A Target SAR Image Expansion Method Based on Conditional Wasserstein Deep Convolutional GAN for Automatic Target Recognition

Jikai Qin, *Student Member, IEEE*, Zheng Liu, Lei Ran, Rong Xie, Junkui Tang, and Zekun Guo, *Student Member, IEEE*

***Abstract*—For the automatic target recognition (ATR) based on synthetic aperture radar (SAR) images, enough training data are required to effectively characterize target features and obtain good recognition performance. However, in practical applications, it is difficult to collect sufficient training data. To tackle the limitation, a novel end-to-end expansion method, called conditional Wasserstein deep convolutional generative adversarial network with gradient penalty (CWDCGAN), is proposed to achieve SAR image expansion with specified category. To be specific, the CWDCGAN innovatively designed a generative adversarial network architecture based on convolutional and deconvolution networks to improve the quality of generated images. At the same time, conditional information is introduced to control the categories of generated images, and Wasserstein distance and gradient penalty are used to modify the loss function, which makes the network training more stable. Besides, feature extraction and classifier design in a typical ATR system often rely heavily on subjective expert knowledge, which seriously affects its generalization performance. Therefore, a joint recognition method of Resnet18 and support vector machine (Renset18-SVM) is adopted to improve the generalization capacity and the recognition performance. Experimental results with public measured data show that the CWDCGAN can generate higher quality SAR images, and by feeding expanded data to Renset18-SVM, the recognition accuracy is improved under different proportions of training samples.**

***Index Terms*—Automatic target recognition (ATR), generative adversarial network (GAN), resnet18, SAR image expansion, support vector machine, synthetic aperture radar (SAR).**

I. Introduction

THE automatic target recognition (ATR) technology is an important subject in the radar surveillance field, which is able to automatically determine the attributes and categories of targets after feature extraction [1]. This technology is mainly applicated on top of radar echo signals, such as 1-D high-resolution range profile (HRRP) [2], [3], [4], [5], [6], [7], [8], [9] and 2-D synthetic aperture radar (SAR) images [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20]. Compared with HRRP, SAR images provide the target scattering information along with both the range and cross range, thus improving the recognition performance dramatically. Therefore, more and more attention has been paid to the SAR-based ATR technology, which can be divided into two stages as follows from the aspect of ATR model training.

In the first stage, the main task is to collect complete target SAR image data with labels. Due to the scattering characteristic, SAR images of targets exhibit certain differences in different azimuth and elevation angles. In addition, collecting the echo information of non-cooperative targets is so hard that the demand of ATR technology for data volume cannot be met. Some researchers try to solve this problem. Feng et al. [21] combined integration parts model and deep learning algorithm, and achieved SAR target classification under limited training data . Zhang et al. [22] proposed a method-based domain knowledge-powered two-stream deep network for few-shot SAR vehicle recognition. In addition to these efforts, many approaches to sample limitation problems belong to SAR image augmentation. In the traditional SAR image expansion methods, SAR image simulation is accomplished via image simulators or raw signal simulators [23], [24], [25], [26]. These simulation ways are capable of acquiring the target SAR images we need, but there are also several shortcomings. First, the quality of the simulation image is related to the geometric precision of the computer-aided design model, especially susceptible to the complex surface conditions of the target. Second, the electromagnetic approximation used in traditional SAR image simulation only applies to electrically large objects. The smaller the target structure is, the lower the computational accuracy becomes. Therefore, other advanced SAR image data expansion methods need to be explored. In recent years, generative adversarial network (GAN) [27] has attracted extensive attention, which generates samples with the same distribution as the real ones through the adversarial games of discriminators and generators. Compared with traditional expansion methods, the GAN method shows incomparable advantages. For example, it does not require any prior target information and can directly generate images based on real samples. However, there are some problems as well, such as the low quality, the uncontrollability of the generated image, and the instability of the network training. Many studies

Manuscript received 6 July 2022; revised 3 August 2022; accepted 12 August 2022. Date of publication 16 August 2022; date of current version 5 September 2022. This was supported by the National Natural Science Foundation of China under Grant 62001346. *(Corresponding author: Zheng Liu.)*

The authors are with the National Lab of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: jkqin@stu.xidian.edu.cn; lz@xidian.edu.cn; rl@xidian.edu.cn; rxie@mail.xidian.edu.cn; tjk_0622@126.com; zekunguo_xidian@163.com).

Digital Object Identifier 10.1109/JSTARS.2022.3199091

try to overcome these problems. Deep convolutional generative adversarial networks (DCGAN) introduce convolution structure into the model and use multiple discriminators, which can improve the quality of generated SAR images by optimizing the GAN structure [28]. Even so, the generated images still have strong uncertainties and the network model is too complex. Mirza et al. [29] used the conditional GAN (CGAN) to generate images of the specified category. However, because of the fully connected structure and the upsampling operation, the generated SAR images are sometimes full of noise and difficult to be understood. Yao [30] attempted to utilize multiple DCGANs to generate multicategory high-quality SAR images, but resulting in a very redundant network structure and an increased workload. Besides, extensive research works have validated that gradient vanishing commonly occurs during the GAN model training, which leads to the instability of network training [31]. Noting the phenomenon, Arjovsky et al. [32] proposed the WGAN. It not only solves the instability problem of the model training, but also provides a reliable indicator of training progress. Regrettably, in some cases, gradient explosion may be triggered. Gulrajani et al. [33] exploited the gradient penalty (GP) on the WGAN to handle this problem, but the generated image was still uncontrollable. Cui et al. [34] adopted the innovative method to generate SAR images from the perspective of the loss function. However, the quality and certainty of SAR images generated cannot achieve the anticipated effect.

Another stage is to extract effective feature and design powerful classifier. MIT's Lincoln Lab has two classical SAR ATR systems: template-based semiautomated image intelligence processing system program [35], [36] and model-based moving and stationary target acquisition and recognition system (MSTAR) program [37]. These two SAR ATR systems are based on manual feature extraction, and employ powerful classifiers have been used to classify the targets. However, they rely on prior knowledge of strong subjective factors and are applied merely in specific situations on account of their low generalization ability. On the contrary, the deep learning methods can extract the inherent features of targets automatically without prior information, and achieve desired recognition performance. In the work of Chen et al. [38], [39], the methods of convolutional neural network and deep convolutional neural network (CNN) were first applied to SAR image target recognition, reducing the reliance of recognition systems on professional knowledge. Unfortunately, as the depth of the network increases, although the extracted features are more and more favorable for recognition, the vanishing gradient phenomenon occurs, which eventually leads to the deterioration of the recognition performance [40].

To deal with the aforementioned problems, a novel recognition scheme called conditional Wasserstein DCGAN with a GP for ATR is proposed in this article. In the SAR image expansion stage, the label condition information is firstly introduced into GAN, then a new network architecture is designed, where the discriminator and generator adopt convolution and deconvolution networks, respectively. Meanwhile, the Wasserstein distance and GP are modified by label information. Furthermore, the quality of generated images is evaluated by qualitative and quantitative experiments. In the recognition stage, an improved recognition method based on a combination of Resnet18 [40] and support vector machine [41] (Resnet18-SVM) is adopted to elevate recognition performance.

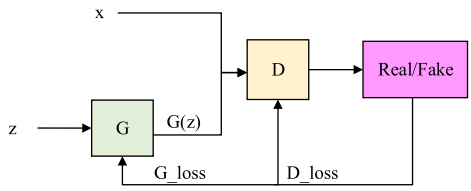


Fig. 1. Framework of the original GAN.

Our contributions can be concluded as follows.

1) An end-to-end intelligent generated model is constructed, which uses label condition information to generate SAR images of specified categories, avoiding the use of multiple networks.

2) The proposed method cleverly designs the convolution and deconvolution network structure, which reduces the loss of spatial structure caused by the full connection layer and upsampling layer, and can directly generate better quality images with specified pixel size.

3) Modified Wasserstein distance and GP by label information are added to the loss function to ensure the network training is more stable.

4) A joint recognition method of Resnet18 and support vector machine (Resnet18-SVM) is adopted to improve the recognition accuracy and generalization ability when facing different proportions of training samples.

The rest of this article is organized as follows. Section II introduces the related work of GAN. Section III introduces the proposed method in detail, and in Section IV, the results and performance analysis of the SAR image expansion experiments and classification experiments are shown. Finally, Section V concludes this article.

II. Related Works

GANs have been widely used to generate images with the same distribution as the real ones through the game between generator and discriminator [42], [43]. In GAN models, the generator learns the characteristics of real-world data and generates fake samples with the same distribution as the training data. The discriminator is responsible for examining samples to determine whether real or not.

The basic structure of the GAN is illustrated in Fig. 1. In general, the model training proceeds in two steps. In the first step, the discriminator D training is carried out by maximizing the probability of assigning the correct label to both real samples and generated samples from the generator G. In the second step, the generator G is trained by minimizing $\log(1 - D(z))$. The essence of the GAN training is to deal with the minimax problem of the function $V(D, G)$. This adversarial process between G and D can be expressed as minimizing the cross-entropy loss

$$\min_G \max_D V(D,G) = E_{x\sim p_{\mathrm{data(x)}}}[\log D(x)] + E_{z\sim p_z}\{\log[1 - D(G(z))]\}. \quad (1)$$

With the rapid development of the GAN, it has been improved in two aspects: network structure and loss function. Radford et al. [44] proposed the DCGAN, which improved the quality of the generated images by modifying the network structure. The core of the DCGAN is to replace the fully connected network of the GAN with a CNN. But the loss function is the same as the original GAN. Wasserstein distance (or earth-mover distance), which measures the distance between two distributions, was used to replace the probability measure of the original GAN to enhance the stability of the network. The formulation expression of the Wasserstein distance is given as follows [32]:

$$W(P_r, P_g) = \inf_{\gamma \in \prod(P_r, P_g)} E_{(x,y)\sim\gamma}[\|x - y\|] \tag{2}$$

where $\prod(P_r, P_g)$ represents the set of all joint distributions $\gamma(x, y)$ whose marginals are P_r and P_g, respectively. Equation (2) cannot be calculated directly, thereby the loss function can be transformed as follows:

$$L \approx \max_{w:\|D_w\|_L \le K} E_{x\sim P_{r(x)}}[D_w(x)] - E_{x\sim P_{g(x)}}[D_w(G(x))] \tag{3}$$

where w represents the parameters of D. The optimal solution of Equation 3 must be obtained on a premise of $\|D_w\|_L \le K$.

The WGAN can easily solve the problem of gradient vanishing during the model training. Unfortunately, the phenomenon of gradient explosion often occurs [33]. Thus, a GP term is introduced to modify the loss function. The final loss function is expressed as follows:

$$L = E_{x\sim P_{r(x)}}[D(x)] - E_{x\sim P_{g(x)}}[D(G(x))] - \underset{\tilde{x}\sim P_{\tilde{x}}}{\lambda E}\left[(\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - 1)^2\right] \tag{4}$$

where x is calculated by $x = \varepsilon x_r + (1 - \varepsilon)x_g$.

