Hindawi
Journal of Mathematics
Volume 2024, Article ID 9865089, 1 page
https://doi.org/10.1155/2024/9865089

Retraction

Retracted: Few Samples of SAR Automatic Target Recognition Based on Enhanced-Shape CNN

Journal of Mathematics

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] M. Huang, F. Liu, and X. Meng, "Few Samples of SAR Automatic Target Recognition Based on Enhanced-Shape CNN," *Journal of Mathematics*, vol. 2021, Article ID 9141023, 16 pages, 2021.

Hindawi
Journal of Mathematics
Volume 2021, Article ID 9141023, 16 pages
https://doi.org/10.1155/2021/9141023

Research Article

Few Samples of SAR Automatic Target Recognition Based on Enhanced-Shape CNN

Mengmeng Huang, Fang Liu, and Xianfa Meng

Automatic Target Recognition Key Lab, National University of Defense Technology, Changsha, Hunan Province 410005, China

Correspondence should be addressed to Fang Liu; smartlf@sina.com

Received 29 October 2021; Accepted 25 November 2021; Published 23 December 2021

Academic Editor: Naeem Jan

Synthetic Aperture Radar (SAR), as one of the important and significant methods for obtaining target characteristics in the field of remote sensing, has been applied to many fields including intelligence search, topographic surveying, mapping, and geological survey. In SAR field, the SAR automatic target recognition (SAR ATR) is a significant issue. However, on the other hand, it also has high application value. The development of deep learning has enabled it to be applied to SAR ATR. Some researchers point out that existing convolutional neural network (CNN) paid more attention to texture information, which is often not as good as shape information. Wherefore, this study designs the enhanced-shape CNN, which enhances the target shape at the input. Further, it uses an improved attention module, so that the network can highlight target shape in SAR images. Aiming at the problem of the small scale of the existing SAR data set, a small sample experiment is conducted. Enhanced-shape CNN achieved a recognition rate of 99.29% when trained on the full training set, while it is 89.93% on the one-eighth training data set.

1. Introduction

High-resolution radar images in range and azimuth can be obtained by Synthetic Aperture Radar (SAR), which includes synthetic aperture principle, pulse compression technology, and signal processing technology. Compared with optical and infrared sensors, SAR has the advantages of day-and-night, all-weather, and the ability to penetrate obstacles such as clouds and vegetation [1–6]. With the increasing SAR imaging resolution, SAR has been diversely utilized in military and civilian fields, such as marine, land monitoring [7], and weapon guidance [8]. Therefore, SAR automatic target recognition (SAR ATR) is becoming a meaningful and challenging research field.

The MIT Lincoln Laboratory proposed to divide SAR ATR into three subsystems: detection, discrimination, and classification [9]. The task of target detection is to determine whether the image contains the target of interest and find the target's position in the image. In the discrimination stage, a discriminator is designed to solve a two-class (target and clutter) classification problem, and the probability of false alarm can be significantly reduced. And then the true target is categorized in the classification and recognition stage.

This paper only focuses on the classification and recognition stage and does not include detection and discrimination. There are three mainstream methods for recognition: template-based, model-based, and deep learning. For template matching, the test sample is matched with certain matching criteria from the template library, which is constructed from the labeled training set [10, 11]. Template-based method is simple but needs to build a large number of template libraries, and the quality of the template library has a great influence on the recognition results.

Due to the unrobustness of the template matching method, a model-based method is proposed. The method extracts the effective features of the training samples and test samples, and then the features extracted from SAR images are fed into the classifier for recognition [12–15]. The features of SAR images primarily include geometric features, transformation features, and electromagnetic features. The geometric features describe the shape and structure of target, such as contour, edge, size, and area. Principal component

analysis (PCA) [16], kernel principal component analysis (KPCA) [17], linear discriminant analysis (LDA) [18], independent component analysis (ICA) [19], and other means are all transformation features that are also applied for SAR target recognition. Due to the unique mechanism of SAR imaging, SAR images have the unique electromagnetic features [20, 21] including polarization mode and scattering centers. After feature extraction, the classifiers are necessary for feature. K-nearest neighbor (K-NN), support vector machine (SVM), and sparse representation-based classification (SRC) are frequently used as classifiers in SAR recognition.

While deep learning is well applied in various fields over years, a great quantity of deep learning methods have also emerged in SAR ATR. Chen et al. [22] proposed that the fully connected layer in convolutional neural network (CNN) is replaced with convolutional layer, which effectively suppresses the overfitting problem and reduces the number of parameters. Since the SAR images are highly sensitive to azimuth angle, Zhou et al. [23] combined three continuous azimuth images of the same target as a pseudocolor image inputting, which are input into CNN. Wang et al. [24] designed a multiview convolutional neural network and long short term memory network (CNN-LSTM) to extract and fuse the features from different adjacent azimuth angles. Zhang et al. [25] utilized CNN with CBAM, which is an attention mechanism to improve recognition rate. The deep-learning method can extract the deep semantic information of the target. Compared with the model-based method, it does not need to extract features manually and has achieved a high recognition rate in the field of SAR target recognition.

More recently, there is a viewpoint that CNN, which is different from human, is more inclined to learn the texture and surface features of the target but pays less attention to deep semantic features such as contour and shape. Contour and shape are the most reliable information in human and biological vision. Geirhos et al. [26] demonstrated that Image Net-trained CNNs are strongly biased towards recognizing textures rather than shapes, which is in stark contrast to human behavioral evidence and reveals fundamentally different classification strategies. Hermann et al. [27] indicated that, on out-of-distribution test sets, the performance of models that like to classify images by shape rather than texture is better than baseline.

Therefore, this paper proposes an enhanced-shape CNN, whose network structure is shown in Figure 1. First, the enhanced-shape CNN strengthened the shape features of the target at the input, constructing a three-channel pseudocolor image as data set, so that the convolutional neural network can tend to pay more attention to target shape. Second, the pooling commonly use in CNNs is maximum pooling and average pooling, and the target information is easily lost when downsampling the feature maps. Thus, we use the SoftPool [28] instead of max pooling to improve the network. Meanwhile, in the above literatures, some attention mechanisms combined with CNNs have been applied to SAR recognition. The channel attention module mechanism, i.e., Squeeze-and-Excitation (SE) module [30], can effectively increase the channel weights that are beneficial for recognition and suppress feature that are less useful in CNNs. However, SE module distributes channel weights more evenly in target recognition, such that there is essentially the same as CNN, as noted in paper [29]. Therefore, SoftPool is utilized by replacing global pooling, which can obtain unbalanced channel weights. Third, it is still troublesome to acquire SAR image data sets with relatively rich conditions of imaging, despite the fact that the acquisition of high-resolution SAR images has become easier. Over these years, a great quantity data sets of SAR ships and vehicles have emerged, but their resolution is not enough to be recognized; hence, the data sets are used for detection. At present, most research of SAR target recognition is based on the Moving and Stationary Target Acquisition and Recognition (MSTAR) [31] data set. From the perspective of less samples, this paper designs experiments to verify that this method has a higher recognition rate compared to existing methods under limited data sets.

The main contributions of this paper are as follows:

(1) Constructing a three-channel pseudocolor image, which is formed by extracting the features of the target and shadow from the original SAR data set, filtering the original SAR images, and the original SAR images. The pseudocolor three-channel images are input to the CNN, enhancing the model to use the shape information of the image.

(2) Improving the pooling of the network and the global pooling of the attention module. Using SoftPool in the network can increase the information of the feature map during the pooling. At the same time, the pooling in the SE module is improved to make the weight distribution of the channel more different, instead of balance.

(3) Training in the full training set, one-half of the training set, one-quarter of the training set, and one-eighth of the training set and testing in full test set based on the MSTAR data set. It is proved that the method proposed in this paper can obtain a higher recognition rate with a few samples.

The remainder of this paper is organized as follows: Section 2 describes the principles of the method, including the extraction method of target and shadow, the principle of lee filter, and the fusion of three-channel pseudocolor image. and a novel pooling method (SoftPool), the Squeeze and Excitation module and Enhanced SE module. Section 3 presents the experimental results to validate the effectiveness of the proposed network, and Section 4 concludes the paper.

2. Methodology

In this section, we will describe some of the principles and structures used in our model.

2.1. Extraction of Target and Shadow. Unlike optical images, SAR images are side-view imaging, so there are shadows in the image in addition to the target. The shadow is the result

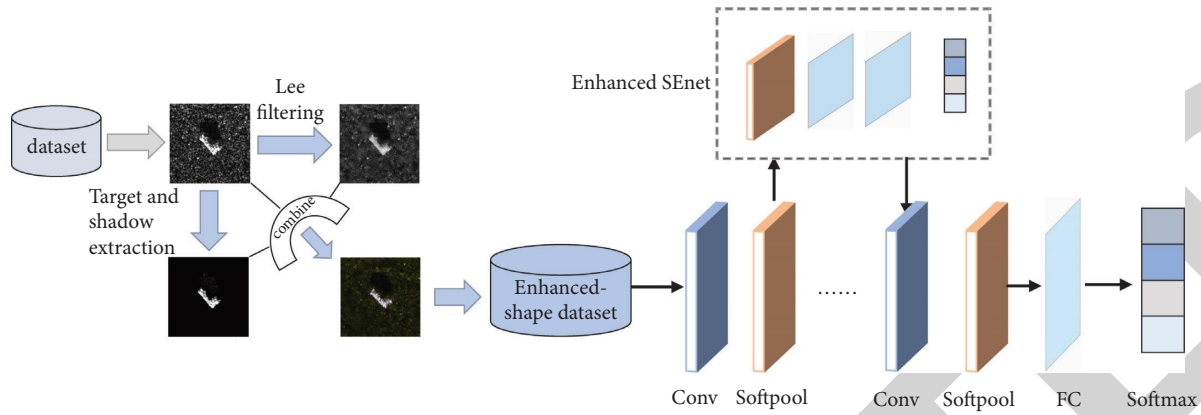


Figure 1: Structure of the enhances-shape network.

of the mutual coupling between the target and the background environment under a specific radar line of sight, and its shape reflects the physical size and shape distribution of the target, so combining joint features of the target and shadow is more helpful for the recognition.

There are many existing segmentation algorithms to extract target and shadow. The focus of our model is not the segmentation algorithm; therefore, the simplest threshold method is used to segment the target and the shadow area. Our threshold setting is based on the threshold proposed by the paper [32]. The main steps are as follows:

(1) Equalize the original SAR image histogram;

(2) Use mean filtering to smooth the result of step 1, and transform the gray dynamic range to [0, 1];

(3) Set the thresholds of the shadow and target area to 0.2 and 0.8, the pixels greater than 0.8 are the target area, and those less than 0.2 are the shadow areas;

(4) Remove the area of total pixels less than 25 to reduce the influence of background noise;

(5) Utilize the morphological closing operation to connect the target area and the shadow area, which obtain a smooth target and shadow contour.

It can be seen that the simple threshold method can achieve good segmentation results and remove a lot of background noise and clutter. However, in real world situations, the common segmentation algorithm may not be able to segment the target and the shadow well, so we set the thresholds 0.1 and 0.9, and 0.3 and 0.7, respectively, to verify that a slightly biased segmentation algorithm works better.

Figure 2 demonstrates the target and shadow images obtained with different segmentation thresholds. (a) is the original image. (b) describes the morphological image of the target and shadow when the threshold is set to 0.8 and 0.2. The target and shadow extracted in (c) are relatively complete, and the pixel value of the shadow is too low to be clear. Relatively, the target area extracted in (d) is redundant, and in (e) it is incomplete.

2.2. Lee Filtering. Due to its special imaging mechanism, SAR images contain more coherent speckle noise. After filtering the SAR image, the shape characteristics of the target can be enhanced, and the texture, especially the interference of noise, can be reduced.

For speckle noise, many filtering methods for the speckle noise of SAR images have been proposed. Our model utilizes lee filtering, which is a classic SAR filtering strategy. The two key aspects of noise suppression are, on the one hand, establishing a true backscatter coefficient estimation mechanism, and on the other hand, formulating a selection plan for pixel samples in homogeneous regions.

Lee filtering is one of the typical methods of image speckle filtering using the local statistical characteristics. It is based on a fully developed speckle noise model. First, a window of a certain length is selected as the local area. Then, it is assumed that the prior mean $\overline{x}$ and variance $\mathrm{var}(x)$ can be calculated by calculating the local mean $\overline{y}$ and the variance $\mathrm{var}(y)$.

$$\hat{x} = a\overline{x} + by, \tag{1}$$

$$\begin{aligned} a &= 1 - \frac{\mathrm{var}(x)}{\mathrm{var}(y)}, \\ b &= \frac{\mathrm{var}(x)}{\mathrm{var}(y)}, \end{aligned} \tag{2}$$

$$\hat{x} = \overline{y} + b(y - \overline{y}), \tag{3}$$

$$\begin{aligned} \mathrm{var}(x) &= \frac{\mathrm{var}(y) - \sigma_v^2 \overline{y}^2}{1 + \sigma_v^2}, \\ \sigma_v^2 &= \frac{1}{N}, \end{aligned} \tag{4}$$

where y signifies the value in the selected window. The window size N selected in this paper is 7.

It can be observed from Figure 3 that the speckle noise in the image is significantly reduced, and the texture features of the target and shadow parts are reduced, but the contour shape is more obvious after lee filtering.

2.3. Fusion. Typically, SAR images are gray images. When recognizing SAR images with CNN, the gray-scale image is generally converted into a three-channel image input. In this

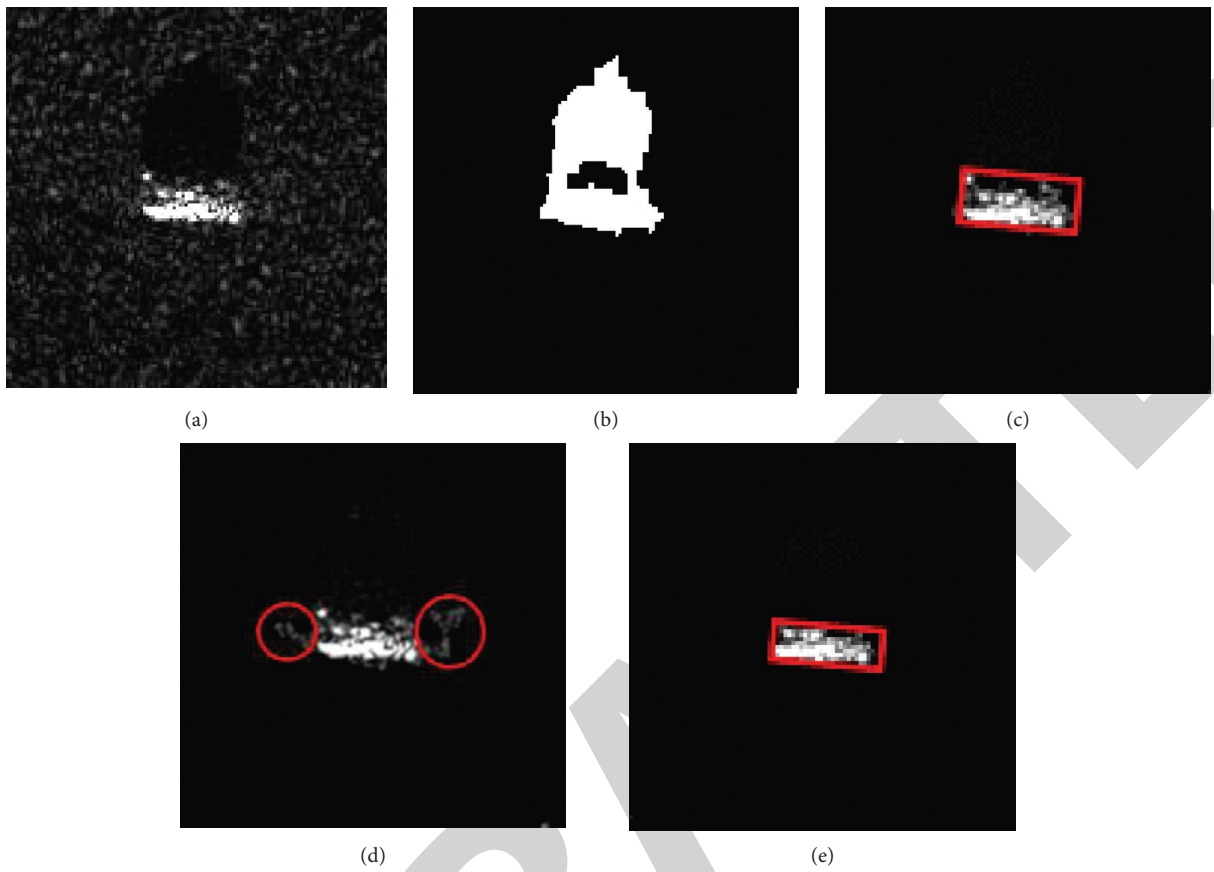


Figure 2: The segmentation of the target and shadow. (a) The original image. (b) Morphological image of target and shadow. (c) Target and shadow when setting the thresholds 0.2 and 0.8. (d) Target and shadow when setting the thresholds 0.1 and 0.8. (e) Target and shadow when setting the thresholds 0.3 and 0.7.

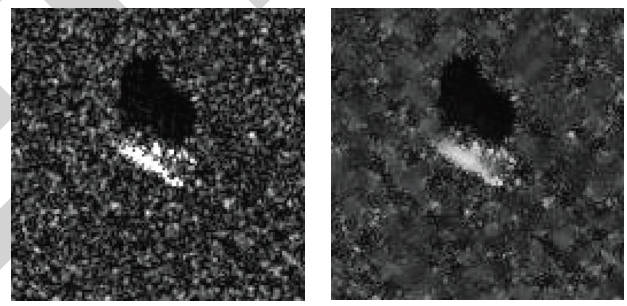


Figure 3: (a) The original image. (b) Image after lee filtering.

paper, the original image is combined with the image of target and shadow and the filtered image in RGB mode to form a three-channel pseudocolor image, as shown in Figure 4. The original image can contain complete target information including shape, contour, and texture, while the image of target and shadow and filtered image can enhance the target shape characteristics. Using pseudocolor images as network input can acquire global information and deep semantic information instead of focusing on texture information.

2.4. SoftPool. The SoftPool is used by us in the network to reduce the loss of target information. Pooling is used in CNN to reduce the size of feature maps to achieve local space invariance and increase convolutional receptive fields. At present, the most commonly used in neural networks is max pooling and average pooling, which will lose the information mapped in the feature map. Therefore, paper [28] proposed SoftPool to reduce the loss of information, while limiting the calculation and memory overhead.

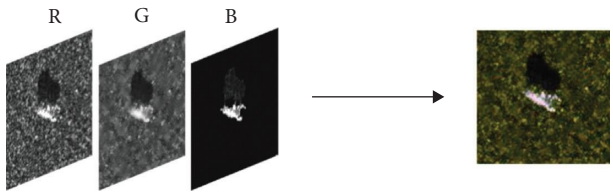


Figure 4: The fusion of multifeature images.

The SoftPool is differentiable. For the pooling kernel $k * k$, we suppose that the output of the pooling operation is $\widetilde{a}_R$, the corresponding gradient is $\Delta\widetilde{a}_i$, R is the maximum approximation in the activation area, and each activation a_i with index i corresponds to a weight w_i. The weight w_i is the ratio of the natural index of the activation to the sum of the natural indices of all activations in the neighborhood R:

$$w_i = \frac{e^{a_i}}{\sum_{j\in R} e^{a_j}}. \tag{5}$$

The weight together with the corresponding activation value is used as a nonlinear transformation. Higher activation is more dominant than lower activation. The output value after the SoftPool is obtained by summing all the weighted activation criteria in the kernel neighborhood R:

$$\widetilde{a} = \sum_{i\in R} w_i * a_i. \tag{6}$$

In the training update phase of SoftPool, the gradient update is proportional to the weight calculated in the forward propagation process, namely, $\nabla a_i = w_i * \nabla\widetilde{a}$. It is realized that the gradient update of the smaller activation is smaller than the gradient update of the larger activation. The forward propagation and backward update of SoftPool are shown in Figure 5.

Compared to max pooling and average pooling, the SoftPool can balance the influence of average pooling and max pooling, while average pooling reduces the effect of activations in the area, and max pooling selects only the highest activation in the area. For SoftPool, all activations in this area contribute to the final output, and higher activations dominate the lower activations. Therefore, in the pooling of CNN, a larger activation value has a greater impact on the output, and the significant details of the feature map can be retained to the greatest extent.

Figure 6 gives the effect of different pooling. The first column is the original image, the second column is the image after max pooling, the third column is the image after average pooling, and the fourth column is the image after SoftPool. The comparison shows that the max pooling activates the pixel points with large gray values in the region, highlighting the target, as well as highlighting scattered noise. The average pooling approximates filtering, reducing the effect of noise, but weakening the structural shape information of the target with it. SoftPooling, on the other hand, retains the relatively intact structural information of the target while removing the effect of scattered noise, making the shape more prominent.

2.5. SE Module and Enhanced SE Module. The core of typical CNN is the convolution operator, and the input feature map is mapped to the new feature map through the convolution kernel. In the convolutional layer, the feature maps of the previous layer are considered to have the same weight for the next layer, but research [30] illustrates that this is not the case. The equal mechanism limits the convolutional neural network to obtain more information. Therefore, paper [30] proposed SE module, which focuses on the relationship between channels and hopes that the model can automatically learn the importance of different channel features.

The network structure of SE module is shown in Figure 7. For input feature map tensor X: $X \in R^{W\times H\times C}$, where $W \times H$ represents the length and width of the feature map, and C represents the number of input channels, and SE module performs a squeeze operation on X to obtain the channel-level global features and then performs an excitation operation on the global features to learn the relationship between each channel and get the weights of different channels. Finally, the output feature map $\widetilde{X}$ is calculated by multiplying the weights and the input feature map X.

As mentioned above, the SE module consists of two steps: squeeze and excitation. For the squeeze F_{sq}, global average pooling is applied to encode the entire spatial feature on a channel as a global feature. The input of average pooling is the feature map tensor X, and the output after a squeeze operation is $z_c \in R^C$, denoting the cth value in the vector z. The mapping relationship between X and z_c is as follows:

$$z_c = F_{sq}(x_c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_c(i,j), \tag{7}$$

where x_c represents the feature map tensor of the cth channel of input X. The squeeze operation gets the global description feature, and then the excitation operation is performed.

$$s = F_{ex}(z,W) = \sigma(g(z,W)) = \sigma\left(W_2\mathrm{ReLU}\left(W_1 z\right)\right), \tag{8}$$

where $W_1 \in R^{(C/r)\times C}$, $W_2 \in R^{C\times(C/r)}$, r is a fixed hyperparameter, σ is the sigmoid activation function, and s indicates the learning weight of different channels. The first FC layer plays the role of dimensionality reduction, and the final FC layer restores the feature map to the original dimensions. After squeeze and excitation, the channel weight is obtained, and finally, the weight is multiplied by the original feature tensor.

$$\widetilde{x}_c = F_{\mathrm{scale}}\left(x_c, s_c\right) = s_c \cdot x_c, \tag{9}$$

where s_c represents the weight of x_c and $F_{\mathrm{scale}}(x_c, s_c)$ represents the product of them.

Essentially, the SE module performs attention operations in the channel dimension. This attention mechanism allows the model to pay more attention to the channel features with the most information, while suppressing those unimportant channel features. However, this advantage is not directly reflected in the experiment on the SAR data set MSTAR. It can be seen from the paper [29] that the channel weights

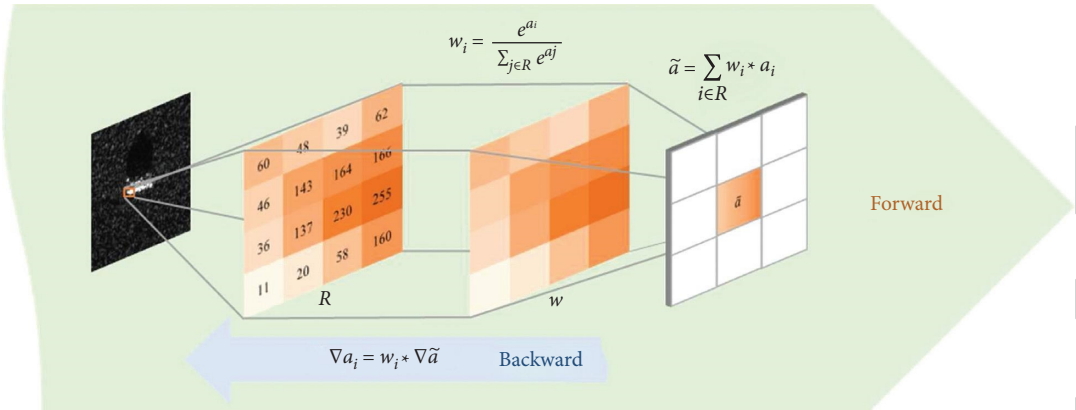


Figure 5: The green part represents forward propagation. The output of pooling is the product of the weights and activation values in region R. The blue part represents backward propagation, and the update of the activation value is also related to the weight.

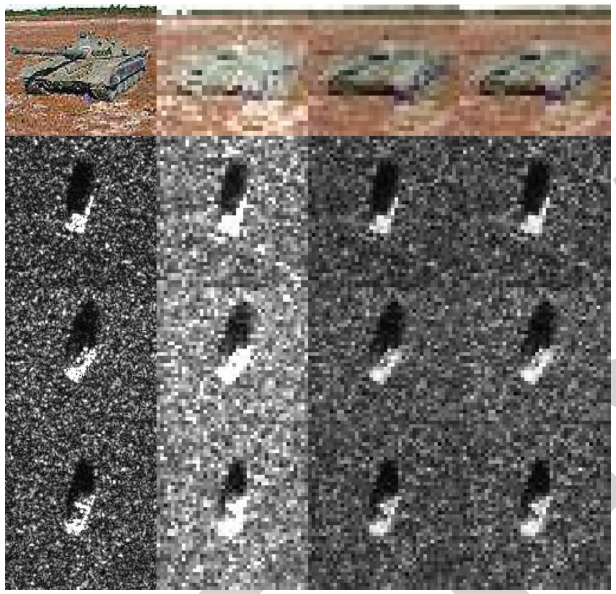


Figure 6: Results for different pooling.

calculated by the SE module are close to 1, which does not reflect the importance of the channel.

Global pooling performs max pooling or average pooling on the entire feature map to obtain a $1 \times 1 \times C$ vector, but this also will lose feature information. Therefore, we think of replacing the global pooling of the SE module with SoftPool to ensure that the dominant feature map has a high weight. Figure 8 gives the calculation results of the two feature matrices under global pooling and soft pooling. (1) can represent the edge information of the target and contains more information amount than (2), but both matrices have the same calculation result, both 4, under global pooling, and cannot distinguish the importance of the channels. When the weight matrix is multiplied with the feature matrix after using soft pooling, the output of (1) is 5.724, and the output of (2) is 3.69, which can make the feature matrix containing more information have greater channel weights and solve the problem of uniform weight distribution of SE module.

