Dual Consistency Alignment Based Self-Supervised Learning for SAR Target Recognition With Speckle Noise Resistance

Yikui Zhai, *Member, IEEE*, Jinrui Liao, *Graduate Student Member, IEEE*, Bing Sun, *Member, IEEE*, Ziyi Jiang, *Student Member, IEEE*, Zilu Ying, Wenqi Wang, Angelo Genovese, *Senior Member, IEEE*, Vincenzo Piuri, *Fellow, IEEE*, and Fabio Scotti, *Senior Member, IEEE*

***Abstract*—Deep-learning-based on convolutional neural networks (CNN) has been widely applied in synthetic aperture radar (SAR) target recognition and made significant progress. However, due to the physical effects of the equipment used to collect images, various degrees of speckle noise will be introduced into SAR images. Traditional CNN-based SAR target recognition methods are premised on the same noise intensity in the training and testing set, which is contrary to the target recognition in practice. To alleviate this problem, we propose a novel speckle noise resistant framework for SAR target recognition, called dual-consistency-alignment-based self-supervised learning. First, original SAR images are randomly added to speckle noise with different thresholds through multiplicative noise, after which contrastive pretraining is performed on unlabeled data. During this period, we combine instance pseudolabel consistency alignment and feature consistency alignment to align multiple threshold speckle noise views with original views under the same targets. Finally, the pretrained model is migrated to the downstream SAR speckle noise target recognition task. In this article, speckle noise modeling is conducted based on moving and stationary target capture and recognition data testing set, and experiment results reveal that this method can adapt to different intensities of speckle noise, is robust to modeled SAR image recognition, and maintains a high recognition rate even in small-sample learning.**

***Index Terms*—Dual consistency alignment (DCA), self-supervised learning (SSL), speckle noise, synthetic aperture radar (SAR).**

Manuscript received 4 September 2022; revised 15 November 2022, 17 January 2023, and 10 March 2023; accepted 12 April 2023. Date of publication 17 April 2023; date of current version 28 April 2023. This work was supported in part by the Key Research Projects for Universities of Guangdong Provincial Education Department under Grant 2020ZDZX3031; in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2019A1515010716, Grant 2021A1515011576, and Grant 2017KCXTD015; in part by the Guangdong, Hong Kong, Macao and the Greater Bay Area International Science and Technology Innovation Cooperation Project under Grant 2021A050530080 and Grant 2021A0505060011; and in part by the Jiangmen City Basic and Applied Basic Research Foundation under Grant 2021030103230006 670. *(Corresponding author: Bing Sun.)*

Yikui Zhai, Jinrui Liao, Ziyi Jiang, Zilu Ying, and Wenqi Wang are with the Department of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, China (e-mail: yikuizhai@163.com; jinrui_liao@163.com; zylofor@foxmail.com; ziluy@163.com; wenqiwang27@163.com).

Bing Sun is with the School of Electronics and Information Engineering, Beihang University, Beijing 100191, China (e-mail: bingsun@buaa.edu.cn).

Angelo Genovese, Vincenzo Piuri, and Fabio Scotti are with the Department of Computer Science and Dipartimento di Informatica, Università Degli Studi di Milano, 20122 Milano, Italy (e-mail: angelo.genovese@unimi.it; vincenzo.piuri@unimi.it; fabio.scotti@unimi.it).

Our code is available at https://github.com/Jordan-Liao/DCA-SSL.

Digital Object Identifier 10.1109/JSTARS.2023.3267824

I. Introduction

SYNTHETIC aperture radar (SAR) image target recognition technology is valuable in military and homeland security, such as identification friend or foe, battlefield target surveillance, maritime characteristic research, disaster assessment, etc. Deep learning with convolutional neural networks (CNNs) as representation have consolidated its status in image recognition in recent years. Researchers have introduced series of remote sensing image processing methods based on CNN and verified the effectiveness of such algorithms [1], [2], [3]. CNN-based algorithms for SAR images are mostly applied in target detection, semantic segmentation, and classification [1], [4], [5], [6], [7], [8]. Chen et al. [1] showed that baseline CNN structure can easily reach more than 97% test accuracy in 10-class classification task of moving and stationary target capture and recognition (MSTAR) dataset. Apart from this, their convolutional networks (A-ConvNets) reach a state-of-the-art average accuracy rate of 99%.

However, a particularly high sensitivity to speckle noise in SAR images was located in deep learning. When the electromagnetic wave encounters reflection from a rough surface, the SAR will be affected by echo interference during the imaging process because of the phase difference, resulting in weaker echo intensity, which generates speckle noise interfering SAR data. The SAR image is filled with noise possibly covering the target information, which might undermine the recognizability of the target and hinder SAR target recognition. Traditionally, both the spatial-domain-based and transform-domain-based methods [9], [10] are mainly utilized to denoise SAR images, while CNN-based SAR target recognition methods have also been used to learn noise [11], [12], [13], [14]. Ding et al. [11] suggested expanding the dataset by a random scattered noise modeling to train the CNN so that the model is invariant to scattered noise variations. On the basis of the data augmentation method ahead, Kwak et al. [12] put forward a speckle noise invariant regularization method in CNN. They regularized the despeckled SAR images by Lee sigma filter to minimize the feature variation dual to speckle noise, therefore optimizing robustness and preciseness. Cho et al. [13] brought up a multifeature-based CNN (MFCNN) that can extract as well as aggregate both strong features with high noise impact and smooth features with low noise

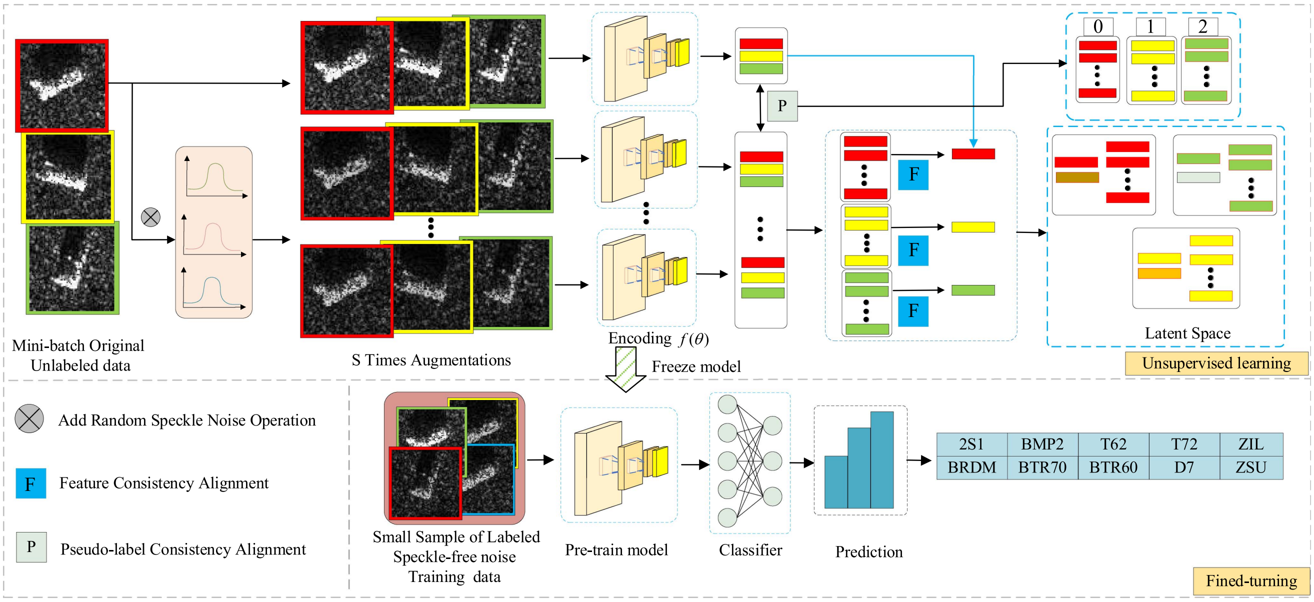


Fig. 1. Framework of the proposed DCA–SSL for SAR target recognition with speckle noise resistance.

impact. While Qin et al. [14] built a multistage wavelet scatter suppression network (Wavelet-SRNet) to effectively solve the problem of SAR target recognition under different noise levels.

The above methods mostly require extra image segmentation or filtering preprocessing steps. And the multilevel Wavelet-SRNet operates under the same speckle noise thresholds of the training and testing sets. In practice, however, the speckle noise threshold of the testing set is uncontrollable dual to the interference of the physical parameters of radar imaging and the ground background, while the noise distribution of training and testing samples does not always congruent. Therefore, it is of necessity to come up with a network that does not require additional preprocessing or filtering, and is capable of discriminating different levels of speckle noise images with once training. To satisfy the above needs, we proposed *dual-consistency-alignment-based self-supervised learning* to track the invariant features between random threshold speckle noise views. Fig. 1 shows the framework of the proposed DCA-SSL for SAR target recognition with speckle noise resistance. After a batch of instances was extracted with encoder *f*(*θ*), we first used pseudolabel consistent alignment loss and then applied the feature consistent alignment loss to align the feature vectors of views from original SAR instance; finally, we froze the pretrain weights and transferred them to the downstream network for robust SAR target recognition. In the fine-tuning phase, the testing set were added with speckle noise, while the training set is without any noise to simulate an actual target recognition scenario. This is because the actual target imaging generates random degree of speckle noise, which is out of control.

This article presents the main innovations as follows.

