ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks

Neelu Verma
MP19AI002
Muskan Dosi
MT19CSE027

Department of Computer Science and Engineering,
Indian Institute of Technology
Jodhpur

1 Critical Analysis

ESRGAN (Enhanced-SRGAN) is capable of generating realistic textures during single image super-resolution. This paper shows improvement of three key components of SRGAN – network architecture, adversarial loss and perceptual loss and derive an Enhanced SRGAN (ESRGAN) to further enhance the visual quality.
This Architecture containing several RDDB blocks without Batch Normalization layers and improve perceptual quality rather than objective quality, such as PSNR.
In ESRGAN Perceptual Index (PI) is a measure of perceptual quality if Higher PI, lower perceptual quality But in PSNR Root mean square error is a measure of objective quality.It means higher RMSE, lower objective quality.
ESRGAN took idea from **relativistic GAN**.It means it is somewhat similar to RGAN which is used to let the discriminator to predict relative realness instead of the absolute value.

1.1 Advantages

1.PSNR-oriented methods such as SRCNN, EDSR, and RCAN, tend to generate blurry results and previous GAN-based methods, i.e., EnhanceNet and SRGAN, tend to produce textures that are unnatural and contain unpleasing noise.But ESRGAN gets rid of these artifacts and produces natural results

2. ESRGAN used **Perceptual loss** that means using the features before activation, which provides stronger supervision for brightness consistency and texture recovery.

3. In ESRGAN Batch Normalization (BN) is removed so by removing BN layers has proved that it increase performance and reduce the computational complexity in different PSNR-oriented tasks, including SR. Because BN brings artifacts.

4. ESRGAN used the dense block to replace the residual block which enhance the network.

5. Paper shows that adversarial loss for the generator contains both real image and fake image. Therefore, the generator benefits from the gradients from both generated data and real data in adversarial training, while in SRGAN, only generated part takes effect.

6. The **interpolated model** is able to produce meaningful results for any feasible without introducing artifacts and we can continuously balance perceptual quality and fidelity without re-training the model.

7.pure GAN-based method produces sharp edges and richer textures but with some unpleasant artifacts, while the pure PSNR-oriented method outputs cartoon-style blurry images. But by employing network interpolation, unpleasing artifacts are reduced while the textures are maintained.

1.2 Disadvantages

1. ESRGAN supports only RGB images, that means it will remove alpha / transparency channels if present and it won't work with grayscale images. But there are ways around both. At the time of writing this there are some popular tools can be used to solve this.

2. ESRGAN used the features before the activation layers which raise some drawbacks:
first is the activated features are very sparse and sparse activation provides weak supervision and thus leads to inferior performance. second is using features after activation also causes inconsistent reconstructed brightness compared with the ground-truth image.

1.3 Technical Improvements

We know from this paper that ESRGAN contain artifacts in reconstruction results,So in the future, we may adopt more appropriate network architectures,a novel and improved algorithm or Methods to combine Texture Loss.