WILEY | Hindawi

Research Article

I-GANs for Infrared Image Generation

Bing Li [iD],[1] Yong Xian,[1] Juan Su,[1] Da Q. Zhang,[1] and Wei L. Guo[2]

[1]Xi'an High-Tech Research Institute, Xi'an 710025, China
[2]Xi'an Satellite Control Center, Xi'an 710043, China

Correspondence should be addressed to Bing Li; libingbenyi@163.com

The making of infrared templates is of great significance for improving the accuracy and precision of infrared imaging guidance. However, collecting infrared images from fields is difficult, of high cost, and time-consuming. In order to address this problem, an infrared image generation method, infrared generative adversarial networks (I-GANs), based on conditional generative adversarial networks (CGAN) architecture is proposed. In I-GANs, visible images instead of random noise are used as the inputs, and the D-LinkNet network is also utilized to build the generative model, enabling improved learning of rich image textures and identification of dependencies between images. Moreover, the PatchGAN architecture is employed to build a discriminant model to process the high-frequency components of the images effectively and reduce the amount of calculation required. In addition, batch normalization is used to optimize the training process, and thereby, the instability and mode collapse of the generated adversarial network training can be alleviated. Finally, experimental verification is conducted on the produced infrared/visible light dataset (IVFG). The experimental results reveal that high-quality and reliable infrared data are generated by the proposed I-GANs.

1. Introduction

Due to the limitations of the application background and support capabilities, the template used in infrared imaging guidance is usually a visible image, while the real-time image itself is infrared. The imaging principles of infrared and visible are different, which results in a large feature disparity between the infrared image and the visible image. As a result, the difficulty of scene matching in infrared imaging guidance increases. If the infrared image is used as the reference image for matching, the matching accuracy and precision can be improved. Moreover, the matching difficulty can be reduced. However, relying solely on an off-site field to obtain infrared reference maps is time-consuming, and it is also arduous to obtain infrared images of targets in complex environments and harsh climates. Compared with testing in the field, the use of infrared image simulation technology to generate the infrared characteristics of the scene in the environment of interest can not only effectively reduce the cost of acquiring infrared data but also generate a large amount of infrared data that is difficult to obtain in the field under a variety of natural environments and scene conditions. In this way, the generated infrared data can be used in the fields of aviation, aerospace, navigation, meteorology, geology, and agriculture by providing basic and reliable data for detection [1], classification [2], positioning, identification, tracking purposes, etc. Therefore, generating infrared reference maps through infrared image simulation technology is highly significant for military and civilian applications.

In recent years, with the continuous improvement of computer performance [3, 4] and the rapid development of deep learning theory, many new neural network-based generation models have been proposed. Among these, generative adversarial networks (GANs) [5] have demonstrated a unique capacity to meet research and application needs in many fields and have accordingly become one of the most critical research hotspots in the field of artificial intelligence [6, 7]. Antipov et al. used conditional generative adversarial networks (CGAN) to generate face images [8]. Through applying GANs to the field of face turning (which refers to a technique for synthetizing high definition (HD) frontal face images from a single-sided face image), Huang

and Tran proposed two-pathway generative adversarial networks (TP-GANs) [9] and disentangled representation learning-generative adversarial networks (DR-GANs) [10], respectively. The Markov-based Markovian generative adversarial networks (MGANs) [11] have the same synthesis speed as texture network [12] in generating image textures. Isola et al. demonstrated that pix2pix approach could realize the conversion of black and white to colour, satellite to map, semantic to street view, and edge to photo [13]. Moreover, the image textures and backgrounds generated by BigGAN [14] are more realistic, although the computation complexity of this approach is high. Subsequently, in order to improve the learning performance by taking advantage of the improvement in image generation quality, Donahue and Simonyan proposed BigBiGAN based on the BigGAN model, extending this approach to the image learning context by adding encoders and modifying the identifier [15]. Image super resolution generative adversarial networks (SRGAN) used residual networks (ResNets) and VGG networks [16] as generators and discriminators, respectively, to attain a better texture detail learning effect [17]. In order to solve the lifelong learning problem of the generative model, Zhai et al. presented the Lifelong GAN [18]. He et al. proposed a dual learning mechanism in which the neural machine translation system can automatically learn from unlabeled data through a dual learning game [19]. Following the idea of dual learning, Yi et al. used the DualGAN model of dual learning to achieve cross-domain image generation [20], and Zhu et al. introduced cycle consistency into GANs to extend the image-to-image conversion work [21]. Choi et al. first proposed a novel and scalable method, StarGAN, which is capable of converting images to images translation for multiple domains from using only one model [22]. Beginning with RGB images from Kinect and curve normal maps, Karras et al. proposed a generative adversarial model called Style-GAN, which takes normal surface as the basis for the generative adversarial networks used to generate images [23]. Based on Style-GAN model, Yang and Lim proposed a framework capable of generating face images that fall into the same distribution as that of a given one-shot example [24]. Besides, Richardson et al. presented a generic image-to-image translation framework Pixel2Style2Pixel (pSp). The pSp framework is based on a new encoder network that directly generates a series of style vectors which are fed into a pretrained Style-GAN generator, forming the extended $W+$ latent space [25]. Chen et al. presented a domain adaptive image-to-image conversion (DAI2I) framework, which is suitable for the I2I model of samples outside the domain [26].

At present, the majority of GANs-based image generation researches have applied GANs to face synthesis, texture generation, sketch-to-photo applications, transforming visible images to night vision images, etc. However, few studies have been published on the use of GANs models in the field of infrared image simulation. In view of the high cost, comparatively small quantities, and the relative difficulty of obtaining infrared data in the off-site field, this paper proposes an infrared image generation method based on generative adversarial networks (infrared generative adversarial networks, or I-GANs), which is capable of simulating and generating infrared images on the basis of visible images. Besides, the generated infrared images can be used to create infrared reference maps, which provide reliable infrared data and expand infrared databases. Based on CGAN architecture, the I-GANs algorithm employs the D-LinkNet network to build the generation network, using visible images and infrared simulation samples as the inputs and outputs, respectively. Then, the real target sample and the generated simulation sample are utilized to train the PatchGAN-based discrimination network, which outputs the probability of a generated sample belonging to the corresponding category. Through alternating iterative training of the generation network and the discriminant network, the final generated infrared simulation samples have essentially the same data distribution as the real samples.