1) A novel SAR target recognition framework named DCA—SSL was proposed to resist speckle noise practically testing scenarios. "DCA," which includes pseudolabel consistency alignment in label space and feature consistency alignment (FA) in feature space, was proposed to perform prompt alignment between original SAR images and their speckle noise images under multilevel thresholds, so as to learn the semantic information relationship between original images and multilevel noise views.
2) Speckle noise was fused in a multiple data augmentation strategy. We applied this strategy for the first time in self-supervised contrastive learning to impel an SAR instance and its multithreshold speckle noise view to be assigned as positive samples.
3) To fit the actual SAR target recognition anti-speckle noise testing, we proposed a more demanding experimental strategy by adding speckle noise to the testing set while nothing to training set. We found that traditional speckle noise resistant methods failed under this testing setting, while our method still performed well in target recognition even with only one training.

II. Related Work

A. SAR ATR Based on Deep Learning

CNN-based SAR automatic target recognition (ATR) application is a key research topic. SAR image target recognition faces many challenges, such as complex background interfering magnetic field, lack of color information, random speckle noise interference, as well as inadequate SAR image dataset. Fortunately, researchers have made significant efforts in deep-learning algorithm architecture to successfully solve these problems. Chen et al. [1] proposed a fully A-ConvNet to alleviate the overfitting problem caused by small training datasets of SAR images. Pei et al. [15] proposed architecture based on CNN that was designed for the problem of limited SAR input data, and CNN with multi-input parallel network topology generates multiview

SAR as input to achieve SAR ATR. In order to balance the effectiveness and robustness of the ATR system, the literature [16] proposed a layered fusion method of global and local features of SAR ATR. The efficient extraction and classification of targets are performed by subtly motivating a classification method based on sparse representation by using random projection features as global features. In contrast to the traditional recognition method of directly feeding SAR data into a classifier, Guo et al. [17] combined adversarial learning and proposed a dual GAN model, which achieved improved performance for small-scale labeled SAR data and robust recognition. Inkawhich et al. [18] focused on the situation of holding 100% synthetic training data, while only measured data were used for testing. Wang et al. [19] designed an effective CNN with channelwise attention mechanism for SAR target recognition in order to reduce the computation and memory consumption of SAR ATR; meanwhile, the network structure was compressed by network pruning and knowledge distillation to improve lightweight network performance. Lin et al. [20] proposed an integrated convolutional highway unit network for processing limited labeled training data in SAR target recognition. In addition, Tai et al. [21] and Zhang et al. [22] explored transfer learning methods to effectively address the overfitting problem of deep CNN training caused by sparse SAR data. A progress has also been made recently in robust SAR ATR based on adversarial learning, and Li et al. [23] conducted some experiments demonstrating that CNNs can cope well with adversarial samples of SAR images.

B. Self-Supervised Learning and Robustness

SSL [24], [25], [26], [27], [28], [29], [30], [31] learns compact semantic data representations by defining and solving pretextual tasks. In these tasks, naturally presented supervised signals were used for training. In recent years, many pretext tasks have been proposed in the field of computer vision, including colorization [32], jigsaw puzzles [33], image restoration [34], context prediction [35], rotation prediction [36], and contrastive learning [37], [38], [39], [40], [42], [43], [44], [45], [46]. Contrastive learning has shown great potential and has become a strong standard for generic feature learning in the field of SSL. Contrastive learning was first studied in sample CNNs [37] and nonparametric instance discrimination (NPID) [38]. The sample CNN [37] learns to distinguish instances using a CNN classifier, where each class represents a single instance and its expansion. While NPID [38] proposed to consider each image as a category in unsupervised representation learning with softmax for classification, i.e., introducing a nonparametric classifier at the individual level. SimCLR [39] and MoCo [40] both adopted the contrastive loss function InfoNCE [41] in need of negative samples. A more progressive step was taken by BYOL [42], which abandoned negative samples in contrastive learning, but adopted momentum encoder to achieve better results. Lately, Chen et al. [43] presented a follow-up work SimSiam and reported a surprising outcome that simple conjoined networks are capable of learning meaningful representations without a momentum encoder. Clustering-based contrastive learning methods have also been proved effective for unsupervised visual representations. For example, SWAV [44], which discards two-by-two view comparisons and uses a method of clustering data while strengthening the consistency between different enhancements of the same view, hence improving the efficiency of memory consumption.

Recently, several works have proved the robustness of SSL to downstream tasks. Chuang et al. [47] reconstructed InfoNCE with Fahrenheit distance as a metric criterion in an attempt to improve the robustness of contrastive learning under noisy views, and provided a rigorous theoretical validation of the proposed contrastive loss function. Goyal et al. [48] provided an overview of recent large-scale studies that visual models perform more robustly and fairly when unprocessed images are pretrained without supervision. Zhong et al. [49] designed systematic testing experiments, which involved downstream task data corruption and pretraining data corruption, with types of data corruption including gamma distortion, global shuffling, local shuffling, synthesized data, and class imbalance. After studying the robustness of contrastive and supervised learning, we discovered interesting robustness behaviors of contrastive learning to different corruption settings.

Traditional deep-learning-based SAR ATR work is based on supervised learning, and the data in the training and testing sets are established to maintain the same data augmentation settings. In practice, traditional supervised learning typically involves data augmentation on batches of data sourced from multiple categories. To achieve multithreshold data augmentation, each training set must undergo preprocessing and saving with its respective threshold value, along with corresponding label information. This method of expanding data distribution entails extensive processing before conducting deep-learning training. By contrast, instance discrimination, as a self-supervised approach, treats both noisy and original views as positive samples, indirectly providing the neural network with experiential knowledge on distinguishing “similar” and “not similar” SAR targets. Through a single training, a variety of noisy views can be learned *a priori*, simplifying the procedures for manual preprocessing and data storage. Moreover, SSL reduces the requirement for a large amount of annotated data, thereby alleviating the burden of data annotation. Given the advantages of SSL in model robustness and the fact that the study of speckle noise in SAR images has not been considered in the field of SSL, this article explores the practical application problem of SAR ATR using self-supervised contrastive learning.

III. Methodology

Aiming to maximize the similarity between the features of speckle noise SAR images and the original images, we proposed joint pseudolabel consistency alignment and feature consistency alignment optimization, named DCA. It took place in the self-supervised pretraining phase, where we only let the model focus on the information of the feature maps under noise interference so that pretraining was performed on unlabeled data. Hendrycks et al. [24] demonstrated that self-supervised contrastive learning contributes to improve model robustness. We relied on the framework of contrastive learning for feature alignment, and our problem set was based on the comparison of image feature similarity under the original SAR image and its multiple speckle noise interference. However, mainstream contrastive learning is performed for two data-augmented samples of one image, so it is not applicable. Inspired by the literature [26], we borrowed the idea of multiple data augmentations, but with the difference to set up images with original images and images with random speckle noise interference. Those samples were treated as positive samples, while the other instances and their noise-interfering pictures were considered as negative samples. After a batch of instances was extracted with encoder, we first used pseudolabel consistent alignment loss and then applied the feature consistent alignment loss to align the feature vectors of views from one instance. Pseudolabel consistency alignment aims to distinguish SAR samples well from positive and negative samples, while feature space consistency alignment is to bring samples of the same class closer together. Finally, we froze the learned weights and migrated them to the downstream network for robust recognition. In the fine-tuning phase, the testing set data were added with speckle noise, while the training set were left unnoised. With the "prior knowledge," the CNN can easily recognize different noise views of the same target and thus obtains robust recognition performance. The implementation process of each part is described in detail separately in the following paragraphs.

A. Speckle Noise Data Augmentation Mechanism

The SAR imaging system introduces the speckle noise of SAR images. Unlike the widely studied additive noise in optical images, it belongs to multiplicative noise, a large proportion of which is of high frequency [12], [13], [14], [50], [51], [52]. To prove the effectiveness of target recognition of the suggested method for speckle noise SAR images, we began with speckle noise modeling referring to the literature [11] for the generation of SAR images with various degrees of speckle noise. The speckle noise in SAR images can be represented by multiplicative noise

$$I(x,y) = Q(x,y) \cdot S(x,y) \tag{1}$$

in which I represents the SAR image produced with speckle noise, x and y denote the coordinates of each resolution unit of the SAR image, Q refers to the original SAR image, and S represents the intensity of speckle noise matching with the exponential distribution. The process goes with the generation of a random noise with mean 0 and variance 1, followed by exponential transformation and multiplication with the original image at last. Because the standard exponential distribution noise possesses some pixel points which can darken the image, we replaced standard exponential distribution with the truncated exponential distribution $e_a = \min(\exp(\mathrm{randn}(M, N), a))$, in which a stands for the truncation parameter, and randn(M, N) means randomly generated two-dimensional (2-D) distribution of $M \times N$. The rationale for $M \times N$ is that SAR images are 2-D grayscale maps. In this way, SAR images with various levels of speckle noise can be obtained by using various parameters a. Fig. 2 presents an original SAR image and four noisy SAR images under different parameters a. With the decrease of the parameter value, the corresponding SAR images can be obtained; however, the target in them shows less visibility and somehow damaged structure. In the algorithm implementation, this part is operated as *Sn(a)*.

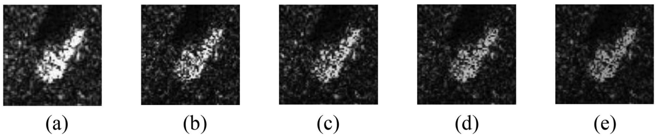


Fig. 2. Speckle noise SAR images. (a) Example of MSTAR image. (b) Speckle noise SAR image with truncated exponential distribution and $a = 1$. (c) Speckle noise image with $a = 0.9$. (d) Speckle noise image with $a = 0.8$. (e) Speckle noise image with $a = 0.7$.

