IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 13, 2019, accepted March 23, 2019, date of publication March 27, 2019, date of current version April 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2907728

Image Data Augmentation for SAR Sensor via Generative Adversarial Nets

ZONGYONG CUI[1], (Member, IEEE), MINGRUI ZHANG[1], (Student Member, IEEE), ZONGJIE CAO[2], (Member, IEEE), AND CHANGJIE CAO[1]

[1]School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Center for Information Geoscience, School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Zongjie Cao (zjcao@ uestc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61801098, and in part by the Fundamental Research Funds for the Central Universities under Grant 2672018ZYGX2018J013.

ABSTRACT As a mission-critical sensor, SAR has been applied in environmental monitoring and battlefield surveillance; moreover, SAR target recognition is one of the most important applications of SAR technology. However, in practical applications, the number of samples available for training is relatively small, so the SAR target recognition can be regarded as a small sample recognition problem. One of the main directions to solve the small sample recognition problem is to realize the data augmentation. Therefore, a SAR image data augmentation method via Generative Adversarial Nets (GAN) is proposed in this paper. The method uses Wasserstein GAN with a gradient penalty (WGAN-GP) to generate new samples based on existing SAR data, which can augment the sample number in training dataset. Meanwhile, the sample selection filters are designed to extract the generated samples with high quality and specific azimuth, which can avoid the randomness of the data augmentation, and improve the quality of the newly generated training samples. The experiments based on MSTAR data show that, for three-class recognition problem, when the training sample is only 108, the proposed method can improve the recognition rate from 79% to 91.6%; and for ten-class recognition problem, when the training sample is only 360, the proposed method can improve the recognition rate from 57.48% to 79.59%. Compared with the traditional data linear generation method, the proposed method shows significant improvement on the quantity and quality of the training samples, and can effectively solve the problem of the small sample recognition.

INDEX TERMS Synthetic aperture radar, target recognition, small sample recognition, data augmentation, Generative Adversarial Nets.

I. INTRODUCTION

Synthetic aperture radar (SAR) is a high-resolution imaging radar, which can be worked at all-day and all-weather conditions. As an important way to use SAR technology, SAR image target recognition has a wide range of applications, such as Social Security, Environmental Monitoring and National Defense etc [1]. Although previous researchers have done a lot of work, SAR image target recognition is still a very difficult and challenging research [2].

The current researches for SAR image recognition are mainly based on the methods from machine learning, like k-nearest neighbor (KNN) [3], support vector machine (SVM) [4], convolutional neural network (CNN) [5], incremental learning [6] and other algorithms. These classification methods have high classification accuracy but need a sufficient number of training samples.

In the field of object recognition in optical images, the ImageNet [7] is a widely used dataset, which contains about 22,000 classes and nearly 15 million labeled images. However, data for SAR target recognition is still very limited in the field of SAR target recognition. Insufficient data limits the research of SAR target recognition algorithms and the wide application of SAR technology.

Under contemporary conditions, the SAR target recognition can be regarded as a small sample recognition problem, because the number of samples available for training is relatively small. To solve the problem of limited SAR image data,

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

it is important to increase the training data rationally via data augmentation.

Data augmentation is a ubiquitous technology that can increase the size of training dataset through specific data transformations. The current data augmentation methods are mainly include three directions: (1) Geometric data augmentation, like Rotation/reflection [8], [9], flip, zoom, shift [10], [11], and scale; (2) Pixels transformation data augmentation, like color jittering, noise added [12], [13]; (3) Linear synthesis.

Dealing with the SAR image target recognition in lack of pose images, it proposed increasing the pose coverage of the training dataset with the pose image synthesis. Based on a small number of known pose images, [14] uses a sparse model to linear synthetic SAR images of specific azimuth angles through a few images with known azimuth angles. However, we found that some linear synthetic images will reduce the recognition rate.

While it is often easy for domain experts to specify individual transformations, constructing and tuning the more sophisticated compositions typically to achieve state-of-the-art results is a time-consuming manual task in practice [15]. SAR image simulators based on computer-aided drawing models play an important role in SAR applications.

RaySAR is a 3D SAR simulator based on ray tracing methods, which means an enhanced version of the open source software POV-Ray used to simulate radar signals, namely azimuth, distance and azimuth elevation [16]. CohRaS$^{@}$ (Coherent Ray Tracing SAR Simulator) is a SAR simulator based on ray tracing methods [17]. The ray tracer itself is based on the concepts developed by Amananatides and Woo (1987). The simulator is used to simulate small scenes with high resolution to create training data for the classifier and sample data for image analyst training. A real-time SAR image simulation system, which is called SARViz is proposed in [18]. SARViz uses a rasterization method implemented on a GPU that allows for very fast simulations, but has certain limitations in geometry and radiation accuracy.

However, simulators accuracy in these methods is easy to be influenced by geometric accuracy and simplification in the process of electromagnetic computation. Therefore, we need to turn our attention to the field of optical image processing again.

Currently, generating natural images via deep learning methods are a research hotpot. Generative adversarial networks can generate the samples which are similar to training samples, by playing games between generative model and discrimination model. However, training of GAN is difficult. The loss of generator and discriminator can ŕt indicate the training process. And the generated samples lack diversity, even sometimes the generated images are full of noise and difficult to understand. A method of Laplacian pyramid expansion shows its ability to generate the high-quality images [19], but the goal function seems to be unstable, mainly due to the introduction of noise when linking multiple models.

Many papers try to improve this problem. Energy Based GAN (EBGAN) aims to model the discriminator D(x) as an energy function [20]. EBGAN also uses its discriminator as an automatic encoder with an error per-pixel. Deep convolutional generative adversarial networks (DCGAN) relied on experimental enumeration of discriminators and generators to find a better set of network architectures [21]. However, it did not fundamentally solve the problem. In [22] an end-to-end GAN model was developed for SAR image simulation-based directly on real images, but the results are not ideal.

Wasserstein GANs (WGAN) introduces a loss function which can be used as a measure of convergence [23]. WGAN uses the Wasserstein distance to measure the distance between the distribution of generated data and the real data. It has made great progress in training stability. But under certain conditions, the convergence can't be satisfied and some generated samples are with low quality.

As the improved version of WGAN, the Wasserstein GAN with a gradient penalty (WGAN-GP) is proposed with implementing Lipschitz constraints instead of the weight clipping of WGAN [24]. This method converges faster and can generate higher quality samples than WGAN with weight clipping.

Based on the WGAN-GP [24], this paper explores the role of generating samples and expands the training images for SAR target recognition. However, the SAR target images have the characteristics of attitude sensitivity, which means that, the SAR images of the same target but under different azimuths will have large differences.

An obvious but important conclusion has been obtained: when recognizing an unknown sample, only the training samples, whose azimuth angles are similar and the labels are the same with the unknown sample, have the major impact [25]. Therefore, it is difficult to use the image or template at a given azimuth to identify the same target at different azimuth angles. It needs the training dataset containing as many target images in all azimuths as possible. So in this paper, the following contributions have been achieved:

(1) Generate the SAR images through WGAN-GP;

(2) Use a binary classification SVM classifier as an image filter to obtain the samples with high quality;

(3) Design an azimuth discriminator to generate the SAR samples with specific azimuth.

The rest of this paper is organized as follows. Section 2 describes the details of the related GAN and WGAN-GP models. Section 3 shows the whole model of the proposed method, and in Section 4 the images generated experiments and recognition experiments are shown. Section 5 is the conclusion of this paper.

II. THE RELATED THEORY OF WGAN-GP

GAN is a new framework, which can generate the new samples through the adversarial process. Two models can be trained at the same time, the generation model G is for capturing data distribution and the discriminant model D is for estimating the probability of samples from training data. The basic structure of GAN is illustrated in Fig.1.

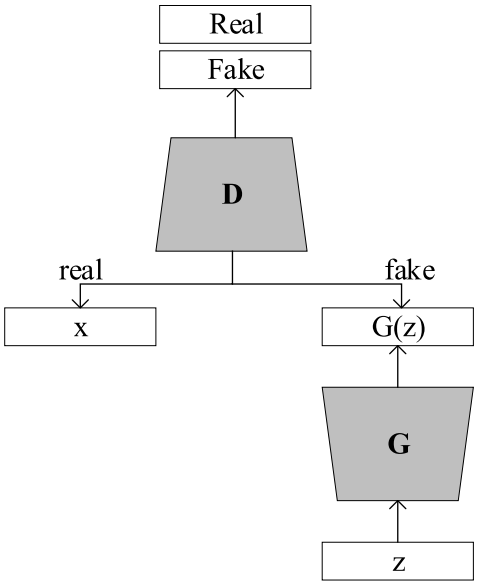

FIGURE 1. The basic GAN structure.