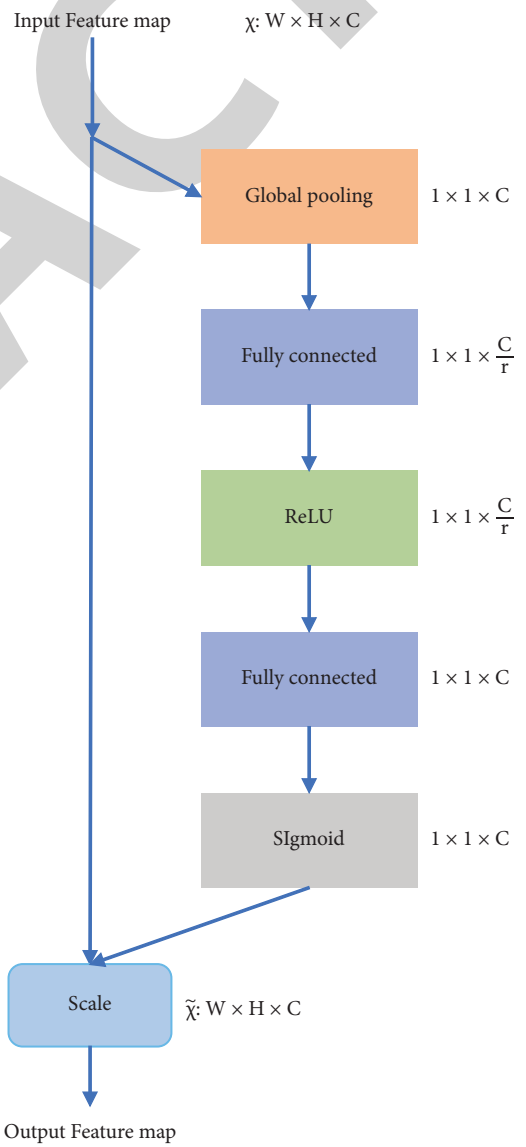


Figure 7: The fusion of multifeature images. The structure of the SE module [30].

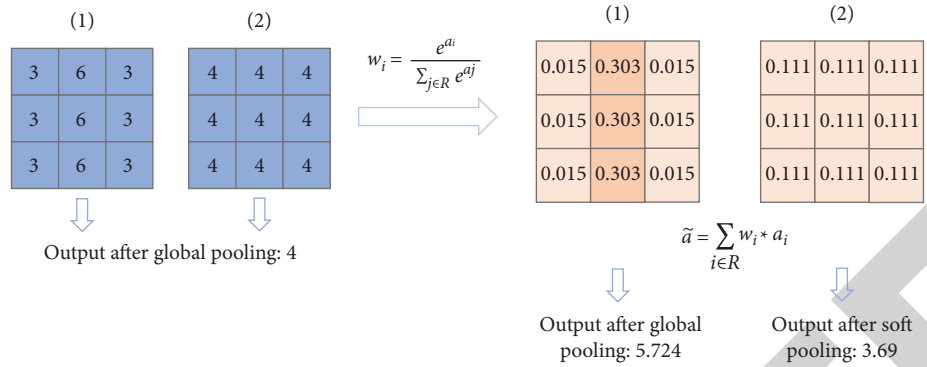


Figure 8: Calculation of global average pooling and soft pooling.

2.6. Analysis with Channel-Wise Activation Maps. Because the deep network will easily lead to overfitting when doing training and recognition with few samples, this paper builds a simple CNN. The structure of the network is designed as Figure 9. (a) is the basic CNN network, and (b) is the shape enhancement network used in this paper.

Figure 10 illustrates the visualization of the features map from the network using the SE module and using the SoftPool-SE module respectively. SoftPool-SEnet clearly highlights certain channels compared to the SE module.

Figure 11 shows the 16 maps of adding different modules in the first convolutional layer. Compared with feature map in (a), that in (b) obviously removes the texture information brought by the background noise and enhances the network's attention to the target's shape. The feature map in (c) adds a lot of information, where SoftPool is used in the network. The network in (d) uses ordinary SE module, but compared with feature map of (d), there are more dark pixels, and more information is lost. The bright pixels of the target in (e) are increased because of the use of enhanced SE module.

2.7. Configuration Specifics in the Enhanced-Shape CNN. The convolutional layer maps the input to a new feature map with a convolutional kernel to perform local perception of the target. Pooling layer is a subsampling to reduce trainable parameters. In order to prevent the problems such as declined convergence speed and poor generalization performance due to the different distributions of the training set and the test set, we adopted batch normalization in the network.

For all convolutional layers, the stride is set to 1, and no spatial zero padding is used in the convolution layer. Meanwhile, the activation function adopts ReLU nonlinearity. Each of the first three convolutional layers is followed by a soft pooling layer with a pooling size of 2×2 and a stride of 1. The size of the input enhance-shape image is $128 * 128$. After the first convolutional layer, where the size of convolution kernel is 5×5, the size of output feature map is $124 * 124$, and their size becomes 62×62 after the first layer of pooling layer. The 62×62 input image was filtered by convolution kernel of size 6×6 in the second convolutional layer, resulting in feature map of size 57×57. After the second pooling, the feature map becomes of size 28×28. At this time, the 28×28 feature map is input into the SoftPool-SE module, and the learning channel has different weights while the output feature map size is still $28 * 28$. The filter kernel of the third convolutional layer is of size 7×7, producing feature map of size 22×22, which becomes 11×11 after pooling and SoftPool-SE module. The convolution kernel of the last layer is $7 * 7$, which brings out 5×5 feature map. Finally, through two fully connected layers and a softmax classifier, 10 vectors are obtained, corresponding to the class probabilities.

In this paper, the loss function is cross entropy loss, and the optimization algorithm uses stochastic gradient descent, with the momentum parameter of 0.9 and the weight decay parameter of 0.005. Subsequently, the learning rate is initially 0.001 and is reduced by a factor of 0.5 after 20 epochs, where epoch denotes the number of times each example has used during training. Finally, batch size is set to 8.

3. Experiments on MSTAR Dataset

3.1. Dataset Description. The experiment data set in this paper is the MSTAR public data set, where the resolution of all images is $0.3\ \text{m} \times 0.3\ \text{m}$, and the polarization mode used is HH polarization mode. The data set contains hundreds of thousands SAR images, covering military targets of different categories, aspect angles, and depression angles, of which only a small part is publicly available. They were collected by *X* band, full aspect coverage (in the range of 0° to 360°).

The disclosed data set includes ten types of ground vehicle targets: armored personnel carrier (BMP-2, BRDM-2, BTR-60, and BTR-70); tank (T-62, T-72); rocket launcher (2S1); air defense unit (ZSU-234); truck (ZIL-131); bulldozer (D7). Figure 12 shows examples of ten types of targets and their corresponding optical images.

When the MSTAR data set is used in SAR ATR, it is often divided into standard operating conditions (SOC) and extended operating conditions (EOC). SOC means that the target configuration and serial number of the test set and training set are the same, and depression angles are different but close. EOC indicates that there is a big difference between the test set and the training set, including target configuration and image clarity.

SOC is a dataset that consists of images with an imaging condition of 17° depression angle as the training set, and 15°

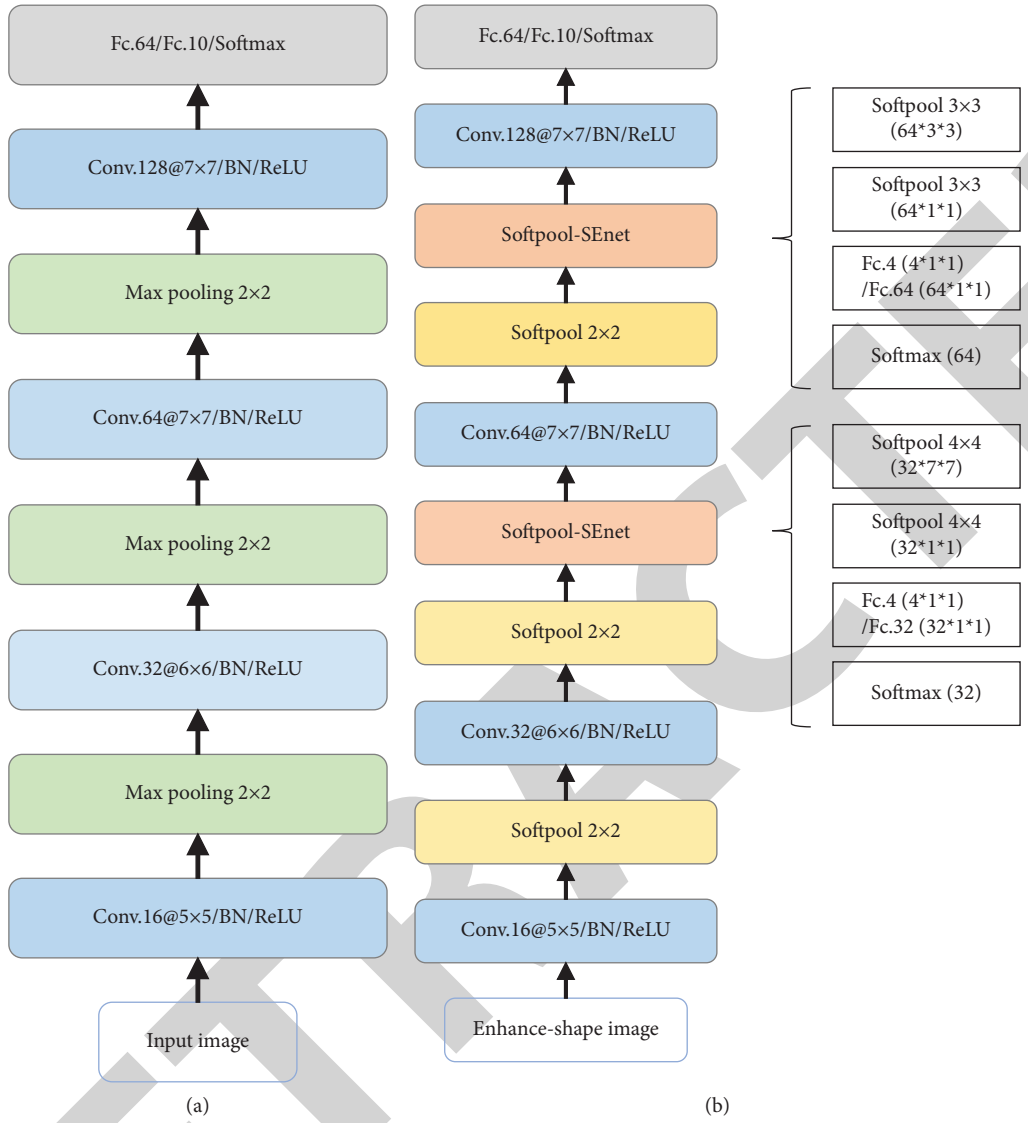


Figure 9: Network structure. Next to the network structure is the size of the feature map. (a) CNN. (b) Enhanced-shape CNN.

depression angle as the test set. The number of test and training samples for each category and the total number of samples are shown in Table 1.

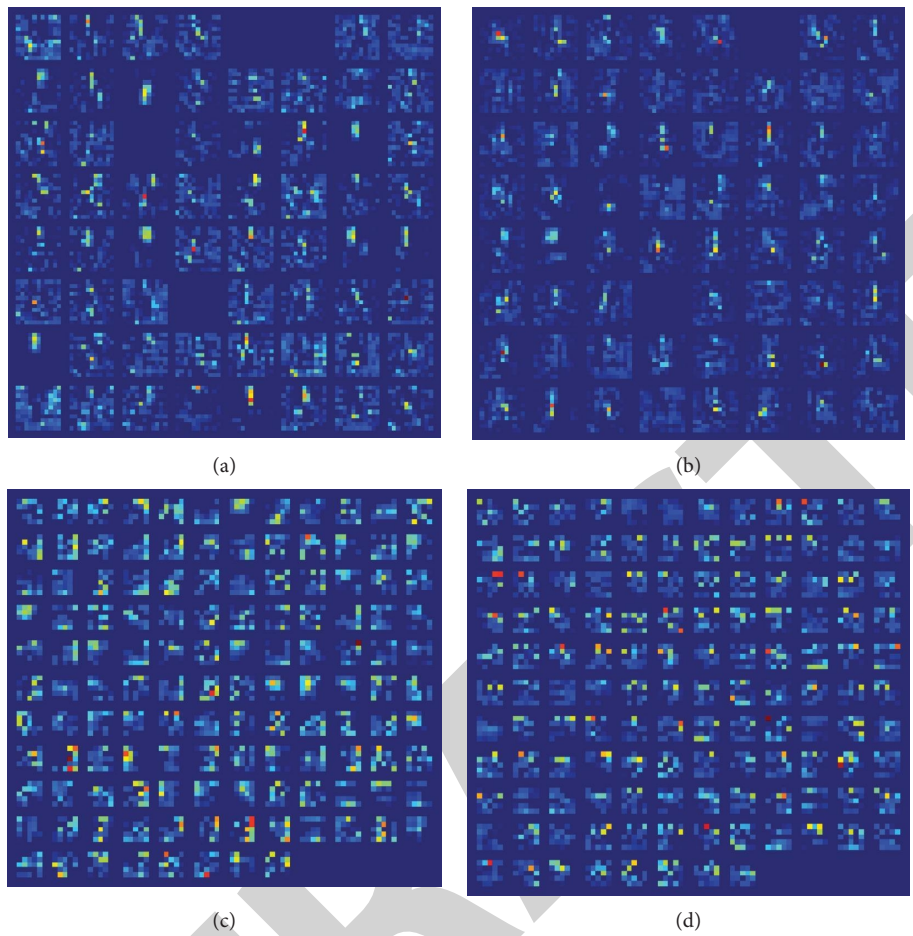
In addition to SOC dataset, we have also set up several EOC datasets. Configuration change refers to the addition or removal of some parts on the vehicle, such as whether the T72 has an oil tank behind the vehicle. In this paper, these two changes are referred to as EOC-1 and EOC-2, i.e., configuration variants and version variants. The specific information of the EOC-1 and EOC-2 data set is listed in Tables 2 and 3. The training set is BMP2, BRDM-2, BTR-70, and T72 with 17° depression, and the test set only includes variants of T72 with 15° depression and 17° depression. The training set of EOC-2 is the same as EOC-1. The test set contains variants of T72 and BMP-2.

Moreover, the image signal-noise ratio of MSTAR is as high as 30 dB, but most images in actual situations contain noise. We set EOC-3 dataset, which adds noise to the MSTAR data [33] to simulate a noisy situation. The method of adding noise is as follows:

$$\mathrm{SNR} = 10\log_{10} f\left[\frac{\mathrm{var}(\text{original image})}{\mathrm{var}(\text{error image})}\right], \tag{10}$$

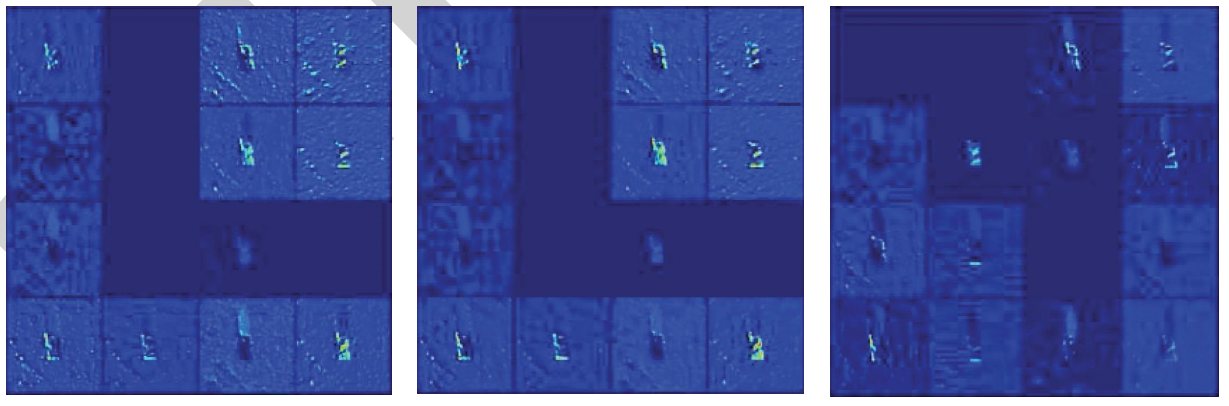
where var is a variance operator. The result is shown in Figure 13.

3.2. Result of SOC. Table 4 shows a confusion matrix, whose row represents the actual target category, and column represents the predicted target category. It is observed that the recognition rate of all targets has reached more than 96%, and the overall recognition rate has reached 99.29%. The recognition rate of each method is listed in Table 5. Compared with other methods, our method got the highest recognition rate, verifying the effectiveness of the proposed method.



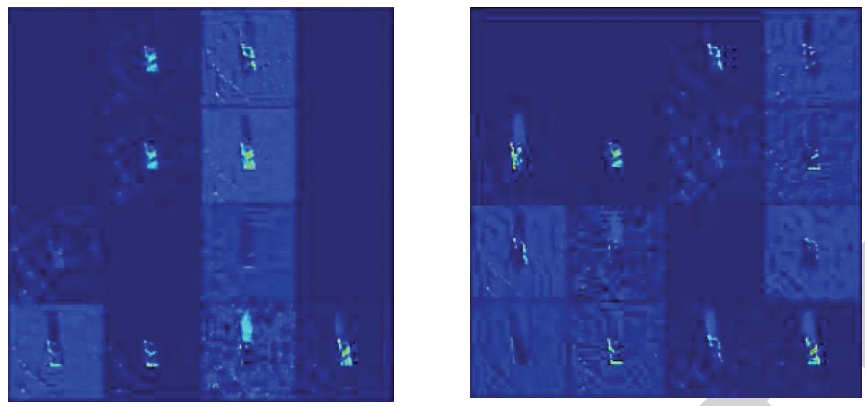
(a) (b) (c) (d)

Figure 10: Visualization of feature maps: (a) output by first SE module; (b) output by first SoftPool-SE module; (c) output by second SE module; (d) output by second SoftPool-SE module.



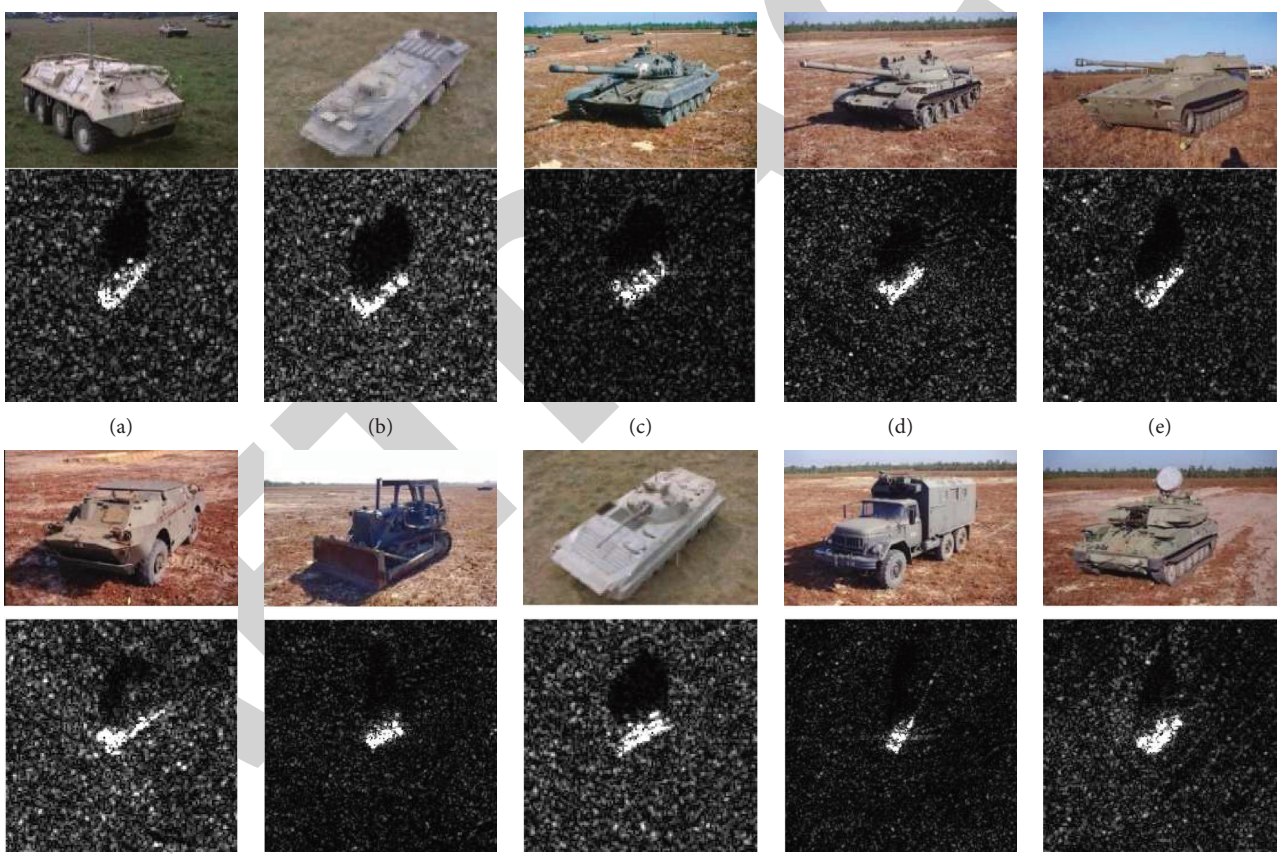
(a) (b) (c)

Figure 11: Continued.



(d) (e)

Figure 11: (a) The basic network. (b) Basic network using enhanced-shape data set. (c) SoftPool is used on the basic network, where enhanced-shape dataset is inputted. (d) SoftPool and SE module are used on the basic network, where enhanced-shape data set is inputted. (e) SoftPool and enhanced SE module are used on the basic network, where enhanced-shape data set is inputted.



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Figure 12: Optical image (top) and SAR image (bottom) of ten types of vehicle targets.

In order to verify that enhanced-shape CNN can also achieve better recognition on a few-sample data set, we set training sets of 100%, 50%, 25%, and 12.5%, respectively, while the size of the testing set remains the same to calculate the recognition rate. The comparison network we used is the basic CNN network pointed out in Figure 9.

As shown in Table 6, in the case of the full training set, the enhanced-shape CNN has reached a recognition rate of more than 99%, which is not much improvement compared to the basic CNN. When we only use 50%, 25%, and 12.5% training sets separately, there will be a corresponding increase of 1.18%, 2.23%, and 4.56%. Compared with the

Table 1: Number of training and test samples for SOC.

Class	Train		Test	
	Depression	Number	Depression	Number
BRDM2	17°	298	15°	274
BTR60	17°	256	15°	195
T72	17°	232	15°	196
2S1	17°	299	15°	274
D7	17°	299	15°	274
BMP2	17°	232	15°	196
ZIL131	17°	299	15°	274
ZSU23/4	17°	299	15°	274
BTR70	17°	233	15°	196
T62	17°	299	15°	273

Table 2: Number of training and test samples for EOC-1 (configuration variants).

	Train			Test	
Class	Depression	Number	Class	Depression	Number
BMP2(9563)	17°	233	T-72(S7)	15° 17°	419
BRDM-2(E71)	17°	298	T-72(S32)	15° 17°	572
BTR-70(c71)	17°	233	T-72(S62)	15° 17°	573
T-72(132)	17°	232	T-72(S63)	15° 17°	573
			T-72(S63)	15° 17°	573

Table 3: Number of training and test samples for EOC-2 (version variants).

	Train			Test	
Class	Depression	Number	Class	Depression	Number
BMP2(9563)	17°	233	T-72(812)	15° 17°	426
BRDM-2(E71)	17°	298	T-72(A04)	15° 17°	573
BTR-70(c71)	17°	233	T-72(A05)	15° 17°	573
T-72(132)	17°	232	T-72(A07)	15° 17°	573
			T-72(A10)	15° 17°	567
			BMP-2/9566	15° 17°	428
			BMP-2/C21	15° 17°	429

experimental results of other methods under small sample data sets, the method proposed in this paper is also far superior to other methods.

Due to the standard of the MSTAR data set, it is relatively simple to segment the target and the shadow area, but the actual situation is often more complicated, so the target and the shadow may not be completely segmented. In order to verify the robustness of our algorithm, we can make a slight deviation when doing threshold segmentation. The deviation image has been given in Figures 2(c) and 2(e), corresponding to set the segmentation threshold to 0.1 and 0.9, and 0.3 and 0.7, respectively.

It can be seen in Figure 14 that even when the segmentation algorithm is not ideal, our method still has a higher recognition rate than CNN on a small number of samples. The shape and shadow area are extracted to highlight the target and enhance the network's learning of target information. Therefore, even when the segmentation algorithm is slightly deviated, it can still achieve better recognition results than the original data.

3.3. Result of EOC. This paper tests the recognition accuracy on two types of data sets, EOC-1 and EOC-2, to further test the effectiveness of the proposed method for refined recognition. The tested confusion matrix is shown in Tables 7 and 8. According to the experimental results, the methods proposed on the EOC-1 and EOC-2 data sets both have achieved good recognition results. The recognition rate reached 99.3% under EOC-1, while it reached 98.85% under EOC-2. It illustrates that when the target changes slightly, such as the addition or removal of fuel tanks, the network can achieve better recognition results.

Figure 15 shows the comparison curves of the recognition rates obtained by the two networks on training sets of different sizes under different noises. It can be seen that our proposed method has achieved a higher recognition rate than ordinary CNN on different data quality. When the signal-to-noise ratio is −5 dB and −10 dB, the recognition rate in enhanced-shape CNN, which uses the 12.5% training set, is improved by nearly 20% compared to that in CNN.

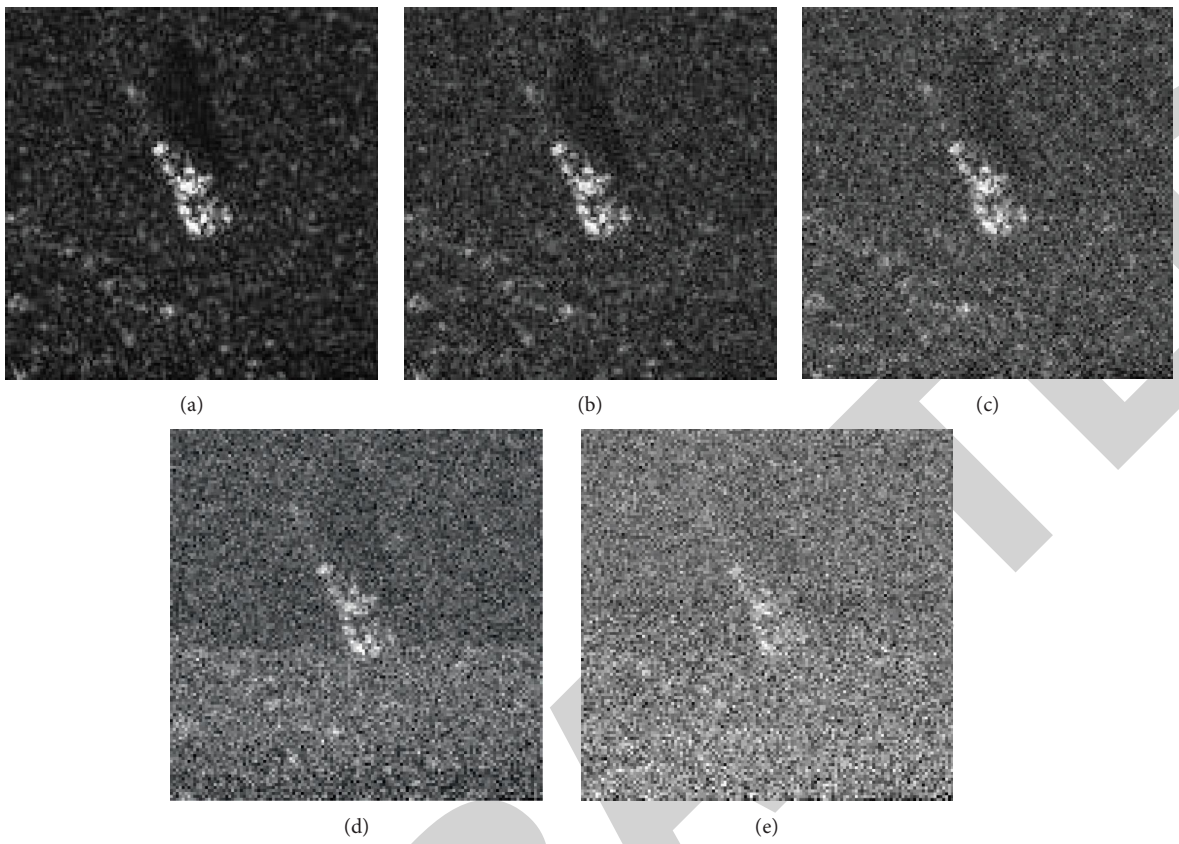


Figure 13: Image after adding different noises.

Table 4: Confusion matrix of Enhanced-shape CNN.

Class	BRDM2	BTR60	T72	2S1	D7	BMP2	ZIL131	ZSU23/4	BTR70	T62	P_{cc} (%)
BRDM2	**272**	0	0	0	0	0	1	1	0	0	99.27
BTR60	4	**189**	0	1	0	0	0	1	0	0	96.92
T72	0	0	**196**	0	0	0	0	0	0	0	100
2S1	0	0	0	**271**	0	0	0	0	0	3	98.91
D7	0	0	0	0	**272**	0	2	0	0	0	99.27
BMP2	0	0	2	0	0	**194**	0	0	0	0	98.98
ZIL131	0	0	0	0	0	0	**274**	0	0	0	100
ZSU23/4	0	0	0	0	1	0	0	**273**	0	0	99.64
BTR70	0	0	0	0	0	0	0	0	**196**	0	100
T62	0	0	0	1	0	0	0	0	0	**272**	99.63
Overall											99.29

Table 5: Performances of different methods.

Method	Accuracy (%)
ACS [34]	95.54
CNN-LSTM [24]	98.78
Multiview-DCNN [35]	98.52
CHU-Net [36]	99.09
A-Convnet [22]	99.13
Enhanced-shape CNN	99.29

Table 6: Recognition rate on different sizes of training set.

Method	Training dataset size			
	100%	50%	25%	12.5%
CNN	99.11	97.65	95.92	85.37
Enhanced-shape CNN	99.29	**98.83**	**98.15**	**89.93**
ARGN [37]	98	97.28	—	—
DS-AE Net [38]	**99.30**	98.06	95.42	—
TAI-SARNET [39]	97.97	93.22%	88.69	76.27

Table 7: Confusion matrix of Enhanced-shape CNN under EOC-1 (configuration variants).

Class	Variants	BMP-2	BTR-70	T-72	BRDM-2	P_{cc} (%)
T-72	S7	6	7	406	0	96.897
	A32	2	0	570	0	99.65
	A62	0	0	573	0	100
	A63	0	0	572	1	99.83
	A64	3	0	570	0	99.48
Total						99.30

Table 8: Confusion matrix of Enhanced-shape CNN under EOC-2 (version variants).

Class	Variants	BMP-2	BTR-70	T-72	BRDM-2	P_{cc} (%)
BMP-2	9566	414	5	7	2	96.73
	c21	420	0	7	2	97.90
T-72	812	2	11	413	0	96.95
	A04	0	0	572	1	99.83
	A05	0	0	573	0	100
	A07	1	0	571	1	99.65
	A10	2	0	565	0	99.65
Total						98.85

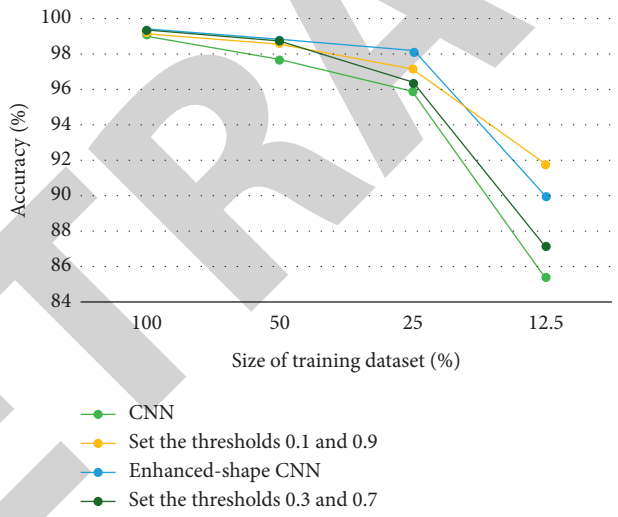


Figure 14: Recognition results under different segmentation thresholds.

3.4. Ablation Experiment. In order to verify the influence of different modules on the performance of the model, ablation experiments are also carried out in this paper. We set up different inputs, respectively, selecting the original image, filtering the image, and extracting the target and shadow image, and the fusion image, to verify that the data enhancement of the fusion of multiple features is effective.

Figure 16 shows the recognition rates obtained for several inputs. The recognition rate of a single filtered image and segmented image is lower than that after fusion. When only the segmented image is input, it is found that the recognition rate is lower than that of the original image input. This is because we extract the target and shadow area only to strengthen the network's attention to the target and shadow. If only the target and shadow are input, the target information will be incomplete owing to the segmentation algorithm, so the recognition rate without inputting the original data is high.

Figure 17 shows the recognition rate using a single module. It can be seen that the different modules used in this

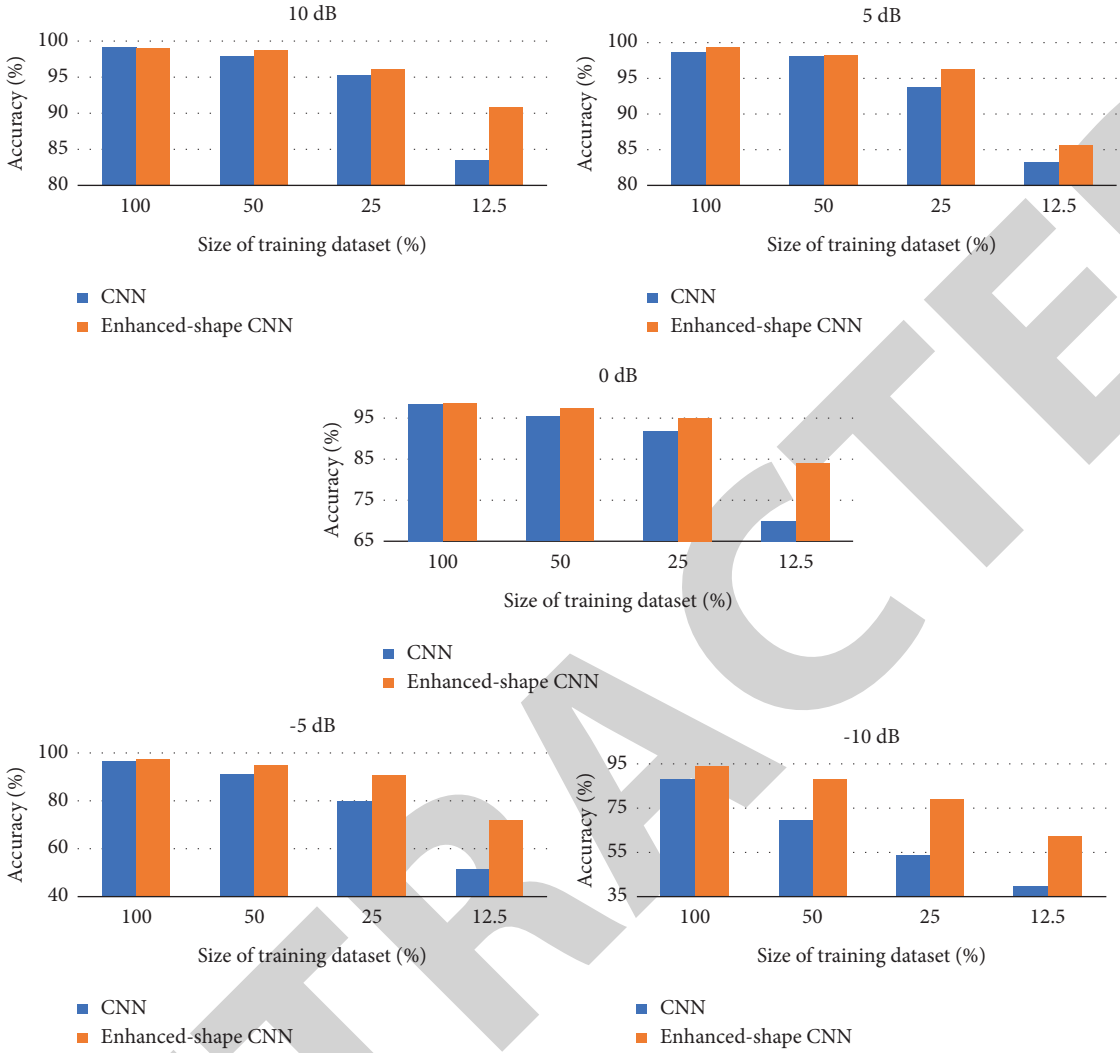


Figure 15: Performance comparison of enhanced-shape CNN and CNN under different signal-to-noise ratios.

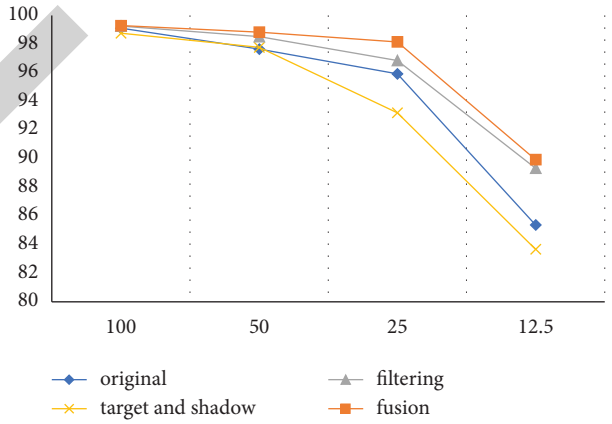


Figure 16: Recognition rate under different input.

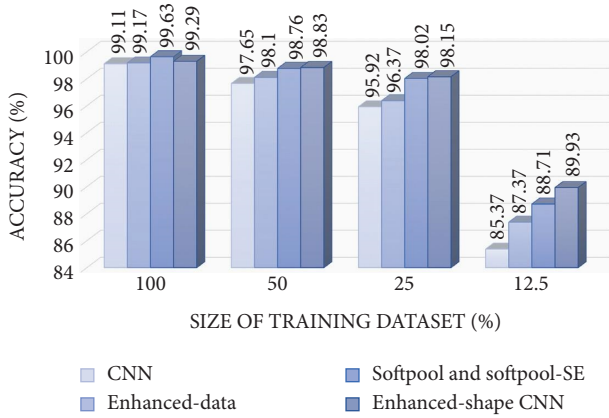


Figure 17: Results of ablation experiments.

paper have an effect on the recognition accuracy of the model.

4. Conclusions

SAR ATR has become an important and promising field of remote sensing image processing. This paper proposed a method from the perspective of shape enhancement with filtering and enhancing target area at the input and synthesizing to strengthen the connection between channels. Simultaneously, the information loss due to ordinary pooling is reduced by the application of SoftPool in CNN. Moreover, the SE module has been improved to highlight the prominent channels for recognition results. As a result, more target information is obtained on a few samples. The experiments verified the accuracy of proposed method, which can achieve an accuracy of 99.29% on ten types of targets, and when the segmentation effect is not good, which is closer to the actual situation, it also has higher performance than CNN. This paper also proved the robustness of the method under noise. In the case of varying degrees of noise, the proposed method is greatly improved compared to CNN when there are few samples. The basic approach proposed in this paper can continue in the future to explore the method of balancing texture features and shape features and guide the directional training of the network based on the attention mechanism.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no competing interest.

Acknowledgments

The authors did not receive specific funding.

References

[1] X. Bai, R. Xue, L. Wang, and F. Zhou, "Sequence SAR image classification based on bidirectional convolution-recurrent network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9223–9235, 2019.

[2] H. Xu, Z. Yang, M. Tian, Y. Sun, and G. Liao, "An extended moving target detection approach for high-resolution multichannel SAR-GMTI systems based on enhanced shadow-aided decision," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, pp. 715–729, 2017.

[3] C. Clemente, L. Pallotta, D. Gaglione, A. De Maio, and J. J. Soraghan, "Automatic target recognition of military vehicles with krawtchouk moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 1, pp. 493–500, 2017.

[4] P. Tait, *Introduction to Radar Target Recognition*, Vol. 18, IET, , London, UK, 2005.

[5] Y. Zhai, W. Deng, T. Lan et al., "MFFA-SARNET: deep transferred multi-level feature fusion attention network with dual optimized loss for small-sample SAR ATR," *Remote Sensing*, vol. 12, no. 9, p. 1385, 2020.

[6] O. Kechagias-Stamatis, "Automatic target recognition on synthetic aperture radar imagery: a survey," 2020, https://arxiv.org/abs/2007.02106.

[7] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 1, pp. 6–43, 2013.

[8] P. Wang, W. Liu, J. Chen, M. Niu, and W. Yang, "A high-order imaging algorithm for high-resolution spaceborne SAR based on a modified equivalent squint range model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, pp. 1225–1235, 2014.

[9] R. L. Dudgeon, "An overview of automatic target recognition," *Lincoln Laboratory Journal*, vol. 6, pp. 3–10, 1993.

[10] L. Novak, G. Owirka, W. Brower, and A. Weaver, "The automatic target-recognition system in SAIP," *Lincoln Laboratory Journal*, vol. 10, pp. 187–201, 1997.

[11] G. Owirka, S. Verbout, and L. Novak, "Template-based SAR ATR performance using different image enhancement techniques," *Proceedings of SPIE*, vol. 3721, pp. 302–319, 1999.

[12] C. Clemente, L. Pallotta, I. Proudler, A. De Maio, J. J. Soraghan, and A. Farina, "Pseudo-Zernike-based multi-pass automatic target recognition from multi-channel synthetic aperture radar," *IET Radar, Sonar & Navigation*, vol. 9, no. 4, pp. 457–466, 2015.

[13] Y. Sun, L. Du, Y. Wang, Y. Wang, and J. Hu, "SAR automatic target recognition based on dictionary learning and joint dynamic sparse representation," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1777–1781, 2016.

[14] L. M. Novak, G. R. Benitz, G. J. Owirka, and L. A. Bessette, "ATR performance using enhanced resolution SAR," in *Algorithms for Synthetic Aperture Radar Imagery III*, pp. 332–337, International Society for Optics and Photonics, Bellingham, WA, USA, 1996.

[15] K. El-Darymli, E. W. Gill, P. McGuire, D. Power, and C. Moloney, "Automatic target recognition in synthetic aperture radar imagery: a state-of-the-art review," *IEEE Access*, vol. 4, pp. 6014–6058, 2016.

[16] Z. He, J. Lu, and G. Kuang, "A Fast SAR target recognition approach using PCA features," in *Proceedings of the*

International Conference on Image and Graphics, pp. 580–585, Chengdu, Sichuan, China, 2007.

[17] P. Han, R. Wu, Z. Wang, and Y. Wang, "SAR automatic target recognition based on KPCA criterion," *Journal of Electronics and Information Technology*, vol. 25, pp. 1297–1301, 2013.

[18] W. Bian and D. Tao, "Asymptotic generalization bound of Fisher's linear discriminant analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 12, pp. 2325–2337, 2014.

[19] N. Besic, G. Vasile, J. Chanussot, and S. Stankovic, "Polarimetric incoherent target decomposition by means of independent component analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, pp. 1236–1247, 2014.

[20] J. Zhou, Z. Shi, C. Xiao, and F. Qiang, "Automatic target recognition of SAR images based on global scattering center model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 3713–3729, 2011.

[21] J. I. Park, S. H. Park, and K. T. Kim, "New discrimination features for SAR automatic target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, pp. 476–480, 2012.

[22] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4806–4817, 2016.

[23] H. Zou, Y. Lin, and W. Hong, "Research on multi-aspect SAR images target recognition using deep learning," *Journal of Signal Processing*, vol. 34, pp. 513–522, 2018.

[24] C. Wang, J. Pei, Z. Wang, Y. Huang, and J. Yang, "Multi-view CNN-LSTM neural network for SAR automatic target recognition," in *Proceedings of the IEEE Geoscience and Remote Sensing Society*, pp. 1755–1758, Waikoloa Village, HI, USA, 2020.

[25] M. Zhang, J. An, D. Yu, L. Yang, and X. Lv, "Convolutional neural network with attention mechanism for SAR automatic target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2020.

[26] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness," 2018, https://arxiv.org/abs/1811.12231.

[27] K. L. Hermann, T. Chen, and S. Kornblith, "The origins and prevalence of texture bias in convolutional neural networks," 2019, https://arxiv.org/abs/1911.09071.

[28] A. Stergiou, R. Poppe, and G. Kalliatakis, "Refining activation downsampling with SoftPool," 2021, https://arxiv.org/abs/2101.00440.

[29] W. Li, B. Xueru, and Z. Feng, "SAR ATR of ground vehicles based on ESENet," *Remote Sensing*, vol. 11, p. 1316, 2019.

[30] S. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," 2017, https://arxiv.org/abs/1709.01507.

[31] E. R. Keydel, S. W. Lee, and J. T. Moore, "MSTAR extended operating conditions: a tutorial," in *Algorithms for Synthetic Aperture Radar Imagery III*, pp. 228–242, International Society for Optics and Photonics, Bellingham, WA, USA, 1996.

[32] P. Xia, "SAR target recognition based on joint use of target region and shadow," *Journal of China Academy of Electronics and Information Technology*, vol. 14, pp. 1062–1067, 2019.

[33] S. Doo, G. Smith, and C. Baker, "Target classification performance as a function of measurement uncertainty," in *Proceedings of the Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, pp. 1–4, Singapore, 2015.

[34] B. Ding, G. Wen, J. Zhong, C. Ma, and X. Yang, "A robust similarity measure for attributed scattering center sets with application to SAR ATR," *Neurocomputing*, vol. 219, pp. 130–143, 2017.

[35] J. Pei, Y. Huang, W. Huo, Y. Zhang, J. Yang, and T.-S. Yeo, "SAR automatic target recognition based on multiview deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2196–2210, 2018.

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

[37] Y. Sun, Y. Wang, H. Liu, N. Wang, and J. Wang, "SAR target recognition with limited training data based on angular rotation generative network," *IEEE Geoscience and Remote Sensing Letters*, vol. 99, pp. 1–5, 2019.

[38] J.-H. Park, S.-M. Seo, and J.-H. Yoo, "SAR ATR for limited training data using DS-AE network," *Sensors*, vol. 21, no. 13, p. 4538, 2021.

[39] Z. Ying, C. Xuan, Y. Zhai et al., "TAI-SARNET: deep transferred atrous-inception CNN for small samples SAR ATR," *Sensors*, vol. 20, no. 6, p. 1724, 2020.