Research Article

Synthetic Aperture Radar Target Recognition Based on Multimodule Image Enhancement Network

Xuan Wang,1,2 Yuliang Lu,1,2 Xuehu Yan,1,2 Da Li,1 and Chunqian He1

1National University of Defense Technology, Hefei 230037, China
2Anhui Province Key Laboratory of Cyberspace Security Situation Awareness and Evaluation, Hefei 230037, China

Correspondence should be addressed to Yuliang Lu; publicluyl@126.com and Xuehu Yan; yanxh17@nudt.edu.cn

Received 15 June 2022; Revised 6 August 2022; Accepted 8 August 2022; Published 21 November 2022

1.Introduction

Synthetic aperture radar (SAR) is an active microwave sensor. It is not affected by a series of external environmental factors such as light and can continuously monitor local ground object scenes [1]. In addition, SAR has a strong penetration ability and high-resolution imaging characteristics so that it can accurately detect objects which are actively camouflaged or blocked by some obstacles. Based on these unique advantages, SAR has been successfully used in many fields, such as geological exploration, environmental monitoring, battlefield reconnaissance, precision strike, and so on [2].

Automatic target recognition (ATR) is dedicated to automatically detecting the region of interest and reasoning the attributes of target categories from complex images. It is an important way to realize intelligent SAR image interpretation. In recent years, with the rapid development of machine learning theory and deep learning technology, SAR ATR technology has made a breakthrough in theoretical research and system design. Target extraction refers to separating the object of interest from the background in a single image or image sequence. It is the operation of recognizing and interpreting meaningful object entities from images and extracting different image features. Nowadays, target extraction has a wide range of applications, such as extracting facial features and fingerprints in computer vision and extracting feature points and lines for image matching and 3D modeling in photogrammetry and remote sensing. However, target extraction from remote sensing images such as SAR images is also called target region extraction in some professional articles. In this paper, the target region refers to ground military vehicles, aircraft, and other military equipment. Since the existing publicly available datasets only contain single target images, and most studies only considered a single target in images, this paper extracts target features from single target images.

SAR images differ from ordinary images in several ways: (1) target size does not change with the distance between...
sensor and target; (2) scene information is determined by the amplitude of radar reflection and scattering; and (3) the image is highly sensitive to changes in target shape and attitude [3]. These characteristics make SAR a reliable instrument for collecting information, giving it important applications in both military and civilian fields [4–6].

In the high-tech local wars such as the Iraq war and the Afghanistan war that took place at the end of the last century, as an important means of information acquisition, SAR has played a great role in battlefield attitude sensing, air defense, and antimissile and precision strike. Obviously, in the new situation of information warfare, SAR ATR provides strong support for accelerating the military information construction and is of great significance in enhancing the military strength of various countries. Meanwhile, some research about SAR ART can also be applied to digital image forensics and other related aspects in real scenes. It can mainly detect and extract suspicious objects from multi-target and large-area digital images in real scenes. The forensics process is mainly divided into the following three steps: firstly, image segmentation is performed on the real scene image, then the quality of the segmented image is enhanced, and finally lock and extract suspicious target areas, so as to achieve the purpose of image forensics.

Many studies focus on SAR ATR [7–9], which can be divided mainly into feature-based methods and model-based methods [3]. Model-based methods focus on model design, but it is difficult to build a general target model for the multiclassification task. This limits the recognition rate of SAR images. Feature-based methods focus on the extraction of target features, and highly descriptive features can generate higher recognition rates. Two directions are generally concerned to improve the ability of the features, one of which is exploring the feature extraction methods [7], including feature fusion [8], and the other is enhancing the input image [9].

In various fields, the application of the deep learning model is more and more extensive, and more and more people pay attention to the research of the model [10, 11]. Among them, a convolutional neural network is a classical representative model of the deep learning model, which is often used for image recognition [12]. Compared with traditional methods, a CNN uses different kernels to convolve images to obtain feature maps, and the feature maps go through stacked layers to obtain deep features [13]. Most of the existing convolutional neural network research, about image feature extraction, does not need to design manual features in advance [14]. It is possible to design a handcrafted feature method better than deep learning-based approaches in the future. However, this paper mainly focuses on the feature extraction scheme of adaptive convolutional neural networks. This simplifies feature extraction without manual intervention and adjustment. Owing to the superiority of deep neural networks, several important studies on the design of the deep learning architecture in SAR ATR have obtained excellent recognition rates [15–17]. Many studies have shown that a convolutional neural network has unique advantages for image recognition. Therefore, this study selects CNNs as a classifier.

The key factors in improving the recognition performance of SAR ATR algorithms that are based on CNN include (1) SAR image preprocessing to extract features more effectively and easily; (2) designing effective network structures that make full use of the extracted features from SAR images.

Chen et al. designed A-CNN, where the number of unknown parameters is greatly reduced by removing the fully connected layer [18]. Lin et al. proposed CHU-Nets, where a convolutional highway unit is inserted into the traditional CNN structure and the classification performance is improved in a limited-labeled training dataset [19]. Pei et al. proposed a multiple-view DCNN (m-VDCNN) to extract the features from target images with different azimuth angles [20]. Wagner enlarged the training set by directly adding distorted SAR images to improve the robustness [21]. He replaced the softmax classifier in the traditional CNN structure with a SVM classifier and achieved high recognition accuracy [21, 22]. Kechagias-Stamatis et al. fused a convolutional neural network module with a sparse coding module under a decision level scheme, which can adaptively alter the fusion weights that are based on the SAR images [23].

All the above researches focus on feature extraction and network optimization, but without considering image quality. Wang et al. used a despeckling subnetwork to suppress speckle noise before inputting SAR images into a classification network while enhancing the target image is more intuitive and effective [24]. Better features can be extracted by enhancing the target area of the image. Another problem for SAR ATR is that most of the existing research is carried out on datasets of standard conditions, which means when the training data and test data acquisition conditions change, the recognition performance of the algorithm will decrease significantly. In addition, the classifiers trained under the standard conditions usually have weak generalization ability, leading to poor performance in practical applications. Li Wang and Bai et al. improved the convolutional neural network and proposed an enhanced squeeze excitation network (ESENet) [25]. Compared with the existing CNN structure designed for SAR ATR, ESENet can extract more effective features from SAR images and obtain better generalization performance. Wang et al. proposed a twin convolutional neural network [26] (twin CNN) based on metric learning and deep learning to measure the similarity probability between samples. The multitask joint learning method is used to train the model, which effectively alleviates the influence of speckle noise on SAR image and reduces the risk of overfitting caused by too much noise. There are also some kinds of literature that use the method of few-shot learning. Zhang et al. proposed a domain knowledge-powered two-stream deep network (DKTS-N) [27], which incorporates SAR domain knowledge related to the azimuth angle, the amplitude, and the phase data of vehicles. This scheme is a pioneering work in few-shot SAR vehicle recognition. Li Wang et al. designed a convolutional bidirectional long short-term memory (Conv-BiLSTM) network as an embedding network to map the SAR images into a new feature space [28].
In practical applications, especially in operational scenarios, SAR ART still has two unsolved problems in the above research. Firstly, it is challenging to build a complete test set by capturing a large number of samples due to the noncooperative characteristics of the target and the working environment of the sensor. Secondly, with the increase in data types and numbers, the existing cumbersome network is inefficient and has poor real-time performance.

In order to solve the problems mentioned above, both the network and the SAR image data are considered. Firstly, we propose a multimodule image enhancement network (MMIE-Net) to solve the problem that the complexity network affects real-time processing. Secondly, we add a migration learning module to optimize the network to solve the problem of fewer datasets in the target domain.

Target enhancement is dedicated to improving the image quality of the target area for better feature extraction. Aiming at finding out the simplest and most effective network, especially solving the problem of recognition rate reduction under extended operation conditions, a suitable network for SAR target recognition is also proposed. From this point of view, we explore the SAR image recognition method based on target extraction and image enhancement along with the suitable feature extractor selected. The specific contributions of this paper are as follows:

1. A multimodule image enhancement network for SAR images is proposed, which is better than threshold segmentation and obtains a complete target image. A multimodule image enhancement network is composed of a new target extraction module based on Markov random field and a new image processing module based on transfer learning.

2. A multimodule image enhancement method is proposed and applied in network training. Compared with original images, using enhanced images can increase the recognition rate by 11.34%.

3. A simple convolutional neural network for SAR ATR is proposed. Our lightweight MMIE-Net not only is superior to some existing deep convolutional neural networks for SAR ATR in terms of parameters and training speed but also has better recognition performance.

The organization of this paper is as follows: Section 2 introduces the MMIE-Net, target extraction method, image enhancement approach, network models, and the steps of the experiment. Section 3 represents the experimental results under EOC1 and SOC datasets. Section 4 discusses the results and gives an explanation. Section 5 clarifies the conclusions and summarizes our method.

2. SAR ATR Based on MMIE-Net

Section 2.1 will give an overall description of the proposed MMIE-Net. We introduce the MMIE-Net’s integral frame. In Section 2.2, a detailed feature extraction module is introduced, including image segmentation, Markov random field correction, and morphological operation. An image enhancement module is based on the idea of transfer learning. An image recognition module is introduced in detail in Section 2.4, mainly including the simple and concise network model we designed.

2.1. MMIE-Net Framework. Before detailing each module, the overall framework of MMIE-Net is shown in Figure 1. The input is SAR images in large scenes. First, the feature extraction module is used to find accurate targets, including target detection and target identification to obtain more accurate targets. Then, the image enhancement module is used to enhance the target definition. Finally, the target recognition module is used to classify the targets to obtain the results. In the feature extraction module, IS refers to image segmentation, MRFC stands for Markov random field correction, and MO represents a morphological operation. The image enhancement module draws on the idea of transfer learning. The design of a concise network makes real-time processing possible in the image recognition module.

We describe the flowchart of the algorithm as shown in Figure 2. This algorithm is mainly divided into three parts. The first part is the target extraction part, including the initial segmentation of the image, Markov random field verification, and morphological reconstruction. The second part is the image enhancement part, which uses the idea of transfer learning to process images. The third part is the target recognition part, which mainly uses a simple lightweight network for target recognition.

2.2. Target Area Extraction. Existing SAR image target extraction methods mainly use threshold segmentation and filtering, which need to manually set the threshold [29]. Meanwhile, the segmentation of the target is not precise enough. To overcome this problem, we propose a target extraction method based on Markov random field.

The specific approaches of our method are as follows:

1. Intercepting the noise regions in the corners of the image and fitting its Rayleigh distribution, then obtaining the threshold, and performing the initial segmentation of the image

2. Using Markov random field to correct the label field after segmentation

3. Using morphological operations to obtain the final complete segmentation

2.2.1. Initial Segmentation of the Image. A SAR image includes three parts of regions: target, background, and shadow. We first verify the pixel distribution of each region. It can be found from Figure 3 that the pixel density distribution is close to the Rayleigh distribution.

\[
f(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right).
\]
After obtaining the statistical characteristics of each region, we make the following assumptions: Each pixel of the image is independent of the others; then the segmentation can be concerned with maximum likelihood estimation when the Rayleigh distribution of each region is determined. The three regions (background area, target area, and shadow area) are marked with labels 1, 2, and 3, respectively.

\[
I_i = \arg \max_{I_i} p(x_i|I_i). \tag{2}
\]

\(i, x_i, \) and \(I_i\) represent the position index, gray value, and label of one pixel. The three situations of \(x_i\) are corresponding to

\[
x_i = \begin{cases} 1 & p(x_i|I_i = 1) > p(x_i|I_i = 2) \& p(x_i|I_i = 1) > p(x_i|I_i = 3) \\ 2 & p(x_i|I_i = 2) > p(x_i|I_i = 1) \& p(x_i|I_i = 2) > p(x_i|I_i = 3) \\ 3 & p(x_i|I_i = 3) > p(x_i|I_i = 1) \& p(x_i|I_i = 3) > p(x_i|I_i = 2) \end{cases}. \tag{3}
\]

Figure 4 shows the Rayleigh distribution of the three different regions. From the statistical distribution, the gray value of the target area is high, the background area is medium, and the shadow area is low. The variance of the three regions of Rayleigh distribution satisfies \(\sigma_2 > \sigma_1 > \sigma_3\); then equation (3) can be simplified as

\[
x_i = \begin{cases} 1 & x_i > T_{12} \& x_i < T_{23} \\ 2 & x_i > T_{23} \\ 3 & x_i < T_{12} \end{cases}. \tag{4}
\]

However, the Rayleigh distribution of the three regions is unclear in the initial segmentation, but the rough distribution of the background can be fitted by collecting part of the background area. Then the thresholds \(T_{12}\) and \(T_{23}\) can be inferred through the false alarm probabilities \(P_{2\text{error}}\) and \(P_{3\text{error}}\).

\[
P_{2\text{error}} = \int_{T_{23}}^{\infty} p(x_i|I_i = 1)dx, \tag{5}
\]

\[
P_{3\text{error}} = \int_{0}^{T_{12}} p(x_i|I_i = 1)dx.
\]

\(P_{2\text{error}}\) represents the probability that the original label 2 is mistakenly judged as label 1, which means the probability of the target area is wrongly classified as the background. \(P_{3\text{error}}\) represents the probability that label 3 is mistakenly recognized as label 1, which is the probability of the shadow area incorrectly noted as the background. In this paper we set \(P_{2\text{error}} = P_{3\text{error}} = 0.05\).

2.2.2. Markov Random Field Correction. Usually, one pixel and its eight-neighborhood pixels in an image form a Markov random field, because they have a correlation with each other.

\(X\) represents the pixel field of a Markov neighborhood and \(\omega\) is its label field. \(\omega\) can be obtained through maximizing the posterior probability:

\[
\omega = \arg \max_{\omega} P(\omega|X). \tag{6}
\]

\(P(X)\) is a constant; according to the Bayesian principle, there is

\[
P(\omega|X) = \frac{P(X|\omega)P(\omega)}{P(X)} \propto P(X|\omega)P(\omega). \tag{7}
\]

So the label field can be obtained by
\[
\omega = \arg\max_\omega P(\omega | X) = \arg\max_\omega P(X | \omega) P(\omega) \\
= \arg\max_\omega \left( \exp(\log P(X | \omega) + \log P(\omega)) \right) \\
= \arg\max_\omega \left( \exp \left( \sum \log P(x_i | l_i) + \log P(\omega) \right) \right). \tag{8}
\]

According to the Hammersley-Clifford theorem [30], only when the label field obeys the Gibbs distribution it can be the Markov field:

\[
P(\omega) = \frac{1}{Z} \exp(-U(\omega)) = \frac{1}{Z} \exp(-\sum_c V_c(\omega)). \tag{9}
\]

\(\omega\) is the label field of the Markov neighborhood with the pixel \(c\) as its center, \(U(\omega)\) is the energy of the label field, and \(Z\) is a normalization constant not affecting \(\omega\); then equation (8) can be written as

\[
\omega = \arg\max_\omega \left( \exp \left( \sum \log P(x_i | l_i) - U(\omega) \right) \right). \tag{10}
\]

The energy \(U(\omega)\) is the sum of the small group potential energy in a Markov random field:

\[
V_c(i, j) = \beta \delta(\omega_i, \omega_j) = \begin{cases} 
-\beta & \text{if } \omega_i = \omega_j, \\
+\beta & \text{if } \omega_i \neq \omega_j,
\end{cases}
\]

\[U(\omega) = \sum_{i,j \in \omega} V_c(i, j). \tag{11}\]

The Markov field correction aims to find the label field which has minimum energy. We use conditional iteration to obtain the optimal solution. Figure 5 compares the changes in the target and shadow area after Markov random field correction.

2.2.3. Morphological Operations. Some small noise areas still exist in the image after Markov field correction; meanwhile, some small hole areas are not marked. In order to obtain the complete target area, it is necessary to perform morphological operations which can be divided into four steps: denoising, dilation, hole filling, and erosion. The results after a series of operations are shown in Figure 6.
In Figure 7, the finally completed region extraction results are as follows.

2.3. Image Enhancement. Ideally using the target image as input can get the best classification results, but two problems exist: (1) There must be information loss when extracting the target area, so obtaining a complete target image is impossible, and (2) in addition to the target areas, shadow areas which are caused by the occlusion of the target also contain useful information.

In order to solve the above problems, we propose to perform image enhancement, superimposing the original image with the processed target image. Here, we draw lessons from the idea of fine-tuning in the transfer learning model [31], which usually freezes most convolution layers close to the input. These convolution layers retain a large amount of underlying information and do not even freeze any full connection layers. Only the remaining convolution layers (usually partial convolution layers close to the output) and full connection layers are trained. Therefore, we combined it with the first layer of the CNNs; then the
superimposing weight and bias can be learned through the network training.

\[ I_{\text{input}}(i, j) = k_1 \times I_O(i, j) + k_2 \times I_T(i, j) + b. \]  

\[ (12) \]

\( I_{\text{input}}, I_O, \) and \( I_T \) represent the enhanced image, the original image, and the target image, respectively. \( k_1 \) and \( k_2 \) are combination weights, and \( b \) is bias. \((i, j)\) indicates the position index of a pixel in the image matrix. Figure 8 shows examples of image enhancement under different parameters (in our experiment, we initialize the parameters \( k_1 = 1, k_2 = 1, b = 0 \)). It can be seen that as \( k_1 \) decreases and \( k_2 \) increases, the noise part of the image is suppressed and the

Figure 8: Enhanced images under different parameters \( k_1, k_2, \) and \( b \).

Figure 9: The structure of MMIE-Net.
target area is enhanced. $b$ controls the overall brightness of the image. The enhancement achieves the best performance when $k_1 = 0.3$, $k_3 = 1$, $b = 0$.

### 2.4. Network Models.
A basic CNN consists of three structures: convolution, activation, and pooling [32, 33]. Large-scale deep convolutional neural networks (CNNs) are usually composed of multiple abovementioned structures. Different CNN’s always have different basic units. We design our networks MMIE-Net (synthetic aperture radar network) and MMIE-Net_plus (synthetic aperture radar network plus).

The network structure of MMIE-Net is shown in Figure 9. The first blue block represents the input data, with the number below representing the image channels, and the following block is the feature maps after convolution. The number below every block indicates the number of the feature maps, and the other two numbers (left and above) represent the size of the feature map. The last purple block is a vector with a dimension of 512, which represents a fully connected layer. There are the BatchNorm layer, ReLu activation layer, and max pooling layer between two adjacent blocks.

Table 1 lists the detailed parameters of MMIE-Net. In the structure column, Conv represents convolution, and Max pool means max pooling. The first parameter of “Conv” is the number of convolution kernels, followed by $n \times n$ which is the size of the convolution kernel. In the output column, the first number and following $n \times n$ are the number and size of feature maps.

<table>
<thead>
<tr>
<th>Layer name</th>
<th>Structure</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prelayer</td>
<td>Conv (64, 3 × 3, stride = 1)</td>
<td>(64, 64 × 64)</td>
</tr>
<tr>
<td></td>
<td>Max pool (2 × 2, stride = 2)</td>
<td>(64, 32 × 32)</td>
</tr>
<tr>
<td>Layer 1</td>
<td>Conv (128, 3 × 3, stride = 1)</td>
<td>(128, 32 × 32)</td>
</tr>
<tr>
<td></td>
<td>Max pool (2 × 2, stride = 2)</td>
<td>(128, 16 × 16)</td>
</tr>
<tr>
<td>Layer 2</td>
<td>Conv (256, 3 × 3, stride = 1)</td>
<td>(256, 16 × 16)</td>
</tr>
<tr>
<td></td>
<td>Max pool (2 × 2, stride = 2)</td>
<td>(256, 8 × 8)</td>
</tr>
<tr>
<td>Layer 3</td>
<td>Conv (512, 3 × 3, stride = 1)</td>
<td>(512, 8 × 8)</td>
</tr>
<tr>
<td></td>
<td>Max pool (2 × 2, stride = 2)</td>
<td>(512, 4 × 4)</td>
</tr>
<tr>
<td>Layer 4</td>
<td>Conv (512, 3 × 3, stride = 1)</td>
<td>(512, 4 × 4)</td>
</tr>
<tr>
<td></td>
<td>Max pool (4 × 4, stride = 1)</td>
<td>(512, 1 × 1)</td>
</tr>
</tbody>
</table>

Stitching of the different color blocks represents the feature map concatenation. The blue arrows represent convolution, activation, and pooling operations. Detailed layers and parameters of the network in Tables 2 and 3 are as follows.

### 2.5. Experimental Setup.
The experiment is designed to verify whether using image enhancement contributes to SAR ATR performance, especially the improvement of the recognition rates under extended operating conditions. The specific steps of the experiment are shown in Figure 11. First, we extract the target image from the original image. Then the target image and the original image are fused through image enhancement. Next, using these three kinds of images as input, we train networks and compare their recognition results. Afterward, we compare different convolutional neural networks on the enhanced image dataset and select the most suitable network. Finally, we verify our method on the SOC dataset.

The image is center-cropped to a standard size 64 × 64 before inputting CNN. In order to improve the generalization ability of the network, the image is randomly rotated between 0 and 15° and flipped horizontally and vertically.

The networks are built and trained using the open source framework Pytorch; its hyperparameters are set as follows: the initial learning rate is $1e^-4$ and will be reduced by 0.5 times every 10 epochs if it is not less than $1e^-8$. Moment momentum and gradient descent are set to 0.9 and 0.1, respectively. The training batch size is 32 and the number of training epochs was 100. When the training finished, the network parameters were applied to the test dataset. All experiments are performed on a device with Intel(R) Core(TM) i5-8300H CPU, 8 GB RAM, and GeForce GTX 1050Ti GPU.

### 3. Results

#### 3.1. Experiment Dataset.
The MSTAR dataset is a SAR radar dataset launched by the US Defense Advanced Research Projects Agency (DARPA) in the mid-1990s [35]. The high-resolution spotlight synthetic aperture radar is used to collect SAR images of various target military vehicles of the former Soviet Union. The MSTAR program carried out SAR ground target testing, including target occlusion, camouflage, configuration changes, and other expandable conditions, forming a more systematic and comprehensive database. The acquisition conditions of the MSTAR dataset can be divided into two categories, Standard Operating Condition (SOC) and Extended Operating Condition.
The SAR image under EOC is generated based on a variety of acquisition conditions, such as different imaging side views, target postures, or target models. Therefore, the MSTAR dataset can test the overall performance of the SAR target recognition method.

Under SOC, the training set and test set have a total of 10 types of targets. These military targets are 2S1 (self-propelled howitzer), BRDM2 (armored reconnaissance vehicle), ZSU234 (self-propelled antiaircraft gun), and T72 (tank). Table 5 lists the detailed information of the dataset. The depression angles of the training set and test set are 17° and 30°, respectively, which makes the target images greatly deformed and the shadow area also changes accordingly. Therefore, this dataset can test the generalization ability of the algorithm.

3.2. Experiments on the EOC Dataset. In this section, the experiments are conducted on the EOC datasets. The original images, the target images, and the enhanced images are used to train the network, respectively. Ten the highest and average recognition rates on the training dataset and test dataset are recorded to analyze the performance.

The detailed information of the dataset under the SOC (target type, image resolution, image number, image acquisition side view, etc.) is shown in Table 4. EOC includes two experimental datasets EOC-1 and EOC-2. EOC-1 mainly contains four types of targets: 2S1 (self-propelled howitzer), BRDM2 (armored reconnaissance vehicle), ZSU234 (self-propelled antiaircraft gun), and T72 (tank). Table 5 lists the detailed information of the dataset. The depression angles of the training set and test set are 17° and 30°, respectively, which makes the target images greatly deformed and the shadow area also changes accordingly. Therefore, this dataset can test the generalization ability of the algorithm.

### Table 3: MMIE-Net_plus’s (main network) detailed parameters.

<table>
<thead>
<tr>
<th>Main network</th>
<th>Structure</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prelayer</td>
<td>Conv (64, 3 × 3, stride = 1)</td>
<td>(64, 64 × 64)</td>
</tr>
<tr>
<td>Layer 1</td>
<td>Concatenation (128, 64 × 64)</td>
<td>(128, 64 × 64)</td>
</tr>
<tr>
<td></td>
<td>Max pool (2 × 2, stride = 2)</td>
<td>(128, 32 × 32)</td>
</tr>
<tr>
<td></td>
<td>Conv (128, 3 × 3, stride = 1)</td>
<td>(128, 32 × 32)</td>
</tr>
<tr>
<td>Layer 2</td>
<td>Concatenation (256, 32 × 32)</td>
<td>(256, 16 × 16)</td>
</tr>
<tr>
<td></td>
<td>Max pool (2 × 2, stride = 2)</td>
<td>(256, 16 × 16)</td>
</tr>
<tr>
<td></td>
<td>Conv (256, 3 × 3, stride = 1)</td>
<td>(256, 16 × 16)</td>
</tr>
<tr>
<td>Layer 3</td>
<td>Concatenation (512, 16 × 16)</td>
<td>(512, 8 × 8)</td>
</tr>
<tr>
<td></td>
<td>Max pool (2 × 2, stride = 2)</td>
<td>(512, 8 × 8)</td>
</tr>
<tr>
<td></td>
<td>Conv (512, 3 × 3, stride = 1)</td>
<td>(512, 4 × 4)</td>
</tr>
<tr>
<td>Layer 4</td>
<td>Max pool (2 × 2, stride = 2)</td>
<td>(512, 4 × 4)</td>
</tr>
<tr>
<td></td>
<td>Conv (512, 3 × 3, stride = 1)</td>
<td>(512, 4 × 4)</td>
</tr>
<tr>
<td>Layer 5</td>
<td>Max pool (4 × 4, stride = 1)</td>
<td>(512, 1 × 1)</td>
</tr>
</tbody>
</table>

(EOC). The SAR image under EOC is generated based on a variety of acquisition conditions, such as different imaging side views, target postures, or target models. Therefore, the MSTAR dataset can test the overall performance of the SAR target recognition method.

Under SOC, the training set and test set have a total of 10 types of targets. These military targets are 2S1 (self-propelled howitzer), BRDM2 (armored reconnaissance vehicle), BTR60 (armored transport vehicle), D7 (Bulldozer), T62 (tank), ZIL131 (cargo truck), ZSU234 (self-propelled artillery), BTR70 (armored transport vehicle), BMP2 (infantry fighting vehicle), and T72 (tank). The training set images were obtained when the radar working elevation angle is 17°, and the test set images were obtained when the radar working elevation angle is 15°. Various types of targets also have different models; the same type but different models have differences in their equipment, but their overall scattering characteristics are similar.

Figure 10: The structure of MMIE-Net_plus.
As shown in Figure 14, using the enhanced image as input data, the recognition rates improve to 99.58% and 99.57% on the test and training; almost all pictures are correctly recognized. Meanwhile, the network is very stable because the curves are close to a horizontal line at the end, which also means no overfitting occurs. In addition, the network training speeds up, beginning to converge at the 30 epochs. It is another benefit of using enhanced images.

Figure 15 compares the performance of the network under different images dataset. When the network is stable, the recognition rate using the original image is the lowest at 90.44%, and its average recognition rate is 87.95%. The blue accuracy curve has strong fluctuations, which means the network is unstable. When using the extracted target image as input, the network stability has obviously improved, the recognition rate also improved to 96.00%, and the average accuracy rate is 95.59%. Finally, the enhanced image is used as input. The recognition rate reaches 99.58% and the average recognition rate is 99.29%, which is 11.34% higher than the original image recognition rate. Meanwhile, the network convergence speed is quicker than in the previous two cases.

At last, we list the recognition rate using different inputs. The “highest rec” in Table 6 refers to the highest one-time recognition rate obtained during the network training process. The “average rec” represents the average recognition rate derived from multiple training after the network is stable. “Test” and “Training” indicate which dataset the recognition rate is inferred from. Using enhanced images improves the recognition rate of SAR images under EOC, and the network is stable. Compared with using the original

<table>
<thead>
<tr>
<th>Vehicle class</th>
<th>17° depression angle</th>
<th>15° depression angle</th>
<th>Image size</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP2</td>
<td>233</td>
<td>195</td>
<td>128 × 128</td>
</tr>
<tr>
<td>BTR70</td>
<td>233</td>
<td>196</td>
<td>128 × 128</td>
</tr>
<tr>
<td>T72</td>
<td>232</td>
<td>196</td>
<td>128 × 128</td>
</tr>
<tr>
<td>BTR60</td>
<td>256</td>
<td>195</td>
<td>128 × 128</td>
</tr>
<tr>
<td>2S1</td>
<td>299</td>
<td>274</td>
<td>158 × 158</td>
</tr>
<tr>
<td>BRDM2</td>
<td>298</td>
<td>274</td>
<td>128 × 129</td>
</tr>
<tr>
<td>D7</td>
<td>299</td>
<td>274</td>
<td>177 × 178</td>
</tr>
<tr>
<td>T62</td>
<td>299</td>
<td>273</td>
<td>172 × 173</td>
</tr>
<tr>
<td>ZIL131</td>
<td>299</td>
<td>274</td>
<td>192 × 193</td>
</tr>
<tr>
<td>ZSU-23-4</td>
<td>299</td>
<td>274</td>
<td>158 × 158</td>
</tr>
<tr>
<td>Total</td>
<td>2747</td>
<td>3202</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vehicle class</th>
<th>Training dataset</th>
<th>Test dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>2S1</td>
<td>299</td>
<td>288</td>
</tr>
<tr>
<td>BRDM2</td>
<td>17° 298</td>
<td>30° 287</td>
</tr>
<tr>
<td>T72</td>
<td>299</td>
<td>288</td>
</tr>
<tr>
<td>ZSU234</td>
<td>299</td>
<td>288</td>
</tr>
</tbody>
</table>

Table 4: MSTAR dataset details on SOC (standard operating condition).

Table 5: MSTAR dataset details on EOC-1 (extended operating condition-1).
image, the average recognition rate increased by 7.94% through target extraction and increased by 11.34% performing image enhancement.

In order to explore the upper limit of the depression angle of the test dataset, we add several sets of experiments to the training set with a depression angle of 17°. Segmenting the depression angle is beneficial to the development of the experiment; we set the depression angle of the test set to 45°, 60°, 75°, and 90° for testing. The specific test results are as follows.

In Table 7, we can clearly see that the depression angle of the test set has a good effect below 60°. When the angle reaches 75°, the depression angle begins to decrease rapidly until the image becomes unrecognizable when the angle reaches 90°. This also shows that our proposed method is also applicable to the case where the image depression angle is large.

3.3. Experiments on the SOC Dataset. In this part, the image enhancement method is verified on the SOC dataset. Figure 16 shows the result of the network training process using three different inputs. Each curve reveals the change in recognition rate as the training epoch increases. The highest recognition rate using the original images is 99.51%, and the average recognition rate is 99.42%. While using target images, the highest recognition rate and the average recognition rate are 97.11% and 96.69%. Therefore, using original images has better performance. When using enhanced images as input, the highest recognition rate is 99.55% and the average recognition rate is 99.39%; it is the best recognition performance we obtained on the SOC dataset.

3.4. Comparison of Results. In Table 8, the ResNet, DenseNet, ESENNet, and SiameseNet_+ are used as comparison algorithms. In the EOC dataset and the SOC dataset, they are compared to MMIE-Net regarding the network recognition performance. Both MMIE-Net and MMIE-Net_plus have high recognition rates. DenseNet is not suitable for the SAR images. ResNet is bad for SAR images in the SOC dataset. Furthermore, we find that a 5-layer
After a series of multimodule enhancements, such as target feature extraction and image enhancement, SAR images are generally pure and fine after denoising and supplementary processing. Although ESENet and SiameseNet all have better results, which benefit from their special network structure, the amount of calculations is much greater than the other networks. Therefore, MMIE-Net is superior to other existing algorithms in terms of prediction accuracy and computational cost.

In order to more clearly see the performance comparison between the proposed network and other existing networks, in Figure 17, we show the convergence and accuracy of each network model on the EOC after several iterations. We find that our proposed network not only has high accuracy but also has fast convergence speed compared with other existing networks.

### 4. Discussion

In the experiment of the EOC dataset, the recognition rate is improved by using the target image. We visualize the feature map of the network to further explain this increment. The number of layers in Figure 18 refers to the convolutional layer. We randomly select 16 feature maps for visualization from the first layer to the third layer. Taking one image of ZSU as an example, this image cannot be correctly identified using the original image but recognized with the target image.

In order to observe the feature map more intuitively, the gray-scale feature maps are converted into pseudo-color maps. As shown in Figure 19, compared to using the original image, the feature maps obtained from the target image are clearer; this phenomenon is especially obvious in the feature maps of the first and second convolutional layers. Actually,
the target image only retains the target regions which have obvious features, causing the feature maps to have a strong degree of discrimination. Therefore, the extracted features are clear in the third layer. However, the feature map obtained from the original image is relatively fuzzy in the third layer, which is confusing and difficult to understand.
Usually, the network feature maps expression is gradually abstracted as the depth increases, and the feature maps of the shallow convolutional layer should include part of the outline of the object. The fuzzy feature maps derived from the original image indicate that little useful feature is extracted. Therefore, the extraction of the target image cannot only remove the interference of noise and background, but also highlight the characteristic information of the target and improve the recognition ability and stability of the network.

However, on the SOC dataset, the recognition performance using the target image is worse than using the original image. Some wrong prediction images are listed to analyze this contradiction. The recognition rate using the target image depends on target extraction. Take BTR_60 as an example; the following images in Figure 16 are wrong prediction images. The recognition rate of BTR60 is 94.36%; in a total of 195 pictures, 184 pictures are correctly recognized and 14 pictures are wrong.

The wrongly recognized target images have the following characteristics:

1. The target image is occluded by shadow, which makes the target merge with the background and shadow, making it difficult to extract the target region
2. Target extraction removes some key parts of the image information, leading to network misclassification

The misclassification caused by incomplete target extraction can be solved through image enhancement. But the first type of problem inevitably affects the quality of data.

In Figure 20, take one image of 2S1 as an example: the image is wrongly recognized using the target image but correctly classified with the enhanced image. Compared with the original image feature maps, the feature map of the enhanced image has more obvious graphics and clearer feature maps. In contrast to the feature maps of the target image, enhanced image feature maps have no mutation at the edges, which increases the compatibility of feature matching. As a result, the enhanced image has better recognition performance than the original image and the target image.

5. Conclusions

Aiming at the problems of poor SAR image quality and complex real-time performance of existing solutions, this study proposed a SAR image recognition method based on multimodule image enhancement. At first, a target extraction method for SAR images was proposed. This method could obtain effectively the target area of SAR images, which has higher accuracy and adaptive capabilities than threshold segmentation. Then, an image enhancement method was applied to improve the recognition performance of SAR images. The original images, target images, and enhanced images were inputted networks, respectively, and the results proved image enhancement could effectively improve the
generalization performance and applicability of the network. Finally, we analyzed it through feature maps and gave an explanation. In addition, we designed a lightweight network for SAR image recognition. The network had better performance for SAR image recognition compared with several other networks. It had a higher recognition rate, fewer parameters, and fewer calculations. Future work will focus on improving model performance and its utility on non-public datasets while researching models for target extraction in multitarget regions of SAR images.

Data Availability

Data are from Sandia National Laboratory of the United States (SAR image was provided by Sandia National Laboratories) (attached: Sandia National Laboratory of the United States, which is subordinate to the National Nuclear Safety Administration of the U.S. Department of energy).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was funded by the Program of the National University of Defense Technology and the National Natural Science Foundation of China (no. 61602491).

References


