Article

Enhanced Prototypical Network with Customized Region-Aware Convolution for Few-Shot SAR ATR

Xuelian Yu [*] , Hailong Yu , Yi Liu , Haohao Ren

Article

Enhanced Prototypical Network with Customized Region-Aware Convolution for Few-Shot SAR ATR

Xuelian Yu [1,*], Hailong Yu [1], Yi Liu [2] and Haohao Ren [1]

[1]   School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

[2]   The 22th Research Institute of China Electronics Technology Group Corporation, Qingdao 266000, China

*   Correspondence: xlyu2012@uestc.edu.cn

Abstract: With prosperous development and successful application of deep learning technologies in the field of remote sensing, numerous deep-learning based methods have emerged for synthetic aperture radar (SAR) automatic target recognition (ATR) tasks over the past few years. Generally, most deep-learning based methods can achieve outstanding recognition performance in condition that an abundant of labeled samples are available to train the model. However, in real application scenarios, it is difficult and costly to acquire and to annotate abundant SAR images due to the imaging mechanism of SAR, which poses a big challenge to existing SAR ATR methods. Thereby, SAR target recognition under the situation of few-shot, where only a scarcely few of labeled samples are available, is a fundamental problem that needs to be solved. In this paper, we put forward a new method named enhanced prototypical network with customized region-aware convolution (CRCEPN) to specially tackle the few-shot SAR ATR tasks. To be specific, We first develop a feature extraction network based on a customized and region-aware convolution, which can adaptively adjust convolutional kernels and their receptive fields according to each SAR image's own characteristics as well as the semantic similarity among spatial regions, thus augmenting its capability to extract more informative and discriminative features. To achieve accurate and robust target identity prediction under the few-shot condition, we then propose an enhanced prototypical network which can improve the representation ability of the class prototype by properly making use of training and test samples together, thus effectively raising the classification accuracy. Meanwhile, a new hybrid loss is designed to learn a feature space with both inter-class separability and intra-class tightness as much as possible, which can further upgrade the recognition performance of the proposed method. Experiments performed on the moving and stationary target acquisition and recognition (MSTAR) dataset and the OpenSARShip dataset demonstrate that the proposed method is competitive with some state-of-the-art methods for few-shot SAR ATR tasks.

Keywords: convolutional neural network (CNN); synthetic aperture radar (SAR); automatic target recognition (ATR); few-shot learning (FSL)

1. Introduction

Synthetic aperture radar (SAR) is a microwave remote sensor used to provide high-resolution images with all-day-and-night and all-weather operating characteristics, which has been widely used in various military and civilian fields [1–3]. Automatic target recognition (ATR) is a fundamental but also challenging task in SAR domain [4]. It consists of two key procedures, i.e., feature extraction and target classification, which are independent of each other in traditional SAR ATR methods. Moreover, traditional methods rely on hand-crafted features, which hinders the development of SAR ATR.

With prosperous development and successful application of deep learning technologies in the field of remote sensing, studies on SAR ATR have gained significant breakthroughs [5]. Numerous deep-learning based SAR ATR methods continue to emerge over the past few years and demonstrate their superiority to traditional methods. To name just a few, Chen et al. [6] were among the first who applied deep convolutional neural network (CNN) to SAR ATR tasks, laying a foundation for follow-up studies in this field. Kechagias-Stamatis and Aouf [7] proposed a SAR ATR method by fusing deep learning and sparse coding, which can achieve excellent recognition performance under different situations. Zhang et al. [8] proposed a multi-view classification with semi-supervised learning

for SAR target recognition. Pei et al. [9] designed a two-stage algorithm based on contrastive learning for SAR image classification. Zhang et al. [10] proposed a separability measure-based CNN for SAR ATR, which can quantitatively analyze the interpretability of feature maps.

One of the biggest challenges for most deep-learning based methods is that they are data-hungry and often require hundreds or thousands of training samples to achieve state-of-the-art accuracy [11]. However, in real SAR ATR scenarios, the scarcity of labeled samples is a common problem due to the imaging mechanism of SAR. Under the situation where a scarcely few labeled SAR images are only available, which is termed as few-shot problem, most existing deep-learning based SAR ATR methods will suffer severe performance decline.

In face of this challenge, a variety of few-shot learning (FSL) methods have been proposed in the past few years. Among them, prototypical network (ProtoNet) [12], relation network (RelationNet) [13], transductive propagation network (TPN) [14], cross attention network (CAN) and transductive CAN[15], graph neural network (GNN) [16], and edge-labeling GNN [17] are some representatives in the field of computer vision. Subsequently, some FSL methods were proposed specifically for SAR ATR under few-shot conditions [18–22]. For instance, Liu et al. [23] put forward a bi-similarity prototypical network with capsule-based embedding (BSCapNet) to solve the problem of few-shot SAR target recognition. Experiments on moving and stationary target acquisition and recognition (MSTAR) dataset show its effectiveness and superiority to some state-of-the-arts. Bi et al. [24] proposed a contrastive domain adaptation based SAR target classification method to solve the problem of insufficient samples. Experimental results on MSTAR dataset demonstrate the effectiveness. Fu et al. [25] proposed a metalearning framework for few-shot SAR ATR (MSAR). Yang et al. [26] came up with mixed loss graph attention network (MGANet) for few-shot SAR target classification. Wang et al. [27] presented a multitask representation learning network (MTRLN) for few-shot SAR ATR. Yu et al. [28] presented a transductive prototypical attention network (TPAN). Ren et al. [29] proposed adaptive convolutional subspace reasoning network (ACSRNet). Liao et al. [30] put forward a model-agnostic meta-learning (MAML) for few-shot image classification. Although some significant achievements have been made, studies on few-shot SAR ATR are yet in its infancy and there remains considerable potential to be explored.

Our goal in this paper is to boost the achievements on few-shot SAR ATR and to further improve the recognition performance by proposing a new method named enhanced prototypical network with customized region-aware convolution (CRCEPN). Extensive evaluation experiments on both the MSTAR dataset and the OpenSARship dataset verify the effectiveness as well as the superiority of the proposed method compared to some state-of-the-arts for few-shot SAR ATR. The main contributions of this paper can be summarized as follows.

1. A feature extraction network based on a customized and region-aware convolution (CRConv) is developed, which can adaptively adjust convolutional kernels and their receptive fields according to each sample's own characteristics and the semantical similarity among spatial regions. Consequently, CRConv can adapt better to diverse SAR images and is more robust to variations in radar view, which augments its capacity to extract more informative and discriminative features. This greatly improves the recognition performance of the proposed method especially under few-shot conditions .

2. To achieve accurate and robust target identity prediction for few-shot SAR ATR, we propose an enhanced prototypical network, which can effectively enhance the representation ability of the class prototypes by utilizing both support and query samples, thereby raising the classification accuracy.

3. We propose a new loss function, namely aggregation loss to minimize the intra-class compactness. With the joint optimization of the aggregation loss and the cross-entropy loss, not only the inter-class differences are enlarged but also the intra-class variations are reduced in the feature space. Thereby, highly discriminative features can be obtained for few-shot SAR ART, thus improving the recognition performance, as supported by the experimental results.

The rest of this paper is organized as follows. Section 2 details the framework and each key component of the proposed method. In section 3, extensive experiments on both the MSTAR and the OpenSARship dataset are performed, and experimental results are analyzed in detail. Section 4 concludes this work.

