Journal of Computers Vol. 32 No. 3, 2021, pp. 249-262
doi:10.3966/199115992021063203018

Generative Adversarial Network in Wavelet Domain for Single Image Super-resolution

Fan Zhang[1,2], Xinwei Wang[1,2], Lin Cao[1,2*], Kangning Du[1,2], Yanan Guo[1,2]

[1] Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument, Beijing Information Science and Technology University, Beijing 100101, China

[2] School of Information and Communication Engineering, Beijing Information Science and Technology University, Beijing 100101, China
zhangfan2015@bupt.cn, {wangxinwei1414, CharLin26}@163.com, kangningdu@outlook.com, yananguo@bistu.edu.cn

Received 17 April 2021; Revised 1 May 2021; Accepted 17 May 2021

Abstract. Nowadays, generative adversarial network for single image super-resolution has achieved superior performance. However, there are few studies on large scaling factor, and the reconstruction performance is relatively poor. Therefore, this paper tackles the above problem by proposing a generative adversarial network in wavelet domain to reconstruct high-resolution image. Specifically, the generator takes low-resolution image as input, and generates wavelet packet decomposition coefficients corresponding to the reconstructed high-resolution image. Then, the discriminator constructs adversarial loss to constrain the generation process of the generator in the wavelet domain. Finally, the high-resolution image is reconstructed based on the inverse wavelet packet transform. The experimental results on CelebA dataset show that our proposed method can achieve better performance than that of comparison methods. The minimum improvements of PSNR and SSIM are 1.783dB and 0.013 on scaling factor ×4, and 0.685dB and 0.016 on scaling factor ×8 when compared with the comparison methods.

Keywords: single image super-resolution, generative adversarial network, wavelet packet transform, deep learning

1 Introduction

As a research hotspot within the field of image processing, single image super-resolution (SISR) is a digital image processing task that reconstructs low-resolution (LR) images into high-resolution (HR) images. Nowadays, with the development of deep learning, image super-resolution algorithms based on deep learning have significantly improved the quality of reconstructed images. Among them, generative adversarial network (GAN) [1], as an algorithm of deep learning, has been used for SISR tasks, such as SRGAN [2], ESRGAN [3], etc. The above methods based on GAN focus on the scaling factor of 4. However, little research has been conducted on SISR task with large scaling factor of 8, and the final reconstruction performance is relatively poor.

In order to solve the above-mentioned problem, this paper proposes a GAN in wavelet domain to perform SISR for facial images, with the advantages of strong visuality and efficiency of wavelet transform in dealing with multi-resolution images. In particular, the LR image is first taken as input, and then feature extraction module of the generator extracts the corresponding features of the LR image, and generates the wavelet packet decomposition coefficients based on the extracted features. Finally, inverse wavelet packet transform is performed on the generated wavelet packet decomposition coefficients to reconstruct the HR image in the wavelet domain. Furthermore, the mapping relationship of generator is constrained by adversarial loss constructed by discriminator in wavelet domain, which minimizes the

* Corresponding Author

distance between the generated wavelet packet decomposition coefficients and the wavelet packet decomposition coefficients of ground-truth (HR image in dataset).To sum up, the proposed method transforms the SISR task from image domain to wavelet domain, and constructs a super-resolution generation adversarial network for the SISR task in wavelet domain, which makes full use of the advantages of wavelet transform to capture low-frequency information and finally realize the face super-resolution reconstruction. Experimental results show that the quality of HR image reconstructed by the proposed method is better, and the details are fully restored.

The rest of this paper is arranged as follows: Section 2 is related work, and Section 3 introduces the network structure and the loss function in detail. Section 4 describes the dataset, experimental setup, experimental results and analysis. Finally, Section 5 concludes the paper.

2 Related Work

Single image super-resolution task can be divided into three categories: the first one is the interpolation-based method, which is widely used within the field of image processing because of the simplicity and real-time performance. However, the HR image reconstructed by interpolation-based method may have jaggies at different levels, resulting in poor reconstruction performance. The second one is the reconstruction-based method. Although reconstruction-based method can reconstruct images with high quality, there exist problems such as high computation complexity and loss of high-frequency details. The third one is the learning-based method, which performs super-resolution reconstruction by learning the mapping relationship between LR images and the corresponding HR images.

Among the learning-based methods, deep learning has been widely applied because of the strong nonlinear fitting and feature extraction capability. C. Dong et al. [4] firstly learned the mapping between the LR image and the corresponding HR image through convolution neural network (CNN), and then HR image was reconstructed based on the mapping. J. Kim et al. [5] proposed VDSR, where the number of network layers was 20. DRRN [6] introduced recurrent neural network (RNN) [7] into the SISR task, where the network depth was increased to 52 because of the shared network parameters. T. Tong et al. proposed SRDenseNet [8], which combined the network structure of DenseNet [9]. B. Lim et al. [10] proposed an enhanced network EDSR with 65 layers. RDN [11] combined ResNet [12] and DenseNet [9] to propose a network with 149 layers, which greatly improved the performance. For the SISR task, wavelet-based methods [13-15] were proposed. Gao et al. [16] proposed a hybrid wavelet convolution network, which used wavelet transform to provide sparse coding [17] and the CNN for sparse coding. The work of [18] showed that using wavelet transform to separate the data at different scales can ensure data linearization and separability.

In addition, Goodfellow et al. firstly proposed GAN within the field of data generation. After that, Ledig et al. [2] introduced GAN into SISR task, and proposed SRGAN, which enriched the high-frequency details of the reconstructed image [17]. Due to the fact that SRGAN uses perceptual loss to constrain the image reconstruction, the PSNR and SSIM of reconstructed image are low, but it has high subjective perception quality. Despite all this, there is still a big gap between the images reconstructed by SRGAN and the ground-truth. Wang et al. [3] improved the network structure and loss function of SRGAN, and proposed an enhanced super-resolution generative adversarial network (ESRGAN). ESRGAN can reconstruct images with more realistic texture than SRGAN but lack of research on SISR task with large scaling factor of 8. Zhang et al. [19] proposed RankerSRGAN with Ranker. RankerSRGAN trained a Ranker which can learn the behavior of perceptual metrics and then introduced a novel rank-content loss to optimize the perceptual quality. Moreover, it can combine the strengths of different super-resolution methods to generate better results. Nevertheless, RankerSRGAN focuses on the scaling factor of 4 and large scaling factor of 8 is not considered. Shang et al. [20] proposed RFB-ESRGAN with receptive field module, which solved the problem of large changes in texture details for SISR task. Besides, RFB-ESRGAN alternately uses different upsampling methods to reduce the computation complexity, and can reconstruct images that contain realistic textures. However, RFB-ESRGAN needs HR inputs, and does not consider the case of LR inputs.

The above-mentioned SISR methods based on GAN can enrich details in the reconstructed HR image and improve the overall visual quality. However, most of the current methods based on GAN focus on the scaling factor of 4 for SISR task. So high-frequency information of the reconstructed HR image cannot be well restored when the scaling factor is large. Note that, wavelet packet transform can extract

information of different frequencies of the image, and the wavelet transform coefficients contain rich high-frequency information. Therefore, the proposed scheme utilizes wavelet packet transform, and GAN in wavelet domain is utilized to reconstruct the HR image, resulting in rich high-frequency details under the condition of large scaling factor.

3 The Proposed Scheme

3.1 Overall Network Architecture

In this paper, the GAN in wavelet domain for SISR task is proposed. The overall network architecture is shown in Fig. 1. The generator consists of feature extraction module and coefficient prediction module, where the feature extraction module extracts features of the LR image, and the coefficient prediction module generates the wavelet packet decomposition coefficients of the reconstructed HR image according to the extracted features.

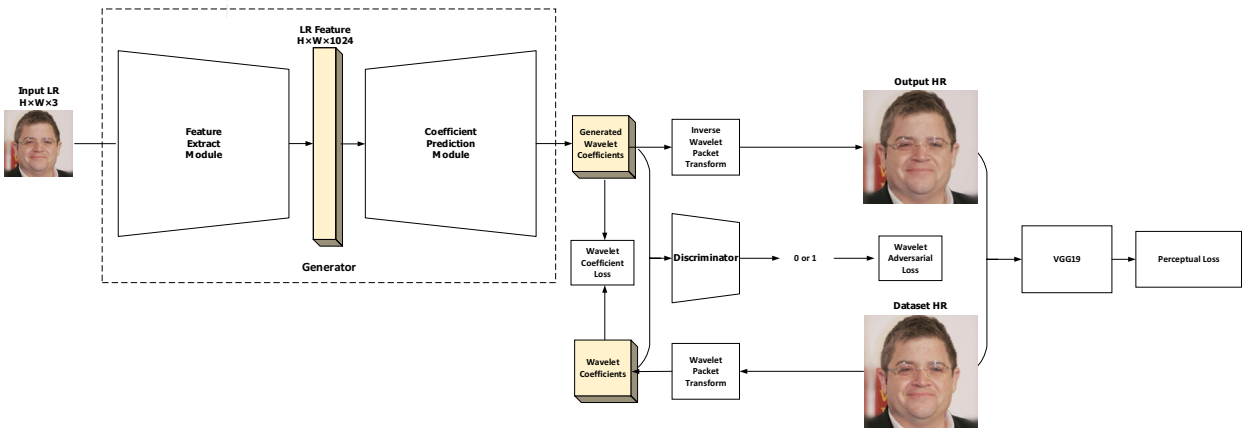

Fig. 1. Overall Network Architecture

The training process of the proposed scheme is as follows. First, the LR image is taken as the input, and then the feature extraction module of the generator extracts the features of the LR image, and the coefficient prediction module generates the wavelet packet decomposition coefficients of the specific scaling factor based on the extracted features. After that, the generated wavelet packet decomposition coefficients and the ones of the ground-truth (HR image in dataset) dataset HR image are input into the discriminator to compute the adversarial loss in wavelet domain. Then, the wavelet coefficient loss between the generated wavelet coefficients and the ones of the ground-truth is calculated. Afterwards, the generated wavelet coefficients are transformed into reconstructed HR images through the inverse wavelet packet transform, and the reconstructed HR image and the HR image in dataset are taken as input into the VGG19 [21] to obtain the perceptual loss. Finally, the total loss is obtained through the weighted sum of above-mentioned three losses (wavelet adversarial loss, wavelet coefficient loss, and perceptual loss), and back propagated to the generator to update the network parameters.

3.2 Network Structure

3.2.1 Generator

Feature Extraction Module.

The feature extraction module takes LR image with size of W×H×3 as input. The features are obtained through a convolution layer, and then features are input into the residual block [12] to obtain $\tilde{f}$ with size of W×H×1024, as shown in Equation (1):

$$\tilde{f} = F(x) \tag{1}$$

where W is the width of the LR image, H is the height of the LR image, $F(\cdot)$ represents the feature extraction module, x represents the LR image, and $f(\sim)$ represents the extracted feature maps.

In the feature extraction module, the size of all convolution kernels is 3×3, stride is 1, and pad is 1. In this way, the height and width of each feature map are the same as the input. The number of channels increases from 128 to 1024, which is used to mine information for the subsequent coefficient prediction module. The network structure is shown in Fig. 2.

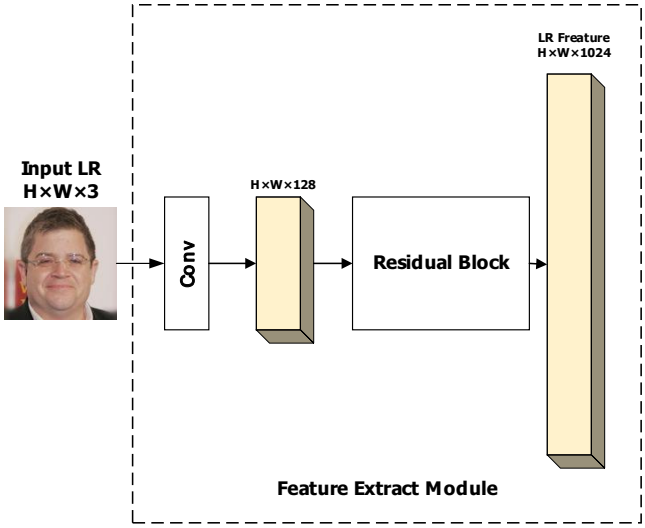

Fig. 2. Feature Extract Module

Coefficient Prediction Module.

The proposed scheme processes images in wavelet domain because of the strong visuality and efficiency of wavelet transform in processing multi-resolution images [22]. In this paper, Haar wavelet packet transform [23-24] is used, which can fully describe facial information with low computational complexity.

When the scaling factor is 8, the level of wavelet packet decomposition is 3. The coefficient prediction module generates 64 sets of wavelet packet decomposition coefficients based on the $\tilde{f}$, as shown in Equation (2):

$$\begin{aligned} C_0 &= P_0(\tilde{f}) \\ C_1 &= P_1(C_0) \\ C_2 &= P_2(C_1) \\ C_3 &= P_3(C_2) \end{aligned} \tag{2}$$

where $P_i(\cdot)$ represents the coefficient prediction module of the i-th level, C_j represents the j-th level wavelet packet decomposition coefficients. When $j = 0$, there exists 1 set of wavelet packet decomposition coefficients. 4 sets when $j = 1$, 16 sets when $j = 2$, and 64 sets when $j = 3$. 64 sets of wavelet packet decomposition coefficients represent the different frequencies contained in the input image respectively.

In the coefficient prediction module, the convolution kernel size of all convolution layers is 3×3, stride is 1, and pad is 1. So, the size of each generated wavelet coefficient is the same as input. In addition, due to the high independence between wavelet transform coefficients, HR image with different scaling factors can be reconstructed according to the different number of generated wavelet packet decomposition coefficients. The structure of the coefficient prediction module is shown in Fig. 3.

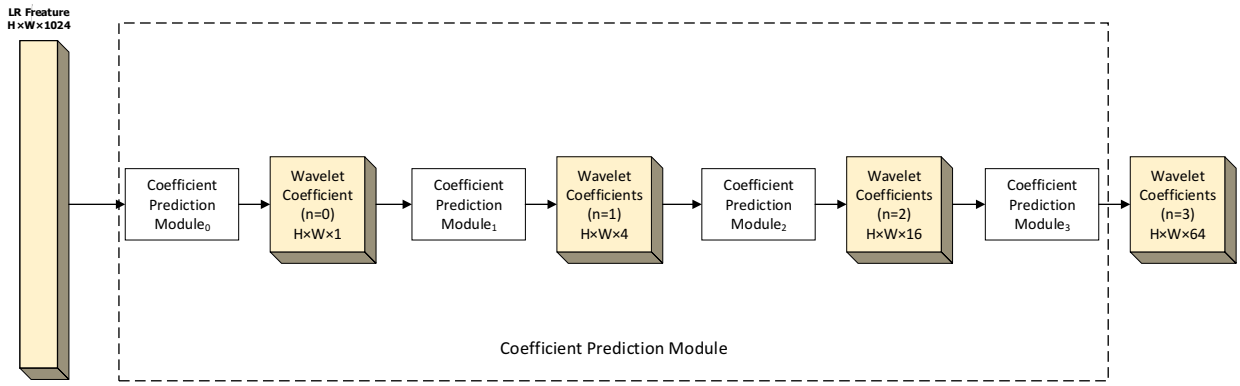

Fig. 3. Coefficient prediction module

3.2.2 Discriminator

In order to identify the wavelet packet decomposition coefficients of HR images from the generated wavelet coefficients, the discriminator in wavelet domain is constructed in this paper. The structure is shown in Fig. 4. Specifically, the discriminator contains 8 convolution layers, the kernel size of each convolution layer is 3×3, and the number of channels increases from 64 to 512. When the number of channels is doubled, dilated convolution is used to keep the image size unchanged. After obtaining the 512 feature maps, two convolution layers, LeakyReLU ($\alpha = 0.2$) and the sigmoid activation function are followed. As the number of channels increases and the size of feature maps is significantly reduced after 8 convolution layers, the obtained feature maps contain more semantic information. In order to avoid the loss of semantic information, the average pooling is used to replace the maximum pooling. Finally, the wavelet adversarial loss is calculated based on the output of discriminator, and back propagated to the generator.

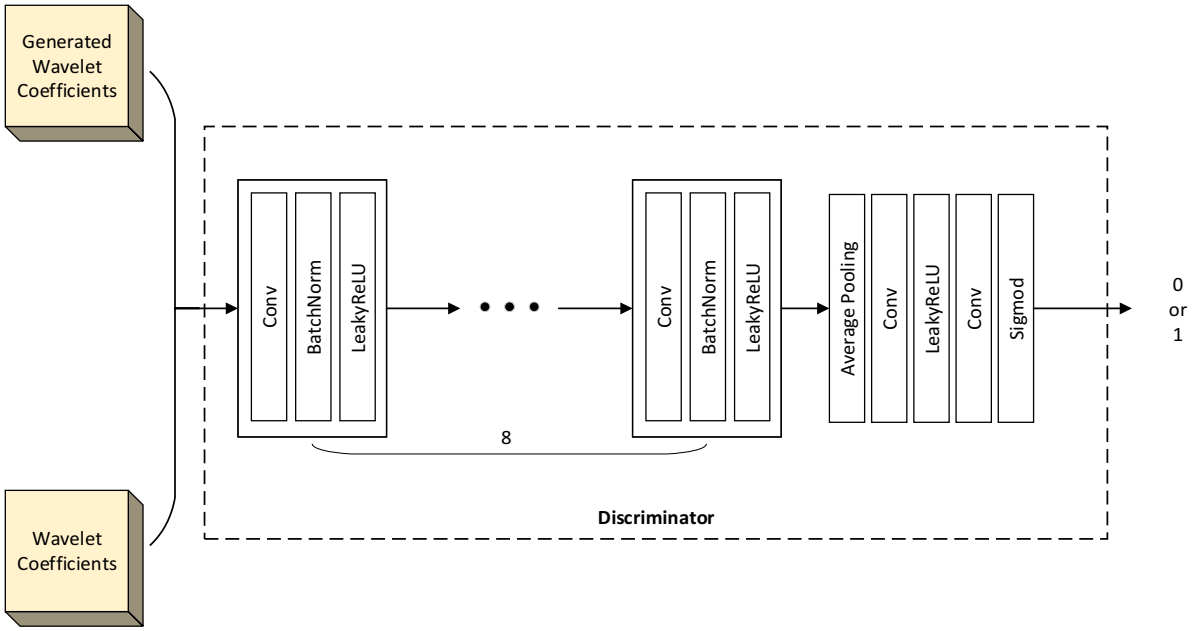

Fig. 4. The Structure of Discriminator

3.3 Loss Function

In this paper, wavelet coefficient loss and wavelet adversarial loss are used to constrain the wavelet coefficients generated by the generator in wavelet domain, which aim to minimize the distance between the generated wavelet packet decomposition coefficients and the ones of HR images in dataset. In

addition, in order to improve the perceptual quality of the reconstructed HR image, the perceptual loss is calculated using VGG19 and back propagated to the generator after obtaining the final reconstructed image. Three loss functions are jointly used to train the network to improve the quality of the reconstructed HR image.

3.3.1 Wavelet Coefficient Loss

The wavelet coefficient loss is shown in Equation (3):

$$\begin{aligned} l_{wavelet}(\hat{C},C) &= \| M \otimes (\hat{C}-C) \|_1 \\ &= \sum_{i=1}^{64} \lambda_i \| \hat{c}_i - c_i \|_1 \\ &= \lambda_1 \| \hat{c}_1 - c_1 \|_1 \\ &\quad + \sum_{i=2}^{64} \lambda_i \| \hat{c}_i - c_i \|_1 \end{aligned} \tag{3}$$

where $C = (c_1, c_2, ..., c_{64})$ and $\hat{C} = (\hat{c}_1, \hat{c}_2, ..., \hat{c}_{64})$ are respectively the wavelet packet decomposition coefficients of the ground-truth and the ones generated by the generator, $M = (\lambda_1, \lambda_2, ..., \lambda_{64})$ is the weight matrix which adjusts the weights of wavelet coefficients at different frequencies, λ_i is the weight of L1 loss between the *i*-th generated wavelet coefficients and the ones of HR image in dataset. Because different wavelet coefficients contain different feature information, the weights are given to make the network focus on the details of image. The wavelet coefficient loss is applied to the whole mapping process between wavelet packet decomposition coefficients generated by the generator and the ones of ground-truth. Using this loss in training can capture overall topology structure and improve the stability of training.

3.3.2 Wavelet Adversarial Loss

The generator generates the wavelet coefficients of the reconstructed HR image under the specific scaling factor. Hence, adversarial loss in wavelet domain is used to constrain the generation process of the generator. The wavelet adversarial loss is shown in Equation (4):

$$\begin{aligned} l_{adv} &= E[\log D(C)] + E[\log(1 - D(G(x)))] \\ &= E[\log D(C)] + E[\log(1 - D(\hat{C}))] \end{aligned} \tag{4}$$

where $G(\cdot)$ and $D(\cdot)$ denote generator and discriminator respectively, x denotes LR image, C represents the wavelet packet decomposition coefficients of HR image in dataset, and $\hat{C}$ denotes wavelet packet decomposition coefficients generated by the generator. The loss is applied to the process of mapping generated wavelet coefficients to real wavelet coefficients, and the adversarial loss adopted in this paper is the original cross entropy loss.

3.3.3 Perceptual Loss

In order to improve the perceptual quality of the reconstructed HR image, perceptual loss is introduced into the loss function. The perception loss is shown in Equation (5):

$$l_{perceptual}(\hat{y}, y) = \frac{1}{C_j \times H_j \times W_j} \| \varphi_j(\hat{y}) - \varphi_j(y) \|_2 \tag{5}$$

where $\varphi_j(\cdot)$ denotes the features extracted from the j-th layer of VGG19, and the output size is $H_j \times W_j \times C_j$. $\hat{y}$ and y respectively refer to the reconstructed HR image and the original HR image. In the training process, the reconstructed HR image and the original HR image are input into the pre-trained VGG19 to obtain the corresponding features respectively, and the perception loss is constructed by calculating the minimum mean square error loss between the features of reconstructed HR image and original HR image.

In essence, perception loss is introduced to further improve the detail information and subjective visual quality of the reconstructed HR image.

4 Experiments

4.1 Dataset

This paper conducts experiments on CelebFaces Attributes Dataset (CelebA). CelebA dataset is a public dataset of the Chinese University of Hong Kong, which contains 202,599 images of 10,177 celebrities. Each image has 40 attribute annotations. The dataset covers large pose changes and background clutter, and remains the characteristics of diversity, large quantity and rich labels. Due to the limitation of equipment performance, 800 images are randomly selected from the CelebA dataset as training data and 50 images as test data. Fig. 5 shows some face images in CelebA dataset.

Fig. 5. Some Face Images of CelebA Dataset

4.2 Experimental Setup

Typically, the size of HR images in the dataset is 512×512, and the training epoch is set to 500. In order to speed up the training, the generator is pre-trained with the training epoch of 30. Wavelet coefficient loss and wavelet adversarial loss are used to constrain the generator. After the pre-training is completed, the discriminator and wavelet adversarial loss are added to train the network. The initial learning rate is set to 0.0001. Adam optimizer, with parameters set to =0. 9 and = 0. 999, is adopted to update the network parameters. In terms of network initialization, weights of convolution layer adopt the random initialization of normal distribution, and bias is initialized to 0. During the training, the batch size is set to 4. All codes in the proposed method are implemented in Python and the repository is PyTorch. The operating system is Ubuntu Linux 18.04.1 LTS. The GPUs used in training and testing is the NVIDIA Geforce RTX 2080 Ti, and the memory is 11GB. In order to objectively evaluate the quality of reconstructed HR images, the proposed scheme utilizes Structural Similarity Index (SSIM) and Peak Signal to Noise Ratio (PSNR). PSNR is used to evaluate the degree of the reconstructed image coloring. The larger the PSNR, the less the distortion. SSIM is used to measure the structure similarity. The larger the SSIM, the higher the similarity between two images.

Algorithm The Proposed Scheme

Input: LR Image

1. Extract feature $\hat{f}$ from LR image and generate wavelet coefficients $\hat{C}\{\hat{c}_i\}$ using generator；
2. Obtain wavelet coefficients $C\{c_i\}$ using HR image in dataset with wavelet packet transform;
3. Calculate wavelet adversarial loss using $\hat{C}\{\hat{c}_i\}$ with discriminator;
4. Calculate wavelet coefficient loss between $C\{c_i\}$ and $\hat{C}\{\hat{c}_i\}$;
5. Reconstruct HR image using $\hat{C}\{\hat{c}_i\}$ by wavelet packet transform
6. Calculate perceptual loss between reconstructed HR image and the one in dataset with VGG19;
7. Calculate total loss and back propagated to the generator;
8. **End**

Fig. 6. The Training Process of Proposed Scheme

4.3 Experimental Results and Analysis

4.3.1 Ablation Experiment

In order to verify the effectiveness of proposed scheme, 50 face images randomly selected from CelebA dataset are utilized. The ablation experiment is performed with scaling factors of 4 and 8. In the ablation experiment, Bicubic [25] and Ours-withoutwa are the baseline methods. Typically, Ours-withoutwa is obtained by removing the wavelet adversarial loss. The reconstructed HR images are obtained by performing ×4 and ×8 image super-resolution reconstruction. Fig. 7 shows the HR face images with the scaling factor of 8. In Fig. 7, the HR images reconstructed by Bicubic have obvious distortion and jaggies, and the details reconstructed by Ours-withoutwa are not recovered, especially in areas such as hair and wrinkles. In short, the loss of details and textures is serious. The experimental results in Fig. 7 show that, compared with the baseline methods, the proposed scheme achieves better performance, which can restore more detail and texture information while ensuring image topology structure.

(a) LR images (b) Ours-withoutwa (c) Ours (d) HR images

Fig. 7. Results of baseline methods with scaling factor of 8

In addition, the performances are quantitatively analyzed by calculating PSNR and SSIM, and the results are shown in Table 1. The results in Table 1 show that the proposed method has better performance in improving quality of reconstructed images, which verifies the effectiveness of the proposed method.

Table 1. Performance of ablation experiment

methods	PSNR/dB		SSIM	
	×4	×8	×4	×8
Ours	**29.325**	**25.968**	**0.789**	**0.719**
Bicubic	28.729	25.823	0.778	0.710
Ours-withoutwa	28.936	25.867	0.782	0.713

4.3.2 Comparative Experiment

In this paper, 50 face images randomly selected from CelebA dataset are used to carry out comparative experiments with the scaling factor of 4 and 8. In order to verify the superiority of the proposed method, SRGAN [2], ESRGAN [3], RankerSRGAN [19] and RFB-SRGAN [20] are selected as comparison methods. Note that, the same training environment is adopted in the comparison methods, and the value in the table is the average.

Table 2 quantitatively compares PSNR and SSIM. The experimental results show that the PSNR of the proposed method is greatly improved compared with the comparison methods, and the performance is better than that of the comparison methods with scaling factors of 4 and 8. The minimum improvements of PSNR and SSIM are 1.783dB and 0.013 on scaling factor ×4, and 0.685dB and 0.016 on scaling factor ×8 when compared with the comparison methods.

Table 2. Performance of comparative experiment

methods	PSNR/dB		SSIM	
	×4	×8	×4	×8
Ours	**29.325**	**25.968**	**0.789**	**0.719**
SRGAN	27.542	25.283	0.762	0.684
ESRGAN	27.426	24.827	0.776	0.703
RankerSRGAN	26.443	24.018	0.700	0.628
RFB-SRGAN	26.425	25.030	0.746	0.702

The visualized comparisons of different methods are shown in Fig. 8 and Fig. 9. Fig. 8 is the visualized results on scaling factor ×4, and Fig. 9 is the visualized results on scaling factor ×8. Specifically, Fig. 8(a) is the LR image, Fig. 8(b) is the HR image reconstructed by SRGAN, Fig. 8(c) is the HR image reconstructed by ESRGAN, Fig. 8(d) is the HR image reconstructed by RankerSRGAN, Fig. 8(e) is the HR image reconstructed by RFB-SRGAN, Fig. 8(f) is the HR image reconstructed by our proposed method, and Fig. 8(g) is the HR image in the CelebA dataset. The experimental results in Fig. 8 show that our method has achieved better performance than the comparison methods, and the recovery of face topology structure and detail information is relatively complete. However, color distortion occurs in Fig. 8(b), Fig. 8(c) and Fig. 8(e) methods, which affect the quality of reconstructed HR images. The results of Fig. 9 show that the current methods based on GAN have poor reconstruction performance when the scaling factor is 8. The Fig. 9(b) method cannot correctly reconstruct detail information of the image. Furthermore, the face area has reconstruction errors, and there are artifacts on the edge. The reconstructed image of Fig. 9(c) is blurred and the details are not restored. The reconstruction performances of Fig. 9(d) and Fig. 9(e) are better than that of Fig. 9(b) and Fig. 9(c), but the reconstruction image of Fig. 9(d) has partial structural distortion, and the authenticity of the reconstructed image is poor. The Fig. 9(e) method has serious checkerboard effect and structural distortion, which affect the quality of the reconstructed image. Compared with the comparison methods, the overall visualized result of our proposed scheme is better due to the use of high-frequency information of the input image. Therefore, our scheme has less distortion in the reconstructed image, recovers more details, and has relatively clearer detail and structure information.

(a) LR image (b) SRGAN (c) ESRGAN (d) RankerSRGAN (e) RFB-SRGAN (f) Ours (g) HR image

Fig. 8. Experimental results on scaling factor ×4

(a) LR image (b) SRGAN (c) ESRGAN (d) RankerSRGAN (e) RFB-SRGAN (f) Ours (g) HR image

Fig. 9. Experimental results on scaling factor ×8

Furthermore, the distributions of images synthesized by different methods under different performance evaluation criteria are calculated in the comparative experiment. Results are shown in Fig. 10, Fig. 11, Fig. 12 and Fig. 13. Fig. 10 is the distribution on scaling factor ×4 under PSNR, Fig. 11 is the distribution on scaling factor ×8 under PSNR, Fig. 12 is the distribution on scaling factor ×4 under SSIM, and Fig. 13 is the distribution on scaling factor ×8 under SSIM. The X-axis represents the value of PSNR or SSIM,

and the Y-axis represents the percentage of images. The higher the percentage of images, the stronger the reconstruction ability of the corresponding model. The experimental results show that under the conditions of different scaling factors, the proportion of the number of high-quality images synthesized by the proposed method is higher than that of other comparative methods, and the ability to synthesize high-quality images is stronger.

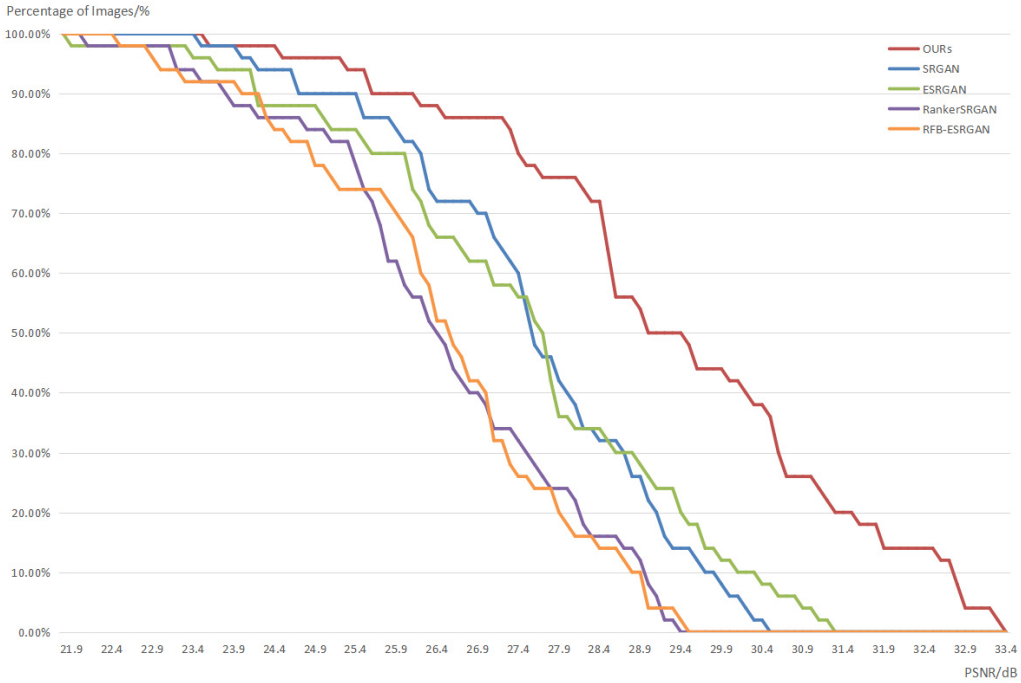

Fig. 10. The distribution of images on scaling factor ×4 under PSNR

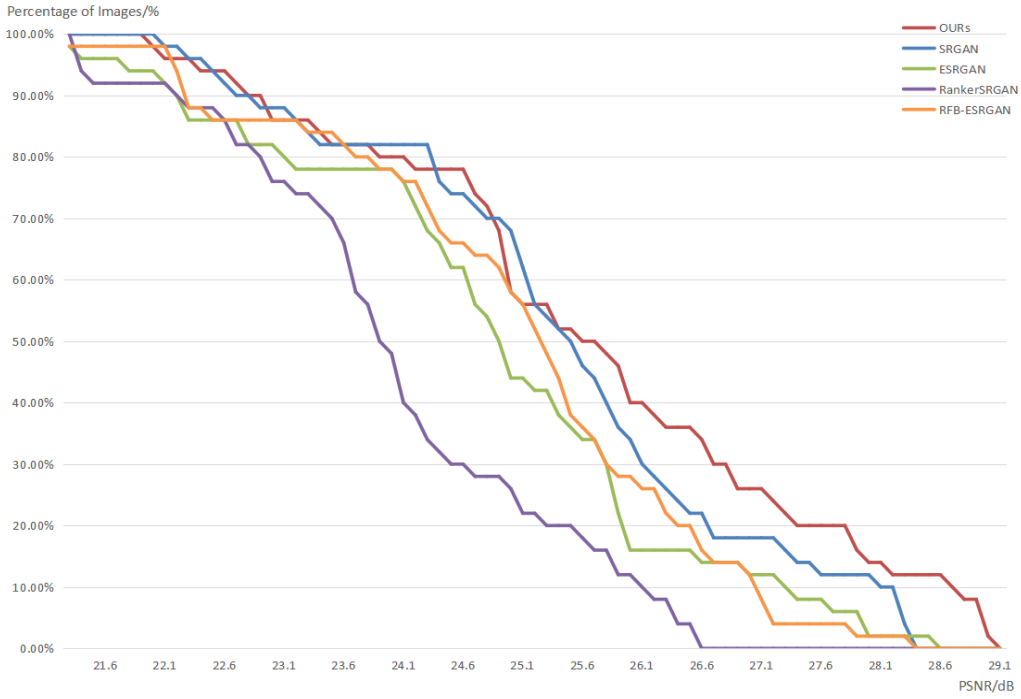

Fig. 11. The distribution of images on scaling factor ×8 under PSNR

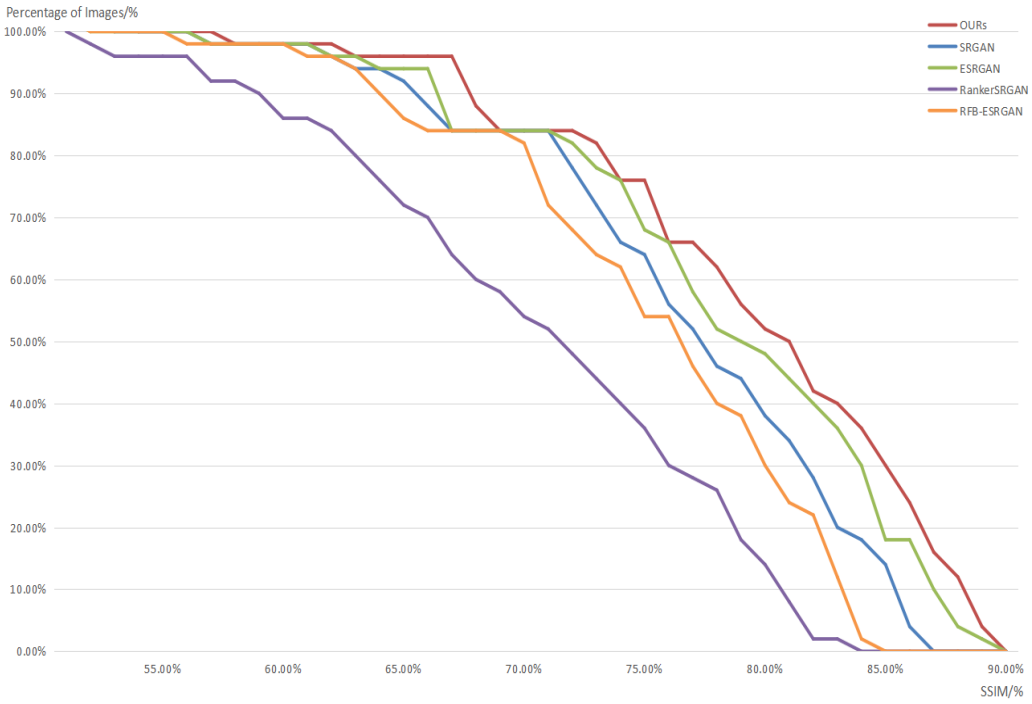

Fig. 12. The distribution of images on scaling factor ×4 under SSIM

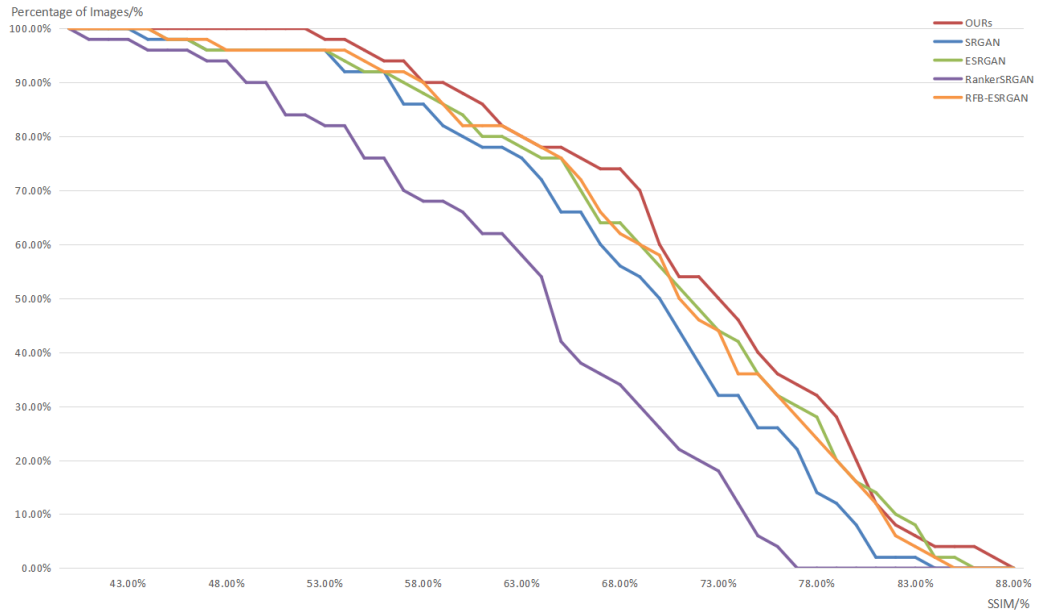

Fig. 13. The distribution of images on scaling factor ×8 under SSIM

5 Conclusion

In this paper, the GAN in wavelet domain for SISR task is proposed, which can reconstruct HR images with better quality when the scaling factor is large. The generator generates the wavelet packet decomposition coefficients corresponding to the reconstructed HR image based on the LR image, and realizes the HR image reconstruction in wavelet domain. The discriminator discriminates the generator performance in wavelet domain and constructs adversarial loss to constrain the generator. In this scheme, wavelet coefficient loss, wavelet adversarial loss and perception loss are used in the training process. Experimental results show that the proposed method has achieved better results in subjective visual

quality and quantitative performance, and the details of reconstructed images are fully restored. As future work, it is of great significance to improve the restoration effect of image details and visual quality.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (No.U20A20163, No. 62001033), Beijing Information Technology University "Qinxin Talent" Cultivation Program (No. QXTCPA201902), Beijing Municipal Education Commission General Project (No. KM202011232021, No.KZ202111232049) and Fundamental Research Funds in Beijing Information Science and Technology University (No. 2025019).

References

[1] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Proc. 2014 Advances in Neural Information Processing Systems, 2014.

[2] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic single image super-resolution using a generative adversarial network, in: Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[3] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C.C. Loy, ESRGAN: Enhanced super-resolution generative adversarial networks, in: Proc. 2018 European Conference on Computer Vision Workshops, 2018.

[4] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: Proc. 2014 European Conference on Computer Vision, 2014.

[5] J. Kim, J.K. Lee, K.M. Lee, Accurate image super-resolution using very deep convolutional networks, in: Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[6] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[7] C. Dyer, A. Kuncoro, M. Ballesteros, N.A. Smith, Recurrent neural network grammars, in: Proc. 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016.

[8] T. Tong, G. Li, X. Liu, Q. Gao, Image super-resolution using dense skip connections, in: Proc. 2017 IEEE International Conference on Computer Vision, 2017.

[9] G. Huang, Z. Liu, L.V.D Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[10] B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, Enhanced deep residual networks for single image super-resolution, in: Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017.

[11] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution, in: Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.

[12] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[13] S.S. Panda, G. Jena, Image Super Resolution Using Wavelet Transformation Based Genetic Algorithm, in: H. Behera, D. Mohapatra (eds) Computational Intelligence in Data Mining, vol 2, Springer Press, New Delhi, 2016.

[14] T. Guo, H.S. Mousavi, T.H. Vu, V. Monga, Deep wavelet prediction for image super-resolution, in: Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017.

[15] T. Wang, W. Sun, H. Qi, P. Ren, Aerial image super resolution via wavelet multiscale convolutional neural networks, IEEE Geoscience and Remote Sensing Letters 15(5)(2018) 769-773.

[16] X. Gao, H. Xiong, A hybrid wavelet convolution network with sparse-coding for image super-resolution, in: Proc. 2016 IEEE International Conference on Image Processing, 2016.

[17] P.M. Sheridan, F. Cai, C. Du, W. Ma, Z. Zhang, W.D. Lu, Sparse coding with memristor networks, Nature nanotechnology 12(8)(2017) 784-789.

[18] S. Mallat, Understanding deep convolutional networks, Philosophical Transactions of the Royal Society A Mathematical, Physical and Engineering Sciences 374(2065)(2016) 20150203.

[19] W. Zhang, Y. Liu, C. Dong, Y. Qiao, RankSRGAN: Generative adversarial networks with ranker for image super-resolution, in: Proc. 2019 IEEE/CVF International Conference on Computer Vision, 2019.

[20] T. Shang, Q. Dai, S. Zhu, T. Yang, Y. Guo, Perceptual extreme super resolution network with receptive field block, in: Proc. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020.

[21] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proc. 2015 International Conference on Learning Representations, 2015

[22] H. Qayyum, M. Majid, S.M. Anwar, B. Khan, Facial expression recognition using stationary wavelet transform features, Mathematical Problems in Engineering (2017) 1-9.

[23] S. Djaballah, K. Meftah, K. Khelil, M. Tedjini, L. Sedira, Detection and diagnosis of fault bearing using wavelet packet transform and neural network, Frattura ed Integrità Strutturale 13(49)(2019) 291-301.

[24] M. Alemohammad, J.R. Stroud, B.T. Bosworth, M.A. Foster, High-speed all-optical Haar wavelet transform for real-time image compression, Optics express 25(9)(2017) 9802-9811.

[25] W. Ruangsang, S. Aramvith, Efficient super-resolution algorithm using overlapping bicubic interpolation, in: Proc. 2017 IEEE 6th Global Conference on Consumer Electronics, 2017.