


Article

Multi-Aspect SAR Target Recognition Based on Non-Local and Contrastive Learning

Xiao Zhou ^{1,2}, Siyuan Li ^{1,2,3}, Zongxu Pan ^{1,2,3} , Guangyao Zhou ^{1,2,*} and Yuxin Hu ^{1,2,3}

¹ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; zhouxiao@aircas.ac.cn (X.Z.); lisiyuan20@mails.ucas.ac.cn (S.L.); zxfan@mail.ie.ac.cn (Z.P.); yxhu@mail.ie.ac.cn (Y.H.)

² Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing 100190, China

³ School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: zhougy@aircas.ac.cn

Abstract: Synthetic aperture radar (SAR) automatic target recognition (ATR) has been widely applied in multiple fields. However, the special imaging mechanism of SAR results in different visual features of the same target at different azimuth angles, so single-aspect SAR target recognition has the limitation of observing the target from a single perspective. Multi-aspect SAR target recognition technology can overcome this limitation by utilizing information from different azimuths and effectively improve target recognition performance. Considering the order dependency and data limitation of existing methods, this paper proposes a multi-aspect SAR recognition method based on Non-Local, which applies a self-attention calculation to feature maps to learn the correlation between multi-aspect SAR images. Meanwhile, in order to improve the generalization ability of the proposed method under limited data, a network based on contrastive learning was designed to pre-train the feature extraction part of the whole network. The experimental results using the MSTAR dataset show that the proposed method has excellent recognition accuracy and good robustness.

Keywords: synthetic aperture radar (SAR); automatic target recognition (ATR); multiview; Non-Local; contrastive learning

MSC: 68T01; 68T07; 68T45



Citation: Zhou, X.; Li, S.; Pan, Z.; Zhou, G.; Hu, Y. Multi-Aspect SAR Target Recognition Based on Non-Local and Contrastive Learning. *Mathematics* **2023**, *11*, 2690. <https://doi.org/10.3390/math11122690>

Academic Editor: Catalin Stoean

Received: 9 May 2023

Revised: 4 June 2023

Accepted: 12 June 2023

Published: 13 June 2023



Copyright: © 2023 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

Synthetic aperture radar (SAR) is a type of high-resolution coherent imaging radar that is not affected by lighting and weather conditions, enabling all-day and all-weather ground detection [1]. SAR is widely used in various fields [2] due to its advantages of long-range and high-resolution imaging. In recent years, with the maturity of SAR technology and the enhancement of data acquisition capabilities, it has become an important problem in SAR applications to extract useful information rapidly from massive high-resolution SAR image data. Nowadays, automatic target recognition (ATR) technology [3] for SAR images, which is aimed at solving this problem, has become a hot research topic.

In the past few decades, SAR ATR technology has made significant progress from theoretical research to practical applications. Classical methods for SAR ATR are based on templates and models. The template-based method can be further divided into direct template-matching methods [4], which calculate the correlation or distance between the test sample and the template obtained from the training sample itself, and feature template-matching methods, which use commonly used classifiers such as SVM [5], KNN [6] and Bayesian classifiers [7], to compare the geometric [8], mathematical [9,10], or electromagnetic scattering features [11,12] extracted from the training and test samples. Although

template-matching methods are simple in principle and easy to implement, it is difficult to establish a complete template library in practical applications and requires high storage and computation space. To overcome these limitations, the model-based SAR target recognition methods [13,14] are proposed, which use electromagnetic simulation software to calculate the electromagnetic scattering images based on the established models of the targets and perform feature matching with test samples to achieve target recognition.

Classical SAR target recognition methods rely on laborious manual-designed feature engineering, which to some extent are limited by data structure and feature extraction capability, leading to unstable recognition performance. Therefore, the automatic feature extraction ability of neural networks enabled the application of deep learning methods in SAR target recognition. Initially, the deep learning model used for computer vision tasks is directly fine-tuned for SAR target recognition [15], and some unsupervised learning methods are also directly used for SAR target feature extraction [16–18]. Subsequently, the methods specifically designed for SAR image amplitude information characteristics are proposed. Chen et al. [19] designed A-ConvNet based on VGG for the MSTAR dataset target recognition task and achieved a recognition rate above 99%; Lin et al. [20] proposed a deep convolutional Highway Unit for SAR ATR with few samples; Gao et al. [21] proposed a dual-branch deep convolutional neural network (Dual-CNN) to extract polarization and spatial features and fuse them together; Jiao et al. [22] designed a Wishart deep stacking network (DSN) specifically for polarimetric SAR target classification, which performed well on real polarimetric SAR data. Li et al. [23] proposed a fully convolutional attention module that focuses on important channels and target areas, improving the computational efficiency and significantly improving the performance of SAR target recognition. Recent studies attempt to combine deep learning with physical models, focusing on the characteristics brought about by the special imaging mechanism of SAR. Zhang et al. [24] introduced a network called DKTS-N to combine deep learning networks with domain-specific knowledge in the context of SAR. Huang et al. [24] proposed Deep SAR-Net, which uses CNN and Convolutional Autoencoders to extract spatial and scattering features of SAR images for classification tasks. Feng et al. [25] proposed a method based on integrated partial models and deep learning algorithms to combine electromagnetic scattering characteristics with deep neural networks.

In practical applications, due to the special imaging mechanism of SAR, the visual features of the same target vary greatly under different observation azimuths, which poses challenges for single-aspect SAR target recognition. As SAR systems advance, the development of multi-aspect SAR techniques, such as Circular SAR, enables the continuous observation of a given target from various viewing angles. Multi-aspect SAR target recognition technology utilizes multiple images of the same target obtained from different observation angles to combine the scattered characteristics from different perspectives. By fully exploiting the complementary and correlated recognition information of the target at different angles, multi-aspect SAR target recognition can significantly improve the accuracy and anti-interference ability of target recognition.

The deep learning methods used for multi-aspect SAR target recognition are mainly based on recurrent neural networks (RNN) and convolutional neural networks (CNN). For example, MA-BiLSTM [26] and BCRN [27] are based on long short-term memory networks, MVDCNN [28] is based on parallel CNN with hierarchical merging and fusion structures, while MVDFLN [29] combines the recurrent unit and convolution. The existing multi-aspect SAR target recognition methods mainly face the following challenges:

1. Multi-aspect SAR image recognition methods based on RNN or CNN are limited by sequence constraints. It is difficult to learn the correlation between two images that are far apart in the multi-aspect SAR image sequence, leading to information loss.
2. The number of SAR image samples is insufficient to meet the needs of deep learning training networks. Most existing methods adopt supervised learning methods combined with data augmentation. The limited SAR data restrict the generalization ability of deep learning models.

To address the limitations of existing methods, a multi-aspect SAR target recognition method based on contrastive learning (CL) and Non-Local is proposed in this paper. After pre-training, the encoder part of the CL network is used to extract feature maps from each image in the multi-aspect SAR image sequence. Based on the obtained feature maps, high-dimensional features are further extracted, while Non-Local computation is inserted between different feature extraction layers to achieve multi-aspect feature learning.

