

Bi-GANs-ST for Perceptual Image Super-resolution

Xiaotong Luo¹, Rong Chen¹, Yuan Xie², Yanyun Qu^{*1}, and Cuihua Li¹

¹ School of Information Science and Engineering, Xiamen University, Xiamen, China
528524296@qq.com, chenrong_mail@qq.com, {yyqu,chli}@xmu.edu.cn

² Research Center of Precision Sensing and Control, Institute of Automation,
Chinese Academy of Sciences, Beijing, China
yuan.xie@ia.ac.cn

Abstract. Image quality measurement is a critical problem for image super-resolution (SR) algorithms. Usually, they are evaluated by some well-known objective metrics, *e.g.*, PSNR and SSIM, but these indices cannot provide suitable results in accordance with the perception of human being. Recently, a more reasonable perception measurement has been proposed in [1], which is also adopted by the PIRM-SR 2018 challenge. In this paper, motivated by [1], we aim to generate a high-quality SR result which balances between the two indices, *i.e.*, the perception index and root-mean-square error (RMSE). To do so, we design a new deep SR framework, dubbed Bi-GANs-ST, by integrating two complementary generative adversarial networks (GAN) branches. One is memory residual SRGAN (MR-SRGAN), which emphasizes on improving the *objective* performance, such as reducing the RMSE. The other is weight perception SRGAN (WP-SRGAN), which obtains the result that favors better *subjective* perception via a two-stage adversarial training mechanism. Then, to produce final result with excellent perception scores and RMSE, we use soft-thresholding method to merge the results generated by the two GANs. Our method performs well on the perceptual image super-resolution task of the PIRM 2018 challenge. Experimental results on five benchmarks show that our proposal achieves highly competent performance compared with other state-of-the-art methods.

Keywords: Image super-resolution · Perceptual image · GAN · Soft-thresholding

1 Introduction

Single image super-resolution (SISR) is a hotspot in image restoration. It is an inverse problem which recovers a high-resolution (HR) image from a low-resolution (LR) image via super-resolution (SR) algorithms. Traditional SR algorithms are inferior to deep learning based SR algorithms on speed and some distortion

* Corresponding author

measures, *e.g.*, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). In addition, SR algorithms based on deep learning can also obtain excellent visual effects [2–8].

Here, SR algorithms with deep learning can be divided into two categories. One is built upon convolutional neural network with classic L1 or L2 loss in pixel space as the optimization function, which can gain a higher PSNR but over-smoothness for lacking enough high-frequency texture information. The representative approaches are SRResNet [5] and EDSR [7]. The other is based on generative adversarial networks (GAN), *e.g.*, SRGAN [5] and EnhanceNet [9], which introduces perceptual loss in the optimization function. This kind of algorithms can restore more details and improve visual performance at the expense of objective evaluation indices. Different quality assessment methods are used in various application scenarios. For example, medical imaging may concentrate on objective evaluation metrics, while the subjective visual perception may be more important for natural images. Therefore, we need to make a balance between the objective evaluation criteria and subjective visual effects.

Blau *et al.* [1] proposed perceptual-distortion plane which jointly quantified the accuracy and perceptual quality of algorithms and also pointed GAN can make the perceptual-distortion tradeoff. In the PIRM-SR 2018 challenge [10], image quality is evaluated by root-mean-square error (RMSE) and perceptual index. Inspired by [1], we design a new SR framework for perceptual image super-resolution which includes two GAN branches. First, we redesign the generator network based on SRGAN in each branch and adopt two-stage adversarial training mechanism in the second branch. Then, soft-thresholding method is used to fuse the two results generated by the two branches. Experimental results show our method can obtain excellent distortion measurement and perceptual quality. The contributions of our algorithm are three-fold:

- 1) We propose a new SR framework named Bi-branch GANs with Soft-thresholding (Bi-GANs-ST) for perceptual image super-resolution which consists of two branches. The one is memory residual SRGAN (MR-SRGAN) which emphasizes on improving the objective performance (*e.g.*, reduce the RMSE value). The other is weight perception SRGAN (WP-SRGAN) which focuses on better subjective perception (*e.g.*, reduce the perceptual index).

