# A Project On

# Image Super Resolution Based on Generative Adversarial Networks

A project paper is submitted to the **Faculty of Engineering and Technology** of **Islamic University** in partial fulfillment of the requirements for the Degree of B.Sc.(Engg.) in **Information and Communication Technology**


**Supervised By:**
**Dr. Md. Alamgir Hossain**
**Associate Professor**
**Dept. of Information and Communication Technology,**
**Islamic University**

**Submitted By:**
**Md Bikasuzzaman**
**Roll no: 1718012**
**Reg. no: 1323**
**Session: 2017-18**

**Department of Information and Communication Technology,**
**Islamic University Kushtia-7003, Bangladesh**


**May 2023**

**Dedicated to...**

**My Beloved Parents**

**&**

**Teachers**

# Certificate

It's my pleasure to certify that Md Bikasuzzaman, Examination Roll No. 1718012, and Registration No. 1323, has performed a project work, **"Image Super Resolution Based on Generative Adversarial Networks"** under my supervision in the academic session 2017-18 for the fulfillment of the partial requirement for B.Sc. (Engg.) in Information and Communication Technology, Islamic University, Kushtia-7003, Bangladesh.

As per I know that this dissertation has not been copied from any other project or submitted elsewhere prior to submission to this department.

...............................................

**Dr. Md. Alamgir Hossain**
Associate Professor,
Department of Information and Communication Technology
Islamic University, Kushtia-7003, Bangladesh.

# Acknowledgement

At first, thanks to Almighty Allah, the creator and sustainer who has given me strength and opportunity to complete the project **"Image Super Resolution Based on Generative Adversarial Networks"**. Regarding the outcome of my project, I would like to express my deepest sense of gratitude and respect to my supervisor **Dr. Md. Alamgir Hossain**, Associate Professor, Department of Information and Communication Technology, Islamic University, Kushtia-7003, Bangladesh for his valuable suggestion to select the topic; and constant guidance, great supervision, advice, encouragement and other fruitful helps throughout the duration of my project. His dedication, collaboration and interaction were key factors in the success of my project. Without his active support and great supervision, I would not be able to complete my project. I am also grateful to my respectable teachers, Department of Information and Communication Technology, for their encouragement and continued supports and help throughout the project.

Finally, my profound respect and thanks to my family for their constant sacrifice encouragement and support over the years.

**Md Bikasuzzaman**
Roll no: 1718012
Reg. no: 1323
Session: 2017-18
Dept. of Information and Communication Technology
Islamic University, Kushtia-7003, Bangladesh.

**Abstract**

Image super-resolution, the process of enhancing the resolution and quality of low-resolution images, has gained significant attention in computer vision. Generative Adversarial Networks (GANs) have emerged as a powerful framework for addressing this problem. In this study, we propose an image super-resolution method based on GANs. Our approach consists of a generator network that generates super-resolution images from low-resolution image. To our knowledge, the low resolution image is converted by the generator model to a super resolution image using four scaling factors. We suggest a perceptual loss function that combines an adversarial loss and a content loss to accomplish this. Our solution is pushed to the natural image manifold by the adversarial loss using a discriminator network that has been trained to distinguish between super-resolved images and the original high-resolution images.Our deep residual network can restore super resolution from images that have been severely downscaled.

**Keyword**: Generator, Discriminator, GAN, Super Resolution.

# Contents

# List of Figures

# Chapter 1
# Introduction

## 1 Introduction

Single image super-resolution (SISR) is a significant branch of image processing. It also aims to recover a high-resolution (HR) image from a low-resolution (LR) image; in the medical field, obtaining higher-quality images can help doctors accurately detect diseases. Thus, studying SISR is important for both academics and industry[15].

Researchers have created a number of techniques based on degradation models of low-level vision tasks to address the SISR problem.[19]. There are three categories for SISR in general, i.e., image itself information, prior knowledge and machine learning. In the image itself information, directly amplifying resolutions of all pixels in a LR image through an interpolation way to obtain a HR image was a simple and efficient method in SISR[11], i.e., nearest neighbor interpolation, bilinear interpolation and bicubic interpolation, etc. It is noted that in these interpolation methods, high-frequency information is lost in the up-sampling process. Alternatively, reconstruction- based methods were developed for SISR, according to optimization methods. That is, mapping a projection into a convex set to estimate the registration parameters can restore more details of SISR.

Despite being able to overcome the shortcomings of image itself information methods, the aforementioned techniques still encountered issues with non-unique solutions, slow convergence speeds, and higher computational costs. To avoid this occurrence, an ideal solution was found to enhance the quality of the predicted SR images by integrating prior knowledge and the image itself into a frame[15]. Using maximum a posteriori (MAP) can regularize a loss function to obtain a maximum probability for improving the efficiency. Besides, machine learning methods can be presented to deal with SISR, according to relation of data distribution. On the basis of ensuring the image SR effect, sparse-neighbor-embedding-based (SpNE) method via partition the training data set into a set of subsets to accelerate the speed of SR reconstruction. There are numerous additional SR techniques that have the benefit of producing flexible and precise detail while frequently using sophisticated prior knowledge to limit the range of potential solutions.

