



## Article

# PeaceGAN: A GAN-Based Multi-Task Learning Method for SAR Target Image Generation with a Pose Estimator and an Auxiliary Classifier

Jihyong Oh and Munchurl Kim \*

Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea; jhoh94@kaist.ac.kr

\* Correspondence: mkimee@kaist.ac.kr

**Abstract:** Although generative adversarial networks (GANs) are successfully applied to diverse fields, training GANs on synthetic aperture radar (SAR) data is a challenging task due to speckle noise. On the one hand, in a learning perspective of human perception, it is natural to learn a task by using information from multiple sources. However, in the previous GAN works on SAR image generation, information on target classes has only been used. Due to the backscattering characteristics of SAR signals, the structures of SAR images are strongly dependent on their pose angles. Nevertheless, the pose angle information has not been incorporated into GAN models for SAR images. In this paper, we propose a novel GAN-based multi-task learning (MTL) method for SAR target image generation, called PeaceGAN, that has two additional structures, a pose estimator and an auxiliary classifier, at the side of its discriminator in order to effectively combine the pose and class information via MTL. Extensive experiments showed that the proposed MTL framework can help the PeaceGAN's generator effectively learn the distributions of SAR images so that it can better generate the SAR target images more faithfully at intended pose angles for desired target classes in comparison with the recent state-of-the-art methods.

**Keywords:** synthetic aperture radar; automatic target recognition; pose angle estimation; deep learning; convolutional neural networks; multi-task learning; generative adversarial networks



**Citation:** Oh, J.; Kim, M. PeaceGAN: A GAN-Based Multi-Task Learning Method for SAR Target Image Generation with a Pose Estimator and an Auxiliary Classifier. *Remote Sens.* **2021**, *13*, 3939. <https://doi.org/10.3390/rs13193939>

Academic Editors: Addabbo Pia, Luca Pallotta, Christos Ilioudis and Qi Wang

Received: 17 August 2021  
Accepted: 29 September 2021  
Published: 1 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Synthetic aperture radar (SAR) is commonly utilized for surveillance systems [1–4]. Since SAR has a compelling characteristic of a penetration, SAR images can be easily obtained, regardless of any weather condition, whether the time is night or daytime or the weather is sunny or cloudy, unlike an optical remote sensing. On the other hand, SAR images generally have serious speckle noise all over the image due to the backscattering of electromagnetic waves [5], therefore making both human and machine learning algorithms hard to interpret semantic features of the SAR images [6,7]. However, due to the recent advent of deep learning methods, deep convolutional neural networks (CNNs) have been widely used for many SAR tasks such as recognition of SAR targets [8–12] and optical image classification [13]. However, it is hard to train the CNNs for SAR-related tasks due to the lack of available SAR images that should be obtained by radar attached to air vehicles and labeled manually with considerable time consumption [8,14]. For this reason, there is a need for generative models that can generate abundant SAR data such as “Big Data” for diverse SAR tasks.

Following the first proposal of the generative adversarial networks (GANs) [15], many variants of GANs have shown generative power of GANs for both natural and synthetic images by attaining the similar attributes of their respective original data set [16–26]. Therefore, GAN-based modeling has been one of the most popular generative models [20]. GANs are usually composed of two elements: a generator (G) that aims to learn a mapping function from some probabilistic input distributions to a real data distribution, and a

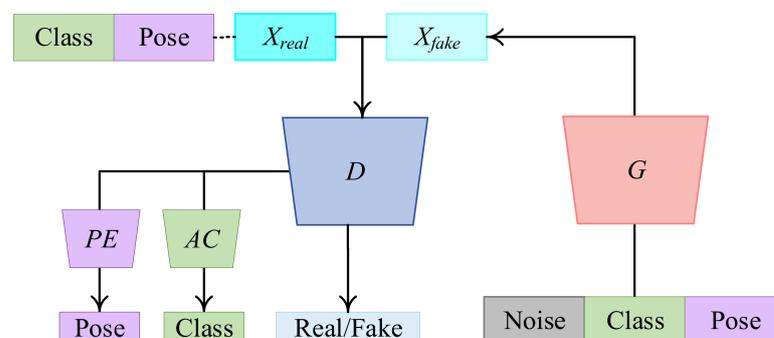
discriminator (D) that distinguishes whether an input sample comes from the real data distribution or a generative one [15]. At the end, ideally after both players (the generator and discriminator) minimize their own losses, the training stage approaches the Nash equilibrium where the generator can produce images that can hardly be distinguishable by the discriminator as to whether the samples are real or fake. It is also known that the Nash equilibrium of GANs can be reached by minimizing a statistical divergence between the real data distribution and the generative one as well [15,27,28]. After deep convolutional GAN (DCGAN) was first successfully trained such that both the discriminator and the generator were trained in a well stabilized manner by several deep learning methods, most GANs have begun to adopt the DCGAN architecture as their base structure [29].

Similar to other fields, GANs have also recently been applied to SAR-related tasks. For example, Guo et al. [30] first utilized GANs for generating SAR target images based on the basic DCGAN, but they reported that the quality of generated images from a generator were often deteriorated by the so-called mode collapse that frequently occurred due to the speckle noise. To prevent those phenomena, they added a pre-processing step, called clutter normalization, for reducing an influence of clutter's speckle noise to boost target recognition performance. On the other hand, Zheng et al. [31] utilized multiple discriminator-based GAN model (MGAN) to generate SAR target images without any additional pre-processing by incorporating a semi-supervised learning method. They realized that inherent target-class features from unlabeled SAR target images and label smoothing regularization (LSR) would help improve the generating stability and the quality of generated SAR target images.

On the other hand, either pose angle estimation or regression of objects via both CNNs and GANs have been treated as an important auxiliary task to boost the main tasks of classification and generation networks for non-SAR data [32–35]. However, research related to SAR tasks tend to only focus on an intensity information of SAR images and have not taken any pose angle information into account for their main tasks.

Multi-task learning (MTL) that optimizes several loss functions jointly to obtain information from multiple tasks is well known to be effective in performing its main task in contrast to a single task learning by benefiting from a regularization effect on the usage of multiple objective functions [36,37]. For example, training a model to predict latent variables in auxiliary classifiers can be hint for a given main task [37]. In addition, MTL makes the model improve a generalization by collecting joint features from related tasks [36].

To tackle all above issues synthetically, we first present a GAN-based MTL method, called PeaceGAN, for SAR target image generation that can estimate their pose angles by its pose estimator and can learn their target-class features by its auxiliary classifier simultaneously, both being located at the side of the discriminator part. Figure 1 shows the overall structure of our proposed PeaceGAN.



**Figure 1.** Overall architecture of our proposed PeaceGAN. The model is composed of four components: a generator (G), a discriminator (D), a pose estimator (PE), and an auxiliary classifier (AC). After training finishes, the generator can produce abundant SAR target images with intended target classes at desired pose angles.

To the best of our knowledge, the PeaceGAN is the first approach to jointly learn both the pose angle and the target class information at the same time so that the main task of SAR target image generation via GAN training becomes more stable, thus improving the quality of the resultant generated SAR target images. The contributions of our works are summarized as follows:

- We first propose a novel GAN-based generative model, called PeaceGAN, that is jointly trained in an end-to-end manner with multi-task learning of estimating both pose angle and target class information of SAR target images. The proposed PeaceGAN explicitly disentangles both pose angles and target classes in learning the distributions of SAR target images, leading to the enlarged diversity and the increased quality of generated SAR target images. Finally, the proposed PeaceGAN can generate SAR target images with both desired pose angles and target classes compared to the recent state-of-the-art methods.
- Moreover, we introduce three indirect evaluation methods as standard metrics to measure (i) the adequacy of generated SAR target images as training data, (ii) the fidelities of them in target class perspective, and (iii) their exactitude in pose angle perspective, thus evaluating the generator's ability in producing the generation diversity of SAR target images at various pose angles for different target classes.

The remainder of this paper is organized as follows: Section 2 briefly reviews related works on GAN-based methods for SAR target image generation, multi-task learning for GANs, and evaluation metrics for GANs; Section 3 introduces the proposed GAN-based MTL method for SAR target image generation, called PeaceGAN; Section 4 describes the proposed indirect evaluation methods for adequacy and quality of the generated SAR target images; Section 5 shows experimental results and analyses to demonstrate the effectiveness of the proposed PeaceGAN for SAR target image generation; and Section 6 concludes this paper.

## 2. Related Works

### 2.1. GAN-Based Methods for SAR Target Image Generation

Since training the GANs is generally to solve a minmax problem on the parameters of deep neural networks, it is severely hard in practice for stable learning, and therefore it often fails to appropriately train the GAN-based networks by facing with *mode collapse* problems [19,28,38]. Moreover, both the standard training methods and the structures of GANs have no significant advantages, neither in SAR image classification nor generation [39]. The main reason is the characteristic of SAR images that have the highest spatial resolution collected by X-band SAR with a resolution of  $0.3 \times 0.3$  m, far lower than those of the optical images. The lower resolution makes both networks and human experience a subtler distinction between different target classes and pose angles of SAR target images. Moreover, speckle noise from the interference of reflected wave at the transducer aperture makes it even harder for the GANs to be stably trained.

Zheng et al. [31] focused on improving SAR target recognition performance of CNN-based framework with semi-supervised GAN learning using multiple discriminators (MGAN) and LSR. The one generator is trained against the feedback aggregated from all the multi-discriminators, which leads the GAN training framework to be more stable. The MGAN also adopts LSR on unlabeled images from the generator in a semi-supervised manner to reduce the confidence of the CNN classifier in obtaining effective supplement; however, multiple discriminators have to be employed with high computation complexities. Moreover, the MGAN cannot even directly fuse pose angle information of SAR images into GAN training. Therefore, the MGAN cannot generate SAR target images at desired pose angles of the intended target classes. However, since our proposed PeaceGAN can disentangle both the pose angle and target class information simultaneously, it can generate SAR target images at any given pose angle in a controlled manner, which can provide a high flexibility of generating the SAR target images.

## 2.2. Multi-Task Learning for GAN-Based Image Synthesis

The MTL for deep learning is motivated from a biological view of human learning. That is, a human naturally learns a main task by utilizing related tasks as additional information. The harder the main task is, the greater importance of MTL in improving both the performance and the generalization of networks [36]. Here, the GAN training is also complex and hard in practice. Tran et al. [35] tried to apply MTL into a GAN training framework for an image synthesis and proposed a disentangled representation GAN, called DRGAN, that adopts an encoder–decoder structure for its generator by providing pose codes of face variations to the decoder and forces its discriminator to perform pose estimation. The class labeling method of DRGAN is similar to a semi-supervised GAN [40] that has  $C$ -class nodes for classification and one additional node for the real/fake decision at the output of  $D$ . However, DRGAN is based on an auto-encoder structure for image synthesis, which is totally different from our PeaceGAN. This is because, while the DRGAN is an image-to-image translation network that takes a face image to be rotated into a desired pose angle for pose-invariant face recognition, our PeaceGAN takes noise input with both an intended pose angle and a target class label and does not translate but generates a corresponding SAR target image of the class at the desired pose angle.

On the other hand, the pose angle information of SAR images has been utilized as *prior* for target recognition, which helps decrease computation complexity and improve target recognition performance [12,41–44]. Moreover, the backscattered intensities of the same target class SAR images are differently observed at different pose angles due to the back-scattering characteristics of speckles [12,45,46]. Therefore, it is essential to consider the pose angle information of SAR images for GAN learning. Motivated by this, we propose our PeaceGAN, which can be effectively trained and generate SAR target images of different classes at intended pose angles. This can be accomplished by training the PeaceGAN for MTL with a GANs loss (adversarial loss), a pose angle loss, and a target class loss (cross entropy loss).

