Improved Few-Shot SAR Image Generation by Enhancing Diversity

Jianghan Bao, *Graduate Student Member, IEEE*, Wen Ming Yu, Kaiqiao Yang, Che Liu, *Member, IEEE*, and Tie Jun Cui, *Fellow, IEEE*

***Abstract*—Due to their remarkable capabilities of generation, deep-learning-based (DL) generative models have been widely applied in the field of synthetic aperture radar (SAR) image synthesis. This kind of data-driven DL methods usually requires abundant training samples to guarantee the performance. However, the number of SAR images for training is often insufficient because of expensive acquisitions. This typical few-shot image generation (FSIG) task still remains not fully investigated. In this article, we propose an optical-to-SAR (O2S) image translation model with a pairwise distance (PD) loss to enhance the diversity of generation. First, we replace the semantic maps used as the input of network in previous studies with more easily available optical images and apply the popular pix2pix model in image-to-image translation tasks as the foundation network. Second, inspired by the FSIG works in the traditional computer vision field, we propose a similarity preservation term in the loss function, which encourages the generated images to inherit the similarity relationship of the corresponding simulated SAR images. Third, the data augmentation experiments on the MSTAR dataset indicates the effectiveness of our model. With only five samples for each target category and six categories in total, the basic O2S network boosts the classification accuracy by 4.81% and 2.27% for data of depression angle of 15° and 17°, respectively. The PD loss is capable of bringing additional 2.23% and 1.78% improvement. The investigation on similarity curves also suggests that the generated images enhanced by the PD loss have closer similarity behaviors to the real SAR images.**

***Index Terms*—Deep learning (DL), few-shot image generation (FSIG), image-to-image translation (I2I), synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) has become an essential tool for land survey, Earth remote sensing, environment monitoring, etc., since it is capable of providing day-and-night, weather-independent, and high-resolution images [1]. As the SAR signal processing and imaging techniques advance, the classification and recognition of targets of interest in SAR images have been paid more and more attention, which leads to the task of automatic target recognition (ATR) [2], [3], [4]. With the rapid development of deep learning (DL) methods recent years [5], many works have been done to apply DL approaches to ATR task [6], [7], [8]. However, training a well-performing DL model needs a large amount of data, which is often limited for SAR ATR task since the acquisition of SAR image is hard in practice. The lack of training samples in DL is usually named as few-shot learning (FSL) problem [9].

Data augmentation (DA) [10] is a common category of methods to tackle the problem of insufficient training images for the DL model. One way of DA is to introduce more data by basic image manipulation of initial samples or utilizing other sources of data. For the SAR ATR task, images given by simulation are often applied to augment the dataset [11], [12], [13]. Another widely used category of DA methods is to generate more data by DL. Among these models, generative adversarial network (GAN) [14] and its various variants have achieved remarkable performance [15], [16], [17], [18], [19]. This kind of models consist of two modules, the generator that attempts to output images as realistic as possible and the discriminator, which tries to distinguish the generated images from the real ones. Because of its impressive capability of generation, we choose the GAN as the foundation of our approach.

The lack of training samples also results in the unrealistic outputs of DL generation models. The generation task under this condition is usually referred as few-shot image generation (FSIG). There exist already several GAN-based works focusing on this problem in the traditional computer vision field [20], [21], [22], [23], [24], [25], but the FSIG of SAR still remains not fully explored. Some representative recent works have been conducted [18], [19]. Inspired by their models, we intend to make further progress to confront more challenging task with fewer samples per target class and larger number of category.

The model proposed in [18] uses semantic maps of SAR images as input of the generator, which are difficult to acquire in practice. Thus, we suggest to replace the semantic maps by the much more easily accessible optical images. This makes it possible to introduce various methods of image-to-image translation (I2I) [26], which focuses on converting images from one domain to another. In our case, we intend to transfer an optical image to

Manuscript received 28 August 2023; revised 4 November 2023 and 15 December 2023; accepted 6 January 2024. Date of publication 10 January 2024; date of current version 23 January 2024. This work was supported in part by the China National Postdoctoral Program for Innovative Talents under Grant BX20220065, in part by the Fundamental Research Funds for the Central Universities under Grant 2242023K5002, in part by the National Natural Science Foundation of China under Grant 61890540, Grant 61890544, and Grant 62301146, and in part by the National Key Research and Development Program of China under Grant 2018YFA0701900 and Grant 2018YFA0701904. *(Jianghan Bao and Wen Ming Yu contributed equally to this work.) (Corresponding authors: Che Liu; Tie Jun Cui.)*

Jianghan Bao, Wen Ming Yu, Kaiqiao Yang, Che Liu, and Tie Jun Cui are with the Institute of Electromagnetic Space, Southeast University, Nanjing 211189, China, and also with the State Key Laboratory of Millimeter Wave, Southeast University, Nanjing 211189, China (e-mail: baojianghan@seu.edu.cn; wmyu@seu.edu.cn; 220220689@seu.edu.cn; cheliu@seu.edu.cn; tjcui@seu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3352237

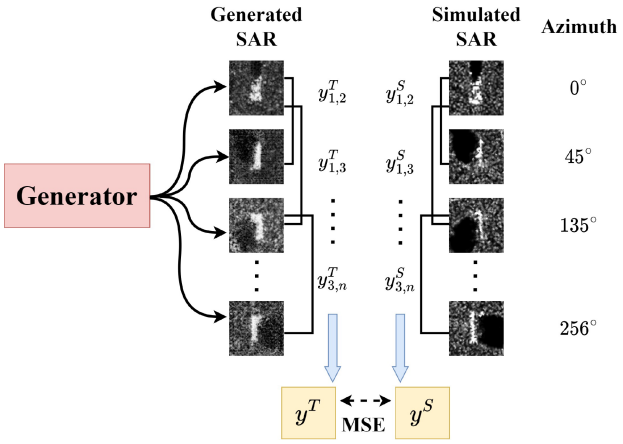


Fig. 1. Simple illustration of the PD loss used in this article. y^T represents the similarity matrix of a batch of generated SAR images, while y^S is the similarity matrix of corresponding simulated SAR images.

the corresponding SAR image while maintaining its content. We apply the popular I2I model pix2pix [27] as the basic network. Following [18] and [19], physics information like target category and angle characterizing the SAR imaging are also utilized as input. To ease the learning of rotation representation for the DL model, we applied the rotated cropping proposed in [18].

In this article, the module tries to mitigate the influence of insufficient training samples is proposed after we revisit previous researches on FSIG [20], [21], [22], [23], [24], [25]. A major problem caused by the small number of training samples is lack of diversity of generated images. Since the quantity of real images used to train distinguish ability of the discriminator is quite limited, the discriminator can simply memorize the few real samples and force the generator to replicate them. One possible way to ameliorate the networks is proposed in [25], which attempts to transfer the diversity information from the large source domain to the small target domain. In their method, the generator will first be well trained with the plenty of samples in the source dataset and their parameters will be recorded as G_S. During the finetuning of the pretrained model on the small target domain, the generator will be forced to preserve the relative similarities and differences of features of G_S for the same input. For the real SAR dataset with hard accessibility, there exists a suitable large dataset that could be used as source domain, the simulated SAR images. Based on their similarity perseverance loss [25], we propose our light-weighted pairwise distance (PD) loss by utilizing simulated data. Instead of calculating the differences of features of all the layers in generator, we compute only the similarity of generated images (i.e., the output of last layer) and regularize it with the similarity matrix obtained by the corresponding simulated SAR images. Thus, it is possible for the model to inherit the diversity in the source domain. A simple illustration of the proposed PD loss is given in Fig. 1.

Our experiments are performed on the moving and stationary target acquisition and recognition (MSTAR) public dataset [28], [29]. Compared with the former work considering orientation angle interval of 25° [18], we concentrate on a more difficult task with only five samples for each equipment class, which results in an angle interval of 72°. Meanwhile, the number of categories under our consideration is twice that of [19], making it a six-way five-shot task. The optical images utilized as the input of network are rendered by Open Graphics Library (OpenGL) [30] and the simulated SAR images are obtained by shooting and bouncing rays (SBR) method based on geometric models. We conduct the experiments on data of both 15° and 17° depression angles in MSTAR. The generated images are used to augment the training of an all-convolution classification network [31]. The classification accuracy of the augmented network on the test set is used as the criterion of the performance of our model. It is shown that the accuracy is improved by 4.81% and 2.27% by the results of basic optical-to-SAR (O2S) translation network for 15° and 17° depression angle, respectively. And the introduction of PD loss will bring another 2.23% and 1.78% elevation of accuracy, which indicates the effectiveness of our approach. The curves of similarity of generated images with respect to orientation angles also show that their similarity behavior can be adjusted closer to that of real SAR by the PD loss. Our contributions can be summarized as the following three aspects.

