

Article

Infrared Dim and Small Target Sequence Dataset Generation Method Based on Generative Adversarial Networks

Leihong Zhang ¹, Weihong Lin ^{2,*} , Zimin Shen ², Dawei Zhang ¹, Banglian Xu ¹, Kaimin Wang ¹ and Jian Chen ³

¹ School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; lhzhang@usst.edu.cn (L.Z.); dwzhang@usst.edu.cn (D.Z.); xblhaq@usst.edu.cn (B.X.); km_wang@usst.edu.cn (K.W.)

² College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai 200093, China; 213332978@st.usst.edu.cn

³ Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China; chenjian@ciomp.ac.cn

* Correspondence: 213332981@st.usst.edu.cn

Abstract: With the development of infrared technology, infrared dim and small target detection plays a vital role in precision guidance applications. To address the problems of insufficient dataset coverage and huge actual shooting costs in infrared dim and small target detection methods, this paper proposes a method for generating infrared dim and small target sequence datasets based on generative adversarial networks (GANs). Specifically, first, the improved deep convolutional generative adversarial network (DCGAN) model is used to generate clear images of the infrared sky background. Then, target-background sequence images are constructed using multi-scale feature extraction and improved conditional generative adversarial networks. This method fully considers the infrared characteristics of the target and the background, which can achieve effective expansion of the image data and provide a test set for the infrared small target detection and recognition algorithm. In addition, the classifier's performance can be improved by expanding the training set, which enhances the accuracy and effect of infrared dim and small target detection based on deep learning. After experimental evaluation, the dataset generated by this method is similar to the real infrared dataset, and the model detection accuracy can be improved after training with the latest deep learning model.

Keywords: generating adversarial networks; infrared dim and small target sequence dataset generation; infrared sky background



Citation: Zhang, L.; Lin, W.; Shen, Z.; Zhang, D.; Xu, B.; Wang, K.; Chen, J. Infrared Dim and Small Target Sequence Dataset Generation Method Based on Generative Adversarial Networks. *Electronics* **2023**, *12*, 3625. <https://doi.org/10.3390/electronics12173625>

Academic Editors: Antonio Fernández-Caballero and Byung-Gyu Kim

Received: 14 June 2023

Revised: 6 August 2023

Accepted: 11 August 2023

Published: 28 August 2023



Copyright: © 2023 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

Infrared images mainly rely on the detector to receive the thermal radiation of the object itself for imaging. Compared with visible light images, infrared imaging conditions are unaffected by light, weather changes and other conditions. The imaging system has a more extended detection range and better penetration ability, so infrared imaging systems are widely used in air defense, the military and other fields. However, collecting infrared data using infrared imaging equipment is costly and time-consuming, and the lack of infrared datasets seriously affects relevant studies based on infrared data.

The main traditional methods for infrared small target detection are filter-based methods [1,2], human eye visual attention-based mechanisms [3–7] and low-rank-based methods [8–14]. With the development of deep learning, infrared dim and small target detection methods based on deep learning have been proposed in recent years. The deep learning-based approach to infrared small target detection uses CNNs to implement feature extraction operations, which allow for deeper semantic information to be obtained from the image. Based on CNNs, Wang et al. [15] proposed a network that uses generative adversarial networks to balance the miss detection rate and false alarm rate in image segmentation,

where the network is trained by adversarial training of three sub-networks to finally achieve the balance of miss detection rate and false alarm rate. Dai et al. [16] proposed the first segmentation-based network, designing an asymmetric background modulation module to aggregate shallow and deep features. Dai et al. [17] then further improved their network by extending local contrast and designing a feature-cyclic transformation scheme to implement trainable local contrast measures. Li et al. [18] proposed a densely nested attention network (DNANet) with a densely nested interaction module and a cascaded channel and spatial attention module designed to implement the interaction between high-level and low-level features as well as the adaptive enhancement of multi-level features, respectively. Zhang et al. [19] proposed the Attention-Guided Pyramid Context Network (AGPCNet) algorithm. Wang et al. [20] proposed a coarse-to-fine internal attention-aware network (IAANet) that uses the semantic contextual information of all pixels within a local region to classify each internal pixel.

The performance of these detection methods depends on their training sets. Due to the insufficient training sets, some methods use cropping, rotation and scaling of the training sets to expand the training sets, which do not fully reflect the real state of the target in the actual scene and have a large amount of data redundancy. Therefore, the study of infrared image generation methods to expand infrared dim and small target image data is of great practical importance for developing infrared dim and small target detection technology.

According to the Society of Photo-Optical Instrumentation Engineers (SPIE) [21], the size of a small target is usually considered to be no more than 9×9 pixels in a 256×256 image. The publicly available datasets are the miss detection vs. false alarm (MDvsFA) [14] and the Single-frame Infrared Small Target (SIRST) [16]. The MDvsFA contains 10,000 images, most of which are close-ups, and the targets are close together and large. The SIRST contains 427 images, with a variety of scenes but too few. Due to the sensitivity of military targets, it is difficult to obtain a sufficient number of publicly available datasets to train the deep learning infrared dim and small target detection algorithm. Therefore, expanding the infrared dataset solves the problem of an insufficient training set for a deep learning-based infrared dim and small target detection algorithm.

The three main models based on image generation are PixelRNN [22], Variational Auto Encoder (VAE) [23], and generative adversarial networks (GANs) [24]. Among them, GANs can extract target features by unsupervised learning, and their strong generalization ability in generative models has developed various improved models [25–28]. From the current research status, the research object of data generation using GANs is still mainly visible images. The research methods in the field of infrared image generation focus on visible light images rather than infrared images. There are fewer studies on generative processing through infrared images, and even fewer studies on the generation of infrared dim and small target data. For example, Uddin M.S. et al. proposed a method for converting optical video to infrared video [29], the basic idea of which is to use attention-generating adversarial networks to focus on target regions by converting a large number of available labeled visible videos to infrared video.

The above methods of generating infrared datasets generate a large area of the target and are single targets. In clustered combat systems, targets usually appear as a group, and thus the need for multi-target tracking is increasingly prominent. Moreover, relying on generative adversarial networks alone suffers from the limitation of not being able to simulate the parameter changes in the target motion process. In order to solve the scarcity of training data, this paper proposes a method to generate an infrared dim and small target dataset, and the generated dataset improves the accuracy and effect of infrared dim and small target detection based on deep learning. In addition, this dataset can provide dataset support for infrared dim and small target detection.

2. Related Principles

2.1. Infrared Background Generation Based on an Improved Deep Convolutional Generative Adversarial Network

2.1.1. Deep Convolutional Generative Adversarial Networks

Deep convolutional generative adversarial networks (DCGANs) [30] were first proposed to combine convolutional neural networks (CNNs) with GANs to improve the unsupervised learning of generative networks by exploiting the powerful feature extraction capability of convolutional networks. A DCGAN consists of two parts, the generator and the discriminator, which continuously learn and improve through zero-sum games and eventually generate data that can be falsified into something that does not exist. The generator is given a noisy input and then generates new sample data by learning the real data's mathematical distribution and feature information. The DCGAN's structure is shown in Figure 1.