## 2.3. Evaluation Metrics for GANs

One of the GAN applications is to generate rich data for data augmentation to train deep neural networks. Therefore, it is essential to evaluate the adequacy of generated data as training and test data for intended applications such as classification (target recognition for SAR images in our case). The Frechet inception distance (FID) is most widely used as a metric to evaluate the generated data during GAN training where intra-class mode dropping can be effectively detected [20]. However, this quantitative metric is well suited for only optical images, not SAR images, because it utilizes the feature maps of the pre-trained InceptionNet [47] using optical images [20,48]. The images generated by trained GANs are often evaluated via a user study by using the Amazon Mechanical Turk (MTurk) [49], which is also not suitable for SAR images because the characteristics of SAR images are very different from those of optical images and their interpretations require the expertise in SAR, which is very costly and time-consuming.

On the other hand, there are indirect evaluation methodologies by using pre-trained deep neural networks, which are effective in checking the adequacy of generated data for intended target applications. Choi et al. [50] trained a facial expression classifier with the Radboud Faces Database, yielding a near-perfect accuracy of 99.55%. They used the pre-trained classifier for quantitative evaluation of their proposed StarGAN [50]. Guo et al. [30] adopted a similar methodology by using the high accuracy of a pretrained target classification model to evaluate a GAN result for SAR target image in the target class-perspective. Intuitively, this kind of indirect evaluation method can be used to judge the suitability of the generated images for the case of either a lower classification error or a larger accuracy rate, that is, whether they can be considered *realistic* with *class-distinguishable* features or not. Therefore, we extend this indirect methodology for two of our quantitative evaluations by *pre-training* both classifier and pose estimator to check the adequacy of the generated SAR target images in terms of both target class- and pose

angle-distinguishable features. Furthermore, another introduced indirect method for evaluating adequacy *as training data* for SAR target recognition using a simple classifier is also described in Section 4 in detail.

#### 2.4. CNN-Based Methods for High-Level SAR Target Tasks

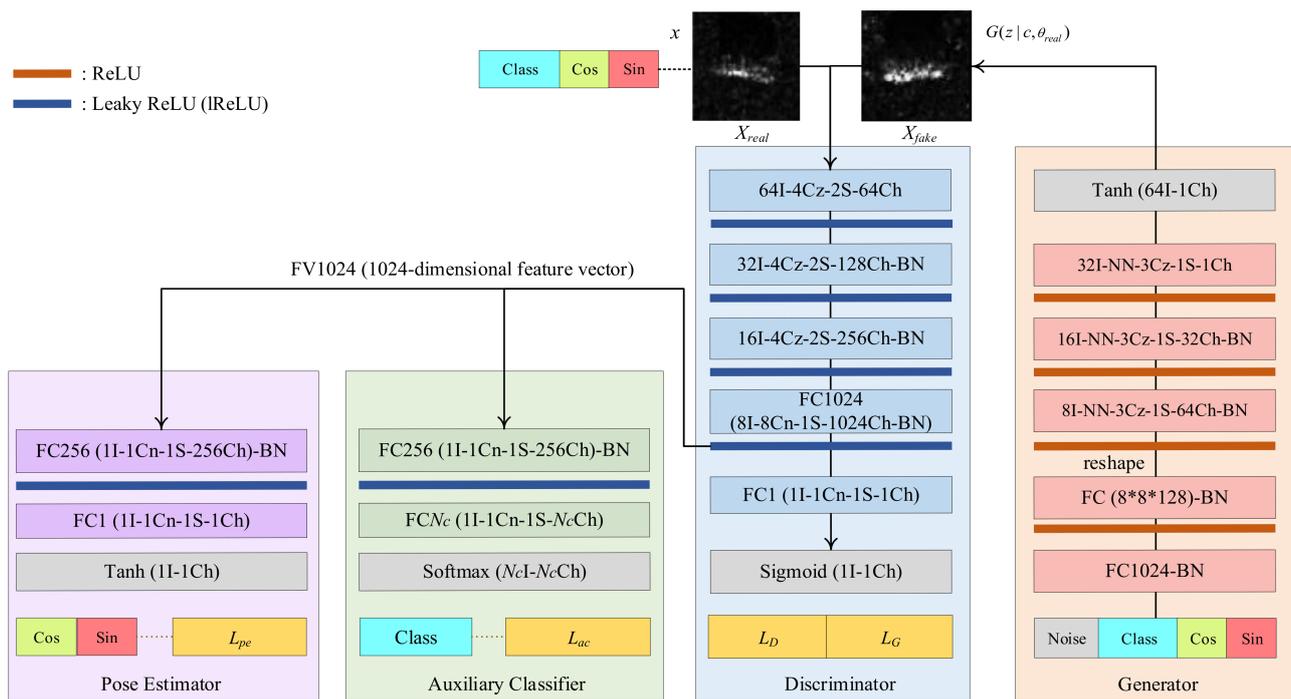
Recently, CNN-based methods for high-level SAR-related tasks, such as recognition, detection, and segmentation, have also shown promising results. Amrani et al. [51] adopted both YOLOv4 [52] fine-tuned with SAR target images and deep CNN to reduce speckle noise. They also proposed multi-canonical correlation analysis (MCCA) to adaptively choose and combine CNN features effectively, which has led their method to achieve state-of-the-art accuracy on SAR target recognition tasks. Feng et al. [53] utilized modified A-ConvNets [11] with two-decoupled head branches to semantically segment SAR vehicle images. They also utilized the ASC (attribute scattering center) model to supplement the lack of annotation for the deep learning-based segmentation models. Their model achieved the best performances in both segmentation and classification for SAR target images. Yang et al. [54] proposed a one-stage object detection framework based on both RetinaNet [55] and a rotatable bounding box (RBox) to solve both the problems of feature misalignment in scales and unbalanced distribution of positive samples that were not handled in previous SAR ship detection methods. Their method achieved the highest average precision (AP) compared with other one-stage RBox-based state-of-the-art methods. Geng et al. [56] proposed a two-stage SAR ship detection method by first segmenting target candidates for detection and by then adopting a lightweight CNN to finally detect the ship targets. They utilized Grad-CAM [57] to analyze attentive pixels in detail by visualizing the model's features. Yue et al. [58] proposed a semi-supervised CNN model to handle insufficient labeled SAR data problems for SAR automatic target recognition (ATR). Their method proposed a loss function based on a scatter matrix that is calculated by the linear discriminant analysis (LDA) method for target class probabilities, showing higher recognition performances than other semi-supervised methods.

### 3. Proposed PeaceGAN—A GAN-Based MTL Method for SAR Target Image Generation

In this section, we explain in detail our proposed GAN-based MTL method for SAR target image generation, called PeaceGAN, with a pose estimator and an auxiliary classifier. The PeaceGAN simultaneously learns both the pose angles by a pose estimator and the target-class features by an auxiliary classifier during training for SAR target image generation. By doing so, the PeaceGAN can successfully generate the SAR target images of various classes with high fidelity at any intended pose angle.

#### 3.1. Overall Structure of Proposed PeaceGAN

An overall structure of the PeaceGAN is shown in Figure 2. The proposed PeaceGAN is basically composed of a discriminator and a generator. In addition, the discriminator additionally has two newly introduced structures: a pose estimator and an auxiliary classifier for stable generation of SAR target images. We adopted several guidelines of DCGAN [29] to design our proposed PeaceGAN, which have been popularly applied to most settings of GAN training [29]. The architecture of our PeaceGAN is partially based on DCGAN that (i) replaces all the pooling layers with strided convolutions for D; (ii) utilizes the batch normalization (BN) [59] in both D and G; (iii) assigns ReLU [60] activation functions to all layers of G except for the output layer where Tanh is used; and (iv) uses Leaky ReLU activation functions [61], called lReLU, with a slope size of 0.2 for all layers in the discriminator [29]. In the next sub-sections, we describe all components of the PeaceGAN in detail.



**Figure 2.** An overall structure of PeaceGAN. PeaceGAN is in essence composed of a discriminator and generator. The generator takes a concatenated input vector of a random noise vector, a target class vector, and a pose angle vector. In addition, the discriminator of PeaceGAN additionally has two newly introduced structures: (i) a pose estimator that is trained to estimate the pose angle values of SAR images in terms of both cosine and sine functions, and (ii) an auxiliary classifier that is trained to determine which target classes the SAR images belong to. Note that a convolution layer with  $vI-wCz-xS-yCh$  indicates having a  $v \times v$ -sized input, a  $w \times w$ -sized convolution filter with zero padding, an  $x \times x$ -sized stride, and  $y$  output channels. In addition,  $wCn$  indicates a  $w \times w$ -sized convolution filter without zero padding, NN means nearest-neighborhood upsampling with scale 2, BN means batch normalization, and  $FCy$  means a fully connected layer with output channel size  $y$ .

### 3.1.1. Generator of PeaceGAN

The generator of the PeaceGAN takes a concatenated input vector that is composed of three vectors: a random noise vector, a target class vector, and a pose angle vector. The random noise vector  $z$  is an  $N_z$ -dimensional vector (noted as “Noise” in Figure 2) whose elements are sampled from  $P_z(z)$ , which is a random uniform distribution in the range of  $[-1, 1]$ . The target class vector  $c$  is an  $N_c$ -dimensional one-hot vector with  $N_c$  types of target classes, having the value “1” at the corresponding target class location and the values “0” at the others, which is indicated as “Class” at the bottom of the generator in Figure 2. Lastly, the pose angle vector is a  $N_p$  ( $=2$ )-dimensional vector where its first and second elements are the cosine and sine component values, respectively, for a given SAR target’s pose angle  $\theta_{real}$  in the range of  $[0, 360]$  at which a desired SAR target image is supposed to be generated. The generator utilizes a nearest neighborhood upsampling with scale 2 followed by a  $3 \times 3$ -sized zero padding convolution filter with a  $1 \times 1$ -sized stride, indicated as “NN-3Cz-1S” in Figure 2, called a NNConv to upsample feature maps. The generator has two fully connected (FC) layers and three NNConvs to finally generate a  $64 \times 64$  fake SAR target image in the range of  $[-1, 1]$  at its output layer.

In sum, the generator of the PeaceGAN can learn a meaningful mapping function from the synthesized vector with both target class and pose angle of the SAR target image and can generate the SAR target images of high fidelity for desired classes at intended pose angles when the training is successfully conducted.

### 3.1.2. Discriminator of PeaceGAN

The discriminator of the PeaceGAN takes two types of input: one is a real SAR target image and the other is a fake SAR target image generated by the generator, alternatively during the training. Both types of SAR target images are of a  $64 \times 64$  size with their values in the range of  $[-1, 1]$ . These inputs are sequentially decreased in spatial dimensions by strided convolution filters without any pooling layer and are increased in channel dimensions to extract useful features. The units of lReLU (Leaky ReLU) are used in the discriminator according to the guidelines of DCGAN [29]. The last two layers of the discriminator, denoted as FC1024 and FC1 in Figure 2, are the fully connected layers with the output channel sizes of 1024 and 1, respectively. Both FC1024 and FC1 are implemented by  $8 \times 8$  and  $1 \times 1$  convolution filters, respectively. At the last layer of the discriminator (after “FC1” denoted as in Figure 2), a sigmoid function yields a scalar output value in the range  $[0, 1]$ , which is used in the adversarial losses, denoted as “ $L_D$ ” and “ $L_G$ ” in Figure 2, that will be described in Section 3.2 in detail.

For the purpose of SAR target image generation by jointly learning both pose angles and target classes of SAR target images, we adopted two additional structures: a pose estimator and an auxiliary classifier, both of which are located at the side of the discriminator as shown in Figure 2. The pose estimator and auxiliary classifier allow for the discriminator not only to distinguish whether the input SAR target images are real or fake, but also to classify the target classes and to estimate pose angles simultaneously via the hard parameter sharing [36] of an MTL framework.

