LDCL: Low-Confidence Discriminant Contrastive Learning for Small-Sample SAR ATR

Jinrui Liao, *Student Member, IEEE*, Yikui Zhai, *Senior Member, IEEE*, Qingsong Wang, Bing Sun, *Member, IEEE*, and Vincenzo Piuri, *Fellow, IEEE*

***Abstract*— Synthetic aperture radar (SAR) target image acquisition presents challenges and incurs high annotation costs. The emergence of self-supervised contrastive learning shows promise for SAR automatic target recognition (ATR) with limited data. However, SAR images suffer from poor discriminability and high sample similarity, hindering instance discrimination in contrastive learning. To address this, we propose low-confidence discriminant contrastive learning (LDCL), which integrates group-instance contrast and batch-mixed training for SAR ATR. LDCL consists of two branches: classical instance discrimination and group-instance discrimination. We refine the SAR-group instance discrimination loss function by incorporating distance calculations to guide feature vectors toward the nearest clusters, enhancing discrimination within the feature space. In addition, we introduce a batch image mixing training strategy to reduce confidence in SAR instance discrimination while preserving intraclass consistency. Experimental results on small sample MSTAR and FUSAR-Ship datasets demonstrate that LDCL outperforms traditional transfer learning and self-supervised learning (SSL) methods, achieving significantly higher recognition rates in SAR ATR tasks.**

***Index Terms*— Batch-mixed training, contrastive learning, group-instance discrimination, synthetic aperture radar automatic target recognition (SAR ATR).**

Manuscript received 20 July 2023; revised 15 November 2023 and 14 December 2023; accepted 29 December 2023. Date of publication 5 January 2024; date of current version 19 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62273365; in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515011576; in part by the Guangdong, Hong Kong, Macao, and the Greater Bay Area International Science and Technology Innovation Cooperation Project under Grant 2021A0505030080 and Grant 2021A0505060011; in part by the Xiaomi Young Talents Program; in part by the Key Research Projects for the Universities of Guangdong Provincial Education Department under Grant 2022ZDZX1032 and Grant 2023ZDZX1029; in part by the Jiangmen Science and Technology Plan Project under Grant 2220002000246; and in part by the Wuyi University and Guangdong–Hong Kong–Macao Greater Bay Joint Research and Development Foundation under Grant 2022WGALH19. *(Jinrui Liao, Yikui Zhai, and Qingsong Wang contributed equally to this work.) (Corresponding author: Bing Sun.)*

Jinrui Liao is with the Department of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, China, and also with the School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen 518107, China (e-mail: jinrui_liao@163.com).

Yikui Zhai is with the Department of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, China (e-mail: yikuizhai@163.com).

Qingsong Wang is with the School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen 518107, China (e-mail: wangqs5@mail.sysu.edu.cn).

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

Vincenzo Piuri is with the Department of Computer Science and Diparti-mento di Informatica, Università Degli Studi di Milano, 20122 Milan, Italy (e-mail: vincenzo.piuri@unimi.it).

Digital Object Identifier 10.1109/TGRS.2024.3350196

I. Introduction

SYNTHETIC aperture radar (SAR) is a sophisticated radar imaging system that utilizes a wide frequency range to acquire high-resolution imagery, irrespective of atmospheric conditions or temporal variations. Unlike conventional radars with a single aperture, SAR overcomes these limitations by utilizing synthetic aperture synthesis. SAR achieves high-resolution image generation by synthesizing an effective aperture that exceeds the physical aperture size. Consequently, SAR can capture detailed information regarding the shape, structure, and scattering properties of targets, even when they are located at significant distances from the sensor [1], [2], [3]. SAR ATR technology is crucial in military applications as it plays a vital role in battlefield reconnaissance, situational awareness, detection, and monitoring [4], [5], [6]. Compared to optical images, single-polarization SAR images are gray-scale and show significant anisotropy. This characteristic renders them vulnerable to clutter interference, posing challenges for ATR tasks [7], [8], [9]. Recently, the field of SAR ATR has witnessed remarkable advancements, particularly with the emergence of convolutional neural network (CNN) methods. These CNN-based approaches have demonstrated excellent performance and have been successfully applied to various SAR ATR applications. An example of such progress is the work by Chen et al. [10], who introduced A-ConvNets, a CNN architecture constructed using fully sparse network layers. Their approach achieved highly competitive recognition performance on the MSTAR dataset, which comprises SAR images with ten different target classes. By leveraging the power of CNNs and exploiting the unique characteristics of SAR data, A-ConvNets demonstrated the possibility of deep learning techniques in SAR ATR, paving the way for further advancements in the field. Zhou et al. [11] conducted a study to combine the electromagnetic scattering properties of SAR images with a CNN. They considered specific attributes of scattering centers to extract discriminative features in SAR automatic target recognition (ATR). Their approach demonstrated the superiority of their method by showcasing improved performance compared to existing techniques. Zhang et al. [12] introduced a feature fusion framework (FEC) that combines the functionality of scattering centers and Deep CNNs to achieve competitive recognition rates in both standard operating condition (SOC) and extended operating condition (EOC) scenarios. However, it's crucial to note that SAR ATR approaches based on deep learning often

demand a substantial volume of training data. In cases where specific target classes only have a limited number of training samples, recognition accuracy can significantly decrease. For instance, studies have shown that deep learning methods yield recognition accuracies below 40% in SAR ATR tasks when there are only a few tens of training samples available for each target [1]. In practical scenarios, especially in military reconnaissance or homeland security applications, obtaining sufficient training samples can be exceptionally challenging, rendering existing SAR ATR algorithms ineffective.

Given the inherent challenges of SAR imagery and the constraints of deep learning due to limited training samples, researchers have embarked on devising innovative methodologies. A notable approach is the CAE-HL-CNN proposed by Qin et al. [13], which synergizes semigreedy convolutional auto-encoders (CAEs) with a hinge loss CNN. This method specifically targets the overfitting issues associated with scarce training data while ensuring robust feature extraction from SAR images. Sun et al. [14] advanced a technique leveraging CAEs for SAR image reconstruction, subsequently fine-tuning the network with minimal annotated data. This strategy capitalizes on the CAEs' reconstruction prowess to augment feature learning, addressing the challenges of limited samples. Another noteworthy method [15] employs a neural network to generate pseudo-labels from sparse training data, facilitating the creation of supplementary training samples. Explorations involving generative adversarial networks (GANs) [16], transfer learning [17], and graph neural network-based strategies [18] have also been conducted in the field of small-sample SAR ATR. However, GANs occasionally grapple with issues like mode collapse and suboptimal image generation, especially with limited data. While transfer learning offers promise, it necessitates a substantial volume of annotated source data and presumes minimal distribution disparity between source and target domains. Graph neural network-based SAR ATR approaches, despite their potential, often overemphasize node relationships, lacking SAR data specialization and potentially compromising performance. Collectively, these endeavors epitomize the ongoing efforts to surmount the challenges of limited training samples in SAR ATR, with the overarching aim of enhancing recognition accuracy in real-world scenarios.

For the challenges posed by limited training samples in small-sample SAR visual tasks, self-supervised learning (SSL) has shown to be a potential option. SSL creates proxy tasks for unsupervised representation learning with vast volumes of unlabeled data [19], [20]. Then, a pretrained model using the learned representations is created and utilized to fine-tune downstream tasks, enabling the completion of visual tasks with insufficient labeled data and mitigating overfitting. Contrastive learning, a widely studied method in SSL, has made significant progress. The core idea behind contrastive SSL is to force view representations of different images further apart (negative pairs) while bringing different view representations of the same image closer together (positive pairs). This approach aims to achieve both invariance and discriminative feature expressions. Contrastive learning methods can be categorized into those with negative samples (e.g., SimCLR [21], CMC [22], and MoCo [23]) and those without (e.g., BYOL [24] and SimSiam [25]). These self-supervised contrastive learning approaches have demonstrated results comparable to supervised learning in downstream tasks. While initial endeavors have been made in applying SSL methodologies to SAR ATR, notable among them are the efforts by Wen et al. [26], who introduced rotation prediction as a surrogate task within an SSL paradigm, subsequently facilitating training in scenarios characterized by a scarcity of SAR ATR samples. Wang et al. [15] advanced this field by integrating a Siamese network architecture for the extraction of representations from SAR target images, achieving significant enhancements in SAR classification efficacy, particularly under conditions of limited training datasets. Furthermore, Liu et al. [27] explored the efficacy of a self-supervised contrastive learning framework in the context of SAR-optical imagery, demonstrating its utility in the classification of multisource image scenes. However, it is worth noting that there are limitations to the direct application of contrast learning methods on SAR images, as they are nonnatural images and significantly challenging. SAR images suffer from speckle noise and lack of strong discriminative features. Different classes in SAR images often exhibit high similarity (as shown in Fig. 1), making instance discrimination more difficult. Traditional contrastive learning treats the contrastive strength of samples within and between classes as equal and disregards the inherent similarity within a class [as shown in Fig. 1(a)]. The geometric shapes of samples within the same class in SAR images may resemble samples from different classes. Moreover, the MSTAR dataset, commonly used in SAR ATR tasks, shows certain similarities between instances of different classes after data augmentation, leading to potential misclassifications. Fig. 1(b) shows an example where the confidence of discriminating a sample of the ZIL131 target at an azimuth angle of 30° from another instance of the same target at an azimuth angle of 60° is considered equal to the confidence of discriminating from other targets at different azimuth angles. Intuitively, the attraction strength of intraclass samples should be higher, indicating lower confidence in instance discrimination; while the strength of interclass samples should be lower, indicating higher confidence. Moreover, the MSTAR dataset exhibits certain similarities between instances of different classes (e.g., the BMP2 and BRDM2 at an azimuth angle of 120° have high similarity in imaging views). Due to their high similarity after data augmentation, different SAR instances can be misclassified as the same instance, which adversely affects the performance of downstream tasks and hinders the generalization of SSL to SAR image recognition tasks.

