Open Set Recognition and Category Discovery Framework for SAR Target Classification Based on K-Contrast Loss and Deep Clustering

Mingyao Chen, Jing-Yuan Xia, Tianpeng Liu, Li Liu, *Senior Member, IEEE*, and Yongxiang Liu, *Member, IEEE*

***Abstract*—Synthetic aperture radar automatic target recognition (SAR ATR) has been widely studied in recent years. Most ATR models are designed based on the traditional closed-set assumption. This type of ATR model can only identify target categories existing in the training set, and it will result in missed detection or misclassification of unseen target categories encountered in battlefield reconnaissance, posing a potential threat. Therefore, it is of great significance to design a model that can simultaneously achieve known class classification and unknown class judgment. In addition, researchers usually use the obtained unknown class data for model relearning to enable it to recognize new categories. However, before this process, it is necessary to manually interpret and annotate the obtained unknown class data, which undoubtedly requires a large time cost and is difficult to meet the timeliness requirements. To solve these problems, we propose a framework that integrates the open-set recognition module and the novel class discovery module. By introducing the K-contrast loss, the open-set recognition module can accurately distinguish unknown class data, classify known class data, and then transfer the known class knowledge through deep clustering for clustering annotation of unknown class data. Extensive experimental results on the MSTAR benchmark dataset demonstrate the effectiveness of the proposed methods.**

***Index Terms*—Category discovery, deep embedded clustering, open set recognition, synthetic aperture radar (SAR) target recognition.**

I. Introduction

THE rapid advancement of synthetic aperture radar (SAR) systems has enabled the collection of vast amounts of image data in a short period of time. The interpretation of this data is a labor-intensive process that often fails to meet the demands of real-time applications. Therefore, the ability to automatically recognize targets within a short timeframe has become an essential factor for successful battlefield reconnaissance and intelligence generation. Currently, most research is based on the ideal assumption that the data used in the training process contain all the target categories to be identified. These studies have achieved high accuracy results on closed-set testing through supervised training [1], [2], [3], [4], [5] or feature based models [6], [7], [8]. However, in practical applications, we are more concerned about targets that have not appeared in the training set. These targets are often missed or misclassified by traditional methods, which can pose potential threats.

The aforementioned issues pose a greater demand on the SAR automatic target recognition (ATR) model. First, the model needs to accurately determine whether the target belongs to a category that has appeared in the training set. We can refer to the categories present in the training set as known classes and those not present as unknown classes. Recognition of a test set containing both known and unknown classes can be defined as an open set recognition (OSR) problem [9]; that is, through training, it can determine whether the test sample is a known class and correctly classify the known class samples. In recent years, researchers have conducted extensive research on OSR in the field of computer vision. In work [10], a novel OpenMax layer is proposed to replace the traditional SoftMax layer, which estimates the probability of an input being from an unknown class. This enables the model to simultaneously determine whether the input sample is known and complete the classification of in-distribution data. To improve the robustness of convolutional neural networks in OSR and maintain their high accuracy in closed-set recognition (CSR), network structures such as convolutional prototype networks (CPN) [11] and generative adversarial networks (GAN) [12], [13], [14] are used in OSR. The above methods have demonstrated excellent performance on optical datasets. However, due to significant differences in imaging mechanisms, target characteristics, and scene applications between radar images and optical images, OSR methods designed for the optical image domain cannot be directly applied to radar target recognition.

Some researchers have proposed OSR methods for the SAR field by combining SAR image characteristics, which can optimize the current ATR system more effectively. Scherreik and Rigling introduced a novel method known as the probabilistic open set support vector machine (POS-SVM) for OSR [15], and later enhanced the POS-SVM and first applied it to the OSR of SAR targets [16]. Inspired by the use of GAN to complete

Manuscript received 21 November 2023; revised 17 December 2023; accepted 6 January 2024. Date of publication 12 January 2024; date of current version 24 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 61921001, Grant 62022091, and Grant 62201588. *(Corresponding authors: Tianpeng Liu; Jing-Yuan Xia.)*

The authors are with the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: cmy@nudt.edu.cn; j.xia16@imperial.ac.uk; liutianpeng2004@nudt.edu.cn; dreamliu2010@gmail.com; lyx_bible@sina.com).

Digital Object Identifier 10.1109/JSTARS.2024.3353453

abnormal detection tasks in [17] and [18], Ma et al. [19] divided the SAR OSR task into two stages: abnormal detection and known class classification. Using targets in the known class to train the generator and discriminator, the scores of the test samples output by the discriminator are used in the abnormal detection stage to judge unknown classes by comparing them with the set threshold [19]. Despite achieving significantly better results than other methods when the number of known classes is three, the method suffers from unstable GAN training, leading to a significant drop in performance when the number of known classes increases. In practical applications, the number of known classes to be classified should be relatively large. In view of this, Ma et al. [20] introduced the multiscale structural similarity (MS-SSIM) loss to obtain a representation that contains more structural information about SAR targets. By constructing clearer boundaries, they improved the distinguishability of test samples and achieved excellent results in scenarios with more known classes.

Since the observed scene is dynamically changing, the number of observed target categories is increasing over time. This puts forward another requirement for the ATR model, that is, further processing of the detected unknown class data. At present, most studies focus on incremental learning. By using limited labeled new category samples to fine-tune the classifier or feature extractor, the model can identify new categories at the lowest possible cost [20], [21], [22]. Such methods of handling detected unknown class data are indeed appealing, but they overlook some essential steps: Determining the number of categories of unknown classes in a batch, followed by classification and labeling. This process, without a doubt, demands considerable human resources. Naturally, we hope that the model can automatically implement the above process, which is called category discovery. It is challenging to determine the number of categories in a dataset composed of unlabeled samples of unknown classes and perform clustering accordingly. However, we can reduce the ambiguity of clustering and enhance the quality of newly discovered classes by transferring the prior knowledge of labeled known-class samples of SAR target [23]. Deep embedded clustering (DEC), as a learning-based clustering method, simultaneously clusters the data and learns a proper data representation, which can effectively meet the requirements of the above-mentioned scenarios [24], [25], [26]. In [23], deep transfer clustering (DTC) is formally proposed and studied as a separate problem. It creatively suggests using a labeled probe subset to better estimate the number of unknown classes and introduces temporal ensembling and π-model consistency loss [27] to improve the performance of the DEC model. Zheng et al. [28] drew inspiration from the above methods and combined DTC with OSR, using the multidomain applicable I-mix contrastive learning for improvement. Although these methods have achieved good results on optical datasets, they perform poorly on SAR datasets due to the inadaptability of the image enhancement methods used by the π-based consistency model in the SAR field and the demand of I-mix for large sample sizes. The work in [29] employed DEC for SAR ship OSR and category discovery. It employs both known and unknown classes for unsupervised training in the OSR stage. Subsequently, the model is fine-tuned using known class label data for novel category discovery. However, this approach faces limitations in practical scenarios where the OSR model training stage does not have access to data of unknown classes.

Inspired by the above limitations analysis, this article proposes an innovative OSR and novel category discovery framework. First, we build a feature extraction network called K-contrast Net, which consists of a trunk network composed of multiple residual blocks and two projection head branches. One branch reduces the empirical risk through the cross entropy loss, and the other branch reduces the open space risk by introducing the K-nearest neighbor (KNN) contrast loss. Both branches participate in network optimization together. Second, the sample representations are extracted from the trained K-contrast Net, and the outlier data detection is realized by a local outlier factor (LOF) classifier [30]. The detected known classes are classified and stored as labeled data in temporary memory. The third step is to divide a portion of the labeled known class samples into a probe set and use it together with the unknown class samples to estimate the number of unknown classes. At last, the number of unknown classes estimated in the previous step is used for deep transfer clustering, and the clustering results are optimized by using the prior of known classes to complete the classification of unknown classes. Extensive comparisons with other methods on the moving and stationary target automatic recognition (MSTAR) dataset verify the effectiveness and extensibility of our method. The main contributions of our work are summarized as follows:

1) *Framework:* A framework combining OSR with unknown class number estimation and unknown class classification is proposed, which makes up for the gap between OSR and incremental learning that requires manual labeling of unknown classes.
2) *OSR:* The proposed K-contrast Net incorporates the K-nearest neighbor contrast loss, mitigating open space risk. Compared to other comparison methods, it demonstrates superior multiclass OSR performance.
3) *DTC+:* By designing the image enhancement method in the consistency loss based on the π-model, the performance of DTC on SAR targets is further improved.

The rest of this article is organized as follows: In Section II, the related works are introduced in detail, including OSR and deep embedded clustering. Section III presents our proposed framework. The experimental results and related discussions are presented in Section IV. Finally, we provide the conclusion of the entire article in Section V.

II. Related Works

A. Open Set Recognition

Traditional classification methods based on the closed-set assumption can only classify objects that have appeared in the training set, and it is difficult to handle unknown class targets that may appear in battlefield reconnaissance. To address this limitation, researchers typically define the classes that have appeared in the training set as known classes and those that have not appeared as unknown classes. The task of determining

whether the class of the test target is known and classifying the known class samples is called OSR.

