

Research Article

Generative Adversarial Network-Based Edge-Preserving Superresolution Reconstruction of Infrared Images

Yuqing Zhao , Guangyuan Fu, Hongqiao Wang, Shaolei Zhang, and Min Yue

Xi'an Research Institute of High-Tech, Shaanxi 710025, China

Correspondence should be addressed to Yuqing Zhao; zoe_rabbi@126.com

Received 7 February 2021; Revised 2 June 2021; Accepted 1 July 2021; Published 21 July 2021

Academic Editor: Quoc-Tuan Vien

Copyright © 2021 Yuqing Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The convolutional neural network has achieved good results in the superresolution reconstruction of single-frame images. However, due to the shortcomings of infrared images such as lack of details, poor contrast, and blurred edges, superresolution reconstruction of infrared images that preserves the edge structure and better visual quality is still challenging. Aiming at the problems of low resolution and unclear edges of infrared images, this work proposes a two-stage generative adversarial network model to reconstruct realistic superresolution images from four times downsampled infrared images. In the first stage of the generative adversarial network, it focuses on recovering the overall contour information of the image to obtain clear image edges; the second stage of the generative adversarial network focuses on recovering the detailed feature information of the image and has a stronger ability to express details. The infrared image superresolution reconstruction method proposed in this work has highly realistic visual effects and good objective quality evaluation results.

1. Introduction

Infrared imaging technology used for passive noncontact detection and identification has the advantages of good concealment, strong transmission ability, no interference from electromagnetic waves, and good low-light and night-vision capabilities [1]. In addition to the main military applications, this technology can also be widely used in civilian fields, such as industry, agriculture, medicine, and public security reconnaissance. However, infrared images have many disadvantages, such as low resolution, low contrast, and blurred edges. Although it is possible to improve the hardware performance of the infrared imaging system by improving the manufacturing process of the infrared detector, it requires tremendous human and financial resources and is difficult to achieve in the short term. Therefore, digital signal processing is an economical and effective way to improve the quality of infrared images [2–4]. Superresolution reconstruction refers to the reconstruction of a high-resolution image or a sequence on single-frame or multiframe low-resolution images and includes three main types, i.e., interpolation-based methods, reconstruction-based methods, and instance-based learning

methods. Instance-based learning methods are flexible in algorithm structure, can provide more details under high magnification, and thus have become a research hotspot of superresolution reconstruction in recent years. Chao et al. [5] proposed the use of a convolutional neural network (CNN) to achieve the superresolution reconstruction of visible light images and to learn the mapping relationship between low-resolution image and high-resolution image by training with a large dataset. Other researchers [6] used the perceptual loss to replace the minimum mean square error and the learned upsampling to replace bicubic interpolation, achieving better results. Subsequently, network structures with more layers, such as the deeply recursive convolutional network (DRCN) [7] and the efficient subpixel convolutional neural network (ESPCN) [8], were proposed to achieve better outcomes. In the field of machine learning, generative models have always been a difficult problem. The proposal of a generative adversarial network (GAN) [9] meets the demands of generative models for research and application in many fields. GAN only uses back propagation and thus avoids the complex Markov chain. At the same time, GAN uses unsupervised learning methods, enabling the production of more clear and true



FIGURE 1: Original image (left) and the image recovered from 4x downsampled images using the proposed method.

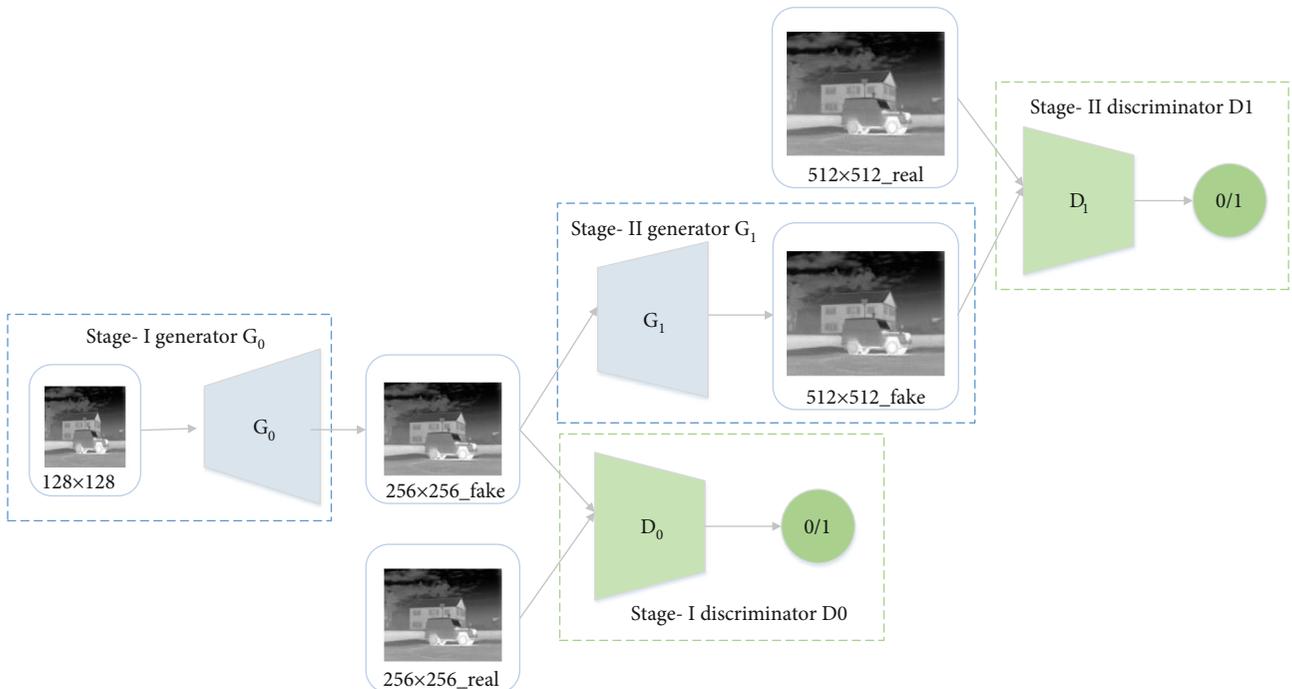
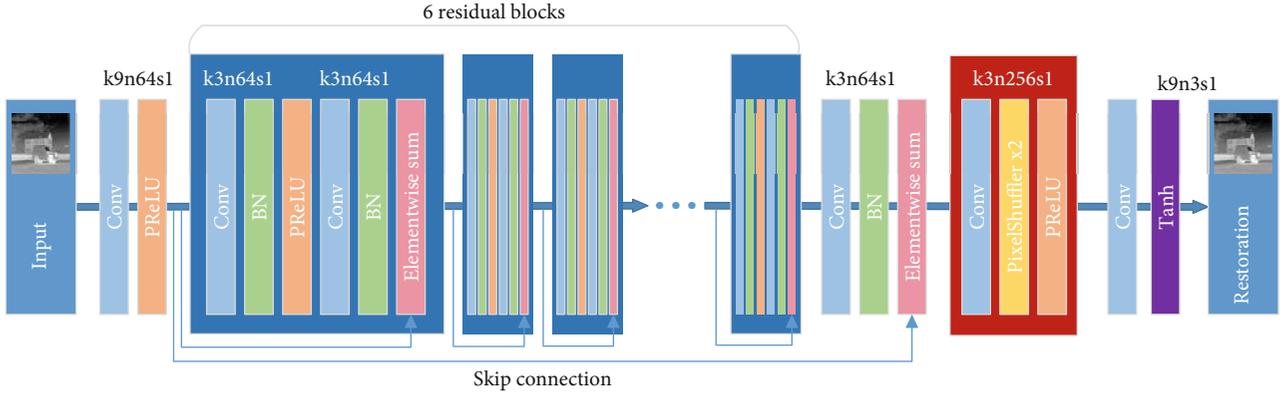
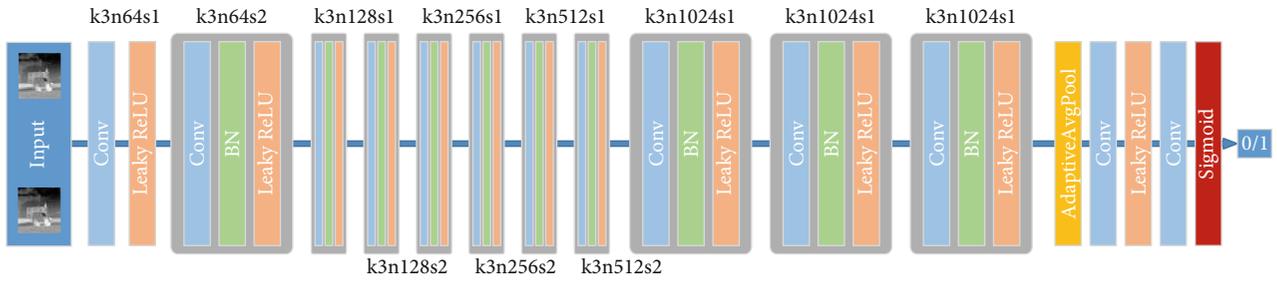
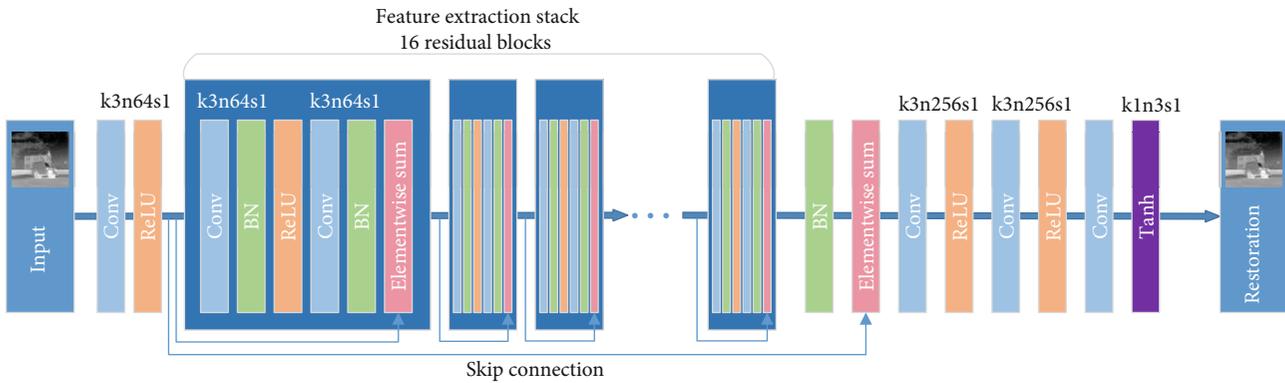


FIGURE 2: Model architecture of the proposed method.

samples. In recent research, GAN has been widely used [10–12]. Among them, the SaliencyGAN model proposed by Wang et al. [10] is a semisupervised salient object detection method for the Internet of Things. Yang et al. [12] proposed a new model based on conditional generative adversarial network (DAGAN) to reconstruct Compressed Sensing Magnetic Resonance Imaging (CS-MRI). Literature [11] proposed a novel fast CS-MRI deep learning architecture based on a conditional generative confrontation network. SRGAN [13] is a GAN for image superresolution (SR) and can recover a photorealistic natural image from four instances of downsampling. However, the amplified details of the

image generated through this method usually show unsightly artifacts. To further improve the visual quality, researchers [14] have proposed a residual-in-residual dense block (RRDB) network unit and made improvements in the perception domain loss. Researchers [15] have also proposed a novel generative adversarial frame to improve the edge structure and texture information in the compressed image. ESRGAN+ [16] designed a network architecture with novel basic blocks to replace the basic structure used by the original ESRGAN. In the field of infrared imaging, SR reconstruction has been mostly achieved using sparse coding methods [17–21].

FIGURE 3: Network structure of the generator G_0 .FIGURE 4: Network structure of the discriminator D_0 .FIGURE 5: Network structure of the generator G_1 .

In this study, we propose a new GAN framework to improve the perceived quality of infrared images. Then, we design a multiconstraint loss function by combining image fidelity loss, adversarial loss, feature fidelity loss, and edge fidelity loss. By continuously updating and iterating the minimization of loss functions, we obtain the reconstructed image with high resolution and sharp edges using the proposed method (Figure 1).

The main contributions of this paper are as follows:

- (i) We propose a GAN-based SR reconstruction framework for edge preservation of infrared images that can enhance the GAN to better restore the edge

structure of the infrared image while maintaining the detailed information

- (ii) To preserve the characteristics and the edge information in the image, we propose a multiple constraints loss function applicable for SR reconstruction
- (iii) We validate the proposed method using images from publicly available datasets and compare the performance of the proposed method with that from other mainstream methods. The results confirm that compared with other methods, the proposed method obtains more realistic reconstructed infrared SR images with sharper edges

