An Efficient Approach to Store Images using ESRGAN

CSE 3200: System Development Project

Developed by:

Nafiul Alam

Roll: 1807005

Year: 3rd

Semester: 2nd

Batch: 2k18

Partho Choudhury Shoumya

Roll: 1807021

Year: 3rd

Semester: 2nd

Batch: 2k18

Supervised by:

Abdul Aziz

Assistant Professor

Department of Computer Science and Engineering

Khulna University of Engineering and Technology

Department of Computer Science and Engineering

Khulna University of Engineering & Technology

Khulna-9203, Bangladesh

Acknowledgement

At first, we would like to thank Almighty god for giving all his blessings on us to complete our task. It is our immense pleasure to express our deepest gratitude to our supervisor Abdul Aziz, Assistant Professor, Department of Computer Science and Engineering (CSE), Khulna University of Engineering & Technology (KUET) for his continuous encouragement, constant guidance and keen supervision throughout of this project. We shall remember his inspiring guidelines in the future. We are also grateful to our parents and friends for their patience, support and encouragement during this period.

Nafiul Alam

Partho Choudhury Shoumya

Abstract

A well-established storage facility is a requirement to accommodate large amount of data to serve users efficiently. The storage mechanisms exist in the world are quite costly since they tend to store image files in their original form. This project report presents a novel approach to image storage and retrieval using the ESR GAN technology. The proposed system is capable of compressing images to 1/10th of the original size, while still maintaining the same level of quality when restored. This is achieved by training a GAN model on a dataset of images, using a combination of a generator network and a discriminator network to learn the features and patterns present in the images. The generator network is then used to generate new images that are similar to the original ones, while the discriminator network evaluates the authenticity of the generated images. The system utilizes a combination of deep learning and generative adversarial networks (GAN) to perform the required task. Experiments were conducted to evaluate the performance of the system and the results showed that the proposed method was able to achieve a compression ratio of 9:1 with a little loss in the quality of the restored images. Furthermore, the system was found to be able to restore the compressed images with very little visible differences from the original images. A web server has been implemented using Flask framework to demonstrate the practical usage of the proposed system. This project report provides evidence that ESR GAN technology could be an effective solution for image storage and retrieval, reducing the need for large amounts of storage space.

Contents

CHAPTER I

Introduction

1.1 Background

Storing and retrieving images on a server is a common task in many applications, such as online photo albums, social media platforms, and e-commerce websites. The efficiency and speed of this process are crucial, as it directly affects the user experience and the overall performance of the system. Though many attempts have been made in blind super-resolution to restore low-resolution images with unknown and complex degradations, they are still far from addressing general real-world degraded images. In this work, we extend the powerful ESRGAN to a practical restoration application (namely, Real-ESRGAN), which is trained with pure synthetic data. Specifically, a high-order degradation modeling process is introduced to better simulate complex real-world degradations. We also consider the common ringing and overshoot artifacts in the synthesis process. In addition, we employ a U-Net discriminator with spectral normalization to increase discriminator capability and stabilize the training dynamics. Extensive comparisons have shown its superior visual performance than prior works on various real datasets. We also provide efficient implementations to synthesize training pairs on the fly.

1.2 Problem Statement

Traditionally, images are stored in their original format on the server, which can lead to a large storage requirement and retrieval time. On the other hand, bandwidth limitations in user's residing areas lead to an inconvenient experience. So here a need arises to store and retrieve image data in a more efficient manner.

1.3 Motivation

In this project, we aim to extend the powerful ESRGAN [1] to restore general real-world LR images by synthesizing training pairs with a more practical degradation process. The real complex degradations usually come from complicate combinations of different degradation processes, such as imaging system of cameras, image editing, and Internet transmission. For example, when we take a photo with our cellphones, the photos may have several degradations, such as camera blur, sensor noise, sharpening artifacts, and JPEG compression. We then do some editing and upload to a social media app, which introduces further compression and unpredictable noises. The above process becomes more complicated when the image is shared several times on the Internet.

This motivates us to extend the classical "first-order" degradation model to "high-order" degradation modeling for real-world degradations, i.e., the degradations are modeled with several repeated degradation processes, each process being the classical degradation model. Empirically, we adopt a second-order degradation process for a good balance between simplicity and effectiveness. A recent work also proposes a random shuffling strategy to synthesize more practical degradations. However, it still involves a fixed number of degradation processes, and whether all the shuffled degradations are useful or not is unclear. Instead, high-order degradation modeling is more flexible and attempts to mimic the real degradation generation process. We further incorporate sinc filters in the synthesis process to simulate the common ringing and overshoot artifacts.

