

ESRGAN

Components that are different from SRGAN:

1. Network architecture:

Basic architecture is of SRResNet

'Basic blocks' or residual blocks in generator have been replaced with Residual-in-residual Dense Block(RRDB)

Batch normalization(BN) blocks have been removed - this has been shown to increase performance and reduce complexity

Lots of problems with BNs

Residual scaling has been done - residuals are multiplied by a constant between zero and one to prevent instability

Smaller initialization - empirically found to make it easier to train

2. Adversarial loss

Use a relativistic discriminator - one that predicts the probability that a real image is relatively more realistic than a fake one

$$D_{Ra}(x_r, x_f) = \sigma(C(x_r) - \mathbb{E}_{x_f}[C(x_f)])$$

where E represents the operation of taking average for all fake data in the mini batch

Discriminator loss:

$$L_D^{Ra} = -\mathbb{E}_{x_r}[\log(D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[\log(1 - D_{Ra}(x_f, x_r))].$$

Adversarial loss:

$$L_G^{Ra} = -\mathbb{E}_{x_r}[\log(1 - D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[\log(D_{Ra}(x_f, x_r))],$$

3. Perceptual loss

$$L_G = L_{\text{percep}} + \lambda L_G^{Ra} + \eta L_1,$$

$$L_1 = \mathbb{E}_{x_i} \|G(x_i) - y\|_1$$

Features before the activation layers have been chosen to define perceptual loss, as they contain more information

Network interpolation:

We first train a PSNR-oriented network G_{PSNR} and then obtain a GAN-based network G_{GAN} by fine-tuning. We interpolate all the corresponding parameters of these two networks to derive an interpolated model G_{INTERP} , whose parameters are:

$$\theta_G^{\text{INTERP}} = (1 - \alpha) \theta_G^{\text{PSNR}} + \alpha \theta_G^{\text{GAN}}$$

This model does not produce artifacts

We can balance perceptual quality and fidelity without having to retrain the model

Directly interpolating images pixel by pixel produces blurry or noisy images

BN removal: We first remove all BN layers for stable and consistent performance without artifacts.

Before activation in perceptual loss: Using features before activation can result in more accurate brightness of reconstructed images. Using features before activation helps to produce sharper edges and richer textures.

RaGAN: RaGAN uses an improved relativistic discriminator, which is shown to benefit learning sharper edges and more detailed textures.

Deeper network with RRDB: Deeper model with the proposed RRDB can further improve the recovered textures.

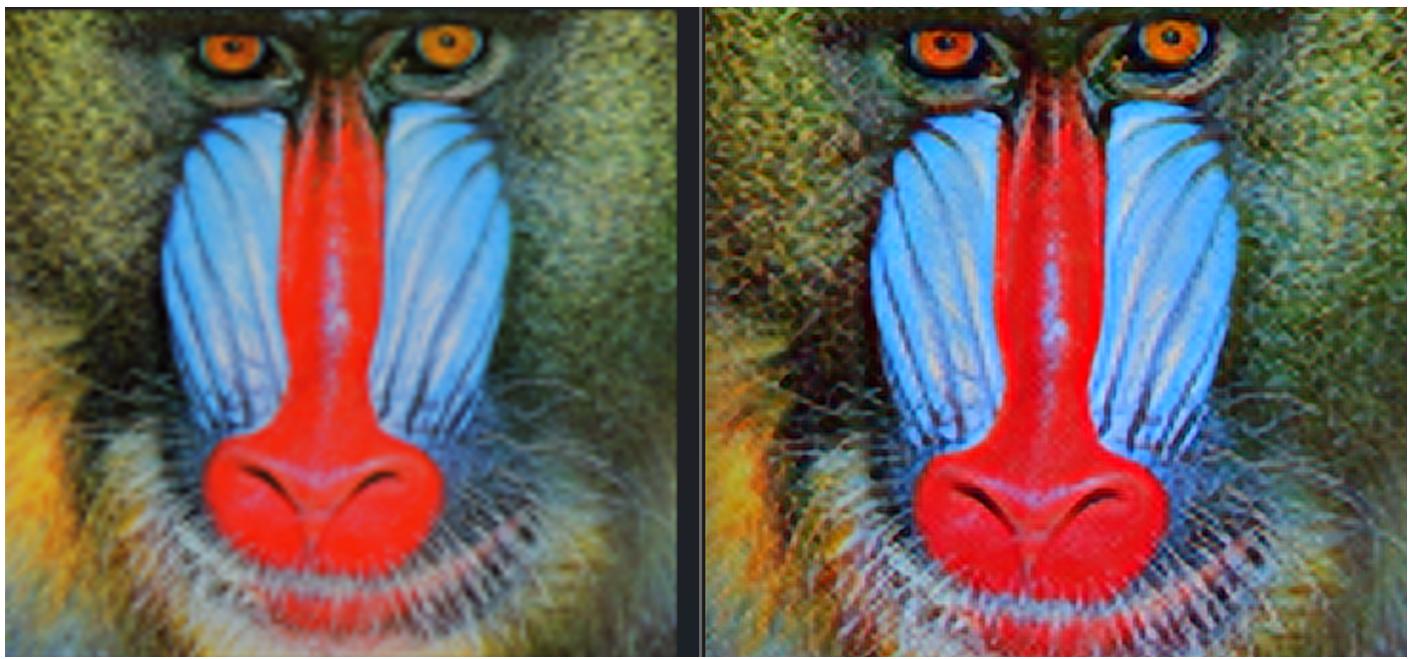
Final model:

We implemented RRDB, residual scaling, and smaller initial weights.

Due to lack of resources, we had to train for 500 epochs instead of 10000 as was mentioned in the paper.

We got the following results:





References:

1. Wang, X. et al. (2019). ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In: Leal-Taixé, L., Roth, S. (eds) Computer Vision – ECCV 2018 Workshops. ECCV 2018. Lecture Notes in Computer Science(), vol 11133. Springer, Cham.
https://doi.org/10.1007/978-3-030-11021-5_5
2. https://youtu.be/ZM4_s5dAWpI?si=-b6JbRF85ZWHneUT