Super-resolution tasks were solved using a variety of deep learning techniques on a large-scale image dataset in order to produce a better and more effective SR model. For instance, Dong et al. proposed a pixel mapping technique based on a super-resolution convolutional neural network (SRCNN) that only required three layers to achieve greater learning capacity than some common machine learning techniques for super-resolution images. The shallow architecture and high complexity of the SRCNN made it problematic even though it had a good SR effect. To overcome challenges of shallow architectures, Kim et al.[8] designed a deep architecture by stacking some small convolutions to improve performance of image super-resolution. Tai et al. relied on recursive and residual operations in a deep network to enhance learning ability of a SR model. To further improve the SR effect, Lee et al. used weights to adjust residual blocks to achieve better SR performance. To extract robust information, the combination of traditional machine learning methods and deep networks can restore more detailed information for SISR. For instance, Wang et al. embedded sparse coding method into a deep neural network to make a trade off between performance and efficiency in SISR. To reduce the complexity, an up-sampling operation is used in a deep layer in a deep CNN to increase the resolution of low-frequency features and produce high-quality images. For example, Dong et al. directly exploited the given low-resolution images to train a SR model for improving training efficiency, where the SR network used a deconvolution layer to reconstructing HR images. There are also other effective SR methods. For example, Lai et al. used Laplacian pyramid technique into a deep network in shared parameters to accelerate the training speed for SISR. Zhang et al. guided a CNN by attention mechanisms to extract salient features for improving the performance and visual effects in image SISR[15].

The above-mentioned SR methods have achieved excellent results in SISR, but the obtained damaged images are insufficient for use in practical applications, which restricts their use with actual cameras. Generative adversarial networks (GANs) used generator and discriminator in a game-like manner to address the issue of small samples and achieve good performance on image applications[16]. Specifically, the generator can generate new samples, according existing samples. The discriminator is used to distinguish the samples from generator. Due to their strong learning abilities, GANs become popular image super-resolution methods[1].

Figure 1: Low resolution image to High resolution image

### 1.1 Motivation

Generative adversarial networks (GANs) have been shown to be effective for a variety of tasks, including image super-resolution. The motivation for using GANs for image super-resolution are...

- In low resolution image, we are unable to directly extract information and comprehend what the image is trying to convey.
- Therefore, we can improve resolution if we want to grasp anything clearly and extract the correct information.

### 1.2 Problem Statement

To recover or restore high resolution image from low resolution image. There are many forms of image enhancement which includes noise-reduction, up-scaling image and color adjustments. We are estimating high-resolution(HR) image from it's low-resolution image(LR) by applying deep network with adversarial network (Generative Adversarial Networks) to produce high resolutions images. Our main target is to reconstruct super resolution image by up-scaling low resolution image such that texture detail in the reconstructed SR images is not lost.

### 1.3 Objective

The objective of image superresolution based on GAN is to reconstruct a high-resolution (HR) image from a low-resolution (LR) image.By enhancing the image resolution, SRGAN can assist medical professionals, the forensic division, and security purposes by making it possible to see what the image actually looks like. This is a challenging problem because the LR image contains only a subset of the information in the HR image. As a result, there are many possible HR images that could be reconstructed from the LR image. I will use generative adversarial networks (GANs) for this project. GANs have been shown to be effective for image super-resolution.

# Chapter 2
# Literature Review

## 2 Literature Review

The authors of [7] discussed image super resolution using GANs and the benefits of various neural network architectures. On the given topic, some network designs are more suitable than others. To make the data more meaningful and to focus on the performance metrics that gauge the caliber of a generated image, they used a generative adversarial network. Some performance metrics, they said, can't be used to gauge quality directly. Even if the quality is low, the most popular performance metrics like PSNR and SSIM may still be higher. This type of issue is caused by loss functions used in neural networks.

Authors of [13] for image super resolution compared the performance of SRGAN and SR-CNN and demonstrate that SRGAN performed better than SR-CNN for their loss function.Later, ESR-GAN improved the architecture to achieve better performance by including more layers, employing a relativistic discriminator, and utilizing preactivation VGG-loss with a weighted average and MSE loss.

Christian Ledig-2017 [9] in his research achieved very impressive results in image super resolution using adversarial network architecture, but it can be challenging to train realistic visual content for deeper network architectures.

Although Kui Fu and his co-authors-2020 [5] in single image super resolution research obtained the most advanced visual performance based on GAN method, their model faced enormous computation for huge parameters in the model and was undoubtedly challenging to train.

Wenlong Zhang and other co-authors of [18] use RankSRGAN to optimize the generator in the direction of perceptual metrics, which enhances the visual quality of super-resolved results. They develop a Ranker that can learn how perceptual metrics behave before introducing a novel rank-content loss to enhance perceptual quality. The method they suggest can improve results by combining the advantages of various SR techniques. They developed RankSRGAN, a flexible framework that can recover more realistic textures and outperform cutting-edge perceptual metric techniques.

The authors of [12] proposed a method in which they used the Enhance Super-resolution Generative Adverserial Networks (ESRGAN) framework and the Residual Channel Attention Network (RCAN) as the generator instead of the Residual-in-Residual Dense Block (RRDB). They trained two RCANs using various discriminators and hyper-parameters in ESRGAN to enhance perceptual quality, and their final SR prediction was a pixel-wise ensemble of those two RCANs.

A deep generative adversarial network for super-resolution was proposed by Manri Cheon, Jun-Hyuk Kim, Jun-Ho Choi, and Jong-Seok Lee of [4] taking into account the trade-off between perception and distortion. To train the EUSR-based GAN model, the researchers used two perceptual content loss functions, namely the DCT loss and the differential content loss. Based on the successful implementation of a recently created super-resolution model, deep residual network using enhanced upscale modules (EUSR), the proposed model is trained to enhance perceptual performance with only a slight increase in distortion.

# Chapter 3
# GANs for Image Super Resolution

## 3 GANs for Image Super Resolution

Image super-resolution (SR) is a technique that aims to reconstruct a high-resolution (HR) image from one or more low-resolution (LR) images. SR has a wide range of applications, including:

- Visualization: SR can be used to enhance the quality of images for visualization applications, including in video surveillance and medical imaging.
- Enhancement: SR can be used in photography and image editing to improve the aesthetic quality of photos.
- Upscaling: SR can be used to increase the resolution of images so they can be viewed on high-definition displays, for example.

### 3.1 Overview of Image Super Resolution

Image super-resolution is a technology used to enhance the resolution and quality of low-resolution images, improving their visual details and sharpness. It is particularly useful in various applications, such as digital photography, medical imaging, surveillance systems, and video processing.