1) We introduce the O2S translation framework to the FSIG task of SAR. The semantic maps with limited availability are replaced by the more easily accessible rendered optical images and can be substituted with real photos in further works.
2) A loss function term based on mutual similarity is proposed in the FSIG task of SAR to enhance the diversity of generated images. This also gives an alternative way of using the simulated SAR images appropriately. The domain gap between the simulated and real SAR has limited the direct use of data given by simulation [13].
3) A more challenging six-way five-shot task compared with previous works is taken into consideration. The basic model of the O2S translation framework has proven its effectiveness by experiments on MSTAR, and the PD loss is able to bring additional improvement.

The rest of this article is organized as follows. Section II reviews important related works of GAN, FSIG, and I2I translation. In Section III, we give detailed descriptions of basic O2S translation network and how to calculate the PD loss. Section IV states how we construct the rendered optical image dataset and simulated SAR image dataset at first, and then gives the augmentation effect, the similarity curves of generated images, and a discussion of characteristics of generation with different settings. We also compare the performance of our approach with current state-of-the-art methods in Section IV. An analysis of ablation study is provided to clarify the effect of the components in our model in Section V. Finally, Section VI concludes this article.

II. Related Works

A. Generative Adversarial Network (GAN)

Due to their capacity of generating realistic and high-quality images, GANs [14] have gained significant attention. GAN models usually consist of two modules, the generator and the discriminator. The former tries to generate synthetic samples that

resemble real data and the later distinguishes real and generated samples. The two parts are trained simultaneously, and they continuously learn from each other. As training progresses, the generator becomes better at generating realistic samples, while the discriminator distinguishing real from fake data. The iterative process eventually leads to the generation of high-fidelity synthetic data.

However, training GANs can be challenging and unstable. A common issue is mode collapse, which refers to the limited variations of samples produced by the generator. A widely used way to mitigate this problem is to replace the Kullback–Leibler (KL) or Jensen–Shannon (JS) divergence in the loss function with the Wasserstein distance, which leads to the popular Wasserstein GAN (WGAN) [32]. Based on this work, an improved strategy to enforcing Lipschitz constraint on the discriminator by gradient penalty (WGAN-GP) is proposed [33], which enables a more stable training process compared with classical GAN models.

In SAR generation task, GAN-based methods have also been widely applied. The conventional GAN and WGAN-GP are utilized in [15] and [16], respectively. To achieve results with higher quality, different kinds of auxiliary information are taken as input of the conditional GAN-based (CGAN) [34] neural networks. For example, Cao et al. [17] make use of label information and Sun et al. [19] need category indexes and aspect angles. In addition, semantic maps of SAR images are recently introduced in [18] to guide the generation.

B. Few-Shot Image Generation (FSIG)

Few-shot learning (FSL) is a subarea of machine learning, which has access to only a few samples with supervised information during the training process [9]. The FSIG tasks mainly concentrate on generating new images under this circumstance. In contrast with the regular image generation tasks with plenty of training samples, the results of FSIG with GAN-based model suffer severely from the lack of diversity. Since the discriminator can simply memorize the extremely limited training samples (e.g., ten images per class) rather than learning the true characteristics of the provided images, the discriminator loses largely the capacity of guiding the generator network [25].

Some works have already been done to tackle this problem. These methods can be broadly divided into two categories: the finetuning-based methods, which try to transfer a source model trained on a large related dataset by finetuning, and the regularization-based methods, which regularize the optimization of model parameters based on the prior knowledge of the dataset. To reduce the overfitting in the few-shot scene, the finetuning-based strategies often attempt to decrease the number of weights to update. Noguchi and Harada [20] optimize only the scale and shift parameters of batch statistics in the generator. Zhao et al. [21] reuse low-level filters of the pretrained model and replace the high-level layers with a smaller network. In [22], the importance of kernels in the pretrained model is measured by Fisher information [23], and the knowledge of important kernels will be preserved while the unimportant ones will be updated. As for the regularization-based methods, additional term is commonly appended to the loss function. In [24], elastic weight consolidation loss term is used to limit the change of important weights when transferring the model. And Ojha et al. [25] introduce a pairwise similarity preservation loss to confront the overfitting.

C. Image-to-Image Translation (I2I)

The I2I refers to the task of converting an image from one domain to another while preserving its content representations [26]. Neural style transfer [35] is a typical example of I2I, which aims to render a content image in different styles. Some of the extensively used I2I models are based on GAN, such as pix2pix [27] and CycleGAN [36]. I2I is also a topic drawing increasing attention in SAR. Recent works include translation between simulated and real SAR images [13] and SAR images with different resolution [37]. The most investigated topic is the translation between optical and SAR images [38], [39], [40], [41], [42]. However, most of these models focus on the remote sensing optical images and airborne or spaceborne SAR, which provide a large amount of training samples. Thus, their methods could not be applied directly to the few-shot SAR image generation. Appropriate modification is needed to introduce I2I translation methods to FSIG task.

III. Proposed Method

In this section, we will give brief explanation of our motivation, detailed descriptions of the proposed model, and the additional loss term to enhance the diversity of generated images.

A. Basic Network

Compared with vanilla GAN, more prior information can be utilized in CGAN [34]. The pix2pix [27] model is a CGAN with an image input as condition. In SAR generation task, Song et al. [18] make use of the semantic maps to get more realistic SAR images. However, their semantic maps are acquired by manual annotation of the bright and dark area of the images, making the method not very convenient in practical scene. To improve the feasibility, we propose to use the more easily obtainable optical image as input of the neural network, which involves an O2S image translation problem. Access to the optical images will be presented thoroughly in Section IV-A. Due to the lack of training samples in FSIG, we choose the popular pix2pix model [27] in I2I as basis, rather than cycleGAN [36], which contains much more parameters by training two translation networks at the same time.

The architecture of the proposed method is given in Fig. 2. As a GAN-based model, it consists of two parts, a generator G and a discriminator D. With the adversarial training progressing, the capacity of distinguishing of the discriminator is strengthened and it will guide the generator to output images with more resemblance to the real ones. In the end, generated results will become indistinguishable for the discriminator. Details of the two subnetwork are given in the following.

As the pix2pix model, the generator G in our method adopts an U-Net structure [43]. In the conventional encoder-decoder networks, the input is passed through a sequence of layers progressively. However, U-Net adds skip connections between each layer i and its mirrored layer $n - i$, where n denotes the

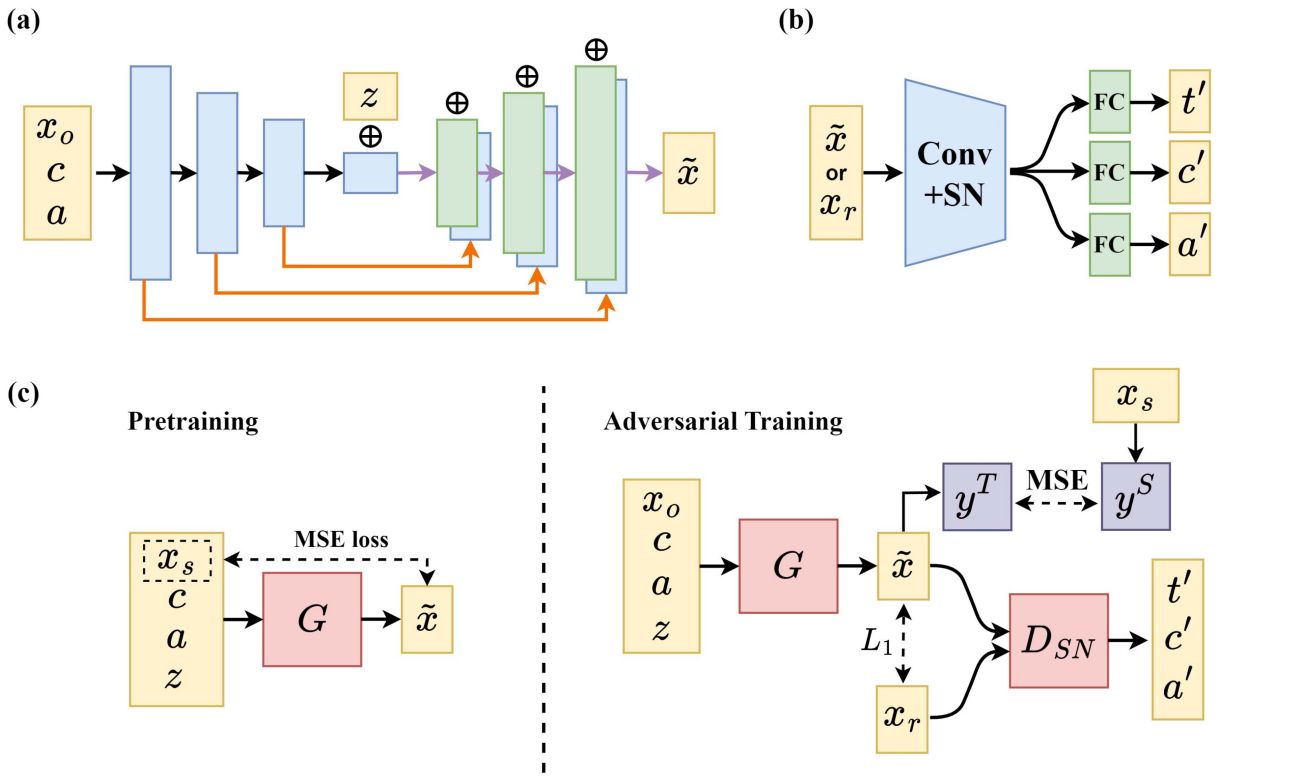


Fig. 2. Architecture of the proposed model. (a) Illustration of the U-Net generator. $\oplus$ stands for the concatenation on the channel. The black arrows represent the downsampling blocks, which consist of LeakyReLU, convolutional layer, and instance normalization. The purple arrows are the upsampling blocks containing ReLU, transposed convolutional layer, and instance normalization. The orange arrows denote the skip connections. x_o, c, a, and z are the input optical image, category label, orientation angle, and noise vector, respectively. $\tilde{x}$ is the generated image. (b) Structure of the discriminator. Conv+SN means convolutional layers with SN. FC represents the fully connected layer. t', c', and a' are the output of the discriminator on the real/fake prediction, classification, and angle prediction tasks, respectively. (c) Training process of the proposed model. It contains two parts, pretraining and adversarial training. During pretraining, the simulated SAR images x_s are used to update the generator G. In adversarial training, besides the GAN loss, the L_1 reconstruction loss and PD loss given by the similarity matrices y^T and y^S are also applied to optimize both the generator G and spectral normalized discriminator D_{SN}.

total number of layers. The connections give it an U-shaped architecture, which forms its name. They also allow the network to preserve both low-level and high-level features, making it more capable than autoencoders without skips. The network blocks in downsampling part of U-Net are composed of LeakyReLU [44], convolutional layers, and instance normalization [45]. In the upsampling part, we use ReLU as activation and transposed convolutional layers. The optical images x_o will be first converted to one-channel grayscale ones, i.e., $x_o \in \mathcal{R}^{1\times M\times M}$. M indicates the size of input images. Besides optical images, category labels c and azimuth angles ϕ are also utilized following the former generation works [18], [19] in our work. The label indexes c are embedded into vectors $\boldsymbol{c} \in \mathcal{R}^k$, where k is the embedding dimension. We expand the vectors $\boldsymbol{c}$ into the same width and height as x_o with k as the channel number, which results in $\boldsymbol{c} \in \mathcal{R}^{k\times M\times M}$. The orientation angle ϕ is represented by $a = (\cos\phi, \sin\phi)$, and then, expanded like $\boldsymbol{c}$, leading to $\boldsymbol{a} \in \mathcal{R}^{2\times M\times M}$. x_o, $\boldsymbol{c}$, and $\boldsymbol{a}$ will be concatenated on channel dimension before input into U-Net generator. With proper choice of layer number n in U-Net and size of x_o, the innermost layer of U-Net will give a 1×1 feature map. The noise vector $z \in \mathcal{R}^m$ drawn from a standard normal distribution will be concatenated with the feature, and then, enter the expansive part of U-Net. The reason why not expanding and concatenating z as what we have done to $\boldsymbol{c}$ and $\boldsymbol{a}$ is that the relatively large value of m may introduce too numerous channels in input. For enhancing the generation results, an L_1 reconstruction loss is also added.

The first part of the discriminator D consists of convolutional neural networks (CNNs). Following [18] and [19], the output feature will be presented to three different fully connected layers, which correspond to three tasks: identification of real and generated images, classification, and orientation angle prediction. To stabilize the training process of the discriminator, a regularization method called spectral normalization (SN) [46] is introduced to all the convolutional layers. The main idea is to enforce Lipschitz continuity of the discriminator's weights by normalizing its spectral norm. Lipschitz continuity ensures that small changes in the input space result in small changes in the output space, which helps improve the overall stability and convergence of GAN training. The weight matrix of a layer is denoted by W, and the corresponding normalized weight W_{SN} is as follows:

$$W_{\text{SN}} = \frac{W}{\sigma(W)} \tag{1}$$

where $\sigma(W)$ indicates the spectral norm obtained by power iteration [47], [48], instead of computation-consuming singular

value decomposition. It is also suggested in [46] that, applying SN and gradient penalty (GP) [33] simultaneously would achieve more improvement since these two methods regularize the discriminator in different ways. So, GP is present during training. In the rest text, we denote the discriminator with SN by D_{SN}. The subscripts t, c, and a are used to represent the real/fake prediction, classification, and azimuth angle prediction given by D_{SN}. To sum up, the loss function of the generator G is

$$L_G = L_{\text{GAN},G} + \lambda_c L_{c,G} + \lambda_a L_{a,G} + \lambda_1 L_{\text{rec}} \tag{2}$$

$$L_{\text{GAN},G} = -\mathbb{E}_{\tilde{x}\sim P_g}[D_{\text{SN},t}(\tilde{x})] \tag{3}$$

$$L_{c,G} = \text{CEL}[D_{\text{SN},c}(\tilde{x}),\ c] \tag{4}$$

$$L_{a,G} = \text{MSE}[D_{\text{SN},a}(\tilde{x}),\ a] \tag{5}$$

$$L_{\text{rec}} = L_1(\tilde{x},\ x_r) \tag{6}$$

where $\tilde{x}$ indicates a generated image, P_g is the generator distribution, and x_r denotes the corresponding real image. $\mathbb{E}_{x\sim P(x)}(f(x))$ represents expectation of $f(x)$ over distribution $P(x)$. λ_c, λ_a, and λ_1 are the coefficients of classification, angle prediction, and reconstruction tasks, respectively. CEL represents the cross entropy loss and MSE stands for mean square error. c and a signify the real category label and angle.

The loss function of D_{SN} is L_D and consists of three parts that relate to the three aforementioned tasks. The detailed expressions are

$$L_D = L_{\text{GAN},D} + \lambda_c L_{c,D} + \lambda_a L_{a,D} \tag{7}$$

$$\begin{aligned} L_{\text{GAN},D} = {} & \mathbb{E}_{\tilde{x}\sim P_g}[D_{\text{SN},t}(\tilde{x})] - \mathbb{E}_{x_r\sim P_r}[D_{\text{SN},t}(x_r)] \\ & + \lambda_{\text{gp}}\mathbb{E}_{\hat{x}\sim P_{\hat{x}}}[(||\nabla_{\hat{x}} D_{\text{SN},t}(\hat{x})||_2 - 1)^2] \end{aligned} \tag{8}$$

$$L_{c,D} = \text{CEL}[D_{\text{SN},c}(x),\ c],\ x \in \{\tilde{x}, x_r\} \tag{9}$$

$$L_{a,D} = \text{MSE}[D_{\text{SN},a}(x),\ a],\ x \in \{\tilde{x}, x_r\} \tag{10}$$

where P_r expresses the real data distribution, λ_{gp} denotes the coefficient of GP, and $\hat{x}$ is given by $\epsilon x_r + (1-\epsilon)\tilde{x}$ with ϵ a random number drawn from uniform distribution $U[0,1]$. x stands for both $\tilde{x}$ and x_r since the discriminator is expected to be able of recognize the category and orientation for both real and generated SAR images.

The aforementioned structures form the basic network of our method. The strategy especially designed for FSIG is presented in the next subsection.

B. Increasing Diversity

As it is stated in [25], when the number of training samples becomes extremely few, the discriminator would simply remember the few examples and force the generator to output images with high similarity to them, which leads to low diversity of the results. A commonly used means is to introduce a related source dataset with plenty of images, and then, attempt to transfer the knowledge from the source dataset to few-shot target dataset. Several works of this category habe been listed in Section II-B, and the cross-domain distance consistency (CDDC) proposed in [25] inspires us the most. Given a large source dataset $\mathcal{D}_S$ and corresponding well-trained generator G_S, we aim to reach an adapted generator G_T initialized with G_S on small target dataset $\mathcal{D}_T$. Besides basic adversarial learning loss, CDDC proposes to enforce preservation of relative distance before and after adaptation. In their method, a batch of $N+1$ noise vectors is input to both G_S and G_T, then the similarity of features between the networks is used to construct an N-way probability distribution. For the ith noise vector, the probability distribution of G_S and G_T is as follows:

$$y_i^{S,l} = \text{Softmax}(\{\text{sim}(G_S^l(z_i), G_S^l(z_j))\}_{\forall i\neq j}) \tag{11}$$

$$y_i^{T,l} = \text{Softmax}(\{\text{sim}(G_T^l(z_i), G_T^l(z_j))\}_{\forall i\neq j}) \tag{12}$$

where l denotes the lth layer of the generator. Softmax is an activation function given by

$$\text{Softmax}(\boldsymbol{x})_i = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \tag{13}$$

where x_i is the ith element of input vector $\boldsymbol{x}$. sim is the cosine similarity that is defined by

$$\text{sim}(\boldsymbol{A}, \boldsymbol{B}) = \frac{\boldsymbol{A}\cdot\boldsymbol{B}}{||\boldsymbol{A}||||\boldsymbol{B}||} = \frac{\sum_i A_i B_i}{\sqrt{\sum_i A_i^2}\sqrt{\sum_i B_i^2}} \tag{14}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ denote the features under consideration. A_i and B_i are the ith components of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively.

The distributions of feature similarity of two generators are encouraged to be close, thus the distance loss is given by

$$L_{\text{dist}}(G_S, G_T) = \mathbb{E}_{z_i\sim p_z(z)} \sum_{l,i} D_{KL}\left(y_i^{T,l} || y_i^{S,l}\right) \tag{15}$$

where D_{KL} indicates the KL-divergence.

Inspired by their work, we propose our own light-weight version PD loss with the simulated SAR images as source dataset. Details of electromagnetic simulation are given in Section IV-A. There exist clear domain gap between the real and simulated SAR images [13], for which the simulated data are hard be directly used to train the network. However, the similarity between simulated images of different orientation and categories probably shares comparable behavior with the real ones.

Taking ZSU234 with depression angle 15° as an example, we show the variations of cosine similarity of different angles with respect to 0° for the optical, simulated, and real SAR images in Fig. 3. Although the variation of simulated SAR is clearly different from that of real SAR regarding the value of similarity, the general tendency remains alike. The similarity drops quite fast for both real and simulated SAR images when the orientation angle begins to change. And for both of them, the most different image appears at about 180°. However, the curve of optical images increases drastically around 180°, which may be caused by the alignment of the bright target area and shadow area around both 0° and 180°. Moreover, the texture of rendering image changes less when the orientation is different. So, the similarity of optical images does not drop as fast as the real and simulated SAR images and its curve is much smoother. For the other five target categories, the characteristics of similarity variations are similar to this situation. Due to the analogous behavior of similarity for real and simulated SAR, we decide to utilize the simulation set to regularize the diversity of generation.

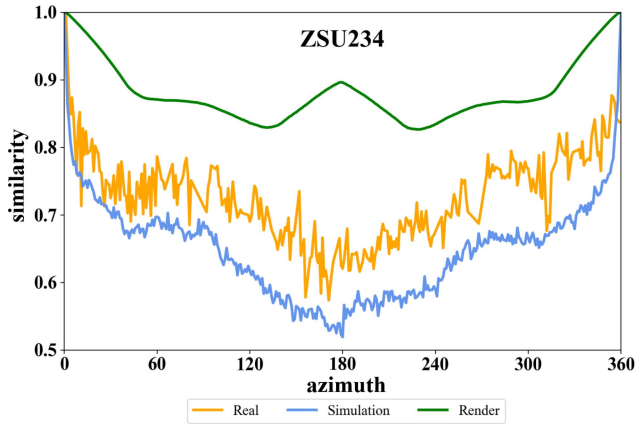


Fig. 3. Cosine similarity variations of ZSU234 of different kinds of images. The orange line stands for the real SAR images, the blue line represents the simulated SAR images, and the green line indicates the optical images given by rendering. The horizontal and vertical axes are the azimuth angle and similarity, respectively.

In the original CDDC, features of every layers in G_S are used to regularize G_T. Since the output images are the final purpose and also for reduction of computation, we propose to constrain only the output of G_T by simulated data and not bother to regularize all the layers of G_T. Given a batch of input $\{z_i, c_i, a_i, x_{o,i}\}_{0\le i\le N}$, we use $G_T(i)$ and $G_T(j)$ to denote the generated images of ith and jth input. The ij-element of the similarity matrix is then given by

$$y^T_{i,j} = \mathrm{sim}(G_T(i), G_T(j)) \; \forall i \neq j \tag{16}$$

where sim refers to the cosine similarity function. With $x_{s,i}$ and $x_{s,j}$ representing the simulated SAR corresponding to the input $x_{o,i}$ and $x_{o,j}$, element of the similarity matrix of simulated images is

$$y^S_{i,j} = \mathrm{sim}(x_{s,i}, x_{s,j}) \; \forall i \neq j. \tag{17}$$

The objective of our strategy is to reduce the difference of two similarity matrices

$$L_{\text{dist}} = \lambda_{\text{dist}} \mathrm{MSE}(y^T, y^S). \tag{18}$$

It should be emphasized that L_{dist} is applied only to the orientations without real SAR training samples. And λ_{dist} is the hyperparameter to adjust the importance of this part in loss function. To summarize, the detailed process of the proposed algorithm is presented in the Algorithm 1.

Algorithm 1: Few-Shot SAR Generation With PD Loss.

Require: Optical image dataset $\mathcal{D}_O$, source dataset given by simulation $\mathcal{D}_S$, small real SAR dataset $\mathcal{D}_T$, normalized discriminator D_{SN} with parameters θ_d, generator G with parameters θ_g, pretraining epoch n_p, total adversarial training steps n_s, discriminator training steps n_d, generator training steps n_g, learning rate l.

for n_p epochs **do**
  sample batches of simulated images $x_s \in \mathcal{D}_S$, corresponding optical images $x_o \in \mathcal{D}_O$, category labels c, angles a
  $\tilde{x} \leftarrow G(x_o, c, a)$,
  $L_p \leftarrow \mathrm{MSE}(\tilde{x}, x_s)$
  $\theta_g \leftarrow \mathrm{RMSprop}(L_p)$
end for
for n_s steps **do**
  for n_d steps **do**
    sample a batch of real images $x_r \in \mathcal{D}_T$, corresponding optical images $x_o \in \mathcal{D}_O$, category labels c, angles a, $\epsilon \sim U[0,1]$
    $\tilde{x} \leftarrow G(x_o, c, a)$, $\hat{x} \leftarrow \epsilon x_r + (1-\epsilon)\tilde{x}$
    $L_D \leftarrow L_{\text{GAN},D} + \lambda_c L_{c,D} + \lambda_a L_{a,D}$
    $\theta_d \leftarrow \mathrm{RMSprop}(L_D)$
  end for
  for n_g steps **do**
    sample a batch of real images $x_r \in \mathcal{D}_T$, corresponding optical images $x_o \in \mathcal{D}_O$, category labels c, angles a
    $\tilde{x} \leftarrow G(x_o, c, a)$
    $L_D \leftarrow L_{\text{GAN},G} + \lambda_c L_{c,G} + \lambda_a L_{a,G} + \lambda_1 L_{rec}$
    $\theta_g \leftarrow \mathrm{RMSprop}(L_G)$
  end for
  sample a batch of simulated images x_s with angle not in $\mathcal{D}_T$, corresponding optical images $x_o \in \mathcal{D}_O$, category labels c, angles a
  $\tilde{x} \leftarrow G(x_o, c, a)$
  $L_{\text{dist}}(G_S, G_T) \leftarrow \lambda_{\text{dist}} \mathrm{MSE}(y^T, y^S)$
  $\theta_g \leftarrow \mathrm{RMSprop}(L_{\text{dist}})$
end for

IV. Experiments

A. Datasets

The widely used moving and stationary target acquisition and recognition (MSTAR) dataset [28], [29] is adapted to our model. This dataset contains SAR images of several targets obtained by the X-band sensor with 0.3 m $\times$ 0.3 m resolution. Six types of targets at 15° and 17° depression and all azimuth angles are used in our experiments. The categories under consideration and corresponding numbers of images are given in Table I. Following [18], rotated cropping is conducted for all the real and simulated SAR and optical images. The first reason is the same as [18]. It is difficult for the neural network to learn the rotation operation that is not linearly accumulative. Second, if the target region in the images rotates along with the azimuth angle, the PD will be mainly contributed by the different locations of the bright area, which makes the similarity unable to represent the critical variation of reflection feature of target with different orientations. A center cropping of 64 $\times$ 64 is also performed since the rest part is largely background. In this article, no

TABLE I
Numbers of Images of Each Equipment Type

Depression	2S1	BRDM2	BTR60	T62	T72A07	ZSU234
15°	274	274	195	273	274	274
17°	299	298	256	299	299	299

Fig. 4. Example images of six military equipment from different datasets. (a) Real photos. (b) Real SAR images. (c) Optical images given by rendering. (d) Simulated SAR images. Photos in (a) are provided in MSTAR dataset website [29]. For images in (b), (c), and (d), the azimuth angle is 144° and depression angle is 15°.

adjustment of pixel value has been performed on the SAR images.

The optical images are rendered by using OpenGL [30], a widely used application programming interface for rendering 2-D and 3-D vector graphics. Given the geometric models of the military equipment, the major purpose of rendering is to get the shadow under parallel light characterized by the depression and azimuth angle. We apply the shadow mapping method [49] to address this problem. The depth map is first created to store distances between points on the surface of target and the light source. Then, during the rendering of scene from the viewer's perspective, each pixel is checked against the depth map to determine if it is shadow or not. The shadow rendered in the condition of optical light shares some common features as that in SAR images. That is why we attempt to generate SAR images by translation from optical ones. An inconvenience of computer rendering is that it needs geometric models in advance. However, this is not the only way to get optical images, real photos, and remote sensing images with proper cropping are potential alternatives [50], which will be investigated in future works.

The simulation of SAR images is based on the SBR method [51], which uses the ray-tracing (RT) [52] approach to compute the equivalent currents. The rays are first shot on surface of target, then with application of physical or geometrical optics, the electromagnetic (EM) fields of each ray are tracked. The summation of EM contributions given by each ray will lead to the total scattered field of the target. The computation of SAR can be time consuming, a fast algorithm based on ray-tube integration formulas is thus introduced [53], [54], and then, expanded to the 3-D SAR simulation [55]. To improve the result of simulation, we also introduce beam-tracing (BT) method [56] to replace the RT used in SBR. The BT approach involves bundles or beams of rays rather than individual rays in RT, and thus, captures the properties of EM wave propagation more accurately.

It should be noted that the real SAR, simulated SAR, and optical images are not strictly coregistered. Thus, the same point on the geometric models is likely to exist at different pixel positions in corresponding images from these three sets. Our model is expected to be capable of overcoming the mismatches. Example images drawn from the three datasets of the military equipment considered in this work are shown in Fig. 4.

B. Experimental Settings

We choose the same setting of extremely few training samples as [19]. Only five real SAR images with 72° angle interval are given for each category, making a six-way five-shot task. The training set drawn from MSTAR contains 30 images, and the test set has 1534 images for depression of 15° and 1720 images for 17°. The source dataset $\mathcal{D}_S$ consists of simulated SAR with orientation interval of 5°, which gives us 432 images for each depression angle. The optical images are rendered for every integer azimuth angle. As to the hyperparameters in the loss function, we set $\lambda_1 = \lambda_c = \lambda_a = 1$, $\lambda_{\text{gp}} = 10$. As to the U-Net generator, the total number of layers is 12, which gives 1×1 1 feature maps in the bottleneck layer for the 64×64 input images. To update the model weights, we choose RMSprop optimizer [57] and the learning rate is 2×10^{-5}. The pretraining process is performed on $\mathcal{D}_S$ with a reconstruction loss by 20 epochs, i.e., $n_p = 20$. In each training step, D_{SN} will be optimized once and parameters of G will be updated three times, *i.e.*, $n_d = 1$, $n_g = 3$. The total number of steps n_s is set to be 4000. Our method is constructed on the Pytorch platform and run on a graphics processing unit (GPU) (Nvidia RTX A5000).

C. Augmentation Effect

The major purpose of FSIG of SAR is to generate more training data for target classification task. So, the classification accuracy after augmented by the generated images is regarded as an important criterion of our method. The A-ConvNets [31], which contains only convolutional networks without fully connected layers, is trained on different training sets and give the corresponding test accuracy. Compared with its original version, appropriate adjustment of the kernel size is applied to fit the image size in our experiments. We use cross entropy loss with Adam optimizer [58] to train the classification network. The

TABLE II
MODULES OF DIFFERENT SETTINGS

Setting	Azimuth and Category	Rendered image	PD (S)	PD (O)
AC	✓			
O2S	✓	✓		
O2S + PD(S)	✓	✓	✓	
O2S + PD(O)	✓	✓		✓

TABLE III
TEST CLASSIFICATION OF DIFFERENT AUGMENTATION SETTINGS

Depression	Setting	Accuracy
15°	Baseline	0.5992
	AC	0.6410
	O2S	0.6473
	O2S + PD(S)	**0.6696**
	O2S + PD(O)	0.5480
	AAE	0.6044
	AGGAN	0.5541
17°	Baseline	0.6167
	AC	0.6376
	O2S	0.6394
	O2S + PD(S)	**0.6572**
	O2S + PD(O)	0.5249
	AAE	0.5782
	AGGAN	0.5183

They indicate that our method outperforms other approaches on these metrics.

learning rate and number of training epochs are set to be 10^{-4} and 800, respectively. It should also be noticed that, the images in training and test sets of classification experiments are all rotated with respect to their orientation angles. The generation results possess a certain degree of randomness, since we have the random noise as input and the model parameters are randomly initialized. And the test accuracy also fluctuates in different experiments even though the training set remains the same. Thus, to eliminate the influence of randomness as much as possible, we generate five outcomes for each generation setting and train the A-ConvNets model on each generated results for five times. The average accuracy of 25 classification experiments is used for evaluation. The results of 15° and 17° depression angles are listed in Table III. We use baseline to denote the accuracy when training only on the five real SAR images per category. AC denotes the scenario where only the azimuth angles and the category labels serve as input of U-Net. O2S stands for the result augmented by the data generated by the O2S translation method, and O2S+PD means the additional PD loss is applied in the augmentation data generation. In the parentheses after PD, S means the simulated SAR images are used to guide the PD and O means that the optical images are used instead. For both O2S+PD situations, λ_{dist} is set to be 2. The modules included by each generation configuration are listed in Table II.

It can be seen from Table III that the classification accuracy on the test set is improved by data augmentation with images generated by both O2S and O2S+PD(S). Augmentation by O2S elevate the accuracy by 0.0481 and 0.0227 for data of 15° and 17° depression angles, respectively. And PD(S) leads to another 0.0223 and 0.0178 improvement. The improvement obtained by O2S is consistent with [18], which involves image translation from semantic maps to SAR images. After being converted to grayscale images, the optical images given by rendering contains certain degree of semantic information. The further improvement achieved by O2S+PD(S) suggests the effectiveness of proposed PD loss. However, if we change the source domain from simulation images to optical images, the accuracy declines significantly and is even smaller than results without any augmentation.

We also attempt to compare our method with the adversarial autoencoder (AAE) approach in [18] and attribute-guided generative adversarial network (AGGAN) in [19]. Since the code of both these two methods is not publicly available, we have done our best to carry out the implementation by ourselves according to the description of the algorithms in the articles. The semantic maps used as input of AAE in [18] are also not accessible for us. Thus, we utilize the rendered optical images in our approach instead. In the original article of the AGGAN approach [19], real SAR images of 7 types of targets in MSTAR are used as source domain to assist the 3-way 5-shot generation task. Since the number of categories under consideration is doubled in our scenario, there are not enough real SAR images in MSTAR that can be applied as source domain. Therefore, we use the same simulated SAR images in our method to aid the AGGAN model. The results of augmentation experiments show that our model outperforms the AAE and AGGAN approaches, at least when using our rendered optical images and simulation data.

D. Enhanced Diversity

To prove that the PD regularized by simulation images indeed enhance the diversity of generated results, we list the similarity variations of distinct image sets for all six target categories of both 15° and 17° depression angles in Fig. 5. It should be noticed that the curves of generated images are given by average of five experiments as the accuracy shown in Table III. In general, the range of variation of simulated SAR is larger than that of real SAR images. And the mutual difference between the optical images is relatively smaller. For the variation of images generated by the O2S method (dark blue line in Fig. 5), it can be seen that, the general tendency is close to the variation of real SAR, even though only five real samples are used to train the model. However, there still exists nonnegligible difference. The major one is that the similarity does not drop as fast as the real SAR. This problem is mitigated by the introduction of PD loss. Notice the result of images obtained by O2S+PD(S) (red line in Fig. 5), for most of the concerned categories, the variation drops faster with the help of simulated SAR. For the images generated PD(O) (violet line in Fig. 5), their variation is clearly closer to the variation of optical images, and thus, largely deviates from the real SAR. This may explain the degradation of classification accuracy augmented by results of O2S+PD(O). On all the orientation angles where there exists real SAR images, the absolute

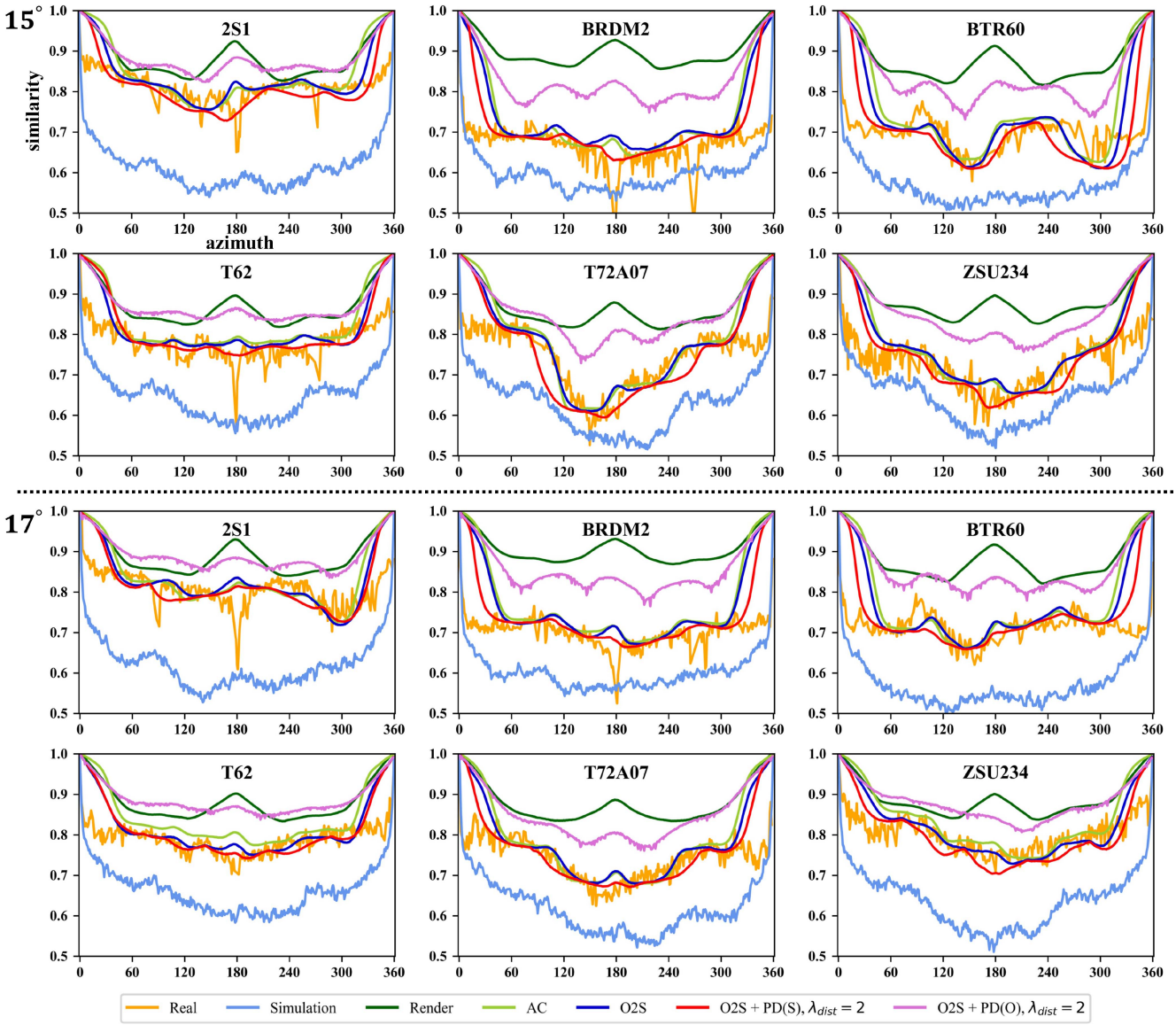


Fig. 5. Illustration of similarity variations of different image sets with respect to orientation angle. The upper part is for images with depression angle equaling 15° and the lower part is for 17°. AC indicates the situation that only azimuth angles and category labels are used as input of U-Net. O2S means the images generated by the O2S translation method. O2S+PD stands for the samples generated by O2S enhanced by pairwise distance loss. S and O in the parentheses indicates that the distance is calculated with help of simulation SAR and optical images, respectively.

difference between the similarity curves of distinct generated image sets and the line of real SAR is calculated. And the mean absolute differences are listed in the third column of Table IV, where "Curve Diff" represents the differences of curves. It can be seen that, the diversity variation tendency of O2S+PD(S) is the closest one to the real SAR and the O2S+PD(O) is the most unlike one. This corresponds to the augmentation results shown in Table III, which means that the image set with most similar diversity variation improves the classification accuracy the most. Compare with the O2S method, the O2S+PD(S) reduces the mean difference from 0.0571 to 0.0563 for the depression angle of 15° and it lowers the difference from 0.0452 to 0.0402 for 17°, indicating that the PD loss improves the diversity behavior once appropriate source domain is selected. The difference of similarity curves for AAE and AGGAN in Table IV also shows that our method is better than them.

E. Generated Samples

To better illustrate the effectiveness of our method, we present several samples of different generation settings for all the equipment categories. Since the average image of different experiments may have some unexpected dissimilarity compare with the real SAR image, the results of O2S+PD(S) with best average augmentation effect of depression angle 15° and 17° are shown in Figs. 6 and 7, respectively. In each subfigure, the upper row is the real SAR and the lower row is the generated ones. From left to right, the azimuth angle change from 72° to 144° with

Fig. 6. Samples of real SAR and images generated by O2S+PD(S) with depression angle 15° of (a) 2S1, (b) BRDM2, (c) BTR60, (d) T62, (e) T72A07, and (f) ZSU234. In each panel, the upper row is real SAR and lower row is generated images. Those with red frame are training samples.

Fig. 7. Samples of real SAR and images generated by O2S+PD(S) with depression angle 17° of (a) 2S1, (b) BRDM2, (c) BTR60, (d) T62, (e) T72A07, and (f) ZSU234. In each panel, the upper row is real SAR and lower row is generated images. Those with red frames are training samples.

TABLE IV
MEAN ABSOLUTE DIFFERENCES OF SIMILARITY CURVES AND SIMILARITY BETWEEN GENERATED AND REAL IMAGES

Depression	Setting	Curve Diff	GSSIM_1	GSSIM_2	MS-SSIM_1	MS-SSIM_2
15°	AC	0.0635	0.5993	0.6172	0.6225	0.7007
	O2S	0.0571	0.6051	0.6208	0.6336	0.7065
	O2S + PD(S)	**0.0563**	**0.6086**	**0.6239**	**0.6409**	**0.7148**
	O2S + PD(O)	0.1068	0.6013	0.6081	0.6183	0.6878
	AAE	0.0799	0.5431	0.5526	0.5059	0.5620
	AGGAN	0.1588	0.3765	0.4263	0.6128	0.6775
17°	AC	0.0503	0.6014	0.6226	0.6255	0.7085
	O2S	0.0452	0.6116	0.6285	0.6314	0.7093
	O2S + PD(S)	**0.0402**	**0.6125**	**0.6289**	**0.6332**	**0.7106**
	O2S + PD(O)	0.1003	0.5994	0.6052	0.6151	0.6844
	AAE	0.0658	0.5215	0.5238	0.4888	0.5528
	AGGAN	0.0861	0.5045	0.5336	0.5987	0.6691

They indicate that our method outperforms other approaches on these metrics.

12° as step. The real SAR images in red boxes are the training samples. By comparing with the real ones, it can be seen that the bright target area and dark shadow region of generated images have high similarity with the real ones. The rotation of shadow and illuminated area with respect to the azimuth angle is clearly displayed in the generated images. This indicates that our model is able to interpolate the few training images along the azimuth angle.

We also introduce gradient-based structural similarity (GSSIM) [59] and multiscale structural similarity (MS-SSIM) [60] as indexes to quantitatively evaluate the generated images. These two indexes are both improved version of widely used structural similarity (SSIM) [61]. Using a and b to denote two images, the traditional SSIM index of a and b consists of three parts: luminance comparison $l(a,b)$, contrast comparison $c(a,b)$, and structure comparison $s(a,b)$. SSIM is defined as

$$\text{SSIM}(a,b) = [l(a,b)]^{\alpha}\,[c(a,b)]^{\beta}\,[s(a,b)]^{\gamma} \tag{19}$$

where α, β, and γ are parameters to adjust the relative importance of each comparison and are all set to be 1 here. To improve SSIM, GSSIM proposed to calculate contrast and structure comparisons based on the gradient maps of images a and b given by the Sobel operator. MS-SSIM first obtains sequences of images with iteratively application of a low-pass filter and downsampling of the filtered images. It provides more flexibility by using a combination of SSIM indexes of corresponding images in the sequences as the final criterion. We calculate the indexes of the generated images with the orientation angles not included in the training set. The output of five different experiments used to augment the classification network and to compute the similarity curves are all taken into consideration and the results listed in Table IV are average values of five experiments. GSSIM_1 and MS-SSIM_1 are performed for the whole image with size of 64×64. GSSIM_2 and MS-SSIM_2 are calculated with only the center target region taken into consideration to relatively reduce the influence of the background region. For getting the pixel range of the target zone, we first select the strong scattering points by the pixel value, and then, regard the minimum bounding rectangle of these points as the target zone. Several examples of the target zone are shown in Fig. 8.

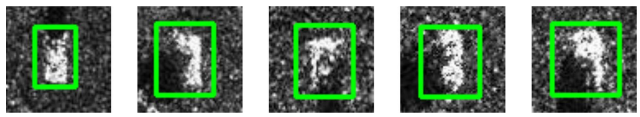


Fig. 8. Several examples of the target zone selected in the real SAR images. The target zone are boxed with green frames.

The mean similarity indexes performed on the generated images of different settings shows that the results enhanced with PD of simulated SAR images are the most similar to the real ones. It also indicates that our model is able to generate more similar images than AAE and AGGAN. This corresponds to the aforementioned results of augmentation experiments. The improvement of similarity indexes is on the general degree, and therefore, hard to discover for human vision. However, we would like to list several examples of generated images in Fig. 9 and try to give some possible aspects that are enhanced by the introduction of rendered optical images and the PD loss with simulated data. Generally speaking, the target area of images given by O2S+PD(O) is more blurred than others, which may cause the too flat similarity curve of optical images. Since the target zone facing the radar are crucial for training classification networks [62], the unclear details of turrets probably results in the negative results of augmentation experiments. Compare results of AC in columns (a) and (c) with the other configurations, we can observe that the orientation and range of shadow area of AC is more dissimilar to the real ones. One potential explanation for this discrepancy is that the neural network finds it more challenging to learn the shape of shadow area solely from azimuth angle and category label, which lack the details provided by optical images. As to the images generated by O2S+PD(S), we can observe relatively darker shadow zone than the output of O2S in columns (a), (b), (d), and (e). The stronger contrast of shadow area may contributes to the faster decrease of similarity curves like the real SAR images. The target zone of the result by O2S+PD(S) in column (c) fades gradually into the shadow area, however the boundaries between target and shadow in the other two images are too sharp. The strong points of O2S in (f) is also slightly too strong compared with O2S+PD(S), which makes the latter the more similar one.