- 2) In MR-SRGAN, we add memory storage mechanism in Generator which can improve the feature selection ability of the model. To further reduce the RMSE, we train MR-SRGAN by removing the logarithm of adversarial losses. In WP-SRGAN, we use two-stage adversarial training mechanism in which we first optimize pixel-wise loss as a pre-training model for obtaining lower RMSE, then optimize perceptual loss for reducing the perceptual index. And we remove Batch Normalization layers in both networks.

- 3) To keep balance between the perceptual index and RMSE, we fuse the results generated by MR-SRGAN and WP-SRGAN via soft-thresholding method. Our proposal achieves competent performance on the task of the PIRM-SR 2018 challenge.

2 Related Work

Abundant single image super-resolution algorithms based on deep learning have been proposed and achieved remarkable performance. Here, we mainly discuss image SR using deep neural networks, image SR using generative adversarial networks and image quality evaluation.

2.1 Image super-resolution using deep neural networks

Dong *et al.* proposed SRCNN [2], which is a preliminary work to apply convolutional neural network into SISR. Although the network contains only three layers, the performance has been greatly improved compared with the traditional reconstructed methods. FSRCNN [4] is an accelerated version of SRCNN, which introduced a deconvolution layer at the end of the network to perform upsampling for reducing the computational complexity. Shi *et al.* proposed ESPCN [11], which mainly utilized the sub-pixel convolutional layer to accelerate the training process. Kim *et al.* proposed VDSR [3], which used cascaded filters and residual learning to obtain a larger receptive field and accelerate convergence. Kim *et al.* [12] first applied the recursive neural network and skip connection [13] to image SR. RED network [14] was composed of symmetric convolutional layers and deconvolution layers to learn the end-to-end mapping from LR to HR image pairs. Lai *et al.* [6] proposed a cascaded pyramid structure with two branches, one is for feature extraction, the other is for image reconstruction. Moreover, Charbonnier loss was applied to multiple levels and it can generate sub-band residual images at each level. Tong *et al.* [15] introduced dense blocks combining low-level features and high-level features to improve the performance effectively. Lim *et al.* [7] removed Batch Normalization layers in residual blocks (ResBlocks) and adopted residual scaling factor to stabilize network training. Besides, it also proposed multi-scale SR algorithm via a single network. However, when the scaling factor is equal to or larger than $4\times$, the results obtained by the aforementioned methods mostly look smooth and lack enough high-frequency details. The reason is that the optimization targets are mostly based on minimizing L1 or L2 loss in pixel space without considering the high-level features.

2.2 Image super-resolution using generative adversarial networks

Super-resolution with adversarial training. Generative adversarial nets (GANs) [16] consist of Generator and Discriminator. In the task of super-resolution, *e.g.*, SRGAN [5], Generator is used to generate SR images. Discriminator distinguishes whether an image is true or forged. The goal of Generator is to generate a realistic image as much as possible to fool Discriminator. And Discriminator aims to distinguish the ground truth from the generated SR image. Thus, Generator and Discriminator constitute an adversarial game. With adversarial training, the forged data and the real data can eventually obey a similar image statistics distribution. Therefore, adversarial learning in SR is important for recovering the image textural statistics.

Perceptual loss for deep learning. In order to be better accordant with human perception, Johnson *et al.* [17] introduced perceptual loss based on high-level features extracted from pre-trained networks, *e.g.* VGG16, VGG19, for the task of style transfer and SR. Ledig *et al.* [5] proposed SRGAN, which aimed to make the SR images and the ground-truth (GT) similar not only in low-level pixels, but also in high-level features. Therefore, SRGAN can generate realistic images. Sajjadi *et al.* proposed EnhanceNet [9], which applied a similar approach and introduced the local texture matching loss, reducing visually unpleasant artifacts. Zhang *et al.* [18] explained why the perceptual loss based on deep features fits human visual perception well. Mechrez *et al.* proposed contextual loss [19, 20] which was based on the idea of natural image statistics, and it is the best algorithm for recovering perceptual results in previous published works currently. Although these algorithms can obtain better perceptual image quality and visual performance, it cannot achieve better results in terms of objective evaluation criteria.