In light of the challenges delineated earlier, our research pivots toward harnessing the potential of self-supervised contrastive learning specifically for SAR ATR. We introduce the low-confidence discriminant contrastive learning (LDCL) framework, a novel instance comparison learning paradigm designed to modulate the confidence level associated with SAR instance discriminations, as illustrated in Fig. 2. LDCL meticulously addresses the dual facets of confidence: the confidence associated with instance discrimination in the feature space and the confidence inherent to label assignment

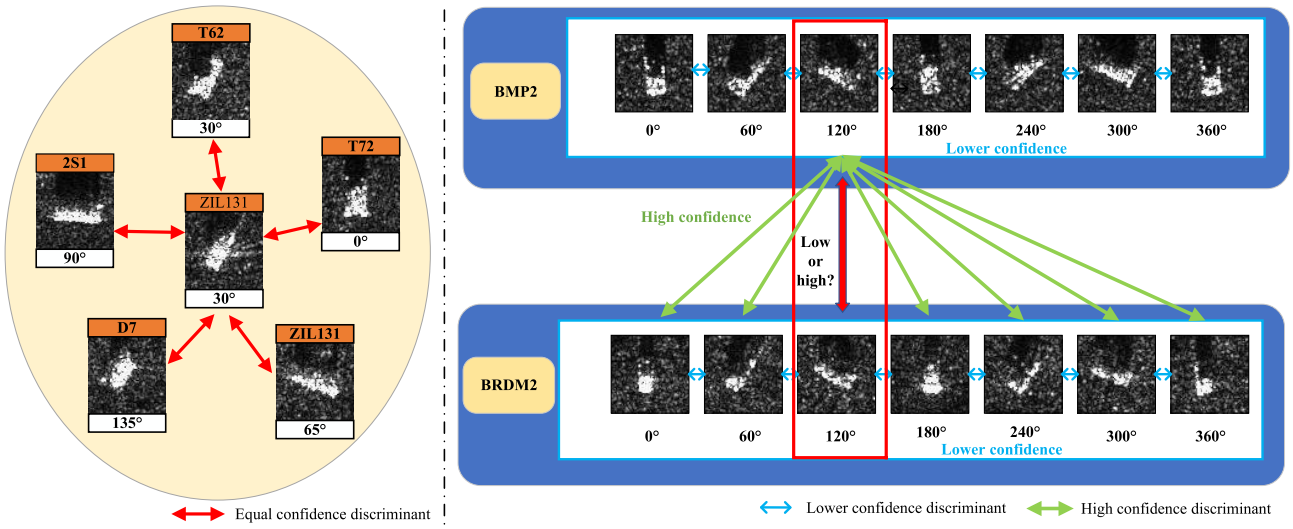

Fig. 1. (a) Traditional method of instance discrimination assigns equal confidence in discrimination to all SAR target samples. (b) In our hypothesis, SAR instances are assigned high and low confidence levels for instance discrimination. However, the high similarity of SAR samples between targets with the same rotation angle introduces confusion in the discriminative self-supervised instance discrimination process.

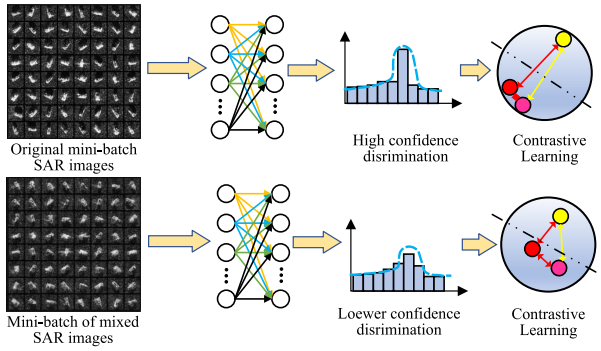

Fig. 2. Concept of low-confidence discrimination for SAR self-supervised instance discrimination.

in the input space. To achieve this, we have architected a batch-mixing SSL framework that seamlessly amalgamates instance discrimination with group discrimination. This design philosophy capitalizes on the inherent similarities among SAR instances, with the overarching goal of tempering the sensitivity of SAR instance discrimination within contrastive learning, thereby bolstering the performance of downstream SAR image representation tasks. By judiciously exploiting instance similarity, our framework augments the robustness and efficacy of the learning process. The salient contributions and innovations presented in this article are as follows.

1) Recognizing the inherent confidence discrepancies among SAR instances in the feature space, we introduce a novel contrastive loss function. This function operates within the feature space and aims to temper the confidence of instance discrimination. By integrating both group discrimination and instance discrimination paradigms, it places a heightened emphasis on challenging SAR samples, ensuring a balanced and nuanced learning process.
2) Venturing into the input space, we put forth a strategic batch-mixing instance training methodology. Predicated on mixing samples from different instances within the same class, this approach retains the intrinsic correlation information and enhances the learning of SAR image representations. Experimental results confirm the effectiveness of the subsequent fine-tuning phase and demonstrate improved accuracy for SAR target recognition tasks. Our strategy enhances the network's discriminative ability in SAR target recognition.
3) Our method achieves competitive small sample recognition performance on the MSTAR and Fusarship datasets and outperforms other methods under various conditions such as model variations, changes in elevation angle, and imbalanced samples. This validates our SSL solution for a small sample SAR ATR, providing promising advancements for SAR image recognition.

II. Related Work

This section thoroughly reviews the existing literature on SAR ATR, specifically highlighting methods tailored for small sample SAR ATR. Furthermore, we will discuss relevant research in SSL and explore its relevance to our proposed method.

A. Small Sample SAR ATR

Deep learning has achieved remarkable success in SAR ATR tasks by leveraging large-scale datasets for training. However, due to challenges in SAR data acquisition and labor-intensive manual annotation, it's challenging to obtain finely annotated large-scale target samples for SAR classification. Limited training samples hinder SAR target classification to perform well and make progress. As a result, there has been an increasing emphasis on small-sample SAR-ATR research. To address the issue of limited samples, researchers have

predominantly utilized data augmentation and transfer learning methods in small-sample learning for SAR-ATR. Approaches such as [28] and [29] tackle this issue from a data perspective. Data augmentation techniques play a crucial role in expanding the dataset by applying traditional or automatic augmentation operations, as well as generating new samples using simulators or generative models. On the other hand, transfer learning utilizes knowledge gained from abundant training data in a source domain to address the sample constraint in the target domain. In the context of small sample SAR ATR, researchers have employed a variety of techniques to overcome the challenge of limited training samples. Data augmentation operations, encompassing both traditional methods and automatic techniques, have been utilized to augment the dataset and enhance sample diversity. In addition, simulators or generative models have been utilized to generate new samples, as discussed in [30] and [31]. Transfer learning has shown advantages in tasks with limited samples, such as [17] and [32]. By leveraging knowledge learned from abundant training data in a source domain, transfer learning addresses sample constraints in the target domain. However, it should be noted that transfer learning still needs substantial annotated data from the source domain during the training phase. In small sample SAR image recognition, successful methods from natural image recognition have been adapted. For instance, Tang et al. proposed improvements to Siamese networks by preserving the classifier and similarity discriminator, effectively utilizing metric learning to reduce network prediction time, as described in [33]. Cai et al. introduced spatial transformation to enhance the prototype network, extracting semantic information from SAR images and enhancing SAR ATR performance under limited sample conditions, as presented in [34]. Wang et al. proposed a novel hierarchically designed lightweight method (HDLM) that addresses the issue of limited data in SAR ATR through label recognition and feature discrimination, as demonstrated in [35]. These methods demonstrate the adaptation and application of successful techniques in natural image recognition to address the challenges of small-sample SAR image recognition. They contribute to the improvement of SAR ATR performance under limited sample conditions.

B. Contrastive Learning

By maximizing mutual information between the learned representations and specific contexts, the original definition of contrastive learning emphasized learning unsupervised representations, as described by scholars [19], [20], [21], [22], [23], [24], [25]. In this approach, an instance sample learns the context of the sample by attracting its own augmented samples (positive samples) in the embedded high-dimensional feature space and repelling augmented samples that do not belong to itself (negative samples), which may contain other instance samples. Instance discrimination, achieved through minimal matching [36], memory banks, or dynamic queues with momentum updates [23], has frequently been employed as a pretext task to differentiate augmented samples. Negative samples play a crucial role in instance discrimination, gradually making the model more discriminative and better at distinguishing challenging samples. However, recent advancements have presented alternative approaches. Grill et al. [24] presented a view representation learning method named BYOL, which eliminates the need for negative samples and solely relies on incremental data from the same view. This highlights the dispensability of negative samples in certain scenarios. In addition, Ren et al. [37] observed the inherent correlation among mixed images sharing the same source image and proposed a data mixing strategy to enhance the performance of contrastive learning. These advancements have widened our comprehension of contrastive learning's mechanism and created fresh avenues for the improvement of learning representations.

III. Methodology

SAR image target recognition encounters challenges such as small interclass sample differences and intraclass imbalance [38], [39]. Directly applying contrastive learning networks from the optical image domain yields limited effectiveness in instance discrimination. To address this issue, we drew inspiration from fine-grained image clustering and recognition tasks [40], and presented a contrastive learning approach that incorporates instance discrimination and group discrimination. This approach explores the relationships between individual instances and group instances to enhance contrastive learning. Instance-group discrimination discriminates between instances and cluster centroids, and transforms "hard discrimination" into "soft discrimination," effectively mitigating erroneous discrimination of highly similar instances by the algorithm model and enhancing the model's feature extraction function. Fig. 3 exhibits the architecture of the group-instance discrimination contrastive learning approach. Initially, a batch of input SAR data is concatenated and blended. This batch comprises SAR target images from diverse categories. Subsequently, random data augmentation techniques are applied to the concatenated mixed samples. The augmented samples are then fed into a feature encoding network, such as ResNet18, which maintains a consistent network architecture [41]. The feature encoding network is responsible for learning feature vectors, which are then subjected to instance contrastive learning in one branch. Simultaneously, the other branch performs feature clustering and calculates the cluster centroids. These centroids represent the feature vectors of "instances" and are utilized for contrastive learning alongside instances from different branches. Next, we will present comprehensive explanations of each component of our proposed method.

A. Instances Discrimination Loss Function

InstDisc, as the prototype of instance discrimination-based contrastive learning, was initially influenced by Deep InfoMax. However, InstDisc only sampled one negative example for each positive sample, which may limited its effectiveness. Meanwhile, SimCLR addressed the issue of insufficient negative samples by incorporating extensive data augmentation and increasing the batch size [21]. It also defined cosine similarity as the measure of representation discrimination. In the context of SAR target images, we considered a set of n instances represented by x_i, where $T(x_i)$ and $T'(x_i)$

