# Novel Category Discovery Without Forgetting for Automatic Target Recognition

Heqing Huang, Fei Gao, Jinping Sun, *Member, IEEE*, Jun Wang, Amir Hussain, and Huiyu Zhou

***Abstract*—In this article, we explore a cutting-edge concept known as class incremental learning (CIL) in novel category discovery for synthetic aperture radar (SAR) targets (CNTs). This innovative task involves the challenge of identifying categories within unlabeled datasets by utilizing a provided labeled dataset as reference. In contrast to the conventional category discover approaches, our method introduces novel categories without relying on old labeled classes and effectively mitigates the issue of catastrophic forgetting. Specifically, to reduce the bias of the established categories toward unknown ones, CNT extracts representational information via self-supervised learning, gleaned directly from the SAR data itself to facilitate generalization. To retain the model's competence in classifying previously acquired knowledge, we employ a dual strategy incorporating the rehearsal of base category feature prototypes and the application of knowledge distillation. Our methodology integrates multiview and pseudolabeling strategies. In addition, we introduce a novel approach that focuses on enhancing the discernibility of class spaces. This strategy primarily ensures distinct separation of the unlabeled classes from base class prototypes, and imposes stringent constraints on the internal relationships among individual samples and their corresponding perspectives. To the best of our knowledge, this is the first study on category discovery in the CIL scenario. The experimental results show that our method significantly improves the performance on SAR images compared to the previous optimal method, which indicates the effectiveness of our method.**

***Index Terms*—Automatic target recognition, class incremental learning (CIL), novel category discovery, synthetic aperture radar (SAR).**

## I. Introduction

IN RECENT years, deep learning [1] has made significant progress in synthetic aperture radar (SAR) automatic target recognition [2], [3], [4], [5], [6], [7], assisting human in Earth observation and military surveillance. Nevertheless, this task typically necessitates the availability of substantial annotated training data. In practical scenarios, obtaining annotations for each target category is only occasionally feasible. For instance, when movable military targets in SAR imagery undergo technological upgrades, such as the addition of turrets to tanks or the replacement of aircraft chassis and wing structures, these modifications often result in the emergence of new strategic target categories. However, annotating all these variants proves to be a costly endeavor. Hence, when a batch of new, unlabeled data from the same domain arrives, it becomes valuable to explore methods for utilizing existing data to facilitate the discovery of new classes.

Novel category discovery (NCD) [8] operates under the assumption that the model is initially provided with a set of labeled images, allowing it to facilitate the clustering of unlabeled images by sharing correspondence representation information. This task shares some similarities with zero-shot learning, as it demands the model to accurately predict classes that were not present during training. However, in zero-shot learning [9], the identification of these unknown categories necessitates prior knowledge of the additional semantic properties associated with the unlabeled classes, which the NCD task does not require. NCD has garnered considerable attention in recent years owing to its practicality and diverse applications across various real-world scenarios. As demonstrated by Han et al. [10], significant improvements have been observed when the labeled class data are jointly trained with labeled class images in the case of NCD. NCD task typically involves a two-phase training approach. In the first phase, pretraining is conducted using existing datasets, through metric learning. The second phase involves the utilization of clustering methods to fine tune the model on the unlabeled data. Notably, this fine-tuning process incorporates both labeled data from old classes and unlabeled data from new classes [see Fig. 1(a)]. Considering the sensitivity and privacy issues in SAR images, labeled data that are provided cannot always be readily used. Therefore, learning new categories without reusing the labeled data makes SAR target recognition more meaningful. Specifically, in the context of new category discovery, when we have a batch of labeled data and a batch of unlabeled data available, the first stage permits flexible usage of both types of data. Then, only the unlabeled data corresponding to the new categories is utilized, and ensure that the model learns the new categories without forgetting old knowledge [see Fig. 1(c)]. However, the training mode suffers from a severe catastrophic forgetting problem [11]. This problem will require rapid adjustment of model weights to accommodate new data, often resulting in the unintentional erasure of previously acquired knowledge. It

Manuscript received 8 November 2023; revised 5 January 2024; accepted 21 January 2024. Date of publication 25 January 2024; date of current version 12 February 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62371022, and in part by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/T021063/1 and Grant EP/T024917/1. *(Corresponding authors: Fei Gao; Jinping Sun.)*

Heqing Huang, Fei Gao, Jinping Sun, and Jun Wang are with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: huangheqing@buaa.edu.cn; feigao2000@163.com; sunjinping@buaa.edu.cn; wangj203@buaa.edu.cn).

Amir Hussain is with the Centre of AI and Robotics, Edinburgh Napier University, EH11 4BN Edinburgh, U.K. (e-mail: a.hussain@napier.ac.uk).

Huiyu Zhou is with the Department of Informatics, University of Leicester, LE1 7RH Leicester, U.K. (e-mail: hz143@leicester.ac.uk).

Digital Object Identifier 10.1109/JSTARS.2024.3358449

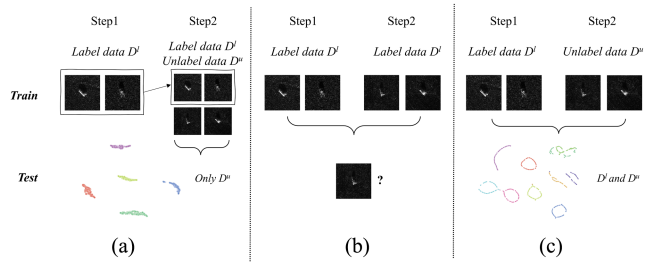



Fig. 1. (a) New category discovery, which identifies new classes through joint training, with a testing emphasis on the accuracy of these newly introduced classes. (b) CIL, staged training with testing covering new and previous classes. (c) CNT, a strategy avoiding base category data for discovering new classes, providing a comprehensive evaluation during testing across both the labeled and unlabeled classes.

is worth noting that the concept of not reusing previous data has been explored in incremental learning (IL) [12], [13]. During the training phase, the model is trained on the initial categories to acquire basic feature representation and classification capabilities. Subsequently, new categories are introduced to the model, and the model is adjusted to accommodate the new labeled data. In the testing phase, the model can be evaluated on data from both the old and new categories [see Fig. 1(b)]. A novel category discovered in incremental scenarios is grounded in the domain of relaxing certain assumptions underlying NCD, a task that proves to be more challenging than NCD/IL due to the unavailability of old class data and the labels of new classes in the incremental phase.

The majority of research in the domain involving old and new class recognition in SAR images has centered around class incremental learning (CIL). For instance, Dang et al. [14] propose a novel approach, which involves a class-boundary-based data reconstruction method. This method is designed to update the exemplars that stores old samples, thereby preserving the model's previous recognition capabilities. Li et al. [15] design new anchored class centers to aggregate the corresponding new class features. These methods improve future learning while attempting to retain what has been learned. Recently, there have been few studies that have examined the problems found in new categories of SAR images. Dai et al. [16] employ semisupervised learning and open-set detection techniques to estimate and identify unknown categories of data. Another study on the discovery of new classes of SAR images is [17], where they propose AutoMix to extend the neighborhood distribution and continuously motivate the discrete sample space. In addition to the majority of general NCD methods, there are studies that have delved into the discovery of optical image classes in IL scenarios. NCDwf [18] for instance, employs a mutually informative regularizer to enhance the unsupervised discovery of new classes using latent representations, aiming to mitigate forgetting. Another approach, ResTune [19] focuses on performing residual fine tuning on unlabeled data to preserve the model's ability to identify known classes while clustering new ones. However, it is crucial to highlight that applying these methods directly to SAR target discovery can be challenging due to differences in imaging methods and properties between optical and SAR images. Notably, the utilization of certain data enhancement strategies, such as color jittering and feature normalization, has a detrimental impact on the performance. Addressing the relatively underresearched problem of NCD tasks in SAR target recognition, we present a novel approach designed for SAR images in incremental scenarios. Our method diverges from the traditional NCD setup by excluding base class during the new category discovery phase, while focusing on the performance of all categories in the test. In the first stage, we work with a batch of labeled and unlabeled categories. The primary goal is to leverage the labeled categories to learn the unlabeled ones, ultimately forming a new set of clusters. We employ self-supervised learning (SSL) [20], [21], [22] to pretrain all available image representations, spanning both labeled and unlabeled datasets. Specifically, the model learns features like the texture structure inherent in the SAR image data, which, as highlighted by Han et al. [10], exhibit remarkable transfer capabilities. As new classes emerge, we anticipate that the model will have a minimal inductive bias toward previously known classes. In the second stage, we refrain from using data of the base class. Instead, we employ the prototype from the previous task's base class as exemplars to mitigate catastrophic forgetting. Sampling of features and labels from these prototypes is done through a defined normal distribution, followed by knowledge distillation conducted at the feature level. Finally, to ensure the performance of new category discovery, we use multiview and overclustering methods [23]. Overclustering encourages the categorization head to produce more refined feature regions, thereby enhancing the quality of unlabeled data representations. Building on this, we introduce a multiview consistency strategy and an enhanced category separability approach. The category separability loss is employed to increase the separation between prototypes and features from unlabeled classes, while consistency across multiple views is achieved by constraining covariance and mean square error (MSE) between samples. In conclusion, the contributions of this work are set out as follows.

