electronics

MDPI

Article

Achieving High-Accuracy Target Recognition Using Few ISAR Images via Multi-Prototype Network with Attention Mechanism

Linbo Zhang [1], Xiuting Zou [1], Shaofu Xu [1], Bowen Ma [1], Wenbin Lu [1,2], Zhenbin Lv [1,2] and Weiwen Zou [1,*]

[1] State Key Laboratory of Advanced Optical Communication Systems and Networks, Intelligent Microwave Lightwave Integration Innovation Center (imLic), Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; linbo525@sjtu.edu.cn (L.Z.); zouxiuting@sjtu.edu.cn (X.Z.); s.xu@sjtu.edu.cn (S.X.); bowenma@sjtu.edu.cn (B.M.); dawen_lu@sjtu.edu.cn (W.L.); lvzhenbin@sjtu.edu.cn (Z.L.)

[2] Shanghai Aerospace Electronic Technology Institute, Shanghai 201109, China

* Correspondence: wzou@sjtu.edu.cn

Abstract: Inverse synthetic aperture radar (ISAR) is a significant means of detection in space of non-cooperative targets, which means that the imaging geometry and associated parameters between the ISAR platform and the detection targets are unknown. In this way, a large number of ISAR images for high-accuracy target recognition are difficult to obtain. Recently, prototypical networks (PNs) have gained considerable attention as an effective method for few-shot learning. However, due to the specificity of the ISAR imaging mechanism, ISAR images often have unknown range and azimuth distortions, resulting in a poor imaging effect. Therefore, this condition poses a challenge for a PN to represent a class through a prototype. To address this issue, we use a multi-prototype network (MPN) with attention mechanism for ISAR image target recognition. The use of multiple prototypes eases the uncertainty associated with the fixed structure of a single prototype, enabling the capture of more comprehensive target information. Furthermore, to maximize the feature extraction capability of MPN for ISAR images, this method introduces the classical convolutional block attention module (CBAM) attentional mechanism, where CBAM generates attentional feature maps along channel and spatial dimensions to generate multiple robust prototypes. Experimental results demonstrate that this method outperforms state-of-the-art few-shot methods. In a four-class classification task, it achieved a target recognition accuracy of 95.08%, representing an improvement of 9.94–17.49% over several other few-shot approaches.

Keywords: inverse synthetic aperture radar; target recognition; few-shot learning; multi-prototype network

check for updates

Citation: Zhang, L.; Zou, X.; Xu, S.; Ma, B.; Lu, W.; Lv, Z.; Zou, W. Achieving High-Accuracy Target Recognition Using Few ISAR Images via Multi-Prototype Network with Attention Mechanism. *Electronics* **2024**, *13*, 4703. https://doi.org/10.3390/electronics13234703

Academic Editor: Elias Stathatos

Received: 24 October 2024
Revised: 25 November 2024
Accepted: 26 November 2024
Published: 28 November 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Inverse synthetic aperture radar (ISAR) images contain important information about long-distance moving targets, such as size, structure, and posture, providing rich features for radar target recognition [1,2]. Hence, target recognition using ISAR imaging plays a vital role in space situational awareness [3,4]. So far, under the condition of large samples of ISAR images, convolutional neural networks (CNNs) have made significant progress [5,6]. However, the observation time of moving non-cooperative targets is usually limited in most cases [7]. Thus, a large number of samples becomes difficult to collect. In this case, CNNs would be overfitting when the available ISAR images are insufficient, resulting in degraded recognition accuracy. Therefore, research on target recognition using only a few ISAR images remains challenging.

Few-shot learning has emerged as an effective method for recognition tasks with few samples [8]. Prototypical networks (PNs), as a representative few-shot learning method, have achieved great success due to their simplicity and effectiveness. In [9], Snell et al. initially proposed a method for representing a class prototype using an average vector, which enhanced the robustness and stability of class assignment for embedded queries.

Based on the PN, Wang et al. proposed a multiscale PN to enhance the discriminative ability of the prototype by fusing the features of different convolutional layers, thereby relieving the problem of high intra-class diversity and inter-class similarity for the images of synthetic aperture radar targets [10]. Cheng et al. proposed a siamese-prototype network with prototype self-calibration and inter-calibration, achieving more effective prototypes through information interaction between support and query samples [11]. Zhu et al. combined meta-learning and contrastive learning to build the hybrid contrastive PN, which enhances the discriminative ability of prototypes through prototype regularization methods [12]. The above method applies a single prototype approach to solve the few-shot problem, easing the stress of data collection.

ISAR images often exhibit unknown range and azimuth distortions as radar observation conditions change given their unique imaging mechanism [13]. In addition, the non-cooperative characteristic of ISAR targets indicates that the imaging geometry and associated parameters between the ISAR platform and the detection target are unknown. In this case, image distortions become difficult to correct using parameter estimation. This condition poses significant challenges for robust feature extraction and effective recognition [14]. While PN can achieve target recognition with few samples, its prototype structure cannot sufficiently deal with the complexities of ISAR images. The limited data per class and significant intra-class variations in ISAR images cause considerable fluctuations in feature representations for the same class [15]. In PN, class representations are derived by averaging global support set information. However, simple averaging will lead to inaccurate prototype construction when dealing with ISAR images [16]. Therefore, a single prototype cannot accurately describe a class, and it is impractical to merge all functionalities into one prototype. To address this problem, we introduce a multi-prototype network (MPN), which generates multiple prototypes for each class given the supporting data.

The main contribution of this study is the achievement of a high-accuracy target recognition scheme under the condition of only a few ISAR images using the MPN with attention mechanisms. This method employs multiple prototypes to represent each class, avoiding the limitations of traditional approaches that rely on obtaining a single prototype through mean calculation [9]. In addition, we improve the prototype generation process in MPN by introducing the CBAM attention mechanism. This module generates attention feature maps along channel and spatial dimensions, thereby maximizing the MPN's feature extraction capability for ISAR images. The key information from the ISAR image is dynamically selected by adaptively adjusting the weights of the MPN, thereby generating multiple robust prototypes. Experiments are conducted on the dataset constructed to validate the superiority of our scheme. The results demonstrate that our scheme achieves a target recognition accuracy of 95.08% in a four-way five-shot task, representing an improvement of 9.94–17.49% over several other few-shot approaches.

2. Related Work

2.1. Few-Shot Learning

Few-shot learning involves learning from existing datasets and leveraging the acquired knowledge to rapidly recognize new classes for subsequent few-shot learning tasks. Generally, few-shot learning methods can be categorized into two types: optimization-based methods and metric-learning-based methods [17]. Optimization-based methods focus on finding a set of excellent network initialization parameters to optimize the model and improve few-shot recognition accuracy. Finn et al. proposed model-agnostic meta-learning [18], which aims to train a set of initial parameters. This approach enables the model to rapidly adapt to new classification tasks with data distributions similar to the training data through one or more gradient adjustments with few samples. Metric-learning-based methods learn a distance or similarity function between image features in an embedding space, enabling the classification of images with limited samples [19]. Vinyals et al. proposed matching networks (MNs), which utilize a long short-term memory structure to automatically select support set features [20]. Then, they employ an attention mechanism

for enhanced prediction of query set samples. However, due to the complexity of the MN, Snell et al. introduced PN [9]. This method calculates the average support set features in the embedding space to obtain prototype representations for each class and then uses Euclidean distance for final classification. Sung et al. proposed the relation network (RN) to avoid manually designing fixed metric methods [21]. This network employs a CNN to automatically learn the distance metric between features. Optimization-based methods systematically learn a parameter initialization function, enabling training to start from a favorable initial point, achieving good classification results with only a few iterations. However, these methods often require fine-tuning for specific target problems or rely on pre-trained models [19]. Compared with optimization-based methods, metric-learning-based methods do not require fine-tuning of the network model, resulting in a simpler training process.

In addition to the aforementioned classification methods, there has been a surge of research focusing on few-shot learning approaches based on attention mechanisms in recent years. Li et al. explored the definition of relations in relation networks and proposed self-attention relation modules for feature and learning ability [22]. The introduction of the visual self-attention mechanism has the function of generating and purifying features, resulting in a target recognition accuracy of up to 89.75%. The KT method [23] included a feature adaptation module based on feed-forward attention to alleviate domain shift, which achieves competitive accuracy on four CDFSL datasets and four FSL datasets compared to state-of-the-art methods. The FSLBR method [24] used an attention-centered memory module in the field of behavior recognition to realize good classification results. The experimental results show that the proposed FSLBR model demonstrates improved performance under the few-shot learning framework compared to standard deep learning methods. In [25], Cheng et al. proposed a GNN framework with triple attention, namely, node self-attention, neighbor attention, and layer memory attention. The method solves the problems of over-fitting and over-smoothing and the authors demonstrated its superiority in few-shot fine-grained and semi-supervised classification tasks through extensive experiments.

2.2. Target Recognition in Radar Scenarios

Radar images are widely used in target recognition and play an important role in military and homeland security. In [26], Wang et al. proposed an RID high-resolution imaging framework to improve the imaging performance for maneuvering targets. The experimental results demonstrate that a hybrid transformer-based high-resolution network can be well applied to ISAR high-resolution imaging. Deng et al. encoded the 3D physical geometry of the target into an implicit neural network, which in turn constructs the 3D geometry based on the physical mechanism of ISAR imaging [27]. The robustness of the method is verified in experiments with low signal-to-noise ratio and noisy projected vectors. In [28], an implicit neural network representation framework for polarized ISAR super-resolution images was constructed based on the idea of implicit neural network representation. Experiments show that the method achieves better super-resolution performance, and the peak signal-to-noise ratio index can be improved by at least 0.93 dB. Recently, Wan et al. proposed a high-accuracy target recognition scheme based on a photonic analog-to-digital converter and CNN [29], achieving high-accuracy recognition for centimeter-level targets. Deng et al. proposed a high-resolution ISAR imaging scheme based on photonics reception [30]. This approach overcomes the limitation of detection range swath and achieves high-accuracy target recognition. While these methods have improved radar target recognition performance, they do not fully consider the problem of difficult image acquisition. Moreover, several studies have explored ISAR image target recognition under few-shot conditions. Li et al. presented a novel domain adaptive few-shot learning method from satellites to ISAR, enhancing transferability during the domain adaptation process [31]. Yuan et al. developed a module capable of extracting rich spatial relationships and effective representations between image regions, enabling robust recognition of maneuvering space targets in ISAR images under few-shot conditions. [32].

Jin et al. investigated the rapid learning and generalization of models through a few-shot framework, achieving test accuracy superior to traditional machine learning methods [33]. However, these methods struggle to effectively represent the internal features of ISAR images, resulting in undesired performance. We propose a novel MPN with an attention mechanism model scheme. The use of multiple prototypes alleviates the uncertainty associated with the fixed structure of a single prototype, allowing for a more comprehensive capture of target information. Furthermore, the introduction of the attention mechanism enables the network to automatically learn and generate more robust prototype representations tailored to the characteristics of the image. As a result, this approach can achieve higher accuracy in target recognition.

3. Methods

We propose an MPN model enhanced by an attention module to achieve high-accuracy target recognition under few-shot conditions, considering the characteristics of ISAR images. The incorporation of the attention module enables the construction of more suitable prototype representations for ISAR images acquired in real-world complex scenarios. The MPN can focus more on the features of the ISAR images by dynamically adjusting network parameters through an attention mechanism. This targeted feature emphasis further enhances the network's ability to distinguish slight differences among target classes.

For few-shot learning, two datasets are typically involved: a support set and a query set, denoted as S and Q, respectively. During network training, the support and query sets share network parameters, but they have no intersection. Generally, if the support set contains N classes, with K samples per class, then this task is referred to as an N-way K-shot task. The overall processing procedure of the MPN is illustrated in Figure 1. Prior to training the network, we initially collect ISAR images in the laboratory environment and divide them into support and query sets. Then, the support and query images are fed into the embedding network module to obtain corresponding feature vector representations in the embedding space. Subsequently, we input this feature vector into the attention module. We can obtain a single prototype representation for each class by initializing the attention mechanism parameters and performing a weighted sum over all support feature vectors. Therefore, we can obtain multiple prototypes for each class by using multiple weights. Finally, the distances between the query set feature vectors and all prototypes are measured to complete the target recognition task. Contrary to the traditional method of a single-prototype network, the proposed method in this study can capture the complex distribution of ISAR images and achieve higher target recognition accuracy.

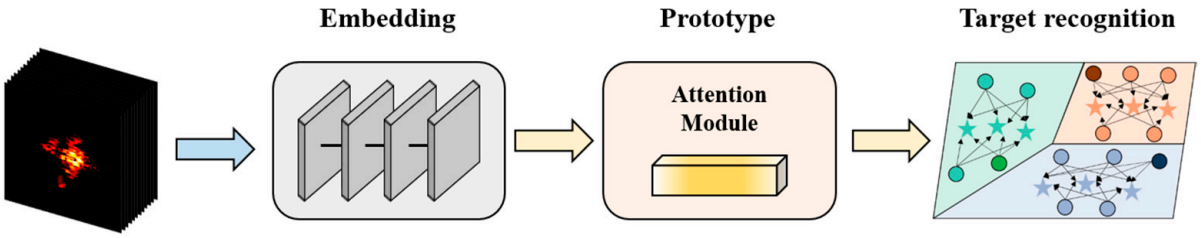


Figure 1. Schematic diagrams of multi-prototype network architectures.

3.1. ISAR Imaging System

Our scheme consists of three modules: a transmitter module, a receiver module, and a neural network. The first two modules are used for ISAR imaging, while the last module performs target recognition. Figure 2 shows the experimental setup of the ISAR imaging system. Furthermore, Table 1 lists the devices and instruments used in the ISAR imaging experiment, including their names and model numbers. In the transmitter, an arbitrary waveform generator (AWG, KEYSIGHT M8195A) generates a baseband linear frequency modulation (LFM) pulse signal with a bandwidth of 1–9 GHz. The pulse width and period of the signal are 3 and 4 μs, respectively. A microwave signal generator (MSG, ROHDE

& SCHWARZ SMA100B) generates a 32 GHz carrier signal. The baseband LFM pulse signal and the carrier signal are mixed by a mixer (Mixer, Marki MM1-1140H) and then filtered by a bandpass filter (BPF, TLBF-36G-8G-A) to produce a radio frequency (RF) LFM pulse signal with a frequency range of 32–40 GHz. After amplification by amplifier 1 (AMP1, CONNPHY CMP-18G-40G-3021-K), the RF LFM pulse signal is transmitted through a horn antenna (TA, INFOMW LB-28-25-C2-KF). The transmitted LFM pulse signal is reflected by a moving target fixed on a turntable rotating at approximately 2 r/s. In the receiver, we use a photonic analog-to-digital converter (PADC) with a sampling rate of 40 GSa/s as the receiver, allowing direct reception of wideband RF signals without complex down-conversion processes. The LFM echo signals received through a receiving antenna (RA, IN-FOMW LB-28-25-C2-KF) and amplified by another amplifier (AMP2, CONNPHY CLN-36G-40G-3030-K) are modulated on a 40 GS/s optical pulse train by a Mach–Zehnder modulator (MZM, EOSPACE AX-OMVS-40-PFA-PFA-LV). The optical pulse train is generated by an actively mode-locked laser (AMLL, Calmar, PSL-40-TT). A dual-output Mach–Zehnder modulator (DO-MZM, EOSPACE AX-1 × 2-OMSS-20-PFA-PFA-LV) demultiplexes the optical pulse train into two channels, each with a sampling rate of 20 GSa/s. This structure reduces the sampling rate requirement for the electronic backend in the PADC [34]. The demultiplexed optical pulse trains are converted to electrical signals by two photodetectors (PD, EOT ET-3500F) with a bandwidth of 10 GHz. Finally, the electrical signals are digitized by an oscilloscope (OSC, KEYSIGHT DSO-S 804A) with a sampling rate of 20 GSa/s and a bandwidth of 10 GHz. In digital signal processing (DSP), we use the back-projection (BP) algorithm [35] to process the digitized LFM echo pulses, ultimately obtaining the ISAR image. In the neural network, we employ an MPN with attention mechanisms to achieve high target recognition accuracy under the condition of only a few ISAR images.

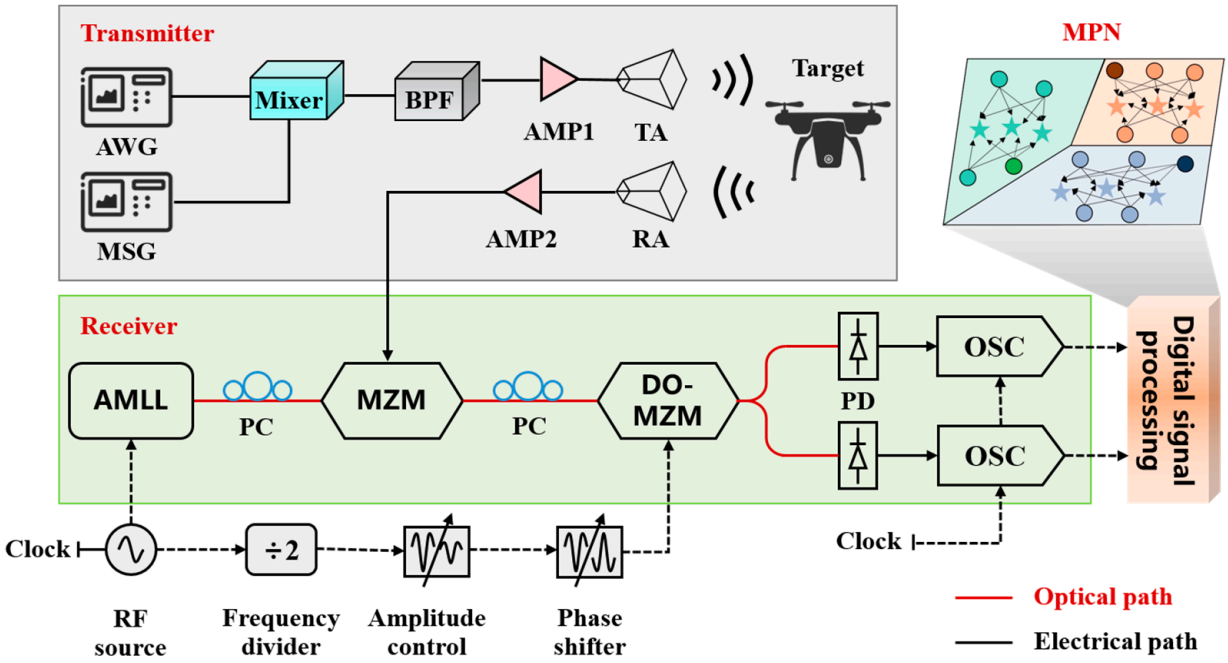


Figure 2. Schematic diagram of the proposed ISAR imaging system.

Figure 3 presents the time domain waveform and the instantaneous frequency of the LFM echo signals. After downsampling by the PADC, the frequency ranges from 0 to 8 GHz. As shown in the figure, the PADC system exhibits broadband reception capabilities, which lays the foundation for high-resolution ISAR imaging.

Table 1. The description of the device name and model number.

Module	Abbreviation	Full Name	Model Number
Transmitter	AWG	arbitrary waveform generator	KEYSIGHT M8195A
	MSG	microwave signal generator	ROHDE & SCHWARZ SMA100B
	BPF	band-pass filter	TLBF-36G-8G-A
	AMP	amplifier	CONNPHY CMP-18G-40G-3021-K
	TA	transmitting antenna	INFOMW LB-28-25-C2-KF
Receiver	RA	receiving antenna	INFOMW LB-28-25-C2-KF
	AMLL	actively mode-locked laser	Calmar PSL-40-TT
	PC	polarization controller	--
	MZM	Mach–Zehnder modulator	EOSPACE AX-OMVS-40-PFA-PFA-LV
	DO-MZM	dual-output Mach–Zehnder modulator	EOSPACE AX-1 × 2-OMSS-20-PFA-PFA-LV
	PD	photodetector	EOT ET-3500F
	OSC	oscilloscope	KEYSIGHT DSO-S 804A

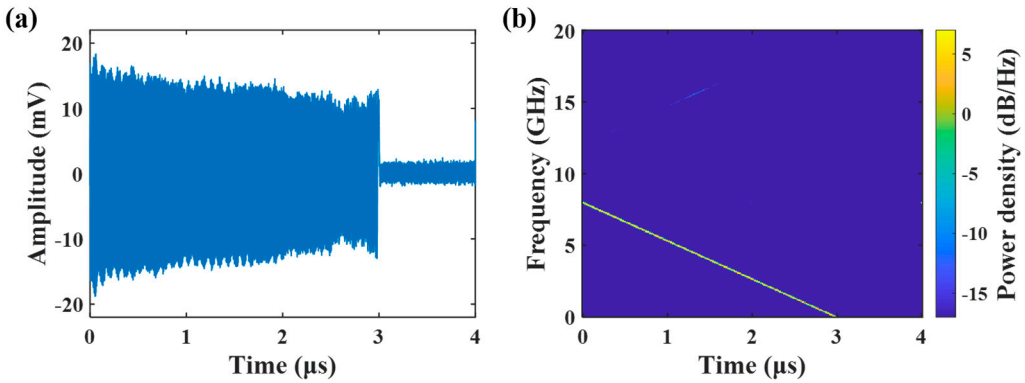


Figure 3. LFM echo signal after PADC at the bandwidth of 8 GHz (0–8 GHz). (**a**) Time domain and (**b**) time–frequency domain of the echo signal.

On the basis of the processing previously described, the transmitted RF LFM signal $S_t(t)$ can be expressed as follows [36]:

$$S_t(t) = rect(\frac{t}{T_P})\cos[2\pi f_c t + \frac{B}{T_P}\pi t^2] \tag{1}$$

where $rect(\cdot)$ is the rectangular function, T_P is the pulse width, f_c is the center frequency, and B is the bandwidth. The LFM echo signal $S_r(t)$ of a single scattering point received by radar can be expressed as follows [37]:

$$S_r(t) = \sigma_p rect(\frac{t-2R/c}{T_P})\cos[2\pi f_c\left(t-\frac{2R}{c}\right) + \frac{B}{T_P}\pi(t-\frac{2R}{c})^2] \tag{2}$$

where σ_p is the backscattering coefficient of the scattering point, R is the distance between the single point target and the radar, and c is the light velocity in vacuum. The exact value of R is supposed to be extracted through pulse compression. Equation (2) illustrates that a fixed time delay, $t_{delay} = \frac{2R}{c}$, exists between the transmitted signal and the echo signal. To maintain the coherence of the radar receiver, pulse compression is represented as the product of the transmitted signal and the echo signal, which can be expressed as follows [38]:

$$\begin{aligned} S_{ps}(t) &= S_r(t) \times S_t^*(-t) \\ &= \sigma_p T_P \frac{\sin\left[\pi k T_P\left(1-\frac{|t|}{T_P}\right)\right]}{\pi k T_P t} rect\left(\frac{t}{2T_P}\right)\cos(2\pi f_c t) \\ &\approx \sigma_p T_P sinc\left[\pi k T_P\left(t-\frac{2R}{c}\right)\right] rect\left(\frac{t-\frac{2R}{c}}{2T_P}\right) \end{aligned} \tag{3}$$

where $k = \frac{B}{T_P}$ is the frequency modulation slope. When $t = \frac{2R}{c}$, the expression in Equation (3) reaches its maximum value. The relative delay t_{delay} can be obtained by identifying the location of the peak. Thus, the distance R can be determined.

3.2. Multi-Prototype Network with Attention Mechanism

Our proposed MPN with attention mechanism consists of three modules: an embedding network module, a prototype learning network module, and a distance metric module, as shown in Figure 4. The embedding network module f_φ contains four convolution blocks and each block includes a convolution layer, a batch normalization layer, a ReLU layer, and a maximum pooling layer. The feature channel number of all four convolution layers is 64. The kernel size of the convolution layer (stride = 1) and maximum pooling layer (stride = 1) is 2 × 2. Thus, we learn a nonlinear mapping of the input into an embedding space using a neural network.

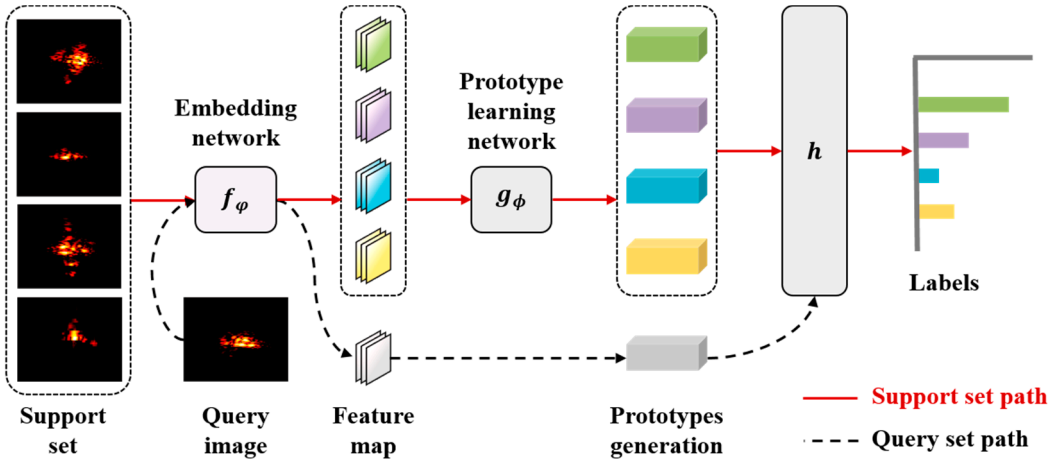


Figure 4. Architecture diagram of our proposed MPN with attention mechanism.

Contrary to traditional methods using a single prototype, we employ an attention mechanism to adaptively learn multiple prototypes. Specifically, the prototype learning module g_ϕ uses CBAM [39,40] as its key component, the architecture of which is depicted in Figure 5. First, a channel attention mechanism is applied to the input feature map to generate channel feature values. These channel feature values are then combined with the original feature map to produce a new feature map, denoted as F'. Next, a spatial attention mechanism is applied to F' to obtain the corresponding spatial attention feature values. Finally, F' is fused with the respective spatial attention feature values to obtain multiple prototype representations for each class. Compared to the squeeze-and-excitation network (SE-Net) model [41], CBAM extracts features from channel and spatial dimensions, making it more suitable for the complex scenarios presented by ISAR images.

Given an input vector F $\in \mathbb{R}^{C\times H\times W}$, CBAM sequentially derives a 1D channel attention map $M_c \in \mathbb{R}^{C\times 1\times 1}$ and a 2D spatial attention map $M_s \in \mathbb{R}^{1\times H\times W}$. The CBAM process can be expressed as follows:

$$\begin{aligned} F' &= M_c(F) \otimes F \\ F'' &= M_s(F') \otimes F' \end{aligned} \tag{4}$$

where $\otimes$ denotes element-wise multiplication. During the multiplication operation, the attention values are appropriately broadcast: channel attention values are broadcast along the spatial dimension, and vice versa. F'' represents the final output result. We generate a channel attention map to enhance the feature representations of different channels by leveraging the inter-channel feature relationships. Generally, the channel attention module can be expressed as follows:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))$$
$$= \sigma\left(W_1\left(W_0\left(F^c_{avg}\right)\right) + W_1(W_0(F^c_{max}))\right) \quad (5)$$

where σ denotes the sigmoid function, MLP represents the multi-layer perceptron, $W_0 \in \mathbb{R}^{C/r \times C}$, and $W_1 \in \mathbb{R}^{C \times C/r}$. Importantly, the MLP weights W_0 and W_1 are shared for both inputs, with W_0 following the ReLU activation function. The spatial attention module connects and convolves the output of the channel attention module through standard convolutional layers, which extract crucial information from different spatial locations. The spatial attention module can be expressed as follows:

$$M_s(F) = \sigma\left(f^{7\times7}([AvgPool(F); MaxPool(F)])\right)$$
$$= \sigma\left(f^{7\times7}\left(\left[F^s_{avg}; F^s_{max}\right]\right)\right) \quad (6)$$

where $f^{7\times7}$ represents a convolution operation with a kernel size of 7×7. For a general N-way K-shot few-shot classification problem, the support set contains K samples per class. Each sample is mapped into a c × h × w feature vector after passing through the embedding network module. We recombine the feature vectors of each class into a tensor $X \in \mathbb{R}^{K\times c\times h\times w}$. Then, we use an axis transformation operation to transform the tensor X into a new feature map $Y \in \mathbb{R}^{c\times(K\times h)\times w}$, where the height is $K \times h$ and the width is w. Finally, we input this new feature map into the CBAM module to obtain multiple prototypes for each class.

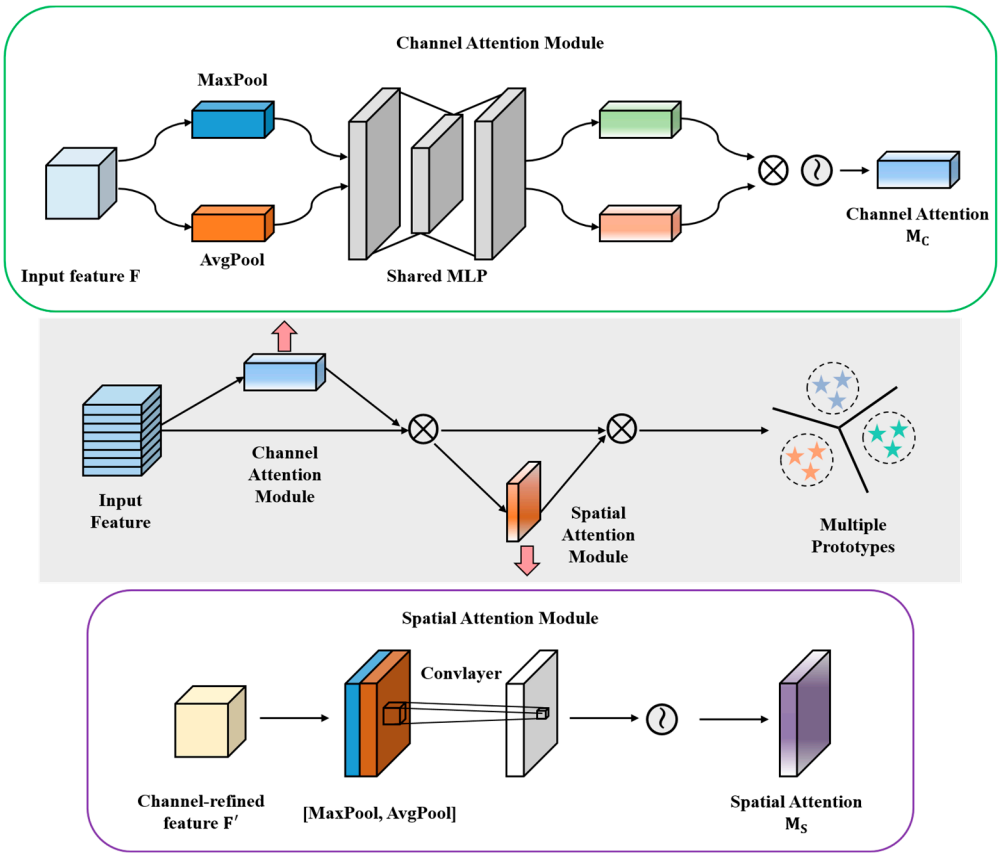


Figure 5. Architecture diagram of CBAM attention.

We obtain multiple prototypes $P = [p_1, p_2, \ldots, p_n]$ for each class and a feature vector $Q = [q_1, q_2, \ldots, q_m]$ for the query sample through the embedding network module and the prototype learning module. We use the Euclidean distance to measure the distance between Q and P and find the most similar class for the query sample Q. For each $q_j\big|_{j=1}^{m}$, we select the most similar prototype p_*^j from the prototypes in P. The distance metric process can be expressed as follows:

$$p_*^j = argmin\left[Euc\left(q_j, p_i\right)\right], \quad p_i \in P \tag{7}$$

where $argmin(\cdot)$ is used to represent the argument value at which a function achieves its minimum value within its domain, $Euc\left(q_j, p_i\right)$ represents the Euclidean distance between q_j and p_i, and p_*^j is the nearest neighbor of q_j in the set of prototypes P. Through Equation (7), we obtain the distance value between a single query sample and its nearest neighbor prototype p_*^j. Subsequently, we calculate the similarity between the query set Q and the nearest neighbor prototype p_*^j by summing the distances. It can be expressed as follows:

$$Sim(Q, P) = \sum_{j=1}^{m} Euc\left(q_j, p_*^j\right) \tag{8}$$

where $Sim(Q, P)$ represents the similarity between Q and P. This method can calculate all the similarities between Q and each class in the support set. Then, we classify Q based on the class with the highest similarity.

4. Results and Discussion

We carried out ISAR imaging experiments with four target models, namely, Target A (aircraft), Target B (helicopter), Target C (unmanned aerial vehicle, UAV), and Target D (bird). All targets were wrapped in tinfoil and surrounded by absorbing material to ensure the targets' scattering intensity without clutter interference. These targets were placed on the rotating platform 1.5 m away from the antenna, as shown in Figure 6a–d. The transmitting RF LFM pulse signal at the bandwidth of 8 GHz (32–40 GHz) is reflected by the rotating targets. In addition, the digital collecting duration of the ADC was 20 ms. The digital signal obtained consists of 5000 pulses, where each pulse is 4 μs. When the turntable rotates at a constant speed, the oscilloscope randomly acquires echo data from the target at a specific angle. Figure 6e–h show the ISAR imaging results of four different targets at a bandwidth of 8 GHz. The figures illustrate that the contour of the targets is evident from the images at a certain angle, consistent with the real target models. However, due to the limitations of ISAR imaging principles, the motion of the target model can affect imaging performance. When the target model is in an inappropriate motion posture, the imaging results will exhibit unknown range and azimuth angle distortions, as shown in Figure 6i–l. Feature extraction using ISAR images and high-accuracy target recognition becomes challenging due to the characteristics of ISAR images, such as unknown distortions and limited samples.

We collected ISAR images through experiments to build a target recognition dataset. Specifically, twenty classes of models were selected as targets, including seven aircraft models, three helicopter models, three UAV models, and seven bird models, with only 20 samples collected for each class. As the target rotated at a constant speed on the turntable, we randomly acquired data at a specific angle using an oscilloscope. Some preprocesses were conducted before the images were fed into the neural network. First, the raw ISAR images obtained have pixel values in linear units, and the accumulation time for a single frame is relatively long. The presence of scattering points results in peak intensities due to the coherent accumulation of zero-phase signals. Conversely, the absence of scatters leads to low noise levels owing to phase cancellation. This disparity produces a significant difference in pixel intensity values, increasing the risk of overfitting during neural network training. To address this, we performed a linear-to-logarithmic value conversion on the images, changing the pixel value unit to decibels. Then, we determined the threshold

point based on the echo signal-to-noise ratio and iterative experience. When the threshold point was set to 45 dB, it was 11 dB below the peak pixel value. Therefore, pixels with intensity values below 11 dB relative to the peak are considered to be non-target scattering points. This parameterization facilitates noise reduction, consequently improving the definition of target features. Finally, as each pixel represents a real-world distance of approximately 0.9 cm, we extracted a 101×101 pixel region centered on the image's centroid, which fully encompasses the target with about 50 cm in size. After converting the image to grayscale, the pixel values were normalized to an interval from 0 to 255. The neural network model uses sampled mini-batches called episodes [20] during training. Episodes are formed by randomly selecting a subset of classes from the training set to act as the support set, and then deciding on a subset of samples within each class of the remainder to serve as the query set. The dataset was divided into two parts: 14 classes for training and 6 classes for validating. Therefore, training episodes comprise 14 classes, where each episode randomly selects 4 classes of targets. There are 5 support samples and the remaining 15 samples from each class of target form the query set. Moreover, the data from six classes were used for validation, performing the same steps as the training episodes. Meanwhile, the initial learning rate was set to 0.001, and the Adam optimizer was used to adaptively optimize the network parameters to minimize the loss. The training of the neural network was carried out on Intel Core i7-10510U and NVIDIA GeForce MX250 GPUs. During training, input images are initially transformed into embedding vectors by an embedding network. Subsequently, the CBAM attention mechanism dynamically adjusts network weights to generate multiple prototypes from the embedding vector. CBAM prioritizes key information within the image, assigning higher weights, while simultaneously suppressing noise. Unlike traditional single-prototype networks that rely on averaging global support information, our method generates more robust prototypes by the CBAM attention mechanism. This enables the network to extract target features with greater precision, leading to higher target recognition accuracy. To further illustrate how the CBAM attention mechanism focuses on specific target features, we visualized the network's learning process, as shown in Figure 7. This visualization shows that CBAM enables the network to focus on key target areas while ignoring background noise, resulting in higher target recognition accuracy.

Figure 8 shows the training loss curves of the MN, PN, and MPN with the same convolutional layers. As training progresses, all training loss curves gradually decrease, approaching zero at the end of training, indicating no overfitting. This demonstrates that all three network models are sufficiently trained. In addition, the loss curve of our proposed MPN descends earlier and faster, ultimately converging to a lower value. This result demonstrates the superior ability of the MPN to effectively extract ISAR image features. Table 2 presents the results of target recognition accuracy and training time for the three models. In the previous presentation of this study, the MPN model utilizes three prototypes, meaning MPN represents MPN-3. To ensure fairness, the embedding network module of the three networks consists of four convolutional layers. MPN achieves a target recognition accuracy of 95.08%, surpassing the two other networks by 17.49% and 11.22%, respectively. Furthermore, while the RN automates the learning of prototypes based on the PN, it still utilizes a single prototype. Its target recognition accuracy surpasses that of the PN by 1.28% but falls short of the MPN by 9.94%. MPN's superior performance can be attributed to two key factors. First, PN and MN primarily focus on the global features of ISAR images. When faced with ISAR images exhibiting significant feature variations, MPN's attention mechanism allows it to focus on slight features, thereby extracting richer target information. Second, by using an attention mechanism, MPN adaptively constructs multiple prototypes for each class, enabling it to capture more comprehensive information compared to single-prototype methods. Furthermore, beyond recognition accuracy, MN's more complex network structure leads to longer testing times. Traditional PNs, which obtain prototypes by averaging global support information, have a simpler computational method and the shortest time among the three models. However, while the MPN outperforms the other two

methods in terms of target recognition accuracy, its testing time is not significantly longer. To further investigate the impact of the number of prototypes on target recognition accuracy, we conducted experiments by varying the number of prototypes. Here, MPN-2 indicates an MPN model with two prototypes, and so on. We provide statistical results with a 95% confidence interval. The results show that as the number of prototypes increases, target recognition accuracy gradually improves, and the network becomes more stable. When the number of prototypes reaches four, target recognition accuracy plateaus. Therefore, selecting three prototypes is considered a reasonable choice.

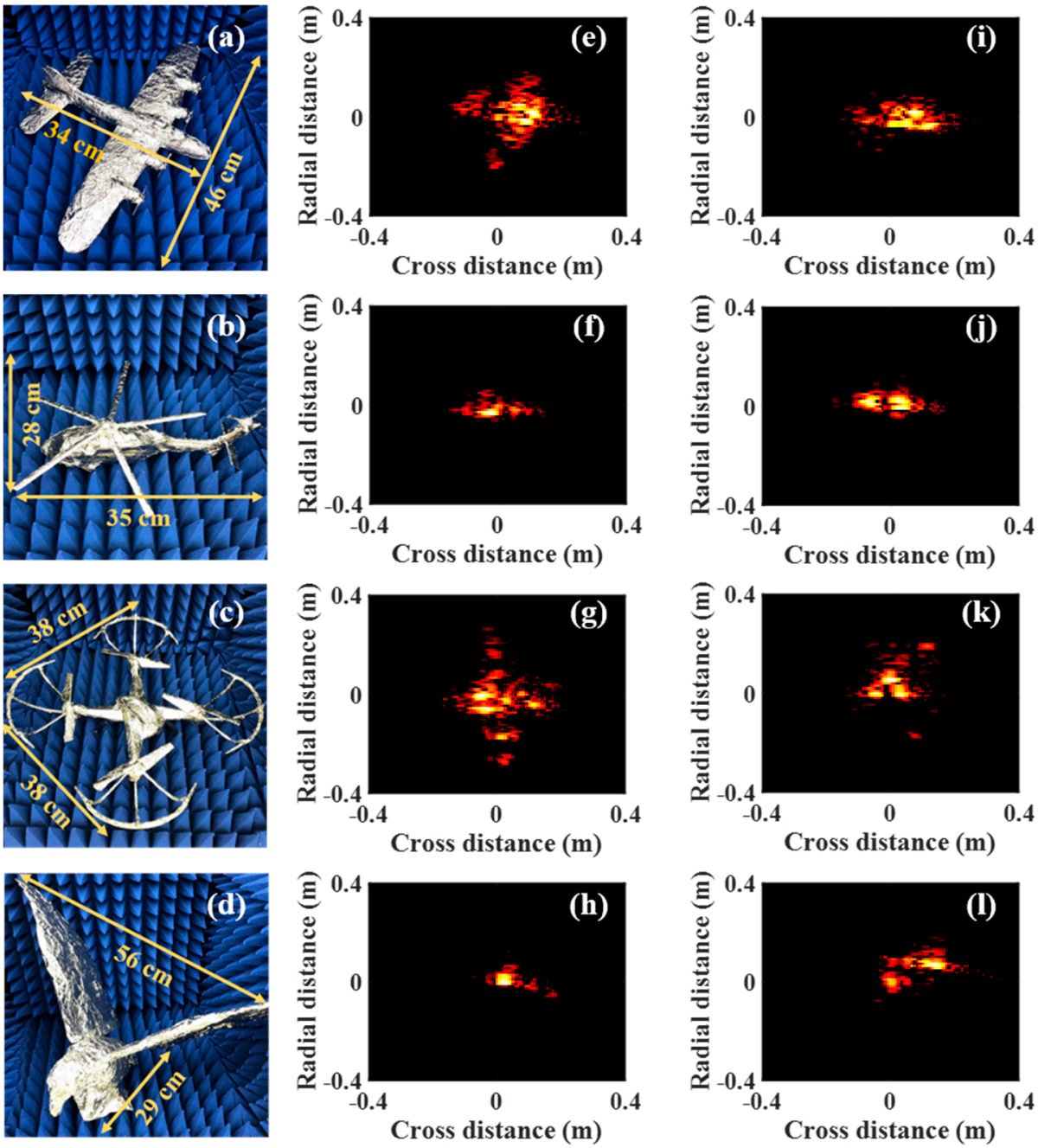


Figure 6. (**a**–**d**) Pictures of the four target models. (**e**–**h**) ISAR images of the four models with appropriate angle. (**i**–**l**) ISAR images of the four models with inappropriate angle.

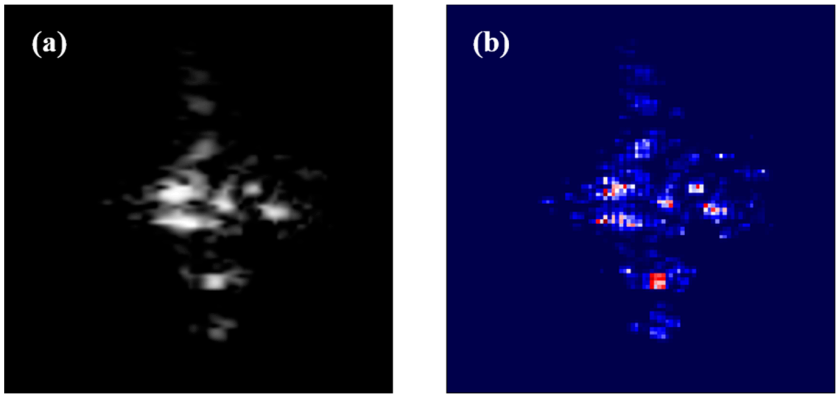


Figure 7. (**a**) Grayscale representation of an ISAR image. (**b**) ISAR image visualization.

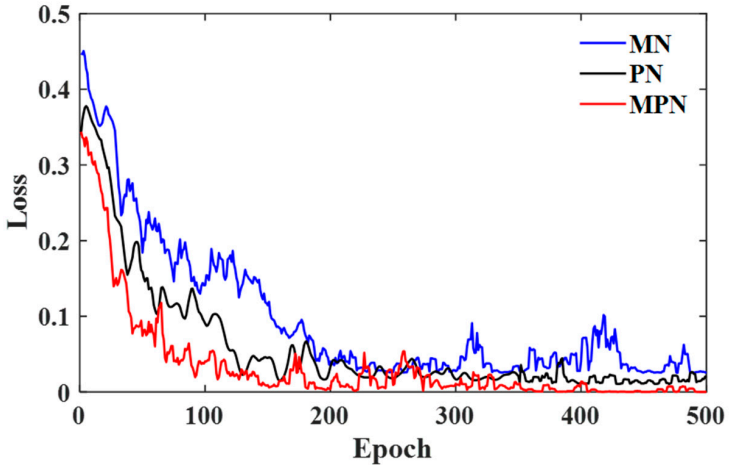


Figure 8. Loss curves for MN, PN, and MPN on ISAR images.

Table 2. Comparison results of target recognition accuracy and testing time on ISAR images with 95% confidence intervals.

Reference	Method	Embed.	Accuracy (%)	Time (ms)
[20]	MN	Conv-4	77.59 ±3.61	326
[9]	PN	Conv-4	83.86 ±2.79	184
[21]	RN	Conv-4	85.14 ±2.51	229
This work	MPN-2	Conv-4	91.53 ±2.34	210
This work	MPN-3	Conv-4	95.08 ±2.02	236
This work	MPN-4	Conv-4	95.56 ±1. 95	243

Furthermore, we also investigated the impact of the number of attention mechanism layers on target recognition accuracy, the results of which are presented in Table 3. In this experiment, the number of prototypes was set to two. As the depth of the attention network increases, so too does the model's computational complexity. Concurrently, a trend of increasing target recognition accuracy is observed. However, when the number of layers in both the channel attention module and the spatial attention module is set to two, a slight decrease in target recognition accuracy occurs. This may be due to slight overfitting in the network. Furthermore, when the number of prototypes is three, increasing the network depth does not result in a significant change in target recognition accuracy. Therefore, to strike a balance between computational complexity and target recognition accuracy, setting the number of layers in both sub-channels to one constitutes a favorable network configuration.

Table 3. The impact of attention layers on target recognition accuracy.

Channel Attention Module Layers	Spatial Attention Module Layers	Accuracy (%)
1	1	91.53
1	2	93.36
2	1	93.28
2	2	92.19

To understand the validity of the results, we visualized the features captured by the three models using the t-SNE algorithm [42], as shown in Figure 9. It can be seen that the network degradation effect of MN, PN, and RN is evident, decreasing the target recognition accuracy. Moreover, our proposed MPN model has clearer feature boundaries and better intra-class clustering than the other three models. Therefore, the proposed MPN model can reduce the intra-class distance to a certain extent while increasing the inter-class distance. This condition is largely attributed to the prototypes constructed. Traditional PN models construct a single prototype by obtaining the mean of the features through a fixed mechanism, which cannot sufficiently represent a complex class distribution. However, MPN automatically learns multiple prototypes suitable for the current model, achieving higher target recognition accuracy under few-shot conditions.

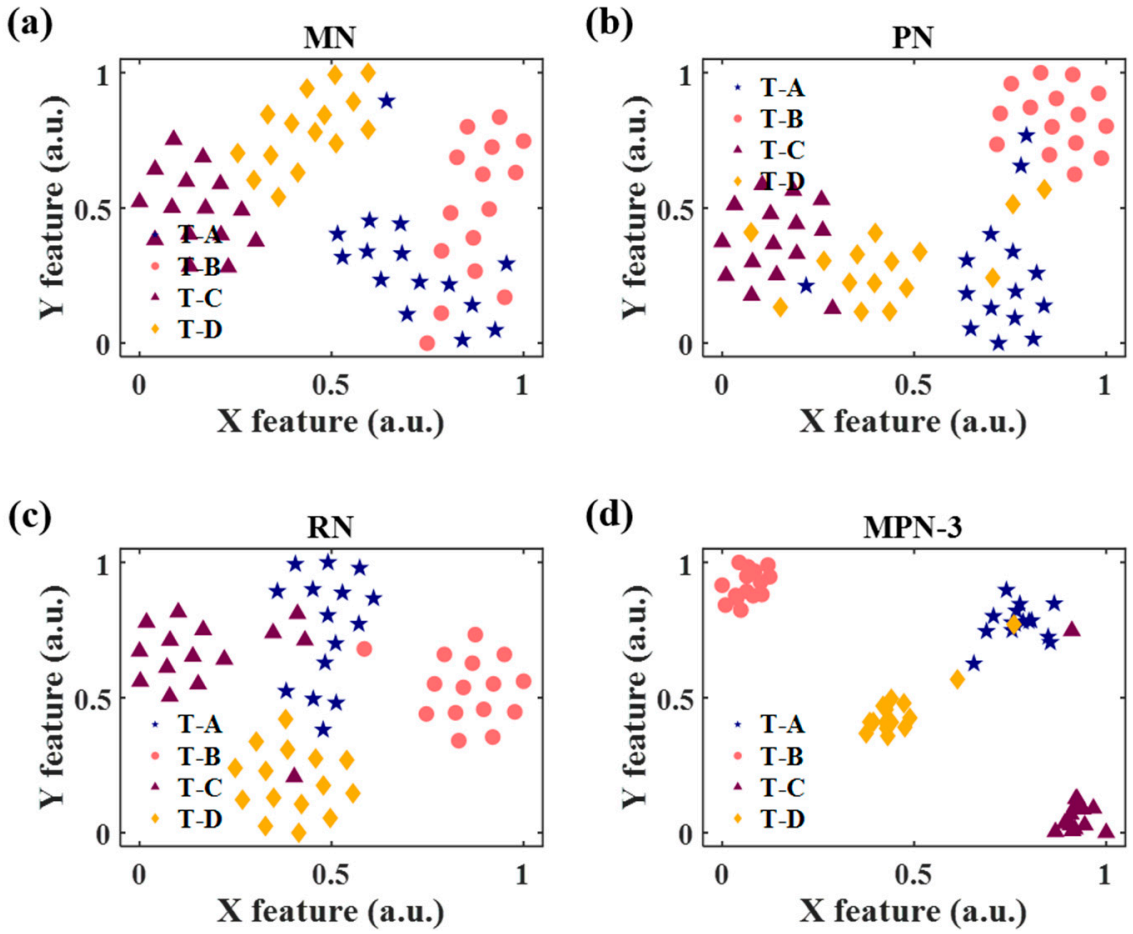


Figure 9. t−SNE feature visualization of the (**a**) MN, (**b**) PN, (**c**) RN, and (**d**) MPN-3 on ISAR images.

The MPN addresses several challenges faced by ISAR image target recognition in practical applications. First, MPN achieves higher target recognition accuracy in recognizing ISAR images. Contrary to traditional single-prototype methods [9], the multi-prototype approach effectively extracts feature information from targets. Second, it alleviates the

difficulty of obtaining large training datasets in real-world scenarios. Compared with CNN-based target recognition methods [26,27], which require hundreds of samples per class, MPN achieves a 95.08% target recognition accuracy with only 20 samples per class. However, despite achieving a relatively high accuracy in target recognition, there is still room for improvement in this approach. Firstly, the difficulty in data acquisition and the lengthy data processing time pose significant challenges. Under laboratory conditions, obtaining the raw waveform for an ISAR image takes approximately 2 min, and transforming this raw waveform into an ISAR image requires about 1 min. To effectively address these issues, it is necessary to develop data parallel processing algorithms suitable for the characteristics of ISAR imaging systems. Concurrently, the application of optical computing methods to ISAR imaging can also offer reduced data processing latency. Secondly, all the data used in this study are collected in the laboratory environment. If applied to real-world scenarios, several foreseeable issues arise. Due to the long range of target detection, the radar echo power is relatively weak. It is difficult for the radar to capture the target feature, leading to low target recognition accuracy. Additionally, the presence of various sources of environmental interference results in mixed signals, which can adversely affect both target detection and signal processing. In the future, to address these issues, further exploration of more robust and stable neural network models for target recognition based on ISAR images is necessary. Thirdly, MPNs require learning multiple prototypes per class, significantly increasing model complexity. Simultaneously, there is no universally perfect method for determining the ideal number of prototypes. It requires task-specific adjustments and experimentation, increasing the difficulty of model tuning.

5. Conclusions

We have achieved a high-accuracy target recognition scheme using an MPN with attention mechanism under the condition of only a few ISAR images. Traditional single prototype network models use a fixed mechanism to obtain prototypes, making it difficult to fully capture the complete target features. Our method introduces an attention mechanism to dynamically adjust network weights, generating robust multiple prototypes. To validate the effectiveness of this approach, we construct a few-shot target recognition dataset using ISAR images. Experimental results demonstrate that our proposed method outperforms state-of-the-art few-shot methods, achieving a target recognition accuracy of 95.08%, surpassing existing few-shot methods by 9.94–17.49%. However, the relatively long ISAR imaging time is tricky. Future work will explore parallel processing or optical computing schemes to reduce processing time. Furthermore, increased target distance and various environmental interferences can weaken target features, reducing the accuracy of target recognition. To address these challenges, future work will investigate the incorporation of additional attention mechanisms to optimize the network architecture with the aim of developing a more robust and stable target recognition neural network model. The performance of this improved model will be evaluated using datasets collected from real-world scenarios.

Author Contributions: Conceptualization, L.Z., X.Z. and S.X.; methodology, L.Z., X.Z. and S.X.; software, L.Z.; validation, L.Z., X.Z., S.X., B.M., W.L., Z.L. and W.Z.; formal analysis, L.Z., X.Z. and S.X.; investigation, L.Z.; resources, L.Z.; data curation, L.Z.; writing—original draft preparation, L.Z.; writing—review and editing, L.Z., X.Z. and S.X.; visualization, L.Z., X.Z. and S.X.; supervision, B.M., W.L., Z.L. and W.Z.; project administration, S.X. and W.Z.; funding acquisition, S.X. and W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the National Natural Science Foundation of China, grant numbers T2225023 and 62205203.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

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

References

1. Lee, S.J.; Lee, M.J.; Kim, K.T.; Bae, J.H. Classification of ISAR images using variable cross-range resolutions. *IEEE Trans. Aerosp. Electron. Syst.* **2018**, *54*, 2291–2303. [CrossRef]
2. Benedek, C.; Martorella, M. Moving target analysis in ISAR image sequences with a multiframe marked point process model. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 2234–2246. [CrossRef]
3. Zhang, S.; Liu, Y.; Li, X. Fast sparse aperture ISAR autofocusing and imaging via ADMM based sparse Bayesian learning. *IEEE Trans. Image Process.* **2019**, *29*, 3213–3226. [CrossRef] [PubMed]
4. Peng, S.; Li, S.; Xue, X.; Xiao, X.; Wu, D.; Zheng, X.; Zhou, B. High-resolution W-band ISAR imaging system utilizing a logic-operation-based photonic digital-to-analog converter. *Opt. Express* **2018**, *26*, 1978–1987. [CrossRef]
5. Zhu, Q.; Zhong, Y.; Wu, S.; Zhang, L.; Li, D. Scene classification based on the sparse homogeneous–heterogeneous topic feature model. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2689–2703. [CrossRef]
6. Zhao, B.; Huang, B.; Zhong, Y. Transfer learning with fully pretrained deep convolution networks for land-use classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1436–1440. [CrossRef]
7. Huang, X.; Ding, J.; Xu, Z. Real-time super-resolution ISAR imaging using unsupervised learning. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
8. Triantafillou, E.; Zemel, R.; Urtasun, R. Few-shot learning through an information retrieval lens. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 2252–2262.
9. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 4077–4087.
10. Wang, S.; Wang, Y.; Liu, H.; Sun, Y. Attribute-guided multi-scale prototypical network for few-shot SAR target classification. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2021**, *14*, 12224–12245. [CrossRef]
11. Cheng, G.; Cai, L.; Lang, C.; Yao, X.; Chen, J.; Guo, L.; Han, J. SPNet: Siamese-prototype network for few-shot remote sensing image scene classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–11. [CrossRef]
12. Zhu, J.; Yang, K.; Guan, N.; Yi, X.; Qiu, C. HCPNet: Learning discriminative prototypes for few-shot remote sensing image scene classification. *Int. J. Appl. Earth Observ. Geoinf.* **2023**, *123*, 103447. [CrossRef]
13. Xue, R.; Bai, X.; Cao, X.; Zhou, F. Sequential ISAR target classification based on hybrid transformer. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [CrossRef]
14. Wehner, D.R. *High Resolution Radar*, 2nd ed; Artech House: Boston, MA, USA, 1995.
15. Liu, Y.; Zhang, X.; Zhang, S.; He, X. Part-aware prototype network for few-shot semantic segmentation. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; pp. 142–158.
16. Wang, W.; Duan, L.; En, Q.; Zhang, B.; Liang, F. TPSN: Transformer-based multi-Prototype Search Network for few-shot semantic segmentation. *Comput. Electr. Eng.* **2022**, *103*, 108326. [CrossRef]
17. Wen, W.; Liu, Y.; Ouyang, C.; Lin, Q.; Chung, T. Enhanced prototypical network for few-shot relation extraction. *Inf. Process. Manag.* **2021**, *58*, 102596. [CrossRef]
18. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the 34th International Conference on Machine Learning (ICML 2017), Sydney, Australia, 6 July 2017; pp. 1126–1135.
19. Wu, Y.; Li, Y.; Zhao, T.; Zhang, L.; Wei, B.; Liu, J.; Zheng, Q. Improved prototypical network for active few-shot learning. *Pattern Recognit. Lett.* **2023**, *172*, 188–194. [CrossRef]
20. Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D. Matching networks for one shot learning. In Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 3630–3638.
21. Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Torr, P.H.; 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, 18–23 June 2018; pp. 1199–1208.
22. Li, Z.; Hu, Z.; Luo, W.; Hu, X. SaberNet: Self-attention based effective relation network for few-shot learning. *Pattern Recognit.* **2023**, *133*, 109024. [CrossRef]
23. Li, P.; Liu, F.; Jiao, L.; Li, S.; Li, L.; Liu, X.; Huang, X. Knowledge transduction for cross-domain few-shot learning. *Pattern Recognit.* **2023**, *141*, 109652. [CrossRef]
24. Mahalakshmi, V.; Sandhu, M.; Shabaz, M.; Keshta, I.; Prasad, K.D.V.; Kuzieva, N.; Soni, M. Few-shot learning-based human behavior recognition model. *Comput. Hum. Behav.* **2024**, *151*, 108038. [CrossRef]
25. Cheng, H.; Zhou, J.T.; Tay, W.P.; Wen, B. Graph neural networks with triple attention for few-shot learning. *IEEE Trans. Multimedia* **2023**, *25*, 8225–8239. [CrossRef]
26. Wang, H.; Chen, Y.; Zhang, Y.; Luo, Y.; Zhang, Q. High-resolution ISAR imaging method for maneuvering targets based on hybrid transformer. *IEEE Trans. Antennas Propag.* **2023**, *71*, 8250–8265. [CrossRef]
27. Deng, J.; Xie, P.; Zhang, L.; Cong, Y. ISAR-NeRF: Neural radiance fields for 3D imaging of space target from multi-view ISAR images. *IEEE Sensors J.* **2024**, *24*, 11705–11722. [CrossRef]
28. Li, M.; Deng, J.; Xiao, S.; Chen, S. Semi-supervised implicit neural representation for polarimetric ISAR image super-resolution. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
29. Wan, J.; Xu, S.; Zou, W. High-accuracy automatic target recognition scheme based on a photonic analog-to-digital converter and a convolutional neural network. *Opt. Lett.* **2020**, *45*, 6855–6858. [CrossRef] [PubMed]

30. Deng, A.; Qian, N.; Hua, S.; Wan, J.; Lv, Z.; Zou, W. High-resolution ISAR imaging based on photonic receiving for high-accuracy automatic target recognition. *Opt. Express* **2022**, *30*, 20580–20588. [CrossRef] [PubMed]
31. Li, B.; Yao, Y.; Wang, Q. Domain adaptive few-shot learning for ISAR aircraft recognition with transferred attention and weighting importance. *Electronics* **2023**, *12*, 2909. [CrossRef]
32. Yuan, H.; Li, H.; Zhang, Y.; Li, M.; Wei, C. Few-shot classification for ISAR images of space targets by complexvalued patch graph transformer. *IEEE Trans. Aerosp. Electron. Syst.* **2024**, *60*, 4896–4909. [CrossRef]
33. Jin, J.; Wang, F. Meta learning-based approach for few-shot target recognition in ISAR images. In Proceedings of the IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023; pp. 6438–6441.
34. Yang, G.; Zou, W.; Yu, L.; Chen, J. Influence of the sampling clock pulse shape mismatch on channel-interleaved photonic analog-to-digital conversion. *Opt. Lett.* **2018**, *43*, 3530–3533. [CrossRef]
35. Gao, B.; Zhang, F.; Sun, G.; Xiang, Y.; Pan, S. Microwave photonic MIMO radar for high-resolution imaging. *J. Light. Technol.* **2021**, *39*, 7726–7733. [CrossRef]
36. Jia, W.; Cao, Y.; Zhang, S.; Wang, W. Detecting high-speed maneuvering targets by exploiting range-doppler relationship for LFM radar. *IEEE Trans. Veh. Technol.* **2021**, *70*, 2209–2218. [CrossRef]
37. Liu, Y.; Deng, A.; Hua, S.; Xu, S.; Zou, W. Photonic ADC-based scheme for joint wireless communication and radar by adopting a broadband OFDM shared signal. *Opt. Lett.* **2022**, *47*, 5421–5424. [CrossRef]
38. Liu, Z.; Guo, Y.; Zhao, X.; Cheng, J.; Zheng, K.; He, C. Development of wideband pulse compression surface wave EMAT. *IEEE Sens. J.* **2024**, *24*, 5073–5085. [CrossRef]
39. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
40. Liao, L.; Wu, S.; Song, C.; Fu, J. PH-CBAM: A parallel hybrid CBAM network with multi-feature extraction for facial expression recognition. *Electronics* **2024**, *13*, 3149. [CrossRef]
41. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
42. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.

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.