applied sciences

MDPI

Article

Cycle Generative Adversarial Network Based on Gradient Normalization for Infrared Image Generation

Xing Yi [1,2,3,4], Hao Pan [1,*], Huaici Zhao [2,3,4,*], Pengfei Liu [2,3,4], Canyu Zhang [1,2,3,4], Junpeng Wang [1,2,3,4] and Hao Wang [1]

[1] School of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China
[2] Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110169, China
[3] Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang 110169, China
[4] Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China
* Correspondence: panhao@syuct.edu.cn (H.P.); hczhao@sia.cn (H.Z.)

Abstract: Image generation technology is currently one of the popular directions in computer vision research, especially regarding infrared imaging, bearing critical applications in the military field. Existing algorithms for generating infrared images from visible images are usually weak in perceiving the salient regions of images and cannot effectively highlight the ability to generate texture details in infrared images, resulting in less texture details and poorer generated image quality. In this study, a cycle generative adversarial network method based on gradient normalization was proposed to address the current problems of poor infrared image generation, lack of texture detail and unstable models. First, to address the problem of limited feature extraction capability of the UNet generator network that makes the generated IR images blurred and of low quality, the use of the residual network with better feature extraction capability in the generator was employed to make the generated infrared images highly defined. Secondly, in order to solve issues concerning severe lack of detailed information in the generated infrared images, channel attention and spatial attention mechanisms were introduced into the ResNet with the attention mechanism used to weight the generated infrared image features in order to enhance feature perception of the prominent regions of the image, helping to generate image details. Finally, to tackle the problem where the current training models of adversarial generator networks are insufficiently stable, which leads to easy collapse of the model, a gradient normalization module was introduced in the discriminator network to stabilize the model and render it less prone to collapse during the training process. The experimental results on several datasets showed that the proposed method obtained satisfactory data in terms of objective evaluation metrics. Compared with the cycle generative adversarial network method, the proposed method in this work exhibited significant improvement in data validity on multiple datasets.

Keywords: cycle generative adversarial networks; spatial attention; channel attention; gradient normalization; residual networks

Citation: Yi, X.; Pan, H.; Zhao, H.;Liu, P.; Zhang, C.; Wang, J.; Wang, H. Cycle Generative Adversarial Network Based on Gradient Normalization for Infrared Image Generation. *Appl. Sci.* **2023**, *13*, 635. https://doi.org/10.3390/app13010635

Academic Editor: Atsushi Mase

Received: 26 November 2022
Revised: 24 December 2022
Accepted: 26 December 2022
Published: 3 January 2023

1. Introduction

The traditional source of infrared image data is through an infrared camera device that uses the temperature difference between the target and the background to capture infrared light and generate infrared images. The advantage of infrared camera image acquisition is that it can be adopted to locate and track targets at night or in harsh weather. However, infrared images acquired in such harsh environments tend to be of poor quality, with severe loss of textured detail, and expensive infrared equipment, and are thus less used in civilian applications. With the rapid development of deep learning, infrared image generation using neural networks has received increasing attention from researchers. Infrared images obtained through using deep learning methods cannot only compensate for defects in infrared cameras employed in harsh environments but also save costs. Therefore, infrared

images generated by neural networks are extensively employed in military research and civilian applications.

In recent years, deep neural networks have been widely used in various fields, such as target detection, computer vision, and image migration [1,2]. Based on the idea of gaming, in 2014, Goodfellow et al. proposed generative adversarial networks [3,4], representing one of the widely applied methods in deep learning at present. Generative adversarial networks include generator networks and discriminator networks. The basic idea is that the task generator generates fake data and the discriminator's task is to discriminate fake data generated by the generator from the real data; in essence, to distinguish real data from fake data. The generative adversarial network achieves the conversion of source-domain images to target-domain images based on this game idea. Nevertheless, researchers have found that the network suffers from easy collapse and limited image generation capability in model training.

In the field of image generation research, in order to achieve better image generation techniques and to obtain more realistic images in the target domain, researchers have proposed many new ideas and methods for generative adversarial networks to overcome problems and difficulties in image generation techniques. In the area of supervised image generation, Alec Dadford and Luke Metz et al. proposed the deep convolutional generative adversarial network [5–7] (DCGAN) in 2015. Compared to traditional generative adversarial networks, deep convolutional generative adversarial networks use leaky ReLU activation function and provide better enhancement of the generated image features. However, the problem with DCGAN networks in image generation is that training requires paired datasets that are small and difficult to collect under normal circumstances. In 2017, P Isola et al. proposed a supervised-learning-based pix2pix network based on pix2pix image generation algorithm [8]. The pix2pix algorithm was based on CGAN, using UNet as the generator of the network and PatchGAN as the discriminator to reconstruct accuracy through reconstruction loss and adversarial joint optimization of low and high frequencies. The method achieves generation from semantically segmented images to visible images and can achieve more desirable results in image generation tasks, yet the method is similar to the DCGAN method in that it requires the use of paired datasets to generate images from the source domain to the target domain, making the preliminary data collection workload particularly high. Meanwhile, the pix2pix algorithm produces unsatisfactory images when the input dataset deviates significantly from the training dataset. To resolve the problems of the pix2pix algorithm, P Isola et al. proposed the pix2pixHD algorithm [9–11] in 2018 to generate high-definition images. The pix2pixHD algorithm was optimized based on the pix2pix algorithm. The authors adopted a global generative network and a local boosting network in the generator and by optimizing the generator network, 2048*1024 HD images could be generated. In the discriminator, the authors utilized three discriminators of the same structure responsible for the discriminative task at different scales and increased the image resolution at a deeper level to obtain high-definition images. However, the training process of the pix2pixHD algorithm was not stable enough. Arjovsky et al. presented the Wasserstein generative adversarial network (WGAN) in 2018 [12,13]. The WGAN method restricts the parameters in the discriminator by cutting the weights of the discriminator to render the discriminator less discriminative, and thus completely solved the instability problem of the generative adversarial network model. In 2018, Zhang H et al. put forward the self-attentive generative adversarial network (SAGAN) [14,15]. The SAGAN method addressed the limitation of convolutional local perceptual field. A self-attentive mechanism was introduced in the discriminator and generator, and global feature information was obtained in this way. Although this method achieved better results with fewer iterations, it was highly unstable and prone to crashing during model training.

In the field of unsupervised learning for image generation, Liu M Y et al. 2016 proposed coupled generative adversarial networks (COGANs) for unpaired datasets [16]. This approach set out to learn the joint distribution of multi-domain images by partial weight sharing. The generators share the weights of the first half, and the discriminators

extract the high-level features of the second half. Taigman Y et al. 2016 presented the domain transfer network (DTN) [17–19]. The DTN network employs a compound loss function to solve the general analogy problem by separating unlabeled samples in the given domain via a multivariate function to learn new mapping relationships for image conversion, while the domain asymmetry and the small amount of information contained in the new source domain leads to unsatisfactory results. In 2017, Zhu J Y et al. put forward a cycle generative adversarial network [20], which used a dual generator with a dual discriminator to implement bidirectional conversion of source-domain images to target-domain images. Zhu J Y et al. concluded that if it was possible to generate target-domain images from source-domain images, then generating source-domain images from target-domain images is also feasible; however, this model does not easily converge during training and the quality of the generated images is usually not high. In 2018, Choi Y et al. proposed StarGAN [21,22]. StarGAN implements transformation between multiple domain images and has the advantage that for x-field transformation, CycleGAN needs to learn x•(x − 1) models, while StarGAN only needs to learn one model. The disadvantage of StarGAN is a large sample size requirement. Huang X et al. proposed the Munit algorithm in 2018 [23]. The Munit algorithm uses both shared content space and differentiated style space for image generation. When generating an image, the same content and different styles were combined and encoded for output, thus allowing multimodal image generation and multimodal image conversion. In 2019, Liu M Y et al. presented the FUNIT algorithm [24,25], which mainly solved the image conversion problem for small sample images and untouched regions. Its network framework is composed of a conditional image generation network and a multitask adversarial discriminative network, which achieves image generation by computing a small number of samples. The method bears strong generalization capability.