1.4 Specific Objective

The main objective of the project is to develop an efficient system for making image store and retrieval process cost and time effective. The specific objectives of the project are to

- Develop a web server to store images after compression and then retrieve it to enhance and get the original image using ESR GAN technology
- Evaluate the proposed evaluation model experimentally and theoretically, and
- Compare the proposed model evaluation with the existing method.

1.5 Conclusion

This chapter is ended with a brief discussion of this project background, problem statement, motivation, specific objective, and methodology. In the next chapter, the related research works in this field will be highlighted.

CHAPTER II

Literature Review

2.1 Introduction

Recently, the researchers are more emphasized on developing store retrieve techniques, especially for the image data. This chapter is going to discuss some existing works related to image store and retrieve mechanisms. Some of them are outcome based and some are not. It will also summarize the most important related works.

2.2 Related Works

We focus on deep neural network approaches to solve the SR problem. As a pioneer work, Dong et al. [11,12] propose SRCNN to learn the mapping from LR to HR images in an end-to-end manner, achieving superior performance against previous works. Later on, the _eld has witnessed a variety of network architectures, such as a deeper network with residual learning, Laplacian pyramid structure, residual blocks, recursive learning, densely connected network, deep back projection and residual dense network. Specifically, Lim et al. propose EDSR model by removing unnecessary BN layers in the residual block and expanding the model size, which achieves significant improvement. Zhang et al. propose to use effective residual dense block in SR, and they further explore a deeper network with channel attention, achieving the state-of-the-art PSNR performance. Besides supervised learning, other methods like reinforcement learning and unsupervised learning are also introduced to solve general image restoration problems.

Several methods have been proposed to stabilize training a very deep model. For instance, residual path is developed to stabilize the training and improve the performance. Residual scaling is first employed by Szegedy et al. and also used in EDSR. For general deep networks, He et al. propose a robust initialization method for VGG-style networks without BN. To facilitate training a deeper network, we develop a compact and effective residual-in-residual dense block, which also helps to improve the perceptual quality.

Perceptual-driven approaches have also been proposed to improve the visual quality of SR results. Based on the idea of being closer to perceptual similarity perceptual loss is proposed to enhance the visual quality by minimizing the error in a feature space instead of pixel space. Contextual loss is developed to generate images with natural image statistics by using an objective that focuses on the feature distribution rather than merely comparing the appearance. Ledig et al. propose SRGAN model that uses perceptual loss and adversarial loss to favor outputs residing on the manifold of natural images. Sajjadi et al. develop a similar approach and further explored the local texture matching loss. Based on these works, Wang et al. propose spatial feature transform to effectively incorporate semantic prior in an image and improve the recovered textures.

Throughout the literature, photo-realism is usually attained by adversarial training with GAN. Recently there are a bunch of works that focus on developing more effective GAN frameworks. WGAN proposes to minimize a reasonable and efficient approximation of Wasserstein distance and regularizes discriminator by weight clipping. Other improved regularization for discriminator includes gradient clipping and spectral normalization. Relativistic discriminator is developed not only to increase the probability that generated data are real, but also to simultaneously decrease the probability that real data are real. In this work, we enhance SRGAN by employing a more effective relativistic average GAN.

SR algorithms are typically evaluated by several widely used distortion measures, e.g., PSNR and SSIM. However, these metrics fundamentally disagree with the subjective evaluation of human observers. Non-reference measures are used for perceptual quality evaluation, including

Ma's score and NIQE, both of which are used to calculate the perceptual index in the PIRM-SR Challenge. In a recent study, Blau et al. find that the distortion and perceptual quality are at odds with each other.

The image super-resolution field has witnessed a variety of developments since SRCNN. To achieve visually-pleasing results, generative adversarial network is usually employed as loss supervisions to push the solutions closer to the natural manifold. Most methods assume a bicubic down sampling kernel and usually fail in real images. Recent works also incorporate reinforcement learning or GAN prior to image restoration.

There have been several excellent explorations in blind SR. The first category involves explicit degradation representations and typically consists of two components: degradation prediction and conditional restoration. The above two components are performed either separately or jointly (iteratively). These approaches rely on predefined degradation representations (e.g., degradation types and levels), and usually consider simple synthetic degradations. Moreover, inaccurate degradation estimations will inevitably result in artifacts.

Another category is to obtain/generate training pairs as close to real data as possible, and then train a unified network to address blind SR. The training pairs are usually 1) captured with specific cameras followed by tedious alignments; 2) or directly learned from unpaired data with cycle consistency loss; 3) or synthesized with estimated blur kernels and extracted noise patches. However, 1) the captured data is only constrained to degradations associated with specific cameras, and thus could not well generalize to other real images; 2) learning fine grained degradations with unpaired data is challenging, and the results are usually unsatisfactory.

