MRI Denoising & Super Resolutions with GANs

CS 598 Project Report

Abhishek Jain
Aj26@illinois.edu

Rahul Sharma
Rahul9@illinois.edu

Wenxi Fan
Wenxif2@illinois.edu

ABSTRACT

The goal of this course project is to explore the potential of deep learning in addressing practical problems of magnetic resonance imaging (MRI). MRI is one of the most important modern medical imaging technologies given its capability in probing rich set of biological information in vivo. However, MRI still suffers from several issues that impede its practical utility, including low signal-to-noise ratio (SNR) and limited spatial resolution. It is still an active research area in the MRI community to push the limit of the imaging capability. Recently, deep learning methods promised to be a powerful tool to advance MRI. In this project, we would like to investigate the possibility of using advance state-of-the-art deep learning architecture, Generative Adversarial Networks, to improve the sensitivity & resolution of MRI images and further compare its performance with traditional techniques.

Keywords

Magnetic Resonance Imaging, Deep Learning, Denoising, Super-Resolution, Generative Adversarial Networks.

1. INTRODUCTION

Magnetic resonance imaging (MRI) is a powerful imaging tool that can noninvasively probe a rich set of biological information, such as organ anatomy, functional activity, blood flows, as well as metabolic activities. Since its inception in 1970s, MRI has been widely used for applications in both clinics and basic science. For example, in clinical applications, MRI can be used for disease diagnosis, treatment assessment, and surgery planning. For scientific research, MRI are widely used to investigate the mechanisms underlying neuronal firing and tissue-level metabolic functions.

1.1 Challenges

Despite its powerfulness and popularity, current MRI methods still have several limitations. One major problem of MRI lies in its low detection sensitivity, which often results in rather low signal-to-noise ratio (SNR) in the acquired images as compared to other modern imaging technology (e.g., positron emission tomography (PET)). In addition, MRI is inherently a very slow imaging method, as compared to computed tomography (CT); hence, high-resolution imaging often requires very long data acquisition time, which is infeasible in practice, especially under clinical settings. Therefore, there are still quite a lot of efforts spent in the MRI community to push the performance limitations of the MRI systems, especially on improving the SNR (i.e., denoising) and spatial resolution (i.e., super-resolution).

1.2 Motivation

Image denoising and super-resolution are old topics in image processing field. Recently, deep learning has become promising in image processing analysis and processing, especially in computer vision community. The success of deep learning in improving the image quality is partially owing to the strong representation power of deep neural networks as well as the availability of large public database. It is a natural question that whether deep learning can be also applied to push MRI to the next level. In our project, we have explored the potential of deep learning in MRI image denoising and super-resolution. This opportunity has helped us in explore and render: 1) gain a better understanding of the powerfulness and potential of deep learning models in healthcare, which is the core of this project, 2) gain hands-on experience of handling practical medical image data, 3) improve our practical skills of data pre-processing, model building, model training, and evaluation.

1.3 Related Work

1.3.1 Traditional Techniques

MRI denoising & super resolution image generation techniques were traditionally based on interpolation-based methods (e.g. Bicubic Interpolation), filtering (e.g., Homomorphic Wavelet, Soft Thresholding), Genetic algorithm and Optimization algorithm. [2] Interpolation methods have failed to recover fine grain details from a low-resolution MRI scans and Optimization techniques required extensive understanding of data representation for modeling. [3]

1.3.2 Convolution Neural Networks

Deep learning techniques have been widely examined in the field of image processing and have seen ground-breaking results in super resolution applications. Recent studies have shown that the performance of MRI denoising can be improved significantly by using deep convolutional neural networks (CNN). Advance CNN architectures like Super Resolution SRNN & Deep Densely Connected Neural Networks DCNN has improvements on faster training, light weighted models and improved accuracy.[3] Since CNN mainly focuses on minimizing the mean squared error, it results in images which lacks high-frequency details.