The core principle of GAN is as following:

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

The whole formula includes two items. x represents a real image, z represents the noise that input to the G network, and $G(z)$ represents the image generated by the G network. $D(x)$ indicates the probability that the D network determines whether the image is true (because x is real, so for D, the closer the value is to 1 the better). And $D(G(z))$ is the probability that the D network determines whether the image generated by G is true.

The purpose of training the discrimination model D is to maximize the accuracy of discriminability. When this data is judged to be from real data, the label 1, otherwise, is marked with 0. To the opposite, the purpose of training the Generative model G is to minimize the discrimination accuracy of the discrimination model D. In the training process, GAN adopts a very direct alternating optimization method, which can be divided into two parts. The first part is to fix the discrimination model D, and then optimize the generative model G, so that the accuracy of the discrimination model is reduced as much as possible. The other part is to fix the generative model G, improving the accuracy of the discrimination model.

WGAN-GP mainly improves GAN from the perspective of loss function. After the loss function is improved, WGAN-GP can achieve better performance even at the fully connected layer. WGAN-GP improves GAN mainly by:

1) The last layer of the discriminator removes the sigmoid.
2) Generator and discriminator loss does not take log.
3) A new type of lipschitz continuity constraint method-gradient penalty is proposed, which solves the problem of the disappearance of the training gradient explosion.
4) Has faster convergence than standard WGAN and can generate higher quality samples.
5) Provides a stable GAN training method, requires little tuning, and successfully trains multiple GAN architectures for image generation and language models.

A unevenly distributed distribution Z of 100-dimensional is set in this paper, and a new vector after Z-input generative model is generated. Let this new vector to be the fake Image, marked as $D(z)$. Selecting a picture randomly from the SAR image dataset and converts the picture into a vector. Let this vector be the Real Image, marked as x. $D(z)$ or x is the input of the discrimination network.

In WGAN-GP, the loss functions of the generator G and discriminator D are:

$$L(G) = -E_{x\sim P_g}[D(x)] \quad (2)$$

Eq. 2 shows that the generator wants to pull the score of the false sample as high as possible. In the entire sample space, the gradient of the $Lp-norm$ discriminator function $D(x)$ is required to be no greater than a finite constant K:

$$\|\nabla_x D(x)\|_p \le K, \quad \forall x \in \chi \quad (3)$$

That is, when the input sample changes slightly, the score given by the discriminator cannot change too sharply.

In WGAN, the restriction is implemented by weight clipping. WGAN-GP is an improved version after WGAN, mainly to improve the conditions of continuity restrictions, because after clipping the weight to a certain range, such as clipping to $[-0.01, +0.01]$, and most of the weights are found to be on -0.01 and 0.01, which means that most of the weights of the network are only two possible numbers.

For deep neural networks, the forced cutting weights tend to cause the gradients disappear or the gradients explode. In other words, the weights can't get updated or the weight changes greatly with each update, which can easily lead to unstable training. Gradient disappearance and gradient explosion cause the selection of the shear range. If the selection range is too small, the gradient will disappear and the range will be slightly larger. The gradient will become larger after each layer of the network. Gradient explosions occur after multiple layers.

In order to solve this problem and find a suitable way to satisfy the *Lipschitz* continuity condition, WGAN-GP uses a gradient penalty method to satisfy this continuity condition. Since the *Lipschitz* constraint requires that the gradient of the discriminator does not exceed K, then a loss function can be used to satisfy this requirement. That is, to find the discriminator gradient $d(D(x))$ at first, and then establish a norm between K and $d(D(x))$ to achieve a simple loss function design. Noticing that the numerical space of the gradient of D is the entire sample space.

For a dataset that contains both the real data set and the generated image set, the dimension and its height are clearly

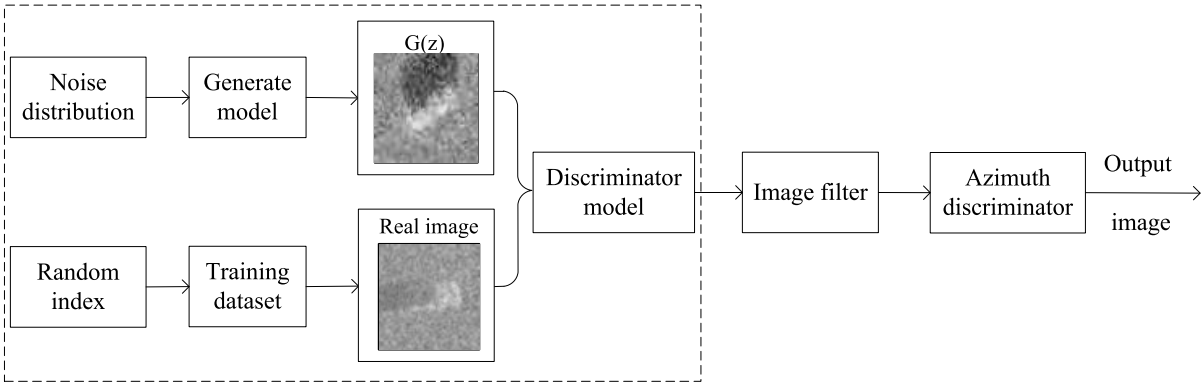

FIGURE 2. The structure of the proposed model, including three parts: samples random generation, samples quality selection and azimuth selection.

unsuitable for calculation. It is not necessary to sample the entire dataset (real images and generated images), as long as the focus is on the areas of the generated sample set, the actual set of samples, and the area sandwiched between them. Specifically, we randomly take a pair of true and false samples, as well as a $0-1$ random number:

$$x_r \sim P_r, x_g \sim P_g, \varepsilon \sim Uniform[0, 1] \tag{4}$$

Then randomly interpolate samples on the lines connecting x_r and x_g,

$$\hat{x} = \varepsilon x_r + (1 - \varepsilon)x_g \tag{5}$$

The loss of WGAN-GP discriminator:

$$L(D) = -E_{x\sim P_r}[D(x)] + E_{x\sim P_g}[D(x)] + \lambda E_{x\sim P_{\hat{x}}}[\|\nabla_x D(x)\|_p - 1]^2 \tag{6}$$

The $P_{\hat{x}}$ is defined as the sampling uniformly along straight lines between pairs of points sampled from the data distribution P_r and the generator distribution P_g. Eq. 6 shows that the discriminator wants to increase the score of the real sample as much as possible, pulling down the score of the false sample.

The WGAN-GP's gradient penalty is only applicable to the true-false sample set area and the intermediate transition zone. However, the gradient norm of the discriminator is directly limited to 1. Therefore, the controllability of the gradient is very strong and it is easy to adjust to a suitable value scale. Since a gradient penalty is applied on each sample independently, the Batch Normalization can't be used in the discriminator's model architecture, because it introduces interdependencies between different samples in the same batch.

III. THE PROPOSED DATA AUGMENTATION MODEL WITH GAN

Based on WGAN-GP, a generative countermeasure network which is suitable for generating usable SAR images is developed in this paper. The structure of the proposed model is shown in Fig.2, including three parts: samples random generation, samples quality selection and azimuth selection. The model cascades a binary classification SVM classifier as an image filter. The images generated by WGAN-GP will be fed into the pre-trained SVM classifier, and the classified images with high confidence can be retained and the unrecognizable images will be discarded. The azimuth discriminator can calculate the target azimuth, and a batch of specific azimuth images can be selected for further training.

A. SAMPLES GENERATED NETWORKS

The generator network of WGAN-GP in this paper follows the structure of DCGAN's generator network, but does not take log for the loss in the network, and adopts the Lipschitz continuity restriction method with gradient penalty. Fig.3 shows the structure of the generator network, in which Deconv is the abbreviation of deconvolutions, bn is the abbreviation of batch normalization, Relu is the abbreviation of Rectified Linear Units, and Tanh represents the hyperbolic tangent function. Relu and Tanh are both activation functions.