The existing OSR methods can be roughly categorized into machine-learning-based (ML-based) methods and deep-neural network-based (DNN-based) methods. ML-based methods have played an important role in previous years due to their good interpretability and domain adaptability. In this kind of method, OSR is described as an open-set risk function minimization problem [31] and studied around it. In [32], Scheirer proposed the Weibull-corrected support vector machine (W-SVM), which combines the SVM with extreme value theory (EVT) for score correction. The PI-SVM algorithm [33] utilizes EVT to model the decision boundary. Subsequently, Scherreik and Rigling [15] proposed the POS-SVM algorithm, in which the algorithm assigns a unique threshold to each training class memory region. The work in [34] proposed the sparse representation-based OSR algorithm, which utilizes EVT to fit the tail of the reconstruction error for both the matched and unmatched classes. Rudd et al. [35] proposed a thresholding algorithm by fitting the marginal distribution of the samples using EVT. With the rapid development of neural networks, researchers have tried to solve the OSR problem through the powerful computing power of deep neural networks. For example, Bendale and Boult [10] replaced the traditional SoftMax layer with OpenMax, which is the first DNN-based method of OSR. Oza and Patel [36] proposed the C2AE algorithm using the class-conditioned autoencoder. The authors in [12], [13], and [14] used the adversarial training advantages of GAN to improve the OSR effect.

Due to the differences in imaging mechanisms, target characteristics, and scene applications between radar images and optical images, the above methods do not perform well in SAR OSR. Researchers have put forward numerous OSR methods for SAR characteristics and have obtained satisfactory results when the number of known classes is small [19], [20], [21], [29], [37]. As the number of known classes increases and the risk of open space further escalates, there is still significant room for improvement in the above-mentioned methods. Therefore, we will try to address this issue in this article.

B. Deep Embedded Clustering

Deep clustering is to combine the clustering model with the network model and jointly optimize representation learning and clustering through clustering loss and network loss. The general paradigm of deep clustering can be expressed as follows:

$$\min L = \alpha L_n + \beta L_c, \quad \alpha \geq 0, \beta > 0 \tag{1}$$

where L_n represents the network loss and L_c represents the clustering loss. We can classify DEC from the perspective of the specific strategy used by the two losses. From the perspective of the neural network models adopted, representative methods include AE based [38], VAE based [39], GAN based [40], GNN based [41], and contrastive learning based [42]. From the perspective of the design of clustering loss L_c, the typical methods are K-means based [43], spectral clustering based [44], subspace clustering based [38], and Kullback–Leibler divergence based (KL based) [23], [24], [25]. Among them, the KL-based methods make the samples closer to the center of the cluster closer by minimizing the KL loss, which makes the data easier to distinguish in the feature space. The authors in [23] and [25] extended it to the field of transfer learning so that it can use the known class data distribution as a priori to cluster the unknown class data, which is very suitable for the application scenario of OSR. However, the application of this kind of method to SAR images is still limited. In [29], deep transfer clustering is used for unknown class clustering of SAR ships, but it needs unknown class sample data in the training stage of the OSR model, which has great limitations in practical applications. Therefore, there is still much room for improvement in the application of DEC in SAR image novel category discovery.

III. Proposed Method

A. Framework

In this section, we propose the OSR and category discovery framework for SAR images. The overall framework is shown in Fig. 1(a). Specifically, the framework is composed of two important parts, including OSR and category discovery. In the first part, the main task can be divided into two stages, namely, representation learning and outlier detection based on the LOF classifier. In the first stage, we propose K-contrast Net for representation learning and use cross-entropy loss and K-contrast loss for joint training. The SAR image $\boldsymbol{x}_i$ in D_{test} is first fed into the trained K-contrast Net to extract feature vectors $\boldsymbol{z}_i$, as follows:

$$\boldsymbol{z}_i = f_{\text{KCN}}(\boldsymbol{x}_i). \tag{2}$$

Here, f_{KCN} denotes the process of extracting features with K-contrast Net. Then, we can obtain its predicted class y_i through the classification head:

$$y_i = \arg\max_y \{h_y(\boldsymbol{z}_i)\}, \tag{3}$$

where $h_y(\boldsymbol{z}_i)$ represents the score of the sample z_i belonging to the class y. We extract all the features of D_{test} for the outlier detection stage. The training set D_{train} is used to fit the LOF classifier so that it can distinguish between known and unknown samples based on the density in the feature space. Then, we select the LOF threshold by calculating the best macro F1-score and accuracy over known classes on the validation set divided during the training phase. We determine the feature vector whose LOF classifier result is greater than the set threshold as an unknown class and store its corresponding sample in the memory buffer D_{unknown}. The feature vectors whose results are lower than the threshold are determined as known classes, and the corresponding samples and prediction labels y are stored in D_{known}, as follows:

$$\begin{cases} D_{\text{known}} \cdot \text{append}(\boldsymbol{x}_i, y_i), & LOF(z_i) \leq \text{threshold} \\ D_{unknown} \cdot \text{append}(\boldsymbol{x}_i), & LOF(z_i) > \text{threshold}. \end{cases} \tag{4}$$

The second part includes two modules: unknown class number estimation $\text{UNE}(\cdot)$ and improved deep transfer clustering $\text{DTC+}(\cdot)$, as shown in Fig. 1(d) and (e). We divide D_{known} (M classes) into probe sets D_l (N_l classes) and D_{probe} ($M - N_l$

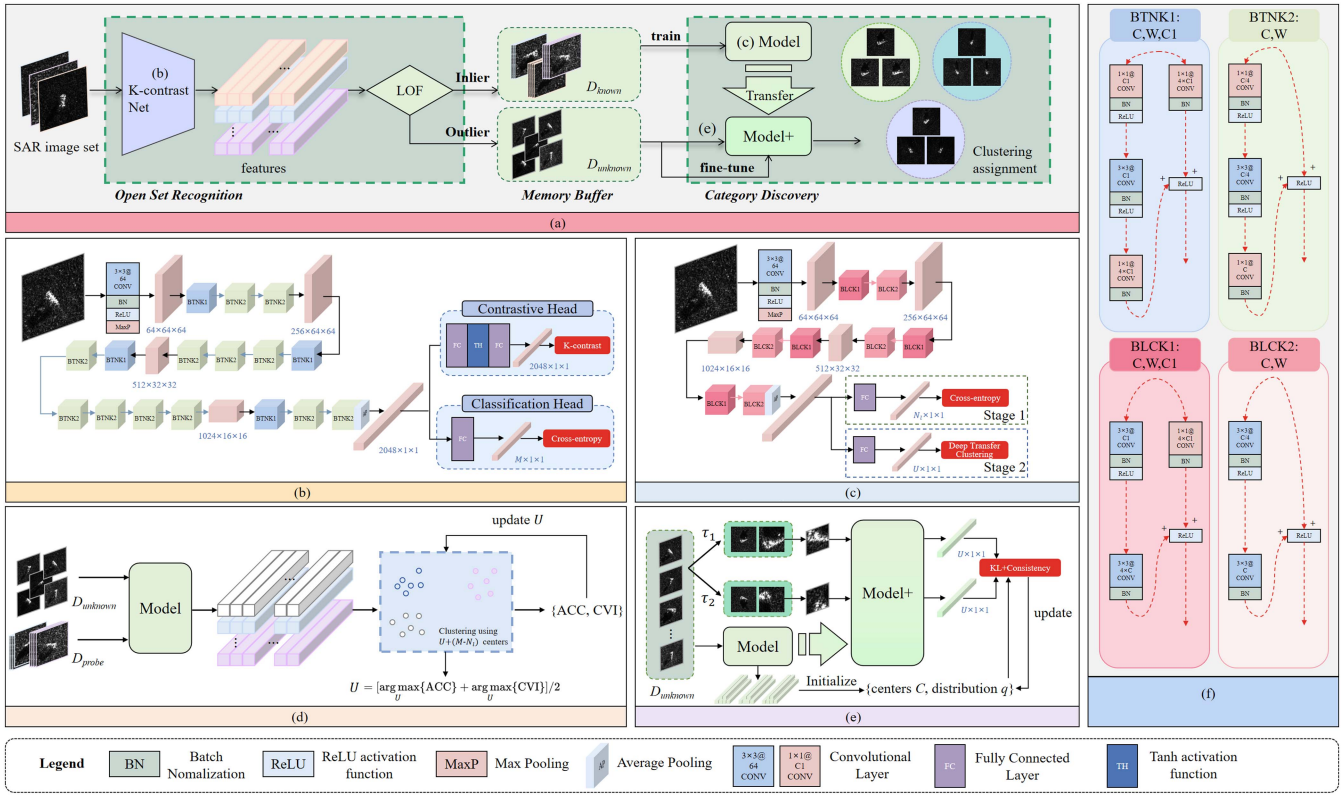


Fig. 1. Framework of the proposed method.

classes), and there is no overlap class between them. Then, we use D_l to train the lightweight model and use the trained model to extract the representations, as follows:

$$z_q = f_L(\boldsymbol{x}_q), \quad \boldsymbol{x}_q \in D_{\text{probe}} \cup D_{\text{unknown}}. \tag{5}$$

We estimate the number of unknown classes U through the unknown class number estimation module $\text{UNE}(\cdot)$:

$$\begin{aligned} K = \text{UNE}\Big(&\{z_1, y_1\}, \{z_2, y_2\}, \ldots, \{z_p, y_p\} \\ &\{z_{p+1}\}, \{z_{p+2}\} \cdots \{z_{p+l}\}\Big) \\ &\{z_1, y_1\}, \ldots, \{z_p, y_p\} \in D^*_{\text{probe}} \\ &\{z_{p+1}\}, \ldots, \{z_{p+l}\} \in D^*_{\text{unknown}} \end{aligned} \tag{6}$$

where D^*_{probe} and D^*_{unknown} are composed of the representations extracted from D_{probe} and D_{unknown}, respectively.