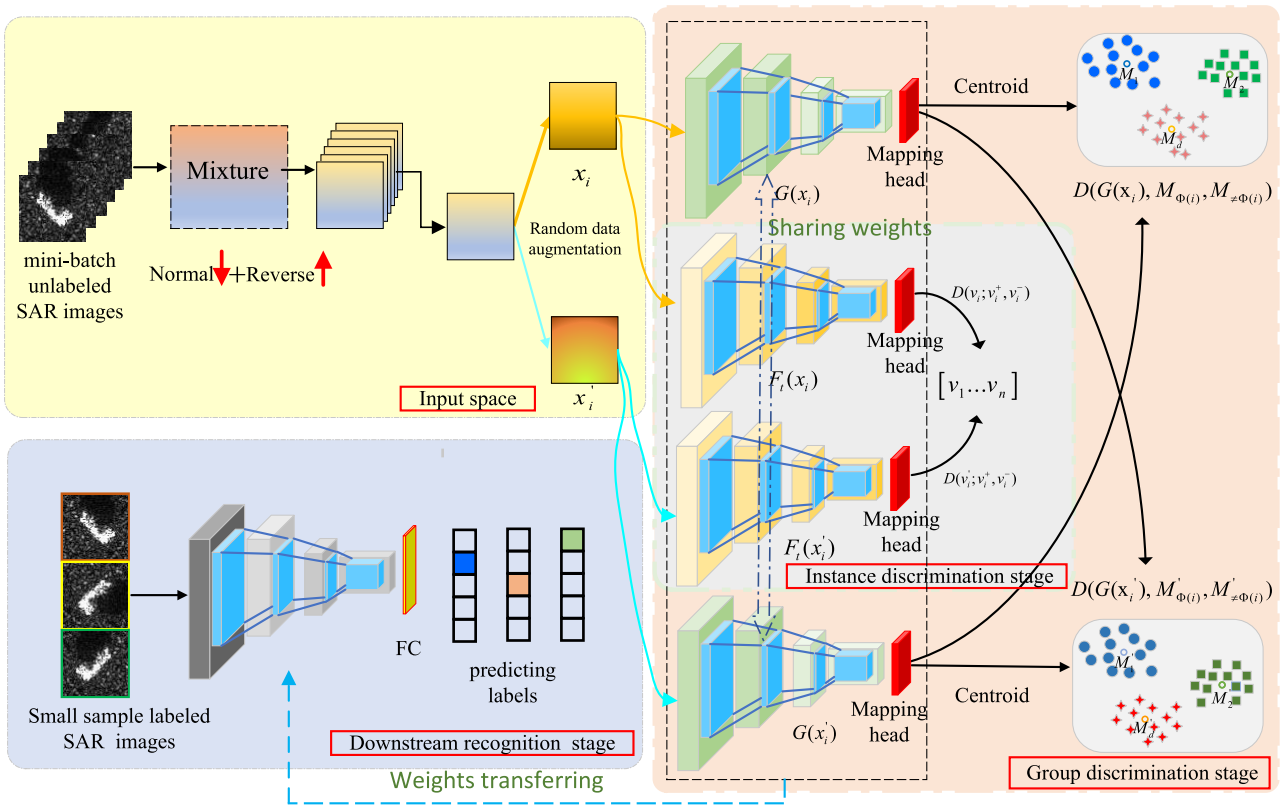

Fig. 3. Overall framework of the proposed LDCL algorithm. The flowchart depicts the LDCL framework, consisting of four distinct stages: data input, instance discrimination, group-instance discrimination, and downstream target recognition. In the data input stage, the predicted label confidence of SAR instance samples is smoothed through batch instance mixing, while augmentation is incorporated to produce target images with two different views. Subsequently, the instance discrimination branch and the group-instance discrimination branch are utilized to collectively diminish the discriminative confidence of contrastive loss in the feature space. This process employs an arbitrary CD backbone network to extract feature representations from the two views using a Siamese approach and map them to the feature space using a normalization mapping head. Finally, the acquired feature parameters are transferred to the downstream network, which includes an additional FC layer, enabling the completion of the small-sample SAR target recognition task by inputting a limited quantity of annotated SAR data.

denote the augmented versions of the images (e.g., through rotation). By applying a CNN with parameters θ, denoted as f, we mapped the $2n$ samples to a D-dimensional hypersphere, represented by $v_i = f(x_i)$, which represents the vectors in the feature space after CNN mapping. Symbols v_i^+ and v_i^- represent positive and negative samples, respectively. Using contrastive learning, we updated the model parameters to minimize the loss function l

$$l\left(v_i, v_i^+, v_{\neq i}^-\right) = -\log \frac{\exp \frac{\text{sim}\left(v_i, v_i^+\right)}{T}}{\exp \frac{\text{sim}\left(v_i, v_i^+\right)}{T} + \sum_{j \neq i} \exp \frac{\text{sim}\left(v_i, v_j^+\right)}{T}}. \tag{1}$$

In the given context, the function sim(,) represents the cosine similarity function, and l corresponds to the noise contrastive estimation (NCE) loss for softmax instance classification. This loss encourages positive samples in the feature space to approach each other while pushing negative samples away. The hyperparameter T is the temperature parameter, which adjusts the discriminative distance.

B. Loss Function for SAR Image Group-Instance Discrimination

SAR images are known for their complexity of formation mechanisms and sensitivity to imaging angles, resulting in significant intraclass differences for the same target and small interclass differences between different targets. Consequently, relying solely on the instance discrimination loss in contrastive learning is insufficient for capturing the discriminative information of SAR images, particularly for samples captured under different azimuth angles. This limitation can significantly impact the performance of downstream SAR tasks. This section aims to address this challenge by thoroughly exploring the similarities between instances and incorporating them into contrastive learning. We proposed the concept of instance-group discrimination, which investigates the relations between group instances and individual instances. By applying contrastive loss to both instance discrimination and group center discrimination, we aimed to enhance self-supervised contrastive learning. To improve the representational performance of challenging positive and negative samples in SAR images and enhance the feature learning capability for SAR image recognition tasks, we introduced a group instance discrimination branch that complements the instance discrimination function f. SAR instances $x_i (i = 1, 2, \ldots 2N)$ were passed through a shared CNN, resulting in feature vectors $G(x_i)$ and $G(x'_i)$ from the group branch for different views. Next, we computed N cluster centers $M_1, \ldots, M_N$ and assigned each instance to its nearest center using a clustering method such as the K-means algorithm. We denoted the assignment of instance i to the jth centroid as $\Phi(i) = j$. Then, we aimed

to predict the correct probability distribution for $G(x'_i)$ being assigned to the label cluster center $M_{\Phi(i)}$. This prediction was accomplished by minimizing the cross-entropy loss function, as shown in the following equation:

$$-E_p\left[\log_q\right] = \sum_i l\left(G\left(\mathrm{x}'_i\right), M_{\Phi(i)}, M_{\neq\Phi(i)}; T_G\right). \quad (2)$$

To enable a soft clustering of similar SAR samples, we extended the assignment of $G(x_i)$ and $G(x'i)$ to the cluster center $M'_\Phi(i)$. This soft clustering approach allows a more flexible representation of similarity between SAR instances. By combining the SAR instance discrimination loss and the group discrimination loss functions, we proposed an optimization loss function as shown in (3), bottom of the next page. This loss function incorporates both the instance-level and group-level discriminative information, leading to improved learning and representation capabilities for SAR image recognition tasks.

In the optimization loss function, α represents the weight assigned to the group discrimination loss function. The SAR instance discrimination term encourages the instance x_i to be dissimilar from other instances x_j, promoting distinct representations. Then, the SAR group discrimination term attracts x_i toward its corresponding group centroid and encourages the alignment of $G(x_i)$ and $G(x'_i)$. This process facilitates the clustering of similar samples together in the representation space, enhancing the discriminative power for challenging samples. By balancing the contributions of both terms, our approach has achieved a more comprehensive and effective learning of SAR image representations.

C. Fusion of Mixed-Batch Unsupervised Training Strategy

In the unsupervised training process, we introduced the concept of distance in the label space to capture the high similarity among SAR image samples. This allowed the model to perceive the soft similarity of positive and negative samples in the feature space and label space. Building upon the work of Shen et al., we incorporated batch mixing as a strategy during training to effectively learn more subtle and robust representations in the label space for SAR images after data augmentation. Specifically, in a batch of SAR images x_i (where $i = 1, 2, \ldots N$) input into a specific branch, each sample undergoes the image mixing strategy described as follows:

$$x_{\text{mix}} = \lambda x_i + (1-\lambda)x_{i+1}. \quad (4)$$

As a result, the distances between positive and negative samples in the label space will be transformed as follows:

$$d\left(x_{\text{mix}}, \hat{x}\right) = \begin{cases} \lambda & \text{if } \hat{x} = \widehat{x_i} \\ (1-\lambda) & \text{if } \hat{x} = \hat{x}_{i+1}. \end{cases} \quad (5)$$

In the mixing strategy, $\widehat{x_i}$ and $\widehat{x_{i+1}}$ represent different views of the same image x_i. The pairs $(x_i, \widehat{x_i})$ and $(x_i, \widehat{x_{i+1}})$ form positive sample pairs in the contrastive learning framework. The mixing rate α is determined by the mixing degree used in the current training iteration. To manage the memory space efficiently, we employed the memory banks strategy, which helped maintain an appropriate number of negative sample pairs. This ensured a balanced positive/negative ratio for contrastive learning and facilitated better invariant mapping. Both the SAR instance discrimination and SAR group discrimination branches utilize the momentum update strategy, which is described by the following equations:

$$\theta_{fkI} \leftarrow m\theta_{fkI} + (1-m)\theta_{fqd} \quad (6)$$

$$\theta_{fkG} \leftarrow m\theta_{fkG} + (1-m)\theta_{fqG}. \quad (7)$$

In the context provided, θ_{fqI} and θ_{fqG} represent the initial parameters of the instance discrimination branch and the group discrimination branch, respectively. During the training process, θ_{fqI} and θ_{fqG} are the branches that underwent backpropagation for instance discrimination and group discrimination, respectively. The parameters of these two branches were updated using the momentum mechanism. The momentum coefficient, denoted as $m \in [0, 1)$, is of vital importance for the final fine-tuning and recognition accuracy of the model.

The mixed training strategy in each batch involves combining the first and last images, as well as the second and second-to-last images (penultimate), in order to adjust the distances between samples in the input space. This mixing order is used for the convenience of experimentation. The visual representation of this mixing order can be seen in Figs. 4 and 5.

After incorporating the batch sample mixing training strategy, the overall loss function $\mathcal{L}_{\text{total}}$ is given by (8). In this equation, $\mathcal{L}_{\text{m}}$ denotes the loss function that incorporates the batch mixing training strategy, where $X_M(\downarrow)$ represents normal sequence mixing, and $X_M(\uparrow)$ indicates reverse sequence mixing. $X_{IM}(\downarrow)$ represents the mixture used for instance discrimination in the normal sequence, while $X_{IM}(\uparrow)$ represents the mixture used for instance discrimination in the reverse sequence. $X_{GM}(\downarrow)$ and $X_{GM}(\uparrow)$ for group discrimination are the same. λ is a hyperparameter used to control the confidence level of SAR image input space

$$\begin{aligned} \mathcal{L}_{\text{total}} &= \mathcal{L}_{\text{ori}}\left(X, \widehat{X}\right) + \lambda\mathcal{L}_m\left(X_M(\downarrow), \widehat{X}\right) \\ &\quad + (1-\lambda)\mathcal{L}_m\left(X_M(\uparrow), \widehat{X}\right) \\ &= \mathcal{L}_{\text{ori}}\left(X_I, X_G; \widehat{X}_I, \widehat{X}_G\right) \\ &\quad + \lambda\mathcal{L}_m\left(X_{IM}(\downarrow), X_{GM}(\downarrow); \widehat{X}_I, \widehat{X}_G\right) \\ &\quad + (1-\lambda)\mathcal{L}_m\left(X_{IM}(\uparrow), X_{GM}(\uparrow); \widehat{X}_I, \widehat{X}_G\right). \end{aligned} \quad (8)$$

D. SAR Instance Sample Normalization Mapping Head

In mainstream contrastive learning, the latent feature $f(x)$ is initially projected onto the unit hypersphere using a projection head and subsequently normalized, yielding the instance feature $V = f(x)$. To ensure the preservation of the invariant mapping of SAR image features following data augmentation, we introduced an SAR instance-based normalization mapping head. The layer weights of the feature extraction network after the linear layer are denoted as W. We projected the SAR instance vector v_i onto W and computed its cosine similarity. After sharing the feature extraction network f or G, the

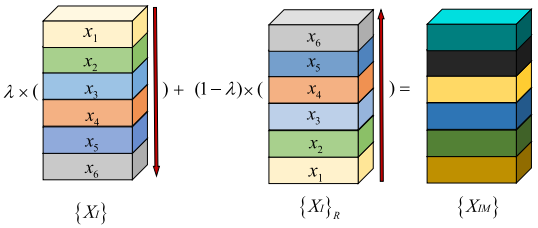

Fig. 4. Batch-instance mixing process.

