






# Self-Supervised Feature Representation for SAR Image Target Classification Using Contrastive Learning

Hao Pei , Graduate Student Member, IEEE, Mingjie Su , Gang Xu , Senior Member, IEEE, Mengdao Xing , Fellow, IEEE, and Wei Hong , Fellow, IEEE

**Abstract**—Nowadays, the developed deep neural networks (DNNs) have been widely applied to synthetic aperture radar (SAR) image interpretation, such as target classification and recognition, which can automatically learn high-level semantic features in data-driven and task-driven manners. For the supervised learning methods, abundant labeled samples are required to avoid the over-fitting of designed networks, which is usually difficult for SAR image applications. To address these issues, a novel two-stage algorithm based on contrastive learning (CL) is proposed for SAR image target classification. In the pretraining stage, to extract self-supervised representations (SSRs) from an unlabeled train set, a convolutional neural network (CNN)-based encoder is first pretrained using a contrasting strategy. This encoder can convert SAR images into a discriminative embedding space. Meanwhile, the optimal encoder can be determined using a linear evaluation protocol, which can indirectly confirm the transferability of pre-learned SSRs to downstream tasks. Therefore, in the fine-tuning stage, a SAR target classifier can be adequately trained using a few labeled SSRs in a supervised manner, which benefits from the powerful pretrained encoder. Numerical experiments are carried out on the shared MSTAR dataset to demonstrate that the model based on the proposed self-supervised feature learning algorithm is superior to the conventional supervised methods under labeled data constraints. In addition, knowledge transfer experiments are also conducted on the openSARship dataset, showing that the encoder pretrained from the MSTAR dataset can support the classifier training with high efficiency and precision. These results demonstrate the excellent training convergence and classification performance of the proposed algorithm.

**Index Terms**—Contrastive learning (CL), convolutional neural network (CNN), self-supervised representation (SSR) learning, synthetic aperture radar (SAR) image, target classification.

Manuscript received 30 March 2023; revised 31 July 2023 and 4 September 2023; accepted 27 September 2023. Date of publication 3 October 2023; date of current version 13 October 2023. This work was supported in part by the National Science Foundation of China (NSFC) under Grant 62071113, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20211559, and in part by the Fundamental Research Funds for the Central Universities under Grant 2242022k60008 and Grant 2242022R40008. (Corresponding author: Gang Xu.)

Hao Pei, Mingjie Su, Gang Xu, and Wei Hong are with the State Key Laboratory of Millimeter Waves, School of Information Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: phww98@seu.edu.cn; 220220896@seu.edu.cn; gangxu@seu.edu.cn; weihong@seu.edu.cn).

Mengdao Xing is with the National Key Laboratory of Radar Signal Processing, Xidian University, Xian 710071, China, and also with the Academy of Advanced Interdisciplinary Research, Xidian University, Xian 710071, China (e-mail: xmd@xidian.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3321769

## I. INTRODUCTION

**S**YNTHETIC aperture radar (SAR) is an advanced remote sensor used to provide high-resolution images, which has many potential applications in military and civilian fields [1], [2], [3], [4], [5]. Nowadays, numerous high-resolution SAR images are available due to extensive satellite exploration activities, making automatic SAR image interpretation an urgent necessity. Target classification is a fundamental SAR interpretation task that can automatically offer target category information for other advanced applications, such as hostile target identification, high-value target surveillance [6], [7], [8], town planning [9], oil spill detection [10], etc.

SAR images inevitably exhibit speckle noise due to their coherent imaging mechanism [11]. This noise presents a significant challenge for the feature extraction capabilities of SAR target classifiers, as it can lead to small interclass differences and large intraclass differences in SAR targets. Therefore, numerous SAR target classification methods mainly focus on feature extraction. For example, principal component analysis (PCA), fisher linear discriminant analysis (LDA) [12], and local discriminant embedding (LDE) [13] directly transform SAR images into feature vectors using a linear or nonlinear transformation automatically. Some researchers have manually designed feature extractors to capture the texture, orientation, and contour features of SAR images, such as HOG [14], SIFT [15], wavelet transform [16]. Nevertheless, these handcrafted features are low-level features or originally developed for optical image textures, resulting in the poor interclass separation and discriminability of the extracted features. These shortcomings restrict the performance of the aforementioned methods.

Fortunately, with the prosperous development of deep learning (DL) technologies, numerous studies on SAR image interpretation with convolutional neural networks (CNNs) have demonstrated that they can automatically learn high-level features from large-scale datasets [17], [18], [19], [20], [21]. Early research mainly applies labeled data to drive model learning with a supervised learning paradigm, which often suffers from over-fitting problems due to the limitation of elaborately labeled SAR samples. Some researchers have proposed alleviating methods such as reducing classifier parameters [22], applying SAR image enhancement [23], and designing special models. For example, Sharifzadeh et al. [24] combined CNN and MLP modules to

propose a novel CNN–MLP hybrid classifier and benefit the SAR ship classification. Samadi et al. [25] suggested a pixel selection approach and a morphological preprocessing on input SAR images to better train a deep belief network, thus alleviating the need for labeled train samples. However, these methods still fail to break out of the supervised learning framework, which is merely a temporary fix. Transfer learning techniques based on SAR image domains have been implemented to overcome the dilemma posed by the lack of labeled SAR images. Huang et al. [26] attempted to transfer knowledge from unlabeled SAR scene data to SAR target classification tasks by feedbacking the reconstruction loss to the classification bypath. Wang et al. [27] and Youk and Kim [28] attempted to prelearn the mapping relationship between the simulated and measured SAR target images. The above methods dramatically alleviate the shortage of labeled SAR samples and achieve considerable performance. However, when and how to transfer knowledge varies by specific tasks and source domains. Such limitations can dilute the extensibility of transfer learning methods.

Self-supervised learning (SSL) is a task-agnostic paradigm that can be easily incorporated with downstream supervised learning tasks through pretext tasks that provide self-supervised signals. It is possible for this technology to learn valuable visual representations from an abundance of unlabeled data. For example, Wen et al. [29] proposed a weak rotation awareness encode method to learn rotational representations from sequence SAR images. Zhang et al. [30] employed a stacked autoencoder to learn spatial representations from SAR images via the denoising pretext task. Contrastive learning (CL) is a cutting-edge self-supervised method that can unlock the potential of self-supervised techniques. Because it can extract more discriminative features from an unlabeled dataset, which can benefit the downstream classification tasks. InstDisc [31] proposed a pretext task to learn instance-level discrimination and a novel nonparametric softmax formulation, which allows the CL model to capture the apparent similarity between instances. Based on instance discrimination, SimCLR [32] has completed the CL framework by studying its components and investigating the effects of different design choices. He et al. [33] proposed momentum contrast (MOCO) to reduce the computational demand in CL. It maintains a dynamic dictionary as a memory bank to access a large number of negative samples. They all establish instance-level discrimination as a pretext task to drive the model to learn discriminative representations. Therefore, based on the instance discrimination task, it is possible to encode raw SAR images as self-supervised representations (SSRs). These SSRs will have a more discriminative embedding space, where the samples with the same category label can be automatically clustered together and separated from other types of targets. In other words, SSRs in this embedding space will have larger interclass differences and smaller intraclass differences. Therefore, the urgency for large annotated datasets can be reduced, and the accuracy of the classifier can be improved when transferring these learned SSRs to downstream SAR target classification tasks.

The pretraining method by implementing a contrasting strategy has been applied to various remote sensing data not only

SAR images [34], [35], but also optical remote sensing images [36] and multimodal remote sensing images [37]. However, during the training of the CL model, the objective of the supervised task in the transfer stage is irrelevant to the pretext task, making it difficult to quantify the encoder’s performance in downstream transfer tasks. To address this paradox, a model-agnostic linear evaluation protocol is proposed to evaluate the state during the training of CL models. This allows the transferability and discriminability of prelearned SSRs to downstream tasks to be indirectly confirmed with a simple linear classifier.

In this article, we propose a two-stage training framework for SAR target classification models. In the pretraining stage, we investigate a CL model to learn SSRs from an unlabeled train set to overcome the insufficiency of annotated samples. Meanwhile, a linear evaluation protocol is proposed to evaluate the learned SSRs during the CL model training, which can indicate the self-supervised training status by verifying the discriminability of the learned SSRs, thereby determining the best pretrained encoder. In the fine-tuning stage, the best encoder is applied to a small set of annotated datasets, where all raw SAR images are encoded as SSRs. Within the SSRs, a supervised classifier can be fitted with a few labeled SAR images and ensure the accuracy of the SAR target classification system.

More specifically, the main contributions of this article are summarized as follows.

- 1) A novel two-stage training framework for SAR target classification is proposed, which overcomes the scarcity of labeled data by contrasting strategies so that the classification task requires only a few labeled SAR image samples to ensure considerable accuracy.
- 2) A linear evaluation protocol is implemented during each pretraining epoch to evaluate the learned SSRs. This guarantees the encoder utilized in the fine-tuning stage is the most efficient one, and further improves the accuracy of SAR target classifiers.
- 3) The proposed algorithm can achieve SOTA performance on the MSTAR dataset with a few labeled SAR samples. Meanwhile, the learned knowledge also has better generalization on the openSARship dataset.

## II. PROPOSED METHOD

In this section, a brief review of our proposed two-stage training framework for SAR target classification is first presented. Then, in the pretraining stage, the novel SSR leaning model using contrastive strategy is described in detail. We also illustrate how to assess the encoder with the proposed linear evaluation protocol. Finally, the transfer learning strategy of fine-tuning the pretrained encoder is introduced in downstream SAR target classification tasks.

