RankSRGAN: Generative Adversarial Networks with Ranker for Image Super-Resolution Supplementary File

Wenlong Zhang¹ Yihao Liu^{1,2} Chao Dong^{1,†} Yu Qiao¹ ¹ShenZhen Key Lab of Computer Vision and Pattern Recognition, SIAT-SenseTime Joint Lab, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China ²University of Chinese Academy of Sciences

{wl.zhang1, yh.liu4, chao.dong, yu.qiao}@siat.ac.cn

Abstract

In this supplementary file, we first present more details and additional experimental results of our proposed Ranker. Then, we provide the curves showing the performance of different RankSRGAN models in the ablation study. Finally, we provide more additional qualitative results to compare our networks with the state-of-the-art methods.

1. Details of Ranker

1.1. Dataset



Figure 1. The convergence curves of RankSRGAN with Ranker1 and Ranker2 in NIQE and PSNR.

Method	Ranker	Dataset	Data Size (k)	SROCC	NIQE	PSNR
RankSRGAN-R-D1	Ranker1	DIV2K	15	0.78	2.53	24.54
RankSRGAN-R-D2	Ranker2	DIV2K+Flickr2K	150	0.88	2.51	25.62

Table 1. The performance of RankSRGAN with Ranker1 and Ranker2. R-D1: rank dataset1 (15 K), R-D2: rank dataset2 (150 K)

To analyze the effects of Ranker on RankSRGAN, we use SRResNet [3], SRGAN [3] and ESRGAN [9] to generate two rank datasets with different sizes. We first employ DIV2K [1] to generate rank dataset1 with **15 K** image pairs. Besides, we use DIV2K+Flickr2k [1] to generate rank dataset2 with **150 K** image pairs. Then, we utilize rank dataset1 and rank dataset2 to train Ranker1 and Ranker2, respectively. Finally, the well-trained Ranker1 and Ranker2 are applied on RankSRGAN. Table 1 shows that more data leads to better SROCC, and the Ranker2 with higher SROCC can reach better performance in NIQE and PSNR. The convergence curves are shown in Figure 1.

[†]Corresponding author (e-mail: chao.dong@siat.ac.cn)

1.2. Network Architecture

The architecture of Ranker is based on the VGG network [8]. We train three VGG networks varying from shallow to deep ones: VGG-8, VGG-12 and VGG-16. Table 2 shows the architecture, the number of parameters, and the performance in different models. Since the VGG-12 can achieve the same accuracy as VGG-16, we apply the VGG-8 and VGG-12 on RankSRGAN. Figure 2 shows the performance of RankSRGAN with different Rankers. The Ranker with higher value of SROCC can achieve better performance when applied on RankSRGAN.

Model	VGG-8	VGG-12 (Ours)	VGG-16				
Architecture	Conv3S1-64	Conv3S1-64	Conv3S1-64, LReLU				
	Conv4S2-64, BN, LReLU	Conv4S2-64, BN, LReLU	Conv4S2-64, BN, LReLU				
			Conv3S1-128, BN, LReLU				
		Conv3S1-128, BN, LReLU	Conv3S1-128, BN, LReLU				
	Conv4S2-128, BN, LReLU	Conv4S2-128, BN, LReLU	Conv4S2-128, BN, LReLU				
			Conv3S1-256, BN, LReLU				
		Conv3S1-256, BN, LReLU	Conv3S1-256, BN, LReLU				
	Conv4S2-256, BN, LReLU	Conv4S2-256, BN, LReLU	Conv4S2-256, BN, LReLU				
			Conv3S1-512, BN, LReLU				
		Conv3S1-512, BN, LReLU	Conv3S1-512, BN, LReLU				
	Conv4S2-512, BN, LReLU	Conv4S2-512, BN, LReLU	Conv4S2-512, BN, LReLU				
			Conv3S1-512, BN, LReLU				
		Conv3S1-512, BN, LReLU	Conv3S1-512, BN, LReLU				
	Conv4S2-512, BN, LReLU	Conv4S2-512, BN, LReLU	Conv4S2-512, BN, LReLU				
	Average pooling						
	FC-100						
	FC-1						
Number of params (K)	7,069	13,734	19,194				
SROCC	0.83	0.88	0.88				

Table 2. The network architecture of Rankers with different depths. The network design draws inspiration from VGG [8] but uses Leaky ReLU activations [5] and strided convolutions instead of pooling layers [7]. Conv3S1-64: Convolutional layer with kernel size 3×3 , stride 1 and channel 64. BN: Batch Normalization. LReLU: Leaky ReLU.



Figure 2. The convergence curves of RankSRGAN with Ranker-VGG-8 and Ranker-VGG-12 in NIQE and PSNR.

1.3. Output Distribution

In Table 5 and 6 of the main paper, we quantitatively evaluate the effects of Ranker and "regression" Ranker on RankSR-GAN. To better understand the effects, we provide the histograms of NIQE [6]/Ma [4] label value in the validation dataset of the rank dataset. Furthermore, we plot the histograms of the output scores of different Rankers ("regression" Ranker and our Ranker) in Figure 3. Comparing Figure 3 (b) and (c), Ranker successfully enlarges the distance between SRGAN and ESRGAN. The "regression" Ranker tends to learn the distribution of the NIQE label, while the NIQE values of SRGAN are close ESRGAN. The same observation is also found in Ma metric as shown in Figure 4.



Figure 3. The histograms of (a) NIQE label value, (b) the regression score of "regression" Ranker and (c) the ranking score of Ranker (ours). These graphs illustrate that Ranker (ours) successfully managed to separate the different perceptual levels. (Better view in color version)



Figure 4. The histograms of (a) Ma label value, (b) the regression score of "regression" Ranker and (c) the ranking score of Ranker (ours). These graphs illustrate that Ranker (ours) successfully managed to separate the different perceptual levels.

2. Details of RankSRGAN

2.1. Convergence curves for RankSRGAN-(NIQE,Ma, and PI)

As mentioned in the main paper, Ranker can guide the SR model to be optimized in the direction of perceptual metrics. We further present the curves showing that our RankSRGAN can achieve a constant improvement compared with the baseline SRGAN. We provide the curves of RankSRGAN-N, RankSRGAN-M, and RankSRGAN-PI (N: Ranker with NIQE [6], M: Ranker with Ma [4], and PI: Ranker with PI [2]) in Figure 5. Besides, we add the curves of RankSRGAN-N-re and RankSRGAN-M-re (re: "regression" Ranker). We observe that Ranker could help RankSRGAN achieve state-of-the-art performance in the chosen metric. This shows that our method can generalize well on different perceptual metrics. Compared with "regression" Ranker, Ranker can accelerate the convergence of RankSRGAN-N. For RankSRGAN-M, Ranker can still reach state-of-the-art performance (less than 1.40 in ESRGAN), while the "regression" Ranker cannot outperform ESRGAN (1.40).



Figure 5. The convergence curves of RankSRGAN_N, RankSRGAN_M and RankSRGAN_{PI}.

2.2. Convergence curves for RankSRGAN-HR

As shown in Figure 6, we present the curves of RankSRGAN and RankSRGAN-HR. To improve the performance of NIQE evaluation, we use (SRResNet, SRGAN, ESRGAN) to generate rank dataset to train Ranker in RankSRGAN. Figure 6 shows that RankSRGAN is consistently better than SRGAN by a large margin. Furthermore, we directly use the ground truth HR to replace ESRGAN . We train our Ranker with the rank dataset (SRResNet, SRGAN, HR) and obtain the new model RankSRGAN-HR. In Figure 6, RankSRGAN-HR achieves better NIQE values than SRGAN. But at the same time, RankSRGAN-HR also constantly improves the PSNR. It achieves a good balance between the perceptual metric and PSNR.



Figure 6. The convergence curves of RankSRGAN-HR in PSNR and NIQE.

3. More Qualitative Results

In this section, we provide additional qualitative results (\times 4 enlargement) to clearly show the effectiveness of our RankSR-GAN. We compare the proposed RankSRGAN with the state-of-the-art perceptual SR methods SRGAN [3] / ESRGAN [9] and PSNR-oriented method SRResNet [3]. We employ NIQE and PSNR to evaluate those SR methods. Lower NIQE value indicates better perceptual quality while higher PSNR indicates that there is less distortion with the Ground-Truth image.

References

- [1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, volume 3, page 2, 2017.
- [2] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. 2018 pirm challenge on perceptual image super-resolution. arXiv preprint arXiv:1809.07517, 2018.
- [3] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, volume 2, page 4, 2017.
- [4] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image superresolution. *Computer Vision and Image Understanding*, 158:1–16, 2017.
- [5] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In Proc. icml, volume 30, page 3, 2013.
- [6] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a" completely blind" image quality analyzer. *IEEE Signal Process*. *Lett.*, 20(3):209–212, 2013.
- [7] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *Computer Science*, 2015.
- [8] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556, 2014.
- [9] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced superresolution generative adversarial networks. In *The European Conference on Computer Vision (ECCV) Workshops*, September 2018.



Img 210

SRGAN (2.70/23.49)

ESRGAN (2.82/23.12) RankSRGAN (Ours) (**2.66**/23.42)



