Team 6 Project II: Neuroimage Registration and Synthesis

Shuyuan Wang
May 15, 2023

Abstract: This project aims to synthesize Fractional Anisotropy (FA) and Apparent Diffusion Coefficient (ADC) maps from T1-weighted (T1w) and T2-weighted (T2w) magnetic resonance imaging (MRI) images. These maps provide crucial information about various tissue properties, some of which are not readily available during the MRI scanning process.[8] To achieve this objective, we will utilize learning-based methods to extract the brain region, align the multi-slice images, and perform pixel-wise synthesis. Additionally, we will explore various optimization strategies, develop state-of-the-art neural networks, and establish balanced evaluation criteria. The effectiveness of the proposed approach will be evaluated comprehensively.

1. Introduction

Magnetic Resonance Imaging (MRI) is a widely used non-invasive imaging technique in medical diagnostics, providing detailed anatomical and functional information about different tissues and organs. To extract comprehensive information from MRI scans, various imaging modalities are utilized, including T1-weighted (T1w) and T2-weighted (T2w) images, as well as diffusion-weighted measurements like Fractional Anisotropy (FA) and Apparent Diffusion Coefficient (ADC) [8]. However, in some cases, the FA and ADC measurements may not be readily available or acquired during the MRI scanning process.

In this project, we propose a learning-based approach combined with traditional feature extraction techniques to synthesize FA and ADC images from the provided T1w and T2w structural MRI images of the brain. Initially, we employ a pre-trained HD-BET model [3] to extract the brain region and correct the images by removing the skull. Subsequently, we utilize the registration function of the ANTsPy library [1] to align the images to the T1w reference. By comparing the results using a minimal Mean Absolute Error (MAE) criterion, we determine the optimal alignment.

For the synthesis process, we employ a U-Net architecture [7], which has demonstrated exceptional performance in pixel-wise segmentation of medical images. Based on the same principle, the U-Net is trained to generate voxel-wise intensity values for the FA and ADC images. Additionally, we incorporate the idea of Conditional Generative Adversarial Networks (CGAN) [6] and explore its application during the training process. To evaluate the effectiveness and accuracy of the proposed solutions, we will conduct a comprehensive evaluation, including quantitative metrics such as Mean Absolute Error (MAE) and discriminator loss, as well as qualitative assessments through visual comparisons.

[1]https://antspy.readthedocs.io/en/latest/registration.html

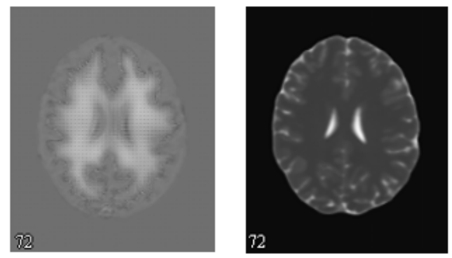

Figure 1: The synthesized ADC image before (left) and after (right) the MinMax normalization

2. Background

2.1. Image Normalization

Both FA and ADC are important measures for characterizing tissue microstructure and have been used in various clinical applications such as stroke diagnosis, tumor detection, and neurodegenerative disease assessment. FA is a scalar value that reflects the degree of anisotropy or directionality of water diffusion within a tissue.[8] The pixel values range from 0 to 1, where 0 indicates isotropic diffusion (no directional preference) and 1 indicates highly anisotropic diffusion (strong directional preference). Since, it is already in the range of [0, 1], we don't need more normalization operation on the output of the network. ADC, on the other hand, is a measure of the magnitude of water diffusion within a tissue, regardless of its directionality.[8] The pixel values typically range between 0 and 3 $\mu m^2/ms$, but the range can vary depending on the imaging protocol used. Hence, as shown in Figure 1, we need a MinMax normalization to help well-define the range of the network result.

