

# MRI enhancements using neural networks

A Project Report Submitted  
in Partial Fulfillment of Requirements  
for the Degree of  
**Bachelor of Technology**

by  
Aditya Agarwal(2019CSB1064)  
Shikhar Soni (2019CSB1119)



Department of Computer Science & Engineering  
Indian Institute of Technology Ropar  
Rupnagar 140001, Punjab, India  
May 2023

# Abstract

Magnetic Resonance Imaging (MRI) is an essential medical imaging technique for diagnosing and monitoring various diseases. However, the low signal-to-noise ratio and long acquisition times of MR images can limit their clinical utility. In recent years, deep learning-based techniques have been proposed to enhance MR images.

In this project, we mainly explore two different methods to enhance MR images, one includes the use of U-Nets to obtain a close to fully sampled MRI from under-sampled MRI and the other includes the use of diffusion models to convert a low-resolution MRI to a higher resolution.

The first proposed approach is for getting a fully sampled MRI image from an under-sampled one. Here we train two U-Net models joined together to form a W-Net model, the input of the model is 3 consecutive slices of an MRI and it uses locality to improve on the previously existing W-net architecture [Souza and Frayne \[2019\]](#).

The second proposed approach for getting a higher resolution image from a lower resolution one works by using diffusion models and tries to improve on the known Super Resolution (SR) models [Saharia et al. \[2021\]](#) to get good results for MRIs while ensuring minimum artefacts.

In conclusion, our study highlights the potential of deep learning-based approaches for enhancing MR images and provides insights into the benefits of using diffusion models and combined U-Net models (W-Net with locality) in this context.

## Acknowledgements

We would like to express our sincere gratitude to our project mentor, Dr Deepti R. Bathula, for her invaluable guidance and support throughout the course of this project. Her expertise in the field of MRI enhancement and her willingness to share her knowledge has been instrumental in the success of this project. Her constant feedback and encouragement helped us to stay motivated and focused.

We would also like to thank our faculty advisor, Dr T.V. Kalyan, for his support and guidance in conducting this project course. His knowledge and experience in the field of computer science have been invaluable in shaping our approach to this project.

We also want to extend our heartfelt gratitude to Dr Puneet Goyal and Dr Abhinav Dhall, who evaluated our project and provided valuable feedback. Their insights and suggestions helped us to improve our project and make it more robust.

Finally, we thank all the staff members of our department who provided us with the necessary facilities and resources to conduct this project. Mr Ashu Kaushik has been helpful in providing us with the resources necessary for the completion of this project and helped sort out any technical issues.

## **Honor Code**

We certify that we have properly cited any material taken from other sources and have obtained permission for any copyrighted material included in this report. We take full responsibility for any code submitted as part of this project and the contents of this report.

Aditya Agarwal ( 2019CSB1064 )

Shikhar Soni ( 2019CSB1119 )

## **Certificate**

It is certified that the B. Tech. project “MRI Enhancements using Neural Networks” has been done by Aditya Agarwal (2019CSB1064), Shikhar Soni (2019CSB1119) under my supervision. This report has been submitted towards partial fulfillment of B. Tech. project requirements.

Dr. Deepti R. Bathula  
Project Supervisor  
Department of Computer Science & Engineering  
Indian Institute of Technology Ropar  
Rupnagar-140001

# Contents

<b>Contents</b>	<b>v</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>Nomenclature</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Existing solutions for the problem space</b>	<b>4</b>
2.1 Enhanced Deep Super-Resolution Network (EDSR) . . . . .	4
2.2 W-Net for MRI Reconstruction . . . . .	5
2.3 Super-Resolution via Repeated Refinements (SR3) . . . . .	6
<b>3 About the Datasets used</b>	<b>8</b>
3.1 Calgary-Campinas Public MR dataset . . . . .	8
3.2 fastMRI dataset . . . . .	9
3.3 Brain Tumor Segmentation(BraTS2020) dataset . . . . .	10
<b>4 Our Solutions</b>	<b>12</b>
4.1 MR Reconstruction Problem . . . . .	12
4.1.1 Solution to reduce training time ( $W - Net Combined$ ) . .	13
4.1.2 Solution to improve image quality ( $W - Net 3 - Layer$ ) .	13
4.2 MR Super Resolution Problem . . . . .	14

## CONTENTS

---

<b>5</b>	<b>Observations</b>	<b>16</b>
5.1	Part 1 – W-Net . . . . .	16
5.1.1	W-Net Combined . . . . .	17
5.1.2	W-Net 3-layer . . . . .	18
5.2	SR3 - Diffusion Model . . . . .	19
<b>6</b>	<b>Conclusion</b>	<b>24</b>
	<b>Appendix A</b>	<b>26</b>
.1	About the SR3 SSIM regularized model . . . . .	26
.2	Architecture details of our model . . . . .	28
	<b>References</b>	<b>29</b>

# List of Figures

1.1	A Brain MRI slice <a href="#">Bakas et al. [2017, 2018]</a> ; <a href="#">Menze et al. [2015]</a> .	2
2.1	ResNet vs SRResNet vs EDSR <a href="#">Lim et al. [2017]</a> . . . . .	4
2.2	W-Net architecture using two U-Nets <a href="#">Souza and Frayne [2019]</a> . .	5
2.3	Super-Resolution via Repeated Refinements on a face taken from Janspiry’s GitHub page . . . . .	7
3.1	Example MRI slice (image domain) from Calgary-Campinas dataset	9
3.2	An example MRI used from the fastMRI dataset . . . . .	10
3.3	An example MRI used from the BraTS 2020 dataset . . . . .	11
5.1	80 % undersampling mask for k-space . . . . .	16
5.2	W-Net Combined results – Predicted (Left), Original (Middle), Undersampled (Right) . . . . .	17
5.3	W-Net 3-layer results – Predicted (Left), Original (Middle), Un- dersampled (Right) . . . . .	18
5.4	SR3 with L2 . . . . .	21
5.5	SR3 with FID . . . . .	21
5.6	SR3 with SSIM . . . . .	22
5.7	SR3 with FIDR . . . . .	22
5.8	SR3 with SSIMR . . . . .	23
1	Inference process in T steps using the model learnt in training <a href="#">Saharia et al. [2021]</a> . . . . .	27



## LIST OF FIGURES

---

2	Outline of the U-Net showing the input $x$ (the bicubic interpolated image), $y_t$ (the noisy image at timestep $t$ ) and the output as $y_{t-1}$ obtained using Algorithm 2 (the noisy image at timestep $t-1$ , less noisy than at timestep $t$ ) <a href="#">Saharia et al. [2021]</a> . . . . .	27
---	---	----

# List of Tables

5.1	Comparison between W-Net models . . . . .	<a href="#">17</a>
5.2	Comparison between loss functions in SR3 . . . . .	<a href="#">20</a>
1	Architecture Details of the Model . . . . .	<a href="#">28</a>

# Nomenclature

## Greek Symbols

$\gamma$       The variance schedule

$\nabla$       Gradient of a function

## Subscripts

$t$       timestep  $t$

## Acronyms

*EDSR*   Enhanced Deep Super-Resolution Network

*FID*   Fréchet Inception Distance

*HR*   High Resolution

*MRI*   Magnetic Resonance Imaging

*NPSNR*   Normalized Peak Signal-to-Noise Ratio

*PSNR*   Peak Signal-to-Noise Ratio

*SR*   Super Resolution

*SR3*   Super-Resolution via Repeated Refinements

*SSIM*   Structural Similarity Index

$T$       Tesla (Measures Magnetic Field Strength)

# Chapter 1

## Introduction

Medical imaging has become a crucial tool in the diagnosis and treatment of various diseases, and the development of new image-processing techniques can help improve the quality and speed of medical imaging, especially MRIs.

