



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

이학석사 학위논문

**Macro to Micro: Brain MR
Image-to-Image Translation between
Structural MRI, Diffusion MRI and
Tractography**

뇌의 미·거시적 특성을 이용한 자기공명영상의 딥러닝
이미지 간 변환 기법

2023년 8월

서울대학교 대학원
뇌인지과학과 뇌인지과학전공
김수영

Macro to Micro: Brain MR Image-to-Image Translation between Structural MRI, Diffusion MRI and Tractography

지도 교수 차 지 욱

이 논문을 이학석사 학위논문으로 제출함

2023년 7월

서울대학교 대학원

뇌인지과학과 뇌인지과학전공

김 수 영

김 수 영의 이학석사 학위论문을 인준함

2023년 7월

위 원 장 _____ 권 준 수 (인)

부위원장 _____ 차 지 욱 (인)

위 원 _____ 유 신 재 (인)

Abstract

The brain consists of the highly localized functions of several brain regions and the integration of these regions through neural connections. These brain neural connections are constantly changing at the systemic and synaptic levels to effectively respond to the ever-changing environment. One of the key factors enabling these dynamic interactions is the structural plasticity of the human brain at the macro and micro scale. Because the brain’s macro- and micro-structures convey different but complementary information, considering both structures is critical to understanding the brain’s structural plasticity and connectivity during cognitive tasks. However, previous studies have not effectively considered this issue.

In this study, a novel deep learning framework, Macro2Micro, is proposed to generate high-quality Diffusion Tensor Imaging (DTI) and tractography from structural MRI (sMRI). The study is premised on the hypothesis that micro-scale structural information can be inferred from macro-scale structures, enabling the generation of different imaging modalities beneficial for disease diagnosis and research, even when only one modality is initially obtained. This approach, unprecedented in the realm of neuroimaging, leverages the benefits of cross-modality image translation, offering significant time and cost savings. The Macro2Micro framework utilizes 3D T1 to generate 2D T1 slices as input, which are then processed through a Generative Adversarial Network (GAN) to produce 2D DTI (FA) slices and subsequently 2D tractography. The key element of this process is the use of Octave Convolutions, which facilitate the analysis of connections between various scale MR modalities. The framework was trained using the Adolescent Brain Cognitive Development (ABCD) dataset, with training losses evaluated through Image Pixel loss, Perceptual loss, GAN loss, and

brain-focused patch GAN loss.

The results not only showed superior performance compared to other algorithms quantitatively and qualitatively but also have significant meaning in neuroscience in that they learned not only the image distribution but also the biological characteristics of the structural and microscopic structures of the brain. The potential application of this image translation model as a data augmentation method could address issues of data imbalance and scarcity. This research underscores the potential of multimodal imaging, specifically the combined use of T1, DTI, and tractography, in advancing disease modeling.

Keywords: Image-to-Image Translation, Generative Adversarial Networks, Deep Learning, Structural MRI, Diffusion MRI, Tractography

Student Number: 2021-22028

Contents

Abstract	i
Chapter 1 INTRODUCTION	1
Chapter 2 RELATED WORK	7
Chapter 3 METHOD	15
3.1 Architecture Overview	16
3.2 Octave Convolution	17
3.3 Networks	18
3.4 Training Losses	19
3.5 Image Quality Metrics	21
3.6 Comparison of generated and real FA images in low-dimensional representation	23
3.7 Prediction of biological and cognitive variables using predicted FA images	23
3.8 Prediction of Tractography from FA images	24
3.9 Experimental Settings	25
Chapter 4 RESULTS	26
4.1 Qualitative Evaluation	26
4.2 Quantitative Evaluation	27
4.3 Generated FA images by Macro2Micro can efficiently predict sex, ADHD and intelligence	29

4.4	Ablation Studies	30
4.5	Effectiveness of Macro2Micro along the distance from the center of the brain	34
4.6	FA Image Translation to Tractography	34
Chapter 5	DISCUSSIONS AND CONCLUSIONS	42
	Bibliography	47
	국문초록	54

List of Figures

Figure 3.1	The overall architecture of Macro2Micro.	16
Figure 3.2	The structure of Octave Convolutions (OctConvs)	17
Figure 4.1	Qualitative Comparison of images made with Macro2Micro (our model), and images made with Pix2Pix and CycleGAN. Note that Pix2Pix and CycleGAN are the baselines.	36
Figure 4.2	Generated realistic FA images from T1-weighted images using Macro2Micro model.	37
Figure 4.3	PCA and T-SNE results for the T1, FA (Ground Truth), and generated FA images from Micro2Macro. (a) PCA (b) T-SNE	38
Figure 4.4	Ablation study Qualitative comparison of the effectiveness of Octave Convolution.	39
Figure 4.5	Ablation study Generated FA images with and without the use of Brain-focused Patch-wise Discriminator.	40
Figure 4.6	Ablation study Qualitative comparison of the effectiveness of pre-trained VGG perceptual loss.	40
Figure 4.7	Ablation study Qualitative comparison of the effectiveness of normalization and activation function.	41
Figure 4.8	Graph depicting the performance with different distances from the middle slice.	41
Figure 4.9	Qualitative comparison of generated 2D Tractography from FA images.	41

List of Tables

Table 4.1	Quantitative comparison with generated whole brain FA images. The best outcomes are shown in bold . \uparrow : Higher is better. \downarrow : Lower is better.	28
Table 4.2	Quantitative comparison with baseline models using only white matter whose FA value is bigger than 0.2 from the generated FA images.	28
Table 4.3	Sex classification performance of real T1, real FA, and generated FA images. GT: Ground Truth.	29
Table 4.4	ADHD classification performance of real T1, real FA, and generated FA images.	30
Table 4.5	Intelligence regression performance of real T1, real FA, and generated FA images.	31
Table 4.6	Ablation study Depending on whether Octave Convolution is applied and the difference in α value.	31
Table 4.7	Ablation study for brain-focused patch discriminator.	32
Table 4.8	Ablation study of perceptual loss using pre-trained VGG network.	32
Table 4.9	Ablation study for different normalization and activation functions.	33
Table 4.10	Quantitative Comparison of generated tractography images with baseline models.	34

Chapter 1

INTRODUCTION

The brain function is primarily determined by its connections. The brain consists of highly localized functioning of several brain regions and integration of these regions through neural connection [1]. The brain connections are constantly adjusted at both systemic and synaptic levels to adapt to the ever-changing environment effectively. Research studying the relationship between the brain's connectivity and its interplay with the environment has yielded novel understandings of adaptive human behavior. One important factor that makes the dynamic interaction possible is *structural plasticity* at the macro and micro scale of the human brain.

The macro-structure of the human brain refers to the architecture of cell bodies and neurite arrangement within the gray matter or axonal properties (e.g., tissue organization, size distribution, density, and myelin content) in the white matter. The macro-structure of the human brain has enormous potential in understanding the structure and organization of the brain, providing unique insights for abnormalities that underpin disease states. The microstructure, on the other hand, refers to the microscopic structure of brain tissue components such as myelin, axons, and dendrites or the quantification of their properties. This microscopic structural architecture includes multiple properties such as cell size, the orientation of axonal fiber bundles, and packing density. Several dysfunctions of the human brain result from pathological features such as demyelination, inflammation, and axonal loss [2]. As both macro- and micro-structure of

the brain convey different yet complementary information of the human brain, considering both structures is extremely important for understanding the *structural plasticity* of the human brain and connectivity during cognitive tasks.

Magnetic resonance imaging (MRI) has had an unprecedented contribution to our understanding of the macro- and micro-structure of the human brain. The macrostructure of the brain could be captured by structural MRIs (sMRI), especially with T1-weighted MRI. T1-weighted images are favorable for observing macro-scale structures such as the overall shape and size of brain tissues. On the contrary, diffusion MRI (dMRI) can be used for studying brain tissue micro-structures [3]. Brownian motion of water molecules in the brain is restricted by factors such as the micro-environment and temperature [4] and this makes the properties of microscopic structural tissue can be observed sensitively at the voxel scale [5]. Thus, dMRI is helpful in recognizing tissue damage or abnormalities that depend on the cell density of the tissue. Unlike sMRI, the contrast of the image in dMRI is determined by the difference in the diffusion rate of water molecules, so neuroanatomy can be seen at the function and physiology level [4]. Diffusion-weighted Image (DWI), one of the most used in dMRI models can sensitively detect the increase in parenchymal signal due to tissue damage caused by cytotoxic edema for diagnosing hyperacute stroke. DWI is used in many advanced dMRI models such as neurite orientation dispersion and density imaging (NODDI), diffusion kurtosis imaging (DKI), and diffusion tensor imaging (DTI). In addition, diffusion-derived scalar maps provide rich information about micro-structural characterization [3].

While these modalities have made great strides in identifying changes at different scale structures, they have generally proven to be very sensitive to tissue abnormalities, albeit with low specificity [2]. The multiplicity of factors contributing to the overall signal prevents a direct correlation between MRI biomarkers and biological substrate.

To address the low specificity and embrace brain structures at different scales, several studies have attempted to combine multiple MRI modalities. Diffusion tensor imaging (DTI) and T1-weighted images provide different yet complementary information. For example, DTI is particularly sensitive in reflecting the structural state of white matter, which has led to active research in white matter diseases such as multiple sclerosis [6, 7], CADASIL [8, 9], amyotrophic lateral sclerosis [10, 11], and Alzheimer’s disease [12, 13]. These studies have demonstrated that DTI can effectively detect pathological alterations in white matter regions that appear normal in T1 and that changes in DTI measures correlate with pathological changes and clinical findings.

Therefore, a multimodal approach that simultaneously utilizes DTI and T1 allows for a more comprehensive understanding of brain information, as recent research supports. Recently, the fusion of diffusion indices and surface information through T1-DTI fused images has shown improved detection power in distinguishing between individuals with mild cognitive impairment (MCI) and healthy controls [14]. Furthermore, a multimodal approach incorporating feature selection from both T1 and DTI data has demonstrated superior performance in age regression compared to the unimodal approach, which utilizes each data modality separately. [15].

Similarly, research to use tractography, which can capture brain connection information, is also in the spotlight. Fractional anisotropy (FA) is the most widely used quantitative index obtained through DTI. However, the FA image has a single scalar value for each voxel. Thus it cannot represent the directionality of tissues in each voxel. On the other hand, tractography utilizes DTI data to tract and reconstruct fiber pathways, providing insights into the structural connections between different brain regions. By tracing the trajectory of neural fibers, tractography can effectively capture and visualize the connectivity patterns in the human brain, which FA alone cannot achieve. Considering and analyzing these three MRI modalities all at once would be

a milestone in neuroscience in that each of them provides distinct yet complementary information.

There have been numerous attempts to integrate these modalities in a multimodal fashion in neuroimaging studies. However, multimodal deep learning poses several challenges.

The major bottleneck for multimodal studies is that obtaining different modalities data is *extremely* costly and time-consuming. For example, Diffusion Tensor Imaging (DTI) requires a much longer acquisition time compared to T1-weighted imaging. In the ABCD (Adolescent Brain Cognitive Development) dataset we utilized, obtaining the DWI (Diffusion Weighted Imaging) data, which serves as the foundation for DTI, took 7 minutes and 30 seconds per single subject. Also in the ABCD study, to obtain a single DTI (Diffusion Tensor Imaging) image, 96 DWI (Diffusion Weighted Imaging) images are required [16]. Therefore, it would take approximately 720 minutes to acquire DTI images for a single patient, while T1 imaging took 5 minutes and 38 seconds using a Phillips scanner [17]. To obtain both T1 and DTI images simultaneously, patients need to remain inside the MRI scanner for an extended period. However, studies have shown that the quality of MRI scans can be compromised in patients with panic disorder or agoraphobia (PD/AG) due to the potentially stressful nature of the MRI environment [18]. Additionally, MRI scans are highly sensitive to motion, leading to the use of sedatives or anesthesia in infants with frequent movement. However, these approaches are not recommended due to ethical concerns and potential long-term risks [19]. Therefore, it is practically impossible to obtain high-quality images for lengthy acquisitions like DWI in patients who cannot tolerate long periods inside the MRI scanner or in infants with frequent motion.

Moreover, when it comes to obtaining tractography, generating 1 million tracks using software such as MRtrix3[20] typically takes around 12 hours per subject. Such

challenges significantly limit the clinical applicability of tractography despite its potential value. Recently, research generating tractography from diffusion-weighted imaging (DWI) has started to emerge, and even that remains rare [21].

In deep learning research where a large amount of data is crucial, only subjects who have data for all modalities can be used for the analysis. As a result, the available data size is inevitably reduced. It can lead to decreased generalizability and an increased risk of overfitting. Furthermore, effective integration of multimodal data requires capturing their interactions accurately, which often necessitates complex model architectures and, consequently, requires more computational resources. This study seeks to propose a model that can generate FA and tractography from T1 without requiring a significant amount of time and resources.

A recent study has shown that when structural brain imaging is divided into high-frequency and low-frequency components, the performance in the image reconstruction task is better compared to the cases without frequency division [22]. Similarly, encoding and decoding the brain based on its frequencies could significantly increase the quality of synthesizing the target image. This study aims to develop a novel image-to-image translation model, Macro2Micro, that can be applied to the brain, utilizing octave convolution commonly used in computer vision.

Firstly, we compare and validate our proposed model rigorously against existing studies in the T1 to FA image translation task. Furthermore, we extend our research by generating tractography from the FA images, allowing us to explore the potential of our model in capturing and representing the structural connectivity information of the brain.

In summary, the main contributions of this research encompass:

- We propose a novel I2I translation framework, Macro2Micro for synthesizing

diffusion MRI and tractography from structural MRI. Our model achieves the highest quality compared to various I2I algorithms qualitatively and quantitatively.

- To effectively synthesize the brain MR images, we introduce several simple yet effective modifications on the discriminator, yielding more satisfying output with fine brain details. To further accelerate the networks' capability to learn subject-independent representation and the connectivity between the macro and micro-structure, Octave Convolutions and prior knowledge from pre-trained convolutional neural networks are used. Comprehensive experimental results demonstrate the validity of our design choice and its superiority to conventional I2I translation techniques.
- To the best of our knowledge, this is the first attempt to translate tractography from structural MR images in medical image modality translation scenarios.

Chapter 2

RELATED WORK

Modalities in Neuroimaging. Magnetic Resonance Imaging (MRI) is a non-invasive imaging technology that produces three-dimensional detailed anatomical images. It is often used for disease detection, diagnosis, and treatment monitoring. It is based on sophisticated technology that excites and detects the change in the direction of the rotational axis of protons found in the water that makes up living tissues. Structural MRI, in particular, provides high-resolution, three-dimensional images of the brain's anatomy. It is commonly used to visualize the brain's structure and detect abnormalities such as tumors, strokes, or developmental anomalies. T1-weighted imaging (T1) is a specific type of MRI that uses the longitudinal relaxation time of tissues to produce images. T1 is particularly useful for visualizing anatomy and pathology in high detail, making it a valuable tool in neuroimaging.

Diffusion Tensor Imaging (DTI) is a type of MRI that measures the random motion of water molecules within a voxel of tissue. It is particularly useful for visualizing white matter tracts in the brain, as the direction of water diffusion can provide information about these fibers. Fractional Anisotropy (FA) is a scalar value between zero and one that describes the degree of anisotropy of a diffusion process. A value of zero means that diffusion is isotropic, i.e., it is unrestricted (or equally restricted) in all directions. A value of one means that diffusion occurs only along one axis and is fully restricted along all other directions. FA is a commonly used measure in DTI studies,

as it can provide important information about the integrity of white matter tracts.

In addition to FA, another commonly used measure in DTI is Mean Diffusivity (MD). MD represents the average magnitude of water diffusion within a voxel, regardless of direction. It is a scalar value that reflects the overall diffusion characteristics of the tissue. Higher MD values indicate greater diffusion, suggesting less restriction or damage to the tissue. Conversely, lower MD values indicate reduced diffusion, which may indicate more restricted or compacted tissue. MD is often used in conjunction with FA to provide a more comprehensive understanding of the microstructural properties of matter tracts. By combining information from both FA and MD, researchers can gain insights into the structural integrity and health of the brain's white matter pathways.

Tractography is a 3D modeling technique used to visually represent neural tracts using data collected by diffusion MRI. It uses special techniques to visualize, analyze, and quantify the diffusion of water molecules in the white matter of the brain, which can provide insights into the structure and connectivity of the brain's neural pathways. Tractography can be used to generate a detailed map of the brain's white matter tracts, which can be useful in neurosurgical planning, diagnosing various neurological conditions, and understanding the brain's connectivity. It can be performed in 2D slices or as 3D reconstructions, providing a comprehensive view of the brain's white matter architecture.

Image-to-Image Translation. Image-to-image translation refers to the task of converting an input image from one domain to another while preserving relevant visual information. It aims to learn a mapping function to transform images from a source domain to a target domain, enabling various applications such as style transfer, colorization, and modality conversion. Its goal is to generate realistic and visually coherent output images that are indistinguishable from the target domain.

Despite advancements in image-to-image translation, several challenges persist. These challenges include:

- **Paired Data Requirement:** Many existing methods rely on paired training data, where corresponding images from the source and target domains are required. However, obtaining such paired data can be time-consuming and expensive in certain domains.
- **Mode Collapse:** Some algorithms may suffer from mode collapse, where the generator produces limited variations of the target domain, resulting in a lack of diversity in the generated images.
- **Preserving Fine Details:** Preserving fine-grained details during the translation process can be challenging, especially when the input and output domains have significant differences in texture, shape, or resolution.

To address the above challenges, several image-to-image translation algorithms have been proposed. Pix2Pix [23] is a conditional generative adversarial network (cGAN) that learns a mapping from input images to output images using paired training data. It combines a generator network with a discriminator network to optimize the image translation process. It produces high-quality translations when trained with paired data and can preserve fine details. However, it heavily relies on paired training data, which can be a limitation in certain scenarios.

CycleGAN [24] is an unsupervised learning approach that utilizes cycle consistency to learn mappings between two domains without paired data. It introduces cycle consistency loss, which enforces the reconstructed image to be close to the original input, ensuring consistency in both directions. It does not require paired training data, making it more flexible for real-world applications. However, it may struggle with preserving fine details and can be sensitive to hyperparameter settings.

These algorithms represent significant advancements in image-to-image translation, each with its own strengths and limitations. Further research and exploration are necessary to address the challenges and improve the performance of cross-modality image-to-image translation techniques.

Structure MR Image to Diffusion Tensor Image Translation. Gu et al. [3] explored the use of GANs, specifically CycleGAN, to generate synthetic diffusion scalar measures from structural T1-weighted images. This approach aimed to address the challenges of cost and time consumption in collecting high-quality diffusion data and the complexity of diffusion data processing pipelines. The CycleGAN model was trained to map a T1 image to FA or MD, and vice versa. The synthetic FA images were then used as a target for non-linear registration to correct for geometric distortions common in diffusion MRI. The results showed good to great visual similarity between synthetic FA and MD images and their ground truth, with the trained CycleGAN performing well across all test subjects. However, the authors noted that the training data used could influence the GAN’s performance, potentially introducing bias.

Yang et al. [25] proposed a cross-modality generation framework that leverages cGANs for Image Modality Translation (IMT) in MR images. The framework was designed to exploit both low-level features (pixel-wise information) and high-level representations (e.g. brain tumors, brain structure) between cross-modalities, addressing the complexity of brain structures. The authors introduced an end-to-end IMT network for cross-modality MRI generation and provided a comprehensive comparison with five datasets representing real-world clinical applications.

The paper also presented two novel methods that utilize the IMT framework. The first was a registration method that augments the fixed image space with translated modalities for atlas-based registration. The second was a segmentation approach, called translated multichannel segmentation (TMS), which performs cross-modality

image segmentation using fully convolutional networks (FCNs). Both methods leverage cross-modality information to improve performance without the need for additional data.

Despite achieving excellent performance, the authors acknowledged that the generated images may not accurately represent tiny structures as compared to real images.

Anctil-Robitaille et al. [26] proposed a novel approach to image-to-image translation for manifold-valued data, specifically in DTI, using a manifold-aware CycleGAN. Traditional GANs have been limited in their application to DTI due to their non-Euclidean nature. The authors addressed this by formulating the objective as a Wasserstein distance minimization problem of data distributions on a Riemannian manifold of symmetric positive definite 3×3 matrices ($\text{SPD}(3)$), using adversarial and cycle-consistency losses. They leveraged the Log-Euclidean metric and the structural information of T1-weighted (T1w) images to generate realistic high-resolution DTI. The generated diffusion tensors were ensured to lie on the $\text{SPD}(3)$ manifold by exploiting the theoretical properties of the exponential and logarithm maps of the Log-Euclidean metric.

The proposed method outperformed both the manifold-aware GAN and the standard CycleGAN in terms of tensor principal orientation estimation, Log-Euclidean distance, and mean squared error (MSE) of derived FA. The results suggested that T1w images may contain information on the high-level geometry of fiber tracts, which can be learned by the network to estimate the diffusion properties and orientation. This work represents a significant contribution to medical image computing, unlocking a vast number of applications on manifold-valued data.

Liu et al. [27] developed a generative self-training (GST) framework for unsupervised domain adaptive (UDA) medical image translation tasks. The framework incorporates a unified uncertainty quantification scheme for both epistemic and aleatoric uncertainties, allowing for adaptive control of generative pseudo-label supervision

based on the reliability of the pseudo-labels. This approach was applied to tagged-to-cine and structure-to-diffusion (e.g., T1-to-FA) image translation tasks, marking the first attempt at such UDA. The results demonstrated superior translation performance compared to popular adversarial UDA methods. The framework also includes a self-attention scheme to emphasize regions of interest and prevent the background region from dominating the training process. The authors also noted that their framework could be applied to other discriminative self-training UDA tasks and that different backbones in the source and target domain could be used.

Anatomy of the Brain and the vmPFC. The human brain is divided into several distinct regions with specific functions and responsibilities. Major divisions of the brain include the cerebrum, cerebellum, and brainstem. The cerebrum, the largest part, is further divided into two hemispheres and is responsible for higher brain functions such as thinking, learning, and consciousness. Each hemisphere is divided into four lobes: the frontal lobe (responsible for reasoning, problem-solving, and skills), the parietal lobe (processing sensory information), the occipital lobe (vision), and the temporal lobe (hearing and memory). Understanding the intricate structure and function of the brain is essential for developing effective image-to-image translation techniques in brain imaging.

The ventromedial prefrontal cortex (vmPFC) is a subregion of the prefrontal cortex, located in the anterior part of the brain. It is situated where the two cerebral hemispheres meet at the bottom (ventral) part of the frontal lobes, hence its name. The vmPFC is bounded by the orbital prefrontal cortex (OFC) laterally, the anterior cingulate cortex (ACC) dorsally, and the subgenual cingulate cortex (sgACC) posteriorly.

The vmPFC is a complex and heterogeneous region with diverse cytoarchitectonic features, indicating a variety of functions. The vmPFC has extensive connections with other brain regions, including the amygdala, hippocampus, hypothalamus, and various

sensory areas, which facilitate its role in integrating emotional, cognitive, and sensory information.

The vmPFC is not only involved in decision-making and emotional processes but also plays a role in self-referential thinking and theory of mind—the ability to understand other’s mental states. It is also implicated in the regulation of autonomic and endocrine responses to stress, and the formation and retrieval of personal memories.

In neuroimaging studies, the vmPFC is often challenging to image due to its location and susceptibility to signal dropout. However, accurate imaging of the vmPFC is crucial given its involvement in various cognitive processes and its implication in numerous psychiatric and neurological disorders. Therefore, advanced image-to-image translation techniques that can accurately capture the structure and function of the vmPFC are of great importance in neuroscience research.

Frequency-based Brain Analysis. Data obtained from brain function measurements, such as EEG (Electroencephalography) and fMRI (Functional Magnetic Resonance Imaging), can be divided into multiple frequency bands. EEG data is characterized by electrical activity in the brain, and it can be divided into different frequency bands to examine different aspects of brain function. The commonly recognized frequency bands in EEG analysis are delta, theta, alpha, beta, and gamma. Delta (0.5-4 Hz) frequency is the lowest frequency range, primarily observed during deep sleep. Theta (4-8 Hz) frequency is primarily observed during sleep but can also be associated with altered states of consciousness. Alpha (8-13 Hz) frequency is typically observed during relaxed states, such as when closing the eyes and resting. Beta (13-30 Hz) frequency is predominantly observed during wakefulness and is associated with alertness. Gamma (30-100 Hz) frequency is a higher frequency range associated with complex brain activities related to cognition, attention, and memory [28, 29, 30]. These frequency bands help us understand the characteristics of EEG signals and investigate

various aspects of brain activity. Frequent band changes can be used to track changes in brain function related to cognition, sleep, attention, and other processes.

On the other hand, fMRI measures changes in blood oxygenation levels to infer brain activity. Although fMRI does not directly provide frequency information like EEG, researchers can analyze fMRI data to study the functional connectivity and fluctuations in different frequency ranges. Resting-state functional MRI research has uncovered that the blood oxygen level-dependent (BOLD) signal in the brain shows slow intrinsic fluctuations (less than 0.1 Hz) that are correlated over time. These fluctuations indicate a relationship within hierarchically organized functional systems known as resting state networks [31].

Furthermore, when applying a fast Fourier transform to fMRI data and examining it in the frequency domain, the scale-free activity demonstrates a division into frequency bands based on the knee frequency of the Lorentzian function [32, 33]. The presence of a low-frequency band suggests that the system exhibits decreased temporal redundancy and operates more efficiently in processing information in real time. Conversely, the high-frequency band indicates a higher level of time-lagged autocorrelation, indicating that past dynamics of the system have a greater impact on its future dynamics, thereby indicating the presence of long-range memory effects [34].

Different frequency ranges in brain signals correspond to different functional systems within the brain. This means that different functions generate signals in different frequency ranges. Therefore, analyzing brain research data by dividing it into frequency bands can help to elucidate various brain functions more clearly.

Chapter 3

METHOD

What is the ideal model for synthesizing the target brain modality while maintaining image quality that is comparable to actual magnetic resonance (MR) imaging data? We argue that the model should be able to synthesize the target brain modality effectively and accurately. Simultaneously, it should learn the delicate representation that describes subject-wise differences and the connectivity between the macro- and the micro-structure of the brain.

In order to accomplish these objectives, the Macro2Micro model (shown in Figure 3.1) is trained with the following core goals: 1) effectively generating the desired image by utilizing a patch discriminator that focuses on the brain and learns the statistical relationships between image patches, 2) learning subject-independent representation and understanding the interconnections between the macro and micro-structures through the utilization of Octave Convolution (OctConv), and 3) mitigating the risk of mode-collapse in the model by incorporating prior knowledge from pre-trained convolutional neural networks like VGG-19 [35].

In this section, we will outline our methodology that effectively enhances the quality of synthetic MR images. The subsequent sections include an exhaustive analysis of the proposed method and its underlying principles.

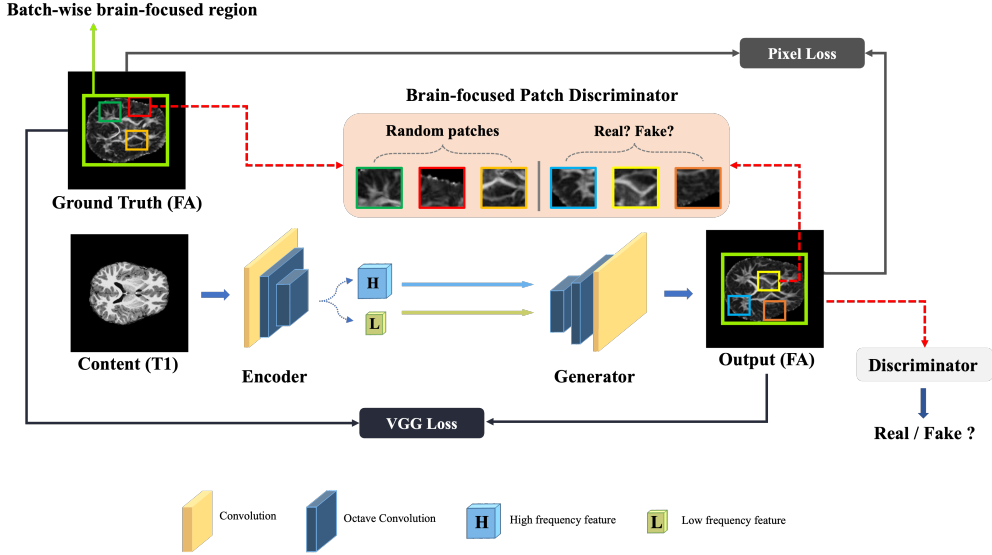


Figure 3.1: The overall architecture of Macro2Micro.

3.1 Architecture Overview

As illustrated in Figure 3.1, the Macro2Micro architecture comprises four primary components: an encoder E , a generator G , a discriminator D , and a brain-focused patch discriminator $D_{brainPD}$. Specifically, the input images are encoded via an encoder E and decomposed into two feature maps containing distinct frequency components. These frequency-decomposed features are subsequently fed to the generator. In the terminal layer of the generator, the synthesized high and low-frequency images are amalgamated to produce the final outputs. To summarize, the overarching pipeline proceeds as follows:

1. Encode two decomposed features f_H, f_L (the high frequency and the low frequency feature map each) from the input image I using the encoder E .

$$f_H, f_L := E(I)$$

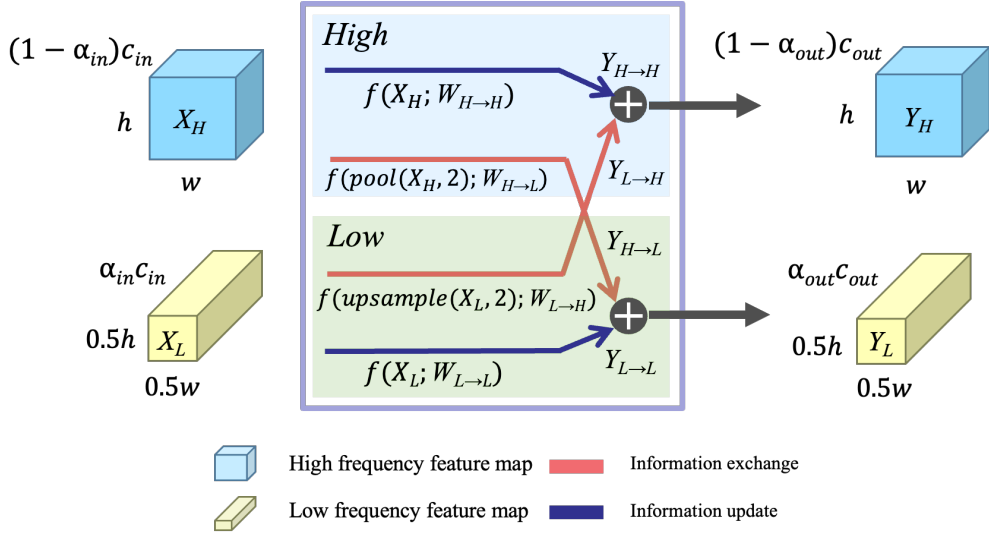


Figure 3.2: The structure of Octave Convolutions (OctConvs)

2. Synthesize the target output I_{out} in the generator G using the two decomposed features f_H, f_L .

$$I_{out} := G(f_H, f_L)$$

3.2 Octave Convolution

One of the key elements of the Octave Convolution (OctConv) [36] operator is its capability to factorize mixed feature maps based on their frequencies, while also facilitating effective communication between high-frequency and low-frequency components. The spatial resolution of low-frequency feature maps in OctConv is decreased by one octave [37]. The term *octave* refers to a spatial dimension that has been divided by a power of two. In this study, a value of 2 was chosen for simplicity, but this is a design choice and can be arbitrarily large.

When the spatial resolution of the low-frequency branch is decreased, the receptive field increases. This makes it possible for smooth, low-frequency maps to be stored in

a lower-resolution tensor. The multi-frequency representation that OctConv provides effectively isolates the input feature maps according to their frequencies, which helps to reduce spatial redundancy. Given input and output of OctConv as $X = \{X_H, X_L\}$ and $Y = \{Y_H, Y_L\}$, the forward pass of OctConv is defined as:

$$Y_H = f(X_H; W_{H \rightarrow H}) + f(\text{upsample}(X_L, 2); W_{L \rightarrow H})$$

$$Y_L = f(X_L; W_{L \rightarrow L}) + f(\text{pool}(X_H, 2); W_{H \rightarrow L}),$$

where $f(X; W)$, $\text{pool}(X, 2)$ and $\text{upsample}(X, 2)$ denote a convolution with parameters W , an average pooling operation with kernel size 2×2 with stride of 2 and an up-sampling operation by 2 using nearest interpolation, respectively.

Empirical findings indicate that employing OctConv with half the channels for each frequency ($\alpha = 0.5$) yields optimal performance. A comprehensive experiment regarding this matter can be found in the section 4.4.

3.3 Networks

The Encoder and The Generator. Both the encoder and the generator improve upon MobileNet [38] by replacing all the convolutions with Octave Convolution (OctConv) to factorize feature maps by their frequency, thereby reducing network redundancy while maintaining simplicity and efficiency. Active information exchange occurs between frequency components during training. The spatial reduction in the low-frequency branch expands the receptive field, capturing more contextual information from distant locations and improving performance. In contrast to the original OctConv [36], the up-sampling order is modified to effectively address checkerboard artifacts [39].

