An Incremental SAR Target Recognition Framework via Memory-Augmented Weight Alignment and Enhancement Discrimination

Heqing Huang, Fei Gao, Jun Wang, Amir Hussain, and Huiyu Zhou

***Abstract*—Synthetic Aperture Radar Automatic Target Recognition (SAR ATR) is one of the most important research directions in SAR image interpretation. While much existing research into SAR ATR has focused on deep learning technology, an equally important yet underexplored problem is its deployment in incremental learning scenarios. This letter proposes a new benchmark approach, termed Memory augmented weights alignment and Enhancement Discrimination Incremental Learning (MEDIL) algorithm to address this issue. Firstly, the attention mechanism is employed as part of the benchmark. Next, we discuss the problem of height deviation of weights at the fully connected layer and design a more suitable alignment of weights by guiding the memory module for contextual data processing. In addition, we leverage the incremental progressive sampling strategy to alleviate the imbalance between old and new classes during the training period. Finally, we propose to enhance the distinction among various classes with an angular penalty loss function to ensure the diversity of incremental instances. The proposed method is evaluated on MSTAR and OpenSARShip under different experimental settings. Experimental results demonstrate that our proposed approach can effectively solve catastrophic forgetting in SAR multiclass recognition problems.**

***Index Terms*—Synthetic aperture radar, automatic target recognition, incremental learning, catastrophic forgetting.**

I. Introduction

SYNTHETIC aperture radar (SAR) is an active earth observation system, usually installed on aircraft or satellites. Because of its remarkable ability to acquire high-resolution microwave images of surface targets in any climatic conditions, it has been widely used in urban planning [1] and image cognitive learnin [2]. Analyzing target characteristics in imaging is the focus of research on SAR image interpretation [3], [4]. Among them, automatic target recognition (ATR) of SAR plays a critical role.

Researchers have made significant progress in recent decades based on different SAR ATR methods [5], [9]. When new data constantly appears in the visual world, the common deep learning paradigm usually uses methods such as transfer learning to fine-tune the new data. The problem is that the approach tends to lose the ability to generalize to old data, termed catastrophic forgetting. Meanwhile, the presence of noise, doppler effect, and other interference signals in SAR images and the unstable data quality pose additional challenges for SAR target recognition and incremental recognition.

In the above context, incremental learning (IL) of SAR ATR is worth discussing [6]–[8]. IL reasons that a model can continuously learn new knowledge from new data and can remember old knowledge that has been previously learned. Recently, many incremental learning methods have been applied to SAR ATR work. Dang et al. [10] proposed a class boundary selection method based on local geometric and statistical information, and introduced a data reconstruction method to update the samples when new classes are added. Tang et al. [11] trained multiple optimal models on the old tasks to correct the cumulative errors and improve the efficiency of model update by pruning. Despite the ideal performance achieved by these methods, we consider that they have the following drawbacks. 1) When using exemplars to store data of old classes, the data of exemplars usually exhibit class imbalance or long-tail data distribution. 2) The feature embedding cannot be explicitly optimized to enhance the intra-class similarity and inter-class inconsistency using the traditional classification loss function.

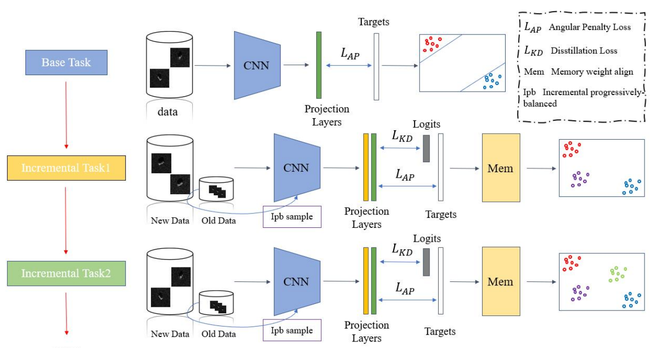

Fig. 1. The framework of our proposed methodology.

Due to the above issues, this letter is concerned with the class incremental learning model capability not only to make the learning incremental, but also to make the incremental model maintain the previous class recognition capability. The overall framework is shown in Fig. 1. The model fixes the number of sample sets for storing old classes in the incremental process and uses knowledge distillation to make the old model guide the current model to train the added classes.

Heqing Huang, Fei Gao 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; wangj203@buaa.edu.cn). (Corresponding author: Jun Wang.)