1.3.3 Deep Modeling Methods

Deep modeling methods (Stacked Auto Encoder (SAE) and Deep Boltzman Machine (DBM)) have been used to segment, annotate MRI images and then learn the functional network of brain [7]. Another area of research is that minimizes the prolonged processing time of MRI images with the help of Deep Residual networks and this approach has achieved better results over the compressed sensing techniques [8]

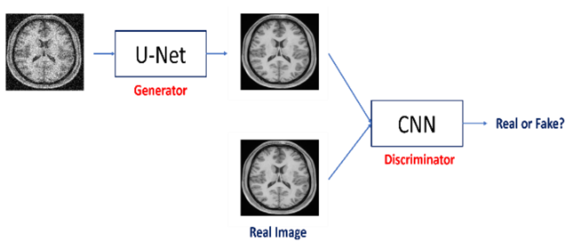
1.3.4 Generative Adversarial Networks

In the recent research Generative Adversarial Networks (GAN) have achieved great performance compared to other data driven methodologies [4][5]. Training with an adversarial network improves the CNNs ability to generate images with improved resolution and quality.[6]

1.4 State of the Art Solution

Recently deep learning community has developed CNN base architecture named UNet [11], which relies on the strong use of data augmentation to annotate the sample data more efficiently. The architecture is structured in two phases a) Contracting path to capture context and b) Expanding path that enables precise localization. Such a network can be trained end-to-end with very few images.

UNets can be augmented with a specialized GAN implementation scheme called Conditional Generative Adversarial Network (cGAN) [12]. This architecture formulation simplifies learning mapping from input image to output image and loss function automatedly which can be used for image to image translation.



2. DATA

The datasets we plan to use in this project are from the MGH-USC Human Connectome Project [1], which is an open database published by Massachusetts General Hospital (MGH) and University of Southern California (USC). The database contains MRI images acquired from 1032 subjects with multiple image modalities such as T1-weighted (T1W) images, T2-weighted (T2W) images, diffusion images, as well as functional MRI images. All these images were acquired from well-controlled experimental conditions and of excellent image quality, thus well-suited to be used as the training data. In this project, we would like to focus on the T1W images, which are commonly used by doctors for disease diagnosis. For each dataset, a 3-dimensional image set was obtained with image size of $260 \times 311 \times 300$ (corresponding to 24 million pixels). Figure 1 illustrates some sample images from this database.

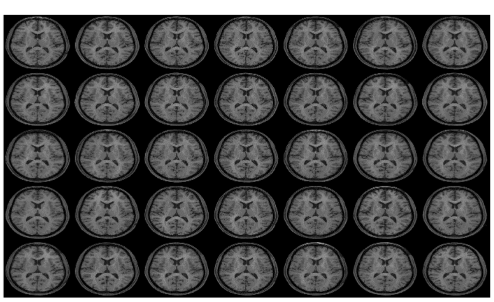


Figure 1. Sample images from the HCP database.

To examine the data quality, we have performed the following investigations:

a) Calculate the mean and standard deviations (STD) images across different subjects. The results are summarized in Fig. 2. As can be seen, the mean image still looks meaningful (with clear brain structures), indicating all the brain images are aligned well (which reduces the complexity of the learning task).

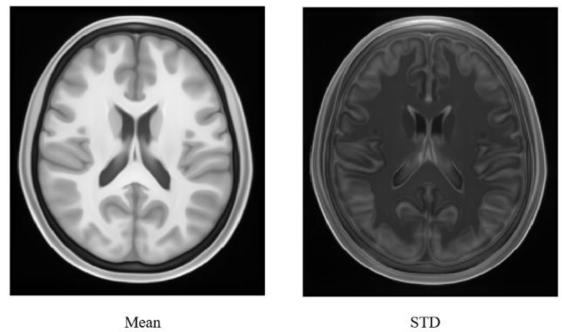


Figure 2. Mean and STD maps calculated across all subjects (1 slice).

b) Histogram plots of image intensities from different pixels and compare among different subjects. As shown in Fig. 3, the intensity distributions for different subjects are similar, which suggests learning from training data is feasible.

