

A Major Project-II Thesis Report

On

IMPROVING IMAGE RESOLUTION USING GENERATIVE ADVERSARIAL NETWORKS

Submitted in partial fulfillment of the requirement for the degree of

Master of Technology

in

Computer Science and Engineering

Submitted By

**Sumit Dhawan
2K18/CSE/17**

Under the guidance of

**Dr. Shailender Kumar
(Associate Professor)**



DELHI TECHNOLOGICAL UNIVERSITY

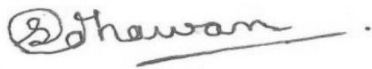
(Formerly Delhi College of Engineering) Shahabad Daultapur,

Main Bawana Road, Delhi-110042

October 2020

DECLARATION

I hereby declare that the Major Project-II work entitled “**Improving Image Resolution Using Generative Adversarial Networks**” which is being submitted to Delhi Technological University, in partial fulfillment of requirements for the award of the degree of Master of Technology (Computer Science and Engineering) is a bonafide report of Major Project-II carried out by me. I have not submitted the matter embodied in this dissertation for the award of any other Degree or Diploma.



Sumit Dhawan
2K18/CSE/17
M. Tech (Computer Science & Engineering)
Delhi Technological University

CERTIFICATE

This is to certify that Project Report entitled “**Improving Image Resolution Using Generative Adversarial Networks**” submitted by **Sumit Dhawan** (Roll No. 2K18/CSE/17) in partial fulfillment of the requirement for the award of degree Master of Technology (Computer Science and Engineering) is a record of the original work carried out by him under my supervision.



Project Guide 19/10/2020

Dr. Shailender Kumar
(Associate Professor)

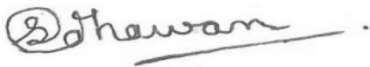
Department of Computer Science & Engineering
Delhi Technological University

ACKNOWLEDGEMENT

First of all, I would like to express my deep sense of respect and gratitude to my project supervisor **Dr. Shailender Kumar** for providing the opportunity of carrying out this project and being the guiding force behind this work. I am deeply indebted to him for the support, advice and encouragement he provided without which the project could not have been a success.

Secondly, I am grateful to **Dr. Rajni Jindal**, HOD, Computer Science & Engineering Department, DTU, for her immense support. I would also like to acknowledge Delhi Technological University library and staff for providing the right academic resources and environment for this work to be carried out.

Last but not the least I would like to express sincere gratitude to my parents and friends for constantly encouraging me during the completion of work.



Sumit Dhawan
2K18/CSE/17
M. Tech (Computer Science & Engineering)
Delhi Technological University

ABSTRACT

Even with all the achievements in precision and speed of various image super resolution models, such as better and more accurate Convolutional Neural Networks (CNN), the results have not been satisfactory. The high resolution images produced are generally missing the finer and frequent texture details. Majority of the models in this area focus on such objective functions which minimize the MSE (Mean Square Error). Although, this produces images with better PSNR (Peak Signal to Noise Ratio) but such images are perceptually unsatisfying and lack the fidelity and high frequency details when seen at a high resolution. Generative Adversarial Networks (GAN), a deep learning model, can be used for such problems. In this work, we present and show how GAN can be used to produce perceptually satisfying images with decent PSNR score as well as good Perceptual Index (PI) when compared to other models. In contrast to existing Super Resolution GAN model, we have introduced various modifications to improve the quality of images, like replacing batch normalization layer with weight normalization layer, modified dense residual block, taking features for comparison before they are fed in activation layer, using the concept of a relativistic discriminator instead of a normal discriminator that is used in vanilla GAN and finally, using Mean Absolute Error in the model.

Keywords: Image Super Resolution, Deep Learning, Generative Adversarial Networks, Convolutional Neural Networks, Super Resolution Models, etc.

TABLE OF CONTENTS

Title Page	i
Declaration	ii
Certificate	iii
Acknowledgement	iv
Abstract	v
Table of contents	vi-vii
List of Figures	viii
List of Tables	ix
List of Abbreviations	x
CHAPTER 1: INTRODUCTION	1
1.1 Background	1
1.1.1 Machine learning	1
1.1.2 Artificial Neural Network	2
1.1.3 Need of Image Super Resolution	4
1.2 Role of Convolutional Neural Network in Image Processing	5
1.3 Convolutional Neural Network	5
1.3.1 Convolution Layer	6
1.3.2 Activation Layer	7
1.3.3 Pooling Layer	7
1.3.4 Dropout Layer	8
1.3.5 Fully Connected Layer	8
1.4 Generative Adversarial Networks	9
1.4.1 Structure of GAN	10

1.4.2 Loss Function	10
1.4.3 Limitations of GAN	11
1.5 Objective of work	12
1.6 Organization of Dissertation	12
CHAPTER 2: LITERATURE REVIEW	13
2.1 The problem of image super resolution	13
2.2 Image Quality Evaluation	14
2.2.1 Peak Signal to Noise Ratio (PSNR)	14
2.2.2 Structural Similarity Index (SSIM)	15
2.2.3 Mean Opinion Score (MOS)	16
2.2.4 Perceptual Index	16
2.3 Image Super Resolution Models	17
CHAPTER 3: IMPLEMENTATION	21
3.1 Problem Statement	21
3.2 Proposed Method	21
3.3 Dataset	24
3.4 Algorithm	24
3.5 Tools and Technologies used	24
CHAPTER 4: RESULTS AND ANALYSIS	25
4.1 Results	25
4.2 Comparison	28
4.3 Analysis	28
CHAPTER 5: CONCLUSION & FUTURE SCOPE	29
5.1 Conclusion	29
5.2 Future Scope	29
REFERENCES	30

LIST OF FIGURES

Fig. 1.1	Architecture of ANN	3
Fig. 1.2.	ReLU function	7
Fig. 1.3.	Architecture of CNN	8
Fig. 1.4	Structure of Generative Adversarial Networks	10
Fig. 3.1	Basic structure of SRResNet is used	21
Fig. 3.2	Structure of a Basic Block	22
Fig. 4.1	Low resolution and high resolution super resolved image - 1	25
Fig. 4.2	Low resolution and high resolution super resolved image - 2	26
Fig. 4.3	Low resolution and high resolution super resolved image - 3	27

LIST OF TABLES

Table 4.1 PSNR Values	28
Table 4.2 Perceptual Scores	28

LIST OF ABBREVIATIONS

1. CNN - Convolutional Neural Network
2. ANN - Artificial Neural Network
3. SR - Super Resolution
4. GPU- Graphics Processing Unit
5. ReLu - Rectified Linear Unit
6. GAN - Generative Adversarial Networks
7. PSNR - Peak Signal to Noise Ratio
8. PI - Perceptual Index
9. NIQE - Natural Image Quality Evaluator
10. SSIM - Structural Similarity Index
11. MOS - Mean Opinion Score
12. IDE- Integrated Development Environment

CHAPTER 1

INTRODUCTION

1.1 Background

In today's world, sharing of information is very important and images are one of the popular mechanisms of doing so. Images are found to be useful in many areas related to human life, such as medical, agriculture, industry and social media. A pixel is the building block of any image and number of these pixels in a square unit area of an image gives the resolution of image, or simply put, it decides the clarity and detail holding capacity of an image. More is the resolution, more is the detail an image can hold and also, it refers to how close two lines can be to each other and still be visibly distinguishable. But due to various reasons such as low cost image capturing device, inability to capture all the details of image or technical fault in the image capturing device, resolution of captured image is not up to the required standard and thus, desirable details, in the form of pixels, are lost. Image super resolution is a technique that comes under the field of image processing, which helps in reproducing the lost details by adding extra pixels in the same one square unit area of the image, thereby increasing the resolution and clarity of image. Deep Learning, which comes under the area of Artificial Neural Networks (a popular and very much in use machine learning technique), has been proven to be very much successful in doing the process of image super resolution.