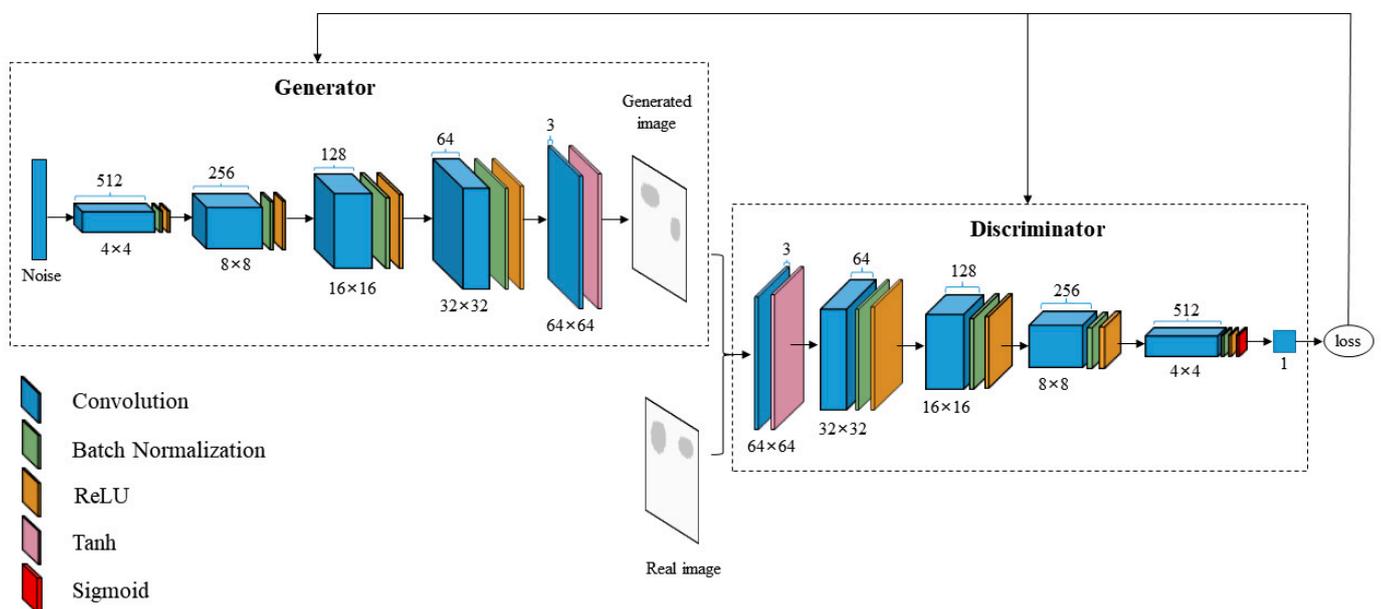


Figure 1. Network structure of DCGAN.

The DCGAN adopts a full convolutional structure based on a GAN to further improve the feature extraction capability of the network. Using the pooling layer for downsampling will lose part of the image information, so the pooling layer in the network is replaced by step convolution. The generator network consists of five deconvolutional layers, using a deconvolution layer with a convolution kernel of 4×4 and a step size of 2, followed by batch normalization (BN). The ReLU activation function is used for all layers except for the last layer, which uses the Tanh activation function. The discriminator network is basically symmetric with the generator network, consisting of five convolutional layers that use a convolutional layer with a convolution kernel of 4×4 and a step size of 2. It uses the LeakyReLU activation function, and the last layer is the Sigmoid function.

2.1.2. ISD-DCGAN Networks

The DCGAN model can perform well in the texture detail of visible light images, and for infrared sky images, its colors do not need to be as rich as those of visible light images. The image size generated by a DCGAN is only up to 64×64 , and larger sizes will have the problem of gradient disappearance. Therefore, this paper proposes the Infrared sky dataset DCGAN (ISD-DCGAN) model by modifying the model based on the original DCGAN, which can significantly improve the stability of model training and obtain high-quality generated images. The ISD-DCGAN differs in the following ways:

(1) The DCGAN generator and discriminator structure is improved by using the ResNet residual module to solve the problem of poor quality of the generated images due to the deepening of the network and the increase in image size. In Figure 1, the DCGAN model has fewer layers, and the generated image size is only 64×64 , which cannot meet the demand, so the image size needs to be further expanded. In this paper, we add two layers of the convolutional network to the original DCGAN structure, and the improved network can generate infrared sky images of size 256×256 , so that it meets the definition of SPIE for infrared dim and small target images. If the number of network layers is directly increased, to a certain extent, more representative image features can be extracted, and the feature expression capability of the network can be improved. However, due to the backpropagation mechanism of the convolutional neural network, the network deepens to increase the number of parameters. If the parameters are extremely large or small, it will lead to the problem of gradient explosion or gradient disappearance during the backpropagation process, and the final result will be the poor quality of the generated images as well as the unstable generation ability of the network. Therefore, the DCGAN is improved by introducing a residual module to deepen the network. The residual module replaces the step-size convolution in the generative and discriminative networks. The residual network can better solve the above problems caused by the deepening of the network layers. Better image generation results are achieved in the case of deeper networks than when directly stacking network layers. It ensures higher-quality images even when the network structure and the number of layers are adjusted. At the same time, introducing the residual network can reduce the number of parameters in the network and further optimize the complexity of the network structure.

(2) The Wasserstein distance is used as a new loss function to enhance the training stability of the network. The loss function of the DCGAN is essentially to make the Jensen Shannon (JS) [31] scatter between P_{data} and P_g as small as possible, but there is a high probability that the two distributions of P_{data} and P_g do not overlap at all. For any two distributions that do not overlap and are sufficiently distant, the JS scatter between them is constant at $\log 2$, causing the gradient to vanish, at which point it is impossible for P_g to move in the direction of P_{data} during the training, and the discriminator cannot be trained. Therefore, the Wasserstein distance [32] is introduced as a loss function in this paper, and the Wasserstein distance achieves a long-range response even when the two distributions do not overlap. The loss function constructed by the Wasserstein distance is introduced to transform the original binary classification task of the discriminant network in the DCGAN into a regression task. Therefore, the last layer of the sigmoid function needs to be removed from the network. The final network structure of the ISD-DCGAN is shown in Figure 2.

In the generator, the DCGAN utilizes multiple deconvolution layers for image generation, while the residual module of the ISD-DCGAN replaces the deconvolution with two convolution operations with a 3×3 convolution kernel and a step size of 1. Each residual unit achieves feature image enlargement in the residual operation by adding up-sampling. The non-residual edges are simply feature maps enlarged using a deconvolution layer with a convolution kernel of 1×1 and a step size of 2 to maintain the same output size as the residual block. After transforming the one-dimensional noise into a 4×4 image, the generating network performs seven consecutive feature map enlargements of the residual module. Finally, a 3×3 convolution kernel is performed once to transform the number of channels, and a 256×256 image is generated using the Tanh activation function. In the discriminator, the modified residual module performs the convolution operation using a convolution layer with a convolution kernel of 3×3 and a step size of 1. Then the feature image reduction is performed by downsampling, and the non-residual edges are reduced using step-size convolution. Finally, the residual edges are stacked with the non-residual edges for output. The discriminative network first performs one step-size convolution, then six residual modules for feature reduction, and finally expands the features into one-dimensional output discriminative results by full connectivity. The improved DCGAN can improve the stability of model training and obtain high-quality generated images.

