International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** I **Month of publication:** January 2020

DOI: http://doi.org/10.22214/ijraset.2020.1058

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 8 Issue I, Jan 2020- Available at www.ijraset.com

Super Resolution of Videos using SRGAN

Jash Shah[1], Deepak Thakur[2], Aradhya Rakvi[3], Shikha Malik[4], Pralhad Gopal Singh[5]
[1, 2, 3, 5]*Electronics and Telecommunication, Atharva College of Engineering Mumbai, India*
[4]*Assistant Professor, EXTC Department, Atharva College of Engineering Mumbai University*

Abstract: The primary objective of our project is to generate the Super resolved images. To enhance the quality of images we are using SRGAN techniques. There are various methods of image transformation where the computing system receives some input and transforms it in output image. Previous techniques of up scaling were based on minimizing the mean squares reconstructed error. Generative Adversarial Networks are the deep neural network architectures comprised of two networks Generator and Discriminator. GANs are about creating, like drawing a portrait or composing a symphony. SRGAN provides several benefits over other techniques it proposes a perceptual loss function which consists of adversarial loss and content loss. The two main blocks are Generator and Discriminator. The Discriminator discriminates between real HR images from generated super resolved images. The Generator function is to train a propagating model. Adversarial loss function uses discriminator network which is already trained to discriminate between the two images. The content loss function uses perceptual similarity instead of pixel space similarity. One of the best things about SRGANs is that they generate data that is similar to real data. SRGANs learn the internal representations of data to generate the upscale images. The neural network is successful in recovering the photo realistic textures from downgraded images. The SRGAN techniques lack high peak to signal noise ratio but provides high efficiency and visual perception. Combining the perceptual and adversarial loss will generate a high quality super resolved image. During training phase perceptual losses measure image similarities more robustly than per-pixel losses. Perceptual loss functions measures the high level perceptual and semantic differences between the images. Our method uses semantic content losses which has a VGG network and an edge promoting adversarial losses for edges.

I. INTRODUCTION

Super Resolution means the generation of high resolution image from low resolution image. Super Resolution provides large computational efficiency and it is now one of the advanced technologies in the field of image processing. Machine Computing need a network for the computation and performing calculations it follows a program to perform a sequences of mathematical and logical operations. Fig.1 shows a bird which is primarily in low resolution and it is converted into a HR image by two different methods. The Bicubic method and SRGAN are two methods used and both of the above methods have tremendous differences. SRGAN result shows higher efficiency and texture realization which is not seen in bicubic. In this paper we have two losses content and adversarial.

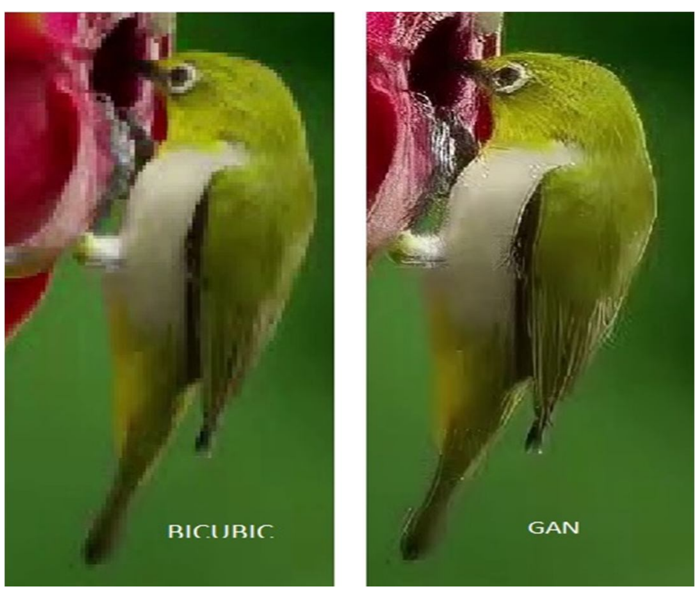


Figure.1.An example of Super-Resolution with Bicubic and SRGAN techniques

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 8 Issue I, Jan 2020- Available at www.ijraset.com

Perceptual loss is based on VGG based content loss .However in the recent development there are many work and research papers which shows that perceptual loss functions are based on the dissimilarities between high resolution level image which are extracted by from trained neural networks. These neural networks are trained before and image generation is based on reducing the loss. Typically SR algorithms are commonly based on the minimizing the MSE between the regenerated High Resolution images and the original one. PSNR should be maximized in order to have lower noise and high perception. Both MSE and PSNR cannot alone determine the image quality and efficiency, it is the per pixel image difference which is also a dominating factor in image efficiency and resolution. In our paper we have shown that the Adversarial training can improve the perceptual quality of the single image. There is a big focus on SRGAN which shows higher efficiency and up scaling which is achievable up to 4x.

