Project Report

Entitled

Super Resolution of Fluorescein Angiography Images

Submitted to the Department of Electronics Engineering in Partial Fulfilment for the Requirements for the Degree of

Bachelor of Technology
(Electronics and Communication)

: Presented & Submitted By :

Yatin Aditya Tekumalla, Dhwaj Kothari, Himanshu Pal

Roll No. (U18EC013, U18EC071, U18EC036)

B. TECH. IV(EC), 8th Semester

: Guided By :

Dr. Kishor Upla

Assistant Professor, DECE

(Year: 2021_22)

DEPARTMENT OF ELECTRONICS ENGINEERING

SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECHNOLOGY

Surat-395007, Gujarat, INDIA.

Sardar Vallabhbhai National Institute Of Technology

Surat - 395 007, Gujarat, India

DEPARTMENT OF ELECTRONICS ENGINEERING

CERTIFICATE

This is to certify that the Project Report entitled "**Super Resolution of Fluorescein Angiography Images**" is presented & submitted by Yatin Aditya Tekumalla, Dhwaj Kothari, Himanshu Pal , bearing Roll Nos. U18EC013, U18EC071, U18EC036, of B.Tech. IV, 8^{th} Semester in the partial fulfillment of the requirement for the award of B.Tech. Degree in Electronics & Communication Engineering for academic year 2021-22.

They have successfully and satisfactorily completed their **Project Examination** in all respects. We, certify that the work is comprehensive, complete and fit for evaluation.

Dr. K. P. Upla
Assistant Professor & Project Guide

PROJECT EXAMINERS:

Name of Examiners	Signature with Date
1. Dr. J. N. Sarvaiya	_____
2. Dr. K. P. Upla	_____
3. Dr. Suman Deb	_____
4. Dr. Raghvendra Pal	_____

Dr. P. N. Patel
Head, DECE, SVNIT

Seal of The Department
(May 2022)

Acknowledgements

We would like to express our profound gratitude and deep regards to our guide Dr. Kishor Upla for his guidance. We are heartily thankful for his suggestion and the clarity of the concepts of the topic that helped us a lot for this work. We would also like to express our gratitude to Dr. Piyush N. Patel, Head of the Electronics Engineering Department, SVNIT and all the faculties of ECED for their co-operation, guidance and support.

We are very grateful to all our classmates and friends for their reviews, suggestions and support.

Yatin Aditya Tekumalla,
Dhwaj Ashishkumar Kothari,
Himanshu Gyansingh Pal
Sardar Vallabhbhai National Institute of Technology
Surat

May 13, 2022

Abstract

Project Topic: Super Resolution of Fluorescein Angiography Images

Medical imaging is a crucial component of medical diagnosis and treatment process. But due to limitations of imaging technology and varying imaging conditions, many medical images have low spatial resolution. This low resolution of medical images makes it difficult to detect minor anatomical features and abnormalities. Super Resolution (SR) is a class of techniques that are used to enhance the resolution of an image.

In this report we have done an imperative study on super resolution methods using deep learning through GAN based models and have tried to propose our own model for obtaining better results. We have also discussed how these methods can be applied in the field medical imaging to enhance the image's resolution so as to provide a more accurate diagnosis on data-set of **Fluorescein Angiographed Retinal Images**. We have also performed experimental analysis to visualise and studied the outputs of these different approaches and compared them both qualitatively and quantitatively.

Yatin Aditya Tekumalla (U18EC013)
Dhwaj Ashishkumar Kothari (U18EC071)
Himanshu Gyansingh Pal (U18EC036)

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CT	Computed Tomography
MRI	Magnetic Resonance Imaging
MR	Magnetic Resonance
PET	Positron Emission Topography
WCE	Wireless Capsule Endoscopy
GI	Gatsro-Intestinal
3D	3-Dimensional
CADe	Computer-Aided Detection
CADx	Computer-Aided Diagnosis
DL	Deep Learning
ROI	Region Of Interest
CNN	Convolutional Neural Network
GAN	Generative Adveserial Network
DSGAN	Down Sampling Generative Adversarial Network
ESRGAN	Enhanced Super Resolution Generative Adversarial Network
ESRGAN-FS	Enhanced Super Resolution Generative Adversarial Network - Frequency Separation
SISR	Single Image Super Resolution
DR	Diabetic Retinopathy
SRResCGAN	Super-Resolution Residual Convolutional Generative Adversarial Network
PReLU	Parametrized Rectified Linear Unit
RGB	Red Green Blue
GPU	Graphics Processing Unit
SRCNN	Super Resolution Convolutional Neural Network
PSNR	Peak Signal to Noise Ratio
SSIM	Structural Similarity Index
SR	Super Resolution
ReLU	Rectified Linear Activation Unit
LR	Low Resolution
HR	High Resolution
SRGAN	Super Resolution Generative Adversarial Network
LFSR	Lesion Focused Super Resolution
AWGN	Additive White Gaussian Noise
SRResNet	Super Resolution Residual Network
VGG	Visual Geometry Group

Chapter 1
Introduction

Medical imaging is an important tool for giving accurate medical diagnosis and prognosis. X-ray radiography, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound Imaging are some of the commonly used techniques. Medical image processing and analysis are complex and diverse tasks. It includes image reconstruction, enhancement, restoration, classification, detection, segmentation, and registration for medical images. Deep learning (DL) has been widely applied in many medical imaging tasks since its emergence, and it has achieved great success in many medical imaging applications, and the foundation for medical artificial intelligence systems and robotic surgeries. In medical images, generally the region of interest (ROI) is small and requires to be accurately observed and the requirement of super resolution arises. Super resolution of image comes under the domain of image enhancement and refers to the process of upscaling and/or improving the details within an image.

1.1 Medical Imaging

Medical imaging is a crucial component of medical diagnosis and treatment process. Medical imaging exploits physical phenomena, such as light, electromagnetic radiation, radioactivity, nuclear magnetic resonance (MR), and sound to generate visual representations or images of external or internal tissues of the human body or a part of the human body in a noninvasive manner or via an invasive procedure. The most commonly used imaging modalities in clinical medicine are X-ray radiography, computed tomography (CT), MR imaging (MRI), ultrasound, and digital pathology [8].

Imaging data makes up over 90% of all healthcare data, making it one of the most valuable sources of information for clinical analysis and medical treatment. Typically, a radiologist examines the medical images and curates a report explaining their findings. Based on the images and the radiologist's report, the referring physician makes a diagnosis and determines a treatment plan. Medical imaging is frequently given as part of a patient's follow-up to ensure that treatment was successful. Furthermore, medical imaging is becoming an increasingly significant part of invasive procedures, as they are used for surgical planning as well as real-time imaging during the surgical procedure.

Medical pictures are widely used in clinical settings, including non-invasive illness diagnosis, anatomic imaging, and treatment planning. Medical imaging makes use of physical phenomena.Light, electromagnetic radiation, radioactivity, and Magnetic Res-

onance(MR) are examples of phenomena and sound to create visual representations of external or internal tissues in a non-invasive or invasive manner of the human body or a component of the human body.

1.1.1 X-Ray Imaging:

The most popular and readily available diagnostic imaging technology is X-rays (radiographs). Radiation that can travel through the body is known as X-rays. The energy from X-rays is absorbed at varying rates by different regions of the body as they pass through it. After the X-rays have passed through, a detector on the other side of the body picks them up and converts them to an image. Bone and other dense components of the body that X-rays have a hard time passing through appear as distinct white spots on the image. Darker regions indicate software components that X-rays can pass through more easily, such as your heart and lungs. The general block diagram of picture data is shown in Fig. 1.1.

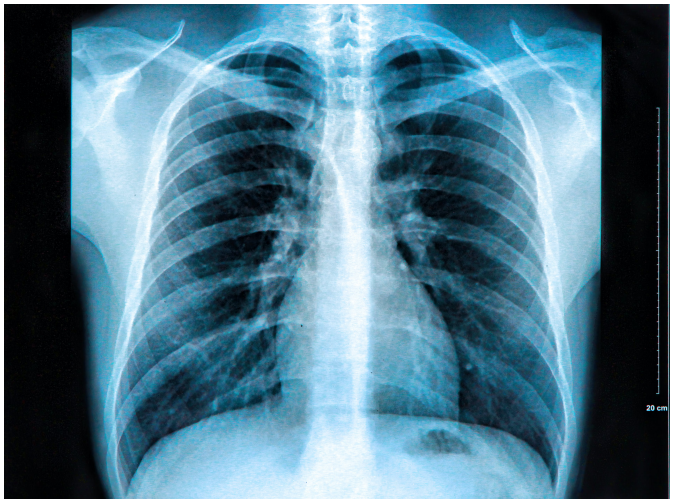


Figure 1.1: X-Ray Imaging [1].

Computed Tomography (CT) :

It's a technique that uses computer-generated X-rays to create tomographic three-dimensional images (virtual slices) of specified portions of a scanned object, allowing the radiologist to view inside without having to cut the object open. One example of the type of image is shown in Fig. 1.2.