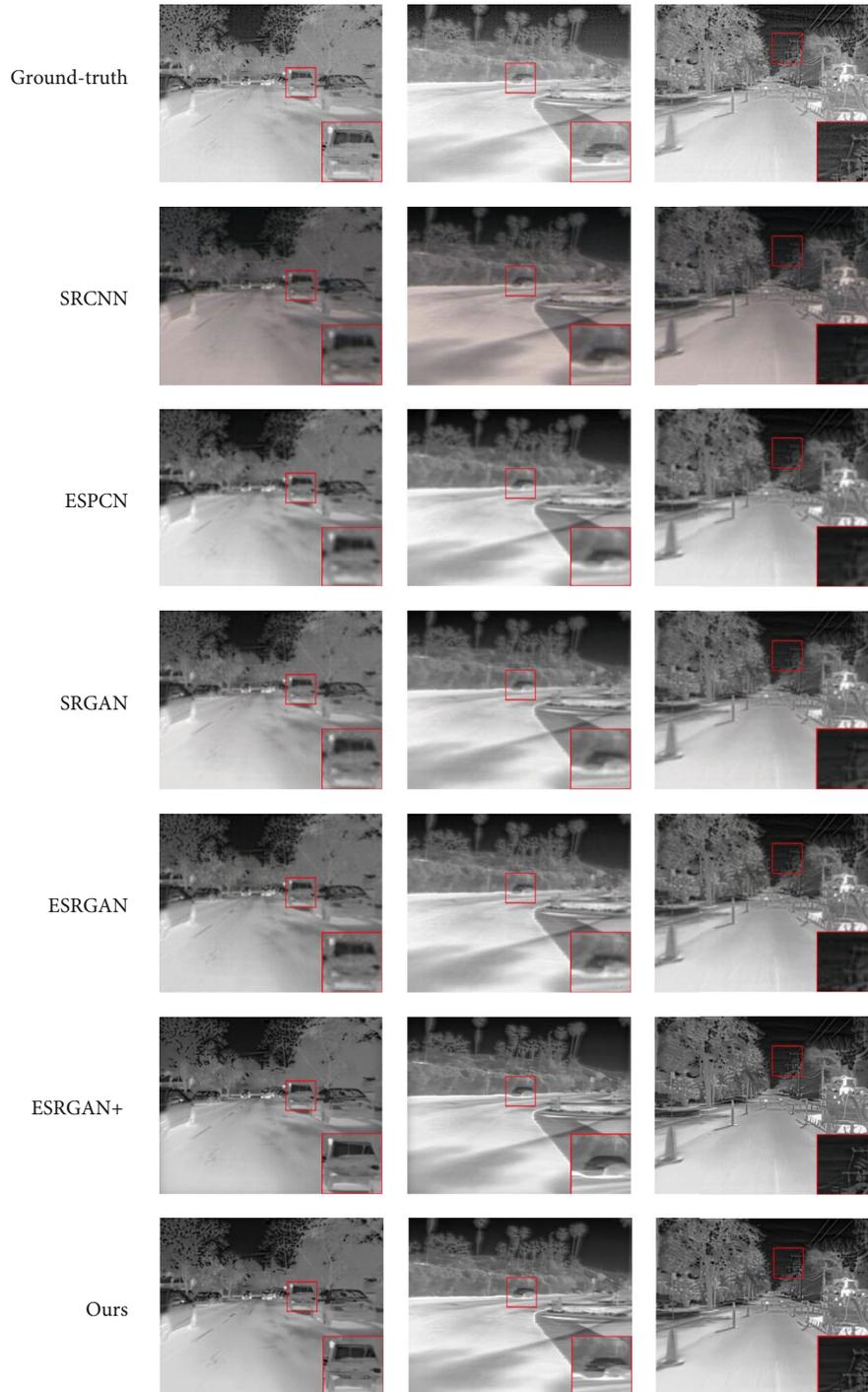


FIGURE 6: Comparison of the reconstructed results of images from the FLIR_ADAS_1_3 validation set. The images in the first row are original images, those in the second row are the reconstruction results using the SRCNN method, those in the third row are the reconstruction results using the ESPCN method, those in the fourth row are the reconstruction results using the SRGAN method, those in the fifth row are the reconstruction results using the ESRGAN method, those in the sixth row are the reconstruction results using the ESRGAN+ method, and those in the last row are the reconstruction results using our proposed method.

We describe the proposed network framework and loss function in Section 2 and their quantitative evaluations using the public datasets in Section 3, followed by recommended future study on the proposed method and the conclusion.

2. Method

To generate a high-resolution image with photograph-grade realistic details and inspired by the literature [22, 23], we propose a simple and effective two-layer GAN, in