Despite many researchers having continuously improve and optimize the algorithms to make the generated images consistent with the target images, the problems of model instability, easy collapse and low quality of image generation have not been sufficiently recognized. The gradient-based normalization method proposed in the current study focuses on solving the aforementioned problems of easy model collapse and insufficient prominent texture detail information in the generated infrared images to improve the quality of image generation and obtain realistic infrared images with realistic effects.

2. Related Theoretical Work

2.1. Cycle Generation of Adversarial Networks

Cycle generative adversarial networks were proposed by Jun-Yan Zhu et al. to implement image generation tasks between unpaired datasets. In contrast to generative adversarial networks, the cycle generative adversarial networks use dual generators and discriminators. The main idea of cycle generative adversarial networks is that after generating a target-domain image from a source-domain image, the source-domain image can be generated again based on the style of target-domain image. In the traditional approach of cycle generative adversarial network, the generators and discriminators use the UNet [26,27] and PatchGAN network structures [28], respectively. General framework diagram of cycle generative adversarial network structure is shown in Figure 1 below. The general flow of network implementation is as follows: the real visible image is input into Generator_A to generate a fake infrared image, and then the fake infrared image is input into Discriminator_B together with the real infrared image for discrimination. The result obtained from the discrimination is fed back to Generator_A. Similarly, the real infrared image is input into Generator_B to generate a fake visible image, the fake visible image is then input into Discriminator_B together with the real visible image to discriminate between the real and fake images, and the results are fed back to Generator_B.

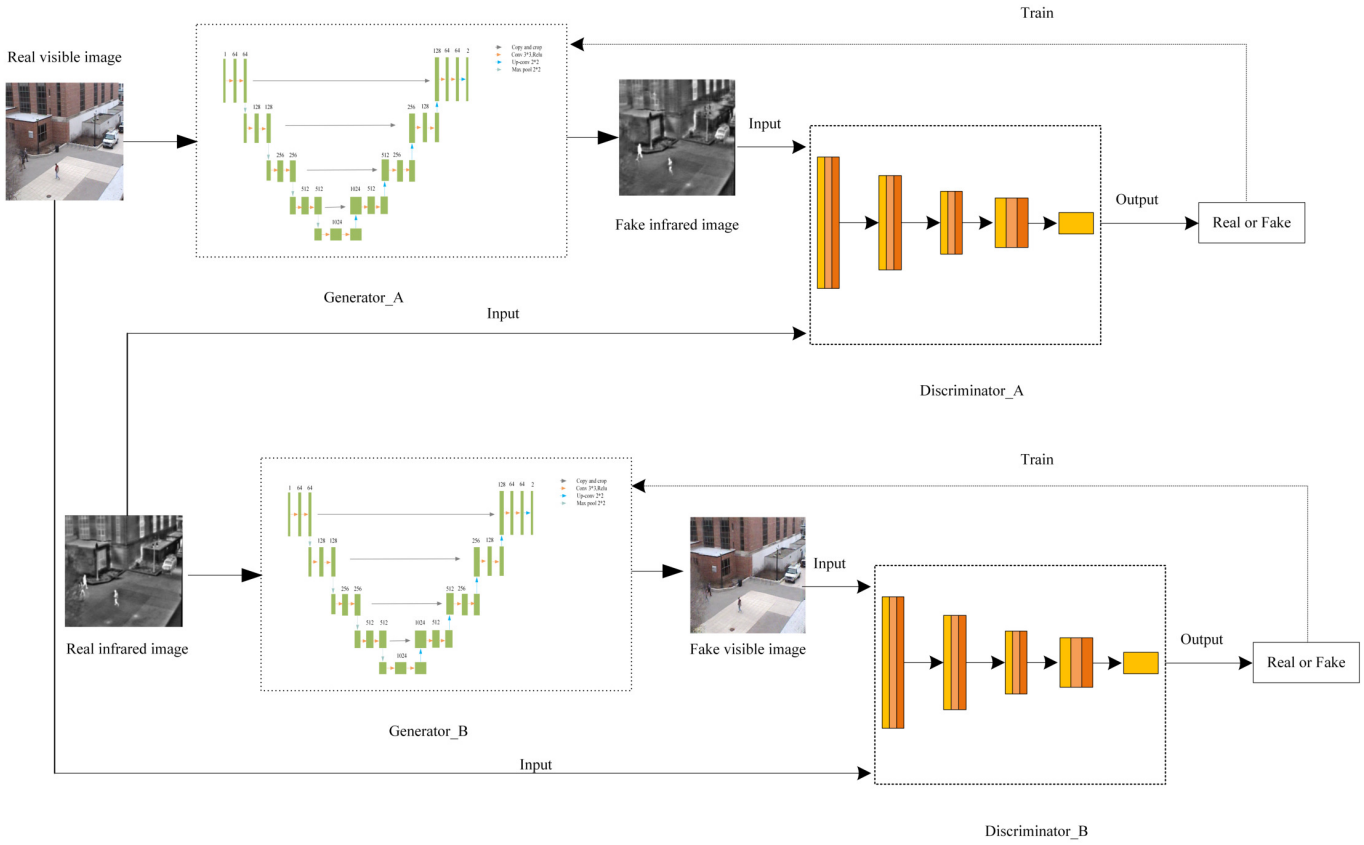

Figure 1. General framework diagram of cycle generative adversarial network structure.

2.2. Channel Attention Mechanism and Spatial Attention Mechanism

In 2018, Woo S et al. proposed a method combining channel attention and spatial attention mechanisms to solve the problem of neural networks failing to focus on the important region features during training [29–31]. The authors used the given feature map to weight attention to the image features in both spatial and channel dimensions. Then they featured matching with the original image to achieve the adaptive adjustment for solving the focusing problem in key feature regions. The structure of attention mechanism is shown in Figure 2 below.

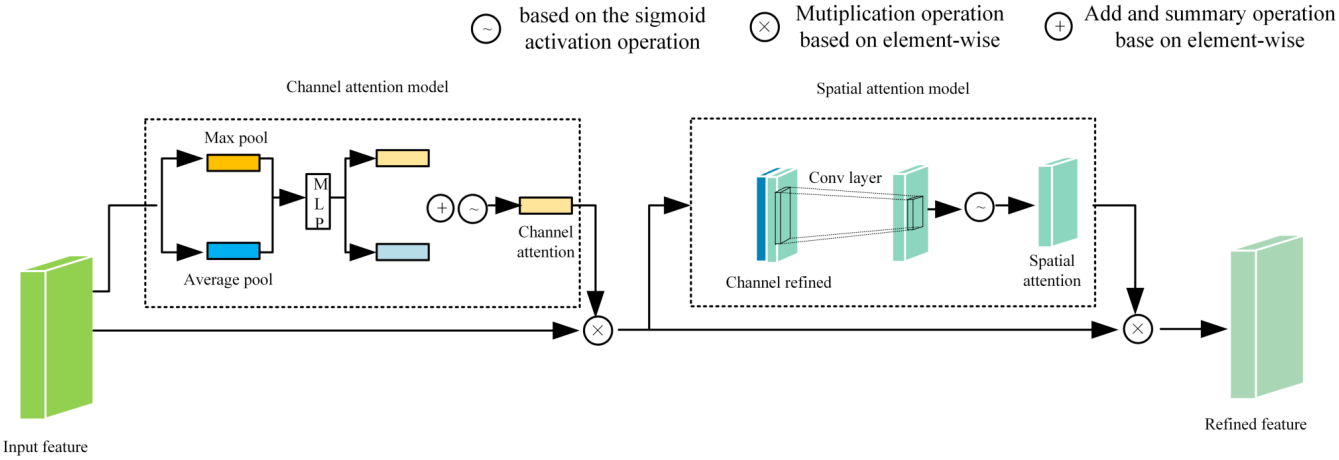

Figure 2. Structural diagrams of the channel and spatial attention mechanisms, which are used to weight the infrared image features to highlight the textural details of the generated image.