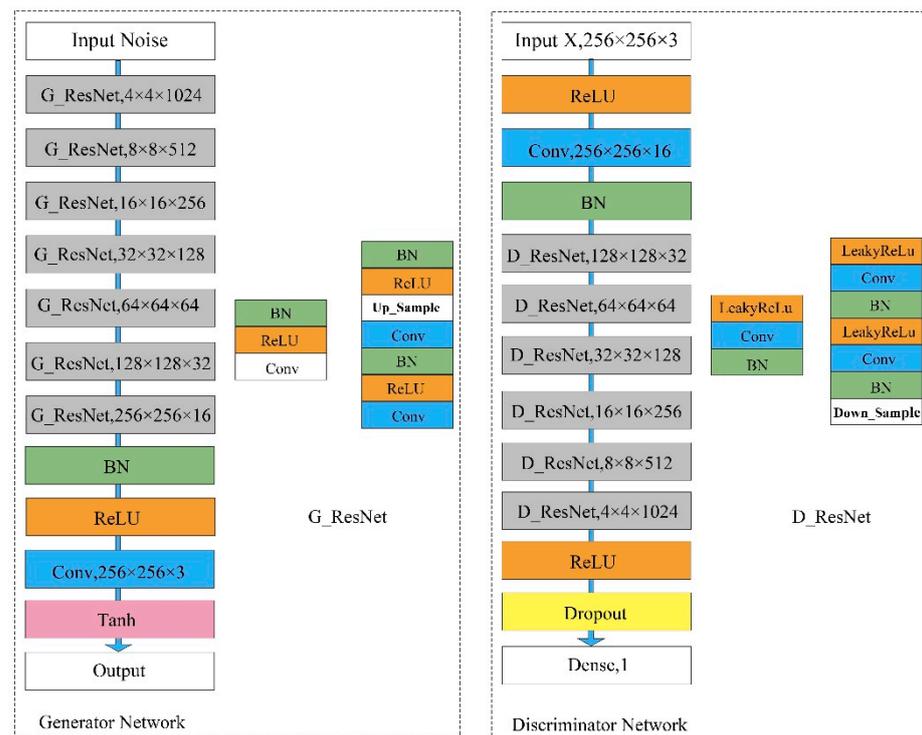


Figure 2. Network structure of ISD-DCGAN.

2.2. Target–Background Image Sequence Construction Based on Improved Conditional Generative Adversarial Networks

Due to the large variability in the main features of the scene and the target, the scene image and the target image need to be generated separately based on two different generation models. After the target and scene images have been generated separately, the target and scene images need to be combined to obtain a reasonable target–background image. The target images generated by the target generation model are of a single scale, whereas in a practical scene, as the motion parameters of the target change, the spatial position and dimensions of the target in the viewing scene will change accordingly. Therefore, the target–background image cannot be directly synthesized in a simple and straightforward manner.

To address the above challenges, this paper proposes a target–background image synthesis model based on an improved conditional generative adversarial network, which combines constraint parameters such as the spatial location and size of the target to achieve a reasonable synthesis of target–background images. To improve the quality of the target–background image generation, a multi-scale feature fusion mechanism and an attention mechanism are incorporated, resulting in a higher fidelity of the generated image.

As shown in Figure 3, target–background image synthesis is achieved by an improved conditional generative adversarial network. As the objects in the target and background images vary in size and shape and their positions are mostly non-fixed, using only a single scale is likely to lose some feature information and affect the detection effect. To address this problem, a multi-scale feature module is designed in this paper, using convolutional kernels of different scales to obtain different ranges of perceptual fields so as to obtain more comprehensive target and scene feature information and strengthen the adaptability of the network to multiple scales. At the same time, an attention mechanism is added to the feature extraction to enable the model to extract more meaningful image features. A multi-scale bidirectional fusion target–background image generator and discriminator is implemented.

Conditional generative adversarial networks are expanded into conditional models by adding constraints to the original generative adversarial network. This is achieved by conditioning the model on additional information, which in turn constrains and guides

the image generation process. The improved generator uses a U-net structure in which the convolutional layer acts as the encoder and the deconvolutional layer as the decoder. In the encoder section, each node to the next undergoes a sequence of a convolutional layer, a normalization layer and an LReLU activation layer. For the decoder part, the input and corresponding decoder mirror layers are stitched before each convolutional layer, with each node to the next undergoing a sequence of a deconvolutional layer, a batch normalization layer and a ReLU activation layer. A jump-join technique is introduced in the encoder–decoder section, whereby the input of each deconvolution layer is the output of the previous layer plus the output of a layer symmetrically convolved with that layer, thus ensuring that the encoder information is constantly remembered at decoder time, allowing the generated image to retain as much of the original image information as possible.

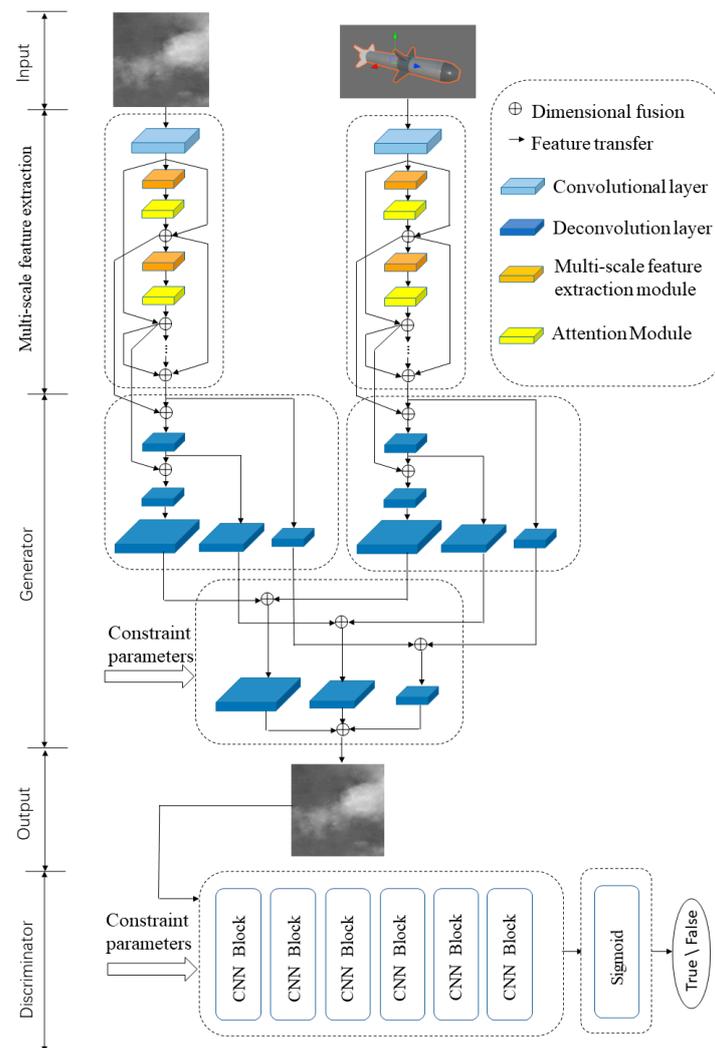


Figure 3. Infrared image sequence generation model.

3. Method

In order to expand the infrared small target dataset, this paper proposes a method for generating infrared dim and small target sequence datasets, including the following steps: (1) Generating an infrared sky background. (2) Creating an infrared small target model. (3) Constructing a target–background image sequence. (4) Generating dataset labels. The flow chart of the infrared dim and small target sequence dataset generation method is shown in Figure 4.

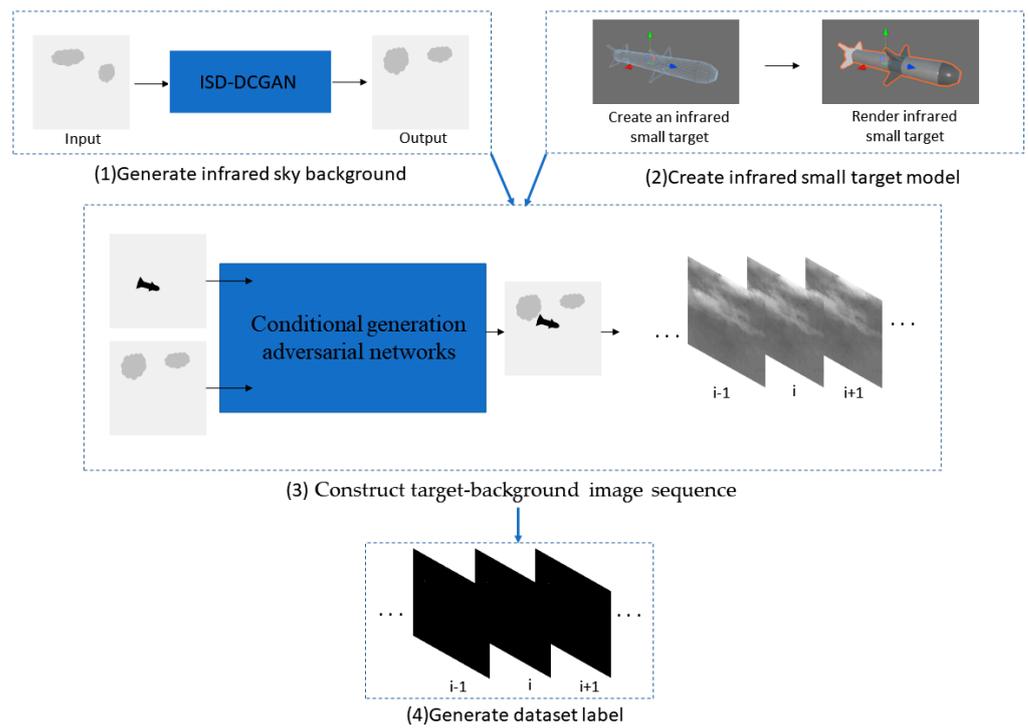


Figure 4. Flowchart of infrared dim and small target sequence dataset generation method.

3.1. Generating an Infrared Sky Background

The ISD-DCGAN network trains the real infrared sky background images to generate 256×256 infrared sky background images. The training set of the network consists of 512 real infrared sky backgrounds with a learning rate of 0.0001 and a batch size of 64. The number of iterations of the generator and discriminator training is 1000 epochs. The loss function converges after 1000 rounds of training, but the effect does not improve. Figure 5 shows the process of generating images for the ISD-DCGAN.

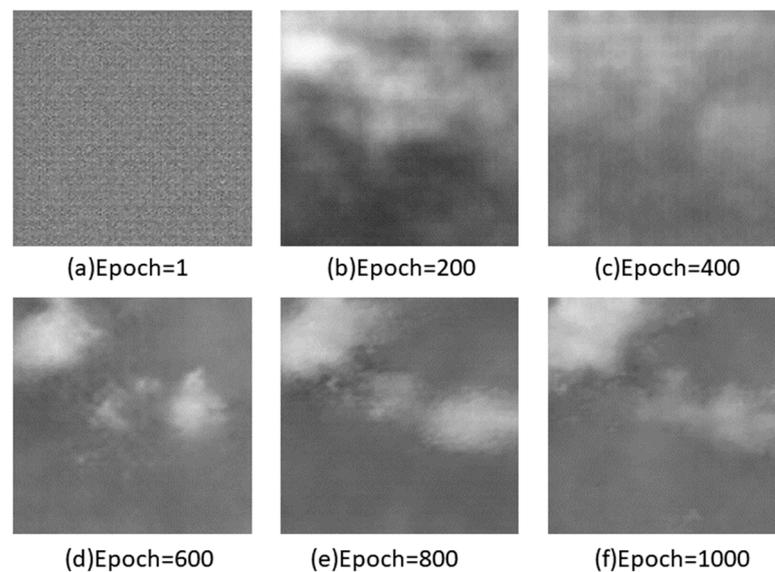


Figure 5. The process of generating images for ISD-DCGAN.

3.2. Creating an Infrared Small Target Model

In this paper, 3ds Max software is used for model building, and the model types are aircraft and missiles. The modeling process uses the missile as an example. Firstly, the

shape proportion, structure and appearance material of the required construction model are analyzed, and the model is modeled separately using 3ds Max tools according to the 1:1 ratio. The model is shown in Figure 6a. Texture drawing technology is then used to better show the details of the model with minimal resource consumption. Finally, the model is rendered to ensure that the target and flight effects are realistic.

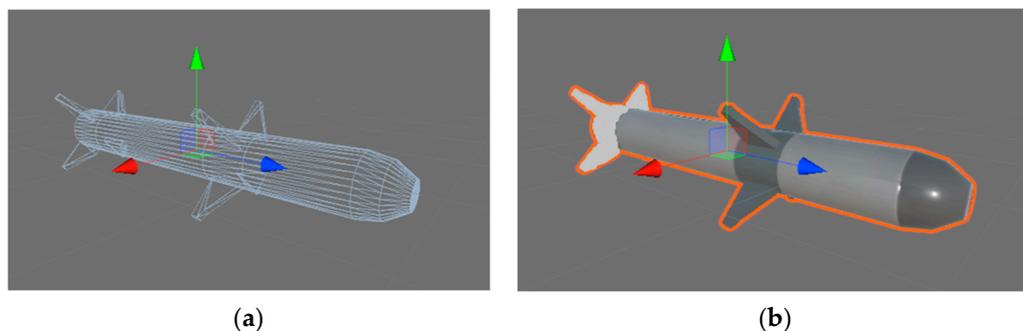


Figure 6. Creating an infrared small target model. (a) Three-dimensional model; (b) heat source effect.

To achieve a more realistic effect, this paper adds the effect of temperature on the model's infrared radiation intensity. Based on the information about the surface material in the field of view, we calculate the radiation of the heat source on the inner surface at a unit distance and then calculate the radiation of each pixel using the depth information in the camera. Finally, the grayscale increment of the material under the influence of the heat source is calculated and saved to the texture. The ComputerShader performs the second calculation of the texture in Unity3D, and the grayscale of the heat transfer from the area radiated by the heat source to the surrounding textures is calculated. The heat source effect obtained is shown in Figure 6b.

3.3. Constructing a Target–Background Image Sequence and Generating Datasets

By improving the conditional generative adversarial network model, multi-scale feature extraction and fusion are performed on the input target and scene images, thus obtaining more comprehensive target and scene feature information and enhancing the network's adaptability to multiple scales and its ability to extract image features. Combining auxiliary constraint parameters such as the target spatial location and the size in the generative adversarial network enables the synthesis of target–background images. The introduction of a jump-join technique in the encoder–decoder part allows the generated image to retain as much information as possible about the original image.

