IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 12, 2019, accepted February 11, 2019, date of publication February 25, 2019, date of current version May 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2900522

Data Augmentation Based on Attributed Scattering Centers to Train Robust CNN for SAR ATR

JUNYA LV[1] AND YUE LIU[2]

[1]School of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450046, China
[2]School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China

Corresponding author: Junya Lv (lvjunya_huel@163.com)

ABSTRACT Driven by the good classification performance of the convolutional neural network (CNN), this study proposes a CNN-based synthetic aperture radar (SAR) target recognition method. In this paper, a novel data augmentation algorithm is proposed via target reconstruction based on attributed scattering centers (ASC). The ASCs reflect the electromagnetic phenomenon of SAR targets, which can be used to reconstruct the target's characteristics. The sparse representation (SR) algorithm is first employed to extract the ASCs from a single SAR image. Afterward, some of the extracted ASCs are selected to reconstruct the target's image. By repeating the process, many new images can be generated as available training samples. In the classification stage, a CNN architecture is designed and trained by the augmented samples. For the test sample, it is also reconstructed using all the extracted ASCs thus relieving the interferences caused by the clutters or noises in the background. Finally, the reconstructed image from the test sample is classified based on the trained CNN. The reconstructed image from the ASCs can reduce the clutters and noises thus enhancing the image quality. More importantly, the generated new training samples could cover more operating conditions, which may probably occur in SAR target recognition. Therefore, the trained CNN can work more robustly in different situations. In the experiments, the moving and stationary target acquisition and recognition (MSTAR) dataset is used to evaluate the performance of the proposed approach. This method could classify the 10 classes of targets with an accuracy of 99.48% under the standard operating condition (SOC). For the extended operating conditions like configuration variance, depression angle variance, noise corruption, and partial occlusion, the proposed method also displays superior performance over some baseline algorithms drawn from state-of-the-art literature.

INDEX TERMS Convolutional neural network (CNN), synthetic aperture radar (SAR), target recognition, attributed scattering center (ASC), data augmentation.

I. INTRODUCTION

Synthetic Aperture Radar (SAR) could work day and night under different weather conditions to produce high-resolution images of the focused area. To help interpretate a large-scene SAR image, the Automatic Target Recognition (ATR) system is designed, which generally consists of three steps: detection, discrimination, and classification [1]. Target detection locates the potential Regions of Interest (ROIs) in the large-scene SAR image thus eliminating the redundancy caused by the background noises or clutters. These ROIs are then processed in the target discrimination stage to reject the false alarms caused by natural clutters such as trees, shrubs, and rivers. In the target classification stage, the remaining ROIs are classified to determine their labels. As a supervised pattern recognition problem, the target classification algorithm has been extensively investigated, which is called the narrow definition of "SAR ATR" [2]. This study also focuses on the target classification stage. Efforts are made to improve the classification performance from two aspects: feature extraction and classifier. Accordingly, the present SAR ATR methods can be summarized as feature-oriented and classifier-oriented ones. Different kinds of features were sought in the previous works including those inherited from the field of optical image processing and those specifically designed for radar images. Typical geometrical features

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

like target outline [3], [4], binary target region [5]–[7], and shadow [8], [9] were used as the basic features in SAR ATR. The projection features could be efficiently extracted by the manifold learning algorithms [10]–[14] like Principal Component Analysis (PCA) [10], Non-negative Matrix Factorization (NMF) [11], etc. Different from the optical images, SAR images reflect the backscattering characteristics of the target, which are sensitive to the aspect angles. Therefore, some related features were also employed for SAR ATR such as Attributed Scattering Centers (ASC) [15], azimuthal sensitivity image [16]. In [17]–[19], several matching strategies were designed for two ASC sets with application to SAR ATR. The classifiers provide decision engines for the extracted features to determine the target labels. Many classifiers have been successfully applied to SAR ATR including the Nearest Neighbor (NN) [10], Adaptive Boosting (AdaBoost) [20], Support Vector Machine (SVM) [21], [22], Sparse Representation-based Classification (SRC) [23], [24], etc.

Recently, the deep learning technique has drawn extensive interests because of its powerful capability of learning discriminative representations from the original data [25]–[27]. As demonstrated in previous works, deep learning methods outperformed the traditional ones significantly in the fields like speech signal processing, face recognition, etc. Motivated by these works, the Convolutional Neural Network (CNN) was also introduced into SAR ATR [28]–[35]. The results demonstrated the superiority of CNN over some traditional methods especially under the Standard Operating Condition (SOC), where the test samples are notably similar with the training ones. However, the application of CNN into SAR ATR actually has some obstacles because the available training samples of the interested targets are quite limited. In addition, there are many Extended Operating Conditions (EOCs) in the real-world environment such as noise contamination, partial occlusion, etc., which can be hardly handled by the network trained by a few training samples. As a remedy, some works sought transfer learning [31], [32] or data augmentation [33]–[35] algorithms to enrich the available training samples. In [31], some SAR scene images were used for transfer learning to address the target classification task in the Moving and Stationary Target Acquisition and Recognition (MSTAR) dataset. Malmgren-Hansen et al. simulated SAR target chips from the target's CAD models via electromagnetic (EM) code, which were then applied to transfer learning for target recognition [32]. Ding *et al.* augmented the SAR training samples by generating the translated, noisy, and pose-synthesized SAR images from the original ones [34]. According to SAR imaging mechanism, Yue generated noisy samples at different Signal-to-Noise Ratios (SNR), multi-resolution representations, and partially occluded images to train CNN [35].

This study develops a novel strategy of data augmentation for SAR images in order to train a more robust CNN for target recognition. In contrast to optical images, SAR images actually record the backscattering field of the target at the high frequency region, e.g., X-band. According to the electromagnetic theories, the high-frequency response of a target can be modeled as the summation of several local phenomenon called scattering centers [15]. In this sense, the target's characteristics can be well maintained and reconstructed by the scattering centers. In addition, when only a part of the scattering centers are used in the reconstruction, the reconstructed image actually embodies the partial properties of the target, which may be caused by EOCs like partial occlusion. First, the Sparse Representation (SR) is used to extract the ASCs from a single SAR image [36], [37]. According to the previous works, SR was capable of estimating the attributes of ASCs with high effectiveness and efficiency compared with the traditional image-domain methods. In addition, because of the optimization during the solution of SR, this ASC extraction algorithm was immune to noise corruption to some extent. Afterwards, some subsets of the extracted ASCs are used to reconstruct the target. So, those reconstructed images describe partial properties of the target, which could supplement the original training samples. Finally, the augmented samples are employed to train the designed CNN architecture in this study for target classification. For the test sample, it is first reconstructed using the ASCs. And the reconstructed image is classified via the trained network. In contrast with the previous data augmentation algorithms, the proposed approach in this study has several outstanding merits. First, ASCs reflect the essential properties of SAR images, i.e., electromagnetic scatterings. Then, target reconstruction based on ASCs can better present the actual deformations of the target caused by different kinds of conditions including SOC or EOCs. Second, the SR-based ASC extraction method can effectively relieve the interferences of the possible noises or clutters in the original SAR images. So, the whole classification scheme performed on the reconstructed images could achieve higher robustness to noise or clutter corruptions. Third, unlike the previous data augmentation methods conducted in the image domain, the combinations of ASCs for target reconstruction provide richer ways of generating the available training samples. Hence, it is promising that the proposed strategy could better enhance the classification capability of a certain CNN architecture.

The rest of this paper is organized as follows. Section 2 introduce the SR-based ASC extraction and ASC-based target reconstruction for data augmentation. In Section 3, the basic theory of CNN is reviewed and the proposed architecture for SAR target recognition is illustrated. Experimental investigations are conducted in Section 4. Finally, Section 5 concludes the whole manuscript.

II. DATA AUGMENTATION BASED ON ASCS

A. SR-BASED ASC EXTRACTION

The ASC model was proposed by Potter et al. to describe the local scattering characteristics of radar targets at high

frequency region. The target characteristics contained in a single SAR image can be modeled as follow:

$$E(f,\phi;\theta)=\sum_{i=1}^{K}E_i(f,\phi;\theta_i) \tag{1}$$

Equation (1) assumes that the whole target's characteristics can be seen as a summation of several individual scattering centers, i.e., ASCs. For each ASC, its scattering field is related to the frequency f and aspect angle ϕ as calculated in equation (2).

$$\begin{aligned}&E_i(f,\phi;q_i)\\&=A_i\cdot(j\frac{f}{f_c})^{\alpha_i}\times\exp(\frac{-j4\pi f}{c}(x_i\times\phi+y_i\sin\phi))\\&\quad\times\mathrm{sinc}(\frac{2pf}{c}L_i\sin(\phi-\overline{\phi_i}))\times\exp(\text{-}2p\,f\;g_i\sin\phi)\end{aligned} \tag{2}$$

where $q=\{q_i\}=[A_i,a_i,x_i,y_i,L_i,\overline{\phi_i},G_i](i=1,2,L,K)$ represents the K ASCs in a single SAR image. Specifically, each parameter in an ASC has concrete physical meanings. Take the ith ASC as an example, A_i records the complex amplitude, which reflects its relative intensity in all the ASCs. (x_i,y_i) mark its physical positions in the image domain; a_i describes the frequency dependence, which has some discrete values, e.g., -1, $-1/2$, 0, 1/2, 1. If the ith ASC is a distributed one, then it has two parameters L_i and $\overline{\phi_i}$ to describe the length and orientation, respectively. Otherwise, if it is a localized ASC, g_i represents its aspect dependence.

For a single SAR image chip containing only one target, the number of the available ASCs is notably less than that of the discretized parameters. In other words, the parameters to be estimated for a single image is sparse from the perspective of the model-parameter domain. In this sense, the ASC extraction problem can be transformed to a SR task [36], [37]. At first, equation (1) is reformulated as a concise matrix multiplication in equation (3).

$$s=D(q)'s+n \tag{3}$$

where s is generated by vectorizing the 2D measurement $E(f,\phi;q)$, which is obtained by transforming the original SAR image to the frequency domain. $D(q)$ represents the redundant dictionary for SR and its each column corresponds to measurement originated from one element in the attribute space. In other words, some attributes of the seven parameters drawn from the discrete parameter domain are input into equation (1) and the resulted scattering data is vectorized as a column in $D(q)$. s is a vector recording the complex amplitudes. And n denotes the noise term, which is experientially modeled as a zero-mean Gaussian process.

Based on aforementioned analysis, the SR-based ASC extraction algorithm is formulated as follow:

$$\hat{s}=\arg\min_s\|s\|_0,\quad \text{s.t. }\left\|s-D(q)'s\right\|_2\,\$e \tag{4}$$

Algorithm 1 SR-Based ASC Extraction

Input: The vectorized measurement s, noise level e, and dictionary $D(q)$.

Initialization: Initialize the ASC set as $\hat{q}=$ Æ, residual $r=s$, and iteration number $t=1$.

1. while $\|r\|_2^2>e$ do

2. Calculate the correlation between the residual and each column in the dictionary: $C(q)=D^H(q)'r$ with $(g)^H$ denoting the conjugate transpose.

3. Obtain the estimated parameters according to the maximum correlation from Step 1 and update the parameter set: $\hat{q}_t=\arg\min_q C(q)$, $\hat{q}=\hat{q}\bigcup\hat{q}_t$.

4. Obtain the estimations of the amplitudes: $\hat{s}=D^{\dagger}(q)'s$, with $(g)^{\dagger}$ denoting the Moore-Penrose pseudo-inverse. $D(\hat{q})$ is the dictionary generated by $\hat{q}$.

5. Update the residual: $r=s-D(\hat{q})'\hat{s}$.

6. $t=t+1$

Output:The estimated ASC set $\hat{q}$.

In equation (4), $e=\|n\|_2$ is an estimate of the noise level, which is adopted as the threshold of the reconstruction error. The l_0-norm optimization problem in equation (4) is nondeterministic polynomial time hard. As an alternative, the greedy algorithms, e.g., Orthogonal Matching Pursuit (OMP), can be used to obtain the approaching solutions. In this study, the SR-based ASC extraction algorithm is utilized with OMP as the solver illustrated in Algorithm 1.

Fig. 1 illustrates the SR-based ASC extraction algorithm with a single MSTAR image using all the extracted ASCs. The ''star'' symbols in Fig. 1(a) mark the locations of the extracted ASCs in the original image, which jointly reconstruct the target as Fig. 1(b). It shows that the reconstructed image well maintain the properties of the target in the original one, such as the geometrical sizes and intensity distribution. In the meantime, the massive noises or clutters in the background around the target can also be effectively reduced so the target characteristics can be clearly observed. As a further verification, some noisy SAR images are simulated to test the robustness of SR-based ASC extraction algorithm as shown in Fig. 2. For the simulated noisy SAR images at the SNRs of 0 dB and -5 dB, their ASC sets can still be extracted with relatively high precision thus reconstructing the original targets well. In comparison with the results in Fig. 1(b), the reconstructed targets in Fig. 2 still keep high precision regarding to the target characteristics and the noises can still be reduced to a large extent.

B. DATA AUGMENTATION

A data augmentation approach is designed based on the ASCs. For a single SAR image from the original training set, its ASCs are extracted using SR-based algorithm to reconstruct the target's images at different levels. At each level, a subset of all the extracted ASCs are employed

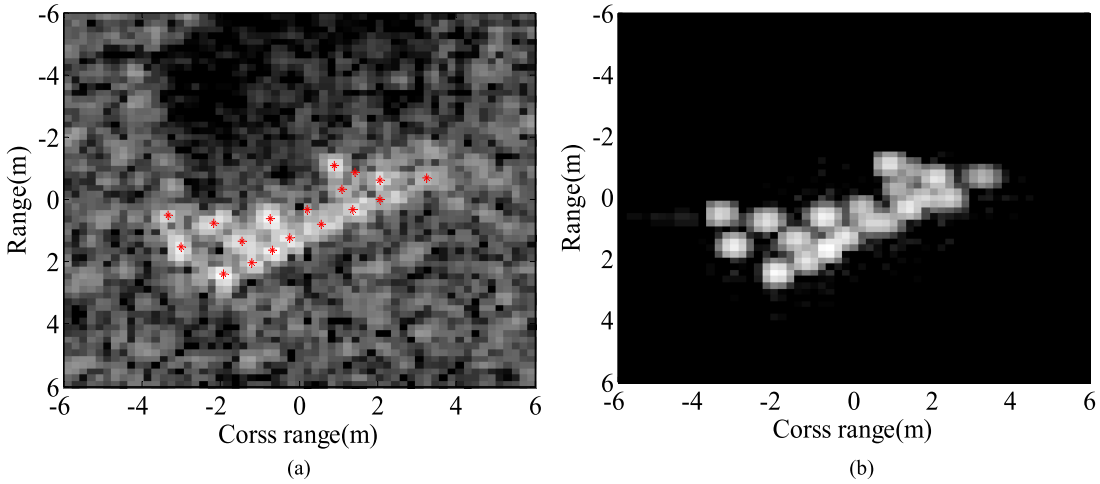

FIGURE 1. Illustration of SR-based ASC extraction. (a) A single MSTAR SAR image marked with the positions of the extracted ASCs; (b) reconstructed image using all the extracted ASCs.

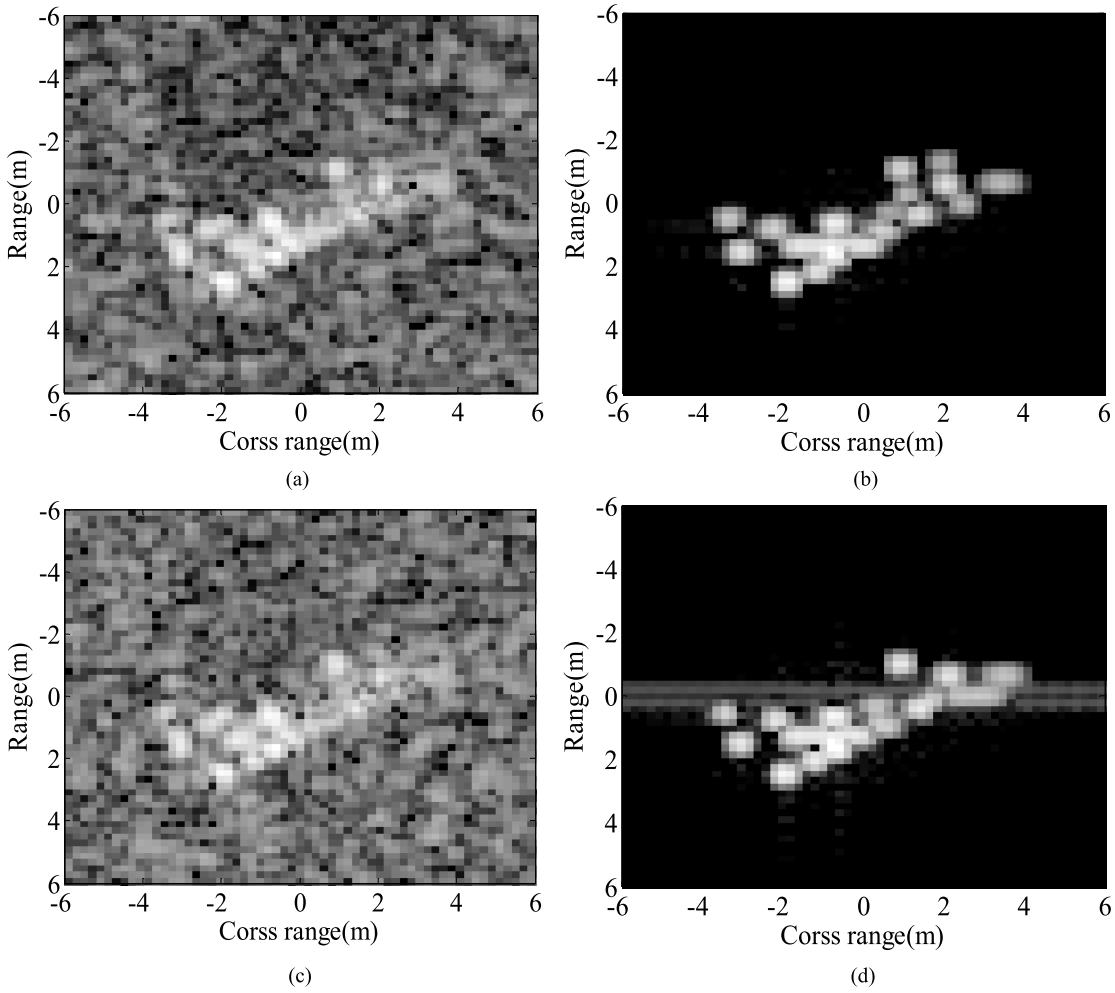

FIGURE 2. SR-based ASC extraction for noisy SAR images. (a) Noisy image at 0 dB; (b) reconstructed image at 0 dB; (c) noisy image at −5dB; (d) reconstructed image at −5 dB.

to reconstruct the target. The parameters of these selected ASCs are inputted into the ASC model in equation (2) and their scattering fields are summed up to obtain the target's image. With more choices of subsets, there will be several

reconstructed images from the original one. By repeating the procedure for each training sample, much more images can be generated.

In fact, there are too many available subsets from the original ASC set. However, some of them are not effective because the reconstructed images are with very low energies compared with the one reconstructed by all the ASCs. Therefore, a threshold is set for the energy of the reconstructed image thus significantly reducing the unnecessary reconstructions from those ineffective subsets. Equation (5) defines the energy factor.

$$f_G = \frac{\sum_{j=1}^{P} |A_{G(j)}|^2}{\sum_{i=1}^{K} |A_i|^2} \tag{5}$$

where the numerator and denominator represent the energies of the selected subset and original ASC set, respectively; the function $G(j)$ finds the index in the original ASC set, which is corresponding to the jth ASC in the selected subset. Hence, f_G is a positive number in the range of [0], [1]. The detailed procedure of constructing one reconstructed image from a single SAR image is illustrated in Algorithm 2. By repeating the process, more reconstructed images can be generated. Specifically, in this study, the energy threshold is set to be $T = 0.5$, assuming that a reconstructed image with lower than half of the original energy make little contributions to the following network training.

Fig. 3 presents some reconstructed images using different subsets of ASCs extracted from the original image shown in Fig. 1 (a). In sequence, the energy factors of these reconstructed images in the subfigures are 0.92, 0.81, 0.75, 0.65, 0.58, and 0.53, respectively. In contrast to the reconstructed image using all the ASCs shown in Fig. 1(b), the results using the subsets reflect the partial characteristics of the targets. Although some ASCs are not considered, the reconstructed images are still available training sources for the following classifier. On one hand, they still reflect the characteristics of the original target so they can help exploit the discriminative features. On the other hand, the EOCs in the practical applications may severely deform the original image, leading into the missing or changing of some ASCs. As a tradeoff between the amount of the training samples and computational load, this study generates 10 reconstructed images from each original training samples, whose energy factors are set to be 1.0 (i.e., using all the extracted ASCs), 0.95, 0.9, 0.85, 0.8, 0.75, 0.7, 0.65, 0.6, and 0.55, respectively. The augmented training samples will be used to train the designed CNN architecture for SAR ATR.

Algorithm 2 Data Augmentation

Input: Extracted ASC set $\hat{q} = [\hat{q}_1\ \hat{q}_2\ L\ \hat{q}_K\ ]$ form a single SAR image; energy threshold T.

1. Randomly select a subset from the original ASC set as $\hat{q}_G = [\hat{q}_{G(1)}\ \hat{q}_{G\,(2)}\ L\ \hat{q}_{G(P)}]$
2. Calculate the energy factor of $\hat{q}_G$ as according to equation (5).
3. If $f_G < T$, repeat from Step 1; else go to Step 4.
4. Calculate the scattering field of each ASC in the selected ASC set and sum up all the scattering field.
5. Conduct the imaging process on the summed scattering field from Step 4 to obtain the reconstruct target.

Output: The reconstructed image I_r.

III. TRAINING CNN FOR TARGET RECOGNITION

A. CNN ARCHITECTURE

In the previous works, CNN has achieved great success in the field of SAR ATR [28]–[35]. A rich set of CNN architectures were developed with very good performance. In this study, a simple CNN architecture is designed as Fig. 4, which comprises of four convolution layers, two pooling layers, and a softmax classifier. Table 1 displays the detailed layout of the network including the sizes of each layer and its outputs. With only four convolution layers, it is actually a simple network, which can be trained with high efficiency. As the main objective of this study, we intend to investigate the effectiveness of data augmentation based on ASCs. Therefore, this simple CNN can effectively reflect the performance of the designed data augmentation algorithm. A brief review of different types of layers are given as follows.

1) CONVOLUTION LAYER

Convolution operation is the core of the whole CNN architecture, which is capable of learning different types of features. In a convolution layer, its inputs (denoted as $O_m^{(l-1)}(m = 1, L, M)$) are connected to all the outputs (denoted as $O_n^{(l)}(n = 1, L, N)$). For convenience, $O_m^{(l-1)}(x, y)$ and $O_n^{(l)}(x, y)$ are used to represent the pixels in the mth input feature map and nth output feature map, respectively. Then, the convolution operation is performed to compute the output feature maps according to the input ones as equation (6).

$$O_n^{(l)}(x, y) = f\Big(\sum_{m=1}^{M} \sum_{p,q=0}^{F-1} k_{nm}^{(l)}(p, q) O_m^{(l-1)} \times (x - p, y - q) + b_n^{(l)}\Big) \tag{6}$$

where $k_{nm}^{(l)}(p, q)$ is the convolution kernel, which connects the input and output feature maps; $f(g)$ represents the nonlinear activation function and $b_n^{(l)}$ denotes the bias.

2) ACTIVATION FUNCTION

In practical applications, the relationship between the inputs and outputs is often nonlinear. To enhance the capability of nonlinear representations, the nonlinear activation functions are implemented in the convolution layers. Some traditional activation functions are the hyperbolic tangent function $f(x) = \tanh(x)$ or sigmoid function $f(x) = 1/(1 + \exp(-x))$ [38]. Recently, the rectified linear unit (ReLU) [39] is validated to be highly effective for training deep

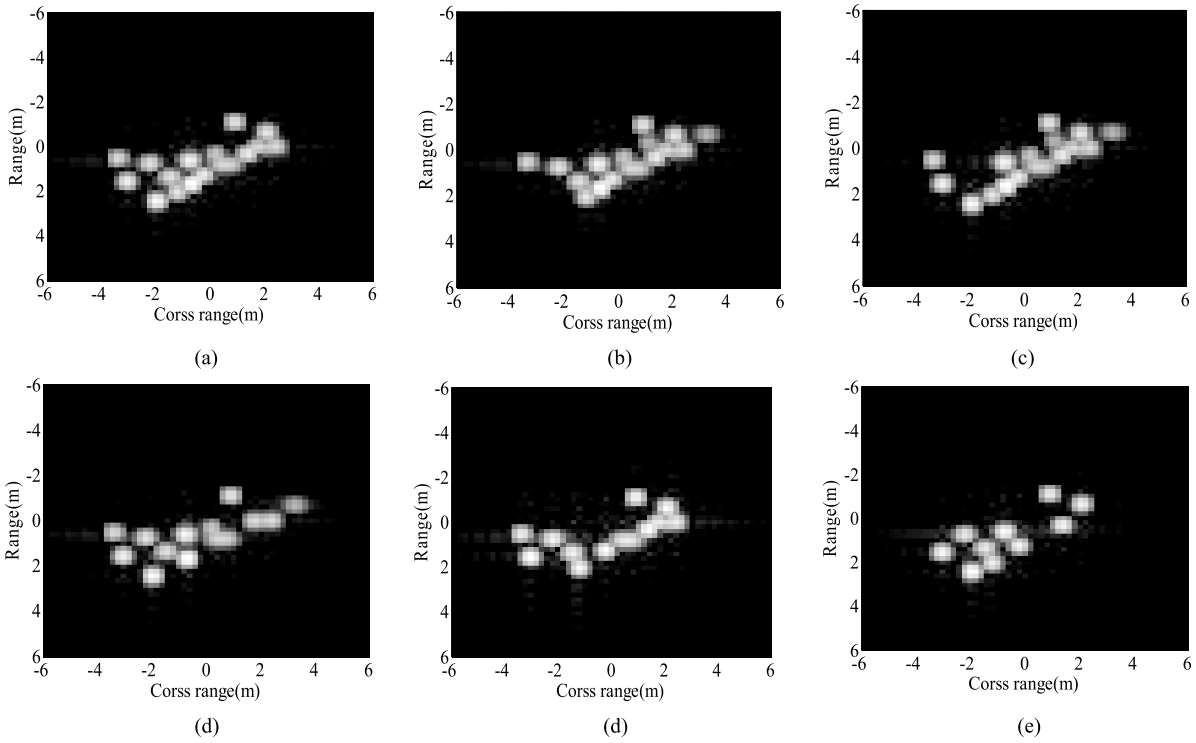

FIGURE 3. SR-based ASC extraction for noisy SAR images. (a) Noisy image at 0 dB; (b) reconstructed image at 0 dB; (c) noisy image at −5dB; (d) reconstructed image at −5 dB.

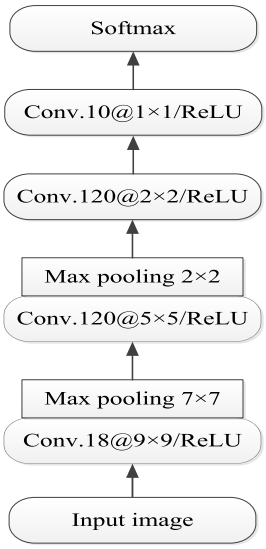

FIGURE 4. The designed CNN architecture in this study.

networks, which could significantly reduce the training time while enhancing the robustness to nonlinear classification problems. The ReLU function has a simple formulation as follow:

$$f(x) = \max(0, x) \tag{7}$$

Owing to its merits, this study uses it as the nonlinear activation functions in our designed CNN architecture.

3) POOLING LAYER

To improve the computational efficiency and robustness of learned features, a pooling layer is often placed after a convolution layer. Although the pooling operating throws away some information in the original feature maps, it can make the trained network robust to shifts or distortions to some extent [40]. There are several pooling operations usually used in current CNN architectures like average pooling and maximum pooling. In this study, we use the max pooling in both pooling layers in Fig. 4, which operates as follow:

$$O_m^{(l+1)}(x, y) = \max_{1<i<h, 1<j<r} (O_m^{(l)}(x + i, y + j)) \tag{8}$$

The max pooling operation in equation (8) works in a window with sizes of $h'w$ and the maximum pixel intensity in the preset window is adopted as the result.

4) SOFTMAX

To enable the network to classify multiple classes, the softmax nonlinearity [41] is often placed at the output layer to produce a K-dimensional weighted vector, whose elements correspond to the posterior probabilities of K different classes, i.e., $p_i = P(y = i|x), i = 1, 2, L, K$. The softmax function is obtained as equation (9).

$$p(i) = \frac{\exp(q_i^T x)}{\mathring{a}_{j=1}^{K} \exp(q_j^T x)} \tag{9}$$

where $q_j^T x$ denotes the jth element in the output vector.

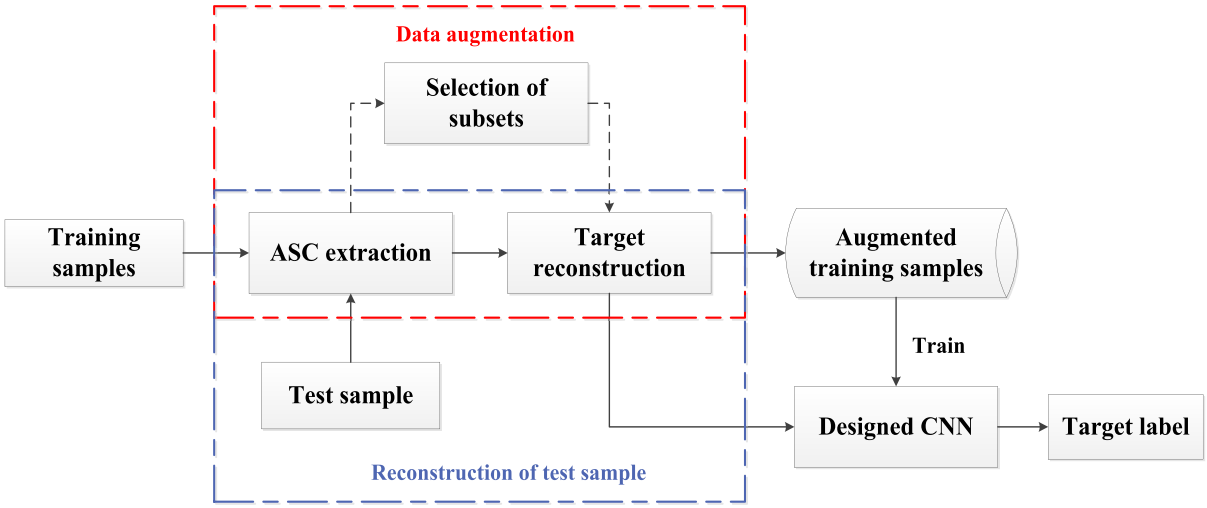

FIGURE 5. General process of target recognition.

When there are N_{cls} labeled training samples, the loss function is defined as:

$$L_{cls}(q_j^T) = -\frac{1}{N_{cls}} \sum_{i=1}^{N_{cls}} \log P(p_i^* | p_i; q_j^T) \tag{10}$$

In equation (10), p_i^* represents the actual label of the ith sample and p_i denotes the predicted one. By adjusting the parameter q_j^T, the loss function is minimized to obtain a better classifier to correctly classify the test sample with a high probability.

To train the designed CNN architecture, the TensorFlow [42] platform is employed, running on a NVIDIA GTX Titan-X graphic processing unit (GPU). And the augmented training set generated in Section 2 are inputted as available samples.

Algorithm 3 Target Recognition

Input: Test sample y; training samples from C classes: $D = [X_1\ X_2\ L\ X_C]$

1. Generate the multiple reconstructed images from each of the training samples using the method in Section 2.2;
2. Construct the augmented training set via collaborating the reconstructed images of all the training samples;
3. Train the designed CNN architecture in Fig. 4 using the augmented training samples;
4. Extract the ASCs from y to reconstruct the target's image;
5. Input the reconstructed image from y to determine its label.

Output: The target label of y.

B. TARGET RECOGNITION

The designed CNN architecture is trained by the augmented training samples generated in Section 2 with application to target recognition. Fig. 5 shows the general procedure of the target recognition method, which could be summarized as Algorithm 3. The original training samples are first reconstructed based on the ASCs to augment the available training images. For the test image, it is first reconstructed using the extracted ASCs and the reconstructed image is input into the network to determine its target label. Compared with the previous CNN-based methods, the proposed one is assumed to have some advantages, which can be analyzed from two aspects. First, the target reconstruction based on ASCs can be seen as a process of image enhancement. As observed in Fig. 1 and Fig. 2, the reconstructed images can effectively relieve the influences caused by the background clutters or noises. In the proposed method, the reconstructed images are used for training and classification so it is promising that the enhanced images tend to achieve better results. Second, much more training samples are available via the data augmentation algorithm used in this study. Different from traditional data augmentation strategies, the essential characteristics of SAR images, i.e., electromagnetic scattering, are adopted as the basis. So, it is predictable that generated images via the proposed approach can better keep the real conditions of the targets under both SOC and EOCs. Therefore, the trained network could obtain better classification ability.

IV. EXPERIMENTS

A. MSTAR DATASET AND BASELINE ALGORITHMS FOR COMPARISON

Since the release of the public version of the MSTAR program, the MSTAR dataset has long been the most popular data source for validating SAR ATR methods. Collected by the X-band sensors, SAR images in the dataset have resolutions of 0.3 m $\times$ 0.3 m. There are ten available targets in the dataset, whose optical images together with exemplar SAR images are shown in Fig. 6. For each target, its SAR images covers the full aspect angles from 0 deg to 360 deg. All the

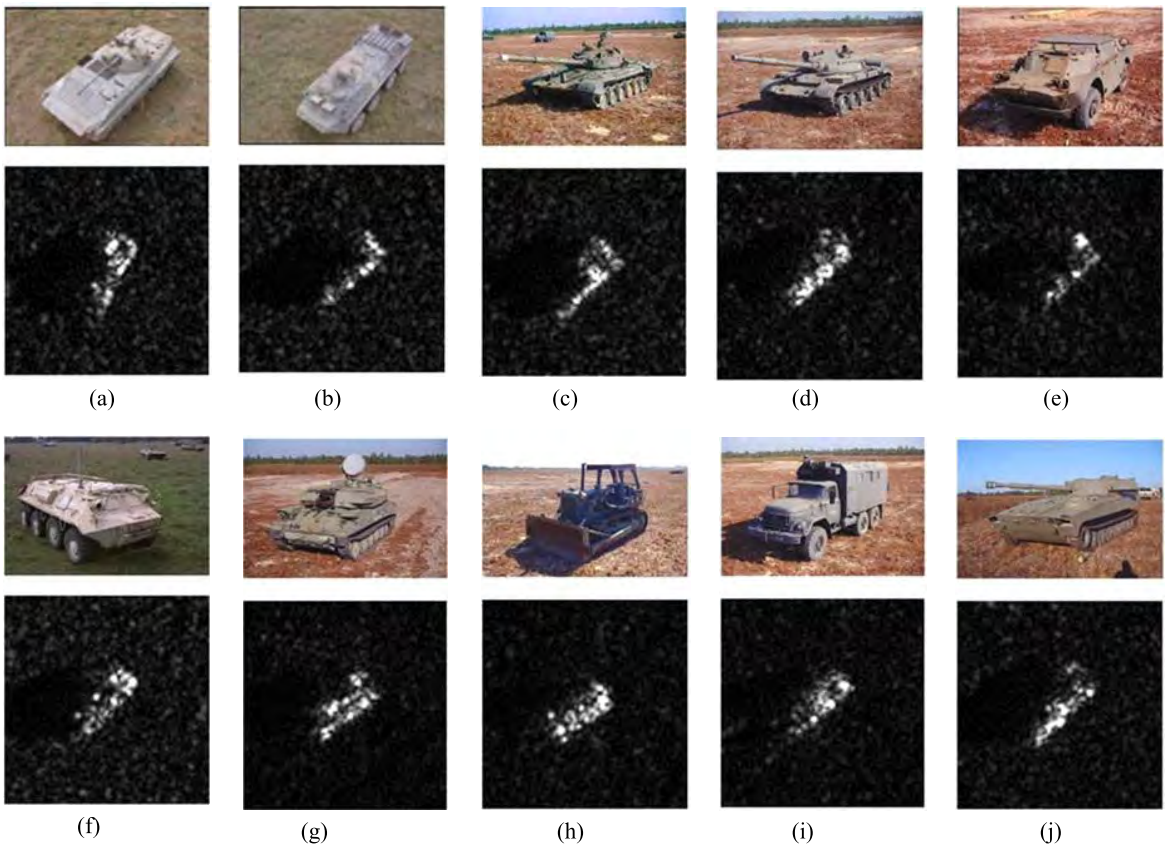

FIGURE 6. Optical and SAR images of the ten targets. (a) BMP2. (b) BTR70. (c) T72. (d) T62. (e) BRDM2. (f) BTR60. (g) ZSU23/4. (h) D7. (i) ZIL131. (j) 2S1.

targets are imaged at the depression angle of 15 deg and 17 deg. And some have extra depression angles like 30 deg and 45 deg. In addition, some targets have several different configurations, which are modified from the standard ones by adjusting some local structures.

Drawn from current literatures, some baseline SAR ATR algorithms are compared in the following experiments. Their detailed descriptions are given as follows.

- SVM: SVM is adopted as the basic classifier as Zhao and Principe [21]. However, rather than using the raw image intensities, PCA is employed to extract feature vectors of 80 dimensions for training and testing.
- SRC: this method uses SRC as the basic classifier as Thiagarajan *et al.* [23]. Random projection is utilized to reduce the original SAR images as 1024-dimension vectors during the classification.
- ASC Matching: ASCs are used as the basic features for SAR target recognition as Ding *et al.* 18]. A one-to-one correspondence is built between two ASC sets from the test sample and template one, respectively. Afterwards, a similarity measure is designed for target recognition.
- A-ConvNets: the all-convolutional networks designed by Chen *et al.* [28], which is widely used as a reference for the CNN-based SAR ATR methods. The raw image intensities are used to train the network as well as for testing.

All these methods are performed on the same platform for fair comparison. In the following parts of this section, detailed experiments are conducted under different operating conditions, which are originally contained in the dataset or simulated. Both the proposed approach and baseline algorithms are tested to fully investigate the performance of the proposed method.

B. EXPERIMENTAL SETTINGS

At first, some typical experimental settings are designed for different operating conditions. As a preliminary validation, SOC is an essential test in almost all the relevant researches. Table 2 presents the training and test sets of the ten targets under SOC, which are from 17 deg and 15 deg depression angles, respectively. For each target, its training and test samples are from the same configurations. Especially, for BMP2 and T72, which containing several different configurations [41], only the standard ones are used (i.e., Sn_9563 and Sn_132, respectively). The training and test samples for the first EOC (EOC-1) is set as Table 3, including four different targets, i.e., BMP2, BDRM2, BTR70, and T72. Among them, there are some configuration differences between the training and test samples of BMP2 and T72. Fig. 7 shows

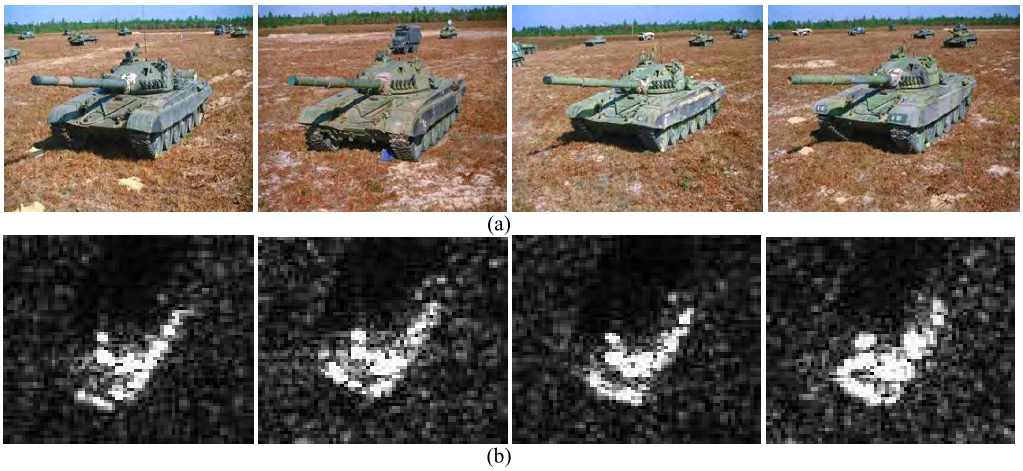

FIGURE 7. Optical and SAR images of four different configurations from T72. (a) Optical images (b) SAR images.

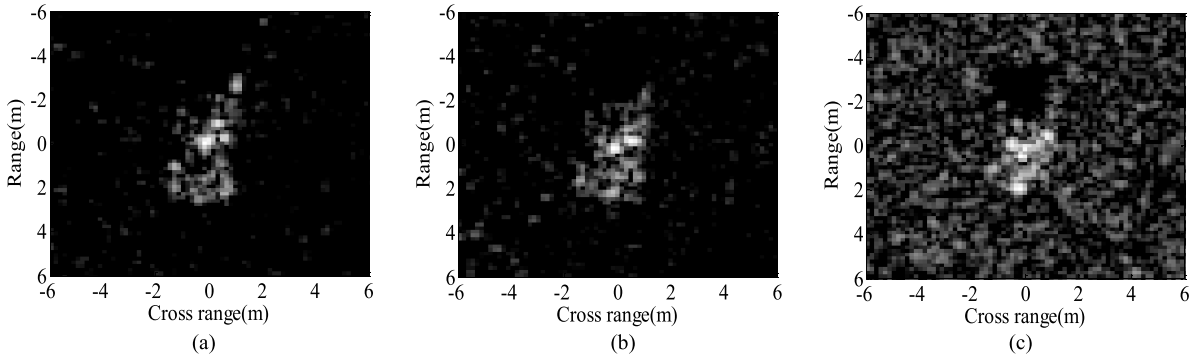

FIGURE 8. SAR images of 2S1 at different depression angles. (a) 17 deg; (b) 30 deg; (c) 45 deg.

the optical and SAR images of four different configurations from T72. It can be observed from the optical images that different configurations have some structural variations, leading into the local variations in the corresponding SAR images. The standard configurations of BMP2 and T72 at 17 deg depression angle are used for training but the test samples are from different configurations collected at both 15 deg and 17 deg depression angles. BDRM2 and BTR70 are only included in the training set as the confuser targets, which further aggravates the difficulty of correct target recognition. The second EOC (EOC-2) investigates the influences of depression angle variance using three different targets [42], i.e., 2S1, BRDM2, and ZSU23/4 as displayed in Table 4. Their training samples are from 17 deg depression angle while the test ones are from 30 deg and 45 deg, respectively. Fig. 8 compares three 2S1 SAR images from different depression angles. As shown, notable differences can be seen between two images with large depression angle differences.

Because of the limited operating conditions in the original MSTAR dataset, this study further simulates two extra EOCs for experimental evaluation. The third EOC (EOC-3) is simulated by adding Gaussian noises [43] into the test samples in Table 2. Different levels of noises are added to generate noisy samples at different SNRs. Fig. 9 shows some exemplar noisy images at different SNRs. Afterwards, the noisy samples are classified based on the original training samples in Table 1. The situation of partial occlusion is simulated as the fourth EOC (EOC-4). Some portions of the target region are removed from each test sample according to the occlusion model in [44] as shown in Fig. 10. Here, the original image to be processed is the same as Fig. 9(a) and the occlusion occurs from different directions. Then, the occluded images are classified for performance evaluation.

TABLE 1. Layout of the designed CNN architecture.

Layer type	Image size	Feature maps	Kernel size
Input	64*64	1	-
Convolution	56*56	18	9*9
Pooling	8*8	18	7*7
Convolution	4*4	120	5*5
Pooling	2*2	120	2*2
Convolution	1*1	120	2*2
Convolution	1*1	10	1*1
Output	1*1	10	-

TABLE 2. Experimental setting for SOC.

	Depr.	BMP2	BTR70	T72	T62	BDRM2	BTR60	ZSU23/4	D7	ZIL131	2S1
Training	17 deg	233 (Sn_9563)	233	232 (Sn_132)	299	298	256	299	299	299	299
Test	15 deg	195 (Sn_9563)	196	196 (Sn_132)	273	274	195	274	274	274	274

TABLE 3. Experimental setting for EOC-1.

	Depr.	BMP2	BDRM2	BTR70	T72
Training	17 deg	233 (Sn_9563)	298	233	232 (Sn_132)
Test	15 deg, 17 deg	428 (Sn_9566) 429 (Sn_c21)	0	0	426 (Sn_812) 573 (Sn_A04) 573 (Sn_A05) 573 (Sn_A07) 567 (Sn_A10)

TABLE 4. Experimental setting forEOC-2.

	Depr.	2S1	BDRM2	ZSU23/4
Training	17 deg	299	298	299
Test	30 deg	288	287	288
	45 deg	303	303	303

C. RESULTS AND ANALYSIS

1) SOC

The experiment is first undertaken under SOC based on the training and test samples in Table 2. The detailed recognition results of the proposed approach are displayed as the confusion matrix in Fig. 11, which is a widely used way to illustrate the classification accuracy. The x and y labels in it mark the actual and predicted labels, respectively. And the elements on the diagonal denote the recognition rates of different classes. As observed, each target can be correctly classified with a recognition rate over 98.5%. Accordingly, the overall recognition rate of all the ten targets is 99.48%. The notably high recognition rate of the proposed method validates its excellent performance under SOC. Table 5 further compares the recognition rates of different methods under SOC. Through quantitative tests, the overall recognition rates of SVM, SRC, ASC Matching, and A-ConvNets are calculated to be 96.54%, 96.38%, 97.12%, and 99.08%, respectively, which is lower than the one achieved by the proposed method. Compared with SVM, SRC, and ASC Matching, the CNN-based methods including the proposal and A-ConvNets achieve much better performance owing to the superior classification ability of deep learning techniques. With a higher recognition rate than A-ConvNets, the proposed approach enhances the classification performance of the designed CNN architecture because of the augmented training samples. In comparison with ASC Matching method, this study makes full use of the ASCs to reconstruct different kinds of conditions, which may occur on the target. The final results demonstrate that the proposed method is much superior over the traditional ways of using ASCs for target recognition, e.g., one-to-one ASC matching.

2) EOC-1

The recognition problems under EOCs begin from EOC-1 according to the experimental setting in Table 3. Table 6 exhibits the classification results of different configurations in the test samples of BMP2 and T72. Although some of them are misclassified to be BRDM2 or BTR70, the recognition rate of each configuration still reaches over 97%. And the overall recognition rate is averaged to be 98.80%. The performance of different methods is compared in Table 7, which reflects the superiority of the proposed approach under configuration variances. With regard to the different configurations from the same class, they have some local variations caused by structural modifications, which can be related to the ASCs to a large extent. Hence, via the target reconstruction based on ASCs, some generated images actually describe the target's characteristics with configuration variance. So, the designed CNN trained by the augmented training samples could better handle this situation. Compared with A-ConvNets, the higher recognition rate of the proposed method under EOC-1 also benefits from the augmentation of training samples.

3) EOC-2

Based on the experimental setting in Table 4, the proposed approach is tested under depression angle variances. The classification results of the three targets at 30 deg and 45 deg depression angles are shown in Table 8. The overall recognition rates at the two depression angles are 98.61% and 74.48%, respectively. The recognition performance degrades severely at 45 deg depression angle because the large depression angle difference significantly changes the appearances of the target in contrast to the reference samples at 17 deg as shown in Fig. 9. As a result, the trained CNN lose some precision when classifying the test samples at 45 deg depression angle. Fig. 12 lists the overall recognition rates of the proposed method and baseline algorithms at the two depression angles. The performance of the baseline algorithms shares similar trend with the proposal at the two depression angles. In comparison, the proposed method defeats the baseline algorithms at both depression angles, which validates its better robustness to depression angle variance. With regard to images from different depression angles, their differences can also be explained by the ASCs.

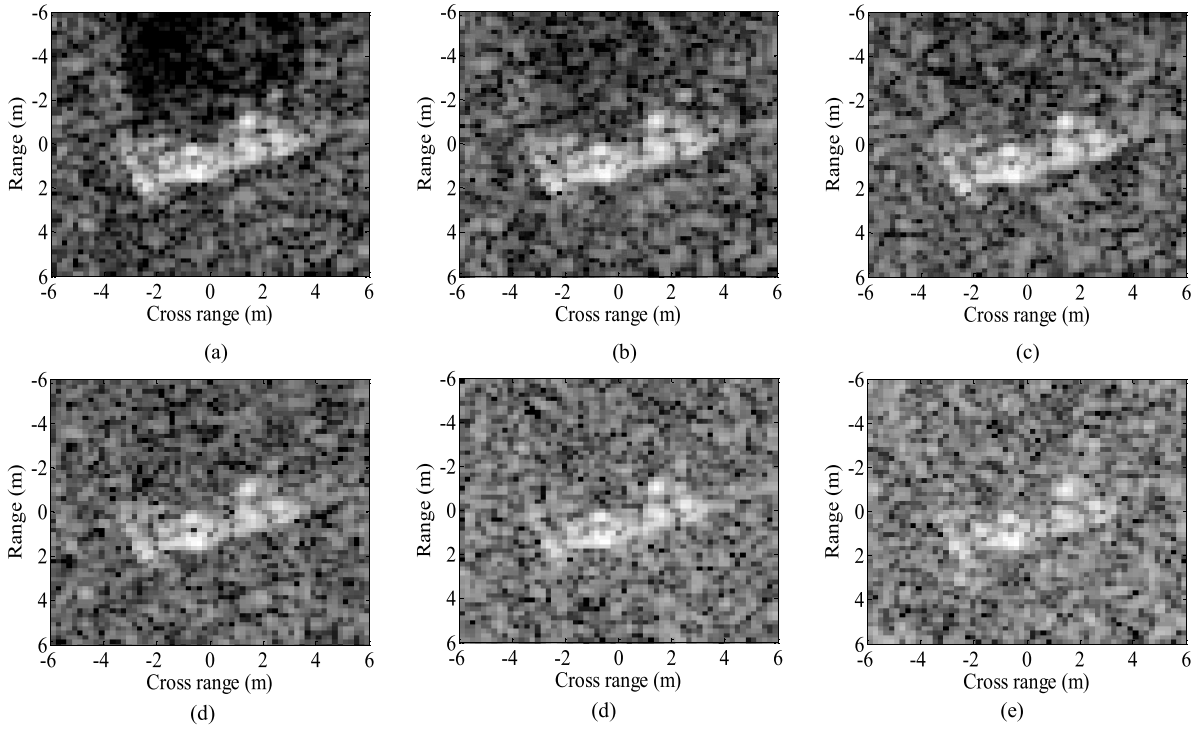

FIGURE 9. Simulated noisy images at different SNR. (a) original image (b)10dB (c) 5dB (d) 0dB (e) −5dB (f) −10dB.

TABLE 5. Overall recognition rates of different methods under SOC.

Method	Proposed	SVM	SRC	ASC Matching	A-ConvNets
Overall recognition rate (%)	99.48	96.54	96.38	97.12	99.08

TABLE 6. Recognition results under EOC-1.

Class	Serial No.	BMP2	BRDM2	BTR70	T72	Recognition rate (%)
BMP2	Sn_9566	416	2	5	5	97.20
	Sn_c21	424	1	3	1	98.83
T72	Sn_812	7	1	3	415	97.42
	Sn_A04	2	5	0	566	98.78
	Sn_A05	0	1	1	571	99.65
	Sn_A07	1	0	1	571	99.65
	Sn_A10	0	0	0	567	100
Overall						98.80

TABLE 7. Overall recognition rates of different methods under EOC-1.

Method	Proposed	SVM	SRC	ASC Matching	A-ConvNets
Overall recognition rate (%)	98.80	95.46	95.08	96.89	96.28

So, the reconstructed images by some subsets of the extracted ASCs can effectively reveal the influences caused by depression angle variances. Therefore, the augmented training samples can enhance the classification ability under EOC-2.

4) EOC-3

To investigate the noise robustness of the proposed approach, the simulated noisy images are classified. The overall recognition rates of the proposed method at different SNRs are plotted in Fig. 13 together with those of the baseline algorithms. The proposed method achieves the highest recognition rates at different noise levels, verifying its superior robustness under EOC-3. It is also evident that the ASC Matching method performs much better than other baseline algorithms especially at lower SNRs than 0 dB. As discussed in Section 2, the ASCs extracted by SR-based method could keep relatively robust under noise corruption. Therefore, the ASC matching can still be smoothly conducted although under noise corruption. In the proposed approach, both the training and test samples are

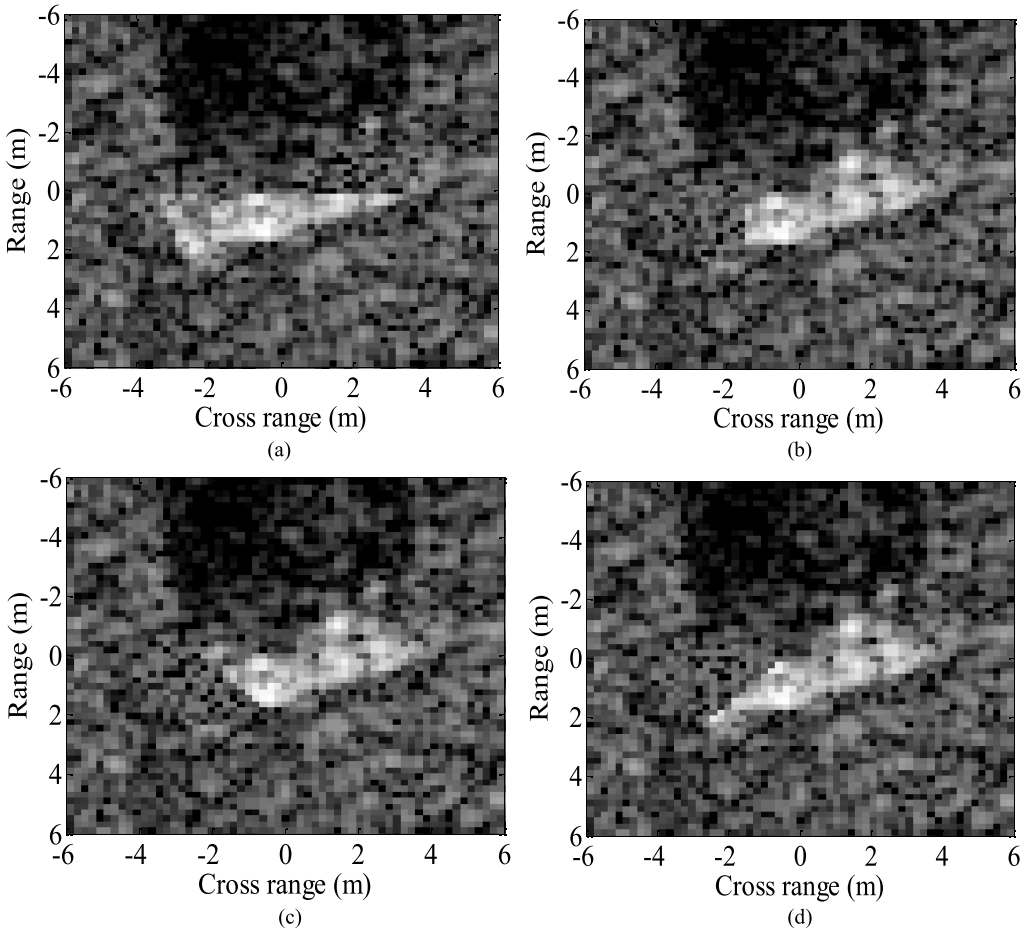

FIGURE 10. Simulated occluded images at 30% occlusion level from different directions. (a) direction 1; (b) direction 2; (c) direction 3; (d) direction 4.

TABLE 8. Recognition results under EOC-2.

Depr.	Actual	Predicted			Recognition rate (%)	Overall (%)
		2S1	BDRM2	ZSU23/4		
30 deg	2S1	283	1	4	98.26	98.61
	BDRM2	1	283	3	98.61	
	ZSU23/4	2	1	285	98.96	
45 deg	2S1	238	41	24	78.55	74.48
	BDRM2	37	226	40	74.59	
	ZSU23/4	53	37	213	70.30	

reconstructed using the extracted ASCs so the influences of noise corruption can be greatly relieved. In addition, the augmented training samples could also improve the classification ability as for the noisy samples. With regard to SVM, SRC, and A-ConvNets, the noisy test samples at low SNRs have remarkable differences with the training ones. As a result, their performance degrades notably under noise corruption.

5) EOC-4

Based on the occluded SAR images, the proposed method and baseline algorithms are tested under different levels of partial occlusions, whose results are simultaneously plotted in Fig. 14. The overall recognition rate of the proposed method tops at each occlusion level. So, the proposed method keeps the most robust under EOC-4 compared with the baseline algorithms. Similar to EOC-3, the ASC Matching method generally outperforms other baseline algorithms under EOC-4. ASCs are local descriptors so they can be used to exploit the stable parts, which are not occluded in the test samples. Therefore, the similarity measure defined based on the one-to-one correspondence between two ASC sets could keep its robustness under partial occlusion. In the proposed method, the ASC-based target reconstruction

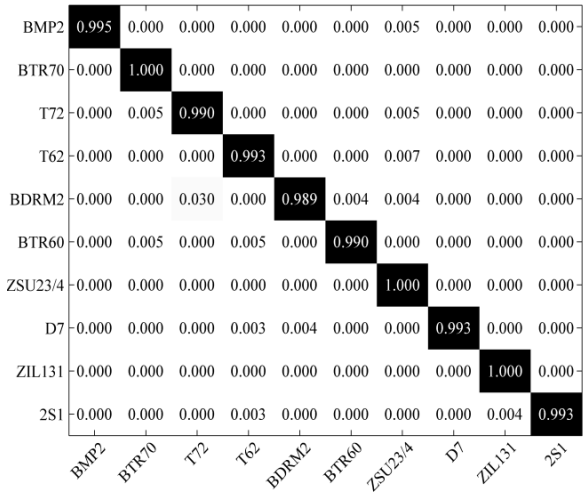

FIGURE 11. Confusion matrix of the proposed method under SOC.

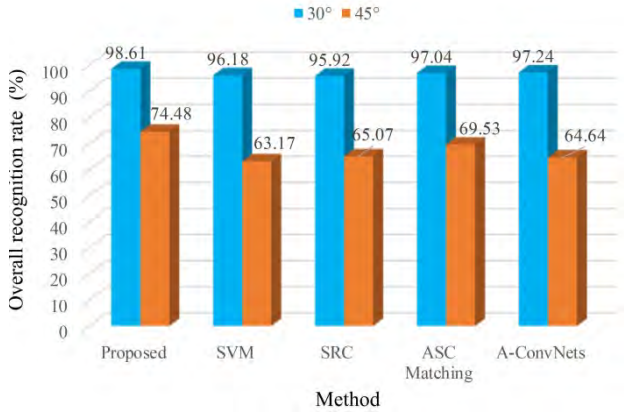

FIGURE 12. Comparison with baseline algorithms under EOC-2.

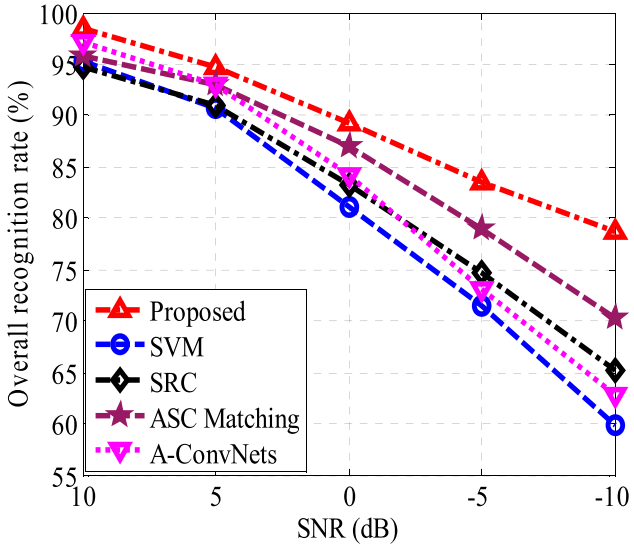

FIGURE 13. Comparison with baseline algorithms under EOC-3.

effectively generates SAR images with some missing ASCs. To some extent, they simulate the partially occluded samples.

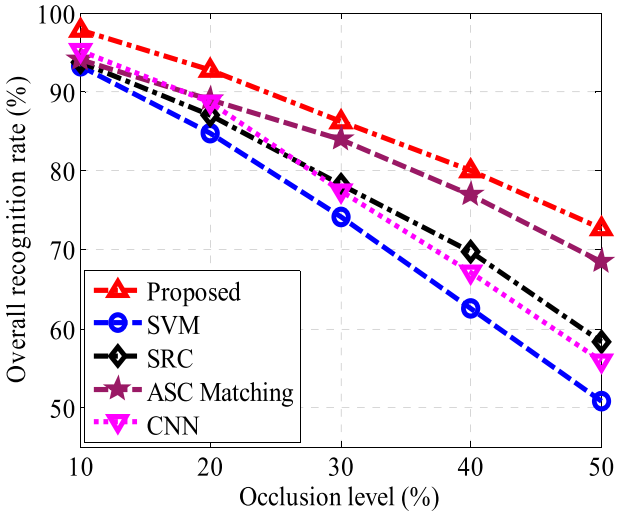

FIGURE 14. Comparison with baseline algorithms under EOC-4.

Therefore, the CNN architecture trained by the augmented training samples could better cope with EOC-4.

V. CONCLUSION

A data augmentation method is proposed in this study through target reconstruction based on ASCs. Then, the augmented training samples are utilized to train a CNN for SAR target recognition. The ASC-based target reconstruction is capable of reducing the clutters and noises from the background in SAR images. So, the reconstructed image could better covey the pure target's characteristics. In this sense, the image quality is enhanced. By reconstructing the target's image using a subset of the extracted ASCs, the generated image represents the partial properties of the target. The repetitions of different subsets produce a large volume of available SAR images for training, which cover more operating conditions. Therefore, the trained CNN based on the augmented training samples could better handle both SOC and various EOCs occurred in practice. Extensive experiments are undertaken on the MSTAR dataset to draw several main conclusions as follows. First, the proposed method inherits the high effectiveness under SOC with an overall recognition rate of 99.48%. Compared with the baseline algorithms, the proposed method achieves the best performance under SOC. Second, for different recognition tasks under EOCs, the proposed approach performs better than the baseline algorithms. On one hand, the image enhancement via ASC-based target reconstruction improves the robustness of the proposed method to the nuisance conditions like noise corruption. On the other hand, the augmented training set covers more operating conditions in SAR ATR such as configuration variance and partial occlusion. Therefore, the trained CNN by the augmented samples could better handle these nuisance conditions.

In fact, this paper provides an open framework for training CNNs using the augmented training set via ASC-based

target reconstruction. So, more suitable networks can also be applied to the augmented training samples to further exploit their potentials in SAR ATR in the future.

REFERENCES

[1] L. M. Novak, G. J. Owirka, and W. S Brower, "Performance of 10- and 20-target MSE classifiers," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1279–1289, Oct. 2000.

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

[3] G. C. Anagnostopulos, "SVM-based target recognition from synthetic aperture radar images using target region outline descriptors," *Nonlinear Anal., Theory, Methods Appl.*, vol. 71, no. 12, pp. e2934–e2939, Dec. 2009.

[4] 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.

[5] M. Amoon and G.-A. Rezai-Rad, "Automatic target recognition of synthetic aperture radar (SAR) images based on optimal selection of Zernike moments features," *IET Comput. Vis*, vol. 8, no. 2, pp. 77–85, Apr. 2014.

[6] C. Clemente, L. Pallotta, D. Gaglione, A. De Maio, and J. J. Soraghan, "Automatic target recognition of military vehicles with Krawtchouk moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 1, pp. 493–500, Feb. 2017.

[7] B. Y. Ding, G. J. Wen, C. H. Ma, and X. L. Yang, "Target recognition in synthetic aperture radar images using binary morphological operations," *Proc. SPIE*, vol. 10, no. 4, Oct. 2016, Art. no. 046006.

[8] S. Papson and R. M. Narayanan, "Classification via the shadow region in SAR imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 2, pp. 969–980, Apr. 2012.

[9] J. J. Cui, J. Gudnason, and M. Brookes, "Automatic recognition of MSTAR targets using radar shadow and superresolution features," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Philadelphia, PA, USA, Mar. 2005, pp. 18–23.

[10] A. K. Mishra, "Validation of PCA and LDA for SAR ATR," in *Proc. IEEE Region 10 Conf.*, Hyderabad, India, Nov. 2008, pp. 1–6.

[11] Z. Cui, Z. Cao, J. Yang, J. Feng, and H. Ren, "Target recognition in synthetic aperture radar images via non-negative matrix factorisation," *IET Radar, Sonar Navigat.*, vol. 9, no. 9, pp. 1376–1385, Dec. 2015.

[12] Y. Huang, J. Pei, J. Yang, B. Wang, and X. Liu, "Neighborhood geometric center scaling embedding for SAR ATR," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 180–192, Jan. 2014.

[13] M. Yu, G. Dong, H. Fan, and G. Kuang, "SAR target recognition via local sparse representation of multi-manifold regularized low-rank approximation," *Remote Sens.*, vol. 10, no. 2, p. 211, Feb. 2018.

[14] X. Liu, Y. Huang, P. Pei, and J. Yang, "Sample discriminant analysis for SAR ATR," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2120–2124, Dec. 2014.

[15] L. C. Potter, and R. L. Mose, "Attributed scattering centers for SAR ATR," *IEEE Trans. Image Process.*, vol. 6, no. 1, pp. 79–91, Jan. 1997.

[16] B. Ding, G. Wen, X. Huang, C. Ma, and X. Yang, "Target recognition in SAR images by exploiting the azimuth sensitivity," *Remote Sens. Lett.*, vol. 8, no. 9, pp. 821–830, Sep. 2017.

[17] H.-C. 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.

[18] B. Ding, G. Wen, X. Huang, C. Ma, and X. Yang, "Target recognition in synthetic aperture radar images via matching of attributed scattering centers," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3334–3347, Jul. 2017.

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

[20] Y. Sun, Z. Liu, S. Todorovic, and J. Li, "Adaptive boosting for SAR automatic target recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 1, pp. 112–125, Jan. 2007.

[21] Q. Zhao and J. C. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 2, pp. 643–654, Apr. 2001.

[22] H. Liu and S. Li, "Decision fusion of sparse representation and support vector machine for SAR image target recognition." *Neurocomputing*, vol. 113, pp. 97–104, Aug. 2013.

[23] J. Thiagarajan, K. Ramamurthy, P. P. Knee, A. Spanias, and V. Berisha, "Sparse representations for automatic target classification in SAR images," in *Proc. Int. Symp. Commun., Control Signal Process. (ISCCSP)*, Limassol, Cyprus, Mar. 2010, pp. 1–4.

[24] H. Song, K. Ji, Y. Zhang, X. Xing, and H. Zou, "Sparse representation-based SAR image target classification on the 10-class MSTAR data set," *Appl. Sci.*, vol. 6, no. 1, p. 26, 2016.

[25] A. Krizhevsky, I. Sutskever, and G. E Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Lake Tahoe, NV, USA, vol. 2, Dec. 2012, pp. 1096–1105.

[26] C. Szegedu *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.

[27] 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.

[28] 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.

[29] S. A. Wagner, "SAR ATR by a combination of convolutional neural network and support vector machines," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 6, pp. 2861–2872, Dec. 2016.

[30] K Du, Y. Deng, R. Wang, T. Zhao, and N. Li, "SAR ATR based on displacement- and rotation- insensitive CNN," *Remote Sens. Lett.*, vol. 7, no. 9, pp. 895–904, Jun. 2016.

[31] 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, p. 907, Aug. 2017.

[32] D. Malmgren-Hansen, A. Kusk, J. Dall, A. A. Nielsen, R. Engholm, and H. Skriver, "Improving SAR automatic target recognition models with transfer learning form simulated data," *IEEE Geosci. Remote Sens Lett.*, vol. 14, no. 9, pp. 1484–1488, Sep. 2017.

[33] H. Furukawa, "Deep learning for target classification from SAR imagery: Data augmentation and translation invariance," IEICE, Tokyo, Japan, Tech. Rep. 182, 2017, vol. 117, pp. 13–17.

[34] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 364–368, Mar. 2016.

[35] Y. Yan, "Convolutional neural networks based on augmented training samples for synthetic aperture radar target recognition," *Proc. SPIE*, vol. 27, no. 2, Apr. 2018, Art. no. 023024.

[36] H. Liu, B. Jiu, F. Li, and Y. Wang, "Attributed scattering center extraction algorithm based on sparse representation with dictionary refinement," *IEEE Trans. Antennas Propagat.*. vol. 65, no. 5, pp. 2604–2614, May 2017.

[37] Y. Cong, B. Chen, H. Liu, and B. Jiu, "Nonparametric Bayesian attributed scattering center extraction for synthetic aperture radar targets," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4723–4736, Sep. 2016.

[38] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control Signals Syst.*, vol. 2, no. 4, pp. 303–314, Dec. 1989.

[39] K. Hara, D. Saito, and H. Shouno, "Analysis of function of rectified linear unit used in deep learning," in *Proc. Int. Joint Conf, Neural Netw.*, Killarney, Ireland, Jul. 2015, pp. 12–17

[40] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, Baltimore, MD, USA, Jun. 2014, pp. 49–54.

[41] B. Ding and G. Wen, "Exploiting multi-view SAR images for robust target recognition," *Remote Sens.*, vol. 9, no. 11, p. 1150, Nov. 2017.

[42] B. Ravichandran, A. Gandhe, R. Simith, and R. Mehra, "Robust automatic target recognition using learning classifier systems," *Inf. Fusion*, vol. 8, no. 3, pp. 252–265, Jul. 2007.

[43] S. H. Doo, G. Smith, and C. Baker, "Reliable target feature extraction and classification using potential target information," in *Proc. IEEE Radar Conf.*, Arlington, VA, USA, May 2015, pp. 628–633.

[44] G. Jones and B. Bhanu, "Recognition of articulated and occluded objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 7, pp. 603–613, Jul. 1999.

JUNYA LV received the M.Sc. degree from Beijing Jiaotong University, in 2008. He is currently a Lecturer with the College of Computer and Information Engineering, Henan University of Economics and Law. His current research interests include image processing, database theory, and technologies.

YUE LIU received the M.Sc. degree from Zhengzhou University, in 2002, where he is currently a Lecturer with the School of Information Engineering. His current research interests include image processing and remote sensing applications.