The specific process is to feed the feature maps of input $H \times W \times C$ into the average pooling layer and maximum pooling layer to obtain the feature maps of two $1 \times 1 \times C$. Then, the feature maps are subjected to the feature extraction by MLP neural network, and the extracted features are weighted and sigmoid-activated to obtain the final feature maps of channel attention. Spatial attention is based on the premise of channel attention, and the input of spatial attention module is the feature map obtained by multiplying the channel attention feature map with the original feature map. First, the input feature maps are subjected to the maximum and average pooling to obtain the two $H \times W \times 1$ feature maps. Then, the two $H \times W \times 1$ feature maps are obtained by 7×7 convolution and channel dimensionality reduction, and the spatial attention feature map is obtained by sigmoid-activation. This feature map is multiplied by the input feature map to form a new feature map with the incorporation of spatial attention and channel attention.

2.3. Gradient Normalization

Cycle generative adversarial networks are volatile, prone to breakdowns during training, and slow to converge, mainly due to the wide gradient space of discriminator. Some researchers have proposed methods such as L2 normalization [32], gradient penalty and weight clipping to address this problem, which can indeed make the network model stable. Still, these methods limit the model capacity of discriminator to a certain extent. Parameter clipping and spectral normalization are similar in the sense, ensuring that the L-constant at each layer of the model is bounded by constraining the parameters so that the total L-constant is also bounded. Whereas the gradient penalization notices that a sufficient condition for $\| f\|_L \leq 1$ is $\nabla_x f(x) \leq 1$ and therefore imposes a soft constraint on the model by using a penalty term $(\| \nabla_x f(x) \| -1)^2$.

The gradient normalization introduced in this paper [33–35] also uses the gradient to transform $f(x)$ into such that it can automatically satisfy $\| \nabla_x \hat{f}(x) \| \leq 1$. Specifically, by regarding leakyReLU as the activation function in which $f(x)$ is actually viewed as a segmented linear function, it shows that, except for the boundary, $f(x)$ is a linear function in a locally continuous region, and the corresponding $\nabla_x f(x)$ then becomes a constant. Therefore, gradient normalization contemplates $\hat{f}(x) = f(x) / \| \nabla_x f(x) \|$, which gives the formula as shown in (1).

$$\| \nabla_x \hat{f}(x) \| = \| \frac{\nabla_x f(x)}{\| \nabla_x f(x) \|} \| = 1 \tag{1}$$

To avoid errors caused by dividing the model by zero, function norm $|f(x)|$ should be introduced into the denominator, thus ensuring that the function is bounded, which gives the formula as shown in (2).

$$\hat{f}(x) = \frac{f(x)}{\| \nabla_x f(x) \| + |f(x)|} \in [-1, 1] \tag{2}$$

In this paper, a gradient-normalized discriminator is introduced to solve the problem of difficult model convergence without limiting the model capacity of discriminator. By weighting the normalized gradient loss with the loss of original discriminator, the model convergence is accelerated and the training process is stabilized. The overall gradient normalization function is shown in Equation (3) below.

$$\hat{f}(x) = |\frac{f(x)}{\| \nabla_x f(x) \| + \zeta(x)}|^2 \tag{3}$$

$\zeta(x)$ is a universal term that can be associated with $f(x)$ or a constant. As $f(x)$ tends to infinity, $\nabla_x f(x)$ tends to 0, and $\zeta(x)$ approximates $|f(x)|$. When the discriminator is saturated by overfitting, the normalized gradient norm approaches 0. This self-control mechanism prevents the generator from acquiring exploding gradients, thus stabilising the model training process of cycle generative adversarial networks.

2.4. Loss Functions

The loss function of the network consists of generator loss and discriminator loss. The former consists of generative adversarial loss, cycle consistency loss and the loss from converting the target domain to the source domain after generating the image. The generative adversarial loss is the loss obtained by the generator from converting the source-domain image to target-domain image. The constructed generative adversarial loss function of generating the target-domain image (infrared image) from the source domain (visible image) is shown in Equation (4).

$$Loss_{GAN}(GA\_B, D_Y, X, Y) = E_{y-P_{data}(y)}[\log_{10} D_Y(y)] + E_{x-P_{data}(x)}[\log_{10}(1 - D_Y(GA\_B(x)))] \quad (4)$$

of which, $E_{y-P_{data}(y)}$ in the generator GA_B is the expected value of the sample image taken in the target domain; $E_{x-P_{data}(x)}$ is the expected value of the sample image taken in the source domain; y in the function is a sample in the Y sample space (infrared image), and x is a sample in the sample space X (visible image). $GA\_B(x)$ is the image generated by the generator GA_B; $D_Y(y)$ is the probability that discriminator D discriminates whether y is a sample from the Y sample space, and $1 - D_Y(G(x))$ is the probability of discriminator D discriminating the image generated by generator GA_B and judging whether the image is a sample taken from the Y sample space.

The generative adversarial loss function with the target domain (infrared image) generating the source domain (visible image) is shown in Equation (5).

$$Loss_{GAN}(GB\_A, D_X, X, Y) = E_{x-P_{data}(x)}[\log_{10} D_X(x)] + E_{y-P_{data}(y)}[\log_{10}(1 - D_X(GB\_A(y)))] \quad (5)$$

This loss function is similar to that is used to generate the target-domain image from the source domain.

The cycle consistency loss function learns both GA_B and GB_A mappings simultaneously. It expects $GA\_B(GB\_A(y))$ to generate an image as close to y as possible and $GB\_A(GA\_B(x))$ to generate an image as close to x as possible. The purpose of cycle consistency loss function is to prevent generator G from overlearning the samples in the Y sample space and thus over-altering the samples in the X sample space. The cycle consistency loss function is shown in Equation (6).

$$Loss_{cycle}(GA\_B, GB\_A) = E_{x-P_{\text{data}}(x)}[||GB\_A(GA\_B(x)) - x||] + E_{y-P_{\text{data}}(y)}[||GA\_B(GB\_A(y)) - y||] \quad (6)$$

Thus, the final loss function is shown in Equation (7).

$$Loss(GA\_B, GB\_A, D_X, D_Y) = Loss_{GAN}(GA\_B, D_Y, X, Y) + Loss_{GAN}(GB\_A, D_X, X, Y) + \lambda Loss_{cycle}(GA\_B, GB\_A) \quad (7)$$

where λ is the weighting factor of the cycle consistency loss function.

3. Methodologies

In order to generate better IR images, we must not only consider the feature extraction capability of the generator to prevent too much detail information from being lost. In addition, we also need to consider the model collapse problem of cycle generative adversarial networks during the model training process, which directly affects the effect of infrared image generation. Therefore, this paper will focus on solving the quality problem of infrared image generation and preventing model collapse. the specific contributions of this paper are as follows:

1. A new generative network with ResNet is used instead of the traditional U-Net network structure, because the feature extraction capability of ResNet is much higher than that of U-Net. The new generative network makes the generated image features richer as well.

2. In this paper, we introduce the spatial attention and channel attention mechanisms to the ResNet network structure. The spatial attention and channel attention mechanism enhances the textured detail information of image generation and reduces the loss of

details. In doing so, this solves the problems of severe textural information loss and low generated-image quality as observed in the traditional method.