3.4. Generating Dataset Labels

Dataset labels are used to manually label the data that need to be identified and discriminated. Deep neural networks learn the features of these labels and eventually achieve the function of autonomous recognition. The current method of labeling infrared dim and small target datasets is to find the dim and small target, then manually label the target area using the Labellmg dataset labeling tool, and finally set the rest of the area to a black background.

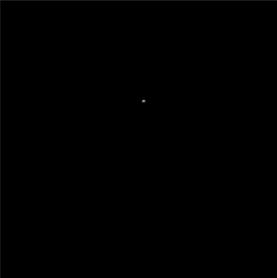
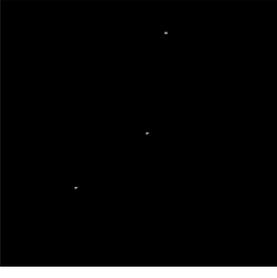
4. Experiments

In this section, we present the experimental results and then introduce the evaluation metrics of the dataset. The experimental hardware included an Intel Core i9-10920X CPU@3.50GHz and an NVIDIA GeForce RTX 3090, and the experimental software included PyCharm 2.3, 3dMax 2020, Unity3d 2019.1.9 and Labellmg.

4.1. Experimental Results

We generated a dataset of infrared dim and small target sequences based on generative adversarial networks. Six synthetic infrared images were generated by varying the parameters of target, noise and wavelength, resulting in 20,000 datasets and 20,000 labels. In Table 1, column (a) indicates single-target and multi-target images in near-infrared light (NIR); column (b) indicates single-target and multi-target images in NIR with added noise; column (c) indicates single-target and multi-target images in far-infrared light (FIR); and column (d) indicates dataset labels.

Table 1. Generated dataset.

	(a) NIR	(b) NIR with Added Noise	(c) FIR	(d) Dataset Labels
Single target				
Multiple targets				

Noise and wavelength can affect the accuracy of target detection methods, so it is essential to simulate the inclusion of different parameters to train target detection models. We can generate the desired infrared background image by generating an adversarial network to make it more consistent with the actual scenario of the target detection process.

4.2. Experimental Evaluation

4.2.1. Comparison between DCGAN and ISD-DCGAN

In order to compare the effect of the model before and after improvement, we plot the discriminator loss of the network in each epoch as a loss function to visualize the change in the loss function during the training of the network. Figure 7 shows the change curves of the discriminant network loss function during the training process before and after the improvement in the network. The loss function of the network generator before improvement decreases in the first 500 epochs and oscillates steadily between 0.5 and 1.5, but the loss starts to oscillate significantly after 500 epochs. The discriminator loss function of the improved ISD-DCGAN network structure is significantly weakened, decreases steadily after 200 epochs and gradually tends to oscillate slightly around 0. From the loss function curves of the training process, the training process of the improved network structure can converge to stability. This indicates that both the generator and the discriminator have finally reached a mutually constrained and balanced state, and the effect is better and more stable than before the improvement.

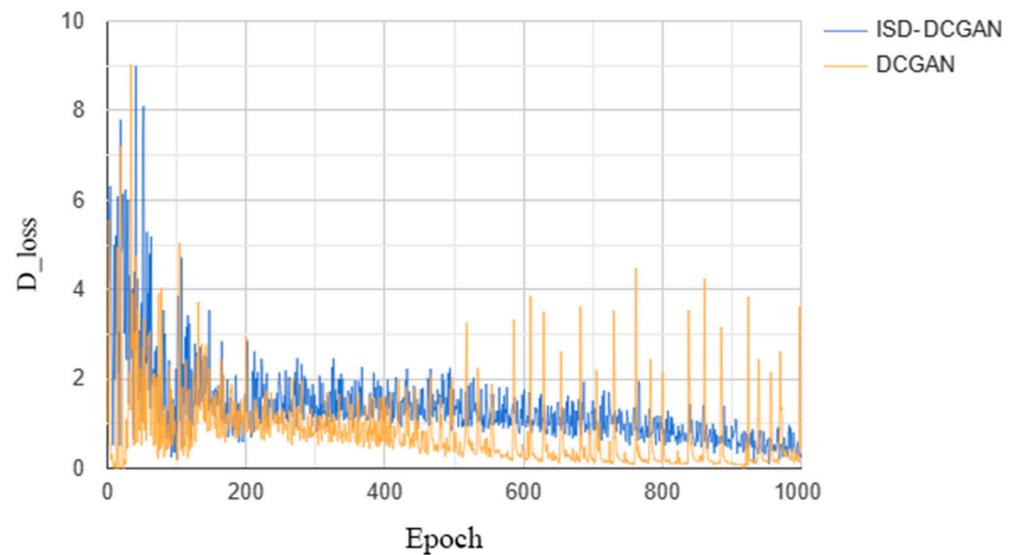


Figure 7. Change in loss value.

4.2.2. Structure Similarity Index Measure

In order to further verify the validity of the experimentally generated images, the generated images are quantitatively analyzed using objective performance metrics. In this paper, the objective evaluation index of the generated image is used as the Structure Similarity Index Measure (SSIM) [33], which formula is:

$$SSIM(x, y) = [l(x, y)^\alpha \cdot c(x, y)^\beta \cdot s(x, y)^\gamma] \quad (1)$$

In Equation (1), $\alpha > 0, \beta > 0, \gamma > 0$ denotes the brightness characteristics of the original image and the simulated image; $c(x, y) = (2\sigma_x\sigma_y + C_2) / (\sigma_x^2 + \sigma_y^2 + C_2)$ denotes the contrast characteristics of the original image and the simulated image; and $s(x, y) = (\sigma_{xy} + C_3) / (\sigma_x\sigma_y + C_3)$ denotes the similarity characteristics of the original image and the simulated image. Among them, μ_x and μ_y represent the average gray values of the original image and the simulated image, respectively, reflecting the luminance information. σ_x and σ_y denote the variance of the gray value of the original and simulated images, respectively, reflecting the contrast information. σ_{xy} denotes the correlation coefficient between the original image and the simulated image, reflecting the similarity of the structural information. C_1, C_2 and C_3 are small quantities greater than zero to prevent overflow of the calculation result when the divisor is zero. During the training of the ISD-DCGAN, the value of SSIM is calculated once every 10 epochs, and the final SSIM curve in the training process is shown in Figure 8.

From the SSIM curve, we can know that the similarity structure with the original image is about 0.34 because the training input is random noise at the beginning. After the continuous game between the generator and the discriminator, the similarity gradually increases and stabilizes at about 0.85. This indicates that the dataset generated in this paper has a high structural similarity with the original image, which can ensure that the generated infrared sky image meets the requirements and also increase the infrared sky background image style. The higher the similarity, the closer it is to the real image. In deep learning, a high-quality training set can improve the performance of the model classifier and help it better detect the targets of real images.