Since the original GAN cannot control the categories of the generated images, this will result in the generated images full of uncertainty. Then, Mirza et al. [29] proposed the CGAN to control the properties of the generated images. The loss function of the CGAN is as follows:

$$\min_G \max_D V(D, G) = E_{x\sim p_{\mathrm{data}(x)}}[\log D(x|y)] + E_{z\sim p_z}\{\log[1 - D(G(z|y))]\}. \tag{5}$$

It is worth noting that CGAN only adds condition information to the inputs of G and D, still using the fully connected network.

III. CWDCGAN FOR ATR

The flowchart of the proposed CWDCGAN expansion method for ATR is presented in Fig. 2. The original SAR image set is divided into two disjoint parts. The first part is original training set, which is used to train the CWDCGAN offline. Another is original test set, which is used to test the final recognition performance online. In the data expansion stage, the well-trained CWDCGAN model comes into being, and the SAR images with the same distribution as the original training set are output. The generated images quality is assessed simultaneously by qualitative and quantitative evaluations. Finally, the generated images are mixed with the original training set to form the extended training dataset. In the target recognition stage, the Resnet18-SVM is trained with the expanded training data, and then, the classification results of the original test set are output through the trained Resnet18-SVM. The specific principles are described in detail in the following.

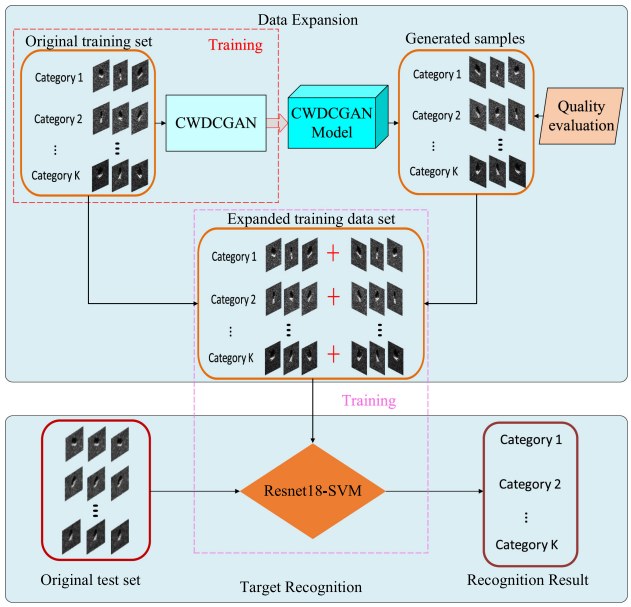


Fig. 2. Flowchart of the CWDCGAN for ATR.

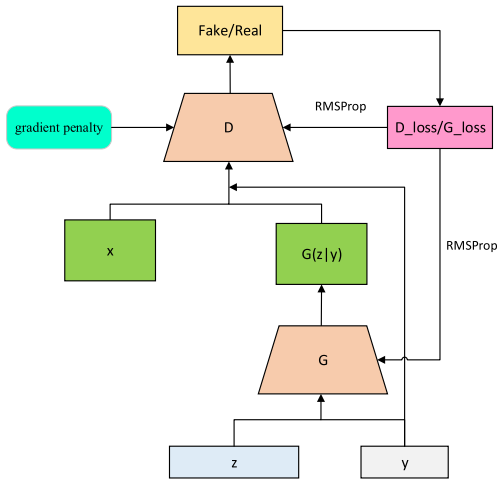


Fig. 3. Framework of the CWDCGAN.

A. Data Expansion Method Based on the CWDCGAN

The CWDCGAN is made up of a deconvolution network G and a convolution network D, including the addition of generated image label information y to control the generated sample category. The framework of the CWDCGAN is shown in Fig. 3, the noise z and the label y are input into G together to get a fake image $G(z|y)$, and then the real image x or the fake image $G(z|y)$ is discriminated by D, respectively. The detailed

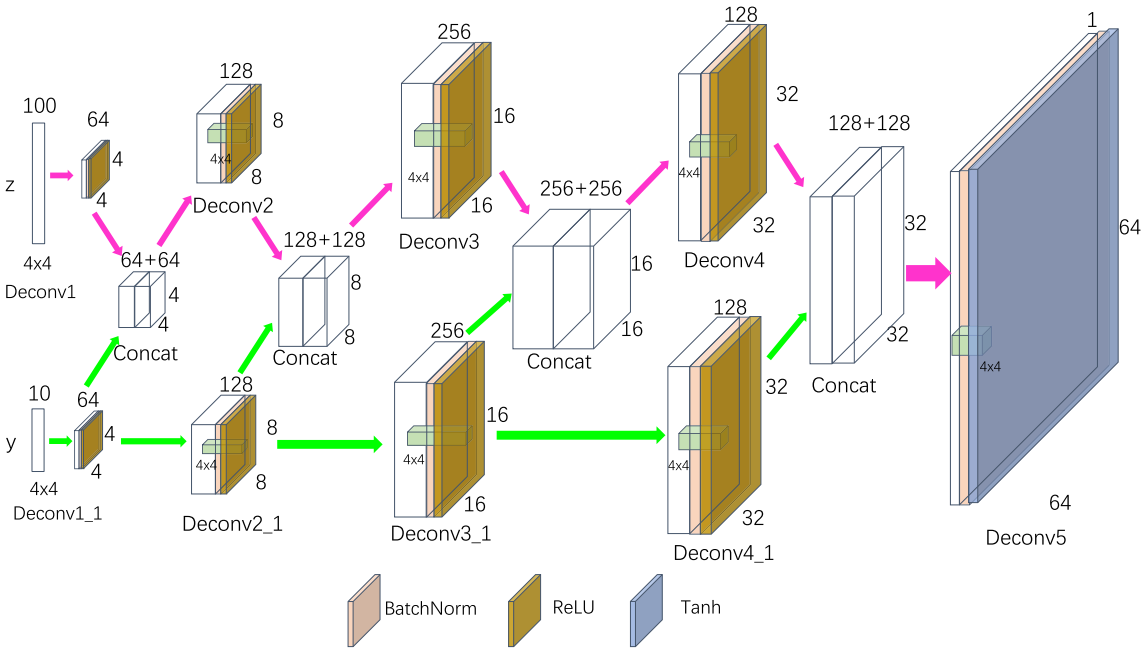


Fig. 4. Network structure of the generator G.

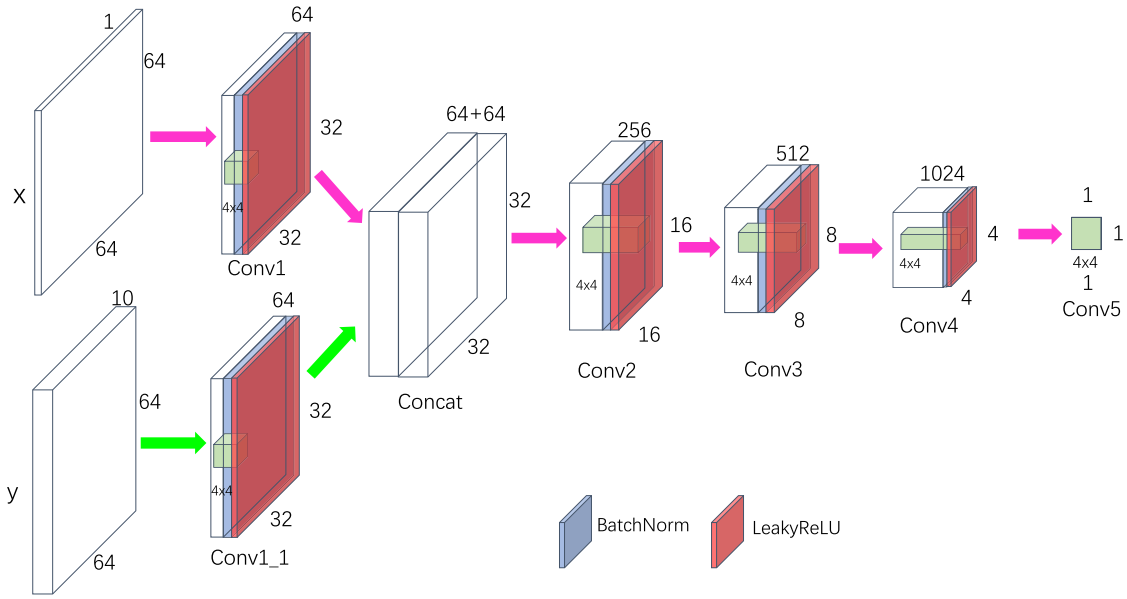


Fig. 5. Network structure of the discriminator D.

introduction of G and D components is shown in Figs. 4 and 5 and Tables I and II.

The generator, called G, contains a five-layer deconvolution structure. In each layer of the network, the noise vector z with dimension 100 and label vector y with dimension 10 are deconvoluted, respectively, and their corresponding outputs are called noise feature map and label feature map. These two feature maps are integrated as an image feature map, combining with the label feature map together as the input of the next layer. The next layer continues to deconvolute and connect the two feature maps until the fifth layer. After deconvoluting at each layer, batch normalization and the ReLU [45] activation function are performed. In particular, the activation function of the last layer is Tanh [46]. In this way, we finally get a SAR image with a pixel size of 64×64. Significantly, the label information is used in every layer to avoid being forgotten as the network depth increases. Table I lists more details of G network architecture.

The discriminator, called D, consists of a five-layer convolution structure. In order to realize a better discrimination performance, in the first layer, the convolution on the label information y and the image x (real image or generated fake image)

TABLE I
ARCHITECTURE OF THE GENERATOR G

Generator G					
Input: Noise vector z, label vector y					
Layer No.	Layer Type	Kernel size	Number of kernels	Stride	Padding
1	Deconvolution Layer (Layer1)	4×4	64	1	0
	Deconvolution Layer (Layer1_1)	4×4	64	1	0
	Concat Layer	-	-	-	-
2	Deconvolution Layer (Layer2)	4×4	128	2	1
	Deconvolution Layer (Layer2_1)	4×4	128	2	1
	Concat Layer	-	-	-	-
3	Deconvolution Layer (Layer3)	4×4	256	2	1
	Deconvolution Layer (Layer3_1)	4×4	256	2	1
	Concat Layer	-	-	-	-
4	Deconvolution Layer (Layer4)	4×4	128	2	1
	Deconvolution Layer (Layer4_1)	4×4	128	2	1
	Concat Layer	-	-	-	-
5	Deconvolution Layer (Layer5)	4×4	64	1	0

TABLE II
ARCHITECTURE OF THE DISCRIMINATOR D

The discriminator G					
Input: image, label vector y					
Layer No.	Layer Type	Kernel size	Number of kernels	Stride	Padding
1	Convolution Layer (Layer1)	4×4	64	2	1
	Convolution Layer (Layer1_1)	4×4	64	2	1
	Concat Layer	-	-	-	-
2	Convolution Layer (Layer2)	4×4	256	2	1
3	Convolution Layer (Layer3)	4×4	512	2	1
4	Convolution Layer (Layer4)	4×4	1024	2	1
5	Convolution Layer (Layer5)	4×4	1	1	0

is performed, and then, the convolution results are connected into an image convolution feature map which is the input of the next layer. In the following four-layer convolution, the label information is no longer used. The batch normalization and the activation function LeakReLU [47] exist in layers 1–4. In order to establish a better connection after the first layer, we expand the label vector y to the size of $10 \times 64 \times 64$ before the first layer. Finally, we get a scalar discrimination result. The details of D are given in Table II.

The loss function of CWDCGAN is modified by introducing condition information, Wasserstein distance, and GP. The final formula is as follows:

$$L = E_{x\sim p_{\text{real}}(x)}[D(x|y)] - E_{z\sim p_z(z)}[D(G(z|y))] - \lambda E_{\tilde{x}\sim p_{\text{penalty}}(\tilde{x})}[(\|\nabla_{\tilde{x}} D(\tilde{x}|y)\| - 1)^2] \quad (6)$$

where x is the input image (the real image or the generated image), y is the label of the input image, $\tilde{x}$ is calculated by $\tilde{x} = \varepsilon x_r + (1-\varepsilon)x_g$, x_r and x_g are the real samples and generated samples with the same label.

Compared with the original GAN, three points are differences, given as follows.

1) The loss function with the condition information y is able to calculate the loss of different categories of targets.
2) Since Wasserstein distance represents the distribution distance between the generated samples and the real samples, the log factor is no longer required for the loss function.
3) GP term makes the network training more stable. Note that the label condition information y is used to modify the Wasserstein distance and GP simultaneously.

Algorithm 1: CWDCGAN Training.

Require: The GP coefficient λ, iteration number N1, the learning rate $lr1$, smoothing constant β, the initial generator parameters ω_g, and the initial discriminator parameters ω_d. All initial parameters need to have a mean of 0 and a variance of 0.02. Initialize the generator G, the discriminator D

for $i = 1$ to N1 **do**
  for $j = 1$ to m_{batch} **do**
    Sample the real samples $x_{real} \sim P_{real}(x)$, corresponding label y, the noise samples $z \sim P_z(z)$
    Combine the one-hot coding y with the input of D and G
    Sample a random number $\varepsilon \sim U[0,1]$
    $x_g \leftarrow G(z|y)$
    $x \leftarrow \varepsilon x_r + (1-\varepsilon)x_g$
    $L_i \leftarrow D(x_g|y) - D(x_r|y) + \lambda((\|\frac{\partial D(x|y)}{\partial x}\|_2 - 1)^2)$
    $\omega_d \leftarrow RMSPorp\left(\frac{1}{m}\sum_{i=1}^{m}\frac{\partial L_i}{\partial \omega_d}, \omega_d, \beta, lr1\right)$
    Sample a batch of prior samples $(z^i)_{i=1}^m \sim P_z(z)$
    Combine the one-hot coding y with the input of D and G
    $\omega_g \leftarrow RMSPorp\left(\frac{1}{m}\sum_{i=1}^{m}\frac{\partial G_{loss}}{\partial \omega_g}, \omega_g, \beta, lr1\right)$
  end for
end for

Return ω_g, ω_d

In the area of deep learning, almost all deep neural networks are nonconvex, general optimization algorithms do not perform well. Therefore, the optimization algorithm selection is very crucial, which usually affects the network training process and whether the objective function can converge to the global optimal value [48]. As is known to all, the root mean square propagation (RMSProp) [49] algorithm can control the amount of historical information obtained by adding attenuation coefficients and ultimately achieve better results. So it is chosen as a suitable candidate for the CWDCGAN training, and the formula is as follows:

$$s_t \leftarrow \beta s_{t-1} + (1-\beta) g_t \odot g_t \tag{7}$$

$$x_t \leftarrow x_{t-1} - \frac{\eta}{\sqrt{s_t + \xi}} \odot g_t \tag{8}$$

where g_t is the gradient of variable x at the time t, β is a hyperparameter between 0 and 1, $\odot$ is Hadamard product, η is the learning rate, and ξ is a constant added to maintain numerical stability, such as 10^{-6}.

In the CWDCGAN model, the discriminator and the generator are alternately trained competitively, requiring no special training ways. The training step of the CWDCGAN model is summarized in Algorithm 1.

B. Image Quality Evaluation

In order to verify the quality of the generated images, the qualitative and quantitative experiments are designed to evaluate the performance of CWDCGAN. First of all, we observe the visual differences between generated images and real images. At the same time, some metrics are used to evaluate the generated SAR images quantitatively, including mean value, variance, information entropy, linear index of fuzziness (LIF), mean gradient, gray level difference (GLD), and Fréchet inception distance (FID) scores [29], [42], [50]. The evaluation metrics are introduced in the following. Suppose the size of image A is $M \times N$.

1) *Mean value:* The mean value of the image represents its total energy, given by:

$$\mathrm{MV} = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N} A(i,j) \tag{9}$$

2) *Variance:* The variance of an image symbolizes the deviation degree of the image from the mean, which can be expressed as

$$V = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N} (A(i,j) - MV) \tag{10}$$

3) *Information entropy:* The information entropy denotes the amount of information in the image and reflects the focusing degree of the image. The smaller the information entropy value is, the more focused the image is. The calculation method can be expressed as

$$H = -\frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}\log p_{ij} \tag{11}$$

where p_{ij} is the probability value of pixels in image $A(i,j)$.

4) *LIF:* The LIF is often used to quantitatively evaluate the enhancement effect of images in spatial domain. It reflects the blurring degree of an image, and the smaller the value is, the sharper the image is. It can be defined by

$$\mathrm{LIF} = \frac{2}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\min(p_{mn}, (1-p_{mn})) \tag{12}$$

$$p_{mn} = \sin\left(\frac{\pi}{2}\times\left[1 - \frac{A(i,j)}{A_{\max}}\right]\right). \tag{13}$$

5) *Average gradient (AG):* The AG can reflect the ability of an image to express details and textures, and is often used to evaluate the clarity of an image. The larger the AG value is, the clearer the edge details of the image are. It can be defined as

$$\mathrm{AG} = \frac{1}{4}\frac{1}{M-1}\frac{1}{N-1}\sum_{i=1}^{M-1}\sum_{j=1}^{N-1} \times\sqrt{[\partial A(i,j)/\partial i]^2 + [\partial A(i,j)/\partial j]^2} \tag{14}$$

where $\partial A/\partial i$ and $\partial A/\partial j$ represent the horizontal and vertical gradient values of the image, respectively.

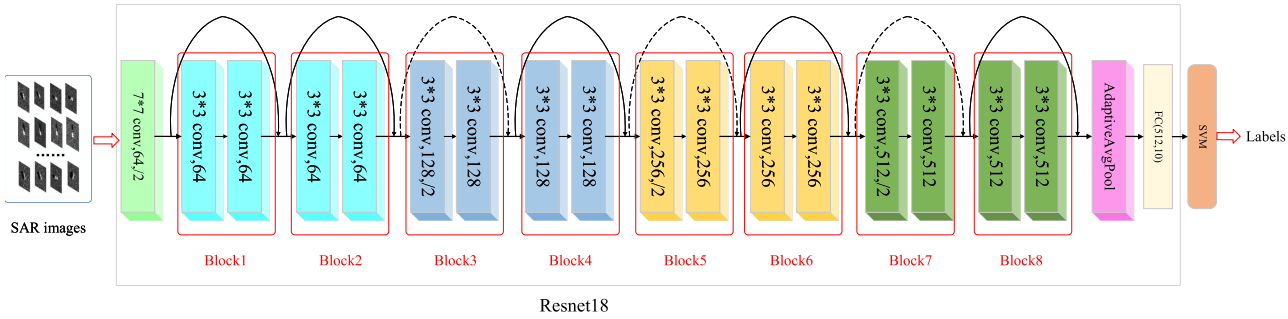


Fig. 6. Resnet18-SVM network structure.

6) *GLD:* The GLD of the image represents the edge sharpness of the target area of interest in the image, which is obtained in the following ways:

$$\mathrm{GLD} = \frac{1}{M-1}\frac{1}{N-1}\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}\left(|A(i,j)-A(i+1,j)| + |A(i,j)-A(i,j+1)|\right). \tag{15}$$

A larger GLD value indicates the image has clearer edge details.

7) *FID:* The FID score is often used to evaluate the image quality of the GAN. FID measures the similarity between two groups of images from computer vision statistical features. It is a measure to calculate the feature vector distance of the real images and the generated images. This visual feature is extracted and calculated by using the inception V3 [51] image classification model. The value of FID in the best case is 0, indicating that the two groups of images are the same. The lower the value is, the more similar the two groups of images are, or the more similar the statistics of the two groups are. In the same way, the smaller the FID is, the better diversity and quality of the image are. The FID is defined as

$$\mathrm{FID}(x,g) = \|MV_x - MV_g\|_2^2 + \mathrm{Tr}\left(\sum\nolimits_x + \sum\nolimits_g - 2\left(\sum\nolimits_x \sum\nolimits_g\right)^{1/2}\right) \tag{16}$$

where x and g represent the real image and the generated image, respectively, $\|\cdot\|_2^2$ is the square of the L2 norm, $\mathrm{Tr}(\cdot)$ is the trace of the matrix, and $\sum_{\cdot}$ is the covariance matrix of the input image features.

C. Feature Extraction and Classification

In order to extract deep features automatically with good representation ability and improve the generalization ability of the recognizer, a Resnet18 and SVM combined recognition method (Resnet18-SVM) is proposed. The Resnet18-SVM network structure is shown in Fig. 6. Here, the input sample is grayscale image, so the number of network channels is set to 1. Resnet18 network includes eight basic blocks shown in the red box. In every block, the input and output are connected by a shortcut, which is curved solid line or dotted line. The curved solid line indicates that the output size is the same as the input, after passing through this basic block, while the curved dotted line indicates that the size is halved. Because of the existence of these shortcuts, the model degradation problem that the gradient disappears due to the excessive number of network layers is solved. The deep features are extracted by Resnet18 network from SAR images and fed into SVM for classification. The detailed structure parameter settings of the Resnet18-SVM network are given in Table III.

Here, an SVM classifier is adopted, whose optimization goal is to maximize the category interval, so the usual cross-entropy loss is a little bit limited. Hinge loss [52] is a loss conducive to max margin classification in the field of machine learning, so it can be used here to replace cross-entropy loss. In the case of binary classification, the formula is as follows:

$$L_{\mathrm{Hinge}}(y) = \max(0, 1 - y \cdot \tilde{y}) \tag{17}$$

where $\tilde{y}$ is the predicted value (between -1 and 1) and y is the target value (1 or -1).

When the data $D(x, y)$ contain N samples of C categories, $\tilde{y}$ is the output of the neural network, y is the real category label $(0 \le y_n \le C-1)$. The loss value $l_{n-\mathrm{Hinge}}$ of the nth sample is calculated as follows:

$$l_{n-\mathrm{Hinge}} = \frac{1}{C}\sum_{i=0\&i\neq y_n}^{C-1} \max\left(0, \mathrm{maigin} - \tilde{y}_n[y_n] + \tilde{y}_n[i]\right)^p \tag{18}$$

where p and maigin are usually 1.

