Hierarchical Generative Adversarial Networks for Single Image Super-Resolution
– *Supplementary Material* –

Weimin Chen[1]*, Yuqing Ma[2]*, Xianglong Liu[2]†, Yi Yuan[1]
[1]NetEase Fuxi AI Lab, Hangzhou, China
[2]State Key Lab of Software Development Environment, Beihang University, China
{chenweimin, yuanyi}@corp.netease.com, {mayuqing, xlliu}@nlsde.buaa.edu.cn

Abstract

In this supplementary material, we provide additional information of our proposed hierarchical generative adversarial networks for single image super-resolution (HSRGAN) and discuss architecture, training, further implementation details and discussions with other multi-scale SR methods. Furthermore, we provide complexity analysis and additional visual experimental results compared with the state-of-the-art CNN-based single image super-resolution (SISR) methods.

1. HSRGAN details

In this section, we show the architecture and training details of our HSRGAN.

1.1. Upsample

There are a lot of ways to increase the resolution of feature maps. While deconvolution (or convolution transpose) may cause checherboard artifacts [9], more recent works use nearest neighbor interpolation followed by a convolution layer or pixel shuffle (or sub-pixel convolution) [11] layer for upsampling. In our HSRGAN, We employ pixel shuffle to amplify the resolution, which is more suitable for SISR problem compared to the deconvolution layer [1]. Pixel shuffle layer has a more flexible way to model the mapping between LR space and HR space, which aggregates feature maps in low dimensional space and a separate upscaling kernel is used to map each feature map to high dimensional space to reconstruct the HR image. In order to guarantee the performance stability of network and generate more realistic SR images, we adopt two pixel shuffle layers to enlarge the feature map by a factor of 2 for each. The comparison among the three upsampling method shown as Table 1 demonstrates that pixel shuffle outperforms others in perceptual quality in our model.

Upsample	PI / PSNR
deconvolution	3.171 / 26.433
conv(nearest)	2.958 / 26.182
pixel shuffle	2.897 / 26.239

Table 1. Comparison among the three upsampling ways on Set14.

1.2. Discriminator

According to [2], $\mathbb{D}_{\text{Ra}}$ denoted in our main paper can be represented by

$$\mathbb{D}_{\text{Ra}}\left(x_r, x_f\right) = \sigma\left(C\left(x_r\right) - \mathrm{E}\left[C\left(x_f\right)\right]\right), \tag{1}$$

where σ is the sigmoid function, $C\left(x\right)$ is the non-transformed discriminator output and $\mathrm{E}\left[\cdot\right]$ is the operation of taking average for all fake data in the mini-batch. In our HSRGAN, x_r is the groundtruth image $\mathbf{I}_{\text{HR}}$ and x_f is the synthetic super-resolved image $\mathbf{I}_{\text{SR}}$.

In our experiment, VGG13 [12] was deployed as the backbone of our discriminator, as other works [6, 17] did. Since max pooling operation may lose some information during feed-forward process, we instead use the convolution kernel with stride set to 2 to downsample the feature maps.

1.3. Training details

Our model was implemented on the PyTorch framework and trained on an NVIDIA GeForce RTX 2080Ti GPU. We empirically chose the hyper-parameters $\lambda = 5 \times 10^{-3}$ and $\eta = 1 \times 10^{-2}$, the the initial learning rate was 10^{-4}. The models were trained with a batch size of 16 with the learning rate reducing to half every 200k iterations.

To accelerate the training process, we first trained a distortion-oriented with the L1 loss and then finetuned the

*The first two authors contributed equally.
†Corresponding author

Methods	Parameters
Bicubic	-
EDSR [8]	43.0M
RCAN [19]	16.0M
SRFBN [7]	3.6M
SRGAN [6]	1.5M
SFTGAN [16]	1.7M
NatSR [13]	4.8M
ESRGAN [17]	15.9M
HSRGAN (Ours)	8.3M

Table 2. Model complexity comparison of our HSRGAN and state-of-the-art methods.

pre-trained model with the overall loss defined in our main paper. Pre-training with pixel-wise loss helps GAN-based methods to obtain more visually pleasing results [17].

1.4. Discussion

Although many of the previous SISR papers leverage the idea of multi-scale [3, 14, 15, 4, 5], they usually use a single-path or recursive network as backbone and pass the shallow feature map directly to the upsampling layer through skip connections to realize multi-scale feature extraction. However, we get rid of complicated convolution blocks and utilize three independent path to extract multi-scale features, which is simple and effective. In addition, we employ HGRM, which amplifies the feature map step by step to stablize the reconstruction procedure and the whole network is trained jointly.