Finally, we use the trained lightweight model to transfer the knowledge of the known class samples to the new class, and use D_{unknown} to fine-tune the model to better perform the classification of unknown class samples:

$$\hat{y}_{p+j} = \text{DTC+}(z_{p+j}), \quad j \in [1, 2, \ldots, l] \tag{7}$$

where $\hat{y}_{p+j}$ is the predicted new category label of z_{p+j}.

B. K-Contrast Net

To simultaneously reduce empirical risk and open space risk, and enhance the accuracy of both inlier classification and outlier detection, we introduce the K-contrast Net, as illustrated in Fig. 1(b). The network consists of a feature extractor composed of many bottlenecks and two projection heads. One is called the classification head, which maps the feature vectors to an M-dimensional space (assuming the number of known classes is M) and calculates the cross-entropy loss. The other is called the contrastive head, which calculates the K-contrast loss after the features pass through two fully connected layers and are compared with other samples. The reasons for this design will be explained in detail below.

First, when the known class samples are used for training, we define the open space as follows [45], [46]:

$$\mathcal{O} = \mathcal{S}_o - \bigcup_{i \in M} \mathcal{S}_r(x_i) \tag{8}$$

where $\mathcal{S}_r(x_i)$ is a closed ball of radius r spanned by the known class training sample x_i. M is the number of all known class samples for training. Let $\mathcal{S}_o$ be a ball that includes all known class training examples as well as the open space $\mathcal{O}$. Then, probabilistic open space risk $\mathcal{R}_{\mathcal{O}}(f)$ can be defined as

$$\mathcal{R}_{\mathcal{O}}(f) = \frac{\int_{\mathcal{O}} f\,(x, \theta_f)\,\mathrm{d}x}{\int_{\mathcal{S}_o} f\,(x, \theta_f)\,\mathrm{d}x} \tag{9}$$

where f represents the discriminator, which is 1 when the sample x is a known class, and 0 otherwise. We extract the representation of SAR images through K-contrast Net for downstream tasks. A good representation should accomplish accurate classification of known class samples and identification of unknown class samples, which correspond to the reduction of empirical risk $\mathcal{R}_{\mathcal{E}}$ and open space risk $\mathcal{R}_{\mathcal{O}}$, respectively. For the empirical risk

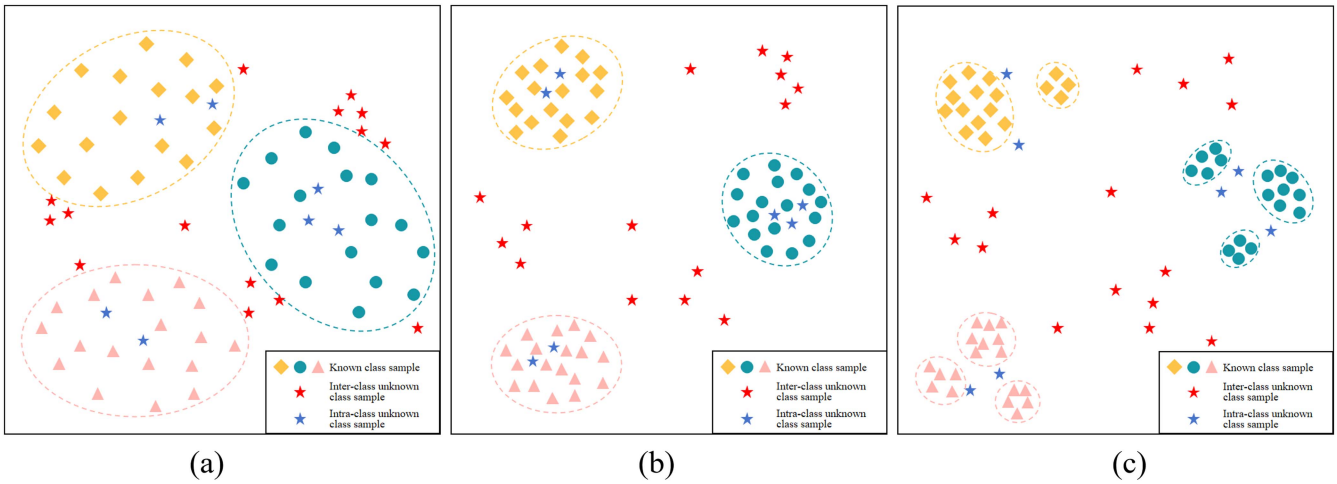


Fig. 2. (a) Original distribution of known class samples and unknown class samples. (b) Distribution of known class samples and unknown class samples when interclass open space risk loss is introduced. (c) Distribution of known class samples and unknown class samples when the joint loss function is introduced.

$\mathcal{R}_{\mathcal{E}}$, we utilize a fully connected layer to perform dimension reduction and incorporate the cross-entropy loss $\mathcal{L}_{ce}$ to optimize the network

$$\mathcal{L}_{ce} = \frac{1}{M}\sum_{i=1}^{M} -\log\frac{\exp\left(h_{y_i}(z_i)\right)}{\sum_{k\in M}\exp\left(h_k(z_i)\right)} \tag{10}$$

where z_i represents the representation extracted by the feature extractor, $h(\cdot)$ represents the classification head, and $h_k(z_i)$ represents the score of the sample z_i in the kth class.

Next, we consider the optimization of the open space risk. According to [46], unknown class samples in the feature space may appear between each known class, as shown by the red asterisks in Fig. 2(a). They may also appear within the space formed by the known class samples, as shown by the blue asterisks in Fig. 2(a). We can reduce the risk of classifying interclass unknown class samples as known classes by bringing similar samples closer together, reducing intraclass variance, and increasing interclass variance. In order to do this, for a single training sample γ and its k-nearest neighbors set $\mathcal{Z}$ (from the same class), we define its interclass open space risk loss in training as follows:

$$\mathcal{L}_{\text{inter},\gamma} = -\sum_{z_i\in\mathcal{Z}}\log\frac{\sum_{z_l\in D^*}\exp\left(\gamma\cdot z_l/\delta\right)}{\sum_{z_v\in D_o\bigcup\{z_i\}}\exp\left(\gamma\cdot z_v/\delta\right)} \tag{11}$$

where D^* denotes all training samples whose classes are the same as γ, and D_o denotes all training samples whose classes are different from γ. δ is a temperature coefficient, which can regulate the tension and repulsion between the contrast samples. Through analysis, it can be seen that the molecular term of the loss reduces the distance between the current sample and all the samples of the same class, and the denominator extends the distance between the current sample and all the samples of different classes. This leads to an interclass variance reduction in the feature space, making the different classes more distinguishable from each other, as shown in Fig. 2(b). As a result, more interclass unknown class samples can be distinguished, but unknown class samples within the known class range will be more difficult to correctly recognize due to the smaller intraclass variance.

To address this issue, we introduce intraclass open space risk loss as follows:

$$\mathcal{L}_{\text{intra},\gamma} = -\sum_{z_i\in\mathcal{Z}}\log\frac{\exp\left(\gamma\cdot z_i/\delta\right)}{\sum_{z_l\in D^*}\exp\left(\gamma\cdot z_l/\delta\right)}. \tag{12}$$

The molecular term of the $\mathcal{L}_{\text{intra},\gamma}$ only reduces the distance between the current sample and its k-nearest neighbor samples of the same class, and the denominator pushes away the distances to other samples of the same class. This forces a single known class that was originally a whole clustering space to split into multiple smaller clusters, and the samples in the cluster are closer, as shown in Fig. 2(c). Therefore, the unknown class samples within the known class range are easier to recognize. The above two loss functions are jointly involved in the optimization, so we combine them into the following loss function, which is named K-contrast loss:

$$\begin{aligned}\mathcal{L}_{k-cnt} &= \mathcal{L}_{\text{intra}} + \mathcal{L}_{\text{inter}} \\ &= -\sum_{j=1}^{M}\frac{1}{|\mathcal{Z}|}\sum_{z_i\in\mathcal{Z}}\left[\log\frac{\exp(\gamma_j\cdot z_i/\delta)}{\sum_{z_l\in D^*}\exp(\gamma_j\cdot z_l/\delta)}\right. \\ &\quad \left.+\sum_{z_i\in\mathcal{Z}}\log\frac{\sum_{z_l\in D^*}\exp(\gamma_j\cdot z_l/\delta)}{\sum_{z_v\in D_o\bigcup\{z_i\}}\exp(\gamma_j\cdot z_v/\delta)}\right] \\ &= -\sum_{j=1}^{M}\frac{1}{|\mathcal{Z}|}\sum_{z_i\in\mathcal{Z}}\log\frac{\exp(\gamma_j\cdot z_i/\delta)}{\sum_{z_v\in D_o\bigcup\{z_i\}}\exp(\gamma_j\cdot z_v/\delta)}.\end{aligned} \tag{13}$$

Combining it with $\mathcal{L}_{ce}$ to jointly optimize the feature extractor, the joint loss function of the training process is as follows:

$$\mathcal{L}_{con} = \mu\cdot\mathcal{L}_{ce} + (1-\mu)\cdot\mathcal{L}_{k-cnt} \tag{14}$$

where μ is a hyperparameter to control the weight of each part of the loss.