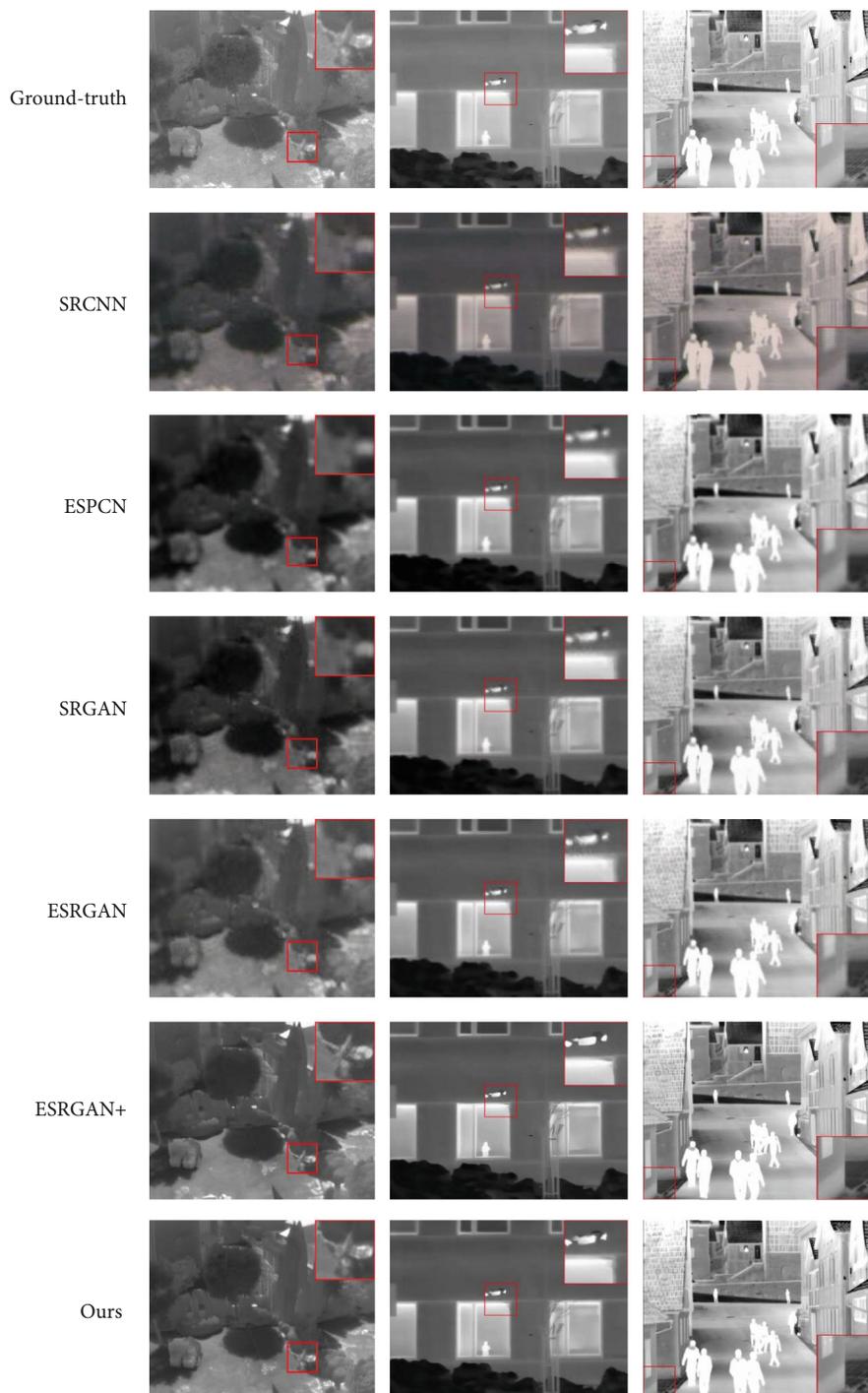


FIGURE 7: Comparison of the reconstructed results of images from the Itir_v1_0 validation set. The images in the first row are original images, those in the second row are the reconstruction results using the SRCNN method, those in the third row are the reconstruction results using the ESPCN method, those in the fourth row are the reconstruction results using the SRGAN method, those in the fifth row are the reconstruction results using the ESRGAN method, those in the sixth row are the reconstruction results using the ESRGAN+ method, and those in the last row are the reconstruction results using our proposed method.

which the image generation process is divided into two stages (Figure 2).

2.1. Stage 1 GAN. In Stage 1, we use the 128×128 low-resolution image I^{LR0} as input and export it to the Stage 1

generator G_0 to generate a false 256×256 image (I^{SR0}), which, together with a real 256×256 image (I^{HR0}), is imported to the Stage 1 discriminator D_0 to identify the image. The core of the network structure of the generator G_0 in the proposed method is shown in Figure 3. We adopt

the network design of Ledig et al. [10] and introduce the skip connection, which has been proven effective in training deep neural networks. We adopt the residual block proposed in the literature [24] to construct a neural network with six residual blocks used as a stack to extract features from the image. Recent work on single-image superresolution [25–28] pointed out that with the deepening of the network and the training and testing phases under the GAN framework, hallucination artifacts often appear. In order to solve this problem, we are following the method in [29] and adding a batch normalization layer (BN) to the network. Each residual block contains two convolutional layers with a kernel size of 3×3 and 64 feature maps, two BN layers, and one parametric rectified linear unit (PReLU) layer [30]. The specific settings of each layer of the generative model are as follows:

$$\begin{aligned} C(PR, 64) &\longrightarrow C(64)BN(PR)C(64)BN(PR)SC \longrightarrow \dots 6 \dots \\ &\longrightarrow C(64)BN(PR)SC \longrightarrow C(PR, 256) \longrightarrow C(t, 3). \end{aligned} \quad (1)$$

Here, $C(PR, 64)$ denotes a set of convolutional layers with 64 feature maps and activation function PReLU; $C(64)BN(PR)C(64)BN(PR)SC$ represents a residual block; $BN(PR)$ is a batch normalization layer with activation function PReLU, and SC denotes a skip connection. There are in total 6 residual blocks. $C(t, 3)$ represents a convolutional layer with 3 feature maps and activation function tanh.

To distinguish the generated SR samples from real high-resolution (HR) samples, we train the discriminator network D_0 , whose overall framework is shown in Figure 4 and which is adopted from the architectural framework summarized by Radford et al. [31], in which LeakyReLU [32] activation is used to avoid maximum pooling of the entire network. The discriminator network includes ten convolutional blocks. All blocks except for the first block contain a convolutional layer, a BN layer, and a LeakyReLU layer. The number of kernels within the filter increases continually, from 64 in the Visual Geometry Group (VGG) network [33] to 1024; then, each time the number of kernels increases, segmented convolution is used to lower the image resolution. After that, a special residual block, containing two convolution layers and a LeakyReLU layer, is connected. The output of the last convolution unit is sent to a dense layer with an S-type activation function to get a true and false result. The structure and parameters of each layer of the discriminator D_0 network are as follows:

$$\begin{aligned} C(LR, 64) &\longrightarrow C(64)BN(LR) \longrightarrow C(128)BN(LR) \\ &\longrightarrow C(256)BN(LR) \longrightarrow C(512)BN(LR) \\ &\longrightarrow C(1024)BN(LR) \longrightarrow C(LR, 1024) \longrightarrow D. \end{aligned} \quad (2)$$

Here, LR denotes the activation function LeakyReLU; $C(LR, 64)$ denotes a set of convolutional layers with 64 feature maps and activation function LeakyReLU; $C(64)BN(LR)$ denotes a set of convolutional layers with 64 feature maps followed by batch-normalization with activation function

TABLE 1: Method comparison: quantitative results of the SRCNN, ESPCN, SRGAN, ESRGAN, and ESRGAN+ methods and the proposed method on the FLIR_ADAS_1_3 validation dataset.

	CC	PSNR (dB)	SSIM	VIF	UIQI	TIME (s)
SRCNN	0.8160	18.3963	0.8655	0.7942	0.8770	2.7112
ESPCN	0.8627	27.7171	0.9349	0.8269	0.9398	1.0206
SRGAN	0.8463	28.0445	0.9497	0.8120	0.8503	0.9202
ESRGAN	0.9547	28.5453	0.9506	0.7918	0.9382	0.1093
ESRGAN+	0.9806	33.7018	0.9855	0.9119	0.9864	0.2713
Ours	0.9964	37.9865	0.9980	0.9083	0.9965	0.1145

TABLE 2: Method comparison: quantitative results of the SRCNN, ESPCN, SRGAN, ESRGAN, and ESRGAN+ methods and the proposed method on the Itir_v1_0 dataset.

	CC	PSNR (dB)	SSIM	VIF	UIQI	TIME (s)
SRCNN	0.7937	20.2460	0.8966	0.7079	0.8746	2.5706
ESPCN	0.8464	22.1708	0.9185	0.8194	0.9877	1.0020
SRGAN	0.8152	22.6732	0.9154	0.7944	0.8610	0.6642
ESRGAN	0.9146	30.3205	0.9608	0.7752	0.9166	0.0937
ESRGAN+	0.9547	34.1024	0.9844	0.8497	0.9541	0.2558
Ours	0.9962	35.2865	0.9943	0.9064	0.9936	0.1059

LeakyReLU, and D is the dense layer for outputting. The feature maps were increased from 64 to 1024.

$$\mathcal{L}_1 = \lambda_1 \mathcal{L}_{\text{adv}} + \lambda_2 \mathcal{L}_{\text{mse}} + \lambda_3 \mathcal{L}_{\text{edge}}, \quad (3)$$

Stage 1. Reconstruction Loss Function. The loss function contains three parts: adversarial loss, image fidelity loss, and edge fidelity loss (Eq. (3)). These parts each capture different perceptual characteristics of the reconstructed image to obtain a more visually satisfactory reconstructed image.

where the weight $\{\lambda_i\}$ is a trade-off parameter that is used to balance multiple loss components. The first part is the adversarial loss between the generator G_0 and the discriminator D_0 of GAN. This part encourages the generator to trick the discriminator network to produce a more realistic HR image as follows:

$$\mathcal{L}_{\text{adv}} = \sum_{n=1}^N -\log D_{\theta_D}(G_{\theta_G}(I^{\text{LR}0})), \quad (4)$$

where $D_{\theta_D}(G_{\theta_G}(I^{\text{LR}0}))$ is the estimated probability of the reconstructed image $G_{\theta_G}(I^{\text{LR}0})$ to be a true HR image. To obtain a better gradient, we use the minimization $-\log D_{\theta_D}(G_{\theta_G}(I^{\text{LR}0}))$ to replace the minimization $\log [1 - D_{\theta_D}(G_{\theta_G}(I^{\text{LR}0}))]$.

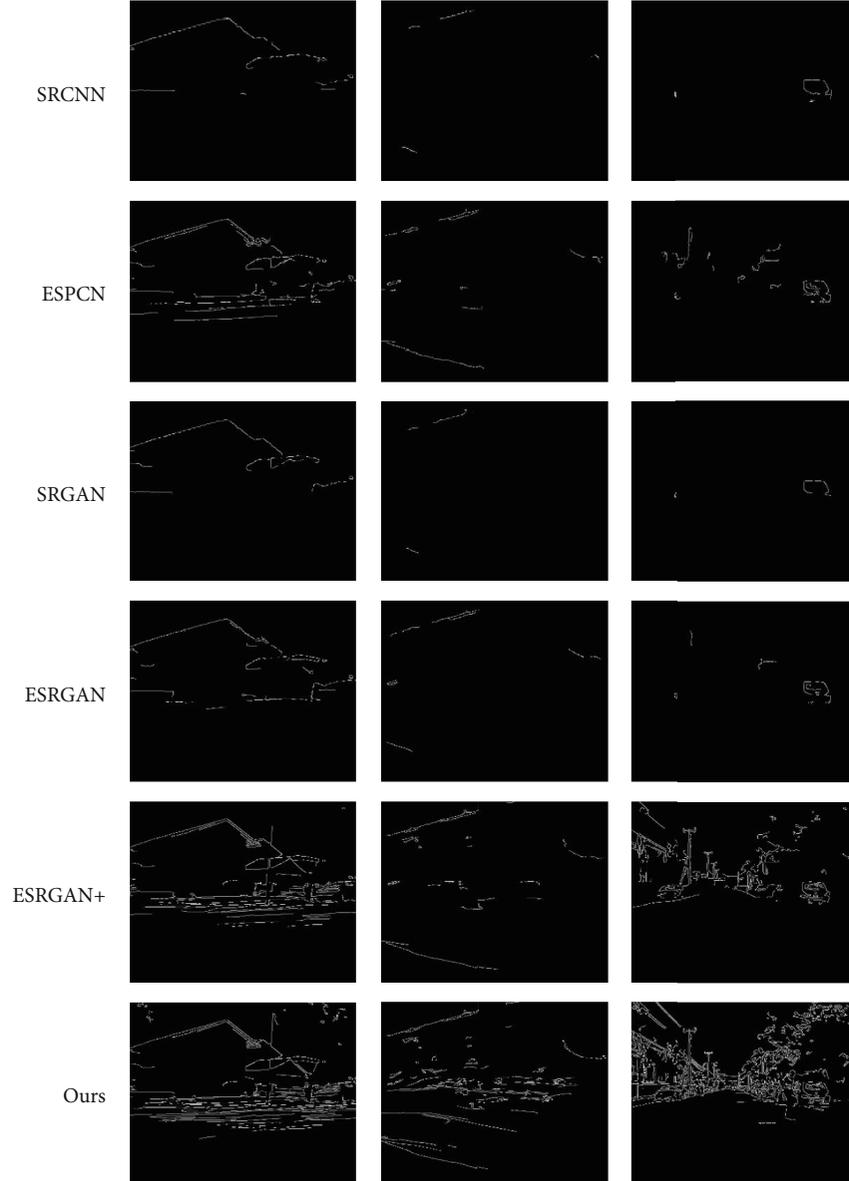


FIGURE 8: Comparison of edge detection results of superresolution reconstruction results.

The second term of Eq. (3), \mathcal{L}_{mse} , ensures the fidelity of the restored image using the pixel-level mean square error (MSE) loss as follows:

$$\frac{\mathcal{L}_{\text{mse}} = 1}{WHC \|I^{\text{HR}0} - I^{\text{SR}0}\|_2^2}, \quad (5)$$

where W , H , and C are the height, the width, and the number of channels, respectively, of the image.

The third term of Eq. (3), $\mathcal{L}_{\text{edge}}$, the edge fidelity loss, is purported to reproduce sharp edge information as follows:

$$\frac{\mathcal{L}_{\text{edge}} = 1}{WH \|I^{\wedge E} - I^E\|_2^2}, \quad (6)$$

where W and H are the width and the height, respectively, of the image. The labeled edge map I^E is extracted by a specific edge filter on the real 256×256 image $I^{\text{HR}0}$, while $I^{\wedge E}$ is extracted by a specific edge filter on the 256×256 image $I^{\text{SR}0}$ generated by the generator G_0 . In our experiments, we chose the Canny edge detection operator. By minimizing the edge fidelity loss, the network continuously guides edge recovery.

2.2. Stage 2 GAN. In the second stage, we use the generated 256×256 low-resolution image $I^{\text{SR}0}$ as input and export it to the Stage 2 generator G_1 to generate a 512×512 image $I^{\text{SR}1}$, which, together with a real 512×512 image $I^{\text{HR}1}$, is exported to the Stage 2 discriminator D_1 to identify if it is false or real. We adopt the network design described in the literature [15], in which a generator model G_1 (Figure 5)

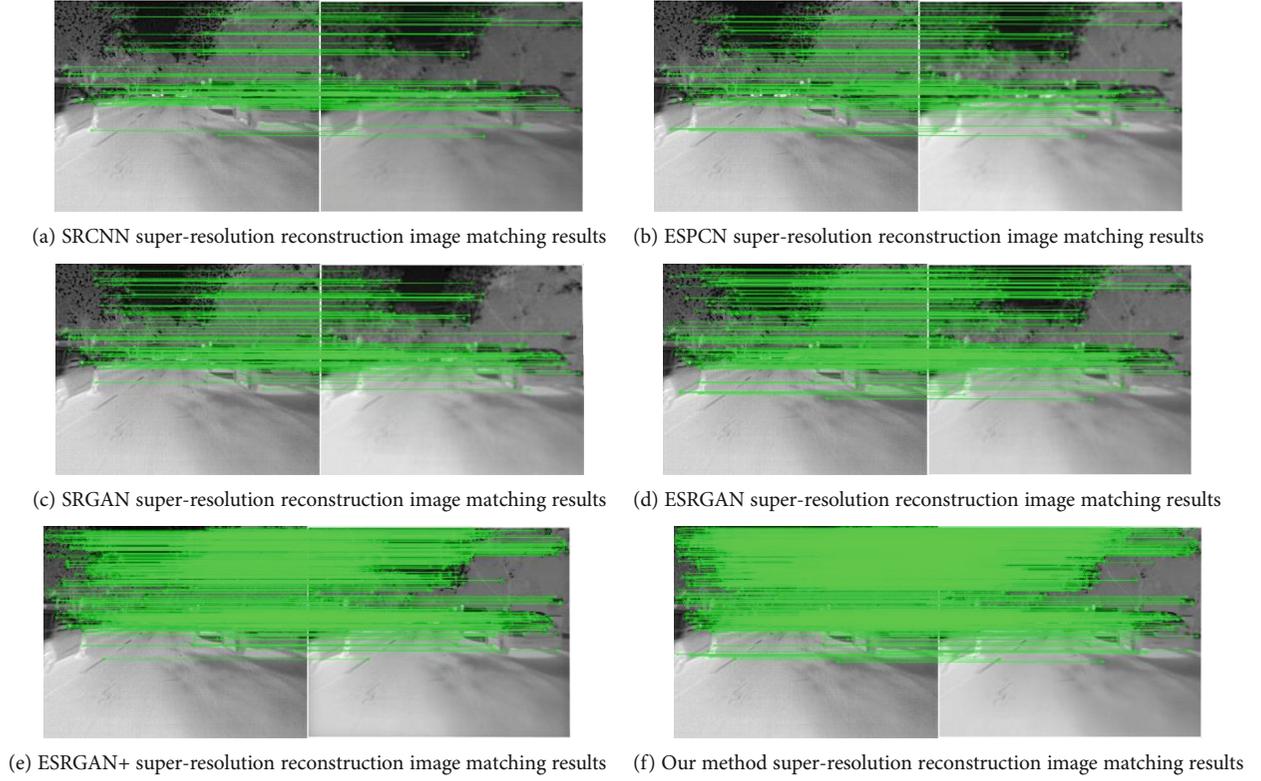


FIGURE 9: Superresolution reconstruction image matching results (the left image is the original high-resolution image).

containing 16 residual blocks is constructed. The output format and parameters of each layer of the generator G_1 are as follows:

$$\begin{aligned} C(R, 64) &\longrightarrow C(64)BN(R)C(64)BN(R)SC \longrightarrow \dots 16 \dots \\ &\longrightarrow C(R, 64)SC \longrightarrow C(256) \longrightarrow C(t, 3), \end{aligned} \quad (7)$$

where $C(R,64)$ denotes a set of convolutional layers with 64 feature maps and activation function ReLU; $C(64)BN(R)C(64)BN(R)SC$ represents a residual block.

The network structure of the discriminator network D_1 in the second stage adopts a network structure similar to that of the discriminator D_0 . Each layer structure and network parameters of D_1 are as follows:

$$\begin{aligned} C(LR, 64) &\longrightarrow C(64)BN(LR) \longrightarrow C(128)BN(LR) \\ &\longrightarrow C(256)BN(LR) \longrightarrow C(512)BN(LR) \\ &\longrightarrow C(1024)BN(LR) \longrightarrow C(LR, 1024) \longrightarrow D. \end{aligned} \quad (8)$$

$$\mathcal{L}_2 = \lambda'_1 \mathcal{L}_{adv1} + \lambda'_2 \mathcal{L}_{mse1} + \lambda'_3 \mathcal{L}_{feature}, \quad (9)$$

where the weight $\{\lambda'_i\}$ is a trade-off parameter that is used to balance multiple loss components. The first term, \mathcal{L}_{adv1} , is the adversarial loss between the generator G_1 and the discriminator D_1 of GAN. The second term, \mathcal{L}_{mse1} , is the image fidelity loss. The third term, $\mathcal{L}_{feature}$, is the feature fidelity

loss, which is defined based on the feature space distance defined in the literature [34], to facilitate the preservation of feature representation in the reconstructed image similar to that of the original image as follows:

$$\frac{\mathcal{L}_{feature} = 1}{WHC \|\phi(I^{HR0} - I^{SR0})\|_2^2}, \quad (10)$$

where W , H , and C are the height, the width, and the number of channels, respectively, of the image and $\phi(\cdot)$ represents the feature space function, which is a pretrained VGG-19 [33] network that maps images to feature space. The fourth pooling layer is used to calculate the feature-activated L2 distance to be used as the feature fidelity loss function.

Stage 2. Reconstruction Loss Function. The loss function contains three parts: adversarial loss, image fidelity loss, and feature fidelity loss (Eq. (9)).

3. Experiments

3.1. Experimental Details. All experiments are performed on a desktop computer with 2.20 GHz $\times 40$ Intel Xeon (R) Silver 4114 CPU, GeForce GTX 1080Ti, and 64 GB of memory. We use the PyTorch deep learning framework to implement the method proposed in this article and all comparison methods. All methods are trained with the FLIR training data set. During the training process, we choose 8,862 images from the FLIR_ADAS_1_3 thermal sensor training dataset released

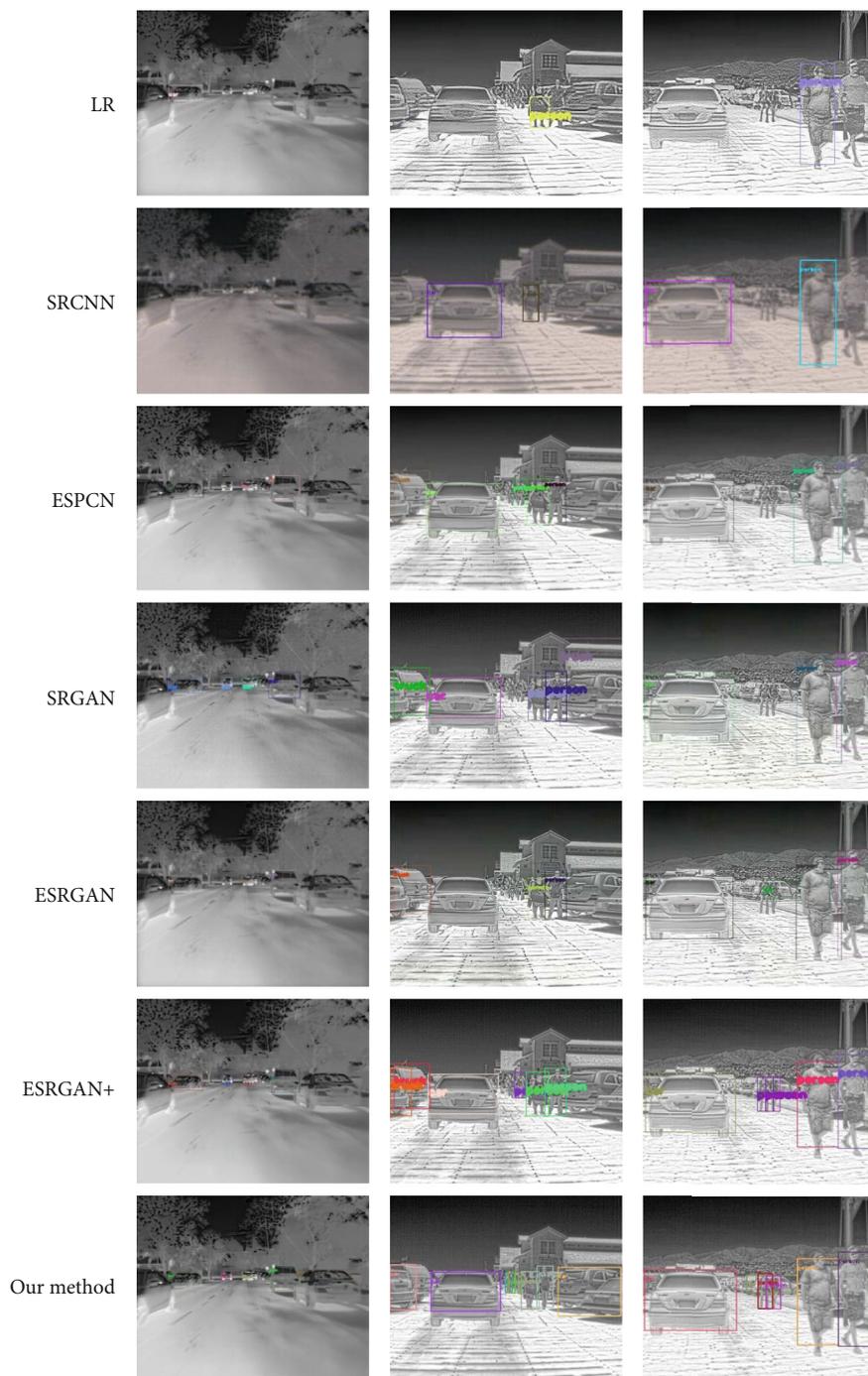


FIGURE 10: Target detection result of superresolution reconstructed image.

by FLIR Systems Inc., a sensor system developer, in 2018. First, we perform the 4x factor downsampling on all experimental data to obtain low-resolution (LR) images by lowering the resolution of the HR images. We set the batch size to 4 and used the Adam [35] with a momentum term of $\beta = 0.9$ as the optimization procedure. To enable the loss function in the same order of magnitude to better balance loss components, we set λ_1, λ_2' , and λ_3' of Eq. (9) to 10^{-3} , 1, and 10^{-6} , respectively. When training the Stage 1 GAN, we set

the learning rate to 10^{-4} , which is lowered to 10^{-5} when training the Stage 2 GAN.

3.2. Experimental Evaluation. To verify the effectiveness of the proposed method, we conduct the validation on two public data sets: the FLIR_ADAS_1_3 (1366 images) validation set and the Itir_v1_0 dataset (11262 images). We compared the proposed method with the most advanced methods, including superresolution using deep convolutional networks (SRCNN)



FIGURE 11: Images with motion blur cannot obtain clear edge information using the ordinary SR reconstruction methods.

[5], ESPCN [8], SRGAN [10], ESRGAN [15], and ESRGAN+ [16] methods. Three images are selected from the verification set of FLIR_ADAS_1_3 and the data set of Itir_v1_0, and the subjective results of several methods are shown in Figures 6 and 7. From the reconstruction results, it is not difficult to see that the reconstruction results of our proposed method produce finer texture and edge details.

To facilitate a fair quantitative comparison, we used the correlation coefficient (CC) [36], the peak-signal-to-noise ratio (PSNR) [37], the Structural Similarity Index Measure (SSIM) [38], the visual information fidelity (VIF) [39], the Universal Image Quality Index (UIQI), and the time consumption, six objective indicators, to evaluate the quality of the reconstructed images and the SR methods. The quantitative results of the comparison of different reconstruction methods are shown in Table 1, which shows that our method is superior to the SRCNN [5], ESPCN [8], SRGAN [13], ESRGAN [15], and ESRGAN+ [16] methods in the indicators of CC, PSNR, SSIM, and UIQI on the FLIR data set. The VIF index is slightly lower than the ESRGAN+ method, and the time consumption is greater than the ESRGAN method.

The quantitative results of comparison of different reconstruction methods on the Itir_v1_0 dataset are shown in Table 2, indicating that the proposed method is superior to the SRCNN [5], ESPCN [8], SRGAN [13], ESRGAN [15], and ESRGAN+ [16] methods on the Itir_v1_0 dataset. Only the time consumption is slightly greater than the ESRGAN method.

In order to more intuitively illustrate the effectiveness of the superresolution method proposed in this work for improving the edge features of infrared images, we show in Figure 8 the comparison of the edge detection results of the superresolution reconstruction results of various methods. It can be seen from the figure that the image reconstructed by our method has more and clearer edge information, which is meaningful for the application of infrared images.

3.3. Use Advanced Vision Tasks to Compare Superresolution Results. Basic vision tasks including image superresolution reconstruction are all for advanced vision tasks. Infrared images are widely used in target detection and target matching tasks, but the shortcomings of low resolution and unclear edges of infrared images affect the accuracy of the above tasks. Therefore, whether the result of infrared image superresolution reconstruction can improve the accuracy of the above tasks has become an evaluation index of the result of superresolution reconstruction. In order to further verify our method, we match the superresolution images generated by several methods with real high-resolution images. Scale

Invariant Feature Transform (SIFT) is a representation of the statistical results of Gaussian image gradients in the field of feature points and is a commonly used image local feature extraction algorithm. In the matching result, the number of matching points can be used as a criterion for matching quality, and the corresponding matching points can also determine the similarity of the local features of the two images. Figure 9 shows the result of matching the superresolution reconstructed image with the original high-resolution image through the SIFT algorithm. It can be seen from the quantity that the reconstructed image produced by our proposed method obtains more correct matching pairs than other methods.

In this experiment, we use the classic YOLO [40] method for image target detection (Figure 10). It can be seen that the superresolution reconstructed image generated by our proposed method has better detection results and can detect more targets.

4. Discussion and Future Work

Through experiments, we demonstrate that compared with other methods, the proposed method has a better perceptual performance. However, during the experiment, we also found that for some images, the reconstruction results are not satisfactory, as shown in Figure 11. By analyzing these images, we found that they have common characteristics, i.e., when the imaging device and the imaging object are moving at high relative speeds, the captured image may contain motion blur. For such images, ordinary SR reconstruction methods cannot achieve effective edge recovery. Therefore, in future studies, we will address these problems.

5. Conclusion

In this study, we propose a two-stage GAN framework that is able to reconstruct SR image by recovering edge structure information and retaining feature information. In the first stage, the image fidelity loss, the adversarial loss, and the edge fidelity loss are combined to preserve the edges of the image. In the second stage, the image adversarial loss, the image fidelity loss, and the feature fidelity loss are combined to mine image visual features. By iteratively updating the generative network and the discriminator network, the SR reconstruction of infrared image with preserved edges is achieved. Experimental verification results show that the proposed method outperforms several other image reconstruction methods in reconstructing SR infrared images.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

- [1] L. Xiang-di et al., "Infrared imaging system and applications," *Laser & Infrared*, vol. 3, 2014.
- [2] X. Sui, Q. Chen, G. Gu, and X. Shen, "Infrared super-resolution imaging based on compressed sensing," *Infrared Physics & Technology*, vol. 63, pp. 119–124, 2014.
- [3] X. Sui, H. Gao, Y. Sun, Q. Chen, and G. Gu, "Infrared super-resolution imaging method based on retina micro-motion," *Infrared Physics & Technology*, vol. 60, pp. 340–345, 2013.
- [4] J. L. Suo, J. I. Xiangyang, and Q. H. Dai, "An overview of computational photography," *Science China(Information Sciences)*, vol. 6, pp. 5–24, 2012.
- [5] D. Chao, C. L. Chen, K. He, and X. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," in *European Conference on Computer Vision*, Springer, 2014.
- [6] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2014.
- [7] J. Kim, J. Lee, and K. Lee, *Deeply-Recursive Convolutional Network for Image Super-Resolution*, 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 1637–1645, 2016.
- [8] W. Shi, J. Caballero, F. Huszar et al., "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016.
- [9] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative Adversarial Networks," *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.
- [10] C. Wang, S. Dong, X. Zhao, G. Papanastasiou, H. Zhang, and G. Yang, "SaliencyGAN: deep learning semisupervised salient object detection in the fog of IoT," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2667–2676, 2020.
- [11] G. Yang, S. Yu, H. Dong et al., "DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1310–1321, 2018.
- [12] S. Yu, D. Hao, G. Yang, G. Slabaugh, and Y. Guo, "Deep De-aliasing for Fast Compressive Sensing MRI," 2017, <https://arxiv.org/abs/1705.07137>.
- [13] C. Ledig, L. Theis, F. Huszar et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2016.
- [14] X. Wang, K. Yu, S. Wu et al., "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," in *Lecture Notes in Computer Science*, Springer, 2018.
- [15] Q. Mao, S. Wang, S. Wang, X. Zhang, and S. Ma, "Enhanced Image Decoding Via Edge-Preserving Generative Adversarial Networks," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*, San Diego, CA, USA, 2018.
- [16] N. C. Rakotonirina and A. Rasoanaivo, "ESRGAN+: further improving enhanced super-resolution generative adversarial network," in *ICASSP 2020 - 2020 IEEE international conference on acoustics, Speech and Signal Processing (ICASSP)*, pp. 3637–3641, Barcelona, Spain, 2020.
- [17] C. Kraichan and S. Pumrin, "Performance Analysis on Multi-Frame Image Super-Resolution Via Sparse Representation," in *2014 International Electrical Engineering Congress (iEECON)*, Chonburi, Thailand, 2014.
- [18] S. Yu-bao, W. Zhi-hui, X. Liang, Z. Zhen-rong, and L. Zhan-Qiang, "Multimorphology Sparsity regularized image super-resolution," *Acta Electronica Sinica*, vol. 38, no. 12, pp. 2898–2903, 2010.
- [19] S. Yang, M. Wang, Y. Chen, and Y. Sun, "Single-image super-resolution reconstruction via learned geometric dictionaries and clustered sparse coding," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 4016–4028, 2012.
- [20] Y. Tang, Y. Yuan, P. Yan, and X. Li, "Greedy regression in sparse coding space for single-image super-resolution," *Journal of Visual Communication & Image Representation*, vol. 24, no. 2, pp. 148–159, 2013.
- [21] L. Qiu-sheng and Z. Wei, "Image super-resolution algorithms based on sparse representation of classified image patches," *Acta Electronica Sinica*, vol. 40, no. 5, pp. 920–925, 2012.
- [22] H. Zhang, T. Xu, H. Li et al., "StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017.
- [23] H. Zhang, T. Xu, H. Li et al., "StackGAN++: realistic image synthesis with stacked generative adversarial networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, pp. 1947–1962, 2018.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.
- [25] J. Zhu, G. Yang, P. Ferreira et al., "A Roi Focused Multi-Scale Super-Resolution Method for the Diffusion Tensor Cardiac Magnetic Resonance," in *Proceedings of the 27th Annual Meeting (ISMRM)*, pp. 11–16, Montréal, QC, Canada, 2019.
- [26] J. Zhu, G. Yang, and P. Lio, "Lesion Focused Super Resolution," in *Medical Imaging: Image Processing 2019*, San Diego, CA, USA, 2019.
- [27] J. Zhu, G. Yang, and P. Lio, "How Can We Make GAN Perform Better in Single Medical Image Super-Resolution? A Lesion Focused Multi-Scale Approach," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy, 2019.
- [28] M. Seitzer, G. Yang, J. Schlemper et al., "Adversarial and Perceptual Refinement for Compressed Sensing MRI Reconstruction," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Springer, 2018.
- [29] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *International conference on machine learning*, PMLR, 2015.

- [30] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, 2012.
- [31] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," 2015, <https://arxiv.org/abs/1511.06434>.
- [32] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of the 30 th International Conference on Machine Learning*, p. 3, Atlanta, Georgia, USA, 2013.
- [33] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *Computer Science*, 2014, <https://arxiv.org/abs/1409.1556>.
- [34] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," in *Computer Vision – ECCV 2016*, Springer, 2016.
- [35] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *Computer Science*, 2014, <https://arxiv.org/abs/1412.6980>.
- [36] M. Deshmukh and U. Bhosale, "Image fusion and image quality assessment of fused images," *International Journal of Image Processing (IJIP)*, vol. 4, no. 5, p. 484, 2010.
- [37] C. Yim and A. C. Bovik, "Quality assessment of deblocked images," *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 88–98, 2011.
- [38] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [39] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Information fusion*, vol. 14, no. 2, pp. 127–135, 2013.
- [40] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2015.