2.3 Image quality evaluation

There are two ways to evaluate image quality including objective and subjective assessment criteria. The popular objective criteria includes the following: PSNR, SSIM, multi-scale structure similarity index (MSSSIM), information fidelity criterion (IFC), weighted peak signal-to-noise ratio (WPSNR), noise quality measure (NQM) [21] and so on. Although IFC has the highest correlation with perceptual scores for SR evaluation [21], it is not the best criterion to assess the image quality. The subjective assessment is usually scored by human subjects in the previous works [22, 23]. However, there is not a suitable objective evaluation in accordance with the human subjective perception yet. In the PIRM-SR 2018 challenge [10], the assessment of perceptual image quality is proposed which combines the quality measures of Ma [24] and NIQE [25]. The formula of perceptual index is represented as follows,

$$Perceptual\ index = \frac{1}{2}((10 - Ma) + NIQE) \quad (1)$$

Here, a lower perceptual index indicates better perceptual quality.

3 Proposed Methods

We first describe the overall structure of Bi-GANs-ST and then construct the networks MR-SRGAN and WP-SRGAN. The soft thresholding method is used for image fusion, as presented in Section 3.4.

3.1 Basic architecture of Bi-GANs-ST

As shown in Fig. 1, our Bi-GANs-ST mainly consists of three parts: 1) memory residual SRGAN (MR-SRGAN), 2) weight perception SRGAN (WP-SRGAN),

3) soft thresholding (ST). The two GANs are used for generating two complementary SR images, and ST fuses the two SR results for balancing the perceptual score and RMSE.

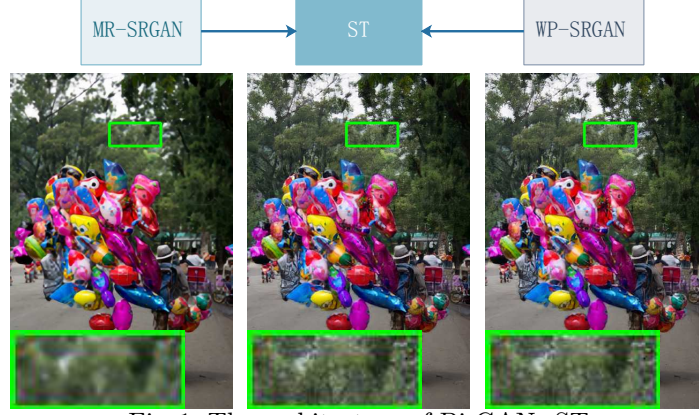


Fig. 1: The architecture of Bi-GANs-ST.

3.2 MR-SRGAN

Network architecture. As illustrated in Fig. 2, our MR-SRGAN is composed of Generator and Discriminator. In Generator, LR images are input to the network followed by one Conv layer for extracting shallow features. Then four memory residual (MR) blocks are applied for improving image quality which help to form persistent memory and improve the feature selection ability of model like MemEDSR [26]. Each MR block consists of four ResBlocks and a gate unit. The former generates four-group features and then we extract a certain amount of features from these features by the gate unit. And the input features are added to the extracted features as the output of MR block. In ResBlocks, all the activation function layers are replaced with parametric rectified linear unit (PReLU) function and all the Batch Normalization (BN) layers are discarded in the generator network for reducing computational complexity. Finally, we restore the original image size by two upsampling operations. n is the corresponding number of feature maps and s denotes the stride for each convolutional layer in Fig. 2 and Fig. 3. In Discriminator, we use the same setting as SRGAN [5].

Loss function. The total generator loss function can be represented as three parts: pixel-wise loss, adversarial loss and perceptual loss, the formulas are as follows,

$$L_{total} = L_{pixel} + \lambda_1 L_{adv} + \lambda_2 L_{vgg} \quad (2)$$

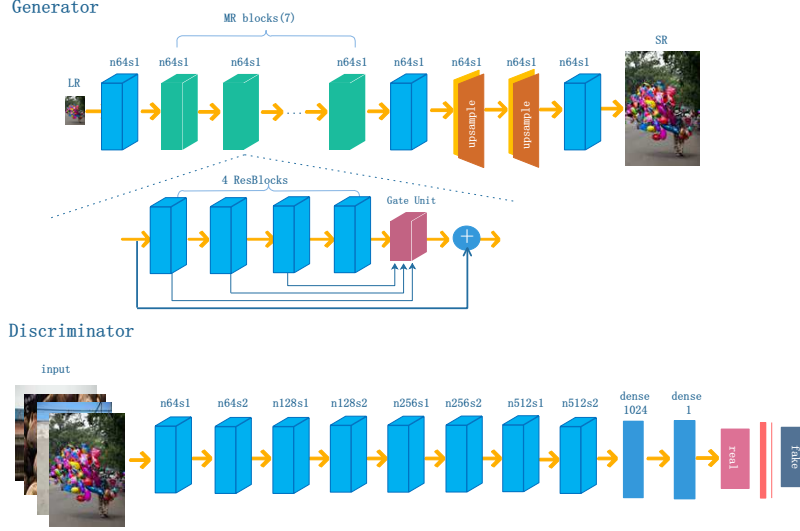


Fig. 2: The architecture of MR-SRGAN.

$$L_{pixel} = \frac{1}{N} \sum_{i=1}^N |x_t^i - G(x_l^i)|^2 \quad (3)$$

$$L_{adv} = -(D(G(x_l))) \quad (4)$$

$$L_{vgg} = \frac{1}{N} \sum_{i=1}^N |\phi(x_t^i) - \phi(G(x_l^i))|^2 \quad (5)$$

where L_{pixel} is the pixel-wise MSE loss between the generated images and the ground truth, L_{vgg} is the perceptual loss which calculates MSE loss between features extracted from the pre-trained VGG16 network, and L_{adv} is the adversarial loss for Generator in which we remove logarithm. λ_1 , λ_2 are the weights of adversarial loss and perceptual loss. x_t , x_l denote the ground truth and LR images, respectively. $G(x_l)$ is the SR images forged by Generator. N represents the number of training samples. ϕ represents the features extracted from pre-trained VGG16 network.

3.3 WP-SRGAN

Network architecture. In WP-SRGAN, we use 16 ResBlocks in the generator network which is depicted in Fig. 3. Each ResBlock is consisted of convolutional layer, PReLU activation layer and convolutional layer. And Batch Normalization (BN) layers are removed in both Generator and Discriminator. The architecture

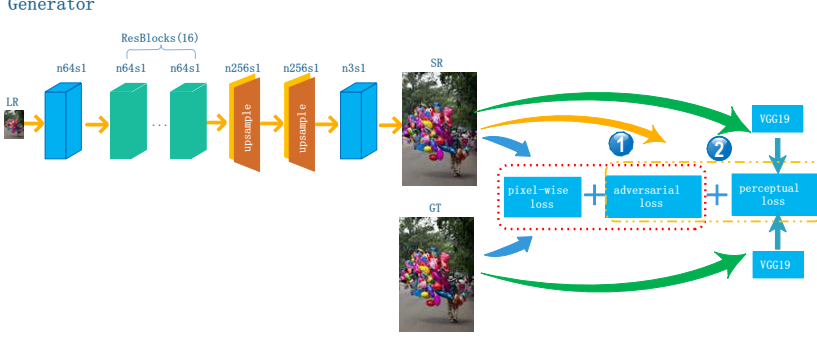


Fig. 3: The generator architecture of WP-SRGAN. The red box shows the generator loss in the first training stage, then the orange box contains the generator loss in the second training stage.

of Discriminator in WP-SRGAN is the same as MR-SRGAN except for removing BN layers.

Loss function. As shown in Fig. 3, a two-stage bias adversarial training mechanism is adopted in WP-SRGAN by using different Generator losses. In the first stage, as the red box shows, we optimize the Generator loss which is consisted of pixel-wise loss and adversarial loss to obtain better objective performance (*i.e.*, reduce the RMSE value). In the second stage, as the orange box shows, we regard the network parameters in the first stage as the pre-trained model and then replace the aforementioned generator loss with perceptual loss and adversarial loss to optimize for improving the subjective visual effects (*e.g.*, reduce the perceptual index). The two-stage losses are represented as Eq. (6) and Eq. (7).

$$L_1 = L_{pixel} + \lambda_1 L_{adv} \quad (6)$$

$$L_2 = \lambda_1 L_{adv} + \lambda_2 L_{vgg} \quad (7)$$

Here, the pixel-wise loss is defined as the Eq. (3), the perceptual loss adopts MSE loss by the features extracted from pre-trained VGG19 network, and the adversarial loss is donated as follows,

$$L_{adv} = -\log(D(G(x_l))) \quad (8)$$

By adopting two-stage adversarial training mechanism, it can make the generated SR image similar to the corresponding ground truth in high-level features space.

3.4 Soft-thresholding

We can obtain different SR results by the two GANs aforementioned. One is MR-SRGAN, which emphasizes on improving the objective performance. The other

is WP-SRGAN, which obtains the result that favors better subjective perception. To balance the perceptual score and RMSE of SR results, soft thresholding method proposed by Deng *et al.* [27] is adopted to fuse the two SR images (*i.e.*, MR-SRGAN, WP-SRGAN) which can be regarded as a way of pixel interpolation. The formulas are shown as follows,

$$I_e = I_G + \text{soft}(\Delta, \xi) \quad (9)$$

$$\text{soft}(\Delta, \xi) = \text{sign}(\Delta) \cdot \max(|\Delta| - \xi, 0) \quad (10)$$

where I_e is the fused image, $\Delta = I_G - I_g$, I_G is the generated image by WP-SRGAN whose perceptual score is lower, I_g is the generated image by MR-SRGAN whose RMSE value is lower. ξ is the adjusted threshold which is discussed in Section 4.2.

4 Experimental Results

In this section, we conduct extensive experiments on five publicly available benchmarks for scaling factor $4\times$ image SR: Set5 [28], Set14 [29], B100 [30], Urban100 [31], Manga109 [32], separately. The first three datasets Set5, Set14, BSD100 mainly contain natural images, Urban100 consists of 100 urban images, and Manga109 is Japanese anime containing fewer texture features. Then we compare the performance of our proposed Bi-GANs-ST algorithm with the state-of-the-art SR algorithms in terms of objective criteria and subjective visual perception.

4.1 Implementation and training details

We train our networks using the RAISE³ dataset which consists of 8156 HR RAW images. The HR images are downsampling by bicubic interpolation method for the scaling factor $4\times$ to obtain the LR images. To analyze our models capacity, we evaluate them on the PIRM-SR 2018 self validation dataset [10] which consists of 100 realistic images including human, plants, animals and so on.

The LR-HR image patches for training are randomly cropped from the corresponding LR and HR image pairs. The crop size for LR patches is 24×24 , and the size of corresponding HR patches is 96×96 . Random flipping is used for image argumentation. The batch size is set to 16.

In our experiments, MP-SRGAN is conducted on the deep learning framework, *i.e.*, Pytorch and WP-SRGAN is conducted on Tensorflow. We implement our method on the platform Ubuntu 16.04, CUDA8.0 and CUDNN6.0 with GTX1080 GPU and 32G CPU Memory.

In Generator of MR-SRGAN, 7 MR blocks are used. The filter size is set to 3×3 . The learning rate is initialized to $1e - 4$ and Adam optimizer with the

³ <http://loki.disi.unitn.it/RAISE/>

momentum 0.9 is utilized. The network is trained for 600 epochs, and we choose the best results according to the metric SSIM.

In Generator of WP-SRGAN, 16 ResBlocks are used and the filter size is 3×3 . The filter size is 9×9 in the first and last convolutional layer. All the convolutional layers use one stride and one padding. The weights are initialized by Xavier method. All the convolutional and upsampling layers are followed by PReLU activation function. The learning rate is initialized to $1e-4$ and decreased by a factor of 10 for 2.5×10^5 iterations and total iterations are 5×10^5 . We use Adam optimizer with momentum 0.9. In Discriminator, the filter size is 3×3 , and the number of features is twice increased from 64 to 512, the stride is one or two, alternately.

The weights of adversarial loss and perceptual loss both in MP-SRGAN and WP-SRGAN (*i.e.*, λ_1 and λ_2) are set to $1e-3$, $6e-3$, respectively. And the threshold (*i.e.*, ξ) for image fusion is set to 0.73 in our experiment.

Table 1: The quantitative results of WP-SRGAN with two stages on the PRIM 2018 self validation dataset for 4 \times enlargement.

Model	Perceptual score/RMSE)
First stage	5.2002 / 14.385
Second stage	2.0815 / 16.2813

4.2 Model analysis

Training WP-SRGAN with Two stages. We analyze the experimental results of the two-stage adversarial training mechanism in WP-SRGAN. The quantitative and qualitative results on PIRM-SR 2018 self validation dataset are shown in Table 1 and Fig. 4.

In Table 1, WP-SRGAN with two stages can achieve lower perceptual score than WP-SRGAN with the first stage. As shown in Fig. 4, the recovered details of WP-SRGAN with two stages are much more than WP-SRGAN with the first stage. And the images generated by two stages look more realistic. Therefore, we use WP-SRGAN with two stages in our model.

Soft thresholding. In the challenge, three regions are defined by RMSE between 11.5 and 16. According to different threshold settings, we draw the perceptual-distortion plane which is shown in Fig. 5, according to the results fused by Eq. (9) and (10). The points on the curve denote the different thresholds from 0 to 1 with an interval of 0.1. Experimental results show that we can obtain excellent perceptual score in Region3 (RMSE is between 12.5 and 16) when ξ is set to 0.73.

Model capacity. To demonstrate the capability of our models, we analyze the SR results of MR-SRGAN, WP-SRGAN and Bi-GANs-ST for the metrics per-

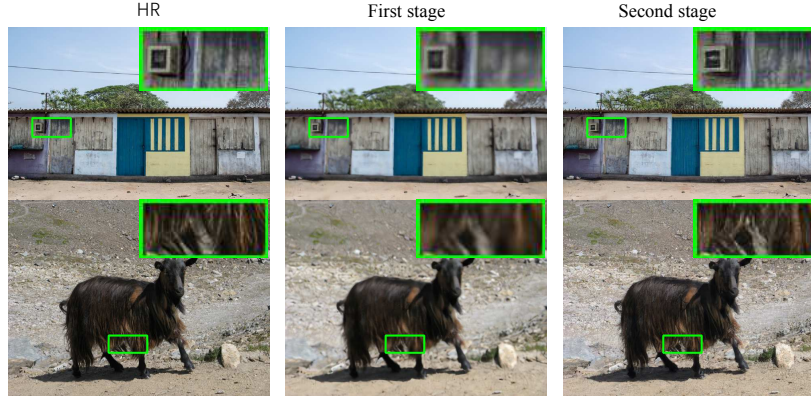


Fig. 4: The visual results of WP-SRGAN with two stages on the PRIM 2018 self validation dataset for $4\times$ enlargement.

ceptual score and RMSE on the PIRM-SR 2018 self validation dataset. The quantitative and qualitative results are shown in Table 2 and Fig. 6. The experimental results show that Bi-GANs-ST can keep balance between the perceptual score and RMSE.

Table 2: The model capacity analysis of the SR results by MR-SRGAN, WP-SRGAN and Bi-GANs-ST for the metrics perceptual score and RMSE on the PIRM-SR 2018 self validation dataset.

Model	Perceptual score/RMSE
MR-SRGAN	4.404 / 11.36
WP-SRGAN	2.082 / 16.28
Bi-GANs-ST	2.139 / 15.77

4.3 Comparison with the state-of-the-arts

To verificate the validity of our Bi-GANs-ST, we conduct extensive experiments on five publicly available benchmarks and compare the results with other state-of-the-art SR algorithms, including EDSR[7], EnhanceNet[9]. We use the open-source implementations for the two comparison methods. We evaluate the SR images with image quality assessment indices (*i.e.*, PSNR, SSIM, perceptual score, RMSE) where PSNR and SSIM are measured on the y channel and ignored 6 pixels from the border.

The quantitative results for evaluating PSNR and SSIM are shown in Table 3. The best algorithm is EDSR, which is on average $1.0dB$, $0.54dB$, $0.34dB$, $0.83dB$

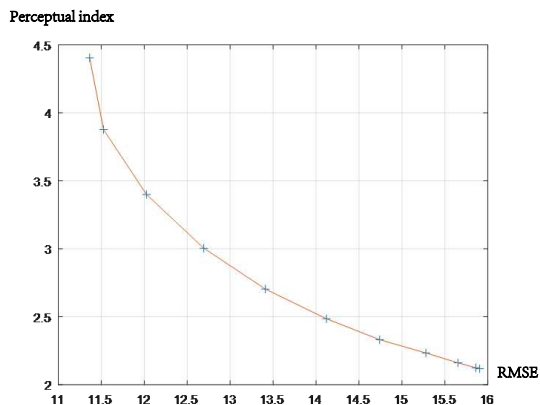


Fig. 5: The perceptual-distortion plane of our method. The points on the curve denote the different thresholds from 0 to 1 with an interval of 0.1.

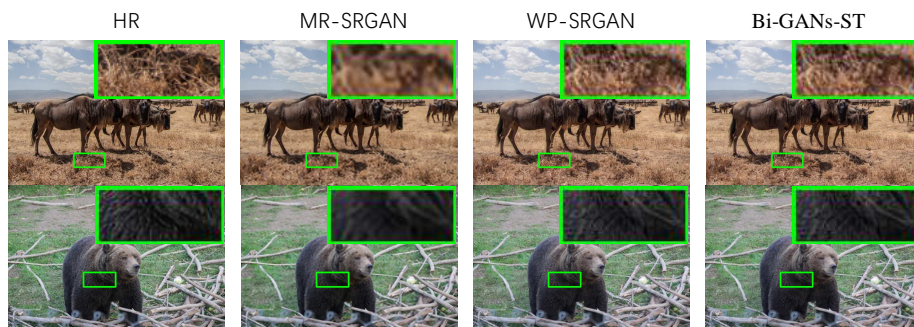


Fig. 6: The visual results on three models (MR-SRGAN, WP-SRGAN and Bi-GANs-ST) for scaling factor $4\times$ for the metrics perceptual score and RMSE on the PIRM-SR 2018 self validation dataset.

and $1.13dB$ higher than our MR-SRGAN. The PSNR values of our Bi-GANs-ST are higher than EnhanceNet on Set5, Urban100, Manga109 approximately $0.64dB$, $0.3dB$, $0.13dB$, respectively. The SSIM values of our Bi-GANs-ST are all higher than EnhanceNet. Table 4 shows the quantitative evaluation of average perceptual score and RMSE. For perceptual score index, our WP-SRGAN achieves the best and Bi-GANs-ST achieves the second best on five benchmarks except for Set5. For RMSE index, EDSR performs the best and our MR-SRGAN performs the second best.

The visual perception results of $4\times$ enlargement of different algorithms on five benchmarks are shown in Fig. 7. These visual results are produced by Bicubic, EDSR, EnhanceNet, MR-SRGAN, WP-SRGAN, Bi-GANs-ST and the ground truth from left to right. EDSR can generate the images which look clear and smooth but not realistic. The SR images of our MR-SRGAN algorithm are like



Table 3: Quantitative evaluation of state-of-the-art SR algorithms on five publicly available benchmarks: average PSNR/SSIM for scaling factor $4\times$ (Red text indicates the best and blue text indicates the second best performance).

Algorithm	Set5	Set14	BSD100	Urban100	Manga109
Bicubic	24.74/0.736	23.47/0.630	23.93/0.602	21.15/0.583	21.82/0.711
EDSR[7]	32.53/0.899	28.82/0.786	27.64/0.740	26.62/0.802	30.95/0.914
MR-SRGAN(ours)	31.53/0.884	28.28/0.771	27.30/0.725	25.79/0.774	29.82/0.895
EnhanceNet[9]	28.90/0.818	26.04/0.685	25.19/0.634	23.60/0.691	26.71/0.827
WP-SRGAN(ours)	29.06/0.834	26.01/0.703	24.38/0.642	23.74/0.700	26.62/0.836
Bi-GANs-ST(ours)	29.54/0.840	26.01/0.706	24.54/0.651	23.90/0.703	26.84/0.839

Table 4: Quantitative evaluation of state-of-the-art SR algorithms on five publicly available benchmarks: average perceptual scores/RMSE for scale $4\times$ (Red text indicates the best and blue text indicates the second best performance).

Algorithm	Set5	Set14	BSD100	Urban100	Manga109
EDSR[7]	5.944/ 6.51	5.256/ 10.92	5.263/ 12.38	4.981/ 14.52	4.714/ 8.63
EnhanceNet[9]	3.141 /9.94	3.009/14.82	2.979/16.14	3.401/19.49	3.259/13.19
MR-SRGAN(ours)	5.304/ 7.44	4.703/ 11.41	4.995/ 12.72	4.466/ 15.64	4.164/ 9.57
WP-SRGAN(ours)	3.317 /9.80	2.824 /14.46	2.226 /18.19	3.260 /19.04	3.195 /13.32
Bi-GANs-ST(ours)	3.531/9.14	2.869 /14.22	2.375 /17.56	3.272 /18.75	3.216 /12.97

to EDSR. EnhanceNet can generate more realistic images with unpleasant noises. The SR images of our WP-SRGAN algorithm obtain more details like EnhanceNet with less noises which are more close to the ground-truth. And our Bi-GANs-ST algorithm has fewer noises than WP-SRGAN.

5 Conclusions

In this paper, we propose a new deep SR framework Bi-GANs-ST by integrating two complementary generative adversarial networks (GAN) branches. To keep better balance between the perceptual score and RMSE of generated images, we redesign two GANs (*i.e.*, MR-SRGAN, WP-SRGAN) to generate two complementary SR results based on SRGAN. Last, we use soft-thresholding method to fuse two SR results which can make the perceptual score and RMSE tradeoff. Experimental results on five publicly benchmarks show that our proposed algorithm can perform better perceptual results than other SR algorithms for $4\times$ enlargement.

Acknowledgements. This work is supported by the National Natural Science Foundation of China under Grant 61876161, Grant 61772524, Grant 61373077 and in part by the Beijing Natural Science Foundation under Grant 4182067.

References

1. Blau, Y., Michaeli, T.: The perception-distortion tradeoff. arXiv preprint arXiv:1711.06077 (2017)
2. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence* **38**(2) (2016) 295–307
3. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2016) 1646–1654
4. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: *European Conference on Computer Vision*, Springer (2016) 391–407
5. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A.P., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: *CVPR*. Volume 2. (2017) 4
6. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep laplacian pyramid networks for fast and accurate superresolution. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Volume 2. (2017) 5
7. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: *The IEEE conference on computer vision and pattern recognition (CVPR) workshops*. Volume 1. (2017) 4
8. Haris, M., Shakhnarovich, G., Ukita, N.: Deep backprojection networks for super-resolution. In: *Conference on Computer Vision and Pattern Recognition*. (2018)
9. Sajjadi, M.S., Schölkopf, B., Hirsch, M.: Enhancenet: Single image super-resolution through automated texture synthesis. In: *Computer Vision (ICCV), 2017 IEEE International Conference on, IEEE* (2017) 4501–4510
10. Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L.: 2018 pirm challenge on perceptual image super-resolution. In: arXiv preprint arXiv:1809.07517. (2018)
11. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2016) 1874–1883
12. Kim, J., Kwon Lee, J., Mu Lee, K.: Deeply-recursive convolutional network for image super-resolution. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2016) 1637–1645
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2016) 770–778
14. Mao, X., Shen, C., Yang, Y.B.: Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: *Advances in neural information processing systems*. (2016) 2802–2810
15. Tong, T., Li, G., Liu, X., Gao, Q.: Image super-resolution using dense skip connections. In: *Computer Vision (ICCV), 2017 IEEE International Conference on, IEEE* (2017) 4809–4817
16. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: *Advances in neural information processing systems*. (2014) 2672–2680

17. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: *European Conference on Computer Vision*, Springer (2016) 694–711
18. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. *arXiv preprint* (2018)
19. Mechrez, R., Talmi, I., Zelnik-Manor, L.: The contextual loss for image transformation with non-aligned data. *arXiv preprint arXiv:1803.02077* (2018)
20. Mechrez, R., Talmi, I., Shama, F., Zelnik-Manor, L.: Learning to maintain natural image statistics. *arXiv preprint arXiv:1803.04626* (2018)
21. Yang, C.Y., Ma, C., Yang, M.H.: Single-image super-resolution: A benchmark. In: *European Conference on Computer Vision*, Springer (2014) 372–386
22. Moorthy, A.K., Bovik, A.C.: Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE transactions on Image Processing* **20**(12) (2011) 3350–3364
23. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing* **21**(12) (2012) 4695–4708
24. Ma, C., Yang, C.Y., Yang, X., Yang, M.H.: Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding* **158** (2017) 1–16
25. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a” completely blind” image quality analyzer. *IEEE Signal Process. Lett.* **20**(3) (2013) 209–212
26. Chen, R., Qu, Y., Zeng, K., Guo, J., Li, C., Xie, Y.: Persistent memory residual network for single image super resolution. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. Volume 6. (2018)
27. Deng, X.: Enhancing image quality via style transfer for single image super-resolution. *IEEE Signal Processing Letters* **25**(4) (2018) 571–575
28. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. (2012)
29. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: *International conference on curves and surfaces*, Springer (2010) 711–730
30. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence* **33**(5) (2011) 898–916
31. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2015) 5197–5206
32. Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., Aizawa, K.: Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications* **76**(20) (2017) 21811–21838