C. Unknown Class Sample Detection Based on LOF

Through the trained K-contrast Net, we extracted the representation of the test samples. Subsequently, we detect unknown class samples among them. Recent studies on SAR OSR typically use statistics-based methods to achieve the rejection of unknown class samples [20], [21], but they need to assume that the data obeys a specific probability distribution, which limits the scope of application. In contrast, we use LOF-based unknown class detection, which does not depend on data distribution. According to [47], we define the reachability distance of an object p w.r.t. object o as follows:

$$reach{-}dist_k(p,o) = \max\{k{-}\text{distance}(o), d(p,o)\} \tag{15}$$

where $d(p,o)$ is the distance between p and o, and $k{-}\text{distance}(o)$ is the distance of o to its kth nearest neighbor. Based on the reachability distance, the local reachability density (lrd) is defined as

$$lrd_k(p) = 1/\left(\frac{\sum_{o\in N_k(p)} reach{-}dist_k(p,o)}{|N_k(p)|}\right) \tag{16}$$

where $N_k(p)$ is the k-nearest neighbors of p. According to the above equation, if a data point is relatively distant from other points, its local reachability density is small. However, for scenes with an imbalanced known class distribution, measuring the degree of anomaly of a data point should not only depend on its absolute local density but also on its relative density with neighboring data points. Therefore, the abnormal score of a sample is defined as follows:

$$LOF_k(p) = \frac{\sum_{o\in N_k(p)} \frac{lrd_k(o)}{lrd_k(p)}}{|N_k(p)|}. \tag{17}$$

The above equation calculates the outlier degree of p. It is easy to see that when the local reachability density of p is lower, the local reachability density of the k-nearest neighbors of p is higher, and the LOF score of p is correspondingly higher. By setting a threshold, we can determine the samples whose LOF scores are greater than the threshold as unknown classes. Finally, we store the corresponding images and labels of the known class samples into D_{known} and the corresponding images of the unknown class samples into D_{unknown}.

D. Unknown Class Number Estimation (UNE)

In our proposed framework, the UNE module is employed to estimate the number of unknown classes by using the prior information provided by the known class samples. The architecture of the UNE module is shown in Fig. 1(d). Suppose that there are M known classes in D_{known}, we take N_l classes to form D_l, and the remaining $M - N_l$ classes are used as the probe set D_{probe} combined with the D_{unknown} for class number estimation. Then we propose a lightweight model, as shown in Fig. 1(c), and use D_l to train it. The training uses cross-entropy loss for supervised classification training. By doing so, the model learns the information of known classes.

Next, we use the trained model to extract the features of samples in D_{probe} and D_{unknown} for the subsequent clustering process. The corresponding feature sets of the above two are D^*_{probe} and D^*_{unknown}. We then further split the D^*_{probe} into a subset $D^*_{\text{probe},v}$ of L classes and a subset $D^*_{\text{probe},a}$ of $(M - N_l - L)$ classes, which we call the validation probe set and anchor probe set, respectively. We use a semisupervised k-means algorithm with $U + (M - N_l)$ centers to estimate the number of classes U in D^*_{unknown}. Specifically, we enforce features in the anchor probe set $D^*_{\text{probe},a}$ to be assigned to the corresponding clusters according to their ground-truth labels during clustering, while features in the validation probe set $D^*_{\text{probe},v}$ are treated as unlabeled data. We conduct experiments by changing the U value several times and evaluate them using average clustering accuracy (ACC) and cluster validity index (CVI), which are defined as follows:

$$ACC = \max_{g\in \text{Sym}(L)} \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{\bar{y}_i = g(y_i)\} \tag{18}$$

$$CVI = \sum_{z\in D^*_{\text{unknown}}} \frac{b(z) - a(z)}{\max\{a(z), b(z)\}} \tag{19}$$

where $\bar{y}_i$ and y_i represent the ground-truth label and clustering assignment for sample $z_i \in D^*_{\text{probe},v}$ and $\text{Sym}(L)$ is the group of permutations of L elements (as a clustering algorithm recovers clusters in an arbitrary order). $a(z)$ represents the average distance between sample z and all other samples within the same cluster. $b(z)$ represents the smallest average distance of sample z to all samples in any other cluster. The ACC on $D^*_{\text{probe},v}$ and the CVI on D^*_{unknown} are recorded in each experiment. Finally, the U value corresponding to the highest accuracy and the U value corresponding to the highest CVI are obtained, and the average of the two is the estimation result of the number of unknown classes. If the average of the above two estimates is a noninteger, round it down. The detailed approach is given in Algorithm 1.

E. Improved Deep Transfer Clustering (DTC+)

Our method is based on the DTC incorporating consistency loss proposed in [23], and it is enhanced by designing specifically tailored enhancement methods for SAR images. According to [23], we use the pretrained lightweight model to extract features of D_{unknown}, use PCA to reduce the feature dimension to the estimated unknown class number U, and use the k-means algorithm to initialize the cluster centers $C = \{\nu_k, k = 1, 2, \ldots, U\}$, as shown in Fig. 1(e). After that, we use PCA as the last layer of the model, replacing the previous classification head. The parameters of this layer will participate in the subsequent fine-tuning process with the pretrained model. We fine-tune the model using samples from D_{unknown}, and the loss function for this process consists of two parts: Kullback–Leibler (KL) divergence loss and consistency loss.

In the first part, we use the KL divergence of two distributions as the loss instead of directly minimizing the k-means objective function. Otherwise, the optimization process will quickly collapse the learned representation vectors to the closest cluster centers, thus losing the representation learning effect.

Algorithm 1: Unknown Class Number Estimation.

Input: anchor probe set $D^*_{probe,a}$, validation probe set $D^*_{probe,v}$, unknown class feature set $D^*_{unknown}$.
Output: Estimation result of unknown class number $\hat{U}$.

Initialized: the maximum value of average clustering accuracy $ACC_{max} = 0$, the maximum value of cluster validity index $CVI_{max} = 0$, $\hat{U}_{acc}$, $\hat{U}_{cvi}$.
for $U \leftarrow 0, 1, \ldots, N$ **do**
  Perform semi-supervised k-means on $D^*_{probe,a} \cup D^*_{probe,v} \cup D^*_{unknown}$ with U+(M-N_l) centers, forcing samples in $D^*_{probe,a}$ to assign to the ground-truth class labels;
  Compute ACC on D^*_{probe} and CVI on $D^*_{unknown}$;
  if $ACC > ACC_{max}$ **then**
    $\hat{U}_{acc} = U$,
    $ACC_{max} = ACC$.
  end
  if $CVI > CVI_{max}$ **then**
    $\hat{U}_{cvi} = U$,
    $CVI_{max} = CVI$.
  end
end
return $\hat{U} = (\hat{U}_{acc} + \hat{U}_{cvi})/2$.

Based on this, we let $P(k|i)$ be the probability of assigning sample $i \in \{1, 2, \ldots, N\}$ to cluster $k \in \{1, 2, \ldots, U\}$, and use the following parametrization of this conditional distribution by assuming a students t distribution:

$$P(k|i) \propto \left(1 + \frac{||z_i - \nu_k||_2^2}{\alpha}\right)^{-\frac{\alpha+1}{2}}. \tag{20}$$

Since the data sampling process is uniform and random, it satisfies $p(i) = 1/N$. Then we can get:

$$P(i,k) = \frac{P(k|i)}{N}. \tag{21}$$

Instead of maximizing the likelihood of P directly, we map it to a sharper distribution Q. Let $Q(k|i) \propto P(k|i) \cdot P(i|k)$, then we can get the following results through the Bayes rule:

$$Q(k|i) \propto \frac{P(k|i)^2}{\sum_{i=1}^{N} P(k|i)}. \tag{22}$$

The molecular part of the above equation makes the distribution sharper by increasing the power, and the denominator part balances it by normalizing according to the frequency per cluster. In this way, we can construct a new distribution Q corresponding to the current P at each epoch. By minimizing the KL divergence between joint distributions $Q(i,k) = Q(k|i)/N$ and $P(i,k)$, we can sharpen the distribution P, so that the samples close to the cluster center become closer and the samples are easier to distinguish in the feature space. The KL divergence loss is defined as

$$\mathcal{L}_{KL} = KL(Q||P) = \frac{1}{N}\sum_i \sum_k Q(k|i) \ln \frac{Q(k|i)}{P(k|i)}. \tag{23}$$

During the fine-tuning process, the distribution Q and the model parameter are alternately updated, and the detailed steps are shown in Algorithm 2.

In the second part, we introduce the consistency loss based on the π-model [27]. By drawing the model output results of two randomly enhanced views of the same data closer, the model can learn some invariances (such as rotation invariance), ignore background, speckle noise, and other secondary factors, and learn more discriminative representations. Suppose that the two enhanced views $\tau_1(\boldsymbol{x})$ and $\tau_2(\boldsymbol{x})$ of the same SAR image $\boldsymbol{x}$ are input into the model, and the sample assignment probabilities output are $P_1(k|i)$ and $P_2(k|i)$, respectively. Then, the loss function is defined as follows:

$$\mathcal{L}_{\pi} = \frac{1}{N}\frac{1}{U}\sum_i \sum_k ||P_1(k|i) - P_2(k|i)||_2^2. \tag{24}$$

Although the image enhancement methods proposed in [23] (random translate with reflect) and [28] (I-mix) perform well in the fields of optical images and natural language processing, they are not effective in SAR images due to the imaging mechanism and data size of the SAR image set. In fact, these methods may even collapse the fine-tuning process into random results. Based on our previous research experience with contrastive learning of SAR images in the SimCLR framework [48], we find that random-resized-crop and random-horizontal-flip can achieve better representations. Therefore, we design the enhanced method as a combination of random-resized-crop and random-horizontal-flip to better learn the structural features of SAR and rotation invariance from a top-down perspective.

Finally, we combine the two loss functions and use $\beta(t)$ to balance them, as follows:

$$\mathcal{L}_{ft} = \beta(t)\mathcal{L}_{KL} + (1 - \beta(t))\mathcal{L}_{\pi}. \tag{25}$$

$\beta(t)$ will decrease with the increase of the training epoch number to adjust the weight of each part of the loss function. Moreover, in order to prevent the training from collapsing to random clustering due to the small number of samples, we set the condition that the KL divergence loss should be disabled. When the iteration reaches a certain number of epochs, only the consistency loss will be used for subsequent optimization.

IV. Experiments

In this section, we conducted extensive experiments on the MSTAR dataset to demonstrate the effectiveness of our framework. First, we provide a detailed introduction to the MSTAR dataset. Second, we compare our approach with existing methods under multiple scenarios to validate the OSR performance on multiclass targets. Third, we verify the novel class discovery performance of our method, i.e., unknown class number estimation and classification capability, under different data distribution scenarios. Finally, we conduct ablation studies and discuss the impact of some hyperparameters on the experimental

Algorithm 2: Improved Deep Transfer Clustering.

Input: unknown class sample set $D_{unknown}$, estimated number of unknown classes $\hat{U}$, augmentation set τ.

Output: predicted new category label of unknown class sample $\hat{y}$.

Initialized: cluster centers C, target distributions Q, weight function $\beta(t_0)$, parameters of pre-trained lightweight model $\boldsymbol{\theta}_L$.

for $t \leftarrow 0, 1, \ldots, N_{epochs}$ **do**

  for $j \leftarrow 1, 2, \ldots, N_{set}$ **do**

    $\boldsymbol{x}_j^{(1)} = \tau^{(1)}(\boldsymbol{x}_j)$,

    $\boldsymbol{x}_j^{(2)} = \tau^{(2)}(\boldsymbol{x}_j)$,

    $P_1(k|j) = f_L(\boldsymbol{x}_j^{(1)}, \boldsymbol{\theta}_L), \quad k = 1, 2, \cdots, \hat{U}$

    $P_2(k|j) = f_L(\boldsymbol{x}_j^{(2)}, \boldsymbol{\theta}_L), \quad k = 1, 2, \cdots, \hat{U}$

  end

  $\mathcal{L}_{ft} = \frac{\beta(t)}{N_{set}} \sum_{j=1}^{N_{set}} \sum_{k=1}^{\hat{U}} Q(k|j) \ln \frac{Q(k|j)}{P(k|j)}$

  $+ \frac{1-\beta(t)}{N_{set}\hat{U}} \sum_{j=1}^{N_{set}} \sum_{k=1}^{\hat{U}} ||P_1(k|j) - P_2(k|j)||_2^2$

  Using stochastic gradient descent to update $\boldsymbol{\theta}_L$.

  Update cluster centers C and target distributions Q.

end

for $j \leftarrow 1, 2, \ldots, N_{set}$ **do**

  Predict $P(k|j)$ for $k = 1, 2, \cdots, \hat{U}$.

  $\hat{y}_j = \arg\max_k P(k|j)$.

end

return $\hat{y}$.

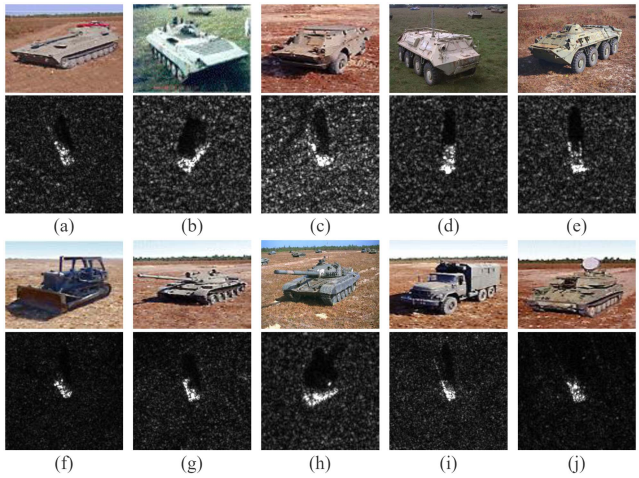


Fig. 3. SAR images and corresponding optical images of ten classes of targets in MSTAR dataset. (a) 2S1. (b) BMP2. (c) BRDM2. (d) BTR60. (e) BTR70. (f) D7. (g) T62. (h) T72. (i) ZIL131. (j) ZSU234.

TABLE I
NUMBER OF 10-CLASS TARGETS

Target	Number (17°)	Number (15°)
2S1	299	274
BMP2	233	196
BRDM2	298	274
BTR60	256	195
BTR70	233	196
D7	299	274
T62	299	273
T72	232	196
ZIL131	299	274
ZSU234	299	274

results. The laptop used in our experiments has an Intel Core i9-13900HX CPU, an NVIDIA GeForce RTX 4060 Laptop GPU, and 16 GB of RAM on the Windows 11 system.

A. Dataset

In this article, we utilize the MSTAR dataset for our experiments. The MSTAR is a standardized database for SAR target recognition that encompasses ten classes of ground targets, including BMP2, BTR70, T72, BTR60, 2S1, BRDM2, D7, T62, ZIL, and ZSU. Among these targets, BMP2 and T72 are distinguished by three variants, namely BMP2-c9563, BMP2-c9566, BMP2-c21, T72-132, T72-812, and T72-s7. For our experiment, we treat BMP2-c9563 and T72-132 as the reference forms for these two classes. The radar operates in the X-band and employs spotlight mode imaging with an image resolution of 0.3 m × 0.3 m. Each image spans an elevation angle of approximately 3°, with data acquisition depression angles set at 15° and 17°, respectively. Similar to the literature [12], [21], we designate 17° as the training set and 15° as the testing set. Fig. 3 illustrates various target objects and their corresponding optical images. It can be seen that different types of SAR images exhibit high similarity and are more difficult to distinguish than optical images. In this article, the images are uniformly cropped to the size of 64 × 64 to reduce the fitting of background information and speckle noise interference during the training process. The numbers of the 10-class targets are shown in Table I.

B. Multiclass OSR

In this section, we focus on verifying the performance of our framework on multiclass OSR. We choose the MS-SSIM autoencoder (MS-SSIM-AE) [20] and EVM [35] as comparison algorithms because MS-SSIM-AE is a recent optimal baseline for OSR in SAR, while EVM is a representative classical algorithm. Similarly, W-SVM [32] and OSmIL [21] are also used to study the OSR of SAR images. Since OSmIL is not publicly available and the results in [20] prove that MS-SSIM-AE is significantly better than OSmIL, we do not take it into account. Moreover, W-SVM is significantly inferior to EVM under all indicators, so we only need to prove the advantages of our method by comparing the results of EVM.

To facilitate comparison, we follow the settings in [20] and select M classes as known classes. In order to fully demonstrate the OSR ability of our method under different number of categories, the number of known classes is taken from 6 to 9, and a new class is added each time compared with the previous setting, rather than randomly selecting the known classes, to minimize the change in training data. The detailed settings are as shown in Table II. We train the K-contrast Net using the known classes in

TABLE II
MSTAR DATASET DIVISION UNDER DIFFERENT KNOWN CLASS NUMBER SCENARIOS

Settings	Dataset Division	
	Known Classes	Unknown Classes
$M = 6$	BMP2, BTR70, T72, BTR60, 2S1, BRDM2	D7, T62, ZIL131, ZSU234
$M = 7$	BMP2, BTR70, T72, BTR60, 2S1, BRDM2, D7	T62, ZIL131, ZSU234
$M = 8$	BMP2, BTR70, T72, BTR60, 2S1, BRDM2, D7, T62	ZIL131, ZSU234
$M = 9$	BMP2, BTR70, T72, BTR60, 2S1, BRDM2, D7, T62, ZIL131	ZSU234

TABLE III
EXPERIMENTAL RESULTS OF THE METHODS UNDER DIFFERENT KNOWN CLASS SETTINGS

Settings	EVM		MS-SSIM-AE		Our method	
	F1-score	ACC	F1-score	ACC	F1-score	ACC
$M = 6$	85.10	85.32	**93.68**	**93.44**	91.66	90.67
$M = 7$	83.72	81.36	92.80	91.55	**94.29**	**92.78**
$M = 8$	82.02	79.51	93.11	91.63	**93.33**	**91.91**
$M = 9$	80.30	80.04	92.07	91.26	**96.57**	**96.44**

The higher values are highlighted in bold.

the training set. During testing, we use all categories for testing, i.e., the testing set includes both known and unknown classes. After training, the K-contrast Net not only needs to recognize and classify the known M classes of targets, but also identify the remaining (10-M) classes of targets as unknown. Similar to [12], [20], [21], we use F1-score and accuracy to evaluate the OSR integral performance of each method, as defined in (28) and (29):

$$\text{precision} = \frac{\sum_{i=1}^{N}\text{precision}_i}{N}, \text{precision}_i = \frac{TP_i}{TP_i + FP_i} \tag{26}$$

$$\text{recall} = \frac{\sum_{i=1}^{N}\text{recall}_i}{N}, \text{recall}_i = \frac{TP_i}{TP_i + FN_i} \tag{27}$$

$$F1 - \text{score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{28}$$

$$ACC = \frac{\sum_{i=1}^{M} TP_{ii}}{N_{All}} \tag{29}$$

where TP and FP denote the counts of correctly and incorrectly recognized test samples to be known classes, and FN is the count of incorrectly identified test samples as unknown classes. N represents the number of classes, and N_{ALL} represents the number of test samples. We set the μ in (14) to 0.9 and set the K value in the K-contrast loss to 40. The number of training epochs for all experiments is set to 100. The experimental results are shown in Table III. It can be seen that our method has better results than the previous optimal baseline in most scenarios. Specifically, when the number of known classes increases, the performance of other methods declines, while our method remains stable. In the cases where the known class number $M = 7, 8, 9$, our method has a higher F1-score than MS-SSIM-AE [20] by 0.65, 0.22, and 4.50, respectively, and higher accuracy than MS-SSIM-AE by 0.15, 0.28, and 5.18, respectively. When $M = 6$, our method is inferior to MS-SSIM-AE, which is due to the insufficient model training caused by the reduction in the number of training samples. Nevertheless, the results obtained by our method are still much higher than those of EVM. In summary, our method exhibits excellent OSR performance in multiclass scenarios.

C. Estimation of the Number of Unknown New Classes

Before clustering the unknown samples, we need to estimate the number of classes in them. We design experiments under four scenarios where the known class numbers $M = 6$, 7, 8, and 9, respectively. When $M = 6$ or 7, we take two classes of known class samples as the probe set D_{probe}, with one class used as the anchor probe set $D_{\text{probe},a}$ and the other as the valid probe set $D_{\text{probe},v}$. When $M = 8$ or 9, we take three classes of known class samples as D_{probe}, and divide the $D_{\text{probe},a}$ and the $D_{\text{probe},v}$ according to a 2:1 classes ratio to transfer more known class information for estimation. The experimental results are shown in Table IV. It can be seen that in the above settings, the number of unknown classes can be estimated without error, which indicates that the method can stably provide accurate estimation and serve the next clustering process.

D. Unknown Class Sample Clustering

In this section, we demonstrate the performance of the proposed method on unknown-class sample clustering. According to the setting of $M = 7$ in Table IV, we fine-tune the pretrained model with the remaining three unknown classes and classify them according to the estimated number of unknown classes in the previous experiment. We directly use k-means [49] as a baseline to demonstrate the difficulty of clustering without optimization and compare our method with KCL [24], MCL [25], Ranking Statistics (RS) [50], DTC, and some variants of DTC [23]. Among them, DTC-Baseline uses DEC loss for model training, DTC-TE incorporates consistency constraints between current prediction and temporal ensemble prediction of each sample, and DTC-π introduces consistency loss between samples and their enhanced views. Different from the DTC-π algorithm, our method replaces the image enhancement method with a more suitable method for SAR targets and sets a threshold for KL

TABLE IV
UNKNOWN CLASSES NUMBER ESTIMATION RESULTS

Settings	Division of D_{known}			Metrics		
	D_l	$D_{\text{probe},a}$	$D_{\text{probe},v}$	GT	Ours	Error
$M=6$	BMP2, BTR70, T72, BTR60	2S1	BDRM2	4.0	4.0	0.0
$M=7$	BMP2, BTR70, T72, BTR60, 2S1	D7	BDRM2	3.0	3.0	0.0
$M=8$	BMP2, BTR70, T72, BTR60, 2S1	BRDM2, D7	T62	2.0	2.0	0.0
$M=9$	BMP2, BTR70, T72, BTR60, 2S1, BRDM2	D7, T62	ZIL131	1.0	1.0	0.0

TABLE V
RESULTS OF TRANSFERRING FROM KNOWN CLASS TO UNKNOWN CLASS

Methods	Metrics		
	ACC_u	NMI	ARI
k-means [50]	0.4300	0.0442	0.0262
KCL [24]	0.5213	0.1818	0.1655
MCL [25]	0.5737	0.1716	0.1718
DTC-Baseline [23]	0.4994	0.1165	0.1101
DTC-TE [23]	0.5335	0.1451	0.1578
DTC-π [23]	0.5518	0.1798	0.1876
RS [51]	0.5521	0.1343	0.1273
Our method	**0.5956**	**0.3276**	**0.3349**

The higher values are highlighted in bold.

TABLE VI
STUDY ON CONSISTENCY LEARNING IMAGE ENHANCEMENT STRATEGIES

Enhancement methods					Metrics		
Crop	Trans	Flip-H	Flip-V	Rotate	ACC_u	NMI	ARI
	✓		✓		0.5311	0.1425	0.1543
	✓			✓	0.5457	0.1672	0.1777
	✓	✓			0.5518	0.1798	0.1876
✓					0.5676	0.2551	0.2698
✓				✓	0.5749	0.2799	0.2884
✓		✓	✓		0.5907	0.2950	0.3055
✓			✓		**0.5968**	0.2992	0.3099
✓		✓			0.5956	**0.3276**	**0.3349**

The higher values are highlighted in bold.

divergence loss. When using the RS method, first, all the samples to be detected are used for unsupervised pretraining of the model simultaneously, and then the known class samples are used for supervised fine-tuning. Finally, positive sample pairs of the same class are constructed using high-similarity samples with the same indices in the feature amplitude top-5, and pseudolabels are created for training. When the training reaches a certain number of epochs, the KL divergence loss is disabled. In addition, we compute the consistency loss between two enhanced views instead of that between an enhanced view and its original image, which is more conducive to the role of consistency constraints in representation learning. The experiments use clustering accuracy (ACC_u), normalized mutual information (NMI) [51], and adjusted rand index (ARI) [52] as metrics, all of which are larger means better effect. The experimental results are shown in Table V. It can be observed that our method outperforms other methods in all metrics, with a 16.56% improvement in clustering accuracy compared to directly using k-means and a 4.38% improvement compared to the optimal DTC variant. The RS method requires constructing pseudolabels from samples that meet the top-5 condition. However, it is challenging for the features extracted from SAR images to meet this condition, resulting in a limited number of positive sample pairs available for training in each epoch, leading to suboptimal clustering performance. Our method has a 4.35% higher accuracy compared to the RS method.

To select the most appropriate combination of image enhancement for SAR targets and explore, which consistency constraint can better guide the clustering, we conduct extensive experiments on combinations between random-resized-crop [48], random-horizontal-flip [48], random-vertical-flip [48], random-translate-with-reflect [23], and random-rotation [53], which are represented by Crop, Flip-H, Flip-V, Trans, and Rotate in subsequent experiments. Crop randomly cuts 0.4–1.0 of the original image area and resizes it to the original image size. Trans cuts a part of the original image and translates it randomly, then fills the remaining part with the original image patches. Rotate randomly rotates the original image in the range of $[-5, 5]$, and the empty part caused by rotation is filled with white. Flip-H and Flip-V flip the original image horizontally and up and down at a probability of 0.5, respectively. We continue to experiment with the setting of $M = 7$, and the results are shown in Table VI.

It can be seen that the combination of Crop with Flip-V or Flip-H can obtain better results, but the combination of these three will make the performance slightly worse. We think that it is due to the excessive combination categories caused by excessive enhancement methods, which makes the representation learning unstable. In summary, using the combination of Crop and Flip-H to enhance the consistency constraint can significantly improve the effect of DTC.

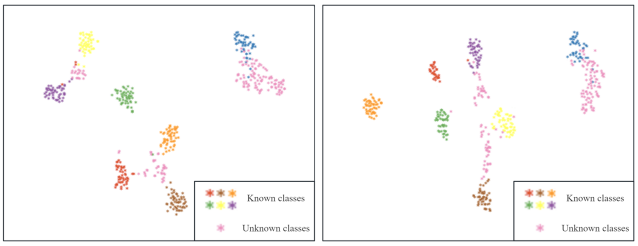
E. Ablation Experiments

In the above experiments, we compared our approach with other state-of-the-art methods. In this part, we conducted ablation studies to demonstrate the effectiveness of the introduced K-contrast loss and the influence of the choice of M.

1) Effect of K-Contrast Loss: To demonstrate the improvement of OSR performance by introducing K-contrast loss, we remove the contrastive head from the K-contrast Net and only use the classification head to train the network. We then use the trained network to extract features, classify test samples, and finally use the LOF classifier to detect unknown class samples. The experimental results of OSR are compared with those obtained when the contrastive head is used, as shown in Table VII. The left column of the table is the result of not introducing K-contrast loss, that is, only using cross-entropy loss for training. The right column is the result of introducing K-contrast loss and joint training with cross-entropy loss. It can be seen that the introduction of K-contrast loss can bring great improvement in all scenarios. Specially, from $M = 6$ to 9, the

TABLE VII
COMPARISON BEFORE AND AFTER THE INTRODUCTION OF K-CONTRAST LOSS

Settings	LOF-CE		Our method	
	F1-score	ACC	F1-score	ACC
$M = 6$	87.31	82.84	91.66	90.67
$M = 7$	89.50	85.36	93.45	91.70
$M = 8$	88.97	86.19	93.33	91.91
$M = 9$	91.05	89.45	96.57	96.44



(a) (b)

Fig. 4. t-SNE visualization of features of some MSTAR samples in different methods. (a) LOF-CE. (b) Our methods.

recognition accuracy has improved by 7.83, 6.34, 5.72, and 5.39, respectively. The F1-score has been improved by 4.35, 3.95, 4.36, and 5.52, respectively.

To visualize the performance of our method more intuitively, we choose the scene of $M = 7$, randomly select samples from each class, and use t-SNE to visualize the results obtained under the two methods, as shown in Fig. 4. Compared with only using cross-entropy loss for training, our method splits the large clusters into multiple small clusters, and the samples in small clusters are closer, which reduces the open space risk. This provides more opportunities to recognize unknown class samples between classes, such as unknown class points within blue points and purple points.

2) Influence of the Choice of K in K-Contrast Loss: To discuss the effect of the choice of K in the K-contrast loss on the performance of OSR, we compare the experimental results under different K values while fixing the other hyperparameters. We chose the scenario with $M = 7$ from Table II for the experiment and set K in the K-contrast loss to 10, 20, 30, 40, 50, and 60, respectively. The results are shown in Fig. 5. We observe that the F1-score and accuracy tend to increase first and then decrease as the value of K increases. This is because when K starts from 0 and increases, the K-contrast loss plays a role in dividing the known classes into smaller clusters, reducing the open space risk and improving the OSR performance. However, when K continues to increase, the semantic space of the known classes is overcompressed, leading to more unknown class samples within the known class range being misclassified as known classes, resulting in a decline in performance. Therefore, we should select an appropriate K value based on the number of known classes and the dataset scale.

V. CONCLUSION

In this article, we propose a framework that integrates SAR image OSR, unknown class count estimation, and unknown class target classification, providing a new approach for ATR and unknown sample processing in battlefield reconnaissance. In the OSR module, we propose the K-contrast Net, which combines the K-contrast loss with the cross-entropy loss, allowing the K-contrast Net to divide a category into smaller, tighter clusters, reducing the open space risk, and leading to a more accurate classification and rejection performance compared with other state of the arts. Under the guidance of the known class knowledge learned by the lightweight model, we estimate the number of unknown classes by deep clustering and classify the unknown class samples, thereby reducing the time cost of manual interpretation and labeling. The KL divergence loss and the consistency loss are introduced into the clustering process. By setting a stop threshold and designing the enhancement method suitable for SAR images, the fine-tuning training process is smoother and the deep clustering effect is improved. The experimental results show that our proposed framework can be well applied to scenarios with high real-time requirements.

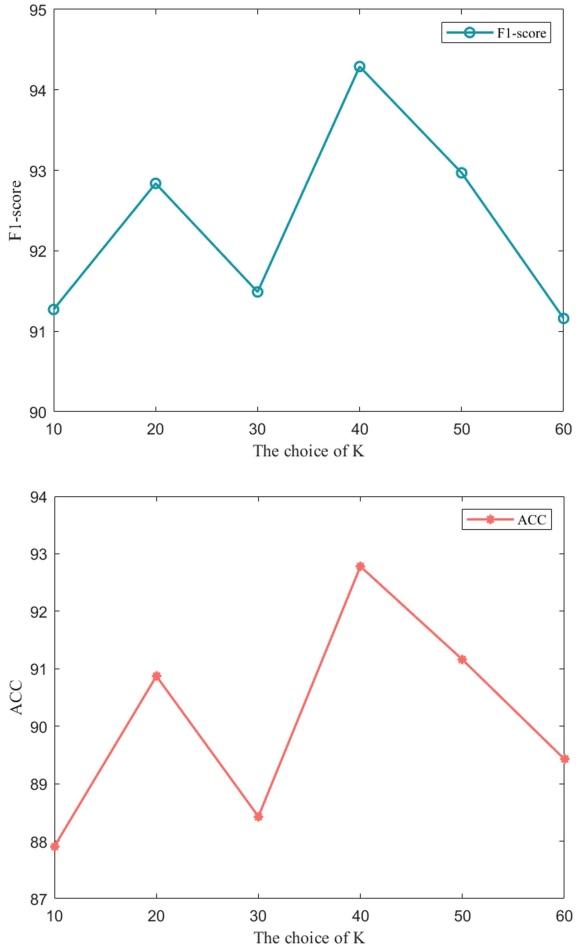


Fig. 5. Influence of the change of K value on the accuracy and F1-score.

In practical application scenarios, sometimes we do not have enough reconnaissance time to collect a large number of observation data of the same unknown class sample, and the obtained unknown class data may be much less than the data of the class used for training, which leads to the problem of

few-shot deep clustering. In the future, we will further expand our work to enable our proposed framework to perform well in the classification and annotation of few-shot unknown classes. Besides, we will conduct research on more advanced SAR image augmentation techniques. Furthermore, if there are SAR datasets with a large number of classes and a balanced distribution of samples in the future, we will carry out research on OSR for such targets.

REFERENCES

[1] C. Li, L. Du, Y. Li, and J. Song, "A novel SAR target recognition method combining electromagnetic scattering information and GCN," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4508705.

[2] 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.

[3] J.-H. Choi, M.-J. Lee, N.-H. Jeong, G. Lee, and K.-T. Kim, "Fusion of target and shadow regions for improved SAR ATR," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5226217.

[4] C. Wang et al., "SAR ATR under limited training data via MobileNetV3," in *Proc. IEEE Radar Conf.*, 2023, pp. 1–6.

[5] C. Mao, L. Huang, Y. Xiao, F. He, and Y. Liu, "Target recognition of SAR image based on CN-GAN and CNN in complex environment," *IEEE Access*, vol. 9, pp. 39608–39617, 2021.

[6] C. Clemente, L. Pallotta, I. Proudler, A. De Maio, J. J. Soraghan, and A. Farina, "Pseudo-Zernike-based multi-pass automatic target recognition from multi-channel synthetic aperture radar," *IET Radar Sonar Navigation*, vol. 9, no. 4, pp. 457–466, 2015. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-rsn.2014.0296

[7] D. Gaglione, C. Clemente, L. Pallotta, I. Proudler, A. De Maio, and J. J. Soraghan, "Krogager decomposition and pseudo-Zernike moments for polarimetric distributed ATR," in *Proc. Sensor Signal Process. Defence*, 2014, pp. 1–5.

[8] C. Clemente, L. Pallotta, D. Gaglione, A. De Maio, and J. J. Soraghan, "Automatic target recognition of military vehicles with Krawtchouk moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 1, pp. 493–500, Feb. 2017.

[9] J. Yang, K. Zhou, Y. Li, and Z. Liu, "Generalized out-of-distribution detection: A survey," ArXiv, vol. abs/2110.11334, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:239049401

[10] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1563–1572.

[11] H.-M. Yang, X.-Y. Zhang, F. Yin, Q. Yang, and C.-L. Liu, "Convolutional prototype network for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 5, pp. 2358–2370, May 2022.

[12] M. Sabokrou, M. Khalooei, M. Fathy, and E. Adeli, "Adversarially learned one-class classifier for novelty detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3379–3388.

[13] S. Kong and D. Ramanan, "OpenGAN: Open-set recognition via open data generation," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, 2022.

[14] S. Pidhorskyi, R. Almohsen, D. A. Adjeroh, and G. Doretto, "Generative probabilistic novelty detection with adversarial autoencoders," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA: Curran Associates Inc., 2018, pp. 6823–6834.

[15] M. D. Scherreik and B. D. Rigling, "Open set recognition for automatic target classification with rejection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 2, pp. 632–642, Apr. 2016.

