ATGAN: A SAR Target Image Generation Method for Automatic Target Recognition

Zhiqiang Zeng, Xiaoheng Tan, Xin Zhang, Yan Huang, *Member, IEEE*, Jun Wan, *Member, IEEE*, and Zhanye Chen, *Member, IEEE*

***Abstract*—The performance of a deep learning-based synthetic aperture radar (SAR) automatic target recognition (ATR) model largely relies on the scale and quality of training samples. However, it is time-consuming and expensive to collect sufficient data in practice. Although generative adversarial network (GAN) provides a way for SAR target image generation, existing GAN-based methods cannot confirm what features the generator learns, thus they struggle in generating precise SAR target images. In this article, we propose an angle transformation GAN (ATGAN) that can generate azimuth-controllable SAR target images while preserving the target details. The key idea of our ATGAN is to reframe the generation task from the perspective of image-to-image translation. To this end, ATGAN consists of two modules, a coarse-to-fine generator that aims to learn the angle transformation in the deep feature space, and then, apply it to manipulate the representation of an input SAR target image to generate a new one, while a spectral-normalized patch discriminator that tries to estimate the probability that an input SAR target image is real rather than fake using a patch-averaged strategy. Combining with spatial transformer and adversarial training paradigm, ATGAN can generate precise SAR target images for ATR. Extensive experiments verify the effectiveness of the proposed ATGAN, and our method outperforms the state-of-the-art method qualitatively and quantitatively.**

***Index Terms*—Angle transformation (AT), automatic target recognition (ATR), generative adversarial network (GAN), synthetic aperture radar (SAR) target image generation.**

I. Introduction

SYNTHETIC aperture radar (SAR) has the capability to generate high-resolution images of the Earth's surface by actively emitting electromagnetic waves, regardless of meteorological and sunlight conditions [1]. SAR automatic target recognition (ATR) is one of the most considerable research topics in SAR image interpretation [2]. Most of recognition models are data driven [3], [4], [5], [6], [7], [8], and their performances rely strongly on the scale and quality of training data [9]. However, due to the principles of coherent imaging and the characteristics of motion imaging, acquiring sufficient SAR target images for practical use is time-consuming and costly [10], [11]. Moreover, the distribution of strong scattering points and shadow areas in SAR target images is highly sensitive to observation parameters. For instance, the scattering characteristics of a target vary with azimuth angle, but due to the restrictions of the flight routes and operating costs, only SAR target images with limited azimuth angles can be collected in a hanging flight test. The data scarcity problem has become a major hindrance to implement deep learning-based methods for SAR ATR.

To alleviate this problem, virtual images obtained by SAR simulators can be served as supplementary data [12], while the typical SAR simulators can be divided into two categories: echo signal simulation [13], [14] and image simulation [15], [16]. The former type focuses on the process of electromagnetic scattering, typically utilizing Kirchhoff physical optics approximation and geometrical optics approximation [14], while the latter type directly establishes a geometric mapping relationship between the target and the image using techniques, such as the rasterization [15] and ray-tracing approaches [16] for SAR target image simulation. However, both of them are developed based on computer-aided drawing (CAD) models, and the quality of the simulated SAR images depends on how accurate the models are [17]. Deviations in modeling and simplifications made to electromagnetic characteristics will make the simulated SAR target image quite different from the real one. But, in most practical scenarios, it is infeasible to design accurate CAD models and parameters for each target within the scene.

In recent years, generative adversarial network (GAN) [18] has achieved striking successes in various computer vision tasks [19], [20], [21]. The capability to produce SAR target images without any electromagnetic calculation has made GAN an important data augmentation method for SAR ATR. For example, Guo et al. [17] proposed a conditional GAN (CGAN) with a clutter normalization method to generate SAR target images while easing the mode collapse. Wang et al. [22] proposed an improved Wasserstein autoencoder for SAR image generation, and its loss function was designed according to the characteristics of SAR images. Cao et al. [23] proposed a label-directed GAN to provide labeled samples for SAR ATR based on the ideas of Wasserstein GAN (WGAN) and CGAN, which can provide labeled

Manuscript received 27 November 2023; revised 21 January 2024 and 5 February 2024; accepted 13 February 2024. Date of publication 27 February 2024; date of current version 13 March 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62271142, Grant 62201099, Grant 62001062, and Grant U20A20157, and in part by the Opening Project of Guangxi Key Laboratory of Wireless Wideband Communication and Signal Processing under Grant GXKL06230203. *(Corresponding authors: Xiaoheng Tan; Jun Wan.)*

Zhiqiang Zeng, Xiaoheng Tan, Xin Zhang, and Jun Wan are with the School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China (e-mail: txh@cqu.edu.cn; xidianwanjun@163.com).

Yan Huang is with the State Key Laboratory of Millimeter Waves, School of Information Science and Engineering, Southeast University, Nanjing 210096, China, and also with Purple Mountain Laboratory, Nanjing 211100, China.

Zhanye Chen is with the State Key Laboratory of Millimeter Waves, School of Information Science and Engineering, Southeast University, Nanjing 210096, China, and also with the Institute of Electromagnetic Space, Southeast University, Nanjing 210096, China.

Digital Object Identifier 10.1109/JSTARS.2024.3370185

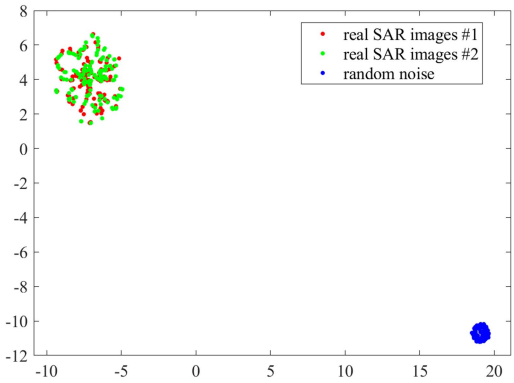


Fig. 1. T-SNE visualization results of feature distribution for real SAR target images and random noise sampled from a Gaussian distribution.

generated samples for SAR target recognition. Meanwhile, Cui et al. [24] employed a Wasserstein GAN with gradient penalty (WGAN-GP) to generate SAR target images and used a support vector machine (SVM) to filter the results. A target recognition method based on constrained naive GAN was proposed in [25] to overcome the problem of low signal-to-clutter-noise ratio. Hu et al. [26] put forward a CVAE-GAN model to generate SAR images with given class labels and observation angles. To generate the SAR target images more faithfully at intended pose angles, a novel GAN-based method for SAR target image generation, called PeaceGAN, was proposed in [27]. Du et al. [8] proposed a multiconstraint GAN to generate high-quality multicategory SAR target images. Recently, Song et al. [9] and [28] proposed an adversarial autoencoder (AAE) for SAR target recognition few-shot learning tasks, which had the ability to generate SAR target images with aspect angular diversity. Despite their SAR target image generation ability, existing GAN-based methods still have difficulty in providing precise results. First, most of GAN-based methods take random noise and category label as input [8], [17], [22], [23], [24], thereby generating the SAR target image with the random azimuth. Second, the significant distribution differences (as shown in Fig. 1) between random noise (sampled from a Gaussian distribution) and real SAR target images make it difficult for these methods to confirm what features the generator learns [28], [29], leading to the degraded performance of target generation. Finally, some GAN-based methods require additional prior knowledge [9], [28] (e.g., semantic maps and pretrained classifiers) as support, which is not available in practical scenarios.

For SAR target image generation, existing GAN-based methods only focus on mapping random noise to the desired images, and thus, they neglect the target feature representation in the training data. Motivated by the notion of representation equivariance [30], we argue that there is a shared knowledge, i.e., angle transformation (AT), between the same target with different azimuth angles in the deep feature space, which can be learned from data and then utilized to enrich samples in the target domain. Thus, different from previous works, we creatively reframe the SAR target image generation task from the perspective of image-to-image translation. Specifically, we propose an angle transformation GAN (ATGAN) that can generate precise SAR target images based on the AT learned from training data. With the aid of spatial transformer and adversarial training manner, ATGAN first learns the AT from training data in the deep feature space. Then, the learned AT is used to manipulate the representation of an input SAR target image to generate a new one in a coarse-to-fine manner. Finally, given a SAR target image with an arbitrary azimuth angle, the trained ATGAN is able to "rotate" the target by a small angle, that is, the new SAR target images with other azimuth angles can be obtained. Compared with existing GAN-based methods, the main idea of the proposed ATGAN is to improve the quality of generated SAR target images by encoding real SAR images. Thus, our ATGAN can generate azimuth-controllable SAR target images while preserving the details of targets without additional classifiers or semantic maps. These generated SAR target images can help to alleviate the data scarcity problem to some degree and promote the recognition performance in SAR ATR. The main contributions of this article are summarized as follows.

1) We creatively reframe the SAR target image generation task from the perspective of image-to-image translation. The proposed scheme can improve the quality of generated images by encoding real samples using the AT learned from training data.
2) We propose an ATGAN, it combines the advantages of spatial transformer and adversarial mechanism. By implementing this method, azimuth-controllable SAR target images can be generated and target details can be preserved.
3) Experimental results demonstrate the effectiveness of generating SAR target images by encoding real SAR images, and our method outperforms the state-of-the-art method qualitatively and quantitatively.

The rest of this article is organized as follows. Section II presents preliminary knowledge about the GAN-based SAR target image generation methods and representation equivariance. The proposed ATGAN is described in detail in Section III, and Section IV presents the experimental results on MSTAR dataset. Section V discusses the contribution of spatial transformer layer (STL) and the transferability of transformation. Finally, Section VI concludes this article.

II. Fundamental

In this section, we briefly review the principles of traditional GAN-based SAR image generation methods and their problems, and then, representation equivariance, i.e., the basis of our idea, is presented for subsequent model designing.

A. GAN-Based SAR Target Image Generation

GAN is arguably one of the most commonly used techniques for generating SAR target images [3]. To control the generated results, traditional GAN-based SAR target image generation methods are based on the idea of CGAN [31], and their architecture can be roughly summarized, as shown in Fig. 2. It mainly consists of two parts: the generator G and the discriminator D. During the process of training, G takes random noise z and class label c as input and generates the corresponding SAR

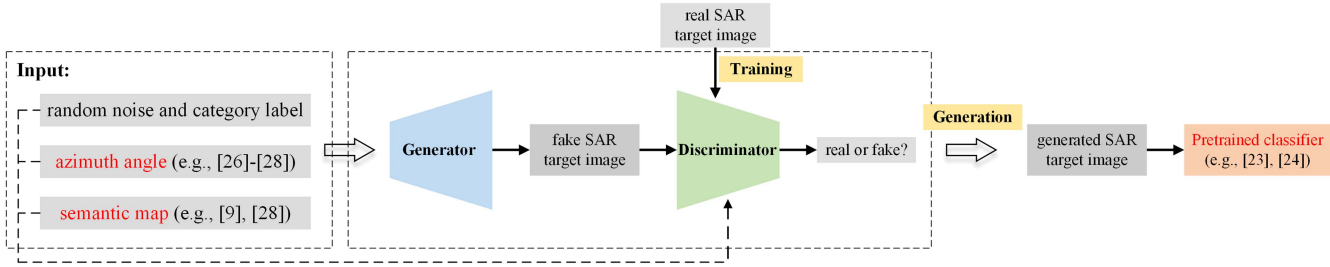


Fig. 2. Architecture of traditional GAN-based SAR target image generation methods. The red font represents additional labels or prior knowledge introduced by existing GAN-based methods to improve the quality of generated SAR target images.

target image to capture the data distribution [17]. For D, it takes both category label c and SAR target images as the input, and outputs a single scalar, which represents the probability that images come from real data rather than the generator G [18]. All in all, the core of these GAN-based methods can still be considered as a two-player minimax game [23], and the generator G and the discriminator D are continuously optimized until they reach Nash equilibrium. As a result, G can generate new SAR target images similar to the real ones for SAR ATR or other applications. The main objective function can be written as

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim P_r}\left[\log D(\boldsymbol{x}|c)\right] + \mathbb{E}_{z\sim P_z}\left[\log\left(1 - D(G(z|c)|c)\right)\right] \quad (1)$$

where $D(\boldsymbol{x}|c)$ denotes the probability that the SAR image $\boldsymbol{x}$ comes from the real data distribution P_r, and $G(z|c)$denotes the sample comes from generated SAR image distribution P_z.

To alleviate the problem of hard training and mode collapse, WGAN-GP was introduced to modify the loss function in (1) [23], [24]. However, this type of GAN-based SAR target image generation method only takes random noise z and target type c as input, resulting in the inability to control the azimuth of the generated SAR image. To address this issue, recent research works introduced azimuth angle $v = [\cos\theta, \sin\theta]$ as an additional input [26], [27], [28]. In addition, the discriminator D is first asked to predict the azimuth angle of its input SAR target image, and then, the purpose of generating a SAR target image with the desired azimuth angle is achieved by minimizing an additional loss function, which can be written as [17]

$$\begin{aligned}\mathcal{L}_{azi} &= 1/N_b \sum_{i=1}^{N_b} \left(|v_r - v_p|\right) \\ &= 1/N_b \sum_{i=1}^{N_b} \left(|\cos\theta_{r,i} - \cos\theta_{p,i}| + |\sin\theta_{r,i} - \sin\theta_{p,i}|\right)\end{aligned} \quad (2)$$

where N_b is the number of SAR target images in a batch, and $v_r = [\cos\theta_r, \sin\theta_r]$ and $v_p = [\cos\theta_p, \sin\theta_p]$ denote the real and predicted azimuth angles, respectively.

However, the regression of the azimuth angles is an underdetermined problem, and sometimes the edges and shapes of targets in SAR images are not sharp or distorted, which will make it even harder for the discriminator D to predict the azimuth angles [17], and too many loss functions may cause the network to be difficult to train. Besides, for better SAR target image generation, some prior knowledge (e.g., semantic maps [9], [28]) or extra methods (e.g., pretrained classifier [23], [24]) are introduced, which are not available in practical application scenarios.

Despite their SAR target image generation ability, existing GAN-based methods still struggle in generating target images with desired azimuth angles while preserving the details of the targets. On the one hand, they take random noise (generally sampled from a standard normal distribution) and conditional information as input, and output a single scalar, tending to generate SAR images that conform to the global distribution, while neglecting local target details [32]. On the other hand, they focus on mapping random noise into a high-dimensional space to gradually approximate the distribution of real SAR images, but neglect the in-depth exploration of random noise. Recently, StyleGAN proposed in [33] also proves that editing random noise vector can control target visual features. More importantly, the distribution differences between random noise and real SAR target images also make it even harder to generate SAR target images with high quality.

Therefore, instead of directly mapping random noise to the desired SAR target images, our idea is to improve the quality of generated SAR target images by using AT to manipulate the features of real samples, which is based on the notion of representation equivariance.

B. Representation Equivariance

Traditional data augmentation techniques, such as cropping, rotation, flipping, and noising [3], can increase the number of the training samples. However, these methods only transform the data in the spatial domain, and the new features cannot be generated to increase the data diversity. Although the effect of these data augmentation approaches is very limited, they give us an inspiration: whether the SAR image features of the same target with different azimuths can be mapped to each other in the deep feature space.

Interestingly, one of the notable properties of representations, that is, equivariance, is presented in [30]. It focuses on how the representation changes upon transformations of the input image. This article demonstrates that most representations, including layers in deep neural networks, change in an easily

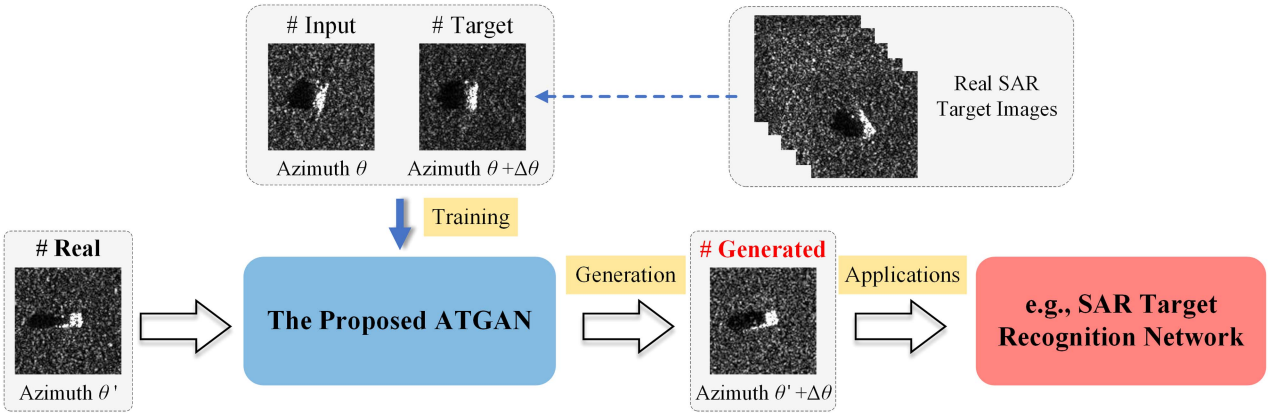


Fig. 3. Schematic of the proposed ATGAN for SAR target image generation and its applications.

predictable manner with the input. More importantly, these equivariant transformations can be learned empirically from available images.

Generally, the convolutional neural networks (CNNs) can be thought of as functions ϕ mapping an image $\boldsymbol{x} \in \chi$ to a vector $\phi(\boldsymbol{x}) \in \mathbb{R}^d$. A representation ϕ is deemed equivariant with a transformation g of the input image when the transformation can be transferred to the representation output. Specifically, equivariance with g is obtained when there exists a map M_g: $\mathbb{R}^d \to \mathbb{R}^d$ such that [30]

$$\forall \boldsymbol{x} \in \chi: \quad \phi(g\boldsymbol{x}) \approx M_g \phi(\boldsymbol{x}). \tag{3}$$

Motivated by representation equivariance, we creatively reframe the SAR target image generation task from the perspective of image-to-image translation. During the training stage, our goal is to obtain the transformed representation of the input SAR target image through the mapping function M_g (i.e., represent the intrinsic AT) satisfying (3). Estimating M_g is naturally formulated as an empirical risk minimization problem [28]. Specifically, given an input image $\boldsymbol{x}_i$ and a target image $\boldsymbol{x}_t$ sampled from a set of SAR target images, learning amounts to optimizing the reconstruction loss

$$E(M_g) = \frac{1}{n} \sum_{i=1}^{n} \ell(\delta(\phi(g\boldsymbol{x}_i)), \delta(M_g \phi(\boldsymbol{x}_i))) \tag{4}$$

where n represents the number of input SAR target images in the training samples, and ℓ denotes reconstruction loss whose choices are described in Section III-C. δ represents the image reconstruction block. According to the assumption of representation equivariance, when $\boldsymbol{x}_t \approx g\boldsymbol{x}_i$, referring to (3), $\boldsymbol{x}_t = \delta(\phi(\boldsymbol{x}_t)) \approx \delta(M_g \phi(\boldsymbol{x}_i))$.

By requiring the same mapping function M_g to work for any SAR image pair, M_g can capture intrinsic AT of the features. Hence, to achieve this goal, our ATGAN is proposed in Section III.

III. Approach

In this section, we first elucidate the framework of the proposed ATGAN for SAR target image generation, as shown in Fig. 3. Then, the three core modules of the proposed network are described in detail. Finally, the loss function of the proposed ATGAN is presented.

A. Framework of Proposed Network

According to the representation equivariance, image representations, such as CNNs change upon transformations of the input image, and these equivariant transformations can be learned empirically from data [30]. Furthermore, the manifold learning theory also demonstrates that the target distribution in SAR images is stable and learnable [34]. Thus, we propose a new GAN-based network, i.e., ATGAN, for SAR target image generation. Benefiting from its specific structure and objective function, the proposed ATGAN can gradually learn the AT through a mapping function M_g in the deep feature space, and then, utilize it to manipulate the representation of the input SAR images to generate realistic SAR target images with specific azimuth angles. The general process of the proposed ATGAN for SAR target image generation is explained as follows.

Given two SAR target images, namely, input SAR image, represented as $\boldsymbol{x}_i$ with azimuth angle θ and target SAR image, denoted by $\boldsymbol{x}_t$ with azimuth angle $\theta + \Delta\theta$. Motivated by (3), our idea is to obtain a transformed representation of the input SAR image $\boldsymbol{x}_i$ through a mapping function M_g. To this end, the generator's encoder first takes $\boldsymbol{x}_i$ as input and maps it from the image to $\phi(\boldsymbol{x}_i)$. Then, the decoder of our generator will generate the fake SAR target image $\boldsymbol{x}_f$ with azimuth $\theta + \Delta\theta$ by reconstructing $M_g \phi(\boldsymbol{x}_i)$, i.e., $\boldsymbol{x}_f = \delta(M_g \phi(\boldsymbol{x}_i))$. In the initial state, the network has not yet learned the AT by the mapping function M_g, resulting in an obvious difference between the fake SAR image $\boldsymbol{x}_f$ and the real SAR image $\boldsymbol{x}_t$. The image pairs, $\{\boldsymbol{x}_i, \boldsymbol{x}_f\}$ and $\{\boldsymbol{x}_i, \boldsymbol{x}_t\}$, are subsequently fed into the discriminator, and the discriminator will determine whether each image pair is real or fake. Through minimizing the reconstruction loss (i.e., $\ell(\boldsymbol{x}_t, \boldsymbol{x}_f = \delta(M_g \phi(\boldsymbol{x}_i)))$) and the adversarial loss, our generator is optimized, and it gradually learns the AT by estimating M_g. Finally, given a real SAR target image $\boldsymbol{x}_r$ with azimuth angle θ', the generator of our ATGAN will generate realistic SAR target image $\boldsymbol{x}_g$ with azimuth angle $\theta' + \Delta\theta$, i.e., $\boldsymbol{x}_g = \delta(M_g \phi(\boldsymbol{x}_r))$.

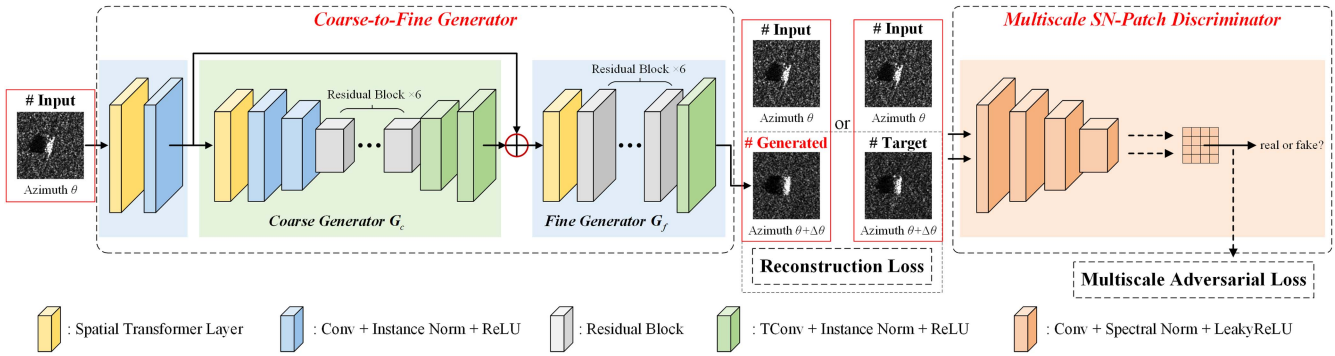


Fig. 4. Architecture of the proposed ATGAN.

Due to the high quality of SAR target images generated by ATGAN, they can be used for data augmentation in SAR ATR or other applications.

B. Specific Implementation of Proposed Network

Given a real SAR target image $\boldsymbol{x}_r$ with azimuth angle θ', we aim to generate its corresponding images $\boldsymbol{x}_g$ with azimuth angle $\theta' + \Delta\theta$. To achieve this goal, the architecture and loss function of our ATGAN are designed, as shown in Fig. 4.

1) Coarse-to-Fine Generator: SAR target images contain rich scattering information of ground objects in a relatively small size [35]. Previous GAN-based methods take random noise z and conditional information c as input and directly generate high-dimensional images [22], [23]. The significant distribution differences between random noise z and real SAR target images make them struggle to generate precise SAR target images. Considering the characteristics of SAR images, we design a coarse-to-fine generator for finer feature representation. Our generator consists of two subnetworks: a coarse generator G_c and a fine generator G_f. The former takes the downsampled real SAR image $\boldsymbol{x}_r$ as input to perform initial coarse generation, while the fine generator G_f generates the final SAR image $\boldsymbol{x}_g$ with better details by taking elementwise sum of two feature maps from both generators. By applying this generator, global and local features can be effectively aggregated for the task of generating SAR target images, while also enlarging the receptive fields and stabilizing training.

To extract the features of SAR target images and learn the AT in the deep feature space, we follow the architecture of encoder–decoder [36] to design our generator. Specifically, both subnetworks mainly consist of three parts: the downsampling block (i.e., a convolutional layer followed with an instance normalization layer [37] and ReLU activation [38]), the residual block [39], and the upsampling block (i.e., a transposed convolutional layer followed with an instance normalization layer and ReLU activation). The residual block is utilized to increase the effective depth of networks. More importantly, we integrate the STL [40] into the encoder–decoder pipeline for better AT learning.

2) Spatial Transformer Layer: To capture the correlation of targets between different azimuth angles, one solution is to estimate a rotation matrix that is given by [41]

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \tag{5}$$

where φ is the angle of the input image $\boldsymbol{x}_i$ with respect to the target image $\boldsymbol{x}_t$. (u, v) are the coordinates of $\boldsymbol{x}_i$, and (u', v') are the coordinates of the transformed SAR images $\hat{\boldsymbol{x}}_t$. This warping operation can be achieved by STL. STL is an opposite of GAN that reconstructs the target SAR image by warping the input image according to the acquired affine matrix. However, for a SAR target image, the gap between input and target is infeasible to bridge by simply warping the pixels from one to another, and thus, STL tends to generate blurry images (see the SAR target images in Fig. 5). Therefore, instead of directly taking $\hat{\boldsymbol{x}}_t$ as the final generation result, we utilize STL as the guidance for feature representations. As a consequence, the proposed generator and designed STL can complement each other for estimating M_g. In this way, both of two strategies can maximize their advantages for generating azimuth-controllable SAR target images while preserving target details. Moreover, STL can automatically select interested regional features in the process of training [40], which is also important for SAR target image generation.

As illustrated in Fig. 6, our STL consists of three parts: localization network, grid generator, and sampler. Specifically, the localization network first takes the feature map $I \in \mathbb{R}^{H\times W\times C}$

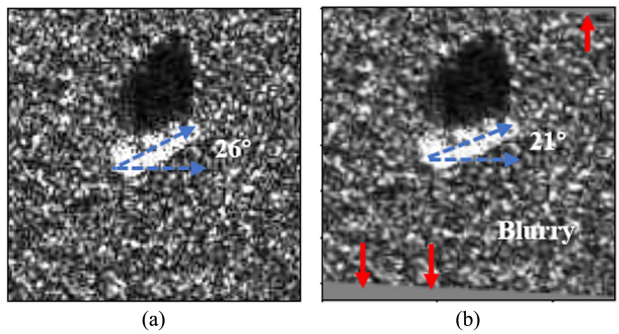


Fig. 5. Example of the real SAR image of the target BTR60 and its corresponding SAR image transformed by the STL. (a) Original image of BTR60. (b) Transformed image of BTR60.

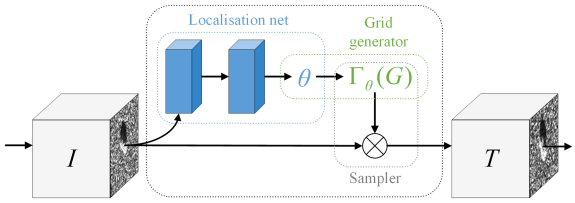


Fig. 6. Architecture of the STL.

as input, and outputs the affine parameters $\theta = f_{\text{loc}}(I)$ through hidden layers. Then, the grid generator utilizes these parameters to create a sampling grid $G = \{G_i\}$. Finally, the input feature map and sampling grid are input to the sampler, generating the target feature map $T \in \mathbb{R}^{H\times W\times C}$. As a result, STL has the ability to actively spatially transform the feature maps or the input SAR target images into the desired form of the next layer. Note that as a differentiable module, our STL can be plugged into the generator and trained with it, without any extra training supervision or modification to the optimization process [40].

3) Multiscale Spectral-Normalized Patch Discriminator: Generating high-quality SAR target images is also a significant challenge for designing discriminators. To this end, we deploy two subnetworks D_1 and D_2 as discriminators, which have the same structure but work at different scales to obtain a larger receptive field. This design also makes our generator much easier to train.

The features of the target in SAR images are highly sensitive to the azimuth angle, which also directly affects the performance of the ATR model [41]. However, existing GAN-based methods rely on a global discriminator, which maps from a whole SAR target image to a single scalar output. To improve the details of the generated targets, we introduce PatchGAN [42] as the critic. As a consequence, each subnetwork maps the input SAR target image to an $N \times N$ matrix by convolutional operation, while each element in this matrix represents a patch in the original SAR target image. Therefore, our discriminator focuses on determining whether these image patches are real or fake rather than the whole SAR target image. Benefited from this strategy, the proposed ATGAN is able to improve the global consistency as well as local details. Moreover, we employ spectral normalization (SN) [43] to further stabilize the network training, which can control the Lipschitz constant of the discriminator, converges faster and outperforms WGAN-GP [44].

C. Loss Function

As the goal of the network, the loss function of the proposed ATGAN mainly consists of two parts: reconstruction loss and multiscale adversarial loss, which evaluate the generated SAR target images from different perspectives.

1) Reconstruction Loss: The L_1 norm is utilized to minimize the pixelwise difference between real target SAR images and generated SAR images to ensure the image content consistency, which also facilitates the estimation of the mapping function M_g. Thus, the reconstruction loss in (4) can be represented as

$$\mathcal{L}_{\text{rec}}(G) = \mathbb{E}_{\boldsymbol{x}_i,\boldsymbol{x}_t\sim\mathcal{P}_r}\left[\|\boldsymbol{x}_t - G(\boldsymbol{x}_i)\|_1\right] \tag{6}$$

where $\|\cdot\|_1$ denotes the L_1 norm, and $\boldsymbol{x}_i$ and $\boldsymbol{x}_t$ are input and target SAR images sampled from real distribution $\mathbb{P}_r$, respectively. $G(\boldsymbol{x}_i)$ is the generated SAR target image conditioned on the input $\boldsymbol{x}_i$. By minimizing the reconstruction loss, the network is trained to estimate the mapping function M_g

2) Multiscale Adversarial Loss: However, using only the L_1 norm tends to produce blurry SAR target images, and it only obtain the outline of the image rather than high-frequency details [45]. To alleviate this problem, we train our generator G and discriminator D in an adversarial manner. Specifically, the generator G is trained to learn a mapping from input image $\boldsymbol{x}_i$ to target image $\boldsymbol{x}_t$, and then generate $G(\boldsymbol{x}_i)$ as realistic as possible, which tricks D into believing that the generated SAR target image $\boldsymbol{x}_g = G(\boldsymbol{x}_i)$ is sampled from the real data [18]. In contrast, the discriminator D is trained to differentiate the real SAR target image $\boldsymbol{x}_t$ and generated SAR target image $\boldsymbol{x}_g$. The adversarial loss of proposed ATGAN can be expressed as

$$\begin{aligned}\mathcal{L}_{\text{adv}}(G,D) = {} & \mathbb{E}_{\boldsymbol{x}_i,\boldsymbol{x}_t\sim\mathbb{P}_r}\left[\log D\left(\boldsymbol{x}_i,\boldsymbol{x}_t\right)\right] \\ & + \mathbb{E}_{\boldsymbol{x}_i\sim\mathbb{P}_r}\left[\log\left(1 - D\left(\boldsymbol{x}_i, G(\boldsymbol{x}_i)\right)\right)\right]\end{aligned} \tag{7}$$

where the generator G tries to minimize this function and the discriminator D tries to maximize it.

With the multiscale discriminators, the objective of the proposed ATGAN in (7) then becomes a multitask objective function, which can be expressed as

$$\min_G \max_{D_1,D_2} \sum_{k=1,2} \mathcal{L}_{\text{adv}}(G, D_k). \tag{8}$$

3) Final Objective: Our final objective function for training the whole model is defined as follows:

$$G^* = \arg\min_G \max_{D_1,D_2} \sum_{k=1,2} \mathcal{L}_{\text{adv}}(G, D_k) + \lambda\mathcal{L}_{\text{rec}}(G) \tag{9}$$

where λ denotes the hyperparameters used to balance losses, which is set as 120 in this article.

IV. Experiments and Results

In this section, after introducing the dataset configuration and hyperparameters settings, the performance of the proposed ATGAN for SAR target image generation will be evaluated qualitatively and quantitatively.

A. Datasets and Settings

1) Experimental Dataset: The moving and stationary target acquisition and recognition (MSTAR) dataset [46] is utilized in this article for the validation of our proposed ATGAN. As a widely used benchmark for SAR algorithm applications, MSTAR includes 0.3 m $\times$ 0.3 m resolution *X*-band SAR images of ten categories of ground stationary targets. In this study, seven types of targets (i.e., 2S1, BRDM2, BTR60, D7, T62, ZIL131, and ZSU234) at 0°–360° azimuth angles and 15° depression angle are selected for the proposed ATGAN performance assessment. All the SAR target images are normalized by min–max scaling to the range [0, 255]. The optical and corresponding 128 $\times$ 128 SAR images of these targets are shown in Fig. 7.

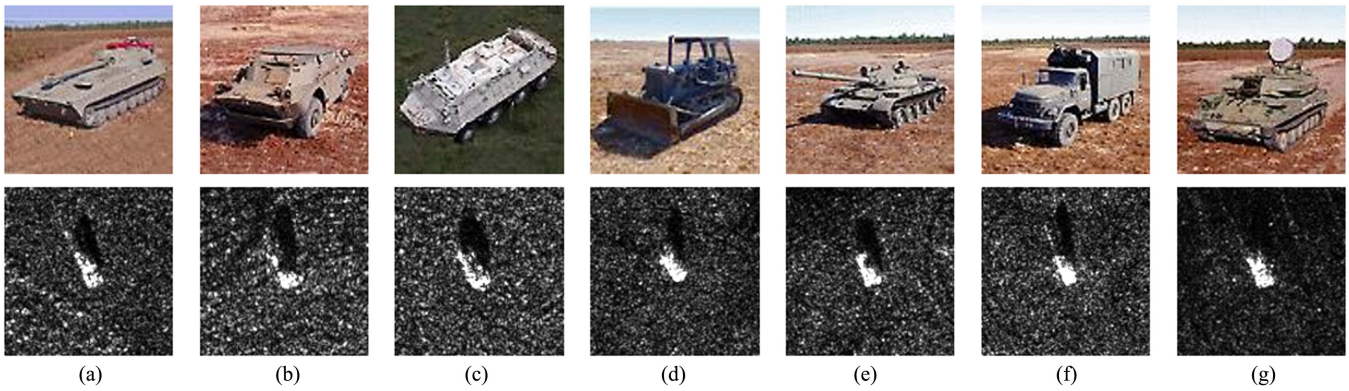


Fig. 7. Optical images and their corresponding MSTAR SAR images for (a) 2S1, (b) BRDM2, (c) BTR60, (d) D7, (e) T62, (f) ZIL131, and (g) ZSU234.

TABLE I
SEVEN TYPES OF TARGETS IN THE MSTAR DATASET

Target Category	Original		Training	Testing
	Depression	Number	Number	Number
2S1	15°	274	45	229
BRDM2	15°	274	46	228
BTR60	15°	195	45	150
D7	15°	274	51	223
T62	15°	273	47	226
ZIL131	15°	274	48	226
ZSU234	15°	274	50	224

Note that the original chips of size 128 × 128 in MSTAR dataset instead of cropped SAR target images are used to train the proposed ATGAN, which is of potential use in SAR ATR.

As mentioned earlier, only SAR target images with limited azimuth angles can be collected in practical applications [17]. Thus, to simulate the actual SAR target image generation task, we first select one sample at every 10° interval as the input image $\boldsymbol{x}_i$, and then find the sample with 10° interval behind the input image $\boldsymbol{x}_i$ as the target image $\boldsymbol{x}_t$. As a result, only 332 of 1838 SAR target images are set as the training dataset, and all the remaining images are only used for testing purposes, as given in Table I.

2) Hyperparameters Settings: Our ATGAN is implemented on the PyTorch framework, performed on NVIDIA GeForce RTX 3080Ti with a compute capability of 8.6. We train our ATGAN with Adam optimizer [47] ($\beta_1 = 0.5$; $\beta_2 = 0.999$), and the batch size and learning rate are set to 4 and 0.0002, respectively.

B. Qualitative Evaluation

The basic goal of a SAR target image generation method is to correctly reproduce the distribution of the dominant scatters in the SAR images. To evaluate the generation capability of the proposed ATGAN, a subjective evaluation of the generated SAR target images by human visual is described, that is, we first compare the ATGAN-generated SAR target images with the real ones in visual similarity.

Fig. 8 presents the qualitative results of the targets BRDM2 and BTR60, where the former is most likely to be confused with the latter. It can be clearly seen that the generated SAR target images can accurately capture the geometric features and morphological structures in real ones. Thus, the ATGAN-generated SAR target images are highly similar to the real SAR target images not only in terms of the azimuth angle but also the target features and shadow areas (as indicated by the red angle rulers and yellow rectangles), which play an important role in SAR ATR. In contrast, their backgrounds are not necessarily the same, indicating that the proposed ATGAN does not merely perform pixel-level mapping. Furthermore, upon comparing the two sets of pictures, it can be observed that although the real SAR target images of BRDM2 and BTR60 targets may be easily confused at certain azimuth angles, their corresponding generated SAR target images are still more similar to themselves than to other targets.

Fig. 9 compares the generated SAR images of the seven types of targets with their two nearest real SAR target images in the training data. The purpose is to intuitively validate the effectiveness of the proposed ATGAN model by using the most similar training samples as a standard, that is, the new SAR target images generation relies on learned AT rather than mere copying or mapping of training samples. It is not difficult to find that the generated SAR target images are different from their

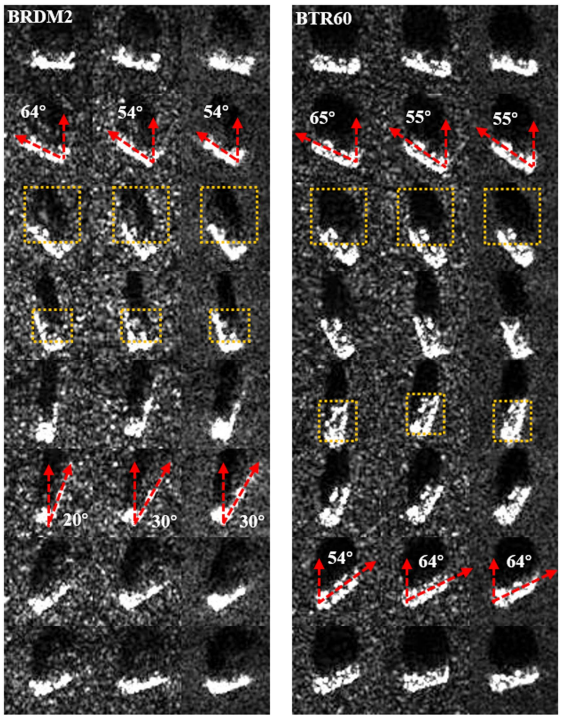


Fig. 8. Input SAR images (left column), target SAR images (middle column) compared with generated SAR images (right column) of (a) BRDM2 and (b) BTR60.

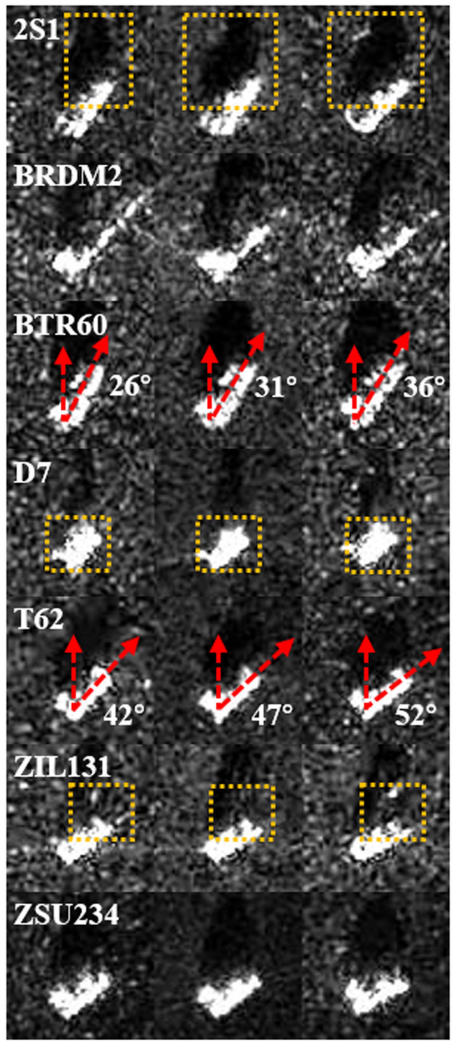


Fig. 9. Generated SAR target images (middle column) compared with the two nearest real SAR target images (left and right columns). From top to bottom are 2S1, BRDM2, BTR60, D7, T62, ZIL131, and ZSU234.

nearest training samples, particularly in terms of local target features and shadow areas (as indicated by the red angle rulers and yellow rectangles). This also demonstrates that our ATGAN can generate unseen SAR target images by "interpolating" along the azimuth angle of the target, just like other SAR target image generation method.

To better verify the effectiveness of the proposed ATGAN, we also select AAE [25], a state-of-the-art SAR target image generation method, for comparison. Furthermore, to facilitate visual similarity comparison, we appropriately modified the training dataset configuration, that is, the azimuth angles of the training samples are scaled down to $0^\circ - 50^\circ$, while azimuth angle interval of adjacent input images is set to 5°. For a fair comparison, we reproduce AAE and set the same sample azimuth angle interval on the training dataset.

Fig. 10 shows a visual similarity comparison between real SAR target images (upper row of each panel) and those generated by AAE (middle row of each panel) and ATGAN (bottom row of each panel) for the six types of targets in the test set. It can be observed that the ATGAN-generated SAR target images have a clear target profile, obvious scattering characteristic, and they are more similar to the real ones in terms of azimuth angle. More importantly, our ATGAN has the ability to generate SAR target images beyond the available azimuth angles to some extent. In other words, the proposed ATGAN can generate new SAR target images by learning the AT. Compared with ATGAN, the quality of the SAR target images generated by AAE is poorer. AAE struggles in reproducing local details and azimuth angles (as indicated by the red angle rulers and yellow rectangles) and fails to generate the desired SAR target images beyond 50° azimuth angle (as shown in the red rectangles). We argue that two main reasons contribute to these problems. On the one hand, AAE takes random noise and conditional information as input and tends to generate SAR target images that conform to the global distribution. Due to the significant difference between the distribution of random noise and real SAR target images (as shown in Fig. 1), AAE relies heavily on "imagine" targets and cannot confirm what features the generator learns, leading to the quality of generated SAR target images degradation. On the other hand, although AAE performs a regression process to obtain the desired azimuth angle, the regression of the azimuth angle is an underdetermined problem [17], and some unsatisfactory target features make it even harder. Furthermore, the semantic maps of the test set, which are not available in the practical scenario, are fed into the AAE as prior knowledge. However, due to the speckle noise in SAR target images, it is difficult for traditional semantic map generation methods to

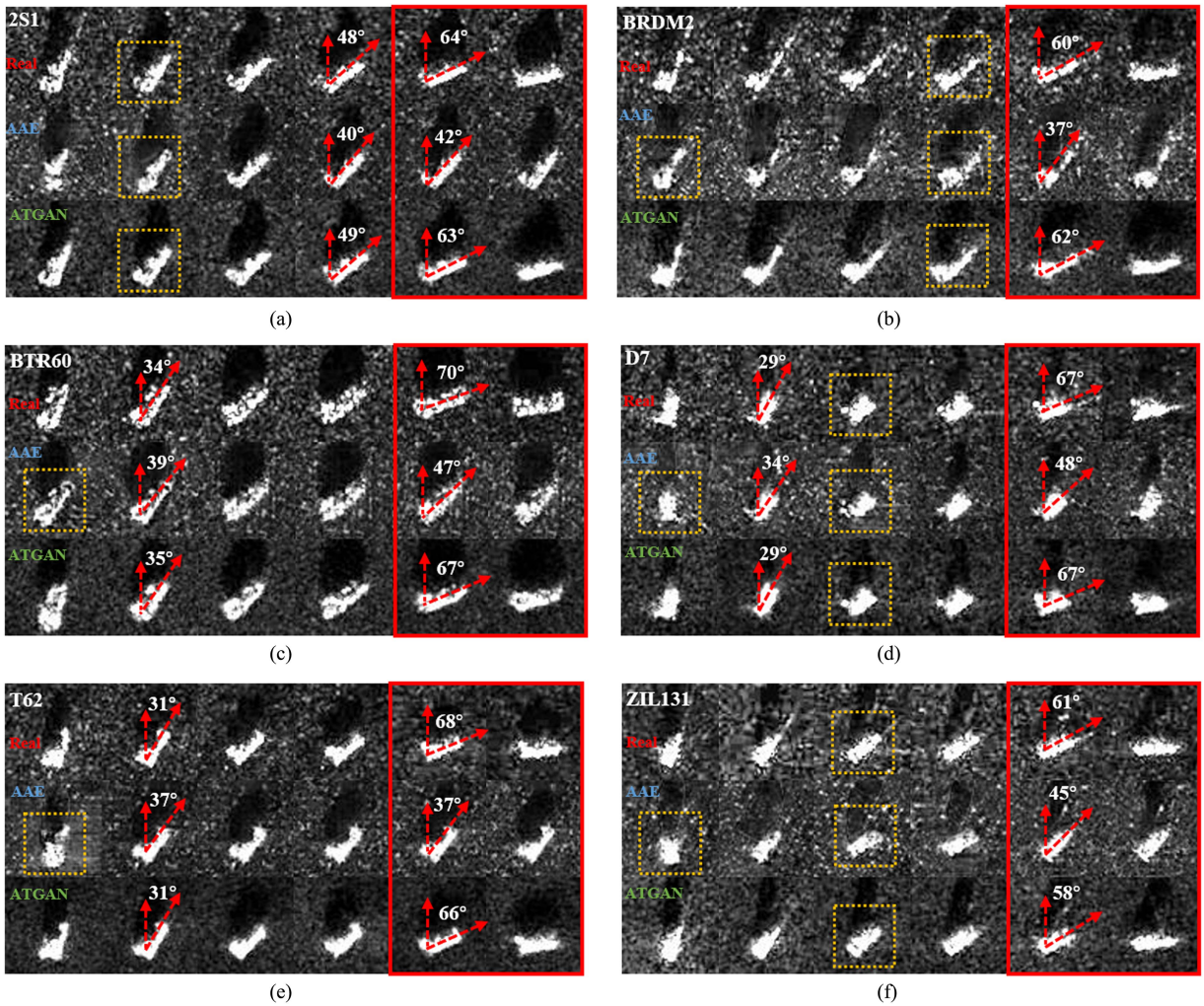


Fig. 10. Comparison of real SAR target images (upper row) of (a) 2S1, (b) BRDM2, (c) BTR60, (d) D7, (e) T62, and (f) ZIL131, with the corresponding generated SAR images via AAE (middle row), and generated SAR images via ATGAN (bottom row).

generate accurate semantic maps for targets and shadows, and thus makes AAE even harder to correctly preserve target shape and shadow areas.

Through the above visual similarity comparison presentation, it demonstrates to some extent that the proposed ATGAN has the ability to generate precise SAR target images with desired azimuth angle meanwhile preserving target details. Although this evaluation criterion is simple and intuitive, it is highly subjective and susceptible to evaluator bias. In other words, it is not enough to rely solely on human eye to intuitively assess the quality of the generated SAR target images. Therefore, an objective evaluation framework is designed to evaluate the quality of generated SAR target images quantitatively in Section IV-C.

C. Quantitative Evaluation

Apart from proposing a meticulously designed CNN architectures and loss functions for SAR target image generation, it is also crucial to establish an objective framework for evaluating the authenticity and reliability of generated SAR target images [48]. The quality of the generated samples directly affects the performance of the SAR ATR model. However, to the best of the authors' knowledge, current works in the literature largely rely on the histogram to approximate the magnitude distribution of the SAR target image [23], [28], and there is no comprehensive framework available that can fully illustrate the accuracy of generated SAR target images. Therefore, in this work, an objective evaluation framework is designed to quantitatively evaluate the quality of the generated SAR target images.

According to the different information concerned, the designed framework consists of three parts: structure level, feature level, and application level. Specifically, these three parts quantitatively measure the similarity between generated and real SAR target images from various perspectives, thereby forming an objective and comprehensive evaluation framework. Different parts calculate the similarity of images using different indicators, each with its own advantages and limitations. The flowchart of the designed evaluation framework is depicted in Fig. 11.

1) Structure Level: This section focuses on measuring the similarity through structural information. Specifically, it consists of two indicators: multiscale structural similarity (MSSIM) and feature similarity index measure (FSIM).

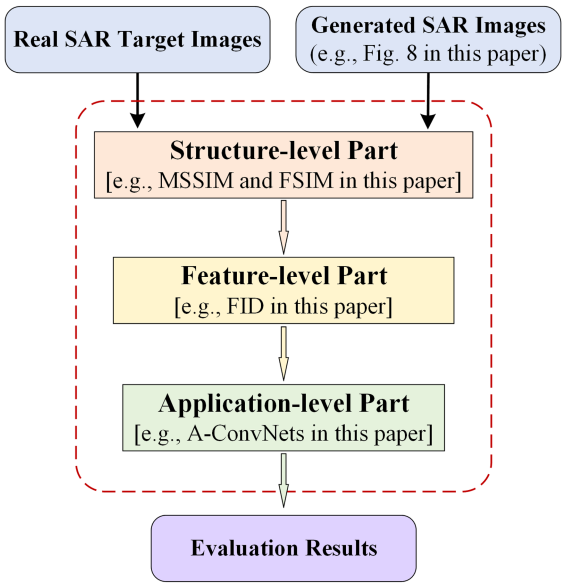


Fig. 11. Flowchart of the designed objective evaluation framework.

On the basis of structural similarity (SSIM) [49], MSSIM [50] measures the similarity between generated and real SAR target images at different scales in terms of contrast and structural information. In this way, it can reduce the interference caused by different image resolutions and obtain more accurate evaluation results. The MSSIM index is calculated as follows:

$$\mathrm{MSSIM}(R,G) = [l_M(R,G)]^{\alpha_M} \sum_{j=1}^{M} [c_j(R,G)]^{\beta_j} [s_j(R,G)]^{\gamma_j} \tag{10}$$

where $l(R,G)$, $c(R,G)$, and $s(R,G)$ are designed to compare luminance, contrast, and structural information of R and G, respectively. $\alpha, \beta, \gamma > 0$ are used to adjust the weights of these information. M denotes the scale reduction factor.

For optical images, the profile and edges of the local target can be well described by gradients [23]. However, extracting gradient features from SAR target images is a challenging task due to the presence of speckle noise. To alleviate this issue, we employ FSIM [51] to capture phase congruency (not SAR phase information is meant here), which has been proved to be robust against variations in noise levels [52]. By applying FSIM, phase consistency and gradient features can complement each other to measure the similarity between generated and real SAR target images. The FSIM index is defined as follows:

$$\mathrm{FSIM} = \frac{\sum_{x\in\Omega} S_L(x) \cdot \mathrm{PC}_m(x)}{\sum_{x\in\Omega} \mathrm{PC}_x(x)} \tag{11}$$

$$S_{\mathrm{PC}}(x) = \frac{2\mathrm{PC}_1(x) \cdot \mathrm{PC}_2(x) + T_1}{\mathrm{PC}_1^2(x) + \mathrm{PC}_2^2(x) + T_1} \tag{12}$$

$$S_G(x) = \frac{2G_1(x) \cdot G_2(x) + T_2}{G_1^2(x) + G_2^2(x) + T_2} \tag{13}$$

$$S_L(x) = [S_{\mathrm{PC}}(x)]^{\alpha} [S_G(x)]^{\beta} \tag{14}$$

$$\mathrm{PC}_m(x) = \max(\mathrm{PC}_1(x), \mathrm{PC}_2(x)) \tag{15}$$

where $\mathrm{PC}_1(x)$ and $\mathrm{PC}_2(x)$ represent the phase consistency of real and generated SAR images, respectively. $G_1(x)$ and $G_2(x)$ represent the gradient of two images. T_1 and T_2 are positive constants. $S_{\mathrm{PC}}(x)$ is the similarity measure for $\mathrm{PC}_1(x)$ and $\mathrm{PC}_2(x)$, and $S_G(x)$ is the similarity measure for $G_1(x)$ and $G_2(x)$. α and β are parameters used to adjust the relative importance of two features. Ω means the whole image spatial domain.

The structure-level part is capable of quantifying the similarity between real and generated SAR target images by means of extracting their structural information. Taking Fig. 8 as an example, the MSSIM and FSIM values of the real and generated BTR60 are 0.6340 and 0.7063, respectively, while the MSSIM and FSIM values of the real BRDM2 and generated BTR60 are 0.3608 and 0.5952, respectively. Similarly, the MSSIM and FSIM values of the real and generated BRDM2 are 0.5267 and 0.6902, compared with 0.3850 and 0.6255 of real BTR60 and generated BRDM2. These results also quantitatively support the analysis of the visual similarity comparison, as shown in Fig. 8.

However, relying solely on structural information extracted by structure-level part is insufficient for evaluating the quality of generated SAR target images. This part mainly focuses on the structural information of the target in SAR images, which may neglect other key features of SAR target images.

2) Feature Level: This section is devoted to evaluating the quality of the generated SAR target images at the feature level. Consequently, we introduce Fréchet inception distance (FID) [53] for this purpose.

FID is a metric that can quantify the similarity between the distributions of real and generated SAR target images.

Specifically, it involves extracting feature representations from both sets of images using a pretrained deep neural network, followed by computing their Fréchet distance. A lower FID score indicates greater resemblance between the generated and real SAR images.

3) Application Level: According to Guo et al.[48], the utility of the generated SAR target images can be evaluated by SAR target recognition networks, which is the initial purpose of SAR target image generation. Therefore, the application-level part utilizes SVM, CNN-SAR [54], and A-ConvNets [55], i.e., multiple SAR target recognition models, as the backbone of the evaluation network. As a consequence, the quality of the generated SAR target image can be evaluated by how close its classification results are to that of the real SAR target image. By applying this part, we can measure the similarity between the generated and real SAR target images under the premise of comprehensively considering the target features and data distribution.

In this section, the designed evaluation framework is used to evaluate the quality of generated SAR target images. The model is trained with the datasets and settings given in Section IV-A. To make a fair comparison, the same datasets and test steps are applied to AAE.

The first part of the designed evaluation framework can measure the similarity between the generated and real SAR target images at the structure level. Table II presents the average values of MSSIM and FSIM computed between the real SAR target images and those generated by AAE and ATGAN on the test set.

TABLE II
STRUCTURE-LEVEL EVALUATION RESULTS OF REAL IMAGES WITH CORRESPONDING AAE-GENERATED AND ATGAN-GENERATED IMAGES ON MSTAR DATASET

	AAE							ATGAN						
	2S1	BRDM2	BTR60	D7	T62	ZIL131	ZUS234	2S1	BRDM2	BTR60	D7	T62	ZIL131	ZSU234
MSSIM (↑)	0.4931	0.3859	0.5542	0.4449	0.4394	0.4162	0.5391	**0.6144**	**0.5129**	**0.6483**	**0.5899**	**0.6166**	**0.5849**	**0.6529**
FSIM (↑)	0.6680	0.6685	0.6839	0.6941	0.6649	0.6801	0.6750	**0.7134**	**0.6869**	**0.7100**	**0.7374**	**0.7241**	**0.7347**	**0.7227**

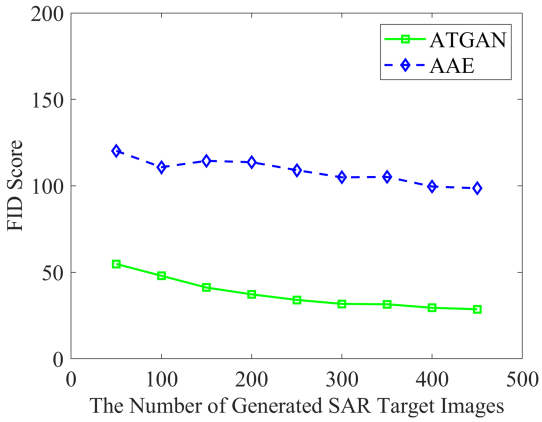


Fig. 12. Comparison results of FID scores for AAE-generated and ATGAN-generated SAR target images.

For these seven types of targets, the average MSSIM values of AAE range from 0.3859 to 0.5542, and the average FSIM values of AAE range from 0.6649 to 0.6839. On the contrary, the average similarity between real and ATGAN-generated SAR target images calculated by MSSIM and FSIM ranges from 0.5129 to 0.6529, and 0.6869 to 0.7374, respectively. The quantitative results presented in Table II further demonstrate that the SAR target image generated by ATGAN is more similar to the real SAR image than AAE-generated SAR image. In other words, our ATGAN model performs competitively against AAE.

For feature-level part, Fig. 12 shows a comparison of FID scores between AAE-generated and ATGAN-generated SAR target images with real ones. The horizontal axis represents the number of generated SAR target images, while the vertical axis is the FID score between the generated and real SAR target images. It can be observed from Fig. 12 that the FID score between the ATGAN-generated and real SAR target images is lower than that of the AAE-generated images. This demonstrates that the SAR target images generated by our ATGAN are more similar to the real SAR images than that of AAE. Specifically, the FID scores of ATGAN range from 54.8435 to 28.1498, compared with 120.2500 to 98.6148 of AAE. Although both generative networks' FID scores decrease as the number of generated SAR target images increases, the FID score curve of ATGAN is smoother than that of AAE, which proves that the proposed network is more stable in SAR target image generation, and the generated SAR target images are closer to the real one in terms of feature distribution.

Based on the above quantitative assessment experiments implemented by two parts in the designed evaluation framework, it is not difficult to find that compared with AAE, the SAR target images generated by ATGAN are more similar to the real SAR target images, and they have reproduced the characteristics, such as the target shape and the shadow areas. As mentioned above, the proposed ATGAN aims to alleviate the scarcity of SAR target images for data-driven SAR ATR. However, it is not clear whether the ATGAN-generated SAR target images have application potential in ATR. Therefore, in application-level part, these generated SAR target images will be utilized as a supplement to the existing training samples for SAR ATR to evaluate their quality. In other words, the higher the classification results of target recognition models obtained using the generated SAR target images, the more similar they are to real SAR target images.

In the application-level part, we first randomly selected 100 real SAR target images as training samples. Then, to evaluate the performance of the generated SAR target images across different scenarios, experiments using augmented data of different sizes are performed. For example, if the training set size is set to 0.9, then 90% of the 500 SAR target images (real or generated) are added to the training samples and retrain the target recognition network. Fig. 13 illustrates the classification results of three target recognition models using real, AAE-generated, and ATGAN-generated SAR target images. The horizontal axis is the size of augmented SAR target images used for training, while the vertical axis represents accuracy, precision, recall, and F1-score, respectively. Taking Fig. 13(c), that is, the classification results of A-ConvNets, as an example. It can be seen clearly that when the training samples are insufficient, these four metrics increase significantly as the three types of SAR target images are gradually added. However, compared with AAE, the role that ATGAN-generated SAR target images play in data augmentation is closer to that of real ones. When the number of augmented SAR target images is 450 (i.e., training set size is 0.9), the classification accuracies of A-ConvNets using real, ATGAN-generated, and AAE-generated SAR target images on the test set are 0.8760, 0.8394, and 0.7870, respectively.

Similarly, the above conclusion can also be verified by the precision, recall, and F1-score of A-ConvNets. In summary, the SAR target image generated by ATGAN is more similar to the real image than that of AAE, indicating that the proposed method has the ability to assist in SAR ATR.

V. DISCUSSION

A. Role of the STL

In this article, the STL [40] is introduced to guide the generator to map the input SAR images to the target SAR images.

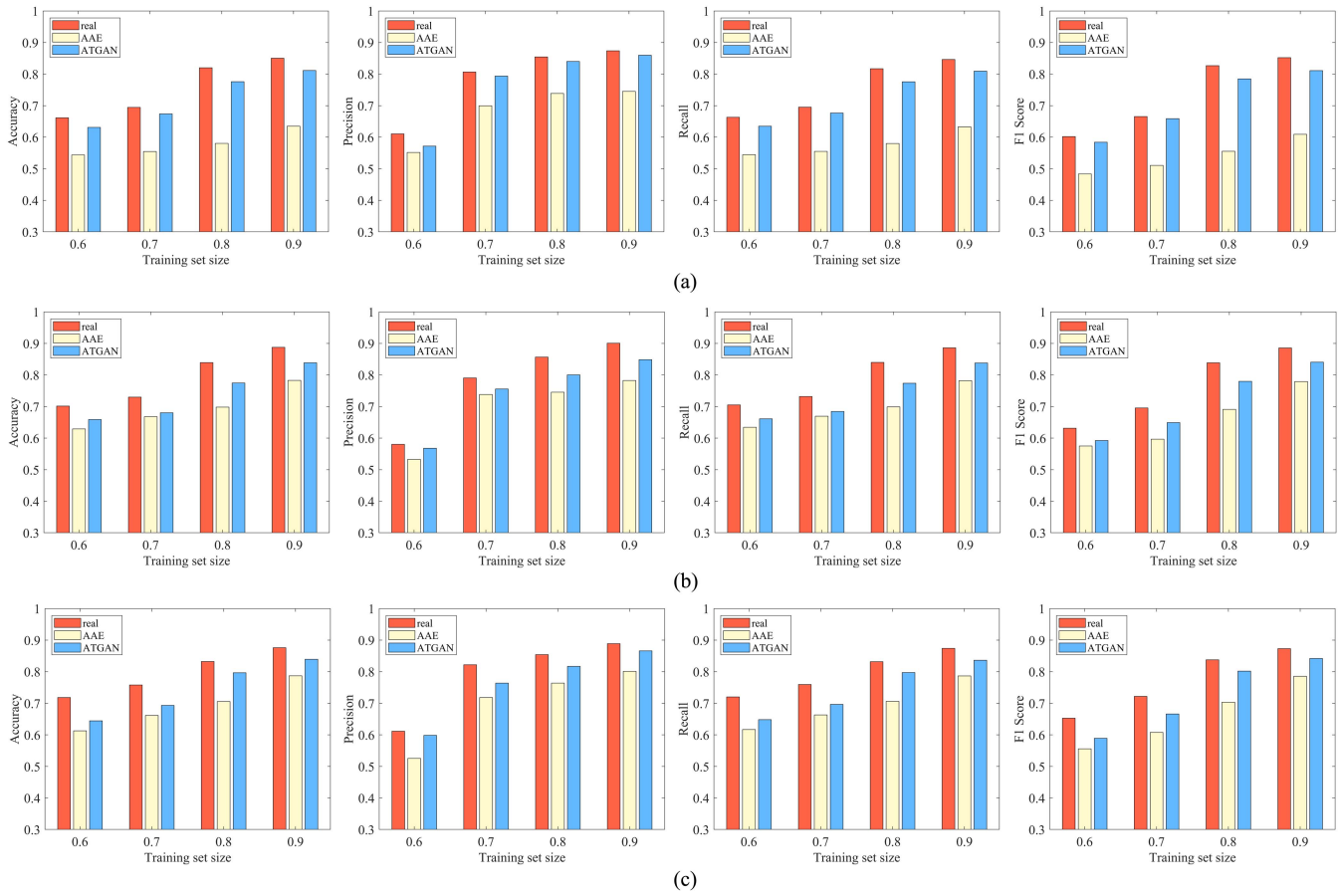


Fig. 13. Classification results of (a) SVM, (b) CNN-SAR, and (c) A-ConvNets, using real, AAE-generated, and ATGAN-generated SAR target images.

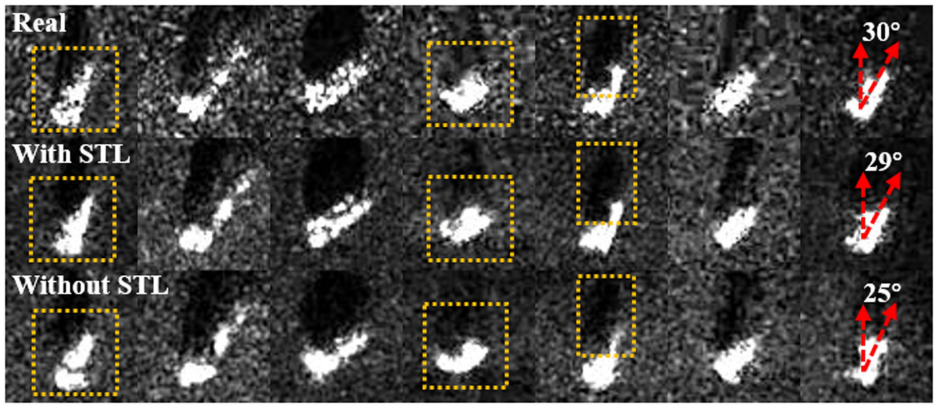


Fig. 14. Comparison of real SAR target images (upper row), with the corresponding generated SAR target images via ATGAN (middle row) with STL, and generated SAR images via ATGAN without STL (bottom row).

To analyze the efficacy of the STL in the proposed ATGAN network, Fig. 14 compares the SAR target images generated by ATGAN with or without STL. As mentioned earlier, the estimation of M_g and the generation of the SAR target images are mainly implemented in the encoder–decoder pipeline. Thus, without STL, ATGAN can still generate SAR target images to a certain extent. However, it can be seen clearly that the shapes and edges of the targets in some SAR target images generated by ATGAN without STL are distorted (e.g., the first and fourth columns of Fig. 14), resulting in deviations from the real targets. Besides, the shadow areas and the azimuth angles of some targets (e.g., the fifth and seventh columns of Fig. 14) are also inconsistent with the real ones. On the contrary, with the STL, the proposed ATGAN can generate correctly desired SAR target images. Table III also gives the quantitative evaluation results between the real SAR target images and images generated by

TABLE III
AVERAGE EVALUATION RESULTS BETWEEN REAL AND ATGAN-GENERATED SAMPLES WITH AND WITHOUT STL

	MSSIM (↑)	FSIM (↑)
Proposed	0.5955	0.7171
Without STL	0.4789	0.6853

TABLE IV
NUMBER OF TRAINING AND TESTING DATASET FOR THE EVALUATION OF THE TRANSFERABILITY

Target Category	Training		Testing	
	Depression	Number	Depression	Number
2S1	15°	274	30°	288
BRDM2	15°	274	30°	287
BTR60	15°	195	30°	/
D7	15°	274	30°	/
T62	15°	273	30°	/
ZIL131	15°	274	30°	/
ZSU234	15°	274	30°	288

ATGAN with and without the STL. It can be observed from Table III that the STL improves the average similarity between the generated image and real image by about 0.0906 to 0.1166, which demonstrates that it contributes a lot to the SAR target image generation. More importantly, STL can make network training more stable in the desired direction, and ATGAN using STL can take less time to achieve convergence.

B. Transferability of Angle Transformation

In this article, we assume that the shared knowledge between the source domain and the target domain is the AT. Motivated by representation equivariance, the main idea of the proposed ATGAN is to learn the AT from training data, and then apply it to reproduce targets with missing azimuth angles in the target domain. To validate the transferability of the AT learned by ATGAN between different domains, we also conduct qualitative and quantitative experiments on the MSTAR dataset. Specifically, given an input SAR image $\boldsymbol{x}_i$ and a target SAR image $\boldsymbol{x}_t$ sampled from training data with 15° depression angle, the proposed ATGAN is first trained to obtain a transformed representation of $\boldsymbol{x}_i$ through the mapping function M_g. Then, by optimizing the loss function (9) and requiring the same M_g to work for any SAR image pair, M_g can capture intrinsic AT of the representation. Finally, given a real SAR target image $\boldsymbol{x}_r$ with 30° depression angle as input, our generator can generate a new SAR target image $\boldsymbol{x}_g$ for ATR. Similarly, the quality of the generated SAR target image $\boldsymbol{x}_g$ will be evaluated qualitatively and quantitatively. The details of the dataset configuration are given in Table IV.

Fig. 15 shows the comparison between real SAR target images (upper row of each panel) and the SAR target images generated by the proposed ATGAN (middle row of each panel). It is not difficult to find that the generated images closely resemble their corresponding real SAR target images in terms of the shape and characteristics of targets and azimuth angle (as indicated by the real angle rulers and yellow rectangles). Moreover, due to the differences in depression angle, there may be variations in shadow range for a given target at same azimuth angle between training and test sets. Therefore, we also compare the generated SAR target images with real training samples (bottom row of each panel) under a depression angle 15°. It can be clearly seen that the shadow range of the generated SAR target image is quite different from that of the real image in the training data, and it is more similar to the expected target SAR image than to the former (as indicated by the yellow rectangles). These comparison results demonstrate that the proposed ATGAN model strives to learn the AT rather than merely duplicating existing training data.

TABLE V
AVERAGE EVALUATION RESULTS BETWEEN REAL AND ATGAN-GENERATED SAMPLES ON THREE TYPES OF TARGETS

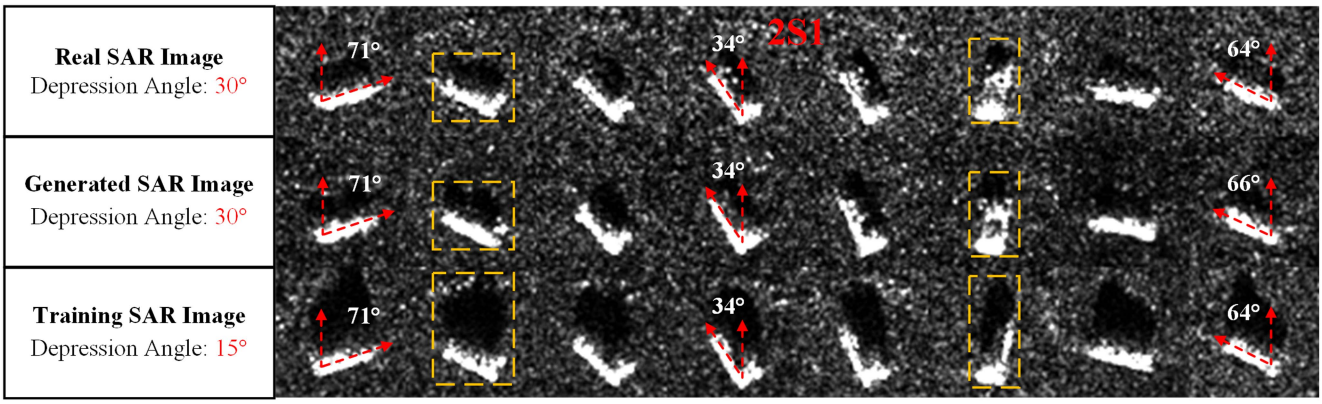
Target Category	Generated versus Real		Generated versus Training	
	MSSIM (↑)	FSIM (↑)	MSSIM (↑)	FSIM (↑)
2S1	0.4792	0.6815	0.3873	0.6530
BRDM2	0.4934	0.6881	0.3844	0.6539
ZSU234	0.5411	0.7174	0.4353	0.6637

TABLE VI
QUANTITATIVE RESULTS OF REAL IMAGES WITH CORRESPONDING IMAGES GENERATED BY ATGAN

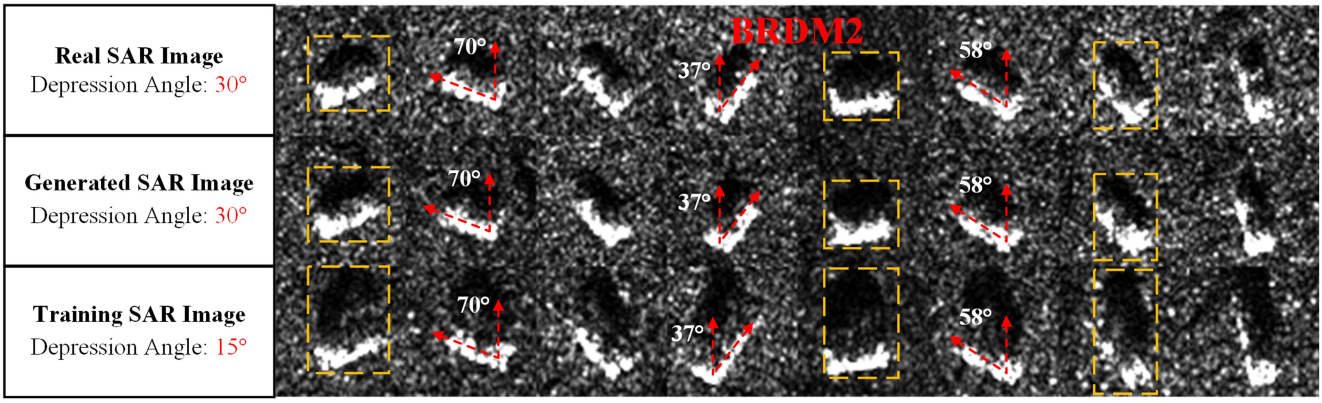
	MSSIM (↑)	FSIM (↑)
Proposed ATGAN	0.5086	0.6933

Table V also gives the average quantitative evaluation results between the generated and real SAR target images on the test set. These quantitative results further demonstrate that the ATGAN-generated SAR target images are close to the real ones. Based on the above results, the transferability of the AT learned by the proposed ATGAN can be verified.

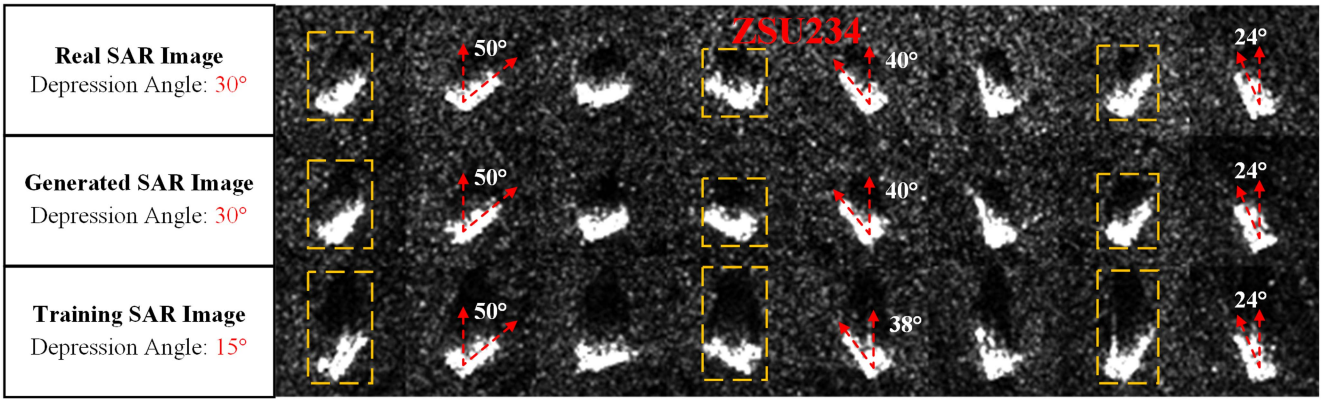
Moreover, we also conducted experiments to validate the generalization ability of ATGAN on azimuth angle generalization. Specifically, following the same setting in [56], the SAR target images with 17° depression angle are divided into two groups according to the azimuth angle, e.g., $[0, \pi)$ and $[\pi, 2\pi]$, in our experiments. As a result, only about 50% azimuth angle is used for training. Fig. 16 illustrates the comparison between the real (upper row of the panel) and generated SAR targets images (bottom row of the panel). It can be observed that the SAR target images generated by ATGAN look similar to the real images not only in terms of azimuth angle (as indicated by the real angle rulers) but also the shape and characteristics of targets (as indicated by the yellow rectangles). In other words, the proposed ATGAN has the ability to generate SAR target images when only about 50% of the azimuth angles are available in training set. The quantitative similarity between the generated SAR target images and the real images given in Table VI further verify the generalization ability of ATGAN on azimuth angle.



(a)



(b)



(c)

Fig. 15. Comparison of real SAR images (upper row) in test set, with the corresponding generated SAR images (middle row) via ATGAN, and corresponding real SAR images (bottom row) in training set. (a) 2S1. (b) BRDM2. (c) ZSU234.

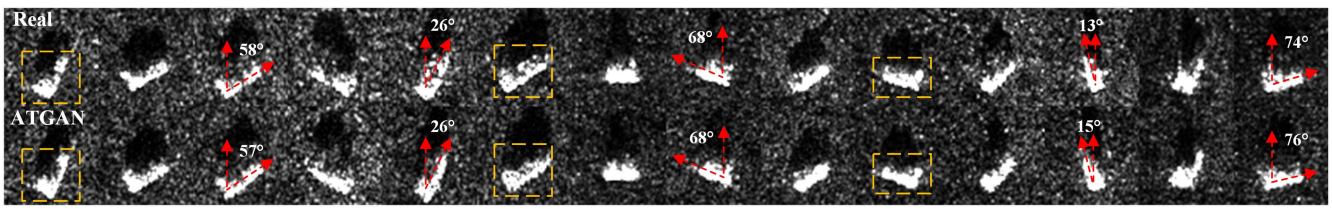


Fig. 16. Comparison of real SAR target images (upper row), with the corresponding generated images via ATGAN (bottom row).

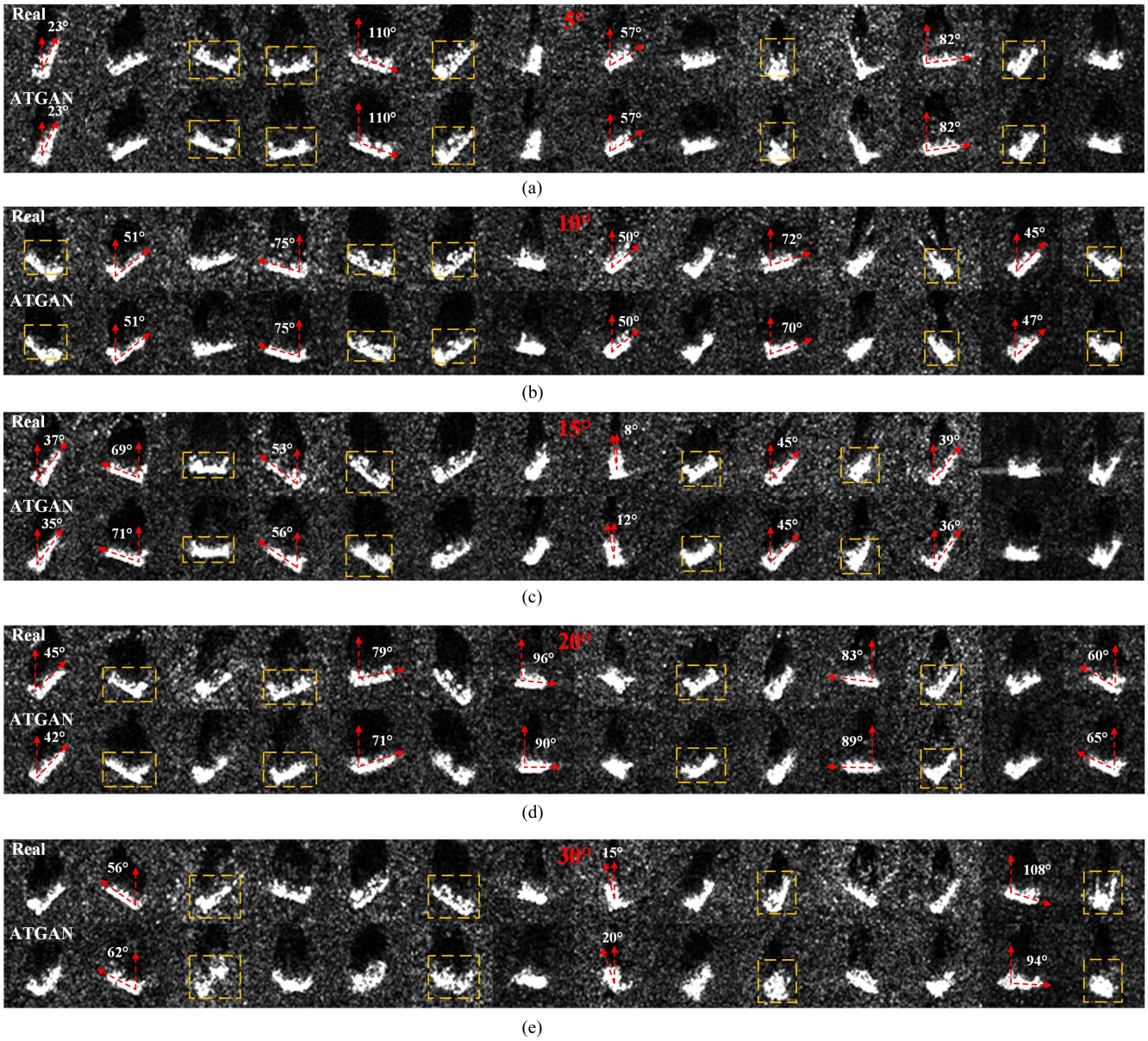


Fig. 17. Generated SAR target images and the real SAR target images with different training azimuth intervals. (a) With 5° azimuth interval. (b) With 10° azimuth interval. (c) With 15° azimuth interval. (d) With 20° azimuth interval. (e) With 30° azimuth interval.

C. Range of Angle Transformation

As mentioned earlier, the distribution of strong scattering points in SAR target images vary rapidly with the azimuth angle. Thus, we have conducted some experiments about the borderline of the training azimuth interval that can generate high-quality SAR target images. The training and testing combinations are listed in Table VII. Fig. 17 illustrates the comparison between generated SAR target images (upper row of each panel) and the real images (bottom row of each panel) with different training azimuth intervals. It can be found that, when the azimuth interval is increasing from 5° to 30°, the quality of the generated SAR target images is decreasing obviously, and when the azimuth interval is 20°, the generated SAR target images start to obviously suffer from degradation in quality and azimuth controllability.

TABLE VII
NUMBER OF TRAINING AND TESTING DATASET FOR THE EVALUATION OF THE RANGE

Azimuth Interval	Training Combination	Testing Combination
5°	287	618
10°	278	592
15°	258	584
20°	241	573
30°	211	527

Furthermore, we have generated SAR target images with the azimuth interval 30°. It can be observed that the generated SAR

TABLE VIII
QUANTITATIVE SIMILARITY RESULTS UNDER INCREASING AZIMUTH INTERVAL

Azimuth Interval	MSSIM (↑)	FSIM (↑)
5°	0.6186	0.7152
10°	0.5933	0.7091
15°	0.5715	0.6945
20°	0.5262	0.6511
30°	0.3958	0.5812

target images are not similar to the real images (as indicated by the red angle rulers and yellow rectangles). The quantitative similarity between the generated SAR target images and the real images given in Table VIII further validates the visual similarity comparison results. Thus, from the results of experiments, the azimuth interval borderline between the input and target SAR images should be between 20° and 30°.

VI. CONCLUSION

In this article, an ATGAN model is first proposed, which can generate azimuth-controllable SAR target images while preserving target details for ATR. Our idea is to learn the AT in the training data, and then, apply it to enrich an incomprehensive set of images in the target domain via image-to-image translation. Specifically, the proposed ATGAN consists of a coarse-to-fine generator and a multiscale SN-patch discriminator. The former aims to learn the AT in the deep feature space and subsequently utilize it to manipulate the representation of an input SAR target image to generate a new one, while the latter tries to estimate the probability that a SAR target image pair is real rather than fake using a patch-averaged strategy. With the aid of spatial transformer and adversarial training manner, the proposed ATGAN can provide precise SAR target images, which avoids the problem of existing GAN-based methods that directly map random noise to the desired SAR target images. Extensive experiments show the superiority of the proposed ATGAN in generating SAR target images, and our method outperforms the state-of-the-art method qualitatively and quantitatively. In SAR ATR test, our ATGAN achieves a score that is closer to real data, proving that it has reliable generation ability, and our ATGAN can be utilized as a new data augmentation approach to alleviate the data scarcity problem in SAR ATR and promote the recognition performance.

We hope that this method can provide new ideas for the generation of SAR target images with desired physical meaning. In the future, how to effectively couple the physical parameters as feedback in the deep generative model for SAR image generation (not limited to a single target) remains to be explored [57]. The physical parameters will be applied as the supervision of the output to ensure the physics consistency, e.g., generating SAR images from different viewing angles given only a single SAR image as input. There are some examples to learn from in the field of hyperspectral remote sensing image synthesis [58], [59]. Moreover, a comprehensive assessment criterion is also supposed to be developed to evaluate the quality of generated SAR images because the evaluation approach from the computer vision field is not completely suitable for SAR images.

APPENDIX
ARCHITECTURE OF GENERATOR IN ATGAN

Fine Generator G_f			
Module	**Output**	**Kernel**	**Stride**
STL	(1, 128, 128)	/	/
Conv + IN + ReLU	(1, 64, 64)	3	2
Coarse Network	(1, 64, 64)	/	/
STL	(1, 64, 64)	/	/
Conv + IN + ReLU	(64, 64, 64)	7	1
Residual Block × 6	(64, 64, 64)	3	1
Conv + IN + ReLU	(1, 128, 128)	3	2
Conv + Tanh	(1, 128, 128)	7	1

Coarse Generator G_c			
Module	**Output**	**Kernel**	**Stride**
STL	(1, 64, 64)	/	/
Conv + IN + ReLU	(64, 64, 64)	7	1
Conv + IN + ReLU	(128, 32, 32)	3	2
Conv + IN + ReLU	(256, 16, 16)	3	2
Residual Block × 6	(256, 16, 16)	3	1
Conv + IN + ReLU	(128, 32, 32)	3	2
Conv + IN + ReLU	(64, 64, 64)	3	2
Conv + Tanh	(1, 64, 64)	7	1

ARCHITECTURE OF DISCRIMINATOR IN ATGAN

SN-Patch Discriminator D_i			
Module	**Output**	**Kernel**	**Stride**
Conv + SN + LeakyReLU	(64, 64, 64)	4	2
Conv + SN + LeakyReLU	(128, 32, 32)	4	2
Conv + SN + LeakyReLU	(256, 16, 16)	4	2
Conv	(1, 15, 15)	4	1

ACKNOWLEDGMENT

The authors would like to thank Dr. Song for providing the source code for AAE.

REFERENCES

[1] Y. Zhang, D. Zhu, X. Mao, X. Yu, J. Zhang, and Y. Li, "Multirotors video synthetic aperture radar: System development and signal processing," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 35, no. 12, pp. 32–43, Dec. 2020.
[2] H. Sun, M. Shimada, and F. Xu, "Recent advances in synthetic aperture radar remote sensing—Systems, data processing, and applications," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2013–2016, Nov. 2017.
[3] X. Sun, B. Wang, Z. Wang, H. Li, H. Li, and K. Fu, "Research progress on few-shot learning for remote sensing image interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2387–2402, 2021.

[4] O. Kechagias-Stamatis and N. Aouf, "Automatic target recognition on synthetic aperture radar imagery: A survey," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 36, no. 3, pp. 56–81, Mar. 2021.

[5] U. Srinivas, V. Monga, and R. G. Raj, "SAR automatic target recognition using discriminative graphical models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 591–606, Jan. 2014.

[6] D. Malmgren-Hansen, A. Kusk, J. Dall, A. A. Nielsen, R. Engholm, and H. Skriver, "Improving SAR automatic target recognition models with transfer learning from simulated data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1484–1488, Sep. 2017.

[7] 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, 2021.

[8] S. Du, J. Hong, Y. Wang, and Y. Qi, "A high-quality multicategory SAR images generation method with multiconstraint GAN for ATR," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4011005.

[9] Q. Song, F. Xu, and Y.-Q. Jin, "SAR image representation learning with adversarial autoencoder networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 9498–9501.

[10] T. Zeng, C. Hu, H. Sun, and E. Chen, "A novel rapid SAR simulator based on equivalent scatterers for three-dimensional forest canopies," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5243–5255, Sep. 2014.

[11] F. Zhang, C. Hu, W. Li, W. Hu, and H.-C. Li, "Accelerating time-domain SAR raw data simulation for large areas using multi-GPUs," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 9, pp. 3956–3966, Sep. 2014.

[12] T. Balz, H. Hammer, and S. Auer, "Potentials and limitations of SAR image simulators–A comparative study of three simulation approaches," *ISPRS J. Photogrammetry Remote Sens.*, vol. 101, pp. 102–109, 2015.

[13] Z. Chen, Z. Zeng, Y. Huang, J. Wan, and X. Tan, "SAR raw data simulation for fluctuant terrain: A new shadow judgment method and simulation result evaluation framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5215018.

[14] G. Franceschetti, M. Migliaccio, D. Riccio, and G. Schirinzi, "SARAS: A synthetic aperture radar (SAR) raw signal simulator," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 1, pp. 110–123, Jan. 1992.

[15] 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.

[16] S. Auer, S. Hinz, and R. Bamler, "Ray-tracing simulation techniques for understanding high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1445–1456, Mar. 2010.

[17] 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.

[18] I. J. Goodfellow et al., "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, 2014, vol. 27, pp. 2672–2680.

[19] F. Liu, L. Jiao, and X. Tang, "Task-oriented GAN for PolSAR image classification and clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2707–2719, Sep. 2019.

[20] X. Song, Y. Chen, Z. H. Feng, G. Hu, D. J. Yu, and X. J. Wu, "SP-GAN: Self-growing and pruning generative adversarial networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2458–2469, Jun. 2021.

[21] C. Zhao, G. G. Yen, Q. Sun, C. Zhang, and Y. Tang, "Masked GAN for unsupervised depth and pose prediction with scale consistency," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5392–5403, Dec. 2021.

[22] K. Wang, G. Zhang, Y. Leng, and H. Leung, "Synthetic aperture radar image generation with deep generative models," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 912–916, Jun. 2019.

[23] 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.

[24] Z. Cui, M. Zhang, Z. Cao, and C. Cao, "Image data augmentation for SAR sensor via generative adversarial nets," *IEEE Access*, vol. 7, pp. 42255–42268, 2019.

[25] C. Mao, L. Huang, Y. Xiao, F. He, and Y. Liu, "Target recognition of SAR image based on CN-GAN and CNN in complex environment," *IEEE Access*, vol. 9, pp. 39608–39617, 2021.

[26] X. Hu, W. Feng, Y. Guo, and Q. Wang, "Feature learning for SAR target recognition with unknown classes by using CVAE-GAN," *Remote Sens.*, vol. 13, no. 18, p. 3554, 2021.

[27] 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, no. 19, p. 3939, 2021.

[28] Q. Song, F. Xu, X. X. Zhu, and Y.-Q. Jin, "Learning to generate SAR images with adversarial autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5210015.

[29] Q. Song and F. Xu, "Zero-shot learning of SAR target feature space with deep generative neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2245–2249, Dec. 2017.

[30] K. Lenc and A. Vedaldi, "Understanding image representations by measuring their equivariance and equivalence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 991–999.

[31] M. Mirza and S. Osindero, "Conditional generative adversarial nets," Accessed: Nov. 01, 2014. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2014arXiv1411.1784M

[32] U. Demir and G. Unal, "Patch-based image inpainting with generative adversarial networks," Mar. 2018, *arXiv:1803.07422*.

[33] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4217–4228, Dec. 2021.

[34] J. Pei, Y. Huang, W. Huo, J. Wu, J. Yang, and H. Yang, "SAR imagery feature extraction using 2DPCA-based two-dimensional neighborhood virtual points discriminant embedding," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2206–2214, Jun. 2016.

[35] X. X. Zhu et al., "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 143–172, Dec. 2021.

[36] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[37] D. Ulyanov, A. Vedaldi, and V. J. A. E.-P. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," Accessed: Jul. 01, 2016. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2016arXiv160708022U

[38] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[40] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, 2015, vol. 2, pp. 2017–2025.

[41] J. Pei, Y. Huang, W. Huo, Y. Zhang, J. Yang, and T.-S. Yeo, "SAR automatic target recognition based on multiview deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2196–2210, Apr. 2018.

[42] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5967–5976.

[43] T. Miyato, T. Kataoka, M. Koyama, and Y. J. A. E.-P. Yoshida, "Spectral normalization for generative adversarial networks," Accessed: Feb. 01, 2018. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2018arXiv180205957M

[44] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.

[45] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2536–2544.

[46] "The air force moving and stationary target recognition database," 2016. [Online]. Available: https://www.sdms.afrl.af.mil/-datasets/mstar/

[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Accessed: Dec. 01, 2014. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2014arXiv1412.6980K

[48] Q. Guo et al., "Recognition rate versus substitution rate curve: An objective utility assessment criterion of simulated training data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5224415.

[49] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[50] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE 37th Asilomar Conf. Signals, Syst. Comput.*, 2003, pp. 1398–1402.

[51] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[52] P. Kovesi., "Image features from phase congruency," *Videre: J. Comput. Vis. Res.*, vol. 1, pp. 1–26, 1999.

[53] G. Baier, A. Deschemps, M. Schmitt, and N. Yokoya, "Synthesizing optical and sar imagery from land cover maps and auxiliary raster data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4701312.
[54] Y. Kwak, W.-J. Song, and S.-E. Kim, "Speckle-noise-invariant convolutional neural network for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 549–553, Apr. 2019.
[55] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806–4817, Aug. 2016.
[56] Z. Wen, J. Liu, Z. Liu, S. Li, and Q. Pan, "Contrastive feature disentangling for partial aspect angles SAR noncooperative target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5103315.
[57] M. Datcu, Z. Huang, A. Anghel, J. Zhao, and R. Cacoveanu, "Explainable, physics-aware, trustworthy artificial intelligence: A paradigm shift for synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 11, no. 1, pp. 8–25, Mar. 2023.
[58] L. Liu, W. Li, Z. Shi, and Z. Zou, "Physics-informed hyperspectral remote sensing image synthesis with deep conditional generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5528215.
[59] L. Liu, B. Chen, H. Chen, Z. Zou, and Z. Shi, "Diverse hyperspectral remote sensing image synthesis with diffusion models," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5532616.

Zhiqiang Zeng received the B.Eng. degree in electronic information engineering from the Wuhan Institute of Technology, Wuhan, China, in 2019, and the M.Eng. degree in electronic and communication engineering in 2022 from Chongqing University, Chongqing, China, where he is currently working toward the Ph.D. degree in information and communication engineering with the School of Microelectronics and Communication Engineering.

His current research interests include synthetic aperture radar (SAR) raw data simulation and SAR image processing.

Xiaoheng Tan received the B.Eng. and Ph.D. degrees in electrical engineering from Chongqing University, Chongqing, China, in 1998 and 2003, respectively.

From 2008 to 2009, he was a Visiting Scholar with the University of Queensland, Brisbane, Australia. He is currently a Professor with the School of Microelectronics and Communication Engineering, Chongqing University. His current research interests include modern communications technologies and systems, radar signal processing, and machine learning.

Xin Zhang was born in 2000. He received the B.Eng. degree in communication engineering in 2022 from Chongqing University, Chongqing, China, where he is currently working toward the M.S. degree in information and communication engineering.

His current research interests include synthetic aperture radar interference suppression and image processing.

Yan Huang (Member, IEEE) received the B.S. degree in electrical engineering and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2013 and 2018, respectively.

From 2016 to 2017, he was a Visiting Ph.D. Student with Electrical and Computer Engineering Department, University of Florida, Gainesville, FL, USA, and with Electrical and Systems Engineering Department, Washington University in St. Louis, St. Louis, MO, USA, from 2017 to 2018. He is currently an Associate Professor with the State Key Laboratory of Millimeter Waves, Southeast University, Nanjing, China. His research interests include machine learning, synthetic aperture radar, image processing, and remote sensing.

Jun Wan (Member, IEEE) received the B.Eng. degree in electrical engineering from Fuzhou University, Fuzhou, China, and the Ph.D. degree in signal processing from Xidian University, Xi'an, China, in 2015 and 2020, respectively.

In 2020, he joined the School of Microelectronics and Communication Engineering, Chongqing University, Chongqing, China, where he is currently a Lecturer. His current research interests include ground moving target indication and imaging, moving/maneuvering target detection, and time–frequency analysis.

Zhanye Chen (Member, IEEE) was born in Hubei, China. He received the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2019.

He is currently an Associate Professor with the State Key Laboratory of Millimeter Waves, Southeast University, Nanjing, China. His research interests include moving target indication, radar imaging, and radar simulation.