Degradation models. Classical degradation model is widely adopted in blind SR methods. Yet, real-world degradations are usually too complex to be explicitly modeled. Thus, implicit modeling attempts to learn a degradation generation process within networks. In this work, we propose a flexible high-order degradation model to synthesize more practical degradations.

2.3 Conclusion

This chapter is concluded with a discussion of existing related works about ESR GAN technology and giving a brief explanation about it. In the following chapter, the details of ESR GAN based image server system will be discussed.

CHAPTER III

Theoretical Background

3.1 Introduction

This chapter will discuss the theoretical background used in this project.

3.2 Theoretical Background

Single image super-resolution (SISR), as a fundamental low-level vision problem has attracted increasing attention in the research community and AI companies. SISR aims at recovering a high-resolution (HR) image from a single low-resolution (LR) one. Since the pioneer work of SRCNN, deep convolution neural network (CNN) approaches have brought prosperous development. Various network architecture designs and training strategies have continuously improved the SR performance, especially the Peak Signal-to-Noise Ratio (PSNR) value. However, these PSNR-oriented approaches tend to output over-smoothed results without sufficient high-frequency details, since the PSNR metric fundamentally disagrees with the subjective evaluation of human observers.

Several perceptual-driven methods have been proposed to improve the visual quality of SR results. For instance, perceptual loss is proposed to optimize super-resolution model in a feature space instead of pixel space. Generative adversarial network is introduced to SR by to encourage the network to favor solutions that look more like natural images. The semantic image prior is further incorporated to improve recovered texture details. One of the milestones in the way pursuing visually pleasing results is SRGAN. The basic model is built with residual blocks and optimized using perceptual loss in a GAN framework. With all these techniques, SRGAN significantly improves the overall visual quality of reconstruction over PSNR-oriented methods.

However, there still exists a clear gap between SRGAN results and the ground-truth (GT) images, as shown in Fig. 1. In this study, we revisit the key components of SRGAN and improve the model in three aspects. First, we improve the network structure by introducing the Residual-in-Residual Dense Block (RDDB), which is of higher capacity and easier to train. We also remove Batch Normalization (BN) layers and use residual scaling and smaller initialization to facilitate training a very deep network. Second, we improve the discriminator using Relativistic average GAN (RaGAN), which learns to judge \whether one image is more realistic than the other" rather than whether one image is real or fake". Our experiments show that this improvement helps the generator recover more realistic texture details. Third, we propose an improved perceptual loss by using the VGG features before activation instead of after activation as in SRGAN. We empirically find that the adjusted perceptual loss provides sharper edges and more visually pleasing results. Extensive experiments show that the enhanced SRGAN, termed ES-RGAN, consistently outperforms state-of-the-art methods in both sharpness and details.

We take a variant of ESRGAN to participate in the PIRM-SR Challenge. This challenge is the first SR competition that evaluates the performance in a perceptual-quality aware manner based on, where the authors claim that distortion and perceptual quality are at odds with each other. The perceptual quality is judged by the non-reference measures of Ma's score and NIQE,
i.e., perceptual index = 1/2 ((10-Ma)+NIQE). A lower perceptual index represents a better perceptual quality.

The perception-distortion plane is divided into three regions defined by thresholds on the Root-Mean-Square Error (RMSE), and the algorithm that achieves the lowest perceptual index in each region becomes the regional champion. We mainly focus on region 3 as we aim to bring the perceptual quality to a new high. Thanks to the aforementioned improvements and some other adjustments, our proposed ESRGAN won the first place in the PIRM-SR Challenge (region 3) with the best perceptual index.

In order to balance the visual quality and RMSE/PSNR, we further propose the network interpolation strategy, which could continuously adjust the reconstruction style and smoothness. Another alternative is image interpolation, which directly interpolates images pixel by pixel. We employ this strategy to participate in region 1 and region 2.

3.3 Conclusion

This chapter is terminated with key features of working principles. The following chapter will illustrate the working diagram and dataset preparation of the project.

CHAPTER IV

Proposed Methodology

4.1 Introduction

In this chapter we shall discuss about the theoretical background in this project. The different aspects of outcome based on technologies such as SRGAN, SRCNN, DCNN, ESRGAN, ResNet will be described

4.2 Flask Server Architecture

This project is based upon Flask server. The architecture of Flask server is shown below:

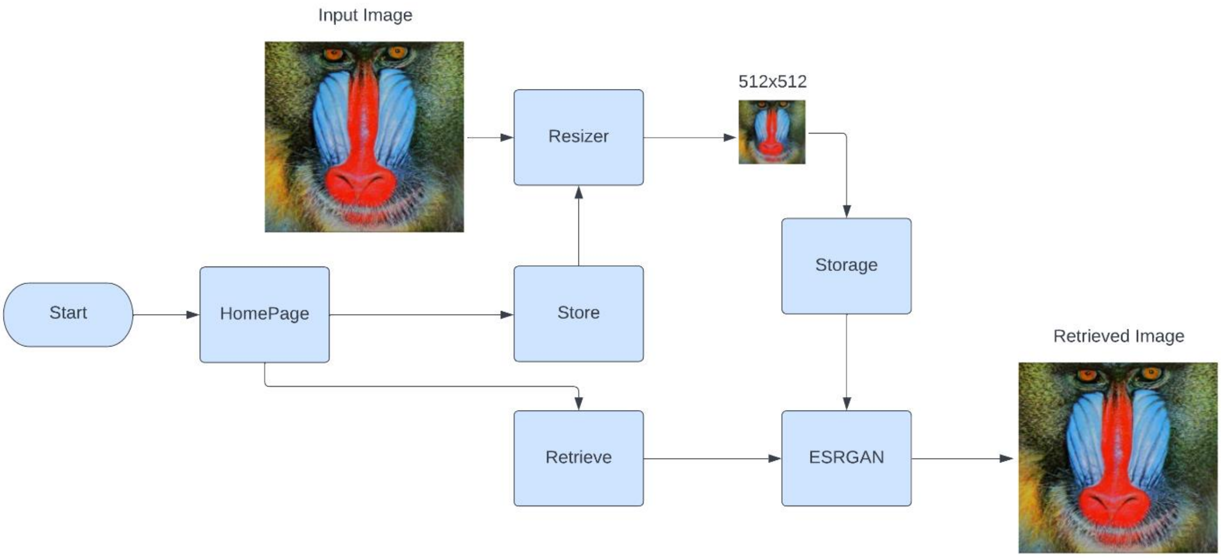

Fig 4.1: Diagram of architecture of Flask server

4.3 Model Architecture

Our main aim is to improve the overall perceptual quality for SR. In this section, werst describe our proposed network architecture and then discuss the improvements from the discriminator and perceptual loss. At last, we describe the network interpolation strategy for balancing perceptual quality and PSNR.

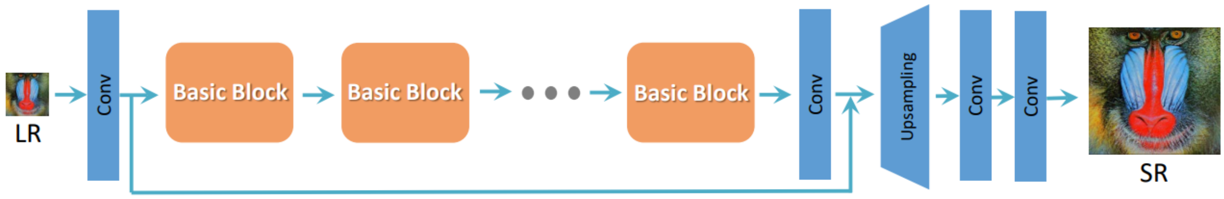

Fig 4.2: We employ the basic architecture of SRResNet where most computation is done in the LR feature space. We could select or design basic blocks" (e.g., residual block, dense block, RRDB) for better performance.

4.4 Network Diagram

In order to further improve the recovered image quality of SRGAN, we mainly make two modifications to the structure of generator G: 1) remove all BN layers; 2) replace the original basic block with the proposed Residual-in-Residual Dense Block (RRDB), which combines multi-level residual network and dense connections as depicted in Fig. 4.1.

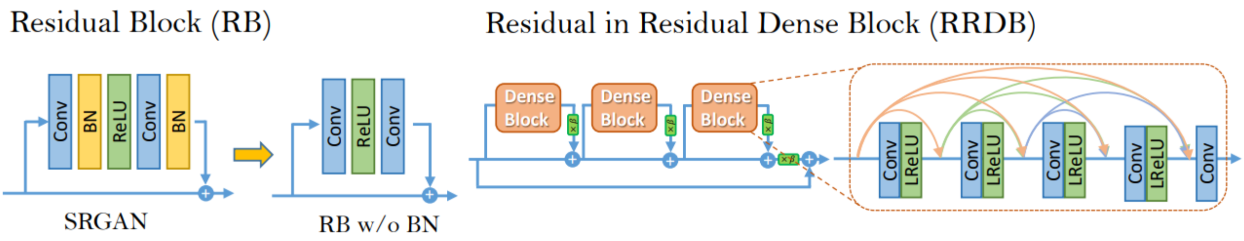

Fig. 4.3: Left: We remove the BN layers in residual block in SRGAN. Right: RRDB block is used in our deeper model and is the residual scaling parameter.