Amir Hussain is with the Cyber and Big Data Research Laboratory, Edinburgh Napier University, Edinburgh EH11 4BN Edinburgh, U.K.(e-mail: A.Hussain@napier.ac.uk).

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

In addition to using an attention mechanism module [12] to direct the output to correct predictions, the method also includes three other components. i.e., weight alignment [13] with memory addressing, incremental progressive sampling, and a discrimination enhancement component. The former can highlight the most representative representations in old category samples, and balance the weight bias of old and new categories. The latter maximizes old and new category separability to obtain highly discriminative features for incremental SAR recognition. Overall, our contributions are summarized as follows:

1) A memory enhancement module is proposed to extract typical representations from old category weights. The weight alignment using the learned representations balances the inductive bias of the old and new category weights.

2) An incremental progressive sampling strategy is designed to address the problem of long-tail data distribution that may occur in a fixed number of exemplars.

3) Finally, we propose a hybrid loss function that guarantees the plasticity-stability of incremental learning of SAR targets through knowledge distillation and angular penalties.

II. Proposed Method

A. Benchmark Incremental Learning with Exemplars Replay

The incremental learning task consists of an original task with base class training data and multiple incremental tasks with new classes. Assume an m-base n-incremental IL task, where m denotes the total number of classes of the old data, and n denotes the total number of classes of the new data. When there are n new data classes, the new model can successfully classify m old and n new classes. Let the base data as $X = \left\{X_{\text{train}}^{0}, X_{\text{train}}^{1}, \ldots, X_{\text{train}}^{m}\right\}$ with the labels $\{0, 1, \ldots, m\}$ for training data. We denote the new data as $X_{new} = \left\{X_{\text{train}}^{m+1}, X_{\text{train}}^{m+2}, \ldots, X_{\text{train}}^{m+n}\right\}$. The model first trains a classifier in the base data X. Then the model selects a fixed number of old samples in X as exemplars and combines them with X_{new} to train a new classifier. During the test, the model can effectively classify all classes seen so far.

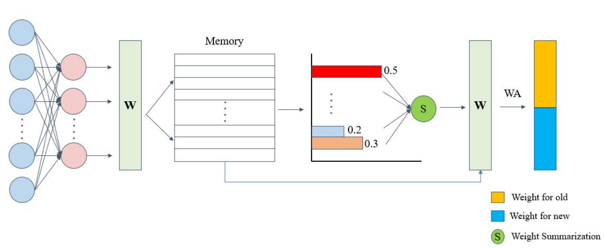

Fig. 2. The memory enhancement module.

Meanwhile, using attention mechanisms in incremental learning tasks can assist the model in better reducing forgetting old lessons and learning new ones. We add an attention module to the feature extract network, which models the tuning of channel relations to improve the network's robustness. The module is divided into two parts-compression and aggregation. Specifically, compression is the scaling of features to the spatial dimension. It represents the global correspondence in the re-channel dimension. The reason for the model's poor performance when a new task arrives is that the new category introduces a different loss optimization space. With the attention mechanism, it is possible to change the model's focus on additional features, allowing it to manage knowledge more flexibly in incremental learning. Specifically, a weight is assigned to each category sample based on its importance, and the more essential weights are then used to control the loss optimization space of the model during training.

B. Weight Alignment Based on Memory Enhancement

As new tasks are introduced, the model tends to classify objects into new classes. The model's parameters will adapt to new features, particularly in the fully connected layer. However, the simple alignment method [13] is insufficient for the SAR ATR incremental task. In addition, the essence of ensuring model stability is to retain old knowledge by not changing parameters as much as possible. Remembering the most representative sample of old classes can alleviate the catastrophic forgetting dilemma that occurs when the model learns old knowledge over a fixed number of exemplars

Based on the above, we present a memory enhancement module between the output of the model and the weighted alignment. As shown in Fig. 2, this module provides richer representational learning capabilities. For a given weight, the memory enhancement module does not encode it directly for weight alignment but instead retrieves the most relevant items in memory as a query. By using the encoded representation as a query, the memory module retrieves the most relevant old category weights in memory employing an attention-based addressing operator. These weights are then aggregated and passed to the alignment module. This helps the model treat the output of old and new classes more fairly. We first define X to represent the data sampling domain and $\widehat{Z}$ to represent the encoding domain of the feature extraction network. Taking X as input, the encoder transforms it into an encoded representation $z \in \widehat{Z}$, using z to retrieve the associated old class weights. Before the weight alignment, we convert $\widehat{Z}$ to an in-memory query. Specifically, memory is a matrix of constructed row vectors, each row vector m representing a memory element with the dimensionality of the model output features. The query $\widehat{Z}$ obtained through the memory network can be represented as:

$$\widehat{\mathbf{Z}} = \mathbf{w}\mathbf{M} = \sum_{i=1}^{N} w_i m_i \tag{1}$$

next, the features' attention coefficient vector w is calculated, where d is the cosine similarity.

