Generative Adversarial Networks for High-Resolution Image Enhancement

Introduction

This project studies how to enhance low-resolution images into high-resolution counterparts using Generative Adversarial Networks (GANs). The Super-Resolution Generative Adversarial Network (SRGAN) was a pioneering Deep Learning approach that combined GANs with perceptual loss, enabling the generation of sharper and more realistic images. Despite its success, SRGAN struggled with recovering fine textures and often introduced artifacts in challenging scenarios.

Building upon this, the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) improved image quality significantly by replacing SRGAN's original building blocks with Residual-in-Residual Dense Blocks (RRDBs). ESRGAN introduced new "basic blocs" as well as a Relativistic GAN, enhancing the discriminator's ability to evaluate realism between real and generated images.

Studying the transition from SRGAN to ESRGAN allows to better understand the motivations behind using this model. This project also aims to highlight some limitations of ESRGAN in practice, which justify the need for further research, broader training, and the introduction of other architectures.

Proposed Methods

Both models are based on the same generative architecture, with 4x-upscaling.

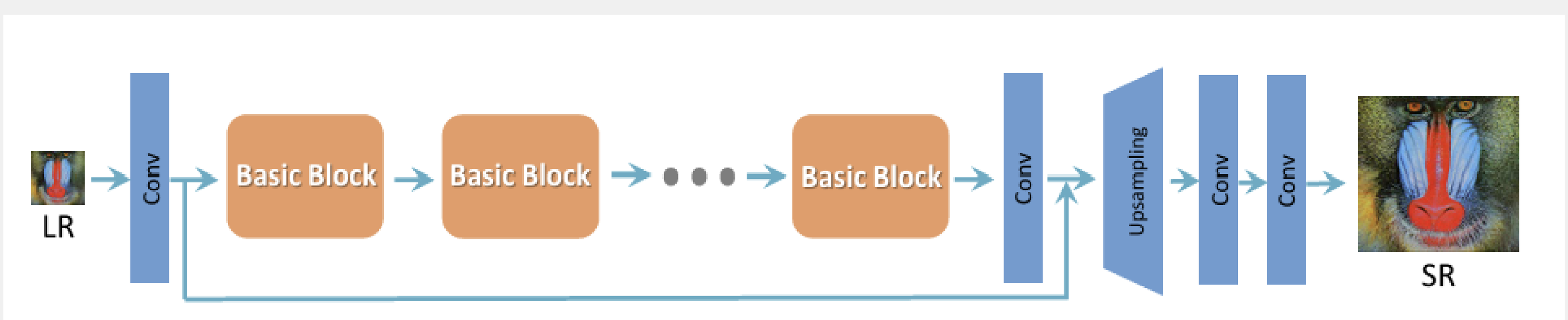

Figure 1. SRResNet Generator Network

One of the main contributions of [2] is to design new "basic blocs". First by removing the Batch Normalization (BN) layers, resulting in **less artifacts, reduced computation needs, and improved generalization.** Second, by introducing RRDB blocs, defined as:

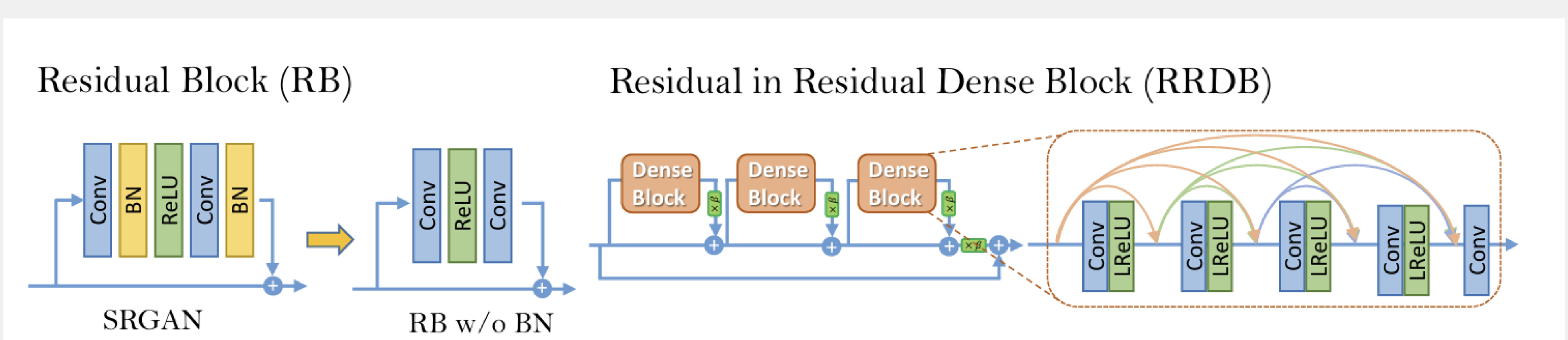

Figure 2. Basic blocks, SRGAN vs. ESRGAN

We now define the Discriminator Network of the SRGAN, and call $C(x)$ its output before the sigmoid, denoted σ:

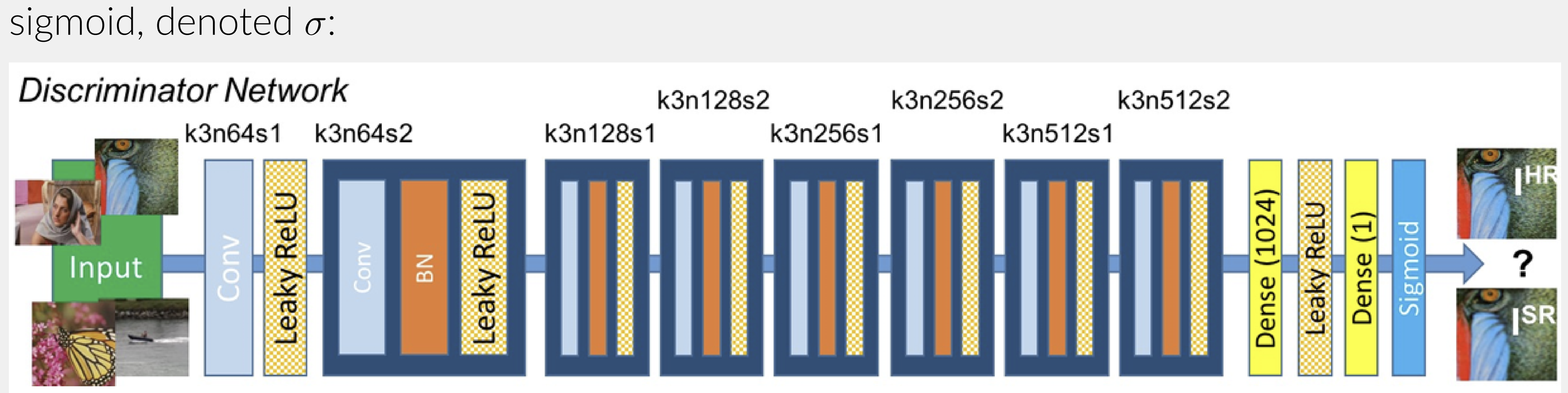

Figure 3. Discriminator Architecture

The authors of [2] slightly modify the output. The relativistic discriminator rather predicts the probability that a real image is more realistic than a fake one (see the definition of D_{Ra}), instead of the probability of an image to be real.

Loss Function

As defined in [2], the loss function is the combination of three terms:

$$Loss = L^{perceptual} + \lambda L_G^{Ra} + \eta L_1.$$

where:

- $L^{perceptual}$ is the Generator loss, defined in [1] as a VGG-based loss, combining a content loss with an adversarial loss. For ESRGAN, [2] apply it **before activation**, avoiding weak supervision (sparsity of activated neurons) and brightness effects.
- L_G^{Ra} is the Relativistic Adversarial loss (for the generator), defined as:
$$L_G^{Ra} = -\mathbb{E}_{x_r}(\log(1 - D_{Ra}(x_r, x_f))) - \mathbb{E}_{x_f}(\log(D_{Ra}(x_r, x_f)))$$
where x_r is the real image and x_f the fake one and $D_{Ra}(x_r, x_f)$ is the relativistic average discriminator, which replaces the classic discriminator: $D_{Ra}(x_r, x_f) = \sigma\left(C(x_r) - \mathbb{E}_{x_f}\left[C(x_f)\right]\right)$, whereas the classic discriminator does not depend on x_r: $D(x_f) = \sigma(C(x_r))$.
- L_1 is a context loss, defined as the mean L_1 distance between the recovered and the ground-truth images.

Executive summary

GANs have proved very useful in high-resolution image enhancement by:

- Outperforming pre-Deep Learning methods (filtering, interpolation, optimization) which produced over-smoothed and blurry images.
- Being more adapted than CNNs (e.g., SRCNN) which relied on pixel-by-pixel loss and are designed for discrimination / segmentation rather than for generation.
- SRGAN first specifically designed for high-resolution. ESRGAN introduces RRDB blocks, relativistic GANs and improves the perceptual loss to reduce artifacts and produce sharper and more realistic images.
- But this method is limited: artifacts, training complexity, computational cost or generalization issues.
- More recent techniques appear: Transformers, Diffusion Models.

Training and experiments

Authors train on DIV2K (800 images), Flickr2K (2650 images) and the OutdoorSceneTraining and test on Set5, Set14, BSD100, Urban100, and the PIRM self-validation.

Evaluation is mostly qualitative, as the authors precise that there is **no standard and reliable metric** to evaluate perceptual-based networks.

We evaluate the **RRDB-ESRGAN-x4 model** in this project on the Set5 dataset, consisting of 256x256 images and on other LR images neither used in training or testing to assess the generalization abilities of the model. We obtained good results on both training and testing sets:

Figure 4. Illustration (train set)

Generalization and limits

Test set and unseen inputs

Let us examine the model outputs on images of the Set5 dataset and on the "model" image:

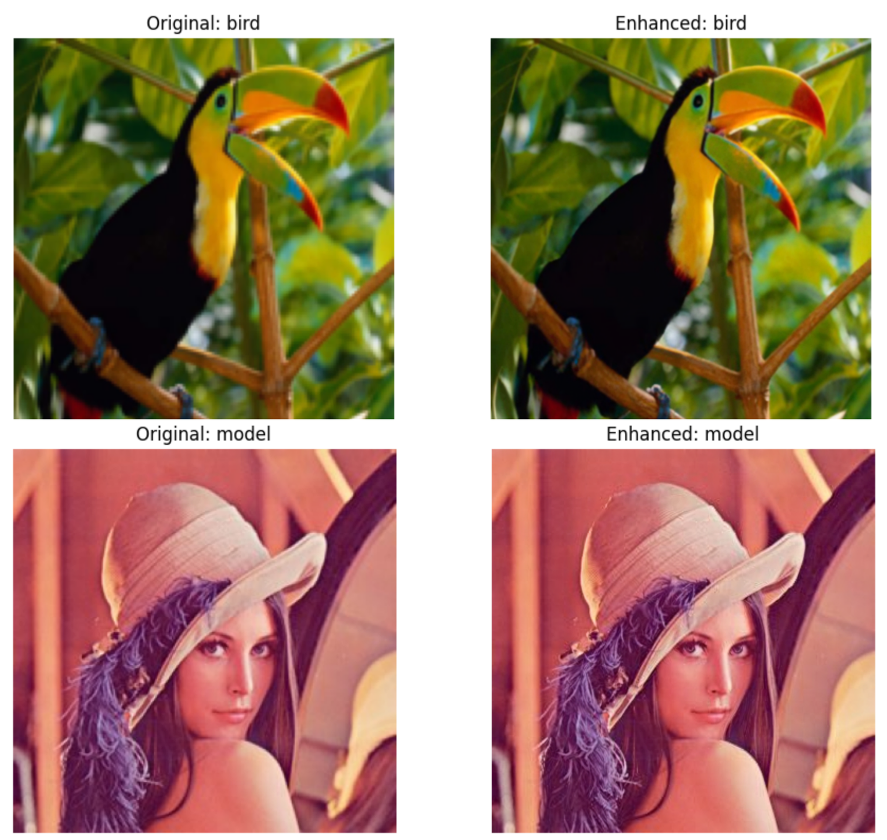

Figure 5. Test set (bird) and "model" image (never seen before)

Limits on unseen data

ESRGAN lacks generalization abilities: the picture is enhanced, but artifacts are introduced, brightness / colors are not well respected and the image remains noisy, as we can see:

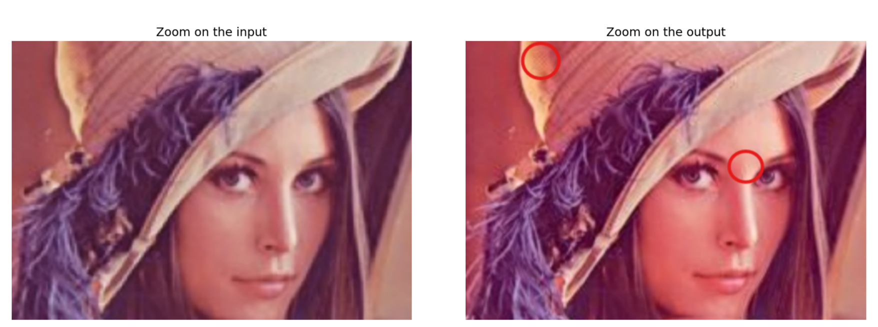

Figure 6. Discussion of the result, model image

Artifacts and noise are systematic problems with the GAN architecture. The color instability comes from activation layers (controlled by the ηL_1 term).

Conclusion

GANs are an exciting area of research for image generation and enhancement, an important topic tackled by several industries. ESRGAN builds on the inefficiencies of SRGAN and showed improved sharpness and lower artifacts generation. However, the model we studied shows poor generalization abilities, justifying the later development of Real-ESRGAN (same network, more training...). Finally, we also discussed structural problems of GANs that required the introduction of new state-of-the-art methods, such as Diffusion Models, more suited for this task.

References

[1] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802*, 2017.

[2] 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 *The European Conference on Computer Vision Workshops (ECCVW)*, September 2018.

[3] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *International Conference on Computer Vision Workshops (ICCVW)*, 2021.

Alexandre Boistard