The goal of super-resolution is to generate a high-resolution image from one or more low-resolution input images. This process involves increasing the level of detail, enhancing fine textures, and improving the overall visual quality. The underlying principle is to estimate the missing high-frequency information that is lost during the downsampling or compression process. There are several approaches to image super-resolution, including both traditional and deep learning-based methods.

### 3.2 Traditional method for Image Super Resolution

In order to increase the resolution of low-resolution images, traditional approaches of image super-resolution concentrate on using mathematical and signal processing techniques. Here are a few such conventional methods:

- Bicubic interpolation: It is a standard method in image interpolation field because of its low complexity and relatively good results. But as it only interpolates in horizontal and vertical directions, edges easily suffer from artifacts such as blocking, blurring and ringing[10].

- Frequency Domain: This method operate on the frequency domain representation of the image, such as the Discrete Fourier Transform (DFT) or Discrete Wavelet Transform (DWT). They exploit the fact that high-frequency information is typically attenuated during downscaling and aim to recover or amplify these high-frequency components to improve image quality. Examples of frequency domain methods include Fourier-based interpolation and wavelet-based methods.

- Edge-based: The perception of visual details depends heavily on edge information. The edges and contours of low-resolution images are enhanced by edge-based super-resolution techniques. Typically, these methods start with edge detection algorithms and then move on to edge-directed interpolation or edge reconstruction procedures.

## 3.3 Generative Adversarial Networks(GANs)

Generative Adversarial Networks, or GANs, are a deep-learning-based generative model. More generally, GANs are a model architecture for training a generative model, and it is most common to use deep learning models in this architecture. The GAN architecture was first described in the 2014 paper by Ian Goodfellow, et al. titled Generative Adversarial Networks. The initial models worked but where unstable and difficult to train. The GAN model architecture involves two sub-models: a generator model for generating new examples and a discriminator model for classifying whether generated examples are real (from the domain) or fake (generated by the generator model).

- Generator: Model that is used to generate new plausible examples from the problem domain.

- Discriminator. Model that is used to classify examples as real (from the domain) or fake (generated).

Generative adversarial networks are based on a game theoretic scenario in which the generator network must compete against an adversary. The generator network directly produces samples. Its adversary, the discriminator network, attempts to distinguish between samples drawn from the training data and samples drawn from the generator[2].

### 3.3.1 The Generator Model

The generator model takes a fixed-length random vector as input and generates a sample in the domain, such as an image. A vector is drawn randomly from a Gaussian distribution and is used to seed or source of noise for the generative process. To be clear, the input is a vector of random numbers. It is not an image or a flattened image and has no meaning other than the meaning applied by the generator model. After training, points in this multidimensional vector space will correspond to points in the problem domain, forming a compressed representation of the data distribution. This vector space is referred to as a latent space, or a vector space comprised of latent variables. Latent variables, or hidden variables, are those variables that are important for a domain but are not directly observable[2].

We often refer to latent variables, or a latent space, as a projection or compression of a data distribution. That is, a latent space provides a compression or high-level concepts of the observed raw data such as the input data distribution. In the case of GANs, the generator model applies meaning to points in a chosen latent space, such that new points drawn from the latent space can be provided to the generator model as input and used to generate new and different output examples[2].

After training, the generator model is kept and used to generate new samples.

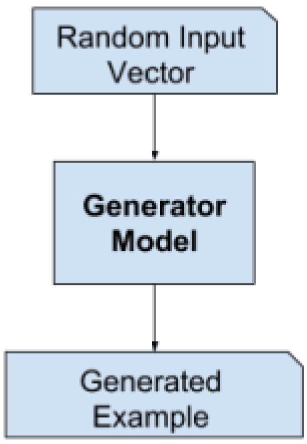



Figure 2: Generator Model

### 3.3.2 The Discriminator Model

The discriminator model takes an example from the problem domain as input (real or generated) and predicts a binary class label of real or fake (generated). The real example comes from the training dataset. The generated examples are output by the generator model. The discriminator is a normal classification model.

After the training process, the discriminator model is discarded as we are interested in the generator. Sometimes, the generator can be repurposed as it has learned to effectively extract features from examples in the problem domain. Some or all of the feature extraction layers can be used in transfer learning applications using the same or similar input data[2].

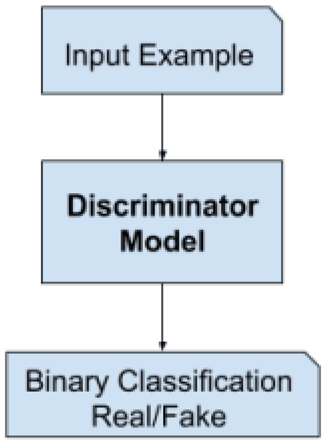



Figure 3: Discriminator Model

## 3.4 Application of GANs in Image Super Resolution

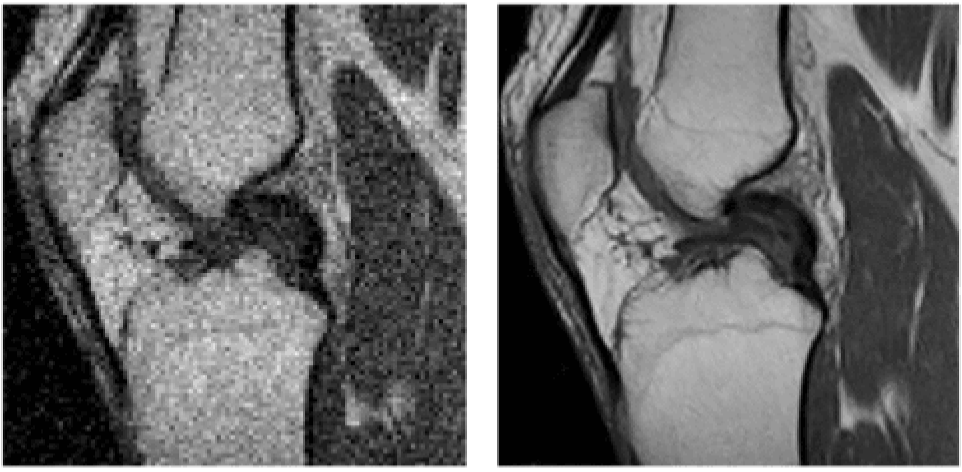
Image Super Resolution refers to the task of enhancing the resolution of an image from low-resolution (LR) to high (HR). It is popularly used in the following applications:

### 3.4.1 Surveillance

To detect, identify, and perform facial recognition on low-resolution images obtained from security cameras.

### 3.4.2 Medical Diagnosis

Various medical imaging techniques can provide anatomical and functional information about the human body structure. However, resolution limits always reduce the relevance of medical imaging in diagnosis. Magnetic resonance imaging (MRI), functional MRI (fMRI), and positron emission tomography (PET) have all utilised SR technology. The goal is to improve the resolution of medical images while retaining true isotropic 3-D imaging. Medical imaging systems can work in highly regulated conditions, allowing for the acquisition of continuous and multi-view pictures[17].



(a) (b)

Figure 4: The single-frame SR result on the MRI knee image. (a) The original LR data. (b) The SR result.

### 3.4.3 Regular video information enhancement

The application of SR techniques has entered our daily life. LR video images can be converted to high-definition images using SR techniques. Hitachi Ltd. achieved the conversion of standard definition TV (SDTV) to high-definition television (HDTV) using SR technology for videos.

### 3.4.4 Biometric information identification

SR is also important in biometric recognition, including resolution enhancement for faces fingerprints and iris images. The resolution of biometric images is pivotal in the recognition and detection process. To deal with the LR observations, a common approach is the development of high-quality images from multiple LR images. Based on the redundancy and similarity in the structured features of biometric images, example-based single-frame SR with an external database is an effective way of resolution enhancement. Using SR, the details of the shapes and structural texture are clearly enhanced, while the global structure is effectively preserved, which can improve the recognition ability in the relevant applications [17].

# Chapter 4
# Methodology

## 4 Methodology

Our proposed method for real-world super-resolution are discussed...

### 4.1 Input Image

In Image Super Resolution(ISR) the aim is to estimate a high-resolution, super resolution image $I^{SR}$ from a low-resolution input image $I^{LR}$. Here $I^{LR}$ is the low-resolution version of its high resolution counterpart $I^{HR}$. The high-resolution images are only available during training. In training, $I^{LR}$ is obtained by resize (32x32) the $I^{HR}$. The LR image as input to the generator model for obtain SR image.[9]

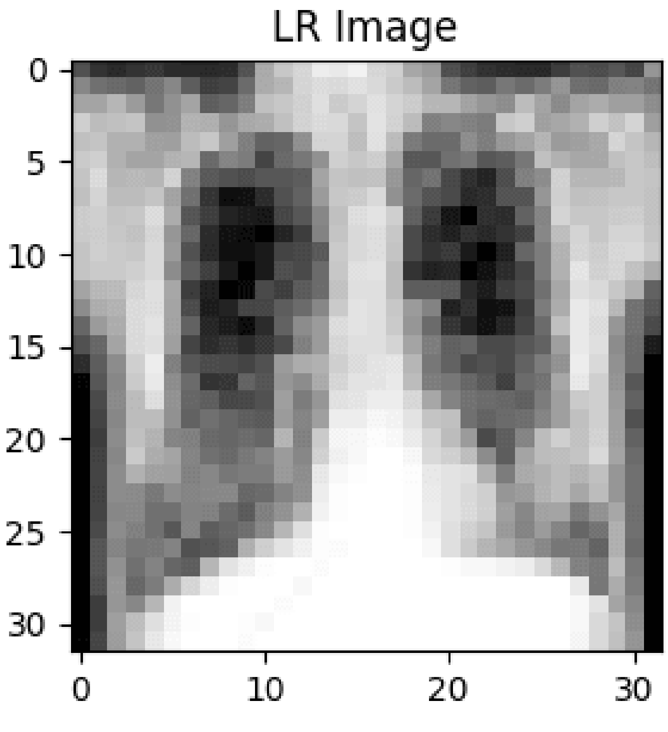



Figure 5: Input Image

## 4.2 Generator Function

Our ultimate goal is to train a generator function that estimates for a given LR input image its corresponding HR counterpart. To achieve this, we train a generator network as a feed-forward CNN[9]. We are providing a LR image to the generator model as an input, and it will create a fake SR image that looks like a real image, then forward the generated image to the discriminator model for classification.

## 4.3 Discriminator Function

A discriminator is used to learn what characteristics make images real. In order to classify SR images, the model will accept both HR and SR images as inputs.The discriminator will provide feedback to the generator model in order for it to create images that are similar to the original image.The discriminator examines both real images (training samples) and produced images separately. It determines whether the discriminator's input image is real or generated.

