



AISSMS
INSTITUTE OF INFORMATION TECHNOLOGY
ADDING VALUE TO ENGINEERING



Department Of Computer Engineering

A SEMINAR REPORT

ON

Image Super Resolution Using GAN

**SUBMITTED TO THE DEPARTMENT OF COMPUTER
ENGINEERING AISSMS IOIT**

TE COMPUTER ENGINEERING

SUBMITTED BY

STUDENT NAME	ERP NO.
AKASH METE	52
HARSH SHAH	67
KAUSTUBH KABRA	38
ONASVEE BANARSE	09



2021 -2022



Department of Computer

Engineering CERTIFICATE

This is to certify that the project report
“Image Super Resolution Using GAN”
Submitted by

STUDENT NAME	ERP NO.
AKASH METE	52
HARSH SHAH	67
KAUSTUBH KABRA	38
ONASVEE BANARSE	09

Our students at this institute and the work has been carried out by them under the supervision of **Prof. Shilpa Pimpalkar** and it is approved for the partial fulfillment of the Department of Computer Engineering AISSMS IOIT.

(Prof. Shilpa Pimpalkar)
Guide
Department of Computer Engineering

(Dr. S. N. Zaware)
Head,
Department of Computer Engineering

Place: AISSMS IOIT, Pune

Date: 30/11/2021

ABSTRACT

Super-resolution reconstruction is an increasingly important area in computer vision. To alleviate the problems that super-resolution reconstruction models based on generative adversarial networks are difficult to train and contain artifacts in reconstruction results, we propose a novel and improved algorithm. Methods: This paper presented the SRGAN (Super-Resolution Generative Adversarial Networks Combining Texture Loss) model which was also based on generative adversarial networks. We redefined the generator network and discriminator network. Firstly, on the network structure, residual dense blocks without excess batch normalization layers were used to form a generator network.

Visual Geometry Group (VGG)19 network was adopted as the basic framework of the discriminator network. Secondly, in the loss function, the weighting of the four loss functions of texture loss, perceptual loss, adversarial loss and content loss was used as the objective function of the generator. Texture loss was proposed to encourage local information matching. Perceptual loss was enhanced by employing the features before the activation layer to calculate. Adversarial loss was optimized based on WGAN-GP (Wasserstein GAN with Gradient Penalty) theory.

Content loss was used to ensure the accuracy of low-frequency information. During the optimization process, the target image information was reconstructed from different angles of high and low frequencies. Results: The experimental results showed that our method made the average Peak Signal to Noise Ratio of reconstructed images reach 27.99 dB and the average Structural Similarity Index reach 0.778 without losing too much speed, which was superior to other comparison algorithms in objective evaluation index.

What is more, SRGAN significantly improved subjective visual evaluations such as brightness information and texture details. We found that it could generate images with more realistic textures and more accurate brightness, which were more in line with human visual evaluation. Conclusions: Our improvements to the network structure could reduce the model's calculation amount and stabilize the training direction. In addition, the loss function we present for the generator could provide stronger supervision for restoring realistic textures and achieving brightness constancy. Experimental results prove the effectiveness and superiority of the SRGAN algorithm.

ACKNOWLEDGEMENT

We present the seminar report as part of the curriculum of the T.E. Computer Engineering. We wish to thank all the people who gave us unending support right from when the idea was conceived. We express sincere and profound thanks to our Seminar Guide **Prof Shilpa Pimpalkar**, and **HOD Mrs S. N. Zaware**, who is always ready to help with the most diverse problems that we have encountered along the way. We express sincere thanks to all staff and colleagues who have helped directly or indirectly in completing this seminar successfully.

AISSMS IOIT, Pune.

INDEX

Contents

ABSTRACT	3
ACKNOWLEDGEMENT	4
INDEX	5
PROBLEM STATEMENT	6
INTRODUCTION	7
DISCUSSION BASE PAPER	10
SRGAN (Super Resolution GAN)	13
LITERATURE SURVEY	17
FUTURE ENHANCEMENT	19
CONCLUSION	20
REFERENCES	21

PROBLEM STATEMENT

To recover or restore a high-resolution image from a low-resolution image. There are many forms of image enhancement which includes noise-reduction, up-scaling image, and color adjustments. This post will discuss enhancing low resolution images by applying deep networks with adversarial networks (Generative Adversarial Networks) to produce high resolutions images.