1) We propose an efficient solution for the task of the SAR novel category discovery within class increment scenarios.
2) We create self-supervised learning tasks aimed at capturing the inherent information within SAR images. Furthermore, we leverage feature prototypes, which can be replayed as examples, to safeguard the model against catastrophic forgetting.
3) We introduce a multiview consistency and enhanced separability strategy to improve the performance of new category discovery. Our approach efficiently leverages the consistency of samples across multiple augmented views and emphasizes the distinctions between prototypical and novel class features.

## II. Related Work

### A. Self-Supervised Learning

Self-supervised learning [24], [25], [26] is a method for acquiring task-agnostic data representations. It involves devising self-supervised tasks specific to the domain and utilizing

abundant unlabeled data for model training, resulting in the extraction of high-level representations well-suited for subsequent tasks. Costa et al. [27] have seamlessly integrated most existing SSL method libraries. In this study, we adopted a contrastive-based SSL methodology [28] to facilitate model training via the comparison of semantically identical inputs. This method primarily involves distinguishing multiple views of a sample, generated through diverse data augmentations, thereby projecting them into the representation space. In the field of SAR image interpretation, Molini et al. [29] introduced a self-supervised Bayesian denoising approach, demonstrating the effectiveness of using only noisy SAR images for training. Wang et al. [30] introduced a pseudolabeled few-shot SAR image classification method, employing a dual network and cross-training strategy. Xu et al. [31] introduced adversarial self-supervised learning to SAR target recognition to maximize the similarity between SAR images enhanced by data and their adversarial examples. These studies collectively illustrate the feasibility of SSL techniques in the domain of SAR target recognition.

### B. Novel Category Discover (NCD)

Novel category discovery is a new research setup proposed by Han et al. [32]. Unlike weakly supervised learning and few-shot learning [33], [34], it involves transferring knowledge from a set of distinct but similarly labeled category data to facilitate learning and discovery on new, unlabeled data. There are two main types of existing NCD work. One type [35], [36] is learned by cross-task migration, first focusing only on labeled target datasets, and then, exploring new classes. The other kind [10], [37], [38] deals with labeled and unlabeled data simultaneously. NCL as introduced in [39], brings forth two enhanced contrast learning terms and formulates the creation of "hard negatives" from similar samples belonging to different categories. This approach significantly improves the learning process. UNO [23] introduced a unified cross-entropy loss function tailored for a novel class of discovery tasks. It is worth highlighting that this method is a joint unified new class discovery framework that combines joint training with labeled and unlabeled data. Our approach does not use labeled data in the new class discovery phase but guarantees the performance of the labeled classes as well. Another obvious difference is that our method exploits the features of SAR images themselves brought by unsupervised learning. At the same time, we propose separable strategies to enhance the distance between old and new classes in feature space. None of these aforementioned approaches consider the problem of forgetting old classes in incremental scenarios. There have also been studies of the incremental category discovery. Class-iNCD [40] trained joint classifiers for both basic and new categories and incorporated feature-level knowledge distillation to retain valuable information from prior knowledge. NCDwF [18] employed the generation of pseudopotential representations as a substitute for labeled data, whereas we utilized prototypes derived from labeled class data while preserving an old model for knowledge distillation. For the discovery of unlabeled classes, NCDwf employed a regularizer based on mutual information to enhance the unsupervised identification of new classes. In contrast, our approach differed by utilizing a multiperspective pseudolabeling strategy embedded within the network structure for the discovery of new classes. ResTune [19] was specifically designed for the category discovery performance, a setup distinct from our study, which took a broader focus on performance across all categories. In ResTune, distillation loss was employed to ensure the similarity of knowledge learned from both old and new categories, facilitating the migration of knowledge from the old model to the new one. Unlike them, our study employed distillation to ensure sustained performance in the old categories. Grow and Merge [41] introduced a multistage category increment and applied it to various incremental discovery settings. This method trained a dynamic network for NCD, where newly discovered categories were subsequently merged with previously known categories into a single model.

### C. Class IL (CIL)

CIL is a paradigm where a model is continuously updated to acquire knowledge about new classes while preserving its performance on the classes it has already learned. This approach can be categorized into three main methods: parameter regularization based, replay based, and network structure based. Parameter regularization-based methods, such as those described in [42] and [43], aim to prevent excessive bias in the learning of new categories by imposing constraints on crucial parameters within the model. Replay-based methods [12], [44] enable the model to retain a small portion of information from old classes. This information can be actual data [45], generated data [46], features [47], or other forms. The network-structure-based approach [48] dynamically extends the network structure during IL, demonstrating excellent performance. In recent years, there has been a surge in methods focusing on IL for SAR targets. Ma et al. [49] employed an autoencoder enhanced with multiscale structural similarity analysis to handle unknown classes of targets. Wang et al. [50] developed an adaptive few-shot CIL framework. This approach utilized hierarchical embedding networks and pseudoincremental strategies for effective model training.

## III. Method

The study's overarching framework is illustrated in Fig. 2. The first step is self-supervised learning, using both labeled and unlabeled categories to obtain implicit abstract features. The second step is supervised learning on the labeled dataset to provide the necessary guidance for clustering. The final step is new category learning, a phase that uses only unlabeled data to help discover new categories multiview consistency loss $L_{\text{consis}}$ and interclass separability loss $L_{\text{separa}}$. To mitigate forgetting and preserve previously acquired knowledge, the framework incorporates prototype replay and knowledge distillation techniques specifically for old categories. Together, these components demonstrate how our approach can effectively reduce the risk of forgetting while enabling new categories of discovery.

*Problem Statement:* In the context of the provided information, a labeled dataset $D^l$ is defined as a collection of data in the form of $\{(x_i^l, y_i^l)\}$, where $i$ ranges from 0 to $N$, and $N$

