

# **Generative AI Approach to Synthesize CT from MRI Using Pix2Pix GAN**

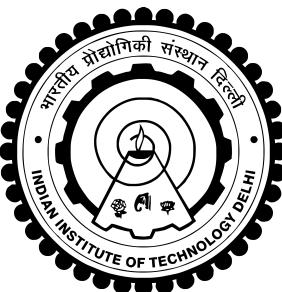
*Thesis submitted by*

**Khushal Shakya  
2022AIB2683**

*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 Yardi School of AI  
INDIAN INSTITUTE OF TECHNOLOGY DELHI**

**May 2024**

## THESIS CERTIFICATE

This is to certify that the thesis titled **Generative AI Approach to Synthesize CT from MRI Using Pix2Pix GAN**, submitted by **Khushal Shakya (2022AIB2683)**, 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 is a crucial component of modern health care, as it enables professionals to analyze the body's internal structures for diagnosis, monitoring, and treatment of various medical conditions. The tools are important in this field and some of them include Magnetic Resonance Imaging (MRI) and Computerized Tomography (CT). Usually, there are several MRI sequences, including T1-weighted, T2-weighted, FLAIR and DWI sequences. However, traditional CT imaging methods are associated with longer examination times, higher resource consumption and increased patient discomfort, posing challenges to workflow optimization. CT imaging involves the use of X-rays, which are a form of ionizing radiation. Exposure to ionizing radiation is known to increase the risk of cancer. The Synthetic Computerized Tomography (sCT) problem gets around these issues by reconstructing CT scans from other input modalities such as MRI images. This research focuses on synthesizing CT scans specifically for brain imaging. The aim is to develop a model that takes an MR image as input and generates the corresponding CT scan. The selected approach utilizes the Pix2Pix Generative Adversarial Network (GAN), a deep learning framework known for its efficacy in image-to-image translation tasks. The project involves training the Pix2Pix GAN model on a paired MRI and CT scan dataset to capture the intricate details of the brain's anatomy. The study shows that the Pix2Pix GAN model effectively creates synthetic CT scans from MR images, improving clarity and accuracy. This work could be useful in medical research, education, and clinical practice by providing a non-invasive way to create CT-like images.

# Contents

<b>ACKNOWLEDGEMENTS</b>	<b>i</b>
<b>ABSTRACT</b>	<b>ii</b>
<b>LIST OF TABLES</b>	<b>v</b>
<b>LIST OF FIGURES</b>	<b>vii</b>
<b>ABBREVIATIONS</b>	<b>viii</b>
<b>1 INTRODUCTION</b>	<b>1</b>
1.0.1 Common MRI sequences . . . . .	1
1.1 Computerized Tomography (CT): . . . . .	3
1.2 Background and preliminaries . . . . .	3
1.3 Problem Statement . . . . .	5
1.3.1 Common concerns associated with CT scans: . . . . .	5
1.4 Objectives . . . . .	5
<b>2 State of the Art</b>	<b>6</b>
2.1 GANs in Medical Imaging . . . . .	6
2.1.1 Generative Adversarial Network . . . . .	6
2.1.2 Pix2Pix GAN . . . . .	7
<b>3 Data and Preprocessing</b>	<b>10</b>
<b>4 Evaluation Metrics</b>	<b>11</b>
<b>5 Optimization Strategies</b>	<b>13</b>
<b>6 Results</b>	<b>16</b>
<b>7 Discussion and Conclusion</b>	<b>31</b>

7.1	Discussion :	31
7.2	Future Work :	33
7.3	Conclusion . . . . .	34

## List of Tables

6.1	RMSE Error on Test Images Using 130 Patients Data for Training with and without Regularizer . . . . .	16
6.2	RMSE Error and SSI on Test Images Using 140 Patients Data for Training with BCE and MSE loss . . . . .	17
6.3	RMSE and SSIM on Test Images Using Different Numbers of Patients for Training . . . . .	18
6.4	RMSE and SSIM on Test Images Using 25,000 Patients for Training and varying lambda . . . . .	20
6.5	RMSE and SSIM on Test Images Using 140 Patients for Training and Varying Batch Size . . . . .	21
6.6	Comparison of Model Performance: Optimal Dataset vs. Full Dataset with Data Augmentation . . . . .	23
6.7	RMSE and SSIM on Test Images with Varying only Sigma Parameters (Epoch = 10) . . . . .	23
6.8	RMSE and SSIM on Test Images with Varying Kernel Sizes (Sigma = 9, Epoch = 10) . . . . .	24
6.9	RMSE and SSIM on Test Images using Different Random Rotations in the Range of $-\theta$ to $\theta$ . . . . .	24
6.10	RMSE and SSIM on Test Images Using Different Augmentation Techniques	24
6.11	Comparing RMSE and SSIM on Test Images with and without vertical flip as Augmentation Technique . . . . .	25
6.12	Comparison of Experimental Settings . . . . .	26
6.13	RMSE and SSIM on Test Images with Increased Training Data and Different Epochs . . . . .	27
6.14	RMSE on Test Images Using Fourier Learning . . . . .	28
6.15	RMSE and SSIM on Test Images Using Patch-Based Learning Approach . . . . .	30
6.16	Comparison Of PSNR Models Trained On Full Image Approach and Patch Image Approach . . . . .	30

# List of Figures

1.1	T1-Weighted Image . . . . .	2
1.2	T2-Weighted Image . . . . .	2
1.3	FLAIR . . . . .	2
1.4	CT Image . . . . .	3
2.1	Generative Adversarial Network Framework . . . . .	6
2.2	Generator architecture from pix2pix . . . . .	7
2.3	Discriminator architecture from pix2pix . . . . .	8
6.1	With Regularization . . . . .	16
6.2	Without Regularization . . . . .	16
6.3	Result at BCE . . . . .	17
6.4	Result at MSE . . . . .	17
6.5	Plots of SSIM and RMSE . . . . .	18
6.6	Generated Images 10,000 images dataset . . . . .	18
6.7	Generated Images 15,000 images dataset . . . . .	18
6.8	Generated Images 20,000 images dataset . . . . .	19
6.9	Generated Images 25,000 images dataset . . . . .	19
6.10	Generated Images 29,000 images dataset . . . . .	19
6.11	Generated results at $\lambda = 100$ . . . . .	20
6.12	Generated results at $\lambda = 500$ . . . . .	20
6.13	Generated results at $\lambda = 1000$ . . . . .	20
6.14	Generated results at $\lambda = 1500$ . . . . .	21
6.15	Generated results at batch size = 30 . . . . .	21
6.16	Generated results at batch size = 50 . . . . .	22
6.17	Generated results at batch size = 80 . . . . .	22
6.18	Generated results at batch size = 100 . . . . .	22

6.19 Generated results at batch size = 120 . . . . .	23
6.20 Generated Image when used Gaussian blur with kernel=29, $\sigma=9$ ,gain = 1 and $\gamma = 0.8$ . . . . .	25
6.21 Result at rotation = $[-25, +25]$ ,kernel size=29, $\sigma=9,\gamma=0.8$ . . . . .	26
6.22 Results generated at 100 epochs and new learning rate . . . . .	27
6.23 Results generated after applying data augmentation . . . . .	28
6.24 Generated image and log magnitude when trained using Fourier learning ap- proach . . . . .	29
6.25 Generated image when trained using Patch-Based learning approach . . . . .	30
7.1 Segmented area where the model is slightly lacking in Real and Generated Images . . . . .	33

## ABBREVIATIONS

<b>IITD</b>	Indian Institute of Technology, Delhi
<b>RTFM</b>	Read the Fine Manual
<b>GAN</b>	Generative Adversarial Network
<b>MRI</b>	Magnetic resonance imaging
<b>CT</b>	Computerized Tomography

# Chapter 1

## INTRODUCTION

In the medical field, imaging has become an important factor in the diagnosis, assessment, and management of various diseases by providing information regarding the internal parameters of the human body. Of the various imaging modalities Magnetic Resonance Imaging (MRI) and Computerized Tomography (CT) are the most important in medical imaging.

CT scans are a form of imaging test that shows clear pictures of inside parts of the body by taking numerous images using an X-ray device that rotates around the patient. This process gives a perspective 3D model which provides an overall look at the various structures of human anatomy. However, different imaging techniques, such as MRI- Magnetic Resonance Imaging, use strong magnetic fields and radio waves to produce the images without ionizing radiation. MRI imaging machine looks somewhat like a large tube containing magnets and the patient can lie inside during imaging.

CT and MRI are both important in medical imaging, the differences in image processing and contrast tend to enhance each other often. Although CT scans are powerful diagnostic tools, the traditional method of obtaining CT images requires longer examination times, higher resource consumption, and patient discomfort. The existing CT imaging procedure proves problematic when considering workflow optimization and patient comfort. This work focuses on the synthetic computerized tomography problem, reconstructing CT images from MRI data. We aim to develop a model that can be trained to transform MRI images into synthetic CT images using the Pix2Pix GAN.

This research is centred on diagnostic imaging of brains, where accurate and non-invasive methods have to be employed. Thanks to the choice of Pix2Pix GAN, which proved its efficiency in image-to-image translation tasks, we will be able to generate synthetic CT images that describe fine features of the brain structure accurately.

### 1.0.1 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.

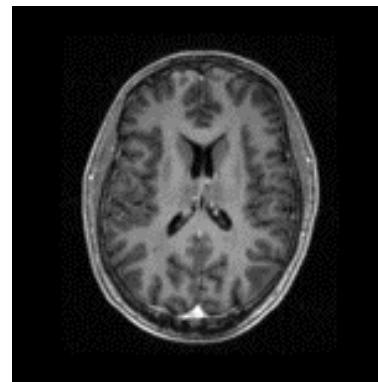


Figure 1.1: T1-Weighted Image

(2) **T2-Weighted (T2W) Sequence:** T2W images highlight differences in water content and are particularly sensitive to fluids. They are useful for detecting abnormalities such as edema, inflammation, and fluid-filled cysts.

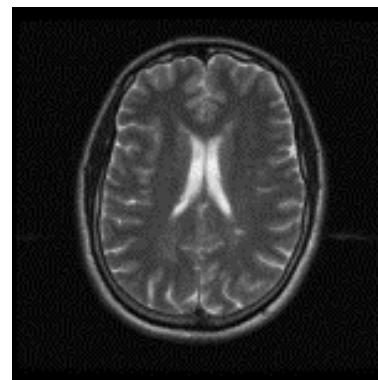


Figure 1.2: T2-Weighted Image

(3) **Fluid-Attenuated Inversion Recovery (FLAIR) Sequence:** FLAIR sequences reduce the signal intensity from CSF which means that it is easier to observe the pathology in the brain as there will be less interference from the CSF.

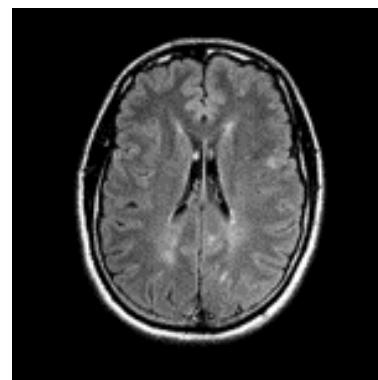


Figure 1.3: FLAIR

These and other imaging techniques assist in further evaluating the body's structures. The technique depends on what has to be diagnosed and some of the tissues/abnormalities being diagnosed. T1 combined with T2 and the FLAIR allows the radiologists to understand the patient and improve the diagnostic and treatment fully.

## 1.1 Computerized Tomography (CT):

A CT (computed tomography) image, also known as a CT scan or CAT scan image, is a detailed cross-sectional view of the body's internal structures. CT imaging is a medical diagnostic technique that uses X-rays and computer processing to create detailed images of different body parts.

CT images are characterized by their high resolution and ability to differentiate between different types of tissues based on their density. The images are typically displayed in grayscale, where different shades of grey represent varying levels of tissue density. For example, bones appear white or light grey, while air-filled spaces, such as the lungs, appear dark.

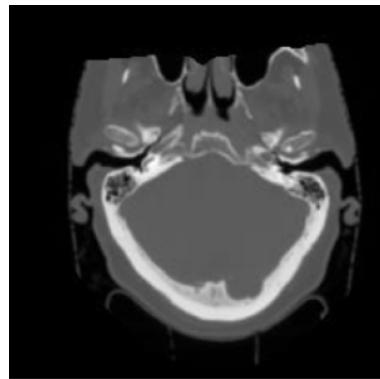


Figure 1.4: CT Image

## 1.2 Background and preliminaries

### 1. Abeer Aljohani, 2022[5]:

In recent years, machine learning and deep learning have been widely applied for medical image processing and have shown incredible achievements. Medical image processing aims to extract abnormal information about a patient's medical condition. Medical images are obtained through several medical technologies, including Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET), and Ultrasound (US). The acquired images are processed using deep learning techniques to extract essential information about the disease visible 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 is well-suited for tasks where obtaining paired data (input-output pairs) is challenging. It can learn translation mappings from unpaired datasets. It is a model based on unsupervised learning. Pix2Pix performs well when there is access to paired input-output data. Pix2pix allows for more control over the translation process. It is a model based on supervised learning. The Pix2pix model outperforms the cycleGAN model in terms of generating more realistic images whenever paired data is available, as it provides more control during translation.

## 1.3 Problem Statement

Although CT scans are powerful diagnostic tools, the traditional method of obtaining CT images requires longer examination times, higher resource consumption, and discomfort to patients. The existing CT imaging procedure proves problematic when considering workflow optimization and patient comfort.

### 1.3.1 Common concerns associated with CT scans:

1. **Ionizing Radiation Exposure:** CT scans use X-rays, which involve ionizing radiation. Excessive exposure to ionizing radiation can increase the risk of cancer, increasing the risk of harm.
2. **Limited Soft Tissue Differentiation:** CT scans provide excellent imaging of bones and dense structures but have limitations in differentiating certain soft tissues. Combining CT with other imaging modalities, such as MRI, can offer a more comprehensive assessment of soft tissues.
3. **Contrast Media Risks:** In some cases, CT scans involve the use of contrast media (iodine-based dye) to enhance image visibility. Allergic reactions or kidney problems can occur in some individuals.
4. **Resource Allocation:** CT scans are expensive pieces of equipment with high demand. Longer scan times for a single patient to get multiple sequences are equally disadvantageous since the scanner is tied down and cannot accommodate other patients who may be waiting.

## 1.4 Objectives

To evaluate the potential of an optimized deep learning model generating synthetic CT images from MRI images of the same person. This research aims to enhance the utility of MRI scans by providing complementary CT-like images without the additional radiation exposure.

# 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

In the case of medical image synthesis, several studies have been explored through GANs. However, for the current approach, this architecture is crucial in cross-modality image synthesis, which is transforming an image of one modality into an image of another modality.

#### 2.1.1 Generative Adversarial Network

A generative Adversarial Network was first introduced in the year 2014 by Ian J. Goodfellow [4] and is a new technique for estimating generative models in an adversarial form. The system is composed by two neural networks a discriminator D, usually a CNN and a generator G that are trained in parallel.

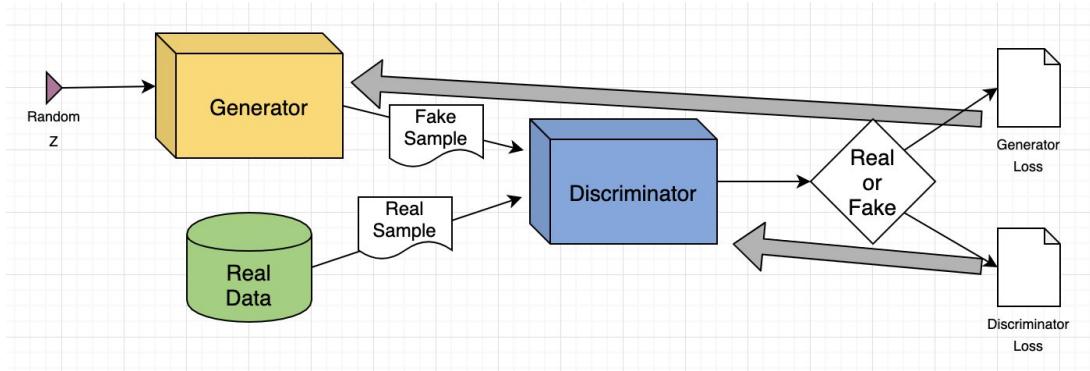


Figure 2.1: Generative Adversarial Network Framework

In particular, the generator is trained to learn the probability distribution of the input data and generate synthetic data that is similar to real data, while the discriminator's role is to determine which images are real or fake.

### 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 :

A U-Net model architecture is used for the generator, instead of the common encoder-decoder model. A U-Net consists of an encoder (downsampler) and a decoder (upsampler). 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, enabling it to bypass the bottleneck.

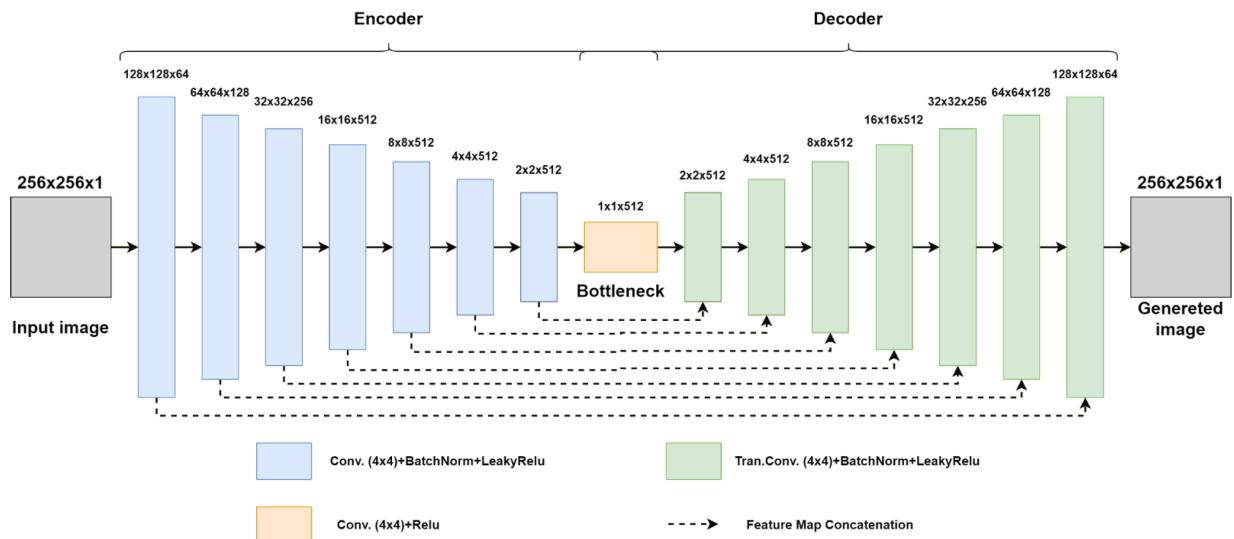


Figure 2.2: Generator architecture from pix2pix

#### Discriminator :

The discriminator model takes an image from the source domain and an image from the target domain and predicts the likelihood of whether the image from the target domain is a real or generated version of the source image.

#### Input :

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

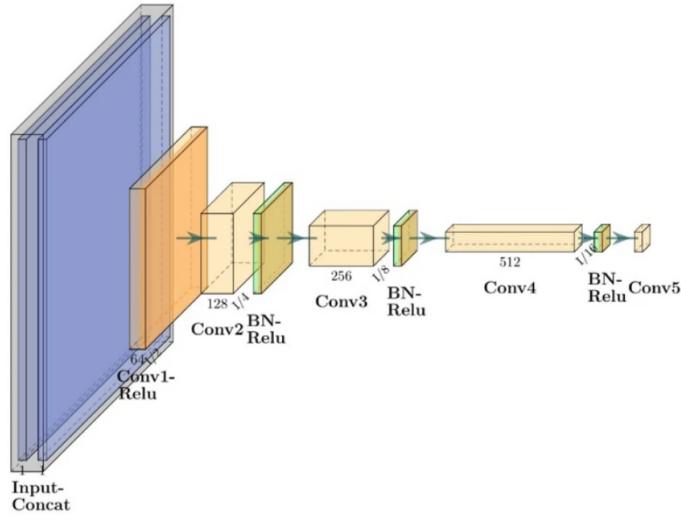


Figure 2.3: Discriminator architecture from pix2pix

### Output :

Probability that the image from the target domain is a real 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. This is a deep convolutional neural network designed to classify patches of an input image as real or fake, rather than the entire 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  $\lambda$  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 synthRAD2023 dataset is used which consists of MR/CT image pairs from 180 patients. For anonymity, we will refer to the three centers with centers A, B, and C without specifying which letter belongs to which center. This dataset consists of MR/CT image pairs. This dataset consists brain images.

Each patient folder contains an MR (mr.nii.gz), a CT (ct.nii.gz), and a binary mask (mask.nii.gz) image. To handle these images a library called nibabel is used. Each image is loaded using nibabel.

T1-W and CT sequences are taken from the dataset only. T1-W is considered as the input image and CT is the targeted image.

By taking into account all of the slices that each image was processed and cut into resulting in 31,800 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 synthRAD2023 dataset, using data from 140 patients for training (29,000 images) and 20 patients for testing (3000 images).

**Data Preprocessing for Fourier learning :** In Fourier-based learning, the input images were first transformed into the frequency domain using the Fourier transform with NumPy python library. This transformation produced real and imaginary components combined to form two channels, resulting in a new input shape of (256,256,2). During evaluation, the images were first converted back from the frequency domain to the spatial domain using the inverse Fourier transform, and then the evaluation was carried out.

**Data Preprocessing for Patch Based learning :** In 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 centre of the image, resulting in 5 patches per input image. Thus, the data set used for training, initially consisting of 29,000 images, is transformed into 145,000 patch images, each size (128,128,3).

## Chapter 4

### Evaluation Metrics

The proposed model's performance is assessed using the test dataset and different key metrics. These measures are selected to offer a thorough evaluation of the resulting CT sequences with respect to both pixel-by-pixel accuracy and structural integrity. This will give an insight into how well the model can synthesize CT sequences as per the 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 CT 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,
- $\mu$  is the mean intensity,
- $\sigma^2$  the variance of the image,
- $\sigma_{xy}$  the covariance of  $x, y$ .

**2. Root Mean Squared Error (RMSE):** RMSE measures the average squared error between the pixel intensities of the generated and actual images. A lower RMSE indicates fewer pixel-wise errors, suggesting better realism of the generated CT images.

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

where:

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

**3. Peak Signal-to-Noise Ratio (PSNR):** The Peak Signal-to-Noise Ratio (PSNR) is a widely used metric for measuring the quality of a synthesized image compared to a

---

true image. It is particularly common in fields such as image compression, denoising, and synthesis.

$$\text{PSNR} = 10 \log_{10} \left( \frac{I_{\max}^2}{\text{MSE}} \right)$$

where:

- $I_{\max}^2$  is the maximum possible pixel value of the image (for an 8-bit image, this is 255).
- MSE (Mean Squared Error) is the mean squared error between the original and the reconstructed image.

A higher PSNR generally indicates that the reconstruction is of higher quality, meaning the image is closer to the original. Conversely, a lower PSNR indicates more distortion.

**4. Histogram Visualization:** Histogram visualization is a method for comparing the distribution of pixel intensity values between generated and real images. By generating histograms that display the number of data points at each intensity level, one can visually compare the intensity distributions of two images.

These histograms allow for the quick identification of variances in brightness and contrast and highlight differences in dynamic range. Such differences may reveal areas in the generated images that do not completely match the originals. Additionally, variations in the shape of the histograms can indicate errors in the generated images that might not be immediately noticeable.

# Chapter 5

## Optimization Strategies

A thorough optimization procedure was carried out to improve the suggested model's ability to produce synthetic CT images from T1 MR images. A training dataset with 60 patients was used in the early phases of the model's development. However, an early finding revealed an overfitting tendency, with the model performing exceptionally well on training data but not preforming on unobserved data. An L2 regularisation method was used to address overfitting issues.

During the training phase, we 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. The purpose of this investigation was to evaluate how different loss functions affect the model's overall performance.

The optimal training set size was explored by changing the amount of the training dataset, going from 25 patients to 140 patients(29,000 images),realizing the crucial role that data quantity has on the process of training the models. The effects of these variations were determined by comparing the evaluation metrics on the test dataset.

We also investigated how the hyperparameter lambda ( $\lambda$ ) 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 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 robustness.

**Random Rotation:** Adding training data with variations in object orientations helps the model learn invariant properties from multiple perspectives due to random rotation. Random rotation improves the model's ability to handle spatial transformations and enhances its capacity to identify objects regardless of their orientation, as it is exposed to more data with rotated samples.

**Gaussian Blur:** To imitate imaging artefacts such as motion blur or lens deficiencies within our dataset, we employ the Gaussian blur filter which improves the dataset quality. Gaussian blurring is performed by convolving the image with Gaussian kernels of different standard deviations providing particular levels of blurriness. When Gaussian blur is applied, the distortions that exist on a faulty image will be magnified hence enhancing the performance of the model.

**Gamma Correction:** Gamma correction allows transforming the pixel intensities with a purpose of getting rid of varying brightness in various contexts. When using gamma correction, we have the model made available to the set of luminance levels which enables the model to make the required predictions based on the different light conditions. This enhances the model's capacity and utilization of information that allows it to perform tremendously under different lighting conditions.

In order to determine how effective these data augmentation methods are, we conduct an evaluation across multiple hyperparameter settings. As for the hyperparameters, we systematically test gamma values, rotation angles, and blur magnitudes to analyze how they affect the model's performance in terms of accuracy and robustness.

First, the study was performed based on the data augmentation technique to distort some of the training samples. Following this, we expanded this by using data augmentation to create new training data samples to see if the model could be trained and perform well on unseen data with a larger number of training samples and epochs.

The approach includes taking the Fourier transform to convert the input images from the spatial domain to the frequency domain and then training the model to enable it to learn the mapping of input and output images in the frequency domain. Transformation and reconstruction processes will be more effective when operating in the frequency domain, enabling the model to capture the essential frequency characteristics and patterns.

The Fourier transform focuses on frequency patterns in data rather than specific spatial details. Because of this, models trained using frequency information tend to be more robust against spatial changes or distortions in input images. This means that using this approach, imaging can be performed reliably and efficiently across different settings and with various modalities.

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

Different patches can be generated separately and combined to create a final image, allowing for greater control over the overall appearance. Synthetic image generation involves variability in data, such as different conditions, dark regions, light regions, and tu-

---

more patches. Patch-based methods can better adapt to such variability by focusing on local features and patterns, allowing for more robust image generation across different body parts.

# Chapter 6

## Results

The evaluation has been done 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 and lambda=100.

Training data =140 patients

When training with only 25 patients data the 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	96.22	79.19

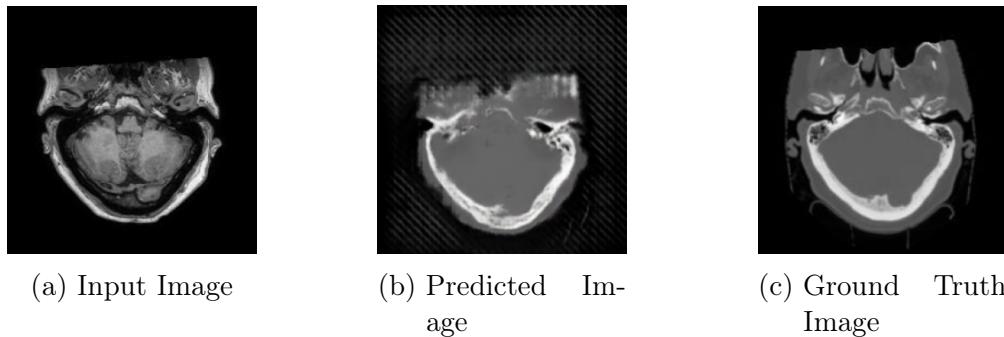


Figure 6.1: With Regularization

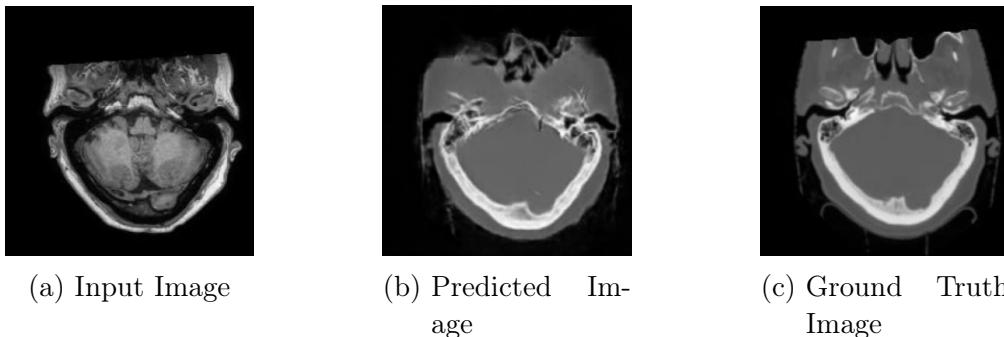


Figure 6.2: Without Regularization

The above fig and table show that there was no benefit in using regularization the performance of the model was decreasing only. As we can see without regularization rmse is

quite low as compared to with regularization. In the above results, we only used 25 patients data. To overcome the issue of overfitting and get better test results the model was trained on a whole dataset of 140 patients (29,000 images).

The performance of the two 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.

Table 6.2: RMSE Error and SSI on Test Images Using 140 Patients Data for Training with BCE and MSE loss

	BCE	MSE
RMSE	69.199	64.030
SSI	0.835	0.859

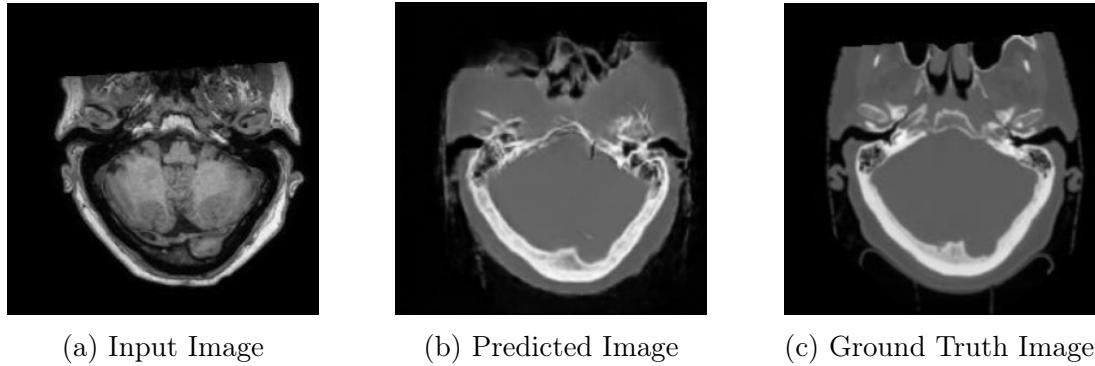


Figure 6.3: Result at BCE

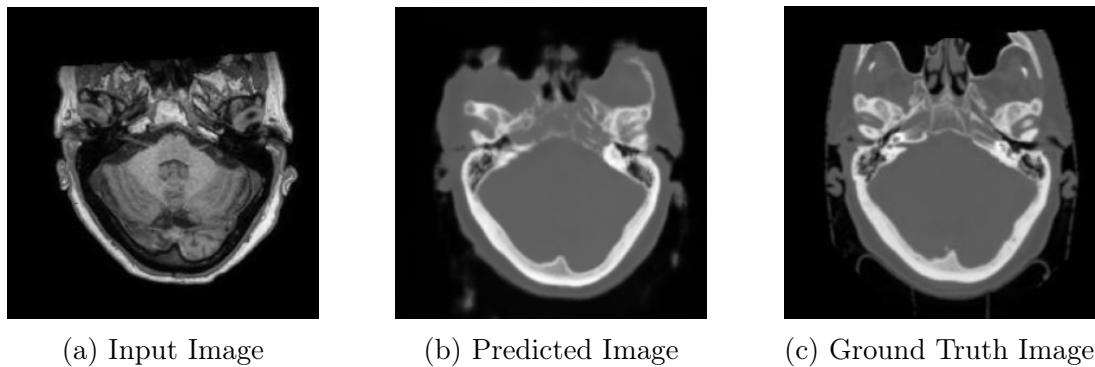


Figure 6.4: Result at MSE

The above fig and table show model that uses MSE performs better than the model that uses BCE. So further analysis has been carried out only on the MSE loss model. The optimal training set size has been explored and the model's performance 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 Images	RMSE	SSI
10,000	64.67	0.826
15,000	65.71	0.846
20,000	59.69	0.854
25,000	59.93	0.856
29,000	64.73	0.848

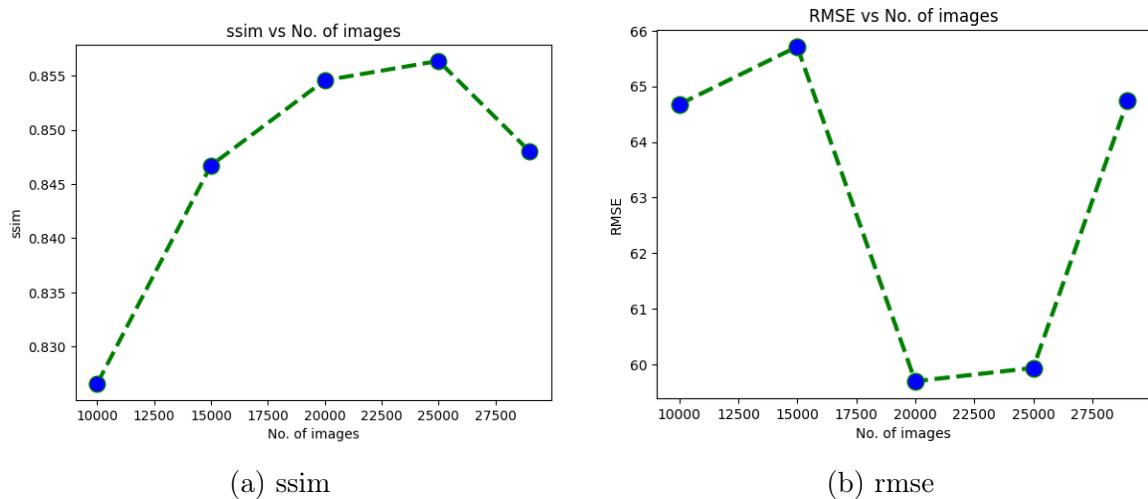


Figure 6.5: Plots of SSIM and RMSE

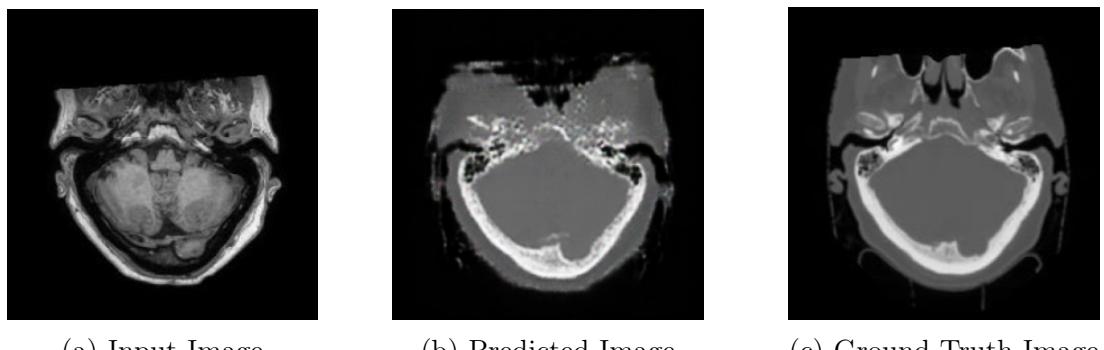


Figure 6.6: Generated Images 10,000 images dataset

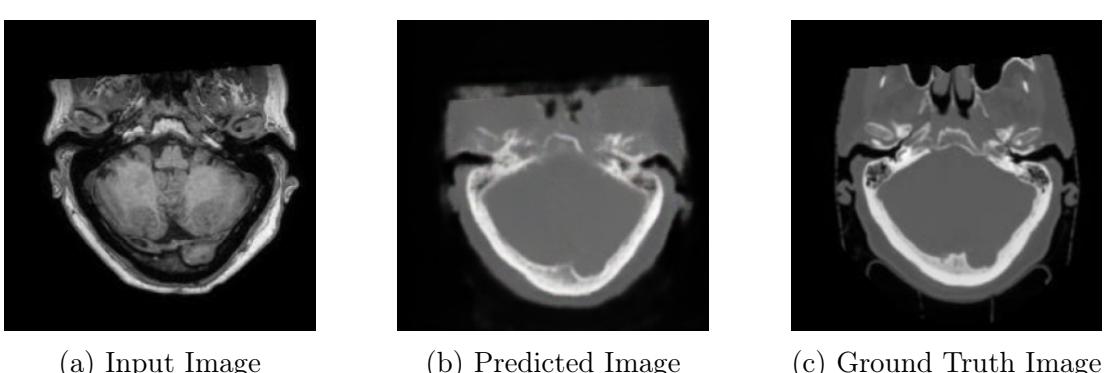


Figure 6.7: Generated Images 15,000 images dataset

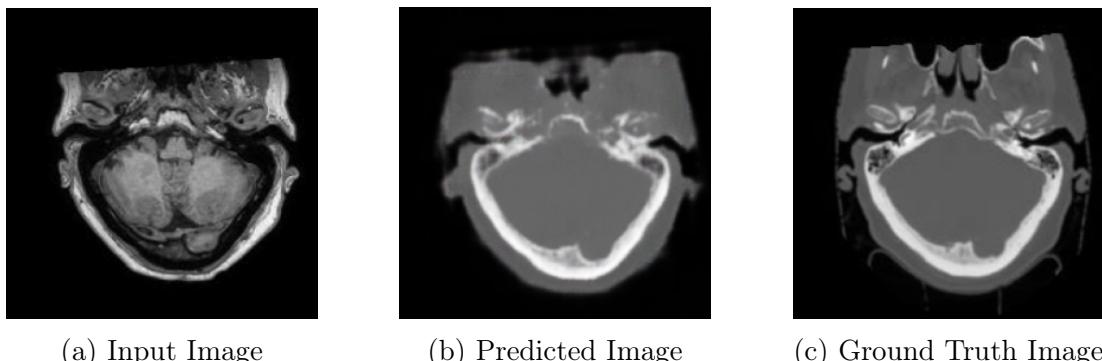


Figure 6.8: Generated Images 20,000 images dataset

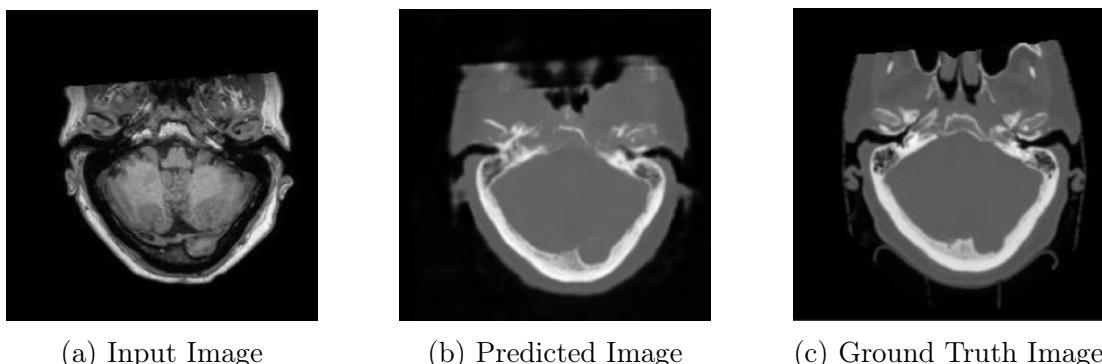


Figure 6.9: Generated Images 25,000 images dataset

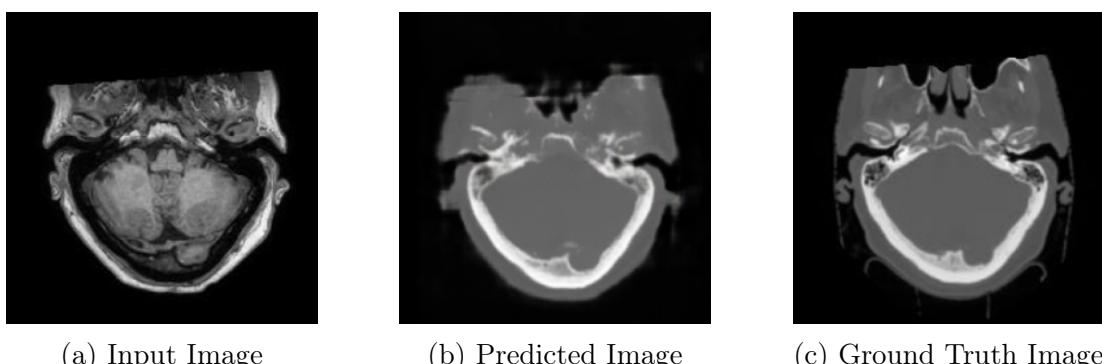


Figure 6.10: Generated Images 29,000 images dataset

The above fig and table show the model trained with 25,000 images gives the best results. So further analysis has been carried out only on the model trained on 25,000 images. 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 25,000 Patients for Training and varying lambda

lambda	RMSE	SSI
100	59.93	0.856
500	64.08	0.854
1000	57.00	0.862
1500	64.76	0.851

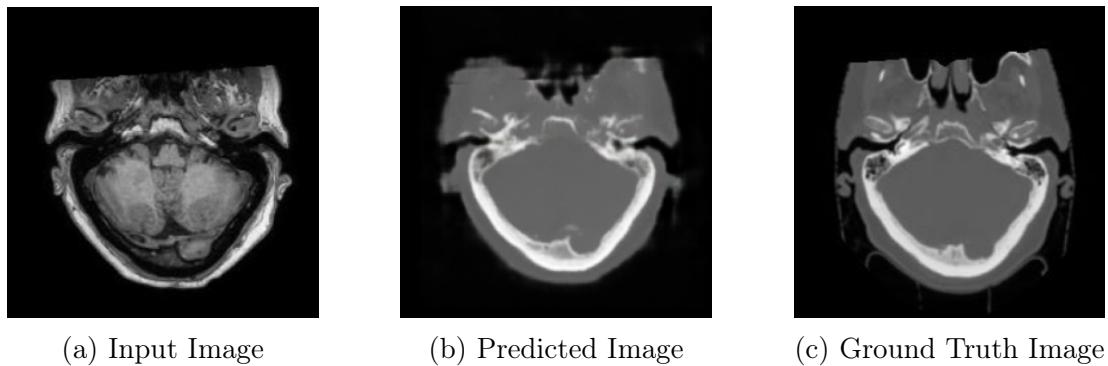


Figure 6.11: Generated results at  $\lambda = 100$

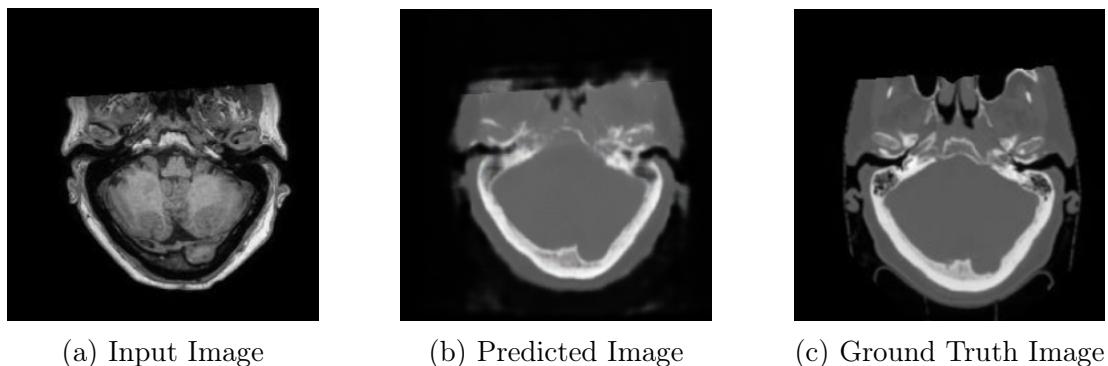


Figure 6.12: Generated results at  $\lambda = 500$

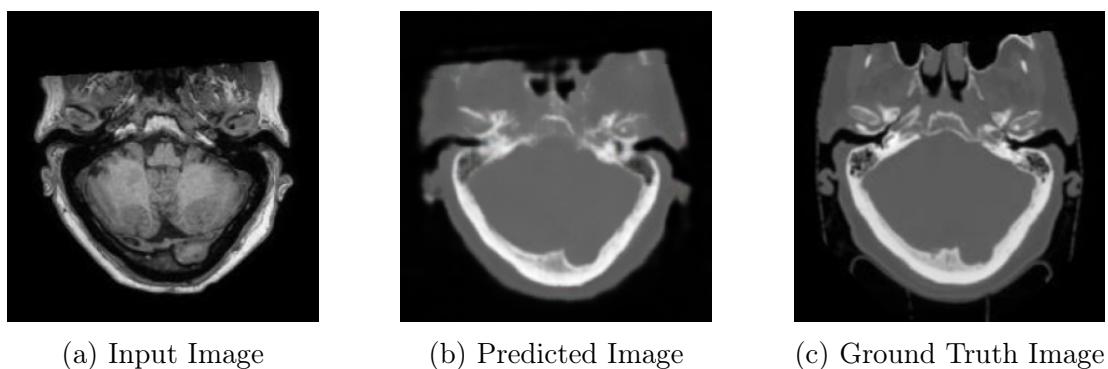


Figure 6.13: Generated results at  $\lambda = 1000$

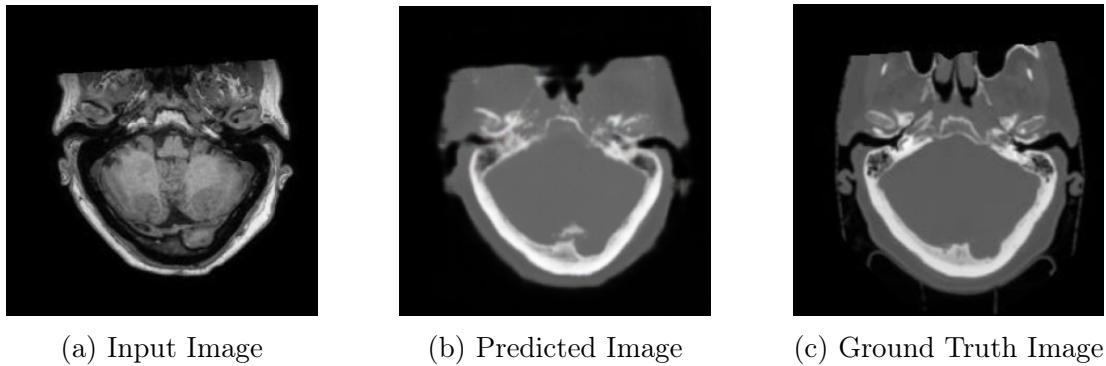


Figure 6.14: Generated results at  $\lambda = 1500$

The above fig and table show the model trained with a lambda value of 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 140 Patients for Training and Varying Batch Size

Batch Size	RMSE	SSI
30	75.99	0.79
50	57.00	0.86
80	66.25	0.85
100	73.66	0.83
120	72.23	0.82

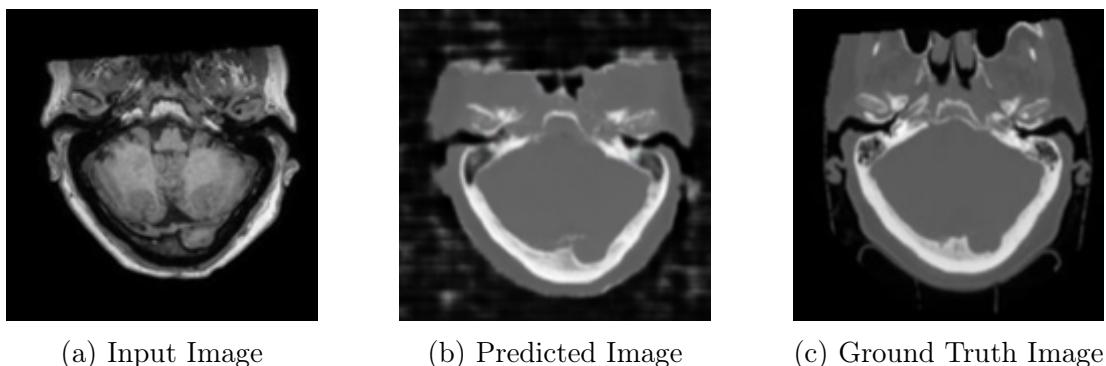


Figure 6.15: Generated results at batch size = 30

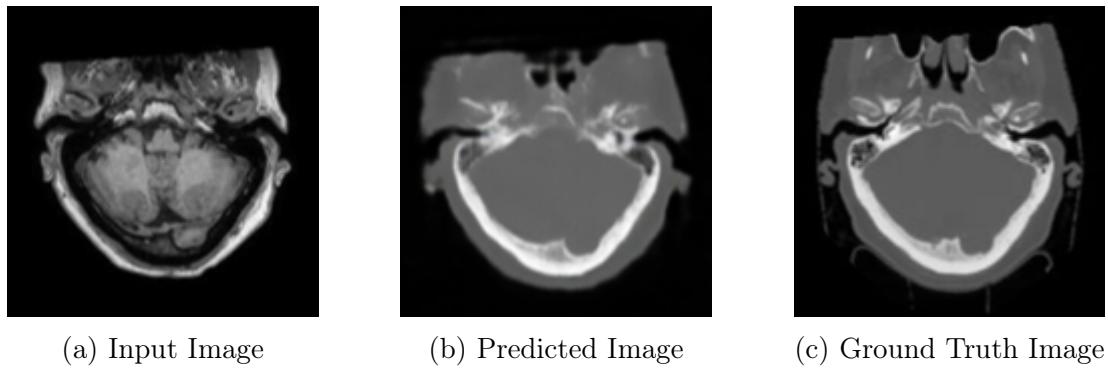


Figure 6.16: Generated results at batch size = 50

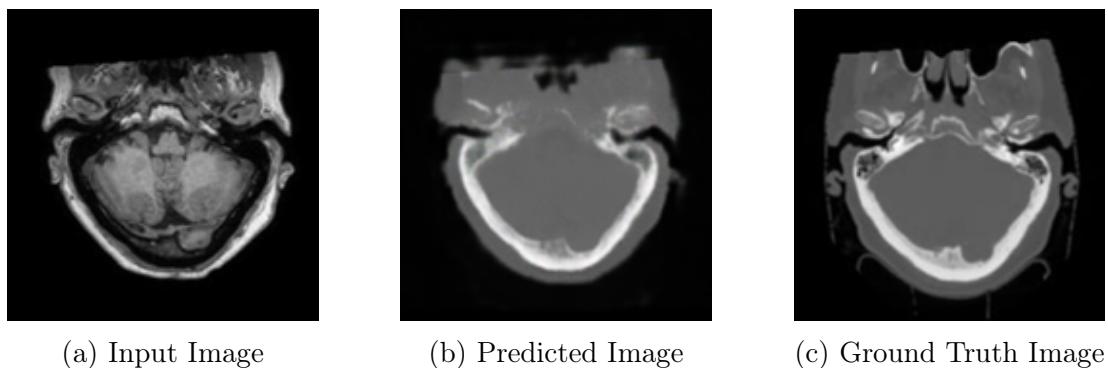


Figure 6.17: Generated results at batch size = 80

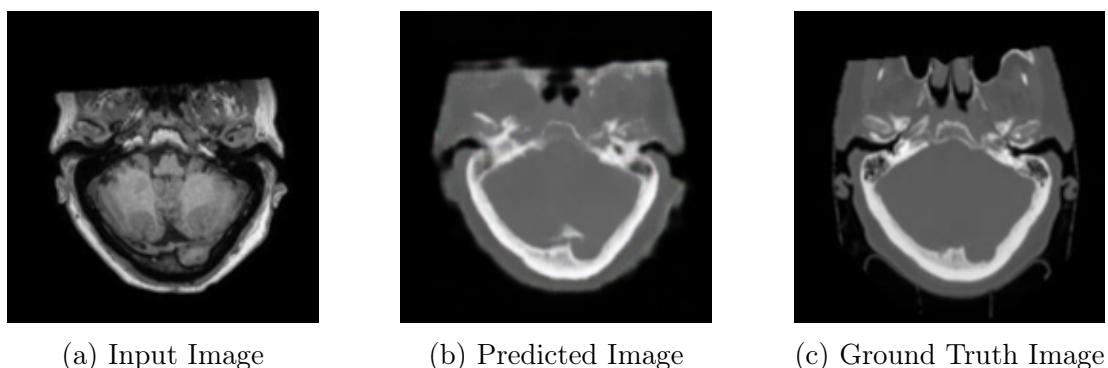


Figure 6.18: Generated results at batch size = 100

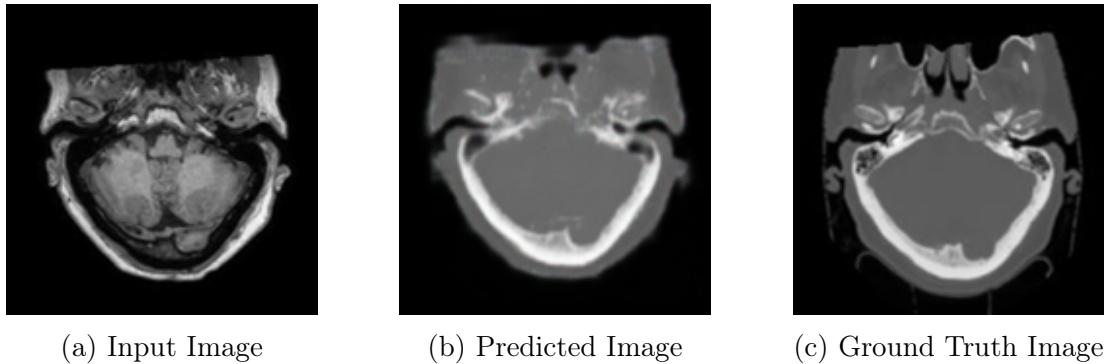


Figure 6.19: Generated results at batch size = 120

The above fig and table show the model trained with a batch size equal to 50 gives the best results. The best results of the model when evaluated on the test dataset were obtained at batch size = 50, lambda = 500, training data = 25,000 images, 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	25k Patients images
RMSE Error	61.00	65.51
Max RMSE Error	170.64	207.49
Min RMSE Error	4.25	17.40
Mean SSIM	0.85	0.85
Min SSIM	0.72	0.744
Max SSIM	0.99	0.99

Initially, the optimal training data was 25,000 patient images. However, when using data augmentation, the model performed better when trained with the whole dataset.

Table 6.7: RMSE and SSIM on Test Images with Varying only Sigma Parameters (Epoch = 10)

Metric	Sigma = 11, Kernel Size = 29	Sigma = 9, Kernel Size = 29	Sigma = 7, Kernel Size = 29	Sigma = 5, Kernel Size = 29
RMSE Error	62.83	59.49	63.05	62.70
Max RMSE Error	244.04	247.52	248.68	251.28
Min RMSE Error	16.43	1.24	13.12	14.33
Mean SSIM	0.85	0.85	0.84	0.85
Min SSIM	0.61	0.60	0.61	0.61
Max SSIM	0.99	0.99	0.99	0.99

Table 6.8: RMSE and SSIM on Test Images with Varying Kernel Sizes (Sigma = 9, Epoch = 10)

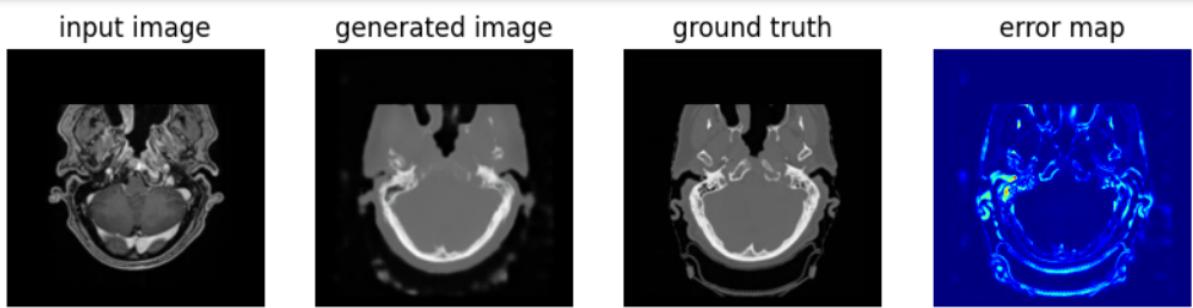
Metric	Kernel Size = 29	Kernel Size = 23	Kernel Size = 17	Kernel Size = 15
<b>Mean RMSE Error</b>	59.49	66.66	62.07	62.45
<b>Max RMSE Error</b>	247.52	1.24	247.21	248.44
<b>Min RMSE Error</b>	25.32	12.41	16.96	15.44
<b>Mean SSIM</b>	0.85	0.83	0.85	0.85
<b>Min SSIM</b>	0.60	0.60	0.61	0.61
<b>Max SSIM</b>	0.99	0.97	0.99	0.99

Table 6.9: RMSE and SSIM on Test Images using Different Random Rotations in the Range of  $-\theta$  to  $\theta$

Metric	$\theta=[-15,+15]$	$\theta=[-20,+20]$	$\theta=[-25,+25]$	$\theta=[-30,+30]$	$\theta=[-35,+35]$	$\theta=[-40,+40]$	$\theta=[-45,+45]$
<b>Mean RMSE Error</b>	62.45	62.86	59.88	65.65	60.41	61.40	69.22
<b>Max RMSE Error</b>	245.05	249.19	255.57	254.28	245.17	248.00	239.57
<b>Min RMSE Error</b>	16.96	16.17	0.84	13.03	15.44	3.23	26.96
<b>Mean SSIM</b>	0.85	0.85	0.86	0.84	0.85	0.84	0.83
<b>Min SSIM</b>	0.61	0.61	0.60	0.61	0.61	0.60	0.57
<b>Max SSIM</b>	0.99	0.99	0.99	0.99	0.99	0.99	0.96

Table 6.10: RMSE and SSIM on Test Images Using Different Augmentation Techniques

Metric	Gaussian Blur	Gamma	Gamma	Gaussian Blur and Gamma
	(kernel=29, $\sigma=9$ )	( $\gamma=0.8$ , gain=1)	( $\gamma=1.2$ , gain=1)	kernel=29, $\sigma=11$ $\gamma=0.8$ , gain=1
Mean RMSE Error	59.49	76.10	66.66	59.08
Max RMSE Error	247.52	161.46	220.67	170.17
Min RMSE Error	1.24	40.01	11.69	13.94
Mean SSIM	0.85	0.82	0.85	0.85
Min SSIM	0.60	0.71	0.73	0.70
Max SSIM	0.99	0.91	0.99	0.98



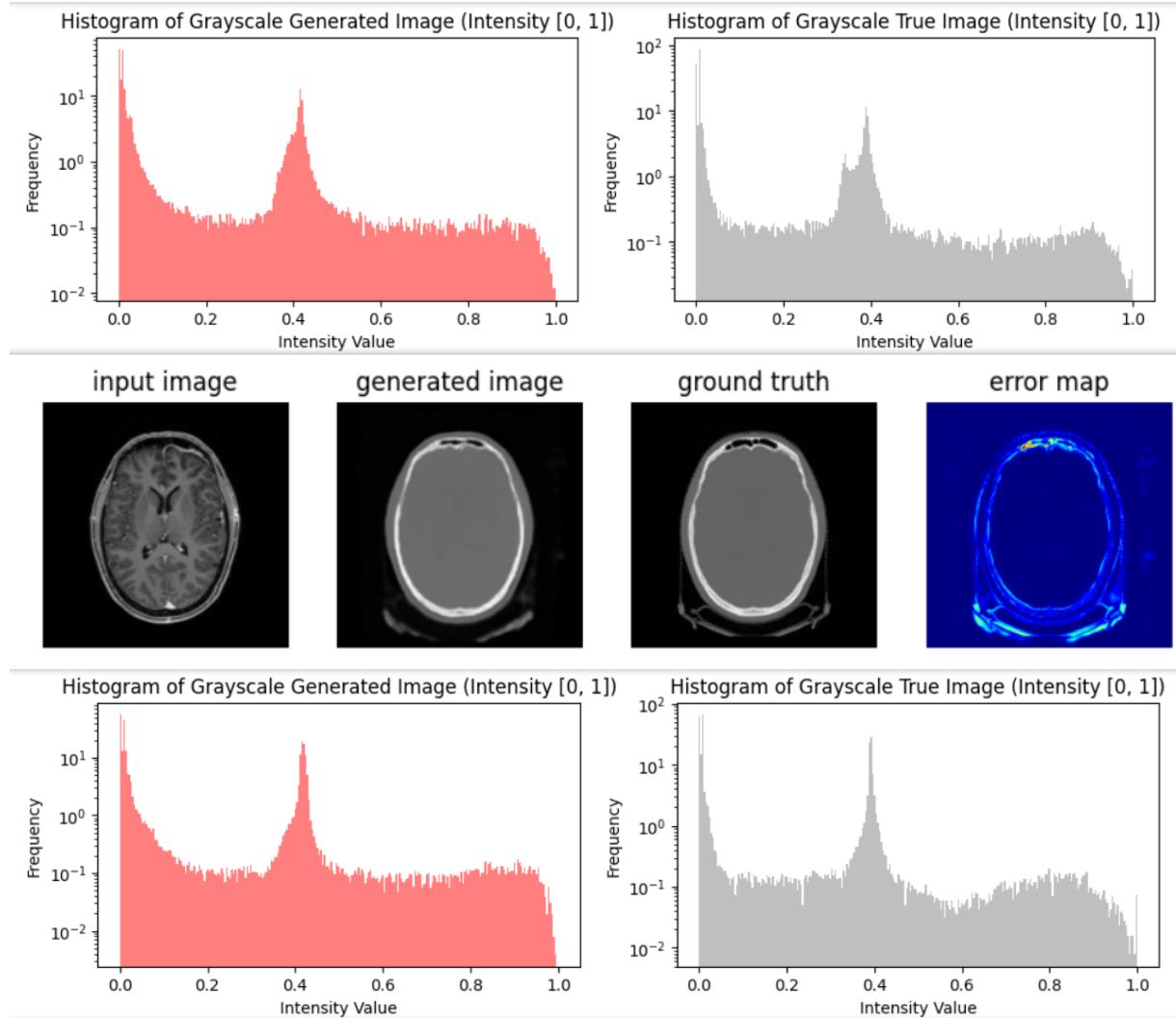


Figure 6.20: Generated Image when used Gaussian blur with kernel=29,  $\sigma=9$ , gain = 1 and  $\gamma = 0.8$

The above tables show that when using Gaussian blur with kernel size =29, $\sigma=9$ ,gain = 1 and  $\gamma = 0.8$  as data augmentation techniques performed better than other settings. The above figures are the results obtained using the same technique.

Table 6.11: Comparing RMSE and SSIM on Test Images with and without vertical flip as Augmentation Technique

Metric	$\theta=[-25,+25]$	$\theta=[-25,+25]$ with vertical flip
Mean RMSE Error	59.88	62.25
Max RMSE Error	255.57	243.23
Min RMSE Error	0.84	0.92
Mean SSIM	0.86	0.84
Min SSIM	0.60	0.60
Max SSIM	0.99	0.99

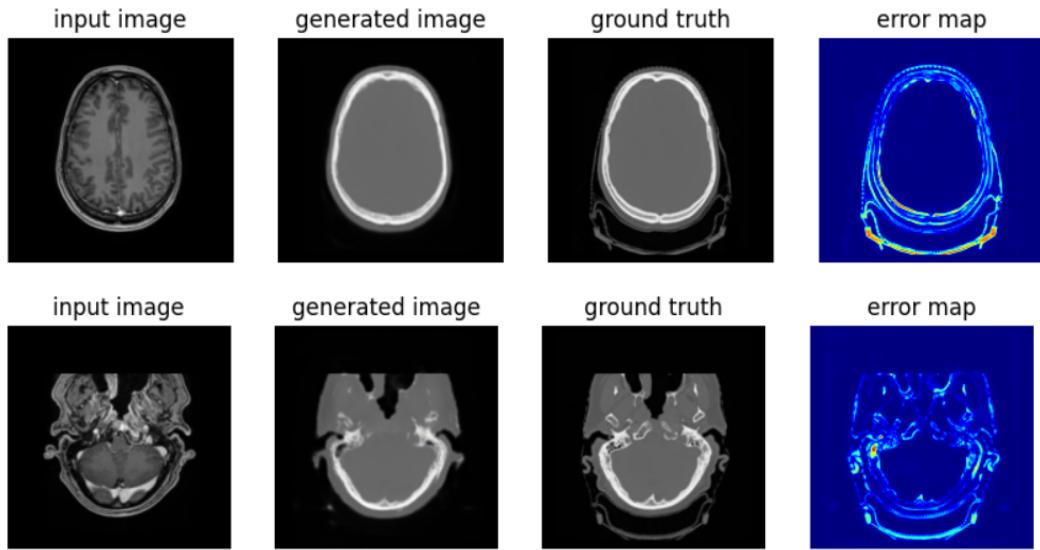


Figure 6.21: Result at rotation =  $[-25, +25]$ , kernel size=29,  $\sigma=9, \gamma=0.8$

The best setting for my model was found when used Gaussian blur with a random rotation function with the angle of rotation ranging from -25 to 25. The effect of adding a random vertical flip function was also explored but the model's performance decreased and hence removed from further analysis.

Table 6.12: Comparison of Experimental Settings

Parameter	Experiment 1	Experiment 2
Disc Learning Rate	2e-4	5e-5
Gen Learning Rate	2e-4	1e-4
Iterations per Plot	10	20
Batch size	50	16
Epochs	10	100
L1 Lambda	500	50
RMSE Error	77.51	57.01
Max RMSE	233.49	255.36
Min RMSE	38.62	1.36
Mean SSIM	0.82	0.87
Min SSIM	0.58	0.60
Max SSIM	0.93	0.99

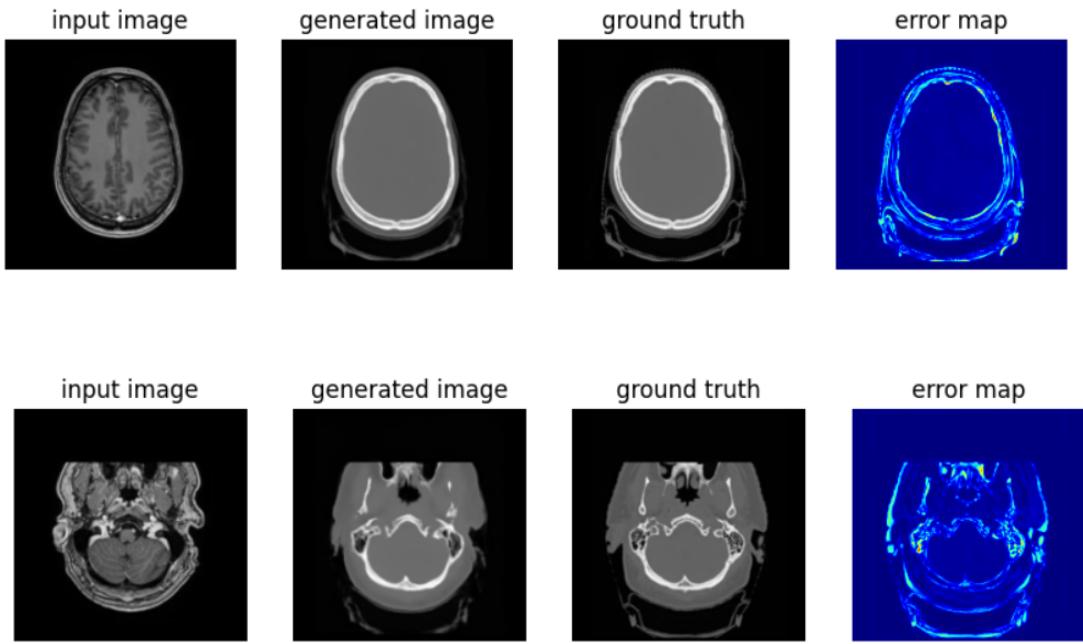


Figure 6.22: Results generated at 100 epochs and new learning rate

I have changed the Discriminator learning rate and Generator learning rate. There is also a change in batch size and I have run my model on 100 epochs. So, In experiment 2 i am seeing improvement in my model while using these parameters rmse went from 77.51 to 57.01 and mean ssim from 0.82 to 0.87 just by changing epoch and learning rate.

Increased the training data from 29286 images to 39048 images using the three transformations gamma ,gaussian blur and rotation each having equal probability and the performance of the model was evaluated at different epochs.

Table 6.13: RMSE and SSIM on Test Images with Increased Training Data and Different Epochs

Parameters	Epoch = 10	Epoch = 30	Epoch = 50
Mean RMSE Error	56.59	57.11	58.36
Max RMSE Error	249.93	321.68	248.06
Min RMSE Error	2.18	25.59	11.63
Mean SSIM	0.86	0.86	0.87
Min SSIM	0.61	0.40	0.57
Max SSIM	0.99	0.96	0.99

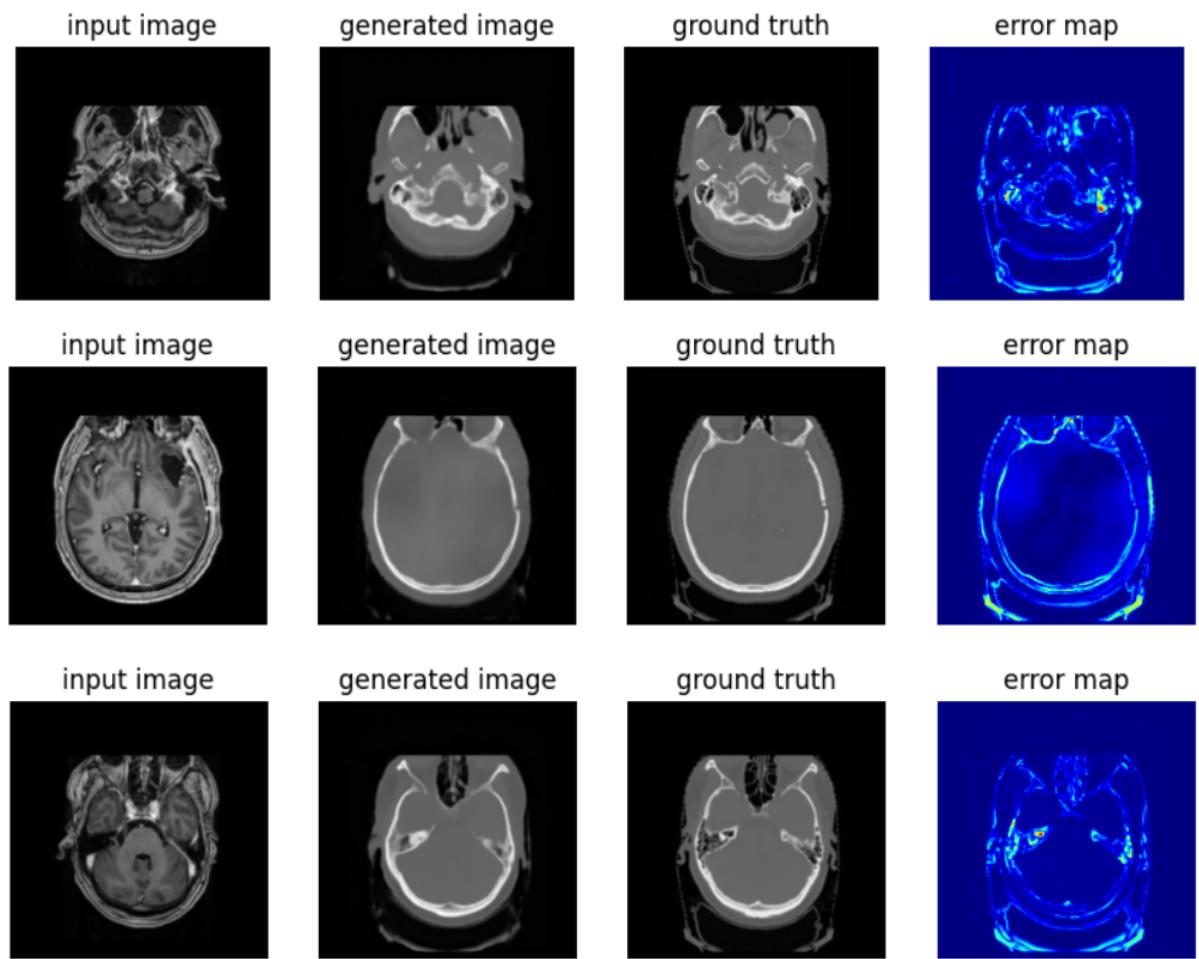
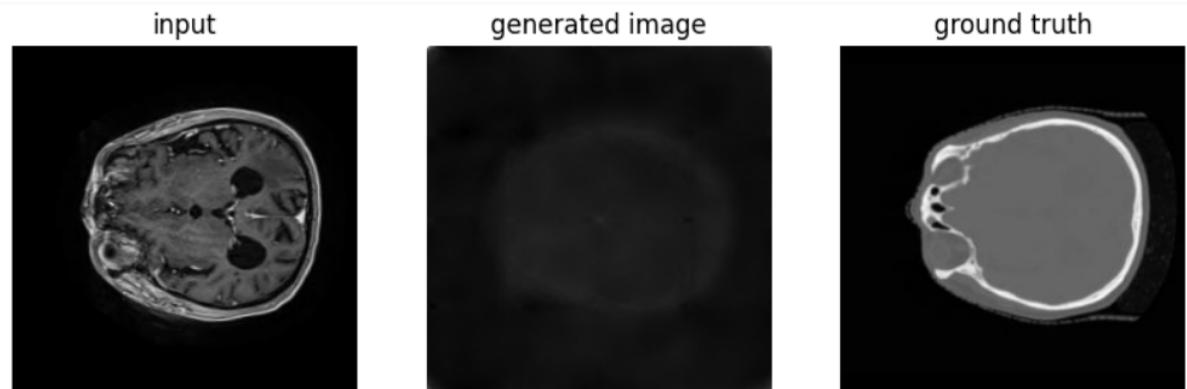


Figure 6.23: Results generated after applying data augmentation

Table 6.14: RMSE on Test Images Using Fourier Learning

Metric	Fourier Learning
Mean RMSE Error	13078.68
Max RMSE Error	21062.77
Min RMSE Error	21.95



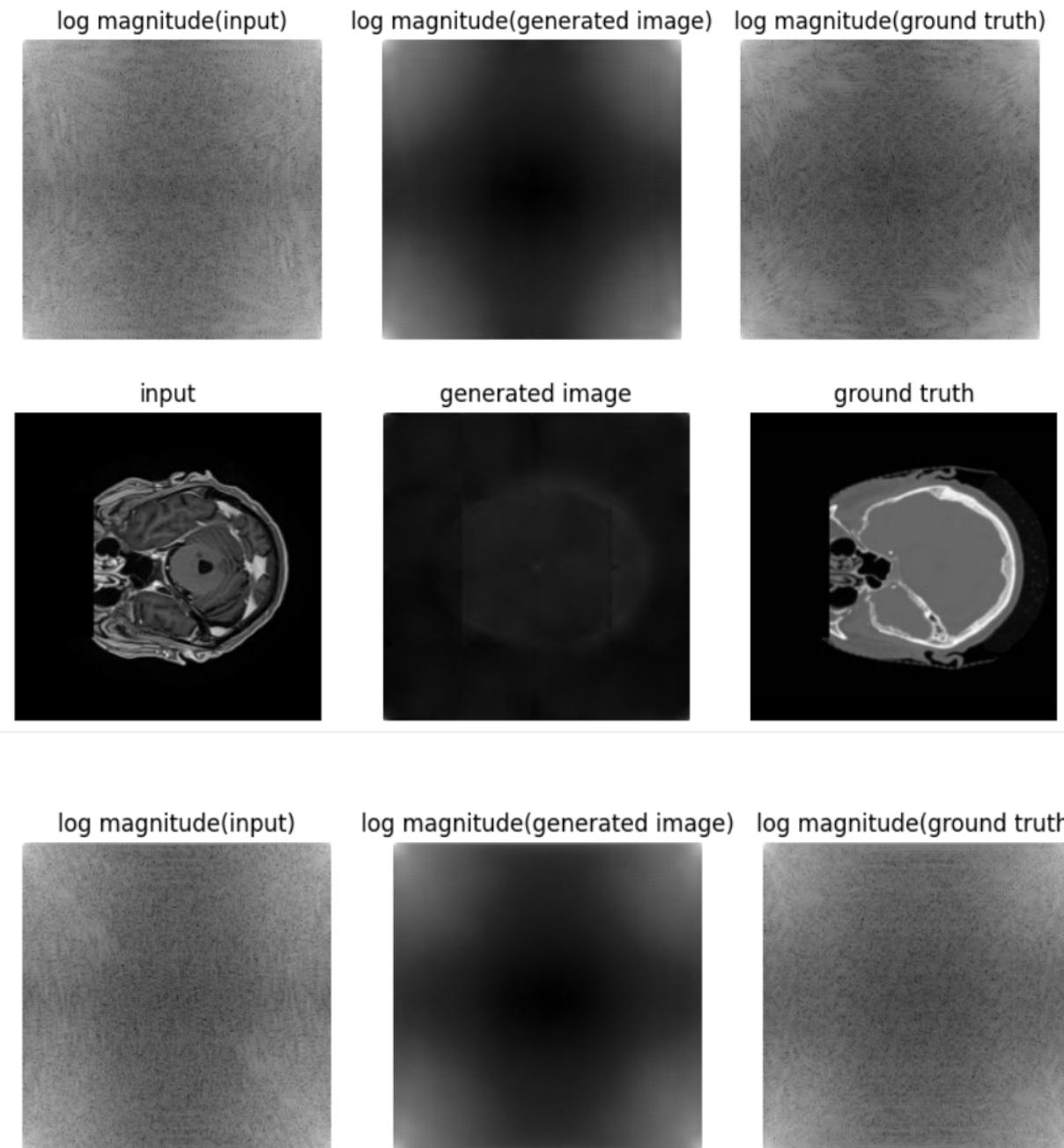


Figure 6.24: Generated image and log magnitude when trained using Fourier learning approach

Table 6.14 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 6.15: RMSE and SSIM on Test Images Using Patch-Based Learning Approach

Metric	Value
Mean RMSE error	61.23
Max RMSE error	227.57
Min RMSE error	4.48
Mean SSIM	0.84
Max SSIM	0.99
Min SSIM	0.63

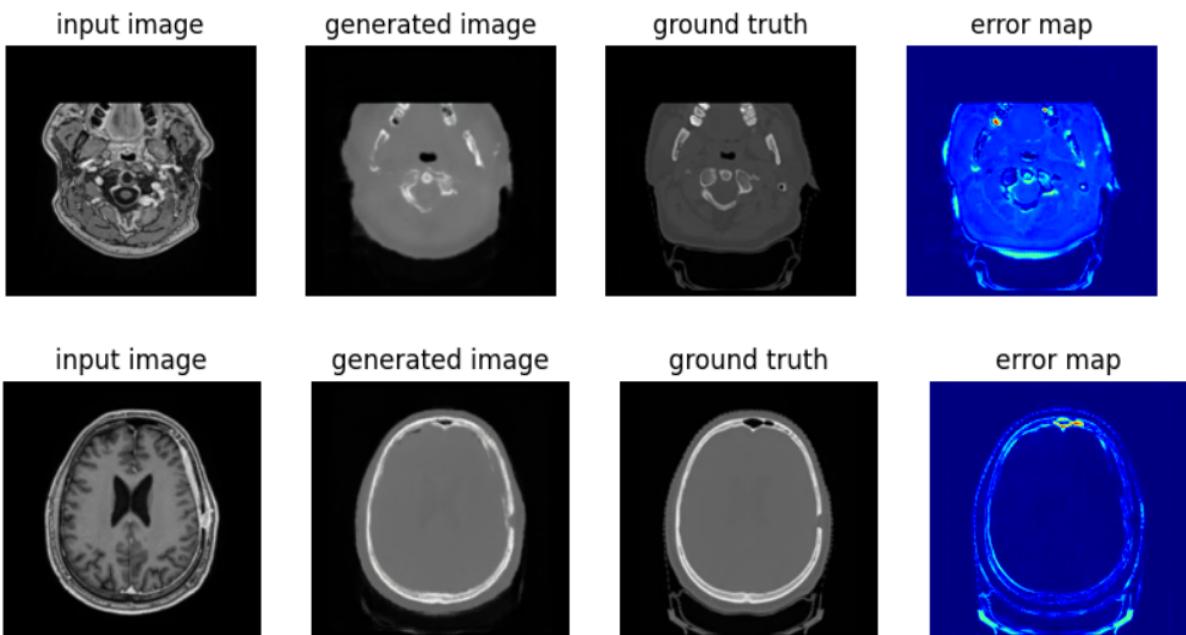


Figure 6.25: Generated image when trained using Patch-Based learning approach

Table 6.15 and the above figure are the results obtained when the patch-based approach was considered. The results obtained are satisfactory, providing more possibilities and applications in the future.

Table 6.16: Comparison Of PSNR 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
PSNR	22.147dB	20.168dB

Table 6.16 shows the comparison of PSNR between the two approaches. The results indicating both approaches were able to create synthetic images.

# Chapter 7

## Discussion and Conclusion

### 7.1 Discussion :

Various works on generating MRI sequences have been noted in the literature, but they often focus on multi-modal image translation tasks. These approaches typically produce a single output sequence with two or more input image sequences. However, we also found some papers that address single-modal image translation tasks. Here are a few key observations from the papers that worked with the pix2pix model:

- In the pix2pix generator, the U-net architecture was then replaced with U-Net++ [8] architecture to accommodate more features, skip connections, and layers than the previous one so as to learn more contextual information.
- In one study, differential image discriminators [8] were employed, 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 offers advantages, such as encouraging the generator to focus on capturing specific details or characteristics of the target domain.
- Several other works also proposed combining MI (Mutual Information) and GD (Gradient Difference) in the loss function to address misalignment problems and for better structural consideration.

A few key differences we observed between our work and existing work are:

- **BCE vs MSE Loss Comparison:** This study uses two loss functions, the binary cross entropy and mean square error, as a loss function and compares their results.
- **Hyperparameter Tuning:** Tuned the hyperparameters like lambda, batch size, and optimal training data size.
- **Data Augmentation and its Hyperparameters Tuning:** Implemented 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:** This study attempted to learn Fourier to Fourier mapping in the frequency domain, hoping the model will learn more minor and more precise details than the one learned in the spatial domain.

- **Patch-based Learning:** Used patch-based learning as a way of getting around this problem. It concentrates on creating small sections and portions of an image rather than the complete image. This yields a more localized understanding of texture and other features, which in turn facilitates constructing more realistic and detailed synthetic images.

[7] Some research papers are doing this patch-based learning because of data scarcity and localized better learning, but no one is doing it with the approach of making the model more robust for structural similarity and generalizing it to different human body parts.

**Best performance Model:** Analyzing the performance, it was clear that the model performs well when tested on synthetic data created through data augmentation. This technique helps the model apply knowledge from various training instances generated under different settings and parameters, enhancing its ability to perform well on new, unseen data. Additionally, increasing the number of epochs during model training leads to even better performance.

**Failure Analysis:** The model captures the outer structure but does not perform satisfactorily in some cases, such as sometimes it cannot capture fine details of the target image. Generally, we observe a loss in fine and localized details in the images. The study also attempts a patch-based approach to overcome this issue for better fine detailing and localized learning. However, the performance of this approach is satisfactory and can be further improved by tuning the hyperparameters.



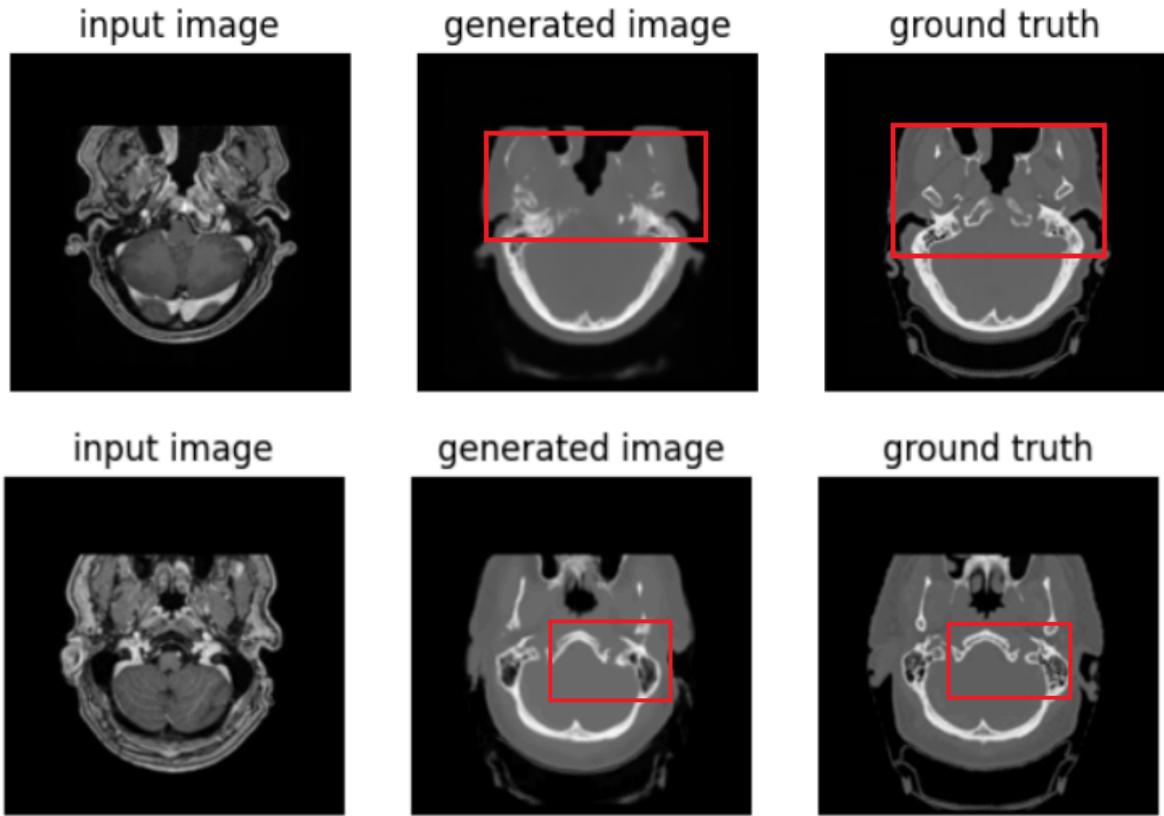


Figure 7.1: Segmented area where the model is slightly lacking in Real and Generated Images

## 7.2 Future Work :

This study explores the generation of CT images from T1-weighted sequences. To further validate the effectiveness and robustness of our approach, we can extend this methodology to other MRI sequence types like FLAIR, T2w, T2, etc. The evaluation in this study is conducted exclusively on the test dataset derived from the SynthRAD2023 Grand Challenge dataset. Consequently, if additional data from different sources becomes available, further evaluation can be performed to provide a more thorough evaluation of the model's generalization capability across diverse datasets, enhancing its robustness and applicability in various clinical settings.

Although the pix2pix GAN produced visually similar images to the CT scans, there remain several limitations and concerns of the problem. Firstly, pathology preservation needs to be ensured. For example, a fracture in a skull shown in the MRI needs to be reflected in the CT scan synthesized by the model. The pathology lost in translation poses a great threat to the patient's overall health and may result in poor diagnoses and treatment. Adding on to the previous point, this is a diagnostic AI, meaning that the AI's performance is one of the factors that will decide a patient's health outcome, so the evaluation and

training of said AI needs to be handled with much care and caution. The outcome of this project shall be translated into an Abstract/manuscript/copyright.

## 7.3 Conclusion

The results in Chapter 6 show that Pix2Pix Generative Adversarial Networks (GANs) effectively create CT images from brain MRI scans. The generated images look realistic and match the details seen in real CT scans. The quality of these images is high, and they are often hard to tell apart from actual CT scans. The model works well because it uses advanced data augmentation techniques to improve the training data. Sometimes the model misses fine details, which indicates more room for improvement before being applied to practical applications. However, these results suggest that GANs could help reduce the need for radiation in medical imaging and make clinical workflows more efficient. Future research should keep improving these techniques to ensure they are reliable and effective in medical settings.

## 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>