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

Supplementary Material

Xintao Wang¹ Liangbin Xie²,³ Chao Dong²,⁴ Ying Shan¹
¹Applied Research Center (ARC), Tencent PCG
²Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
³University of Chinese Academy of Sciences ⁴Shanghai AI Laboratory
{xintaowang, yingsshan}@tencent.com ¹lb.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.

![Blur kernels with different shape parameters in general Gaussian distribution and plateau-shaped distribution. Zoom in for best view](image)

Figure 2: Blur kernels with different shape parameters in general Gaussian distribution and plateau-shaped distribution. 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 $\beta$ controls kernel shapes. Empirically, we found that including these blur kernels produces sharper outputs for several real samples.

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.

![Visual comparisons of Gaussian and Poisson noises. Poisson noise has low noise intensity in dark areas. Zoom in for best view](image)

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.

|--------------------------|---------|-------------|---------|------------|----------|-------------|-------------|-------------|

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

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).

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 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