[16] M. Scherreik and B. Rigling, "Multi-class open set recognition for SAR imagery," in *Automatic Target Recognition XXVI*. Bellingham, WA, USA: SPIE, 2016. [Online]. Available: https://doi-org-s.libyc.nudt.edu.cn:443/10.1117/12.2224384

[17] T. Schlegl, P. Seeboeck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Proc. 25th Biennial Int. Conf. Inf. Process. Med. Imag.*, 2017, pp. 146–157.

[18] S. Plakias and Y. S. Boutalis, "Exploiting the generative adversarial framework for one-class multi-dimensional fault detection," *Neurocomputing*, vol. 332, pp. 396–405, 2019. [Online]. Available: https://www-sciencedirect-com-s.libyc.nudt.edu.cn:443/science/article/pii/S092523121831498X

[19] X. Ma, K. Ji, L. Zhang, S. Feng, B. Xiong, and G. Kuang, "An open set recognition method for SAR targets based on multitask learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4014005.

[20] X. Ma, K. Ji, S. Feng, L. Zhang, B. Xiong, and G. Kuang, "Open set recognition with incremental learning for SAR target classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5106114.

[21] S. Dang, Z. Cao, Z. Cui, Y. Pi, and N. Liu, "Open set incremental learning for automatic target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4445–4456, Jul. 2019.

[22] N. Inkawhich, J. Zhang, E. K. Davis, R. Luley, and Y. Chen, "Improving out-of-distribution detection by learning from the deployment environment," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2070–2086, 2022.

[23] K. Han, A. Vedaldi, and A. Zisserman, "Learning to discover novel visual categories via deep transfer clustering," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 8400–8408.

[24] Y.-C. Hsu, Z. Lv, and Z. Kira, "Learning to cluster in order to transfer across domains and tasks," in *Proc. Int. Conf. Learn. Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=ByRWCqvT-

[25] Y.-C. Hsu, Z. Lv, J. Schlosser, P. Odom, and Z. Kira, "Multi-class classification without multi-class labels," in *Proc. Int. Conf. Learn. Representations*, 2019. [Online]. Available: https://openreview.net/forum?id=SJzR2iRcK7

[26] G. Huang, H. Larochelle, and S. Lacoste-Julien, "Centroid networks for few-shot clustering and unsupervised few-shot classification," *CoRR*, vol. abs/1902.08605, 2019. [Online]. Available: http://arxiv.org/abs/1902.08605

[27] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," in *Proc. Int. Conf. Learn. Representations*, 2017. [Online]. Available: https://openreview.net/forum?id=BJ6oOfqge

[28] J. Zheng, W. Li, J. Hong, L. Petersson, and N. Barnes, "Towards open-set object detection and discovery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2022, pp. 3960–3969.

[29] L. Dai, W. Guo, Z. Zhang, and W. Yu, "Discovering novel categories in SAR images in open set conditions," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 1932–1935.

[30] T.-E. Lin and H. Xu, "Deep unknown intent detection with margin loss," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*. Florence, Italy: Assoc. Comput. Linguistics, 2019, pp. 5491–5496. [Online]. Available: https://aclanthology.org/P19-1548

[31] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1757–1772, Jul. 2013.

[32] W. J. Scheirer, L. P. Jain, and T. E. Boult, "Probability models for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2317–2324, Nov. 2014.

[33] L. P. Jain, W. J. Scheirer, and T. E. Boult, "Multi-class open set recognition using probability of inclusion," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 393–409.

[34] H. Zhang and V. M. Patel, "Sparse representation-based open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1690–1696, Aug. 2017.

[35] E. M. Rudd, L. P. Jain, W. J. Scheirer, and T. E. Boult, "The extreme value machine," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 762–768, Mar. 2018.

[36] P. Oza and V. M. Patel, "C2AE: Class conditioned auto-encoder for open-set recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2302–2311.

[37] C. Wang et al., "An entropy-awareness meta-learning method for SAR open-set ATR," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 4005105.

[38] P. Ji, T. Zhang, H. Li, M. Salzmann, and I. Reid, "Deep subspace clustering networks," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 23–32.

[39] L. Yang, N.-M. Cheung, J. Li, and J. Fang, "Deep clustering by Gaussian mixture variational autoencoders with graph embedding," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6439–6448.

[40] S. Mukherjee, H. Asnani, E. Lin, and S. Kannan, "ClusterGAN: Latent space clustering in generative adversarial networks," in *Proc. 33rd AAAI Conf. Artif. Intell./31st Innov. Appl. Artif. Intell. Conf./9th AAAI Symp. Educ. Adv. Artif. Intell.*, 2019, doi: 10.1609/aaai.v33i01.33014610.

[41] X. Li, H. Zhang, and R. Zhang, "Adaptive graph auto-encoder for general data clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9725–9732, Dec. 2022.

[42] G. Sheng, Q. Wang, C. Pei, and Q. Gao, "Contrastive deep embedded clustering," *Neurocomputing*, vol. 514, pp. 13–20, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231222012085
[43] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Proc. 15th Eur. Conf. Comput. Vis.*, Munich, Germany, 2018, pp. 139–156, doi: 10.1007/978-3-030-01264-9_9.
[44] S. Huang, K. Ota, M. Dong, and F. Li, "MultiSpectralNet: Spectral clustering using deep neural network for multi-view data," *IEEE Trans. Comput. Social Syst.*, vol. 6, no. 4, pp. 749–760, Aug. 2019.
[45] A. Bendale and T. Boult, "Towards open world recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1893–1902.
[46] Y. Zhou, P. Liu, and X. Qiu, "KNN-contrastive learning for out-of-domain intent classification," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, 2022, pp. 5129–5141. [Online]. Available: https://aclanthology.org/2022.acl-long.352
[47] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," *ACM SIGMOD Rec.*, vol. 29, pp. 93–104, 2000.
[48] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020.
[49] J. Macqueen, "Some methods for classification and analysis of multivariate observations," in *Proc. Symp. Math. Statist. Probability*, 1967, vol. 1, no. 14, pp. 281–297.
[50] K. Han, S.-A. Rebuffi, S. Ehrhardt, A. Vedaldi, and A. Zisserman, "Automatically discovering and learning new visual categories with ranking statistics," in *Proc. Int. Conf. Learn. Representations*, 2020. [Online]. Available: https://openreview.net/forum?id=BJl2_nVFPB
[51] A. Strehl and J. Ghosh, "Cluster ensembles—A knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, Mar. 2003, doi: 10.1162/153244303321897735.
[52] S. Wang, J. Yang, J. Yao, Y. Bai, and W. Zhu, "An overview of advanced deep graph node clustering," *IEEE Trans. Comput. Social Syst.*, to be published, doi: 10.1109/TCSS.2023.3242145.
[53] Y. Liu, W. Yang, W. Li, S. Zhang, W. Zhang, and Q. Shen, "Robust recognition method and device of ground target SAR image based on contrastive learning," U.S. Patent CN115187862A, 2022.

Mingyao Chen received the B.E. degree in information and communication engineering from the College of Information and Communication, National University of Defense Technology, China, in 2022. He is currently working toward the master's degree, majoring in information and communication engineering, with the College of Electronic Science, National University of Defense Technology.

His research interests include radar automatic target recognition, pattern recognition, and open set recognition.

Jing-Yuan Xia received the B.Sc. and M.Sc. degrees in electric and electronic engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2014 and 2016, respectively, and the Ph.D. degree in electric and electronic engineering from Imperial College London (ICL), London, U.K, in 2020.

He is currently an Associate Professor with the College of the Electronic Science, NUDT. His current research interests include low level image processing, nonconvex optimization, and machine learning for signal processing.

Tianpeng Liu received the B.E., M.E., and Ph.D. degrees in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2008, 2011, and 2016 respectively.

He is currently an Associate Professor with the College of Electronic Science and Technology, Chengdu, China. He has authored or coauthored numerous papers in respected journals, including IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS and International Conference on Signal Processing. His research interests include radar signal processing, electronic countermeasure, and cross-eye jamming.

Li Liu (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2012.

She is currently a Full Professor with NUDT. She has held visiting appointments with the University of Waterloo, Waterloo, ON, Canada, The Chinese University of Hong Kong, Hong Kong, and University of Oulu, Oulu, Finland, respectively. Her research interests include computer vision, pattern recognition, and machine learning. She was a Co-Chair of many International Workshops along with major venues, such as CVPR and ICCV. She was the Leading Guest Editor of the special issues for IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and *International Journal of Computer Vision*. She was also the Area Chair of several respected international conferences. She is currently an Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY AND PATTERN RECOGNITION. Her papers currently have more than 10 000 citations, according to Google Scholar.

Yongxiang Liu (Member, IEEE) received the B.S. and Ph.D. degrees in information and communication engineering from the College of Electronic Science, National University of Defense Technology, Changsha, China, in 1997 and 2004, respectively.

He is currently a Full Professor with the National University of Defense Technology. He has authored or coauthored numerous papers in respected journals, including IEEE TRANSACTIONS ON IMAGE PROCESSING and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. His research interests include radar imaging, SAR image interpretation, and artificial intelligence.