$$x' = \frac{x - min(x)}{max(x) - min(x)}$$

2.2. Brain Extraction

Brain extraction, commonly known as skull stripping, is to separate the brain tissue from non-brain tissue, including extracerebral structures like the skull and skin. Accurate analysis and further processing of MRI data heavily rely on the quality of brain extraction. HD-BET is a more recent algorithm that uses a deep learning-based approach.[3] The key advantage of HD-BET lies in its ability to learn complex patterns and variations in brain structure, making it robust in the presence of tissue alterations caused by pathology or treatment. By capturing high-level features and spatial information, the network can accurately distinguish brain tissue from surrounding non-brain structures. This deep learning-based approach has demonstrated superior performance in various brain extraction tasks.

2.3. Image registration

The image registration is another critical process in the neuroimaging pipeline. It aligns different sets of data into one coordinate system, facilitating the comparison and integration of different modalities. Here we want to align multiple images with the T1w. The idea is to call ANTsPy library, which registers a pair of images through the simplified interface to the ANTs registration method. For the T2w images, we utilize the "BOLDRigid" registration method, which performs a rigid transformation, typically used for BOLD to BOLD intrasubject registration. For the FA and ADC images, we use the "SyN" (Symmetric Normalization) method, a more complex transformation method that includes both an affine transformation and a deformable transformation. The deformable transformation can model more complex, non-linear variations in brain anatomy, allowing for better alignment of the images.

2.4. Image synthesis

In light of the limited number of training and testing sets available, we propose the use of the U-Net model for segmentation-related tasks, as it has demonstrated remarkable performance even with a small number of samples and data augmentation.[7] Considering the pure generator structure always fails to produce satisfactory results with MAE loss, we include a more comprehensive Conditional GAN architecture.[6] Another ideas is to use, for example, the Riemannian network[1] to generate the diffusion images via T1w and T2w. FA and ADC can be then gained by the Diffusion Tensor Imaging (DTI) studies.[8]

Some other learning-based methods are developed to fit the diffusion maps. The pix2pix method explores conditional adversarial networks as a general-purpose solution for image-to-image translation problems.[4] The MedGAN method, as shown in Figure 2, merges the adversarial framework with a new combination of non-adversarial losses.[2] The BCI method proposes a pyramid pix2pix image generation model that has demonstrated superior translation results.[5]

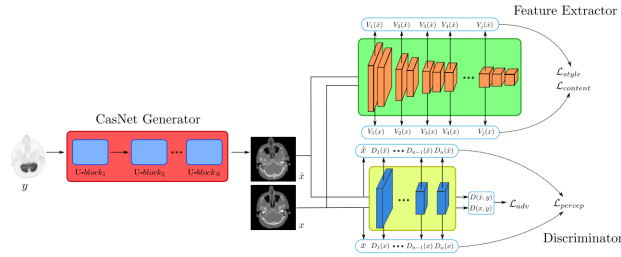

Figure 2: MedGan Architecture

3. Preliminary Method

The proposed network architecture is based on the u-net structure. As shown in Figure 3, it consists of an encoder-decoder structure with skip connections, allowing for effective feature extraction and preservation of spatial information.

In our approach, we preprocess the T1w and T2w MRI images by applying skull stripping using the HD-BET algorithm and aligning them using the ANTsPy library. To leverage the complementary information from the T1w and T2w MRI modalities, we combine them into a two-channel input. Unlike traditional approaches that select key slices or crop multiple patches, our network takes the entire volume as input. By doing so, we enable the network to capture global information, facilitating better localization and synthesis performance.

There is a trade-off between localization accuracy and the use of context. A large patch would provide more context that helps the network to learn and analysis the neighbour information while low down the localization accuracy by frequent downsamplings, vice versa. This concern would be the central idea for designing the following operations. Then two layers of 3D convolution are applied in each horizontal layer to utilize the context and capture the abstract information. A maxpooling operation is set to reduce the patch size that helps improving the localization accuracy. On the upsampling side, a transpose convolution is called to match the resolution, ensuring the synthesized images have the same spatial dimensions as the input one. We copy and combine high and low resolution information. This allows the network to leverage both detailed local information and global contextual information to facilitate localization and synthesis performance. Finally, we use linear interpolation to generate FA and ADC maps of the desired dimensions.

A normalization step is applied to the T1w and T2w images before inputting them into the network. The convolution parameters are initialized using Xavier initialization with a normal distribution. Batch normalization are called to stabilize the training of neural networks and improve their accuracy. ReLU and LeakyReLu activation functions

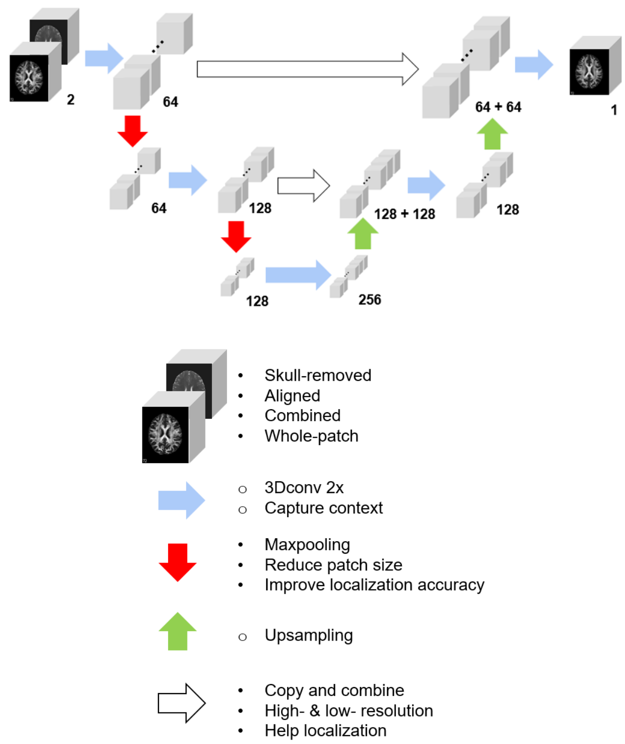

Figure 3: U-Net Architecture

are incorporated after each convolution layer to introduce non-linearity. The Adam optimizer with a learning rate of 0.001 is utilized, and a scheduler is employed to decrease the learning rate by 0.1 every 10 epochs. The total number of epochs is set to 30. Due to the large 3D nature of the MRI data and memory constraints, a batch size of 1 is chosen, although it may result in slower training. The MAE loss function is selected as the criterion due to its sensitivity to outliers and gradient stability properties.

4. Experimental Results

The dataset of 200 subfolders are divided into training and validation set by a ration of 9:1. As shown in Figure 4, the FA network converges and get the minimal MAE loss 0.0079 at epoch 28. For ADC learning, there is still spike and plummet after 30 epochs. But they are of very small scale and hence be regarded as convergence. We find the minimal MAE loss 0.0348 at epoch 24.

The ground truth and demos during iteration will be shown in Figure 5.

5. Discussion and Conclusion:

5.1. Further Discussion

In the conditional Generative Adversarial Network (CGAN) framework, two key components are utilized: a generator and a discriminator. The generator is responsible for generating synthesized ”fake” images, while the discriminator’s role is to differentiate between the ”fake” synthesized images and the ”real” ground truth images. Throughout the training process, the generator and discriminator engage in a competitive interplay, resulting in the generation of high-quality synthesized images. This adversarial setup facilitates the improvement of the generator’s ability to produce realistic and visually appealing outputs.

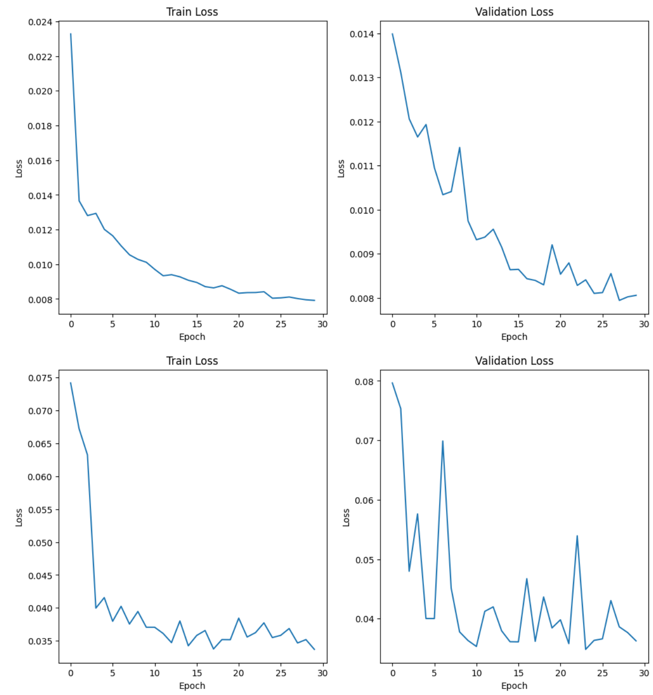

Figure 4: Training and Validation Loss of FA (above) and ADC (below)

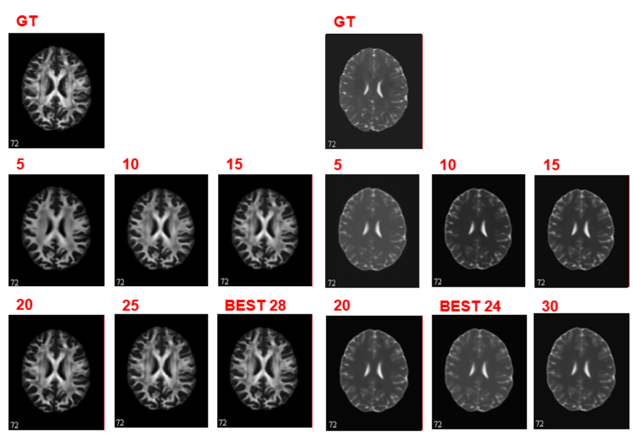

Figure 5: Iteration Demostration of FA (left) and ADC (right)

We copy the design of the U-Net architecture to set up

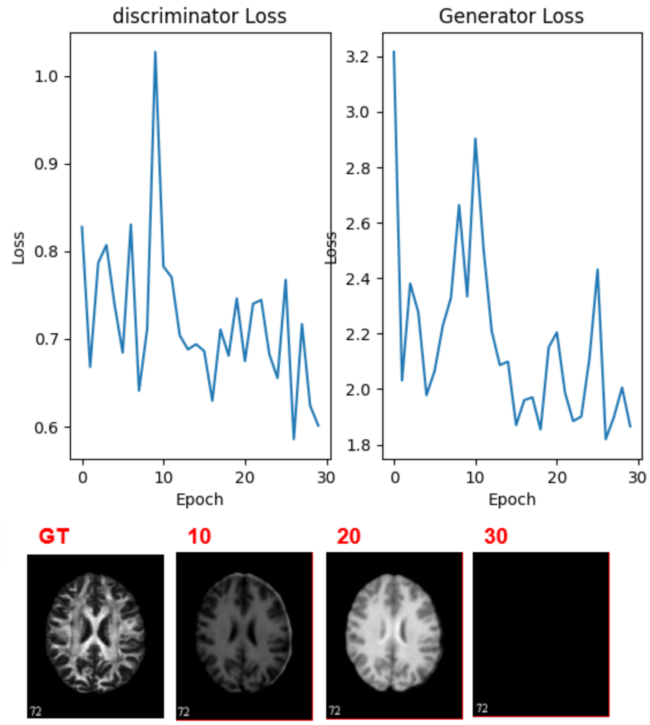

Figure 6: CGAN Iteration Demostration

the generator and develop five layers of 3D convolution to distinguish the images. The challenging part is to design a balanced loss function. We use MAE loss to evaluate the image quality of the generator and BCE loss to quantify the real and fake judgement. A smoothing label strategy is applied to prevent the model from being too confident about the binary calss labels. The "real" and "fake" labels are set to be slightly less than 1 and greater than 0 respectively. The total loss is combined with the weight ratio of 5:1 so that the network relies more on the MAE than the BCE.

