Attribute-Guided Generative Adversarial Network With Improved Episode Training Strategy for Few-Shot SAR Image Generation

Yuanshuang Sun, Yinghua Wang, *Senior Member, IEEE*, Liping Hu, Yuanyuan Huang, Hongwei Liu, *Senior Member, IEEE*, Siyuan Wang, and Chen Zhang

***Abstract*—Deep-learning-based models usually require a large amount of data for training, which guarantees the effectiveness of the trained model. Generative models are no exception, and sufficient training data are necessary for the diversity of generated images. However, for synthetic aperture radar (SAR) images, data acquisition is expensive. Therefore, SAR image generation under a few training samples is still a challenging problem to be solved. In this article, we propose an attribute-guided generative adversarial network (AGGAN) with an improved episode training strategy for few-shot SAR image generation. First, we design the AGGAN structure, and spectral normalization is used to stabilize the training in the few-shot situation. The attribute labels of AGGAN are designed to be the category and aspect angle labels, which are essential information for SAR images. Second, an improved episode training strategy is proposed according to the characteristics of the few-shot generative task, and it can improve the quality of generated images in the few-shot situation. In addition, we explore the effectiveness of the proposed method when using different auxiliary data for training and use the Moving and Stationary Target Acquisition and Recognition benchmark dataset and a simulated SAR dataset for verification. The experimental results show that AGGAN and the proposed improved episode training strategy can generate images of better quality when compared with some existing methods, which have been verified through visual observation, image similarity measures, and recognition experiments. When applying the generated images to the 5-shot SAR image recognition problem, the average recognition accuracy can be improved by at least 4%.**

***Index Terms*—Few-shot image generation, generative adversarial network (GAN), meta-learning, synthetic aperture radar (SAR), transfer learning.**

I. Introduction

SINCE generative adversarial networks (GANs) [1] are proposed, various variants of GAN-based models [2], [3], [4], [5] have been proposed for image generation. An important application of the generated images is enriching the initial training set for recognition. Synthetic aperture radar (SAR) target recognition usually requires adequate training data. In practice, SAR images are hard to acquire, and SAR image generation can serve as one solution to deal with the problem of insufficient training data. However, the lack of training data deteriorates the performance of not only the recognition model but also the generative model itself, which results in poor quality of the generated images. Consequently, we mainly focus on few-shot SAR image generation in this article.

For a well-trained GAN-based model, the data distribution of the generated data is trained to be similar to that of the real data. When only a small number of training samples can be obtained for training a GAN-based model, for example, only five samples per category, the generated data may not learn the data distribution well in this situation. The aspect-angle-related image information is severely missing, which results in the poor generalization capability of the trained generative model for generating images under new aspect angles, and the generated images lack diversity. The Moving and Stationary Target Acquisition and Recognition (MSTAR) public dataset [6] and the auxiliary classifier GAN (AC-GAN) [4] are used to illustrate the problem as shown in Fig. 1. The generated images are supposed to be at different aspect angles, and nevertheless, the generated images look similar to the few training data. We can see that the generated images lack the diversity of aspect angles. Therefore, one of the bottlenecks for training a deep generative model is the difficulty of acquiring enough diversified training data. Few-shot image generation is one of the few-shot learning problems [7].

Considering the few-shot image generation problem for optical datasets, there are already some relevant works, which can be divided into three categories [8], i.e., the optimization-based methods [9], [10], the fusion-based methods [11], [12], [13], [14], and the transformation-based methods [15], [16]. The optimization-based methods apply the optimization strategy for the few-shot classification to the few-shot image generation, such as the reptile [17] training strategy and the model-agnostic meta-learning [18] strategy, to improve the performance of the generative model. As for the fusion-based methods, different conditional features, such as the different colors and shapes, can be fused to generate new images, which do not exist in the training set. The transformation-based methods refer to the idea

Manuscript received 18 July 2022; revised 8 December 2022; accepted 17 January 2023. Date of publication 24 January 2023; date of current version 13 February 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61671354, in part by the stabilization support of the National Radar Signal Processing Laboratory under Grant KGJ202206, and in part by the 111 Project. *(Corresponding author: Yinghua Wang.)*

Yuanshuang Sun, Yinghua Wang, Yuanyuan Huang, Hongwei Liu, Siyuan Wang, and Chen Zhang are with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: sunsun5544@126.com; yhwang@xidian.edu.cn; hyy19970610@163.com; hwliu@xidian.edu.cn; wangsiyuan@stu.xidian.edu.cn; czhang_17@stu.xidian.edu.cn).

Liping Hu is with the Science and Technology on Electromagnetic Scattering Laboratory, Beijing Institute of Environmental Features, Beijing 100854, China (e-mail: fox_plh@163.com).

Digital Object Identifier 10.1109/JSTARS.2023.3239633

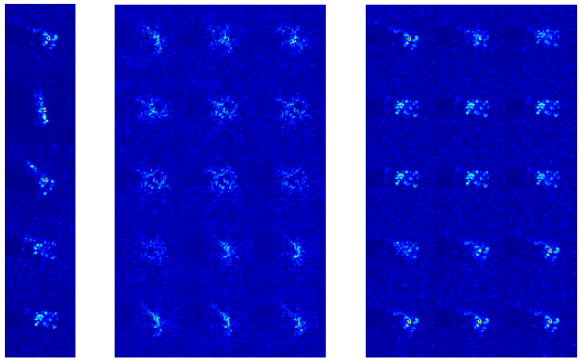


Fig. 1. Generation results of a trained AC-GAN [4] model. The training set contains 5 samples for each of the 3 categories, and thus, a total of 15 samples are used for training. The model is trained via the minibatch gradient descent method. (a) Visualization of the only five training samples of the BMP2 category. (b) Generated images of BMP2 category; trained by 2000 epochs. (c) Generated images of BMP2 category; trained by 20 000 epochs.

that a class label should be invariant to a particular transformation, and then, the transformation can be used to generate new images with the same label. The optimization-based methods have the potential to be directly applied to the few-shot SAR image generation problem. However, unlike optical images, which usually have some obvious visible features, for SAR images, there are no apparent visible features, and even the category information is not distinguishable. The fusion-based methods and the transformation-based methods for the few-shot image generation problem are mainly manually designed for optical images according to the visible features, and they may not be suitable for SAR data. The main difficulty of applying fusion-based and transformation-based methods to SAR data is the manual design of the fusion or transformation ways according to the special characteristics of the SAR target images.

Currently, the research works on few-shot image generation for SAR datasets are rare. There are some research works [19], [20], [21], [22], [23], [24], [25] on the SAR image generation problem but not in the few-shot situation, and the smallest sample size used for training is 90 samples from the 7 classes of the MSTAR dataset in [24], while more samples are used for training in other generative models. In practical applications, the available SAR images for training a deep generative model may be far less than 90. A detailed introduction of these related research works is presented in Section II-C.

To deal with the SAR image generation problem under extremely few samples, like only five samples for each category, we propose an attribute-guided GAN (AGGAN) with an improved episode training strategy. The whole framework is illustrated in Fig. 2, and the main improvements of our methods are marked by the red boxes and introduced from the following three aspects.

1) For the few-shot SAR image generation problem, we design the AGGAN model and the attribute labels of AGGAN to be the category and aspect angle labels to control the generated images. These two kinds of label information are essential for generating images, which can be better mined and used in the few-shot situation, and spectral normalization [26] is also applied to stabilize network training when the training data are insufficient. An earlier version of this model has been reported in [27].
2) The idea of transfer learning [28], [29] is applied in our framework, which aims to improve the diversity of the generated images when lacking training samples. In this article, two different cases of transfer learning are explored. Case A: The source data are the real SAR data with different categories from the target data. Case B: The source data are the electromagnetically simulated data with the same categories as the target data. Considering the need to obtain a large amount of source data for training, the simulated SAR data are relatively easier to obtain than the real SAR data. Thus, we also explore the application of the simulated data for the few-shot image generation problem.
3) An improved episode training method is proposed to increase the sampling probability of the target domain classes, which can be applied to both transfer learning cases. We set all the auxiliary data to another same category label to distinguish them from the target data to improve the simplicity and suitability to fit the learning tasks.

The remainder of this article is organized as follows. Section II briefly introduces the transfer learning and meta-learning ideas, the simulated data, and the existing SAR generative models. Section III briefly reviews the related GAN-based works. Section IV describes the network structure and loss function of our proposed AGGAN in detail. The traditional training strategies and the proposed improved episode training strategy are introduced in Section V. Extensive experimental results based on the MSTAR dataset and a simulated dataset are provided in Section VI. Finally, Section VII provides the conclusion and outlines future works.

II. Related Work

A. Brief Introduction of Transfer Learning and Meta-Learning

Transfer learning relaxes the hypothesis that the training data must be independent and identically distributed (i.i.d.) with the test data [29]. Therefore, transfer learning is often used to solve the problem of insufficient training data. We follow the definition given in [28] and give the definition of transfer learning as follows:

Given a target task T_t in the target domain D_t and the source task T_s in the source domain D_s, transfer learning aims to improve the performance of T_t by the learning of T_s, where $D_s \neq D_t$ and/or $T_s \neq T_t$ and the data size in the source domain D_s is often much larger than that in the target domain D_t. If the knowledge transfer is realized by deep neural networks, the transfer learning becomes deep transfer learning [29].

There are already some works [30], [31], [32], [33] applying the transfer learning methods for few-shot image generation using optical datasets. However, there are few research works applying the transfer learning method for few-shot SAR image generation. In this article, we apply the transfer learning method in our framework, which aims to improve the diversity and

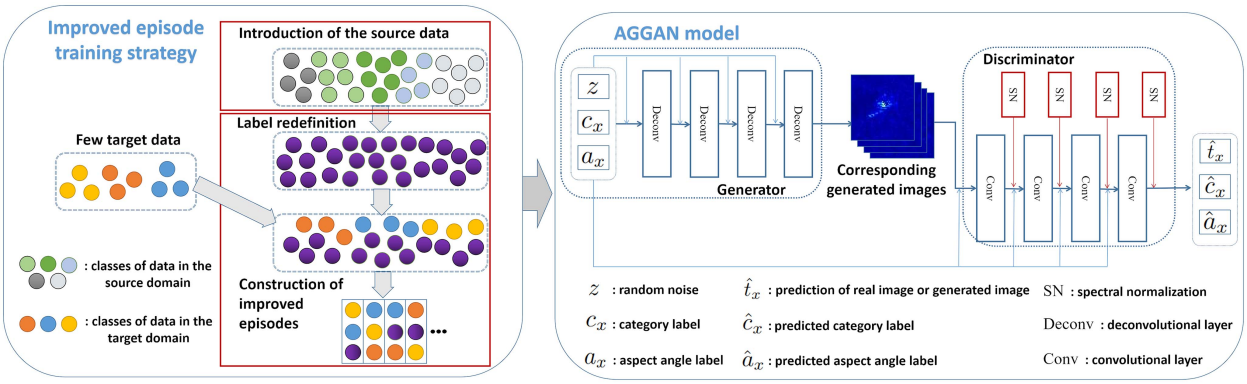


Fig. 2. Illustration of our framework.

quality of the generated images with the help of sufficient source data. Different from transferring the parameters of a pretrained model, we use the meta-learning method to gain experience from the source data.