II. RELATED WORK

Convolution Neural Networks. CNN based SR has shown great performance in many applications. There are different types of CNN based SR such as SRCNN(Super Resolution Convolutional Neural Networks), FSRCNN(Fast Super Resolution Convolutional Network), VDSR(Very Deep Super Resolution Convolutional Networks), and DRCN(Deeply Recursive Convolutional Networks) for the single image super resolution. SRCNN is the first method for the SISR; FRSCNN is the next version of SRCNN which is based on increasing the velocity of the HR reconstruction. Hence the structure of it is very complicated.VDSR shows some improvement over the previous methods and increases the depth of the network. DRCN uses intermediate layers with multisupervised methods. The methods used by CNN shows both qualitative and quantitative improvement. All the methods which uses CNN does not use the guiding HR image for the transformation of the LR image SRResNet. It employs Residual network architecture and contains the series of resblocks**.** These blocks modify the input LR image by up sampling, SRResnet employs 16 residual blocks with 64 channels and provides a global skip connection. It also provides pixel shuffle upsampling.

III. METHOD

For Video super resolution we require a lower resolution video as the input .The LR video converted into frames and then individual frame are enhance by using neural networks method. Finally enhance frames are converted into video. The format of frame is in png or jpeg and format of video are mp4/av. In 30 second of video have approximate 770 frames are present. Video Super resolution requires a High Resolution frames from the lower resolution frames. The lower resolution frames should generate the continues frames. So in our work we are going to Use SRGAN algorithm to generate super resolved videos. The previous works were based on Convolutional Neural Networks approaches which have shown the fine results. The Convolutional Neural Networks based approaches were unable to generate the texture details which are very stringent part of our Super resolution technique. The Deep learning methods were based on the high PSNR and motional compensation. These methods fails to remove the blurring effect and the output High Resolution frames consist of less texture realistic features. As we move to the SRGAN based approach our super resolution output generates more texture related and realistic features which favours it over the previous approaches. Super Resolution using Generative Adversarial Networks uses adversarial loss function and content loss function which are the part of the perceptual loss functions; it eliminates the averaging error in the output part. For the Single image super resolution we require a high resolution image and a low resolution image. SRGAN provides a stringent architecture consisting of various blocks and loss functions. The previous methods were based on MSE based loss. But in GAN networks we replace it with VGG architecture. Hence SRGAN consists of MSE loss plus Perceptual loss plus GAN loss; we will describe all the three losses in the further sections. We require a Gaussian filter which has high resolution image which is downsampled to get a low resolution image, the downsampling conversion factor is denoted by coefficient r. The training phase starts with the training of the generator function G that has input as low resolution image and its high resolution image. We train the generator network as feed forward architecture of CNN.

A. GAN Architecture

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

The GAN architecture follows the equation given below (1)

The above equation consists of intensity of low resolution and a high resolution images. The intensity is described by theta functions consisting of min and max of generator and discriminator. The logarithmic function determines the discriminator and generator's output. The basic goal is to train a generative model and fool the discriminator model however the discriminator differentiates between the high resolution and super resolved images. The generator can be trained by various methods and hence its

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 8 Issue I, Jan 2020- Available at www.ijraset.com

solutions are varying so that discriminator cannot differentiate between real and fake images. Hence the adversarial network architecture is very different comparing to the MSE based super resolution models. The Generator and discriminator architecture figure will be described in the next page. The Generator consists of various residual blocks and a specialized skip connection. Skip connection are utilized when the function has to skip from one point to another without having its effect in the output. The input to generator is a low resolution image which is followed by convolution operation; the convolution operation will convolved the LR image with a certain pattern which consists of a matrix. The k9n64s1 means that the kernel value is 9 and there are 64 feature maps, s1 means only one stride convolution is applied so that input that is changed can be regenerated. The next blocks are residual blocks which consist of various layers and batches. The first layer of the residual block is convolution layer for the convolution operation. The second layer consists of batch normalization function and the third one is the activation function SRGAN uses PReLU as its activation unit. Then there is a repetition of these blocks and the final one is element wise sum. The residual blocks change its kernel size and its size changes to 3 and there is no change in the feature maps due to the system requirements. Skip connection as described earlier will skip its step if required within the residual blocks. At the final stage of the generator there is change in no of feature maps which change its value to 4 times the previous feature map value. A final layer of convolution is required to convert the image into super resolved image which consists of kernel of size as of input, however the featuremaps decreases. The next architecture is of the discriminator which needs to be trained so that it can discriminate the HR images and the generated images by generator. The Activation function used by the Discriminator is LeakyReLU which uses different mathematical methods for activation. The Equation 1. Need to be solved by discriminator. The discriminator consists of eight convolution layers, the size of kernel is 3x3 and the no of features maps are increased to 512.Sigmoid function is used to predict the probability at the output of discriminator.