Removing BN layers has proven to increase performance and reduce computational complexity in different PSNR-oriented tasks including SR and deblurring. BN layers normalize the features using mean and variance in a batch during training and use estimated mean and variance of the whole training dataset during testing. When the statistics of training and testing datasets differ a lot, BN layers tend to introduce unpleasant artifacts and limit the generalization ability. We empirically observe that BN layers are more likely to bring artifacts when the network is deeper and trained under a GAN framework. These artifacts occasionally appear among iterations and different settings, violating the needs for a stable performance over training. We therefore remove BN layers for stable training and consistent performance. Furthermore, removing BN layers helps to improve generalization ability and to reduce computational complexity and memory usage.

4.5 Preprocessing

Here we have processed the input image and stored in the server. By using this method, we not only stored efficiently but also preprocessed the image for model.

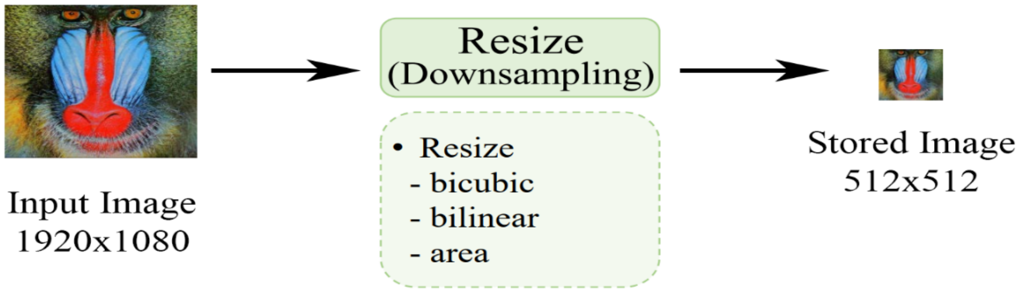

Fig 4.4: Preprocessing image using resize (bicubic, bilinear, area)

4.6 Working Process

The above model has been implemented using a website named “ESRGAN STORE”. The website is built using the Flask framework. When we enter the website, we get to see the homepage:

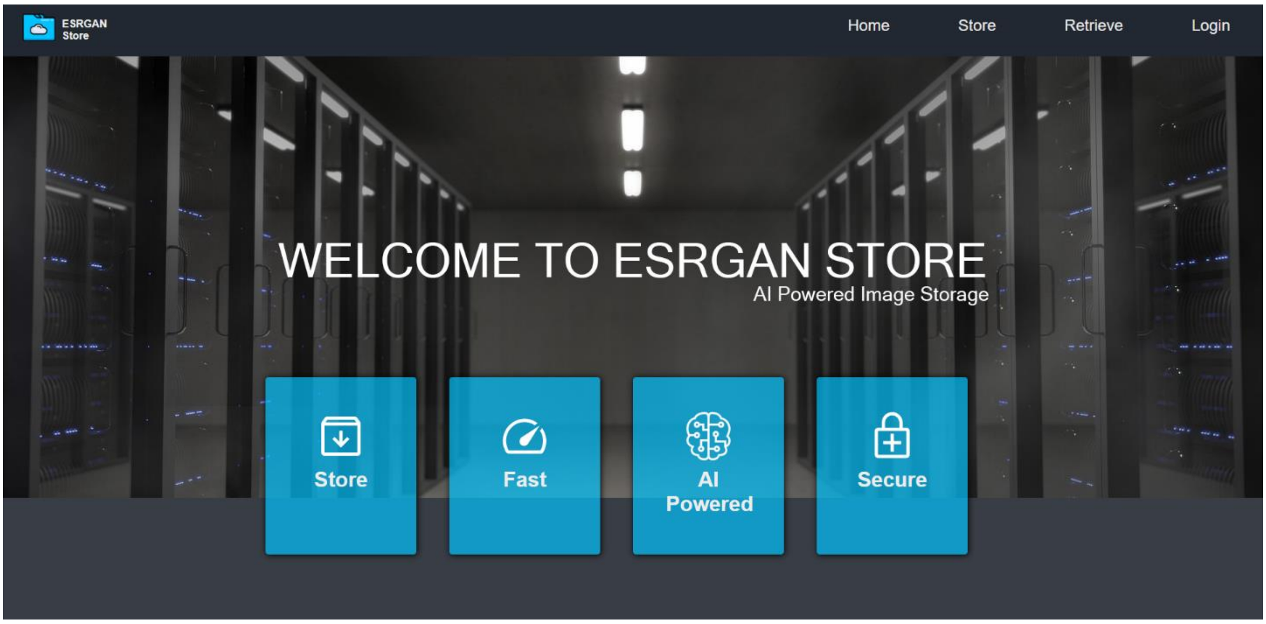

Fig 4.5: Homepage of “ESRGAN STORE” website which is developed using Flask framework to implement the above model.

