Super Pixel

1st Dhruv Kabariya
Under Prof. Mehul Raval
Ahmedabad University
Ahmedabad, India
dhruv.k1@ahduni.edu.in

2nd Shivam Thakker
Under Prof. Mehul Raval
Ahmedabad University
Ahmedabad, India
shivam.t1@ahduni.edu.in

3rd Pranav Gandhi
Under Prof. Mehul Raval
Ahmedabad University
Ahmedabad, India
pranav.g@ahduni.edu.in

***Abstract*—The primary aim of this project is to construct a high-resolution (HR) image from a corresponding low-resolution (LR) input image. The key pain point behind the super-resolution is to recover the finer texture details when super-resolved at large upscaling factor. We are trying to build a SRGAN[1], a generative adversarial network (GAN) for image super-resolution which takes care of every finer details and generate the output image accordingly.**

***Index Terms*—Generative Adversarial Network(GAN), Super Resolution, Generator Network, Discriminator Network**

I. Introduction

The super important task of estimating a High Resolution(HR) image from it's low resolution(LR) image is known as Super Resolution(SR). A lot of research has been carried out in this field in the computer vision field. Super resolution tries to reconstruct high quality clean image from it's low quality degraded image. For this, Convolution Neural Network has become the base for Image Reconstruction. The most challenging task in image restoration is to restore every fine texture details of the image.

Most of the CNN based models focus on increasing the architecture design such as residual learning[2],[3] and dense connections[4],[5]. Using the above methods, the performance has increased significantly but it suffers from two basic problems. First the interactions between images and convolution kernels are content-independent. Means they are using same convolution kernel to restore different image regions which may not be the suitable choice. Second, under the principle of local processing, convolution is not effective for long-range dependency modelling.

A. Abbreviations and Acronyms

- GAN - Generative Adversarial Network
- SR - Super Resolution.
- I^{LR} - Low Resolution Image
- I^{HR} - High Resolution Image
- I^{SR} - Super Resolution
- SNR - Signal to Noise Ratio

II. Problem Formulation

- Creating a model which will convert Low resolution image into High resolution by taking care of every small pixel level detail.
- Minimizing the Mean Squared Error(MSE) between the produced High Resolution image and the original image.
- Designing a deep neural network architecture for generating High Resolution image

III. Contribution

GAN is a very powerful method for generating images close to the real image with high perceptual quality. Here in this project we have used deep ResNet architecture using the concept of GANs to form a perceptual loss function for photo-realistic SISR. Our main contributions are:

- We set a new state of the art for image SR with high upscaling factors (4×) as measured by PSNR and structural similarity (SSIM) with our 16 blocks deep ResNet (SRResNet) optimized for MSE.
- We propose GAN which is a GAN-based network optimized for a new perceptual loss. Here we replace the MSE-based content loss with a loss calculated on feature maps of the VGG network [49], which are more invariant to changes in pixel space

IV. Network Model

We want to generate a High Resolution Image(I^{HR}) from a Low Resolution(I^{LR}) input image. In training, I^{LR} is obtained by applying a Gaussian filter to I^{HR} followed by a downsampling operation with downsampling factor r. For an image with C color channels, we describe I^{LR} by a real-valued tensor of size W × H × C and I^{HR}, I^{SR} by rW × rH × C respectively.

Our main aim is to train a generating function G that estimates for a given LR input image its corresponding HR image. To achieve this, we train a generator network as a feed-forward CNN G_{θ_G} parametrized by θ_G. Here θ_G = $W_{1:L}; b_{1:L}$ denotes the weights and biases of a L-layer deep network and is obtained by optimizing a SR-specific loss function l_{SR}. For training images I^{HR}_n n = 1, . . . , N with corresponding I^{LR}_n , n = 1, . . . , N, we solve:

$$\hat{\theta}_G = \arg\min_{\theta_G} \frac{1}{N} \sum_{n=1}^{N} l^{SR}(G_{\theta_G}(I_n^{LR}), I_n^{HR}) \quad (1)$$

We also define a discriminator network D_{θ_D} which we optimize in an alternating manner along with G_{θ_G} to solve the adversarial min-max problem:

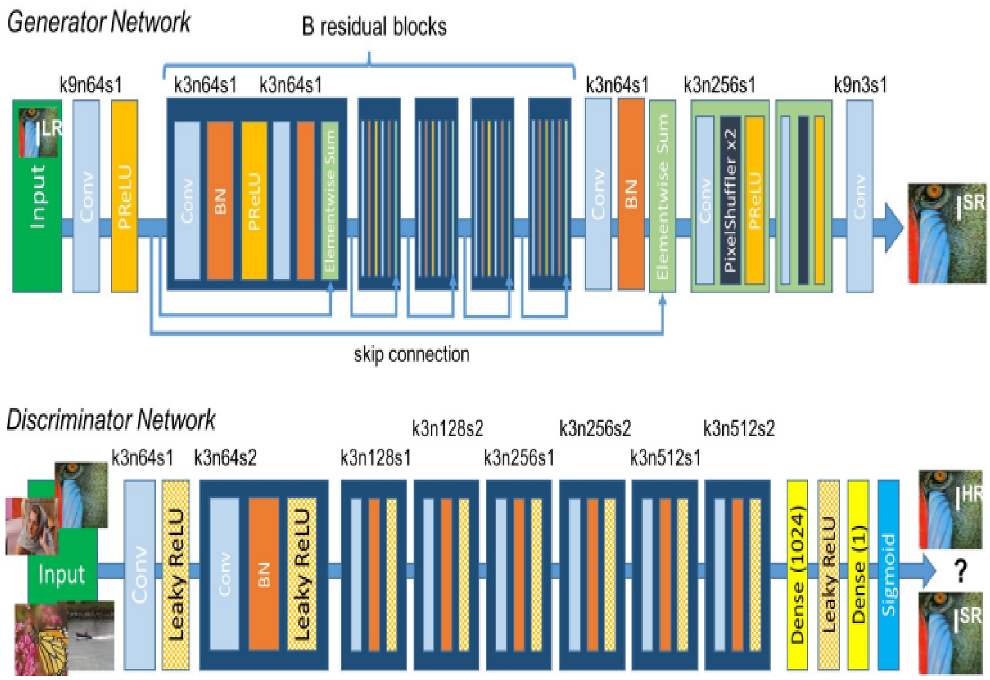

Fig. 1. Generator and Discriminator Model

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{\text{train}}(I^{HR})}[\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})}[\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))] \tag{2}$$

This discriminator network helps in distinguising between super resolved images and real images. Fig. 1. shows the Network model of Generator and Discriminator model. The Generator model Block uses 3x3 kernel and 64 feature maps with convolution method followed by batch normalization, and Paramteric Relu as the activation function. We increase the resolution of the input image with two trained sub-pixel convolution layers. In Discriminator model, uses 3x3 kernel with convolution method followed by batch normalization, and Leaky Relu as the activation function and avoid max-pooling throughout the network. 8 blocks are connected sequentially with Dense Connected Layer (MLP) having 1024 node and Relu Activation. Sigmoid activation is used for last layer.

V. Results

Here in Fig. 2. we have shown the comparison of different models such as Bicubic, SResNet, GAN with the original image. It is evitable from the figure that the GAN produces image close to the original image which proves that GAN works very well as compared to Bicubic or SRResNet.

VI. Conclusion and Future Works

Proposed GAN method is providing excellent result and great image quality. Some image are preserving texture information but when image with more details are given the network was not able to clear those fine details in High Resolution image.

Fig. 2. Comparison of image generated by Bicubic, SResNet and GAN with original image

So in next step we would like to add technique which will carry fine detail of image to end of network and add to HR image.The steps would be,

- Adding Self Attention to Model (for better feature extraction).
  - Better performance in Machine translation
  - Visual self attention give better result too
  - Mechanism of focus on texture thing
- Creating graphs of epoch loss and noise in generated image.
- Graph of noise in our and SRResNet and other work.
- Graph of Content Loss and adversarial Loss.

References

[1] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang et al., "Photo-realistic single image super-resolution using a generative adversarial network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4681–4690.

[2] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Ca- ´ ballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In IEEE Conference on Computer Vision and Pattern Recognition, pages 4681–4690, 2017

[3] Wenbo Li, Kun Zhou, Lu Qi, Nianjuan Jiang, Jiangbo Lu, and Jiaya Jia. Lapar: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. arXiv preprint arXiv:2105.10422, 2021

[4] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image superresolution. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2472–2481, 2018

[5] 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 European Conference on Computer Vision Workshops, pages 701–710, 2018