This paper proposes an innovative approach to exploit the correlation between multi-aspect SAR images by utilizing Non-Local [30]. Self-attention calculation is not affected by the order of images in the sequence; thus, it can mine the correlation information between images more effectively. As a classic application of self-attention in computer vision, Non-Local directly operates on two-dimensional images and achieves pixel-level self-attention calculation between feature maps with a simple and flexible structure. In consideration of the loss of local detailed information in self-attention calculation, the ResNet [31] structure is designed to extract feature maps. Convolutional operations and pre-training based on CL are employed to reduce the requirement for sample quantity and improve the generalization ability of the network.

Compared to existing methods, the novelty and contribution of the proposed method in this paper can be summarized as follows.

1. A Non-Local structure is introduced for multi-aspect SAR feature learning. By implementing self-attention calculation multiple times in feature spaces of different dimensions, Non-Local can effectively capture the correlation information among multi-aspect SAR images.
2. The lightweight contrastive learning network is applied to SAR image feature extraction tasks in order to fully utilize the limited SAR data to train an effective feature extraction network.
3. Compared with existing methods, our method achieves higher recognition accuracy on the MSTAR dataset and demonstrates better generalization performance in case of few samples and strong interference.

The remainder of this paper is organized as follows. A comprehensive description of the proposed network structure is provided in Section 2. Section 3 outlines the experimental details and discusses the obtained results. Section 4 discusses the benefits of the proposed method and outlines potential future work. Section 5 provides a summary of the entire paper.

2. Proposed Method

The overall architecture of the proposed network includes four parts, i.e., sequence construction, feature extraction based on pre-training by contrastive learning, multi-aspect feature learning based on Non-Local, and classification, as shown in Figure 1.

The multi-aspect SAR image sequences are constructed based on the single-aspect SAR images, which are also used for pre-training. In the CL network for pre-training, traditional data augmentation methods are used to generate two enhanced views for a single SAR image, each serving as input for the two branches. The optimization of the pre-train network is achieved by reducing the differences between the output features of two branches. After pre-training, the encoder part of the upper branch is transferred to extract the feature map of each image in the multi-aspect SAR image sequences. Then, during the multi-aspect feature learning process, the extracted feature maps of each image are input into the multi-aspect encoder based on Non-Local and ResNet to learn the correlation between multi-aspect SAR images. The output features of the multi-aspect encoder are dimensionally reduced and then averaged along the sample dimension for feature fusion. Finally, the softmax classifier is used to obtain the prediction probability. The following sections will provide details and training process of the proposed method.

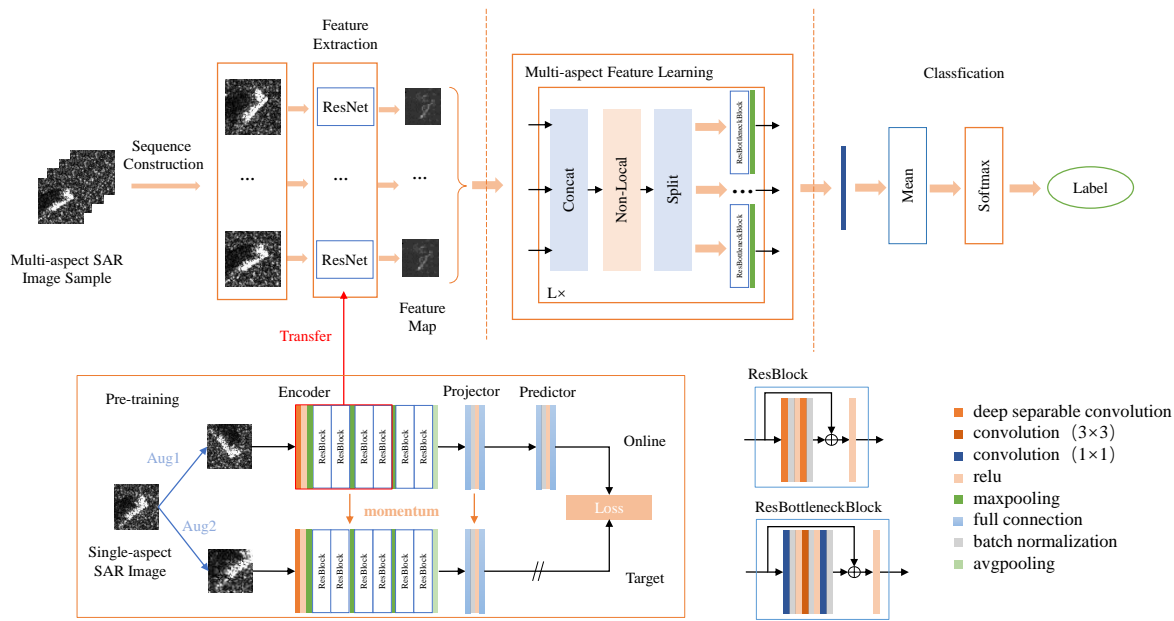


Figure 1. The overall architecture of the proposed network.

2.1. Sequence Construction

From different azimuth and depression angles, multi-aspect SAR images of the target can be obtained through one or more platforms, which can be used to construct multi-aspect SAR image sequences based on the following steps [32]. The original SAR image set $X^r = \{X^1, X^2, \dots, X^C\}$ consists of multiple categories X^i . The images in each category are sorted according to the azimuth, which is denoted as $X^i = \{x_1^i, x_2^i, \dots, x_{n_i}^i\}$. C represents the total number of target categories, n_i represents the number of images contained in each class and $\varphi(x_j^i)$ is assumed to be the azimuth angle of the image. For the given sequence length k , $k + 1$ images are selected from the original image set by a sliding window with the step size of 1, which are combined in different permutations to obtain sequences. The sequences in which the azimuth difference between any two images is less than the given angle range θ are selected as the experimental sample. Assume that the final constructed sequences of a certain class are denoted as $X_S^i = \{X_{s_1}^i, X_{s_2}^i, \dots, X_{s_{N_i}}^i\}$, where N_i represents the number of sequences. The construction process above is shown in Algorithm 1, and an example of multi-aspect SAR image sequence construction is shown in Figure 2.

Algorithm 1 Sequence construction algorithm

Initialization: The angle range θ , the sequence length k

Input: Original images $X^r = \{X^1, X^2, \dots, X^C\}$ and the number of classes C

Output: The constructed sequence set $X^S = \{X_S^1, X_S^2, \dots, X_S^C\}$

for $i = 1$ to C do

 for $j = 1$ to $n_i - k$ do

 if $|\varphi(x_j^i) - \varphi(x_{j+k}^i)| \leq \theta$

 Combine to add all possible sequences of length k except $\{x_{j+1}^i, x_{j+2}^i, \dots, x_{j+k}^i\}$

 else if $|\varphi(x_j^i) - \varphi(x_{j+k-1}^i)| \leq \theta$

 Add the sequence $\{x_j^i, x_{j+1}^i, \dots, x_{j+k-1}^i\}$

 end for

 if $|\varphi(x_{j+1}^i) - \varphi(x_{n_i}^i)| \leq \theta$

 Add the sequence $\{x_{j+1}^i, x_{j+2}^i, \dots, x_{n_i}^i\}$

end for

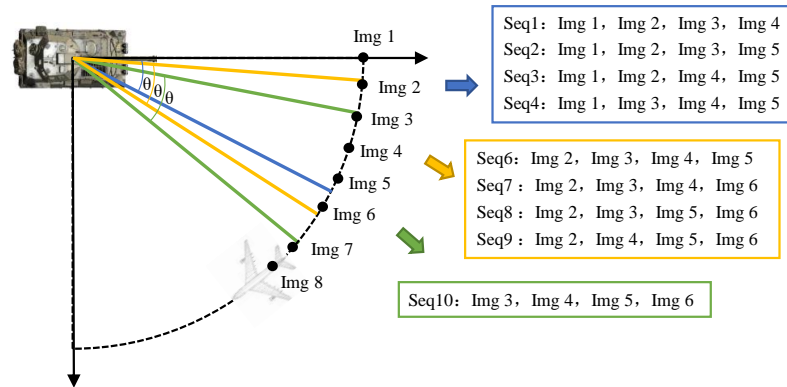


Figure 2. Example of sequence construction.

