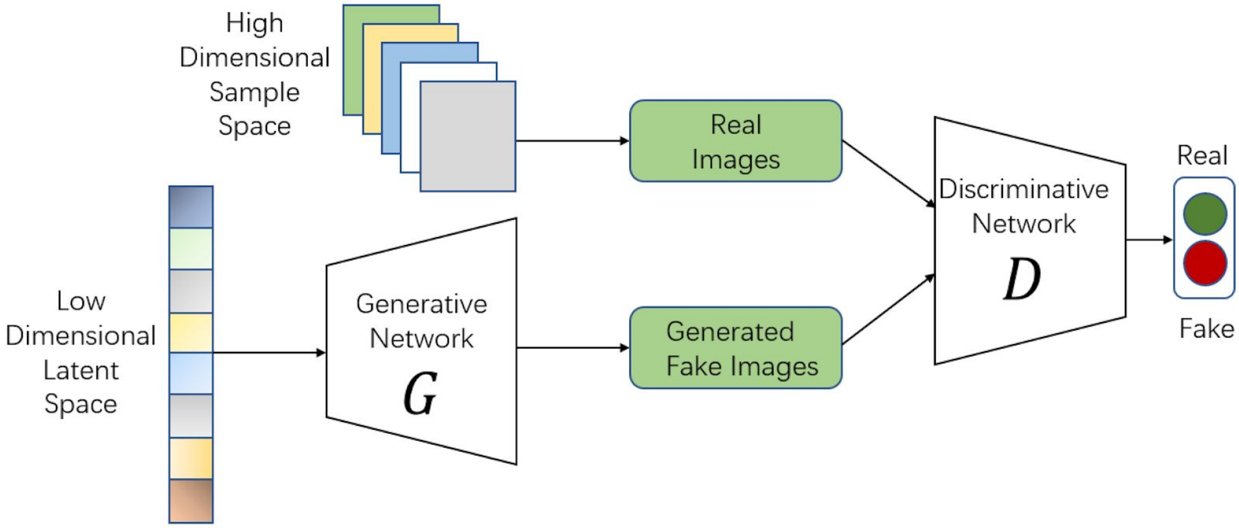
Super Resolution Generative Adversarial Network

Super Resolution is referred to the task of estimating a Higher resolution image from its counterpart Lower Resolution image. SRGAN inspired through Generative Adversarial Network (GAN) is an architecture proposed by researchers at Twitter whose motive was to recover fine texture from image when we super resolve the images at large upsampling factors without the quality of image being compromised.Other works were mainly focused on minimizing the mean squared reconstruction error which lacks high-frequency details and are perceptually unsatisfying in the sense that they fail to match the fidelity expected at the higher resolution. It is concluded that SRGAN has better accuracy and generates image more pleasing to eyes as compared to previous work which is achieved through a perceptual loss function which consists of an adversarial loss and a content loss.

Architecture:

Inspired through GAN even SRGAN has two parts of its architecture Generator and Discriminator where Generator tries to generate refined images as much as possible in order to fool the generator ,while the Discriminator tries to differentiate between the Images and decides weather it is coming from input dataset or it one of those images which are generated by Generator. The image below shows the architecture of Generative Adversarial Network:



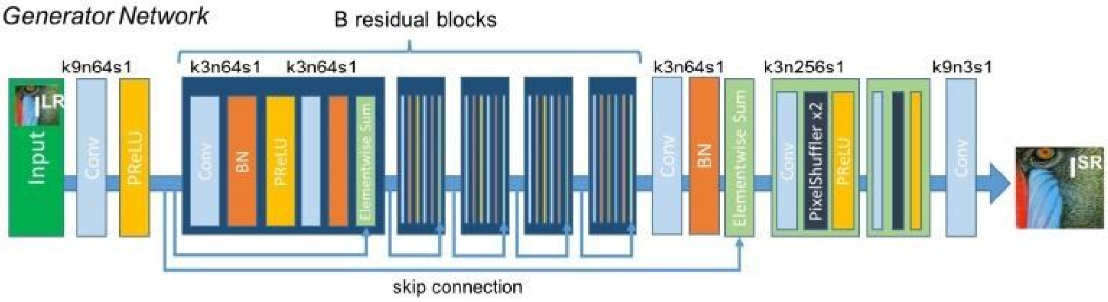
The loss function used in Gan's architecture also plays a vital role apart from Generator and Discriminator. The three main pillars of any GAN's architecture is its Generator, Discriminator and Loss function used.

Generator:

In SRGAN we use a GAN architecture which uses residual network instead of deep convolutional network because of the skip connections that residual network allows due to which we can go train our model could substantially deeper while being easy to train.

Super Resolution Generative Adversarial Network

Let's have a look at the Generator Architecture used in the paper proposed in SRGAN



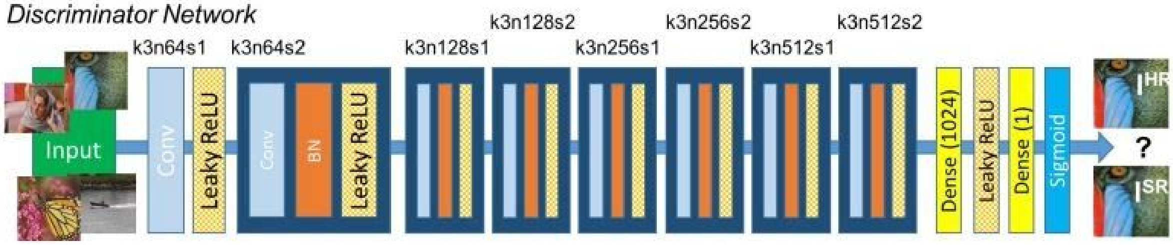
I^{SR} : Super-resolved Image (rW x rH x C)
I^{LR}: Low Resolution Image (W x H x C)
I^{HR}: High Resolution Image (rW x rH x C)

The term $\mathbf{I^{SR}}$ denotes super-resolved image from a low-resolution input image $\mathbf{I^{LR}}$. Here we describe $\mathbf{I^{LR}}$ by a real-valued tensor of size **W × H × C** and $\mathbf{I^{HR}}$, $\mathbf{I^{SR}}$ by **rW × rH × C** respectively where r is the upsampling factor. At the core of our very deep generator network there are B residual blocks with identical layout.Specifically, they have used two convolutional layers with small 3×3 kernels and 64 feature maps followed by batch-normalization layers and ParametricReLU as the activation function. We increase the resolution of the input image with two trained sub-pixel convolution layers which could be done through Depth_to_space layer in Tensorflow also referred to as Spatial upsampling.

During training, a High resolution image is taken from the input dataset and then downsample through a factor of r and then that image is used to create a super-resolved image through a generator($\mathbf{I^{SR}}$). This image is then passed on to the Discriminator which in turn tries to discriminate in between a Super-resolved image and a High-Resolution Image and generates Adversarial loss.

Discriminator:

To discriminate real HR images from generated SR samples they train a discriminator network.
Let's have a look at the Discriminator Network used in SRGAN:



Super Resolution Generative Adversarial Network

The discriminator architecture used in this paper is similar to DC- GAN architecture with LeakyReLU (α = 0.2) as activation and max-pooling is avoided throughout the network. The network contains eight convolutional layers of 3×3 filter kernels, increasing by a factor of 2 from 64 to 512 kernels as in the VGG network. Strided convolutions are used to reduce the image resolution each time the number of features is doubled. The resulting 512 feature maps are followed by two dense layers and a leakyReLU applied between and a final sigmoid activation function to obtain a probability for sample classification.

Loss Function:

The Loss Function used in SRGAN is a perceptual loss function which consists of an adversarial loss and a content loss. The adversarial loss pushes our solution to the natural image manifold using the discriminator network discussed above that is trained to differentiate between the super-resolved images and original photo-realistic images. In addition, we use a content loss motivated by perceptual similarity instead of similarity in pixel space.
We formulate the perceptual loss as the weighted sum of a content loss ($\mathbf{L^{SR}_{X}}$) and an adversarial loss component as:

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

- **Content Loss:** Most Image Super Resolution Techniques uses Pixel wise MSE Loss which many state-of-the-art approaches rely and is given as :

$$l^{SR}_{MSE} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I^{HR}_{x,y} - G_{\theta_G}(I^{LR})_{x,y})^2$$

However MSE Loss results in frequency content due to which the achieved images are perceptually unsatisfying with overly smooth textures.Instead of relying on pixel-wise losses we use a loss function that is closer to perceptual similarity. We define the VGG loss based on the ReLU activation layers of the pre-trained 19 layer VGG network.We define the VGG loss as the euclidean distance between the feature representations of a reconstructed image $\mathbf{G_{\theta G}}$ **($\mathbf{I^{LR}}$)** and the reference image $\mathbf{I^{HR}}$:

Super Resolution Generative Adversarial Network

$$l^{SR}_{VGG/i.j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2$$

$G_{\theta G}$ (I^{LR}) : Represents Generator’s Output with Lower Resolution Image as Input
I: Denotes Image
L: Denotes Loss

Here $W_{i,j}$ and $H_{i,j}$ describe the dimensions of the respective feature maps within the VGG network.

- **Adversarial Loss:** Adversarial Loss encourages our network to favor solutions that reside on the manifold of natural images, by trying to fool the discriminator network. The generative loss **L^{SR}_{Gen}** is defined based on the probabilities of the discriminator **$D_{\theta D}(G_{\theta G}(I_{LR}))$** over all training samples as:

$$l^{SR}_{Gen} = \sum_{n=1}^{N} -\log D_{\theta_D}(G_{\theta_G}(I^{LR}))$$

$D_{\theta D}(G_{\theta G}(I_{LR}))$: Probability that the reconstructed image **$G_{\theta G}(I_{LR})$** is a natural HR image

For better gradient behavior they minimize **$- \log D_{\theta D}(G_{\theta G}(I_{LR}))$** instead of **$\log[1 - D_{\theta D}(G_{\theta G}(I_{LR}))]$**

Results:

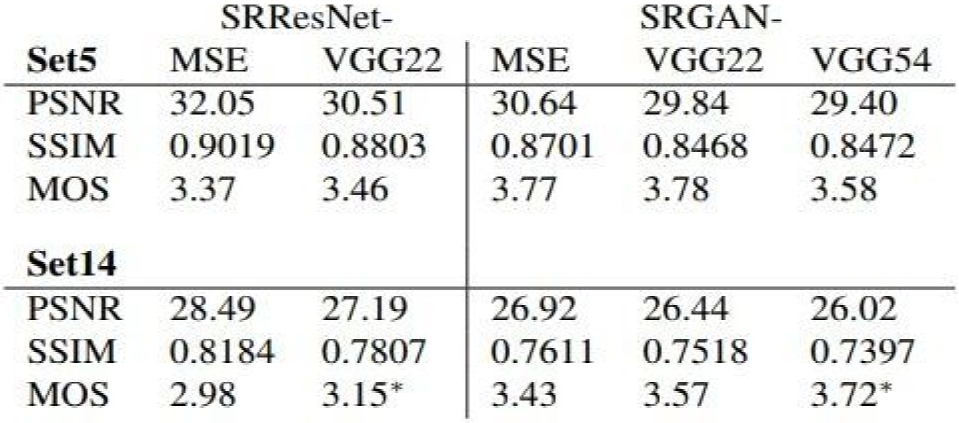
The given architecture is then tested on 3 widely used benchmark datasets **Set5** , **Set14** and **BSD100** and then further compared with the results obtained through SRResNet.These experiments performed on 4x up sampling of both rows and columns.The metrics used for comparison are:

- **PSNR:** Peak Signal to Noise Ratio
- **SISM:** Structure Similarity Index
- **MOS:** Mean Opinion Score

Super Resolution Generative Adversarial Network

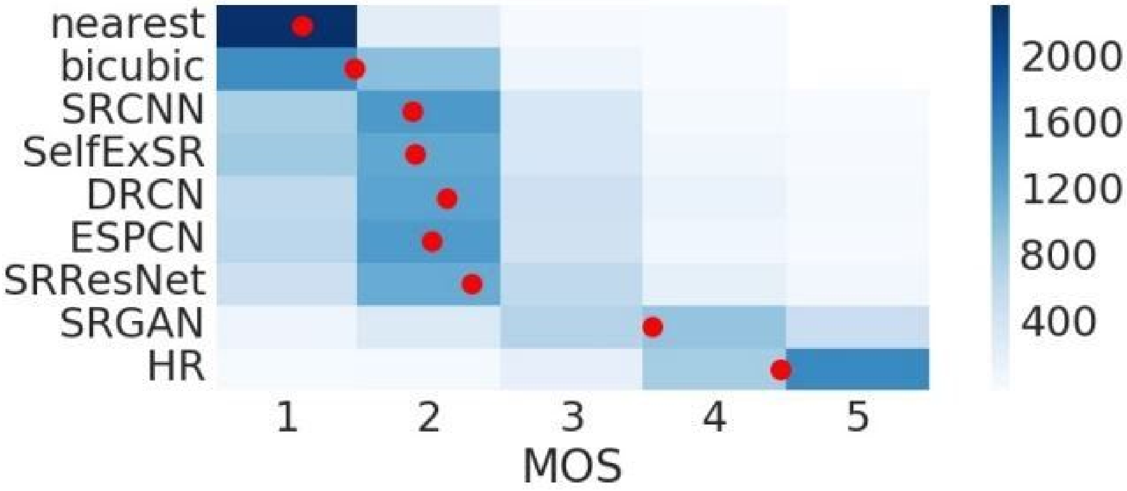
Also the authors have used MSE: Content loss based on simple Mean Squared Error , VGG22: Content loss defined on feature maps representing lower-level features , VGG54: Content loss defined on feature maps of higher level features from deeper network layers with more potential to focus on the content of the images.

The Table below shows and compares results of various experiments done by the authors of SRGAN's:



	SRResNet-			SRGAN-	
Set5	MSE	VGG22	MSE	VGG22	VGG54
PSNR	32.05	30.51	30.64	29.84	29.40
SSIM	0.9019	0.8803	0.8701	0.8468	0.8472
MOS	3.37	3.46	3.77	3.78	3.58
Set14					
PSNR	28.49	27.19	26.92	26.44	26.02
SSIM	0.8184	0.7807	0.7611	0.7518	0.7397
MOS	2.98	3.15*	3.43	3.57	3.72*

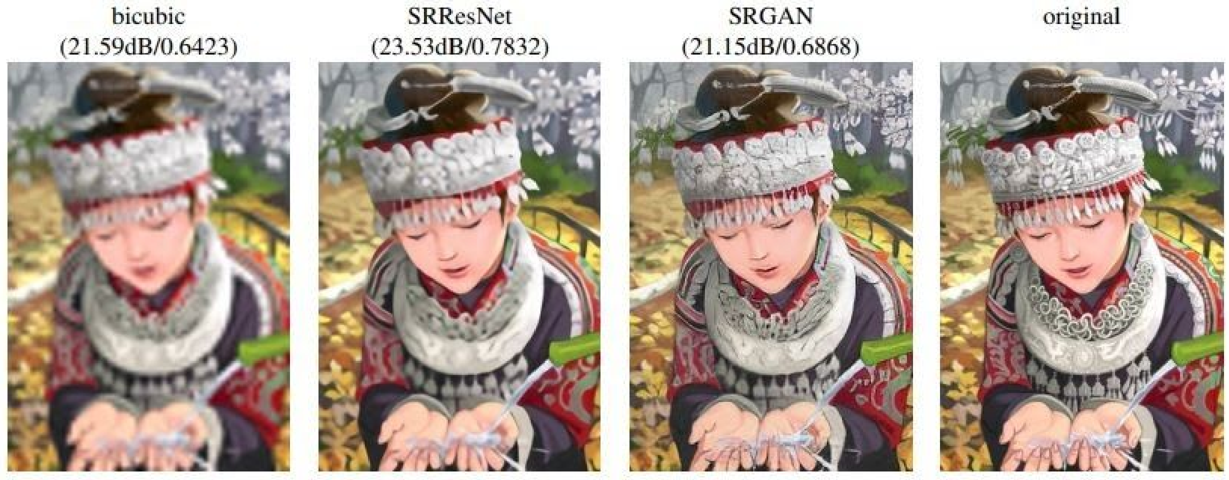
The distribution below shows the MOS results of different techniques along with SRGAN trained on BSD100 dataset. For each method 2600 samples (100 images × 26 raters) were assessed. Mean shown as red marker.



SRGAN was able to generate state-of-the-art results which was validated with extensive Mean Opinion Score (MOS) tests on three public benchmark datasets.

Super Resolution Generative Adversarial Network

Let’s have a look at practical results of SRGAN which was uploaded by the authors of SRGAN.



bicubic (21.59dB/0.6423)

SRResNet (23.53dB/0.7832)

SRGAN (21.15dB/0.6868)

original

We can observe that the image produced through SRResNet has a smoothness involved while the obtained image through SRGAN is pretty sharp and clear.