2. Experiments

2.1. Complexity analysis

In this section, we discuss the complexity of our proposed model. In general, for the model without recursive connection, the number of parameters of the model is positively correlated with operations [18, 10] and inference time. As shown in Table 2 and Table 2 in main paper, in comparison with other state-of-the-art networks, especially those with a large number of parameters, such as ESRGAN and EDSR, our proposed HSRGAN can achieve competitive visual results, while only needs the 52% and 19% parameters of ESRGAN and EDSR, respectively. This demonstrates our method can well balance the number of parameters and the reconstruction performance.

2.2. Comparison with the state-of-the-art

In this section, we employ Bicubic, EDSR [8], RCAN [19], SRFBN [7], SRGAN [6], SFTGAN [16], NatSR [13], ESRGAN [8] as our comparison methods. We retrained these models with their published codes and run them on the test datasets. The methods can be divided into 3 categories, where Bicubic is a baseline for the others. Compared to distortion-oriented methods, such as EDSR, RCAN and SRFBN, GAN-based or perception-oriented methods generate clear edges of images to some extent. Among all the GAN-based methods, our HSRGAN outperforms the other methods on denoising, details recovery and texture reality, as shown in Figure 1 to Figure 4.

References

[1] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer, 2016. 1

[2] Alexia Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard gan. *arXiv preprint arXiv:1807.00734*, 2018. 1

[3] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016. 2

[4] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 624–632, 2017. 2

[5] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. 2

[6] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 1, 2

[7] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3867–3876, 2019. 2

[8] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 2

[9] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and checkerboard artifacts. *Distill*, 1(10):e3, 2016. 1

[10] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018. 2

[11] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. 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*, pages 1874–1883, 2016. 1

[12] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1

[13] Jae Woong Soh, Gu Yong Park, Junho Jo, and Nam Ik Cho. Natural and realistic single image super-resolution with explicit natural manifold discrimination. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2

[14] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 3147–3155, 2017. 2

[15] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *Proceedings of the IEEE international conference on computer vision*, pages 4539–4547, 2017. 2

[16] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 606–615, 2018. 2

[17] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018. 1, 2

[18] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 6848–6856, 2018. 2

[19] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 286–301, 2018. 2

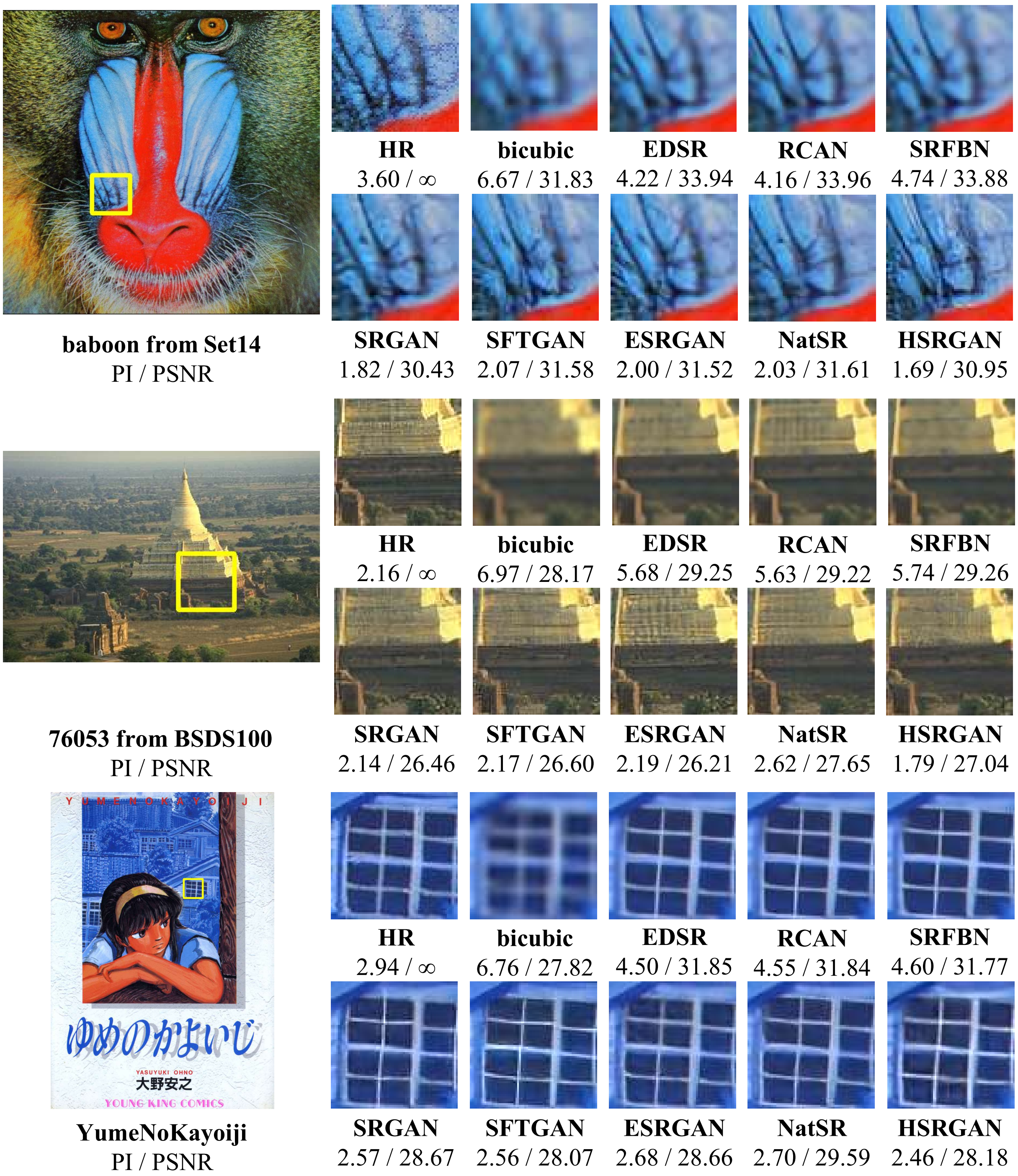

Figure 1. Qualitative results of our HSRGAN and state-of-the-art methods. HSRGAN generates more realistic textures and less artifacts.

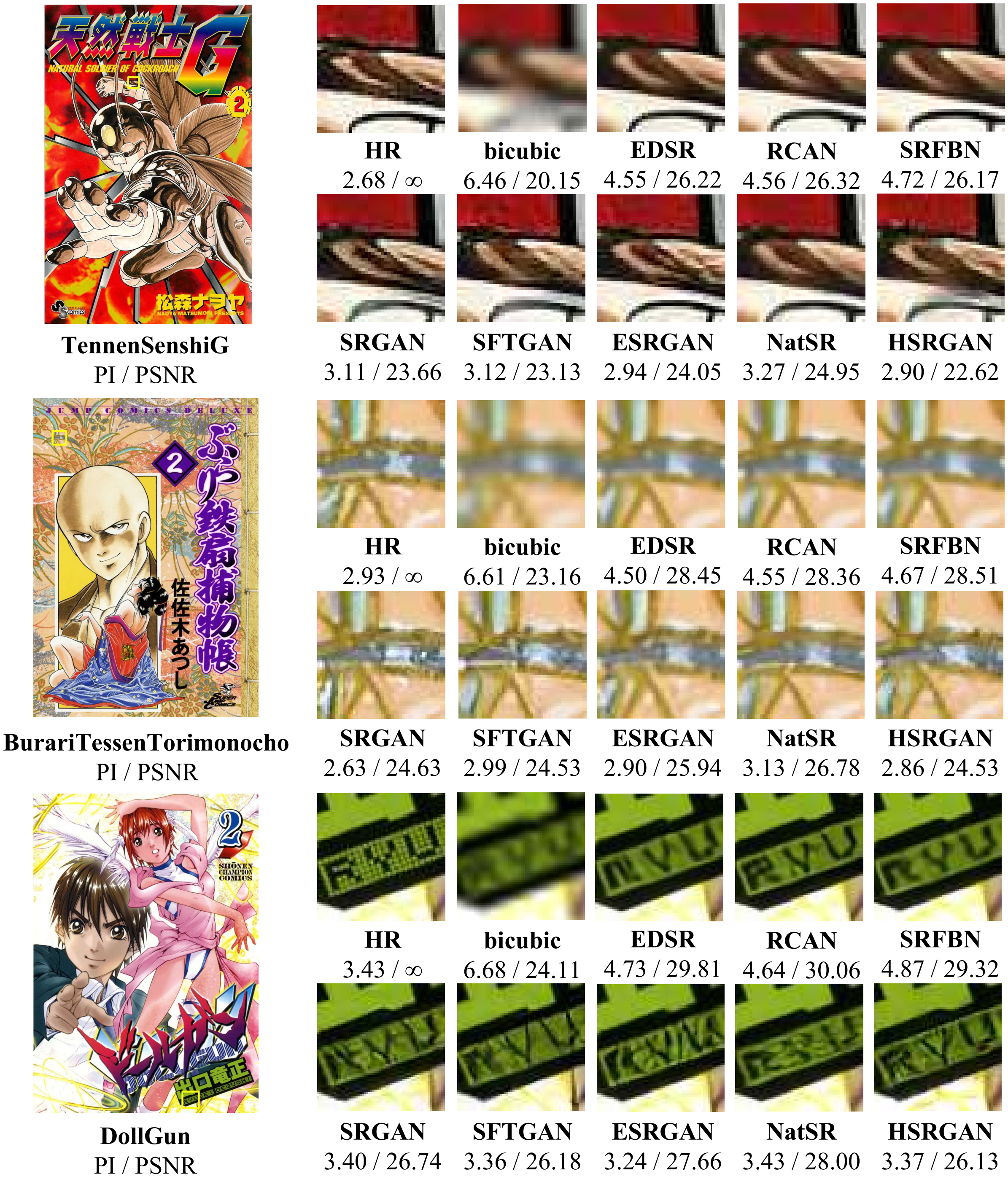

Figure 2. Qualitative results of our HSRGAN and state-of-the-art methods. HSRGAN generates more realistic textures and less artifacts.

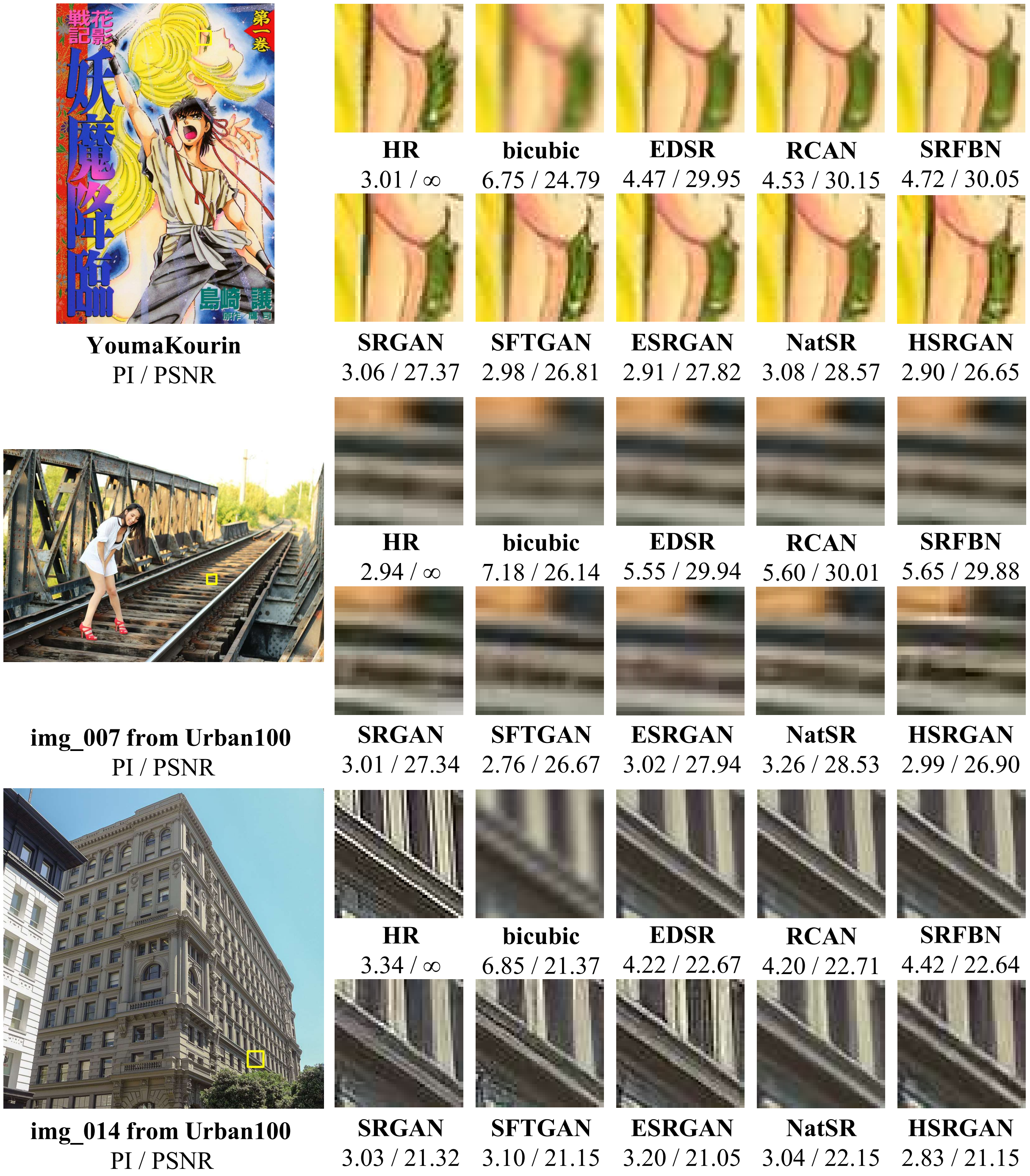

Figure 3. Qualitative results of our HSRGAN and state-of-the-art methods. HSRGAN generates more realistic textures and less artifacts.

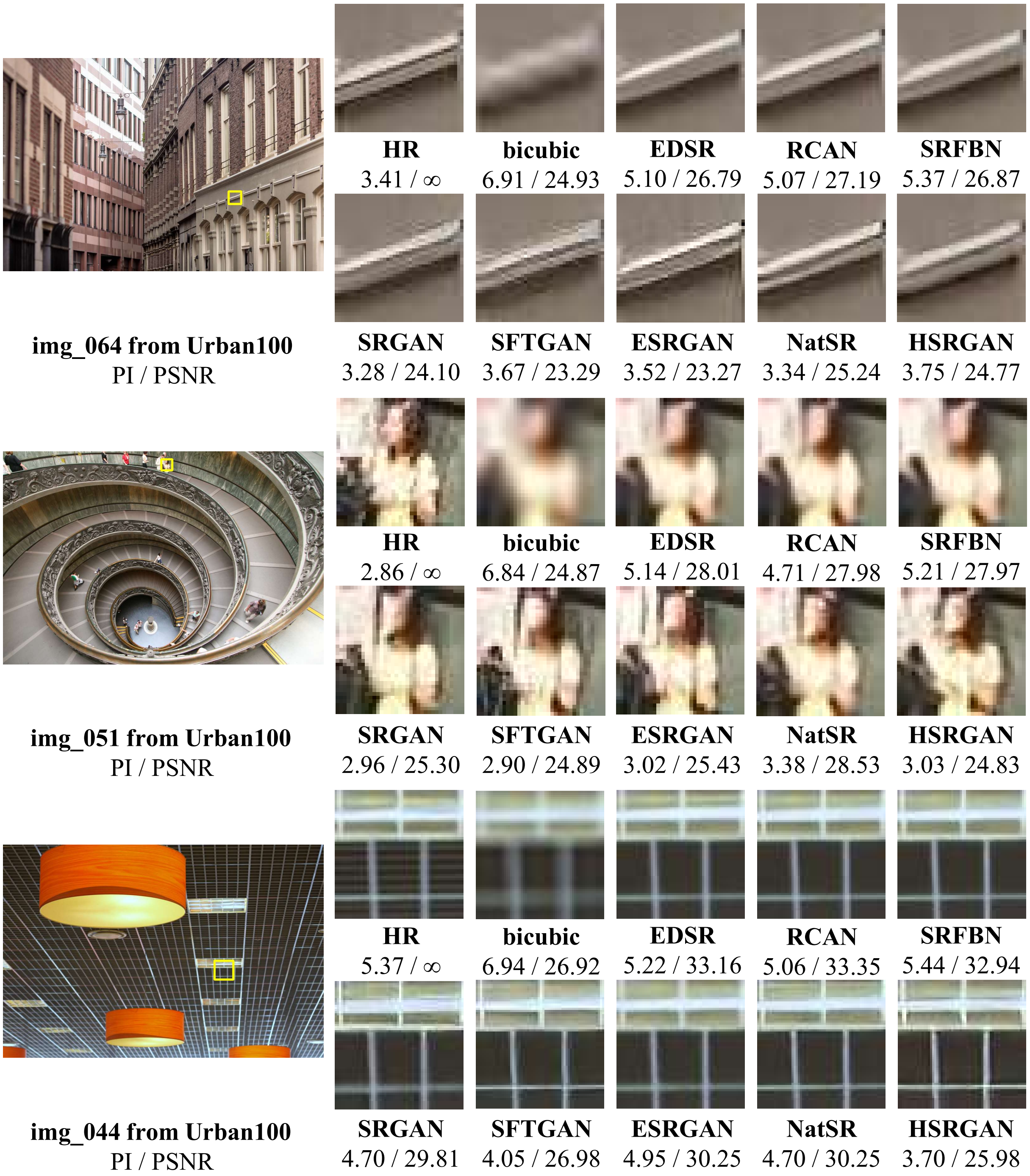

Figure 4. Qualitative results of our HSRGAN and state-of-the-art methods. HSRGAN generates more realistic textures and less artifacts.