1.1.1 Machine learning

As our dependency on machines is increasing day by day, machine learning has become a very important and useful aspect of this age. Machine learning is simply a process through which machine is made to learn to perform a specific kind of task. Based on the dataset present, it can be supervised or unsupervised. In the task of image processing, machine learning plays an important role. A machine can be made to learn the mapping between low resolution image and corresponding high resolution image, and when put to use, resolution of any new image can be expected to increase. The obtained high resolution image varies in perceptual quality and clarity, depending on the image super resolution technique used.

Artificial Neural Network is one popular and very much in use machine learning technique which has proven to give outstanding results and has outperformed other machine learning models for various problem statements..

1.1.2 ANN (Artificial Neural Network)

Artificial Neural Network is modeled after the neural setup in our brain. It mimics the human biological neural connection and tries to copy the learning procedure as is followed by human brain. The complex network of billions of neurons in a biological brain comprises of

- a) Axon - Neurons are connected to each other through the axon.
- b) Dendrites - Receives input from the previous layer of neurons.
- c) Synapses – Sends output to the next layer of neurons

Nodes in ANN mimic the neurons and links between the nodes mimic the dendrites and synapses. The links have weights which is a learnable parameter as it is in biological brain. After receiving an input, a node applies an activation function to it and if a threshold value is crossed then the node is said to be activated and it fires the signal to the next layer of nodes connected to it. If threshold value is not crosses then no action is performed and the node remains inactivated. After learning process is completed, if a node is found to be inactive, that is, if the output of activation function for that node never reaches the threshold value, the node or that link is said to be dead. It is based on the fact that when a human brain learns any specific task, say, to distinguish dogs apart from cats, some specific connections are made in the brain and during the process of distinguishing, only those specific neurons will be fired and others will take no part in the process. It is exactly this newly formed connection between those specific neurons or pattern of connection that helps in performing the task that has been learned, again and again.

1.1.2.1 Advantages of ANN

- ANN can be made to learn any relationship between input and output, on its own and can even learn complex relationships which are nonlinear in nature. Such relationships are found in real life scenarios and thus, are very useful in learning the mapping between inputs to outputs. This is possible because of the structure and depth or multiple layers of neural network.

- ANN can be trained to be as smart as humans by making it learn the mapping from a huge dataset and thus letting it form unseen relationships. It is expected to perform well when in actual usage, it encounters data that is never seen before
- ANN never restricts input data to be of a particular type or format. Also, it gives better results for heterogeneous data when compared to other models.

1.1.2.2 Disadvantages of ANN

- Huge amount of data is required – As complexity of the model increases with increasing number of weighted links, it requires more and more data to completely saturates the weights and biases according to input. Also it is helps in increasing the accuracy of model.
- Longer training time is required – Machine Learning is quite a complex task in itself. And when ANN comes into picture, it increases the complexity exponentially. Trying to mimic the human brain, it requires lot of computing power and takes time for the model to be trained as many iterations are needed for the model to get stable. GPUs helps in reducing the time requirement when compared to CPUs
- Fine tuning of parameters of architecture – As there are so many parameters associated to a neural network, like weights, biases, learning parameter, number of iterations or epochs, batch size, etc., fine tuning of these parameters is very much needed so as to get a stable and efficient network that produces accurate results.

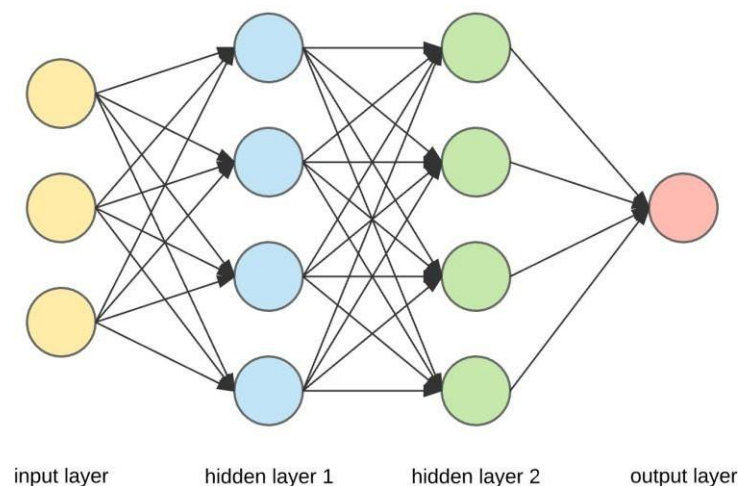


Figure 1.1 Architecture of ANN

1.1.3 Need of Image Super Resolution

Increasing resolution of an image has got many applications in real world. Few are listed below:

- a) **Satellite Image Processing:** Due to the distance and technology constraint images captured by satellites are not crisp and clear. Thus, it requires restoration of the lost information and enhancement by increasing number of pixels. It increases the visual interpretation of the image and enhance the geographical information by removing distortions.
- b) **Medical Image Processing:** It is well known that in medical imaging, especially MRI or CT Scan, clear and high resolution images are required which helps in early diagnosis and accurate treatment. Ultrasound images have lot of noise and X-ray images have low contrast. Converting these images to high resolution images will be beneficial in terms of medical research and treatment.
- c) **Microscopic Image Processing:** Image super resolution plays a significant role in analyzing and studying the microscopic entities such as cells and cellular structure. These techniques can enhance the visually poor image by adding information upto nanometer of scale, to the image. Many images taken in the switching mode, when put together, produce clearer and high resolution image.
- d) **Multimedia and Video Enhancement:** The usage and importance of multimedia applications is well known in today's world. It gives an immersive experience to the user. High resolution images enhance the experience of the user by making the visuals clearer, high definition and less noisy and less blurry. Old computer games are being re-launched with better visuals and high resolution images using the Super Resolution techniques on the visuals of old games. As a video is a collection of images, thus by increasing the resolution of images, resolution of whole video can be improved.
- e) **Astronomical Image Processing:** As the heavenly objects are far away and are tightly grouped together in clusters, their images are usually blurry or not so much clear that objects can be visually distinguished from each other in the image. Super Resolution techniques can help in increasing the visual perceptiveness of the astronomical images, thus helping in astronomical research and study purpose.
- f) **Other Applications:** Besides all these applications, there are many other applications too for Image Super Resolution, like, in forensics, military, surveillance, scanning, real time processing, automotive industry, object detection, etc.

Among all the neural networks CNN (Convolutional Neural Network) is the most used and most trending neural network. As, nowadays everything works upon artificial vision and automation of tasks. People want to make self-drive cars, robots that can travel by themselves. All this is possible only with the help of CNN. CNN proves to be the best method to process images, to identify objects in it, to detect scenes, human faces and other details in images.

1.2 Role of Convolutional Neural Network in Image Processing

CNN (Convolutional Neural Network) is one of the most used and most efficient neural networks when it comes to image processing. It can be used for object identification in images or detecting faces or other details in any image. CNN is a deep learning model. Deep Learning helps in making the machine learn by example, just like humans do. For example, it is being used in driverless cars, making them to distinguish between a stop sign and a lamppost and taking necessary action in all the types of conditions that occur while driving. It is able to achieve very good results that were impossible before as deep learning got a boost with the improved computing power. The term deep in deep learning corresponds to number of hidden layers in neural network. In traditional ANN, there were only 2-3 hidden layers whereas in deep learning models, hidden layers can reach up to even 150 in number.

Traditionally for image processing, filters for the image were designed corresponding to the problem statement. This was done manually and required a lot of manpower, time, expertise and energy. CNN, on the other hand, improves on this as it doesn't require any manual intervention. CNN itself extracts the required features from the image and in every iteration it learns how what features should be focused upon and what not and if it is present in the image or not. So many hidden layers in CNN help in this process as with every consecutive layer, abstraction of features extracted increases. For example, initial layers might extract lines and curves from an image while final layers might extract whole shapes and figures which are built using those lines and curves.