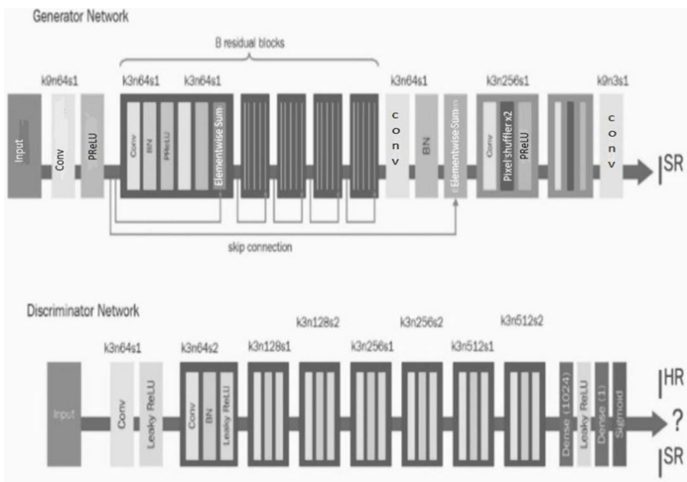


Fig.2.Architecture of Generator and Discriminator

B. Perceptual Loss Function

The Perceptual loss function consists of two losses adversarial and content loss

$$l^{SR} = \underbrace{\underbrace{l^{SR}_{X}}_{\text{content loss}} + \underbrace{10^{-3} l^{SR}_{Gen}}_{\text{adversarial loss}}}$$

(2)

The above equation shows that the perceptual loss is the sum of the both losses.

1) *Content Loss:* The content loss is described by the Equation 3.

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

(3)

The equation 3 describes the MSE loss which is pixel wise only. The content loss focuses on the perceptual similarity instead of relying on the Mean Square loss.. The Activation function used in Discriminator is LeakyReLU which defines the VGG based loss.

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

The VGG loss is described in Equation 4

(4)

2) Adversarial Loss: The second type of the perceptual loss is called adversarial loss

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

(4)

The most important function of adversarial loss is to make the Super Resolved Images with high photo realistic content. It also aims on high perceptual quality. The Equation 4 describes the adversarial loss at the generator which is a mathematical function. The logarithmic function describes the logarithmic behavior of the generator output. The logarithmic function should be minimized in order to get better gradient behavior. The Adversarial loss function will train the Generator network to get an output of highly efficient super resolved images. The Sigmoid Activation as described earlier in GAN Architecture will produce the adversarial loss.

IV. FUTURE WORK

In this paper we have proposed an interesting solution namely SRGAN. It is observed that PSNR value fails to attain image quality and perception. The Generator and Discriminator are highly trained networks to obtain high efficiency. The trends will change further by employing Enhanced methods using SRGAN which will have high perceptual similarity. The future work will also depend on the loss functions proposed by the GAN, it can further uses identity loss which can sum up with all types of loss. We can use the perceptual loss for the other transformation objectives such as segmentation and colorization. For the Video Super Resolution a similar algorithm can be used. Especially for the Video games the SRGAN are used to provide high and realistic graphics with real time experience.If you have pictures taken from a low-resolution camera, GANs can help you generate high-resolution images without losing any essential details. This can be useful on websites. For Face Ageing GAN is very useful tool which can be very useful for both the entertainment and surveillance industries. It is particularly useful for face verification because it means that a company doesn't need to change their security systems as people get older. An age-cGAN network can generate images at different ages, which can then be used to train a robust model for face verification.

V. CONCLUSION

SRGAN is one of the advanced super resolution techniques. SRGAN involves training of GAN network and have adversarial loss and content loss functions augmentation. SRGAN can achieve the 4x upscaling with high perception. However the PSNR value of SRGAN is quite less than previous methods but it does not affect the image quality and superiority. Hence the proposed super resolution technique is effective.

REFERENCES

[1] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningh. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network , 2017 IEEE Conference on Computer Vision and Pattern Recognition.

