1 Introduction

Single Image Super resolution has gained a lot of traction in the deep learning community recently. Once the pioneer work of SRCNN was proposed, many other deep convolutional neural networks have brought prosperous development. The main focus of these approaches are to improve PSNR(Peak signal to noise ratio), but we have seen in the SRGAN paper that this metric fundamentally disagrees with the subjective evaluation of human observers.

The amazing work implemented by SRGAN paper which proposes a GAN based approach and perceptual loss function encourages the network to favor the solutions that are more photo realistic or more like natural images. Having said that there is still some clear difference between the generated images and the natural images.

This paper really takes the concept of SRGAN and tries to improve the model by incorporating the following changes.

1) Network structure introduces Residual-in-Residual Dense Block (RDDB), which has higher capacity and easier to train.
2) Removal of Batch normalization layers, and using residual scaling and smaller initialization to help training a very deep network.
3) Introduction of RaGAN (Relativistic average GAN), which focusses on identifying more realistic images, rather than if an image is fake or real.
4) Improvement in perceptual loss by using VGG features before activation, which was done after activation in SRGAN.

2 Proposed Approach

The main focus is to improve the overall perceptual quality for SR. We will try to explain the different components in the subsections below.

2.1 Network Architecture

The main changes in the structure of the generator of SRGAN are:
1) Remove all the Batch Normalization layers
2) Replace the original basic block with the proposed Residual-in-Residual Dense Block(RDDB).

Impact of removing the BN layers made a huge impact as reported in the paper. It increased performance and reduced computational complexity in different PSNR oriented tasks including super resolution and debarring. The reason for this can be that BN layers normalize the features using means and variance in a batch during training and use estimated mean and variance of the whole training set during testing. If there is lot of difference in the training and testing set, BN layers introduce unpleasant artifact and hamper generalization. One observation was that BN layers bring artifacts for deep networks and if it trained over GAN framework.

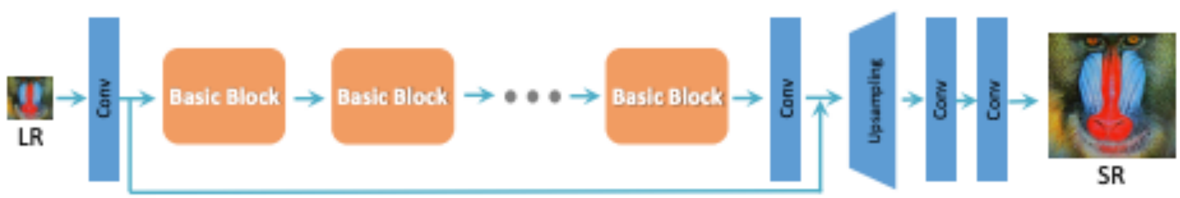
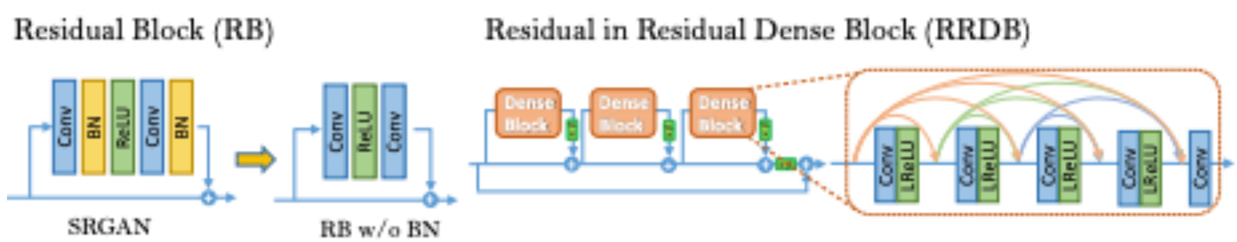


Fig 1. High level SRResnet, Generator for SRGAN

The high level architecture proposed is similar to SRGAN, they used basic RDDB block with a deeper and more complex structure which is different from the normal Residual block used in SRGAN. Both the architectures is depicted in the images below.



Other than the changes in the architecture, they also did residual scaling i.e scaling the residuals by multiplying them with a constant between 0 and 1, to make them stable.

2.2. Relativistic discriminator

Apart from making changes in the generator architecture they also changed the structure of the discriminator by introducing relativistic GAN, which focusses on identifying the relative difference between a realistic image and a fake one. The below figure depicts the difference of this from a standard discriminator.

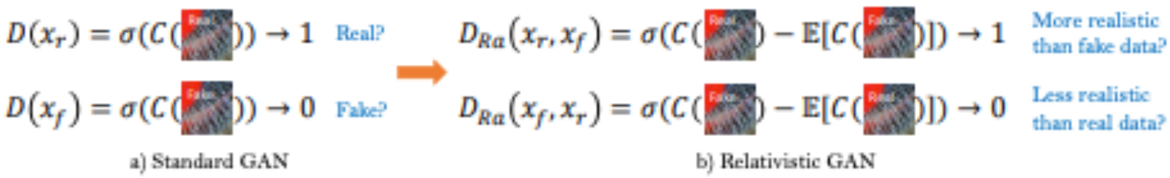


Fig3: Diff between standard and relativistic discriminator

2.3 Perceptual Loss

A more effective perceptual loss by constraining on features before activation rather than after activation as practiced in SRGAN. The advantage of using the features before activation overcomes drawbacks in the original design like having sparse activated features in a very deep network, using features after activation also causes inconsistent reconstructed brightness compared with the ground truth image. Fine tuning the VGG network for material recognition also helped in identifying textures which is very critical in case of Super Resolution.

2.4 Network Interpolation

They also did network interpolation which helped in removing unpleasant noise and keeping intact the perceptual quality of the image. To achieve this the approach is to first train the PSNR-oriented network and then obtain a GAN-based network by fine tuning. It helps in producing meaningful results and maintains the perceptual quality and fidelity without the need to retrain the model.

3 Qualitative Results

The results obtained by the approach proposed in the paper were compared on some benchmark datasets with state of the art models like SRCNN, EDSR, RCAN and SRGAN to name a few. The difference can be seen in the figure 4.

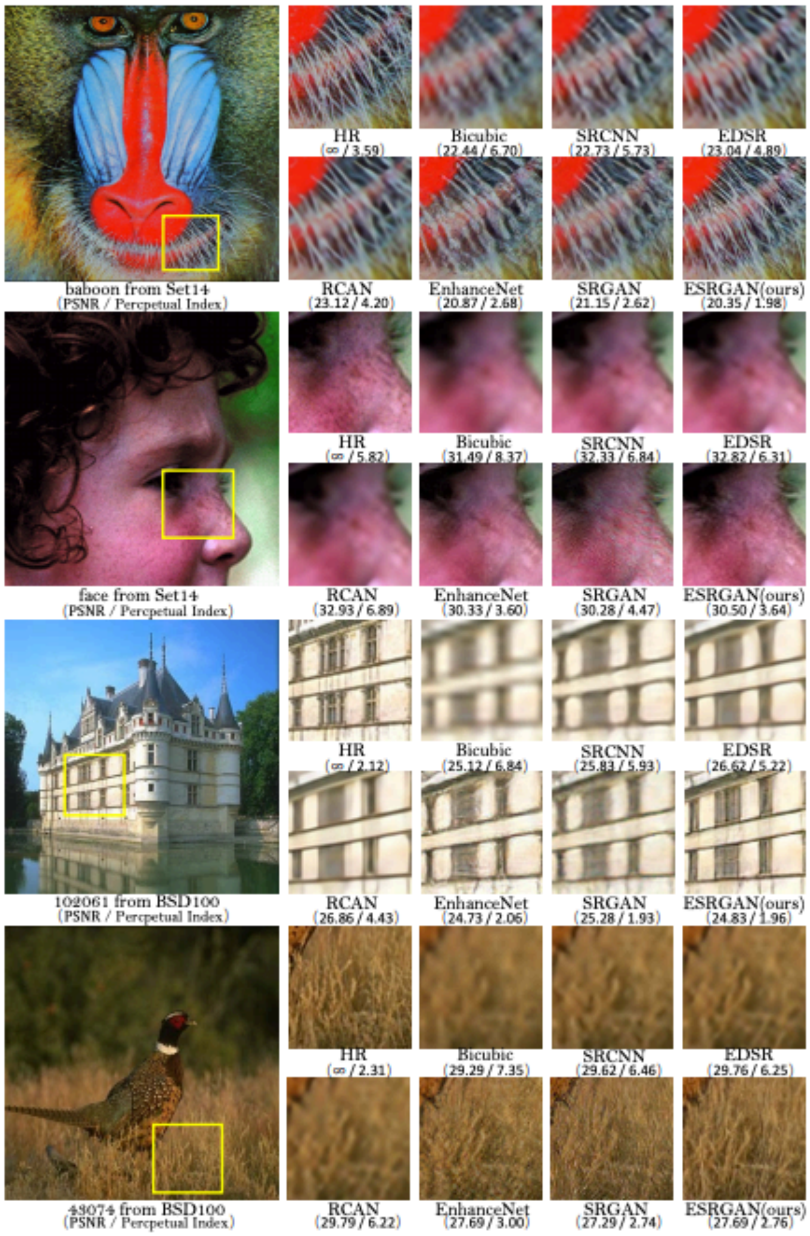


Fig4. Comparing results of ESRGAN.

4 Conclusion

The results shown in the paper present that ESRGAN model achieves consistently better perceptual quality than previous SR methods. Some takeaways from the approach are, RDDB block, Relativistic GAN for discriminator, VGG features before the activation and Network Interpolation.