Titre: Title:	Development of a Deep Learning-Based Method for Pediatric Brain Segmentation on Ultra-Low-Field MRI Validated Through Neurodevelopmental Trajectories
Auteur: Author:	Alexia Mahlig
Date:	2025
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Mahlig, A. (2025). Development of a Deep Learning-Based Method for Pediatric Brain Segmentation on Ultra-Low-Field MRI Validated Through Neurodevelopmental Trajectories [Master's thesis, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/67668/

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Development of a Deep Learning-Based Method for Pediatric Brain Segmentation on Ultra-Low-Field MRI Validated Through Neurodevelopmental Trajectories

ALEXIA MAHLIG
Institut de génie biomédical

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

Génie biomédical

Septembre 2025

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Ce mémoire intitulé :

Development of a Deep Learning-Based Method for Pediatric Brain Segmentation on Ultra-Low-Field MRI Validated Through Neurodevelopmental Trajectories

présenté par **Alexia MAHLIG**
en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Lama SÉOUD, présidente
Eva ALONSO ORTIZ, membre et directrice de recherche
Benjamin DE LEENER, membre et codirecteur de recherche
Tomas PAUS, membre

ACKNOWLEDGEMENTS

My deepest gratitude goes to my codirector, Benjamin De Leener, whom I cannot thank enough for his support, his guidance, and the many insightful discussions along the way. Most of all, thank you for believing in me.

I would also like to thank my research director, Eva Alonso Ortiz, for giving me the opportunity to begin my Master's journey at Polytechnique Montréal and for introducing me to the world of medical imaging.

A huge thank you to all my lab mates at NeuroPoly for their help throughout my Master's and for sharing with me the daily life of graduate studies. A special thank you to Andjela, for her support and for the many discussions we shared about our projects.

Thank you to Lama Séoud for accepting to be part of my jury, and to Tomas Paus, for both your role in the evaluation and the valuable collaboration that helped shape this project.

And finally, I would like to thank my friends and family for their constant and unconditional support through every stage of this journey.

RÉSUMÉ

L'imagerie par résonance magnétique (IRM) à très bas champ magnétique représente une modalité prometteuse en neuroimagerie pédiatrique, notamment en raison de son accessibilité et de son coût plus faible par rapport aux systèmes conventionnels. Toutefois, sa qualité d'image réduite présente des défis importants pour la segmentation des structures cérébrales, ce qui limite son utilité dans les études neurodéveloppementales qui reposent sur des mesures morphométriques précises et longitudinales. Ce projet vise à évaluer si la reconstruction d'images IRM de très bas champ magnétique en images de plus haute résolution permet d'améliorer la qualité des segmentations, et de proposer une méthode de validation longitudinale capable de mesurer la cohérence des résultats avec les trajectoires développementales attendues.

Une base de données synthétique de 844 cartes T1 pédiatriques simulant une IRM à 64mT a été générée à partir du Baby Connectome Project (BCP), en utilisant un échantillonnage gaussien basé sur des valeurs de T1 issues de la littérature, puis dégradée pour reproduire les caractéristiques réalistes du très bas champ. Malgré un accès limité aux données réelles, la base synthétique a réussi à reproduire des contrastes et des résolutions typiques du très bas champ magnétique. Un U-Net 3D optimisé a ensuite été entraîné pour reconstruire des images de haute résolution à partir de ces images dégradées. Les reconstructions ont montré des améliorations notables de la segmentation pour la matière grise corticale et les ventricules, ainsi que des performances stables pour la matière blanche et la matière grise profonde.

Au-delà de la segmentation, une méthode de validation longitudinale a été développée afin d'évaluer si les volumes tirés des segmentations suivent les trajectoires neurodéveloppementales attendues. Appliquée aux images reconstruites et aux images synthétiques de très bas champ magnétique, cette méthode révèle que les données synthétiques suivent plus fidèlement les tendances développementales, malgré une précision de segmentation légèrement inférieure. Ces résultats soulignent l'importance d'évaluer les méthodes de reconstruction non seulement à l'aide de métriques voxel par voxel, mais aussi en fonction de leur impact sur les tendances longitudinales.

Des travaux complémentaires seront nécessaires pour généraliser ces résultats à des données réellement acquises à très bas champ, notamment par l'acquisition d'images additionnelles et l'exploration de modèles de reconstruction plus avancés. Néanmoins, cette étude propose une approche prometteuse pour améliorer la qualité des segmentations en neuro-imagerie pédiatrique à très bas champ. La méthode de validation introduite constitue par ailleurs

un outil pertinent pour évaluer la fiabilité des segmentations longitudinales et guider leur interprétation dans les études de neurodéveloppement. Ensemble, ces contributions appuient le potentiel de l'IRM à très bas champ magnétique pour les études populationnelles pédiatriques, en particulier lorsqu'elle sont couplées à des stratégies de reconstruction et de validation adaptées aux défis spécifiques de l'imagerie cérébrale en développement.

ABSTRACT

Ultra-low-field (ULF) MRI is emerging as a promising modality in neuroimaging, particularly in pediatric populations where accessibility and cost are key considerations. However, its lower image quality presents significant challenges for structural brain segmentation, limiting its utility in neurodevelopmental studies that require accurate morphometric measurements over time. This project aims to evaluate whether reconstructing high-resolution pediatric brain images from synthetic ULF MRI can improve segmentation accuracy and to propose a longitudinal validation framework capable of assessing whether segmentation results align with expected developmental trajectories.

To support this goal, a synthetic dataset of 844 pediatric 64mT T1 maps was generated based on the images from the Baby Connectome Project (BCP), using a Gaussian sampling approach informed by literature-derived T1 values and further degraded to simulate realistic ULF characteristics. Despite limited data availability, the resulting dataset successfully replicated key features of ultra-low-field imaging. A 3D U-Net model optimized for this task was trained to reconstruct high-resolution images from the synthetic ULF inputs. The reconstructed images showed improved segmentation accuracy for cortical grey matter and ventricles, with stable performance for white matter and deep grey matter structures.

Beyond segmentation accuracy, a longitudinal validation framework was developed to assess whether the segmentation outputs follow known neurodevelopmental trajectories. When applied to both the reconstructed and the synthetic ULF datasets, the framework revealed that the synthetic ULF data more closely adhered to expected developmental trends, despite exhibiting slightly lower segmentation accuracy. These findings highlight the importance of evaluating reconstruction methods not only through voxel-wise metrics, but also in terms of their impact on longitudinal patterns.

While additional work is needed to generalize these results to real ULF data, including age-diverse acquisitions and more advanced models architecture, this study demonstrates a promising approach for improving segmentation in pediatric ULF neuroimaging. The proposed validation framework further provides a valuable tool for assessing the longitudinal reliability of segmentation outputs and guiding interpretation in growth-based analyses. Together, these contributions support the potential of ULF MRI for population-scale pediatric imaging, particularly when paired with reconstruction and validation strategies that account for the challenges of early brain development.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

B0	Main Magnetic Field
BCP	Baby Connectome Project
CDF	Cumulative Distribution Function
CNN	Convolutional Neural Networks
CSF	Cerebrospinal Fluid
dHCP	Developing Human Connectome Project
GAM	Generalized Additive Model
GAMLSS	Generalized Additive Model for Location, Scale and Shape
GAMM	Generalized Additive Mixed Model
GAN	Generative Adversarial Network
GM	Grey Matter
GT	Ground Truth
HF	High-Field
MANTiS	Morphogical Adaptive Neonate Tissue Segmentation
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
PSNR	Peak Signal-to-Noise Ratio
RF	Radiofrequency
ROI	Region of Interest
SNR	Signal-to-Noise Ratio
SSIM	Structural Similarity Index
STD	Standard Deviation
T1	Longitudinal Relaxation Time
T2	Transverse Relaxation Time
T1w	T1-weighted
T2w	T2-weighted
TE	Echo Time
TR	Repetition Time
ULF	Ultra-Low-Field
VAE	Variational Autoencoder
WM	White Matter

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

1.1 Problem Statement

Understanding how the human brain operates remains a central question in neuroscience. Despite major advances, many aspects of brain development and activity are still not fully understood. From just a few cells, the brain develops into a highly complex organ responsible for coordinating thoughts, behaviors, and actions. Understanding this development can shed light not only on how the brain functions, but also on how and when developmental abnormalities arise. One way to study these processes is by acquiring precise anatomical information throughout the different phases of development and tracking how these structures evolve over time. Collecting this data from a large number of individuals allows for robust datasets and can lead to more reliable patterns and hypotheses to emerge. Various imaging modalities can provide such data, but in pediatric populations, the non-invasive nature of the technology becomes critical, especially when imaging large cohorts. While modalities such as CT and ultrasound are available, they either expose subjects to ionizing radiation or lack sufficient soft tissue contrast. In contrast, magnetic resonance imaging (MRI) is non-ionizing and offers excellent soft tissue contrast, making it particularly well suited for studying the developing brain. However, conventional MRI remains expensive and logistically demanding. It requires advanced infrastructure, specialized personnel, and remains largely inaccessible in low-resource settings. These limitations restrict the diversity of sampled populations and the scale of longitudinal studies, introducing potential biases and reducing generalizability.

In recent years, portable ultra-low-field (ULF) MRI scanners such as the Hyperfine Swoop [1] have become commercially available. They are substantially more affordable than standard MRI units, require minimal installation infrastructure, and are designed for ease of use. However, this promising technology comes with trade-offs. To achieve portability, they use a much lower magnetic field strength of 64 mT compared to the 1.5 T or 3T of conventional clinical scanners, which results in reduced image quality. This in turn limits the amount of information that can be reliably extracted for neurodevelopmental studies. Despite these challenges, technology continues to evolve, and new methods are being developed to address these limitations.

One area of particular interest is image segmentation; the process of delineating different brain tissues, which then enables the extraction of region-specific metrics. Segmentation can be done manually by expert radiologists, but this approach is time-consuming and impractical for large-scale datasets. As a result, automated segmentation tools are commonly used, but

most have been developed and optimized for adult brains, which differ significantly from pediatric anatomy. In addition, existing tools often struggle with the low image quality typical of ULF scans, limiting their reliability in downstream analyses. Improving image quality prior to segmentation is one strategy to mitigate these limitations. However, with a wide range of available tools, with their own strengths and weaknesses, it remains difficult to determine which approaches are most suitable for longitudinal studies, and how segmentation errors may influence derived metrics over time.

This thesis addresses these challenges by investigating whether the reconstruction of high-quality images from synthetic ULF scans can enhance segmentation performance in pediatric brain MRI. It also proposes a longitudinal validation framework to assess how segmentation quality influences neurodevelopmental trajectory analyses, providing a means to compare and interpret methods more effectively in long-term studies.

1.2 Thesis Outline

The remainder of this thesis is organized as follows. **Chapter 2** provides an overview of the key concepts related to this project, including relevant aspects of brain anatomy, neurodevelopmental changes during early childhood, and fundamental principles of MRI physics. It also includes a literature review covering existing segmentation and reconstruction methods adapted for pediatric and ultra-low-field imaging. In addition, this chapter briefly reviews neurodevelopmental analysis techniques and discusses existing validation strategies, particularly in the context of longitudinal studies. **Chapter 3** begins by outlining the three main objectives of this project and describes the methodology used to address each one. **Chapter 4** presents the results corresponding to each objective. **Chapter 5**, also structured around the three objectives, discusses the results in greater detail, highlights the limitations of the methods used, and proposes potential improvements and future directions. **Chapter 6** concludes the thesis with a brief summary of the overall project.

CHAPTER 2 LITERATURE REVIEW

2.1 Brain Anatomy

The brain is commonly divided into three main components: grey matter, white matter, and cerebrospinal fluid (CSF). In neurodevelopmental studies [2–4], brain structures are often grouped according to these categories, as they share common anatomical and functional characteristics. Grey matter is mostly composed of neuronal cell bodies, dendrites, and unmyelinated axons [5], and can be divided into two sub-categories; cortical grey matter, and deep grey matter. Cortical grey matter is located in the outer regions of the brain, such as the cerebral cortex. In contrast, deep grey matter refers to subcortical regions located beneath the cortex, including structures such as the thalamus and basal ganglia. While both are composed of similar tissue types, they are anatomically distinct in location and organization. White matter, located more centrally, consists mainly of the myelinated axons of the neurons [6]. CSF, a clear fluid mostly composed of water, fills the space between the brain and the skull and also occupies internal cavities known as ventricles. An overview of the anatomical positions of all four brain structures and the composition of a neuron is provided in Figure 2.1.

2.2 Neurodevelopment

The human brain undergoes major structural changes during early development. By the age of two, it reaches approximately 80% of its adult volume and around 90% by age five [3, 7]. This growth, however, is not uniform across all brain regions. As illustrated in Figure 2.2, grey matter and white matter follow distinct developmental trajectories, each peaking at different stages. Grey matter volume increases rapidly during early childhood, after which it gradually declines. In contrast, white matter volume continues to grow throughout adolescence and into midlife [2].

Other morphometric features also follow unique developmental patterns. Surface area expands rapidly during early childhood, following a trajectory similar to grey matter volume, whereas cortical thickness peaks much earlier, around two years of age, and then steadily decreases as white matter continues to expand [7]. These overlapping but non-identical trajectories underscore the importance of considering multiple morphometric metrics when studying brain development.

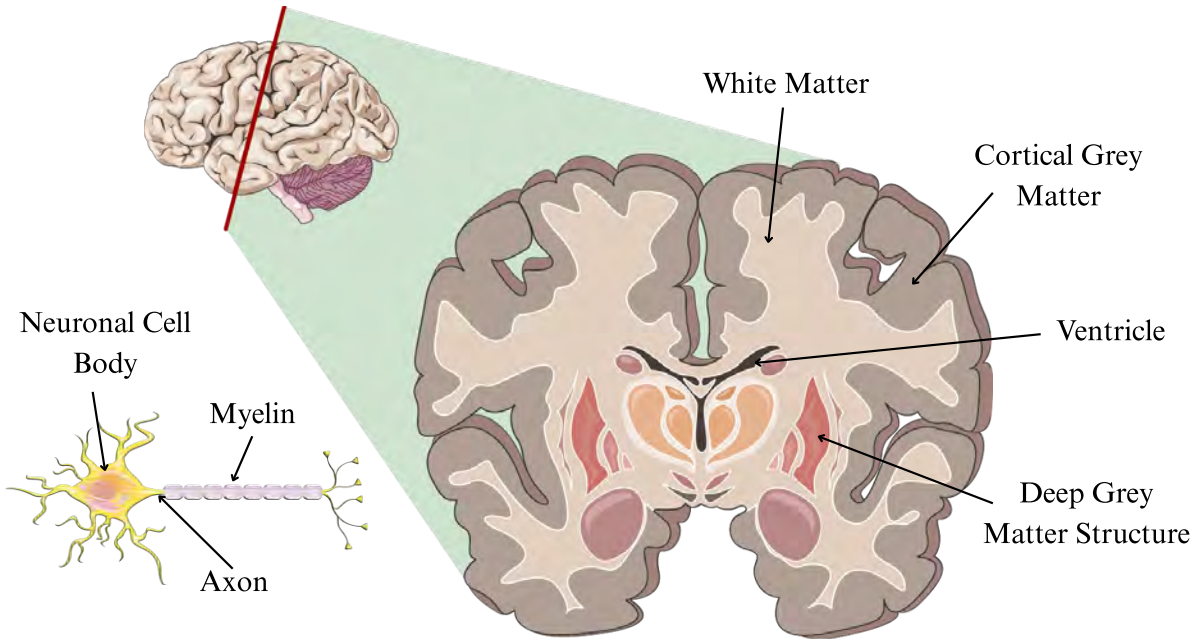

Figure adapted from Servier Medical Art `https://smart.servier.com/`, licensed under CC BY 4.0.

Figure 2.1 Coronal brain section showing cortical grey matter, white matter, deep grey matter structures, and ventricles. On the left, an illustration depicts the composition of a neuron.

Within the deep grey matter, growth is also heterogeneous across structures during childhood. Some regions, such as the pallidum, show rapid volumetric increases shortly after birth, followed by a pronounced decline. Others, like the amygdala, grow more slowly and decrease at a slower rate. The putamen, caudate, thalamus, and hippocampus follow similar developmental patterns, with fast growth after birth followed by a gradual decline, though the timing and rate differ slightly across structures [8].

In addition to volume changes, tissue properties also evolve over time. Among these, myelination is particularly relevant to this project. It progresses rapidly in the first two years of life [9–12] and has a major influence on how brain tissue appears on MRI. These microstructural changes contribute to the evolving appearance of brain images over time and are essential to consider in studies using imaging-based markers of development. Overall, early brain development is a highly dynamic process in which different tissues and structures mature at their own pace. This variability must be considered when analyzing growth trajectories or interpreting developmental brain images.

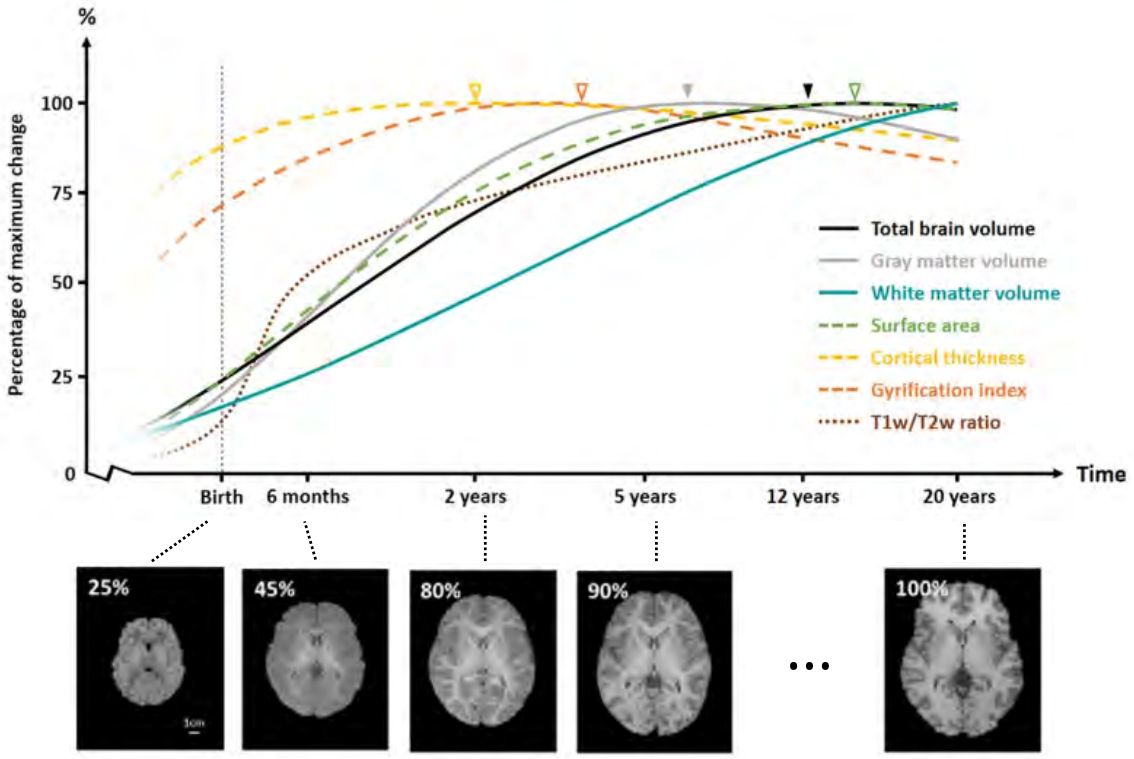

Figure adapted from [7]

Figure 2.2 Developmental trajectories of key brain metrics from birth to adulthood, expressed as percentage of maximum change. MRI images illustrate developmental stages and their corresponding percentages of total brain volume.

2.2.1 Growth Curves

Given these multiple and complex changes, and the importance of studying brain growth, not only from a research perspective to better understand the developing brain, but also from a clinical perspective to help diagnose or monitor neurological conditions, various techniques have been developed to track brain development over time. One commonly used approach is the creation of growth curves [2, 8, 13–15], which makes it possible to evaluate whether an individual is following a typical developmental trajectory or showing significant deviations that may require further investigation. These charts can be applied to a range of metrics, including weight, height, and head circumference, as well as brain-specific measures such as total brain volume [16].

A widely used reference for brain growth is *Brain charts for the human lifespan* [2], which provides normative centiles and rates of change across the lifespan and identifies the timing

of developmental milestones. Another study by Alex et al. [8] focuses on early childhood and aggregates multi-cohort data from birth to six years to model subcortical volumes and their associations with demographic and cognitive factors. As with most growth curve studies, these analyses are based on large, high-quality datasets and primarily emphasize cross-sectional rather than longitudinal analysis. Furthermore, they do not directly address the potential impact of image processing choices, such as segmentation pipelines, on the derived developmental trajectories.

Fitting and Modeling Approaches

An important aspect of growth curves is the type of fit used to model the data, as it directly impacts the shape of the curve and any analysis derived from it. The choice of fitting method can influence how developmental trends are interpreted, particularly when identifying deviations or estimating normative trajectories. A variety of fitting approaches are used in neurodevelopmental studies, depending on the nature of the data and the research objectives.

These methods can be broadly grouped into parametric, semi-parametric, and non-parametric models. Parametric models assume a specific functional form (e.g., polynomial, exponential), which makes them easier to interpret but potentially less flexible. Semi-parametric models represent a balance between maintaining some structure in the curve shape and allowing flexibility. Non-parametric models offer maximum flexibility but may require larger sample sizes and are more sensitive to noise. Semi-parametric models have generally been shown to perform best for brain development data. This is due to their ability to capture the complex, non-linear trajectories typical of neurodevelopment, while still providing stable estimates when sample sizes are small or data is unevenly distributed [17]. Large studies such as [2, 14] employed a Generalized Additive Model for Location, Scale and Shape (GAMLSS) for this purpose. Simplified versions, such as Generalized Additive Mixed Model (GAMM) or Generalized Additive Model (GAM) can also produce robust results, as demonstrated respectively by [15] and [18].

2.3 Magnetic Resonance Imaging

To monitor brain anatomy and its changes over time with accuracy, magnetic resonance imaging (MRI) is widely regarded as the modality of choice for generating large datasets suitable for longitudinal neurodevelopmental studies. Its ability to provide excellent soft tissue contrast without the use of ionizing radiation makes it particularly advantageous for pediatric populations. The physics underlying MRI is highly complex, therefore a simplified

overview will be presented in this section to introduce the key concepts that will be used throughout this thesis.