Based on the DCGAN discriminator network, we remove the last layer's sigmoid function and do not take log for the loss in the network, as shown in Fig.4. Conv is the abbreviation of convolutional, ln is the abbreviation of layer normalization. LeakyReLu is activation functions.

On the discrimination network D, a value between 0 and 1 is input to represent the probability that the input picture is a Fake Image or Real Image. Originally, both the generative model and the discrimination model are not trained, and the two models became stronger and stronger during the confrontational training and eventually reached the steady state. During this process, the input type of the discrimination model is $(x_{fake}, 0)$ or $(x_{real}, 0)$. At the beginning, both the generative model and the discrimination model are not trained. The two models work together in order to confrontation training. The generation model generates a picture to deceive the discrimination model. Then the discrimination model decides whether the SAR image is true or false.

The generate network is used to generate image samples, and its purpose is to generate an image that is sufficiently mixed the spurious with the genuine, through continuous

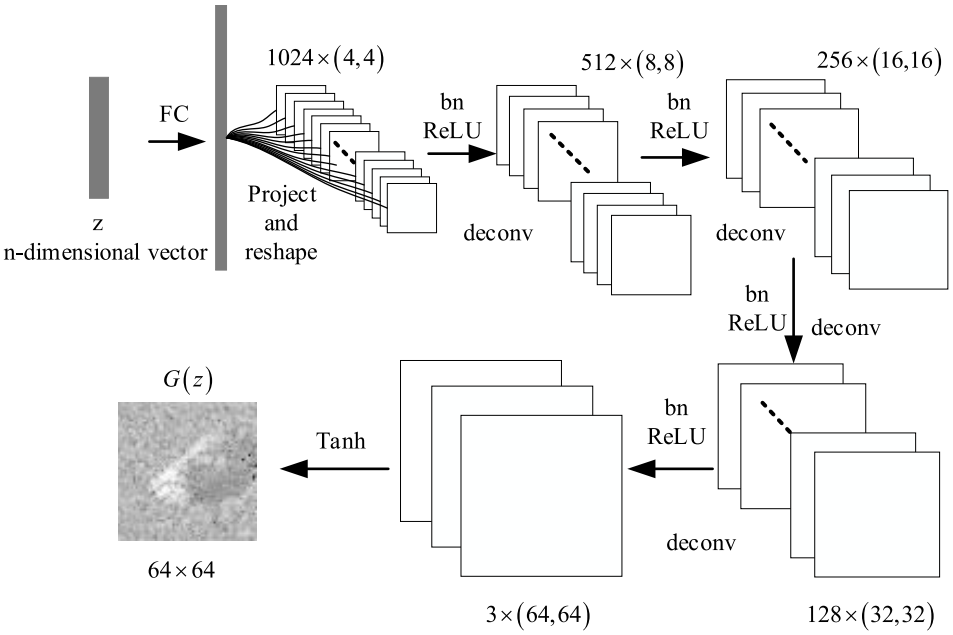

FIGURE 3. The network structure of the generator network. The generation network of WGAN-GP in this section follows the structure of DCGAN's generation network, but the loss in the network is not taken. Deconv represents the transposition convolution, bn represents the batch normalization, and Relu (Rectified Linear Units) and Tanh (the hyperbolic tangent function) are activation functions.

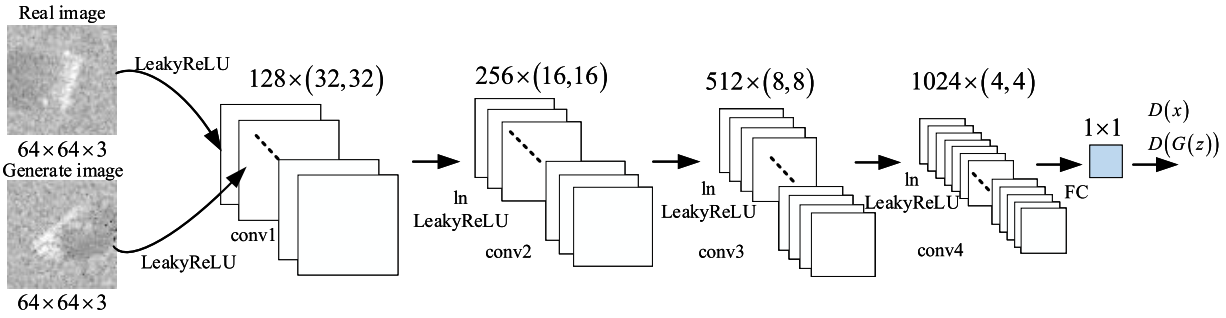

FIGURE 4. The network structure of the discriminator network. The discriminative network of WGAN-GP adopted in this section is improved on the basis of DCGAN's discrimination network, and the last layer of sigmoid function is removed. Conv represents convolution, ln (Layer Normalization) surface standardization, and LeakyReLu is an activation function.

training to maximize the probability that the network discriminates the generated image into a real image. That is, let the discrimination network make error, judges the input image is a real image instead of judges it as a fake image generated by the network. In order to learn the distribution P_g of the generated network on the data x, define an input noise variable z, mapping z into the data space, and then training, generating, and optimizing. The discrimination network is used to determine the true or false of the generated image. During the training process, half of the data input from the discrimination network comes from the real training data, and the other half comes from the fake image generated by the generate network. In the training process, the discrimination network attempts to assign a probability value close to 1 to the real data; and assigns a probability value close to 0 to the pseudo sample generated by the generating network, thereby separating the true and false samples. The structure of the generated network and the discrimination network is shown in Fig.3 and Fig.4.

Inside the algorithm, the loop trains the discrimination network so that it can discern the real sample from the input sample data, and the loop will eventually converge to $D(x) = P_{data}(x)/(P_{data}(x)+P_g(x))$. The parameters of the discrimination network are then fixed and the training generate network is updated. After the update is completed, the gradient of the generation network will direct the generated image to tend to be judged by the discriminating network as the direction of the real sample. Unlike the single iteration update of the original GAN network, the network used in this chapter is trained multiple times. After many training, if the generate network and the discrimination network reach sufficient complexity, the generate network and the discrimination network reach

an equilibrium point:$P_g = P_{data}$, and the probability density function of the generated sample is equal to the probability density function of the real sample. At this time, the generation network can generate an image of ''falsely true'', and the network cannot determine the authenticity of the generation network generated image, that is, $D(x) = 1/2$.

B. SAMPLE SELECTION VIA CLASSIFICATION

In this stage, the model cascades a binary classification SVM classifier as an image filter, and the images generated by WGAN-GP are input to the pre-trained SVM classifier [26]. The classified images with high confidence can be retained and the unrecognizable images will be discarded. The resulting images are pre-screened to increase the availability of the generated images.

In order to improve the SVM classification efficiency, extract the target features of SAR images is extracted by Principal Component Analysis (PCA). PCA is an optimal orthogonal transformation based on the characteristics of the target [27]. Fig.5 shows the specific process of our SVM-based SAR image classification. To ensure the consistency of the experiment, the input for training the SVM discriminator is the same as the input of GAN. Then the samples generated by GAN are feed to the trained SVM, if the recognition rate higher than 90%, these samples can be judged as positive samples, which will be used for further recognition.

C. SAMPLE SELECTION VIA AZIMUTH FILTER

The azimuth discriminator model includes target segmentation, edge extraction, and azimuth discrimination. The specific process is as follows:

1) TARGET SEGMENTATION

The level-set segmentation method is used to segment the SAR image generated by the Generation Adversarial Networks, and represent the extracted target by a binary matrix. By writing the energy function of the curve C as an energy function about the level-set function ϕ [28], the target areas can be obtained by following formula:

$$\begin{aligned} F(c_1, c_2, \phi) = \mu \int_{\Omega} \delta(\phi(x, y))|\nabla\phi(x, y)|dxdy \\ + v \int_{\Omega} H(\phi(x, y))dxdy \\ + \lambda_1 \int_{\Omega} |u_0(x, y) - c_1|^2 H(\phi(x, y)dxdy \\ + \lambda_2 \int_{\Omega} |u_0(x, y) - c_2|^2 (1 - H(\phi(x, y))dxdy \end{aligned} \tag{7}$$

in which, x and y are the pixel of the image, μ, v, λ_1, λ_2 are the fixed parameters, and $\mu \geq 0$, $v \geq 0$, $\lambda_1, \lambda_2 \geq 0$. ϕ is level set function. Ω represents bounded open subset of R^2. u_0 represents a mapping from $\Omega \rightarrow R$, here represents an image consisting of two regions. The constantsc_1 andc_2 are the mean of the pixel points inside and outer the curve C.

The segmentation algorithm steps are as follows:

1) Initialize $\phi^0 = \phi_0$ randomly and n=0, where n is the number of iterations;
2) According to the formula:

$$\begin{aligned} c_1(\phi) &= average(u_0) in\{\phi \geq 0\} \\ c_2(\phi) &= average(u_0) in\{\phi < 0\} \end{aligned} \tag{8}$$

calculate the average c_1 and c_2;
3) According to the iterative formula:

$$\begin{aligned} \frac{\phi_{i,j}^{n+1} - \phi_{i,j}^{n}}{\Delta t} = \delta_h(\phi_{i,j}^{n})[\mu curvature_{ij} - v - \lambda_1(u_{0,i,j} - c_1)^2 \\ + \lambda_2(u_{0,i,j} - c_2)^2] \end{aligned} \tag{9}$$

solve ϕ^{n+1}, where $1 \leq i \leq M$, $1 \leq j \leq N$, $n > 0$ and $M \times N$ is the number of image pixels.

After the above process, the SAR target after segmentation and the target region represented with a binary matrix can be obtained.

2) EDGE EXTRACTION

The edge extraction is performed on the extracted target area. If the number of target points is fewer than the critical value, it is determined as an edge point; otherwise, it is a non-edge point. The number of target points in the surrounding area for each point in the extracted target matrix is calculated, and the target edge information is represented as a binary edge matrix. If the number of target points is fewer than the preset threshold, the point will be regarded as an edge point and assigned a value of 1, otherwise the non-edge point is assigned a value of 0.

3) AZIMUTH CALCULATION

The azimuth discriminator can calculate the target azimuth, and a batch of specific azimuth images can be selected for further training. The minimum bounding rectangle is added to the minimum edge area, and the angle between the north end of the vertical direction and the longest side of the minimum bounding rectangle clockwise is defined as the azimuth of the target.

After obtaining the azimuth angle information, according to the actual need, the samples with specific azimuth angle will be chosen as the new training samples, to improve the recognition performance. The azimuth selection process is shown in Fig.6.

IV. EXPERIMENTS

The experiments are mainly based on the MSTAR dataset. The SAR images used in our experiments include X-band and HH polarized SAR images with a resolution of 0.3 m and are suitable for multiple targets, including BMP2 (tank), BTR70 (armored vehicle), and T72 (tank)., BTR60 (armored vehicle), 2S1 (cannon), BRDM2 (truck), D7 (bulldozer), T62 (tank), ZIL131 (truck) and ZSU23 / 4 (cannon).

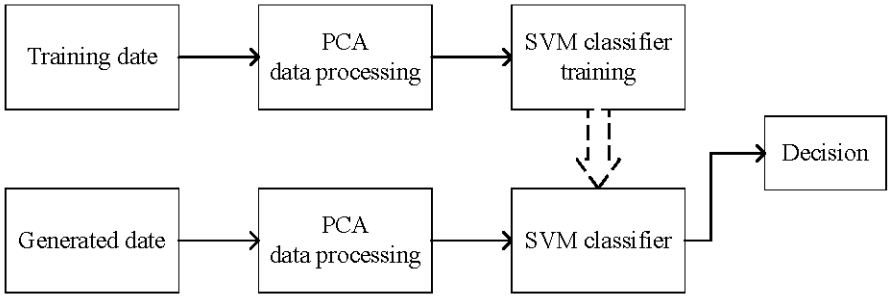

FIGURE 5. SAR generated images discrimination model to select the samples with high quality.

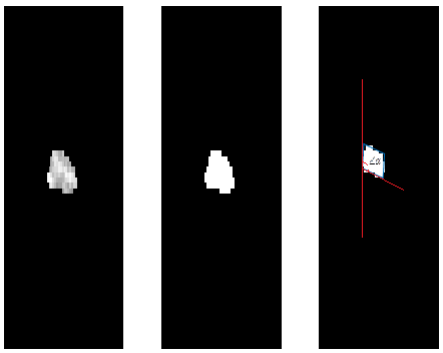

FIGURE 6. The azimuth selection process: segmentation, binaryzation and azimuth extraction.

The images are captured at two different depression angles of 15° and 17°. Similar to the earlier experiments, the images with depression angle of 17° are used as training set, and the images with depression angle of 15° are used for test [25]. All images in the following experiments were cropped by extracting 64 × 64 patches from the center of the image, as shown in Fig. 7.

For the preprocessing of training images, the training images are scaled to the range of $[-1, 1]$ by the activation function. The model is trained with mini-batch stochastic gradient descent. The weighted initialization is a normal distribution with a mean of 0 and a variance of 0.02. The model learning rate is set to 0.0001, and the momentum term is set to 0.5. The original type and sample number of training and testing set is shown in TABLE 1. The network input is generated by 64 n-dimensional vectors z, and output an image of 64 × 64. The number of the filter of the first transposed convolution layer is 64, the convolution kernel is 4 × 4, and the step of the transposition convolution is $s = 2$, after each transposition convolution, the sliding step is 2, the feature map will be expanded to the original $2 \times 2 = 4$ times. For the discrimination network, the input is the image generated by the generate network, and the output is the probability that the input image is a real image. The step of discriminating the network convolution is $s = 2$. That is, after each transposition of the convolution, the sliding step is 2, the feature map is reduced to the original $2 \times 2 = 4$ times. The model diagram of the generate network is shown in Fig.3. The schematic diagram of the discrimination model is shown in Fig.4. The parameter details of the entire model are shown in TABLE 2. deconv is a deconvolution layer, conv is a convolutional layer, bn stands for Batch Normalization, and ln stands for Layer Normalization. The relu (Rectified Linear Units function) and the Tanh hyperbolic tangent function are both activation functions.

TABLE 1. The original type and sample number of training and testing set.

Training set(17°)	Sample number	Testing set(15°)	Sample number
BMP2 c21	232	BMP2 9566	196
BTR70 c71	233	BMP2 9563	195
T72 132	233	BMP2 c21	196
		BTR70 c71	196
		T72 132	196
		T72 812	195
		T72 s7	196
BTR60	256	BTR60	195
2S1	299	2S1	274
BRDM2	298	BRDM2	274
D7	299	D7	274
T62	299	T62	273
ZIL131	299	ZIL131	274
ZSU23	299	ZSU23	274

A. IMAGE AUGMENTATION EXPERIMENTS

The WGAN-GP network model with an azimuth discriminator is proposed to generate a SAR image with orientation. SAR images in MSTAR are used as the training dataset of WGAN-GP, to generate new SAR images by GAN. Some generated samples without selection is shown in Fig. 8.

FIGURE 7. Optical images (top) and SAR images (bottom) of ten kinds of target samples. According to the order: BMP2, BTR70, T72, BTR60, 2S1, BRDM2, D7, T62, ZIL131, and ZSU23/4.

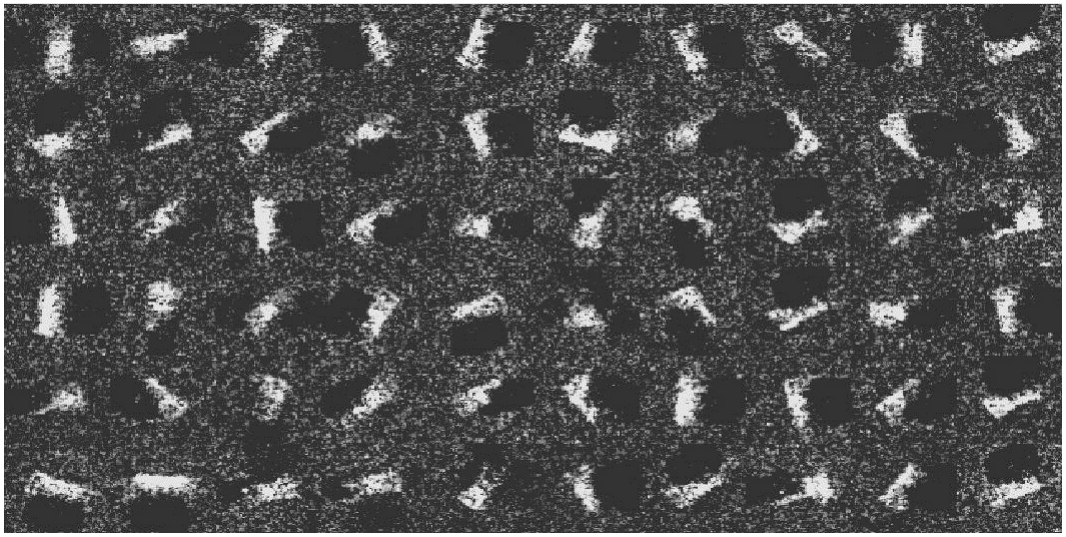

FIGURE 8. The result of WGAN-GP with epoch=500, and the images generated by WGAN-GP but without filter.

TABLE 2. The detail of Model parameter for generating SAR images based on WGAN-GP.

network	Number of layers	type	kernel size	Output feature map dimension	stride	Normalized	Activation function
generation network	0	FC and reshape	\	\	\	\	\
	1	Deconv.	4	4	2	BN	ReLu
	2	Deconv.	4	16	2	BN	ReLu
	3	Deconv.	4	32	2	BN	ReLu
	4	Deconv.	4	64	2	BN	ReLu
discrimination network	0	Conv.	4	32	2	\	LeakReLu
	1	Conv.	4	16	2	\	LeakReLu
	2	Conv.	4	8	2	LN	LeakReLu
	3	Conv.	4	4	2	LN	LeakReLu
	4	FC	\	1	2	LN	\

Then the target segmentation and contour extraction on the generated SAR image are performed. The minimum bounding rectangle of the target contour is built, and the azimuth angle of this target by calculating the angle between the north end of the vertical direction and the clockwise angle of the longest side of the minimum bounding rectangle is calculated. This method of directional generation of SAR images plays a significant role in data expansion of SAR image dataset.

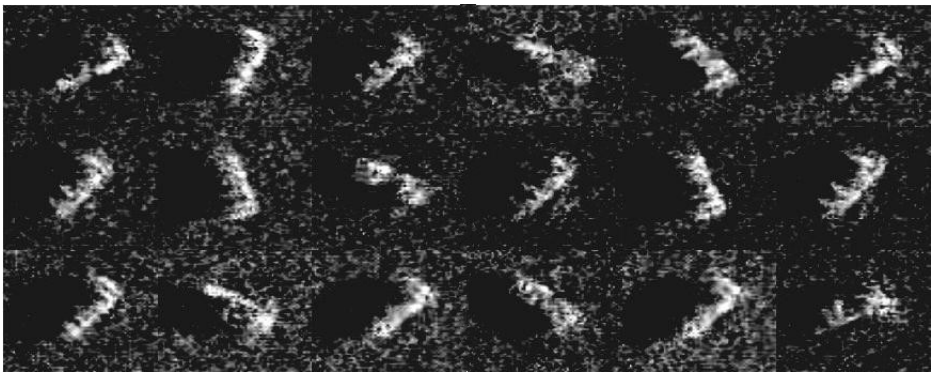

FIGURE 9. The result of WGAN-GP cascading an image filter (epoch=500).

1) SAMPLE SELECTION VIA SVM CLASSIFICATION

In this part, the target BMP-2 9566 in MSTAR dataset is used as the training sample. Through the process in Fig. 5, a series of preliminary filtered images can be obtained, as shown in Fig.9.

a: QUALITY EVALUATION BASED ON OBJECTIVE PARAMETERS

In addition to subjectively visually judging the quality of the generated images, Statistical feature histogram, gradient feature histogram and four representative objective parameters are selected the measure the quality of the selected images: Mean, Variance, Equivalent Number of Looks(ENL), Radiometric Resolution(RR), which are the efficient and intuitive assessment for the quality of SAR images.

In reference [29], to evaluate the quality of SAR images, 4 parameters are chosen from the perspective of image acquisition, including mean, variance, Equivalent Number of Looks and radiation resolution. It is pointed out that these four parameters can initially reflect the relative intensity of the speckle noise of the SAR image and the ability of the SAR system to distinguish the target backscattering coefficient.

We use these four parameters to reflect whether there is a commonality between the generated image and the real image in terms of the relative intensity of the speckle noise and the ability of the SAR system to distinguish the target backscatter coefficient.

At TABLE 3, a real SAR image and a generated image are select randomly to calculate their objective parameters. Form the results, it can be seen that the mean of the generated image is close to the real SAR image. However, the difference between the variance of the generated image and the real image is large. In TABLE 3 ENL is the abbreviation for Equivalent Number of Looks, RR is the abbreviation for Radiance resolution.

TABLE 3. The Objective parameters of generated SAR image.

BMP2	Mean	Variance	ENL	RR
Original image	187.03	406.09	86.14	0.440
Generated by DCGAN	151.28	968.9	23.29	0.87
our method	182.82	405.96	87.82	0.444

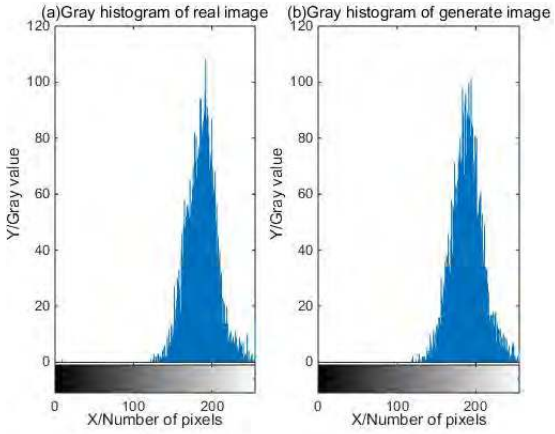

FIGURE 10. Gray histogram of real image and image generated by WGAN-GP based directional expansion method.

b: QUALITY EVALUATION BASED ON GRAY HISTOGRAM AND GRADIENT HISTOGRAM

The gray histogram is a function of the gray level distribution and is a statistic about the gray level distribution of the image. Grey histogram represents the number of pixels in a grey level in an image. Fig.10 shows a comparison of the real image of the BMP2 9566 target type in the MSTAR dataset with the generated BMP2 9566 gray histogram. From the comparison, we can find that the majority of the grey images generated by the WGAN-GP-based directional expansion method appear similar to the real image in the image. At the gray level, the image generated by the directional expansion method based on WGAN-GP has certain credibility.

The gradient histogram is used to calculate the statistical value of the direction information of the local image gradient. In this section, the similarity between the two is compared by calculating the horizontal (x-direction) and vertical (y-direction) gradient histograms of the real sample image and the generated image. Fig.10 shows the real image

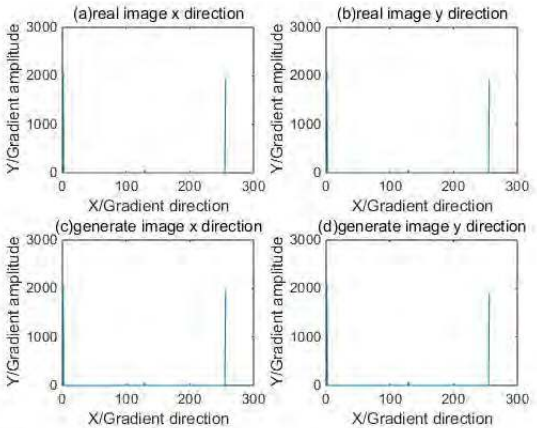

FIGURE 11. Gradient histogram of real image and image generated by WGAN-GP based directional expansion method.

of the BMP2 9566 target type in the MSTAR data set and the horizontal and vertical gradient histogram results of the generated 9566.

Calculated by the Bhattacharyya coefficient: the similarity between the image generated by the WGAN-GP based directional expansion method and the true image x-direction gradient histogram is 0.9982, and the similarity of the y-direction gradient histogram is 0.9984. Combined with the gradient histogram feature results in Fig.10, it can be found that the image generated by the WGAN-GP based directional expansion method is very close to the real image and has high credibility.

c: COMPARE WITH CAD-BASED SIMULATION METHOD

SAR image simulation method based on CAD model is a very popular simulation method. Firstly, we use FEKO to build tank model and LEPO modeling method to realize simulation. Simulation results are shown at Fig.12 and Fig.13. Observing the simulation of the tank model: it can be seen from the simulation results that the simulation result of the tank model has a phenomenon of 'top and bottom inversion'.

However, geometric model-based methods have several drawbacks: First of all, establishing a precise CAD model for each target in the scene in practical applications is a very labor-intensive process. Secondly, in the actual environment, the target may be rusted or coated, which also changes the electromagnetic properties of the target. Thirdly, for non-cooperative targets, researchers cannot measure their surface condition in practice. In addition, the electromagnetic approximation is only effective for large size targets. For small structures in the target, the accuracy of the calculation method is very low.

For the above reasons, it is expected that some information about surface conditions can be obtained directly from an image to achieve image generation.

At last, the samples are selected by azimuth filter. the images generated in previous section are put into the proposed azimuth discriminant model and SAR images with specific angles will be exported. After obtaining the azimuth angle information, the samples with specific azimuth angle will be chosen as the new training samples, to fed to the classifier again. Fig.14 shows the some selected samples with specific azimuth, which are generated via the proposed model.

B. RECOGNITION PERFORMANCE

Different from CNNs based methods [30], our experiments are based on a small sample of SVM recognition experiments.

1) SOC CONDITION RECOGNITION EXPERIMENT

(1)Assume that the depression 17° training sample has an azimuthal spacing of 10° (equivalent to only 36 training samples). At this time, the classifier recognition rate is 79.0134%, and when using GAN to augment the sample with 2 times, 3times, 6 times, 8 times, 9 times and 10 times, the recognition rates are shown in Fig.15.

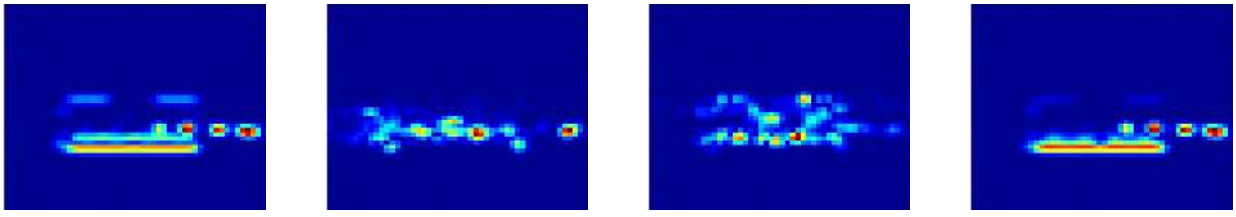

FIGURE 12. Fortress 0°, barrel 5°.

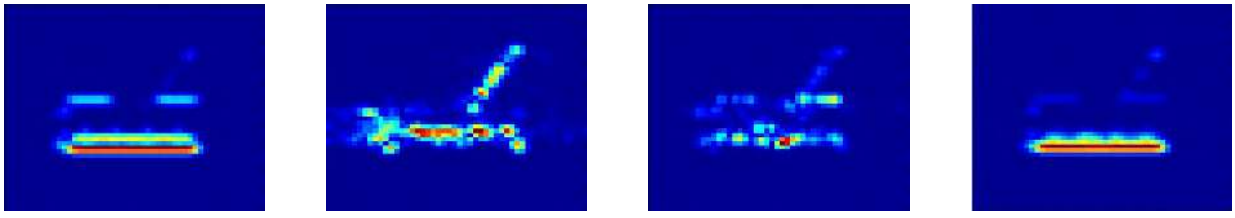

FIGURE 13. Fortress 60°, barrel 7°.

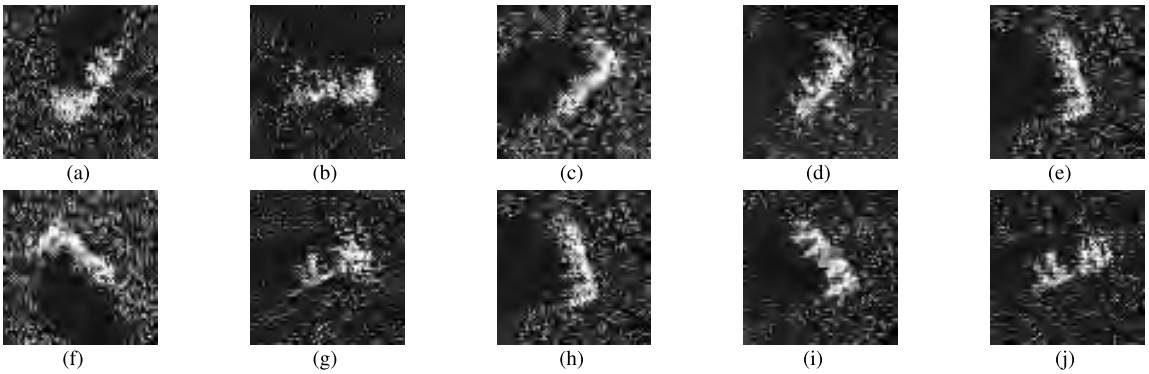

FIGURE 14. Generate specified azimuth SAR image with our method.

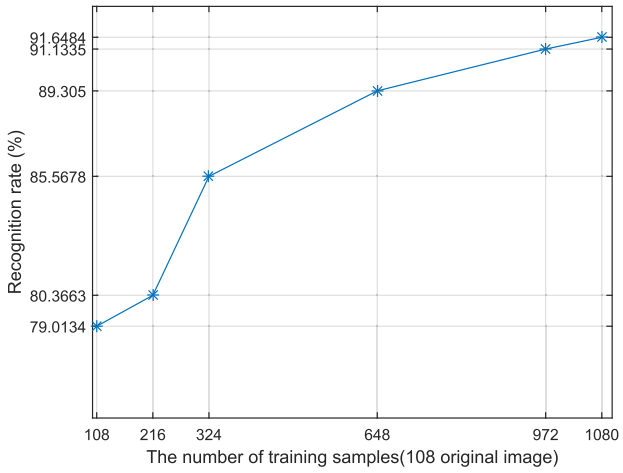

FIGURE 15. Small sample recognition after sample augmentation.

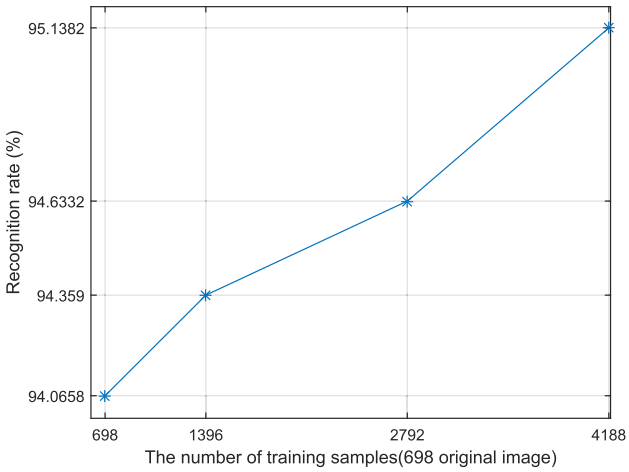

FIGURE 16. Small sample recognition after sample augmentation.

From Fig.15, it can be seen that, in the case of small samples, the recognition rate can be improved by adding the SAR image generated by GAN, which is similar to the actual 698 training samples of 94.0658%, but we only use a small amount of real samples to conduct the confrontation. After the sample is expanded, the recognition rate of more than 91% can be achieved, which is a gratifying progress.

(2)The recognition rate of all the depression 17° training samples was 94.0658%. On this basis, when using GAN to augment the sample 1 times, 3 times, and 5 times, the recognition rate is shown in Fig.16.

(3)To verify the validity and usefulness of the azimuth image we generated. We have expanded the original image from 0-180° for each type of azimuth to 181 images (one for each degree), and compared with the real existing 0-180° images for recognition rate experiments.

In the BMP-2, BTR-70 and T-72 samples, there were 349 training samples with a target azimuth of 0-180° (121 BMP-2 c21, 109 BTR-70 c71, and 119 T-72 132), and, there were 1365 test samples, the recognition rate of SVM three-class experiment was 83.0518%. We only completing BMP-2 c21 to 181, only complete BTR-70 c71 to 181,only complete T-72 132 to 181, and complete all three categories individually. The experiment results show in the TABLE 4, in which OC is the abbreviation for only complete, AC is an abbreviation for All three types are completed. In TABLE 4 and TABLE 5 OC is the abbreviation for only complete, AC is an abbreviation for All three types are completed.

We have expanded the original image of each type of azimuth angles from 0-180° (the image whose azimuth cannot be determined in the actual situation is 180-360°) to 181, and the sample recognition rate can still be improved in the case of only one piece per degree. If you increase the sample per degree, you can still further increase the sample recognition rate.

2) EOC CONDITION RECOGNITION EXPERIMENT

The above experiments were all carried out based on SOC condition. In order to better verify the validity of the

TABLE 4. Recognition experiment after azimuth image expansion via GAN.

Training set(17°)	Sample number	Testing set(15°)	Sample number
Original	349	1365	83.058%
OC c21	409	1365	84.322%
OC c71	421	1365	84.469%
OC 132	411	1365	85.134%
AC	543	1365	87.476%

generated images, this section also carried out the recognition experiment under EOC condition:

(1) Under EOC condition, assume that the depression 17° training sample has an azimuthal spacing of 10° (equivalent to only 36 training samples). The sample was expanded by 1 time, 2 times, 3 times, 4 times, and 5 times using WGAN-GP, the recognition results are shown in Fig.17.

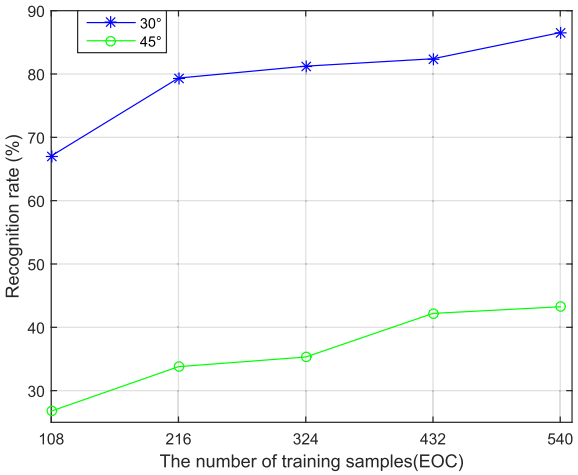

FIGURE 17. Recognition results after small sample expansion (under EOC condition).

The experiment result in Fig.17 shows that the directional expansion method based on WGAN-GP has a positive influence on the recognition result. When the depression is 30°, the recognition rate increased from 67.0557% to 86.535%. And when the depression is 45° the recognition rate increased from 26.7655% to 43.2432%, but due to the noise and other factors in the image under EOC condition, the extent of the increase is not as good as the SOC condition.

(2) Selecting a 0-180° original training sample with a depression of 17° every 5° as a new training set, and expanding each type of training set to 181 pieces using the WGAN-GP-based directional expansion method. Under EOC condition (only 2S1, BRDM-2, ZSU-23-4), 87 training samples with a target azimuth of 0-180° were selected (including 2S1 29, 29 BRDM-2, 29 ZSU-23-4). A total of 1114 test set samples with a depression of 30° and a total of 1147 test set samples with a depression of 45°. The recognition results of the 0-180° completion of the training samples is shown in Fig.18.

The experiment result in Fig.18 shows that the directional expansion method based on WGAN-GP has a positive influence on the recognition result. In the process of complementing the 0-180° image under EOC condition, the recognition rate can be gradually increased. The recognition rate increased from 57.5386% to 70.2513% at depression 30°, and the recognition rate increased from 28.2066% to 40.1613% when the depression was 45°.

C. COMPARISON EXPERIMENTS

1) THREE-CLASS COMPARISON EXPERIMENT

The traditional method of linearly synthesizing azimuth images expands the original image from 0-180° for each type

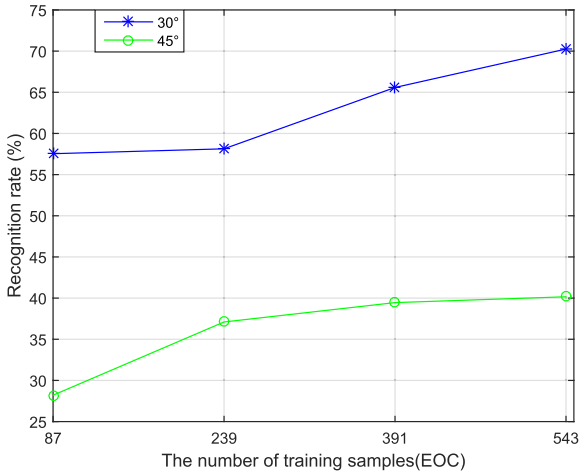

FIGURE 18. Recognition experiment of 0-180° azimuth image completion (under EOC condition).

of azimuth to 181 images (one for each degree), and the result of SVM recognition is as shown in the table TABLE 5:

TABLE 5. Recognition experiment after linear synthetic azimuth image expansion.

Training(17°)	Sample number	Testing(15°)	Sample number
Original	349	1365	83.058%
OC c21	409	1365	82.1978%
OC c71	421	1365	84.1026%
OC 132	411	1365	84.5424%
AC	543	1365	82.4908%

From the experimental results, we can know: We only completing BMP-2 c21 to 181, the recognition rate is 82.1978%; only complete BTR-70 c71 to 181, the recognition rate is 84.4689%; only complete T-72 132 to 181 the recognition rate is 84.1026%; and complete all three categories individually the recognition rate is 82.4908%. The linear synthesis method has a positive effect on the expansion results of the BTR70-C71 and T72-132 categories, and can improve the recognition rate to some extent. However, due to the negative influence of BMP2-C21, the overall recognition rate is reduced. Compared to the linear synthesis method, our method improves the linear synthesis problem and improves the overall recognition rate.

2) TEN-CLASS COMPARISON EXPERIMENT

Assume that the depression 17° training sample has an azimuthal spacing of 10° (equivalent to only 36 training samples). At this time, the classifier recognition rate is 57.4784%, When the sample was expanded by 1 time, 2 times, 3 times, and 5 times using WGAN-GP, linear synthesis method and simple rotating method, the comparison results of the recognition rate are shown in Fig.19.

From the results of the comparative experiment shown at Fig.19, it can be seen that by using the WGAN-GP method to

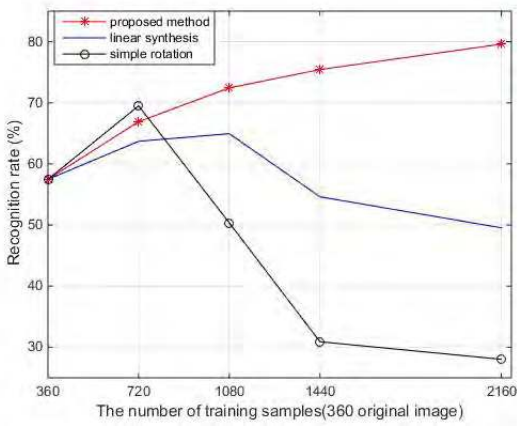

FIGURE 19. Small sample recognition after sample augmentation.

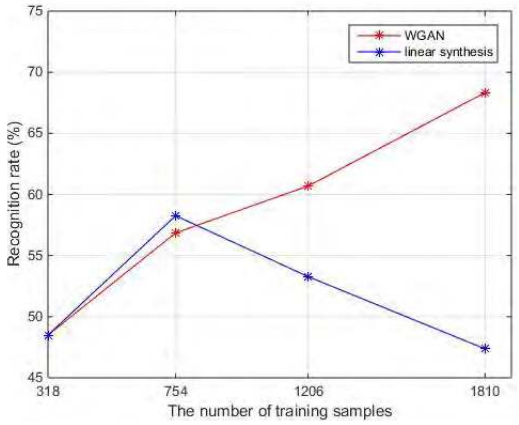

FIGURE 20. Target recognition after azimuth complement.

expand the sample to 1 times, 2 times, 3 times, and 5 times, the recognition rate is on the upward trend. After linear synthesis, the sample was expanded to 1 times, 2 times, 3 times, and 5 times, and the recognition result showed a trend of rising first and then decreasing, and the recognition result was lower than that of WGAN-GP. We also compare the synthesis method based on simple rotating. The results of the recognition experiment can be found at Fig.19.

With the simple rotation method, the recognition result after doubling the sample is higher than the WGAN-GP based method and the linear synthesis method. This indicates that the first rotation may generate some useful samples, which improves the recognition rate to some extent. However, this advantage has not been maintained. After expanding to 2 times, 3 times, and 5 times, the recognition rate has dropped rapidly. This also shows that our method is better than the simple rotation method.

We select each type of azimuth from 0 to 180°. The original picture will be selected every 5°, and will be expanded to 181 frames. There were 318 training samples with a target azimuth of 0-180°, which were selected from the 10 samples (including 35BMP2-c21, 37BTR70-c71, 35 T72-132, 33 BTR60, 2S1 29, BRDM-2 29, D7 32, T62 30,ZIL-131 29,ZSU-23-4 29). There were 2427 test samples, and each type of sample was expanded to 181 by WGAN-GP and linear synthesis. The recognition results of the original sample, 3 types of azimuth complement, 6 types of azimuth complement and 10 types of azimuth complement are shown in Fig.20.

It can be seen from the comparison experiment results Fig.20 that the influence of WGAN-GP on the recognition result is positive, and in the process of complementing the 0-180° image, the recognition rate can be gradually increased. The linearly synthesized samples have a very serious negative impact after a small increase in the recognition rate, which greatly reduces the recognition rate and is lower than the initial recognition rate.

V. CONCLUSION

In this paper, we solve the small sample recognition problem by using WGAN-GP to generate specific azimuth SAR images. The main difficulty encountered in the experimental process of generating images is the instability of the model. The clutters and speckles in SAR images led to failure of the generating. Therefore, a SVM classifier is used to filter out the failed images in real time, which can reduce the redundant calculations and improve the success rate for acquiring a specific azimuth image. Meanwhile, the quality of the generated image is measured from three aspects: the objective parameters, the accuracy of the binary classifications, and the accuracy of the three classifications. The computational results show that the mean of the generated image is close to the real SAR image, and the generated images can improve the recognition efficiency of limited sample target recognition.

Experiments based on MSTAR data show that, for three-class recognition and ten-class recognition problems, the proposed method can improve the recognition rate from 79% to 91.6% and 57.48% to 79.59% respectively, which verify the effectiveness of the proposed method. In the future, the authors will try to apply this method to other mission critical sensors, to solve the problem of data lacking.

REFERENCES

[1] J. Wu, Z. Sun, H. An, J. Qu, and J. Yang, "Azimuth signal multichannel reconstruction and channel configuration design for geosynchronous spaceborne-airborne bistatic SAR," *IEEE Trans. Geosci. Remote Sens.*, to published.

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

[3] N. D. Cagatay and M. Datcu, "Scene recognition based on phase gradient InSAR images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 5162–5166.

[4] S. Ding, X. Nie, H. Qiao, and B. Zhang, "Online classification for SAR target recognition based on SVM and approximate convex hull vertices selection," in *Proc. 11th World Congr. Intell. Control Automat.*, Jul. 2014, pp. 1473–1478.

[5] D. A. E. Morgan, "Deep convolutional neural networks for ATR from SAR imagery," *Proc. SPIE*, vol. 9475, p. 94750F, May 2015.

[6] S. Dang, Z. Cui, Z. Cao, and N. Liu, "SAR target recognition via incremental nonnegative matrix factorization," *Remote Sens.*, vol. 10, no. 3, p. 374, 2018.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[8] X. Fu and Y. Wan, "Accurate image rotation using DCT transformation," in *Proc. IEEE Adv. Inf. Technol., Electron. Automat. Control Conf. (IAEAC)*, Dec. 2015, pp. 537–541.
[9] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. ICDAR*, vol. 3, Aug. 2003, pp. 958–962.
[10] A. J. R. Joe and N. Rama, "Scaling transform methods for compressing a 2D graphical image," *Adv. Comput. Int. J.*, vol. 4, no. 2, pp. 41–50, 2013.
[11] T. Zhang, J. Liang, Y. Yang, G. Cui, L. Kong, and X. Yang, "Antenna deployment method for multistatic radar under the situation of multiple regions for interference," *Signal Process.*, vol. 12, pp. 292–297, Feb. 2018.
[12] Y. Yang, *et al.*, "Deployment of multistatic radar system using multi-objective particle swarm optimisation," *IET Radar, Sonar Navigat.*, vol. 12, no. 5. pp. 485–493, 2018.
[13] J. Wu, W. Pu, Y. Huang, J. Yang, and H. Yang, "Bistatic forward-looking SAR focusing using $\omega-\kappa$ based on spectrum modeling and optimization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4500–4512, Nov. 2018.
[14] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 364–368, Mar. 2016.
[15] A. J. Ratner, H. Ehrenberg, Z. Hussain, J. Dunnmon, and C. Ré, "Learning to compose domain-specific transformations for data augmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3239–3249.
[16] S. J. Auer, "3D synthetic aperture radar simulation for interpreting complex urban reflection scenarios," Ph.D. dissertation, Technische Univ. München, München, Germany, 2011.
[17] H. Hammer and K. Schulz, "Coherent simulation of SAR images," *Proc. SPIE*, vol. 7477, Sep. 2009, Art. no. 74771G.
[18] 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.
[19] E. L. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a Laplacian pyramid of adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1486–1494.
[20] J. Zhao, M. Mathieu, and Y. LeCun. (2016). "Energy-based generative adversarial network." [Online]. Available: https://arxiv.org/abs/1609.03126
[21] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. Image Graph.*, Sep. 2017, pp. 97–108.
[22] J. Guo, B. Lei, C. Ding, and Y. Zhang, "Synthetic aperture radar image synthesis by using generative adversarial nets," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1111–1115, Jul. 2017.
[23] M. Arjovsky, S. Chintala, and L. Bottou. (2017). "Wasserstein GAN." [Online]. Available: https://arxiv.org/abs/1701.07875
[24] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5769–5779.
[25] Z. Cui, Z. Cao, J. Yang, J. Feng, and H. Ren, "Target recognition in synthetic aperture radar images via non-negative matrix factorisation," *IET Radar Sonar Navigat.*, vol. 9, no. 9, pp. 1376–1385, 2015.
[26] C. Tison and N. Pourthie, and J.-C. Souyris, "Target recognition in SAR images with support vector machines (SVM)," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2007, pp. 456–459.
[27] S. Wang and Z. He, "The fast target recognition approach based on PCA features for SAR images," *J. Nat. Univ. Defense Technol.*, vol. 30, no. 3, pp. 136–140, 2008.
[28] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
[29] Q. Zhang, "Quality assessment and target recognition in SAR image," Ph.D. dissertation, Dept. Electron. Eng. Inf. Sci., Univ. Sci. Technol. China, Hefei, China, 2011.
[30] 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. 1–12, Aug. 2016.

ZONGYONG CUI (S'13–M'15) received the B.E. and Ph.D. degrees in signal and information processing from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2007 and 2015, respectively.

From 2013 to 2014, he was a Visiting Student with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. He is currently a Lecturer with the School of Information and Communication Engineering, UESTC. His research interests include radar image processing, target recognition, and machine learning.

Dr. Cui is a Reviewer of the IEEE Geoscience and Remote Sensing Letters, the *IET Radar, Sonar & Navigation*, and the *Journal of Electronic Imaging*.

MINGRUI ZHANG (S'16) received the B.E. degree in electronic and information engineering from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2016. She is currently pursuing the M.S. degree in electronics and communication engineering with the University of Electronic Science and Technology of China (UESTC), Chengdu, China.

Her research interests include data augmentation, image processing, and deep learning.

ZONGJIE CAO (M'10) received the B.E. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1999 and 2005, respectively.

From 2006 to 2008, he was a Postdoctoral Researcher with the Communication and Information System Postdoctoral Center, University of Electronic Science and Technology of China (UESTC). In 2008, he joined the School of Electronic Engineering, UESTC, where he is currently a Professor with the School of Information and Communication Engineering. He has authored or coauthored over 50 papers. His current research interests include radar signal processing, synthetic aperture radar imaging, and target recognition.

Dr. Cao is a Reviewer of several international journals and conferences, such as the IEEE Transactions on Geoscience and Remote Sensing, the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, the IEEE Geoscience and Remote Sensing Letters, the *IET Radar, Sonar & Navigation*, and *Remote Sensing*.

CHANGJIE CAO received the B.E. degree in information and computing science from the Chengdu University of Technology, Chengdu, China, in 2017. He is currently pursuing the Ph.D. degree in the field of signal and information processing with the University of Electronic Science and Technology of China (UESTC), Chengdu, China.

His research interests include pattern recognition, machine learning, and image processing.