





SAR Target Classification Based on Integration of ASC Parts Model and Deep Learning Algorithm

Sijia Feng , Kefeng Ji , *Member, IEEE*, Linbin Zhang , Xiaojie Ma ,
and Gangyao Kuang, *Senior Member, IEEE*

Abstract—Automatic target recognition of synthetic aperture radar (SAR) images has been a vital issue in recent studies. The recognition methods can be divided into two main types: traditional machine learning methods and deep-learning-based methods. For most traditional machine learning methods, target features are extracted based on electromagnetic scattering characteristics which are interpretable and stable. However, the extraction process of effective recognition features is often complex and the computational efficiency is low. Compared with the traditional methods, the deep learning methods can directly learn the high-dimensional features of the target to obtain higher target recognition accuracy. However, these algorithms have poor generalization performance and are difficult to explain. In order to comprehensively consider the advantages of the two kinds of methods, this article proposes a novel method for SAR target classification based on integration parts model and deep learning algorithm. First, part convolution and modified bidirectional convolutional-recurrent network are used to extract local feature of target through parts model which is calculated based on attribute scattering centers. Then, modified all-convolutional networks are used to extract the global feature of the target. The final classification result is achieved through decision fusion of local and global features. Experimental results on the moving and stationary target acquisition and recognition show the superiority of the proposed method, especially under complex conditions. Besides, a brief analysis of target key parts with part occlusion method is given, which is helpful to the interpretability of the deep learning network.

Index Terms—Attribute scattering center (ASC) model, deep learning, part model, synthetic aperture radar (SAR), target classification.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) has strong applicability in both military and civilian fields with the advantages of all-time, all-weather, long-range, and high-resolution imaging ability [1], [2]. Because of its various advantages, SAR is widely used in the field of remote sensing image target detection [3]–[6], target recognition [7], [8], terrain classification [9], change

Manuscript received June 30, 2021; revised August 5, 2021 and September 17, 2021; accepted September 24, 2021. Date of publication October 1, 2021; date of current version October 18, 2021. This work was supported by the National Natural Science Foundation of China under Grant 62001480 and Grant 61372163. (*Corresponding author: Kefeng Ji.*)

The authors are with the State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System and the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China. (e-mail: fengsijia12@nudt.edu.cn; jikefeng@nudt.edu.cn; zlbndt@163.com; mxj286@foxmail.com; kuangyeats@hotmail.com).

Digital Object Identifier 10.1109/JSTARS.2021.3116979

detection [10], and so on. In these applications, automatic target recognition (ATR) plays an important role in SAR civil and military fields. A standard SAR–ATR system consists of three main parts proposed by the MIT Lincoln Laboratory: detection, discrimination, and classification [11]. The purpose of target classification is to automatically classify each input target image which is obtained by target detection and discrimination.

SAR target classification methods can be divided into two main types: traditional machine learning methods and deep-learning-based methods. Traditional machine learning methods mean recognize targets through manually extracted target features and traditional classifiers such as support vector machine [12] and sparse-representation-based classification (SRC) [13]. Some commonly used target features include geometric features [14], transform domain features [15], and electromagnetic scattering features [16]. Among the traditional methods, the target classification methods based on attribute scattering center (ASC) model [17] have received much attention in recent years. The model is based on the electromagnetic scattering characteristics of the target and can reflect the local structure of the target. The traditional target recognition method combines the structural characteristics of the target closely when calculating the target features, which is interpretable. Its shortcomings are also obvious, mainly reflected in the complexity of feature design, low computational efficiency, and low recognition accuracy.

In recent years, deep learning techniques applied in many technical fields have made a breakthrough [18], [19], making deep learning a valuable choice for SAR target classification [20]. For example, convolutional neural network (CNN) [21] can learn complex features from the original image with convolution layers and pool layers. VGGNet [22], ResNet [23], and DenseNet [24] are the major CNN architectures which are always used as backbone structures in SAR image interpretation methods [25]. Besides CNN, RNN [26] is another major type of deep network. Long short-term memory (LSTM) [27] is one of the most popular architectures in RNNs. This structure is widely used in the research of multiview SAR images [28]. At present, there are three main research targets of deep-learning-based SAR image target recognition methods: aircraft target [29], ship target [30], and vehicle target [31]. All these researches need the support of abundant target data. The moving and stationary target acquisition and recognition (MSTAR) dataset [32] serves as a benchmark for SAR vehicle target classification algorithms evaluation and comparison. Many pieces of research based on deep learning are carried out on this dataset because of its

standardized data format and various operating conditions. The advantage of deep learning method is that it can automatically learn the high-dimensional features of the target, and does not need the complex work of artificial feature extraction. Moreover, in the same training and testing environment, a higher target recognition accuracy can be obtained. However, compared with the traditional methods, the deep learning model is not interpretable enough. At the same time, the robustness of the model is insufficient.

Through the analysis and comparison of the above two kinds of methods, we believe that the performance of SAR image target recognition algorithm can be further improved by fusing traditional features and deep learning model.

A. Related Work

In the traditional recognition methods, the method based on electromagnetic scattering features is a research focus. This type of method reflects the specific structure of the target and describes the scattering center of SAR image from the perspective of SAR imaging. Theoretical and experimental analysis shows that the high-frequency scattering response of complex target can be well modeled as the sum of the response of multiple ASCs [33]. Therefore, the reflection behavior model of the target can be described by ASCs, which can be used as candidate features for target recognition [34]. Chiang *et al.* [35] proposed a Bayesian method to establish a one-to-one correspondence between two sets of ASCs and used *a posteriori* probability to evaluate their similarity. Dungan *et al.* [36] used the least trimmed square Hausdorff distance to match the scattering center sets. Fu *et al.* [37] proposed a novel template matching-based aircraft recognition method with scattering structure feature to improve classification accuracy in SAR images. Ding *et al.* [38] proposed a hierarchical fusion method of the global and local features for SAR ATR. SRC is used to process global features, and local descriptors are calculated by ASC. Hungary algorithm is then used for target matching. It can be found from the above literature that traditional methods need complex feature extraction and classifier design, which is inefficient in the use of large-scale data.

Different from traditional recognition methods, the deep learning method does not consider the electromagnetic scattering characteristics of the target and learns high-dimensional target features directly through the network. Chen *et al.* [39] first proposed a fully all-convolutional network (A-ConvNet) for SAR vehicle target classification which replaces all the fully connected layers with convolutional layers. Shang *et al.* [40] also proposed a network structure named M-net for vehicle target classification. The network is trained with a two-step method which can reduce the training time and make the training more stable. The structure of convolutional highway unit is proposed in [41], which enables the network to obtain deeper feature expression. When the training data is reduced to 30%, it can still maintain good target classification performance. In [42], a transfer-learning-based method is proposed, which makes knowledge learned from sufficient unlabeled SAR scene images transferrable to labeled SAR target data. It can be seen from the

above research results that deep learning method can achieve higher classification accuracy, but a large number of network parameters and the limited number of training samples are easy to cause overfitting. Therefore, data augmentation and parameters transfer are needed to improve the network robustness. However, at present, the construction of deep-learning-based SAR image target recognition algorithm lacks the guidance of target electromagnetic scattering characteristics. The performance of the algorithm will decline significantly under complex conditions.

According to the existing literature, the fusion methods can be divided into three main strategies: early feature fusion (EFF), late feature fusion (LFF), and decision level fusion (DLF) [43]. Eff refers to fusion using low-level features calculated based on original data and LFF refers to fusion using higher level features. The level of the feature is defined based on the depth of the network used. DLF means that the output probability of each fused method is averaged to obtain the final output probability. Cui *et al.* [44] proposed an updated learning method for SAR unlabeled target recognition based on the decision fusion of CNN and other assistant classifiers. Zhang *et al.* [45] proposed a feature fusion framework based on ASC feature and deep learning feature for the first time. The bag of visual word (BOVW) model and discrimination correlation analysis are introduced in the article. The proposed method proves superior effectiveness and robustness under different operation conditions.

B. Contributions

In order to combine the advantages of traditional methods and deep learning methods, we propose a novel method for SAR target classification based on the integration of ASC parts model and deep learning model. Compared with recent SAR target classification methods, the main contributions of the article are as follows:

- 1) We propose a novel parts model for SAR target based on ASC model. The parts are calculated based on electromagnetic scattering characteristics of target local structure which are the basis of the whole target recognition algorithm. Part convolution (Part Conv) operation is introduced based on parts model, which creatively transforms the traditional electromagnetic scattering features into the data that can be calculated by convolution neural network.
- 2) Modified bidirectional convolutional-recurrent network (MBCRN) is proposed to extract local feature of targets based on the sequence of electromagnetic scattering features which synthesizes the scattered target parts. Besides, a global feature extraction architecture based on modified A-ConvNets is proposed. A decision fusion strategy is introduced to fuse the results achieved from local feature extraction and global feature extraction which improves the classification result compared with every single branch.
- 3) We innovatively use the part occlusion method to analyze the key parts that affect the target recognition results. The visualization results can clearly show the knowledge learned by our method for different types of targets. In addition, the results are consistent with our expert knowledge, which further shows the rationality of our method.

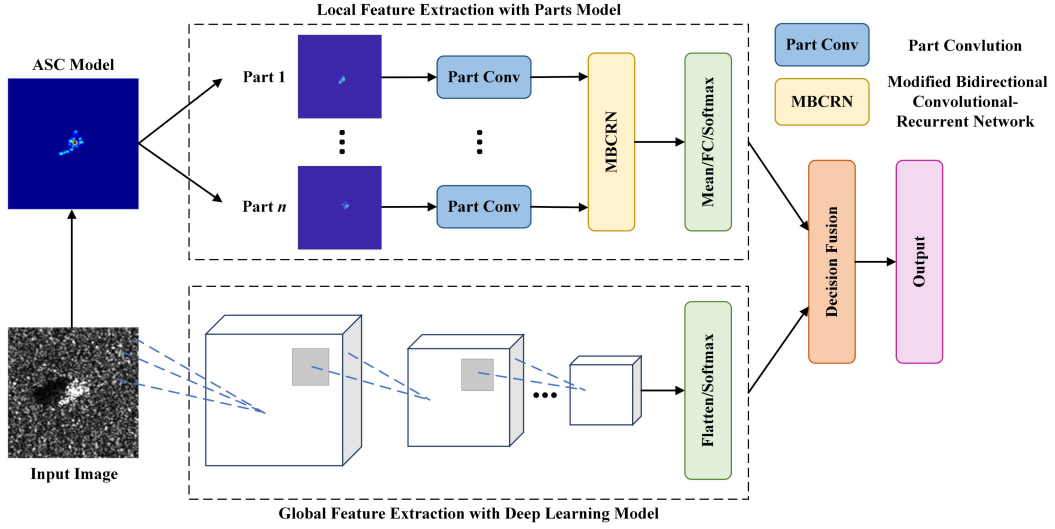


Fig. 1. Overall structure of proposed integration method.

We carry out a lot of experiments and the experimental results show the superiority of the proposed method which achieves higher classification accuracy than the method used in [45]. Different from their previous work, our method combines discrete ASCs into target parts to better reflect the local structural characteristics of the target. Compared with the BOVW model, target parts model contains more electromagnetic scattering information and provides support for the interpretability analysis of the algorithm.

The rest of this article is organized as follows. Section II gives the detailed structure of proposed method and the total training process. Section III presents the experimental results of proposed method compared with recent algorithms and gives a brief analysis on target key parts which are important for classification. Finally, Section IV concludes the article.

II. METHODOLOGY

The overall structure of the proposed method is shown in Fig. 1. The detailed process of the target classification method is described in this section. In Section II-A, we first introduce the target parts extraction method based on ASC model. Sections II-B II-C show the methods of local feature extraction with parts model and global feature extraction with deep learning model, respectively. Decision fusion method based on these two kinds of features is described in Section II-D. Section II-E gives total training and test process.

A. Target Parts Extraction With ASC Model

ASC model is widely used in target recognition field. The model assumes that the backscattering of a target can be well approximated as a sum of responses from individual scattering centers as given in the following:

$$E(f, \varphi; \Theta) = \sum_{i=1}^q E_i(f, \varphi; \theta_i) \quad (1)$$

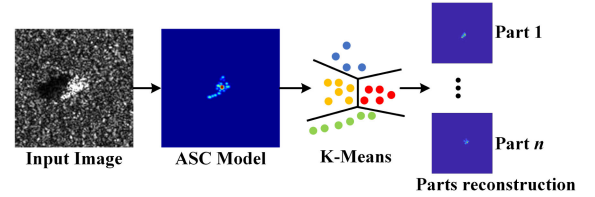


Fig. 2. Target parts extraction process based on ASC model.

where q is the total number of individual scattering centers. For an individual ASC, the backscattered field is parameterized as a function of frequency f and aspect φ :

$$E_i(f, \varphi; \theta_i) = A_i \cdot \left(j \frac{f}{f_c} \right)^{\alpha_i} \cdot \exp[-j \frac{4\pi f}{C} (x_i \cos \varphi + y_i \sin \varphi)] \cdot \text{sinc} \left(\frac{2\pi f}{C} L_i \sin(\varphi - \bar{\varphi}_i) \right) \cdot \exp(-2\pi f \gamma_i \sin \varphi) \quad (2)$$

where f_c is the radar center frequency and C is the propagation velocity of electromagnetic wave.

The $\Theta = \{\theta_i = [A_i, \alpha_i, x_i, y_i, L_i, \bar{\varphi}_i, \gamma_i]\} (i = 1, 2, \dots, q)$ in (2) is the parameter set in which (x_i, y_i) are the spatial locations, A_i is the relative amplitude, α_i denotes the frequency and geometry dependence, L_i and $\bar{\varphi}_i$ represent the length and orientation of the distributed ASC, respectively, and γ_i is the aspect dependence of the localized ASC. The seven parameters in ASC model can be predicted by many parameter optimization algorithms. In this article, genetic algorithm [46] is adopted for ASC calculation.

After parameters prediction, we can extract target parts images based on ASC results according to the following process shown in Fig. 2. We use a BMP2 target in the MSTAR dataset for presentation.

First, the K-means clustering method [47] is applied to select scattering centers belonging to different parts. The clustering

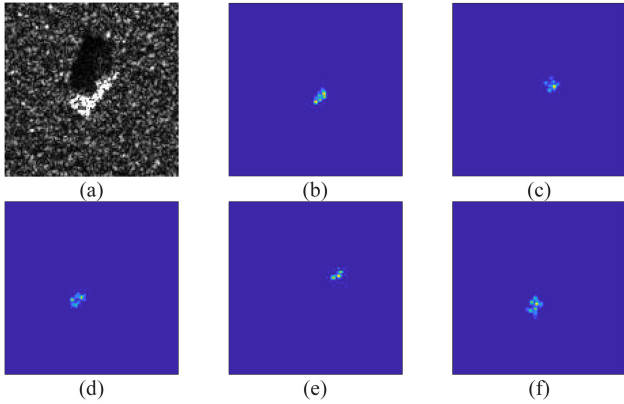


Fig. 3. Parts extraction result of a vehicle target. (a) Original image. (b)–(f) Different parts.

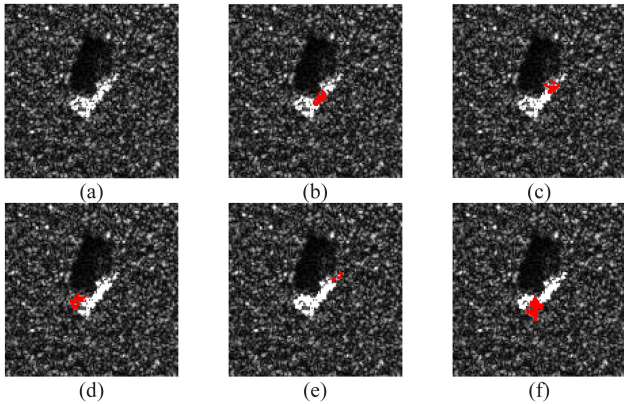


Fig. 4. Positions of target parts on original image. (a) Original image. (b)–(f) Position of each target part.

object of K-means is the seven parameters of each ASC. Then, the clustering algorithm divides the original ASCs into multiple sets. For each scattering centers cluster, we calculate the reconstructed image based on the predicted parameter set as in (3):

$$E_R(f, \varphi; \Theta_R) = \sum_{i=1}^p E_{Ri}(f, \varphi; \theta_{Ri}). \quad (3)$$

E_R denotes the reconstructed data in the frequency domain by the extracted ASCs. $\Theta_R = \{\theta_{Ri}\} (i = 1, 2, \dots, p)$ is the predicted parameter set. p represents the scattering centers number in certain cluster. Then, two-dimensional (2-D) inverse Fourier transform is applied to E_R to gain the reconstructed part image.

Fig. 3 shows parts extraction result of a vehicle target in MSTAR dataset with the parts number set to 5. Fig. 4 shows the positions of target parts extracted in Fig. 3 on original image. It can be seen from the figure that the target parts extracted by this method cover the main area of the target. For BMP 2 target, the main recognition feature is that the turret is in the middle and rear. It can be seen from Fig. 4 that parts (b) and (f) constitute the main structure of the turret.

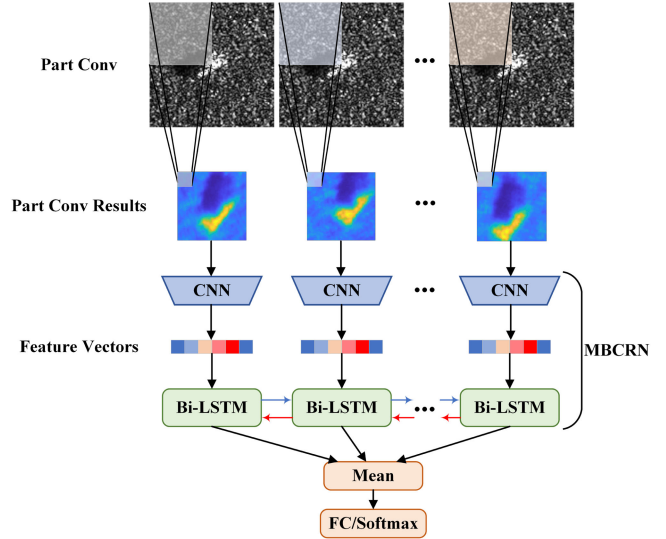


Fig. 5. Detailed structure of local feature extraction method. “FC” means fully connected layer.

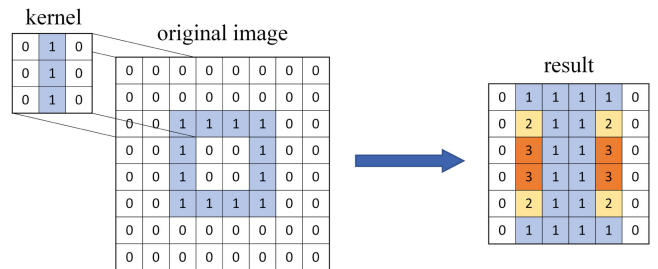


Fig. 6. Simple illustration of the convolution process and result.

B. Local Feature Extraction With Part Conv and MBCRN

As shown in Fig. 1, local feature extraction method is mainly composed of two parts: Part Conv and MBCRN. The detailed structure of local feature extraction method is shown in Fig. 5.

Since ASC model provides the physical and geometric description of the target, the reconstructed parts images reflect the distribution of different attribute features of the target. From the reconstructed image, we can know the position and structure of each part of the target. In order to integrate that information into the algorithm, we need to transform those descriptive features into network recognizable features. According to our understanding of convolution operation in deep learning, it can extract specific features according to different convolution kernels. Fig. 6 shows a simple calculation process and result of convolution. The convolution kernel is a component structure of the original image. It can be seen from the figure that the convolution result highlights the characteristics of the structure corresponding to the convolution kernel in the original figure, including the shape and position. That is to say, through this operation, we can convert the descriptive information of a certain structure at a location in the original map into visualized information that can be recognized by the network. Therefore, using the reconstructed parts images as the convolution kernels

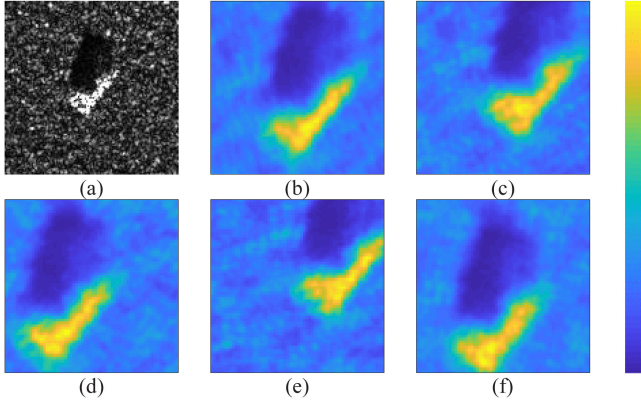


Fig. 7. Part convolution results of a vehicle target. (a) Original image. (b)–(f) Part convolution results with different parts kernels.

to convolute with the corresponding SAR image can extract the local attribute features of the image. We call this process Part convolution.

During this process, we first cut the reconstructed parts images to 65×65 . Then, Part Conv results of corresponding SAR image are calculated with part kernels. Fig. 7 shows the Part Conv results of a vehicle target in MSTAR dataset with five parts. The brighter the area, the larger the value after convolution. From convolution results, we can find that under different convolution kernels, the convolution results are obviously different. We can see that the results not only reflect the scattering values of the parts but also reflect the position of the parts. Combining with Figs. 4 and 7, it can be found that when the target part changes, the convolution result will change at the corresponding position of the part. At the same time, the convolution result image will shift to the position of the part in the target. Through this operation, we can convert the descriptive information of a certain structure at a location in the original map into visualized information that can be recognized by the network.

To integrate electromagnetic scattering features with deep learning algorithm, we set the Part Conv results as the input of MBCRN. Our MBCRN structure is designed for single-channel data which is different from the original BCRN proposed in [28]. Since the Part Conv results are single-channel data, many parameters of original network need to be revised. First, a few convolutional layers are applied to convert Part Conv results into feature vectors. Detailed parameter setting of the convolutional layers is presented in Fig. 8.

Through the network, the corresponding 1-D feature of each part image can be obtained. In order to comprehensively consider the feature of each component, we use the bidirectional LSTM (Bi-LSTM) network [48] structure. It consists of a forward sequence and a backward sequence of the LSTM network. LSTM network has cell architecture and the detailed structure of a single cell is shown in Fig. 9. $X = \{x_n\}, n = 1, 2, \dots, N$ means the input vector sequence of LSTM, $Y = \{y_n\}, n = 1, 2, \dots, N$ means the output vector sequence, and $C = \{C_n\}, n = 1, 2, \dots, N$ means the state of each cell, where N is the length of the image sequence. $f_n, i_n,$ and o_n represent

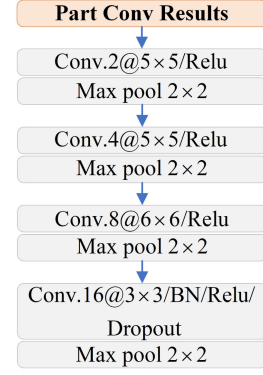


Fig. 8. Detailed parameter setting of the convolutional layers. “BN” means batch normalization module.

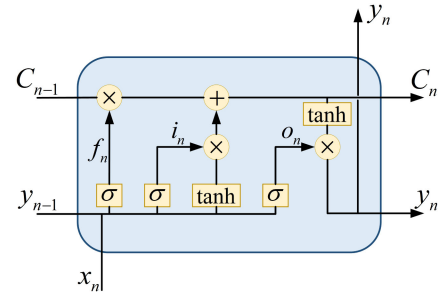


Fig. 9. Detailed structure of a LSTM network cell [28].

the output values of three different gates defined in LSTM: forget gate, input gate, and output gate. σ and \tanh are the logistic sigmoid function and the hyperbolic tangent function, respectively.

Then, at step n , the forward propagation in the forward LSTM network is described as follows:

$$\begin{aligned}
 f_n &= \sigma(W_f[y_{n-1}, x_n] + b_f) \\
 i_n &= \sigma(W_i[y_{n-1}, x_n] + b_i) \\
 C_n &= f_n \times C_{n-1} + i_n \times \tanh(W_C[y_{n-1}, x_n] + b_C) \\
 o_n &= \sigma(W_o[y_{n-1}, x_n] + b_o) \\
 y_n &= o_n \times \tanh(C_n)
 \end{aligned} \tag{4}$$

where each W and b represents the weight and offset of the corresponding network layer. For reverse LSTM networks, except that the sequence parameters are entered in the opposite order, the other calculation methods are similar to forward networks. We denote the output vector sequence of reverse LSTM networks as $Y' = \{y'_n\}, n = 1, 2, \dots, N$. Then, the output of the Bi-LSTM network is the combination of the output for the forward and reverse networks, that is,

$$Y_{\text{Bi}} = [Y; Y'] \tag{5}$$

The sequence size of the Bi-LSTM layer we used is the same as the parts number and the hidden size of each step is set to 32. Since we use a bidirectional LSTM structure, the final output size of each step in Bi-LSTM is 64. Then, we calculate the average of

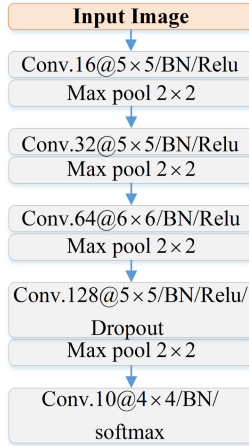


Fig. 10. Detailed architecture of modified A-ConvNets. “BN” means batch normalization module.

the output vector sequence. Finally, a fully connected layer and softmax function are used to calculate the probabilities of the predicted categories. The output of the local feature extraction branch is a 1-D label probability distribution vector.

C. Global Feature Extraction With Modified A-ConvNets

To extract global features of each target, we use a modified network structure based on the structure of A-ConvNets [39] which can achieve high classification accuracy on MSTAR dataset. The detailed architecture is shown in Fig. 10.

Different from the local feature extraction branch, the input of this modified network is the whole target image. We use five trainable layers to extract features of different targets. The main difference between the original A-ConvNets and our modified architecture is network parameter setting. The input image size of the original network is 88×88 pixels. To simplify the image preprocessing, the input size of our network is 128×128 . We add max pool layers and change convolution kernel size to satisfy the input size. Besides, we add a batch normalization module [49] after each convolution layer. The advantage of the modified network structure is that it needs neither cropping the input images nor conducting data augmentation steps during the training process.

From the above network, we can also get a 1-D label probability distribution vector which is learned from the whole target image with a global feature extraction branch.

D. Target Recognition Based on DLF

After we get two 1-D probability distribution vectors from the local feature extraction branch and global feature extraction branch, respectively, we need to fuse the two vectors to get the final recognition result. There are many fusion strategies such as EFF, LFF, and DLF [43] as mentioned in the previous introduction. The difference between these three strategies is that the fusion operation is located at different levels of the algorithm. DLF is the simplest form of fusion between different approaches where the output probabilities of each individual approach are

averaged. Considering the simplicity of calculation, the final target recognition result is achieved based on the DLF of these two vectors as shown in (6):

$$P_{\text{final}} = \frac{1}{2} \sum_{i=1}^{\text{classnum}} \left(p_i^{\text{local}} + p_i^{\text{global}} \right). \quad (6)$$

P_{final} means the final output probability distribution vector of the proposed method. p_i^{local} and p_i^{global} denote the probability of a certain class calculated through different feature extraction branches.

E. Training and Test Process

The complete training process of the proposed method is as follows. In preprocessing step, we first predict the parameters of the ASC model of each target image. Then, target parts images are reconstructed based on prediction results and clustering operation. After that, we take the whole target image and parts images sequence as the input of our network to extract two 1-D probability distribution vectors based on local feature and global feature, respectively. The final output of the proposed method is the mean of two probability distribution vectors. In this article, the cross entropy loss [50] is used to calculate the error for backward propagation, and the stochastic gradient descent [51] is used to optimize parameters. During the test process, parameters of the ASC model and parts images of each test target also need to be calculated at first. Through forward propagation of the whole network, we can get the final label probability distribution vector for each test target. Then, the predicted label is the one with the highest probability. Compared with true labels, the classification accuracy of the algorithm is obtained.

III. EXPERIMENTS

To verify the performance of the proposed method, we first introduce the images of the MSTAR dataset in Section III-A, and generate the training and test dataset under the standard operating condition (SOC) and extended operating conditions (EOCs), respectively. In Section III-B, we first analyze the influence of different parameter settings on the performance of our method. Then, we compare the classification performance of the proposed algorithm and other algorithms under different conditions. In Section III-C, ablation experiments are conducted to illustrate the effect of the proposed modules. In addition, we also perform performance verification under more complex conditions such as limited training data and random occlusions presented in Section III-D. Finally, we briefly discuss the key parts that affect the results of target classification in Section III-E.

All experiments are carried out with both Matlab platform and the Pytorch framework on a NVIDIA GeForce RTX 2080 Ti GPU card.

A. Datasets

The experiments are conducted on the MSTAR public release dataset. This dataset is collected by the Sandia National

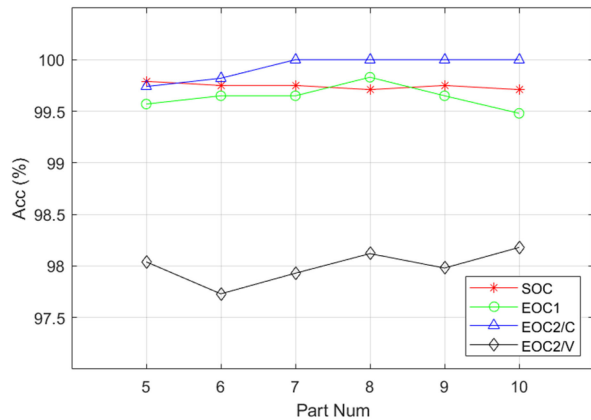


Fig. 11. Classification accuracy of proposed method with different parts numbers.

Laboratory using a Twin Otter SAR sensor operating at X-band HH polarization with $0.3 \text{ m} \times 0.3 \text{ m}$ resolution [32]. It is taken at multiple depression angles (15° , 17° , 30° , 45°) over a full 0° – 360° range of viewing angles. The dataset includes 10 classes of military targets: BMP2, BTR70, T72, T62, BTR60, 2S1, BRDM2, D7, ZIL131, and ZSU23/4. The dataset can be used under different operation conditions with different data selections. The data compositions under SOC and EOCs are presented in Tables I–IV.

SOC refers to the training data and test data under similar imaging conditions, using 17° depression angle data of 10 kinds of targets as training data and 15° depression angle data as test data. EOC1 corresponds to large depression angle variation, where the training set has 17° depression angle and the test set has 30° depression angle. EOC2/C represents configuration variation of vehicle targets, where training data and test data differ in components. EOC2/V means target version variation, where some structures have relative motion among the training and test sets [32].

B. Results of SOC and EOCs

1) *Analysis of Different Parts Number*: We first experiment with the proposed method under different parts numbers to show the influence of parts numbers on the classification performance of the algorithm. From Section II-A, we can find that under different cluster numbers, we can get different numbers of target parts. Since the results of the ASC prediction method used in this article include 30–40 main scattering centers, we only do experimental analysis on five to ten parts. More parts will lead to too detailed division and a lack of representativeness. The parameters of the proposed method change with different parts numbers, which are mainly reflected in the sequence size of Bi-LSTM. Experiment results of SOC and EOCs are shown in Fig. 11 and Table V with different parts numbers.

From the results, it can be found that different experimental conditions have different preferences for the number of target parts. For SOC, the classification performance of the algorithm is less affected by the number of target parts. However, for EOC1,

TABLE I
TRAINING AND TEST EXAMPLES FOR SOC

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

TABLE II
TRAINING AND TEST EXAMPLES FOR EOC1

Train			Test		
Type	Depression	Num	Type	Depression	Num
T72	17°	299	T72	30°	288
2S1	17°	299	2S1	30°	288
BRDM2	17°	298	BRDM2	30°	287
ZSU23/4	17°	299	ZSU23/4	30°	288

TABLE III
TRAINING AND TEST EXAMPLES FOR EOC2/C

Train			Test		
Type/ Series	Depression	Num	Type/ Series	Depression	Num
BMP2/9563	17°	233	T72/A32	$15^\circ, 17^\circ$	572
BTR70/C71	17°	233	T72/A62	$15^\circ, 17^\circ$	573
T72/132	17°	232	T72/A63	$15^\circ, 17^\circ$	573
BRDM2/E71	17°	298	T72/A64	$15^\circ, 17^\circ$	573
			T72/S7	$15^\circ, 17^\circ$	419

the classification performance of the algorithm first increases and then decreases with the change of the number of target parts. We consider that the training data and test data in SOC are with the same serial number of each target type and the same depression angle. The recognition of different types of targets depends more on large-scale features than on detailed features. Due to the large depression angle change in EOC1, the target parts with a certain size will show stable characteristics, while the more detailed structure will have larger changes in electromagnetic scattering characteristics. Therefore, under EOC1, the classification performance of the proposed method improves first when the number of target parts increases. When the target parts are divided too much, the performance will be degraded. For EOC2s, although there are some fluctuations, the overall classification performance is on the rise with the increase in the number of parts. We consider that the configuration of the target changes significantly in EOC2s, so the large-scale parts of targets of different versions in the same category are quite

TABLE IV
TRAINING AND TEST EXAMPLES FOR EOC2/V

Train			Test		
Type/ Series	Depression	Num	Type/ Series	Depression	Num
BMP2/9563	17°	233	BMP2/9566	15°,17°	428
BTR70/C71	17°	233	BMP2/C21	15°,17°	429
T72/132	17°	232	T72/A04	15°,17°	573
BRDM2/E71	17°	298	T72/A05	15°,17°	573
			T72/A07	15°,17°	573
			T72/A10	15°,17°	567
			T72/812	15°,17°	426

TABLE V
COMPARISON OF CLASSIFICATION ACCURACY BETWEEN DIFFERENT PARTS
NUMBERS UNDER SOC AND EOCs

Part Num	Accuracy (%)			
	SOC	EOC1	EOC2/C	EOC2/V
5	99.79	99.57	99.74	98.04
6	99.75	99.65	99.82	97.73
7	99.75	99.65	100	97.93
8	99.71	99.83	100	98.12
9	99.75	99.65	100	97.98
10	99.71	99.48	100	98.18

TABLE VI
COMPARISON OF CLASSIFICATION ACCURACY BETWEEN DIFFERENT
METHODS UNDER SOC AND EOCs

Method	Accuracy (%)			
	SOC	EOC1	EOC2/C	EOC2/V
A-ConvNets [39]	99.13	96.12	98.93	98.60
VDCNN [52]	98.52	94.61	95.46	95.45
M-Net [40]	99.71	97.48	99.67	98.74
FEC [45]	99.59	99.19	-	98.48
BCRN [28]	99.97	99.54	99.81	98.92
Proposed	99.79	99.57	99.74	98.04

different, but the details are similar. Thus, large parts number performs better in EOC2s.

Because the number of parts has no obvious influence on the performance of the method under SOC and we focus on SOC data in the following experiments, we set the parts number of the proposed method as five for convenience and higher computational efficiency.

2) *Analysis of Different Methods*: In order to illustrate the performance of the proposed method, we compare it with the existing methods. Table VI presents the classification accuracy of different methods under SOC and EOCs. From the results of the table below, we can find that our algorithm can achieve high accuracy which has reached the advanced level in this field.

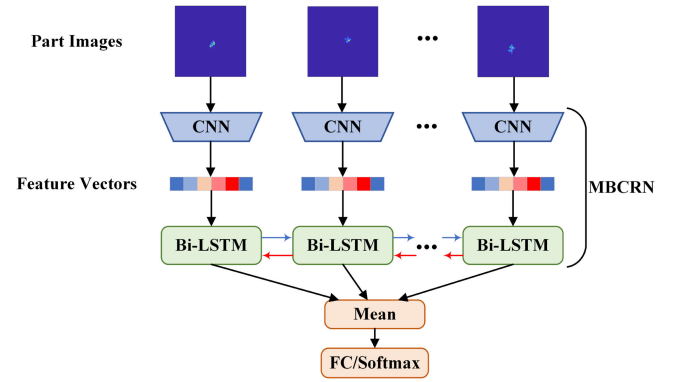


Fig. 12. Detailed modified local feature extraction structure. “FC” means fully connected layer.

The experimental results show that both BCRN and the proposed method can achieve high accuracy under different conditions. BCRN is only slightly higher than our method under some conditions. However, we compare the computational efficiency of the proposed algorithm with that of BCRN and find that BCRN is less efficient. We employ frame per second (FPS) to measure the time efficiency of different methods. In this article, it means the number of target chip images classified per second. The FPS of BCRN is 11.66, whereas our proposed method is 33.68, which is significantly higher than that of BCRN.

C. Results of Different Architectures

In this section, in order to verify the effectiveness of the proposed Part Conv module and decision fusion strategy in the target recognition, we conduct ablation experiments with different variants of the proposed structure for comparison.

The first variant is the network structure proposed in Section II-C which is used to extract the global feature of SAR targets, denoted as Global. We only use this branch for data training and testing under different conditions. Comparing the experimental results with the complete method, we can analyze the role of local features learned from part images. The second variant is to remove the Part Conv module in the proposed method, while the global feature extraction branch is maintained, denoted as Global-Part. The detailed modified local feature extraction structure of this variant is shown in Fig. 12.

Different from the original local feature extraction branch proposed in Section II-B, the variant changes the input of MBCRN structure from Part Conv results into part images of 128×128 size. Thus, the CNN architecture also needs to be modified to satisfy the input size. The detailed modified CNN parameters are shown in Fig. 13.

Besides, we also test the performance of local feature extraction branches in target recognition. Local means the structure using only the local features mentioned in Section II. Global-Local represents the complete structure of the algorithm proposed in this article. We test those variants under SOC and EOCs. Then, the classification results of different architectures are compared in Table VII. From the table, we can find that our algorithm has higher classification accuracy than other variants.

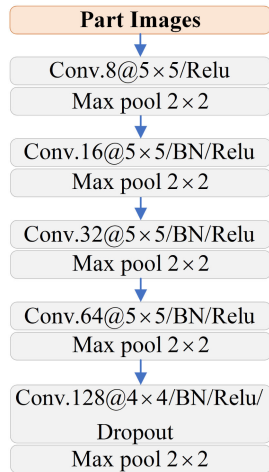


Fig. 13. Detailed parameter setting of modified CNN. “BN” means batch normalization module.

TABLE VII
COMPARISON OF CLASSIFICATION ACCURACY BETWEEN DIFFERENT ARCHITECTURES UNDER SOC AND EOCs

Method	Accuracy (%)			
	SOC	EOC1	EOC2/C	EOC2/V
Global	99.79	99.48	98.56	97.03
Global-Part	99.38	99.22	98.43	75.09
Local	76.82	91.66	99.93	92.97
Global-Local	99.79	99.57	99.74	98.04

Experimental results show that using parts images for feature extraction directly does not improve the performance, but reduces the classification accuracy. Therefore, the proposed Part Conv module can extract more useful features than the ordinary convolution layer. Comparing these experimental results, we can find that decision fusion can indeed improve classification performance, especially in EOCs. In the EOCs experiment, the classification accuracy after fusion is significantly higher than that using only global branches. In addition, in EOC1 and EOC2/V, the classification accuracy after fusion is higher than that of any branch, which shows that the part features and overall features of the target are complementary. By fusing the classification probabilities based on the calculation results of different branches, the accuracy of target classification under complex conditions can be effectively improved.

D. Results of Complex Conditions

Since different algorithms can both achieve high classification accuracy under standard experimental condition, there is a risk of overfitting. To further compare the performance of the proposed method with others, we conduct experiments under more complex conditions including limited training data and random occlusion on target images.

1) *Limited Training Data*: First, experiments with limited training data are analyzed. Classification results of different algorithms are shown in Table VIII. SOC column means test

TABLE VIII
COMPARISON OF CLASSIFICATION ACCURACY BETWEEN DIFFERENT METHODS WITH LIMITED TRAINING DATA

Method	Accuracy (%)			
	SOC	100	50	10
A-ConvNets [39]	99.13	98.52	92.04	63.75
ResNet-18 [23]	99.42	98.14	93.36	61.81
MVGGNet [45]	99.27	96.45	90.19	58.19
BCRN [28]	99.97	99.79	98.72	72.08
Proposed	99.79	99.51	99.09	82.06

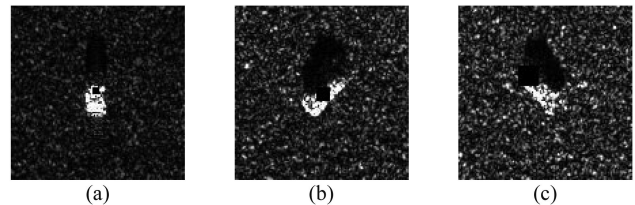


Fig. 14. Target images after random occlusion. Cover sizes of (a)–(c) are 5×5 , 10×10 , and 15×15 .

accuracy when training with full data under SOC and other columns mean test accuracy when training with corresponding number of samples in each target category under SOC. All the sample data are obtained by random sampling in the SOC training dataset. Then, we use all the test data to verify the classification performance of different algorithms.

From the table, we can find that our algorithm can still maintain a relatively stable classification performance in the case of limited training samples. Although the classification accuracy of the proposed method is slightly lower than that of BCRN under the condition of 100 samples per class, the performance of the proposed method is more robust when the sample size is smaller.

2) *Random Occlusions*: In view of the actual situation, we cover part of the target image to further test the generalization of each algorithm. We consider that the target is mainly concentrated in the central region on the slice image. Therefore, we only cover the middle area of the image and the range of central region is 64×64 . We use different cover sizes for experiments: 5×5 , 10×10 , and 15×15 . Fig. 14 shows the target image after random occlusion. The black squares in the target area are the cover area.

Because our algorithm needs to use the target parts images, for the rationality of the experiment, we occlude the parts images used in the algorithm with the same position as the corresponding target image shown in Fig. 15. It can be seen from the figure that some of the target parts are occluded and some of them are not occluded.

The classification results of different methods are shown in Table IX. It can be seen from the changes in classification accuracy of various methods in the table below that our algorithm is still very robust under different cover sizes.

When the cover size is small, both our method and BCRN can achieve very high classification accuracy. However, our method

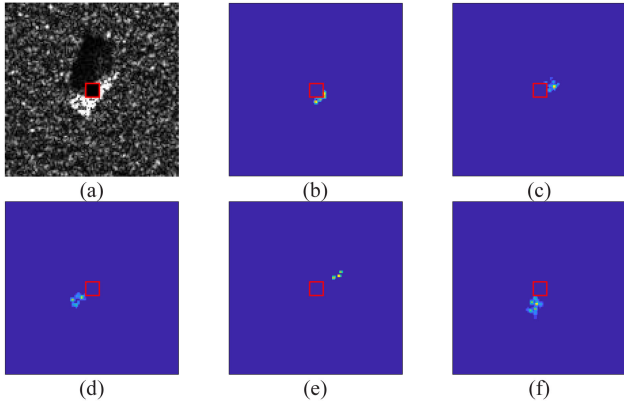


Fig. 15. Target image and parts images after random occlusion. Red box marks the position of the cover.

TABLE IX
COMPARISON OF CLASSIFICATION ACCURACY BETWEEN DIFFERENT METHODS WITH RANDOM OCCLUSIONS

Method	Accuracy (%)			
	SOC	5×5	10×10	15×15
A-ConvNets [39]	99.13	96.99	83.75	70.19
BCRN [28]	99.97	99.97	98.31	88.08
Proposed	99.79	99.71	99.26	93.73

is obviously superior to other algorithms when the size becomes larger. We believe that the advantage of the proposed algorithm is that it uses the local features related to the target parts to improve the stability of the classification. When the random occlusion does not cover the key parts that affect the target classification, the performance of our algorithm will not be greatly degraded. However, other methods all depend on the overall features of the target. Therefore, when the target area is incomplete, the performance of the algorithm will be significantly reduced.

According to the above experimental results, the proposed algorithm can achieve excellent classification performance under the conditions of limited training samples or random occlusions. Thus, we can conclude that the proposed method is more generalized and robust than other methods.

E. Analysis on Key Parts With Part Occlusion

From the above experimental comparison, we can find that the part features of proposed algorithm have a certain impact on the classification accuracy. In this part, we use the visualization method to analyze the key parts that affect the target classification. In Belloni *et al.* [53], researchers use a black square to occlude target images and the percentage of correctly classified images is used as the new intensity of the pixels located in the center of the black square in the classification map. Different from it, we occlude target parts instead of pixel squares. The visualization process is shown in Fig. 16.

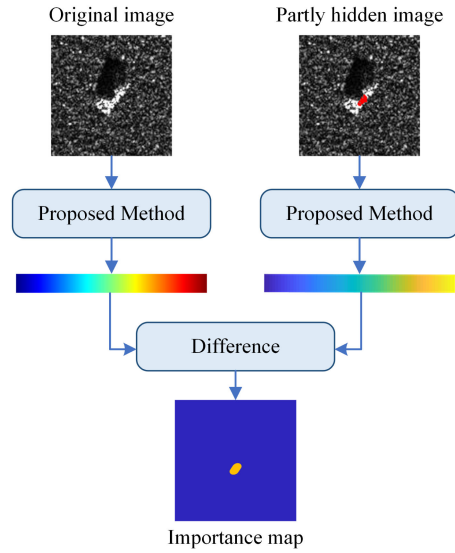


Fig. 16. Calculation process of the importance map of different target parts with part occlusion.

We use test data under SOC for analysis. First, we calculate the class probability distribution vector of each test image. Then, after occluding a part from the image, the class probability distribution vector is recalculated.

We calculate the difference of the probability values of the two probability distribution vectors corresponding to the real class labels, and take it as the importance value of the target part. The larger the difference is, the lower the recognition rate will be after removing this part of the image, that is, the more important this part is for target recognition.

We calculate the importance of each target part in turn and then visualize it to get the thermodynamic diagram of the key parts of target recognition. Fig. 17 shows the thermal distribution diagram of key parts of different types of targets at similar azimuth angles. The visualization result shows that different targets do have differences in the distribution of key parts. In order to make the conclusion more convincing, we normalize the target angle. Then, we average the distribution map of key parts of all targets in the same category and get the result as shown in Fig. 18. It can be seen from the graphic results that some key parts have important impacts on SAR vehicle target recognition. At the same time, there are separable differences in the distribution of key parts of different targets.

Fig. 19 shows examples of optical images of ten types of targets at similar aspect angles. Combined with Figs. 18 and 19, we can find that the key parts learned by the proposed algorithm are consistent with the expert knowledge of vehicle target recognition in SAR images. For example, for BMP2, the focus of manual interpretation is that the turret is obvious and located in the middle and rear of the target. These are exactly the features learned by the proposed algorithm as shown in Fig. 18(a). For ZIL131, this is a truck target, and the key to interpretation lies in the joint between the front and the body which is also can be seen in Fig. 18(i). These conclusions further demonstrate the rationality of our method.

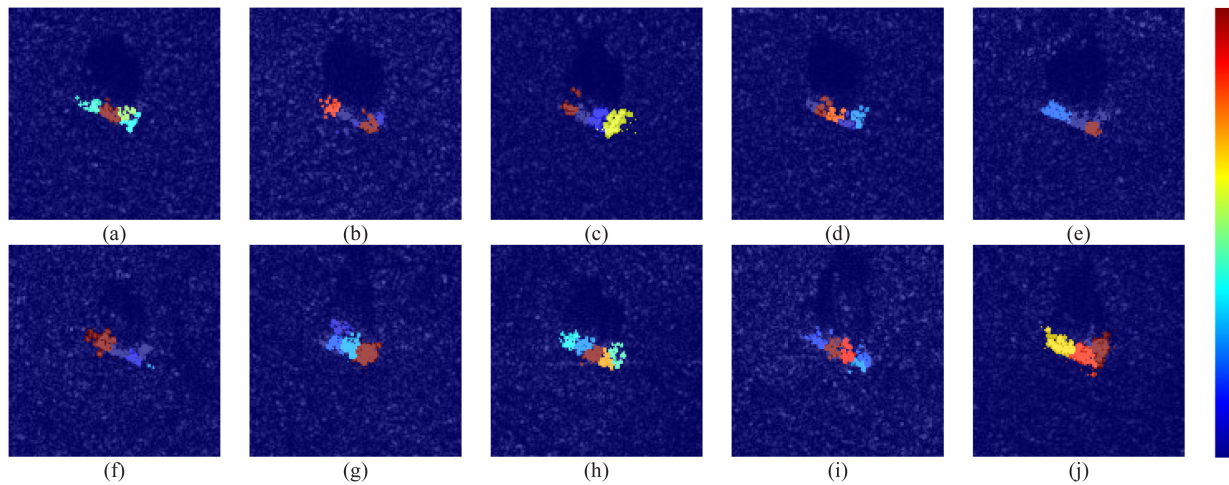


Fig. 17. Thermal distribution diagram of key parts of different types of targets at similar azimuth angles: (a) BMP2, (b) BTR70, (c) T72, (d) BTR60, (e) 2S1, (f) BRDM2, (g) D7, (h) T62, (i) ZIL131, and (j) ZSU23/4.

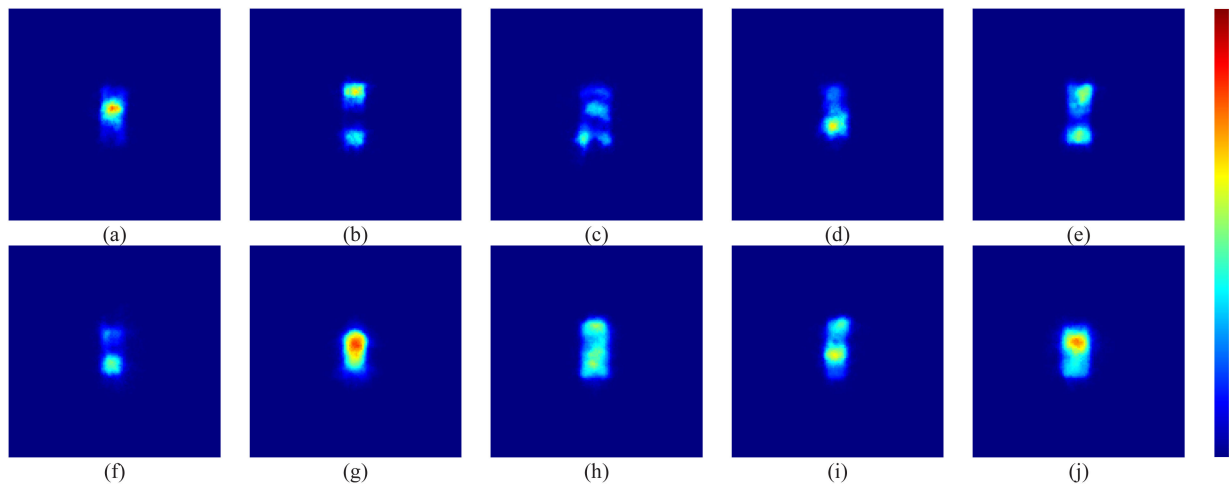


Fig. 18. Average thermal distribution diagram of key parts of different types of targets with angle normalization: (a) BMP2, (b) BTR70, (c) T72, (d) BTR60, (e) 2S1, (f) BRDM2, (g) D7, (h) T62, (i) ZIL131, and (j) ZSU23/4.

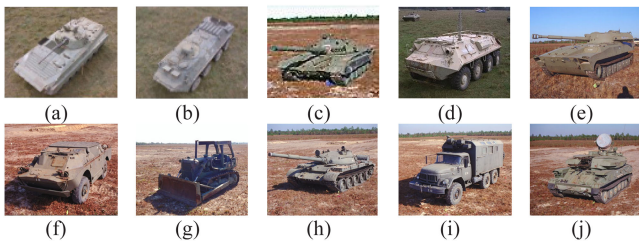


Fig. 19. Examples of optical images of MSTAR targets at similar aspect angles: (a) BMP2, (b) BTR70, (c) T72, (d) BTR60, (e) 2S1, (f) BRDM2, (g) D7, (h) T62, (i) ZIL131, and (j) ZSU23/4.

IV. CONCLUSION

Considering the advantages and disadvantages of both traditional machine learning methods and deep-learning-based methods in SAR ATR, an integration method based on

electromagnetic scattering mechanism of target and deep learning features is proposed in this article. First, target parts extraction method based on the ASC model is introduced. Then, target local feature is calculated with Part Conv and MBCRN. Modified A-ConvNets is used to extract target global features. The final classification result is the combination of the results based on both local and global features with decision fusion. Experimental results on MSTAR dataset show that our method can achieve excellent classification accuracy under SOC and EOCs. At the same time, for complex experimental conditions such as limited training samples and random occlusion, the proposed method is still stable and superior to other existing algorithms. Based on the part occlusion method, we briefly analyze the key parts that affect the results of target classification. The visualization results further prove the rationality of the proposed method and provide an experimental basis for further research on the interpretability of learning networks.

In this article, a preliminary study on the integration of the target parts model and deep learning model is carried out. In future article, we will further study the electromagnetic scattering mechanism of the target parts model and explore more effective fusion methods of different classification algorithms.

REFERENCES

- [1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.
- [2] O. Kechagias-Stamatis and N. Aouf, "Automatic target recognition on synthetic aperture radar imagery: A survey," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 36, no. 3, pp. 56–81, Mar. 2021.
- [3] Y. Zhao, L. Zhao, B. Xiong, and G. Kuang, "Attention receptive pyramid network for ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2738–2756, May 2020.
- [4] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751–755, May 2019.
- [5] X. Hua, Y. Ono, L. Peng, Y. Cheng, and H. Wang, "Target detection within nonhomogeneous clutter via total bregman divergence-based matrix information geometry detectors," *IEEE Trans. Signal Process.*, vol. 69, pp. 4326–44340, Jul. 2021.
- [6] X. Hua and L. Peng, "MIG median detectors with manifold filter," *Signal Process.*, vol. 188, 2021, Art. no. 108176.
- [7] J. Liu, M. Xing, H. Yu, and G. Sun, "EFTL: Complex convolutional networks with electromagnetic feature transfer learning for sar target recognition," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2021.3083261](https://doi.org/10.1109/TGRS.2021.3083261).
- [8] J. Zhang, M. Xing, G. Sun, and Z. Bao, "Integrating the reconstructed scattering center feature maps with deep CNN feature maps for automatic SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: [10.1109/LGRS.2021.3054747](https://doi.org/10.1109/LGRS.2021.3054747).
- [9] X. Li, L. Lei, Y. Sun, M. Li, and G. Kuang, "Multimodal bilinear fusion network with second-order attention-based channel selection for land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1011–1026, Mar. 2020.
- [10] Y. Sun, L. Lei, X. Li, X. Tan, and G. Kuang, "Patch similarity graph matrix-based unsupervised remote sensing change detection with homogeneous and heterogeneous sensors," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4841–4861, Jun. 2021.
- [11] D. E. Dudgeon and R. T. Lacoss, "An overview of automatic target recognition," *Lincoln Lab. J.*, vol. 6, no. 1, pp. 3–10, 1993.
- [12] M. Liu, Y. Wu, Q. Zhang, F. Wang, and M. Li, "Synthetic aperture radar target configuration recognition using locality-preserving property and the gamma distribution," *IET Radar Sonar Navigation*, vol. 10, no. 2, pp. 256–263, 2016.
- [13] G. Dong, G. Kuang, N. Wang, and W. Wang, "Classification via sparse representation of steerable wavelet frames on Grassmann manifold: Application to target recognition in SAR image," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2892–2904, Jun. 2017.
- [14] J.-I. Park, S.-H. Park, and K.-T. Kim, "New discrimination features for SAR automatic target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 476–480, May 2013.
- [15] Z. Cui *et al.*, "Target recognition in synthetic aperture radar images via non-negative matrix factorization," *IET Radar Sonar Navigation*, vol. 9, no. 9, pp. 1376–1385, 2015.
- [16] B. Ding and G. Wen, "Target reconstruction based on 3-D scattering center model for robust SAR ATR," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3772–3785, Jun. 2018.
- [17] L. C. Potter and R. L. Moses, "Attributed scattering centers for SAR ATR," *IEEE Trans. Image Process.*, vol. 6, no. 1, pp. 79–91, Jan. 1997.
- [18] W. Ge, X. Lin, and Y. Yu, "Weakly supervised complementary parts models for fine-grained image classification from the bottom up," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3029–3038.
- [19] L. Zhao, X. Peng, Y. Chen, M. Kapadia, and D. N. Metaxas, "Knowledge as priors: Cross-modal knowledge generalization for datasets without superior knowledge," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6527–6536.
- [20] X. Zhu *et al.*, "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, to be published, doi: [10.1109/MGRS.2020.3046356](https://doi.org/10.1109/MGRS.2020.3046356).
- [21] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–14.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [24] G. Huang, Z. Liu, K. Weinberger, and L. Maaten, "Densely connected convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.
- [25] S. Feng, K. Ji, X. Ma, L. Zhang, and G. Kuang, "Target region segmentation in SAR vehicle chip image with ACM net," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: [10.1109/LGRS.2021.3085188](https://doi.org/10.1109/LGRS.2021.3085188).
- [26] B. A. Pearlmutter, "Learning state space trajectories in recurrent neural networks," *Neural Comput.*, vol. 1, no. 2, pp. 263–269, 1989.
- [27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [28] X. Bai, R. Xue, L. Wang, and F. Zhou, "Sequence SAR image classification based on bidirectional convolution-recurrent network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9223–9235, Nov. 2019.
- [29] C. He *et al.*, "A component-based multi-layer parallel network for airplane detection in SAR imagery," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1016.
- [30] C. Bentes, D. Velotto, and B. Tings, "Ship classification in TerraSAR-X images with convolutional neural networks," *IEEE J. Ocean. Eng.*, vol. 43, no. 1, pp. 258–266, Jan. 2018.
- [31] X. Ma, K. Ji, L. Zhang, S. Feng, B. Xiong, and G. Kuang, "An open set recognition method for SAR targets based on multitask learning," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: [10.1109/LGRS.2021.3079418](https://doi.org/10.1109/LGRS.2021.3079418).
- [32] E. R. Keydel, S. W. Lee, and J. T. Moore, "MSTAR extended operating conditions: A tutorial," in *Proc. 3rd SPIE Conf. Algorithms SAR Imagery*, 1996, pp. 228–242.
- [33] B. Ding, G. Wen, X. Huang, C. Ma, and X. Yang, "Data augmentation by multilevel reconstruction using attributed scattering center for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 6, pp. 979–983, Jun. 2017.
- [34] M. Ghasemi and A. Sheikhi, "Joint scattering center enumeration and parameter estimation in GTD model," *IEEE Trans. Antennas Propag.*, vol. 68, no. 6, pp. 4786–4798, Jun. 2020.
- [35] H. Chiang, R. L. Moses, and L. C. Potter, "Model-based classification of radar images," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1842–1854, Aug. 2000.
- [36] K. E. Dungan and L. C. Potter, "Classifying transformation-variant attributed point patterns," *Pattern Recognit.*, vol. 43, no. 11, pp. 3805–3816, 2010.
- [37] K. Fu, F. Dou, H. Li, W. Diao, X. Sun, and G. Xu, "Aircraft recognition in SAR images based on scattering structure feature and template matching," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4206–4217, Nov. 2018.
- [38] B. Ding, G. Wen, C. Ma, and X. Yang, "An efficient and robust framework for SAR target recognition by hierarchically fusing global and local features," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5983–5995, Dec. 2018.
- [39] S. Chen, H. Wang, F. Xu, and Y. 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.
- [40] R. Shang, J. Wang, L. Jiao, R. Stolkin, B. Hou, and Y. Li, "SAR targets classification based on deep memory convolution neural networks and transfer parameters," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 8, pp. 2834–2846, Aug. 2018.
- [41] 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 Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1091–1095, Jul. 2017.
- [42] Z. Huang, Z. Pan, and B. Lei, "Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data," *Remote Sens.*, vol. 9, no. 9, 2017, Art. no. 907.
- [43] R. Theagarajan *et al.*, "Integrating deep learning-based data driven and model-based approaches for inverse synthetic aperture radar target recognition," *Opt. Eng.*, vol. 59, no. 5, 2020, Art. no. 051407.
- [44] Z. Cui, C. Tang, Z. Cao, and S. Dang, "SAR unlabeled target recognition based on updating CNN with assistant decision," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1585–1589, Oct. 2018.

- [45] J. Zhang, M. Xing, and Y. Xie, "FEC: A feature fusion framework for SAR target recognition based on electromagnetic scattering features and deep CNN features," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2174–2187, Mar. 2021.
- [46] M. Jing and G. Zhang, "Attributed scattering center extraction with genetic algorithm," *IEEE Trans. Antennas Propag.*, vol. 69, no. 5, pp. 2810–2819, May 2021.
- [47] S. Xia *et al.*, "A fast adaptive k-means with no bounds," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: [10.1109/TPAMI.2020.3008694](https://doi.org/10.1109/TPAMI.2020.3008694).
- [48] F. Eyben, S. Böck, B. Schuller, and A. Graves, "Universal onset detection with bidirectional long-short term memory neural networks," in *Proc. 11th Int. Soc. Music Inf. Retrieval Conf.*, 2010, pp. 589–594.
- [49] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [50] L. Zhang, C. Zhang, S. Quan, H. Xiao, G. Kuang, and L. Liu, "A class imbalance loss for imbalanced object recognition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2778–2792, May 2020.
- [51] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*. [Online]. Available: <https://arxiv.org/abs/1609.04747>
- [52] 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 Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2196–2210, Apr. 2018.
- [53] C. Belloni, A. Balleri, N. Aouf, J.-M. Le Caillec, and T. Merlet, "Explainability of deep SAR ATR through feature analysis," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 1, pp. 659–673, Feb. 2021.



ages.

Linbin Zhang received the B.S. degree in information engineering from the National University of Defense Technology, Changsha, China, in 2018. He is currently working toward the Ph.D. degree in information and communication engineering from the State Key Laboratory of Complex Electromagnetic Environment Effects, National University of Defense Technology, Changsha, China.

His research interests include object detection, image classification, few-shot learning, and machine learning and its applications to remote sensing images.



chine learning.

Xiaojie Ma received the B.S. degree in measurement and control technology and instrument and M.S. degree in measuring and testing technologies and instruments from Yanshan University, Qinhuangdao, China, in 2016 and 2019, respectively. He is currently working toward the Ph.D. degree in information and communication engineering from the State Key Laboratory of Complex Electromagnetic Environment Effects, National University of Defense Technology, Changsha, China.

His research interests include SAR ATR and machine learning.



Sijia Feng received the B.S. degree in information engineering and M.S. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2016 and 2019, respectively. She is currently working toward the Ph.D. degree in information and communication engineering with the State Key Laboratory of Complex Electromagnetic Environment Effects, National University of Defense Technology.

Her research interests include SAR image interpretation, feature extraction, and machine learning.



Gangyao Kuang (Senior Member, IEEE) received the B.S. and M.S. degrees in geophysics from the Central South University of Technology, Changsha, China, in 1988 and 1991, respectively, and the Ph.D. degree in communication and information from the National University of Defense Technology, Changsha, China, in 1995.

He is currently a Professor with the School of Electronic Science, National University of Defense Technology. His research interests include remote sensing, SAR image processing, change detection,

SAR ground moving target indication, and classification with polarimetric SAR images.



Kefeng Ji (Member, IEEE) received the B.S. degree in aerospace engineering from Northwestern Polytechnical University, Xi'an, China, in 1996 and the M.S. and Ph.D. degrees in information and telecommunication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 1999 and 2003, respectively.

In 2003, he joined the College of Electronic Science and Technology, NUDT, where he is currently a Professor. He has authored or coauthored more than 80 papers. His research interests include signal

processing, machine learning, pattern recognition, remote sensing information processing, SAR image interpretation, target detection, recognition, feature extraction, and marine surveillance.

Prof. Ji is a reviewer of several international journals and conferences, such as IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, IEEE ACCESS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, *International Journal of Remote Sensing*, *Remote Sensing Letters*, *European Journal of Remote Sensing*, and *Remote Sensing*.