remote sensing

MDPI

Communication

Improving Pre-Training and Fine-Tuning for Few-Shot SAR Automatic Target Recognition

Chao Zhang [1,†], Hongbin Dong [1,†] and Baosong Deng [2,*]

[1] College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China
[2] Defense Innovation Institute, Academy of Military Sciences, Beijing 100071, China
[*] Correspondence: dbs310@163.com
[†] These authors contributed equally to this work.

Abstract: SAR-ATR (synthetic aperture radar-automatic target recognition) is a hot topic in remote sensing. This work suggests a few-shot target recognition approach (FTL) based on the concept of transfer learning to accomplish accurate target recognition of SAR images in a few-shot scenario since the classic SAR ATR method has significant data reliance. At the same time, the strategy introduces a model distillation method to improve the model's performance further. This method is composed of three parts. First, the data engine, which uses the style conversion model and optical image data to generate image data similar to SAR style and realize cross-domain conversion, can effectively solve the problem of insufficient training data of the SAR image classification model. Second is model training, which uses SAR image data sets to pre-train the model. Here, we introduce the deep Brownian distance covariance (Deep BDC) pooling layer to optimize the image feature representation so that the model can learn the image representation by measuring the difference between the joint feature function of the embedded feature and the edge product. Third, model fine-tuning, which freezes the model structure, except the classifier, and fine-tunes it by using a small amount of novel data. The knowledge distillation approach is also introduced simultaneously to train the model repeatedly, sharpen the knowledge, and enhance model performance. According to experimental results on the MSTAR benchmark dataset, the proposed method is demonstrably better than the SOTA method in the few-shot SAR ATR issue. The recognition accuracy is about 80% in the case of 10-way 10-shot.

Keywords: synthetic aperture radar (SAR); few-shot learning; automatic target recognition (ATR); image style transfer

check for updates

Citation: Zhang, C.; Dong, H.; Deng, B. Improving Pre-Training and Fine-Tuning for Few-Shot SAR Automatic Target Recognition. *Remote Sens.* **2023**, *15*, 1709. https://doi.org/10.3390/rs15061709

Academic Editor: Andrzej Stateczny

Received: 23 January 2023
Revised: 18 March 2023
Accepted: 20 March 2023
Published: 22 March 2023

1. Introduction

1.1. Background

Synthetic aperture radar (SAR) is a sensor based on radar wave imaging. Because of its excellent imaging ability, it has been frequently used in the military field. As an application of SAR, the automatic target recognition (ATR) system is significant. It aims to identify the type of interested targets in SAR images. With the emergence of neural networks, researchers are more interested in integrating neural network models into ATR to improve the recognition ability of the system. In SAR image target recognition, because the neural network model usually needs a lot of labeled data for training, obtaining labeled SAR image data is usually challenging. Therefore, how to train a reliable SAR image target recognition model using the minimum amount of SAR image data is a critical issue in this field.

Two main methods to solve the above problems are the few-shot learning method and the data augmentation method. The method proposed in this paper belongs to the former. In order to solve the problem of insufficient training and learning in the training phase of the deep neural network due to the small amount of data in the SAR image target recognition task, a style transfer model is introduced to generate a large-scale SAR image base class data set. A good training model is obtained using the base class data set for training. The idea of transfer learning is introduced, and a small amount of SAR image target recognition data is used to fine-tune the model so that the model has the recognition ability of the SAR image target. Finally, the knowledge distillation method is introduced to train the model iteratively to improve its performance of the model further.

1.2. Related Work

Automatic target recognition. A SAR ATR system is designed to detect and locate targets encompassing the target of interest (ROI) with high precision and without human interaction. Numerous SAR ATR algorithms and systems have been suggested, which may typically be divided into three categories: classification methods based on templates [1], classification methods based on models [2], and classification methods based on deep learning [3]. In order to minimize the loss function, deep learning-based classification algorithms and deep neural networks learn and identify the best characteristics of the input during the training phase. The test phase extracts and classifies SAR image target characteristics from beginning to finish. ATR approaches based on deep learning have been an important area of study in the discipline. The approach suggested in this work is also based on research on methods for deep learning.

Style transfer. The goal of image style transfer is to move the source picture's content style across to the destination image's content style. Initially, techniques focused mostly on stroke rendering [4], picture analogy [5], and image filtering [6] were developed for artistic style transmission before neural network models for image style transfer were established. These techniques often involve trade-offs in terms of effectiveness, efficiency, and generality of style transmission and are unable to provide flawless outcomes. However, since gram loss [7] was established to evaluate the loss of style transfer, several style transfer methodologies were built on neural network models. Single style transfer models [8], multiple style transfer models [9], and generalized style transfer models [10] are the three categories into which image style transfer models may be divided.

Few-shot learning. Learning from a small number of labeled samples is the goal of few-shot learning. Research on few-shot learning techniques has gained popularity in recent years. Three broad categories may be used to categorize research on few-shot learning techniques: model based, metric based, and fine-tuning based. Model-based methodology: through the design of the model structure, aim to swiftly update network model parameters on a limited number of samples [11–13]. Metric-based approach: by recommending good metrics that calculate the sample distance between the query set and the support set, classification is carried out [14]. Method based on fine-tuning: given that it is difficult to directly adapt the common model to few-shot circumstances, the model is improved by adding a limited number of labeled unique samples after pre-training on a large-scale dataset [15].

Active learning. Active learning methods can generally be divided into two parts: a learning engine and a selection engine [16]. The learning engine is responsible for maintaining the benchmark classifier and using supervised learning algorithms to learn the set of labeled samples provided by the system, thereby improving the performance of the classifier. The selection engine is responsible for running a sample selection algorithm to select unmarked samples and submitting them to a human expert for marking. Then, add the labeled sample to the labeled sample set, and the learning and selection engines work alternately. After multiple cycles, the performance of the benchmark classifier gradually improves. When the preset conditions are met, the entire process ends [17]. The method proposed in this paper refers to active learning by selecting data sets similar to SAR image target data sets to pre-train the model so that the benchmark classifier performs well on new class data sets.

Knowledge distillation. Knowledge distillation is mostly used to shrink the size of models without sacrificing their performance in order to deploy them on mobile computing platforms. Its fundamental concept is to train a small student model using a large teacher model. Distillation techniques now in use can be categorized based on several task-oriented needs. We may broadly categorize the known distillation techniques into single-point [18] and multi-point [19] distillation based on the quantity of distillation points. Distillation can also be classified into homogeneous distillation [20] and heterogeneous distillation [21,22], with heterogeneous distillation being considered more difficult due to its enormous structural variances. This is based on the structural similarities between teachers and pupils [18]. Knowledge distillation can result in several problem contexts. KD also results in diverse issue contexts, including mutual distillation [23], self-distillation [24], knowledge fusion [25], and data-free distillation [26], in addition to the conventional teacher–student distillation.

1.3. Motivation

Compared with optical images, SAR images are more difficult to obtain and label. Therefore, there are few public high-quality large-scale data sets that can provide a complete mark of SAR targets. When the ATR system encounters a new target to be identified, it cannot quickly obtain a large number of labeled samples to train an effective model. Therefore, it is necessary to develop a few-shot learning method to achieve an accurate classification of the target. Aiming at the above challenges, the method proposed in this paper is mainly based on pre-training large-scale related datasets and then fine-tuning using a small amount of SAR target data to achieve high-performance recognition. Therefore, the performance of this method depends on three aspects: (1) What data can help training to obtain a high-performance pre-trained model? (2) What algorithm can better extract the common features of the same category of images and increase the distance between the features of different categories? (3) What strategies can fine-tune the model on novel datasets and achieve high accuracy? The method proposed in this paper is based on the above three problems and studies the SAR image target recognition task.

1.4. Contribution

The improved pre-training and fine-tuning for few-shot automatic target recognition method proposed in this study achieves reliable recognition on the SAR image target recognition dataset by fine-tuning the model using only a small amount of SAR image target data after training the model using the generated SAR-like image target recognition dataset as the base class dataset. The main contributions are as follows:

- A novel framework for few-shot SAR image target recognition is proposed.
- A style transfer model is introduced to style transfer the optical image target recognition dataset to generate a SAR-like image style target recognition dataset. This dataset is used as the base class dataset for the model, and the model is trained to successfully eliminate the cross-domain problem between the base class dataset and the novel class dataset.
- The introduction of the deep BDC pooling layer method effectively fits the characteristics of SAR image targets and increases the importance of target contour features, thus effectively improving the accuracy of target recognition.

2. Methodology

Figure 1 depicts the method's general structure as it is described in this paper. It is primarily made up of three stages: the data engine, the training stage, and the fine-tuning stage. This section will provide a summary of the knowledge distillation approach that was used, as well as discussing the fundamental operating idea behind each step in further depth.

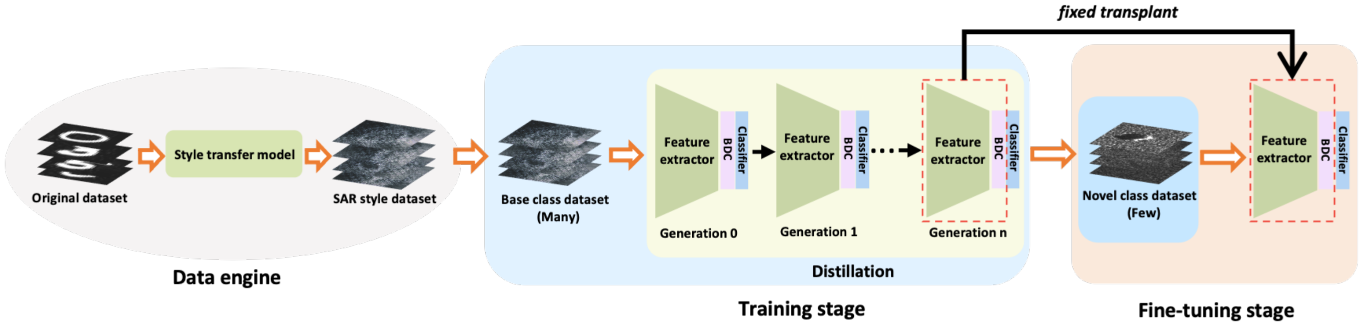

Figure 1. Framework of the proposed FTL.

2.1. Data Engine

Which data may aid training to produce a high-performing pre-trained model? The optimal approach to this issue is to pre-train the model using a large SAR image target identification data set. Based on the training and testing mechanism of deep learning models, it is known that the bigger the amount of training data utilized by the model and the greater the similarity between the training and testing datasets, the better the model's final testing results will be. Due to the difficulties in gathering SAR image target identification data, the existing SAR image target data set has a limited variety and volume. To tackle this issue, a data engine module is suggested in this study. The specific workflow of the data engine module consists of selecting an optical image dataset with image content similar to the novel SAR image dataset and employing the trained image style transfer model to migrate the style of this optical image dataset to SAR images to generate a SAR-like style image dataset. First, employing this dataset for model pre-training may successfully mitigate the cross-domain issue resulting from the inconsistency of styles between the pre-trained dataset and the novel dataset. Secondly, using a big volume dataset for pre-training may help the model to learn great network parameters and facilitate future fine-tuning on the novel dataset.

2.2. Deep Brownian Distance Covariance

What algorithm can better extract the common features of the same category of images and increase the distance between the features of different categories? For this problem, this paper introduces the deep Brownian distance covariance (DeepBDC) [27] theory to solve the problem, which was first proposed in 2009. It can represent the Euclidean distance between the joint feature functions of random variables and the product of their edge distributions (for a detailed definition of the BDC metric, refer to [27]).

In this framework, the computation of BDC is designed as a pooling layer, and the BDC matrix is obtained by convolving features after the pooling layer. The specific computation of the BDC layer is as follows:

1. The reduced-dimensional convolutional features are rearranged to produce the expression $\mathbf{X} \in \mathbb{R}^{hw \times d}$, where h and w are the spatial height and width, respectively, and d is the number of channels. Each column in the table $\chi_k \in \mathbb{R}^{hw}$ may be seen as an observation of the random vector X.
2. Calculate the Euclidean distance square matrix $\widetilde{\mathrm{A}} = (\tilde{a}_{kl})$, where $\tilde{a}_{kl}$ represents the squared Euclidean distance between the kth and lth columns of X, and then square it to produce the Euclidean distance matrix $\widehat{\mathrm{A}} = \left(\sqrt{a_{kl}}\right)$.
3. After removing the row and column means as well as the overall mean from the distance matrix, the BDC matrix may be derived.

For a set of convolutional features of the support image, the feature map at a specific location is considered a random variable x_1 and the feature map at the same location of the query image is y_1 so that the set of (x, y) can be obtained. The convolutional features are input into this pooling layer, and the BDC matrices A and B for support and query can be calculated separately using the formula. Then, based on the inner product, the similarity of the pair of images can be calculated. Then, in this framework, by inputting the output of the convolutional feature from the backbone directly into the BDC layer, the computed BDC matrix is sent to the tail classifier for training as the embedding features of the input image, and the cross-entropy loss function optimizes the network.

2.3. Sequential Self-Distillation

The process of moving the information that is incorporated in one model to another single model or from one bigger teacher model to one smaller student model is referred to as knowledge distillation. In the process of sequential self-distillation, we do not directly use the trained model for testing but extract the knowledge of the model into a new model with the same architecture and train on the same training data set. Here, the model of learning embedded knowledge is recorded as f_φ by minimizing the weighted sum of the cross entropy loss between the prediction tag and the real tag and the KL (Kullback–Leibler) divergence of the prediction tag distribution probability.

$$\varphi' = \arg\min_{\varphi'}\left(\alpha\mathcal{L}^{ce}\left(D^{train};\varphi'\right) + \beta KL\left(f\left(D^{train};\varphi'\right), f\left(D^{train};\varphi\right)\right)\right) \tag{1}$$

Here, $\mathcal{L}^{ce}$ represents the cross entropy loss function between the predicted value and the real value. β and α are equilibrium coefficients, and $\beta + \alpha = 1$, usually setting $\beta = \alpha = 0.5$. D^{train} refers to the data set used for training.

In the approach suggested in this research, the Born gain [28] strategy is used, and KD is employed to create numerous generations, as indicated in the knowledge distillation module of Figure 1. In each phase, the embedded model of the generation is trained using the transmitted information from the embedded model of the generation:

$$\varphi_k = \arg\min_{\varphi}\left(\alpha\mathcal{L}^{ce}\left(\mathcal{D}^{train};\varphi\right) + \beta KL\left(f\left(\mathcal{D}^{train};\varphi\right), f\left(\mathcal{D}^{train};\varphi_{k-1}\right)\right)\right) \tag{2}$$

Assuming K iterations, the embedded model used is expressed as φ_K to extract features for testing.

2.4. Baseline

What strategies can fine-tune the model on novel datasets and achieve high accuracy? A baseline is given here. The process of this baseline is the same as that of most migration learning, which is mainly divided into two processes: pre-training and fine-tuning. Its overall process is shown in Figure 1.

Training stage. The cross-entropy classification loss function is minimized to optimize the overall model, and a pre-trained dataset created by the data engine module is used to train a classification model. The feature extractor of this model comprises a backbone and a BDC layer. A backbone may incorporate Resnet12, Resnet18, and other similar networks, but this is not a limitation of the backbone (which will be explained in the next section). The classifier is made up of two components: a linear layer and a softmax function.

Fine-tuning stage. Freeze the network parameters of feature extractor after the training stage. In order to provide the model with the ability to distinguish novel classes, the novel class data (support set) with labels are used to only retrain the classifier. What needs to be explained here is that the data used in the test (query set) are also novel class data, and the support set and query set describe the same targets but do not contain each other.

3. Experiments

In this section, we first describe the dataset that was used in the experiment before performing ablation tests on each module of the suggested FTL approach to confirm its effectiveness. Simultaneously, the stage outcomes are visualized. Then, replace the backbone with other network architectures to determine the impact of network architecture on FTL performance. Ultimately, the comparison between FTL and SOTA demonstrates that the proposed approach is better.

3.1. Experimental Dataset

The MSTAR dataset is a benchmark dataset that is used for evaluating the performance of SAR ATR. SAR picture slices of 0.3 m × 0.3 m targets are included in this collection. There are 10 distinct categories of ground targets included in this dataset. Figure 2 displays the optical pictures and the matching SAR images for the ten different classes of objects that were included in the MSTAR dataset. The EMINIST dataset is an expanded version of the MINIST dataset. It takes the original 10 classes of handwritten numbers and adds 26 classes for lowercase letters and 26 classes for uppercase letters, bringing the total number of classes to 62. The total number of pictures in the dataset is 814,255.

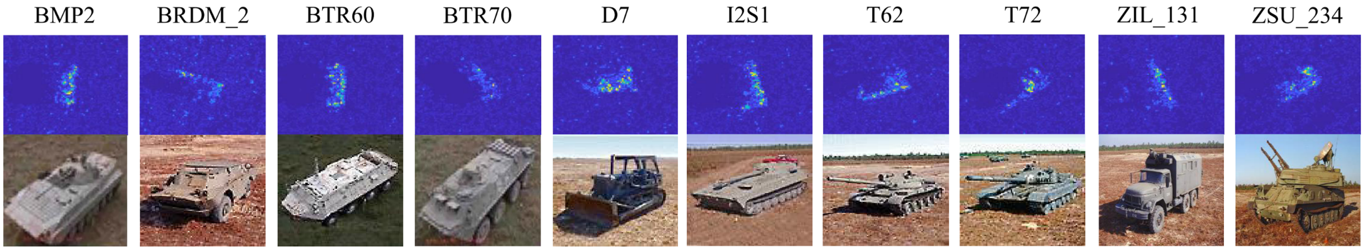

Figure 2. Optical and SAR pictures of ten target types from the MSTAR collection. The image's relevant targets are BMP2, BRDM 2, BTR60, BTR70, D7, I2S1, T62, T72, ZIL 131, and ZSU234.

3.2. Experimental Setup

The deployment software environment for the experiment was implemented in Ubuntu 20.04 using python 3.7 and python 1.7. The hardware environment is mainly supported by GPU: A100 for model training. The backbone of the proposed method is tested with ResNet series. The calculation of recognition accuracy in the experimental results is based on the setting of the number of few-shot, and the average value and standard deviation of 2000 tests are calculated.

3.3. Style Transfer Visualization

Considering the characteristics of the SAR image, the target does not have color and texture features, and the target area's main target feature for SAR image recognition is the target area's contour feature. Furthermore, the SAR image target background area is noisy, and the recognition model needs to be suitable for target recognition under such a background. Because of the above two points, we selected the EMINIST dataset in this experiment and used the ArtFlow style transfer model [9] to convert it to a SAR-like style and used it as a pre-training dataset. The reasons for choosing the EMINIST dataset are as follows: Firstly, the recognition of handwritten characters also relies on the contour features of the target region, and the handwritten character image also does not have texture and color features. Secondly, the amount of EMINIST data is large and the dataset contains many classes. As a pre-training dataset, it can help the model learn the initialization parameters with superior performance in the pre-training process. Last but not least, the EMINIST dataset after style migration is very similar to the SAR image dataset. As shown in Figure 3, the handwritten digits 0, 2, 3 and 6, as well as the handwritten letters t, g, k, s and R after style transfer are selected for display. The original handwriting 0, 2, 3, 6, t, g, k, s and R can be found to have accurate contour and clean background. After the style migration of MSATR data, the contour of the characters becomes blurred, but the contour of the image target area can still be identified, and the background generates a lot of clutter. The generated image is very similar to the SAR image as a whole. The visible EMINIST dataset after style migration effectively reduces the domain difference with the MSTAR dataset.

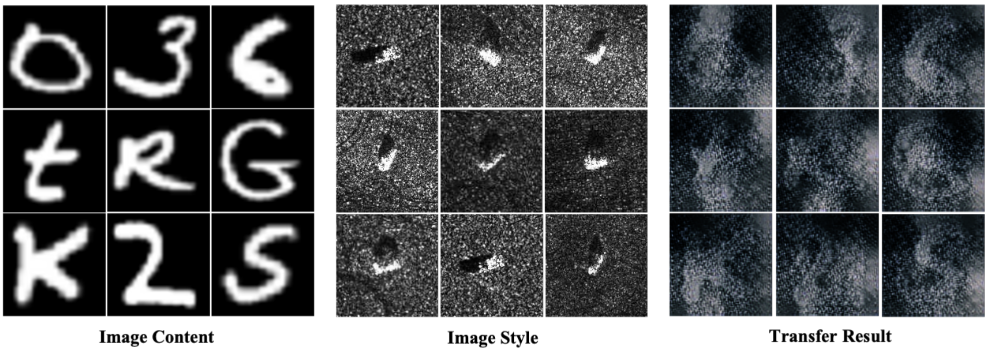

Figure 3. A partial display of the results of the style transfer from the MSTAR dataset to the EMINIST dataset.

3.4. Ablation Study

To determine the efficacy of each module, we eliminated each module and conducted ablation contrast studies. In order to determine the efficacy of each module, we used ResNet18 as the backbone network, eliminated each module and conducted a comparative study of ablation. As stated in Table 1, the following combinations apply to each module: (1) Baseline; (2) BDC Layer+Baseline; (3) Data Engine+Baseline; and (4) BDC Layer+Data Engine+Baseline. Comparing the experimental outcomes of Case 1 and Case 2 reveals that when just the BDC layer is applied to the baseline without the data engine module, the EMINIST dataset is utilized instead of the style transformation to pre-train the baseline with the BDC layer. We can see a considerable improvement in 1-shot, 5-shot, and 10-shot compared to the first testing findings, indicating that the features become more obvious after the BDC layer. By comparing the experimental results of Cases 1 and 3, it can be determined that the comparison of the experimental results when only adding the data engine to the baseline and when only adding the baseline reveals significant improvements in each case, proving the efficacy of the data engine and the cross-domain problem of the model resulting from the inconsistent image styles of the base dataset and the novel dataset. Simultaneously, the data engine may efficiently minimize the domain spread of the base dataset and the novel dataset in order to enhance the model's performance on the novel dataset. Observing the experimental results of Case 4 reveals that the model's performance is maximized when the data engine and BDC layer are included, demonstrating the compatibility of each module.

Table 1. Results of ablation experiments.

Case	Data Engine	BDC Layer	Baseline	10-Way 1-Shot	10-Way 5-Shot	10-Way 10-Shot
1	✗	✗	✓	20.13 ± 0.75	27.72 ± 0.76	31.62 ± 0.74
2	✗	✓	✓	32.57 ± 0.32	52.24 ± 0.37	56.04 ± 0.31
3	✓	✗	✓	28.90 ± 0.79	42.47 ± 0.69	49.65 ± 0.72
4	✓	✓	✓	**51.62 ± 0.40**	**59.47 ± 0.40**	**77.92 ± 0.32**

3.5. Comparative Experiment of Replacing Different Backbones

Considering that the framework proposed in this paper is a few-shot learning method based on fine-tuning strategy, the selection of a backbone network is very important, so we set up the following experiment. Replace the network of different depths to compare the experimental results. ResNet series networks with different depths are selected here: ResNet10, ResNet18, ResNet34, ResNet50, and ResNet101. From the experimental results in Table 2, it can be seen that the deeper the network depth is in the cases of 1-shot, 2-shot, 5-shot, and 10-shot, the overall performance of the fine-tuned model shows a downward trend, especially in the case of 10-shot. This is contrary to the traditional idea that the deeper the network structure, the better the performance of the model. From this, we can conclude that the few-shot learning method based on fine-tuning is not that the more complex the backbone network is, the better its performance on the novel class dataset after fine-tuning. Generally, the deeper the model of the neural network, the more rigid the model is after pre-training and it is difficult to migrate to novel class data and obtain excellent results on the novel class dataset. Therefore, the selection of a backbone network should not only consider the performance of the model itself but also refrain from selecting too deep of a network.

Table 2. Experimental results of replacing backbone.

Backbone	10-Way 1-Shot (%)	10-Way 2-Shot (%)	10-Way 5-Shot (%)	10-Way 10-Shot (%)
ResNet10	49.93 ± 0.40	58.64 ± 0.39	70.53 ± 0.37	79.46 ± 0.32
ResNet18	51.62 ± 0.40	59.47 ± 0.40	69.38 ± 0.37	77.92 ± 0.32
ResNet34	43.56 ± 0.37	51.32 ± 0.36	61.34 ± 0.35	69.82 ± 0.34
ResNet50	47.14 ± 0.38	55.22 ± 0.38	66.34 ± 0.36	74.72 ± 0.33
ResNet101	45.41 ± 0.37	52.59 ± 0.37	62.19 ± 0.36	69.26 ± 0.34

3.6. Comparative Experiment

As shown in Table 3, we compared the approach provided in this section with other few-shot learning methods described in recent years in the 1-shot, 2-shot, 5-shot, and 10-shot scenarios using the MSTAR dataset in order to further confirm the proposed method's progressivity. DeepEMD [29]: a few-shot learning method that introduces EMD by splitting the image into multiple blocks. DeepEMD grid [30] and DeepEMD sample [30] are improved DeepEMD methods by adding grid selection and random sampling strategies, respectively, and it can be found that the two improved methods for DeepEMD are not applicable to the few-shot target recognition task of SAR images. The performance is reduced compared with the original method. DN4 [31] is also a few-shot learning method based on metric learning, which extracts multiple local descriptive features from an image and classifies the image based on the similarity of multiple local descriptive features between metric images. Prototypical [32] and Relation [33] are both meta-learning methods, which learn a feature vector representation for each class through different network structures and calculate the distance between the feature representation of the image to be classified and the features of each class to perform classification. Comparing all the methods mentioned above, we can find that the proposed method (FTL) achieves the highest performance in all scenarios (this result is based on the experimental results of ResNet10 for the backbone network), especially when the fine-tuned sample size is increased to 5-shot and 10-shot, with a nearly 10% improvement over the second ranked performance. Meanwhile, the performance of the FTL-dis model is improved by nearly 2% in all scenarios when the sequential distillation method is introduced. In general, the method proposed in this paper achieved very strong competitive performance in the few-shot target recognition task of SAR images.

Table 3. Comparative analysis of other few-shot learning methods.

Method	10-Way 1-Shot (%)	10-Way 2-Shot (%)	10-Way 5-Shot (%)	10-Way 10-Shot (%)
DeepEMD [29]	36.19 ± 0.46	43.49 ± 0.44	53.14 ± 0.40	59.64 ± 0.39
DeepEMD [30] grid	35.89 ± 0.43	41.15 ± 0.41	52.24 ± 0.37	56.04 ± 0.31
DeepEMD sample [30]	35.47 ± 0.44	42.39 ± 0.42	50.34 ± 0.39	52.36 ± 0.28
DN4 [31]	33.25 ± 0.49	44.15 ± 0.45	53.48 ± 0.41	64.88 ± 0.34
Prototypical [32]	40.94 ± 0.47	54.54 ± 0.44	69.42 ± 0.39	78.01 ± 0.29
Relation [33]	36.19 ± 0.46	43.49 ± 0.44	53.14 ± 0.40	59.64 ± 0.39
FTL (Ours)	**49.93 ± 0.40**	**58.64 ± 0.39**	**70.53 ± 0.37**	**79.46 ± 0.32**
FTL-dis (Ours)	**51.91 ± 0.40**	**60.76 ± 0.36**	**72.13 ± 0.40**	**81.21 ± 0.31**

4. Conclusions

An improved pre-training and fine-tuning for few-shot SAR recognition method (FTL) is proposed in this paper. Its structure can be decomposed into three main phases: data engine—using the style transfer model to generate a large-scale SAR-like style base class dataset, which can effectively attenuate the domain differences between the novel class dataset and the base class dataset; model training phase—the generated base class dataset is used to pre-train the model, while a DeepBDC pooling layer is added in the feature extraction phase for enhanced representation of the contour features of SAR image targets; and fine-tuning phase—a small amount of novel class data is fed to fine-tune only the classifier layer of the model. In order to enhance the performance of the model even further, the knowledge distillation approach is implemented concurrently. The usefulness of each module in the aforementioned architecture is experimentally proven, and the entire performance is assessed on the MSTAR dataset, where an accuracy of about 80% can be attained at 10-way 10-shot. Thus, FTL demonstrates its competitive performance.

Author Contributions: Conceptualization, C.Z.; methodology, C.Z.; software, C.Z.; validation, C.Z.; writing—original draft preparation, C.Z.; writing—review and editing, C.Z.; visualization, C.Z.; supervision, H.D. and B.D.; funding acquisition, H.D. and B.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of Heilongjiang Province under Grant LH2020F023.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SAR	Synthetic aperture radar
ATR	Automatic target recognition
ROI	Region containing the target of interest
DeepBDC	Deep Brownian distance covariance

References

1. Novak, L.M.; Owirka, G.J.; Brower, W.S.; Weaver, A.L. The automatic target-recognition system in SAIP. *Linc. Lab. J.* **1997**, *10*.
2. Ikeuchi, K.; Wheeler, M.D.; Yamazaki, T.; Shakunaga, T. Model-based SAR ATR system. In Proceedings of the Aerospace/Defense Sensing and Controls, Orlando, FL, USA, 8–12 April 1996.
3. Goodfellow, I.J.; Bengio, Y.; Courville, A.C. Deep Learning. *Nature* **2015**, *521*, 436–444.
4. Hertzmann, A. Painterly rendering with curved brush strokes of multiple sizes. In Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, Orlando, FL, USA, 19–24 July 1998.
5. Frigo, O.; Sabater, N.; Delon, J.; Hellier, P. Split and Match: Example-Based Adaptive Patch Sampling for Unsupervised Style Transfer. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July2016; pp. 553–561.
6. Winnemöller, H.; Olsen, S.C.; Gooch, B. Real-time video abstraction. *ACM Trans. Graph.* **2006**, *25*, 1221–1226. [CrossRef]
7. Gatys, L.A.; Ecker, A.S.; Bethge, M. Texture Synthesis Using Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems 28 (NIPS 2015), 2015. Available online: https://papers.nips.cc/paper/2015 (accessed on 10 March 2023).
8. Li, C.; Wand, M. Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2479–2486.
9. An, J.; Huang, S.; Song, Y.; Dou, D.; Liu, W.; Luo, J. ArtFlow: Unbiased Image Style Transfer via Reversible Neural Flows. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 862–871.
10. An, J.; Li, T.; Huang, H.; Shen, L.; Wang, X.; Tang, Y.; Ma, J.; Liu, W.; Luo, J. Real-time Universal Style Transfer on High-resolution Images via Zero-channel Pruning. *arXiv* **2020**, arXiv:2006.09029.

11. Santoro, A.; Bartunov, S.; Botvinick, M.M.; Wierstra, D.; Lillicrap, T.P. Meta-Learning with Memory-Augmented Neural Networks. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016.
12. Zhao, X.; Lv, X.; Cai, J.; Guo, J.; Zhang, Y.; Qiu, X.; Wu, Y. Few-Shot SAR-ATR Based on Instance-Aware Transformer. *Remote Sens.* **2022**, *14*, 1884. [CrossRef]
13. Yue, Z.; Gao, F.; Xiong, Q.; Sun, J.; Hussain, A.; Zhou, H. A novel few-shot learning method for synthetic aperture radar image recognition. *Neurocomputing* **2021**, *465*, 215–227. [CrossRef]
14. Vinyals, O.; Blundell, C.; Lillicrap, T.P.; Kavukcuoglu, K.; Wierstra, D. Matching Networks for One Shot Learning. In Proceedings of the Advances in Neural Information Processing Systems 29 (NIPS 2016), 2016. Available online: https://researchr.org/publication/nips-2016 (accessed on 10 March 2023).
15. Xie, J.; Long, F.; Lv, J.; Wang, Q.; Li, P. Joint Distribution Matters: Deep Brownian Distance Covariance for Few-Shot Classification. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 7962–7971.
16. Settles, B. *Active Learning Literature Survey*; Department of Computer Sciences, University of Wisconsin-Madison: Madison, WI, USA, 2009.
17. Aggarwal, C.C.; Kong, X.; Gu, Q.; Han, J.; Philip, S.Y. Active learning: A survey. In *Data Classification*; Chapman and Hall/CRC: London, UK, 2014; pp. 599–634.
18. Tian, Y.; Krishnan, D.; Isola, P. Contrastive Representation Distillation. In Proceedings of the 2020 International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
19. Adriana, R.; Nicolas, B.; Ebrahimi, K.S.; Antoine, C.; Carlo, G.; Yoshua, B. Fitnets: Hints for thin deep nets. *Proc. ICLR* **2015**, *2*.
20. Yim, J.; Joo, D.; Bae, J.; Kim, J. A gift from knowledge distillation: Fast optimization, network minimization and transfer learning. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4133–4141.
21. Zhang, L.; Song, J.; Gao, A.; Chen, J.; Bao, C.; Ma, K. Be your own teacher: Improve the performance of convolutional neural networks via self distillation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3713–3722.
22. Mahmoud, A.S.; Mohamed, S.A.; Moustafa, M.S.; El-Khorib, R.A.; Abdelsalam, H.M.; El-Khodary, I.A. Training compact change detection network for remote sensing imagery. *IEEE Access* **2021**, *9*, 90366–90378. [CrossRef]
23. Zhang, Y.; Xiang, T.; Hospedales, T.M.; Lu, H. Deep mutual learning. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4320–4328.
24. Zhang, Z.; Sabuncu, M. Self-distillation as instance-specific label smoothing. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 2184–2195.
25. Shen, C.; Wang, X.; Song, J.; Sun, L.; Song, M. Amalgamating knowledge towards comprehensive classification. In Proceedings of the AAAI Conference on Artificial Intelligence, 2019; Volume 33, pp. 3068–3075. Available online: https://ojs.aaai.org/index.php/AAAI/issue/view/246 (accessed on 10 March 2023).
26. Fu, Y.; Li, S.; Zhao, H.; Wang, W.; Fang, W.; Zhuang, Y.; Pan, Z.; Li, X. Elastic knowledge distillation by learning from recollection. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**. [CrossRef] [PubMed]
27. Sz'ekely, G.J.; Rizzo, M.L. Brownian distance covariance. *Ann. Appl. Stat.* **2009**, *3*, 1236–1265. [CrossRef] [PubMed]
28. Furlanello, T.; Lipton, Z.; Tschannen, M.; Itti, L.; Anandkumar, A. Born again neural networks. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 1607–1616.
29. Zhang, C.; Cai, Y.; Lin, G.; Shen, C. DeepEMD: Few-Shot Image Classification With Differentiable Earth Mover's Distance and Structured Classifiers. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 12200–12210.
30. Zhang, C.; Cai, Y.; Lin, G.; Shen, C. DeepEMD: Differentiable Earth Mover's Distance for Few-Shot Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**.. [CrossRef] [PubMed]
31. Li, W.; Wang, L.; Xu, J.; Huo, J.; Gao, Y.; Luo, J. Revisiting Local Descriptor Based Image-To-Class Measure for Few-Shot Learning. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 7253–7260.
32. Snell, J.; Swersky, K.; Zemel, R.S. Prototypical Networks for Few-shot Learning. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017. Available online: https://papers.nips.cc/paper/2017 (accessed on 10 March 2023).
33. Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Torr, P.H.S.; Hospedales, T.M. Learning to Compare: Relation Network for Few-Shot Learning. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1199–1208.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.