Our main target is to reconstruct super resolution images or high-resolution images by up-scaling low resolution images such that texture detail in the reconstructed SR images is not lost.

INTRODUCTION

Image Resolution -

The term resolution in image processing corresponds to the amount of information contained in an image that can be used to judge the quality of the image and image acquisition/ processing devices. Resolution can be classified into several categories such as Pixel or Spatial resolution, Spectral Resolution, Temporal resolution, Radiometric resolution. For this project, we will be dealing with spatial resolution, and the term resolution used henceforth will imply spatial resolution. Spatial Resolution is the number of pixels that are used to construct the image and is measured by some pixel columns (width) \times number of pixel rows (height), say for, e.g., 800×600 .

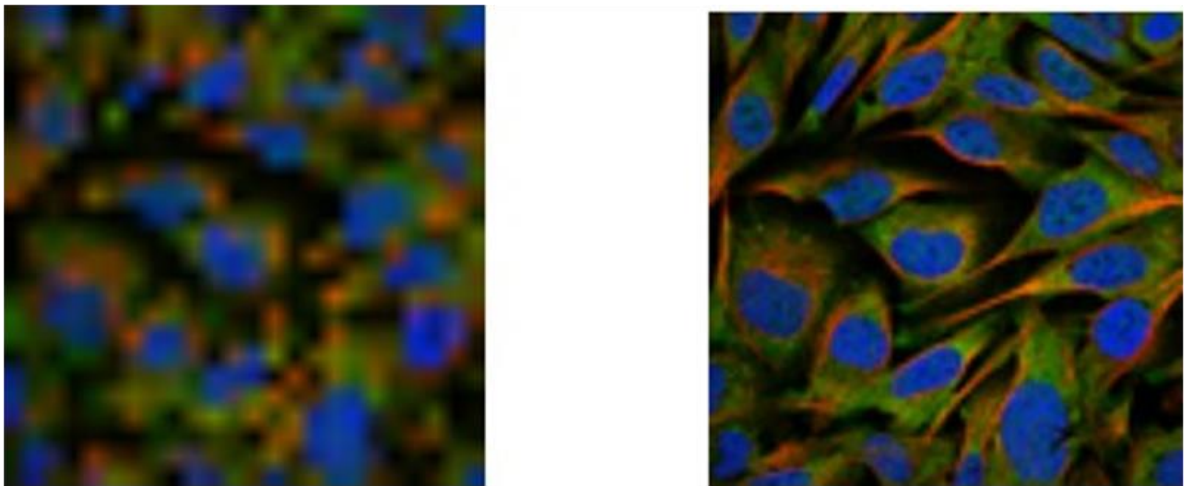


Figure 2.1: As can be seen in the above figures, L has more spatial resolution than R.

Pixels -

They are the smallest addressable parts of an image. Each image can be considered as a matrix consisting of several pixel values. Every pixel stores a value proportional to the light intensity at a particular location, and for an 8-bit grayscale image, the pixel can take values from 0 to 255.

Low resolution -

A low-resolution image implies that the pixel density of the image is small thereby giving fewer details.

High resolution -

A high-resolution image implies that the pixel density of the image is high leading to more details.

Super-resolution -

SR is constructing an HR image from a single/multiple LR image.

Super resolution methods can be categorized into two categories based on the number of images involved

-

- a) Multiframe super-resolution
- b) Single image super- resolution.

Multiframe super-resolution -

This method utilizes multiple LR images to re- construct an HR image. These multiple images can come from various cameras at separate locations capturing a scene or several pictures of the same scene. These multiple input LR images contain the same information, however the information of interest is the subpixel shifts that occur due to movement of objects, scene shifts, motion in imaging systems (e.g., satellites) If the different LR image inputs have different subpixel shifts then this unique information contained in each LR image can be leveraged to reconstruct a good HR image.

Single image super-resolution (SISR) -

In SISR, the super resolving algorithm is applied to only one input image. Since in most cases there is no underlying ground truth, the significant issue is to create an acceptable image. The majority of the SISR algorithms employ some learning algorithms to hallucinate the missing details of the output HR image utilizing the relationship between LR and HR images from a training database.

The SR reconstruction problem can be formulated regarding an observation model as shown in Figure which relates the HR image with the input LR images.

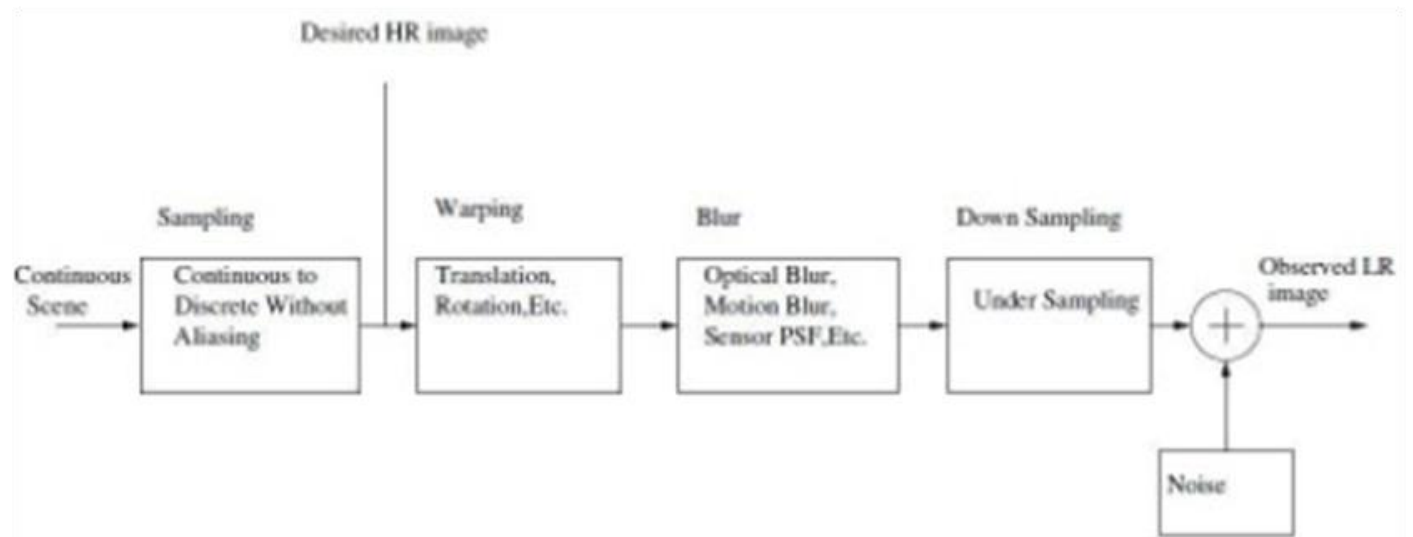


Figure: Observation model between an LR and HR image for a real imaging system.

First by continuous signal sampling the desired HR image is produced which is then subjected to translation, rotation leading to blurring due to optical, motion, imaging system motion, etc. Next, LR observation images are achieved by down sampling the blurred image.

Strategies to increase image resolution

The resolution of an image can be increased by either increasing the hardware capabilities of imaging devices or using a software/algorithmic approach.

- **Hardware Approach -**

One direct way to increase the spatial resolution is to increase the number of pixels per unit area by reducing the pixel size from sensor manufacturing techniques. But reducing the pixel size beyond a threshold (which is already achieved by current technologies) leads to the generation of shot noise as less amount of light is available for the decreasing number of pixels, degrading the image quality severely.

- **Software Approach-**

Techniques such as image interpolation, restoration, rendering, etc. are widely used in enhancing spatial resolution. Image interpolation approximates the color and intensity of a pixel based on the neighboring pixels values but fails to reconstruct the high-frequency details as noise is introduced in the HR image. Image restoration works by applying deblurring, sharpening and removing sources of corruption such as motion blur, noise, camera misfocus, etc. keeping the size of the input and output images the same. In image rendering, a model of an HR scene with imaging parameters is given which is used to predict the HR observation of the camera. Image super-resolution is a signal processing technique which considers a single/multiple LR image to construct an HR image. Apart from costing less than the hardware-based approaches, SR techniques can be applied to the existing imaging systems.

DISCUSSION BASE PAPER

Neural Networks

A neural network consists of an input layer, an output layer and at least one intermediate layer called the hidden layer. The layers consist of units called as neurons which are connected to neurons of the preceding layer in a directed acyclic graph, i.e., the neuron outputs from the previous can become the neuron inputs in the next layer. The most common layer type used in neural networks is the fully connected layer wherein all the neurons in the adjacent layers are pairwise connected with each other and connections between neurons of the same layer is prohibited.