[2] Yang Yang, Ping Bi, Super-Resolution. Plate Image Super-Resolution Based on Convolutional Neural Network, 2018 3rd IEEE International Conference on Image, Vision and Computing

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition,Microsoft Research

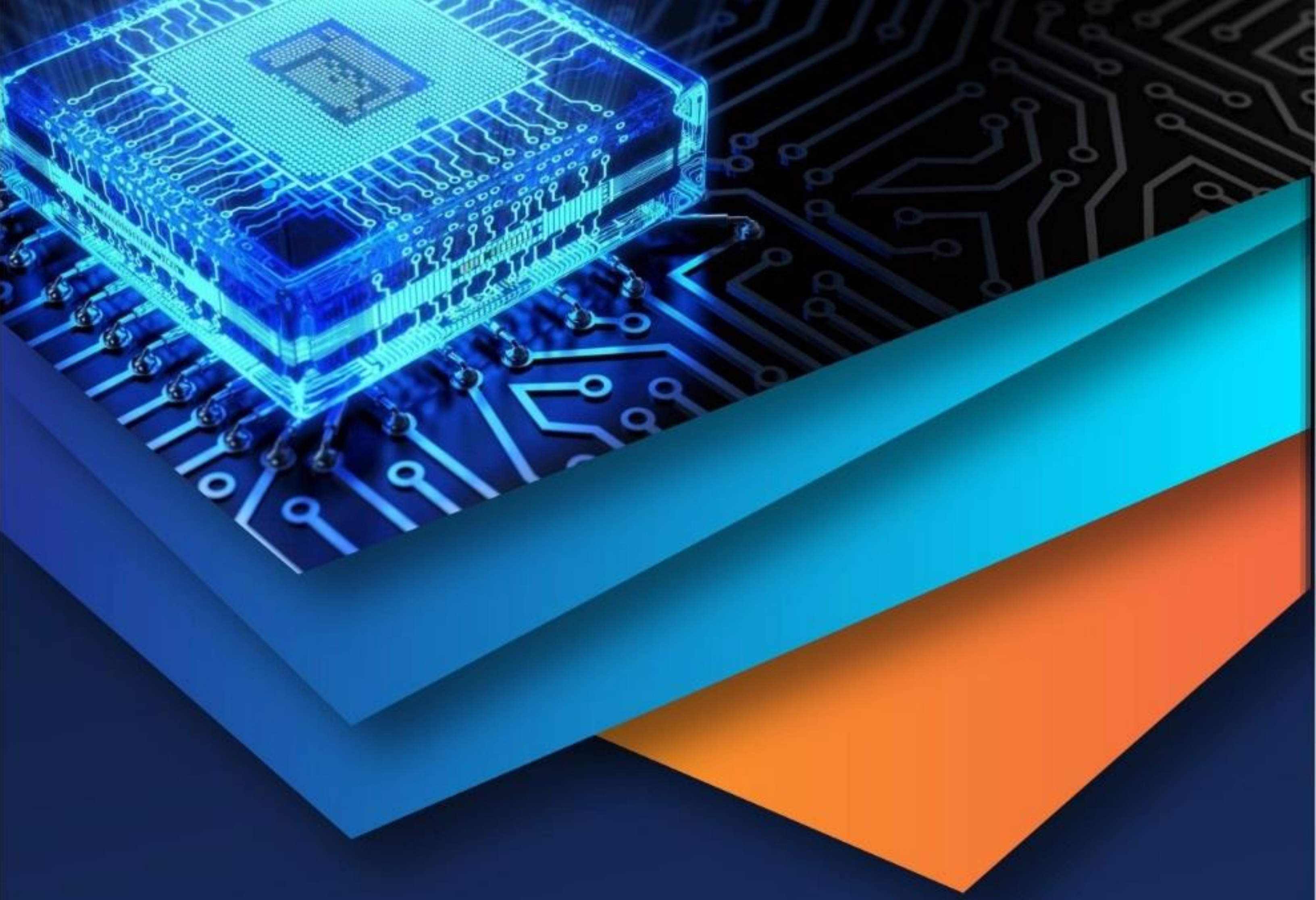
[4] Justin Johnson, Alexandre Alahi, Li Fei-Fei. Perpetual losses for Real Time Style Transfer and super resolution Standford University,27 March 2016

[5] Berk Dogan, Shuhang Gu, Radu Timofte. Exemplar Guided Face Image Super- Resolution without Facial Landmarks, Computer Vision Lab, D-ITET, ETH Zurich.

[6] Wenning Yanf,Xuechen Zhang, Yapeng Tian, Wei Wang,Jing-Hao Zue, Oqingmin Liao. Deep Learning for Single Image Super-Resolution :A Brief Review, 12 Jul 2019 IEEE

[7] Mengyu Chu, You Xie, Laura Leal-Taix´e and Nils Thuerey. Temporally Coherent GANs for Video Super-Resolution(TecoGAN), Technical University of Munich.

[8] Jinaan Li, Xiadong Liang, Yunchao Wai, Tingfa Fu, Jiashi Feng, Shuicheng Yen. Perceptual Generative Adversarial Networks for Small Object Detection, CVPR 2017



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY