ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0012

An Enhanced Image Super-Resolution Via Generative Adversarial Network

Dongmei Ma[1, 2, a], Yu Li[1, 2, b, *]

[1]School of Physics and Electronic Engineering Northwest Normal University, Gansu Lanzhou, China

[2]Engineering Research Center of Gansu Province for Intelligent Information Technology and Application, Gansu Lanzhou, China

[a]madongmei@nwnu.edu.cn, [b]2633103777@qq.com

Abstract

Image super-resolution algorithms have problems such as excessive smoothing of reconstructed images and poor visual quaity. Based on the existing image super-resolution algorithm SRGAN, an improved algorithm named PA-SRGAN is proposed. The generator uses residual dense connection blocks instead of traditional residual units to ensure the input low-resolution image features are fully transmitted throughout the generator, and pyramid attention is introduced to further improve the model performance. The discriminator adopts a relativistic average discriminator to improve the quality of the generative images and spectral normalization is used to enhance the training stability of the discriminant network. The whole model uses perceptual loss to prevent the image from being over-smoothed and improve the perceptual quality of generated images. The experimental results show that the algorithm can obtain better perception coefficients in Set5, Set14, BSD100, Urban100 general test sets, the generative images are more in line with human visual perception.

Keywords

Image super-resolution; Generatative adversarial network; Residual dense connection; Relativistic average discriminator; Perceptual loss.

1. INTRODUCTION

Single image super-resolution reconstruction is a common method to improve image resolution，which purpose is to restore the high resolution image (HR) from a given single low resolution image (LR)[1]. Recently years, generative adversarial network(GAN) gradually attracted attention of researchers and applied in image super-resolution. Ledig et al[2].proposed the first image super-resolution based on generative adversarial network SRGAN, which used SRResNet as generator, and introduced perceptual loss to optimized model, which made the model paid more attention to the high frequency details. Wang et al[3]. proposed an enhanced super-resolution generative adversarial network ESRGAN, used RRDB as basic module of generative network, the relativistic average discriminator is introduced to improve the quality of generative images. Wang et al[4]. used semantic segmentation images as prior information to prevent structural distortion and proposed SFTGAN. Soh et al[5]. proposed NatSR, which used natural flow discriminator to restore the texture of images.

Although existing algorithms based on generative adversarial network, such as SRGAN, NatSR, SFTGAN, can obtain more richer details, they are easy to generate redundant details in the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0012

reconstruction image. For this problem, we propose an enhanced image super-resolution algorithm based on generative adversarial network named PA-SRGAN.The generator uses residual dense block instead of residual block and add pyramid attention module to enhance the performance of the generator. The discriminator optimizes by using relativistic average discriminator and uses spectral normalization instead of batch normalization. The purposed methord can obtain images with better visual effect.

2. METHORD

In this part, we will show all details of our PA-SRGAN model, including overall framework of PA-SRGAN, residual dense block, pyramid attention module and the overall objective function.

2.1. Structural of generator

The whole structure of generator is shown in figure1.The generator includes two stages: feature extraction and reconstruction. Input a low-resolution image and get 64 feature maps throuth a 3×3 convolutional layer, the output of 64 feature maps as input of RDB modules and the pyramid attention module, the output feature of the last RDB module add with the output feature of the first convolutional layer, then use a 3×3 convolutional layer to extract useful feature. The reconstruction stage consists two sub-pixel layer with upsample factor ×2, after upsample, a 3 channel feature map can be obtained by using a 3×3 convolutional layer. Last, use bicubic to generate bicubic image and add with the output of the last convolutional layer.Figure1 shows the structual of the generator.

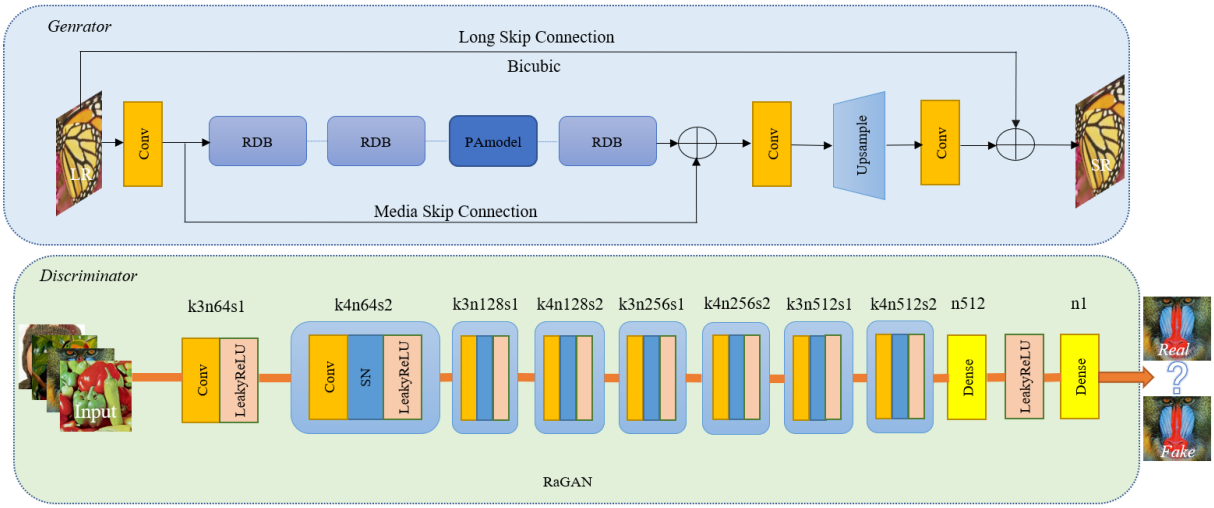


Figure 1. Overall framework of PA-SRGAN

2.1.1 Residual dense block

Figure2 is residual dense block. LR image has strong relevance with HR image, the shallow features of LR image are more helpful to reconstruct HR image. However, some information will be lost in the process of feature extract by convolutional layer, and the network usually prefers to use the deep feature information of LR image, the shallow feature of LR image cannot be fully utilized[6-8].

For this problem, residual dense block is introduced instead of residual block, which make full use of the shallow features of LR image in each convolutional layer. Movever, PReLU is used to solve the neuronal death in ReLU function, meanwhile, the output mean value of the PReLU function is approximately zero, which convergence speed is faster than relu.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0012

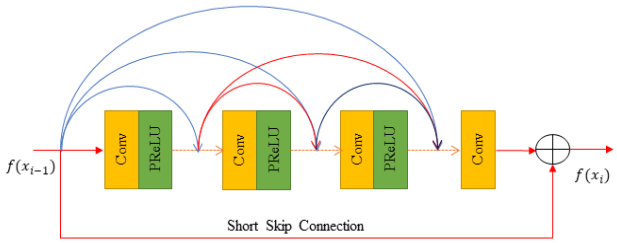


Figure 2. Resifual dense block

2.1.2 Pyramid attention

The existing image super-resolution reconstruction method based on depth learning extracts prior information on a single scale through self-attention mechanism, for example, non-local attention is a self-attention. However, the self-attention mechanism cannot fully utilize the feature information of different scales. Therefore, the pyramid attention module[9] is introduced to expand non-local attention to multi-scale space, making it easier for the model to capture multi-scale feature responses. Figure3 is overall structual of pyramid attention module.

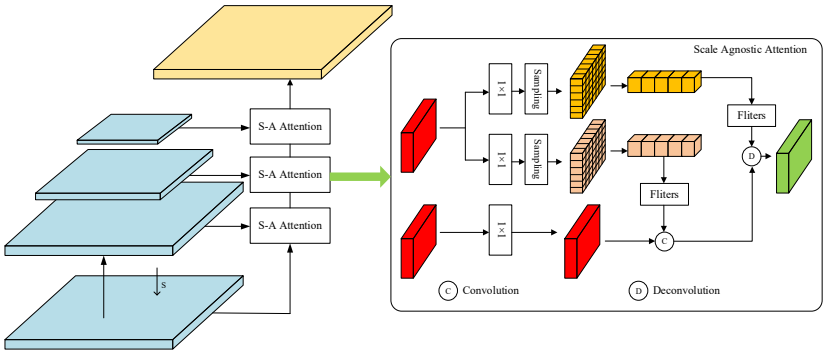


Figure 3. Structual of pyramid attention

x is denote input feature, $\Phi(.)$ is used to Calculate correlation information for different locations, $\theta(.)$ is feature conversion function, $\sigma(x)$ is normalize function. Use eq.1 to calculate the characteristic number of the j-th position:

$$y^i = \frac{1}{\sigma(x)} \sum_j \Phi(x^i, x^j)\theta(x^j) \tag{1}$$

$S = \{1, S_1, S_2, \ldots, S_n\}$ is denote multi-scale feature, eq.1 is expanded to multi-scale non-local attention:

$$y^i = \frac{1}{\sigma(x)} \sum_{s \in S} \sum_j \Phi(x^i, x^j_{\omega(s)})\theta(x^j_{\omega(s)}) \tag{2}$$

Where $\omega(s)$ is the x^2 field centered on point j. Downsampling input $x^j_{\omega(s)}$ and obtain the downsample feature z on dimension s. Scale agnostic attention[9] can be obtained by eq.3:

$$y^i = \frac{1}{\sigma(x,z)} \sum_j \Phi(x^i, z^j)\theta(z^j) \tag{3}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0012

Expanding this scale agnostic attention to all pyramid attention and calculating the relevance of multiple scales. $F = \{F_1, F_2, \dots, F_n\}$ is denote feature pyramid consisted by scale S,where $F_i = (\frac{H}{S_i}, \frac{W}{S_i})$

The pyramid attention can be expressed as:

$$y^i = \frac{1}{\sigma(x)} \sum_{z \in F} \sum_{j \in z} \Phi(x^i_{\omega(r)}, z^j_{\omega(r)}) \theta(z^j) \tag{4}$$

Where $\omega(.)$ and $\theta(.)$is similarity transformation function and feature conversion function. According to the Mei et al. $\omega(.)$ choose the embedding function, $\theta(.)$ choose the Gauss function.

2.2. Structual of the discriminator

As shown in figure1, our discriminator is roughly similar as VGG, but the discriminator of proposed methord uses 4×4 convolutional layers with stride 2 to downsample feature. Spectral normalization (SN)[10] is used in the discriminator to ensure the training stability of the discriminator.

If the discriminator function $D(.)$ is satisfied :

$$| D(x_1) - D(x_2) | \leq K | x_1 - x_2 | \tag{5}$$

the discriminator content with Lipschitz condition. In figure1, LeakyReLU has already satisfied Lipschitz condition, therefore, SN only used on convolutional layers. The weight of the convolutional layer after use SN can be expressed as:

$$\tilde{W} = \frac{W_n}{\sigma(W_n)} \tag{6}$$

Where $\sigma(.)$ is spectral norm of weight matrix W_n. $\sigma(.)$ is given by eq.7:

$$\sigma(W) = \max_{||h|| \neq 0} \frac{|| Wh ||_2}{|| h ||_2} = \max_{||h||=1} || W ||_2 \tag{7}$$

2.3. Loss function

The overall loss function is given by eq.8:

$$L = L_{percept} + \alpha L_G^{RaGAN} + \eta L_{pixel} \tag{8}$$

Where $L_{percept}$ is perceptual loss[11]. Figure4 shows the perceptual loss network of PA-SRGAN, use VGG19 network "Conv1_1" , "Conv2_1" , "Conv3_1" and "Conv4_1" as featre extractor , extract HR images' and SR images' feature , $\varphi(.)$ is denote VGG19 network, calculate feature loss between HR feature map and SR feature map via eq.9:

$$L_{percept} = \frac{1}{n} \sum_{i=1}^{n} | \varphi(y_i) - \varphi(\hat{y}_i) | \tag{9}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0012

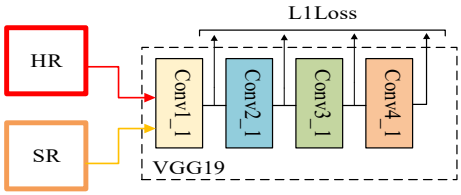


Figure 4. VGG perceptual loss network

L_G^{RaGAN} is adversarial loss of generator. This paper introduces relativistic average discriminator to improve the quality of generative images. x_r and x_f is real images and fake images, σ is sigmoid, the relativistic average discriminator(RaD)[12] can be expressed as:

$$\begin{cases} D(x_r, x_f) = \sigma(C(x_r) - E(C(x_f))) \\ D(x_f, x_r) = \sigma(C(x_f) - E(C(x_r))) \end{cases} \tag{10}$$

After use relativistic average discriminator, the adversarial loss between generator and discriminator is given by eq.11:

$$\begin{cases} L_D^{RaGAN} = -E_{x_r}[\log D(x_r, x_f)] - E_{x_f}[1 - \log(D(x_f, x_r))] \\ L_G^{RaGAN} = -E_{x_r}[1 - (\log D(x_r, x_f))] - E_{x_f}[\log D(x_f, x_r)] \end{cases} \tag{11}$$

L_{pixel} is pixel loss between HR image and SR image, use L1loss to calculate pixel loss:

$$L_{pixel} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{12}$$

In eq.8 , according Wang et al[3]. Set α=0.005, η=0.1.

3. EXPERIMENTS

3.1. Datasets

The train dataset is DIV2K, which consists 800 train images, 100 eval images and 100 test images. Because there are few samples in the DIV2K train dataset, the images in DIV2K train dataset is processed as follows:crop 800 HR images to 480×480 sub-images via crop stride 24, after crop complete, there are 32592 HR sub-images in the DIV2K train dataset, using random rotation and random crop to enhance train dataset. The input of the generator is 32×32.The test dataset select Set5,Set14,BSD100 and Urban100.

3.2. Experiments details

Firstly,use L1loss as objective function,training a PSNR-drivern model.The optimizer use Adam with β =(0.9, 0.999),initial learning rate is set to 2.0×10^{-4} and descend a half at [200K, 300K, 400K, 500K], set batchsize=16 ,train the PSNR-driven model total 600K iters. Then, load the PSNR-drivern model's parameters to train the whole generative adversarial network, the generator and the discriminator select Adam optimizer with β =(0.9, 0.999), initial learning rate is 1.0×10^{-4} and descend a half at [50K, 100K, 150K, 200K], set batchsize=16,train the GAN model total 300K iters. There are 16 RDB modules in the generator and the location of the PA-model is the output of 8-th RDB module. We design all experiments based on Pytorch on NVIDIA RTX3080 GPUs.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0012

For model performance, PSNR and SSIM is used to evaluate objective indicators. Before calculate, every RGB images should be converted to YCbCr space and calculate PSNR and SSIM only on Y channel. PI and LPIPS is selected to evaluate perceptual quality.Lower PI and LPIPS value means the higher perceptual quality. In this paper, we define algorithms optimized by using L1 or L2 loss as PSNR-driven algorithms.

3.3. Experiment results

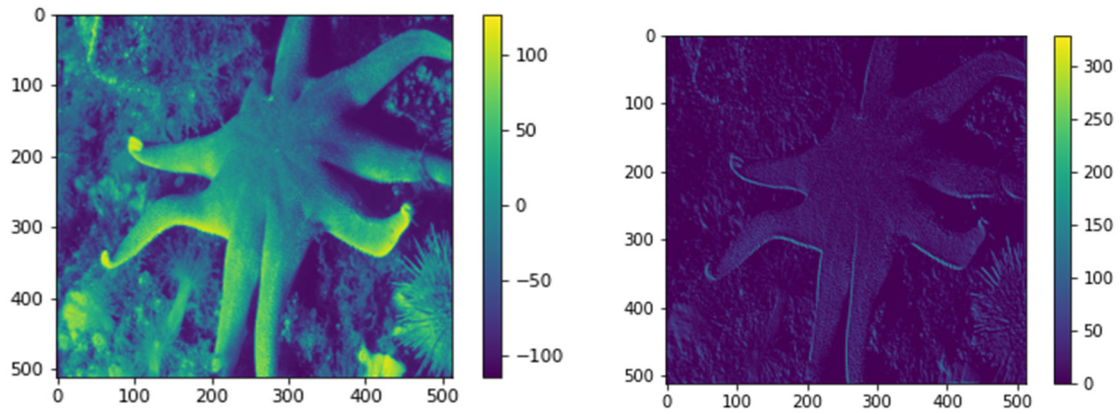
Validation of PA-model: For the effectiveness of PA-model in the proposed methord, an ablation experiment is designed as follows: P is denote the output of first RDB module, M is the output of 8-th RDB module, L is the output of the last RDB module. Use L1loss as objective function, calculate PSNR values when PA-model is not used in the generated network and PA-model is used in different locations of the generated network on the Set5 dataset. The results of this ablation experiment are shown in table1.

Table 1. Results of PA-model ablation experiment

	None	P	M	L
PSNR	32.03	32.18	32.44	32.39

In table1, the PSNR is 32.03 when the generator is not used PA-model, the lowest PSNR is 32.18 when the generator contain the PA-model, which is improved 0.15 compared with the generator without PA-model. For the location of the PA-model in the generator, the best PSNR is 32.44 when the PA-model located at M, this result is higher than P and L 0.26 and 0.05. The ablation experiment shows that when the generation network uses PA-model and is located in the middle of the 16 RDB modules of the generation network, the generation network performance is significantly improved.

Validation of perceptual loss: The purpose of this ablation experiment is to verify the output of VGG convolution layer can obtain clearer images. Figure5 shows the visualize results of output of ”Conv1_1” and ”ReLU1_1” in VGG19 network.In figure5(a), the feature map value of convolutional layer output is [-100,100], in figure5(b), the feature map value of ReLU output is [0,300]. The output of convolutional layer retains a lot of detailed features while the ReLU gets the sparse feature. Sparse features are not conducive to reconstruction.



(a) Conv1_1 (b) ReLU1_1

Figure 5. Visualize results of Conv1_1 and ReLU1_1

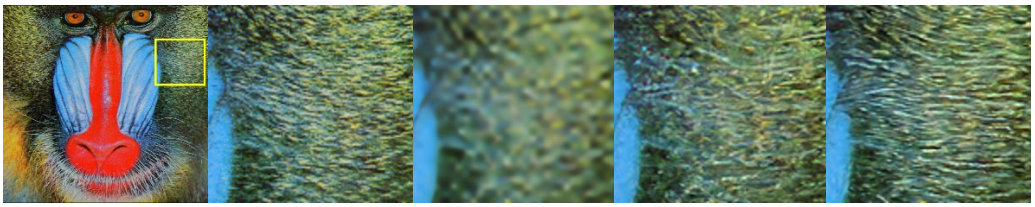
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0012

Use eq.8 as objective function, train GAN model and calculate PI value on Set14 dataset, table2 gives the results of the validation of perceptual loss:

Table 2. Results of effects of perceptual loss

	Baseline	GAN+after activation	GAN+before activation
PI	5.9806	3.0832	2.9013

In table2, the PI value is 5.9806 when only use generator, the PI value when use "GAN+after activation" is 3.0832 and the 2.9013 when use "GAN+before activation". The best PI appears in "GAN+before activation", which is lower than baseline 3.0793 and lower than "GAN+after activation" 0.1819.



baboon.png HR baseline after before

Figure 6. Visualize results of ablation experiment

Figure6 shows the results of baboon.png in Set14 dataset. The details of the image are lost seriously,which make the naked eye can hardly observe the details of the image when only use generator. Generated images by using "GAN+after activation" can produce richer details, while the images generated by using "GAN+before activation" has better visual effect and more closer to the original image.

Validation of relativistic average discriminator : The adversarial loss of standard GAN(SGAN) is :

$$\begin{cases} L_D = \log(D(x_r)) + \log(1 - D(x_f)) \\ L_G = \log(D(x_f)) \end{cases} \tag{13}$$

Now we use L_G in eq.13 as adversarial loss of generator and repleace the second item in eq.8, L_D as adversarial loss of discriminator. Calculate PI value when GAN model use SGAN and RaGAN on Set5 and Set14 test dataset, all results are shown in table3:

Table 3. Results of effects of relativistic average discriminator(PI)

	SGAN	RaGAN
Set5	3.8536	3.6841
Set14	3.0678	2.9013

In table3, the PI value of SGAN on Set5 test dataset is 3.8536 while the PI value of RaGAN is 3.6841. On the Set14 test dataset, the PI value of SGAN is 3.0678 and the RaGAN obtain 2.9013 PI value. This result shows that the performance of generator is improved when the relativistic average discriminator is used in the discriminator. It means that the relativistic average discriminator is helpful to generate higher perceptual quality images.

Compare with other algorithm: We choose some classical algorithms compare with the propose algorithm. All results are shown in table4 and table5.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0012

Table 4. Objective index of different algorithms(PSNR/SSIM)

	Set5	Set14	BSD100	Urban100
Bicubic	28.38/0.8148	26.01/0.7033	25.88/0.6700	23.13/0.7710
EDSR	32.46/0.8966	28.79/0.7874	27.70/0.7417	26.64/0.8031
SRResNet	32.02/0.9007	28.47/0.7905	27.55/0.7616	26.18/0.7582
SRGAN	29.16/0.8613	26.17/0.7841	25.46/0.6485	24.40/0.6988
DBPN	32.47/0.8980	28.82/0.7859	27.73/0.7401	26.37/0.7944
RCAN	32.63/0.9002	28.87/0.7889	27.77/0.7435	26.81/0.8088
RDN	32.46/0.8990	28.81/0.7871	27.72/0.7419	26.61/0.8028
SFTGAN	29.93/0.8665	26.22/0.7854	25.51/0.6549	24.01/0.7068
NatSR	30.99/0.8800	27.51/0.8140	26.45/0.6831	25.46/0.7432
PA-SRGAN	30.19/0.8624	26.44/0.7880	25.82/0.6598	24.60/0.7219

Table 5. Perceptual index of different algorithms(PI/LPIPS)

	Set5	Set14	BSD100	Urban100
Bicubic	7.3604/0.2798	6.9684/0.3639	6.9490/0.3696	6.8799/0.4814
EDSR	5.9819/0.2088	5.2594/0.2963	5.2625/0.3249	4.9844/0.2923
SRResNet	6.0075/0.2160	5.3057/0.3058	5.4252/0.3328	5.1817/0.2478
SRGAN	3.9820/0.0822	3.0851/0.1663	2.5459/0.1980	3.6980/0.1551
DBPN	6.1324/0.2108	5.4596/0.2985	5.4915/0.3250	5.1360/0.2748
RCAN	6.3749/0.2158	5.7127/0.3106	5.7588/0.3317	5.4181/0.2944
RDN	6.0092/0.2134	5.4633/0.3039	5.5412/0.3299	5.2502/0.3162
SFTGAN	3.7587/0.0890	2.9063/0.1480	2.3774/0.1469	3.6136/0.1433
NatSR	4.1648/0.0939	3.1094/0.1758	2.7801/0.2114	3.6523/0.1500
PA-SRGAN	3.6841/0.0721	2.9013/0.1350	2.3054/0.1335	3.4977/0.1108

In table4, RCAN obtain the best PSNR value on every test dataset, SRResNet obtain the best SSIM value on Set14 and BSD100. SRGAN, SFTGAN, NatSR and our PA-SRGAN use perceptual loss to optimize the model, however, the PSNR and SSIM value of SRGAN, SFTGAN, NatSR and PA-SRGAN is lower than PSNR-driven algorithms, because the definition of PSNR is :

$$PSNR = 10 \times \log_{10}(\frac{(2^n - 1)^2}{MSE}) \tag{14}$$

Eq.14 means that the lower MSE value, the higher PSNR value obtain. PSNR-driven algorithms always use L1 or L2 loss, minimze L1 or L2 loss can increase PSNR value, therefore, the PSNR value of PSNR-driven is higher. SRGAN obtain the lowest PSNR value except Urban100 dataset, because the loss function of SRGAN only consists perceptual loss and adversarial loss, thus, optimize SRGAN can not reduce MSE value and get lower PSNR value.

In table5, the PI and LPIPS value of SRGAN, SFTGAN, NatSR and PA-SRGAN is lower than other PSNR-driven algorithms. Our PA-SRGAN obtain the best PI and LPIPS on every test datasets.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0012

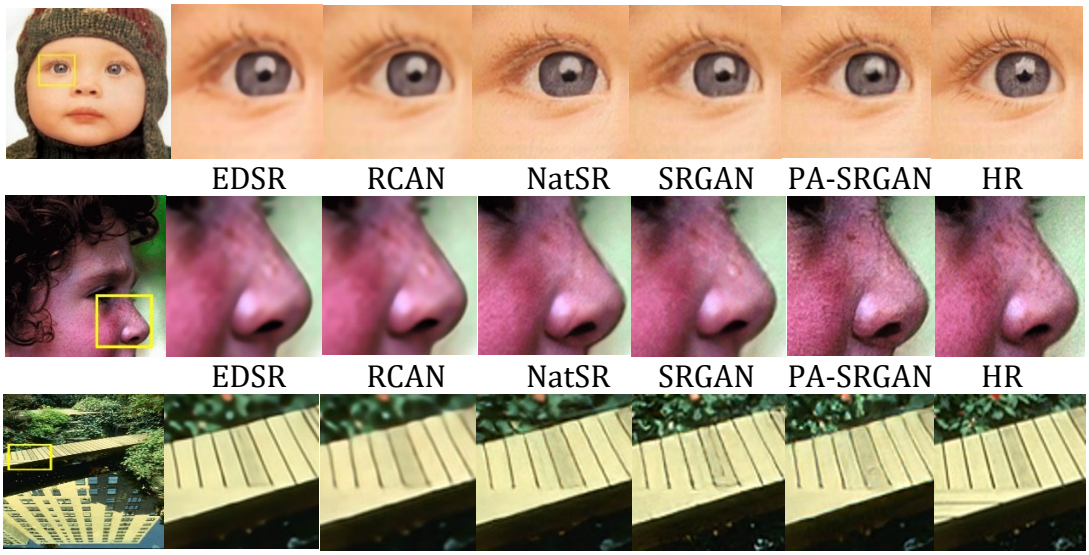


Figure7. Visualize results of some algorithms, the first row is baby.png from Set5, the second row is face.png from Set14 and the last row is 148026.png from BSD100

Figure 7 show the visualize results of PSNR-driven algorithms EDSR and RCAN, generative adversarial networks SRGAN and NatSR. EDSR and RCAN can obtain the better PSNR and SSIM value, however, the visualize results show that EDSR and RCAN reconstructed images are too smooth, the high-frequency details of the image are seriously lost. The PSNR value of SRGAN, NatSR and PA-SRGAN are obviously lower than EDSR and RCAN. However, compare with PSNR-driven algorithms EDSR and RCAN, SRGAN, NatSR and PA-SRGAN can obtain more clearer images, but SRGAN and NatSR generate redundant details while PA-SRGAN generates higher perceptual quality images.

4. CONCULUSION

This paper proposed an enhanced image super-resolution algorithm PA-SRGAN.The generator uses residual dense block to ensure the shallow feature of the LR image can effectively transfer between convolutional layers and pyramid attention module is introduced to improve the ability of obtain multi-scale features. The discriminator uses relativistic average discriminator and spectral normalization convolutional layers to enhance the quality of generate images and the stability of discriminator. The results of experiments show that the proposed methord can obtain the better PI and LPIPS and visualize images.

ACKNOWLEDGMENTS

This paper was financially supported by National Natural Science Foundation of China(No.61961037).

REFERENCES

[1] Tang Yanqiu, Pan Hong, Zhu Yaping, et al. A survey of image super-resolution reconstruction[J].Joural of electronic, 2020, 48(7):1407-1419.

[2] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proc of the Conference on Computer Vision and Pattern Recognition(CVPR). USA:IEEE, 2017:105-114.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202303_9(3).0012

[3] Wang Xintao, Yu Ke, Wu Shixiang, et al. ESRGAN:Enhanced super-resolution generative adversarial network[C]//Proc of the European Conference on Computer Vision(ECCV). Germany: Springer, 2018:63-79.

[4] Wang Xintao, Yu Ke, Dong Chao, et al. Recovering Realistic Texture in Image Super-Resolution by Deep Spatial Feature Transform[C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).USA:IEEE, 2018:606-615.

[5] Soh J W, Park G Y, Jo J, et al. Natural and realistic single image super-resolution with explicit natural maniford discrimination[C]//Proc of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR).USA:IEEE, 2019:8114-8123.

[6] Zhang Yulun, Tian Yapeng, Kong Yu, et al. Residual dense network for image super-resolution[J].arXiv preprint arXiv:1802.08797,2018.

[7] Huang Gao, Liu Zhuang, Laurens van der M, et al. Densely connected convolutional networks [J]. arXiv preprint arXiv:1608.06993, 2018.

[8] Liu Jie, Zhang Wenjie, Tang Yuting, et al. Residual feature aggregation network for image super-resolution[C]//Proc of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA: IEEE, 2020:2359-2368.

[9] Mei Yiqun, Fan Yuchen, Zhang Yulun, et al. Pyramid attention networks for image restoration[J]. arXiv preprint arXiv : 2004.13824, 2020.

[10] Miyato T, Kataoka T, Koyama M, etal. Spectral normalization for generative adversarial networks[J].arXiv preprint arXiv:1802.05957,2018.

[11] Johnson J , Alahi A ,Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[J]. arXiv preprint arXiv:1603.08155, 2016.

[12] Alexia J M. The relativistic discriminator: a key element missing from standard GAN[J].arXiv preprint arXiv:1807.00734v3,2018.