

**FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO**



# **AI-based Conditional Generation of Diffusion MR Images**

**Pedro Sousa**

**FOR JURY EVALUATION**

**Master in Bioengineering - Biomedical Engineering**

**1<sup>st</sup> Supervisor: Chantal Tax**

**2<sup>nd</sup> Supervisor: Jaime Cardoso**

**September 28, 2023**

© Pedro Sousa, 2023

# **AI-based Conditional Generation of Diffusion MR Images**

**Pedro Sousa**

Master in Bioengineering - Biomedical Engineering

September 28, 2023

# Resumo

A técnica imagiológica de ressonância magnética (RM) quantitativa é utilizada em todo o mundo com o intuito de medir, de forma reproduzível, a microestrutura e conectividade dos tecidos do cérebro. Porém, um único contraste de RM contém uma quantidade limitada de informação, daí a necessidade de combinar esta modalidade de imagiologia médica com outras. É com esta motivação que é introduzida a modalidade de ressonância magnética por difusão (RMd), ainda que com um senão: quanto maior a dimensionalidade do espaço de aquisição RM, maiores serão os tempos de aquisição, tornando-se proibitivos e fazendo com que os procedimentos de RMd sejam *ad hoc* e ineficientes, especialmente se usados num contexto clínico.

O projeto desta dissertação foca-se tanto em conseguir reduzir a duração e custo destes procedimentos, como em expandir e aumentar a quantidade de dados obtidos para esta modalidade imagiológica através da produção artificial e condicional de scans. Isto envolve obter um pequeno conjunto de scans com definições de aquisição preestabelecidas e usar inteligência artificial para gerar qualquer outra varredura de RMd, com quaisquer outras definições que sejam clinicamente relevantes no contexto de diagnóstico de condições como tumores cerebrais e doenças neurodegenerativas.

O fluxo de trabalho deste projeto divide esta geração condicional de imagens considerando dois grau de liberdade, quanto ao número de scans necessários para usar como contexto e que serão obtidos e alimentados aos modelos, e quanto ao número de scans que serão gerados usando esse mesmo input. Ao utilizar 3 abordagens diferentes ao problema, cada uma mais condicional que a última, são tomados passos para garantir que ambos estes números são diminuídos.

Os resultados obtidos mostram que, utilizando apenas 100 scans de RMd, os modelos são capazes de gerar scans para quaisquer outras definições de aquisição com grande eficácia, mesmo nunca tendo tido acesso às mesmas. Para além disso, das duas implementações abordadas, o desempenho é maior para a que trata as imagens voxel por voxel, e não como imagens 3D inteiras ou cortes das mesmas. Esta última implementação causa o aumento do desfoque das imagens geradas, bem como resultados menos favoráveis no que toca a geração da matéria branca e cinzenta. Porém, estes resultados são consideravelmente piores quando é pedido que o modelo faça conversões de 1 para 1, isto é, quando tenta gerar uma combinação de definições de aquisição através de uma outra qualquer. Esta análise leva à conclusão de que o problema requer uma solução que não passe pelo uso de aprendizagem máquina superficial. Assim, um meio-termo é alcançado, requerendo, como contexto, um número tão baixo quanto 100 scans, mas impossibilitando a ideia de utilizar os métodos propostos para gerar um scan a partir de qualquer outro.

Em suma, os resultados obtidos são um valioso contributo para a área da geração condicional de imagens médicas, principalmente tendo em conta a falta de estudos presentes na literatura que consigam aplicar todas as características deste projeto a dados de imagiologia médica.

**Palavras-Chave:** Difusão por Ressonância Magnética; Inteligência Artificial; Criação Condisional de Imagens; Aquisição de RM

# Abstract

Quantitative Magnetic Resonance Imaging (MRI) is used throughout the globe with the intent of extracting reproducible measures in regard to brain tissue microstructure and connectivity. However, a single MRI contrast is only able to provide so much information, hence the need for combinations with other modalities for further exploration. It is here that Diffusion MRI (dMRI) is introduced, albeit with a catch: the bigger the dimensionality of the MRI Acquisition Space, the more the scan acquisition times become prohibitive, thus making multi-contrast dMRI experiments ad hoc and time-inefficient, especially for clinical usage.

The project's focus is both on shortening dMRI procedures' length and costs, as well as further expanding and augmenting dataset of this scan modality by artificially generating dMRI scans in a conditional manner. This involves obtaining only a subset of scans with specific acquisition settings and using artificial intelligence tools to generate any other dMRI volumes with any other settings that clinicians might deem useful in the context of diagnosis of ailments such as brain tumours or neurodegenerative diseases.

The project's workflow divides this conditional image generation issue considering the degrees of freedom regarding the number of scans to obtain and use as context, and the number of volumes generated using that same input. By utilising 3 approaches, each more conditional than the last, steps are made to ensure that both numbers are reduced iteratively.

Obtained results show that, if given as few as 100 acquired dMRI scans to work with, models are able to generate scans for any other combination of acquisition settings present in the dataset with great accuracy and precision, even if these have not seen before by said algorithms. Moreover, this mapping's performance is greatly increased by the adoption of a voxel-wise approach to training, rather than having the computer vision designs analyse entire 3D scans or its slices. The latter has proved to increase the amount of blurring in fake scans, as well as less accuracy when generating the white and grey matter sections of these when compared to the target ground truth image. This performance does, however, take a turn for the worse when the pipeline asks for a 1-to-1 conversion, i.e., from a combination of source acquisition settings to a target one, leading to believe that this dMRI scan mapping approach is an issue not resolvable by conventional machine learning means. And so, a middle ground is reached, requiring procedures to acquire as little as 100 scans, but not creating the possibility of generating any acquisition setting scan from any other single scan, at least using the proposed methods.

All things considered, the results obtained can be recognised as a valuable contribution in the field of conditional medical image generations, especially taking into consideration that literature in this field is, in its great majority, a novelty which is both not applied to medical imaging data and lacking a great deal of the features included in this project.

**Keywords:** Diffusion MRI; Artificial Intelligence; Conditional Image Generation; MRI Acquisition

# Acknowledgements / Agradecimentos

Firstly, some words of appreciation towards the mentors behind this project. A huge thanks to prof. Chantal Tax, prof. Jaime Cardoso and Ishaan Bhat for all the availability shown and the guiding provided throughout this last year. It's clear to me that without their continued firm support in the makings of this project, none of this would have been possible.

Secondly, for receiving me with open arms in a new country and overall setting, I would extend this thank you note also to the people of Biltstraat, Kimmade and the ISI group, which, against all odds, have helped me overcome this past year and all the challenges it entailed.

Não poderia restar um enorme obrigado à minha família, pelo suporte, carinho e paciência para comigo. Poderei sempre dizer que, apesar de não estar presente em muitas instâncias, são e serão sempre uma âncora à qual me posso agarrar e uma luz que tento usar não só para seguir em frente, mas também para me lembrar de que, por vezes, o mais indicado pode ser voltar para trás. Destas raízes que criei e me foram criadas, tiro tudo o que sou, porém deixo um mero obrigado.

Por último, fica a restar o agradecimento às pessoas com quem partilhei estes 5 anos, os que foram, os que ficaram e os que levo comigo. Os que foram casa em mais sentidos do que um e os que poderão dizer o mesmo de mim. A vós espero deixar de tudo um pouco: companheirismo, amor, mágoa, saudade. De vós, "fica um tempo que não passa, neste passar de um tempo que não volta".

Pedro Sousa

# Academic Output

## Best Oral Communications Distinction:

Sousa, P., Tax, C. M. W., Cardoso, J., Bhat, I. (2023). AI-based Conditional Generation of Diffusion MR Images (21092). In: IJUP 2023 – 16th Meeting of Young Researchers of University of Porto. Topic – Engineering I. Session III/Room A4 (Wednesday 10th - 14:30-18:00), Rectory of the University of Porto, Porto (Portugal), 10th-12th May.



Fundação para a Ciéncia e a Tecnologia



REPÚBLICA  
PORTUGUESA



UNIÃO EUROPEIA  
Fundo Europeu  
de Desenvolvimento Regional

*“He just knows he needs to get out on the water.  
From there, Jesus will take the hamster wheel and guide him where he needs to go.”*

Charles White Jr.

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>                       | <b>1</b>  |
| 1.1      | Context and Motivation . . . . .          | 1         |
| 1.2      | Objectives . . . . .                      | 2         |
| 1.3      | Dissertation Structure . . . . .          | 3         |
| <b>2</b> | <b>Literature Review</b>                  | <b>4</b>  |
| 2.1      | Diffusion MRI Fundamentals . . . . .      | 4         |
| 2.1.1    | Functioning Principles . . . . .          | 4         |
| 2.1.2    | Image Acquisition . . . . .               | 6         |
| 2.2      | State-of-the-art . . . . .                | 8         |
| 2.2.1    | Conditional Image Generation . . . . .    | 9         |
| 2.2.2    | MRI Data Generation . . . . .             | 12        |
| 2.2.3    | Previous Related Works . . . . .          | 15        |
| 2.3      | Summary . . . . .                         | 16        |
| <b>3</b> | <b>Methodology</b>                        | <b>17</b> |
| 3.1      | Dataset Preparation . . . . .             | 17        |
| 3.1.1    | Dataset Acquisition . . . . .             | 17        |
| 3.1.2    | Scan Preprocessing . . . . .              | 19        |
| 3.1.3    | Parameter Handling . . . . .              | 19        |
| 3.2      | Task Partitioning . . . . .               | 21        |
| 3.2.1    | Baseline: Some-to-Some Approach . . . . . | 21        |
| 3.2.2    | Some-to-Any Approach . . . . .            | 22        |
| 3.2.3    | Any-to-Any Approach . . . . .             | 24        |
| 3.3      | Model Design . . . . .                    | 26        |
| 3.3.1    | Baseline: Some-to-Some Approach . . . . . | 26        |
| 3.3.2    | Some-to-Any Approach . . . . .            | 27        |
| 3.3.3    | Any-to-Any Approach . . . . .             | 29        |
| 3.4      | Evaluation Framework . . . . .            | 32        |
| <b>4</b> | <b>Results and Discussion</b>             | <b>35</b> |
| 4.1      | Baseline: Some-to-Some Results . . . . .  | 35        |
| 4.2      | Some-to-Any Results . . . . .             | 41        |
| 4.2.1    | Voxel-Wise Strategy . . . . .             | 41        |
| 4.2.2    | Slice-Wise Strategy . . . . .             | 46        |
| 4.3      | Any-to-Any Results . . . . .              | 51        |
| 4.4      | Overall Comparison . . . . .              | 55        |
| 4.4.1    | Test Parameter Subset Results . . . . .   | 57        |

|   |           |
|---|-----------|
| <b>5 Conclusions and Future Work</b>    | <b>59</b> |
| 5.1 Summary and Conclusions . . . . .   | 59        |
| 5.2 Future Work . . . . .               | 60        |
| <b>References</b>                       | <b>62</b> |
| <b>A Parameter Subset Distributions</b> | <b>65</b> |

# List of Figures

|     |  |    |
|-----|--|----|
| 2.1 | Examples of 3 different maps created by way of DTI-associated imaging methods.<br>A: DTI colormap; B: DTI data visualisation using ellipsoids; C: tractography reconstruction of neural connections created via DTI . . . . .  | 6  |
| 2.2 | A: Philips Achieva 3.0T Diffusion MRI Scanner; B: Example of a patient's positioning during a dMRI procedure inside the scanner's magnetic bore and laying on the scanner's table. . . . .   | 7  |
| 2.3 | Workflow of the CcGAN algorithm [1]. . . . .   | 9  |
| 2.4 | Examples of Conditional Continuous Generative Adversarial Network Generated Images for both the Chair Rotation Angle Dataset (A) and the UTKFace Dataset created with the purpose of allowing for image-based age regression and progression (B) [1]. . . . .  | 10 |
| 2.5 | OpenAI's DALLE.E 2 algorithm training pipeline [2]. . . . .  | 11 |
| 2.6 | Image and metric results from the MoTran article [3], in which their pipeline is compared to others in the field. Above, the ground truth (right) and reconstructions of a T2-weighted scan. Below, reconstruction performance for T1 and T2-weighted acquisitions at $R = \{4, 8\}$ (acceleration factor). . . . .  | 12 |
| 2.7 | Example results of Image-to-Image translation model mapping from 3T MRI scans to 7T ones using both simply the generator ( <i>FCN</i> ) and the full extent of the GAN methodology. The source 3T scan is included in the left, the example generated images in the middle and the ground truth 7T scan in the right. Below, two difference maps between the fake image and GT, for each of the two approaches [4]. . . . .  | 14 |
| 3.1 | Distribution of all 4 settings, across volume acquisition process. The third gradient orientation direction is represented by the bubble's radius. . . . .   | 18 |
| 3.2 | Visual Representation of the MUDI Dataset's Organisation: a full masked volume including all intensities for all the patient's image voxel positions (blue); all voxels pertaining to the same position, though for different volumes, meaning different parameter combinations (orange). In reality, for voxel-wise implementations, the actual organization of data is much better represented in the scheme to the right, seeing as masked volumes will extend into all the different voxel positions dictated by the patient's mask. . . . . | 19 |
| 3.3 | Baseline / Some-to-Some Approach's Input (blue) and Output (orange) Data Organisation within the space represented as the MUDI Dataset . . . . .   | 22 |
| 3.4 | Voxel-Wise Implementation of Some-to-Any Approach's Input (blue) and Output (orange) Data Organisation within the space represented as the MUDI Dataset . . . . .  | 23 |
| 3.5 | Slice-Wise Implementation of Some-to-Any Approach's Input (blue) and Output (orange) Data Organisation within the space represented as the MUDI Dataset . . . . .  | 24 |

|   |    |
|---|----|
| 3.6 Voxel-Wise Any-to-Any Approach Model's Input (blue) and Output (orange) Data Organisation within the space represented as the MUDI Dataset . . . . .  | 25 |
| 3.7 Baseline / Some-to-Some Approach's Model Architecture, including number of Neurons for all Layers depending on the Experiment. All the input layer's nodes will be reserved for measurements / voxel intensity values. . . . .  | 26 |
| 3.8 Voxel-Based Implementation of Some-to-Any Approach's Model Architecture, including number of Neurons for all Layers. Of the input layer's nodes a total of 5 are reserved for parameter values and the remaining are measurements / voxel intensity values. . . . .   | 27 |
| 3.9 Slice-Wise Some-to-Any Approach Model's Architecture [5]. . . . .   | 29 |
| 3.10 Voxel-Wise Any-to-Any Approach Model's Architecture, including number of Neurons . . . . .   | 30 |
| 3.11 Example Imagery for Structural Similarity Index Explanation: to the left are the original images and to the right are the distorted ones for both examples. . . . .  | 34 |
| <br>  |    |
| 4.1 Boxplot of MSE Loss, MAE Loss and SSIM Index Results for all Experiments of the Some-to-Some Approach, by segregation of the generated scans for Source and Target Parameter Combinations. The boxplot includes the 1 <sup>st</sup> and 3 <sup>rd</sup> quartile values and the minimum and maximum values (excluding outliers), as well as the median (line inside the box) and mean values (cross-shaped marker), and finally, the amount of samples for each plot. . . . .   | 36 |
| 4.2 On the left, a distribution of MSE Loss and SSIM Index results according to the scan's b Value for all 3 experiments performed for the some-to-some approach; on the right, 2 examples with a low b Value of 0 and a high b Value of 3000 both containing the respective ground truth (above), generated (below) scans and MSE, MAE and SSIM Index results. . . . .   | 38 |
| 4.3 On the left, a distribution of MSE Loss and SSIM Index results according to the scan's TI Value for all 3 experiments performed for the some-to-some approach; on the right, 2 examples with low TI values of 176.6 and 706.5, both containing the respective ground truth (above), generated (below) scans and MSE, MAE and SSIM Index results. . . . .  | 38 |
| 4.4 Example results for the same parameter combination scan present in the target subset of all 3 Some-to-Some experiments, including generated scan (above), MSE and MAE loss heatmaps (middle) and SSIM index masks (below), as well as marked points of discrepancy scan's and results in comparison to the experiment's median and mean values for all metrics. . . . .   | 39 |
| 4.5 Visual representations of mean loss values per voxel position according to both metric (horizontally) and experiment's number of source input parameter combinations (vertically). All heatmaps for the same metric were given the same scale.  | 42 |
| 4.6 On the left, area plots displaying mean evaluation metrics results versus volume slice index for all three experiments performed. The idea of using area plots and not simple line plots is aimed at better understanding of differences between experiments. To the right, two examples containing both the ground truth (above), the generated scan (below) and the heatmap for structural similarity and mean squared error respectively, as well as metric values and slice index. These examples are, in order, a low SSIM example and a high MSE one. . . . . | 43 |

- 4.7 Boxplot of MSE and MAE losses for all experiments of the Some-to-Any voxel-wise approach, by segregation of the different tissue segmentations (below) performed on the target generated scans for white matter, grey matter and cerebrospinal fluid. The boxplot includes the 1<sup>st</sup> and 3<sup>rd</sup> quartile values and the minimum and maximum values (excluding outliers), as well as the median (line inside the box) and mean values (cross-shaped marker), and finally, the amount of samples for each plot. . . . .

4.8 Example results for the same parameter combination scan present in the target subset of all 3 Some-to-Any Voxel-Wise experiments, including generated scan (above), MSE and MAE loss heatmaps (middle) and SSIM index masks (below), as well as marked points of discrepancy scan's and results in comparison to the experiment's median and mean values for all metrics. . . . .

4.9 On the left, area plots displaying mean evaluation metrics results versus volume slice index for all three experiments performed. The idea of using area plots and not simple line plots is aimed at better understanding of differences between experiments. To the right, three examples containing both the ground truth (above) and generated scan (below), as well as metric values and slice index. These examples are, in order, two low SSIM examples and a high MSE one. . . . .

4.10 Three examples of target parameter scans, including original and generated scans, each marked with the model experiment they belong according to the pre-established color scheme. Each example's metrics results and slice index are also included as labels. . . . .

4.11 Boxplot of MSE and MAE losses for all experiments of the Some-to-Any slice-wise approach, by segregation of the different tissue segmentations (below) performed on the target generated scans for white matter, grey matter and cerebrospinal fluid. The boxplot includes the 1<sup>st</sup> and 3<sup>rd</sup> quartile values and the minimum and maximum values (excluding outliers), as well as the median (line inside the box) and mean values (cross-shaped marker), and finally, the amount of samples for each plot. . . . .

4.12 Example results for three chosen slices of the same parameter combination scan present in the target subset of all 3 Some-to-Any slice-wise experiments, including ground truth (left) and generated scan for each experiment, as well as mean results in comparison to the experiment's median and mean values for all metrics. . . . .

4.13 Boxplot of MSE loss and SSIM index results for both experiments of the Any-to-Any Approach. The boxplot includes the 1<sup>st</sup> and 3<sup>rd</sup> quartile values and the minimum and maximum values (excluding outliers), as well as the median (line inside the box) and mean values (cross-shaped marker). . . . .

4.14 Examples for 2 source-target combinations in which the only mutation between the 2 acquisition settings amalgamations is the TE value, including the source scan, the target scan and the generated image, as well as the values for the evaluation metrics for this generation. Below, a table of all other parameters which suffer no change between the source index and the target one, for both exemplary sets of imagery. . . . .

|   |    |
|---|----|
| 4.15 Examples for 2 source-target combinations in which the only mutation between the 2 acquisition settings amalgamations are the gradient orientation values, including the source scan, the target scan and the generated image, as well as the values for the evaluation metrics for this generation. Below, a table of all other parameters which suffer no change between the source index and the target one, for both exemplary sets of imagery. . . . .      | 53 |
| 4.16 Mean validation set results for both metrics used as back-propagation and evaluation along the number of training epochs for both experiments ran using this approach. . . . .   | 54 |
| 4.17 Boxplot of MSE loss, MAE loss and SSIM index Results for all experiments of the Some-to-Some and Some-to-Any Approaches, regarding the target parameters being mapped to. The boxplot includes the 1 <sup>st</sup> and 3 <sup>rd</sup> quartile values and the minimum and maximum values (excluding outliers), as well as the median (line inside the box) and mean values (cross-shaped marker), and finally, the amount of samples for each plot. . . . .     | 55 |
| 4.18 Boxplot of MSE loss and SSIM index results for all experiments of the Some-to-Any Approach, by segregation of the generated scans for target and test Parameter Combinations. The boxplot includes the 1 <sup>st</sup> and 3 <sup>rd</sup> quartile values and the minimum and maximum values (excluding outliers), as well as the median (line inside the box) and mean values (cross-shaped marker), and finally, the amount of samples for each plot. . . . . | 58 |
| A.1 Parameter subset distributions for all acquisition settings. The grey line is centered around the dataset's percentage of scans with specific parameter values and helps understand how the subset's distribution is fairly performed. Gradient orientations have been displayed using unique index values, of which there are 106 in total. . . . .  | 66 |

# List of Tables

|     |  |    |
|-----|--|----|
| 3.1 | MUDI dataset's variable & static acquisition settings labels, including their value / interval and distribution . . . . .                            | 18 |
| 3.2 | Parameter Manual Normalisation Methods . . . . .   | 20 |
| 3.3 | Tasks and approaches included in the methodology chapter, including experiments created for the models and their specificities . . . . .             | 31 |
| 3.4 | Tasks and approaches included in the methodology chapter, including experiments created for the models and their number of training epochs . . . . . | 32 |

# Abbreviations & Symbols

|        |  |
|--------|--|
| AI     | Artificial Intelligence  |
| API    | Application Programming Interface                                  |
| CAD    | Computer-Aided Diagnostic  |
| CNN    | Convolutional Neural Network                                       |
| CSF    | Cerebrospinal Fluid  |
| CVAE   | Convolutional Variational AutoEncoder                              |
| DL     | Deep Learning  |
| dMRI   | Diffusion Magnetic Resonance Imaging                               |
| DTI    | Diffusion Tensor Imaging   |
| DWI    | Diffusion Weighted Imaging   |
| FCN    | Fully Convolutional Network  |
| FMRIB  | Functional MRI of the Brain  |
| FSL    | FMRIB's Software Library   |
| GAN    | Generative Adversarial Network                                     |
| GPP    | Gaussian Process Prior   |
| GT     | Ground Truth   |
| IoT    | Internet of Things   |
| IQR    | InterQuartile Range  |
| MAE    | Mean Absolute Error  |
| MICCAI | Medical Image Computing and Computer Assisted Intervention Society |
| ML     | Machine Learning   |
| MRI    | Magnetic Resonance Imaging   |
| MSE    | Mean Squared Error   |
| MUDI   | MultiDimensional Diffusion   |
| MUSHAC | Multi-Shell Diffusion MRI Harmonisation and Enhancement Challenge  |
| NN     | Neural Network   |
| ReLU   | Rectified Linear Unit  |
| SNR    | Signal-to-Noise Ration   |
| SOLID  | Slice-wise OutLIer Detection                                       |
| SSIM   | Structural Similarity Index  |
| TE     | Echo Time  |
| TI     | Inversion Time   |
| VAE    | Variational AutoEncoder  |
| WWW    | <i>World Wide Web</i>  |

# Chapter 1

## Introduction

### 1.1 Context and Motivation

Clinical magnetic resonance imaging (MRI) methods typically display qualitative tissue-contrast with intensities arbitrarily scaled in function of the different magnetic resonance phenomena being captured. In opposition, quantitative MRI aims to extract reproducible measures directly related to tissue properties. Diffusion MRI (dMRI), on the other hand, is one of the newer MRI methodologies that has been adopted as the common tool for quantification of tissue microstructure. Seeing as it functions by monitoring the random molecular motion of water molecules inside the body, it is the ideal imaging tool to probe structures, including tumours, in great detail. To this end, in a clinical setting, a number of diffusion sensitisation strengths and directions are usually mixed and matched to obtain a plurality of different scans, each of which will target specific structures. Furthermore, the recently discovered abilities of clinically used dMRI have driven its potential to greater lengths, consequently increasing the amount of research studies on this matter. These include deriving quantitative features such as fractional anisotropy from such scans using diffusion tensor imaging techniques, and virtually reconstructing pathways with tractography .

However, and despite dMRI's exquisite sensitivity to disease and to microstructural characteristics, this modality is often paired with advanced encoding and other non-dMRI modalities to provide complementary information, such as T1/T2 relaxation and susceptibility. This is due to the fact that conventional dMRI alone cannot handle more complex issues, such as the specificities of some microstructural characteristics and the differentiation of tissue compartments. These are, on the other hand, areas in which other biophysical models excel in. Even the study of white matter, which has been extensively researched into in the past, proves to be quite challenging using the current dMRI-only models and usually requires the inclusion of further tissue information. The solution found for this was the joining of these models' efforts with those of other techniques, consequently allowing for the obtainal of more detailed and unbiased investigation into tissue constituents in the areas of health and disease, by use of a technique which was dubbed multi-contrast diffusion MR imaging. The information contained in these is, ultimately, of great relevance, although also of difficult usage in a clinical setting.

This approach has since become extremely valuable in the areas of oncology and neuroimaging. It enhances the information obtained using typical dMRI by fusing it with the acquisition of multiple contrasts or, in other words, multiple sets of diffusion-weighted images with different gradient sensitivities and directions. However, as the size and dimensionality of the accessible MRI space increase, the acquisition times for multi-contrast dMRI become prohibitive. So much so that experimental designs are, more often than not, *ad hoc* and severely time-inefficient. The current multi-contrast dMRI protocol used in the Netherlands lasts a total of 2.5 hours, which is both clinically impracticable and an obstacle in the way of patient well-being. Moreover, such a lengthy protocol will also require that hospitals and privatised healthcare units incur additional costs to be able to provide such a service. This results in a decrease in the usage of what can be considered an exceptionally-rich imaging technique, especially in countries in which socialised healthcare is not the *status quo*.

In light of all these hindrances, this project will aim to substantially reduce the amount of time necessary for multi-contrast dMRI volume acquisition, by coming up with a tool that can allow clinicians to obtain only a select few scans and have any other necessary ones be generated synthetically using machine learning and computer vision methods. In other words, it will allow the procedure to acquire only a portion of the 3D scans it currently does, have the tool analyse that obtained portion and create any others, with any other acquisition settings, that a trained clinician might find useful in the way of ailment diagnosis.

In turn, this will make the referred imaging technique more broadly available to the public as a diagnosis tool, by decreasing the scan acquisition times and allowing medical institutions to both cut costs and provide a greater quality service. The resolution of acquisition times would also not require patients to lay inside a very noisy and claustrophobia-inducing machine for either one really large session or several sessions across multiple days. Furthermore, by creating a tool for synthetic multi-contrast dMRI volumes, there is also the possibility to further expand this area's research field by creating a tool for dataset augmentation, in the sense that the tool's output deals in realistic and accurate dMRI scans.

## 1.2 Objectives

This dissertation focuses on the development of tools for the generation of synthetic multi-contrast diffusion MRI scans by manipulation of its acquisition settings. It follows a workflow in which a dataset of dMRI volumes are acquired using a scanner and recorded together with the settings used. A portion of these is then used as input for a tool to be able to generate a number of other scans with settings other than the ones chosen for input. With this pipeline in mind, various objectives were defined:

- Development of an automatic, 'plug-and-play'-like tool for the generation of synthetic patient-accurate, as well as precise and realistic diffusion MRI 3D volumes;

- Reduction of the number of scans required to function as input for the tool to acquire information about the patient dMRI dataset's spatial and structural organization;
- Increase in the freedom of choice regarding the scanner acquisition settings of the scan being built and the amount of these the tool is able to correctly create images for.

Given the premise and motivation for the project, the ideal situation would include using the least possible amount of scans as a reference to generate any single, highly-specific parameter combination. Expressed differently, the ideal solution would imply that the created tool would work much like a MRI scanner, meaning that it would be fed all the acquisition settings of a scan the user wishes to be acquired. This, however, is not without the added requirement of having to previously obtain a small set of 3D dMRI volumes with fixed settings with which the tool can get an understanding of the patient's brain anatomy. By utilising this data-driven pipeline, the amount of time the patient is required to spend inside the scanner is greatly diminished.

### 1.3 Dissertation Structure

Beyond the Introduction, this dissertation is composed of 4 more chapters. Chapter 2 provides an overview on the background necessary for understanding the fundamentals behind diffusion MRI and how procedures for this imaging technique are like, as well as an explanation of the state-of-the-art studies on conditional image generation and MRI-based generative issues. Afterwards, Chapter 3 presents the framework implemented to address the objectives set for development of a tool with the uttermost degree of freedom as it pertains to the generation of accurate, precise and realistic dMRI scans, including detailed exposition of the steps behind the building, training and evaluation of such pipelines. The obtained results of the implemented framework are presented on Chapter 4. The outcomes of the objectives set are exposed in a variety of different ways, followed by a discussion of those same results. Finally, the main conclusions drawn from this dissertation are stated in Chapter 5, as well as a number of suggestions on how the project's implemented algorithms could be improved upon in future follow-up project ideas.

# Chapter 2

## Literature Review

In this chapter, an understanding of the basics behind both the data acquisition techniques and the state-of-the-art methods that inspired the realisation of this project are presented. Section 2.1 provides a succinct introduction of the phenomena behind the diffusion MRI and diffusion tensor imaging techniques behind the dataset used in the project. Section 2.2 expands the knowledge on current state-of-the-art published works focused around generative approaches, some of which focused on dMRI data and some which introduce conditionality as an obstacle. Afterwards, section 2.2.3 details previous works leading up to this project and whose outputs were greatly appreciated and used throughout the whole project. Finally, section 2.3 summarises this chapter's contents.

### 2.1 Diffusion MRI Fundamentals

Currently, there is a vast number of imaging techniques which are ingrained into common health practices performed on a daily basis all across the globe, primarily for diagnosis reasons. This widespread use can be vouched for by the fact that they allow one to visualise sections of the human body which would otherwise require extremely invasive procedures. Of these, dMRI is perhaps one of the modalities that has had the most success in recent time, due to both the richness of the information it provides as well as the scale which it targets the most and which coincides with that of tumours.

#### 2.1.1 Functioning Principles

**Diffusion Magnetic Resonance Imaging** [6] does share a great deal of similarity with regular, structural MRI, but it is fundamentally different when it pertains to the physical laws of motion it abides by. For convenience purposes, and despite these methods' versatility, this section will only refer to brain MRI scans, which is the one modality the project mostly focuses on.

Summarised, this technique utilises the random thermal motion of water molecules inside biological tissues to model the appearance of structures around them. This type of movement is commonly known as **Brownian motion** [7] and it is influenced by the local microenvironment, i.e.

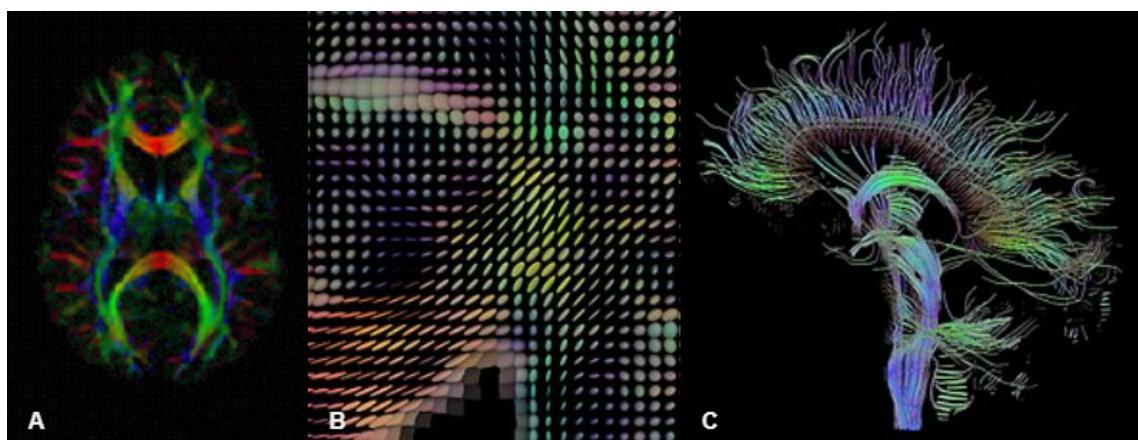
axons, tissue segregations and other barriers which might restrict the water molecules' movement. This molecular diffusion is then quantified by **Diffusion Tensor Imaging (DTI)** [8], as being an ellipsoid. This representation, in the form of a vector, can be further decomposed into eigenvalues and eigenvectors as a way to obtain the primary diffusion directions with which to characterise the water molecules' movement.

However, and since this would not be enough information to build a conclusive model with which to understand tissue microstructure and connectivity inside the brain, the dMRI technique expands upon this in two different ways. Firstly, it derives a measure called **fractional anisotropy** [9] from DTI used to quantise the anisotropy / directionality of water diffusion within a single voxel, and which is commonly valuable in analysis of tissue integrity and organization. Lastly, information on the diffusion process is also encoded by using a series of sensitising gradients [10]. This encoding process results in the attenuation of the resonance signal, which can then be used to infer certain tissue properties.

While both dMRI and standard MRI techniques use magnetic fields and radio waves to generate scans, they differ on the means of capturing such information and on the information itself. These differences are, one by one, specified below:

- Conventional MRI techniques mostly provide information on anatomical structure and tissue density discrepancies within the brain, making it perfect for visualisation of soft tissues, bones and abnormalities, such as tumours. dMRI, on the other hand, by capturing the diffusion direction of water molecules, focuses on tissue microstructure and connectivity as a whole.
- The images generated by way of typical MRI are based on T1 and T2-weighted contrasts, thus highlighting, for example, proton density and relaxation times. Conversely, dMRI images include a variety of different measurements, such as fractional anisotropy, mean diffusivity and directional color maps, which provide more insight into microstructures and the orientation of fibers, rather than anatomical structures alone.
- In contrast to standard MRI's high resolution and excellent tissue contrast for all types of different anatomical structure, dMRI's interest is broader, in the sense that it looks for the presence of cell membranes, axonal fibers or any other barriers that might hinder the diffusion of water through it or around it.

Having established the functioning of dMRI as an imaging technique, it should be clear that it poses certain benefits when compared to others, but also that it is subject to a fair share of limitations as well. One such benefit is the fact that it allows for in-vivo probing of living biological tissue to generate a complex and rich map of the brain's features by capturing how the radio waves' signal decays across space. Furthermore, because the scale that water molecules move on coincides with that of most tumours, according to Brownian movement theory, this tool becomes increasingly valuable as a method for diagnosis.



**Figure 2.1:** Examples of 3 different maps created by way of DTI-associated imaging methods. A: DTI colormap; B: DTI data visualisation using ellipsoids; C: tractography reconstruction of neural connections created via DTI

However, this imaging application is bound to face some challenges, one of which being the ongoing research to decrease the inherently low Signal-to-Noise Ratio (*SNR*) [11]. Moreover, because fiber orientation is, in some areas of the brain, very complex and presents a lot of crossings, it is something DTI will struggle with [12] [13]. Nevertheless, this is an issue which is addressed by more advanced diffusion models, such as High Angular Resolution Diffusion Imaging (*HARDI*) [14] [15]. Finally, and as it has been mentioned before, it is a complex technique, and one which produces data which is both complex and of computationally intensive analysis.

Therefore, this data can be analysed in a number of different ways and for a number of different procedures or techniques. It is commonly combined with functional MRI (*fMRI*) to obtain some more insight into the functional connectivity inside the brain, or in other words, to understand how the different brain regions communicate and cooperate with one another. However, perhaps its most common use dwells in **tractography**, which is essential for neurosurgical planning, seeing as it reconstructs the white matter pathways in the brain, allowing for the mapping of neural connections. Both applications further contribute to pathological diagnosis of neurological conditions (e.g. multiple sclerosis, strokes and brain tumours) and have had a strong impact in development of these ailments' early diagnosis, since dMRI can detect changes in the tissue microstructure before any macroscopic structural changes in the brain actually occur.

### 2.1.2 Image Acquisition

To comprehend the functioning principles behind dMRI image acquisition, it is first required to understand the basics behind magnetic resonance. MR will rely on detection of the signals emitted by the hydrogen atoms' nuclei when exposed to strong magnetic field and radiofrequency pulses [16]. Although the dMRI technique applies this same function as a part of its scanner's features, it also involves using diffusion-weighted imaging (*DWI*), i.e. the application of a gradient magnetic field so as to allow for the measurement of signal attenuation resulting from Brownian motion.

As evident by the complexity of all the features ingrained into acquiring 3D dMRI volumes, the hardware behind such scanner is a product of highly-specialised medical imaging equipment, which, despite sharing common traits with conventional MRI scanners, does require specific components and features tailored for DWI. These include, but are not limited to, high-field 3T and even 7T magnets used to enhance signal-to-noise ratio (SNR) and image quality, gradient coils for the enabling of diffusion measurements in three orthogonal directions and radiofrequency coils which may or may not be specialised if the intent is to optimise sensitivity for brain regions of interest. Of all these, it is the gradient coils that generate immense amounts of noise, this being one external factor usually associated to the discomfort felt inside a scanner.



**Figure 2.2:** A: Philips Achieva 3.0T Diffusion MRI Scanner; B: Example of a patient's positioning during a dMRI procedure inside the scanner's magnetic bore and laying on the scanner's table.

The procedure behind dMRI acquisition is also complex in and of itself, being comprised of various different steps. Firstly, the patient is positioned on the scanner table so that the region of interest (e.g., the brain) can be within the magnetic area of influence. After the gradient magnetic field is established in various directions at a time, MRI pulse sequences are selected, each of which will create a controlled diffusion-sensitising gradient to manipulate the movement of water molecules inside the tissue. As the radiofrequency coils excite the water molecules' nuclei and they emit energy when returning to the equilibrium state, these signals are captured and analysed to obtain information on tissue microstructure and diffusion directions [13]. The described process is repeated a number of times with varying gradient strengths and directions so that a computer system can then reconstruct diffusion-weighted images and fractional anisotropy maps. While these steps do make up most of the process, often a post-processing method is also applied to derive certain quantitative parameters, such as the apparent diffusion coefficient and other DTI metrics. It is also because of the amount of steps that need to be repeated that this procedure of image acquisition is usually deemed both lengthy and uncomfortable, requiring the patient to lay inside the rigid and often small scanner table for long periods of time.

## 2.2 State-of-the-art

Although generative tasks are not the most common problem usually tackled by way of machine learning (*ML*) and artificial intelligence (*AI*) methodologies, they are still very well documented in literature and make a substantial branch of the innovation in the computer vision technologies developed in recent years. However, the problem provided in the introductory chapter is one which needs to leverage various different requirements, some of which not common in current research. Given the motivation and the objectives of the tool to be designed by the project, a few different features can be identified:

- The ability to withstand and comprehend the intrinsic structures captured within 3D dMRI scans and how to correctly construct and mimic them. This can either be done by breaking down the 3D scan to simpler elements or by utilising them as a whole.
- The conditionality, or the means of generating scans with specific acquisition settings, rather than generic, non-specific ones or reconstructions of the original source images. This will imply understanding the differences between the scans provided and those which the user wishes to generate, and how such changes impact the look of the dMRI volumes and its structures.
- The support for multiple continuous image labels which will be defined as the scanner acquisition settings included in the dataset. These parameters are the sole conditional aspect of the scan generation process, meaning that it is imperative that the designed tool has an understanding of both each setting's relevance and effect on contrast, luminance, etc., as well as how their continuous variations impact the scans themselves.

Much like it has been stated before, the problem at hand is a generative one, i.e., one in which the algorithms are requested to create data instances, in this case to specificities of the user. In that sense, while the most common machine learning models, typically discriminative in nature, will try to predict to which of the seen class types the instance belongs to, conditional generative models work in inverse fashion, being requested to, given the target class, predict what the data will look like based on some manner of input. Mathematically, this will mean that predictive models will try to capture the conditional probability  $P(Y|X)$  that data instance  $X$  belongs to class  $Y$ . Generative-like approaches, on the other hand, model a joint probability to the likes of  $P(X, Y)$ , having to understand the distribution of the entire dataset to capture how likely the value of an instance  $X$  is if the condition set for it is  $Y$ .

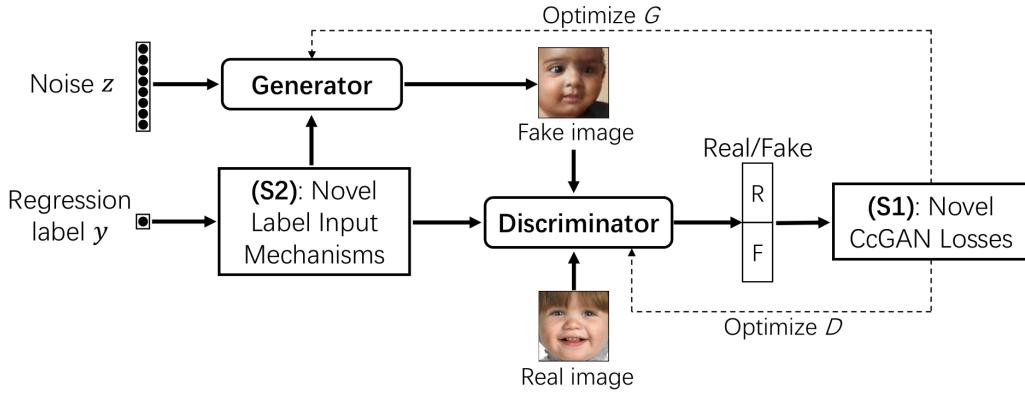
Having said this, this section will overview the main works in the field stated, dividing those which apply all the features discussed from those which engage in generative-like approaches using MRI data.

### 2.2.1 Conditional Image Generation

#### Conditional Continuous Generative Adversarial Network

While a number of studies in generative issues exist, most focus on discrete, categorical conditions, or class labels. The first not to has been proposed by Ding *et al.* in [1] (2023) and it seems to bridge the gaps created by the continuous nature of its dataset labels. Therefore, one could consider the proposed Conditional Continuous Generative Adversarial Network (*CcGAN*) the first and still one of the few algorithms tasked with a problem similar to the one this project faces, i.e., a generative image generation problem in which conditions are established by multiple continuous-valued parameters.

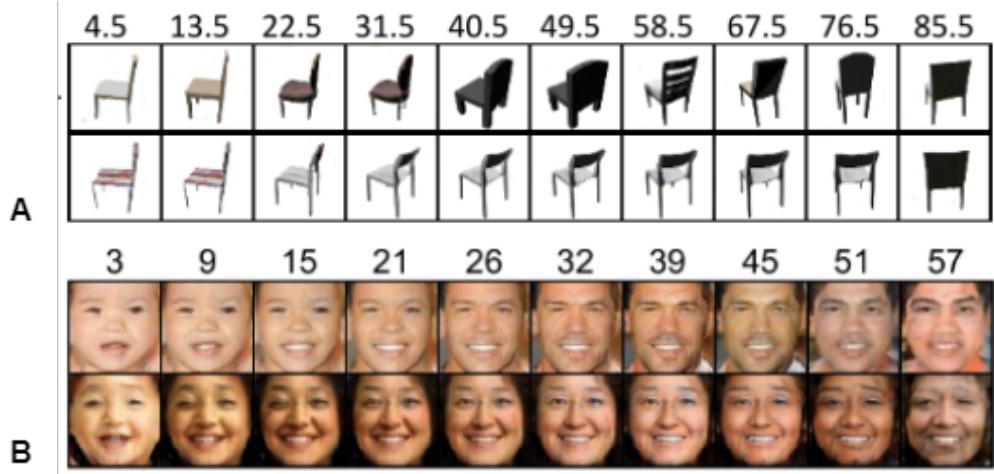
The proposed model's design was adapted from the original Generative Adversarial Network [17]. This architecture has since become a staple of image-based generative problems because of the advantages of the adversarial training concept. Explained simply, it is a training method in which two models, the generator and the discriminator, are trained against one another. The first, the generator, will be tasked with creating synthetic images from random, usually gaussian, noise vectors. The latter, the discriminator, will pit these generated images against real data in order to score them on how likely the image provided is to be real. This training method is then alternated in between the generator and the discriminator until the generator becomes able to generate realistic images that can fool the discriminator's judgement. With some adjustments on the input side, it is possible to turn such a pipeline into that of the *CcGAN* model, as seen in figure 2.3.



**Figure 2.3:** Workflow of the *CcGAN* algorithm [1].

In light of this problem requiring data-driven approaches, it is important to state that the proposed model was trained with two different datasets and, therefore, two different condition categories. The first included 49 3D chair object models which were rendered in order to create different images according to yaw axis rotation angle. Therefore, the model was tasked to create an image of the same chair according to a rotation angle condition. Secondly, the UTKFace dataset was used. This dataset includes over 55000 facial images of about 13000 subjects with ages ranging from 16 to 77 years old and was created and used by Zhang *et al.* in [18] (2017) to train a similar Conditional Adversarial Autoencoder on how to execute age progression and regression, thus making it one of the precursors of the *CcGAN* study. Therefore, the model was trained and

tested on the UTKFace dataset on how well it could model the subject's age to generate accurate and realistic-looking synthetic facial image data according to the target age and the person being provided as input. Results were included in figure 2.4 as reference material.



**Figure 2.4:** Examples of Conditional Continuous Generative Adversarial Network Generated Images for both the Chair Rotation Angle Dataset (A) and the UTKFace Dataset created with the purpose of allowing for image-based age regression and progression [1].

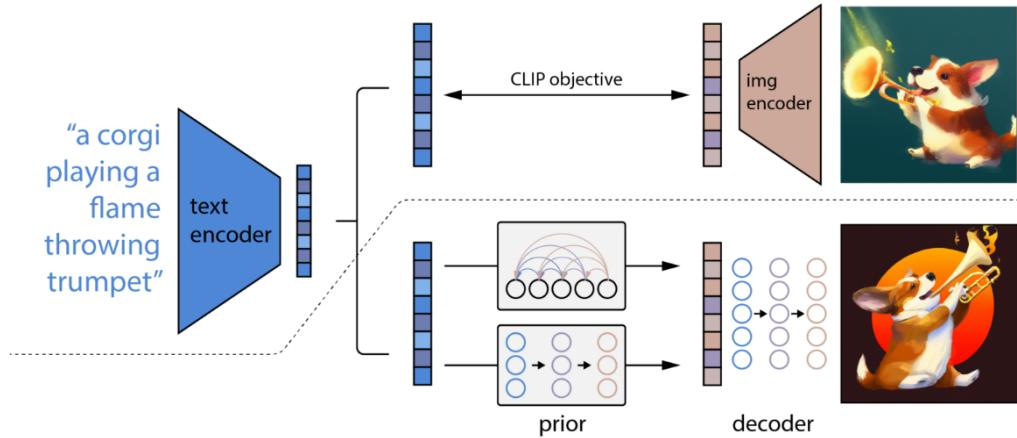
Despite dealing with only a single condition at a time, the study's proposed model proves to be an important precursor to the attempted task, seeing as it implements most of the features designated by the objectives. However, and perhaps because of the nature of data usually used, most conditional generation algorithms such as this study's CcGAN create concerns on whether or not they would be fit to deal with medical data, which is both more complex and more sensitive to error. And so, while certain facial features might be wrong or deformed in the fake images displayed, the same distortions could prove fatal in a clinical context.

### Text-to-Image Synthesis

In recent years, a number of different generative models have been thoroughly researched into in the area of conditional image generations. A vast majority of it, such as the DALL.E model developed by OpenAI [19] (2021), has been dedicated to the field of AI-generated art. These excel in combining natural language processing with computer vision in order to generate images using input textual descriptions. In other words, it is an approach which requires users to describe what the image they want to have generated includes and looks like, so the model can base its results on a semantic understanding of the text prompt. Commonly, such models are subject to large-scale pretraining datasets in an internet-of-things (*IoT*) environment and also a very well-calibrated fine-tuning of the task and images it is allowed to have access to, in order to generate accurate results and understand prompts clearly.

Usually described as a large language model, the second version of the DALL.E (figure 2.5) algorithm combines an encoder-decoder training pipeline to be able to deconstruct the text prompt it is given with a diffusion model, which will decode this embedding into the final image. This

embedding is done using OpenAI’s custom language model named Contrastive Language-Image Pretraining (*CLIP*) [20] (2022), whose task is to use a pre-trained neural network to encode both text prompts and images into a high dimensional vector that the decoder can then use to understand the concepts requested and the relationships between them. Before that, however, the CLIP embedding is first passed through a prior model which will ensure that the generated image is of high quality and that it matches the prompt provided.



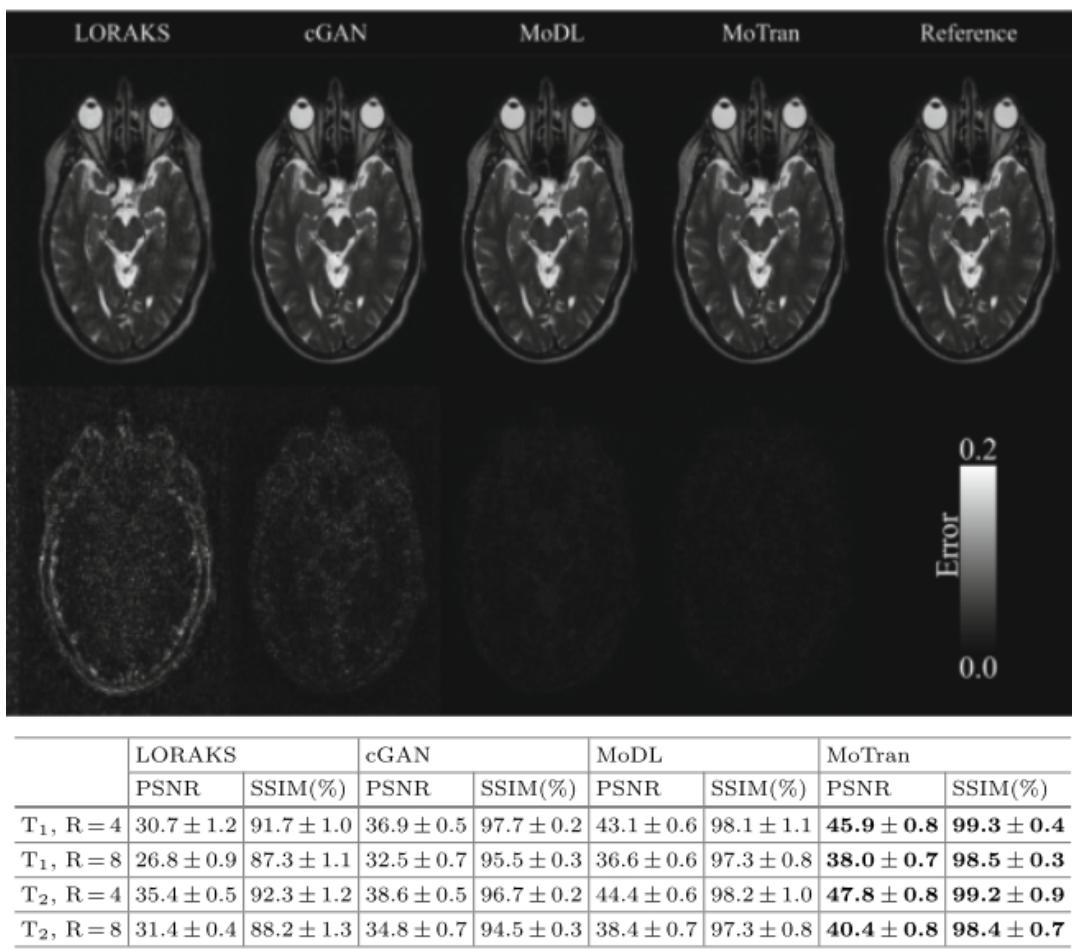
**Figure 2.5:** OpenAI’s DALL-E 2 algorithm training pipeline [2].

Although these approaches to image generation can have a variety of different practical applications, they mostly apply to the field of creative content generation and concept art creation, and have not yet delved into the medical domain. And so, despite this being a possibility in the future, this genre of conditional image generation is, for now, far removed from what the project aims to achieve. For example, the DALL-E algorithm’s methods of dealing with the user input are quite dissimilar since the input itself is nothing alike. Because models like the DALL-E deal with natural language processing they must reach a semantic understanding of the input, both when it comes to the concepts being requested and also how these relate to one another, so as to deliver on the text prompt received. The project’s synthetic generation, on the other hand, deals with acquisition setting values, meaning that it must understand how different parameters and their variations affect the look of the dMRI scan.

### 2.2.2 MRI Data Generation

#### Scan Reconstruction

The field of generative AI is still only making its way into the domain of medical imaging, so it is expectable for most approaches on the market to be primarily corrective, and not predictive. In other words, most works in the field of generative issues applied to medical imaging data are usually either on reconstruction, distortion correction or segmentation. Nonetheless, they rarely ever delve into truly generative problems, such as conditional image generation and even image inpainting, which refers to studies in which AI is used to fill in or reconstruct sections of a damaged image. MRI reconstruction, for example, is a field in and of its own, seeing as it is necessary for the visualisation of high-quality MRI data and, by default, incorporated in the pipeline of acquisition as post-processing step. Most procedures include reconstruction methods which are not as extensive and customisable as AI-based ones, but the latter have began making its way into the scene of this post-processing step.



**Figure 2.6:** Image and metric results from the MoTran article [3], in which their pipeline is compared to others in the field. Above, the ground truth (right) and reconstructions of a T2-weighted scan. Below, reconstruction performance for T1 and T2-weighted acquisitions at R = {4, 8} (acceleration factor).

One such example by Korkmaz *et al.* [3] (2022) describes the use of a number of different methods, namely transformer models, for high-performance MRI data reconstruction. Their proposed transformer implementation (*MoTran*) aims to correct for the lack of generalisation and sensitivity to long-range context that other state-of-the-art CNN-based architecture in this field suffer from. This reconstruction is done on both T1 and T2-weighted acquisitions and results display both peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), which is a metric relevant to this particular project and one which allows a comparison term.

As shown in figure 2.6, it is evident that reconstruction pipelines are well established and produce good results overall, although all the shown methods are supplanted by the proposed *MoTran*. This is further corroborated by the high PSNR and SSIM values, which assert the quality, accuracy and precision of the generated scans.

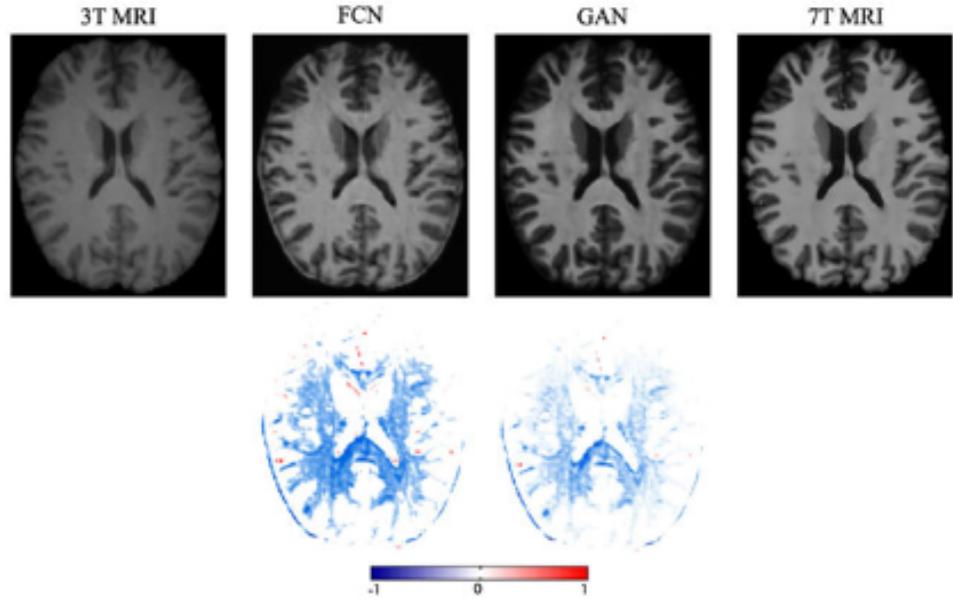
### **Image-to-Image Translation**

One other problem for which some AI and Computer Vision-based studies try to find solutions for is Image-to-Image synthesis. The solution itself encompasses a wide range of tasks and applications for which there is a need to model an input image from one domain to another, while preserving its essential and intrinsic content. This will include the transformation of an image from one style, modality, appearance, etc. to another, which may or may not be conditional in implementation. Among others, this is a computer vision modality greatly appreciated for issues such as image colourisation, resolution upscaling and even, to a certain extent, semantic segmentation.

This is the case for the pipeline developed by Nie *et al.* in [4] (2018), whose objective is to incur in medical image synthesis given a source and a target imaging modality. In particular, the models suggested are evaluated on their ability to generate CT scans from MRI data and 7T MRI scans from 3T ones. The proposed methodology includes a Generative Adversarial Network (*GAN*) architecture composed of a fully convolutional network (*FCN*) generator and a Convolutional Neural Network (*CNN*) posing as a discriminator, as well as adversarial training, residual connections inside the generator and an image gradient difference loss function. As a consequence of this architecture, its results are quite satisfactory in performing the task, as can be seen in figure 2.7, which includes an example of the translation done from 3T MRI images to 7T ones.

Having described the intricacies behind this computer vision task, it should be evident that it is similar to what the project intends to achieve, seeing as both intend to capture information which is intrinsic to the image but present it as seen from a different light. And so, while the study presented aims to freely transform images from one imaging modality to another, the project at hand strives to, within the same imaging modality of multi-contrast dMRI, create 3D volumes with specific scanner acquisition settings from a set of other volumes for which these parameters are known. This dissertation's and the referred article's goals can, therefore, be considered similar, since the mapping described in the motivation follows a common train of thought, although with some major differences, namely on the processing of the labels. The study in this subsection uses what can be considered class / categorical labels indicating what imaging modality to map the input to,

whereas our issue requests that labels include a number of continuous scanner acquisition settings instead.



**Figure 2.7:** Example results of Image-to-Image translation model mapping from 3T MRI scans to 7T ones using both simply the generator (*FCN*) and the full extent of the GAN methodology. The source 3T scan is included in the left, the example generated images in the middle and the ground truth 7T scan in the right. Below, two difference maps between the fake image and GT, for each of the two approaches [4].

### 2.2.3 Previous Related Works

One aspect of literature that this project does take input on is the data itself. As described in [21][22], previous studies have been performed using dMRI and acquisition setting labels with the intent of optimising the acquisition procedure's duration. Both endeavours describe the use of concrete autoencoders to select different-sized optimal subsets of dMRI measurements from a vast number of these settings' combinations. These were composed of: an encoder defined by a concrete selection layer, tasked to select the most relevant subset; and a decoder, which was, for the first experiment a simple neural network (*NN*) and for the second a physics-informed *NN*.

This approach yields outputs which can be applied to the project at hand, seeing as the subsets of combinations are such that they specifically choose the most representative dMRI scans and eliminate any redundancies, thus optimising the acquisition procedure. Therefore, there is a chance that, if the methodologies implemented in this dissertation use the scans designated in these subsets as source material to be able to map to other targets, their performance or at least their training curve is optimised as well. And so, while the presented research decreases the time necessary for dMRI acquisition, it does not have the capability to generate scans with target settings which clinicians might deem useful.

It is, however, capable of generating a fake dataset containing all scans for all different acquisition settings using the two decoders within in the concrete autoencoder architecture. For what is described in the methodology, this is done in a voxel-wise manner, by feeding voxel intensity sequences for the same voxel position in different scans, rather than whole images. While unorthodox, it is a sound implementation, which could prove useful in a conditional setting much like the one stated in this project's motivation.

Finally, it would be important to state the existence of studies which also aim to increase the efficiency of dMRI procedures and, thus decrease the amount of acquisition time, while maximising the amount of information contained in its features [23]. However, this is done by leveraging *a priori* assumed models and representations with new parameter values for settings such as gradient strengths which can be optimised to obtain greater contrast of certain elements in the brain. As such, the input these studies provide are of limited use, seeing as the project wishes to work within a range of these acquisition settings and should not be bound by static settings anymore than it needs to be.

## 2.3 Summary

This chapter presents an overview of generative systems, their applications in medicine and unrelated fields, and their fundamental operating workflows. As seen by the literature reviewed, while generative models which implement conditionality requirements have become commonplace in recent years, those that have been applied to the medical field are few and far between.

And so, while state-of-the-art works in conditional image generation achieve good results in their tasks, their performance is entirely reliant on the quality, complexity and nature of both the data and the labels provided. On the other hand, few studies exist on generative issues similar to that established by the project and also applied to medical imaging data and the ones that do focus on imaging tend to deal with non-conditional image-to-image translations. Expressed differently, they mostly focus on how the interactions found within the input data itself and not, for example, on how such interactions will relate to label parameter variations, which is something the project aims to do.

This project strives to intertwine both modalities together, hence there are few terms of comparison to be relied upon when evaluating the tools designed in this dissertation. For this reason, the models designed can only be put to the test against each other and against modifications of others similar to those presented in this chapter, and not against results previously obtained in other studies.

Lastly, the one study this model received valuable input from was a precursor to the project itself which established the parameter subsets used all throughout the experiments detailed, as well as the intrinsic value of a voxel-wise approach to the matter of dMRI scan generation.

# Chapter 3

## Methodology

In this section, we present the proposed techniques for training and testing all the models considered, including their creation process and the reasoning behind their use. The objective is to evaluate the best method for generating accurate, precise and realistic scans with on-demand scanner acquisition settings.

In section 3.1, the utilised Multi-Dimensional Diffusion MRI (*MUDI*) dataset is presented. Following this, there is the need to, in section 3.2, state the main tasks being approached in furtherance of the presented final goal. The methodologies and models employed for each task are then explained in more detail in section 3.3. At last, a series of metrics employed to evaluate every model's performance is also disclosed, as of section 3.4.

### 3.1 Dataset Preparation

#### 3.1.1 Dataset Acquisition

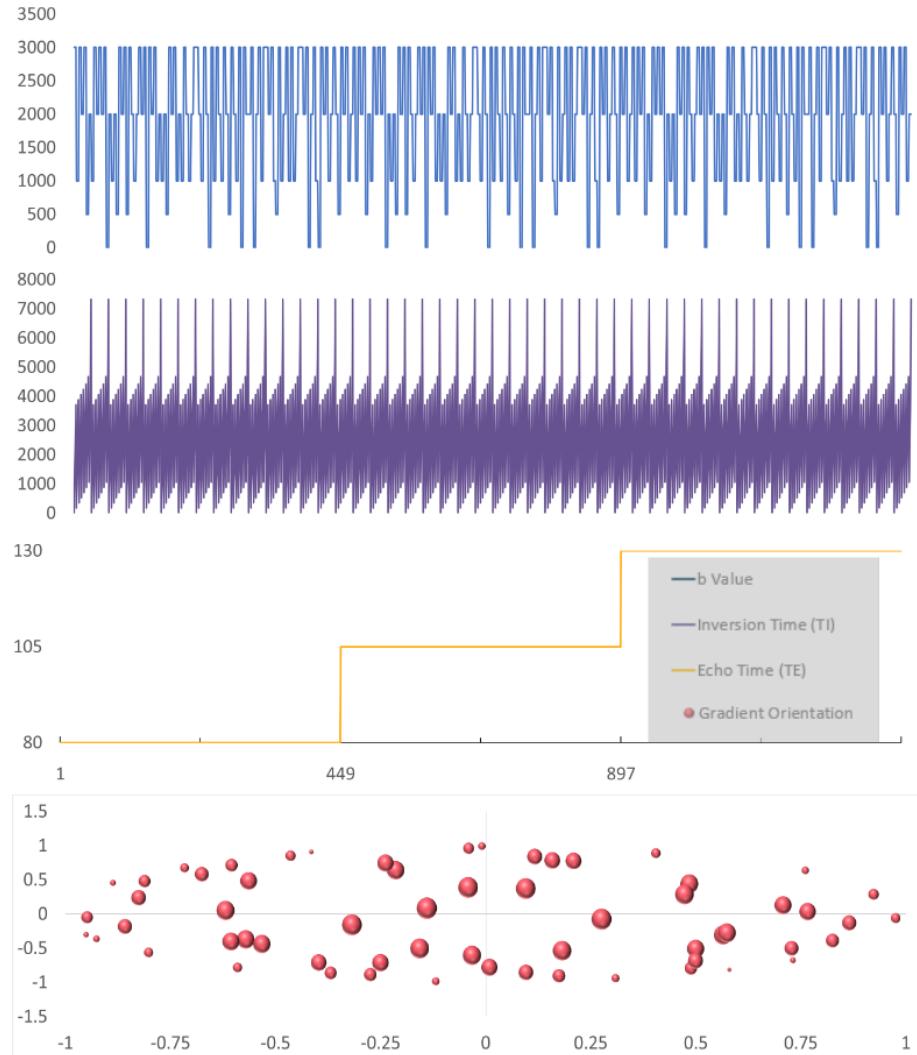
Organised as part of the 2019 Medical Image Computing and Computer Assisted Intervention Society (*MICCAI*) Conference, the MUDI Challenge [24] aimed to reduce the amount of images acquired by dMRI procedures by exploiting information redundancies. It then used these scan subsets to predict the entirety of the densely sampled multidimensional data from only sub-sampled versions of itself. As per a continuation of this challenge, the project has used the same dataset, containing MultiDimensional Diffusion MRI scans, for which acquisition settings are also registered.

The complete dataset is, as of now, comprised of the scans for five healthy human volunteers, of which 3 males and 2 females, with ages ranged in between 19 and 46 years. This acquisition was approved and performed with due informed consent (REC code: 12/LO/1247). For each patient, 1344 fully 3D DWI brain volumes were obtained on a clinical 3T *Philips Achieva* scanner in Best, Netherlands, utilising a 32-channel adult head coil. Each volume can be described as being obtained with a different combination of the scanner's parameter settings, of which a specific few are variable, while others remain unchanged.

**Table 3.1:** MUDI dataset's variable & static acquisition settings labels, including their value / interval and distribution

| Acquisition Parameter        | Variability | Value / Interval  | Distribution           |
|------------------------------|-------------|---|------------------------|
| b Value                      | Variable    | $b \text{ Value} \in \{0, 500, 1000, 2000, 3000\} \text{s/mm}^2$      | Non-Uniform            |
| Echo Time (TE)               | Variable    | $\text{TE} \in \{80, 105, 130\} \text{ms}$                            | Fully Balanced         |
| Inversion Time (TI)          | Variable    | $\text{TI} \in [20, 7322.7] \text{ms}$ in 28 mostly regular intervals | Fully Balanced         |
| Gradient Orientation         | Variable    | 106 Uniformly spread Directions                                       | Balanced w/ Exceptions |
| Repetition Time (TR)         | Static      | 7.5 seconds   |                        |
| Resolution                   | Static      | 2.5mm Isotropic   |                        |
| Field-of-View (FOV)          | Static      | 220 x 230 x 140 mm  |                        |
| Sensitivity Encoding (SENSE) | Static      | 1.9   |                        |
| Halfscan                     | Static      | 0.7   |                        |
| Multiband Factor             | Static      | 2   |                        |

A list of all the variable and static acquisition settings can be found in table 3.1, as well as some insight into the organization of the data itself, as it comes to the variable parameters being included as labels to the dMRI volumes, in the form of figure 3.1.

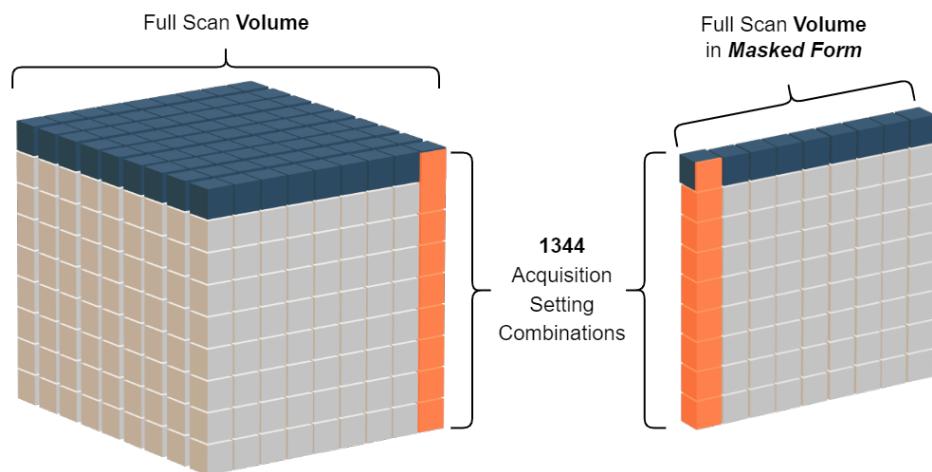
**Figure 3.1:** Distribution of all 4 settings, across volume acquisition process. The third gradient orientation direction is represented by the bubble's radius.

### 3.1.2 Scan Preprocessing

After all 1344 volumes were acquired, they were reconstructed and also denoised in the complex domain [25]. Since all scans were acquired with the exact same scanner and performed by the same investigator, there was no need for data harmonisation to correct for any possible variations caused by such external factors. However, it should be noted that the scans were collected with reversed phase-encode blips [26], to correct any susceptibility-induced distortions, using the Software Library designed by the Oxford Center for Functional MRI of the Brain (*FSL*) [27].

It should also be stated that scans are different in shape when comparing one patient's dataset to another, which will also lead to a different number of voxels in datasets' masked form. Therefore, using the provided patient-specific masks and as a way to allow for image reconstruction, the full dataset has been used, in all instances, by segregating one patient from another. A further, more visual explanation of how the data is organized, per patient, can be found in figure 3.2.

The dataset was, ultimately, divided into training, validation and test sets in a patient-wise manner, which were kept throughout all experiments. This three-way split was made so that the training set would keep the first 3 patients, and validation and test sets 1 patient each. Various sub-selections of the acquisition parameters were also created.



**Figure 3.2:** Visual Representation of the MUDI Dataset's Organisation: a full masked volume including all intensities for all the patient's image voxel positions (blue); all voxels pertaining to the same position, though for different volumes, meaning different parameter combinations (orange). In reality, for voxel-wise implementations, the actual organization of data is much better represented in the scheme to the right, seeing as masked volumes will extend into all the different voxel positions dictated by the patient's mask.

### 3.1.3 Parameter Handling

Taking the task into account, each volume is paired with a set of labels composed of the variable acquisition settings being experimented with, i.e., b Value, Inversion Time (TI), Echo Time (TE) and Gradient Orientation.

Although at first 3 unit vectors were provided as the volume's gradient orientation for the model to work with, in the form of Cartesian coordinates, much like  $(\alpha_x, \alpha_y, \alpha_z)$ , these were changed into angles or, in other words, 2 polar coordinates,  $(\theta, \phi)$ . The formula utilised to perform this conversion is presented in equations 3.1 and 3.2.

$$\theta : \arctan2\left(\frac{\alpha_y}{\alpha_x}\right) \quad (3.1)$$

$$\phi : \arctan2\left(\frac{\sqrt{\alpha_x^2 + \alpha_y^2}}{\alpha_z}\right) \quad (3.2)$$

It should also be noted that all parameters / labels used were also manually normalised, to allow a fine-tuning of their values and ensure that they are given their due relevance when being mapped to and from the employed models. Table 3.1.3 displays the normalisation method applied to the different parameters, as well as their initial and final range after conversion / normalisation.

By looking at the distribution for all the normalised variables being provided for modelling, it is clear that no two parameters have the exact same range, and so, will not be treated exactly alike during modelling. This makes settings such as gradient orientation much more variable and, thus, potentially problematic, than b Value, whose values are, in the dataset, almost discrete.

**Table 3.2:** Parameter Manual Normalisation Methods

| Acquisition Parameter | Initial Range   | Normalisation Method | Final Range  |
|-----------------------|---|----------------------|--|
| b Value               | [0, 3000]   | $bValue/10000$       | [0, 0.3]   |
| Echo Time (TE)        | [0, 7000]   | $TE/10000$           | [0, 0.7]   |
| Inversion Time (TI)   | [80, 130]   | $(TI - 80)/200$      | [0, 0.25]  |
| Gradient Orientation  | $\begin{cases} \theta & [-\pi, \pi] \\ \phi & [0, \pi] \end{cases}$ | $\alpha + \pi$       | $\begin{cases} [0, 2\pi] \\ [\pi, 2\pi] \end{cases}$ |

## 3.2 Task Partitioning

This section will display how the different tasks were stipulated in an attempt to achieve the main goal by varying two different factors: the amount of information the model has on the scans being generated and its acquisition settings; and whether or not the model can leverage information on the input images' MRI combination parameters. Therefore, each sub-task will be different in the way it tackles the main goal in regards to how much liberty they allow the input and output sides, when it comes to the number and choice of the scans required and generated.

It should also be said that this was mostly attempted in a **voxel-wise** manner, meaning one in which the model is expected to receive a combination of voxel intensities, either of the same scan, or the same position in different scans and find the ideal function to map them to the voxel intensity corresponding to the same image position in different scans. Then, these must be unmasked to get the actual scan to be displayed in a comprehensible, image-like representation, also creating the need to generate a mask specific to each patient in the dataset. However, one of the approaches does include an **image-wise** term of comparison, in which samples can be described as being volume slices, thus allowing further insight into the anatomical points of interested captured by dMRI scans, which a voxel-based approach does not.

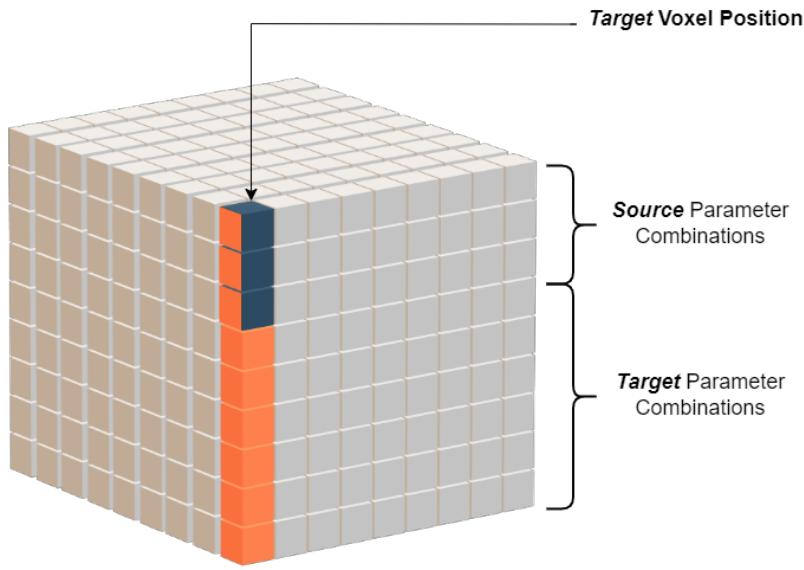
### 3.2.1 Baseline: Some-to-Some Approach

This project is a direct continuation of the methodology proposed by Tax *et al.* in [21] and [22], and as such, it would be advisable to use the task proposed in that paper as a starting point and as a general term of comparison. In view of the very satisfactory results of that endeavour, this choice becomes even more obvious.

This baseline task is one in which a varying number of scans are sub-selected and provided in their masked form, to which the name of 'source scans' is given. These are fed to the model in linear fashion and, in this case, in a voxel-wise manner, preventing any model who is given the task to have a full scope of the scan's spatial information, seeing as the same sample will only ever have voxel intensities for the same position but not for the same scan. This variable number of scan measurements will then be used to predict a total of 1344 features, which one can train the model to compare to the entirety of the scans a patient dataset is comprised of.

It is, therefore, expected of the model to generate every intensity for that same voxel position and all scans contained within the dataset, from the intensity of a single voxel position for all the sub-selected training scans. If this process is repeated for every single voxel in the patient's 3D image, the full scope of the dataset can be recreated, including the 'target scans', or scans for the parameter combinations which were not selected to be part of the source data. Based on previously shown schemes, a visual interpretation of the input and output data can also be found in figure 3.3.

This will also mean that the task is very limited to the scans' parameter values, meaning that if one were to change either the input or the target scans' parameter settings, the model would not be able to accurately predict the scan's appearance anymore, seeing as it has been trained on fixed parameters for both input and output. This can also be demonstrated by the fact that in no



**Figure 3.3:** Baseline / Some-to-Some Approach's Input (blue) and Output (orange) Data Organisation within the space represented as the MUDI Dataset

way, shape or form does the model ever actually have access to what these parameters' values are, nor are they provided at any point. This proves what can only be considered an **absolute lack of conditionality**, or, in other words, the inability to map from, and also to, a single scan of a specific parameter combination, instead of mapping to all scans contained within the dataset.

### 3.2.2 Some-to-Any Approach

Taking this into account, the project moved on to include some conditionality on the output side, by allowing the choice of a single target parameter combination to which the input would have to be mapped to.

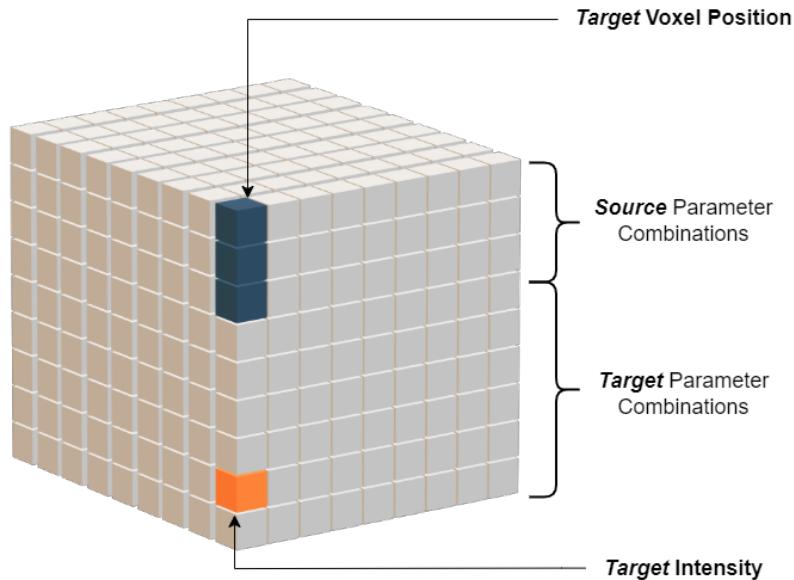
#### 3.2.2.1 Voxel-Wise Implementation

Mimicking the way data is fed in the baseline approach, there will also be a varying number of source acquisition setting combinations provided as input. However, these will be mapped to a single target, instead of to all the ones available in the dataset. Being this an approach in which the data is still provided in its masked form, it should still be evident that no spatial correlation inside the scan can ever be asserted by this model, seeing as the samples can only establish relationships between the source parameters and between the source and the target combinations and their values.

The model should, consequently, be able to map to the intensity for any chosen target parameter amalgamation, given the necessary ingredients: the intensities of all voxels for that same position and corresponding to the sub-selected source setting combinations the model was trained on, and the target parameter values one wishes to map to.

Mapping to every single parameter would require, for every single voxel position in the somewhat big volumes the MUDI dataset contains, a lengthy process due to the enormous amount of iterations this entails. For this reason, on occasion experiments with this model do use a random portion of the voxels, different for every single one of the training patients and mutable at the end of the training epoch. It should also be stated that various experiments were created, each aiming to decrease the amount of sub-selected source parameter combinations and, consequently, the amount of data the model must learn from in order to map to any of the ones in the target subset, making the task's achievement harder, but also allowing more freedom from the input side.

Similarly to what has been done for the previous sub-task, a visual explanation of how the input and output data relate to each other and the rest of each patient's dataset can be found in figure 3.4.

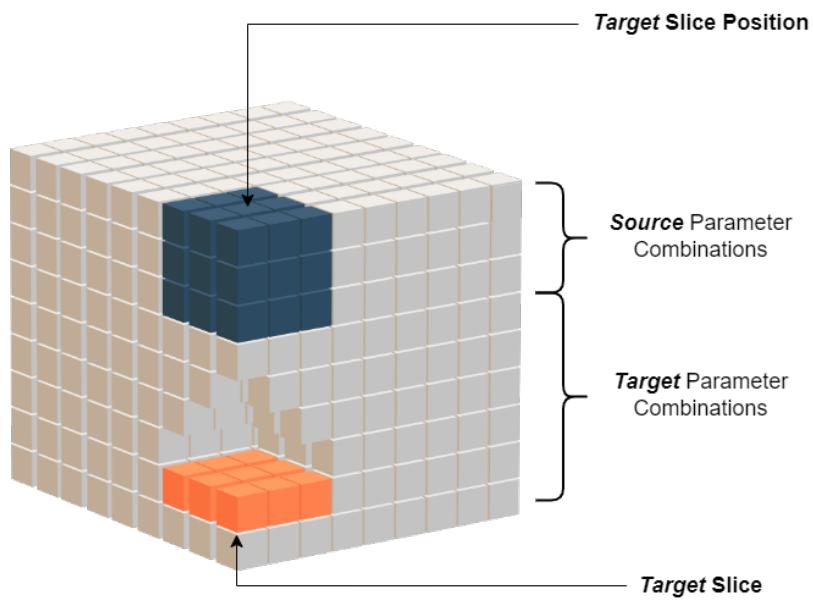


**Figure 3.4:** Voxel-Wise Implementation of Some-to-Any Approach's Input (blue) and Output (orange) Data Organisation within the space represented as the MUDI Dataset

### 3.2.2.2 Slice-Wise Implementation

This task was not only performed in a voxel-wise manner as described here, but also in an image-based one, implying that the samples provided are 2D volume slices, and the task itself, although the same, shifts so that instead of voxel positions, the mapping becomes centered around slice index inside the 3D volume. Apart from this minor difference, the main idea for this approach is still to create a design which can map that input to a target settings combination, given a set slices of the same index in the 3D image and pertaining to the sub-selected source parameters. This dataset feeding process and its similarities with the voxel-wise implementation for this task can also be visualized in figure 3.5.

This secondary approach is also aided by the fact that even though different patient's volumes will have different 3D shapes, if turned to an axial orientation, they do all have the exact same number of 2D slices: 56, which correlate to one another in a way that allows for direct mapping in between each other, much like what has been done previously in a voxel-wise manner. However, the shape of each slice does vary from patient to patient, as evidenced by the volume masks provided in the dataset. Hence, **zero-padding** has been employed as a preprocessing method to correct for that inconsistency. All 3D scans were therefore made to resemble a [56, 96, 96] shape for this specific sub-task.



**Figure 3.5:** Slice-Wise Implementation of Some-to-Any Approach's Input (blue) and Output (orange) Data Organisation within the space represented as the MUDI Dataset

Contrary to what is true for this approach's voxel-wise counterpart, the number of slices also makes it infeasible to consider leaving some out of the training process, meaning all were used for all subsets of the MUDI dataset. Having said this, three experiments were setup, in which the number of scans provided to the model and of sub-selected source parameters vary, just the same as they did for the voxel-based implementation of this approach.

### 3.2.3 Any-to-Any Approach

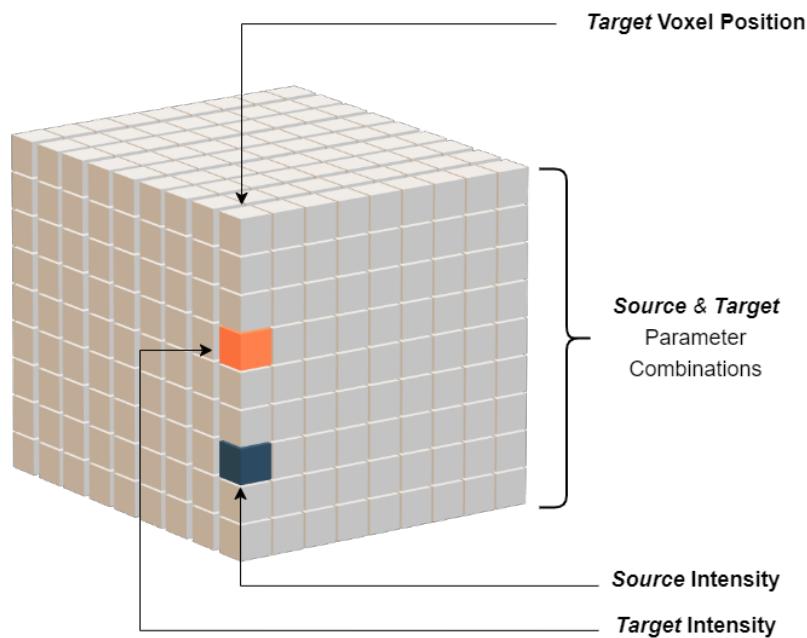
Following the presented line of reasoning, the project progressed onto also including full conditionality on the input side, by allowing the choice of not only to which single parameter combination to map to, but also which one to map from.

Therefore, given the premise of an any-to-any approach, the task asks for a learning process on how to map from any of the available mixture of acquisition settings to any other, which multiplies the number of interactions and the learning curve by a great amount. It also forces a model to truly

understand the effect of every variable on the look of the dMRI scan, not only individually, but also as a whole and when combined.

Similarly to what was done for the some-to-any approach, the input requirements of this approach can be stated to be: the intensity of the voxel for the chosen image position and for the source parameter combination; and the parameter values for both the source volume as well as the target one. The algorithm should, given this, be able to generate an accurate prediction for the intensity of the voxel, for the given position, on the target scan. Due to the enormous endeavour that would be to map from every single one of the 1344 combinations to every single other, for every single voxel in each patient's full dataset, this task's model was trained, and also tested, using only a randomly selected portion of the available samples. This intents to maximise the reach of the model's understanding, while also minimizing the computational time required for both the training and testing process, which would otherwise be impractical.

A more discernible explanation of these requirements is also in figure 3.6, so as to understand how the approach chooses which values to map to and from within the dataset's established structure.



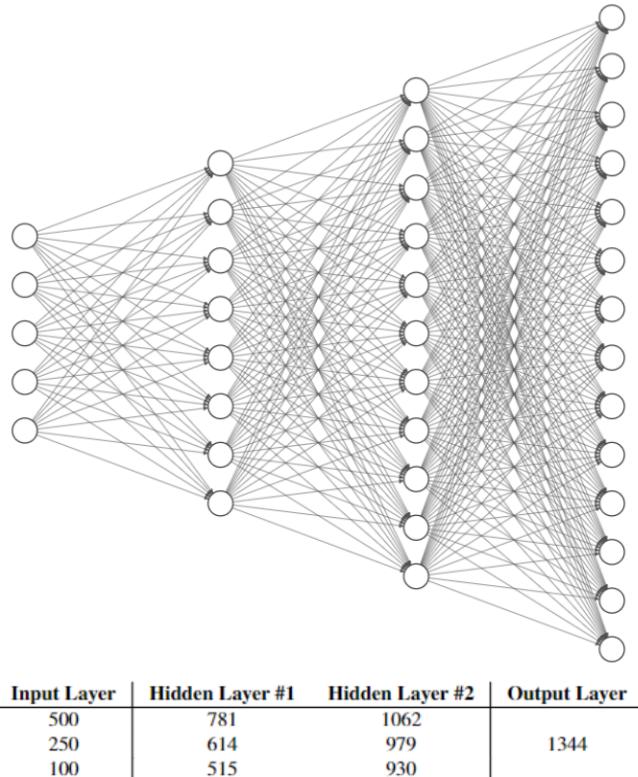
**Figure 3.6:** Voxel-Wise Any-to-Any Approach Model's Input (blue) and Output (orange) Data Organisation within the space represented as the MUDI Dataset

### 3.3 Model Design

This section details the design and architectures of all models used for each of the approaches in 3.2, including how they relate to the different sub-tasks and how the different experiments make an attempt at iteratively increasing the training process's difficulty by reducing the amount of data the model has to work with. A summary of all the different implementations is displayed in table 3.3.3.

One aspect of all the constructed models to state beforehand is that all networks lack any type of activation in the last output-producing layer. This is standard for generative problems, since it allows the output to not be constrained within a specific range or distribution, and to fit into the complex patterns and structures the model has found within the data, which could not be fully represented by a single activation function. Moreover, all voxel-wise implementations are done so using fully linear and non-convolutional models, so as to avoid the phenomenon of overfitting, which would see its likelihood greatly increased by the small size and nature of these approaches' input. Finally, batch normalization is a common occurrence among the architectures as a way to increase the training and weight initialization times.

#### 3.3.1 Baseline: Some-to-Some Approach



**Figure 3.7:** Baseline / Some-to-Some Approach's Model Architecture, including number of Neurons for all Layers depending on the Experiment. All the input layer's nodes will be reserved for measurements / voxel intensity values.

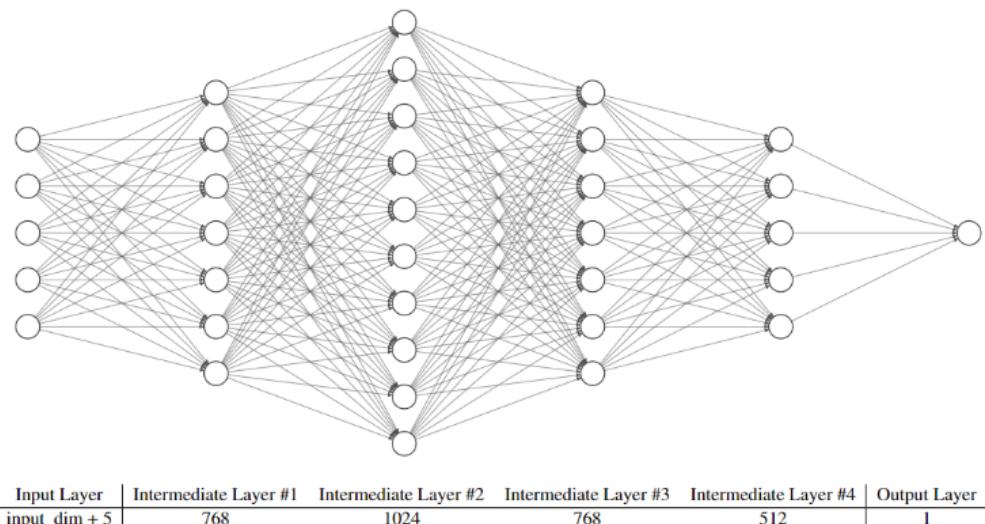
The baseline model can be described as a fully connected NN, seeing as it was constructed in a linear fashion, composed of an various modules, each with increasing number of nodes, and each with a linear layer followed by batch normalization and, finally, a ReLU activation function. Therefore, it is implied that it will first start learning low-level features. These can then be used in the last layer to construct an array of intensities containing all 1344 parameter combination for that voxel position, which, as is common to generative models, possesses no activation of its own. As such, its architecture can be described as something to the likes of figure 3.7.

Do take into account that the model's architecture, namely the number of neurons in the intermediate hidden layers will vary depending on the experiment being performed, of which 3 have been done. These numbers follow the same pattern, seeing as the size of each layer follow an equal spacing starting at the input layer and ending in the 1344 which represents the total number of target parameter combinations. In each experiment detailed in table 3.3.3, the number of sub-selected source setting combinations is decreased, meaning the model will have progressively less input combinations to map from. This makes the acquisition process easier for newcomer patients, as the amount of necessary volumes the model needs to be fed becomes increasingly lower.

### 3.3.2 Some-to-Any Approach

#### 3.3.2.1 Voxel-Wise Implementation

The Some-to-Any voxel-based model can still be described as a fully connected NN, although there are clear differences between it and the baseline approach's architecture. The functions of which each layer of the network is composed of, however, do remain the same: a linear layer followed by one-dimensional batch normalization and activated by ReLU. This approach's design is summarised in a visual form by figure 3.8.



**Figure 3.8:** Voxel-Based Implementation of Some-to-Any Approach's Model Architecture, including number of Neurons for all Layers. Of the input layer's nodes a total of 5 are reserved for parameter values and the remaining are measurements / voxel intensity values.

With the intent to compare this approach's performance to the baseline one, 3 experiments were also performed for this model, in which there is a decrease in the number of sub-selected source parameter combinations provided (described in image 3.8 as  $num\_input \in \{500, 250, 100\}$ ). Nevertheless, these values do not affect the rest of the design and the number of neurons for any other layers other than the input one. For all the experiments described, the percentage of possible voxel positions used was of 60% (randomly selected and changed every epoch) for training patients and 100% for validation ones, so as to assure results are representative. These experiments and further ones are described in detail in table 3.3.3.

However, when compared to the baseline approach, there is a big difference regarding the organisation of the layers themselves and in how there is an attempt to first expand and then contract the network, for a few different reasons:

- While not being used to capture spatial and hierarchical features, as is the case for most **expansion and contraction patterns**, there is still a level of utility served by this choice when it comes to the processes of feature transformation and dimensionality reduction [28] ;
- The dichotomy of an expansion followed by a contraction into the final output allows the design to learn successive layers of **feature abstraction**. These are relevant if, as is suspected for the data at hand, the underlying relationships are not entirely linear, which will be a task for the intermediate layers to deconstruct ;
- By forcing the network to work under such a constrained structure, there is also a level of **implicit regularization** being done ;

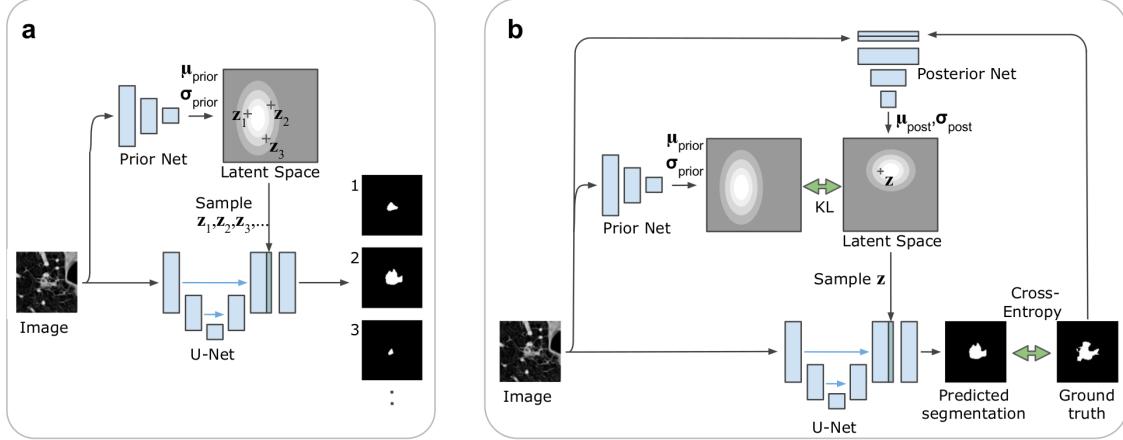
### 3.3.2.2 Slice-Wise Implementation

The design chosen to be the most suitable for this task was detailed by Kohl *et al.* in [5] (2018) as a combination of a probabilistic U-Net with a Conditional Variational Autoencoder (CVAE) trained for generative segmentation purposes.

The proposed work involves sampling the images to a low-dimensional latent space, in which each position aims to best encode a segmentation variant. A 'prior net' will then estimate the probability distribution for each of these variants given the input image, which can be used to draw random samples to be used in the U-Net's feature map. At the same time, the so-called 'posterior-net' will learn to recognise the segmentation variants by comparing the input image to the ground truth segmentation, which it can then map inside the latent space, allowing for the generation of a final segmentation output. Similarly to CVAEs, the training will also use a **Kullback-Leibler Divergence** term to both minimize the variational lower bound and align the posterior and prior nets' distributions.

In order to fit the task provided, the original model was modified in specific sites. By including the acquisition parameters as channels inside the provided slice, it is asserted that, with each of the voxels, all 5 of its scanner settings are coupled with the voxel's intensity. This provides the model

more awareness about how the target ground truth cross section sample relates to these values and to the intensity of each of the voxel intensities, for the same position slices, for all the sub-selected source volumes.



**Figure 3.9:** Slice-Wise Some-to-Any Approach Model’s Architecture [5].

Furthermore, in an attempt to make the model purely generative, the number of segmentations was also fixed at 1, tricking the design into generating a prediction of what the target sample’s cross section would look like. The training loop then uses this to condition the model into learning the spatial representations in-between slice pixels and the interactions in between the fixed source parameters and the target one.

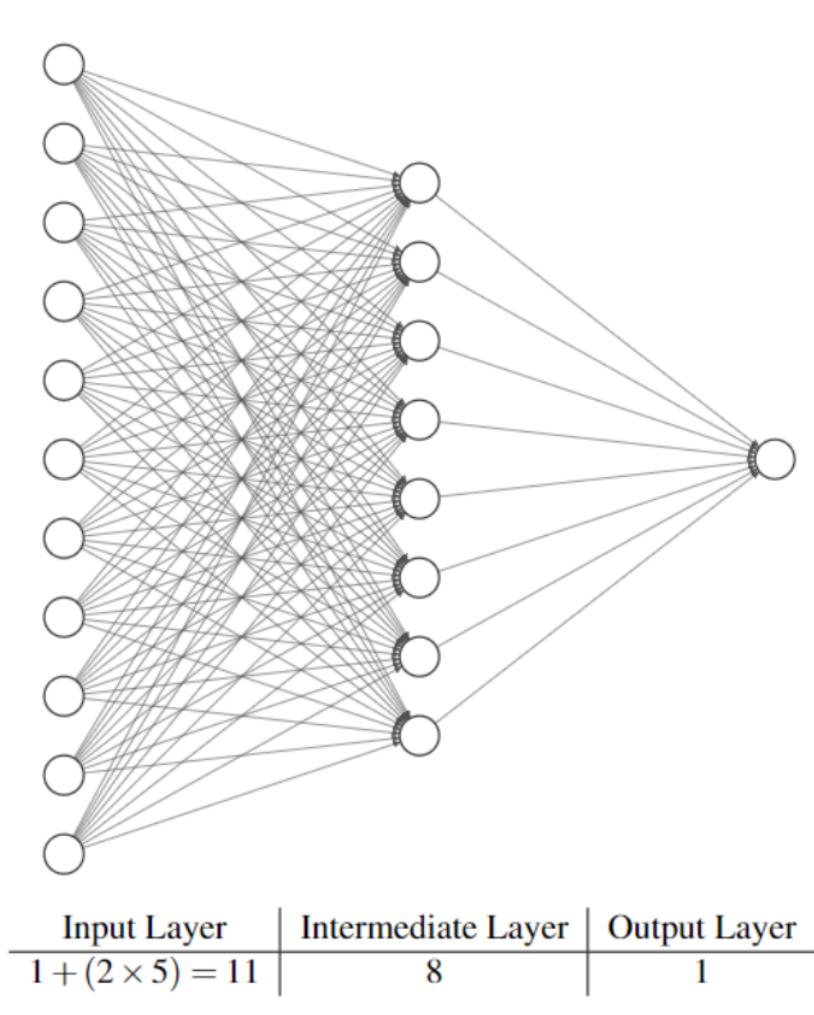
And still, it is noticeable that the elements of the input requirements for both blueprints still locks them down on needing to be trained and then used with the same set of source parameters. This concept will impossibilitate the idea of being able to feed the model any available volumes without heeding what their acquisitions settings are.

Similarly to its voxel-based counterpart, this architecture was also subject to 3 experiments, each with an input number of source parameters of  $num\_input \in \{500, 250, 100\}$ . The concept of this approach also allows for the introduction of a test set to the target parameters, meaning combinations which have not been mapped to before in the process of training, and, subsequently, which the model has not seen yet. With this, one can test the model’s adaptability and its grasp on the actual values of each of the 5 acquisition settings and what the underlying correlations and structures between these and the volumes are. These experiments were also set up using the voxel-wise build, and a subset of 100 random test parameters created, much like table 3.3.3 shows.

### 3.3.3 Any-to-Any Approach

Following the voxel-wise approaches’ trend of using linear neural networks, this model’s design is, understandably, quite simple. This architecture also accommodates the fact that the input for this model will only ever be comprised of a single intensity value and the corresponding values for all 5 shown variables for both the volume the intensity is from and the volume one intends to map that intensity to. Only one intermediate layer is created and both it and the input one do have,

as per what was used previously, batch normalization and a ReLU activation, just as figure 3.10 displays.



**Figure 3.10:** Voxel-Wise Any-to-Any Approach Model’s Architecture, including number of Neurons

A total of two experiments were delineated using this architecture. The first sees the model faring against all possible combinations of both source and target parameter combinations present in the dataset, while the second eases the task at hand by only allowing the model to draw the source settings from a pool of 500 combinations and the target ones from the remaining 844. As previously mentioned, only a portion of each of these source-target mixtures and possible voxel positions was actually used in each experiment, these being described in table 3.3.3.

Despite its lack of complexity, this sub-task’s approach can still be considered an over-arching ideal voxel-wise solution to what the posed problem is. In other words, there is no lack of conditionality to be had here, and the user can control both which scans are acquired for any new patients, and which to acquire through usage of the model. Moreover, it would also be possible to, for example, only slightly adjust a single parameter and obtain the requested volume, thus proving a clear-cut fit for the situation.

Lastly, and as a conclusion to the methodologies, a table containing all the approaches, their tasks and what experiments have been outlined using each one can be found in [3.3.3](#). For each experiment, there is also detailed information on the number of parameter in all 3 subsets: source, target and test, as well as the percentage of used voxels / slices and combinations of source-to-target used for both training and validating each model.

**Table 3.3:** Tasks and approaches included in the methodology chapter, including experiments created for the models and their specificities

| Sub-Task                        | Approach   | Parameter Subsets |                     |             | Train Set               |                            | Validation Set          |                            |
|---------------------------------|------------|-------------------|---------------------|-------------|-------------------------|----------------------------|-------------------------|----------------------------|
|                                 |            | Source Subset     | Target Subset       | Test Subset | Voxel / Slice Positions | Source-Target Combinations | Voxel / Slice Positions | Source-Target Combinations |
| Some-to-Some<br>(Source ->1344) | Voxel-Wise | 500<br>250<br>100 | 1344                | 0           | 100%                    | 100%                       | 100%                    | 100%                       |
| Some-to-Any<br>(Source ->1)     | Voxel-Wise | 500<br>250<br>100 | 844<br>1094<br>1244 | 0           | 60%                     | 100%                       | 100%                    | 100%                       |
|                                 |            | 500<br>250<br>100 | 744<br>944<br>1144  | 100         | 60%                     | 100%                       | 100%                    | 100%                       |
|                                 |            | 500<br>250<br>100 | 844<br>1094<br>1244 | 0           | 100%                    | 100%                       | 100%                    | 100%                       |
| Any-to-Any<br>(1 ->1)           | Voxel-Wise | 1344<br>500       | 1344<br>844         | 0           | 30%                     | 2.5%<br>10%                | 100%                    | 0.5%<br>1%                 |

### 3.4 Evaluation Framework

For the purpose of fairness and computational equity, all models were trained using the exact same learning rate of  $\alpha = 1e - 4$ , weight decay step of  $w_d = 1e - 5$  and, finally, stochastic optimization algorithm, this being ***AdamW***, which works much like its original counterpart ***Adam***, with the added argument that weight decay is decoupled from the gradient update step. In other terms, by performing the weight decay not as an extension of the cost function but inside it, this optimizer allows for a regularization term which will not be considered within the moving averages, and will only ever depend on the value of the weight itself. This creates a better environment for a faster, more concise training process, as proven by Loshchilov and Hutter [29] (2017).

Moreover, with the intent of providing each of the shown designs the same opportunity to perform, their training process was not constrained on a specific number of epochs, but rather on how their performance increases over time and over the iterative task. As such, the project has utilized **Early Stopping** as a way to determine when to conclude the training of each of the presented architectures, which, as per the parameters stipulated, will make it so that the model will have to stop once it has reached 5 epochs with no clear improvement in the training loss for the considered validation set. Improvement is also specified to only be qualified as such under a threshold of  $\delta_{min} = 1e - 5$ .

**Table 3.4:** Tasks and approaches included in the methodology chapter, including experiments created for the models and their number of training epochs

| Sub-Task                               | Approach          | Parameter Subsets |               |             | No. Training Epochs |
|--|-------------------|-------------------|---------------|-------------|---------------------|
|  |                   | Source Subset     | Target Subset | Test Subset |                     |
| <i>Some-to-Some</i><br>(Source ->1344) | <i>Voxel-Wise</i> | 500               |               |             | 23                  |
|  |                   | 250               | 1344          | 0           | 31                  |
|  |                   | 100               |               |             | 22                  |
| <i>Some-to-Any</i><br>(Source ->1)     | <i>Voxel-Wise</i> | 500               | 844           |             | 33                  |
|  |                   | 250               | 1094          | 0           | 37                  |
|  |                   | 100               | 1244          |             | 29                  |
|  | <i>Slice-Wise</i> | 500               | 744           |             | 33                  |
|  |                   | 250               | 944           | 100         | 36                  |
|  |                   | 100               | 1144          |             | 32                  |
| <i>Any-to-Any</i><br>(1 ->1)           | <i>Voxel-Wise</i> | 500               | 844           |             | 42                  |
|  |                   | 250               | 1094          | 0           | 47                  |
|  |                   | 100               | 1244          |             | 49                  |
| <i>Any-to-Any</i><br>(1 ->1)           | <i>Voxel-Wise</i> | 1344              | 1344          | 0           | 30                  |
|  |                   | 500               | 844           |             | 29                  |

At last, it should also be said that the same losses were used for both back-propagating and evaluating each approach's performance, and that all have used GPU accelerating and multi-processing for the dataset loading. Seeing as the object being evaluated are, ultimately, the generated scans, the employed losses make an effort to reflect not only how accurate they are when compared to the ground truth for each patient, but also how realistic, considering the normal dMRI scan visual distribution. Each model's number of training epochs is included in table 3.4.

### Mean Squared Error (*MSE*)

The first presented function and the one which was uniquely used for the back-propagation of every single model presented was Mean Squared Error or L2 Loss, which, as the name implies, follows a very simplistic function, presented in equation 3.3:

$$mse\_loss = \frac{1}{n} \sum_n^{i=1} (y_i - \hat{y}_i)^2 \quad (3.3)$$

*n*: No. of Data Points     $y_i$ : Observed Value     $\hat{y}_i$ : Predicted Value

As per its nature, *MSE* is a very analytical loss function, seeing as it only considers the values being presented and how they are numerically distant from the considered ground truth. In that regard, it is a loss used very often and in any context, especially if, as is the case during the training process, the presented sample is not a dMRI scan or slice, but just a randomly assorted combination of voxels.

### Mean Absolute Error (*MAE*)

As of the context of this project, the function of Mean Absolute Error, or it is more commonly known, L1 Loss, was used mostly with the intent of evaluating the designs' performance in such a way that it would compliment the analysis provided by the previous MSE loss. Similarly to this squared counterpart, it also follows a very simplistic formula, as described in equation 3.4:

$$mae\_loss = \frac{1}{n} \sum_n^{i=1} |y_i - \hat{y}_i| \quad (3.4)$$

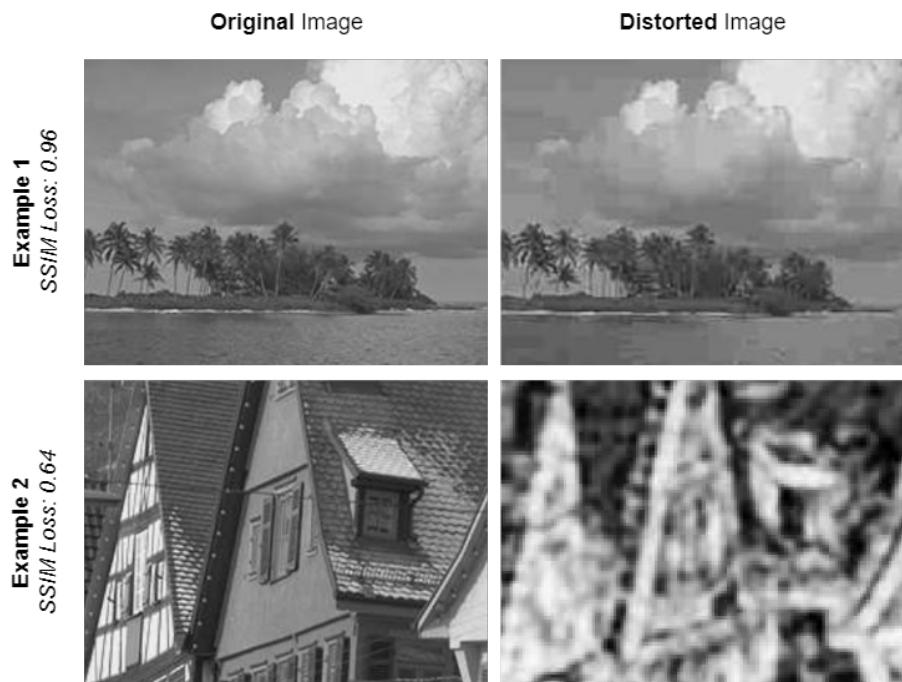
*n*: No. of Data Points     $y_i$ : Observed Value     $\hat{y}_i$ : Predicted Value

Similarly to MSE, MAE is also quite analytical in its role, however, it allows for a broader analysis of results in the sense that, unlike MSE, it does not overly punish large errors, and so it becomes easier to interpret and assess performance using it.

## Structural Similarity Index (SSIM)

This perceptual metric was designed by Wang *et al.* [30] to quantify image quality degradation. As such, it focuses on visible structures on the generated image, comparing them to the ones in the provided ground truth, so as to create a more holistic, visual-based and human-like distinction between the two. This is done by taking some different aspects such as luminance, contrast and structure organization into account, and doing so in clusters of voxels, thus allowing a broader overall vision of what the differences might be. It also utilizes a moving-windows methodology to do so, allowing it to look at patches of pixels rather than single ones, which is much closer to the human way of perceiving imagery.

For better understanding of how the SSIM index assesses images and correlates them to one another in a perceptual way, some examples where included in 3.11. As can be seen, 2 cases are presented: the first sees very little distortion when it comes to the final image, and, to the naked eye, the 2 images look quite similar, thus the structural similarity index is close to 1, the maximum. The second example displays 2 very different images, even though one stems from the other and still contains many of its characteristics, such as limits and some of the contrast, proving a smaller structural similarity index, closer to the minimum of 0.



**Figure 3.11:** Example Imagery for Structural Similarity Index Explanation: to the left are the original images and to the right are the distorted ones for both examples.

Although not being used as an actual back-propagating loss function for any of the approaches, it was monitored quite frequently as a performance evaluator. Due to the overall goal of the project, this becomes a relevant choice, since it allows some insight into how the generated images could, from the human perspective, be considered realistic or not.

# Chapter 4

## Results and Discussion

This chapter displays and discusses the results obtained through the methods developed in Chapter 3. Firstly, both the robust and fragile points of all algorithms developed for each task will be presented. This includes the sole voxel-wise approaches for the Baseline and Any-to-Any endeavours in sections 4.1 and 4.3 respectively, but also a comparison between the 2 techniques created for the Some-to-Any issue, in section 4.2. Finally, section 4.4 deals with an overall comparison of the results for all tasks as a way to achieve the project's ultimate goal. Do take into account that all results displayed stem from the same patient, selected from the 5 in the MUDI dataset to remain off the training and validation processes, and that most scans displayed can be traced back to the intermediate-most slice of the 3D volumes, seeing as it presents the largest amount of information to be had from all available 56 cross sections in the axial cut.

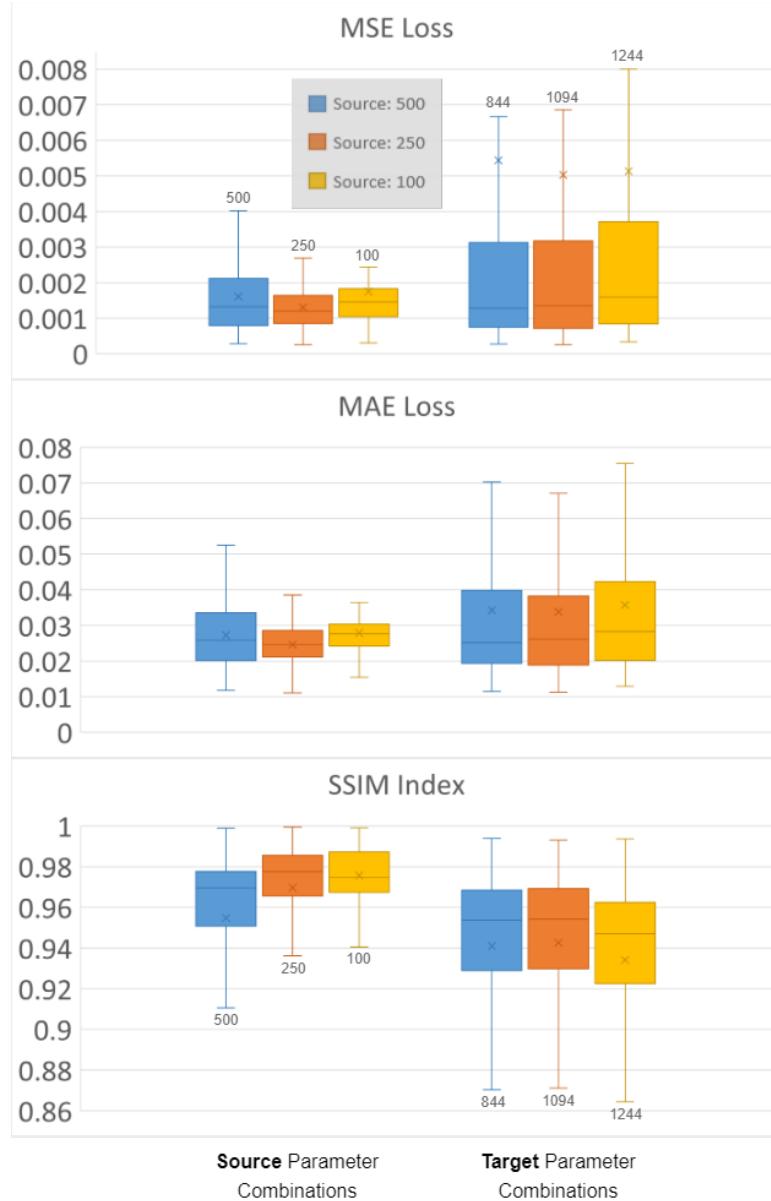
### 4.1 Baseline: Some-to-Some Results

The baseline approach, and its design were chosen due to having proven before to be very effective in its task and for the dataset also used in this project [22][21]. As such, results for this architecture were improved upon by introducing new experiments and metrics with which to analyse them.

Given the fact that this is the only approach, of the ones presented, in which the model is forced to generate scans for the source subset of parameter combinations with which the model was trained with, results shown should be constrained by this fact, so as to not over amplify this approach's quality results. This is validated by analysis 4.1, seeing as the results for all 3 experiments show considerably higher interquartile range (IQR) MSE and MAE values and lower SSIM overall ones for scans obtained corresponding to the source parameter combinations which the model has been fed as training data. This discrepancy is also accentuated by the amount of target settings, which become exceedingly larger than their source counterparts along the experiments, contributing to overall worse results for the graph's right side.

Overall, it should be clear to see that the second experiment, of 250 source parameters, was the one to broadly yield the best results, and not the first, more well-informed one. This can be traced

to how this experiment displays lower or comparable MSE and MAE median values to those of the 1<sup>st</sup> experiment, despite having only half the number of the scans as context. This idea implies that the first subset of parameter possesses a lot of unnecessary information which the models perform just as well without. Moreover, SSIM further demonstrates this, by the fact that source scans are generated mostly flawlessly and that target scans, despite numbering quite a few less than they do in the 1<sup>st</sup> experiment, do just as well. However, it should also be said that the third experiment also fares quite well against these results, despite being dealt the worse hand and displaying clear result dissimilarities when compared to the other two experiments.



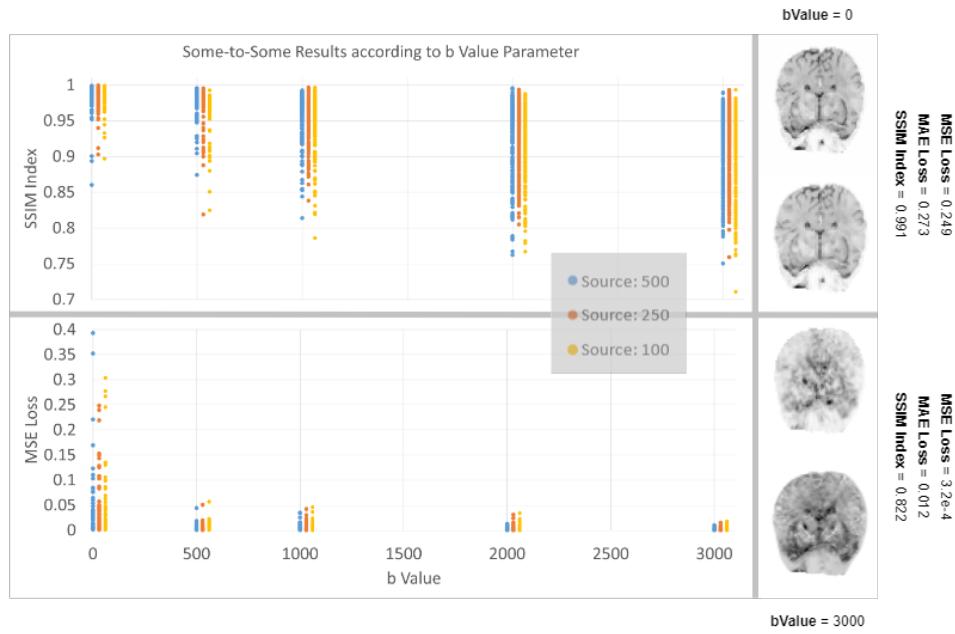
**Figure 4.1:** Boxplot of MSE Loss, MAE Loss and SSIM Index Results for all Experiments of the Some-to-Some Approach, by segregation of the generated scans for Source and Target Parameter Combinations. The boxplot includes the 1<sup>st</sup> and 3<sup>rd</sup> quartile values and the minimum and maximum values (excluding outliers), as well as the median (line inside the box) and mean values (cross-shaped marker), and finally, the amount of samples for each plot.

Figure 4.1 also allows one to take some more detailed conclusions about the overall results of each experiment and how the number of source parameters fed to the model influence the quality and accuracy of the generated target images. For example, as it pertains to the MSE values, it should be noted that percentile and median values for the target scans are similar to that of the source ones. However, by inspecting the mean values, which is a less robust measure, the existence of faulty outliers seems to be pointed out. These are, however, not detected by the structural similarity index, and to a certain extent, nor by the mean absolute error analysis, which, due to being more broad in its analysis of the images, does not consider them as being outliers, as much as it does just score them as being worse. Given the quality of all results except these outliers, it would then make sense to analyse in detail what parameter values does the model struggle with, to possibly find a connection to the existence of these outlier scans.

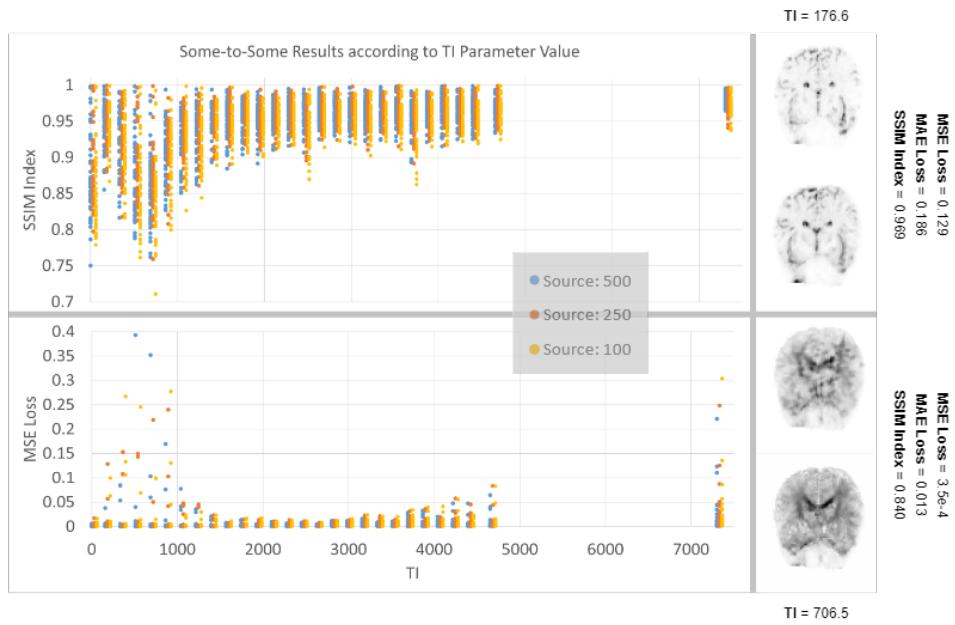
Effectively, as evidenced by figure 4.2, which correlates both MSE and SSIM index values to the scans' b value, there is an abundance of outliers for both a b value of 0 and of 3000, although both metrics seem to disagree on what to consider such an outlier. This can be a product of both combinations' distributions, in light of the fact that, in all subsets of parameters, the latter b value is much more present than the first. What is more, if the scans are analysed in further detail, it is clear that the MSE metric is considering the generated scan for a b Value of 0 as being very faulty. Despite some differences in contrast, to the naked eye, it can, however, be considered very similar to its dataset counterpart. SSIM, on the other end, seems to be correctly assessing this generation's value. On the other hand, the second image example proves the opposite: that the 2 scans, despite being quite analytically similar, as proved by the low mean squared error, look clearly different, thus agreeing to the idea that MSE might, for some of these outlier cases, be performing a poor evaluation, or at least one which is not conducive to human perspective.

Neither gradient orientation, nor TE value distributions were found to produce such an explicit outlier correlation, however TI on the other hand, did. Diagram 4.3 further expands this by suggesting that lower TI values, below 1000, might be creating an issue when it comes to the scan generation for this approach, as evidenced by their simultaneously high MSE and low SSIM results. In effect, and by analysis of all results within these values for this acquisition setting, one can find that these outliers are not, in fact, simultaneous and that there is, much like there was for the b Value analysis, a discrepancy in the 2 evaluation methods. This will mean that, and as proven by the 2 provided examples in chart 4.3, which happen to be both for low values of TI, a good SSIM outcome might fail to realise some minor contrast differences between the ground truth and the output scan, although ultimately they will look similar to one another. On the other hand, a favourable MSE output might in fact be the result of a faulty target scan present in the dataset.

In turn, from analysis of a great deal of these outliers, one can reach the conclusion that the vast majority of result outliers fall within two categories: beneficial outliers with a high MSE value but low SSIM, which stem only from contrast differences, which, although fixable, do not alter the human similarity perspective one expects this task to provide any possible user; true outliers with low MSE and high SSIM, which stem from distorted, faulty scans present in the dataset and

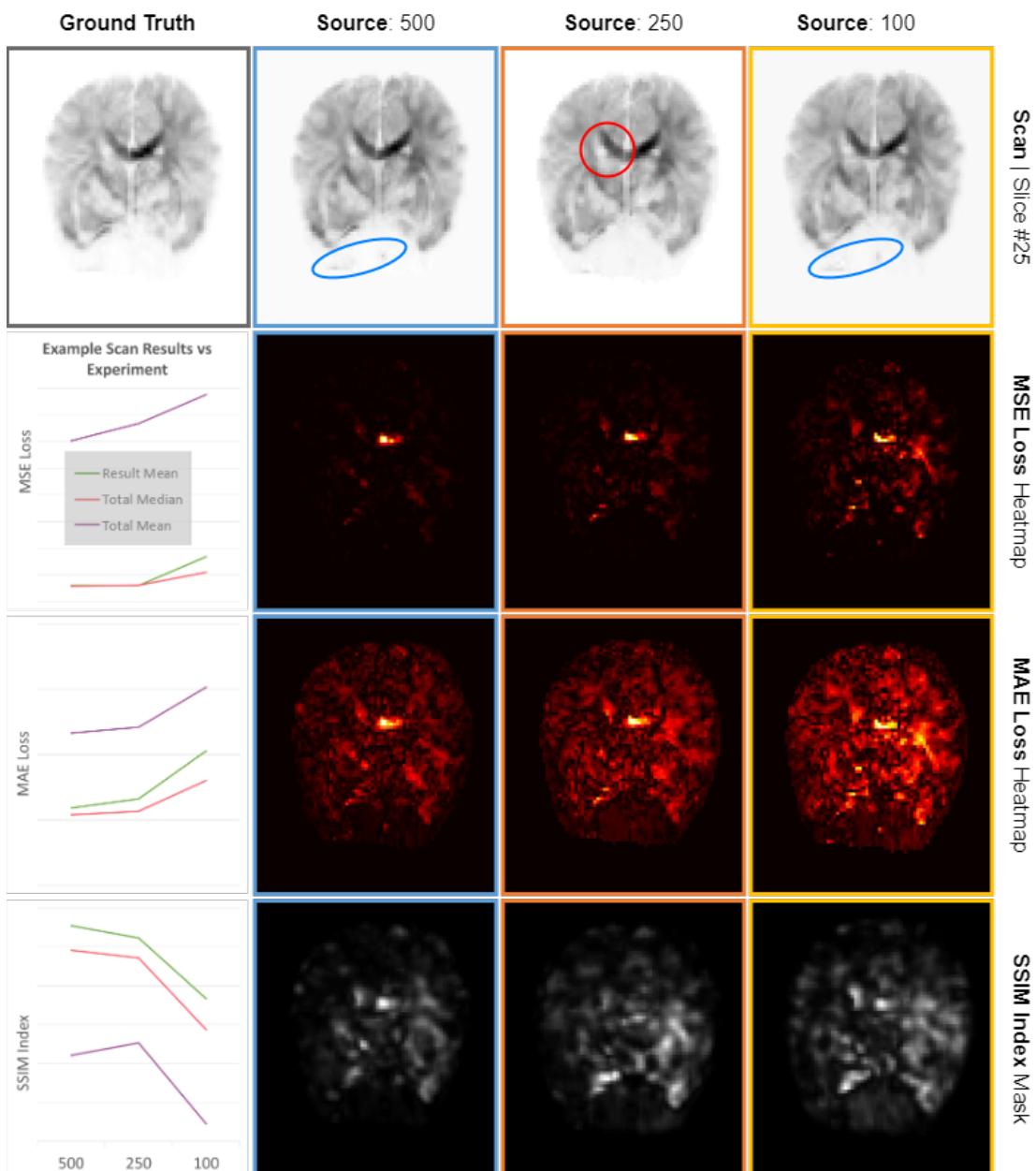


**Figure 4.2:** On the left, a distribution of MSE Loss and SSIM Index results according to the scan's b Value for all 3 experiments performed for the some-to-some approach; on the right, 2 examples with a low b Value of 0 and a high b Value of 3000 both containing the respective ground truth (above), generated (below) scans and MSE, MAE and SSIM Index results.



**Figure 4.3:** On the left, a distribution of MSE Loss and SSIM Index results according to the scan's TI Value for all 3 experiments performed for the some-to-some approach; on the right, 2 examples with low TI values of 176.6 and 706.5, both containing the respective ground truth (above), generated (below) scans and MSE, MAE and SSIM Index results.

to which the training process learns to model to quite well analytically, but which do not offer enough value to be considered from a clinical standpoint. This last set of faulty results, found mainly in combinations with lower TI and the highest b Values, make the design's task harder, but they do not, however, invalidate the quality of all the other results, for which there is agreement between the MSE and SSIM analysis, such as the ones present in figure 4.4.



**Figure 4.4:** Example results for the same parameter combination scan present in the target subset of all 3 Some-to-Some experiments, including generated scan (above), MSE and MAE loss heatmaps (middle) and SSIM index masks (below), as well as marked points of discrepancy scan's and results in comparison to the experiment's median and mean values for all metrics.

The chosen target scan and the fact that its metric results are of similar value to each experiment's median attests to the model's quality, even for lower numbers of source parameters. And although some visible discrepancies can be asserted, such as near the frontal sinus (blue marked areas) and around the left side of the choroid plexus (red marked area), most of these are visually hard to spot in the dMRI images but can be appreciated in MSE heatmaps.

Having been provided the easiest task, in which no degree of conditionality is expected of the model, it would make sense that these results be quite satisfactory and that the approach fares quite well against all the target scans it needs to map to, seeing as it has been trained with similar healthy scans for evidently equal setting combinations. It might also be said that the existence of malformed outlier scans in the dataset itself simulates a clinical environment, which despite not being corrected for by the various experiments, does not seem to be detrimental to the overall results regarding all other accurate and realistic dMRI volumes.

## 4.2 Some-to-Any Results

Much like it was discussed previously, the some-to-any approach was the only one to have both a voxel and a slice-wise implementation, so as to allow for a comparison between the novel strategy being experimented with and the well-established norm for such generative tasks, such as the one the project deals with. To all intents and purposes, these will be dealt with separately in sections 4.2.1 and 4.2.2.

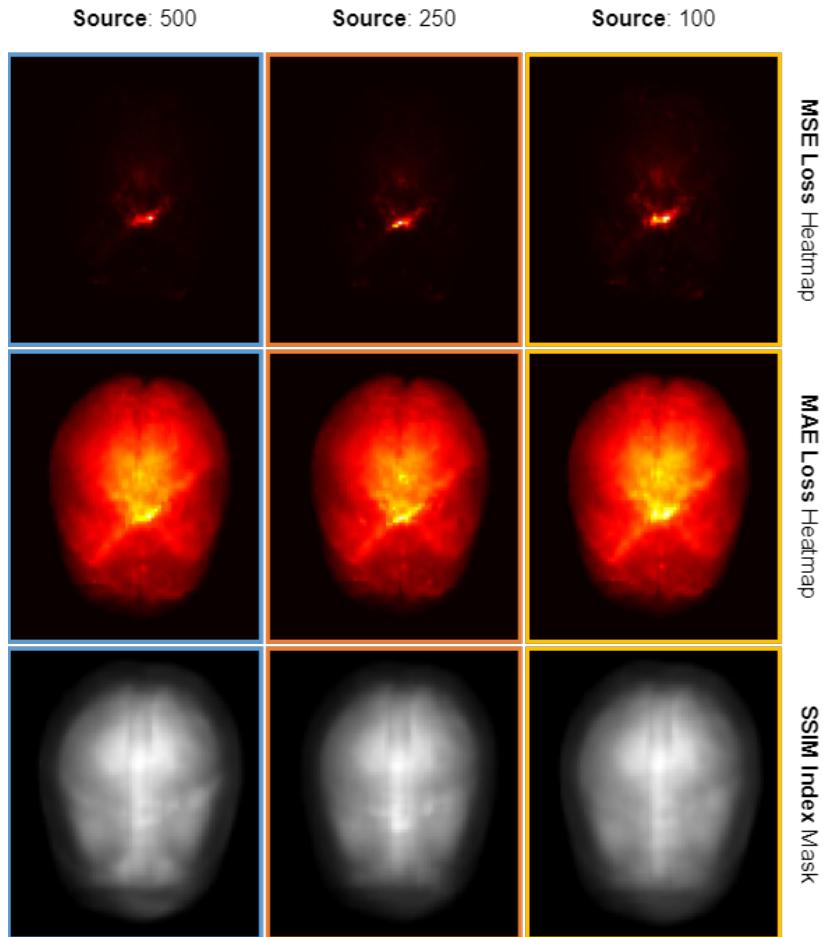
### 4.2.1 Voxel-Wise Strategy

Provided the feeding method used as a way to achieve this approach's partial conditionality state, and seeing as this is the one single approach in which the voxel-wise implementation itself can have its performance compared to a more well-established image-wise one, there is necessity to experiment with results tailored to each voxel position the model needs to map. Since only the intensity is provided and not the actual position, the designs cannot learn how different voxels interact with each other, but only how parameters do.

This will mean that rather than targetting specific positions individually, the architectures can focus on capturing how intensity value is modeled directly for each of the target parameter sets. This is evidenced strongly by image 4.5's heatmaps, seeing as, because of the axial cut of the brain being provided, the error maps, especially for the mean absolute error, display a centered-focus radial distribution, these being positions in which the model will have more intensities to map to. And although no positioning is being provided, the models learning capabilities seem to allow for an understanding that the intensities in the periphery of the brain's mask will be mapped to zero more often than the ones in its center, despite there being only slight value differences between the two. The reasoning behind this center-focused error values could also be linked to misregistration during the acquisition procedure due to ventricular pulsation, leading prediction to fail at capturing these very malleable and ever-changing structures.

However, further inquiry into the faults of this implementation require a more specific inspection of the acquired metrics and so, the same methodology was applied to a slice-per-slice comparison, as of the graphs in figure 4.6. Through it, a number of shortcomings are evidenced. Apart from peripheral slices being the ones to show a larger accuracy and precision for all metrics, which would be expectable seeing as they contain the least amount of non-null voxel intensity values, some lack-indicating peaks can be found in the area plots for some of the experimental versions of this implementation. These include slices 15 to 21 having a non-expected higher mean squared error for all experiments and slices 25 to 33 displaying low structural similarity and possibly, as seen for the baseline approach, low semblance to realistic scans. Indeed, when these specific peaks are analysed in detail, some examples of these dissimilarities can be found, these being presented to the right of the plots.

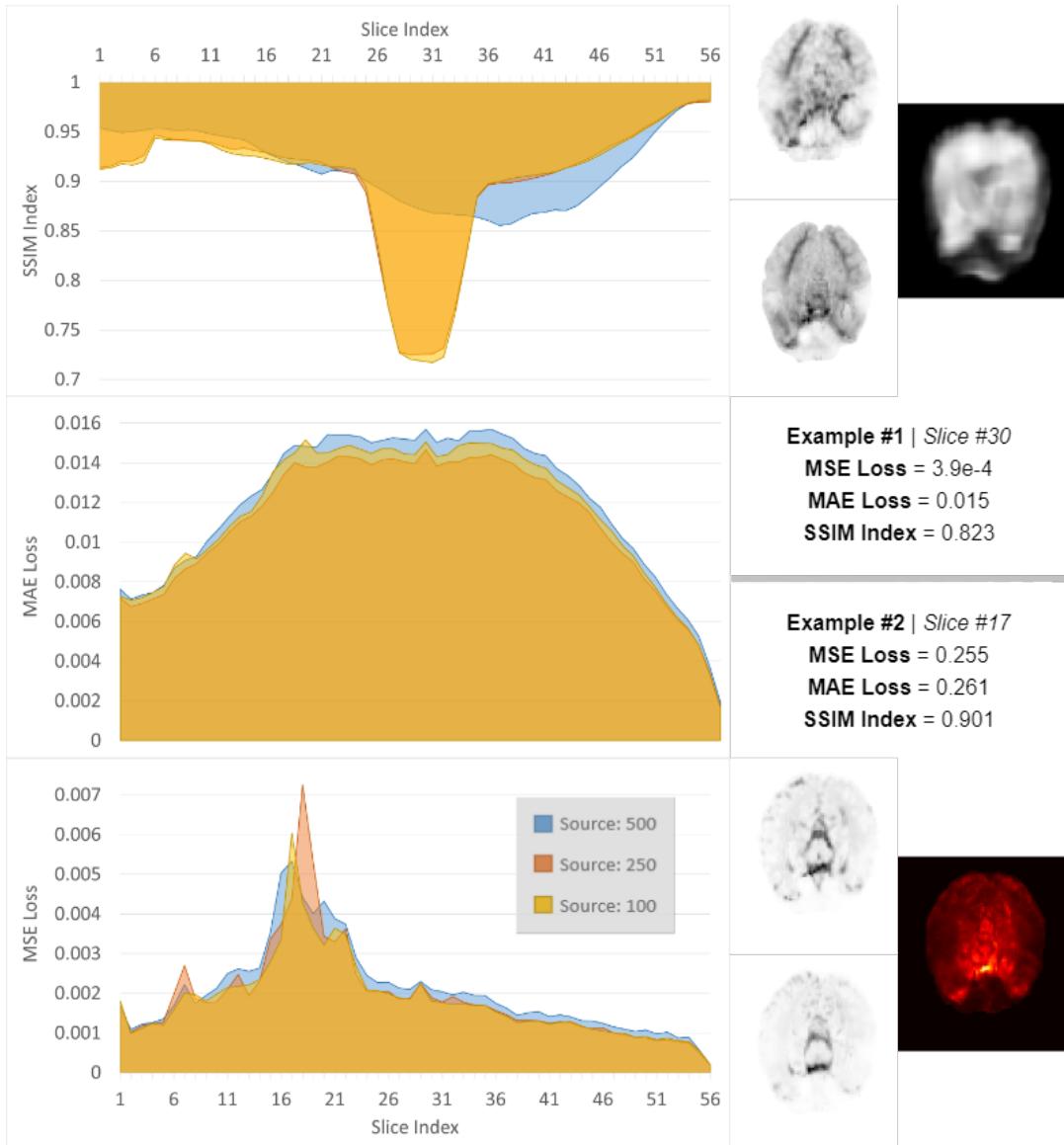
It could be said that both the same discrepancy of judgement between metrics and the model's tendency to learn the mapping of the previously mentioned faulty outlier scans present in the



**Figure 4.5:** Visual representations of mean loss values per voxel position according to both metric (horizontally) and experiment's number of source input parameter combinations (vertically). All heatmaps for the same metric were given the same scale.

dataset originally do still exist to at least the same extent as they did for the some-to-some approach. However, the examples shown suggest the existence of a broader issue not being correctly addressed by the models' learning capabilities, namely in the translation of certain structures. The first ground truth / generated scan duo, for example, displays a scan that would be, by default, a faulty parameter combination in the dataset, and has been modeled to look realistic to the target parameters but not similar to the actual outlier. Therefore, one can assume there are, in both exemplary figures and throughout the test results, structures not correctly generated and understood by the employed architecture, such as those in the frontal areas of the second generated scan.

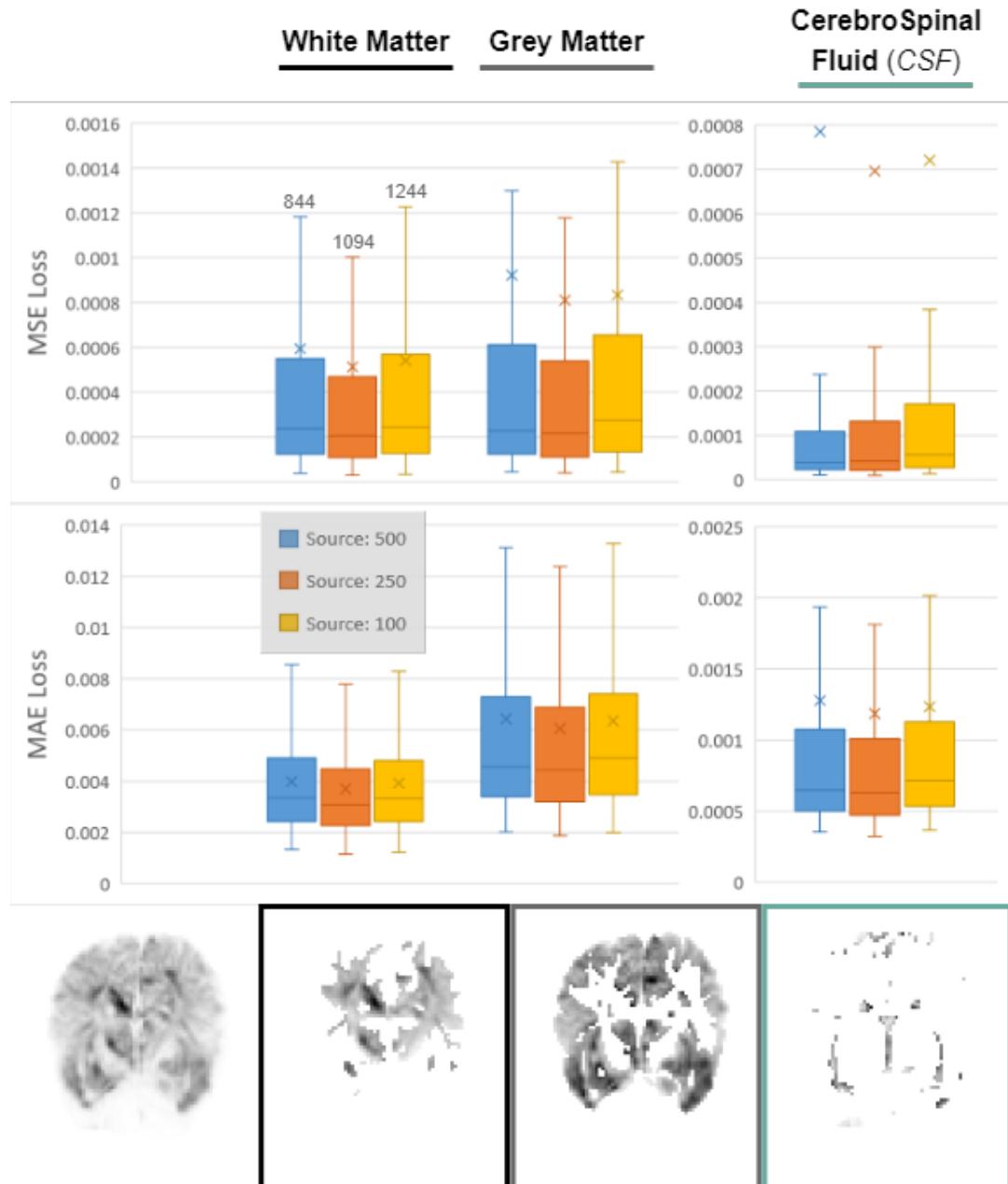
As a way to expand upon this analysis methodology, the project moved on to base the assessment on the brain tissues for which there were segmentation masks for, i.e., white and grey matter and cerebrospinal fluid (CSF). What is more, since these compose the bulk of the structures present in the brain scans captured by dMRI techniques, this analysis also allows for evaluation of the implementation's results on a target parameter level and, therefore, of the design's overall performance. Although the boxplots in 4.7 for all three tissue types include the 2 most analytical



**Figure 4.6:** On the left, area plots displaying mean evaluation metrics results versus volume slice index for all three experiments performed. The idea of using area plots and not simple line plots is aimed at better understanding of differences between experiments. To the right, two examples containing both the ground truth (above), the generated scan (below) and the heatmap for structural similarity and mean squared error respectively, as well as metric values and slice index. These examples are, in order, a low SSIM example and a high MSE one.

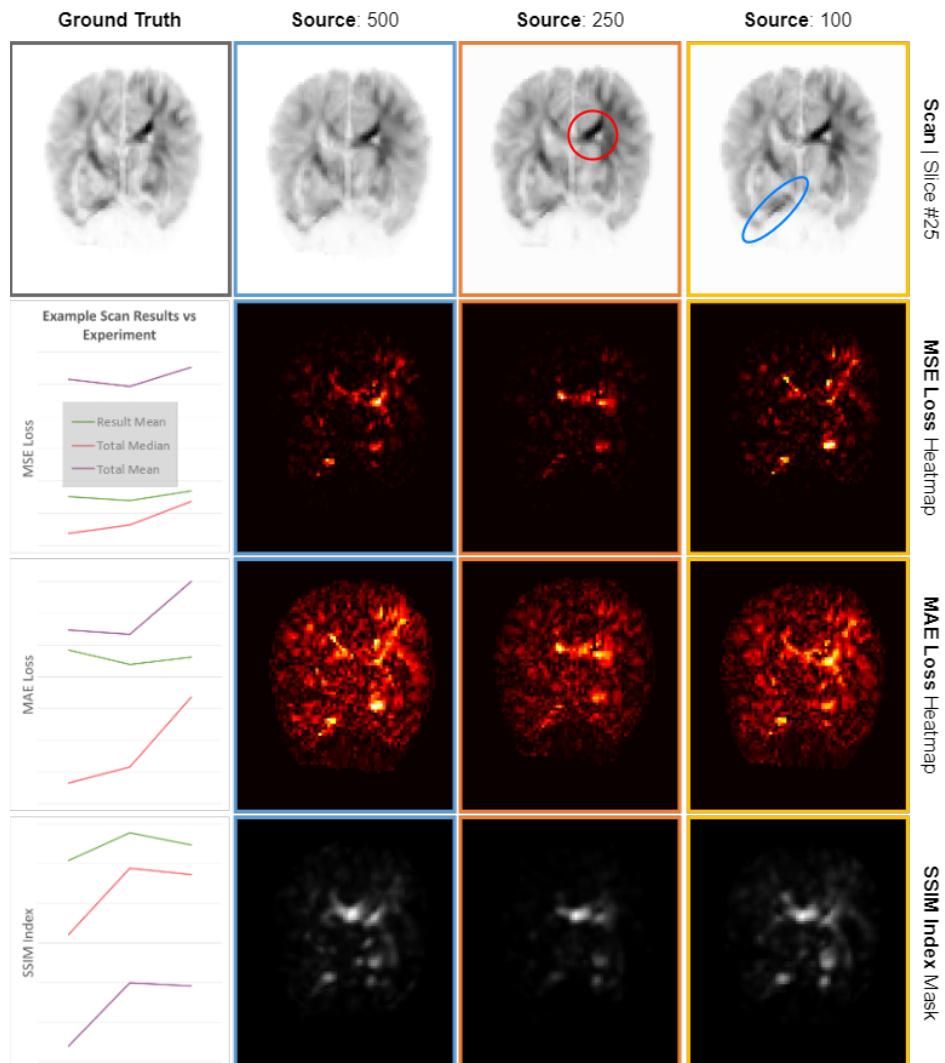
metrics, they do not refer back to the structural similarity results for this testing approach, due to the fact that the used masks, especially the white matter and CSF ones, providing very fragmented and thin image sections, which do not pair well with this algorithm's moving window technique. It should also be noted that any and all results for the cerebrospinal fluid tissue are destined to have lower squared and absolute error values due to the lower amount of voxels pertaining to this tissue. This does not exclude, however, the extremely high mean MSE values displayed for this tissue which suggest either that this tissue is the source of outlier scans or that, despite the median results, there are certain targets this implementation struggles with. This is further expanded

upon by the discrepancies between the MSE boxes for both the white and grey matter tissues. In other words, despite clear differences between the number of voxels considered within the white matter and the grey matter masks, there is little deviation between the MSE boxes between these two tissues, implying that white matter is being much inferiorly translated and understood by all experiments for this design.



**Figure 4.7:** Boxplot of MSE and MAE losses for all experiments of the Some-to-Any voxel-wise approach, by segregation of the different tissue segmentations (below) performed on the target generated scans for white matter, grey matter and cerebrospinal fluid. The boxplot includes the 1<sup>st</sup> and 3<sup>rd</sup> quartile values and the minimum and maximum values (excluding outliers), as well as the median (line inside the box) and mean values (cross-shaped marker), and finally, the amount of samples for each plot.

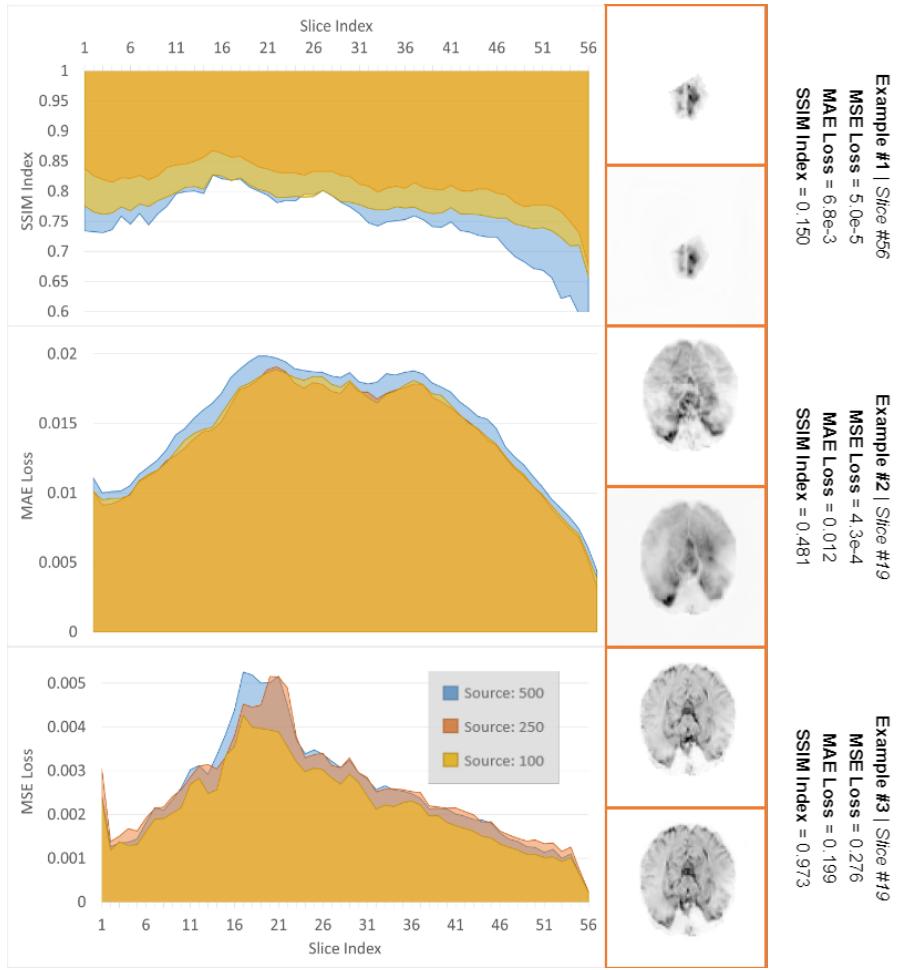
Having said all this, and considering all the shown graphs, including the voxel heatmaps and the differences between amounts of input scans and parameter combinations, the one experiment which shows the most promise is the second one, cementing the idea that the 250 parameter subset is, much like it was for the some-to-some approach, the one most representative of the overall landscape of the acquisition settings. This accuracy and precision is asserted further by the example provided in figure 4.8 and the fact that its results can be found, for the most part, in between the mean and median values for the target generated scans. As these display, apart from minor structure differences, for example near the choroid plexus' right side and trigeminal nerve (red marked area) and along the temporal muscle (blue marked area), the model seems to be able to generate accurate and visually very similar scans for the great majority of target parameter combinations it is tasked to map to.



**Figure 4.8:** Example results for the same parameter combination scan present in the target subset of all 3 Some-to-Any Voxel-Wise experiments, including generated scan (above), MSE and MAE loss heatmaps (middle) and SSIM index masks (below), as well as marked points of discrepancy scan's and results in comparison to the experiment's median and mean values for all metrics.

### 4.2.2 Slice-Wise Strategy

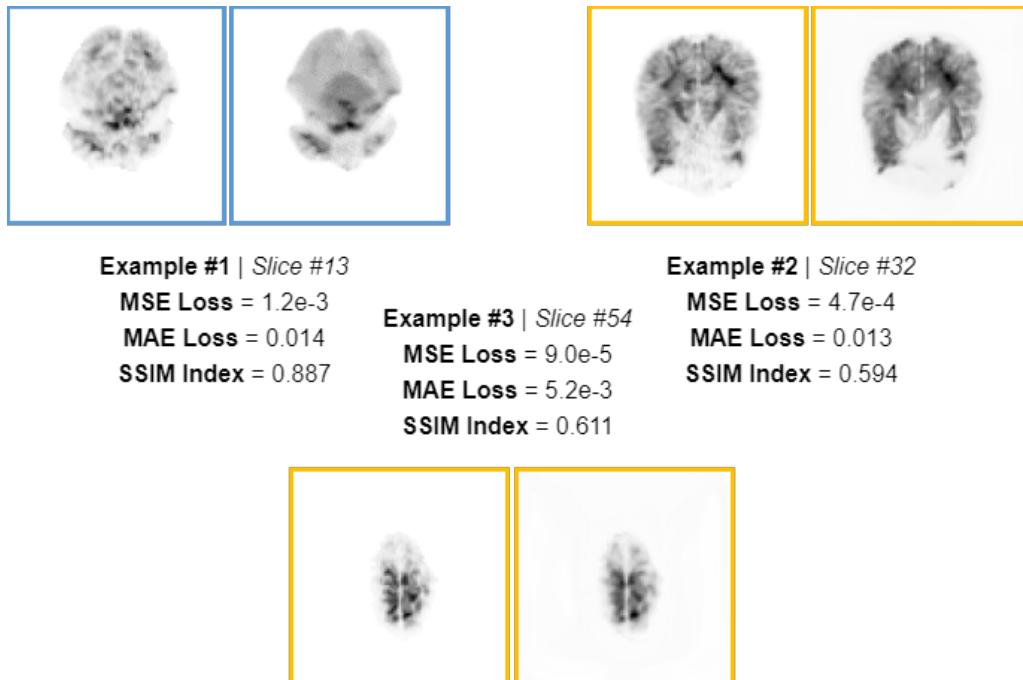
Similarly to this approach's voxel-wise implementation, so would it make sense to analyse results in a slice-wise way, seeing as this is the way this model is receiving its input data. Once more, no index of the slice is being fed to the architecture at any point in the pipeline, forcing it to learn strictly from the images' organization when distinguishing different slices from one another. As such, the model is bound to have slices it deals with better than it does other, much like the area plots in figure 4.9 prove. Through analysis of what results display, it seems clear that, although lower / better than they are for the voxel-wise implementation of this task, both MSE and MAE behave quite similarly to what they did in 4.6. Structural similarity, on the other hand, creates very counter-intuitive results by assuming worse results for peripheral slices which, much like it has been seen before, contain a lot less non-null voxel intensities than intermediate ones, this fact being even truer due to the usage of zero-padding on the volumes.



**Figure 4.9:** On the left, area plots displaying mean evaluation metrics results versus volume slice index for all three experiments performed. The idea of using area plots and not simple line plots is aimed at better understanding of differences between experiments. To the right, three examples containing both the ground truth (above) and generated scan (below), as well as metric values and slice index. These examples are, in order, two low SSIM examples and a high MSE one.

Indeed, by inclusion of some examples, it becomes apparent that the analytical metrics tend to disagree with the structural similarity index and that this might be causing the seen outlier values. In particular, the first and single peripheral slice example displays an extremely low SSIM result, but completely adequate MSE and MAE ones, despite the generated scan looking, at least visually, quite similar to the original. In other words, despite the existence of clear structural faults in between the two images, the attributed score seems to be inadequate and too strict, given especially the scale of such missing details. Furthermore, the third example, of an intermediate slice expands upon this given the fact that, despite minor dissimilarities accounted for by SSIM, the squared error is much lower than one would normally think it to be. And so, these two examples display a clear antithesis, seeing as both generated scans display clear and visible resemblances to their corresponding target ground truth volumes, although this is only captured in an expected way by the latter example.

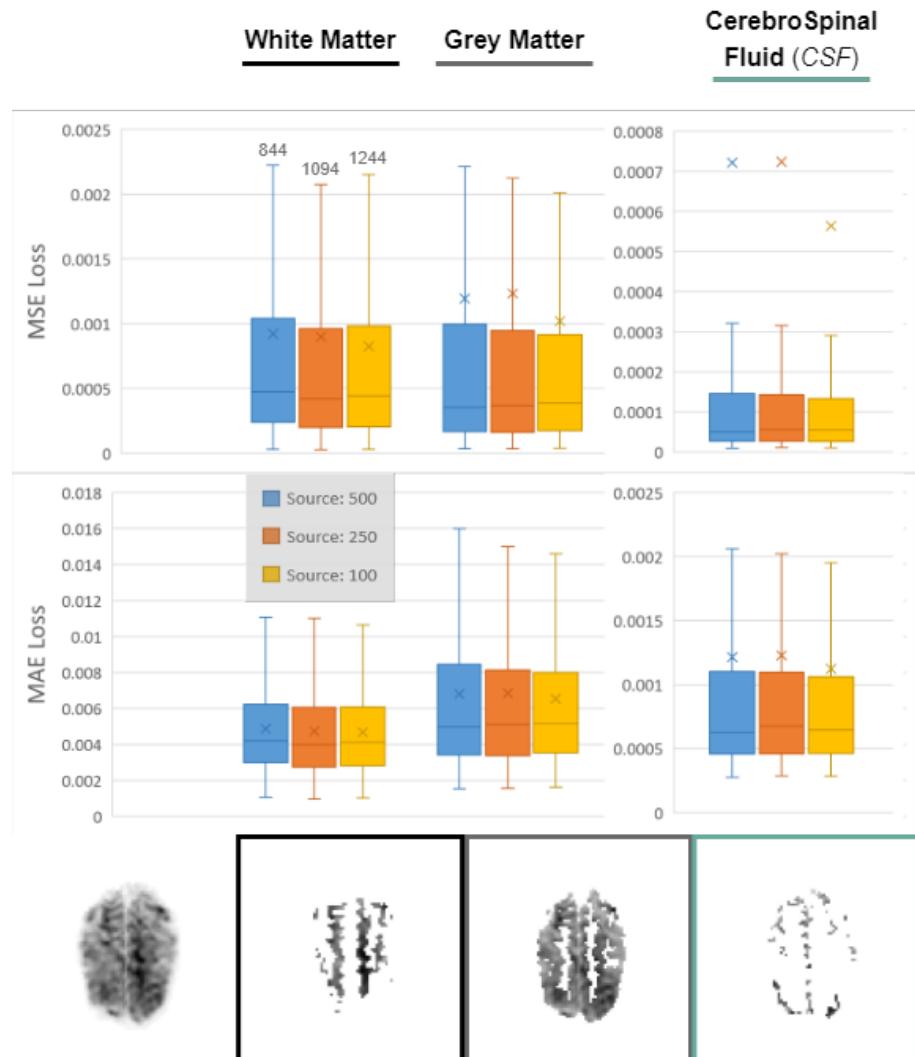
The second example additionally implies that this inaccuracy from the SSIM index metric might be reserved to the peripheral slices and that this might be caused by the scale of structures being analysed in these. This is because the produced scan defies the faulty and corrupted nature of the original and can be considered a realistic-looking scan when compared to other similar parameter amalgamations scans for the same slice index. This defiance, however, is not and should not be excused by SSIM's evaluation, which this example accurately complies with, as proven by the low score it's been given. What is more, if this matter is looked further into, a number of examples for several different slice indexes can be found to produce this effect, some of which are displayed in the [4.10](#) compilation.



**Figure 4.10:** Three examples of target parameter scans, including original and generated scans, each marked with the model experiment they belong according to the pre-established color scheme. Each example's metrics results and slice index are also included as labels.

All three examples display some degree of blurring and aliasing effect as a result of the convolutional method being applied. This is most notable for the first example, in which the blurring can be seen across all the different tissues and throughout the whole generated scan, erasing a number of different complex structures and reducing them to a vague and indistinct mesh. However, this is also the one example best scored by the SSIM metric among all three. This is because, as one can assume, structural similarity is having trouble understanding the different patterns being created by the convolutional layers inside this approach's design. In other words, because the generated image maintains, to a certain degree, all the structures the original has, although in a barely recognisable manner, these are being picked up on by SSIM. The remaining two examples, on the other hand, display a lot of missing structures and, in some cases, the union of different structures by bridges created by convolutional means, which the metric recognises as being harshly uncharacteristic when compared to the original. The scale of these is also quite a lot lower than that of the blurring effect happening in the first example, leading one to believe that SSIM is, perhaps, looking at too broad a picture. Since these effects are, to some extent, concentrated on some areas of the brain scan, such as in the white matter tissue for the second example, an analysis of results according to this segmentation is, similarly to what has been done for this approach's voxel-wise implementation, in order.

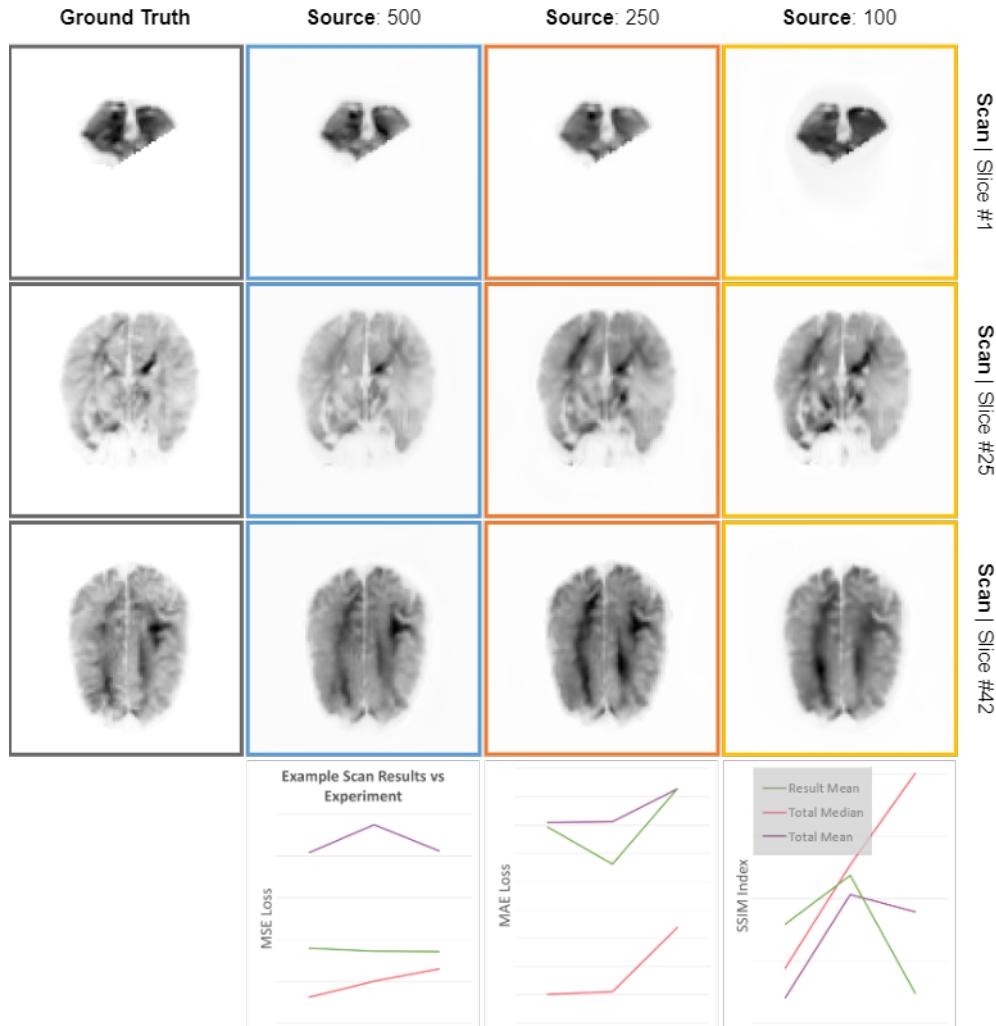
Indeed, by comparison of the boxplots and loss scale axis in figure 4.11 to the ones of the voxel-wise implementation in graph 4.7, one can tell that all results for white and grey matter tissue are substantially worse. Due to the fact that these two compose the over-arching majority of the voxels for all slices, including peripheral ones much like the one shown in the example beneath the boxplots, a conclusion can be reached that this implementation's performance is hindered by a great amount of blurring across all the target parameter combinations being mapped to.



**Figure 4.11:** Boxplot of MSE and MAE losses for all experiments of the Some-to-Any slice-wise approach, by segregation of the different tissue segmentations (below) performed on the target generated scans for white matter, grey matter and cerebrospinal fluid. The boxplot includes the 1<sup>st</sup> and 3<sup>rd</sup> quartile values and the minimum and maximum values (excluding outliers), as well as the median (line inside the box) and mean values (cross-shaped marker), and finally, the amount of samples for each plot.

This effect can be felt a lot when analysing all the test patient's results, however it should not demean this implementation's successful results, one of which being included in 4.12. This example was, once again, sourced by comparing its mean results to the median and mean values for all target combination scans, and three representative slices were included for visualisation purposes. It is important to notice that different slices will have different metric results, which are not being displayed in the figure, seeing as the most crucial point of interest was measured around the mean volume results for all slices and how this value correlates to the expected for each experiment of this approach. This examples, together with all the analysis methods employed should prove both this method's applicability and quality in performing the given task, as well as the existence of clear lacking points in such performance. Furthermore, by side-to-side comparisons between

all three experiments, it could be said that the second experiment's source subset is still the one producing the best results, although this might be hard to see when examining the slice-wise study and figure 4.9, but is much clearer in the boxplots provided for the brain tissue segregation one.



**Figure 4.12:** Example results for three chosen slices of the same parameter combination scan present in the target subset of all 3 Some-to-Any slice-wise experiments, including ground truth (left) and generated scan for each experiment, as well as mean results in comparison to the experiment's median and mean values for all metrics.

### 4.3 Any-to-Any Results

Being the last approach used by the project's pipeline, the any-to-any procedures faced the hurdle of having an extensive amount of results, ones which could not be analysed only from the perspective of the parameter combination being mapped to, but also the one being mapped from. For this reason, it was computationally unfeasible to either train, validate or test the model utilising all possible source-target combinations, which, given the task's premise, would number nearly 2 million 3D volumes, all of which would have to be analysed through the different evaluation metrics used for all other approaches.

As such, only a select few source-target examples were chosen for evaluation of the design's performance for both the first mentioned experiment, which would use a true version of an any-to-any parameter approach, utilising as source and target scans all 1344 available volumes, and the second experiment, which attempts to tone this down by using only 500 source and 844 target combinations, thus significantly reducing the possible amount of source-target compositions to slightly over 400 000. In order for this analysis to be as fair as possible, a strive was made to utilise the same amount of source-target outputs for both experiments, resulting in the generation of a random 35 possible target scans for the first experiment and 100 for the second. Each was generated and had its results measured by providing the model all single source scans available for the experiment at hand. This would then mean that, in total, 47040 (35 Targets x 1344 Sources) 3D volumes were generated by the model in the first study and 50000 (100 Targets x 500 Sources) in the second. Results for both endeavours can be ascertained from the boxplots in figure 4.13.

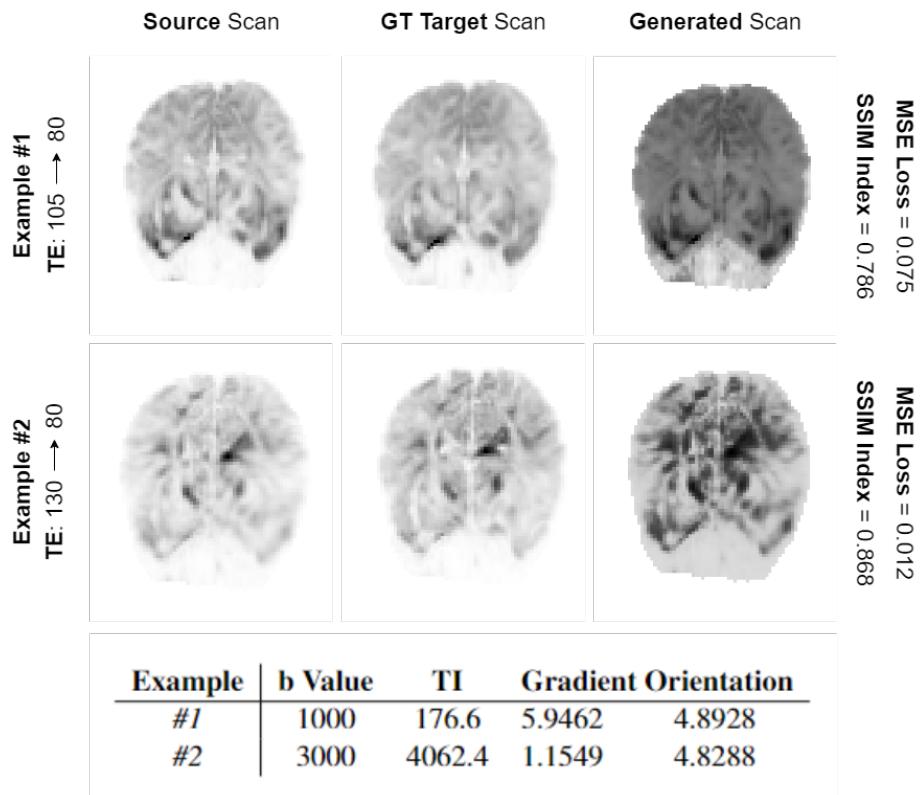


**Figure 4.13:** Boxplot of MSE loss and SSIM index results for both experiments of the Any-to-Any Approach. The boxplot includes the 1<sup>st</sup> and 3<sup>rd</sup> quartile values and the minimum and maximum values (excluding outliers), as well as the median (line inside the box) and mean values (cross-shaped marker).

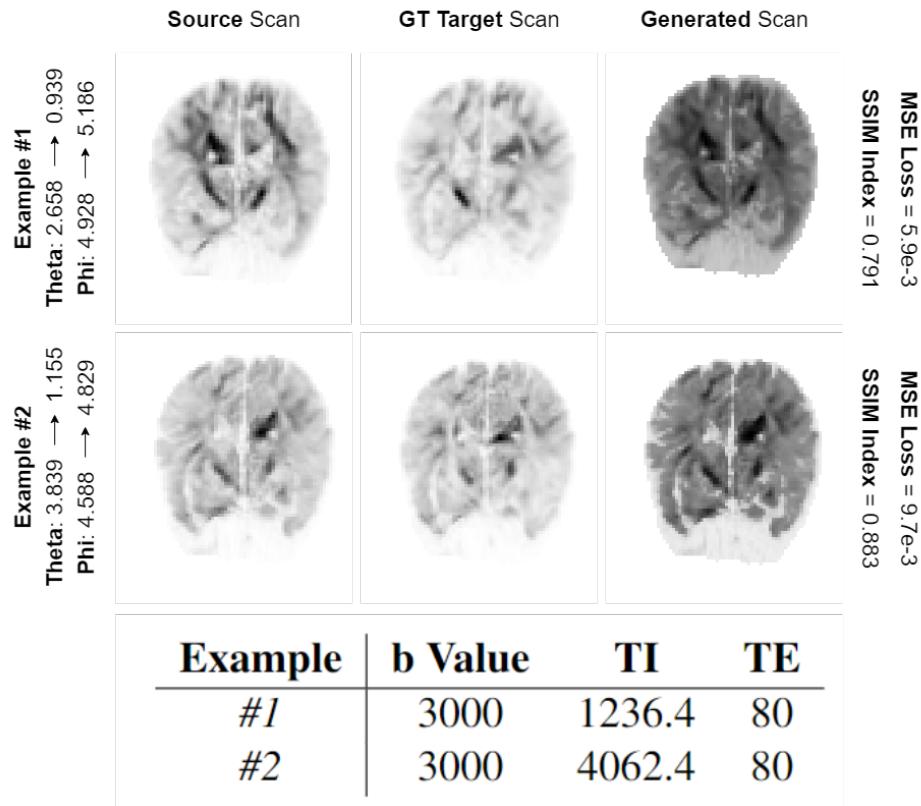
In effect, the first experiment, despite more free in its analysis of the provided source-target combinations, does include less target scans to map to. This is something strongly denoted by the distribution of its results, which suggest that not all source volumes will be conducive to generate the chosen targets with a high degree of fidelity, and that while some might be suitable due to

closer proximity when it comes to parameter values, some others might present too much of a step for the model to make, thus the larger range overall, especially for the structural similarity index. This experiment does, however, still compete with the second, more constrained study, whose results indicate the presence of a larger number of outliers, on account of how the mean value for the mean squared error distances itself from the main core of the results, which would be a product of the larger number of targets tested, which would increase the number of previously referenced harder-to-model-from source mixtures of parameters.

Given this result landscape, and with the intent of evaluating this model's learning capabilities and its grasp on each acquisition setting's value and how these affect the overall aspect of the dMRI scan for all voxel positions, an analysis was made of the design's reaction upon the mutation of a single parameter's value. In other words, by analysing the set of target parameters used in the test set, some of the source scans used in the generation of results could be identified, namely ones in which all parameters' values are the same, except for one, consequently allowing for the evaluation of the design's performance in capturing single label value alterations. Due to the nature of dMRI acquisition settings, this was an experiment only possible the echo time (TE) and gradient orientation segmented labels, of which outcomes can be found, respectively, in images 4.14 and 4.15.



**Figure 4.14:** Examples for 2 source-target combinations in which the only mutation between the 2 acquisition settings amalgamations is the TE value, including the source scan, the target scan and the generated image, as well as the values for the evaluation metrics for this generation. Below, a table of all other parameters which suffer no change between the source index and the target one, for both exemplary sets of imagery.

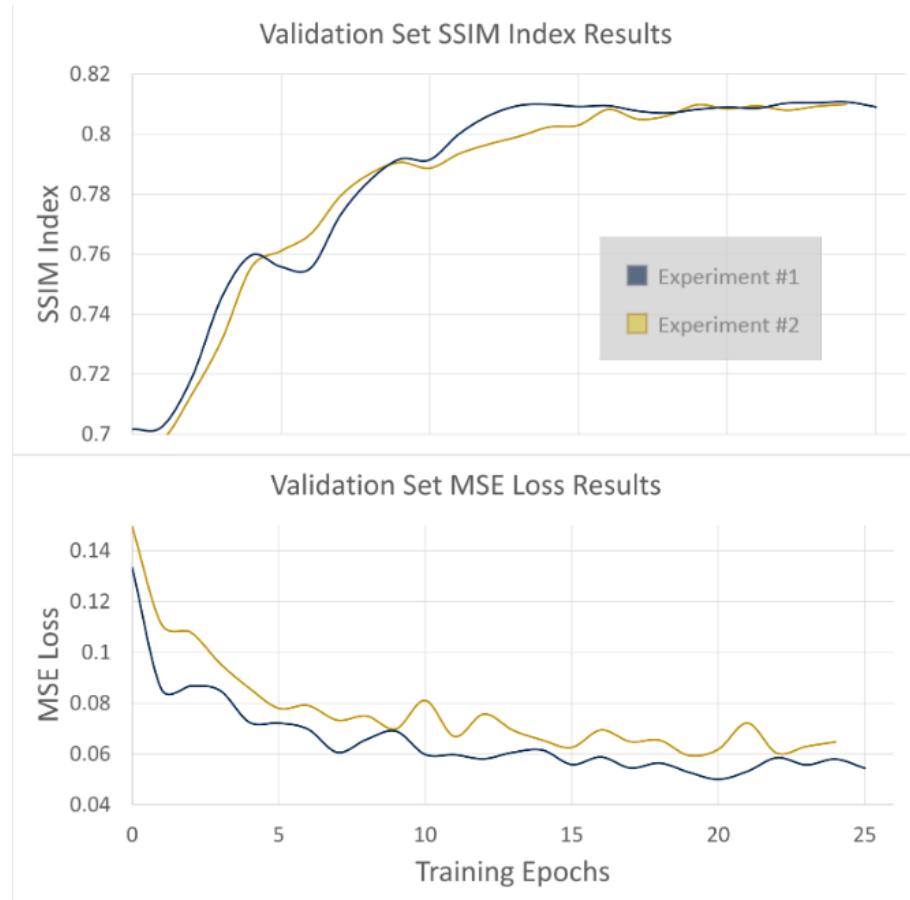


**Figure 4.15:** Examples for 2 source-target combinations in which the only mutation between the 2 acquisition settings amalgamations are the gradient orientation values, including the source scan, the target scan and the generated image, as well as the values for the evaluation metrics for this generation. Below, a table of all other parameters which suffer no change between the source index and the target one, for both exemplary sets of imagery.

Given the observed results of this experimentation method, it becomes apparent that the model struggles in capturing the intricacies present when attempting to map from any source to any target parameter, even more so than its numeric results seem to betray. Expressed differently, the outcomes shown hint to an inability when it comes to the model's learning capabilities of source-target interactions, whatever they may be, and even in the case of simple parameter changes, leading one to assume that the model is incapable of overcoming the source parameter's influence on the output, seeing as most generated scans look, to a degree, very similar to their source input. Even though this might be the correct formula to use in some specific cases, figures suggest that little to no feature map has been correctly ascertained in between these parameters' values and the intensities of all single voxels for all different slices inside the image. As such, the modelling process will be very much influenced by the source scan's appearance, which is a further proved by the fact that certain distortions in the source scan are often translated into the final project with a similar pattern to the original.

The displayed lack of accuracy can only be attributed to both the task's intricacy and the model's lack of complexity and depth. And so, while trying to capture all different possible interactions provided to it, the design ultimately fails and learns a very broad mixture of parameters, which when applied to a great number of situations and source-target combinations, results in the

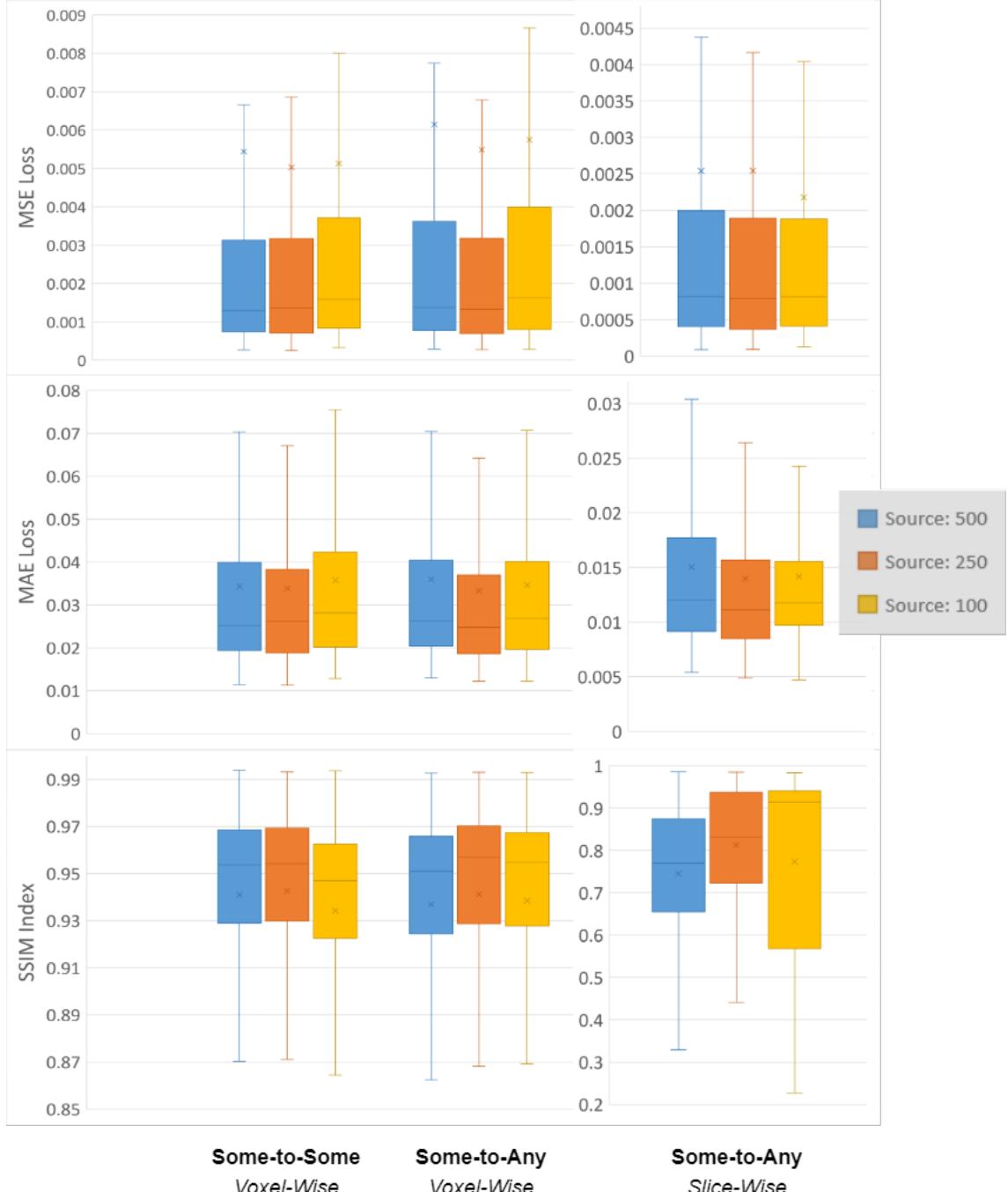
either the darkening or enlightening of all voxels of the source scan, seeing as its ability to determine which voxels to darken and which to lighten is hindered by the amount of information it needs to cope with, this being how to change every single voxel's intensity in function of the value of 5 different and completely separate continuous labels. This can be traced back to the design for this approach, and its lack of depth needed to capture all the information described, which can be confirmed by the overall stagnation of its results for the validation patient along the number of training epochs in figure 4.16, even considering this validation was done using the same combinations the architecture was tasked to model during training. Although not able to compete with its less conditional counterparts, it is an approach which still shows the promise to, if such issues were to be fixed, be able to improve its capacity to not only correctly mimic specific source-target conversions, but also do so for an exponentially greater number of pairs.



**Figure 4.16:** Mean validation set results for both metrics used as back-propagation and evaluation along the number of training epochs for both experiments ran using this approach.

## 4.4 Overall Comparison

At last, a comparison of all the implementations was also performed. In the interest of fairness, any and all analysis is done firstly between approaches for the same experiment and later considering all the different approaches and sub-tasks, as can be seen in Figure 4.17.



**Figure 4.17:** Boxplot of MSE loss, MAE loss and SSIM index Results for all experiments of the Some-to-Some and Some-to-Any Approaches, regarding the target parameters being mapped to. The boxplot includes the 1<sup>st</sup> and 3<sup>rd</sup> quartile values and the minimum and maximum values (excluding outliers), as well as the median (line inside the box) and mean values (cross-shaped marker), and finally, the amount of samples for each plot.

Having said this, results shown reach various conclusions:

- When comparing the voxel-wise implementations to the image-wise one, it is notable that there is a clear distinction in the scale of results, proving the analytical metrics to be much lower and SSIM much broader for the latter. This, in turn, has been stated through different methods of inspection, to be undesirable and caused by a blurring effect and a number of distortions happening upon the scan generation process. And so, even though analytical results are better than for its sub-task's voxel-wise counterpart, in depth analysis and the case built around how SSIM behaves has proven this to be deceptive and not conducive to the idea of better generated scans.
- When comparing the two voxel-wise methodologies, each in the context of their own task, the some-to-any's successful performance should be clear to see, by the fact that it provides similar results for every metrics, with the added challenge of having partial output-sided conditionality.
- Any-to-Any results were not included, due to the way its experiments are built, however they have been shown to be significantly worse when it comes to MSE loss and about similar to the slice-wise method in terms of SSIM index.
- Once again, the second experiment's 250-sized subset of source parameters has shown, for all different approaches, to be a good representative of all scans and an ideal middle ground between good results and having a low amount of input scans. It is also true, however, that the third experiment's setup is quite beneficial performance-wise, seeing as it allows for a higher number of target parameter combinations and a larger learning pool, which most models seem to take advantage of as well, although its task is significantly harder.

As demonstrated by multiple works in the past such as the one in [3], state-of-the-art unconditional experiments on the matter of MRI reconstruction have been the most successful utilising Deep Learning methods, most of which based on CNN architectures with a variety of different implementations. However, it is also demonstrably true that a CNN backbone will, despite providing a lower complexity model of faster inference than more complex autoencoder architectures, generally speaking, have a more limited receptive field and trouble capturing an insightful overview and a grasp on the underlying patterns in the image pixel, or in this case, voxel long-range interactions.

Results obtained prove that the issue at hand, due to the nature and intricacy of both the structures and the task of conditionality itself, might therefore be a point of struggle to such a deep learning model architecture. And so, while it is justifiable that an image-wise method is used, seeing as it is an industry-standard and also allows for a larger inherent spatial knowledge within the samples, this project's problem seems to resonate better with a voxel-wise solution. This, in turn, does forbid designs for this approach the usage of any complex or convolutional layers, which would, given the size and nature of the training samples composed of voxel groupings being fed, most likely lead to overfitting. Not only would this be the case but it would double the hardship faced by the model when trying to distinguish similar parameter combinations from one another

and the patterns each tend to contain, without overfitting to the most common patterns. On the other hand, for a convolutional image-wise solution to succeed it would need to take both the spatial and parameter information into account, e.g. by working in the SE(3) space.

#### 4.4.1 Test Parameter Subset Results

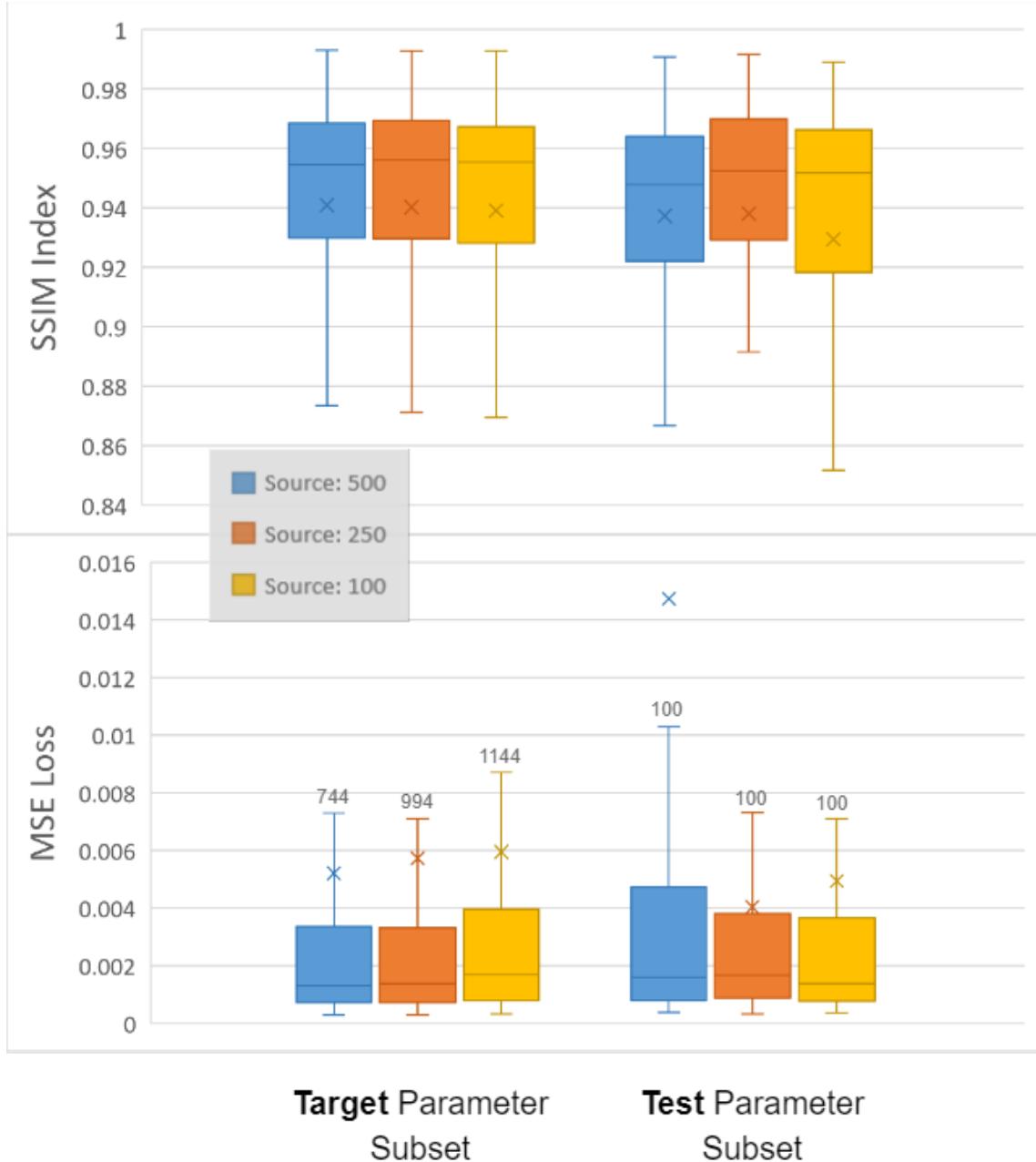
These results encompass all 3 experiments designed to use the some-to-any approach's voxel-wise method for which there were a subset of 100 parameter combinations not included in either training or validation, or in other words, combinations the model has not seen before the testing phase and which it needs to map to not ever having had access to the scans they normally produce. It requires, therefore, that the model possess intrinsic knowledge on how to map all the source voxel intensities provided, to the target by assuming and learning from the changes it has seen between these source combinations and other target ones with similar values as the ones in this test subset.

The attached boxplots depict the results for these test subsets when compared to the same design's output metrics for the target subset of parameter combinations it was trained with. This assessment was performed for all three experiments described previously, in which the number of combinations included in the source parameter subset the architecture was trained with is increasingly reduced. As figure 4.18 implies, and according to what would be expectable, both metrics' outputs are generally, worse for the test subset of scans than they appear to be for the target volumes which the model has had access to previously. However, in the grand scheme of these metrics, this difference is quite negligible, with exception to the first experiment of 500 input scans, which does clearly display a large amount of outliers, as proven by the large structural similarity range and off-course MSE mean value.

This is also a product of the smaller number of samples used in the measurements for the test subset, seeing as this only includes 100 parameters, a lot less than the model has had as target practice during training. However, and as a result of this, it appears that, as the number of inputs provided to the model decreases, so does the amount of fixed interactions it needs not learn, forcing it on a harder mapping curve, but also allowing for the model to have more insight into how to map to more and more parameter combinations. This is beneficial when trying to include a test subset of parameter, since it proves that designs are able to interpret all 5 label values independently and how their combination ultimately affects a voxel's intensity based on the structures on the fixed input scans. Moreover, it does so in a manner that can compete with the same design's results for the target set of combinations it was trained with.

And so, despite the existence of outlier, faulty scans in the test scans, much like there was for the other subsets, the model actively corrects for these distortions, never having seen the ground truth for such combinations. It does, therefore, provide a take on what these scans would look like if they were not corrupted, unbiased by the errors that took place in the acquisition and that led to distortion in the first place. This enforces that, rather than treating these combinations as monolithic and discrete, the design learns to understand how their combined values affect the scan

at a voxel level. It then applies this knowledge to generate either accurate synthetic scans or, when the GT is not of much use, realistic and non-distorted looking ones.



**Figure 4.18:** Boxplot of MSE loss and SSIM index results for all experiments of the Some-to-Any Approach, by segregation of the generated scans for target and test Parameter Combinations. The boxplot includes the 1<sup>st</sup> and 3<sup>rd</sup> quartile values and the minimum and maximum values (excluding outliers), as well as the median (line inside the box) and mean values (cross-shaped marker), and finally, the amount of samples for each plot.

# Chapter 5

## Conclusions and Future Work

### 5.1 Summary and Conclusions

The present work has brought a comprehensive and broad study on different approaches regarding conditional dMRI image generation. This objective plays a crucial role in decreasing the length and cost incurred by health providers when performing acquisition procedures for this imaging modality. Thus it creates a possibility for faster diagnosis and overall treatment of conditions such as brain tumours or neurodegenerative diseases. With this in mind, this dissertation aimed to push the boundaries of conditionality by decreasing the amount of scans and acquisition setting combinations which procedures would need to acquire, and expanding those to which models could map to with accuracy and precision.

The reviewed literature on the fields of conditional image generation and medical imaging-applied generative issues proved to be very enlightening regarding the capabilities of the currently employed techniques. Moreover, the previous works on this specific field and dataset served as a great inspiration and provided this project three different-sized ideal parameter subsets, which have been extensively used throughout all experiments as a means to lose the least possible amount of information while reducing the amount of input-sided information.

When comparing the results of the first two approaches, the Some-to-Some and Some-to-Any, which utilise the same amount of context to generate either a full dataset or a specific target scan, it is evident that while the first sub-task has an easier job by default, the latter's performance is comparable, especially for the voxel-wise implementation. This reveals that given the same amount of information and input scans, the designs need only leverage the contrast, luminance and which structures the target combination of acquisition settings emphasises to be able to accurately generate a fake image. This is asserted by the fact that, much like is the case for image-to-image synthesis issues, all scans contain the same information and structures, although captured in a different lens. This will mean that both approaches will only ever need to learn how to map from the source scans to the one chosen target and what that implies when it comes to which structures get the spotlight. Furthermore, the different experiments on both approaches demonstrate the

subset's efficiency in capturing the most relevant information contained within the full dataset, as well as the models' good understanding of what different variation in each setting entails.

As for the two implementations of the Some-to-Any approach, obtained results verify that the voxel-wise strategy is, in the instance of medical imaging data to the likes of dMRI 3D volumes, superior when compared to the slice-wise one. The reasoning behind this relates to the scale of data being provided. On the one hand, the voxel-wise implementation can focus on structures on a minor scale, disregarding the need to identify them firstly. The slice-wise implementation, on the other hand, due to having to learn about 500 full 2D, must firstly identify and create boundaries for all structures at a time, which, despite providing greater spatial insight, also ends up hurting its performance. This is the reason for both the blurriness effect and the discrepancies between the analytical and the structural similarity index metrics, which, as proven before, need to be in accordance for results to be truly considered accurate and precise. This is especially notable when juxtaposing the two performances according to the brain tissue generations.

Coincidentally, it is for the opposite of this line of reasoning that the faulty results of the Any-to-Any approach's implementation occur. These can be traced back to the models being unable to interpret and identify all the cerebral structures that dMRI can capture from a single input scan in which only a few are emphasised. As such, it is expectable that these designs struggle with any-to-any mapping when the scans for the two setting combinations display and emphasise completely different brain portions. Expressed differently, while a greater output-sided degree of freedom will contribute to the project's goal of creating a virtual dMRI scanner, bottlenecking information on the input side has proven to be demonstrably disadvantageous.

Finally, test subset results helped corroborate the validity of the proposed tools. By testing the Some-to-Any approach's voxel-wise implementation, using a target set of parameter combinations which have never been seen, a clear understanding of these setting's values is attested for. This indicates that models have an intrinsic understanding of how the parameter themselves affect contrast, luminance, and structures on a voxel level, which they can apply to any new combination that is presented to them. This especially marks a great success point in the project's pipeline, seeing as it proves that a mere 100 dMRI volumes, which are of fast acquisition, can be used to generate any other scans clinicians might deem necessary, no matter the acquisition settings defined.

## 5.2 Future Work

Future work based on this dissertation must, first and foremost, extend the tools' capabilities to new and more natural databases, i.e., non-healthy patients. Because results prove themselves able to capture intrinsic data structures using a minimal amount of information from a subset of other scans, if the dataset was ever expanded to include patients with tumours, there is a likely possibility that constructed models would fare somewhat successfully in capturing, for example, brain tumours and in providing clinicians with different contrast views of those same ailments. While perhaps not the ultimate test to this dissertation's results, it would be a natural outcome of the research proposed, and one which creates a greater window of application. In that sense,

it also opens the discourse on whether or not the models are able to generate accurate scans for parameter values it has not seen, and which might be a far cry from those contained in the dataset. Furthermore, as it pertains to the dataset, perhaps one approach for tackling the faulty scans in it and improve overall performance would be the implementation of an outlier detection algorithm specialised in dMRI, to the likes of the Slice-wise OutLIER Detection *SOLID* methodology in [31].

In addition to this, this dissertation opens the door into fully conditional models using dMRI data, which, for the most part, was not successful. However, if further explored, it could be possible that a single 3D dMRI volume could contain and transmit all the information necessary for targeted mapping into any other acquisition settings. This, in turn, would allow patients to lay in the scanner for only a minute or two, cutting costs and the need for hospitals to have multiple scanners, while having deep learning methods do the heavy lifting. However, such a convoluted issue could only be fixed by a Deep Learning-based any-to-any approach, one which would most likely need to have full knowledge on the 3D scan, so as to allow for more spatial insight and the building of 3D feature representations.

While this concept of a practically virtual, software-only dMRI scanner would be ideal for the resolution of this dissertation's problem, it is important to notice that the presented approach provides a legitimate and proven way to generate patient-specific target-accurate scans from only a very reduced number of acquired dMRI images. This, by definition, enables clinicians to pick and choose what acquisition settings they wish to see, without requiring the patient to still be in the building, thus automating the process of acquisition and analysis itself.

# References

- [1] Xin Ding, Yongwei Wang, Zuheng Xu, William J. Welch, and Z. Jane Wang. Continuous conditional generative adversarial networks: Novel empirical losses and label input mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45:8143–8158, 7 2023.
- [2] Ajitesh Kumar. How does dall-e 2 work?, 2023.
- [3] Yilmaz Korkmaz, Muzaffer Ozbey, and Tolga Cukur. MRI reconstruction with conditional adversarial transformers.
- [4] Dong Nie, Roger Trullo, Jun Lian, Li Wang, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with deep convolutional adversarial networks. *IEEE transactions on bio-medical engineering*, 65:2720, 12 2018.
- [5] Simon A.A. Kohl, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R. Ledsam, Klaus H. Maier-Hein, S. M. Ali Eslami, Danilo Jimenez Rezende, and Olaf Ronneberger. A probabilistic u-net for segmentation of ambiguous images. *Advances in Neural Information Processing Systems*, 2018-December:6965–6975, 6 2018.
- [6] Heidi Johansen-Berg and Timothy E.J. Behrens. Diffusion mri. *Diffusion MRI*, 2009.
- [7] Ioannis Karatzas and Steven E. Shreve. Brownian motion and stochastic calculus. 113, 1998.
- [8] Andrew L. Alexander, Jee Eun Lee, Mariana Lazar, and Aaron S. Field. Diffusion tensor imaging of the brain. *Neurotherapeutics*, 4:316, 7 2007.
- [9] Paul E. Kaplan. Fractional anisotropy. *Encyclopedia of Clinical Neuropsychology*, pages 1074–1074, 2011.
- [10] Filip Szczepankiewicz, Carl Fredrik Westin, and Markus Nilsson. Gradient waveform design for tensor-valued encoding in diffusion mri. *Journal of Neuroscience Methods*, 348:109007, 1 2021.
- [11] Jacques Donald Tournier, Susumu Mori, and Alexander Leemans. Diffusion tensor imaging and beyond. *Magnetic Resonance in Medicine*, 65:1532–1556, 6 2011.
- [12] Derek K. Jones, Thomas R. Knösche, and Robert Turner. White matter integrity, fiber count, and other fallacies: the do’s and don’ts of diffusion mri. *NeuroImage*, 73:239–254, 6 2013.
- [13] DC Alexander, KK Seunarine, and DK Jones. Diffusion mri: theory, methods, and application. 2010.

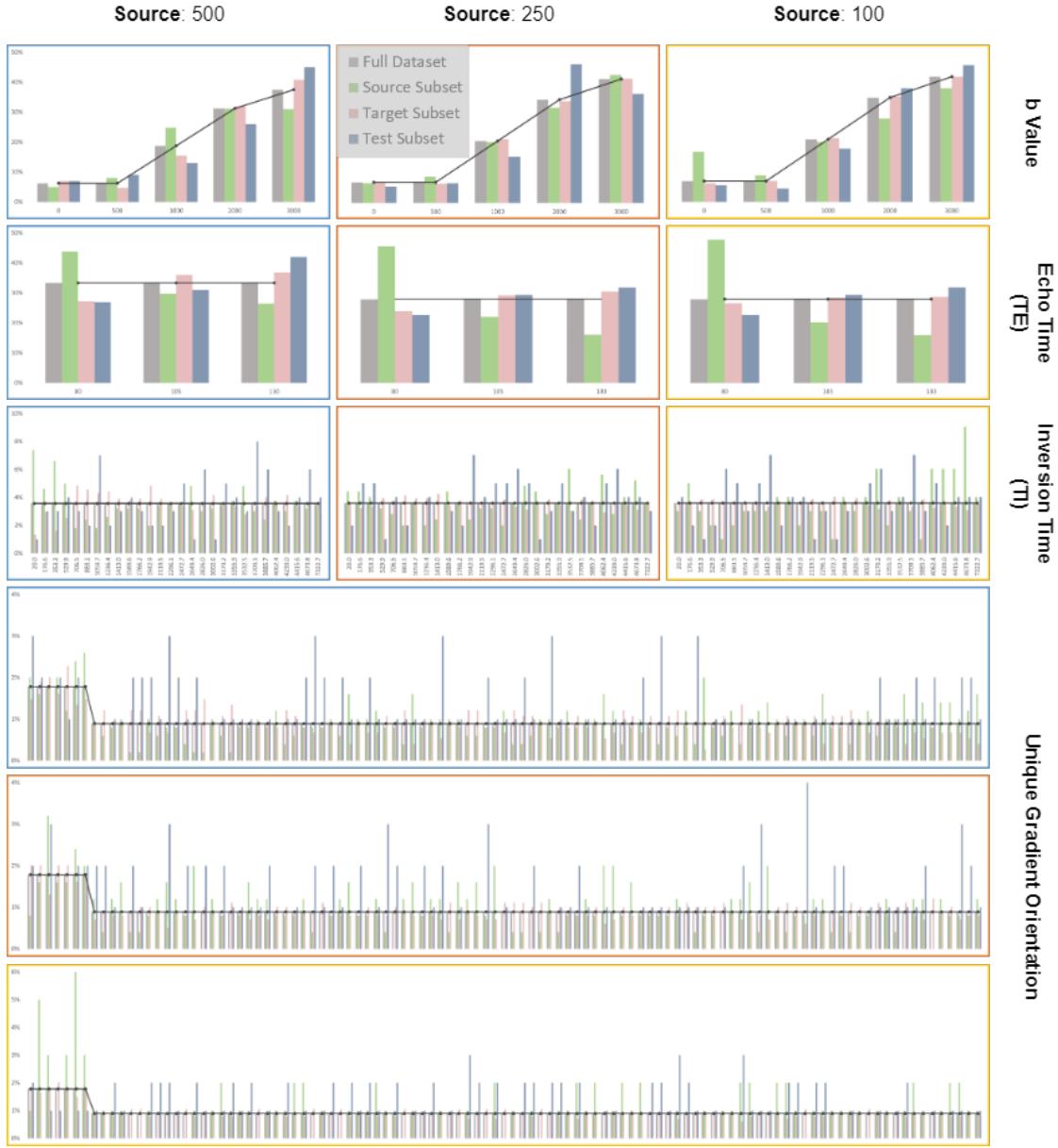
- [14] J. I. Berman, M. R. Lanza, L. Blaskey, J. C. Edgar, and T. P.L. Roberts. High angular resolution diffusion imaging probabilistic tractography of the auditory radiation. *AJNR: American Journal of Neuroradiology*, 34:1573, 8 2013.
- [15] Maxime Descoteaux, Rachid Deriche, Thomas R. Knösche, and Alfred Anwander. Deterministic and probabilistic tractography based on complex fibre orientation distributions. *IEEE transactions on medical imaging*, 28:269–286, 2 2009.
- [16] Robert W. Brown, Yu-Chung N. Cheng, E. Mark Haacke, Michael R. Thompson, and Ramesh Venkatesan. Magnetic resonance imaging : physical principles and sequence design.
- [17] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63:139–144, 6 2014.
- [18] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-January:4352–4360, 2 2017.
- [19] OpenAI. DALL·E: Creating images from text. <https://openai.com/research/dall-e>, 2021.
- [20] Xuran Pan, Tianzhu Ye, Dongchen Han, Shiji Song, and Gao Huang. Contrastive language-image pre-training with knowledge graphs. 10 2022.
- [21] CMW Tax, Hugo Larochelle, J De Almeida Martin, Jana Hutter, Derek K Jones, Maxime Chamber-land, and Maxime Descoteaux. Optimising multi-contrast mri experiment design using concrete autoencoders. In *Proc. 29th Annual Meeting of the ISMRM, Online*, page 1240, 2021.
- [22] Álvaro Planchuelo-Gómez, Maxime Descoteaux, Santiago Aja-Fernández, Jana Hutter, Derek K Jones, and Chantal M W Tax. Optimising multi-contrast mri experiment design using concrete autoencoders, 2020.
- [23] Daniel C. Alexander. A general framework for experiment design in diffusion mri and its application in measuring direct tissue-microstructure features. *Magnetic resonance in medicine*, 60:439–448, 2008.
- [24] Marco Pizzolato, Marco Palombo, Elisenda Bonet-Carne, Chantal M.W. Tax, Francesco Grussu, Andrada Ianus, Fabian Bogusz, Tomasz Pieciak, Lipeng Ning, Hugo Larochelle, Maxime Descoteaux, Maxime Chamberland, Stefano B. Blumberg, Thomy Mertzanidou, Daniel C. Alexander, Maryam Afzali, Santiago Aja-Fernández, Derek K. Jones, Carl Fredrik Westin, Yogesh Rathi, Steven H. Baete, Lucilio Cordero-Grande, Thilo Ladner, Paddy J. Slator, Joseph V. Hajnal, Jean Philippe Thiran, Anthony N. Price, Farshid Sepehrband, Fan Zhang, and Jana Hutter. Acquiring and predicting multidimensional diffusion (mudi) data: An open challenge. *Mathematics and Visualization*, pages 195–208, 2020.
- [25] Lucilio Cordero-Grande, Daan Christiaens, Jana Hutter, Anthony N. Price, and Jo V. Hajnal. Complex diffusion-weighted image estimation via matrix recovery under general noise models. *NeuroImage*, 200:391–404, 10 2019.

- [26] Jesper L.R. Andersson, Stefan Skare, and John Ashburner. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage*, 20:870–888, 10 2003.
- [27] Stephen M. Smith, Mark Jenkinson, Mark W. Woolrich, Christian F. Beckmann, Timothy E.J. Behrens, Heidi Johansen-Berg, Peter R. Bannister, Marilena De Luca, Ivana Drobniak, David E. Flitney, Rami K. Niazy, James Saunders, John Vickers, Yongyue Zhang, Nicola De Stefano, J. Michael Brady, and Paul M. Matthews. Advances in functional and structural mr image analysis and implementation as fsl. *NeuroImage*, 23:S208–S219, 1 2004.
- [28] Christopher M. Bishop. Pattern recognition and machine learning. *Information Science Statistics*, page 738, 2006.
- [29] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *7th International Conference on Learning Representations, ICLR 2019*, 11 2017.
- [30] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13:600–612, 4 2004.
- [31] Viljami Sairanen, A. Leemans, and C.M.W. Tax. Fast and accurate slicewise outlier detection (solid) with informed model estimation for diffusion mri data. *NeuroImage*, 181:331–346, 2018.

## **Appendix A**

# **Parameter Subset Distributions**

In this appendix, the distribution of all acquisition settings for all used source, target and test parameter combination subsets is observable. It also includes a comparison between the distributions in these subsets and that of the full MUDI dataset. Obtained histograms prove that, despite some minor discrepancies, all subsets have a balanced and equitable amount of scans for all instances of any of the four acquisition settings described in the methodology. This does not imply that all parameter values have the exact same representation in the full dataset, as stated before, but that their representation in all subsets successfully mimics that of the overall MUDI dataset. It should also be noted that, for experiments in which the test subset is suppressed, its scans will be contained within the target subset. Because these are many times larger than the 100 scans defined for testing, this results only in a slight alteration of the histograms.



**Figure A.1:** Parameter subset distributions for all acquisition settings. The grey line is centered around the dataset's percentage of scans with specific parameter values and helps understand how the subset's distribution is fairly performed. Gradient orientations have been displayed using unique index values, of which there are 106 in total.