Our report is mostly concerned with the use of MRIs, they use magnetic fields and radio waves to create detailed images of the inside of the body, especially useful for soft tissues such as the brain, muscles, and organs. MRIs are non-invasive and do not use ionizing radiation, making them a safer alternative to other imaging technologies such as X-rays or CT scans. However, MRIs can also be expensive and time-consuming to perform, and some patients may not be able to tolerate the confined space of the MRI machine.

In recent years, there has been an increased interest in using neural networks to help with this process. We propose two solutions to tackle two of the most important problems concerning MRI.

The first problem involves speeding up the process of taking an MRI. To speed up an MRI, we must be able to use the under-sampled MRI that takes lesser time to construct than a fully-sampled MRI output i.e, MRI Reconstruction problem (Section 4.1 discusses the problem in more detail). Our solution to this problem involves the use of the pre-existing W-Net architecture [Souza and Frayne \[2019\]](#). The W-Net architecture involves the use of two U-Nets joined together in such a

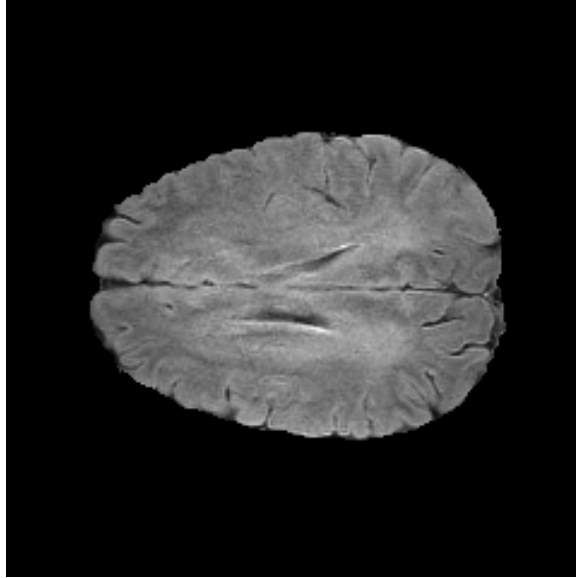


Figure 1.1: A Brain MRI slice [Bakas et al. \[2017, 2018\]](#); [Menze et al. \[2015\]](#)

way that the output of one is supplied to the other as input. The first residual U-Net works with the k-space domain MRI data and the second residual U-Net works with the image domain MRI data.

Our improvement to this model works with the following idea in mind – neighbouring MRI slices show very little variations between each other. In our approach, we take only 3 neighbouring slices to show any improvements. The model now takes 3 slices as input and outputs a single close-to-fully-sampled MR image for the middle slice. The neighbours help in providing local information relevant to constructing the middle slice’s fully sampled version.

The next problem being explored is the well-known problem of super-resolution in MRIs (Section 4.2 discusses the problem in more detail). In MRIs, however, it’s harder to perform super-resolution because we have to ensure that there are minimal artefacts introduced during the super-resolution process. Our approach to solving this problem involves the use of the pre-existing SR3 network [Saharia et al. \[2021\]](#). Here we don’t use MSE (Mean Square Error) as a judgement pa-

---

parameter for the likeness of two images because it doesn't ensure minimisation of artefacts, instead, we propose the use of SSIM (Structural Similarity Index) Wang et al. [2004] and FID (Fréchet Inception Distance) Mathiasen and Hvilshøj [2021]; Yu et al. [2021] as both of them promote higher structural similarity. We also want to minimize the noise in the model, a common occurrence in the result of diffusion models such as SR3, so we must also give some weightage to MSE as well.

Keeping the above in mind, we experimented with variations of the above to study the best use case for MRIs. Our report covers how the difference impacts MRI quality and tries to study the introduction of artefacts.

In the report, we aim to show the usefulness of both solutions and how they stand against their baseline counterparts.

## Chapter 2

# Existing solutions for the problem space

### 2.1 Enhanced Deep Super-Resolution Network (EDSR)

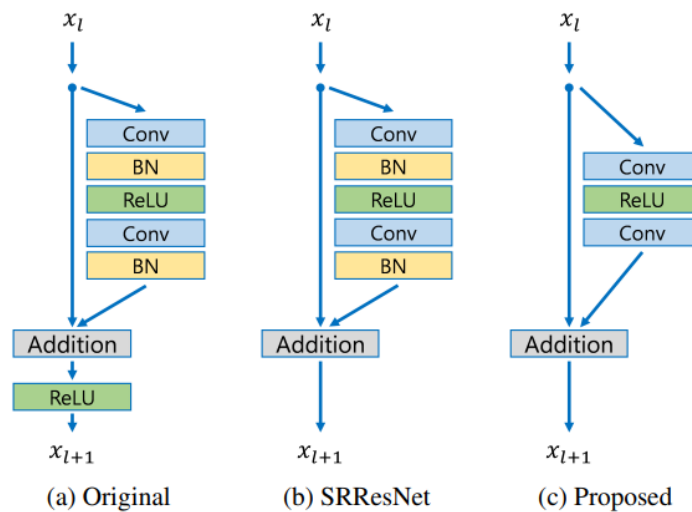


Figure 2.1: ResNet vs SRResNet vs EDSR [Lim et al. \[2017\]](#)

The EDSR architecture [Lim et al. \[2017\]](#) is based on the SRResNet architecture. It gives performance better than most state-of-the-art super-resolution models because it doesn't utilize the unnecessary modules present in conventional residual networks, and by increasing the model size.

An important feature of the EDSR network is that it can be used for higher magnifications such as 4X and 8X and has been used extensively for various kinds of super-resolution tasks.

Despite its large number of training parameters, EDSR is very simple to train and provides a very good PSNR for super-resolution, especially for medical imaging such as MRI which makes it a good baseline model for our work.

## 2.2 W-Net for MRI Reconstruction

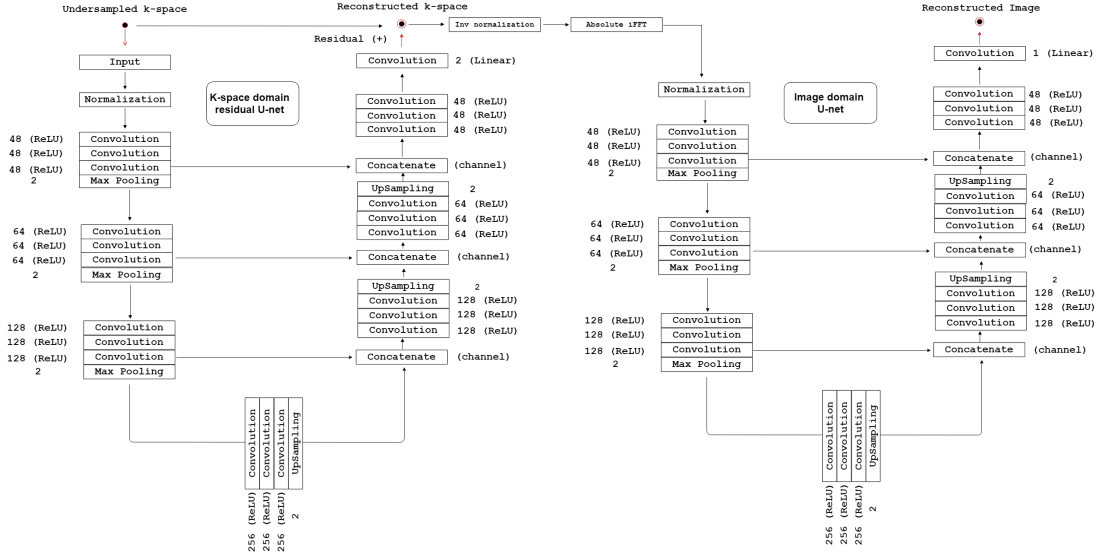


Figure 2.2: W-Net architecture using two U-Nets [Souza and Frayne \[2019\]](#)

W-Net is an architecture that comprises two similar U-Nets and works both in the k-space (or frequency-domain) and the image (or spatial) domains. The



---

model uses a complex-valued residual U-Net in the k-space domain, an iFFT operation, and a real-valued U-Net in the image domain (Souza and Frayne [2019]). This model is particularly suited to MR Images where there is an availability of raw k-space data to work with before it is converted into the image domain. The overall architecture can be seen in figure 2.2 borrowed from Xia and Kulis [2017].

This architecture displayed excellent results for reconstructing MR Images while utilizing lesser resources than other models such as GANs and Deep-Cascade networks. We aim to further enhance this model and tailor it further for MR Images.

## 2.3 Super-Resolution via Repeated Refinements (SR3)

Super-Resolution via Repeated Refinements (SR3) is a super-resolution diffusion model. Diffusion models have sprung up in recent years (Saharia et al. [2021]) as a model that provides accurate results, even more so than the popular Generative-Adversarial Networks (GANs Goodfellow et al. [2014]) while being easier to train (Dhariwal and Nichol [2021]).

Diffusion models have found their way into medical imaging and have become a popular option for the Super-Resolution of medical images. The SR3 model is trained by taking an image and adding noise to it progressively until only pure noise remains. The model learns to reverse this process so that it can create a high-resolution image from pure noise, iteratively, with guidance from the low-resolution image. Figure 2.3 shows how SR3 works by removing noise to produce a high-resolution image, given the low-resolution image for reference (taken from Janspiry’s GitHub page).

Some popular implementations already exist and use well-known loss functions such as L1 and L2 for training to provide good results. This model showed very promising results, and due to it being a relatively very new model, it is under-

---

researched. We aim to use this model for the super-resolution of MRIs.

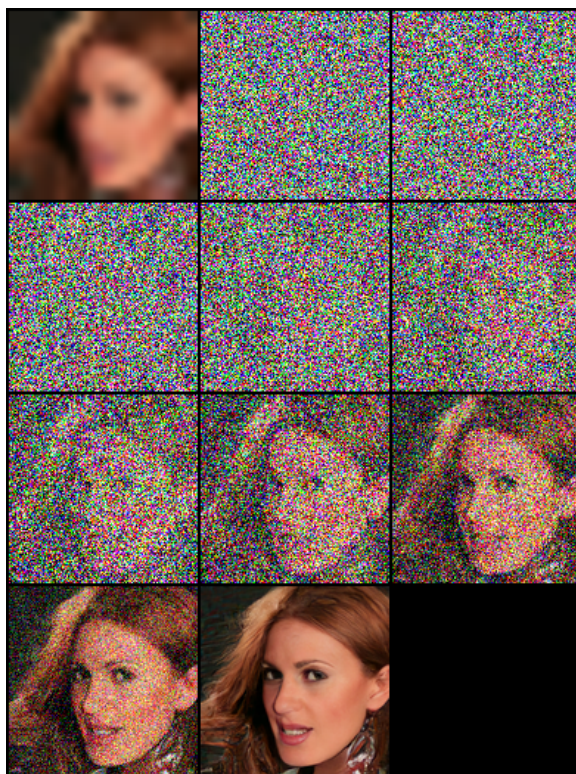


Figure 2.3: Super-Resolution via Repeated Refinements on a face taken from Janspiry's GitHub page

# Chapter 3

## About the Datasets used

### 3.1 Calgary-Campinas Public MR dataset

A collaborative effort between researchers at the Vascular Imaging Lab located at the University of Calgary and the Medical Image Computing Lab located at the University of Campinas (UNICAMP) resulted in the Calgary-Campinas Public MR dataset [Souza et al. \[2018\]](#).

The Calgary-Campinas Public MR dataset consists of 359 MRIs of healthy adults in the range of 29-80 years old. It consists of images from MRI scanners with two magnetic field strengths 1.5T (Tesla) and 3T (Tesla) and also contains scanners from 3 different companies, they are GE, Philips and Siemens. The data is also well divided among all the six combinations formed by the company name and magnetic field strength with approximately 60 images in each. The dataset is also gender and age-balanced as much as possible with the subjects at hand.

We have used 45 sideways brain MRIs from the Calgary-Campinas dataset (Single-channel coil data).

The MRIs have been divided into 25 for training (4524 slices), 10 for validation (1700 slices) and 10 for testing (1700 slices). Each MRI has approx 170 slices and the per-slice resolution is 256 X 256.

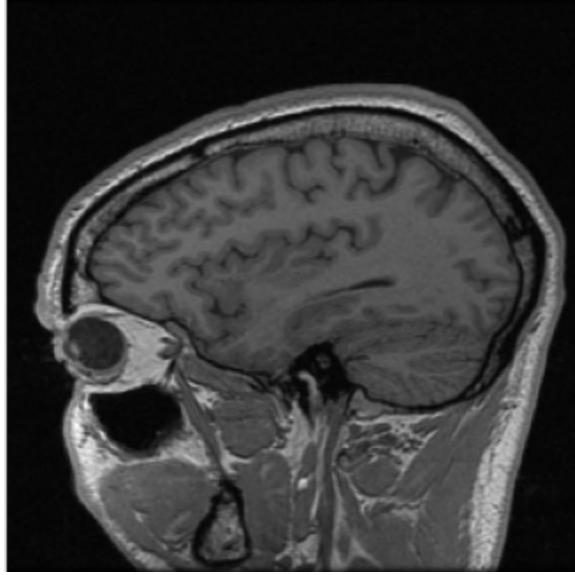


Figure 3.1: Example MRI slice (image domain) from Calgary-Campinas dataset

The purpose of using the single coil Calgary-Campinas dataset is to have a dataset that gives sharp images of healthy individuals and also is small enough to easily allow us to work with it, given the limited memory allocated to us on the provided hardware. The dataset was primarily used to train the W-Net, W-Net combined and W-Net 3-layer models, discussed in more detail in [4.1](#).

## 3.2 fastMRI dataset

The fastMRI dataset [Zbontar et al. \[2019\]](#) is a large-scale collection of both raw MR measurements and clinical MR images, that can be used for training and testing various machine learning approaches including MRI reconstruction, super-resolution, etc.

We used a small portion of the brain multicoil MRI scan data containing 10-30 slices in each MRI scan and with variable resolutions. We purposed the provided

### 3. About the Datasets used

---

data to give us a fixed resolution of 248 X 248 while ensuring no loss of relevant data and managed to obtain 5473 slices, where 80% was allotted to training and 20% was allotted to validation with 50 remaining for training.

We used this data initially to train known models such as ESPCN and EDSR to study the existing SR (Super Resolution) models and have a perspective of their capabilities when working with MRIs and the problems that often appear with them.

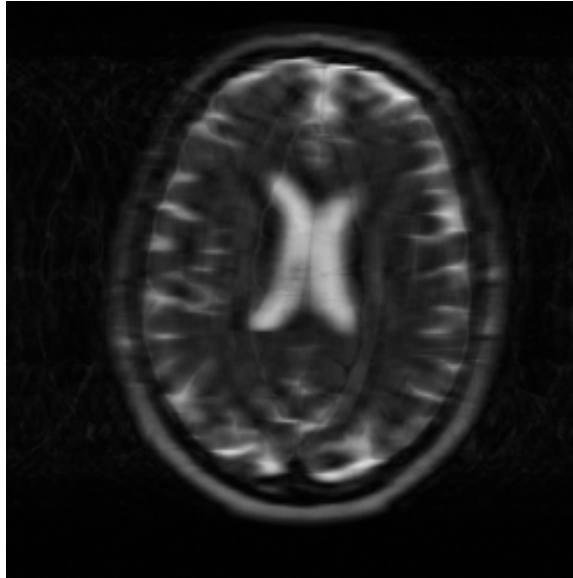


Figure 3.2: An example MRI used from the fastMRI dataset

### 3.3 Brain Tumor Segmentation(BraTS2020) dataset

The Brain Tumor Segmentation(BraTS2020) dataset [Bakas et al. \[2017, 2018\]](#); [Menze et al. \[2015\]](#) includes pre-operative MRI scans of patients with various types of brain tumours and is often used for brain tumour segmentation and analysis.

### 3. About the Datasets used

---

One MRI scan in the dataset has four different MRI modalities, they are native (T1) scans, post-contrast T1-weighted, T2-weighted and T2-flair. These scans were sourced from many scanners from 19 institutions.

The dataset has been purposed to provide us with 3312 images of 512 X 512 MR images. Brain MRI data is preferred for comparison as compared to other regions because of the very fine details present in the brain MRIs along with minimal artefacts that come in especially handy when it comes to Super Resolution tasks.

We preferred this dataset over fMRI dataset [3.1](#) because we observed a ringing effect in the fMRI multicoil scans and hence didn't want to compromise the results of super-resolution because of those artefacts. Calgary-Campinas dataset [3.2](#) on the other hand works with healthy humans but instead we want the SR problem to focus on a real-world environment. The BraTS dataset provides us with brain tumour MRIs to ensure that the results we obtain are reproducible in a real-world environment for this specific problem. This is because SR is extremely sensitive to varied environments and hence we preferred this dataset for the SR problem.

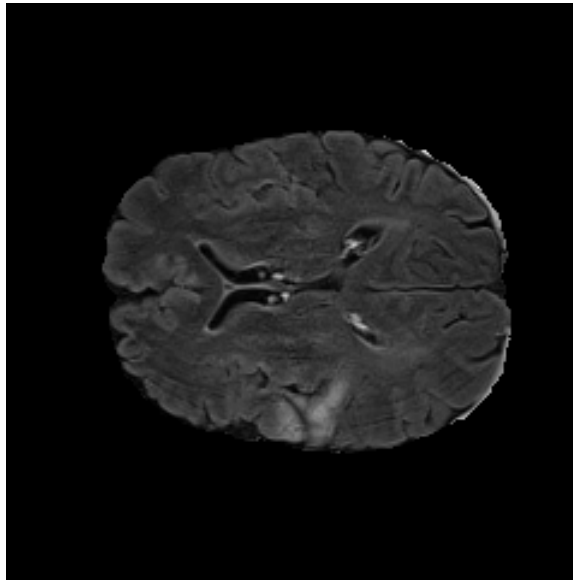


Figure 3.3: An example MRI used from the BraTS 2020 dataset

# Chapter 4

## Our Solutions

### 4.1 MR Reconstruction Problem

MR Reconstruction problem is a key area of research where the objective is to speed up the process of taking an MRI, the long wait time is problematic for a large group of patients such as children, people with fear of enclosed spaces, etc. A way to effectively tackle this problem is to take under-sampled MRI scans from a patient and try to fill in the gaps left behind by an incomplete scan with the help of neural networks.

Our Solution builds upon the pre-existing W-Net [Souza and Frayne \[2019\]](#). The working of the W-Net architecture goes as follows, it accepts a single k-space slice of an MRI. There are two components of the W-Net – the first U-Net working in the k-space domain and the other U-Net working in the image domain connected to the output of the previous U-Net (before the information is passed over to the other it is converted back to the image domain). This process works together using a combined loss function, i.e., NRMSE (Normalized Root Mean Square) and gives a certain weightage to each loss (1% weightage to the U-Net in the k-space domain and 99% weightage to the U-Net in image domain). The priority here lies in reducing image domain loss.

---

#### 4.1.1 Solution to reduce training time (*W-Net Combined*)

Having limited resources and time is a common problem for medical imaging and we tried to think of a simple way to reduce training time for the model by coming up with the idea to split the W-Net back into two U-Nets and train them independently on two separate machines and then train the two independently trained models together.

Individual U-Nets converge in lesser time than the complete W-Net and the idea was that combined training will help the model to converge quicker.

The combined network will give results similar to W-Net and train quicker than W-Net.

#### 4.1.2 Solution to improve image quality (*W-Net 3-Layer*)

Having discussed a method to reduce training time, we propose a method to improve on the idea of the W-Net.

Our solution uses the pre-existing W-Net architecture [Souza and Frayne \[2019\]](#) and adds another component to it, i.e., to use the localised information of the neighbouring slices to help learn the fully sampled MRI reconstruction. Here, instead of feeding one single slice of the MRI to our W-Net, we feed it 3 consecutive slices of an MRI (3 consecutive slices of an MRI are very similar in structure and look, and the information from the nearby slices act as additional information helpful to reconstruct a fully sampled MRI scan slice).

We limit our solution to only 3 slices because otherwise the model becomes too large to train given the resources available to us and while it is possible to adjust this parameter to obtain better results, it can be a point to be explored in detail separately.



---

## 4.2 MR Super Resolution Problem

Super Resolution is a well-known problem in deep learning and it has numerous applications in medical imaging as well.

Super Resolution of MRIs is a problem for multiple reasons given below.

1. MR images are more prone to noise because of reasons such as the motion of the patient, and noise caused by nearby electronics.
2. Large amount of data is required for training a super-resolution model but it's difficult to come by a large amount of medical imaging data.
3. MRIs contain complex features such as tissue contrast, noise, etc. that make it difficult to predict the high-resolution from the low-resolution image.

Our goal is to tackle the main bottleneck that appears in medical imaging and that is the requirement of a large dataset size. To tackle this issue we used a pre-trained SR3 model, this pre-trained model had already been trained on human face data and performed well in SR tasks in reference to human faces.

The issue of SR quality in MR images is another concern that we tackle by changing the likeness criterion for judging images. It's popular to use criteria such as L1 and L2 loss; they work well enough in general cases. But for MRIs, it's essential to capture the subtle details and show a contrast between different structures.

To achieve the above, we propose the use of judgement criteria such as the use of SSIM (Structural Similarity Index) Wang et al. [2004] and FID (Fréchet Inception Distance) Yu et al. [2021]. While FID and SSIM work well in terms of promoting structural similarity between the SR and the HR MR image, it causes noise to creep in and doesn't work as well as L1 or L2 loss when it comes to reducing the noise in the final resultant image.

---

The idea is to use a mixture of both the structural similarity promoting criterion and the criterion that promotes the reduction of noise, hence a mixture of both is used and we end up with two new approaches that we shall refer to as SSIM Regularized SR3 and FID Regularized SR3.

Both of the approaches use a combination of MSE with either SSIM or FID. The objective here is to study the impact of these on the resultant image in comparison to the original SR3.

The approaches discussed in this chapter will be discussed in more detail in Chapter 5 and there we will also compare the approaches to their original counterparts and present the observations related to them.

# Chapter 5

## Observations

### 5.1 Part 1 – W-Net

We experimented with two different model architectures; W-Net Combined and W-Net 3-layer. We discuss the observations relating to these below, against W-Net<sup>1</sup>. The input was corrupted in the k-space domain for training and testing using a random 80% undersampling mask (example in fig 5.1) to retain only 20% of the k-space data.

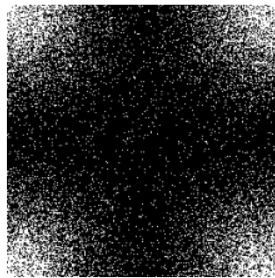


Figure 5.1: 80 % undersampling mask for k-space

---

Comparison table – W-Net models			
	PSNR	SSIM	Parameters
<b>W-Net</b>	$33.24 \pm 2.92$	0.8829	4M
<b>W-Net Combined</b>	$31.22 \pm 3.28$	0.7832	2+2+4 M
<b>W-Net 3-layer</b>	$32.21 \pm 0.53$	0.8385	4M

---

Table 5.1: Comparison between W-Net models

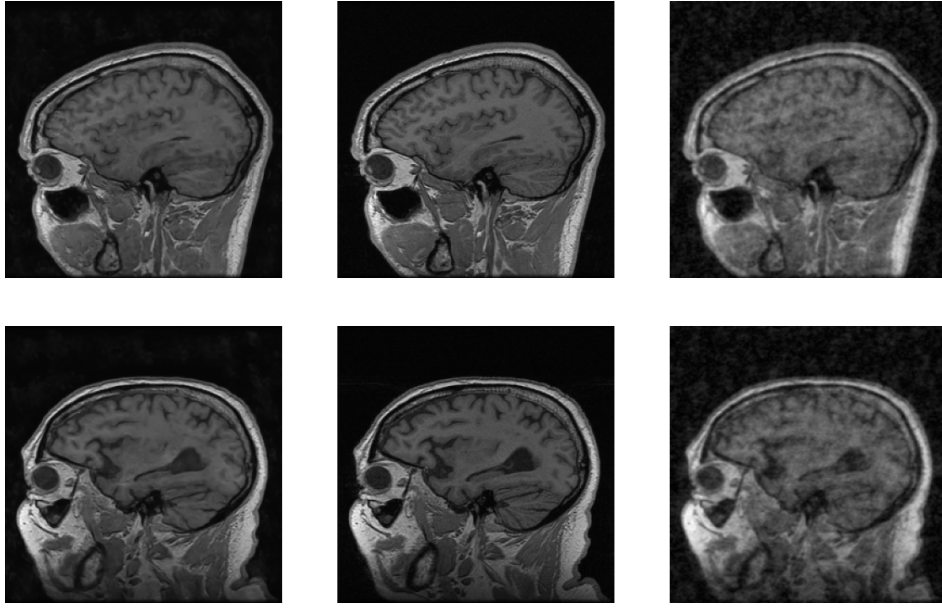


Figure 5.2: W-Net Combined results – Predicted (Left), Original (Middle), Undersampled (Right)

### 5.1.1 W-Net Combined

We trained Part 1 (k-space domain) and Part 2 (image domain) of the models separately for 20 epochs each on the k-space data for 25 MRIs, with 10 for validation, and finally trained the whole combined model by joining the two previously trained parts for 15 epochs. Each MRI contained around 170 image slices. The training was done on a Tesla T4 GPU.

---

<sup>1</sup>W-Net comparison model taken from [rmsouza01's GitHub page](#)

---

We don't have comparisons to other super-resolution algorithms such as EDSR or Bicubic interpolation because the image was corrupted in the k-space domain and retains the same pixel dimensions as the original. This is MR reconstruction rather than super-resolution.

We can see from figure 5.2 that a lot of the fine details are visible in the prediction and is very true to the original. The PSNR is very close (table 1) to the original W-Net, all while using significantly lesser computational time because of the split training technique. We believe that this model is an improvement over the original W-Net as it gives comparable results while reducing the training time and hardware requirements.

### 5.1.2 W-Net 3-layer

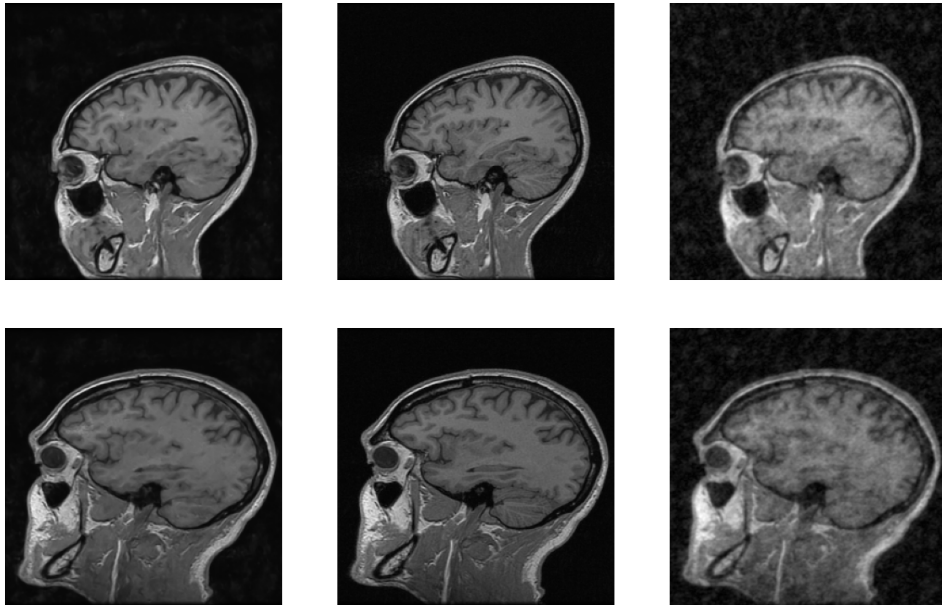


Figure 5.3: W-Net 3-layer results – Predicted (Left), Original (Middle), Under-sampled (Right)

---

We trained the model for 25 MRIs and validated it with 10 MRIs. Each MRI has about 170 slices so there are  $170-(3-1)=168$  inputs (3 slices combine to form one input). Training was done on a Tesla T4 GPU for 30 epochs.

As evident from figure 5.3 and table 1, the model performed on par with the original W-Net and was able to reproduce the fine details such as the folds in the cerebellum, despite the much lesser number of parameters (to fit with the limitations of current hardware) and less number of epochs (75 vs 30). If given better hardware and more time, this model will definitely be able to outperform the original W-Net on the given dataset. This model has a lot of potential and needs to be explored further using better hardware to get exceptional results.

## 5.2 SR3 - Diffusion Model

The SR3 model was trained over a number of different loss functions to improve image super-resolution quality.

1. L2 loss
2. FID
3. SSIM
4. FIDR - FID regularized (75% normalized PSNR, 25% normalized FID)
5. SSIMR - SSIM regularized (75% normalized PSNR, 25% SSIM)

The following are the results –

FIDR, SSIMR, and Bicubic Interpolation – all perform at a similar level if seen from the perspective of PSNR (as seen in table 5.2) but PSNR is not able to distinguish well when it comes to scenarios where one image may have distortions or artefacts while another may not and they end up having the same PSNR when compared with the true image.

---

Comparison table – SR3		
	PSNR	SSIM
<b>Bicubic</b>	$35.10 \pm 1.22$	0.8275
<b>L2</b>	$34.25 \pm 2.44$	0.8233
<b>FID</b>	$35.58 \pm 0.63$	0.8354
<b>SSIM</b>	$36.12 \pm 0.59$	0.8618
<b>FIDR</b>	$36.20 \pm 0.35$	<b>0.8632</b>
<b>SSIMR</b>	<b><math>36.77 \pm 0.34</math></b>	0.8569

---

Table 5.2: Comparison between loss functions in SR3

Hence, we will be giving more weightage when it comes to the judgement of quality to SSIM that can capture the perceptual quality of an image in a way that is more similar to humans as compared to PSNR.

SSIMR and FIDR models almost perform on a similar level with SSIMR showing the best qualitative results out of all. This shows that having a parameter that gives more weightage to structural similarity performed better as compared to vanilla SR3.

The L2 model i.e., the vanilla SR3, and bicubic interpolation showed a decently high PSNR but lagged behind when it came to factors such as structural similarity. Even with the naked eye, one can observe that L2 in Fig. 5.4 wasn't able to give a structurally accurate result when compared to, say, SSIMR in Fig. 5.8. In the original SR3 paper, they used human subjects and fool rate to get an idea of what image is original when it comes to the judgement of human subjects. Here, we leave the judgement to the reader to observe the difference between the results of these and SSIMR or FIDR.

On the other hand, we observed that SSIM and FID seemed to give worse performance than SSIMR or FIDR and also had more noise in their result as compared to the others.

Finally, we would like to conclude our observations by saying that the results

---

seem to show that there definitely is a correlation between having a loss function that gives weightage to structural similarity and MRI results that are more structurally similar to the original.

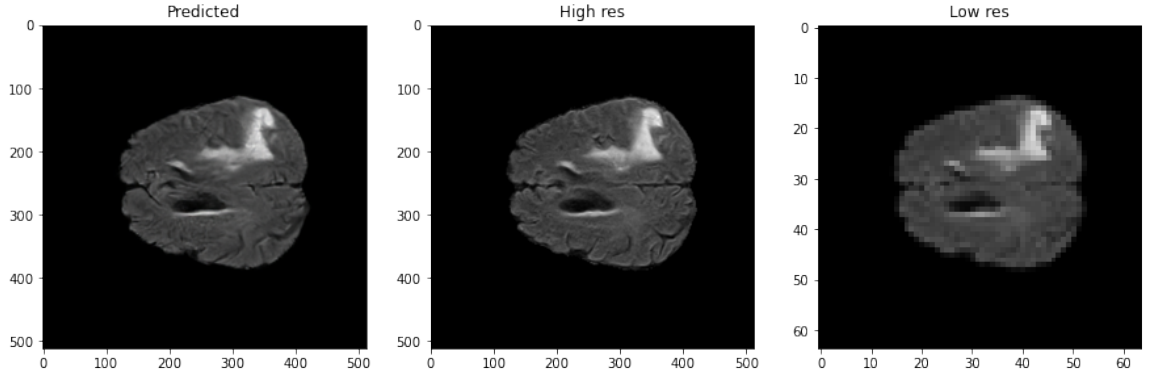


Figure 5.4: SR3 with L2

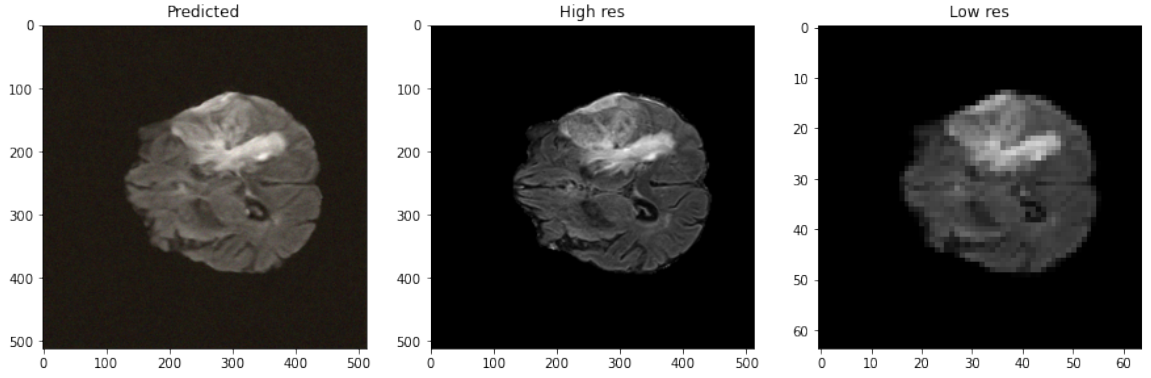


Figure 5.5: SR3 with FID



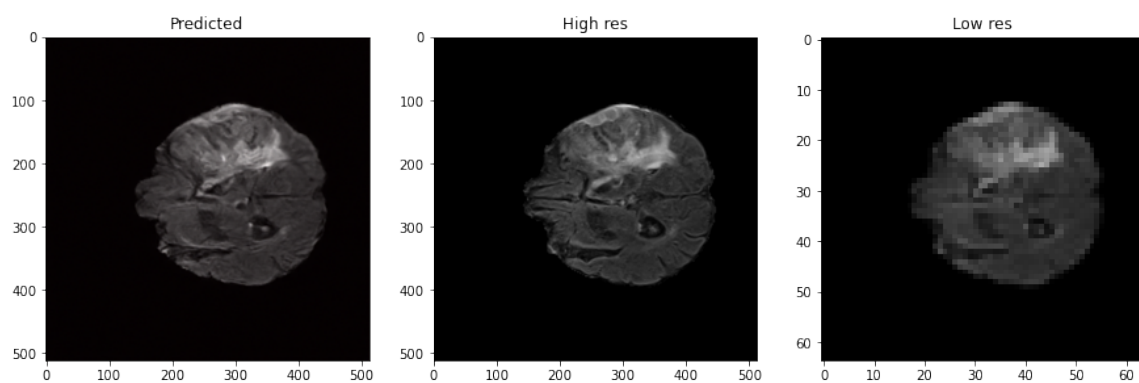


Figure 5.6: SR3 with SSIM

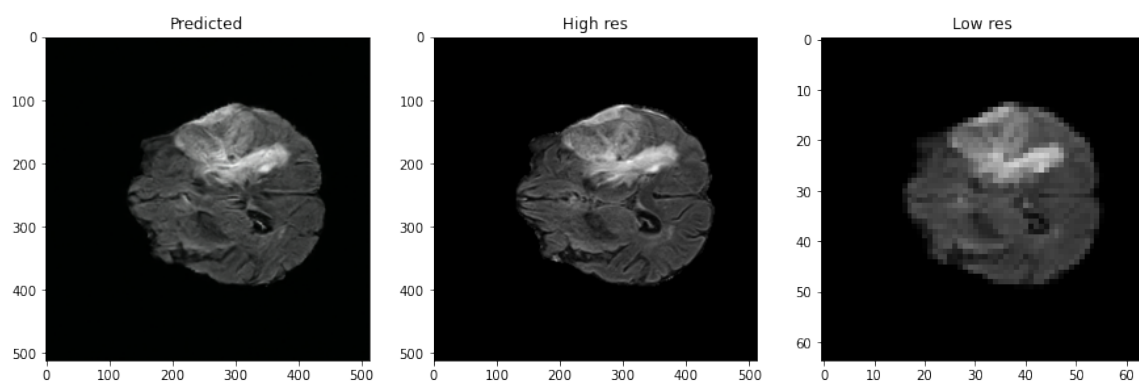


Figure 5.7: SR3 with FIDR

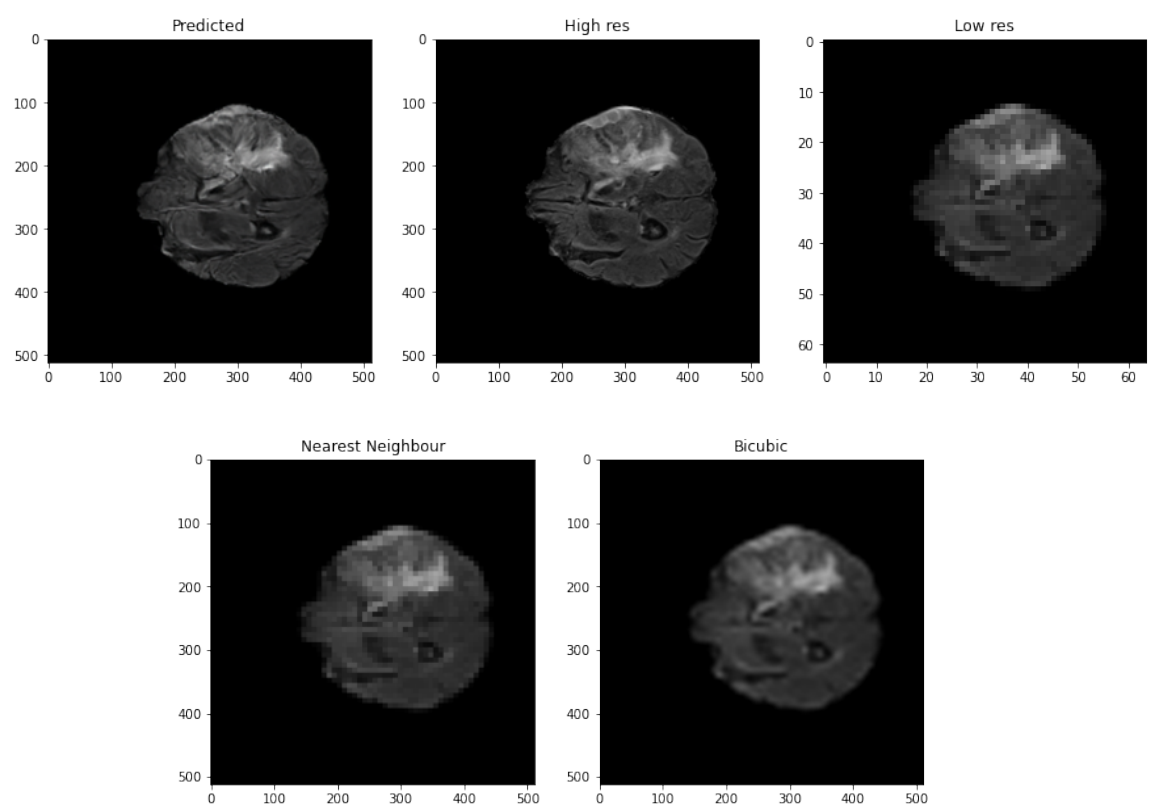


Figure 5.8: SR3 with SSIMR

# Chapter 6

## Conclusion

From the observations made earlier, we can conclude the following for each of the solutions discussed in Chapter 4.

Solution 1 discussed in section 4.1 talks about the MR reconstruction problem and our solution to the two issues faced in medical imaging were W-Net Combined and W-Net 3-Layer using the pre-existing W-Net Architecture Souza and Frayne [2019].

From the observations made about W-Net Combined in 5.1, it can be seen that the model performed at par with the original W-Net while consuming less time to train than the original W-Net.

The observations made about the W-Net 3-Layer show that the model also performed at par with the W-Net but with reduced convolution layers to adjust for the low specs of the hardware allotted to us. Despite the following issue, the model showed great performance with very low variance and showed a promise for improvement if trained in a better environment.

Solution 2 discussed in section 4.2 talks about the MR Super Resolution Problem and our solution aims to tackle the issue of quality in MR images, while also aiming to reduce training time and training dataset by using previously trained models on different problems (such as human face super-resolution). The solu-

---

tion proposed by us is the use of FID and SSIM regularization in SR3 algorithm [Saharia et al. \[2021\]](#).

The qualitative observations made in the section [5.2](#) show that the SSIM Regularized SR3 performed the best but the difference is small as far as we could observe for 100K iterations and it is possible that the results may go in another direction if the model is trained for more iterations. It is also important to note that the model has a small batch size that may be the cause of high variations thus affecting the results.

Overall, while our experiments were limited by resources and time, we still managed to show that the newly proposed models are able to perform at least at par with the existing models or slightly better. We managed to reduce training time with the help of W-Net Combined, showed promise of improved performance with W-Net 3-layer, and showed improved performance with the help of FIDR and SSIMR in SR3 when compared to their baseline models respectively.

# Appendix A

## .1 About the SR3 SSIM regularized model

SR3 algorithm works very similar to the original DDPM paper [Ho et al. \[2020\]](#). The training can be described as follows:

1. repeat
2.  $(x, y_0) \sim p(x, y)$
3.  $\gamma \sim p(\gamma)$
4.  $\epsilon \sim N(0, I)$
5. Take a gradient descent step on:

$$-(\nabla_{\theta} NPSNR(f_{\theta}(x, y_t, \gamma_t), \epsilon) + \lambda * SSIM(f_{\theta}(x, y_t, \gamma_t), \epsilon))$$

6. until converged

The above describes the training process, where a random sample is picked,  $x$  is the bicubic interpolated image from the LR image and  $y_0$  is the HR image,  $\gamma_t$  is the variance schedule constant and it is assumed to be a constant in the model for a given  $t$ .

The model is used to obtain the SR image back using a  $T$  step refinement process as described below in the original SR3 paper.

---

**Algorithm 2** Inference in  $T$  iterative refinement steps

---

```

1:  $\mathbf{y}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
2: for  $t = T, \dots, 1$  do
3:    $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  if  $t > 1$ , else  $\mathbf{z} = \mathbf{0}$ 
4:    $\mathbf{y}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{y}_t - \frac{1-\alpha_t}{\sqrt{1-\gamma_t}} f_{\theta}(\mathbf{x}, \mathbf{y}_t, \gamma_t) \right) + \sqrt{1-\alpha_t} \mathbf{z}$ 
5: end for
6: return  $\mathbf{y}_0$ 

```

---

Figure 1: Inference process in  $T$  steps using the model learnt in training Saharia et al. [2021]

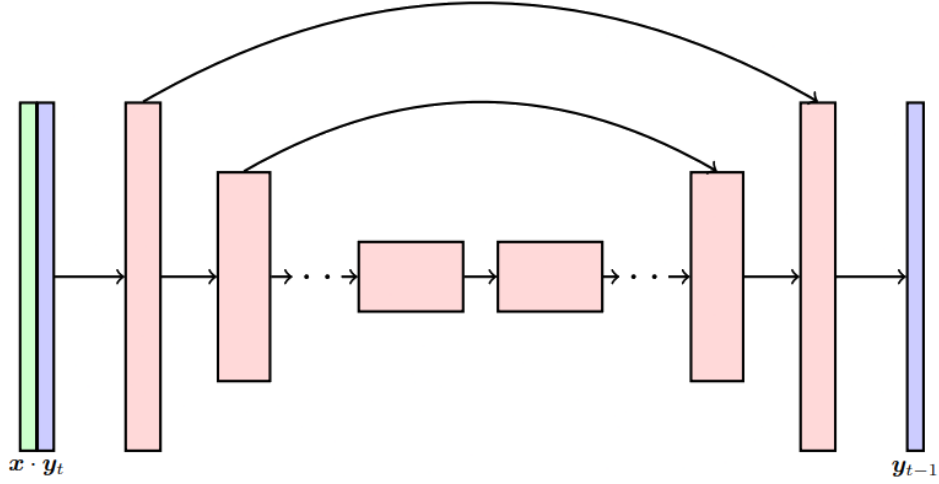


Figure 2: Outline of the U-Net showing the input  $x$  (the bicubic interpolated image),  $y_t$  (the noisy image at timestep  $t$ ) and the output as  $y_{t-1}$  obtained using Algorithm 2 (the noisy image at timestep  $t-1$ , less noisy than at timestep  $t$ ) Saharia et al. [2021]

---

## .2 Architecture details of our model

We have trained a variation of the SR3 i.e., SSIM regularized SR3 network with the following details.

1. We have focused on the task of converting 64 X 64 MR images to 512 X 512 image datasets.
2. We have used a pre-trained model (taken from Janspiry’s GitHub page), trained on human faces using FFHQ-CelebaHQ dataset [Karras et al. \[2018, 2019\]](#).
3. Other details are included in the table below

---

<b>Channel Multipliers</b>	{1,2,4,8,16}
<b>Timestep T</b>	2000
<b>Variance Schedule</b>	Cosine Variance Schedule
<b>Batch Size</b>	1 (Low GPU memory)
<b>Res Blocks</b>	1
<b>Dropout</b>	0
<b>Model Parameters</b>	155M
<b>Train steps</b>	100K

---

Table 1: Architecture Details of the Model

The FID model follows the same pattern as SSIM except for the function SSIM being replaced by FID and the architecture details also follow similarly to SSIM. The simple SR3 model also follows the same architecture and parameters but with the absence of regularization.

# References

- Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, and et al. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Nature Scientific Data*, 4:170117, 2017. doi: 10.1038/sdata.2017.117. [vii](#), [2](#), [10](#)
- Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, and et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018. [vii](#), [2](#), [10](#)
- Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis, 2021. [6](#)
- Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014. [6](#)
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020. [26](#)
- Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation, 2018. [28](#)
- Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks, 2019. [28](#)
- Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee.



## REFERENCES

---

- Enhanced deep residual networks for single image super-resolution, 2017. [vii](#), [4](#), [5](#)
- Alexander Mathiasen and Frederik Hvilshøj. Backpropagating through fréchet inception distance, 2021. [3](#)
- Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, and et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015. doi: 10.1109/TMI.2014.2377694. [vii](#), [2](#), [10](#)
- Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement, 2021. [i](#), [vii](#), [viii](#), [2](#), [6](#), [25](#), [27](#)
- Roberto Souza and Richard Frayne. A hybrid frequency-domain/image-domain deep network for magnetic resonance image reconstruction. In *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 257–264, Oct 2019. doi: 10.1109/SIBGRAPI.2019.00042. [i](#), [vii](#), [1](#), [5](#), [6](#), [12](#), [13](#), [24](#)
- Roberto Souza, Oeslle Lucena, Julia Garrafa, David Gobbi, Marina Saluzzi, Simone Appenzeller, Letícia Rittner, Richard Frayne, and Roberto Lotufo. An open, multi-vendor, multi-field-strength brain mr dataset and analysis of publicly available skull stripping methods agreement. *NeuroImage*, 170:482–494, 2018. ISSN 1053-8119. doi: <https://doi.org/10.1016/j.neuroimage.2017.08.021>. URL <https://www.sciencedirect.com/science/article/pii/S1053811917306687>. Segmenting the Brain. [8](#)
- Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861. [3](#), [14](#)
- Xide Xia and Brian Kulis. W-net: A deep model for fully unsupervised image segmentation, 2017. [6](#)

## REFERENCES

---

- Yu Yu, Weibin Zhang, and Yun Deng. Frechet inception distance (fid) for evaluating gans, 09 2021. [3](#), [14](#)
- Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdal, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova, James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastmri: An open dataset and benchmarks for accelerated mri, 2019. [9](#)