

# A Coarse-to-Fine Hierarchical Feature Learning for SAR Automatic Target Recognition With Limited Data

Yan Wen , Xihao Wang , Lihe Peng , and Yu Qiao

## I. INTRODUCTION

**Abstract**—With the rapid advancements in deep learning, Synthetic Aperture Radar (SAR) Automatic Target Recognition (ATR) has seen significant improvements in performance. However, the effectiveness of even the most advanced deep-learning-based ATR methods is limited by the scarcity of training samples. This challenge has sparked growing interest in SAR ATR under data-constrained conditions in recent years. Most current approaches for SAR ATR with limited data enhance recognition through data augmentation, specialized modules, or contrastive learning-based loss functions. However, effectively utilizing limited supervision signals to identify key features remains a significant challenge that existing methods have not thoroughly addressed. In our research, we introduce a novel coarse-to-fine hierarchical feature learning strategy for SAR ATR with limited data. Starting with a feature extractor that produces multi-level features, we implement a coarse-to-fine gradual feature constraint to optimize each level using limited supervision signals. This approach simplifies parameter search and ensures effective feature utilization from coarse to fine granularity. Additionally, our method enhances the compactness within classes and the separability between classes of features at various levels. This is achieved by capitalizing on the consistency of features across multiple levels, thereby progressively enhancing the features and, in turn, boosting the model's overall performance. To validate our approach, we conducted recognition and comparative experiments on the MSTAR and OpenSARShip datasets. The results demonstrate our method's exceptional performance in limited-sample recognition scenarios. Moreover, ablation studies confirm the robustness of our approach, underscoring its potential in addressing the challenges of SAR ATR with limited data.

**Index Terms**—Automatic target recognition (ATR), coarse-to-fine, hierarchical feature learning, synthetic aperture radar (SAR).

Manuscript received 22 July 2023; revised 20 December 2023; accepted 26 June 2024. Date of publication 4 July 2024; date of current version 8 August 2024. This work was supported in part by the Sichuan Province 2022 Industry-Education Integration Demonstration Project-18 “Intelligent Agricultural Machinery Equipment Manufacturing Technology Application Innovation Service Platform” and in part by the Chengdu Agricultural College Institutional Project “Big Data Storage and Analysis of Facility Vegetables Based on Hadoop” under Grant 23ZR213. (Corresponding author: Yan Wen.)

Yan Wen is with the School of Mechanical, Electrical Information, Chengdu Agricultural College, Chengdu 611130, China (e-mail: cldgdxwenyan@126.com).

Xihao Wang is with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: wangxihao0414@126.com).

Lihe Peng is with the Suzhou Tongyuan Software & Control Technology Company, Suzhou 215123, China (e-mail: lihepeng2jh@foxmail.com).

Yu Qiao is with the College of Computer Science, Chongqing University, Chongqing 400044, China (e-mail: wzmxcg@outlook.com).

Digital Object Identifier 10.1109/JSTARS.2024.3423377

SYNTHETIC aperture radar (SAR), a pivotal microwave remote sensing system, plays an essential role in both military and civilian contexts [1], [2], [3]. Its implementation in automatic target recognition (ATR) is fraught with challenges but stands as a key element in SAR applications. Numerous esteemed researchers have spearheaded diverse deep learning strategies, leading to substantial advancements in SAR ATR applications [4], [5], [6], [7], [8].

Yet, a common limitation in these algorithms is the need for large amounts of labeled samples per target type for training, a condition that's difficult to meet in real-world applications [9], [10], [11], [12], [13]. In certain scenarios, such as earthquake and sea rescue operations, the volume of SAR images can be sparse, causing potential shortcomings in existing SAR ATR methodologies. As a result, attention has shifted towards ATR methods that operate efficiently with limited SAR data [14], [15], [16], [17], [18], [19]. The focus is on building robust classifiers using minimal labeled SAR images.

The existing strategies for handling SAR ATR with limited data are broadly classified into three main categories: data augmentation, metric-based, and model-based methods [20], [21], [22].

Data augmentation strategies play a critical role in overcoming the challenges posed by sparse data in SAR ATR [4], [6], [14], [23], [24]. These techniques focus on enhancing recognition performance by increasing both the number and variety of training SAR images. A notable example is the work of Zheng et al., who introduced a semisupervised SAR ATR method utilizing a multidiscriminator generative adversarial network (GAN). This approach achieved a remarkable recognition rate of 85.23% with a minimal dataset containing only 20 samples per target [15]. Similarly, Gao et al. [25] proposed a semisupervised GAN method with multiple generators, attaining an efficiency of over 92% using approximately 40 samples for each target type. These innovations highlight the effectiveness of data augmentation in significantly enhancing SAR ATR performance, even when dealing with limited data.

Metric-based approaches, in contrast to other methods, focus on learning class representations that can effectively generalize to new, unseen classes [18], [26], [27], [28], [29]. The primary aim of these strategies is to develop more robust classification systems. An example is the work of Wang et al. [30], who

introduced a prototypical network. In this network, the classifier is designed based on the Euclidean distance between training samples and the prototype of each class. This distance-centric approach ensures that the classifier is capable of accurately identifying and categorizing novel samples. In a similar vein, a distinct hard task mining technique has been applied to enhance meta-learning. This technique demonstrated significant improvements, achieving absolute gains of 1.7% and 2.3% in 1-shot and 5-shot settings, respectively [31]. These metric-based methods are proving to be effective in creating versatile and generalizable class representations.

Model-based methods, the third category, harness prior knowledge to construct the embedding space and regulate the complexity of the model [19], [32], [33]. These approaches prioritize incorporating domain knowledge and past experiences into the model design. These model-based methods underscore the value of incorporating prior knowledge and structural information into the modeling process.

Despite these advancements, the key challenge in SAR ATR with limited data is determining how to use rare supervisory information to enable the model to identify effective recognition features and achieve accurate identification. Generally, models undergo an end-to-end training process, but the scarcity of data often prevents the provision of sufficient supervisory information. Consequently, it becomes challenging for the model to find optimal hypotheses in the whole hypotheses space and extract effective recognition features from SAR images. Therefore, achieving precise recognition with limited SAR data in an ATR model is quite difficult [27], [34], [35].

In this article, we introduce a novel coarse-to-fine hierarchical feature learning approach for SAR ATR, specifically designed to work with limited data. Our method effectively utilizes the scarce supervisory signals available from limited samples and progressively refines features at various levels, aligning with the granularity of these features. Additionally, by leveraging the consistency across different granularities of features as supplementary supervisory input, our approach aids the model in navigating towards the optimal hypothesis during training, thereby enhancing the feature effectiveness. This strategy facilitates accurate target identification even with a limited number of SAR samples. More specifically, our approach begins by employing a multistage feature extractor to obtain initial multilevel features from limited SAR images, each representing different levels of granularity. We then compute a multilayer deep supervisory signal, which is derived from the consistency between features of varying granularities. This computation includes two primary components: the consistency loss in probability distribution and the consistency loss in feature distribution. Throughout the training process, this multilayer deep supervisory signal is utilized to progressively refine the features at each level, applying different weights to different granularities. This methodology enables precise recognition capabilities even with a constrained set of SAR training data. The key contributions of our method can be summarized as follows.

- 1) Our framework introduces an innovative approach to handle the limited SAR training data challenge, focusing on extracting highly discriminative features. This framework

is structured around hierarchical feature embedding, implementing a coarse-to-fine gradual constraint mechanism and a dual consistency constraint that targets features of different granularities. This design is meticulously crafted to amplify the effectiveness of feature extraction, even with the limited availability of SAR data.

- 2) The core of our method lies in the coarse-to-fine gradual feature constraint. This strategy prioritizes the extraction of broader, more effective regions in SAR images initially, before delving into the extraction of finer, more discriminative features. Furthermore, we propose a consistency constraint aimed at not only improving the compactness within classes and the separability between classes at different granularity levels but also ensuring uniformity across features of various granularities.
- 3) Our approach demonstrates exceptional performance in the recognition of MSTAR and OpenSARship datasets, even with limited training data. The effectiveness of our method is substantiated through rigorous verification, showcasing its robustness and reliability in handling constrained SAR data scenarios. The detailed outcomes and analyses of these verification processes, highlighting our approach's capabilities, are elaborated in the subsequent sections of this article.

The rest of this article is organized as follows. Section II provides an in-depth presentation of the proposed method. Section III validates the effectiveness of our approach through various experiments. Finally, Section IV concludes this article.

## II. METHOD

In this section, the proposed method is introduced. First, we elucidate the framework of our method. Then, the coarse-to-fine gradual feature constraint is described in detail. The consistency in features of different granularities is also presented.

### A. Framework

The motivation behind the proposed method is to leverage limited supervisory signals to progressively enhance the effectiveness of multilevel features in a coarse-to-fine manner. Furthermore, the consistency across multilevel features serves as additional prior information, further addressing the problem of inadequate feature discriminability in small sample scenarios. Given practical constraints, ATR methods often confront a reduction in recognition performance with limited SAR data.

However, in real-world scenarios, there often exists knowledge that can serve as prior information, which can help augment the originally limited supervisory signals.

This can also enhance the utilization of limited supervisory signals, thereby acquiring more discriminative recognition features even with a small sample size. Hence, by exploring the utilization of supervisory signals and effective prior information, the recognition performance of the ATR method with limited SAR data can be improved and achieve state-of-the-art performance.

As shown in Fig. 1, our method can be divided into three stages. First, a multilevel feature extractor is constructed. SAR

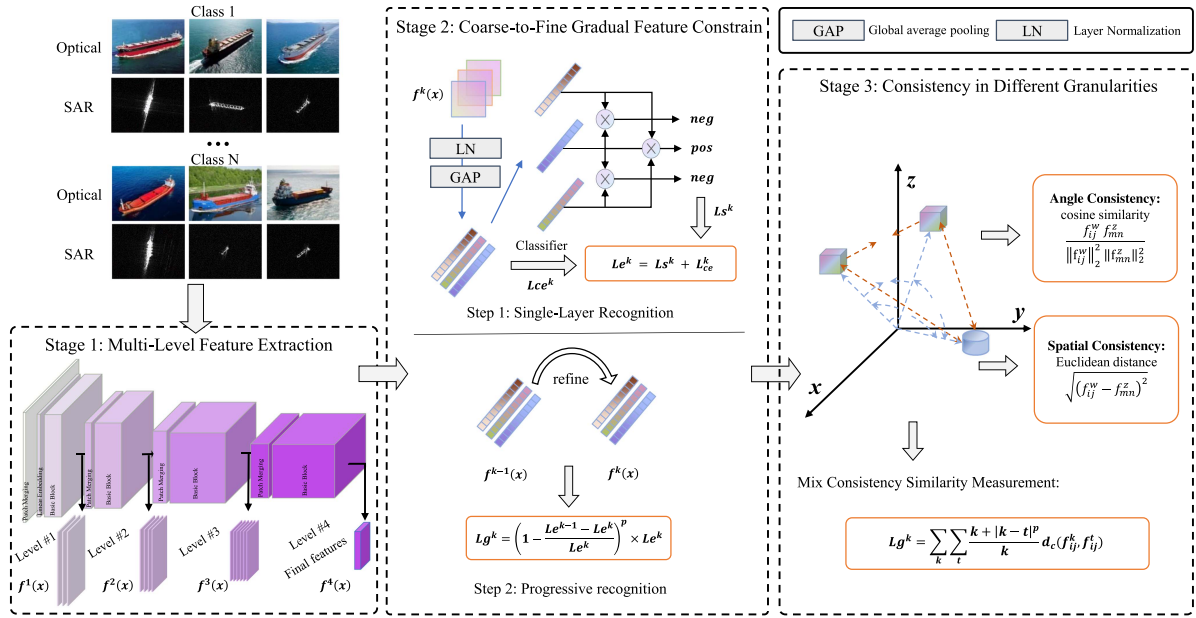


Fig. 1. Whole framework of the proposed method. The proposed method uses the sparse supervisory signals from limited samples to iteratively optimize multilevel features based on their granularity. It employs consistency across different feature granularities as extra guidance, assisting the model in finding the optimal hypothesis during training and improving feature effectiveness. This leads to precise identification, even with limited SAR samples.

images are inputted and corresponding multilevel features are obtained.

Second, the coarse-to-fine gradual feature constraint is proposed to progressively enhance the effectiveness of the multilevel features from coarse to fine in recognition, ensuring that features at both coarse and fine levels have sufficient recognition effectiveness.

Then, the consistency in features across different levels further enhances the consistency of features at different levels, working in conjunction with the progressive enhancement of effectiveness of the coarse-to-fine gradual feature constraint. It not only filters features at multiple levels, but also effectively enhances the separability and compactness within classes of features. Moreover, it limits the search range of the model's hypothesis space, making it easier for the model to find the optimal hypothesis with small sample sizes. The pipeline of our method can be formulated as follows.

Given the training SAR image datasets  $X = \{x_{11}, \dots, x_{ij}, \dots, x_{CN}\}$ ,  $C$  is the total class number, and  $N$  is the number of SAR images for each class.  $x_{ij}$  is the  $j$ th SAR image of the  $i$ th class, and  $y_{ij}$  is the corresponding class label.

First, the feature extractor is constructed to obtain the multilevel features  $F_{ij} = \{f_{ij}^1, \dots, f_{ij}^K\}$  from the input SAR image  $x_{ij}$ . Here,  $f_{ij}^k$  represents the features at the  $k$ th level of  $x_{ij}$ . Then, the coarse-to-fine gradual feature constraint calculates the recognition effectiveness of each level feature on the  $K$  levels of features  $\{L_g^1, \dots, L_g^K\}$ . By further proposing a progressive optimization constraint for  $\{L_g^1, \dots, L_g^K\}$  and calculating the final loss of features at  $K$  levels  $L_{gc}$ , it realizes the optimization of single-layer features using the inadequate supervisory signals provided by limited samples, and progressively enhances the recognition effectiveness of features with the increase of feature levels.

To maximize the efficacy of limited supervisory signals, our method “consistency in features of different levels” calculates the consistency measures for features at  $K$  levels, focusing on both feature distribution and recognition probability distribution. This calculation results in an optimization loss,  $L_c$ . This approach necessitates strong consistency between features at adjacent levels, ensuring that recognition-effective features are progressively enhanced through each layer. This strategy works in tandem with the previously established progressively enhanced recognition effectiveness optimization loss, facilitating better coordination of multilevel features for accurate recognition in scenarios with limited sample sizes.

The total loss of our method is formulated as

$$L_{\text{total}} = L_{ce} + L_{gc} + L_c. \quad (1)$$

In this equation,  $L_{ce}$  represents the basic loss for recognition, encompassing the foundational aspect of our loss function.

Our proposed method systematically augments the recognition effectiveness of multilevel features using the constrained supervisory signals. This is achieved through the implementation of multilevel consistency constraints and progressive effectiveness constraints, which are critical for ensuring accurate recognition in contexts characterized by small sample sizes. In the subsequent sections, we delve into the intricacies of the coarse-to-fine gradual feature constraint and the consistency in features of different levels, providing a comprehensive overview of these integral components of our methodology.

## B. Coarse-to-Fine Gradual Feature Constraint

The coarse-to-fine gradual feature constraint aims to optimize the features of  $K$  levels individually, using only limited supervisory signals, so that the recognition effectiveness of features

at  $K$  levels progressively increases with the enhancement of the level.

Most previous methods used the same limited supervisory signals to optimize the entire model simultaneously, making it difficult to obtain ultimately effective recognition features. This is because of the following. 1) The entire model has a large number of parameters, corresponding to a large hypothesis space, and limited supervisory signals find it hard to help the model search for the optimal hypothesis. 2) The process of feature extraction by the entire model gradually transitions from coarse to fine levels. When limited supervisory signals are used to optimize the entire model simultaneously, the optimization of single-layer features is not controlled. This might lead to a situation where a certain level of feature has high recognition effectiveness, but the next level of feature has low recognition effectiveness. These locally optimal parameters limit the overall performance of the recognition model. Therefore, we propose the coarse-to-fine gradual feature constraint to ensure the recognition effectiveness of features at each level among the  $K$  levels and a progressive effectiveness constraint to help the model gradually learn effective recognition features. This not only improves the way the model searches for optimal hypotheses in the model's hypothesis space, but also enhances the recognition performance of the model.

The coarse-to-fine gradual feature constraint method is structured into two primary steps: 1) measurement of single-layer recognition effectiveness; and 2) progressive constraint of recognition effectiveness across  $K$  levels. The overarching structure of our proposed approach is depicted in Fig. 1. The detailed methodology is as follows.

We start with SAR image datasets  $X = \{x_{11}, \dots, x_{ij}, \dots, x_{CN}\}$  and extract the features  $F_{ij} = \{f_{ij}^1, \dots, f_{ij}^K\}$  of each image  $x_{ij}$ .

*Step 1:* Measurement of single-layer recognition effectiveness. This step focuses on evaluating the effectiveness of single-layer features in two dimensions: the distinction between interclass features and the accuracy of recognition. Initially, cosine similarity is employed to assess the similarity between two sample features

$$d(x_{ij}, x_{mn}) = \frac{f_{ij}^k \cdot f_{mn}^k}{\|f_{ij}^k\|_2 \|f_{mn}^k\|_2}. \quad (2)$$

Here,  $\|\cdot\|_2$  denotes L2 normalization. Subsequently, for a given feature  $x_{ij}$ , we identify the most similar interclass sample feature  $x_{ij}^n$  and the least similar intraclass sample feature. The separation degree of interclass features for  $x_{ij}$  is then calculated

$$Ls_{ij}^k = \max(d(x_{ij}^k, x_{ij}^n) + \theta - d(x_{ij}, x_{ij}^p), 0). \quad (3)$$

In this equation,  $\theta$  is a margin, a hyperparameter that defines the desired level of feature separation. By measuring the cosine similarity between pairs of features, we can effectively determine the most similar interclass and the least similar intraclass features.

Therefore, the measurement of the degree of separation of interclass features at the single layer can be calculated as follows:

$$Ls^k = - \sum_j \sum_i L_{ij}^k. \quad (4)$$

In assessing the accuracy of single-layer feature recognition, we employ the cross-entropy loss as a metric

$$L_{ce}^k = - \sum_i \sum_j y_{ij} \log(p(y_{ij}|f_{ij}^k)). \quad (5)$$

Here,  $p(y_{ij}|f_{ij}^k)$  represents the probability of the feature  $f_{ij}^k$  being correctly classified into its corresponding class label  $y_{ij}$ . To maintain consistency in classification across all granularity levels, we utilize a shared-weight multilayer perceptron (MLP) as the classifier. The overall effectiveness of the recognition using single-layer features is quantified by

$$Le^k = Ls^k + L_{ce}^k. \quad (6)$$

*Step 2:* Progressive recognition effectiveness constraint for  $K$  levels. After calculating the recognition effectiveness of single-layer features, we progressively constrain the features at adjacent, gradually deepening levels. For features at level  $k-1$  and  $k$ , their progressive constraint is calculated as follows:

$$Lg^k = \left(1 - \frac{Le^{k-1} - Le^k}{Le^k}\right)^p \times Le^k \quad (7)$$

where  $p$  is the parameter to adjust the degree of progression. Through the proposed coarse-to-fine gradual feature constraint, we use limited supervisory signals to not only optimize the features at each level, but also progressively constrain them layer by layer, completing the progressive enhancement of feature effectiveness. Next, the consistency of features at different levels is described in detail.

### C. Consistency in Features of Different Granularities

While the coarse-to-fine gradual feature constraint method successfully enhances feature effectiveness incrementally, its performance is limited due to the inadequacy of supervisory signals. This limitation leads to a discrepancy in the features extracted across different layers, resulting in the neglect or loss of effectively extracted features, ultimately impacting the overall performance of the recognition model. To address this, we introduce the ‘‘consistency in features of different levels’’ approach. This method aims to fortify the consistency among features at various levels, diminish the disparity between these features, and thereby augment the collaborative effectiveness of multilevel features, which is crucial for accurate recognition in small sample scenarios.

The ‘‘consistency in features of different levels’’ involves two main steps: 1) measuring the consistency of features across different levels; and 2) assessing the consistency of features at  $K$  levels. The process, depicted in Fig. 1, unfolds as follows.

Starting with SAR image datasets  $X = \{x_{11}, \dots, x_{ij}, \dots, x_{CN}\}$  and the associated features  $F_{ij} = \{f_{ij}^1, \dots, f_{ij}^K\}$  for each image  $x_{ij}$ :

*Step 1:* We measure the consistency between features at levels  $t$  and  $k$  for  $x_{ij}$  by unifying their scales using lightweight convolution and then employing a mixed method to gauge their similarity, calculated as

$$d_c(f_{ij}^k, f_{ij}^t) = d(f_{ij}^k, f_{ij}^t) + \sqrt{(f_{ij}^k - f_{ij}^t)^2} \quad (8)$$

where  $d(f_{ij}^k, f_{ij}^{k-1})$  represents the cosine similarity.

*Step 2:* For the consistency measurement of features at  $K$  levels for  $x_{ij}$ , we compute

$$L_c(x_{ij}) = \sum_k \sum_t \frac{k + |k - t|^p}{k} d_c(f_{ij}^k, f_{ij}^t) \quad (9)$$

where  $|\cdot|$  computes the absolute value.

The final consistency measurement across  $K$  levels is thus

$$L_c = \sum_i \sum_j L_c(x_{ij}). \quad (10)$$

In addition, we use the cross-entropy loss as the basic recognition loss

$$L_{ce} = - \sum_i \sum_j y_{ij} \log(p(y_{ij}|x_{ij})). \quad (11)$$

The  $K$ th feature of  $x_{ij}$  post dense layer processing is utilized as the final recognition feature for calculating  $L_{ce}$ .

By employing these methods, our approach leverages limited supervisory signals to optimize and progressively constrain features at  $K$  levels while enhancing consistency among level features, effectively utilizing the limited supervisory signals. Diverging from traditional overall optimization methods, our approach to single-layer optimization increases the likelihood of obtaining optimal features at each layer. The progressive level constraints and consistency constraints collectively boost feature effectiveness, enabling accurate recognition even with limited sample sizes. The complete process is summarized in Algorithm 1.

### III. EXPERIMENTAL RESULTS

In the upcoming section, we will first introduce the dataset that we have selected for our experimental evaluation. Subsequently, to authenticate the efficacy of our proposed method, we will conduct a series of performance assessments of the recognition capability under a variety of sample number scenarios. These tests will further illustrate the robustness and versatility of our method under varying conditions.

#### A. Dataset and Network Setup

The moving and stationary target acquisition and recognition (MSTAR) dataset serves as a benchmark for evaluating SAR ATR performance. Comprising SAR images of ten distinct ground target classes, each image in the dataset measures  $0.3 \text{ m} \times 0.3 \text{ m}$ . Fig. 2 displays optical images along with corresponding SAR images for all ten target classes within the MSTAR dataset. The training and testing data both feature the same ten target classes, but differ in terms of depression angles. While the training data were captured at a depression angle

#### Algorithm 1: Enhanced Recognition With Multilevel Feature Consistency.

---

**Input:** Training SAR image dataset  
 $X = \{x_{11}, \dots, x_{ij}, \dots, x_{CN}\}$ ; Total class number  $C$ ; Number of SAR images per class  $N$ .

**Output:** Optimized recognition model.

```

1 for each image  $x_{ij}$  in  $X$  do
2    $F_{ij} \leftarrow \text{ExtractFeatures}(x_{ij})$ ;
3   for  $k = 1$  to  $K$  do
4      $Le^k, Ls^k \leftarrow$ 
       MeasureSingleLayerEffectiveness( $F_{ij}, k$ );
5      $Lg^k \leftarrow \text{ProgressiveConstraint}(Le^{k-1}, Le^k, k)$ ;
6   end
7    $L_{gc} \leftarrow \sum_k Lg^k$ ;
8    $L_c \leftarrow \text{ConsistencyMeasurement}(F_{ij})$ ;
9    $L_{ce} \leftarrow \text{CrossEntropy}(F_{ij}, i)$ ;
10   $L_{total} \leftarrow L_{ce} + L_{gc} + L_c$ ;
11  Optimize the model using  $L_{total}$ ;
12 end
13 return Optimized recognition model
```

---

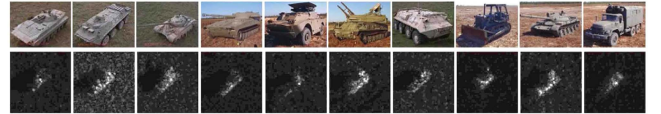


Fig. 2. Optical images and corresponding SAR images of ten classes of objects in the MSTAR database. (From left to right: BMP2, BTR70, T72, 2S1, BRDM2, ZSU234, BTR60, D7, T62, and ZIL131.).

TABLE I  
ORIGINAL NUMBER OF IMAGES FOR DIFFERENT DEPRESSION ANGLES FOR SOC

| Class      | Training |            | Testing |            |
|------------|----------|------------|---------|------------|
|            | Number   | Depression | Number  | Depression |
| BMP2-9563  | 233      | 17°        | 195     | 15°        |
| BRDM2-E71  | 298      |            | 274     |            |
| BTR60-7532 | 256      |            | 195     |            |
| BTR70-c71  | 233      |            | 196     |            |
| D7-92      | 299      |            | 274     |            |
| 2S1-b01    | 299      |            | 274     |            |
| T62-A51    | 299      |            | 273     |            |
| T72-132    | 232      |            | 196     |            |
| ZIL131-E12 | 299      |            | 274     |            |
| ZSU234-d08 | 299      |            | 274     |            |

of 17°, the testing data were acquired at an angle of 15°. The distribution of training and testing images is detailed in Table I.

OpenSARShip is one key benchmark dataset in the field of SAR ship recognition. The OpenSARShip dataset is specifically designed to facilitate the development of advanced ship detection and classification algorithms that can perform under high interference. The data comprising this dataset were collected from 41 Sentinel-1 images under diverse environmental conditions. The dataset encompasses a total of 11 346 ship chips representing 17 types of SAR ships, which are combined with

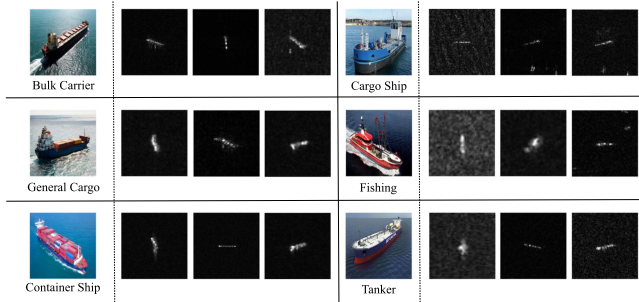


Fig. 3. SAR ship images and corresponding optical ship images of six classes in the OpenSARShip dataset.

TABLE II  
RECOGNITION ACCURACY (%) WITH DIFFERENT NUMBERS OF LABELED IMAGES EACH CLASSES UNDER MSTAR

| Class   | Labeled Number in Each Class |       |        |        |        |        |        |
|---------|------------------------------|-------|--------|--------|--------|--------|--------|
|         | 5                            | 10    | 20     | 40     | 80     | 100    | All    |
| BMP2    | 38.46                        | 62.05 | 82.56  | 98.46  | 99.49  | 100.00 | 99.49  |
| BRDM2   | 74.09                        | 71.17 | 97.81  | 97.81  | 100.00 | 99.27  | 100.00 |
| BTR60   | 61.03                        | 54.87 | 85.64  | 97.44  | 100.00 | 98.97  | 100.00 |
| BTR70   | 45.92                        | 81.12 | 86.22  | 97.45  | 98.47  | 99.49  | 100.00 |
| D7      | 94.89                        | 92.70 | 98.18  | 96.72  | 100.00 | 100.00 | 100.00 |
| 2S1     | 81.02                        | 86.86 | 99.64  | 98.54  | 99.64  | 99.64  | 99.64  |
| T62     | 74.36                        | 82.78 | 100.00 | 98.53  | 100.00 | 100.00 | 100.00 |
| T72     | 58.16                        | 95.41 | 97.96  | 98.47  | 100.00 | 100.00 | 100.00 |
| ZIL121  | 95.26                        | 96.35 | 100.00 | 99.27  | 100.00 | 100.00 | 100.00 |
| ZSU235  | 94.89                        | 99.64 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| Average | 74.52                        | 83.46 | 95.67  | 98.31  | 99.79  | 99.75  | 99.92  |

automatic identification system (AIS) information. The reliability of the labels within this dataset stems from the fact that they are based on AIS information [36]. In our experiments, we utilized the ground range detected (GRD) data, which boasts a resolution of  $2.0 \text{ m} \times 1.5 \text{ m}$  and a pixel size of  $10 \text{ m} \times 10 \text{ m}$  in azimuth and distance directions under the Sentinel-1 IW mode. The lengths of the ships in this dataset range from 92 to 399 m, with widths spanning from 6 to 65 m. The SAR images of six different ship classes within this dataset are illustrated in Fig. 3.

The experimental setup and the architecture of the network are detailed in this section. The SAR input images are resized to dimensions of  $224 \times 224$  using bilinear interpolation on the raw data. We set the  $k$  value to 4, which implies we have 4 granularities. We initialize the batch size to 32. The learning rate is initially set to 0.01, and it is subsequently decreased by a factor of 0.1 every 25 epochs. We also incorporate a warming-up period of 10 epochs prior to actual training. Our proposed method was implemented and tested on a GPU cluster featuring an Intel Xeon CPU E5-2698 v4 @ 2.20 GHz, and eight Tesla V100 units, each equipped with 32 GB memory. The implementation leveraged the open-source PyTorch framework, utilizing a single Tesla V100 unit.

### B. Recognition Performance Under MSTAR Dataset

The recognition results of the proposed method are shown in Table II. The number of training samples per class ranges from 5 to all to validate the effectiveness of our method. From

TABLE III  
IMAGE NUMBER AND IMAGING CONDITIONS OF DIFFERENT TARGETS IN OPENSARSHIP

| Class          | Imaging Condition   | Training | Testing | Total  |
|----------------|---|----------|---------|--------|
|                |   | Number   | Number  | Number |
| Bulk Carrier   | VH and VV, C band<br>Resolution = 5–20m<br>Incident angle = $20^\circ$ – $45^\circ$ | 200      | 475     | 675    |
| Container Ship |   | 200      | 811     | 1011   |
| Tanker         | Elevation sweep angle = $\pm 11^\circ$<br>Rg20m $\times$ az22m                      | 200      | 354     | 554    |
| Cargo          |   | 200      | 557     | 757    |
| Fishing        | Rg20m $\times$ az22m  | 200      | 121     | 321    |
| General Cargo  |   | 200      | 165     | 365    |

the table, facing 5-shot each target, our method can achieve a 74.52% overall recognition ratio, and when the training number of each class in a few labeled SAR samples is 20 and 10, respectively, the recognition rate just decreases from 95.67% to 83.46%. It indicates that our method can greatly improve the recognition performance when the labeled samples increase. When the number of each class in a few labeled SAR samples is 40, it gets above 98.31%, and most classes of all targets are above 98.00%. It shows that the innovations proposed in our method are beneficial for the recognition of a few labeled samples, though the one class of the target is more sensitive to the decreased labeled samples. In addition, it is clear that, when the number of each class in a few labeled SAR samples is larger than 40, the recognition rate can get above 98.00%. It demonstrates the effectiveness of the proposed method in SAR ATR under few labeled SAR samples and relatively sufficient SAR samples. When all the training samples are employed, the proposed method also achieves state-of-art performance in SAR ATR.

Through the experiments of different sample numbers under MSTAR, it has been illustrated that our method can achieve superior recognition performance facing a large range of number of training sample each class. Then the recognition performance under OpenSARship dataset is presented.

### C. Recognition Performance Under OpenSARship Dataset

In this section, we conduct two sets of recognition experiments using the OpenSARShip dataset to thoroughly assess the efficacy of our method. The OpenSARShip dataset, which includes several ship classes prevalent in the international shipping market, constitutes a significant portion of this market, as referenced in [37]. The experiments are carried out for both three and six-class categorizations, in line with [37], [38], and [39]. Specifically, the three-class experiments involve bulk carriers, container ships, and tankers, whereas the six-class trials additionally include cargo ships, fishing vessels, and general cargo ships, as detailed in Table III. The image preprocessing for these experiments aligns with the protocols used in the MSTAR dataset.

Table IV showcases our method's results in the three-class OpenSARShip experiment. Impressively, with 200 training samples per class, our method attains an overall recognition rate

TABLE IV  
RECOGNITION ACCURACY (%) OF THREE CLASSES UNDER DIFFERENT TRAINING DATA IN OPENSARSHIP

| Class          | Labeled Number in Each Class |       |       |       |       |       |       |       |       |       |
|----------------|------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|                | 10                           | 20    | 30    | 40    | 50    | 60    | 70    | 80    | 100   | 200   |
| Bulk Carrier   | 47.79                        | 53.68 | 54.53 | 58.32 | 64.63 | 71.58 | 65.47 | 70.74 | 68.84 | 78.53 |
| Container Ship | 84.22                        | 80.02 | 83.35 | 80.89 | 86.81 | 83.35 | 87.30 | 83.72 | 89.15 | 89.15 |
| Tanker         | 68.08                        | 82.77 | 84.18 | 82.49 | 78.25 | 83.62 | 83.05 | 86.72 | 84.75 | 87.01 |
| Average        | 69.76                        | 72.99 | 73.96 | 76.34 | 78.54 | 80.00 | 80.06 | 80.61 | 82.32 | 85.61 |

TABLE V  
RECOGNITION ACCURACY (%) OF SIX CLASSES UNDER DIFFERENT TRAINING DATA IN OPENSARSHIP

| Class          | Labeled Number in Each Class |       |       |       |       |       |       |       |       |       |
|----------------|------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|                | 10                           | 20    | 30    | 40    | 50    | 60    | 70    | 80    | 100   | 200   |
| Bulk Carrier   | 50.74                        | 62.53 | 56.00 | 53.26 | 64.42 | 56.63 | 68.00 | 59.58 | 68.42 | 69.47 |
| Container Ship | 65.35                        | 65.60 | 71.15 | 73.74 | 71.27 | 75.83 | 77.81 | 73.49 | 79.53 | 81.26 |
| Tanker         | 50.28                        | 47.18 | 45.76 | 39.83 | 56.50 | 56.50 | 47.74 | 49.72 | 48.02 | 50.28 |
| Cargo          | 40.04                        | 42.37 | 47.04 | 52.60 | 49.37 | 52.24 | 44.70 | 58.53 | 53.14 | 61.76 |
| Fishing        | 81.82                        | 82.64 | 86.78 | 91.74 | 90.91 | 83.47 | 90.08 | 88.43 | 91.74 | 92.56 |
| General Cargo  | 17.58                        | 30.91 | 44.85 | 39.39 | 39.39 | 38.79 | 40.61 | 41.82 | 32.12 | 46.67 |
| Average        | 52.36                        | 55.70 | 58.24 | 58.84 | 61.78 | 62.02 | 62.34 | 62.71 | 64.44 | 68.47 |

TABLE VI  
ABLATION EXPERIMENTS: AVERAGE RECOGNITION ACCURACY (%) OF SIX CLASSES UNDER DIFFERENT TRAINING DATA IN OPENSARSHIP WITH DIFFERENT LEVEL OF GRANULARITY

| Levels | Labeled Number in Each Class |              |              |              |              |              |
|--------|------------------------------|--------------|--------------|--------------|--------------|--------------|
|        | 10                           | 20           | 30           | 40           | 80           | 100          |
| 1      | 51.04                        | 53.21        | 56.45        | 56.70        | 61.78        | 63.13        |
| 2      | 51.55                        | 54.35        | 57.27        | 57.34        | 62.65        | 63.56        |
| 3      | 52.13                        | 54.99        | 58.03        | 57.99        | <b>62.83</b> | 63.89        |
| 4      | <b>52.36</b>                 | 55.70        | <b>58.24</b> | <b>58.84</b> | 62.71        | <b>64.44</b> |
| 5      | 52.21                        | <b>56.32</b> | 57.94        | 58.35        | 62.65        | 64.05        |
| 6      | 51.40                        | 56.01        | 56.35        | 58.11        | 62.47        | 63.87        |

Bold values indicate the best results for the respective experimental metrics within the same category.

of 85.61%. Remarkably, the recognition rates only slightly decrease from 82.32% to 80.00% when training samples are reduced from 100 to 60 per class, underscoring our method's robustness and effectiveness. A further reduction in training samples from 30 to 20 leads to a stable recognition rate around 72.99%, a commendable achievement in SAR ship target recognition. Even under stringent conditions with just ten samples per class, the method maintains a robust recognition rate of 69.76%.

In the 6-class OpenSARShip experiment, as shown in Table V, the challenge is amplified due to the high similarity among different ship classes and the significant variation within the same class. Here, with 200 training samples per class, the recognition rate stands at 68.47%. The rates remain relatively stable, marginally dropping from 64.40% to 62.02%, as the training data reduces from 100 to 60 samples. This robustness is consistent with the three-class performance. With a further reduction in training samples from 40 to 20, the recognition

rates show slight variation, ranging from 58.84% to 55.70%. Even with as few as ten samples per class, the method achieves a recognition rate of 52.39%, reflecting the typical challenges faced in practical applications.

Our analysis reveals that for each dataset, the performance initially improves rapidly with an increase in the number of samples. During this phase, the diversity of data supplements the model's learning of key SAR image features, leading to swift performance enhancement. However, as the sample size reaches a certain threshold, the addition of key features becomes less pronounced. The model learns more about the same category in different environments, providing auxiliary information that aids in classification, hence the slower rate of improvement.

The quantitative results from the OpenSARShip experiments demonstrate that our method not only achieves high accuracy in SAR ship target recognition but also maintains robustness and effectiveness even with diminishing numbers of training samples.

#### D. Ablation Experiments

In this section, we assess the effectiveness of our method through ablation experiments conducted on the OpenSARShip and MSTAR datasets.

Initially, we perform these experiments with four configurations of our method: each using 40 samples per class for OpenSARShip and ten samples per class for MSTAR. As detailed in Table VII, the configurations are: V1, the basic vanilla model without our proposed innovations; V2, which incorporates the coarse-to-fine gradual feature constraint into the vanilla model; V3, which applies consistency in features of different granularities to the vanilla model; and ours, the full version of our method. The recognition performances of these configurations are presented in Tables VII and VIII.

TABLE VII  
ABLATION EXPERIMENTS: RECOGNITION PERFORMANCE (%) OF DIFFERENT ABLATION CONFIGURATIONS UNDER 40 TRAINING SAMPLES IN OPENSARSHIP

| Method | CFGFC | CFDG | Bulk Carrier | Container Ship | Tanker       | Cargo        | Fishing      | General Cargo | Average      |
|--------|-------|------|--------------|----------------|--------------|--------------|--------------|---------------|--------------|
| V1     | ×     | ×    | 49.68        | 76.33          | <b>50.85</b> | 36.98        | 82.64        | 37.58         | 56.50        |
| V2     | ✓     | ×    | 47.58        | <b>76.57</b>   | 38.42        | 50.45        | 90.91        | <b>44.85</b>  | 58.32        |
| V3     | ×     | ✓    | 49.26        | 74.85          | 45.48        | 40.75        | 89.26        | 43.03         | 56.71        |
| Ours   | ✓     | ✓    | <b>53.26</b> | 74.74          | 39.83        | <b>52.60</b> | <b>91.74</b> | 39.39         | <b>58.84</b> |

CFGFC stands for coarse-to-fine gradual feature constraint and CFDG stands for consistency in features of different granularities.

Bold values indicate the best results for the respective experimental metrics within the same category.

TABLE VIII  
ABLATION EXPERIMENTS: RECOGNITION PERFORMANCE (%) OF DIFFERENT ABLATION CONFIGURATIONS SHOT UNDER TEN TRAINING SAMPLES IN MSTAR

| Method | CFGFC | CFDG | BMP2         | BRDM2        | BTR60        | BTR70        | D7           | 2S1          | T62          | T72          | ZIL121       | ZSU235       | Average      |
|--------|-------|------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| V1     | ×     | ×    | 61.54        | 64.60        | 70.77        | 69.90        | 89.78        | 74.45        | 79.12        | 82.14        | 89.78        | 98.91        | 79.01        |
| V2     | ✓     | ×    | 72.82        | 62.04        | 48.21        | <b>73.47</b> | <b>95.62</b> | 77.37        | <b>89.38</b> | <b>90.31</b> | 91.97        | 99.27        | 81.20        |
| V3     | ×     | ✓    | <b>81.03</b> | 65.33        | 48.72        | 70.41        | 90.51        | 71.90        | 86.45        | 85.71        | 84.67        | <b>99.64</b> | 79.34        |
| Ours   | ✓     | ✓    | 57.95        | <b>68.98</b> | <b>73.33</b> | 70.92        | 94.89        | <b>94.53</b> | 86.81        | 89.80        | <b>93.43</b> | 98.18        | <b>84.16</b> |

CFGFC stands for coarse-to-fine gradual feature constraint and CFDG stands for consistency in features of different granularities.

Bold values indicate the best results for the respective experimental metrics within the same category.

The comparative analysis of these configurations highlights the enhancement in recognition performance due to our innovations. Specifically, the comparison between V1 and V2 demonstrates the effectiveness of the coarse-to-fine gradual feature constraint. Moreover, the performance difference between V2 and V3 indicates that the coarse-to-fine feature constraint is more impactful than consistency in features across different granularities. However, the combined implementation of these two features in the “Ours” configuration illustrates that the consistency of features across different granularities can synergistically work with the coarse-to-fine feature constraint to significantly enhance recognition performance.

Further, detailed ablation experiments in Table VI for the OpenSARShip model reveal the impact of granularity consistency across various shot settings. The model shows the best average performance at level 4 granularity across different shot settings. Generally, model performance increases with granularity up to level 4, but beyond this point, granularity levels 5 and 6 lead to a decrease in performance due to model overfitting. This pattern validates the necessity of multilevel granularity consistency and, by extension, the efficacy of our approach.

Building on these findings, we will next present a comparison of our methods with other state-of-the-art approaches in SAR ATR under limited data conditions.

### E. Comparison

In this section, the comparisons with other state-of-the-art methods are presented under MSTAR and OpenSARShip datasets.

TABLE IX  
RECOGNITION ACCURACY (%) UNDER DIFFERENT NUMBERS OF LABELED IMAGES

| Method              | Labeled Number |              |              |              |       |              |
|---------------------|----------------|--------------|--------------|--------------|-------|--------------|
|                     | ALL            | 100          | 80           | 40           | 30    | 20           |
| PCA+SVM [43]        | 94.32          | –            | 92.48        | 87.95        | –     | 76.43        |
| ADaboost [43]       | 93.51          | –            | 91.45        | 86.45        | –     | 75.68        |
| LC-KSVD [37]        | 95.13          | –            | 93.23        | 87.39        | –     | 78.83        |
| DGM [43]            | 96.07          | –            | 92.85        | 88.14        | –     | 81.11        |
| Gauss [44]          | 97.16          | –            | 94.10        | 87.51        | –     | 80.55        |
| DNN1 [45]           | 95.54          | –            | 93.04        | 86.98        | –     | 77.86        |
| DNN2 [46]           | 96.50          | –            | 93.76        | 87.73        | –     | 79.39        |
| CNN [43]            | 97.03          | –            | 93.88        | 88.35        | –     | 81.80        |
| GAN-CNN [43]        | 97.53          | –            | 94.91        | 90.13        | –     | 84.39        |
| MGAN-CNN [43]       | 97.81          | –            | 94.91        | 90.82        | –     | 85.23        |
| TMDC-CNNs [34]      | –              | 99.09        | –            | –            | 86.93 | –            |
| ARGN [35]           | 98.00          | –            | 95.98        | –            | 87.20 | –            |
| Semisupervised [47] | –              | –            | 98.65        | 97.11        | –     | 92.62        |
| The Proposed        | <b>99.92</b>   | <b>99.75</b> | <b>99.79</b> | <b>98.31</b> | –     | <b>95.67</b> |

Bold values indicate the best results for the respective experimental metrics within the same category.

We compare our approach with other SAR ATR methods, incorporating five traditional methods (PAC+SVM, ADaboost, LC-KSVD, DGM, and Gauss) and several few-shot learning methods. The latter includes three data augmentation



TABLE X  
COMPARISON OF PERFORMANCE (%) OF THREE CLASSES UNDER  
OPENSARSHIP

| Methods              | Number range of training images in each class |                    |                    |
|----------------------|---|--------------------|--------------------|
|                      | 1–50  | 51–100             | 101–338            |
| Semi-Supervised [37] | 61.88 (20)                                    | 68.67 (80)         | 71.29 (120)        |
|                      | 64.73 (40)                                    |                    | 74.96 (240)        |
| Supervised [37]      | 58.24 (20)                                    | 65.63 (80)         | 68.75 (120)        |
|                      | 62.09 (40)                                    |                    | 70.83 (240)        |
| CNN [40]             | 62.75 (50)                                    | 68.52 (100)        | 73.68 (200)        |
| CNN+Matrix [40]      | 72.86 (50)                                    | 75.31 (100)        | 77.22 (200)        |
| PFGFE-Net [41]       | -   | -                  | 79.84 (338)        |
| MetaBoost [42]       | -   | -                  | 80.81 (338)        |
| Proposed             | <b>69.76 (10)</b>                             | <b>80.00 (60)</b>  | <b>85.61 (200)</b> |
|                      | <b>72.99 (20)</b>                             | <b>80.06 (70)</b>  |                    |
|                      | <b>73.96 (30)</b>                             | <b>80.61 (80)</b>  |                    |
|                      | <b>78.54 (40)</b>                             | <b>82.32 (100)</b> |                    |

The number in parentheses is the number of the training samples for each method. Bold values indicate the best results for the respective experimental metrics within the same category.

methods (GAN-CNN, MGAN-CNN, and ARGN), a metric-based method (TMDC-CNNs), and four model-based methods (DNN1, DNN2, CNN, and semisupervised).

When testing using 20, 40, 80, or all target samples from each label as training images, it is evident from Table IX that our proposed method outperforms the other methods in terms of recognition rate. Remarkably, even when the training dataset for each target comprises only 20 samples, our method achieves a recognition rate exceeding 94%, whereas most other methods fall below 86%. These findings attest not only to the effectiveness of our method when leveraging all training images in the dataset but also to its proficiency in few-shot learning scenarios.

Table X presents a comparison of our method with several other approaches used for SAR ship recognition. The methods we compare include the following. 1) Semisupervised learning [37], which leverages unlabeled data during training to reduce the reliance on labeled data. 2) Supervised [37], a variant of semisupervised learning that is fully supervised. 3) CNN [40], a conventional classification framework. 4) CNN+matrix [40], which combines a convolutional neural network with a matrix for improved performance. 5) PFGFE-Net [41] that achieves polarization fusion at input, feature, and decision levels to address the issue of insufficient utilization of polarization information. 6) MetaBoost [42], a method that employs a two-stage filtration system focusing on the generation and combination of “good and different” base classifiers. For the purpose of our comparison, following [42], we divided the number of training images for each class into three bands: 1–50, 51–100, and 101–338.

From Table X, it is evident that our method further enhances the recognition performance across different sample quantity intervals. In the 1–50 quantity range, among the compared methods, semisupervised learning achieves an excellent performance of 61.88% when each class has 20 shots, whereas our method reaches 71.46%. In the 51–100 quantity range, CNN [40] and semisupervised learning [37] both deliver impressive performance, exceeding 68.50% when each class has 80 shots. Our method, on the other hand, reaches a recognition rate of 79.33%. Through comparison with other methods on the OpenSARShip dataset, it is clear that our method can achieve state-of-the-art performance.

From the above comparison, it is evident that our method can attain state-of-the-art recognition performance across a broad spectrum of training sample sizes. More significantly, our method clearly surpasses all other methods in both types of recognition tasks, regardless of the number of classes involved.

#### IV. CONCLUSION

In conclusion, this article addresses the significant challenges in the increasingly explored field of SAR ATR under the constraints of limited data, primarily caused by the scarcity of abundant training samples. We have conducted a thorough analysis of the limitations inherent in existing SAR ATR methods, such as those relying on data augmentation, specialized modules, or contrastive learning-based loss functions. Our analysis underscores the vital importance of effectively utilizing limited supervision signals for precise target identification. We introduce a groundbreaking approach in the form of coarse-to-fine hierarchical feature learning tailored for SAR ATR with limited data. Our method is distinguished by its feature extractor that yields multilevel features and a unique coarse-to-fine gradual feature constraint. This approach enables individualized optimization of features at each level, significantly reducing the complexity of model parameter search and ensuring the robustness of features across all granularity levels. A key aspect of our methodology is the optimization of intraclass compactness and interclass separability of features at different levels, achieved by leveraging the consistency of features across these levels. This strategy of progressively enhancing feature effectiveness substantially elevates the overall performance of the model. Empirical experiments conducted on the MSTAR and OpenSARShip datasets have demonstrated the superior performance of our approach in scenarios with limited sample sizes. The robustness of our method has been further validated through comprehensive ablation studies. This research contributes meaningful insights to the ongoing advancements in SAR ATR with limited data, paving the way for more sophisticated, efficient, and effective solutions in the future.

#### REFERENCES

- [1] G. Akbarizadeh, “A new statistical-based kurtosis wavelet energy feature for texture recognition of SAR images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4358–4368, Nov. 2012.
- [2] B. Bhanu, “Automatic target recognition: State of the art survey,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 4, pp. 364–379, Jul. 1986.

- [3] W.-T. Chen, K.-F. Ji, X. -W. Xing, H. -X. Zou, and H. Sun, "Ship recognition in high resolution SAR imagery based on feature selection," in *Proc. Int. Conf. Comput. Vis. Remote Sens.*, 2012, pp. 301–305.
- [4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [5] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.
- [6] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.
- [7] C. Wang, J. Pei, X. Liu, Y. Huang, and J. Yang, "A deep deformable residual learning network for SAR image segmentation," in *Proc. IEEE Radar Conf.*, 2021, pp. 1–5.
- [8] L. Li, X. Yao, X. Wang, D. Hong, G. Cheng, and J. Han, "Robust few-shot aerial image object detection via unbiased proposals filtration," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5617011.
- [9] K. Fu, T. Zhang, Y. Zhang, Z. Wang, and X. Sun, "Few-shot SAR target classification via metalearning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 2000314.
- [10] 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 Geosci. Remote Sens. Lett.*, vol. 17, no. 11, pp. 1928–1932, Nov. 2020.
- [11] F. Zhang, Y. Wang, J. Ni, Y. Zhou, and W. Hu, "SAR target small sample recognition based on CNN cascaded features and adaboost rotation forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1008–1012, Jun. 2020.
- [12] 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, 2018, Art. no. 2929.
- [13] J. C. Curlander and R. N. McDonough, *Synthetic Aperture Radar*. NY, New York, USA: Wiley, 1991.
- [14] A. Howard et al., "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1314–1324.
- [15] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," 2016, *arXiv:1602.07360*.
- [16] C. Wang, J. Pei, J. Yang, X. Liu, Y. Huang, and D. Mao, "Recognition in label and discrimination in feature: A hierarchically designed lightweight method for limited data in SAR ATR," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5239613.
- [17] C. Wang et al., "An entropy-awareness meta-learning method for SAR open-set ATR," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 4005105.
- [18] X. Wang, Z. Cao, and Y. Pi, "Semisupervised classification with adaptive anchor graph for PolSAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 4009705.
- [19] Z. Yue et al., "A novel semi-supervised convolutional neural network method for synthetic aperture radar image recognition," *Cogn. Comput.*, vol. 13, no. 4, pp. 795–806, 2021.
- [20] S.-I. Hwang and K. Ouchi, "On a novel approach using MLCC and CFAR for the improvement of ship detection by synthetic aperture radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 391–395, Apr. 2010.
- [21] X. Xing, K. Ji, H. Zou, W. Chen, and J. Sun, "Ship classification in TerraSAR-X images with feature space based sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1562–1566, Nov. 2013.
- [22] H. Lang and S. Wu, "Ship classification in moderate-resolution SAR image by naive geometric features-combined multiple kernel learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1765–1769, Oct. 2017.
- [23] C. Wang et al., "SAR ship target recognition via multiscale feature attention and adaptive-weighted classifier," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 4003905.
- [24] C. Wang, Y. Huang, X. Liu, J. Pei, Y. Zhang, and J. Yang, "Global in local: A convolutional transformer for SAR ATR FSL," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4509605.
- [25] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 31, no. 1, 2017.
- [26] C. Cao, Z. Cao, and Z. Cui, "LDGAN: A synthetic aperture radar image generation method for automatic target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3495–3508, May 2020.
- [27] C. Li, T. Xu, J. Zhu, and B. Zhang, "Triple generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [28] C. Wang, X. Liu, J. Pei, Y. Huang, Y. Zhang, and J. Yang, "Multiview attention CNN-LSTM network for SAR automatic target recognition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 12504–12513, 2021.
- [29] C. Wang, J. Pei, M. Li, Y. Zhang, Y. Huang, and J. Yang, "Parking information perception based on automotive millimeter wave SAR," in *Proc. IEEE Radar Conf.*, 2019, pp. 1–6.
- [30] C. Wang et al., "Semi-supervised SAR ATR framework with transductive auxiliary segmentation," *Remote Sens.*, vol. 14, no. 18, 2022, Art. no. 4547.
- [31] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1251–1258.
- [32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.
- [33] C. Wang et al., "When deep learning meets multi-task learning in SAR ATR: Simultaneous target recognition and segmentation," *Remote Sens.*, vol. 12, no. 23, 2020, Art. no. 3863.
- [34] F. Wang and C. Zhang, "Label propagation through linear neighborhoods," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 985–992.
- [35] C. Persello and L. Bruzzone, "Active and semisupervised learning for the classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6937–6956, Nov. 2014.
- [36] I. M. O. (IMO), "Guidelines for the onboard operational use of shipborne automatic identification systems (AIS)," Int. Maritime Org. (IMO), London, UK, 2002.
- [37] C. Wang et al., "Semisupervised learning-based SAR ATR via self-consistent augmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4862–4873, Jun. 2021 .
- [38] T. Zhang et al., "HOG-ShipCLSNet: A novel deep learning network with hog feature fusion for SAR ship classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5210322.
- [39] J. He, Y. Wang, and H. Liu, "Ship classification in medium-resolution SAR images via densely connected triplet CNNs integrating fisher discrimination regularized metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3022–3039, Apr. 2021.
- [40] Y. Li, X. Li, Q. Sun, and Q. Dong, "SAR image classification using CNN embeddings and metric learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2020, Art. no. 4002305.
- [41] T. Zhang and X. Zhang, "A polarization fusion network with geometric feature embedding for SAR ship classification," *Pattern Recognit.*, vol. 123, 2022, Art. no. 108365.
- [42] H. Zheng, Z. Hu, J. Liu, Y. Huang, and M. Zheng, "Metaboost: A novel heterogeneous DCNNs ensemble network with two-stage filtration for SAR ship classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4509005.
- [43] C. Zheng, X. Jiang, and X. Liu, "Semi-supervised SAR ATR via multi-discriminator generative adversarial network," *IEEE Sensors J.*, vol. 19, no. 17, pp. 7525–7533, Sep. 2019.
- [44] J. A. O'Sullivan, M. D. DeVore, V. Kedia, and M. I. Miller, "SAR ATR performance using a conditionally gaussian model," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 1, pp. 91–108, Jan. 2001.
- [45] Q. Qian, L. Shang, B. Sun, J. Hu, H. Li, and R. Jin, "Softtriplet loss: Deep metric learning without triplet sampling," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6450–6458.
- [46] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 364–368, Mar. 2016.
- [47] F. Gao, Y. Yang, J. Wang, J. Sun, E. Yang, and H. Zhou, "A deep convolutional generative adversarial networks (DCGANs)-based semi-supervised method for object recognition in synthetic aperture radar (SAR) images," *Remote Sens.*, vol. 10, no. 6, 2018, Art. no. 846.



**Yan Wen** received the bachelor's degree in management from the Chengdu University of Technology, Chengdu, China, and another bachelor's degree in engineering from the University of Electronic Science and Technology of China, Chengdu, in 2004, and the master's degree in engineering from Southwest Jiaotong University, Chengdu, in 2014.

She is currently employed with Chengdu Agricultural College, Chengdu. Her primary research interests include the development and application of deep learning algorithms in smart agriculture.



**Xihao Wang** received the B.Sc. degree in mathematics and applied mathematics in 2022 from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, where he is currently working toward the M.Sc. degree in statistics.

His research interests include state estimation, multimanoeuvre target tracking, and multisource information fusion.



**Yu Qiao** received the bachelor's degree in computer science from Chongqing University (CQU), Chongqing, China, in 2022.

He is currently employed with Citibank, Shanghai, China, where he is engaged in deep learning research. His primary research interests include object detection, data mining, and few-shot learning.



**Lihe Peng** received the B.Sc. degree in information and computing science from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2022.

He is currently a Numerical Software Engineer with Suzhou Tongyuan Software & Control Technology Company, Suzhou, China.