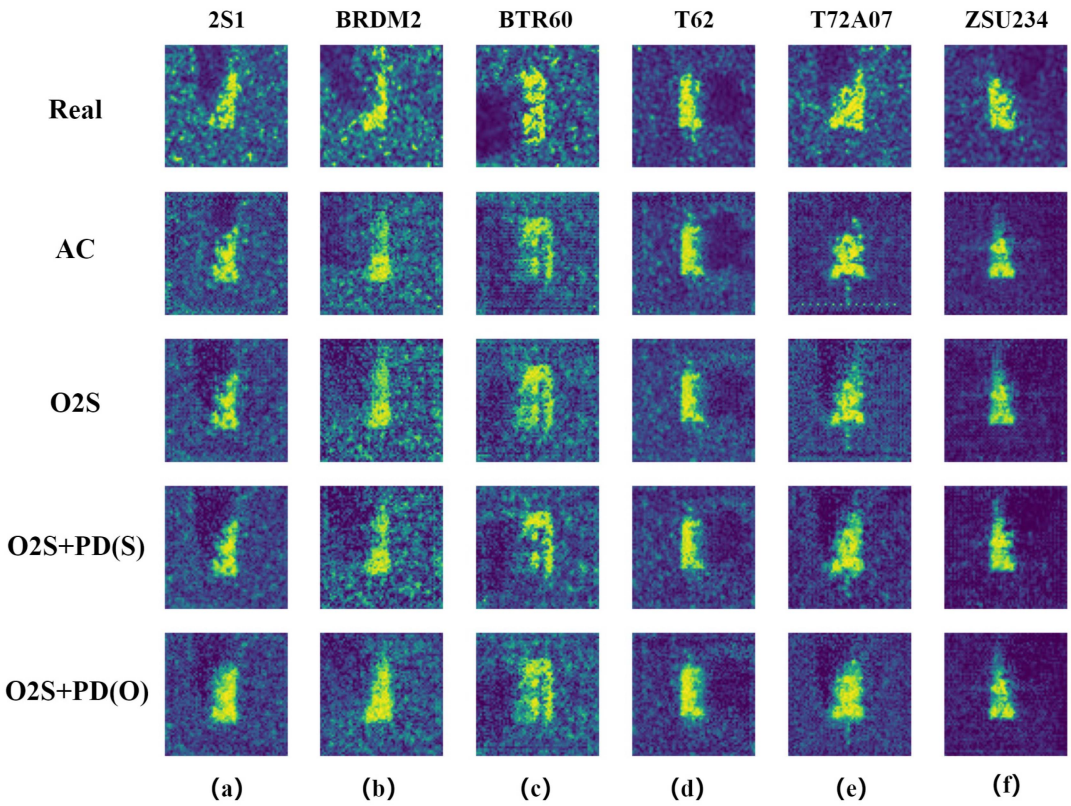


Fig. 9. Several generated samples to explain which part is possibly improved by the introduction of rendered optical images and PD loss on simulated data. A different color map is chosen to make the pixels of the shadow area more visible.

V. Discussion

We would like to give an analysis of the ablation study in this section to elucidate which component is able to bring improvement in the proposed model. As mentioned in Section I, the primary components that play a pivotal role in our proposed model are the rendered optical images and the PD loss with simulated data. By comparing the performances between the AC and O2S configurations, we can illustrate the effect of utilizing rendered images as input for the U-Net. The data augmentation experiments in Table III and the similarity indexes in Table IV both reveal the improvement of O2S compared with AC. This is quite intuitive since optical images introduce much more information such as the geometric size of the military targets and the shape of shadow region. In a sense, the optical image serves as a semantic map akin to the approach in [18]. Similarly, the efficiency of PD loss can be illustrated by the comparison between O2S and O2S+PD(S). The improvement in diversity is evident when observing Fig. 5, since the curve of O2S+PD(S) (red line) exhibits closer behavior to the real SAR (orange line) than O2S (dark blue line). This indicates that direct constraint on the mutual relationship of the output images with simulated SAR images is capable of enhancing the diversity of generation. Notably, O2S+PD(O) performs worse than the baseline, indicating that the domain used to guide the mutual similarity of generation has to possess alike characteristics to the real SAR images. It also should be noticed that, the improvement on similarity indexes exhibits a nonlinear relationship with the increase in classification accuracy. A minor improvement of the former can result in a relatively large rise of the latter.

VI. Conclusion

In this article, we proposed an DL model for FSIG of SAR. The basic part of our approach is a pix2pix network, a model intensively used in I2I tasks that attempt to convert an optical image given by rendering to an SAR image. The application of optical images avoids the difficult availability of semantic maps used in previous works and introduces more physics information than the simple target category labels and angle value of orientation. To mitigate the lack of diversity that frequently occurs in FSIG tasks, we propose a PD loss to regularize the similarity behavior of generated SAR images with corresponding simulated SAR images. Experiments on MSTAR dataset indicates that the images generated by our model is capable of augmenting the classification accuracy of an all-convolution network. The basic O2S part induce 4.81% and 2.27% for the tasks of 15° and 17° depression angles, respectively. The introduction of PD loss will bring another 2.23% and 1.78% improvement. As it is shown in the similarity curves of different image sets, the similarity behavior can also be adjust closer to the real SAR images with the help of PD loss. Under certain condition, the optical images given by rendering is also not easily obtainable since it is in need of geometric models. We intend to replace them with real photo or appropriately cropped remote sensing images in the future works.

REFERENCES

[1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.

[2] Q. Zhao and J. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 2, pp. 643–654, Apr. 2001.

[3] Y. Sun, Z. Liu, S. Todorovic, and J. Li, "Adaptive boosting for SAR automatic target recognition," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 1, pp. 112–125, Jan. 2007.

[4] 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.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[6] 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.

[7] S. Kazemi, B. Yonel, and B. Yazici, "Deep learning for direct automatic target recognition from SAR data," in *Proc. IEEE Radar Conf.*, 2019, pp. 1–6.

[8] U. Majumder, E. Blasch, and D. Garren, *Deep Learning for Radar and Communications Automatic Target Recognition*. Norwood, MA, USA: Artech House, 2020.

[9] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Comput. Surv.*, vol. 53, no. 3, Jun. 2020, Art. no. 63.

[10] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, Jul. 2019, Art. no. 60.

[11] 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.

[12] Q. Song, H. Chen, F. Xu, and T. J. Cui, "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.

[13] L. Liu, Z. Pan, X. Qiu, and L. Peng, "SAR target classification with cycleGAN transferred simulated samples," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 4411–4414.

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

[15] 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.

[16] 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.

[17] 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.

[18] 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, Jun. 2022, Art. no. 5210015.

[19] Y. Sun et al., "Attribute-guided generative adversarial network with improved episode training strategy for few-shot SAR image generation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1785–1801, Jan. 2023.

[20] A. Noguchi and T. Harada, "Image generation from small datasets via batch statistics adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 2750–2758.

[21] M. Zhao, Y. Cong, and L. Carin, "On leveraging pretrained GANs for generation with limited data," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, pp. 11340–11351.

[22] Y. Zhao, K. Chandrasegaran, M. Abdollahzadeh, and N.-M. M. Cheung, "Few-shot image generation via adaptation-aware kernel modulation," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2022, vol. 35, pp. 19427–19440.

[23] A. Ly, M. Marsman, J. Verhagen, R. P. Grasman, and E.-J. Wagenmakers, "A tutorial on Fisher information," *J. Math. Psychol.*, vol. 80, pp. 40–55, 2017.

[24] Y. Li, R. Zhang, J. Lu, and E. Shechtman, "Few-shot image generation with elastic weight consolidation," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 15885–15896.

[25] U. Ojha et al., "Few-shot image generation via cross-domain correspondence," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 10743–10752.

[26] Y. Pang, J. Lin, T. Qin, and Z. Chen, "Image-to-image translation: Methods and applications," *IEEE Trans. Multimedia*, vol. 24, pp. 3859–3881, Sep. 2022.

[27] 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.

[28] J. R. Diemunsch and J. Wissinger, "Moving and stationary target acquisition and recognition (MSTAR) model-based automatic target recognition: Search technology for a robust ATR," *Proc. SPIE*, vol. 3370, pp. 481–492, 1998.

[29] "The air force moving and stationary target recognition database," Accessed : Jun. 1, 2023. [Online]. Available: https://www.sdms.afrl.af.mil/-datasets/mstar/

[30] "Open graphics library," Accessed : Jun. 1, 2023. [Online]. Available: https://www.opengl.org/

[31] 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.

[32] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, vol. 70, pp. 214–223.

[33] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5769–5779.

[34] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[35] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song, "Neural style transfer: A review," *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 11, pp. 3365–3385, Nov. 2020.

