Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data
Supplementary Material

Xintao Wang[1] Liangbin Xie*[2,3] Chao Dong[2,4] Ying Shan[1]
[1]Applied Research Center (ARC), Tencent PCG
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[3]University of Chinese Academy of Sciences [4]Shanghai AI Laboratory
{xintaowang, yingsshan}@tencent.com {lb.xie, chao.dong}@siat.ac.cn
https://github.com/xinntao/Real-ESRGAN

Abstract

We provide more details (especially examples) of each degradation type used in the classical degradation model in Sec. 1. Quantitative comparisons are presented in Sec 2. We show more qualitative comparisons with previous works in Sec. 3.

1. Details of Classical Degradation Model

1.1. Blur

Isotropic and anisotropic Gaussian filters are the common choices for blur kernels. We show several Gaussian kernels and their corresponding blurry images in Fig. 1.

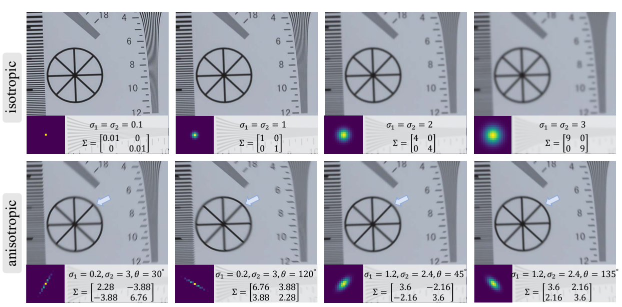

Figure 1: Examples of Gaussian kernels (kernel size 21) and their corresponding blurry images. Zoom in for best view

To include more diverse kernel shapes, we further adopt generalized Gaussian blur kernels [6] and a plateau-shaped distribution. Fig. 2 shows how the shape parameter β controls kernel shapes. Empirically, we found that including these blur kernels produces sharper outputs for several real samples.

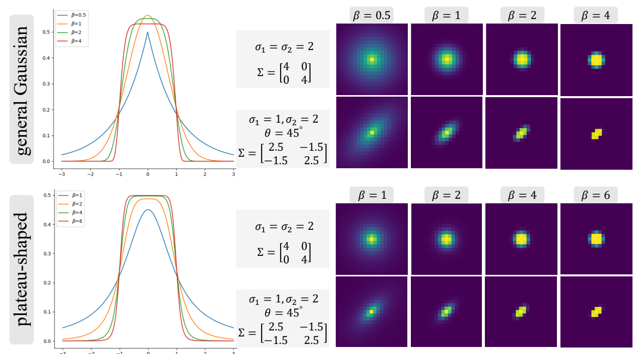

Figure 2: Blur kernels with different shape parameters in general Gaussian distribution and plateau-shaped distribution. Zoom in for best view

1.2. Noise

Fig. 3 depicts the additive Gaussian noise and Poisson noise. Poisson noise has an intensity proportional to the image intensity, and the noises at different pixels are independent of one another. As shown in Fig. 3, the Poisson noise has low noise intensity in dark areas.

Figure 3: Visual comparisons of Gaussian and Poisson noises. Poisson noise has low noise intensity in dark areas. Zoom in for best view

1.3. Resize

There are several resize algorithms. We compare the following resize operations: nearest-neighbor interpolation, area resize, bilinear interpolation and bicubic interpolation. We examine the different effects of these resize operations. We first downsample an image by a scale factor of four and then upsample to its original size. Different downsampling and upsampling algorithms are performed, and the results of different combinations are shown in Fig. 4. It is observed

Table 1: NIQE scores on several diverse testing datasets with real-world images. The lower, the better.

	Bicubic	ESRGAN [10]	DAN [7]	RealSR [5]	CDC [11]	BSRGAN [12]	**Real-ESRGAN**	**Real-ESRGAN+**
RealSR-Canon [2]	6.1269	6.7715	6.5282	6.8692	6.1488	5.7489	4.5899	**4.5314**
RealSR-Nikon [2]	6.3607	6.7480	6.6063	6.7390	6.3265	5.9920	5.0753	**5.0247**
DRealSR [11]	6.5766	8.6335	7.0720	7.7213	6.6359	6.1362	4.9796	**4.8458**
DPED-iphone [4]	6.0121	5.7363	6.1414	5.5855	6.2738	5.9906	5.4352	**5.2631**
OST300 [9]	4.4440	3.5245	5.0232	4.5715	4.7441	4.1662	2.8659	**2.8191**
ImageNet val [3]	7.4985	**3.6474**	6.0932	3.8303	7.0441	4.3528	4.8580	4.6448
ADE20K val [13]	7.5239	3.6905	6.3839	**3.4102**	6.9219	3.9434	3.7886	3.5778

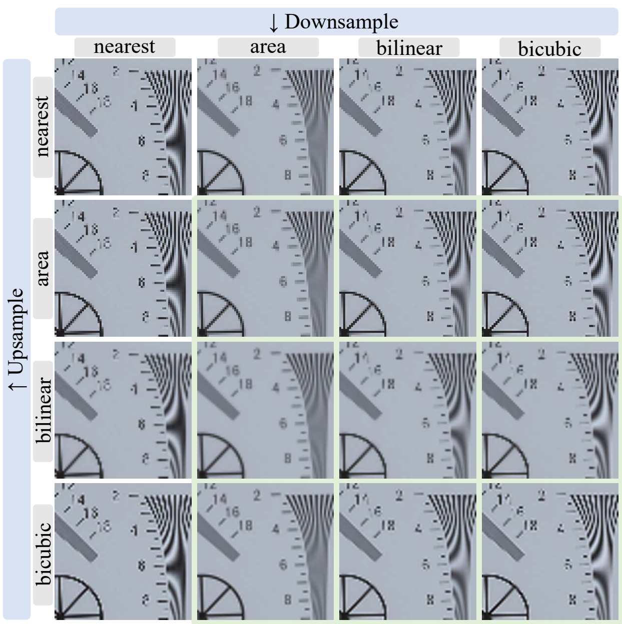

Figure 4: Effects of different combinations of down- and up-sampling algorithms. The images are first downsampled by a scale factor of four and then upsampled to its original size. Zoom in for best view

that different resize operations result in very different effects - some produce blurry results while some may output over-sharp images with overshoot artifacts.

1.4. JPEG compression

We use the PyTorch implementation - `DiffJPEG`. We observe that the compressed images by `DiffJPEG` are a bit different from those compressed by the `cv2` package. Fig. 5 shows the typical JPEG compression artifacts and the difference caused by using different packages. Such a difference may bring an extra gap between synthetic and real samples. In this work, we only adopt `DiffJPEG` for simplicity, and this difference will be addressed later.

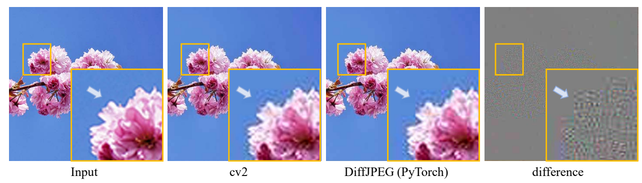

Figure 5: JPEG compressed images by `cv2` and `DiffJPEG`, with quality factor $q = 50$. They produces slightly different results. Zoom in for best view

2. Quantitative Comparisons

We provide the non-reference image quality assessment - NIQE [8] for reference. Note that existing metrics for perceptual quality cannot well reflect the actual human perceptual preferences on the fine-grained scale [1].

We compare our Real-ESRGAN with several state-of-the-art methods, including ESRGAN [10], DAN [7], CDC [11], RealSR [5] and BSRGAN [12]. We test on several diverse testing datasets with real-world images, including RealSR [2], DRealSR [11], OST300 [9], DPED [4], ImageNet validation [3] and ADE20K validation [13]. The results are shown in Tab. 1. Though our Real-ESRGAN+ does not optimize for NIQE scores, it sill produces lower NIQE scores on most testing datasets.

3. More Qualitative Comparisons

We show more qualitative comparisons with previous works. As shown in Fig. 6, our Real-ESRGAN outperforms previous approaches in both removing artifacts and restoring texture details. Real-ESRGAN+ (trained with sharpened ground-truths) can further boost visual sharpness. Other methods typically fail to remove complicated artifacts (the 1st sample) and overshoot artifacts (the 2nd, 3rd sample), or fail to restore realistic and natural textures for various scenes (the 4th, 5th samples).