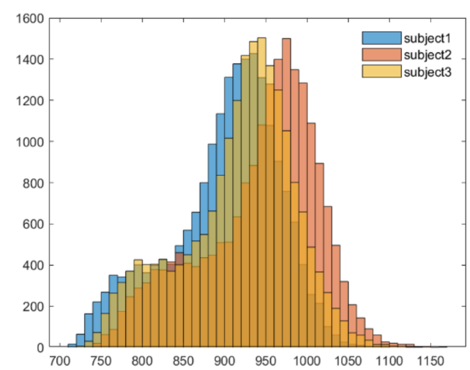


Figure 3. Intensity distributions of different subjects.

3. APPROACH

The goal of this project is to remove the additive noise of MRI images using an end-to-end deep neural network. Our project consists of 3 major stages: a) Data Pre-Processing – 2D noiseless MRI images are sliced from original 3D data and noisy 2D images are generated by adding noise to the original ones b) Model Training – setup & train a GAN based deep learning architecture with generator & discriminator components for image denoising tasks. c) Model Validation – trained generator model is used to generate denoised MRI images by passing noisy counterparts;

results are compared with baseline models. Below are some more details on these different stages:

3.1 Pre-Processing

We treat the raw MRI images from the HCP database as the ground truth (i.e., noiseless) given their excellent image qualities. The original images from HCP database are 3D images, which cannot be handled by common deep networks due to memory issue; therefore, we transformed them to 2D images for model training. Also further to setup data for experimentation both a high and low quality image pair is required for GANs training.

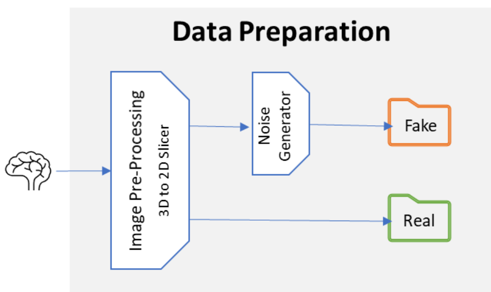


Figure 4. Data pre-processing components required for setting up the experiment.

These pre-processing requirements are addressed by below two components:

3.1.1 2D Image Slicer

2D images are extracted by treating every slice from each 3D dataset as an independent image. To save computational memory, we only focused on the central 15 out of 300 slices for each subject.

3.1.2 Low Resolution Image Generator

To generate low resolution images degraded images, we first add Gaussian distributed noise to the images to produce noisy images. All these operations are reasonably close to practical image perturbations. Some typical degrade images are shown in Fig. 4.

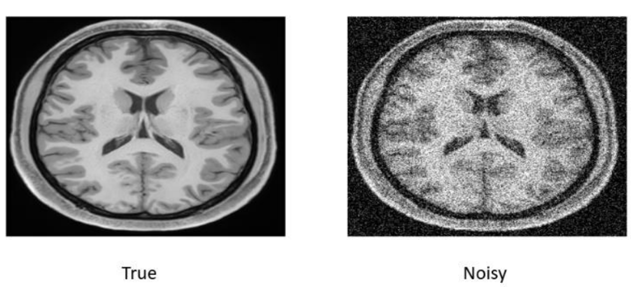


Figure 4. Typical degraded images generated from HCP database.

3.2 Model Training

In this project, we plan to use one of the state-of-the-arts GAN based neural network architecture in image processing community to address the image denoising and super-resolution problems.

The architecture contains two components: the generator and the discriminator. The generator takes the degraded image as the input and tries to produce the improved image that is as close to the ground truth image as possible. The discriminator is a component that distinguish whether an image comes from the real training data or generated from the generator. These two components essentially play a zero-sum game: the generator tries to fool the discriminator by generating as real images as possible, while the discriminator provides important critiques (or feedbacks) to the generator to push it improve its parameters.