The novelty of the work in this paper can be summarized as follows: (1) innovation of research background. We present a novel generation adversarial network algorithm (i.e., I-GANs) with infrared image simulation as the research background, which has a reliable reference value for the subsequent infrared image generation researches; (2) we introduce a D-LinkNet module into conditional GANs. Armed with D-LinkNet, the generator can better preserve the spatial details of the images and achieve multiscale feature fusion.

2. Related Work

Generative adversarial networks (GANs) were first proposed by Goodfellow et al. at the 28th International Conference on Neural Information Processing Systems in 2014 [5]. The generative adversarial networks are a new generative model developed on the basis of a deep generative model. The significant difference between this model and other generative models lies in its use of an adversarial approach. It first learns the difference between the generated sample and the training sample through the discriminator and then guides the generator to reduce this difference rather than to directly target the differences between the data distribution and the model distribution. At present, GANs are one of the most significant research hotspots in the field of artificial intelligence.

2.1. Generative Adversarial Networks. The key concept behind GANs involves setting up a zero-sum game to achieve learning through the confrontation between two players. In the zero-sum game, one player acts as the generator while the other acts as the discriminator. The generator's main task is to generate samples that appear as identical as possible to the training samples, thereby deceiving the other player. For the discriminator, the goal is to accurately determine whether the input samples belong to the set of real training samples. In GANs, the generation network and the adversarial network are often thought of as analogous to a counterfeiter of banknotes and a detector of forged currency. The GANs training process thus resembles the following

procedure: the counterfeiter continues to increase the sophistication of their forged banknotes in order to produce counterfeit banknotes that are as identical as possible to real currency, in the hope that the forgery detector will fail to spot the forgery; for their part, the money detector constantly improves their ability to identify counterfeit banknotes. As the GANs training process continues, both the counterfeiter's ability to manufacture convincing counterfeit notes and the money detector's ability to identify forgeries will continually increase [20].

The GANs consist of two networks, a generative network (generator G) and an adversarial network (discriminator D), which corresponds to the generative and the adversarial model, respectively. The basic framework of the original generative adversarial networks is illustrated in Figure 1.

In the original GANs, the value function $V(G, D)$ [5, 27] is defined as follows:

$$V(G, D) = \mathbb{E}_{\mathbf{x} \sim p_{data}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z}[\log(1 - D(G(\mathbf{z})))], \tag{1}$$

where $x \sim p_{data}$ represents the distribution of x taken from real data, $z \sim p_z$ indicates that the random noise z comes from simulated data (such as a Gaussian noise distribution), $\mathbb{E}(\cdot)$ is the expected value, and G tries to minimize this objective while an adversarial D tries to maximize it; i.e., $G* = \arg\min_G \max_D V(G, D)$.

2.2. Conditional Generative Adversarial Networks.

With the goal of remedying the original GANs' inability to generate pictures with specific attributes, Mirza and Osindero proposed the conditional generative adversarial networks (CGAN) [28]. The core concept of the CGAN involves integrating condition information y into the generator and discriminator. Condition y can be any label information, such as the facial expressions of face images and image categories. The CGAN network structure is presented in Figure 2.

The objective of a CGAN can be expressed as follows:

$$\ell_{CGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))]. \tag{2}$$

3. Methods

3.1. Objective.

In this section, based on the CGAN framework, we proposed the I-GANs algorithm which uses images as input rather than random noise. In order to make better use of the structural information contained in the input image, the L1 objective function is introduced into the loss function as follows:

$$\ell_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1]. \tag{3}$$

The loss function of I-GANs is then finally defined as follows:

$$G^* = \arg\min_G \max_D \ell_{CGAN}(G, D) + \lambda \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1]. \tag{4}$$

3.2. Generative Networks.

The network of the common encoder-decoder structure operates by first downsampling to a low dimension and then upsampling to the original resolution. By contrast, D-LinkNet [29], which uses LinkNet as the basic framework and then introduces a residual network [30], has the advantages of employing skip connection (used to retain pixel-level detailed information at different resolutions), residual blocks, and encoder-decoder systems, thus increasing the receptive fields of the network, retaining the spatial detail information of the image, and realizing multiscale feature fusion.

In the proposed I-GANs algorithm, D-LinkNet is used to construct a generative network. More specifically, in this article, D-LinkNet is designed to receive images of size 256×256 as input. As shown in Figure 3, D-LinkNet is composed of three parts, A, B, and C, which are the encoder part, the central part, and the decoder part, respectively. In the encoder part, ResNet34 [30], which is trained on the ImageNet dataset, is used as the encoder. In the central part, dilated convolution with shortcut is added to enhance the network's recognition ability, expand the receptive field, and fuse multiscale information. Finally, the decoder part uses transposed convolution [31] layers to conduct upsampling, restoring the resolution of the feature map from 8×8 to 256×256.

The center dilation part of this D-LinkNet can be unrolled into the structure illustrated in Figure 4. From top to bottom in the figure, if the dilation rates of the stacked dilated convolution layers are 2, 1, and 0, respectively, then the corresponding numbers of receptive fields are 7, 3, and 1; finally, the results of each branch are added together, and the characteristics of the fusion are obtained. Since the encoder part of the D-LinkNet contains five downsampling layers, while the size of the input data is 256×256, the encoder output feature map will be of the size 8×8. In this case, D-LinkNet uses dilated convolution layers with a dilation rate of 1 and 2 in the center part. Thus, the feature points on the last center layer will yield 7×7 points on the first center feature map, covering the main part of the first center feature map.