$$w_i = \frac{\exp\left(d\left(\mathbf{Z}, m_i\right)\right)}{\sum_{j=1}^{N} \exp\left(d\left(\mathbf{Z}, m_j\right)\right)} \tag{2}$$

$$d\left(\mathbf{Z}, m_i\right) = \frac{\mathbf{Z} m_i^T}{\|\mathbf{Z}\| \, \|m_i\|} \tag{3}$$

We use this approach to remember the most representative sample prototypes better, picking the most similar task weights from the memory module for alignment. During the training phase, the memory module uses only a tiny number of addressed memory items to record typical old sample weights. Thus, memory supervision forces the memory to record the most representative prototype patterns in the old class data. When testing, both old and new class samples can index to the class with the highest probability.

C. Incremental Progressive Sampling

Most of the incremental learning tasks suffer from data long-tail distribution and class imbalance. In this study, the number of exemplars is fixed at 200, i.e., the sum of the samples stored in all old classes does not exceed 200. Typically, upon the arrival of the subsequent incremental lesson, a conspicuous disparity arises in the number of samples between the newly introduced classes and the pre-existing classes stored in the exemplars. This conspicuous difference results in a situation where the most recent data sets carry more weight in determining the loss value and therefore, tend to dominate the learning of the parameters.

To solve this problem, we design an incremental progressive sampling strategy. A combination of the instance-balanced and class-balanced sampling methods is performed. By giving each training sample a sampling weight, this sampling weight changes incrementally with the increase of epochs. As training proceeds, the sample sampling weights of the exemplars are incrementally increased, gradually transforming the data from the original instance distribution to a balanced distribution. Compared with random sampling, the incremental progressive sampling strategy ensures that the data distribution in each epoch matches the sample dataset of the old and new classes. The general sampling approach can be expressed in the following form:

$$p_j = \frac{n_j^q}{\sum_{i=1}^{C} n_i^q} \tag{4}$$

there are four parameters. p_j represents the sampling probability of the j-th class, n_j represents the total number of samples in the j-th class. C represents the total number of classes for all classes. q is a hyperparameter used to control the sampling mode. The strategy for progressive sampling is given by the following equation,

$$p_j^{\mathrm{PB}}(t) = \left(1 - \frac{t}{T}\right) p_j^{\mathrm{IB}} + \frac{t}{T} p_j^{\mathrm{CB}} \tag{5}$$

where t and T denote the current epoch and the total number of epochs, respectively. p_j^{IB} is the strategy obtained when q of Eq. (4) is taken as 1, specifically, an increase in the number of samples in a category, leads to a higher probability of sampling. p_j^{CB} is the strategy obtained when q of Eq. (4) is taken as 0. At this point, the sampling probabilities p_j of the samples in the category are all equal to $1/C$.

At the beginning of training, each training example has the same probability of being selected. For incremental learning, instance-balanced sampling causes the model parameters to be more inclined to learn X_{new} with an enormous amount of data. Learning old categories in the exemplars is insufficient to occupy a favorable position for updating the neural network parameters. Incremental progressive sampling is used to mitigate this discrepancy. As learning proceeds, ”interpolation” between the X_{new} and exemplars is continuously performed.

D. Design of hybrid loss function

A certain level of stability-plasticity is required in an ideal incremental learning environment. The model can overcome catastrophic forgetting in previous tasks and acquire knowledge in newly added tasks. The softmax activation function is often post-connected to the last layer of the neural network for the classification task. The benchmark uses softmax to convert all predicted values into predicted probability values. However, the new category requires a more expansive feature space in incremental scenarios. When new data arrives, the network will selectively catastrophically forget or not update parameters for new classes if the feature space is insufficient. The softmax loss function is shown as follows:

$$L_1 = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{y_i}}{\sum_{j=1}^{n} e^{y_i}} \tag{6}$$

where n is the number of classes, and N is the batch size. y_i represents the output of the full connection. Inspired by [14], the inner product/dot product of the feature vectors contains information about the angle of the inter-vector pinch. Therefore, we denote y_i as:

$$y_i = W_i^T f = \|W_i\| \, \|f\| \cos(\theta_i) = \cos(\theta_i), \tag{7}$$

where f is the feature of the input image and W is the weight parameter.

We employ the pinch angle information to add constraints to the loss function, aiming to compress similar data into a more compact space while widening the gap between different classes. First, we use cosine similarity as a distance metric to measure the similarity of the data and calculate the score. It measures the angular similarity of an image toward its class, which indicates the likelihood that the image belongs to the class. Usually, cosine similarity prediction is used with cross-entropy loss to separate features from dissimilarity by maximizing the probability of ground truth. The final loss function is:

$$Loss_{AP} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{s(\cos(y_i+m))}}{e^{s(\cos(y_i+m))} + \sum_{j=1, j\neq y_i}^{n} e^{s\cos\theta_j}} \tag{8}$$

where θ is the angle between the weight W and the feature x, m is an additive angular margin between the feature x and the target weight W. s is used to re-scale the normalized features. The proposed additive angle margin penalty enhances both intra-class compactness and inter-class variability.

III. Experiments

A. Implementation Details

Data sets: The data sets utilized in this letter are taken from the MSTAR [15] and two classes of OpenSARShip database [16]. A total of 10 types of military targets are collected in this database, which mainly includes trucks (ZIL131), bulldozers (D7), cannons (ZSU234), armored cars (BTR70, BTR60), tanks (T72, BMP2, T62), truck (BRDM2), cannon (2S1). OpenSARShip is a database that can be used for SAR image ship classification. We will select a portion of data from two categories(tanker, cargo) to validate the model, and the experiments will be conducted on a 12 class dataset jointly constructed by MSTAR and OpenSARShip.

TABLE I
TOP-1 ACCURACY (%)↑ FOR EACH INCREMENTAL SESSION

	0	1	2	3	4	5
iCaRL [17]	99.45	98.18	94.27	93.23	80.13	74.56
BiC [18]	99.64	97.46	88.36	88.42	82.35	85.27
WA [13]	99.45	98.55	89.14	82.19	73.13	77.49
DER [19]	99.45	97.13	97.39	92.34	90.41	85.19
PodNet [20]	99.64	98.35	93.99	91.05	87.11	82.13
Ours	99.96	98.81	97.73	96.83	93.28	91.51

TABLE II
HARMONIC ACCURACY (%)↑ FOR EACH INCREMENTAL SESSION

	0	1	2	3	4	5
DER	0	99.32	98.23	94.13	90.74	87.38
Ours	0	99.63	98.60	95.57	93.91	92.60

Training details: The benchmark in this letter is described in Section II. The experimental code is run on the Win10 operating system. Our approach is mainly built based on the PyTorch deep learning library. The training is optimized using an SGD optimizer with momentum. The number of training epochs per incremental stage is uniformly 50, the learning rate is 0.01, and the weight decay is 0.0005. Each session training uses the batch size of 32. The temperature T used to calculate the knowledge distillation loss is set to 2. The number of exemplars storing samples of old classes is 200. In the specific experiments, we developed two settings: 1-base 1-incremental/ 2-base 2-incremental.

Evaluation protocol: The average incremental precision is used as one of this letter's incremental learning evaluation metrics. This metric reflects the overall precision of the old and new classes identified. Although this method is standard incremental learning evaluation metrics, using accuracy alone as an evaluation method is not sufficient.

A model performing well on the base class but poorly on subsequent incremental sessions can still have good average accuracy. For example, the model has the precision of 100 on having 60 base classes and shows 0% accuracy in 2-incremental. However, the precision is still rated as 96.7%.

TABLE III
ABLATION STUDY UNDER THE 1-BASE 1-INCREMENTAL

	0	1	2	3	4	5	6	7	8	9	10	11
	Precision											
Variation 0	100.0	99.64	96.72	96.86	97.77	93.07	97.16	89.21	92.24	72.61	72.48	67.62
Variation 1	100.0	99.37	99.15	99.02	97.35	95.95	94.65	92.28	90.14	88.54	86.51	83.21
Variation 2	100.0	99.45	99.30	99.41	98.92	97.59	95.21	95.53	92.99	92.87	91.48	90.26
Variation 3	100.0	99.63	99.60	99.25	98.70	97.61	96.27	95.05	93.44	93.20	94.37	93.40
Ours	100.0	99.45	99.39	99.51	99.75	98.79	98.51	97.65	96.23	96.29	96.70	93.87
	Harmonic Accuracy											
Variation0	0	99.45	98.60	96.57	95.38	89.17	90.79	90.41	89.17	87.96	89.24	86.14
Variation1	0	99.39	98.35	97.89	96.79	88.93	93.56	90.23	75.41	87.96	81.29	80.17
Variation2	0	99.35	99.27	98.26	98.19	93.83	94.76	93.71	90.87	88.53	87.13	86.94
Variation3	0	99.45	99.33	99.63	99.35	97.55	95.73	94.36	93.79	93.20	94.11	89.29
Ours	0	99.63	99.63	99.75	99.70	97.61	96.27	95.05	93.44	94.25	96.27	93.40

Therefore, we use an additional evaluation method, harmonic accuracy, to assess whether or not the model learns the new task. If an incremental learning model has high average precision but low harmonic accuracy, it would indicate that the model's performance is mainly provided by the underlying classes and could be better for learning new classes. The formula for this method is shown as follows:

$$A_h = \frac{2 \times A_b \times A_i}{A_b + A_i} \tag{9}$$

where A_b is the average precision in the primary classes, and A_i is the average precision of the incremental session classes. The ideal incremental learning model must be balanced in precision and harmonic accuracy.

B. Comparison with the State-of-the-art Methods

We compared our method with several representative methods in the SAR dataset, including iCaRL [17], BiC [18], WA [13], DER [19] and PodNet [20]. According to Tables I and II, the highest average accuracy and harmonic accuracy were achieved by our method in experimental scenarios with the 2-base 2-incremental settings. For the scenario setup with 2-base 2-incremental, we obtain the average precision of 91.51% and the harmonic accuracy of 92.60%, which are 6.32% and 5.22% higher than DER's method, respectively. In the 5th incremental session, BiC's performance produced a significant difference. We believe that the linear fitting process generates substantial errors and does not apply to ATR incremental learning. The incremental version of WA is more stable but shows more precision reduction per task. In the last session, the precision of our method is 14.02% higher.

It is worth noting that the results of the iCaRL model that we used to perform experiments on the MSTAR dataset are somewhat different from [11]. We think this may be due to two reasons: different samples retaining the old categories and different order of the incremental categories. These reasons may lead to a different feature space learned by the model.

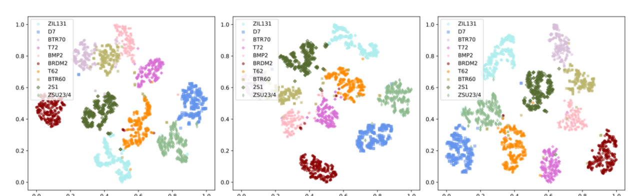

Fig. 3. Feature embedding visualization using t-SNE in the MSTAR dataset.

C. Ablation Study

To verify the effectiveness of each part of our method, we performed an ablation study on the SAR image dataset with the setting of 1-base 1-incremental. Table III shows each incremental class's accuracy and the ablation study's experimental results. **Variant 0**, benchmark training. **Variant 1**, training with the attention mechanism. **Variant 2**, training with a memory enhancement module and a weight alignment module via variant 1. **Variant 3**, training with incremental progressive sampling strategy via variant 2. **Ours**, training with the angle penalty loss based on variant 3.

When the attention mechanism was used in the benchmark test, the model showed varying degrees of accuracy improvement in both incremental learning classes. This reflects the training stability and model robustness resulting from the attention mechanism, which effectively ameliorates the catastrophic forgetting problem. However, when validating the harmonic accuracy, we observe that the training of the model was unstable. We argue the attention mechanism introduced additional attention weight parameters that have a slight effect on the network gradient for learning new classes, thus affecting the overall network optimization process. Further, the incremental progressive sampling strategy proposed in this letter ensures more stable training of the model to effectively solve this problem. The high precision training models use memory-based incremental weight alignment and show over-specialization. Their good precision performance is mainly attributed to the balance of old and new class weights. The overall performance at the end of class incremental learning is improved by more than 7% compared to the benchmark. We then design the incremental progressive sampling strategy to balance more training data to compensate for the loss of fewer samples of old classes in the training data. With the help of this method, in the last incremental session, the precision and harmonic accuracy improved by 3.74% and 2.35%. The model pre-agings different virtual prototypes with enhanced discrimination items, which reserves the embedding space for new classes by angle penalty loss. The precision (harmonic accuracy) achieved 93.87% (93.40%) in the last session. Ablations validate that forward-compatible training is helpful for MEDIL.

Fig.3 shows the feature embedding of the MSTAR dataset visualized using t-SNE. The different colors represent different categories. The models visualized from left to right are: the benchmark, the weight alignment module with memory enhancement added, and MEDIL. As can be seen from the figure, the feature embeddings of three categories, BTR60, T72 and BTR70, are mixed together in the benchmark. Following use the memory-enhanced weight alignment, the feature points of the same category are seen to be more compact. In MEDIL, the feature embeddings of different categories are clearly distinguishable, which proves that the angular penalty loss can enhance the generalization ability of features.

IV. Conclusion

This letter proposes MEDIL, a simple and effective approach for dealing with catastrophic forgetting in class incremental learning. As new classes continue to emerge, MEDIL is able to protect previous knowledge in the form of selecting and updating class weights. It also provides sufficient boundary distance to provide learning space for new classes. We demonstrate that modifying the model structure can help reduce the old class catastrophic forgetting problem in incremental learning. Meanwhile, the recognition performance loss in incremental data processing that mainly arises from changes in the training data distribution and weight parameters. In the future, we will focus more on going to data in the real world.

References

[1] 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 Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–17, 2022.

[2] Z. Yue, F. Gao, Q. Xiong, J. Wang, T. Huang, E. Yang, and H. Zhou, "A novel semi-supervised convolutional neural network method for synthetic aperture radar image recognition," *Cognitive Computation*, vol. 13, no. 4, pp. 795–806, 2021.

[3] Z. Lin, K. Ji, M. Kang, X. Leng, and H. Zou, "Deep convolutional highway unit network for sar target classification with limited labeled training data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 1091–1095, 2017.

[4] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target classification using the deep convolutional networks for sar images," *IEEE transactions on geoscience and remote sensing*, vol. 54, no. 8, pp. 4806–4817, 2016.

[5] 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 Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[6] Y. Zhou, S. Zhang, X. Sun, F. Ma, and F. Zhang, "Sar target incremental recognition based on hybrid loss function and class-bias correction," *Applied Sciences*, vol. 12, no. 3, p. 1279, 2022.

[7] 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 Transactions on Geoscience and Remote Sensing*, 2023.

[8] B. Li, Z. Cui, Z. Cao, and J. Yang, "Incremental learning based on anchored class centers for sar automatic target recognition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[9] F. Gao, Y. Huo, J. Sun, T. Yu, A. Hussain, and H. Zhou, "Ellipse encoding for arbitrary-oriented sar ship detection based on dynamic key points," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–28, 2022.

[10] S. Dang, Z. Cao, Z. Cui, Y. Pi, and N. Liu, "Class boundary exemplar selection based incremental learning for automatic target recognition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5782–5792, 2020.

[11] J. Tang, D. Xiang, F. Zhang, F. Ma, Y. Zhou, and H. Li, "Incremental sar automatic target recognition with error correction and high plasticity,"*IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 1327–1339, 2022.

[12] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[13] B. Zhao, X. Xiao, G. Gan, B. Zhang, and S.-T. Xia, "Maintaining discrimination and fairness in class incremental learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13 208–13 217.

[14] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 212–220.

[15] E. R. Keydel, S. W. Lee, and J. T. Moore, "Mstar extended operating conditions: A tutorial," *Algorithms for Synthetic Aperture Radar Imagery III*, vol. 2757, pp. 228–242, 1996.

[16] L. Huang, B. Liu, B. Li, W. Guo, W. Yu, Z. Zhang, and W. Yu, "Opensarship: A dataset dedicated to sentinel-1 ship interpretation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 1, pp. 195–208, 2017.

[17] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "icarl: Incremental classifier and representation learning," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, pp. 2001–2010.

[18] Y. Wu, Y. Chen, L. Wang, Y. Ye, Z. Liu, Y. Guo, and Y. Fu, "Large scale incremental learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 374–382.

[19] S. Yan, J. Xie, and X. He, "Der: Dynamically expandable representation for class incremental learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3014–3023.

[20] A. Douillard, M. Cord, C. Ollion, T. Robert, and E. Valle, "Podnet: Pooled outputs distillation for small-tasks incremental learning," in *European Conference on Computer Vision*. Springer, 2020, pp. 86–102.