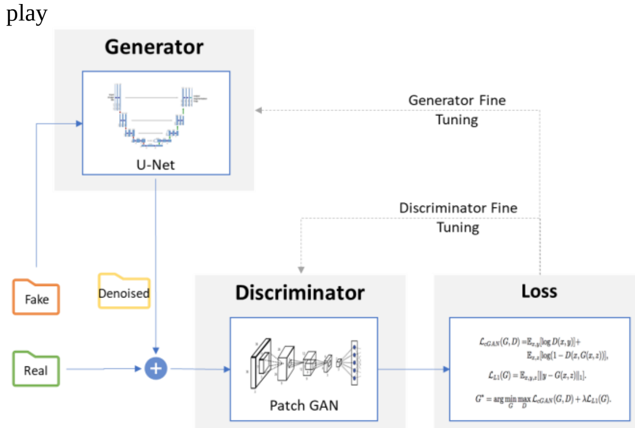


Figure 5. Model Architecture

3.2.1 U-Net based Generator

In this project, we plan to use U-Net [Fig 5] as the generator, given its capability in extracting multi-scale image features. The U-Net model is an encoder-decoder model for image translation where skip connections are used to connect layers in the encoder with corresponding layers in the decoder that have the same sized feature maps [11].

This architecture includes two sections, contraction and expansion. Both sections consist set of steps. The aim of contraction phase is to perform multichannel features extraction using sequence of convolution layers, each followed by nonlinear Leaky Relu activation function and then apply 2*2 max-pool on each channel. At each step we expand channels factor by 2. As an input we supply single channel greyscale noisy image [4.1.2] and expand channel from 1 to 128 to learn granular level image features.

Next Expansion phase, we use transposed convolution with the convolution step (same as contraction phase) to improve features resolution through back propagation. At each step image channel is reduced factor by 2 and finally it returns single channel grayscale image.

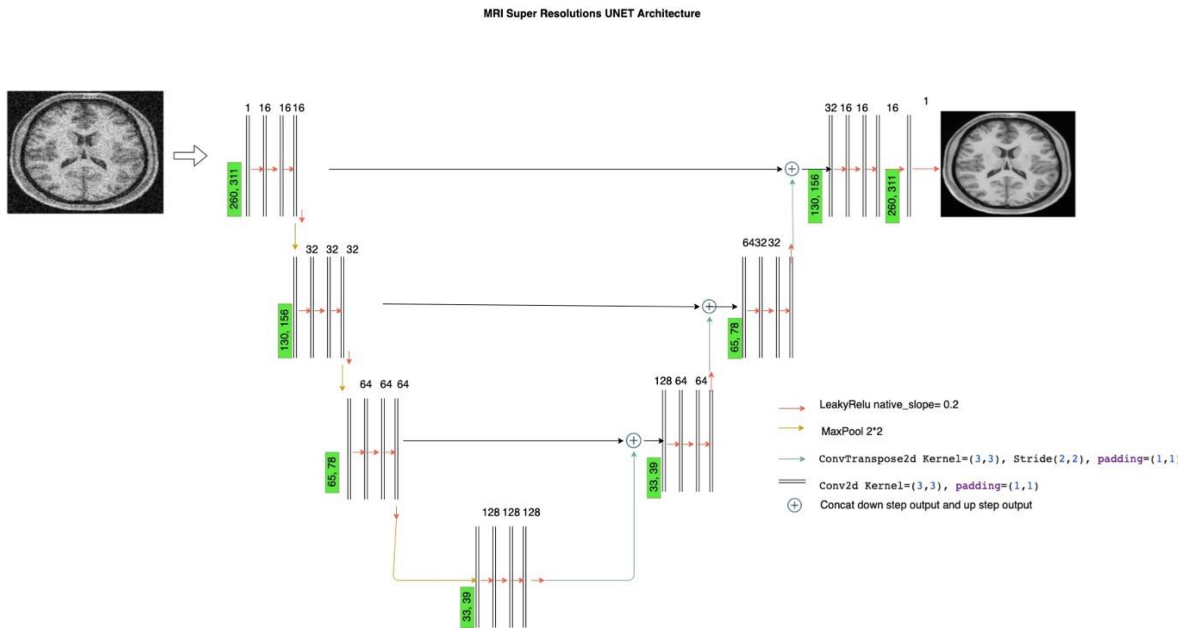
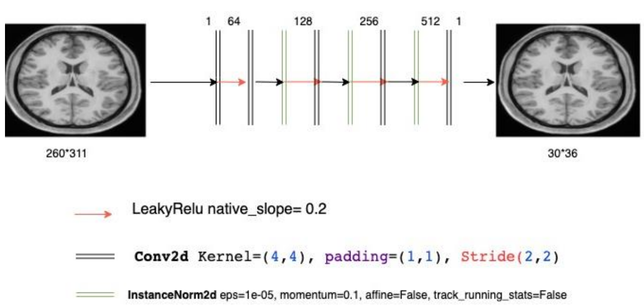


Figure 5. UNET Architecture

3.2.2 PatchGAN Discriminator

We implemented N-Layer-Discriminator using PatchGAN architecture, it is a type of discriminator for generative adversarial networks which only penalizes structure at the scale of local image patches. The PatchGAN discriminator tries to classify if each NxN patch in an image is real or fake. This discriminator scans across image convolutionally and further it averages responses to provide the classification result. It can be understood as a type of texture/style loss.



3.2.3 Loss Functions

Discriminator Loss

The discriminator model is trained in a standalone manner in the same way as a traditional GAN model, minimizing the negative log likelihood of identifying real and fake images, although

conditioned on a source image. The training of the discriminator is too fast compared to the generator, therefore the discriminator loss is halved in order to slow down the training process.[12]

Generator Loss

The generator model is trained using both the adversarial loss for the discriminator model and the L1 or mean absolute pixel difference between the generated translation of the source image and the expected target image. The adversarial loss and the L1 loss are combined into a composite loss function, which is used to update the generator model.[12]

Conditional-Adversarial Loss (Generator versus Discriminator)

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,z}[\log(1 - D(x,G(x,z)))],$$

L1-loss function is denoted by:

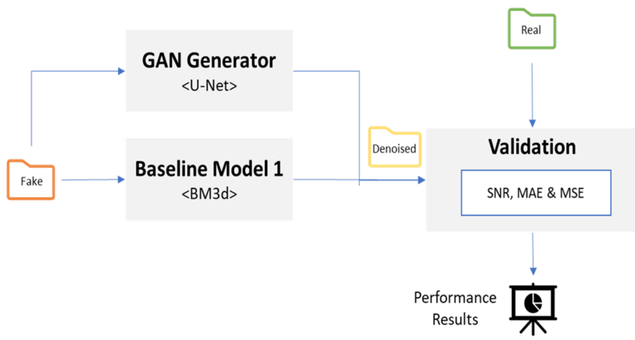
$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x,z)\|_1].$$

Combining these two functions together:

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda\mathcal{L}_{L1}(G).$$

3.3 Model Performance Validation

To evaluate the performance of the deep model, we will compare with the traditional image denoising methods such as Gaussian filter-based method and BM3d as well as traditional image deblurred method.



3.3.1 *Baseline 1: Block Matching & 3D Filtering*

BM3D is a denoising strategy based on an enhanced sparse representation in transform domain using a collaborative filter technique. Image enhancement is performed in three successive steps: 3D transformation of a group, shrinkage of the transform spectrum, and inverse 3D transformation. The result is a 3D estimate that consists of the jointly filtered grouped image blocks. By attenuating the noise, the collaborative filtering reveals even the nest details shared by grouped blocks and at the same time it preserves the essential unique features of each individual block.[9]

For the model comparison and evaluation, we tested a range of parameters and chose best performing model with sigma as 0.1

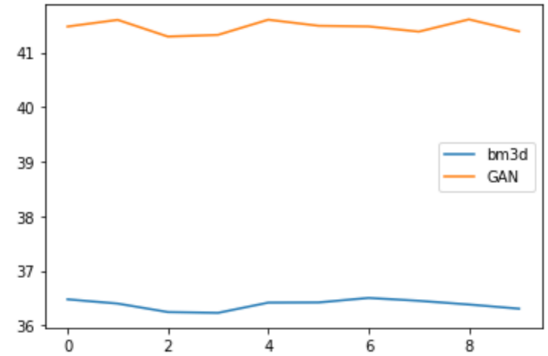
3.3.2 *Target Model: U-Net based Generator*

Approach already discussed in section 4.2

3.3.3 *Performance Metrics*

We will use the following performance metrics for a quantitative evaluation:

- Peak signal-to-noise ratio (PSNR) as compared with the ground truth.
- Mean absolute error (MAE) as compared with the ground truth.
- Mean squared error (MSE) as compared with the ground truth.
- Model runtime: the time takes



to train and predict.

4. EXPERIMENTAL SETUP

We are using Pytorch and the python machine learning packages such as Sklearn, in addition to the basic packages used for data analytics (e.g., NumPy, Scipy ,Pandas, etc.).

We are using Google Colab for the model training and performance evalution.

5.1. Model Training

We inherit the initial setup from Pix2Pix[13] and implemented UNet generator. We extracted 90 images from HCP dataset [3] for training and trained model with 32 batch_size and 2k epoch on Google Colab.

Experiment stats:

GPU: Nvidia P100

Network G Structure: Unet

Network D Structure: PatchGAN

Input Size: 900

Batch size: 32

Epoch: 400

Network initialized parameters:

Network G: 1.025 M

Network D: 2.763

Complete training stats are available here.

5.2. Model Results

One main objective is to evaluate if our model can produce realistic denoised images with the similar properties to the test label. We compared the performance between the baseline model (BM3D) and GAN on 10 test images, with performance measured by PSNR, MAE, MSE. The results are as follows:

5.2.1 The Comparison of the Overall Metrics

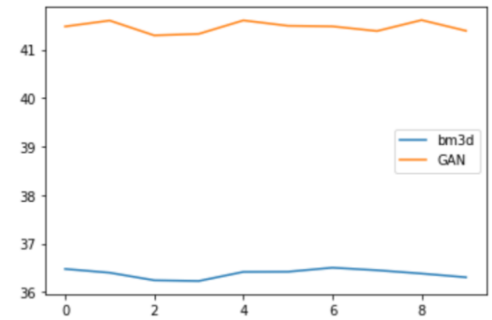
Model	PSNR	MAE	MSE

BM3D	36.380832	0.009840	0.000230
GAN (UNet)	41.464526	0.005844	7.1398e-05

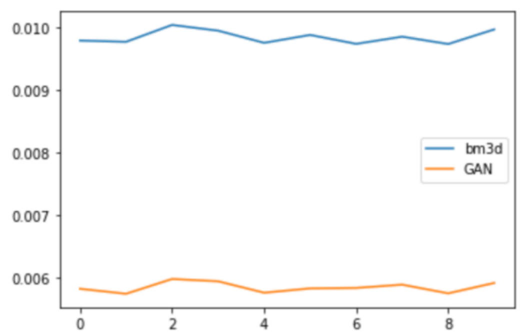
*The metric value was calculated as the average of the 10 test objects.

5.2.2 The Comparison by Test Object

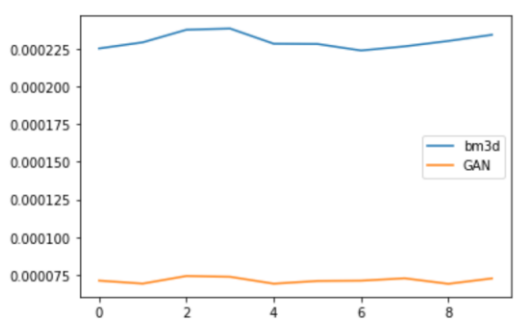
- The PSNR of the 10 test objects by model



- The MAE of the 10 test objects by model

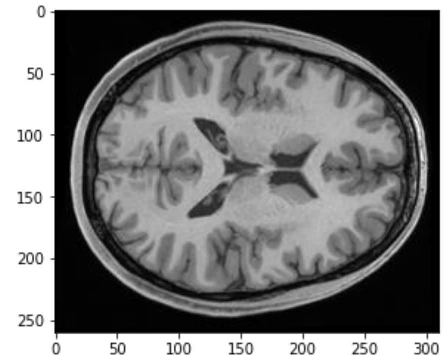


- The MSE of the 10 test objects by model

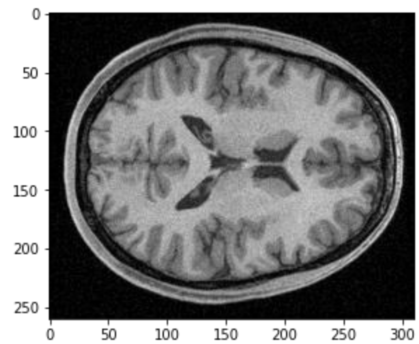


5.2. Sample Results:

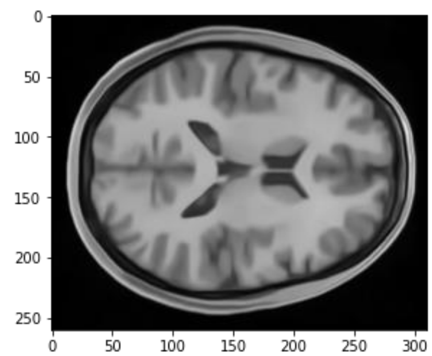
Below is a sample output demonstrating improvements GAN based model achieves and its comparison to baseline model.



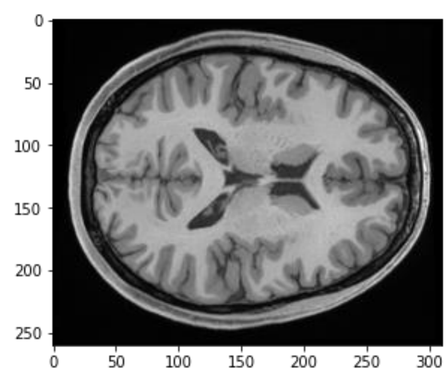
Original Image



Fake Image (Input)



BM3d (Baseline Output)



GAN (Output)

5. CONCLUSION

Our initial experiment results suggest GAN based approach to MRI denoising & super resolution has considerable improvements and outperforms traditional BM3D based approach. It has been observed using both quantitative measures & validating images visually as discussed in section [5].

6. REFERENCES

[1] Path to the dataset used for this project: http://www.humanconnectomeproject.org/data/

[2] Prabhpreet Kaur, Gurvinder Singh a`nd Parminder Kaur. A Review of Denoising Medical Images Using Machine Learning Approaches. Current Medical Imaging Reviews, 2018, 14, 675-685.

[3] Yuhua Chen, Yibin Xie, et al. Brain MRI Super Resolution using 3D Deep Densely Connected Neural Network

[4] Yuhua Chen, Feng Shi, Anthony G Christodoulou, et al. Efficient and Accurate MRI Super-Resolution using a Generative Adversarial Network and 3D Multi-Level Densely Connected Network.

[5] Guang Yang et al. Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction

[6] Jelmer M. Wolterink et al. Generative Adversarial Networks for Noise Reduction in Low-Dose CT

[7] Dinggang Shen, Guorong Wu, Heung-ll Suk. Deep Learning in Medical Image Analysis.

[8] Dongwook Lee, Jaejun Yoo, Sungho Tak and Jong Chul Ye. Deep Residual Learning for Accelerated MRI using Magnitude and Phase Networks.

[9] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3D transform-domain collaborative ltering

[10] Ian J. Goodfellow et al. Generative Adversarial Networks

[11] Olaf Ronneberger, Philipp Fischer, Thomas Brox U-Net: Convolutional Networks for Biomedical Image Segmentation

[12] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks