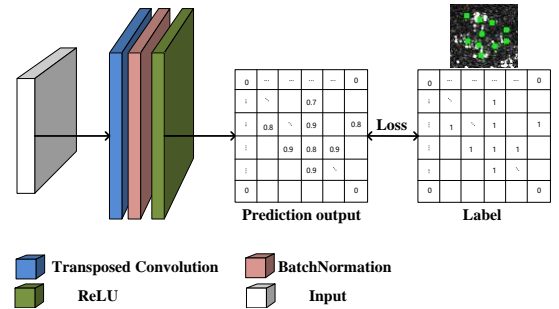


Fig. 5: The network structure of the SEM. Strong scattering points are pre-extracted by Harris-Laplace detector and serve as supervision information.

The output of the second convolution block in the feature extraction module is served as the input of the scattering extraction branch. Generally, the keypoint prediction can be considered a pixel-level classification requiring us to decode the input feature map and resume it to a larger size. As a result, the deconvolution structure is introduced in this paper to predict key points. The deconvolution process has learnable parameters which can increase the network's fitting ability compared with up pooling and upsampling operations. The structure of the scattering extraction branch is shown in Fig. 5.

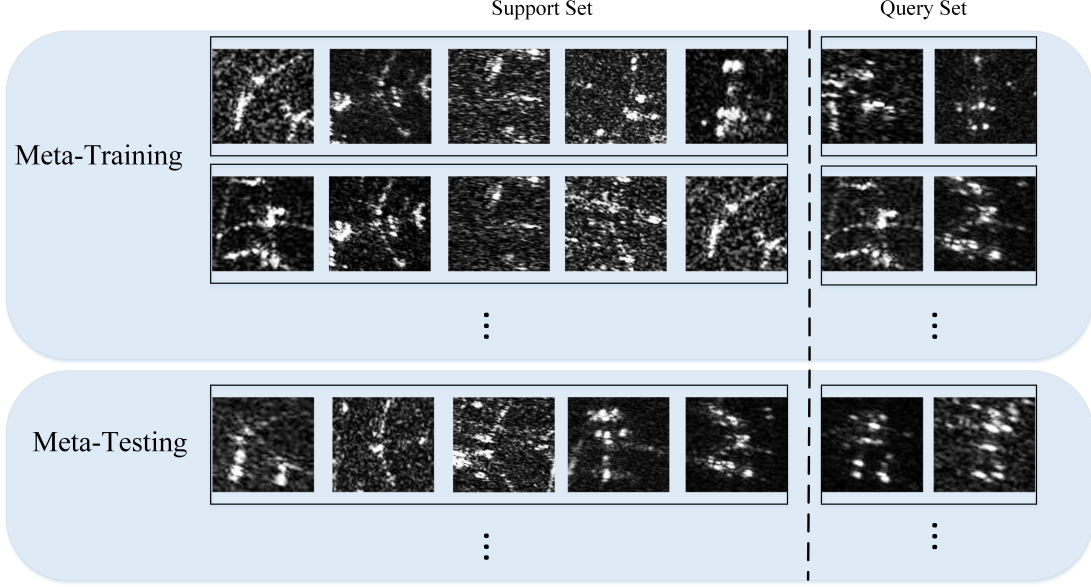


Fig. 4: The 5way 1-shot setup. There is no intersection between the training categories and test categories. In each episode, one image is chosen from each of the five selected categories to form the support set, and 15 images are chosen from each of the five selected categories to form the query set.

The coordinate conversion formula is used to calculate the position of the keypoint on the output. It is calculated as

$$y_2 = \frac{h_2}{h_1} \cdot y_1, \quad x_2 = \frac{w_2}{w_1} \cdot x_1 \quad (2)$$

where (w_1, h_1) and (w_2, h_2) illustrate the original image size and scaled size, respectively. (x_1, y_1) and (x_2, y_2) represent the original coordinates and the scaled coordinates, respectively.

Assume the corresponding label key point map is $H \in \mathbb{R}^{50 \times 50}$ and one of the corresponding scaled coordinates of the strong scattering point is (x_2, y_2) , then $H(x_2, y_2) = 1$ and other areas are set to 0. The prediction key point feature map is $\hat{H} \in \mathbb{R}^{50 \times 50}$. The keypoint prediction loss is written as:

$$Loss_{SEM} = \text{sigmoid} \left(\sum_{(i,j)} \hat{H}(i,j) - H(i,j) \right) \quad (3)$$

To encourage features representation learning via explicit supervision, the model parameters are restricted to update towards the set trajectory to generate more representative features.

C. The Angle Self-adaption Classifier(ASC)

1) *Algorithm Description*: In the few-shot SAR aircraft classification tasks, N labeled examples $S = (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ of the support set are given, where each feature vector $x_i \in \mathbb{R}^D$ and the corresponding labels is $y_i \in 1, \dots, K$. The initial feature vectors are then mapped to M -dimension through an embedding function $f_\phi : \mathbb{R}^D \rightarrow \mathbb{R}^M$ with learnable pa-

rameters ϕ [65]. Center representation vectors c_k are computed as follows:

$$c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\phi(x_i) \quad (4)$$

where S_k denotes the sample set containing samples labeled with class k .

Given a distance function $Dis : \mathbb{R}^M \times \mathbb{R}^M \rightarrow [0, +\infty)$, the distances from query samples to the center representation vectors in the mapping space can be calculated [65]. The original distribution without the angle information is shown in Equation 5:

$$p_\phi(y = k|x) = \frac{\exp(-Dis(f_\phi(x), c_k))}{\sum_{k'} \exp(-Dis(f_\phi(x), c_{k'}))} \quad (5)$$

There are many distance metric functions, such as Euclidean Distance, Manhattan Distance, Chebyshev Distance, Minkowski Distance. Experiments show that using squared Euclidean distance is better.

SAR target imaging result is angle-sensitive. For the same type of aircraft, when the incident angle of the SAR image is the same or very close, the imaging results show significant differences with the increase of the TAA difference. Therefore, the angle self-adaption classifier is designed to make the network be more concerned with the positive sample pairs hard to distinguish. Naturally, the network will learn more representative and distinguishing feature representations.

2) *Strong scattering point extraction*: The aircraft is basically composed of subcomponents such as the wings, the tail, the fuselage, the nose, and the engines. It is found during the experiment that the Harris-Laplace detector shows better robustness and scattering point detection

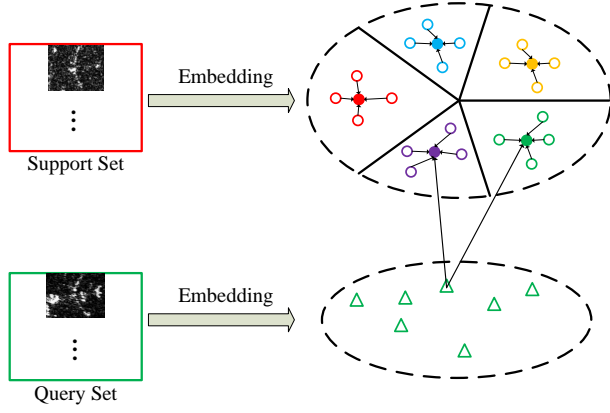


Fig. 6: The classification process for query images. The center representation vector of each class is got by calculating the mean vectors of the support set. The query vectors are classified according to the distance between them and the center representation vectors.

accuracy. Hence, the Harris-Laplace detector is selected to extract strong scattering points in this paper. Strong scattering point detection results for different types of aircraft are given in Fig. 7. For different sizes and types of aircraft, such as Boeing and Airbus, it verifies the rationality of the distribution of the extracted scattering points.

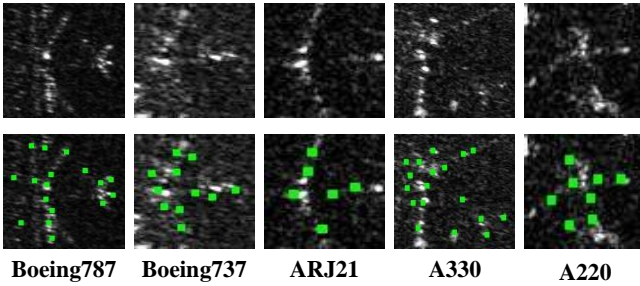


Fig. 7: Strong scattering point extraction results of different types of aircraft. Initial images are shown in the upper line, and detected images are shown in the bottom line.

3) Acquisition of the TAAs by the least square method:

It is laborious to obtain the corresponding TAAs by manual annotation for an aircraft instance. A novel method for calculating the TAAs is devised with the extracted strong scattering points. There are three types of typical SAR aircraft echo characteristics: one is the "V" type, with swept-back wings; the second is the "T" type, which is a flat-wing aircraft; the third is the double dot or "peanut" type, which is the helicopter. Considering the SAR aircraft can be regarded as a series of strong scattering points and the symmetry of the above three structures, it is reasonable to use the least square method for linear regression on these strong scattering points.

The slope of the line obtained by regression corresponds to the TAA. As shown in Fig. 7, assume n points are extracted by Harris-Laplace detector, define these points as $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$. We then calculate the respective average values of the abscissa and ordinate of them.

$$\begin{aligned}\bar{x} &= \frac{x_1 + x_2 + \dots + x_n}{n} \\ \bar{y} &= \frac{y_1 + y_2 + \dots + y_n}{n}\end{aligned}\quad (6)$$

Then the intercept and slope of linear equation can be calculated according to the average value above,

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (7)$$

$$k = \bar{y} - b\bar{x} \quad (8)$$

where b and k are the intercept and slope of the linear equation, respectively. The final angle is:

$$\theta = \arctan(k) \quad (9)$$

4) *The angle-guided classification loss function:* For the 5-way 1-shot tasks, the support set S contains $5 \times 1 = 5$ images. Suppose that there are N_Q examples in the query set Q for each episode. the original classification loss is denoted as

$$Loss = -\log\text{softmax} \left(\sum_{q \in Q} \text{sqr}t(\text{sum}((q - s_{y_q})^2)) \right) \quad (10)$$

where $q \in Q$ and $s \in S$ are the embedding vectors obtained from the feature extraction module. y_q is the corresponding category label of q , and s_{y_q} is the support vector of the corresponding category label of q . The angle-guided classification loss combined can be written as

$$Loss_{ASC} = -\log\text{softmax} \left(\sum_{q \in Q} f(\theta_q - \theta_{s_{y_q}}) \text{sqr}t(\text{sum}((q - s_{y_q})^2)) \right) \quad (11)$$

where θ_q and $\theta_{s_{y_q}}$ are the TAA of q and s_{y_q} , respectively. The $f(\cdot)$ is denoted as

$$\begin{aligned}f(x) &= \text{softmax}(\text{abs}(x)) \\ x &= [\theta_q - \theta_{s_{y_q}}], q \in Q\end{aligned} \quad (12)$$

In this situation, the network will focus more on the positive sample pairs with the same category but have large differences in the distribution of strong scattering points.

The total loss is:

$$J = Loss_{ASC} + \frac{1}{N_Q} Loss_{SEM} \quad (13)$$

By minimizing the loss of the query samples at each episode via Adam to achieve meta-learning.

D. Frequency Embedded Module(FEM)

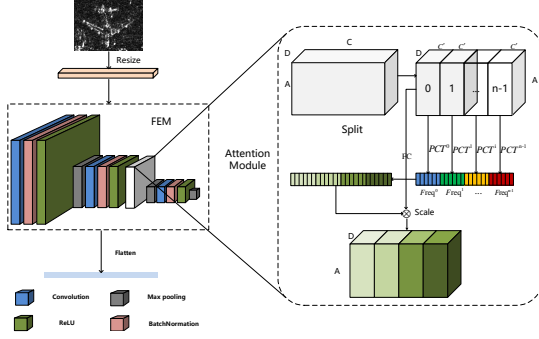


Fig. 8: The structure of the frequency embedded module. The resized images are sent to the embedding network to obtain the flattened feature vectors.

Frequency domain learning has always been a powerful measure for image processing. Frequency domain analysis should be more meaningful for SAR target interpretation considering the special imaging mechanism of SAR images.

1) *Basic Frequency Channel Attention Network*: Assume the input image is $I \in \mathbb{R}^{3 \times A \times D}$, $X \in \mathbb{R}^{C' \times A' \times D'}$ is the output feature map, D' and A' are the width and height of the feature map, respectively. C' is the number of channels. Different channels represent the image components on different convolution kernels. Similar to time-frequency transformation, the convolution operation is similar to the Fourier transform. Thus, the information of one channel can be decomposed into signal components on C convolution kernels [66]. Typically, the two-dimensional (2D) DCT is formulated as:

$$f_{a,d}^{2d} = \sum_{i=0}^{A-1} \sum_{j=0}^{D-1} x_{i,j}^{2d} \cos\left(\frac{\pi a}{A}\left(i + \frac{1}{2}\right)\right) \cos\left(\frac{\pi d}{D}\left(j + \frac{1}{2}\right)\right), \quad (14)$$

$s.t. \ a \in 0, 1, \dots, A-1, d \in 0, 1, \dots, D-1,$

where $f_{a,d}^{2d} \in \mathbb{R}^{A \times D}$ is the 2D DCT frequency spectrum, $x_{i,j}^{2d} \in \mathbb{R}^{A \times D}$ is the input, A is the height of $x_{i,j}^{2d}$, and D is the width of $x_{i,j}^{2d}$.

In [66], it has been proved that the other frequency components and information are discarded in the global average pooling except for the lowest frequency component. The larger coefficient of dimensional discrete cosine transform indicates that the visual features of the corresponding frequency and direction appear more frequently in space.

2) *PCT*: Pulse Cosine Transform (PCT) model suppresses the visual features that appear more frequently in the visual space by whitening (normalized) the amplitude values of the discrete cosine transform (DCT) coefficients. It has been proven effective for SAR image interpretation [67], [68].

After obtaining the frequency domain characteristics through the two-dimensional DCT, the transformed output will be quantized into 1, 0, and -1 in the frequency domain. This operation imitates the biological visual

mechanism. 1 and -1 transformations respectively indicate the activation and deactivation of the signal at the corresponding position, and 0 transformation means that the information is inhibited. The quantization process of the transformed feature map in the frequency domain can be expressed as:

$$I_n(u_i, v_i) = \text{sign}(I_o(u_i, v_i)) \begin{cases} -1, & \text{if } I_o(u_i, v_i) < 0 \\ 0, & \text{if } I_o(u_i, v_i) = 0 \\ 1, & \text{if } I_o(u_i, v_i) > 0 \end{cases} \quad (15)$$

where (u_i, v_i) represents the coordinates of pixel i and $I_o(u_i, v_i)$ and $I_n(u_i, v_i)$ represent the original value and the transformed value of pixel i , respectively. Besides, $\text{sign}(\cdot)$ represents the symbolic function, the positive DCT coefficient in the feature map is quantized to +1, the negative DCT coefficient is quantized to -1, and the original data is retained when the DCT coefficient is 0. Fig. 8 displays the architecture of the frequency embedded network. The feature map is partitioned into n parts in channel dimension, and each part corresponds to different parameters.

E. Combination with the Pre-training Phase

Due to the limitation of SAR data, it is meaningful to take full advantage of the existing data. Therefore, a pre-training phase is introduced to get more prior knowledge and better model initialization parameters. As described in Fig. 4 for meta-learning, the meta-learning dataset consists of the training and test set. The pre-training phase is implemented on the training set before the meta-training stage, and details are displayed in Fig. 9.

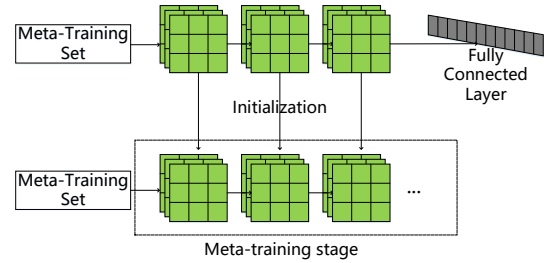


Fig. 9: The meta-learning process combined with pre-training.

IV. EXPERIMENT

In this section, we will perform different groups of experiments to verify the effectiveness of the meta-learning principle and the SCAN. According to the data set partition form, we conduct two groups of experiments.

In the first group of an experiment, the meta-test data set and meta-training data set are made up of 6 and 14 categories, respectively. For the second group of an experiment, all 20 categories are used in both the training and test sets. The images of each class are partitioned

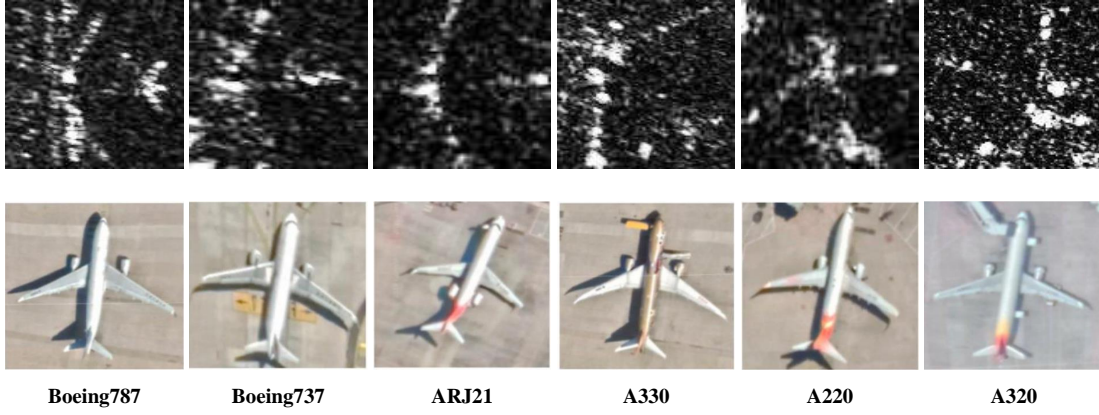


Fig. 10: Civil aircraft in SAR-ACD and the corresponding optical images.

into the test set and training set based on the ratio of 2:8. The model adopts the 20-way 1-shot meta-training module. To simplify, the two groups of experiments are named **Experiment I** and **Experiment II**, respectively.

A. Data Set Description

1) **Data Sources**: Similar to GF3-ADD in [37] and FUSAR-Ship in [69] where Gaofen-3 C-band images are utilized for SAR ship recognition, an experimental data set named SAR-ACD is constructed to illustrate the effectiveness of the proposed network. A total of 11 Gaofen-3 C-band images with 1m resolution working in the Spotlight mode are used in this experiment. These Gaofen-3 satellite images cover the Shanghai Hongqiao Airport area, Beijing Capital International Airport area, and another airport in different time phases with HH polarization. The data set contains 6 civil aircraft categories and 14 other aircraft categories. A total of 4322 aircraft clips are contained in SAR-ACD. It should be emphasized that SAR-ACD contains complex scenes at different airports. Meanwhile, the targets in SAR-ACD are rich in categories and have large size differences. These factors leads to a very challenging classification task.

2) **Category Labelling**: This SAR data is manually labeled according to the corresponding optical images. Generally, the aircraft stand is fixed. The ground truth of aircraft can be considered reliable and accurate by combining expert interpretation and large manual annotation.

After collecting the initial SAR images, normalization and radiation calibration are implemented to correct the images.

Civil SAR aircraft and the corresponding optical images are shown in Fig. 10. The specific categories of civil aircraft and the target number of each category are shown in Fig. 11. Several data expansion methods such as contrast transformation, brightness transformation, and sharpness transformation are adopted to address the issue of the imbalanced number of instances. Due to the imaging result of SAR target varies greatly as the TAA

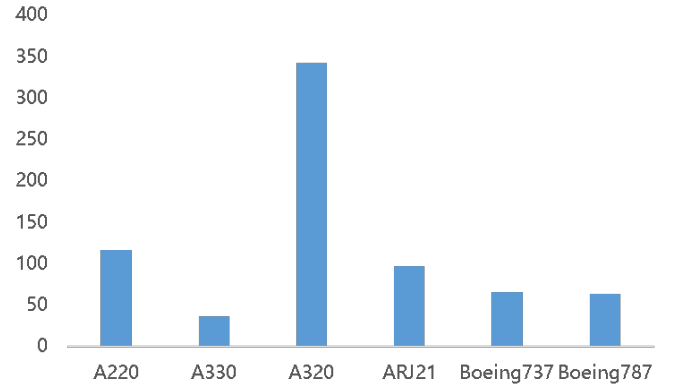


Fig. 11: The original number of each type of civil aircraft cropped from the original large-scale images without data expansion. The number of each category of aircraft after expansion is around 400.

TABLE I: The ablation study of FEM and SEM (Experiment I).

SEM	FEM	5-way 1-shot(%)	5-way 5-shot(%)
×	×	53.18 ± 0.62	68.66 ± 0.51
×	✓	53.54 ± 0.63	69.9 ± 0.50
✓	×	53.57 ± 0.62	70.04 ± 0.52
✓	✓	55.9 ± 0.64	71.04 ± 0.52

TABLE II: The classification accuracy combined with pre-training on the 5-way 1-shot condition (Experiment I).

SEM	FEM	No pre-training(%)	Pre-training(%)
×	×	53.18 ± 0.62	55.21 ± 0.63
×	✓	53.54 ± 0.63	55.28 ± 0.64
✓	×	53.57 ± 0.62	55.55 ± 0.64
✓	✓	55.9 ± 0.64	56.05 ± 0.65

TABLE III: The classification accuracy combined with pre-training on the 5-way 5-shot condition (**Experiment I**).

SEM	FEM	No pre-training(%)	Pre-training(%)
×	×	68.66 ± 0.51	69.04 ± 0.52
×	✓	69.9 ± 0.50	71.41 ± 0.50
✓	×	70.04 ± 0.52	70.50 ± 0.50
✓	✓	71.04 ± 0.52	71.84 ± 0.50

changes, the operations like rotation and reversal are not used.

B. Implementation Detail

1) *Data Setting details*: In **Experiment I**, the ablation studies on the SEM and FEM before and after pre-training are performed, respectively. The experimental results also prove the significance of the pre-training phase. The pre-trained model is obtained by the 14-class classification experiment on the meta-training dataset, which can fully utilize the prior information of the existing samples.

In addition, the meta-training stage is trained with the supervised learning. We The Adam optimizer is used for training where the learning rate is set to 0.001. The model is trained for 10000 epochs, and if the verification results do not improve after 200 epochs, the training process will be terminated. Besides, 100 train episodes and 100 test episodes are carried out per epoch. In the meta-test stage, the best model obtained in the meta-training stage is tested on 1000 randomly selected N-way K-shot tasks. There are two types of tasks: 5-way 5-shot and 5-way 1-shot tasks. As Table V shows, the number of query examples per class remains 15 for both of them.

In **Experiment II**, we change the proportion of training images from 10% to 100% to fully verify the effectiveness of the meta-learning way. Similarly, the training mode and the optimizer setting are the same as **Experiment I**. The N-way K-shot tasks are sampled from the training set of all 20 categories. In **Experiment II**, the meta-training and meta-test ways need special attention. Unlike **Experiment I** where the sampled tasks are 5-shot tasks or 1-shot tasks with 5 classes, the basic task unit in **Experiment II** is 20-way 5-shot and 20-way 1-shot. To ensure a fair comparison with classical fully connected networks (FCN), the number of query images per class for meta-training and meta-testing is 15 and 2, respectively.

2) *Network Structure Details*: The construction of the FEM is exhibited in Fig. 8. FEM consists of 3 convolutional blocks. For each convolutional block, it consists of a convolutional layer with 64 filters and 3×3 convolution, batch normalization (BN), ReLU nonlinearity and 2×2 max-pooling. The single-channel input images of different sizes are first resized to $1 \times 100 \times 100$ uniformly, and the output size of the three convolution blocks of the

feature extraction module are $64 \times 25 \times 25$, $64 \times 12 \times 12$ and $64 \times 6 \times 6$, respectively. The output of the feature extraction module is then flattened and sent into the AG-classifier 6. For the SEM branch, the input size is $64 \times 25 \times 25$, and the stride of the transposed convolution layer is set to 2. It is noticeable that the size of the SEM output is $1 \times 50 \times 50$, while the keypoints are extracted on the initial input images with varying sizes. For the value in the keypoint feature map, 1 denotes the corresponding position is a keypoint, and 0 denotes the corresponding position is non-critical components or background.

The details of the meta-training and meta-test process are shown in Algorithm 1 and Algorithm 2, respectively.

Algorithm 1: SCAN for Meta-Training

```

1 Input: Meta-training data set
    $D_{\text{train}} = \{(x_1, y_1), \dots, (x_{N_{\text{tr}}}, y_{N_{\text{tr}}})\}$ , where each
    $y_i \in \{1, \dots, C_{\text{tr}}\}$ .  $D_k$  contains all samples belonging
   to class  $k$  in  $D_{\text{train}}$ .
2 for each episode do
3    $\text{RANDOMSAMPLE}(\{1, \dots, C_{\text{tr}}\}, N)$ 
4   for  $k$  in  $\{1, \dots, N\}$  do
5      $\text{RANDOMSAMPLE}(D_k, K) \rightarrow S_k$ 
6      $\text{RANDOMSAMPLE}(D_k, Q) \rightarrow Q_k$ 
7     Compute center representation vector  $c_k$ 
       following Equation 4.
8   end
9   for  $k$  in  $\{1, \dots, N\}$  do
10    for  $(x_q, y_q), (x_s, y_s)$  in  $Q_k, S_k$  do
11      Compute the TAAs of  $x_q$  and  $x_s$ 
        according to Equations. 6-9.
12    end
13    Compute  $\text{Loss}_{\text{angle}}$  according to Equation 11.
14    Compute  $\text{Loss}_{\text{pre}}$  according to Equation 3.
15    Compute total loss  $J$  following Equation 13.
16  end
17 end

```

C. Experiment I on SEM and FEM

This section performs the ablation study with and without pre-training on the two modules to verify their effectiveness. The data setup for the meta-training and meta-test stage is shown in Table V.

1) *Ablation Study*: As is shown in Table I, the model shows better performance with the help of the two modules. When only with FEM, the test accuracy increased by 0.4% (1-shot) and 1.3% (5-shot), respectively. When only with SEM, the test accuracy is increased by 0.4% (1-shot) and 1.4% (5-shot), respectively. For 1-shot, the classification accuracy even increased by 2.8% with both modules. The ablation results fully verify the effectiveness of the two modules.

Besides, all channels of the last convolution layer are visualized to understand what the network has learned with SEM intuitively. As shown in Fig. 12, there are 64

Algorithm 2: SCAN for Meta-Testing

```

1 ***** Meta test *****
2 for each episode do
3   RANDOMSAMPLE( $\{1, \dots, C_{te}\}, N$ )
4   for  $k$  in  $\{1, \dots, N\}$  do
5     RANDOMSAMPLE( $D_k, K$ )  $\rightarrow S_k$ 
6     RANDOMSAMPLE( $D_k, Q$ )  $\rightarrow Q_k$ 
7     Compute center representation vector  $c_k$ 
       following Equation 4.
8   end
9   for  $(x_q, y_q)$  in  $Q_k$  do
10    for  $k$  in  $\{1, \dots, N\}$  do
11      Compute Distance( $x_q, c_k$ )
12    end
13    Result:  $x_q \in$ 
      CLASS  $\arg \min_k (\text{Distance}(x_q, c_k))$ 
14  end
15 end

```

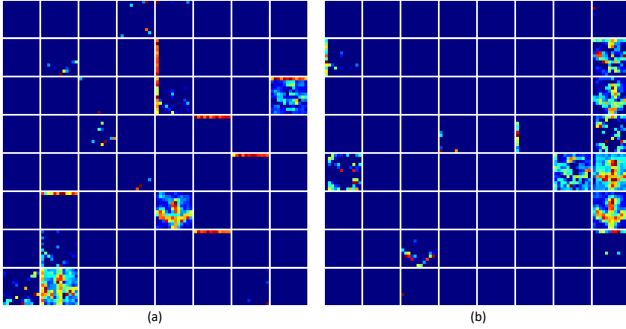


Fig. 12: Visualization of all channels in the last convolution layer (**Experiment I**). (a) Baseline network. (b) Network with SEM.

channels in the last convolutional layer. The network focuses more on the critical parts of the aircraft and less on the edge and background noise.

2) Interpretability analysis for application on SAR data:

As shown in Fig. 13, the two shared convolution layers of the network with SEM and FEM are visualized. The top row is A220 and the bottom row is Boeing787. After adding SEM and FEM, the network pays more attention to the key parts such as the nose, wings and tail. Meanwhile, the feature maps of different aircraft based on different models are also visualized. As shown in Fig. 14, (b) and (c) represent the visualization of the first convolutional layer of the baseline model and the model with SEM and FEM, respectively. (d) and (e) represent the visualization of the second convolutional layer of the baseline model and the model with SEM and FEM, respectively. Red represents places with high network attention. From the visualization of the two shared convolutional layers, it can be seen that the networks attention to the background decreases, and the attention to the aircraft tail increases.

3) *Combination with the Pre-training Phase:* Beyond that, a comparison experiment with or without the pre-training stage is also conducted to prove the availability of the pre-training stage. The comparison of the test results is shown in Table II and Table III. For both the 5-shot and 1-shot case, the test accuracy is improved to a certain extent. The experimental result also reflects that pre-training fully excavates the prior information of the existing data.

TABLE IV: Comparative experiment on AG-classifier with the combination of pre-training (**Experiment I**).

FEM	SEM	No ASC(%)	With ASC(%)
×	×	55.21 ± 0.64	57.01 ± 0.64
✓	×	55.28 ± 0.63	57.82 ± 0.63
×	✓	55.55 ± 0.64	57.85 ± 0.62
✓	✓	56.05 ± 0.65	57.92 ± 0.60

D. Experiment I on ASC

The effectiveness of ASC is further illustrated by contrasting the results with or without this module. What needs illustration is that the weight of classification loss per positive sample pair is determined by the angle difference. While in the 5-way 5-shot condition, the center representation vector is obtained by averaging five support samples, and the average angle value is meaningless. As a consequence, only the comparative experiment on the 5-way 1-shot condition is conducted. The model with AG-classifier yields good classification performance as shown in Table IV.

TABLE V: Details of data setup during training and testing (**Experiment I**).

Setup	5-way 5-shot		5-way 1-shot	
	support set	query set	support set	query set
meta-training	5	15	1	15
meta-testing	5	15	1	15

TABLE VI: Detailed settings for meta-training and meta-testing in the 20-class classification evaluation (**Experiment II**).

Setting	Support Set	Query Set
meta-training	20*1=20	20*15=300
meta-testing	20*1=20	20*2=40

E. Comprehensive performance comparison analysis of **Experiment I**

The confusion matrixes for six new test categories are visualized in Fig. 15. It indicates that the classification accuracy of "Type1" and "Type2" is relatively high. For

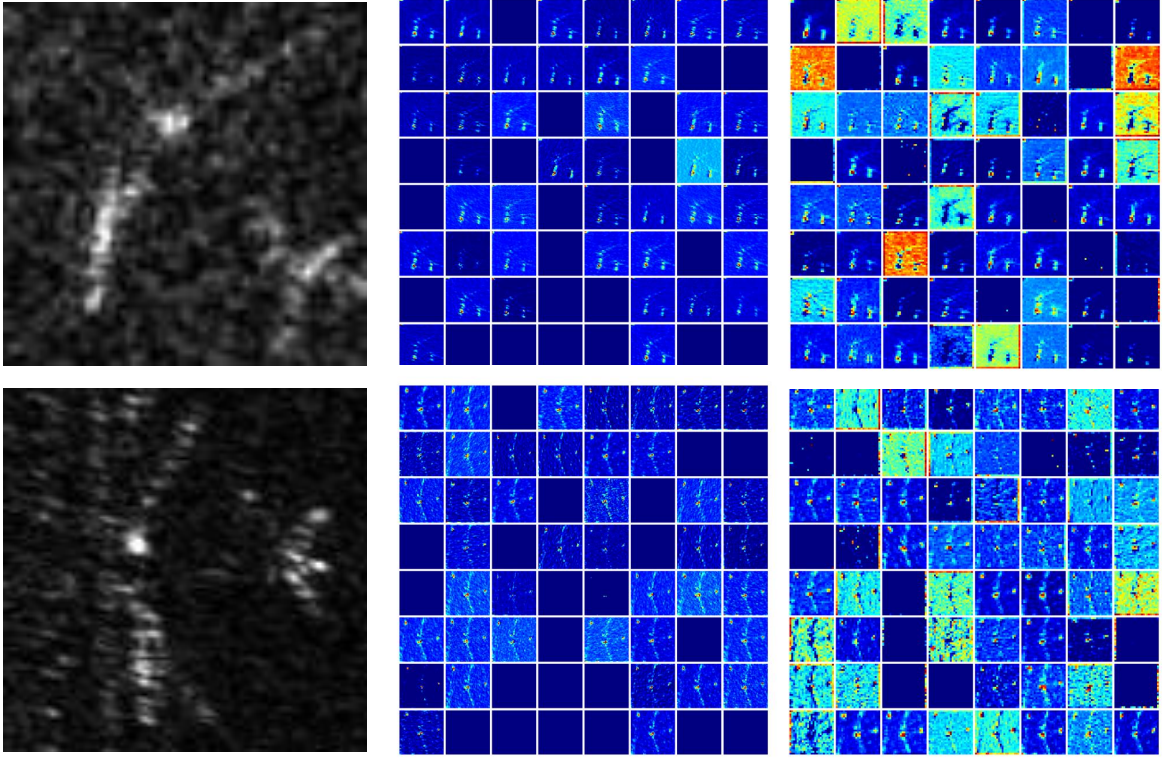


Fig. 13: Visualization of all the channels in the two shared convolution layers of the model with SEM and FEM (**Experiment I**). The first row is Boeing787 and the second row is A220.

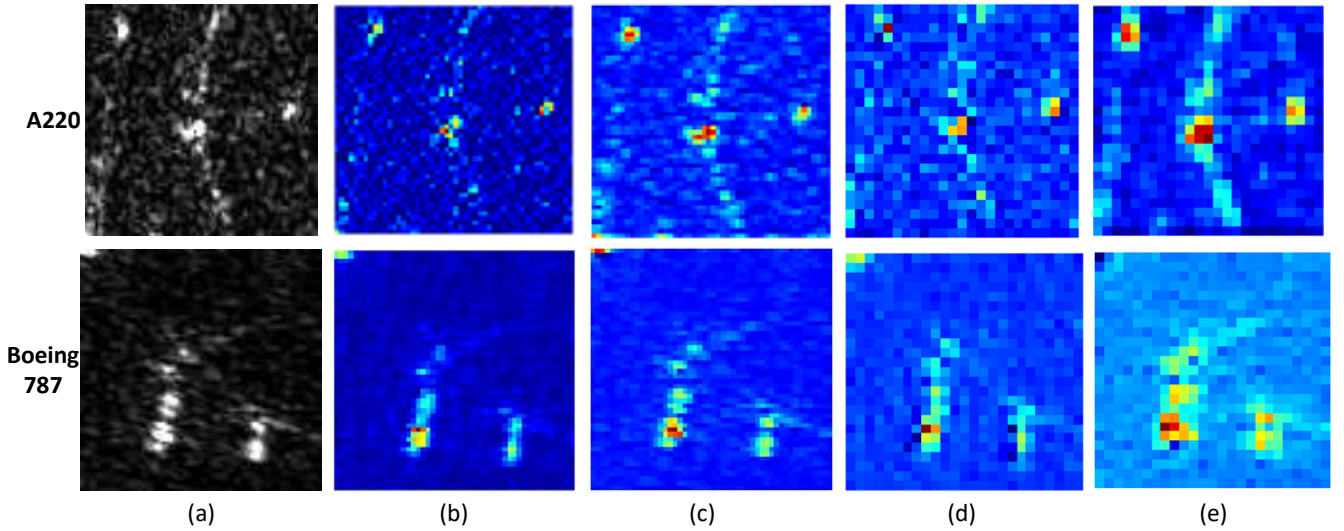


Fig. 14: Feature map visualization comparison of the baseline model and the model with SEM and FEM (**Experiment I**). (a) illustrates the original input images. (b) illustrates the visualization of the first convolutional layer of the baseline model. (c) illustrates the visualization of the first convolutional layer of the model with SEM and FEM. (d) illustrates the visualization of the second convolutional layer of the baseline model. (e) illustrates the visualization of the second convolutional layer of the model with SEM and FEM.

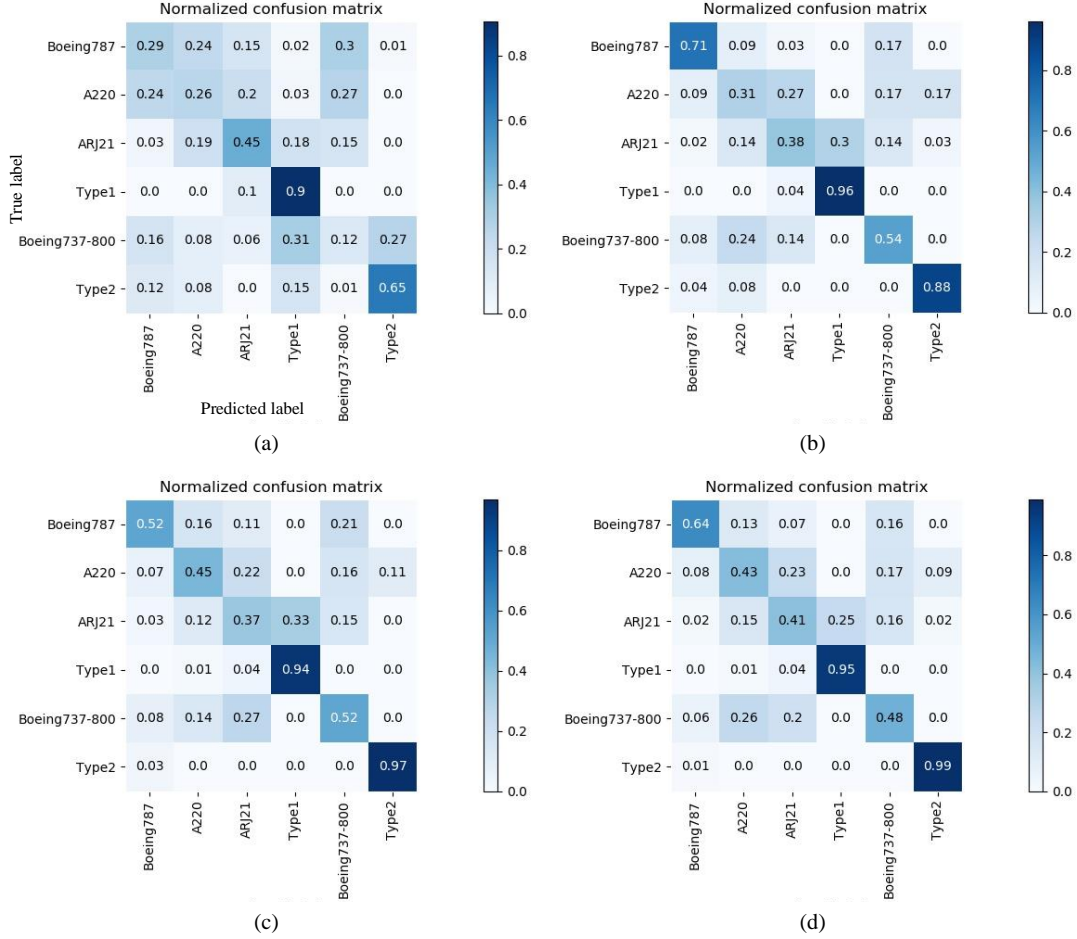


Fig. 15: Visualization of the classification results by confusion matrixes. The matrixes show the test results of the meta-testing data set with different modules for 5-shot (**Experiment I**). (a). The confusion matrix of baseline network. (b). The confusion matrix with FEM. (c). The confusion matrix with SEM. (d). The confusion matrix with both FEM and SEM.

the rest of civil aircraft, it's relatively difficult to classify accurately due to the imaging interference existing in civil airports. Finally, the classification accuracy of civil aircraft increased by at least 20%, which demonstrates the advancement of all modules. The classification accuracy curves with different modules are displayed in Fig. 16. The classification results of several test patches in this experiment are shown in Table VII. SCAN can classify the aircraft patches with different TAAs more accurately. We also test our model on large-scale SAR images. Parts of the Hongqiao Airport and Capital Airport are selected to do the slice classification experiment, and several typical sliding windows are highlighted in Fig. 17. The classification result verifies the stability and effectiveness of the proposed model.

F. 20-class Classification Evaluation of Experiment II

In this section, meta-learning is used for classification among all 20 categories. The percentage of meta-test images is fixed at 20% for each category. Differen-

t proportions will be selected for the remaining 80% images to form the meta-training set. To validate the effectiveness of the meta-learning rule, we contrast the 20-class classification results using meta-learning with those classical fully connected classification networks. In the meta-test phase, the 20-way 1-shot task with 2 query examples is taken as the basic task unit where the total number of query examples is $20 * 2 = 40$. To ensure that the experiment is as fair as possible, the batch size of the fully connected layer network equals the number of meta-testing query examples 40. The detailed settings are shown in Table VI.

1) *Ablation study on ASC*: The effect of ASC on experiment performance is specifically studied. The classification accuracy in the 1-shot condition is shown in Table VIII. The accuracy of the meta-learning classification with ASC is far more than that of a classic classification network with fully connected layers. It can achieve 90% accuracy with only 50% data. Besides, the ablation study of FEM and SEM with different proportions in

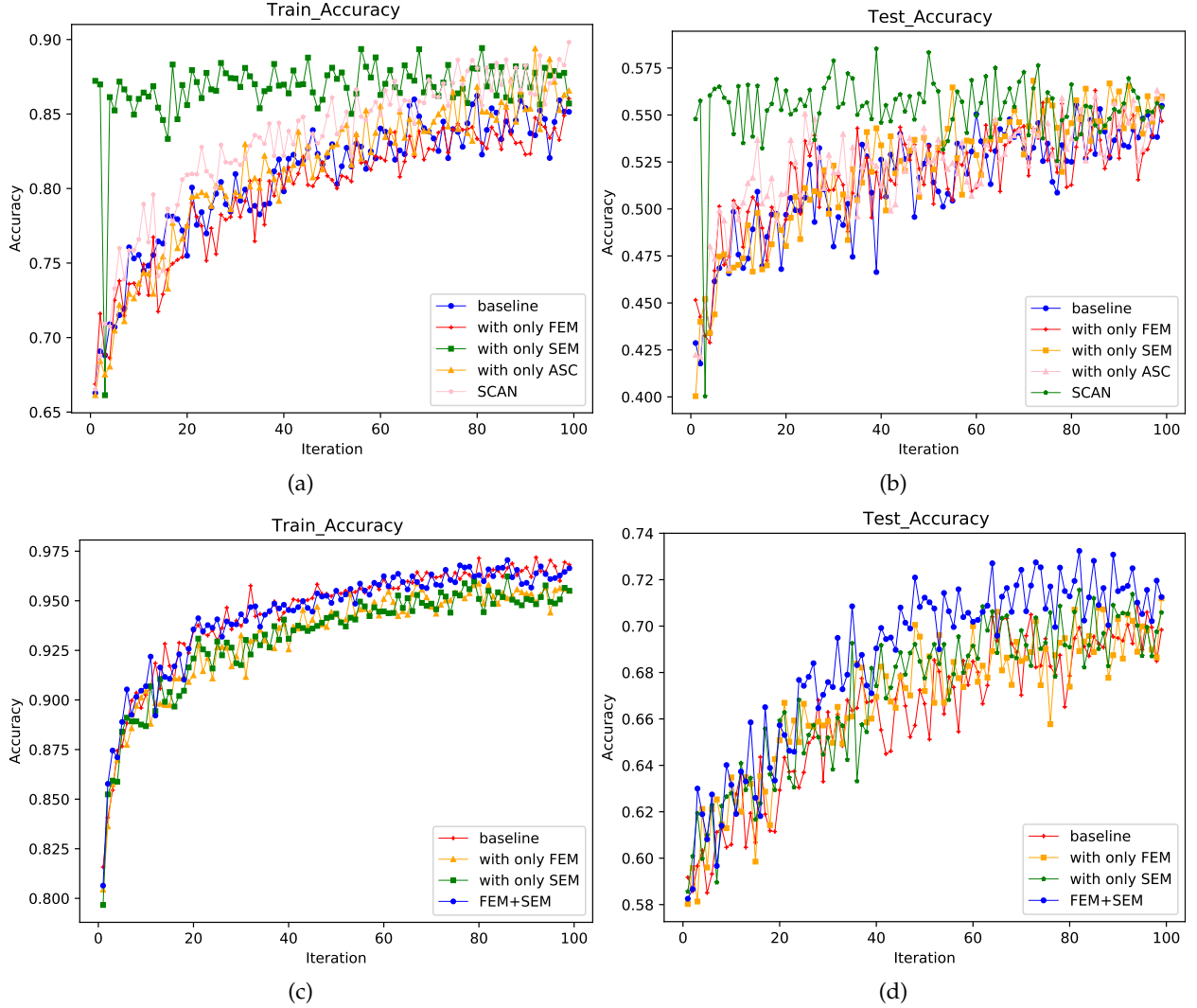


Fig. 16: The classification accuracy with different modules during training and testing (**Experiment I**). (a). Training accuracy in the 5-way 1-shot condition. (b). Testing accuracy in the 5-way 1-shot condition. (c). Training accuracy in the case of 5-way 5-shot. (d). Testing accuracy in the case of 5-way 5-shot.

TABLE VIII: Classification accuracy with different proportions of the training set on the 20-way 1-shot condition (**Experiment II**).

Percent	FCN(%)	baseline(%)	With ASC(%)
10%	32.09	69.18 ± 0.26	75.31 ± 0.26
30%	42.98	81.24 ± 0.27	89.32 ± 0.26
50%	56.75	93.34 ± 0.27	94.47 ± 0.27
70%	58.36	95.30 ± 0.27	96.01 ± 0.27
100%	62.33	95.75 ± 0.26	97.22 ± 0.27

Experiment II is also carried out. As is explained in Fig. 18, The test results verify the efficiency of both the modules.

G. Comparison With Other Methods and

As shown in Table IX, SCAN achieves the best accuracy on both 5-way 1-shot and 5-way 5-shot tasks among

the compared methods. Compared with the previous state-of-the-art method [39], SCAN improves by 1.35% and 6.69% on 5-way 1-shot and 5-way 5-shot tasks, respectively. In addition, in order to fully verify the

TABLE IX: Comparison with other few-shot learning models.

Methods	5-way 1-shot(%)	5-way 5-shot(%)
kNN	22.50	36.30
RelationNet	56.05 ± 0.80	66.71 ± 0.70
MAML [70]	52.2 ± 0.62	67.5 ± 0.52
Matching network	54.54 ± 0.63	58.9 ± 0.50
MSAR [39]	56.57 ± 0.62	65.15 ± 0.50
SCAN	57.92 ± 0.60	71.84 ± 0.52

generalization ability of the proposed recognition model, we randomly select different groups of test sets for

TABLE VII: The test results on some aircraft patches of the meta-test set (**Experiment I**).

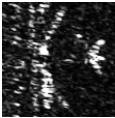
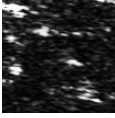


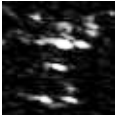

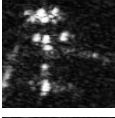
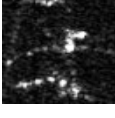
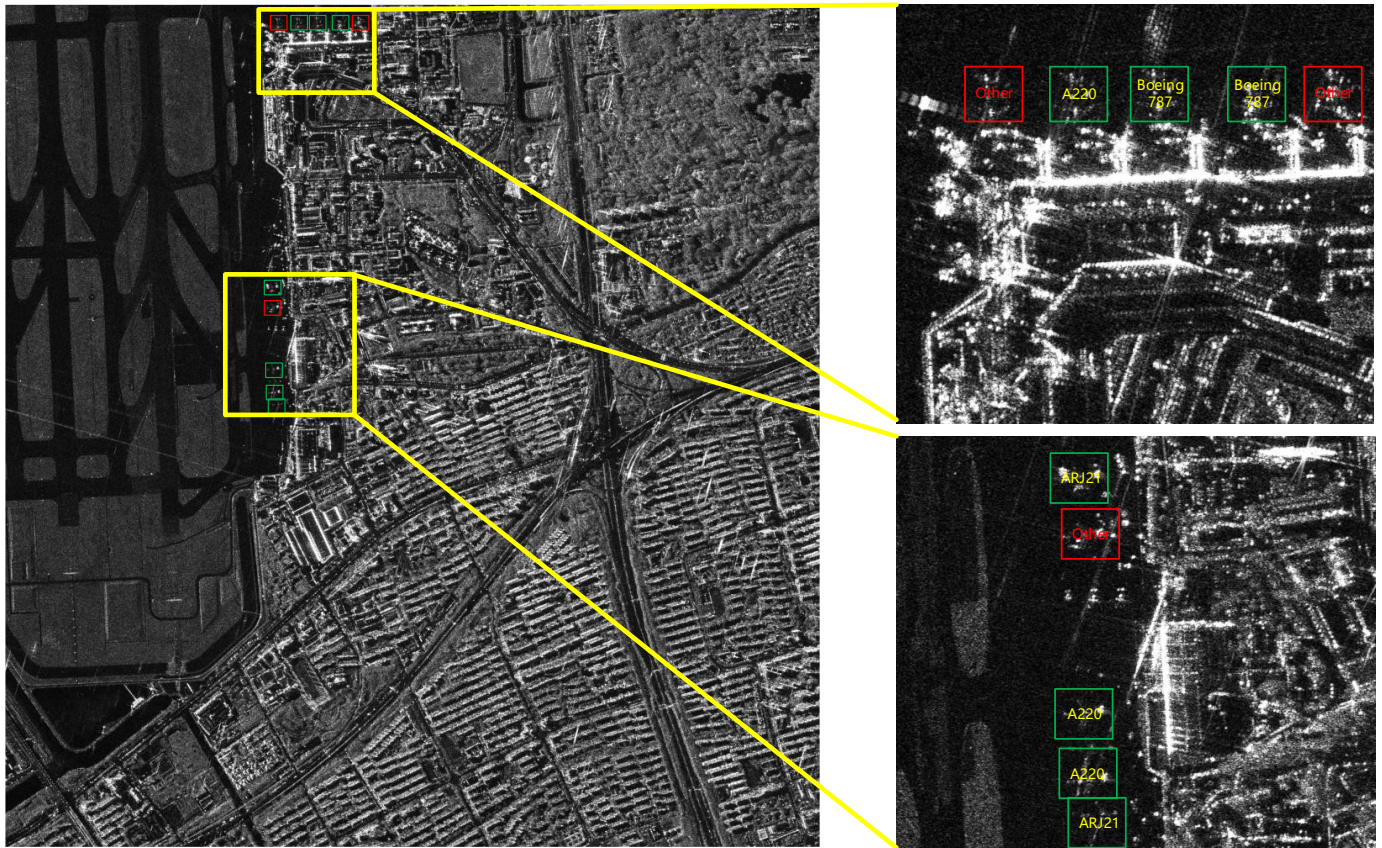
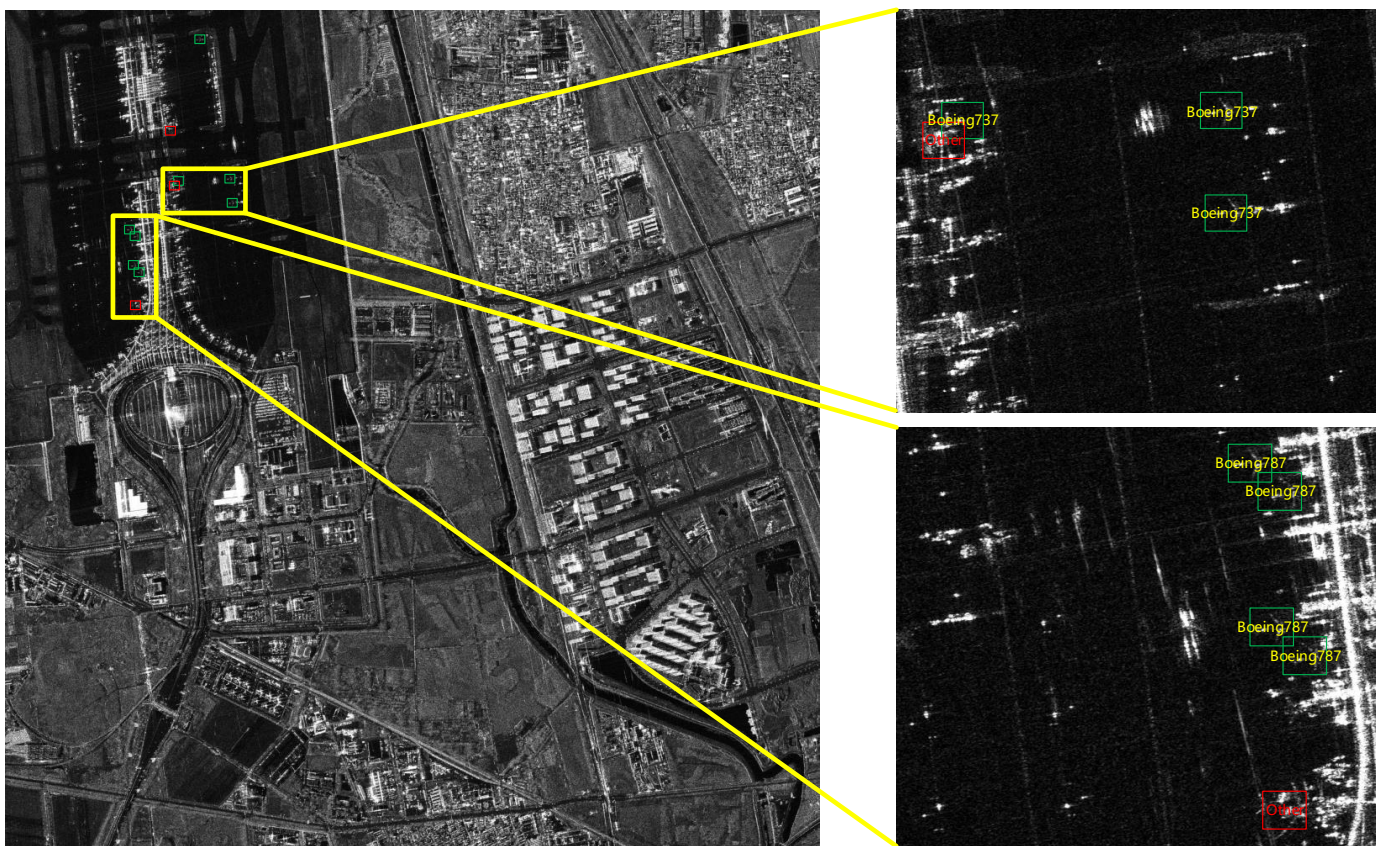
N-way K-shot	TAAAs	Images	Ground Truth	baseline	With only SEM	With only ASC	SCAN
5-way 1-shot	−92°		Boeing787	ARJ21 ×	Boeing787 ✓	Boeing787 ✓	Boeing787 ✓
	−41°		Boeing787	A220 ×	Boeing737 ×	A220 ×	Boeing787 ✓
	−45°		A220	Boeing787 ×	A220 ✓	A220 ✓	A220 ✓
	−122°		A220	Boeing787 ×	A220 ✓	A220 ✓	A220 ✓
5-way 5-shot	5°		ARJ21	A220 ×	Boeing737 ×	Boeing737 ×	A220 ×
	8°		ARJ21	Boeing737 ×	Boeing737 ×	ARJ21 ✓	ARJ21 ✓
	0°		Boeing737	ARJ21 ×	Boeing737 ✓	Boeing737 ✓	Boeing737 ✓
	42°		Boeing737	A220 ×	ARJ21 ×	Boeing737 ✓	Boeing737 ✓

TABLE X: Ablation study on three randomly selected meta-test sets. The fluctuation ranges of the accuracy of the three groups of experiments are similar. The fluctuation ranges of the 5-way 1-shot task and 5-way 5-shot task are 0.0060-0.0064 and 0.0062-0.0065, respectively.

	Group1					Group2					Group3			
	SEM	ASC	FEM	Acc(%)		SEM	ASC	FEM	Acc(%)		SEM	ASC	FEM	Acc(%)
5-way	×	×	×	53.18	5-way	×	×	×	68.75	5-way	×	×	×	61.94
	×	✓	✓	57.82		×	✓	✓	69.84		×	✓	✓	62.86
1-shot	✓	×	×	55.55	1-shot	✓	×	×	69.24	1-shot	✓	×	×	62.47
	✓	✓	✓	57.92		✓	✓	✓	70.21		✓	✓	✓	63.83
5-way	×	-	×	68.66	5-way	×	-	×	80.91	5-way	×	-	×	74.51
	×	-	✓	69.90		×	-	✓	82.84		×	-	✓	76.33
5-shot	✓	-	×	70.04	5-shot	✓	-	×	82.65	5-shot	✓	-	×	76.28
	✓	-	✓	71.04		✓	-	✓	83.37		✓	-	✓	77.13



(a) Shanghai Hongqiao International Airport



(b) Beijing Capital International Airport

Fig. 17: Large scene classification results of different airports.

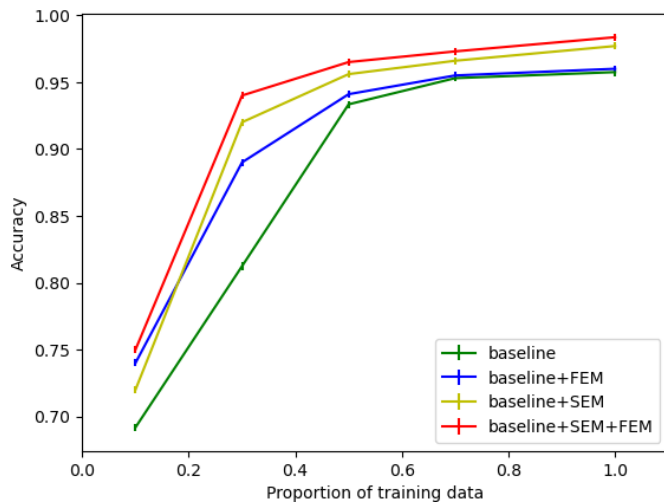


Fig. 18: The ablation study of FEM and SEM with different data proportions on the 20-way 1-shot condition. The error bar for each method fluctuates between 0.0026-0.0027. We can see that the fluctuation of accuracy decreases with the increase of the number of test categories. (Experiment II).

experimental verification. The experimental results are shown in Table X. The performance of the model has been effectively improved on several randomly selected meta-test sets containing new categories.

V. CONCLUSION

In this article, a novel framework named SCAN is proposed to solve the few-shot classification problems in the SAR domain. In contrast with the existing few-shot classification models, SCAN efficiently integrates the scattering characteristics. First, SEM is proposed to combine the target imaging mechanism with the network. Second, ASC is designed to solve the problem caused by the sensitivity of the SAR imaging results to TAAs. Third, we leverage the proposed FEM to exploit frequency domain information. Experiments on SAR-ACD demonstrates the validation of the proposed method. The model performance has increased by 4.8% for 5-way 1-shot after adding these modules. Compared with other methods with similar structure, our method achieves state-of-the-art performance for 5-way 5-shot.

VI. REFERENCES

REFERENCES

- [1] G. Dong and G. Kuang, "Classification on the monogenic scale space: application to target recognition in sar image," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2527–2539, 2015.
- [2] L. M. Novak, G. J. Owirka, and W. S. Brower, "Performance of 10-and 20-target mse classifiers," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 4, pp. 1279–1289, 2000.
- [3] S. A. Wagner, "Sar atr by a combination of convolutional neural network and support vector machines," *IEEE transactions on Aerospace and Electronic Systems*, vol. 52, no. 6, pp. 2861–2872, 2016.
- [4] R. Hummel, "Model-based atr using synthetic aperture radar," pp. 856–861, 2000.
- [5] B. Ding, G. Wen, C. Ma, and X. Yang, "Target recognition in synthetic aperture radar images using binary morphological operations," *Journal of Applied Remote Sensing*, vol. 10, no. 4, pp. 1–14, 2016.
- [6] Q. Zhao and J. C. Principe, "Support vector machines for sar automatic target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 2, pp. 643–654, 2001.
- [7] Y. Sun, Z. Liu, S. Todorovic, and J. Li, "Adaptive boosting for sar automatic target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 1, pp. 112–125, 2007.
- [8] Z. Lin, K. Ji, M. Kang, X. Leng, and H. Zou, "Deep convolutional highway unit network for sar target classification with limited labeled training data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 1091–1095, 2017.
- [9] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target classification using the deep convolutional networks for sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4806–4817, 2016.
- [10] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-resolution sar image classification via deep convolutional autoencoders," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2351–2355, 2015.
- [11] P. Wang, X. Sun, W. Diao, and K. Fu, "Fmssd: Feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3377–3390, 2019.
- [12] X. Sun, Y. Liu, Z. Yan, P. Wang, W. Diao, and K. Fu, "Sraf-net: Shape robust anchor-free network for garbage dumps in remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.
- [13] X. Sun, P. Wang, C. Wang, Y. Liu, and K. Fu, "Pbnet: Part-based convolutional neural network for complex composite object detection in remote sensing imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 50–65, 2021.
- [14] X. Sun, A. Shi, H. Huang, and H. Mayer, "bas⁴net: Boundary-aware semi-supervised semantic segmentation network for very high resolution remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5398–5413, 2020.
- [15] Q. He, X. Sun, Z. Yan, and K. Fu, "Dabnet: Deformable contextual and boundary-weighted network for cloud detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.
- [16] Y. Yang, Z. Zhang, W. Mao, Y. Li, and C. Lv, "Radar target recognition based on few-shot learning," *Multimedia Systems*, no. 3, 2021.
- [17] A. Kusk, A. Abulaitijiang, and J. Dall, "Synthetic sar image generation using sensor, terrain and target models," in *Proceedings of EUSAR 2016: 11th European Conference on Synthetic Aperture Radar*. VDE, 2016, pp. 1–5.
- [18] Q. Song, F. Xu, X. X. Zhu, and Y.-Q. Jin, "Learning to generate sar images with adversarial autoencoder," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.
- [19] Y. Ma, Y. Liang, W. Zhang, and S. Yan, "Sar target recognition based on transfer learning and data augmentation with lsgans," in *2019 Chinese Automation Congress (CAC)*. IEEE, 2019, pp. 2334–2337.
- [20] K. Wang, G. Zhang, and H. Leung, "Sar target recognition based on cross-domain and cross-task transfer learning," *IEEE Access*, vol. 7, pp. 153 391–153 399, 2019.
- [21] W. Zhang, Y. Zhu, and Q. Fu, "Semi-supervised deep transfer learning-based on adversarial feature learning for label limited sar target recognition," *IEEE Access*, vol. 7, pp. 152 412–152 420, 2019.
- [22] C. Lu and W. Li, "Ship classification in high-resolution sar images via transfer learning with small training dataset," *Sensors*, vol. 19, no. 1, p. 63, 2019.
- [23] C. Kang and C. He, "Sar image classification based on the multi-layer network and transfer learning of mid-level representations," pp. 1146–1149, 2016.
- [24] D. Marmanis, W. Yao, F. Adam, M. Datcu, P. Reinartz, K. Schindler, J. D. Wegner, and U. Stilla, "Artificial generation of big data for improving image classification: A generative adversarial network approach on sar data," *arXiv preprint arXiv:1711.02010*, 2017.
- [25] Z. Huang, Z. Pan, and B. Lei, "Transfer learning with deep convolutional neural network for sar target classification with limited labeled data," *Remote Sensing*, vol. 9, no. 9, p. 907, 2017.

- [26] D. Malmgren-Hansen, A. Kusk, J. Dall, A. A. Nielsen, R. Engholm, and H. Skriver, "Improving sar automatic target recognition models with transfer learning from simulated data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1484–1488, 2017.
- [27] D. Zhang, J. Liu, W. Heng, K. Ren, and J. Song, "Transfer learning with convolutional neural networks for sar ship recognition," in *IOP Conference Series: Materials Science and Engineering*, vol. 322, no. 7. IOP Publishing, 2018, p. 072001.
- [28] W. Zelong, X. Xianghui, and Z. Lei, "Study of deep transfer learning for sar atr based on simulated sar images," *Journal of University of Chinese Academy of Sciences*, vol. 37, no. 4, p. 516, 2020.
- [29] L. Wang, X. Bai, and F. Zhou, "Few-shot sar atr based on conv-bilstm prototypical networks," in *2019 6th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*. IEEE, 2019, pp. 1–5.
- [30] L. Wang, X. Bai, C. Gong, and F. Zhou, "Hybrid inference network for few-shot sar automatic target recognition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 11, pp. 9257–9269, 2021.
- [31] 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, 2021.
- [32] M. Yang, X. Bai, L. Wang, and F. Zhou, "Mixed loss graph attention network for few-shot sar target classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.
- [33] 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, pp. 1–4.
- [34] D. Liu, X. Gao, and Q. Shen, "Prototypical network for radar image recognition with few samples," *Journal of Physics: Conference Series*, vol. 1634, no. 1, p. 012116 (10pp), 2020.
- [35] Y. Li, X. Li, Q. Sun, and Q. Dong, "Sar image classification using cnn embeddings and metric learning," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2020.
- [36] K. Wang, G. Zhang, Y. Xu, and H. Leung, "Sar target recognition based on probabilistic meta-learning," *IEEE Geoscience and Remote Sensing Letters*, 2020.
- [37] Q. Guo, H. Wang, and F. Xu, "Scattering enhanced attention pyramid network for aircraft detection in sar images," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–18, 2020.
- [38] K. Fu, J. Fu, Z. Wang, and X. Sun, "Scattering keypoints guided network for oriented ship detection in high-resolution and large-scale sar images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021.
- [39] K. Fu, T. Zhang, Y. Zhang, Z. Wang, and X. Sun, "Few-shot sar target classification via metalearning," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.
- [40] C. Cao, Z. Cui, Z. Cao, L. Wang, and J. Yang, "An integrated counterfactual sample generation and filtering approach for sar automatic target recognition with a small sample set," *Remote Sensing*, vol. 13, no. 19, p. 3864, 2021.
- [41] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
- [42] J. Guo, B. Lei, C. Ding, and Y. Zhang, "Synthetic aperture radar image synthesis by using generative adversarial nets," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 1111–1115, 2017.
- [43] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
- [44] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," *arXiv preprint arXiv:1704.00028*, 2017.
- [45] S. J. Auer, "3d synthetic aperture radar simulation for interpreting complex urban reflection scenarios," Ph.D. dissertation, Technische Universität München, 2011.
- [46] Z. Cui, M. Zhang, Z. Cao, and C. Cao, "Image data augmentation for sar sensor via generative adversarial nets," *IEEE Access*, vol. 7, pp. 42 255–42 268, 2019.
- [47] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for sar target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 364–368, 2016.
- [48] Z. Wang, X. Fu, and K. Xia, "Target classification for single-channel sar images based on transfer learning with subaperture decomposition," *IEEE Geoscience and Remote Sensing Letters*, 2020.
- [49] Y. Wang, C. Wang, and H. Zhang, "Ship classification in high-resolution sar images using deep learning of small datasets," *Sensors*, vol. 18, no. 9, p. 2929, 2018.
- [50] Z. Ying, C. Xuan, Y. Zhai, B. Sun, J. Li, W. Deng, C. Mai, F. Wang, R. D. Labati, V. Piuri *et al.*, "Tai-sarnet: Deep transferred atrous-inception cnn for small samples sar atr," *Sensors*, vol. 20, no. 6, p. 1724, 2020.
- [51] M. Rostami, S. Kolouri, E. Eaton, and K. Kim, "Sar image classification using few-shot cross-domain transfer learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.
- [52] 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.
- [53] Z. Huang, Z. Pan, and B. Lei, "What, where, and how to transfer in sar target recognition based on deep cnns," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2324–2336, 2019.
- [54] C. Zhong, X. Mu, X. He, J. Wang, and M. Zhu, "Sar target image classification based on transfer learning and model compression," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 3, pp. 412–416, 2018.
- [55] Y. Sun, Y. Wang, H. Liu, N. Wang, and J. Wang, "Sar target recognition with limited training data based on angular rotation generative network," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 11, pp. 1928–1932, 2019.
- [56] 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*. IEEE, 2019, pp. 1342–1345.
- [57] S. Balakrishnama and A. Ganapathiraju, "Linear discriminant analysis-a brief tutorial," *Institute for Signal and Information Processing*, vol. 18, no. 1998, pp. 1–8, 1998.
- [58] T. Mensink, J. Verbeek, F. Perronnin, and G. Csurka, "Metric learning for large scale image classification: Generalizing to new classes at near-zero cost," in *European Conference on Computer Vision*. Springer, 2012, pp. 488–501.
- [59] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
- [60] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [61] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [62] K. Wang and G. Zhang, "Sar target recognition via meta-learning and amortized variational inference," *Sensors*, vol. 20, no. 20, p. 5966, 2020.
- [63] 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, no. 13, p. 4333, 2021.
- [64] Y. Li, X. Li, Q. Sun, and Q. Dong, "Sar image classification using cnn embeddings and metric learning," *IEEE Geoscience and Remote Sensing Letters*, 2020.
- [65] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," *arXiv preprint arXiv:1703.05175*, 2017.
- [66] Z. Qin, P. Zhang, F. Wu, and X. Li, "Fcanet: Frequency channel attention networks," *arXiv preprint arXiv:2012.11879*, 2020.
- [67] Y. Yu, B. Wang, and L. Zhang, "Hebbian-based neural networks for bottom-up visual attention and its applications to ship detection in sar images," *Neurocomputing*, vol. 74, no. 11, pp. 2008–2017, 2011.
- [68] J. Zhao, "Research on sar image object/land fine interpretation methods," Ph.D. dissertation, Shanghai Jiao Tong University, 2020.
- [69] 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," *Science China Information Sciences*, vol. 63, no. 4, pp. 1–19, 2020.
- [70] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1126–1135.



Xian Sun (Senior Member, IEEE) received the B.Sc. degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2004, and the M.Sc. and Ph.D. degrees from the Institute of Electronics, Chinese Academy of Sciences (CAS), Beijing, in 2009.

He was a Visiting Scholar, Karlsruher Institut für Technologie, Karlsruhe, Germany, in 2013. He is currently a Professor with the Aerospace Information Research Institute, CAS. His research interests include computer vision, geospatial

data mining, and remote sensing image understanding.

Dr. Sun was a recipient of the Outstanding Science and Technology Achievement Prize of the CAS in 2016 and the First Prize for The State Scientific and Technological Progress of China in 2019. He also serves as an Associate Editor for IEEE ACCESS and a Guest Editor for the special issue of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS) and other journals.



Yixuan Lv received the B.Sc. degree from Xidian University, Xian, China, in 2019. She is pursuing the master's degree with the University of Chinese Academy of Sciences, Beijing, China, and the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

Her research interests include computer vision and deep learning, especially on SAR target classification and object detection.



Zhirui Wang (Member, IEEE) received the B.Sc. degree from the Harbin Institute of Technology, Harbin, China, in 2013, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2018.

He is an Assistant Researcher with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. His research interests include synthetic aperture radar (SAR) terrain classification, and SAR target detection and recognition.



Kun Fu (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees from the National University of Defense Technology, Changsha, China, in 1995, 1999, and 2002, respectively.

He is currently a Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences (CAS), Beijing, China. His research interests include computer vision, remote sensing image understanding, and geospatial data mining and visualization.

Dr. Fu was a recipient of the First Prizes for The State Scientific and Technological Progress of China in 2015 and 2019, the Outstanding Science and Technology Achievement Prize of the CAS in 2016, the Scientific and Technological Innovation Leading Talent by the National High-Level Talents Special Support Plan in 2017, and the Distinguished Young Scholars from the National Natural Science Foundation of China in 2017.