The proposed two-stage algorithm are mainly consists of two stages: pretraining and fine-tuning. In stage one, we use a variant CL model to pretrain a CNN-based encoder in a self-supervised manner. The encoder will learn hierarchical features and transform the whole unlabeled train set into discriminable SSRs. Then, partial labeled SSRs will be used in the subsequent linear evaluation protocol to evaluate the transferability and

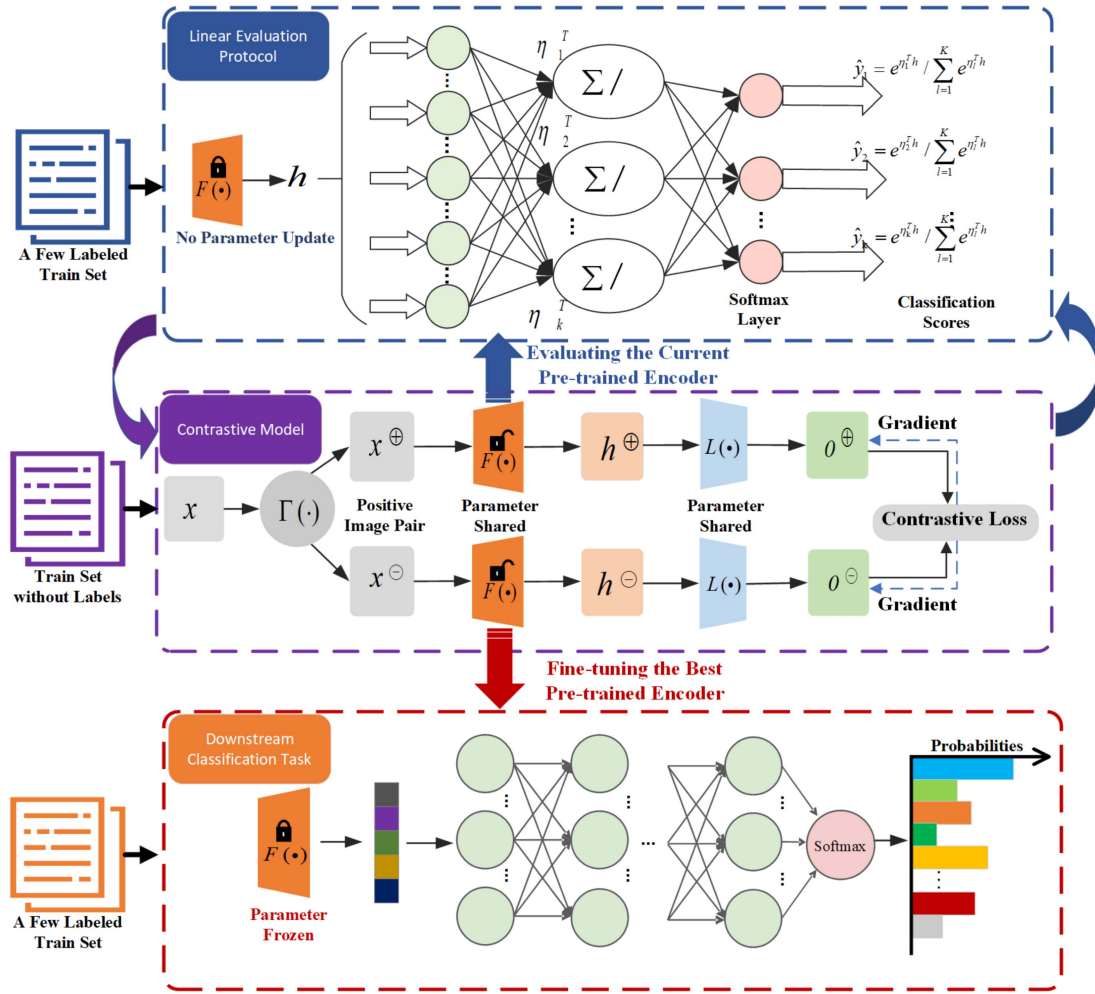


Fig. 1. Block diagram of the proposed two-stage training framework for SAR target classification models.

discriminability of current learned SSRs. In the fine-tuning stage, the best pretrained encoder determined by the linear evaluation protocol will be applied to fine-tune a SAR target classifier with a small number of annotated samples. The detailed block diagram of the proposed two-stage algorithm is demonstrated in Fig. 1.

#### A. Pretraining the Network With CL

Generally, supervised learning has a clear training purpose or task, but the SSL needs a pretext to motivate the model to learn meaningful representations from an unlabeled dataset. Compared with supervised classification tasks that learned the category discrimination with the assistance of category labels, we set the instance discrimination [31] as the pretext task, which means learning to discriminate between individual instances without any semantic categories. Hence, based on the instance discrimination pretext task, a CL paradigm with three modules can be designed. 1) The augmentation module, which can provide two different views of one raw sample; 2) The CNN-based encoder module is pretrained with a contrastive strategy to learn

discriminative representations; 3) The nonlinear module, which can project all the representations into a contrastive space, where the contrast loss can be calculated with the projected samples. Besides, some preprocessing of SAR images is performed, including despeckling, normalization, clip max, etc., before the raw SAR images are input into the above modules. The main components of the CL model applied in this article are illustrated in Fig. 2.

*A Stochastic Data Augmentation Module  $\Gamma(\cdot)$ :* Since many researchers have found that various complex data augmentation benefits CL [38]. We adopt various image transformation methods that can be applied to one-channel SAR images. The transform is formed with random cropping followed by resizing back to the original size, random horizontal flip, random grayscale, random color jitter and random Gaussian blur. The augmentation module can transform any SAR images into two random augmented images, which can note as a positive pair  $x^{\oplus}$  and  $x^{\ominus}$ . Meanwhile, other transformed SAR images can be viewed as negative samples against the positive pair  $\{x^{\oplus}, x^{\ominus}\}$ .

*A CNN Based Feature Encoder  $F(\cdot)$ :* The CNN encoder can project the augmented SAR images into an embedding space

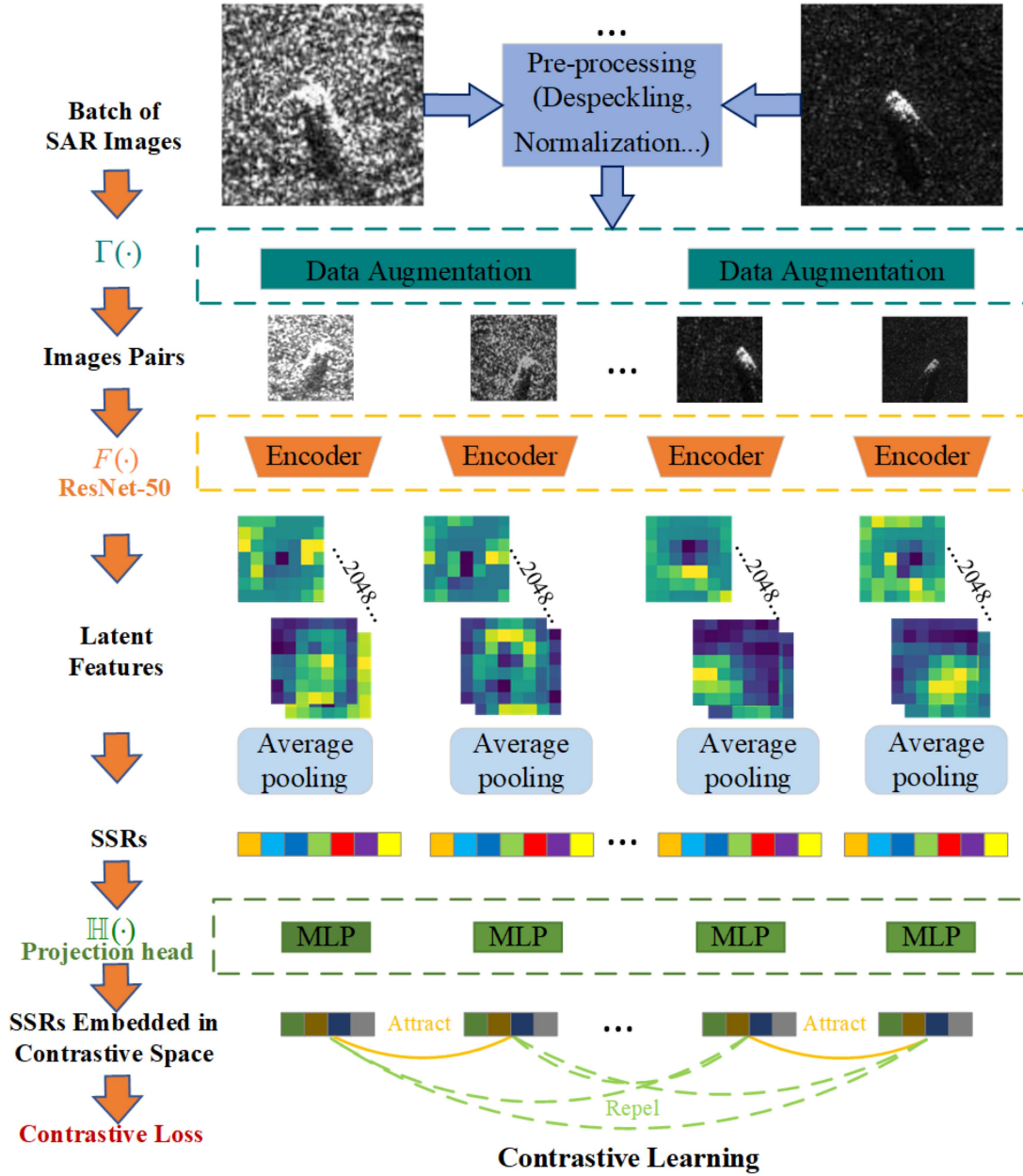


Fig. 2. Detailed structure of the CL model.