In essence, for the few-shot learning problem, meta-learning [34] and transfer learning [28], [29] both attempt to improve the performance of the learning tasks in the target domain based on the source data. Meta-learning, also known as learning-to-learn, aims to improve the learning ability of the learning algorithm itself [34]. The meta-learning methods can be used to further improve the performance of the transfer learning methods as well as other problems [34].

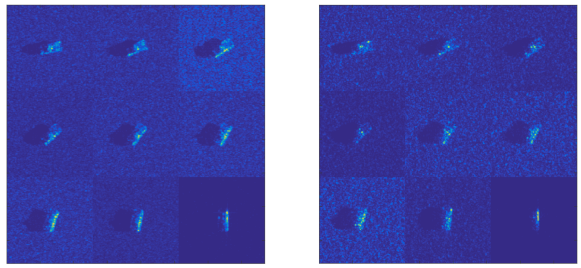
For optical datasets, meta-learning methods have been widely used for few-shot classification tasks [17], [18], [35], [36], [37], and there are also some few-shot image generation works [8], [9] based on meta-learning methods in recent years. To the best of our knowledge, a few research works explore the few-shot SAR image generation problem based on meta-learning methods. The episode training strategy utilized in [35] for few-shot classification is a well-known meta-learning method. We modify the episode training strategy and propose an improved episode training strategy for the few-shot SAR image generation problem.

B. Simulated SAR Data

Compared with the real SAR data, the acquisition cost of the simulated SAR data is lower. The simulated data cannot be used directly for generation or recognition tasks since there is a domain difference between the simulated and real data. To illustrate the differences between the electromagnetically simulated data used in this article and the MSTAR data, pairs of images under the same categories and the same aspect angles are shown in Fig. 3.

Comparing the paired data in Fig. 3, we can see more noises in the background area for the MSTAR data, and the target areas of the simulated data are clearer. In addition, the strong scattering centers of real images and the corresponding simulated images are obviously different. Therefore, there is indeed a distribution gap between the simulated and real images.

There are some research works [38], [39], [40], [41], [42], [43], [44], [45] utilizing the simulated data for the recognition



(a) (b)

Fig. 3. Comparison of the simulated data and the corresponding real data. (a) Simulated images of the BMP2 category. (b) Corresponding real images of the BMP2 category in the MSTAR dataset.

tasks. In [39], the simulated SAR data are used to pretrain a network, and then, the pretrained network is fine-tuned for the recognition tasks on the real SAR data. In [38], [40], and [41], the simulated SAR data are refined and then used for corresponding downstream tasks. In [42], [43], [44], and [45], the extreme situation is considered, which is training fully on the simulated data and testing on the real SAR data, and the distribution gap between the simulated data and the real data is bridged using different methods.

Considering the real-world situation, it is usually challenging to obtain a large amount of source data that follow the similar distribution as the few target data for training. Comparatively, the simulated data is more accessible as the auxiliary source data. To our knowledge, simulated data are rarely used for the generation task. Thus, we explore the usage of simulated SAR data as the source data for few-shot image generation, and the corresponding downstream recognition tasks using the generated data are also explored, which has great significance for practical applications.

C. Existing SAR Generative Models

In this section, the existing SAR generative models are further analyzed, and we make a comparison between the proposed AGGAN and the existing SAR generative models.

The SAR image generative model proposed in [19] is based on the idea of CGAN, with a relatively simple network structure.

A clutter normalization method is proposed to stabilize the training. The conditional information used in [19] is not explicitly stated. The problem of image generation under limited training samples is mentioned, and the quality of the generated images degrades in this situation. In total, 25% of the complete training data are selected according to the specified aspect angles and used to train the model, and there are observable losses in detail of the generated images. According to the sample number of the 10-type MSTAR dataset under the depression angle of 17°, 25% of the whole dataset means about 93 samples for each category.

A SAR adversarial autoencoder network is proposed in [20] and [24]. The architecture of the generator and discriminator follows that of progressive growing of GANs [3]. The category label, the aspect angle label, and a segmentation map are used as conditional information. In [24], the SAR image generation problem in the case of a small number of training samples is considered, and seven classes of target data from the MSTAR dataset are used to evaluate the performance. In total, 90 chips are selected from the seven-class targets to train the model, which implies that there are about 13 samples for each category, and the aspect angles of the training data are chosen at regular intervals, covering the range of 0° to 360°. The results demonstrate its ability to generate SAR images with aspect angle diversity, and finally, the test accuracy of A-ConvNets [46] for target recognition can be boosted by 5.77%. However, an extra method called rotated cropping is introduced to address the challenge of representing the target orientation when there are only a few data for training. Moreover, the semantic segmentation maps are needed as the prior knowledge for image generation in the test phase, which is practically impossible to be obtained.

Label-directed GANs proposed in [21] are based on the ideas of WGAN and CGAN. Only the category labels are used as conditional information for generating images. The recognition results in [21] are achieved by a support vector machine [47] classifier, which is not the state-of-the-art recognition model. The problem of insufficient training data is also mentioned, and 400 training samples from the complete training set, equivalent to about 40 samples for each category, are selected for training. Nevertheless, the generation results in this situation are not presented.

To sum up, we think that the problems discussed in the aforementioned papers are somewhat different from the few-shot image generation problem we considered. The problems mentioned in papers [19] are [21] are more like the "limited data" situations but not the "few-shot" cases, in which there are extremely few training samples, like five samples for each category. The problems mentioned in papers [20] and [24] use much fewer samples but are still not an extremely small case. For SAR datasets, there are already some advanced research works [48], [49], [50], [51], [52] for few-shot SAR target recognition but research on few-shot SAR image generation is rare.

III. Preliminary Knowledge

GANs are famous generative models. The generator and discriminator are essential for a GAN model. The generator takes the random noises as inputs to generate a fake image. The discriminator judges whether an input image is a real image or a fake image. With the adversarial training between the generator and the discriminator, finally, the generated images become indistinguishable from real images for the discriminator, which means that the distribution of generated data has nearly matched the distribution of the training data. Simple illustrations of CGAN and AC-GAN are shown in Fig. 4, left and right, respectively. We briefly introduce them in the following sections.

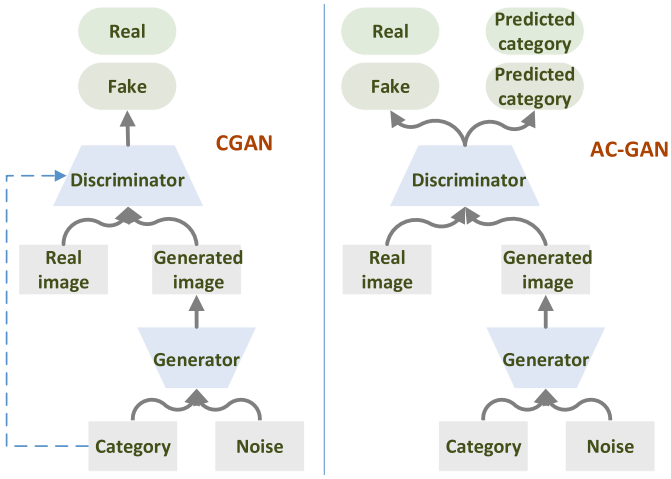


Fig. 4. Illustration of CGAN and AC-GAN, drawn by referring to [2], [4], and [53].

A. CGAN

CGAN is one of the earliest variants of GANs, which controls the content of the generated images by introducing an additional condition, and the conditional information includes but is not restricted to category labels.

As we can see in Fig. 4 left, for the generator G, the noise vector z and the category label y are concatenated as the input, and the output is the corresponding generated image. For the discriminator D, the category labels and images are both used as the input while the output is a single scalar representing the probability that the image is a real image. In terms of network structure, the multilayer perceptrons (MLPs) structure is used in the original CGAN model. While other network structures, like convolutional layers, can also be used to implement CGAN. The objective function of CGAN is expressed as follows:

$$\min_G \max_D V(D,G) = E_{x_{\text{real}}}[\log D(x,y)] + E_{x_{\text{fake}}}[\log(1 - D(G(z,y),y))] \quad (1)$$

where $V(D,G)$ represents the loss function of the entire CGAN network. The real image is denoted as x_{real} while the generated image is denoted as $x_{\text{fake}} = G(z,y)$. $D(x,y)$ is the probability that the sample x comes from the real data distribution rather than the generated data distribution.

B. AC-GAN

AC-GAN also controls the generated images by introducing the condition information. Meanwhile, it also completes an auxiliary classification task, which is shown in Fig. 4 right. For the generator G, the noise vector z and the category label c_x are concatenated as the input of the generator, and the output of the generator is the corresponding generated image, i.e., there is $G(z, c_x) = x_{\text{fake}}$. For the discriminator D, the real image x_{real} or generated image x_{fake} are used as the input. The output gives the prediction probability $\hat{t}_x$ that the image is real and the predicted category label $\hat{c}_x$, i.e., $D(x) = [\hat{t}_x, \hat{c}_x]$. The objective function has two parts: 1) the loss $L_t(x)$ for determining whether the input sample x is generated or real and 2) the loss $L_c(x)$ for determining the category label

$$L_c(x) = E_x \left[\log P(\hat{c}_x = c_x))\right] \tag{2}$$

$$L_t(x) = E_{x_{\text{real}}}[\log(\hat{t}_x)] + E_{x_{\text{fake}}}[\log(1 - \hat{t}_x)] \tag{3}$$

D is trained to maximize $L_c(x) + L_t(x)$ while G is trained to maximize $L_c(x) - L_t(x)$.

Compared with CGAN, there are three aspects of modification in AC-GAN. First, there is an auxiliary classification task to determine the image category. Second, the category label is not used as the input for the discriminator. And third, in terms of network structure, the deep convolutional network is applied in the AC-GAN model.

C. Spectrally Normalized GAN

One of the challenges in the study of GANs is the instability of its training. The training of the discriminators affects the performance of GANs crucially, and the derivative of the optimal discriminator can be unbounded or even incomputable theoretically [59], which is not good for convergence. Thus, some regularity condition to the derivative of the optimal discriminator is introduced. The spectral normalization proposed in [26] controls the Lipschitz constant of the discriminator by normalizing the weight matrices using the technique devised by [60], which aims to stabilize the training of the discriminator and promotes the performance of GANs.