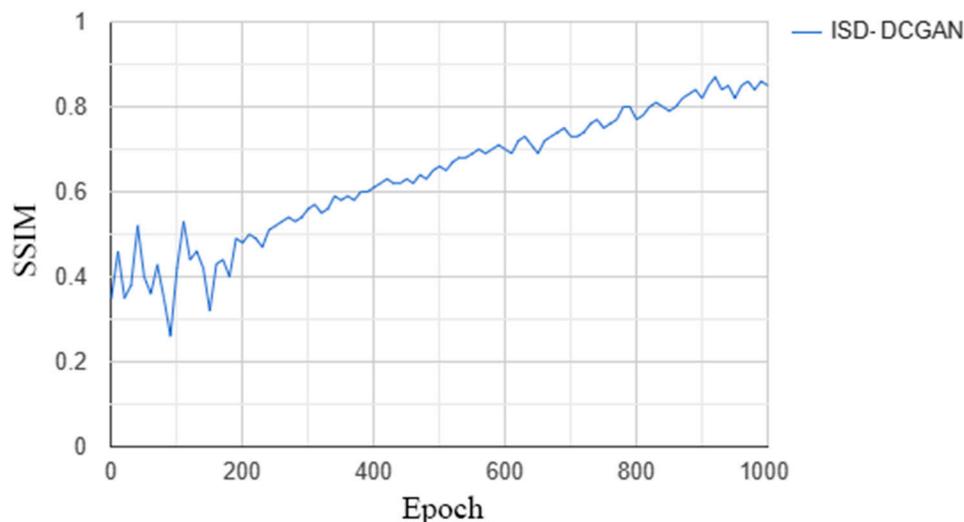


Figure 8. SSIM curve in the training process.

4.2.3. Comparative Analysis with Other Datasets

There are very few existing open datasets in infrared dim and small target detection, and most of the traditional detection methods are evaluated on their internal datasets. Only a few infrared small target datasets are published by CNN-based methods. The first open one is the MDvsFA dataset. This dataset consists of 10,000 training images, a significant portion of which are synthesized. Another dataset developed is the SIRST, which has 427 images and is suitable for testing. Although these open datasets have greatly contributed to the development of infrared dim and small target detection, they suffer from limited data capacity and poor labeling. Figure 9 shows the MDvsFA, SIRST and ISD-DCGAN datasets and their 3D plots. Row (a) represents the MDvsFA dataset, and row (b) its 3D plot. Row (c) represents the ISD-DCGAN dataset, and row (d) its 3D plot. Row (e) represents the SIRST dataset, and row (f) its 3D plot.

In this paper, the generated infrared dim and small target sequence dataset is applied to the infrared dim and small target detection method to verify the effectiveness of the generated dataset. Firstly, the MDvsFA dataset and the 10,000 images in this paper were used to train in the Dense Nested Attention Network (DNANET), the Attention-Guided Pyramid Context Network (AGPCNet) and the Interior Attention-Aware Network (IAANET). The object detection accuracy was then tested on all datasets of the SIRST. Figure 10 shows a plot of the detection results trained using the MDvsFA, and Figure 11 shows a plot of the detection results trained using the ISD-DCGAN. Table 2 shows the detection accuracies of the three detection methods after training on different datasets. The target detection rate P_d and the false detection rate F_d for target detection are calculated as follows:

$$P_d = \frac{\text{Number of real targets detected}}{\text{Actual target number}} \tag{2}$$

$$F_d = \frac{\text{Number of false targets detected}}{\text{Actual target number}} \tag{3}$$

From the table and figures above, it can be seen that, firstly, different datasets train the model to obtain different levels of precision, indicating that different quality datasets have an influence on the detection model. Secondly, the dataset in this paper is more in line with the real image. By comparing the precision of our dataset with the MDvsFA dataset after different model training, the test results in the SIRST dataset show that our dataset has a certain enhancement over the MDvsFA after training, which illustrates its effectiveness. The enhancement brought by the dataset to the accuracy of the model can be visualized in Figures 10 and 11.

Table 2. Precision of the model after training on different datasets.

Model	Data Set	P_d	F_d
IAANET	MDvsFA	0.642	0.811
IAANET	Ours	0.705	0.753
AGPCNET	MDvsFA	0.593	0.282
AGPCNET	Ours	0.634	0.244
DNANET	MDvsFA	0.883	0.883
DNANET	Ours	0.904	0.794

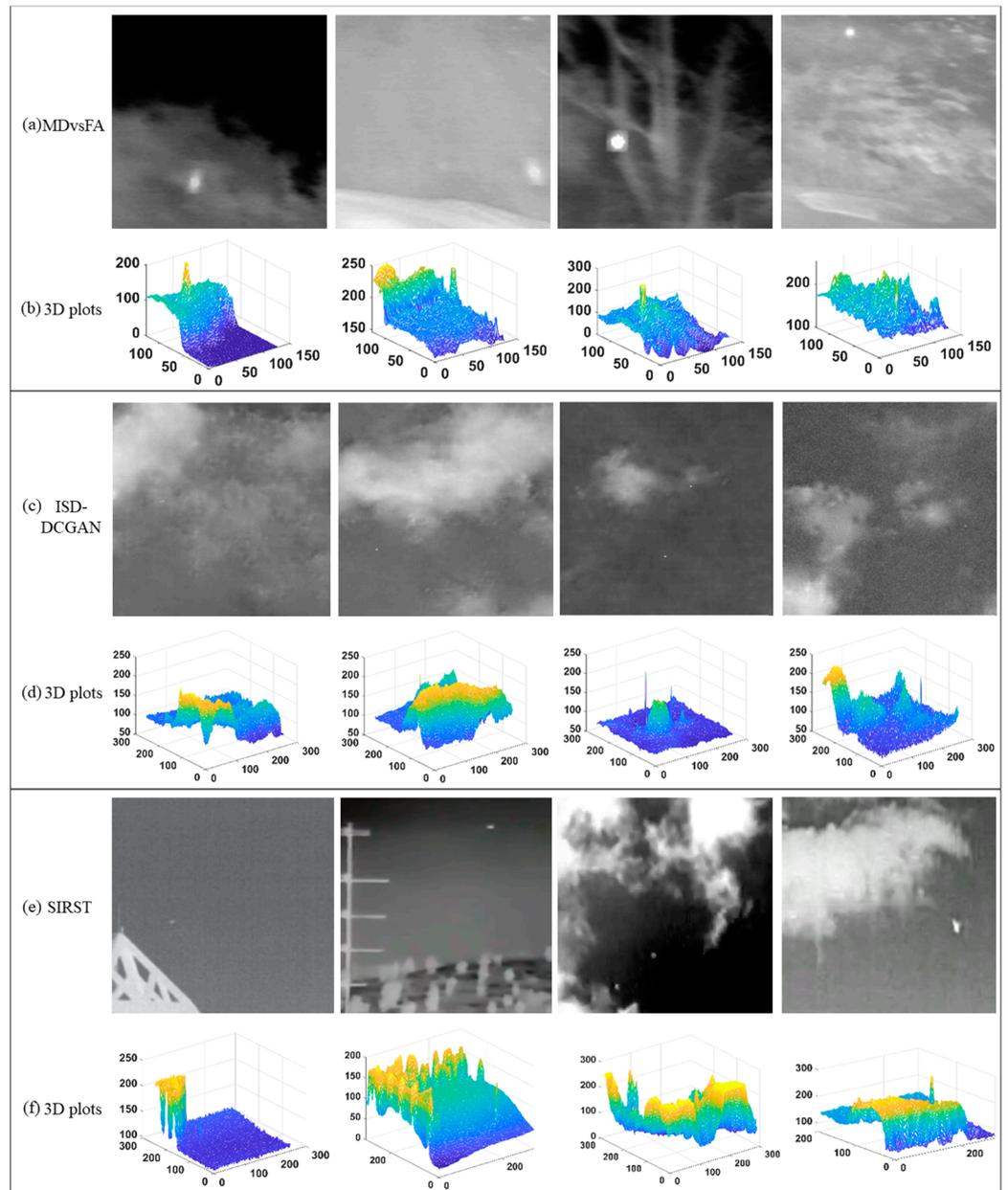


Figure 9. MDvsFA, ISD-DCGAN and SIRST datasets and their 3D plots.

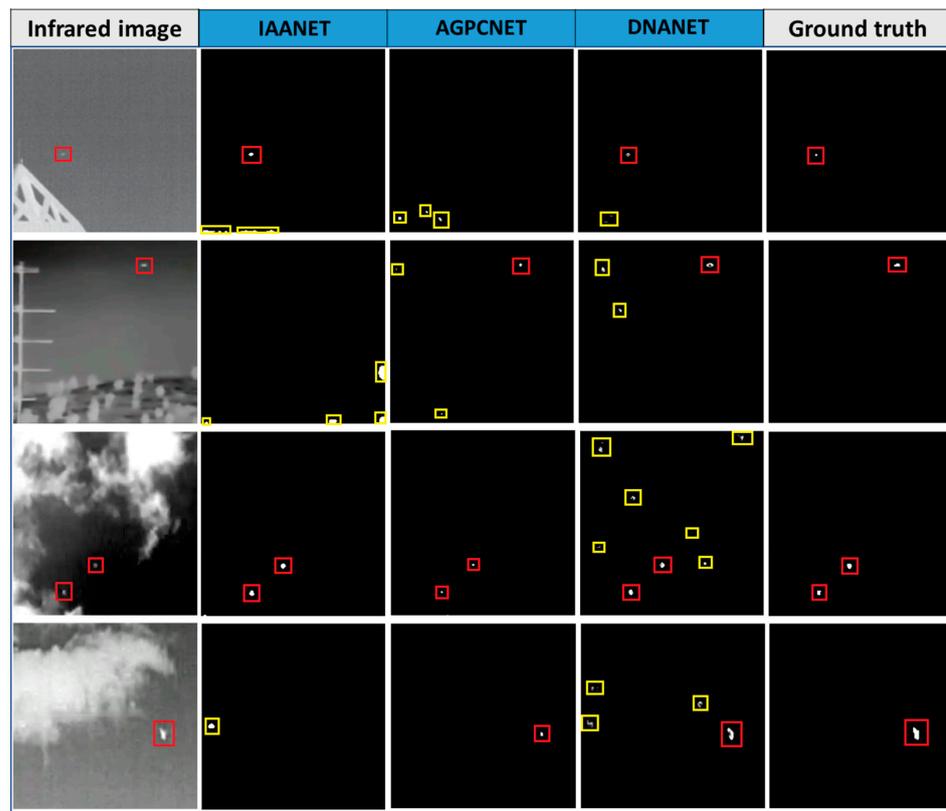


Figure 10. Detection results of different algorithms after training on MDvsFA dataset. Red boxes indicate true targets, yellow releases indicate misdetections.

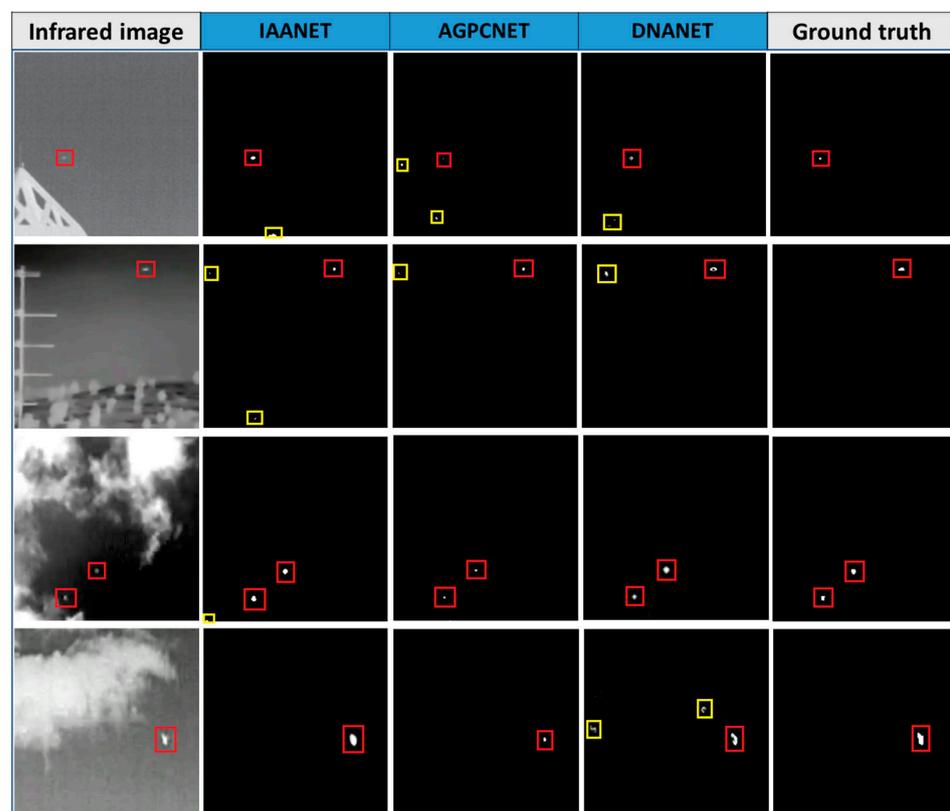


Figure 11. Detection results of different algorithms after training on ISD-DCGAN dataset. Red boxes indicate true targets, yellow releases indicate misdetections.

5. Discussion

In recent years, deep learning-based infrared dim and small target detection algorithms have been proposed by an increasing number of researchers. Due to the sensitivity of military targets, it is difficult to obtain a sufficient number of publicly available datasets for training deep learning-based infrared dim and small target detection algorithms. Currently, the only publicly available datasets are the MDvsFA and the SIRST. Although these open-source datasets have greatly contributed to the development of infrared dim and small target detection, they suffer from limited data capacity, non-compliant targets and manual annotation, and better methods of dataset expansion are needed. Datasets are generally expanded by rotating, cropping and mirroring, which does not result in completely new datasets, and manual annotation is problematic. To solve the problem of insufficient infrared dim and small target datasets and better improve the accuracy and effectiveness of infrared dim and small target detection based on deep learning, this paper proposes a method for generating infrared dim and small target sequence datasets based on deep convolutional generative adversarial networks to generate new data on the basis of the original datasets.

In this paper, we have fully validated the effectiveness of this dataset through experiments. Firstly, the impact of the improved network on the generated images is analyzed. Secondly, the similarity metric of the generated images is analyzed. Finally, the impact of training is compared between our dataset and other datasets through different model training.

In summary, the dataset in this paper enriches the infrared dim and small target datasets and is useful for deep learning focused on small target models. We will expand the dataset to include different scenarios in the future.

6. Conclusions

In this paper, a method for generating infrared dim and small target sequence datasets based on deep convolutional adversarial networks is proposed. First, we improve the deep convolutional generation adversarial network model to generate compliant infrared sky background images. Then, the target with the generated infrared sky background image is added to an improved conditional generation adversarial network to generate a different dataset of infrared dim and small target sequences. After experimental analysis, we conclude that: (1) The improved deep convolutional generation adversarial network solves the problem of gradient disappearance due to increasing image size and improves the quality of the generated images. (2) The datasets generated are valid and can be applied to training infrared dim and small target detection models. (3) Compared with the MDvsFA dataset, the precision of the dataset generated in this paper has improved after training infrared dim and small target detection models in recent years. In summary, this paper mainly investigates the method of generating infrared dim and small target sequence datasets based on generative adversarial networks and provides a new method for expanding infrared dim and small target datasets.

Author Contributions: Conceptualization, L.Z. and W.L.; methodology, Z.S.; software, K.W.; validation, B.X., D.Z. and L.Z.; formal analysis, J.C.; investigation, Z.S.; resources, L.Z.; data curation, D.Z.; writing—original draft preparation, W.L.; writing—review and editing, W.L.; visualization, J.C.; supervision, L.Z.; project administration, L.Z.; funding acquisition, L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number No. 62275153) and the Development Fund for Shanghai Talents (grant number No. 2021005).

Data Availability Statement: The 20,000 datasets for this paper will be available at <https://github.com/LWH1115> (accessed on 1 August 2023).

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

References

1. Deng, L.; Zhang, J.; Xu, G.; Zhu, H. Infrared small target detection via adaptive M-estimator ring top-hat transformation. *Pattern Recognit.* **2021**, *112*, 107729. [[CrossRef](#)]
2. Bae, T.W. Small target detection using bilateral filter and temporal cross product in infrared images. *Infrared Phys. Technol.* **2011**, *54*, 403–411. [[CrossRef](#)]
3. Chen, C.P.; Li, H.; Wei, Y.; Xia, T.; Tang, Y.Y. A Local Contrast Method for Small Infrared Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 574–581. [[CrossRef](#)]
4. Han, J.; Ma, Y.; Zhou, B.; Fan, F.; Liang, K.; Fang, Y. A Robust Infrared Small Target Detection Algorithm Based on Human Visual System. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 2168–2172.
5. Han, J.H.; Liang, K.; Zhou, B.; Zhu, X.Y.; Zhao, J.; Zhao, L.L. Infrared Small Target Detection Utilizing the Multiscale Relative Local Contrast Measure. *IEEE Trans. Geosci. Remote Sens.* **2018**, *15*, 612–616. [[CrossRef](#)]
6. Wei, Y.; You, X.; Li, H. Multiscale patch-based contrast measure for small infrared target detection. *Pattern Recognit.* **2016**, *58*, 216–226. [[CrossRef](#)]
7. Deng, H.; Sun, X.; Liu, M.; Ye, C.; Zhou, X. Small Infrared Target Detection Based on Weighted Local Difference Measure. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4204–4214. [[CrossRef](#)]
8. Zhang, L.; Peng, L.; Zhang, T.; Cao, S.; Peng, Z. Infrared small target detection via non-convex rank approximation minimization joint $l_{2,1}$ norm. *Remote Sens.* **2018**, *10*, 1821. [[CrossRef](#)]
9. Zhang, T.; Wu, H.; Liu, Y.; Peng, L.; Yang, C.; Peng, Z.; Zhang, T. Infrared small target detection based on non-convex optimization with L_p -norm constraint. *Remote Sens.* **2019**, *11*, 559. [[CrossRef](#)]
10. Tz, A.; Zp, A.; Hao, W.A.; Yh, A.; Cl, B.; Cy, A. Infrared small target detection via self-regularized weighted sparse model. *Neurocomputing* **2021**, *420*, 124–148.
11. Zhang, C.; He, Y.; Tang, Q.; Chen, Z.; Mu, T. Infrared Small target detection via interpatch correlation enhancement and joint local visual saliency prior. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5001314. [[CrossRef](#)]
12. Cao, Z.; Kong, X.; Zhu, Q.; Cao, S.; Peng, Z. Infrared dim target detection via mode-k1k2 extension tensor tubal rank under complex ocean environment. *ISPRS J. Photogramm. Remote Sens.* **2021**, *181*, 167–190. [[CrossRef](#)]
13. Zhang, P.; Zhang, L.; Wang, X.; Shen, F.; Pu, T.; Fei, C. Edge and Corner Awareness-Based Spatial-Temporal Tensor Model for Infrared Small-Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10708–10724. [[CrossRef](#)]
14. Kong, X.; Yang, C.; Cao, S.; Li, C.; Peng, Z. Infrared small target detection via nonconvex tensor fibered rank approximation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 3068465. [[CrossRef](#)]
15. Wang, H.; Zhou, L.; Wang, L. Miss Detection vs. False Alarm: Adversarial Learning for Small Object Segmentation in Infrared Images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8509–8518.
16. Dai, Y.; Wu, Y.; Zhou, F.; Barnard, K. Asymmetric contextual modulation for infrared small target detection. In Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2021; pp. 950–959.
17. Dai, Y.; Wu, Y.; Zhou, F.; Barnard, K. Attentional local contrast networks for infrared small target detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 9813–9824. [[CrossRef](#)]
18. Li, B.; Xiao, C.; Wang, L.; Wang, Y.; Lin, Z.; Li, M.; An, W.; Guo, Y. Dense nested attention network for infrared small target detection. *IEEE Trans. Image Process.* **2023**, *32*, 1745–1758.
19. Zhang, T.; Cao, S.; Pu, T.; Peng, Z. Agpnet: Attention-guided pyramid context networks for infrared small target detection. *arXiv* **2021**, arXiv:2111.03580.
20. Wang, K.W.; Du, S.Y.; Liu, C.X.; Cao, Z.G. Interior Attention-Aware Network for Infrared Small Target Detection. *IEEE Geosci. Remote Sens.* **2022**, *60*, 5002013. [[CrossRef](#)]
21. Tartakovsky, A.; Kligys, S.; Petrov, A. Adaptive sequential algorithms for detecting targets in a heavy IR clutter. In Proceedings of the SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, Denver, CO, USA, 18–23 July 1999.
22. van den Oord, A.; Kalchbrenner, N.; Kavukcuoglu, K. Pixel Recurrent Neural Networks. In Proceedings of the 33rd International Conference on International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 1747–1756.
23. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv* **2013**, arXiv:1312.6114.
24. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Commun. ACM* **2020**, *63*, 139–144.
25. Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Wang, X.; Huang, X.; Metaxas, D.N. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5907–5915.
26. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv* **2014**, arXiv:1411.1784.
27. Karras, T.; Laine, S.; Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
28. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
29. Uddin, M.S.; Hoque, R.; Islam, K.A.; Kwan, C.; Gribben, D.; Li, J. Converting Optical Videos to Infrared Videos Using Attention GAN and Its Impact on Target Detection and Classification Performance. *Remote Sens.* **2021**, *13*, 3257. [[CrossRef](#)]

30. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2015**, arXiv:1511.06434.
31. Dong, T.; Shang, W.; Zhu, H. Naive Bayesian Classifier Based on the Improved Feature Weighting Algorithm. In *Advanced Research on Computer Science and Information Engineering*; Shen, G., Huang, X., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 142–147.
32. Zhang, Y.; Fang, Q.; Qian, S.; Xu, C. Knowledge-aware Attentive Wasserstein Adversarial Dialogue Response Generation. *ACM Trans. Intell. Syst. Technol.* **2020**, *11*, 37. [[CrossRef](#)]
33. Zhou, W.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.