1.3 CNN (Convolutional Neural Network)

CNN is used heavily in image processing. The input images of CNN have 3 dimensions i.e. height, width and number of channels. Height and width represent the image resolution and third dimension, represents the number of channels, RGB or pixel values for red, green and blue colors respectively.

Following are the layers of a CNN model:

- Convolution layer
- Activation layer (e.g. ReLu or LeakyReLu)
- Pooling Layer
- Batch Normalization Layer
- Dropout Layer
- Fully Connected Layer

1.3.1 Convolution Layer:

This is the first layer to which input image is given. A kernel or filter (of square size) which is used to create feature map, glides over the image in fixed gap intervals called strides. To achieve desired results, tuning the size of stride is crucial. While running this kernel over the image, dot product is calculated between a pixel value of filter and corresponding pixel value from that part of image on which filter is present. Then resultant sum of all such values of product matrix is copied to the corresponding position in convolved feature map matrix. Therefore, with this process, a feature map with reduced dimensions is obtained. Filter can be of any shape or value, as it can extract different features from the image. For example, one filter can extract lines and curves from the image, another filter can extract some specific color intensity parts of the image. With every layer, complexity of filters increases as complex filter can extract the feature which is a combination of features extracted by previous filters.

1.3.1.1 Parameters which helps in adjusting CNN's performance:

- Stride – This is the number of pixels which we can pass while moving the filter over the image. Stride's value corresponds to the amount of loss we can suffer while reducing the image to a filter map.
- Padding - It means padding of zeroes all around the edges of input image. This makes the feature map output to be of required size. Generally, it is done in order to conserve the image's dimensions after convolution.
- Filters – Also known as kernels. These are of many different kinds. Each filter helps in extracting some feature from the image. It is a learnable parameter.

1.3.2 Activation Layer

This layer helps in deciding what values to pass to the next layer and what to reject. It simply consists of an activation function like ReLU (Rectified Linear Unit) or Leaky ReLU which passes activates the output connection only when threshold value is reached.

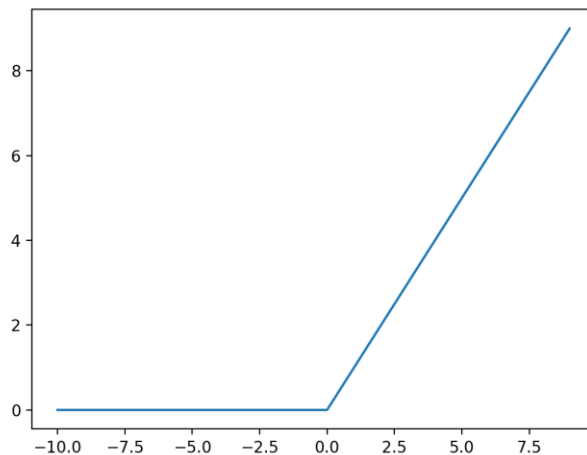


Figure 1.2 ReLU function

1.3.3 Pooling Layer

It is used to choose the pixel from a group of pixels which can contribute maximum to the result and discard other pixels. It further reduces the size of feature map by reducing the unnecessary sparse cells in the resultant matrix, which will be of no use. Though it reduces the size of the matrix, making the process faster and easier, it results in some loss of information. The idea behind these is that nearby or adjacent pixels can be approximated by the maximum information carrying pixel. It is generally of three types:

- Max Pooling: Selects the pixel having maximum value present in that group of pixels.
- Average Pooling: Takes the average of all the pixel values present in that group of pixels.
- Sum Pooling: Takes the sum of all pixel values present in that group of pixels

1.3.4 Dropout Layer

Dropout Layer plays a significant role in preventing overfitting of a neural network. When similar inputs are given to a model in training, it might lead to a fixed pattern of learning and inter-node connections which will fail if the input is changed to some other type. Dropout layer randomly selects any node and makes weight of its connection as 0 so as to dropout that node or make it dead. This will make the model learn a new path and use other nodes and connections to learn the mapping. It helps in increasing the efficiency of network. It makes the network learn in true sense such that it can perform even when the input is changed slightly and not just learn the direct mapping between given input and output.

1.3.5 Fully Connected Layer

The pattern of number of parameters in a CNN model follows an inverted pyramid structure as the parameters keeps on converging until they correspond to the desired output classes. The filter matrix is flattened to a single vector and it is fed to the fully connected layer which further gives its output to the activation functions such as sigmoid or softmax and thus helps in classifying the image.

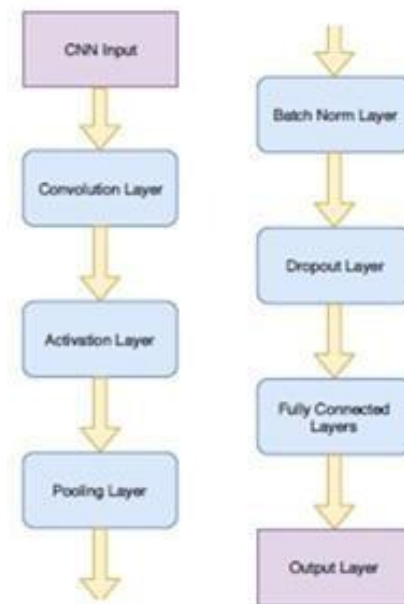


Figure 1.3 Architecture of CNN

1.4 Generative Adversarial Networks

With the increase in availability of processing power, machine learning has taken a giant leap. Deep Learning, a subset of machine learning, which requires lot of computing power, was once a distant possibility but is being used very much in today's world. Deep Learning outperforms various other models in learning any task because it itself extracts the features ranging from simple to high level, from the input given to it and it learns to extract more relevant features in every iteration, thus reducing the manual intervention of humans. It is also known as representation learning and this is the approach that human mind follows to learn anything. A subset of machine learning models, known as generative models lays the foundation of Generative Adversarial Networks or GANs in short. GANs are deep learning models as they uses CNN model. Variational Auto Encoder [47] and Boltzmann Machine [48] are few examples of initial generative models which were conceptualized on maximum likelihood estimation and markov chains.

A generative models work by estimating the distribution of output data points based on the distribution of input data points. But their performance is not as good as GANs, the concept of which was proposed by Goodfellow et al. [1] in the year 2014. GAN comprises of two adversaries or enemies of one another, one is called generator (G), another is called discriminator (D). Their only aim is to outperform each other by achieving their defined goal and thus they improve themselves in every iteration. It learns the joint probability distribution of input data and corresponding output data.

Image generation is a problem which has been a topic of research as it helps to generate an image based on some implicit and explicit features of an already existing image. As GANs uses CNN, which is an already state of the art model when it comes to image processing, they also work really well with mages and being a generative model, GANs can synthesize images based on the input images given. It is made up of two sub models, generator and discriminator. Generator is assigned the duty to generate new images based on the input images provided to it and fool the discriminator whereas discriminator is assigned the duty to use its discriminative ability and catch the bluff of generator by classifying real images as real and fake images as fake. It has been found out that GAN outperforms other models in image generation tasks by generating pretty good and life like images.

1.4.1 Structure of GAN

The basic concept of GAN states that it is a model of two adversaries which are playing a non-cooperative game where each player is well aware of the strategies of the other player, therefore even if any player modifies its strategy, it will gain nothing. This is the concept behind Nash equilibrium of Game Theory. GAN strives to reach this equilibrium only. The functions which are used for generator and discriminator should be differentiable. If G represents the generator function and D represents the discriminator function, then according to Figure 1.4, input to D is x or real data and $G(z)$ or fake/generated data, output of D is the classification of data to either real or fake, input to generator is just simple noise data or z .

If data is real, discriminator should classify it as real and label it as 1, otherwise it should classify it as fake and label it as 0. Discriminator will try to catch the lie of generator whereas generator will try to produce images so real that discriminator can't classify it as fake. This game between these two opponents is the core idea behind GANs and improves both the sub-models as generator is trying to produce better and more real looking images whereas discriminator is trying to further improve the generator by catching its bluff.

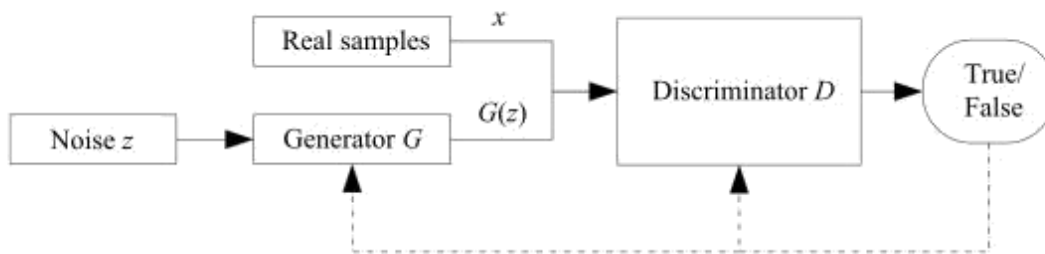


Figure 1.4 Structure of Generative Adversarial Networks

1.4.2 Loss Function

$$\min_G \max_D V(D, G) = E_{x \sim p(x)}[\log D(x)] - E_{z \sim p(z)}[\log(1 - D(G(z)))] \quad (1)$$

where,

x - Obtained by sampling the distribution of real data or $p(x)$.

z - Obtained by sampling the distribution of prior data or $p(z)$ (which can be Gaussian or Uniform distribution).

$E[x]$ - Expected value of any random variable x .

$D(x)$ - The probability that x is sampled from actual data and not from the generated data.

Discriminator has to maximize the loss function or objective function by minimizing $D(G(z))$ because it has to classify fake data as fake, and maximizing $D(x)$ because it has to classify real data as real. Generator has to minimize the loss function or objective function by maximizing $D(G(z))$ because it has to fool the discriminator by making it classify fake data as real. In initial iterations as generator is very weak in generating good images, discriminator will win in every iteration making $\log(1 - D(G(z)))$ tends to 0 as $1 - D(G(z))$ tends to 1. It makes the derivative tends to 0 as saturation happens. Only to prevent this unwanted scenario, objective function is modified to have $D(G(z))$ instead of $(1-D(G(z)))$.

Many variants have been proposed for GANs, each pertaining to a specific problem, like object detection, converting image to text or text to image, style transfer from one image to other, image enhancement, etc.

1.4.3 Limitations of GAN

Following are few limitations of GANs:

- Vanishing gradient problem: A shortcoming of using many layers, the gradient in back-propagation doesn't reach the initial layers and thus the change in initial weights is minimal, making it hard for the network to learn and converge to local minima.
- Mode collapse: Sometimes it so happens that GANs produces images related to a single or few kind of images only as it starts focusing on one mode only.
- Non convergence: Model doesn't converge as models parameters keep on oscillating and don't get stable.
- Lack of universal performance evaluation metric: Some evaluation metrics have been proposed for GANs like Average Log Likelihood [49], Inception Score [50], Wasserstein metric [51], Frechet Inception Distance [52]. But there is no consensus on which evaluation metric is best for GANs performance evaluation.
- Sensitivity of GAN models towards hyper-parameters also causes problems.

1.5 Objective of work

Here, in this work we are going to use state of the art technique – Generative Adversarial Networks for converting a low resolution image to a high resolution image. First, the model is trained on a training dataset of images and then it is evaluated by calculating Peak Signal to Noise Ratio and Perceptual Index of the generated super resolved image. Many changes have been proposed in existing super resolution models and incorporated in the proposed model. Further, comparison is drawn among existing super resolution models and proposed model, based on images produced and scores assigned. As GANs harness the power of CNN and combined with proposed changes in existing models, proposed model is expected to deliver accurate results fast when compared to existing models.

1.6 Organization of Dissertation

Chapter 1 deals with the introduction part and describe the background of this work, which comprises of the models used and concept behind the methodology being used. The usage and relevance of this work is explained at length. Working, uses and limitations of models are described. Then in Chapter 2 a detailed literature review is presented which contains all the literature and work done in this field. The problem is described along with existing methodologies to solve it.

In Chapter 3, implementation related details are shared along with proper explanation and reasoning behind proposed changes. Complete algorithm is stated stepwise as well as evaluation of generated images is also performed. In Chapter 4, a detailed analysis and comparison of few of the popular existing image super resolution models is done with proposed model. All results are properly discussed and stated in tables. In Chapter 5, the conclusion and the future scope of this work has been discussed. Then the references to all the resources used in gathering the information for this project are given in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

2.1 The problem of image super resolution

The problem of image super resolution, which means extracting high resolution image from the low resolution image, has attracted a lot of attention over the last decade, and has found multiple real world applications, like, reading sign and numbers plates from the low resolution CCTV footage, medical image processing, satellite and aerial imaging, low resolution facial image or textual image analysis. All this work has been discussed in Yang et al. [8] and also Nasrollahi and Moeslund [11]. It is a very challenging problems in the way that for a single low resolution images, there are multiple high resolution images to which mapping can be done. It has to be decided as to which mapping is to be followed and which high resolution image will give the better results. There have been many super resolution techniques proposed such as, sparse representation, statistical, edge based, patch based, prediction based, etc.

Let low resolution image be I^{LR} and high resolution image counterpart be I^{HR} then,

$$I^{LR} = D(I^{HR}; P_D) \quad (2)$$

where,

D is the degradation mapping function

P_D is the parameters of degradation process

In this process of blind SR, degradation process is not known and I^{SR} or super resolved image which is an approximation of the I^{HR} has to be obtained from the I^{LR} as:

$$I^{SR} = F(I^{LR}; P_F) \quad (3)$$

where,

F is the super resolution mapping function

P_F is the parameters of F

Conversion is from I^{HR} to I^{LR} to I^{SR} . The degradation process is not in our hands and can be impacted by many factors lie noise or compression but researchers have tried modeling the same by downsampling I^{HR} to I^{LR} by a scaling factor, s . Initially, bicubic interpolation was used as downsampling operation.

In the starting of super resolution techniques, interpolation methods based on pixel values were used, like bilinear interpolation (nearest 2x2 neighborhood for calculating value of unknown pixel) or bicubic interpolation (nearest 4x4 neighborhood for calculating value of unknown pixel), which are very fast but not that much efficient as they are PSNR oriented methods and produce extra smooth image and thus can't recover the finer and frequent texture details from the image, making it look less real like and more artificial. With the advent of machine learning, various approaches were proposed that simply rely on the mapping between low resolution image and its counterpart high resolution image. Such approaches have proven to be much more powerful than the previous methods used for SR problem. Freeman et al. [19, 20] has explained few super resolution methods.

2.2 Image Quality Evaluation

To assess and evaluate the quality of any image, one has to focus on visual attributes and perceptual quality of image. There are broadly two kind of image quality assessment techniques available, one is subjective, another is objective.. Subjective evaluation refers to how humans perceive the image and how clear and real it looks. Objective evaluation is done using scores and computation and no subjective preference or biases are involved. Subjective evaluation is generally better as humans can assess the quality of images better than artificially intelligent machines but it requires lot of time and energy therefore objective evaluation is generally preferred if there is a time constraint or budget constraint. It should be noted here that these two methods may lead to different results because the quality parameters of an objective evaluation need not to be consistent with the actual visual perceptiveness of an image as perceived by a human. Objective evaluation techniques are full reference (use reference images), reduced reference (use extracted features for comparison), no reference (devoid of any reference images).

2.2.1 Peak Signal to Noise Ratio (PSNR)

It is one of the widely used and popular image quality evaluation metric. It is impacted by Mean Squared Error (MSE) between images I^{SR} and I^{HR} and maximum pixel value (p_{max}).

Then, PSNR can be calculated as:

$$\begin{aligned}
PSNR &= 10.\log_{10}((p_{max})^2/MSE) \\
&= 10.\log_{10}((p_{max}/\sqrt{MSE})^2) \\
&= 20.\log_{10}(p_{max}/\sqrt{MSE})
\end{aligned} \tag{4}$$

PSNR focuses on the pixel value differences between two images and not the perceptual difference which can be judged by humans; therefore PSNR gives poor performance while evaluating reconstructed image's quality. But as is it has been used so much that it is still being used to compared with the works given in literature and also because there is lack of universal accurate evaluation metric for image quality.

2.2.2 Structural Similarity Index (SSIM)

The human brain is adapted to extract the structure images, which is the concept behind SSIM [7], which measures structural similarity between two images comparing structures, contrast and luminance.

Let say C_l denotes luminance comparison and C_c denotes contrast comparison between I^{SR} and I^{HR} , μ_{HR} is the mean of image pixels intensity and σ_{HR} is the standard deviation of image pixels intensity, then,

$$C_l(I^{HR}, I^{SR}) = (2. \mu_{SR}.\mu_{HR} + C_1)/(\mu_{HR}^2 + \mu_{SR}^2 + C_1) \tag{5}$$

where,

$$\mu_{HR} = 1/N (\sum_{i=1}^N I^{HR}(i))$$

$I^{HR}(i)$ is the intensity of i^{th} pixel

N is the number of pixels

$$C_1 = (k_1 p_{max})^2$$

and,

$$C_c(I^{HR}, I^{SR}) = (2.\sigma_{SR}.\sigma_{HR} + C_2)/(\sigma_{HR}^2 + \sigma_{SR}^2 + C_2) \tag{6}$$

where,

$$\sigma_{HR} = 1/(N-1) (\sum_{i=1}^N (I^{HR}(i) - \mu_{HR})^2)^{1/2}$$

$$C_2 = (k_2 p_{max})^2$$

C_1 and C_2 are constants to prevent instability.

Let C_s be the structure comparison function, then,

$$\begin{aligned} \sigma_{HR SR} &= 1/(N-1) (\sum_{i=1}^N (I^{HR}(i) - \mu_{HR})(I^{SR}(i) - \mu_{SR})), \\ C_s(I^{HR}, I^{SR}) &= (\sigma_{HR SR} + C_3)/(\sigma_{HR} \sigma_{SR} + C_3) \end{aligned} \quad (7)$$

where,

$\sigma_{HR SR}$ is the covariance between I^{HR} and I^{SR}
 C_3 is a constant for stability

and,

$$SSIM(I^{HR}, I^{SR}) = [C_l(I^{HR}, I^{SR})]^{\alpha} [C_c(I^{HR}, I^{SR})]^{\beta} [C_s(I^{HR}, I^{SR})]^{\gamma} \quad (8)$$

where,

α , β and γ are control parameters for assigning weights according to the importance.

2.2.3 Mean Opinion Score (MOS)

Mean Opinion Score comes under the subjective evaluation techniques where humans are asked to rate the images by assigning scores depending on the perceptual quality of image then arithmetic mean is taken of all the scores. As humans are involved in this technique it gives better results in judging the perceptual clarity of the image but it has some defects too like variance and biases as subjective evaluation is done. Models which perform poorly according to PSNR may perform very good according to MOS.

2.2.4 Perceptual Index

Perceptual index a very important metric when it comes to perceptual relevance of a restored image. Even if PSNR value is very high, it doesn't mean that it will be perceptually satisfying to the humans. It might be extra smooth thus making the image look more artificial and less real. Perceptual Index avoids such scenarios by trying to accurately predict the image quality based on how humans may perceive it.

Perceptual Index or perceptual score is defined as,

$$PI = 1/2 ((10 - Ma) + NIQE) \quad (9)$$

Ma refers to the score explained in Ma et al. [45] in which they have proposed a new evaluation no reference metric that learns from visual perceptual scores given by humans. Three kinds of statistical features, namely, global frequency features, local frequency feature and spatial features are extracted from frequency and spatial domains of an image and 2 stage regression model is used along with predefined perceptual scores given by humans, to predict the proposed metric without any ground truth image.

NIQE is Natural Image Quality Evaluator as introduced in [46]. This metric is also known as blind Image Quality Assessment and it uses only such deviations which are measurable, from the statistical irregularities found in real images. It is based on a group of statistical features that are quality aware and are derived from a collection of undistorted and natural images. It is a distance metric that refers to the distance between distorted image and model statistics.

2.3 Image Super Resolution Models

Image Resolution models have been categorized in following categories:

- Linear Networks (ex: SRCNN, VDSDR, IRCNN)
- Residual Networks (ex: EDSR, FormResNet)
- Recursive Networks (ex: DRCN, DRRN).
- Densely Connected Networks (ex: SR-DenseNet, RDN)
- GAN Models (ex: SRGAN)

Proposed model falls in the category of GAN models. All these models are explained below.

Dong et al. [4, 5] proposed a popular SR model, known as SRCNN, in which upsampling is done using bicubic interpolation and then the mapping of that upsampled low resolution image with high resolution image is learnt with the help of fully convolutional neural network that is three layers deep. It has been very successful when compared to previous SR models as it enabled the model to learn how mapping is done using filters of upscaled image. That made the model more accurate and increased its speed. VDSDR (Very Deep Super Resolution), as described in [42], is based on the Convolutional Neural Network with many layers or simply deep CNN. Its performance shows that deeper networks are beneficial in learning general

parameters used in image super resolution. Another model, IRCNN (Image restoration CNN) [43] performs tasks like deblurring and denoising using CNN based image denoisers which are composed of 7 stacked dilated convolution layers along with batch normalization layers and ReLu layers.

EDSR (Enhanced Deep Super Resolution), as described in [41], uses ResNet architecture and shows significant improvements by removing the batch normalization layers which are present in residual blocks and ReLu layers which are present outside the residual blocks. It performs better and gets more PSNR score when compared to VDSR. Another residual network is FormResNet, introduced in [13], which has two sub model networks, both of which are just like DnCNN, but loss layers are different. Euclidean and perceptual loss is included in the first network, known as Formatting Layer and it removes high frequency unwanted elements. Second network which learns the structured regions is known as DiffResNet.

Kim et. al [17] proposed a deeply-recursive convolutional network or DRCN which causes the impact of far off pixels on the unknown pixel being calculated. DRCN uses convolution layers and is made up of embedding net (used for input conversion to feature maps), inference net (performs super resolution) and reconstruction net (converts feature maps to super resolved image). Another recursive network is DRRN [44], which is even deeper than previous architectures and has 52 layers in it. To solve the problem of complexity induced by so many layers and to increase the speed, residual learning is performed by making local identity connections.

SRDenseNet [12] is a densely connected network which is based on DenseNet and in between the layers, it has dense connections. It means that a layer will get output directly from all the previous layers. It helps in flowing of information from low level feature layer to high level feature layer. It increases the speed by removing vanishing gradient problem. MSE loss trains the SRDenseNet and it shows significant improvement when compared to models which lacks densely connected layers. RDN or Residual Dense Network [38] uses dense connections from SRDenseNet and skip connections from SRResNet. It is based on the concept that hierarichal feature representation shall be utilized to learn local patterns. Along with MSE loss, it also uses MAE loss. It shows resistance against degradation of images and produces better SR images.

Tai et. Al [22] proposed an edge-directed SR method, that will try to recover the finer texture details without producing the unnecessary edge artifacts or the bands that appear around edges of a recovered or generated high resolution image. Yue et al. [21] fetches corresponding high resolution image from the internet and proposes a structure alignment matching criterion for the original image and retrieved high resolution image. In another technique [24, 25], embedding of the neighborhood patches is done. To super-resolve any low resolution image, its contained patches are looked for in other low resolution images and their corresponding patches in high resolution counterpart is fetched and thus upsampling is done. As discussed by Kim and Kwon [23], such approaches have the tendency of causing overfitting. Bruna et al. [10] and Johnsn et al. [9] proposed the models that relied on loss function, which focuses on perceptual similarity and quality of image rather than MSE of pixel values and thus produces better looking images that looks convincingly real.

The success of SR using CNN has shown that these models improve the accuracy and output of the model as highly complex mapping of images is done but at the cost of difficulty in training the model as these models are many layers deep. [6, 18] To stabilize the networks while training, concept of batch normalization [26] was introduced which works by subtracting the output of previous activation layer with batch mean and dividing by batch standard deviation. It makes the layer to learn the mapping somewhat independent of the effect of other layers and reduces covariance shift of each hidden layer. Another concept used in such CNN based SR models is that of residual blocks [28] and skip connections [17, 27], taken from the idea of ResNet or Residual Networks. Using skip connections in residual blocks, proposed network can be made to learn identity mapping and also reduce the problem of vanishing gradient by giving the gradient to initial layers more quickly, thus making the network train faster. Moreover it has been shown that learning the upscaling as well as mapping of upscaled low resolution image to high resolution image is better in terms of speed and accuracy [29, 30, 31], rather than giving the upsampled image to the network and let it learn only the mapping.

Loss functions used in SR models, like pixel wise loss function, where mean squared error of corresponding pixel values in super resolved image and original high resolution image, is calculated and minimized to produce an image that seems more like the actual high resolution image. It generally makes the image extra smooth and decreases the perceptual quality of the

image. PSNR (Peak Signal to Noise Ratio) has been a popular metric to gauge the quality of an image, even though later it was found out to be not much useful as it just focuses on MSE and not the actual perceptual quality of an image. To tackle this problem, various models have been proposed. Denton et al. [34] and Mathieu et al. [32] used Generative Adversarial Networks for super resolution of an image. Yo and Porikli [33] combined pixel wise loss along with GAN's discriminator loss.

Bruna et al. [10] and Johnson et al. [9] introduced a technique in which rather than focusing on pixel wise loss, focus is on minimizing the Euclidian distance between the extracted features from upsampled super resolved image and original high resolution image. This feature extraction is done by VGG16 model, given by K. Simonyan and A. Zisserman [6]. Similar to this approach, in SRGAN by Ledig et al. [2], the objective function is composed of two losses, content loss and adversarial loss. Content loss is based on the popular MSE loss used in many SR models, but in this, it is calculated on the extracted features from both images and not the pixels. Adversarial loss is based on the GAN model, which helps in making the image looks more real and natural. This combination of both the losses is known as perceptual loss and minimizing this perceptual loss makes the image much more perceptually appealing and increases finer texture details.

$$l^{SR} = l^{SR}_{VGG/i,j} + 10^{-3} l^{SR}_{Gen} \quad (10)$$

where l^{SR} is perceptual loss. $l^{SR}_{VGG/i,j}$ is content loss. l^{SR}_{Gen} is adversarial loss

VGG loss or content loss is calculated as:

$$l^{SR}_{VGG/i,j} = 1/W_{i,j} H_{i,j} (\sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\Phi_{i,j}(I^{HR})_{x,y} - \Phi_{i,j}(G_{\theta(G)}(I^{LR}))_{x,y})^2) \quad (11)$$

where,

$W_{i,j}$ and $H_{i,j}$ are dimensions of feature maps extracted by VGG network.

$G_{\theta(G)}(I^{LR})$ is the reconstructed image for the reference image I^{HR} .

$\Phi_{i,j}$ is the feature map produced by VGG before i^{th} maxpooling layer and after the j^{th} convolution layer.

Adversarial loss is calculated as:

$$l^{SR}_{Gen} = \sum_{n=1}^N (-\log D_{\theta(D)}(G_{\theta(G)}(I^{LR}))) \quad (12)$$

where $D_{\theta(D)}(G_{\theta(G)}(I^{LR}))$ is the probability that recovered or super resolved image is natural high resolution image.

CHAPTER 3

IMPLEMENTATION

3.1 Problem Statement

The main goal is to enhance the image quality of super resolution method using GAN, that too, mainly in terms of perceptual quality of image, not just making image clearer when viewed at a high resolution. This section explains about the proposed method, model architecture and then few modifications in the existing SRGAN model that improves the image quality in terms of perceptual index and PSNR.

3.2 Proposed Method

As shown in Figure 3.1, low resolution image is first given to the convolutional block that is used to apply filters on the input and thus produce feature maps. Then, its output is given to the residual basic block.. Output of residual basic blocks is again provided to another convolutional block and then it is upsampled with the factor ($r=4$) and then again is given to two consecutive convolutional blocks. Thus, the final output produced is a super resolved high resolution image.

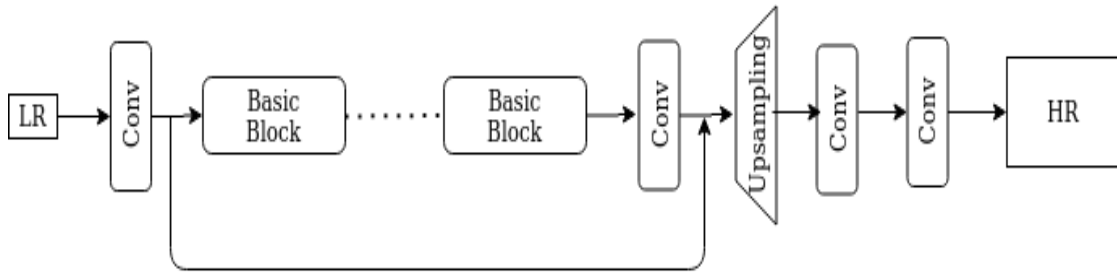


Figure 3.1 Basic structure of SRResNet is used

The goal is to generate super resolved high resolution image (I^{SR}) from given low resolution image (I^{LR}). For model training purpose this I^{LR} is obtained by downsampling the high resolution image (I^{HR}) by a factor ($r = 4$, following the same value used in SRGAN [2], as higher factors were shown to produce undesirable image artifacts). To get I^{SR} output from the input of I^{LR} , generator network is used. Using GAN, where generator is fooling the discriminator, it learns to produce images which are very similar to real images and thus

emphasis is always on such solutions which are perceptually better than those solutions which are just based on minimizing pixel loss using MSE.

In the structure of SRGAN, batch normalization layers are replaced by weight normalization layers [35] because the latter incurs a smaller cost of calculation on Convolutional Neural Networks (or generator model), thus it is easier and faster to train, as well as, not many unpleasant artifacts are produced like in the case of batch normalization layers. Also this modification is robust to the scale of weight vector and focuses on weight initialization. Moreover, batch normalization is useful and stable if the batch size is large whereas, in this work, stochastic gradient descent (SGD) is used instead of batch gradient descent (BGD), because SGD is computationally faster than BGD, thus allowing for more number of iterations and making the super resolution model efficient. Also, SGD doesn't average out the noise in input like it is done in BGD, therefore letting the model learn accordingly because in the problem of single image super resolution, a single image, generally having some noise, is given to model.

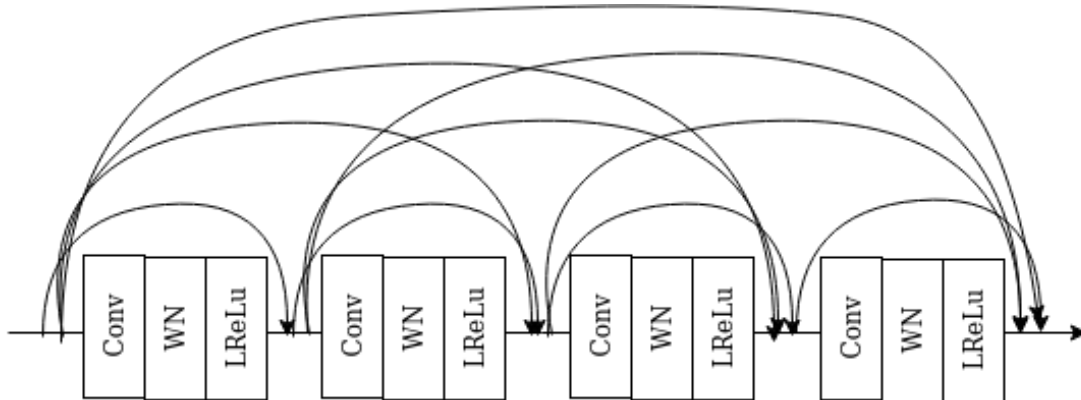


Figure 3.2 Structure of a Basic Block (Residual Dense Block with WN or Weight Normalization Layer)

In the content loss function, instead of using L2 loss or Mean Squared Error (MSE), L1 loss or Mean Absolute Error (MAE) is used because MAE tries to reduce the average error by minimizing the sum total of differences between I^{HR} and I^{SR} whereas MSE will try to find such a solution out of all plausible solutions, that results in the lowest loss and is easiest to converge to and thus will produce extra smooth image, which is not perceptually appealing. Also, MSE doesn't work well with outliers in the data. The residual block used in generator

model of SRGAN is replaced by a deeper and more complex residual dense block which is introduced by Zhang et al. [38]. It helps in extracting a number of local features with the help of fully connected and dense convolutional layers, exploiting the hierarchical features from all the layers.

Here, as shown in Figure 3.2., one residual block is connected to all the residual blocks behind it and thus local features fetched from all those blocks combined with the current local feature creates a kind of contiguous memory system, which helps in effective learning of more features. It helps in learning hierarchical features globally in the whole network and also causes a more stable training of model. Generator loss is weighted average of content loss, adversarial loss and Mean Absolute Error loss between I^{SR} and I^{HR} . Along with the modification in generator, one change in discriminator is also introduced, that is, the concept of relativistic discriminator [39] is used instead of standard discriminator. As discussed in the [39], generator should be trained to reduce the probability that real data is real along with increasing the probability that fake data is real. And discriminator is trained to catch this lie of generator.

Relativistic discriminator loss is given as:

$$L_D^{Ra} = -E_{x_r} [\log (D_{Ra}(x_r, x_f))] - E_{x_f} [\log(1 - D_{Ra}(x_f, x_r))] \quad (13)$$

Generator loss is given as:

$$L_G^{Ra} = -E_{x_r} [\log (1 - D_{Ra}(x_r, x_f))] - E_{x_f} [\log(D_{Ra}(x_f, x_r))] \quad (14)$$

where,

$E_{x_f} []$ represents average of all the fake data in a batch.

D^{Ra} represents the relativistic discriminator.

x_f represents output of generator or super resolved image, ($x_f = G(x_i)$; x_i is the input low resolution image)

x_r represents real data or high resolution image.

Here, generator contains both x_r and x_f , so in training, it uses the gradients of generated as well as real data instead of only gradient of generated data in SRGAN. This will train the model to detect sharp edges and finer texture details.

3.3 Dataset

Training is done on three datasets: Street View Images by Google [3, 40] (a collection of 1000 images), Div2k [36] (a collection of 800 images), BSD100 [16] (a collection of 100 images). Super resolved images of all datasets are compared with corresponding high resolution images. PSNR value and Perceptual Score (or, Perceptual Index, PI) is calculated based on I^{SR} and I^{HR} . Testing is done on three datasets: BSD100 [16], Set5 [14] and Set 14 [15]. As it is a single image super resolution model, only one image was taken from every dataset for testing purpose.

3.4 Algorithm

To convert I^{HR} to I^{LR} , bicubic downsampling with a factor of $r = 4$, is done. Following process is repeated for every batch in every epoch:

1. Generator is trained and a super resolved image (I^{SR}) is generated from the I^{LR} .
2. Pixel wise loss based on MAE as well as MSE is calculated between I^{SR} and I^{HR}
3. PSNR value is found using above calculated MSE.
4. Adversarial loss is calculated on real and fake image predictions by discriminator using Binary Cross Entropy Loss followed by Sigmoid function.
5. Content loss is calculated based on the MAE between features of real and fake image. Features are extracted using the pre-trained VGG19 model.
6. Generator loss is calculated using the weighted average of content loss, adversarial loss and pixel wise loss (based on MAE).
7. Discriminator is trained using the relativistic discriminator loss function.

Proposed model is then compared with previous SR models based on the given evaluation metrics and same testing datasets. Scores of other models is referred from Zhu et al. [37].

3.5 Tools and Technologies used

Language used for developing the project: **Python**

Platform used for developing the project: **Google Colab**

Library used: **PyTorch**

CHAPTER 4

RESULTS AND ANALYSIS

The images produced using the proposed model, were compared with the original high resolution images based on various metrics like PSNR and Perceptual Index. Thus, scores obtained are followed by every set of images composed of low resolution (downsampled) and corresponding super resolved image.

4.1 Results

A. Google Street View Images Dataset



Fig 4.1 Low resolution and high resolution super resolved image

PSNR: 29.396287

PI: 3.8805

PSNR and PI obtained are highest for this image when compared to other images, therefore explaining that the produced image is smoother than other resultant images

B. Div2K Dataset



Fig 4.2 Low resolution and high resolution super resolved image

PSNR: 27.140278

PI: 3.7762

PSNR obtained is lowest and PI obtained is moderate for this image when compared to other images, therefore explaining that the produced image has finer texture details and has least blurriness.

C. BSD100 Dataset



Fig 4.3 Low resolution and high resolution super resolved image

PSNR: 28.086177

PI: 2.8758

PSNR obtained is moderate and PI obtained is least for this image when compared to other images, therefore explaining that the produced image is perceptually satisfying with some amount of blurriness.

4.2 Comparison

	VDSR	EDSR	SRGAN	Proposed Model
Set5	31.349	32.630	30.666	31.680
Set14	26.090	28.015	28.953	30.015
BSD100	25.957	27.287	27.796	28.086

Table 4.1 PSNR values

	VDSR	EDSR	SRGAN	Proposed Model
Set5	6.297	5.906	3.844	4.044
Set14	5.696	5.514	3.064	3.423
BSD100	5.700	5.559	2.802	2.875

Table 4.2 Perceptual Score

4.3 Analysis

It can be seen from the comparison tables above that proposed model has shown a significant improvement when compared to basic SR models, like VDSR and EDSR. Proposed model was trained on images in the magnitude of thousands whereas original SRGAN was trained on images in the magnitude of lakhs. Still, the performance is at a similar scale as that of SRGAN as it gives nearby similar Perceptual Scores and better PSNR values than SRGAN. The perceptual score of proposed model, which is a better metric than PSNR in terms of image perceptual quality, is shown to be way better than VDSR and EDSR. For all datasets, especially BSD100, perceptual score of proposed model is very much near to perceptual score of original SRGAN.

CHAPTER 5

CONCLUSION & FUTURE SCOPE

5.1 Conclusion

An image super resolution model is presented, that is based on GAN with various modifications to the existing SR models. The new structure with weight normalization layers instead of batch normalization layers, dense residual blocks, relativistic discriminator, stochastic gradient descent instead of batch gradient descent and mean absolute error instead of mean squared error, has given a boost in performance of the model when compared to other SR models. Downsampled or low resolution images are shown along with corresponding super resolved generated images and PSNR and PI is given for every dataset. Further, comparison is done between SR models on same datasets and evaluation metrics

The proposed model outperform previous SR models like VDSR and EDSR as better Perceptual Score (lower is better) and more PSNR (higher is better) are obtained when compared to these two models. Perceptual loss is way more efficient as a loss function rather than pixel wise MSE loss, to get an image with better perceptual quality, fidelity and high frequency details. Along with it, proposed modifications in the network increased the speed, stability, accuracy and performance of the network.

5.2 Future Scope

There is always some space for innovation, improvement, and modification of existing technique in any research area. So much good and thorough research work is present in this area, still there is still a lot of work to be done in the making of these Single Image Super Resolution models more accurate and reliable. Reconstruction of structured scenes or text which should be at least partially convincing is challenging and thus lies in the future scope of this work. The formulation of such content loss functions that focus more on spatial content and less on changes in pixel space will increase the perceptual relevance of results. Many more Image Quality Assessment metrics like SSIM can be used to measure the performance of this model with others. Accuracy of these systems shall be improved so that they can be used in real time and crucial tasks, high quality images, though required are unavailable due to any kind of limitation be it distance or technology or cost.

CHAPTER 6

REFERENCES

- [1] I. Goodfellow et al., “Generative adversarial nets,” in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 2672–2680.
- [2] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: CVPR. (2017)
- [3] Google Street View.<https://www.google.com/maps/streetview/>
- [4] Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: ECCV. (2014)
- [5] Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. TPAMI 38(2) (2016) 295–307.
- [6] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015
- [7] A. Horé and D. Ziou, "Image Quality Metrics: PSNR vs. SSIM," 2010 20th International Conference on Pattern Recognition, Istanbul, 2010
- [8] C.-Y. Yang, C. Ma, and M.-H. Yang. Single-image super-resolution: A benchmark. In European Conference on Computer Vision (ECCV), pages 372–386. Springer, 2014
- [9] Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: ECCV. (2016).
- [10] Bruna, J., Sprechmann, P., LeCun, Y.: Super-resolution with deep convolutional sufficient statistics. In: ICLR. (2015)
- [11] K. Nasrollahi and T. B. Moeslund. Super-resolution: A comprehensive survey. In Machine Vision and Applications, 2014.
- [12] T. Tong, G. Li, X. Liu, and Q. Gao, “Image super-resolution using dense skip connections,” in ICCV, 2017.
- [13] J. Jiao, W.-C. Tu, S. He, and R. W. Lau, “Formresnet: formatted residual learning for image restoration,” in CVPRW, 2017.
- [14] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. BMVC, 2012.
- [15] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In Curves and Surfaces, pages 711–730. Springer, 2012.
- [16] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In IEEE International Conference on Computer Vision (ICCV), 2001
- [17] J. Kim, J. K. Lee, and K. M. Lee. Deeply-recursive convolutional network for image super-resolution. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016
- [18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1–9, 2015.
- [19] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. IEEE Computer Graphics and Applications, 2002.
- [20] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. International Journal of Computer Vision, 40(1):25–47, 2000.
- [21] H. Yue, X. Sun, J. Yang, and F. Wu. Landmark image super-resolution by retrieving web images. IEEE Transactions on Image Processing, 22(12):4865–4878, 2013
- [22] Y.-W. Tai, S. Liu, M. S. Brown, and S. Lin. Super Resolution using Edge Prior and Single Image Detail Synthesis. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2400–2407, 2010.
- [23] K. I. Kim and Y. Kwon. Single-image super-resolution using sparse regression and natural image prior. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(6):1127–1133, 2010.
- [24] R. Timofte, V. De Smet, and L. Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In Asian Conference on Computer Vision (ACCV), pages 111–126. Springer, 2014
- [25] R. Timofte, V. De, and L. Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In IEEE International Conference on Computer Vision (ICCV), pages 1920–1927, 2013.

- [26] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of The 32nd International Conference on Machine Learning (ICML)*, pages 448–456, 2015.
- [27] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In *European Conference on Computer Vision (ECCV)*, pages 630–645. Springer, 2016
- [28] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016
- [29] Y. Wang, L. Wang, H. Wang, and P. Li. End-to-End Image Super-Resolution via Deep and Shallow Convolutional Networks. arXiv preprint arXiv:1607.07680, 2016.
- [30] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1874–1883, 2016
- [31] C. Dong, C. C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. In *European Conference on Computer Vision (ECCV)*, pages 391–407. Springer, 2016.
- [32] M. Mathieu, C. Couprie, and Y. LeCun. Deep multi-scale video prediction beyond mean square error. In *International Conference on Learning Representations (ICLR)*, 2016.
- [33] X. Yu and F. Porikli. Ultra-resolving face images by discriminative generative networks. In *European Conference on Computer Vision (ECCV)*, pages 318–333. 2016.
- [34] E. Denton, S. Chintala, A. Szlam, and R. Fergus. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1486–1494, 2015
- [35] Tim Salimans and Diederik P Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. arXiv preprint arXiv:1602.07868, 2016.
- [36] E. Agustsson and R. Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPR Work-shops*, July 2017
- [37] X. Zhu, Y. Cheng, J. Peng, M. Wang, R. Le, and X. Liu, “Super-Resolved Image Peceptual Quality Improvement via Multi-Feature Discrimiantors,” CoRR, abs/1904.10654, 2019.
- [38] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: CVPR (2018)
- [39] Jolicoeur-Martineau, A. The relativistic discriminator: a key element missing from standard GAN. In *ICLR*, 2019
- [40] A. R. Zamir, T. Wekel, P. Agrawal, C. Wei, J. Malik, and S. Savarese. Generic 3d representation via pose estimation and matching. In *Proc. of the European Conf. on Computer Vision (ECCV)*, 2016
- [41] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *CVPR Workshops*, 2017.
- [42] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646–165.
- [43] K. Zhang, W. Zuo, S. Gu, and L. Zhang, “Learning deep cnn denoiser prior for image restoration,” in *CVPR*, 2017
- [44] Y. Tai, J. Yang, and X. Liu, “Image super-resolution via deep recursive residual network,” in *CVPR*, 2017.
- [45] Ma, C., Yang, C.Y., Yang, X., Yang, M.H.: Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding* 158,1–16 (2017)
- [46] Mittal, A., Soundararajan, R., Bovik, A.C.: Making a “completely blind” image quality analyzer. *IEEE Signal Process. Lett.* 20(3), 209–212 (2013)
- [47] Kingma, Diederik P., and Max Welling. “Auto-Encoding Variational Bayes.” ArXiv:1312.6114 [Cs, Stat], May 2014. arXiv.org, <http://arxiv.org/abs/1312.6114>.
- [48] Fischer, Asja, and Christian Igel. “An Introduction to Restricted Boltzmann Machines.” *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, edited by Luis Alvarez et al., Springer, 2012, pp. 14–36. Springer Link.
- [49] Hamid Eghbal-zadeh, Gerhard Widmer, “Likelihood estimation for generative adversarial networks”, 2017.
- [50] Salimans, Tim, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. "Improved techniques for training gans." In *Advances in Neural Information Processing Systems*, pp. 2234-2242. 2016.
- [51] Arjovsky, Martin, Soumith Chintala, and Léon Bottou. "Wasserstein gan." arXiv preprint arXiv:1701.07875, 2017.
- [52] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, Gans trained by a two time-scale update rule converge to a local nash equilibrium, in: *Advances in Neural Information Processing Systems*, 2017, pp. 6629–6640.