



Feature Learning for SAR Target Recognition with Unknown Classes by Using CVAE-GAN

Xiaowei Hu^{1,2}, Weike Feng², Yiduo Guo^{2,*} and Qiang Wang³

- ¹ Key Lab for Information Science of Electromagnetic Waves (MoE), Fudan University, Shanghai 200433, China; xwhu@fudan.edu.cn
- ² Early Warning and Detection Department, Air Force Engineering University, Xi'an 710051, China; feng.weike.q4@dc.tohoku.ac.jp
- ³ Experimental Training Base of College of Information and Communication, National University of Defense Technology, Xi'an 710106, China; wq10196@nudt.edu.cn
- * Correspondence: guoyiduo111@126.com

Abstract: Even though deep learning (DL) has achieved excellent results on some public data sets for synthetic aperture radar (SAR) automatic target recognition(ATR), several problems exist at present. One is the lack of transparency and interpretability for most of the existing DL networks. Another is the neglect of unknown target classes which are often present in practice. To solve the above problems, a deep generation as well as recognition model is derived based on Conditional Variational Auto-encoder (CVAE) and Generative Adversarial Network (GAN). A feature space for SAR-ATR is built based on the proposed CVAE-GAN model. By using the feature space, clear SAR images can be generated with given class labels and observation angles. Besides, the feature of the SAR image is continuous in the feature space and can represent some attributes of the target. Furthermore, it is possible to classify the known classes and reject the unknown target classes by using the feature space. Experiments on the MSTAR data set validate the advantages of the proposed method.

Keywords: feature learning; SAR-ATR; unknown classes; deep learning; CVAE-GAN

1. Introduction

Automatic target recognition (ATR) is a challenging task for synthetic aperture radar (SAR) [1,2]. In general, SAR-ATR may consist of two stages: first extract the representative features of SAR image and then assign the image to a predetermined set of classes using a classifier. The discriminative features are crucial for SAR-ATR and can significantly influence the success of the latter classifier. In the last few decades, lots of feature-extracted technologies have been proposed in SAR-ATR, including but not limited to the principal component analysis (PCA)-based method [3,4], the scattering center model (SCM)-based method [5,6], sparse representation method [7,8], Krawtchouk moments method [9], and multi-features fusion method [10]. Different from the traditional technologies which extract the image features manually, the deep learning (DL) technology can automatically extract the image features by combining the feature extractor and the classifier, so the remarkable performance on target recognition can be achieved. Several results on some public data sets for SAR-ATR by using DL (deep learning) have been reported and are far beyond the results by using traditional technologies [11–13]. However, there are some obvious limitations on the DL-based SAR-ATR studies at present. On one hand, many of the existing DL models are black boxes that do not explain their predictions in a way that humans can understand, so it is hard to ensure the rationality and reliability of the predictive models. On the other hand, most of the existing SAR target recognition methods suppose that the target classes in the test set have appeared in the training set. However, in a realistic scenario, there probably exist some unknown target classes. The traditional SAR-ATR methods will not work or deteriorate seriously for the unknown classes.



Article

Citation: Hu, X.; Feng, W.; Guo, Y.; Wang, Q. Feature Learning for SAR Target Recognition with Unknown Classes by Using CVAE-GAN. *Remote Sens.* 2021, *13*, 3554. https:// doi.org/10.3390/rs13183554

Academic Editor: Lefei Zhang

Received: 27 July 2021 Accepted: 4 September 2021 Published: 7 September 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/).

In order to cope with the above challenges, some researches have been done. In [14,15], the transfer learning is studied to transfer the knowledge from some SAR data sets with abundant labeled samples to the target task with limited labeled samples. In [16], the recognition model is firstly trained on the simulated SAR data, and then is transferred to the transferred target task. In [17,18], the simulated SAR images are refined to reduce the difference between the simulated image and the real image. Besides the transfer learning methods, the generative networks [19,20] are also used to produce the SAR images that are not contained in the training data. All the above studies are helpful to improve the classification results for the case that only few samples of the SAR target classes are available. However, for the target classes with no one sample, the few-shot learning (FSL) methods are valid. To address this issue, several zero-shot learning (ZSL) methods have been studied in SAR-ATR. In [21,22], the optical images or the simulated SAR images of the unknown class targets are utilized to provide the aided information for zero-shot target recognition. Without adding extra information, a feature space for SAR-ATR is built based on the known class targets in [23]. By mapping the SAR images of the unknown class targets into the feature space, the attribute of the target can be explained in some extent. Moreover, the feature space is continuous and is able to generate the target images that are not in the training data, although the generated images are usually obscure. In [24], a more stable feature space is built which has the better interpretability. The problem in [23,24] is the feature space can only explain the unknown class qualitatively, but can't distinguish the unknown class from the known classes effectively. Open set recognition (OSR) is a similar task as zero-shot recognition. Besides several known classes, OSR takes all the unknown classes as an extra class. In test, it can classify targets of the known classes and reject targets of the unknown classes. Several studies have been done on OSR for SAR targets, such as the support vector machine (SVM)-based methods [25] and the edge exemplar selection methods [26,27]. These methods don't exploit the strong ability of feature extraction in DL. Recently, a OSR method for SAR-ATR is proposed based on Generative Adversarial Network (GAN) [28]. A classification branch is added to the discriminator, so the discriminator can provide the target class as well as the score of probability that the target belongs to the known class. However, the recognition accuracy decreases obviously as the number of know classes increases. What is more, the proposed model also lacks transparency and interpretability, like most of existing DL models.

