

MRI T2 Sequence Synthesis from T1: A Pix2Pix Approach for Enhanced Imaging

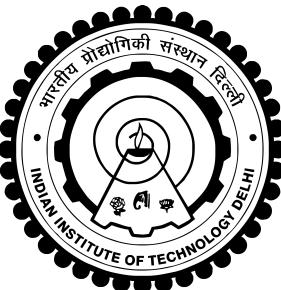
Thesis submitted by

Sahil Jain
2022AIB2677

under the guidance of
Prof. Anup Singh
Indian Institute of Technology Delhi

*in partial fulfilment of the requirements
for the award of the degree of*

Master of Technology



**Department Of School of AI
INDIAN INSTITUTE OF TECHNOLOGY DELHI**

May 2024

THESIS CERTIFICATE

This is to certify that the thesis titled **MRI T2 Sequence Synthesis from T1: A Pix2Pix Approach for Enhanced Imaging**, submitted by **Sahil Jain (2022AIB2677)**, to the Indian Institute of Technology, Delhi, for the award of the degree of **Master of Technology**, is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof.

Dr.Anup Singh

Dept. of Biomedical Engineering
IIT-Delhi, 110016

Place: New Delhi

Date: 28rd May 2024

ACKNOWLEDGEMENTS

First of all, I would like to thank Prof. Dr Anup Singh, my supervisor and guide throughout this challenging journey, for their excellent availability during these months of work but also for their patience, continuous support, the fundamental ideas and constructive discussions that I allowed this thesis to be realized. Thank you for stimulating me and making me more passionate about this subject. Thanks to my mom and dad, who never stopped encouraging me, and they have supported me in every way possible during these years, allowing me to concentrate on studying full-time and live a wonderful experience. All this, without them, wouldn't be possible.

ABSTRACT

Medical imaging plays a crucial role in diagnosing and understanding various health conditions. Magnetic Resonance Imaging (MRI) is a powerful imaging modality that provides detailed anatomical information. In general multiple MRI sequences such as T1-W, T2-W, FLAIR, DWI, etc. are acquired. However, this process take long acquisition time and sometime a particular image sequence might be corrupted or missing. Therefore, an alternative technique to synthesize an MRI sequence from another would be of great use. In this study, we propose a novel approach for generating one MRI sequence from other MRI sequence (primarily, generating T2 sequence from T1 sequence) using a pix2pix architecture. The motivation behind this work stems from the need to enhance the efficiency of MRI imaging by synthesizing additional sequences without the need for separate acquisitions. The T2 sequence provides valuable information about tissue characteristics, complementing the T1 sequence. Our proposed method leverages the power of adversarial learning to train a generator network to transform T1 sequences into realistic T2 sequences.

Contents

| | |
|---|-------------|
| ACKNOWLEDGEMENTS | i |
| ABSTRACT | ii |
| LIST OF TABLES | vi |
| LIST OF FIGURES | viii |
| ABBREVIATIONS | ix |
| 1 INTRODUCTION | 1 |
| 1.1 Common MRI sequences | 1 |
| 1.2 Background and preliminaries | 3 |
| 1.3 Problem Statement | 4 |
| 1.3.1 Challenges in obtaining different sequences | 4 |
| 1.4 Objectives | 4 |
| 2 State of the Art | 5 |
| 2.1 GANs in Medical Imaging | 5 |
| 2.1.1 Generative Adversarial Network | 5 |
| 2.1.2 Pix2Pix GAN | 6 |
| 3 Data and Preprocessing | 9 |
| 4 Evaluation Metrics | 10 |
| 5 Optimization Strategies | 12 |
| 6 Results | 15 |
| 7 Discussion and Conclusion | 35 |
| 7.1 Discussion : | 35 |

| | |
|-----------------------------|----|
| 7.2 Future Work : | 37 |
| 7.3 Conclusion : | 37 |

List of Tables

| | | |
|------|---|----|
| 6.1 | RMSE Error on Test Images Using 130 Patients Data for Training with and without Regularizer | 15 |
| 6.2 | RMSE Error and SSI on Test Images Using 338 Patients Data for Training with BCE and MSE loss | 16 |
| 6.3 | RMSE and SSIM on Test Images Using Different Numbers of Patients for Training | 17 |
| 6.4 | RMSE and SSIM on Test Images Using 250 Patients for Training and varying lambda | 19 |
| 6.5 | RMSE and SSIM on Test Images Using 250 Patients for Training and Varying Batch Size | 20 |
| 6.6 | Comparison of Model Performance: Optimal Dataset vs. Full Dataset with Data Augmentation | 22 |
| 6.7 | RMSE and SSIM on Test Images Using Different Augmentation Techniques | 22 |
| 6.8 | RMSE and SSIM on Test Images using different Random Rotations in with Range of $-\theta$ to θ | 24 |
| 6.9 | Comparing RMSE and SSIM on Test Images with and without vertical flip as Augmentation Technique | 24 |
| 6.10 | RMSE and SSIM on Test Images for Different Epochs with range of rotation from -45 to 45 degree | 24 |
| 6.11 | RMSE and SSIM on Test Images for Different Learning Rates and Epochs with range of rotation from -45 to 45 degree | 25 |
| 6.12 | RMSE and SSIM on Test Images Using Synthetic Data Generated with Augmentation Techniques at Different Epochs | 26 |
| 6.13 | RMSE and SSIM on Test Images Using Fourier Learning | 27 |
| 6.14 | RMSE and SSIM on Test Images Using Patch-Based Learning Approach | 28 |
| 6.15 | Comparison Of Relative CNR Models Trained On Full Image Approach and Patch Image Approach | 29 |
| 6.16 | Model trained by adding synthetic data generated using augmentation technique for 50 epochs (T2 to T1) | 31 |
| 6.17 | RMSE and SSIM on Test Images Using Patch-Based Learning Approach(T2 to T1) | 32 |

| | |
|---|----|
| 6.18 Comparison Of Relative CNR Models Trained On Full Image Approach and Patch Image Approach(T2 to T1) | 33 |
|---|----|

List of Figures

| | | |
|------|---|----|
| 1.1 | T1-Weighted Image | 1 |
| 1.2 | T2-Weighted Image | 2 |
| 1.3 | FLAIR | 2 |
| 2.1 | Generative Adversarial Network Framework | 5 |
| 2.2 | Generator architecture from pix2pix | 6 |
| 2.3 | Discriminator architecture from pix2pix | 7 |
| 6.1 | Generated Image With Regularization | 15 |
| 6.2 | Generated Image Without Regularization | 15 |
| 6.3 | Generated Image when trained with BCE loss function | 16 |
| 6.4 | Generated Image when trained with MSE loss function | 16 |
| 6.5 | RMSE and SSI with varying no of patients data for training. | 17 |
| 6.6 | Generated Images trained with 150 patients | 17 |
| 6.7 | Generated Images trained with 200 patients | 17 |
| 6.8 | Generated Images trained with 250 patients | 18 |
| 6.9 | Generated Images trained with 300 patients | 18 |
| 6.10 | Generated Images trained with 338 patients | 18 |
| 6.11 | Generated Images at $\lambda = 100$ | 19 |
| 6.12 | Generated Images at $\lambda = 500$ | 19 |
| 6.13 | Generated Images at $\lambda = 1000$ | 19 |
| 6.14 | Generated Images at $\lambda = 1500$ | 20 |
| 6.15 | Generated Images at batch size = 30 | 20 |
| 6.16 | Generated Images at batch size = 50 | 21 |
| 6.17 | Generated Images at batch size = 80 | 21 |
| 6.18 | Generated Images at batch size = 100 | 21 |
| 6.19 | Generated Images at batch size = 120 | 22 |

| | |
|--|----|
| 6.20 Generated Image when used Gaussian blur with kernel=29 and $\sigma=11$ | 23 |
| 6.21 Generated Image at $\theta=[-45,+45]$ and 20 epochs | 25 |
| 6.22 Generated Images when trained along with synthetic data at 30 epochs | 27 |
| 6.23 Generated image and log magnitude when trained using Fourier learning approach | 28 |
| 6.24 Generated image when trained using Patch-Based learning approach | 29 |
| 6.25 Segmented Tumor Area in Real and Generated Images | 30 |
| 6.26 Generated image when trained along with synthetic data at 50 epochs(T2 to T1) | 32 |
| 6.27 Generated Images when used patch based learning approach (T2 to T1) | 33 |
| 6.28 Segmented Tumor Area in Real and Generated Images | 34 |
| 7.1 Cases when generated image is not able to capture fine details and model performs poorly | 36 |

ABBREVIATIONS

| | |
|--------------|---------------------------------------|
| IITD | Indian Institute of Technology, Delhi |
| MRI | Magnetic resonance imaging |
| GAN | Generative Adversarial Network |
| BRATS | Brain Tumor Image Segmentation |

Chapter 1

INTRODUCTION

Medical imaging is a fundamental component of modern healthcare, providing medical professionals with essential tools for diagnosing illnesses, monitoring the course of treatment, and visualizing internal structures. Magnetic Resonance Imaging (MRI) is one of the most important imaging modalities. Using radio waves and strong magnetic fields, this technology creates detailed images of the inside of the body.

MRI delivers its multi-functionality utilizing many sequences, where each sequence is created to highlight different parts of the body or the characteristics of the tissue. The clinical problem in front of us determines the MRI sequence, which in turn ensures that the image produced is rich in detail and has a wide view of various health problems. Because of its accuracy and versatility, magnetic resonance imaging (MRI) is a vital tool in modern medicine, greatly advancing our ability to diagnose and monitor patients. However, traditional MRI imaging approaches often requires separate scans for each desired sequence, leading to extended scan times, increased resource utilization, and potential patient discomfort. In response to these challenges, this research explores advanced image synthesis techniques to generate T2 sequences directly from existing T1 data, eliminating the need for separate acquisitions.

1.1 Common MRI sequences

(1) **T1-Weighted (T1W) Sequence:** T1W images differentiate tissues with different relaxation times. They provide anatomical details and are used for structural imaging.

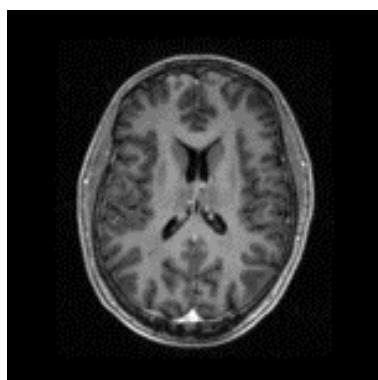


Figure 1.1: T1-Weighted Image

(2) **T2-Weighted (T2W) Sequence:** T2W MR images illustrate differences in the amount of fluid and are mainly used for the detection of fluids. They are used in the diagnosis of illnesses through imaging such as the detection of edema, inflammation, and fluid-filled cysts.

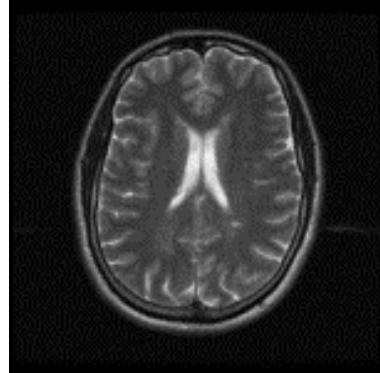


Figure 1.2: T2-Weighted Image

(3) **Fluid-Attenuated Inversion Recovery (FLAIR) Sequence:** FLAIR sequences suppress the signal from the CSF, thereby making the brain lesions and abnormalities visible by minimizing the CSF-related artifacts.

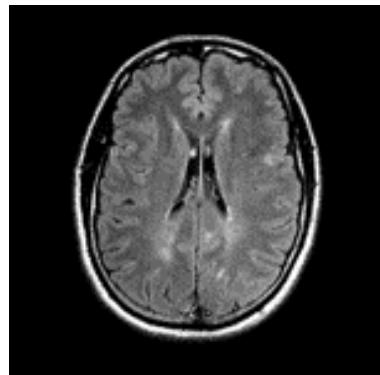


Figure 1.3: FLAIR

Besides these, many other types of sequences are used to study different body parts and medical conditions. The sequence of tests is determined by the testing purposes and the features of the tissues or abnormalities being evaluated. With the help of T1, T2, and advanced sequences like FLAIR, radiologists can get a complete insight into the patient's condition which is useful for accurate diagnosis and treatment planning.

1.2 Background and preliminaries

1. Abeer Aljohani, 2022[5]:

Deep learning and Machine learning approaches have been increasingly popular in recent years. It was thought to be impossible, but medical image processing has been accomplished. The primary goal of medical image processing is to depict and extract abnormal information present in the patient's medical data. Medical images are obtained through several medical technologies, including Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET), and Ultrasound (US). These obtained images are processed via deep learning techniques for deriving the most important diagnosis of the presented disease case in the image.

2. Goodfellow, 2014[4]:

One of the research fields in medical image processing is generating synthesized images based on GAN. The framework architecture includes two networks: one that generates fake images and the second one differentiates the original and synthetic images among them. GANs have attracted a lot of achieved attention in medical image analysis systems and various GAN models have recently focused on generating high-quality synthetic images. Recently, the GAN framework has been applied to several medical imaging tasks. Most of the research has been accomplished by the GAN, which is an image-to-image framework to generate image translation.

3. Han, 2018[3]:

In this study performance of DCGAN and WGAN is compared for generating synthetic multi-sequence brain MR images. In which WGAN outperforms DCGAN and can generate realistic multi-sequence brain MR images, possibly leading to valuable clinical applications: data augmentation because of more stable training as compared to DCGAN.

4. Isola, 2017[1]:

We investigate conditional adversarial networks as a general-purpose solution to image-to-image translation problems. These networks not only learn the mapping from input image to output image but also learn a loss function to train this mapping.

5. Nripendra Kumar Singh, 2020[2]:

Conditional GANs (cGANs) like CycleGAN and pix2pix have been highly successful in various image-to-image translation tasks, including converting satellite images to maps and vice versa. CycleGAN can be effectively used for those tasks where data collection is difficult (e.g. input-output pairs) is not applicable. It can learn translation mappings from unpaired data and represents an unsupervised learning framework. Paired data is necessary for the Pix2Pix model to function. Pix2Pix facilitates this translation process better than CycleGAN and is a model that is based on supervised learning. Pix2pix model performs superior to that of CycleGAN in producing much more realistic images whenever we have the paired data due to the fact this model has more control during translation.

1.3 Problem Statement

Even though MRI is an indispensable part of modern medical diagnostics, In general, additional scans are needed to provide the specific imaging sequences, like T1 and T2 leading to a longer time of examination, higher resource consumption, and the patient's discomfort as well. Workflow optimization and patient comfort are challenged by the inefficiencies of the conventional MRI imaging procedure.

1.3.1 Challenges in obtaining different sequences

1. **Time-Consuming Process:** Acquiring multiple sequences requires additional scan time. This can be particularly challenging for patients who may find it uncomfortable to remain still for extended periods.
2. **Patient Cooperation and Comfort:** Some patients like those with claustrophobia, anxiety and pain may find the MRI procedure distressing. Having multiple sequences could worsen their discomfort and consequently diminish their desire to continue the method.
3. **Resource Allocation:** MRI machines are expensive high-frequency machines having high demand. Scheduling scans that last for longer times for one patient and acquiring various kinds of sequences could reduce the current scanner availability for the other patients waiting in the queue.
4. **Radiation Concerns:** As is in X-ray or CT scans, MRI doesn't use ionizing radiation, thus being a safer choice. Moreover, for these patients who require repeated scans, exposure to the strong magnetic fields of the MRI machine could be a matter of concern.

1.4 Objectives

The primary focus is on developing a deep learning model adept at translating one MRI sequence into another with utmost precision, emphasizing both high resolution and accuracy.

Chapter 2

State of the Art

In this chapter we first present the related work on the Generative Adversarial Networks in the literature (2.1) and then, in Section 2.2, we introduce the theoretical aspects our work is based upon, while giving the reader an overview of the tools and techniques we applied in our models, in order to better understand the setting in which our work takes place.

2.1 GANs in Medical Imaging

Several studies about medical image synthesis were approached using GANs. In particular, cross-modality image synthesis (the conversion of the input image of one modality to the output of another modality) is the most critical application of this architecture.

2.1.1 Generative Adversarial Network

Generative Adversarial Network was proposed in 2014 by Ian J. GoodFellow [4] and represents a new framework for estimating generative models in an adversarial setting. The system is composed by two neural networks: a discriminator D, typically a CNN, and a generator G that are trained simultaneously.

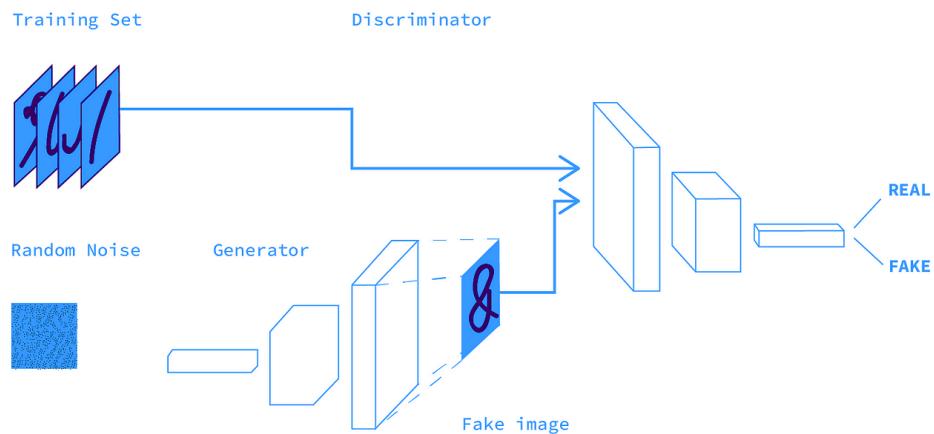


Figure 2.1: Generative Adversarial Network Framework

In particular, G is trained to learn the probability distribution of the input data and generate synthesized data that exhibits characteristics similar to those of authentic data. At the same time, the discriminative model estimates the probability that a sample came from the training data rather than G.

2.1.2 Pix2Pix GAN

The pix2pix model is a specific type of conditional generative adversarial network (cGAN) designed for image-to-image translation tasks.

Generator :

Instead of using the typical encoder-decoder model, the generator uses a U-Net model architecture. An encoder (downsampler) and a decoder (upsampler) make up a U-Net. The encoder-decoder generator architecture involves taking an image as input and downsampling it over a few layers until a bottleneck layer, where the representation is then upsampled again over a few layers before outputting the final image with the desired size with links or skip-connections are made between layers of the same size in the encoder and the decoder, allowing the bottleneck to be circumvented.

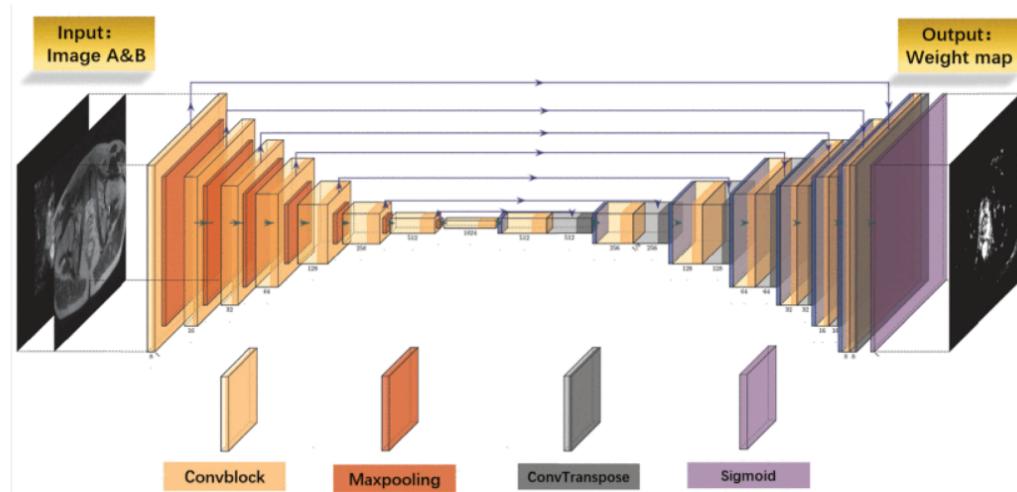


Figure 2.2: Generator architecture from pix2pix

Discriminator :

The discriminator model's job is to evaluate images. It takes two inputs: one image from the source domain and one from the target domain. Its goal is to predict the likelihood that

the target domain image is a real, authentic version of the source domain image, or if it's a generated fake.

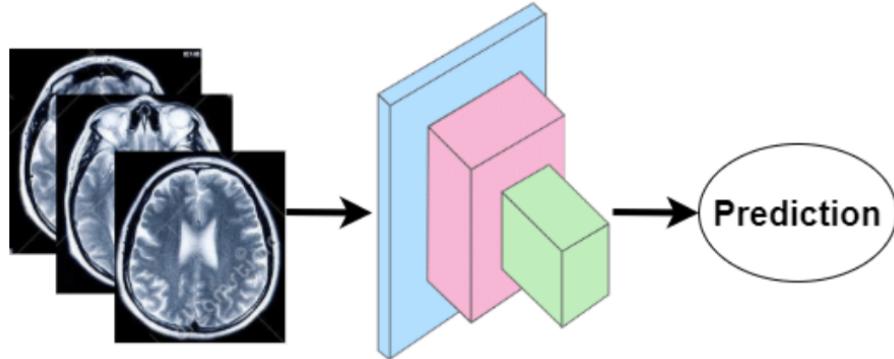


Figure 2.3: Discriminator architecture from pix2pix

Input :

Image from a source domain and Image from the target domain.

Output :

The probability that the image from the target domain is an authentic translation of the source image. Unlike the traditional GAN model that uses a deep convolutional neural network to classify images, the Pix2Pix model uses a PatchGAN. Rather than classifying an entire input image as real or fake, this deep convolutional neural network is intended to identify individual patches of the image.

Generator loss :

The generator loss is a sigmoid cross-entropy loss of the generated images and an array of ones. The Pix2Pix has L1 loss, which is an MAE (mean absolute error) between the generated image and the target image. This allows the generated image to become structurally similar to the target image. The formula to calculate the total generator loss is The total loss is given by:

$$\text{Total loss} = \text{Gan_loss} + \lambda \times \text{L1_loss} \quad (2.1)$$

where λ is a hyperparameter.

Discriminator loss

The discriminator loss function takes 2 inputs: real images and generated images. the real loss is a sigmoid cross-entropy loss of the real images and an array of ones(since these are the real images). generated loss is a sigmoid cross-entropy loss of the generated images and an array of zeros (since these are fake images).

$$\text{Total loss} = \text{Real_loss} + \lambda \times \text{Generated_loss} \quad (2.2)$$

Chapter 3

Data and Preprocessing

For training BRATS 2020 dataset of MRI scans from 369 patients with a type of brain tumour called diffuse glioma is utilized. Each patient's scan has four types of images: T1-weighted (T1), post-contrast T1-weighted (T1ce), T2-weighted (T2), and T2 Fluid-Attenuated-Inversion-Recovery (FLAIR). The dataset contains both training and validation data. To handle the images a library called nibabel is used. Every image is loaded using nibabel which transforms it into an input array of shape 248*248*155 which is the shape of the each input.

Only T1W and T2W sequences are of specific interest so they have been taken from the dataset only. T1W is considered as the input image and T2W as the targeted image. By taking into account all of the slices that each image was processed and cut into, 155 images of shapes 248*248* are obtained, resulting in 57,040 images in the dataset. The pix2pix model expects images to be a certain size, so each image has been resized to 256*256*1 and then converted to RGB using the PIL library to get an image of shape 256*256*3 and normalized accordingly.

For training and testing the model, we split the BRATS dataset, using data from 338 patients for training (52,390 images) and 20 patients for testing (3,100 images).

DATA Preprocessing for Fourier learning

For Fourier-based learning, the images were first changed to the frequency domain using the Fourier transform by NumPy. This transformation resulted in two channels, real and imaginary, which were merged together to form an input shape of 256*256*2. During the evaluation process, the images were brought down from the frequency domain to a spatial domain using the inverse Fourier transform and were subjected to the analysis.

DATA Preprocessing for Patch Based learning

For patch-based learning, the input images are divided into small patches on which the model is trained. The input image of shape 256*256*3 is divided into 5 patches of shape 128*128*3 . This division includes 4 patches at the corners and 1 patch at the center of the image, resulting in 5 patches per input image. Thus the data set used for training initially consisting of 52,390 images is transformed into 261,950 patch images, each of size 128*128*3.

Chapter 4

Evaluation Metrics

Using the test dataset, the proposed model's performance is assessed using following key metrics: The Structural Similarity Index (SSIM), the Root Mean Squared Error (RMSE), Relative Contrast To Noise Ratio(CNR), and Histogram Visualization These approaches were chosen to provide a comprehensive evaluation of the resulting MRI T2 sequences by comparing their accuracies pixel-by-pixel, structural integrity, frequency of each pixel, and the relative contrast of tumor. This provides insight into how well the model can synthesize T2 sequences in accordance with ground truth data.

1. Structural Similarity Index (SSIM): SSI measures the structural similarity between the generated and ground truth images. A higher SSI indicates better structural similarity between the synthetic and real T2 sequences.

$$SSIM = \frac{(2\mu_x + \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (4.1)$$

where:

- x, y are the two images to be compared,
- μ is the mean intensity,
- σ^2 the variance of the image,
- σ_{xy} the covariance of x, y .

2. Root Mean Squared Error (RMSE): RMSE quantifies the average squared difference between pixel intensities in the generated and ground truth images. A lower RMSE suggests reduced pixel-wise error, indicating better fidelity of the generated T2 sequences.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (4.2)$$

where:

- y_i is the original image,
- \hat{y}_i is the synthesized image.

3. Contrast-to-Noise Ratio (CNR): The Contrast-to-Noise Ratio (CNR) is a metric utilized in medical imaging to define how distinguishable a malignant tumor is from the adjacent healthy tissue. It is described as the signal intensity difference between the background tissue and the tumour, normalised by the background tissue's noise level. High contrast, as indicated by higher CNR values means better differentiate the tumor is from the surrounding tissue.

$$\text{CNR} = \frac{|\mu_t - \mu_b|}{\sigma_b} \quad (4.3)$$

where:

- μ_t is the mean signal intensity of the target (the object of interest),
- μ_b is the mean signal intensity of the background,
- σ_b is the standard deviation of the background signal (which represents the noise level).

4. Relative CNR: Utilization of relative CNR, the absolute difference between the real target image and a generated image as a metric of evaluation yields a fair assessment of the synthetic image quality. This method aids in evaluating the ability of the generated image to maintain a high degree of contrast of tumor relative to the background as compared with real images.

A low absolute difference in CNR suggests that the generated image is of good quality because it preserves the same degree of contrast between the tumour and the backdrop as the original image.

$$\text{relative CNR} = |\text{CNR}_{\text{real}} - \text{CNR}_{\text{generated}}| \quad (4.4)$$

where:

- CNR_{real} is the Contrast-to-Noise Ratio of the real target image,
- $\text{CNR}_{\text{generated}}$ is the Contrast-to-Noise Ratio of the generated image.

5. Histogram Visualization: A method for comparing the distribution of pixel intensity values between produced and real images is called histogram visualisation. One can visually compare the intensity distributions of two images by generating histograms that show the number of data points at each intensity level.

The histograms can be used visually to quickly identify variances in brightness, contrast, and render the dynamic range for difference of images which might create some area that not completely match with the originals, and also variations in the shape of the histograms could show errors or artifacts created due the generated images that may not be noticeable in first glance.

Chapter 5

Optimization Strategies

Various optimization process and strategies were performed to enhance the proposed model's capacity to generate MRI T2 sequences from T1 data. At the beginning of the model's development, a training dataset comprising 130 patients was utilized. However, an early finding revealed an overfitting tendency, with the model performing exceptionally well on training data but finding it difficult to generalize to new, unobserved data. An L2 regularisation method was used to address overfitting issues.

During the training phase, the study investigated the effects of replacing the conventional Binary Cross Entropy (BCE) loss function in GANs with Mean Squared Error (MSE) for the generator and discriminator. This investigation evaluated how the two loss functions affect the model's convergence behavior and overall performance.

The optimal training set size was explored systematically by changing the amount of the training dataset, going from 150 patients(23,250 images) to 338 patients(52,390 images), realizing the critical significance that data quantity plays in model training. The effects of these variations were determined by comparing the evaluation metrics on the test dataset.

The study investigated how the hyperparameter lambda (λ) determines the relative importance of L1 loss in the generator loss function throughout the training of our model. By carefully examining the test data evaluation metrics, we were able to assess the effect of methodically changing lambda values on the model's performance.

Then, the focus was shifted to how batch size affected overall performance. We carefully assessed the test dataset using important evaluation indicators, systematically changing the batch sizes during training.

This study incorporates three data augmentation methods to improve model performance: gamma correction, Gaussian blur, and random rotation. Techniques for augmenting data are essential for improving the training dataset's quality and the model's capacity to generalize outside of the observed data distribution. By adding variances to the training set, these methods expose the model to various cases, which reduces overfitting and increases resilience.

Random Rotation: Adding variances in object orientations into the training data set makes it possible for the model to learn the invariant properties of the objects from multiple perspectives. Imposing certain random rotations increases the robustness of the model against spatial distortions and improves its capability to identify objects from various viewpoints as it trains with enhanced data that comprises the rotated samples.

Gaussian Blur: We utilize the Gaussian blur filter to simulate imaging artefacts like motion blur or lens defects within our dataset, thereby enhancing its overall quality. Gaussian blurring, which is done by the convolutional operation of Gaussian kernels with different standard deviations, gives specific levels of blur. The imperfections that Gaussian blur adds to the model make it more robust against any visual distortions occurring on a faulty image, improving its performance.

Gamma Correction: Gamma correction modifies the pixel intensity levels to compensate for differences in illumination across various contexts. With gamma correction, we make the model available to the set of luminance levels that will help it to make better predictions in different light conditions. This leads to an increase in the model's capacity to generalize and perform excellently within diverse lighting situations.

To fully assess the effectiveness of these data augmentation methods, we perform a comparative study over a range of hyperparameters. We evaluate the effects of changing hyperparameters like gamma values, rotation angles, and blur magnitudes in a systematic way on metrics related to the performance of the model, such as accuracy and robustness.

In the beginning, the study was conducted through the process of data augmentation by applying that to some of the existing training samples. Subsequently, we have expanded this method by data generation through data augmentation to find out whether the model is capable of learning and generalization on unseen data given an enhanced number of training samples and epochs.

The study involves taking the Fourier transform to transform input images from the spatial to frequency domain then train the model to learn mapping between input and output images in the frequency domain. Transformation and reconstruction processes will be more effective by operating in the frequency domain which enables the model to capture the essential frequency characteristics and patterns.

The Fourier transform disregards spatial details and represents frequency patterns in the data. Therefore, models trained in the frequency domain may have greater robustness to spatial variations or deformations encountered in the input images. A benefit of this approach is that imaging can be done reliably and efficiently in a variety of settings and with different modalities.

To enhance local feature learning and robustness to structure, a patch-based learning approach model was also studied. Patch-based methods which focus on generating small patches or regions within an image rather than the image as a whole allow a more localized understanding of texture and other features, which can result in more realistic and detailed synthetic images.

Separate patches can be produced independently of one another and then combined to obtain a final image – this provides additional flexibility for the final image's appearance.

The problem of synthetic image generation entails the presence of data in different conditions – dark regions, light regions, and tumor patches. Patch-based methods are more suitable for such variability since they concentrate on local details of the image and generate them for any particular body part with higher accuracy.

Chapter 6

Results

The evaluation has been carried out by considering 200 images in the test dataset. Parameters Used during Training: Adam optimizer for both generator and discriminator with learning rate=2e-4,betas=(0.5, 0.999) is used with batch size=10, lambda=100 and Training data =130 patients

When training with only 130 patients data the pix2pix model was overfitting to the training dataset and was not giving good results on the test dataset. To overcome this L2 regularization was used and the results obtained on the test dataset were compared.

Table 6.1: RMSE Error on Test Images Using 130 Patients Data for Training with and without Regularizer

| | With Regularization | Without Regularization |
|------|---------------------|------------------------|
| RMSE | 55.04 | 54.23 |

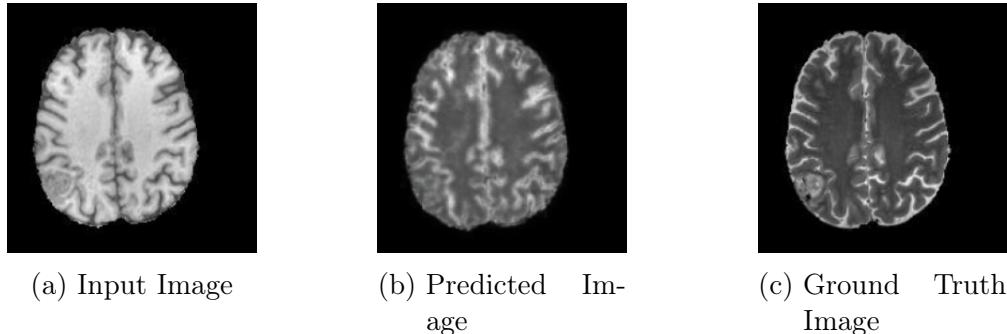


Figure 6.1: Generated Image With Regularization

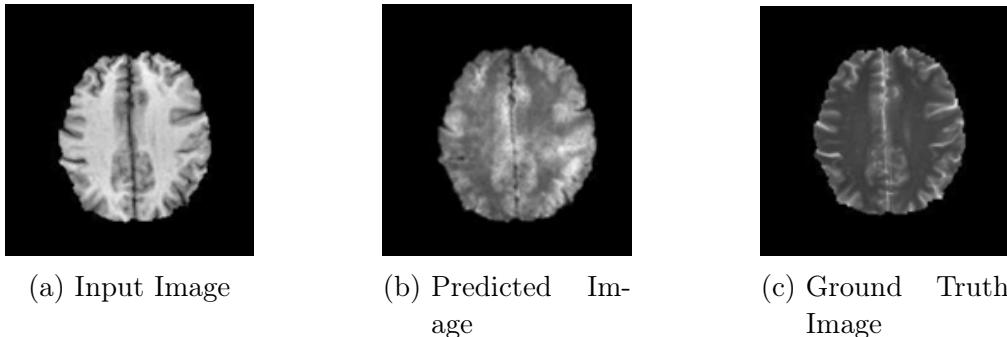


Figure 6.2: Generated Image Without Regularization

The above fig and table show that there was no benefit in using regularization the performance of the model was decreasing only.

To overcome the issue of overfitting and get better test results the pix2pix model was trained on a whole dataset of 338 patients (52,390 images).

The performance of the two pix2pix models that use BCE and MSE loss in calculating the generator loss and discriminator loss respectively are compared on the basis of evaluation metrics of the test dataset. The above fig and table show pix2pix model that uses MSE

Table 6.2: RMSE Error and SSI on Test Images Using 338 Patients Data for Training with BCE and MSE loss

| | BCE | MSE |
|------|-------|-------|
| RMSE | 36.10 | 34.56 |
| SSI | 0.911 | 0.917 |

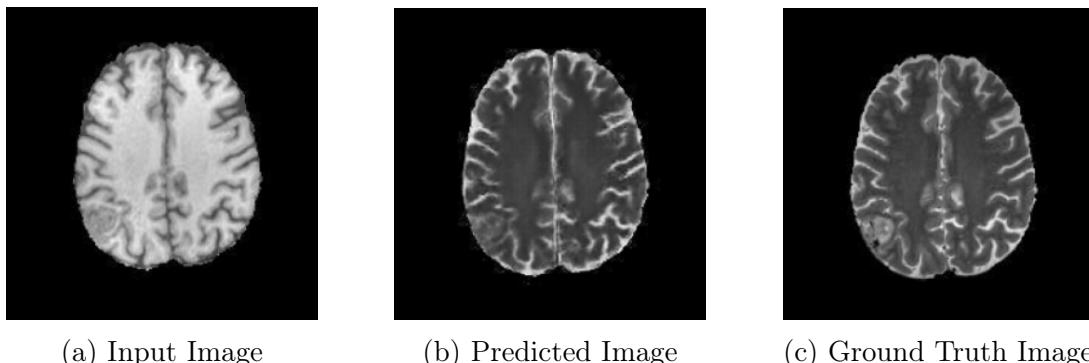


Figure 6.3: Generated Image when trained with BCE loss function

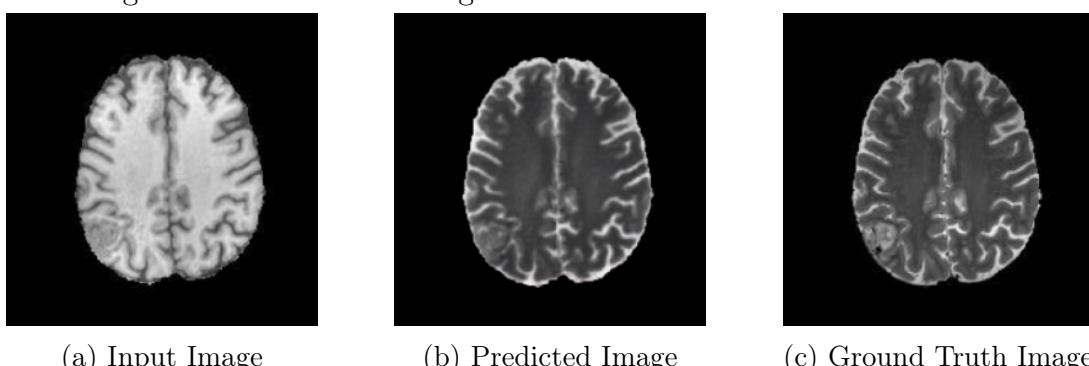


Figure 6.4: Generated Image when trained with MSE loss function

performs better than the model that uses BCE. So further analysis has been carried out only on the model that uses MSE loss.

The optimal training set size has been explored and the performance of the model has been compared on the test dataset.

Table 6.3: RMSE and SSIM on Test Images Using Different Numbers of Patients for Training

| No. of Patients | RMSE | SSI |
|-----------------|-------|-------|
| 150 | 36.44 | 0.915 |
| 200 | 34.08 | 0.917 |
| 250 | 34.12 | 0.920 |
| 300 | 34.92 | 0.913 |
| 338 | 36.34 | 0.911 |

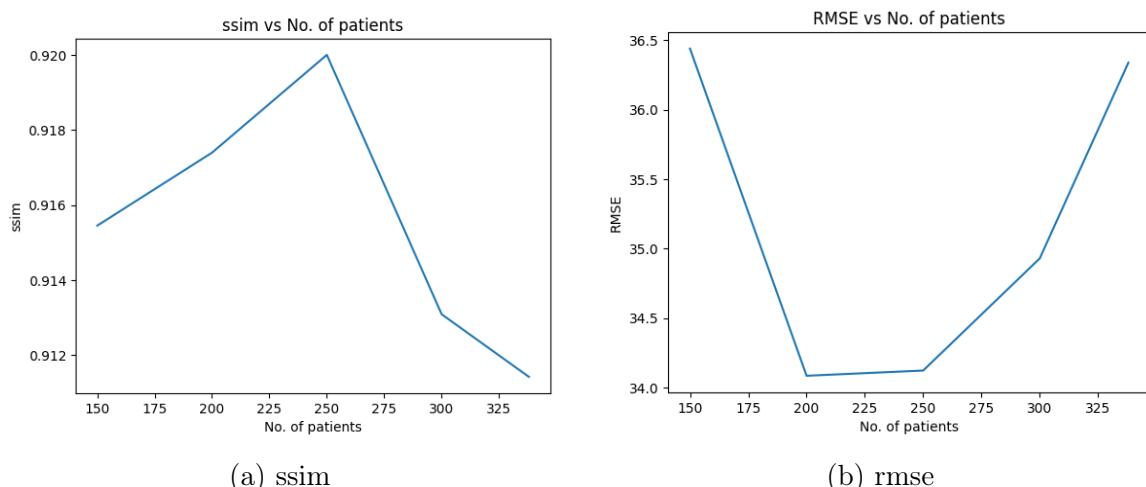


Figure 6.5: RMSE and SSI with varying no of patients data for training.

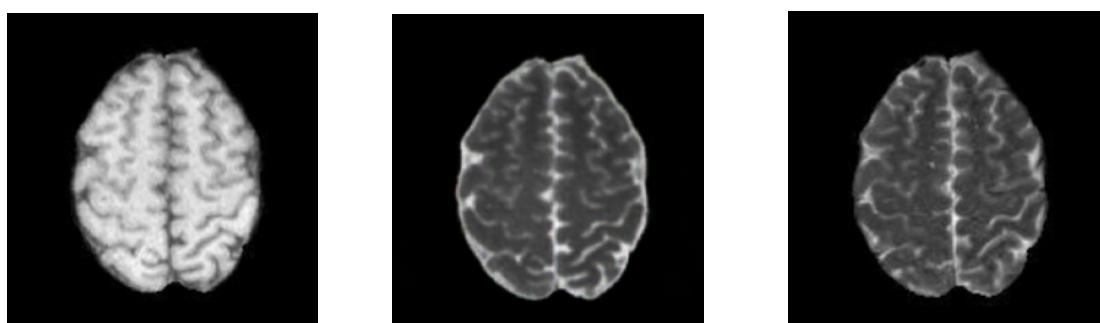


Figure 6.6: Generated Images trained with 150 patients

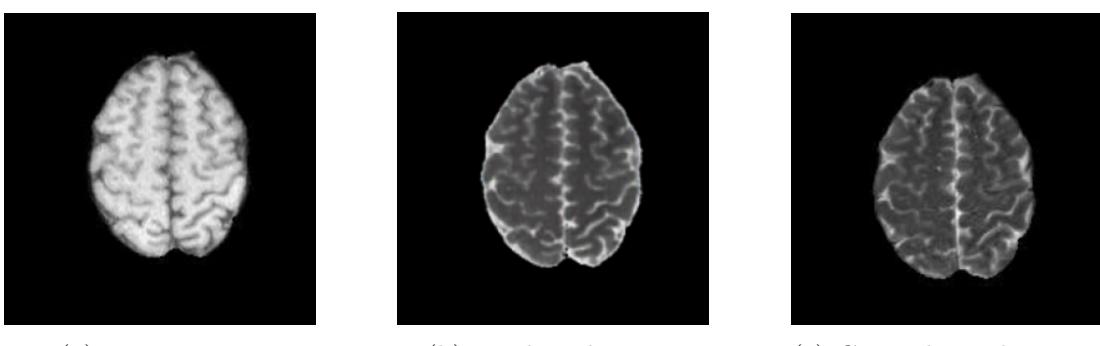


Figure 6.7. Gaseous and liquid water in equilibrium at 200 °C.

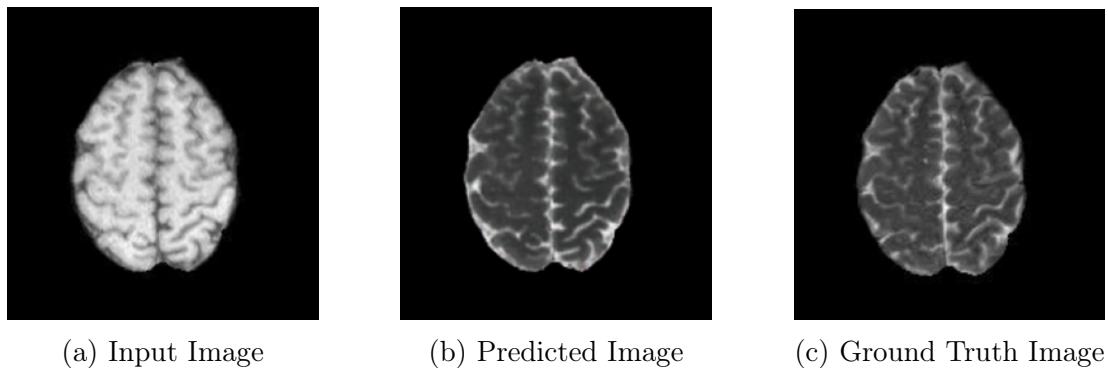


Figure 6.8: Generated Images trained with 250 patients

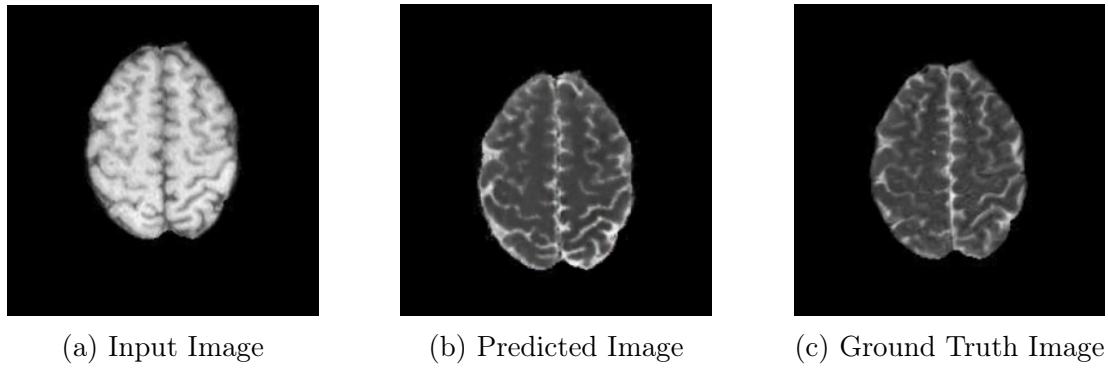


Figure 6.9: Generated Images trained with 300 patients

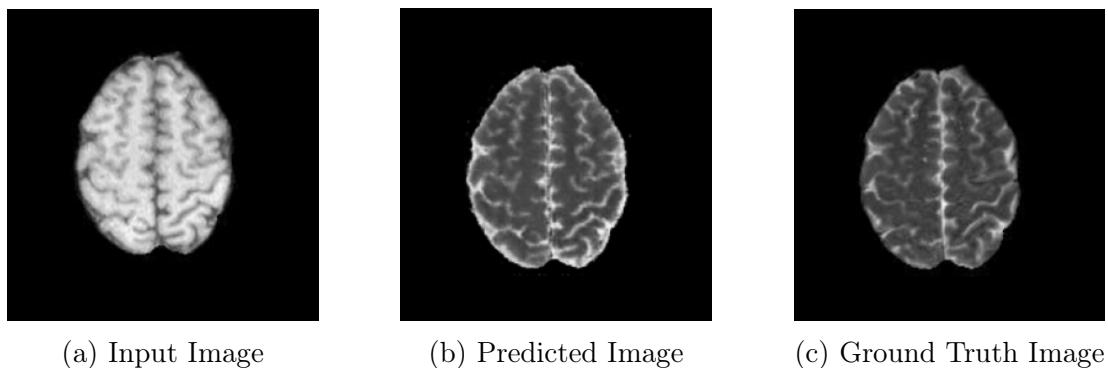


Figure 6.10: Generated Images trained with 338 patients

The above fig and table show the pix2pix model trained with 250 images gives the best results. So further analysis has been carried out only on the model trained on 250 patient data. The optimal value of lambda has been explored and the performance of the model has been compared on the test dataset.

Table 6.4: RMSE and SSIM on Test Images Using 250 Patients for Training and varying lambda

| lambda | RMSE | SSI |
|--------|-------|-------|
| 100 | 34.12 | 0.920 |
| 500 | 33.73 | 0.916 |
| 1000 | 33.93 | 0.923 |
| 1500 | 35.68 | 0.912 |

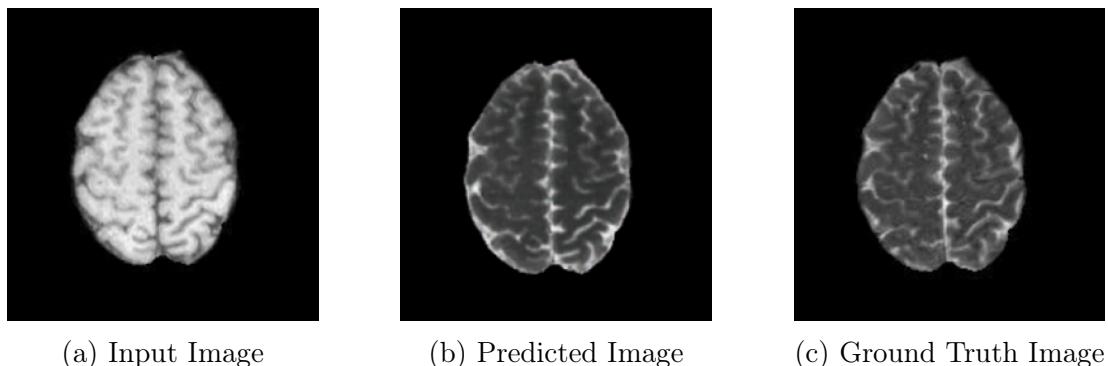


Figure 6.11: Generated Images at $\lambda = 100$

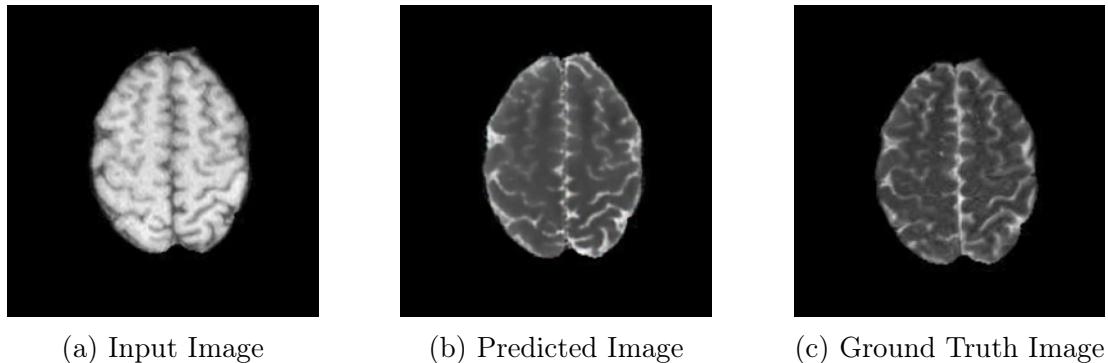


Figure 6.12: Generated Images at $\lambda = 500$

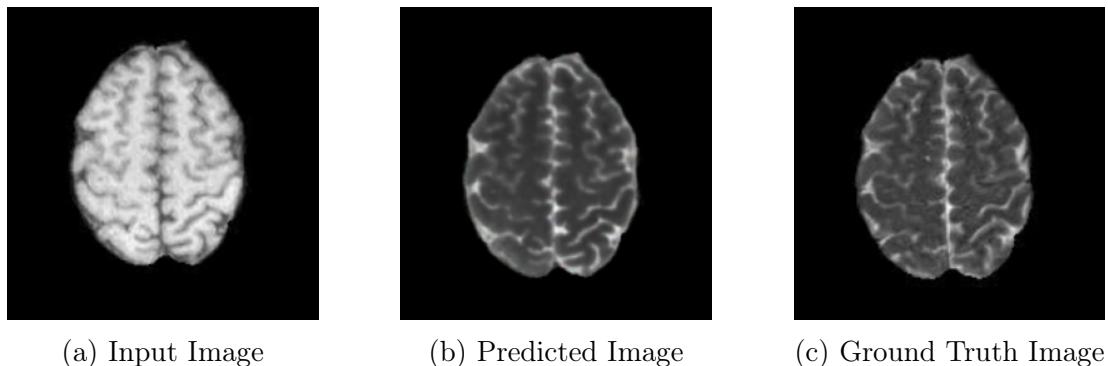


Figure 6.13: Generated Images at $\lambda = 1000$

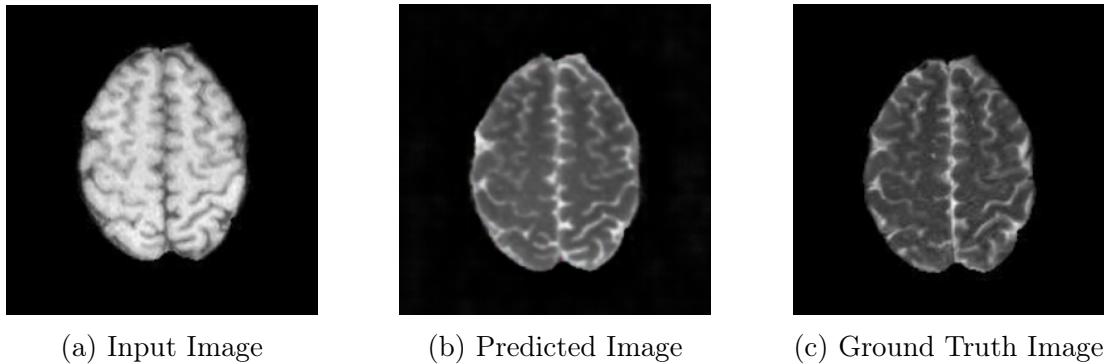


Figure 6.14: Generated Images at $\lambda = 1500$

The above fig and table show the pix2pix model trained with lambda value 0f 500 gives the best results. So further analysis has been carried out only on the model with lambda equal to 500.

The optimal value of batch size during training has been explored and the performance of the model has been compared to the test dataset.

Table 6.5: RMSE and SSIM on Test Images Using 250 Patients for Training and Varying Batch Size

| Batch Size | RMSE | SSI |
|------------|-------|-------|
| 30 | 33.72 | 0.917 |
| 50 | 33.73 | 0.916 |
| 80 | 35.75 | 0.915 |
| 100 | 35.18 | 0.914 |
| 120 | 35.63 | 0.913 |

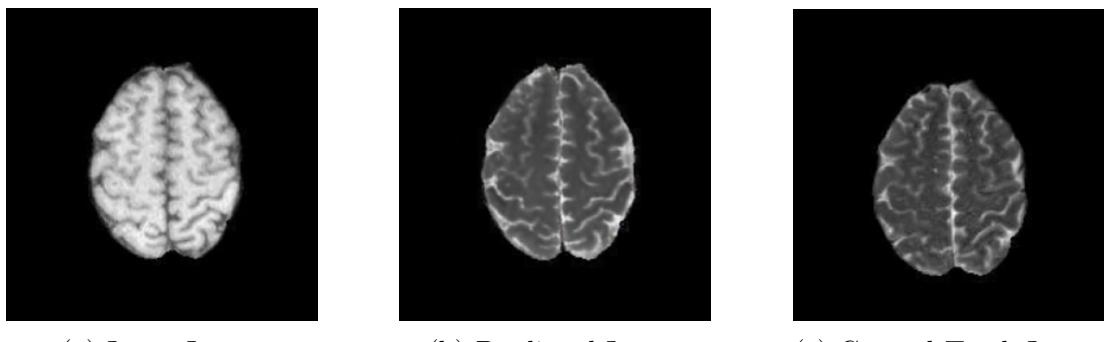


Figure 6.15: Generated Images at batch size = 30

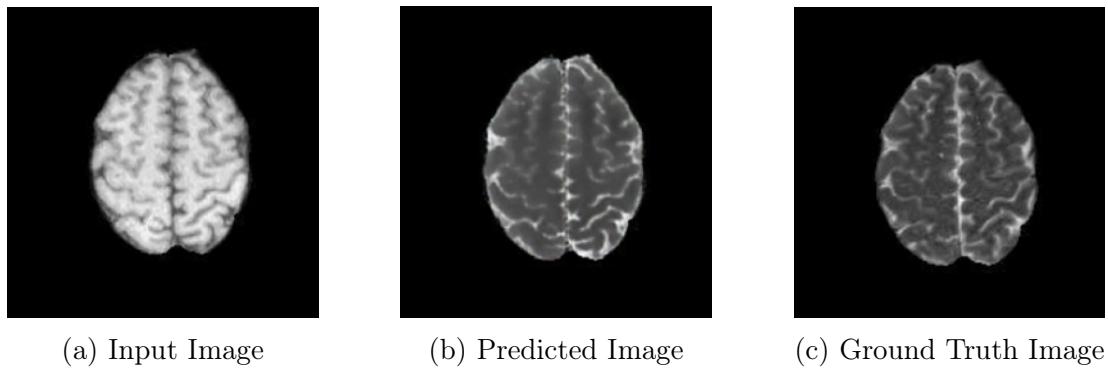


Figure 6.16: Generated Images at batch size = 50

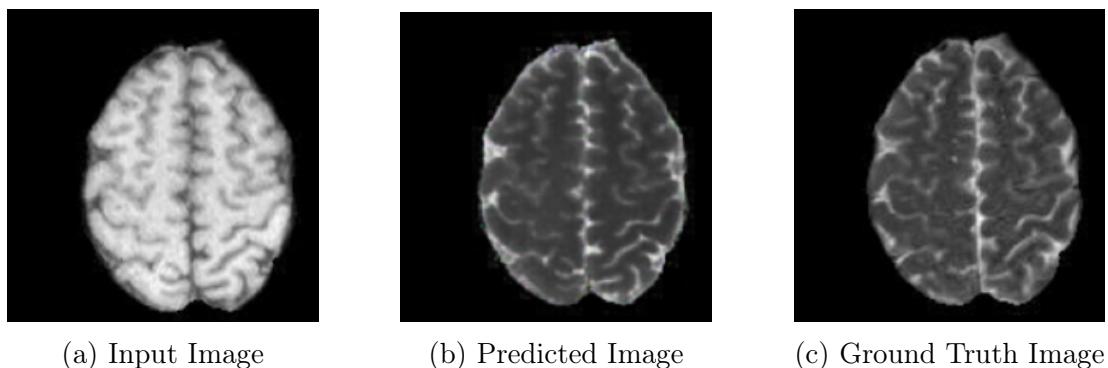


Figure 6.17: Generated Images at batch size = 80

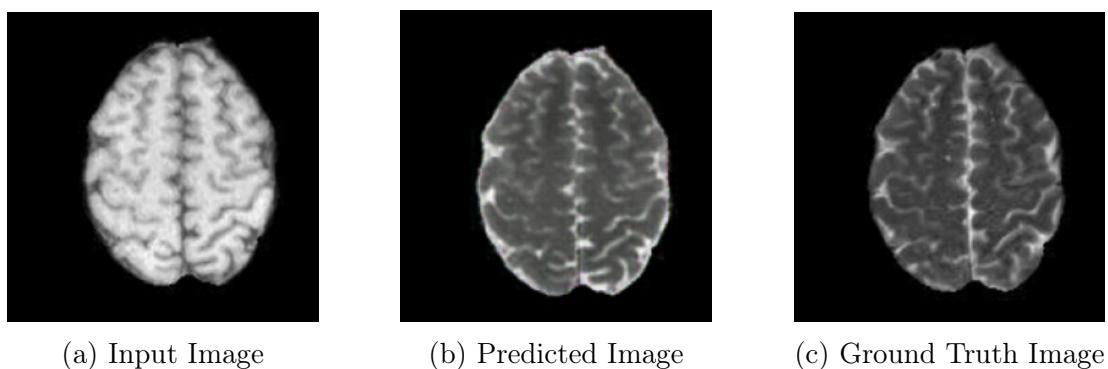


Figure 6.18: Generated Images at batch size = 100

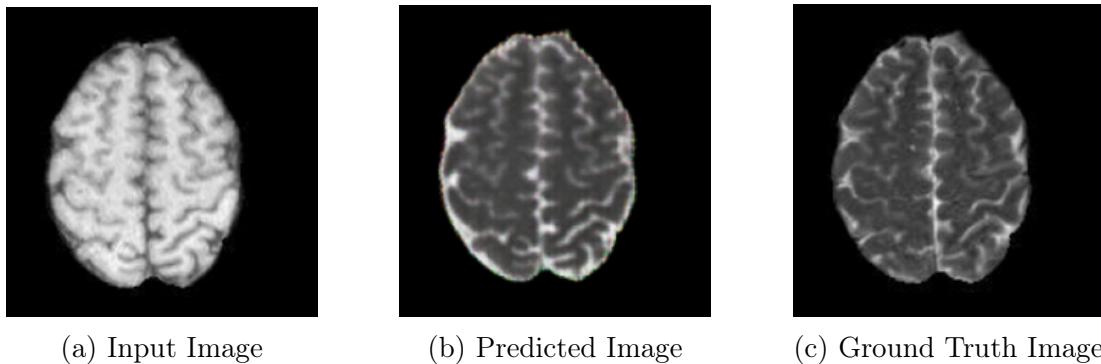


Figure 6.19: Generated Images at batch size = 120

The above figure and table show the pix2pix model trained with a batch size equal to 30 gives the best results.

The best results of the model pix2pix when evaluated on the test dataset were obtained at batch size = 30, lambda = 500, training data = 250 patients, and generator loss and discriminator loss when calculated using MSE.

Table 6.6: Comparison of Model Performance: Optimal Dataset vs. Full Dataset with Data Augmentation

| Metric | Full Dataset | 250 Patients |
|----------------|--------------|--------------|
| RMSE Error | 40.83 | 31.48 |
| Max RMSE Error | 107.24 | 114.79 |
| Min RMSE Error | 22.27 | 0.0003 |
| Mean SSIM | 0.88 | 0.92 |
| Min SSIM | 0.703 | 0.69 |
| Max SSIM | 0.95 | 1.00 |

Initially, the optimal training data was 250 patients. However, when used data augmentation the model performed better when trained with whole dataset (368 patients).

Table 6.7: RMSE and SSIM on Test Images Using Different Augmentation Techniques

| Metric | Gaussian Blur (kernel=11, $\sigma=5$) | Gaussian Blur (kernel=29, $\sigma=11$) | Gamma ($\gamma=0.8$, gain=1) | Gamma ($\gamma=1.2$, gain=1) | Gaussian Blur and Gamma kernel=29, $\sigma=11$ $\gamma=0.8$, gain=1 |
|-----------------|---|--|-----------------------------------|-----------------------------------|--|
| Mean RMSE Error | 33.09 | 31.48 | 32.93 | 33.69 | 33.29 |
| Max RMSE Error | 113.18 | 114.79 | 117.85 | 124.75 | 120.63 |
| Min RMSE Error | 0.401 | 0.0003 | 0.0011 | 1.70 | 1.53 |
| Mean SSIM | 0.918 | 0.92 | 0.920 | 0.917 | 0.92 |
| Min SSIM | 0.69 | 0.69 | 0.69 | 0.692 | 0.70 |
| Max SSIM | 0.99 | 1.00 | 1.00 | 0.998 | 0.99 |

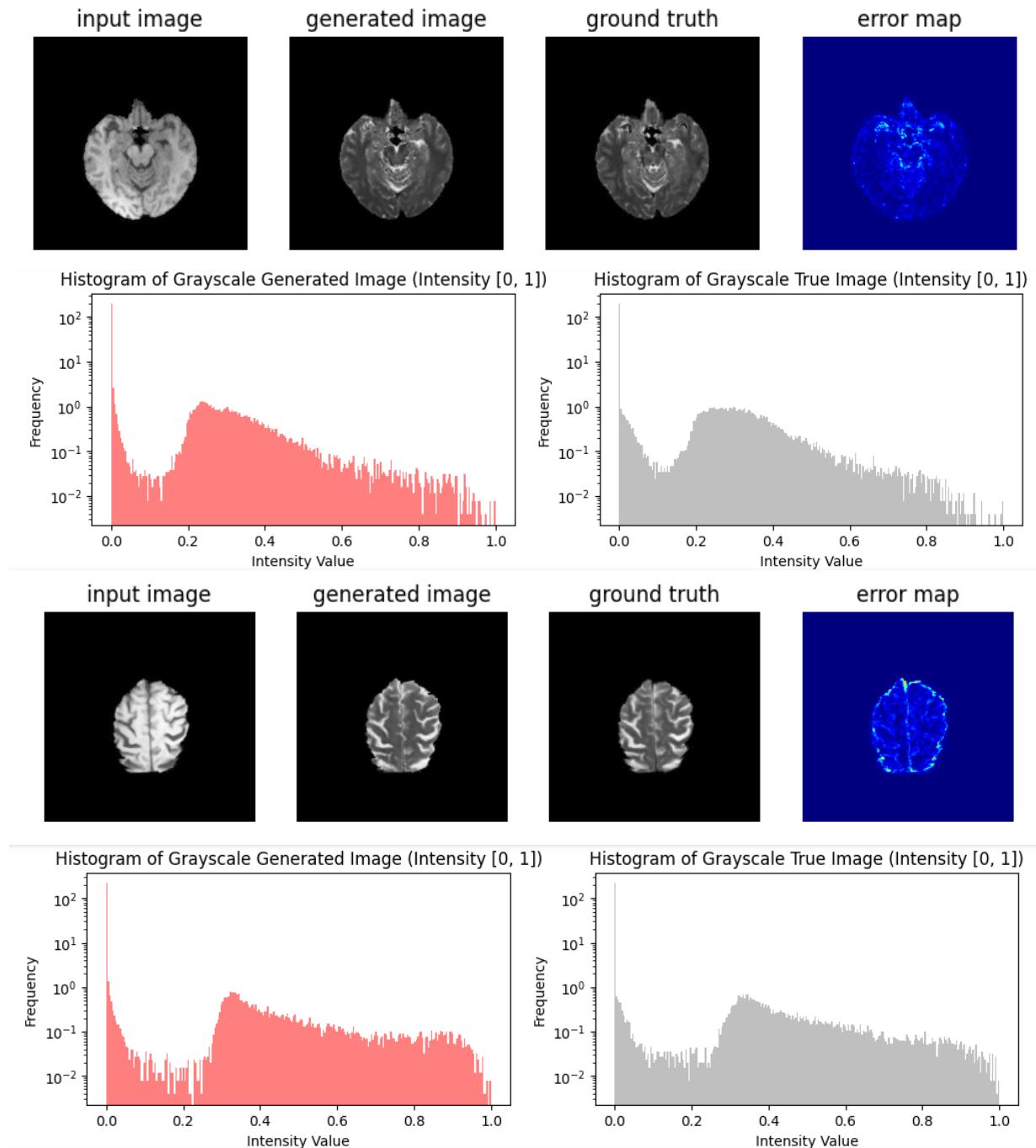


Figure 6.20: Generated Image when used Gaussian blur with kernel=29 and $\sigma=11$

The above table 6.7 shows that when used gaussian blur with kernel size =29 and $\sigma=11$ as data augmentation techniques performed better than other settings. The above figures are the results obtained using the same technique. Then the same setting was tried with different ranges of rotations using the Random Rotation function and the performance was evaluated.

Table 6.8: RMSE and SSIM on Test Images using different Random Rotations in with Range of $-\theta$ to θ

| Metric | $\theta=[-10,+10]$ | $\theta=[-20,+20]$ | $\theta=[-30,+30]$ | $\theta=[-40,+40]$ | $\theta=[-45,+45]$ | $\theta=[-50,+50]$ |
|-----------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Mean RMSE Error | 69.71 | 34.97 | 33.52 | 32.90 | 32.25 | 33.37 |
| Max RMSE Error | 137.87 | 125.01 | 123.93 | 110.95 | 141.70 | 122.32 |
| Min RMSE Error | 33.28 | 5.9e-05 | 0.001 | 0.83 | 0.0004 | 6.3e-05 |
| Mean SSIM | 0.762 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 |
| Min SSIM | 0.513 | 0.66 | 0.65 | 0.64 | 0.64 | 0.65 |
| Max SSIM | 0.913 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 |

Table 6.9: Comparing RMSE and SSIM on Test Images with and without vertical flip as Augmentation Technique

| Metric | $\theta=[-45,+45]$ | $\theta=[-45,+45]$ with vertical flip |
|-----------------|--------------------|--|
| Mean RMSE Error | 32.25 | 37.26 |
| Max RMSE Error | 141.7 | 134.91 |
| Min RMSE Error | 0.0004 | 6.98 |
| Mean SSIM | 0.92 | 0.907 |
| Min SSIM | 0.64 | 0.613 |
| Max SSIM | 1.00 | 0.99 |

The best setting for my model was found when used Gaussian blur with a random rotation function with the angle of rotation ranging from -45 to 45. The effect of adding a random vertical flip function was also explored but the performance of the model decreased and hence removed from further analysis.

Table 6.10: RMSE and SSIM on Test Images for Different Epochs with range of rotation from -45 to 45 degree

| Metric | $\theta=[-45,+45]$ and 10 epochs | $\theta=[-45,+45]$ and 20 epochs | $\theta=[-45,+45]$ and 30 epochs |
|-----------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Mean RMSE Error | 32.25 | 31.86 | 33.5 |
| Max RMSE Error | 141.7 | 122.81 | 152.72 |
| Min RMSE Error | 0.0004 | 0.06 | 0.013 |
| Mean SSIM | 0.92 | 0.92 | 0.92 |
| Min SSIM | 0.64 | 0.64 | 0.65 |
| Max SSIM | 1.00 | 1.00 | 1.00 |

Table 6.11: RMSE and SSIM on Test Images for Different Learning Rates and Epochs with range of rotation from -45 to 45 degree

| Metric | $\theta = [-45, +45]$ | $\theta = [-45, +45],$ 20 epochs | $\theta = [-45, +45],$ 10 epochs, changed lr | $\theta = [-45, +45],$ 20 epochs, changed lr |
|-----------------|-----------------------|-------------------------------------|---|---|
| | 20 epochs | 10 epochs, changed lr | 20 epochs, changed lr | |
| Mean RMSE Error | 31.86 | 32.48 | | 121.82 |
| Max RMSE Error | 122.81 | 125.82 | | 200.87 |
| Min RMSE Error | 0.06 | 0.010 | | 68.206 |
| Mean SSIM | 0.92 | 0.92 | | 0.68 |
| Min SSIM | 0.64 | 0.64 | | 0.45 |
| Max SSIM | 1.00 | 1.00 | | 0.80 |

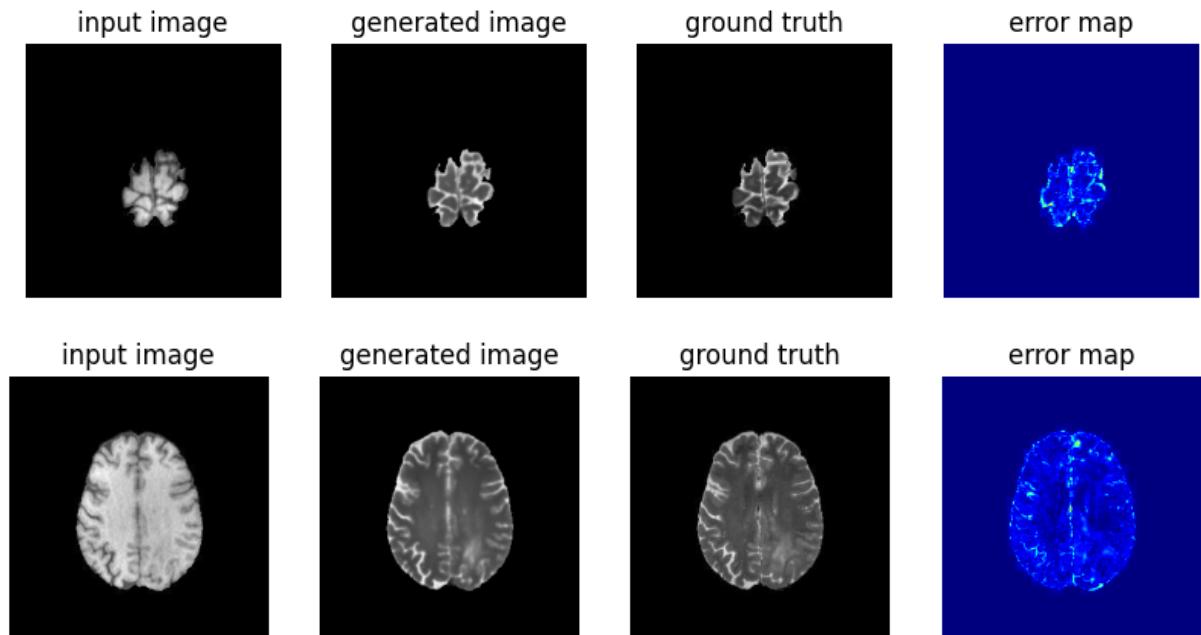


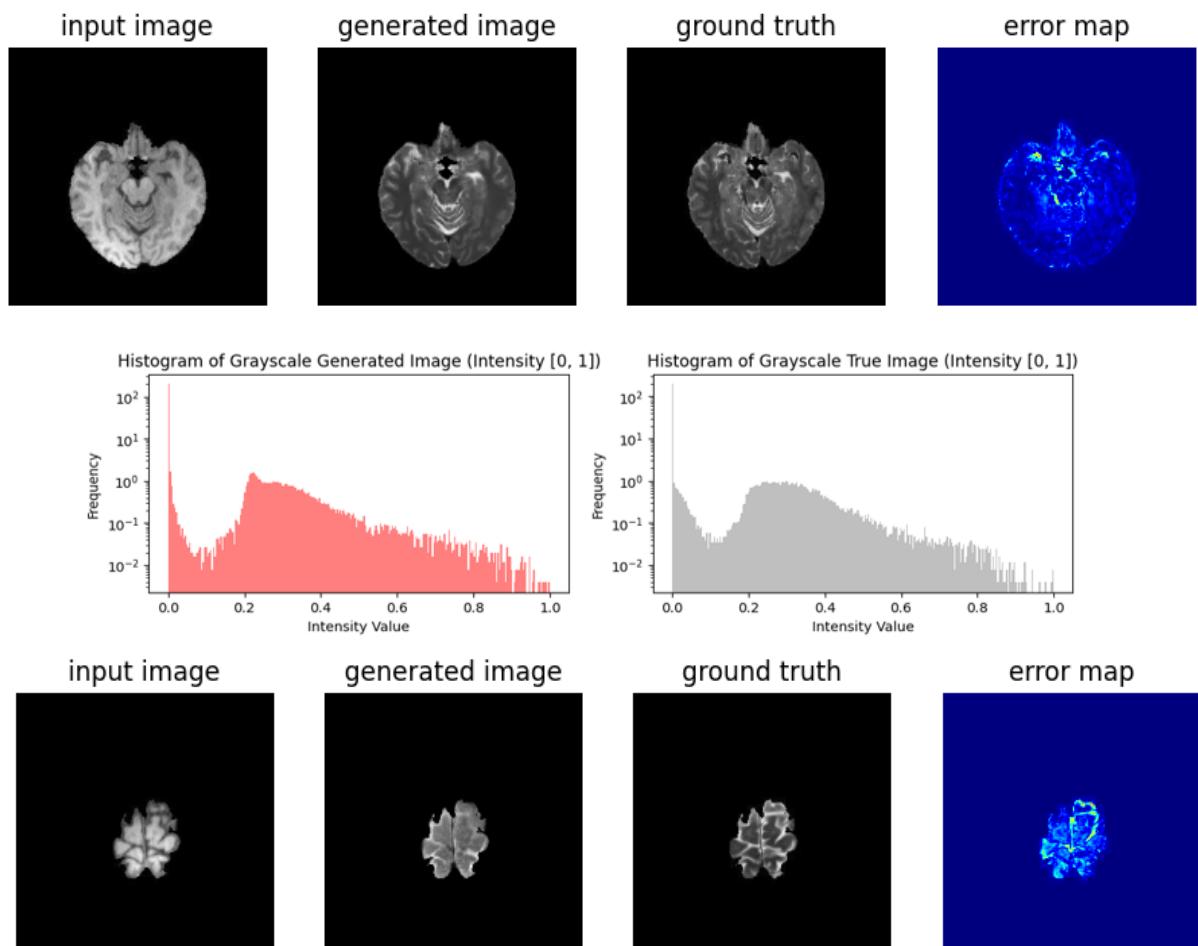
Figure 6.21: Generated Image at $\theta = [-45, +45]$ and 20 epochs

The above tables 6.10 and 6.11 shows the performance of the model with varying epochs and the effect of changing learning rate of the generator to 1e-4 and the discriminator to 5e-5. The best settings were obtained with 20 epochs and with learning rate as 2e-4 for both generator and discriminator. The above pics are the results generated at that setting.

10,478 synthetic data images were generated using the data augmentation techniques to increase the training samples from 52,390 images to 62,868 images and the performance of the model was evaluated at different epochs.

Table 6.12: RMSE and SSIM on Test Images Using Synthetic Data Generated with Augmentation Techniques at Different Epochs

| Metric | With synthetic data (10 ep) | With synthetic data (20 ep) | With synthetic data (30 ep) | With synthetic data (50 ep) |
|-----------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| Mean RMSE Error | 327.26 | 108.87 | 31.07 | 33.27 |
| Max RMSE Error | 446.22 | 389.73 | 140.00 | 122.52 |
| Min RMSE Error | 112.63 | 0.0001 | 0.012 | 0.00057 |
| Mean SSIM | 0.256 | 0.77 | 0.925 | 0.914 |
| Min SSIM | -0.02 | 0.091 | 0.64 | 0.65 |
| Max SSIM | 0.72 | 1.00 | 1.00 | 1.00 |



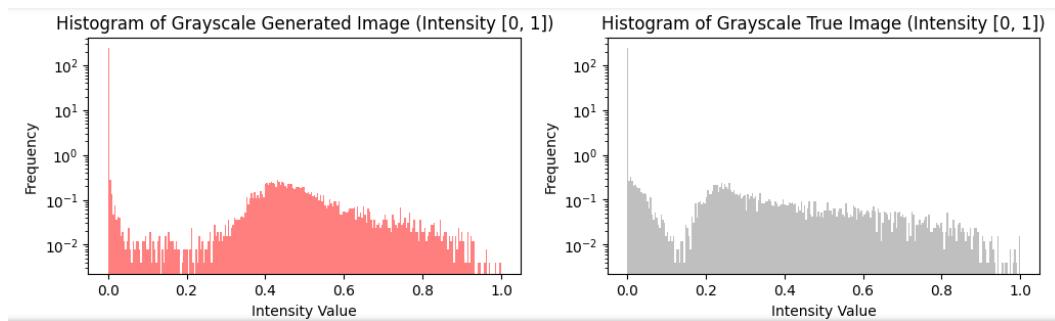
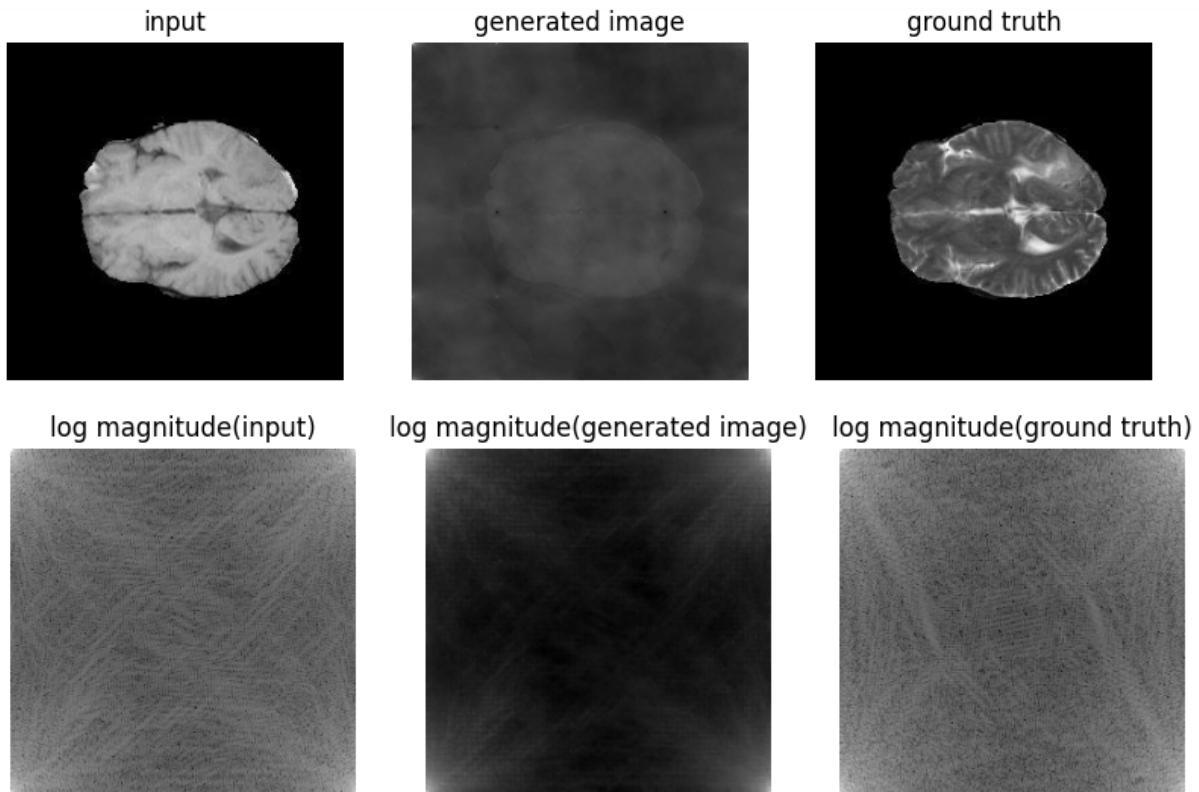


Figure 6.22: Generated Images when trained along with synthetic data at 30 epochs

The optimal setting for the model was found at 30 epochs and the above figures are the results obtained at that setting.

Table 6.13: RMSE and SSIM on Test Images Using Fourier Learning

| Metric | Fourier Learning |
|-----------------|------------------|
| Mean RMSE Error | 25250.95 |
| Max RMSE Error | 115764.20 |
| Min RMSE Error | 264.089 |



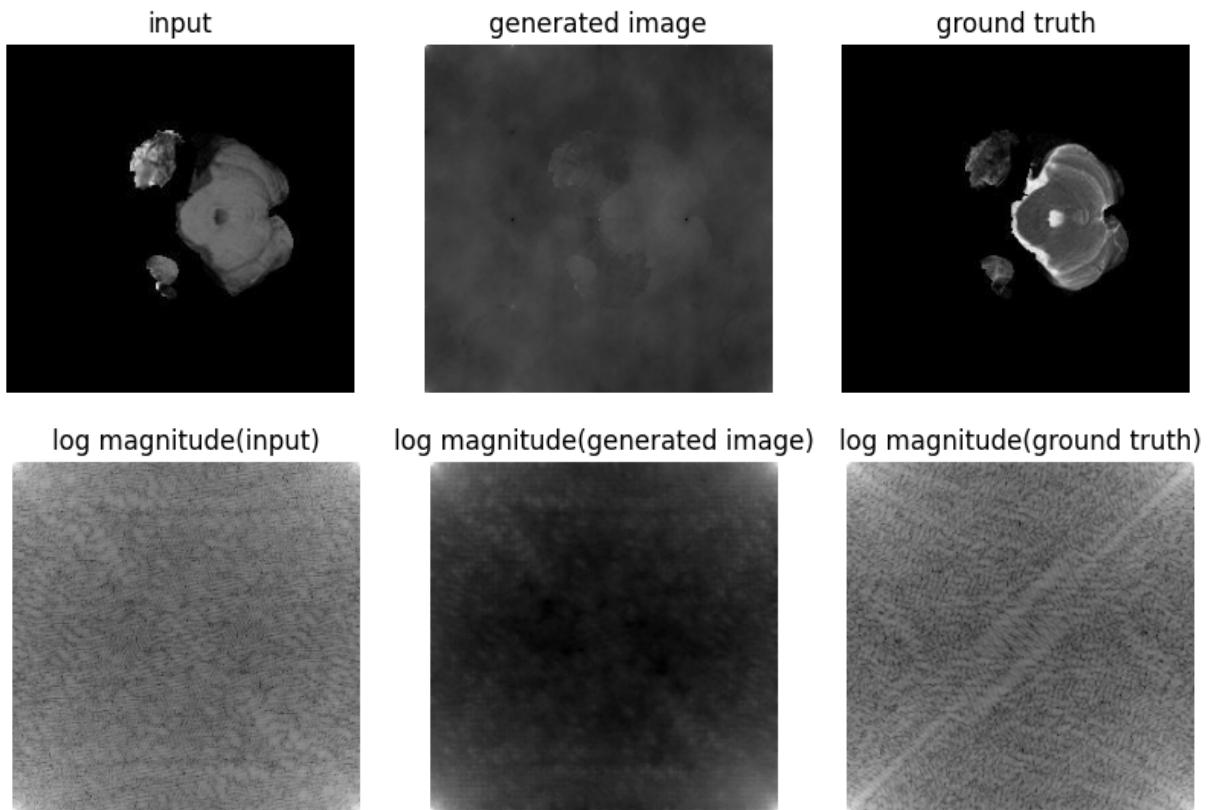


Figure 6.23: Generated image and log magnitude when trained using Fourier learning approach

Table 6.13 and the above figure are the results obtained when the model was trained to learn Fourier to Fourier mapping. The results obtained were not satisfactory Hence the approach was not considered for further analysis.

Table 6.14: RMSE and SSIM on Test Images Using Patch-Based Learning Approach

| Metric | Value |
|-----------------|--------|
| Mean RMSE error | 35.40 |
| Max RMSE error | 127.57 |
| Min RMSE error | 0.48 |
| Mean SSIM | 0.92 |
| Max SSIM | 0.99 |
| Min SSIM | 0.63 |

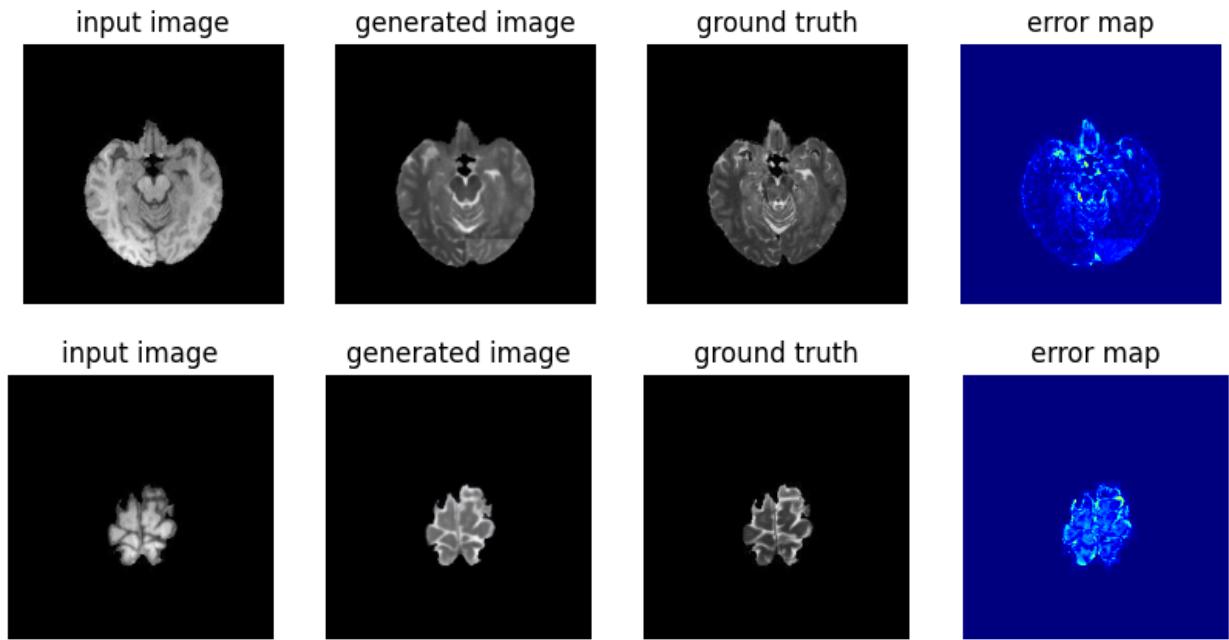


Figure 6.24: Generated image when trained using Patch-Based learning approach

Table 6.14 and the above figure are the results obtained when the patch-based approach was considered. The results obtained are satisfactory giving more future possibilities and applications.

Table 6.15: Comparison Of Relative CNR Models Trained On Full Image Approach and Patch Image Approach

| Metric | Optimal model based on Whole Image Approach | Model based on Patch-Based Learning Approach |
|--------------|--|---|
| Relative CNR | 0.118 | 0.201 |

Table 6.15 shows the comparison of relative CNR between the two approaches. The results indicating both approaches were able to create synthetic images with tumors having the same proper contrast as observed in real images.

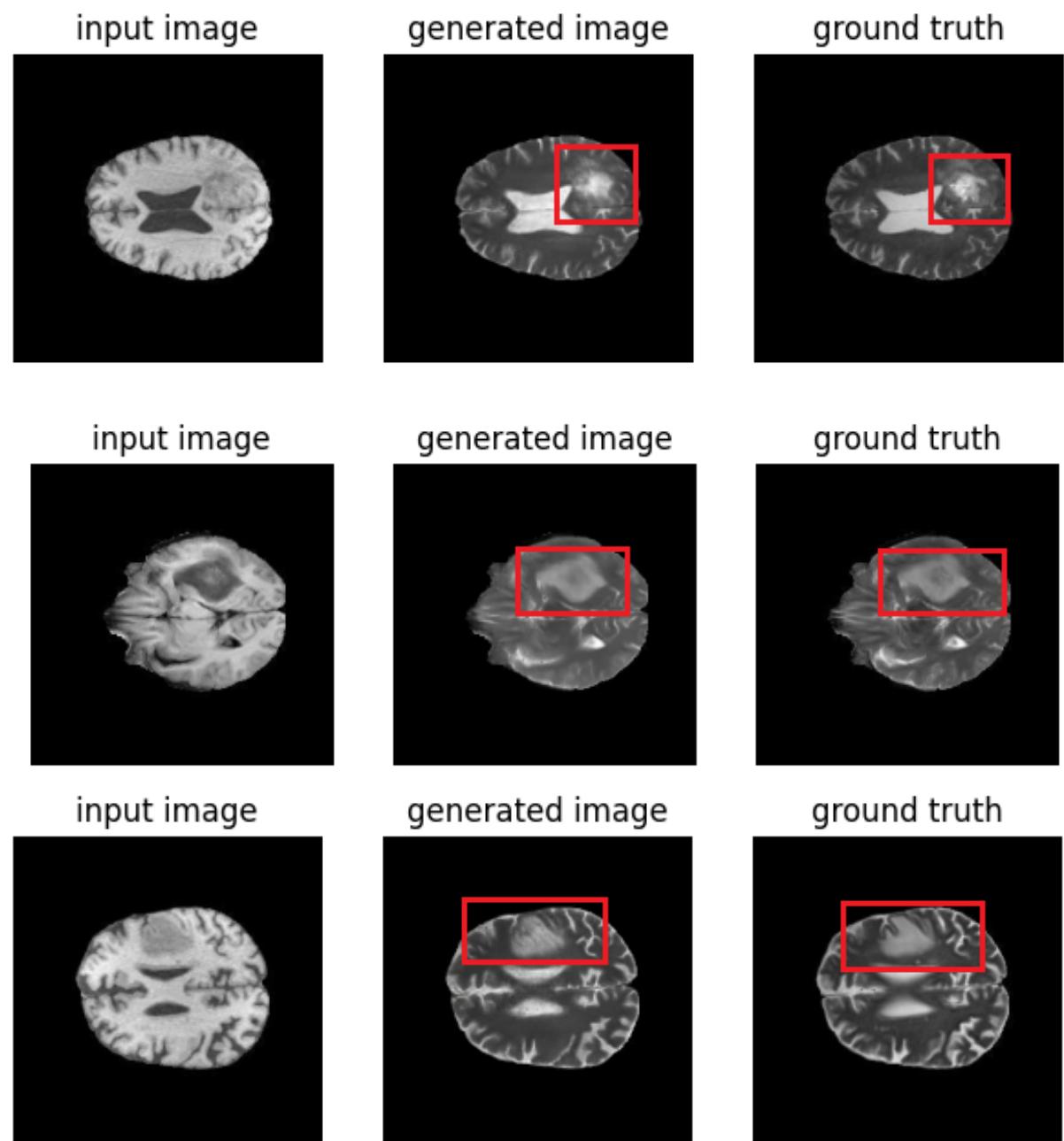
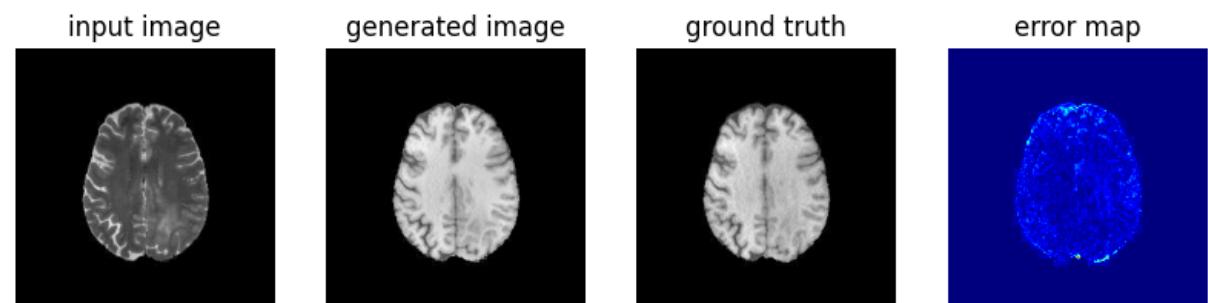
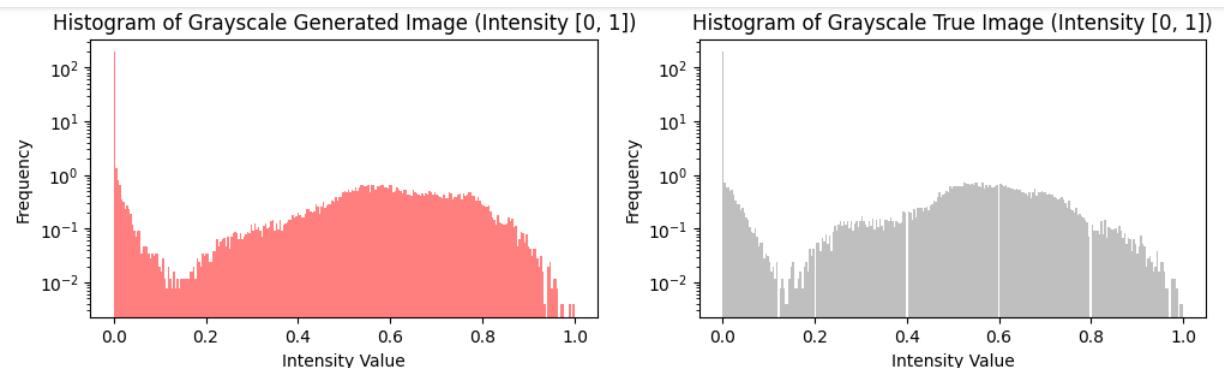
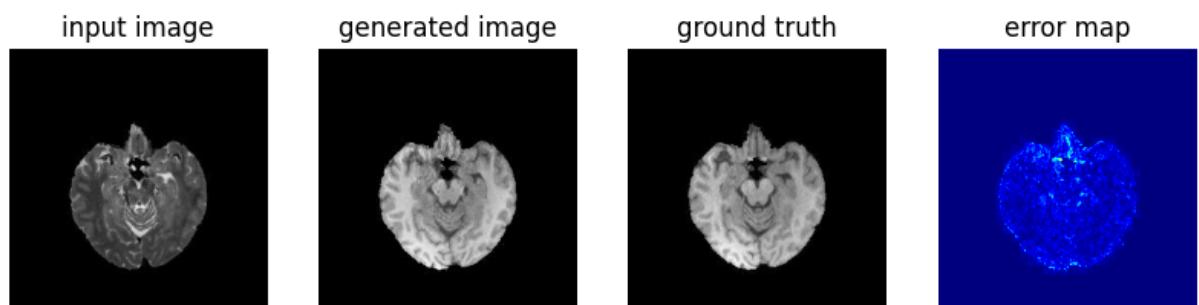


Figure 6.25: Segmented Tumor Area in Real and Generated Images

Results for T1 synthesis using T2 images :

Table 6.16: Model trained by adding synthetic data generated using augmentation technique for 50 epochs (T2 to T1)

| Metric | Value |
|-----------------|--------|
| Mean RMSE error | 38.16 |
| Max RMSE error | 261.14 |
| Min RMSE error | 0.0006 |
| Mean SSIM | 0.90 |
| Max SSIM | 1 |
| Min SSIM | 0.47 |



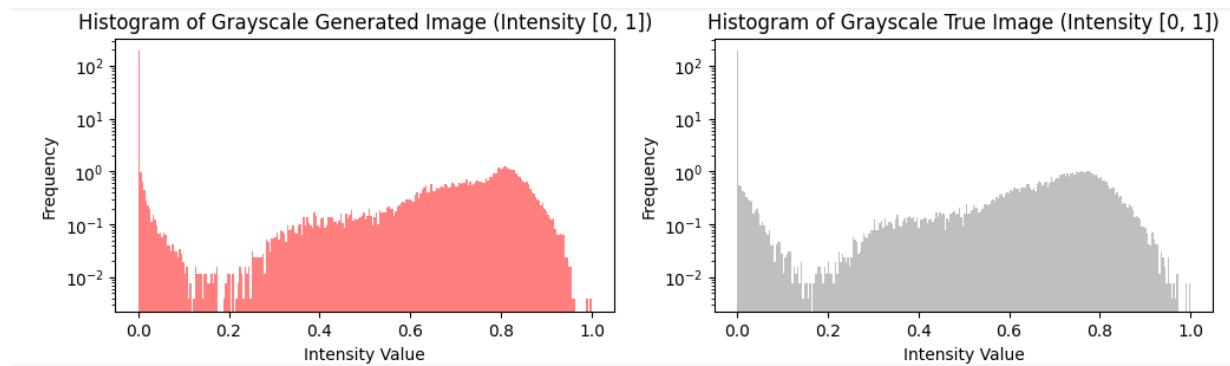
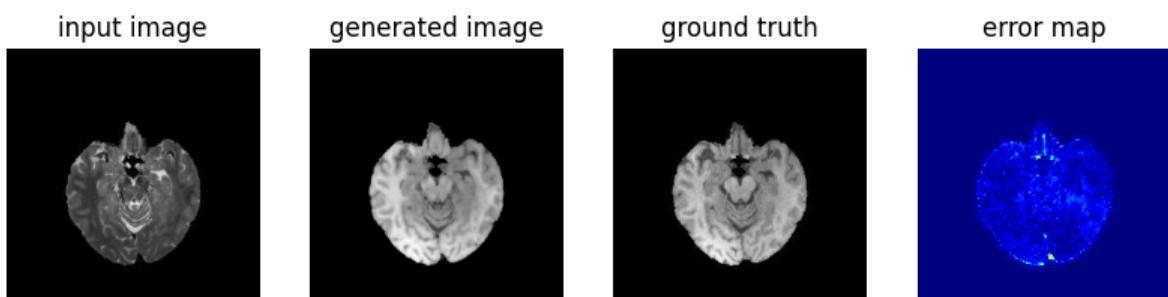


Figure 6.26: Generated image when trained along with synthetic data at 50 epochs(T2 to T1)

Table 6.16 and the above figure are the results obtained for T1 sequence generation from the T2 sequence when the model was trained along with synthetic data generated using data augmentation technique for 50 epochs.

Table 6.17: RMSE and SSIM on Test Images Using Patch-Based Learning Approach(T2 to T1)

| Metric | Value |
|-----------------|--------|
| Mean RMSE error | 40.76 |
| Max RMSE error | 145.83 |
| Min RMSE error | 0.32 |
| Mean SSIM | 0.90 |
| Max SSIM | 0.99 |
| Min SSIM | 0.60 |



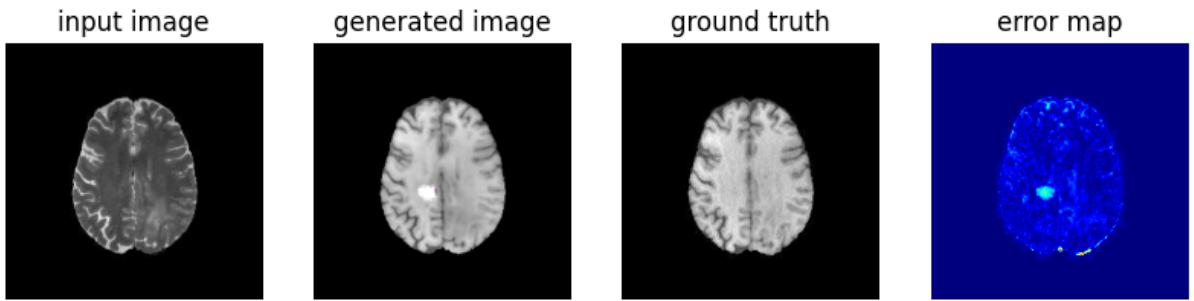


Figure 6.27: Generated Images when used patch based learning approach (T2 to T1)

Table 6.17 and the above figure are the results obtained when the patch-based approach was considered. The results obtained are satisfactory giving more future possibilities and applications.

Table 6.18: Comparison Of Relative CNR Models Trained On Full Image Approach and Patch Image Approach(T2 to T1)

| Metric | Optimal model based on Whole Image Approach | Model based on Patch-Based Learning Approach |
|--------------|---|--|
| Relative CNR | 0.087 | 0.138 |

Table 6.18 shows the comparison of relative CNR between the two approaches. The results indicating both approaches were able to create synthetic images with tumors having the same proper contrast as observed in real images.

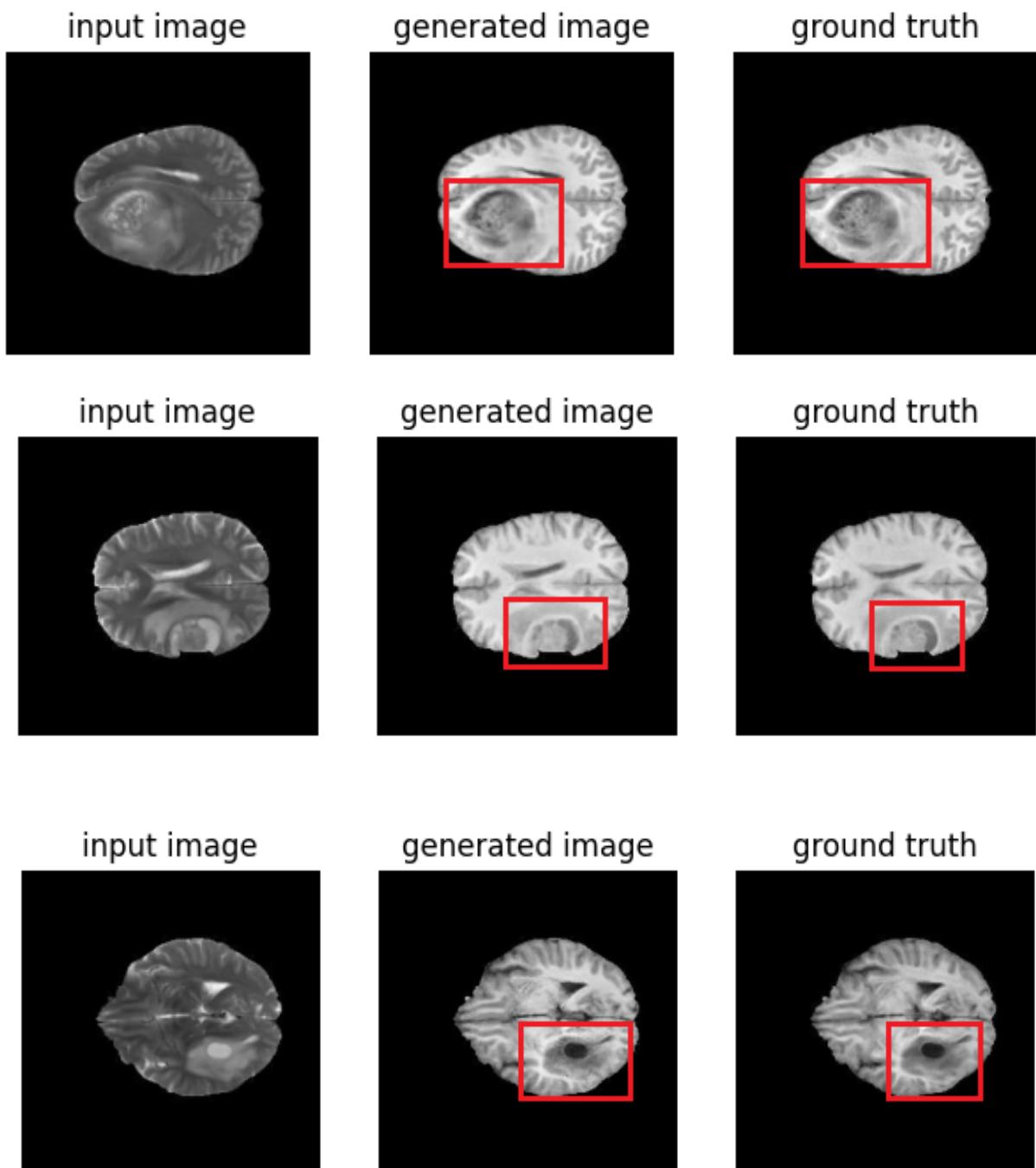


Figure 6.28: Segmented Tumor Area in Real and Generated Images

Chapter 7

Discussion and Conclusion

7.1 Discussion :

Similar work for generating MRI sequences were observed in the literature. In few studies multi-modal image translation tasks was mentioned where using multiple input sequence were used and translated into a single output sequence. Few studies also focused on single-modal image translation tasks in which by adding small changes in the model architecture of pix2pix they tried to improve the model performance.

Few key highlights from the existing studies are:

- [8] The U-net architecture of pix2pix generator was replaced by U-Net ++ architecture which is a bigger network than the previous one with more skipped connections in between and more layers than the previous one in the hope of learning more contextual information.
- [8] One study used differential image discriminators which try to differentiate the real difference between the input image and the real target image and the fake difference between the image and the generated target image. This approach offer advantages such as encouraging the generator to focus on capturing specific details or characteristics of the target domain.
- One study also suggested adding MI(Mutual Information) and GD (Gradient Difference) in the loss function to overcome misalignment and for structural accuracy.

Few key differences between this study and the existing studies are:

- **BCE vs MSE Loss Comparison:** This study uses two loss function, the binary cross entropy and mean square error and compared their performance and results.
- **Hyperparameter Tuning:** Tuned the hyperparameters like lambda, batch size, and optimal training data size.
- **Data Augmentation and its Hyperparameters Tuning:** The study incorporates data augmentation techniques like Gaussian blur, gamma correction, and rotation at various values and compared them to obtain the best parameters like kernel size for Gaussian blur, gamma for gamma correction, and random rotation angle range ($-\theta$ to $+\theta$).
- **Fourier Transform:** The study attempted to learn Fourier to Fourier mapping in the frequency domain in the hope that the model will be able to learn more minor and more precise details than the one learned on the spatial domain.

- **Patch-based Learning:** Implemented a patch-based learning approach. It focuses on generating small patches or regions within an image rather than the image as a whole. This allows a more localised understanding of texture and other features, which can result in more realistic and detailed synthetic images.

[7] Some studies have implemented patch-based learning because of data scarcity and low resources but none with the aim of making the model more robust of anatomic structure and generalize it to different human body parts.

The best performance on the test results were obtained when the model was trained along with synthetic data generated using data augmentation technique as the model is trained on various instances of training data obtained on various settings and different parameters which makes the model more robust and more generalize to unseen data.

The model performs poorly in some cases like when the input image is blurred because of which the model is not able to capture fine details of the target image and also generally we can see loss in some fine and localized details in the few test images. To overcome this, the study also attempts patch based approach for better fine detailing and localized learning but for now the performance of this approach is only satisfactory but it can be further improved by optimizing the hyperparameters for the approach.

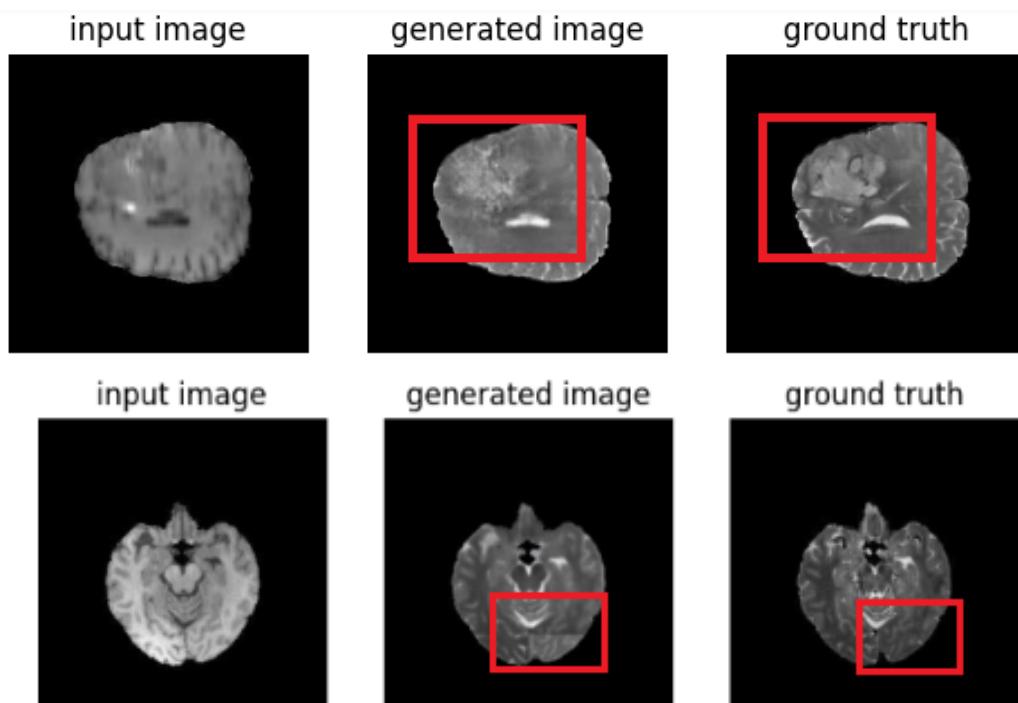


Figure 7.1: Cases when generated image is not able to capture fine details and model performs poorly

7.2 Future Work :

This study involves how T1-weighted MRI images are produced from T2-weighted sequences and vice versa. To assess the accuracy and the robustness of our approach even more, we can extend our methodology to other combinations of MRI sequence types like FLAIR, T1ce, STIR, etc.

The evaluation of this study is done only on the test data which is part of the BRATS2020 dataset. Therefore, considering the data from the different sources would be advantageous, and further evaluation could be done to establish the model's generalization ability with respect to the diverse datasets, ensuring the model's robustness and applicability across the different clinical settings.

Our patch-based learning strategy can be expanded even further to make the model more robust towards body structure and learn a mapping from one MRI sequence to the other regardless of the body part. Training the model on a data set, including the MRI sequences and images of different body parts, allows the model to learn more about local features regardless of the overall anatomical structure.

The project's output will be turned into an abstract, a text, and a copyright.

7.3 Conclusion :

The results presented in chapter 6 obtained are after training the GAN pix2pix model on thousand of brain MRI slices at different hyperparameter settings of the model and using different optimization techniques. The generated images in many cases are highly accurate and visibly it is difficult to distinguish between the real and generated images but in few cases the results obtained are not satisfactory and indicates more room for improvement before applying it to practical applications. The results obtained using evaluation metrics are satisfactory with the possibility for more optimization in the future. The use of data augmentation techniques increased the model performance and generalization of the model to different MRI sequences obtained under different conditions. These findings validate the usefulness of GANs in neuroimaging and are very encouraging for future research.

References

- [1] Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. "Image-to-Image Translation with Conditional Adversarial Networks." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5967-5976, 2017 <https://doi.org/10.48550/arXiv.1611.07004>
- [2] Nripendra Kumar Singh, Nripendra Kumar, and Khalid Raza. "Medical Image Generation using Generative Adversarial Networks." 2020. <https://doi.org/10.48550/arXiv.2005.10687>
- [3] Han, Changhee, Hideaki Hayashi, Leonardo Rundo, Ryosuke Araki, Wataru Shimoda, Shinichi Muramatsu, Yujiro Furukawa, Giancarlo Mauri, Hideki Nakayama. "GAN-Based Synthetic Brain MR Image Generation." 2018 IEEE 15th International Symposium on Biomedical Imaging.2018. <https://doi.org/10.1109/ISBI.2018.8363678>
- [4] Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. "Generative Adversarial Networks". 2014. <https://doi.org/10.48550/arXiv.1406.2661>
- [5] Abeer Aljohani, Nawaf Alharbe. "Generating Synthetic Images for Healthcare with Novel Deep Pix2Pix GAN". 2022. <https://doi.org/10.3390/electronics11213470>
- [6] Yang, Q., Li, N., Zhao, Z. et al. MRI Cross-Modality Image-to-Image Translation. Sci Rep 10.2020. <https://doi.org/10.1038/s41598-020-60520-6>
- [7] Wang C-C, Wu P-H, Lin G, Huang Y-L, Lin Y-C, Chang Y-P, Weng J-C. "Magnetic Resonance-Based Synthetic Computed Tomography Using Generative Adversarial Networks for Intracranial Tumor Radiotherapy Treatment Planning".Journal of Personalized Medicine. 2022. <https://doi.org/10.3390/jpm12030361>
- [8] Xi Zhao, Haizheng Yu, Hong Bian."Image to Image Translation Based on Differential Image Pix2Pix Model".Comput Mater Contin. 2023. <https://doi.org/10.3390/jpm12030361>