Here we can create an account or login to an existing account to use more features of this site.

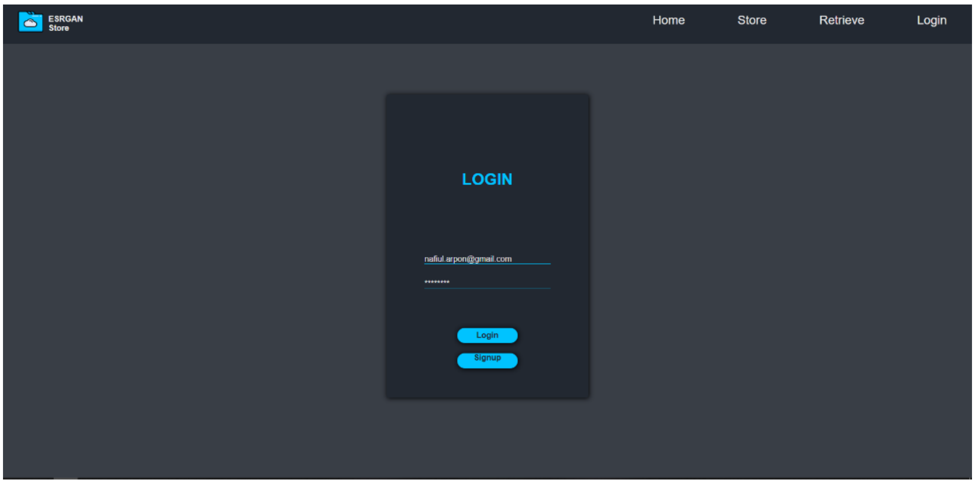

Fig 4.6: Login screen of “ESRGAN STORE” website

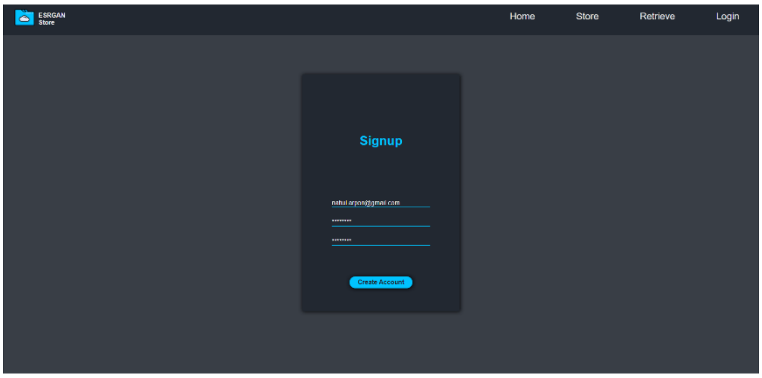

Fig 4.7: Signup page of "ESRGAN STORE" website

After authentication completed, a user can store images in the server by "drag and drop" method or by browsing in the file system. The selected images will then be uploaded and stored against that user's identity.

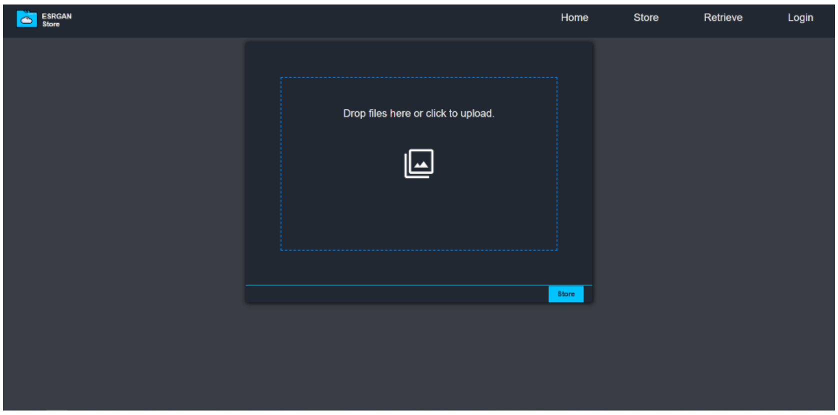

Fig 4.8: Image selection portion of "ESRGAN STORE" website

After that if a user wants, he/she can select the uploaded images from their account and click on the restore button to download the selected image(s).