2.3.1 MRI Basics

Magnetic Resonance Imaging (MRI) in medical applications leverages the abundance of hydrogen protons in biological tissues, and their intrinsic property to posses a spin. When a subject is placed in the scanner, the application of an external magnetic field (B0) tends to align the spins of hydrogen protons with the direction of the field, resulting in the formation of a net magnetization vector. A radiofrequency (RF) pulse tuned to the resonance frequency of these spins (the Larmor frequency) is then applied perpendicular to the magnetic field, tipping the net magnetization away from its alignment. This process, illustrated in Figure 2.3, produces the magnetic resonance effect: as the magnetization relaxes back toward equilibrium, it induces a change in magnetic flux, which generates an electromotive force detected by a receiver coil tuned to the same frequency [19].

The amplitude of the resulting signal is proportional to the strength of the external magnetic field, as stronger fields induce a larger net magnetization. At lower magnetic field strengths, reduced magnetization leads to a lower signal-to-noise ratio (SNR), resulting in noisier images and decreased overall image quality [20]. Spatial localization of the signal is achieved by applying controlled magnetic field gradients, which cause systematic variations in spin frequency and phase. Using this information, the signal can be reconstructed into a volumetric image of the subject [21].

2.3.2 Image Constrast

Building on the basic principles of signal generation, MRI enables the acquisition of different image contrasts by exploiting the fact that the resonance properties of protons vary according to the molecular environment in which they are bound. These differences can be quantified by measuring the time it takes for the net magnetization to return toward equilibrium after an RF excitation pulse. Two main relaxation time constants are defined: the longitudinal relaxation time (T1) and the transverse relaxation time (T2). T1 corresponds to the time required for the longitudinal component of the magnetization to recover to approximately 63% of its equilibrium value, whereas T2 corresponds to the time required for the transverse component to decay to about 37% from its initial value [19]. These relaxation times vary across tissue types and depend strongly on the magnetic field strength.

The measured MRI signal, however, is not determined by relaxation time alone. It is heavily

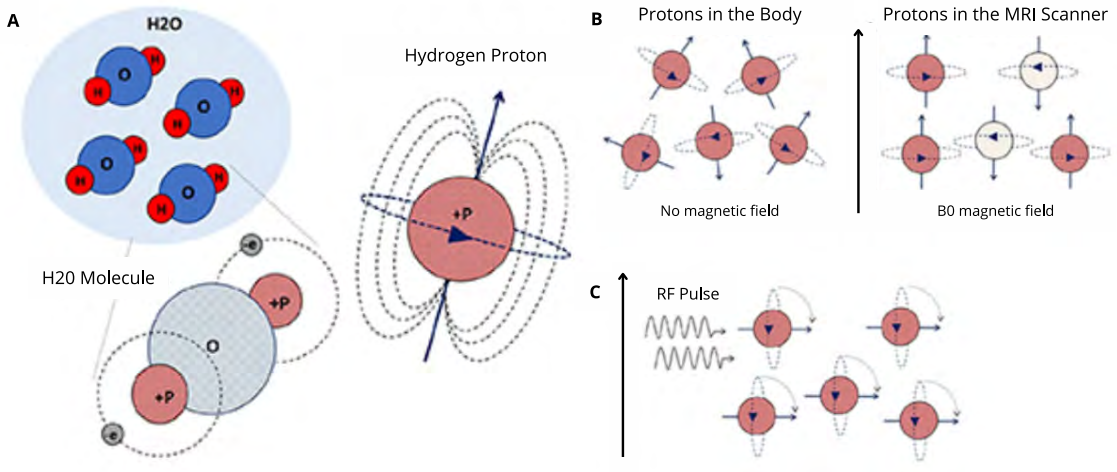

Figure adapted from [22], licensed under CC BY.

Figure 2.3 (A) Structure of the hydrogen molecule. (B) Alignment of proton spins with and without an external magnetic field. (C) Tipping phenomenon induced by a radiofrequency pulse.

influenced by acquisition parameters such as the repetition time (TR) and echo time (TE), as well as by scanner hardware, field inhomogeneities, and the type of RF coil used [21]. As a result, voxel intensity values in MRI images can be difficult to reproduce consistently, even in the same subject. For the purposes of this study, the T1-weighted (T1w) contrast acquisition sequence is particularly relevant. This sequence uses specific TR and TE parameters to enhance the contribution of the T1 signal to the final image. T1w imaging is commonly used as the contrast of choice for structural imaging of the brain due to its high tissue contrast and anatomical detail [23]. However, in young infants, the ongoing process of myelin maturation alters the relaxation properties of brain tissues, reducing the T1w contrast between white and gray matter. As a result, T2w imaging often provides better differentiation of tissue types than T1w imaging in subjects under six months of age. [24].

2.3.3 Quantitative MRI

As outlined above, conventional MRI images such as T1w scans are qualitative in nature and therefore difficult to reproduce. A growing field of research, known as quantitative MRI (qMRI), concentrates on leveraging the link between the measured MRI signal and the biophysical characteristics of the tissues [25]. This field is particularly relevant to the present study, as it enables the generation of more reproducible MRI images. Within this framework, a T1 map can be viewed as the quantitative equivalent of a T1w image. By acquiring multiple T1w images with varying acquisition parameters, T1 values can be fitted and mapped into

an image [26]. This process minimizes the undesired variability of the T1w images, leaving mostly the intrinsic T1 values of the tissues [27]. An example of a T1w image with its derived T1 map is shown in Figure 2.4.

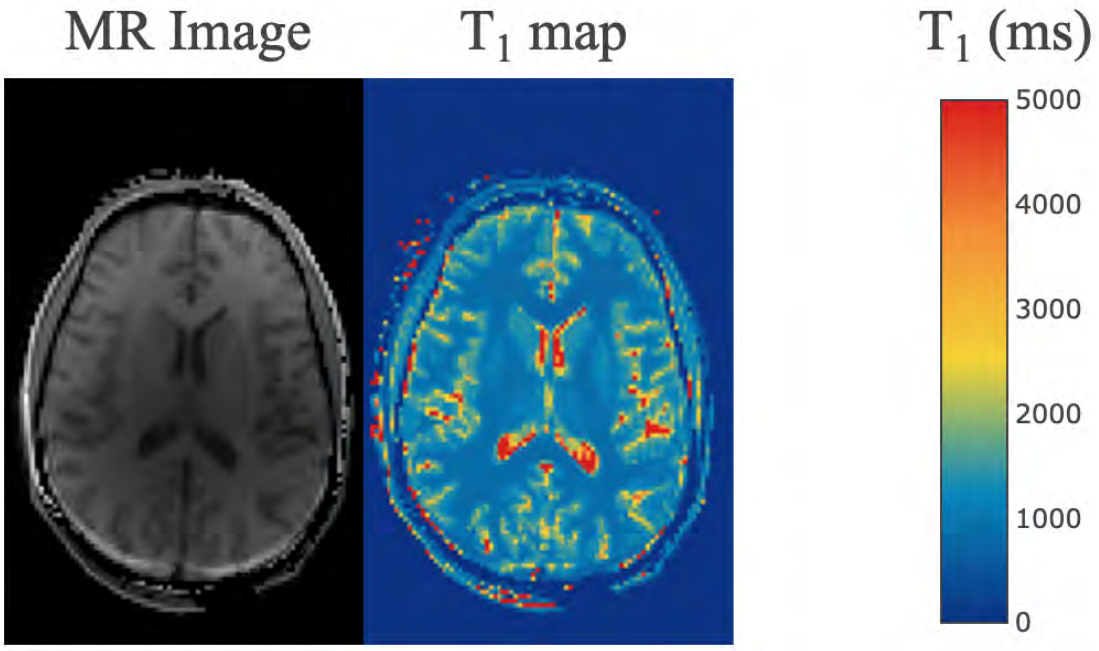

Figure reproduced from `https://qmrlab.org/t1_book/01/ir_blog/IR_DataFitting.html`

Figure 2.4 Example of a structural T1w MR image and its corresponding quantitative T1 map. The T1w image represents one of the acquired images used for the fitting process, shown here with an inversion time of 1030 ms. The color bar indicates T1 values of the T1 map in milliseconds.

2.4 Segmentation

Another central aspect of this project is the segmentation of MRI images. In this context, segmentation is used to delineate the boundaries between different brain structures, such as white matter, grey matter, or smaller regions like the hippocampus. Manual segmentation, while considered the gold standard, requires the expertise of a trained radiologist and is extremely time-consuming, particularly for large datasets. This limitation has motivated the development of numerous automatic segmentation methods [28], which can generally be grouped into three main categories: atlas-based, voxel-based, and deep learning [29].

Most existing brain segmentation methods were developed for adult brains and often perform poorly on pediatric data because of differences in brain anatomy and tissue properties dur-

ing development [30, 31]. Pediatric brains introduce additional challenges, including smaller structures that are harder to identify and lower contrast between white and grey matter in early childhood. The use of ULF MRI introduces additional complexity, with its lower image quality and distinct signal characteristics. Only a few methods have begun to address these challenges, with deep-learning approaches generally reporting the best performance [32, 33]. The specific limitations associated with each approach are discussed in the following sections. Given the focus of this project on ULF pediatric segmentation, atlas-based and voxel-based methods are reviewed briefly, with particular emphasis placed on deep learning approaches.

2.4.1 Voxel-Based and Atlas-Based

The first category is voxel-based segmentation, where each voxel is assigned a probability of belonging to a tissue class based on its intensity, and often incorporating spatial priors. Most widely used voxel-based pipelines, such as ANTs Atropos [34] and FSL FAST [35], were developed for adult brains and generalize poorly to pediatric populations [36], as the underlying intensity models may not accurately reflect the changing signal characteristics seen in younger populations. While these methods are sometimes simpler to implement and do not require large training datasets, their performance deteriorates further on ultra-low-field (ULF) images, where lower SNR and increased artefacts make tissue classes less distinct and segmentation less reliable.

The second category, atlas-based segmentation, relies on a reference atlas, which is a template brain annotated with labeled regions of interest representing the average anatomy of a particular population. The subject image is registered to the atlas, and the anatomical labels are then propagated to the subject space to produce the segmentation [37]. In adult cohorts, atlas-based pipelines often employ multi-atlas registration with label fusion to derive robust segmentations, as implemented in AutoSeg [38]. However, these methods typically do not translate well directly to pediatric data, as the rapid anatomical changes during early brain development necessitate age-specific atlases covering small age ranges to maintain accuracy. To address this limitation, pediatric pipelines such as dHCP [39, 40], Infant Free Surfer [36], ChildMetrix [41] and MANTiS [42] incorporate age-specific atlases and use a combination of atlas-based and voxel-based approaches to address more effectively the unique challenges of developing brains. However, even with pediatric pipelines, atlas-based segmentation methods are limited by their reduced flexibility in the presence of anatomical anomalies or when the subject's brain diverges from the atlas [37]. These challenges increase with ULF images, where poor image quality can result in propagated labels reflecting the atlas more than the individual subject, reducing segmentation accuracy.

2.4.2 Deep Learning Methods

The third approach is based on deep learning models, which have become increasingly common for segmentation in recent years. These methods typically rely on convolutional neural networks (CNNs), a type of deep learning architecture designed for image analysis [43]. A simplified overview of how a CNN works is shown in Figure 2.5. In short, convolution layers first detect image patterns such as edges or textures. Pooling layers then reduce the size of the resulting feature maps, making the model less sensitive to small variations. When combined, these layers enable the network to learn more complex features that can be used for segmentation. As a result, deep learning–based segmentation methods go beyond simple voxel intensities and are able to capture sophisticated anatomical patterns. When trained with realistic data augmentation, they can be robust to noise and variations in image contrast [44]. However, their performance often declines when applied to data that differ from the domain used during training.

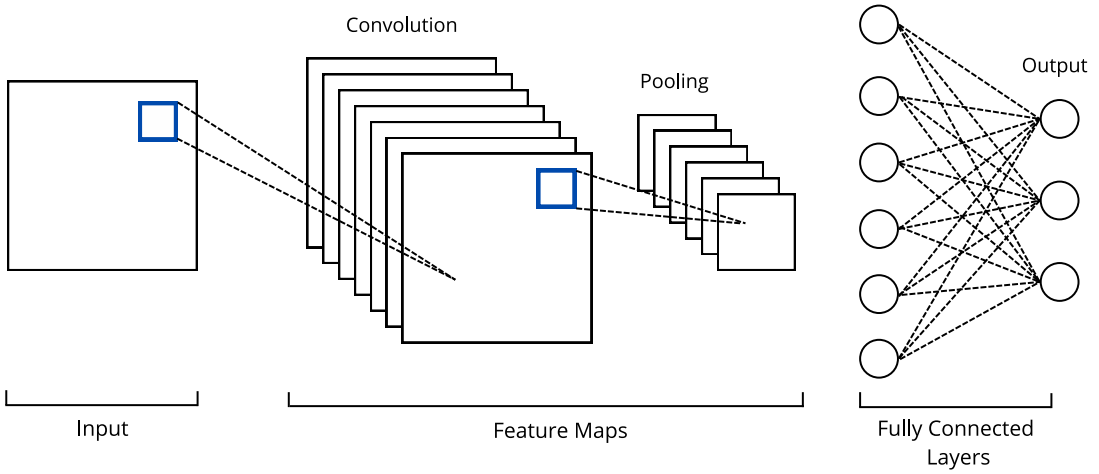

Figure 2.5 Schematic representation of the functioning of a Convolutional Neural Network.

Several deep learning approaches have been developed specifically for pediatric brain segmentation [45–47]. Notable examples include iBEAT [48] and BIBSNet [49], which have demonstrated strong performances on high-field pediatric datasets by leveraging large, age-specific training sets and incorporating preprocessing steps and model architectures designed for early brain anatomy. These methods, however, are generally limited in their applicability to ULF MRI, as they have not been trained on images acquired at these lower magnetic field strengths. While recent work has begun to incorporate ULF images into training [50, 51], these pipelines are typically optimized for adult populations, which limits their ability to generalize to the unique characteristics of pediatric ULF data.

Despite these limitations, deep learning methods remain the most effective approach for segmenting ULF MRI images, as they can be highly specialized when trained on data that closely match the target acquisition characteristics. A major constraint, however, is their reliance on large, well-labeled datasets that closely resemble the images the pipeline aims to segment, which are particularly difficult to obtain at ultra-low field strengths. ULF MRI has only recently gained popularity, resulting in much smaller available datasets compared to more commonly used magnetic field strengths such as 1.5T or 3T [52]. Moreover, the reduced image quality at ultra-low field makes manual segmentation more challenging and less reliable. To address this lack of data, some pipelines have adopted the strategy of generating synthetic images for training, which is the approach taken in this project. The use of synthetic data as a form of data augmentation has already shown promising results in many pipelines, both for brain imaging and for other types of applications [53–55].

The algorithm SynthSeg [56] illustrates these principles well. Rather than relying on conventional training datasets, it generates a large number of synthetic images directly from label maps, with varying image contrast, resolution, field of view, and anatomical characteristics. These synthetic images are then highly deformed to increase the model's robustness to anatomical variability and scanner differences. This training strategy allows SynthSeg to generalize well across diverse imaging conditions, including pediatric datasets, which motivated its selection for specific steps in this project. However, since the model is not specifically trained on pediatric or ULF images, its performance on directly ULF pediatric data remains suboptimal. This limitation, present on most available segmentation methods, leads to the need for an additional step to improve the quality of ULF images before using existing segmentation methods, which is the goal of the reconstruction approach described in the next section.

2.5 Reconstruction

Recent progress in deep learning, particularly in image-based applications, has further facilitated the generation of images in medical imaging. Many methods have been developed to reconstruct higher-field MRI images either to compensate for the lack of available data or to enhance the quality of existing data [57, 58]. Unlike deep learning models used for segmentation, which are classification tasks that assign a discrete label to each voxel, models for image reconstruction are formulated as regression tasks that estimate the continuous intensity at each voxel. Several generative model frameworks are well suited for this task, and the most common are summarized below.

Generative Adversarial Networks (GANs) consist of two models trained jointly: a generator,

which produces synthetic images, and a discriminator, which learns to distinguish real images from generated ones. This adversarial loop iteratively improves the realism of the generated images. GANs are widely used for image generation but can be unstable during training, making them more complex to use than simpler models [59]. Diffusion models represent another popular approach. They generate images by learning to progressively remove noise, starting from pure noise. They can produce highly detailed images but are computationally demanding to train and use at inference [60]. Variational Autoencoders (VAEs) learn a latent space representation of the input images and can generate new images by sampling from this distribution. While VAEs are generally easier to train and useful for modeling the variability of the data, their outputs can sometimes be blurrier than those generated by GANs or diffusion models [61].

Finally, the U-Net, a CNN architecture, is among the most widely used models for image segmentation because of its flexibility and simplicity. Although initially designed for segmentation tasks, it can also be effectively adapted for image generation. As shown in Figure 2.6, it follows a U-shaped encoder–decoder design with skip connections between the two paths. These skip connections allow the model to recover fine details lost during the downsampling process in the encoder, making the U-Net particularly effective at preserving global trends while maintaining local details [62]. U-Nets are often used as the backbone within the generative frameworks discussed above or as standalone models. They offer a simple yet flexible architecture that can be adapted and refined to improve performance, making them well suited for the reconstruction task addressed in this project.

Reconstruction pipelines developed to approximate images of higher MRI field strength vary widely in their complexity and objectives. Some methods focus only on super-resolution [63, 64], which essentially increases the resolution of the input images without addressing other characteristics of the images. While this approach is simpler, it overlooks important aspects of MRI physics. Lower-field MRI images typically exhibit lower SNR, increased artifacts, and altered intensity distributions, as magnetic field strength directly influences the measured signal. Since images acquired at lower magnetic field strengths are not simply low-resolution versions of higher-field images, the potential of these approaches to ultimately generalize to real ULF data is limited. Other methods rely on generative models to learn mappings between images acquired at different field strengths, such as LoHiResGAN [51] who employs a GAN trained on paired 64 mT– 3T adult data to synthesize high-field quality images from ultra-low-field inputs. These methods attempt to correct differences in contrast and noise properties but typically require access to large datasets in which each subject has been scanned at both field strengths, a type of data that is difficult and costly to acquire.

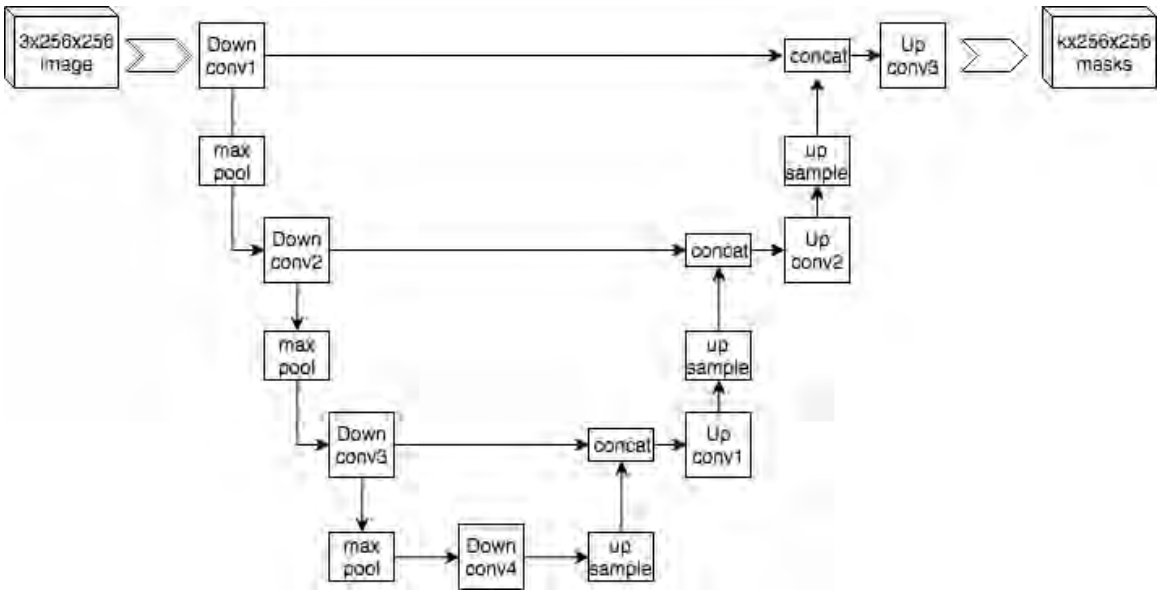

Figure reproduced from Mehrdad Yazdani, Wikimedia Commons, `https://commons.wikimedia.org/wiki/File:Example_architecture_of_U-Net_for_producing_k_256-by-256_image_masks_for_a_256-by-256_RGB_image.png`, licensed under CC BY-SA 4.0.

Figure 2.6 U-Net architecture showing the downsampling and upsampling paths with skip connections.

The SynthSR algorithm proposed by Iglesias et al. [65] addresses these limitations by generating synthetic data that accounts for MRI characteristics beyond differences in spatial resolution. The synthetic images are created with the basis of the intrinsic values of the target image contrast, thereby eliminating the need for paired datasets. It has been shown to reconstruct more realistic images across different types of contrasts and to improve segmentation accuracy when applied to adult ULF data [32]. For these reasons, this algorithm was chosen as the foundation for the present study.

2.6 Validation Methods

Validating the reliability of the processing methods used is a critical step in neuroimaging studies. This section focuses on three types of validation implemented in this project. The first is image reconstruction validation, which ensures that the reconstructed images accurately represent those they are intended to simulate. The second is segmentation validation; since all analyses rely on metrics derived from segmentation, it is essential to understand both the accuracy and the limitations of the method employed. Finally, longitudinal validation evaluates the extent to which the processing pipeline affects the final results and their interpretation, a necessary component in longitudinal studies.

2.6.1 Reconstructed Images Validation

As mentioned by [66], there is currently no standardized procedure to evaluate the quality and accuracy of reconstructed neuroimages. For images generated using deep learning approaches, metric such as mean squared error (MSE) is often used during training to guide the model toward a more accurate representation of the images [67]. However, this metric only assess voxel-level intensity differences and does not capture the overall fidelity of the reconstructed image. Complementary metrics such as the structural similarity index (SSIM) or the peak signal-to-noise ratio (PSNR), as used by [64], can also be employed to evaluate perceived image quality or key MRI characteristics such as SNR.

The validation of deep learning models can be performed in several ways but typically involves the use of both a validation set and a test set. The validation set is used to fine-tune the model's hyperparameters, while the test set, which contains data never seen during training, is used for the final evaluation. This setup allows the assessment of the model's ability to generalize to unseen data. In practice, the dataset can be split directly into training, validation, and test sets. However, this approach may be suboptimal for limited datasets as it reduces the number of samples available for training. In such cases, cross-validation is often preferred [68]. This technique divides the dataset into k folds and trains the model k times, each time using a different fold as the validation set and the remaining k–1 folds for training. The final performance metric is obtained by averaging the results across all folds, providing a more robust estimate than a single train–validation split. This ensures that validation is consistently performed on unseen data while maximizing the use of the entire dataset. The quality of reconstructed images is often further assessed by examining their impact on downstream tasks, such as segmentation accuracy.

2.6.2 Segmentation Validation

Several segmentation validation metrics exist to compare the predicted segmentation to the ground truth, either from the method itself or from the reconstructed image. Some metrics are boundary-based, such as the Hausdorff Distance, which measures the maximum distance from a point on the predicted boundary to the closest point on the ground truth boundary. However, overlap-based metrics are more commonly used [69]. The Jaccard Index, also known as the Intersection over Union (IoU), measures the ratio between the intersection and the union of the predicted and ground truth labels. The most widely used metric is the Dice–Sørensen coefficient [70], commonly referred to as the Dice Score. This metric ranges from 0 to 1, where 0 indicates no overlap and 1 indicates a perfect match, and is computed

as follows:

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|}$$

where A is the set of predicted voxels and B is the set of ground truth voxels. As it is both widely adopted and easily interpretable, the Dice score is a strong metric for comparing methods across studies.

2.6.3 Longitudinal Validation

The metrics presented in the previous sections evaluate each image independently and therefore do not capture longitudinal changes that may occur when a subject is scanned at multiple time points. Yet, assessing the impact of image processing on longitudinal measures is essential to neurodevelopment studies, as it provides an understanding of the method's reliability over time. This type of evaluation can also serve as a basis for comparing different methods and determining which are best suited for studies that rely on repeated measurements. To my knowledge, very few validation approaches address these challenges, and those that do are not widely used in the field. Most studies focus on evaluating the results in a cross-sectional manner, as [71] who uses a test-retest evaluation. While this approach is useful for confirming that the method is consistent when no biological change is expected, it does not assess whether the changes measured by the method reflect true longitudinal patterns. Other methods [72,73] aim to improve consistency across time points through joint registration and segmentation, but they do so without a clear validation framework to support their longitudinal reliability. Alternatively, a study by Long et al. [74] assesses the biological plausibility of longitudinal metrics using spaghetti plots, providing a qualitative overview of individual trajectories. However, this approach lacks quantitative measures of deviation at the subject level, limiting its ability to capture subtle inconsistencies in longitudinal accuracy.

2.7 Research Gap

Ultra-low-field (ULF) MRI holds great promise for imaging large pediatric populations, making it a practical and scalable choice for neurodevelopmental studies. However, its lower image quality poses challenges for segmentation accuracy, which is essential for extracting brain volumes and deriving metrics to track brain growth. Although progress has been made, most existing segmentation methods are not optimized for the combined challenges of pediatric brain anatomy and the unique characteristics of ULF MRI, thereby limiting their performance. Furthermore, very few frameworks address the need to evaluate segmentation methods from a longitudinal perspective and ensure that their outputs align with expected

developmental trajectories, an essential consideration as segmentation approaches continue to be developed and applied in neurodevelopmental studies. These considerations form the foundation of my research objective, which is to develop a method for pediatric brain volume extraction from ULF MRI with a validation based on neurodevelopmental growth trajectories.

CHAPTER 3 METHODOLOGY

A review of the literature showed that improvements in ultra-low-field (ULF) brain pediatric segmentation methods are still needed to allow for accurate analysis. One way to improve segmentation accuracy is to first enhance the quality of the images before applying the segmentation method. This project adapts the framework proposed by Iglesias et al. [65] to the pediatric population, using a deep learning approach on synthetic images to reconstruct higher quality images prior to segmentation, and extends it with a novel validation method based on neurodevelopmental growth trajectories. To achieve this goal, the project was divided into three sub-objectives:

- **O1** : Generate a synthetic ultra-low-field pediatric brain MRI dataset
- **O2** : Develop and validate a reconstruction method for high-field images from the synthetic ultra-low-field dataset
- **O3** : Create a longitudinal validation framework based on neurodevelopmental growth trajectories

A summary of the steps included in each sub-objective and the overall project workflow is shown in Figure 3.1.

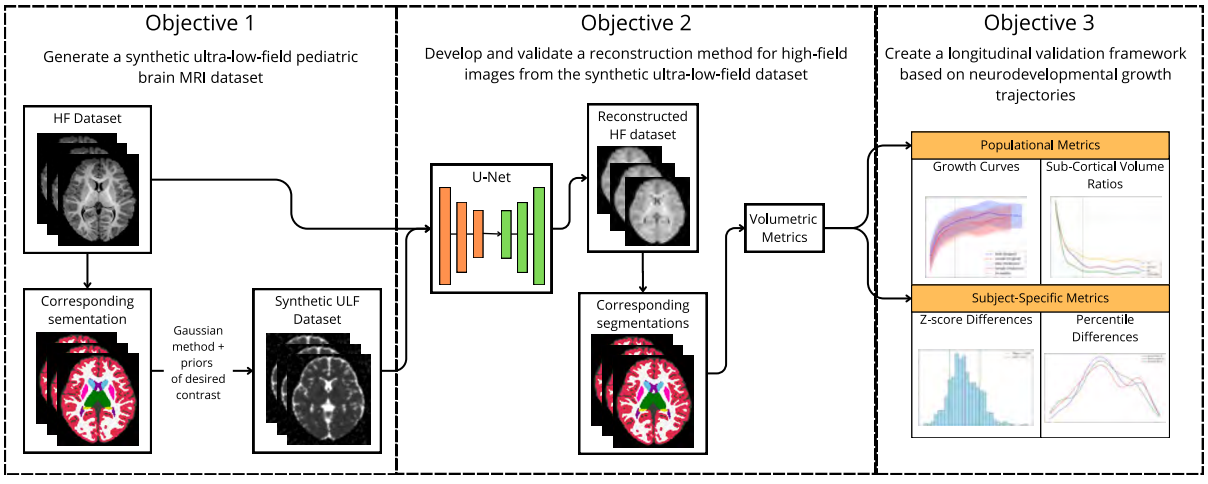

Figure 3.1 Overview of the main steps of the project, organized according to the three objectives.

3.1 O1 : Generation of a Synthetic Dataset

The creation of the synthetic ULF dataset lays the foundation for the rest of the project as it will be used all throughout the second and third objectives. With limited ULF data available, a choice was made to create a dataset of synthetic 64 mT T1 map images, as the values of quantitative MRI images could be more easily adapted to the needs of the task. However, a similar approach could be used with a different contrast, provided the availability of more data. An overview of the steps taken in the generation of the dataset are shown in Figure 3.2.

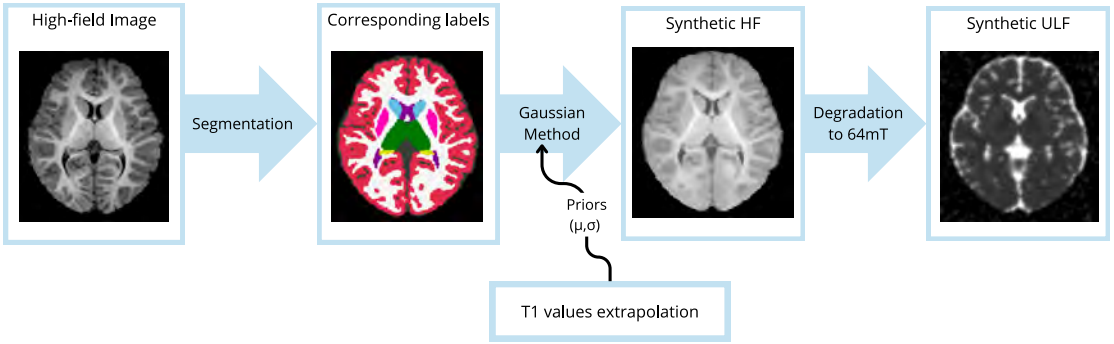

Figure 3.2 Workflow diagram illustrating the main steps of Objective 1, shown for a single image and applied consistently across the entire dataset.

The brain morphometrics of multiple subjects are extracted from a real qualitative T1w dataset by segmenting their images. T1 values are then extracted from three public studies and applied to these segmentations in order to create a synthetic high-field T1 map dataset. This dataset is then degraded to a 64mT field strength. Although magnetic field strength affects several factors of image formation, the degradation of the images was centered around three key factors: spatial resolution, signal-to-noise ratio (SNR), and field strength–dependent variations in T1 relaxation times. The details of these steps are explained in the following section.

3.1.1 Data

The Baby Connectome Project (BCP) dataset [75] was chosen as the high-field T1w dataset used in this project to generate the synthetic images and subsequently reconstruct them into high-field images. The images were acquired on a 3T Siemens Prisma MRI scanner for 371 subjects across multiple time points, covering an age range from 2 weeks to 81 months.

The majority of scans were obtained before 24 months of age. As part of this project, all images underwent quality control to assess the presence of artefacts or abnormalities. 40 images were removed from the initial dataset after this inspection, leaving the total number of images available at 903. A summary of the characteristics of this dataset can be found in Table 3.1.

Despite the changing contrast in infant brain MRI, in this study, only T1w images were used across all age groups. This choice was made to ensure consistency across sessions, as T1w typically show better contrast for segmentation for subjects older than 6 months old [24], and also to reduce the risk of registration errors and to focus the evaluation on the performance on the training pipeline instead of these potentially introduced biases.

Table 3.1 Characteristics of the Baby Connectome Project (BCP) dataset

	BCP dataset
Number of subjects	371
Age range	2 weeks – 81 months
Number of images removed with QC	40
Number of images left after QC	903
Average number of sessions per subject	2.53
Average age of sessions	17 months
% Female	51.88 %
% Male	47.85 %

A T1w image acquired from a 2-month-old infant at Sainte-Justine's hospital, in Montréal, Canada, with the 64mT MRI Swoop scanner from Hyperfine [1] was also used as a reference for the noise found in MRI images at this field strength. The details of the acquisition protocol used can be found in Appendice A.

In order to create a T1 map dataset, data from three public studies were combined to derive T1 values across the entire age range: (i) the first dataset [76] contained the T1 values for the entire age range but only for a subset of the regions of interest, (ii) the second dataset [77] contained the images in which the values for all the regions of interest could be extracted but covered only part of the age range, (iii) the third study [9] presented the T1 values for the

cortical maturation in infants up to 6 years old. Each study contributed a different portion of the data and, together with some extrapolation described in more detail in Section 3.1.3, provided the necessary T1 values. A summary of the regions and age ranges corresponding to each study is presented in Table 3.2.

Table 3.2 Summary of the data available from the three datasets used for T1 extrapolation within the selected age range for this study.

Dataset	**i**	**ii**	**iii**
Age Range	3 months–6 years	3–6 years	1–6 years
Number of Subjects	94	5	215
Data Type	Values	Images	Values
Regions Available	Brainstem, Thalamus, Putamen, White Matter, Caudate	Brainstem, Thalamus, Putamen, White Matter, Caudate, CSF, Ventricles, Ventral DC, Pallidum, Hippocampus, Amygdala, Accumbens Area	Cortex

3.1.2 Image Segmentation

Segmentation of the BCP dataset is the first step necessary in generating the synthetic dataset. The resulting segmentation labels are also used later on as the ground truths in the analysis of the validation method. The segmentation was performed on all 903 images using SynthSeg [56], followed by quality control on all outputs. Eight segmentations were excluded due to major errors, and another eight were manually corrected to address minor errors after the automatic segmentation. Segmentation was unsuccessful for subjects aged 2 months and younger because of significant anatomical differences at this age. Other segmentation methods [39,78] were tested on these images but produced subpar results, so these cases were ultimately excluded from the dataset. After curation of the images and their corresponding segmentations, the total number of images was reduced from 903 to 844.

3.1.3 Synthetic Image Generation Process

With the initial segmentations completed and curated, the next step was the creation of the synthetic images. The method used for generating these images was adapted from the method proposed by the SynthSR algorithm [65]. The concept behind the approach resembles a simplified version of a gaussian mixture model. The voxel intensities in each label in the segmented image are assumed to follow their own gaussian distribution. Therefore, a mean and standard deviation are attributed to each label and following this gaussian distribution, random values are assigned to each voxel in the segmented image. This ensures the preservation of the specific morphological characteristics of the subject's brain, while allocating intensity values closer to the aimed image contrast. The intensity ranges were extrapolated from T1 values reported in the literature, then adjusted to simulate the expected T1 values at lower magnetic field strengths. After applying these values to the segmentation, the images are degraded to match the lower resolution and signal-to-noise (SNR) ratio typically found in 64mT images, resulting in synthetic ULF images that depict more accurately what would be found in real acquisitions. Further details of these steps are provided in the following sections.

T1 Values Extrapolation

As T1 values change with age in developing children and ULF images were not available for every age range, it was not possible to simply extract T1 values from a real ULF image and apply their mean and standard deviation to the Gaussian model to generate the entire dataset. In addition, voxel intensity values of T1w images are qualitative and do not represent the intrinsic properties of the tissues, making it impossible to extract values directly from the BCP dataset. Instead, quantitative T1 maps, which more closely reflect the required tissue properties, were chosen as the reference contrast for creating the synthetic ULF dataset. Although this method could be adapted to other contrasts if ULF images of those contrasts were available, for this project the required values had to be extrapolated from the literature to construct the dataset.

The dataset 1 [76] directly contained the T1 values for the entire age range for a subset of the regions of interest. The dataset 2 [77], included 73 subjects aged 3 to 17 years, with 5 subjects falling within the target age range of this study (3 to 6 years). This dataset contained the full brain images and was segmented using the same method described in Section 3.1.2. The third study [9] contained the information for the region of the cortex.

In order to fill the missing information, a two-step smoothing fit was first applied to the

data of each label to model the developmental trajectories. A sliding window was applied to reduce noise, followed by penalized B-spline smoothing using Scipy's Univariate Spline function [79]. The labels present in the two datasets were then evaluated to confirm that similar regions would show similar age-related trends across datasets. They were found to have similar intensities, confirming that assumption. Labels of certain deep grey matter regions (ventral DC, pallidum, hippocampus, amygdala, accumbens area) were absent from the dataset containing the full age range. A Pearson coefficient test found a perfect correlation between the spline-based growth curves of the three grey matter regions present. One region was then selected as a representative to model the trajectories for the missing deep grey matter values. Although the regions may approximately follow similar trends, their absolute values can differ. To account for this, an offset was computed between the two available datasets and added to the fitted values for the missing regions.The third study was used to directly derive the values for the cortex for the target age range and fit a univariate spline.

Figure 3.3 shows the final trajectories of all the regions modeled in the synthetic images for the entire age range. A change in contrast between white and grey matter is visible around 3 months, consistent with the findings of [7] that T1w images can appear isointense between 3 and 6 months postnatal age.

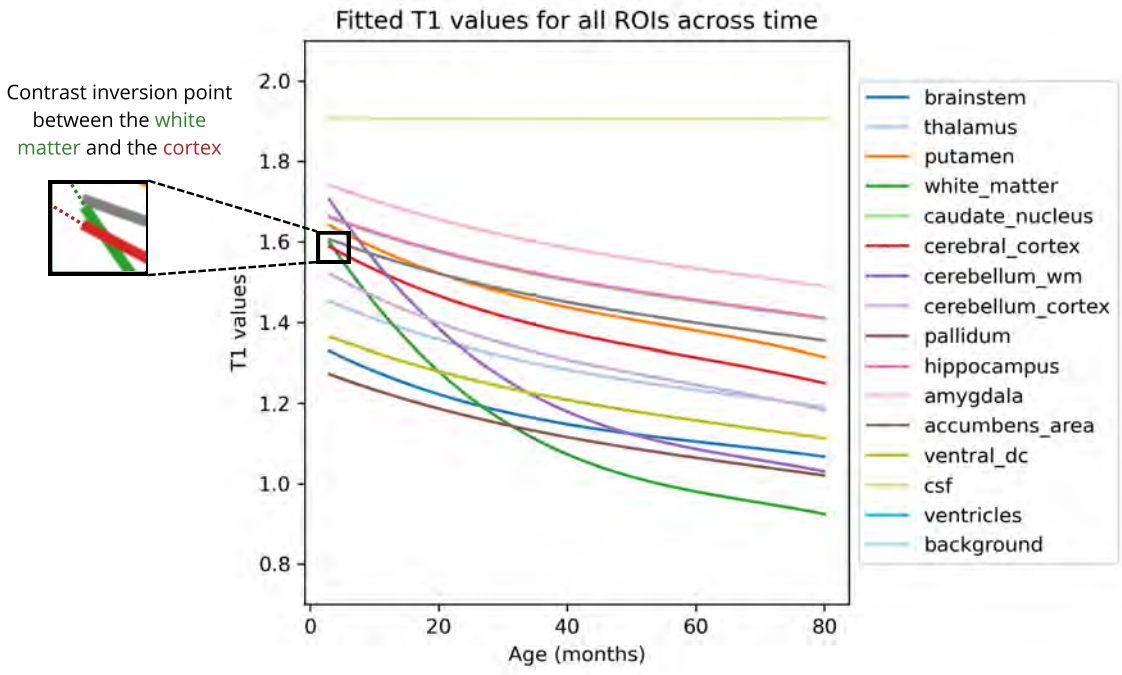

Figure 3.3 Fitted T1 values for the pediatric brain across time, derived from literature data.

Intra-Tissue Variability

As a lot of variability is present even among the same tissue type in a real brain, modeling this variability aids the realism of the synthetic images. The standard deviation of these values is an indicator of the amount of variability found in each label and can then be imputed into the gaussian distribution.

As only dataset 2 [77] contained the full images, the standard deviation of voxel intensities within each label was computed across its five subjects in the desired age range. The mean of these five standard deviations was computed and found to be relatively constant with age. The age-specific mean and the standard deviation values were then used to generate synthetic images by assigning voxel intensities sampled from the corresponding Gaussian distributions to all the segmentations of the BCP dataset.

T1 Relation to B0

T1 values are intrinsically linked to the strength of B0, and this dependence must be taken into account when defining voxel values for synthetic images at a specific field strength. The T1 values extrapolated in the previous section correspond to a B0 of 3T. A relationship therefore needs to be applied to adjust these values to match a B0 field of 64mT. This relationship can be modeled by the following equation described in [80]:

$$T_1 = C \cdot (\gamma B_0)^{\beta} \tag{3.1}$$

Where γ is the gyromagnetic constant, and C and β are constants approximated using values from [81]. The computed constants for each region can be found in Table 3.3. Starting from the T1 values across ages at a field strength of 3T, the scaling ratio was applied to estimate the corresponding T1 values at 64 mT and applied on the synthetic dataset. As T1 values were unavailable for certain deep grey matter regions, another ratio between their known intensity values and those of regions with defined constants was applied to the scaling ratio.

Noise Addition

To better simulate one of the essential characteristics of real MRI images, their signal-to-noise ratio (SNR), noise was added to the synthetic images to approximate the SNR typically observed at ultra-low field strengths. The noise level in an ULF image was estimated by computing the standard deviation within a background region of an image acquired with the Hyperfine 64mT scanner. The same procedure was then applied to each image in the

Table 3.3 T_1 relaxation time as a function of magnetic field strength for each ROI

ROIs	**T_1 model**
Caudate nucleus	$T_1 = 0.954(B_0)^{0.325}$
Globus pallidus	$T_1 = 0.664(B_0)^{0.367}$
Putamen	$T_1 = 0.855(B_0)^{0.352}$
Brainstem	$T_1 = 0.459(B_0)^{0.508}$
Thalamus	$T_1 = 0.817(B_0)^{0.357}$
White matter	$T_1 = 0.583(B_0)^{0.376}$
Cerebral cortex	$T_1 = 0.857(B_0)^{0.376}$
Cerebellum WM	$T_1 = 0.583(B_0)^{0.376}$
Cerebellum cortex	$T_1 = 0.857(B_0)^{0.376}$
Hippocampus	$T_1 = 0.817(B_0)^{0.357}$
Amygdala	$T_1 = 0.817(B_0)^{0.357}$
Accumbens area	$T_1 = 0.817(B_0)^{0.357}$
Ventral DC	$T_1 = 0.817(B_0)^{0.357}$
Ventricles	$T_1 = 4.322(B_0)^{-0.006}$
CSF	$T_1 = 4.322(B_0)^{-0.006}$

synthetic BCP dataset. To replicate the noise expected at 64mT, Gaussian noise was added to the synthetic images based on the difference in background standard deviation between the two datasets.

However, unlike qualitative MRI sequences such as T1w images, T1 map images do not directly reflect raw signal intensity. Instead, their voxel intensities represent estimated T1 relaxation times derived from multiple acquisitions. As such, the metric referred to as SNR in this thesis is not a true signal-to-noise ratio, but rather a derived approximation, computed in a way that mirrors conventional SNR estimation (Equation 3.2). For each of the 15 regions of interest, the mean T1 value within the region was divided by the standard deviation of a uniform region, used as an estimate of the image's noise level. Given that T1 maps reflect relaxation properties rather than signal amplitude, this derived SNR cannot be directly compared to that of a 64mT T1w image. Instead, the derived SNR values from the synthetic 64mT T1 maps were compared to those of the available 3T T1 maps to assess whether the overall range of values was consistent with lower field strength and whether the expected

variation across regions was preserved.

$$\text{SNR} = \frac{\mu_{\text{ROI}}}{\sigma_{\text{background}}} \tag{3.2}$$

Downsampling

The resolution of the synthetic image was downsampled using linear interpolation to match the typical resolution of images acquired with the Hyperfine scanner (2.0 x 2.0 x 2.0 mm). The image was then resampled back to its original resolution to facilitate the remaining processing steps, while preserving the loss of information introduced during the downsampling process.

3.1.4 Visual Validation

In the absence of access to real 64mT T1 map MRI images, a qualitative evaluation of the synthetic images was performed by visually comparing them to a real ULF T1 map image. Padormo et al. [82] presented a newborn T1 map image acquired with a 64mT Hyperfine scanner, which was used as a reference for this comparison. The visual assessment focused on prominent image characteristics, including the intensity of major structures, the ability to delineate boundaries between white and grey matter, levels of noise, blurring and general visual resemblance.

3.2 O2 : Reconstruction of High-Field T1w Images from Synthetic Ultra-Low-Field Images

This second objective focuses on reconstructing the synthetic dataset created in the first objective to high-resolution T1w images.

First, the synthetic dataset and the BCP dataset are preprocessed and used to train a 3D U-Net model. The validation of the model is achieved with a cross-validation, which enables the reconstruction and evaluation of the whole dataset. Segmentations are performed and serve, in this objective, as the basis for the evaluation of the reconstruction. A visual representation of the workflow is shown in Figure 3.4.

3.2.1 Pre-processing

Prior to training the U-Net model, a few pre-processing steps were applied to the images to optimize model performance. The ground truth images, consisting of the initial BCP dataset, were skullstripped by applying their corresponding segmentation masks to ensure the images

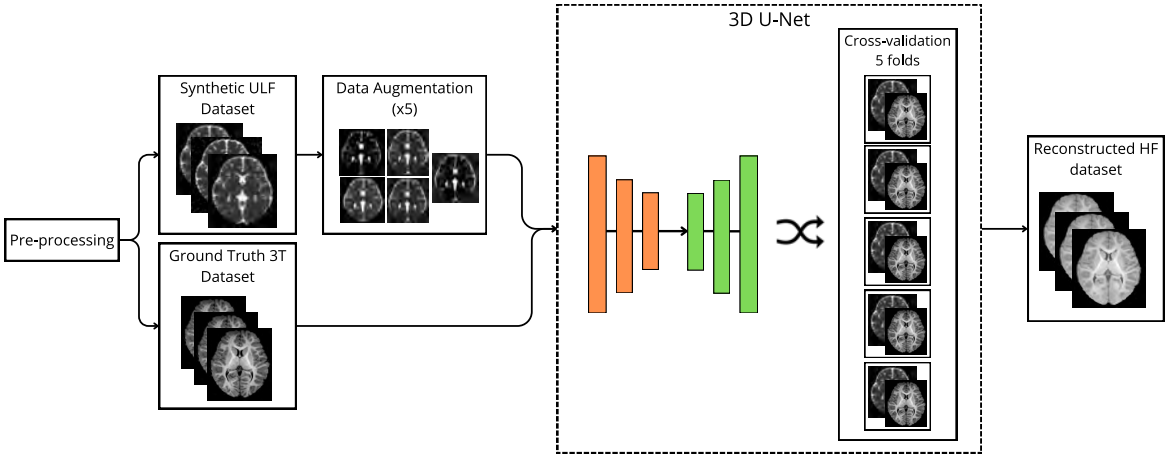

Figure 3.4 Workflow diagram illustrating the main steps of Objective 2.

used in training contained exactly the same structures. The synthetic dataset, which was generated from these same segmentations, by design did not contain the skull.

Both the ground truth and synthetic images were cropped around the brain to a fixed size of 192 x 192 x 160. This cropping step was performed to reduce computational demands during training and ensured that all images had a uniform size to facilitate the training process.

3.2.2 Data Augmentation

To enhance the performance of the model, data augmentation was performed on the synthetic images to increase the size and variability of the training dataset. New synthetic images were created to augment the quantity of data the model was trained on. This process involved generating additional synthetic images by introducing slight variations in the intensity values, allowing the model to better generalize and adapt to a wider range of cases while focusing on the main task of reconstructing brain images from synthetic ULF MRI. Specifically, four new sets of images were created by multiplying the mean intensity of each region by a factor of its standard deviation, as demonstrated in Equation 3.3

$$\mu = [\mu_a + c \cdot \sigma_a,\ \mu_b + c \cdot \sigma_b,\ \mu_c + c \cdot \sigma_c,\ \ldots], \quad c \in \left\{-1,\ -\tfrac{1}{2},\ 0,\ \tfrac{1}{2},\ 1\right\} \tag{3.3}$$

This multiplied the number of available images for training by 5 and brought the total number of training images to 4220.

3.2.3 Model Architecture

In addition to the training dataset, the choice of model and its architecture is another deciding factor in the reconstruction performance. For this project, a model similar to SynthSR [65] was selected as a starting point as it has proven to yield good results in reconstructing similar types of images. The model's hyperparameters were then adjusted to match the characteristics of the data used and optimize the reconstruction quality.

The final 3D U-Net architecture retained the initial five levels, with each level consisting of two convolutional layers. This depth was sufficient to reconstruct the level of detail required for this type of image. A batch size of 2 was selected as a compromise between the available computational resources and the need to support generalization. A dropout rate of 0.1 was added after the layers to reduce overfitting tendencies. Each convolution used 3x3x3 kernels, followed by group normalization and a ReLU activation. Group normalization was chosen instead of batch normalization due to the small batch size and the presence of dropout. After testing, ReLU activation was found to perform better in this context than the eLU used in the original architecture and was therefore adopted as the final activation function. The number of feature channels started at 24 and increased by 12 at each level of the encoder, then decreased by 12 per level in the decoder. These values were optimized through testing and provided sufficient detail in the reconstructed images. The final layer applied a convolution with a linear activation, as is standard practice. The model was optimized with the Adam optimizer and compiled using a mean squared error loss function, which is commonly used in regression tasks such as image reconstruction. The U-Net was implemented in Keras with a TensorFlow backend and trained on an RTX A6000 GPU.

3.2.4 Validation

Validation is a crucial step following the training of a deep learning model to assess whether the model is able to generalize to unseen data and evaluate its accuracy. For this study, the training was validated using a five-fold cross-validation as it allows for the subsequent analysis to be done on all the dataset. Subjects were evenly distributed across folds to ensure balanced representation of all age groups within each fold. All sessions belonging to the same subject were assigned to a single fold, to avoid the risk of testing data from subjects the model has already seen and introduce bias in the performance metrics. The resulting distribution of subjects across the folds is illustrated in Figure 3.5.

For each fold, predictions were generated on the subset of images that were not used during training. Final results were aggregated across all folds to cover the entire dataset for analysis.

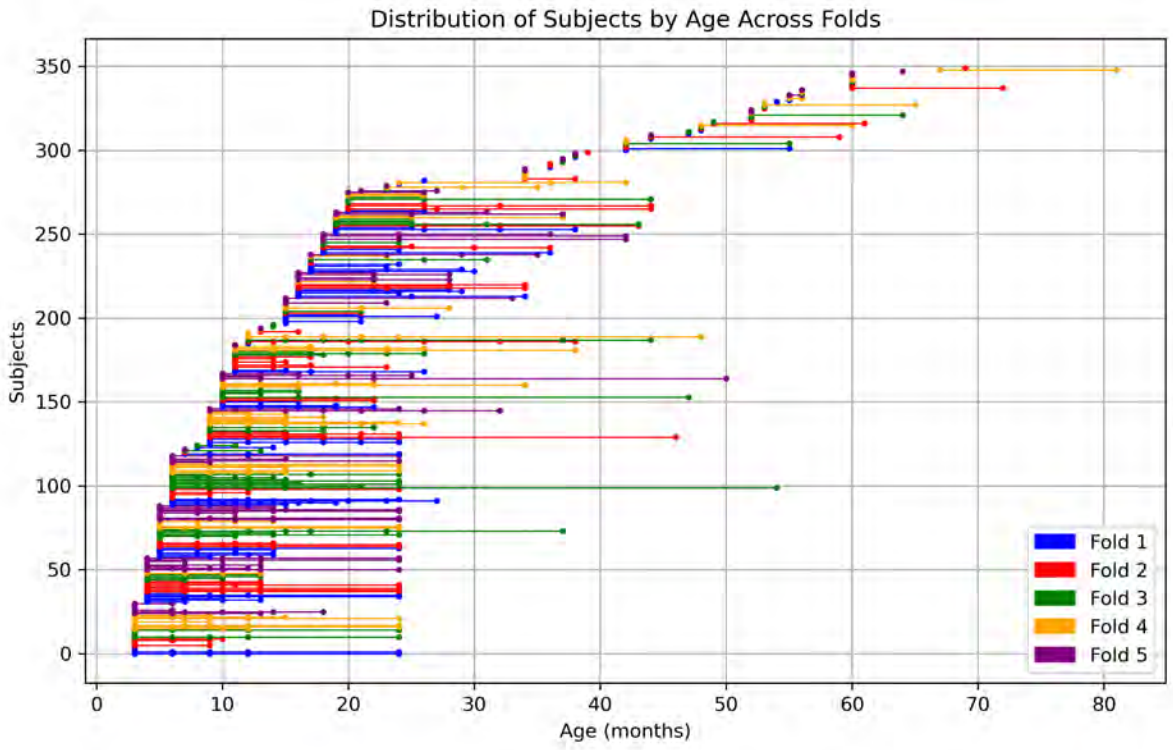

Figure 3.5 Distribution of subjects from the BCP dataset across the five folds used for training and validating the reconstruction model.

The performance of the model was evaluated using the Dice score computed over the regions of interest. Statistical analysis was performed to determine whether the reconstruction method significantly improved segmentation accuracy compared to the synthetic ULF images.

3.3 O3 : Growth Trajectories Validation Method

As the third objective of this project is to create a validation method, the reconstructed images created in the previous section will be used to introduce this validation and show how accurately the previously reconstructed images reflect individual growth trajectories. The method is separated into populational and subject-specific metrics. Each of the metrics measures a different characteristic that is believed to be important in the evaluation of segmentation methods from a neurodevelopmental standpoint. They are separated as follow:

- Populational metrics:
  - Growth curves: Assess the impact of the method on the overall population growth

trajectories of the four main brain structures.

  - Sub-cortical volume ratios: Evaluate the ability of the method to capture differences in growth rates on smaller structures.

- Subject-specific metrics:

  - Z-score differences: Measure the extent to which the method alters each subject's deviation from its expected trajectory.
  - Percentile differences: Quantify the change in percentile position resulting from the method for each subject.

3.3.1 Growth Curves

Growth curves are often used to portray how brain structures evolve across age at a population level. In this study, growth curves are used to model the trajectories of the cerebral volumes for 4 structures of interest. The labels provided by the segmentation method SynthSeg [56] have been aggregated into these structures as follows:

- White Matter: cerebral white matter, brain stem, cerebellum white matter, and ventral DC
- Deep Grey Matter: thalamus, caudate, putamen, pallidum, accumbens area
- Cortical Grey Matter: cerebral cortex, cerebellum cortex, hippocampus, amygdala
- Ventricles: lateral ventricle, inferior lateral ventricle, 4th ventricle, 3rd ventricle

The CSF was excluded from the analysis due to low segmentation performance, which could have introduced bias and affected the reliability of the results.

The volumes of all images corresponding to these regions were plotted as a function of age and modeled using a Generalized Additive Model (GAM) to characterize the growth trajectories. This model was chosen as it was more robust to the limited amount of data and was found to best capture the growth trends compared to interpolated splines, polynomial fits, or rolling means. The same methodology was applied to generate the growth curves from the segmentation results of the predicted images. The growth trajectories derived from the initial high-field dataset is then used as the reference standard against which the performance of the proposed method is evaluated.

3.3.2 Sub-Cortical Regions Growth

A similar approach was applied to subcortical regions of interest to evaluate the performance of the method on specific regions. This was done to determine whether the reconstructions allowed the segmentations to capture region-specific changes in volume as the structures grow, rather than simply applying a uniform scaling of the brain labels. The regions included in this analysis were chosen because their growth rates differ both from one another and from the overall brain volume. Their volumes were normalized by the total brain volume to assess how regional growth deviates from total volume growth. The analysis focused on the putamen, caudate, hippocampus, thalamus, pallidum, and amygdala, following [15].

3.3.3 Z-score

To quantify subject-specific trajectories, a modified version of the Z-score was developed. This metric represents the deviation of an individual subject relative to the expected growth trajectories of the studied population and can be used to evaluate the performance of a model in relation to these trajectories. It relies on the volume of a given brain region for a subject at a specific age and the GAM population fit described in Section 3.3.1. The standard deviation for each age group was calculated across the entire population and then fitted using a sliding window approach. The fit was performed using only non-zero standard deviations to avoid biased estimates and used a window size of 15. This step was applied to preserve the overall trend while reducing fluctuations caused by the small sample sizes in certain age groups. The relationship can be expressed using the following equation for a subject i at age j:

$$Zscore_{i,j} = \frac{volume_{i,j} - \mu_{\text{population fit},j}}{\sigma_{\text{population fit},j}} \tag{3.4}$$

Where $volume_{i,j}$ is the volume of the desired region for the subject i at the time point j, $\mu_{populationfit_j}$ represents the volume interpolated from the GAM fit at this specific age and $\sigma_{populationfit_j}$ represents the fitted standard deviation of the population at the age j.

Z-Score Difference

After computing the Z-score for all subjects, the resulting values are compared to the Z-scores derived from the ground truth to evaluate the deviation of each subject's metrics from its own expected developmental path. Since the ground truth is assumed to reflect both intra-subject variability and the variability introduced by the segmentation method, comparing the two Z-scores provides an estimate of the additional deviation caused by the method.

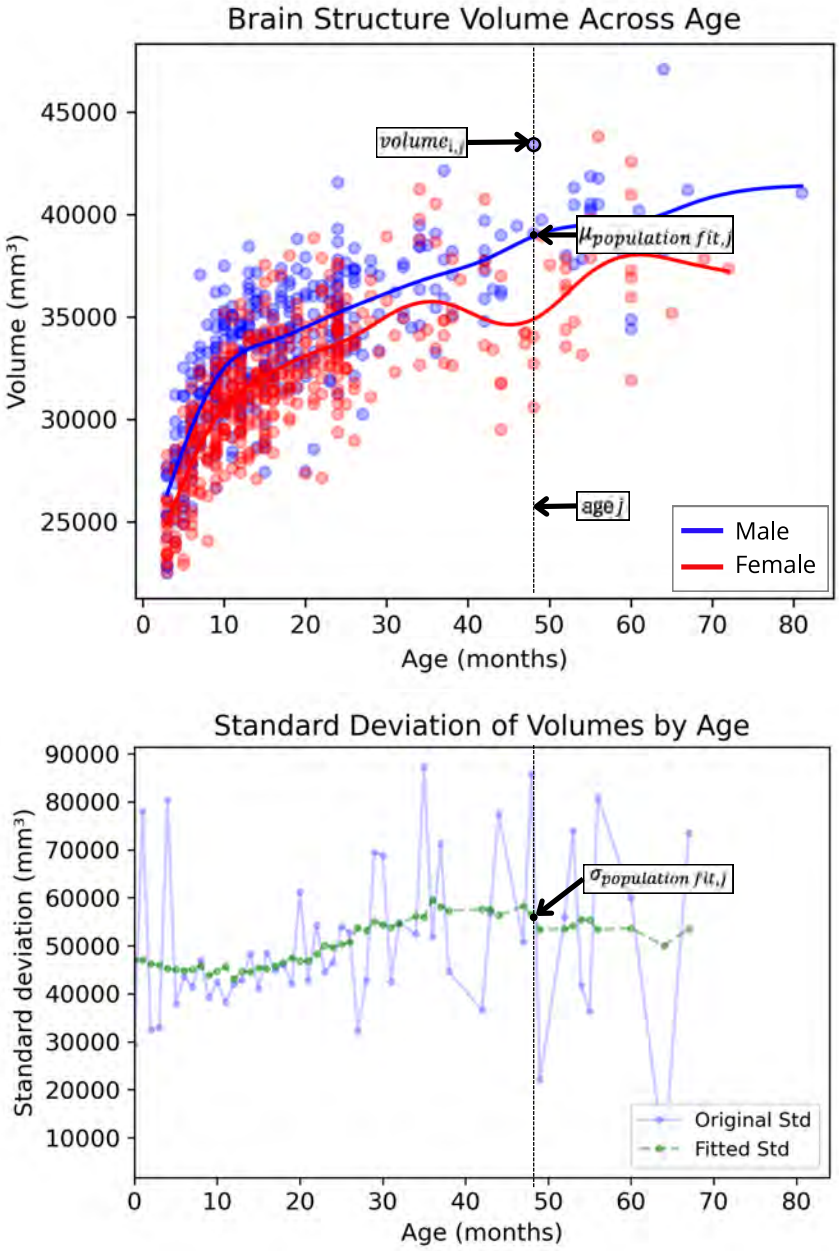

Figure 3.6 Visual representation of the variables defined in Equation 3.4.

Another important consideration in the analysis is that each subject follows its own growth trajectory. As a result, the observed variation must be interpreted relative to this individual trajectory. A subject's Z-score can naturally change between two time points, as growth is not constant throughout development. This needs to be taken into account because directly measuring the difference between the Z-scores at two time points does not accurately repre-

sent the deviation of interest; what is desired is the added deviation from the subject's natural trajectory. This natural trajectory, similar to the percentile that is often used in clinical practice, was estimated by averaging the subject's Z-scores across all sessions. This provides an estimate of the subject's general position relative to the population growth curves, indicating, for example, whether they consistently fall within a higher or lower range in terms of brain volume. The complete formulation of this metric, incorporating all these considerations, is presented in Equation 3.5.

$$\Delta Z_i = [(Z_{i,\text{recon}} - \mu_{i,\text{GT}}) - (Z_{i,\text{GT}} - \mu_{i,\text{GT}})] \tag{3.5}$$

Figure 3.7 provides a visual representation of the intended variation in the Z-score trajectories for an example subject, and illustrates how the mean Z-score estimation is used.

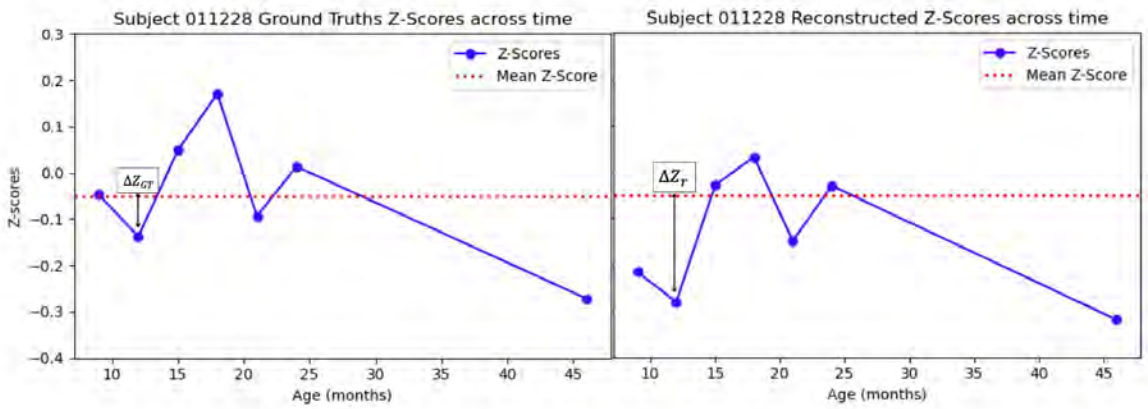

Figure 3.7 Definition of the variable representing the Z-score difference, illustrated with data from subject 011228.

This difference in Z-scores can also be used on the synthetic ULF data. This allows for a comparison of the influence of the reconstruction compared to the deviation that is expected directly from the ULF data.

3.3.4 Percentile Change

In the previous section, the concept of mean Z-score was introduced to represent the individual position of the subject along the growth curve, similar to the percentile used in clinical neurodevelopmental evaluation. Since Z-scores can be directly converted into percentiles, this measure can offer a more intuitive understanding of a subject's developmental position. Comparing the percentile obtained from the reconstruction method to that derived from the ground truth can provide insight into how the method may influence this clinical indicator.

Additionally, comparing this difference to the percentile derived from ULF data can help evaluate how the reconstruction has influenced meaningful growth patterns. The mean Z-scores were directly converted into percentiles using the Gaussian cumulative distribution function (CDF). The difference between the percentiles is shown in the form of the mean Euclidean difference as given in Equation 3.6:

$$D_a = \frac{1}{N_a} \sum_{i \in S_a} \sqrt{(x_i - y_i)^2} \tag{3.6}$$

The Euclidean distance between each pair of corresponding points from the ground truth and the reconstructed dataset is first computed for the four regions of interest. Then, for each age range, the mean distance across all subjects within that range is calculated. This approach accounts for the unequal number of subjects across age ranges and avoids the bias that could be introduced by simply aggregating distances without considering the size of each group. The same method is applied to the synthetic ULF dataset to enable a direct comparison between the two sets of results.

CHAPTER 4 RESULTS

This chapter presents the results of the objectives of this project. First, the synthetic ultra-low-field (ULF) dataset is presented along a comparison to key metrics previously identified from real ULF MRI images. Next, the reconstructed high-resolution images are introduced, and their corresponding segmentations are evaluated against the ground truth. Finally, these findings are used to support the proposed neurodevelopmental validation method.

4.1 Synthetic Images

Figure 4.1 illustrates the progression of synthetic images across multiple time points for the same subject. Visually, the images preserve the morphometric characteristics of the brain as it grows, with the overall shape, sulcal patterns, and volume changes remaining consistent at each developmental stage, highlighting the ability of the synthetic dataset to accurately capture structural growth trends over time. Padormo et al. [82] presented a T1 map image of a newborn acquired on a 64mT scanner, which shows visual characteristics very similar to the images in Figure 4.1, including the hyperintensities of the CSF and the difficulty in distinguishing the structures and the boundaries between white and grey matter. For this reason, the change in contrast that occurs in young infants cannot be easily discerned visually in this figure. Nevertheless, these contrast changes are captured in the voxel intensity values, even if they are not perceptible to the naked eye. The images were downsampled to a resolution of 2x2x2 mm, which matches the typical acquisition resolution of a 64 mT ULF scanner. This explains the reduced sharpness and the blurred tissue boundaries observed, which are consistent with real ULF acquisitions.

4.1.1 T1 Values

A critical aspect in generating the synthetic database is accurately modeling the evolution of T1 relaxation times across age. As the brain develops, microstructural changes influence these relaxation times, which are reflected in MRI signal intensities. To produce realistic longitudinal images, it is essential that these age-related variations are properly captured. Figure 4.2 illustrates how the intensities in the synthetic T1 map images evolve with age within the same subject. A closer examination reveals that white matter exhibits more rapid changes over time compared to grey matter, consistent with the trends identified in Section 3.1.6.1. However, the degradation process simulating 64mT acquisition introduces a more

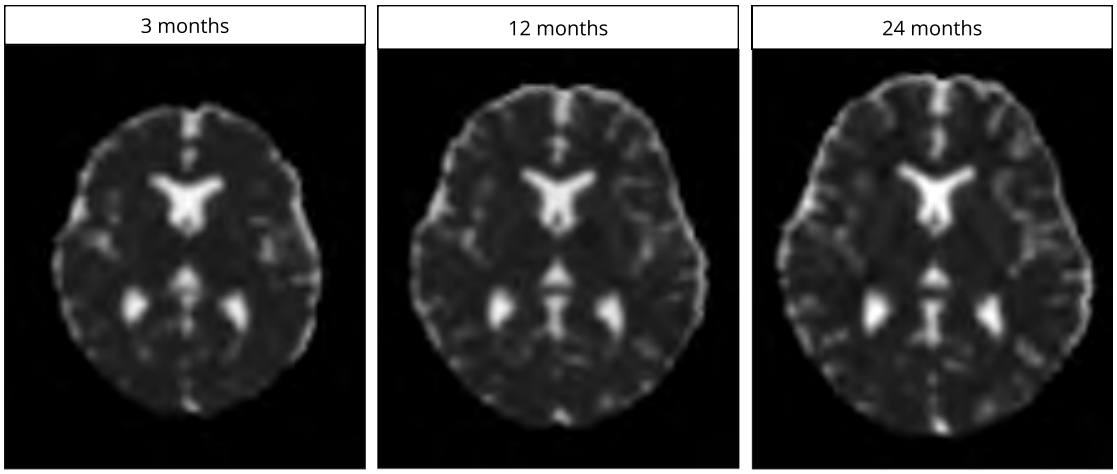

Figure 4.1 Evolution of synthetic images over time for subject 231205 from the BCP dataset. Pixel intensities represent synthetic T1 values, ranging from 0 (black) to 2.5 s (white).

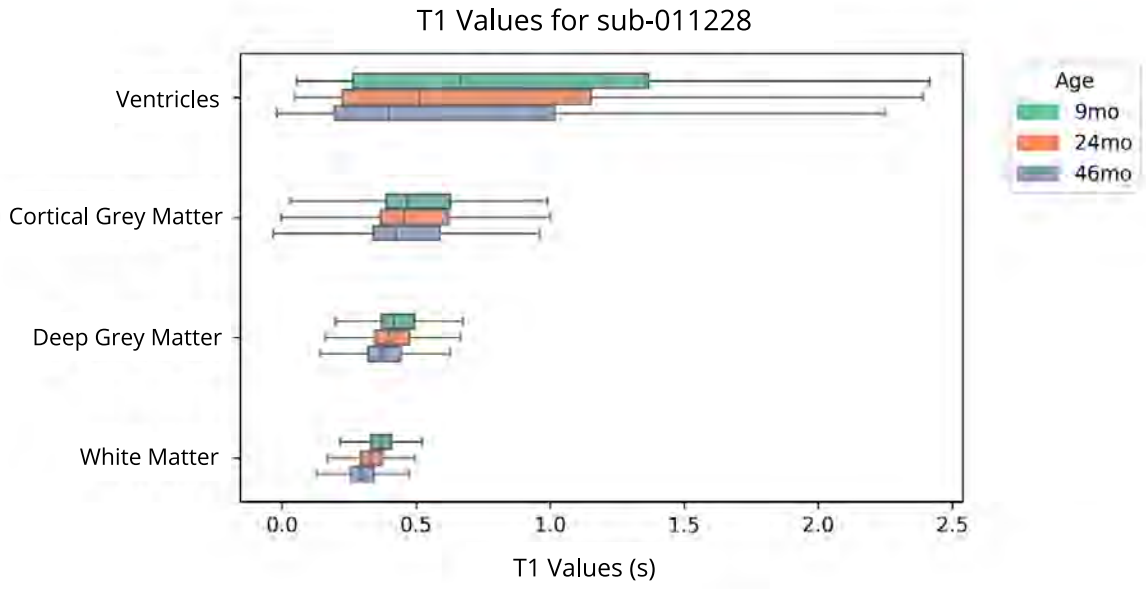

Figure 4.2 Evolution of T1 values across three time points for the synthetic 64 mT T1 map images of subject 011228.

subtle variation of contrast between the two structures. The ventricles show the highest intensities of the four regions, which matches what we see in Figure 4.1, where they appear much brighter than the other structures. Even though their imputed values stayed the same with age, partial volume effects in the segmentations could explain the slight changes in values over time.

4.1.2 SNR Values

A commonly used metric to characterize MRI images is the signal-to-noise ratio (SNR). Unlike qualitative MRI images such as T1-weighted scans, T1 map images do not represent raw signal intensity. The metric computed in this section approximates the SNR by taking the ratio of the T1 relaxation time to the noise present in the image. Figure 4.3 compares the derived SNR between the synthetic ULF images and the original high-field T1 map images across the 15 regions defined earlier. As expected, the synthetic ULF images exhibit substantially lower SNR values than their high-field counterparts, in line with the observations reported in the literature [51]. The ventricular region shows the highest SNR, which is consistent with their notably higher intensity values in both datasets.

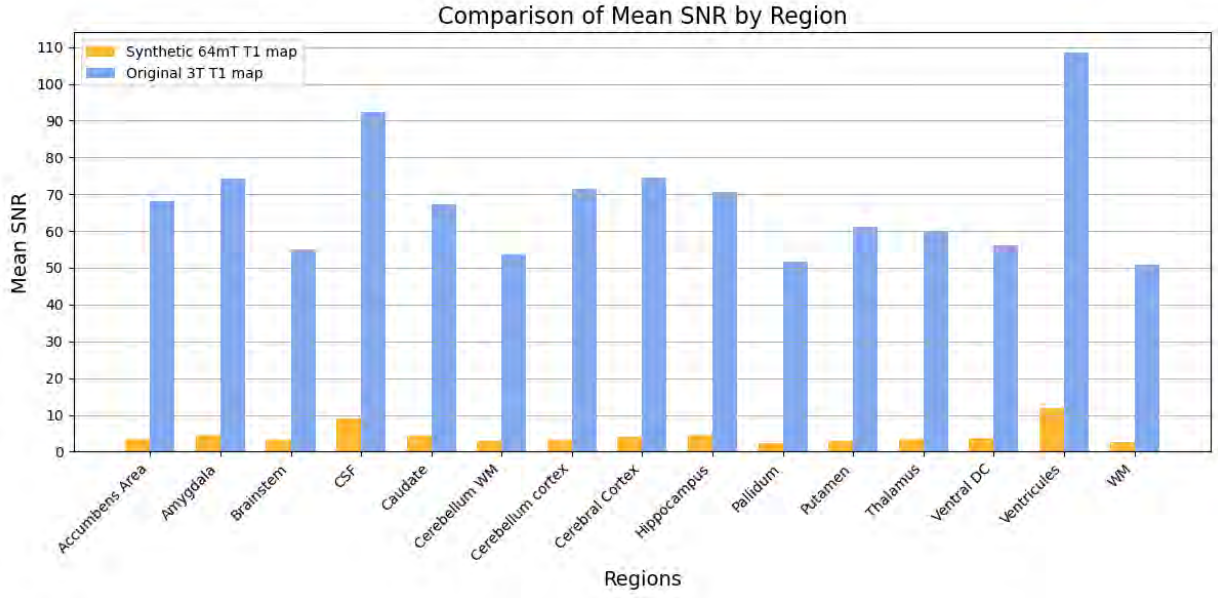

Figure 4.3 Comparison of derived SNR between the synthetic 64 mT dataset and the original 3T T1 map across the 15 regions of interest.

4.2 Reconstructed Images

This section presents the reconstructed images generated by the deep learning model using the synthetic ULF dataset introduced in the previous section. Figure 4.4 shows the reconstructed images for the same subject presented in Figure 4.1. For the younger ages, the reconstructions have more difficulty clearly defining the boundaries between white and grey matter. This is partly due to the naturally reduced contrast between these tissues in early brain development, which already poses a challenge in the original high-field images. The

effect is further amplified by the Gaussian method used to create the synthetic images. This method tends to homogenize tissue intensities, making it even more difficult to distinguish between structures, particularly for the subjects aged closer to the point of contrast inversion. As a result, the reconstructed images deviate more from the original images at these early time points, which is the case for other subjects in the same age range in the dataset. For older subjects, the reconstructed images better preserve the structural characteristics of the original images and show a closer match in contrast. The cerebellum and occipital cortex, however, appears less detailed and can introduce slight variations in the sulcal pattern, as shown in the figure. These differences may influence the subsequent analysis of this region in the reconstructed images.

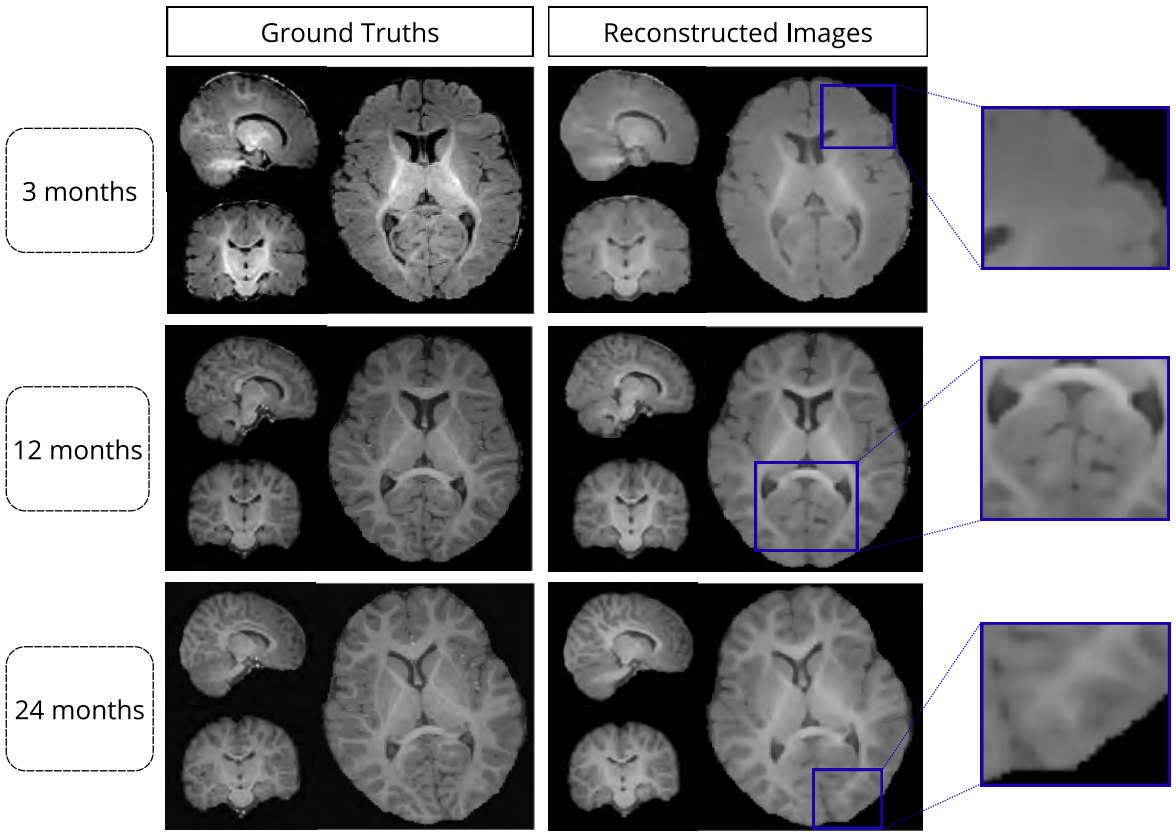

Figure 4.4 Comparison of reconstructed and ground truth images across multiple time points for subject 231205 from the BCP dataset.

4.2.1 Segmentations

To assess how well the reconstructed images reproduce neurodevelopmental metrics, the synthetic and reconstructed datasets were segmented using the same approach as the original images. As shown in Figure 4.5, and in accordance with the observations from figure 4.4, segmentation performance for the reconstructed images is lower in younger subjects. This effect can be attributed in part to the reduced contrast at this age, which already affects the segmentation accuracy for the ground truth images and also slightly more the reconstructed images as discussed in the previous section. The discrepancy can also be attributed to the morphological differences of younger brains which are underrepresented in the training data used by the segmentation method. Visually, it is clear the performance of the model improves as subject age increases. A quantitative evaluation of the segmentation performance is described in the following section, using the dice score as the metric.

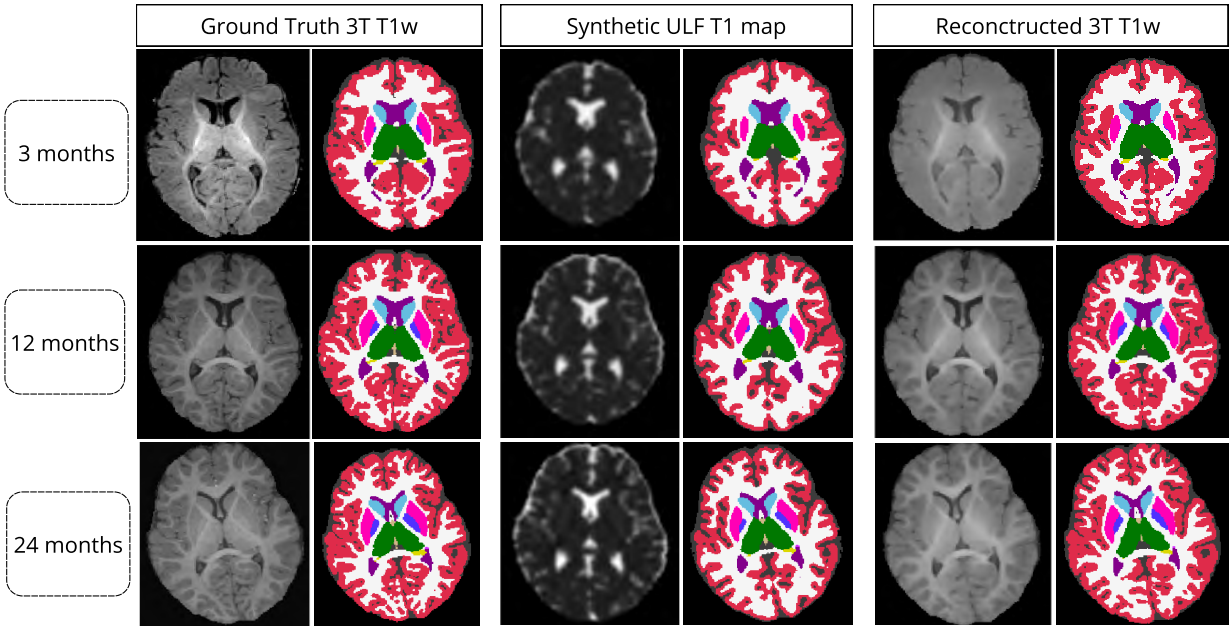

Figure 4.5 Segmentation results for subject 231205 from the BCP dataset at 3, 12, and 24 months, comparing reconstructed images, synthetic ultra-low-field images, and ground truth segmentations.

4.2.2 Dice Scores

The dice score measures the overlap between the predicted segmentation and the reference segmentation, providing a value that reflects the accuracy of the prediction. Values range from 0 to 1, with higher scores indicating a greater degree of overlap. To quantitatively compare the segmentations of the reconstructed images and the synthetic ULF images, dice scores were computed for the segmentations of both datasets across the four structures of interest, as shown in Figure 4.6.

A Shapiro-Wilk test confirmed that the distributions of dice scores were not normally distributed. Consequently, a Wilcoxon test was used to compare the two distributions, revealing statistically significant differences for all four structures, even after applying a Bonferroni correction to control for family-wise error. Overall, the reconstructed images yielded higher segmentation accuracy for white matter, cortical grey matter, and ventricles, while the synthetic ULF images performed slightly better for deep grey matter. However, despite the statistical significance, the median dice score differences for deep grey matter and white matter were relatively small, below 0.0006 and 0.008 respectively, suggesting limited impact between the methods for these structures.

Among the four structures analyzed, cortical grey matter achieved the highest segmentation accuracy, with a mean dice score of 0.90, representing a slight improvement over the 0.87 obtained with the synthetic images. White matter showed comparable performance to the synthetic images, with both achieving a dice score of 0.84. Deep grey matter was the only structure to perform slightly worse than the synthetic images, with a dice score of 0.87 compared to 0.88. In contrast, the ventricles demonstrated a substantial improvement, with an overall dice score of 0.88 compared to 0.78 for the synthetic dataset. These results suggest that performing reconstruction prior to segmentation can improve segmentation accuracy for cortical grey matter and ventricles, while producing relatively similar outcomes for white matter and deep grey matter. Figure 4.6 also shows that dice scores are slightly lower for subjects under 12 months of age, then remain relatively stable across the rest of the age range. This decrease in performance for younger subjects is consistent with the reduced tissue contrast discussed in the previous section.

Since the analysis of the reconstructed images was conducted using cross-validation, Table 4.1 presents the dice scores for each fold to assess the variability introduced by the fold division. While the folds were designed to be evenly distributed by age, the results also indicate a relatively uniform distribution of segmentation performance across folds.

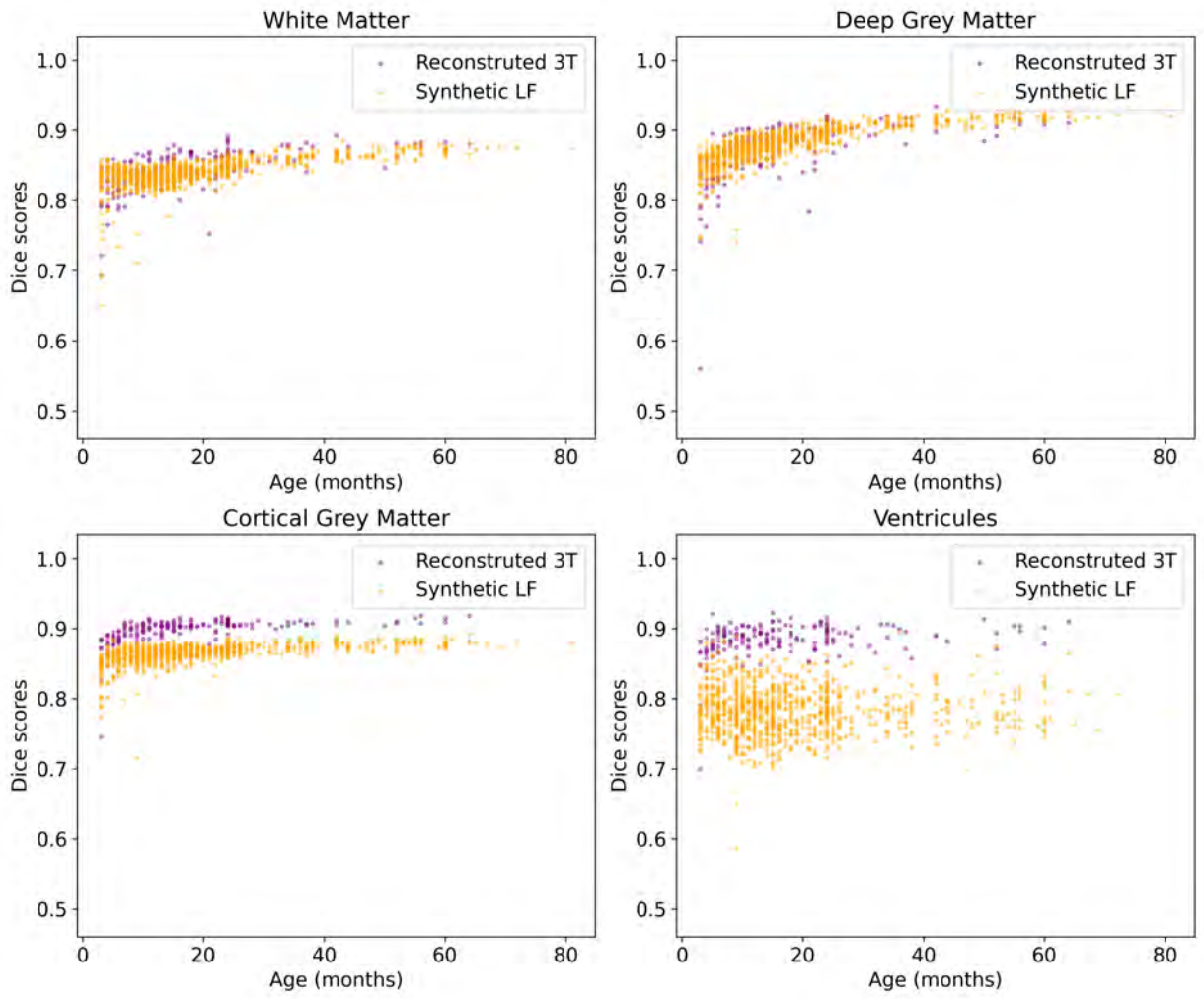

Figure 4.6 Comparison of Dice scores for white matter, deep grey matter, cortical grey matter, and ventricles between segmentations from the reconstructed (purple) and synthetic ULF (orange) datasets.

Table 4.1 Dice scores for each structure across the five folds, along with the overall Dice score for the full dataset and the synthetic ULF images.

White Matter

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Total	Synthetic
Mean	0.8382	0.8387	0.8518	0.8489	0.8438	0.8443	0.8406
Std	0.0353	0.0312	0.0291	0.0308	0.0301	0.0313	0.0203

Deep Grey Matter

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Total	Synthetic
Mean	0.8704	0.8669	0.8817	0.8805	0.8743	0.8748	0.8812
Std	0.0477	0.0506	0.0465	0.0361	0.0415	0.0445	0.0243

Cortical Grey Matter

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Total	Synthetic
Mean	0.8960	0.8938	0.9008	0.8985	0.8980	0.8974	0.8652
Std	0.0224	0.0149	0.0117	0.0168	0.0179	0.0167	0.0144

Ventricles

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Total	Synthetic
Mean	0.8837	0.8757	0.8922	0.8840	0.8853	0.8842	0.7827
Std	0.0339	0.0242	0.0283	0.0204	0.0229	0.0259	0.0339

4.3 Validation of Neurodevelopmental Trajectories

This section presents the results of the reconstructed images discussed in the previous section, now evaluated in terms of their alignment with neurodevelopmental trajectories. As the final objective of this project was to develop a trajectory-based validation method, the results shown here demonstrate how the proposed validation method operates when applied to this concrete example. At the same time, this provides an additional assessment of the reconstructed images based on their ability to reflect meaningful developmental patterns.

4.3.1 Growth Curves

The first step in this validation method is to directly compare the growth trajectories of the reconstructed images to those of the ground truth. As can be visually observed in Figure 4.7a, the reconstruction tends to under-segment grey matter and over-segment white matter relative to the ground truth, while ventricular volumes remain relatively stable. Despite these discrepancies, the overall age-related trends for both male and female subjects are well preserved across all four structures. This analysis provides a rapid way to assess whether the reconstruction method captures the major developmental patterns present in the ground truth trajectories. However, as illustrated in Figure 4.7b, the amount of data available for analysis significantly drops after 24 months, which may explain some of the deviations observed, particularly in regions where the fitted trajectories appear less representative of expected developmental trends.

4.3.2 Sub-Cortical Regions Growth Curves

After examining the trends in the four main brain structures, it is valuable to take a closer look at the smaller deep grey matter structures, which are often of particular interest in neurodevelopmental studies [15]. These structures are noteworthy not only for their developmental relevance, but also because their growth does not necessarily follow the same trajectory as overall brain volume. Since segmentation and reconstruction methods may rely on global scaling and fail to account for regional differences in growth rates, this analysis serves to validate whether the reconstruction method accurately reflects the development patterns of these specific regions.

To this effect, Figure 4.8 presents the growth rate across age for each structure, normalized by the total brain volume, to better capture how the method represents regional growth differences in comparison to overall brain growth. These growth rates are then compared to those of the synthetic ULF images to evaluate whether the synthetic reconstruction offers

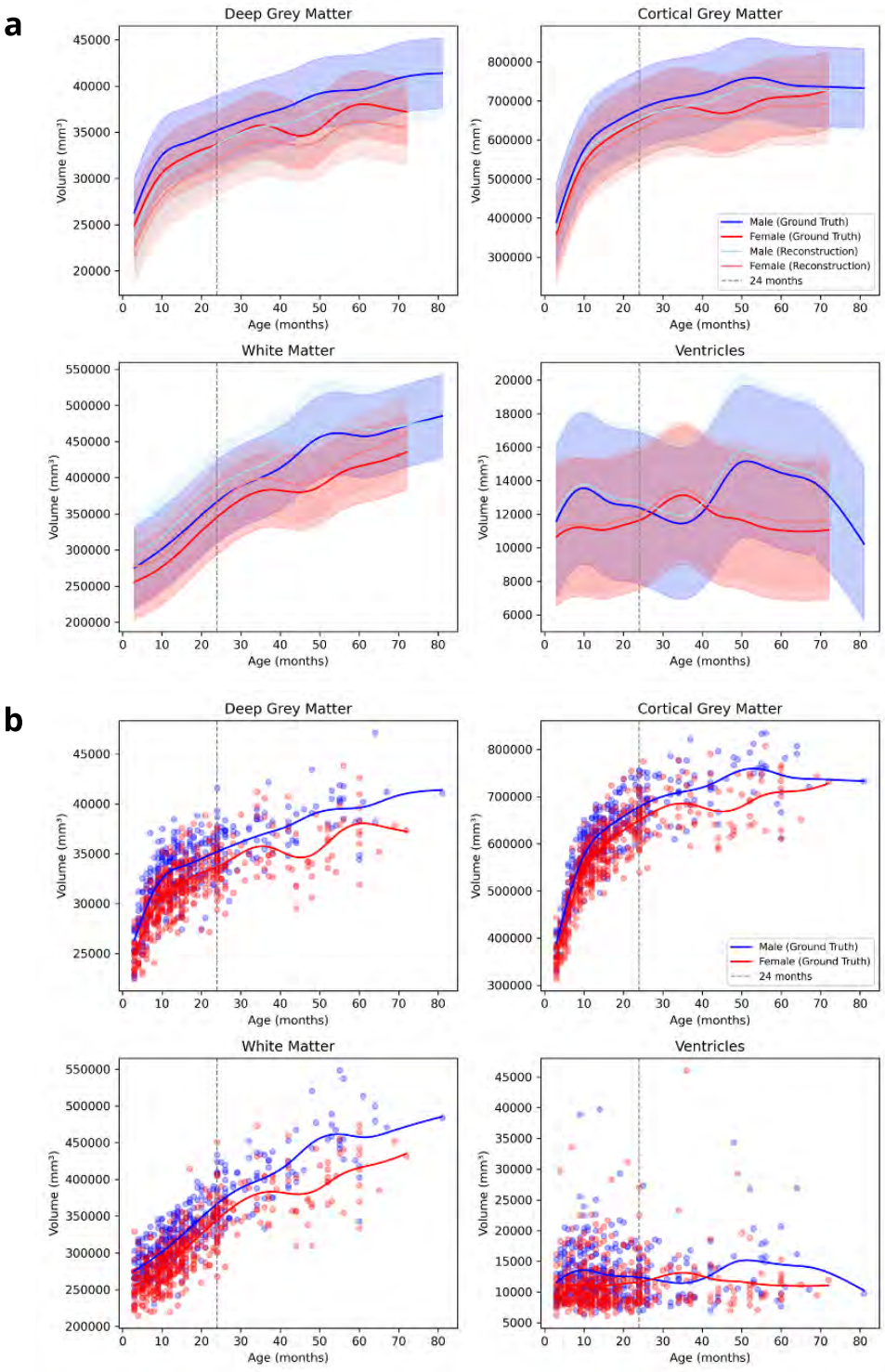

Figure 4.7 (a) Growth curve comparison for the four structures of interest, showing reconstructed images versus ground truth, with results separated by sex. (b) Data distribution of the original dataset, highlighting limited data availability after 24 months.

improvements in this intricate aspect. As shown in Figure 4.8, the reconstructed images (purple) follow the expected growth trends (green) more closely than the synthetic ULF images (orange) for subjects under 24 months of age in the caudate, putamen, hippocampus, and amygdala, and show similar improvements for the putamen, while performing worse for the pallidum. After 24 months, however, the reconstructed images tend to perform more poorly than the synthetic ULF images. Overall, while the reconstructed images provide a closer approximation to the ground truth for certain structures and age ranges, notable discrepancies in growth rates remain, indicating that further refinement of the reconstruction process is needed.

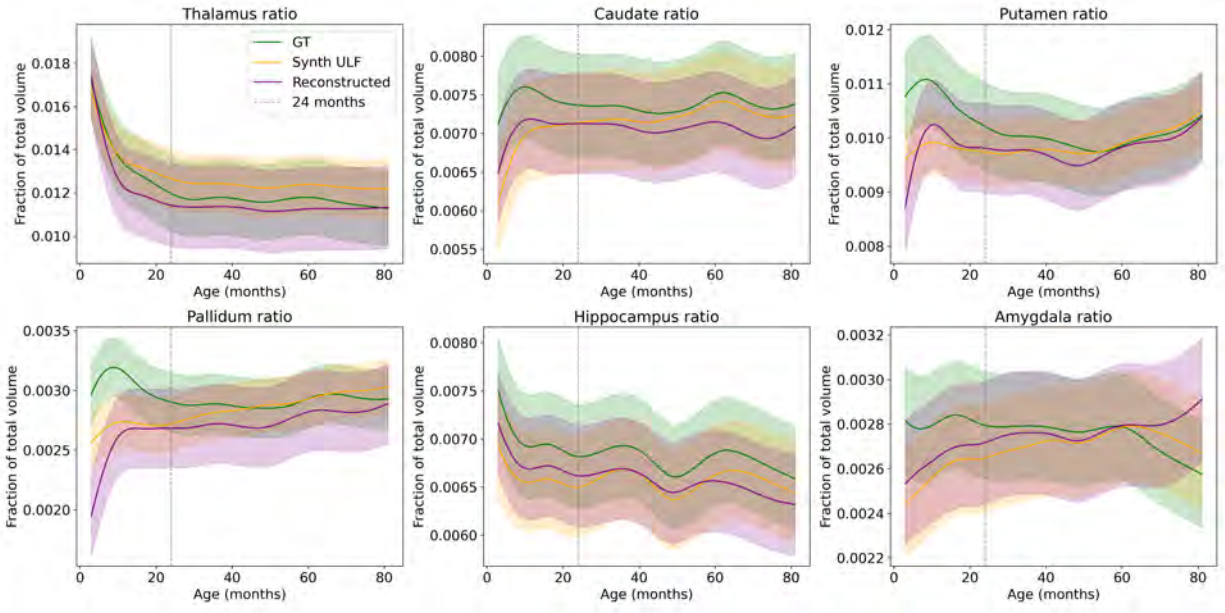

Figure 4.8 Subcortical region growth ratios, normalized by total brain volume, for the ground truth (green), reconstructed images (purple), and synthetic ULF images (orange).

4.3.3 Z-Score Analysis

The quantitative metric used to evaluate how much each reconstructed image deviates from its expected trajectory is the specialized Z-score described in Section 3.3.3. Figure 4.9 illustrates the deviation of reconstructed images from the expected trajectories for each subject who has at least two acquisitions at different time points. Since the ground truth incorporates the subject-specific intra-subject variability in development, this Z-score reflects the additional deviation introduced by the reconstruction process. While segmentation accuracy also contributes to this deviation, this factor is mitigated by applying the same segmentation method to both the ground truth and reconstructed images. Since the Z-scores reflect the

deviation from the fitted trajectories, it is important to note that a mean deviating from zero reflects limitations in the fitting method itself. As shown in Figure 4.9, the mean Z-scores are furthest from zero in the deep grey matter regions, suggesting that the model fit was less representative for these structures. Deep grey matter also exhibits the widest spread of values, indicating greater variability in their deviations. In contrast, the ventricles show both a mean and a distribution of values closest to zero, suggesting that the reconstructed images do not introduce significant additional deviations beyond those already present in the ground truth. Cortical grey matter and white matter display similar distributions, with moderate deviation and variability, indicating a more consistent performance of the fitting method for these tissues and that the reconstruction introduced a moderate level of additional deviation.

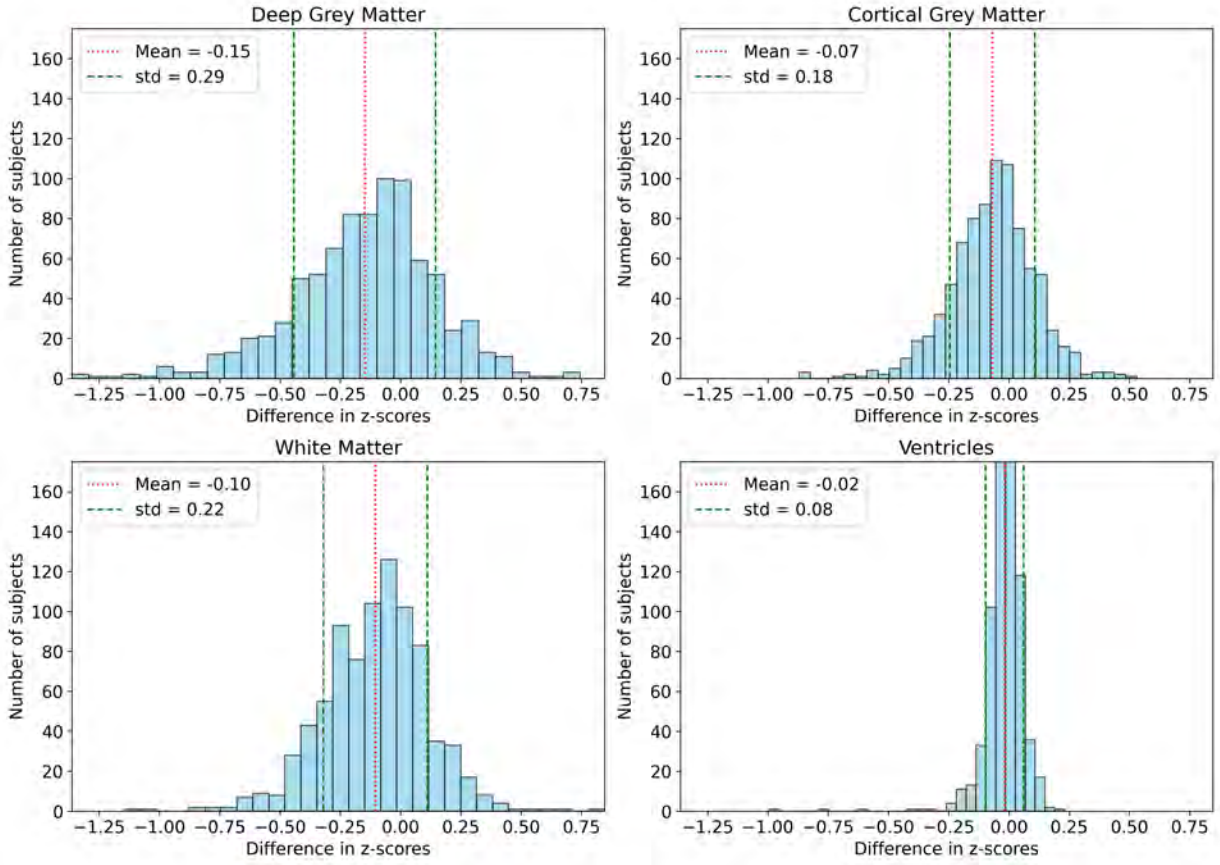

Figure 4.9 Distribution of Z-score differences for the four regions of interest, representing the additional deviations introduced by the reconstruction method from the expected growth trajectories, for subjects with at least two time points.

Comparison with Synthetic Ultra-Low-Field Data

Building on the previous analysis, applying the same approach to the synthetic ULF dataset reveals similar trends across the four regions of interest, as shown in Figure 4.10. The means of all structures are closer to zero, indicating that the ground truth fit better captures the variability in this type of data. Additionally, the standard deviations are slightly smaller across all structures, suggesting, based on this metric, that the segmentations of the synthetic images introduce less deviation from the expected developmental trajectories than those from the reconstructed images.

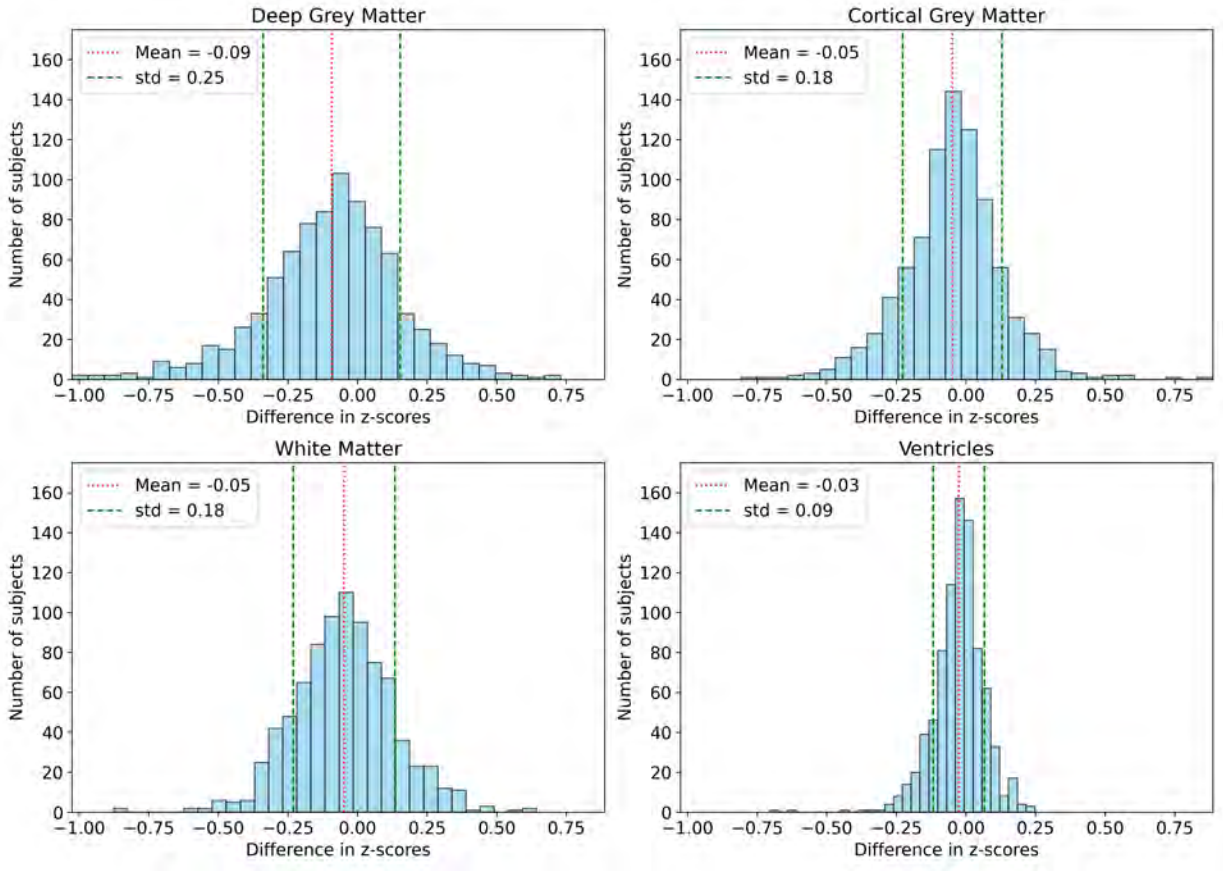

Figure 4.10 Distribution of Z-score differences for the four regions of interest, representing the additional deviations of the synthetic ultra-low-field data from the expected growth trajectories, for subjects with at least two time points.

4.3.4 Variations in Percentiles

The final metric used to quantify discrepancies between the reconstructed data, the ground truth, and the synthetic ULF data is the variation in subjects' percentiles. Since percentiles

are widely used in clinical settings, understanding how these methods affect percentile changes provides valuable insight. Previously, the mean Z-score of each subject across multiple time points was approximated as their intrinsic Z-score, representing the trajectory they would follow in the absence of intra-subject variability. As Z-scores can be easily converted into percentiles, these mean Z-scores were transformed, and the differences between the computed percentiles for each subject were compared to those of the ground truth with a mean euclidean distance for each age.

The distribution of these differences for the four regions of interest is presented in Figure 4.11. Overall, the synthetic ULF dataset exhibits a lower mean Euclidean distance, indicating that its values follow the expected developmental trends more closely than those of the reconstructed images, consistent with the findings of the previous section. However, a more detailed examination reveals that, for subjects under 24 months, the ventricles from the reconstructed images perform slightly better than those from the synthetic dataset. Additionally, the variability in performance across age ranges appears to increase beyond 24 months. In summary, the relative longitudinal performance of each dataset varies depending on the structure and the age range under study. Nonetheless, based on the current metric, the synthetic ULF dataset demonstrates a more consistent alignment with the ground truth growth trajectories.

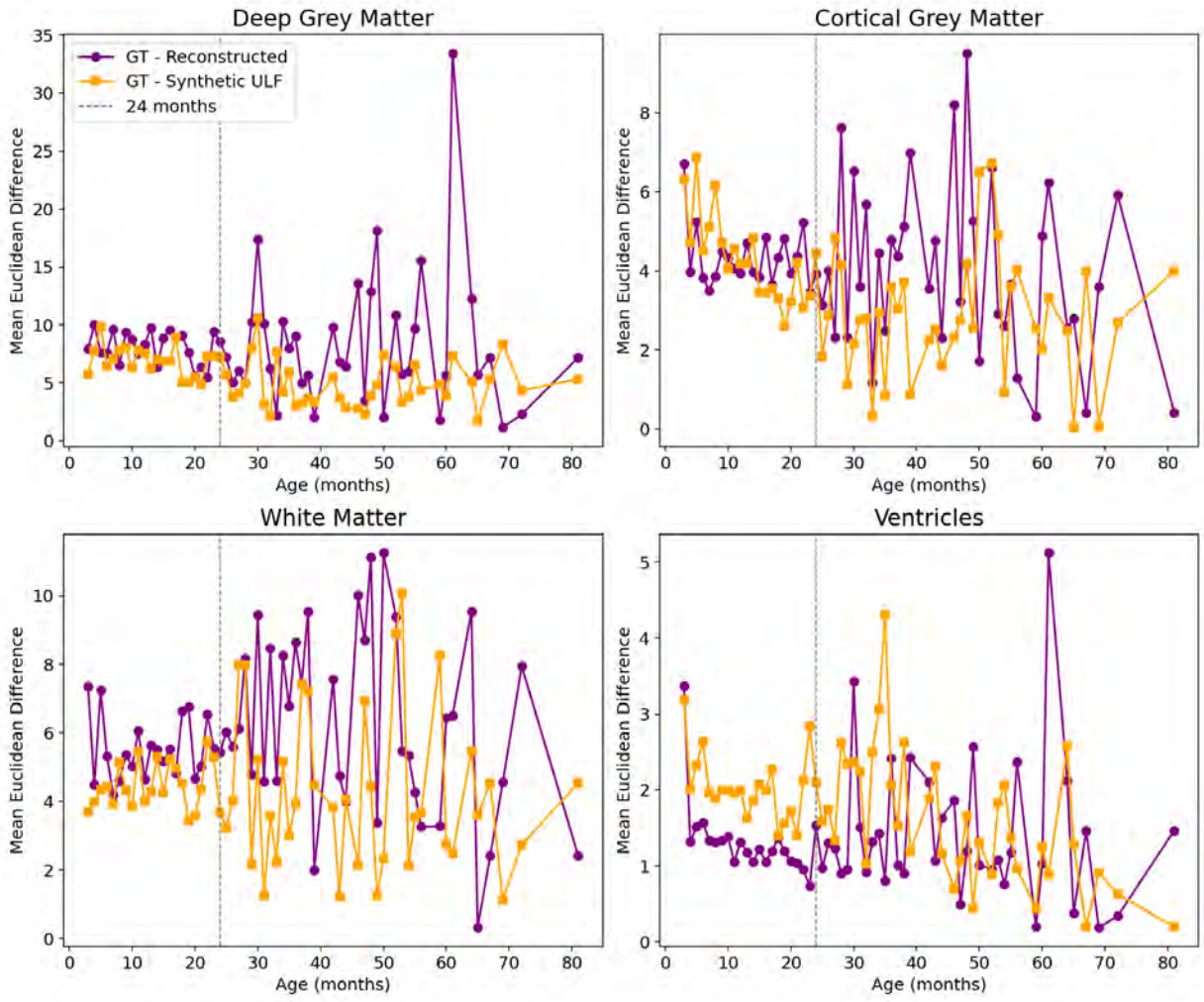

Figure 4.11 Comparison of percentile changes for the four structures of interest between the reconstructed images (purple) and the synthetic ULF data (orange).

CHAPTER 5 DISCUSSION

The main contributions of this project are the development of a method for enhancing ultra-low-field (ULF) brain MRI to facilitate segmentation and the introduction of a validation framework based on longitudinal growth trajectories. To achieve this, a complete pipeline was implemented, covering three objectives: **(1)** the creation of a synthetic dataset that reproduces the main characteristics of ULF pediatric brain MRI, **(2)** the reconstruction of these images into high-resolution images to enable more accurate segmentation, and **(3)** the development of a trajectory-based validation framework to evaluate performance in a longitudinal context. The results, limitations, and potential improvements for each objective are presented in the following section.

5.1 Synthetic Dataset Creation

A synthetic dataset of 64 mT T1 maps was generated from the BCP T1w 3T images using a Gaussian-based approach. The resulting images were then further degraded to replicate the noise characteristics typically observed in ULF acquisitions. The synthetic dataset showed intensity patterns consistent with lower field images while preserving the morphometric growth of the subjects. A visual validation comparing the synthetic images to a real 64mT T1 map image supported their plausibility as ULF synthetic images. However, several limitations related to their creation remain. The T1 values used to generate the images were fitted and approximated from publicly available images and values reported in the literature. These estimates may not accurately reflect true T1 values, as the fitting process introduces errors and represents a simplified version of the tissue properties within an image. Due to limited data availability, additional estimations had to be made for several regions, which could affect the accuracy of the contrast and reduce the overall realism of the synthetic images.

The Gaussian-based method used to generate the synthetic images also entailed significant limitations, especially regarding contrast inversion. As voxel intensities are assigned based on Gaussian distributions within anatomical labels, any overlap in mean values between structures, such as grey and white matter near the age of contrast inversion, makes those structures harder to distinguish. In real MRI images, subtle signal differences often remain detectable even at low contrast, but in synthetic images generated through this method, the tissue distributions become nearly indistinguishable. The results indicate that the Gaussian method, when paired with fitted T1 values, is poorly suited to model contrast inversion and contributes to the lower performance observed for younger subjects. Furthermore, the

intra-tissue variability introduced by the method, by modeling a Gaussian distribution with a fixed standard deviation, does not accurately reflect the variability observed in real images. In real data, variability is influenced by surrounding structures and exhibits a degree of spatial coherence, whereas this method introduces random variability within each structure independently.

In addition, the overall degradation process, involving the adjustments of the T1 values to the desired magnetic field strength, downsampling and adding noise, represent a simplified version of the complex physics that underlie real MRI acquisition. The strength of the magnetic field influences directly the signal that is captured by the MRI scanner and then translated into an image. The influence of magnetic field strength on image contrast and signal is not easily replicable using post-processing alone. The method does not fully replicate noise patterns, inhomogeneities and motion artefacts, typically present in ULF acquisitions, which makes the synthetic images appear cleaner than real ULF data and simplifies the subsequent reconstruction task. This simplification poses challenges when generalizing the results of the reconstruction model to real ULF data.

Furthermore, the visual validation was conducted only on one newborn image and not across the full age range of the dataset, limiting the ability to assess whether age-related changes in brain growth and tissue contrast were accurately reproduced. In addition, due to the lack of access to real ULF T1 maps and limited information available in the literature, the approximated SNR and the evaluation of T1 values could not be directly compared to expected values. Only the overall trends could be assessed, reducing the strength of these comparisons.

Despite these limitations, the pipeline developed in this study provides a valuable tool for use in the absence of paired ULF/high-field MRI data and can serve as a foundation for further analysis and testing of image-processing pipelines. As more data becomes available, the approach can also be refined to produce more realistic synthetic datasets. Overall, the dataset successfully reproduces key information such as brain morphometrics and major ULF image characteristics and was sufficient to assess the performance of the pipeline when applied to data resembling ULF MRI.

One possible improvement would be to acquire images of subjects across different age ranges using the portable ULF scanner. These acquisitions could then serve as more representative priors for the Gaussian-based synthetic image generation method. This approach would still avoid the need for a fully paired dataset but would significantly improve the realism of the resulting contrast, which would be based on actual ULF image characteristics rather than simulated T1 values. Producing synthetic images that more closely reflect the contrast of

actual ULF acquisitions would bring the pipeline closer to the level of realism required for generalization to real-world data. Access to such data could also support more accurate refinement of image metrics, such as SNR, and provide a more realistic basis for comparison. Furthermore, incorporating a broader range of deformations during image synthesis could further enhance realism by better modeling the variability typically observed in ULF acquisitions.

5.2 Image Reconstruction

The synthetic 64 mT T1 map images were reconstructed into 3T T1w images using a 3D U-Net specifically optimized for this task. The resulting reconstructions showed strong similarities with the ground truth high-field images. They produced visually comparable contrasts and preserved the overall morphometrics, achieving dice scores above 0.84 for the main structures, with cortical grey matter reaching the highest score of 0.90. The ventricles also benefited from the reconstruction step, with the dice score improving to 0.88 compared to 0.78 for the synthetic images. These results are promising, especially given the clinical relevance of cortical grey matter in neurodevelopmental research. However, the reconstructions reproduced fewer fine details, particularly in the anterior regions and deep grey matter. Reconstruction quality was also lower in younger subjects, which can be explained in part by the reduced tissue contrast of the Gaussian synthesis method, but can also be caused by the deep learning strategy used in this study. No separation by age was implemented in the training data or the hyperparameter optimization, despite the age-related differences in brain morphology. Incorporating age-specific considerations could be an important avenue for future work.

The main limitations of the reconstruction method stem from its strong dependence on the synthetic dataset. The model was trained on a single dataset, which limited the anatomical variability it was exposed to and may reduce its ability to generalize beyond the original images. The analysis of the performance of the reconstructions is also influenced by the segmentation method used, as all the segmentations were performed using SynthSeg [56]. While SynthSeg proved reliable overall, inaccuracies and biases were still observed, especially for younger subjects, which could have impacted the reported performance despite the inclusion of a visual quality control step. Another important limitation is that the synthetic images follow a fixed pattern because of the way they are generated. This homogeneity makes the reconstruction task easier for the U-Net model since the degraded and high-resolution images share predictable mappings. Real ultra-low-field images are much more variable in terms of quality, noise, and artifacts.

Overall, the reconstruction method showed promising results in preserving key brain structures and improving segmentation accuracy for several regions, but its dependence on synthetic data and reduced performance in younger subjects highlight the need for refinements. Future work could explore the use of more advanced model architectures to further improve reconstruction performance, in combination with the proposed improvements to the synthetic dataset. Approaches such as generative adversarial networks (GANs), diffusion models, or combinations of multiple models have shown promising results in generating realistic medical images and could be worth investigating in this context. Another avenue would be to investigate whether training age-specific models yields better performance compared to a single model trained across all age groups. Additionally, acquiring real ULF MRI images would be an important next step to assess the validity of both the synthetic image generation process and the quality of the reconstructed images.

5.3 Longitudinal Validation Method

A longitudinal validation framework integrating both population-level and subject-specific metrics was developed and applied to the reconstructions obtained with the proposed method. Analysis of growth curves derived from the reconstructed images revealed a tendency to slightly undersegment deep and cortical grey matter, while oversegmenting white matter and ventricles compared to the ground truth. By highlighting these systematic segmentation patterns introduced by the reconstruction process, the subsequent analysis deriving from these metrics can be adapted accordingly. For the subcortical volume ratios, the caudate, putamen, hippocampus, and amygdala showed improvements compared to the synthetic images for subjects under 24 months of age, while the improvement was less apparent beyond this age range. This suggests that the reconstruction method may be more effective at segmenting certain small regions that do not follow the same growth patterns as overall brain volume. In contrast, the pallidum and thalamus did not show any improvement over the synthetic images. The analysis of z-score deviations confirmed that the reconstructed images generally followed the expected population growth trends, but not more so than the direct synthetic images. In fact, the percentile change metric suggests that the synthetic images deviated less from the ground truth trajectories than the reconstructed ones. In addition, greater variability was observed in the deep grey matter, indicating that the reconstructed images introduced more fluctuations in the expected growth patterns for this structure compared to the population model.

While the longitudinal validation approach provides a valuable framework for assessing how well the reconstructed images preserve neurodevelopmental trajectories, it has two major

limitations. The first is its dependence on population-based growth models. The approach assumes that the reference population accurately represents typical growth patterns, which may not always be the case. As a result, subjects might be considered outliers even if they have valid anatomical variations, therefore potentially biasing the analysis. For this project, the BCP dataset was used as the reference for the populational growth. However, subjects over 24 months of age are underrepresented in this dataset, which may reduce the accuracy of the results for this age range, as there is insufficient data to fully generalize growth trends. Additionally, the dataset provides longitudinal time points that vary in number and in the exact ages across subjects. This inconsistency can bias the comparisons for subject-specific analysis, as the quantity and distribution of data are not uniform between individuals.

The second major limitation is the method's sensitivity to segmentation accuracy. While all images are segmented using the same method, which helps reduce the impact of global segmentation bias when comparing to the ground truth, differences in performance across image types, such as synthetic or reconstructed images, can still influence the validation metrics. For example, consistent over or under segmentation in one image type but not the other could introduce bias into the longitudinal analysis. To better understand and mitigate this effect, the pipeline could be tested with multiple segmentation methods to evaluate how performance varies across datasets and adjust the analysis accordingly.

To further strengthen the approach, it would be valuable to involve clinical and neurodevelopmental researchers in identifying which metrics are most relevant to track in practice. The current framework was developed from an analytical perspective and would benefit from clinical input to ensure its relevance and applicability in real-world contexts.

CHAPTER 6 CONCLUSION

6.1 Summary of works

The main objective of this project was to evaluate whether reconstructing high-resolution pediatric brain MRI images from synthetic ultra-low-field (ULF) data could enhance segmentation accuracy, and to develop a validation method capable of assessing this improvement from a longitudinal perspective. To support this goal, a synthetic dataset of pediatric ULF T1 map images was successfully created. Although the limited availability of data posed challenges, the resulting dataset replicates key characteristics of ULF imaging, making it suitable for evaluating the proposed pipeline on synthetic ULF data. Using this dataset, a deep learning model specifically optimized for reconstruction was applied to generate high-resolution images. The reconstructed images improved segmentation accuracy for cortical grey matter and ventricles, and maintained performance for deep grey matter and white matter. While further work is needed to confirm generalization to real ULF data, these findings suggest a promising approach to improving segmentation accuracy in pediatric ULF neuroimaging studies. Finally, a longitudinal validation framework was proposed to evaluate the ability of segmentation methods to capture expected neurodevelopmental trajectories. When applied to both the synthetic ULF and reconstructed datasets, the validation framework revealed that, overall, the synthetic ULF images more closely preserved developmental trends than the reconstructed images, despite lower segmentation accuracy. However, this advantage varied depending on the specific age range and brain structure examined. These findings demonstrate the utility of the proposed framework for interpreting segmentation results and assessing the impact of image processing on longitudinal analyses.

6.2 Limitations

A major limitation in this project lies in the generation of the synthetic dataset, which served as the foundation for both the reconstruction process and the subsequent analysis. Although a preliminary evaluation of the synthetic images was performed, the absence of real ULF data limited the ability to fully assess their realism. The synthetic generation method, and consequently the reconstruction, was also limited at the point of contrast inversion, an important factor in pediatric brain imaging. In addition, the model was trained on a single dataset without age group separation, potentially reducing its ability to reconstruct age specific features. In terms of validation, all metrics were derived from segmentation volumes produced

by a single method. While this ensured consistency across datasets, it may have introduced bias into the analysis. Furthermore, the longitudinal validation framework relied heavily on these segmentation outputs, and the results assumed that the studied population accurately reflected typical neurodevelopmental trends, which may not be fully representative.

6.3 Future Research

To address these limitations, future work should begin by acquiring a small number of ULF images across different age ranges. This would support the generation of more accurate synthetic images and provide a stronger basis for evaluating their realism. With more diverse and representative training data, more advanced model architectures could also be explored to improve reconstruction quality and better capture anatomical detail. Using age specific models could further improve performance by accounting for developmental variability. Additionally, acquiring paired ULF and high-field images would enable more direct evaluation of anatomical accuracy and allow segmentation performance and developmental metrics to be interpreted relative to real ground truths. Finally, integrating clinical input into the design and selection of validation metrics would help ensure that the outcomes remain relevant and interpretable in real world clinical settings.

Overall, this work presents a promising pipeline for improving segmentation from pediatric brain ULF MRI and introduces a novel longitudinal validation strategy grounded in developmental trajectories. With further refinement and access to real data, this approach has the potential to support more accessible pediatric neuroimaging while ensuring that processing methods remain aligned in neurodevelopmental trajectories.

REFERENCES

[1] Hyperfine, Inc., "The Swoop® Portable MR Imaging® System," [Online]. Available: https://hyperfinemri.com/. Accessed: 2025-08-09.

[2] R. a. I. Bethlehem, J. Seidlitz, S. R. White, J. W. Vogel, K. M. Anderson, C. Adamson, S. Adler, G. S. Alexopoulos, E. Anagnostou, A. Areces-Gonzalez, D. E. Astle, B. Auyeung, M. Ayub, J. Bae, G. Ball, S. Baron-Cohen, R. Beare, S. A. Bedford, V. Benegal, F. Beyer, J. Blangero, M. Blesa Cábez, J. P. Boardman, M. Borzage, J. F. Bosch-Bayard, N. Bourke, V. D. Calhoun, M. M. Chakravarty, C. Chen, C. Chertavian, G. Chetelat, Y. S. Chong, J. H. Cole, A. Corvin, M. Costantino, E. Courchesne, F. Crivello, V. L. Cropley, J. Crosbie, N. Crossley, M. Delarue, R. Delorme, S. Desrivieres, G. A. Devenyi, M. A. Di Biase, R. Dolan, K. A. Donald, G. Donohoe, K. Dunlop, A. D. Edwards, J. T. Elison, C. T. Ellis, J. A. Elman, L. Eyler, D. A. Fair, E. Feczko, P. C. Fletcher, P. Fonagy, C. E. Franz, L. Galan-Garcia, A. Gholipour, J. Giedd, J. H. Gilmore, D. C. Glahn, I. M. Goodyer, P. E. Grant, N. A. Groenewold, F. M. Gunning, R. E. Gur, R. C. Gur, C. F. Hammill, O. Hansson, T. Hedden, A. Heinz, R. N. Henson, K. Heuer, J. Hoare, B. Holla, A. J. Holmes, R. Holt, H. Huang, K. Im, J. Ipser, C. R. Jack, A. P. Jackowski, T. Jia, K. A. Johnson, P. B. Jones, D. T. Jones, R. S. Kahn, H. Karlsson, L. Karlsson, R. Kawashima, E. A. Kelley, S. Kern, K. W. Kim, M. G. Kitzbichler, W. S. Kremen, F. Lalonde, B. Landeau, S. Lee, J. Lerch, J. D. Lewis, J. Li, W. Liao, C. Liston, M. V. Lombardo, J. Lv, C. Lynch, T. T. Mallard, M. Marcelis, R. D. Markello, S. R. Mathias, B. Mazoyer, P. McGuire, M. J. Meaney, A. Mechelli, N. Medic, B. Misic, S. E. Morgan, D. Mothersill, J. Nigg, M. Q. W. Ong, C. Ortinau, R. Ossenkoppele, M. Ouyang, L. Palaniyappan, L. Paly, P. M. Pan, C. Pantelis, M. M. Park, T. Paus, Z. Pausova, D. Paz-Linares, A. Pichet Binette, K. Pierce, X. Qian, J. Qiu, A. Qiu, A. Raznahan, T. Rittman, A. Rodrigue, C. K. Rollins, R. Romero-Garcia, L. Ronan, M. D. Rosenberg, D. H. Rowitch, G. A. Salum, T. D. Satterthwaite, H. L. Schaare, R. J. Schachar, A. P. Schultz, G. Schumann, M. Schöll, D. Sharp, R. T. Shinohara, I. Skoog, C. D. Smyser, R. A. Sperling, D. J. Stein, A. Stolicyn, J. Suckling, G. Sullivan, Y. Taki, B. Thyreau, R. Toro, N. Traut, K. A. Tsvetanov, N. B. Turk-Browne, J. J. Tuulari, C. Tzourio, Vachon-Presseau, M. J. Valdes-Sosa, P. A. Valdes-Sosa, S. L. Valk, T. van Amelsvoort, S. N. Vandekar, L. Vasung, L. W. Victoria, S. Villeneuve, A. Villringer, P. E. Vértes, K. Wagstyl, Y. S. Wang, S. K. Warfield, V. Warrier, E. Westman, M. L. Westwater, H. C. Whalley, A. V.

Witte, N. Yang, B. Yeo, H. Yun, A. Zalesky, H. J. Zar, A. Zettergren, J. H. Zhou, H. Ziauddeen, A. Zugman, X. N. Zuo, E. T. Bullmore, and A. F. Alexander-Bloch, "Brain charts for the human lifespan," *Nature*, vol. 604, no. 7906, pp. 525–533, Apr. 2022, doi:10.1038/s41586-022-04554-y.

[3] R. K. Lenroot and J. N. Giedd, "Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging," *Neuroscience & Biobehavioral Reviews*, vol. 30, no. 6, pp. 718–729, Jan. 2006, doi:10.1016/j.neubiorev.2006.06.001.

[4] M. R. Peterson, V. Cherukuri, J. N. Paulson, P. Ssentongo, A. V. Kulkarni, B. C. Warf, V. Monga, and S. J. Schiff, "Normal childhood brain growth and a universal sex and anthropomorphic relationship to cerebrospinal fluid," *Journal of Neurosurgery: Pediatrics*, vol. 28, no. 4, pp. 458–468, Jul. 2021, doi:10.3171/2021.2.PEDS201006.

[5] T. Paus, "Tracking development of connectivity in the human brain: Axons and dendrites," *Biological Psychiatry*, vol. 93, no. 5, pp. 455–463, 2023, doi:10.1016/j.biopsych.2022.08.019.

[6] R. L. Stephens, B. W. Langworthy, S. J. Short, J. B. Girault, M. A. Styner, and J. H. Gilmore, "White matter development from birth to 6 years of age: A longitudinal study," *Cerebral Cortex (New York, NY)*, vol. 30, no. 12, pp. 6152–6168, Jun. 2020, doi:10.1093/cercor/bhaa170.

[7] R.-K. Chen, M.-Y. Li, Z.-Y. Zhao, H.-A. Xu, C.-L. Ning, J. Lu, Q.-F. Zhu, Y.-Q. Huang, R.-K. Zhao, L.-X. Zhang, T.-L. Tao, K.-B. Tang, J.-L. Cheng, C. Lai, H.-X. Zhang, X.-H. Ma, H. Kim, Z.-H. Ren, C.-F. Lian, F. Wang, G. Chen, M.-H. Ouyang, A.-Q. Qiu, H. Zhang, X.-Q. Huang, X.-J. Duan, G. Li, J. Yang, H. Huang, K. Oishi, E. Grant, W.-L. Lin, D. Xu, and D. Wu, "Advances in magnetic resonance imaging of the developing brain and its applications in pediatrics," *World Journal of Pediatrics*, May 2025, doi:10.1007/s12519-025-00905-7.

[8] A. M. Alex, F. Aguate, K. Botteron, C. Buss, Y.-S. Chong, S. R. Dager, K. A. Donald, S. Entringer, D. A. Fair, M. V. Fortier, N. Gaab, J. H. Gilmore, J. B. Girault, A. M. Graham, N. A. Groenewold, H. Hazlett, W. Lin, M. J. Meaney, J. Piven, A. Qiu, J. M. Rasmussen, A. Roos, R. T. Schultz, M. A. Skeide, D. J. Stein, M. Styner, P. M. Thompson, T. K. Turesky, P. D. Wadhwa, H. J. Zar, L. Zöllei, G. de los Campos, and R. C. Knickmeyer, "A global multicohort study to map subcortical brain development and cognition in infancy and early childhood," *Nature Neuroscience*, vol. 27, no. 1, pp. 176–186, Jan. 2024, doi:10.1038/s41593-023-01501-6.

[9] S. C. Deoni, D. C. Dean, J. Remer, H. Dirks, and J. O'Muircheartaigh, "Cortical maturation and myelination in healthy toddlers and young children," *Neuroimage*, vol. 115, pp. 147–161, 2015, doi:10.1016/j.neuroimage.2015.04.058.

[10] T. K. Turesky, J. Vanderauwera, and N. Gaab, "Imaging the rapidly developing brain: Current challenges for MRI studies in the first five years of life," *Developmental Cognitive Neuroscience*, vol. 47, p. 100893, Feb. 2021, doi:10.1016/j.dcn.2020.100893.

[11] J. B. Girault, E. Cornea, B. D. Goldman, R. C. Knickmeyer, M. Styner, and J. H. Gilmore, "White matter microstructural development and cognitive ability in the first 2 years of life," *Human Brain Mapping*, 2019, doi:10.1002/hbm.24439.

[12] C. Lebel and S. Deoni, "The development of brain white matter microstructure," *NeuroImage*, vol. 182, pp. 207–218, 2018, doi:10.1016/j.neuroimage.2017.12.097.

[13] T. K. Turesky, E. S. Escalante, M. Loh, and N. Gaab, "Longitudinal trajectories of brain development from infancy to school age and their relationship with literacy development," *Proceedings of the National Academy of Sciences*, vol. 122, no. 24, p. e2414598122, Jun. 2025, doi:10.1073/pnas.2414598122.

[14] H. Zhang, J. Li, X. Su, Y. Hu, T. Liu, S. Ni, H. Li, X.-N. Zuo, J. Fu, T.-F. Yuan, and Z. Yang, "Growth charts of brain morphometry for preschool children," *NeuroImage*, vol. 255, p. 119178, Jul. 2022, doi:10.1016/j.neuroimage.2022.119178.

[15] L. Chen, Y. Wang, Z. Wu, Y. Shan, T. Li, S.-C. Hung, L. Xing, H. Zhu, L. Wang, W. Lin, and G. Li, "Four-dimensional mapping of dynamic longitudinal brain subcortical development and early learning functions in infants," *Nature Communications*, vol. 14, no. 1, p. 3727, Jun. 2023, doi:10.1038/s41467-023-38974-9.

[16] D. L. Zimmerman, V. Núñez-Antón, T. G. Gregoire, O. Schabenberger, J. D. Hart, M. G. Kenward, G. Molenberghs, G. Verbeke, M. Pourahmadi, P. Vieu, D. L. Zimmerman, V. Núñez-Antón, T. G. Gregoire, O. Schabenberger, J. D. Hart, M. G. Kenward, G. Molenberghs, G. Verbeke, M. Pourahmadi, and P. Vieu, "Parametric modelling of growth curve data: An overview," *Test*, vol. 10, no. 1, pp. 1–73, 2001, doi:10.1007/BF02595823.

[17] R. González-Val, "Parametric, semiparametric and nonparametric models of urban growth," *Cities*, vol. 132, p. 104079, 2023, doi:10.1016/j.cities.2022.104079.

[18] G. Y. Lee, Y.-A. Youn, Y. H. Jang, H. Kim, J. Y. Lee, Y. J. Lee, M. Jung, and H. J. Lee, "Structural development and brain asymmetry in the fronto-limbic regions in preschool-aged children," *Frontiers in Pediatrics*, vol. 12, 2024, doi:10.3389/fped.2024.1362409.

[19] M. A. Brown and R. C. Semelka, *MRI: Basic Principles and Applications*, 3rd ed. Hoboken, NJ: Wiley-Liss, 2003.

[20] T. Guallart-Naval and J. Alonso, "Electromagnetic noise characterization and suppression in low-field MRI systems," arXiv preprint arXiv:2507.20413, Jul. 2025.

[21] D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince, *MRI from Picture to Proton*, 2nd ed. Cambridge, UK; New York, USA: Cambridge University Press, 2006.

[22] K. M. Broadhouse, "The physics of mri and how we use it to reveal the mysteries of the mind," *Frontiers for Young Minds*, vol. 7, p. 23, 2019, doi:10.3389/frym.2019.00023.

[23] W. S. Tae, B. J. Ham, S. B. Pyun, and B. J. Kim, "Current clinical applications of structural MRI in neurological disorders," *Journal of Clinical Neurology*, vol. 21, no. 4, pp. 277–293, Jul. 2025, doi:10.3988/jcn.2025.0185.

[24] T. Paus, D. L. Collins, A. C. Evans, G. Leonard, B. Pike, and A. Zijdenbos, "Maturation of white matter in the human brain: a review of magnetic resonance studies," *Brain Research Bulletin*, vol. 54, no. 3, pp. 255–266, Feb. 2001, doi:10.1016/S0361-9230(00)00434-2.

[25] S. Eminian, S. D. Hajdu, R. A. Meuli, P. Maeder, and P. Hagmann, "Rapid high resolution t1 mapping as a marker of brain development: Normative ranges in key regions of interest," *PLoS ONE*, vol. 13, no. 6, p. e0198250, Jun. 2018, doi:10.1371/journal.pone.0198250.

[26] M. Boudreau, K. E. Keenan, and N. Stikov, "Quantitative T1 MRI," *NeuroLibre Reproducible Preprints*, vol. 19, 2023, doi:10.55458/neurolibre.00019.

[27] D. Gräfe, J. Frahm, A. Merkenschlager, D. Voit, and F. W. Hirsch, "Quantitative t1 mapping of the normal brain from early infancy to adulthood," *Pediatric Radiology*, vol. 51, no. 3, pp. 450–456, Mar. 2021, doi:10.1007/s00247-020-04842-7.

[28] A. Makropoulos, S. J. Counsell, and D. Rueckert, "A review on automatic fetal and neonatal brain mri segmentation," *NeuroImage*, vol. 170, pp. 231–248, 2018, doi:10.1016/j.neuroimage.2017.06.074.

[29] I. Despotović, B. Goossens, and W. Philips, "Mri segmentation of the human brain: Challenges, methods, and applications," *Computational and Mathematical Methods in Medicine*, vol. 2015, p. 450341, 2015, doi:10.1155/2015/450341.

[30] G. Li, L. Wang, P.-T. Yap, F. Wang, Z. Wu, Y. Meng, P. Dong, J. Kim, F. Shi, I. Rekik, W. Lin, and D. Shen, "Computational neuroanatomy of baby brains: A review," *NeuroImage*, vol. 185, pp. 906–925, 2019, doi:10.1016/j.neuroimage.2018.03.042.

[31] T. V. Phan, D. Smeets, J. B. Talcott, and M. Vandermosten, "Processing of structural neuroimaging data in young children: Bridging the gap between current practice and state-of-the-art methods," *Developmental Cognitive Neuroscience*, vol. 33, pp. 206–223, 2018, doi:10.1016/j.dcn.2017.08.009.

[32] R. Cooper, R. A. Hayes, M. Corcoran, K. N. Sheth, T. C. Arnold, J. M. Stein, D. C. Glahn, and M. Jalbrzikowski, "Bridging the gap: improving correspondence between low-field and high-field magnetic resonance images in young people," *Frontiers in Neurology*, vol. 15, 2024, doi:10.3389/fneur.2024.1339223.

[33] J. E. Iglesias, R. Schleicher, S. Laguna, B. Billot, P. Schaefer, B. McKaig, J. N. Goldstein, K. N. Sheth, M. S. Rosen, and W. T. Kimberly, "Quantitative brain morphometry of portable low-field-strength MRI using super-resolution machine learning," *Radiology*, vol. 306, no. 3, p. e220522, Mar. 2023, doi:10.1148/radiol.220522.

[34] B. B. Avants, N. J. Tustison, J. Wu, P. A. Cook, and J. C. Gee, "An open source multivariate framework for n-tissue segmentation with evaluation on public data," *Neuroinformatics*, vol. 9, no. 4, pp. 381–400, 2011, doi:10.1007/s12021-011-9109-y.

[35] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001, doi:10.1109/42.906424.

[36] L. Zöllei, J. E. Iglesias, Y. Ou, P. E. Grant, and B. Fischl, "Infant freesurfer: An automated segmentation and surface extraction pipeline for t1-weighted neuroimaging data of infants 0–2 years," *NeuroImage*, vol. 218, p. 116946, 2020, doi:10.1016/j.neuroimage.2020.116946.

[37] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. Bach Cuadra, "A review of atlas-based segmentation for magnetic resonance brain images," *Computer Methods and Programs in Biomedicine*, vol. 104, no. 3, pp. e158–e177, 2011, doi:10.1016/j.cmpb.2011.07.015.

[38] J. Wang, C. Vachet, A. Rumple, S. Gouttard, C. Ouziel, E. Perrot, G. Du, X. Huang, G. Gerig, and M. Styner, "Multi-atlas segmentation of subcortical brain structures via the autoseg software pipeline," *Frontiers in Neuroinformatics*, vol. 8, p. 7, 2014, doi:10.3389/fninf.2014.00007.

[39] A. Makropoulos, E. C. Robinson, A. Schuh, R. Wright, S. Fitzgibbon, J. Bozek, S. J. Counsell, J. Steinweg, K. Vecchiato, J. Passerat-Palmbach, G. Lenz, F. Mortari, T. Tenev, E. P. Duff, M. Bastiani, L. Cordero-Grande, E. Hughes, N. Tusor, J.-D. Tournier, J. Hutter, A. N. Price, R. P. A. G. Teixeira, M. Murgasova, S. Victor, C. Kelly, M. A. Rutherford, S. M. Smith, A. D. Edwards, J. V. Hajnal, M. Jenkinson, and D. Rueckert, "The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction," *NeuroImage*, vol. 173, pp. 88–112, 2018, doi:10.1016/j.neuroimage.2018.01.054.

[40] A. Makropoulos, I. S. Gousias, C. Ledig, P. Aljabar, A. Serag, J. V. Hajnal, A. D. Edwards, S. J. Counsell, and D. Rueckert, "Automatic whole brain mri segmentation of the developing neonatal brain," *IEEE Transactions on Medical Imaging*, vol. 33, no. 9, pp. 1818–1831, 2014, doi:10.1109/TMI.2014.2322280.

[41] T. V. Phan, D. M. Sima, C. Beelen, J. Vanderauwera, D. Smeets, and M. Vandermosten, "Evaluation of methods for volumetric analysis of pediatric brain data: The childmetrix pipeline versus adult-based approaches," *NeuroImage: Clinical*, vol. 19, pp. 734–744, 2018, doi:10.1016/j.nicl.2018.05.030.

[42] R. J. Beare, J. Chen, C. E. Kelly, D. Alexopoulos, C. D. Smyser, C. E. Rogers, W. Y. Loh, L. G. Matthews, J. L. Y. Cheong, A. J. Spittle, P. J. Anderson, L. W. Doyle, T. E. Inder, M. L. Seal, and D. K. Thompson, "Neonatal brain tissue classification with morphological adaptation and unified segmentation," *Frontiers in Neuroinformatics*, vol. 10, p. 12, 2016, doi:10.3389/fninf.2016.00012.

[43] G. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017, doi:10.1016/j.media.2017.07.005.

[44] T. Islam, M. S. Hafiz, J. R. Jim, M. M. Kabir, and M. Mridha, "A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions," *Healthcare Analytics*, vol. 5, p. 100340, 2024, doi:10.1016/j.health.2024.100340.

[45] J. Simarro, M. I. Meyer, S. Van Eyndhoven, T. V. Phan, T. Billiet, D. M. Sima, and E. Ortibus, "A deep learning model for brain segmentation across pediatric and adult populations," *Scientific Reports*, vol. 14, no. 1, p. 11735, 2024, doi:10.1038/s41598-024-61798-6.

[46] L. Henschel, D. Kügler, L. Zöllei, and M. Reuter, "Vinna for neonates: Orientation independence through latent augmentations," *Imaging Neuroscience*, vol. 2, p. 00180, 2024, doi:10.1162/imag_a_00180.

[47] S. Kumar, S. Conjeti, A. Guha Roy, C. Wachinger, and N. Navab, "Infinet: Fully convolutional networks for infant brain mri segmentation," arXiv preprint arXiv:810.05735, 2018.

[48] L. Wang, Z. Wu, L. Chen, Y. Sun, W. Lin, and G. Li, "iBEAT v2.0: A multi-site applicable, deep learning-based pipeline for infant cerebral cortical surface reconstruction," *Nature protocols*, vol. 18, no. 5, pp. 1488–1509, May 2023, doi:10.1038/s41596-023-00806-x.

[49] T. J. Hendrickson, P. Reiners, L. A. Moore, A. J. Perrone, D. Alexopoulos, E. G. Lee, M. Styner, O. Kardan, T. A. Chamberlain, A. Mummaneni, H. A. Caldas, B. Bower, S. Stoyell, T. Martin, S. Sung, E. Fair, J. Uriarte-Lopez, A. R. Rueter, E. Yacoub, M. D. Rosenberg, C. D. Smyser, J. T. Elison, A. Graham, D. A. Fair, and E. Feczko, "BIBSNet: A deep learning baby image brain segmentation network for MRI scans," *bioRxiv*, May 2023, doi:10.1101/2023.03.22.533696.

[50] P. Laso, S. Cerri, A. Sorby-Adams, J. Guo, F. Mateen, P. Goebl, J. Wu, P. Liu, H. Li, S. I. Young, B. Billot, O. Puonti, G. Sze, S. Payabavash, A. DeHavenon, K. N. Sheth, M. S. Rosen, J. Kirsch, N. Strisciuglio, J. M. Wolterink, A. Eshaghi, F. Barkhof, W. T. Kimberly, and J. E. Iglesias, "Quantifying white matter hyperintensity and brain volumes in heterogeneous clinical and low-field portable MRI," arXiv preprint arXiv:2312.05119, Dec. 2023.

[51] K. T. Islam, S. Zhong, P. Zakavi, Z. Chen, H. Kavnoudias, S. Farquharson, G. Durbridge, M. Barth, K. L. McMahon, P. M. Parizel, A. Dwyer, G. F. Egan, M. Law, and Z. Chen, "Improving portable low-field MRI image quality through image-to-image translation using paired low- and high-field images," *Scientific Reports*, vol. 13, no. 1, p. 21183, Dec. 2023, doi:10.1038/s41598-023-48438-1.

[52] W. T. Kimberly, A. J. Sorby-Adams, A. G. Webb, E. X. Wu, R. Beekman, R. Bowry, S. J. Schiff, A. de Havenon, F. X. Shen, G. Sze, P. Schaefer, J. E. Iglesias, M. S.

Rosen, and K. N. Sheth, "Brain imaging with portable low-field mri," *Nature Reviews Bioengineering*, vol. 1, no. 9, 2023, doi:10.1038/s44222-023-00086-w.

[53] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, and M. Michalski, "Medical image synthesis for data augmentation and anonymization using generative adversarial networks," in *Simulation and Synthesis in Medical Imaging*, A. Gooya, O. Goksel, I. Oguz, and N. Burgos, Eds., vol. 11037, pp. 1–11. Springer, 2018, doi:10.1007/978-3-030-00536-8_1.

[54] K. A. Wahid, J. Xu, D. El-Habashy, Y. Khamis, M. Abobakr, B. McDonald, N. O'Connell, D. Thill, S. Ahmed, C. S. Sharafi, K. Preston, T. C. Salzillo, A. Mohamed, R. He, N. Cho, J. Christodouleas, C. D. Fuller, and M. A. Naser, "Deep-learning-based generation of synthetic high-resolution mri from low-resolution mri for use in head and neck cancer adaptive radiotherapy," *medRxiv*, 2022, doi:10.1101/2022.06.19.22276611.

[55] T. Fujioka, M. Mori, J. Oyama, K. Kubota, E. Yamaga, Y. Yashima, L. Katsuta, K. Nomura, M. Nara, G. Oda, T. Nakagawa, and U. Tateishi, "Investigating the image quality and utility of synthetic mri in the breast," *Magnetic Resonance in Medical Sciences*, vol. 20, no. 4, pp. 431–438, 2021, doi:10.2463/mrms.mp.2020-0132.

[56] B. Billot, D. N. Greve, O. Puonti, A. Thielscher, K. Van Leemput, B. Fischl, A. V. Dalca, and J. E. Iglesias, "SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining," *Medical Image Analysis*, vol. 86, p. 102789, May 2023, doi:10.1016/j.media.2023.102789.

[57] J. M. Molina-Maza *et al.*, "Development of a super-resolution scheme for pediatric brain mri," *Frontiers in Neuroscience*, vol. 16, p. 9641213, 2022, doi:10.3389/fnins.2022.9641213.

[58] L. Qu, Y. Zhang, S. Wang, P.-T. Yap, and D. Shen, "Synthesized 7t MRI from 3t MRI via deep learning in spatial and wavelet domains," *Medical image analysis*, vol. 62, p. 101663, May 2020, doi:10.1016/j.media.2020.101663.

[59] M. M. Saad, R. O'Reilly, and M. H. Rehmani, "A survey on training challenges in generative adversarial networks for biomedical image analysis," *Artificial Intelligence Review*, vol. 57, p. 19, 2024, doi:10.1007/s10462-023-10624-y.

[60] A. Bandi, "The power of generative ai: A review of requirements, capabilities, and challenges," *Future Internet*, vol. 15, no. 8, p. 260, 2023, doi:10.3390/fi15080260.

[61] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan, "Deep learning approaches for data augmentation in medical imaging: A review," *Image Processing*, vol. 9, no. 4, p. 81, 2023, doi:10.3390/ipj9040081.

[62] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," arXiv preprint arXiv:1505.04597, 2015.

[63] M. Payne, I. Mali, T. Mueller, M. Cain, R. Segev, and S. H. Bossmann, "Super-resolution reconstruction in ultrahigh-field mri," *Biophysical Reports*, vol. 3, no. 2, p. 100107, Jun. 2023, doi:10.1016/j.bpr.2023.100107.

[64] A. Kalluvila, D. Bhutto, N. Koonjoo, and M. Rockenbach, "Synthetic low-field mri super-resolution via nested u-net architecture," arXiv preprint arXiv.2211.15047, 2022.

[65] J. E. Iglesias, B. Billot, Y. Balbastre, A. Tabari, J. Conklin, R. Gilberto González, D. C. Alexander, P. Golland, B. L. Edlow, and B. Fischl, "Joint super-resolution and synthesis of 1 mm isotropic MP-RAGE volumes from clinical MRI exams with scans of different orientation, resolution and contrast," *NeuroImage*, vol. 237, p. 118206, Aug. 2021, doi:10.1016/j.neuroimage.2021.118206.

[66] J. Wang, J. Wang, S. Wang, and Y. Zhang, "Deep learning in pediatric neuroimaging," *Displays*, vol. 80, p. 102583, Dec. 2023, doi:10.1016/j.displa.2023.102583.

[67] M. K. Sherwani and S. Gopalakrishnan, "A systematic literature review: deep learning techniques for synthetic medical image generation and their applications in radiotherapy," *Frontiers in Radiology*, vol. Volume 4, 2024, doi:10.3389/fradi.2024.1385742.

[68] F. Renard, S. Guedria, N. D. Palma, A. Mitra, and O. Steichen, "Variability and reproducibility in deep learning for medical image segmentation," *Scientific Reports*, vol. 10, p. 13724, 2020, doi:10.1038/s41598-020-69920-0.

[69] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, p. 210, 2022, doi:10.1186/s13104-022-06096-y.

[70] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, p. 29, 2015, doi:10.1186/s12880-015-0068-x.

[71] S. Cerri, D. N. Greve, A. Hoopes, H. Lundell, H. R. Siebner, M. Mühlau, and K. Van Leemput, "An open-source tool for longitudinal whole-brain and white

matter lesion segmentation," *NeuroImage: Clinical*, vol. 38, p. 103354, 2023, doi:10.1016/j.nicl.2023.103354.

[72] B. Li, W. J. Niessen, S. Klein, M. de Groot, M. A. Ikram, M. W. Vernooij, and E. E. Bron, "Longitudinal diffusion MRI analysis using segis-net: A single-step deep-learning framework for simultaneous segmentation and registration," *NeuroImage*, vol. 235, p. 118004, 2021, doi:10.1016/j.neuroimage.2021.118004.

[73] B. Li, W. J. Niessen, S. Klein, M. de Groot, M. A. Ikram, M. W. Vernooij, and E. E. Bron, "A hybrid deep learning framework for integrated segmentation and registration: Evaluation on longitudinal white matter tract changes," in *Medical Image Computing and Computer Assisted Intervention*. Springer International Publishing, 2019, doi:10.1007/978-3-030-32248-9_72.

[74] M. Long, C. Ostertag, J. E. Reynolds, J. Zheng, B. Landman, Y. Huo, N. D. Forkert, and C. Lebel, "Few sex differences in regional gray matter volume growth trajectories across early childhood," *Imaging Neuroscience*, vol. 2, pp. 1–26, 2024, doi:10.1162/imag_a_00154.

[75] B. R. Howell, M. A. Styner, W. Gao, P.-T. Yap, L. Wang, K. Baluyot, E. Yacoub, G. Chen, T. Potts, A. Salzwedel, G. Li, J. H. Gilmore, J. Piven, J. K. Smith, D. Shen, K. Ugurbil, H. Zhu, W. Lin, and J. T. Elison, "The UNC/UMN baby connectome project (BCP): An overview of the study design and protocol development," *NeuroImage*, vol. 185, pp. 891–905, Jan. 2019, doi:10.1016/j.neuroimage.2018.03.049.

[76] F. Kühne, W.-J. Neumann, P. Hofmann, J. Marques, A. M. Kaindl, and A. Tietze, "Assessment of myelination in infants and young children by t1 relaxation time measurements using the magnetization-prepared 2 rapid acquisition gradient echoes sequence," *Pediatric Radiology*, vol. 51, no. 11, pp. 2058–2068, 2021, doi:10.1007/s00247-021-05109-5.

[77] D. Romascano, G. F. Piredda, S. Caneschi, T. Hilbert, R. Corredor, B. Maréchal, T. Kober, J.-B. Ledoux, E. Fornari, P. Hagmann, and S. Denervaud, "Normative volumes and relaxation times at 3t during brain development," *Scientific Data*, vol. 11, p. 429, Apr. 2024, doi:10.1038/s41597-024-03267-3.

[78] S. Cerri, D. N. Greve, A. Hoopes, H. Lundell, H. R. Siebner, M. Mühlau, and K. Van Leemput, "An open-source tool for longitudinal whole-brain and white matter lesion segmentation," *NeuroImage: Clinical*, vol. 38, p. 103354, Jan. 2023, doi:10.1016/j.nicl.2023.103354.

[79] P. Virtanen, R. Gommers, T. E. Oliphant, and et al., "Scipy 1.0: Fundamental algorithms for scientific computing in python," *Nature Methods*, vol. 17, pp. 261–272, 2020, doi:10.1038/s41592-019-0686-2.

[80] P. A. Bottomley, T. H. Foster, R. E. Argersinger, and L. M. Pfeifer, "A review of normal tissue hydrogen nmr relaxation times and relaxation mechanisms from 1–100 mhz: Dependence on tissue type, nmr frequency, temperature, species, excision, and age," *Medical Physics*, vol. 11, no. 4, pp. 425–448, 1984, doi:10.1118/1.595535.

[81] W. D. Rooney, G. Johnson, X. Li, E. R. Cohen, S.-G. Kim, K. Ugurbil, and C. S. J. Springer, "Magnetic field and tissue dependencies of human brain longitudinal $1h_2o$ relaxation in vivo," *Magnetic Resonance in Medicine*, vol. 57, no. 2, pp. 308–318, 2007, doi:10.1002/mrm.21122.

[82] F. Padormo, P. Cawley, L. Dillon, E. Hughes, J. Almalbis, J. Robinson, A. Maggioni, M. De la Fuente Botella, D. Cromb, A. Price, L. Arlinghaus, J. Pitts, T. Luo, D. Zhang, S. C. L. Deoni, S. Williams, S. Malik, J. O'Muircheartaigh, S. J. Counsell, M. Rutherford, T. Arichi, A. D. Edwards, and J. V. Hajnal, "In vivo t_1 mapping of neonatal brain tissue at 64 mt," *Magnetic Resonance in Medicine*, vol. 89, no. 3, pp. 1016–1025, 2023, doi:10.1002/mrm.29509.

APPENDIX A ACQUISITION PARAMETERS

Table A.1 Acquisition parameters for the Hyperfine Unity neonatal T1w sequence.

Parameter	**Value**
Sequence	Hyperfine Unity Neonatal T1w
Resolution (mm)	2×2×2
FOV (mm)	220×180×180
Matrix size	110×90×90
Bandwidth (Hz)	32 000 / 64 000
Flip RF	180/90/180/30/90/90/90
TE (s)	0.00764
TI (s)	0.5
TR (s)	1.25
Sampling Path	Radial
Scan Plan	XYZ
Total duration (s)	893.9