2. Methodology

The overall framework of the proposed method is illustrated in Figure 1. Following the idea of meta-learning, the proposed method consists of two stages, i.e., meta-training and meta-test. Below, three main components, including feature extraction module, classification module, and loss function, are elaborated.
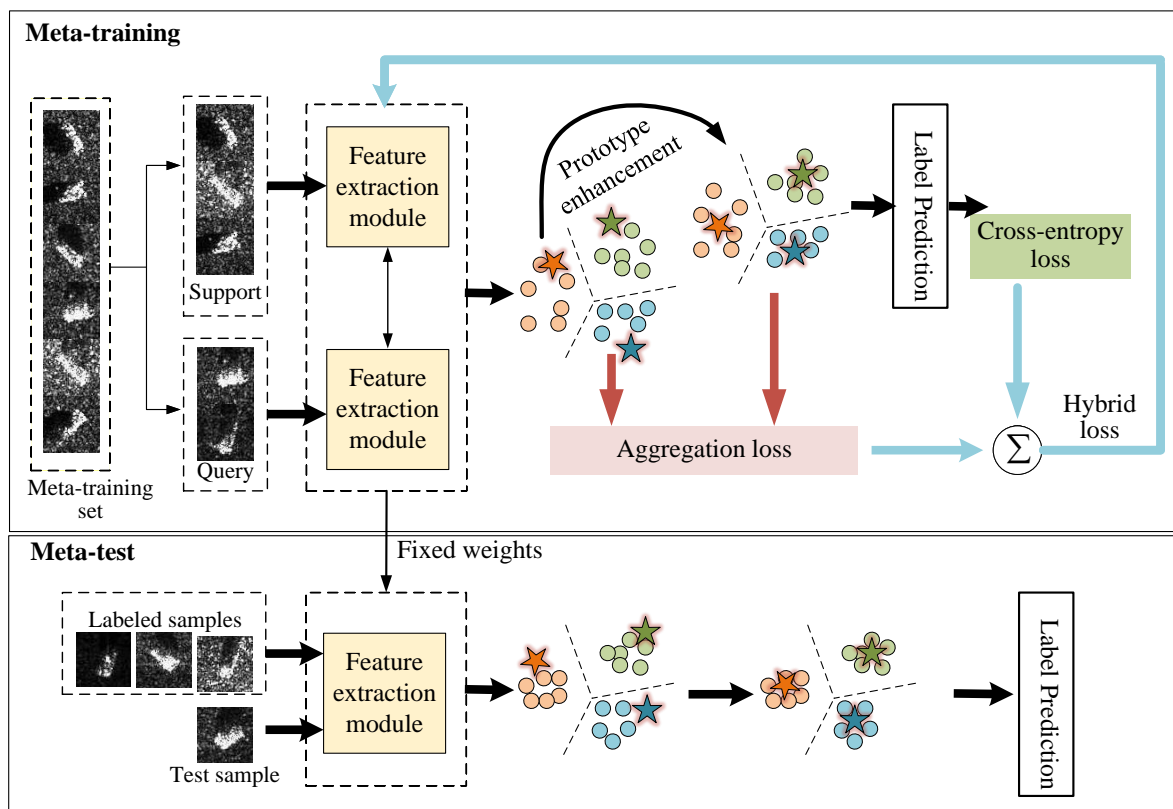

Figure 1. The overall framework of the proposed method.

2.1. Feature Extraction Module

As a vital component of SAR ATR systems, feature extraction plays an important role in improving the effectiveness and robustness of the whole ATR system. The convolutional neural network (CNN) stands as a cornerstone in the development of deep learning technology, primarily due to its ability to efficiently process and extract rich features from images.

However, the traditional CNN is characterized by using standard static convolutional kernels, which exists two deficiencies in nature. First, standard convolution performs in kernel sharing manner across spatial domain, thus limiting its representation ability due to single kernel's poor capacity. Second, static convolutional kernels are randomly generated and shared among all input images, which is not effective to capture each image's uniqueness.

Owing to SAR image's sensitivity to variations in radar view, different images of the same target may appear variable spatial information distribution, while images from different targets may appear similarly. Under few-shot condition, this problem may be more prominent. Consequently, standard static convolution may not be able to effectively extract discriminative features for few-shot SAR ATR [29].

Inspired by dynamic mechanism [29,31–33], we introduce a customized and region-aware convolution (CRConv) in this paper, and then construct the feature extraction network by cascading four layers of CRConv. There are two distinct features in CRConv that are different from traditional convolution. First, different SAR images no longer share the same convolutional kernels in CRConv, but use their own unique kernels, that is, the convolutional kernels are customized for each image according to their own characteristics. Second, for each SAR image, different spatial regions may adopt different kernels and also different kernel size according to their semantical similarity, that is, the convolutional kernels are region-aware. This is useful since the target area and the background area usually display much different semantic, and the center area and edge area of the target may also be semantically different.

Concretely, by generating multiple customized and region-aware convolutional kernels for each SAR image and dynamically assign them to corresponding spatial regions, CRConv is capable of capturing specific feature of each sample and handling variable spatial information distribution. It can therefore adapt better to diverse SAR images, which not only can augment its representation capacity but also makes it more robust to variations in radar view. We can thus anticipate that the feature extraction network based on CRConv is effective to extract more informative and discriminative features for few-shot SAR ATR, thus improving the recognition performance.

As shown in Figure 2, CRConv comprises three main parts: dynamic region division, customized kernel generation, and dynamic convolution. In the part of dynamic region division, each SAR image is divided into several regions across spatial dimension according to their semantical similarity. In the part of customized kernel generation, multiple customized kernels are generated specifically for each SAR image and each region. Dynamic convolution is finally performed between each individual region and its corresponding convolutional kernel. In the following, each part is elaborated.
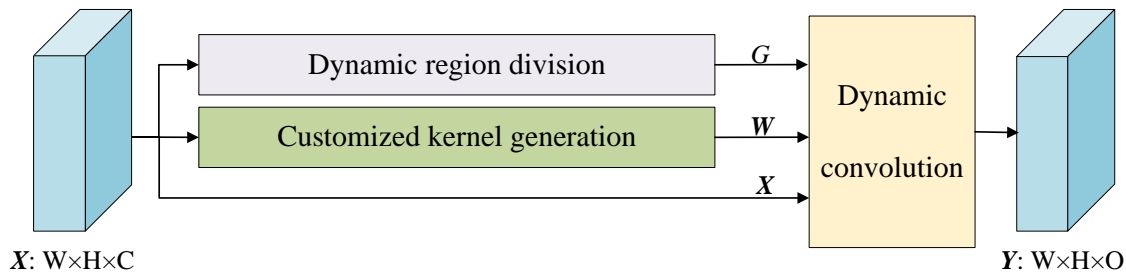

Figure 2. Flowchart of CRConv.

2.1.1. Dynamic Region Division

The implementation procedure of dynamic region division is shown in Figure 3. For each input image $\mathbf{X}$, a group of standard convolutions with n different sizes $k_s \times k_s (s = 1, \cdots, n)$ are firstly applied to produce a set of feature maps,

$$\mathbf{F} = \{\mathbf{F^1}, \cdots, \mathbf{F^p}\} \tag{1}$$

where $p = m \times n$, m is the kernel number of each size, and n is the number of different kernel sizes. One can see that in CRConv, convolutional kernels with different receptive field are used to adapt better to multiscale semantical features. This is different from [31] where the kernel size is single.

Then, a region division map $\mathbf{M}$ is obtained by

$$M_{i,j} = \text{argmax}(F_{i,j}^1, \cdots, F_{i,j}^p) \tag{2}$$

where $\text{argmax}(\cdot)$ outputs the index of maximum value, and (i, j) represents spatial position. It is easy to see that the values in $\mathbf{M}$ vary from 1 to p, i.e., $M_{i,j} \in [1, p]$, which can be expressed by one-hot-form. For instance, $M_{i,j} = 2$, $p = 4$, and the one-hot-form is $[0, 1, 0, 0]$.

To make the region division map learnable, in backward propagation, the softmax operation is applied to feature map **F** in order to approximate the one-hot-form of **M** [31],

$$\hat{F}_{i,j}^t = \frac{\exp(F_{i,j}^t)}{\sum_{v=1}^{p} \exp(F_{i,j}^v)}, \quad 1 \le t \le p. \tag{3}$$

Finally, based on the region division map, the entire spatial pixels of the input SAR image can be divided into p regions,

$$G = \{G_1, \cdots, G_p\} \tag{4}$$

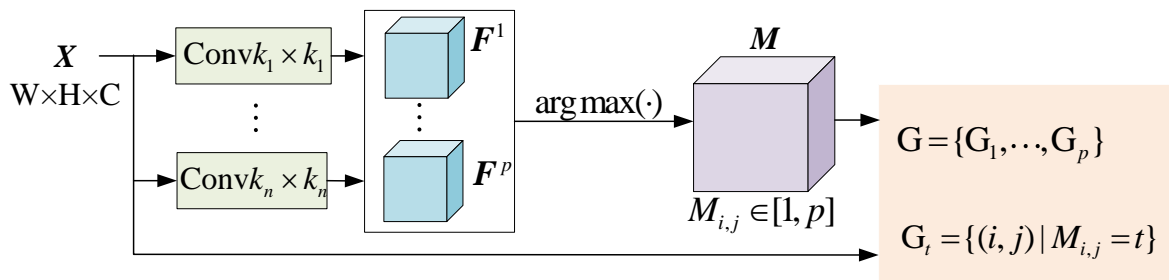with $G_t = \{(i,j) | M_{i,j} = t\}, 1 \le t \le p$.

Figure 3. Dynamic region division.

2.1.2. Customized Kernel Generation

Figure 4 illustrates the implementation procedure of customized kernel generation, where multiple customized kernels are generated specifically for each SAR image and each region. As shown in Figure 4, the input image **X** is first average pooled such that it is down-sampled to n different sizes, $k_s \times k_s (s = 1, \cdots, n)$. Then, a 1×1 convolution including sigmoid as activation function is applied, followed by another 1×1 convolution with group=m. Finally, a set of p kernels are generated for each input-output channel pair, denoted as $\mathbf{W} = \{\mathbf{W_1}, \cdots, \mathbf{W_p}\}$.

Apparently, the convolutional kernels of CRConv are customized and region-aware, which can be adjusted adaptively according to each sample's own characteristics and can make full use of the diversity of spatial information. As a result, the CRConv adapts better to diverse SAR images and is more robust to variations in radar view, which is helpful for improving the recognition performance, especially under few-shot conditions.