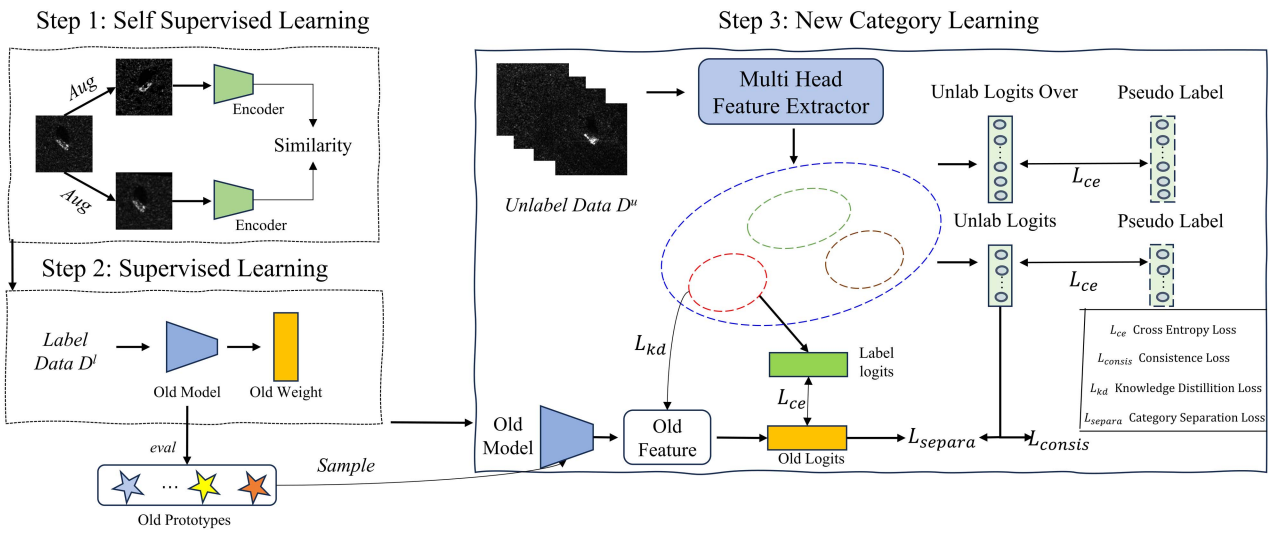



Fig. 2. Illustration of the proposed methodological framework. "Aug" for augmented data portfolio, "Encoder" represents the feature extractor in VICReg [28], and "Old Model" represents ResNet-18 [51] feature extractor.

represents the number of categories within this labeled dataset. In addition, there is an unlabeled dataset $D^u$ consisting of data in the form of $\{(x_i^u)\}$, where each data are associated with a category label $y^u$ that varies from 0 to $M$, with $M$ being the number of categories in the unlabeled dataset, $y^u$ is not provided in training. It is important to highlight that the labeled dataset $D^l$ and the unlabeled datasets $D^u$ do not share any common categories, meaning $\mathcal{D}^l \cap \mathcal{D}^u = \emptyset$. The primary objective of NCD is to merge the information from $D^l$ and $D^u$ to categorize $D^u$ into $M$ clusters. In this particular context of new category discovery without forgetting, special emphasis is placed on maintaining the accuracy of the labeled dataset $D^l$. This represents a notable departure from traditional NCD tasks where preserving the accuracy of existing categories is often not a primary concern.

### A. Self-Supervised Learning and Supervised Learning

The dataset used in this study comprises SAR images obtained from various orientation angles. In contrast to optical images, the slices within this dataset display significant uniformity concerning their location, color, and texture. As a result, the model needs to focus more on the specific details of the target within the SAR images. Due to the similarity and consistency in background characteristics across these SAR images, the model must emphasize the unique and distinct features of the actual target itself for accurate identification and classification. During the pretraining phase, when only supervised information is provided for the labeled dataset $D^l$, it can become challenging to achieve superior clustering results for new categories. Experiments have shown that the model has a strong bias toward the labeled data in $D^l$, which makes it challenging for the model to generalize effectively to new, unlabeled categories. This inductive bias introduced by the labeling of the data in $D^l$ can hinder the model's ability to adapt and learn to categorize new data points in the desired manner.

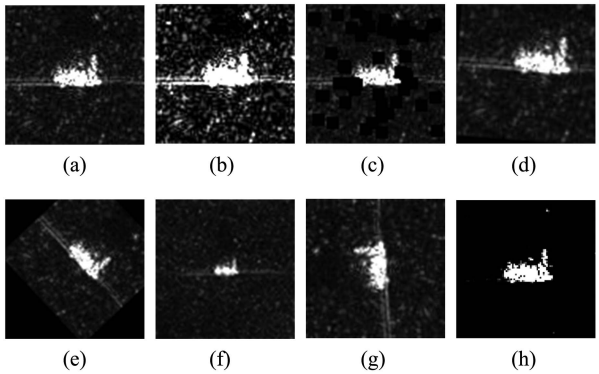



Fig. 3. Combination of data enhancements selected in the self-supervised learning task. (a) Original. (b) ColorJitter. (c) CoarseDropout. (d) ElasticTransform. (e) ShiftScaleRotate. (f) Resize. (g) Rotate. (h) Normalize.

Specifically, we use VICReg [28] as our self-supervised learning model and train it on the concurrent sets on $D^l$ and $D^u$. In this self-supervised learning task, the image batch is systematically generated by initially inputting a collection of images, which are subsequently subjected to a series of augmentations aimed at diversifying their perspectives. To enhance the richness of these image views, we incorporate eight distinct forms of data augmentation techniques using the albumentations library [52], implementing them with stochasticity based on predetermined human-defined probabilities, as illustrated in Fig. 3. These augmented image batches are subsequently channeled through an encoder, with the objective of deriving the corresponding embedding features. The overarching training loss, fundamental to our approach, can be succinctly formulated as follows:

$$l(Z, Z') = \lambda s(Z, Z') + \mu [v(Z) + v(Z')] + \nu [c(Z) + c(Z')] \quad (1)$$

where $\lambda$, $\mu$, and $\nu$ are the weighting factors. $Z$ and $Z'$ are the different viewpoints of the image after enhancement. $s$ is to compute the mean of the $L2$ paradigm between the vectors, $v$ computes the variance of the features, and $c$ is used to constrain

the expression of the mean of the diagonal elements of the matrix.

The utilization of supervised learning in new category discovery tasks ensures the availability of high-level semantic information, thereby enhancing the distinctiveness of the clustering outcomes. In our methodology, we employ the ResNet18 architecture [51] and fine tune the final layer of the network to execute the computation of cross-entropy loss on the dataset $D^l$ as

$$L_{ce} = -\frac{1}{N}\sum_{i=1}^{N}\log \eta_{y_i}^{l}\left(z_i^l\right) \tag{2}$$

where $z_i^l$ is the features of the input image and $\eta$ is the supervised classifier. Constraints through the labels help the new class to produce unique clustering results.

### B. Prototype Replay to Alleviate Catastrophic Forgetting

Once the pertinent representation necessary for clustering is acquired, the subsequent phase involves amalgamating the data from $D^u$ into a new set of categories. It is important to note that in this context, only data from $D^u$ are employed, representing an IL scenario. Notably, we do not impose any constraints by freezing any network structures, thus facilitating the seamless adaptation and learning of new classes. To sustain the performance of identifying old classes without relying on the historical class data, we draw inspiration from the work of iCaRL [12]. In our approach, we employ the concept of exemplars, which serve as repositories for retaining past knowledge, thus ensuring the preservation of the model's proficiency in recognizing categories that have been encountered previously. This mechanism allows us to maintain the performance of the previously seen categories even in the absence of their original data. Based on this, we extract features by performing feature extraction on the classes of $D^l$ and save prototypes of the features. Subsequently, when acquiring new classes, we employ a statistical approach by defining a normal distribution, utilizing the mean and variance parameters of the existing class prototypes. The formulation for this normal distribution can be described as follows:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(\mathrm{x}-\mu)^2}{2\sigma^2}\right) \tag{3}$$

where $\mu$ and $\sigma$ are the mean and variance of each type of prototypes. We randomly collect 20 samples from each category prototype as input to the network to get the prototype features, and compute the loss of the features to the sampled labels.

The phenomenon of catastrophic forgetting in models poses a substantial challenge, and relying solely on prototype playback may not be sufficient to mitigate this issue. Notably, when gradient backtracking occurs, the learning of new categories tends to dominate the model's behavior. To address this, we incorporate knowledge distillation as a means to enhance the performance on the previously learned classes. This approach involves a continual penalization of the model's output using logits derived from the initial model trained on the old classes. The knowledge distillation loss function is formally defined as follows:

$$\mathcal{L}_{kd} = \sum_{k=1}^{N}\pi_k^{\text{old}}(\mathrm{x})\log\pi_k^{\text{new}}(\mathrm{x}) \tag{4}$$

where $\pi_k^{\text{old}}$ are logits of the old copy mode, and $N$ is the head count of the network

$$\pi_k^{\text{old}} = \frac{e^{\mathbf{o}_k(\mathrm{x})/T}}{\sum_{l=1}^{N} e^{\mathbf{o}_l(\mathrm{x})/T}} \tag{5}$$

where $T$ can help solve the problem of too high a probability of being in the right category.