2.2. Feature Extraction Based on Contrastive Learning

To introduce contrastive learning into SAR image feature extraction, considering that the visual features presented in SAR images are often quite similar, we choose the Bootstrap Your Own Latent (BYOL) [33] network, which is based on the asymmetry of two branches instead of negative samples, for pre-training the ResNet model used for single-aspect SAR image feature extraction.

As shown in Figure 3, the online branch of the pre-training network based on BYOL consists of the encoder based on ResNet, the projector and predictor based on multi-layer perceptron (MLP) with the same structure. The target branch only includes the encoder and projector with the same structure as the online branch.

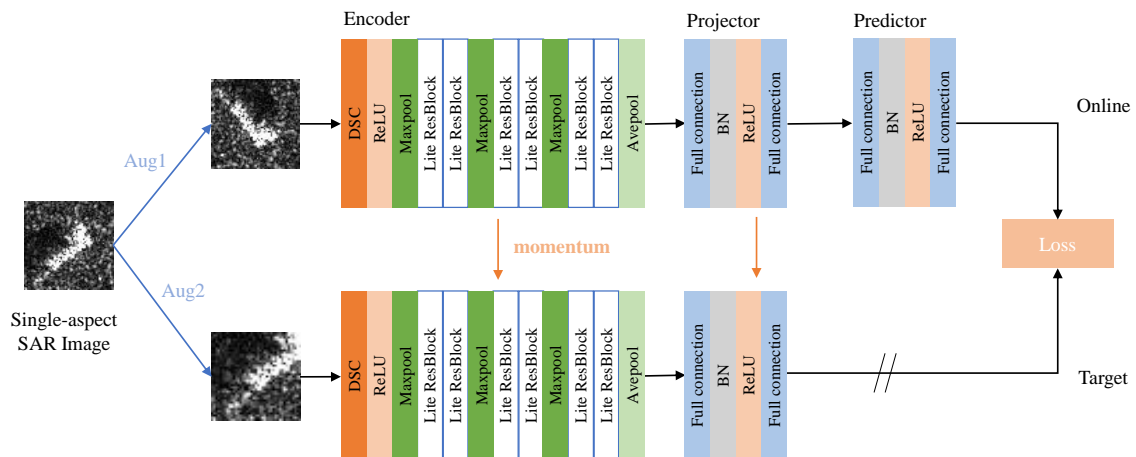


Figure 3. The structure of the pre-training network.

The encoder based on ResNet is used to output the feature representation of the input image. To obtain a lightweight model, Deep Separable Convolution (DSC) [34] is designed to replace the original convolution layer in ResNet, which consists of deep convolution and pointwise convolution, as shown in Figure 4. Deep convolution applies a separate convolution kernel to each channel of the input, while pointwise convolution uses the 1×1 convolution to fuse the output of deep convolution across channels and change the channel number of the final output. Compared to original convolution, DSC achieves a significant reduction in parameter and computational costs.

The input convolutional layer of the encoder contains a DSC layer and ReLU activation function. Then, the core structure consists of several units with the same structure stacked together, each of which contains a max pooling layer and two Lite ResBlocks. The Lite ResBlock includes two DSC layers and the residual connection as shown in Figure 5. The

output pooling layer of the encoder uses average pooling, which computes the average of values within the pooling window as the value after pooling.

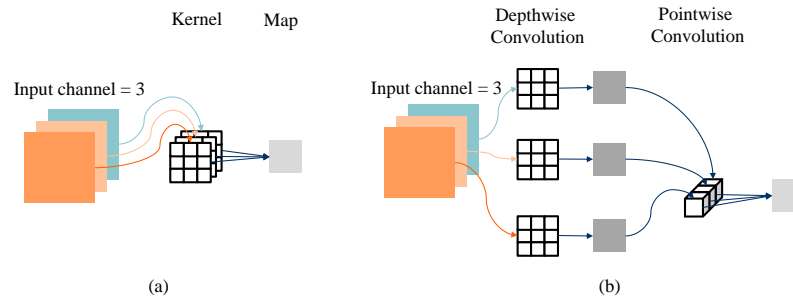


Figure 4. Schematic diagram of (a) convolution and (b) DSC process.

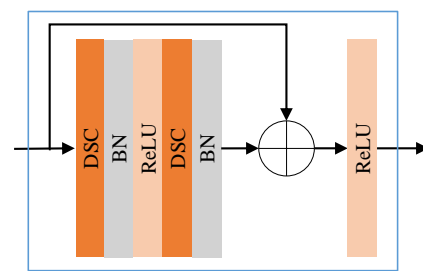


Figure 5. The structure of Lite ResBlock.

Assuming that the input image of the encoder is x , the DSC operation is denoted as $dscnv$ and σ represents the ReLU activation function, the calculation of the input convolution layer can be expressed as:

$$x^0 = \sigma(dscnv(x)) \tag{1}$$

Then, suppose the input of the n th unit stacked is x^{n-1} and the output is x^n , where the input of the first unit is x^0 . Suppose Φ_{Res} denotes the computation of Lite ResBlock, and f_{Down} denotes the max pooling operation. The operation of the n th unit can be represented as:

$$x^n = \Phi_{Res}(\Phi_{Res}(f_{Down}(x^{n-1}))) \tag{2}$$

Suppose Φ_{BN} is the batch normalization operation; then, the calculation process of Φ_{Res} is defined as:

$$\Phi_{Res}(z) = z + \Phi_{BN}(dscnv(\sigma(\Phi_{BN}(dscnv(z)))))) \tag{3}$$

The output of the last unit x_N passes through the average pooling layer, denoted as f_{AVG} , to obtain the final output x_r of the encoder, which can be expressed as:

$$x_r = f_{AVE}(x^N) \tag{4}$$

The projector and predictor are based on the multi-layer perceptron with the same structure, consisting of two fully connected layers separated by batch normalization and ReLU activation functions, which expand and reduce the dimensionality of the feature vector. The input of the projector is the output of the encoder x_r , and the output of the projector is denoted as x_{pro} . The predictor takes the output of the projector as its input and is denoted as x_{pre} . The calculation process of the projector and predictor is as follows:

$$x_{pro} = \Phi_{FC2}(\sigma(\Phi_{BN}(\Phi_{FC1}(x_r)))) \tag{5}$$

$$x_{pre} = \Phi_{FC4}(\sigma(\Phi_{BN}(\Phi_{FC3}(x_{pre})))) \tag{6}$$

where Φ_{FC1} and Φ_{FC2} represent the two fully connected sub-layers in the projector, while Φ_{FC3} and Φ_{FC4} represent the two fully connected sub-layers in the predictor.

The pre-training network takes the single-aspect SAR image as input and generates two augmented views through methods such as rotation, flipping, cropping, scaling, and brightness adjustment. The online branch goes through the encoder, projector, and predictor to obtain the output vector, while the target branch obtains the output vector only through the encoder and projector. During pre-training, the network is optimized by minimizing the error between the output vectors from the two branches, which will be introduced in Section 2.5.1. After pre-training, the input convolutional layer and the first N stacked units of the online branch encoder will be transferred for extracting feature maps of each image in the sequence. The obtained feature maps are denoted as $x_{(i)}^N, i = 1, \dots, k$.

2.3. Multi-Aspect Encoder Based on Non-Local

The multi-aspect encoder, which combines Non-Local and ResNet, is designed to achieve multi-aspect feature learning. The main structure and details of the multi-aspect encoder are shown in Figure 6.

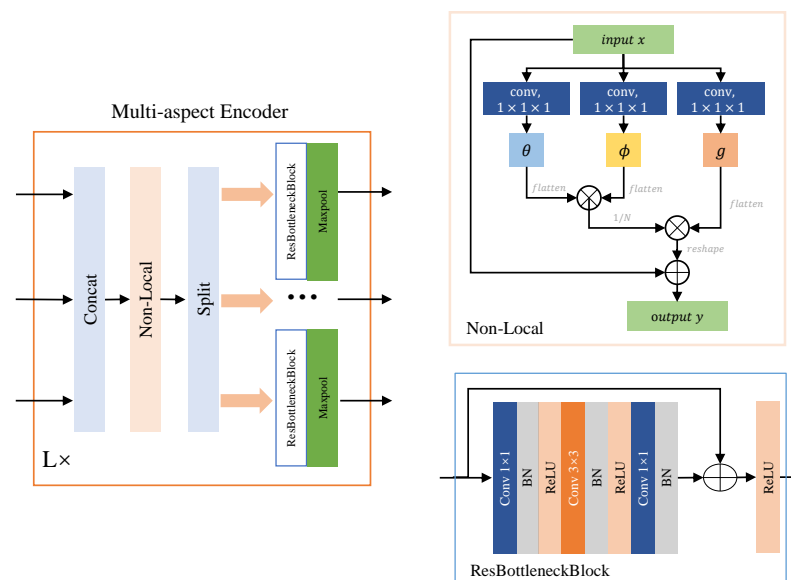


Figure 6. The main structure and details of the multi-aspect encoder.

The multi-aspect encoder is composed of multiple layers. Each layer concatenates the feature maps extracted from each SAR image in the input multi-aspect SAR image sequence and performs self-attention calculation via the Non-Local layer. Then, the output of Non-Local is split along the sample dimension to extract higher-level features separately for each image by ResBottleneckBlock. Finally, the output is downsampled through the maxpool layer. Suppose there are M layers in the multi-aspect encoder. The detailed calculation process for each layer will be described next.

Assuming the m th layer of the multi-aspect encoder takes input $x_{(i)}^{m-1} \in \mathbb{R}^{d_{m-1} \times d_{m-1} \times c_{m-1}}, i = 1, \dots, k$ and output $x_{(i)}^m \in \mathbb{R}^{d_m \times d_m \times c_m}, i = 1, \dots, k$, where k is the length of the multi-aspect SAR image sequence, d_{m-1} and d_m are the feature map sizes of the input and output of the m th layer, and c_{m-1} and c_m represent the number of channels of the input and output in the m th layer. The input in the first layer is the output $x_{(i)}^N, i = 1, \dots, k$ from the feature extraction part of the network. Firstly, we concatenate the input feature maps into $X_{m-1} = [x_{(1)}^{m-1}, \dots, x_{(i)}^{m-1}]$, which serves as the input vector for Non-Local.

Non-Local is a residual structure that computes the output by adding the input vector with the result of the self-attention computation. Given an input vector $X_{m-1} \in$

$\mathbb{R}^{k \times d_{m-1} \times d_{m-1} \times c_{m-1}}$, three convolution operations with kernel size $1 \times 1 \times 1$ are performed to obtain query vector θ_m , key vector ϕ_m , and value vector g_m with the same dimensions as the input vector. The convolution kernels are denoted as W_θ^m , W_ϕ^m , and W_g^m , and the corresponding biases are denoted as b_θ^m , b_ϕ^m and b_g^m . The calculation process can be expressed as:

$$\theta_m = X_{m-1} * W_\theta^m + b_\theta^m \tag{7}$$

$$\phi_m = X_{m-1} * W_\phi^m + b_\phi^m \tag{8}$$

$$g_m = X_{m-1} * W_g^m + b_g^m \tag{9}$$

The shapes of θ_m , ϕ_m and g_m are flattened into (v_{m-1}, c_{m-1}) , where $v_{m-1} = k \times d_{m-1} \times d_{m-1}$. Next, the original correlation matrix $\hat{A}_m \in \mathbb{R}^{v_{m-1} \times v_{m-1}}$ is calculated through the dot product, which means transposing the key vector ϕ_m and multiplying it with the query vector θ_m . Then, the weight matrix A_m is obtained by normalizing \hat{A}_m . The calculation process is as follows:

$$\hat{A}_m = (\phi_m)^T \theta_m \tag{10}$$

$$A_m = \hat{A}_m / N_m \tag{11}$$

where N_m represents the position number of the input vector, that is, v_{m-1} . The weight matrix A_m is multiplied with the value vector g_m to obtain the output \hat{Y}_m of self-attention, which is given by:

$$\hat{Y}_m = g_m A_m \tag{12}$$

Finally, reshape \hat{Y}_m into $(k, d_{m-1}, d_{m-1}, c_{m-1})$ and add it to the input vector X_{m-1} of Non-Local to complete the residual calculation and obtain the output Y_m , which can be expressed as:

$$Y_m = X_{m-1} + \hat{Y}_m \tag{13}$$

The output of Non-Local Y_m is split back into individual feature maps $y_{(i)}^m \in \mathbb{R}^{d_{m-1} \times d_{m-1} \times c_{m-1}}$, $i = 1, \dots, k$, which are then fed into ResBottleneckBlock with shared parameters for further feature extraction. ResBottleneckBlock is a residual structure where the input vector passes through a 1×1 convolution layer with batch normalization (BN) and ReLU, which is followed by a 3×3 convolution layer with BN and ReLU and finally a 1×1 convolution layer with BN only. The output of ResBottleneckBlock is then added to the input vector and passed through another ReLU activation function. The calculation process can be expressed as follows:

$$\hat{x}_{(i)}^m = \sigma(y_{(i)}^m + \Psi_{1 \times 1}(\Phi_{3 \times 3}(\Phi_{1 \times 1}(y_{(i)}^m)))) \tag{14}$$

where $\Phi_{3 \times 3}$ represents the 3×3 convolution with BN and ReLU, $\Phi_{1 \times 1}$ represents the 1×1 convolution with BN and ReLU, and $\Psi_{1 \times 1}$ represents the 1×1 convolution with only BN. The output of the ResBottleneckBlock is downsampled using the maxpool layer to obtain the output of the m th layer in the multi-aspect encoder.

$$x_{(i)}^m = f_{\text{DOWN}}(\hat{x}_{(i)}^m) \tag{15}$$

2.4. Feature Dimensionality Reduction and Classification

The output of the M th layer of the multi-aspect encoder is k 1-D feature vectors, which are denoted as $x_{(i)}^M$, $i = 1, \dots, k$. They are concatenated along the sample dimension to obtain $X_M = [x_{(1)}^M, \dots, x_{(k)}^M]$, with the size $k \times c_M$. The feature dimension of X_M is reduced by a 1×1 convolution layer to obtain the output $Z \in \mathbb{R}^{k \times C}$, where C is the number of sample classes. After averaging Z along the sample dimension to achieve feature fusion,

the softmax classifier is applied to obtain the predicted probabilities of the input samples output by the network.

2.5. Training Process

2.5.1. Pre-Train Based on CL and Layer Transfer

As described in Section 2.2, the CL network is optimized by minimizing the difference between the outputs of the online and target branch. The MSE loss function is used to calculate the error between the two branches, which is computed as the distance between the L2 normalized output vectors of the two branches. The loss function can be formulated as:

$$L_{Aug1} = \|x_{pre}^1 - x_{pro}^2\|_2^2 = 2 - 2 \frac{\langle x_{pre}^1, x_{pro}^2 \rangle}{\|x_{pre}^1\| \cdot \|x_{pro}^2\|} \quad (16)$$

where x_{pre}^1 represents the output of the online branch for the augmented view Aug1, and x_{pro}^2 represents the output of the target branch for the augmented view Aug2. Due to the asymmetry of the structure, the network needs to exchange input views of the two branches; then, the complete loss function of the network is:

$$L_{Pretrain} = L_{Aug1} + L_{Aug2} \quad (17)$$

The parameter of the online branch is updated by backpropagation (BP), while the parameter of the target branch is updated using the momentum update mechanism, meaning that the update is determined based on the corresponding parameter of the online branch. The momentum update mechanism can be expressed as:

$$\zeta = m\zeta + (1 - m)\eta \quad (18)$$

where ζ represents the parameter of the target branch and η represents the updated parameter of the online branch. m is the weight coefficient which is usually large, so that the parameter of the target branch changes slowly and steadily approaching η .

After the pre-training network converges, the first few layers of the online branch encoder are transferred to the entire network to extract feature maps from each image in sequences. During the training process of the entire network, the parameters of the feature extraction part are fixed and unchanged.

2.5.2. Training of Overall Network

Considering that DSC used in the Lite ResBlock may lead to the loss of model accuracy under certain conditions [35], Knowledge Distillation (KD) [36] is introduced into the overall network training. The supervised information of better performing but more complex models can be involved in the training of lightweight models through KD, thereby improving the performance of lightweight models.

The proposed lightweight network is the student model, and the network using convolutional layers instead of DSC is the teacher model. One of the key points of KD is to add a temperature parameter T to the softmax classifier, which can be described as:

$$q_i = \frac{e^{z_i/T}}{\sum_j e^{z_j/T}} \quad (19)$$

In the training process of the student model, the parameters of the teacher model that has been trained based on the cross-entropy loss function remain unchanged. The loss function of the student model includes the cross-entropy loss between the output and the sample label and the distillation loss that measures the gap at the same temperature t

between the student model and the teacher model using Kullback–Leibler divergence. The loss function can be formulated as:

$$L_{soft} = - \sum_j^N p_j^t \log(q_j^t) \quad (20)$$

$$L_{hard} = - \sum_j^N c_j \log(q_j^1) \quad (21)$$

$$L = \alpha L_{soft} + (1 - \alpha) L_{hard} \quad (22)$$

where N represents the number of class types, p_j^t and q_j^t represent the prediction probability for class j output by the teacher model and student model at temperature t , c_j represents the value of the sample label corresponding to class j , and α is the proportion coefficient.

3. Experiments and Results

3.1. Network Architecture Setup

In the experiment, the input SAR images are cropped to 64×64 . The structure of each layer in the online branch of the pre-train network is shown in Table 1, while the target branch does not include the projector. The Conv1 to ResBlock2 layers of the pre-trained model are transferred for feature extraction after pre-training. The number of channels in each layer of the multi-aspect encoder is 256, and the size of the maxpool layer in the ResBottleneckBlock is 3×3 with the stride 2.

Table 1. The structure of each layer in the online branch of the pre-train network.

Part	Layer Name	Input Size	Parameters
Encoder	Conv1	$64 \times 64 \times 1$	DSC 7×7 , 64, stride 2
	Maxpool1	$32 \times 32 \times 64$	Maxpool 3×3 , 64, stride 2
	ResBlock1	$16 \times 16 \times 64$	Lite ResBlock-128 $\times 2$
	Maxpool2	$16 \times 16 \times 128$	Maxpool 3×3 , 64, stride 2
	ResBlock2	$8 \times 8 \times 128$	Lite ResBlock-256 $\times 2$
	Maxpool3	$8 \times 8 \times 256$	Maxpool 3×3 , 64, stride 2
	ResBlock3	$4 \times 4 \times 256$	Lite ResBlock-512 $\times 2$
	Avepool3	$4 \times 4 \times 512$	output size = (1, 1)
Projector	FC1	$1 \times 1 \times 512$	2048-d
	FC2	$1 \times 1 \times 2048$	512-d
Predictor	FC3	$1 \times 1 \times 512$	2048-d
	FC4	$1 \times 1 \times 2048$	512-d

Our proposed network is implemented using Pytorch 1.9.1. The key hardware used for all experiments in this paper are an Intel Core i7-9750H CPU and NVIDIA GeForce RTX 2060 GPU. The weight coefficient of the pre-train network is 0.99. With the mini-batch size of 32, the learning rate is 0.001 for pre-training and 0.0001 for overall training.

3.2. Dataset

The MSTAR dataset [37] mainly comprises high-quality SAR images captured using the X-band high-resolution Spotlight SAR from 10 stationary military vehicles with the resolution of 0.3×0.3 m and HH polarization. The azimuth angle of images for each type of target covers $0 \sim 360^\circ$, and the interval between images is $5 \sim 6^\circ$. The optical and corresponding SAR images of ten targets are shown in Figure 7.

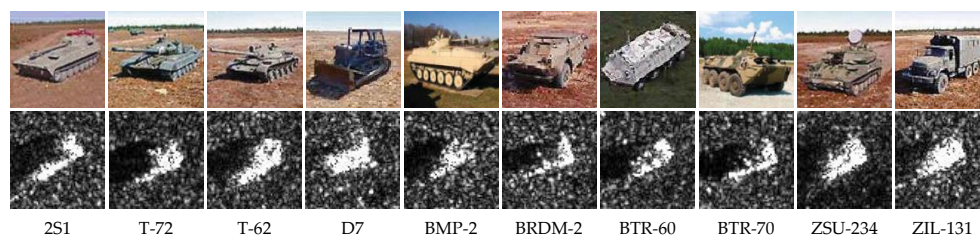


Figure 7. The optical and SAR images of ten targets.

Standard Operating Condition (SOC) and Extended Operating Condition (EOC) are two categories of the acquisition conditions in the MSTAR dataset. SOC indicates that the target category of training and testing data is the same with the similar imaging conditions, while the data difference is greater and the recognition difficulty is higher under EOC.

3.3. Results under SOC and EOC

The training and testing data in the experiment under SOC are of the same category with the depression angles of 17° and 15° , respectively. Considering the actual radar imaging situation and the trade-off between data acquisition and network training costs [28], the angle range is set to 45° when constructing the sequences. Table 2 shows the class types used in the experiment and the number of training and testing samples with different sequence lengths.

Table 2. Dataset of SOC.

Class Type	Image Samples	Training			Image Samples	Testing		
		2-Aspect Sequences	3-Aspect Sequences	4-Aspect Sequences		2-Aspect Sequences	3-Aspect Sequences	4-Aspect Sequences
2S1	299	578	840	1084	274	525	755	956
BRDM2	298	576	837	1080	274	525	755	956
BTR60	256	489	709	906	195	362	508	589
D7	299	578	840	1084	274	525	755	959
T72	232	443	640	814	196	364	499	569
BMP2	233	443	642	812	195	362	494	558
BTR70	233	442	639	817	196	363	496	568
T62	299	578	840	1084	273	522	749	954
ZIL131	299	578	840	1084	274	525	755	956
ZSU234	299	578	846	1087	274	525	755	959
Total	2747	5283	7673	9852	2425	4598	6521	8024

Compared with SOC, the target recognition tasks under EOC are more challenging due to the greater differences between the training and testing data, which mainly include the configuration variation (EOC-C) and version variation (EOC-V). EOC-V refers to the same type of target built based on different blueprints, while EOC-C denotes targets of same manufacturing method with different post-production equipment added. The training data for EOC consists of four types of targets with the depression angle 17° . The testing data for EOC-C include seven different configurations for two types of targets at 17° and 15° , while the testing data for EOC-V consist of images of five different versions of T72. The training and testing data of experiments under EOC are listed in Tables 3–5.

Based on the data shown in Tables 3–5, the recognition performance of the method proposed is experimentally verified. When the input is the single-aspect SAR image or the image sequence with different lengths, the experimental results on SOC and EOC data are shown in Table 6.

Table 3. The training data of EOC.

Class Type	Depression Angle	Image Samples	2-Aspect Sequences	3-Aspect Sequences	4-Aspect Sequences
BMP2	17°	233	443	642	812
BRDM2	17°	298	576	837	1080
BTR70	17°	233	442	639	817
T72	17°	232	443	640	814
Total	17°	996	1904	2758	3523

Table 4. The testing data of EOC-V.

Class Type	Depression Angle	Image Samples	2-Aspect Sequences	3-Aspect Sequences	4-Aspect Sequences
T72/A32	17° & 15°	572	1103	1595	2046
T72/A62	17° & 15°	573	1105	1604	2050
T72/A63	17° & 15°	573	1105	1598	2044
T72/A64	17° & 15°	573	1105	1598	2050
T72/S7	17° & 15°	419	789	1116	1349
Total	17° & 15°	2710	5207	7511	9539

Table 5. The testing data of EOC-C.

Class Type	Depression Angle	Image Samples	2-Aspect Sequences	3-Aspect Sequences	4-Aspect Sequences
T72/A04	17° & 15°	573	1105	1598	2044
T72/A05	17° & 15°	573	1105	1598	2050
T72/A07	17° & 15°	573	1105	1604	2050
T72/A10	17° & 15°	567	1092	1577	2001
T72/812	17° & 15°	426	803	1133	1369
BMP2/9566	17° & 15°	428	807	1145	1401
BMP2/C21	17° & 15°	429	811	1143	1381
Total	17° & 15°	3569	6828	9798	12,296

Table 6. Recognition accuracy of the proposed method under different inputs.

Input	Recognition Accuracy (%)		
	SOC	EOC-C	EOC-V
single-aspect	98.10	95.21	98.52
2-aspect	99.52	96.68	99.27
3-aspect	99.64	97.75	99.69
4-aspect	99.94	98.81	99.75

Based on the experimental results shown in Table 6, the proposed method in this paper achieves higher recognition accuracy compared to single-aspect SAR target recognition when inputting multi-aspect SAR image sequences under both SOC and EOC. Moreover, the recognition accuracy gradually increases as the length of the multi-aspect SAR image sequence increases. The experimental results show that the proposed method based on CL and Non-Local for multi-aspect SAR target recognition can effectively learn the correlation information between features of multi-aspect SAR images.

3.4. Recognition Performance Comparison

To further validate the recognition performance, the proposed method is compared with six existing multi-aspect SAR target recognition methods, including three classical methods, i.e., JSR [38], SRC [39] and data fusion [40], and three deep learning methods, i.e., MVDCNN [28], BCRN [27] and MVDFLN [29]. The results are shown in Table 7 while ensuring that the inputs of each method are as close as possible. The experimental

results show that the proposed method achieves higher recognition accuracy compared to existing multi-aspect SAR target recognition methods, which confirms the effectiveness of the proposed method.

Table 7. Recognition accuracy of the proposed method and existing methods.

Method	SOC	Recognition Accuracy (%)	
		EOC-C	EOC-V
JSR	94.69	-	-
SRC	98.94	96.78	-
Data Fusion	98.32	-	-
MVDCNN	98.52	95.45	95.46
BCRN	99.50	97.21	98.59
MVDFLN	99.62	97.84	99.10
Our Method	99.94	98.81	99.75

3.5. Discussion

The proposed multi-aspect SAR target recognition method learns multi-aspect features by performing multiple Non-Local calculations on the feature map. To verify the effectiveness of this structure, the experiment is designed to compare the network recognition performance under different Non-Local layers, and the experimental results are shown in Table 8. When the number of Non-Local layers increases, the recognition accuracy of the network also increases, proving that multiple self-attention calculations help to focus more on the correlation information between features in multi-aspect SAR images, thereby improving the recognition accuracy of the network.

Table 8. Recognition accuracy with different layers of Non-Local.

Number of Layers	SOC	Recognition Accuracy (%)	
		EOC-C	EOC-V
1	98.87	97.63	98.58
2	99.41	98.32	99.34
3	99.94	98.81	99.75

To verify the lightweight effect of the model, the comparison of model size and FLOPs between the proposed method and other existing methods is shown in Table 9. It can be seen that our proposed lightweight design can effectively reduce the number of parameters and FLOPs of the proposed method. After lightweight, the proposed method significantly reduced the number of parameters compared to existing methods. However, as for the FLOPs, which reflects the inference speed of the network, our proposed method can still be further optimized.

Table 9. Model size and FLOPs comparison with 4-aspect input sequences.

Method	BCRN	MVDCNN	Our Method without Lightweight	Our Method
Model Size(M)	135.25	11.49	5.28	1.63
FLOPs(G)	1.894	2.654	45.759	10.984

In order to experimentally verify the robustness of the proposed method, first, the training sample sequences are downsampled to simulate the few-sample condition. The results of MVDCNN, BCRN, and our method with different sampling ratios are summarized in Table 10, which show that when the current sampling ratio is reduced to 2%, the recognition accuracy of our method can still be maintained at above 90%, demonstrating good robustness with few samples.

Next, in order to quantitatively verify the robustness of this method under noise, Table 11 shows the recognition accuracy of different methods when adding Gaussian white noise with variance from 0.01 to 0.05 to the testing data. The experimental results show that our method can still maintain a higher recognition accuracy than BCRN under strong noise and is basically on par with MVDCNN, proving that the proposed network has good anti-noise ability.

Table 10. Performance comparison between this and existing methods with few samples.

Methods	Recognition Accuracy of Different Sampling Ratios (%)					
	100%	50%	25%	10%	5%	2%
MVDCNN	99.09	98.75	98.45	97.33	95.00	87.99
BCRN	99.50	99.43	98.99	94.34	91.43	77.70
Our Method	99.94	99.77	99.31	98.68	96.27	91.84

Table 11. Comparison of anti-noise performance between this and existing methods.

Methods	Recognition Accuracy with Different Variance of Noise (%)				
	0.01	0.02	0.03	0.04	0.05
BCRN	98.46	88.73	70.31	54.78	44.63
MVDCNN	98.17	94.64	90.32	83.21	72.49
Our Method	98.72	94.35	89.64	81.59	70.57

Finally, the simulated occlusion images are generated by blocking the peak points [41] in the testing images to test the recognition performance of the proposed method under partial target occlusion. Specifically, after preprocessing, the multi-aspect SAR images are rotated to a uniform orientation based on the azimuth angle. Each SAR image is then scanned to identify peak points, whose grayscale value is greater than that of all its eight neighbors. The total number of peak points in each image is recorded to determine the number of peak points based on the occlusion ratio. Each SAR image is scanned along four vertical directions denoted as d_1 to d_4 , and the peak points encountered first are occluded by setting their grayscale values to zero to simulate the occlusion of targets. Taking an occlusion ratio of 50% as the example, the simulated SAR occluded images generated in the directions of d_1 to d_4 are shown in Figure 8. The performance of our method under different occlusion ratios with different inputs is shown in Table 12. It can be seen that a 4-aspect input provides more learnable feature information for target recognition than the single-aspect input under target occlusion, thus achieving better recognition accuracy.

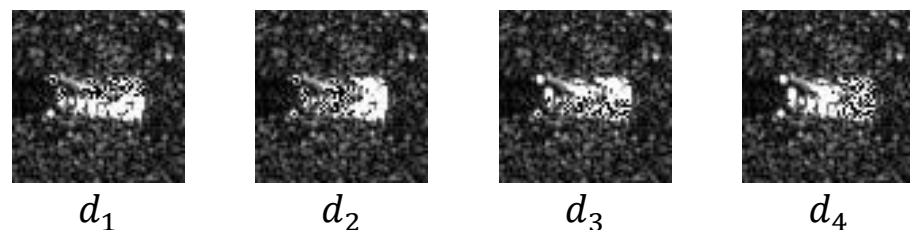


Figure 8. Example of simulated SAR occlusion target image.

Table 12. Comparison under different occlusion ratios and different inputs.

Input	Recognition Accuracy with Different Occlusion Ratios (%)				
	10%	20%	30%	40%	50%
single-aspect	98.03	97.75	96.94	95.76	93.28
4-aspect	99.91	99.86	99.73	99.54	99.09

4. Discussion

4.1. Advantages

The experimental results in Section 3.4 indicate that the proposed method achieves higher recognition accuracy under both SOC and EOC compared to existing methods, which proves the feasibility of achieving multi-aspect SAR target recognition through self-attention calculation between 2D feature maps under various complex conditions.

The proposed method can achieve higher recognition rates than other methods under few samples as shown in Section 3.5. Considering the high cost of SAR image acquisition and annotation in practical applications, our method is more practical due to its ability to achieve more effective target recognition under limited data.

Meanwhile, the results of experiments under noise and occlusion verify the good generalization performance of the proposed method. In practical SAR target recognition tasks, noise and occlusion are very common but pose significant difficulties for target recognition. The good anti-noise and anti-interference performance for ground vehicle target recognition makes the proposed method more valuable in practice.

4.2. Future Work

Although the proposed method has been experimentally validated for the effectiveness, there are still some issues that can be further studied and improved from the following two aspects in the future.

Firstly, the experiments in this paper are all based on the MSTAR dataset and simulation data. On one hand, the construction process of multi-aspect SAR sequences requires further experiments and optimization, such as the impact of angle range and sliding window step size on the subsequent experiments. On the other hand, considering the ideality of simulation methods and the singularity of existing data on the target, the actual application performance and effectiveness against other targets of the model of the proposed method still need to be further verified after collecting data in real environments.

In addition, the characteristics brought about by the special imaging mechanism of SAR are currently an important direction that needs to be studied. The proposed method only focuses on the amplitude information of SAR images, but it ignores the additional information contained in complex SAR images which can also be learned through networks to further improve model performance.

5. Conclusions

This paper proposes a multi-aspect SAR target recognition method based on contrastive learning and a Non-Local structure. Specifically, the lightweight ResNet model is pre-trained using a contrastive learning network based on single-aspect SAR images, after which it is transferred for SAR image feature extraction. Then, Non-Local is used to apply self-attention to two-dimensional feature maps to fully learn the correlation between multi-aspect SAR image features and improve the recognition accuracy by performing multiple pixel-level self-attention calculations in different dimensional feature spaces. Experiments conducted on the MSTAR dataset validate that the proposed multi-aspect SAR target recognition method achieves satisfactory recognition accuracy and good generalization performance.

Author Contributions: Conceptualization, Z.P. and X.Z.; methodology, Z.P. and S.L.; software, S.L.; validation, X.Z. and G.Z.; investigation, S.L.; resources, Y.H.; writing—original draft preparation, S.L.; writing—review and editing, Z.P., S.L. and X.Z.; supervision, Z.P., X.Z. and G.Z.; project administration, X.Z. and G.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Youth Innovation Promotion Association, CAS under number 2022119.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: (<https://www.sdms.afri.af.mil/>, accessed on 20 April 2023).

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

References

1. Wang, P.; Liu, W.; Chen, J.; Niu, M.; Yang, W. A high-order imaging algorithm for high-resolution spaceborne SAR based on a modified equivalent squint range model. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1225–1235. [[CrossRef](#)]
2. Scheer, J.; Holm, W.A. *Principles of Modern Radar: Radar Applications*; The Institution of Engineering and Technology: London, UK, 2013.
3. Bhanu, B. Automatic target recognition: State of the art survey. *IEEE Trans. Aerosp. Electron. Syst.* **1986**, *22*, 364–379. [[CrossRef](#)]
4. Novak, L.M.; Owirka, G.J.; Brower, W.S.; Weaver, A.L. The automatic target-recognition system in SAIP. *Linc. Lab. J.* **1997**, *10*, 187–202.
5. Zhao, Q.; Principe, J.C. Support vector machines for SAR automatic target recognition. *IEEE Trans. Aerosp. Electron. Syst.* **2001**, *37*, 643–654. [[CrossRef](#)]
6. Hou, B.; Kou, H.; Jiao, L. Classification of polarimetric SAR images using multilayer autoencoders and superpixels. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3072–3081. [[CrossRef](#)]
7. Ma, W.; Wu, Y.; Gong, M.; Xiong, Y.; Yang, H.; Hu, T. Change detection in SAR images based on matrix factorisation and a bayes classifier. *Int. J. Remote Sens.* **2019**, *40*, 1066–1091. [[CrossRef](#)]
8. Papsion, S.; Narayanan, R.M. Classification via the shadow region in SAR imagery. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 969–980. [[CrossRef](#)]
9. Lu, X.; Han, P.; Wu, R. Research on mixed PCA/ICA for SAR image feature extraction. In Proceedings of the 2008 9th International Conference on Signal Processing, Beijing, China, 26–29 October 2008; pp. 2465–2468.
10. Song, S.; Xu, B.; Yang, J. SAR target recognition via supervised discriminative dictionary learning and sparse representation of the SAR-HOG feature. *Remote Sens.* **2016**, *8*, 683. [[CrossRef](#)]
11. Cong, Y.; Chen, B.; Liu, H.; Jiu, B. Nonparametric Bayesian attributed scattering center extraction for synthetic aperture radar targets. *IEEE Trans. Signal Process.* **2016**, *64*, 4723–4736. [[CrossRef](#)]
12. Zhang, H.; Tian, X.; Wang, C.; Wu, F.; Zhang, B. Merchant vessel classification based on scattering component analysis for COSMO-SkyMed SAR images. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1275–1279. [[CrossRef](#)]
13. Chiang, H.C.; Moses, R.L.; Potter, L.C. Model-based classification of radar images. *IEEE Trans. Inf. Theory* **2000**, *46*, 1842–1854. [[CrossRef](#)]
14. Zhou, J.; Shi, Z.; Cheng, X.; Fu, Q. Automatic target recognition of SAR images based on global scattering center model. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3713–3729.
15. Wang, Y.; Wang, C.; Zhang, H.; Zhang, C.; Fu, Q. Combing single shot multibox detector with transfer learning for ship detection using Chinese Gaofen-3 images. In Proceedings of the 2017 Progress in Electromagnetics Research Symposium-Fall (PIERS-FALL), Singapore, 19–22 November 2017; pp. 712–716.
16. Xie, H.; Wang, S.; Liu, K.; Lin, S.; Hou, B. Multilayer feature learning for polarimetric synthetic radar data classification. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 2818–2821.
17. Geng, J.; Fan, J.; Wang, H.; Ma, X.; Li, B.; Chen, F. High-resolution SAR image classification via deep convolutional autoencoders. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2351–2355. [[CrossRef](#)]
18. Qin, F.; Guo, J.; Sun, W. Object-oriented ensemble classification for polarimetric SAR imagery using restricted Boltzmann machines. *Remote Sens. Lett.* **2017**, *8*, 204–213. [[CrossRef](#)]
19. 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. [[CrossRef](#)]
20. Lin, Z.; Ji, K.; Kang, M.; Leng, X.; Zou, H. Deep convolutional highway unit network for SAR target classification with limited labeled training data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1091–1095. [[CrossRef](#)]
21. Gao, F.; Huang, T.; Wang, J.; Sun, J.; Hussain, A.; Yang, E. Dual-branch deep convolution neural network for polarimetric SAR image classification. *Appl. Sci.* **2017**, *7*, 447. [[CrossRef](#)]
22. Jiao, L.; Liu, F. Wishart deep stacking network for fast POLSAR image classification. *IEEE Trans. Image Process.* **2016**, *25*, 3273–3286. [[CrossRef](#)]
23. Li, R.; Wang, X.; Wang, J.; Song, Y.; Lei, L. SAR target recognition based on efficient fully convolutional attention block CNN. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 4005905. [[CrossRef](#)]
24. 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 Trans. Geosci. Remote Sens.* **2021**, *60*, 5215315. [[CrossRef](#)]
25. Feng, S.; Ji, K.; Zhang, L.; Ma, X.; Kuang, G. SAR target classification based on integration of ASC parts model and deep learning algorithm. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 10213–10225. [[CrossRef](#)]
26. Zhang, F.; Hu, C.; Yin, Q.; Li, W.; Li, H.C.; Hong, W. Multi-aspect-aware bidirectional LSTM networks for synthetic aperture radar target recognition. *IEEE Access* **2017**, *5*, 26880–26891. [[CrossRef](#)]
27. Bai, X.; Xue, R.; Wang, L.; Zhou, F. Sequence SAR image classification based on bidirectional convolution-recurrent network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9223–9235. [[CrossRef](#)]
28. Pei, J.; Huang, Y.; Huo, W.; Zhang, Y.; Yang, J.; Yeo, T.S. SAR automatic target recognition based on multiview deep learning framework. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 2196–2210. [[CrossRef](#)]

29. Pei, J.; Huo, W.; Wang, C.; Huang, Y.; Zhang, Y.; Wu, J.; Yang, J. Multiview deep feature learning network for SAR automatic target recognition. *Remote Sens.* **2021**, *13*, 1455. [[CrossRef](#)]
30. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
32. Li, S.; Pan, Z.; Hu, Y. Multi-Aspect Convolutional-Transformer Network for SAR Automatic Target Recognition. *Remote Sens.* **2022**, *14*, 3924. [[CrossRef](#)]
33. Grill, J.B.; Strub, F.; Altché, F.; Tallec, C.; Richemond, P.; Buchatskaya, E.; Doersch, C.; Avila Pires, B.; Guo, Z.; Gheshlaghi Azar, M.; et al. Bootstrap your own latent—a new approach to self-supervised learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21271–21284.
34. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
35. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.
36. Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. *arXiv* **2015**, arXiv:1503.02531.
37. Ross, T.D.; Worrell, S.W.; Velten, V.J.; Mossing, J.C.; Bryant, M.L. Standard SAR ATR evaluation experiments using the MSTAR public release data set. In *Algorithms for Synthetic Aperture Radar Imagery V*; SPIE: Bellingham, WA, USA, 1998; Volume 3370, pp. 566–573.
38. Zhang, H.; Nasrabadi, N.M.; Zhang, Y.; Huang, T.S. Multi-view automatic target recognition using joint sparse representation. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 2481–2497. [[CrossRef](#)]
39. Ding, B.; Wen, G. Exploiting multi-view SAR images for robust target recognition. *Remote Sens.* **2017**, *9*, 1150. [[CrossRef](#)]
40. Ruohong, H.; Keji, M.; Yanjing, L.; Jiming, Y.; Ming, X. SAR target recognition with data fusion. In Proceedings of the 2010 WASE International Conference on Information Engineering, Qinhuangdao, China, 14–15 August 2010; Volume 2, pp. 19–23.
41. Jones, G.; Bhanu, B. Recognition of articulated and occluded objects. *IEEE Trans. Pattern Anal. Mach. Intell.* **1999**, *21*, 603–613. [[CrossRef](#)]

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.