CS230

Denoising Diffusion MRI

Ariel J. Hannum
Department of Bioengineering
Stanford University
ahannum@stanford.edu
SUNET ID: ahannum
Computer Vision

Abstract

Diffusion-weighted Imaging (DWI) is an MRI technique that maps the diffusion process of molecules to microstructure in biological tissue, which is limited by image resolution and the data size. Here, we utilized a deep learning model to infer fractional anisotropy maps from few diffusion weighted images.

1 Introduction

Diffusion MRI is an imaging technique that capitalizes on the sensitivity of MRI to the diffusion (movement) of water molecules in order to infer the microstructure of tissue [1]. Diffusion-weighted images are acquired via MRI and are implemented in a diffusion tensor model to infer tissue properties such as molecular diffusion rate or the direction of diffusion (Fractional Anisotropy - FA)[2], which can be indicative of underlying health. In order to obtain more complicated metrics, many more diffusion directions are required, which will increase scan time. In addition, the diffusion decreases the overall signal across a voxel, so these images can also be very noisy [3]. Deep learning has been proposed as a method to mitigate the shortcomings of traditional diffusion MRI reconstruction, denoising, and the tensor model.

In this project, we implemented a U-Net convolutional neural network in order to create a high quality, reduced noise, Fractional Anisotropy map (FA) from fewer diffusion directions [4]. The inputs to this model are 6 optimized Diffusion-Weighted MRI images and one non-diffusion weighted MRI image. The output to this model is then a high-quality FA Map, which could be indicative of the underlying tissue health.

2 Related Work

Obtaining higher-resolution diffusion information and the application of machine learning have been jointly investigated over the past several years. To traditionally estimate Diffusion Tensor metrics, non-ML algorithms are used, but estimating parameters with noisy data suffers inaccuracies [5, 6]. Different approaches that denoise and improve metric computation have emerged that utilize the deep learning framework. Some approaches focus on denoising the MRI images themselves and then apply these images to the existing model to calculate diffusion metrics, such as in a self-supervised learning model[5] or 3D convolutional neural networks (CNN) [7]. Other approaches output the diffusion parameter maps themselves [8] [4]. We compare our networks to previous work mapping sub-sampled diffusion directions to higher quality diffusion parameter maps. [4]

CS230: Deep Learning, Fall 2021, Stanford University, CA.

3 Data Features

The dataset we have used is constructed from a 760µm open source diffusion dataset. [9] This unique dataset contains 18 separate Brain diffusion MRI files of 760 µm resolution. Each file has 2D slice matrix size 292 x 288, 192 slices, and 141 diffusion directions + 16 non-diffusion weighted images. The dataset has already gone through several MRI post-acquisition processing steps and brain masking. For the purposes of this project, we have down sampled the 760 µm data to be a 128x128 matrix size (84 slices), via Fourier Cropping.

We used one Brain diffusion MRI file (128x128 matrix X 84 slices X 141 diffusion directions + 16 non-diffusion weighted images) for training, testing, and validation. Training inputs were generated by selecting one diffusion direction at random and then selecting the five other diffusion directions using an optimization algorithm [7] and then selecting one of the 16 random non-diffusion weighted images. The optimization algorithm yielded 87 optimal diffusion direction combinations. Thus in total giving 7,308 samples in total. 20 random diffusion directions were then selected from the 141 diffusion directions and underwent the DiPy tensor reconstruction [5] to generate ground truth samples. Therefore, we had 5040 training (75%), 840 testing (12.5%), and 840 validation examples (12.5%).

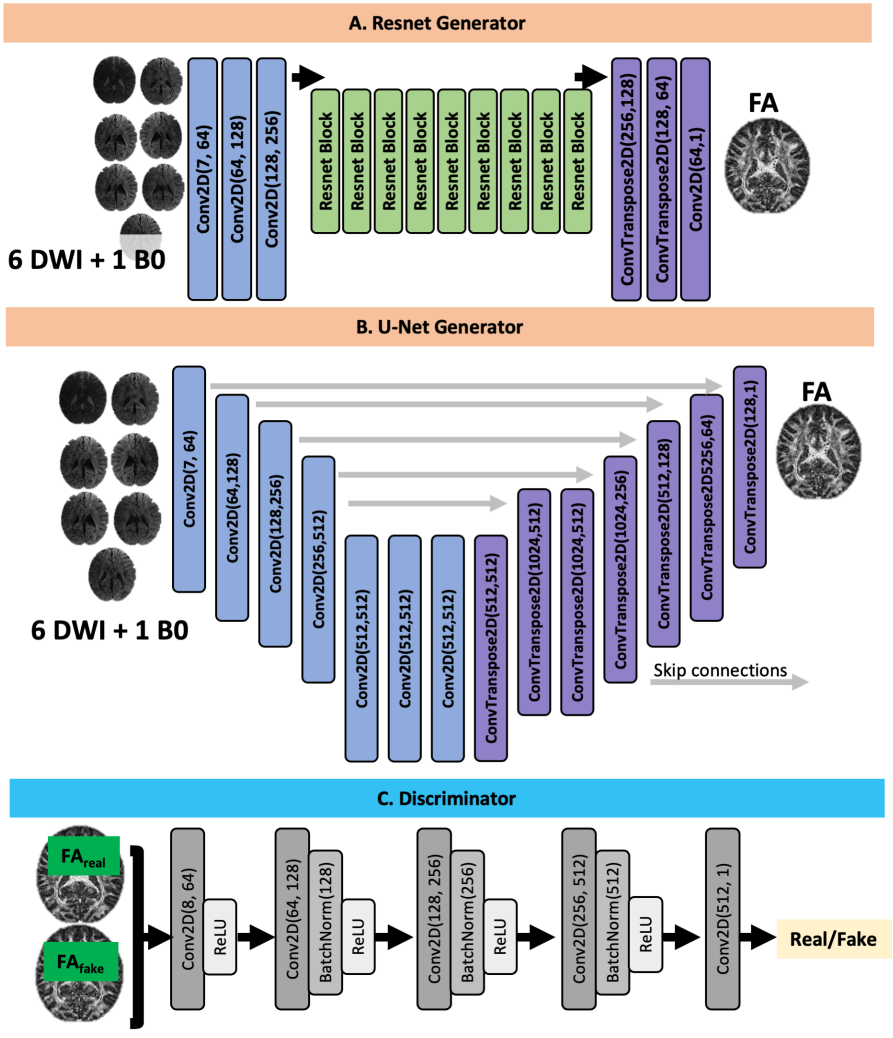

Figure 1: The different deep learning models that were evaluated: (A) Resnet and (B) U-net were evaluated seperately and as generators in a GAN model. (C) Discriminator layout utilized in GANs.

4 Methods

The models were implemented using the Pytorch library. Training was performed on 1 GPU NVIDIA Titan RTX, 24 GB. The number of input channels into the model was 7 and the number of output channels were 1. The number of iterations was 100 epochs and optimizer was Adam. A batch size of 1 was used for all training cases as GAN framework [10] in previous implementations had worked best with this batch size number. The pixel wise loss, the L1 distance between real and fake images, were computed for the Resnet and Unet frameworks, between the real image, FA_{real}, and the fake image, FA_{fake}: $(E|FA_{real} - FA_{fake}|)$, where E is the expected value. The adversarial loss, the squared log likelihood, was also incorporated:$[E(Input, FA_{real}) - 1]^2 - [E(Input, FA_{fake})^2]$.

4.1 Resnet

The Resnet has three convolutional encoders, batch normalization and ReLU activation function. A residual network follows, containing 9 Resnet Blocks made of 2D Convolution, Batch Normalization, ReLU activation function, and Dropout segments. Deconvolution blocks make up the decoder,which contain ReLU and Batch normalization. The output image is FA map. map.[11]

4.2 U-Net

The U-net FA generator has an architecture of a downsampling and upsampling path with 4x4 convolutional operators, batch normalization, and Reactivation function. Skip connections are included to aid in merging features together.