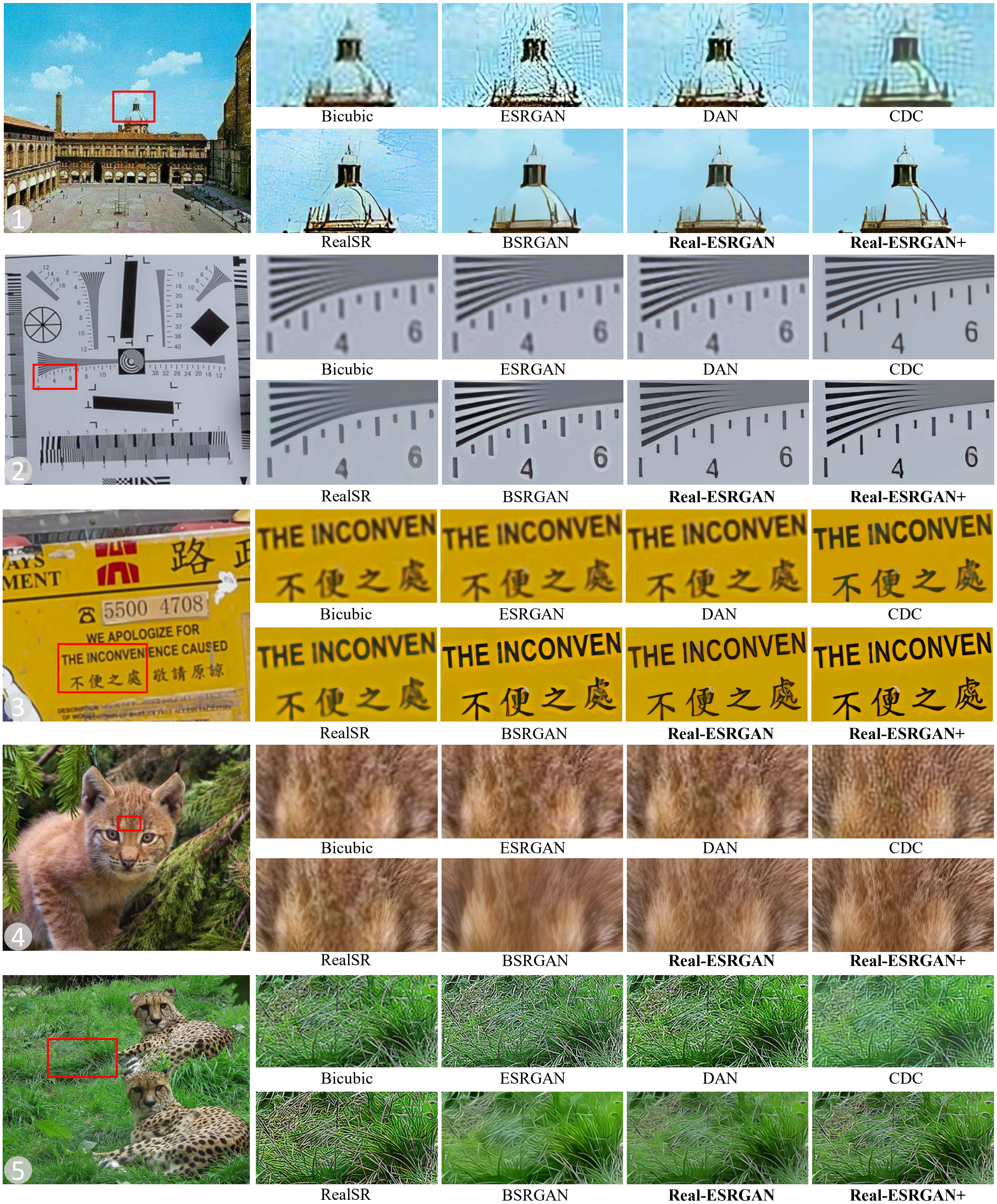

Figure 6: Qualitative comparisons on several representative real-world samples with upsampling scale factor of 4. Our Real-ESRGAN outperforms previous approaches in both removing artifacts and restoring texture details. Real-ESRGAN+ (trained with sharpened ground-truths) can further boost visual sharpness. Other methods typically fail to remove complicated artifacts (the 1st sample) and overshoot artifacts (the 2nd, 3rd sample), or fail to restore realistic and natural textures for various scenes (the 4th, 5th samples). (**Zoom in for best view**)

References

[1] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on perceptual image super-resolution. In *ECCVW*, 2018. 2

[2] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *ICCV*, 2019. 2

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009. 2

[4] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *ICCV*, 2017. 2

[5] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In *CVPRW*, 2020. 2

[6] Yu-Qi Liu, Xin Du, Hui-Liang Shen, and Shu-Jie Chen. Estimating generalized gaussian blur kernels for out-of-focus image deblurring. *IEEE Transactions on circuits and systems for video technology*, 2020. 1

[7] Zhengxiong Luo, Yan Huang, Shang Li, Liang Wang, and Tieniu Tan. Unfolding the alternating optimization for blind super resolution. In *NeurIPS*, 2020. 2

[8] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a completely blind image quality analyzer. *IEEE Signal Process. Lett.*, 20(3):209–212, 2013. 2

[9] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *CVPR*, 2018. 2

[10] 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 *ECCVW*, 2018. 2

[11] Pengxu Wei, Ziwei Xie, Hannan Lu, ZongYuan Zhan, Qixiang Ye amd Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In *ECCV*, 2020. 2

[12] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. *arXiv preprint arXiv:2103.14006*, 2021. 2

[13] Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Semantic understanding of scenes through the ade20k dataset. *International Journal of Computer Vision*, 2019. 2