Introduction

Despite breakthroughs in accuracy and speed of single image super-resolution using faster and deeper CNN architectures, there is one central problem which remains unsolved: how do we recover finer texture details when we super resolve at large up scaling factors? SRGAN proposes an adversarial objective function that promotes super resolved image that lie close to the manifold of natural images.
The main highlight of the approach taken is multi-task loss formulation that consists of three main parts: 1) a MSE loss that encodes pixel wise similarity, 2) a perceptual similarity metric in terms of distance metric defined over high-level image representation, and 3) an adversarial loss that balances a min-max game between a generator and a discriminator.
This framework basically favors outputs that are perceptually similar to the high resolution images. To quantify the results a new metric is introduced which is Mean Opinion Score(**MOS**) which is basically the average of the scores assigned manually by human raters on the quality of the Super-Resolved image. And as other techniques optimize data dependent measures, it outperforms them by a significant margin on the proposed metric.

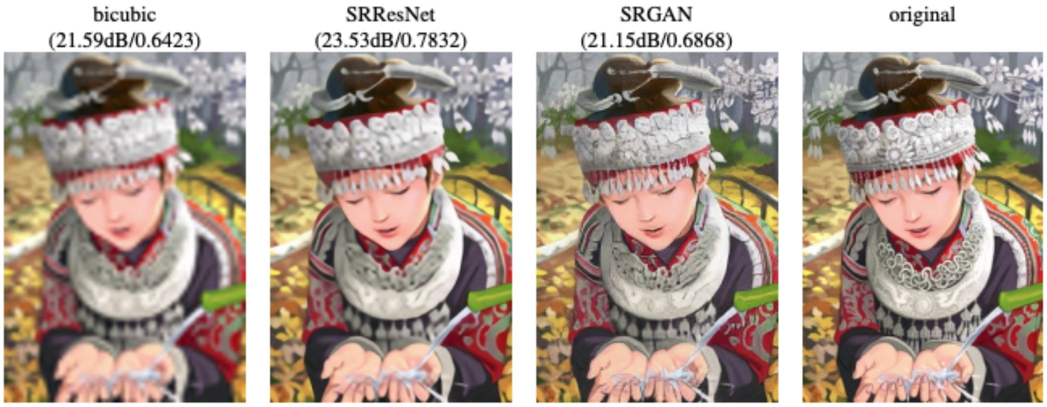
Contribution of the technique

The GAN approach moves the reconstructions to the space with high probability of creating photo-realistic images which means it is closer to the natural image manifold. The main take away from this approach is first GAN based approach with a Deep ResNet architecture to train the generator. And achieving the state of the art performance in PSNR using a No GAN approach with just a Deep ResNet and having MSE as our loss function. We summarize the main contributions of the approach below:

- A new of the state of the art for Image Super resolution with high upscaling factors (4x) as measured by PSNR and SSIM with the 16 blocks deep resnet (SRResNet) optimized for MSE.

- Confirmation with extensive MOS test on images from three benchmark datasets set5 set14 and BDS100 that SRGAN is the new state of the art by a long distance, for the estimation of photo realistic Super resolution images.

Just an example from the paper where we can clearly see what is the meaning of photo realistic image and using MOS metric for evaluation of performance.


bicubic (21.59dB/0.6423)

SRResNet (23.53dB/0.7832)

SRGAN (21.15dB/0.6868)

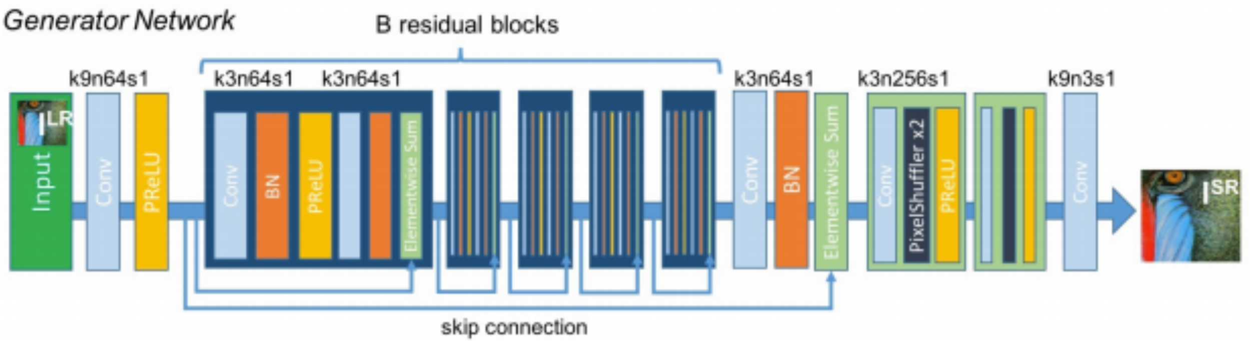
original

Architecture and Method

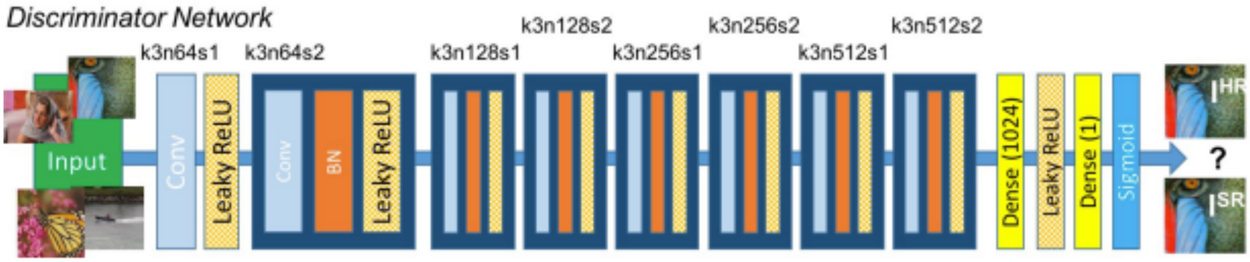
Take the High Resolution images which is only available during training, apply Gaussian filters to the HR images followed by a down sampling operation with a down sampling factor ***r,*** An image with tensor dimensions of WxHxC is downsampled to rW * rH * C. The end goal is to train a generator which estimates a High Resolution image from its Low resolution counter part.

The Architecture is based on the GAN approach where a discriminator network is trained which is optimized in an alternating manner with the generator network. With the approach we try to generate images which are almost same as real images and make it difficult for the discriminator to classify. This encourages a perceptually superior image residing in the manifold of the natural images. This kind of approach is totally different from the other SR solutions where they try to minimize the pixel wise measurement loss such as Mean Squared Error.

Generator architecture:
The core of the network is a number of **residual** blocks which have identical layout. There are two convolutional layers with small **3x3** kernels and **64** feature maps followed by batch normalization layers. The activation function used is **ParamtericRELU.** The architecture is depicted in the figure below.



Discriminator architecture:
It contains eight convolutional layers with an increasing number of **3x3** kernels which is similar to a **VGG** network. The increment is by a factor of 2 from **64** to **512** kernels. Strides are used to reduce resolution when the number of features is doubles. The activation used in the beginning is **leakyRELU** and in the end the network has two dense layers with a final **sigmoid** activation function. The architecture is depicted in the figure below.



Loss Function(Perceptual)

One of the main focus of the approach is the definition of the perceptual loss function which played a major role in enhancing the performance of the generator network. The loss is a sum of two loss functions **content** loss and **adversarial** loss. The formula for the same is given below:

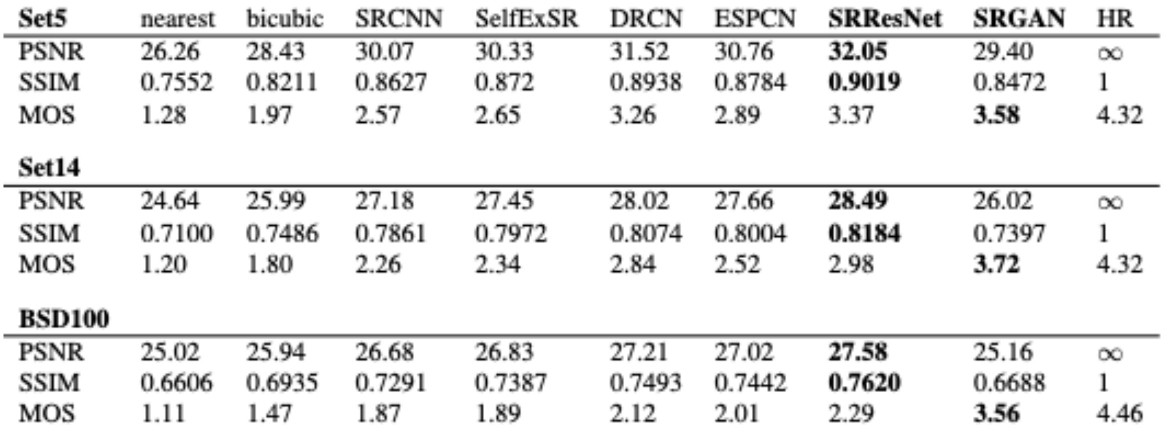
$$l^{SR} = \underbrace{\underbrace{l^{SR}_{\mathrm{X}}}_{\text{content loss}} + \underbrace{10^{-3} l^{SR}_{Gen}}_{\text{adversarial loss}}}_{\text{perceptual loss (for VGG based content losses)}}$$

Let's talk about content loss. The most commonly used loss function for SR is pixel wise MSE loss. However it lacks the frequency content which in turn results in perceptually unsatisfactory solutions. The loss used in the paper is closer to the perceptual similarity. It is basically VGG loss which is the euclidean distance between the feature representations of the a reconstructed image and the reference image as in HR image.

With this we also add the generative loss too the network to get the overall perceptual loss. The loss is defined based on the probabilities of the discriminator over all the training examples.

Performance of the networks

When the performance of various networks like NN, bicubic interpolation and other state of the art methods were compared to SRGAN and SRResNet on three benchmarks sets i.e set5, set14, BDS100 SRResNet gave a new state of the art in terms of PSNR/SSIM score as mentioned in the paper. One other rating which they used was MOS on the BDS100 using all the reference methods for super resolution and SRGAN outperforms all the others on this metric by a large margin. The result is depicted in the table below. This is directly from the paper which we are referencing.



Set5	nearest	bicubic	SRCNN	SelfExSR	DRCN	ESPCN	**SRResNet**	**SRGAN**	HR
PSNR	26.26	28.43	30.07	30.33	31.52	30.76	**32.05**	29.40	∞
SSIM	0.7552	0.8211	0.8627	0.872	0.8938	0.8784	**0.9019**	0.8472	1
MOS	1.28	1.97	2.57	2.65	3.26	2.89	3.37	**3.58**	4.32
Set14									
PSNR	24.64	25.99	27.18	27.45	28.02	27.66	**28.49**	26.02	∞
SSIM	0.7100	0.7486	0.7861	0.7972	0.8074	0.8004	**0.8184**	0.7397	1
MOS	1.20	1.80	2.26	2.34	2.84	2.52	2.98	**3.72**	4.32
BSD100									
PSNR	25.02	25.94	26.68	26.83	27.21	27.02	**27.58**	25.16	∞
SSIM	0.6606	0.6935	0.7291	0.7387	0.7493	0.7442	**0.7620**	0.6688	1
MOS	1.11	1.47	1.87	1.89	2.12	2.01	2.29	**3.56**	4.46

Key takeaways:

1) **New state of the art with SRResNet in terms of PSNR/SSIM ratings which is basically a no GAN version of the network proposed where it just uses the generator with MSE loss function.**
2) **With the output of the images and MOS scores, the paper has highlighted some limitations of PSNR/SSIM metrics for evaluation.**
3) **SRGAN architecture which uses a GAN architecture with residual blocks and augments the content loss with adversarial loss.**
4) **One of the main things is that the MOS scores of SRGAN for large up scaling factors is a new state of the art as compared to the reference methods mentioned in the paper.**