as latent representation  $h^{\oplus}$  and  $h^{\ominus}$ . As known that deep neural network (DNN) can obtain more generalized and high-level semantic representations, when there are enough data for training. Though various complex data augmentation for CL can provide more data patterns to help adequately activate the parameters in the neural network. The DNNs will also degrade due to the gradient exploding or vanishing. Therefore, we use the ResNet50 [39] as a CNN encoder to address the degradation problem. ResNet stacks the residual blocks to build a deep network. Meanwhile, the residual blocks ensure that the model will not degrade during the deepening by at least learning identity mapping. In addition, ResNet50 is more profound and has more channels in the same block layer compared with ResNet18 and ResNet34. Because of concerns about the training time we can afford, a

“bottleneck” layer is added to the vanilla residual block to reduce the additional parameters associated with the increased number of layers and channels. Fig. 3 shows the bottleneck residual block in ResNet50 when inputting the output  $x_l \in \mathbb{R}^{H_l \times W_l \times C_l}$  of the previous layer or layers, assuming that  $\mathcal{F}(x_l)$  indicates the transformed intermediate features of current layers, then the residue block can be represented as  $\mathcal{F}(x_l) + x_l$  before the activation layer  $\delta$  by shortcut connection operation. However, in practice,  $\mathcal{F}(x_l)$  may have different channels with  $x_l$ , when a  $1 \times 1$  convolution layer  $s(\cdot)$  should be used to reshape  $x_l$  before the element-wise add operation. Meanwhile, a stack of three layers with  $1 \times 1$ ,  $3 \times 3$  and  $1 \times 1$  convolutions build a bottleneck block in the above procedure. The first  $1 \times 1$  convolution is responsible for reducing the channel dimension, and the second



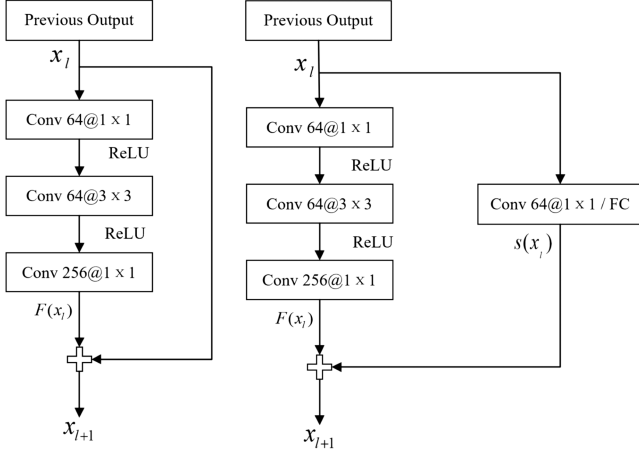


Fig. 3. Bottleneck residual blocks in ResNet50. The right figure shows the saturation when  $x_l$  has different shape with  $F(x_l)$ .

can restore the dimensions, which can leave the  $3 \times 3$  layer a bottleneck with smaller input/output dimensions. The above processing can be given as

$$x_{l+1} = \begin{cases} \delta(\mathcal{F}(x_l) + s(x_l)) & C_{x_l} \neq C_{\mathcal{F}(x_l)} \\ \delta(\mathcal{F}(x_l) + x_l) & C_{x_l} \equiv C_{\mathcal{F}(x_l)} \end{cases} \quad (1)$$

in which  $C_{x_l}$  and  $C_{\mathcal{F}(x_l)}$  denote the number of channels for intermediate feature map  $x_l$  and  $\mathcal{F}(x_l)$ , respectively.

A *Nonlinear Projection Module*  $L(\cdot)$ : Chen et al. [32] find that a nonlinear projection head improves the representation quality of the layer before it. In this work, this module is simply structured by an MLP with one hidden layer and one ReLU nonlinearity layer, which is formulated as

$$o = L(h) = W^{(2)} \text{ReLU}(W^{(1)}h) \quad (2)$$

where  $W^{(1)}$  and  $W^{(2)}$  represent the weight parameters of the two fully connected layers, respectively.

Based on the three above modules, a CL training procedure can be established. When a mini-batch of  $N$  original SAR images  $\{S\}_i^N$  pass through the data augmentation module. One positive pair is generated by the data augmentation module accompanied with  $2(N-1)$  negative pairs. Then, the  $2N$  pair-labeled samples are encoded into an embedding space as latent features (3), which can also be called as SSRs in this article

$$H = \{h_i = (F \circ \Gamma(s_i)) | s_i \in S\} \quad (3)$$

where  $h_i = (h_i^{\oplus}, h_i^{\ominus})$ .  $h_i^{\oplus}$  and  $h_i^{\ominus}$  denote the encoded SSRs for positive pair  $(x_i^{\oplus}, x_i^{\ominus})$ .

At last, the nonlinear projection module  $L(\cdot)$  will project SSRs into a same contrastive space with the same dimension as

$$O = \{o_i = (L(h_i^{\oplus}), L(h_i^{\ominus})) | h_i \in H\} \quad (4)$$

where  $o_i = (o_i^{\oplus}, o_i^{\ominus})$ .  $o_i^{\oplus}$  and  $o_i^{\ominus}$  are obtained from the encoded SSR pair  $(h_i^{\oplus}, h_i^{\ominus})$ , which are processed by the nonlinear projection module  $L(\cdot)$ .

Then, in the contrastive space, the similarity between positive pair vector  $o_i^{\oplus}$  and vector  $o_i^{\ominus}$  against other negative samples can be measured by Info-NCE loss [33] formed as (5). It means that

given a nonlinear coding vector  $o_i^{\oplus}$ , we want to query the most similar SSR  $o_i^{\ominus}$  in the contrastive space

$$\ell_i^{\oplus} = -\log \frac{\exp(\text{Cosim}(o_i^{\oplus}, o_i^{\ominus})/\tau)}{\sum_{k=1}^N \mathbb{I}_{[\hat{o}_k \neq o_i^{\oplus}]} \exp(\text{Cosim}(o_i^{\oplus}, \hat{o}_k)/\tau)}. \quad (5)$$

Among the contrastive loss (5),  $\tau$  is an adjustable temperature parameter that can be used to control the distribution shape of (5) and the discrimination of the model to negative samples;  $\mathbb{I}_{[\hat{o}_k \neq o_i^{\oplus}]}$  is an indicator function, and the value is 1 only if  $\hat{o}_k$  is different from  $o_i^{\oplus}$ , otherwise, it is 0. In Info-NCE loss, the instance-wise discrimination can be regarded as the similarity or distance between two features generated by the nonlinear projection module. We use the cosine similarity formulated as (6) to measure the distance, where the vectors  $u, v$  are Euclidean normalized.

$$\text{Cosim}(u, v) = \frac{u \cdot v}{|u| * |v|}. \quad (6)$$

Therefore, for mini-batch of  $N$  original SAR images  $\{S\}_{i=1}^N$ , the batch loss can be calculate by

$$\mathcal{L} = \frac{1}{2N} \sum_{i=1}^N (\ell_i^{\oplus} + \ell_i^{\ominus}). \quad (7)$$

### B. Evaluation of the Pretrained Encoder

When pretraining the encoder with the CL model, the objective of the encoder  $F(\cdot)$  is to project the original SAR images into a discriminative embedding space as SSRs, which is agnostic to the objective of downstream SAR target classifier. To monitor the process of CL, we proposed a linear classification evaluation protocol to evaluate the encoder in each training epoch. It applies partial currently learned SSRs and the corresponding class labels to fit a linear classifier. Then, quantitatively evaluating the linear classifier with some classification metrics that can represent the efficiency of the current encoder. It should be noted that the SSL process requires no labeled samples, but a few labeled samples are necessary for ensuring the optimal performance of the pretrained encoder by linear evaluation protocol.

In detail, for each training epoch, given a train set of  $N$  unlabeled samples  $\mathbb{D}_t = \{x_n\}_{n=1}^N$ , and a few labeled train set of  $M$  samples  $\mathbb{D}_t = \{x_m, y_m\}_{m=1}^M$ , where  $y_m \in \{1, 2, 3, \dots, k\}$  and  $M \leq N$ . The weight parameters  $\zeta^*$  of current CNN encoder  $F(\cdot)$  pretrained by CL model can be given as

$$\zeta^* = \arg \min_{\zeta} \frac{1}{N} \sum_{n=1}^N \mathcal{L}[F(x_n; \zeta)]. \quad (8)$$

Then, we can use  $F(x_m; \zeta^*)$  to denote the current learned SSRs, and a multinomial logistic regression classifier [40] of  $K$  categories is used to estimate the probability that  $F(x_m; \zeta^*)$  belongs to each category, which can be given as

$$p(y_m = k | F(x_m; \zeta^*); \eta) = \frac{e^{\eta_k^T F(x_m; \zeta^*)}}{\sum_{l=1}^k e^{\eta_l^T F(x_m; \zeta^*)}}. \quad (9)$$

Moreover, the logical probability of sample  $(x_m, y_m)$  shows as follows:

$$h_\eta(F(x_m; \zeta^*)) = \begin{bmatrix} p(y_m = 1 | F(x_m; \zeta^*); \eta) \\ p(y_m = 2 | F(x_m; \zeta^*); \eta) \\ \vdots \\ p(y_m = k | F(x_m; \zeta^*); \eta) \end{bmatrix} = \frac{1}{\sum_{l=1}^k e^{\eta_l^T F(x_m; \zeta^*)}} \begin{bmatrix} e^{\eta_1^T F(x_m; \zeta^*)} \\ e^{\eta_2^T F(x_m; \zeta^*)} \\ \vdots \\ e^{\eta_k^T F(x_m; \zeta^*)} \end{bmatrix} \quad (10)$$

where  $\eta$  denote the learnable parameters of the multinomial logistic regression classifier. Therefore, for total labeled train set, the  $K$  categories logistic regression classifier can be fit with

$$\eta^* = \arg \max \frac{1}{M} \left[ \sum_{m=1}^M \sum_{k=1}^K \mathcal{I} \log \frac{e^{\eta_k^T F(x_m; \zeta^*)}}{\sum_{l=1}^k e^{\eta_l^T F(x_m; \zeta^*)}} \right]$$

$$\mathcal{I}(y_m, k) = \begin{cases} 1, & y_m = k \\ 0, & y_m \neq k. \end{cases} \quad (11)$$

Finally, with the solution of  $\zeta^*$  and  $\eta^*$ , the overall classification accuracy in the test set can be used as metric for assessing the current epoch learned SSRs.