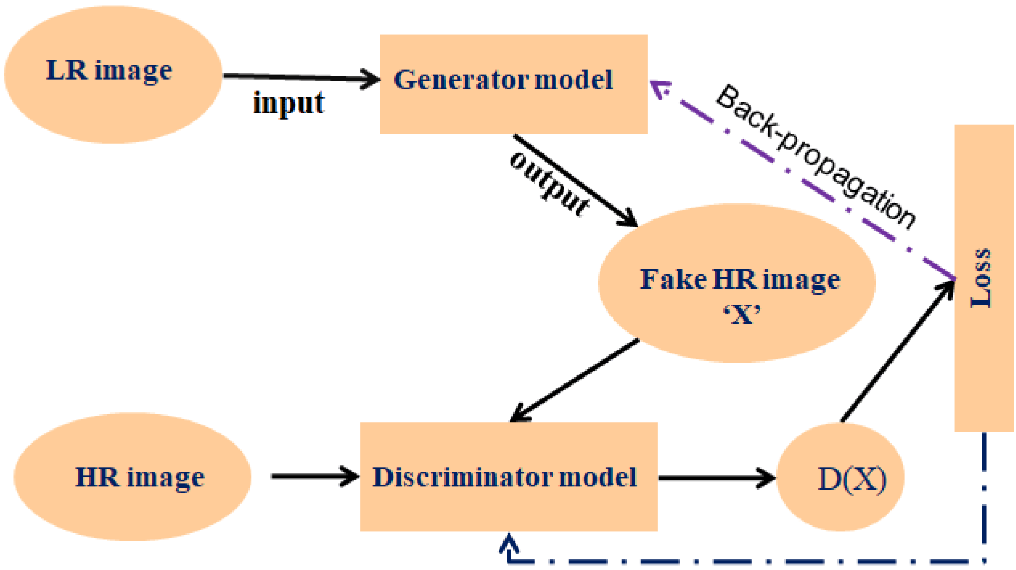



Figure 6: SRGAN model from LR image to HR image

## 4.4 Loss

The discriminator outputs a value D(x) indicating the chance that x is a real image. The discriminator objective is to maximize the chance to recognize real images as real and generated images as fake. The generator model parameters will be updated and the discriminator parameter won't change if the discriminator determines that the generated image is fake at that point. In contrast, when the discriminator is unable to identify a fake image, the parameters of the discriminator model are adjusted, while those of the generated model are left unchanged. This method will be continued until the discriminator model loss reaches one and the generator model loss is close to being too small.

# Chapter 5
# Implementations

## 5 Implementations

### 5.1 Dataset

The first stage is to acquire a dataset of low-resolution (LR) and high-resolution (HR) images. The LR images will be used as input to the GAN, and the HR images will be used as ground truth. The dataset should be large enough to train the GAN effectively. DIV2K dataset consists of high-quality images collected on the Internet, and it contains 1000 RGB images with a resolution of 2 K. The DIV2K dataset contains a variety of categories of content, including people, natural environments, flora and fauna, and more. The dataset is divided into a training set, a validation set, and a test set, and the numbers are 800, 100, and 100, respectively[5].

### 5.2 Hardware and Software Configuration

Depending on the particular project, the hardware and software configuration for an image super-resolution (ISR) project based on GANs can change. However, a few standard prerequisites are as follows:

- A powerful GPU: GANs are computationally expensive to train, so a powerful GPU is essential for achieving good results.
- A large amount of memory: GANs require a large amount of memory to store the training data and the model parameters.
- A high-speed internet connection: The training data and the model parameters can be large, so a high-speed internet connection is necessary for downloading and uploading them.

  We used Google Colab to achieve this Hardware goal.

The software setup must also meet the following criteria in addition to the hardware requirements:

- A deep learning framework: A deep learning framework, such as Keras, PyTorch or TensorFlow, is needed to train the GAN.

- TensorFlow: TensorFlow is a powerful deep learning framework that is used by many large companies.

- Keras: Keras is a high-level API that can be used with either PyTorch or TensorFlow.

## 5.3 Image Data Processing

We collected the DIV2K dataset, and the sizes of those images were not the same, so we resized them. For low-resolution and high resolution images, resize all of the gathered images to $32 \times 32$ and $128 \times 128$ pixels, respectively, and save those images in two folders.

## 5.4 Network Architecture Design

### 5.4.1 SRGAN Generator

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

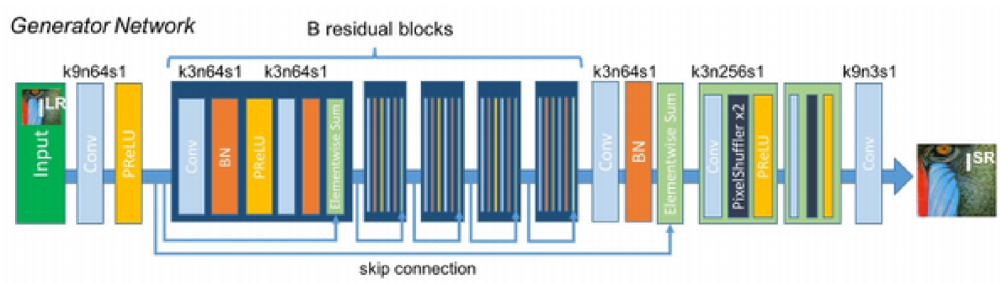



Figure 7: Generator Network[9]

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[9]. 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[6].

During the training, A high-resolution image (HR) is downsampled to a low-resolution image (LR). The generator architecture than tries to upsample the image from low resolution to super-resolution. After then the image is passed into the discriminator, the discriminator and tries to distinguish between a super-resolution and High-Resolution image and generate the adversarial loss which then backpropagated into the generator architecture.

### 5.4.2 SRGAN Discriminator

The task of the discriminator is to discriminate between real HR images and generated SR images. We follow the architectural guidelines summarized by Radford et al. and use LeakyReLU activation ( = 0.2) and avoid max-pooling throughout the network[9]. It contains eight convolutional layers with an increasing number of $3 \times 3$ filter kernels, increasing by a factor of 2 from 64 to 512 kernels as in the VGG network. Strided convolutions are used to reduce the image resolution each time the number of features is doubled. The resulting 512 feature maps are followed by two dense layers and a final sigmoid activation function to obtain a probability for sample classification[6].

