Article

A THz Passive Image Generation Method Based on Generative Adversarial Networks

Guan Yang ^{1,2,3,*}, Chao Li ^{1,2,3,*} , Xiaojun Liu ^{1,2} and Guangyou Fang ^{1,2,3}

¹ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China; yangguan17@mails.ucas.edu.cn (G.Y.); lxjdr@mail.ie.ac.cn (X.L.); gyfang@mail.ie.ac.cn (G.F.)

² Key Laboratory of Electromagnetic Radiation and Sensing Technology, Chinese Academy of Sciences, Beijing 100190, China

³ School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100190, China

* Correspondence: cli@mail.ie.ac.cn; Tel.: +86-010-5888-7238

Abstract: A terahertz (THz) passive imager with automatic target detection is an effective solution in the field of security inspection. The high-quality training datasets always play a key role in the high-precision target detection applications. However, due to the difficulty of passive image data acquisition and the lack of public dataset resources, the high-quality training datasets are often insufficient. The generative adversarial network (GAN) is an effective method for data augmentation. To enrich the dataset with the generated images, it is necessary to ensure that the generated images have high quality, good diversity, and correct category information. In this paper, a GAN-based generation model is proposed to generate terahertz passive images. By applying different residual connection structures in the generator and discriminator, the models have strong feature extracting ability. Additionally, the Wasserstein loss function with gradient penalty is used to maintain training stability. The self-developed 0.2 THz band passive imager is used to carry out imaging experiments, and the imaging results are collected as a dataset to verify the proposed method. Finally, a quality evaluation method suitable for THz passive image generation task is proposed, and classification tests are performed on the generated images. The results show that the proposed method can provide high-quality images as supplementary.

Keywords: terahertz passive images; target detection; data augmentation; generative adversarial network

Citation: Yang, G.; Li, C.; Liu, X.; Fang, G. A THz Passive Image Generation Method Based on Generative Adversarial Networks. *Appl. Sci.* **2022**, *12*, 1976. <https://doi.org/10.3390/app12041976>

Academic Editor: Mira Naftaly

Received: 25 January 2022

Accepted: 8 February 2022

Published: 14 February 2022

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

Copyright: © 2022 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

Terahertz (THz) waves can easily penetrate clothing; compared to X-rays, its photon energy is relatively low, so it is non-ionizing and harmless to the biological tissue. Compared to the microwave band, it can achieve a higher imaging resolution [1,2]. Due to the above advantages, the THz imaging systems have attracted extensive attention in the field of security inspection, non-destructive testing (NDT) and other applications [3,4].

In the field of security inspection, THz imaging can be divided into active imaging and passive imaging. The active imaging systems apply the transmitter to emit electromagnetic waves in the terahertz band and use the echoes of the target for image reconstruction. The passive systems apply the natural radiation emitted by the targets. Thus, active and passive imaging systems show different advantages in imaging performance and scene visualization [5]. For the hidden object detection task, the radiation brightness temperature between the human body and the hidden object always has a significant difference, so the target boundary is more obvious in the terahertz passive image, which is beneficial to the target detection task. Furthermore, passive systems have some other advantages in terms of safety concerns and privacy issues in security inspection applications.

The dataset always plays a key role in model training and generalization performance. As we all know, THz security inspection equipment has just begun to be set in a few security inspection scenarios. At present, there is no public dataset available for training models to identify specific dangerous targets, and this dataset involves sensitive issues such as public safety. Therefore, using data augmentation technology to fully mine the existing dataset is a necessary means to improve the performance of the model. Most of the existing data augmentation techniques involve random cropping, splicing, rotation and other operations on the original image during the training process [6,7]. However, such technologies cannot generate some unseen images and do not fundamentally solve the problem. Especially in the case of few-shot learning, the improvement of network performance is limited.

With the continuous progress of deep-learning algorithms, image generation methods based on deep convolutional neural networks provide a solution for data augmentation. The generative adversarial network (GAN) [8] is a typical image generation method. GAN models learn feature distribution of real data and improve the quality of generated images through game learning between the generator and the discriminator. The early GAN models performed feature extraction through the fully connected layer, which ignored the spatial correlation of the image. Thus, the quality of the generated image was poor. Then the deep convolutional network structure was introduced into the GAN model (DCGAN) [9], which improved the quality of generated image. Subsequently, to generate images of multiple categories during one training process, a supervised learning GAN with image classification information is proposed [10]. However, the introduction of classification labels combined with the dynamic training process has a negative impact on the generated image categories [11].