The SGD optimization algorithm with the learning rate $lr2$ and momentum term γ is utilized to optimize the Resnet18 network. The training step of the Renset18-SVM model are summarized in Algorithm 2.

IV. Experiment Results

In this section, the effectiveness of the proposed ATR method based on SAR image expansion is demonstrated by experiments. First, the quality of SAR images generated by the CWDCGAN method and the stability of this model are evaluated. Then, the recognition performance of the proposed Resnet18-SVM method is explored in the case of the different number of training samples. At the same time, the experiments are carried out on MSTAR to check whether the generated SAR image can improve the performance of ATR.

TABLE III
RESNET18-SVM NETWORK STRUCTURE PARAMETER

Layer name	Output size	18-layers
Conv1	32×32	7×7, 64, stride 2
Basic block1_2	32×32	$\begin{bmatrix} 3 \times 3 & 64 \\ 3 \times 3 & 64 \end{bmatrix} \times 2$
Basic block3_4	16×16	$\begin{bmatrix} 3 \times 3 & 128 \\ 3 \times 3 & 128 \end{bmatrix} \times 2$
Basic block5_6	8×8	$\begin{bmatrix} 3 \times 3 & 256 \\ 3 \times 3 & 256 \end{bmatrix} \times 2$
Basic block7_8	4×4	$\begin{bmatrix} 3 \times 3 & 512 \\ 3 \times 3 & 512 \end{bmatrix} \times 2$
AdaptiveAvgPool	10×1	AdaptiveAvgPool 10-d FC
SVM	1×1	Kernel=rbf (Gaussian)

Algorithm 2: Resnet18-SVM Training.

Require: iteration number $N2$, the learning rate $lr2$, the hyperparameters $\beta 2$, the initial Resnet18 parameters and SVM parameters ω_R, ω_S

STEP 1 Resnet18 Training

for $i = 1$ to N2 **do**

  for $j = 1$ to m_{batch} **do**

    Sample the training samples $(x, y) \sim D_{training}(x, y)$

    $$l_{n-Hinge} = \frac{1}{C} \sum_{i=0 \& i \neq y_n}^{C-1} \max\left(0, maigin - \tilde{y}_n[y_n] + \tilde{y}_n[i]\right)^p$$

    $$\omega_R \leftarrow SGD\left(\frac{1}{m} \sum_{i=1}^{m} \frac{\partial l_{n\_Hinge}}{\partial \omega_R}, \omega_R, \gamma, lr2\right)$$

  end for

end for

Return ω_R

STEP 2 SVM Training

for $k = 1$ to m_{batch} **do**

  Sample the training samples $(x, y) \sim D_{training}(x, y)$

  $F \leftarrow \text{Resnet18(x)}$

  $w_S \leftarrow Fit\ SVM\ use\ (F, y)$

end for

Return ω_S

Return ω_R, ω_S

A. Dataset Description

The experiment adopts the measured SAR image data from the MSTAR program [53]. The sensor collecting the dataset is a high-resolution spotlight SAR with a resolution of 0.3×0.3 m. It contains X-band and HH polarization SAR slice images of various targets at the 128×128 pixel size. Most of the SAR slice images are about stationary vehicles, including various vehicle target images in various azimuth and depression angles. The azimuth range is 0° to 360° with intervals about of 1° to 2°, and the depression is 15° or 17°.

B. Target SAR Image Expansion Experiment

1) Data Adoption and Parameter Setting: As described in Section III, we divide the data into the original training set and the original test set. These two sets are composed of ten different categories of ground stationary vehicle targets, such as armored vehicles (BMP2, BRDM2, BTR60, and BTR70), tanks (T62 and T72), rocket launchers (2S1), air defense units (ZSU234), trucks (ZIL131), and bulldozers (D7). The depression angle of the original training set is 17° and the original test set is 15°. The optical images of these targets and the corresponding SAR images are shown in Fig. 7. It is obvious that there are speckles, shadows, and clutters in the real SAR images, which cannot be observed in generated images by traditional simulation methods [54].

In the following experiments, to reduce the influence of speckle noise and the calculation burden, the images are cropped to the pixel size of 64×64 from 128×128. The details of various target samples in MSTAR are given in Table IV.

CWDCGAN model training: The model training is performed with the minibatch 64 and the noise vector z with dimension 100. The label y is converted by one-hot encoding as a vector with dimension 10. All the experiments are implemented on the open-source library Pytorch 1.90. The interface is Python 3.8 and the platform is a Dell Precision 5820 workstation with an Intel i9-10920X CPU, a GeForce GTX 3090, and 64 GB RAM under Windows 10. In the training procedure, we use an RMSProp-based optimizer with a learning rate $lr1 = 0.0002$ and a smoothing constant $\beta = 0.99$. Simultaneously, the training iteration number is 1000, and the G will be updated once after each update of the D.

2) Experimental Results of the CWDCGAN: Furthermore, in order to analyze the advantages of the CWDCGAN, we design

TABLE IV
IMAGE NUMBER OF EXPERIMENTAL DATASET

Category	Original training set (17°)	Original test set (15°)
2S1	299	274
BMP2(sn9566)	232	196
BRDM2	298	274
BTR60	256	195
BTR70(c71)	233	196
D7	299	274
T62	299	273
T72(sn132)	232	196
ZIL131	299	274
ZSU234	299	274
Total	**2746**	**2426**

The bold values show the total number of samples.

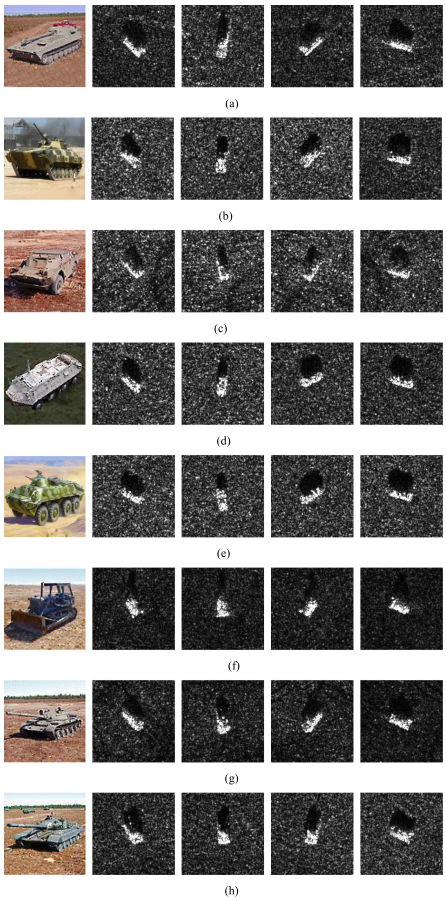


Fig. 7. Optical images and the corresponding SAR images of ten targets. (a) 2S1. (b) BMP2. (c) BRDM2. (d) BTR60. (e) BTR70. (f) D7. (g) T62. (h) T72.

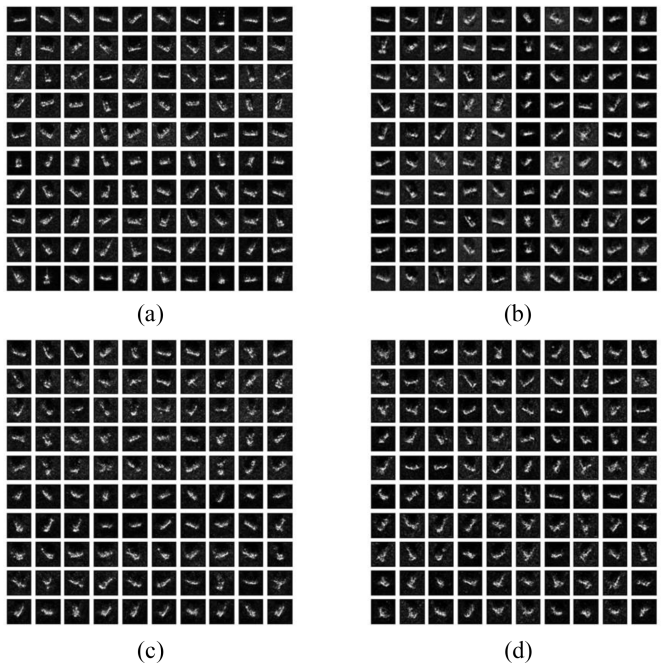


Fig. 8. Real SAR images and the generated SAR images with three different models. (a) Real SAR images. (b) Generated SAR images by CGAN. (c) Generated SAR images by CDCGAN. (d) Generated SAR images by CWDCGAN. The column of each subgraph represents the category of the target, the order from left to right is as follows: 2S1, BMP2, BRDM2, BTR60, BTR70, D7, T62, T72, ZIL131, ZSU234, and the rows are different sample images of the current target category.

another two comparison models: the CGAN and the CDCGAN. The CGAN is constructed by introducing label condition information into the original GAN model, which adopts a five-layer full connection structure. The CDCGAN uses a 5-layer convolution and deconvolution structure.

Comparison of the generated images with the real images: As described in Section III-B, we first qualitatively evaluate the generated images by observing the visual differences. As shown in Fig. 8, the three models can generate SAR images of corresponding categories, which proves that the introduced label condition information is effective. Therefore, it avoids training multiple network models and saves workload.

In order to qualitatively evaluate the quality of the generated image clearly, the generated images by each model separately are selected and compared with the real images in various azimuths. As shown in Figs. 9–11, taking 2S1, D7, and BDRM2 as examples, the first row of each figure is the real SAR images, and the other three rows are the generated SAR images by

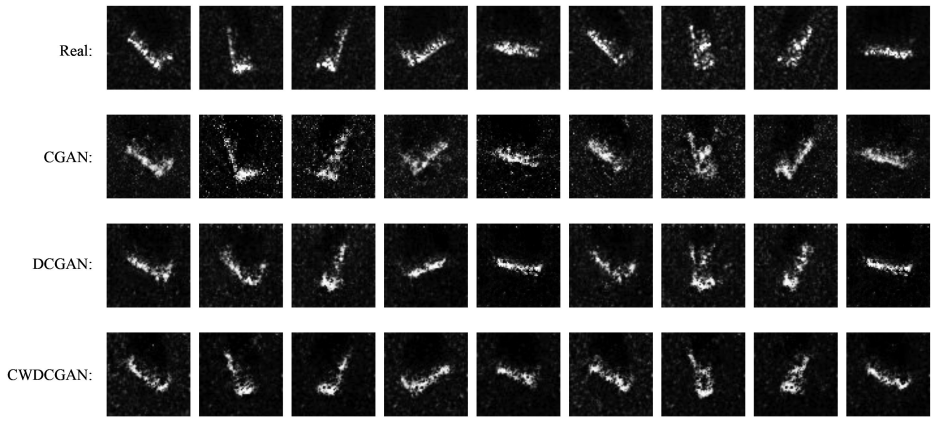


Fig. 9. Real images and generated SAR images of 2S1.