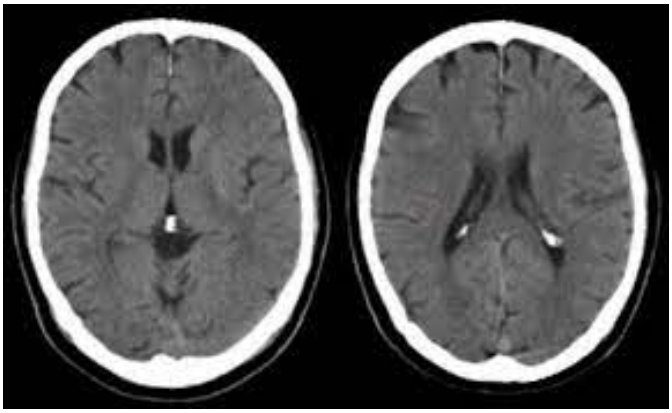


Figure 1.2: CT Imaging [2].

Positron Emission Tomography (PET) :

PET (Positron Emission Tomography) is a nuclear medicine treatment that assesses the metabolic activity of cells in the body. PET can accurately detect tumors in their early stages by detecting the varied absorption capacities of radioisotope-labeled compounds in different areas of the body. PET images looks like as shown in Fig. 1.3.

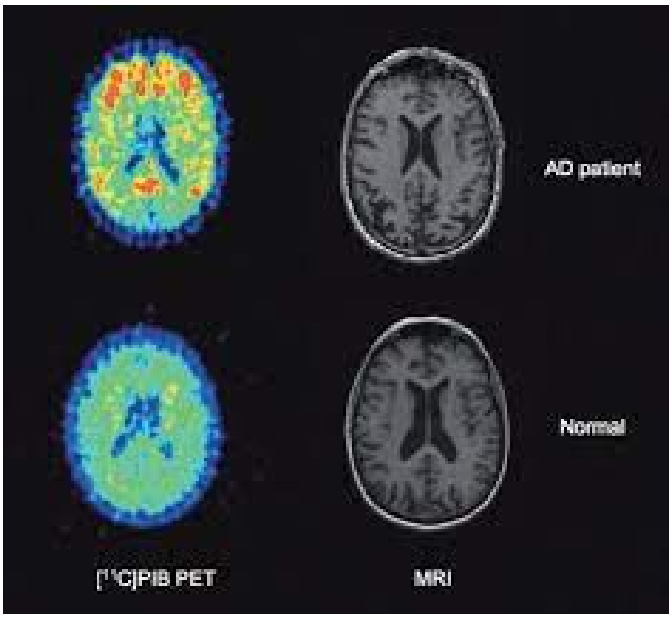


Figure 1.3: PET Imaging [3].

Limitations:

• During CT scanning, radiation harm is unavoidable. Low dose CT is a clinical recommended technology for reducing the radiation injury to patients, but at the expense of low image resolution and noise contamination.

• PET imaging is based on detecting the quantity of photon pairs created by a radioisotope's annihilation reaction. The low resolution problem in PET imaging is also a challenge because of the paucity of photon pairs and noise interference.

1.1.2 Magnetic Resonance Imaging (MRI)

Cryogenic superconducting magnets in the range of 0.5 Tesla (T) to 1.5 Tesla (T) are used in current diagnostic MRI scanners. The majority of clinical trials used a field strength of 1.5 T. 3 T systems, on the other hand, are now widely available and are routinely employed in research settings, where their capabilities are being investigated and maximized as shown in Fig. 1.4.

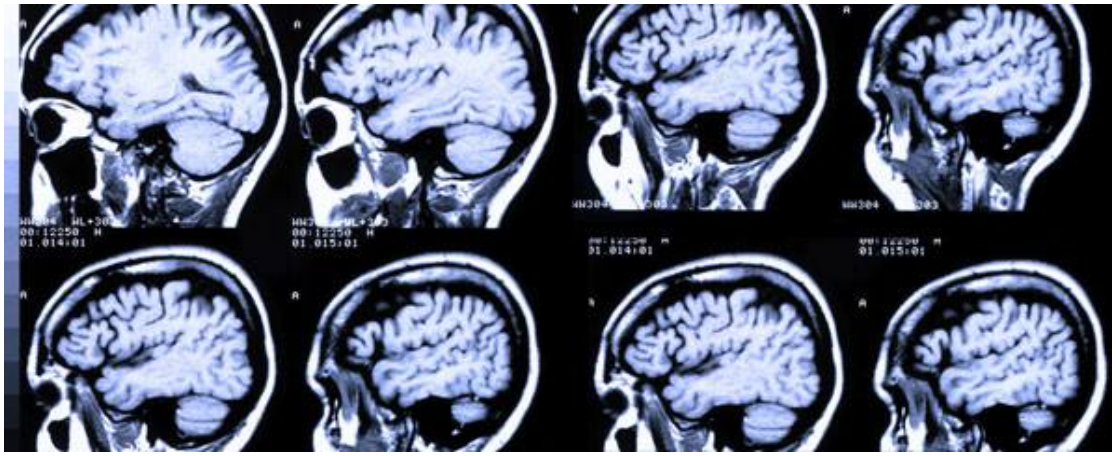


Figure 1.4: MRI Imaging [4].

Limitations:

• Magnetic susceptibility, eddy current artifacts, and magnetic field instability are all disadvantages of MRI procedures.Eddy currents may lead to perturbations in the gradient field, reducing resolution of the subsequent MR Image.

• A key disadvantage of MRI is the extended acquisition time, which leaves picture quality vulnerable to patient movement.Low magnetic field of the MRI scanner also constrains the spatial resolution of the MRI.

1.1.3 Ultrasound

High-frequency sound waves are rebounded off areas of the body during ultrasound, and the echoes are collected as images. Ultrasound employs sound with a frequency that is higher than that of the human ear. The transducer sends and receives echoes at such a fast rate that the reconstructed images appear to be in real time. Because

ultrasound images are collected in real time, they can show internal organ movement as well as blood flowing via blood vessels. Ultrasound imaging does not expose you to ionizing radiation like X-ray imaging does. A transducer (probe) is put directly on the skin or within a bodily orifice during an ultrasonic examination.The ultrasonic waves are delivered from the transducer into the body through a small layer of gel placed to the skin. The ultrasound image is created by the waves reflecting off of the body structures. The information needed to create an image is provided by the power (amplitude) of the sound signal and the time it takes for the wave to travel through the body as shown in Fig. 1.5.

Figure 1.5: Ultrasound Imaging [5].

Limitations:

• For effective imaging, a lower frequency is required. As a result, the ultrasound image resolution is reduced.

• It has low air or bone penetration, and images are difficult to interpret. In comparison to CT and MRI, image resolution is lower.

1.1.4 Endoscopy:

Endoscopy is a common imaging method used for both diagnostic and minimally invasive surgical procedures. WCE stands for Wireless Capsule Endoscopy. cutting-edge imaging equipment that provides direct visualization of the gastrointestinal tract. A typical colon WCE test yields approximately 8 hours of RGB video footage. WCE is a non-invasive treatment for detecting gastrointestinal (GI) tract anomalies in humans. It has an optical dome, an illuminator, an imaging sensor, a battery, and an RF transmitter in a capsule-shaped transmitter with a length of 26 mm and a diameter of 11 mm. The

patient drinks the WCE and then slides it during the examination gradually along the small intestine, photographing the entire gastrointestinal tract. Finally, these images are transferred wirelessly to a data-recording device so that doctors can study the photos later for diagnosis as shown in figure 1.6. While moving through the gastrointestinal tract, an average of 50k to 60k images are taken.

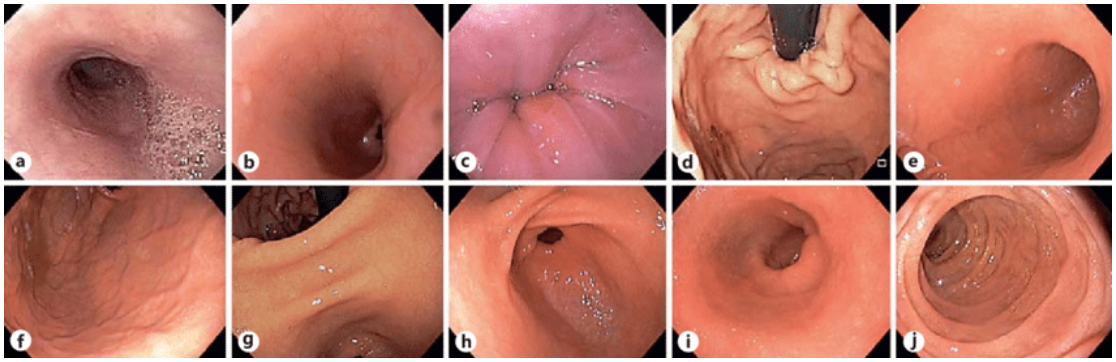


Figure 1.6: Endosocpy Imaging [6].

Limitations:

• Artifacts such as motion blur, bubbles, specular reflections, floating objects, and pixel saturation make visual interpretation and automated analysis of endoscopic recordings difficult.

• Images are captured using capsule endoscopes with poor lighting and limited power. Videos with modest resolutions and frame rates are wirelessly transferred to a recorder outside the human body. This deterioration in image quality can make it more difficult to make a correct diagnosis.