- (a) Pose Estimator (PE). As shown in Figure 2, the pose estimator takes as input a 1024-dimensional feature vector (FV1024), which is the output of the FC1024 layer of the discriminator. The FV1024 input is passed to the FC256 layer followed by the FC1 layer, which yields a scalar output, called  $T_{pe}$ , with the value range of  $[-1, 1]$  after *Tanh*. For a pose estimator loss (denoted as  $L_{pe}$  in Figure 2), we first calculate  $\theta_{pe} = (T_{pe} + 1) \times 180$ , which can be viewed as an appropriate pose angle value if the network is well trained with an adequate loss. Then,  $\theta_{pe}$  can be converted into the cosine and sine values, each of which is denoted as “Cos” and “Sin”, respectively, at the last part (bottom) of the pose estimator in Figure 2.  $L_{pe}$  is described in Section 3.2 in detail. The newly introduced pose estimator for SAR target image generation encourages both the discriminator and the generator to learn the pose-angle features of SAR target images.
- (b) Auxiliary Classifier (AC). Similar to the pose estimator, the auxiliary classifier passes the FV1024 as an input to the FC256 layer followed by the FC  $N_c$  layer, yielding an  $N_c$ -dimensional output, denoted as  $S_{ac}$ , after a *softmax* activation function [62] for  $N_c$  classes of SAR target images. For an auxiliary classifier loss, denoted as  $L_{ac}$  in Figure 2, we use the cross-entropy loss that calculates the difference between the true class probability and the estimated output probability stochastically.  $L_{ac}$  is described in Section 3.2 in detail. The auxiliary classifier for SAR target image generation guides both the discriminator and the generator to learn the scattering characteristics of target-class features of SAR target images.
- (c) Spectral Normalization (SN). Miyato et al. [18] originally applied the spectral normalization (SN) to the discriminator only for stable training of GAN. After that, Zhang et al. [17] argued that applying the SN to the generator also benefits from the GAN training with more stable learning. However, in our SAR target image generation, applying the SN only to the discriminator is almost the same as applying it to both the discriminator and the generator. Therefore, for simplicity, the SN is only applied in the discriminator, including both the pose estimator and the auxiliary classifier with one round of the power iteration method that is known for a sufficient approximation of SN [18].

### 3.2. Combined MTL Loss for PeaceGAN

The PeaceGAN uses the combined loss of the adversarial losses ( $L_D$  and  $L_G$ ), the pose estimator loss ( $L_{pe}$ ), and the auxiliary classifier loss ( $L_{ac}$ ) for training. Here, we describe these three kinds of losses in detail in the following subsections.

#### 3.2.1. Adversarial Loss

In the standard GAN [15], the input of the generator  $G$  is generally a random noise vector  $z$  with a probability distribution  $P_z(z)$ , and  $G$  generates a fake image  $G(z)$ , attempting to follow the target data distribution  $P_{data}$  as closely as possible. The discriminator  $D$  alternatively takes two kinds of images: one is a real image  $x(=X_{real})$  from  $P_{data}(x)$  and the other is a fake image  $G(z)(=X_{fake})$ .  $D$  is trained to distinguish between them well. Finally, both  $G$  and  $D$  try to solve an adversarial minmax optimization problem with a value function  $V(G, D)$  as follows:

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

It is known that if both  $G$  and  $D$  have a sufficient learning capacity, they finally reach the point where neither can improve anymore because  $P_{data} = P_{G(z)}$  [15], which is called the Nash equilibrium where  $D$  cannot make a distinction between real images and fake images. It means that we have  $D(x) = D(G(z)) = 0.5$  [15]. In addition, Equation (1) can be solved in the form of a non-saturating (NS) loss as follows [15]:

$$L_D = -E_{x \sim P_{data}(x)} [\log D(x)] - E_{z \sim P_z(z)} [\log(1 - D(G(z)))], \quad (2)$$

$$L_G = -E_{z \sim P_z(z)} [\log D(G(z))]. \quad (3)$$

The final output of  $D$  is usually a sigmoid function value so that a sigmoid cross-entropy loss is generally utilized.

For our proposed PeaceGAN, since we adopted an MTL framework for GAN training with the auxiliary classifier and the pose estimator, it has two additional components: pose angle value  $\theta_{real}$  and target class  $c$ , concatenated to the noise vector  $z$  at the input part of the generator that generates a fake SAR target image  $G(z|c, \theta_{real})$ . The real images  $x$  from  $P_{data}(x)$  and fake images  $G(z|c, \theta_{real})$  are alternatively fed into the discriminator  $D$ . Additionally, we adopted the WGAN-GP loss [38] as the adversarial loss for the PeaceGAN for good training, which can be expressed in our case as:

$$L_D = -E_{x \sim P_{data}(x)} [D(x|c, \theta_{real})] + E_{z \sim P_z(z)} [D(G(z|c, \theta_{real}))] + \lambda_{gp} E_{\hat{x} \sim P_{\hat{x}}(\hat{x})} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2], \quad (4)$$

$$L_G = -E_{z \sim P_z(z)} [D(G(z|c, \theta_{real}))], \quad (5)$$

where  $\hat{x}$  is defined as  $\hat{x} = \varepsilon x + (1 - \varepsilon)G(z|c, \theta_{real})$ ;  $\varepsilon$  is randomly sampled from the uniform distribution, i.e.,  $\varepsilon \sim U[0, 1]$ ; and  $\lambda_{gp}$  is a weight parameter for WGAN-GP [39]. The gradient norm  $E_{\hat{x} \sim P_{\hat{x}}(\hat{x})} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]$  for the penalty is calculated on a linear interpolation between the pairs of real data points  $P_{data}(x)$  and those of the generator distribution  $G(z|c, \theta_{real})$  [38]. It should be noted that WGAN-GP utilizes the *critic* directly for losses rather than using the sigmoid cross-entropy loss [28]. The adversarial loss guides the generator of the PeaceGAN to generate fake SAR target images that are not distinguishable from the real ones by the discriminator of the PeaceGAN. Note that we directly use FC1 output (i.e., *critic*), not "Sigmoid" of Figure 2, for Equation (4).

#### 3.2.2. Pose Estimator Loss

The loss  $L_{pe}$  of the pose estimator has two components of  $L_{pe,real}$  and  $L_{pe,fake}$  for the input data of two types into the discriminator. The two losses are calculated by

$$L_{pe,real} = E_{x \sim P_{data}(x)} [(\cos \theta_{real} - \cos \theta_{pe})^2 + (\sin \theta_{real} - \sin \theta_{pe})^2] |x], \quad (6)$$

$$L_{pe, fake} = E_{z \sim P_z(z)} [(\cos \theta_{real} - \cos \theta_{pe})^2 + (\sin \theta_{real} - \sin \theta_{pe})^2 | G(z|c, \theta_{real})]. \quad (7)$$

$L_{pe, real}$  in Equation (6) guides the discriminator to identify the pose angles of SAR target images as correctly as possible, and  $L_{pe, fake}$  in Equation (7) allows for the generator to have an ability to generate SAR target images at the intended pose angles.

### 3.2.3. Auxiliary Classifier Loss

The loss  $L_{ac}$  of the auxiliary classifier is calculated by a cross-entropy measure to calculate the difference between estimated target class probabilities  $S_{ac}$  and true target-class probabilities  $P_c(c)$ , which is calculated as

$$L_{ac, real} = E_{x \sim P_{data}(x)} [\log P(S_{ac} = c | x)], \quad (8)$$

$$L_{ac, fake} = E_{z \sim P_z(z)} [\log P(S_{ac} = c | G(z|c, \theta_{real}))]. \quad (9)$$

$L_{ac, real}$  in Equation (8) forces the discriminator to classify the input SAR target images as accurately as possible, while  $L_{ac, fake}$  in Equation (9) has the generator to produce the SAR target images of the desired classes as precisely as possible.

### 3.2.4. Total Combined Losses

The total combined losses for the PeaceGAN are denoted as  $L_{PeaceGAN, D}$  for its discriminator and  $L_{PeaceGAN, G}$  for its generator, respectively, which are given by

$$L_{PeaceGAN, D} = \lambda_{adv} L_D + \lambda_{mtl} (\lambda_{pe} L_{pe, real} + \lambda_{ac} L_{ac, real}), \quad (10)$$

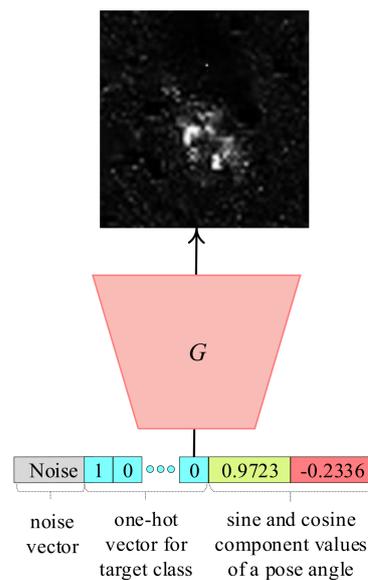
$$L_{PeaceGAN, G} = \lambda_{adv} L_G + \lambda_{mtl} (\lambda_{pe} L_{pe, fake} + \lambda_{ac} L_{ac, fake}), \quad (11)$$

where  $\lambda_{adv}$ ,  $\lambda_{pe}$ ,  $\lambda_{ac}$ , and  $\lambda_{mtl}$  are weight parameters. These two total combined losses are used to train the proposed PeaceGAN in an MTL framework, which allows for effective GAN learning. As a result, if the training of the PeaceGAN is successfully finished, the generator can produce abundant SAR target images of desired classes at any pose angle.

## 4. Introduced Indirect Evaluation Methods for Adequacy and Quality of Generated SAR Target Images

In order to inspect the effectiveness of the PeaceGAN, we introduced new indirect evaluation methods as standard metrics in a perspective of GAN-based SAR target image generation. It is important to evaluate both the adequacy of the generated SAR target images by the generator as training data and their fidelities of pose-angle features and target-class features. Nevertheless, none of the methods has addressed this issue in a quantitative manner. In this paper, we propose three ways of evaluating the generated SAR target images: (i) the *adequacy* as training data for SAR target recognition using a simple classifier, called the SC method; (ii) the *quality* of generated SAR target images using an over-fitted deep classifier for real SAR target images of specific target classes, called the DC method; and (iii) the *quality* of generated SAR target images using an over-fitted deep pose-estimator for real SAR target images at specific pose angles, called the DP method. Figure 3 shows the three types of network structures used for the indirect evaluation methods. From here onwards, we assume that GAN's generators have finished learning the distribution of real SAR target images with a depression angle of  $17^\circ$ , called  $D_{17}$ .





**Figure 4.** An example of how to generate the intended class of the SAR image with the desired pose angle by PeaceGAN: the “BMP2”-class SAR image (first class) with pose angle  $346.49^\circ$  of  $D_{17}$ .

#### 4.2. Indirect Evaluation Using the DC Method

Figure 3b shows a deep classifier for the DC method, called DC-Net. Unlike the SC method, the DC-Net is *pre-trained* to indirectly evaluate (inspect) the *quality* of  $D_{17,GAN}^X$  in the perspectives of target-class features, which are inspired from [30,50]. First, the DC-Net is intentionally trained with a training set of 10,000 samples augmented from  $D_{17}$  to yield 100% target recognition performance for  $D_{17}$ . Here, the DC-Net has a deep structure with a large number of convolution layers, each of which has a large number of feature channels, to be over-fitted for  $D_{17}$ . Therefore, in a testing phase, a higher target recognition performance of the pre-trained DC-Net for  $D_{17,GAN}^X$  implies that the better quality of  $D_{17,GAN}^X$  is generated by the GAN’s generator, and the generator is well trained in the perspective of target-class features. We pre-trained the DC-Net five times with a random weight initialization each time to have the 100% classification rates for all experiments. Thus, the quality of  $D_{17,GAN}^X$  is measured as the average of the resulting five target classification rates tested by the five pre-trained DC-Nets in the perspective of target-class features.

#### 4.3. Indirect Evaluation Using the DP Method

Figure 3c shows a deep pose-estimator for the DP method, called DP-Net. Moreover, the DP-Net is designed almost in the same manner as the DC-Net. The only difference is the usage of *Tanh* at the output where the DP-Net is trained by the same square difference loss in Equation (6). Further, using the same training set of 10,000 samples augmented from  $D_{17}$ , we pre-trained the DP-Net five times with a random weight initialization each time for pose angle estimation where the DP-Net is trained to be over-fitted to the pose-angle features of  $D_{17}$  each time. It should be noted that, unlike the DC-Net, it is hard to train the DP-Net to be over-fitted to perfectly predict the pose angles of the samples in  $D_{17}$ . This is because the pose angles of the SAR target images have continuous values in the range of  $[0^\circ, 360^\circ)$ . From the five-times training of over-fitting, the DP-Net yielded the pose angle estimation performance in terms of the two averages of the resulting five MAD and STD values with  $2.22^\circ$  and  $3.28^\circ$ , respectively, for  $D_{17}$ , where the MAD and STD indicate the mean absolute difference and the standard deviation difference between the true pose angles ( $\theta_{real}$ ) and their estimated ones ( $\theta_{pe}$ ), respectively. Therefore, when the quality of  $D_{17,GAN}^X$  in the perspective of pose-angle features is measured, smaller MAD and STD values measured by the pre-trained DP-Net for  $D_{17,GAN}^X$  imply that a better quality of  $D_{17,GAN}^X$  is generated and the corresponding generator is also trained well in the perspective of pose-angle features.

The SC-Net, DC-Net, and DP-Net in Figure 3 are trained with the same hyper-parameters. All weights of the convolution filters are initialized by Xavier initializer [63]. The mini-batch size is set to 100; the weight decay coefficient for L2 loss is 0.004; the total epoch is 50; and the initial learning rate is set to 0.01 and is multiplied by 0.1 cumulatively at the 10th, 25th, and 40th epochs. In addition, the mini-batch stochastic gradient descent method is utilized with a momentum parameter value of 0.9.

## 5. Experiments Results

### 5.1. Experiments Settings

#### 5.1.1. MSTAR Dataset

The Moving and Stationary Target Acquisition and Recognition (MSTAR) public dataset was utilized for experiments of the SAR target image generation. The MSTAR was gathered by the Sandia National Laboratory (SNL) SAR sensor platform that was supported by Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory (AFRL) [64]. Each piece of MSTAR data includes information about azimuth angles, depression angles, versions, configurations, and target classes, among others. The MSTAR dataset is a benchmark for experiments related to various problems handling SAR images [8–11,30,31] and is composed of  $0.3 \times 0.3$  m resolution of target images captured by X-band SAR sensor. Figure 5 shows some SAR target images of four classes with different pose angles of two different depression angles taken from the MSTAR data. As shown in Figure 5, it is hard to infer either the exact target classes or the corresponding pose angles of the SAR target images due to both their speckle noise and low spatial resolutions. For experiments, we used a data set of the central-cropped  $64 \times 64$ -sized 2747 SAR target images with a depression angle of  $17^\circ$  for  $D_{17}$  from MSTAR data to train GANs. The size of  $64 \times 64$  is known to be sufficient to contain both the clutters and the targets adequately for SAR target image generation [30].  $D_{15}$  also consists of the  $64 \times 64$ -sized 2425 SAR target images of MSTAR data obtained under a depression angle of  $15^\circ$ . Table 1 summarizes the data set used for our experiments. There were 10 target classes that are composed of one bulldozer (D7), one truck (ZIL131), one air defense unit (ZSU234), one rocket launcher (2S1), two tanks (T62, T72), and four armored personnel carriers (BMP2, BTR60, BTR70, BRDM2).



**Figure 5.** Examples of MSTAR dataset: (a) “BMP2”-class with 26.49 pose angles and 17 depression angles, (b) “2S1”-class with 143.33 pose angles and 17 depression angles, (c) “T62”-class with 256.52 pose angles and 17 depression angles, (d) “ZSU234”-class with 347.99 pose angles and 15 depression angles.

**Table 1.** The information of MSTAR data for experiments.

Target Names (Classes)	Serial #	For GANs, DC and DP (SAR Image Size: $64 \times 64$ )		Only Used for SC (SAR Image Size: $64 \times 64$ )	
		Depression Angles	# of Data (Total: 2747)	Depression Angles	# of Data (Total: 2425)
BMP2	9563	17	233	15	195
BTR70	C71	17	233	15	196
T72	132	17	232	15	196
BTR60	K10	17	256	15	195
2S1	B01	17	299	15	274
BRDM2	E71	17	298	15	274
D7	92V	17	299	15	274
T62	A51	17	299	15	273
ZIL131	E12	17	299	15	274
ZSU234	D08	17	299	15	274

Since there were limited numbers of available MSTAR SAR target images for GAN training, we utilized data augmentation methods of [10,11]: (i) *rotation*, (ii) *pose synthesis*, and (iii) *speckle noising* to prevent from the *mode collapse* and a memorization problem of the GAN generators, which are the severe issues of GAN training. *Rotation* data augmentation randomly rotates SAR target images to produce new SAR target images within a small 15-degree angle to preserve the backscattering characteristics [10,11]. *Pose synthesis* data augmentation combines two neighboring SAR images that have similar pose angles to generate new SAR target images by a weighted sum [10]. *Speckle noising* data augmentation generates new data SAR images by newly adding speckle noises sampled from an exponential distribution [10]. To conduct experiments at sufficient performance levels for better comparisons, we produced 5000 SAR target images for each of the 10 classes on the basis of the above data augmentation methods. As a result, our training data set had  $64 \times 64$ -sized 50,000 SAR target images with the  $17^\circ$  depression angle for all 10 classes.

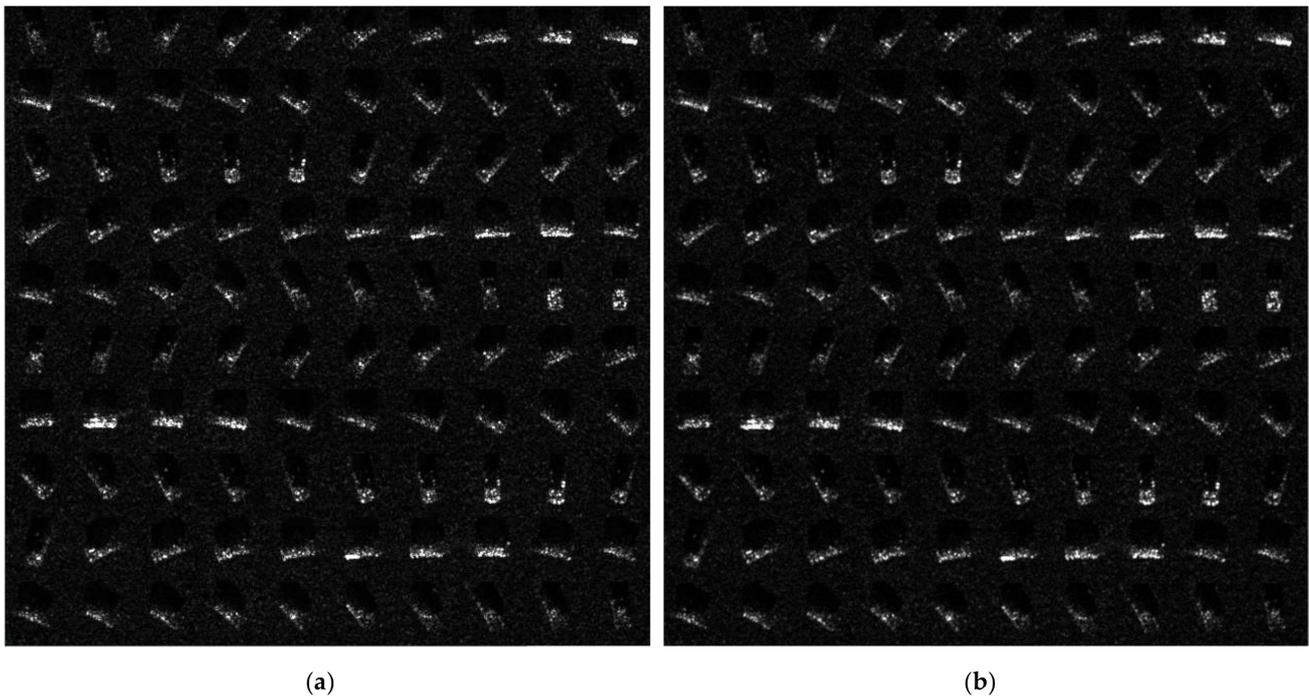
### 5.1.2. Implementation of PeaceGAN

We used two Adam optimizers [65] to minimize  $L_{PeaceGAN,G}$  and  $L_{PeaceGAN,D}$  with learning rates of  $lr_g$  and  $lr_d$ , respectively. The initial value of  $lr_d$  was set to 0.0005, which started to decay linearly to zero from the 40th epoch to the last epoch, the 60th. In addition, we found that a relationship of  $lr_g = 5 \times lr_d$  was appropriate for stable training in our experiments. The generator and the discriminator with the two additional structures were trained alternatively by the two corresponding Adam optimizers. The momentum term  $\beta_1$  of Adam optimizers was set to 0.5, being widely known to make the GAN learning more stable [29], and  $\beta_2$  was set to 0.999 as a default setting [65]. All weights of the convolution filters were initialized by random normal distribution with a zero mean and a standard deviation of 0.02. The mini-batch size was 25. The dimension of the input noise vector was  $N_z = 64$ , the number of SAR target image classes was  $N_c = 10$ , and  $N_p = 2$  since the pose angle input consisted of two sine and cosine angle components. For the weight parameters for losses, we empirically set  $\lambda_{gp} = 5$ ,  $\lambda_{adv} = 1$ ,  $\lambda_{pe} = 0.2$ ,  $\lambda_{ac} = 0.8$ , and  $\lambda_{mtl} = 1$ . The size of all SAR target images for experiments was  $64 \times 64$ .

## 5.2. Evaluations on Generated SAR Target Images by the Proposed PeaceGAN

### 5.2.1. Subjective Comparison between Generated and Real SAR Target Images

Figure 6 shows some generated SAR target images in  $D_{17,GAN}^{Peace}$  generated by the PeaceGAN and their corresponding real SAR target images of a same target class at the same pose angles in  $D_{17}$ . As shown in Figure 6, it can be noted that the PeaceGAN can produce SAR target images visually very similar to the actual SAR target images. The generated SAR target images contain similar speckle noises to those of the real ones and similar shadows around the targets.



**Figure 6.** The comparison between  $D_{17,GAN}^{Peace}$  and D17: (a) first 100 “BMP2”-class SAR images generated by PeaceGAN ( $D_{17,GAN}^{Peace}$ ) at pose angle increments with  $5^\circ$ ; (b) first original 100 “BMP2”-class SAR images from D17 at pose angle increments with  $5^\circ$ .

### 5.2.2. Analyses on PeaceGAN Variants Using SC and DC Performances

Table 2 shows the target recognition performance (%) for the several PeaceGAN variants indirectly evaluated by the SC and DC. For this, each  $D_{17,GAN}^{Peace}$  generated by PeaceGAN variant was utilized to train the SC-Net three times with a random weight initialization to obtain stable average results tested with  $D_{15}$ , according to the SC described in Section 4.1, and the each  $D_{17,GAN}^{Peace}$  was tested by the five pre-trained DC-Nets to yield an average target recognition performance, according to the DC described in Section 4.2.

**Structural Variants.** We first compared three PeaceGAN variants with respect to their structures using the SC and DC (methods) from a structural point of view. In Table 2, the “PeaceGAN with NNConv &  $CS_g$ ”, which is our baseline structure shown in Figure 2, indicated that (i) its generator used the nearest neighborhood upsampling method as described in Section 3.1.1, and (ii) the pose angle input to the generator consisted of separate cosine and sine component values, called  $CS_g$ , trained with Equations (6) and (7) for the pose estimation losses. Instead of the NNConv, the “PeaceGAN with USConv &  $CS_g$ ” utilized the fractionally-strided convolution filters for two-times upsampling, called USConv, used in the DCGAN’s generator [29].

Finally, instead of using the  $CS_g$  as a pose angle input, the “PeaceGAN with NNConv &  $DN_g$ ” replaced the  $CS_g$  with two same values, each of which was the directly normalized pose angle by  $\theta_{norm} = (\theta_{real} - 180)/180$  into the range of  $[-1, 1]$  to keep  $N_p = 2$  for a fair comparison to  $CS_g$ , denoted as  $DN_g$ , with this variant being trained by Equations (12) and (13) for the pose estimation losses instead of Equations (6) and (7) to keep a scale of losses for the fair comparison:

$$L_{pe,real} = E_{x \sim P_{data}(x)} [2 \times (\theta_{norm} - \theta_{pe})^2 | x], \quad (12)$$

$$L_{pe,fake} = E_{z \sim P_z(z)} [2 \times (\theta_{norm} - \theta_{pe})^2 | G(z|c, \theta_{norm})], \quad (13)$$

It is also should be noted that both Equations (12) and (13) directly calculate mean squares error between  $\theta_{norm}$  and  $\theta_{pe}$ , not utilizing both cosine and sine functions. From the

results in Table 2, we can see that NNConv was more effective than USConv, and the usage of  $CS_g$  in the form of separate cosine and sine components trained with Equations (6) and (7) loss functions also helped the generator effectively learn the target-class features with respect to the target's pose angles better than the usage of  $DN_g$ . It is also worthwhile to mention that applying the spectral normalization (SN) to the discriminator of the PeaceGAN was also effective for GAN-based SAR target image generation, as shown in Table 2.

**GAN techniques.** On the other hand, in terms of GAN techniques, we found that the proposed PeaceGAN ("PeaceGAN with NNConv &  $CS_g$ ") that is trained by WGAN-GP yields superior results in comparison with PeaceGAN variants trained by any other GAN techniques [66–69], as shown in Table 2. For these experiments, we set weight parameters for gradient penalty to 5 for both WGAN-LP [66] and DRAGAN [67] including RaDRAGAN [68], which means utilizing DRAGAN loss combined with RaGAN loss [68], respectively. The other settings are the same for fair comparisons. It should be noted that the PeaceGAN variant trained with standard GAN losses of Equations (2) and (3) alone cannot be trained by any combination of hyperparameters and the mode collapse has always occurred. Therefore, the results in Table 2 indicate that it is important to select appropriate GAN techniques applied to the given structure for better quality of SAR image generation.

**Table 2.** Target recognition performance (%) for PeaceGAN variants indirectly evaluated by the SC and DC methods.

Variants	with SN		without SN	
	SC	DC	SC	DC
<b>PeaceGAN with NNConv &amp; <math>CS_g</math></b>	<b>81.20</b>	<b>99.42</b>	78.93	98.46
PeaceGAN with USConv & $CS_g$	77.61	98.04	76.25	98.01
PeaceGAN with NNConv & $DN_g$	76.21	96.70	74.56	96.07
PeaceGAN by WGAN-LP [66]	79.59	99.20	77.11	97.31
PeaceGAN by RaDRAGAN [68]	62.60	92.25	58.60	87.94
PeaceGAN by DRAGAN [67]	58.47	90.16	56.37	89.63
PeaceGAN by LSGAN [69]	33.61	79.11	32.25	69.01
PeaceGAN with $\lambda_{pe} = 0.05$	75.34	98.67	67.18	94.12
PeaceGAN with $\lambda_{pe} = 0.1$	77.69	98.87	73.08	96.69
PeaceGAN with $\lambda_{pe} = 0.4$	77.90	98.90	76.87	98.05
PeaceGAN with $\lambda_{pe} = 0.8$	78.43	98.67	74.89	97.11

**Pose Estimator Loss.** It should also be noted that the influence of  $\lambda_{pe}$  to the quality of PeaceGAN's outputs is quite substantial for both the class-features' point of view and the adequacy as training data for SAR target recognition. Thus,  $\lambda_{pe}$  should be carefully determined as shown in Table 2 ( $\lambda_{pe} = 0.2$  for the final PeaceGAN).

**Other Hyperparameters.** Since the GAN framework is generally difficult to train well, we conducted various experiments to search appropriate hyperparameters settings. To empirically determine several hyperparameters, we first denoted ( $\bullet$ ,  $\bullet$ ) as (SC, DC) performances, respectively. Note that our final PeaceGAN showed performances of (**81.20, 99.42**). All variants of PeaceGAN have the same setting for training with their own variant values (e.g., hyperparameter values such as  $\lambda_{mtl}$ ,  $lr_d$ ,  $\lambda_{gp}$ ,  $lr_g$ , and batch sizes), and are trained separately from scratch. The PeaceGAN variants trained with the settings of  $\lambda_{mtl} = 2.5, 5, 10$  yielded (77.23, 98.90), (76.08, 98.03), and (61.81, 93.58), respectively ( $\lambda_{mtl} = 1$  for PeaceGAN). The PeaceGAN variants trained with the initial learning rates for the discriminator as  $lr_d = 0.0001, 0.001$  showed (72.99, 91.97) and (70.47, 98.29), respectively ( $lr_d = 0.0005$  for PeaceGAN). The PeaceGAN variants trained with  $\lambda_{gp} = 1, 10$  performed as (79.18, 97.38) and (76.12, 98.76), respectively ( $\lambda_{gp} = 5$  for PeaceGAN). The variants trained with the relationships of  $lr_g = 3 \times lr_d$  and  $7 \times lr_d$  yielded (74.19, 98.98) and (79.18, 97.38), respectively ( $lr_g = 5 \times lr_d$  for PeaceGAN). On the other hand, the variants trained with batch sizes of 10 and 50 yielded (75.59, 94.61) and (77.00, 98.42), respectively (25 for PeaceGAN).

In summary, our final PeaceGAN was empirically determined with the optimal setting that yielded the best performance via all the above experiments.

Finally, in order to see the adequacy of the generated SAR target image set ( $D_{17,GAN}^{Peace}$ ) as a training dataset compared to its corresponding original SAR target dataset ( $D_{17}$ ), we also trained the SC-Net with  $D_{17}$ , which yielded 83.71% target recognition performance when it was tested with  $D_{15}$ . On the other hand, when the SC-Net was trained with  $D_{17,GAN}^{Peace}$  and tested with  $D_{15}$ , the resulting target recognition rate turned out to be 81.20%, as shown in Table 2, which was very close to 83.71%. From this, it can be noted that the PeaceGAN is capable of generating adequate SAR target images as training data. As mentioned before, the DC-Net was pre-trained in an intended over-fitted manner for  $D_{17}$ , thus yielding 100% target recognition performance for  $D_{17}$  itself. The intention of this over-fitting is that the DC-Net is pre-trained to memorize the target-class features of  $D_{17}$ . When  $D_{17,GAN}^{Peace}$  was tested by the five pre-trained DC-Nets, the averaged target recognition performance turned out to be 99.42%, as shown in Table 2, which was very close to 100%. For the target recognition performance of 99.42% in Table 2, we show in detail in Table 3 the averaged per-class target recognition performance (confusion matrix) of the five pre-trained DC-Nets for  $D_{17,GAN}^{Peace}$ . As shown in Table 3, the generated SAR target images of “D7”, “ZIL131”, and “ZSU234” classes were classified remarkably with 100% accuracy, and those of the other classes were also classified nearly close to 100% accuracy. From this, it can also be noted that the PeaceGAN is capable of generating the SAR target images with very similar target-class features of  $D_{17}$ .

**Table 3.** Confusion matrix for target recognition performance (%) on  $D_{17,GAN}^{Peace}$  tested by five pre-trained DC-Nets.

Classes	BMP2	BTR70	T72	BTR60	2S1	BRDM2	D7	T62	ZIL131	ZSU234	Recognition Rates (%) / Std.
BMP2	230.6	1	0	1.2	0.2	0	0	0	0	0	98.97/0.210
BTR70	1	229.4	0	2.6	0	0	0	0	0	0	98.45/0.210
T72	0	1	231	0	0	0	0	0	0	0	99.57/0.000
BTR60	0	4.2	0	251.8	0	0	0	0	0	0	98.36/0.156
2S1	0	1	0	0	296.6	0	0	1.4	0	0	99.20/0.164
BRDM2	0.2	0	0	1	0	296.8	0	0	0	0	99.60/0.134
D7	0	0	0	0	0	0	299	0	0	0	100.0/0.000
T62	0	0	1	0.2	0	0	0	297.8	0	0	99.60/0.134
ZIL131	0	0	0	0	0	0	0	0	299	0	100.0/0.000
ZSU234	0	0	0	0	0	0	0	0	0	299	100.0/0.000
Averages											<b>99.42/0.025</b>

### 5.2.3. Indirect Evaluation by DP for Pose-Angle Features of $D_{17,GAN}^{Peace}$

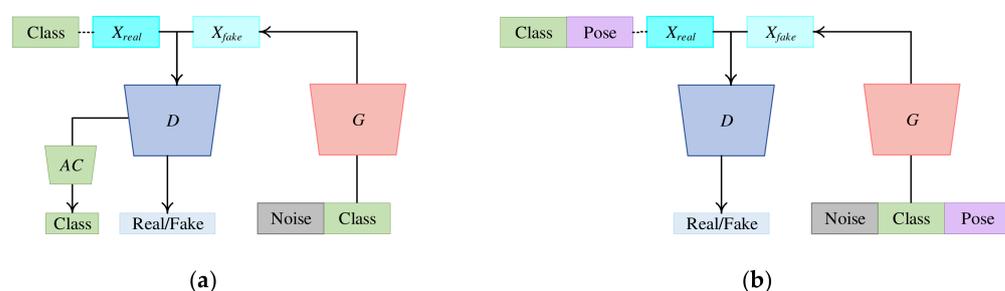
One of the PeaceGAN’s advantages is a capability to generate SAR target images at intended pose angles. Thus, it is necessary to evaluate whether or not the PeaceGAN can appropriately generate SAR target images in a perspective of pose angles. For this, we used an indirect evaluation method using the DP method as described in Section 4.3. Table 4 shows the pose estimation performance of the pre-trained DP-Net on  $D_{17}$  and  $D_{17,GAN}^{Peace}$ . The estimation accuracy values in Table 4 were measured in terms of MAD and STD as their averages of the same tests five times by the DP-Net with five-times independent training with random weight initialization each time. The DP-Net shows the average  $2.22^\circ$  MAD for  $D_{17}$  and the average  $6.83^\circ$  for  $D_{17,GAN}^{Peace}$ . It can be noted in Figure 6b that the generated SAR target images at neighboring pose angles with  $5^\circ$  differences were visually very similar. Thus, the pose-angle features of  $D_{17,GAN}^{Peace}$  were found to be sufficiently similar to those of  $D_{17}$ .

**Table 4.** Pose estimation performance of the pre-trained DP-Net on  $D_{17}$  and  $D_{17,GAN}^{Peace}$ .

Classes	$D_{17}$		$D_{17,GAN}^{Peace}$	
	MAD	STD	MAD	STD
BMP2	2.14	3.00	7.03	11.49
BTR70	2.31	3.67	8.48	12.04
T72	2.61	3.89	6.71	10.44
BTR60	2.16	3.47	6.71	9.55
2S1	1.90	2.75	6.84	10.29
BRDM2	2.12	2.99	6.77	10.73
D7	2.37	3.83	7.45	10.38
T62	2.11	2.66	6.39	10.27
ZIL131	2.16	2.97	5.83	8.49
ZSU234	2.37	3.58	6.10	8.99
Averages	<b>2.22</b>	<b>3.28</b>	<b>6.83</b>	<b>10.27</b>

### 5.3. Comparison between PeaceGAN, Modified ACGAN, and Modified CGAN

To highlight the novelty of our PeaceGAN in a structural point of view, we compared it with the most relevant GANs, ACGAN [70], and CGAN [32]. The ACGAN is based on the standard GAN and has an additional component of a classifier, which is called the auxiliary classifier [70]. The original ACGAN only utilizes class information that is concatenated to the input vector of its generator and has the same forms of classification losses in Equations (8) and (9) but is directly connected to the end of its discriminator, not independently such as the structure of the auxiliary classifier of the PeaceGAN. We compared the generation capabilities of SAR target images by the PeaceGAN that utilizes both target class and pose angle information and the ACGAN that only uses target class information. For a fair comparison, we modified the original ACGAN to have the same generator, discriminator, and auxiliary classifier as the PeaceGAN, called ACGAN<sub>m</sub>, which became equivalent to a PeaceGAN variant without the pose estimator, as shown in Figure 7a. For experiments, the parameters of the ACGAN<sub>m</sub> were set with the same values as the PeaceGAN, except for the two parameters with  $N_z = 66 (= 64 + 2)$ ,  $N_c = 10$ , and  $N_p = 0$  to keep the original 76-dimensional combined input vector, and with  $\lambda_{pe} = 0$  and  $\lambda_{ac} = 1.0 (= 0.8 + 0.2)$  to keep the same scales of gradients. Since the ACGAN<sub>m</sub> cannot incorporate the pose angle information directly into SAR target image generation, we were only able to constitute the input vector with a one-hot vector of target classes and a random noise vector, where the resulting generated SAR target images were denoted as  $D_{17,GAN}^{AC}$  for the depression angle of  $17^\circ$ . Thus,  $D_{17,GAN}^{AC}$  generated SAR target images at random pose angles with the same number of SAR target images per each class corresponding to  $D_{17}$ .

**Figure 7.** The conceptual architectures of (a) modified ACGAN [70] (ACGAN<sub>m</sub>) and (b) modified CGAN [32] (CGAN<sub>m</sub>) [30,71].

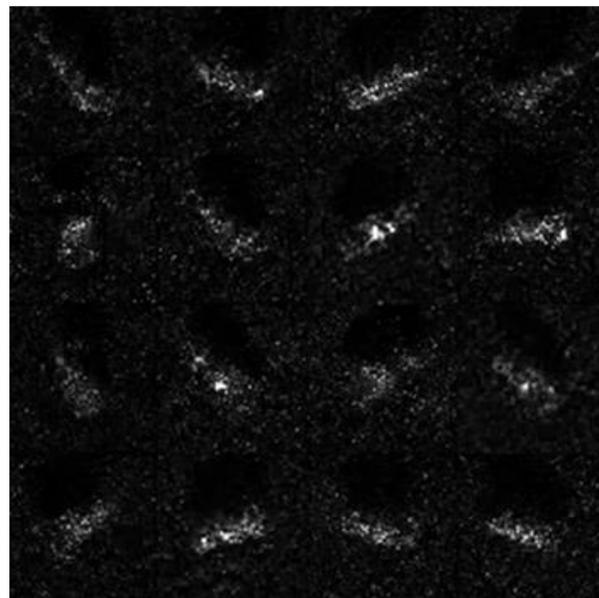
On the other hand, we also designed a modified CGAN-based network, called CGAN<sub>m</sub>, as shown in Figure 7b. For a fair comparison, we firstly modified the original CGAN [32] to have the same generator and discriminator of the PeaceGAN without both pose estimator and auxiliary classifier. Secondly, to follow the definition of CGAN [32],

we designed a discriminator of  $CGAN_m$  additionally having conditioned inputs that were composed of the  $X_{real}$  or  $X_{fake}$  concatenated with a stretched (filling same values in  $64 \times 64$  spatial dimensions) one-hot target class, a cosine value, and a sine value of a pose angle. For experiments, all parameters of the  $CGAN_m$  were set with the same values as the PeaceGAN, except the input channel of the discriminator increased to 13 ( $=1 + 10 + 2$ ) from 1 and  $\lambda_{mtl} = 0$ .

Table 5 shows the target recognition performance evaluated by both the SC and DC on  $D_{17,GAN}^{AC}$  and  $D_{17,GAN}^C$  in comparison with  $D_{17,GAN}^{Peace}$ . As shown in Table 5, the target recognition performances on  $D_{17,GAN}^{Peace}$  by both SC and DC were always higher than those on  $D_{17,GAN}^{AC}$  and  $D_{17,GAN}^C$ , which implies that the PeaceGAN can generate better SAR target images than the  $ACGAN_m$  and  $CGAN_m$  in terms of both the “adequacy as training data” and the fidelity of “target-class features”, respectively. Figure 8 shows some of generated SAR target images by the  $ACGAN_m$ . As shown in Figure 8, the shapes of the targets generally appeared dim, and sometimes speckle noises looked dissimilar to  $D_{17}$ . From this observation in Figure 8 and the results in Table 5, we noticed that the proposed PeaceGAN can not only control pose angles to generate SAR target images but also can learn the joint distributions of SAR target images more reliably for various target classes and pose angles.

**Table 5.** The comparison between  $D_{17,GAN}^{Peace}$ ,  $D_{17,GAN}^{AC}$ , and  $D_{17,GAN}^C$ : target recognition performance (%) by SC and DC.

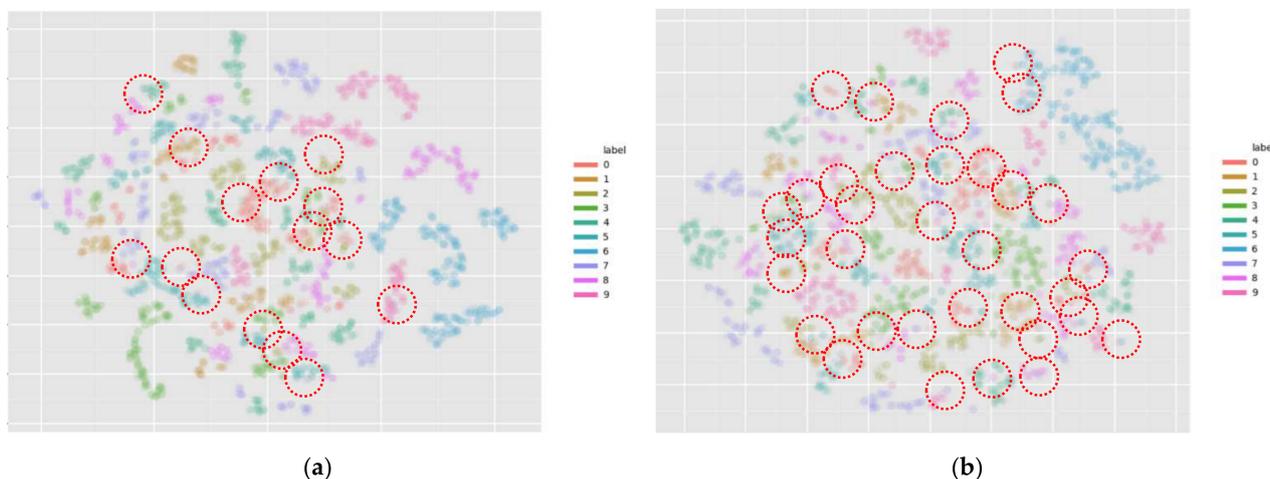
Structures	SC	DC
PeaceGAN	81.20	99.42
$ACGAN_m$	73.03	96.23
$CGAN_m$	77.53	95.63



**Figure 8.** The 16 “BMP2”-class SAR images generated by  $ACGAN_m$ . It should be noted that the generator of  $ACGAN_m$  generated SAR target images at random pose angles with the same number of SAR target images per each class corresponding to  $D_{17}$ .

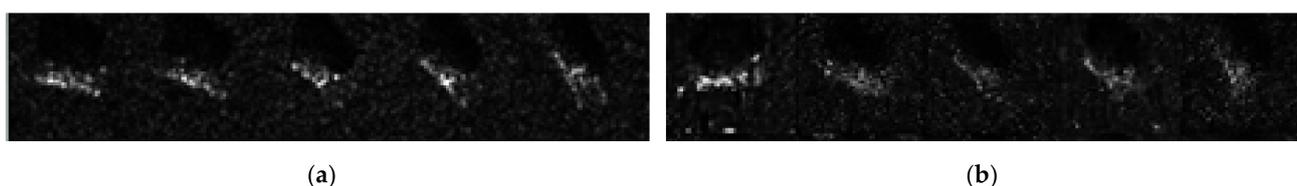
We also used 2D t-SNE [72] to visualize the distributions of generated SAR target images of  $D_{17,GAN}^{Peace}$  and  $D_{17,GAN}^{AC}$  in a two-dimensional space. Figure 9 shows the distribution visualizations of 10 kinds of feature points for  $64 \times 64$ -sized 2747 generated SAR target images (data samples) using the 2D t-SNE, where the data samples in the same class were represented by the same color. As shown in Figure 9, the data samples in  $D_{17,GAN}^{AC}$  were spread with larger overlaps across different classes, as indicated by red circles, in

comparison with those of  $D_{17,GAN}^{Peace}$ . In other words, the data samples of  $D_{17,GAN}^{Peace}$  belonging to the same classes exhibited higher concentrations than those of  $D_{17,GAN}^{AC}$ . On the basis of these observations, we are also able to conclude that taking advantage of pose angle information for SAR target image generation can help generate SAR target images more reliably with higher fidelities for various target classes.



**Figure 9.** The comparison between 2D t-SNE of (a)  $D_{17,GAN}^{Peace}$  and (b)  $D_{17,GAN}^{AC}$  by indicating overlapped points with red circles. Best viewed in zoom.

On the other hand, Figure 10 shows the examples of  $D_{17,GAN}^{Peace}$  generated by the PeaceGAN and  $D_{17,GAN}^C$  generated by the CGAN<sub>m</sub>. As shown in Figure 10, it also can be easily noted that the PeaceGAN can produce SAR target images visually better than CGAN<sub>m</sub>. Furthermore, the generated SAR target images of  $D_{17,GAN}^C$  tend to contain unnatural speckle noises and abnormal scattering characteristics. Furthermore, the DC result of CGAN<sub>m</sub> (95.63%) was slightly lower than that of ACGAN<sub>m</sub> (96.23%), as shown in Table 5, which implies that the fidelity of target-class features of ACGAN<sub>m</sub> was better than that of CGAN<sub>m</sub>. On the other hand, note that CGAN<sub>m</sub> can also control the pose angles to generate  $D_{17,GAN}^C$  and therefore it is reasonable that the  $D_{17,GAN}^C$  has more diverse pose angles than  $D_{17,GAN}^{Peace}$  and thus the SC result of CGAN<sub>m</sub> (77.53%) is better than that of ACGAN<sub>m</sub> (73.03%) from the perspective of adequacy as training data (SC). Lastly, Table 6 shows the pose estimation performance on  $D_{17,GAN}^{Peace}$  and  $D_{17,GAN}^C$  evaluated by DP to compare the fidelity of pose-angle features. As shown in Table 6, the DP results of both MAD and STD of  $D_{17,GAN}^{Peace}$  were much better than those of  $D_{17,GAN}^C$ . As a result, the proposed structure of PeaceGAN including pose estimator and auxiliary classifier also outperformed the CGAN-based structure in both quantitative and qualitative manners.



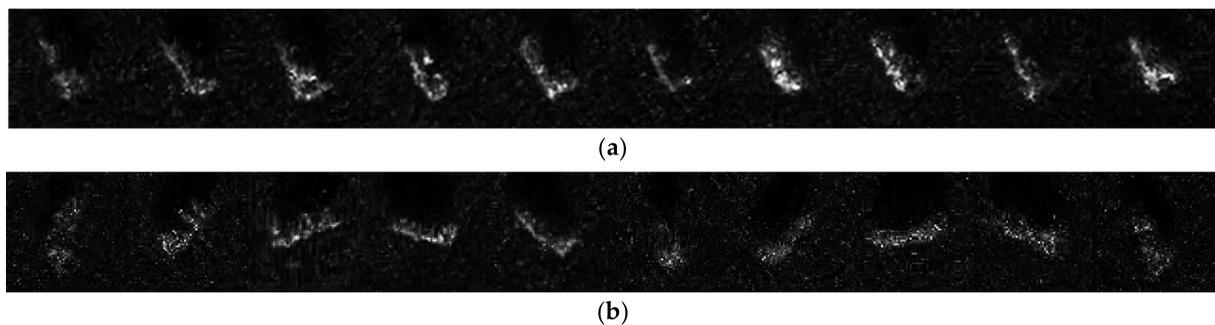
**Figure 10.** The comparison between examples of (a) 41th to 45th “BMP2”-class SAR images of  $D_{17,GAN}^{Peace}$  and (b) 41th to 45th “BMP2”-class SAR images of  $D_{17,GAN}^C$ .

**Table 6.** Pose estimation performance of the pre-trained DP-Net on  $D_{17,GAN}^{Peace}$  and  $D_{17,GAN}^C$ .

Classes	$D_{17,GAN}^{Peace}$		$D_{17,GAN}^C$	
	MAD	STD	MAD	STD
BMP2	7.03	11.49	11.59	19.28
BTR70	8.48	12.04	10.95	18.07
T72	6.71	10.44	11.59	19.67
BTR60	6.71	9.55	9.50	16.05
2S1	6.84	10.29	9.67	16.99
BRDM2	6.77	10.73	10.61	18.38
D7	7.45	10.38	10.61	16.41
T62	6.39	10.27	9.61	16.98
ZIL131	5.83	8.49	8.85	14.23
ZSU234	6.10	8.99	8.72	13.32
Averages	<b>6.83</b>	<b>10.27</b>	<b>10.17</b>	<b>16.94</b>

#### 5.4. SAR Image Generation with Varying Target Classes and Pose Angles

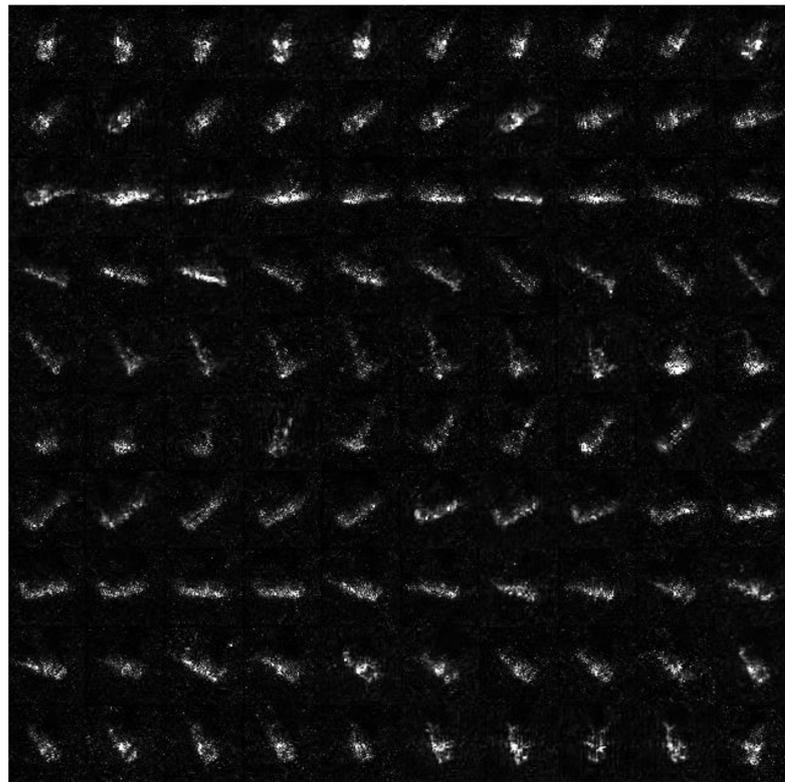
Figure 11 shows the ability of PeaceGAN, as a generator, that can generate SAR target images for given desired pose angles and target classes. The SAR target images in Figure 11a were obtained by inputting into the PeaceGAN generator the target class information from “BMP2” to “ZSU234” sequentially while fixing the pose angle at  $151.2^\circ$ . Figure 11b shows the generated SAR target images of “BTR70”-class for varying pose angles from  $0^\circ$  to  $324^\circ$  uniformly by a pose angle increment of  $36^\circ$ . As shown in Figure 11, it is clear that the PeaceGAN can faithfully generate SAR target images with two controllable variables such as pose angles and target classes.



**Figure 11.** The generated SAR images with varying two variables, target classes, and pose angles: (a) fixed pose angle  $151.2^\circ$  and varying 10 target classes; (b) fixed target class “BTR70” and varying 10 pose angles from  $0^\circ$  to  $324^\circ$  uniformly by a pose angle increment of  $36^\circ$ .

#### 5.5. Discussion about Applications of PeaceGAN for Training SAR Target Recognition Network

Since the generator of the PeaceGAN can be trained to generate SAR target images at intended pose angles for desired target classes, it can also be utilized to train a SAR target recognition network by generating abundant SAR images at any intended pose angle. For example, if we want to generate  $N_{TR}$  numbers of SAR target images, let us take them at uniformly spaced pose angles between  $0^\circ$  and  $360^\circ$  per target class by first calculating both cosine and sine values of the  $N_{DA}$  angles at equally spaced angles from  $0^\circ$ ,  $(360^\circ/N_{TR}) \times 1$ ,  $(360^\circ/N_{TR}) \times 2$ ,  $\dots$ ,  $(360^\circ/N_{TR}) \times (N_{TR} - 1)$ , without loss of generality. Then, with the pre-calculated cosine and sine component values and an one-hot vector for each target class as shown in Figure 4, we can generate  $D_{17,GAN}^{Peace}$  with  $N_{TR}$ -synthesized SAR target images per target class at  $N_{TR}$  pose angles, which is denoted as  $D_{17,GAN}^{PeaceTR}$ . Figure 12 shows some examples of generated SAR target images with  $N_{TR} = 100$  for “ZIL131”-class. The generated SAR target images in Figure 12 are displayed for various angles with each incremented angle of  $3.6^\circ$  from  $0^\circ$  to  $356.4^\circ$ .



**Figure 12.** Examples of generated SAR target images at various pose angles with  $N_{TR} = 100$ ; “ZIL131”-class SAR images of  $D_{17,GAN}^{PeaceTR}$  to  $D_{17}$ .

For a simple verification of the PeaceGAN’s ability for training SAR target recognition network, we utilized the same structure of the DC-Net shown in Figure 3b with the same implementation settings for training. For this experiment, the new DC-Net was trained by  $D_{17,GAN}^{PeaceTR}$  and then tested by  $D_{15}$ . Table 7 shows the target recognition performance of the DC-Net trained by  $D_{17,GAN}^{PeaceTR}$  and then tested by  $D_{15}$ , and the total number of  $N_{TR} \times N_c$  generated by PeaceGAN’s generator where  $N_c = 10$  is the number of target classes. It should be noted that, although the *simple* architecture of DC-Net is newly trained *only* with  $D_{17,GAN}^{PeaceTR}$  by setting  $N_{TR} \times N_c$  to  $10,000 \times 10$ , the DC-Net performed with similar precision levels of 98.10% compared to the previous SAR target recognition methods that are dedicatedly designed to yield high performances such as 93.66% of MSRC [73], 97.84% of DCNN-SVM [74], and 98.81% of M-PMC [75]. From these observations, it can be noted that the generated SAR target images by the PeaceGAN can be utilized to train the SAR target recognition network, which would be useful for generating such SAR target images that are often very costly for acquisition.

**Table 7.** Target recognition performance (%) of a DC-Net newly trained by  $D_{17,GAN}^{PeaceTR}$  for testing with  $D_{15}$ .

$N_{TR} \times N_c$	Accuracy (%)
100 × 10	20.57
200 × 10	46.43
300 × 10	69.27
400 × 10	79.34
500 × 10	90.06
1000 × 10	95.55
1500 × 10	97.36
5000 × 10	97.81
10,000 × 10	<b>98.10</b>

## 6. Conclusions

In this paper, we propose a novel GAN structure that can be jointly trained in an end-to-end manner with multi-task learning of estimating both pose angle and target class information of SAR target images, called PeaceGAN. As a result, the PeaceGAN explicitly disentangles the pose angles and the target class, which leads to both the diversity and the improved quality of the generated SAR target images that are visually very similar to real SAR target images. We also introduce three indirect evaluation methods as standard metrics to evaluate the adequacy of generated SAR target images as training data and the fidelities of target-class features and pose-angle features of them generated by our PeaceGAN. In intensive experiments, the PeaceGAN showed a high potential to generate SAR target images with good fidelity at intended pose angles for desired target classes.

**Author Contributions:** J.O. proposed the PeaceGAN framework; J.O. performed the experiments; J.O. and M.K. analyzed the results; J.O. and M.K. wrote and reviewed the manuscript together. Both authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2016-0-00288) supervised by the IITP (Institute for Information and Communications Technology Planning and Evaluation).

**Data Availability Statement:** The SAR targets used for the experiments are the publicly available dataset MSTAR Public Targets. The descriptions and archive for the MSTAR SAR Target dataset can be found at <https://www.sdms.af.mil/index.php?collection=mstar&page=targets>, accessed on 16 September 2021.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Liu, C.; Li, H.; Liao, W.; Philips, W.; Emery, W. Variational textured Dirichlet process mixture model with pairwise con-straint for unsupervised classification of polarimetric SAR images. *IEEE Trans. Image Process.* **2019**, *28*, 4145–4160. [[CrossRef](#)]
- Kayabol, K.; Zerubia, J. Unsupervised amplitude and texture classification of SAR images with multinomial latent model. *IEEE Trans. Image Process.* **2012**, *22*, 561–572. [[CrossRef](#)]
- Dong, G.; Kuang, G. Classification on the monogenic scale space: Application to target recognition in SAR image. *IEEE Trans. Image Process.* **2015**, *24*, 2527–2539. [[CrossRef](#)]
- Dong, G.; Kuang, G.; Wang, N.; Wang, W. Classification via sparse representation of steerable wavelet frames on Grassmann Manifold: Application to target recognition in SAR image. *IEEE Trans. Image Process.* **2017**, *26*, 2892–2904. [[CrossRef](#)]
- Lee, J.S. Speckle suppression and analysis for synthetic aperture radar. *Opt. Eng.* **1986**, *25*, 636–643. [[CrossRef](#)]
- Nie, X.; Qiao, H.; Zhang, B. A variational model for PolSAR data speckle reduction based on the Wishart Distribution. *IEEE Trans. Image Process.* **2015**, *24*, 1209–1222. [[CrossRef](#)] [[PubMed](#)]
- Deledalle, C.-A.; Denis, L.; Tupin, F.; Reigber, A.; Jager, M. NL-SAR: A unified nonlocal framework for resolution-preserving (Pol)(In)SAR denoising. *IEEE Trans. Geosci. Remote Sens.* **2014**, *53*, 2021–2038. [[CrossRef](#)]
- Chen, S.; Wang, H. SAR target recognition based on deep learning. In Proceedings of the IEEE International Conference on Data Science and Advanced Analytics, Shanghai, China, 30 October–2 November 2014; pp. 541–547. [[CrossRef](#)]
- Li, X.; Li, C.; Wang, P.; Men, Z.; Xu, H. SAR ATR based on dividing CNN into CAE and SNN. In Proceedings of the IEEE 5th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), Singapore, 1–4 September 2015; pp. 676–679. [[CrossRef](#)]
- Ding, J.; Chen, B.; Liu, H.; Huang, M. Convolutional neural network with data augmentation for SAR target recognition. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 364–368. [[CrossRef](#)]
- Chen, S.; Wang, H.; Xu, F.; Jin, Y.-Q. Target classification using the deep convolutional networks for SAR images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4806–4817. [[CrossRef](#)]
- Oh, J.; Youm, G.-Y.; Kim, M. SPAM-net: A CNN-based SAR target recognition network with pose angle marginalization learning. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 701–714. [[CrossRef](#)]
- Zhou, Y.; Wang, H.; Xu, F.; Jin, Y.-Q. Polarimetric SAR image classification using deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1935–1939. [[CrossRef](#)]
- Pei, J.; Huang, Y.; Huo, W.; Zhang, Y.; Yang, J.; Yeo, T.-S. SAR automatic target recognition based on multiview deep learning framework. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 2196–2210. [[CrossRef](#)]
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Montreal, QC, Canada, 8–13 December 2014.

16. Chan, E.R.; Monteiro, M.; Kellnhofer, P.; Wu, J.; Wetzstein, G. Pi-GAN: Periodic Implicit Generative Adversarial Networks for 3D-Aware Image Synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 19–23 June 2021; pp. 5799–5809.
17. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-attention generative adversarial networks. In Proceedings of the International Conference on Machine Learning (ICML), Long Beach, CA, USA, 11–13 June 2019.
18. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral normalization for generative adversarial networks. In Proceedings of the International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 30 April–3 May 2018.
19. Lucic, M.; Kurach, K.; Michalski, M.; Gelly, S.; Bousquet, O. Are GANs created equal? A large-scale study. In Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), Montreal, QC, Canada, 3–8 December 2018.
20. Kurach, K.; Lucic, M.; Zhai, X.; Michalski, M.; Gelly, S. The GAN landscape: Losses, architectures, regularization, and normalization. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019.
21. Chen, X.; Xu, C.; Yang, X.; Song, L.; Tao, D. Gated-GAN: Adversarial gated networks for multi-collection style transfer. *IEEE Trans. Image Process.* **2018**, *28*, 546–560. [[CrossRef](#)]
22. Wang, C.; Xu, C.; Wanga, C.; Tao, D. Perceptual adversarial networks for image-to-image transformation. *IEEE Trans. Image Process.* **2018**, *27*, 4066–4079. [[CrossRef](#)]
23. Hsu, C.; Lin, C.; Su, W.; Cheung, G. SiGAN: Siamese generative adversarial network for identity-preserving face hallucination. *IEEE Trans. Image Process.* **2019**, *28*, 6225–6236. [[CrossRef](#)]
24. Zhang, K.; Luo, W.; Zhong, Y.; Ma, L.; Liu, W.; Li, H. Adversarial spatio-temporal learning for video deblurring. *IEEE Trans. Image Process.* **2018**, *28*, 291–301. [[CrossRef](#)]
25. Li, Y.; Tang, S.; Zhang, R.; Zhang, Y.; Li, J.; Yan, S. Asymmetric GAN for unpaired image-to-image translation. *IEEE Trans. Image Process.* **2019**, *28*, 5881–5896. [[CrossRef](#)]
26. Wang, P.; Bai, X. Thermal infrared pedestrian segmentation based on conditional GAN. *IEEE Trans. Image Process.* **2019**, *28*, 6007–6021. [[CrossRef](#)] [[PubMed](#)]
27. Nowozin, S.; Cseke, B.; Tomioka, R. F-Gan: Training generative neural samplers using variational divergence minimization. In Proceedings of the 30th International Conference on Neural Information Processing Systems (NeurIPS), Barcelona, Spain, 5–10 December 2016.
28. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning (ICML), Sydney, Australia, 6–11 August 2017; pp. 214–223.
29. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. In Proceedings of the International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2–4 May 2016.
30. Guo, J.; Lei, B.; Ding, C.; Zhang, Y. Synthetic aperture radar image synthesis by using generative adversarial nets. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1111–1115. [[CrossRef](#)]
31. Zheng, C.; Jiang, X.; Liu, X. Semi-supervised SAR ATR via multi-discriminator generative adversarial network. *IEEE Sens. J.* **2019**, *19*, 7525–7533. [[CrossRef](#)]
32. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
33. Kanazaki, A.; Matsushita, Y.; Nishida, Y. RotationNet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
34. Rogez, G.; Schmid, C. Image-based synthesis for deep 3D human pose estimation. *Int. J. Comput. Vis.* **2018**, *126*, 993–1008. [[CrossRef](#)]
35. Tran, L.; Yin, X.; Liu, X. Disentangled representation learning gan for pose-invariant face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, Hawaii, 21–26 July 2017; pp. 1415–1424.
36. Sebastian, R. An overview of multi-task learning in deep neural networks. *arXiv* **2017**, arXiv:1706.05098.
37. Mostafa, A.; Yaser, S. Learning from hints in neural networks. *J. Complex.* **1990**, *6*, 192–198. [[CrossRef](#)]
38. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved training of Wasserstein GANs. In Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–9 December 2017.
39. Gao, F.; Ma, F.; Wang, J.; Sun, J.; Yang, E.; Zhou, H. Semi-supervised generative adversarial nets with multiple generators for SAR image recognition. *Sensors* **2018**, *18*, 2706. [[CrossRef](#)]
40. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved techniques for training gans. In Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), Barcelona, Spain, 5–10 December 2016; pp. 2234–2242.
41. Sun, Y.; Liu, Z.; Todorovic, S.; Li, J. Adaptive boosting for SAR automatic target recognition. *IEEE Trans. Aerosp. Electron. Syst.* **2007**, *43*, 112–125. [[CrossRef](#)]
42. El-Darymli, K.; Gill, E.W.; McGuire, P.; Power, D.; Moloney, C. Automatic target recognition in synthetic aperture radar imagery: A state-of-the-art review. *IEEE Access* **2016**, *4*, 6014–6058. [[CrossRef](#)]
43. Zhao, Q.; Xu, D.; Principe, J. Pose estimation of SAR automatic target recognition. In Proceedings of the Image Understanding Workshop, Monterey, CA, USA, 20–23 November 1998.

44. Principe, J.C.; Xu, D.; Iii, J.W.F. Pose estimation in SAR using an information theoretic criterion. In Proceedings of the SPIE 3370, Algorithms for Synthetic Aperture Radar Imagery V, Orlando, FL, USA, 15 September 1998.
45. Jianxiong, Z.; Zhiguang, S.; Xiao, C.; Qiang, F. Automatic target recognition of SAR images based on global scattering center model. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3713–3729. [[CrossRef](#)]
46. Zhao, Q.; Principe, J.C. Support vector machines for SAR automatic target recognition. *IEEE Trans. Aerosp. Electron. Syst.* **2001**, *37*, 643–654. [[CrossRef](#)]
47. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 12 June 2015.
48. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Klambauer, G.; Hochreiter, S. GANs trained by a two time-scale update rule converge to a Nash equilibrium. In Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–9 December 2017.
49. Amazon Mechanical Turk. Available online: <https://www.mturk.com> (accessed on 16 September 2021).
50. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. StarGAN: Unified generative adversarial networks for multidomain image-to-image translation. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2017.
51. Amrani, M.; Bey, A.; Amamra, A. New SAR target recognition based on YOLO and very deep multi-canonical correlation analysis. *Int. J. Remote Sens.* **2021**, 1–20. [[CrossRef](#)]
52. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
53. Feng, S.; Ji, K.; Ma, X.; Zhang, L.; Kuang, G. Target region segmentation in SAR vehicle chip image with ACM net. *IEEE Geosci. Remote Sens. Lett.* **2021**, 1–5. [[CrossRef](#)]
54. Yang, R.; Pan, Z.; Jia, X.; Zhang, L.; Deng, Y. A novel CNN-based detector for ship detection based on rotatable bounding box in SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1938–1958. [[CrossRef](#)]
55. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
56. Geng, X.; Shi, L.; Yang, J.; Li, P.; Zhao, L.; Sun, W.; Zhao, J. Ship detection and feature visualization analysis based on lightweight CNN in VH and VV polarization images. *Remote Sens.* **2021**, *13*, 1184. [[CrossRef](#)]
57. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626.
58. Yue, Z.; Gao, F.; Xiong, Q.; Wang, J.; Huang, T.; Yang, E.; Zhou, H. A novel semi-supervised convolutional neural network method for synthetic aperture radar image recognition. *Cogn. Comput.* **2019**, *13*, 795–806. [[CrossRef](#)]
59. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning (ICML), Lille, France, 7–9 July 2015; pp. 448–456.
60. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human level performance on imagenet classification. In Proceedings of the International Conference on Computer Vision (ICCV), Las Condes, Chile, 11–18 December 2015; pp. 1026–1034.
61. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the ICML Workshop on Deep Learning for Audio, Speech, and Language Processing, Atlanta, GA, USA, 16 June 2013.
62. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
63. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
64. Keydel, E.R.; Lee, S.W.; Moore, J.T. MSTAR extended operating conditions: A tutorial. In Proceedings of the SPIE 2757, Algorithms for Synthetic Aperture Radar Imagery III, Orlando, FL, USA, 10 June 1996.
65. Kingma, D.; Ba, J. A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.
66. Petzka, H.; Fischer, A.; Lukovnikov, D. On the regularization of Wasserstein GANs. In Proceedings of the International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 30 April–3 May 2018.
67. Kodali, N.; Abernethy, J.; Hays, J.; Kira, Z. On convergence and stability of GANs. *arXiv* **2017**, arXiv:1705.07215.
68. Jolicoeur-Martineau, A. The relativistic discriminator: A key element missing from standard gan. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019.
69. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Paul Smolley, S. Least squares generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2794–2802.
70. Odena, A.; Olah, C.; Shlens, J. Conditional image synthesis with auxiliary classifier GANs. In Proceedings of the International Conference on Machine Learning (ICML), Venice, Italy, 22–29 October 2017.
71. Huang, H.; Zhang, F.; Zhou, Y.; Yin, Q.; Hu, W. High resolution SAR image synthesis with hierarchical generative adversarial networks. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019.

- 
72. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
  73. Dong, G.; Wang, N.; Kuang, G. Sparse representation of monogenic signal: With application to target recognition in SAR images. *IEEE Signal Process. Lett.* **2014**, *21*, 952–956. [[CrossRef](#)]
  74. Gao, F.; Huang, T.; Sun, J.; Wang, J.; Hussain, A.; Yang, E. A new algorithm for SAR image target recognition based on an improved deep convolutional neural network. *Cogn. Comput.* **2018**, *11*, 809–824. [[CrossRef](#)]
  75. Park, J.-I.; Kim, K.-T. Modified polar mapping classifier for SAR automatic target recognition. *IEEE Trans. Aerosp. Electron. Syst.* **2014**, *50*, 1092–1107. [[CrossRef](#)]