4.3 GAN

Both the Resent and U-net are implemented as a Generator in a Generative Adversarial Network (GAN) [10].This GAN model was previously implemented to leverage information from different images contrasts to synthesize clinical contrast images, such as T1, T2, and PD-weighted. The GAN leverages two neural networks in order to produce a realistic output image. The Generator network was either the Resnet or U-Net architectures described above, which train on 6 DWIs and 1 non-diffusion-weighted image to generate FA maps of the same quality as utilizing 20 diffusion directions. The produced image is then compared to the ground truth FA map (generated from the diffusion tensor reconstruction with 20 directions) using a discriminator network. This discriminator included four convolutions with 4x4 kernals, leaky ReLu activation function, and batch normalization. The adversarial loss (squared log-likelihood loss above) was computed for both the generator and discriminator between the real FA image, FA_{real}, and the fake image).

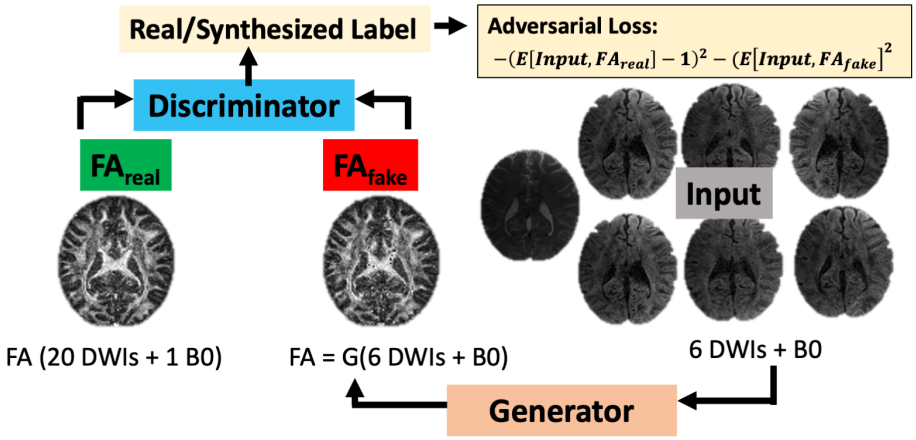

Figure 2: Overview of GAN network, where two neural networks contest with one another. The input to a generator (the Resnet or U-net) are 6 DWIs and 1 non-diffusion weighted image. The "fake" FA map produced by the discriminator network (a classifier) is then evaluated against a "real" FA map,generated by 20 DWIs.

4.4 DiffNet: Pretrained U-Net

A pretrained U-Net for slice-to-slice mapping [4] was used for a comparison to current literature. The model also used 6 diffusion directions and 1 non-diffusion image as input into the model to generate high resolution FA maps. The architecture used has a down sampling and upsampling path with 3x3 and 2x2 convolution operators, exponential linear activation function, and a max pooled operator. The upsampling path has 2x2 convolutional operators followed by 3x3 with exponential linear unit activation functions. Skip connections are included to help merge together features.

4.5 Analysis

For assessing the training over the epochs, L1 loss ($|FA_{real} - FA_{fake}|$), Mean Squared Error (MSE) ($\frac{1}{n}\sum_{i=1}^{n}(FA_{real} - FA_{fake})^2$), Peak Signal-to-Noise Ratio (PSNR) ($10log_{10}\frac{(L-1)^2}{MSE}$) and Structural Similarity Index (SSIM) ($\frac{(2_{real}fake+c_1)(2_{real,fake}+c_2)}{(^2_{real}+^2_{fake}+c_1)(^2_{real}+^2_{fake}+c_2)}$) were computed using Scikit Image Processing. [12] [13]

As further evaluation, a separate Brain Diffusion MRI file (128x128 matrix X 84 slices X 141 diffusion directions) were tested with each of the networks. The dataset was upsampled to 1.73 mm iso resolution and only six optimized diffusion directions were given to the networks. Gold Standard FA reference images are generated from the 141 diffusion directions with the DiPy [5] diffusion tensor reconstruction. Absolute differences between network output and the Gold Standard were computed.

5 Experiments/Results/Discussion

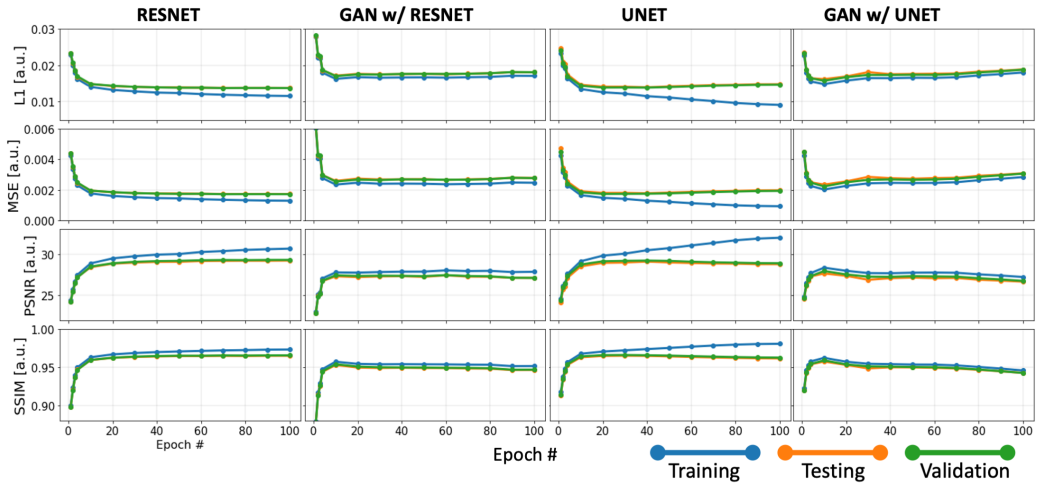

Figure 3: Quantitative plots of mean L1, MSE, PSNR, and SSIM for the different deep learning architectures during Training, testing, and validation for 100 epochs. Convergence is achieved in all examples. Non-GAN models achieve higher SSIM and lower errors, but have larger differences in training versus testing and validation curves.

5.1 Training Neural Networks

We successfully trained, validated, and tested each of the four deep learning models. All final training was had 100 total epochs; however, 50 epochs had a constant learning rate (0.0002), and 50 epochs had a decreasing learning rate (from 0.0002 to 0). The rate for cycle loss was set to 100. These parameters were selected as they had preliminary shown promise in this model in early results and had been utilized in a previous application of this architecture. [10]

For both GAN models, convergence was reached before the 5th epoch, in which the loss oscilated around a steady state. This behavior would be very typical of a GAN model as GANs have decreased

stability. The non-GAN architectures have lower PSNR-SSIM but may actually have higher image quality. Thus, it is likely that non-GANs have smoother images, but the GAN may have more realistic images.

5.2 Evaluation to Gold Standard Data

As further analysis, both the networks and the conventional tensor reconstructions were compared to a reconstruction from 141 diffusion directions, which we regard as the "Gold Standard". The Deep learning models have reduced difference error in comparison to the 6 direction FA map. There is large noise reduction in these models as well. Structures are more well defined in the deep learning models as opposed to the conventional tensor reconstruction for six directions. The non-GAN models have lower difference error in comparison to the GAN models. The GAN networks have smoother FA maps in comparison to the non-GAN cases. Interestingly, the FA map that was produced from DiffNet [4] was much smoother than the FA maps produced from our Deep learning models. This is likely due to the model being trained on multiple subjects, such that the resultant FA maps are smoother.

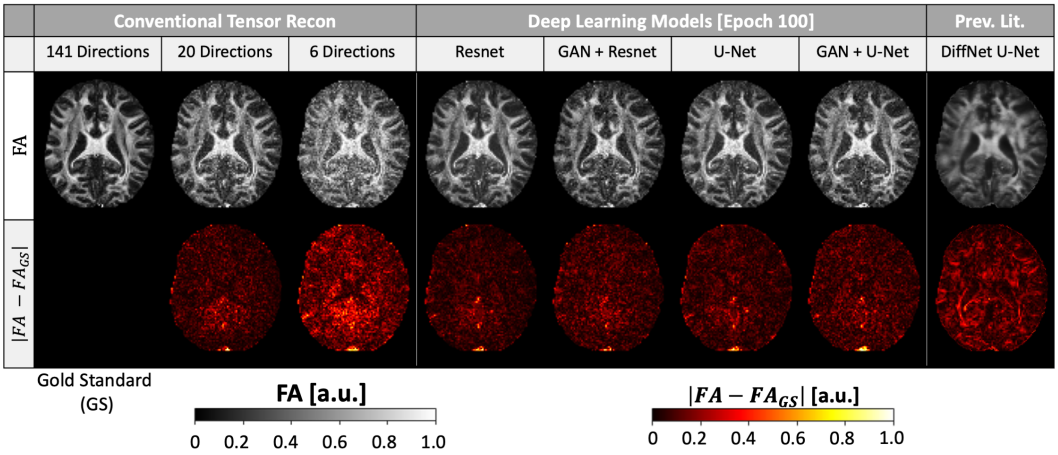

Figure 4: Qualitative results of the generated FA maps to the conventional diffusion tensor reconstruction and a previously published U-Net model.[4] Networks have reduced noise in comparison to conventional tensor recon with six directions and less blurring in comparison to the DiffNet.

6 Conclusion and Future Work

In conclusion we have evaluated the performance of U-Net, Resnet, and GAN architectures on producing high quality FA parameter maps from fewer DWIs. We have observed that Resnet, U-net, and a GAN framework with these models does perform comparatively to the conventional reconstruction from 20 diffusion images.

Future directions would include increasing the number of diffusion directions to generate ground truth data, training the existing models on higher resolution data (760µm) and evaluating the performance through similar analysis. It would also be interesting to investigate how the models perform with only three diffusion directions provided, as we cannot obtain FA parameter maps from three diffusion directions in conventional analysis. The networks could also be trained on exampls of other diffusion parameter maps, such as Mean diffusivity (MD), Eigenvectors, solving 3-way crossing fibers, or Neurite Orientation and Dispersion Parameters (NODDI).

We anticipate that this work provides a foundation for evaluating existing denoising techniques and further improving upon the computation of diffusion tensor metrics with reduced diffusion directions at higher resolutions.

7 Contributions

Ariel:

- Processed training data by upsampling image resolution
- Trained, tested, and validated the four different model architectures,
- Analyzed network performance compared to recent literature and traditional techniques
- Summarized recent literature and experimental results in project report and presentation.

There are several lab members I would like to acknowledge:

- Mahmut Yurt: Insight into network implementation, base repository to build generators and GAN models architectures
- Julio Oscanoa: Project idea generation, configuring lab resources to use for configuring the dataset and training the models

I would also like to credit my advisors, Professors Daniel Ennis and Kawin Setsompop, for their feedback.

References

[1] Jennifer Shane Williamson Campbell and Gilbert Bruce Pike. Diffusion magnetic resonance imaging. In *Encyclopedia of Biomedical Engineering*, volume 1-3, pages 505–518. Elsevier, 1 2019.

[2] Samantha J. Holdsworth, Rafael O'Halloran, and Kawin Setsompop. The quest for high spatial resolution diffusion-weighted imaging of the human brain in vivo, 4 2019.

[3] Denis Le Bihan, Cyril Poupon, Alexis Amadon, and Franck Lethimonnier. Artifacts and Pitfalls in Diffusion MRI. *J. Magn. Reson. Imaging*, 24:478–488, 2006.

[4] Eric Aliotta, Hamidreza Nourzadeh, and Sohil H. Patel. Extracting diffusion tensor fractional anisotropy and mean diffusivity from 3-direction DWI scans using deep learning. *Magnetic Resonance in Medicine*, 85(2):845–854, 2 2021.

[5] Eleftherios Garyfallidis, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan van der Walt, Maxime Descoteaux, and Ian Nimmo-Smith. Dipy, a library for the analysis of diffusion MRI data. *Frontiers in Neuroinformatics*, 8(FEB), 2 2014.

[6] FSL - FslWiki.

[7] Qiyuan Tian, Berkin Bilgic, Qiuyun Fan, Congyu Liao, Chanon Ngamsombat, Yuxin Hu, Thomas Witzel, Kawin Setsompop, Jonathan R. Polimeni, and Susie Y. Huang. DeepDTI: High-fidelity six-direction diffusion tensor imaging using deep learning. *NeuroImage*, 219:117017, 10 2020.

[8] Eric K. Gibbons, Kyler K. Hodgson, Akshay S. Chaudhari, Lorie G. Richards, Jennifer J. Majersik, Ganesh Adluru, and Edward V.R. DiBella. Simultaneous NODDI and GFA parameter map generation from subsampled q-space imaging using deep learning. *Magnetic Resonance in Medicine*, 81(4):2399–2411, 4 2019.

[9] Fuyixue Wang, Zijing Dong, Qiyuan Tian, Congyu Liao, Qiuyun Fan, W. Scott Hoge, Boris Keil, Jonathan R. Polimeni, Lawrence L. Wald, Susie Y. Huang, and Kawin Setsompop. In vivo human whole-brain Connectom diffusion MRI dataset at 760µm isotropic resolution. *Scientific Data 2021 8:1*, 8(1):1–12, 4 2021.

[10] Salman U.H. Dar, Mahmut Yurt, Levent Karacan, Aykut Erdem, Erkut Erdem, and Tolga Cukur. Image Synthesis in Multi-Contrast MRI with Conditional Generative Adversarial Networks. *IEEE Transactions on Medical Imaging*, 38(10):2375–2388, 10 2019.

[11] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9906 LNCS:694–711, 3 2016.

[12] scikit-learn: machine learning in Python — scikit-learn 1.0.1 documentation.

[13] Stéfan Van Der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. Scikit-image: Image processing in python. *PeerJ*, 2014(1), 2014.

A Additional Network Figures

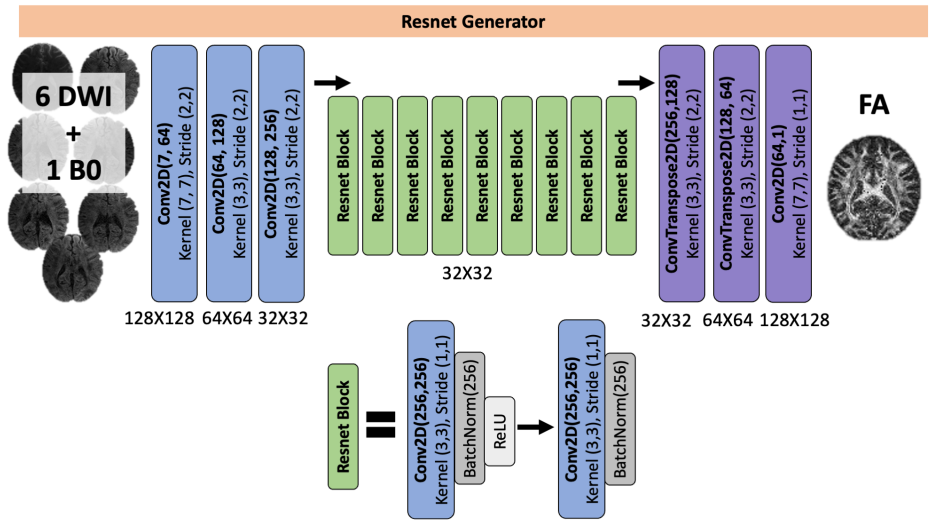

Figure 5: In-depth diagram of Resnet framework. Each Resnet block has a convolution operator followed by batch normalization and ReLU activation function.

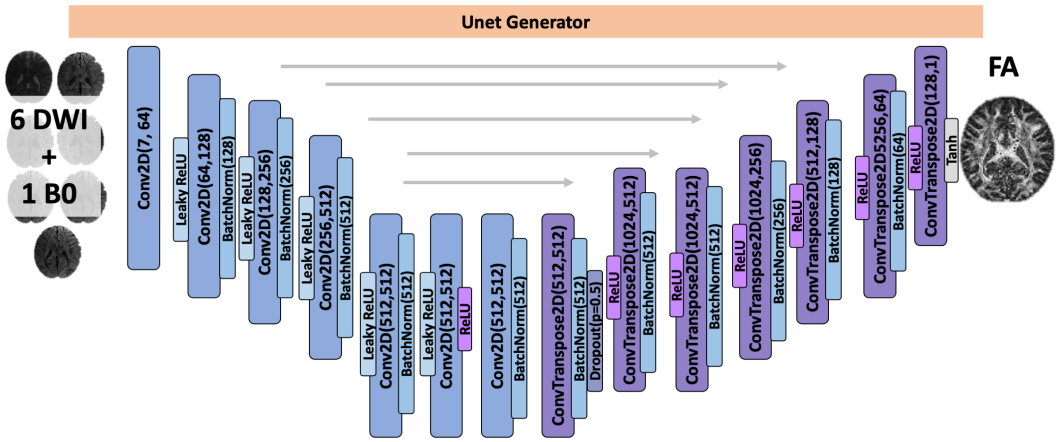

Figure 6: In-depth diagram of U-net framework. Convolutional kernals are preceded by leaky ReLU and followed by batch normalization while in the up sampling path the kernels are preceded by ReLU activation and followed by Batch Normalization. Tanh activation function is at the end.

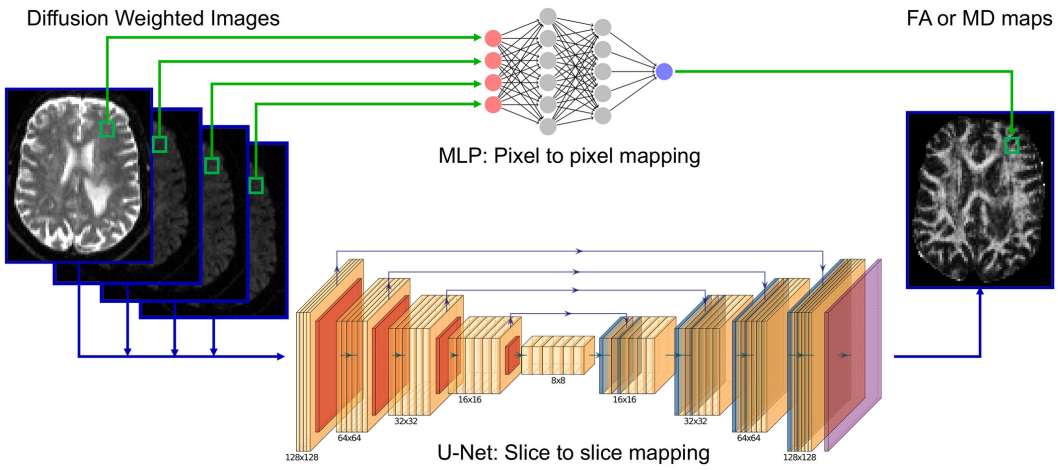

Figure 7: Overview of Diffnet U-net framework that was previously published.[4]

B Additional Results

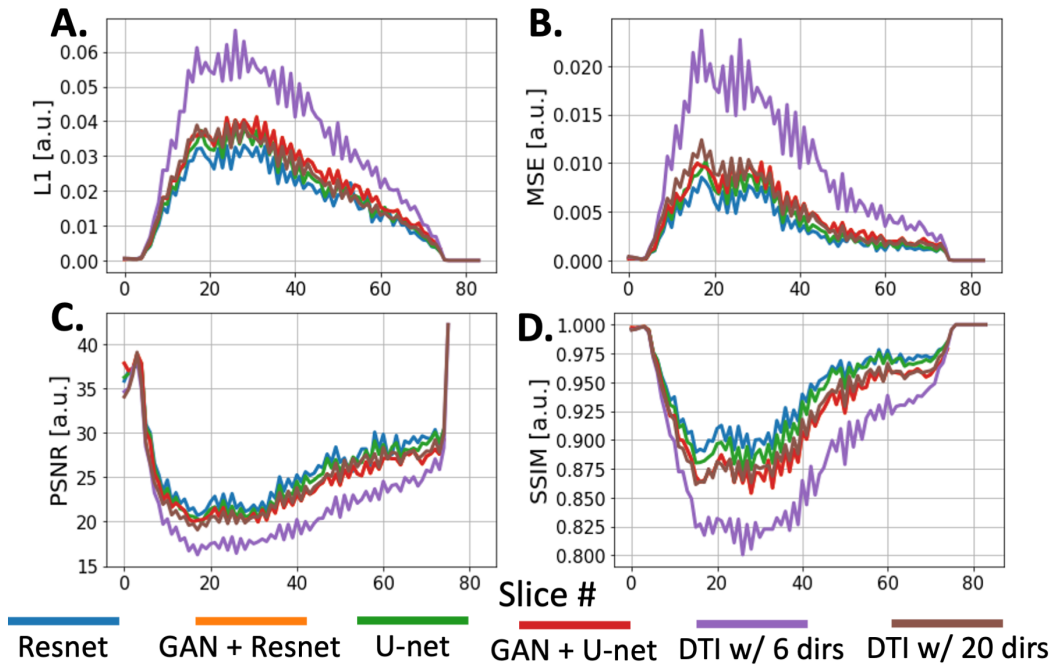

Figure 8: Further analysis of evaluation to Gold Standard Data. (A) L1, (B) MSE, (C) PSNR, and (D) SSIM as a function of the slice number. It is expected that most extreme slices would contain very little information. Excluding these early values, it appears that there is increased errors at lower slice positions relative to upper slice positions. Network performance in minimizing error and maximizing PSNR and SSIM is better than the typical reconstruction with 6 diffusion direction images.

C Original Dataset

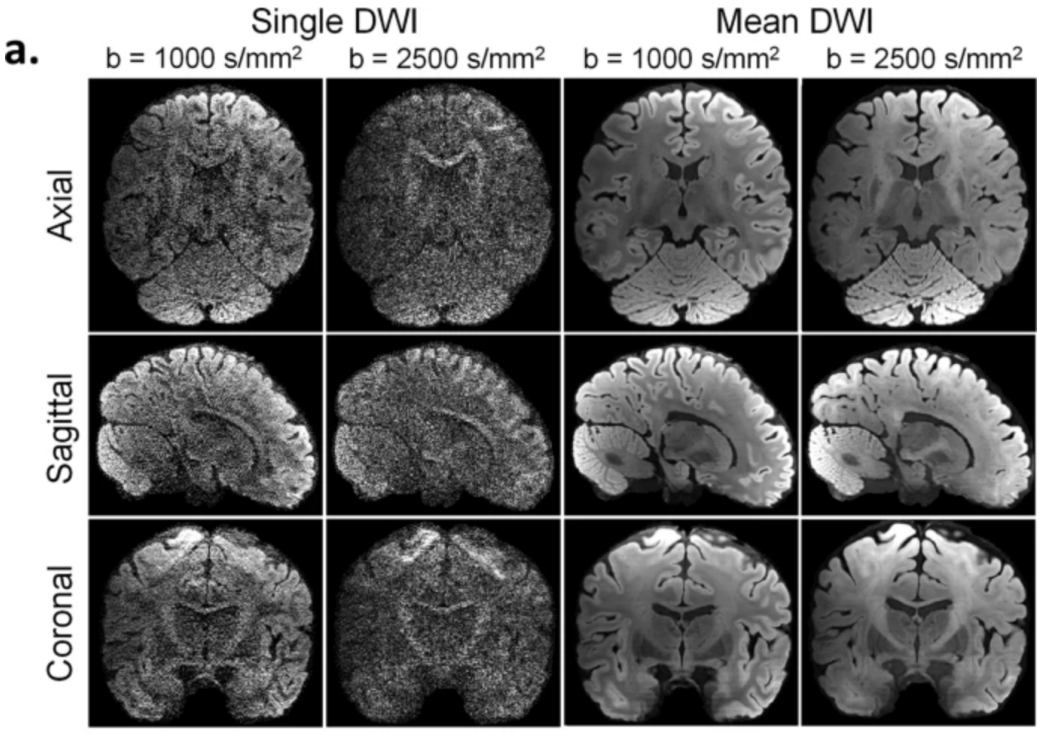

Figure 9: Diffusion-weighted images from the original 760μm dataset, which were downsampled onto a 128X128 matrix. [9]