3. To address the problem of the model being unstable and prone to collapse during the training process of cycle generative adversarial networks, we introduce a gradient normalization module to the discriminator to stabilize the training process of the model and increase its convergence speed in the training process. In addition, the gradient normalization module ensures that the model does not easily collapse.

Network Framework Structure

This paper adopts a cycle generative adversarial network-based framework structure. The generator in this paper adopts the ResNet network with a stronger feature extraction capability and introduces the channel and spatial attention mechanisms to the ResNet network, improving the quality of generated images and solving the severe problem of lacking texture information in the generated images.In the discriminator, we introduce a gradient normalization module to stabilize the training process of the model and prevent the model from collapsing due to the unstable state during the training process. The overall network framework structure is shown in Figure 3 below. The general flow of network implementation is as follows: the real visible image is input into Generator_A to generate a fake infrared image, and then the fake infrared image is input into Discriminator_B together with the real infrared image for discrimination. The result obtained from the discrimination is fed back to Generator_A. Similarly, the real infrared image is input into Generator_B to generate a fake visible image, the fake visible image is then input into Discriminator_B together with the real visible image to discriminate between the real and fake images, and the results are fed back to Generator_B.

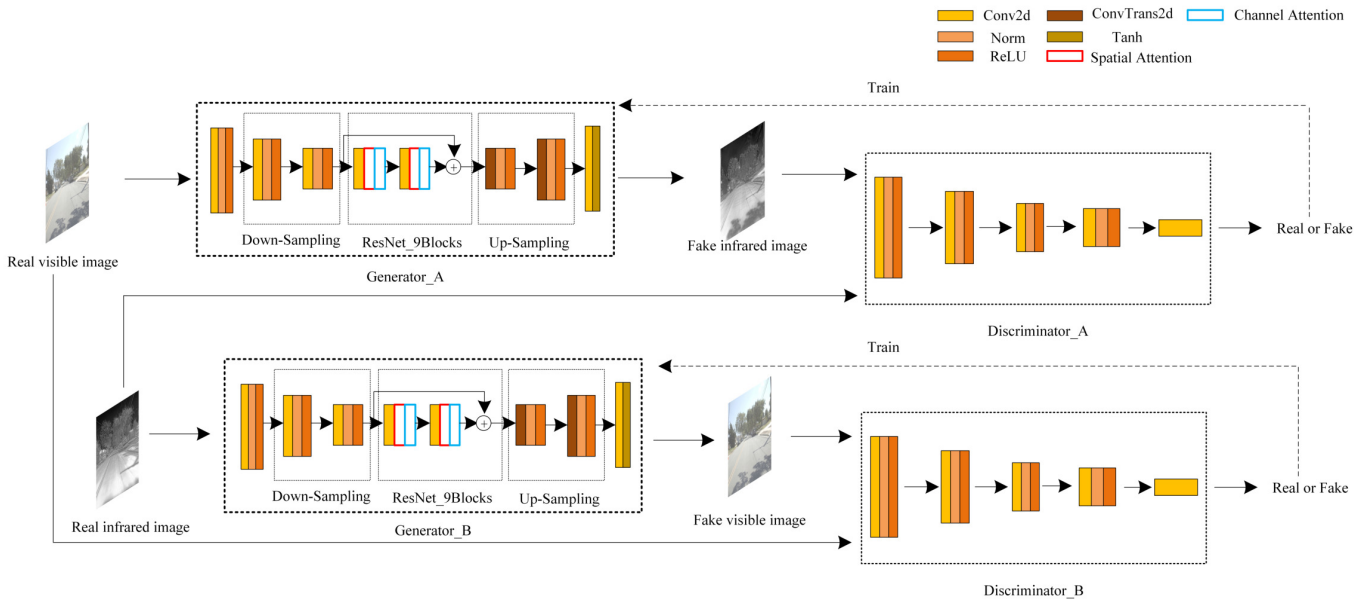

Figure 3. Flowchart of the overall framework of the cycle generative adversarial network.

In this paper, the specific training process of the network model is represented in a pseudocode form, as shown in Table 1. The prerequisite to ensure the output of infrared images is whether the discriminator can accurately distinguish between true and false infrared images. For this, a batched cycle training form of training the dual discriminator K times in line and then training the generator once, and continuing to iterate to the maximum training number M, is adopted.

Table 1. Training process of the network model.

	Input: Infrared image $/I_i$,visible image $/I_v$ Output: False infrared image $/I_v$ generated from visible image $/I_j$
step 1	For M epochs do
step 2	For K steps do
step 3	n samples taken from the IR image distribution $/\{I_i^1, I_i^2, I_i^3, ..., I_i^n\}$
step 4	n samples taken from the visible image distribution $/\{I_j^1, I_j^2, I_j^3, ..., I_j^n\}$
step 5	Training the discriminator Discriminator_A and updating the parametric model.
step 6	Training the discriminator Discriminator_B and updating the parametric model.
step 7	End for
step 8	Training the generator Generator_A and updating the model parameters.
step 9	Training the generator Generator_B and updating the model parameters.
step 10	End for

4. Analysis of Experimental Results

4.1. Dataset and Experimental Procedure

The platform configuration used to train the deep learning model in this paper is as follows: the graphics card is a GeForce RTX 2080Ti, the memory is 32 GB, and the framework is PyTorch. This model was built on a Linux system using Python 3.7. The experimental results after model testing and the experimental ablation data were obtained using MATLAB R2019a on a Windows 10 operating system.

The two datasets used in the training process were the visible and infrared images from the OSU colour thermal dataset [36] and visible-infrared images from the Flir dataset [37]. Five hundred pairs of infrared-visible image sequence pairs from each dataset were selected as the training dataset, and another 20 pairs were selected as the test set used in this experiment. The training epoch was set to 200, and the learning rate was set to 0.0002 from 1 to 100 epochs and to a linear gradient descent at 100 to 200 epochs, with a total of 1450 learned steps. The batch_size was set to 1, and the Adam optimization strategy was adopted.

4.2. Experimental Results

This paper uses objective evaluation metrics to evaluate the quality of generated IR images. The objective evaluation uses the peak signal-to-noise ratio [38,39] and structural similarity [40,41] to assess the sharpness and richness of image texture detail. In the field of image generation, structural similarity is a more authoritative image evaluation metric widely used within the current image processing field. It is used to obtain information about the image structure in visible region by the strong correlation between image pixels and to obtain approximate information about the image using whether the perceived structural information has changed to express the similarity difference between images.

The peak signal-to-noise ratio is commonly used in image compression, image fusion, and image generation evaluation methods, mainly to compare the difference between two images. A higher signal-to-noise ratio value indicates better quality of the generated image and less difference between images, in which PSNR is defined by the mean square error (MSE). Given a noise-free mn monochrome image I and its noise approximation K, the MSE is defined as shown in Equation (8).

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2 \quad (8)$$

The PSNR is therefore defined as shown in Equation (9).

$$PSNR = 10 \cdot \log_{10}(\frac{MAX_i^2}{MSE}) \tag{9}$$

where MAX indicates the possible maximum pixel in the image.