In addition, the abovementioned GAN model uses the Jensen–Shannon (JS) divergence to measure the difference between the generated image and the real image. The JS divergence loss function will be a constant in most cases, resulting in invalid gradient backpropagation and unstable training. Then, Arjovsky et al. used the Wasserstein distance with weight clipping to replace the JS divergence, and proposed the WGAN model [12], which improved the training stability. Then, Gulrajani et al. proposed an improved WGAN model by using gradient penalty to replace the weight clipping operation in WGAN, named WGAN-GP, which avoided the problem of extreme weight parameter distribution due to weight clipping [13]. In the abovementioned GAN models, simple stacked deep convolutional network structures are used. Generally speaking, to enhance the feature extraction ability of the model, the most direct way is to increase the network depth, but some studies found that blindly increasing the network depth will cause the problem of network degradation.

To enhance the feature-learning ability of deep convolutional network, avoid network degradation and vanishing gradient problems in the deep network structure, a generative adversarial network model using residual structure [14] is proposed in this paper, named Res-WGAN-GP. Moreover, the loss function of WGAN-GP is used to the proposed model to maintain the training stability. According to the special requirements of the loss function, different residual block structures are designed for the generator and discriminator. In addition, a quality evaluation method suitable for a THz passive image generation task is proposed. Finally, the proposed model can generate high-quality THz passive images to augment the dataset. The main contribution of this paper is the first attempt to apply deep-learning technology to achieve low-cost terahertz passive image data augmentation, which aims to help the further application of terahertz passive imaging systems in the field of security inspection.

This rest of paper is organized as follows. A detailed description of the Res-WGAN-GP model is given in Section 2. Then the experimental results based on 0.2 THz band passive imaging dataset are given in Section 3. Then, the resulting analyses of the generated images are discussed in Section 4. Finally, a conclusion is drawn in Section 5.

2. Methodology

In this section, the overall framework and network architecture of the Res-WGAN-GP model will be introduced in detail.

2.1. Overall Framework

To avoid the instability of training caused by category labels, an unsupervised GAN model is proposed to generate high-quality THz passive images by improving the model structure. Figure 1 shows the overall framework of the proposed Res-WGAN-GP. There are two deep convolutional networks in the model: a generator (G) and a discriminator (D). The generator can be thought of as analogous to a counterfeiter, trying to make counterfeit money and use it without detection, while the discriminator is analogous to the police trying to detect counterfeit currency. The competition in this game pushed both sides to refine their methods until the fakes were indistinguishable from the genuine articles [8].

Figure 1. Overall framework of Res-WGAN-GP.

Specifically, the generator model converts the input of a certain distribution to the output closest to the real sample distribution by game learning. Here, the normal distribution random noise is employed as the input of the generator. Random noise is used to ensure that the images generated by the generator are different each time.

Then, the generated images and real images are mixed and sent to the discriminator. The discriminator is equivalent to a binary classifier, which is used to determine which of the mixed input images are real and which are fake. Then, the discriminator feeds the judgment result to the generator to improve the quality of the generated images through the continuous adversarial process.

Early GAN models suffer from training instability [13]. To solve the problem, the loss function based on Wasserstein distance was proposed in WGAN model [12] to improve the training stability. In the WGAN model, the weight of each layer of the discriminator is controlled by a fixed threshold to make it satisfy the Lipschitz constraint, which acts on all possible input spaces. The loss functions of the discriminator (D) and generator (G) in the WGAN model are expressed as

$$\begin{cases} \text{Loss}_D = E_{\tilde{x} \sim P_G} [D(\tilde{x})] - E_{x \sim P_{\text{data}}} [D(x)] \\ \text{Loss}_G = -E_{\tilde{x} \sim P_G} [D(\tilde{x})] \end{cases} \quad (1)$$

where P_{data} is the distribution of real data; P_G denotes the distribution of generated image data. $\tilde{x} = G(z)$ is the output of the generator, z denotes the random noise vector.

Such a discriminator in the WGAN model would be inclined to learn a simple mapping function and cause the problem of the gradient vanishing or exploding due to the weight clipping. In order to avoid this problem, the loss function with a gradient penalty was proposed in the WGAN-GP model [13]. Different from the sample space which the WGAN model focuses on, the WGAN-GP model only focuses on the region in the generated sample set region, the real sample set region, and the regions in between. Specifically, random sampling produces a pair of true and false samples x_{data} , $\tilde{x}_G$, and then generates a random number ε between 0 and 1.

$$x_{data} \sim P_{data}, \tilde{x}_G \sim P_G, \varepsilon \sim Uniform[0, 1] \quad (2)$$

Then random interpolation sampling is carried out on the line between x_{data} and $\tilde{x}_G$ to obtain random data $\hat{x}$

$$\hat{x} = \varepsilon x_{data} + (1 - \varepsilon) \tilde{x}_G \quad (3)$$

The distribution of $\hat{x}$ is defined as $P_{\hat{x}}$. In WGAN-GP model, the gradient of the discriminator is limited to no more than 1. The discriminator hopes to increase the score difference between true and false samples, so after the discriminator is fully trained, the gradient value of the gradient will be around 1. Therefore, the added gradient penalty loss term can be set as the modulus of the discriminator gradient minus 1. Then, the improved discriminator loss function can be expressed as

$$Loss_D_{gp} = E_{\tilde{x} \sim P_G} [D(\tilde{x})] - E_{x \sim P_{data}} [D(x)] + \lambda E_{\hat{x} \sim P_{\hat{x}}} [\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1]^2 \quad (4)$$

In summary, the generator and discriminator loss functions used in the WGAN-GP model are

$$\begin{cases} Loss_G = -E_{\tilde{x} \sim P_G} [D(\tilde{x})] \\ Loss_D_{gp} = E_{\tilde{x} \sim P_G} [D(\tilde{x})] - E_{x \sim P_{data}} [D(x)] + \lambda E_{\hat{x} \sim P_{\hat{x}}} [\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1]^2 \end{cases} \quad (5)$$

The last term of the discriminator loss function is the added gradient penalty term. The penalty coefficient λ is set to 10 according to the practice. The abovementioned loss function (5) is also employed as the loss function in our model, so the WGAN-GP model will be a baseline for comparison in the experiment.

2.2. Network Architecture

The most direct way to improve network performance includes increasing network depth, but after research by many scholars, it was found that with the deepening of the network, the fitting effect of the model actually declined. This phenomenon is called the network degradation problem. From the perspective of information theory, the possible reason for this phenomenon is due to the existence of the “data processing inequality”, that is, in the process of forward propagation, as the network deepens, the original image information contained in the feature map will decrease layer by layer. The neural network is not easy to fit the identity map due to nonlinear activation. He K. M. et al. [14] proposed the concept of residual learning, which well alleviated the degradation problem of deep networks. The main contribution was to construct a natural identity mapping relationship between neural network layers through the proposed residual module.

According to the basic idea of residual learning, the detailed structures of the proposed generator and discriminator networks are shown in Figure 2. The numbers above each color block represent the number of output channels.

Figure 2. The proposed Res-WGAN-GP module structure.

In the generator structure, a residual block named “ResBlock Upsample” is designed to achieve feature map upsampling and a standard residual block “ResBlock” which does not change the size of input feature map. The detailed structures are shown in Figure 3a,b, respectively. The “ResBlock Upsample” block includes the convolution layer (Conv), batch normalization (BN) [15] layer, leaky rectified linear unit (LeakyReLU) activation function [16] and deconvolution (Deconv) layer [17]. The output feature map size of each operation is listed next to each color block. In the “ResBlock Upsample” block, a deconvolution layer is used to achieve double upsampling of the input feature map. The dimension of the input random noise vector is set to 256. After the fully connected layer, the input noise is mapped to the initial feature map size of 8×8 . Then, the output image of 256×256 is obtained after 5 times of the $2 \times$ upsampling process.