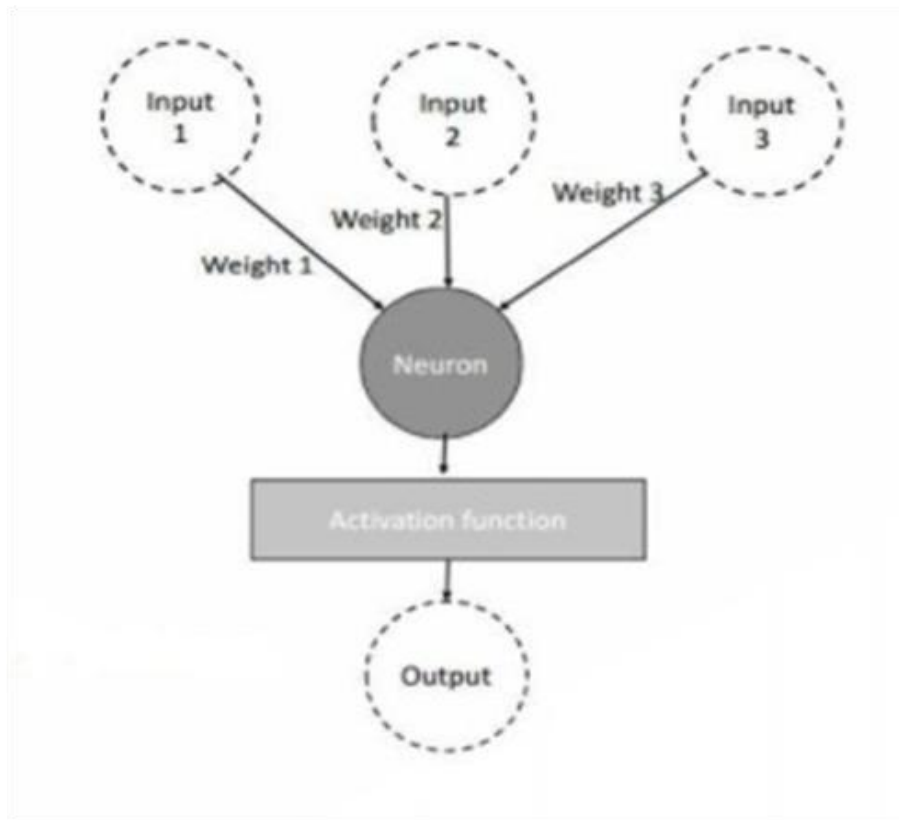


Figure: A graphical representation of a neuron with three inputs (Input1, Input2, Input3), their corresponding weights (Weight1, Weight2, Weight3), activation function and the resulting output

As shown in Fig a neuron computes the weighted sum of its inputs and a bias which is then applied to a linear/ nonlinear activation function. To represent the process formally, for given inputs x_i with its respective weights w_{ij} , a neuron y_j computes the weighted sum w_{ij} along with the bias b_j and applies an activation function f to the whole sum as shown below -

$$Y_j = f(w_{ij}x_i + b_j)$$

The activation function ' f ' introduces nonlinearity to the output of neuron y_j . This comes in handy since we want the network to account for the nonlinear patterns in the data and most of the real-world data has nonlinear structure.

Convolutional Neural Networks

Convolutional neural networks are a category of neural networks that have been proven to be remarkably effective in computer vision and classification applications such as object detection, self-driving cars, super-resolution, etc.

A convolutional layer consists of two-dimensional filters/kernels. The idea is to organize the neurons into units with inputs from local neighborhoods in the image which results in this filter. These filters are learned during the training of the algorithm, unlike the custom handcrafted features that are used in conventional machine learning algorithms. This operation is like the standard mathematical concept of convolution and is called that. The learned filters are convolved with the input image, and the feature responses that result from this are passed in the next processing layer as input. Neural networks that have such convolutional layers as cascaded stacks are called Deep convolutional neural networks. Some well-known architectures are Alexnet which uses five convolutional layers winning the best recognition performance at ILSVRC 2012, ResNet a 152-deep residual network winner of the best performance at ILSVRC 2015 almost entirely consists of convolutional layers.