The spectral normalization controls the Lipschitz constant of the discriminator by literally constraining the spectral norm of each layer g, and there is $g(\boldsymbol{h}) = W\boldsymbol{h}$, where $\boldsymbol{h}$ is the input and W is the corresponding weight of this layer. The spectral normalization normalizes the weight matrix W of each layer as follows:

$$W_{\text{SN}} := W/\sigma(W) \tag{4}$$

where W_{SN} is the normalized weight matrix, and $\sigma(W)$ is the spectral norm of the matrix W and is equivalent to the largest singular value of W.

Computing $\sigma(W)$ by singular value decomposition is computationally expensive, and in [26], the power iteration method proposed in [60] and [61] is used for estimating $\sigma(W)$. $\tilde{\boldsymbol{u}}$ is initialized as a random vector, which is sampled from isotropic distribution. The power iteration method for an unnormalized weight W is expressed as follows:

$$\begin{aligned} \tilde{\boldsymbol{v}} &\leftarrow W^{\mathrm{T}}\tilde{\boldsymbol{u}}/\left\|W^{\mathrm{T}}\tilde{\boldsymbol{u}}\right\|_2 \\ \tilde{\boldsymbol{u}} &\leftarrow W\tilde{\boldsymbol{v}}/\left\|W\tilde{\boldsymbol{v}}\right\|_2 \end{aligned} \tag{5}$$

and the spectral norm $\sigma(W)$ is calculated as follows:

$$\sigma(W) = \tilde{\boldsymbol{u}}^{\mathrm{T}} W \tilde{\boldsymbol{v}}. \tag{6}$$

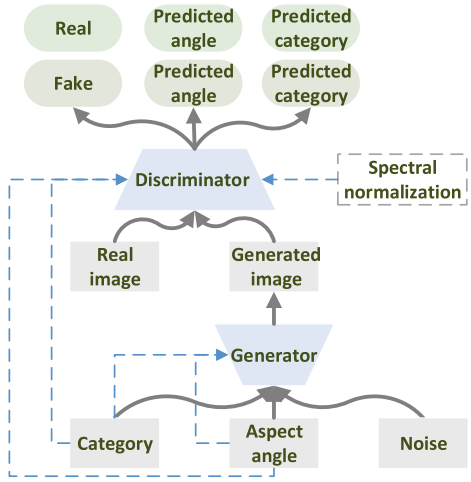


Fig. 5. Illustration of the proposed AGGAN.

IV. Network Structure and Loss Function of AGGAN

A. Network Structure of AGGAN

The illustration of the proposed AGGAN is shown in Fig. 5. First, two kinds of conditional information, which are category label and aspect angle label, are used to control the generated image, which aims to make better use of the potential information contained in the SAR images. Second, the conditional information is added to each layer of both the generator and the discriminator, as shown in Fig. 5. For the generator, it is used to improve the quality of the generated images; for discriminator, it is used to make a better prediction of the conditional information when lacking the training data, and as a result, the trained discriminator cannot be used directly for classification. Third, spectral normalization is used to stabilize the training of the discriminator, which is beneficial for improving the image generation quality. The specific network structure and parameters of the generator and discriminator for AGGAN are shown in Table I.

In Table I, the input dimension of the generator is 106, which is the concatenation of a 100-dimensional random noise vector, a category label vector, and an angle label vector. The angle label vector is composed of the sine and cosine values of the angle label. The output of every layer for both the generator and the discriminator is concatenated with the conditional information. The detailed introduction of the labels and the loss functions is given in the following section.

TABLE I
SPECIFIC NETWORK STRUCTURE OF THE GENERATOR AND DISCRIMINATOR FOR AGGAN

The network structure of the generator	The network structure of the discriminator
Linear:106-1024-ReLU Concat(?,4,4,64)-(?,4,4,70) Deconv5*5-s:2-128-InstanceNorm-ReLU Concat(?,8,8,128)-(?,8,8,134) Deconv5*5-s:2-64-InstanceNorm-ReLU Concat(?,16,16,64)-(?,16,16,70) Deconv5*5-s:2-32-InstanceNorm-ReLU Concat(?,32,32,32)-(?,32,32,38) Deconv5*5-s:2-16-InstanceNorm-ReLU Conv5*5-s:1-1-Tanh	Concat(?,64,64,1)-(?,64,64,7) CvSN5*5-s:2-16-LeakyRelu Concat(?,32,32,16)-(?,32,32,22) CvSN5*5-s:2-32-InstanceNorm-LeakyReLU Concat(?,16,16,32)-(?,16,16,38) CvSN5*5-s:2-64-InstanceNorm-LeakyReLU Concat(?,8,8,64)-(?,8,8,70) CvSN5*5-s:2-128-InstanceNorm-LeakyReLU Logits_adv=Linear-SN:2048-1 Logits_cls=Linear-SN:2048-4 Logits_angle=Linear-SN:2048-1

The deconvolution layer is represented as "Deconv(kernel size)-(s:stride size)-(number of feature maps)". The convolution layer is represented as " Conv(kernel size)-(s:stride size)-(number of feature maps)". The convolution layer with spectral normalization is represented as "CvSN(kernel size) -(s:stride size)-(number of feature maps)". The fully connected layer is represented as "Linear:(number of units for this layer)-(number of units for next layer)", and "Linear-SN" is the fully connected layers with spectral normalization. "Concat" means concatenation of multiple feature maps, a-nd it is represented as "Concat(size of current feature map)-(size of the feature map after concatenation)". "LeakyReLU" [54],"ReLU" [55, 56] and "Tanh" [57] represent the corresponding activation functions respectively. And "InstanceNorm" is the instance normalization method [58]. "?" illu-strates the number of samples used in an episode. "Logits_adv" gives the predicted probability that the input image is real. "Logits_cls" gives the predicted category label probability vector and "Logits_angle" gives the predicted value of the aspect angle.

B. Loss Function of AGGAN

A SAR image x is described by three labels in this article, which are category label c_x, aspect angle label a_x and the label t_x to denote whether it is a generated image. c_x is of the one-hot form, and the dimension is determined by the number of categories. The input aspect angle label a_x is a value in the range of [0°, 360°). The sine and cosine values of a_x are used in the training process. A SAR image is labeled as $l_x = [c_x, a_x, t_x]$ in this article. The real image is denoted as x_{real} while the generated image is denoted as x_{fake}.

The input of the generator G includes a random noise vector z, the category label c_x and the angle label a_x, and the corresponding generated data x_{fake} can be expressed as $x_{\text{fake}} = G(z, c_x, a_x)$. The loss function to be maximized for the generator is expressed as follows:

$$L_G = L_c(x_{\text{fake}}) - L_a(x_{\text{fake}}) + E_{x_{\text{fake}}}[\log(\hat{t}_x)] \tag{7}$$

where L_c is the log-likelihood corresponding to the correct category label c_x, which assures the classification accuracy of each category, and L_a gives the prediction error for the aspect angle. $\hat{t}_x$ is the predicted probability that the sample x comes from the real data distribution rather than the generated data distribution:

$$L_c(x) = E_x\left[c_x \cdot \log(\hat{c}_x))\right] \tag{8}$$

$$L_a(x) = E_x\left[\|\hat{a}_x - a_x\|_2^2\right] \tag{9}$$

Among them, $\hat{c}_x$ is the predicted category label of x, and $\hat{a}_x$ is the predicted value of aspect angle.

The input and the output of discriminator D can be shown as $D(x, a_x, c_x) = [\hat{c}_x, \hat{a}_x, \hat{t}_x]$, in which the prediction of three labels is given. The loss L_D to be maximized for discriminator D is expressed as follows:

$$L_D = L_c(x) - L_a(x) + L_t(x) \tag{10}$$

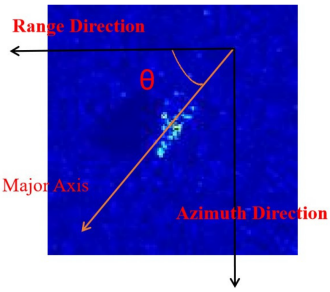


Fig. 6. Illustration of the aspect angle.

where L_t is the cross-entropy loss for determining whether the input sample is generated or real and expressed as follows:

$$L_t(x) = E_{x_{\text{real}}}[\log(\hat{t}_x)] + E_{x_{\text{fake}}}[\log(1 - \hat{t}_x)] \tag{11}$$

All in all, like all GANs models, D and G look like playing the two-player minimax game to make the generated images realistic in terms of categories and aspect angles.

V. TRADITIONAL TRAINING STRATEGIES AND IMPROVED EPISODE TRAINING STRATEGY

In [62], the shared image transformation pattern related to different aspect angles in the source domain helps to improve the recognition performance in the target domain. In this article, to improve the generation performance under the few-shot situation, an auxiliary dataset in the source domain with enough training samples is also introduced into the training process to help improve the few-shot image generation in the target domain. The definition of aspect angle is the angle between the azimuth or range direction and the major axis of the target, as shown in Fig. 6, and for SAR images of different categories, the aspect angle is defined in the same way. Thus, when the training samples in the target domain are inadequate and the aspect-angle-related image information is missing, the source

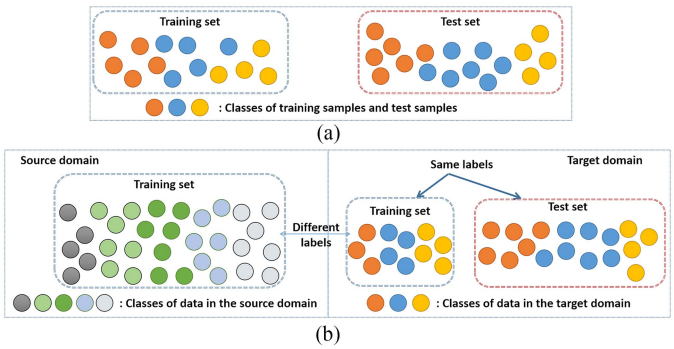


Fig. 7. Differences of datasets in traditional classification and few-shot classification. The samples of the same category are given the same color. (a) Traditional classification. (b) Few-shot classification.

data with complete aspect-angle-related image information may be helpful for image generation in the target domain.

Different from traditional transfer learning, which commonly implements vanilla learning on the source task and then parameter transfer plus finetuning [63], we use the meta-learning method [34] to gain experience from the source data. The episode training strategy utilized in [35] for few-shot classification is a well-known meta-learning method, we modify it and propose an improved episode training strategy for the better quality of the generated images, and the modification is made according to the special characteristics of the few-shot image generation task.

A. Traditional Episode Training Strategy

Since the episode training strategy was originally utilized for the few-shot classification problem in [35], the few-shot classification problem is explained at first. Referring to the work in [51], the differences between the datasets in traditional classification and few-shot classification are shown in Fig. 7.

The traditional classification problem is illustrated in Fig. 7(a). When few labeled samples are available for training the classifier, the trained classifier is easy to overfit. The few-shot classification problem is shown in Fig. 7(b), and an auxiliary dataset of a relevant source domain is introduced to assist the classification in the few-shot situation. However, the class labels of the source domain and the target domain are disjoint, and the classifier trained by minibatch gradient descent (MGD) on the training set in the source domain cannot generalize well to the unseen test samples. The episode training strategy can be utilized in this situation, and it is a meta-learning method.

The details of the episode training strategy are introduced as follows. In each iteration, a task (episode) is constructed from the source data to mimic the final recognition task in the target domain, and the episode is of N-way K-shot form. To form an episode for training, sampling is performed twice for the source data. The first time is to pick N categories from the source data, where N is the number of categories in the target domain, and the second time is to sample inside the selected category. In the selected categories, K samples of each category are randomly selected. K is usually set to be the number of samples in each category of the target data.

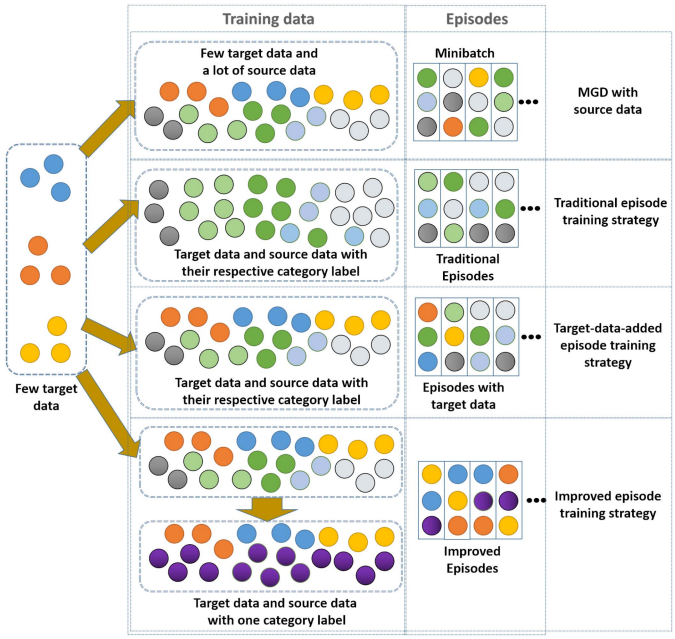


Fig. 8. From top to bottom, MGD with source data, traditional episode training strategy, target-data-added episode training strategy, and improved episode training strategy are illustrated, respectively. Suppose there are three training samples for each category in the target domain, and each color represents a category. The source data are added as auxiliary data. For the MGD, the batchsize is set to 3, which means that 3 samples are selected randomly as a batch for training. For episode-based training strategies, 3-way 1-shot episodes are used for training, which means that 1 sample is selected from each of the 3 randomly selected categories to form the episode for training.

The advantages of the episode training strategy can be summarized as improving the generalization capability of the trained model, and it can be explained as follows: During the training process, the training episodes are designed to mimic the few-shot target tasks, which makes the training scenario more faithful to the test environment, and each task may contain different combinations of categories. By training the model iteratively on each task, the model learns the common knowledge of different tasks; then, in the testing stage, when facing the new category combination in the target domain, the well-trained model can fit tasks that are new and unfamiliar, and the knowledge transfer can improve the classification result. In fact, the application of the episode training strategy in [36] makes it popularized for few-shot learning problems.

B. Improved Episode Training Strategy

In this section, we discuss different training strategies, which are MGD [64], traditional episode training strategy [36], target-data-added episode training strategy [37], and improved episode training strategy, as shown in Fig. 8.

MGD is used for training in most of the existing GAN-based models. However, due to the introduction of the source data, the MGD training method does not work well in this situation, and it will be verified in Section VI-C.

When the traditional episode training strategy is applied for training a generative model, as shown in Fig. 8, only the source

data are used for training. Then, in the testing stage, the learned model will not generate fake images related to the target category in the target domain since the generative model is trained to learn the distribution of the source data but not the target data. The images of the unseen classes (i.e., classes in the target domain) cannot be well generated in this situation.

The target-data-added episode training strategy has been utilized in [37] for the few-shot classification, in which the episodes are constructed using both the source and target data. When applying the target-data-added episode training strategy to the few-shot SAR image generation problem, as shown in Fig. 8, we can see that both the target data and the source data are used to form the episode for training while the source data keep their original category label.

In our article, two different kinds of source data are used, denoted by Case A and Case B, as mentioned in Section I. Directly using the target-data-added episode training strategy for these two cases will have different problems. For case A, there is sufficient source data with different category spaces. Directly sampling categories from both the source and target data will cause a low sampling probability of the target domain. Thus, the target data cannot be learned sufficiently. For case B, the source data are the simulated data with the same categories as the target data. In this case, the data in the source and target domains cannot be distinguished from each other, and the sampling of categories cannot be realized in the episode construction. To deal with these problems, we modify the target-data-added episode training strategy and propose an improved episode training strategy for the task of few-shot image generation.

The improved episode training strategy uses the episodes constructed by the source and target data together with the specified class label definition. For improved episode training strategy, all the images in the source domain are given the same category label, and the images in the target domain retain their respective category labels, which means that all the image data in the source domain are considered as one category distinguished from the target domain. As shown in Fig. 6, the aspect angle label definition of the source and the target data is in the same way. Therefore, the aspect-angle-related image information is shared between the source domain and the target domain, and the complete aspect-angle-related image information contained in the source domain is expected to compensate for that in the target domain. For category labels, the category-related image information contained in the source domain is relatively unimportant, since we want to generate images of the target categories but not the auxiliary source categories. An illustration of different training strategies is shown in Fig. 8.

The advantages of improved episode training strategy can be summarized as follows: First, the redefinition of category labels enables the application of different source data for training, no matter what label space the source data has; second, by defining category labels in this way, the few training data in the target domain are more likely to be selected to form the N-way K-shot training episode. As a result, the target data are used more frequently in the training process, and the unnecessary category-related image information in the source domain is reduced for target tasks. It will be beneficial for the image generation tasks in the target domain, which will be verified by the following experimental results.

TABLE II
TARGET DOMAIN: THREE TYPES OF TARGET IN THE MSTAR DATASET

Target Type	BMP2			BTR70	T72		
	SNC21	SN9563	SN9566	C71	SN132	SN812	SNS7
Training	0	233	0	233	232	0	0
Testing	196	195	196	196	196	195	191

TABLE III
SOURCE DOMAIN A: SEVEN TYPES OF TARGETS IN THE MSTAR DATASET

Target Type	BTR60	2S1	BRDM2	D7	T62	ZIL131	ZSU23/4
Training	256	299	298	299	299	299	299

TABLE IV
SOURCE DOMAIN B: THREE TYPES OF TARGETS IN THE SIMULATED SAR DATASET

Target Type	BMP2	BTR70	T72
Training	360	360	360

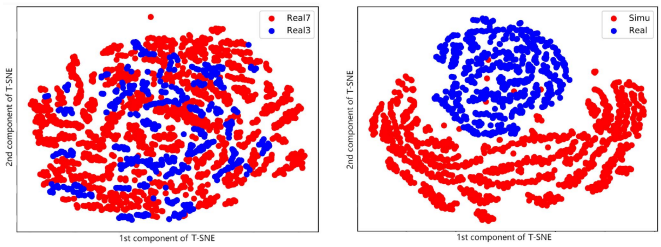
Fig. 9. Optical images of three categories in the target domain. BMP2 means the infantry fighting vehicle; BTR70 means the armored transport vehicle; T72 means the tank.

VI. EXPERIMENTAL RESULTS

A. Experimental Datasets

The MSTAR public dataset and a simulated SAR dataset are used to evaluate the performance of the proposed method. The MSTAR dataset consists of X-band SAR images with 0.3 m×0.3 m resolution of multiple targets. Tables II and III show the three and seven types of targets. The aspect angles of each category range from 0° to 360°. The depression angles of the training data in Tables II and III are 17°, and the depression angle of the testing data in Table II is 15°. The simulated data [65] are provided by the Science and Technology on Electromagnetic Scattering Laboratory, Beijing Institute of Environmental Features, China. The simulated data is obtained by 3-D modeling and electromagnetic simulation, and the simulation details can be found in [65]. There are three categories of targets in the simulated dataset, as shown in Table IV. The resolution is 0.2 m×0.2 m, and there are 360 images in each category while the aspect angles range from 0° to 360° while the depression angle is 17°. The optical images of three categories in the target domain are shown in Fig. 9.

There are three types of data in the target domain as shown in Table II, and two experimental settings are considered.



(a) (b)

Fig. 10. T-SNE visualization results of the data distribution for case A and case B. (a) Case A: "Real7" indicates the seven types of SAR data in Table III, and "Real3" indicates the three types of training data in Table II. (b) Case B: "Simu" indicates simulated SAR data in Table IV, and "Real" indicates three types of training data in Table II.

1) *Case A:* Seven types of MSTAR data in Table III are used as the source data, and three types of MSTAR data in Table II are used as the target data. It can be seen that the source and target data are in different category spaces.

2) *Case B:* Three types of simulated SAR data in Table IV are used as source data, and three types of MSTAR data in Table II are used as target data. In this situation, the source and target data own the same category space but the data distributions are different.

The t-distributed Stochastic Neighbor Embedding (t-SNE) [66] has been widely used to visualize the high-dimensional data by mapping them into the low-dimensional space and is used to compare the data distributions of case A and case B, as shown in Fig. 10. It is noticed from Fig. 10(a) that the data distributions of the real and simulated data are superimposed, which means that the data distributions of source and target data in case A are similar to each other. However, we can see that the real and simulated data distributions do not completely overlap in Fig. 10(b), which means that there is a domain shift between the real and simulated data, and it is not suitable to use the simulated data directly for image generation. Therefore, case A and case B are two different cases of transfer learning.

B. Experimental Settings

Data preprocessing: In the following experiments, all the images in the MSTAR dataset and simulated dataset are cropped to the size of 64×64 to exclude the influence of background and normalized by min-max scaling to the range [0, 255]. Referring to the experimental setting in [62], by reducing the number of training data in the target domain, the scenario of few-shot SAR image generation is realized.

Model Training: The source data and the selected target data are used to train generative models with different training strategies, the episodes utilized in the following experiments are 3-way 5-shot tasks, and the batchsize of the MGD method is 15 accordingly. Unless otherwise stated, the generative models are trained by 20 000 epochs or 20 000 episodes. When dealing with the 3-way k-shot recognition tasks, $k = 10, 15, 20$, we explore the shot setting in training by experiments and find that the 5-shot setting and k-shot setting, $k = 10, 15, 20$, have little influence on the recognition results. Thus, 3-way 5-shot setting is also used in the training of 3-way k-shot, $k = 10, 15, 20$, recognition tasks.

Evaluation of the generation results: In addition to visual observation and image similarity measures between the generated images and the corresponding real images, the category-related information contained in the generated images should also be verified. Since the category-related information is difficult to be observed by our naked eyes, we refer to the quantitative comparison method used in [19] and design two kinds of recognition experiments to analyze the generated images, which are authenticity verification and few-shot recognition, as shown in Fig. 11. If the generated images are of better quality, the recognition rates will be higher. Referring to the work in [46], we design our recognition model, which is a 5-layer fully convolutional structure, and the network parameters are designed to match the input image size.

1) *Authenticity verification:* Only the generated data are used to train the recognition model, which follows two steps: The first step is generative model training and image generation. Five samples of each category in the target domain are randomly selected and together with the source data are used to train the generative models. In total, 1000 image samples for each category in the target domain are randomly generated with the trained generative models. Second, as shown in Table II, a recognition model is trained by the 3000 generated images and tested on the corresponding test data in the target domain. When the case A experimental setting is used, the model name is appended with A, and the same for B.

2) *Few-shot recognition:* The selected target data and the generated data are used for training together. The first step is also the AGGAN model training, and only 360 images for each category are generated for recognition, which aims to match the number of the simulated data shown in Table IV; then, the generated images and the few real training data (15 samples here) in the target domain are used together to train the recognition model, and the trained model is used to test the testing data in the target domain as shown in Table II. The generated images act as augmented data in this situation.

Illustration of recognition results: To avoid the randomness of the recognition experiment, each group of experiments is carried out 10 times to calculate the averages and the standard deviations as the final results, and the target data of 10 experiments for different generative models are the same. However, since there are only a few samples for training, the target samples selected greatly impact the recognition rate, leading to a large standard deviation in the multiple recognition experiments shown in the following sections.

Arrangements of the following sections: In Section VI-C, 5-shot SAR image generation results using different training strategies are performed to verify the superiority of the improved episode training strategy. In Section VI-D, with AGGAN and the improved episode training strategy, the generation results of training cases A and B are obtained and compared. Then, in Section VI-E, the generation results of different generative models are compared to prove the superiority of our AGGAN model, and the recognition results under different target sample sizes are also analyzed. Finally, Section VI-F further verifies

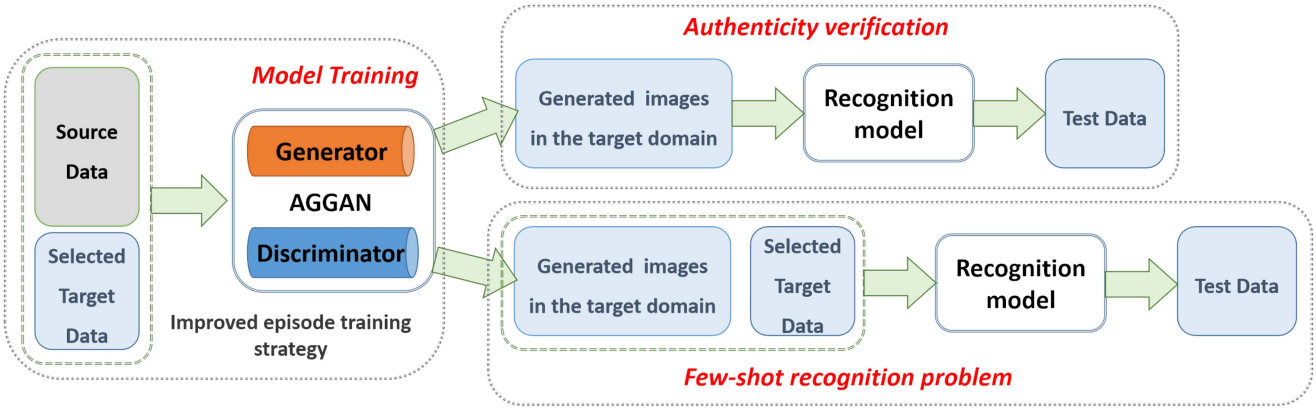


Fig. 11. Illustration of the two kinds of recognition experiments, which are authenticity verification and few-shot recognition.

the benefit of applying generated and augmented data to the few-shot SAR target recognition problem.

C. Verification of the Improved Episode Training Strategy

Different training strategies are compared in this section to verify the effectiveness of the improved episode training strategy, and the experiments are carried out under case A in this section. The generation results of different training strategies are shown in Fig. 12.

The generation results using MGD with source data are shown in Fig. 12(b) and (c). The generation results shown in Fig. 12(b) illustrate that the network has not been well-trained. It can be seen that the quality of the generated images is still inferior by increasing the training epochs, as shown in Fig. 12(c), as the generated images look unclear and the target outlines are not obvious. Therefore, when introducing source data and applying MGD for training, the difficulty of target task learning increases, and the image quality of generated images in the target domain is not improved.

As for the traditional episode training strategy [see Fig. 12(d)], the generated images are also unsatisfactory since the target data are not used in the training process. Consequently, the generated images of the target categories are not so realistic. For the target-data-added episode training strategy, the generation results are generally acceptable. Comparing the 4th, 5th, and 12th images of Fig. 12(e) and (f) marked by the red boxes, we can see that the target areas of Fig. 12(f) look clearer. This implies that the improved episode training strategy is beneficial for image generation. Most importantly, the improved episode training strategy makes the application of different source data possible.

Comparing all the generation results in Fig. 12, we can see that the generation results in Fig. 12(e) and (f) are similar and relatively better than other generation results, and thus, the AGGAN generation results of target-data-added episode training strategy and improve episode training strategy are further compared in Fig. 13.

Fig. 13 shows the BTR70 and T72 generation results of AGGAN using different training strategies. Comparing the generation results in Fig. 13(a) with (b) and Fig. 13(c) with (d), we can see that the target areas of images in Fig. 13(b) and (d) are more apparent. For the results of target-data-added episode training strategy shown in Fig. 13(a) and (c), there are more noises in the background areas of the generated images. Overall, the proposed improved episode training strategy performs better for few-shot SAR image generation in the target domain.

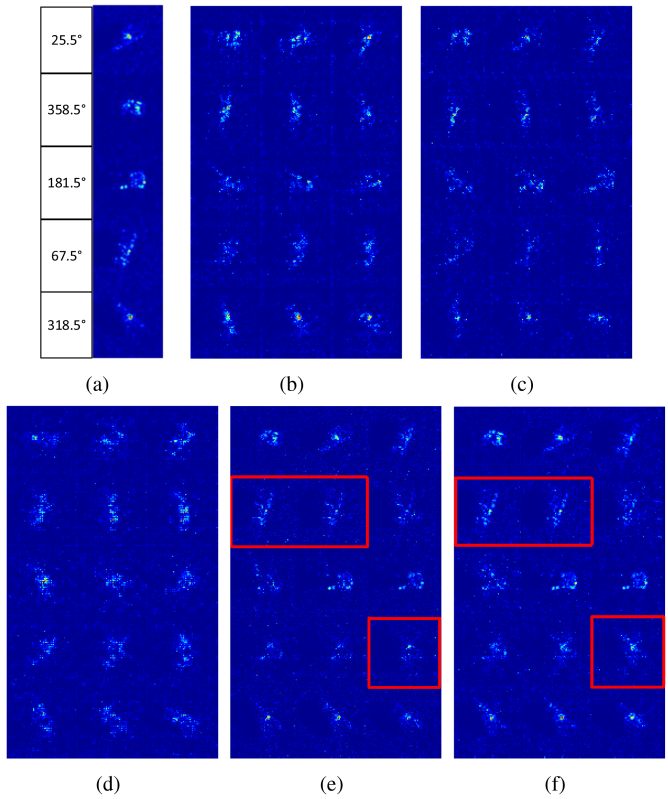


Fig. 12. BMP2 generation results of AGGAN using different training strategies. For the ith image counted from left to right and top to bottom, the aspect angle is $24 \times (i-1)$. The aspect angles of generated images shown below are all counted in this way. (a) Five training images of the BMP2 category in the target domain. (b) MGD with source data. (c) MGD with source data: 50 000 epochs. (d) traditional episode training strategy. (e) target-data-added episode training strategy. (f) improved episode training strategy.

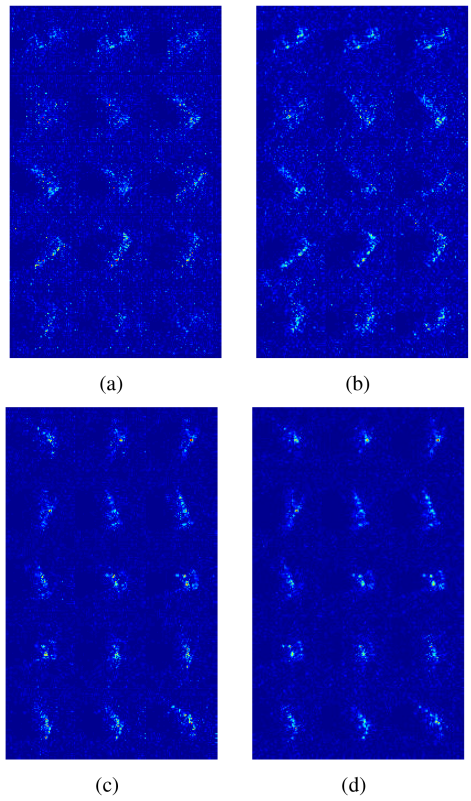


Fig. 13. BTR70 and T72 generation results of AGGAN using different training strategies. (a) BTR70: Target-data-added episode training strategy. (b) BTR70: Improved episode training strategy. (c) T72: Target-data-added episode training strategy. (d) T72: Improved episode training strategy.

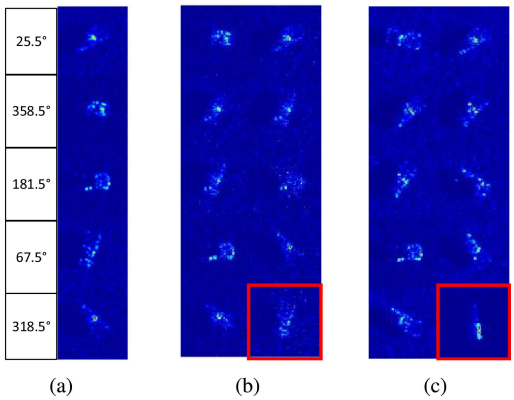


Fig. 14. Comparison between the generated images and the corresponding real images for the BMP2 category. (a) Five given training images of the BMP2 category in the target domain. (b) generated images. (c) corresponding real images.

We have also compared the generated images with the corresponding real images, shown in Fig. 14. We can see that the trained model can generate the images at the aspect angles that do not exist in the training data of the target domain. In Fig. 14(b), there are 9 valid ("valid" means the aspect angle of the generated image looks correct and the generated image is relatively clear.) generated images. Compared with the corresponding real images shown in Fig. 14(c), there is a relatively obvious error in the aspect angle for the 10th generated image marked by the red box. Furthermore, the strong scattering centers of the generated images do not look consistent with the corresponding real images. Admittedly, differences exist for the strong scattering points between some generated images and the corresponding real images. More discussions about the effectiveness of the improved episode training strategy will be given as follows.

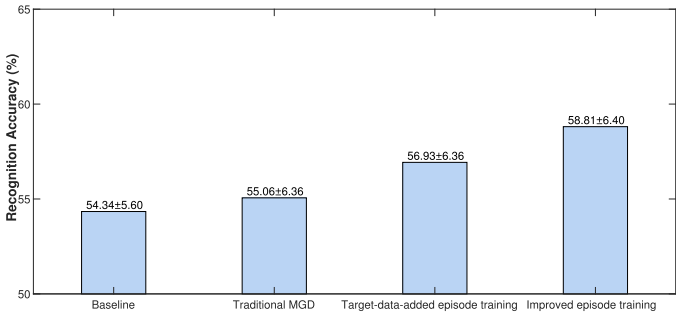


Fig. 15. Case A: 5-shot recognition results of AGGAN trained by different training strategies.

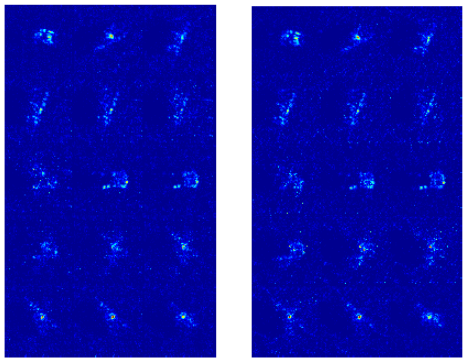
We carry out 5-shot recognition experiments, as shown in the second row of Fig. 11, to verify the effectiveness of the proposed improved episode training strategy. The experiments are carried out in the case A setting. The recognition results are shown in Fig. 15. The baseline is the recognition result when only a few real target data are used to train the recognition model, represented as "Baseline" in Fig. 15.

We can see that the recognition result of AGGAN trained by the improved episode training strategy gets the best performance, and it confirms that the authenticity of the generated images of AGGAN trained by the improved episode training strategy outperforms others. The generated images of the model trained by traditional MGD training strategy are not so helpful for the recognition performance improvement when compared with "Baseline."

D. 5-Shot Generation Results Using Different Source Data

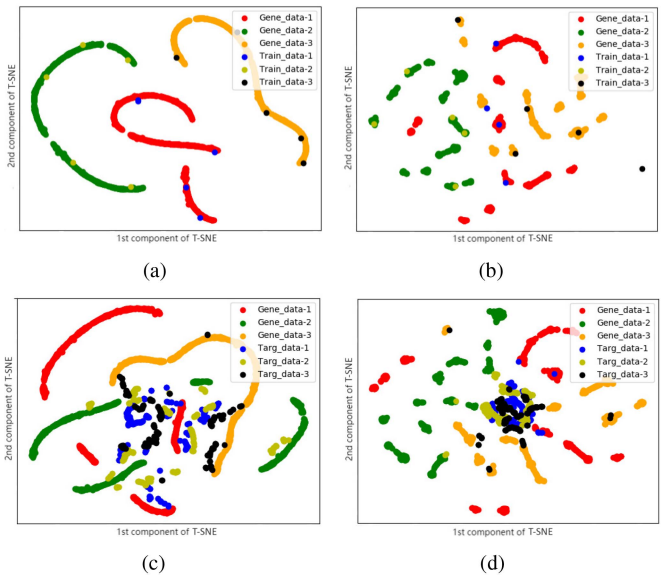
This section gives the generation results of AGGAN with improved episode training strategy under case A and case B settings. The generation results are shown in Fig. 16. To further analyze the generated images, the t-SNE visualization results of the target data and the generated data are also shown in Fig. 17. In Fig. 17, the generated data are represented as "Gene_data," and it is appended with the category number; the target data means all the training data shown in Table II, and is represented as "Targ_data" appended with the category number; the training data means the selected 15 training samples (5 samples per category) in the target domain used for training the AGGAN and is represented as "Train_data" appended with the category number.

In Fig. 16, we can see that the generation results of both case A and case B settings are satisfactory. Thus, the simulated data can be used as the source data to increase the diversity of the generated images. Fig. 17(a) and (b) shows the distribution of the generated images and the corresponding 15 training samples



(a) (b)

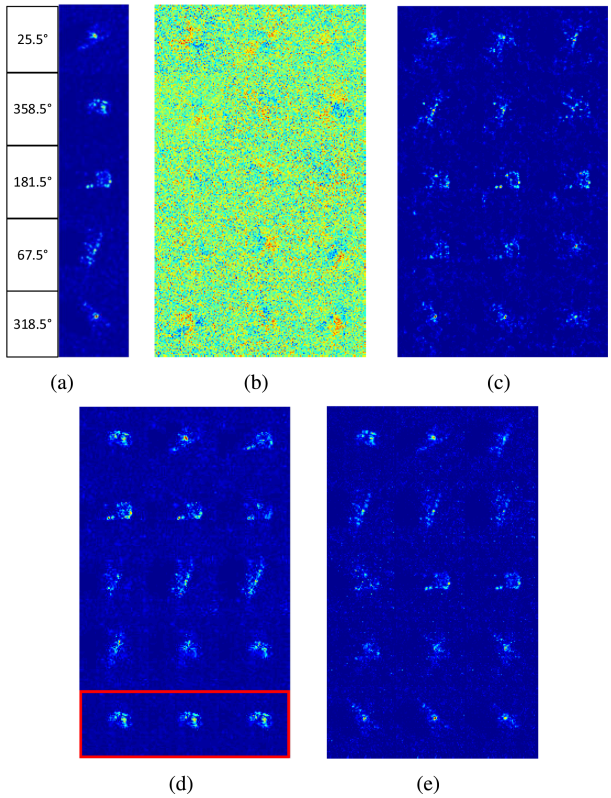
Fig. 16. BMP2 generation results of AGGAN with an improved episode training strategy using different source data. (a) Case A. (b) Case B.



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

Fig. 17. T-SNE visualization result. (a) Case A: 15 training samples and the corresponding generated data. (b) Case B: 15 training samples and the corresponding generated data. (c) Case A: The generated data and the target data. (d) Case B: The generated data and the target data. Here, the target data means the three types of training data in Table II.

in the target domain, under case A and B settings, respectively. From Fig. 17(a) and (b), we can see that the training data of a certain category basically distribute inside the area of the generated data of the same category. This implies that the generated data catches the distribution of the few real training samples (15 samples of the target data). Since only 15 samples of real data are used for training, the generated data cannot well catch the distribution of the other real data (i.e., the other training data in the target domain). Therefore, the distributions of the generated data do not entirely overlap with all the target data, as shown in Fig. 17(c) and (d). In addition, by comparing the distributions of generated data shown in Fig. 17(a) and (b), we can see that the different source data affect the distribution of generated data.



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

Fig. 18. BMP2 generation results of different models using the improved episode training strategy. (b)–(e) Generation results. (a) 5 given training images of the BMP2 category in the target domain. (b) CGAN-MLP. (c) CGAN-DN. (d) AC-GAN. (e) AGGAN.

E. 5-Shot Generation Results of Different Models

In this section, we compare the generation results of AGGAN, CGAN-MLP, CGAN-DN, and AC-GAN under the case A experimental setting using the improved episode training strategy. CGAN-MLP represents CGAN with MLP network structures. The discriminator and the generator of CGAN-MLP are implemented by the two-layer MLP networks, as mentioned in [2]. CGAN-DN represents the CGAN with deep convolutional network structures and is implemented by almost the same structures, as shown in Table I, without adding conditional information per layer and spectral normalization. For CGAN-MLP and CGAN-DN, their loss functions follow (1). For AC-GAN, the overall network is similar to that in Table I, and spectral normalization is not used. The loss function of AC-GAN follows (2) and (3). The two conditional labels are both used in the CGAN-MLP, CGAN-DN, and AC-GAN structures. The comparison of generation results is shown in Fig. 18.

Note that the results of CGAN-MLP shown in Fig. 18(b) are similar to those in [21], and it can be concluded that compared with the MLP network, the convolutional network is beneficial for learning the detailed images. The network structures of CGAN-DN, ACGAN, and AGGAN models are similar. When they are well-trained, there is not much difference of the test speeds for CGAN-DN, ACGAN, and AGGAN. While the overall quality of images generated by CGAN-DN [see Fig. 18(c)]

TABLE V
SIMILARITY MEASURES BETWEEN GENERATED IMAGES AND REAL IMAGES

Similarity \ Model&Category	CGAN-DN			AC-GAN			AGGAN		
	BMP2	BTR70	T72	BMP2	BTR70	T72	BMP2	BTR70	T72
COSS	0.474	0.430	0.488	0.520	0.522	0.583	**0.575**	**0.525**	**0.606**
SSIM	0.229	0.162	0.292	0.234	0.223	0.358	**0.343**	**0.255**	**0.391**
NCC	0.261	0.191	0.332	0.272	0.248	0.408	**0.384**	**0.283**	**0.453**

is also poor, and the targets are not clear enough. Comparing the generation results of AC-GAN in Fig. 18(d) with other generative models, the aspect angles of the images generated by AC-GAN are more difficult to control. Specifically, the aspect angles from the 13th to the 15th images are 288°, 312°, and 336°, respectively. We can see that the aspect angles of the 13th to the 15th images in Fig. 18(d) are obviously wrong, and the aspect angles of generated images in Fig. 18(c) and Fig. 18(e) look more reasonable. To sum up, the visual quality of images generated by AGGAN is better than that of CGAN-DN, and the aspect angles of images generated by AGGAN are more reasonable than that of AC-GAN. Therefore, the generation results of AGGAN are relatively better.

We analyze the generation results of AGGAN further. It is noticed from Fig. 18(e) that there are still some unclear generated images, and only the outlines of the targets can be seen, and some generated images are even with wrong aspect angles. These results imply that image generation with extremely few training samples is, indeed, a very difficult task. In [24], an additional method called rotated cropping is introduced as a mechanism to address the challenge of representing the target orientation, which also illustrates that it is difficult to learn an effective angle representation when using only a few data for training. With only 5 samples per category for training, even assisted by the auxiliary source data, the proposed model still cannot learn the aspect angles and strong scattering centers well enough. Although the generation results of the proposed AGGAN still have some defects, the quality of the generated images is already better than the compared existing methods.

To illustrate the superiority of the images generated by the AGGAN, we make an additional quantitative comparison between the generated images and the real images for different methods. To measure the similarity between the generated data and the corresponding real data, we use Cosine Similarity (COSS) [45], Structural Similarity (SSIM) [45], and Normalized Cross Correlation (NCC) [24] to measure the similarity between the generated images and the real images. The detailed computation of the similarity measures can be found in [24] and [45]. The similarity experiments are carried out as follows: 100 generated images of each category are compared with the corresponding real images with the same class and the same aspect angle to calculate the similarity value, and the mean similarity values of different generative models for different categories are given in Table V.

By comparing various similarity measures, we can see that the similarity values of our proposed model are higher than those of other generative models, which verifies that the images generated by AGGAN are of higher quality.

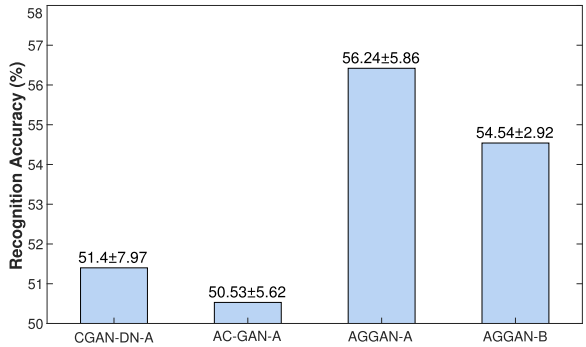


Fig. 19. Recognition results of different models with improved episode training strategy using different source data.

TABLE VI
CLASSIFICATION ACCURACY AND STANDARD DEVIATION UNDER DIFFERENT NUMBER OF TARGET DATA (%)

Model \ Num	5	10	15	20
CGAN-DN-A	51.40±7.97	53.44±**3.97**	50.54±6.28	46.04±7.56
AC-GAN-A	50.53±5.62	49.92±6.65	54.70±5.00	51.51±6.41
AGGAN-A	**56.24**±5.86	64.24±4.57	**70.74**±4.21	**73.20**±**3.02**
AGGAN-B	54.54±**2.92**	**65.77**±3.98	70.18±**2.54**	72.46±5.05

The authenticity verification of 5-shot generation results, which are shown in the first row of Fig. 11, are also conducted as quantitative measures to further compare the quality of the generated data. Fig. 19 gives the results.

We can see that the recognition results of AGGAN outperform others, and the experimental result of AGGAN for case A is slightly better than that for case B. Overall, in both case A and case B experimental settings, the recognition results of AGGAN outperform others.

According to the above analysis, the results verify that the generated images of AGGAN outperform others by visual observation and quantitative comparisons. We consider the authenticity verification under different sample sizes for further evaluation. The image quality of generated images under different "Num" settings is further evaluated, and "Num" means the number of labeled training samples of each category in the target domain. The generated images of CGAN-DN, AC-GAN, and AGGAN are compared. The results are given in Table VI.

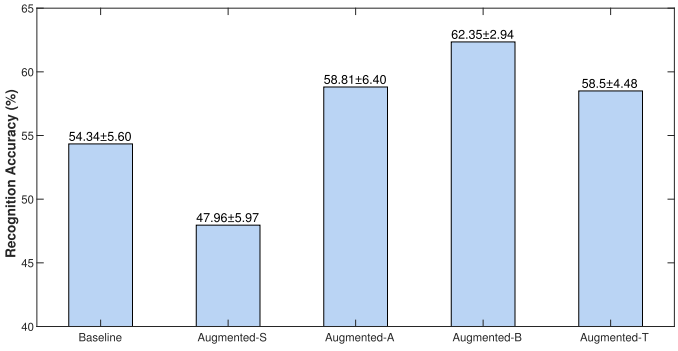


Fig. 20. 5-shot recognition results with different augmentation methods.

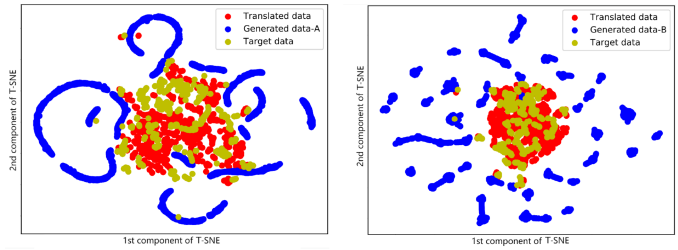


Fig. 21. T-SNE visualization results of translational augmented data, generated data, and target data.

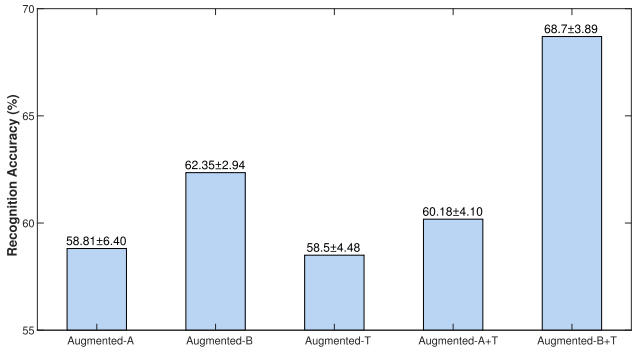


Fig. 22. Recognition results when the generated data and the translational augmented data are both used as the augmented data.

From Table VI, we can see that when the training samples in the target domain increase, for AGGAN, the recognition rates increase obviously; for CGAN-DN, on the contrary, the recognition rates decrease. Thus, the data generated by CGAN-DN may not be beneficial for the recognition task. And for AC-GAN, the recognition rates are fluctuant. It illustrates that the additional classification task for the category may be beneficial for conditional image generation, since the AC-GAN and AGGAN both complete an additional classification task for the category label in the training process. Observing the recognition results, we can infer that the category-related information contained in the images generated by AGGAN is more accurate and beneficial for the recognition tasks. Overall, in both case A and case B settings, the recognition results of AGGAN outperform other generative models.

F. Effectiveness of Different Augmentation Methods

In this section, the effectiveness of different augmentation methods is evaluated. "Baseline" is the recognition result when only a few real target data are used to train the recognition model. The generated images under different source data are used as the augmented data. And for case A and case B, the recognition results are represented as "Augmented-A" and "Augmented-B," respectively. "Augmented-S" represents the recognition result when the few target and simulated data are used to train the recognition model together. Since SAR data augmentation is commonly achieved by image translation in recent works [46], [67], we also consider the translation augmentation method. The implementation of translation augmentation refers to the work in [67]. At the same time, 360 translational augmented images (360 matches the number of samples in the simulated data set) for each category are used for recognition, and the recognition results are represented as "Augmented-T." The final results are the average results shown in Fig. 20.

For "Baseline," the recognition model fails to converge occasionally, and the recognition result of "Baseline" shown in Fig. 20 excludes the cases of nonconvergence. It is worth mentioning that the recognition model converges whenever adding the generated images as the augmented data, and the average recognition result is improved. Directly adding the simulated data declines the recognition rates, illustrating the domain shift of using simulated data directly for recognition. It is also beneficial to augment the training data by translation. Nevertheless, the diversity of the translational augmented data may not be improved effectively. The t-SNE visualization is utilized to explore the diversity of the generated data and the translational augmented data as shown in Fig. 21, in which the target data means all the training data in the target domain shown in Table II.

We can see that the distribution of translational augmented data is similar to that of the target data while the generated data show greater diversity. Thus, we consider using both the translational augmented and generated data for further improvements. In total, 180 generated images and 180 translational augmented images for each category are used to augment the target data together, matching the number of augmented data in Fig. 20, "Augmented-A+T" represents augmentation by the images generated in case A scenario and the translational augmented images, and "Augmented-B+T" represents augmentation by the images generated in case B scenario and the translational augmented images. Fig. 22 shows the experimental results.

The recognition rates increase effectively when the generated and the translational augmented data are both used as the augmented data for training a recognition model. As shown in Fig. 21, the data distribution of the translational augmented data is similar to that of the target data but the generated data show greater diversity. The recognition results in Fig. 22 verify that the translational augmented data and the generated data increase the diversity of the images from different aspects. They have complementary effects in boosting the few-shot recognition performance. When only the generated and target data are used

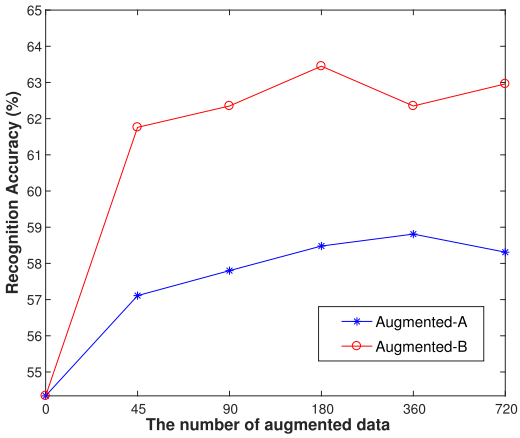


Fig. 23. Recognition rates with different number of augmented data.

for recognition, the learned classifier may have classification bias since there are only a few target samples, which makes the classifier prone to classify the generated data. When the generated and the translational augmented data are both used as the augmented data, the classification bias may be offset to some extent, resulting in higher classification accuracy.

Furthermore, an exploration of the amount of augmented data is also presented, and the results are shown in Fig. 23. It can be concluded that blindly increasing the number of generated data for recognition is not so helpful, which is also mentioned in [21]. Since there are only a few target samples, the learned classifier may have a classification bias. When there are increased generated data for recognition, it may lead to a more severe classification bias and deteriorate the recognition performance. Comparing cases A and B in Figs. 22 and 23, the consistent category information contained in the simulated data may have a good effect on the recognition in the target domain. Overall, the recognition results obtained in the case B scenario are better. This section mainly verifies that the generated data helps improve the recognition rate and image diversity. The results also imply that it may be a feasible way to make use of the simulated data.

VII. Conclusion

In this article, for the few-shot SAR image generation problem, an attribute-guided GAN is designed, and an improved episode training strategy is proposed. We test on both the real data and the simulated to verify the effectiveness of our proposed method. The experimental results demonstrate the effectiveness of our method, and the improved episode training strategy enables the application of different source data for training. When there are only five samples per category, the better quality of the generated data is verified through visual observation, image similarity measures, and two kinds of recognition experiments. For the 5-shot SAR target recognition problem, the recognition rate can be boosted by at least 4% when using the generated images of the proposed AGGAN model. The experimental results also show that, for the few-shot SAR image generation problem, there is still room to improve the image quality of the generated images. For example, more attribute information can be used as conditional labels to generate corresponding images. At present, we mainly focus on the ground vehicle target in our article, and the application of our method in satellite SAR data for ship target generation can be explored in our future work. The application of the generated data and simulated data also deserves more exploration.

References

[1] I. Goodfellow et al., "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[2] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.
[3] T. Karras et al., "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. Learn. Representations*, 2018.
[4] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2642–2651.
[5] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.
[6] The Air Force Moving and Stationary Target Recognition Database. Accessed: Mar. 8, 2011. [Online]. Available: https://www.sdms.afrl.af.mil/datasets/mstar/
[7] Y. Wang et al., "Generalizing from a few examples: A survey on few-shot learning," *ACM Comput. Surv.*, vol. 53, no. 3, Jun. 2021, Art. no. 63.
[8] "Awesome few-shot image generation," [Online]. Available: https://github.com/bcmi/Awesome-Few-Shot-Image-Generation
[9] L. Cloutre and M. Demers, "FIGR: Few-shot image generation with reptile," 2019, *arXiv:1901.02199*.
[10] W. Liang, Z. Liu, and C. Liu, "DAWSON: A domain adaptive few shot generation framework," 2020, *arXiv:2001.00576*.
[11] D. Wertheimer, O. Poursaeed, and B. Hariharan, "Augmentation-interpolative autoencoders for unsupervised few-shot image generation," 2020, *arXiv:2011.13026*.
[12] Y. Hong, L. Niu, and J. Zhang, "F2GAN: Fusing-and-filling GAN for few-shot image generation," in *Proc. ACM Int. Conf. Multimedia*, 2020, pp. 2535–2543.
[13] Y. Hong, L. Niu, and J. Zhang, "MatchingGAN: Matching-based few-shot image generation," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2020, pp. 1–6.
[14] S. Bartunov and D. Vetrov, "Few-shot generative modelling with generative matching networks," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2018, pp. 670–678.
[15] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," 2017, *arXiv:1711.04340*.
[16] Y. Hong, L. Niu, J. Zhang, and L. Zhang, "DeltaGAN: Towards diverse few-shot image generation with sample-specific delta," in *Proc. Euro. Conf. Comp. Vis.*, Tel Aviv, Israel, Oct. 23–27, 2022, pp. 259–276.
[17] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," 2018, *arXiv:1803.02999*.
[18] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1126–1135.
[19] J. Guo et al., "Synthetic aperture radar image synthesis by using generative adversarial nets," *IEEE Trans. Geosci. Remote. Sens. Lett.*, vol. 14, no. 7, pp. 1111–1115, Jul. 2017.
[20] Q. Song, F. Xu, and Y. Jin, "SAR image representation learning with adversarial autoencoder networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 9498–9501.
[21] C. Cao, Z. Cao, and Z. Cui, "LDGAN: A synthetic aperture radar image generation method for automatic target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3495–3508, May 2020.
[22] Z. Cui et al., "Image data augmentation for SAR sensor via generative adversarial nets," *IEEE Access*, vol. 7, pp. 42255–42268, 2019.
[23] C. Cao et al., "An integrated counterfactual sample generation and filtering approach for SAR automatic target recognition with a small sample set," *Remote Sens.*, vol. 13, no. 19, pp. 38–64, 2021.
[24] Q. Song et al., "Learning to generate SAR images with adversarial autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, Jun. 2021.
[25] J. Oh and M. Kim, "PeaceGAN: A GAN-based multi-task learning method for SAR target image generation with a pose estimator and an auxiliary classifier," *Remote Sens.*, vol. 13, 2021, Art. no. 3939.

[26] T. Miyato et al., "Spectral normalization for generative adversarial networks," in *Proc. Int. Conf. Learn. Representations*, 2018.
[27] Y. Huang, "Research on SAR image generation models," M.S. thesis, Sch. Electron. Eng., Xidian Univ., Xi'an, China, 2021.
[28] S. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[29] C. Tan et al., "A survey on deep transfer learning," in *Proc. Int. Conf. Artif. Neural Netw.*, 2018, pp. 270–279.
[30] A. Noguchi and T. Harada, "Image generation from small datasets via batch statistics adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 2750–2758.
[31] Y. Wang, A. Gonzalez-Garcia, and D. Berga, "MineGAN: Effective knowledge transfer from GANs to target domains with few images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9329–9338.
[32] Y. Li, R. Zhang, and J. Lu, "Few-shot image generation with elastic weight consolidation," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 15885–15896.
[33] U. Ojha et al., "Few-shot image generation via cross-domain correspondence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 10738–10747.
[34] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-learning in neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5149–5169, Sep. 2022.
[35] O. Vinyals et al., "Matching networks for one shot learning," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 3637–3645.
[36] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2017.
[37] A. Li et al., "Few-shot learning with global class representations," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 9714–9723.
[38] L. Liu et al., "SAR target classification with CycleGAN transferred simulated samples," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 4411–4414.
[39] D. Malmgren-Hansen et al., "Improving SAR automatic target recognition models with transfer learning from simulated data," *IEEE Trans. Geosci. Remote. Sens. Lett.*, vol. 14, no. 9, pp. 1484–1488, Sep. 2017.
[40] M. Cha et al., "Improving SAR automatic target recognition using simulated images under deep residual refinements," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2018, pp. 2606–2610.
[41] B. Camus, E. Monteux, and M. Vermet, "Refining simulated SAR images with conditional GAN to train ATR algorithms," in *Proc. Actes de la Conf. CAID*, 2020, pp. 160–167.
[42] B. Lewis et al., "A SAR dataset for ATR development: The synthetic and measured paired labeled experiment (SAMPLE)," *Proc. SPIE*, vol. 10987, 2019, Art. no. 109870.
[43] N. Inkawhich et al., "Bridging a gap in SAR-ATR: Training on fully synthetic and testing on measured data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2942–2955, Feb. 2021.
[44] Q. Song et al., "EM simulation-aided zero-shot learning for SAR automatic target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1092–1096, Jun. 2020.
[45] C. Zhang et al., "SAR target recognition using only simulated data for training by hierarchically combining CNN and image similarity," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Jan. 2022.
[46] S. Chen et al., "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806–4817, Aug. 2016.
[47] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM. Trans. Intel. Syst. Technol.*, vol. 2, pp. 27:1–27:27, 2011. [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm
[48] X. Sun et al., "SCAN: Scattering characteristics analysis network for few-shot aircraft classification in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, Apr. 2022.
[49] Q. Song, F. Xu, and H. Wang, "Target representation and classification with limited data in synthetic aperture radar images," in *Proc. Gen. Assem. Sci. Symp. Int. Union Radio Sci.*, 2020, pp. 1–4.
[50] Q. Guo and F. Xu, "A deep feature transformation method based on differential vector for few-shot learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 407–410.
[51] S. Wang et al., "Attribute-guided multi-scale prototypical network for few-shot SAR target classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 12224–12245, Nov. 2021.
[52] K. Fu et al., "Few-shot SAR target classification via metalearning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Feb. 2021.
[53] "Tensorflow-generative-model-collections," [Online]. Available: https://github.com/hwalsuklee/tensorflow-generative-model-collections
[54] A. Maas, A. Hannun, and A. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn.*, 2013.
[55] V. Nair and G. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
[56] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
[57] T. Epelbaum, "Deep learning: Technical introduction," 2018, *arXiv:1709.01412v2*.
[58] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," 2016, *arXiv:1607.08022*.
[59] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," in *Proc. Int. Conf. Learn. Representations*, 2017.
[60] Y. Yoshida and T. Miyato, "Spectral norm regularization for improving the generalizability of deep learning," 2017, *arXiv:1705.10941*.
[61] G. Golub and H. Van der Vorst, "Eigenvalue computation in the 20th century," *J. Comput. Appl. Math.*, vol. 123, pp. 35–65, 2000.
[62] Y. Sun et al., "SAR target recognition with limited training data based on angular rotation generative network," *IEEE Trans. Geosci. Remote. Sens. Lett.*, vol. 17, no. 11, pp. 1928–1932, Nov. 2020.
[63] J. Yosinski et al., "How transferable are features in deep neural networks?," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.
[64] F. Fitzek, F. Granelli, and P. Seeling, *Computing in Communication Networks*, 1st ed. Cambridge, MA, USA: Academic, 2020.
[65] C. Dong et al., "Efficient simulation method for high quality SAR images of complex ground vehicles," *J. Radars*, vol. 4, no. 3, pp. 351–360, 2015.
[66] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.
[67] J. Ding et al., "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 364–368, Mar. 2016.

Yuanshuang Sun received the B.S. degree in communication engineering in 2015 from Xidian University, Xi'an, China, where she is currently working toward the Ph.D. degree in information and communication engineering.

Her research interests include SAR target recognition with small samples, heterogeneous SAR data recognition, deep generative models, etc.

Yinghua Wang (Senior Member, IEEE) received the B.S. degree in information engineering and the Ph.D. degree in control science and engineering from Xi'an Jiaotong University, Xi'an, China, in 2004 and 2010, respectively.

In 2007, she joined the Image and Signal Processing Department, Telecom Paris, Paris, France, as a Visiting Student. She is currently a Professor with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an. Her research interests include synthetic aperture radar (SAR) automatic target recognition, polarimetric SAR data analysis and interpretation, and SAR image processing.

Liping Hu received the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2009.

She is currently with Science and Technology on Electromagnetic Scattering Laboratory, Beijing Institute of Environmental Features, Beijing, China. Her research interests include synthetic aperture radar target simulation and recognition.

Yuanyuan Huang received the M.S. degree in information and communication engineering from Xidian University, Xi'an, China, in 2021.

She is currently with Qingdao Haier Technology Co., Ltd., Qingdao, China. Her research interests include few-shot learning and synthetic aperture radar image generation.

Hongwei Liu (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electronic engineering from Xidian University, Xi'an, China, in 1995 and 1999, respectively.

From 2001 to 2002, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. He is currently a Professor with the National Laboratory of Radar Signal Processing, Xidian University. His research interests include radar automatic target recognition, radar signal processing, and adaptive signal processing.

Siyuan Wang received the B.S. degree in electronic information engineering from Zhengzhou University, Zhengzhou, China, in 2018. She is currently working toward the Ph.D. degree in information and communication engineering with Xidian University, Xi'an, China.

Her research interests include few-shot learning and synthetic aperture radar target recognition.

Chen Zhang received the B.S. degree in electronic and information engineering in 2019 from Xidian University, Xi'an, China, where he is currently working toward the Ph.D. degree in signal and information processing with the National Laboratory of Radar Signal Processing.

His research interests include synthetic aperture radar automatic target recognition and deep learning.