3.3. Adversarial Networks.

In the I-GANs, the adversarial network is constructed using the convectional PatchGAN classifier. The main idea behind PatchGAN is as follows: since GANs are used to build high-frequency information, there is no need to input the entire image into the discriminator; instead, the discriminator can make true or false judgments about each block of the image, which penalises the structure only on the scale of the image block. Therefore, the I-GANs' discriminator only needs to pay attention to the local structure of the image (which can effectively reduce the number of parameters in training), model the high-frequency components of the image, and rely on the L1 items to ensure the accuracy at low frequencies.
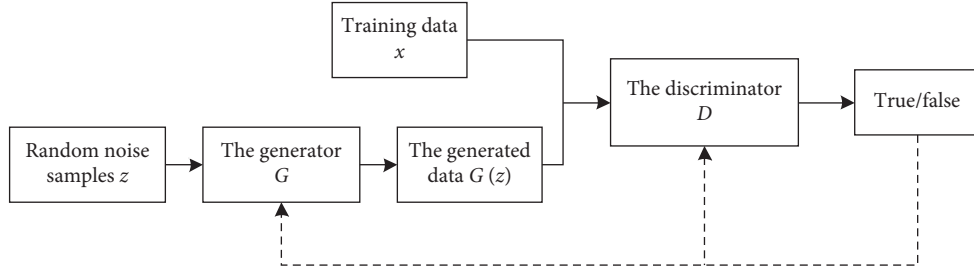
FIGURE 1: The basic framework of the original GANs. The GANs consist of two networks: a generative network (generator G) and an adversarial network (discriminator D).
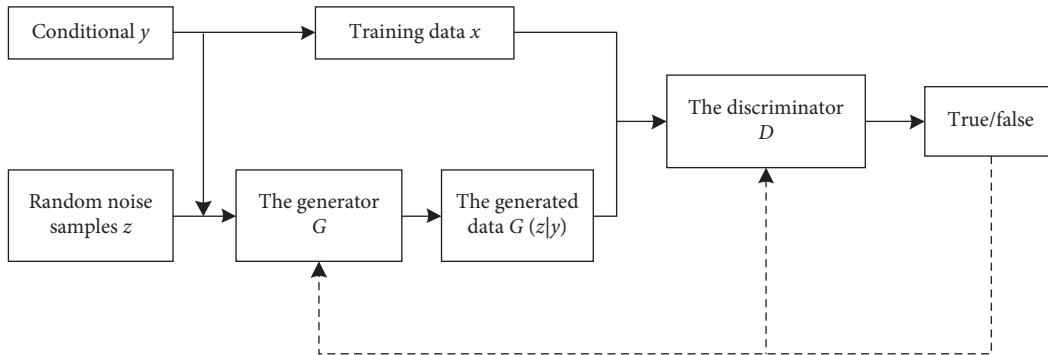

FIGURE 2: The basic framework of CGAN. CGAN has an additional condition y.

4. Results and Discussion

4.1. Datasets. UAV is equipped with a thermal infrared camera and a visible camera (both of which are coaxially installed) to capture the desired target and scene in the designated area. In brief, the designated area is photographed using a coaxial infrared camera and a visible-light camera simultaneously. Targets in the data include buildings (with materials including steel, concrete, cement, and various types of bricks), vehicles (including trucks and buses), radar covers, power stations (e.g., thermal and hydroelectric), oil depots, highways (with materials including cement and asphalt), runways, grasslands (both real and artificial), trees, and rivers (or ponds). Scenes in the data include cities, campuses, streets, factories, residential areas, transportation hubs, and rivers. Meteorological conditions identified in the data collection include sunny, cloudy, hazy, and rainy. We name this dataset "IVFG."

4.2. Subjective Evaluation. In order to evaluate the proposed I-GANs methods, we conducted a large number of experiments on the IVFG dataset. The generation effect of infrared-generated images is evaluated by means of subjective observation and objective index verification.

Next, infrared-generated images of buildings, chimneys, and cooling towers, generated by the I-GANs algorithm, are presented in Figures 5–7. The building materials in Figure 5 include steel, concrete, cement, and various types of bricks. Through visual interpretation and subjective evaluation, it can be determined that the grey information and contour

information of the infrared-generated images are closer to those of the real infrared images. In addition, the similarity between the two is higher, and the infrared generation effect is superior.

4.3. Objective Evaluation. Generally speaking, the greater the similarity of the grey characteristics between generated infrared images and those obtained in real time, the better the infrared image generation results. In order to objectively evaluate the I-GANs algorithm's effectiveness at generating infrared images, we calculate the Root Mean Square Error (RMSE) and feature similarity (Feature SIMilarity, FSIM) [32] between infrared generation-based templates (which are split off from infrared-generated results via human-computer interaction) and infrared real-time maps, respectively.

The RMSE is a measure of the degree of information change between the two images, which reflects the difference in grey values. In general, the smaller the RMSE value, the smaller the greyscale difference between the two, that is, the better the generation effect of the infrared-generated images. On the contrary, the larger the RMSE value, the worse the generation effect of the infrared-generated image. Moreover, FSIM represents an improvement of structural similarity, which not only uses phase consistency to extract rich texture, edge, and structure information, but also introduces the contrast information of the gradient amplitude to extract images, enabling the structural differences between images to be evaluated. Generally speaking, the greater the FSIM value, the higher the similarity between images (i.e., the better
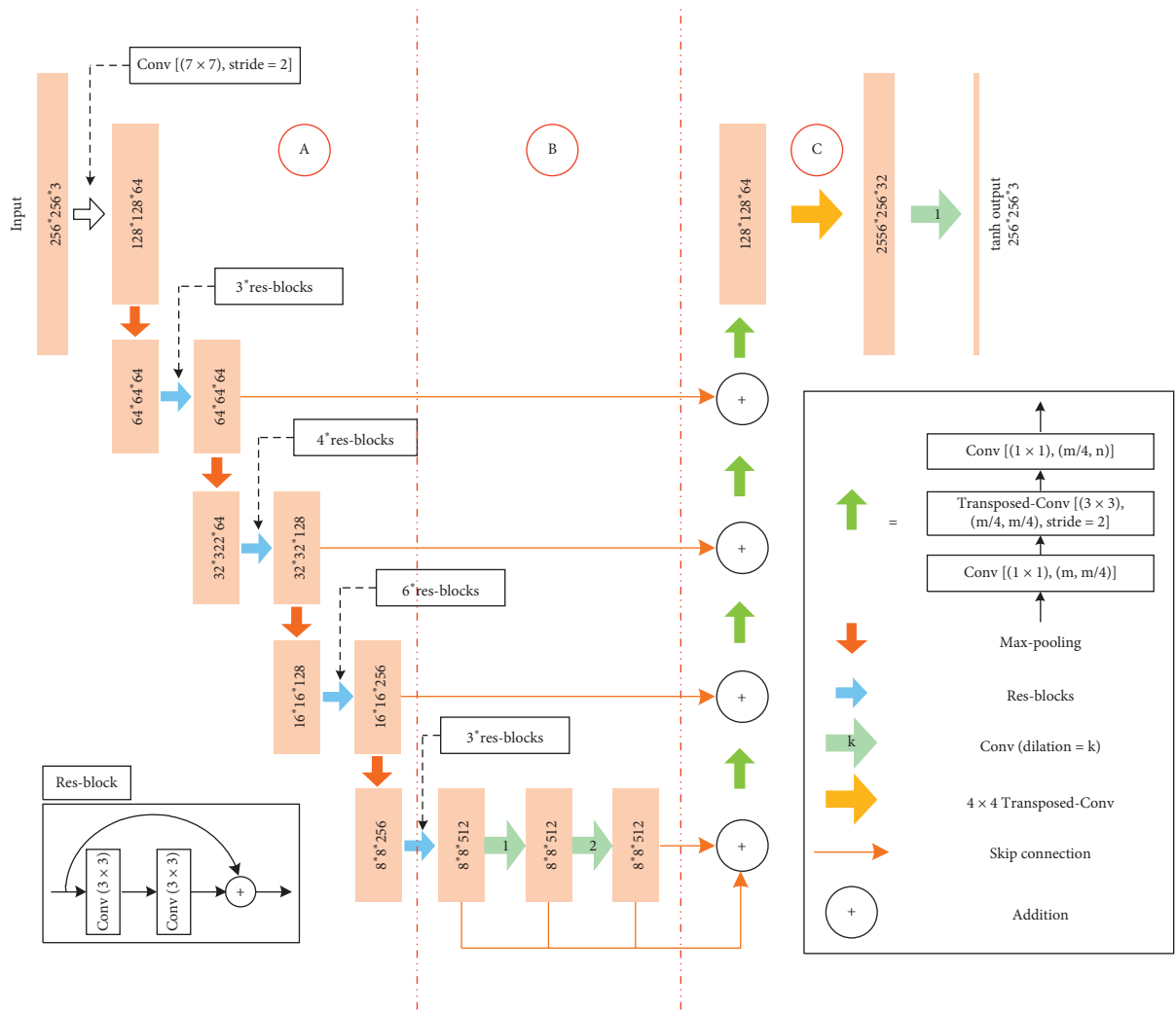
FIGURE 3: A diagram of D-LinkNet. Each orange rectangular block represents a multichannel features map. Part A and Part C are the encoder and the decoder of D-LinkNet, respectively. Here, D-LinkNet uses ResNet34 as encoder.
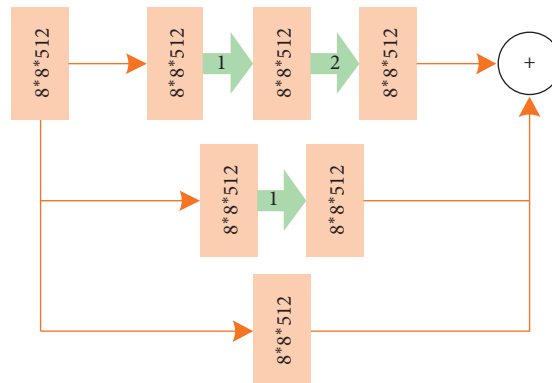

FIGURE 4: The unrolled structure of the center dilation part of D-LinkNet. From top to bottom, the receptive fields are 7, 3, and 1, respectively.

the infrared generation). Because the user tends to pay more attention to the infrared generation effect of the target, this paper only calculates the RMSE and FSIM between the target's infrared real-time map and the infrared generation map. The RMSE and FSIM are calculated according to the following equations:
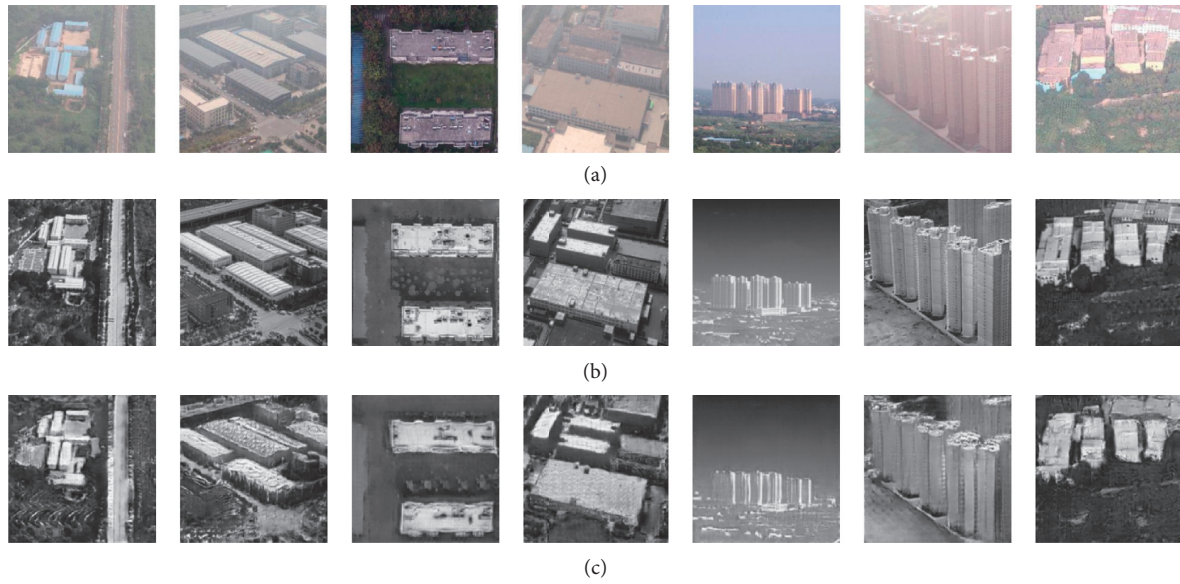
(a)

(b)

(c)

FIGURE 5: Examples of infrared-generated images of buildings produced on the basis of the I-GANs algorithm. (a) Visible images. (b) Real infrared images. (c) Infrared-generated images.
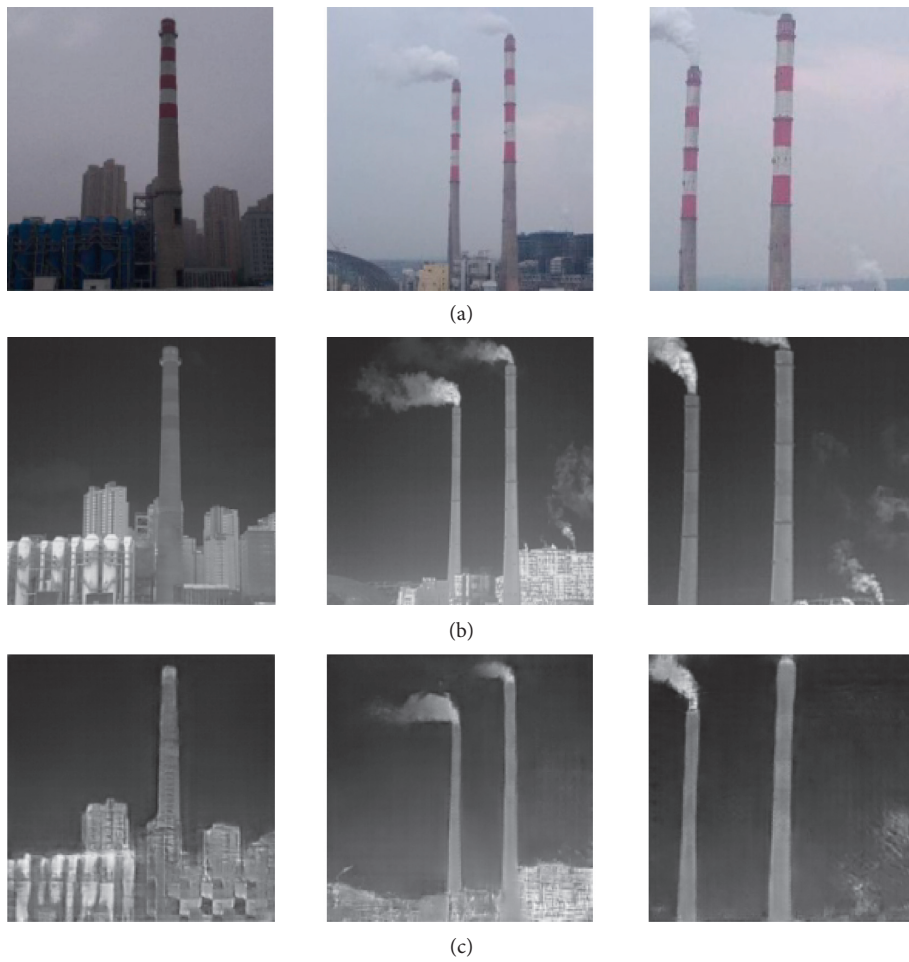

(a)

(b)

(c)

FIGURE 6: Examples of infrared-generated images of chimneys produced on the basis of the I-GANs algorithm. (a) Visible images. (b) Real infrared images. (c) Infrared-generated images.
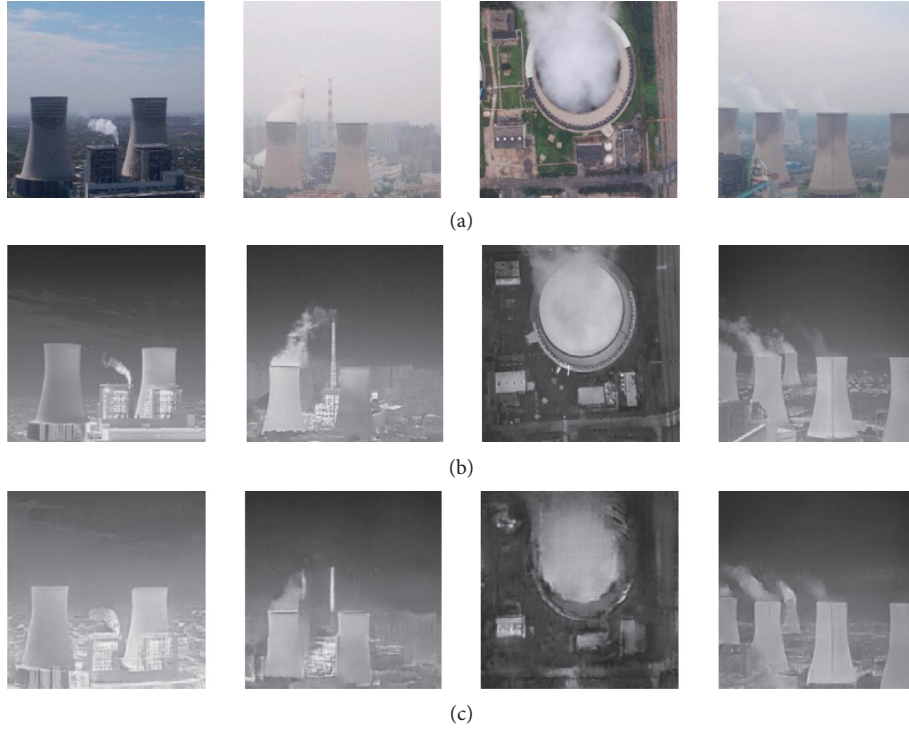
(a)

(b)

(c)

FIGURE 7: Examples of infrared-generated images of cooling towers produced on the basis of the I-GANs algorithm. (a) Visible images. (b) Real infrared images. (c) Infrared-generated images.

$$\text{RMSE} = \left(\frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} \left(I(x, y) - S(x, y) \right)^2 \right)^{1/2},$$

$$\begin{cases} S_{PC} = \dfrac{2\text{PC}_1(I) \cdot \text{PC}_2(S) + T_1}{\text{PC}_1^2(I) + \text{PC}_2^2(S) + T_1}, \\[3mm] S_G = \dfrac{2G_1(I) \cdot G_2(S) + T_2}{G_1^2(I) + G_2^2(S) + T_2}, \\[3mm] S_L = \left[S_{PC} \right]^{\alpha} \cdot \left[S_G \right]^{\beta}, \\[3mm] \text{FSIM} = \dfrac{\sum S_L \cdot \max\left(\text{PC}_1(I), \text{PC}_2(S) \right)}{\sum \max\left(\text{PC}_1(I), \text{PC}_2(S) \right)}, \end{cases} \quad (5)$$

where I and S represent the infrared measure of the target and the infrared simulation chart, respectively. Moreover, $\text{PC}_1(I)$ and $\text{PC}_2(S)$ represent the phase consistency of I and S, respectively, while $G_1(I)$ and $G_2(S)$ represent the gradient amplitude of I and S, respectively.

In this paper, in order to verify the generation results, the proposed I-GANs algorithms are compared with three GANs-based algorithms, the generators of which are U-Net256, ResNet9, and ResNet34, respectively. Among them, the algorithm with U-Net256 as generator is the classic Pix2pix algorithm [13], and the following are all described with "Pix2pix". Besides, in the following, the GANs-based algorithms construct generators with ResNet9 and

ResNet34, respectively, are called "Resnet9" and "Resnet34," respectively. The network structure of the four algorithms participating in the experimental comparison is shown in Table 1.

There are 1374 sets of infrared/visible light images (1374 infrared images and 1374 visible images) in the dataset involved in the experiment in this paper. The training samples and test samples are constructed according to the ratio of 1070 : 304. For the RMSE index, smaller value is superior; among the FSIM index, larger value is superior. We make statistics on the number of superior and inferior values of the actual values of the image quality evaluation indexes and define the statistical result as the ratio of superiority and inferiority (RSI).

We count the RMSE and FSIM values between all infrared images generated by these four algorithms and the corresponding real infrared images. We also calculate the average value of each index value (represented by mRMSE and mFSIM) and the RSI of the index values between the four algorithms. The statistical results are shown in Table 2. RMSE needs to consider the grey value of the corresponding points of the two images. However, there are differences (such as scale transformation, rotation, and angle) between the visible image and the real infrared image—it is not possible to fully pair the corresponding points of the target's infrared generation reference map and the same coordinates in the real infrared image. This affects the calculation of the square root error, which may lead to a larger RMSE value.

According to the experimental data given in Table 2, it can be concluded that

TABLE 1: The network structure of the four GANs algorithms.

Method	Networks				
	Generator				Discriminator
	U-net256	ResNet9	ResNet34	D-LinkNet34	PatchGAN
Pix2pix	√				√
Resnet9		√			√
Resnet34			√		√
Our method				√	√

TABLE 2: The average and the superior/inferior sample numbers of the evaluation indexes.

Method	The average of the evaluation indexes		The superior/inferior sample numbers of the evaluation indexes			
			RMSE samples		FSIM samples	
	mRMSE	mFSIM	Our'<Other'	Our'>Other'	Our'>Other'	Our'<Other'
Pix2pix	35.01	0.721	207	97	220	84
Resnet9	34.42	0.722	180	124	220	84
Resnet34	37.04	0.700	228	76	243	61
Our method	33.82	0.737	—	—	—	—

(a). Among the four algorithms, our method has the smallest mRMSE value of 33.82 and the largest mFSIM value of 0.737, which means that the quality of the infrared images generated by our method is the best;

(b) In the 304 groups of comparative data, the numbers of samples where our method's RMSE index values are better than Pix2pix, Resnet9, and Resnet34 are 207, 180, and 228, respectively;

(c) In the 304 groups of comparative data, the numbers of samples where our method's FSIM index values are better than Pix2pix, Resnet9, and Resnet34 are 220, 220, and 243, respectively.

According to the above analysis, the quality of the infrared image generated by our method is better than the other three GANs-based algorithms.

4.3.1. Statistical Results of RMSE. In order to express the experimental results more intuitively, based on the ascending order of the 304 RMSE values obtained by our algorithm, a comparison chart of the experimental results of our method and Pix2pix is drawn. As shown in Figure 8, the experimental results of our method are represented by the curve "", and the experimental results of Pix2pix are represented by the scattered points "".

It can be seen from Figure 8 that the number of "" above the curve "" is obviously more than those below the curve. Among the RMSE index results of our method, 207 index values are superior to the Pix2pix, and 97 index values are inferior to the Pix2pix. That is, the RMSE index RSI of the two algorithms is 207 : 97, indicating that, among the infrared images generated by our method, 207 images are with better quality than the Pix2pix algorithm.

According to the drawing standard in Figure 8, the RMSE index results obtained by our method, Resnet9, and Resnet34 algorithms are drawn, as shown in Figure 9. In Figure 9, the RMSE values of our method, Resnet9, and Resnet34 are represented by the curve "", the scattered point "", and the scattered point "", respectively.

As demonstrated in Figure 9, the number of "" and "" distributed above the curve "" is obviously more than those below the curve. The RMSE index RSI of our method and Resnet9 algorithm is 180 : 124, and the RSI of our method and Resnet34 algorithm is 228 : 76. These illustrate that the quality of infrared images generated by our method is significantly better than Resnet9 and Resnet34 algorithms.

4.3.2. Statistical Results of FSIM. According to the drawing standard in Figure 8, the FSIM index results obtained by our method and Pix2pix are drawn, as shown in Figure 10. In Figure 10, the FSIM values of our method and Pix2pix are represented by the curve "————" and the scattered point "■", respectively.

As shown in Figure 10, the number of "●" below the curve "▲" is obviously more than those above the curve. Among the FSIM index results of our method, 220 index values are superior to the Pix2pix, and 84 index values are inferior to the Pix2pix. This indicates that the FSIM index RSI of the two algorithms is 220 : 84, which means that among the infrared images generated by our method, 220 images are with better quality than the Pix2pix algorithm.

Similarly, we draw the FSIM index results obtained by our method, Resnet9, and Resnet34 algorithms. As shown in Figure 11, the FSIM values of our method, Resnet9, and Resnet34 are represented by the curve "", the scattered point "", and the scattered point "", respectively.

As shown in Figure 11, the number of "" and "" distributed below the curve "" is obviously more than those above the curve. The FSIM index RSI of our method and Resnet9 algorithm is 220 : 84, and the RSI of our method and Resnet34 algorithm is 243 : 61. These also show that the
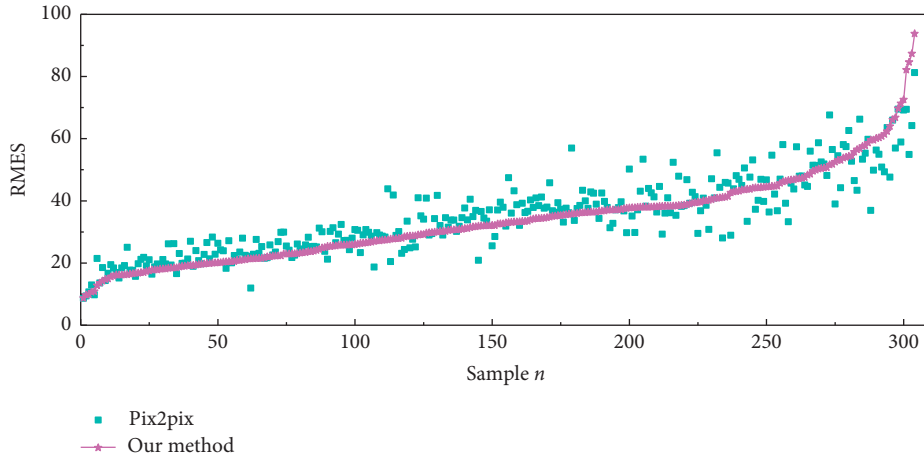
FIGURE 8: The RMSE results between the real infrared images and the infrared images generated by our method and the Pix2pix. The *X*-axis represents different test samples and the *Y*-axis represents the RMSE value corresponding to the sample.
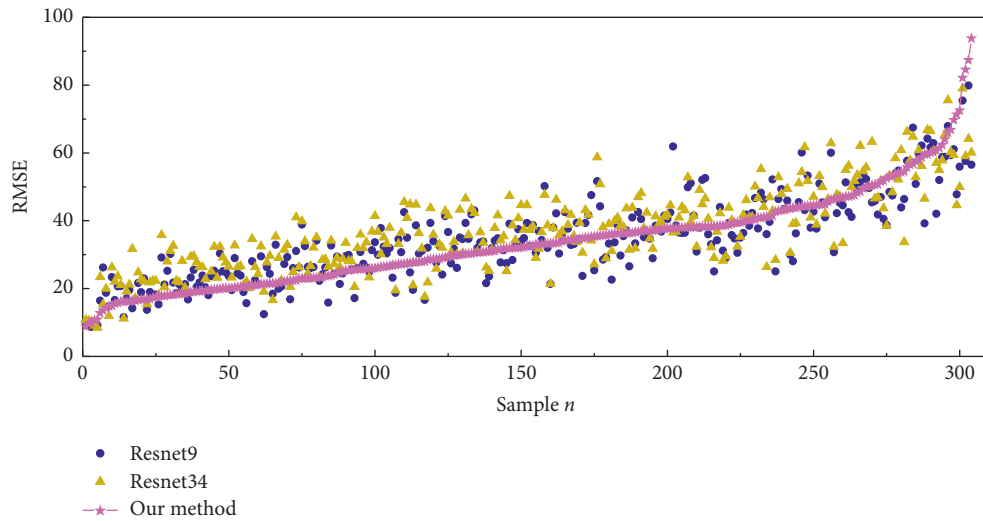

FIGURE 9: The RMSE results between the real infrared images and the infrared images generated by our method and the Resnet9 and Resnet34. The *X*-axis represents different test samples and the *Y*-axis represents the RMSE value corresponding to the sample.
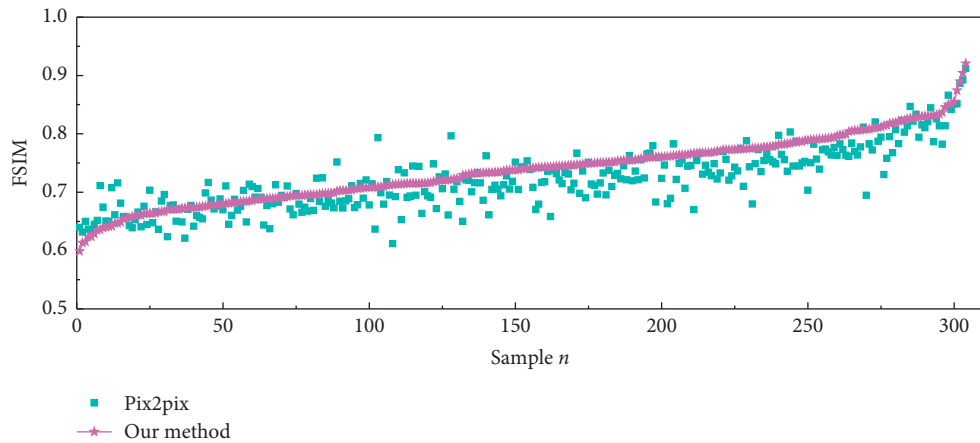

FIGURE 10: The FSIM results between the real infrared images and the infrared images generated by our method and the Pix2pix. The *X*-axis represents different test samples and the *Y*-axis represents the FSIM value corresponding to the sample.
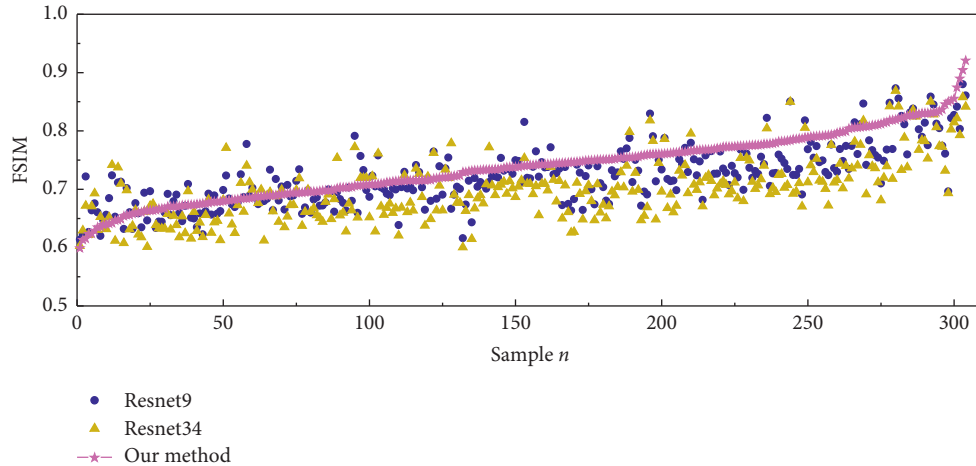
FIGURE 11: The FSIM results between the real infrared images and the infrared images generated by our method and the Resnet9 and Resnst34. The *X*-axis represents different test samples and the *Y*-axis represents the FSIM value corresponding to the sample.

quality of infrared images generated by our method is significantly better than Resnet9 and Resnet34 algorithms.

Based on subjective interpretation and objective analysis, it can be determined that the infrared images generated by our method (that is, I-GANs algorithm) are similar to the real infrared images; i.e., the infrared generation effect is well.

5. Conclusions

Infrared reference map preparation plays an important role in improving the accuracy and precision of infrared imaging guidance. This paper proposes an infrared image generation algorithm based on generative adversarial networks, which is named I-GANs. The algorithm introduces the D-LinkNet network to build a generation network for the purpose of learning image textures and discovering the dependencies between images. Furthermore, PatchGAN is adopted to construct a discriminant model, which can effectively process the high-frequency components of the image and reduce the amount of calculation required. In the training process, batch normalization and the Adam are utilized to optimize the training process in order to alleviate training instability and mode collapse. The simulation on the produced infrared/visible light image data (IVFG) reveals that the proposed I-GANs algorithm can generate high-quality infrared images, which are more realistic and similar to the real infrared images.

Data Availability

The data used to support this research was collected by the authors through UAV, which is equipped with a thermal infrared camera and a visible camera (both of which are coaxially installed) to capture the desired target and scene in the designated area; in brief, the designated area is photographed using a coaxial infrared camera and a visible-light camera simultaneously. Targets in the data include buildings (with materials including steel, concrete, cement, and various types of bricks), vehicles (including trucks and buses),

radar covers, power stations (e.g., thermal and hydroelectric), oil depots, highways (with materials including cement and asphalt), runways, grasslands (both real and artificial), trees, and rivers (or ponds). Scenes in the data include cities, campuses, streets, factories, residential areas, transportation hubs, and rivers. Meteorological conditions identified in the data collection include sunny, cloudy, hazy, and rainy.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. Xi, C. Wang, X. Yang, and B. Yang, "Limited-budget output consensus for descriptor multiagent systems with energy constraints," *IEEE Transactions on Cybernetics*, vol. 50, no. 11, pp. 4585–4598, 2020.

[2] X. Yang, G. Lin, Y. Liu, F. Nie, and L. Lin, "Fast spectral embedded clustering based on structured graph learning for large-scale hyperspectral image," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2020.

[3] N. Cai, M. He, Q. Wu, and M. Khan, "On almost controllability of dynamical complex networks with noises," *Journal of Systems Science and Complexity*, vol. 32, pp. 1125–1139, 2017.

[4] Z.-Y. Tan, N. Cai, J. Zhou, and S.-G. Zhang, "On performance of peer review for academic journals: analysis based on distributed parallel system," *IEEE Access*, vol. 7, pp. 19024–19032, 2019.

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets,," in *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS)*, pp. 2672–2680, Montreal, Canada, December 2014.

[6] L. Wang, J. Xi, M. He, and G. Liu, "Robust time-varying formation design for multiagent systems with disturbances:

extended-state-observer method," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 7, pp. 2796–2808, 2020.

[7] J. Xi, L. Wang, J. Zheng, and X. Yang, "Energy-constraint formation for multiagent systems with switching interaction topologies," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 6, pp. 2442–2454, 2020.

[8] G. Antipov, M. Baccouche, and J. L. Dugelay, "Face aging with conditional generative adversarial networks," in *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP)*, pp. 2089–2093, Beijing, China, September 2017.

[9] R. Huang, S. Zhang, T. Y. Li, and R. He, "Beyond face rotation: global and local perception GAN for photorealistic and identity preserving frontal view synthesis," in *Proceedings of the 16th IEEE International Conference on Computer Vision (ICCV)*, pp. 2458–2467, Venice, Italy, October 2017.

[10] L. Tran, X. Yin, and X. M. Liu, "Disentangled representation learning GAN for pose-invariant face recognition," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1283–1292, Honolulu, HI, USA, July 2017.

[11] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *Proceedings of the 14th European Conference on Computer Vision (ECCV)*, pp. 702–716, Amsterdam, The Netherlands, October 2016.

[12] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. Lempitsky, "Texture networks: feed-forward synthesis of textures and stylized images," in *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, pp. 1349–1357, New York, NY, USA, June 2016.

[13] P. Isola, J. Y. Zhu, T. H. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5976, Honolulu, HI, USA, June 2017.

[14] A. Brock, J. Donahue, and K. Simonyan: Large Scale GAN Training for High Fidelity Natural Image Synthesis," 2018, https://arxiv.org/abs/1809.11096.

[15] J. Donahue and K. Simonyan, "Large scale adversarial representation learning," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NIPS)*, Vancouver, Canada, February 2019.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, May 2015.

[17] C. Ledig, L. Theis, F. Huszår et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 105–114, Honolulu, HI, USA, July 2017.

[18] M. Zhai, L. Chen, F. Tung et al., "Lifelong GAN: continual learning for conditional image generation," in *Proceedings of the 32th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2759–2768, Long Beach, CA, USA, September 2019.

[19] D. He, Y. Xia, T. Qin et al., "Dual learning for machine translation," in *Proceedings of the 29th International Conference on Neural Information Processing Systems (NIPS)*, pp. 820–828, San Juan, Puerto Rico, December 2016.

[20] Z. Yi, H. Zhang, P. Tan, and M. L. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proceedings of the 16th IEEE International Conference on Computer Vision (CVPR)*, pp. 2868–2876, Venice, Italy, July 2017.

[21] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the 16th IEEE International Conference on Computer Vision (CVPR)*, pp. 2223–2232, Venice, Italy, July 2017.

[22] Y. Choi, M. Choi, M. Kim et al., "StarGAN: unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the 31th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8789–8797, Salt Lake City, UT, USA, June 2018.

[23] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the 32th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4401–4410, Long Beach, CA, USA, September 2019.

[24] C. Yang and S.-N. Lim, "One-Shot domain adaptation for face generation," in *Proceedings of the 33th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5920–5929, Seattle, WA, USA, August 2020.

[25] E. Richardson, Y. Alaluf, O. Patashnik et al., Encoding in Style: A StyleGAN Encoder for Image-To-Image Translation," 2020, https://arxiv.org/abs/2008.00951.

[26] Y. T. Chen, X. G. Xu, and J. Y. Jia, "Domain adaptive image-to-image translation," in *Proceedings of the 33th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5273–5282, Seattle, WA, USA, August 2020.

[27] I. J. Goodfellow, "NIPS 2016 tutorial: generative adversarial networks," in *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS)*, San Juan, Puerto Rico, December 2016.

[28] M. Mirza and S. Osindero: Conditional Generative Adversarial Nets," 2014, https://arxiv.org/abs/1411.1784.

[29] L. Zhou, C. Zhang, and M. Wu, "D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, June 2018.

[30] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings f the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[31] M. D. Zeiler, G. W. Taylor, and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in *Proceedings of the 13th IEEE International Conference on Computer Vision (ICCV)*, pp. 2018–2025, Barcelona, Spain, March 2011.

[32] L. Zhang, L. Zhang, X. Q. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 2378–2386, 2011.