Figure 3. Residual structure in generator. (a) “ResBlock Upsample” block. (b) Traditional residual block.

In the discriminator, the traditional convolution operation with stride 2 is used to reduce the size of the input feature map. In the discriminator loss function, the gradient penalty term penalizes each sample independently, and the BN layer will introduce correlation between samples. Thus, the BN layer cannot be used in the discriminator. The residual structures used in the discriminator are shown in Figure 4. In addition, in the process of GAN training, the discriminator network is easier to train than the generator. An over optimum discriminator makes the training of the generator harder, and even fail [11]. Therefore, the structure of the discriminator is designed to be simpler.

Figure 4. Residual structures used in discriminator. (a) "ResBlock Downsample" block. (b) Traditional residual block.

3. Experiments

In this section, the practical performance of the proposed model will be discussed. First, the acquisition method of the dataset used for model training is introduced; then, the model training strategy is given; finally, the model performance is comprehensively evaluated.

3.1. Dataset

As mentioned in Section 1, terahertz passive imaging is a cutting-edge security technology and has not been widely used yet, so there are no public dataset resources at present. To verify the method proposed in this paper, the self-developed 0.2 THz band passive imager which combined a simple quasi-optical scanning and a single-channel radiometric receiver was used for imaging experiments (the schematic is shown in Figure 5). A two-dimensional (2D) field of view (FOV) is obtained based on a raster scan of the beam spot along the X and Y directions. The sub-reflector is a rotating one with a small size, capable of steering the beam quickly in a horizontal X-direction through its rotation around the Y-axis. A concave main reflector is employed to focus the terahertz signal emitted from the target to the feed horn of the radiometric receiver. The imager operates at a center frequency of 0.2 THz with a bandwidth of 30 GHz.

Figure 5. Simple architecture of the imager.

A pistol model and a mobile phone are used as the hidden objects. The imager and the imaging results are shown in Figure 6.

Figure 6. Typical terahertz passive imaging system. (a) The passive imaging system. (b) Imaging targets. (c) The person to be tested. (d) Passive imaging results.

As shown in Figure 7, there are three categories of images in the dataset: carrying a mobile phone, carrying a pistol model, and carrying above two objects. After a long-term imaging experiment, a total of 1800 images of three categories are collected. The experimenters include different body types, and the number of each category is almost equal. The original size of the terahertz passive image is a rectangle of 392×192 and the physical size corresponding to one pixel is 5 cm. For the convenience of model design, the original images need to be preprocessed. Firstly, the original image is zero-padded to obtain a 392×392 square image. There are no valid pixels in the zero-padded area, so it does not affect the network performance. Then, the filled images are scaled to 256×256 as the final dataset.

Figure 7. Samples of THz passive image dataset. (a) Category 1: carrying a gun model. (b) Category 2: carrying a cell phone. (c) Category 3: carrying two objects.

3.2. Training Strategy

In the training process, the discriminator is first trained for five iterations, and the parameters of the generator are fixed at this time. Then, the discriminator parameters are fixed to train the generator. Game learning is achieved through alternating the optimization of the network.

Each complete training process uses only a single type of image data; that is, the unsupervised method is used to train three models to generate three types of terahertz images, and the parameters of the three models are saved separately. Since there are only three types of images at present, the separate training process is not complicated, and it is easier to obtain high-quality generated images without the interference of label information.

The optimizer used in the generator and discriminator is an Adam optimizer [18], the learning rate is set to 0.0001, the batch size is 4, and the beta1 and beta2 parameters in Adam are set to 0.5 and 0.999, respectively.

The experiment environments for model training are listed in Table 1.

Table 1. The experiment environments.

Environments	Environment Configuration
Software Development Environment	Windows 10
	Python 3.7.9
	PyTorch 1.7.1
	CUDA 10.1
	CuDNN 7.6.5
Hardware Environment	CPU: Intel core i3-8100
	GPU: Nvidia 2080Ti
	RAM: 32 GB

4. Results Analyses

In this section, the performance of the proposed Res-WGAN-GP model and the original WGAN-GP model on the terahertz passive image generation task is evaluated and compared. It is worth noting that due to the particularity of terahertz passive images, there is no public dataset; and to the best of our knowledge, there are currently no research reports on terahertz passive image generation methods, which means that no general evaluation method for terahertz passive image generation can be directly applied. In addition, the Inception Score [19] and Fréchet Inception Distance [20] evaluation metrics, commonly used in the optical imagery field, cannot be used in the terahertz passive image generation task. Thus, this section draws on relevant literature to propose an objective evaluation method suitable for the current task.

The proposed GAN model aims to provide data augmentation for object detection tasks; thus, we hope that the generated images meet the following requirements:

- (1) Visual quality: the generated images should be high-quality;
- (2) Category consistency: the generated image must represent the desired class;
- (3) Diversity: the generated images must not be repetitive;
- (4) Usability: the generated images must be different from the real images already in the training set.

The proposed Res-WGAN-GP model and the original WGAN-GP model were trained three times to obtain the generative models of the three types of images, and then eight groups of random noise inputs were used to obtain the generated images. The comparison is shown in Figure 8. In terms of visual effects, the images generated by the proposed model are comparable in quality to real images, and basically meet the first requirement. The images generated by the original WGAN-GP model are mostly blurry, the desired object area is not clear enough or even missing, which is quite different from the real dataset.

Figure 8. Three types of terahertz images generated by the two generators. (a) carrying a gun model generated by Res-WGAN-GP. (b) Carrying a gun model generated by WGAN-GP. (c) Carrying a cell phone generated by Res-WGAN-GP. (d) Carrying a cell phone generated by WGAN-GP. (e) Carrying two objects generated by Res-WGAN-GP. (f) Carrying two objects generated by WGAN-GP.

To verify that the three separately trained models can accurately generate realistic images of the corresponding classes, the cross and hybrid tests are performed, respectively [11]. First, the original dataset is classified according to categories and then train a classification model. Here, the GoogLeNet [21] model is used as the pretrained classification model. Then, 100 fake images for each category are used as the test set. The classification accuracy is the result of the cross test. Secondly, a hybrid test is performed. A batch of generated images is mixed into the real dataset for retraining, and then another batch of generated images is used as a test set to verify whether the classification performance of the network has been improved, which is closer to the actual application scenario. The statistical results of the cross test and the hybrid test are shown in Table 2. It can be seen that the images generated by Res-WGAN-GP model has higher classification accuracy than the original WGAN-GP model, which indicates that the generated images basically belong to the desired category and satisfy the second requirement.

Table 2. The performance of the generated image on the classification model.

Model	Cross Test	Hybrid Test
WGAN-GP	73.7%	90.3%
Res-WGAN-GP	92.7%	96.7%

In order to verify the diversity of generated images, that is, that the model does not generate one or several images repeatedly, the SSIM indicator [22] is used to evaluate the similarity between the generated images. The calculation formula of SSIM can be expressed as

$$\text{SSIM}(g_1, g_0) = \frac{(2\mu_1\mu_0 + C_1)(2\sigma_{1,0} + C_2)}{(\mu_1^2 + \mu_0^2 + C_1)(\sigma_1^2 + \sigma_0^2 + C_2)} \quad (6)$$

where g_0 is the reference image; g_1 denotes the image to be evaluated. $(\mu_1, \mu_0), (\sigma_1, \sigma_0)$ and $\sigma_{1,0}$ represent the mean, variance and covariance of the two images, respectively. C_1 and C_2 are constants, which are used to ensure the stability of the calculation process and can be expressed as

$$C_1 = (K_1 L)^2, C_2 = (K_2 L)^2 \quad (7)$$

where $K_1 = 0.01$, $K_2 = 0.03$ and $L = 255$.

Generally, SSIM is a value between 0 and 1. The larger the SSIM value, the higher the similarity between the image to be evaluated and the reference image. If the two images are exactly the same, the SSIM is equal to 1.

Specifically, a batch of images was generated for each category, and randomly selected a pair of images to calculate the SSIM value. If the SSIM of the two is equal to 1, it means there is a repeated generation. On the contrary, if the SSIM value is relatively small, it means that the generated images are obviously different. In this paper, 400 images for each category are generated, 100 pairs of generated images were randomly extracted from the current category to calculate the SSIM values, and its mean value is defined as the interclass SSIM. If the interclass value is small, it means that the diversity of the generated images is better. Then, the average SSIM value of the real image in the same verification method is calculated as a reference; the real image is more variable than the generated image, that is, it usually corresponds to a smaller interclass SSIM value [23]. The test results are shown in Figure 9. The average intraclass SSIM of the three types of real data is 0.287, indicating that real data has good diversity. The average intraclass SSIM of the three types of samples generated by Res-WGAN-GP was 0.323, which means have a good diversity. In addition, the trend of the intra-class SSIM values of the three types of samples is similar to that of the real samples, indicating that the data generated by the proposed model has the same category information as the real samples. Although the test results of the images generated by WGAN-GP have very low SSIM values, even lower than the real samples, this is due to the lack of network learning ability, resulting in the overall low quality of the generated

images. It is manifested that the gray distribution is uneven, and the difference between the generated images is large, which can also be seen from the visual comparison of the images in Figure 8. In summary, the image generated by the proposed model satisfies the third requirement.

Figure 9. Diversity evaluation by interclass SSIM.

For the fourth requirement, generally speaking, the input of the GAN model is random noise sampled from a normal distribution. Although the model has a strong learning ability, the probability of completely fitting the real data distribution is very small. Therefore, the phenomenon of overfitting is not common in the GAN field, but in order to avoid this unexpected situation and generate useless data, the SSIM indicator is also used for evaluation. Specifically, 100 pictures are generated for each of the 3 categories, and then the SSIM values are calculated for all the generated pictures and all the real pictures to obtain the maximum SSIM value. If it is less than 1, it means that the generated pictures and the real dataset do not overlap. After calculation, the maximum SSIM values of each category are 0.77, 0.81 and 0.73, respectively. Therefore, it can be judged that there is no rare overfitting phenomenon in the model, and the generated data can be used for data enhancement of target detection after simple screening and eliminating abnormal conditions.

5. Conclusions

In this paper, a Res-WGAN-GP generation model and a quality evaluation method are proposed for THz passive image data augmentation application. Based on the framework of deep convolutional generative adversarial network, the generator and discriminator models based on residual structure are designed. The WGAN-GP loss function is used to ensure the stability of the training process. The generated images are evaluated in terms of visual quality, category consistency, diversity, and usability. The results show that the proposed model can improve the quality of generated images and meet the requirements of data augmentation applications. In the classification tests, the classification accuracy can be improved by applying the augmented dataset, so it is expected to apply the augmented data in the object detection task with more target categories.

It is worth noting that the main contribution of this paper is the first attempt to use deep-learning methods for data augmentation of terahertz passive image datasets. This paper aims to provide a low-cost solution for poor detection performance caused by insufficient data volume.

In the future, the network structure, training strategy and loss function need to be optimized, and the proposed method also needs to be verified with a dataset containing more categories. In addition, some effective clarity evaluation methods suitable for GAN models need to be investigated.

Author Contributions: G.Y. performed theoretical study, conducted the experiments, processed the data and wrote the manuscript. C.L. designed the imaging system and revised the manuscript. X.L. and G.F. provided the experiment equipment and funds for the research. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key-Area Research and Development Program of Guangdong Province (2019B010157001, 2020B0101110001); the National Natural Science Foundation of China (61988102, 61731020, 61671432); the National Key Research and Development Program of China (2017YFA0701004, 2018YFF01013004); the Project of Equipment Pre-Research (WJ2019G00019); and the Key Program of Scientific and Technological Innovation from Chinese Academy of Sciences (KGFZD-135-18-029).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the editors and reviewers for their efforts to help the publication of this work.

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

References

1. Yang, G.; Li, C.; Wu, S.; Liu, X.; Fang, G. MIMO-SAR 3-D Imaging Based on Range Wavenumber Decomposing. *IEEE Sens. J.* **2021**, *21*, 24309. [CrossRef]
2. Yang, G.; Li, C.; Wu, S.; Zheng, S.; Liu, X.; Fang, G. Phase Shift Migration with Modified Coherent Factor Algorithm for MIMO-SAR 3D Imaging in THz Band. *Remote Sens.* **2021**, *13*, 4701. [CrossRef]
3. Liu, H.; Wu, S.; Zhao, M.; Li, C.; Liu, X.; Fang, G. Terahertz Spiral Spatial Filtering Imaging. *Appl. Sci.* **2021**, *11*, 2526. [CrossRef]
4. Heimbeck, M.S.; Kim, M.K.; Gregory, D.A.; Everitt, H.O. Terahertz digital off-axis holography for non-destructive testing. In Proceedings of the 36th International Conference IRMMW-THz, Houston, TX, USA, 1–2 October 2011. [CrossRef]
5. Li, R.; Li, C.; Wu, S.; Fang, G. Study of Automatic Detection of Concealed Targets in Passive Terahertz Images for Intelligent Security Screening. *IEEE Trans. Terahertz Sci. Technol.* **2019**, *9*, 165. [CrossRef]
6. Bochkovskiy, A.; Wang, C.; Liao, H. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. Available online: <https://arxiv.org/abs/2004.10934> (accessed on 24 January 2022).
7. Zhang, H.; Cisse, M.; Dauphin, Y.; Lopez-Paz, D. mixup: Beyond Empirical Risk Minimization. *arXiv* **2017**, arXiv:1710.09412. Available online: <https://arxiv.org/abs/1710.09412> (accessed on 24 January 2022).
8. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *15*, 2672–2680. Available online: <https://dl.acm.org/doi/10.5555/2969033.2969125> (accessed on 24 January 2022).
9. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434. Available online: <https://arxiv.org/abs/1511.06434> (accessed on 24 January 2022).
10. Odena, A.; Olah, C.; Shlens, J. Conditional image synthesis with auxiliary classifier GANs. *arXiv* **2017**, arXiv:1610.09585. Available online: <https://arxiv.org/abs/1610.09585v4> (accessed on 24 January 2022).
11. Du, S.; Hong, J.; Wang, Y.; Qi, Y. A High-Quality Multicategory SAR Images Generation Method with Multiconstraint GAN for ATR. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
12. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv* **2017**, arXiv:1701.17875. Available online: <https://arxiv.org/abs/1701.17875> (accessed on 24 January 2022).
13. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of Wasserstein GANs. *Adv. Neural Inf. Process. Syst.* **2017**, *11*, 5767–5777. Available online: <https://dl.acm.org/doi/10.5555/3295222.3295327> (accessed on 24 January 2022).
14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *IEEE Conf. Comput. Vis. Pattern Recognit.* **2016**, *12*, 770–778. [CrossRef]
15. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Int. Conf. Mach. Learn.* **2015**, *12*, 448–456.
16. Andrew, L.M.; Awani, Y.H.; Andrew, Y.N. Rectifier nonlinearities improve neural network acoustic models. *Int. Conf. Mach. Learn.* **2013**, *12*, 1152–1160.
17. Dumoulin, V.; Visin, F. A guide to convolution arithmetic for deep learning. *arXiv* **2016**, arXiv:1603.07285. Available online: <https://arxiv.org/abs/1603.07285> (accessed on 24 January 2022).
18. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2015**, arXiv:1412.6980. Available online: <https://arxiv.org/abs/1412.6980> (accessed on 24 January 2022).
19. Barratt, S.; Sharma, R. A Note on the Inception Score. *arXiv* **2018**, arXiv:1801.01973. Available online: <https://arxiv.org/abs/1801.01973> (accessed on 24 January 2022).

20. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *arXiv* **2017**, arXiv:1706.08500. Available online: <https://arxiv.org/abs/1706.08500> (accessed on 24 January 2022).
21. Szegedy, C.; Liu, W.; Jia, Y.Q.; Sermanet, P.; Rabinovich, A. Going deeper with convolutions. *IEEE Conf. Comput. Vis. Pattern Recognit.* **2015**, *11*, 594. [[CrossRef](#)]
22. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Processing* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
23. Antoniou, A.; Storkey, A. Edwards, H. Data Augmentation Generative Adversarial Networks. *arXiv* **2017**, arXiv:1711.04340. Available online: <https://arxiv.org/abs/1711.04340s> (accessed on 24 January 2022).