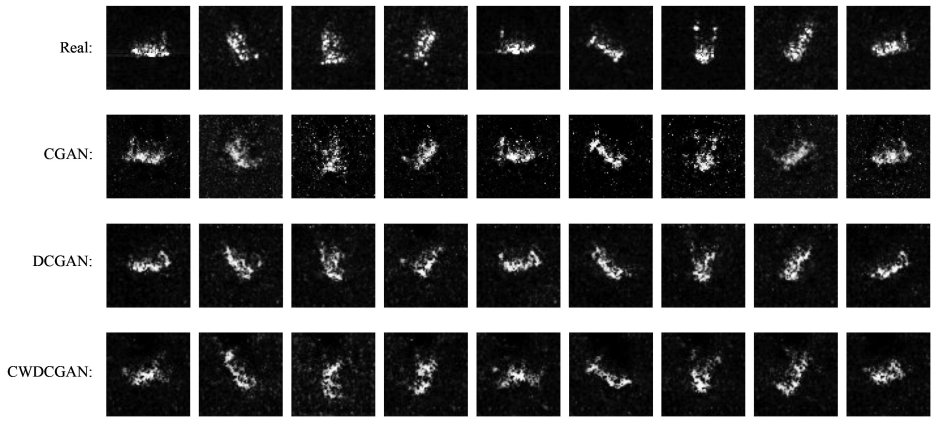


Fig. 10. Real images and generated SAR images of D7.

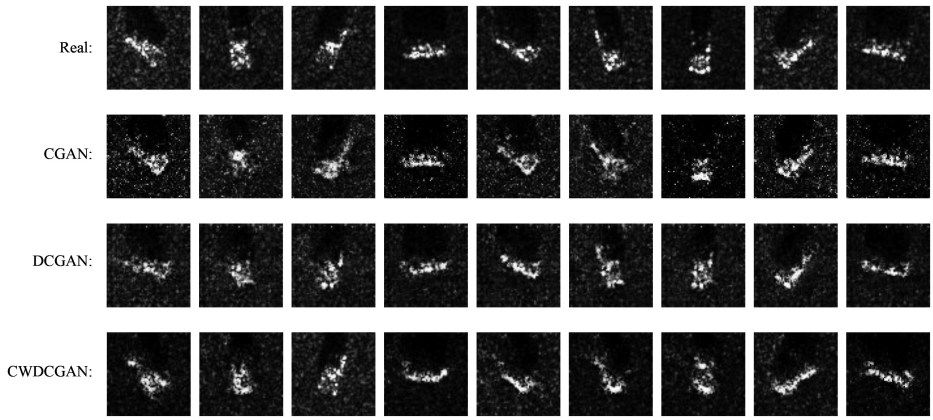


Fig. 11. Real images and generated SAR images of BMP2.

the CGAN, CDCGAN, and CWDCGAN models in order. The images with approximately the same azimuth are put in the same column.

Obviously, SAR images with different azimuth angles can be generated by three models, which are beneficial to improve the diversity of samples. However, compared with the real images, the SAR images generated by the CGAN model are blurred and full of random bright spots. In contrast, the images generated by other methods have almost no bright spots, and the images are clearer, especially CWDCGAN. At the same time, the background and target in the image are more similar to the real image. According to the abovementioned comparison, it can be found

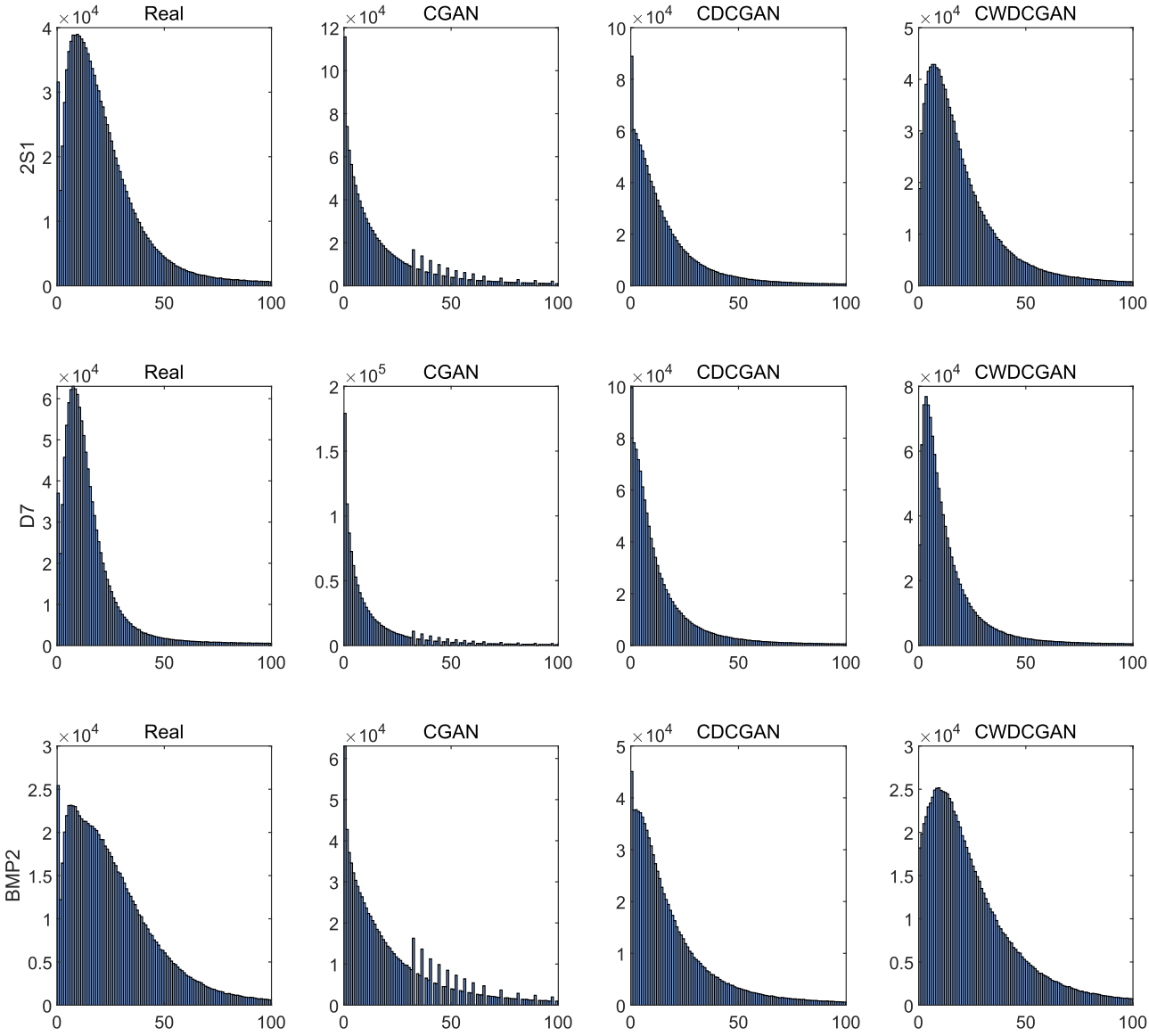


Fig. 12. Gray histograms of four types of images. Each row represents a category of target: 2S1, D7, or BMP2. The order of the gray histograms in every row from left to right is: the real images, images generated by CGAN, images generated by CDCGAN, and images generated by CWDCGAN.

that introducing the designed convolution and deconvolution network structure can make full use of the spatial structure information of the real images, and generate SAR images closer to the real ones.

However, due to the poor interpretability of SAR images, Fig. 12 compares the statistical feature differences of generated images and real images in the form of a gray histogram. In terms of the shape, the gray histogram of the CWDCGAN image is most similar to the one of the real image. It demonstrates that the CWDCGAN model has mastered the distribution of real images from statistical characteristics to a greater extent and generated more realistic images. Moreover, to add credibility to the experimental results, the histogram similarity index is adopted to measure the distance between different histograms of the same target. Table V tabulates the comparison results of the gray histogram similarity between the generated images and the real images. By means of comparing the values of these histogram similarity indexes, it is easy to find evidence that the images generated by the CWDCGAN most resemble the real images in terms of statistical characteristics.

It is not enough to compare statistical features between generated and real images by direct observation alone. In order to make the experimental results more convincing, a qualitative evaluation experiment as stated in Section III-B is carried out to quantitatively evaluate the quality of the generated SAR images. In terms of the former six indexes, the quality of generated SAR images can be judged by comparing the similarity between generated SAR images and real images. The smaller the difference between the index values, the closer the generated SAR images are to the real images. As for FID, a smaller value indicates a higher quality and diversity of the generated images. It can be seen from Table VI that the generated images of the CWDCGAN are closest to the real ones, and the quality and diversity are optimal.

TABLE V
HISTOGRAM SIMILARITY INDEXES OF GENERATED IMAGES AND REAL IMAGES

Target category / Comparative object	CGAN and Real	CDCGAN and Real	CWDCGAN and Real
2S1	0.9286	0.9667	0.9928
D7	0.9176	0.9685	0.9826
BMP2	0.9364	0.9724	0.9954

TABLE VI
EVALUATION INDEX VALUES FOR THE SAMPLES GENERATED BY THE THREE MODELS

	Mean	Variance	Entropy	LIF	AG	GLD	FID
Real samples	**0.1089**	**0.0221**	**0.9229**	**0.0243**	**0.3807**	**0.1046**	**—**
Generated samples (CGAN)	0.0953	0.0269	1.0143	0.0315	0.4518	0.1614	57.6534
Generated samples (CDCGAN)	0.1057	0.0241	1.0126	0.0274	0.4261	0.1240	37.3382
Generated samples (CWDCGAN)	**0.1096**	**0.0238**	**0.9022**	**0.0244**	**0.3932**	**0.1137**	**30.4716**

The bold values highlight that the evaluation values of the samples generated by the proposed method are closer to the real samples.

By comparison, it can be concluded that the quality and diversity of images generated by the CWDCGAN and CDCGAN are higher than those generated by the CGAN. It further shows that the convolution and deconvolution structures we designed can learn the distribution of the real images and extract the spatial structure information more intelligently.

Loss changes during the model training: In order to verify the convergence stability of our proposed model, the change curves of the loss value in the model training process are plotted. Since the output of discriminator D uses the activation function, the CGAN and CDCGAN models' losses refer to the probability difference between the generated samples and real samples. As the model training progresses, these losses may converge to a very small value, indicating that network training has stabilized. In terms of the CWDCGAN model, the activation function is removed in the output of discriminator D and the Wasserstein distance is used. In addition, a GP term is also introduced. Therefore, the loss of CWDCGAN represents the sum of the distribution difference (the distribution difference between the generated images and the real images) and the absolute value of the GP term. This loss is not going to converge to a very small value like the others. When this loss value converges to a stability status, it means that the network training is completed.

According to Fig. 13, D_Loss and G_Loss of the CGAN model drop rapidly at the beginning, but G_Loss gradually became unstable with the progress of training. In the CDCGAN model, G_Loss shows similar changes, which implies the quality of generated images is getting worse and worse. However, in CWDCGAN model, it can be observed that G_Loss and D_Loss converge rapidly and maintain a stable state. It indicates that the introduced Wasserstein distance and GP do good to the network training.

In summary, in this section, we can come to several conclusions, given as follows.

1) The introduction of label condition information is beneficial for generating target SAR images of specified categories, which is beneficial to reduce the number of required networks and the workload of network training.
2) Introducing convolution and deconvolution structures into the GAN designed in this article, the quality of the generated SAR images is significantly improved.
3) The modified Wasserstein distance and GP terms make the network more stable and easier to train.

C. Classification Experiments

Previous expansion experiments have proved that the images generated by the proposed method are of high quality and diversity, but the data expansion is not the final goal. Therefore, the following experiment will attempt to not only verify the classification performance of Resnet18-SVM, but also judge whether the generated SAR images are conducive to improve the ATR performance.

1) Data Adoption and Parameter Setting: To verify the performance of our proposed recognition method, the different classifiers are trained with the extended training dataset and then tested with the original test set. In the proposed method, the loss function of the classification model is hinge loss, optimization algorithm is SGD, learning rate is 0.01, momentum is 0.9, and iteration number is 1000. The SVM parameter is set to the Gaussian kernel and the experimental equipment is the same as in Section IV-B.

2) Experimental Results of Resnet18-SVM: In this section, we conduct the experiments on an original training dataset and three extended training datasets. Every extended dataset is a one-to-one blend of the original training samples and the generated data corresponding by one of the models. The changes of recognition accuracy of ten different recognition methods are

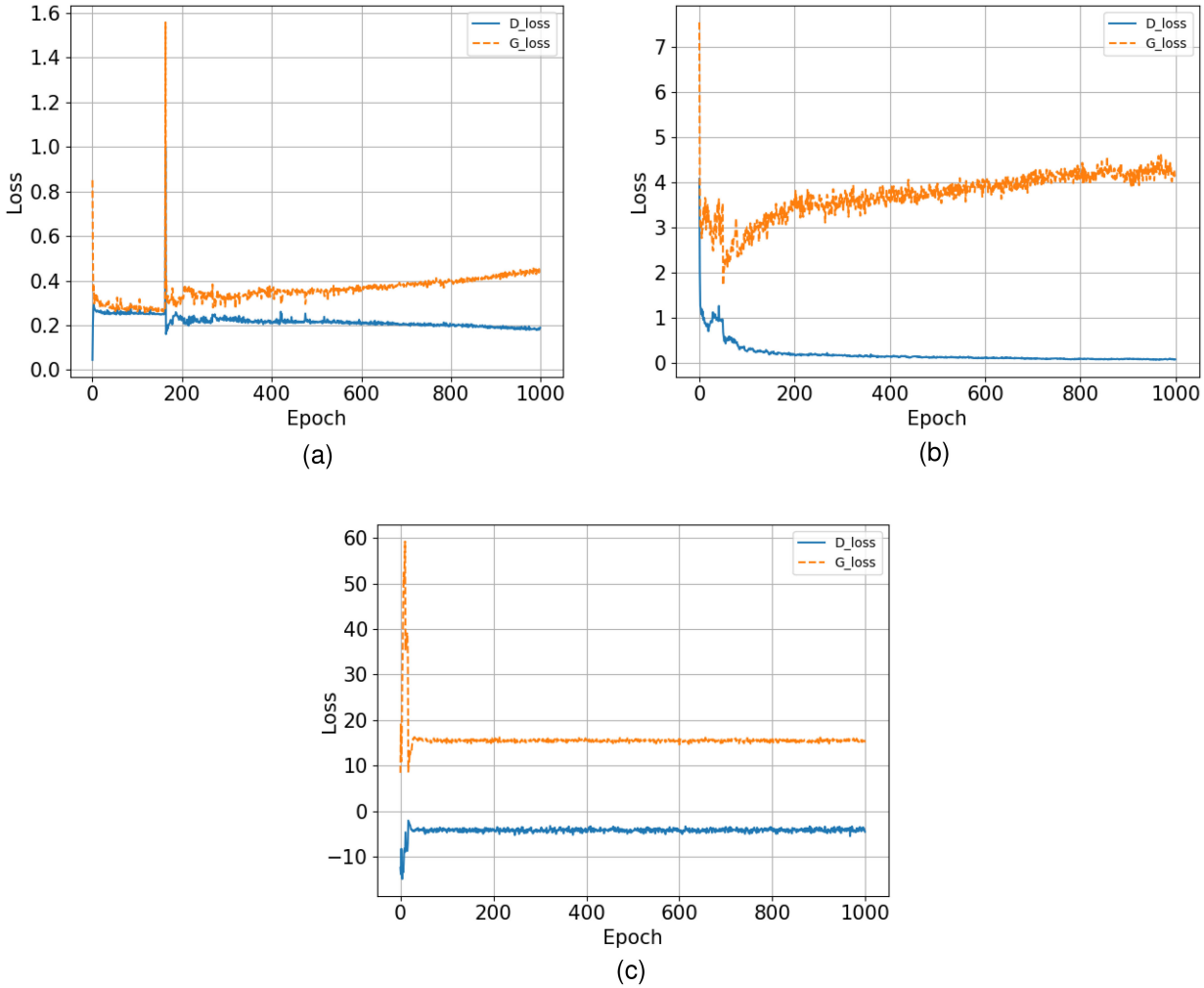


Fig. 13. Loss value changes with the training process proceeding in three models. (a) Loss corresponding to the CGAN model. (b) Loss corresponding to the CDCGAN model. (c) Loss corresponding to the CWDCGAN model. D_Loss and G_Loss represent the loss of discriminator and generator, respectively. (a) CGAN (b) CDCGAN (c) CWDCGAN.

TABLE VII
RECOGNITION RESULTS (100% OF THE ORIGINAL TRAINING SET)

Method	Recognition accuracy (%)						
	Original	Original+ CGAN	Δ	Original+ CDCGAN	Δ	Original+ CWDCGAN	Δ
Logistic	84.61	83.59	-1.02	83.67	-0.94	84.66	0.05
KNN	91.71	91.70	-0.01	91.76	0.05	91.84	0.13
SVM	94.35	93.95	-0.40	94.27	-0.08	94.48	0.13
MLP	84.92	84.46	-0.46	85.37	0.45	86.73	1.81
VGG16-SVM	97.61	97.55	-0.06	97.78	0.16	97.82	0.20
Resnet50-SVM	92.79	91.80	-0.98	92.33	-0.45	92.83	0.04
Densenet121-SVM	94.64	95.38	-0.74	95.41	0.77	95.49	0.85
MobeilnetV1-SVM	95.22	93.65	-1.57	94.31	-0.90	95.21	-0.01
Resnet18-softmax	98.49	97.85	-0.30	97.97	-0.14	98.60	0.11
Resnet18-SVM	98.05	97.89	-0.16	97.98	-0.07	98.10	0.05

recorded in Table VII. Δ represents the change in the recognition accuracy compared to the original training set.

By observing Table VII, it can be found that the Resnet18-softmax and Resnet18-SVM recognition methods are better than the other eight common recognition methods, indicating that features extracted by Renset18 network are conducive to recognition. Moreover, by horizontally comparing the classification performance of four training sets corresponding to different methods, it can be checked that the classification performance of the expanded training data based on the CGAN and the CDCGAN declines slightly, which is consistent with what we suspected. Low-quality images generated by these two methods

TABLE VIII
RECOGNITION RESULTS (80% OF THE ORIGINAL TRAINING SET)

Method	Recognition accuracy (%)						
	Original	Original+ CGAN	Δ	Original+ CDCGAN	Δ	Original+ CWDCGAN	Δ
Logistic	84.37	80.67	-3.70	84.83	0.46	84.38	0.01
KNN	88.57	88.95	0.38	89.53	0.96	89.74	1.17
SVM	93.53	94.11	0.58	92.95	-0.85	93.73	0.20
MLP	83.20	80.31	-2.89	83.97	0.77	84.67	1.47
VGG16-SVM	97.24	97.15	-0.08	97.68	0.44	97.71	0.48
Resnet50-SVM	90.31	84.13	-6.18	88.71	-1.61	91.61	1.29
Densenet121-SVM	94.19	93.65	-0.54	93.73	-0.45	94.54	0.35
MobeilnetV1-SVM	92.46	93.03	0.58	93.44	0.98	94.44	1.98
Resnet18-softmax	96.23	97.33	1.10	97.72	1.49	97.83	1.60
Resnet18-SVM	97.61	96.13	-1.30	97.20	-0.41	97.98	0.37

TABLE IX
RECOGNITION RESULTS (60% OF THE ORIGINAL TRAINING SET)

Method	Recognition accuracy (%)						
	Original	Original+ CGAN	Δ	Original+ CDCGAN	Δ	Original+ CWDCGAN	Δ
Logistic	83.25	78.35	-4.90	83.43	0.18	84.20	0.95
KNN	86.07	86.64	0.57	85.49	-0.58	86.89	0.82
SVM	91.26	91.95	0.69	91.43	0.17	92.87	1.61
MLP	82.61	83.51	0.90	81.04	-1.57	84.35	1.74
VGG16-SVM	95.67	95.03	-0.64	96.02	0.34	97.04	1.37
Resnet50-SVM	81.82	80.40	-1.42	82.23	0.41	86.40	4.58
Densenet121-SVM	91.10	90.31	-0.78	91.51	0.41	94.31	3.22
MobeilnetV1-SVM	92.87	88.50	-4.37	90.60	-2.27	93.50	0.63
Resnet18-softmax	95.10	97.11	2.01	96.16	1.06	95.88	0.78
Resnet18-SVM	97.28	96.62	-0.66	96.35	-0.93	97.86	0.58

TABLE X
RECOGNITION RESULTS (40% OF THE ORIGINAL TRAINING SET)

Method	Recognition accuracy (%)						
	Original	Original+ CGAN	Δ	Original+ CDCGAN	Δ	Original+ CWDCGAN	Δ
Logistic	78.96	76.30	-2.66	80.54	1.58	80.90	1.94
KNN	78.32	80.59	2.27	80.09	1.77	80.59	2.27
SVM	87.63	89.57	1.94	87.42	-0.21	90.48	2.85
MLP	79.64	78.40	-1.24	80.88	1.24	81.12	1.48
VGG16-SVM	93.40	92.63	-0.77	94.81	1.40	95.50	2.10
Resnet50-SVM	82.03	77.38	-4.65	80.29	-1.74	84.20	2.17
Densenet121-SVM	89.74	86.81	-2.93	90.79	1.05	92.15	2.41
MobeilnetV1-SVM	89.75	85.37	-4.38	86.44	-3.31	90.35	0.60
Resnet18-softmax	93.62	89.28	-4.34	94.18	0.56	95.83	2.21
Resnet18-SVM	95.74	90.20	-5.54	94.65	-1.09	96.81	1.07

are not conducive to the promotion of recognition performance. Although the augmented training dataset of the CWDCGAN can improve the classification performance, the impact is very slight. Moreover, for the original training set, the Resnet18-SVM has a lower recognition performance than the Resnet18-softmax. We have considered relevant reasons for this: first, the original training set is too complete, so the generated samples contribute little to the diversity of the samples, which leads to a small improvement in the recognition performance of the proposed method. Second, the advantage of SVM is to deal with limited sample data. In the case of a complete original training set, the effect may be not as good as softmax.

Therefore, a series of classification experiments are designed by gradually reducing the proportion of the original training samples for further validation. The experimental results are shown in Tables VIII–XI.

From Tables VII–XI, we can find that as the original training samples go down, the performance of each recognition method continue to decline. However, it is a joy that the recognition accuracy of the Resnet18-SVM is still optimal. Moreover, the proposed SAR image expansion method improves the performance of almost all classifiers.

To more intuitively compare experimental results, these data are arranged into a line chart, as shown in Fig. 14. By comprehensively analyzing, it can be seen that the advanced deep learning models combined with SVM classifier can achieve better recognition results than traditional machine learning methods. Moreover, VGG16-SVM, Resnet18-softmax, and Resnet18-SVM have the most outstanding recognition performance when using the original training set, which is much higher than the other seven recognition methods. The same results are also shown on the extended datasets, which fully illustrate that the VGG16

TABLE XI
RECOGNITION RESULTS (20% OF THE ORIGINAL TRAINING SET)

Method	Recognition accuracy (%)						
	Original	Original+ CGAN	Δ	Original+ CDCGAN	Δ	Original+ CWDCGAN	Δ
Logistic	73.33	70.03	-3.30	75.10	1.77	75.87	2.54
KNN	63.40	65.38	1.98	65.54	2.14	65.81	2.41
SVM	74.24	63.27	-10.97	78.85	4.61	78.70	4.60
MLP	64.96	70.86	5.90	70.12	5.16	75.93	10.97
VGG16-SVM	88.29	80.16	-8.14	89.74	1.44	93.79	5.50
Resnet50-SVM	65.25	58.90	-6.35	62.45	-2.80	74.90	9.65
Densenet121-SVM	81.57	80.59	-0.98	84.75	3.17	88.00	6.42
MobeilnetV1-SVM	83.06	79.78	-3.28	79.39	-3.67	85.78	2.72
Resnet18-softmax	89.03	79.43	-9.60	88.66	-0.37	90.23	1.20
Resnet18-SVM	91.57	81.33	-10.24	90.35	-1.22	95.01	3.44

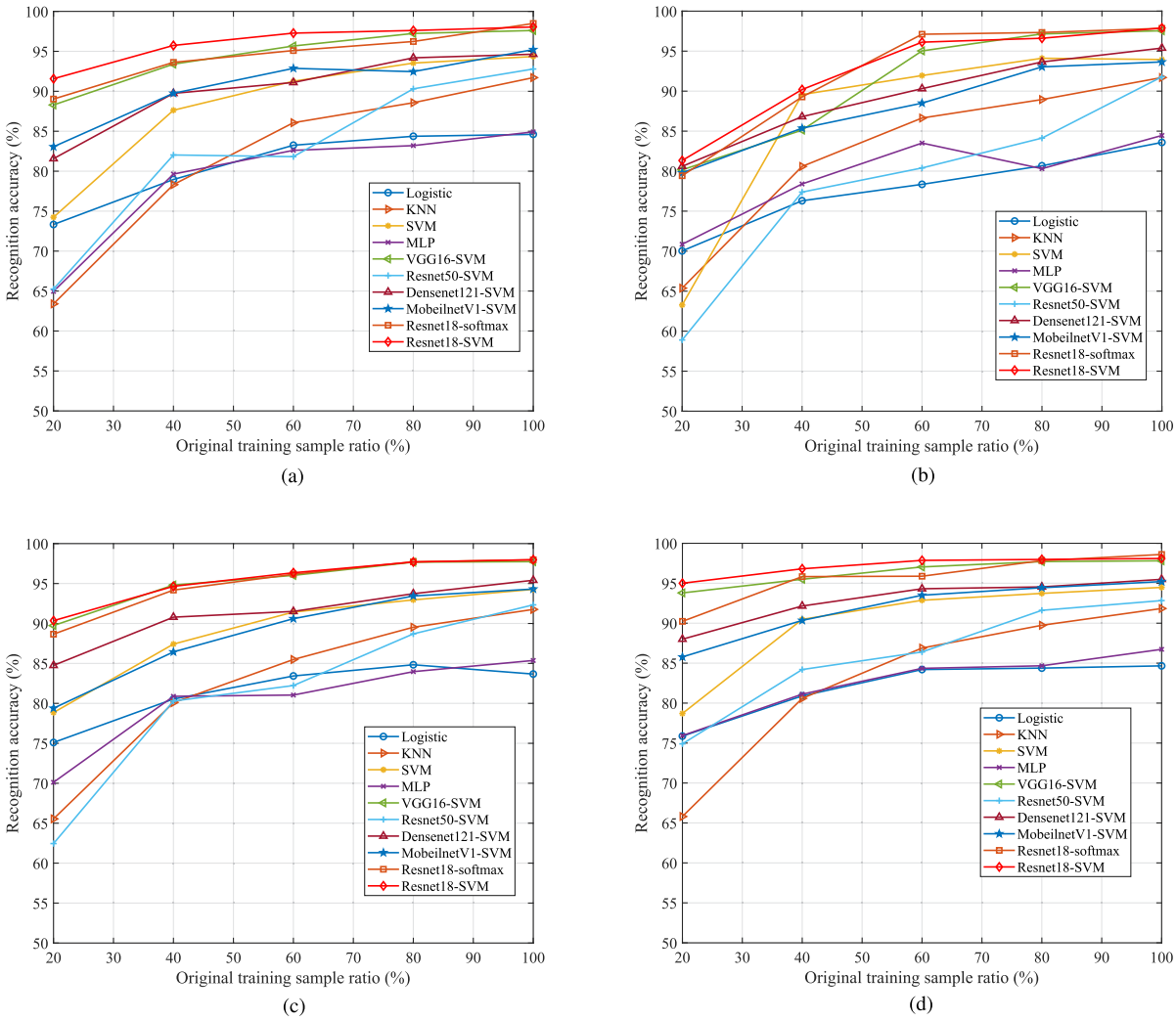


Fig. 14. Recognition accuracy of ten kinds of training sets. (a) Original training set. (b) Dataset expanded by using the CGAN method. (c) Dataset expanded by using the CDCGAN method. (d) Dataset expanded by using the CWDCGAN method. (a) The original training set. (b) The expended training set (CGAN). (c) The expended training set (CDCGAN). (d) The expended training set (CWDCGAN).

and Resnet18 networks can extract features with strong representation ability and beneficial to classification. As expected, the advantages of our method become apparent as the training sample was reduced. This is because SVM has strong generalization capacity in the case of limited data, which is its inherent characteristic. In addition, we found that the comprehensive application of SVM and Resnet18 network can play a better effect than the combination of SVM and VGG16 network. More surprisingly, the proposed recognition method can adapt well to a wide range of training sample ratio. Especially for the dataset

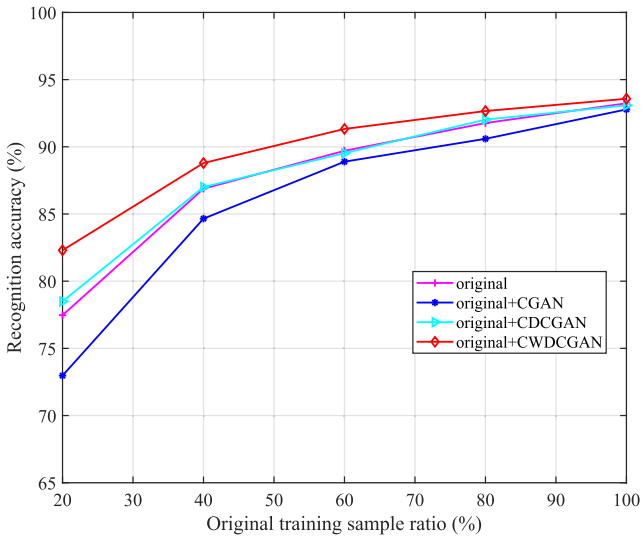


Fig. 15. Recognition accuracy of the different expansion methods. "Original" represents the original training set, "original+CGAN" represents the extended training set by the CGAN, "original+CDCGAN" represents the extended training set by the CDCGAN, and "original+CWDCGAN" represents the extended training set by the CWDCGAN.

expanded by the CWDCGAN, the recognition probability of the proposed classification method can remain above 95%, which implies strong robustness.

The aforementioned experiments mainly verify the beneficial performance of the proposed recognition method on different datasets and original training sets with different scales, and also confirm that the proposed expansion method is beneficial to ATR. In order to alleviate the influence of different recognition methods, the average recognition accuracy of the abovementioned six categories of methods under the four datasets are plotted as a line chart, as shown in Fig. 15.

It can be clearly seen that as the amount of original training data is reduced, the recognition accuracy of the proposed expansion method is always in the leading position, while the CDCGAN is slightly behind. It is disappointing that the extended training set by the CGAN hold the worst performance among all the recognition results, even lower than those with the original training set. When the original training sample ratio is less than 60%, the recognition accuracy is significantly improved on the expanded data based on the CWDCGAN. This also agrees with the conclusion in Section IV-B that the SAR image quality and diversity of our proposed expansion method is better and more favorable for classification.

V. Conclusion

In this article, a SAR image ATR method based on the CWDCGAN and Resnet18-SVM is proposed under limited dataset conditions. In this method, the end-to-end generation model the CWDCGAN is used to learn the image data distribution through the confrontation between the generator and discriminator, and multiple types of extended samples are generated using only one network model. Generated SAR images are able to enhance the diversity of samples. Furthermore, this article utilizes the Resnet18 network to extract deep features with strong representational ability. Finally, the extracted features are sent to the SVM for classification. The Resnet18-SVM recognition method improves the generalization of the recognition model with no need to expert knowledge. According to the experiments on MSTAR, this article proves the quality of the SAR images generated by our proposed method is the best, which makes a great contribution to ATR. Also, a strong advantage of the Resnet18-SVM recognition method is confirmed in the different proportions of training data.

This article presents a SAR image ATR method based on data expansion. In the future work, we will further explore the target SAR image generation with specified angles and the visual description of features in the classification stage.

Acknowledgment

The authors would like to thank MIT's Lincoln Lab for providing the MSTAR dataset for free downloading. The authors would also like to thank the anonymous reviewers for their valuable suggestions.

References

[1] C. Alexandrov, A. Draganov, and N. Kolev, "An application of automatic target recognition in marine navigation," in *Proc. IEEE Int. Radar Conf.*, 1995, pp. 250–255.
[2] L. Li and Z. Liu, "Radar high resolution range profile recognition via dual-SVDD classifier," in *Proc. CIE Int. Conf. Radar*, 2016, pp. 1–4.
[3] L. Li, Z. Liu, and T. Li, "Radar high resolution range profile recognition via multi-SV method," *J. Syst. Eng. Electron.*, vol. 28, no. 5, pp. 879–889, 2017.
[4] L. Li and Z. Liu, "Noise-robust HRRP target recognition method via sparse-low-rank representation," *Electron. Lett.*, vol. 53, no. 24, pp. 1602–1604, 2017.
[5] J. Wang, Z. Liu, L. Ran, and R. Xie, "Feature extraction method for DCP HRRP-based radar target recognition via $m-\chi$ decomposition and sparsity-preserving discriminant correlation analysis," *IEEE Sens.*, vol. 20, no. 8, pp. 4321–4332, Apr. 2020.
[6] J. Wang, Z. Liu, R. Xie, and L. Ran, "Radar HRRP target recognition based on dynamic learning with limited training data," *Remote Sens.*, vol. 13, no. 4, 2021, Art. no. 750.
[7] J. Wang, Z. Liu, T. Li, L. Ran, and R. Xie, "Radar HRRP target recognition via statistics-based scattering centre set registration," *IET Radar Sonar Navigation*, vol. 13, no. 8, pp. 1264–1271, 2019.
[8] P. Guo, Z. Liu, and J. Wang, "HRRP multi-target recognition in a beam using prior-independent DBSCAN clustering algorithm," *IET Radar Sonar Navigation*, vol. 13, no. 8, pp. 1366–1372, 2019.
[9] G. Pengcheng, L. Zheng, and W. Jingjing, "Radar group target recognition based on HRRPs and weighted mean shift clustering," *J. Syst. Eng. Electron.*, vol. 31, no. 6, pp. 1152–1159, 2020.
[10] S. K. Chaturvedi, "Study of synthetic aperture radar and automatic identification system for ship target detection," *J. Ocean. Eng. Sci.*, vol. 4, no. 2, pp. 173–182, 2019.
[11] Y. Sun, Z. Liu, S. Todorovic, and J. Li, "Adaptive boosting for SAR automatic target recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 1, pp. 112–125, Jan. 2007.
[12] H. Liu and S. Li, "Decision fusion of sparse representation and support vector machine for SAR image target recognition," *Neurocomputing*, vol. 113, pp. 97–104, 2013.
[13] T. Zhang and X. Zhang, "A polarization fusion network with geometric feature embedding for SAR ship classification," *Pattern Recognit.*, vol. 123, 2022, Art. no. 108365.
[14] Y. Zhang, X. Guo, H. Ren, and L. Li, "Multi-view classification with semi-supervised learning for SAR target recognition," *Signal Process.*, vol. 183, 2021, Art. no. 108030.
[15] G. Dong, G. Kuang, N. Wang, L. Zhao, and J. Lu, "SAR target recognition via joint sparse representation of monogenic signal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3316–3328, Jul. 2015.

[16] U. Srinivas, V. Monga, and R. G. Raj, "SAR automatic target recognition using discriminative graphical models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 591–606, Jan. 2014.
[17] F. Gao, F. Ma, J. Wang, J. Sun, E. Yang, and H. Zhou, "Semi-supervised generative adversarial nets with multiple generators for SAR image recognition," *Sensors*, vol. 18, no. 8, 2018, Art. no. 2706.
[18] Y. Li, J. Chen, M. Ke, L. Li, Z. Ding, and Y. Wang, "Small targets recognition in SAR ship image based on improved SSD," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process*., 2019, pp. 1–6.
[19] Y. Tian, J. Sun, P. Qi, G. Yin, and L. Zhang, "Multi-block mixed sample semi-supervised learning for SAR target recognition," *Remote Sens.*, vol. 13, no. 3, 2021, Art. no. 361.
[20] Y. Wang, Z. Ding, P. Xu, K. Chen, T. Zeng, and T. Long, "Strip layering diagram-based optimum continuously varying pulse interval sequence design for extremely high-resolution spaceborne sliding spotlight SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6751–6770, Aug. 2021.
[21] S. Feng, K. Ji, L. Zhang, X. Ma, and G. Kuang, "SAR target classification based on integration of ASC parts model and deep learning algorithm," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10213–10225, Oct. 2021, doi: 10.1109/JSTARS.2021.3116979.
[22] L. Zhang et al., "Domain knowledge powered two-stream deep network for few-shot SAR vehicle recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art no. 5215315.
[23] G. Franceschetti, M. Migliaccio, and D. Riccio, "The SAR simulation: An overview," in *Proc. Int. Geos. Remote Sens. Symp., Quan. Remote Sens. Sci. Appl.*, 1995, vol. 3, pp. 2283–2285.
[24] T. Balz, H. Hammer, and S. Auer, "Potentials and limitations of SAR image simulators–A comparative study of three simulation approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 101, pp. 102–109, 2015.
[25] S. Auer, S. Hinz, and R. Bamler, "Ray-tracing simulation techniques for understanding high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1445–1456, Mar. 2010.
[26] T. Balz and U. Stilla, "Hybrid GPU-based single-and double-bounce SAR simulation," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 10, pp. 3519–3529, Oct. 2009.
[27] I. Goodfellow, "NIPS 2016 tutorial: Generative adversarial networks," 2016, *arXiv:1701.00160*. [Online]. Available: https://arxiv.org/abs/1411.1784
[28] F. Gao, Y. Yang, J. Wang, J. Sun, E. Yang, and H. Zhou, "A deep convolutional generative adversarial networks (DCGANs)-based semi-supervised method for object recognition in synthetic aperture radar (SAR) images," *Remote Sens.*, vol. 10, no. 6, 2018, Art. no. 846.
[29] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: https://arxiv.org/abs/1411.1784
[30] X. Yao et al., "GMT-WGAN: An adversarial sample expansion method for ground moving targets classification," *Remote Sens.*, vol. 14, no. 1, 2021, Art. no. 123.
[31] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," 2017, *arXiv:1701.04862*. [Online]. Available: https://arxiv.org/abs/1411.1784
[32] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.
[33] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
[34] Z. Cui, M. Zhang, Z. Cao, and C. Cao, "Image data augmentation for SAR sensor via generative adversarial nets," *IEEE Access*, vol. 7, pp. 42255–42268, 2019.
[35] L. M. Novak, G. J. Owirka, W. S. Brower, and A. L. Weaver, "The automatic target-recognition system in SAIP," *Lincoln Lab. J.*, vol. 10, no. 2, pp. 187–202, 1997.
[36] L. Novak, "State-of-the-art of SAR automatic target recognition," in *Proc. IEEE Int. Radar Conf*., 2000, pp. 836–843.
[37] J. Wissinger, R. Ristroph, J. R. Diemunsch, W. E. Severson, and E. Fruedenthal, "MSTAR's extensible search engine and model-based inferencing toolkit," *Proc. SPIE*, vol. 3721, pp. 554–570, 1999.
[38] S. Chen and H. Wang, "SAR target recognition based on deep learning," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal.*, 2014, pp. 541–547.
[39] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806–4817, Aug. 2016.
[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*., 2016, pp. 770–778.
[41] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[42] T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*., 2018, pp. 8798–8807.
[43] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Pro. IEEE Conf. Comput. Vis. Pattern Recognit*., 2017, pp. 4681–4690.
[44] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: https://arxiv.org/abs/1411.1784
[45] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
[46] W. Malfliet and W. Hereman, "The tanh method: I exact solutions of nonlinear evolution and wave equations," *Phys. Scripta*, vol. 54, no. 6, pp. 563–568, 1996.
[47] A. Salam, A. El Hibaoui, and A. Saif, "A comparison of activation functions in multilayer neural network for predicting the production and consumption of electricity power," *Int. J. Elect. Comput. Eng.*, vol. 11, no. 1, pp. 163–170, 2021.
[48] E. P. Wigner, "On the distribution of the roots of certain symmetric matrices," *Ann. Math.*, vol. 67, pp. 325–327, 1958.
[49] T. Tieleman et al., "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.
[50] W. Fan, F. Zhou, Z. Zhang, X. Bai, and T. Tian, "Deceptive jamming template synthesis for SAR based on generative adversarial nets," *Signal Process.*, vol. 172, 2020, Art. no. 107528.
[51] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*., 2016, pp. 2818–2826.
[52] J. Luo, H. Qiao, and B. Zhang, "Learning with smooth Hinge losses," *Neurocomputing*, vol. 463, pp. 379–387, 2021.
[53] "MSTAR public targets dataset," Accessed: Nov. 23, 2016. [Online]. Available: https://www.sdms.afrl.af.mil/index.php?collection=mstar
[54] A. Kusk, A. Abulaitijiang, and J. Dall, "Synthetic SAR image generation using sensor, terrain and target models," in *Proc. IEEE 11th Eur. Conf. Synthetic Aperture Radar*, 2016, pp. 1–5.

Jikai Qin (Student Member, IEEE) was born in 1994. He received the B.S. degree in agricultural electrification and automation from the Henan University of Science and Technology, Luoyang, China, in 2016. He is currently working toward the Ph.D. degree in information and communication engineering with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an, China.

His research interests include automatic target recognition based on high-resolution range profile and SAR image.

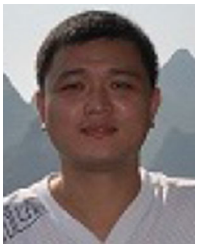
Zheng Liu was born in 1964. He received the B.S. degree in radio technology from the Shaanxi Institute of Technology, Xian, China, in 1985, and the M.S. and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, in 1991 and 2000, respectively.

He is currently a Professor, the Doctoral Director, and the Vice Director of the National Laboratory of Radar Signal Processing, Xidian University. His research interests include the theory and system design of radar signal processing, precision guiding technology, and multisensor data fusion.

Lei Ran was born in Shandong, China, in 1989. He received the B.S. degree in electronic information science and technology from the Qingdao University of Technology, Qingdao, China, in 2012, and the Ph.D. degree in pattern recognition and intelligent system from Xidian University, Xi'an, China, in 2018.

In 2018, he was selected into the National Postdoctoral Innovative Talent Support Program. In 2019, he went to the University of Birmingham as a Visiting Scholar. He is currently an Associate Professor with the National Laboratory of Radar Signal Processing, Xidian University. He has presided over six projects including the National Natural Science Foundation of China (NSFC) Equipment Pre-Research Postdoctoral Fund, participated in a number of vertical and horizontal projects as the backbone of scientific research, authored or coauthored more than ten academic papers, and applied for many invention patents. His research interests include radar imaging on high-speed platforms, multistation radar imaging, multichannel/array radar imaging and target detection.

Dr. Ran was awarded the Excellent Doctoral Dissertation of Shaanxi Province in 2020.



Rong Xie received the B.S. degree in electronic and information engineering, the M.S. degree in signal and information processing, and the Ph.D. degree in pattern recognition and intelligent system from Xidian University, Xi'an, China, in 2003, 2006, and 2011, respectively.

He is currently an Associate Professor with the National Laboratory of Radar Signal Processing, Xidian University. His research interests include multiple-input multiple-output radar signal processing, target motion parameter estimation, and real-time implementation.

Junkui Tang received the B.S. degree in communication engineering from Henan University, Henan, China, in 2017. He is currently working toward the Ph.D. degree in pattern recognition and intelligent system with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an, China.

His research interests include array radar signal processing and multichannel forward-looking radar imaging.

Zekun Guo (Student Member, IEEE) received the M.S. degree in mechanical engineering in 2020 from Xidian University, Xi'an, China,where he is currently working toward the Ph.D. degree in information and communication engineering with the National Laboratory of Radar Signal Processing.

His research interests include pattern recognition and radar intelligent information processing.