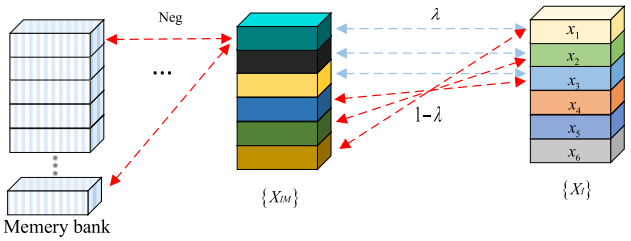

Fig. 5. Relationship between the confidence of the batch-mixed samples and the original samples.

normalized mapping of the tth component of the SAR image feature $F_t(x_i)$ is given by the provided equation

$$F_t(x_i) = < \frac{W_t}{\|W_t\|}, \frac{v_i}{\|v_i\|} > . \tag{9}$$

In this context, $F_t(x_i)$ represents the tth component of the SAR image feature, v_i corresponds to the SAR instance vector, and W indicates the layer weights of the feature extraction network after the linear layer. $\|v_i\|$ and $\|W_t\|$ denote the L2 norm (Euclidean norm) of vector v_i and W, respectively. Suppose the feature vector v_i has a dimension of d, and the weight matrix W has dimensions of $k \times d$, where k represents the dimension of the projected feature. During projection, the feature vector v_i is projected onto the weight matrix W using matrix multiplication, resulting in the projected feature vector n, which has a dimension of k. The normalization mapping head plays a crucial role in extracting robust image information. It ensures that the features extracted from SAR images are normalized and remain invariant even with the application of data augmentation techniques, enhancing the network's capacity to capture distinctive information and thus improving the overall effectiveness of tasks related to SAR image recognition.

The pretraining strategy of the proposed pseudocode is depicted in Algorithm 1. The batch-mixing training strategy is applied in the input space, whereas the group-instance feature discrimination is conducted in the feature space. By combining these two strategies, the algorithm facilitates SSL of low-confidence SAR instances.

Algorithm 1 Pseudocode of LDCL

Input: Unlabeled SAR ATR task dataset $D = \{x_i\}_{i=1}^N$, feature extracting network f batch size N,

for sampled minibatch $\{x_k\}_{k=1}^N$ **do**
  $i\_r = \text{flip}\{x_k\}_{k=1}^N$ // batch image reverse order processing
  $m\_batch = lam * X_k + (1 - lam) * i\_r$
  for all $k \in \{1, \ldots, N\}$ **do**
    //draw two augmentation functions $t \sim T, t' \sim T$
    $x_I = t(x_k)$, $\widetilde{x}_I = t'(x_k)$, $x_G = t(x_k)$, $\widetilde{x}_G = t'(x_k)$
    $h_I = f(x_I)$, $z_I = g(h_I)$
    $\widetilde{h}_I = f(\widetilde{x}_I)$, $\widetilde{z}_I = g(\widetilde{h}_I)$//repersentation and projection for instance branch
    $h_G = f(x_G)$,$z_G = g(h_G)$
    $\widetilde{h}_G = f(\widetilde{x}_G)$, $\widetilde{z}_G = g(\widetilde{h}_G)$//repersentation and projection for group branch
    //draw a cluster functions C, M_Φ is the cluster centroid
    $M_{\Phi(i)} \leftarrow C(z_G^1, \ldots z_G^N)$, $M'_{\Phi(i)} \leftarrow C(z_G^1, \ldots z_G^N)$
  end for
  for all $i \in \{1, \ldots, 2N\}$ $j \in \{1, \ldots, 2N\}$ **do**
    $sim = z_i^{\mathrm{T}} z_j / (\|z_i\| \|z_j\|)$
  end for
  define $l(v_i, v_i^+, v_i^-, T)$ based on Equation (1)
  Update networks f and g to minimize loss $\mathcal{L}_{total}$ based on Equation (8)
  //momentum update
  Update θ_{fkI} based on Equation (6)
  Update θ_{fkG} based on Equation (7)
end for
return encoder network $f(\cdot)$, and discard $g(\cdot)$.

IV. Experiment and Discussion

We evaluated our algorithm's performance in small-sample SAR target recognition using a unified set of MSTAR dataset and a subset of FUSAR-Ship dataset.

A. Dataset and Experimental Settings

1) MSTAR Dataset: The Defense Advanced Research Projects Agency (DARPA) supports the collecting of ground-based stationary target data known as the MSTAR dataset using SAR [42]. This dataset comprises multiple vehicle targets captured at various aspect angles. It includes ten distinct ground targets under SOCs, such as armored vehicles (BRDM2, BTR60, BTR70, BMP2, D7, and ZIL131), artillery (ZSU234 and 2S1), and tanks (T62 and T72). Some targets also have variations, for instance, T72 with three variants (812, 132, and S7) and BRDM2 with three variants (9563, C21,

$$\mathcal{L}(v, G(x); T_I; T_G; \alpha) = \sum_i \underbrace{l\left(v_i, v_i^+, v_i^-; T_I\right) + l\left(v'_i, v_i^+, v_i^-; T_I\right)}_{\text{SAR-ID}} + \alpha \sum_i \underbrace{l\left(G\left(\mathrm{x}'_i\right), M_{\Phi(i)}, M_{\neq\Phi(i)}; T_G\right) + l\left(G(\mathrm{x}_i), M'_{\Phi(i)}, M'_{\neq\Phi(i)}; T_G\right)}_{\text{SAR-IGD}} \tag{3}$$

TABLE I
MSTAR DATASET

Target	Serial	Trainging Set		Serial	Tesing 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
BRDM2	E71	17	298	E71	15	274
2S1	B01	17	299	B01	15	274

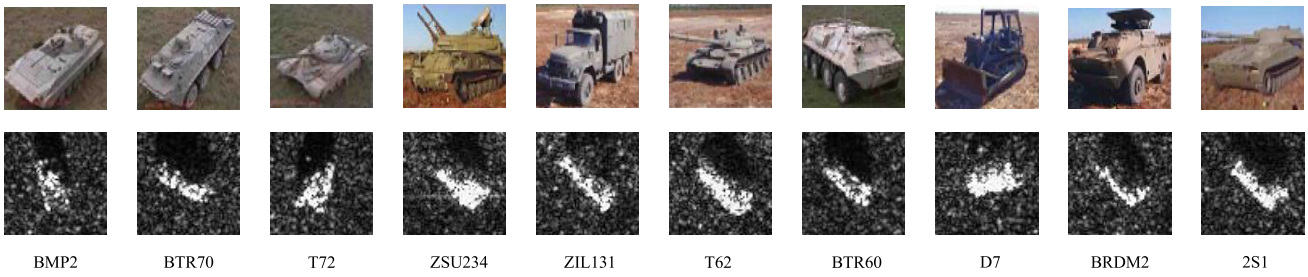

Fig. 6. MSTAR dataset consists of various classes of SAR targets along with their corresponding optical target images.

and 9566). The SAR target images in the MSTAR dataset have a resolution of 0.3 × 0.3 m, with a pixel size of 128 × 128. To mitigate background interference, the images were cropped to 64 × 64 with the target centered. Fig. 6 displays optical target images along with their corresponding SAR target images. Further information on the ten target classes can be found in Table I.

2) FUSAR-Ship Dataset: The FUSAR-Ship dataset, introduced by Hou et al. [43], is derived from the Gaofen-3 (GF-3) satellite and specifically designed for SAR maritime remote sensing and ocean monitoring applications. This dataset encompasses a wide range of ship categories and provides data on targets in various scenes, including ocean, land, coastline, river, and island scenes. For our study, we selected a subset of this dataset following the experimental setup described in the referenced papers. The chosen subset consists of seven categories of nearshore targets, namely, bridges, land patches, coastal land, and reefs, strong false alarm targets (such as buoys, windmills, and offshore fish farms), ocean patches, ships, and ocean waves. More specific details about the dataset are presented in Table II and Fig. 7. It is worth noting that this dataset exhibits notable interclass similarity among the different categories.

3) Small Sample Target Recognition Tasks: The small-sample SAR target recognition task data is obtained by undersampling the original training samples at ratios of 1:2, 1:3, 1:4, 1:8, 1:16, and 1:32. Fig. 8 depicts a schematic illustration of the undersampling process. The datasets generated through this undersampling method are referred

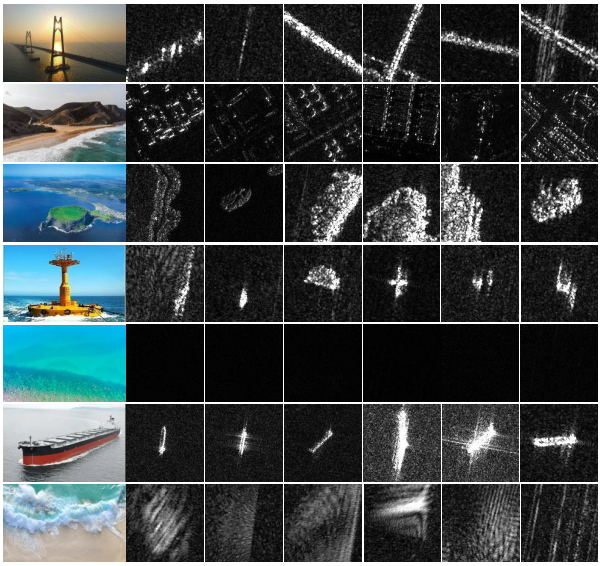

Fig. 7. FUSAR maritime object targets. From top to bottom, they are bridge, land fast, coastal land and islands, strong false alarm target, ocean patch, ship, and waves and sea clutter, totaling seven targets.

TABLE II
MARITIME TARGETS IN THE FUSAR-SHIP DATASET

	Ships	Strong scatterers	Bridges&coastlines	Coastal lands&islands	Sea clutter waves	Sea patches	Land patches
Training data	1296	229	1023	707	1377	1250	1137
Testing data	555	128	438	303	590	535	487
Dataset	1851	427	1461	1010	1967	1785	1624

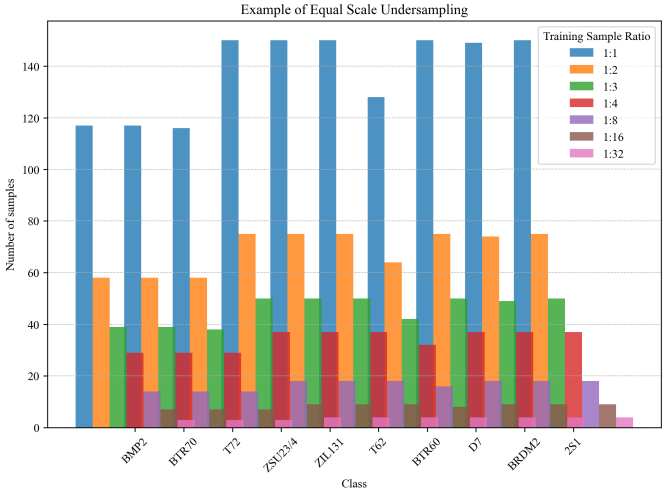

Fig. 8. Undersampling method employed in this study.

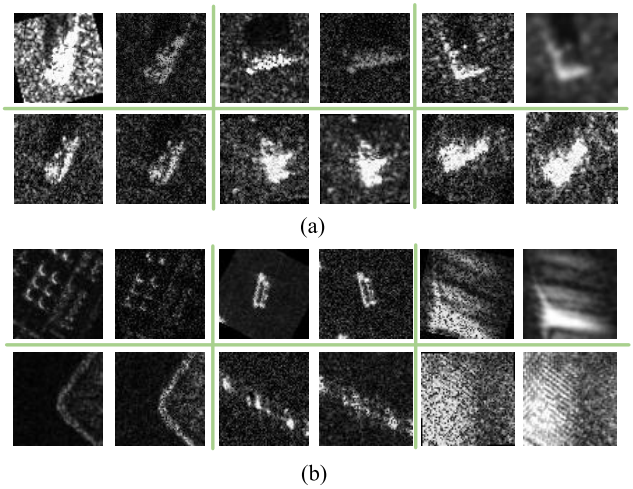

Fig. 9. Examples of augmented sample pairs were selected from both (a) MSTSAR database and (b) FuSARShip database.

to as small-sample conditions 1–6 (simplified notation as S1–S6). It is important to note that the small-sample approach applies exclusively to the training samples. In addition to the classical SAR target recognition task under SOC, this article introduces additional tasks to increase the difficulty and adapt to various application scenarios of SAR target recognition. These additional tasks include SAR fine-grained condition, large elevation angle variation condition, and generalization condition. Table III provides detailed information regarding the datasets required for the small sample target recognition tasks in different scenarios and task conditions. Moreover, apart from conducting experiments on equi-proportional undersampling, we also explored the utilization of unbalanced data to form small sample data for SAR targets, as discussed in Section IV-F.

4) Augmentations for SAR Data: In contrastive learning, data augmentation plays a crucial role by introducing variations and perturbations to the samples, thereby providing a broader range of positive and negative pairs. Models learn more diverse and richer feature representations with this augmentation strategy, leading to improved performance in contrastive learning tasks. Data augmentation serves two main purposes: first, it increases the number of positive pairs, enabling the pretrained model to capture the inherent consistency within samples and enhance its discriminative ability to recognize the same object or category. Second, data augmentation also augments the number of negative pairs, facilitating the model to capture the distinctive characteristics between samples and improve its discriminative power for distinguishing different objects or categories.

The data augmentation operations were employed in this study to increase the diversity of positive samples and capture the variations observed in real-world SAR imaging. The utilized techniques encompassed blurring, color enhancement, rotation, flipping, Gaussian noise, and speckle noise interference. By applying these operations, augmented samples with diverse characteristics were generated. The torchvision module in the PyTorch framework was employed to facilitate the implementation of these augmentation methods. Fig. 9 illustrates the positive sample pairs generated through this approach. During the pretraining phase of LDCL, the positive sample pairs were encouraged to exhibit proximity in the feature space, thereby promoting the learning of consistent representations.

5) Experimental Settings: In the LDCL experiments, ResNet18 architecture was employed as the backbone feature extraction network for contrastive learning. The learning rate was set to 0.03, and a batch size of 512 was utilized for training. The unsupervised pretraining phase consisted of 500 iterations. For instance discrimination, the temperature coefficient was set to 0.7, while for group discrimination, it was set to 0.5. The momentum coefficient was chosen as 0.999. The weight α for the impactful group-instance discrimination loss was thoughtfully chosen as 0.2, amplifying its significance. In a similar vein, the weight λ attributed to data batch mixing was carefully determined as 0.25. To enrich the positive sample pairs in the neural network memory, various data augmentation techniques were employed in the experiments. All experiments in this article were conducted on a hardware configuration consisting of an Intel Xeon[1]

[1]Registered trademark.

TABLE III
DETAILED INFORMATION OF DATASETS FOR SMALL SAMPLE SAR TARGET RECOGNITION TASKS IN DIFFERENT SCENARIOS AND TASK CONDITIONS

Tasks	Dataset	Class	S1	S2	S3	S4	S5	S6	Testing Set	Depression
SOC	MSTAR	BMP2	117	78	59	30	15	8	195	17°, 15°
		BTR70	117	78	59	30	15	8	196	17°, 15°
		T72	116	78	58	29	15	8	196	17°, 15°
		ZSU23/4	150	100	75	38	19	10	274	17°, 15°
		ZIL131	150	100	75	38	19	10	274	17°, 15°
		T62	150	100	75	38	19	10	273	17°, 15°
		BTR60	128	86	64	32	16	8	195	17°, 15°
		D7	150	100	75	38	19	10	274	17°, 15°
		BRDM2	149	100	75	38	19	10	274	17°, 15°
		2S1	150	100	75	38	19	10	274	17°, 15°
Fine-grained	MSTAR	BMP2(sn_9563)	117	78	59	30	15	8	195	17°, 15°
		BMP2(sn_9566)	117	78	59	30	15	8	196	17°, 15°
		BMP2(sn_c21)	117	78	59	30	15	8	196	17°, 15°
		BTR60	150	100	75	38	19	10	195	17°, 15°
		T62	150	100	75	38	19	10	273	17°, 15°
		T72(sn_132)	128	86	64	32	16	8	196	17°, 15°
		T72(sn_812)	150	100	75	38	19	10	195	17°, 15°
		T72(sn_s7)	149	100	75	38	19	10	191	17°, 15°
Large pitch angle change	MSTAR	2S1	150	100	75	38	19	10	288	17°, 30°
		BRDM2	149	100	75	38	19	10	287	17°, 30°
		ZSU23/4	150	100	75	38	19	10	288	17°, 30°
Generalization	FUSAR-Ship	Ships	648	432	324	162	81	40	555	_
		Strong scatters	114	76	57	28	14	7	128	_
		Bridge&coastlines	511	341	255	127	63	31	438	_
		Coastal lands&islands	353	256	176	88	44	22	303	_
		Sea clutter waves	688	459	344	172	86	43	590	_
		Sea patches	625	416	312	156	78	39	535	_
		Land patches	568	379	284	142	71	35	487	_

E5-2620 CPU with 72 GB memory and an RTX2080Ti GPU with 11 GB memory. The software environment comprises the Ubuntu 20.04 operating system and a deep learning workstation equipped with the PyTorch 1.8 deep learning framework.

B. Ablation Studies

This section presents intriguing ablation experiments that delve into the profound impact of reducing instance discrimination confidence on SAR image target recognition. These experiments meticulously probe the influential role of discrimination strength on recognition performance. The analysis zooms in on two pivotal domains: feature space and input space, shedding light on the captivating effects of low-confidence discrimination. Through this comprehensive exploration, we unravel the intricate dynamics of low-confidence discrimination and its profound implications for SAR image target recognition.

Low-confidence discrimination aims to reduce the discriminative strength of traditional contrastive learning. Fig. 10 presents the t-SNE visualization of the ten-class MSTAR target data after multiple rounds of self-supervised pretraining iterations using the proposed method [44]. After 500 iterations, the targets were successfully differentiated, indicating the completion of SAR target representation learning. However, because of the high similarity among SAR targets, the assumption of instance discrimination based on natural images was not applicable, resulting in slow training and poor instance discrimination performance, as shown in Fig. 10. MoCo exhibits slightly lower training stability than LDCL on the MSTAR dataset, but it demonstrates slightly better discrimination accuracy than LDCL (better aggregation of target samples with the same color in MoCo). This difference could be attributed to the influence of the low-confidence discrimination strategy. However, considering the availability of labeled information for downstream tasks, these influences can be disregarded.

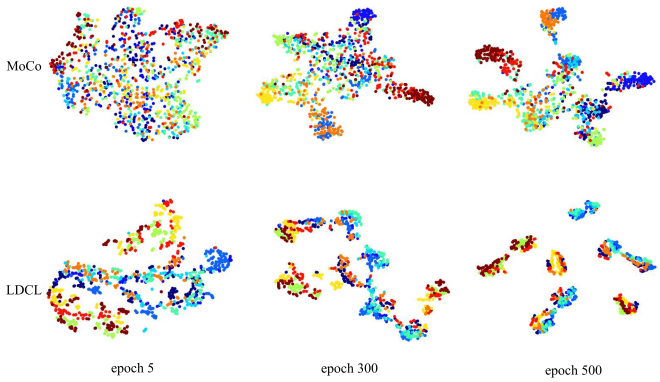

Fig. 10. Visualization of embedding features on MSTAR using MoCo and LDCL with t-SNE.

Batch size impacts the performance of LDCL. We conducted 500 iterations of training with LDCL using various batch sizes, followed by fine-tuning the undersampled dataset set up in MSTAR's S6 scenario. Table IV displays the experimental results. It is observed that as the batch size increases, the expansion in the quantity of SAR instances per training amplifies the complexity of the pretext task,

TABLE IV
FINE-TURNING WITH DIFFERENT BATCH SIZES UNDER S6 SET OF MSTAR DATABASE

Batch Size	32	64	128	256	512
Accuracy (%)	80.45	82.51	83.85	87.09	**88.83**

TABLE V
CLASSIFICATION ACCURACY OF ABLATION EXPERIMENTS

Methods	Low-confidence Discrimination		Accuracy(%)
	BMT	GID	
MoCo	×	×	81.27
LDCL	✓	×	86.80
LDCL	×	✓	86.28
LDCL	✓	✓	**88.83**

and softens the probability distribution, thereby diminishing the confidence in SAR instance discrimination. Given the experimental hardware conditions of this study, which include an NVIDIA RTX2080Ti with 11 GB RAM, larger batch sizes lead to memory overflow.

To demonstrate the remarkable effectiveness of LDCL in small sample SAR ATR for downstream tasks, we performed comprehensive comparative experiments to evaluate the influence of low-confidence discrimination in both the input space (batch instance mixing strategy) and the feature space (group-instance discrimination). The insightful experimental findings are presented in Table V, where we use the abbreviations "BMT" for the batch instance mixing strategy and "GID" for the group-instance discrimination method, ensuring brevity and clarity in the table. We performed ablation experiments on a small-sample MSTAR dataset under the S6 configuration, employing the low-confidence discriminative treatments of "BMT" and "GID." These treatments resulted in the highest SAR ATR accuracy.

We compared the classification test accuracy and training loss of LDCL and MoCo on the MSTAR dataset under the small sample conditions S1 and S6. Fig. 11 displays the findings. On SAR data with high similarity and low discriminability, MoCo demonstrated stable performance when a larger amount of annotated data was available (S1). However, it demonstrated notable volatility in performance when confronted with a smaller data scale. In stark contrast, LDCL showcased enhanced stability, as evidenced by a consistently smooth training loss and a classification accuracy curve that remained steady throughout.

We investigated the sensitivity of the recognition accuracy to the hyperparameters λ, which control the confidence strength in the SAR image input space, and α, which control the confidence strength in the SAR image feature space. The results, shown in Fig. 12 for the small sample condition S6, were obtained using the MSTAR database. Due to the similarity of intraclass samples and the similarity among interclass samples (different target images under the same elevation angle appearing similar), the highest accuracy was achieved when $\alpha = 2.5$ and $\lambda = 0.5$. In contrast, the FuSARShip dataset, which consists of maritime targets, exhibits high intraclass

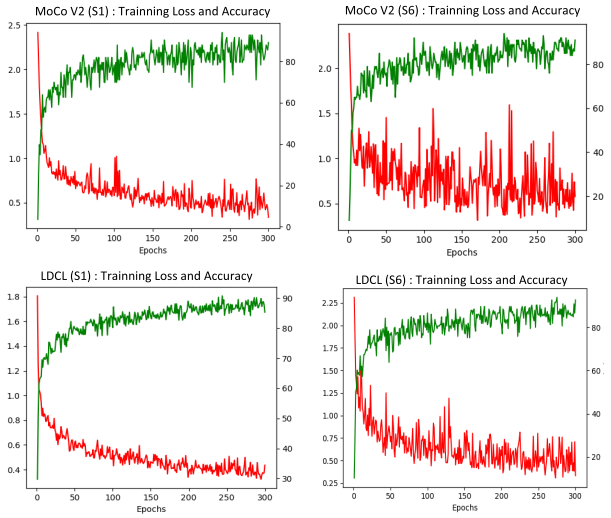

Fig. 11. Classification test accuracy and training loss of LDCL and MoCo on the MSTAR dataset under the small sample conditions S1 and S6.

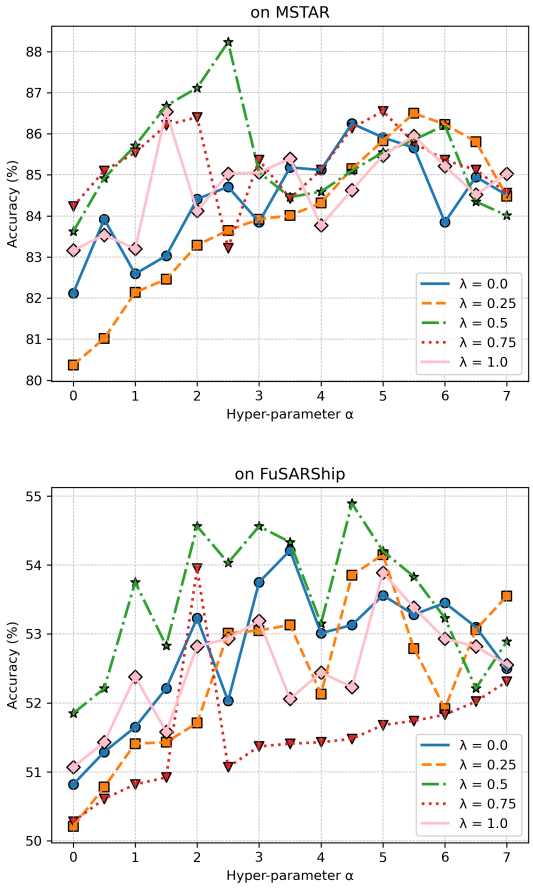

Fig. 12. Sensitivity of the recognition accuracy to the hyperparameters λ and α.

similarity and weak interclass similarity. Hence, the highest accuracy was obtained when $\alpha = 5.5$ and $\lambda = 0.5$.

C. Performance Evaluation Under SOC

This section uses the original MSTAR dataset to validate the effectiveness of LCDL in small-sample SAR target

TABLE VI
COMPARISON WITH SUPERVISED LEARNING CLASSIFICATION NETWORK PERFORMANCE UNDER SOC

Methods	S1	S2	S3	S4	S5	S6
LeNet	87.25	83.91	81.09	62.98	49.55	32.60
AlexNet	92.22	91.51	89.87	80.59	60.25	39.62
VGG16	63.34	56.52	47.10	39.82	32.54	15.30
RestNet18	78.51	76.32	63.69	58.82	43.85	35.52
ResNet50	69.91	59.51	38.55	35.6	29.37	20.20
InceptionV3	83.60	73.11	61.32	50.83	44.25	18.23
DenseNet	84.2	87.5	82	60.8	50	17
A-ConvNet	92.12	85.52	83.41	76.28	62.45	35.69
AlexNet+PT	92.50	93.65	81.09	68.54	51.36	33.62
RestNet18+PT	93.88	92.70	83.62	77.30	50.58	34.79
MFFA-SARNET	96.62	**95.84**	93.09	70.25	52.18	39.51
LCDL	**98.38**	95.30	**93.71**	**89.90**	**89.02**	**88.56**

TABLE VII
RECOGNITION AVERAGE ACCURACY OF DIFFERENT ALGORITHMS IN SMALL SAMPLE SETTING UNDER SOC

Methods	S1	S2	S3	S4	S5	S6
Fine-tuned:						
SimCLR	94.94	94.54	93.88	93.19	88.04	84.86
MoCoV2	94.31	94.22	94.15	93.23	86.51	81.27
NPID	92.36	88.91	85.68	82.43	80.55	78.60
Byol	92.6	91.33	88.12	87.89	85.78	80.89
SimSame	95.8	92.52	93.56	90.78	88.78	83.59
SwAV	93.88	92.85	91.56	88.52	86.25	84.03
RotANet	87.78	85.69	82.63	79.85	76.82	73.71
MS-SSL	90.50	86.58	86.06	85.59	78.90	75.81
CLPL-SAR	90.32	88.21	81.90	60.58	52.56	51.83
Barlow Twins	87.22	80.56	76.71	72.19	71.86	68.15
LDCL	**97.85**	**95.29**	**92.6**	**93.96**	**89.33**	**88.83**
Linear evaluation:						
SimCLR	91.41	91.78	91.5	91.66	87.88	84.38
MoCoV2	88.69	88.47	87.17	86.59	82.37	79.46
NPID	90.83	85.50	82.01	80.63	73.88	71.90
Byol	89.88	88.68	85.23	83.55	79.78	78.89
SimSame	93.89	90.54	88.83	86.56	82.35	81.26
SwAV	93.13	91.56	88.58	85.89	83.25	81.5
RotANet	60.50	58.06	58.61	53.85	49.29	38.51
MS-SSL	70.50	66.58	58.76	55.59	48.90	45.81
CLPL-SAR	88.56	83.55	76.32	52.65	51.69	48.73
Barlow Twins	85.29	78.79	71.54	66.25	65.57	62.83
LDCL	**95.28**	**90.93**	**90.33**	**89.25**	**88.28**	**85.11**

recognition. To create the small-sample dataset, training samples with an original elevation angle of 17° were randomly selected at proportions of S1–S6. The testing set consisted of samples from the ten classes with an elevation angle of 15°. For comparison of the proposed method with classical supervised learning methods for SAR ATR and classical self-supervised contrastive learning methods, we adopted LeNet [45], AlexNet [46], VGG16 [47], ResNet50 [41], InceptionV3 [48], DenseNet [49], A-ConvNet, and MFFA-SARNET for the supervised learning comparison. These methods involved random initialization of parameters. In the experiments, AlexNet and ResNet18 were additionally supplemented with a ten-class MSTAR pretrained model[2] [50], [51], referred to as AlexNet+PT and ResNet+PT. In addition, SimCLR, MoCoV2, NPID, Byol, SimSame, SwAV, RotANet [26], MS-SSL [27], CLPL-SAR [15], and Barlow Twins [52] serve as benchmarks for SSL comparisons. The results are presented in Table VI. Our method underwent 500 epochs of pretraining using the Adam optimizer, with a learning rate adjustment to 0.3 and a batch size of 128. The temperature coefficient, for instance, for discrimination was set to 0.7, and the temperature coefficient for group discrimination was set to 0.5. Both the linear evaluation and fine-tuning stages involved training the model for 500 epochs. Table VII demonstrates our method performs well in small sample SAR target recognition compared to classical contrastive learning methods.

D. Fine-Grained SAR Target Recognition Performance Evaluation

For the evaluation of fine-grained target recognition, we focused on samples exclusively from the MSTAR dataset. The training set includes samples with an elevation angle of 17° and the test set includes samples with an elevation angle of 15°. To represent fine-grained target categories, we carefully selected the BMP2, T72, BTR60, and T62 target models, known for their distinct characteristics. In addition, to create a highly challenging fine-grained SAR target recognition task with similar SAR imaging views, we included pairs of target categories with comparable scattering characteristics and appearances, specifically T72/T62 and BTR60/BMP2.To assess the discriminative capability of our proposed LDCL algorithm for small sample SAR target recognition, we compared it against traditional self-supervised contrastive learning algorithms as baselines. The task in this section posed even greater challenges compared to Section IV-B, where LDCL outperformed traditional methods. As a result, traditional self-supervised contrastive learning algorithms exhibited subpar performance, while LDCL demonstrated remarkable recognition accuracy in both the linear evaluation and fine-tuning stages. For a detailed understanding of the dataset used for the fine-grained target recognition task, refer to Table III. Moreover, the recognition results of traditional contrastive learning methods on the SAR dataset can be found in Table VIII.

E. Performance Evaluation for Large Pitch Angle Change Conditions

The imaging views of SAR targets exhibit significant differences due to variations in elevation angles. As the elevation angle difference increases, the similarity of target imaging

[2]Sourced from https://github.com/Alien9427/SAR_specific_models.

TABLE VIII
RECOGNITION AVERAGE ACCURACY OF DIFFERENT ALGORITHMS IN SMALL SAMPLE SAR FINE-GRAINED IMAGE RECOGNITION

Methods	S1	S2	S3	S4	S5	S6
Fine-tuned:						
SimCLR	62.06	61.30	60.11	54.81	53.82	51.56
NPID	57.33	53.76	51.69	45.28	43.91	40.27
CPCV2	60.95	58.72	50.69	46.19	43.33	40.60
MoCoV2	62.92	61.85	60.72	57.42	52.53	50.59
Byol	65.09	60.33	58.12	57.89	55.78	50.89
SimSame	65.88	62.56	63.36	60.78	58.70	53.60
SwAV	63.88	62.35	60.19	58.52	56.54	54.03
Barlow Twins	60.85	60.51	58.62	58.29	54.20	53.81
RotANet	62.36	57.30	55.58	53.25	50.69	48.83
MS-SSL	61.52	58.95	58.76	51.90	47.12	43.30
CLPL-SAR	63.89	62.15	59.35	54.24	53.09	52.80
LCDL	**68.55**	**65.29**	**62.60**	**63.69**	**59.89**	**55.91**
Linear evaluation:						
SimCLR	61.66	61.20	60.03	53.61	52.88	50.95
NPID	56.09	53.69	50.56	43.69	41.60	38.50
CPCV2	60.15	55.60	48.59	43.05	40.86	38.55
MoCoV2	60.69	58.47	57.25	56.45	50.37	48.30
Byol	63.82	58.58	55.89	51.33	49.68	48.39
SimSame	63.92	60.88	58.85	56.03	55.60	51.36
SwAV	63.05	61.69	58.15	55.81	53.65	51.38
Barlow Twins	58.33	58.10	56.91	55.74	52.56	51.02
RotANet	60.25	55.21	54.23	48.90	45.78	41.26
MS-SSL	60.59	57.71	56.60	50.38	45.56	38.21
CLPL-SAR	60.84	57.85	55.24	51.50	50.87	49.33
LCDL	**68.32**	**64.55**	**60.36**	**60.25**	**57.85**	**55.06**

views decreases. In this section, we assessed the robustness of the small-sample SAR target recognition algorithm to changes in elevation angle. Specifically, we selected three targets: rocket launcher (2S1), armored vehicle (BRDM2), and anti-aircraft unit (ZSU23/4). The training set comprises data with an elevation angle of 17°, while the test set consists of data with an elevation angle of 30°. The substantial difference in elevation angles between the training and test sets imposes higher demands on the model's generalization ability. The dataset for elevation angle variations was described in detail in the accompanying table. We compared the proposed algorithm with traditional supervised learning methods, as presented in Table IX, to access its recognition ability when SAR targets' elevation angles varied significantly. Regarding the recognition rate, our algorithm outperforms traditional supervised learning methods, highlighting its effectiveness in handling SAR targets with large variations in elevation angle.

F. Recognition Average Accuracy of Different Algorithms in Small Sample Setting Under Unbalanced Data Condition Performance Evaluation

For SAR target recognition, imbalanced target categories are commonly encountered due to the uneven occurrence

TABLE IX
RECOGNITION AVERAGE ACCURACY OF DIFFERENT ALGORITHMS IN SMALL SAMPLE SETTING UNDER LARGE PITCH ANGLE CHANGE CONDITIONS

Method	S1	S2	S3	S4	S5	S6
LeNet	87.4	84.1	80.9	63	48	31.6
AlexNet	83.2	82.3	79.5	70.2	60.5	38.2
VGG16	82.7	55.5	46.1	39.2	31.4	14
ResNet50	60.9	25.1	28.5	35.6	59.7	69
InceptionV3	86	71.1	60.2	49.8	43.5	17.7
DenseNet	86.2	87.5	82	60.8	50	17
A-ConvNet	91.38	82.35	73.88	62.82	60.15	32.56
MFFA-SARNET	76.6	65.4	63.9	55.5	53.8	38.9
LDCL	**93.08**	**91.38**	**90.59**	**88.09**	**83.90**	**78.51**

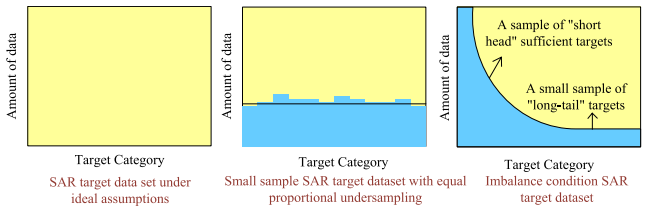

Fig. 13. Schematic of SAR target dataset: ideal scenario, equal-scale undersampling, and "long-tail" distribution.

frequency of natural targets. In Sections IV-C–IV-E, experiments were conducted using undersampling data processing, where an equal number of samples were randomly selected for training, ensuring consistent proportions across categories. However, this approach may not fully exploit the sample information from minority classes. Moreover, traditional neural network-based target recognition models tend to favor majority-class samples with higher empirical risk, further exacerbating the imbalanced nature of the data.

In this section, we evaluated the performance of small sample SAR target recognition in the context of imbalanced data distribution. This issue is akin to the "long-tail distribution" problem commonly observed in computer vision. As depicted in Fig. 13, the SAR target dataset exhibits a high occurrence frequency for certain targets, resulting in a concentrated distribution in the head with an abundance of data. On the other hand, the tail of the distribution comprises targets with low occurrence frequency, which are more difficult to collect and belong to the small sample category.

Traditional neural network-based target recognition models encounter two challenges when handling imbalanced training datasets: insufficient information and a bias toward the majority class samples with higher empirical risk. The limited number of samples from minority classes may not provide enough discriminative information to effectively represent those classes. Moreover, neural networks tend to prioritize the empirical risks associated with the majority class samples, which can result in decreased recognition accuracy for the minority classes.

In contrast, self-supervised models treated each sample as a separate "category" during the pretraining phase. For

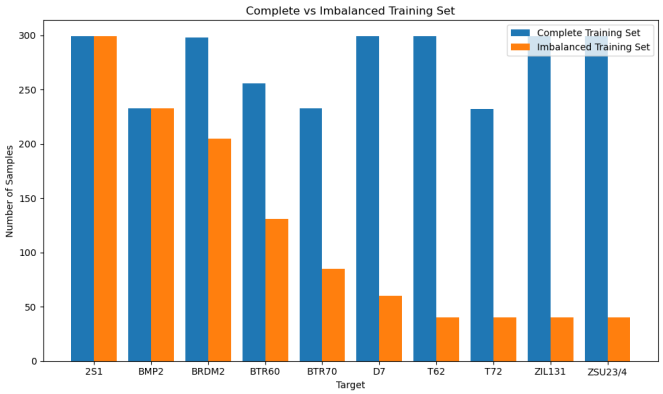

Fig. 14. Complete versus imbalanced training set.

TABLE X
RECOGNITION AVERAGE ACCURACY OF DIFFERENT ALGORITHMS IN SMALL SAMPLE SETTING UNDER IMBALANCED DATASET CONDITIONS

Methods	Accuracy (%)	
	Complete Dataset	Imbalanced Dataset
Restnet18	93.25	86.89
Restnet50	94.32	68.72
InceptionV3	93.55	66.09
DenseNet	94.50	73.82
A_ConvNet	92.83	74.91
MFCNN	95.57	83.71
MFFA_SarNet	93.30	88.56
LDCL	**97.41**	**92.07**

example, the self-supervised network generated 2747 pseudo-labels with the ten-class MSTAR training dataset consisting of 2747 samples. This approach ensured that the self-supervised model did not overlook the recognition accuracy of minority class samples and mitigated the cost-sensitive problem encountered by traditional target recognition networks when dealing with imbalanced SAR data. By not solely focusing on minimizing the empirical risk of the target recognition model, the self-supervised approach addressed the imbalanced nature of the dataset.

To construct an imbalanced training dataset, this section employed a random selection process for each of the ten MSTAR target classes at an elevation angle of 17°. The distribution of samples for each class is as follows: 299, 233, 205, 131, 85, 60, 40, 40, 40, and 40, as illustrated in Fig. 14. The test set comprises a complete ten-class target dataset at an elevation angle of 15°. Table X compares the performance of our algorithm on this unbalanced dataset with that of the traditional SAR-based target recognition algorithms. While traditional supervised learning methods maintain an accuracy of over 90% on the complete dataset, they suffer from varying degrees of accuracy decline under the imbalanced data condition, with a significant drop. In contrast, the proposed algorithm in this chapter only experiences a 5% decrease in accuracy, which confirms its effectiveness in effectively mitigating the problem of empirical risk preference caused by the "long-tail" data distribution.

TABLE XI
RECOGNITION PERFORMANCE OF LDCL WITH DIFFERENT ALGORITHMS ON SMALL SAMPLE FUSAR-SHIP DATASET

Methods	S1	S2	S3	S4	S5	S6
Fine-tuned:						
SimCLR	70.94	64.05	53.52	51.79	48.04	43.97
NPID	66.32	60.53	56.89	51.30	44.79	31.29
CPCV2	71.58	63.52	58.90	46.80	41.23	32.09
MoCoV2	74.31	64.62	60.56	58.62	50.81	45.37
Byol	72.88	70.50	68.59	63.56	55.02	41.54
SimSame	65.78	62.29	53.58	48.62	46.25	39.35
SwAV	63.59	62.15	59.38	52.66	42.30	34.32
Barlow Twins	71.93	70.58	68.45	60.85	40.81	35.28
RotANet	62.95	60.88	59.56	55.71	50.83	41.26
CLPL-SAR	68.90	65.88	58.61	52.85	41.28	38.56
MS-SSL	72.83	68.25	60.19	57.05	52.93	46.35
LDCL	**76.52**	**73.39**	**72.69**	**63.61**	**58.21**	**50.33**
Linear evaluation:						
SimCLR	65.51	61.43	51.29	41.69	39.02	30.55
NPID	62.09	59.54	52.51	46.32	40.57	29.65
CPCV2	63.61	52.72	55.71	43.62	39.06	30.32
MoCoV2	68.59	58.47	51.29	41.90	32.57	30.06
Byol	69.12	48.68	45.23	43.55	39.78	38.89
SimSame	73.21	70.94	68.38	66.21	62.35	81.26
SwAV	63.61	61.16	58.58	54.89	33.52	31.51
Barlow Twins	68.95	64.24	59.47	54.33	32.81	30.55
RotANet	60.50	58.06	58.61	53.85	49.29	38.51
CLPL-SAR	62.13	61.92	52.82	48.32	39.90	35.30
MS-SSL	70.50	66.58	58.76	55.59	48.90	45.81
LDCL	**75.28**	**70.63**	**65.03**	**59.55**	**58.38**	**46.55**

G. Generalization

To assess the generalizability of our algorithm on other SAR datasets, here we present experiments performed on the FUSAR-Ship dataset using the same undersampling method as described in Table III. Given the characteristics of high intraclass similarity and large interclass differences in the maritime target samples of FUSAR-Ship, we adjusted the weight α of the group-instance discriminative loss to 0.5, while keeping the other initialization parameters unchanged. The proposed algorithm's recognition performance on the small-sample FUSAR-Ship dataset is summarized in Table XI, and it surpasses that of classical SSL algorithms on the same dataset. This preliminary result confirms the generalization ability of our algorithm across different SAR datasets.

Furthermore, to validate the generalization capability of LDCL, we also selected aircraft targets (model types dzx-72 and T504N-72) [53], ship targets (models bulk carrier, container, and fishing) [43], and vehicle targets (model types 2S1,

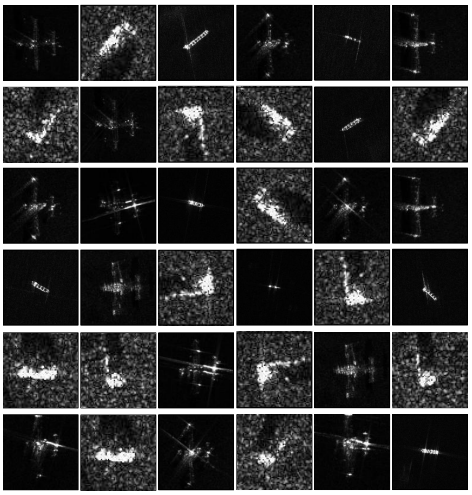

Fig. 15. Mixed SAR target samples for LDCL self-supervised pretraining.

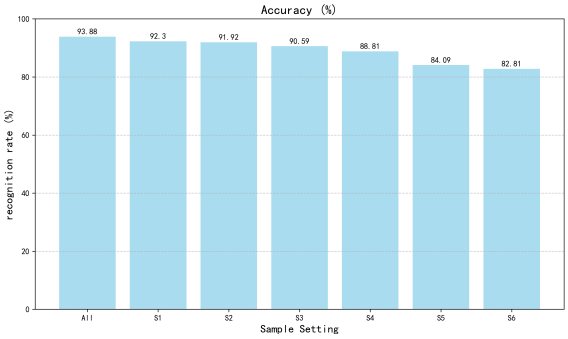

Fig. 16. Small-sample recognition rate of LDCL under generalization conditions.

BRDM_2, and BTR60) for pretraining [42]. The mixed data samples used for pretraining are illustrated in Fig. 15. The weight α of group-instance discrimination was set to 0.5, and the weight λ attributed to data batch mixing was also set to 0.5. Experiments resulted in an identification rate of 93.88%, as shown in Fig. 16. When training samples were gradually reduced using an equidistant method, the identification rate remained above 82.80%. This indicates that LDCL possesses exceptional generalization capability when dealing with SAR targets in various application scenarios, even under conditions of limited sample labeled data.

V. Conclusion

This article introduced LDCL, a method specifically developed to address the challenges encountered in small sample SAR target recognition tasks. The proposed LDCL aims to improve the performance of small sample SAR target recognition by reducing the confidence of contrastive learning from the input space and feature learning space and improving the self-supervised representation learning of SAR target images with high similarity.

Our algorithm was thoroughly evaluated on the MSTAR dataset, covering various tasks such as small sample fine-grained target recognition, target recognition under significant elevation angle variations, and target recognition under imbalanced data conditions. Across all these tasks, LDCL consistently outperforms classical target recognition algorithms as well as conventional CNN-based SSL algorithms, showcasing its superior recognition performance.

In addition, LDCL's applicability has been further validated on the FUSAR-Ship dataset, which focuses on maritime target recognition. This validation has demonstrated LDCL's effectiveness in small sample SAR target recognition tasks that involve high intraclass similarity and low interclass similarity. The potential utility of the algorithm has been reinforced by its generalization ability observed on different datasets.

Although SSL proves effective in small sample annotated SAR target recognition scenarios, it heavily relies on a substantial amount of unlabeled SAR data related to downstream tasks. The future research direction is to explore the possibility of small-sample SAR target learning without heavy reliance on downstream task-related data.

References

[1] 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.

[2] K. Fu, T. Zhang, Y. Zhang, Z. Wang, and X. Sun, "Few-shot SAR target classification via metalearning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022, Art. no. 2000314, doi: 10.1109/TGRS.2021.3058249.

[3] L. Zhang et al., "Domain knowledge powered two-stream deep network for few-shot SAR vehicle recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022, Art. no. 5215315, doi: 10.1109/TGRS.2021.3116349.

[4] F. Ma, X. Sun, F. Zhang, Y. Zhou, and H.-C. Li, "What catch your attention in SAR images: Saliency detection based on soft-superpixel lacunarity cue," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–17, 2023, Art. no. 5200817, doi: 10.1109/TGRS.2022.3231253.

[5] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.

[6] Y. Zhou, H. Liu, F. Ma, Z. Pan, and F. Zhang, "A sidelobe-aware small ship detection network for synthetic aperture radar imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–16, 2023, Art. no. 5205516, doi: 10.1109/TGRS.2023.3264231.

[7] 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.*, Jul. 2018, pp. 625–628.

[8] 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.

[9] R. Qin, X. Fu, J. Chang, and P. Lang, "Multilevel wavelet-SRNet for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[10] 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.

[11] Y. Zhou, Y. Li, W. Xie, and L. Li, "A convolutional neural network combined with attributed scattering centers for SAR ATR," *Remote Sens.*, vol. 13, no. 24, p. 5121, Dec. 2021.

[12] J. Zhang, M. Xing, and Y. Xie, "FEC: A feature fusion framework for SAR target recognition based on electromagnetic scattering features and deep CNN features," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2174–2187, Mar. 2021.

[13] R. Qin, X. Fu, J. Dong, and W. Jiang, "A semi-greedy neural network CAE-HL-CNN for SAR target recognition with limited training data," *Int. J. Remote Sens.*, vol. 41, no. 20, pp. 7889–7911, Oct. 2020.

[14] Z. Sun, J. Li, P. Liu, W. Cao, T. Yu, and X. Gu, "SAR image classification using greedy hierarchical learning with unsupervised stacked CAEs," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5721–5739, Jul. 2021.

[15] 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, pp. 1–5, 2022.

[16] F. Gao, Y. Yang, J. Wang, J. Sun, E. Yang, and H. Zhou, "A deep convolutional generative adversarial networks (DCGANs)-based semi-supervised method for object recognition in synthetic aperture radar (SAR) images," *Remote Sens.*, vol. 10, no. 6, p. 846, May 2018.

[17] Y. Zhai et al., "MFFA-SARNET: Deep transferred multi-level feature fusion attention network with dual optimized loss for small-sample SAR ATR," *Remote Sens.*, vol. 12, no. 9, p. 1385, Apr. 2020.

[18] R. Yang, X. Xu, X. Li, L. Wang, and F. Pu, "Learning relation by graph neural network for SAR image few-shot learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 1743–1746.

[19] I. Misra and L. van der Maaten, "Self-supervised learning of pretext-invariant representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6706–6716.

[20] X. Liu et al., "Self-supervised learning: Generative or contrastive," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 1, pp. 857–876, Jan. 2023.

[21] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[22] Y. Tian, D. Krishnan, and P. Isola, "Contrastive multiview coding," in *Proc. 16th Eur. Conf. Comput. Vis.* Berlin, Germany: Springer-Verlag, Aug. 2020, pp. 776–794.

[23] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9726–9735.

[24] J.-B. Grill et al., "Bootstrap your own latent-a new approach to self-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 21271–21284.

[25] X. Chen and K. He, "Exploring simple Siamese representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15745–15753.

[26] 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.

[27] 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, p. 4632, Sep. 2022.

[28] 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.

[29] T. Jiang, Z. Cui, Z. Zhou, and Z. Cao, "Data augmentation with Gabor filter in deep convolutional neural networks for SAR target recognition," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 689–692.

[30] S. Du, J. Hong, Y. Wang, K. Xing, and T. Qiu, "Physical-related feature extraction from simulated SAR image based on the adversarial encoding network for data augmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[31] Y. Sun, Y. Wang, H. Liu, N. Wang, and J. Wang, "SAR target recognition with limited training data based on angular rotation generative network," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 11, pp. 1928–1932, Nov. 2020.

[32] Y. Xu and H. Lang, "Ship classification in SAR images with geometric transfer metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6799–6813, Aug. 2021.

[33] J. Tang, F. Zhang, Y. Zhou, Q. Yin, and W. Hu, "A fast inference networks for SAR target few-shot learning based on improved Siamese networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 1212–1215.

[34] J. Cai, Y. Zhang, J. Guo, X. Zhao, J. Lv, and Y. Hu, "ST-PN: A spatial transformed prototypical network for few-shot SAR image classification," *Remote Sens.*, vol. 14, no. 9, p. 2019, Apr. 2022.

[35] C. Wang, J. Pei, J. Yang, X. Liu, Y. Huang, and D. Mao, "Recognition in label and discrimination in feature: A hierarchically designed lightweight method for limited data in SAR ATR," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, Art. no. 5239613, doi: 10.1109/TGRS.2022.3225187.

[36] 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.*, Jun. 2018, pp. 3733–3742.

[37] S. Ren et al., "A simple data mixing prior for improving self-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 14575–14584.

[38] 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, Art. no. 5204317, doi: 10.1109/TGRS.2021.3066195.

[39] C. Cao, Z. Cui, L. Wang, J. Wang, Z. Cao, and J. Yang, "Cost-sensitive awareness-based SAR automatic target recognition for imbalanced data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022, Art. no. 5205316, doi: 10.1109/TGRS.2021.3068447.

[40] J. Yu, M. Tan, H. Zhang, Y. Rui, and D. Tao, "Hierarchical deep click feature prediction for fine-grained image recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 2, pp. 563–578, Feb. 2022.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[42] C. Coman and R. Thaens, "A deep learning SAR target classification experiment on MSTAR dataset," in *Proc. 19th Int. Radar Symp. (IRS)*, Jun. 2018, pp. 1–6.

[43] X. Hou, W. Ao, Q. Song, J. Lai, H. Wang, and F. Xu, "FUSAR-ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition," *Sci. China Inf. Sci.*, vol. 63, no. 4, pp. 1–19, Apr. 2020.

[44] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.

[45] H. Fırat, M. E. Asker, M. I. Bayindir, and D. Hanbay, "Spatial–spectral classification of hyperspectral remote sensing images using 3D CNN based LeNet-5 architecture," *Infr. Phys. Technol.*, vol. 127, Dec. 2022, Art. no. 104470.

[46] Z.-W. Yuan and J. Zhang, "Feature extraction and image retrieval based on alexnet," in *Proc. 8th Int. Conf. Digit. Image Process. (ICDIP)*, vol. 10033. Bellingham, WA, USA: SPIE, 2016, pp. 65–69.

[47] X. Zhang, J. Zou, K. He, and J. Sun, "Accelerating very deep convolutional networks for classification and detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 1943–1955, Oct. 2016.

[48] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," in *Proc. 2nd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2017, pp. 783–787.

[49] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[50] Z. Huang, Z. Pan, and B. Lei, "What, where, and how to transfer in SAR target recognition based on deep CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2324–2336, Apr. 2020.

[51] Z. Huang, C. O. Dumitru, Z. Pan, B. Lei, and M. Datcu, "Classification of large-scale high-resolution SAR images with deep transfer learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 107–111, Jan. 2021.

[52] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, "Barlow twins: Self-supervised learning via redundancy reduction," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 12310–12320.

[53] W. Ruyi et al., "Multiangle SAR dataset construction of aircraft targets based on angle interpolation simulation," *J. Radars*, vol. 11, no. 4, pp. 637–651, 2022.

Jinrui Liao (Student Member, IEEE) received the B.S. and M.S. degrees from Wuyi University, Jiangmen, China, in 2019 and 2023, respectively. He is currently pursuing the Ph.D. degree with the School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen, China.

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

Yikui Zhai (Senior Member, IEEE) received the B.S. and master's degrees 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 June 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.

Qingsong Wang received the B.S. and Ph.D. degrees from the National University of Defense Technology, Changsha, China, in 2006 and 2011, respectively.

Since 2019, he has been working with the School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen, China, where he is mainly engaged in basic theory and engineering application research in the field of precision processing of microwave mapping. His research interests include multisystem radar 3-D positioning, image matching, and UAV vision navigation.

Bing Sun (Member, IEEE) received the B.S. and Ph.D. degrees from Beihang University, Beijing, China, in 2003 and 2008, respectively.

He became a Lecturer after finishing his post-doctoral research in 2010. He was a Visiting Scholar with The University of Texas–Pan American, Edinburg, TX, USA, from November 2013 to November 2014. He has been an Associate Professor with Beihang University since 2015. His research interests include synthetic aperture radar top-level design and simulation, signal processing and quality evaluation, and pattern recognition.

Vincenzo Piuri (Fellow, IEEE) received the M.S. and Ph.D. degrees in computer engineering from Politecnico di Milan, Milan, Italy, in 1984 and 1988, respectively.

He was an Associate Professor with Politecnico di Milan from 1992 to 2000, a Visiting Professor with The University of Texas at Austin, Austin, TX, USA, from 1996 to 1999, and a Visiting Researcher with George Mason University, Fairfax, VA, USA, from 2012 to 2016. He was the Department Chair at the University of Milan, Milan, from 2007 to 2012, where he has been a Full Professor, since 2000. He founded a startup company, SENSURE SRL, Bergamo, Italy, in the area of intelligent systems for industrial applications (leading it from 2007 to 2010) and was active in industrial research projects with several companies. His main research and industrial application interests include intelligent systems, computational intelligence, pattern analysis and recognition, machine learning, signal and image processing, biometrics, intelligent measurement systems, industrial applications, distributed processing systems, the Internet-of-Things, cloud computing, fault tolerance, application-specific digital processing architectures, and arithmetic architectures.

Dr. Piuri is an ACM Fellow.