### C. Fine-Tuning the Classifier Designed for SAR Target

Fine-tuning is a prevalent technique for transfer learning, which aims to transfer knowledge learned by pretrained models using downstream tasks and target domain datasets. In our work, the encoder is pretrained by the proposed CL model on unlabeled source domain samples. Then, in downstream tasks, it will be fine-tuned with a few labeled samples. In the following, we illustrate how to use the pretrained encoder during the fine-tuning stage.

When the CL model has been trained for multiple epochs, the linear evaluation protocol is adopted to evaluate the performance of the encoder. Then, the optimal CNN encoder  $F(\cdot)$  can be determined, and the encoder's parameters  $\zeta^*$  are frozen in the subsequent procedure. Consequently, given a small annotated dataset  $\mathbb{D}_a = \{x_i, y_i\}$ , where the samples can be encoded as a labeled representation set  $R = \{r_i = (F(x_i), y_i) | x_i \in \mathbb{D}_a\}$ . Then, an MLP-based classification head  $\mathbb{H}(\cdot)$  is directly applied on this labeled representation set to fine-tune the encoder using downstream SAR target classification task and the representation set  $R$ , which can be expressed as follows:

$$\chi^* = \arg \min \frac{1}{N} \sum_{i=1}^N \mathcal{L}_{\text{sup}} \{\mathbb{H}(r_i; \chi), y_i\}. \quad (12)$$

Here,  $\chi$  indicates the learned parameters for classification head  $\mathbb{H}$ ;  $\mathcal{L}_{\text{sup}}$  indicates the supervised loss function for downstream tasks. Especially, it is cross-entropy loss for SAR image classification tasks.

The detailed architecture of the MLP-based  $\mathbb{H}(\cdot)$  shows in Table I. We stack two Linear-BatchNorm-ReLU modules and add a linear layer behind them to convert the number of output

TABLE I  
STRUCTURE OF THE MLP-BASED CLASSIFICATION HEAD  $\mathbb{H}(\cdot)$

Layer	Input	Linear	BN2D	ReLU	Linear
Params	B, 2048	2048, 128	128	–	128, 64
Layer	BN2D	ReLU	Linear	Softmax	Output
Params	64	–	64, $n_{\text{class}}$	–	B, $n_{\text{class}}$

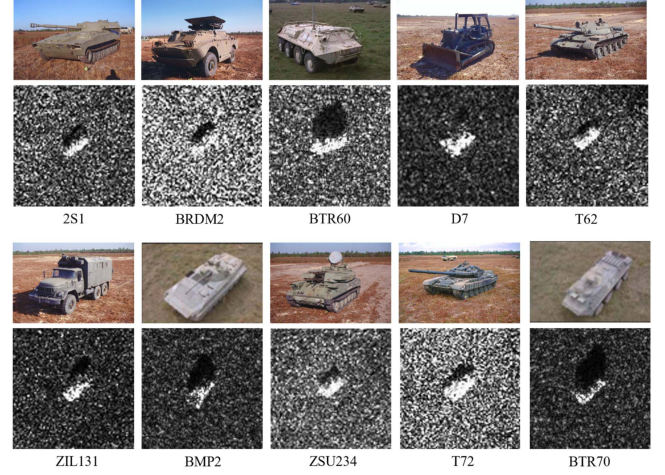


Fig. 4. Military ground targets in the MSTAR dataset. SAR images (bottom) and corresponding optical images (top).

channels to the number of categories. Then, following by a Softmax layer to predict the classification scores for input sample  $x_i$ . The shape of the input batch of average pooled SSRs is (B, 2048). The first linear module's output units are 128, and the second linear module's output units are 64. Finally,  $n_{\text{class}}$  output units for the last classification layer.

## III. EXPERIMENTAL RESULTS

### A. Experiment Setup

1) *Dataset*: The experiments about the SAR target classifier's performance and linear evaluation protocol are carried out on moving and stationary target acquisition and recognition (MSTAR) [41] database, which is collected by a 10 GHz SAR platform with  $0.3 \times 0.3$  m resolution. It published thousands of SAR images for 10 military ground targets, including tank: BMP2, T72, T62; armored vehicle: BTR70, BTR60; truck: BRDM, ZIL131; cannon: 2S1, ZSU234 and bulldozer D7. Fig. 4 shows the samples of 10 targets, including SAR and corresponding optical images. In addition, there are  $15^\circ$  and  $17^\circ$  two depressing angles for each target, and all of them are full aspect coverage (in the range of  $0^\circ$  to  $360^\circ$ ). Similar to [22], we applied the MSTAR 10 class targets classification benchmark, in which all of the data with  $17^\circ$  depressing angle are divided as the train set  $\mathbb{D}_{\text{train}}$ , and other data with  $15^\circ$  depressing angle are divided as the test set  $\mathbb{D}_{\text{test}}$ . Although  $\mathbb{D}_{\text{train}}$  and  $\mathbb{D}_{\text{test}}$  are all annotated, when pretraining the encoder with CL, just using  $\mathbb{D}_{\text{train}}$  without category labels, while full labeled  $\mathbb{D}_{\text{train}}$  during the linear evaluation. Table II shows the number of available SAR

TABLE II  
AVAILABLE SAR IMAGES OF THE TRAIN SET(17°) AND THE TEST SET(15°) IN THE MSTAR DATASET

DataSet	Tank		BMP2	Armored Vehicle		Truck		Cannon		Bulldozer D7
	T72	T62		BTR70	BTR60	BRDM2	ZIL131	2S1	ZSU234	
Train(17°)	232	299	233	233	256	298	299	299	299	299
Test(15°)	196	273	194	196	195	274	274	274	274	274

TABLE III  
AVAILABLE SAR SHIP IMAGES OF THE TRAIN SET AND THE TEST SET IN THE  
OPENSARSHIP DATASET

	Cargo	Bulk Carrier	Container Ship
Train	788	393	533
Test	337	169	229

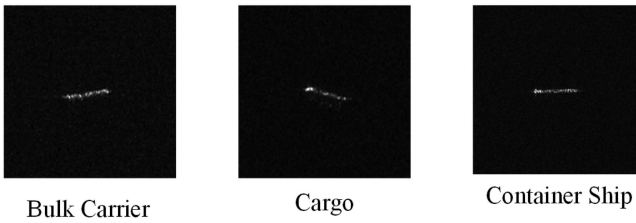


Fig. 5. Three types of SAR ship in the prepared openSARship dataset.

images of the train set and the test set with different depress angles.

After pretraining the ResNet50-based encoder on the MSTAR dataset, a special knowledge transfer experiment is also set up on another dataset named openSARship [42] to verify the transferability of the SSRs extracted by the encoder. The openSARship dataset is a well-organized shared dataset collected from the Sentinel-1 satellite and widely used for marine surveillance. The publisher has provided thousands of SAR ship target images and corresponding ground truth, including tens of categories, while also having serious category imbalance issues (8470 in cargo but 4 in towing). Therefore, only the three most numerous ship types, including cargo, bulk carrier, and container ship, are used to form the classification dataset, and only the VH polarization and ground range detected (GRD) products are sampled to compose the specific openSARship dataset. Finally, we randomly split the dataset into a train set and a test set in a ratio of 7:3. Table III shows the number of samples in each category, and Fig. 5 illustrates the three classes of SAR ships.

2) *Hyperparameters and Metrics*: In this article, only the pretraining stage is trained on two NVIDIA RTX 3090 GPUs with the PyTorch DL framework on an Ubuntu 20.04 Linux system, and one NVIDIA RTX 3090 GPU is used for other experiments, when training the classifier with the proposed two-stage algorithm mentioned in Section II-C. In the pretraining stage, the ResNet50-based encoder  $F(\cdot)$  is self-supervised per-trained by the proposed CL model. In the data argumentation module  $\Gamma(\cdot)$ , the original SAR images are first resized to  $158 * 158$  and randomly cropped back to  $128 * 128$ . After that, three augmentation policies are applied: random flip, grayscale, and Gaussian blur. Then, we set the epochs to 1000 and batch size

to 256 (128 per GPU), using the Adam optimizer with  $10^{-3}$  initial learning rate and  $10^{-6}$  weight decay. In addition, since the CL model is very unstable at the beginning of training, we apply CosineAnnealing learning rate scheduler [43] with 100 epoch period to overcome it. Finally, we trained the CL model with Info-NCE loss which temperature  $\tau$  is 0.2. During the CL model training period, we evaluated the current learned SSRs with the linear evaluation protocol mentioned in Section II-B for each epoch. In the fine-tuning stage, the learnable parameters of ResNet50 encoder  $F(\cdot)$  are frozen for the downstream SAR target classification task. We just trained the classification head  $\mathbb{H}(\cdot)$  with cross-entropy loss and Adam optimizer, meanwhile, setting the epochs to 100, batch size to 64 and initial learning rate to 0.01 with OneCycle learning rate scheduler.

In the experiment on fine-tuning the pretrained encoder, we use openSARship to transfer the knowledge learned from MSTAR dataset into a new dataset, and evaluate the efficiency of knowledge transferring compared with training on openSARship dataset from scratch. When training with fine-tuning scheme, we use the OneCycle learning rate scheduler with maximum learning rate of  $10^{-4}$ , but  $10^{-3}$  for training from scratch. Then, keeping the other hyperparameters be the same. For example, training models with cross-entropy loss and Adam optimizer, and setting the epochs to 200, batch size to 64. In addition, no data enhancements are applied during the experiments.

Since all experiments are related to classification tasks, two common classification evaluation indicators, including overall accuracy (OA) and average accuracy (AA) are adopted in this article. OA refers to the percentage of all test samples that are correctly categorized. AA refers to the average accuracy for each category

## B. Performance of the Proposed Pretrained Network

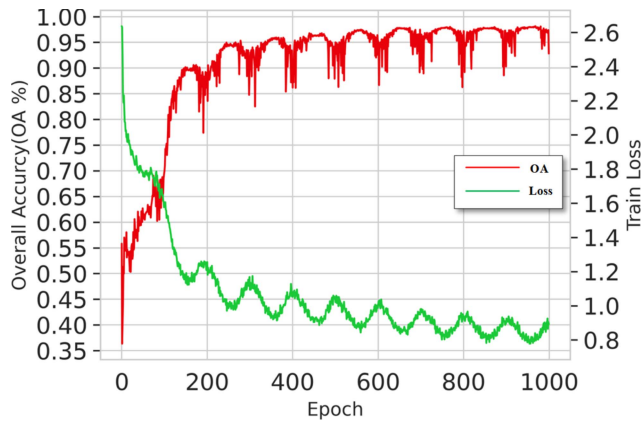
Here, we design a series of experiments in the pretraining stage. 1). The training states of the CL model are illustrated by plotting the CL model's training loss and the LR classifier's OA at each epoch; 2). In addition to LR, three other linear classifiers, including KNN, DT, and SVM, are considered for testing the efficacy of various linear classifiers used in the linear evaluation protocol; 3). T-Stochastic Neighbor Embedding (T-SNE) algorithm [44] is utilized to visualize the distribution of SSRs in order to evaluate the separability and discriminability of SSRs in the embedding space.

1) *Linear Evaluation*: In the pretraining stage, the whole train set  $\mathbb{D}_{\text{train}}$  without category labels is employed to train the CL model proposed in Section II-A. For each training epoch, the encoder  $F(\cdot)$  in current epoch can transform the whole  $\mathbb{D}_{\text{train}}$  into SSRs. Then, the logistical regression classifier can be fitted with

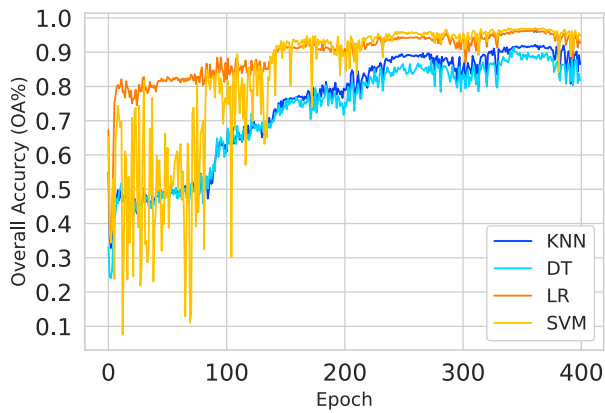


TABLE IV  
DETAILED CLASSIFICATION RESULTS ON THE TEST SET OF THE MSTAR DATASET

Category	T72	T62	BMP2	BTR70	BTR60	BRDM	ZIL131	2S1	ZSU234	D7	AA(%)
T72	194	0	2	0	0	0	0	0	0	0	98.98
T62	0	270	0	0	0	1	0	1	1	0	98.90
BMP2	1	0	192	1	0	0	0	0	0	0	98.97
BTR70	0	0	0	196	0	0	0	0	0	0	100
BTR60	3	3	0	0	185	0	0	4	0	0	94.87
BRDM	0	0	0	0	0	269	0	1	3	1	98.17
ZIL131	0	0	0	0	0	0	273	0	0	0	99.64
2S1	0	3	0	0	1	1	3	266	0	0	97.08
ZSU234	0	0	0	0	0	0	0	0	273	1	99.64
D7	0	1	0	0	0	0	2	0	2	269	98.17



(a)



(b)

Fig. 6. Training states in the pretraining stage. (a) The train loss and OA curves during the pretraining stage. (b) OA curves of four different types of linear classifier.

the SSRs and category labels in  $\mathbb{D}_{\text{train}}$ . Finally, the fitted classifier can assess the performance of the pretrained network on the  $\mathbb{D}_{\text{test}}$  using OA and AA metrics. Fig. 6(a) plots the train loss and OA in linear evaluation protocol for each epoch. In the figure, the best encoder for the subsequent fine-tuning stage is obtained in epoch 968, in which the linear evaluation has achieved the highest OA of 98.47%. Meanwhile, the AA of the best logistical regression classifier and detailed classification result is shown in Table IV. The result indicates that when the Info-NCE loss decreases,

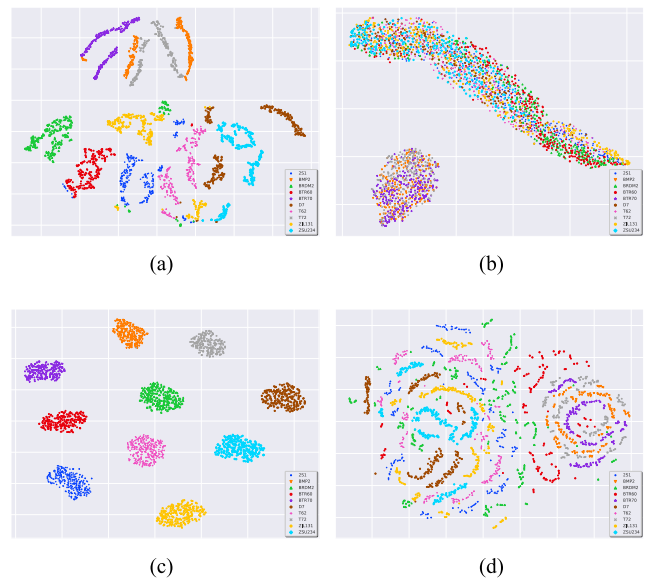


Fig. 7. Results of T-SNE visualization for different feature extractors. (a) The optimal encoder. (b) The randomly initialized encoder. (c) Classification head  $\mathbb{H}(\cdot)$  in the fine-tuning stage. (d) PCA.

better SSRs are learned, meanwhile, the OA of the logistical regression classifier is gradually improved. It demonstrates that the linear evaluation protocol proposed in Section II-B can provide guidelines for evaluating the pretrained encoder's performance in the process of SSL.

2) *Different Linear Classifiers*: This section examines four types of classifiers: LR, KNN, DT, and SVM. The four classifiers listed above are tested for 1000 epochs. Fig. 7 shows the OA-Epoch curve for four different linear classifiers at the first 400 epochs, and the detailed metrics for four kinds of classifiers in epoch 968 are listed in Table V. It is clear that the OA of LR and SVM classifiers are obviously better than KNN and DT classifiers with more than 2%–5% improvement, while in the initial stage of training the CL model, the performance of the SVM classifier fluctuates more than the LR classifier. It implies that the multinomial logistic regression classifier would be more appropriate as a linear evaluator in the pretraining stage.

3) *Visualization With T-SNE*: In this experiment, the best encoder  $F(\cdot)$  in the pretraining stage is compared with several feature extractors, including randomly initialized encoder



TABLE V  
CLASSIFICATION RESULTS OF FOUR KINDS OF LINEAR CLASSIFIER IN EPOCH 968

	T72	T62	BMP2	BTR70	BTR60	BRDM2	ZIL131	2S1	ZSU234	D7	OA(%)
KNN	98.97	94.55	97.95	99.49	96.84	99.25	96.07	96.26	98.52	93.03	96.91
DT	95.43	92.42	94.71	94.50	91.49	94.66	92.41	89.24	88.04	96.73	92.86
SVM	97.94	98.50	99.44	100.0	100.0	99.16	98.87	98.49	98.51	99.58	98.55
LR	97.98	97.47	98.97	99.49	99.46	99.26	98.20	97.79	97.84	98.90	98.47

$\hat{F}(\cdot)$ , proposed classification head  $\mathbb{H}(\cdot)$  in the fine-tuning stage and PCA. We use the T-SNE algorithm to reduce the high-dimensional feature vectors to two dimensions and visualize their distributions. Fig. 7(a) and (b) demonstrates that with the iterative training of the CL model, samples with the same category can automatically gather together and separate from the rest in the embedding space. Comparing Fig. 7(a) with Fig. 7(d) indicates that the encoder pretrained by the CL model can provide better interclass separability in the embedding space than the conventional machine learning method (PCA). Fig. 7(c) shows that the discrimination of the previously preextracted SSRs in the feature space will be further improved after the bootstrapping of the downstream supervised classification task. The above results show that the pretrained encoder is capable of embedding raw SAR images into a more discriminative feature space, thus facilitating the convergence of the downstream SAR target classifier during the fine-tuning stage.

### C. Fine-Tuning Using Limited Annotated Data

In the fine-tuning experiments, the parameters of the best encoder obtained in the pretraining stage is fixed. To begin with, we evaluated its performance under a few labeled  $\mathbb{D}_{\text{train}}$  in MSTAR. The results show that our proposed algorithm can achieve an optimal OA of 90.71% when only samples 10% labeled  $\mathbb{D}_{\text{train}}$  and 99.34% when samples 30%. Fig. 8 shows the above classification results with a confusion matrix. Each row is the true category, and each column is the predicted category. In addition, the elements of the confusion matrix represent the number of targets identified as a certain category, and the elements on the diagonal line represent the number of true identifications of a category. It shows that our fine-tuned classifier performs comparably to the state-of-the-art (SOTA) supervised model in the MSTAR 10 targets classification benchmark, even when the available labeled dataset is constrained.

To further demonstrate the efficiency of the proposed fine-tuning approach, several SAR targets classification methods are compared with the proposed two-stage algorithm, including PCA with an LR classifier, vanilla ResNet50 [39] and A-ConvNet [22]. In the PCA-based method, the raw SAR images are firstly flattened and then reduced its dimensionality to 128 using the PCA algorithm. The 128-dimensional vectors are then used to train an LR classifier comparable to the one proposed; In the vanilla ResNet50 model, just adding a linear layer to reshape the number of output channels to 10 categories; For A-ConvNet, only reshaping the raw SAR images as the model required, and do not use the enhancement methods mentioned in the model's paper. Fig. 9 shows the OA for the proposed classification model and other aforementioned methods trained with 10%, 20%, 30%,

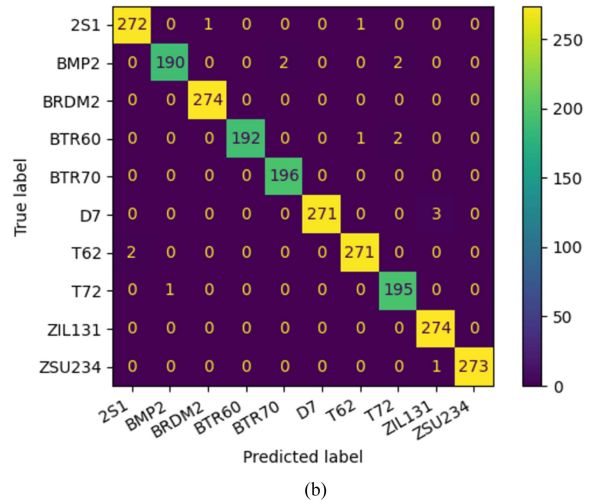
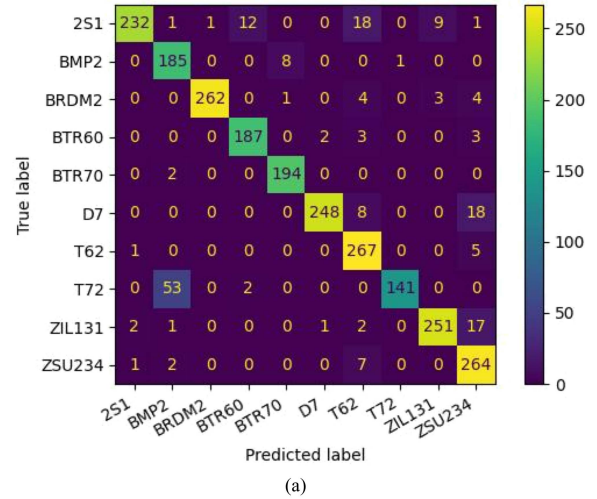


Fig. 8. Confusion matrix for fine-tuning experiments in MSTAR dataset. (a) 10% train set. (b) 30% train set.

50%, and 100% annotated  $\mathbb{D}_{\text{train}}$ . The result shows that the PCA-based approach can work better than Vanilla ResNet50 and A-ConvNet under very few(10%) labeled datasets. However, as the size of the annotated train set increases, the performance of supervised models based on DL can significantly outperform traditional methods (PCA). In contrast to the above approaches, the proposed two-stage algorithm using SSRs is better than other comparison methods under any ratio of the labeled train set. Especially our proposed algorithm can achieve the best OA of 99.34% with 30% of the labeled train set, which is comparable to the performance of other methods under 50% and 100% of

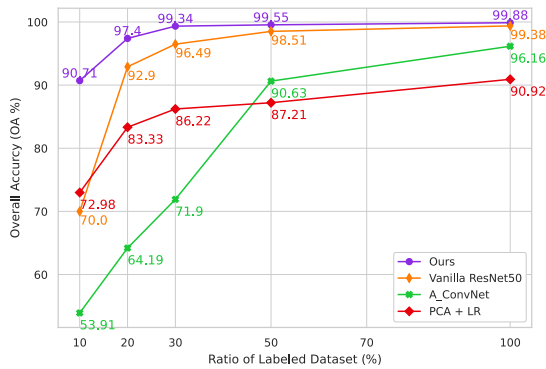


Fig. 9. OA for different SAR classification models under various train set ratio.

the train set. The above analysis indicates that our proposed algorithm has very significant advantages in insufficiency of annotated samples. Meanwhile, the classification performance can also be rapidly improved by slightly increasing the proportion of annotated train set.

#### D. Knowledge Transfer Under Different Datasets

In the experiments of this section, we attempt to transfer the knowledge learned from MSTAR dataset to training a openSARship classifier. Some classifiers are trained on the prepared openSARship dataset under the following different settings.

- 1) Supervised training of the classifier from scratch with ResNet18.
- 2) Supervised training of the classifier from scratch with ResNet50.
- 3) Training the classifier with ResNet50 pretrained by the CL model in Section III-B on the  $\mathbb{D}_{\text{train}}$  in MSTAR. In addition, the pretrained ResNet50-based encoder is frozen during the knowledge transfer experiments.
- 4) Training the classifier with ResNet50 pretrained by the CL model in Section III-B on the  $\mathbb{D}_{\text{train}}$  in MSTAR. But the pretrained ResNet50-based encoder isn't fixed during the knowledge transfer experiments.

In addition, the class head behind the pretrained ResNet50-based encoder is identical to Table I, except that the output units of the softmax layer are three. Meanwhile, the class layer in ResNet18 and ResNet50 is replaced with the abovementioned class head to maintain fairness.

Fig. 10 shows the evaluation results when training the above four ship classifiers. When supervised training from scratch, the trends of the evaluation curve during model training are almost similar, and the best OA for ResNet18 and ResNet50 are quite close, which are 71.17% and 70.72%, respectively. However, fine-tuning with the pretrained encoder can achieve the best OA of 77.03% and a smoother and faster training state than training from scratch. It indicates that training a classifier on the openSARship dataset with the pretrained encoder is more easier than training from scratch, even though the encoder is pretrained using the MSTAR dataset. Combining this experiment result with the above experiments in Section III-C, it seems that compared with supervised learning models that tend to capture the

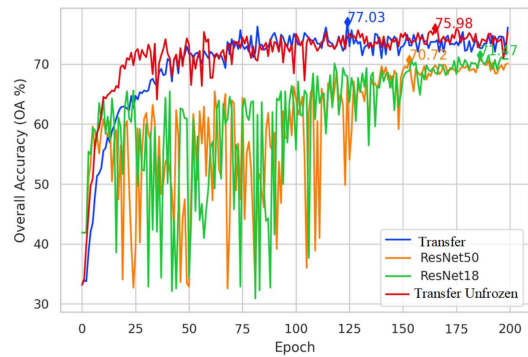


Fig. 10. Training states for different SAR classification models in openSARship dataset.

class similarities among samples driven by category labels, the CL model based on instance discrimination pretext task attempts to pretrain an encoder to grasp the instance similarities among samples so that the encoded features can be more generalized and high-level representations and make the knowledge transfer between different datasets easier.

#### E. Compare With SOTA Methods

This section compares some SOTA SAR target classification methods with our proposed algorithm. We unify the encoder as ResNet50 and compare their classification accuracy using 10% and 30% randomly sampled labeled MSTAR dataset. Besides, SAR target classifiers are trained without data augmentation. The SOTA and baseline methods used for comparison are listed as follows.

- 1) *Baseline methods*: Three traditional supervised methods are selected as baselines, including feature extraction-based methods: PCA and wavelet energy [45]. supervised CNN-based model A-ConvNet [22]. We use a three-layer Haar wavelet transform to extract normalized wavelet coefficients from SAR images. The energy, kurtosis, and skewness of nine high-frequency subgraphs were used to form a 27-dimensional feature vector for training a SVM classifier.
- 2) *SAR domain transfer learning methods*: One attempts to learn knowledge from unlabeled SAR images by reconstructing these images [26]. We have replicated the experiment by reconstructing the MSTAR dataset. Another attempts to learn knowledge from simulating SAR images [27]. We borrowed their experimental results.
- 3) *SSL methods*: InstDisc [31] and MOCO [33]. We use these method to pretrain a ResNet50 with 1000 epochs. Then fine-tuning the ResNet50 with 100 epochs. Since InstDisc and MOCO do not use a linear evaluation protocol to determine the best pretrained encoder. In the fine-tuning stage, the encoder trained in the last pretraining epoch is applied to train a SAR target classifier.

The experimental results shown in Table VI prove the superiority of our proposed algorithm on SAR target classification tasks when labeled SAR images are scarce. It also demonstrates

TABLE VI  
OA OF SOTA METHODS TRAINED ON 10% AND 30% LABELED MSTAR DATASET

Methods	Overall Accuracy (OA%)	
	10%	30%
Supervised Methods		
PCA+LR	72.98	86.22
Wavelet+SVM [45]	45.28	51.49
A-ConvNet [22]	53.91	71.90
SAR Domain Transfer Learning Methods		
Huang et al.(Reconstruct) [26]	<b>91.13</b>	93.73
Wang et al.(Simulate) [27]	88.90	95.00
Self-supervised Learning Methods		
InstDisc [31]	85.85	92.37
MOCO [33]	86.82	96.87
Ours	90.71	<b>99.34</b>

that our two-stage method is more efficient than other self-supervised methods. This further emphasizes the importance and practicality of our proposed additional linear evaluation protocol.

#### IV. CONCLUSION

In this article, a contrasting strategy is proposed for learning SSRs. Based on this, a two-stage algorithm is designed to train a SAR image target classifier under the constraints of the annotated dataset. In the pretraining stage, a CL model has investigated to pretrain a CNN-based encoder with an unlabeled train set. This encoder can transfer the raw SAR images into a discriminative embedding space in which the samples with the same category label can be automatically clustered together and separated from other types of targets. In the fine-tuning stage, the classifier can be adequately trained with a few transformed SSRs, and corresponding labels because the SSRs already have some discriminatory properties. In addition, we also apply a linear evaluation protocol to indicate the CL model training state in each training epoch. It can quantitatively evaluate the performance of the pretrained CNN encoder and the corresponding transformed SSRs, which can indirectly indicate the merit of the downstream classifier trained in the stage of fine-tuning. For SAR target classification tasks, experimental results show that our proposed algorithm achieves SOTA quantitative results with an accuracy of 90.71% when sampling only 10% of the labeled dataset and 99.34% when sampling 30% on the MSTAR dataset. Meanwhile, it achieves the best performance and a smoother, faster training state in knowledge transfer experiments, where MSTAR learned knowledge is transferred to the openSARship dataset. On the one hand, it shows that our proposed algorithm has excellent potential in small sample learning and SSR learning. On the other hand, this suggests that the proposed pretraining strategy can take full advantage of many unlabeled SAR datasets to pretrain a SAR image feature extractor with high generalization performance. In future work, based on the contrasting strategy, we will attempt to pretrain a more generalized feature extractor using vast quantities of unlabeled SAR images

and then transfer it to classification, detection, segmentation, tracking, and other tasks.

#### ACKNOWLEDGMENT

The authors would like to thank the anonymous editors and reviewers for their constructive suggestions, which have helped improve the quality of this article.

#### REFERENCES

- [1] J. Chen, H. Yu, G. Xu, J. Zhang, B. Liang, and D. Yang, "Airborne SAR autofocus based on blurry imagery classification," *Remote Sens.*, vol. 13, no. 19, Sep. 2021, Art. no. 3872, doi: [10.3390/rs13193872](https://doi.org/10.3390/rs13193872).
- [2] B. Zhang, G. Xu, R. Zhou, H. Zhang, and W. Hong, "Multi-channel back-projection algorithm for mmwave automotive MIMO SAR imaging with doppler-division multiplexing," *IEEE J. Sel. Topics. Signal Process.*, vol. 17, no. 2, pp. 445–457, Mar. 2023, doi: [10.1109/JSTSP.2022.3207902](https://doi.org/10.1109/JSTSP.2022.3207902).
- [3] L. Zhang, G. Gao, D. Duan, X. Zhang, L. Yao, and J. Liu, "A novel detector for adaptive detection of weak and small ships in compact polarimetric SAR," *IEEE J. Miniat. Air Space Syst.*, vol. 3, no. 3, pp. 153–160, Sep. 2022, doi: [10.1109/JMASS.2022.3204772](https://doi.org/10.1109/JMASS.2022.3204772).
- [4] J. Chen and H. Yu, "Wide-beam SAR autofocus based on blind RS," *Sci. China Inf. Sci.*, vol. 66, no. 4, 2023, Art. no. 140304, doi: [10.1007/s11432-022-3574-7](https://doi.org/10.1007/s11432-022-3574-7).
- [5] J. Chen et al., "Blind NCS-Based autofocus for airborne wide-beam SAR imaging," *IEEE T. Comput. Imag.*, vol. 8, pp. 626–638, 2022, doi: [10.1109/TCI.2022.3194745](https://doi.org/10.1109/TCI.2022.3194745).
- [6] C. Zhang et al., "Performance evaluation of data enhancement methods in SAR ship detection," *IEEE J. Miniat. Air Space Syst.*, vol. 3, no. 4, pp. 249–255, Dec. 2022, doi: [10.1109/JMASS.2022.3211256](https://doi.org/10.1109/JMASS.2022.3211256).
- [7] S. Gao and H. Liu, "RetinaNet-Based compact polarization SAR ship detection," *IEEE J. Miniat. Air Space Syst.*, vol. 3, no. 3, pp. 146–152, Sep. 2022, doi: [10.1109/JMASS.2022.3203214](https://doi.org/10.1109/JMASS.2022.3203214).
- [8] Y. Zhao and M. Jiang, "Integration of optical and SAR imagery for dual PolSAR features optimization and land cover mapping," *IEEE J. Miniaturization Air Space Syst.*, vol. 3, no. 2, pp. 67–76, Jun. 2022, doi: [10.1109/JMASS.2022.3195955](https://doi.org/10.1109/JMASS.2022.3195955).
- [9] Z. Tirandaz, G. Akbarizadeh, and H. Kaabi, "PolSAR image segmentation based on feature extraction and data compression using weighted neighborhood filter bank and hidden Markov random field-expectation maximization," *Measurement*, vol. 153, Mar. 2020, Art. no. 107432, doi: [10.1016/j.measurement.2019.107432](https://doi.org/10.1016/j.measurement.2019.107432).
- [10] F. Mahmoudi Ghara, S. B. Shokouhi, and G. Akbarizadeh, "A new technique for segmentation of the oil spills from synthetic-aperture radar images using convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8834–8844, Oct. 2022, doi: [10.1109/JSTARS.2022.3213768](https://doi.org/10.1109/JSTARS.2022.3213768).
- [11] G. Xu, B. Zhang, H. Yu, J. Chen, M. Xing, and W. Hong, "Sparse synthetic aperture radar imaging from compressed sensing and machine learning: Theories, applications, and trends," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 4, pp. 32–69, Dec. 2022, doi: [10.1109/MGRS.2022.3218801](https://doi.org/10.1109/MGRS.2022.3218801).
- [12] A. K. Mishra, "Validation of PCA and LDA for SAR ATR," *IEEE Region 10 Conf.*, 2008, pp. 1–6, doi: [10.1109/TENCON.2008.4766807](https://doi.org/10.1109/TENCON.2008.4766807).
- [13] X. Liu, Y. Huang, J. Pei, and J. Yang, "Sample Discriminant Analysis for SAR ATR," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2120–2124, Dec. 2014, doi: [10.1109/LGRS.2014.2321164](https://doi.org/10.1109/LGRS.2014.2321164).
- [14] J. Geng, H. Wang, J. Fan, and X. Ma, "Deep supervised and contractive neural network for SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, pp. 2442–2459, Apr. 2017, doi: [10.1109/TGRS.2016.2645226](https://doi.org/10.1109/TGRS.2016.2645226).
- [15] Y. Xiang, F. Wang, and H. You, "OS-SIFT: A robust SIFT-Like algorithm for high-resolution Optical-to-SAR image registration in suburban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3078–3090, Jun. 2018, doi: [10.1109/TGRS.2018.2790483](https://doi.org/10.1109/TGRS.2018.2790483).
- [16] K. Ni, M. Zhai, Q. Wu, M. Zou, and P. Wang, "A wavelet-driven subspace basis learning network for high-resolution synthetic aperture radar image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1900–1913, 2023, doi: [10.1109/JSTARS.2023.3241944](https://doi.org/10.1109/JSTARS.2023.3241944).
- [17] J. Zhang, J. Chen, H. Yu, D. Yang, X. Xu, and M. Xing, "Learning an SAR image despeckling model via weighted sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7148–7158, Jul. 2021, doi: [10.1109/JSTARS.2021.3097119](https://doi.org/10.1109/JSTARS.2021.3097119).



- [18] H. Yu, T. Yang, L. Zhou, and Y. Wang, "PDNet: A lightweight deep convolutional neural network for InSAR phase denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5239309, doi: [10.1109/TGRS.2022.3224030](https://doi.org/10.1109/TGRS.2022.3224030).
- [19] S. Fu, F. Xu, and Y. Jin, "Reciprocal translation between SAR and optical remote sensing images with cascaded-residual adversarial networks," *Sci. China Inf. Sci.*, vol. 64, 2021, Art. no. 122301, doi: [10.1007/s11432-020-3077-5](https://doi.org/10.1007/s11432-020-3077-5).
- [20] J. Kang et al., "DisOptNet: Distilling semantic knowledge from optical images for weather-independent building segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4706315, doi: [10.1109/TGRS.2022.3165209](https://doi.org/10.1109/TGRS.2022.3165209).
- [21] L. Zhou, H. Yu, V. Pascasio, and M. Xing, "PU-GAN: A One-step 2-D InSAR phase unwrapping based on conditional generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5221510, doi: [10.1109/TGRS.2022.3145342](https://doi.org/10.1109/TGRS.2022.3145342).
- [22] S. Chen, H. Wang, F. Xu, and Y. Y.-Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806–4817, Aug. 2016, doi: [10.1109/TGRS.2016.2551720](https://doi.org/10.1109/TGRS.2016.2551720).
- [23] 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, doi: [10.1109/LGRS.2015.2513754](https://doi.org/10.1109/LGRS.2015.2513754).
- [24] F. Sharifzadeh, G. Akbarizadeh, and Y. Seifi Kavian, "Ship Classification in SAR images using a new hybrid CNN-MLP Classifier," *J. Indian. Soc. Remote Sens.*, vol. 47, no. 4, pp. 551–562, Apr. 2019, doi: [10.1007/s12524-018-0891-y](https://doi.org/10.1007/s12524-018-0891-y).
- [25] F. Samadi, G. Akbarizadeh, and H. Kaabi, "Change detection in SAR images using deep belief network: A new training approach based on morphological images," *IET Image Process.*, vol. 13, no. 12, pp. 2255–2264, Oct. 2019, doi: [10.1049/iet-ipr.2018.6248](https://doi.org/10.1049/iet-ipr.2018.6248).
- [26] Z. Huang, Z. Pan, and B. Lei, "Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data," *Remote Sens.*, vol. 9, no. 9, Aug. 2017, Art. no. 907, doi: [10.3390/rs9090907](https://doi.org/10.3390/rs9090907).
- [27] K. Wang, G. Zhang, and H. Leung, "SAR target recognition based on cross-domain and cross-task transfer learning," *IEEE Access*, vol. 7, pp. 153391–153399, 2019, doi: [10.1109/ACCESS.2019.2948618](https://doi.org/10.1109/ACCESS.2019.2948618).
- [28] G. Youk and M. Kim, "Transformer-based synthetic-to-measured SAR image translation via learning of representational features," *Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–18, 2023, doi: [10.1109/TGRS.2023.3267480](https://doi.org/10.1109/TGRS.2023.3267480).
- [29] Z. Wen, Z. Liu, S. Zhang, and Q. Pan, "Rotation awareness based self-supervised learning for SAR target recognition with limited training samples," *IEEE Trans. Image Process.*, vol. 30, pp. 7266–7279, Aug. 2021, doi: [10.1109/TIP.2021.3104179](https://doi.org/10.1109/TIP.2021.3104179).
- [30] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, "Detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 24–41, Jun. 2016, doi: [10.1016/j.isprsjprs.2016.02.013](https://doi.org/10.1016/j.isprsjprs.2016.02.013).
- [31] Z. Wu, Y. Xiong, S. Yu, and D. Lin, "Unsupervised feature learning via non-parametric instance discrimination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3733–3742.
- [32] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.
- [33] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Seattle, WA, USA, 2020, pp. 9726–9735, doi: [10.1109/CVPR42600.2020.00975](https://doi.org/10.1109/CVPR42600.2020.00975).
- [34] C. Wang, H. Gu, and W. Su, "SAR image classification using contrastive learning and pseudo-labels with limited data," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4012505, doi: [10.1109/LGRS.2021.3069224](https://doi.org/10.1109/LGRS.2021.3069224).
- [35] Y. Zhai et al., "Weakly contrastive learning via batch instance discrimination and feature clustering for small sample SAR ATR," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022, doi: [10.1109/TGRS.2021.3066195](https://doi.org/10.1109/TGRS.2021.3066195).
- [36] J. Kang, Z. Wang, R. Zhu, X. Sun, R. Fernandez-Beltran, and A. Plaza, "PiCoCo: Pixelwise contrast and consistency learning for semisupervised building footprint segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10548–10559, Oct. 2021, doi: [10.1109/JS-TARS.2021.3119286](https://doi.org/10.1109/JS-TARS.2021.3119286).
- [37] X. Sun et al., "From single- to multi-modal remote sensing imagery interpretation: A survey and taxonomy," *Sci. China Inf. Sci.*, vol. 66, no. 4, 2023, Art. no. 140301, doi: [10.1007/s11432-022-3588-0](https://doi.org/10.1007/s11432-022-3588-0).
- [38] P. Goyal, D. Mahajan, A. Gupta, and I. Misra, "Scaling and benchmarking self-supervised visual representation learning," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6390–6399, doi: [10.1109/ICCV.2019.00649](https://doi.org/10.1109/ICCV.2019.00649).
- [39] S. Xie, R. Girshick, P. Dollr, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 5987–5995, doi: [10.1109/CVPR.2017.634](https://doi.org/10.1109/CVPR.2017.634).
- [40] M. Khodadadzadeh, J. Li, A. Plaza, and J. M. Bioucas-Dias, "A subspace-based multinomial logistic regression for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2105–2109, Dec. 2014, doi: [10.1109/LGRS.2014.2320258](https://doi.org/10.1109/LGRS.2014.2320258).
- [41] E. R. Keydel, S. W. Lee, and J. T. Moore, "MSTAR extended operating conditions: A tutorial. Algorithms for Synthetic Aperture Radar Imagery III," *SPIE*, 1996, pp. 228–242, doi: [10.1117/12.242059](https://doi.org/10.1117/12.242059).
- [42] L. Huang et al., "OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 195–208, Jan. 2018, doi: [10.1109/JS-TARS.2017.2755672](https://doi.org/10.1109/JS-TARS.2017.2755672).
- [43] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in *Proc. Int. Conf. Learn. Representation*, 2017, pp. 1–16.
- [44] V. L. der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605 2008.
- [45] S. Dua, U. R. Acharya, P. Chowriappa, and S.V. Sree, "Wavelet-based energy features for glaucomatous image classification," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 1, pp. 80–87, Jan. 2012, doi: [10.1109/TITB.2011.2176540](https://doi.org/10.1109/TITB.2011.2176540).



**Hao Pei** (Graduate Student Member, IEEE) was born in Chengdu, China, in 1998. He received the B.S degree in industrial engineering from Ningbo University, Ningbo, China, in 2020. He is currently working toward the Ph.D. degree in electrical engineering with the State Key Laboratory of Millimeter Waves, School of Information Science and Engineering, Southeast University, Nanjing, China.

His current research interests include synthetic aperture radar (SAR) imagery interpretation and deep learning.



**Mingjie Su** was born in Linyi, Shandong, China, in 2000. She received the B.S. degree in electrical engineering from the Nanjing University of Science and Technology, Nanjing, China, in 2022. She is currently working toward the M.S. degree in electrical engineering with the State Key Laboratory of Millimeter Waves, School of Information Science and Engineering, Southeast University, Nanjing, China.

Her research interests include radar imagery detection and tracking.



**Gang Xu** (Senior Member, IEEE) was born in Za-zhuang, China, in 1987. He received the B.S. and Ph.D. degrees in electrical engineering from Xidian University, Xi'an, China, in 2009 and 2015, respectively.

From 2015 to 2016, he was a Full-Time Post-doctoral Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He is currently a Professor with the State Key Laboratory of Millimeter Waves, School of Information Science and Engineering, Southeast University, Nanjing, China. His current research interests are synthetic aperture radar (SAR), SAR interferometry (InSAR), inversed synthetic aperture radar (ISAR), sparse signal processing, microwave remote sensing, and millimeter wave radar.

Dr. Xu guest edited several special issues in *Remote Sensing*, *Electronics* and *Frontiers in Signal Processing*.



**Mengdao Xing** (Fellow, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Xidian University, Xian, China, in 1997 and 2002, respectively.

He is currently a Professor with the National Laboratory of Radar Signal Processing, Xidian University, Xian, China. He holds the appointment of the Dean with the Academy of Advanced Interdisciplinary Research Department, Xidian University. He has authored or coauthored more than 200 refereed scientific journal papers and two books about synthetic aperture

radar (SAR) signal processing. The total citation times of his research are greater than 10000 (H-index 50). He was rated as a Most Cited Chinese Researcher by Elsevier. He has achieved more than 50 authorized China patents. His research has been supported by various funding programs, such as the National Science Fund for Distinguished Young Scholars. His research interests include SAR, interferometric SAR, inversed SAR, sparse signal processing, and microwave remote sensing.

Dr. Xing guest edited several special issues in *IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE* and *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING*. He currently serves as an Associate Editor for radar remote sensing of *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING* and the Editor-in-Chief of *MDPI Sensors*.



**Wei Hong** (Fellow, IEEE) received the B.S. degree in radio engineering from the University of Information Engineering, Zhengzhou, China, in 1982, and the M.S. and Ph.D. degrees in radio engineering from Southeast University, Nanjing, China, in 1985 and 1988, respectively.

In 1993, he joined the University of California at Berkeley, Berkeley, CA, USA, as a Short-Term Visiting Scholar. From 1995 to 1998, he was a Short-Term Visiting Scholar with the University of California at Santa Cruz, Santa Cruz, CA, USA. Since 1988, he has been with the State Key Laboratory of Millimeter Waves, Southeast University, where he has been the Director since 2003. He is currently a Professor with the School of Information Science and Engineering, Southeast University. He has authored or coauthored more than 300 technical publications and authored two books. His current research interests include numerical methods for electromagnetic problems, millimeter-wave theory and technology, antennas, electromagnetic scattering, and RF technology for mobile communications.

Dr. Hong was an Elected IEEE MTT-S AdCom Member from 2014 to 2016. He is a Fellow of CIE. He was a recipient of the National Natural Prizes twice, the First-Class Science and Technology Progress Prizes thrice, issued by the Ministry of Education of China and Jiangsu Province Government, and the Foundations for China Distinguished Young Investigators and “Innovation Group” issued by the NSF of China. He is currently a Vice President of the CIE Microwave Society and Antenna Society and the Chair of IEEE MTTs/APS/EMCS Joint Nanjing Chapter. He was an Associate Editor for *IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES* from 2007 to 2010 and one of the guest editors for the 5G Special Issue of *IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION* in 2017.