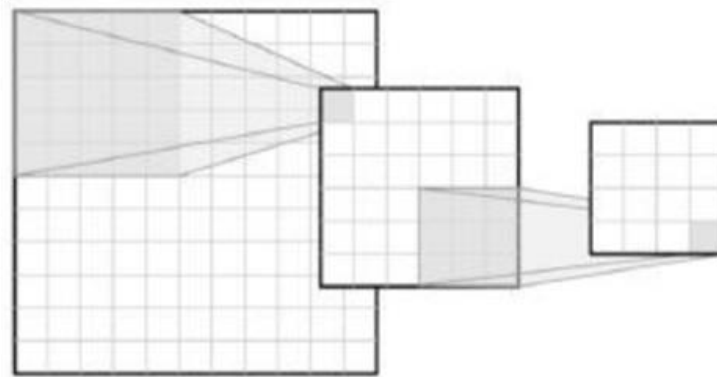


Figure 2.6: Visual representation of a convolutional neural network subsequently creating two feature maps, first with a filter of size 5 and then 3.

Generative Adversarial Networks

Generative adversarial networks (GAN) are a class of generative models used in unsupervised machine learning consisting of two networks (the generator and the discriminator) competing against each other in a zero-sum game framework. GAN uses a latent code that describes everything that's generated later. GAN's are asymptotically consistent, meaning if one can find the equilibrium point of the game-defining a GAN, it's guaranteed that the real distribution that generates the data is recovered and given an infinite amount of training data, the correct distribution is eventually rescued.

To describe the working of the GAN framework, we have two competing models in the sense of game theory where there is a game that has defined payoff functions with each player trying to maximize their

payoffs.

Within this game, one of the networks is the generator which is our primary model of interest that produces samples (generated samples/fake samples) with the aim of mimicking those that were from the real training distribution (real samples). The other competing model is the discriminator which inspects the sample and determines whether it's real or fake. During the training, images or any other samples are fed to the discriminator. The discriminator can be any differentiable function (usually a deep neural network) whose parameters can be learned by gradient descent. When the discriminator is applied to samples/images that come from the training set (real samples), its objective is to yield a value close to one, representing a high probability that the input was real rather than fake.

The discriminator is also applied to the samples generated from the generator (fake samples), and the goal of the discriminator in this scenario is to make the output as close to zero as possible implying the sample was fake. The generator is a differentiable function (usually a deep neural network) whose parameters can be learned by gradient descent. The generator function is applied on a sampled latent vector ' z ' which is nothing but noise at the start acting as a source of randomness helping the generator in producing a wide range of outputs. The generated images by the generator are then fed to the discriminator, and the generator tries to make the discriminator output one, fooling it into thinking the generated image is real when it is not. The readers can find more detailed technical information on GAN's here.

On a higher level, the generator can be viewed as a counterfeiter trying to create fake currency while the discriminator can be viewed as the police trying to ban fake currency while allowing real currency. As these two adversaries are forced to compete against each other, the counterfeiter must create more and realistic currency samples with the ultimate objective of fooling the police into believing that generated fake currency is real.

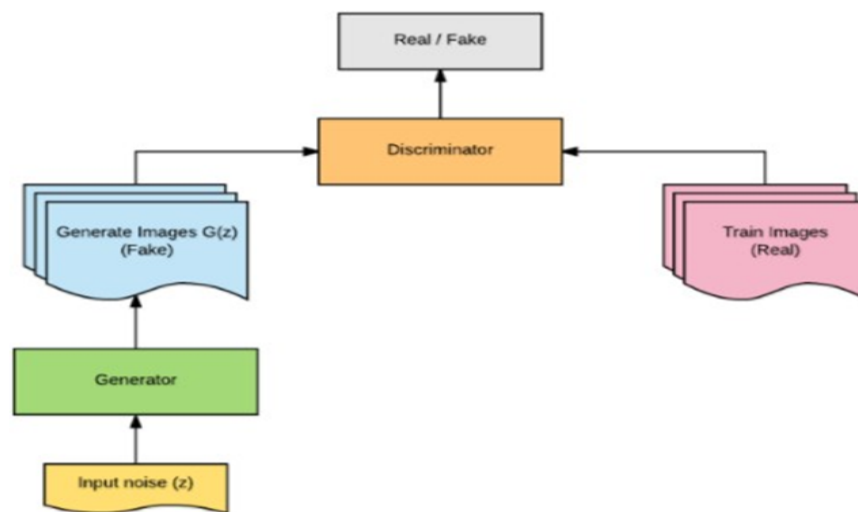


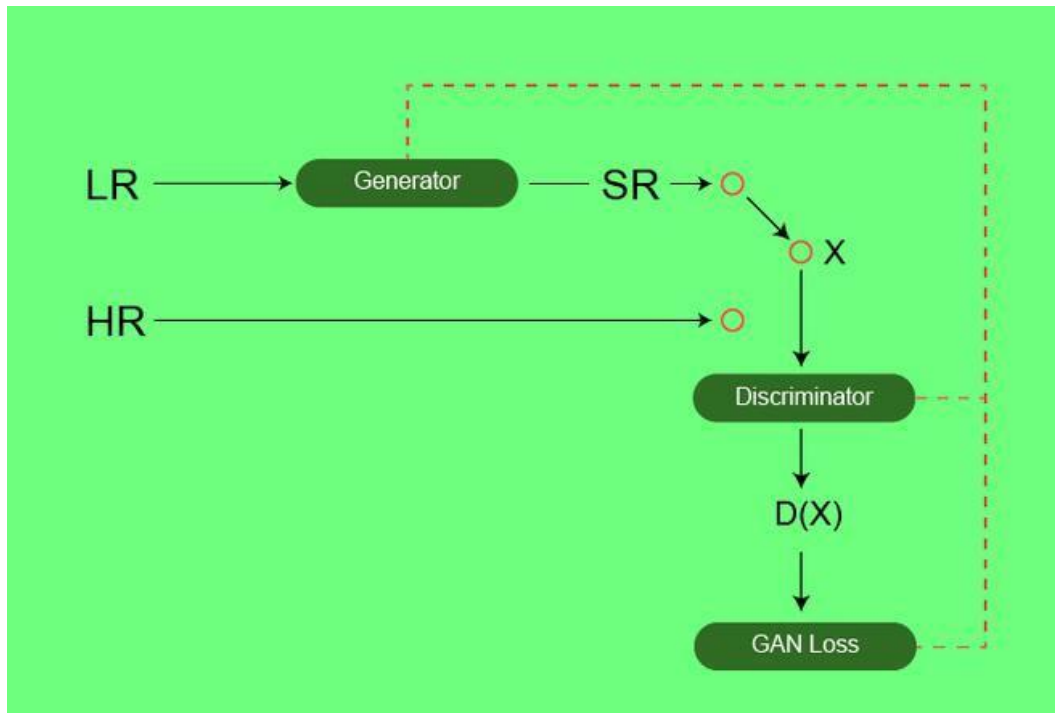
Figure 2.7: Visual representation of a generative adversarial network in action.

SRGAN (Super Resolution GAN)

SRGAN was proposed by researchers at Twitter. The motive of this architecture is to recover finer textures from the image when we upscale it so that it's quality cannot be compromised. There are other methods

such as Bilinear Interpolation that can be used to perform this task, but they suffer from image information loss and smoothing. In this paper, the authors proposed two architectures, one without GAN (SRResNet) and one with GAN (SRGAN). It is concluded that SRGAN has better accuracy and generates images more pleasing to eyes as compared to SRGAN.

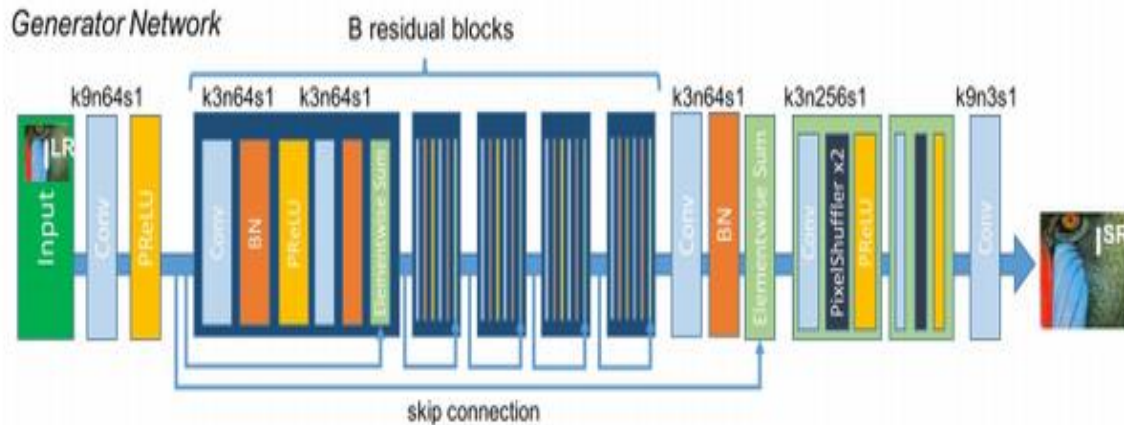
Architecture: Like GAN architectures, the Super Resolution GAN also contains two parts Generator and Discriminator where generator produces some data based on the probability distribution and discriminator tries to guess whether data coming from input dataset or generator. Generator then tries to optimize the generated data so that it can fool the discriminator. Below are the generator and discriminator architectural details:



SR-GAN architecture

Generator Architecture:

The generator architecture contains residual networks instead of deep convolutional networks because residual networks are easy to train and allows them to be substantially deeper to generate better results. This is because the residual network used a type of connection called skip connections.



There are B residual blocks (16), originated by ResNet. Within the residual block, two convolutional layers are used, with small 3×3 kernels and 64 feature maps followed by batch-normalization layers and ParametricReLU as the activation function.

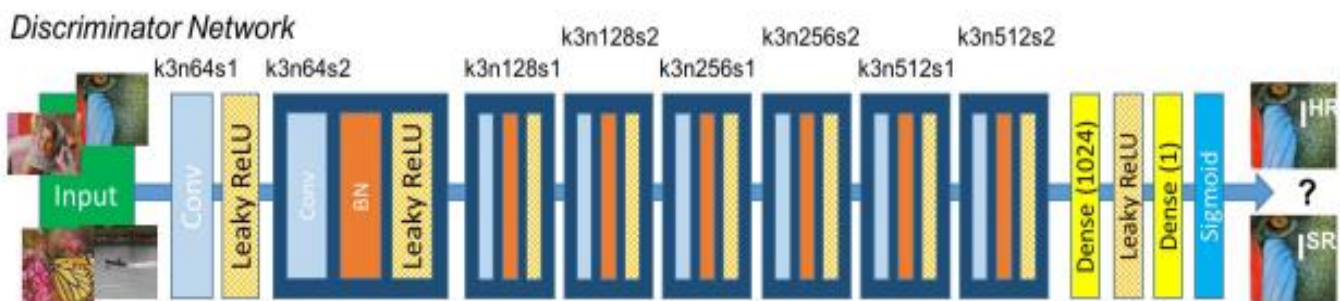
The resolution of the input image is increased with two trained sub-pixel convolution layers.

This generator architecture also uses parametric ReLU as an activation function which instead of using a fixed value for a parameter of the rectifier (alpha) like LeakyReLU. It adaptively learns the parameters of rectifier and improves the accuracy at negligible extra computational cost

During the training, A high-resolution image (HR) is down sampled to a low-resolution image (LR). The generator architecture then tries to up sample the image from low resolution to super-resolution. After then the image is passed into the discriminator, the discriminator tries to distinguish between a super-resolution and High-Resolution image and generate the adversarial loss which then back propagates into the generator architecture.

Discriminator Architecture:

The task of the discriminator is to discriminate between real HR images and generated SR images. The discriminator architecture used in this paper is like DC- GAN architecture with LeakyReLU as activation. The network contains eight convolutional layers with 3×3 filter kernels, increasing by a factor of 2 from 64 to 512 kernels. Strided convolutions are used to reduce the image resolution each time the number of features is doubled.



LITERATURE SURVEY

A literature survey represents a study of previously existing material on the topic of the report. This includes (in this order) –

- a. Existing theories about the topic which are accepted universally.
- b. Books written on the topic, both generic and specific.
- c. Research done in the field is usually in the order of oldest to latest.
- d. Challenges being faced and ongoing work, if available.

S.No.	Name/Title of Paper Referred	Name & Year of Publication	Description
1	Photo Realistic Single Image Super Resolution Using a Generative Adversarial Network	Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham Wenzhe Shi (CVPR) 2015	In this paper, the authors have discussed estimating a high-resolution image from a low-resolution image. They have discussed GANs and their types. They have given a detail study about adversarial neural networks and have also given a detailed account for the loss functions for the network.

S.No.	Name/Title of Paper Referred	Name & Year of Publication	Description
2	CCTV Surveillance Camera's Image Resolution Enhancement using SRGAN	Prof. M. Seshaiyah, Abhijith R Nair, Ektha Mallya, Shiv Dev (IRJET) JUNE 2020	In this paper, the authors have implemented the enhancement method for images that have a poor scale low resolution. Implemented method uses a machine learning algorithm, Super Resolution Generative Adversarial Networks (SRGAN) for achieving the goal of enhancement of images obtained from surveillance cameras. Super resolution of images allows us to obtain images with better resolution and less noise and hence provides the users with better experience of using the surveillance system.
3	Generative adversarial networks for single image super resolution in microscopy images	SAURABH GAWANDE (Master's Thesis at KTH Information and Communication Technology)2018	In this paper, the author has discussed image super-resolution techniques. He has given a detailed study about neural networks and GANs. He has accounted for different types of GANs and the algorithm for generation of each of them. He has also compared the loss functions of different GANs and has expressed the pros and cons of GANs.

FUTURE ENHANCEMENT

GANs were originally proposed to produce plausible synthetic images and have achieved exciting performance in the computer vision area. GANs have been applied to some other fields, (e.g., time series generation and natural language processing) with some success. Compared to computer vision, GANs research in other areas is still somewhat limited. The limitation is caused by the different properties inherent in image versus non-image data. For instance, GANs work to produce continuous value data but natural languages are based on discrete values like words, characters, and bytes, so it is hard to apply GANs for natural language applications. As this is also a very promising area, success in this area will lead to lots of applications such as generating subtitles and generating comments to live streaming, research areas such as neuroscience may have some privacy issues on the data and successful data augmentation will have significant impact in these areas. However, generation for other modal data such as time-series data using GANs is limitedly explored even though there is a lack of efficient evaluation metrics for evaluating the performance of GANs in those areas. More research is encouraged to be carried out in those areas. Since the first GAN was proposed in 2014, the development of GANs has brought lots of benefits to us either in research or in our real life. However, improper use of GANs can also bring hidden concerns to society e.g., GANs can be used to generate a tampered video for specific people and inappropriate events, creating images that are detrimental to a particular person, and may even affect that personal safety. We should also focus on developing forgery detector to detect the AI-generated images (including using GANs) efficiently and effectively

CONCLUSION

In this work, we presented a first Generative adversarial network SRGAN for performing super-resolution exclusively on high content screening microscopy images. We highlight the inefficiency of using MSE alone for generating super-resolved images that are not amiable to the human visual system. Our work was also motivated by the fact that the existing state of the art in super-resolution, SRGAN, trained on natural images, might potentially suffer from efficiently transferring feature representations in a distant domain (CCTV Images). SRGAN also pays disproportionate attention in optimizing for the perceptual quality of the images by using content loss and entirely discarding MSE which leads to poor psnr values.

We conclude that Generative adversarial networks optimized for carefully designed loss functions can lead to aesthetically pleasing SR images (notwithstanding the convergence issues) and implementing such creative loss functions geared towards the application domain holds an important key towards generating more realistic images.

REFERENCES

- [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.
- [2] M.-Y. Liu and O. Tuzel, “Coupled generative adversarial networks,” in *Advances in neural information processing systems*, 2016, pp. 469–477.
- [3] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training GANs,” in *Advances in Neural Information Processing Systems*, 2016, pp. 2234–2242.
- [4] M. O. Turkoglu, L. Spreeuwers, W. Thong, and B. Kicanaoglu, “A layer-based sequential framework for scene generation with GANs,” in *Thirty-Third AAAI Conference on Artificial Intelligence*, Honolulu, Hawaii, United States, 2019.
- [5] H. Wu, S. Zheng, J. Zhang, and K. Huang, “GP-GAN: Towards realistic high-resolution image blending,” *arXiv preprint arXiv:1703.07195*, 2017.
- [6] J. Pan, C. C. Ferrer, K. McGuinness, N. E. O’Connor, J. Torres, E. Sayrol, and X. Giro-i Nieto, “SalGAN: Visual saliency prediction with generative adversarial networks,” *arXiv preprint arXiv:1701.01081*, 2017.
- [7] G. K. Dziugaite, D. M. Roy, and Z. Ghahramani, “Training generative neural networks via maximum mean discrepancy optimization,” *arXiv preprint arXiv:1505.03906*, 2015.