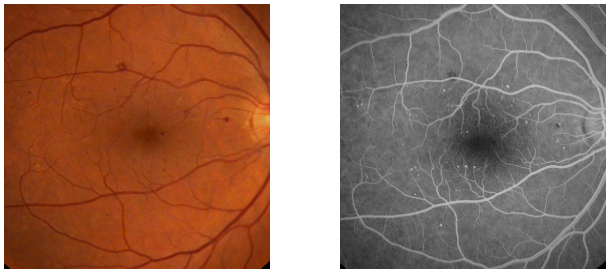
• With the increased use of endoscopy in many clinical applications, recognising artifacts and automating the repair of damaged video frames has become a major medical imaging challenge.

• Increasing the size of the optics and sensor array to improve picture resolution is not always a viable solution because cost and essential space constraints are prohibitive in many endoscopic applications. Better disease/abnormality identification, region segmentation, and 3D reconstruction will all benefit from improved image quality.

1.1.5 Retinal Fundoscopy and Fluorescein Angiography

Retinal fundus photography is a non-invasive imaging procedure that takes only a few seconds to accomplish. Retinal vascular and pigment epithelial-choroidal disorders are diagnosed using Fluorescein Angiography (FA) and Retinal Funduscopy as shown in

Fig. 1.7. A fluorescent dye is injected into the optic vein, which appears in 8-12 seconds depending on the age and circulatory structure of the eye and lasts up to 10 minutes. Optical Coherence Tomography and basic image processing techniques are the only contemporary alternatives to flourecein angiography (FA). In general, these systems are relatively costly.



(a) Retinal Fundoscopy

(b) Fluorescein Angiography

Figure 1.7: Retinal Fundoscopy and Fluorescein Angiography of same eye

Limitations:

• Using Fluorescein Angiography (FA) to diagnose retinal vascular degeneration is a time-consuming method that can have major negative consequences for the patient.

• Angiography necessitates the injection of a dye that can have serious side effects and possibly be fatal.

• There are currently no non-invasive technologies that can produce Fluorescein Angiography images.

• Although the dye is generally regarded safe, allergic responses to it have resulted in mild to severe problems. Nausea, vomiting, anaphylaxis, heart attack, anaphylactic shock, and death are all common adverse effects.

• Without a computationally effective and financially viable technique for generating reliable and repeatable flourecein angiograms, the only option for differential diagnosis is to use retina funduscopy.

Image interpretation by humans, however, is limited due to human subjectivity, the large variations across interpreters, minute variation in imaging hardware and hardware fatigue. Radiologists who review cases have limited time to review an ever-increasing number of images, which leads to missed findings, long turn-around times, and a scarcity of numerical results or quantification. Furthermore, medical images can often be challenging and time-consuming to analyze due to the shortage of radiolo-

gists. This, drastically limits the medical community's ability to advance toward more evidence-based personalized healthcare. In the past decade, with the development of technologies related to the image acquisition process, imaging devices have improved in speed and resolution. Also the advent of deep learning technology has created an opportunity to reduce some of these problems.

1.2 Deep Learning in Medical Images

Deep learning is the leading method for a wide range of computer vision tasks, including medical imaging. It is the state of the art for several medical imaging tasks like the classification of Alzheimer's, breast cancer detection, lung cancer detection, retinal disease detection, etc. Deep Learning has been extensively used in various medical imaging applications such as medical image reconstruction, medical image enhancement, medical image segmentation, medical image registration, computer aided detection and diagnosis [9].

Medical image reconstruction: It is the process of converting signals obtained by a medical imaging instrument, such as a CT or MRI scanner, into a image. Reconstruction of high-quality images from low dosage and quicker acquisitions, has been made possible due to deep learning [8, 10].

Medical image enhancement: It is the process to modify the intensities of an image to make it more acceptable for presentation or further investigation. De-noising, super resolution, MR bias field correction, and image harmonisation are some of the techniques to enhance the image.Image enhancement is a vital step of medical image analysis and image recognition. X-ray and ultrasound imaging are the most preferred medical imaging technologies which are important for diagnosis of disease.

Medical image segmentation: It is the process of assigning labels to pixels such that pixels with the same label can be combined to generate a segmented entity. Clinical measurement, therapy, and surgical planning all benefit from segmentation. Image registration in medicine is the process of aligning the spatial coordinates of one or more images into a shared coordinate system. Registration is extensively employed for image segmentation via label transfer and is widely utilised in population analysis, longitudinal analysis, and multimodal fusion [11].

CADe and CADx (computer-aided detection and diagnosis): CADe tries to locate or locate a bounding box that contains an object of interest (usually a lesion). The goal of CADx [12] is to further identify the localised lesion as benign/malignant or one of the many different forms of lesion. Landmark detection [13], image or view recognition, automatic report generation, are also some of the application of Deep Learning in

medical imaging.

Super Resolution: SR is a class of techniques that are used to enhance the resolution of an image. Image super-resolution (SR) is the process of recovering high-resolution (HR) images from low-resolution (LR) images. Super Resolution (SR) has a wide range of real-world applications, such as medical imaging, satellite imaging, surveillance and security, astronomical imaging, amongst others. In medical imaging, details of minor anatomical markers are crucial accurate diagnosis and proper prognosis.In Radiology, MRI is a commonly used method to produce medical imaging but the limitations of lab equipment and health hazard of being in an MRI radiation environment to obtain good quality scans lead to lower quality scans and also it takes a lot of time to get a high-resolution data. Medical imaging system developers strive to increase resolution since higher resolution is the key to more accurate understanding of the anatomy, it can support early detection of abnormalities and can increase the accuracy in the assessment of size and morphology of organs and pathologies. Hence, super resolution is much desired in medical imaging to get high resolution images. Traditional method for super resolution uses interpolation techniques such as Nearest-neighbor Interpolation, Bilinear Interpolation, and Bicubic Interpolation. But traditional interpolation-based methods often introduce some side effects such as computational complexity, noise amplification, blurring results, etc. Hence, Deep Learning-based super resolution is recently very popular which is used to overcome the shortcomings of interpolation-based methods. Deep learning-based SR models have been intensively researched in recent years and they frequently produced state-of-the-art performance on several SR benchmarks. .

1.3 Motivation

Details of minor anatomical landmarks and diseases are crucial in medical science for correct illness analysis. Small alterations in the microvasculature (arterioles, capillaries, and venules) surrounding a tumour, for example, are a key bio-markers for cancer detection, and the unappreciated soft exudates (fluid that leaks out of blood vessels into nearby tissue) are important pathologies for the identification of retinal conditions. High-resolution (HR) Fundus Fluorescein Angiography images become a vital tool in ophthalmology screenings because the HR images can provide doctors with more indicators, such as micro-aneurysms, hemorrhages, and small veins, to assist their diagnosis decision. However, due to limitations of imaging technology and different imaging conditions, many medical images have low spatial resolution. This low resolution makes it difficult to detect or segment minor anatomical features and abnormalities, as well as to make an early diagnosis of some critical disorders. A Radiologist is a medical profes-

sional who is responsible for analysing medical images. However, as the population's health deteriorates, the strain placed on these specialists grows dramatically. There is also a scarcity of such experts in certain disciplines. Hence many people do not have access to medical treatments in a timely and efficient manner. Despite advancements in collection technology and the performance of optimal reconstruction algorithms over the last decade, obtaining an image at the necessary resolution remains difficult due to imaging surroundings, physical imaging system limits, and factors such as noise and blur. Machine learning, particularly deep learning algorithms, have revolutionised image super resolution reconstruction in recent years. These techniques have greatly aided in obtaining better and more detailed images of the medical results than what we now have, as well as in better understanding the source of disease. Adapting these techniques will also aid in offering easier detection facilities in the medical area.

Hence in this report we aim to study existing DL methods for image super-resolution and contrast them with older, traditional methods and propose some better approach.

1.4 Objectives

Super resolution plays a very crucial part in many aspects by providing highly resolved image for the low resolution image which is way comparable to the original truth image. This particular feature of this technique can have a major impact in the sector of medical science and can be very helpful in recognising the disease or any symptom regarding the same by providing fine detailing of the images. As the current techniques that we use does not provide a very high resolution image, but using this technique that problem can be solved to some extent.

Despite some successes through supervised learning based methods, particularly in terms of high PSNR or other quantitative metrics, a major drawback is that the SR images constructed are often not realistic in fine details and may leave blurred impressions perceptually. These methods do not tackle the problem of artifacts created in the SR process from the noise in input images i.e. inherent sensor noise, stochastic noise, compression artifacts, possible mismatch between image degradation process and camera device. The most major contributor of noise in medical images is due to the motion of patient, which is non-uniform. Hence, they do not perform well for real time applications.

Through this report, we aim:

- To study and compare different state-of-the-art methods for super resolution of medical images.

- To come up with a GAN based approach for real world super resolution of retinal images.
- To compare its qualitative and quantitative results with other models.

1.5 Report Outline

In the beginning of the report at Chapter 1, we have introduced about the topic of super resolution in medical images along with the discussion about what is medical imaging and what are the impacts of deep learning in medical imaging. Further we have described about the various factors that have motivated us to research on this topic and finally the objectives of this report.

In Chapter 2, we have written about the survey that we did on different research papers and have given a brief about different methods used in each one of them. There we compared various topics like the traditional methods and the deep learning methods used. We have also discussed about the research gap that we observed while going through the papers.

In Chapter 3, we have discussed about what our model architecture is and all other functioning details about the architecture.

Chapter 4, presents the experimental results we obtained from running the proposed deep learning model and have compared the results with some other architectures that have been already proposed in obtaining better results for data-sets for medical images.