Moreover, utilizing multiscale kernels is able to further enhance the representation capacity of CRConv. Intuitively, by accommodating features of varying scales, these kernels ensure that both fine-grained details and broader spatial patterns are effectively captured. This is particularly beneficial for SAR images, where targets may vary greatly in size and shape, and background clutter may obscure critical details.
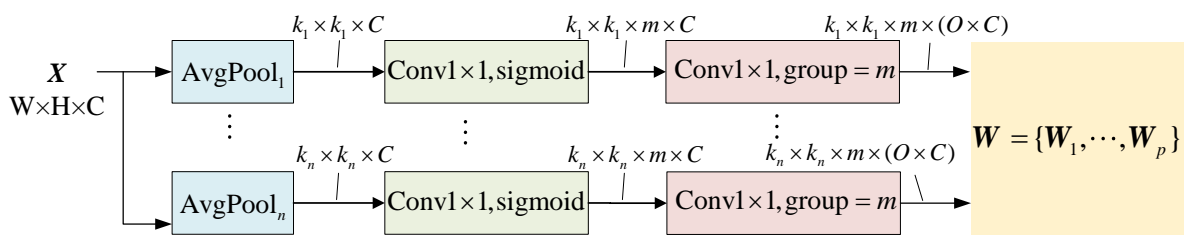

Figure 4. Customized kernel generation.

2.1.3. Dynamic Convolution

With the process of the above two parts, i.e., dynamic region division and customized kernel generation, the input image is divided into p regions, each region corresponding to a customized convolutional kernel. Then, the convolution operation in CRConv can be expressed as

$$Y_{i,j}^{(r)} = \sum_{c=1}^{C} X_{i,j}^{(c)} * W_t^{(c,r)}, \quad (i,j) \in G_t, \quad 1 \le t \le p \tag{5}$$

where $X_{i,j}^{(c)}$ is the c-th channel input, $Y_{i,j}^{(r)}$ is the r-th channel output, $W_t^{(c,r)}$ represents the convolutional kernels from the c-th input to the r-th output, and $*$ is 2D convolution operation. From (5), we see that in CRConv different regions use different customized convolutional kernels, and in each region a standard convolution operation is performed. Specifically, spatial pixels in the region G_t uses the convolutional kernel W_t.

2.2. Classification Module

Prototype learning is a popularly adopted and widely used classification method in the realm of few-shot learning. It has also been utilized for few-shot SAR ATR [23]. To put it simply, the core of prototype learning involves the computation of a class prototype that essentially represents the average feature vector of all samples within a given class. Mathematically, the prototype for the i-th class, denoted as m_i, is calculated by

$$m_i = \frac{1}{K} \sum_{x \in S_i} \phi(x), \quad i = 1, \cdots, N \tag{6}$$

where $\phi(\cdot)$ represents the feature extraction network, S_i is the support set belonging to the i-th class, N is the number of target categories, and K is the number of labeled samples per category. The target identity of a sample in the query set, $z \in Q$, is then predicted according to the nearest distance between $\phi(z)$ and each class prototype.

Clearly, the true prototype may not be obtained accurately with only a few labeled samples under few-shot condition. Besides, owing to SAR image's sensitivity to variations in radar view, the representation ability of the prototype for each SAR target will be further weakened, resulting in lowered classification accuracy [29]. To improve the classification accuracy for few-shot SAR ATR, we develop an enhanced prototypical network, which can effectively enhance the representation ability of class prototype by utilizing both support and query samples.

Firstly, the initial probability of a query sample z belonging to the i-th class is predicted by

$$w_{z,i} = \frac{\exp(-d(\phi(z), m_i))}{\sum_{j=1}^{N} \exp(-d(\phi(z), m_j))}, \quad z \in Q \tag{7}$$

where $d(\cdot, \cdot)$ represents the similarity between $\phi(z)$ and m_i which is obtained by an exponential mapping from the Euclidean distance, i.e.,

$$d(\phi(z), m_i) = \exp\left(\|\phi(z) - m_i\|^2\right) - 1. \tag{8}$$

By using (7) and (8), we elaborately assign larger weights to samples that are more similar to initial class prototypes and smaller weights to samples which are less similar to initial class prototypes.

Then, an enhanced class prototype with augmented representative ability is generated based on both support and query sets as below

$$m_i^* = \frac{\sum_{x \in S_i} \phi(x) + \sum_{z \in Q} w_{z,i} \phi(z)}{K + \sum_{z \in Q} w_{z,i}}. \tag{9}$$

Based on the class prototype obtained via (9), the final probability of a query sample z belonging to the i-th class is calculated by

$$p(y = i|z) = \frac{\left\|\phi(z) - m_i^*\right\|^2}{\sum_{j=1}^N \left\|\phi(z) - m_j^*\right\|^2}.$$ (10)

And the target identity of z is predicted as

$$c_0 = \arg\max(p(y = i|z)), \quad i = 1, \cdots, N.$$ (11)

From above process, one can see that by utilizing both support and query samples dexterously, the proposed enhanced prototypical network can effectively enhance the representation ability of class prototype and achieve robust target identity prediction, which will help to raise the classification accuracy for few-shot SAR ATR.

2.3. Loss Function

The proposed method utilizes a hybrid loss function which is composed of two distinct components: the cross-entropy loss and the aggregation loss.

The cross-entropy loss is utilized to train the model at classification level, which can be expressed as

$$L_c = -\frac{1}{M} \sum_{z \in Q} \sum_{i=1}^N y_{z,i} \log(p(y = i|z))$$ (12)

where M is the total number of samples in query set, and $y_{z,i} = 1$ if z belongs to the i-th class, otherwise $y_{z,i} = 0$.

In order to promote the discriminative ability of the extracted features, we propose an aggregation loss, which is defined as

$$L_a = \frac{1}{2N} \sum_{i=1}^N \left\|m_i^* - m_i\right\|^2.$$ (13)

As can be seen from (13), the aggregation loss can update the prototype of each class and simultaneously penalize the distances between initial and enhanced prototypes. Thereby, a hybrid loss which combines (12) and (13) is formulated to train the proposed model, i.e.,

$$L = L_c + \lambda L_a$$ (14)

where λ is used for balancing the two losses.

On one hand, the cross-entropy loss plays a crucial role in ensuring that the model can accurately classify input samples by minimizing the disparity between the predicted class probabilities and the actual class labels. It encourages the model to create clear boundaries between different classes, thereby enhancing inter-class separability. On the other hand, the aggregation loss focuses on the inner clustering within each class by minimizing the distances between initial and enhanced prototypes. Intuitively, the former forces the features of different classes staying apart, while the latter pulls features of the same class staying together. With joint optimization of both losses, we train a robust network to obtain features with both inter-class separability and intra-class tightness as much as possible, which can beneficially improve the recognition performance of the proposed method.

3. Experimental Results and Analysis

3.1. Dataset Description

1) MSTAR dataset: The MSTAR dataset was publicly released by the Defense Advanced Research Projects Agency (DARPA). It includes ten categories of ground military targets, i.e., BMP2, BTR60, BTR70, T62, T72, D7, ZIL131, 2S1, BRDM2, and ZSU23/4. SAR images of each target were collected using an X-band SAR in spotlight mode under three different depression angles (15°, 17°, and 30°) over full aspect angles from 0° to 360°. All images are 128×128 pixels in size. Figure 5 shows some optical images and corresponding SAR images of ten categories of targets from the MSTAR dataset.
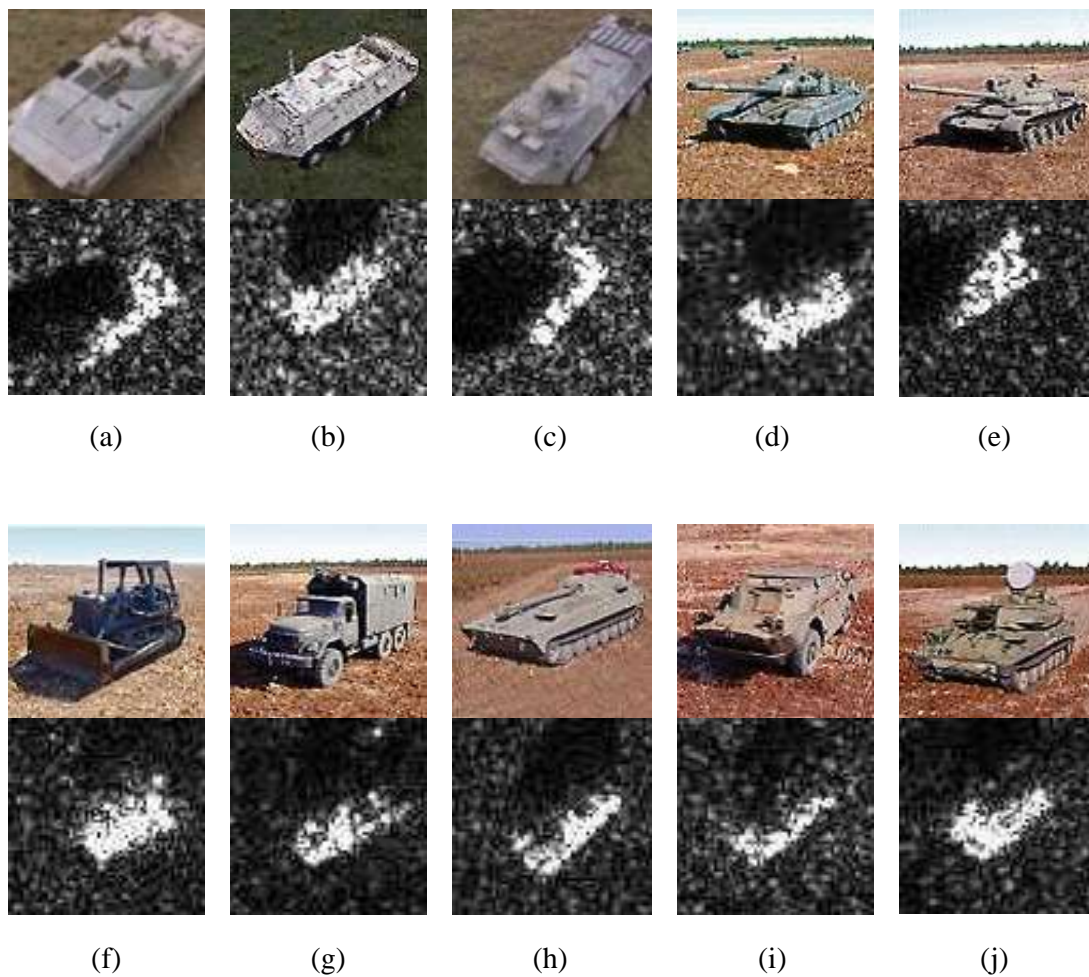
Figure 5. Some optical images and their corresponding SAR images from the MSTAR dataset. (a) BMP2, (b) BTR60, (c) BTR70, (d) T60, (e) T70, (f) D7, (g) ZIL131, (h) 2S1, (i) BRDM2, (j) ZSU23/4.

Based on the idea of meta learning, we split the MSTAR dataset into meta-training set and meta-test set. Referring to existing studies [22,23,25,27,28], SAR images of seven targets including BMP2, BTR60, BTR70, T62, T72, D7, and ZIL131 collected at 17° depression angle constitute the meta-training set, while SAR images of the other three targets, i.e., 2S1, BRDM2, and ZSU23/4, collected at 15°, 17°, and 30° depression angles act as the meta-test set. Table 1 summarizes the detailed information.

Table 1. Details of the MSTAR dataset for meta-training and meta-test.

Stage	Depression	Class	Number	Stage	Depression	Class	Number
		BMP2	233			2S1	274
		BTR60	256		15°	BRDM2	274
		BTR70	233			ZSU23/4	274
Training	17°	T62	299	Test		2S1	299
		T72	232		17°	BRDM2	298
		D7	299			ZSU23/4	299
		ZIL131	299			2S1	288
					30°	BRDM2	287
						ZSU23/4	288

2) OpenSARShip dataset: The OpenSARShip dataset was publicly released by Shanghai Jiaotong University, which is widely used as a benchmark for evaluating SAR target detection and recognition algorithms. The dataset contains 11,346 SAR images from 17 categories of ship targets, and all images were derived from 41 Sentinel-1 images with four polarization modes. The resolution of each SAR image is 10m×10m. In this paper, SAR images with vertical-vertical (VV) and vertical-horizontal (VH) polarization are utilized for experiments. Figure 6 illustrates some optical and corresponding SAR images of six ship targets in the OpenSARShip dataset. Referring to previous work [8,21], SAR images of three ship targets, i.e., Dredger, Fishing, and Tug are used as the meta-traning data, while images of Bulk carrier, Container ship, and Tanker are used as the meta-test data. Table 2 lists the number of each target in the meta-training set and meta-test set.
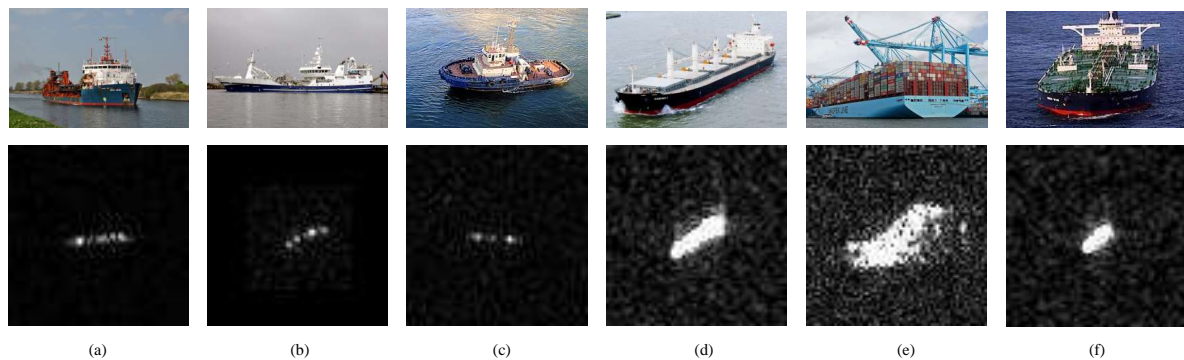
Figure 6. Some optical images and their corresponding SAR images from the OpenSARShip dataset. (a) Dredger, (b) Fishing, (c) Tug, (d) Carrier, (e) Container, (f) Tanker.

Table 2. Details of the OpenSARship dataset for meta-training and meta-test.

Stage	Class	Number	Stage	Class	Number
	Dredger	300		Carrier	80
Training	Fishing	300	Test	Container	80
	Tug	300		Tanker	80

3.2. Experimental Detail

Following previous work [22,23,25,27,28], we simulate two few-shot SAR ATR tasks, i.e., 3-way 1-shot and 3-way 5-shot, on both the MSTAR dataset and the OpenSARShip dataset. On the MSTAR dataset, each few-shot task is conducted under five different experimental scenarios so as to evaluate the robustness of the proposed method. Table 3 lists the detailed information of five different experimental scenarios. In the following experiments, each SAR image is cropped into 64×64 pixels

in size. And each experiment is conducted 1000 independent runs in order to obtain statistically significant results.

To demonstrate the superiority of the proposed method for few-shot SAR ATR tasks, several state-of-the-art few-shot methods are employed for performance comparison, including ProtoNet [12], RelationNet [13], TPN [14], CAN and TCAN [15], MSAR [25], BSCapNet [23], MTRLN [27],TPAN [28], 2SCNet[22], ACSRNet [29], and MAML [30]. To be fair, the feature extraction network in RelationNet, TPN, CAN, TCAN, MSAR, and 2SCNet keeps the same with that of ProtoNet [12], i.e., consisting of four convolutional layers. BSCapNet, MTRLN, TPAN, ACSRNet, and MAML are reproduced according to the setting in original papers.

The Adam optimizer [34] is adopted to optimize the proposed method, and the learning rate is set to 0.01. The hyperparameter λ in (14) is fine-tuned to 0.01 for a relatively optimum balance between the cross-entropy loss and the aggregation loss. In CRConv, four different sizes of convolutional kernels are generated, i.e., 1×1, 3×3, 5×5, and 7×7, and the kernel number for each size is set to 4. This results in a total of sixteen customized kernels for each SAR images, that is, $m = n = 4$ and $p = m \times n = 16$.

All experiments are implemented in the PyTorch framework, making use of its dynamic computational graph and extensive library of tools for deep learning research. All experiments are conducted on a high-performance server equipped with a 16-core AMD Ryzen 9 7950X CPU and an NVIDIA GeForce RTX 3090 Ti GPU.

Table 3. Five different experimental scenarios on the MSTAR dataset.

| Experimental scenario | Depression angle | | | |
| | Training Stage | | Test Stage | |
	Support	Query	Support	Query
1			15°	15°
2			15°	17°
3	17°	17°	17°	17°
4			17°	30°
5			30°	30°

3.3. Recognition Results on the MSTAR Dataset

In this section, extensive experiments are conducted on the MSTAR dataset in order to evaluate the performance of the proposed method for few-shot SAR ATR tasks. Table 4 lists the recognition results of the proposed method and other competitors under five different experimental scenarios. Table 5 shows the confusion matrices of the average recognition accuracy of the proposed CRCEPN under different experimental scenarios.

As can be seen from Table 4, the proposed CRCEPN consistently performs better than all competitors for either 1-shot or 5-shot task under different experimental scenarios on the MSTAR dataset. One can also see that TPAN performs the second best on the whole. Particularly, in the first four experiments for both few-shot tasks, the recognition rates of the proposed CRCEPN are about 2%-9% higher than those of TPAN, and the performance improvements compared with other competitors are more remarkable. In the fifth experiment, where the depression variation between the meta-training set and the meta-test set is enlarged and thus the SAR ATR task is more challenging, the proposed CRCEPN still surpasses TPAN by about 5% for 5-shot task and 2% for 1-shot task, respectively. Overall, these experimental results manifest the effectiveness and superiority of the proposed CRCEPN for few-shot SAR ATR tasks.

It is worth noting that in TPAN, a cross-feature spatial attention module is designed following the feature extractor to get more discriminative features [28]. While in the proposed CRCEPN, we use only a feature extractor consisting of 4 layers of CRConv.

Moreover, from the confusion matrices in Table 5, one can see that the recognition performance of the proposed method on each target is relatively balanced under different experimental scenarios for both 1-shot and 5-shot tasks on the MSTAR dataset, which indicates huge potential of the proposed method for few-shot SAR target recognition.

Table 4. Recognition performance comparison on the MSTAR dataset.

Setting	Method	Experimental scenario				
		1	2	3	4	5
1-shot	ProtoNet [12]	71.51 ± 0.65	66.81 ± 0.66	69.90 ± 0.62	64.05 ± 0.65	66.76 ± 0.79
	RelationNet [13]	74.22 ± 0.67	67.50 ± 1.22	69.95 ± 0.62	64.41 ± 0.70	62.78 ± 0.84
	TPN [14]	75.51 ± 0.83	70.79 ± 0.69	77.48 ± 0.69	73.70 ± 0.73	68.26 ± 0.70
	CAN [15]	77.36 ± 0.65	70.31 ± 0.76	75.79 ± 0.80	72.82 ± 0.86	78.30 ± 0.85
	TCAN [15]	79.44 ± 0.73	72.73 ± 0.82	77.96 ± 0.91	75.47 ± 0.87	78.71 ± 0.94
	MSAR [25]	67.71 ± 0.85	66.59 ± 0.85	68.43 ± 0.88	63.71 ± 1.04	64.51 ± 1.12
	2SCNet [22]	78.02 ± 0.68	70.02 ± 0.63	72.14 ± 0.66	68.21 ± 0.62	68.26 ± 0.71
	BSCapNet [23]	72.01 ± 0.80	70.84 ± 0.71	74.37 ± 0.65	73.98 ± 0.47	67.22 ± 0.68
	MTRLN [27]	86.11 ± 0.70	74.32 ± 0.79	81.59 ± 0.89	78.20 ± 0.73	70.17 ± 0.76
	ACSRNet [29]	79.55 ± 0.81	74.58 ± 0.72	75.54 ± 0.69	74.02 ± 0.91	70.79 ± 0.78
	MAML [30]	74.23 ± 0.46	74.36 ± 0.10	76.08 ± 0.10	73.90 ± 0.42	72.92 ± 0.45
	TPAN [28]	86.90 ± 0.64	81.03 ± 0.61	87.14 ± 0.68	84.15 ± 0.62	78.87 ± 0.75
	CRCEPN(ours)	**94.05 ± 0.48**	**89.70 ± 0.52**	**91.26 ± 0.49**	**87.62 ± 0.50**	**80.86 ± 0.67**
5-shot	ProtoNet [12]	80.28 ± 0.42	81.29 ± 0.41	82.22 ± 0.42	78.24 ± 0.43	78.59 ± 0.49
	RelationNet [13]	79.54 ± 0.40	80.46 ± 0.42	81.21 ± 0.39	74.36 ± 0.43	74.60 ± 0.57
	TPN [14]	82.87 ± 0.56	76.98 ± 0.55	85.04 ± 0.52	80.65 ± 0.55	79.31 ± 0.48
	CAN [15]	85.53 ± 0.40	77.42 ± 0.47	84.70 ± 0.51	82.24 ± 0.47	86.57 ± 0.52
	TCAN [15]	85.89 ± 0.41	78.52 ± 0.52	85.75 ± 0.52	82.46 ± 0.52	86.92 ± 0.43
	MSAR [25]	82.62 ± 0.58	82.21 ± 0.63	83.83 ± 0.61	77.89 ± 0.67	77.55 ± 0.69
	2SCNet [22]	84.21 ± 0.40	82.40 ± 0.41	84.91 ± 0.39	81.08 ± 0.41	81.13 ± 0.45
	BSCapNet [23]	82.65 ± 0.45	78.71 ± 0.47	86.60 ± 0.48	83.20 ± 0.44	80.51 ± 0.49
	MTRLN [27]	93.78 ± 0.50	82.72 ± 0.56	91.23 ± 0.43	87.22 ± 0.67	78.72 ± 0.52
	ACSRNet [29]	89.77 ± 0.52	83.08 ± 0.48	87.96 ± 0.56	83.82 ± 0.49	81.53 ± 0.67
	MAML [30]	93.31 ± 0.50	85.89 ± 0.86	93.90 ± 0.45	89.18 ± 0.65	88.45 ± 0.69
	TPAN [28]	92.73 ± 0.31	85.97 ± 0.42	94.27 ± 0.29	90.27 ± 0.36	87.34 ± 0.47
	CRCEPN(ours)	**95.74 ± 0.23**	**94.92 ± 0.33**	**97.24 ± 0.18**	**92.68 ± 0.27**	**93.00 ± 0.26**

Table 5. Confusion matrix of the recognition accuracy of CRCEPN on the MSTAR dataset.

Scenario		Setting					
		1-shot			5-shot		
		2S1	BRDM2	ZSU23/4	2S1	BRDM2	ZSU23/4
1	2S1	**94.09**	3.62	2.34	**95.50**	2.15	1.93
	BRDM2	2.98	**93.33**	2.94	2.43	**95.70**	2.04
	ZSU23/4	2.93	3.05	**94.72**	2.07	2.15	**96.03**
2	2S1	**89.83**	5.10	5.54	**95.20**	2.34	2.86
	BRDM2	4.82	**89.50**	4.67	2.44	**94.91**	2.48
	ZSU23/4	5.35	5.40	**89.79**	2.36	2.75	**94.66**
3	2S1	**91.18**	4.68	4.62	**97.33**	1.40	1.31
	BRDM2	3.63	**91.39**	4.15	1.46	**96.99**	1.27
	ZSU23/4	5.19	3.93	**91.23**	1.21	1.61	**97.42**
4	2S1	**86.68**	5.39	6.36	**91.26**	4.02	2.74
	BRDM2	6.46	**88.47**	5.93	3.82	**93.11**	3.53
	ZSU23/4	6.86	6.14	**87.71**	4.92	2.87	**93.73**
5	2S1	**80.98**	9.15	9.81	**92.41**	3.35	3.50
	BRDM2	9.81	**81.37**	9.95	3.41	**93.44**	3.34
	ZSU23/4	9.21	9.48	**80.24**	4.18	3.21	**93.16**

3.4. Recognition Results on the OpenSARship Dataset

This section performs evaluation experiments on the OpenSARShip dataset so as to further verify the recognition performance of the the proposed method for few-shot SAR ATR tasks. Table 6 lists the recognition results of each method for both 1-shot and 5-shot tasks on the OpenSARShip dataset.

It can be seen from the experimental results in Table 6 that the proposed CRCEPN still exhibits the best performance for either 1-shot or 5-shot task on the challenging OpenSARShip dataset. In particular, the recognition rate of CRCEPN comes up to 65% and 53% for 5-shot and 1-shot settings respectively, which are about 5% and 6% higher than those of the second best method, i.e., 2SCNet[22] and ACSRNet [29].

Table 6. Recognition performance comparison on the OpenSARShip dataset.

Method	Experimental setting	
	1-shot	**5-shot**
ProtoNet [12]	46.27 ± 0.60	58.44 ± 0.53
RelationNet [13]	42.80 ± 0.64	49.64 ± 0.54
TPN [14]	42.88 ± 0.68	52.43 ± 0.57
CAN [15]	41.04 ± 0.70	48.15 ± 0.68
TCAN [15]	42.33 ± 0.65	49.47 ± 0.61
MSAR [25]	45.04 ± 0.69	57.65 ± 0.57
2SCNet [22]	47.37 ± 0.72	59.66 ± 0.62
BSCapNet [23]	42.12 ± 0.64	48.96 ± 0.56
MTRLN [27]	46.98 ± 0.69	58.13 ± 0.71
ACSRNet [29]	46.46 ± 0.76	60.21 ± 0.65
MAML [30]	46.11 ± 0.28	59.41 ± 0.71
TPAN [28]	46.98 ± 0.69	58.13 ± 0.71
CRCEPN(ours)	$\mathbf{53.21 \pm 0.84}$	$\mathbf{65.15 \pm 0.60}$

Table 7 lists the confusion matrices of the average recognition accuracy of the proposed CRCEPN on the OpenSARShip dataset. Likewise, it indicates that the recognition performance of the proposed method on each target is properly balanced for both 1-shot and 5-shot tasks.

Table 7. Confusion matrix of the recognition accuracy of CRCEPN on OpenSARship dataset.

	Experimental setting					
	1-shot			**5-shot**		
	Carrier	**Container**	**Tanker**	**Carrier**	**Container**	**Tanker**
Carrier	**52.54**	24.76	22.76	**64.21**	17.58	17.42
Container	24.44	**52.07**	24.45	17.35	**65.32**	17.06
Tanker	23.02	23.17	**52.79**	18.44	17.10	**65.52**

From the above extensive experimental results on both the MSTAR dataset and the OpenSARShip dataset, we can state that the proposed method shows huge potential and significant superiority in solving the problem of few-shot SAR ATR compared with other state-of-the-art competitors.

3.5. Effectiveness Analysis

In this section, a series of experiments are designed and conducted to comprehensively evaluate the effectiveness of the proposed methods. First, ablation experiments are performed to investigate the efficacy of each component of the proposed CRCEPN, i.e., feature extraction network based on CRConv, enhanced prototypical network (EPN), and hybrid loss (HL) function. Then, enumeration experiments are carried out to examine the influence of different values of the hyperparameter λ on the recognition performance of the proposed method. Finally, two visualization methods, i.e., the

t-SNE [35] and the Grad-CAM[36], are leveraged to display respectively the feature distributions and feature maps of the proposed method and several competitors.

3.5.1. Ablation Experiment

In this section, we quantitatively investigate the effectiveness of the feature extraction network based on CRConv, the enhanced prototypical network (EPN), and hybrid loss (HL) function for improving the recogntion performance of the proposed CRCEPN. Below, ablation experiments are carried out on both the MSTAR dataset and the OpenSARShip dataset.

The basic architecture of the proposed method is similar to that of ProtoNet [12], so we use ProtoNet as the baseline. ProtoNet also consists of three components: feature extraction network based on standard convolution, prototypical network (PN), and cross-entropy loss (CL). By replacing one or more components of ProtoNet with corresponding counterparts of CRCEPN, we get different variants, as shown in the first column of Table 8, so as to evaluate the effectiveness of corresponding components of CRCEPN. Specifically, ProtoNet-CRConv represents a variant where the standard convolution in ProtoNet is replaced by CRConv, ProtoNet-EPN means PN is replaced by EPN, ProtoNet-EPN-HL means PN and CL are replaced by EPN and HL respectively, and so forth. Particularly, if all three components of ProtoNet are replaced correspondingly, it yields ProtoNet-CRConv-EPN-HL, which is just our proposed CRCEPN.

Table 8 gives the ablation experimental results on the MSTAR dataset under three different scenarios for both 1-shot and 5-shot tasks. From the results in Table 8, one can observe the following three points.

Table 8. Results of ablation experiments on the MSTAR dataset.

Method	Experimental setting and scenario					
	1-shot			5-shot		
	1	3	5	1	3	5
ProtoNet	71.51	69.90	66.76	80.28	82.22	78.59
ProtoNet-EPN	74.41	76.50	67.12	84.90	86.54	79.68
ProtoNet-EPN-HL	77.10	79.22	67.51	85.52	88.70	80.40
ProtoNet-CRConv	88.58	87.39	76.42	95.21	95.86	89.85
ProtoNet-CRConv-EPN	90.20	88.76	77.32	95.46	96.30	90.89
ProtoNet-CRConv-EPN-HL (CRCEPN)	**94.05**	**91.26**	**80.86**	**95.74**	**97.24**	**93.00**

1) The recognition rates of ProtoNet-CRConv are about 9%-18% higher than those of ProtoNet under each experimental scenario for either 1-shot or 5-shot task. Also, ProtoNet-CRConv-EPN-HL (i.e., CRCEPN) outperforms ProtoNet-EPN-HL by about 9%-17% in recognition rates. It demonstrates convincingly that the CRConv-based feature extraction network is capable of extracting more informative and discriminative features for few-shot SAR ART, thus improving the recognition performance of the proposed method.

2) By comparing ProtoNet-EPN with ProtoNet, and ProtoNet-CRConv-EPN with ProtoNet-CRConv, one can see that the recognition rates are increased by about 0.4%-6.6% and 0.3%-1.7%, respectively. So we can state that the proposed enhanced prototypical network (EPN) is better than traditional prototypical network (PN) for solving the problems of few-shot SAR target recognition.

3) Furthermore, the recognition rates of ProtoNet-EPN-HL exceed those of ProtoNet-EPN by 0.4%-2.7%, and ProtoNet-CRConv-EPN-HL (i.e., CRCEPN) surpasses ProtoNet-CRConv-EPN by 0.4%-3.9%, which indicate explicitly that the proposed hybrid loss (HL) function is beneficial to further enhance the recognition performance.

The ablation experimental results on the OpenSARShip dataset are listed in Table 9. As can be seen, similar results can be concluded for both 1-shot or 5-shot tasks. Broadly speaking, the CRConv-

based feature extraction network brings 2.5%-6% improvement for the recognition rates, the enhanced prototypical network (EPN) gives 0.6%-1.5%, and the hybrid loss (HL) function contributes about 0.1%-3%.

Table 9. Results of ablation experiments on the OpenSARShip dataset.

Method	Experimental setting	
	1-shot	5-shot
ProtoNet	46.27	58.44
ProtoNet-EPN	47.75	59.08
ProtoNet-EPN-HL	50.75	60.66
ProtoNet-CRConv	51.76	64.48
ProtoNet-CRConv-EPN	52.95	65.04
ProtoNet-CRConv-EPN-HL (CRCEPN)	**53.21**	**65.15**

From these above ablation experiments, we can state that each key component, that is, the feature extraction network based on CRConv, the enhanced prototypical network (EPN), and hybrid loss (HL) gives their own individual contribution to promoting the recognition performance of the proposed CRCEPN. And CRConv plays an explicitly dominant role in the performance improvement. These comprehensive evaluation on both the MSTAR dataset and the OpenSARShip dataset not only demonstrates the effectiveness of each key component of the the proposed CRCEPN, but also partly indicates that the proposed method is relatively robust and is able to handle diversified few-shot SAR data.

3.5.2. Hyperparameter λ Analysis

λ is key parameter of the proposed method that is used to balance the importance of the cross-entropy loss and the aggregation loss in the hybrid loss function. In this section, we conducts enumeration experiments in order to investigate the influence of different values of λ on the recognition performance of the proposed method. The values of λ are set to {0.05, 0.02, 0.01, 0.005, 0.002, 0.001}. All experiments are conducted on the MSTAR dataset under the first scenario for both 1-shot and 5-shot tasks.

Figure 7 displays the recognition rates of the proposed method with different values of λ. One can see from Figure 7 that for both few-shot tasks, as the decrease of the value of λ from 0.05 to 0.001, the recognition rate of the proposed method increases first and then decreases gradually. This indicates that an appropriate value of λ is important for the proposed method. Specifically, in our experiments, the proposed method obtains the best recognition performance when λ is set to 0.01. When λ is greater than 0.01, the recognition rate shows significant degradation, while as λ decreases to 0, it also declines but slightly and gradually. As a matter of fact, if the value of λ is greater than 0.05, the recognition performance of the proposed method will drop dramatically.
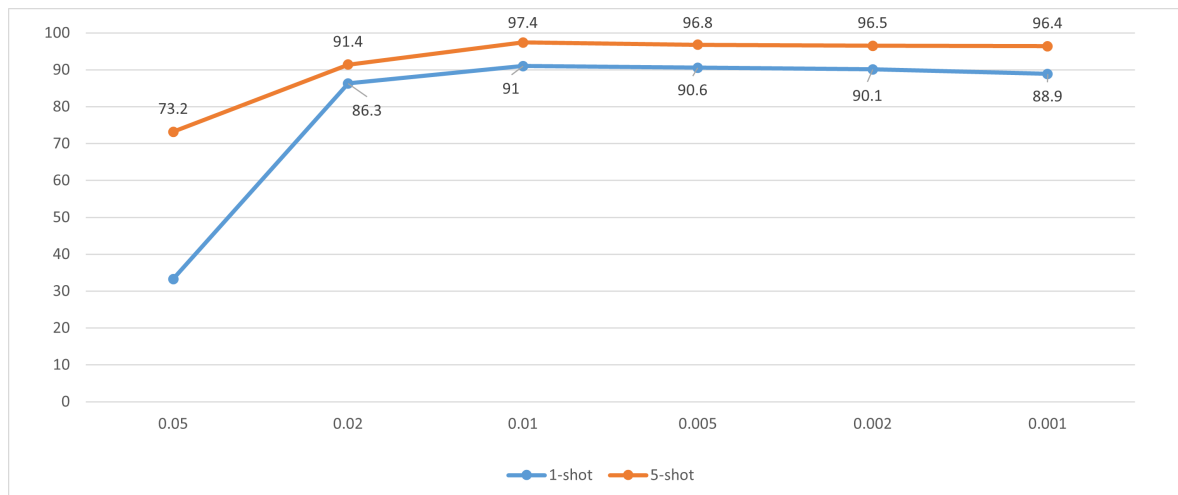
Figure 7. Recognition rates of the proposed method with different values of λ on MSTAR dataset.

As mentioned in section 2.3, the cross-entropy loss forces the features of different classes staying as far apart as possible, while the aggregation loss pulls features of the same class gathering as much as possible. Greater value of λ means the aggregation loss plays more role, while smaller value of λ means the cross-entropy loss dominates the hybrid loss function. For SAR ATR problems, the ultimate goal is to distinguish one class from another, so it is perfectly reasonable that the cross-entropy loss play a leading role while the aggregation loss act as a supplementary in the proposed method. That is to say, a reasonably smaller value of λ is necessary for the proposed method to obtain satisfactory recognition performance.

3.5.3. Feature Distribution Visualization

In this section, the performance of the proposed method is further demonstrated from the perspective of feature distribution. Specifically, we employ the t-SNE tool [35] in the Python package to visualize the distribution of original SAR images as well as the features extracted by ProtoNet, TPAN, and our proposed CRCEPN, respectively. For this purpose, SAR images of three targets in the MSTAR dataset for test, i.e., 2S1, BRDM2, and ZSU23/4, collected at a 17° depression angle, are selected for visualization.

Figure 8 (a) displays that the distribution of original SAR images is widely scattered, exhibiting severe within-class dispersion and inter-class overlap. It underscores the inherent challenge to distinguish between different targets based solely on the original SAR image data. By contrast, Figure 8 (b) shows that the features extracted by ProtoNet are broadly clustered, suggesting an improvement in the ability to separate different classes. However, the within-class dispersion is still relatively wide and a noticeable overlap remains between the features of BRDM2 and 2S1 (represented by purple and green dots, respectively). Then, Figure 8 (c) reveals that the feature space of TPAN is more separable than that of ProtoNet and the between-class overlap is reduced, suggesting an enhancement in the recognition performance of TPAN, which has been confirmed through the comparison experiments in section 3.3.
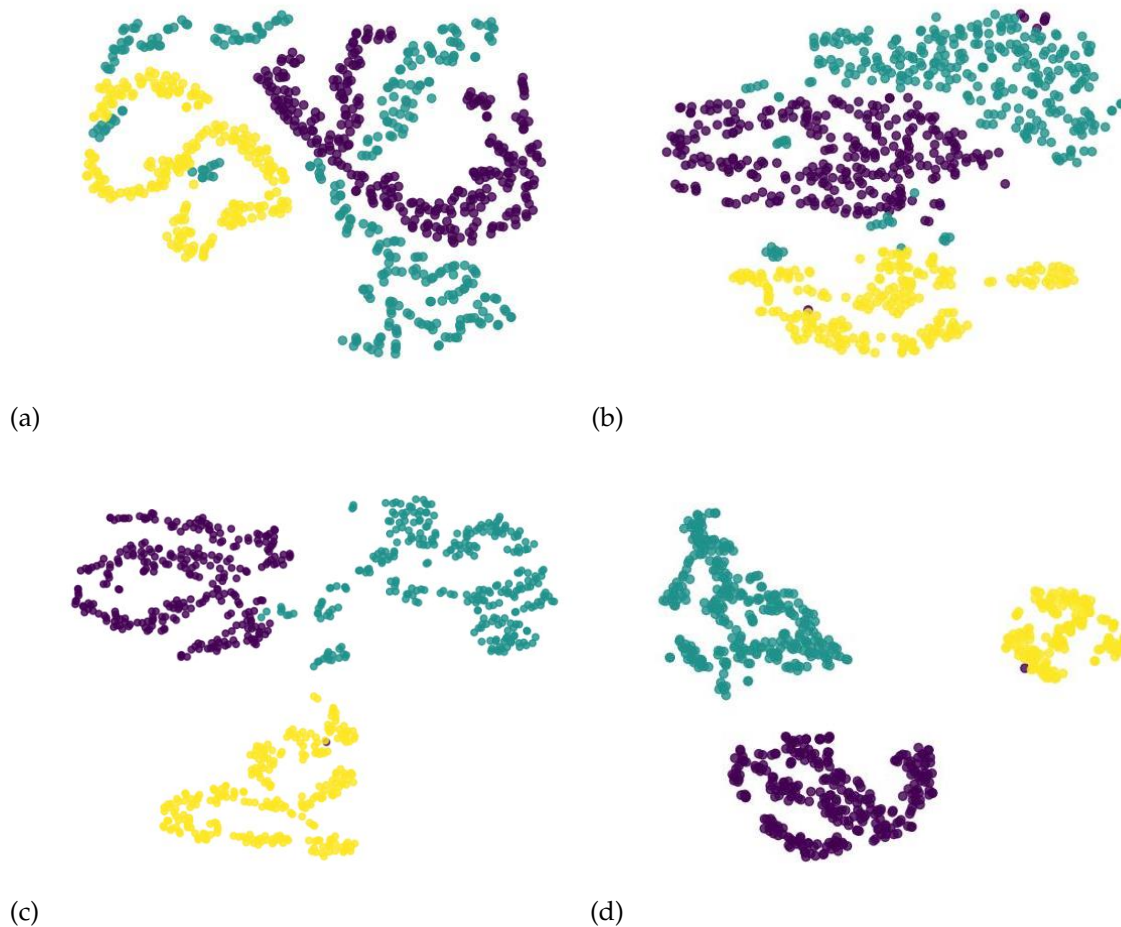
Figure 8. T-SNE visualization results on the MSTAR dataset. (a) original SAR images. (b) Features of ProtoNet. (c) Features of TPAN. (d) Features of CRCEPN.

Finally, Figure 8 (d) demonstrates that the proposed CRCEPN can yield a more distinguishable feature space compared to TPAN, characterized by better both inter-class separability and intra-class compactness. This visualization result of feature distribution is also a perfect illustration to what we discussed in section 2.3, that is, with joint optimization of both the cross-entropy loss and the aggregation loss, we train a robust network to obtain features with both inter-class separability and intra-class tightness. And this trait contributes directly to the superior recognition performance of the proposed CRCEPN for few-shot SAR ATR tasks, also being verified by the experimental results in section 3.3.

3.5.4. Feature Map Visualization

In this section, we continue to investigate the effectiveness and superiority of the proposed method intuitively. For this purpose, a visualization method, i.e., Grad-CAM [35] is utilized to display the extracted feature map of the proposed CRCEPN. ProtoNet and TPAN are still used as competitors. The visualization results are displayed in Figure 9, where the first row is the results of one image of T62 and the second row lists the results of one image of BTR60, both from the MSTAR dataset.

The results in Figure 9 (b) show that ProtoNet has relatively poor ability to focus on the target area in SAR images and the edge area of target is blurred by surrounding backgrounds, which as a result gives very limited recognition performance. Figure 9 (c) indicates TPAN can locate better on target area than ProtoNet and the edge of target area is clearly visible, thus improving the recognition performance. This may benefit from the elaborately-designed region-awareness-based feature extractor

in TPAN [28]. Nevertheless, one can also see from Figure 9 (c) on the second row that TPAN may possibly highlight some extra background area that is explicitly irrelevant to the target.
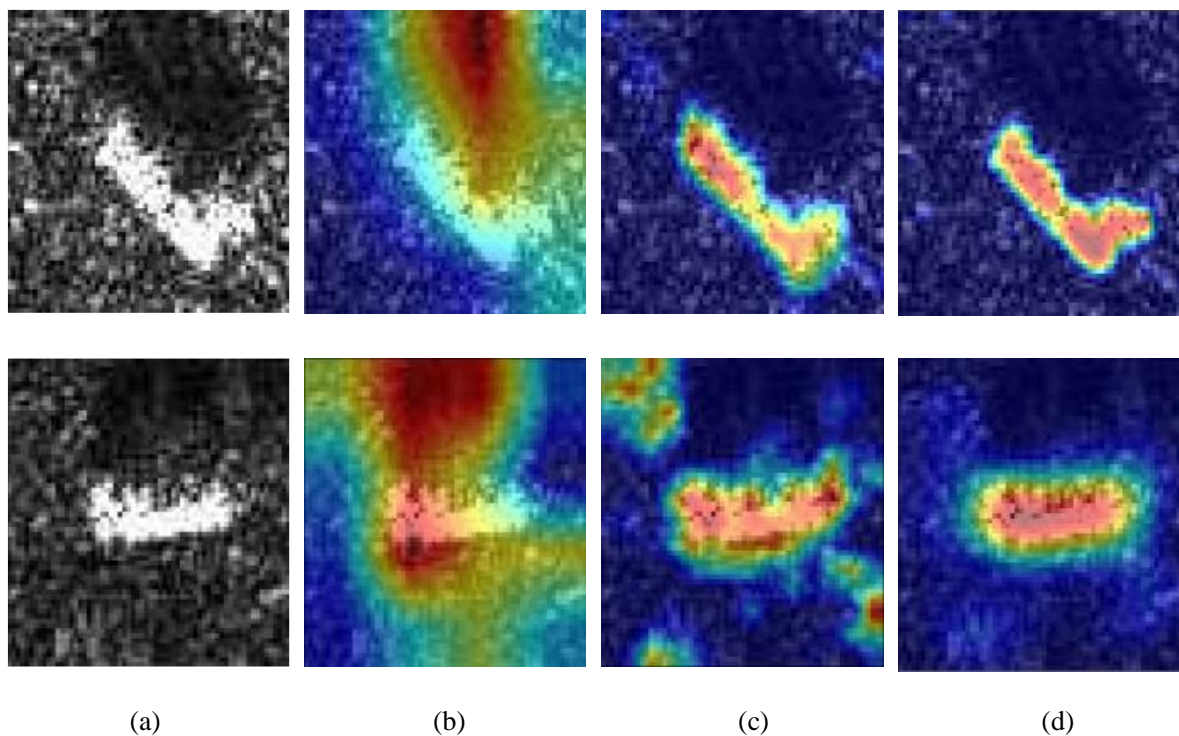

| (a) | (b) | (c) | (d) |

Figure 9. Feature map visualization, first row: T62, second row: BTR60. (a) original SAR image, (b) ProtoNet, (c) TPAN, (d) CRCEPN.

By comparison, we can see from Figure 9 (d) that the proposed CRCEPN can not only focus on the target area but also properly suppress the redundant background, which helps to obtain higher recognition performance than TPAN. This superiority mainly owns to the use of CRConv in CRCEPN which can adaptively adjust convolutional kernels as well as their receptive fields according to each SAR image's own characteristics and semantical similarity of spatial regions, thereby augmenting the capability to extract more informative and discriminative features that significantly aids the target classification. This is much valuable for few-shot SAR target recognition, where the scarcity of labeled images may degrade severely a network's representation ability and then how to make the most out of every piece of available information to amplify as much as possible the representation capacity of the network appears particularly important.

4. Conclusions

This paper proposes a new method called enhanced prototypical network with customized region-aware convolution (CRCEPN) to solve the problem of few-shot SAR target recognition where a scarcely few labeled samples are only available. The contributions of this paper includes three aspects. First, a feature extraction network with customized and region-aware convolutional kernels is developed to extract more informative and discriminative features for few-shot SAR target recognition, which can adapt better to diverse SAR images and remarkably improve the recognition performance of the proposed method. Second, an enhanced prototypical network is proposed to achieve more accurate and robust target identity prediction, which can effectively enhance the representation ability of class prototypes and in turn raise the classification accuracy especially for the situation of few-shot SAR target recognition. Third, a new loss function namely aggregation loss is proposed to pull features of the same class gathering together as much as possible and a hybrid loss is then designed to learn a feature space with both inter-class separability and intra-class tightness, which can further upgrade the

recognition performance of the proposed method. Extensive experiments on both the MSTAR dataset and the OpenSARship dataset demonstrate that the proposed method is superior to some state-of-the-art methods for few-shot SAR target recognition. In the future research, we will further exploit the few-shot SAR ATR algorithm in dynamic environments where the number of target categories continues to increase.

References

1. Curlander, J.C.; McDonough, R.N. *Synthetic aperture radar:Systems and Signal Processing*; Wiley: New York, NY, USA, 1991.
2. Henderson, F.M.; Lewis, A.J. *Principles and applications of imaging radar*; John Wiley and Sons: New York, NY, USA, 1998.
3. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine* **2013**, *1*, 6–43.
4. Dudgeon, D.E.; Lacoss, R.T. An overview of automatic target recognition. *The Lincoln Laboratory Journal* **1993**, *6*, 3–10.
5. Majumder, U.K.; Blasch, E.P.; Garren, D.A. *Deep learning for radar and communications automatic target recognition*; Norwood, MA, USA: Artech House, 2020.
6. Chen, S.; Wang, H.; Xu, F.; Jin, Y.Q. Target classification using the deep convolutional networks for SAR images. *IEEE Trans Geosci. Remote Sens.* **2016**, *54*, 4806–4817.
7. Kechagias-Stamatis, O.; Aouf, N. Fusing deep learning and sparse coding for SAR ATR. *IEEE Transactions on Aerospace and Electronic Systems* **2018**, *55*, 785–797.
8. Zhang, Y.; Guo, X.; Ren, H.; Li, L. Multi-view classification with semi-supervised learning for SAR target recognition. *Signal Processing* **2021**, *183*, 108030.
9. Pei, H.; Su, M.; Xu, G.; Xing, M.; Hong, W. Self-supervised Feature Representation for SAR Image Target Classification Using Contrastive Learning. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2023**, *16*, 9461–9476.
10. Zhang, Y.; others. SM-CNN: Separability Measure based CNN for SAR Target Recognition. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*.
11. Inkawhich, N. A Global Model Approach to Robust Few-Shot SAR Automatic Target Recognition. *IEEE Geosci. Remote Sens. Lett.* **2023**.
12. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. *Advances in neural information processing systems* **2017**, *30*.
13. Sung, F.; others. Learning to compare: Relation network for few-shot learning. *Proc. IEEE Conf. Compu. Vis. Pattern Recognit. (CVPR)* **2018**, pp. 1199–1208.
14. Liu, Y.; others. Learning to propagate labels: Transductive propagation network for few-shot learning. *arXiv:1805.10002* **2018**.
15. Hou, R.; Chang, H.; Ma, B.; Shan, S.; Chen, X. Cross attention network for few-shot classification. *Proc. Adv. Neural inf. Process. Syst.* **2019**, *32*.
16. Garcia, V.; Bruna, J. Few-shot learning with graph neural networks. *arXiv preprint arXiv:1711.04043* **2017**.
17. Kim, J.; Kim, T.; Kim, S.; Yoo, C.D. Edge-labeling graph neural network for few-shot learning. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.(CVPR)* **2019**, pp. 11–20.
18. Zhang, L.; Leng, X.; Feng, S.; Ma, X.; Ji, K.; Kuang, G.; Liu, L. Domain knowledge powered two-stream deep network for few-shot SAR vehicle recognition. *IEEE Transactions on Geoscience and Remote Sensing* **2021**, *60*, 1–15.

19. Wang, S.; Wang, Y.; Liu, H.; Sun, Y. Attribute-guided multi-scale prototypical network for few-shot SAR target classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2021**, *14*, 12224–12245.

20. Wang, L.; Bai, X.; Gong, C.; Zhou, F. Hybrid inference network for few-shot SAR automatic target recognition. *IEEE Transactions on Geoscience and Remote Sensing* **2021**, *59*, 9257–9269.

21. Wang, C.; Pei, J.; Yang, J.; Liu, X.; Huang, Y.; Mao, D. Recognition in label and discrimination in feature: A hierarchically designed lightweight method for limited data in sar atr. *IEEE Transactions on Geoscience and Remote Sensing* **2022**, *60*, 1–13.

22. Ren, H.; Yu, X.; Wang, X.; Liu, S.; Zou, L.; Wang, X. Siamese subspace classification network for few-shot sar automatic target recognition. IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2022, pp. 2634–2637.

23. Liu, S.; others. Bi-similarity prototypical network with capsule-based embedding for few-shot sar target recognition. Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), 2022, pp. 1015–1018.

24. Bi, H.; Liu, Z.; Deng, J.; Ji, Z.; Zhang, J. Contrastive Domain Adaptation-Based Sparse SAR Target Classification under Few-Shot Cases. *Remote Sensing* **2023**, *15*, 469.

25. Fu, K.; Zhang, T.; Zhang, Y.; Wang, Z.; Sun, X. Few-shot SAR target classification via metalearning. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14.

26. Yang, M.; Bai, X.; Wang, L.; Zhou, F. Mixed loss graph attention network for few-shot SAR target classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–13.

27. Wang, X.; others. Multi-Task Representation Learning Network for Few-Shot Sar Automatic Target Recognition. Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), 2022, pp. 2618–2621.

28. Yu, X.; others. Transductive Prototypical Attention Network for Few-shot SAR Target Recognition. Proc. IEEE Radar Conf. (RadarConf), 2023, pp. 1–5.

29. Ren, H.; others. Adaptive Convolutional Subspace Reasoning Network for Few-shot SAR Target Recognition. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*.

30. Liao, R.; Zhai, J.; Zhang, F. Optimization model based on attention mechanism for few-shot image classification. *Machine vision and Applications* **2024**, pp. 1–14.

31. Chen, J.; Wang, X.; Guo, Z.; Zhang, X.; Sun, J. Dynamic region-aware convolution. Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.(CVPR), 2021, pp. 8064–8073.

32. Han, Y.; Huang, G.; Song, S.; Yang, L.; Wang, H.; Wang, Y. Dynamic neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2021**, *44*, 7436–7456.

33. Ye, Z.; Xia, M.; Yi, R.; Zhang, J.; Lai, Y.K.; Huang, X.; Zhang, G.; Liu, Y.j. Audio-driven talking face video generation with dynamic convolution kernels. *IEEE Transactions on Multimedia* **2022**, *25*, 2033–2046.

34. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* **2014**.

35. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *Journal of machine learning research* **2008**, *9*.

36. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE international conference on computer vision* **2017**, pp. 618–626.