In network design, the strategy involves the use of multiclassification heads. The incorporation of a multiclassification head is pivotal, working in tandem with backpropagation techniques. This synergy serves to mitigate the abrupt impact of newly introduced class learning on the overall model. Furthermore, an essential aspect of our methodology involves the imposition of knowledge distillation on each of these output headers. This measure ensures that the wealth of knowledge previously acquired is effectively utilized to constrain the model's learning process.

### C. SAR Novel Category Discover

To facilitate the acquisition of knowledge about new categories, we adopt the strategy proposed by Fini et al. [23]. In our approach, we employ the ResNet18 feature extractor and create two distinct classification heads for handling labeled and unlabeled classes separately. Initially, a batch of images is input into the network, allowing it to extract feature vectors from the data. These feature vectors serve as the basis for subsequent classification tasks. To encourage the network to generate consistent predictions, we introduce the exchange of predictions from multiple perspectives or views. This approach compels the network to produce consistent predictions by leveraging diverse views of the same target to facilitate the exchange of pseudolabels. Notably, our approach diverges from prior methods in that we exclusively rely on unlabeled data. We do not utilize labeled data but, instead, focus our loss calculations on the logits generated from the unlabeled data.

Overclustering allows the model to better capture small differences in the data, improving its performance and generalization over new categories. We use overclustering to aid in the learning of new categories. The purpose of this overclustering header is to extend the output of the clustering neurons by a factor of $N$, where $N$ represents a scaling factor that amplifies the clustering output. In this study, $N = 2$. By amplifying the output of the clustering neurons through overclustering, our model becomes more sensitive to nuanced variations in the data. This increased sensitivity allows the model to better capture the finer details and intricacies present in the data, resulting in improved overall performance and a heightened capacity to generalize effectively when faced with new categories or data points. Both clustering and overclustering are constrained using the cross-entropy loss function, the pseudolabel is provided by the Sinkhorn-Knopp algorithm [53]. In our baseline approach, we incorporate a multiview strategy to generate pseudolabels

for images that have undergone different enhancement effects. This strategy allows us to leverage multiple views of the same image to generate pseudolabels for unlabeled data. However, the baseline approach lacks constraints on the consistency of classification results within these multiviews. To address this limitation and further enhance the model's performance, we introduce a consistency loss. The consistency loss plays a critical role in improving the compactness within a class and, in turn, enhances the model's performance in discovering new classes. This loss ensures that the classification results of different enhanced images, representing the same target within the unlabeled data, are consistent with each other. By imposing such consistency constraints, the model is encouraged to produce coherent and reliable predictions across the various views of the same image. For enforcing this consistency constraint, our study utilizes the MSE as the loss function. The MSE measures the squared difference between the predicted outputs for different enhanced views of the same image, thereby encouraging the network to produce consistent predictions within the multiview setting. This, in turn, promotes better compactness within classes, which is particularly beneficial for discovering and correctly classifying new categories within the data

$$\mathcal{L}_{\text{consis}} = \frac{1}{M}\sum_{i=1}^{M}(z_i^u - \overline{z}_i^u)^2 \tag{6}$$

where $z$ and $\overline{z}$ are the representations of different augmented views, respectively.

During the training phase, each batch exclusively comprises unlabeled images, devoid of any class annotations. The information of labeled class images is obtained through normal distribution prototype features. Generally, the inherent similarity between the distributions of these two components is attributed to their shared domain provenance. Indeed, in order to enhance the effectiveness of new category discovery, it is imperative that the representations of these categories within the feature space exhibit a more pronounced dissimilarity. The extent of dissimilarity between two probability distributions sharing the same feature space can be quantified using the Kullback–Leibler (KL) divergence distance, which is computed as follows:

$$D(P\|Q) = \sum_{x\in X} P(x)\log\frac{P(x)}{Q(x)}. \tag{7}$$

When two probability distributions exhibit similarity, their KL distance tends toward 0. Consequently, a logical approach is to maximize the value of the KL distance. To achieve this, we have formulated a separability-enhancing strategy explicitly aimed at expanding the distance between newly introduced categories and the reenacted feature prototypes. This strategy is designed to accentuate the divergence between the representations of new categories and the prototypes from previously features

$$\mathcal{D}_{\text{all}} = \frac{1}{2}\left(D\left(\boldsymbol{p}_i^l\|\boldsymbol{p}_j^u\right) + D\left(\boldsymbol{p}_j^u\|\boldsymbol{p}_i^l\right)\right). \tag{8}$$

Among them, $p_j^u$ represents the prototype probability distribution of the example set, and $p_l^i$ denotes the probability distribution derived from unlabeled samples. Let $N$ be the batch size.

TABLE I
SUMMARY OF THE MSTAR DATASETS

| Target Type | Train datasets | Test datasets |
|---|---|---|
| 2S1 | 299 | 274 |
| BMP2 | 233 | 195 |
| BRDM2 | 298 | 274 |
| BTR60 | 256 | 195 |
| BTR70 | 233 | 196 |
| ZIL131 | 299 | 274 |
| D7 | 299 | 274 |
| ZSU234 | 299 | 274 |
| T72 | 232 | 196 |
| T62 | 299 | 273 |

The category separation loss is defined as follows:

$$\mathcal{L}_{\text{separa}} = -\frac{1}{N}\sum_{i=1}^{N}\mathcal{D}_{\text{all}}. \tag{9}$$

The comprehensive loss function for the CNT is expressed as follows:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{ce}} + \mathcal{L}_{\text{kd}} + \mathcal{L}_{\text{separa}} + \mathcal{L}_{\text{consis}}. \tag{10}$$

## IV. EXPERIMENTAL RESULTS

### A. Experimental Setup

*1) Datasets Setup:* The datasets employed in this study have been sourced from the MSTAR database [54]. An example of an image in the dataset is shown in Fig. 4. The images contained within the MSTAR database are a result of high-azimuth resolution multibeam SAR acquisitions, utilizing the horizontal-to-horizontal (HH) polarization mode. The images were taken at two different pitch angles (17°, 15°), with the number of images in each class ranging from 190 to 300. We use the instances with a pitch angle of 17° in the database as the training set and the instances with a pitch angle of 15° as the test set. In the dataset comprising ten distinct categories, the labels were applied to five of these categories, while the remaining five categories were left unlabeled. A comprehensive depiction of the dataset's particulars and distribution is presented in the ensuing Table I. For the experiments on the MSTAR dataset, we utilize a controlled proportion of unlabeled data. Initially, the training dataset is constructed by dividing the categories into 50% old categories as the labeled dataset and 50% novel categories as the unlabeled dataset. Subsequently, we adjust the proportions to evaluate validity at different CNTs, specifically 40%, 60%, and 70%, 30%, respectively.

*2) Training Details:* All experiments in this study were conducted using an Ubuntu operating system environment running on a GeForce RTX 4090 graphics card. The experiments were carried out within a virtual environment configured with Python 3.8, Torch 1.9.0, and PyTorch-Lighting 1.8.0. The self-supervised model was trained on the solo-learn [27] framework. We configured the experiments with the batch size of 128 and the learning rate of 0.3. It is important to mention that Vicreg is not the exclusive method employed in this context, as there are other self-supervised techniques [24], [25] that can yield comparable

Fig. 4. (Top) Optical images and (bottom) their corresponding SAR images. These images are arranged in the following order: ZIL131, D7, ZSU234, BTR70, T72, BMP2, BRDM2, T62, and BTR60, 2S1. These images are commonly used in research and training applications related to remote sensing, radar, and target recognition.

results on the MSTAR dataset. In the supervised learning phase, the learning rate is initialized at 0.01 and gradually reduced to a minimum of 0.001. The learning rate optimization strategy employed is a combination of linear warmup and cosine annealing. The batch size for this phase is set to 12, and the self-supervised learning model is fine tuned over the course of 200 epochs. During the new class discovery phase, we utilize a learning rate of 0.01 with a larger batch size of 128. This phase is focused on the exploration and discovery of unlabeled classes within the dataset. In order to construct different perspectives to learn new classes, we use a combination of data enhancement using random flip, random crop, and random translate with reflect.

*3) Evaluation Metric:* To evaluate NCD methods in class incremental scenarios, it is imperative to evaluate the performance of metrics pertaining to the existing classes not only on the $D^u$ but also on the entire dataset. Agglomerative clustering accuracy (ACC) [55] is a primary method for evaluating NCD tasks. It quantifies the accuracy of the clustering process by measuring the percentage of samples that are correctly assigned to their respective clusters out of the total number of samples in the clustering result. In certain situations, the number of the clusters generated by a model may not correspond directly to the number of the actual categories, and as a result, there is a need to map the predicted labels to the actual labels in an optimal way. The Hungarian algorithm [56] can be used for this mapping task. It is employed to find an optimal assignment or mapping between the predicted clusters and actual categories. The formula for ACC is as follows:

$$\text{ACC} = \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}\left[y_i^u = \text{map}\left(\hat{y}_i^u\right)\right]. \tag{11}$$

Among them, $M$ is the total number of test samples, and map is the specific mapping method. In addition, we incorporate two additional settings, namely task aware and task agnostic, to evaluate the model. Task aware involves utilizing task-specific information to differentiate outputs unrelated to the current sample. For instance, outputs with labeled categories are exclusively considered for labeled classifiers, and for images with unlabeled categories, only the outputs of unlabeled classifiers are taken into account. Typically, in practical applications, task agnostic provides a more accurate reflection of the model's performance. In this evaluation setup, the model cannot differentiate the relationship between labeled and unlabeled categories but integrates them to generate the final output.

### B. Comparison to the State-of-the-Art

Other researches have not explored the implementation involving within the SAR novel category discovery task, specifically under conditions where the class incremental scenario. In this section, we commence by conducting novel class discovery using solely unlabeled data in the frameworks of the state-of-the-art NCD task. We set up three experimental conditions for comparison in the MSTAR dataset: 40% known categories, 60% novel categories (40%K / 60%N), 50% known categories, 50% novel categories (50%K / 50%N), 70% known categories, and 30% novel categories (70%K / 30%N). Subsequently, we apply these methods in the same setting on the dataset used in this study and compared it.

In this experiment, both UNO and AutoNovel were pretrained without utilizing any labeled data. The settings for FRoST and ResTune remained consistent with the configuration used in the study. Lwf [11] is an incremental learning algorithm designed to regulate model parameters. To ensure a fair comparison within the NCD setting, we align the new class identification method in the incremental phase with the approach used in AutoNovel. The Lwf + proto variant incorporates a rehearsal strategy in addition to this alignment. The experimental results can be found in the Table II. At the 50%K / 50%N setting, the AutoNovel recognizes old classes with an accuracy of 11.2%. In the specific implementation of AutoNovel, the self-supervised algorithm employed is Rotnet [25]. However, SAR images from different azimuthal variations that may not effectively recognize simple rotation transformations. The complexity of SAR image data, stemming from differences in acquisition conditions, can make it challenging for a model trained on rotation-based self-supervised tasks to generalize well to new classes. It is notable that UNO manages to achieve a clustering accuracy of 53.0% even after discarding the labeled samples. This indicates that UNO is effective at clustering unlabeled data. In this context, it might be beneficial to consider complementary modifications to UNO that enhance its ability to adapt to new classes. FRoST exhibits a degree of efficacy in alleviating the phenomenon of catastrophic forgetting, as evidenced by the achieved accuracy of 74.9% for the previously learned class. The utilization of similar feature comparison learning methodologies fails to effectively segregate dissimilar samples within the context of SAR images. The contrast in accuracy between our approach and ResTune is notable, with our method displaying a superiority of 8.8% in performance (92.2% compared to 83.4%). The observation reveals a substantial inclination in the model's predictions

TABLE II
COMPARATIVE EXPERIMENTS BETWEEN THIS STUDY AND OTHER NCD METHODS

| Method | 40%K / 60%N | | | 50%K / 50%N | | | 70%K / 30%N | | |
|---|---|---|---|---|---|---|---|---|---|
| | Old | New | All | Old | New | All | Old | New | All |
| LwF | 43.8 | 70.7 | 57.2 | 44.2 | 68.5 | 56.3 | 30.5 | 60.6 | 45.5 |
| LwF wo Proto | 53.7 | 73.2 | 63.4 | 48.4 | 61.9 | 59.1 | 32.9 | 50.4 | 41.6 |
| AutoNovel | 25.1 | 43.3 | 34.3 | 11.2 | 27.3 | 19.2 | 29.3 | 43.2 | 36.2 |
| UNO | 78.7 | 70.8 | 74.8 | 35.9 | 53.0 | 44.5 | 43.6 | 59.8 | 51.7 |
| FRoST | 68.0 | 73.7 | 70.8 | 74.9 | 47.3 | 61.1 | 54.1 | 60.9 | 57.5 |
| ResTune | 92.7 | 35.6 | 64.1 | 83.4 | 36.0 | 59.7 | 93.6 | 55.9 | 74.7 |
| CNT | **95.9** | **84.6** | **90.2** | **92.2** | **95.1** | **93.6** | **98.6** | **97.7** | **98.1** |

The best performance are in bold values.


Fig. 5. Results of UMAP visualization on MSTAR train sets. (a) FRoST. (b) ResTune. (c) CNT.

toward the old categories. We hypothesize that this bias may stem from the supervised learning framework, consequently impeding the model's capacity to effectively assimilate and learn new classes. Our method also achieves the best performance in the 40%K / 60%N and 70%K / 30%N settings. We show feature visualization [57] images of the two comparison methods and the CNT on the training dataset to visualize the performance of these comparison methods, as shown in Fig. 5. It can be seen that our method has better category differentiation.

### C. Ablation Study

*1) Effectiveness of Self-Supervised Learning:* Primarily, the analysis of self-supervised learning's significance merits attention. SAR images characterized by a uniform background and pervasive speckle noise. This distinct trait often results in the model's predisposition to encapsulate noise rather than fundamental target attributes. The generalization ability of self-supervised learning improves the model's focus on the target. Our aim is to evaluate the performance of the MSTAR dataset within the baseline framework. We endeavor to assess the influence of self-supervised learning on the discovery of new classes in SAR images. Our investigative approach adheres to the benchmark set forth by UNO, where both newly labeled and unlabeled classes are jointly considered. The results in Table III show that the accuracy of the old class was reduced by 1.4% when using self-supervised learning for pretraining, but the performance of discovering new classes improved by 14.5%. To provide a visual representation of this effect, Fig. 6 illustrates the combined training of both methods using labeled and unlabeled data, visualized through uniform manifold approximation and projection (UMAP) [57] visualization. The

TABLE III
EXPERIMENTS ON THE WITH OR WITHOUT OF SELF-SUPERVISED PRETRAINING

| Methods | Labeled | Unlabeled |
|---|---|---|
| Baseline w/o SSL | 95.7 | 70.3 |
| Baseline with SSL | 94.3 | 84.8 |
| CNT w/o SSL | 87.5 | 80.7 |
| CNT with SSL | 92.2 | 95.1 |

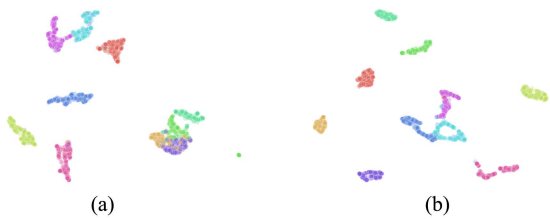



Fig. 6. Experiments with and without self-supervised learning, with visualization effects on all test data. In this experiment, labeled data are available when novel classes are discovered. (a) Baseline without self-supervised pretraining. (b) Baseline including self-supervised pretraining.

discernible enhancement in the model's capacity to identify new classes underscores the potent representational capabilities of self-supervised learning. Consequently, we have deemed it to integrate these techniques into our forthcoming research endeavors. The experiment in CNT also demonstrated the effectiveness of using self-supervised learning for pretraining. The experiment conducted within the context of CNT serves as additional evidence of the efficacy of self-supervised learning as a pretraining strategy. As indicated in Table III, the CNT model, when pretrained with self-supervised learning, achieved

TABLE IV
EFFECTS OF DIFFERENT COMPONENTS ON THE CNT

| Method | Prototype | KD | Consis | Separa | Task-aware | | | Task-agnostic | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Old | New | All | Old | New | All |
| CNT | ✗ | ✗ | ✗ | ✗ | 35.9 | 53.0 | 44.5 | 35.4 | 51.6 | 43.5 |
| | ✓ | ✗ | ✗ | ✗ | 78.7 | 70.8 | 74.8 | 75.3 | 66.7 | 71.0 |
| | ✓ | ✓ | ✗ | ✗ | 86.1 | 71.1 | 78.6 | 82.8 | 69.7 | 76.2 |
| | ✓ | ✓ | ✓ | ✗ | 97.8 | 85.7 | 91.8 | 94.2 | 84.6 | 89.4 |
| | ✓ | ✓ | ✓ | ✓ | 92.2 | **95.1** | **93.6** | 89.7 | **93.5** | **91.6** |

The best performance are in bold values.

95.1% ACC, with an 14.4% increase compared to the method that omitted the self-supervised pretraining step.

*2) Validity of Model Components:* We will delve into an analysis of the notable improvements achieved by the CNT model. Specifically, we observe old labeled class precision in the task-aware setting, transitioning from an initial 35.9% to 92.2%. Furthermore, the precision for new class discoveries rise from 53.0% to 95.1%. These compelling advancements underscore the effectiveness of the CNT model's capabilities in both labeled classes and novel categories. The results of the ablation test are shown in Table IV. In the task-agnostic setting, it can be seen that there is a 54.3% and 41.9% improvement in the accuracy of the old class and the new class compared to the original baseline. Next, we will analyze the experiments specifically under the task-aware setting.

At the very beginning, we learn new classes in an CIL scenario. The accuracy of the labeled classes decreases rapidly after running a few epochs. This can be visually confirmed by referencing the green curve in Fig. 7(a). Consistent with our expectation, the model suffers from catastrophic forgetting. Furthermore, the accuracy of recognizing the novel classes is also at a relatively low level. This phenomenon can be attributed to the fact that both old and new classes are subjected to a uniform loss objective during training. This uniform loss objective may not effectively differentiate between the requirements of old and new class recognition, which can impact the accuracy of both class types. The application of prototype replay to yield favorable results in the CIL scenario. By employing this technique, the accuracy of the model for recognizing the old class experiences a substantial increase, reaching 78.7%. This improvement in old class accuracy signifies the positive impact of prototype replay on the model's ability to better recall and distinguish the characteristics of classes.

Following the enhancement achieved with prototype replay, the new model's weight output for the old class is constrained using knowledge distillation loss. The distillation process does not appear to have detrimental effects on the model's capability to discover new classes. Instead, it results in a further increase in old class accuracy by 7.4%.

By incorporating a loss function focused on ensuring consistency across multiple views, this approach strengthens the relationships and compactness between different classes. As a result, the model's accuracy increase by 13.2% percentage points on the entire test set (78.6% versus 91.8%).

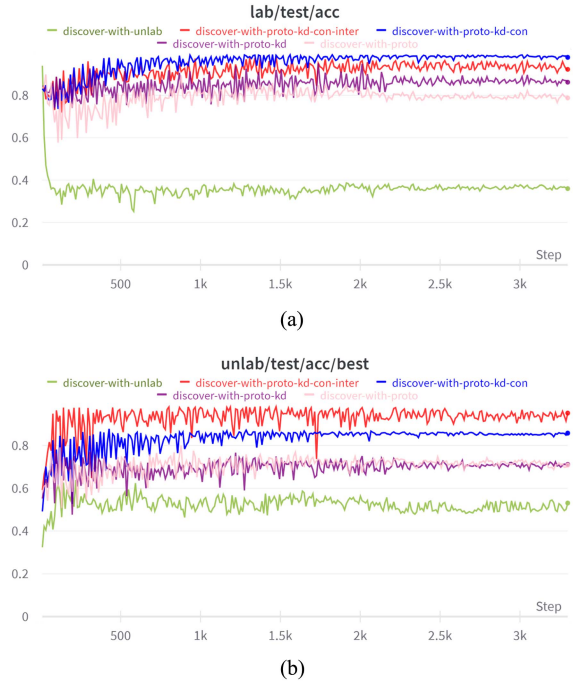



Fig. 7. Accuracy change curves for labeled and unlabeled data as training progresses. (a) Progression of accuracy for labeled data. (b) Trends for unlabeled data. The green curve is the result of the model without using labeled data for new class discovery. The pink and purple curves are with prototype playback and knowledge distillation loss added, respectively. The blue curve is the model with consistency loss added. Finally, the red curve encapsulates the ultimate iteration of the CNT model, consolidating all improvements and enhancements.

Following the enhanced separable strategy enhancement, the CNT model achieves accuracy of 92.2% for labeled classes and 95.1% for the unlabeled classes, underscoring its ability to successfully recognize and classify previously novel classes.

Finally, Fig. 8 illustrates the visual representation of the model across various versions using UMAP, providing an overview of the model's evolution and the impact of different strategies and enhancements on its performance. In Fig. 8(a), we present a visualization of the dataset distribution utilizing the ResNet18 pretrained model. Fig. 8(b) showcases the results obtained using the new class discovery method provided by UNO, without involving labeled classes at this stage. The feature space is obtained through unsupervised learning and pretraining weights. Fig. 8(c) and (d) depicts the results after introducing the old class prototype and distillation loss, respectively. It is evident

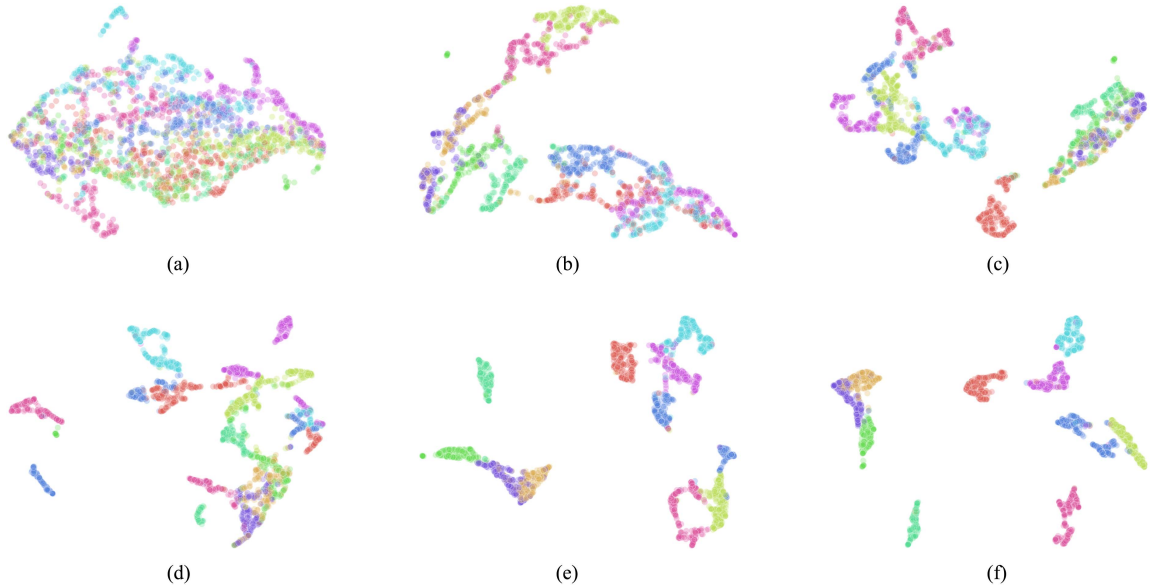



Fig. 8. Results of UMAP visualization on all test sets. (a) Initial feature space. (b) Baseline. (c) Baseline with prototype replay. (d) Baseline with prototype replay and knowledge distillation. (e) Add loss of consistency. (f) CNT.

TABLE V
SUMMARY OF THE FUSAR-SHIP DATASETS

| Target type | Train datasets | Test datasets |
|---|---|---|
| BulkCarrier | 85 | 37 |
| Fishing | 65 | 29 |
| Tanker | 32 | 14 |
| CargoShip | 110 | 48 |

TABLE VI
COMPARATIVE EXPERIMENTS BETWEEN THIS STUDY AND OTHER NCD METHODS

| Methods | Old | New | All |
|---|---|---|---|
| AutoNovel | 8.7 | 39.6 | 24.1 |
| FRoST | 81.2 | 47.3 | 64.2 |
| ResTune | 73.9 | 25.0 | 49.4 |
| CNT | **86.4** | **70.1** | **78.2** |

The best performance are in bold values.

TABLE VII
IMPLICATIONS OF DIFFERENT NETWORK ARCHITECTURES (IN 50%K / 50%N SCENARIOS)

| Methods | Old | New | All |
|---|---|---|---|
| ResNet50 | 90.5 | 93.7 | 92.1 |
| ResNet101 | 86.4 | 91.4 | 88.9 |
| DenseNet121 | 82.8 | 79.5 | 81.1 |
| ResNet18 | **92.2** | **95.1** | **93.6** |

The best performance are in bold values.

that some categories are distinguished, but a certain level of ambiguity persists in most category features. Fig. 8(e) shows that the incorporation of multiattempt consistency brings about performance improvement, representing a more compact feature space for the majority of categories. Finally, Fig. 8(f) introduces a separable strategy on top of the aforementioned approaches, and the visualized images demonstrate the effectiveness of this strategy.

### D. Discovery of Different Domain Categories

We also delve into cross-domain category discovery, examining the migration of MSTAR data to the FUSAR-ship dataset [58] to assess the impact of discover novel SAR objectives from diverse domains and under varying imaging conditions. The FUSAR-ship dataset comprises GF-3 satellite-acquired high-resolution AIS data for ship detection and identification. For our experiment, we selected four subclasses of targets (BulkCarrier, Fishing, Tanker, and CargoShip). Details regarding the specific training set and the number of test sets are presented in the Table V.

In this experiment, we compare our model with three state-of-the-art new category discovery methods. The experimental results are shown in Table VI. Specifically, our model attains a new category discovery accuracy of 70.1%, surpassing ResTune by 45.1%. In addition, our accuracy on the ship test set is higher than that of FRoST by 14.0%. This further validates the notable advantages of the method proposed in this article in terms of effectiveness and generalization ability.

### E. Experiments With Different Network Structures

To further examine the impact of the algorithm presented in this article, various network structures undergo analysis within the same experimental framework. Table VII displays the test outcomes of the ResNet18 model alongside other neural network models (ResNet50, ResNet101, and DenseNet121), revealing the superior clustering accuracy of ResNet18. The comparison results indicate that employing a deeper network structure does not yield better outcomes on the MSTAR dataset. This analysis suggests that an increased number of parameters might introduce redundancy within the model, potentially causing overfitting

and inhibiting the model's ability to accurately capture the true data distribution. Consequently, this could hinder the deep network's capacity for effective new category discovery due to poor generalization.

## V. Discussion

Our method demonstrates significant performance enhancements in SAR images compared to the state-of-the-art method. This indicates that our proposed CNT is more effective in addressing the challenge of discovering novel categories in incremental learning scenarios. We observe that the performance improvement primarily stems from the following factors: self-supervised learning enabling access to data for insightful representational information; ensuring a distinct separation between unlabeled categories and base category prototypes by enhancing the discriminate of the category space; and restricting internal relationships between individual samples and their corresponding perspectives. Although the CNT approach achieves commendable results in this task, there are still noteworthy limitations.

*1) Novel Category Discovery for Different Domains/Datasets:* In our current experiments, we noticed unsatisfactory results in the FUSAR-ship dataset when using MSTAR as the labeled category. Despite both belonging to SAR images, variations in acquisition methods, imaging techniques, and resolutions result in a limited generalization ability. In our future research, we intend to delve deeper into domain disparities among diverse datasets and explore better ways to transfer shared information between them.

*2) Unified Framework:* This study devised a multistage approach, however, addressing the resulting complexity requires future efforts toward creating simpler and more efficient unified frameworks with fewer loss functions and hyperparameters for optimal performance.

## VI. Conclusion

In this study, we focused on SAR image category discovery within incremental learning, aiming to identify new classes while retaining knowledge of existing ones in situations where not all old classes are available. We employed self-supervised learning for representational information, supervised learning for annotation uniqueness in labeled class samples, and a prototype replay method with knowledge distillation to preserve discriminative power in labeled classes. CNT incorporated multiview consistency loss and separability strategies to enhance recognition accuracy for both old and new classes. Our goal was to advance research on new class discovery in SAR images.

## References


[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] 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, 2022, Art. no. 5200817.

[3] Z. Yue et al., "A novel semi-supervised convolutional neural network method for synthetic aperture radar image recognition," *Cogn. Comput.*, vol. 13, pp. 795–806, 2021.

[4] Z. Huang, X. Yao, Y. Liu, C. O. Dumitru, M. Datcu, and J. Han, "Physically explainable CNN for SAR image classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 190, pp. 25–37, 2022.

[5] 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, doi: 10.1109/TGRS.2023.3264231.

[6] F. Gao et al., "SAR target incremental recognition based on features with strong separability," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–13, 2024, doi: 10.1109/TGRS.2024.3351636.

[7] F. Gao et al., "Cross-modality features fusion for synthetic aperture radar image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, 2023, doi: 10.1109/TGRS.2023.3307825.

[8] 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. 8401–8409.

[9] W. Wang, V. W. Zheng, H. Yu, and C. Miao, "A survey of zero-shot learning: Settings, methods, and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–37, 2019.

[10] K. Han, S.-A. Rebuffi, S. Ehrhardt, A. Vedaldi, and A. Zisserman, "AutoNovel: Automatically discovering and learning novel visual categories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6767–6781, Oct. 2022.

[11] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2935–2947, Dec. 2018.

[12] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "ICARL: Incremental classifier and representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2001–2010.

[13] D.-W. Zhou, Q.-W. Wang, Z.-H. Qi, H.-J. Ye, D.-C. Zhan, and Z. Liu, "Deep class-incremental learning: A survey," 2023, *arXiv:2302.03648*.

[14] S. Dang, Z. Cao, Z. Cui, Y. Pi, and N. Liu, "Class boundary exemplar selection based incremental learning for automatic target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5782–5792, Aug. 2020.

[15] B. Li, Z. Cui, Z. Cao, and J. Yang, "Incremental learning based on anchored class centers for SAR automatic target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, doi: 10.1109/TGRS.2022.3208346.

[16] 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.

[17] Q. Wei, Z. Cui, Y. Deng, B. Li, and Z. Cao, "Research on novel class discovery of SAR target," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 7479–7482.

[18] K. Joseph et al., "Novel class discovery without forgetting," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 570–586.

[19] Y. Liu and T. Tuytelaars, "Residual tuning: Toward novel category discovery without labels," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 7271–7285, Oct. 2022.

[20] X. Chen, H. Fan, R. Girshick, and K. He, "Improved baselines with momentum contrastive learning," 2020, *arXiv:2003.04297*.

[21] 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.

[22] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 16000–16009.

[23] E. Fini, E. Sangineto, S. Lathuiliere, Z. Zhong, M. Nabi, and E. Ricci, "A unified objective for novel class discovery," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 9284–9292.

[24] A. Ng et al., "Sparse autoencoder," *CS294 A Lecture Notes*, vol. 72, no. 2011, pp. 1–19, 2011.

[25] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," 2018, *arXiv:1803.07728*.

[26] 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.

[27] V. G. T. Da Costa, E. Fini, M. Nabi, N. Sebe, and E. Ricci, "Solo-learn: A library of self-supervised methods for visual representation learning," *J. Mach. Learn. Res.*, vol. 23, no. 1, pp. 2521–2526, 2022.

[28] A. Bardes, J. Ponce, and Y. LeCun, "VICReg: Variance-invariance-covariance regularization for self-supervised learning," 2021, *arXiv:2105.04906*.

[29] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, "Speckle2void: Deep self-supervised SAR despeckling with blind-spot convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022, doi: 10.1109/TGRS.2021.3065461.

[30] 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, doi: 10.1109/LGRS.2021.3069224.
[31] Y. Xu, H. Sun, J. Chen, L. Lei, K. Ji, and G. Kuang, "Adversarial self-supervised learning for robust SAR target recognition," *Remote Sens.*, vol. 13, no. 20, 2021, Art. no. 4158.
[32] 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.*, Oct. 2019, pp. 8401–8409.
[33] X. Xie, G. Cheng, X. Feng, X. Yao, X. Qian, and J. Han, "Attention erasing and instance sampling for weakly supervised object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–10, 2024, doi: 10.1109/TGRS.2023.3339956.
[34] L. Li, X. Yao, X. Wang, D. Hong, G. Cheng, and J. Han, "Robust few-shot aerial image object detection via unbiased proposals filtration," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–11, 2023, doi: 10.1109/TGRS.2023.3300071.
[35] Y.-C. Hsu, Z. Lv, J. Schlosser, P. Odom, and Z. Kira, "Multi-class classification without multi-class labels," 2019, *arXiv:1901.00544*.
[36] H. Chi et al., "Meta discovery: Learning to discover novel classes given very limited data," 2021, *arXiv:2102.04002*.
[37] Y. Qing, Y. Zeng, Q. Cao, and G.-B. Huang, "End-to-end novel visual categories learning via auxiliary self-supervision," *Neural Netw.*, vol. 139, pp. 24–32, 2021.
[38] Z. Zhong, L. Zhu, Z. Luo, S. Li, Y. Yang, and N. Sebe, "Openmix: Reviving known knowledge for discovering novel visual categories in an open world," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 9462–9470.
[39] Z. Zhong, E. Fini, S. Roy, Z. Luo, E. Ricci, and N. Sebe, "Neighborhood contrastive learning for novel class discovery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 10867–10875.
[40] S. Roy, M. Liu, Z. Zhong, N. Sebe, and E. Ricci, "Class-incremental novel class discovery," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 317–333.
[41] X. Zhang et al., "Grow and merge: A unified framework for continuous categories discovery," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 27455–27468, 2022.
[42] J. Kirkpatrick et al., "Overcoming catastrophic forgetting in neural networks," *Proc. Nat. Acad. Sci.*, vol. 114, no. 13, pp. 3521–3526, 2017.
[43] F. Zenke, B. Poole, and S. Ganguli, "Continual learning through synaptic intelligence," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 3987–3995.
[44] L. Wang et al., "Memory replay with data compression for continual learning," 2022, *arXiv:2202.06592*.
[45] S. Hou, X. Pan, C. C. Loy, Z. Wang, and D. Lin, "Learning a unified classifier incrementally via rebalancing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 831–839.
[46] H. Yin et al., "Dreaming to distill: Data-free knowledge transfer via deepinversion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8715–8724.
[47] A. Iscen, J. Zhang, S. Lazebnik, and C. Schmid, "Memory-efficient incremental learning through feature adaptation," in *Proc. 16th Eur. Conf. Comput. Vis.*, 2020, pp. 699–715.
[48] S. Yan, J. Xie, and X. He, "DER: Dynamically expandable representation for class incremental learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 3014–3023.
[49] 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.
[50] L. Wang, X. Yang, H. Tan, X. Bai, and F. Zhou, "Few-shot class-incremental SAR target recognition based on hierarchical embedding and incremental evolutionary network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5204111.
[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[52] A. Buslaev, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, 2020, Art. no. 125. [Online]. Available: https://www.mdpi.com/2078-2489/11/2/125
[53] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," *Adv. Neural Inf. Process. Syst.*, vol. 26, pp. 2292–2300, 2013.
[54] E. R. Keydel, S. W. Lee, and J. T. Moore, "Mstar extended operating conditions: A tutorial," *Algorithms Synthetic Aperture Radar Imagery III*, vol. 2757, pp. 228–242, 1996.
[55] Y. Yang, D. Xu, F. Nie, S. Yan, and Y. Zhuang, "Image clustering using local discriminant models and global integration," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2761–2773, Oct. 2010.
[56] H. W. Kuhn, "The hungarian method for the assignment problem," *Nav. Res. Logistics Quart.*, vol. 2, no. 1/2, pp. 83–97, 1955.
[57] L. Mcinnes and J. Healy, "UMAP: Uniform manifold approximation and projection," *J. Open Source Softw.*, vol. 3, no. 29, p. 861, 2018.
[58] 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.s*, vol. 63, pp. 1–19, 2020.

**Heqing Huang** received the B.S. degree in computer science and technology from Zhengzhou University, Zhengzhou, China, in 2019, and the M.S. degree in agricultural engineering and information technology from the South China Agricultural University, Guangzhou, China, in 2022. He is currently working toward the Ph.D. degree in signal and information processing with the Beijing University of Aeronautics and Astronautics, Beijing, China.

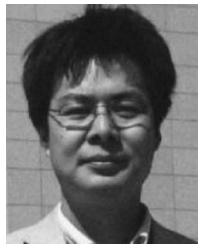
His research interests include radar signal processing, incremental learning, and image processing.

**Fei Gao** received the B.S. degree in industrial electrical automation, and the M.S. degree in electromagnetic measurement technology and instrument from the Xi'an Petroleum Institute, Xi'an, China, in 1996 and 1999, respectively, and the Ph.D. degree in signal and information processing from the Beihang University, Beijing, China, in 2005.

He is currently a Professor with the School of Electronic and Information Engineering, Beihang University. His research interests include target detection and recognition, image processing, and deep learning for applications in remote sensing.

**Jinping Sun** (Member, IEEE) received the M.Sc. degree in communication and electronic system and the Ph.D. degree in signal and information processing from Beihang University (BUAA), Beijing, China, in 1998 and 2001, respectively.

He is currently a Professor with the School of Electronic and Information Engineering, BUAA. His research interests include statistical signal processing, high-resolution radar signal processing, target tracking, image understanding, and robust beamforming.



**Jun Wang** received the B.S. degree in electronic engineering from the North Western Polytechnical University, Xi'an, China, in 1995, and the M.S. and Ph.D. degrees in information and communication engineering from the Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 1998 and 2001, respectively.

He is currently a Professor with the School of Electronic and Information Engineering, BUAA. His research interests include signal processing, digital signal processor (DSP)/field-programmable gate array (FPGA) real-time architecture, target recognition and tracking, and so on.

**Amir Hussain** received the B.Eng. and the Ph.D. degrees in electronic and electrical engineering from the University of Strathclyde, Scotland, U.K., in 1992 and 1997, respectively.

Following Postdoctoral and Academic positions with the University of West of Scotland (1996–98), Paisley, U.K.; University of Dundee (1998–2000), Dundee, U.K.; and University of Stirling (2000–18), Stirling, U.K., respectively. He is currently a Professor and Founding Head with the Cognitive Big Data and Cybersecurity (CogBiD) Research Lab, Edinburgh Napier University, Edinburgh, U.K. His research interests include cognitive computation, machine learning, and computer vision.

**Huiyu Zhou** received the B.Eng. degree in radio technology from the Huazhong University of Science and Technology, Wuhan, China, in 1990, the M.Sc. degree in biomedical engineering from the University of Dundee, Dundee, U.K., in 2002, and the Ph.D. degree in computer vision from HeriotWatt University, Edinburgh, U.K., in 2006.

He is currently a Full Professor with the School of Computing and Mathematical Sciences, University of Leicester, Leicester, U.K. He has authored or co-authored more than 380 peer-reviewed papers in the field.