In Chapter 5, at the end, we have concluded the report and have discussed about how this model have helped us in satisfying our objective.

Chapter 2

Literature Survey

In the recent years, there has been a lot of research for effective super resolution of images. Super-resolution methods can be broadly divided into two main categories: traditional and deep learning methods. Classical algorithms have been around for decades now, but are out-performed by their deep learning based counterparts. Therefore, most recent algorithms rely on data driven deep learning models to reconstruct the required details for accurate super-resolution. Some prominent studies of these methods have been mentioned in this chapter.

2.1 Traditional Super Resolution methods

Traditional methods for super resolution include interpolation-based upscaling methods with Nearest neighbour interpolation, bi-linear interpolation, and cubic spline interpolation are a few well-known examples. Yang *et al.* [14] have presented sparse representation for image super resolution. Various regularization methods have been proposed to further stabilize the output of these methods [15–17].

Yang *et al.* [18] give a detailed overview of the different approaches to single image super-resolution (SISR). Common SISR methods can be divided into two categories: interpolation-based methods, reconstruction-based methods. Aktar *et al.* [19] have used bicubic interpolation [20] to get simplicity and speedy high resolution image. Dai *et al.* [21] represented the local image patches using the background or foreground descriptors and reconstructed the sharp discontinuity between the two. Sun *et al.* [22] explored the gradient profile prior for local image structures and applied it to super-resolution.

2.2 Super Resolution using Deep Learning

According to different purposes, these methods can be mainly divided into two main categories: Convolutional Neural Network (CNN) based approaches which try to achieve high peak signal-to-noise ratio (PSNR) and or structural similarity (SSIM) and Generative Adversarial Network (GAN) based approaches which aims to get high perceptual quality. GAN based models researches are gradually getting more attention in the medical imaging research communities in recent years.

2.2.1 Convolutional Neural Network based Approaches

Convolutional neural networks (CNN) have recently exploded in popularity, thanks in part to their effectiveness in picture classification. Methods with higher PSNR or SSIM were mainly achieved by network structure innovations such as deeper networks, various feature connections and feature flowing strategies inside networks.

CNN based Super Resolution for Non-Medical Images

ChaoDong *et al.* applied a convolutional neural network to image super-resolution reconstruction, which is called a super-resolution convolutional neural network (SRCNN) [23]. The well network implements end-to-end image reconstruction. The hidden layer of the convolutional neural network (CNN) completes all feature extraction and aggregation, but SRCNN does not consider any self-similarities. SRCNN was the first implementation on SR problem using CNN. Since then, more people have proposed image super-resolution algorithms based on CNN. For example, another algorithm a Accelerating the Super-Resolution Convolutional Neural Network (FSRCNN) proposed by ChaoDong *et al.* [24]. Compared with SRCNN, this algorithm does not need to use interpolation to enlarge the original LR images, but process these directly. Its advantage is to shorten the training time by reducing the picture size.

The Accurate Image Super-Resolution Using Very Deep Convolutional Networks (VDSR) [25] model was proposed by Jiwon Kim *et al*. Its input is divided into two parts. One part is to take the entire original images as input, and the other part is to train the residuals and get input. We add the two parts together to get HR images, which greatly speeds up the training and converges better. Afterward, Jiwon Kim *et al* proposed Deeply-Recursive Convolutional Network for Image Super-Resolution (DRCN) [25]. This network consists of three parts. However, at the layer of non-linear mapping, it uses a recursive network. The data can loop through the layer multiple times. Unrolling this loop is equivalent to multiple convolutional layers in series using the same set of parameters. Ying Tai *et al.* [26] proposed the Image Super-Resolution via Deep Recursive Residual Network (DRRN) model.

CNN based Super Resolution for Medical Images

Junyoung Park et al. [27] have developed a convolutional neural network (CNN) based model for computed tomography (CT) image super-resolution. The network learns an end-to-end mapping between low and high resolution images using the modified U-Net. *Shi et al.* [28] have proposed residual learning-based SR algorithm for MRI, which combines both multi-scale GRL and shallow network block-based local residual learning (LRL). It minimizes the partial loss of image details that usually happens in a very deep network due to the degradation problem.

Jiang et. al. were the first to propose super resolution for fundus fluorescein angiography (FFA) imaging in 2018 [29]. They used transfer learning on SRCNN architecture. Based on a clinical FFA dataset collected from Second Affiliated Hospital to Xuzhou Medical University, each SR method is implemented and evaluated for the pipeline to improve the resolution of FFA images.

With the minimisation of mean square error (MSE) loss, they have got prominent results on PSNR and SSIM metrics but SR images still contradict with human observation because human eyes are good at catching high frequency components rather than telling pixel-level difference. On the other hand, GAN based methods considers perceptual loss. These methods performed well on perceptual index metrics but compromise on PSNR and SSIM metrics.

2.2.2 GAN based Approaches

Super-resolution (SR) for image enhancement using GAN can be broadly classified into two methods, one requires multiple low resolution (LR) images from different views of the same object to be reconstructed to the high resolution (HR) output, and the other one relies on the learning from a large amount of training datasets, i.e., LR-HR pairs. In real time scenario, labelled and homogeneous image datasets for Super resolution in medical imaging is still not sufficiently available hence single image super resolution are much popular.

GAN based Super Resolution for Non- Medical Images

Christian Ledig *et al.* [30] first presented SRGAN based on GAN architecture. The SRGAN approach has shown to produce quick and accurate SR outputs. As, SRGAN has been developed for natural images and has not been completely optimized for SR in medical imaging.

GAN based Super Resolution for Medical Images

Yuchong *et al.* [31] presented a GAN-based SR framework for common medical images (MedSRGAN), which included Residual Whole Map Attention Network (RWMAN) as generator, discriminator for image pairs. This model focus on improving the resolution of CT and MRI images, and has shown remarkable improvement in results on perceptual index metrics.

Various models for specific medical imaging for specific applications have also been developed. For example, Jin *et al.* [32], have presented a lesion focused SR (LFSR) method that leverage the merits of GAN based models to generate perceptually more realistic SR results and also avoid introducing non-existing features into the lesion area

after SR. By performing simulation based studies on the Multimodal Brain Tumor Segmentation Challenge (BraTS) datasets, Jin *et al.* demonstrated the efficacy of SR model in application of spatial resolution enhancement of brain tumor MRI images to potentially maintain crucial diagnostic information for following clinical tasks: lesion detection and abnormality enhancement. LFSR includes a lesion detection neural network LD before GAN. LD aims to detect the region of interest (ROI, e.g. brain tumors) and the pass that ROI image for super resolution.

Jin *et al.* [32] have proposed a max pooling residual block and an input-scale free residual neural network LD. Compared to the residual blocks and skip connection have been widely used, a max pooling layers is added after two residual blocks, which include two skip connections between four convolution and batch normalization layers. This can help accelerate the training process, and reduce the memory cost of the ROI detection task.

2.3 Real-World Super Resolution Approaches

Despite their success, the algorithms outlined above are trained with HR/LR image pairings on bicubic down-sampling, which limits their performance in real-world scenarios. Lugmayr et al. [33] provided the real-world challenge series [34], which described the impacts of bicubic downsampling and independent degradation learning for super resolution. Fritsche et al. [35] later suggested the DSGAN to learn degradation by training the network in an unsupervised manner, as well as modifying the ESRGAN as ESRGAN-FS [36] to improve its performance in real-world conditions.

2.4 Research gap

In the field of super resolution, research has led to multiple methods to perform resolution in order to enhance the low resolution images to higher resolution. But there have been some gaps that we could observe while going through these evolution in super resolution methods and techniques.

- Today, super resolution can be employed in many sectors and can be way beneficial than the current technologies employed. Among these the most crucial sector is the healthcare sector. Because of the lack of research advancement in using super resolution in this particular sector, many problems which could have been solved using this technique is still left unattended. Many latest methods available for natural images have still not been translated for their use in medical imaging modalities.

- Despite some successes among these supervised learning based methods, particularly in terms of high PSNR or other quantitative metrics, a major drawback is that the SR images constructed are often not realistic in fine details and may leave blurred impressions perceptually. The deep learning models trained with synthetic dataset generalize poorly on the real-world or natural data where the degradation characteristics cannot be fully modelled. These methods do not tackle the problem of artifacts created in the SR process from the noise in input images i.e. inherent sensor noise, stochastic noise, compression artifacts, possible mismatch between image degradation process and camera device. The most major contributor of noise in medical images is due to the motion of patient, which is non-uniform. Hence, they do not perform well for real time applications.
- Fluorescein Angiography (FA) is used to diagnose retinal vascular degeneration. Angiography necessitates the injection of a dye that can have serious side effects in higher concentration. There are currently no non-invasive technologies that can produce high resolution Fluorescein Angiography images. Very less research work has been done for super resolution of Fluorescein Angiography images so that doctors can identify retinal vascular structure easily.

Chapter 3

Our Methodology

Most current deep learning based single image super resolution (SISR) methods focus on designing deeper/wider models to learn the non-linear mapping between low resolution (LR) inputs and the high-resolution (HR) outputs from a large number of paired (LR/HR) training data. They usually take as assumption that the LR image is a bicubic down-sampled version of the HR image. However, such degradation process is not available in real-world settings i.e. inherent sensor noise, stochastic noise, compression artifacts, possible mismatch between image degradation process and camera device. The most major contributor of noise in medical images is due to the motion of patient, which is non-uniform. It reduces significantly the performance of current SISR methods due to real-world image corruptions. Here, we have proposed a GAN based real world super resolution model inspired by SRResCGAN [7] and ESRGAN-FS [35] for medical images.

3.1 Network Architecture

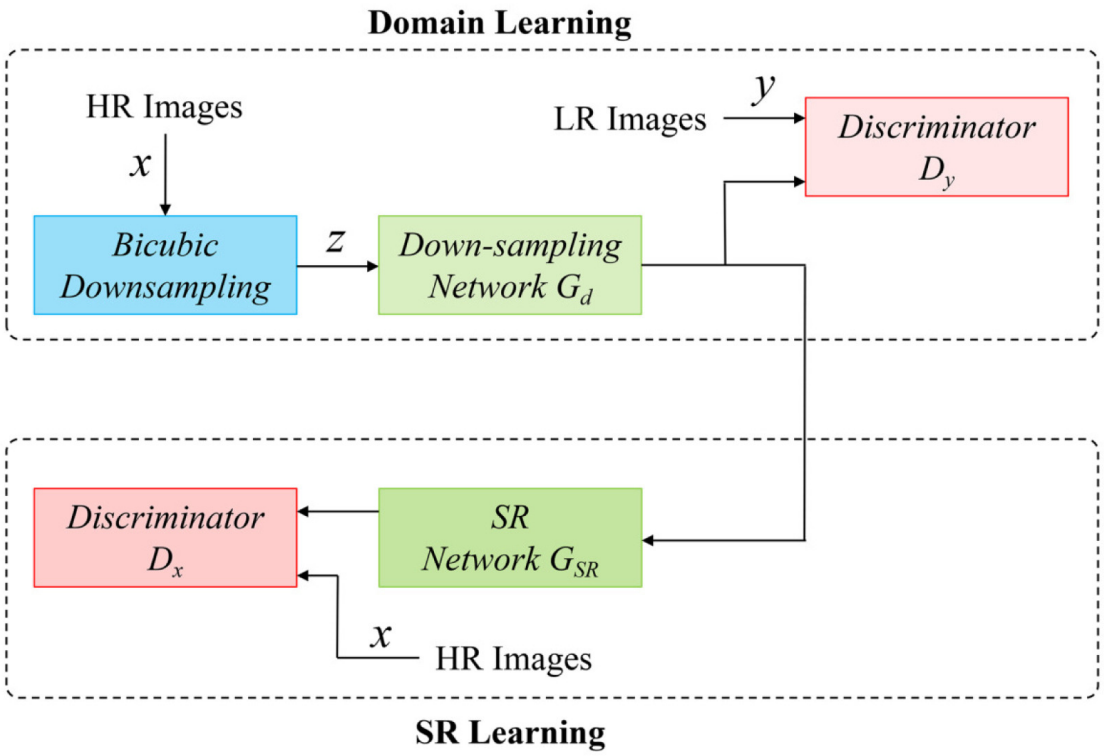
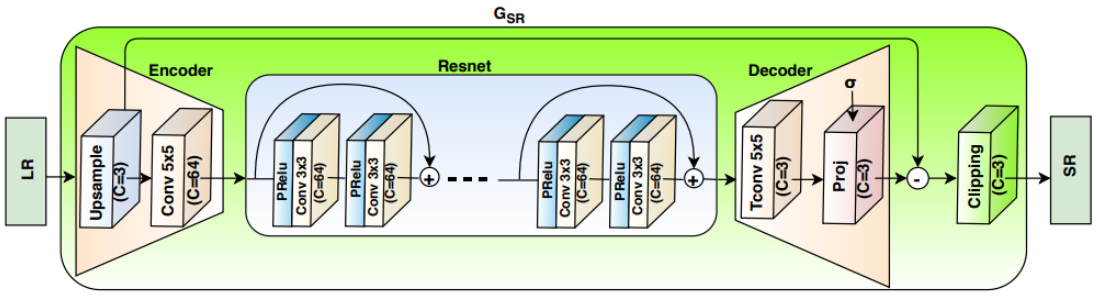


Figure 3.1: The architecture of our Proposed Approach based on SRResCGAN [7].

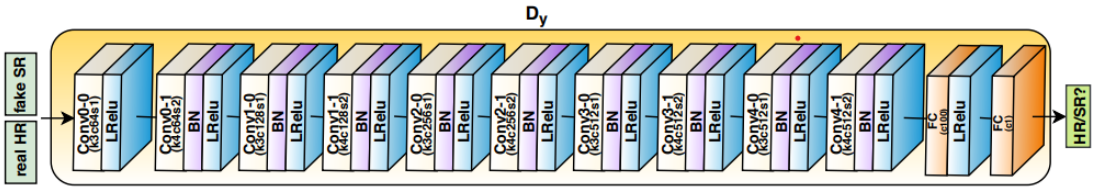
The general architecture of SRResCGAN [7] is as follows:

The down sampling generator network (G_d) consists of 8 Resnet blocks (two Conv layers and PReLU activations in between) that are sandwiched between two Conv layers. All Conv layers have 3 × 3 kernel support with 64 feature maps. Finally, sigmoid nonlinearity is applied on the output of the G_d network. While, the discriminator network (D_x) consists of a three-layer convolutional network that operates on a patch level. All Conv layers have 5 × 5 kernel support with feature maps from 64 to 256 and also applied Batch Norm and Leaky ReLU (LReLU) activations after each Conv layer except the last Conv layer that maps 256 to 1 features.

Generator (G_{SR}): In the G_{SR} network (Fig 3.2-(a)), both Encoder (Conv, refers to L_k filters) and Decoder (TConv, refers to L_k Tfilters) layers have 64 feature maps of 5 × 5 kernel size with C × H × W tensors, where C is the number of channels of the input image. Inside the Encoder, LR image (Y) is upsampled by the Bilinear kernel with Upsample layer, where the choice of the upsampling kernel is arbitrary.



(a) Generator



(b) Discriminator.

Figure 3.2: The architectures of Generator and Discriminator networks. The k, c, s denotes kernel size, number of filters, and stride size.

Resnet consists of 16 residual blocks with two Pre-activation Conv layers, each of 64 feature maps with kernels support 3 × 3. The preactivations are the parameterized rectified linear unit (PReLU) with 64 out feature channels support. The trainable projection (proj) layer inside Decoder computes the proximal map with the estimated noise standard deviation and handles the data fidelity and prior terms. Moreover, the proj layer parameter is fine-tuned during the training via a back-propagation. The noise realiza-

tion is estimated in the intermediate Resnet that is sandwiched between Encoder and Decoder. The estimated residual image after Decoder is subtracted from the LR input image. Finally, the clipping layer incorporates our prior knowledge about the valid range of image intensities and enforces the pixel values of the reconstructed image to lie in the range [0, 255]. Reflection padding is also used before all convolution layers to ensure slowly-varying changes at the boundaries of the input images.

Discriminator (D_y): The Fig. 3.2-(b) shows the architecture of discriminator network that is trained to discriminate real HR images from generated fake SR images. The raw discriminator network contains 10 convolution layers with kernels support 3 × 3 and 4 × 4 of increasing feature maps from 64 to 512 followed by Batch Norm (BN) and leaky ReLU as do in SRGAN [30].

3.1.1 Proposed Approach

With aim to improve upon this model, we did ablation study with various combination of losses and changes in network architecture based on other papers. We changed the architecture of degradation generator G_d by reducing the 8 residual blocks to 4. We did so in hope to reduce the training time.

To compensate the decreased number of residual block, we switched Leaky Relu with parameterized rectified linear unit (PReLU) in the degradation generator G_d. Because Leaky ReLUs allow a small, non-zero gradient when the unit is not active whereas Parametric ReLUs take this idea further by making the coefficient of leakage into a parameter that is learned along with the other neural network parameters.

We changed the number of residual blocks to 12 for super resolution generator network G_{SR} with residual connections after every 2 blocks, to reduce vanishing gradient issue for a very deep network. This allows us to obtain better results due to more number of learnable features in the network.

Also the loss function of the super resolution generator G_{SR} network is given as:

$$L_{GSR} = L_{per} + L_{GAN} + 10 * L_1 + L_{tv} \tag{3.1}$$

Here L_{per} refers to perceptual loss, L_{GAN} is texture loss, L_1 is content loss, and L_1 is total variation loss. We selected the learning rate α as 0.0001 and halved at [5k, 10k, 20k, 30k] iterations. We ran this model for 200 epochs on NVIDIA GTX 1070, a 8GB GPU for 34 hours.

3.2 Network Losses

To learn the super-resolution for the target domain, we train the proposed (GSR) network in a GAN framework [7] with the following loss functions:

$$L_{GSR} = L_{per} + L_{GAN} + 10 * L_1 + L_{tv} \tag{3.2}$$

To learn the domain distribution corruptions from the source domain (x), we train the network G_d in Down-samapling GAN [35] with the following loss function:

$$L_{Gd} = L_1 + 0.01 * L_{per} + 0.005 * L_{GAN} \tag{3.3}$$

where, these loss functions are defined as follows: Perceptual loss (L_{per}): It focuses on the perceptual quality of the output image and is defined as:

$$L_{per} = \frac{1}{N}\sum_{i=1}^{N} L_{VGG} = \frac{1}{N}\sum_{i=1}^{N}(||\phi(G_{SR}(\hat{x}_i)) - \phi(\hat{y}_i)||_1) \tag{3.4}$$

where, ϕ is the feature extracted from the pretrained VGG19 network at the same depth as our network.

Texture loss (L_{GAN}): It focuses on the high frequencies of the output image and is defined as:

$$L_{GAN} = \textbf{\textit{LRaGAN}} = -Ey[log(1-D_y(y, G_{SR}(x)))] - \hat{E}y[log(D_y(G_{SR}(\hat{x}), y))] \tag{3.5}$$

where, Ey and Eyˆ represent the operations of taking average for all real (y) and fake (yˆ) data in the mini-batches respectively. We employed the relativistic discriminator used in the ESRGAN [35] that provides the relative GAN score of real HR and fake SR image patches and is defined as:

$$\textit{Dy(y,)(C)} = \sigma(C(y) - E[C(\hat{y})]) \tag{3.6}$$

where, C is the raw discriminator output and is the sigmoid function.

Content loss (L1): It is defined as:

$$L_1 = \frac{1}{N}\sum_{i=1}^{N}(||(G_{SR}(\hat{x}_i) - (y_i)||_1) \tag{3.7}$$

where, N is represents the size of mini-batch.

TV (total-variation) loss (L_{tv}): It focuses to minimize the gradient discrepancy and produce sharpness in the output image and is defined as:

$$L_{tv} = \frac{1}{N}\sum_{i=1}^{N}((||\nabla_h G_{SR}(\hat{x_i}) - \nabla_h(y_i)||_1 + ||\nabla_v G_{SR}(\hat{x_i}) - \nabla_v(y_i)||_1) \tag{3.8}$$

where, ∇h and ∇v denote the horizontal and vertical gradients of the images.

3.3 Evaluation Metrics

Metrics are used to track and measure a model's performance (during training and testing). We evaluate the trained model under the Peak Signal to-Noise Ratio (PSNR), Structural Similarity (SSIM) metrics.

3.3.1 PSNR

The most frequent technique for determining the quality of outcomes is the Peak Signal to Noise Ratio. Peak Signal to Noise Ratio is a metric that is widely used to assess the quality of lossy compression codec reconstruction (PSNR). In Super Resolution research, this metre has become the de facto norm. It determines how far the deformed (potentially lower quality) image differs from the original high-quality image. The formula below can be used to determine it straight from the MSE, where L is the greatest pixel value that can be calculated (255 for an 8-bit image).

$$\textit{PSNR} = 10.\log_{10}(\frac{L^2}{\frac{1}{N}\sum_{i=1}^{N}(I(i) - \hat{I}(i))^2}) \tag{3.9}$$

Where m and n are the dimensions of the image I and its noisy approximation $\hat{I}$.

3.3.2 SSIM

The structural similarity index measure (SSIM) is a technique for estimating the perceived quality of digital television and film visuals, as well as other digital images and videos. The SSIM algorithm is used to determine how similar two photos are. The SSIM index is a full reference metric, which means that image quality is measured or predicted using an uncompressed or distortion-free image as a baseline. SSIM is a perception-based model that analyzes image deterioration as a perceived change in structural information, as well as crucial perceptual phenomena such as brightness and

contrast masking. Using the formula below, this measure is used to compare the perceptual quality of two photographs, given the mean (μ), standard deviation (σ), and correlation (c) of both images.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{3.10}$$

Where x is the input image and y is output image.

μ_x the average of x; μ_y the average of y; σ_x^2 the variance of x; σ_y^2 the variance of y; σ_{xy} the covariance of x and y; $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ two variables to stabilize the division with weak denominator; L the dynamic range of the pixel-values (typically this is $2^{\#bits\ per\ pixel} - 1$); $k_1 = 0.01$ and $k_2 = 0.03$ by default.

3.4 Dataset

Diabetic Retinopathy Detection Dataset contains a large set of high-resolution retina images taken under a variety of imaging conditions. A left and right field is provided for every subject. Images are labeled with a subject id as well as either left or right [37].

(a) Retinal Fundus Image

(b) Florescein Angiography Image

Figure 3.3: Retinal Fundas image and the Florescein Angiography image

Fig. 3.3 shows a patient's retina scan from the dataset. A clinician has rated the presence of diabetic retinopathy in each image on a scale of 0 to 4, according to the following scale: No DR, Mild, Moderate, Severe, Proliferate DR.

Images contain artifacts, be out of focus, underexposed, or overexposed. For training, we use the fundus and angiography data-set provided by Hajeb et al. [38]. The data-set consists of 30 pairs of diabetic retinopathy and 29 pairs normal of angiography and fundus images from 59 patients. The images are either perfectly aligned or nearly aligned. The resolution for fundus and angiograms are as follows 576×720. Fundus

photographs are in RGB format, whereas angiograms are in gray-scale format. Due to shortage of data, we take 50 random crops of size 128 x 128 from each images for training our model out of angiograms and used them. We downsampled these images to create LR - HR pair. We also augmented our data by using Fundus2Angio [39] to convert fundus images to florescein angiography images and then downsampled them to create LR. We also augmented our data by adding blurring effect, rotation, horizontal and vertical flips to LRs.

3.5 Hyperparameter Tunings

For optimizer, we used Adam with learning rate = 0.0001 and halved at [5k, 10k, 20k, 30k] iterations, 1 = 0.5 and 2 = 0.999. We trained with mini-batches with batch size, b = 4 for 200 epochs.We trained our model on NVIDIA GTX 1070 for 36 hours. Our input resolution was 128x128 pixels and our output resolution was 512x512 pixels.

Chapter 4

Experimental Results

In this chapter we have shown qualitative and quantitative study of our proposed model. We will also compare it with the results of standard benchmark SRGAN and state-of-the-art ESRGAN.

4.1 Qualitative analysis

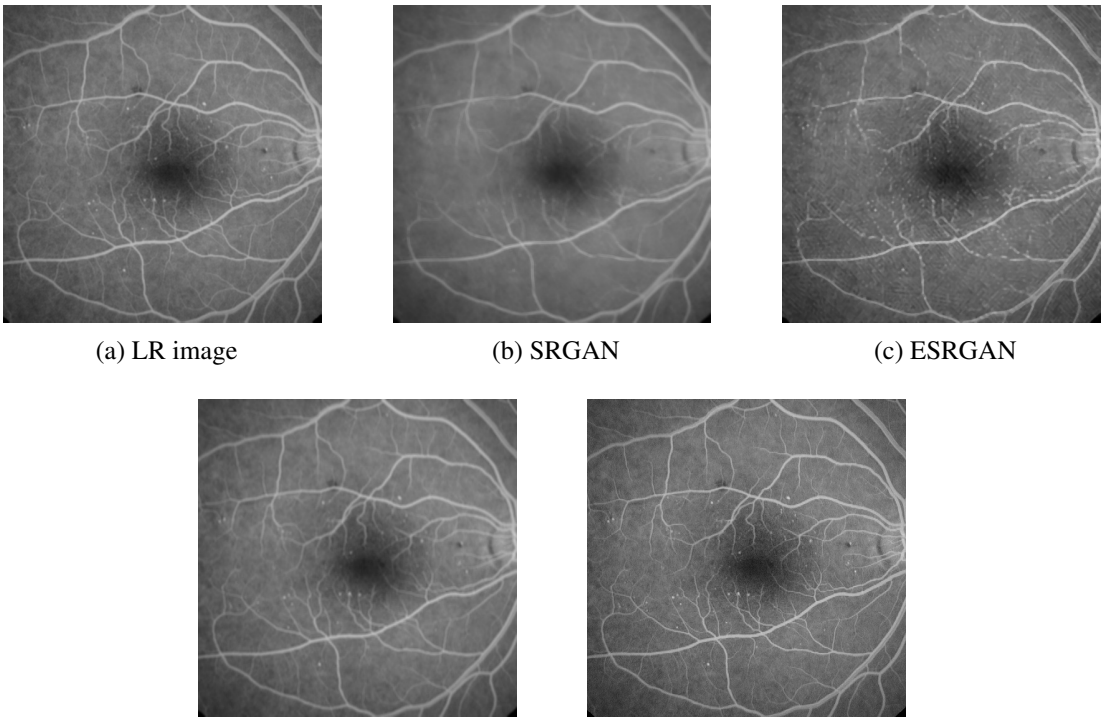
We generated Fluorescein Angiography Images from Retinal Fundus images by passing them through Fundus2Angio [39](refer Fig. 4.1). These images were then down sampled and added to the dataset to augment our data to train the model:

(a) Input Retinal Fundus Image

(b) Generated Florescein Angiography (LR) image

Figure 4.1: Input Retinal Fundas image and the generated Florescein Angiography image

We fed the LR image through SRGAN, ESRGAN, and our proposed approach separately. The results for x4 upscaling are as shown in the Fig. 4.2. Here we trained SRGAN for 200 epochs on the same dataset with NVIDIA GTX 1070. Also we trained ESRGAN on the same dataset with NVIDIA GTX 1080. The total epochs were 200 epochs. The training time was around a day. The learning rate for SRGAN was 10^{-4}, the padding was 4, iterations 10, batches 72, batch size 15 and the learning rate decay 0.5.



(a) LR image (b) SRGAN (c) ESRGAN

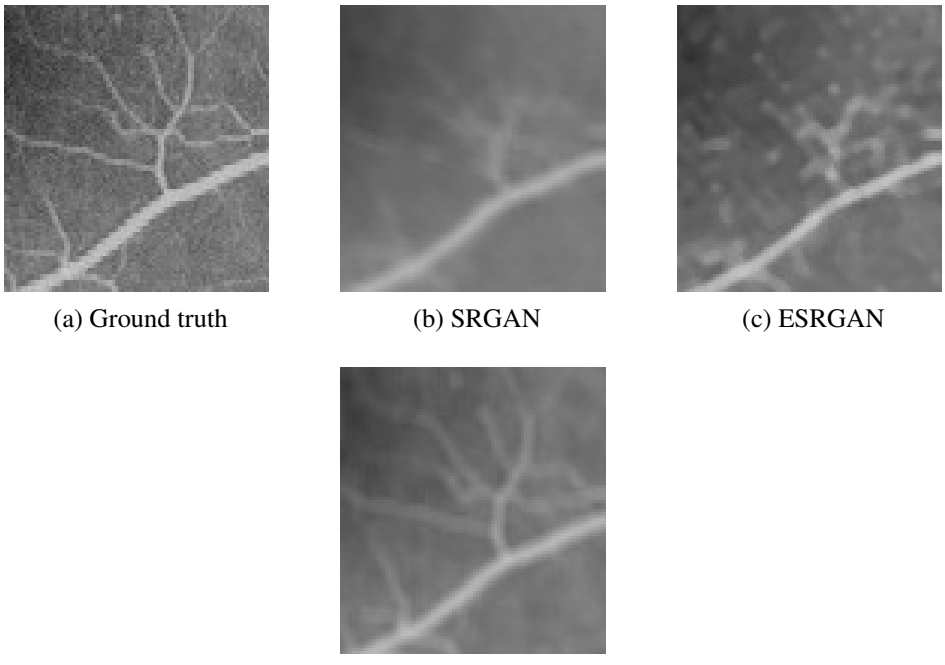
(d) Proposed approach (e) Ground truth

Figure 4.2: Ground Truth compared with SRGAN, ESRGAN and Proposed approach

Here in Fig. 4.3, we observe a specific ROI from a Fluorescein Angiography Image and compare the ROI of super resolved images using SRGAN (Fig. 4.4 -(b)), ESRGAN (Fig. 4.4 -(c)), and our proposed approach (Fig. 4.4 -(d)).

Figure 4.3: ROI on GT

The results of x4 upscaling for SRGAN, ESRGAN and Proposed approach (refer Fig. 4.4):

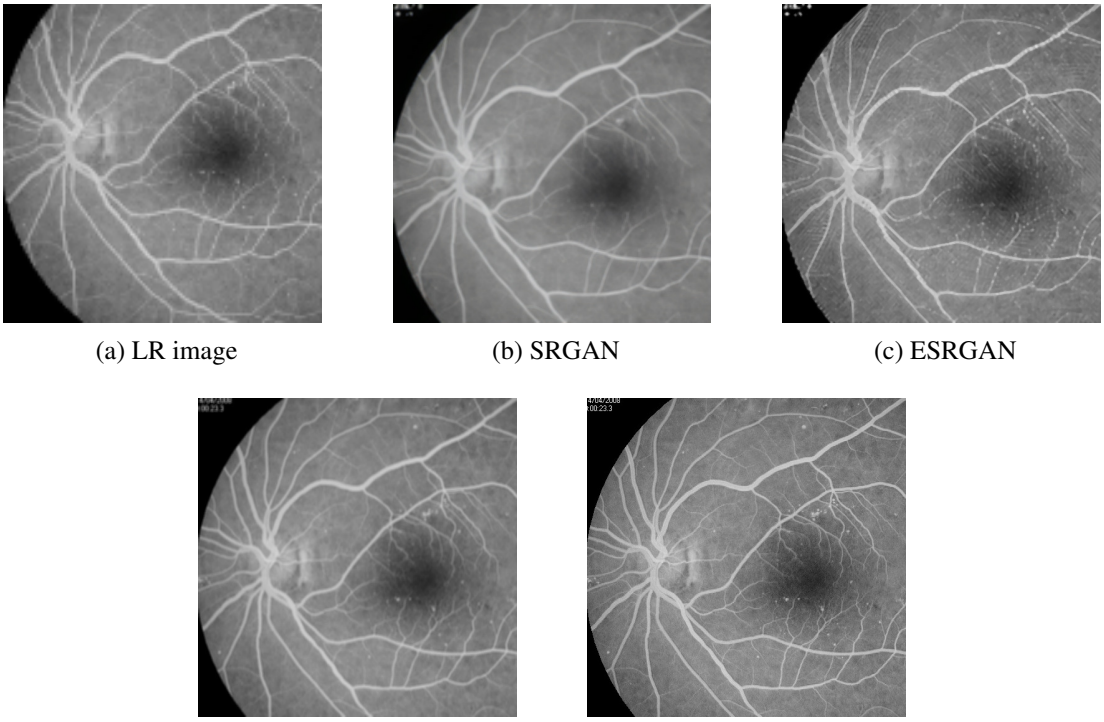


(a) Ground truth (b) SRGAN (c) ESRGAN

(d) Proposed approach

Figure 4.4: Ground Truth compared with SRGAN, ESRGAN, Proposed approach

Another example for another eye image (refer Fig. 4.5):



(a) LR image (b) SRGAN (c) ESRGAN

(d) Proposed approach (e) Ground truth

Figure 4.5: Ground Truth compared with SRGAN Proposed approach

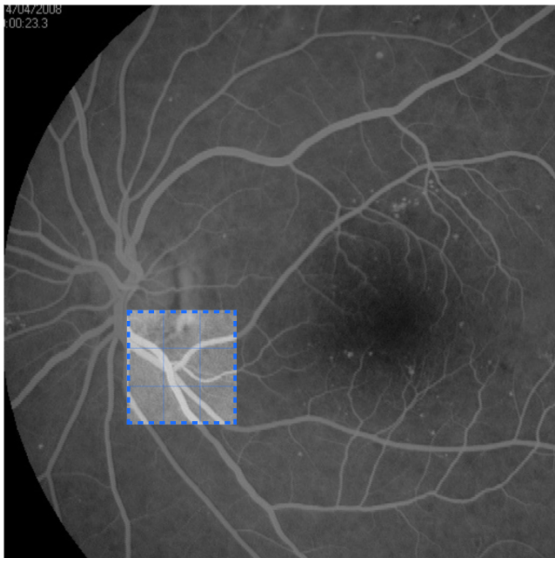
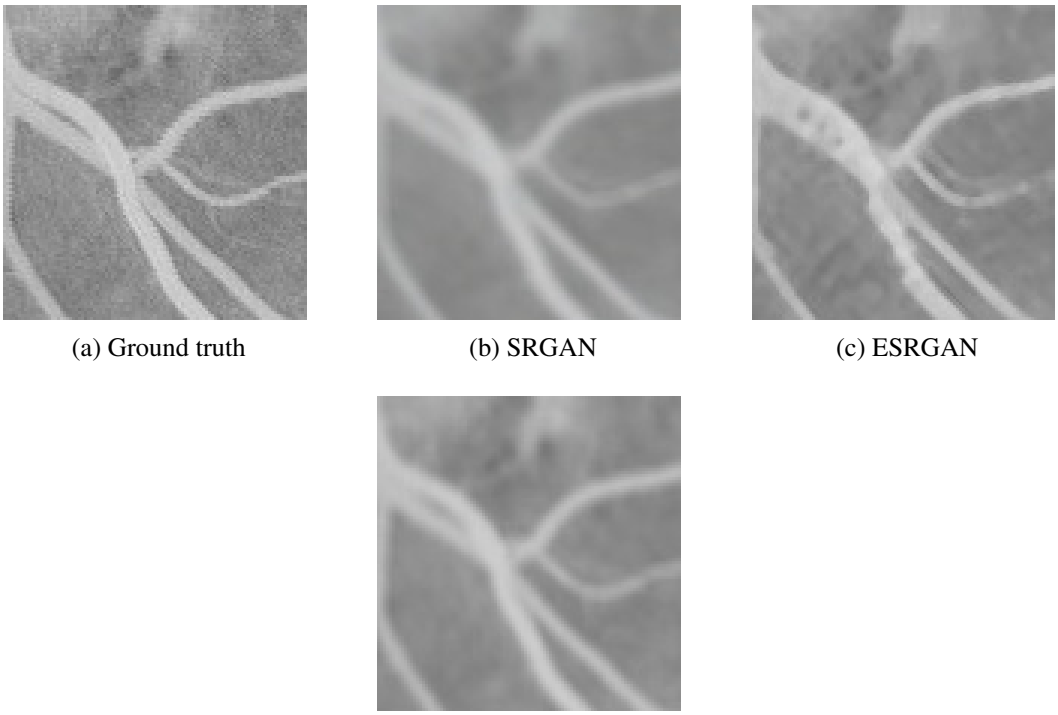


Figure 4.6: ROI on GT



(a) Ground truth (b) SRGAN (c) ESRGAN

(d) Proposed approach

Figure 4.7: Ground Truth compared with SRGAN, ESRGAN, Proposed approach

Here in Fig. 4.6, we select a small region of interest(ROI) from the a retinal angiograph and compare the ROI of super resolved images using SRGAN (Fig. 4.7-(b)), ESRGAN (Fig. 4.7-(c)), and our proposed approach (Fig. 4.7-(d)).

As we can clearly see our proposed approach performs better than SRGAN and ESRGAN in terms of visual clarity and allows for better determination of the vascular structure of the retina (refer Fig. 4.7-(a) and Fig. 4.7-(b), Fig. 4.7-(c), Fig. 4.7-(d). In the case of ESRGAN we see more textural noise added to the image while in the case of SRGAN we note that the image is more blurry. Our proposed approach is less blurry as compared SRGAN while it has less textural noise as compared to ESRGAN. The ESRGAN and SRGAN also misses out on thin capillaries in the eye that our proposed approach captures. These thin retino-vascular capillaries are essential for proper diagnosis of the eye. ESRGAN tends to cut off these capillaries arbitrarily while SRGAN tends to fade them out. In both cases studying them becomes harder. Our model on the other hand does a much better job at this and shows these capillaries better.

4.2 Quantitative analysis

We calculated the PSNR and SSIM values of the results of passing the image though our proposed approach, ESRGAN, and SRGAN with respect to the HR image:

Table 4.1: Average PSNR and SSIM for SRGAN, ESRGAN, and Our model

	SRGAN	ESRGAN	Our
PSNR	31.42	31.08	32.18
SSIM	0.618	0.644	0.693

From Table 4.1, we can observe that the PSNR and SSIM values are the best for our proposed approach. ESRGAN's SSIM is better than that of SRGAN while SRGAN's PSNR is better than ESRGAN's. Higher PSNR value indicates higher peak signal to noise ratio. Our model's output is also clearer as compared to SRGAN and ESRGAN. The SSIM value (structural similarity index measure) indicates that our model is more structurally similar to the ground truth than SRGAN and ESRGAN both. ESRGAN is more structurally similar to the ground truth than SRGAN.

Chapter 5
Conclusion and Future Work

5.1 Conclusion

In this report, we analysed various architectures and models based on CNN and GAN for super resolution. We proposed our model for super resolution based on SRResC-GAN architecture and performed ablation study. After implementing our own proposed method we observed that the architecture that we proposed produces better results as compared to SRGAN and ESRGAN on the basis of the high PSNR and SSIM values obtained for our specific dataset. While we trained our model on retinal angiograms, there are still a lot of issues with reconstructing high frequency and texture details while keeping sensitive information like blood vessels and thin details intact.

5.2 Future Work

Deep learning-based SR methods have shown outstanding performance on diverse medical imaging modalities, thanks to the rapid development of Convolutional Neural Network and Generative Adversarial Network.

• For retinal angiography, super-resolution technologies should be improved. There are still a lot of issues with reconstructing high frequency and texture details while keeping sensitive information like blood vessels and thin details intact.

• GAN based SR can generate extra information in the output that is not present in the LR image. As a result, it is vital to improve these techniques so as to avoid false generation.

• Further work on this model could be done to improve the hyper-parameters to result in better super resolution with lesser training time and resources.

• In future, we will try to extend our model to other modality of medical imaging such as Computer Tomography (CT) and MRI to better generalize our model for medical image super resolution.

References

[1] A. Warren, "X-ray imaging," https://informaconnect.com/the-science-behind-x-ray-imaging/, 2017, accessed: 2021-11-7.

[2] D. C. Preston, "Ct imaging," https://case.edu/med/neurology/NR/CT

[3] D. D. World, "Pet imaging," https://www.ddw-online.com/positron-emission-tomography-pet-shining-a-light-on-human-disease-1322-200704/, 2022, accessed: 2022.

[4] R. A. H. M. Unit, "Mri imaging," https://www.myvmc.com/investigations/3d-magnetic-resonance-imaging-3d-mri/, 2005, accessed: 2021-11-7.

[5] F. US, "Ultrasound imaging," https://www.livescience.com/32071-history-of-fetal-ultrasound.html, 2014, accessed: 2021-11-7.

[6] S. Marques, M. Bispo, P. Pimentel-Nunes, C. Chagas, and M. Dinis-Ribeiro, "Image documentation in gastrointestinal endoscopy: Review of recommendations," *GE - Portuguese Journal of Gastroenterology*, vol. 24, 06 2017.

[7] R. Muhammad Umer, G. Luca Foresti, and C. Micheloni, "Deep generative adversarial residual convolutional networks for real-world super-resolution," in *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[8] Y. Zhang and M. An, "Deep learning-and transfer learning-based super resolution reconstruction from single medical image," *Journal of healthcare engineering*, vol. 2017, 2017.

[9] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Proceedings of the IEEE*, 2021.

[10] H.-M. Zhang and B. Dong, "A review on deep learning in medical image reconstruction," *Journal of the Operations Research Society of China*, pp. 1–30, 2020.

[11] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," *Medical Image Analysis*, vol. 63, p. 101693, 2020.

[12] H.-P. Chan, L. M. Hadjiiski, and R. K. Samala, "Computer-aided diagnosis in the era of deep learning," *Medical physics*, vol. 47, no. 5, pp. e218–e227, 2020.

[13] D. Liu, K. S. Zhou, D. Bernhardt, and D. Comaniciu, "Search strategies for multiple landmark detection by submodular maximization," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 2831–2838.

[14] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[15] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, 1997.

[16] M. Tipping and C. Bishop, "Bayesian image super-resolution in advances in neural information processing systems," 2003.

[17] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE transactions on image processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[18] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Transactions on Multimedia*, vol. 21, no. 12, pp. 3106–3121, 2019.

[19] P. Akhtar and F. Azhar, "A single image interpolation scheme for enhanced super resolution in bio-medical imaging," in *2010 4th International Conference on Bioinformatics and Biomedical Engineering*. IEEE, 2010, pp. 1–5.

[20] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE transactions on acoustics, speech, and signal processing*, vol. 29, no. 6, pp. 1153–1160, 1981.

[21] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong, "Soft edge smoothness prior for alpha channel super resolution," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[22] J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.

[23] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European conference on computer vision*. Springer, 2014, pp. 184–199.

[24] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European conference on computer vision*. Springer, 2016, pp. 391–407.

[25] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1637–1645.

[26] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3147–3155.

[27] J. Park, D. Hwang, K. Y. Kim, S. K. Kang, Y. K. Kim, and J. S. Lee, "Computed tomography super-resolution using deep convolutional neural network," *Physics in Medicine & Biology*, vol. 63, no. 14, p. 145011, jul 2018. [Online]. Available: https://doi.org/10.1088/1361-6560/aacdd4

[28] J. Shi, Q. Liu, C. Wang, Q. Zhang, S. Ying, and H. Xu, "Super-resolution reconstruction of MR image with a novel residual learning network algorithm," *Physics in Medicine & Biology*, vol. 63, no. 8, p. 085011, apr 2018. [Online]. Available: https://doi.org/10.1088/1361-6560/aab9e9

[29] Z. Jiang, Z. Yu, S. Feng, Z. Huang, Y. Peng, J. Guo, Q. Ren, and Y. Lu, "A super-resolution method-based pipeline for fundus fluorescein angiography imaging," *BioMedical Engineering OnLine*, vol. 17, no. 1, pp. 1–19, 2018.

[30] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. 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*, 2017, pp. 4681–4690.

[31] Y. Gu, Z. Zeng, H. Chen, J. Wei, Y. Zhang, B. Chen, Y. Li, Y. Qin, Q. Xie, Z. Jiang *et al.*, "Medsrgan: medical images super-resolution using generative adversarial networks." *Multimedia Tools & Applications*, vol. 79, 2020.

[32] J. Zhu, G. Yang, and P. Lio, "Lesion focused super-resolution," in *Medical Imaging 2019: Image Processing*, vol. 10949. International Society for Optics and Photonics, 2019, p. 109491L.

[33] A. Lugmayr, M. Danelljan, and R. Timofte, "Unsupervised learning for real-world super-resolution," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 2019, pp. 3408–3416.

[34] A. Lugmayr, M. Danelljan, R. Timofte, M. Fritsche, S. Gu, K. Purohit, P. Kandula, M. Suin, A. Rajagoapalan, N. H. Joon *et al.*, "Aim 2019 challenge on real-world

image super-resolution: Methods and results," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 2019, pp. 3575–3583.

[35] M. Fritsche, S. Gu, and R. Timofte, "Frequency separation for real-world super-resolution," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 2019, pp. 3599–3608.

[36] ——, "Frequency separation for real-world super-resolution," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 3599–3608.

[37] C. H. Foundation, "Diabetic retinopathy detection dataset," https://www.kaggle.com/c/diabetic-retinopathy-detection/overview, 2015, accessed: 2021-11-10.

[38] S. H. M. Alipour, H. Rabbani, and M. R. Akhlaghi, "Diabetic retinopathy grading by digital curvelet transform," *Computational and Mathematical Methods in Medicine*, vol. 2012, 2012.

[39] S. A. Kamran, K. Fariha Hossain, A. Tavakkoli, S. Zuckerbrod, S. A. Baker, and K. M. Sanders, "Fundus2angio: a conditional gan architecture for generating fluorescein angiography images from retinal fundus photography," in *International Symposium on Visual Computing*. Springer, 2020, pp. 125–138.