IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 22, 2021, accepted October 2, 2021, date of publication October 5, 2021, date of current version October 20, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3118034

Synthetic Aperture Radar SAR Image Target Recognition Algorithm Based on Attention Mechanism

BAODAI SHI, QIN ZHANG, DAYAN WANG, AND YAO LI
Air and Missile Defense College, Air Force Engineering University, Xi'an 710051, China
Corresponding author: Baodai Shi (shi19085096879@163.com)

This work was supported by the National Natural Science Foundation of China under Grant 61971438.

ABSTRACT SAR images contain a large amount of noise, and related algorithms will cause high complexity when increasing the accuracy. To overcome this problem, a neural network model based on the attention mechanism was proposed in this paper. The model extracted information in two stages. It gradually extracts high-level features by reducing noise first and then adding hybrid attention. First, use dual-channel one-dimensional convolution to reconstruct the residual shrinkage network to construct a lightweight and efficient feature module, which improved the information extraction of the module with the consumption of a small amount of computing resources. Then, it was used as the backbone for model construction. Subsequently, mixed adaptive pooling was adopted to improve the maximum pooling. After that, dimensionality was reduced by pooling and linear interpolation was used to increase dimensionality, so as to generate feature weights of mixed dimension. Tests were performed on MSTAR dataset. The results showed that compared with the advanced algorithms, the proposed model in this paper can greatly reduce the amount of parameters and complexity while ensuring accuracy. The robustness test demonstrated that the model can effectively identify images with noise being added.

INDEX TERMS SAR image, one-dimensional convolution, attention mechanism, mixed adaptive pooling, robustness.

I. INTRODUCTION

As mentioned in [1]–[4], synthetic aperture radar (SAR) is a kind of active microwave imaging radar, and it has been extensively applied to military and civil fields for the advantages of full-time and all-weather work. The demand for military reconnaissance has stimulated SAR image automatic target recognition (ATR) technology, that is generally divided into three stages: image preprocessing, feature extraction, and target classification and recognition. Finally, the model or the category of the mission target is given, so that corresponding measures can be taken.

With the development of big data, traditional machine learning methods no longer meet demands. Since deep learning algorithms have entered the computer field, various networks have been created, such as GoogleNet series [5]–[7], ResNet series [8]–[10], and VGG [11], which have achieved good recognition results. Conventional image classification and recognition algorithms are also applicable to SAR. For example, in 2015, Guo *et al.* [12] combined the depth confidence network with polarimetric SAR data and proposed a new classification method based on depth learning, which achieved good classification accuracy. In 2017, Zhang *et al.* [13] introduced the amplitude and phase information of SAR images into CNN to further reduce the classification error.

Due to the high complexity and a large number of parameters, deep neural networks could consume tremendous computing resources, and the feature extraction efficiency is at a low level. The attention mechanism can help to solve this problem to a certain extent. For example, in 2017, Hu *et al.* [14] compressed the input feature map globally, and then completed the adaptive calibration of the weight from the channel dimension through excitation. In the same year, based on the idea of cross-layer connection of residual networks (ResNets), Wang *et al.* [15] realized the attention of space and channel domains (mixed domains) at the same time on soft branches. In 2018, on the basis of the channel

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

attention, Woo *et al.* [16] studied the position information of the feature map and proposed the Convolutional Block Attention Module (Convolutional Block Attention Module). In the same year, Wang *et al.* [17] used non-local blocks to capture global long-range dependencies, but a large amount of GPU resources were consumed. There are also cross-attention [18], self-attention [19], etc.

With the popularity of smart devices, accuracy is no longer the only indicator, and in the realm of military reconnaissance, when accuracy is ensured, the speed of SAR image target recognition will have a great impact on the result of the mission. Therefore, a lighter model that guarantees accuracy is desired. There are various methods to reduce the weight of the model, such as adding an attention mechanism, network pruning [20], and separable convolution [21]. For example, in [22], in 2019, Zhao *et al.* proposed a lightweight CNN model to avoid the over fitting problem caused by the lack of data, and achieved 98.30% accuracy on MSTAR data set. In [23], Ying used self-attention and knowledge distillation to achieve weight reduction of the model, and good results were achieved on the MSTAR dataset. However, there are still the following problems: (1) The existing models have a high complexity, and the feature extraction efficiency is low, or there is room for further improvement. (2) The SAR image contains high noise, and needs to be processed by a model with strong robustness.

In this paper, a SAR image target recognition algorithm-Model T was proposed based on attention mechanism. The model is divided into two stages. Stage 1: the two-channel S-DRSN module is used to initially extract features and remove noise, and the intermediate feature map was input into Stage 2. Stage 2 is divided into a trunk branch and a mask branch, which is responsible for implementing the hybrid attention mechanism. The trunk branch extracts mainstream features. The mask branch combines down-sampling and up-sampling to add hybrid attention. The contribution of this paper can be divided into the following three aspects:

1) In this paper, the improvement of residual shrinkage network is completed. A two channel adaptive one-dimensional convolution method is proposed to avoid dimensionality reduction and only conduct an appropriate amount of channel interaction, and the improved module is named S-DRSN. This method improves the information transmission efficiency of the module while only consuming a small amount of parameters.
2) Aiming at the problem that the weights generated in the hybrid attention mechanism proposed by Wang *et al.* are not accurate enough, an adaptive hybrid pooling method is proposed to improve the feature representation ability of down-sampling. This method takes into account both background information and texture information, and improves the accuracy of mask branches.
3) This model has two characteristics: lightweight and strong robustness. S-DRSN is a lightweight and high-efficiency feature extraction module. It is used as the backbone to build a model, which not only ensures accuracy, but also reduces the consumption of computing resources. And the model has strong resistance to random noise and salt and pepper noise.

II. RELATED WORK

The attention mechanism has become an important means to improve the performance of SAR classification models. This section is devoted to discussing the literature related to the method in this article.

A. A LIGHTWEIGHT CONVOLUTIONAL NEURAL NETWORK

In 2018, Jiaqi Shao *et al.* proposed a lightweight SAR classification model [24]. The model is improved based on ResNet50 whose innovations can be divided into three aspects. First, the channel attention mechanism and spatial attention mechanism were added to the main branch of ResNet50 to enhance the feature extraction ability; Then, the standard convolution in resnet50 was replaced by Depthwise Separable Convolution [21] to reduce the amount of parameters in the model; After that, a new weighted distance measure loss function was used to reduce the negative impact of unbalanced data on accuracy. In the end, the recognition rate on the MSTAR data set reached 99.54%. Compared with ResNet [8] and A-ConvNets [25], the recognition rate of this model is higher, But the model is still large, reaching 24.2Mb, and the network has strict requirements on the size of the input image, which is not practical in reality, so there is room for improvement.

B. CONVOLUTIONAL NEURAL NETWORK WITH ATTENTION

In 2020, Ming Zhang *et al.* also proposed a lightweight CNN model [26]. Based on the concept of A-ConvNets [25] full convolution, the model did not add any full connection layers, but only stacked eight convolution layers to extract features, which reduced trainable parameters. Moreover, in order to improve the feature extraction ability of the model, a convolutional block attention module (CBAM) was added after each convolutional layer. Similar to paper [24], the method completed the re-calibration of weights from channel domain and spatial domain respectively. Finally, the recognition rate of this model on MSTAR data set reached 99.35%, Compared with A-Convnet [25] and TAI-Sarnet [27], this model not only effectively reduced the number of parameters, but also achieved a higher recognition rate. Even so, the complexity of the model is still relatively large, reaching 5.12M, and there is still much room for improvement. It can be seen that although the direct addition of the existing attention mechanism can reduce the size of the model to a certain extent, it still cannot achieve excellent results. Therefore, a more efficient feature extraction module must be developed.

C. SELF-ATTENTION MULTISCALE FEATURE FUSION NETWORK

In [23], Ying *et al.* proposed a Self-attention Multiscale Feature Fusion Network for Small Sample SAR Image Recognition. The innovations of this model mainly include

three aspects: Firstly, a lightweight self-attention ghost module was constructed by combining self-attention mechanism [28] with ghost Module [29]. This module can efficiently extract target features and input to the next layer; Secondly, the channel shuffle [30] unit was added to the network structure to promote information interaction; Thirdly, knowledge distillation was carried out on the network to reduce the model size; Finally, the recognition rate on the MSTAR dataset reached 98.22%. However, the parameters of the model still reached 9×10^6, which does not achieve a good lightweight effect, and there is still room to improve the recognition rate, therefore, in order to achieve good recognition effect, not only attention mechanism should be applied, but also more efficient feature extraction module should be constructed.

III. PROPOSED METHOD

A. IMPROVED RESIDUAL SHRINKAGE NETWORK

In [31], Zhao *et al.* proposed a Deep Residual Shrinkage Network (DRSN) to solve the problem of fault diagnosis of vibration signals. This network is an improved version of residual networks (ResNets). The soft thresholding module is embedded in the residual network, which enhances the ability to learn task target features from noise, and effectively reduces the occupancy of computing resources by redundant features. The soft thresholding module is the core of the model, and its structure is shown in Figure 1, and the soft thresholding module can be divided into two steps here. Threshold generation and threshold screening, the threshold generation is accomplished by global average pooling (GAP) and two consecutive fully connected layers. Threshold screening is done by soft threshold functions, and the process of soft threshold function processing is shown in formula 1.

$$\lambda' = \begin{cases} \lambda - \gamma & \lambda > \gamma \\ 0 & -\gamma \le \lambda \le \gamma \\ \lambda + \gamma & \lambda < -\gamma \end{cases} \tag{1}$$

In Formula 1, λ is the input of the soft thresholding module, γ is the generated threshold, and λ' is the threshold after screening. In figure 1, the soft thresholding module is equivalent to the channel attention mechanism. The gray area in the figure is used to generate threshold. The size of the input image is $H \times W \times C$, and the output threshold is γ ($\gamma = \alpha \times \beta$). It passes through the global average pooling layer (GAP), and the absolute value is taken to obtain the real value α, that is processed by two fully connected layers, and then normalized by the Sigmoid function, so that the threshold β ($\beta \subseteq (0, 1)$) corresponding to the feature map is acquired. After γ and λ are processed by soft threshold function, the output feature λ' is obtained. Its principle is to remove the characteristic value whose absolute value is lower than a certain threshold, and shrink the features that larger than the threshold to about 0. The mechanism of the threshold function is not the highlight in this study, and thus it will not be further introduced here. Interested readers can consult reference [31].

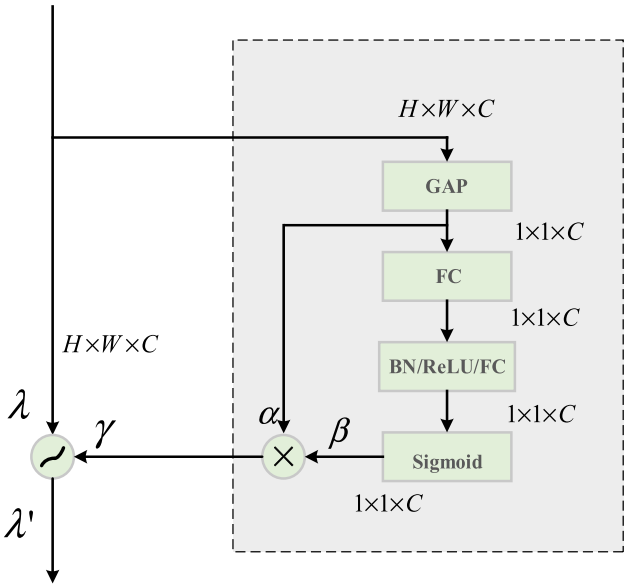

FIGURE 1. Soft thresholding of residual shrinkage module.

DRSN uses two fully connected layers to generate threshold. The fully connected layer reduces the dimensionality of the model while completing the cross-channel interaction. However, the extensive use of the fully connected layer will produce a large number of parameters, thereby reducing the processing efficiency, and previous study [32] has shown that: Reducing the dimensionality will weaken the accuracy of the model to a certain extent, and it is not necessary for the fully connected layer to obtain the connection between all channels. The following is an experiment to illustrate the effect of dimensionality reduction, stack 10 layers of residual shrinkage block and its two variants as a model, and verify the performance on the MSTAR data set, the experimental results are shown in Table 1, and the corresponding scatter diagram is shown in Figure 2.

In Table 1, DRSN represents residual shrinkage network, DRSN removing soft thresholding module is DRN, which is equivalent to residual network, DRSN (-FC) represents removing one fully connected layer in soft threshold module, and DRSN (-2FC) represents removing two fully connected layers in soft threshold module. It can be seen from the table that the accuracy of the residual shrinkage network is 87.65%, which is 10.20% higher than that of the residual network. It can be seen that the soft threshold module is very useful. When a fully connected layer is removed, part of the dimension reduction is reduced, the accuracy is increased by 4.28%, and the complexity and the number of parameters are slightly reduced. After that, the two fully connected layers are removed and the global average pooling (GAP) value was output as the threshold value, the accuracy was reduced by 2.84%, it can be seen that blindly reducing the dimension while ignoring the cross-channel information interaction

TABLE 1. The effect of different fully connected layers on the result.

Network	Params	FLOPs	Accuracy /%
DRN	27938	134720	77.45
DRSN	31458	150140	87.65
DRSN(-FC)	30018	143720	91.93
DRSN(-2FC)	28578	137320	89.09

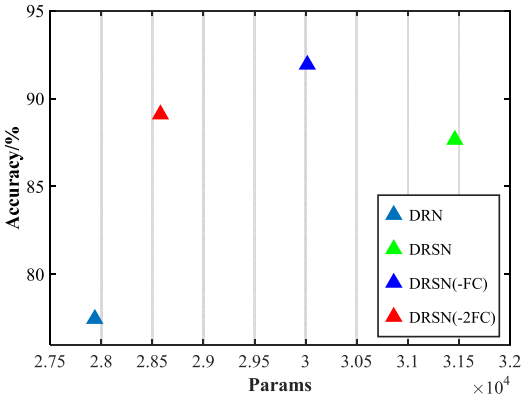

FIGURE 2. Scatter plot of the effect of different fully connected layers on the experimental results.

brought by the fully connection layer is not conducive to improving network performance. Moreover, the connection between channels decreases with the increase of the distance between channels, so it is unnecessary and inefficient for the full connection layer to obtain the connection between all channels. Therefore, we can only obtain the connection between each channel and the surrounding k channels, instead of obtaining the connection between all channels, and the K value should be tuned according to different situations.

Therefore, this research used two-channel adaptive one-dimensional convolution to complete the information interaction between local channels. The use of one-dimensional convolution will not reduce dimensionality, and only the interaction between each channel and its neighbors is obtained, instead of the connection between the global channels. In this way, not only the negative effects brought by the fully connected layer can be avoided, but also parameter amount of the DRSN module can be reduced, so that the feature extraction efficiency of the network is improved, as shown in Figure 3.

As shown in Figure 3, all channels of the input feature map are independently processed by global average pooling and global maximum pooling (GMP). Then, two positive real values are obtained and input to one-dimensional convolution layer with a convolution kernel size k. The k value here changes according to the change of the channel dimension. For unknown mapping $k = \varphi(c)$, it is generally difficult to find the optimal mapping relationship, After research,

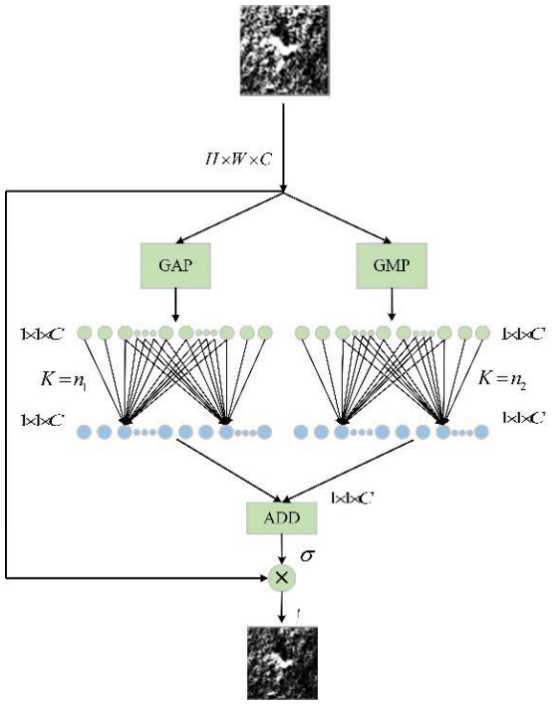

FIGURE 3. Two-channel one-dimensional convolution.

Formula 2 was used as the approximate mapping.

$$k = \left[\left|\frac{-b \pm \sqrt{b^2 - 4a(d - \log_2^{\frac{C}{A}})}}{2a}\right|\right] \quad b^2 - 4a(d - \log_2^{\frac{C}{A}}) \geq 0 \tag{2}$$

The derivation process is as follows:

It is very important to find the appropriate K value, which determines the range of channel interaction. When the number of channels increases, the range of interaction will certainly be different. Assuming $k = \varphi(c)$, c can also be expressed by the formula containing k. The number of channels c is generally an integer power of 2. In order to simplify the problem, the highest power of k was set to 2, as shown in Formula 3.

$$C = A2^{ak^2+bk+d} \tag{3}$$

Reverse solution leads to Formula 4:

$$ak^2 + bk + d - \log_2^{\frac{C}{A}} = 0 \tag{4}$$

Formula 2 was solved, and the nearest integer that did not exceed the absolute value of the calculated value was selected as the value of k. In order to simplify the problem, the values of A, a, b, and d were set to 2, 1, 2, and 0, respectively. Then, the output threshold $\gamma = \sigma(CID(x))$, $C1D(\cdot)$ represents one-dimensional convolution and x is the input feature. Let all the channels share the parameters, and the threshold generation module has a total of $2k$ parameters, which are reduced compared with the parameters $2c^2$ of the double-layer FC.

Next, stack 10 layers of blocks as a network structure to experiment with the number of adaptive one-dimensional convolutions and how to add them. The test results are shown in Table 2, the corresponding scatter plot is shown in Figure 4.

TABLE 2. Experimental results of the number of adaptive one-dimensional convolutions and how to add them.

Network	Params	FLOPs	Accuracy /%
S1-DRSN	28018	134860	92.05
S2-DRSN	28098	135000	93.63
S3-DRSN	28258	135280	90.58
SS2-DRSN	28098	135000	91.25
SS3-DRSN	28258	135280	90.23

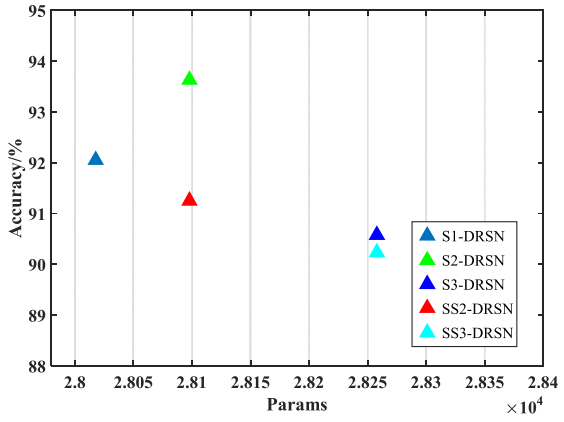

FIGURE 4. Scatter plot of experimental results of the number of adaptive one-dimensional convolutions and adding methods.

In Table 3, S1-DRSN represents single-channel one-dimensional convolution, S2-DRSN represents double-channel one-dimensional convolution, S3-DRSN stand for three-channel one-dimensional convolution, SS2-DRSN represents continuously stacked two-layer one-dimensional convolution, and SS3-DRSN represents continuously stacked three-layer one-dimensional convolution. It can be seen from the table that the accuracy of S2-DRSN is the highest, reaching 93.63%, and the number and complexity of parameters are similar to that of residual network in Table 1. The accuracy of S2-DRSN is the highest, reaching 93.63%, but the number and complexity of parameters are similar to that of residual network in Table 1. It can be seen that soft threshold module is an efficient lightweight module. Moreover, the accuracy of continuous stacked one-dimensional convolution is generally low. Therefore, in this paper, S2-DRSN serves as the backbone module for network construction and is named S-DRSN.

B. IMPROVED HYBRID ATTENTION

The second stage is based on the hybrid attention mechanism (HAM) proposed by Wang *et al.*, and its implementation is shown in Figure 5.

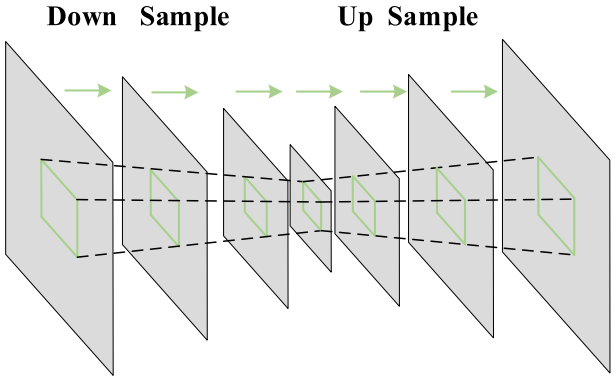

FIGURE 5. Schematic diagram of mixed attention.

As shown in the figure, its structure is similar to that of Fully Convolutional Networks (FCN) in [33]. It combines the feedforward scanning mechanism with the top-down feedback mechanism through down-sampling and up-sampling. First, n ($n > 0$) times Global Max Pooling was used to search global features to increase the receptive field. In order to make the size of the output feature map of the mask branch match the size of the output feature map of the trunk branch, deconvolution or linear interpolation was adopted to enlarge the feature map n times. After passing through two convolutions and the sigmoid function, the features were normalized and the weight value was extracted.

Although pooling reduces the resolution and increases the receptive field, it will cause the loss of information. Global Max Pooling will discard all activation values except the maximum value, thus ignoring the background information in the figure. The use of Global Average pooling may also lead to the situation that positive and negative activation values cancel each other out, while texture information is ignored, thus resulting in the loss of information. Therefore, only using max pooling is not accurate enough. In order to make the weights output by the mask branch more accurate, the down-sampling method was improved.

The input feature map matrix was set to F, the size of the pooling domain matrix P is $c \times c$, with the bias of b_2, and the processed feature map is S, then the calculation expression of max pooling is shown in Formula (5), and the calculation expression of average pooling is shown in Formula (6).

$$S_{ij} = \max_{i=1,j=1}^{c} (F_{ij}) + b_2 \tag{5}$$

$$S_{ij} = \frac{1}{c^2}\left(\sum_{i=1}^{c}\sum_{j=1}^{c} F_{ij}\right) + b_2 \tag{6}$$

In this paper, max pooling and average pooling were combined linearly, and a new pooling method-mixed adaptive

pooling (MA pooling) was proposed. The calculation expression is shown in Formula (7).

$$S_{ij} = \beta \frac{1}{c^2}\left(\sum_{i=1}^{c}\sum_{j=1}^{c} F_{ij}\right) + (1-\beta)\max_{i=1,j=1}^{c}(F_{ij}) + b_2 \quad (7)$$

where the value of β is calculated by Formula (8).

$$\beta = \frac{\frac{1}{c^2}\left(\sum_{i=1}^{c}\sum_{j=1}^{c} F_{ij}\right) + b_2}{\max_{i=1,j=1}^{c}(F_{ij}) + \frac{1}{c^2}\left(\sum_{i=1}^{c}\sum_{j=1}^{c} F_{ij}\right)} \quad (8)$$

When β is assigned to the value of average pooling and the ratio of the values of the two, it is equivalent to completing the normalization of the ratio, which is reasonable. When $\beta \rightarrow 0$, the value degenerates to the maximum pooling, and when $\beta \rightarrow 1$, the value is close to the average pooling. In this way, the texture information is considered, and the background information is protected. Besides, the back-propagation process can also be adjusted according to the value of β without affecting it negatively. Excessive reduction of the dimensionality will affect the accuracy of the model. Even if the dimensionality can be increased later through linear interpolation, the internal value will also have irreversible changes. Therefore, in this paper, two improved pooling methods were adopted to reduce the dimensionality. The structure of the second stage is exhibited in Figure 6. The performance of the pooling method was verified in the following test.

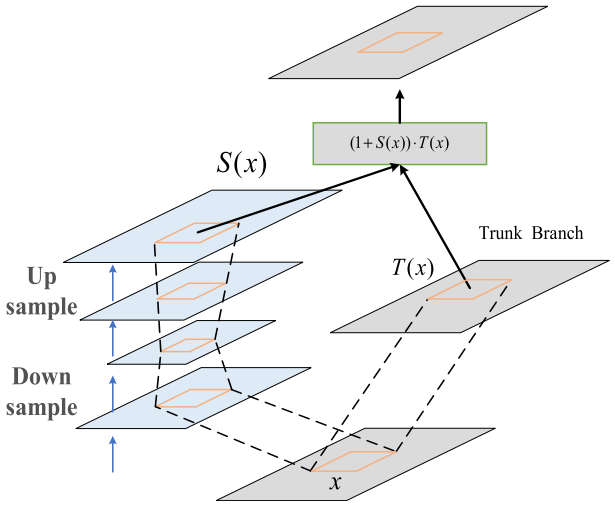

FIGURE 6. Second stage structure diagram.

For complex SAR image classification tasks, it is necessary to continuously extract high-level features. Especially, after the down-sampling and convolution, the feature value of the backbone network gradually reduces. In order to avoid network degradation during the training process, residual learning was used in the second stage. The parallel structure of trunk branch and mask branch was used for feature extraction. The mask branch is responsible for adding the hybrid attention to the trunk branch. The output of the trunk branch is $T(x)$, and the output of the mask branch is $S(x)$. The outputs of the two branches were multiplied and then added, and the total output is shown in Formula (9).

$$H_{i,c}(x) = (1 + S_{i,c}(x)) \cdot T_{i,c}(x) \quad (9)$$

where x is the input of the second part; $H_{i,c}(x)$ is the output of the i-th layer c channel through the hybrid attention mechanism, and the same is true for $S_{i,c}(x)$ and $T_{i,c}(x)$; The value of $S_{i,c}(x)$ is between [0,1], and it can be considered as the feature selector of $T_{i,c}(x)$ to suppresses noise and enhances the weight of effective features.

C. MODEL STRUCTURE

Limited by the imaging principles, SAR images will inevitably contain speckle noise. Therefore, in this study, a model was built from the perspectives of noise removal and high-efficiency feature extraction. The improved S-DRSN module and hybrid attention in this paper are both based on these two perspectives. In the design of the network, the accuracy was not improved by stacking the network, but from the perspective of enhancing the weight of effective features. The model is divided into two parts, both of which use the S-DRSN module as the backbone to gradually extract high-level features. In the first stage, the S-DRSN module is made two-way parallel to widen the network, achieve high-level feature expression, and improve network performance. Moreover, the improved residual shrinkage network also completes the feature screening of the channel dimension. After that, the networks of the two channels are fused and input to the second stage. Both the improved residual shrinkage network and the soft branch of the second stage can be used as attention mechanism, which can not only complete the recalibration of the forward feature weights, but also can update the gradient of back propagation, so that the model is stronger. The overall network structure is shown in Figure 5.

As shown in Figure 7, the model contains three parameters, namely q, w and p, where both p and q represent the number of S-DRSN modules in the branch, and w represents the number of S-DRSN modules in the middle of two stages. The model can be trained from end-to-end. The image is first input to Stage 1, and both branches are connected to two 3×3 convolutions with the convolution kernel of 32 and 8 respectively, and then S-DRSN modules are connected, whose number is q ($q = 1$). The value of q is adjustable, including the value of w and p, that should be optimized according to different task requirements. Then after batch normalization and Relu activation, two 3×3 convolutions are used to extract deep features. S-DNSNs are added between the two stages, whose number is w ($w = 1$). Then, the processed feature map is input to the second stage. The trunk branch contains only 2 S-DRSN modules. After down-sampling and up-sampling, the soft branch passes through two 1×1 convolutions, and subjects to the Sigmoid function for normalization, and completes the weight calibration together with the trunk branch. Similarly, after passing through w S-DRSNs, and

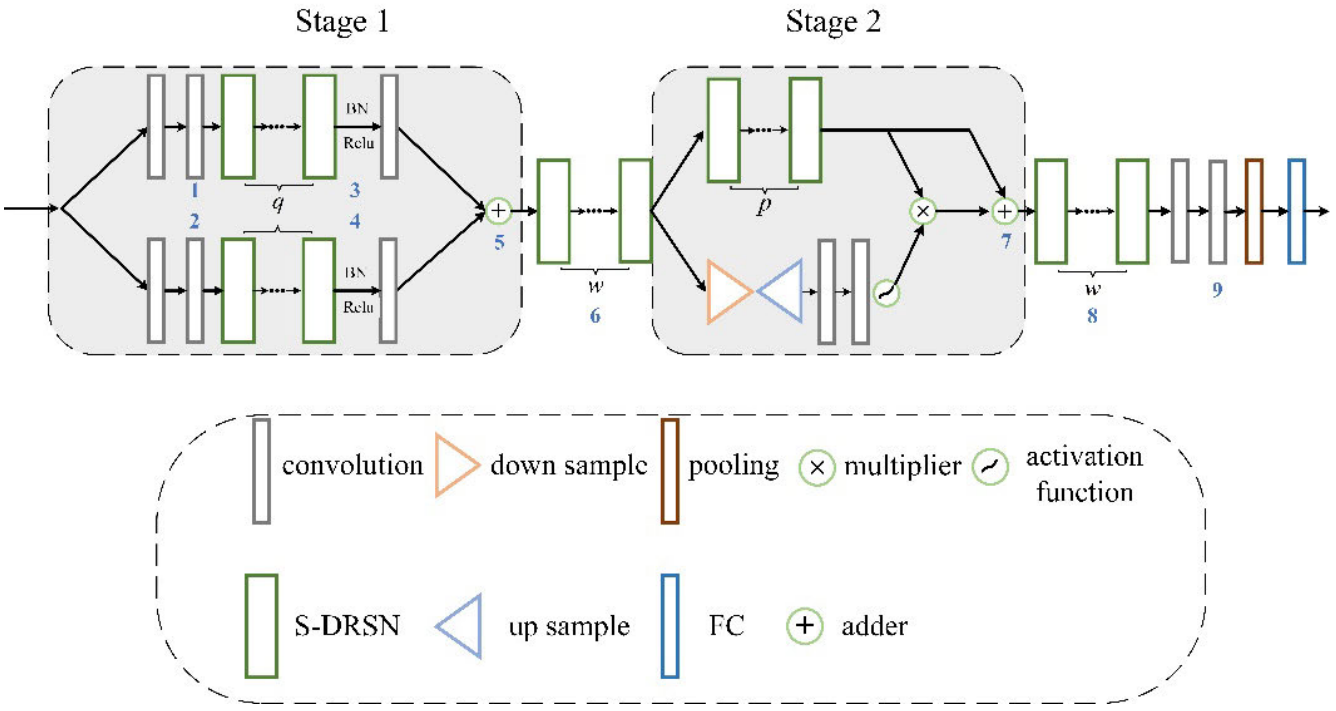

FIGURE 7. Structure diagram of the model.

then a 3 × 3 convolutional layer, one-dimensional convolution is used to complete the channel integration, and then global average pooling is used to replace the FC layer, which reduces the parameter amount and makes the model stronger. Finally, the classification result is output through the softmax function. The network structure is shown in Table 3, and the structure of the mask branch (HAM) is shown in Table 4.

This model has strong robustness to SAR images containing noise, see below for related test. However, due to the complex electromagnetic environment for SAR imaging and the cluttered scenes, this model is not suitable for all situations, and a new attention mechanism needs to be developed according to different task objectives.

TABLE 3. Overall network structure parameters.

Input	Operater	parameter
96×96×3	Conv2D Conv2D	(3×3,32) (3×3,8)
96×96×32 96×96×3	Conv2D Conv2D	(3×3,8) (3×3,8)
96×96×8 96×96×8	S-DRSN S-DRSN	-- --
48×48×8 48×48×8	Conv2D Conv2D	(3×3,32) (3×3,32)
48×48×32 48×48×32	Add	-- --
48×48×32	S-DRSN	--
48×48×32	S-DRSN	--
48×48×32 48×48×32	S-DRSN HAM	-- --
48×48×32	S-DRSN	--
48×48×32	Conv2D	(3×3,64)
48×48×64	Conv2D	(1×1,32)
48×48×32	GAP	
32	Dense/Softmax	10

IV. TEST ANALYSIS

A. MSTAR DATASET

The Moving and Stationary Target Acquisition and Recognition (MSTAR) project jointly launched by the Air Force Research Laboratory (AFRL) and the Defense Advanced Research Project Agency (DARPA) [34], [35] combines a model-based target recognition algorithm with the ATR system, which can effectively identify ground targets, and has better robustness than Semi-Automated Image Intelligence Processing (SAIP) [36]. The project published the MSTAR dataset for research. At present, there are hundreds of papers based on this dataset, such as references [37]–[40].

The performance of the model was tested on the MSTAR dataset. The dataset contains two types of data collected under Standard Operating Condition (SOC) and Extended Operating Condition (EOC). The dataset collected under SOC contains three major types of data, with a total of 10 types of targets: BMP2, BRDM2, BTR60, BTR70, ZIL131, D7, T62, T72, 2S1 and ZSU234. The data without variants being collected at a pitch angle of 17° were used as the training set, and the data containing variants collected at a pitch angle of 15° were used as the test set. The image size is 128 × 128. Optical images of 10 types of targets and the corresponding SAR image are displayed in Figure 8. The training data collected under EOC includes three types of military targets: BMP2 (SN_c21), BTR60, and T72 (SN_132).

TABLE 4. Structure parameters of HAM.

HAM	parameter
MA Pooling	3×3
MA Pooling	3×3
Interpolation	--
Interpolation	--
BN+Relu	--
Conv2D	(1×1,32)
BN+Relu	--
Conv2D	(1×1,32)
Sigmoid	--

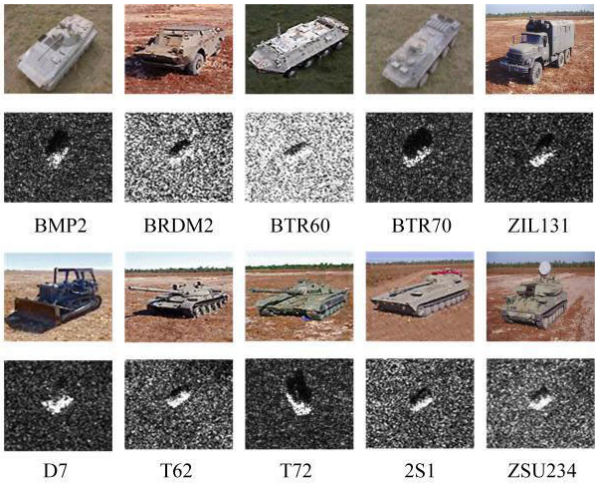

FIGURE 8. Optical images of 10 types of targets and corresponding SAR images.

Four variants were added to the test samples, and a total of seven models were used to test the ability of the model to identify variants, including BMP2 (SN_9563, SN_9566, SN_c21), BTR60 and T72 (SN_132, SN_812, SN_s7). Two different scenes were set to test the model performance more comprehensively.

B. TEST UNDER SOC

1) TEST PROCESS

a: TEST PLATFORM

The hardware graphics card in this study is Nvidia GeForce RTX2060 (4G) graphics card, and CPU is i7-10750H (16G); Win 10 GPU version of Tensorflow 1.14 deep learning framework was adopted, python 3.7. Training details: The ADAM optimizer was used for realizing adaptive learning rate, with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 10e - 8$. The cross-entropy loss function was used, 70 Epochs were trained, and the Batch size was set to 64.

The center of the ten types of target images was cropped to 96 × 96 for test, and training was performed according to the above settings. The confusion matrix of the proposed model's recognition results of the ten types of targets under SOC is shown in Table 5, and each row represents the classification result of the test samples and corresponding accuracy rate. The comparison between the proposed model and other algorithms is shown in Table 5. It can be seen from the table that the average recognition rate of the ten types of targets in this paper has reached 99.42%, and the accuracy of BRDM2, ZIL131, D7, T62 and 2S1 targets has reached 100%. It can be seen that the features of interest in the model are consistent with the target characteristics. Moreover, BTR60 and BTR70 are easy to misjudge each other, which is the main reason for the reduction of the average accuracy, In the next step, fine-tuning can be performed according to the difference between the characteristics of the two types of targets. Overall, the proposed algorithm in this paper has achieved good results for ten types of targets under the SOC, which verifies the effectiveness of the algorithm. The accuracy curve and loss curve of the training and testing process are shown in Figure 9. It can be seen that the model converges quickly, and the accuracy of the test set is higher than that of the training set, which proves that the model is effective.

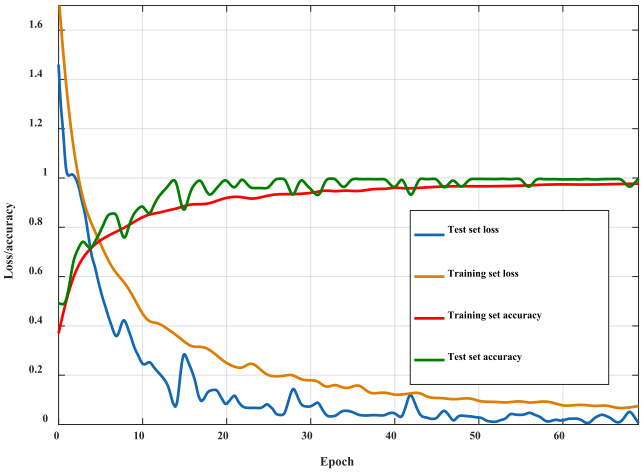

FIGURE 9. Experimental process curve under SOC condition.

The comparison between the proposed model and other algorithms is shown in Table 6. In terms of accuracy, Model T that was proposed in this paper, VGG-S1-DCA, YOLOv4-MCCA,'Model S, A-CNN and CMNet all reach more than 99%, at the same level. Among them, the recognition rate of YOLOv4-MCCA and VGG-S1-DCA respectively reached 99.7% and 99.9%, which were the highest in accuracy, but there were too many parameters in these two models. However, Model T has the least amount of parameters, with only 0.1m, and the complexity is also the lowest, which is two orders of magnitude less than that of Model S, 38.97% of the complexity of A-CNN and 73.61% of the complexity of CMNet. Although these models have a small number of network layers, each layer has a large number of convolutional kernels, and the efficiency of information extraction is low, resulting in a large amount of complexity. This guarantees a lighter model, and greatly reduces the occupation of computing resources. However, it is not enough to test the

TABLE 5. Confusion matrix of experiments under SOC.

Target	BMP2	BRDM2	BTR60	BTR70	ZIL131	D7	T62	T72	2S1	ZSU234	P/%
BMP2	**193**	0	0	0	0	0	0	1	1	0	98.97
BRDM2	0	**274**	0	0	0	0	0	0	0	0	100
BTR60	0	1	**191**	2	0	0	0	0	0	1	97.94
BTR70	0	0	2	**193**	0	0	0	1	0	0	98.47
ZIL131	0	0	0	0	**274**	0	0	0	0	0	100
D7	0	0	0	0	0	**274**	0	0	0	0	100
T62	0	0	0	0	0	0	**273**	0	0	0	100
T72	0	1	0	0	0	0	0	**195**	0	0	99.49
2S1	0	0	0	0	0	0	0	0	**274**	0	100
ZSU234	0	0	0	0	0	2	0	0	0	**272**	99.27
Average/%					--						99.42

TABLE 6. Comparison table of different algorithms on the SOC data set.

Network	Params /10^6	FLOPs /10^6	Accuracy /%
VGG-S1-DCA[41]	551.97	/	99.9
ResNet50	25.68	82.52	97.9
Residual Attention Network[15]	32.21	96.78	98.9
YOLOv4-MCCA[42]	20.02	/	99.7
Model S [43]	15.68	50.50	99.6
A-ConvNets [25]	0.37	0.74	99.1
A-CNN [44]	0.30	1.36	99.4
CMNet [45]	0.16	0.72	99.3
Model T	0.11	0.53	99.4

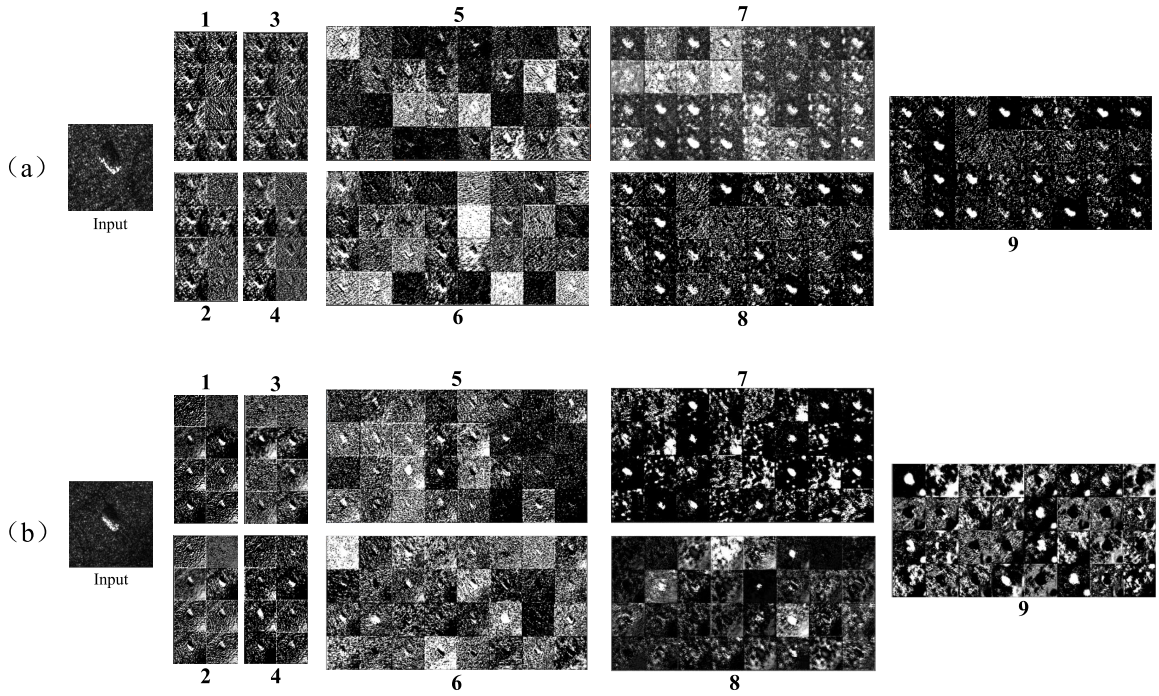

FIGURE 10. Visualization of the intermediate feature map.

performance of the model under standard condition. Next, tests were conducted under extended condition and in an environment with added noise, see below.

2) VISUALIZATION OF FEATURE MAPS

In order to intuitively understand the learning results inside the network, visual methods were used to output the feature maps of the important nodes of Model T. BRDM_2 and T72 in the test set were selected, and the visual feature map of the operation corresponding to the blue numbers in Figure 5 was extracted. The results are shown in Figure 10. The visualization results of BRDM_2 and T72 correspond to (a) and (b) in Figure 7 respectively. It can be seen that the feature map after each layer processing is quite different, but

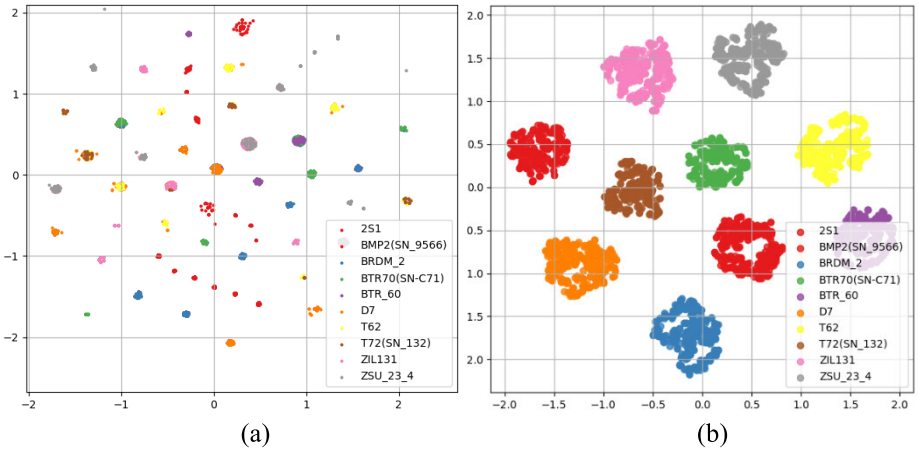

FIGURE 11. Data distribution diagram before and after model processing.

the recognition rate of these two types of targets has reached 100%, so the feature extraction is effective.

3) DATA DISTRIBUTION VISUALIZATION

This section uses the t-SNE algorithm proposed by Geoffrey [46] to visualize the distribution of the MSTAR data set before and after classification, so as to visually observe the processing effect of the model. The original MSTAR data distribution is visualized, and the results are shown in Figure 11(a). The data distribution after model processing is visualized, and the results are shown in Figure 11(b). It can be seen from the two figures that the distribution of original data is chaotic and scattered, and it is difficult to distinguish. After processing by model T proposed in this chapter, 10 types of targets in MSTAR data have been grouped together, which proves that the model is effective.

C. TEST UNDER EOC

The model's ability to identify variants was tested. Unlike the test of the recognition of ten types of targets, this experiment has a small number of samples, with only 698. First, the dataset was expanded. The batch size was set to 12, and 40 Epochs were trained. Other parameter settings remained the same as above, and the confusion matrix of the test results is shown in Table 7. From the results, it can be seen that after adding variants, the accuracy rate is slightly lower than that of the recognition of ten types of targets, but it still reaches 97.66%. This verifies the model is still effective and has strong robustness.

D. CIFAR-10 EXPERIMENT

This model is not only applicable to SAR images, but also to general images. This experiment was used to test the generalization of the model on general optical datasets. CIFAR-10 is a commonly used dataset in the field of image classification, such as references [47] and [48]. It is divided into 10 categories, with a total of 60,000 images, including 50,000 training samples, and 10,000 test samples, and the image size is 32×32.

TABLE 7. Confusion matrix of experiments under EOC.

Target	BMP2	BTR60	T72	P/%
BMP2(SN_9563)	192	1	2	
BMP2(SN_9566)	189	3	4	97.27
BMP2(SN_c21)	190	2	4	
BTR60	2	190	3	97.44
T72(SN_132)	0	0	196	
T72(SN_812)	4	2	189	98.28
T72(SN_)	3	1	187	
Average		--		97.66

Considering the various types of data and the difficulty of training, let the value of w be 2, and the number of convolution kernels was fine-tuned. The model structure remains unchanged. The comparison of the recognition results of the models is shown in Table 8. It can be seen from the table that compared with the existing advanced lightweight neural networks, the model in this paper has achieved considerable accuracy, but the parameter amount and complexity are reduced by an order of magnitude. The model in this paper is still effective for conventional optical images.

TABLE 8. Comparison of different algorithms on the CIFAR-10.

Network	Params/10^6	FLOPs /10^6	Accuracy/%	
			Top-1	Top-5
MobileNetV2 [49]	2.21	6.18	90.16	99.02
SqueezeNet [50]	5.72	7.79	88.20	97.69
Model T	0.26	0.95	89.51	98.28

V. MODEL ROBUSTNESS TEST

Robustness is an important indicator to measure the performance of the model, and it is also the key to whether the model can be extended to practical applications. Especially, the current electromagnetic environment of the battlefield is

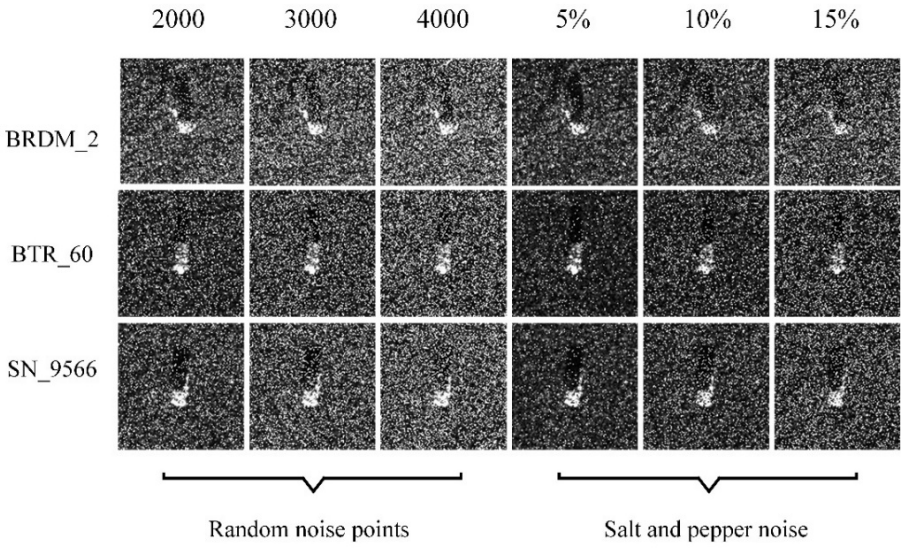

FIGURE 12. Noise adding effect picture.

TABLE 9. Recognition result of noisy image.

Scale / P/% / Target	2000	3000	4000	5%	10%	15%
BRDM_2	94.89	96.72	98.91	98.54	91.97	94.89
BTR_60	99.49	98.97	96.92	100	100	98.46
SN_9566	84.18	81.63	72.45	88.78	78.57	67.56
Average/%	92.85	92.44	89.43	95.77	90.18	86.97

complex, and it is easy to bring additional noise to SAR images. Model T was proposed to deal with data containing noise. In this research, random noise points, salt and pepper noise were added to the MSTAR data, and the BRDM_2, BTR_60 and the variant target of BMP2 - SN_9566 were selected for test. Random noise: The gray value of pixel points was set to 0 randomly, and the random noise points with the number of n were added to the image. The value of n is 2000, 3000, and 4000 respectively. Salt and pepper noise: The form of this kind of noise is similar to the speckle noise. In the experiment, the noise ratio P_r was given, and then the random function was used to generate a random number between [0, 1]. The value greater than $1 - P_r$ was set as the salt noise, and the value less than $1 - P_r$ was set as pepper noise. The ratio of noise addition is 5%, 10%, and 15% respectively. The effect of the two noise addition methods are shown in Figure 12, and the confusion matrix of the recognition results is shown in Table 9.

The results of the average recognition rate showed that the model has strong robustness to both random noise and salt and pepper noise. When the random noise is 4000 and the noise ratio is 15%, the recognition rate reaches over 86%. The recognition accuracy of the three types of targets basically decreases with the increase of noise, and SN_9566 is greatly affected by noise. However, the recognition rate of the target

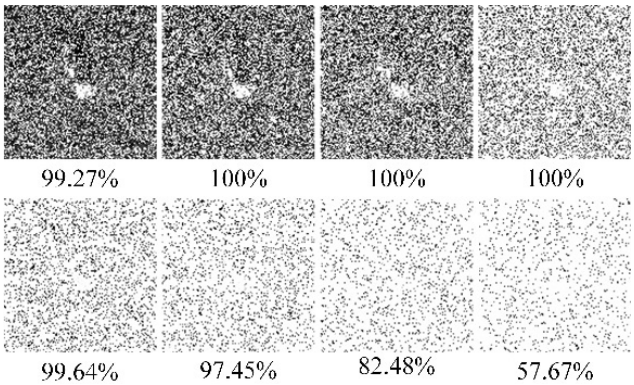

FIGURE 13. BRDM_2 noisy image and corresponding recognition rate.

BRDM_2 increases with the increase of random noise. When the number of random noise is 4000, the accuracy rate is 98.91%. To further explore the reason, random noise was further added to the target BRDM_2. The number of noise points and the corresponding recognition rate are displayed in Figure 13. The number of noise points from left to right, and from top to bottom is 6000, 8000, 10000, 20000, 30000, 35000, 40000, and 45000, respectively. When the number of noise points is 30,000, the recognition rate is more than 99%, indicating that Model T has strong robustness to the target. Supposing that what Model T is interested in BRDM_2 is a certain prominent feature E, and feature E can be enhanced with the number of noise points until it reaches a certain critical value, and this critical value is between 20000 and 30000.

VI. CONCLUSION

Due to the technical advantages of SAR and a large amount of noise contained in the image, the classification and recognition of SAR images has always been a hot and difficult research problem. Convolutional neural network has made

a breakthrough in the field of image, but its application in the field of SAR image classification is not mature enough, and the current number of algorithm layers is too deep and the network structure is complex. Although some achievements have been made, the algorithm complexity is high. Compared with other algorithms, model T greatly reduces the complexity of the algorithm on the basis of maintaining a high recognition rate, which belongs to a lightweight SAR classification model.

The improved S-DRSN module in this paper can efficiently extract features from noise images. The Model T built with S-DRSN, and suppresses the transmission of noise, so that more advanced accuracy can be obtained with a small consumption of computing resources. In addition, the model has good robustness, and it is applicable to both SAR images and conventional optical color datasets. The improved S-DRSN module in this paper can efficiently extract features from images containing noise. And it can be embedded in various networks. Nevertheless, the model still has the following limitations:

1) The deep learning algorithm relies on a large number of data, but there are few SAR data, which limits the application of the model to a certain extent.

2) At present, the electromagnetic environment is complex, which brings many kinds of noise to the image, so the performance of the model needs further verification.

ACKNOWLEDGMENT

The authors would like to thank the Air Force Engineering University of China for helpful on this work. They also thank the associate editor and the reviewers for their useful feedback that improved this paper.

REFERENCES

[1] A. Pasmurov and J. Zinoviev, "Radar imaging application," in *Radar Imaging and Holography*. England, U.K.: IET Digital Library, 2005, pp. 191–230. [Online]. Available: https://digital-library.theiet.org/content/books/10.1049/pbra019e_ch9?__cf_chl_jschl_tk__=pmd_tCB6YoNqF5uEGZvaHOy41TtvgxbBozWWvuSqvWCXdcs-1633955605-0-gqNtZGzNAiWjcnBszQj9

[2] X. Zhang, L. Jiao, F. Liu, L. Bo, and M. Gong, "Spectral clustering ensemble applied to SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2126–2136, Jul. 2008.

[3] J.-M. Nicolas and F. Adragna, "The principles of synthetic aperture radar," in *Processing of Synthetic Aperture Radar Images*, H. Maître, Ed. U.K.: Wiley, 2008, pp. 25–55.

[4] L. M. H. Ulander, A. Barmettler, B. Flood, P-O. Frolind, A. Gustavsson, T. Jonsson, E. Meier, J. Rasmusson, and G. Stenstrom, "Signal-to-clutter ratio enhancement in bistatic very high frequency (VHF)-band SAR images of truck vehicles in forested and urban terrain," *IET radar, Sonar Navigation*, vol. 4, no. 3, pp. 438–448, 2010.

[5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 1–9.

[6] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2818–2826.

[7] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, 2017, pp. 4278–4284.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[9] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1492–1500.

[10] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 4700–4708.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[12] Y. Guo, S. Wang, C. Gao, D. Shi, D. Zhang, and B. Hou, "Wishart RBM based DBN for polarimetric synthetic radar data classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Milan, Italy, Jul. 2015, pp. 1841–1844.

[13] Z. Zhang, H. Wang, F. Xu, and Y.-Q. Jin, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, Dec. 2017.

[14] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, vol. 42, Jun. 2018, pp. 7132–7141.

[15] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 3156–3164.

[16] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, 2018, pp. 3–19.

[17] X. Wang, R. B. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, Jun. 2018, pp. 7794–7803.

[18] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: Criss-cross attention for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 603–612.

[19] R. Zhang, Y. Zou, and J. Ma, "Hyper-SAGNN: A self-attention based graph neural network for hypergraphs," 2019, *arXiv:1911.02613*. [Online]. Available: http://arxiv.org/abs/1911.02613

[20] N. T. Siebel, J. Botel, and G. Sommer, "Efficient neural network pruning during neuro-evolution," in *Proc. Int. Joint Conf. Neural Netw.*, Atlanta, GA, USA, 2009, pp. 2920–2927.

[21] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861

[22] Y. Zhai *et al.*, "A novel lightweight SARNet with clock-wise data amplification for SAR ATR," *Prog. Electromagn. Res. C*, vol. 91, pp. 69–82, Mar. 2019.

[23] Z. Ying, Y. Zhai, C. Xuan, and F. Wang, "Self-attention multiscale feature fusion network for small sample SAR image recognition," *J. Signal Process.*, vol. 36, no. 11, pp. 1846–1858, Nov. 2020, doi: 10.16798/j.issn.1003-0530.2020.11.007.

[24] J. Shao, C. Qu, J. Li, and S. Peng, "A lightweight convolutional neural network based on visual attention for SAR image target classification," *Sensors*, vol. 18, no. 9, p. 3039, Sep. 2018.

[25] S. Chen, H. Wang, F. Xu, and Y. Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806–4817, Aug. 2016.

[26] M. Zhang, J. An, D. H. Yu, L. D. Yang, L. Wu, and X. Q. Lu, "Convolutional neural network with attention mechanism for SAR automatic target recognition," *IEEE Geosci. Remote Sens. Lett.*, early access, Nov. 2, 2021, doi: 10.1109/LGRS.2020.3031593.

[27] Z. Ying, C. Xuan, Y. Zhai, B. Sun, J. Li, W. Deng, C. Mai, F. Wang, R. D. Labati, V. Piuri, and F. Scotti, "TAI-SARNET: Deep transferred atrous-inception CNN for small samples SAR ATR," *Sensors*, vol. 20, no. 6, p. 1724, Mar. 2020.

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[29] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1580–1589.

[30] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 6848–6856.
[31] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4681–4690, Jul. 2019.
[32] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks, 2020 IEEE," presented at the IEEE Comput Soc. Conf. Comput. Vis. Pattern Recognit., Virtual, USA, Apr. 2020.
[33] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 3431–3440.
[34] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. New York, NY, USA: SciTech, 2004.
[35] H. Cheng, Q. Yu, J. Teah, and J. Liu, "Based on wavelet bridge target detection in SAR image segmented by support vector machine," *Huazhong Sci. Technol. J. Univ., Natural Sci. Ed.*, vol. 34, no. 4, pp. 52–55, 2006.
[36] L. M. Novak, G. J. Owirka, W. S. Brower, and A. L. Weaver, "The automatic target-recognition system in SAIP," *Lincoln Lab. J.*, vol. 10, no. 2, pp. 1–16, 1997.
[37] H. Wang, S. Chen, F. Xu, and Y.-Q. Jin, "Application of deep-learning algorithms to MSTAR data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 3743–3745.
[38] C. Coman and R. Thaens, "A deep learning SAR target classification experiment on MSTAR dataset," in *Proc. 19th Int. Radar Symp. (IRS)*, Bonn, Germany, Jun. 2018, pp. 1–6.
[39] M. Amrani, K. Yang, D. Zhao, X. Fan, and F. Jiang, "An efficient feature selection for SAR target classification," in *Proc. Pacific Rim Conf. Multimedia*, 2017, pp. 68–78.
[40] M. Amrani, F. Jiang, Y. Xu, S. Liu, and S. Zhang, "SAR-oriented visual saliency model and directed acyclic graph support vector metric based target classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3794–3810, Oct. 2018.
[41] M. Amrani and F. Jiang, "Deep feature extraction and combination for synthetic aperture radar target classification," *J. Appl. Remote Sens.*, vol. 11, no. 4, p. 1, Oct. 2017.
[42] M. Amrani, A. Bey, A. Amamra, "New SAR target recognition based on YOLO and very deep multi-canonical correlation analysis," *Int. J. Remote Sens.*, pp. 1–20, 2021.
[43] B. Shi, Q. Zhang, Y. Li, and Y. Li, "SAR image target recognition based on improved residual attention network," *Laser Optoelectronics Prog.*, vol. 58, no. 8, 2021, Art. no. 0810008, doi: 10.3788/LOP202158.0810008.
[44] Y. Chen, L. Yu, and X. Xie, "SAR image target classification based on all convolutional neural network," *Radar Sci. Technol.*, vol. 16, no. 3, pp. 242–248, 2018.
[45] X. Hu, Q. Yao, B. Hou, H. Song, and H. Lei, "Target recognition using convolutional neural network for SAR images," *Sci. Technol. Eng.*, vol. 19, no. 21, pp. 228–232, 2019.
[46] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.
[47] Z. Zhang, Y. Peng, and G. Cheng, "Optimization of image classification model based on CIFAR-10," *Comput. Appl. Softw.*, vol. 35, no. 3, pp. 177–181, 2018.
[48] D. Bankman, L. Yang, B. Moons, M. Verhelst, and B. Murmann, "An always-on 3.8 μJ/86% CIFAR-10 mixed-signal binary CNN processor with all memory on chip in 28-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 54, no. 1, pp. 158–172, Oct. 2018.
[49] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 4510–4520.
[50] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," 2016, *arXiv:1602.07360*. [Online]. Available: http://arxiv.org/abs/1602.07360

BAODAI SHI was born in 1996. He received the B.S. degree from the Xi'an University of Electronic Science and Technology, Xi'an, China, in 2019. He is currently pursuing the M.S. degree with the Air and Missile Defense College, Air Force Engineering University. At present, he mainly studies the related technologies of synthetic aperture radar target recognition and deep learning.

QIN ZHANG was born in1973. He received the B.S. and M.S. degrees from the Xi'an University of Electronic Science and Technology, Xi'an, China. He is currently a Professor with the Air and Missile Defense College, Air Force Engineering University. His research interests include radar signal processing and deep learning.

DAYAN WANG received the B.S. degree in mechanical engineering from China Air Force Engineering University, in 2019. He is currently pursuing the M.S. degree with the Air Defense Anti-Missile Academy, Air Force Engineering University, Xi'an. His research interests include HMI design, cognitive psychology, and computer simulation.

YAO LI was born in 1996. She received the B.S. degree from the Xi'an University of Post and Telecommunications, Xi'an, China, in 2018. She is currently pursuing the M.S. degree with the Air and Missile Defense College, Air Force Engineering University. At present, she mainly studies the related technologies of radar gesture recognition and deep learning.