Learning to Generate SAR Images With Adversarial Autoencoder

Qian Song, *Member, IEEE*, Feng Xu, *Senior Member, IEEE*, Xiao Xiang Zhu, *Fellow, IEEE*, and Ya-Qiu Jin, *Life Fellow, IEEE*

***Abstract*— Deep learning-based synthetic aperture radar (SAR) target recognition often suffers from sparsely distributed training samples and rapid angular variations due to scattering scintillation. Thus, data-driven SAR target recognition is considered a typical few-shot learning (FSL) task. This article first reviews the key issues of FSL and provides a definition of the FSL task. A novel adversarial autoencoder (AAE) is then proposed as an SAR representation and generation network. It consists of a generator network that decodes target knowledge to SAR images and an adversarial discriminator network that not only learns to discriminate "fake" generated images from real ones but also encodes the input SAR image back to target knowledge. The discriminator employs progressively expanding convolution layers and a corresponding layer-by-layer training strategy. It uses two cyclic loss functions to enforce consistency between the inputs and outputs. Moreover, rotated cropping is introduced as a mechanism to address the challenge of representing the target orientation. The moving and stationary Target recognition (MSTAR) 7-target dataset is used to evaluate the AAE's performance, and the results demonstrate its ability to generate SAR images with aspect angular diversity. Using only 90 training samples with at least 25° of orientation interval, the trained AAE is able to generate the remaining 1748 samples of other orientation angles with an unprecedented level of fidelity. Thus, it can be used for data augmentation in SAR target recognition FSL tasks. Our experimental results show that the AAE could boost the test accuracy by 5.77%.**

***Index Terms*— Adversarial autoencoder (AAE), deep learning (DL), few-shot learning (FSL), image representation, synthetic aperture radar (SAR).**

I. Introduction

THE performance of deep learning (DL)-based automatic target recognition (ATR) from synthetic aperture radar (SAR) images is often limited by the diversity and scale of the available training sample images [1]. On the one hand, SAR images are extremely sensitive to observation configurations; for example, target features such as the distribution of strong scattering points in SAR images vary rapidly with the view angle. This is because SAR images are formed via coherent summation of the electromagnetic scattering field; thus, they are affected by the scintillation phenomenon of the radar cross section. The intensity of SAR images is determined by the amplitude of the backscattered waves from targets and thus depends on wave incidence, target geometry, and material. On the other hand, it is difficult to obtain adequate SAR image samples due to the inaccessibility of SAR sensors [2], so-called few-shot learning (FSL) problem. SAR images are much less commonly available than optical images. The scarcity and high variability of SAR images have become a major hindrance to implement data-driven approaches for SAR ATR.

To address this issue, supplementary data can be acquired from other sources, such as simulation data [3]–[5], optical images [6], [7], and low-resolution SAR images [8]. These data sources are easier to obtain but contain the same semantic information as high-resolution SAR images. Thus, these supplementary data are often used to pretrain a deep neural network (DNN), which is later fine-tuned on the real target dataset using a process known as transfer learning. Transfer learning effectively improves model generalizability because it avoids having to train a network from scratch starting from a randomly initialized state. Other methods that have been proposed link high-resolution SAR images with other sources of data [5], [7], [9]. However, the differences between the different data sources cause an underlying nonnegligible feature shifting problem [5].

Alternatively, the number of unknown network parameters to be trained can be reduced to avoid overfitting due to the limited training data. Chen *et al.* [10] proposed using an all-convolution network (AConvNet) to reduce the number of trainable parameters. Regularization terms such as the center loss [11] were introduced to restrict model complexity. Recently, generative networks such as generative adversarial networks (GANs) [12] have been used to generate SAR-like images [4], [13], [14]; then the recognition network is trained on the real images as well as the generated images. This approach can be considered as a type of data augmentation. However, as we will discuss in this article, the essential problem behind it is the lack of diversity of the training samples in physical dimensions. Physics-based generative

Manuscript received January 23, 2021; revised March 19, 2021 and May 8, 2021; accepted May 31, 2021. This work was supported by the Natural Science Foundation of China under Grant 61991422 and Grant 61822107. *(Corresponding author: Feng Xu.)*

Qian Song was with the Key Laboratory for Information Science of Electromagnetic Waves (MoE), Fudan University, Shanghai 200433, China. She is now with the German Aerospace Center, Remote Sensing Technology Institute, 82234 Wessling, Germany.

Feng Xu and Ya-Qiu Jin are with the Key Laboratory for Information Science of Electromagnetic Waves (MoE), Fudan University, Shanghai 200433, China (e-mail: fengxu@fudan.edu.cn).

Xiao Xiang Zhu is with the German Aerospace Center, Remote Sensing Technology Institute, 82234 Wessling, Germany, and also with the Data Science in Earth Observation, Technical University of Munich, 80333 Munich, Germany.

Digital Object Identifier 10.1109/TGRS.2021.3086817

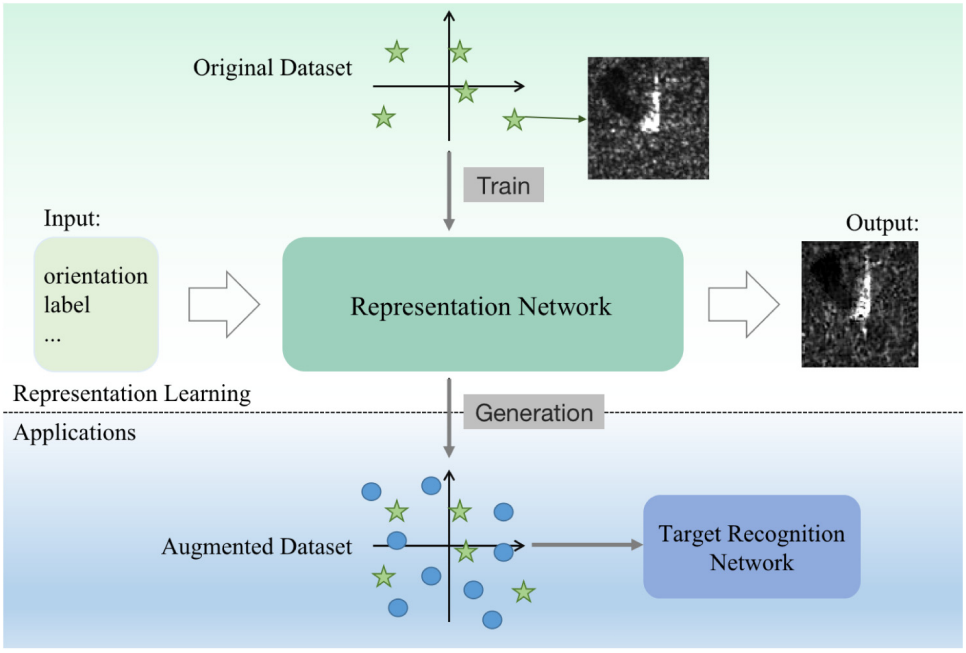


Fig. 1. Framework of the proposed AAE for representation learning and its applications.

networks are critical for SAR applications. These generative networks do not consider the physical dimensions of the generated target; thus, they are not able to perform interpolation along physical dimensions such as the target orientation angle.

In this article, we propose an SAR image generation network that learns a high-dimensional abstract representation to address the FSL issue, as shown in Fig. 1. It includes two parts, representation learning and applications. At first, the representation network, i.e., the proposed adversarial autoencoder (AAE), takes an attribute vector as input consisting of target intrinsic features and radar observation conditions and outputs the corresponding generated SAR image. After training with limited samples (pentacles in Fig. 1), the network is able to generate not only suitable training samples but also new samples with other observation conditions (circles in Fig. 1), such as interpolated viewing angles. Using this representation network, data augmentation can be performed via interpolation along the attribute vector dimensions. Then, the augmented data can be directly applied to FSL tasks, such as target recognition. The contributions of this article are: 1) we reformulate and redefine the FSL problem in SAR target recognition; 2) a novel generative network that is able to improve the diversity of the training samples in the physical dimension is proposed; 3) rotated cropping is proposed to deal with rotation representation problem of generation network.

The remainder of this article is organized as follows. Section II reviews the generative network and the concept of the FSL problem. The architecture of the proposed AAE network is explained in Section III, while Section IV presents the experimental results on the moving and stationary Target recognition (MSTAR) dataset [16]. Section V discusses the contribution of the semantic map and rotated cropping and whether additional generated images improve the model's FSL ability. Finally, Section VI concludes this article.

II. Related Works

A. Generative Network

Generative networks are an approach often adopted to improve model generalizability in SAR applications. The main objective for using a generative network is to compensate for missing information to make the distribution of training samples more similar to that of the testing samples [17]. Generally, the generative network takes the category label, orientation, and other target parameters as input and outputs an SAR-like image of the target. After training, the model can be used to generate abundant images of the target under various conditions.

GAN models [12], which are known for their ability to generate high-fidelity images, are one of the most commonly used types of generative networks. The two subnetworks in a GAN (i.e., a generator and a discriminator), respectively, minimize and maximize the prediction accuracy of the discrimination function. Experiments on several datasets have shown that using this type of optimization objective instead of a one-to-one mapping loss, such as the L_2 norm, can improve the sharpness of the generated images. The original GAN suffers from model collapse easily; thus, several modifications of the network architecture were suggested in [18]. In [19], the authors replaced the Jensen–Shannon divergence in the objective function with the Wasserstein distance, i.e., the so-called Wasserstein GAN (WGAN). In [20], a gradient penalty was introduced to replace the weight clipping operation; this improved model stability compared with a classical GAN.

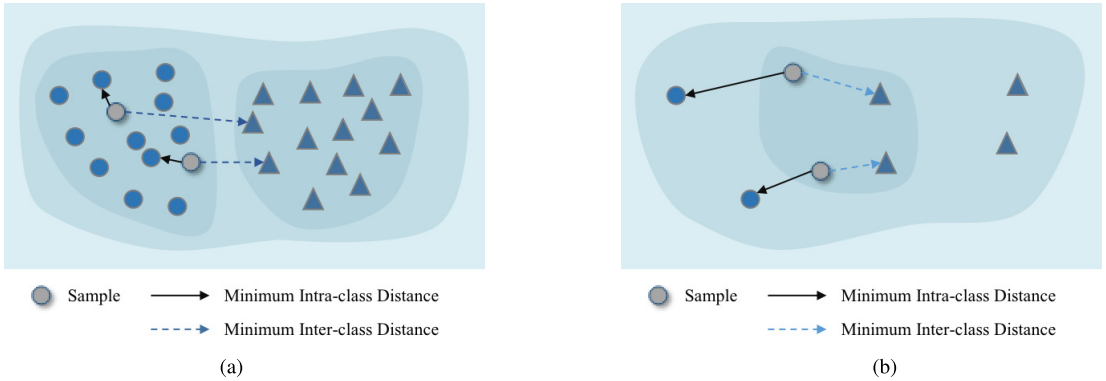


Fig. 2. Comparison of the common-setting (a) and FSL problem (b). The solid and dashed lines indicate the distances between the current sample and the nearest intraclass and interclass samples, respectively. (a) Classical learning problem. (b) FSL problem.

Guo *et al.* [21] proposed using a GAN to generate synthetic SAR images. Cui *et al.* [13] used a WGAN model to generate SAR images and then used a trained support vector machine (SVM) to eliminate meaningless images. Liu *et al.* [4] proposed using tools such as RaySAR to generate simulated images to increase the training set size. However, there were gaps between the simulated and real images. Therefore, they proposed using CycleGAN to adjust the simulation images. Their experiments indicated that using CycleGAN can improve the classification accuracy by as much as 9.3%. Zheng *et al.* [14] proposed Multidiscriminator GAN (MDGAN) model to improve training stability. The MDGAN trains multiple discriminators in parallel and uses a parameter to balance training stability with the training effect. Based on previous experimental results, the MDGAN model is superior to other generative networks. The generative network has also been applied to translation between optical and SAR images [15]. In this study, a generative network is applied for SAR target representation learning. However, despite their image generation ability, generative networks have difficulty representing physically plausible target features such as target orientation. In Section III-B, we reveal that this problem occurs because rotation is a nonlinear and discontinuous transformation; therefore, we propose rotated cropping to address this issue.

B. Few-Shot Learning

FSL is a type of machine learning task in which only a limited number of samples are available during the training stage [22]. Limited training sample availability is a common problem that hinders the improvement of SAR ATR applications. To our knowledge, although FSL has been a hot topic in fields such as machine learning, computer vision, and SAR-ATR, none of the existing works provides a specific definition of FSL. In [23], the authors define the FSL problem as a special case of traditional machine learning in which little supervised information is provided for the target. However, this definition does not answer the question: how many samples can be considered as a "few"?

In this article, we define FSL as a type of machine learning task that encounters low-density isolation ambiguity. More specifically, FSL refers to a task in which, after considering all the similarity and distance measures defined in original spaces, the average maximum intraclass similarity is less than the average maximum interclass similarity in the training set or in which the average minimum intraclass distance is larger than the average minimum interclass distance in the training set. These conditions can be notated as follows:

$$\begin{aligned} &\mathop{\mathbb{E}}_{i}\max_{k}\mathbb{S}(X_i,X_k) < \mathop{\mathbb{E}}_{m}\max_{o}\mathbb{S}(X_m,X_o) \\ \text{or}\ &\mathop{\mathbb{E}}_{i}\min_{k}\mathbb{D}(X_i,X_k) > \mathop{\mathbb{E}}_{m}\min_{o}\mathbb{D}(X_m,X_o) \\ &X_k \in \{X_j, y_i = y_j\}, \quad X_o \in \{X_n, y_m \neq y_n\} \end{aligned} \tag{1}$$

where X_i is the ith SAR image, and y_i the corresponding label; $\mathbb{E}$, $\mathbb{S}$, and $\mathbb{D}$ denote expectation, similarity metric, and distance metric, respectively.

Fig. 2 compares classical machine learning to FSL settings. The circles and the triangles indicate two types of targets, each of which corresponds to a sample. The solid and dashed lines show the minimum distances between the current sample and the intra and interclass samples, respectively. The figure shows that under the FSL setting, the high-density region lies in the border area between two types of targets. Thus, for a sample that lies in the border area, the nearest sample may be a different category.

Under this definition, FSL does not satisfy the traditional cluster assumption [24]. The cluster assumption states that the same types of targets are distributed close to and can be grouped into clusters; thus, the algorithms based on this assumption such as the low-density isolation [24] and transductive support vector machine (TSVM) [25] try to place the decision boundaries on the low-density regions. However, these algorithms are no longer applicable in FSL. In addition, deep convolutional neural network (CNN) models are prone to overfitting in FSL situations.

We believe that a generative network can be used to increase the density of the intraclass targets, thus provides a solution to the FSL problem.

III. Adversarial Autoencoder

In this section, we begin by introducing the proposed AAE network. We first elaborate on the network's two core modules and then explain the loss function and the learning algorithm.

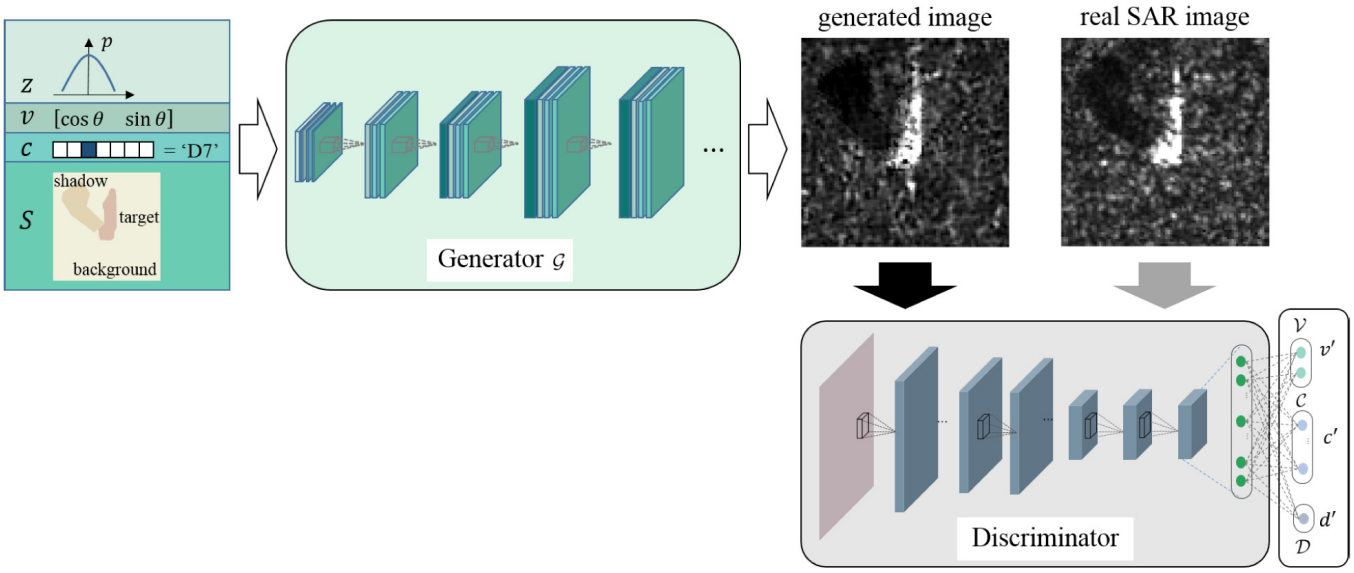


Fig. 3. Architecture of the proposed AAE.

A. AAE Architecture

The architecture of the proposed AAE is shown in Fig. 3. The AAE consists of two parts (i.e., a generator and a discriminator). The generator generates SAR images as realistically as possible, while the discriminator learns to discriminate these "fake" generated images from real SAR images. These two subnetworks are trained in an adversarial manner, as in a typical GAN [12].

The four generator inputs of $\mathcal{G}$ are random samples drawn from a standard normal distribution z, whose orientation is represented by $v = [\cos\theta, \sin\theta]$, where c is a one-hot label corresponding to the target type, and the semantic map is S. The semantic map $S \in \Re^{M*M*3}$ has three channels that correspond to the background mask, the shadow area, and the target area, respectively. The generator generates image I' with the same size as the real SAR images I. Note that in this article, we use the original amplitude data instead of adjusted grayscale values to train the network. Thus, the generated images are expected to retain the statistical characteristics of real SAR images, which are of potential use in object detection. The discriminator has three outputs: $d' = \mathcal{D}(X) \in \{0, 1\}$ indicates whether the input X is a real or fake image; $v' = \mathcal{V}(X)$ is the predicted orientation representation of the input image X; and $c' = \mathcal{C}(X) \in \Re^K$ is the predicted label for the input image, where K denotes the number of target types in the datasets.

When the discriminator is presented with generated images I', the proposed network is expected to predict the input orientation representation and the type label (i.e., $v' = \mathcal{V}(I') = v$ and $c' = \mathcal{C}(I') = c$). Then, the generator and discriminator can be regarded as a decoder and an encoder, respectively. Accordingly, the network is called an AAE.

However, because the orientation representation used here is discontinuous and because of the nonlinearity of rotation, it is difficult for the network to accurately represent the target's orientation feature. In Section III-B, we propose using rotated cropping instead of the traditional horizontal cropping to address this issue.

B. Rotated Cropping

It is challenging for neural networks to learn physically plausible feature representations such as target orientation because the rotation operation is not linearly accumulative, that is,

$$\begin{bmatrix}\cos(\theta+\beta) & -\sin(\theta+\beta)\\ \sin(\theta+\beta) & \cos(\theta+\beta)\end{bmatrix} \neq \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix} + \begin{bmatrix}\cos\beta & -\sin\beta\\ \sin\beta & \cos\beta\end{bmatrix}. \quad (2)$$

Thus, the features that depend on target orientation are not linearly interpolatable. As illustrated in Fig. 4, after being rotated by 45° in the clockwise direction around its center point, the rotated square becomes a diamond. Moreover, the change of the components' position is related to the distances between the center of rotation and the components' positions. As a result, the rotated green square is located further away from the original square than the yellow square. These two types of deformations are difficult for naive neural networks to learn.

In [26], the authors pointed out that no representation in four or fewer dimensions in a Euclidean space is continuous for 3-D rotation. A representation is deemed continuous when the mapping from the original space to the representation space (as well as its inverse mapping from the representation space to the original space) is continuous. Considering 2-D rotation, any legitimate rotation matrix M can be solely determined by the rotation angle $\theta \in [0, 2\pi]$. The mapping from the rotation matrix to the rotation angle is discontinuous at point $M = \mathbb{I}$, where $\mathbb{I}$ is the identity matrix because the mapping of M approaches either 0 or 2π. Such

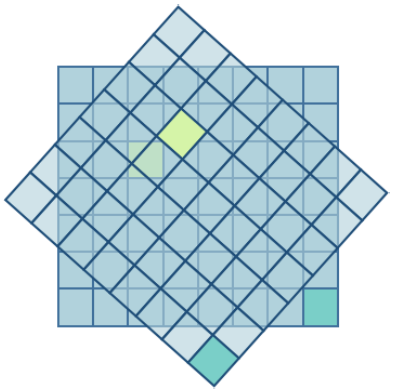


Fig. 4. Example image and the same image after a 45° rotation.

a discontinuous mechanism would need to be represented by multiple neurons; thus training a neural network would require many more samples and iterations [27].

The same target may appear in different orientations in SAR images depending on both the direction of the platform trajectory and the target status. Thus, a powerful SAR image representation network requires a large number of nonlinear units. When trained on sparsely distributed training samples, such a large-scale network could easily overfit. To alleviate this problem, we propose using rotated squares to crop the image when preparing the training set, as shown in Fig. 5. The edges of rotated squares are parallel/perpendicular to the target's orientation, instead of the range/azimuth direction. Hence, the bright areas of the same target in different orientations located in the same region. As will be shown later in the experimental results, the proposed rotated cropping can alleviate the discontinuous representation issue to some extent and improve the representation performance of the network.

C. Progressive Network

In this study, the architecture of the network is built progressively, a technique first used in progressive growing GAN (PGGAN) [28] to improve the training stability. The training process is divided into several stages. Initially, we build a shallow network as a generator ($\mathcal{G}$) that outputs an $m \times m$ image. This shallow network, which has few parameters to tune, can reach an optimal solution within a few iterations. In the next stage, we append convolution layers to the generator, increasing the size of the output image by 2×2, as shown in Fig. 6. By the final stage, the generator outputs images that are the same size as the real full-scale SAR images.

Accordingly, the discriminator is grown incrementally in the same manner. By adding more convolution layers at the front of the discriminator (left side in Fig. 3), the input grows from a low-resolution image to a full-resolution image. To form the training images, the real SAR images are downsampled in the earlier stages as follows:

$$I_{\text{low}}(i, j) = \sqrt{\sum U_{\Leftarrow}^2(I(\sigma \times i, \sigma \times j), \sigma)} \tag{3}$$

where $U_{\Leftarrow}(a, b)$ is the upper-left neighborhood of a with radius b. Here, σ is the scaling factor of the current stage.

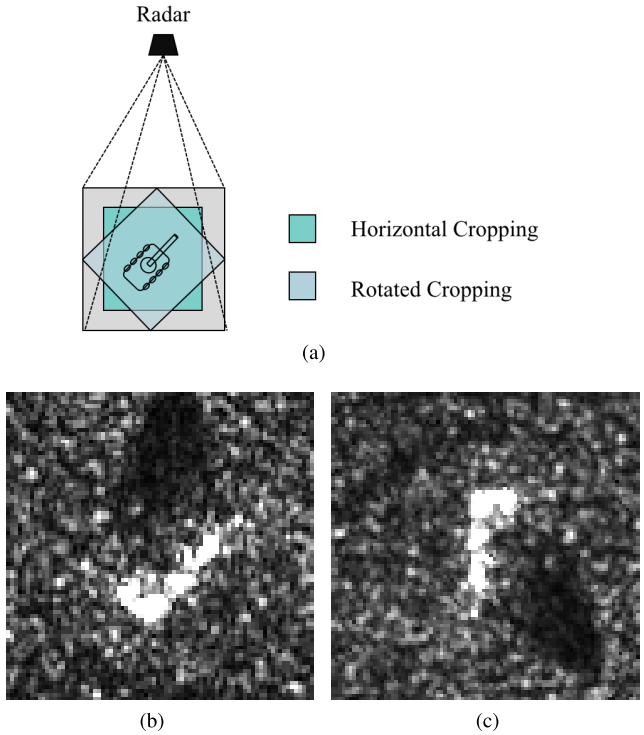


Fig. 5. Example of rotated-square image cropping of BRDM2. (a) Illustration of horizontal and rotated cropping. (b) Horizontal cropping of BRDM2. (c) Rotated cropping of BRDM2.

To smoothly apply the new layers, we use the weighted sum of the generated high-resolution image and enlarged low-resolution images as the generator output during the even-number iterations; we preserve the network architecture from the previous stage, but output the high-resolution images directly

$$\begin{aligned} o &= (1-\alpha)d_{l-1} + \alpha d_l, \quad \text{when } l \in \{2n, n \in \mathbb{N}^+\} \\ o &= d_l, \quad \text{when } l \in \{2n+1, n \in \mathbb{N}^+\} \end{aligned} \tag{4}$$

where α is the weight; l is the number of stages, d_{l-1} is an enlarged low-resolution image generated from the previous stage and expanded by a factor of two using nearest-neighbor interpolation, and d_l is the generated high-resolution image, as shown in Fig. 6. During training, α changes from 0 to 1 linearly, ensuring that the network is trained smoothly.

Note that the proposed generator has a dual-path structure: it contains two paths that generate two images of equal size. These two paths have the same number of parameters, but the parameters are initialized independently. The dual-path design increases the amount of variation in the generated images and thus improves the network's representation ability.

D. Loss Function

A loss or objective function defines a network's goal. In GANs, the generator and the discriminator are trained adversarially in a min-max game fashion. The objective of GAN [12] is defined as follows:

$$\min_{\mathcal{G}} \max_{\mathcal{D}} \ \mathbb{E}_X[\mathcal{D}(X)] - \mathbb{E}_{z \sim p(\text{data})}\{\mathcal{D}[\mathcal{G}(z, v, c, S)]\} \tag{5}$$

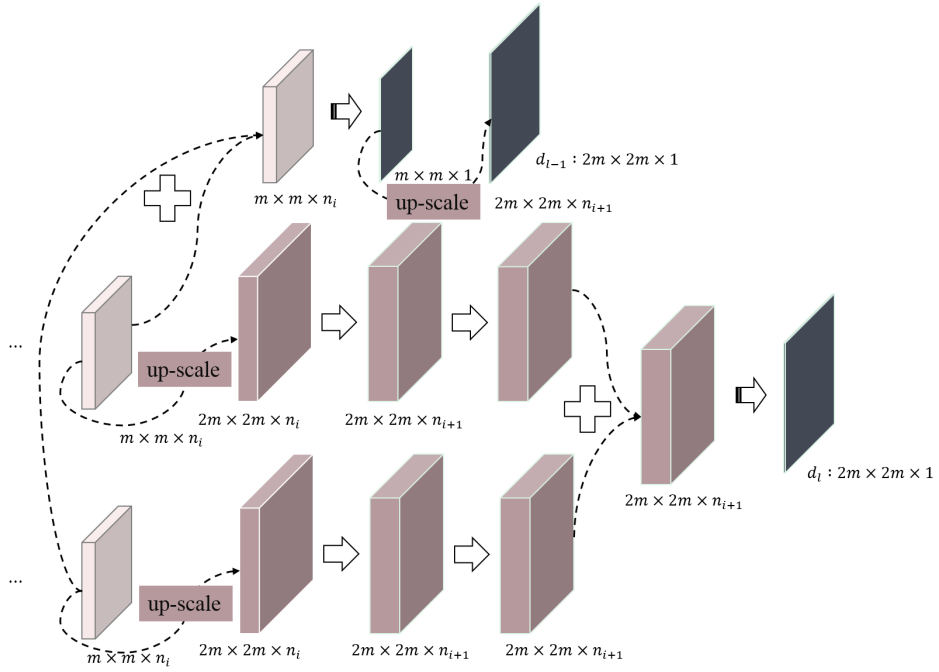


Fig. 6. Progressive architecture network used in this study. Here, n_i and n_{i+1} represent the number of channels at the ith and $(i+1)$th layers, respectively, and l denotes the number of training stages.

where $\mathbb{E}_a[f(a)]$ is the expected value of $f(a)$. To minimize the objective function, the output, $\mathcal{D}[\mathcal{G}(z, \upsilon, c, S)]$, is expected to be 1. Then, the generator is trained to generate images that are as realistic as possible, which improves its representation capability. In contrast, the discriminator is trained to maximize the objective function to continuously improve its ability to identify discrepancies between real and fake SAR images, thus extending the upper limit of representation ability.

The most important difference between GANs and other generative networks is the loss function. Traditional generative networks often use direct distances between generated data and real data as the loss measure, such as the L_1 and L_2 norms, or Kullback-Leibler (KL) divergence. The loss in (5) is more intuitive, which causes the generated images to appear much sharper. However, GANs suffer from mode collapse [30], particularly when insufficient samples are available. Thus, in this article, we also use the L_1 norm between each specific generated image and the corresponding real image as an additional loss term to impose one-to-one mapping regularity

$$\mathcal{L}_g[\mathcal{G}(z, \upsilon, c, S)|X] = \mathbb{E}_X[\|\mathcal{G}(z, \upsilon, c, S) - X\|_1] \tag{6}$$

where the inputs υ, c, and S, respectively, correspond to the orientation representation, labels, and semantic maps of real SAR images X.

In addition, when a generated image I' is input into the discriminator, the predictors $\mathcal{V}$ and $\mathcal{C}$ are expected to be consistent with the inputs of $\mathcal{G}$. Therefore, we further introduce the L_2 norm and cross entropy to ensure the consistency of the orientation representation and label prediction, respectively, which are defined as follows:

$$\mathcal{L}_\upsilon(\upsilon'|\upsilon) = \mathbb{E}_\upsilon\left(\|\upsilon - \upsilon'\|_2^2\right) \tag{7}$$

$$\mathcal{L}_c(c'|c) = \mathcal{D}_{\text{entropy}}(c'|c) = \mathbb{E}_c(c \log c'). \tag{8}$$

The loss function of the generator is summarized by

$$\begin{aligned}\arg\min_{\mathcal{G}} \ & \lambda_\upsilon \mathcal{L}_\upsilon\{\mathcal{V}[\mathcal{G}(\upsilon|z, c, S)]|\upsilon'\} \\ & + \lambda_c \mathcal{L}_c\{\mathcal{C}[\mathcal{G}(c|z, \upsilon, S)]|c'\} + \lambda_g \mathcal{L}_g[\mathcal{G}(z, \upsilon_t, c_t, S_t)|X] \\ & - \mathbb{E}_{z\sim p(\text{data})}\{\mathcal{D}[\mathcal{G}(z, \upsilon, c, S)]\}\end{aligned} \tag{9}$$

where the hyperparameters $\lambda_\upsilon = 5$, $\lambda_c = 10$, and $\lambda_g = 1$ are used to balance the four loss terms, and υ_t, c_t, and S_t are correspond, respectively, to the orientation, labels and semantic maps of the training set.

The loss function of the discriminator is summarized as follows:

$$\begin{aligned}\arg\min_{\mathcal{D}} \ & \lambda_\upsilon \mathcal{L}_\upsilon[\mathcal{V}(X)|\upsilon_t] + \lambda_c \mathcal{L}_c[\mathcal{C}(X)|c_t] \\ & + \mathbb{E}_{z\sim p(\text{data})}\{\mathcal{D}[\mathcal{G}(z, \upsilon, c, S)]\} - \mathbb{E}_X[\mathcal{D}(X)] \\ & + \lambda_{\text{gp}}\left[\left\|\nabla_{\hat{X}}\mathcal{D}(\hat{X})\right\|_2 - 1\right]^2\end{aligned} \tag{10}$$

where the hyperparameters are set to $\lambda_\upsilon = 5$, $\lambda_c = 10$, and $\lambda_{\text{gp}} = 10$. Note that the last term is the gradient penalty proposed in [20], where $\hat{X}$ is the interpolated image, and $\nabla_{\hat{X}}\mathcal{D}(\hat{X})$ is the gradient of $\mathcal{D}$'s outputs with respect to $\hat{X}$. This ensures that when the inputs are modified only slightly, the weights will not change much, thus ensuring the stability of the training process.

IV. Experimental Results

A. Datasets

We adopt the widely used MSTAR dataset [16] to validate the representation ability of the proposed AAE network. Seven types of targets at all azimuth angles and a 15° depression angle were obtained using an X-band one-foot resolution

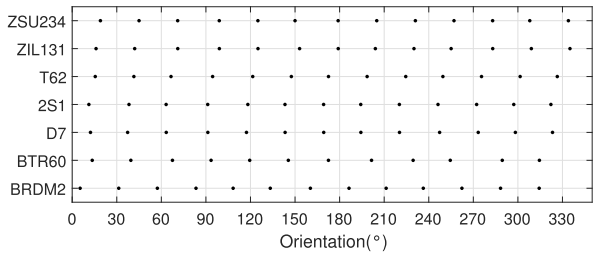


Fig. 7. Orientation distribution of the training images.

SAR sensor. Note that in this article, we use the original amplitude values instead of adjusted grayscale images to train the network. Thus, the generated images are expected to retain the statistical characteristics of real SAR images, which are of potential use in object detection. As one of the few datasets that provides full-aspect high-resolution real SAR data, MSTAR is our optimal choice for examining the aspect angular diversity of generated images with the proposed network.

The original chips of size 128×128 in MSTAR are rotated in reverse by their orientation to the aligned angles to conduct rotated cropping. After rotation, the targets are still distributed within the same area but exhibit different features. Note that the shadow areas of chips under different orientations are also different, but as the results will reveal, the shadowed areas are easier for the AAE to represent. The semantic maps are generated by manually annotating the bright and dark areas in the image as the target and shadow area, respectively, and the rest as background. In the future, this procedure can be automated using conventional image segmentation approaches.

In practical applications, it is difficult or even impossible to obtain all-aspect data samples of the same target. Usually, one can collect only limited samples at a few aspect angles. To mimic the practical target recognition task, we selected one sample at every 25° interval as the training set, which means that the orientation interval between any two samples for the same target type is at least 25° in the training set. Only 90 of 1838 chips are used as the training set; all the remaining 1748 chips were only used for testing purposes. Fig. 7 shows the orientation distribution of the training samples; the samples are sparsely distributed along the orientation dimension, which makes representation learning a challenging FSL task indeed.

B. Representation Ability

Using only the 90 sparsely distributed training samples, the proposed AAE is first trained to learn a general representation for MSTAR targets for all aspect angles. It can then generate an infinite number of synthetic "fake" samples of the seven types of targets with arbitrary orientations. When provided with the orientation angles and semantic maps of the test set, the trained AAE can generate 1748 samples for the test set. Note that the semantic maps of the test set are used as prior knowledge input into the AAE, which is not possible in real scenarios. In future work, we plan to implement a semantic map generation method that, would ideally be able to generate semantic maps for different targets and orientations. We believe that such a task would be much less challenging than generating SAR images. Next, we compare the generated sample images with the real images to evaluate the representation ability and generalizability of the AAE.

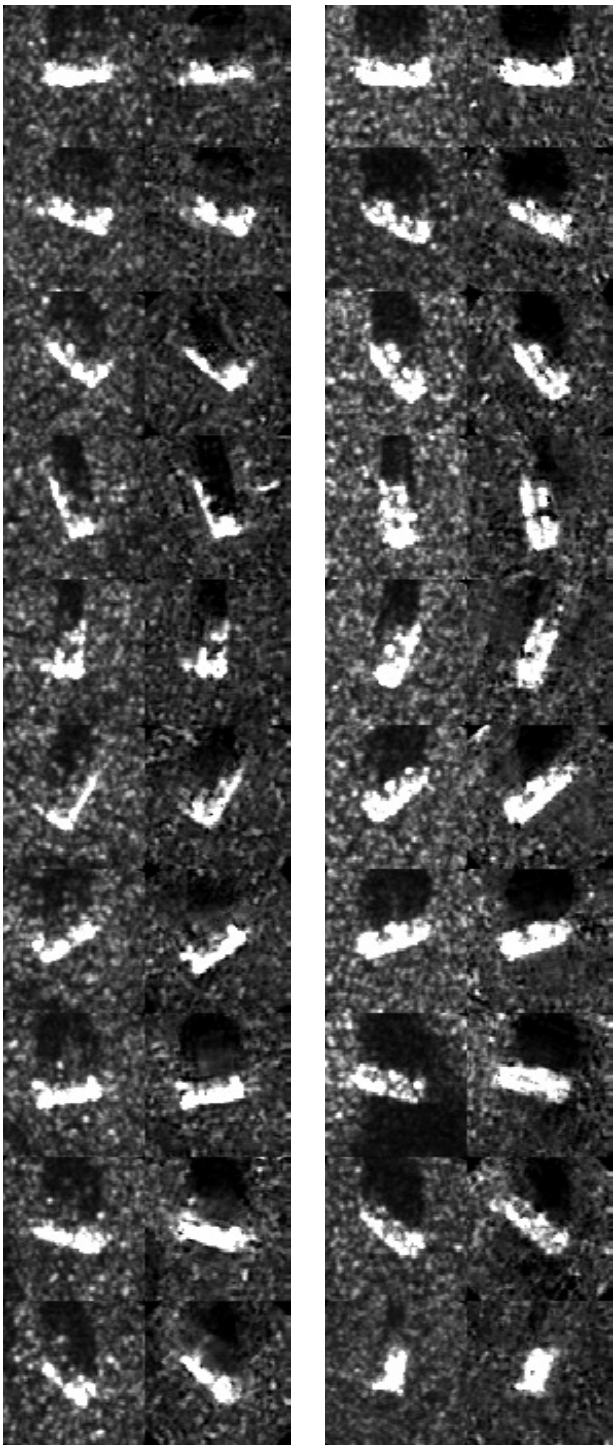


Fig. 8. Real SAR images (left column) compared to generated images (right column) of: (a) BRDM2 and (b) BTR60. Note that the outputs of AAE and the original data are normalized and adjusted for visualization.

The two sets of pictures in Fig. 8 show examples of the real SAR images of the targets BRDM2 and BTR60 and their corresponding generated images. Note that the generated

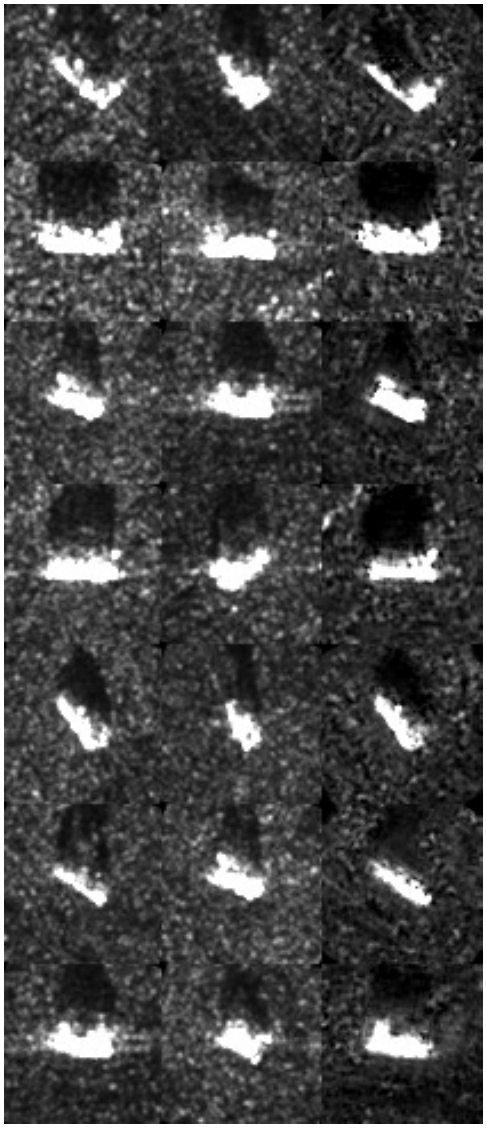


Fig. 9. Real SAR images (left column) compared with the nearest image of another type (middle column) and the corresponding generated images (right column). From top to bottom are BRDM2, BTR60, D7, 2S1, T62, ZIL131, and ZSU234.

images look similar to the real images not only in terms of image style but also the shape and characteristics of targets. Moreover, these AAE-generated images appear to be much sharper than those of previous results that used traditional generative networks [29]. Although the features in the target and shadow areas of the two images are quite similar, their backgrounds are not necessarily the same. This reveals that the generator learned the invariant and intrinsic high-level features of the targets rather than simply low-level pixel values. This demonstrates for the first time that DNNs are able to represent random speckle effects in SAR images.

Fig. 9 compares real images of the seven types of targets with their most similar confused targets and with their corresponding generated images. The purpose is to examine the similarity of the real and generated SAR images visually using the most similar confused target as the standard. For example, BTR60 is most likely to be confused with BRDM2; thus, we examine whether the generated BTR60 looks more similar to the real BTR60 than it does to the real BRDM2, using images with the same orientation. Compared with the confused images in the middle column, the generated images may lack certain features of the real samples, however, they still appear more similar to the real images. This indicates that the generated images reflect the key features that can be used to correctly classify different targets.

To quantitatively evaluate the representation ability of the AAE, we used three indexes to assess the similarity between the generated and real images.

1) Normalized Cross Correlation (NCC): Normalized cross correlation (NCC) is calculated as follows:

$$\mathrm{NCC} = \frac{\sum (A(x,y) - \bar{A})(B(x,y) - \bar{B})}{\sqrt{\sum (A(x,y) - \bar{A})^2 \sum (B(x,y) - \bar{B})^2}} \quad (11)$$

where A and B are two images, and $\bar{A}$ and $\bar{B}$ are the means of A and B. Subtracting the average value of the entire image from each pixel value avoid situations where the similarity index is not robust due to the overall "brightness" or "darkness". The NCC values are $\in [-1, 1]$, and a larger value means that the two images are more similar.

2) Mean Gradient Structural Similarity (MGSM): Mean gradient structural similarity(MGSM) is an advanced version of the Structural Similarity Index (SSIM) [31]. MGSM first calculates the gradient structural similarity (GSM) of each pixel between the two images and then uses the average gradient similarity of all pixels as the similarity index of the two images. By evaluating the gradient similarity rather than the means or standard deviations of the images, MGSM reflects the geometric structure similarity of the two images. The gradient similarity is calculated as follows:

$$\mathrm{GSM}(i) = \frac{2g_A(i)g_B(i) + C}{g_A^2(i) + g_B^2(i) + C}, \quad g(i) = \sqrt{dx_i^2 + dy_i^2} \quad (12)$$

where $g_A(i)$ is the gradient of image A at the ith pixel, and dx_i and dy_i are the horizontal and vertical gradients, respectively, which can be calculated using the Prewitt operator. C is a small constant used to avoid a zero denominator. MGSM $\in [0, 1]$, and two images are highly similar when the MGSM value is close to 1.

3) Normalized Gradient Structural Correlation (NGSC): The NGSC is a combination of the two above similarity indexes. The NGSC first calculates the gradients of two images and then derives the NCC of the two gradient images

$$\mathrm{NGSC} = \frac{\sum \left(g_A(x,y) - \overline{g_A}\right)\left(g_B(x,y) - \overline{g_B}\right)}{\sqrt{\sum \left(g_A(x,y) - \overline{g_A}\right)^2 \sum \left(g_B(x,y) - \overline{g_B}\right)^2}}. \quad (13)$$

These three indexes are able to indicate the scattering similarity between SAR images. For example, NCC, MGSM, and NGSC of real and generated D7 are 0.8597, 0.7497, 0.8233, compared with 0.6871, 0.7138, and 0.6545 of real D7 and its confused target in Fig. 9; And NCC, MGSM, and NGSC of real and generated ZIL131 are 0.8226, 0.7521, 0.8016, but that of real ZIL131 and its confused target in Fig. 9 are 0.5846, 0.7213, and 0.6073. This result agrees with the fact that generated D7 are very similar to the real D7 in terms of

TABLE I
AVERAGE SIMILARITY OF GENERATED AND REAL IMAGES AND AVERAGE INTERCLASS SIMILARITY

	Fake vs. Real on Training			Fake vs. Real on Test			Inter-class on All Dataset		
	NCC	MGSM	NGSC	NCC	MGSM	NGSC	NCC	MGSM	NGSC
BRDM2	0.6828	0.7426	0.6733	0.5394	0.7269	0.5182	0.4062	0.7124	0.4058
BTR60	0.6823	0.7314	0.6550	0.5547	0.7088	0.5279	0.3833	0.6877	0.3755
D7	0.8597	0.7556	0.8080	0.7764	0.7434	0.7230	0.4331	0.7128	0.4431
2S1	0.8338	0.7555	0.8015	0.7187	0.7446	0.6796	0.4469	0.7154	0.4510
T62	0.7819	0.7525	0.7668	0.7117	0.7422	0.6631	0.5086	0.7198	0.4963
ZIL131	0.7935	0.7473	0.7703	0.7261	0.7429	0.6917	0.4724	0.7134	0.4654
ZSU234	0.8913	0.7518	0.8415	0.7534	0.7374	0.6841	0.4770	0.7116	0.4519

scattering features, but the similarity of real ZIL131 and its confused target is low.

Table I shows the average similarity between the real SAR images and the generated images among the training set compared with the average interclass similarity. Note that when calculating the interclass similarity, only sample pairs with the same orientation are considered. Usually, the largest similarity between different target types is obtained when they are acquired under similar azimuths; this fact is validated in Fig. 9. According to Table I, the average NCC, MGSM, and NGSC values between the interclass targets are less than 0.51, 0.72, and 0.50, respectively, while the values of the generated and real images are larger than 0.68, 0.73, and 0.65, respectively. In other words, the generated images are more similar to the real SAR images.

C. Generalizability

Note that the AAE-generated samples are quite similar to the real images. Next, we examine the similarity of the generated samples when generalized toward the test set. Fig. 10 compares example real SAR images (upper row of each panel) with generated images (bottom row of each panel) for the seven types of targets in the test set. In each panel, the orientation of the target increases from left to right; the target chip on the two sides in the red rectangles are from the training set and have a 25° difference in orientation angle. The AAE-generated samples rotate in the same way as the real samples, demonstrating the exceptional ability of the AAE to learn physically plausible mechanisms, such as view aspect angle variation. Additionally, note that the generated images and the real samples have high similarity in terms of the target profile and even scattering features. In other words, the proposed AAE is able to interpolate the training data along physical dimensions such as target orientation. Such nonlinear interpolation of the physical mechanism was not hitherto possible with traditional linear interpolation methods.

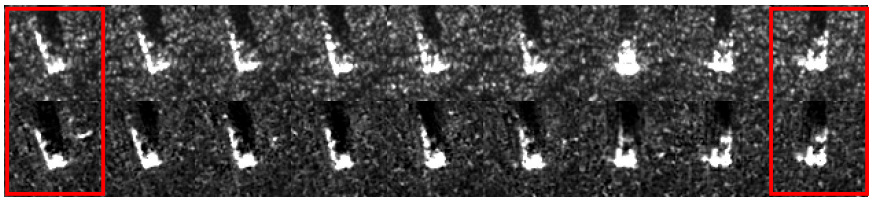
Table I lists the average similarity between the generated and real test samples using the aforementioned three indexes. The average NCC values for the generated and real images range from 0.5394 to 0.7764; the average MGSM values exceed 0.7198 (maximum average interclass similarity) for most types of targets; and the average NGSC values range from 0.5182 to 0.7230. Compared with the training set results, the similarity between the generated images and the test images is somewhat less than that between the generated images and the training samples but it is still larger than the average interclass similarity among the training set. That is, introducing the generated images to the training set can bridge the gap between the training and test sets, which should certainly improve the generalizability of an ATR classification network.

D. FSL Ability

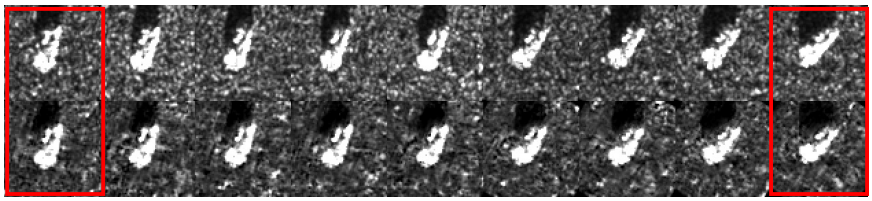
The generated images increase the density of intraclass samples, thus providing a solution to the FSL problem. In this section, we investigate the contribution of the generated images via the proposed AAE to the FSL task.

A definition of FSL was given in (1). Here, we use practical SAR datasets [16] to verify the definition. Fig. 11 plots the average maximum intra and interclass similarity calculated by NCC and MGSM as defined in (11) and (12) compared with the ratio of the number of training sets to the total number of samples. Note that trend of NGSC is similar to that of NCC. From this figure, we can see that as the orientation interval increases, the number of training data decreases; both the average maximum intra and interclass similarity decrease, but the average maximum intraclass similarity decreases faster. Based on our definition, whether a problem belongs to the FSL problem is also related to the similarity index in use; however, both indexes show that when the orientation interval is larger than 12°, it is an FSL task in our case. This result reveals that in FSL, the samples are distributed so sparsely that the average minimum intraclass distance is larger than that of the interclass distance.

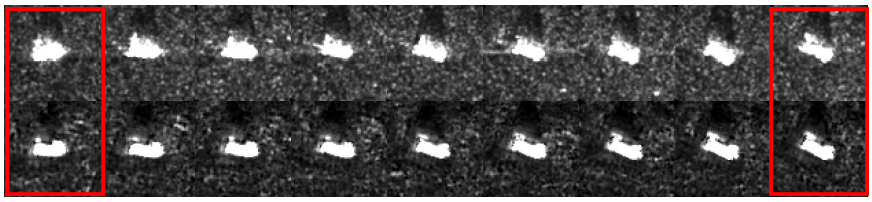
Using the AAE-generated images can increase the distribution density of the training set. In this section, we use 90 real samples to train the proposed AAE and then use the trained AAE to generate 1748 test samples and use them



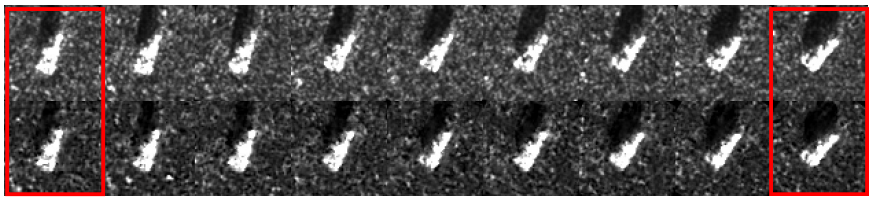
(a)



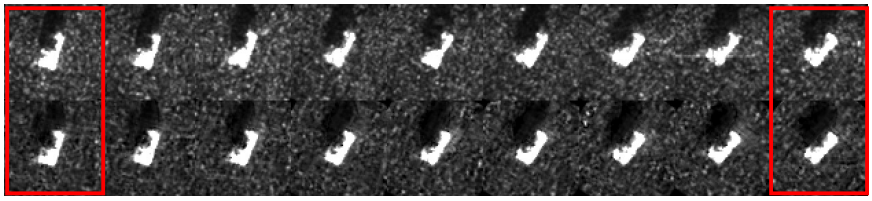
(b)



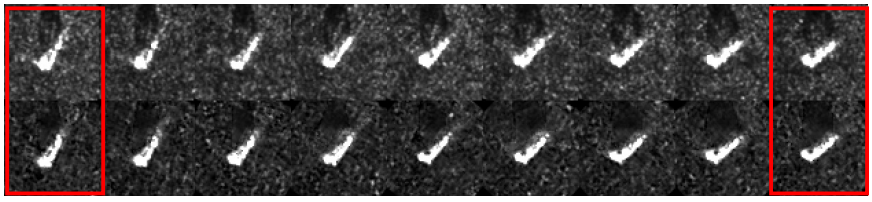
(c)



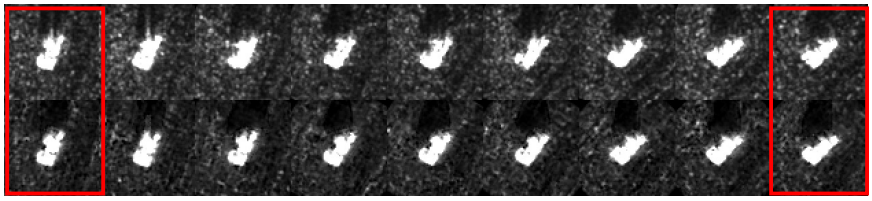
(d)



(e)



(f)



(g)

Fig. 10. Image interpolation along the orientation: two adjacent training images (in the rectangles) and their interpolations. The top and bottom rows are, respectively, the SAR real and generated images of (a) BRDM2, (b) BTR60, (c) D7, (d) 2S1, (e) T62, (f) ZIL131, and (g) ZSU234.

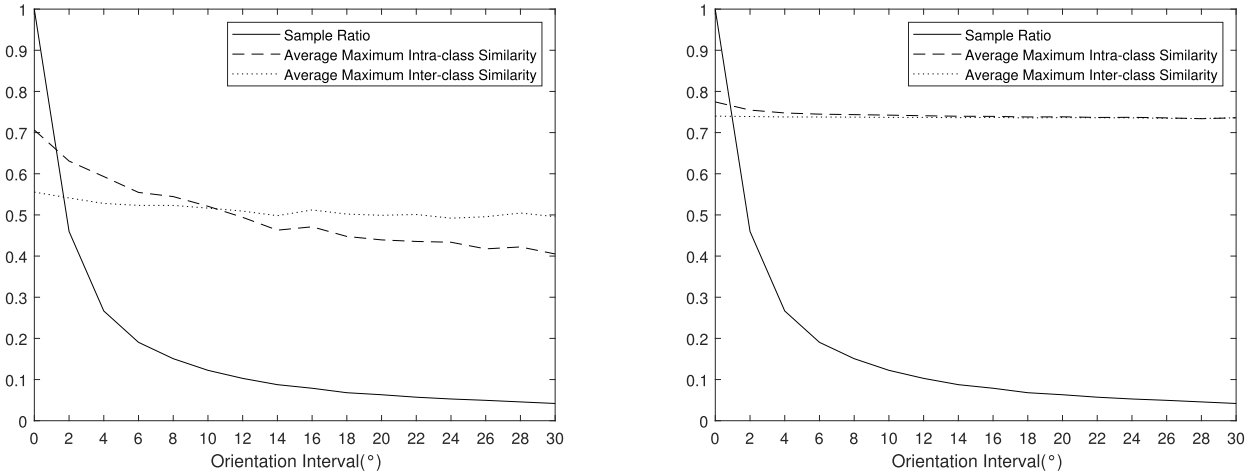


Fig. 11. Average maximum intra and interclass similarity of MSTAR compared to the sample ratio. (a) NCC. (b) MGSM.

TABLE II
AVERAGE MAXIMUM INTRA AND INTERCLASS SIMILARITIES WITH AND WITHOUT IMAGES GENERATED BY THE AAE

	Without AAE			With AAE		
	NCC	MGSM	NGSC	NCC	MGSM	NGSC
Average Maximum Intra-class Similarity	0.5734	0.7325	0.5618	0.9368	0.7843	0.9224
Average Maximum Inter-class Similarity	0.6598	0.7333	0.6435	0.7370	0.7412	0.7045

to augment the training data. Table II compares the average maximum intra and interclass similarities within the training set before and after augmentation using the generated images. After data augmentation, both the average maximum intra and interclass similarities increase. For example, the intraclass NGSC increases from 0.5618 to 0.9224, and the interclass NGSC increases from 0.6435 to 0.7045. However, the increase in the intraclass similarity (0.3606) is larger than that of interclass similarity (0.061). In the augmented data, all the indexes reveal that the average maximum intraclass similarity is larger than the average maximum interclass similarity, reflecting the increased density of the dataset.

Fig. 12 compares the test accuracy of A-ConvNets [10] using the data-augmented training set and of A-ConvNets using the original 90 training samples. The horizontal axis is the minimum orientation interval between the training and test sets. Clearly, both curves fall as the orientation interval increases. However, the test accuracy is improved when using the augmented data. Under the same FSL setting, the classification accuracy of A-ConvNets using augmented data on the test data is 91.19%—3.6% higher than the accuracy when using only the original data (87.59%).

E. Transferability

The trained AAE can be used as an initial representation model to generate other types of targets. In this section, we use the bulk carrier chips from the FUSAR-ship dataset [33] to fine-tune the trained MSTAR representation network. We chose eight different bulk carrier chips with different orientations, as shown in Fig. 13. All the chips are first cropped and resized to 128×128; then, they are rotated based on their orientation and cropped to 88×88 to match the input size of the network. The orientation and semantic map of each SAR image are annotated manually. During training, the trained AAE is fine-tuned with the eight images using a small initial learning rate of 0.0001.

The second and third rows in Fig. 13 show the generated bulk carriers. Although only a few samples were used for training, the trained model is able to grasp the unique features of the bulk carrier target – the box-like hatches – which verifies the transferability of the proposed AAE. Using the pretrained AAE, it takes only approximately 1.6 h to train the bulk carrier generator, compared with the over 6 h needed to train the original AAE. However, note that the trained AAE can be successfully transferred only for isolated targets such as ships and airplanes.

V. DISCUSSION

A. Role of the Semantic Map

In this article, the semantic map is introduced as a strong regularization that restricts the profile of generated images. To evaluate the contribution of the semantic map, Fig. 14

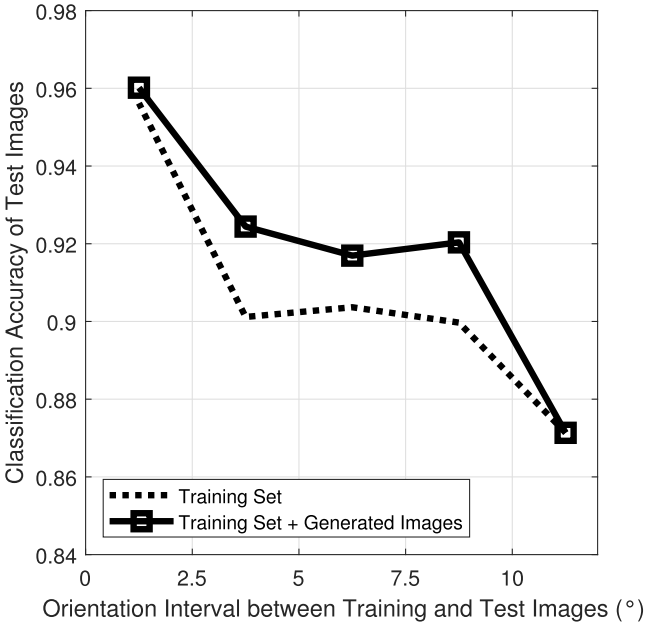


Fig. 12. Average test accuracy versus the smallest orientation difference between the training and test samples when using A-ConvNets with and without AAE-generated augmented data.

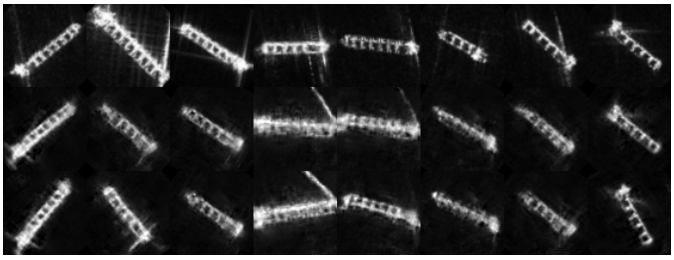


Fig. 13. Fine-tuning the trained AAE network with bulk carriers: 8 real SAR images of bulk carriers (top row), generated bulk carriers (middle row), and generated images from different aspects (bottom row).

compares AAE-generated images with and without the semantic map. For some targets (for example, the second and last columns in Fig. 14), there is no evident difference between the two generated images; for other targets (the fourth and fifth columns, for example), the shadow areas in the images from AAE with the semantic maps agree better with the same areas of the real images. The first two rows in Table III list the average similarity between the real images and the AAE-generated images with and without the use of the semantic maps. These results demonstrate that the semantic map improves the representation ability of the proposed AAE; it increases the average similarity between the real and generated images by 0.001 to 0.0051. Based on this result, we can conclude that the semantic map is able to boost the similarity but to a limited extend. Note that this conclusion is drawn based on the MSTAR dataset. Semantic map may play a more important role in other cases.

In this article, a semantic map consists of three parts, corresponding to the background, shadow, and target regions, respectively. This is because the profiles of target and shadow areas change with orientation angle in significantly

TABLE III
AVERAGE SIMILARITY BETWEEN REAL AND AAE-GENERATED SAMPLES WITH AND WITHOUT SEMANTIC MAPS AND ROTATED CROPPING

	NCC	MGSM	NGSC
Proposed	0.6944	0.7383	0.6502
Without semantic maps	0.6909	0.7373	0.6451
Without rotated Cropping	0.5256	0.7243	0.5016

different ways. In other applications (such as ship generation), the shadow may be ignorable. Besides, one plausible way approach to generating the semantic map is the segmentation of simulated images. According to [5], the simulated SAR targets are similar to real images in terms of profiles but different in the distributions of strong scattering points.

B. Rotation Is Hard to Represent

As mentioned above, neural networks have difficulty representing target orientation due to the nonlinearity and noncontinuity of the rotation operation. Our previous work in [1] showed that the generated test images fail to agree with the real images in terms of target orientations. In this article, we introduce the rotated cropping technique for SAR images to train the AAE, which reduces the high nonlinearity demand of the representation network to some extent.

Fig. 15 compares images generated via the AAE with and without rotated cropping; otherwise, the AAE architectures and other training parameters are identical. The comparison shows that the target orientation in the images generated by the AAE without rotated cropping deviates greatly from the real targets. However, with rotated cropping, the proposed AAE is able to generate correctly oriented target images. As shown in Table III, rotated cropping improves the average similarity between the generated and real images by 0.1688, 0.014, and 0.1486 as evaluated by NCC, MGSM, and NGSC, respectively.

C. Is More Better?

In Section IV-D, we proved that using the AAE representation network can increase the density of the training set, thus contributing to an improved classification performance under the FSL setting. Intuitively, one might predict that when the training set is augmented by even more generated images, A-ConvNets would achieve an even higher test accuracy. Remember that the AAE is controlled by the random vector z, which follows a normal distribution. By inputting many samples of z, AAE can generate an infinite number of samples under the same conditions. These generated images have a similar profile and semantic features but differ in nonessential targets and background details. Thus, the AAE can generate as many images as necessary, which means that the training data can be augmented by an infinite amount.

Fig. 16 compares the test accuracy of A-ConvNets when using many data augmentations. When using 5 and 25 generated images, the test accuracy is 93.02% and 93.36%,

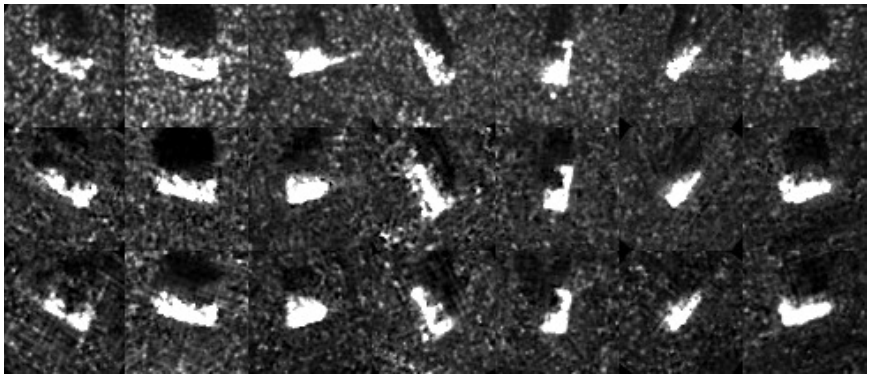


Fig. 14. Comparison of practical SAR images (upper row), with the corresponding generated images via an AAE (middle row) trained with semantic maps, and images generated via an AAE trained without semantic maps (bottom row).

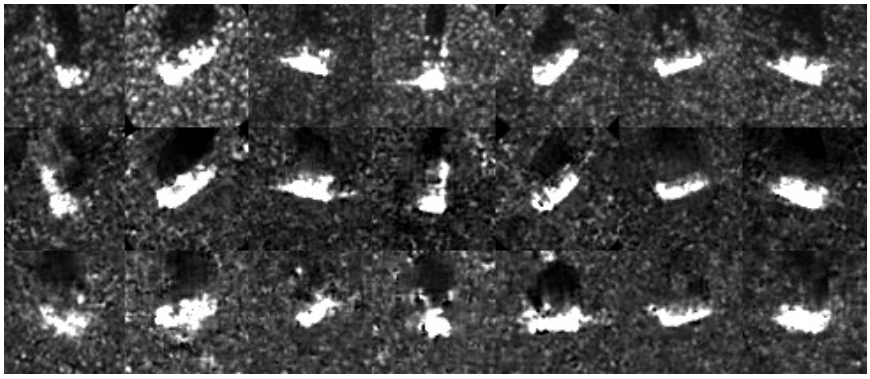


Fig. 15. Practical SAR images (upper row), corresponding generated images via AAE with rotated cropping (middle row), and generated images via AAE with horizontal cropping (bottom row).

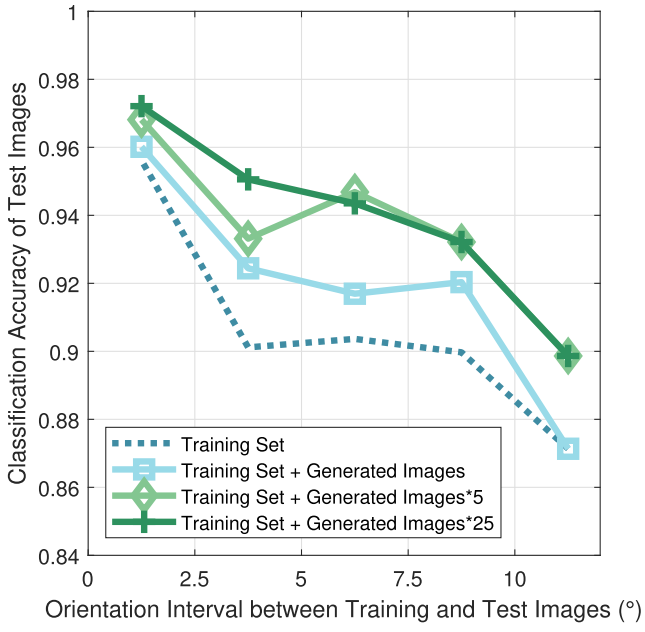


Fig. 16. A-ConvNets average test accuracy versus the least orientation difference between training and test samples when trained with the original training set and with data augmented by 1, 5, and 25 generated images, respectively.

respectively, i.e., the test accuracy increases as the number of augmented images grows. However, when trained with 50 generated images, A-ConvNets achieves a test accuracy of only approximately 87%. We suspect that this result occurs because the training samples are dominated by the massive number of generated images, causing the trained classification network to be biased.

VI. Conclusion

In this article, we proposed an SAR image generation network based on an AAE, which consists of a generator (that also functions as a decoder) that outputs SAR-like images, and a discriminator (encoder), which takes either a real or generated image as input and predicts its type and orientation and whether it is a fake or real SAR image. The two subnetworks are trained adversarially to increase the fidelity of the generated images. Adding generated images to the real SAR images increases the density of SAR images in the sampling space. Thus, the proposed AAE can potentially be applied to FSL tasks. In this study, we introduced rotated cropping to address the challenge of rotation representation in neural networks. The experimental results show that using rotated cropping increases the average similarity of the real and generated images by 0.1688 and that the AAE is able to learn to represent a physically plausible rotation mechanism. The maximum average intraclass similarity of the dataset is increased by at least 0.0518 via AAE, which in turn increases the test accuracy of A-ConvNets by 5.77%. Note that the proposed AAE establishes a framework for SAR image representation and generation. It could be extended to generate SAR images for other purposes in the future.

ACKNOWLEDGMENT

The data generated in this study are available at https://github.com/fudanxu/SAR-Images-Generation.

REFERENCES

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

[2] F. Xu, C. Hu, J. Li, A. Plaza, and M. Datcu, "Special focus on deep learning in remote sensing image processing," *Sci. China Inf. Sci.*, vol. 63, no. 4, pp. 1–2, Apr. 2020.

[3] D. Malmgren-Hansen *et al.*, "Improving SAR automatic target recognition models with transfer learning from simulated data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1484–1488, 2017.

[4] 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. (IGARSS)*, Jul. 2018, pp. 4411–4414.

[5] 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.

[6] C. Kang and C. He, "SAR image classification based on the multi-layer network and transfer learning of mid-level representations," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 1146–1149.

[7] T. Toizumi, K. Sagi, and Y. Senda, "Automatic association between SAR and optical images based on zero-shot learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 17–20.

[8] Z. Huang, Z. Pan, and B. Lei, "Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data," *Remote Sens.*, vol. 9, no. 9, p. 907, Aug. 2017.

[9] Y. Sun, Y. Wang, H. Liu, N. Wang, and J. Wang, "SAR target recognition with limited training data based on angular rotation generative network," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 11, pp. 1928–1932, Nov. 2020.

[10] 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.

[11] F. Zhang, Z. Fu, Y. Zhou, W. Hu, and W. Hong, "Multi-aspect SAR target recognition based on space-fixed and space-varying scattering feature joint learning," *Remote Sens. Lett.*, vol. 10, no. 10, pp. 998–1007, Oct. 2019.

[12] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[13] 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.

[14] C. Zheng, X. Jiang, and X. Liu, "Semi-supervised SAR ATR via multi-discriminator generative adversarial network," *IEEE Sensors J.*, vol. 19, no. 17, pp. 7525–7533, Sep. 2019.

[15] 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, no. 2, pp. 1–38, Feb. 2021.

[16] *The Air Force Moving and Stationary Target Recognition Database*. Accessed: Sep. 1, 2016. [Online]. Available: https://www.sdms.afrl.af.mil/-datasets/mstar/

[17] X. Liu, Y. Qiao, Y. Xiong, Z. Cai, and P. Liu, "Cascade conditional generative adversarial nets for spatial-spectral hyperspectral sample generation," *Sci. China Inf. Sci.*, vol. 63, no. 4, pp. 1–16, Apr. 2020.

[18] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434

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

[20] I. Gulrajani *et al.*, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.

[21] J. Guo *et al.*, "Synthetic aperture radar image synthesis by using generative adversarial nets," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1–5, May 2017.

[22] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 594–611, Apr. 2006.

[23] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Comput. Surveys*, vol. 53, no. 3, pp. 1–34, Jul. 2020.

[24] O. Chapelle and A. Zien, "Semi-supervised classification by low density separation," in *Proc. AISTATS*, 2005, pp. 57–64.

[25] V. Vapnik, *Statistical Learning Theory*. Hoboken, NJ, USA: Wiley, 1998.

[26] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li, "On the continuity of rotation representations in neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5738–5746.

[27] B. Llanas, S. Lantarón, and F. J. Sáinz, "Constructive approximation of discontinuous functions by neural networks," *Neural Process. Lett.*, vol. 27, no. 3, pp. 209–226, Jun. 2008.

[28] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*. [Online]. Available: http://arxiv.org/abs/1710.10196

[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.

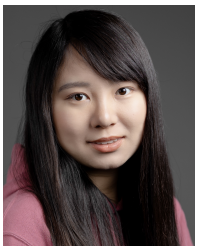
[30] Y. Shen *et al.*, "Invertible zero-shot recognition flows," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 614–631.

[31] 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.

[32] J. Donahue and K. Simonyan, "Large scale adversarial representation learning," 2019, *arXiv:1907.02544*. [Online]. Available: http://arxiv.org/abs/1907.02544

[33] X. Hou, W. Ao, Q. Song, J. Lai, H. Wang, and F. Xu, "FUSAR-ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition," *Sci. China Inf. Sci.*, vol. 63, no. 4, pp. 1–19, Apr. 2020.

[34] X. Zhu *et al.*, "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, to be published, doi: 10.1109/MGRS.2020.3046356.



Qian Song (Member, IEEE) received the B.E. degree (Hons.) from the School of Information Science and Technology, East China Normal University, Shanghai, China, in 2015, and the Ph.D. degree (Hons.) from Fudan University, Shanghai, in 2020.

She is a Post-Doctoral Fellow with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany. Her research interests include advanced deep learning technologies and their applications in synthetic aperture radar image interpretation.

Dr. Song has been awarded as the International Union of Radio Science (URSI) Young Scientist Award in 2020.

Feng Xu (Senior Member, IEEE) received the B.E. degree (Hons.) in information engineering from Southeast University, Nanjing, China, in 2003, and the Ph.D. degree (Hons.) in electronic engineering from Fudan University, Shanghai, China, in 2008.

From 2008 to 2010, he was a Post-Doctoral Fellow with the National Oceanic and Atmospheric Administration (NOAA) Center for Satellite Application and Research, Camp Springs, MD, USA. From 2010 to 2013, he worked with Intelligent Automation Inc., Rockville, MD, USA, and with the NASA Goddard Space Flight Center, Greenbelt, MD, USA, as a Research Scientist. In 2012, he was selected into China's Global Experts Recruitment Program and subsequently returned to Fudan University, Shanghai, in 2013, where he is a Professor with the School of Information Science and Technology and the Vice Director of the Key Laboratory for Information Science of Electromagnetic Waves (MoE). He has published over 40 articles in peer-reviewed journals, coauthored 2 books, and 2 patents, as well as publishing many conference papers. His research interests include electromagnetic scattering modeling, SAR information retrieval, and radar system development.

Dr. Xu received the Second-Class National Nature Science Award of China in 2011 among other honors. He was a recipient of the Early Career Award of the IEEE Geoscience and Remote Sensing Society (GRSS) in 2014 and the SUMMA Graduate Fellowship in the advanced electromagnetics area in 2007. He is the Founding Chair of the IEEE GRSS Shanghai Chapter. He serves as an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Xiao Xiang Zhu (Fellow, IEEE) received the M.Sc., Dr.Ing., and Habilitation degrees in signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

She is the Professor of data science in earth observation with TUM and the German Aerospace Center (DLR), the Head of the Department "EO Data Science" with the DLR's Earth Observation Center, and the Head of the Helmholtz Young Investigator Group "SiPEO" at DLR and TUM. She was a Guest Scientist or a Visiting Professor with the Italian National Research Council, Naples, Italy, in 2009, Fudan University, Shanghai, China, in 2014, The University of Tokyo, Tokyo, Japan, in 2015, and the University of California at Los Angeles, Los Angeles, CA, USA, in 2016, respectively. Her main research interests are remote sensing and earth observation, signal processing, machine learning and data science, with a special application focus on global urban mapping.

Dr. Zhu is a member of young academy (Junge Akademie/Junges Kolleg) with the Berlin-Brandenburg Academy of Sciences and Humanities, the German National Academy of Sciences Leopoldina, and the Bavarian Academy of Sciences and Humanities. She is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Ya-Qiu Jin (Life Fellow, IEEE) received the bachelor's degree in electrical engineering and computer science from Peking University, Beijing, China, in 1970, and the M.S., E.E., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1982, 1983, and 1985, respectively.

In 1985, he joined Atmospheric Environmental Research, Inc., Cambridge, as a Research Scientist. From 1986 to 1987, he was a Research Associate Fellow with the City University of New York, New York, NY, USA. In 1993, he joined the University of York, York, NY, U.K., as a Visiting Professor, sponsored by the U.K. Royal Society. He is the Te-Pin Professor and the Director of the Key Laboratory for Information Science of Electromagnetic Waves (MoE), Fudan University, Shanghai, China. He has authored more than 720 articles in refereed journals and conference proceedings and 14 books, including *Polarimetric Scattering and SAR Information Retrieval* (Wiley and IEEE, 2013), *Theory and Approach of Information Retrievals from Electromagnetic Scattering and Remote Sensing* (Springer, 2005), and *Electromagnetic Scattering Modeling for Quantitative Remote Sensing* (World Scientific, 1994). His research interests include electromagnetic scattering and radiative transfer in complex natural media, microwave satellite-borne remote sensing, theoretical modeling, information retrieval and applications for Earth terrain and planetary surfaces, and computational electromagnetics.

Dr. Jin was awarded a Senior Research Associateship in the National Oceanic and Atmospheric Administration (NOAA)/NESDIS by the USA National Research Council in 1996. He was a recipient of the IEEE GRSS Distinguished Achievement Award in 2015, the IEEE GRSS Education Award in 2010, the China National Science Prize in 1993 and 2011, the Shanghai Sci/Tech Gong-Cheng Award in 2015, and the First-Place MoE Science Prizes in 1992, 1996, and 2009, among many other prizes. He is an Academician of the Chinese Academy of Sciences, a fellow of the World Academy of Sciences and the International Academy of Astronautics. He was a Co-Chair of TPC for IGARSS2011 in Vancouver, BC, Canada, and the Co-General Chair of IGARSS2016 in Beijing, China. He was an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2005 to 2012, a member of the IEEE GRSS ADCOM, and the Chair of the IEEE Fellow Evaluation of GRSS from 2009 to 2011. He is an IEEE GRSS Distinguished Speaker and an Associate Editor of IEEE ACCESS.