The control experiments in this paper were conducted using the same CycleGAN framework structure by implementing control experiments in different generators. The control experiments mainly used UNet_256, ResNet_6blocks and ResNet_9blocks, CUT in the two datasets. CUT [42] is proposed in 2020. CUT is a state-of-the-art image translation implemented using contrast learning. CUT does not require paired datasets as training data, and generates clear quality images. CUT is now widely used in the field of image generation. This paper's method comprises ResNet_9blocks, channel attention, spatial attention mechanism, and gradient normalization. In order to make it easier to read, the method is named GN_CycleGAN, and the method composed of ResNet_9blocks, channel attention, and spatial attention mechanism is called CBAM_CycleGAN. It will be applied in the later experimental part, and the experimental results are shown in Table 2. The visualized images are shown in Figures 3 and 4. Compared with the original CycleGAN method in the Flir dataset, the PSNR and SSIM metrics of the proposed method in this paper improved by 2.3% and 9.7%, respectively. The PSNR metric improved by 32.2% on the OSU colour thermal dataset, and the SSIM metric improved from 0.2479 to 0.7491. The objective metrics of the two datasets in Data Table 2 show that the quality pf infrared images generated by the method are significantly improved, and based on this proposed method, the desired results in terms of objective evaluation metrics have also been obtained, which objectively validates the effectiveness of algorithm in this paper.

Table 2. Objective evaluation results of the comparative experiment.

Mathod	**OSU**		**Flir**	
Evaluation indicators	PSNR/dB	SSIM	PSNR/dB	SSIM
CycleGAN+unet_128	13.6867	0.2479	13.2131	0.4167
CycleGAN+unet_256	13.7892	0.2686	12.7318	0.4444
CycleGAN+resnet_6blocks	17.3956	0.7071	13.3745	0.4700
CycleGAN+resnet_9blocks	17.0300	0.6900	13.3275	0.4368
CUT [42]	13.3502	0.3168	13.1813	0.4214
Ours(GN_CycleGAN)	18.0699	0.7491	13.5195	0.4572

As seen from the visualization results in the first data set (first row) of Figure 4, the image generated by original method suffers from blurred white shadows and a severe lack of detail in the region marked in red, with a poor degree of infrared image effect. In the ResNet experiment, there is a small amount of white shading in the edge information at the image details, and the ability to generate more texture detail information is significantly improved compared to the original method, with a small amount of detail information loss. From our proposed method, the generated infrared image is rich in texture detail, with more apparent contour information, and is close to the real picture. In the second set of experiments, the original method has more white shadows and severe feature loss in the region marked in red. In the ResNet experiments, there are only a few white shadows, which are more blurred in the image, and there is a large loss of detailed information. In the cut method, there is no good generation for all three sample datasets; the generated images' quality is more blurred, and the infrared effect is not apparent. In comparison with the two previous methods, the method in this paper produces images with no white shadows, significant feature information, and no loss of detail information. In the third data group, the red-marked area is the road line. Compared with the real infrared image, the original method and the method based on ResNet network cannot generate the features in the red area well. While the method in this paper can show better area feature information, correctly-represented texture details, and a better image-generation effect.

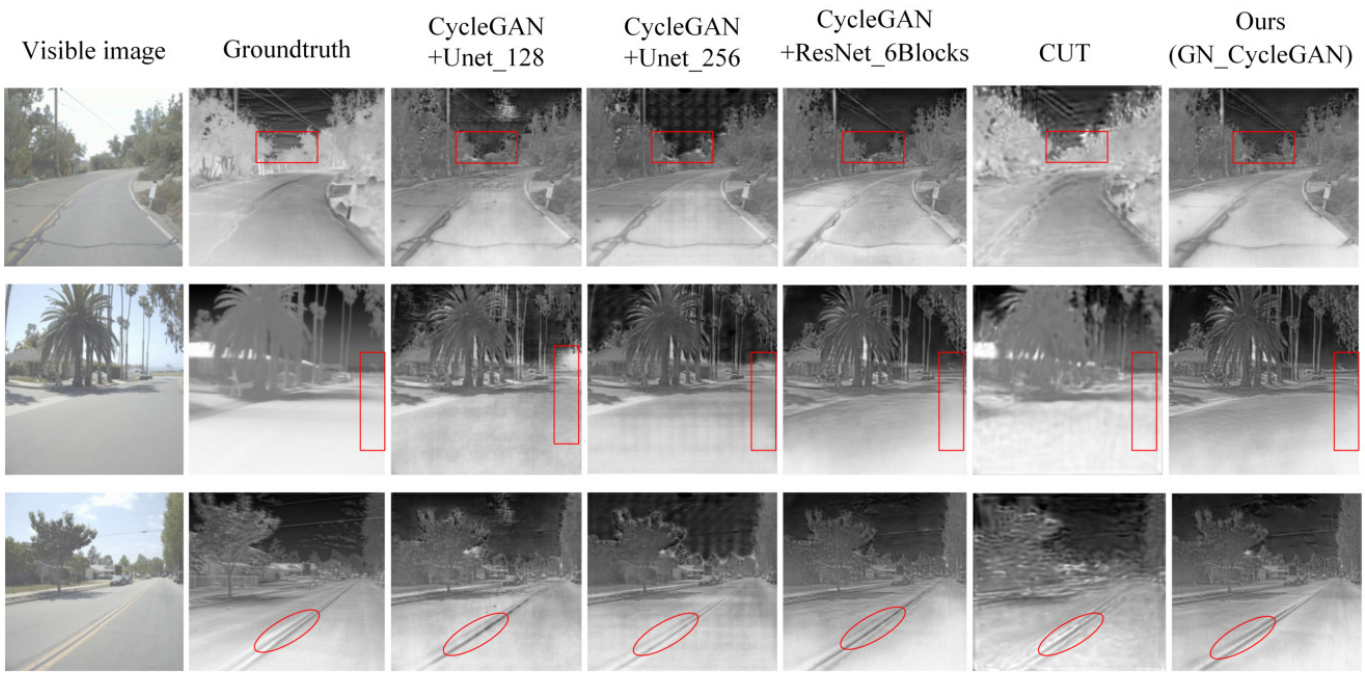

Figure 4. Comparison of CycleGAN and our generation results on the Flir dataset.

As seen from the visualization results in the first data set (first row) of Figure 5, in the CycleGAN+Unet_128 (original) experiment, there is a severe lack of detailed information within the lower left red annotated region compared to the real IR image, and the features in the image are not recognizable. In the ResNet generator network experiments, the image is able to exhibit good discrimination, but there are more heavily shaded overlapping feature sections within the upper right red-labeled area. In the cut method, the resulting image is blurred, and the texture detail is severely missing. In our proposed method, the shortcomings of the two experiments mentioned above are overcome within the red-labelled regions, generating highly discriminative images, rich in detailed information generation, and closest to the real IR image. In the red region of the second data set, the other methods generate images with severe information loss in the red region, while our proposed method is able to avoid the loss of detailed information. The red area of the third data set shows the ground texture detail information. Through observation, it suggests that the other methods that generate the ground image texture clarity are not as good as the proposed method.

To address the problem prone to collapse of the current model training process, a gradient normalization method is used to verify the stability of the model training in this paper by testing the overall direction of the cyclic loss function values of generator A and generator B during the training process. In this paper, the original CycleGAN method and the CBAM_CycleGAN method, which introduces an attention mechanism, are used as control groups. The following visualization results of the loss function obtained during model training are shown in Figures 6 and 7: the cycle loss of generator A and generator B shows that the cycle loss value of the CycleGAN method is high and low until 400 step, which has poor stability; the CBAM_CycleGAN method has less local fluctuations, which has average stability; the cycle loss value of the generator in the proposed method shows a regular gradient downward trend, converges faster and the overall process is smoother. Between 400 step and 1000 step, all three methods tend to be smooth, but the CycleGAN method and CBAM_CycleGAN are more volatile with poor model generation ability. After 1000 steps, compared with CycleGAN and CBAM_CycleGAN methods, the loss value of the proposed gradient cycle generation adversarial network (GN_CycleGAN) tends to stabilize as the gradient decreases, and the cycle loss is lower than that of cycleGAN and CBAM_CycleGAN methods, laterally reflecting that the model has a better generative abil-

ity. On the whole, the gradient-based cyclic adversarial network method (GN_CycleGAN) has an obvious gradient descent pattern and converges quickly, which reflects the excellent stability of the generator network, and, therefore, can prove the effectiveness of the proposed method.

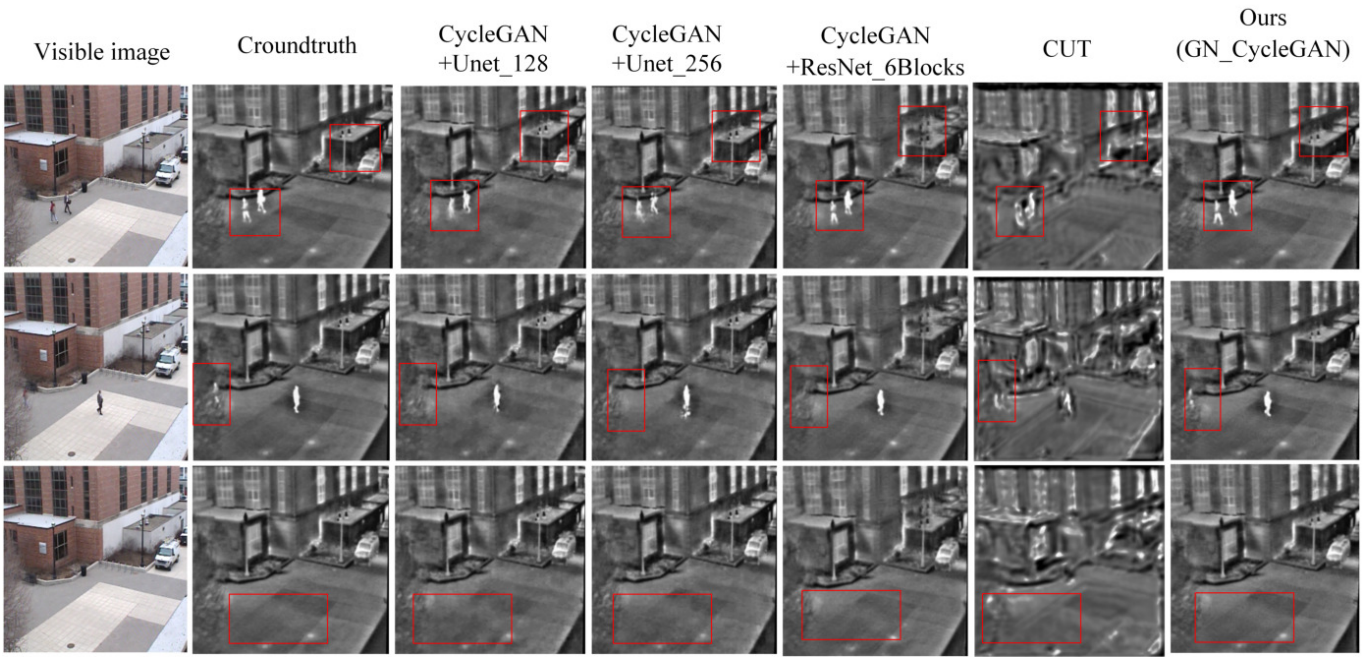

Figure 5. Comparison of CycleGAN and our generation results on the OSU coloured hot dataset.

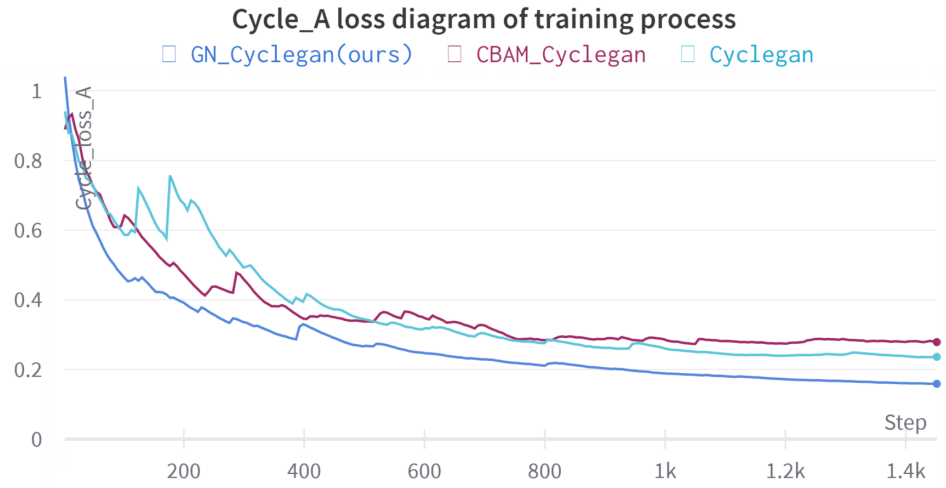

Figure 6. Visualisation of the cycle loss gradient results for generator_A.

Through the experimental data of image evaluation metrics on several datasets and the visualization results of the training model, it can be concluded that the gradient normalization method proposed in this paper has better generative performance in infrared image generation. To address the problem of poor quality and texture details loss in the infrared images obtained by infrared devices in harsh environments, this paper uses a network with stronger feature extraction capability in the generator network, and introduces channel attention and spatial attention mechanisms into the network structure, which have better effects on the extraction of texture details in image generation. To prevent the problem of easy model collapse during the training process, we added a gradient normalization module into the discriminator to further strengthen the convergence speed of model train-

ing. Through experiments, it is shown that the model's convergence speed is faster after introducing this module, and the model training presents gradient descent with better stability. Meanwhile, the experimental data in Table 2 shows that, for the image data obtained from the introduced residual network and attention mechanism, the effect is significantly improved, indicating that the effect of image generation is very satisfactory.

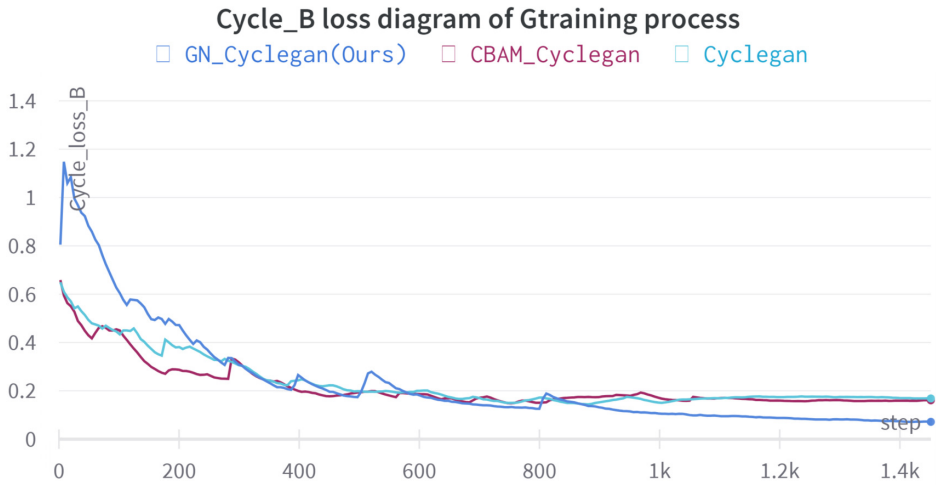

Figure 7. Visualisation of the cycle loss gradient results for generator_B.

4.3. Ablation Experiments

This paper conducts ablation experiments using the Flir dataset to analyze the role of the proposed method in the network structure. The experiments consist of 4 parts.

(1) CycleGAN (baseline): the original method network.

(2) CycleGAN+ResNet: a residual network used in the generator apparatus.

(3) CycleGAN+ResNet+CBAM: a residual network incorporating an attention mechanism is used in the generator.

(4) CycleGAN+ResNet+CBAM+Gradients Normalization(Ours): the gradient normalized cycle generative adversarial network method proposed in this paper.

The statistical results are shown in Table 3. The table shows that, compared to the original method, the PSNR and SSIM metrics, where the residual network was introduced, improved slightly in the experiments. In the experiments with the residual network incorporating the attention mechanism, the PSNR values improved slightly, and the SSIM values improved more than the original method. In the experimental method proposed in this paper, there is an improvement of 2.32% and 9.72% in the PSNR and SSIM metrics, respectively, compared to other methods.

Table 3. Results of objective evaluation indicators for ablation experiments.

Method	PSNR/dB	SSIM
Baseline(CycleGAN)	13.2131	0.4167
CycleGAN+ResNet	13.3275	0.4368
CycleGAN+ResNet+CBAM	13.2403	0.4530
Ours (GN_CycleGAN)	13.5195	0.4572

The results of the ablation experiments are shown in Figure 8. As seen from the figure, the first group data (first row) in the CycleGAN experiment, compared with the original real IR image, differs significantly from the real data in the red annotated region, and the CycleGAN method has a significant feature loss to generate images with a large degree of blurring. Compared with the original method, the red annotated regions of the first

group data in the experiment with the introduction of ResNet generator show less feature loss, and the image details are not sufficiently detailed. By introducing our proposed gradient normalization network method, the loss in the red region of the first dataset is minimal, and the generated image features are more adequate and closest to the original real image. In the second (second row) and third (third row) datasets, compared with the real data, the images generated by the Cyclegan method have insufficient texture detail information in the red annotated area as well as a large loss, and the generated images are blurred with some white shadows resulting in poor image visibility. In the experiments with the introduction of ResNet generator, the quality of generated images is better, with no white shadows and good image visibility. In the experiments after introducing the ResNet generator, the quality of generated images is better. There is no white shadow, and the image visibility is good, but in terms of detail generation, the images from the ResNet generator experiments differ significantly from the real images. In our proposed gradient normalization method, the loss of texture detail information in the generated image quality is minimal, the best image quality is obtained, and the detail information generation is more prosperous.

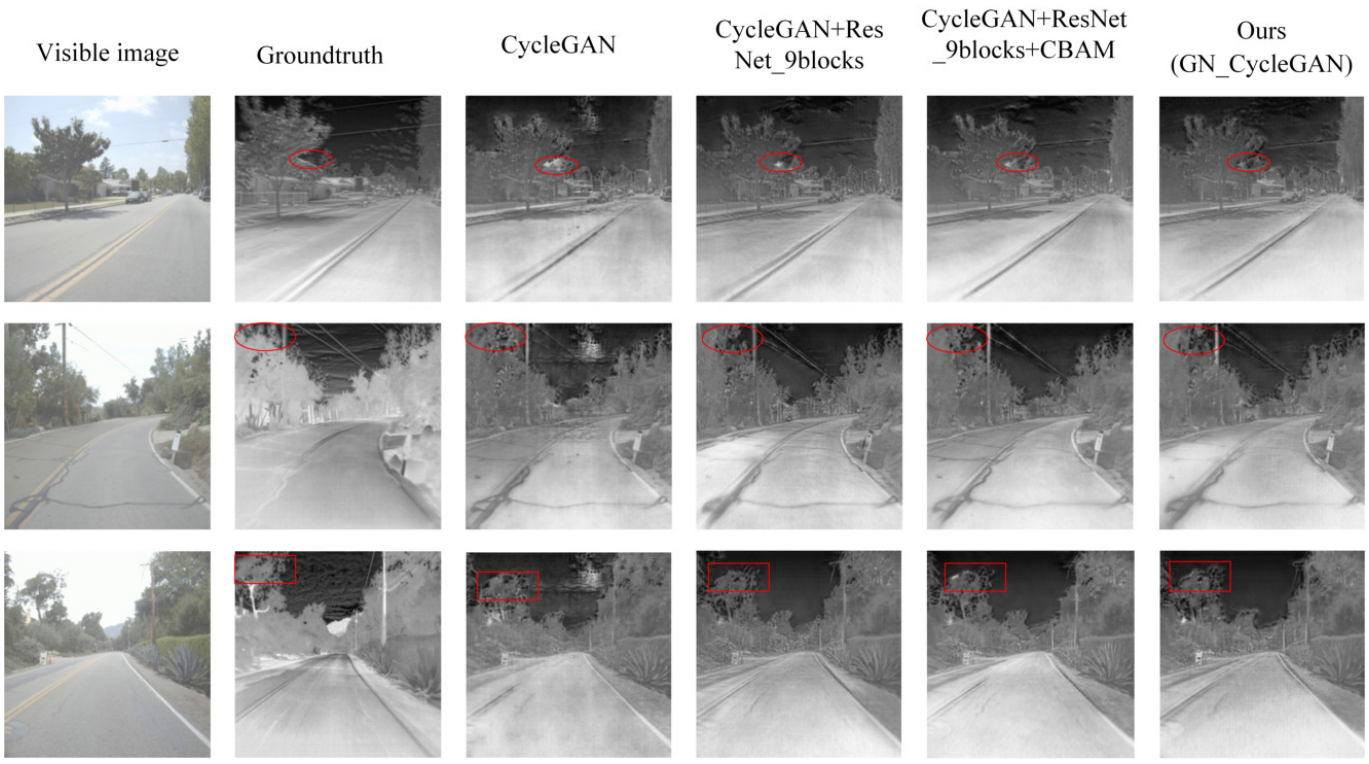

Figure 8. Results of ablation experiments for the Flir dataset.

The visualization results of ablation tests and the evaluation index data show that our proposed method has a more remarkable improvement in the generation of image texture details, and that the generated infrared images are closest to the real infrared images, thus better validating the effectiveness of our method.

5. Discussion

In the current mainstream algorithms for generating infrared images from visible images, we have found that the current methods suffer from poor image quality, loss of texture detail and lack of model stability during the training process. Researchers have made outstanding contributions to address some of these problems. Our work focuses on the inability of infrared cameras to acquire high-quality infrared images with rich texture detail in harsh environments, and uses deep learning methods to compensate for the shortcomings. We propose a gradient-normalized recurrent generative adversarial network

approach for IR images by a ResNet residual network with better image feature extraction capabilities. To reduce the loss of texture details, we add a spatial attention and channel attention mechanism to the generator network, highlighting the ability to perceive image detail regions to improve image generation quality, and to prevent the training model from crashing. In this paper, we introduce a gradient normalization module in the discriminator to speed up the model convergence and maintain the model training stability. The method achieves excellent performance in terms of image quality, texture detail and model stability. It outperforms the current mainstream CycleGAN method in the structural similarity metrics and peak signal-to-noise ratio metrics. In particular, it yields more satisfactory data results on the OSU colour-heat dataset with a 32.2% improvement in peak signal-to-noise ratio metric and an improvement from 0.2479 to 0.7491 in structural similarity. Meanwhile, on the FLIR dataset, our proposed method achieves a 2.3% improvement in peak signal-to-noise ratio, and improves the peak S/N ratio by 2.3% and the structural similarity by 9.7%.

In order to address the problems of low image quality, lack of detail in texture information and unstable model training in the field of infrared image generation, the solution proposed in this paper is compared to the current mainstream generative networks. For example, the method has an advantage in data results compared to the Sparse GANS IR image generation method [43], which mainly focuses on the lack of prominent texture details resulting in poor image quality. The method in this paper is more stable than the conditional generative adversarial networks [44], which have better image generation capabilities but prone to collapse during model training. In contrast to previous work on image generation, we introduce for the first time a gradient normalization module into the field of image generation methods for recurrent generative adversarial networks to stabilise the model training process. The approach in this paper, focusing on the quality of image generation and the model stability, can achieve a good image quality to a certain extent, but the proposed approach also has shortcomings, such as neglecting the cost of model training. The future work will focus on improving the model efficiency with a balance between image quality and model stability.