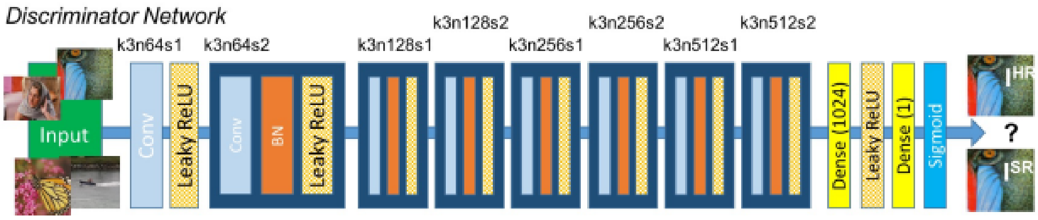



Figure 8: Discriminator Network[9]

Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.
Few things to note from Network architecture:

- Residual blocks: Since deeper networks are more difficult to train. The residual learning framework eases the training of these networks, and enables them to be substantially deeper, leading to improved performance. More about Residual blocks and Deep Residual learning can be found in paper given below. 16 residual blocks are used in Generator.

- PixelShuffler x2: This is feature map upscaling. 2 sub-pixel CNN are used in Generator. Upscaling or Upsampling are same. There are various ways to do that. In code keras inbuilt function has been used.

- PRelu(Parameterized Relu): We are using PRelu in place of Relu or LeakyRelu. It introduces learn-able parameter that makes it possible to adaptively learn the negative part coefficient.

- k3n64s1 this means kernel 3, channels 64 and strides 1.

## 5.5 Loss Function

The definition of our perceptual loss function $l^{SR}$ is critical for the performance of our generator network. While $l^{SR}$ is commonly modeled based on the MSE, we improve on Johnson et al. and Bruna et al. and design a loss function that assesses a solution with respect to perceptually relevant characteristics. We formulate the perceptual loss as the weighted sum of a content loss ($l^{SR}$) and an adversarial loss component.This loss is very important for the performance of the generator architecture:

### 5.5.1 Content Loss

In this study, we employ two types of content loss: pixelwise MSE loss for the SRResnet architecture, which is the most prevalent MSE loss for picture Super Resolution. However, MSE loss is unable to handle with high frequency material in the image, resulting in extremely smooth images. As a result, the authors of the research opted to use loss of several VGG layers. This VGG loss is based on the ReLU activation layers of the pre-trained 19-layer VGG network[9]. A pretrained VGG network output is compared pixel-by-pixel in order to determine the content loss, which is defined as a VGG loss. Only when the input images themselves are comparable will the outputs of the real image VGG and the fake image VGG be similar. This is done with the idea that pixel-by-pixel comparison will strengthen the main goal of achieving super-resolution. The results are really favorable when the GAN loss and the content loss are combined.

### 5.5.2 Adversarial Loss

The Adversarial loss is the loss function that forces the generator to image more similar to high resolution image by using a discriminator that is trained to differentiate between high resolution and super resolution images[6].

## 5.6 Training

Train the GAN model using the paired low-resolution and high-resolution images from the dataset. The generator and discriminator networks are trained alternately in a min-max game. The generator tries to generate high-resolution images that deceive the discriminator, while the discriminator tries to accurately classify real and fake images.

## 5.7 Optimization

Use an optimization algorithm such as stochastic gradient descent (SGD) or Adam to update the weights of the generator and discriminator networks during training. Adjust the learning rate and other hyperparameters to ensure effective convergence and stability.

## 5.8 Batch Normalization

Batch Normalization (BN) is a normalization technique used on the input of each layer to reduce covariance shift (change of distribution of activations in intermediate layers). Wang, et al. claim that removing the BN improves the image quality, but it is also known to allow a wider range of selections for hyper-parameters. Determining hyper-parameters increases the experimental budget exponentially when we change the image size, and BN helps alleviate this cost. Using batch normalization helps the network to converge more easily[14].

## 5.9 Batch Size

Batch size is a hyperparameter in machine learning that specifies how many training examples must be processed before the model parameters are updated. The batch size can be any positive integer. There are several factors to consider when choosing a batch size. The model's accuracy may increase with a larger batch size, but this may come at the expense of additional memory usage and processing time. Even though it may require less memory and processing time, a smaller batch size may result in less accuracy. In general, the GPU memory size is a good place to start when determining batch size. The model won't fit in the GPU memory if the batch size is too large, making training impossible. The model won't be able to utilize the GPU's parallel processing capabilities if the batch size is too small, which will result in a slower training process.

## 5.10 Learning Rate

The learning rate, a hyperparameter in deep learning, regulates how frequently the model parameters are updated during training. A high learning rate will speed up the model's learning, but it also runs the risk of overfitting the training set. The model will learn slowly if the learning rate is low, but it also increases the likelihood that the model will underfit the training set. Depending on the particular model and the training set, the ideal learning rate will be determined.

# Chapter 6
# Results and Discussion

## 6 Results and Discussion

### 6.1 Results

On the widely used benchmark dataset known as DIV2K, we performed experiments. These experiments involved $4 \times$ *upsampling* of both rows and columns. The main output of our project is a super resolution $128 \times 128$ image that was created from a $32 \times 32$ low resolution image. The Peak Signal-to-Noise Ratio (PSNR) is then calculated to assess the quality, and the output result, a super resolution image, is compared to the original $128 \times 128$ image using the Structural Similarity Index (SSIM).

A batch size of 10 and 220 epochs were employed during the model's training. As the number of epochs was increased, the generator loss gradually decreased. The generator loss during training is depicted in figure 9.

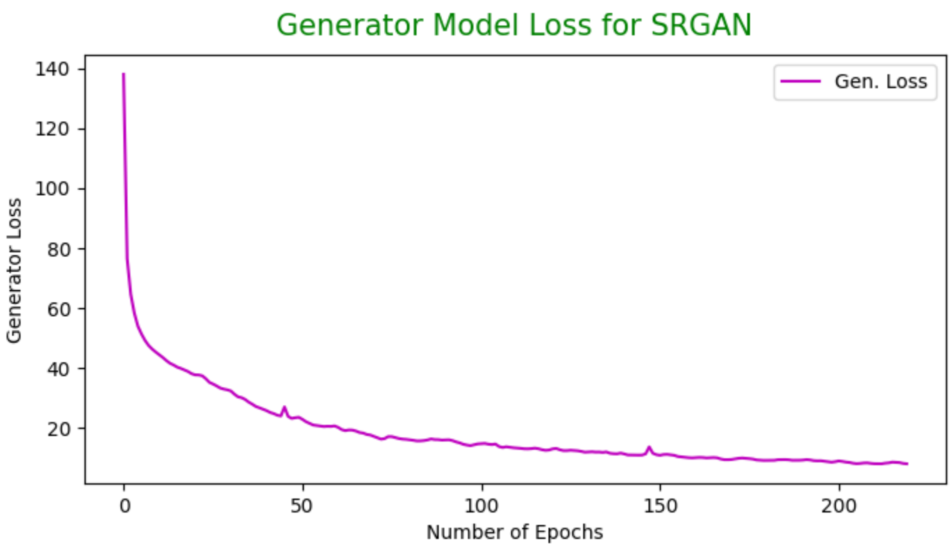



Figure 9: Generator Loss on each Epochs

Figure 10 shows the category of image we will use as input for the generator.

Figure 10: $32 \times 32$ input image

Figure 11: Zoom Figure 10 image

The generator function takes an input of $32 \times 32$ and produces the output $128 \times 128$ at figure 12, which has been scaled up by 4 times from the input image size $32 \times 32$.

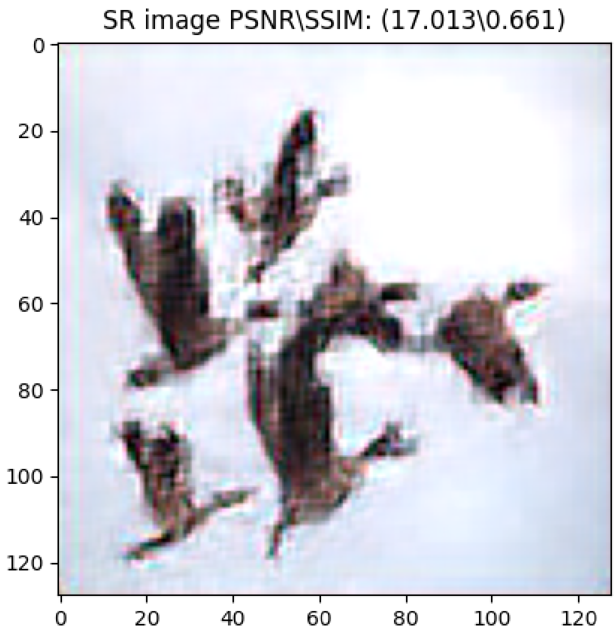



Figure 12: Super Resolution image

We used the original High Resolution image in addition to the Super Resolution image to calculate the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) at Figure 13.

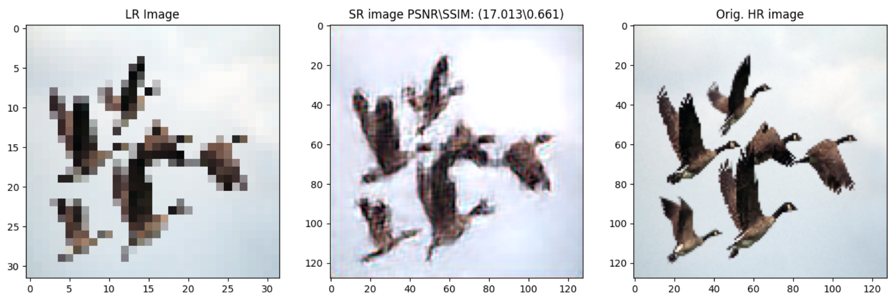



Figure 13: LR, SR & Orig. HR image

Figure 12 shows an SSIM value of 0.661, indicating that the generated super-resolution image is 66% similar with the original high-resolution image.

#### 6.1.1 Peak Signal-to-Noise Ratio (PSNR)

Currently, PSNR-oriented network structures have been used as the generators of GAN models in the majority of GAN based perceptual SISR studies[5]. Image quality is influenced by the ratio between noise distortion and the maximum signal strength possible. related to the mean squared pixel errors. PSNR provides a quantitative measure of the quality of the generated high-resolution images compared to the ground truth high-resolution images. A higher PSNR value suggests that the two images are more similar.Because of its straight forward structure, PSNR, which assesses the effectiveness of image compression, is still widely used, particularly for video images. When comparing super resolution images, it is one of the most trustworthy techniques[7].

### 6.1.2 Structural Similarity Index (SSIM)

The Structural Similarity Index (SSIM) is a full-reference image quality assessment metric that compares two images based on their structural similarity. SSIM was first introduced in 2004 by Wang et al. and has since become one of the most popular image quality assessment metrics[3]. It is employed to assess how structurally similar two images are. calculated by accounting for the brightness, contrast, and structural aspects of both images. More similar images are indicated by a higher SSIM value[7]. The SSIM values ranges between 0 to 1, 1 means perfect match of the reconstruct image with the original one. Generally, higher SSIM values indicate good quality reconstruction techniques.

## 6.2 Discussion

The perceptual quality of super-resolved images was prioritized over computational efficiency in this work. The model presented here is not optimized for real-time video SR. However, preliminary network architecture experiments indicate that shallower networks have the potential to provide very efficient alternatives at a minor reduction in qualitative performance. Deeper networks ($B > 16$) can improve model performance even further, but at the cost of longer training and testing times. Because of the appearance of high-frequency artifacts, SRGAN variants of deeper networks are becoming increasingly difficult to train. In this project, we decided to focus on the Single Image Super Resolution based on GANs models. The reason behind this is that while under the same broad term of SISR they are structured and optimized for entirely different tasks. In our observations, we deduced that a SISR model that has been built to work on LR images that have a lower amount of information than others tend to use sparsity based SISR structures. On the other hand, the GAN based SISR models dominated the research areas that worked on images with similar attributes, like human faces, medical imagery, etc. GAN based models were by far the most popular way of implementing SISR, this comes from the ease of implementation and decent performance in every type of research area. When it comes to the best possible performance in any type of objective image.

### 6.2.1 Challenges of GANs for Image Super Resolution

Variations of GANs have achieved excellent performance in image super resolution. Although GANs perform well in image super resolution, they suffer from the following challenges:

- Unstable training. Due to the confrontation between generator and discriminator, GANs are unstable in the training process.

- Large computational resources and high memory consumption. A GAN is composed of a generator and discriminator, which may increase computational costs and memory consumption.

- Complex image super-resolution. Most of GANs can deal with a single task, i.e., image super resolution and synthetic noisy image super-resolution, etc. However, collected images by digital cameras in the real world suffer from drawbacks, i.e., low-resolution and dark-lighting images, complex noisy and low-resolution images. Besides, digital cameras have higher requirement on the combination of image low-resolution and image recognition[15].

# Chapter 7
# Conclusion & Future Works

## 7 Conclusion & Future Works

### 7.1 Conclusion

GANs popularity is growing every day. Two networks competing against one another simultaneously is the GANs most potent feature. At the conclusion of the procedure, more realistic data can be produced, similar to mini-max games. Finally, the loss functions that are employed by GAN are its most crucial component. As The generator and discriminator models have more computational parameters than other models, so tuning the parameters takes more time when training the model. In order to see the results from the low resolution image, we used the DIV2K train dataset for training & testing used DIV2K test dataset by converting into the low resolution and predict the SR outputs that outputs are $4 \times upsampling$ than low resolution images .

### 7.2 Future Works

There are several ways to enhance SRGAN's performance in the future, including:

- Using more data: To enhance SRGAN's performance, more data can be used for training. This might entail gathering more low- and high-resolution images, or it might entail generating more data by using data augmentation techniques.

- Improved Generator Architecture: The deep convolutional neural network (CNN) architecture of the VGG network serves as the foundation for the generator network used by SRGAN. Developing more sophisticated generator architectures that more effectively upsample low-resolution images and capture high-frequency details may be the main focus of future research. This might entail investigating new network architectures, incorporating attention mechanisms, or utilizing methods from other cutting-edge image generation models like StyleGAN.

- Discriminator Enhancement: For the SRGAN generator to produce high-quality super-resolved images, the discriminator network is of utmost importance. The perceptual quality and output resolution could be improved by improving the discriminator's capacity to tell real high-resolution images from super-resolved ones.

- Using more Epochs: The model can produce better results with more training on the datasets.

# References


[1] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. *Advances in Neural Information Processing Systems*, 32, 2019.

[2] Jason Brownlee. Generative adversarial networks with python deep learning generative models for image synthesis and image translation. pages 8–10, 2019.

[3] Dominique Brunet, Edward R. Vrscay, and Zhou Wang. On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing*, 21(4):1488–1499, 2012.

[4] Manri Cheon, Jun-Hyuk Kim, Jun-Ho Choi, and Jong-Seok Lee. Generative adversarial network-based image super-resolution using perceptual content losses. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.

[5] Kui Fu, Jiansheng Peng, Hanxiao Zhang, Xiaoliang Wang, and Frank Jiang. Image super-resolution based on generative adversarial networks: a brief review. 2020.

[6] https://www.geeksforgeeks.org/super-resolution-gan srgan/. Super resolution gan (srgan).

[7] Hürkal HÜSEM and Zeynep ORMAN. A survey on image super-resolution with generative adversarial networks. *Acta Infologica*, 4(2):139–154, 2020.

[8] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

[9] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[10] Jing Liu, Zongliang Gan, and Xiuchang Zhu. Directional bicubic interpolation—a new method of image super-resolution. In *3rd International Conference on Multimedia Technology (ICMT-13)*, pages 463–470. Atlantis Press, 2013.

[11] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 20(3):21–36, 2003.

[12] Haoyu Ren, Amin Kheradmand, Mostafa El-Khamy, Shuangquan Wang, Dongwoon Bai, and Jungwon Lee. Real-world super-resolution using generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 436–437, 2020.

[13] Egypt Ahmed181532@bue.edu.eg Super Resolution Enhancement With Generative Adversarial Network Ahmed Tammaa Artificial Intelligence Major The British University in Egypt Cairo, Egypt Ahmed189292@bue.edu.eg Ahmed Ramy Artificial Intelligence Major The British University in Egypt Cairo. Super resolution.

[14] Nao Takano and Gita Alaghband. Srgan: Training dataset matters. *arXiv preprint arXiv:1903.09922*, 2019.

[15] Chunwei Tian, Xuanyu Zhang, Jerry Chun-Wei Lin, Wangmeng Zuo, Yanning Zhang, and Chia-Wen Lin. Generative adversarial networks for image super-resolution: A survey. *arXiv preprint arXiv:2204.13620*, 2022.

[16] Zhengwei Wang, Qi She, and Tomas E Ward. Generative adversarial networks in computer vision: A survey and taxonomy. *ACM Computing Surveys (CSUR)*, 54(2):1–38, 2021.

[17] Linwei Yue, Huanfeng Shen, Jie Li, Qiangqiang Yuan, Hongyan Zhang, and Liangpei Zhang. Image super-resolution: The techniques, applications, and future. *Signal processing*, 128:389–408, 2016.

[18] Wenlong Zhang, Yihao Liu, Chao Dong, and Yu Qiao. Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3096–3105, 2019.

[19] Lu Zhou, Xiaobo Lu, and Li Yang. A local structure adaptive super-resolution reconstruction method based on btv regularization. *Multimedia tools and applications*, 71(3):1879–1892, 2014.