Robust SAR Automatic Target Recognition Via Adversarial Learning

Yuchen Guo, Lan Du, *Senior Member, IEEE*, Di Wei, and Chen Li

***Abstract*—The traditional denoising methods in noise robust synthetic aperture radar (SAR) automatic target recognition research are independent of the recognition model, which limits the robust recognition performance. In this article, we present a robust SAR automatic target recognition method via adversarial learning, which could integrate data denoising, feature extraction, and classification into a unified framework for joint learning. Different from the common recognition methods of directly inputting the SAR data into the classifiers, we add a dual-generative-adversarial-network (GAN) model between the SAR data and the classifier for data translation from a noise-polluted style to a relatively clean style to reduce the noise from SAR data. In order to ensure the target information in the SAR data can be retained during the data style translation, reconstruction constraint and label constraint are also used in the dual-GAN model. Then, the more reliable transferred SAR data are fed into the classifier. The parameters of the dual-GAN and classifier are learned through joint optimization in our method. Thus, the data separability is guaranteed in the process of denoising and feature extraction, which greatly improves the recognition performance of the method. In addition, our method can be easily extended to a semisupervised method by using different objective functions for labeled and unlabeled training data, which is more suitable for practical application. Experimental results on MSTAR dataset and Gotcha dataset show that our method can get the encouraging performance in the case of low signal-to-noise ratio and small labeled data size.**

***Index Terms*—Adversarial learning, automatic target recognition (ATR), generative adversarial networks (GAN), noise robust, semisupervised learning, synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) is an active ground observation system that can be installed on aircraft, satellites, spacecraft, and other flight platforms [1]. Also, it can perform observations on the ground at all times and all days and has the ability to penetrate the ground. Therefore, the SAR has unique advantages in the application of disaster monitoring, environmental monitoring, marine monitoring, resource exploration, crop estimation, mapping, and military and can play a role that other remote sensing methods cannot play [2]–[5]. Thus, it is increasingly being used by countries around the world. SAR processing can produce high-resolution imagery containing 2-D information [1]. Besides, SAR images contain both shape information and scattering information, presenting a good representation of the objects. For this reason, SAR images are also becoming a useful tool for radar automatic target recognition (ATR) [4], [5], [38]–[40].

However, when electromagnetic wave encounters rough-surface reflection, due to the phase difference, the echo interference occurs, resulting in the echo strength becoming stronger and weaker and the speckle pattern forming the interference noise to the SAR data. The noise is distributed over the whole SAR image and will mask the target information and, thus, reducing the identifiability of the target and increasing the difficulty of SAR target recognition [6]. Experimental results presented in [7] showed that the classification performance of SAR images dramatically deteriorates under the high noise level. Therefore, the noise robustness of a recognition algorithm is very important in the SAR ATR research.

In practice, there are usually three approaches to improve the recognition performance of SAR data under the high noise level.

1) One approach is to extract the noise-robust feature of SAR data. Xu *et al.* [8] proposed a robust principal component analysis (PCA) method to get noise-robust features of SAR data by extracting the main component features. In [7], sparse-representation-based classification (SRC) and attributed scattering center (ASC) matching were hierarchically fused to ensure that the SAR recognition framework can extract robust features. Nevertheless, since the traditional methods of extracting noise-robust features are learned separately from the classifiers, the classifiable distances between the features extracted from different categories of targets cannot be guaranteed, which may affect the final recognition performance.
2) A popular approach in recent research is data augmentation [9], [10]. Both the expanded data, which are generated by adding noise to the original SAR training samples, and the original SAR training samples participate in the training of the recognition model to improve the noise robustness [10]. However, in order to extract the noise-robust features for the noised SAR data and guarantee the recognition performance, it is necessary to greatly increase the complexity of the recognition model along with data augmentation.
3) The last but also the more natural choice is to remove the noise component in SAR data before classification. These

Manuscript received June 27, 2020; revised August 27, 2020 and October 13, 2020; accepted November 4, 2020. Date of publication November 19, 2020; date of current version January 6, 2021. This work was supported in part by the National Science Foundation of China under Grant 61771362, Grant U1833203, and Grant 61671354 and in part by 111 Project (B18039) *(Corresponding author: Lan Du.)*

The authors are with the National Lab of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: ycguo_xd@163.com; dulan@mail.xidian.edu.cn; xdweidi@126.com; chenli@stu.xidian.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3039235

methods can be classified into two categories: the image-domain methods, i.e., average filtering, median filtering, refined Lee filtering [50], and Gamma-MAP filtering [11], and frequency-domain methods, i.e., RELAX [12] and OMP [13] algorithms. Nevertheless, the goal of these denoising methods is to remove noise from SAR images rather than for the recognition task. In addition, both of RELAX and OMP require a prior knowledge of the noise level, which is hard to obtain in practice.

Generative adversarial network (GAN) [14] is a family of successful models which learn to generate with the constraint functions learned jointly with the model itself. GAN have recently demonstrated impressive performance for image generation and received widespread attention. Models with adversarial constraint have been used in a wide range of cross-domain applications. The cross-domain research study mainly focuses on the translation and fusion of data from different domains, such as words in different languages [35], words and images [19], and images of different styles [17], [18], [20]–[22]. In computer vision, domain adaptation methods for image classification task have already been widely studied. Recent works [41]–[43] implement domain confusion with gradient reversal layer or GAN loss for feature alignment through adversarial learning. Besides that, inspired by the success of image-to-image translation techniques, other pixel-level works [44] aim to achieve domain adaptation through translating images from the source domain to the target domain. Hertzmann *et al*.'s image analogies [21] first proposes the idea of image-to-image translation. Most common methods implement image-to-image translation by using input–output training image pairs in the training phase [21], [22]. Different from other traditional methods that require one-to-one image pairs, GAN [14], [23] learned to capture special characteristics of one image domain and to figure out how these characteristics could be translated into the other image domain without any paired training examples. Inspired by the GAN, it is desirable to introduce the image-to-image translation by adversarial learning into recognition to reduce the noise in SAR images. It occurred to us that if the SAR images in low signal-to-noise ratio (SNR) could be transformed into an image style such as the SAR images in high SNR without losing target information, we could reduce the nuisance noise in SAR images.

Deep learning, which is inspired by the human learning system, has launched a profound reformation and even been applied to many real-world tasks [24]–[26], in which image classification is one of the most important applications. The two steps of feature extraction and classification in the recognition framework are integrated into one model in deep learning methods, which greatly improves the classification performance of the model. Many researchers have applied deep learning to SAR ATR research and achieved good recognition performance [4], [26]. For example, Xu *et al*. [26] obtained excellent recognition results by training the deep network with data-augmented SAR images. Naturally, we want to apply the deep learning method to our SAR recognition model. Nonetheless, recognition via deep learning methods obviously falls into the scope of supervised learning, which means that a lot of labeled data are provided for the training processes. In practice, we may measure a lot of data in real environment which is difficult to labeled accurately. Fully supervised deep learning methods cannot use these data to learn more abundant information for extracting more robust features. In the traditional machine learning fields, many semisupervised learning algorithms are introduced to achieve the label prediction by training with limited samples, such as semisupervised SVM [36] and semisupervised sparse representation [37]. With the development of deep learning, some research studies incorporating deep learning with semisupervised algorithms are proposed recently [34]. In semisupervised deep learning, there are two ways to use unlabeled data to participate in model learning: One is to add a new loss function, such as reconstruction loss [45], and the other is to use the well-learned network with labeled data to predict the pseudolabel for the unlabeled data [46]. By reducing the requirement of labeled samples, these semisupervised algorithms are able to improve the training and recognition capability of the deep learning system with limited dataset to some extent. Therefore, in the SAR ATR via deep learning research, it would be satisfying if the model could have the semisupervised learning ability to utilize the information of unlabeled data.

In this article, we present an image recognition method via adversarial learning, which could integrate data denoising, feature extraction, and classification into a unified framework for joint learning. We use a deep learning framework as the classifier. And a dual-GAN model is added between the images and the classifier for image translation. Different from traditional GAN, the dual GAN has two generators and two discriminators to realize image translation in two symmetrical directions. By training the dual-GAN model with low SNR SAR data and high SNR SAR data, the dual GAN can reduce the noise irrelevant to the recognition task in SAR data. Furthermore, to ensure that the transferred SAR images retain the information related to recognition, reconstruction constraint and label constraint are also used in the dual-GAN model. The reconstruction constraint can ensure that the target information can be retained during image transformation, whereas the label constraint can ensure that the transferred image is as similar as possible to the same category of the original SAR data. Then, the more reliable transferred SAR images are used to train a classification deep network. The dual GAN and the classification deep network are in a joint learning framework. Furthermore, we extend our method to a semisupervised learning method by using different objective functions for labeled data and unlabeled data, so that it is more suitable for practical applications. In this article, in order to verify the performance of our method, we use a simple network LeNet as the classification network. With applications on the MSTAR dataset and Gotcha dataset, our proposed method gets encouraging results in the case of low SNR ratio and small labeled data size.

The main contributions of this article are summarized as follows.

1) The proposed SAR ATR method integrates data denoising and feature extraction and classification into a unified framework for joint learning.
2) A dual GAN is proposed for SAR data denoising via adversarial learning. Our dual GAN can provide more

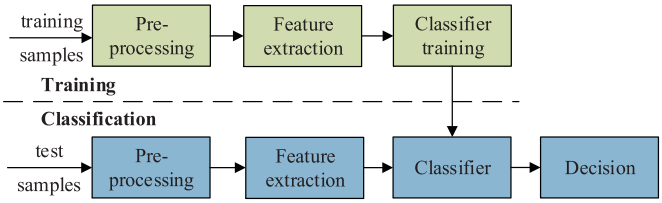


Fig. 1. Typical scheme of SAR ATR with training stage and classification stage.

reliable image translation results than traditional GAN via both of the reconstruction constraint and label constraint.

3) Different from traditional denoising methods, most of which rely on prior knowledge of noise level, our method uses the idea of cross-domain to remove noise information, which is more reliable when the noise level prior information is not accurate.
4) We extend this method to semisupervised learning. A large amount of unlabeled data will enable the deep network to learn more reliable robust features.

The rest of the article is organized as follows. In Section II, SAR ATR and GAN are described. In Section III, our adversarial learning method is proposed. Experimental results compared with other methods on MSTAR dataset and Gotcha dataset are shown in Section IV. Finally, Section V concludes this article

II. Background

A. SAR ATR Framework

A typical SAR ATR scheme is depicted in Fig. 1. As we can see, there are two stages, i.e., training stage and classification stage, in the recognition procedure, both of which include preprocessing and feature extraction steps. For robust SAR ATR by traditional denoising methods, as discussed in Section I, the processing technique of classification stage is needed to reduce the noise of SAR data. And in the deep learning methods, the feature extraction step and classifier step are usually combined in one model. For example, the first several layers of convolution neural network (CNN) are the feature extraction layer, and the last one is classification layer.

B. Generative Adversarial Networks

As an excellent generative model, the GANs have exploded many interesting applications of image generation [23], [28]. The key to GANs' success is the idea of an adversarial loss that forces the generated images to be indistinguishable from real images.

In the traditional GAN, for mapping $\mathbf{X} \rightarrow \mathbf{Y}$, in which $\mathbf{X}$ and $\mathbf{Y}$ present two different data domains, the generator G and a discriminator D compete in a two-player minimax game. The discriminator tries to distinguish real training data from fake data; and the generator tries to fool the discriminator. Concretely, the network D tries to minimize the loss function

$$L_D = -E_{\mathbf{y}}[\log D(\mathbf{y})] - E_{\mathbf{x}}[log(1 - D(G(\mathbf{x})))] \quad (1)$$

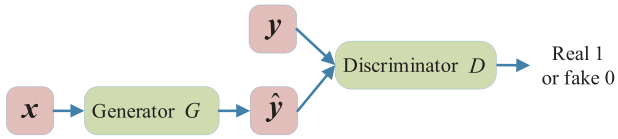


Fig. 2. Traditional GAN framework, and $\hat{\mathbf{y}} = G(\mathbf{x})$. It pits two adversaries against each other in a game. Each player is represented by a differentiable function controlled by a set of parameters.

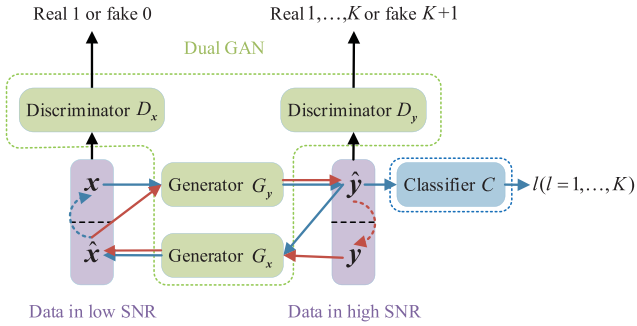


Fig. 3. Illustration of our method structure. Our proposed method contains two parts : 1) A dual GAN capable of two-way domain transfer and contains two generators G_Y and G_X, two discriminators D_Y and D_X; 2) The classification network C. For a more intuitive understanding, we use blue and red lines to represent the two data flows $\mathbf{x} \rightarrow \hat{\mathbf{y}} \rightarrow \hat{\mathbf{x}}$ and $\mathbf{y} \rightarrow \hat{\mathbf{x}} \rightarrow \hat{\mathbf{y}}$, respectively.

whereas network G tries to minimize

$$L_G = E_{\mathbf{x}}[log(1 - D(G(\mathbf{x})))] \quad (2)$$

$E_{\mathbf{y}}[\log D(\mathbf{y})]$ indicates the expectation that the discriminator D judges $\mathbf{y}$ as a real sample, and $E_{\mathbf{x}}[log(1 - D(G(\mathbf{x})))]$ indicates the expectation that the discriminator D judges the generated sample $G(\mathbf{x})$ as a fake sample. In the process of the game, the performance of both generator G and discriminator D is getting better and better. The process of the game is described in detail in Fig. 2: For updating D, $\mathbf{y}$ is sent to D, and the expected output is 1, whereas $\hat{\mathbf{y}} = G(\mathbf{x})$ is sent to D and the expected output is 0; for updating G, the $\hat{\mathbf{y}} = G(\mathbf{x})$ is sent to D and the expected output is 1, which means the G tries to fool the D that the $\hat{\mathbf{y}} = G(\mathbf{x})$ is from the domain $\mathbf{Y}$.

Compared with other generative models, GANs have two main characteristics: 1) GANs do not depend on any priori hypotheses. Many traditional methods assume that the data obey a distribution and then use maximum likelihood to estimate the data distribution [27]. However, the idea of GANs is simple. A discriminator is used to measure the distance between the distribution generated by the generator and the real data distribution without maximum likelihood. 2) The way to generate real-like samples is very simple. The way GANs generate real-like samples is propagated forward by generator, whereas the traditional sampling method is very complex, such as nearest neighbor resampling, bilinear resampling, and bicubic resampling.

III. Method

In this section, we introduce the proposed SAR ATR method via adversarial learning. As shown in Fig. 3, our proposed method contains two parts: 1) the dual GAN and 2) the classification network C. The dual GAN consists of four components:

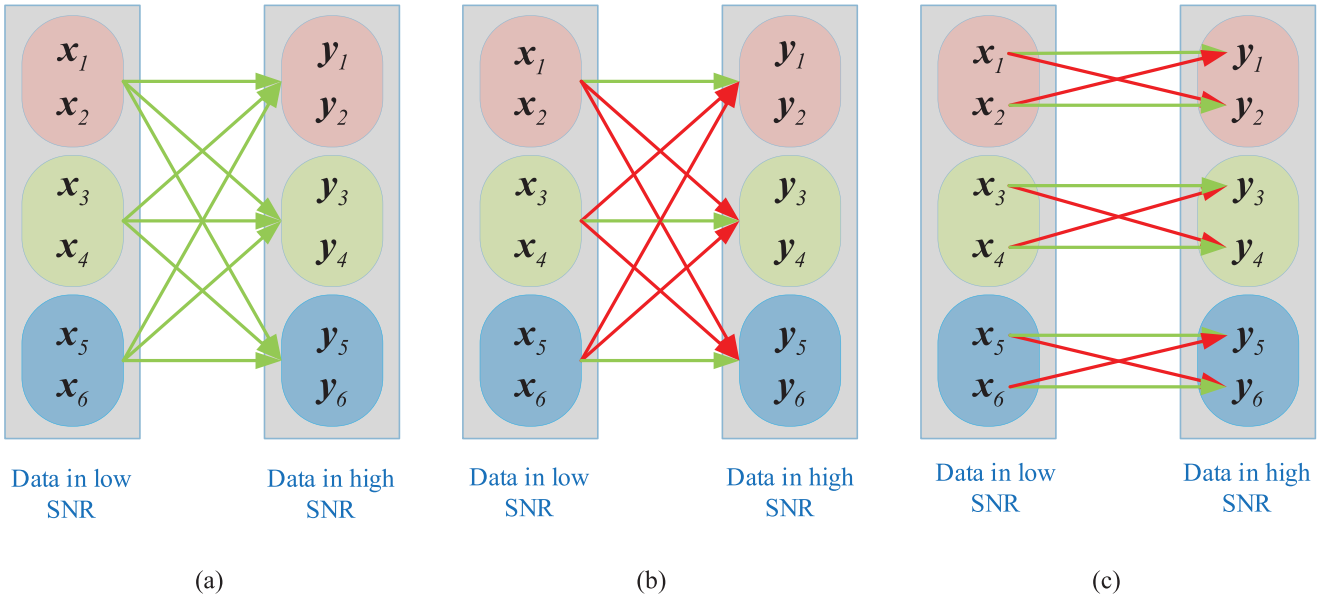


Fig. 4. Illustration of the mapping of different constraints. The background color of the data represents the category the data belongs to, of which the same background color means the same target category. The color of the arrow indicates whether the mapping is allowed: Green indicates yes, and red indicates no. The number in the subscript presents the corresponding image, i.e., $\mathbf{x}_1$ and $\mathbf{y}_1$ are the same image with different SNRs. (a) Original GAN. (b) GAN with the label constraint. (c) Dual GAN with the label constraint and the reconstruction constraint.

denoising generator G_Y for image denoising, image-restoration generator G_X for image reconstruction to make sure the target information can be retained during the image denoising, discriminator D_Y designed to assist the denoising generator G_Y to generate high SNR image, and discriminator D_X designed to assist the image-restoration generator G_X to reconstruct the original low SNR image. More specifically, the denoising generator G_Y first maps $\mathbf{x}$ to a fake image $\hat{\mathbf{y}}$. Then, the image-restoration generator G_X maps $\hat{\mathbf{y}}$ to a fake image $\hat{\mathbf{x}}$, and the classifier C uses $\hat{\mathbf{y}}$ to predict the label l.The quality of G_Y is judged by how well $\hat{\mathbf{y}}$ fools the discriminator D_Y, how well the $\hat{\mathbf{x}}$ fools the discriminator D_X, and how well the label predicts. For a more intuitive understanding, we use blue and red lines to represent the two data flows $\mathbf{x} \to \hat{\mathbf{y}} \to \hat{\mathbf{x}}$ and $\mathbf{y} \to \hat{\mathbf{x}} \to \hat{\mathbf{y}}$ in Fig. 3, respectively. During the training stage, low SNR data $\mathbf{x}$ is fed into G_Y to generate $\hat{\mathbf{y}} = G_Y(\mathbf{x})$ to fool the D_Y into thinking that $\hat{\mathbf{y}}$ is high SNR data. Then, the $\hat{\mathbf{y}}$ is fed into G_X to generate $\hat{\mathbf{x}} = G_X(\hat{\mathbf{y}})$ to fool the D_X into thinking that $\hat{\mathbf{x}}$ is low SNR data. At the same time, the $\hat{\mathbf{x}} = G_X(\hat{\mathbf{y}})$ is expected to be able to be the same as $\mathbf{x}$. It is the same process for data $\mathbf{y}$. For the classification network C, the $\hat{\mathbf{y}} = G_Y(\mathbf{x})$ is fed into C to predict the label of $\mathbf{x}$. It should be pointed out that G_X, D_X, and D_Y only participate in the training phase. In the test phase, our method contains only two parts: 1) image translation through generator G_Y to get an SAR image with high SNR and 2) label prediction through classifier C. We introduce the details of our method in the next part of this section.

A. Dual-GAN Model

As discussed in Section II, the GAN realizes the image translation by reducing the distribution discrepancy between two data domains that forces the generated images to be indistinguishable from real images. However, the traditional GAN constraint only considers the marginal (global) distributions between domains. in the SAR ATR applications, the marginal (global) distributions and conditional (local) distributions between domains are often contributing differently to the image translation. As shown in Fig. 4(a), the traditional GAN constraint allows the model to learn the difference between high and low SNR in SAR data, but cannot guaranteed that the generated image will correspond to the original image. More specifically, if we input a T72 SAR data with low SNR, the GAN model may output BMP2 SAR data with high SNR. It is not expected for the recognition task.

In our dual GAN, the generators G_Y, G_X and the discriminators D_Y, D_X form two dual GANs and enables our dual GAN that it could reduce the noise from the SAR data in low SNR $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N}$. We present the SAR data in high SNR as $\mathbf{Y} = \{\mathbf{y}_j\}_{j=1}^{M}$. For the mapping $\mathbf{Y} \to \mathbf{X}$, we express the objective as

$$L_{D_X} = -E_{\mathbf{x}}[\log D_X(\mathbf{x})] - E_{\mathbf{y}}[log(1 - D_X(G_X(\mathbf{y})))] \quad (3)$$

$$L_{G_X} = E_{\mathbf{y}}[log(1 - D_X(G_X(\mathbf{y})))] \quad (4)$$

where G_X tries to generate images $G_X(\mathbf{y})$ that look similar to images from domain $\mathbf{X}$, whereas D_X aims to distinguish between translated samples $G_X(\mathbf{y})$ and real samples $\mathbf{x}$. In (3), the D_X is trying to distinguish $\mathbf{x}$ and fake sample $\hat{\mathbf{x}} = G_X(\mathbf{y})$. In (4), $\mathbf{y}$ is sent to the generator G_X to fool the discriminator D_X. Temporarily, we introduce a similar adversarial loss for the mapping $\mathbf{X} \to \mathbf{Y}$

$$L_{D_Y} = -E_{\mathbf{y}}[\log D_Y(\mathbf{y})] - E_{\mathbf{x}}[log(1 - D_Y(G_Y(\mathbf{x})))] \quad (5)$$

$$L_{G_Y} = E_{\mathbf{x}}[log(1 - D_Y(G_Y(\mathbf{x})))]. \quad (6)$$

The loss function let G_Y map $\mathbf{x}$ to a fake image $\hat{\mathbf{y}}$ identically distributed as target domain $\mathbf{Y}$. However, as discussed above,

the adversarial losses cannot guarantee that the learned G_Y can map an individual input $\mathbf{x}$ to a desired output $\mathbf{y}$, which means the information of $\mathbf{x}$ may have nothing to do with $\hat{\mathbf{y}}$ meanwhile $\mathbf{x}$ and $\hat{\mathbf{y}}$ may be in different categories for our recognition task. According to our needs, we argue that our dual GAN should ensure that the transferred SAR data can retain the target information during the image translation. Therefore, to further improve the performance of the model, we make two improvements to the objective function: add the label loss to the mapping $\mathbf{X} \to \mathbf{Y}$ and introduce the reconstruction loss [33] with L1 distance.

First, we introduce the label constraint to the generator G_Y and discriminator D_Y. Assume $\mathbf{X}$ and $\mathbf{Y}$ have K categories. Then, we rewrite the adversarial losses for the mapping $\mathbf{X} \to \mathbf{Y}$

$$\begin{aligned} L_{D_Y} = & -E_{\mathbf{y},l}[\log p(D_Y(\mathbf{y}) = k)] \\ & -E_{\mathbf{x}}[\log p(D_Y(G_Y(\mathbf{x})) = K+1)] \end{aligned} \tag{7}$$

$$L_{G_Y} = E_{\mathbf{x},l}[\log p(D_Y(G_Y(\mathbf{x})) = k)]. \tag{8}$$

We add a new class label $y = K+1$, which corresponds to the fake data [34]. Compared with (5) and (6), the new adversarial losses can constrain the fake image $G_Y(\mathbf{x})$ to be closer to the distribution of high SNR data of the same class. We visualize the mapping in Fig. 4(b). We can intuitively see that compare with adversarial constraint, the label constraint preserves the correct mapping while suppressing the mapping between different categories.

But we are not satisfied with the current results. Although the category is preserved during the image transformation, there is no guarantee that the target information in the SAR data can also be preserved, i.e., the mapping of $\mathbf{x}_1 \to \mathbf{y}_2$ is still allowed.

The reconstruction loss is expressed as

$$L_{\mathrm{L1}} = E_{\mathbf{x}}[\|G_X(G_Y(\mathbf{x})) - \mathbf{x}\|_1] + E_{\mathbf{y}}[\|G_Y(G_X(\mathbf{y})) - \mathbf{y}\|_1]. \tag{9}$$

We can see from the reconstruction loss that for each SAR data, the generators should ensure the ability to return the image back to the original condition. Further explanation, in order to complete (9), the denoising generator G_Y must only remove the noise information without losing target information in the original image during the process of denoising. Otherwise G_X will not be able to reconstruct $\mathbf{x}$ from $G_Y(\mathbf{x})$. Intuitively, we can see in Fig. 4(c) that the dual GAN with the label constraint and the reconstruction can guarantee that the corresponding relationship of the image target in the image transformation remains unchanged.

Because of the benefits of the reconstruction constraint, it is desirable to add the constraint to our model.

B. Full Objective and Training Procedure

For the traditional supervised GANs [15], [16], the discriminator can also implement the task of classification. However, in our model, the discriminator is trained by the original training SAR data $\mathbf{Y}$, so that it is not suitable for the classification of $G_Y(\mathbf{x})$. Thus, a classifier C is linked to the dual GAN and takes the $G_Y(\mathbf{x})$ as input. Concretely, the classifier Ctries to minimize the loss function

$$L_C = -\sum_{i=1}^{N} l_i \ln C(G_Y(\mathbf{x}_i)). \tag{10}$$

The adversarial losses L_{D_X}, L_{G_X}, L_{D_Y}, and L_{G_Y} can be rewritten to two min–max problems (L_{GAN_X}, L_{GAN_Y}). Then, we give our full objective as follows:

$$\begin{aligned} & \min_{G_X, G_Y, C} \max_{D_X, D_Y} V(G_X, G_Y, D_X, D_Y, C) \\ & \qquad = \lambda_1 L_{\mathrm{GAN}_X} + \lambda_2 L_{\mathrm{GAN}_Y} + \lambda_3 L_{\mathrm{L1}} + \lambda_4 L_C \end{aligned} \tag{11}$$

where λ_1, λ_2, λ_3, and λ_4 control the relative importance of the four objectives. By solving the min–max problem V, the low SNR SAR data can be transformed to the simple domain of high SNR SAR data. And by the constraints of reconstruction and label, the image translation can retain the target information in the original SAR images and provide more reliable images for the classifier.

In our method, we first use the classifier pretrained by the high SNR SAR data to initialize the parameters of classifier C. The training process of our method is shown in Algorithm 1. Since the $G_Y(\mathbf{x})$ is not well enough at the first iterations to fine-tune the classifier, we do not update the parameters of C in the first $T_{\mathrm{burn-in}}$ iterations.

Algorithm 1: Training Step.

Input: $\{(\mathbf{x}_i, l_i)\}_{i=1}^{N}$, $\{(\mathbf{y}_j, l_j)\}_{j=1}^{M}$, T, $T_{\mathrm{burn-in}}$, λ_1, λ_2, λ_3, λ_4
Output: Parameters G_X, G_Y, D_X, D_Y, C;
1: Initialize the parameter G_X, G_Y, D_X, D_Y, C.
2: for $t = 1$ to T do
3: Sample minibatch of n samples $\{(x_1, l_1), \ldots, (x_n, l_n)\}$ from data distribution $p_{\mathrm{data}}(x, l)$.
4: Sample minibatch of m samples $\{(y_1, l_1), \ldots, (y_m, l_m)\}$ from data distribution $p_{\mathrm{data}}(y, l)$.
5: Update the discriminators by maximizing the full objective:
6: $D_X \leftarrow \max_{D_X} V(D_X)$; $D_Y \leftarrow \max_{D_Y} V(D_Y)$;
7: Update the generators by minimizing the full objective:
8: $G_X \leftarrow \min_{G_X} V(G_X)$; $G_Y \leftarrow \min_{G_Y} V(G_Y)$;
9: if $t > T_{\mathrm{burn-in}}$,do
10: $C \leftarrow \min_{C} -\sum_{i=1}^{n} l_i \ln C(G_Y(\mathbf{x}_i))$;
11: end if
12: end for

C. Discussion

In our method, we use a dual GAN and a classifier to build robust SAR ATR method via adversarial learning. By adversarial learning, our method can automatically remove the noise from the SAR data without any prior information. Compared with

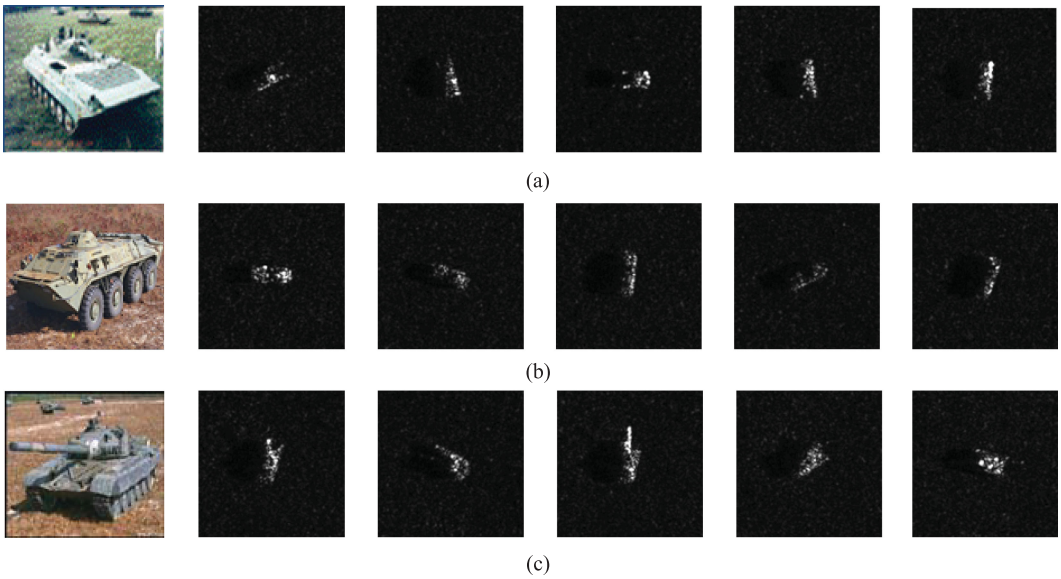


Fig. 5. SAR and optical images of the three-category data. The first column shows the optical images. The remaining columns gives some SAR imagery examples of the three targets. (a) BMP2 target. (b) BTR70 target. (c) T72 target.

the classical denoising algorithms, i.e., OMP and RELAX, our method is more reliable when the noise level prior information is not accurate. And by integrating data denoising, feature extraction and classification steps into a joint learning framework, the denoising process also ensures that the distances between denoised image distributions of different categories is as large as possible.

It is popular to improve the recognition performance of SAR data under the high noise level by data augmentation. In order to improve the performance of the model, the deep learning methods always increase the complexity of the classifier. In our method, we remove the noise information of the SAR data first. In this way, the noise robustness of the model is greatly improved, whereas the classifier is not burdened. Therefore, even though the classifier structure is very simple, our method can still get satisfactory recognition results.

IV. Experimental Results

In this section, we evaluate the performance of our proposed model in the MSTAR dataset and Gotcha dataset from three aspects: denoising effect, fully supervised recognition, and semisupervised recognition.

A. Data Description

1) MSTAR Dataset: In this section, MSTAR dataset with three-category data is tested using our proposed model. The data contain two types of tanks, T72 and BMP2, and a vehicle BTR70. Some SAR and optical imagery examples of these three types of targets are shown in Fig. 5. Each of the target has full aspect coverage from 0° to 360° and different views at 15° and 17° depression angles. The data at depression 17° are used for training and those at depression 15° for test. The number of aspect views available for these targets is listed in Table I. As shown in Table I, we only use the images of BMP2-9563, BTR70-C71, and T72-132 at depression angle 17° as the training data, whereas all images of BMP2-C21, BMP2-9566, BMP2-9563, BTR70-C71, T72-132, T72-S7, and T72-812 at depression angle 15° are used as the test data. MSTAR images available are of around 128 × 128 pixels and cropped to 63 × 63 pixels region of interest in our experiments. The amplitudes of all images differ a lot in the MSTAR dataset.

TABLE I
Type and Number of Training and Test Samples for Three-Category Dataset

Dataset	BMP2			BTR70	T72		
	C21	9566	9563	C71	132	S7	812
Training samples (17°)	233	0	0	233	232	0	0
Test samples (15°)	196	196	195	196	196	191	195

Since it is not convenient to accurately control the SNR of the SAR image after adding multiplicative noise, we consider two ways to add noise to the SAR data, so that we can more fully analyze the effectiveness of our proposed method. First, the noisy SAR data are simulated by adding additive Gaussian noises to the original data according to the predefined SNR [47], [48]. We simulate data at −10 dB, −5 dB, 0 dB, 5 dB, and10 dB for training and −10 dB, −7.5 dB, −5 dB, −2.5 dB, 0 dB, 2.5 dB, 5 dB, 7.5 dB, and 10 dB for test. In order to verify the generalization ability of the methods compared with the training data, there are four more SNRs in the test data: −7.5 dB, −2.5 dB, 2.5 dB, and 7.5 dB. Fig. 6 shows the noised images at different SNRs. With the deterioration of noise contamination, target characteristics are submerged in the noise, which will definitely increase the difficulty of target recognition. Then, in order to prove the effectiveness of the proposed method in the case of more realistic SAR data, we noise the SAR data with multiplicative noise modeled by Gamma distribution, which

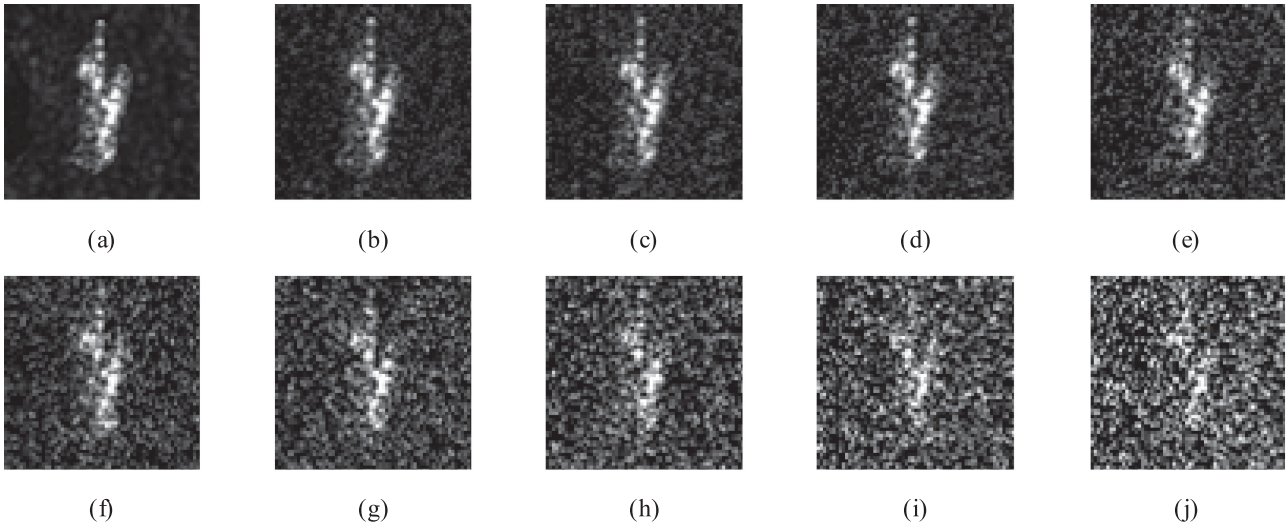


Fig. 6. Noisy images at different SNRs. (a) Original image. (b) 10 dB. (c) 7.5 dB. (d) 5 dB. (e) 2.5 dB. (f) 0 dB. (g) −2.5 dB. (h) −5 dB. (i) −7.5 dB. (j) −10 dB.

TABLE II
TYPE AND NUMBER OF TRAINING AND TEST SAMPLES OF GOTCHA DATASET

Type	Size of vehicle	Training samples	Test samples
ChevyMalibu	4.8 × 1.8 × 1.48	83	20
ToyotaCamry	4.8 × 1.8 × 1.48	83	25
FordTaurusWag	5 × 1.8 × 1.4	116	32
CASEtractor	5 × 2.7 × 3.1	86	28
HysterForkLift	2.2 × 1.1 × 2	60	17
NissanMaxima	4.9 × 1.8 × 1.4	93	31
NissanSentra	4.6 × 1.7 × 1.5	85	24
HyundaiSantaFe	4.7 × 1.9 × 1.7	85	25
ChevyPrizm	4.4 × 1.6 × 1.3	61	19

contains shape parameter α and scale parameter β. We set $\alpha = 10,\ \beta = 1$ for training data and $\alpha = 5,\ \beta = 5$ for test data.

2) Gotcha Dataset: Gotcha dataset [51] is called Gotcha volumetric SAR data set and collected by a high-resolution spotlight SAR with X-band. The Gotcha data are SAR image of the whole parking lot in an omnidirectional manner at eight different pitch angles. The imaging scene contains nine types of civil vehicles, namely Chevy Malibu, toyotacamry, Ford Taurus wag, casetracker, hysterforklift, Nissan maxima, Nissan centra, Hyundai isantafe, and Chevy prizm. We cut out the single target image from the original SAR image and divided the training and test sets. The type and number of training and test samples of Gotcha dataset is shown in Table II. We noise the Gotcha SAR data with multiplicative noise modeled by Gamma distribution, which contains shape parameter α and scale parameter β. We set $\alpha = 10,\ \beta = 1$ for training data and $\alpha = 5,\ \beta = 5$ for test data.

B. Network Architectures and Parameter Settings on MSTAR Dataset

In our experiments, we use a six-layer network for the generators with two convolutional layers as encoding blocks, two convolutional layers as residual blocks, and two deconvolutional layers as decoding blocks. The corresponding discriminator D_X is a four-layer two-class DCGAN [16] and D_Y is a four-layer $K+1$-class DCGAN. For the classifier, to prove that our method can avoid increasing complexity of classifier, we use LeNet as our classifier. we use two classical networks, LeNet and ResNet-18 [50], as our classifier. Fig. 7 shows the network architecture of our method. Since the structure of generator G_X and G_Y, as well as the structure of discriminator D_X and D_Y, are the same, we only give the structure diagram of G_Y, D_Y, and classifier C (LeNet) in the structure diagram. The number on the feature map represents the size of the feature map, the number below the feature map represents the number of channels, and "FC" represents the fully connection.

The parameter setting in our experiments is $\lambda_1 = \lambda_3 = 0.5$, $\lambda_2 = \lambda_4 = 1, T = 50\,000, T_{\mathrm{burn-in}} = 5000$. We use minibatch SGD and apply the Adam solver [29], with learning rate of 0.0001 and a batch size of 64.

C. Experimental Results on MSTAR Dataset With Gaussian Noise

1) Visualization of Denoising Quality: In this section, we first analyze the denoising quality of our algorithm. Nine targets with different SNRs from test samples are selected for the denoising visualization. As shown in Fig. 8, the left six columns are the denoising results of BMP2 targets, the middle six columns are the results of BTR70 targets, and the right six columns are the results of T72 targets. Fig.8 clearly shows that our method can indeed transform the low SNR SAR data to high SNR SAR data with high quality. The noise information is obviously removed in the denoised images, and the target information is completely retained. Furthermore, it can be found that the denoising effect of our method has little difference in SAR data at different SNRs.

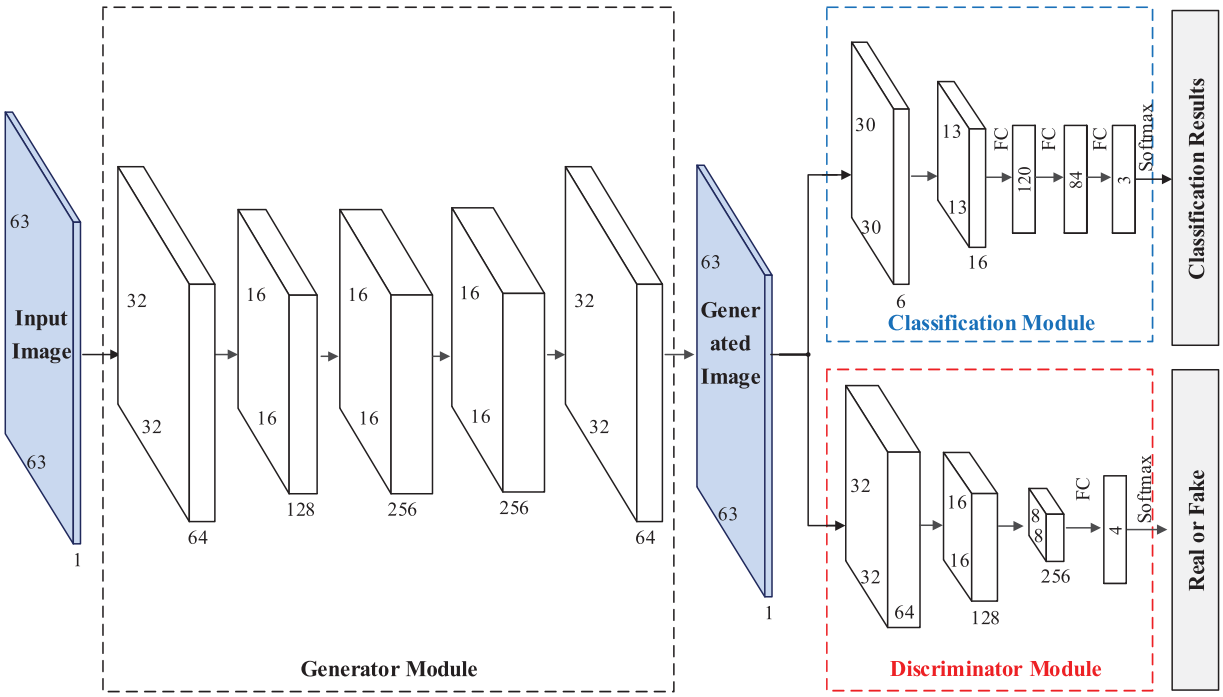


Fig. 7. Network architecture of generator G_Y, discriminator D_Y, and classification C (LeNet).

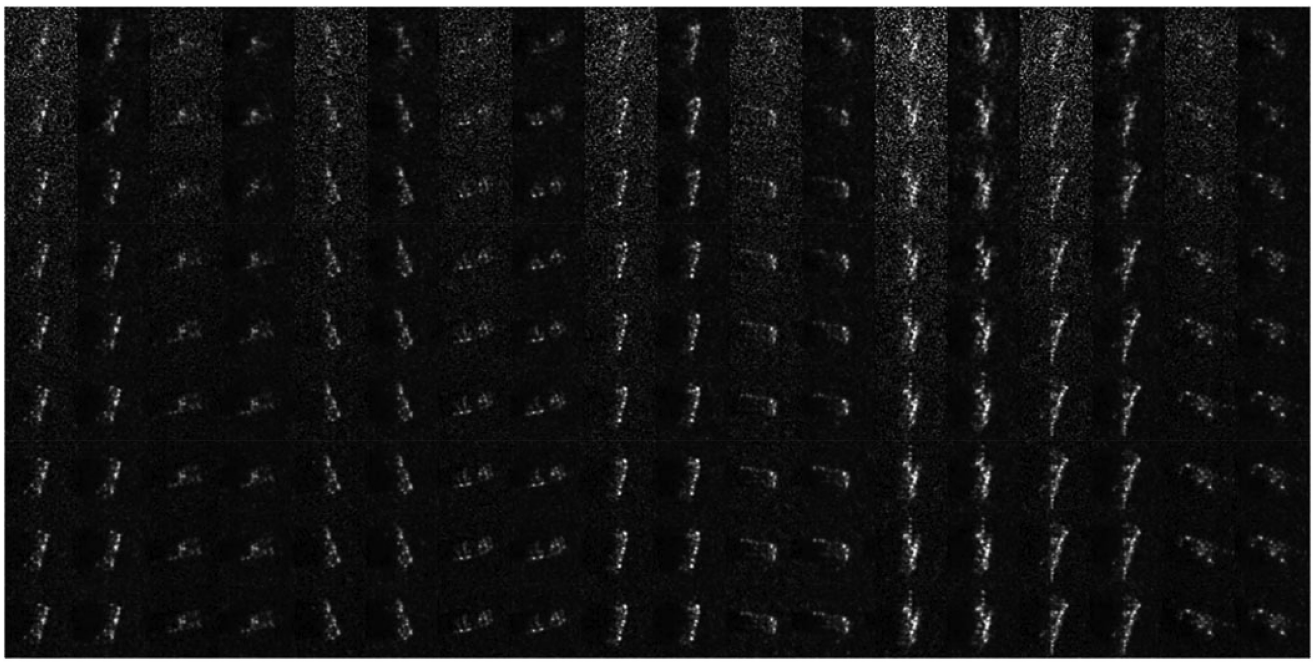


Fig. 8. Denoising results of our method on MSTAR dataset with Gaussian noise. Odd-numbered columns: the original SAR data with low SNRs (from top to bottom: −10 dB, −7.5 dB, −5 dB, −2.5 dB, 0 dB, 2.5 dB, 5 dB, 7.5 dB, and 10 dB). Even-numbered columns: transformed SAR data with high SNR generated from the corresponding one. Observe that noise information has been removed from the SAR data obviously. On the other hand, target information that is classification related has been well preserved.

For the robust SAR automatic recognition task, the denoising capability of our algorithm can undoubtedly guarantee the noise robustness.

Second, we also compared our algorithm with the traditional denoising algorithm based on the OMP method and refined Lee filter. As an example, we pick up a test target to visualize the performance of the two methods. Fig. 9 shows the denoising results of the target data at different SNRs by our algorithm, OMP method, and refined Lee filter respectively. Since the OMP method is essentially to find the strong scattering points in the SAR image, it loses the ability of denoising under low SNRs. And by observing the results of refined Lee filter, we

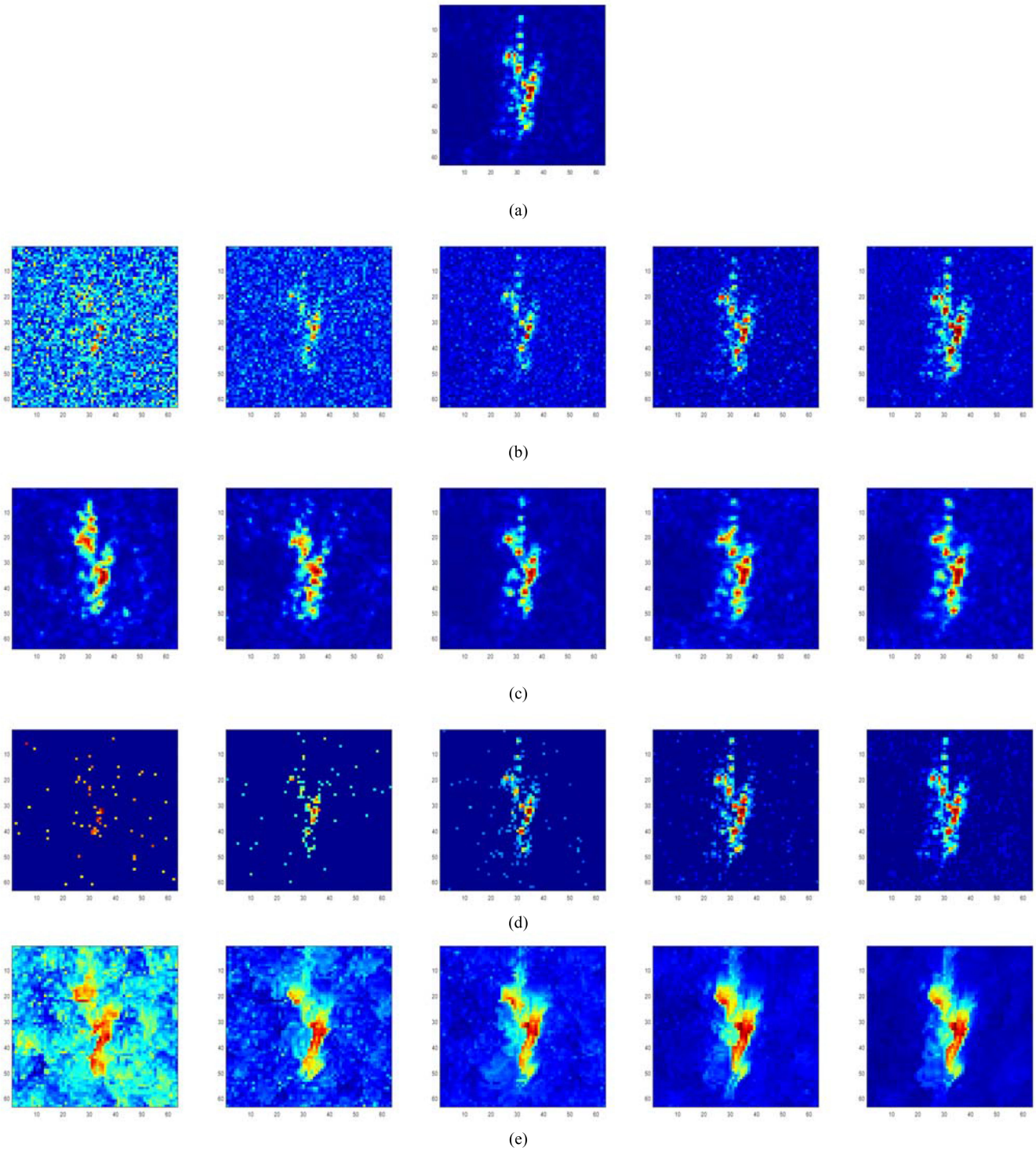


Fig. 9. Denoising results of our algorithm and the OMP method at different SNRs with Gaussian noise on MSTAR dataset. From left to right are the results with an SNR of −10 dB, −5 dB, 0 dB, 5 dB, and 10 dB. (a) Original test SAR image at high SNR. (b) Noised SAR images. (c) Results of our algorithm. (d) Results of the OMP method. (e) Results of the refined Lee filter.

find that the target information in the denoised SAR image is obviously modified. By comparing the denoising results with the original SAR data, we can find that our algorithm has a better denoising effect and more stable performance under different SNRs than both the OMP method and the refined Lee filter. Moreover, what is surprising is that although the image has been seriously damaged by noise at 10 dB, our algorithm can still get a satisfactory denoising result, which will be of great help to our robust SAR ATR research.

2) Analysis of Model Convergence: The convergence is an important problem in the training of GAN, which is affected by many factors and difficult to adjust parameters. Failure to

TABLE III
RESULTS OF ABLATION EXPERIMENTS ON MSTAR DATASET WITH GAUSSIAN NOISE

Method	-10dB	-7.5dB	-5dB	-2.5dB	-0dB	2.5dB	5dB	7.5dB	10dB
Baseline	0.7377	0.8190	0.8645	0.8872	0.9062	0.9187	0.9290	0.9289	0.9355
Label	0.7805	0.8300	0.8854	0.8962	0.9024	0.9047	0.9051	0.9010	0.9145
Label + Reconstruction	0.8006	0.8514	0.8921	0.9091	0.9176	0.9297	0.9464	0.9423	0.9482
Label + Reconstruction + Joint learning	**0.8190**	**0.8684**	**0.9082**	**0.9151**	**0.9223**	**0.9339**	**0.9487**	**0.9467**	**0.9494**

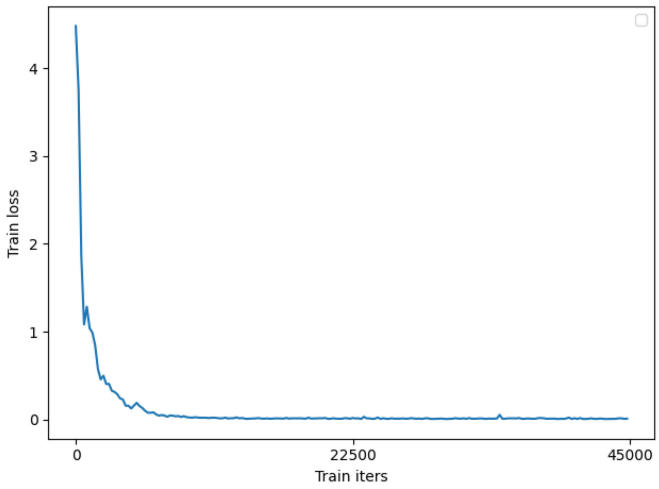


Fig. 10. Convergence curve of the classifier's loss during training on MSTAR dataset with Gaussian noise.

converge will lead to the mode collapse of the GAN model. The essence of the mode collapse problem is that GAN cannot reach the optimal Nash equilibrium state through training [14]. Since the generator and discriminator are trained against each other, the loss function of the traditional GAN is not monotonically decreasing in convergence.

In our method, we can observe whether the loss of classifier decreases and converges. The convergence curve of the classifier's loss during training is given in Fig. 10. It can be seen that the loss curve is convergent with the increase in iteration, which proves that the training of our model is convergent. Moreover, by observing the generated images in Fig. 8, our method can indeed transform the low SNR SAR data to high SNR SAR data with high quality, which also proves that the dual GAN has been trained well.

3) Ablation Study: To verify the effectiveness of our method, we first conducted ablation study on three modules of the algorithm: reconstruction constraint, label constraint, and joint learning. In order to verify the effectiveness of the various modules of the algorithm, we perform ablation experiments in this section. In particular, the baseline indicates the model LeNet. The experimental results are shown in Table III.

As can be seen from Table III, reconstruction constraint, label constraint, and joint learning are all effective at the low SNRs. First, the label constraint makes the model has the ability to denoise trough GAN, which makes the model recognition performance greatly improved at the low SNRs (accuracy of the model on the test set improves from 73.77% to 78.05% at -10 dB). However, since the label constraint alone cannot ensure that the target information in the SAR images is not lost, the model performance is reduced at high SNRs. When reconstruction constraint is added to the model, the accuracies at all SNRs have been significantly improved, which show that reconstruction constraint brings better recognition performance. The dual GAN retains the target information more completely and accurately in the process of data denoising, so that the classifier can more accurately identify the target. Similarly, by jointly training the dual GAN and classifier, the recognition performance of the model can also be improved. By jointly learning the three steps of denoising, feature extraction and classification, each step can make the data as separable as possible when processing the SAR data. In the ablation experiments, under the "Label + Reconstruction + Joint learning" experimental setting, the best recognition accuracies are achieved at all SNRs.

TABLE IV
TOTAL RECOGNITION ACCURACIES OF FULLY SUPERVISED LEARNING OBTAINED BY DIFFERENT SAR ATR METHODS ON MSTAR DATASET WITH GAUSSIAN NOISE

Methods	Accuracy
SVM	0.8007
PCA	0.8122
LeNet	0.8807
ResNet-18	0.9004
ResNet-34	0.9045
VGGNet	0.9088
OMP+ ResNet-18	0.8867
Refined Lee+ResNet-18	0.9035
Proposed method (LeNet)	0.9140
Proposed method (ResNet-18)	**0.9268**

4) Recognition Performance Comparison: In this section, we compare the proposed method with other representative SAR ATR , such as SVM, PCA [30], LeNet [31], ResNet-18 [49], ResNet-34 [49], VGGNet [32], OMP+ResNet-18, and refined Lee+ResNet-18. For the PCA method that can extract noise robust features, we use the original high SNR data as the training data, which is the common way for the PCA method [8]; for the deep learning methods and SVM method, we use the expanded data of different SNRs (-10 dB, -5 dB, 0 dB, 5 dB, 10 dB) as the training data.

First, we use all the training data for training and then test the recognition performance on all the expanded test data. As shown in Table IV, our method achieves the highest recognition accuracy in MSTAR dataset. Compared with SVM, our method

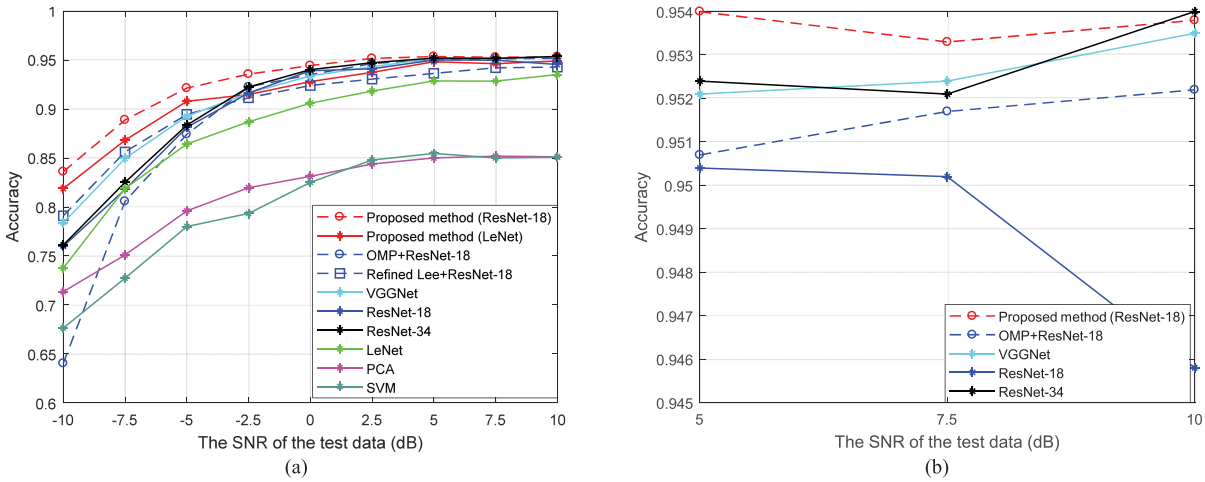


Fig. 11. Variation of the recognition performance with SNR of fully supervised learning for different methods on MSTAR dataset with Gaussian noise. (a) Results of all methods under all SNR. (b) Details of several methods with high recognition accuracies in (a) under the SNRs from 0 to 10 dB.

with LeNet is 11.33% higher, which indicate that our method extracts more separable features than the original SAR data. And compared with traditional robust recognition method PCA, the recognition accuracy of our method with LeNet also has an advantage of 10.18% and 7.08%. The recognition accuracy of our method with LeNet is 3.33% higher than that of LeNet due to the ability of the dual GAN reducing noise information from SAR data.

Since our algorithm has a better denoising effect and more stable performance under different SNRs than the OMP method and refined Lee filter, our method with LeNet is 2.73% higher than that of OMP+ResNet-18 and 1.05% higher than that of refined Lee+ResNet-18. Surprisingly, our method with LeNet outperforms ResNet-18 by 1.36% in recognition performance. However, the experiment results show that our method with LeNet has little improvement in recognition accuracy compared with ResNet-34 and VGGNet (accuracy rate improved by 0.95% and 0.52%). The reason for this result is that the LeNet is much simpler than ResNet-34 and VGGNet, which greatly limits the classification performance of the model. When we replace the LeNet with ResNet-18 in our method, the recognition accuracy of our method with ResNet-18 is 2.23% higher than that of ResNet-34 and 1.80% higher than that of VGGNet. Moreover, we represent the test results with different SNRs in Fig. 11 to make further comparison. It can be clearly observed that our method with ResNet-18 achieves the highest recognition accuracy in all SNRs except for 10 dB, in which the recognition accuracy of our method is only 0.02% lower than that of ResNet-34.

Second, we test the classification performance of our method in the case of semisupervised learning. And the results of VGGNet under fully supervised learning are compared in Table V. The different columns in Table V represent the percentages of the labeled training data used in model learning. In experiments with low SNR and few labeled training dataset, since our proposed method can use both labeled and unlabeled data for model learning to ensure the denoising effect, it can be observed that our proposed method has a significant improvement in recognition

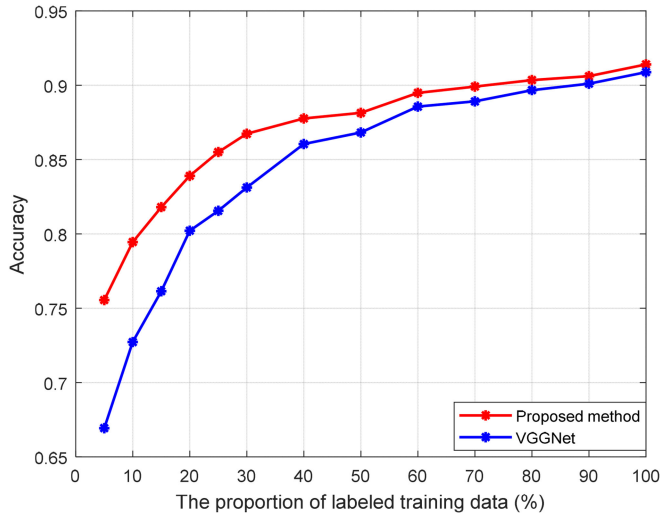


Fig. 12. Variation of the recognition performance with labeled training dataset of different methods on MSTAR dataset with Gaussian noise.

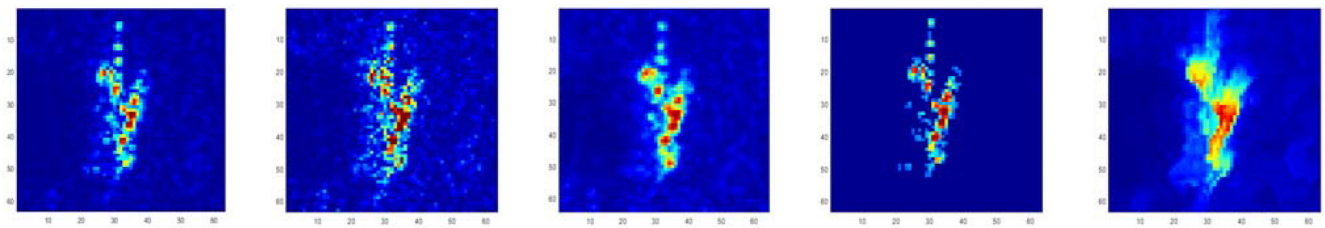
performance compared with VGGNet. More intuitively, we draw the recognition results of all test data recognition accuracies under different proportions of labeled training data in Fig. 12. As the semisupervised learning ability can utilize the information of unlabeled data, the advantage of our proposed method is more obvious with the reduction of the labeled training dataset. In the case of a large number of labeled training samples and high SNR situations, the VGGNet achieves better recognition accuracies because of that the VGGNet is much more complex than our classification network LeNet. However, our method still obtains the results that are competitive with the VGGNet.

D. Experimental Results With Multiplicative Noise

In order to prove the effectiveness of the proposed method in the case of more realistic SAR data, we test our method on SAR data with multiplicative noise modeled by Gamma distribution in this section.

TABLE V
RECOGNITION ACCURACIES OF SEMISUPERVISED LEARNING ON MSTAR DATASET WITH GAUSSIAN NOISE

SNRs	Methods	5	10	15	20	25	30	40	50	60	70	80	90	100
-10dB	VGGNet	0.5897	0.6330	0.6725	0.6945	0.6974	0.7150	0.7216	0.7377	0.7429	0.7377	0.7656	0.7770	0.7839
	Proposed method	**0.6878**	**0.7119**	**0.7558**	**0.7599**	**0.7608**	**0.7695**	**0.7746**	**0.7868**	**0.7892**	**0.7938**	**0.7982**	**0.8004**	**0.8190**
-7.5dB	VGGNet	0.6286	0.6967	0.7231	0.7480	0.7451	0.7736	0.7963	0.8103	0.8227	0.8300	0.8374	0.8344	0.8498
	Proposed method	**0.7249**	**0.7712**	**0.7855**	**0.7961**	**0.8173**	**0.8166**	**0.8206**	**0.8306**	**0.8375**	**0.8474**	**0.8508**	**0.8523**	**0.8684**
-5dB	VGGNet	0.6623	0.7055	0.7641	0.7817	0.8007	0.8195	0.8322	0.8469	0.8689	0.8733	0.8733	0.8806	0.8930
	Proposed method	**0.7504**	**0.7844**	**0.8174**	**0.8291**	**0.8453**	**0.8518**	**0.8652**	**0.8760**	**0.8852**	**0.8893**	**0.8959**	**0.9040**	**0.9082**
-2.5dB	VGGNet	0.6791	0.7363	0.7766	0.8154	0.8374	0.8440	0.8696	0.8835	0.8952	0.9055	0.9077	0.9122	0.9165
	Proposed method	**0.7754**	**0.8008**	**0.8290**	**0.8435**	**0.8619**	**0.8727**	**0.8939**	**0.9038**	**0.9000**	**0.9089**	**0.9150**	**0.9141**	**0.9191**
0dB	VGGNet	0.6806	0.7377	0.7795	0.8212	0.8381	0.8535	0.8872	0.8901	**0.9179**	0.9187	**0.9209**	**0.9289**	**0.9342**
	Proposed method	**0.7693**	**0.8046**	**0.8312**	**0.8555**	**0.8728**	**0.8870**	**0.8931**	**0.8992**	0.9160	**0.9206**	0.9204	0.9218	0.9283
2.5dB	VGGNet	0.6952	0.7524	0.7810	0.8293	0.8498	0.8627	0.8960	0.8982	0.9245	0.9253	0.9297	**0.9317**	**0.9436**
	Proposed method	**0.7778**	**0.8156**	**0.8364**	**0.8565**	**0.8825**	**0.8967**	**0.8977**	**0.9023**	**0.9290**	**0.9269**	**0.9313**	0.9316	0.9379
5dB	VGGNet	0.6960	0.7612	0.7853	0.8410	0.8564	0.8659	0.9077	**0.9084**	0.9289	**0.9319**	**0.9451**	**0.9473**	**0.9521**
	Proposed method	**0.7738**	**0.8187**	**0.8363**	**0.8710**	**0.8831**	**0.9011**	**0.9112**	0.9014	**0.9299**	0.9302	0.9366	0.9433	0.9487
7.5dB	VGGNet	0.6967	0.7582	0.7824	0.8418	0.8542	0.8659	0.9150	**0.9194**	**0.9326**	**0.9385**	**0.9421**	**0.9466**	**0.9524**
	Proposed method	**0.7711**	**0.8188**	**0.8301**	**0.8686**	**0.8858**	**0.9073**	**0.9172**	0.9187	0.9314	0.9366	0.9392	0.9411	0.9467
10dB	VGGNet	0.6952	0.7641	0.7883	0.8462	0.8608	0.8803	0.9179	**0.9194**	**0.9363**	**0.9407**	**0.9487**	**0.9503**	**0.9535**
	Proposed method	**0.7687**	**0.8244**	**0.8403**	**0.8709**	**0.8858**	**0.9030**	**0.9252**	0.9135	0.9353	0.9382	0.9436	0.9458	0.9494

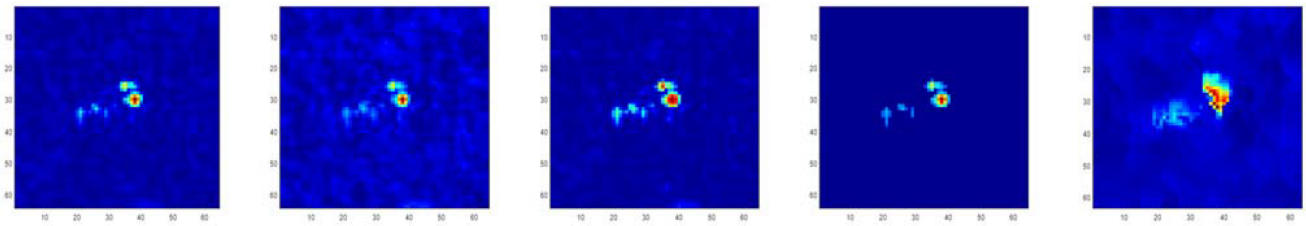


(a) (b) (c) (d) (e)

Fig. 13. Denoising results of our algorithm and the OMP method on MSTAR dataset with multiplicative noise. (a) Original test SAR image at high SNR. (b) Noised SAR image. (c) Denoising result of our method. (d) Denoising result of the OMP method. (e) Denoising result of the refined Lee filter.

1) MSTAR Dataset: We first analyze the denoising quality of our method. Fig. 13 shows the denoising results of the SAR data with multiplicative noise by our algorithm. It can be clearly observed that compared with the original noised SAR image, the noise interference in the denoised image is greatly reduced, and the target is more significant. By comparing with the denoising results of OMP method in Fig. 13(d) and the denoising results of refined Lee filter in Fig. 13(e), we can find that our algorithm has a better denoising effect than the OMP method and refined Lee filter.

Second, as presented in the previous section, we compare the proposed method with other representative SAR ATR methods. The recognition results are shown in Table VI. The recognition accuracy of our method with LeNet is obviously higher than SVM and PCA. And compared with LeNet, the recognition accuracy of our method with LeNet is 2.59% higher due to the ability of the dual GAN reducing noise information from SAR data. However, due to the simplicity of the LeNet, our method with LeNet can only get comparable results as ResNet-18, ResNet-34 and VGGNet. When we replace the LeNet with ResNet-18 in our method, the recognition accuracy of our method with ResNet-18 is 2.22% higher than that of ResNet-18, 1.50% higher than that of ResNet-34, and 1.28% higher than that of VGGNet. Since our algorithm has a better denoising effect and more stable performance under different SNRs than the OMP method and refined Lee filter, our method with ResNet-18 is 0.83% higher than that of OMP+ResNet-18 and 3.63% higher than that of refined Lee+ResNet-18.



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

Fig. 14. Denoising results of our algorithm and the OMP method on Gotcha dataset with multiplicative noise. (a) Original test SAR image at high SNR. (b) Noised SAR image. (c) Denoising result of our method. (d) Denoising result of the OMP method. (e) Denoising result of the refined Lee filter.

TABLE VI
RECOGNITION ACCURACIES OF DIFFERENT SAR ATR METHODS ON MULTIPLICATIVE NOISED MSTAR DATASET

Methods	Accuracy
SVM	0.8220
PCA	0.8212
LeNet	0.8974
ResNet-18	0.9165
ResNet-34	0.9237
VGGNet	0.9259
OMP+ ResNet-18	0.9305
Refined Lee+ ResNet-18	0.9024
Proposed method (LeNet)	0.9233
Proposed method (ResNet-18)	**0.9387**

TABLE VII
RECOGNITION ACCURACIES OF DIFFERENT SAR ATR METHODS ON MULTIPLICATIVE NOISED GOTCHA DATASET

Methods	Accuracy
SVM	0.7714
PCA	0.7691
LeNet	0.8630
ResNet-18	0.8718
ResNet-34	0.8763
VGGNet	0.8806
OMP+ResNet-18	0.8817
Refined Lee+ResNet-18	0.8547
Proposed method (LeNet)	0.8831
Proposed method (ResNet-18)	**0.8922**

2) Gotcha Dataset: We first analyze the denoising quality of our method. Fig. 14 shows the denoising results of the SAR data with multiplicative noise by our algorithm. It can be clearly observed that compared with the original noised SAR image, the noise interference in the denoised image is greatly reduced, and the target is more significant. By comparing with the denoising results of the OMP method in Fig. 14(d) and the denoising results of the refined Lee filter in Fig. 14(e), we can find that our algorithm has a better denoising effect than the OMP method and the refined Lee filter.

Second, as presented in the previous section, we compare the proposed method with other representative SAR ATR methods. The recognition results are shown in Table VII. The recognition accuracies of our method with ResNet-18 and LeNet are the highest accuracies among all the methods. The recognition accuracies of our method with ResNet-18 is obviously higher than SVM and PCA. And compared with LeNet, the recognition accuracy of our method with LeNet is 2.01% higher due to the ability of the dual GAN reducing noise information from SAR data. When we replace the LeNet with ResNet-18 in our method, the recognition accuracy of our method with ResNet-18 is 2.04% higher than that of ResNet-18, 1.59% higher than that of ResNet-34, and 1.16% higher than that of VGGNet. Since our algorithm has a better denoising effect and more stable performance under different SNRs than the OMP method and the refined Lee filter, our method with ResNet-18 is 1.05% higher than that of OMP+ResNet-18 and 3.75% higher than that of refined Lee+ResNet-18.

V. CONCLUSION

To integrate data denoising and feature extraction and classification into a unified framework for joint learning, we proposed a robust SAR ATR method via adversarial learning. Different from traditional denoising methods, most of which rely on prior knowledge of noise level, our method uses the idea of cross-domain to remove noise information, which is more reliable when the noise level prior information is not accurate. By using reconstruction constraint and label constraint, the denoised data are more reliable for recognition than the original SAR data. In the experiments, we use MSTAR dataset and Gotcha dataset to validate our method and obtain gratifying results in both fully supervised learning and semisupervised learning.

REFERENCES

[1] C. Tison, N. Pourthie, and J.-C. Souyris, "Target recognition in SAR images with support vector machines (SVM)," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Barcelona, Spain, Jul. 2007, pp. 456–459.
[2] A. Sara, "Extracting impervious surfaces from full polarimetric SAR images in different urban areas," *Int. J. Remote Sens.*, vol. 41, no. 12, pp. 4642–4661, Jun. 2020.
[3] H. S. Srivastava, P. Patel, and Y. Sharma, "Large-area soil moisture estimation using multi-incidence-angle RADARSAT-1 SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2528–2535, Aug. 2009.
[4] S. Deng, L. Du, C. Li, J. Ding, and H. Liu, "SAR automatic target recognition based on Euclidean distance restricted autoencoder," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3323–3333, Jul. 2017.
[5] K. Tang, X. Sun, H. Sun, and H. Wang, "A geometrical-based simulator for target recognition in high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 958–962, Sep. 2012.

[6] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 4, no. 2, pp. 157–166, Mar. 1982.

[7] B. Ding, G. Wen, C. Ma, and X. Yang, "An efficient and robust framework for SAR target recognition by hierarchically fusing global and local features," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5983–5995, Dec. 2018.

[8] G. Xu, X. Wang, Y. Huang, L. Cai, and Z. Jiang, "Joint multi-channel sparse method of robust PCA for SAR ground moving target image indication," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1709–1712.

[9] J. Ding, B. Chen, H. W. Liu, and M. Y. 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.

[10] J. T. Lv and Y. Liu, "Data augmentation based on attributed scattering centers to train robust CNN for SAR ATR," *IEEE Access*, vol. 7, pp. 25459–25473.

[11] A. Baraldi and F. Parmiggiani, "A refined gamma-map-SAR speckle filter with improved geometrical adaptivity," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 5, pp. 1245–1257, Sep. 1995.

[12] B. Yong, Y. Zhou, and C. Li, "Improved RELAX-PHAF autofocus algorithm of hybrid SAR," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2005, pp. 4659–4661.

[13] A. Hossain and I. Elshafiey, "Compressed sensing approach for UWB-OFDM SAR imaging using greedy algorithms," *J. Optoelectron. Adv. Mater.*, vol. 17, no. 5/6, pp. 785–789, May/Jun. 2015.

[14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, X. Bing, D. Warde-Farley, and S. Ozair, "Generative adversarial nets," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[15] E. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a Laplacian pyramid of adversarial networks," in *Proc. 28th Conf. Neural Inf. Process. Syst.*, 2015, vol. 1, pp. 1486–1494.

[16] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv:1511.06434*.

[17] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 2962–2971.

[18] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 1180–1189.

[19] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *Proc. 33rd Int. Conf. Int. Conf. Mach. Learn.*, Jun. 2016, vol. 48, pp. 1060–1069.

[20] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, vol. 2, pp. 1033–1038.

[21] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, "Image analogies," in *Proc. SIGGRAPH*, 2001, pp. 327–340.

[22] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2014.

[23] M. Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, 2016, pp. 469–477.

[24] Y. S. Chen, Y. Wang, Y. F. Gu, X. He, P. Ghamisi, and X. P. Jia, "Deep learning ensemble for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1882–1897, Jun. 2019.

[25] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[26] S. Z. Chen, H. P. Chen, 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.

[27] R. Rosales, K. Achan, and B. Frey, "Unsupervised image translation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2003, pp. 472–478.

[28] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.

[29] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *,arXiv:1412.6980*.

[30] Y. Wang, P. Han, X. Lu, R. Wu, and J. Huang, "The performance comparison of Adaboost and SVM applied to SAR ATR," in *Proc. CIE Int. Conf. Radar*, 2006, pp. 1–4.

[31] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*.

[33] T. Z. Krähenbühl, P. M. Aubry, Q. Huang, and A. A. Efros, "Learning dense correspondence via 3D-guided cycle consistency," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 117–126.

[34] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. 30th Conf. Neural Inf. Process. Syst.*, 2016, pp. 2234–2242.

[35] C. Jia, X. B. Liang, and Y. Zhang, "Cross-domain NER using cross-domain language modeling," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, Florence, Italy, Jul./Aug. 2019, pp. 2464–2474.

[36] C. Persello and L. Bruzzone, "Active and semisupervised learning for the classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6937–6956, Nov. 2014.

[37] Z. Pan, X. Qiu, Z. Huang, and B. Lei, "Airplane recognition in TerraSAR-X images via scatter cluster extraction and reweighted sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 112–116, Jan. 2017.

[38] H. Y. Liu, F. H. Shang, S. Y. Yang, M. G. Gong, T. W. Zhu, and L. C. Jiao, "Sparse manifold-regularized neural networks for polarimetric SAR terrain classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 8, pp. 3007–3016, Aug. 2020.

[39] W. Yan, W. Yang, H. Sun, and M. S. Liao, "Unsupervised classification of PolInSAR data based on Shannon entropy characterization with iterative optimization," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 4, no. 4, pp. 949–959, Dec. 2011.

[40] F. Biondi, "Multi-chromatic analysis polarimetric interferometric synthetic aperture radar (MCA-PolInSAR) for urban classification," *Int. J. Remote Sens.*, vol. 40, no. 10, pp. 3721–3750.

[41] G. Yaroslav, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030.

[42] S. Kuniaki, "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018. pp. 3723–3732.

[43] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017. pp. 7167–7176.

[44] T. Luan, S. Kihyuk, and Y. Xiang, "Gotta adapt'em all: Joint pixel and feature-level domain adaptation for recognition in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2672–2681.

[45] W. D. Xu, H. Z. Sun, C. Deng, and Y. Tan, "Variational autoencoder for semi-supervised text classification," in *Proc. 31st AAAI Conf. Artif. Intell.*, Feb. 2017, pp. 3358–3364.

[46] D. H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. ICML Workshop.*, pp. 1–6.

[47] B. Y. Ding, G. J. Wen, C. H. Ma, and X. L. Yang, "An efficient and robust framework for SAR target recognition by hierarchically fusing global and local features," *IEEE Trans. Image Process*, vol. 27, no. 12, pp. 5983–5995, Dec. 2018.

[48] C. J. Jiang and Y. Zhou, "Hierarchical fusion of convolutional neural networks and attributed scattering centers with application to robust SAR ATR," *Remote Sens.*, vol. 10, Art. no. 819.

[49] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2015, pp. 770–778.

[50] Y. Aiyeola Sikiru, L. Rongke, and S. Wu, "SAR image despeckling using refined Lee filter," in *Proc. 7th Int. Conf. Intell. Hum.-Mach. Syst. Cybern.*, Hangzhou, China, Aug, 2015, pp. 260–265.

[51] "Gotcha volumetric SAR data set," [Online]. Available: https://www.sdms.afrl.af.mil/datasets/gotcha.index.php