Another tricky point is the collapse issue. It refers to the situation where the generator learns to produce a limited number of output samples that fool the discriminator, but not the entire target distribution (in our case, only producing images with all pixel values 0). The generator finds a particular point in the space that the discriminator can't distinguish as fake, and it just keeps generating that. The reason is inadequate training and architecture. Hence we set a ratio of 1:20 to train generator for one batch and discriminator for the next twenty batches. The ration will increase with iteration and finally reach 1:2. This setting would allow the CGAN reach a minimal generator loss of 1.8190 at epoch 25, as shown in Figure 6. The collapse happens at epoch 30 with a relatively high loss. We still need to finedefine a more balanced loss and try different architectures.

5.2. Conclusion

In conclusion, this project aimed to address the task of skull removal and alignment of MRI images, followed by the synthesis of FA and ADC maps using T1w and T2w MRI images. A comprehensive procedure, including HD-BET and ANTsPy, was developed to successfully remove the skull and achieve image alignment. We proposed a U-Net architecture that effectively synthesized FA and ADC maps with a notable reduction in loss. Additionally, we explored the design and fine-tuning strategies of the Conditional Generative Adversarial Network (CGAN) to further enhance the synthesis results. These advancements contribute to the development of improved techniques for MRI image synthesis and analysis, showcasing potential for future research and applications in medical imaging.

6. Management

6.1. Team Contribution

In our team, Tong Mu and Shuyuan Wang are responsible for the Project II, where Tong focus on the skull stripping, image registration, and network dataloader. Shuyuan is in charge of designing, tuning, and data-collecting the neural networks.

Junyi Liu and Chenyu Jin are responsible for the Project I.

6.2. Acknowledgement

I would like to express my deepest gratitude to Dr. Jerry Prince for his invaluable guidance, support, and encouragement throughout the lecture. His extensive knowledge, insightful feedback and constructive criticism have been instrumental in improving the quality of our study and work.

Honored as the co-champion of the demo competition. I would like to express my sincere gratitude to my teammates, namely Tong Mu, Junyi Liu, and Chenyu Jin. Their dedication, hard work, and collaboration were instrumental in achieving our goals.

References

[1] B. Anctil-Robitaille, A. Théberge, P.-M. Jodoin, M. Descoteaux, C. Desrosiers, and H. Lombaert. Manifold-aware synthesis of high-resolution diffusion from structural imaging. *arXiv preprint arXiv:2108.04135*, 2021.

[2] K. Armanious, C. Jiang, M. Fischer, T. Küstner, T. Hepp, K. Nikolaou, S. Gatidis, and B. Yang. Medgan: Medical image translation using gans. *Computerized medical imaging and graphics*, 79:101684, 2020.

[3] F. Isensee, M. Schell, I. Pflueger, G. Brugnara, D. Bonekamp, U. Neuberger, A. Wick, H.-P. Schlemmer, S. Heiland, W. Wick, et al. Automated brain extraction of multisequence mri using artificial neural networks. *Human brain mapping*, 40(17):4952–4964, 2019.

[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[5] S. Liu, C. Zhu, F. Xu, X. J. Z. Shi, and M. Jin. Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix supplementary material.

[6] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[7] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[8] J. M. Soares, P. Marques, V. Alves, and N. Sousa. A hitchhiker's guide to diffusion tensor imaging. *Frontiers in neuroscience*, 7:31, 2013.