Given the dataset $D = \{X, Y\}$, herein $x \in X$ stands for the training samples and $y \in Y$ refers to the corresponding labels. We started with a random speckle noise addition operation on a batch of data. The noise interference intensity was defined between [0.71] to ensure relatively complete target. The mathematical expression for the random speckle noise interference is

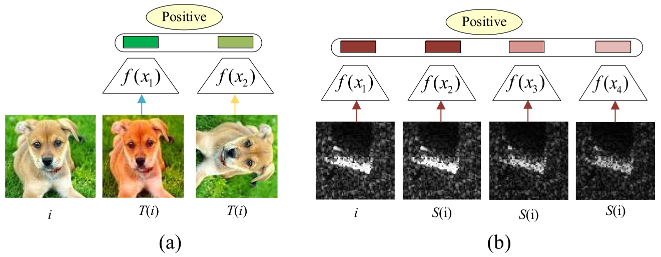
$$\overline{x}_i = x_i^{\mathrm{Sn}(\mathrm{rand}(a))},\ a \in R : \|a\| \le \varepsilon,\ i \in \{1, \ldots, N\} \tag{2}$$

where $\overline{x}_i$ is a set of speckle noise SAR images after noise enhancement, ε is the range of interference, and N is the sample size of a batch of data.

B. Dual-Consistency-Alignment-Based Self-Supervised Learning

For the data under the influence of speckle noise, we have tried to filter out SAR image noise before target recognition [51], but unfortunately, it is challenging to achieve SOTA results. From the experimental performance in Fig. 9, the RestNet18 network with discrete wavelet transform (DWT) for filtering still suffers from unstable recognition results. We conjectured that purely using the filter for noise reduction can damage the feature representation of the target because the speckle noise often belongs to the high-frequency part of the SAR image, and the original target features tend to be damaged if speckle noise is denoised with force, which can affect the learning of high-frequency semantic information of the SAR images in the neural network. We shifted our perspective to noise tolerance, in which the neural network should be able to learn how the target samples maximize the related information relationship between the target samples and their augmented images so that the CNN learns the invariant consistency between multiple augmented images. The literature [16] demonstrates that SSL can improve the robustness of the model. Mainstream comparison learning always use two data augmentations to transform one sample, however, our setting is to compare original images with multiple levels speckle noise interference pictures, so mainstream comparison learning cannot be adopted. We proposed using S different thresholds of speckle noise data augmentations for one target sample, thus maximizing the similarity representation of original SAR images and their speckle noise enhancing images, as shown in Fig. 3.

First and foremost, we recalled the classical loss function of contrastive learning: InfoNCE loss. Given two vectors V_1 and



(a) (b)

Fig. 3. (a) Generic contrastive loss conception, where both data augmentations of the same image are considered as positive samples. (b) Our proposed contrastive loss conception based on SAR images with multiple noise views, where the original image and various degrees of speckle noise views are classified as positive samples, with the help of which the pseudolabels achieve alignment.

V$_2$ in the feature space, we interpreted contrastive learning as a binary classification problem with sample pairs (*V*$_1$, *V*$_2$), where the label is 1 if the sample pair is from the same joint sample instance of the data augmented view, and −1 if it is from a different instance of the data augmented view. In general, the contrastive learning loss function is based on two data augmentations, so that we can clearly write InfoNCE loss as follows:

$$L_{\text{InfoNCE}}(x, v, i) = -\frac{1}{n}\sum_{i=1}^{n}\log\frac{e^{f(x)^T g(v)/t}}{e^{f(x)^T g(v)/t} + \sum_{i=1}^{K} e^{f(x)^T g(v_i)/t}} \quad (3)$$

where $f(x)$ and $g(v)$ are embedded by CNN, *t* is a positive value of temperature to avoid gradient saturation, while *K* is the value of the negative sample pair. For the specific task of this article, the same SAR samples and their noise views are labeled as label 1, thus achieving pseudolabel alignment of SAR noise instances.

After *S* thresholds of speckle process a mini-batch data of size *N*, the original SAR images and the speckle-processed samples are passed into the feature extraction network with (*S*+1)*N* samples. The original SAR images and the speckle-processed samples $\{x^s \in \mathbb{R}^{n\times s}, s = 1, 2, \ldots S\}$are passed into the feature extraction network with a total of (*S*+1)*N* samples. The category pseudolabel consistency alignment is performed on the label space so that the instance and their noise images (noted as positive samples) are close to each other in the feature space, while the SAR images of different instances and their noise images (noted as negative samples) are separated. This process is called pseudolabel consistency alignment, and the loss function is denoted as

$$L_{LA}(r_s^+, r_i^-, i) = -\frac{1}{n}\sum_{i=1}^{n}\log\frac{e^{r_s^+}}{e^{r_s^+} + \sum_{i=1}^{K} e^{r_i^-}}$$

$$:= -\frac{1}{n}\sum_{i=1}^{n}\log\frac{e^{f(x)^T g(v_s)/t}}{e^{f(x)^T g(v_s)/t} + \sum_{i=1}^{K} e^{f(x)^T g(v_i)/t}}. \quad (4)$$

Herein, $r = \{\{r_s^+\}_{s=1}^S, \{r_i^-\}_{i=1}^K\}$, *r* denotes a batch of SAR data, whereas $r^+$and $r_i^-$are the scores of relevant SAR samples (positive samples) and uncorrelated SAR samples (negative samples), and *t* is the temperature parameter introduced to avoid gradient saturation. The loss function learns to classify whether a pair (*x*, v_s) is a positive or negative sample by maximizing/minimizing the positive scores and negative scores. When *i*th vector x_i is extracted from CNN, there are *S* vectors similar to it, i.e., *x* has *S* positive samples. For a minibatch (*S*+1)*N* samples x_i (*i* = 1,2…,(*S*+1)*N*), our goal is to make them grouped into *N* corresponding categories.

Algorithm 1: Pseudocode of DCA–SSL for SAR ATR.

```
Input: Batch size N, S times speckle noise augmentation, full connected layer $f_{fc,}$ feature extraction $f_\theta$ , learning rate $\eta$ , temperature T.
Pretraining: pretrain the network with $L_{LA}$ and $L_{FA}$
For all $t \in \{1, \ldots, epochs\}$do:
for each minibatch do
Initialize the parameters of $f_{fc}$
For all $s \in \{1, \ldots, S\}$do
For sampled minibatch $\{x_i\}_{i=1}^N$do
For all $i \in \{1, \ldots, N\}$do
Assign instance label $y_i^k$ to $\widetilde{x}_i^s$ where$y_i^k = i$
# Set the random level speckle noise
 $\widetilde{x}_i^k = a^k(x_i)$
 $h_i^k = f_{fc}^t(f_\theta^t(\widetilde{x}_i^k))$
end for
end for
define
 $L_{LA} = -\frac{1}{nS}\sum_{s=1}^{S}\sum_{i=1}^{n}\log\frac{e^{s^+}}{e^{s^+}+\sum_{i=1}^{K} e^{s_i^-}}$
Update network $f_\theta^t$and $f_{fc}^t$to minimize $L_{LA}$
end for
for all $s \in \{1, \ldots, S\}$ and $i \in \{1, \ldots, N\}$do
 $x_i^k = f_\theta^t(\widetilde{x}_i^k)$
end for
define
 $L_{FA} = \frac{1}{DN}\sum_{i=1}^{N}\sum_{d=1}^{D}\|X_{id}^s - X_{id}\|_2^2 W_i^s$#dth dimension
 vector alignment
$W_i^s =$
 $\{ \begin{array}{c} 1, \text{the} s - \text{th speckle noise view of the } i - \text{th instance,} \\ 0, \text{otherwise,} \end{array}$
update network $f_\theta^t$to minimize $L_{FA}$
end for
#momentum update network
 $\theta_t \leftarrow m\theta_{t-1} + (1-m)\theta_t$
end for
Fine-Turned: Computational identification of similar category under speckle noise interference for minibatch
 $\{x_i^s\}$
do
 $l_c(f_\theta^t(t(x))) \to c$
where c is the logical labels values of the same category
update network $f_\theta^t$to maximized between loss $l_c$ and c.
end for
```

Then, FA loss function is introduced to closely draw the similarity relationship between *S* noise-interfered images and their original noise images on the feature space for feature alignment. Due to the SAR imaging characteristics, there is little variability between the speckle noise interfered SAR image and the original SAR images in the high dimension. Thus, we compare the noisy picture (s views) feature $X_i^s = f(x_i^s)$ on the given latent space after feature extraction with the original feature X_i. Then, the Euclidean distance was calculated in each of the $d(d = 1{,}2\ldots{,}D)$ dimensions with the aim of weakening the gap between the original image and the speckle noise image of different degrees. The consistent alignment loss function can be expressed as

$$L_{\text{FA}} = \frac{1}{DN}\sum_{i=1}^{N}\sum_{d=1}^{D}\|X_{id}^s - X_{id}\|_2^2 W_i^s \tag{5}$$

where subscript d means the dth dimension of a feature vector. The network learning parameters are updated by minimizing L_{FA}. W_i^s serves to closely link instances i and the semantic features with the same instances are to be aligned. We define this as the following equation:

$$W_i^s = \begin{cases} 1, \text{ the } s\text{th speckle noise view of the } i\text{th instance,} \\ 0, \qquad \text{otherwise.} \end{cases} \tag{6}$$

Our model is optimized by instance discriminant alignment and consistency alignment loss functions, and the total loss function can be written as

$$L_{\text{DCA}} = L_{\text{LA}} + L_{\text{FA}}. \tag{7}$$

In general, the gradient optimization is updated by backpropagation with the objective of reducing the value of L_{DCA}. The parameters of network learning are defined as the current epoch θ_t, m stands for the momentum confidence from 0 to1, and θ_{t-1} refers to the historical parameters of the model. To enhance the stability and timeliness of the model, we used the momentum update learning strategy by

$$\theta_t \leftarrow m\theta_{t-1} + (1-m)\theta_t. \tag{8}$$

Further, we migrated the above pretrained parameters to the ResNet18 network for learning with speckle noise only added to the testing set. For images disturbed by speckle noise, the classifier is able to assemble enhanced views of the same target image into adjacent regions. For the input x belonging to category c, we predict the category of x by computing its transformed expectation distribution

$$S(x) = \underset{c\in Y}{\text{argmax}}\, \mathbb{E}_{t\sim\Gamma}(l_c(f(t(x))) = c). \tag{9}$$

Here, $l_c(.)$ is the logical value of the category. The category distribution expectation $\mathbb{E}_{t\sim\Gamma}$is maximized by aggregating the SAR image features through multiple speckle noise processing. Algorithm 1 is a pseudocode based on DCA–SSL, which summarizes the ideology of this article.

IV. Experiment and Discussion

A. Experiment Setup

In our experiment, the MSTAR dataset [1], [53] is introduced to prove our method's effectiveness. The database includes X-band SAR images of multiple targets with 0.3 m × 0.3 m resolution. Before the experiment, we cropped all images from a size of 128 × 128 to 64 × 64 to remove background interference. In this case, samples with a 17° pitch angle are categorized as the training set and those with a 15° pitch angle as the testing set. The quantity of each type samples in the experiments is concluded in Table I. And the corresponding SAR targets and their natural images are displayed in Fig. 4. To fit the actual SAR target recognition anti-speckle noise testing, we proposed a more demanding experimental strategy by adding speckle noise to the testing set while nothing to training set. Since the original SAR data are distributed with local speckle noise, our experimental setup is to simulate the problem of inconsistent noise distribution between the training set and the testing set. The experiment results from subsections B, C, and D are all based on the inconsistent noise setting.

The optimization of this model is conducted through Adam optimizer. The pretraining learning rate is regulated at 0.3, while the momentum parameter at 0.9 with the batch size of 512 and the trained epochs of 300. In the fine-tuning phase, the learning rate is 0.03 and 200 epochs are trained. The basis for the selection of these parameters is described in detail in the subsequent Section B.

Since computational consumption of SSL is associated with the backbone network, the computational complexity of our method depends on the depth of the backbone feature extraction network. Subsection B, Part 3 concludes our method has the best performance for SAR image target recognition with RestNet18 as feature extraction network. We calculated the consumption performance of our method under this setting, and Table II displays the consumption performance of our method under the condition of a single image input. Params is the total number of network parameters, and Flops is the amount of floating point arithmetics. It is notable that in addition to the parameter values and FLOPs, commonly used by CNNs to consume computational metrics, we also presented the amount of multiplication metrics and the amount of memory usage during the computation. They are denoted as "Madd" and "Memory usage," respectively. We may set the sample batch size of 512 to accommodate unburdened computations on our computer workstation that owns Nvidia 2080Ti with 11G RAM.

B. Ablation Studies

1) Times of Data Augmentation: The time of data augmentation for unsupervised pretraining affects the positive and negative sample contrastive pairs for network contrastive learning. Traditionally, self-supervised contrastive learning is performed twice for data augmentation, but the effect of two data augmentations is not ideal due to the diversity of noisy views. We would like to prove the consistency of an instance sample under different levels of noise views. To study the influence of the quantity of data augmentation on our method, we did the experimental comparison shown in Table III. Setting the extreme condition that the noise interference intensity is $a = 0.7$, it can be observed that the fine-tuned recognition performs best when

TABLE I
CATEGORIES AND QUANTITIES OF MSTAR DATASET

Targets	Serial	Training set		Serial	Testing set	
		Depr	Number		Depr	Number
BMP2	9563	17°	233	9563	15°	195
				9566	15°	196
				c21	15°	196
BTR70	c71	17°	233	c71	15°	196
T72	132	17°	232	132	15°	196
				812	15°	195
				S7	15°	191
ZSU23/4	D08	17°	299	D08	15°	274
ZIL131	E12	17°	299	E12	15°	274
T62	A51	17°	299	A51	15°	273
BTR60	K10yt7532	17°	256	K10yt7532	15°	195
D7	92v13015	17°	299	92v13015	15°	274
BDRM2	E71	17°	298	E71	15°	274
2S1	B01	17°	299	B01	15°	274

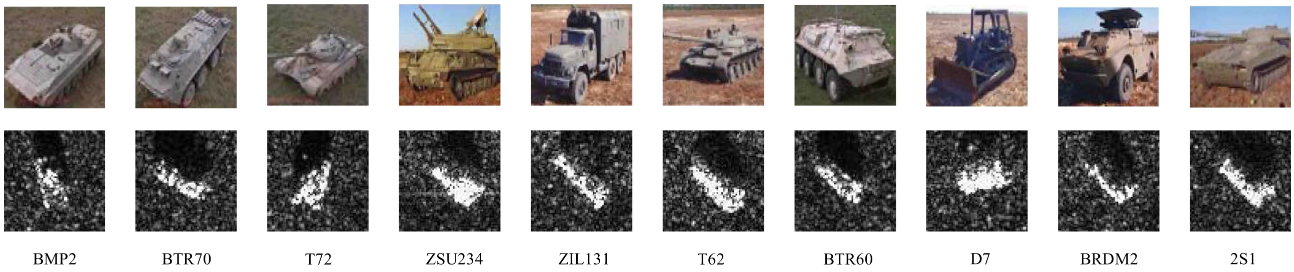


Fig. 4. Selection of samples from the MSTAR database and their natural images.

TABLE II
COMPUTATIONAL OF DCA–SSL WITH SINGLE IMAGE INPUT

Computational metrics	**Value**
Params	10.7 M
Flops	148.65 M
Madd	296.97 M
Memory usage	46.96 M

TABLE III
RECOGNITION ACCURACY FOR THE TIMES OF DATA AUGMENTATION AND ITS FINE-TUNED WITH THE NOISE INTENSITY OF THE MSTAR DATA VALIDATION SET AS 0.7

Augmentation times	**Accuracy**
2	78.23
3	91.10
4	88.79
5	94.91
6	**95.92**
7	94.04
8	93.09

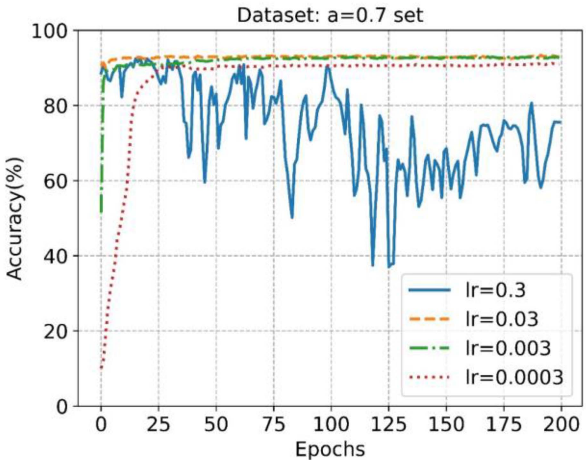


Fig. 5. Learning rate and its recognition rate for 200 iterations fine-tuned when the MSTAR data validation set noise intensity is adjusted to 0.7.

the number of data augmentation for pretraining is 6. Therefore, we chose to carry out six data augmentations.

2) Learning Rate: To determine the target recognition rate of the model for speckle noise interference, we did a set of comparison experiments to verify this and it was demonstrated that the greater the intensity of the speckle noise, the worse the recognition performance. We set the extreme noise condition of 0.7. In the fine-tuning stage, we selected the learning rate of 0.3, 0.03, 0.003, and 0.0003 to determine which parameter shows the best performance. Fig. 5 shows the results of comparison. It was noticed that the model could converge quickly and keep a high recognition rate when the learning rate was 0.03, so we chose 0.03 as the learning rate in the fine-tuning stage.

3) Backbone: Since representation extraction of unsupervised feature learning is a crucial procedure, we used MSTAR data as input to study the unsupervised deep CNN which is suitable for learning SAR images. Because the residual network is a classical CNN network, RestNet (RestNet18, RestNet34,

TABLE IV
RECOGNITION ACCURACY OF THE THREE METHODS WHEN THE TESTING SET SPECKLE NOISE INTENSITY IS 1.0

Method	Unsupervised pretraining	Dual consistent alignment	Acc
RestNet18			74.65
MoCo	✓		82.77
DCA-SSL	✓	✓	**93.44**

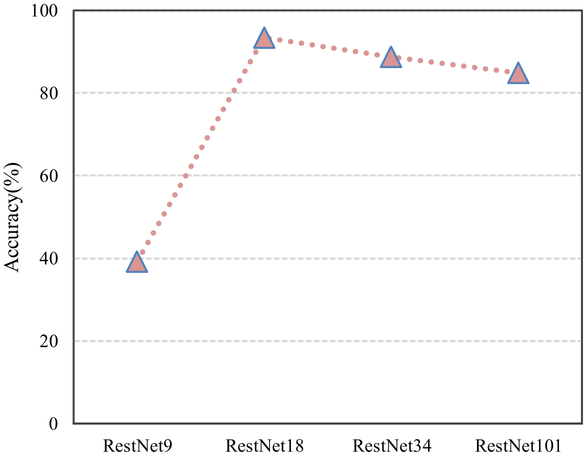


Fig. 6. Recognition rate of the fine-tuned feature extraction network for unsupervised learning using different depths when the testing set speckle noise interference intensity is set to 0.7.

RestNet50, and Re-stNet101) in PyTorch deep-learning architecture was employed in this article for recognition performance at different depths. Also, we designed RestNet9 based on deep residual. Fig. 6 shows the effect of unsupervised representation learning after performing fine-tuning, and the speckle noise in the testing was set at the fine-tuning stage set to 0.7. Equipped with the best effect, RestNet18 is naturally selected as the architecture for representation learning.

4) Iteration Number of Pretraining: The iteration number of pretraining is likewise a hyperparameter of unsupervised learning. On this device, we visualized SAR image representation learning with t-SNE [54] to turn high-dimensional vectors into 2-D vectors, as shown in right of Fig. 7. When the iteration of pretraining reached 300, the representations of the same category can all be close together, indicating that feature alignment has been well performed, so we set the iteration number of pretraining to 300.

5) Pretraining and Feature Alignment: We now study the effect of pretraining model with dual alignment idea on SAR speckle noise view recognition. First, we verified the testing accuracy of the backbone network RestNet18 after a random degree of speckle noise augmentation. As our method set up a momentum update strategy and used a self-supervised pretraining strategy, the MoCo with feature extractor RestNet18 was chosen for study. In this article, random speckle noise augmentation was also added in the pretraining phase, while noise was only added to the testing set not the training set in the fine-tuning phase. Table IV demonstrates our ablation learning of pretraining and feature alignment. We set the noise intensity of the testing set to 1.0 for proving the effectiveness of pretraining and feature alignment. In particular, the unsupervised pretraining alone does not distinguish the learned representations well, but with the idea of dual alignment, the unsupervised representation learning can be distinguished well. Fig. 7 shows the t-SNE plot of MoCo after 300 iterations of unsupervised noise view learning. Although some of the features are clustered, it did not have an obvious effect; on the contrary, the DCA demonstrated a significant effect, which leads to the success of the downstream task.

C. Recognition Performance of Supervised Networks Under Different Noise Intensities

Each experiment was repeated 10 times and the average test accuracy was regarded as the recognition result. When training labeled data, we apply some data augmentation operations, such as random rotation, flipping, and cropping. Fig. 8 shows the confusion matrix of the fine-tuned recognition rates for the testing set after the same pretraining with different degrees of speckle noise interference. Results show the excellent recognition rates are maintained by the model under diverse noisy interference, even just under one training round. Fig. 9 presents the image results of the comparison algorithm on the original image and the four noisy image sets. The mainstream neural networks maintain high recognition rates for recognizing SAR targets without noise interference; however, the recognition performance plummeted after the addition of a little noise to the testing set. When the parameter a is set to 1, the recognition rates of ResNet18, RestNet50, SIN-CNN [12], EfficieNetV2, Wavelet-SRNet [14], and RestNet18 with DWT, MFFA-SARNet [55], A_ConvNet [1], and MFCNN [13] drop sharply from 96.19%, 94.88%, 98.81%, 99.30%, 97.5%, 97.81%, 95.78%, 95.21%, and 96.88% to 74.65%, 61.47%, 36.30%, 57.5%, 50.33%, 56.28%, 53.96%, 63.59%, and 52.15%, respectively. In the same condition, the recognition rates of our method only show a mild decrease from 98.21% to 93.60%. Moreover, traditional methods perform worse in recognition when a is set to 0.9, 0.8, and 0.7, indicating a great impact of speckle noise on the recognition performance of the neural network. Nevertheless, in this case, our method can still maintain the recognition rate over 93.4%, which probably thanks to the fact that the neural network has learned to maximize the relationship between the category information in different threshold cases.

D. Recognition Performance of Self-Supervised Networks Under Different Noise Intensities

Comparison between the proposed method and the mainstream SSL methods(SimCLR [39], MoCo-v2 [40], BYOL [42], SimSame [43], SwAV [44], AdCo [25], and RLRS [26]) as well

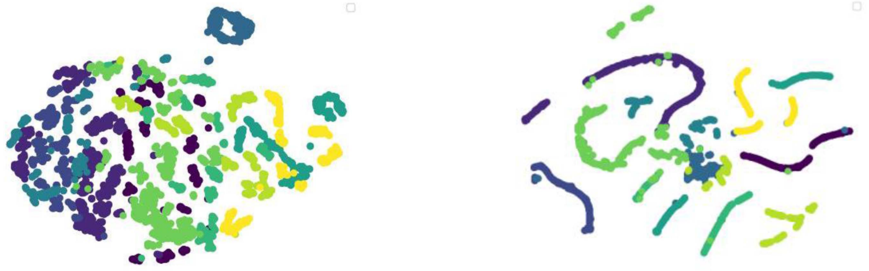


Fig. 7. T-SNE visualization of unsupervised learning representations after 300 iterations of the MSTAR training set. Left: MoCo; right: DCA-SSL (ours). Colors represent categories.

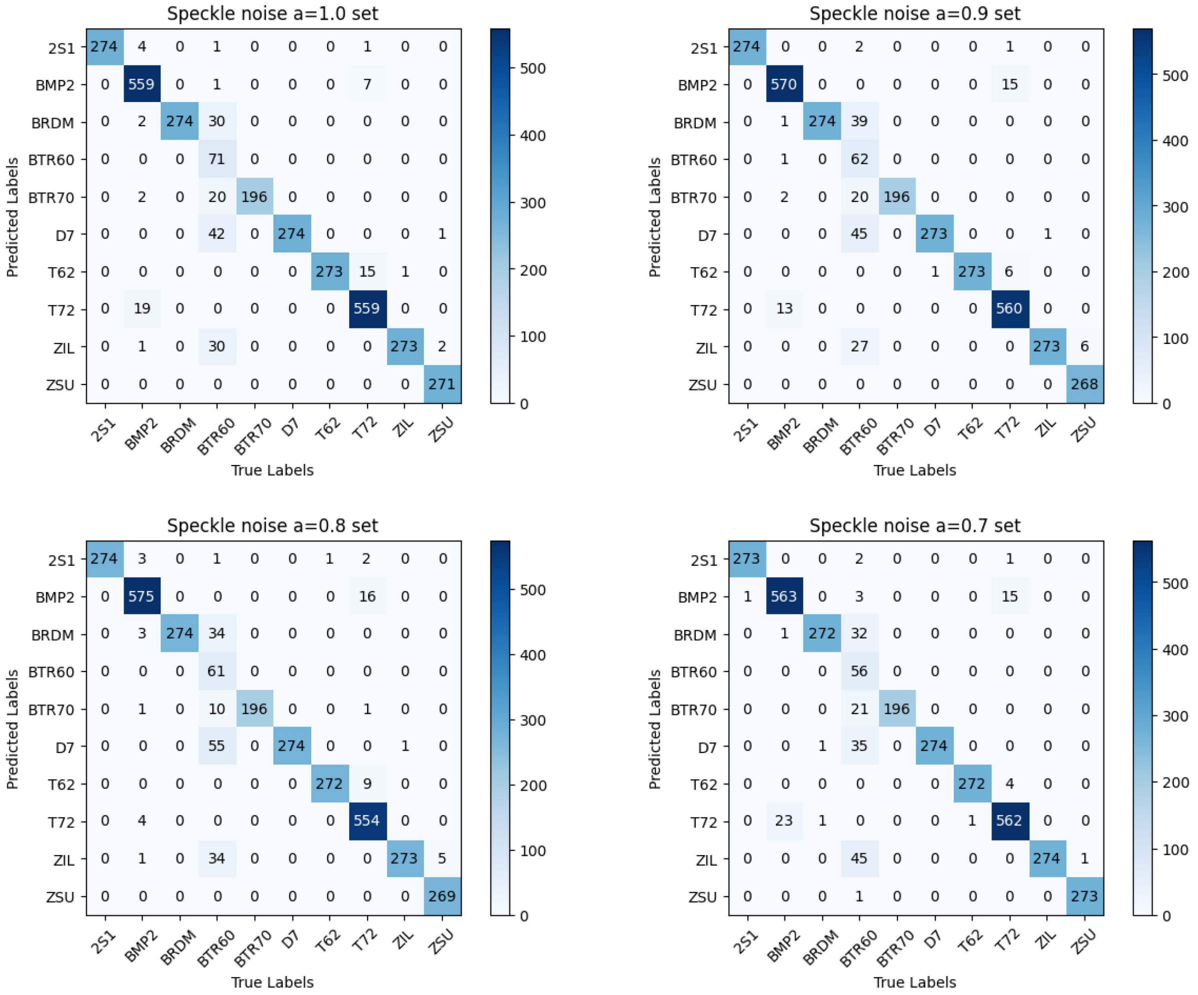


Fig. 8. Confusion matrix of fine-tuned recognition rates for testing sets with different degree of speckle noise interference after once pretraining.

as SSL methods based on SAR ATR (RotANet [56], CLPL-SAR [57], and MS-SSL [58]) was still conducted with the dataset set in experiment setting. To ensure a fair comparison, the unlabeled SAR images were also mixed with speckle noise in the pretraining phase with 300 epochs while with 200 epochs in the fine-tuning phase. During the fine-turning phase, we used only 10% of the training data. In addition to fine-tuning, linear evaluation was applied to this experimental evaluation as an alternative way of SSL model evaluation. After a pretrained model was obtained in the SSL phase, the linear evaluation model was updated with parameters only for the fully connected layer of the model by virtue of a few labeled data, and the other weights of the model were kept fixed. The literature [39] found that linear evaluation was slightly less accurate than the strategy of fine-tuning, and all parameters were updated for the downstream recognition task. Table V records the performance

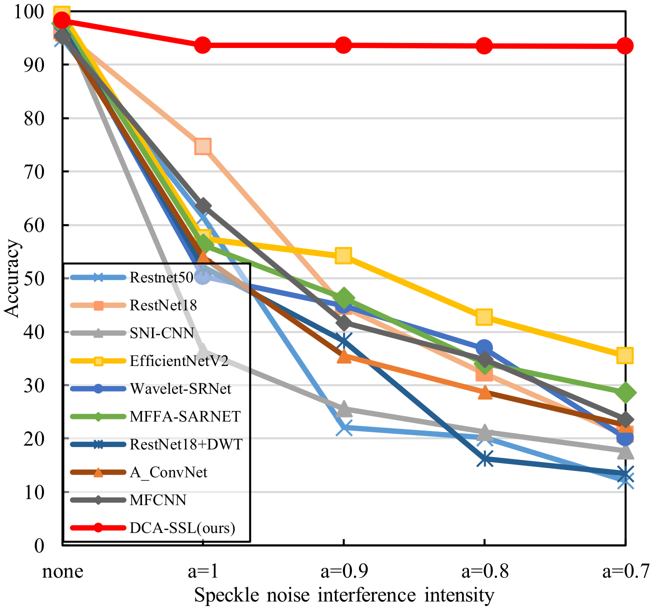


Fig. 9. Recognition rates of 10 methods under four intensities of speckle noise interference.

TABLE V
COMPARISON OF RECOGNITION ACCURACY (%) OF MSTAR DATASET ON SELF-SUPERVISED LEARNING METHODS UNDER DIFFERENT THRESHOLDS OF SPECKLE NOISE

Method	None	a = 1.0	a = 0.9	a = 0.8	a = 0.7
Fine-turned:					
SimCLR	95.47	78.31	52.23	44.52	37.96
MoCo v2	96.15	82.77	62.35	57.93	49.29
Byol	**98.91**	86.11	79.44	61.79	36.27
SimSame	98.03	80.36	69.40	51.85	47.04
SwAV	97.50	79.55	67.51	58.25	40.10
AdCo	96.89	87.58	66.21	59.29	52.13
RLRS	93.52	70.10	56.32	52.85	49.21
RotANet	96.86	65.31	50.05	43.56	35.23
CLPL-SAR	95.83	84.05	65.72	58.02	41.88
MS-SSL	90.80	78.31	55.83	48.71	40.39
DCA-SSL	98.21	**93.66**	**93.62**	**93.47**	**93.44**
Linear evaluation:					
SimCLR	92.83	66.83	48.33	41.58	36.82
MoCo v2	93.48	80.49	62.08	55.45	48.93
Byol	93.88	85.60	78.51	60.66	36.19
SimSame	94.85	79.95	69.33	51.51	37.89
SwAV	92.57	79.29	67.08	57.95	39.02
AdCo	94.59	85.95	65.33	58.81	50.25
RLRS	91.12	65.96	52.83	50.48	46.59
RotANet	93.28	62.13	46.50	40.88	34.13
CLPL-SAR	92.81	81.51	63.40	57.38	39.12
MS-SSL	88.92	76.28	52.56	47.83	39.89
DCA-SSL	**95.19**	**92.95**	**92.83**	**92.76**	**92.69**

of various SAR target recognition under different degrees of speckle noise interference, which explicitly demonstrates the superiority of our method in SSL.

TABLE VI
RECOGNITION ACCURACY (%) OF DCA–SSL WITH VARIOUS LEVEL OF SPECKLE NOISE DATA AND SMALL-SAMPLE NUMBER

Level	**1:2**	**1: 3**	**1:4**	**1:8**	**1:16**	**1:32**
Fine-turned:						
None	94.19	93.73	93.35	92.69	92.19	91.44
a=1.0	93.72	92.91	91.98	90.45	88.57	87.54
a=0.9	93.66	92.58	91.76	90.40	88.28	87.26
a=0.8	93.51	92.62	91.72	90.67	88.97	87.23
a=0.7	93.44	92.33	91.60	90.49	88.51	87.19
Linear evaluation:						
None	93.16	93.38	93.35	92.58	92.02	91.15
a=1.0	93.06	91.85	91.28	90.06	88.33	86.49
a=0.9	93.04	91.72	91.06	90.08	88.05	86.12
a=0.8	93.01	91.82	90.85	89.92	88.13	86.05
a=0.7	92.85	91.68	90.66	89.81	88.06	85.98

E. Recognition Performance of DCA–SSL With Various Level of Speckle Noise Data and Small Sample

It was noticed that the recognition rate of our method can still be maintained at a high level when the training samples were reduced. We took the MSTAR data training set as 1:2, 1:3, 1:4, 1:8, 1:16, and 1:32, respectively, and added different thresholds of speckle noise to the testing set as in the previous experimental setup method, while left the training set untreated. Table VI and Fig. 10 show the recognition rates of different levels of speckle noise data in small-scale training samples. When the training samples were gradually reduced, the recognition performance of our method did not drop significantly, which showed that our method can be well applied in SAR target recognition scenarios with insufficient data volume and is less sensitive to speckle noise.

V. DISCUSSION

We selected some SAR testing set samples and added different levels of speckle noise interference to perform interpretability analysis of the DCA–SSL method. Fig. 11 shows the Grad-cam [59] attention maps of randomly selected 2S1 targets and their speckle noise interference images. 2S1(a) refers to the result of layer4 pretrained by RestNet18 and 2S1(b) stands for the result of MoCo (feature extractor is RestNet18) pretrained and fine-turned. We found that RestNet18 and MoCo can focus on the target without speckle noise interference; on the contrary, after being added with speckle noise, both networks have difficulty in the attention of the target center area, which indicates that the noise interference has a greater impact on the target recognition. MoCo is able to focus on some target areas or near the target areas, but as the interference intensity increases, the larger the focus bias of the network will be. While DCA–SSL is able to focus on the target regions, as shown in Fig. 12, which is the Grad-cam attention map of MSTAR 10 categories of targets after DCA pretraining and fine-tuning. These Grad-cam attention maps validate that our method can learn the similarity of SAR images under multiview speckle noise and maintain the consistency induction bias, which improves the recognition rate of SAR ATR for downstream tasks.

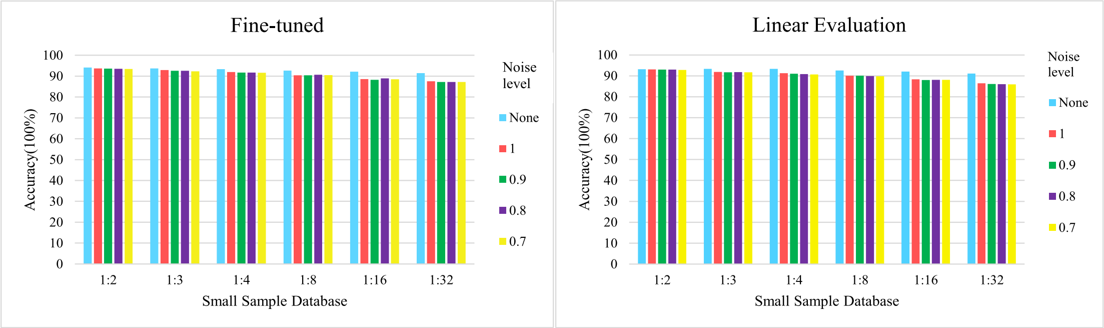
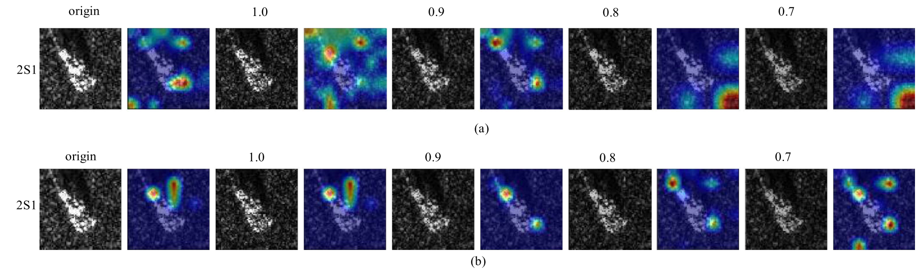


Fig. 10. Fine-tuned/linear evaluation on small-sample database with speckle noise disturbances.



(a)

(b)

Fig. 11. Visualization of 2S1 targets under speckle noise interference and their Grad-Cam attention maps, upper: RestNet18 layer4 Grad-Cam attention maps, lower: Moco fine-tuning stage layer4 Grad-Cam attention maps.

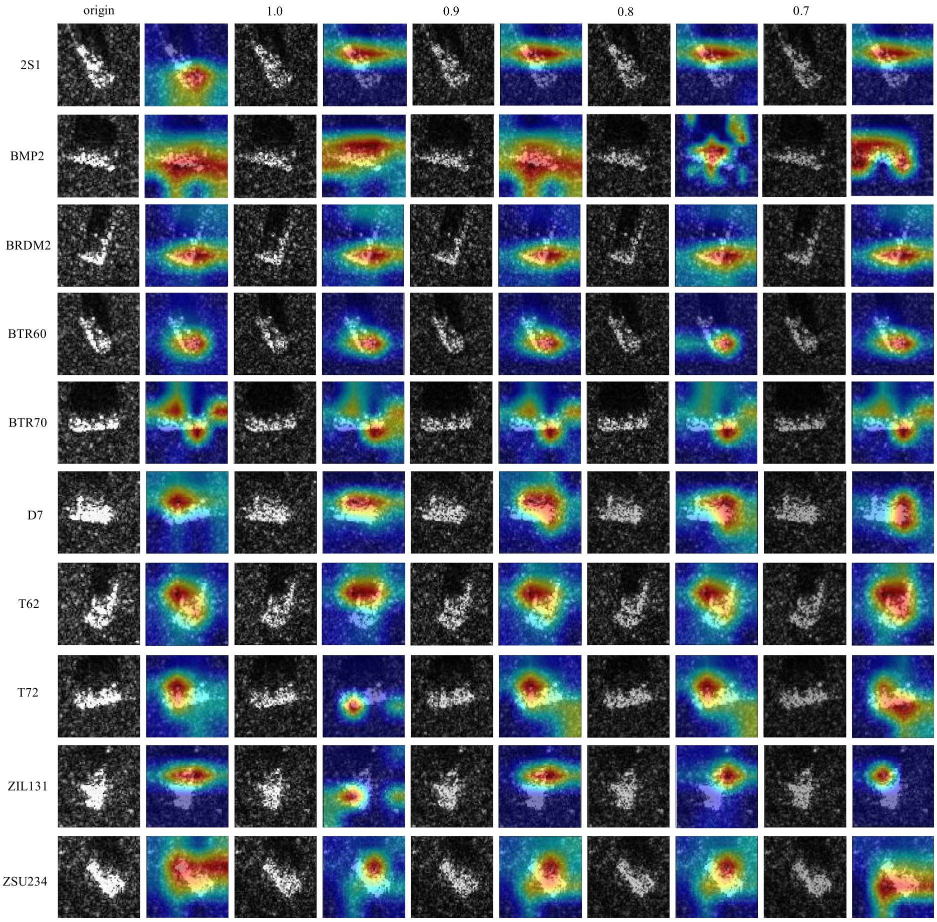


Fig. 12. Some MSTAR testing sets of speckle noise views and their corresponding Grad-CAM attention maps, which is the outcome of DCA-SSL in fine-tuned stage by selecting layer.

DCA–SSL still performs better in the case of small samples, which proves the superiority of SSL on small-sample learning. Classical contrastive learning methods [38], [39], [40], [41], [42] require large amounts of unlabeled data for comparison in upstream tasks. Tens of thousands unlabeled data are required in unsupervised learning, which are crucial to improve the network's ability to discriminate between "similar" and "dissimilar." Thus, when unsupervised learning knowledge is transferred to downstream tasks, high classification accuracy can be achieved with few labeled data. We performed six times data augmentation of the SAR samples among the limited number of training. We indirectly expanded the number of samples for unsupervised pretraining, and considered the original SAR samples with views under different threshold speckle noise as positive samples, which is definitely quite helpful for the excellent performance of small-sample SAR target recognition for downstream tasks.

Despite the good performance of our method on speckle noise resistant SAR target recognition, the limitation of DCA–SSL is that hundreds and thousands of unlabeled SAR categories similar to downstream task are required during pretraining stage. In some extreme SAR target recognition domains, obtaining these data may not be easy.

VI. Conclusion

In this article, aiming to cope with the challenge of speckle noise on SAR target recognition, we proposed to use DCA, including pseudolabel consistency alignment and FA to weaken the gap between speckle noise views and original views. We began with self-supervised representation learning from speckle noise tolerant features and then undertook a target recognition task under interference conditions. We verified the well performance of the proposed SAR image target recognition method when various threshold of speckle noise was introduced. Compared with the traditional supervised networks used for speckle noise resistance target recognition and the mainstream self-supervised networks, our method is optimal and has the advantage of being highly resistant to speckle noise. Extensive experiments conducted on MSTAR dataset demonstrate the proposed method achieves optimal recognition results and possesses the advantage of strong resistance to speckle noise. Despite of small-sample condition, satisfying recognition performance can still be located in the proposed method, which provides feasible solutions for application in SAR target recognition scenarios with insufficient data. In the future, anti-speckle noise SAR target recognition with extremely small amount of SAR target samples during unsupervised learning will be studied, hoping that excellent anti-speckle noise SAR target recognition performance would still be achieved.

References

[1] S. Chen, H. Wang, F. Xu, and Y. - Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806–4817, Aug. 2016.

[2] J. Liu, M. Xing, H. Yu, G. Sun, and G. Sun, "EFTL: Complex convolutional networks with electromagnetic feature transfer learning for SAR target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5209811.

[3] Q. Dai, G. Zhang, Z. Fang, and B. Xue, "SAR target recognition with modified convolutional random vector functional link network," *IEEE Trans. Geosci. Remote Sens.*, vol. 19, 2022, Art. no. 4502205.

[4] J. Zhang et al., "Design and implementation of raw data compression system for subsurface detection SAR based on FPGA," *J. Geovis. Spatial Anal.*, vol. 4, 2020, Art. no. 2.

[5] A. O. de Albuquerque et al., "Dealing with clouds and seasonal changes for center pivot irrigation systems detection using instance segmentation in sentinel-2 time series," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8447–8457, 2021.

[6] R. Cao, Y. Wang, B. Zhao, and X. Lu, "Ship target imaging in airborne SAR system based on automatic image segmentation and ISAR technique," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1985–2000, 2021.

[7] H. Guo, X. Yang, N. Wang, and X. Gao, "A centernet++ model for ship detection in SAR images," *Pattern Recognit.*, vol. 112, no. 7, pp. 1–10, Apr. 2021.

[8] R. Shang et al., "SAR image segmentation based on constrained smoothing and hierarchical label correction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5102216.

[9] J. S. Lee, T. L. Ainsworth, and Y. Wang, "A review of polarimetric SAR speckle filtering," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 5303–5306.

[10] A. Aghabalaei, Y. Amerian, H. Ebadi, and Y. Maghsoudi, "Speckle noise reduction of time series SAR images based on wavelet transform and Kalman filter," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 625–628.

[11] 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.

[12] Y. Kwak, W. J. Song, and S. E. Kim, "Speckle-noise-invariant convolutional neural network for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 549–553, Apr. 2019.

[13] J. H. Cho and C. G. Park, "Multiple feature aggregation using convolutional neural networks for SAR image-based automatic target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1882–1886, Dec. 2018.

[14] R. Qin, X. Fu, J. Chang, and P. Lang, "Multilevel wavelet-srnet for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 4009005.

[15] J. Pei, Y. Huang, W. Huo, Y. Zhang, J. Yang, and T. S. Yeo, "SAR automatic target recognition based on multiview deep learning frame work," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2196–2210, Apr. 2018.

[16] Y. Li, L. Du, and D. Wei, "Multiscale CNN based on component analysis for SAR ATR," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5211212.

[17] Y. Guo, L. Du, D. Wei, and C. Li, "Robust SAR automatic target recognition via adversarial learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 716–729, 2020.

[18] N. Inkawhich et al., "Bridging a gap in SAR-ATR: Training on fully synthetic and testing on measured data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2942–2955, 2021.

[19] Z. Wang, L. Du, and Y. Li, "Boosting lightweight CNNs through network pruning and knowledge distillation for SAR target recognition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8386–8397, 2021.

[20] Z. Lin, K. Ji, M. Kang, X. Leng, and H. Zou, "Deep convolutional highway unit network for SAR target classification with limited labeled training data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1091–1095, Jul. 2017.

[21] Y. Tai, Y. Tan, S. Xiong, Z. Sun, and J. Tian, "Few-shot transfer learning for SAR image classification without extra SAR samples," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2240–2253, 2022.

[22] M. Zhang, W. Li, R. Tao, and S. Wang, "Transfer learning for optical and SAR data correspondence identification with limited training labels," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1545–1557, 2021.

[23] H. Li et al., "Adversarial examples for CNN-based SAR image classification: An experience study," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1333–1347, 2021.

[24] D. Hendrycks et al., “Using self-supervised learning can improve model robustness and uncertainty,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 1–12.

[25] Q. Hu, X. Wang, W. Hu, and G. J. Qi, “AdCo: Adversarial contrast for efficient learning of unsupervised representations from self-trained negative adversaries,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 1074–1083.

[26] M. Patacchiola and J. S. Amos, “Self-supervised relational reasoning for representation learning,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 4003–4014.

[27] D. Guo, Y. Xia, and X. Luo, “Self-supervised GANs with similarity loss for remote sensing image scene classification,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2508–2521, 2021.

[28] W. Li, H. Chen, and Z. Shi, “Semantic segmentation of remote sensing images with self-supervised multitask representation learning,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6438–6450, 2021.

[29] K. Li, Y. Qin, Q. Ling, Y. Wang, Z. Lin, and W. An, “Self-supervised deep subspace clustering for hyperspectral images with adaptive self-expressive coefficient matrix initialization,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3215–3227, 2021.

[30] J. Yue, L. Fang, H. Rahmani, and P. Ghamisi, “Self-supervised learning with adaptive distillation for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2022, Art. no. 5501813.

[31] Y. Yuan and L. Lin, “Self-supervised pretraining of transformers for satellite image time series classification,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 474–487, 2021.

[32] P. Zhang and A. A. Efros, “Colorful image colorization,” in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 649–666.

[33] M. Noroozi and P. Favaro, “Unsupervised learning of visual representations by solving jigsaw puzzles,” in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 69–84.

[34] W. Wang, J. Li, and H. Ji, “Self-supervised deep image restoration via adaptive stochastic gradient Langevin dynamics,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1989–1998.

[35] I. Misra and L. Maaten, “Self-supervised learning of pretext-invariant representations,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6707–6717.

[36] S. Gidaris, P. Singh, and N. Komodakis, “Unsupervised representation learning by predicting image rotations,” in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–9.

[37] R. Hadsell, S. Chopra, and Y. LeCun, “Dimensionality reduction by learning an invariant mapping,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 1735–1742.

[38] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, “Unsupervised feature learning via non-parametric instance discrimination,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3733–3742.

[39] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple frame-work for contrastive learning of visual representations,” in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[40] 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.*, 2020, pp. 9729–9738.

[41] D. T. Hoffmann et al., “Ranking info noise contrastive estimation: Boosting contrastive learning via ranked positives,” in *Proc. Amer. Assoc. Artif. Intell.*, 2022, pp. 897–905.

[42] J.-B. Grill et al., “Bootstrap your own latent—A new approach to self-supervised learning,” in *Proc Adv. Neural Inf. Process. Syst.*, 2020, pp. 21271–21284.

[43] X. Chen and K. He, “Exploring simple Siamese representation learning,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 15750–15758.

[44] M. Caron et al., “Unsupervised learning of visual features by contrasting cluster assignments,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 9912–9924.

[45] R. J. Chen, K. Chen, H. Chen, W. Li, Z. Zou, and Z. Shi, “Contrastive learning for fine-grained ship classification in remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4707916.

[46] Z. Jiang et al., “Improving contrastive learning on imbalanced data via open-world sampling,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 5997–6009.

[47] C. Chuang et al., “Robust contrastive learning against noisy views,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 16670–16681.

[48] P. Goyal et al., “Vision models are more robust and fair when pretrained on uncurated images without supervision,” 2022, *arXiv:2202.08360*.

[49] Y. Zhong et al., “Is self-supervised contrastive learning more robust than supervised learning?,” in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 1–16.

[50] M. Gierszewska and T. Berezowski, “On the role of polarimetric decomposition and speckle filtering methods for C-Band SAR wetland classification purposes,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2845–2860, 2022.

[51] R. Farhadiani, S. Homayouni, and A. Safari, “Hybrid SAR speckle reduction using complex wavelet shrinkage and non-local PCA-based filtering,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 5, pp. 1489–1496, May 2019.

[52] M. Yahia, T. Ali, M. M. Mortula, R. Abdelfattah, S. E. Mahdy, and N. S. Arampola, “Enhancement of SAR speckle denoising using the improved iterative filter,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 859–871, 2020.

[53] H. Wang, S. Chen, F. Xu, and Y. -Q. Jin, “Application of deep-learning algorithms to MSTAR data,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 3743–3745.

[54] L. van der Maaten and G. Hinton, “Visualizing data using t-SNE,” *Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[55] Y. Zhai et al., “MFFA-SARNET: Deep transferred multi-level feature fusion attention network with dual optimized loss for small-sample SAR AT R,” *Remote Sens.*, vol. 12, no. 9, pp. 1385–1404, Apr. 2020.

[56] 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, 2021.

[57] 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.

[58] C. Liu, H. Sun, Y. Xu, and G. Kuang, “Multi-source remote sensing pretraining based on contrastive self-supervised learning,” *Remote Sens.*, vol. 14, no. 18, pp. 4632–4651, Sep. 2022.

[59] R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.

Yikui Zhai (Member, IEEE) received the B.S. degree in optoelectronics and communications and the master’s degree in signal and information processing from Shantou University, Shantou, China, in 2004 and 2007, respectively, and the Ph.D. degree in signal and information processing from Beihang University, Beijing, China, in 2013.

Since October 2007, he has been with the Department of Intelligence Manufacturing, Wuyi University, Jiangmen, China, where he is currently a Professor. From June 2016 to June 2017, he was a Visiting Scholar with the Department of Computer Science, University of Milan, Milan, Italy. His research interests include image processing, deep learning, and pattern recognition.

Jinrui Liao (Graduate Student Member, IEEE) received the B.S. degree in electronic information engineering from the Wuyi University, Jiangmen, China, in 2019. He is currently working toward the master’s degree in computer technology with the Department of Intelligence Manufacturing, Wuyi University.

His research interests include image analysis, self-supervised learning, and pattern recognition.

Bing Sun (Member, IEEE) received the B.S. degree in electronic information engineering and Ph.D. degree in communication and information systems from the Beihang University (Beijing University of Aeronautics and Astronautics, BUAA), Beijing, China, in 2003 and 2008, respectively.

He has been with the School of Electronics and Information Engineering, Beihang University, since 2008. His research interests include signal processing, image processing, synthetic aperture radar system design, and processing algorithms.

Ziyi Jiang (Student Member, IEEE) received the B.S. degree in computer network in 2019 from the Wuyi University, Jiangmen, China, where he is currently working toward the master's degree in information and communication engineering with the Department of Intelligence Manufacturing.

His research interests include image processing, pattern recognition, and automatic measuring.

Zilu Ying received the B.S., M.S., and Ph.D. degrees in electrical information engineering from the Beihang University, Beijing, China, in 1985, 1988, and 2009, respectively.

He is currently a Full Professor with the Wuyi University, Jiangmen, China. He is also an Executive Director of the Guangdong Society of Image and Graphics and a member of the Signal Processing Branch of the Chinese Institute of Electronics. His research interests include biometric extraction and pattern recognition.

Wenqi Wang received the B.S. degree in communication engineering from the Henan University of Engineering, Kaifeng, China, in 2019. He is currently working toward the master's degree in information and communication engineering with the Department of Intelligence Manufacturing, Wuyi University, Jiangmen, China.

His research interests include computer vision, self-supervised contrastive learning, image processing, and pattern recognition.

Angelo Genovese (Senior Member, IEEE) received the Ph.D. degree in computer science from the Università degli Studi di Milano, Crema, Italy, in 2014.

He has been a Postdoctoral Research Fellow in computer science with the Università degli Studi di Milano, since 2014. He has been a Visiting Researcher with the University of Toronto, Toronto, ON, Canada. He has authored/coauthored more than 30 papers in international journals, proceedings of international conferences, books, and book chapters. His research interests include signal and image processing, 3-D reconstruction, computational intelligence technologies for biometric systems, industrial and environmental monitoring systems, and design methodologies and algorithms for self-adapting systems.

Dr. Genovese is an Associate Editor of the *Journal of Ambient Intelligence and Humanized Computing* (Springer).

Vincenzo Piuri (Fellow, IEEE) received the Ph.D. degree in computer engineering from the Politecnico di Milano, Milan, Italy, in 1989.

He was an Associate Professor with the Politecnico di Milano, from 1992 to 2000, and a Visiting Professor with the University of Texas at Austin, Austin, TX, USA (summers 1996–1999). He has been a Full Professor in computer engineering since 2000 and the Director of the Department of Information Technology, Università degli Studi di Milano, Milan, from 2007 to 2012. He has participated in several national and international research projects funded by the European Union, the Italian Ministry of Research, the National Research Council of Italy, the Italian Space Agency, and industries. He has authored/coauthored more than 350 papers in international journals, proceedings of international conferences, and book chapters. His research interests include biometrics, signal and image processing, pattern analysis and recognition, theory and industrial applications of neural networks, machine learning, intelligent measurement systems, industrial applications, fault tolerance, digital processing architectures, embedded systems, cryptographic architectures, and arithmetic architectures.

Prof. Piuri is a Distinguished Scientist of ACM and a Senior Member of INNS. He has been an Associate Editor of IEEE TRANSACTIONS ON NEURAL NETWORKS and IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT. He is an ACM Fellow.

Fabio Scotti (Senior Member, IEEE) received the Ph.D. degree in computer engineering from the Politecnico di Milano, Milan, Italy, in 2003.

He was an Assistant Professor with the Department of Information Technologies, Università degli Studi di Milano, Milan, Italy, from 2002 to 2015, where he was an Associate Professor with the Department of Computer Science from 2015 to 2020. He has been a Full Professor with the Università degli Studi di Milano, since 2020. He has authored/coauthored 150 papers in international journals, proceedings of international conferences, books, book chapters, and patents. His research interests include biometric systems, machine learning and computational intelligence, signal and image processing, theory and applications of neural networks, 3-D reconstruction, industrial applications, intelligent measurement systems, and high-level system design.

Prof. Scotti is an Associate Editor of IEEE TRANSACTIONS ON HUMAN–MACHINE SYSTEMS and IEEE OPEN JOURNAL OF SIGNAL PROCESSING. He is serving as a Book Editor (Area Editor, section Less-Constrained Biometrics) of the Encyclopedia of Cryptography, Security, and Privacy (3rd Edition, Springer). He has been an Associate Editor of IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, *Soft Computing* (Springer) and a Guest Coeditor of IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.