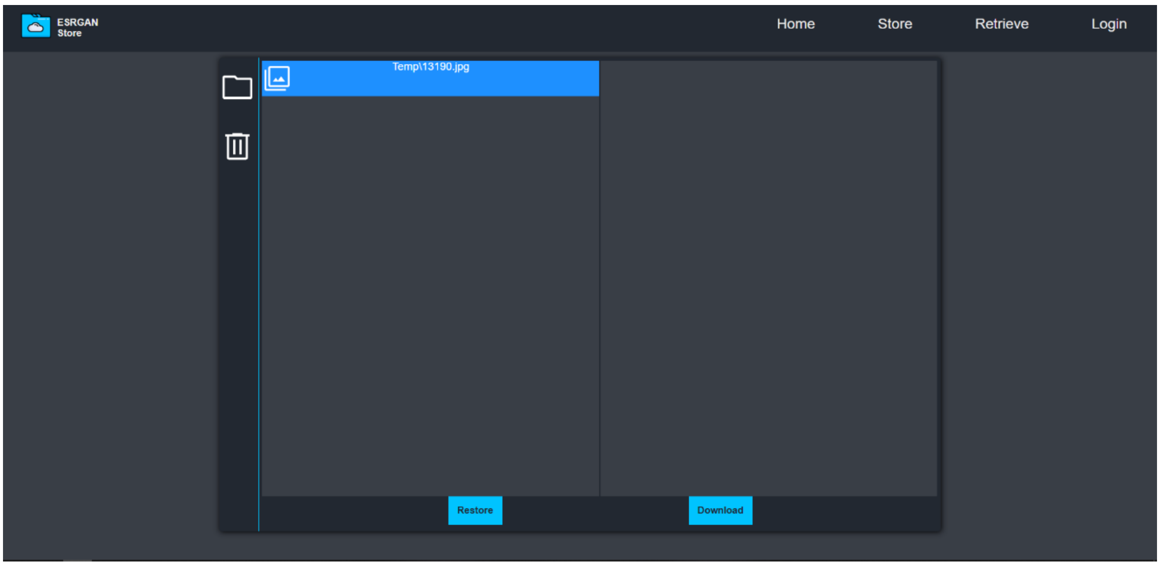

Fig 4.9: Image selection option to download images in "ESRGAN STORE" website

4.7 System Diagram

Following image shows the working procedures of the entire system:

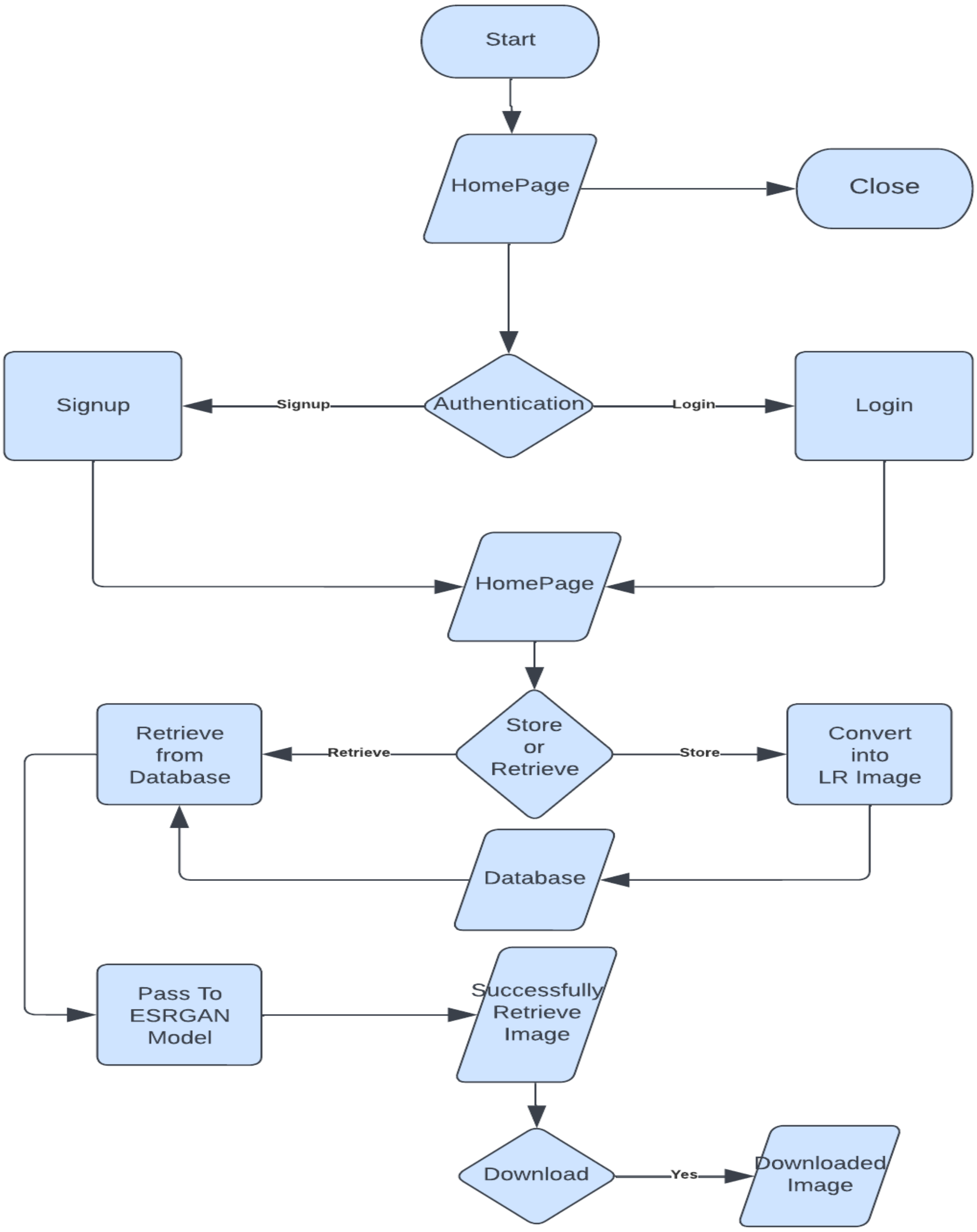

Fig 4.10: Flowchart of the entire system of "ESRGAN STORE" website

4.8 Conclusion

This chapter is terminated with discussion about working diagram and dataset preparation. In next chapter we shall conclude our project report.

CHAPTER V

Conclusion

5.1 Summary

This is the last chapter where it is going to conclude the project study. The summary and the recommendation for the future work of this project will also be displayed.

5.2 Conclusion

In this project, we have used the state-of-the art image upscaling technique ESRGAN to solve the storage problem of large images where traditional method like JPEG, PNG, GIF can't reduce the image size a lot without compromising image quality. Our ESRGAN not only restores saved low-resolution images into their original form but also in some cases it has improved the image quality. ESRGAN requires a lot of computational power for processing. So, it is very hard to enable everyone to use this. But by providing this feature in a website, this technology can be available to everyone. Even though sometimes the output of the project is not up to the mark but with more data provided to the model and more training of the model using those data will significantly improve the output.

5.3 Recommendations for future work

The future work in this field may go in different directions which are compactly aligned to ESR GAN based process and evaluation. The future aspirants may extend this project work with the following task:

- Designing a desktop application or web-based software for the automation of the entire system.
- Identifying the system's progress in performance and reports graphically for each task performed.

REFERENCES

1. "ESRGAN". https://github.com/xinntao/ESRGAN (accessed: Dec. 27, 2022).
2. Xintao Wang, Liangbin Xie, Chao Dong, Ying Shan. "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data" https://arxiv.org/abs/2107.10833 (accessed: Dec. 27, 2022).
3. Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, Xiaoou Tang. "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks" https://arxiv.org/abs/1809.00219 (accessed: Dec. 27, 2022).
4. "Flask web development, one drop at a time". https://flask.palletsprojects.com/en/2.2.x/ (accessed: Dec. 27,2022)
5. "Scikit-image" https://scikit-image.org/ (accessed: Dec. 27,2022)
6. "MongoDB" https://www.mongodb.com/docs/ (accessed: Dec. 27, 2022)
7. "Matplotlib" https://matplotlib.org/stable/index.html (accessed: Dec. 27, 2022)
8. Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, L. Lin. "Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks." In: CVPRW. (2018)
9. T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida. "Spectral normalization for generative adversarial networks". arXiv preprint arXiv:1802.05957 (2018)
10. C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang. "Photo-realistic single image super-resolution using a generative adversarial network". In: CVPR (2017)
11. C. Dong, C.C. Loy, K. He, X. Tang. "Learning a deep convolutional network for image super-resolution". In: ECCV. (2014)
12. C. Dong, C.C. Loy, K. He, X. Tang. "Image super-resolution for image restoration by deep reinforcement learning". In CVPR. (2018)