TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PURWANCHAL CAMPUS

COMPARISON OF DIFFERENT IMAGE SUPER RESOLUTION TECHNIQUES

BY:

ABHINAV JAISHI (PUR076BEI001)

MEGHA THAKUR (PUR076BEI021)

SANGHARSHA DAHAL (PUR076BEI029)

KABIN POUDEL (PUR076BEI047)

A MID TERM PROGRESS REPORT TO THE DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE BACHELOR’S DEGREE IN ELECTRONICS, COMMUNICATION AND INFORMATION ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

PURWANCHAL CAMPUS

DHARAN, NEPAL

January 7, 2024

ABSTRACT

The instructors are responsible to ensure the smoothness of the classroom activities alongside with the monitoring the student's attendance, attention and activities. Manual observation is a tedious job and affects the whole learning process. With the incorporation of IOT devices and computational algorithms

The collected preliminary data-set from area around Kathmandu valley and the country's major lines are able to map some interesting features and environmental proxies that are visualised and the patterns and variations in it are explored using various models namely such as Autoregressive integrated moving average (ARIMA), Recurrent neural network (RNN), which worked best for time series database.

Keywords: a, b, c, d

TABLE OF CONTENTS

List of Figures

List of Tables

1. INTRODUCTION

Image super resolution is a technique used to enhance the resolution and quality of image beyond its original size. It is useful technique in the field of image processing. It can be used to enhance the quality of image captured in night, shaky image etc. There are various techniques and algorithm used for image super resolution, including both traditional methods and deep learning-based approaches. Traditional super image resolution techniques are based on mathematics techniques. These techniques have been used for many years. Some of the traditional super resolution techniques include: Interpolation, Edge-Directed interpolation, Bayesian methods etc. Traditional methods have certain limitations compared to deep learning approaches. They may struggle to capture complex patterns and textures, and their performance is often constrained by the hand-crafted features and assumptions used in the algorithms. Deep learning based super resolution typically employs Convolution Neural Network (CNN) to learn the mapping between low-resolution and high-resolution images. The network is trained on large dataset of paired low-resolution and high- resolution images. During training, the networks learn to identify patterns and features that enables it to generate high-resolution details from low-resolution inputs.

1.1. Statement of problem

The problem of super image resolution arises when we encounter low-resolution images that lack the desired level of detail and clarity. These low-resolution images can be the result of various factors such as limitations in capturing hardware, or resizing operations. The inadequate resolution in these images hinders their effective use in various applications, where higher resolution and finer details are essential for accurate analysis, interpretation, and visualization.

1.2. Objectives

- Compare different Super Resolution algorithms.
- Recommend

1.3. Scope and Applications

Solving the super image resolution problem has numerous practical applications in field like photography, surveillance, and digital art where enhanced image quality is crucial for accurate decision-making, analysis, and interpretation. The development of efficient and accurate super-resolution method involves exploring traditional signal processing techniques, as well as cutting-edge deep learning approaches, to achieve optimal results while considering computational efficiency and resource constraints.

Super image resolution can be used for following things:

1. Super resolution can be used in digital cameras and smartphones to provide better zoom capabilities without significant loss of image quality.
2. Super resolution can enhance the resolution of individual frames in videos, leading to improved video quality and better extraction of information.
3. Super resolution can be used to improve the resolution of old and degraded images, preserving historical photographs, artworks, and documents.
4. In surveillance systems, low resolution camera feeds can be upscaled to improve the ability to identify faces, license plates, or other critical details, helping in investigations.

2. LITERATURE REVIEW

In the timeline of image scaling process, the first used methods were different interpolation techniques and sparse representation-based methods. But with the advent of deep-learning based image scaling techniques interpolation methods were less commonly used in image enhancement.

D. Han (2013) **“Comparison of commonly used image interpolation methods”** discusses multiple interpolation methods like nearest neighbor, bilinear, bicubic, etc. Although interpolation is fast but have some disadvantages. But using these methods was not optimal for our use case as they may lead to the loss of fine details and sharpness in image and can amplify noise present in image which may not be accurate in real-world images.[1]

C. Dong, et. al, **“Image super-resolution using deep convolutional networks”** is a seminal paper on use of Deep Learning for the use of for the task of SR. It only consists of three layers and requires the LR image to be up-sampled using bicubic interpolation prior to being processed by the network, but it was shown to outperform the state-ofthe-art methods of that time.[2]

C. Ledig et al., **“Photo-realistic single image super-resolution using a generative adversarial network”** discusses about a Generative Adversarial Network (GAN) for image super-resolution (SR). To our knowledge, it is the first framework capable of inferring photo-realistic natural images for 4x upscaling factors. To achieve this, it uses a perceptual loss function which consists of an adversarial loss and a content loss. The adversarial loss pushes the solution to the natural image manifold using a discriminator network that is trained to differentiate between the super-resolved images and original photo-realistic images. In addition, it uses a content loss motivated by perceptual similarity in VGG space instead of similarity in pixel RGB space. This deep residual network is able to recover photo-realistic textures from heavily down sampled images on public benchmarks. An extensive mean-opinion-score (MOS) test shows hugely significant gains in perceptual quality using SRGAN. The MOS scores obtained with SRGAN are closer to those of the original high-resolution images than to those obtained with any state-of-the-art method at the time.[3]

3. HARDWARE AND SOFTWARE REQUIREMENTS

3.1. Hardware Requirements

We used the system with following specifications for training:

- PC with Windows 11
- 16 GB RAM
- Nvidia Geforce GTX 1050(4GB)
- 256 GB storage
- Processor: Intel(R) Core(TM) i5-8300H CPU @ 2.30GHz

3.2. Software Requirements

The software requirements for developing, training, evaluating, and deploying the model effectively are listed below:

1. **Integrated Development Environment (IDE):** We extensively used Visual Studio Code (VS Code) as IDE for training, testing and documentation also.
2. **Version Control System:** We used Git and Github.
3. **Web framework:** We simply used Streamlit framework.
4. **Deep-Learning Frameworks:** We used PyTorch.

4. METHODOLOGY

Image Super Resolution is a branch of Artificial Intelligence that deals with upscaling a Low-Resolution Image to High Resolution Image, filling in the missing pixels with the different techniques. There are other simpler methods to upscale images like Linear or Bi-cubic interpolation, but they do not generate any new information based on the environment and hence are not super useful to upscale an LR image. Deep Learning based methods require huge amounts of data so that the model is not overfitted, the dataset needs to contain an HR LR version of the same image that are perfectly aligned to each other. We need to synthetically create LR images from HR images. The 'classical' degradation model is the most used, which considers down-sampling, blurring, and noise:

$$I^{LR} = (I^{HR} \circledast k) \downarrow_s +n$$

where $\circledast$ represents a convolution operation, k is a kernel (typically a Gaussian blurring kernel, but it can also represent other functions such as the Point Spread Function (PSF)), n represents additive noise, and $\downarrow$s is a downscaling operation that is typically assumed to be bicubic down-sampling with scale factor s. SR aims to then reverse whichever degradation process is considered, to retrieve the original underlying high-fidelity image. We have implemented SSResNet and SRGAN till now. The methodology for them is explained below:

1. SRResNet: The SRResNet is a fully convolutional network designed for 4x super-resolution. It incorporates residual blocks with skip connections to increase the optimizability of the network despite its significant depth. The SRResNet is trained and used as a standalone network and provides a nice baseline for the SRGAN – for both comparison and initialization. We use the following architecture for this network.

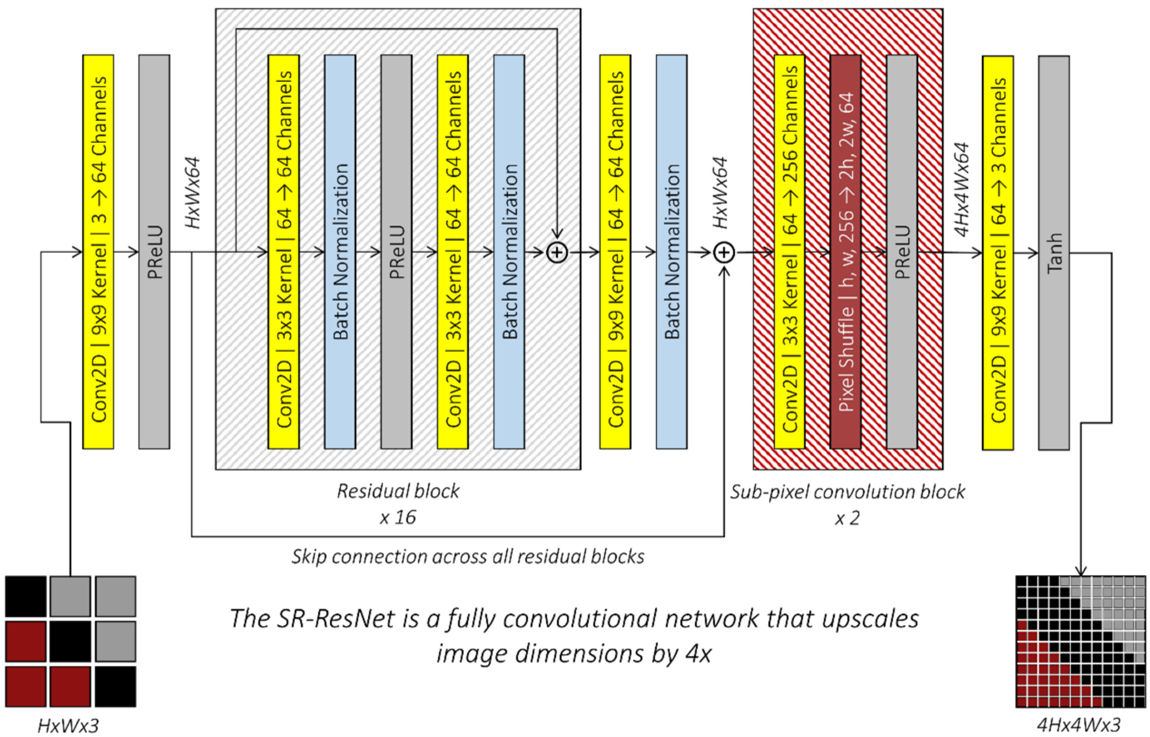

Figure 1: SRResNet Architecture

The SRResNet first contains convolution block of large Kernel Size 9x9, stride of 1 and 64 channels with PreLU activation. There are 16 residual blocks with convolution layer of Kernel size 3x3 followed by batch normalization, PreLU same conv layer again and batch normalization again. The output then is passed through same 3x3 conv layer and batch normalized. Two subpixel convolution blocks are used followed by PreLU activation each of which provides two times upscaling. Finally, a convolution with large kernel size of 9x9 and 1 stride with 3 out channels for RGB is done with Tanh activation to get super resolved image.

In Forward Pass, SRResNet produces a 4x upscaled image from provided low-res image using the above architecture. Mean-Squared Error (MSE) is used as the loss function to compare the upscaled image and original high-quality image. MSE is a type of content loss but here it only looks in RGB space of predicted and target images. Minimizing the MSE by changing the parameters of the network will make the model produce images closer to the original images.

2. SRGAN: Super-Resolution Generative Adversarial Network consist of two adversary networks Generator and Discriminator which are trained in tandem. The goal of the Generator is to learn to super-sample an image such that Discriminator can't tell difference between artificial and natural origins. The interplay between these two networks leads to the improvement of both over time.

   The Generator learns not only by minimizing content loss, as in the case of the SRResNet but also by spying on the Discriminator's methods. By providing the Generator access to the Discriminator's inner workings in the form of the gradients produced therein when backpropagating from its outputs, the Generator can adjust its parameters in a way that alters the Discriminator's outputs in its favour. As the Generator produces more realistic high-resolution images, we use these to train the Discriminator, improving its discriminating abilities.

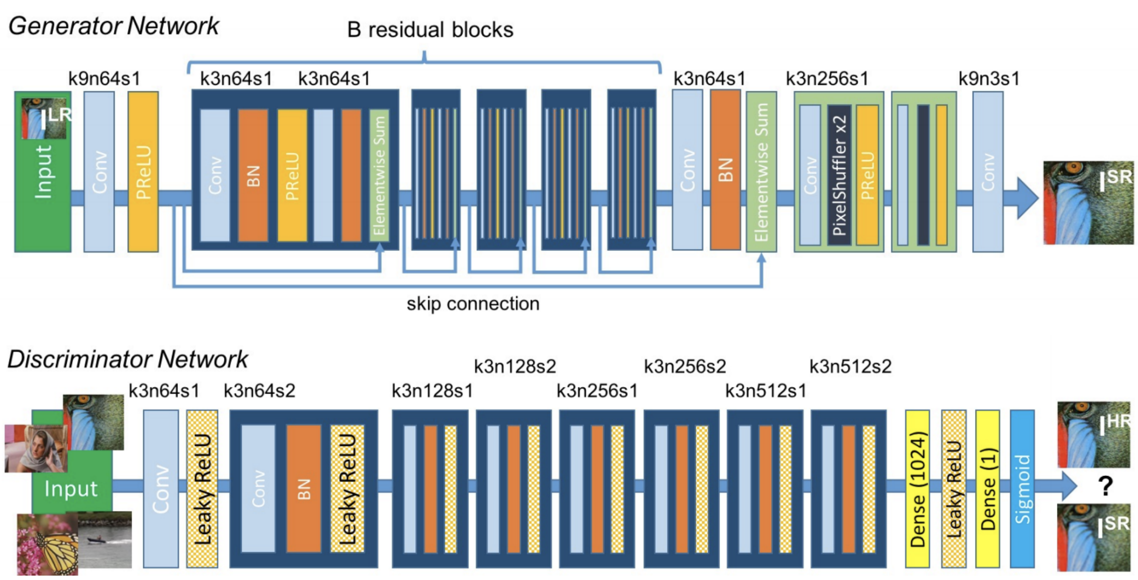

Figure 2: SRGAN Architecture

The Generator has same architecture as SRResNet. Training Discriminator is not different from any binary image classifier. It is trained on both real images and generated images from generator. In Forward Pass, it outputs the probability score P_{HR}. If the provided image was real, we desire the P_{HR} to be as high as possible towards 1. And for generated image as low as possible towards 0. We use Binary

Cross Entropy loss $-logP_{HR}$ when input image is real and $-logP_{SR}$ when image is generated. Here $P_{SR} = 1 - P_{HR}$. Minimizing these losses by changing the parameters will make discriminator predict higher probability for real images and lower probability for generated images.

We use better content loss than that of SRResNet. In ResNet we use MSE loss in RGB space which produces overly smooth images with no finer details. There are a lot of possibilities of similar pixel combination that can be formed from low resolution image patch. When we use content loss in RGB space, it averages the output rather than choosing one of the combinations which would produce better result as an overly smooth "averaged" prediction will always have lower MSE.

We can use CNN trained to classify images to find deeper meaning of patterns in images. This new representation space is more suitable for calculating content loss and can hallucinate new details showing creativity. We specifically use the VGG19 network as recommended in the paper. We use MSE-based content loss in this VGG space to compare the images. The use of content loss is only one component of the generator update, we use adversarial loss obviously. The super-resolved image is passed through the Discriminator with its weight frozen not to update the discriminator but to get the probability score P_{HR} with misleading label and use the BCE loss $-logP_{HR}$ and resulting gradient information to update the Generator's weights. From this loss formulation, we are using gradient information in the Discriminator not to update the Discriminator but to rather get gradient information in the generator via backpropagation to update the Generator weight such that it produces images closer to its natural origin. GAN is trained in an interleaved fashion, where the generator and discriminator are alternatively teained for short periods of time. In this paper each component network is updated just once before making the switch.

5. FEASIBLITY ANALYSIS

We are planning to use deep learning to develop a system that can increase the resolution of image. The system is trained on dataset of high-resolution images and low-resolution images generated using degradation model to generate high resolution images from low resolution images.

Deep learning is shown to be very effective for ISR. There are many models developed for this task we can take inspiration from. The technical feasibility of this project is therefore high. The cost of developing the system will depend on the size of the dataset used to train the model, and the complexity of the model. As we are planning to train the model on server and deploy the system on cloud-based service. So, server costs will be our main cost for this project.

The technical and economic feasibility analysis show that we can complete this project successfully.

6. RESULT AND DISCUSSION

6.1. Work Completed

6.2. Work Remaining

6.3. Limitations

6.4. Problem Faced

6.5. Budget Analysis

6.6. Work Schedule

Action plan	2023							2024		
	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
Research										
Data Acquisition										
Model Development and Evaluation										
Model Deployment										
Prepare the report and presentation										

References

[1] Dianyuan Han. Comparison of commonly used image interpolation methods. In *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*. Atlantis Press, 2013/03.

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks, 2015.

[3] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. *CoRR*, abs/1609.04802, 2016.