6. Conclusions

To solve problems concerning poor quality, lack of detail, unstable training models, and easy collapse during infrared image generation, we propose a cycle generative adversarial network method based on gradient normalization. First, the ResNet residual network is employed in the generator to improve image-feature extraction ability. Second, channel attention and spatial attention modules are introduced in the residual network to enhance the image features and texture information highlighted during model training, enhance image generation details, and improve the quality of generated images. Finally, in the discriminator network, a gradient normalization module is adopted to constrain the discriminator network training model and thus ensure its stability. The experimental results show that the gradient-normalised recurrent generative adversarial network method proposed in this paper bears better detail generation capability and model stability.

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

Funding: This research is funded by the National Equipment Development Department of China (Grant No. 41401040105).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Shenyang University of Chemical Technology and Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences for the support.

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

References

1. Wang, Y. Survey on Deep Multi-modal Data Analytics: Collaboration, Rivalry, and Fusion. *ACM Trans. Multimed. Comput. Commun. Appl.* **2021**, *17*, 1. [CrossRef]
2. Wang, Y.; Peng, J.; Wang, H.; Wang, M. Progressive Learning with Multi-scale Attention Network for Cross-domain Vehicle Re-identification. *Sci. China Inf. Sci.* **2022**, *65*, 16103. [CrossRef]
3. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 1–7.
4. Creswell, A.; White, T.; Dumoulin, V. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [CrossRef]
5. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
6. Singh, N.K.; Raza, K. Medical image generation using generative adversarial networks: A review. *Health Inform.* **2021**, *932*, 77–96.
7. Suárez, P.L.; Sappa, A.D.; Vintimilla, B.X. Infrared image colorization based on a triplet dcgan architecture. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017.
8. Isola, P.; Zhu, J.Y.; Zhou, T. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
9. Wang, T.C.; Liu, M.Y.; Zhu, J.Y. High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–21 June 2018.
10. Cui, J.; Zhong, S.; Chai, J. Colorization method of high resolution anime sketch with Pix2PixHD. In Proceedings of the 2021 5th Asian Conference on Artificial Intelligence Technology (ACAIT), Haikou, China, 29–31 October 2021.
11. Dash, A.; Ye, J.; Wang, G. High Resolution Solar Image Generation using Generative Adversarial Networks. *Ann. Data Sci.* **2022**, 1–17. [CrossRef]
12. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 11–15 August 2017.
13. Zhou, C.; Zhang, J.; Liu, J. Lp-WGAN: Using Lp-norm normalization to stabilize Wasserstein generative adversarial networks. *Knowl.-Based Syst.* **2018**, *161*, 415–424. [CrossRef]
14. Zhang, H.; Goodfellow, I.; Metaxas, D. Self-attention generative adversarial networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach Convention Center, Long Beach, CA, USA, 10–15 June 2019.
15. Lin, Z.; Feng, M.; Santos, C.N. A structured self-attentive sentence embedding, Computing Research Repository. *arXiv* **2017**, arXiv:1703.03130.
16. Liu, M.Y.; Tuzel, O. Coupled generative adversarial networks. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 469–477.
17. Taigman, Y.; Polyak, A.; Wolf, L. Unsupervised cross-domain image generation. *arXiv* **2016**, arXiv:1611.02200
18. Mao, X.; Wang, S.; Zheng, L. Semantic invariant cross-domain image generation with generative adversarial networks. *Neurocomputing* **2018**, *293*, 55–63. [CrossRef]
19. Benaim, S.; Wolf, L. One-shot unsupervised cross domain translation. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 2108–2118
20. Zhu, J.Y.; Park, T.; Isola, P. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 21–30 October 2017.
21. Choi, Y.; Choi, M.; Kim, M. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 18–21 June 2018.
22. Choi, Y.; Uh, Y.; Yoo, J. Stargan v2: Diverse image synthesis for multiple domains. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
23. Huang, X.; Liu, M.Y.; Belongie, S. Multimodal unsupervised image-to-image translation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
24. Liu, M.Y.; Huang, X.; Mallya, A. Few-shot unsupervised image-to-image translation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019.
25. Murez, Z.; Kolouri, S.; Kriegman, D.; Ramamoorthi, R.; Kim, K. Image to Image Translation for Domain Adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–21 June 2018.
26. Wang, M.; Li, H.; Li, F. Generative adversarial network based on resnet for conditional image restoration. *arXiv* **2017**, arXiv:1707.04881.
27. Cao, K.; Zhang, X. An improved res-unet model for tree species classification using airborne high-resolution images. *Remote Sens.* **2020**, *12*, 1128. [CrossRef]
28. Demir, U.; Unal, G. Patch-based image inpainting with generative adversarial networks. *arXiv* **2018**, arXiv:1803.07422.

29. Woo, S.; Park, J.; Lee, J.Y. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018.
30. Ma, B.; Wang, X.; Zhang, H.; Li, F.; Dan, J. CBAM-GAN: Generative adversarial networks based on convolutional block attention module. In Proceedings of the International Conference on Artificial Intelligence and Security, New York, NY, USA, 26–28 July 2019.
31. Gul, M.S.K.; Mukati, M.U.; Bätz, M. LightField View Synthesis Using A Convolutional Block Attention Module. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021.
32. Aytekin, C.; Ni, X.; Cricri, F. Clustering and unsupervised anomaly detection with L2 normalized deep auto-encoder representations. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazi, 8–13 July 2018.
33. Wu, Y.L.; Shuai, H.H.; Tam, Z.R. Gradient normalization for generative adversarial networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–18 October 2021.
34. Bhaskara, V.S.; Aumentado-Armstrong, T.; Jepson, A.D. GraN-GAN: Piecewise Gradient Normalization for Generative Adversarial Networks. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 2–5 March 2022.
35. Karras, T.; Laine, S.; Aila, T.A. style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–21 June 2019.
36. Davis, J.; Sharma, V. Background-Subtraction using Contour-based Fusion of Thermal and Visible Imagery. *Comput. Vis. Image Underst.* **2007**, *106*, 162–182. [CrossRef]
37. Sagan, V.; Maimaitijiang, M.; Sidike, P. UAV-based high resolution thermal imaging for vegetation monitoring, and plant phenotyping using ICI 8640 P, FLIR Vue Pro R 640, and thermomap cameras. *Remote Sens.* **2019**, *11*, 330. [CrossRef]
38. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010.
39. Winkler, S.; Mohandas, P. The evolution of video quality measurement: From PSNR to hybrid metrics. *IEEE Trans. Broadcast.* **2008**, *54*, 660–668. [CrossRef]
40. Sara, U.; Akter, M.; Uddin, M.S. Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study. *J. Comput. Commun.* **2019**, *7*, 8–18. [CrossRef]
41. Setiadi, D.I.M. PSNR vs SSIM: Imperceptibility quality assessment for image steganography. Multimed Too. PSNR vs. SSIM: Imperceptibility quality assessment for image steganography. *Multimed. Tools Appl.* **2021**, *80*, 8423–8444. [CrossRef]
42. Park, T.; Efros, A.A.; Zhang, R.; Zhu, J.Y. Contrastive learning for unpaired image-to-image translation. In Proceedings of the European Conference on Computer Vision, Online, 23–28 August 2020; pp. 319–345.
43. Qian, X.; Zhang, M.; Zhang, F. Sparse gans for thermal infrared image generation from optical image. *IEEE Access* **2020**, *8*, 180124–180132. [CrossRef]
44. Chen, F.; Zhu, F.; Wu, Q. InfraRed Images Augmentation Based on Images Generation with Generative Adversarial Networks. In Proceedings of the 2019 IEEE International Conference on Unmanned Systems (ICUS), Beijing, China, 17–19 October 2019.