SRFBGAN: Image super resolution using Feedback loops in Generative Adversarial Networks

Sagar Vakkala
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India

Aniket Mishra
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India

Senthil Kumar
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India

Punitha D
Asst. Professor, Department of Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India.

Abstract— **Image super resolution has always had a great deal of proportion in the problem set solved by neural networks. The quality of the generated SR image has always been proportional to the time and resources put into the training of the network. With many methods like SR-GANs and Feedback loops for super resolution imaging all have their disadvantages, the biggest being the performance-time tradeoff. What SRFBGAN intends to do is use the best parts of both worlds, the high quality SR image provided by SR-GANs and the utilization of high resolution features by Feedback loops combined to produce a high quality SR image comparable to the other state-of-the-art architectures. This is accomplished by altering the structure of the Generator used to create the SR image. The novel generator architecture consists of feedback loops that ensure high level information is utilized when generating super resolution images.**

Keywords— SR:- Super Resolution, SISR:- Single Image Super Resolution, CNN: Convolution Neural Networks, GANs:- Generative Adversarial Networks, Feedback Loops, SRFBGAN:- Super Resolution Feedback Generative Adversarial Network

I. INTRODUCTION

Image super-resolution (SR) are methods by which a lower resolution image can be converted to a higher resolution image so that when it is enlarged exponentially, it is not distorted and does not lose any of its features. The application of image super-resolution has drastically changed over the decades. Since it got attention from the computer vision research community, other fields such as biotechnology have been trying to use SR techniques to enhance the visualization of the mechanisms of the cell, for tissue engineering and so on. With all the new challenges in many domains that can be solved by super-resolution imaging, a more novel approach that solves the SR problem of efficiency and the performance time tradeoff is required.

Since high frequency information in a HR image is lost or degraded due to aliasing during the sampling process of a single LR image is a challenging SR problem and because there was no effective way to reduce high frequency information, the single image SR problem was usually considered as an added problem. In addition to this, for high upscaling factors, the texture detail constructed SR images is almost always very low. To counter these problems, this paper has restructured the generator in the SR-GAN by Christian Ledig et al.[3] with a novel generator architecture by incorporating it with a feedback loop[3] in the convolution neural network as designed by Jiuxiang Gua et al.[6] as a combined network of sorts.

The discriminator is a general GAN discriminator, which gives a 'True' value for the original image and a 'False' value for a LR image. The combined novel GAN will produce 'True' for a SR image created by the generator if it manages to create a HR replica of the provided LR image or it will be rejected. The architecture of the feedback loop in the generator enhances the LR image exponentially, the GAN produces a very high quality HR image in a very short training time reducing the performance-time trade off.

II. RELATED WORK

A. Image Superresolution

It is a set of processes used to create HR images from LR images. They provide a much sharper and pixel correct image that looks exactly like the original even when changing the size exponentially. Thus, after super-resolution, a smaller image looks similar and not distorted even after increasing the size exponentially. Usually, neural networks are used to train each image by breaking it into a matrix of numbers with each point in the matrix representing each pixel in the image. Articles of Nasrollahi and Moeslund [7] are a recent overview on image SR Imaging. One of the first methods to tackle SISR was Prediction-based. The filtering approaches like Lanczos, linear or bicubic filtering can be quick to produce results but they yield images with inordinately smooth textures as they tend to oversimplify the SISR problem. The proposed methods put particular focus on edge-preservation. There are other more powerful methods that aim to set up a complex mapping between LR and HR image information, which usually rely on the training data. Overlapping patches of images reduce consistency and so we process the whole image, thereby improving consistency.

B. Convolution Neural Networks

Following the success of the work by Jin Yamanakal, Shigesumi Kuwashima1 and Takio Kurita [9] for many image manipulation and computer vision problems, deep convolutional neural networks (DNN) was set as the state-of-the-art. It has been demonstrated that these deeper network architectures are easier and faster to train than the former DNN networks and achieve strong performance, SISR being no exception. These concepts of residual blocks and skip-connections are a very powerful design choice that enhances the training of DNNs while not sacrificing on speed or accuracy. This is an improvement over the research done by C. Dong et al.[10] where the image was upscaled using bicubic interpolation before being fed to CNN.

C. SRGAN and ESRGAN

The development of SRGAN[4] by Christian Ledzig et al. established a new standard for SR Imaging by generating photo realistic images of downsampled HR images. SRGAN employs a dense ResNet that is 16 blocks deep for generating images upscaling them up to 4 times. The use of a generative adversarial network vastly improves the quality of images generated compared to older SISR CNN models. Following the development of SRGAN, Enhanced SRGAN (ESRGAN) by Xintao Wang et al.[11]. aims to improve upon the network architecture, and the loss functions of SRGAN. ESRGAN utilizes a Residual - in - residual block (RDRB) for their network architecture along with a realtistivistic discriminator borrowed from relativistic GAN [12]. ESRGAN also improves the existing loss functions by utilizing the features before activation layers.

III. NETWORK ARCHITECTURE

A. The GAN Architecture

We use the traditional network architecture of a Generative Adversarial Network (GAN) as established by Goodfellow et al. (2014)[1] consisting of Generator (G) and Discriminator (D).

$$\min\max[\log(D(x)) + \log(1-D(G(z)))]$$

1. The min-max algorithm used in the Generator G.

The generator converts Low Resolution (LR) images to Super Resolution (SR) images. The discriminator is a convolutional network that is used to classify or detect if an image is real or fake. The generator and discriminator are connected in an adversarial network where the job of the discriminator is to classify images from the generator as fake and images that belong to the original ground truth HR as real. The job of the generator is to pass the discriminator's test by using the images generated by the generator.

B. Generator G

Borrowing the idea of feedback networks discussed in 'Feedback Networks' by Amir R et al. [2] and the recent development of SRFBN by Zhen Li et al.[3] we employ the

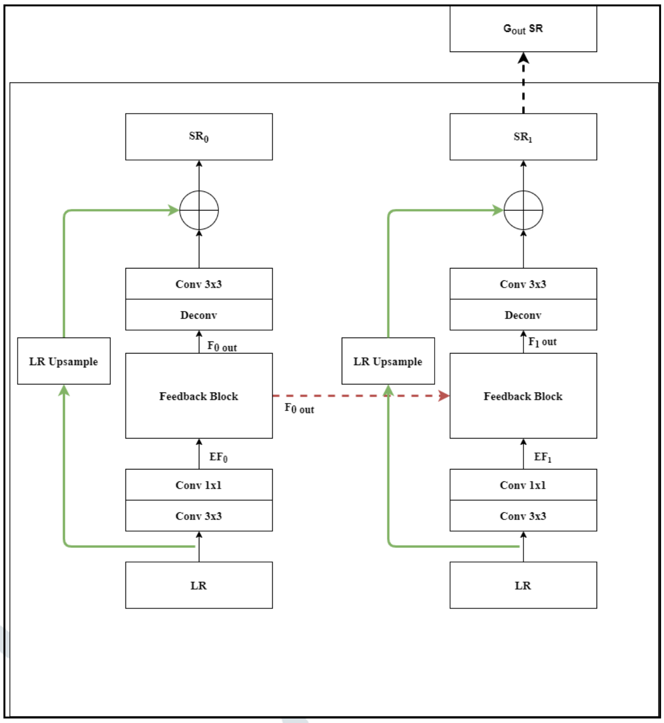

Fig 3. The Generator architecture

feedback loops as discussed in SRFBN for our Generator Architecture. This allows for the generator layers to learn from the following layers

The generator begins with convolutional layers of sizes Conv(3x3) and Conv(1x1) that were used to extract features from the Low Resolution(LR) image. These extracted features (EF) are passed onto the following Feedback Block. The output from the feedback block(FBout) is used as the input for the feedback block of the next iteration along with EF of that block.

$$FB_{input} = FB_{prev} + EF_{curr}$$

Fig 4. The Feedback block input.

FB out is then passed onto a Deconvolution Block that is to upscale the LR features to HR features followed by a Conv(3x3) to generate a residual image I_{res} that is combined with an upscaled image of the LR image using a skip connection. The final SR image is obtained as follows

$$SR_{img} = Upscale(LR) + Conv(Deconv(FB_{out}))$$

Fig 4. The final SR Image

To simplify training we have assumed the final SR image at the last iteration as the output of the generator.

C. The Feedback Block

At the beginning of the block EF_{curr} is concatenated with FB_{prev} using a Conv(1x1) layer. This allows us to capture the HR features of the previous feedback block iteration along with the extracted LR features of the current iteration. This concatenated input is passed through projection groups that project HR features to LR consist of alternating up

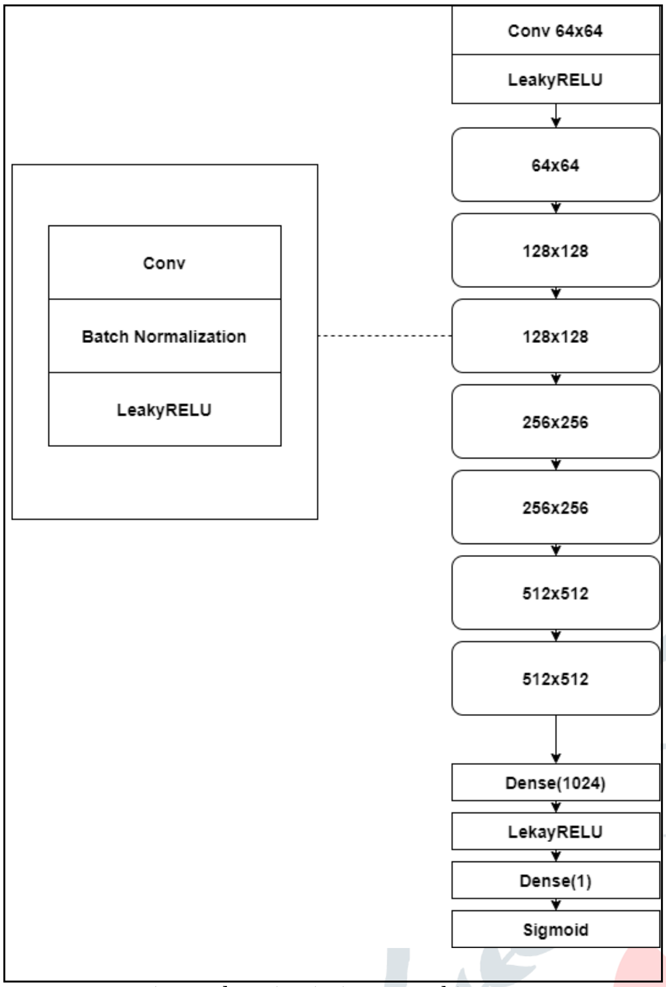

Fig 5. The Discriminator architecture

sampling and downsampling operations. Each projection consists of skip connections that pass HR and LR features separately. LR features are obtained by downsampling the previous HR features while the HR features are obtained by up sampling the previous LR features. After the final projection block LR features from all the previous groups are concatenated to form FB_{out}.by using a Conv(1x1)

$$FB_{out} = Conv(LR_0, LR_1, LR_2 ......LR_n)$$

Fig . Feedback output.

D. Discriminator

The discriminator in Fig 6 is a detection network consisting of Convolutional layers and pooling layers that is used in training the GAN by classifying the output from the generator as real or fake. The architecture consists of 8 convolutional layers that are each followed by a leakyRELU activation. The final layer is connected to a dense layer followed by a sigmoid activation. The sigmoid function is used to find the probability of realness or fakeness of the input provided to the discriminator.

IV. Implementaiton /Method

We make the use of a combination of generator losses and an adversarial loss to find the total loss of the network. While the SRFBN uses a L1 loss for its architecture we instead use a MSE loss that compares the final SR image generated at the last iteration to the original High resolution (HR) image.

A. Adversarial Loss

We make the use of the traditional method of calculating adversarial loss in GAN by trying to minimize the function of [1- D(G(LR))]

Where G(LR) is the output of the generator i.e. the fake image generator and D(G(LR)) is the discriminator's probability classification of the fake image.

B. MSE loss

A simple loss function that is used to compare the SR and HR images by comparing the images either by pixel wise or comparing the Euclidean distances of the feature maps.

$$SR_{MSE} = \sum_{x=1}^{m}\sum_{y=1}^{n}(SRx,y - HRx,y)$$

Fig 7. The MSE Loss function

$SR_{x,y}$ is the Super resolution image generated by the generator G and $HR_{x,y}$ is the original ground truth image.

C. Training

We train our network on a NVIDIA GTX 1080 using the DIV2K data set of images with an upscale factor of 4 with an Adam optimizer set = 0.09 over 100 epochs with a learning rate of 10^{-4}. We set the value of 'T', the number of feedback loop iterations as 2 as discussed earlier.

V. Experimental Results

A. Datasets and settings

The network uses DIV2K dataset for training. To make sure the network behaves properly with corner cases and also to use the data efficiently, The data is augmented.

B. Image degradation methods

To achieve consistency and also to build a fair comparison system for the network with existing systems we use "Bicubic downsampling" also known as **BI** to degrade our training images to a lower resolution and also to remove a great deal of high level features from the original image. This gives a very accurate representation of an actual lower resolution image while giving the same consistency throughout the dataset.

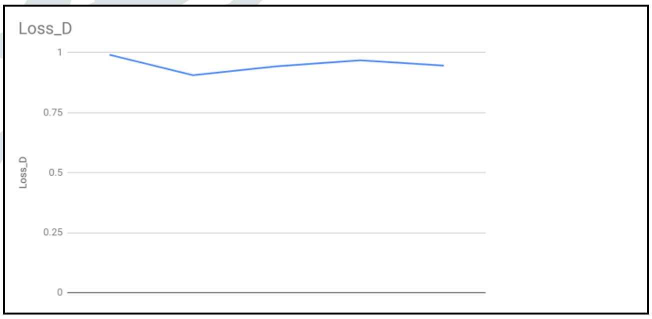

Fig 7. Discriminator Loss

To see how the network progresses in it's SR capabilities we feed the patches of RGB synthetic LR images in the training and check the D_LOSS for the first 5 epochs.

We also compare the performance of the generator and discriminator through Generator score (Score_G) and Discriminator score (Score_D) so that we may avoid common failure modes of GAN training such as convergence failures or mode collapse.

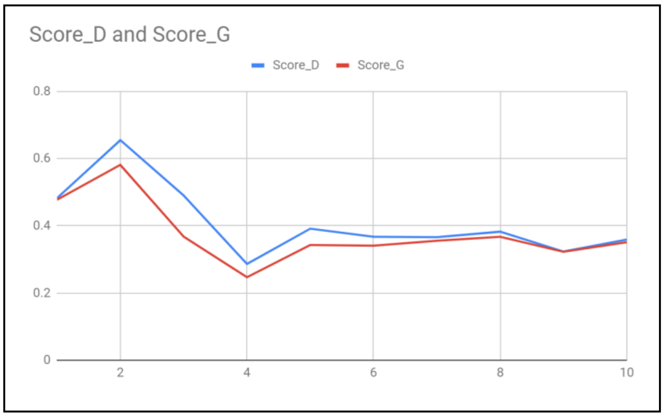

Fig 8. Generator score and Discriminator Score over 10 epochs.

Table 1 PSNR, SSIM Values

EPOCH	PSNR	SSIM
1	24.01056	0.708006
2	24.0147	0.708892
3	24.0112	0.707293
4	24.00392	0.704376
5	24.0161	0.709347
6	24.00966	0.706577
7	24.01143	0.707305
8	24.00926	0.706412
9	24.00926	0.705553
10	24.00926	0.702917

Table 2: PSNR, SISM Values for state of the art SR techniques on the set14 dataset.

SR Method	PSNR	SISM
SRGAN	26.02	0.7397
SRResNet	28.49	0.8184
SRFBN	26.60	0.7101

VI. 6. Conclusion and Future work

SRFBGAN introduces a new model for generating super resolution images. It shows a Generative Adversarial network with a novel generator architecture in the form of feedback loops that obtains performance that is comparable to that of existing systems. For future work, the impact of the number of feedback loops or iterations to increase the image quality can be considered at the cost of training time and memory space.

Fig 9. From left to right: - Bicubic interpolation, original HR image and Super Resolution image produced by SRFBGAN

References

[1] Generative Adversarial Networks, Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio

[2] Feedback Networks, Amir R. Zamir, Te-Lin Wu, Lin Sun, William Shen, Jitendra Malik, Silvio Savarese

[3] Feedback Network for Image Super-Resolution, Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, Wei Wu

[4] Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi

[5] Efficient Module Based Single Image Super Resolution for Multiple Problems, Dongwon Park, Kwanyoung Kim, Se Young Chun, School of ECE, Ulsan National Institute of Science and Technology, 44919, Ulsan, South Korea

[6] Recent Advances in Convolutional Neural Networks, Jiuxiang Gua , Zhenhua Wangb , Jason Kuenb , Lianyang Mab , Amir Shahroudyb , Bing Shuaib , Ting Liub , Xingxing Wangb , Li Wangb , Gang Wangb , Jianfei Caic , Tsuhan Chenc chill.

[7] Super-resolution: A comprehensive survey. In Machine Vision and Applications, K. Nasrollahi and T. B. Moeslund.

[8] Image deblurring and superresolution by adaptive sparse domain selection and adaptive regularization. IEEE Transactions on Image Processing, W. Dong, L. Zhang, G. Shi, and X. Wu. 2011

[9] Fast and Accurate Image Super Resolution by Deep CNN with Skip Connection and Network in Network, Jin Yamanaka1, Shigesumi Kuwashima1 and Takio Kurita

[10] Learning a deep convolutional network for image super-resolution, C. Dong, C. C. Loy, K. He, and X. Tang. 2014.

[11] ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks, Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy , Yu Qiao2 , Xiaoou Tang

[12] The relativistic discriminator: a key element missing from standard GAN, Alexia Jolicoeur-Martineau