In this paper, we focus on the problem of OSR in SAR-ATR where there is no information about the unknown classes, including the training samples, class number, semantic information, etc. The feature learning method is studied for SAR-ATR, which should classify the known classes, reject the unknown classes and is also explainable in some degree. Inspired by [23,24], a feature space which can well represent the attributes of the SAR target and also has the ability of image generation is expected in the paper. Variational Auto-encoder (VAE) [29,30] and GAN [31,32] are the two most popular deep generative models, which can both generate images by inputting random noise. The noise in VAE is supposed to have some certain distribution and can be seen as the representation of images. VAE is rigorous in theory, but the generated images are often blurry. GAN has no explicit assumption on the noise distribution and can usually produce clear images. However, the input noise in GAN can't represent the images and the interpretability is not enough. What is more, the traditional VAE and GAN can't conduct the conditional generation. It is possible to improve the generated image by combining VAE and GAN [33,34]. In the paper, a conditional VAE (CVAE) together with GAN (CVAE-GAN) is proposed for feature learning in SAR-ATR. The latent variable space in VAE, where the noise is produced, is selected as the feature space. Both the target representation and classification are conducted in this feature space. The attribute and class label of the target are used as the conditions to control the image generation. An extra discriminator is used to improve the quality of the generated image. Compared with the existing approaches, the proposed CVAE-GAN model can obtain a feature space for SAR-ATR, which has three main advantages.

- 1. **Condition generation for SAR images**. By using the encoder network and the generation network in CVAE-GAN jointly, clear SAR images, which are absent in the training data set, can be generated with the given target class label and target attributes, while the network in [23] only produces the obscure images. It can be used in data augmentation and benefit for SAR-ATR with limited labeled data.
- 2. **Feature interpretability in some degree**. Different from that of auto-encoder, the latent variable of VAE is continuous in the variable space. Therefore, the latent variable obtained via the encoder network can act as the representation of SAR images and reflects some attributes of the target. The network in [28] is incapable of the interpretability.
- 3. **Recognition of unknown classes.** Because of the additive angular margin in the feature space, the intra-class distance of the known classes is decreased and the interclass distance is increased. As a result, it is possible to classify the known classes and recognize the unknown target classes whose features are outside the angular range of the known classes. The networks in [23,24] can't recognize the unknown classes.

The rest of this paper is organized as follows. In Section 2, the proposed CVAE-GAN is introduced in detail, including the model derivation, loss function and network structure. In Section 3, experimental results are shown to demonstrate the effectiveness of the proposal. In Section 4, the results are further discussed compared with the previous studies. The work is concluded in Section 5.

2. CVAE-GAN for SAR-ATR with Unknown Target Classes

The main objective in this paper is to find an explicable feature space, in which the attributes of the SAR images can be well represented and the unknown classes can be distinguished. The generative model is a good choice to achieve the interpretability. We select VAE as the generative model, because it not only has the generation ability, but also can obtain the feature representation of images. Considering only random images can be generated in the original VAE, in this section, we will firstly derive a conditional VAE model by embedding the target attributes. And then integrate the classification loss and GAN loss into the CVAE model and achieve the CVAE-GAN model. The resultant model encourages the abilities of conditional image generation, feature interpretability and target classification simultaneously. A deep network will be built based on the proposed model in the last.

2.1. CVAE-GAN Model

We expect to generate the corresponding target SAR image with a given class label and some attribute parameters such as the observation angle, resolution, frequency and polarization mode. Let x, y, θ denote the target image, the class label and the related attribute parameters, respectively. The generation model can be realized by maximizing the conditional log-likelihood function

$$\max E_{x \sim p(x)} \log p(x|y,\theta) \tag{1}$$

where p(x), $p(x|y,\theta)$ are the prior distribution and posterior distribution of x. According to the idea of VAE, an auxiliary latent variable z is added to represent log $p(x|y,\theta)$. Let q(z|x) be the posterior distribution of z, and log $p(x|y,\theta)$ can be expressed as

$$\log p(x|y,\theta) = \int_{z} q(z|x) dz \times \log p(x|y,\theta) = \int_{z} q(z|x) \log p(x|y,\theta) dz = \int_{z} q(z|x) \log \frac{p(x,z|y,\theta)}{q(z|x,y,\theta)} dz = \int_{z} q(z|x) \log \frac{p(x,z|y,\theta)}{q(z|x)} \frac{q(z|x)}{q(z|x,y,\theta)} dz = \int_{z} q(z|x) \log \frac{p(x,z|y,\theta)}{q(z|x)} dz + \int_{z} q(z|x) \log \frac{q(z|x)}{q(z|x,y,\theta)} dz = \int_{z} q(z|x) \log \frac{p(x,z|y,\theta)}{q(z|x)} dz + KL[q(z|x)||q(z|x,y,\theta)]$$

$$(2)$$

where $KL[q(z|x)q(z|y,\theta)]$ refers to the (Kullback–Leibler, KL)-divergence between the distributions q(z|x) and $q(z|y,\theta)$. Because of the non-negativity of KL-divergence, we can get the variational lower bound of log $p(x|y,\theta)$ which is written as

$$\log p(x|y,\theta) \ge \int_{z} q(z|x) \log \frac{p(x,z|y,\theta)}{q(z|x)} dz$$
(3)

Therefore, the maximization of log $p(x|y, \theta)$ can be converted to maximize its variational lower bound, which is further derived as

$$\int_{z} q(z|x) \log \frac{p(x,z|y,\theta)}{q(z|x)} dz = \int_{z} q(z|x) \log \frac{p(x|y,\theta,z)p(z|y,\theta)}{q(z|x)} dz$$

$$= \int_{z} q(z|x) \log \frac{p(z|y,\theta)}{q(z|x)} dz + \int_{z} q(z|x) \log p(x|y,\theta,z) dz \qquad (4)$$

$$= -KL[q(z|x)||p(z|y,\theta)] + \int_{z} q(z|x) \log p(x|y,\theta,z) dz$$

Consider the latent variable *z* can be generated with *y*, θ by $p(z|y, \theta)$, so the term log $p(x|y, \theta, z)$ in (4) is simplified as log p(x|z). That means the latent variable *z* contains the information of target class label and target attribute, so it can be used to generate the corresponding target image. As a result, the generation problem in (1) can be re-expressed as

$$\max_{\substack{p(z|y,\theta), \\ q(z|x), \\ p(x|z)}} - KL[q(z|x) \| p(z|y,\theta)] + E_{z \sim q(z|x)} \log p(x|z)$$
(5)

In Equation (5), the expectation sign $E_{x \sim p(x)}$ is omitted for simplicity. The first term in Equation (5) aims to minimize the difference between the distribution q(z|x) and $p(z|y,\theta)$. As a result, the latent variable z produced from x will have the information of y, θ . This makes it possible for conditional image generation and the latter target classification compared with the original VAE. The purpose of the second term in Equation (5) is to maximize the probability of generating the image x from the corresponding z which is produced by q(z|x). The two losses above are called as the distribution loss and the reconstruction loss, respectively, in the following.

Since the latent variable *z* contains the target class information, it can be used as the recognition feature for target classification. Therefore, a classification loss $E_{z \sim q(z|x)} \log p(y|z)$ is added in Equation (5) and the problem is further expressed as

$$\max_{\substack{p(z|y,\theta), \\ q(z|x), \\ p(y|z)}} -KL[q(z|x)||p(z|y,\theta)] + E_{z \sim q(z|x)} \log p(x|z) + E_{z \sim q(z|x)} \log p(y|z)$$
(6)

Considering the generated images via VAE are often blurry, GAN is combined with CVAE in this paper to improve the image quality. Let $D(\cdot)$ be an extra discriminator, and then the problem of the generation model is expressed as

$$\max_{\substack{p(z|y,\theta), \\ q(z|x), \\ p(y|z)}} -KL[q(z|x) || p(z|y,\theta)] + E_{z \sim q(z|x)}[\log p(x|z)] + E_{z \sim q(z|x)}[\log p(y|z)] + E_{x \sim p(x|z)}[\log D(x)]$$
(7)
(7)

The adversarial loss for the discriminator is expressed as

$$\max_{D} L_{D} = E_{x \sim p(x)}[\log D(x)] + E_{x \sim p(x|z)}[\log(1 - D(x))]$$
(8)

Equations (7) and (8) are the final objective functions for the CVAE-GAN model, which will be solved alternately in training.

2.2. Realization of Loss Functions

The concrete form of the loss function in Equation (7) is derived in the following.

2.2.1. Distribution Loss

We suppose q(z|x) and $p(z|y,\theta)$ in the first term of Equation (7) have the normal distribution $q(z|x) = N(\mu_1(x), \sigma_1(x))$ and $p(z|y,\theta) = N(\mu_2(y,\theta), \sigma_2(y,\theta))$, respectively. The mean value $\mu_1(x)$ and standard deviation $\sigma_1(x)$ are the functions of x, while $\mu_2(y,\theta), \sigma_2(y,\theta)$ are the functions of y, θ . Different from the standard normal distribution used for all samples in the original VAE, the distributions for different samples can be various based on the target attributes. Therefore, it is possible to distinguish the samples with different target attributes in the feature space. Mark $\mu_1(x), \sigma_1(x), \mu_2(y,\theta), \sigma_2(y,\theta)$ as $\mu_1, \sigma_1, \mu_2, \sigma_2$, respectively, for simply, and the distribution loss in (7) can be expressed as

$$L_{1} = -KL[N(\mu_{1},\sigma_{1})||N(\mu_{2},\sigma_{2})]$$

$$= -\int_{x} \frac{1}{\sqrt{2\pi\sigma_{1}^{2}}} e^{-\frac{(x-\mu_{1})^{2}}{2\sigma_{1}^{2}}} \log \frac{e^{-(x-\mu_{1})^{2}/2\sigma_{1}^{2}}/\sqrt{2\pi\sigma_{1}^{2}}}{e^{-(x-\mu_{2})^{2}/2\sigma_{2}^{2}}/\sqrt{2\pi\sigma_{2}^{2}}} dx$$

$$= -\int_{x} \frac{1}{\sqrt{2\pi\sigma_{1}^{2}}} e^{-\frac{(x-\mu_{1})^{2}}{2\sigma_{1}^{2}}} \left[\log \frac{\sigma_{2}}{\sigma_{1}} + (x-\mu_{2})^{2}/2\sigma_{2}^{2} - (x-\mu_{1})^{2}/2\sigma_{1}^{2}\right]} dx$$

$$= -\log \frac{\sigma_{2}}{\sigma_{1}} - \frac{\sigma_{1}^{2} + (\mu_{1} - \mu_{2})^{2}}{2\sigma_{2}^{2}} + 0.5$$
(9)

If both σ_1 and σ_2 are alterable, the possible results of σ_1 and σ_2 may be both close to 0 in order to maximize L_1 . In this case, the randomness which is expected in VAE will have disappeared. To avoid this problem, we fix σ_2 and let $\sigma_2 = 1$ in this paper. And then (9) is revised as

$$L_1 = \log \sigma_1 - \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2} + 0.5$$
(10)

As the latent variable is multidimensional, the final loss is the sum of L_1 in all the dimensions.

2.2.2. Reconstruction Loss

Suppose q(x|z) in the second loss of Equation (7) has the normal distribution $N(\mu_3(z), \sigma_z)$, where $\mu_3(z)$ is a function of z and σ_z is a constant. Thus, the reconstruction loss is expressed as

$$L_{2} = \log p(x|z) = \log \frac{1}{\sqrt{2\pi\sigma_{z}^{2}}} e^{-(x-\mu_{3}(z))^{2}/\sigma_{z}^{2}} = -(x-\mu_{3}(z))^{2}/2\sigma_{z}^{2} - \log \sqrt{2\pi\sigma_{z}^{2}}$$
(11)

It can be seen from Equation (11) that, the maximization of L_2 is equal to the minimization of $x - \mu_3(z)^2$.

2.2.3. Classification Loss

The cross-entropy loss (CEL) is the commonly used classification loss. For the *i*th target sample and the corresponding feature z_i , whose class label is y_i , the CEL (maximization) is expressed as

$$CEL = \log \frac{e^{f_{y_i}}}{\sum_{i=1}^C e^{f_j}} \tag{12}$$

where *C* is the number of the known target classes, $f_j = W_j^T z_i + b_j$ is the score that z_i belongs to the *j*th class, W_i , b_j are the weight and bias in the classifier corresponding to the

*j*th class. However, just as stated in the introduction, there may exist some unknown classes in practice and the traditional CEL can't distinguish them from the known classes. In order to recognize the unknown classes, an idea is to incorporate the margin in the softmax loss functions, which can push the classification boundary closer to the weight vectors and add the inter-class separability. It is a popular line of research for face recognition in recent years [35–40]. In this paper, we apply the ArcFace [40] to add the angular margin for the known classes in the feature space. Let $|W_j| = 1$, $b_j = 0$ and $|z_i| = s$, s is a predetermined constant. And then the classification loss is expressed as

$$L_{3} = \log \frac{e^{f_{y_{i}}}}{e^{f_{y_{i}}} + \sum_{j=1, j \neq y_{i}}^{C} e^{f_{j}}}$$
(13)

where $f_j = s \cos(\phi_j)$, $f_{y_i} = s \cos(\phi_{y_i} + m)$, ϕ_j and ϕ_{y_i} are the angles between z_i and W_j, W_{y_i} , respectively, *m* is the additive angular margin. In Equation (13), the different classes can be separated in angle. For the simplest binary classification, the classification boundaries for Class 1 and Class 2 are $s \cos(\phi_1 + m) = s \cos(\phi_2)$ and $s \cos(\phi_1) = s \cos(\phi_2 + m)$, respectively. The schematic diagram is shown in Figure 1. For m = 0, the two classification boundaries coincide with each other and the unknown classes can't be identified, while m > 0 the two classification boundaries are separated and the unknown classes are possible to be distinguished from the known.





The fourth term in Equation (7) is realized by using the traditional GAN loss and more details can be found in [31].

2.3. Network Architecture

Based on the above model, the overall network of the proposed CVAE-GAN is shown in Figure 2a. It is composed of 5 sub-networks, i.e., the encoder networks E_1 and E_2 , the generation network G, the classification network C and the discrimination network D. E_1, E_2, G, C are expected to fit the distributions $q(z|x), p(z|y, \theta), p(x|z)$ and p(y|z) in Equation (7), respectively. More specifically, E_1 is used to realize the functions of $\mu_1(x), \sigma_1(x)$ in L_1 , E_2 is used to realize the function $\mu_2(y, \theta)$ in L_1 , and G is used to realize the function $\mu_3(z)$ in L_2 . The function of condition generation is realized by jointly utilizing the encoder network E_2 and the generation network G. By projecting the images into the feature space with E_1 and generating the images from the features with G, the attributes of the target can be explained in some extent. Besides, by using E_1 and C, we can classify the known classes and recognize the unknown classes.



(b)

Figure 2. Network architecture of CVAE-GAN: (a) Overall framework; (b) Particular architectures of sub-networks.

The particular architectures of the sub-networks used in the experiment section are shown in Figure 2b. E_1 , G and D are all modified from the popular U-Net [41] which is widely used in natural image processing. E_2 is constructed by using the multi-layer perceptron. In the figure, Conv denotes a convolution layer, while ConvT means a transposed

convolution layer. Conv(c_in , c_out , k, s, p) indicates the layer who has c_in input channels, c_out output channels and a $k \times k$ sized convolution filter with stride = s and padding = p. Linear(a_in , a_out) indicates a full-connection layer with input size a_in and output size a_out . BN and DP are the abbreviations of BatchNorm and Dropout. C is not shown in Figure 2b, which is simply constructed according to (13). The total number of parameters in CVAE-GAN is about 33.46 million.

In the network training, a reparameterization trick [29] is used to sample *z*. It aims to make the sampling step derivable, which is expressed as

2

$$z = \mu_1 + \sigma_1 \times \varepsilon \tag{14}$$

where ε is the noise with the standard normal distribution $\varepsilon \sim N(0, 1)$, $\mu_1 = \mu(x)$, $\sigma_1 = \sigma(x)$. Thus, the sampled *z* has the normal distribution $q(z|x) = N(\mu(x), \sigma(x))$. The second term in Equation (7) can be expressed in the summation form with the samples of *z* as

$$E_{z \sim q(z|x)}[\log p(x|z)] = \sum_{k=1}^{K} \log p(x|z_k), \ z_k \sim q(z|x)$$
(15)

Since the network is usually trained with mini-batch, the sample number for *z* in Equation (15) can be set as K = 1 for each iteration. Similarly, one sample of *z* is also used for $E_{z \sim q(z|x)}[\log p(y|z)]$ and $E_{x \sim p(x|z)}[\log D(x)]$ in Equation (7).

3. Experiments

In this section, the commonly used public MSTAR data set is applied to test the proposed feature learning method. There are 10 types of vehicle targets in the data set, including 2S1 (Self-Propelled Howitzer), BDRM2 (Armored Reconnaissance), BTR60 (Armored Personnel Carrier), D7 (Bulldozer), T62 (Tank), ZIL131 (Military Truck), ZSU234 (Air Defense Unit), T72 (Main Battle Tank), BMP2(Infantry Fighting Vehicle) and BTR70(Armored Personnel Carrier), which are numbered from Class 0 to Class 9 in order. The former 7 classes are selected as the known target classes and the latter 3 classes are the unknown classes. The target images of the 7 known classes with depression angle of 15° are used for training, while the images of all the 10 classes with depression angle of 17° are used for testing. In the training, inputs of the network contain the SAR image, the azimuth angle and the corresponding class label. All the SAR images are cropped to the size of 128×128 . The azimuth angle information is represented as a 2-D vector $[\cos \phi, \sin \phi]^T$, where $\phi \in 0^\circ, 360^\circ$) denotes the azimuth angle. The class labels are encoded into the 7-D one-hot vectors. The dimension of the latent variable is set as 512. The constants in the classifier are set as s = 64, m = 1 in the experiment.

In order to make the training of CVAE-GAN easier, a stepwise training strategy is applied. In the first, the loss $(L_1 + L_2)$ is utilized to train the CVAE networks E_1 , E_2 and G. Next, the classification network C together with the CVAE networks will be trained by using the loss $(L_1 + L_2 + L_3)$. In the last, the loss $L_1 + L_2 + L_3 + L_4$ and L_D will be trained adversarially to improve the clarity of generated images. The network is initialized with the default parameters in Pytorch and is trained by using the Adam optimization with a learning rate of 0.0001. All the experiments are performed on a workstation with Intel Core i9-9900 CPU and NVIDIA GeForce RTX 2080 Ti GPU.

3.1. Visualization of the Learned Feature Space

The latent variables, i.e., the extracted features, are high-dimensional of 512, which are difficult to visualize. Therefore, the t-distributed stochastic neighbor embedding (t-SNE) [42], a widely used dimension reduction technique, is utilized to embed the high-dimensional features into a 2-D feature space. Since only the angle of the feature is used for classification in Equation (13), we normalize all the features before visualization. Figure 3 visualizes the extracted features of the 7 known classes with depression angle of 15°.

Colors for Class 0 ~ Class 6 are 'blue', 'orange', 'green', 'red', 'purple', 'brown' and 'pink', respectively. It can be seen that different classes can be separated from each other easily in this feature space. The azimuth angles of the features in Figure 3 are marked in Figure 4. It shows that the feature of SAR image moves continuously in the space along with the gradual change of azimuth angle. Thus, the extracted feature is useful to explain the angle attribute of the SAR image.



Figure 3. Visualized features of 7 known classes with depression angle 15°.



Figure 4. Azimuth angles corresponding to the features in Figure 3: (a) Overall view of 7 classes; (b) Enlarged view of one class.

Furthermore, the features of the 7 known classes in the testing data with depression angle of 17° are also examined. All these extracted features are marked in gray and shown in Figure 5. It can be seen the features of the testing data have the similar distribution as that of the training data on the whole. The features of the 3 unknown classes together with the known classes are shown in Figure 6, where the unknown Class 7, 8 and 9 are marked in 'olive', 'cyan' and 'gray', respectively. We can see most samples of the unknown classes are separated from that of the known classes in the feature space, which makes it possible

to distinguish the unknown from the known. We can also see that Class 7 (T72) is close to Class 4 (T62) in the feature space. It indicates that there may be high similarity between the two classes. It is consistent with the physical truth that T72 tank approximates to T62 tank in shape and structure. The analogous conclusion can be made for Class 9 (BTR70) and Class 2 (BTR60). Therefore, the feature space is useful to explain the target attribute of SAR image for the unknown classes.



Figure 5. Visualized features of 7 known classes with depression angle and 15° and 17°.



Figure 6. Visualized features of 7 known classes and 3 unknown classes with depression angle 15°.

3.2. Results of Image Generation

In this part, the image generation ability of the proposed model is tested. Figure 7 shows the generated images with given conditions. Figure 7a is the result of one random sample. *x* is a real SAR image randomly selected from the training data set. $\mu_3(\mu_1(x))$ denotes the generated image of *x* after going through encoding $z = \mu_1(x)$ and decoding $\mu_3(z)$ successively. Take the class label *y* and the azimuth angle θ of the real SAR image *x*, and conduct encoding $z = \mu_2(y, \theta)$ and decoding $\mu_3(z)$, we can obtain the conditional generation image $\mu_3(\mu_2(y, \theta))$. It can be seen both the encoding-decoding image and the conditional generation image are very similar to the original SAR image. The additive noise is also considered for the conditional generation. The feature now is $z = \mu_2(y, \theta) + \varepsilon$,

where ε is the uniformly distributed noise within the range of $[-U_{\varepsilon}, U_{\varepsilon}]$. The noisy images with $U_{\varepsilon} = 1,3,5$ are denoted as $\mu_3(\mu_2(y,\theta) + \varepsilon_1)$, $\mu_3(\mu_2(y,\theta) + \varepsilon_2)$ and $\mu_3(\mu_2(y,\theta) + \varepsilon_3)$ in Figure 7a. We can see the randomly generated images are still similar to the original images although some slight differences exist. Some other random samples are shown in Figure 7b, which are consistent with the above result. The normalized features of the noisy images (marked in gray) are visualized in Figure 8. It demonstrates the generation ability of the continuous feature space.

Images with the azimuth angles that are absent in the training data are generated in the following. The obvious angle gaps for Class 0~6 in the training data are (83.3, 93.3), (275.5, 289.5), (200.3, 210.3), (84.5, 94.5), (83.1, 93.1) and (83.0, 93.0), respectively. Sample the angle with equal interval in the gaps and generate the SAR images with the sampled angles and class labels. The results are shown in Figure 9. Figure 9a shows the result of Class 0. $x(\theta_1)$ and $x(\theta_2)$ denote the real images with the angle $\theta_1, \theta_2, \mu_3(\mu_2(y, \theta_1))$ and $\mu_3(\mu_2(y, \theta_2))$ denote the generated images with the present angle $\theta_1, \theta_2, \mu_3(\mu_2(y, \theta'))$ denotes the generated image with the absent angle θ' and $\theta' \in \{\theta'_1, \theta'_2, \theta'_3, \theta'_4\}$ where $\theta'_1, \theta'_2, \theta'_3, \theta'_4$ are the interpolated absent angles between θ_1 and θ_2 . The results of Class 1~6 are shown Figure 9b and the visualized features of the generated images (marked in gray) are shown in Figure 10. It can be seen the generated images are changed smoothly and can well fill the angle gaps in the feature space.



Figure 7. Cont.



(b)

Figure 7. Examples of conditional generation images: (a) One sample; (b) Some other samples.



Figure 8. Visualized features of the conditional generation images.



(b)

Figure 9. Generated images with angles absent in the training data: (a) Class 0; (b) Class 1~ Class 6.

Further, the weight vectors of the classifier for each class are also visualized in the feature space. Figure 11 shows the result, where the weight vectors are marked as gray '9'. We can see the weight vectors are close to features of the samples in the corresponding classes. In the classifier, the images whose features are identical to the weight vectors will obtain the highest classification score for the corresponding class. Next, we will show the generated images by using the weight vectors as features. The feature amplitude may influence the result of the generated image. The mean amplitude of the original weight

vectors is about 198 and the mean amplitude of features in the training data is about 70. Herein, we set the weight vectors with different amplitudes of (198, 127, 81, 52). The generated images for Class 0, as an example, are shown individually in Figure 12a for clarity. The others for Class 1~6 are shown in Figure 12b. The results may reflect the most recognizable image feature for each class. For example, the generated images for Class 2 all have strong clutter background. It reflects the fact that, in some SAR images for Class 2, the clutter is strong, while for the other classes this characteristic is not obvious. Therefore, for a SAR image who has the strong clutter background, the classifier will classify it into Class 2 with high probability. According to the generated images sampled in the feature space, the actions of the classifier can be understood in some degree.



Figure 10. Visualized features of generated images in Figure 9.



Figure 11. Visualized weight vectors of the classifier.



Generated image via the weight vector

Generated image via the weight vector with amplitude 3 (81)



Generated image via the weight vector with amplitude 4 (52)







(b)

Figure 12. Generated images based on the weight vectors of the classifier.

3.3. Results of SAR Target Recognition with Unknown Classes

In this section, we will test the classification ability of the proposed model on unknown classes. The SAR images with depression angle of 17° are used for testing. There are 7 known classes with 2049 samples and 3 unknown classes with 698 samples in the test data set. As illustrated in Figure 1, the target which is located inside the classification boundary will be classified to the corresponding known class (Class 0 ~ Class 6), while the target outside all the classification boundaries of the known classes will be classified to the unknown class. The rate of the samples correctly classified in the known classes is denoted as the true positive ratio (TPR), and the rate of the samples that are falsely classified in the unknown classes is denoted as the false positive ratio (FPR). The classification boundary is adjustable in the test by using a classification threshold m_{th} . The recognition results with different m_{th} are shown in Table 1. It can be found, with the increase of m_{th} , both TPR and FPR are reduced. The overall accuracy achieves the best result 0.9236 when $m_{th} = 0.7$.

Table 1. Recognition results with different classification thresholds. TPR denotes the true positive ratio and FPR denotes the false positive ratio.

Threshold $m_{\rm th}$	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
TPR	0.9956	0.9893	0.9810	0.9702	0.9492	0.9097	0.8414	0.7252
FPR	0.7063	0.4542	0.3281	0.2264	0.1519	0.0888	0.0501	0.0143
accuracy	0.8173	0.8766	0.9024	0.9203	0.9236	0.9101	0.8689	0.7914

To further illustrate the result, the confusion matrix with $m_{\text{th}} = 0.7$ is shown in Figure 13. It can be seen there are no falsely classified samples among the 7 know classes, but only some falsely classified samples exist between the known and unknown classes. Based on the confusion matrix, the average of precision, recall and F1_score on the 8 classes can be calculated by Equation (16), where \bar{c} is the total number of classes. The precision, recall and F1_score with $m_{\text{th}} = 0.7$ are 0.9410, 0.9359 and 0.9369, respectively.

	0	288	0	0	0	0	0	0	11
	1	0	291	0	0	0	0	0	7
	2	0	0	232	0	0	0	0	23
label	3	0	0	0	297	0	0	0	2
true	4	0	0	0	0	292	0	0	7
	5	0	0	0	0	0	270	0	29
	6	0	0	0	0	0	0	274	25
	7	72	14	14	0	6	0	0	592
		0	1	2	3	4	5	6	7
				F	predicte	ed labe	1		

Figure 13. Confusion matrix with $m_{\rm th} = 0.7$.

$$precision = \frac{1}{\overline{C}} \sum_{i=0}^{\overline{C}-1} precision_i, precision_i = \frac{TP_i}{TP_i + FP_i}$$

$$recall = \frac{1}{\overline{C}} \sum_{i=0}^{\overline{C}-1} recall_i, recall_i = \frac{TP_i}{TP_i + FN_i}$$

$$F1_score = \frac{1}{\overline{C}} \sum_{i=0}^{\overline{C}-1} F1_score_i, F1_score_i = \frac{2 \times precision_i \times recall_i}{precision_i + recall_i}$$
(16)

In the above, the additive angular margin is set as m = 1. The recognition results with different *m* are also tested, which are shown in Table 2. We can see all of the accuracy, precision, recall and F1_score are improved with the increase of *m*. For m = 1.4 and

 $m_{\text{th}} = 0.9$, the overall accuracy is up to 0.9363. It should be also noted that the network will be hard to convergence as *m* is too large.

Table 2. Recognition results with different additive angular margins.

Angular Margin m	$0.5 \ (m_{\rm th} = 0.4)$	$0.8 \ (m_{\rm th} = 0.6)$	$1 (m_{\rm th} = 0.7)$	$1.2 (m_{\rm th} = 0.8)$	1.4 ($m_{\rm th} = 0.9$)
TPR	0.9370	0.9390	0.9492	0.9531	0.9590
FPR	0.1991	0.1447	0.1519	0.1461	0.1304
accuracy	0.9024	0.9177	0.9236	0.9279	0.9363
precision	0.9238	0.9382	0.9410	0.9444	0.9505
recall	0.9194	0.9278	0.9359	0.9400	0.9469
F1_score	0.9194	0.9316	0.9369	0.9401	0.9473

4. Discussion

In this section, we will discuss the above results compared with the existing studies.

With regard to the explainable feature space for unknown SAR target recognition, the related studies have been done in [23,24]. The feature spaces built in these papers only represent the categories of the target, but can't show the other target attributes, such as the azimuth angle, effectively. It can be seen clearly from the Figures 9 and 11 in [23] and Figures 6 and 7 in [24]. In this paper, the target attributes including the azimuth angle can be explained in the continuous feature space, which is useful to understand the result of deep network in some extent and benefits for the conditional image generation.

About the image generation, the network proposed in [23] is able to generate the SAR image with the given class and angle. However, the generated images are blurry as shown in Figures 5 and 7 of [23]. Besides, some variations of GAN have been proposed to generate the images with given class labels [43,44]. Herein, we also test this kind of methods on SAR image generation. Figure 14 shows the generated SAR images with class labels by using the conditional GAN in [44]. It can be seen the generated images are clear. However, as well as we know, the conditional GANs only deal with the concrete conditions well, such as the class label, but don't work for the continuous conditions at present, such as the azimuth angle. Moreover, the conditional VAE, which is without the discriminator compared with the proposed CVAE-GAN, is also tested. The result is shown in Figure 15. We can see it is able to deal with the continuous angle, but the generated images are not as clear as the conditional GANs. The proposed CVAE-GAN performs well on both the image quality and the conditional generation, as shown in Figures 7, 9 and 12.

With regard to the SAR target recognition with unknown classes, the open set recognition method proposed in [28] is also tested. The test condition is the same as that in Section 3.3. The network is built according the Figure 1 in [28] and the input images are cropped into the size of 64×64 as [28] did. The network is trained in 300 epochs and all the networks with different epochs are used for the OSR test. The best result is obtained by using the network in epoch 275 and with the threshold 0.1. The confusion matrix is shown in Figure 16 and the recognition metrics are shown in Table 3. By comparing with the results in Section 3.3, we can see the proposed method has the better performance for OSR.



Figure 14. Generated images with conditional GAN.

	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
angle 1		ţra	6/2	114	40	4	*
angle 2	Carak	iyente			خليو:	منتلة	in-
angle 3	1	4	J	6. ¹	d d	4	si.
angle 4	4	*		j.	1	and the second s	*
angle 5			607		.74	4	
angle 6	have	2-4-	and a	أهري		1 miles	24
angle 7	Aug	-	67.5	and the	ing.	1	-

Figure 15. Generated images with conditional VAE.

	0	277	0	1	0	0	0	0	21
	1	0	287	1	1	0	0	0	9
	2	0	0	242	0	0	0	0	13
label	3	0	0	0	295	0	0	0	4
true	4	1	0	0	0	285	0	0	13
	5	0	0	0	0	0	299	0	0
	6	0	0	0	0	0	0	270	29
	7	42	7	106	0	18	0	0	525
		0	1	2	3	4	5	6	7
				r	redict	ed labe	1		

Figure 16. Confusion matrix by using the OSR method in [28].

TPR	FPR	Accuracy	Precision	Recall	F1_score
0.9546	0.2479	0.9031	0.9157	0.9292	0.9190

Table 3. Recognition results with the method in [28].

5. Conclusions

In this paper, a deep generation and recognition network is proposed for SAR-ATR based on CVAE-GAN model. The proposed network can form a feature space, in which both the conditional generation task and the open-set recognition task can be realized. More importantly, some actions of the deep network can be explained in the feature space. In the next work, we will pay attention to the explainable feature space which can show the confidence coefficient for classification. Moreover, to classify the unknown classes with the help of some aided target information is also a meaningful job in the further. The unbalanced data is a critical issue for the classification of unknown classes.

Author Contributions: X.H. and Y.G. proposed the model and network; W.F. and Q.W. conceived and designed the experiments; X.H. and W.F. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the National Natural Science Foundation of China under Grants 61701526, 62001507 and 62001479, by the China Postdoctoral Science Foundation under Grant 2019M661356, and by the School Scientific Research Program of National University of Defense Technology ZK19-28.

Data Availability Statement: The public MSTAR data set used in this paper can be found here: https://www.sdms.afrl.af.mil/datasets/mstar (accessed on 26 July 2021).

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

References

- 1. Tait, P. Introduction to Radar Target Recognition; IEE Radar Series; The Institution of Engineering and Technology: London, UK, 2005.
- NovakL., M.; Benitz, G.R.; Owirka, G.J.; Bessette, L.A. ATR performance using enhanced resolution SAR. In Proceedings of the SPIE Conference on Algorithms for Synthetic Aperture Radar Imagery III, Orlando, FL, USA, 22 March 1996; pp. 332–337.
- Mishra, A. Validation of PCA and LDA for SAR ATR. In Proceedings of the IEEE Region 10 Conference, Hyderabad, India, 19–21 November 2008; pp. 1–6.
- 4. Pei, J.; Huang, Y.; Huo, W.; Wu, J.; Yang, J.; Yang, H. SAR Imagery Feature Extraction Using 2DPCA-Based Two-Dimensional Neighborhood Virtual Points Discriminant Embedding. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2016, *9*, 2206–2214. [CrossRef]
- 5. Zhou, J.; Shi, Z.; Cheng, X.; Fu, Q. Automatic Target Recognition of SAR Images Based on Global Scattering Center Model. *IEEE Trans. Geosci. Remote Sens.* 2011, 49, 3713–3729.
- Li, T.; Du, L. SAR Automatic Target Recognition Based on Attribute Scattering Center Model and Discriminative Dictionary Learning. *IEEE Sensors J.* 2019, 19, 4598–4611. [CrossRef]
- Sun, Y.; Du, L.; Wang, Y.; Wang, Y.H.; Hu, J. SAR Automatic Target Recognition Based on Dictionary Learning and Joint Dynamic Sparse Representation. *IEEE Geosci. Remote Sens. Lett.* 2016, 13, 1777–1781. [CrossRef]
- 8. Zhou, Z.; Cao, Z.; Pi, Y. Subdictionary-Based Joint Sparse Representation for SAR Target Recognition Using Multilevel Reconstruction. *IEEE Trans. Geosci. Remote Sens.* 2019, 57, 6877–6887. [CrossRef]
- 9. Clemente, C.; Pallotta, L.; Gaglione, D.; Maio, A.D.; Soraghan, J.J. Automatic Target Recognition of Military Vehicles with Krawtchouk Moments. *IEEE Trans. Aerosp. Electron. Syst.* 2017, 53, 493–500. [CrossRef]
- 10. Srinivas, U.; Monga, V.; Raj, R.G. SAR Automatic Target Recognition Using Discriminative Graphical Models. *IEEE Trans. Aerosp. Electron. Syst.* 2014, *50*, 591–606. [CrossRef]
- 11. 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]
- 12. Huang, Z.; Dumitru, C.O.; Pan, Z.; Lei, B.; Datcu, M. Classification of Large-Scale High-Resolution SAR Images with Deep Transfer Learning. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 107–111. [CrossRef]
- 13. Huang, X.; Yang, Q.; Qiao, H. Lightweight Two-Stream Convolutional Neural Network for SAR Target Recognition. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 667–671. [CrossRef]
- 14. Huang, Z.; Pan, Z.; Lei, B. What, Where, and How to Transfer in SAR Target Recognition Based on Deep CNNs. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 2324–2336. [CrossRef]
- 15. Huang, Z.; Pan, Z.; Lei, B. Transfer Learning with Deep Convolutional Neural Network for SAR Target Classification with Limited Labeled Data. *Remote Sens.* 2017, *9*, 907. [CrossRef]

- 16. Malmgren-Hansen, D.; Kusk, A.; Dall, J.; Nielsen, A.A.; Engholm, R.; Skriver, H. Improving SAR Automatic Target Recognition Models with Transfer Learning from Simulated Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1484–1488. [CrossRef]
- Cha, M.; Majumdar, A.; Kung, H.T.; Barber, J. Improving sar automatic target recognition using simulated images under deep residual refinements. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 2606–2610.
- 18. Liu, L.; Pan, Z.; Qiu, X.; Peng, L. SAR target classification with CycleGAN transferred simulated samples. *IEEE Int. Geosci. Remote Sens. Symp.* **2018**, 4411–4414.
- 19. Sun, Y.; Wang, Y.; Liu, H.; Wang, N.; Wang, J. SAR Target Recognition with Limited Training Data Based on Angular Rotation Generative Network. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 1928–1932. [CrossRef]
- Song, Q.; Xu, F.; Jin, Y.Q. SAR Image Representation Learning with Adversarial Autoencoder Networks. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 9498–9501.
- Toizumi, T.; Sagi, K.; Senda, Y. Automatic association between SAR and optical images based on zero-shot learning. In Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 17–20.
- 22. Song, Q.; Chen, H.; Xu, F.; Cui, T.J. EM simulation-aided zero-shot learning for SAR automatic target recognition. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1092–1096. [CrossRef]
- 23. Song, Q.; Xu, F. Zero-shot learning of SAR target feature space with deep generative neural networks. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2245–2249. [CrossRef]
- 24. Wei, Q.R.; He, H.; Zhao, Y.; Li, J.-A. Learn to Recognize Unknown SAR Targets From Reflection Similarity. *IEEE Geosci. Remote Sens. Lett.* **2021**, 1–5. [CrossRef]
- 25. Scherreik, M.; Rigling, B. Multi-class open set recognition for SAR imagery. In Proceedings of the Automatic Target Recognition XXVI, Baltimore, MD, USA, 18–19 April 2016; Volume 9844.
- 26. Dang, S.; Cao, Z.; Cui, Z.; Pi, Y.; Liu, N. Open set incremental learning for automatic target recognition. *IEEE Trans. Geosci. Remote Sens.* 2019, *57*, 4445–4456. [CrossRef]
- 27. Dang, S.; Cao, Z.; Cui, Z.; Pi, Y. Open set SAR target recognition using class boundary extracting. In Proceedings of the 6th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), Xiamen, China, 26–29 November 2019; pp. 1–4.
- 28. Ma, X.; Ji, K.; Zhang, L.; Feng, S.; Xiong, B.; Kuang, G. An Open Set Recognition Method for SAR Targets Based on Multitask Learning. *IEEE Geosci. Remote Sens. Lett.* **2021**, 1–5. [CrossRef]
- 29. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. arXiv 2013, arXiv:1312.6114v10.
- 30. Rezende, D.J.; Mohamed, S.; Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. *arXiv* **2014**, arXiv:1401.4082.
- 31. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *arXiv* **2014**, arXiv:1406.2661, 2672–2680.
- 32. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
- 33. Larsen, A.B.L.; Sonderby, S.K.; Larochelle, H.; Winther, O. Autoencoding beyond pixels using a learned similarity metric. *Int. Conf. Int. Conf. Mach. Learn.* **2016**, *48*, 1558–1566.
- 34. Bao, J.; Chen, D.; Wen, F.; Li, H.; Hua, G. CVAE-GAN: Fine-grained image generation through asymmetric training. *arXiv* 2017, arXiv:1703.10155, 2745–2754.
- 35. Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A Discriminative Feature Learning Approach for Deep Face Recognition. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 499–515.
- 36. Wang, F.; Xiang, X.; Cheng, J.; Yuille, A.L. NormFace: L2 hypersphere embedding for face verification. In Proceedings of the ACM Multimedia Conference, Mountain View, CA, USA, 23–27 October 2017.
- 37. Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. SphereFace: Deep Hypersphere Embedding for Face Recognition. *arXiv* 2017, arXiv:1704.08063, 6738–6746.
- 38. Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; Liu, W. Cosface: Large margin cosine loss for deep face recognition. *arXiv* 2018, arXiv:1801.09414, 5265–5274.
- 39. Wang, F.; Cheng, J.; Liu, W.; Liu, H. Additive margin softmax for face verification. IEEE Sig. Proc. Lett. 2018, 25, 926–930. [CrossRef]
- 40. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. ArcFace: Additive angular margin loss for deep face recognition. *arXiv* 2018, arXiv:1801.07698.
- Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
- 42. Maaten, L.V.D.; Hinton, G. Visualizing high-dimensional data using t-SNE. J. Mach. Learn. Res. 2008, 9, 2579–2605.
- 43. Mirza, M.; Osindero, S. Conditional generative adversarial nets. arXiv 2014, arXiv:1411.1784.
- 44. Odena, A.; Olah, C.; Shlens, J. Conditional image synthesis with auxiliary classifier GANs. In Proceedings of the 34th International Conference on Machine Learning, Sydney, NSW, Australia, 6–11 August 2017; Volume 70, pp. 2642–2651.