The Discriminator and The Brain-focused Patch Discriminator. Macro2Micro uses two discriminators: the discriminator and the brain-focused patch discriminator. While using the discriminator solely seems sufficient for our model to synthesize the

target modality, the results still suffer from the checkerboard artifacts [39] and undesired artifacts (see the section 4.4 for the details).

To tackle this issue, we have adopted a patch co-occurrence discriminator from the swapping autoencoder [40] and applied a simple yet effective pre-processing algorithm. We dubbed this discriminator as the brain-focused patch discriminator $D_{brainPD}$. Similar to [40], we encourage the cropped patch from the synthesized output to maintain the same representation and expressions in any cropped ground truth images. Consequently, the generator aims to generate an output image such that any patch from the output cannot be distinguished from a group of patches from the ground truth images.

Unlike computer vision tasks, the majority of MR brain images contained redundant background regions. Cropping patches from such regions and feeding them to the patch discriminator are inefficient and computationally expensive. Moreover, such redundant background regions are mostly zero-values. This could lead to the degradation of output image quality (e.g., blurring or dimmer images, pixelization) as the model would learn the background zero values or the abrupt changes in pixels.

To effectively cope with this, we first calculate the *valid* brain regions in the training mini-batch, which are then used to crop the valid region from the given training mini-batch. By doing so, our brain-focused patch discriminator serves to focus on the effective regions of the brain and enforce that the joint statistics of a learned representation consistently followed the ground truth modality.

3.4 Training Losses

To effectively guide our model to learn subject-independent representation and the connectivity between the macro and micro-structure of the human brain while synthesizing the target image with desired image quality, we impose several constraints on Macro2Micro. The following sections provide a detailed description of our training

objectives.

The Image Pixel Loss. The image pixel loss is Mean Square Error (MSE) between the synthesized output and the ground truth, which can be denoted as:

$$\mathcal{L}_{\text{pix}} = \|I_{\text{out}} - I_{\text{GT}}\|_1 \quad (3.1)$$

The Perceptual Loss. To prevent the model from falling the mode-collapse and generating skull-like artifacts (see the details in figure 4.4), we utilize prior knowledge from a pre-trained convolutional neural network, such as VGG-19 [35]. The perceptual loss was originally proposed by [41], yet has not been actively addressed in the Magnetic Resonance Imaging domain to cope with the mode collapse. Given I_{out} representing the synthesized output and I_{GT} denoting the ground truth image, the final perceptual objective we used is as follows:

$$\mathcal{L}_{\text{perct}} = \sum_{n=1}^4 \|f_n(I_{\text{out}}) - f_n(I_{\text{GT}})\|_2 \quad (3.2)$$

where f_n symbolizes the n -th layer in the VGG-19 model. The perceptual loss is computed at the $\{\text{conv1_1}, \text{conv2_1}, \text{conv3_1}, \text{conv4_1}\}$.

Discriminator Losses. The discriminator loss follows that of Swap-AE [40], which can be denoted as:

$$\mathcal{L}_{\text{GAN}} = E[-\log(D(I_{\text{out}}))] \quad (3.3)$$

For the brain-focused patch discriminator, we follow the loss of Swap-AE [40], but with slight changes. The final GAN loss for the brain-focused patch discriminator is as follows:

$$\mathcal{L}_{\text{patch}} = E[-\log(D_{\text{patch}}(\text{crops}(\text{valid}(I_{\text{out}})), \text{crops}(\text{valid}(I_{\text{GT}}))))] \quad (3.4)$$

where *crops* operator selects a random patch of size $1/2$ to $1/3$ of the full image dimension on each side and *valid* operator calculate the valid brain regions in the given training mini-batch and then crop according to them.

Total Loss. Considering all the aforementioned losses, the total loss is formalized as:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{pix}}\mathcal{L}_{\text{pix}} + \lambda_{\text{perct}}\mathcal{L}_{\text{perct}} + \lambda_{\text{GAN}}\mathcal{L}_{\text{GAN}} + \lambda_{\text{patch}}\mathcal{L}_{\text{patch}} \quad (3.5)$$

where λ_{pix} , λ_{perct} , λ_{GAN} , and λ_{patch} are the weighting hyper-parameters for each loss.

3.5 Image Quality Metrics

We compared generated and ground truth FA images and tractography using the following metrics.

- **Mean Absolute Error (MAE):** MAE measures the average magnitude of the errors in a set of predictions, without considering their direction [42]. It is the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. Given two monochrome images I and K where they have the same dimensions of $m \times n$ and one of the images is considered a noisy approximation of the other, the MAE can be defined as:

$$MAE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |I(i, j) - K(i, j)| \quad (3.6)$$

- **Mean Squared Error (MSE):** MSE measures the average of the squares of the errors between the estimated values and the actual value [42]. Given two monochrome images I and K where they have the same dimensions of $m \times n$

and one of the images is considered a noisy approximation of the other, the MSE can be defined as:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (3.7)$$

- **Peak Signal-to-Noise Ratio (PSNR):** PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation [43]. It is most easily defined using the MSE as shown below:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (3.8)$$

where MAX_I is the maximum possible pixel value of the image

- **Structural Similarity Index (SSIM):** The SSIM index is a perception-based model that considers image degradation as perceived change in structural information, while also incorporating important perceptual phenomena, including both luminance masking and contrast masking terms [43]. The SSIM index is calculated on various windows of an image. The measure between two windows x and y of common size $N \times N$ is:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (3.9)$$

where μ_x and μ_y are the average of x and y , σ_x^2 and σ_y^2 are the variance of x and y each, σ_{xy} is the covariance of x and y and c_1, c_2 are variables to stabilize the division with weak denominator respectively.

3.6 Comparison of generated and real FA images in low-dimensional representation

To assess the effectiveness of the transformation from original T1-weighted images to FA images, we conducted principal component analysis (PCA) and T-SNE on three types of images: T1-weighted images, real FA images, and FA images generated by Macro2Micro. We analyzed and visualized 1499 subjects in the test dataset. After flattening $65,536 = 256 \times 256$ voxels from MRI slices, we removed the non-brain background, leaving 26,891 voxels for each brain modality, PCA and T-SNE were then applied to the 26,981 features. The incremental PCA and T-SNE were implemented using Python’s ‘scikit-learn’ package. A batch size of 200 was utilized for the incremental PCA, and the visualization was performed using 2 principal components. Furthermore, a learning rate of 100 and a perplexity of 30 were applied in the T-SNE algorithm.

3.7 Prediction of biological and cognitive variables using predicted FA images

The image-to-image translation approach is powerful in terms of image transformation; however, there is a major concern about potential damage to biological characteristics [44]. To ensure that biological information is preserved during the transformation process and that biological features are not lost during the T1 to FA image conversion, we conducted a task of predicting the sex, intelligence, and Attention-Deficit/Hyperactivity Disorder (ADHD) diagnosis of children using both predicted and real FA images. The total intelligence score in the NIH toolbox was used for the intelligence prediction task. The parent KSAD (Kiddie Schedule for Affective Disorders and Schizophrenia) was employed as the diagnostic tool for assessing ADHD in children. Children who did not receive a diagnosis of KSAD and obtained a total

problem score below 65 on the Child Behavior Checklist (CBCL) were designated as the healthy control group. We assessed the accuracy of classification tasks using FA images generated by CycleGAN, Pix2Pix, and Macro2Micro. Additionally, we compared the predictive performance of these generated FA images with real FA images or T1 images. Logistic regression with L2 norm was employed for gender classification, and PCA was used as the feature extraction method. PCA was applied to flattened, background-removed brain features, and feature reduction was performed by selecting the top 1,000 components (sex) and 500 components (Total intelligence, ADHD) that accounted for more than 90% of the explained variance. The analysis was conducted on the test dataset of the image generation task. Among these datasets, 20% were held out as a test set, while the remaining 80% of the subjects were used for a training set in a 5-fold cross-validation. The models trained on these five folds were used to evaluate the performance on the held-out test set, and the results were averaged. Accuracy and AUROC were utilized for evaluating performances in classification tasks. Pearson product-moment correlation coefficients and mean square error (MSE) were chosen for evaluating the regression task.

3.8 Prediction of Tractography from FA images

We follow the model design and hyper-parameters of the section 3.3 to synthesize the 2D tractography of the vmPFC region from FA images. However, in most cases, the regions of interest (ROI) from 2D tractography are very sparse. This makes our brain-focused patch discriminator hard to train. As a result, we exclude the brain-focused patch discriminator for this task.

3.9 Experimental Settings

Data. We use the Adolescent Brain Cognitive Development (ABCD) dataset [17], which consists of comprehensive developmental data including structural brain MRI in children across various sites in the United States. The data utilized in this study is obtained from ABCD, and the image acquisition protocol and minimal processing pipeline can be found in previous studies [17, 16]. For image-to-image translation between T1-weighted images and DTI, the T1 image and fractional anisotropy (FA) for DTI have dimensions of $256 \times 256 \times 256$ and a voxel size of 1mm. We use quality controlled 7,669 subjects. For image-to-image translation between FA and tractography, the FA image has dimensions of $190 \times 190 \times 190$ and a voxel size of 1.25mm. The analysis includes a total of 6,365 subjects. Tractography is performed using 1 million tracks derived from DTI data. Only the streamlines passing by vmPFC (ventromedial prefrontal cortex) based on the aparc+aseg atlas [45] is extracted, and the shape is adjusted to match the FA image.

Implementation Details. During training, all images are loaded as 256x256 pixels and scaled to $[0, 1]$. The model is trained using the Adam optimizer [46] with a learning rate of 0.0002 and a batch size of 8 for 200 epochs. The encoder feature map has dimensions of (128, 64, 64) for high and (128, 32, 32) for low-frequency components. Baseline and our experiments are conducted using the PyTorch framework [47] on a single NVIDIA RTX A5000(24G) GPU. All the experiments for the ablation study use 4 NVIDIA RTX A5000(24G) GPUs.

Chapter 4

RESULTS

4.1 Qualitative Evaluation

Figure 4.1 shows the qualitative comparison between our model and other image-to-image translation algorithms such as Pix2Pix [23] and CycleGAN [24]. Note that all these results are generated from brain slices of different subjects. In terms of white matter and overall brain structure, all models generate brain images with comparable quality. However, Pix2Pix and CycleGAN could not recognize and generate intricate micro-scale structures in many cases. For example, baseline models mostly neglected and underexpressed the microscopic white matter at the boundaries of the brain (e.g., 4th and 5th columns in Figure 4.1), whereas our model does not. The Macro2Micro not only generates the most comparable images to the ground truths but also captures structural details lost in the original and reconstructs parts that were previously disconnected or absent.

It is worth noting that our model learns both macro-structure from T1-weighted images and micro-structure from FA images. As shown in Figure 4.1, our results learn the presence of the white matter from the FA image while determining the morphology of the white matter from the T1 image (see the second row of the figure). Similarly, our results bend the straight white matter line in ground truth FA to its more specific endpoint by referencing T1’s macro-structure (see the 8th row of the figure).

Also, our model preserves the structure from T1 which is not present in FA. In most cases, the structure is detected from the FA images whereas the T1 images reveal the connection of the white matter. Generated FA image of our model connects the white matters like the T1 scan, whereas the ground truth FA image does not (see the upper left corner of the 4th row).

Moreover, as our model conserves the white matter which is not clearly visible in most FA images (e.g., the middle of the 6th row), Macro2Micro has strong robustness to noise and can generate high-resolution images compared to other image-to-image translation methods.

Compared to actual ground truth FA images, results of Macro2Micro show outstanding performance (Figure 4.2). To test whether our model has mode collapse issues, we tested on the same slice of different subjects. It is shown that the results demonstrate the absence of mode collapse and the successful generation of a wide range of realistic FA images that accurately depict the subject’s diversity. We generate not only the whole brain structure of actual FA images but also the microscopic aspect of white matter in the brain’s periphery without noise. The structural location and FA value are both well-learned, and the isotropic or anisotropic movement of water molecules is similar to reality.

4.2 Quantitative Evaluation

Baseline. Compared to the two baselines, Macro2Micro shows the best performance quantitatively. As demonstrated in Table 4.1, our results achieve the best SSIM, PSNR, MAE, and MSE among various image-to-image translation algorithms. In addition, our results show the state-of-the-art performance even if we compared just the white matter where the FA value is greater than 0.2 (Table 4.2).

Table 4.1: Quantitative comparison with generated whole brain FA images. The best outcomes are shown in **bold**. \uparrow : Higher is better. \downarrow : Lower is better.

Methods	SSIM (\uparrow)	PSNR (\uparrow)	MAE (\downarrow)	MSE(\downarrow)
Pix2Pix	0.8310	24.7738	0.1469	0.1292
Cycle GAN	0.8332	24.6660	0.1477	0.1299
Ours (Macro2Micro)	0.8600	25.7560	0.1383	0.1226

Table 4.2: Quantitative comparison with baseline models using only white matter whose FA value is bigger than 0.2 from the generated FA images.

Methods	SSIM (\uparrow)	PSNR (\uparrow)	MAE (\downarrow)	MSE(\downarrow)
Pix2Pix	0.8354	24.8992	0.1374	0.1200
Cycle GAN	0.8369	24.7585	0.1383	0.1208
Ours (Macro2Micro)	0.8627	25.8493	0.1300	0.1146

PCA and T-SNE. In Figure 4.3, the application of PCA and T-SNE resulted in the representation of T1, FA, and generated FA images within two principal components. In the case of PCA, the two principal components individually accounted for 13.7 % and 2.7 % of the variances. Our validation confirmed the distinctiveness of the original T1 images from the real FA images. Notably, both PCA and T-SNE demonstrated that the generated FA images exhibited significant dissimilarity compared to the original T1 images while displaying an overlap with the real FA images in the low-dimensional representation.

Table 4.3: Sex classification performance of real T1, real FA, and generated FA images. GT: Ground Truth.

Input	AUROC (\uparrow)	ACC (\uparrow)
T1	0.7820	0.7133
FA (GT)	0.7641	0.7066
Pix2Pix FA	0.7565	0.68
Cycle GAN FA	0.7534	0.6766
Ours (Macro2Micro FA)	0.7726	0.6866

4.3 Generated FA images by Macro2Micro can efficiently predict sex, ADHD and intelligence

We present the performance of the generalized linear model (GLM) on sex, intelligence, and ADHD prediction tasks. In Table 4.3, The sex classification performance of T1 images yielded an AUROC of 0.782, slightly higher than that of real FA images (0.7641). The FA images generated by Pix2Pix and CycleGAN exhibited relatively lower sex classification performance than real FA images. However, Macro2Micro demonstrated relatively higher performance in AUROC (0.7726) and accuracy (0.6866) than the other models, surpassing the AUROC of real FA images (0.7641). In Table 4.4, the FA images generated by Macro2Micro exhibited comparable performance in predicting ADHD diagnosis (AUROC of 0.4926) to real FA images (AUROC of 0.4812).

Table 4.4: ADHD classification performance of real T1, real FA, and generated FA images.

Input	AUROC (\uparrow)	ACC (\uparrow)
T1	0.5034	0.5479
FA (GT)	0.4812	0.5342

Pix2Pix FA	0.5532	0.5821
Cycle GAN FA	0.4445	0.4794
Ours (Macro2Micro FA)	0.4926	0.5136

Pix2Pix showed the best performance in predicting ADHD diagnosis (0.5532), which outperforms the performances of real T1 (AUROC of 0.5034) and FA images (AUROC of 0.4812). In Table 4.5, though FA images generated by Pix2pix exhibited the best intelligence prediction performances (Correlation Coefficient of 0.187), predicted FA images from Macro2Micro also showed better performances (Correlation Coefficient of 0.166) than real T1 images (0.159) and FA images (0.124).

4.4 Ablation Studies

Octave Convolution. In this section, we will show how Octave Convolution influences our suggested model. When Octave Convolution was utilized, both the SSIM and the PSNR improved, as demonstrated in Table 4.6. The α value in Octave Convolution represents the percentage of low-frequency features relative to total features and verifies

Table 4.5: Intelligence regression performance of real T1, real FA, and generated FA images.

Input	Corr. Coef. (\uparrow)	MSE (\downarrow)
T1	0.159	0.777
FA (GT)	0.124	0.832

Pix2Pix FA	0.187	0.784
Cycle GAN FA	0.066	0.836
Ours (Macro2Micro FA)	0.166	0.797

Table 4.6: **Ablation study** Depending on whether Octave Convolution is applied and the difference in α value.

Methods	SSIM (\uparrow)	PSNR (\uparrow)	MAE (\downarrow)	MSE(\downarrow)	Time(\downarrow)
No Octave	0.8598	25.8797	0.1280	0.1122	0.0056
Oct $\alpha = 0.25$	0.8640	26.1565	0.1417	0.1264	0.0126
Oct $\alpha = 0.75$	0.8597	25.9887	0.1414	0.1258	0.0129
Macro2Micro ($\alpha = 0.50$)	0.8631	26.0478	0.1374	0.1221	0.0110

the difference. In Figure 4.4, at an α of 0.25, it is clear that the low-frequency image has little information, while the high-frequency image contains a lot of information. Low-frequency images reveal more information with an α of 0.75 than at values of 0.5 and 0.25. On the other hand, if α is set to 0.5, it's clear that the image-based separation

Table 4.7: **Ablation study** for brain-focused patch discriminator.

Methods	SSIM (\uparrow)	PSNR (\uparrow)	MAE (\downarrow)	MSE(\downarrow)
No brain PD	0.8663	26.2560	0.1440	0.1288
Macro2Micro	0.8631	26.0478	0.1374	0.1221

Table 4.8: **Ablation study** of perceptual loss using pre-trained VGG network.

Methods	SSIM (\uparrow)	PSNR (\uparrow)	MAE (\downarrow)	MSE(\downarrow)
No VGG loss	0.8565	25.7207	0.1407	0.1251
Macro2Micro	0.8631	26.0478	0.1374	0.1221

of high and low frequencies is optimal. The best results in terms of MAE, MSE, and inference time were achieved with an α value of 0.5, while an α value of 0.25 yielded the higher SSIM and PSNR but achieved similar image quality with an α value of 0.5 (Table 4.6).

Brain-focused Patch Discriminator. The brain-focused Patch Discriminator contributes to sophisticated images. In Figure 4.5, generated images not using brain-focused Patch Discriminator have checkerboard patterns (e.g., middle of the 2th row). Also, they have artifacts like white dots along the borders of the brain. Our Macro2Micro with the brain-focused Patch Discriminator outperforms the model without the brain-focused patch discriminator and generates higher-resolution images without artifacts.

Table 4.9: **Ablation study** for different normalization and activation functions.

Methods	SSIM (\uparrow)	PSNR (\uparrow)	MAE (\downarrow)	MSE(\downarrow)
Batch Norm	0.8521	25.4675	0.1425	0.1258
Instance Norm	0.8605	25.8663	0.1419	0.1260
ReLU	0.8575	25.8609	0.1399	0.1242
Macro2Micro	0.8631	26.0478	0.1374	0.1221

VGG Perceptual Loss. The ablation study of perceptual loss using a pre-trained VGG network shows that Macro2Micro outperformed the model that did not use VGG loss in every evaluation metric (Table 4.8.) Also, all of the results without VGG perceptual loss have severe artifacts that look like a brain skull at the boundary of the brain (Figure 4.6).

Normalization & Activation Function. Normalization and activation functions affect the quality of the generated images. In Table 4.9, we found that the normalization including batch normalization or instance normalization, and the ReLU activation function don't aid our Macro2Micro in synthesizing high-quality images. The combination of not using normalization and using leaky ReLU as an activation function performs the highest SSIM and PSNR scores and the lowest MAE and MSE scores. Also, the results using normalization or ReLU function have artifacts like dotted lines at the borders of the brain (Figure 4.7).

Table 4.10: Quantitative Comparison of generated tractography images with baseline models.

Methods	SSIM (\uparrow)	PSNR (\uparrow)	MAE (\downarrow)	MSE(\downarrow)
Pix2Pix	0.8206	15.0687	0.0654	0.0620
CycleGAN	0.8048	15.7659	0.1270	0.1236
Macro2Micro	0.8551	17.0572	0.0521	0.0474

4.5 Effectiveness of Macro2Micro along the distance from the center of the brain

Although our method was trained using only the central slice of the brain, it works well not only in the center but also in its periphery. Figure 4.8 depicts how each evaluation metrics shift from the center to its periphery. As shown, our model demonstrates relatively robust performance in both the peripheral and the center slice, demonstrating our model’s superior performance in generability and robustness. It is worth noting that the score is higher towards the extreme end of the brain than in the center. I believe this is because the size of the brain in the image itself is small in comparison to the central slice, therefore the metric includes more backgrounds, resulting in improved performance.

4.6 FA Image Translation to Tractography

Our model’s results were most comparable to the ground truth tractography. In Figure 4.9, the overall position of the fiber passing through the vmPFC region and the dense region where many streamlines pass were well predicted. Compared to Pix2Pix

and CycleGAN, which produce more noisy results around the target tractography, our results have little noise and generate the target regions more accurately while showing the state-of-the art performance in SSIM, PSNR, MAE and MSE score (Table 4.10).

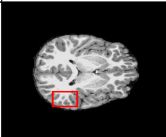
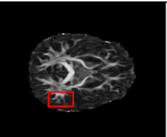
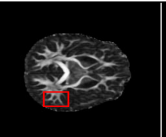
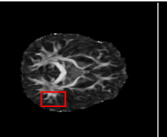
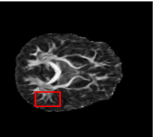





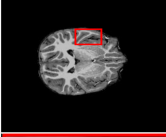
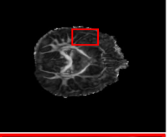
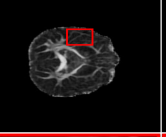
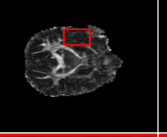
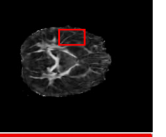






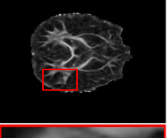
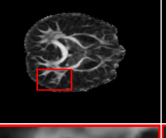
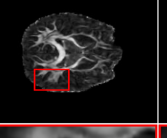
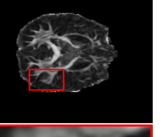





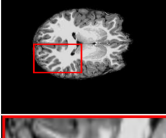
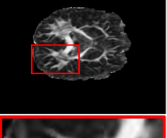
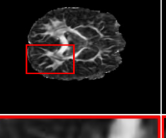
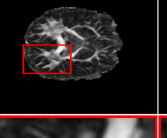
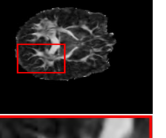

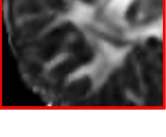
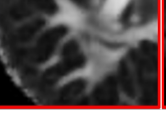


	T1	FA (GT)	Macro2Micro (Ours)	Pix2Pix	Cycle GAN
Synthesized Image					
Magnified Image					
Synthesized Image					
Magnified Image					
Synthesized Image					
Magnified Image					
Synthesized Image					
Magnified Image					

Figure 4.1: Qualitative Comparison of images made with Macro2Micro (our model), and images made with Pix2Pix and CycleGAN. Note that Pix2Pix and CycleGAN are the baselines.

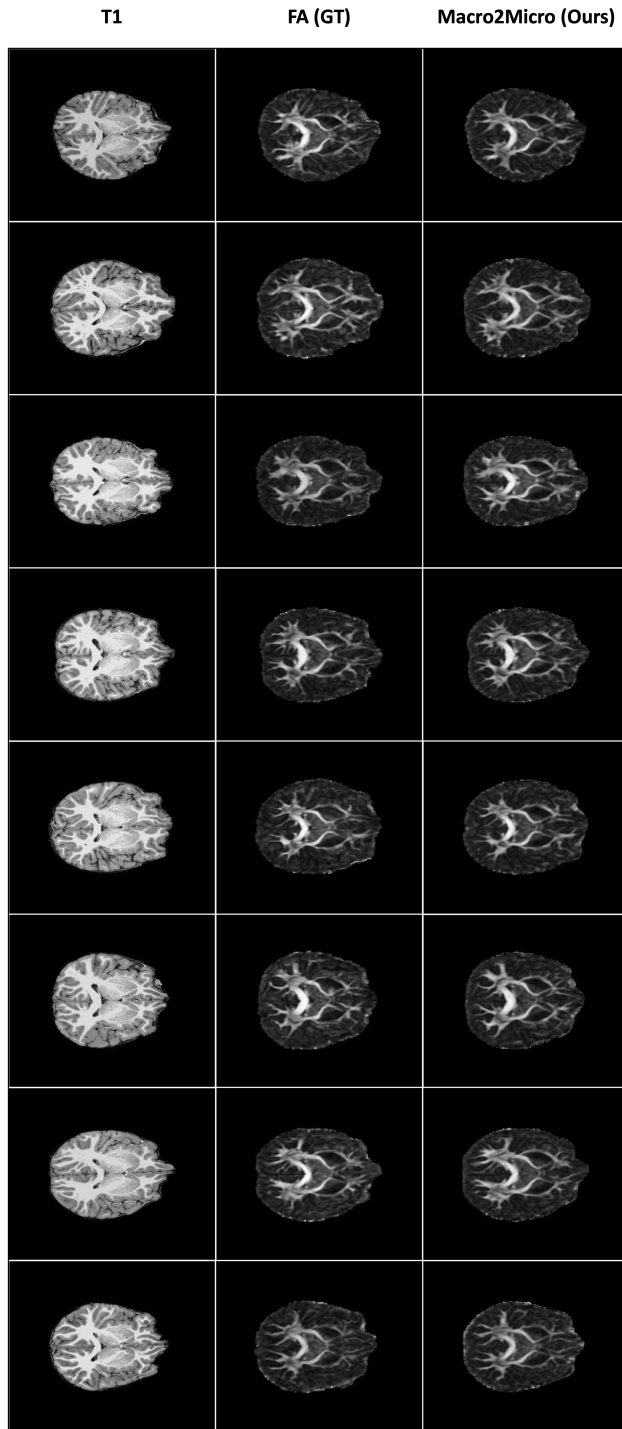


Figure 4.2: Generated realistic FA images from T1-weighted images using Macro2Micro model.

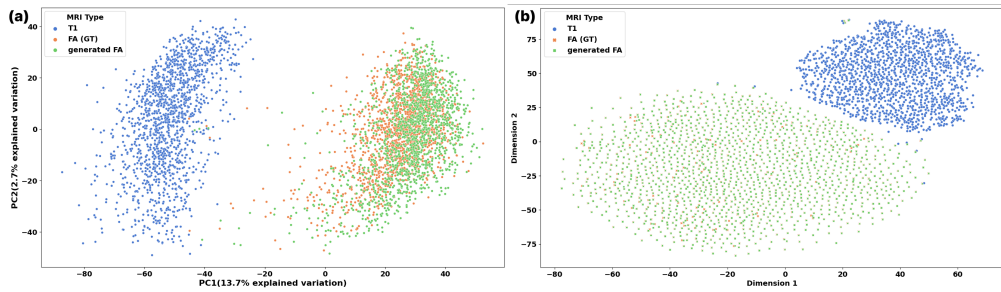


Figure 4.3: PCA and T-SNE results for the T1, FA (Ground Truth), and generated FA images from Micro2Macro. **(a)** PCA **(b)** T-SNE

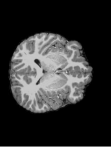
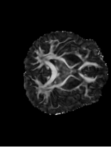
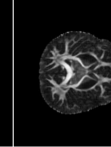
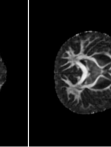
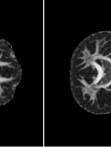
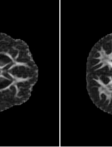
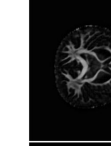
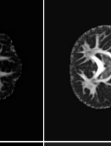
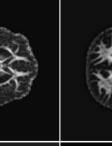

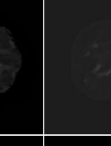
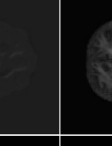
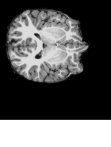
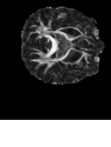
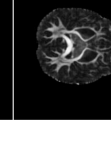
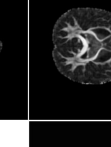
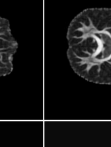
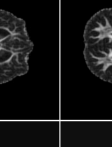
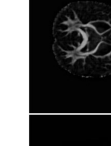
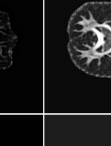
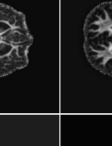
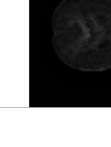


	T1	FA (GT)	No Octave	Macro2Micro $\alpha=0.5$	Octave $\alpha=0.25$	Octave $\alpha=0.75$
Synthesized Image						
High Frequency Image						
Low Frequency Image						
Synthesized Image						
High Frequency Image						
Low Frequency Image						

Figure 4.4: **Ablation study** Qualitative comparison of the effectiveness of Octave Convolution.

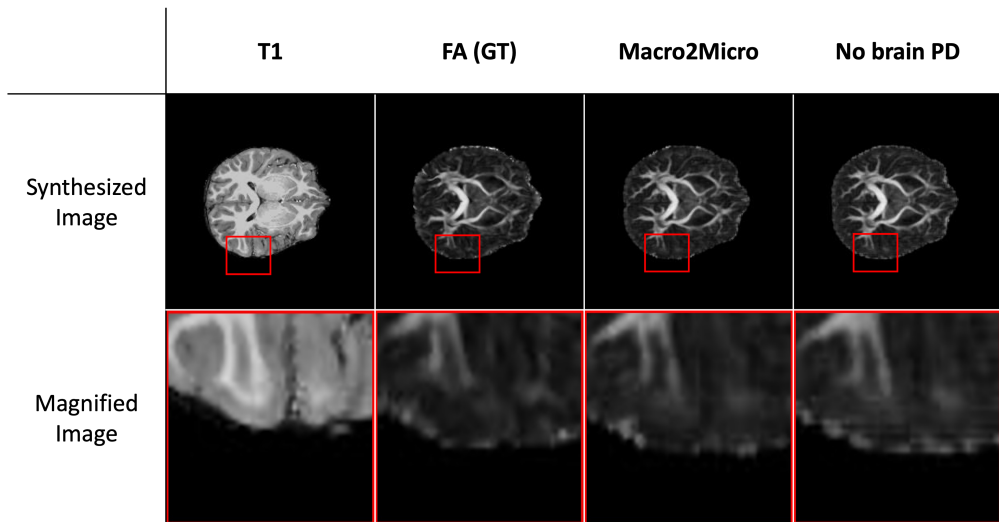


Figure 4.5: **Ablation study** Generated FA images with and without the use of Brain-focused Patch-wise Discriminator.

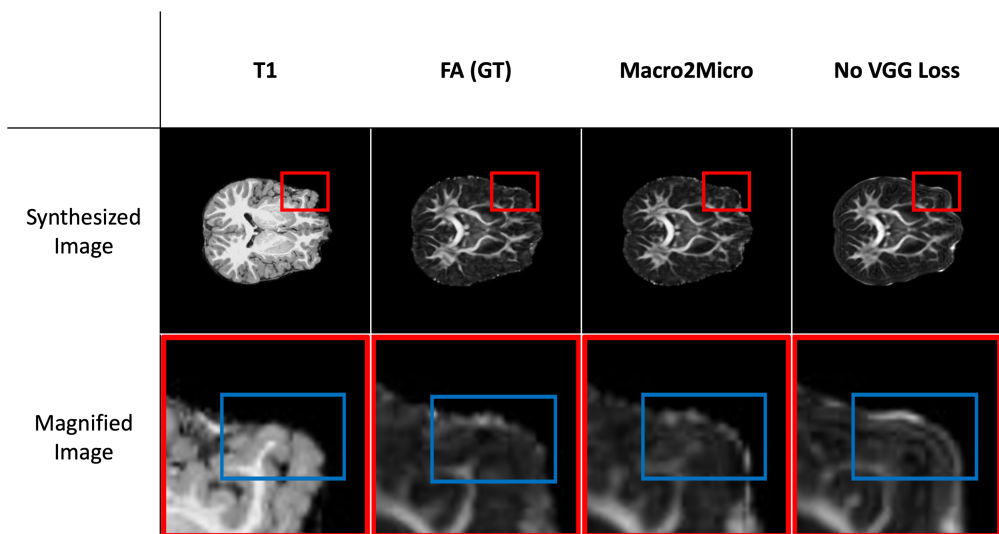


Figure 4.6: **Ablation study** Qualitative comparison of the effectiveness of pre-trained VGG perceptual loss.

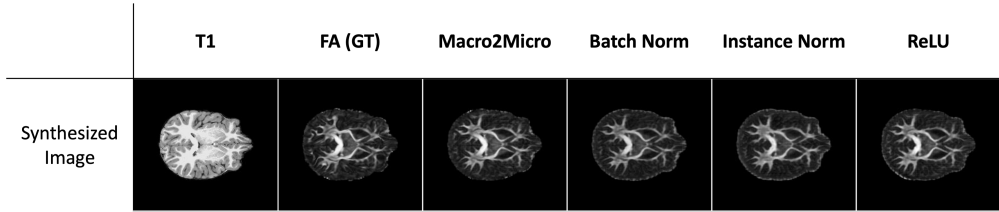


Figure 4.7: **Ablation study** Qualitative comparison of the effectiveness of normalization and activation function.

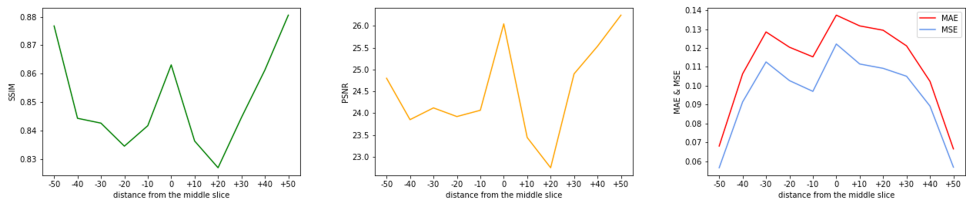


Figure 4.8: Graph depicting the performance with different distances from the middle slice.

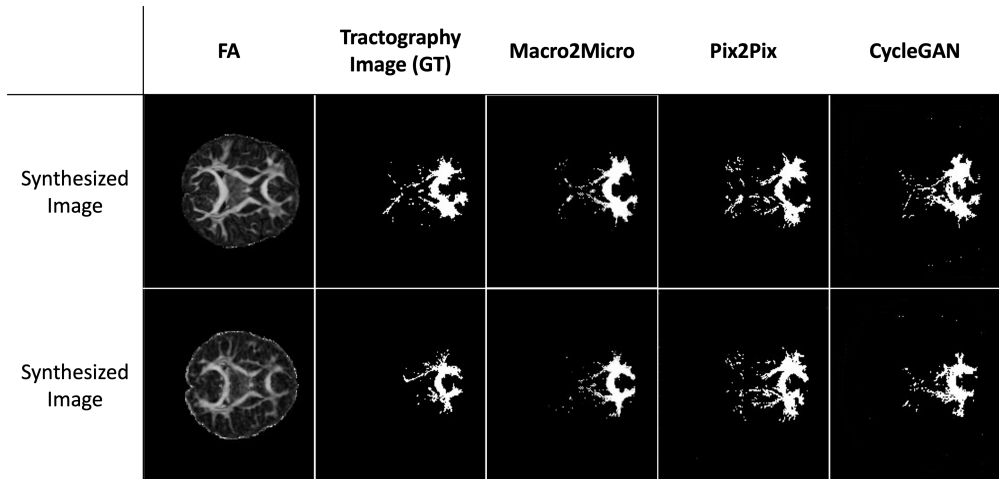


Figure 4.9: Qualitative comparison of generated 2D Tractography from FA images.

Chapter 5

DISCUSSIONS AND CONCLUSIONS

In this study, we propose a new model called Macro2Micro to effectively predict micro-structure based on the macro-structure of brain anatomy. We validate the capability of Macro2Micro to decode FA images containing micro-structure information from T1-weighted MRI, which measures the macro-structure of the brain. Octave Convolution, which separates the structural characteristics of different hierarchies in the brain by frequency, is employed for this purpose. Various metrics that evaluate the quality of generated images are utilized for quantitative assessment. Furthermore, feature reduction methods such as PCA and T-SNE are applied to evaluate the similarity between the FA images generated from T1 images in low-dimensional image representations and real FA images.

The results reveal the superior quantitative performance of Macro2Micro over widely used models such as CycleGAN and Pix2Pix. Additionally, qualitative analysis confirms the similarity between the generated images and real FA images, surpassing the quality of conventional T1 images. Moreover, through a sex, intelligence, and ADHD prediction task, it is confirmed that the generated images perform modality conversion without compromising biological characteristics. While previous studies have been limited to the task of transforming T1 to DTI, our study demonstrates the potential of Macro2Micro in predicting tractography from FA images. The effective mapping of nonlinear relationships between brain structures from macro to micro us-

ing generative models in deep learning offers a new perspective in understanding the relationship between brain structure and behavior.

Our model preserves the structural information from T1 to generate FA, resulting in generated FA images that not only exhibit a high level of agreement with the ground truth FA but often surpass it in terms of structural information at a high resolution. It accurately generates high-resolution white matter structures in locations where the ground truth FA exhibits disconnected or distorted regions. This indicates that our model learns the association between macro-scale and micro-scale structures, allowing it to extract additional information from T1 that is missing in the ground truth FA, resulting in a more accurate generation of microscopic white matter structures. It demonstrates that macro-scale knowledge is useful for designing micro-structures, in addition to micro-scale information.

Unlike images that can be easily encountered in the computer vision domain, brain images focus more on structural information, contrast, and edges rather than color information. Octave Convolution, which divides images into high-frequency and low-frequency components, can be used to efficiently encode and generate brain images, effectively maintaining high image qualities such as contrast or structure.

When dealing with brain images, it is common to train models from scratch rather than relying on pre-trained models. However, the pre-trained VGG model with computer vision data used for perceptual loss of images removes skull-shaped artifacts in the FA images. This suggests that, on some level, generic image data and brain images may share image features. This finding indicates the potential for extending pre-trained models from general data to the brain domain, opening up possibilities for further exploration.

In most cases, the brain images include a significant portion of the redundant back-

ground. Even if the loss values of background are small, the background potentially impedes the learning of the brain region. A patch discriminator is employed to address the problem by cropping the image around the brain. The patch discriminator extracts patches from images and determines whether each patch belongs to the actual brain data or not. This approach focused on the loss of only brain regions during training. As a result, the quality of generated images gets higher in the detailed region, and artifacts such as checkerboard patterns or dotted lines at the brain boundary could be effectively solved. The brain boundary becomes more similar to the actual boundaries. Background and overall image quality can be adequately learned using pixel loss and VGG perceptual loss.

Microscopic quality and artifacts specific to brain images are improved by the brain-focused Patch Discriminator. This approach helps to alleviate the challenges posed by the background dominance and allows the model to better capture and generate high-quality brain images, with improved details, accuracy in brain region representation, and reduced artifacts.

Through various quantitative and qualitative metrics, it is demonstrated that the FA images generated from T1 images via the proposed model show high similarity with actual FA images, exceeding the similarity with T1 images. However, even if the generated images show distributional resemblance to actual FA images in techniques such as PCA or T-SNE, there exists a risk of compromising biological information from the actual T1 images during the image-to-image translation process [44]. If the generated FA images result in the loss of important individual differences, the clinical application would be significantly limited. We demonstrated classification and regression performance in the generated FA images, which is similar to that of actual FA images. These similarities indicate that the fundamental individual biological information was conserved or amplified during the image translation process. Macro2Micro exhibits higher performance in predicting sex, intelligence and ADHD compared to actual FA, sug-

gesting that it operated in a manner that preserved or enhanced biological information during the transformation from T1 to FA. Given that real T1_w images exhibit superior predictive performance for sex, intelligence and ADHD compared to real FA images, we propose that the structural information present in T1 images was preserved and positively influenced the generated FA images. Furthermore, the heightened performance of generated FA images might be attributed to their resilience against scanner effects. Training the model on multiple subjects can enhance its capacity to establish a consistent mapping between macro-structure and micro-structure. This robustness to versatile factors contributes to the emergence of more prominent biological characteristics in the generated images. Evaluation of generated images using such approaches has not been frequently conducted in previous generative research, emphasizing the high value of utilizing the actual generated images. If future studies extend their research to variables beyond sex, the utilization of generated images can provide researchers with various valuable benefits.

Limitations. On the other hand, in the Macro2Micro, only a single slice positioned in the middle along the z-axis was used for training. Consequently, it did not perform well on slices located at the ends of the z-axis. In future research, incorporating all slices of the brain and training the model could lead to generating high-quality 2D slices and concatenating them to create realistic 3D images. Although the results of predicting tractography from FA images outperformed other baseline models, there were still noticeable differences compared to actual tractography images. However, this study demonstrated the ability to generate tractography from macro structures using image-to-image translation. We anticipate that further research will enable us to create more accurate and sophisticated images efficiently.

Future Impacts. This research successfully generated microscopic structures from

macro-structures of the human brain, demonstrating the interrelation between brain structures at different scales. Furthermore, it effectively learned subject-independent structures by accommodating individual variations. In addition, when there is not enough time to scan various modalities in clinical practice, this approach enables to translate different microscopic modalities from structural MRI, which is more readily obtainable. This significantly reduces the time required for diagnosis and decreases financial costs. Additionally, this approach facilitates multi-modality research by generating data for individuals who only have access to structural MRI or for populations where acquiring imaging data is challenging, such as pediatric patients.

Bibliography

- [1] K. J. Friston, “Functional integration in the brain,” *Human Brain Function 2nd edn Academic Press, San Diego*, pp. 971–997, 2004.
- [2] M. Cercignani and S. Bouyagoub, “Brain microstructure by multi-modal mri: Is the whole greater than the sum of its parts?,” *Neuroimage*, vol. 182, pp. 117–127, 2018.
- [3] X. Gu, H. Knutsson, M. Nilsson, and A. Eklund, “Generating diffusion mri scalar maps from t1 weighted images using generative adversarial networks,” in *Image Analysis: 21st Scandinavian Conference, SCIA 2019, Norrköping, Sweden, June 11–13, 2019, Proceedings 21*, pp. 489–498, Springer, 2019.
- [4] N. Ibarretxe-Bilbao, C. Junque, M. J. Marti, and E. Tolosa, “Brain structural mri correlates of cognitive dysfunctions in parkinson’s disease,” *Journal of the neurological sciences*, vol. 310, no. 1-2, pp. 70–74, 2011.
- [5] G. Parker, “Analysis of mr diffusion weighted images,” *The British journal of radiology*, vol. 77, no. suppl_2, pp. S176–S185, 2004.
- [6] M. A. Horsfield, M. Lai, S. L. Webb, G. J. Barker, P. S. Tofts, R. Turner, P. Rudge, and D. H. Miller, “Apparent diffusion coefficients in benign and secondary progressive multiple sclerosis by nuclear magnetic resonance,” *Magnetic resonance in medicine*, vol. 36, no. 3, pp. 393–400, 1996.
- [7] S. Roosendaal, J. J. Geurts, H. Vrenken, H. Hulst, K. S. Cover, J. Castelijns, P. J. Pouwels, and F. Barkhof, “Regional dti differences in multiple sclerosis patients,” *Neuroimage*, vol. 44, no. 4, pp. 1397–1403, 2009.

- [8] H. Chabriat, S. Pappata, C. Poupon, C. Clark, K. Vahedi, F. Poupon, J. Mangin, M. Pachot-Clouard, A. Jobert, D. Le Bihan, *et al.*, “Clinical severity in cadasil related to ultrastructural damage in white matter: in vivo study with diffusion tensor mri,” *Stroke*, vol. 30, no. 12, pp. 2637–2643, 1999.
- [9] M. O’sullivan, T. Barrick, R. Morris, C. Clark, and H. Markus, “Damage within a network of white matter regions underlies executive dysfunction in cadasil,” *Neurology*, vol. 65, no. 10, pp. 1584–1590, 2005.
- [10] F. Segawa, J. Kishibayashi, K. Kamada, N. Sunohara, and M. Kinoshita, “Mri of paraventricular white matter lesions in amyotrophic lateral sclerosis—analysis by diffusion-weighted images,” *No to Shinkei= Brain and Nerve*, vol. 46, no. 9, pp. 835–840, 1994.
- [11] H.-P. Müller, M. R. Turner, J. Grosskreutz, S. Abrahams, P. Bede, V. Govind, J. Prudlo, A. C. Ludolph, M. Filippi, and J. Kassubek, “A large-scale multicentre cerebral diffusion tensor imaging study in amyotrophic lateral sclerosis,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 87, no. 6, pp. 570–579, 2016.
- [12] C. D. Mayo, M. A. Garcia-Barrera, E. L. Mazerolle, L. J. Ritchie, J. D. Fisk, J. R. Gawryluk, and A. D. N. Initiative, “Relationship between dti metrics and cognitive function in alzheimer’s disease,” *Frontiers in aging neuroscience*, vol. 10, p. 436, 2019.
- [13] Y. Chen, Y. Wang, Z. Song, Y. Fan, T. Gao, and X. Tang, “Abnormal white matter changes in alzheimer’s disease based on diffusion tensor imaging: A systematic review,” *Ageing Research Reviews*, p. 101911, 2023.
- [14] Y. Lao, B. Nguyen, S. Tsao, N. Gajawelli, M. Law, H. Chui, M. Weiner, Y. Wang, and N. Leporé, “A t1 and dti fused 3d corpus callosum analysis in mci subjects

- with high and low cardiovascular risk profile,” *NeuroImage: Clinical*, vol. 14, pp. 298–307, 2017.
- [15] H. Cai, A. Li, G. Yu, X. Yang, and M. Liu, “Brain age prediction in developing childhood with multimodal magnetic resonance images,” *Neuroinformatics*, vol. 21, no. 1, pp. 5–19, 2023.
- [16] D. J. Hagler Jr, S. Hatton, M. D. Cornejo, C. Makowski, D. A. Fair, A. S. Dick, M. T. Sutherland, B. Casey, D. M. Barch, M. P. Harms, *et al.*, “Image processing and analysis methods for the adolescent brain cognitive development study,” *Neuroimage*, vol. 202, p. 116091, 2019.
- [17] B. J. Casey, T. Cannonier, M. I. Conley, A. O. Cohen, D. M. Barch, M. M. Heitzeg, M. E. Soules, T. Teslovich, D. V. Dellarco, H. Garavan, *et al.*, “The adolescent brain cognitive development (ab cd) study: imaging acquisition across 21 sites,” *Developmental cognitive neuroscience*, vol. 32, pp. 43–54, 2018.
- [18] U. Lueken, M. Muehlhan, H.-U. Wittchen, T. Kellermann, I. Reinhardt, C. Konrad, T. Lang, A. Wittmann, A. Ströhle, A. L. Gerlach, *et al.*, “(don’t) panic in the scanner! how panic patients with agoraphobia experience a functional magnetic resonance imaging session,” *European neuropsychopharmacology*, vol. 21, no. 7, pp. 516–525, 2011.
- [19] A. Copeland, E. Silver, R. Korja, S. J. Lehtola, H. Merisaari, E. Saukko, S. Sinisalo, J. Saunavaara, T. Lähdesmäki, R. Parkkola, *et al.*, “Infant and child mri: a review of scanning procedures,” *Frontiers in neuroscience*, vol. 15, p. 666020, 2021.
- [20] J.-D. Tournier, R. Smith, D. Raffelt, R. Tabbara, T. Dhollander, M. Pietsch, D. Christiaens, B. Jeurissen, C.-H. Yeh, and A. Connelly, “Mrtrix3: A fast, flex-

- ible and open software framework for medical image processing and visualisation,” *Neuroimage*, vol. 202, p. 116137, 2019.
- [21] I. Benou and T. Riklin Raviv, “Deeptract: A probabilistic deep learning framework for white matter fiber tractography,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III* 22, pp. 626–635, Springer, 2019.
- [22] T. Nguyen-Duc, T. M. Quan, and W.-K. Jeong, “Frequency-splitting dynamic mri reconstruction using multi-scale 3d convolutional sparse coding and automatic parameter selection,” *Medical image analysis*, vol. 53, pp. 179–196, 2019.
- [23] 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*, pp. 1125–1134, 2017.
- [24] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.
- [25] Q. Yang, N. Li, Z. Zhao, X. Fan, E. Chang, and Y. Xu, “Mri cross-modality image-to-image translation,” *Scientific Reports*, vol. 10, p. 3753, 02 2020.
- [26] B. Ancil-Robitaille, C. Desrosiers, and H. Lombaert, “Manifold-aware cyclegan for high-resolution structural-to-dti synthesis,” 2020.
- [27] X. Liu, J. L. Prince, F. Xing, J. Zhuo, T. Reese, M. Stone, G. El Fakhri, and J. Woo, “Attentive continuous generative self-training for unsupervised domain adaptive medical image translation,” *Medical Image Analysis*, vol. 88, p. 102851, 2023.

- [28] W. Klimesch, “Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis,” *Brain research reviews*, vol. 29, no. 2-3, pp. 169–195, 1999.
- [29] E. Başar, C. Başar-Eroglu, S. Karakaş, and M. Schürmann, “Gamma, alpha, delta, and theta oscillations govern cognitive processes,” *International journal of psychophysiology*, vol. 39, no. 2-3, pp. 241–248, 2001.
- [30] G. Buzsaki and A. Draguhn, “Neuronal oscillations in cortical networks,” *science*, vol. 304, no. 5679, pp. 1926–1929, 2004.
- [31] G. Doucet, M. Naveau, L. Petit, N. Delcroix, L. Zago, F. Crivello, G. Jobard, N. Tzourio-Mazoyer, B. Mazoyer, E. Mellet, *et al.*, “Brain activity at rest: a multi-scale hierarchical functional organization,” *Journal of neurophysiology*, vol. 105, no. 6, pp. 2753–2763, 2011.
- [32] K. Diba, H. A. Lester, and C. Koch, “Intrinsic noise in cultured hippocampal neurons: experiment and modeling,” *Journal of Neuroscience*, vol. 24, no. 43, pp. 9723–9733, 2004.
- [33] K. J. Miller, L. B. Sorensen, J. G. Ojemann, and M. Den Nijs, “Power-law scaling in the brain surface electric potential,” *PLoS computational biology*, vol. 5, no. 12, p. e1000609, 2009.
- [34] A. Eke, P. Herman, L. Kocsis, and L. Kozak, “Fractal characterization of complexity in temporal physiological signals,” *Physiological measurement*, vol. 23, no. 1, p. R1, 2002.
- [35] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

- [36] Y. Chen, H. Fan, B. Xu, Z. Yan, Y. Kalantidis, M. Rohrbach, S. Yan, and J. Feng, “Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution,” in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3435–3444, 2019.
- [37] T. Lindeberg, *Scale-space theory in computer vision*, vol. 256. Springer Science & Business Media, 2013.
- [38] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.
- [39] A. Odena, V. Dumoulin, and C. Olah, “Deconvolution and checkerboard artifacts,” *Distill*, vol. 1, no. 10, p. e3, 2016.
- [40] T. Park, J.-Y. Zhu, O. Wang, J. Lu, E. Shechtman, A. Efros, and R. Zhang, “Swapping autoencoder for deep image manipulation,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 7198–7211, 2020.
- [41] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*, pp. 694–711, Springer, 2016.
- [42] C. J. Willmott and K. Matsuura, “Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance,” *Climate Research*, vol. 30, pp. 79–82, 2005.
- [43] A. Hore and D. Ziou, “Image quality metrics: Psnr vs. ssim,” in *2010 20th international conference on pattern recognition*, pp. 2366–2369, IEEE, 2010.

- [44] J. P. Cohen, M. Luck, and S. Honari, “Distribution matching losses can hallucinate features in medical image translation,” May 2018.
- [45] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. Van Der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, *et al.*, “Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain,” *Neuron*, vol. 33, no. 3, pp. 341–355, 2002.
- [46] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [47] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, “Pytorch: An imperative style, high-performance deep learning library,” *Advances in neural information processing systems*, vol. 32, 2019.

국문초록

뇌는 여러 뇌 영역의 고도로 국소화된 기능과 신경 연결을 통한 영역의 통합으로 구성된다. 이러한 뇌의 신경 연결은 끊임없이 변화하는 환경에 효과적으로 대응하기 위해 시스템 및 시냅스 수준에서 지속적으로 변화한다. 이러한 동적 상호작용을 가능하게 하는 중요한 요소 중 하나는 인간 두뇌의 거시적 및 미시적 규모에서의 구조적 가소성이다. 뇌의 거시적 및 미세적 구조는 서로 다르지만 보완적인 정보를 전달하기 때문에 두 구조를 모두 고려하는 것은 뇌의 구조적 가소성과 인지 작업 중 연결성을 이해하는데 매우 중요하다. 하지만 기존의 연구는 이를 효과적으로 고려하지 못했다.

본 연구에선 인간 뇌의 거시적 및 미세적 구조를 통합하여 기존에 알지 못했던 인간 뇌의 의미와 새로운 표현형을 얻기 위해 구조 MRI (sMRI)에서 고품질 확산 텐서 이미징 (DTI) 및 트랙토그래피를 생성하도록 설계된 새로운 딥러닝 프레임워크인 Macro2Micro를 제시한다. 본 연구는 거시적 구조로부터 미시적 구조 정보를 유추할 수 있다는 가설을 전제로 하여, 초기에 한 가지 영상 기법만 획득하더라도 질병 진단 및 연구에 유익한 추가적인 영상 기법을 생성한다. 신경 영상 영역에서 전례가 없는 이 접근 방식은 멀티모달 이미지 번역의 이점을 활용하여 상당한 시간과 비용을 절감한다. Macro2Micro는 3D T1을 입력으로 사용하여 2D T1 슬라이스를 생성한 다음 적대적 생성망(GAN)을 통해 처리되어 2D DTI (FA) 슬라이스와 2D 트랙토그래피 이미지를 생성한다. 이 프로세스의 핵심 요소는 이미지 특성을 주파수 대역에 따라 분리하는 옥타브 합성곱을 사용하는 것이다. 이 프레임워크는 청소년 뇌 인지 발달(ABCD) 데이터를 사용하여 훈련되었으며, 이미지 픽셀 손실, 지각 손실, GAN 손실 및 뇌 중심 패치 GAN 손실을 통해 훈련 손실이 정의되었다. 본 연구의 결과는 정량적 및 정성적으로 뛰어난 성능을 보였으며 단순히 이미지 분포만을 학습한 것이 아닌 뇌의 구조적 및 미시적 구조의 생물학적인 특성까지 학습했다는 점에서

의의가 있다. 이미지 변환 모델을 데이터 증대 방법으로 잠재적으로 적용하면 데이터 불균형 및 희소성 문제를 해결할 수 있다. 이 연구는 발전하는 질병 모델링에서 다중 모드 이미징, 특히 T1, DTI 및 트랙토그래피의 조합 사용의 잠재력을 제시한다.

주요어: 이미지 변환, 딥러닝, 적대적 생성 네트워크, 자기공명영상, 뇌 미세구조, 뇌 거시구조

학번: 2021-22028