[36] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2242–2251.

[37] D. Ao, C. O. Dumitru, G. Schwarz, and M. Datcu, "Dialectical GAN for SAR image translation: From Sentinel-1 to TerraSAR-X," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1597.

[38] L. Wang et al., "SAR-to-Optical image translation using supervised cycle-consistent adversarial networks," *IEEE Access*, vol. 7, pp. 129136–129149, 2019.

[39] M. Fuentes Reyes, S. Auer, N. Merkle, C. Henry, and M. Schmitt, "SAR-to-Optical image translation based on conditional generative adversarial networks-optimization, opportunities and limits," *Remote Sens.*, vol. 11, no. 17, 2019, Art. no. 2067.

[40] S. Fu, F. Xu, and Y.-Q. Jin, "Reciprocal translation between SAR and optical remote sensing images with cascaded-residual adversarial networks," *Sci. China Inf. Sci.*, vol. 64, 2021, Art. no. 122301.

[41] M. Zhang, C. He, J. Zhang, Y. Yang, X. Peng, and J. Guo, "SAR-to-Optical image translation via neural partial differential equations," in *Proc. Int. Joint Conf. Artif. Intell.*, 2022, pp. 1644–1650.

[42] X. Yang, J. Zhao, Z. Wei, N. Wang, and X. Gao, "SAR-to-optical image translation based on improved CGAN," *Pattern Recognit.*, vol. 121, 2022, Art. no. 108208.

[43] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2015, pp. 234–241.

[44] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn.*, 2013, Art. no. 3.

[45] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," 2016, *arXiv:1607.08022*.

[46] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Proc. Int. Conf. Learn. Representations*, 2018, p. 26.

[47] G. H. Golub and H. A. Van der Vorst, "Eigenvalue computation in the 20th century," *J. Comput. Appl. Math.*, vol. 123, no. 1, pp. 35–65, 2000.

[48] Y. Yoshida and T. Miyato, "Spectral norm regularization for improving the generalizability of deep learning," 2017, *arXiv:1705.10941*.

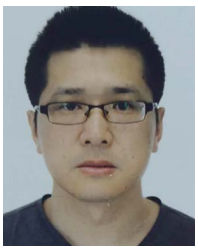
[49] L. Williams, "Casting curved shadows on curved surfaces," in *Proc. Annu. Conf. Comput. Graph. Interact. Techn.*, 1978, pp. 270–274.

[50] X. Sun, Y. Lv, Z. Wang, and K. Fu, "Scan: Scattering characteristics analysis network for few-shot aircraft classification in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Apr. 2022, Art. no. 5226517.

[51] H. Ling, R.-C. Chou, and S.-W. Lee, "Shooting and bouncing rays: Calculating the RCS of an arbitrarily shaped cavity," *IEEE Trans. Antennas Propag.*, vol. 37, no. 2, pp. 194–205, Feb. 1989.
[52] T. Whitted, "An improved illumination model for shaded display," *Commun. ACM*, vol. 23, no. 6, pp. 343–349, Jun. 1980.
[53] R. Bhalla and H. Ling, "A fast algorithm for signature prediction and image formation using the shooting and bouncing ray technique," *IEEE Trans. Antennas Propag.*, vol. 43, no. 7, pp. 727–731, Jul. 1995.
[54] R. Bhalla and H. Ling, "Image domain ray tube integration formula for the shooting and bouncing ray technique," *Radio Sci.*, vol. 30, no. 5, pp. 1435–1446, 1995.
[55] R. Bhalla, L. Lin, and D. Andersh, "A fast algorithm for 3D SAR simulation of target and terrain using xpatch," in *Proc. IEEE Int. Radar Conf.*, 2005, pp. 377–382.
[56] P. S. Heckbert and P. Hanrahan, "Beam tracing polygonal objects," in *Proc. Annu. Conf. Comput. Graph. Interact. Techn.*, New York, NY, USA, 1984, pp. 119–127.
[57] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*.
[58] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, p. 13.
[59] G.-H. Chen, C.-L. Yang, and S.-L. Xie, "Gradient-based structural similarity for image quality assessment," in *Proc. Int. Conf. Image Process.*, 2006, pp. 2929–2932.
[60] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 2003, vol. 2, pp. 1398–1402.
[61] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[62] C. Belloni, A. Balleri, N. Aouf, J.-M. Le Caillec, and T. Merlet, "Explainability of deep SAR ATR through feature analysis," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 1, pp. 659–673, Feb. 2021.

Jianghan Bao (Graduate Student Member, IEEE) was born in Chizhou, Anhui, China, in 1996. He received the B.Eng. degree in nuclear science and engineering from Sun Yat-sen University, Guangzhou, China, in 2017, and the M.Sc. degree in condensed matter physics from Nanjing University, Nanjing, China, in 2020. He is currently working toward the Ph.D. degree in electromagnetic fields and microwave technology with the State Key Laboratory of Millimeter Wave, Southeast University, Nanjing.

His research interests include application of deep learning in computational electromagnetic and metamaterial.



Wen Ming Yu was born in Zhuji, Zhejiang, China, in 1980. He received the B.Sc. and Ph.D. degrees in electromagnetic fields and microwave technology from the Nanjing University of Science and Technology, Nanjing, China, in 2002 and 2007, respectively .

He is currently a Lecturer with the School of Information Science and Engineering, Southeast University, Nanjing. His research interests include computational electromagnetics.

Kaiqiao Yang was born in Shaanxi, China, in 1999. He received the B.Sc. degree in electromagnetic fields and microwave technology from the Beijing University of Posts and Telecommunications, Beijing, China, in 2022. He is currently working toward the Ph.D. degree in electric engineering with Southeast University, Nanjing, China.

His current research interests include intelligent and parallel methods for electromagnetic computing.

Che Liu (Member, IEEE) was born in Suzhou, Jiangsu, China, in 1993. He received the B.Eng. degree in information science and technology and the Ph.D. degree in electromagnetic fields and microwave technology from Southeast University, Nanjing, China, in 2015 and 2022, respectively .

He is currently a Zhishan Postdoctoral Fellow with Southeast University. He is committed to use artificial intelligence technology solving electromagnetic issues, including ISAR imaging, holographic imaging, inverse scattering imaging, automatic antenna design, and diffraction neural network. His research interests include computational electromagnetic, metamaterial, and deep learning.

Tie Jun Cui (Fellow, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electromagnetic fields and microwave technology from Xidian University, Xi'an, China, in 1987, 1990, and 1993, respectively.

In 1993, he joined the Department of Electromagnetic Engineering, Xidian University, and was promoted to an Associate Professor in November 1993. From 1995 to 1997, he was a Research Fellow with the Institut fur Hochstfrequenztechnik und Elektronik (IHE), University of Karlsruhe, Karlsruhe, Germany. In 1997, he joined the Center for Computational Electromagnetics, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Champaign, IL, USA, first as a Postdoctoral Research Associate, and then, as a Research Scientist. In 2001, he was a Cheung-Kong Professor with the Department of Radio Engineering, Southeast University, Nanjing, China. In 2018, he became the Chief Professor with Southeast University. He is currently an Academician with the Chinese Academy of Science. He is the first author of the books *Metamaterials: Theory, Design, and Applications* (Springer, November 2009), *Metamaterials: Beyond Crystals, Noncrystals, and Quasicrystals* (CRC Press, March 2016), and *Information Metamaterials* (Cambridge University Press, 2021). He has authored or coauthored more than 600 peer-reviewed journal articles, which have been cited by more than 62000 times (H-Factor 122), and licensed more than 150 patents.

Dr. Cui was an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and a Guest Editor of *Science China Information Sciences, Science Bulletin*, IEEE JOURNAL ON EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS, and *Research*. He is the Chief Editor for *Metamaterial Short Books* (Cambridge University Press), an Editor for *Materials Today Electronics*, an Associate Editor for *Research*, and an Editorial Board Member for *National Science Review, eLight, PhotoniX, Advanced Optical Materials*, *Small Structure*, and *Advanced Photonics Research*. He presented more than 100 Keynote and Plenary Talks in Academic Conferences, Symposiums, or Workshops. From 2019 to 2023, he was ranked in the top 1% for the highly cited articles in the field of Physics by Clarivate Web of Science (Highly Cited Researcher). His research has been selected as one of the most exciting peer-reviewed optics research Optics in 2016 by Optics and Photonics News Magazine, ten Breakthroughs of China Science in 2010, and many Research Highlights in a series of journals. His work has been widely reported by *Nature News*, *MIT Technology Review*, *Scientific American*, *Discover*, and *New Scientists*. He was the recipient of the Research Fellowship from Alexander von Humboldt Foundation, Bonn, Germany, in 1995, the Young Scientist Award from the International Union of Radio Science in 1999, Cheung Kong Professor by the Ministry of Education, China, in 2001, and the National Science Foundation of China for Distinguished Young Scholars in 2002. He was also the recipient of the Natural Science Award (first-class) from the Ministry of Education, China, in 2011, and the National Natural Science Awards of China (second-class, twice) in 2014 and 2018.