



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

---

Year : 2023

## Deep learning methods for diffusion MRI in early development of the human brain: resolution enhancement and model estimation

Kebiri Hamza

Kebiri Hamza, 2023, Deep learning methods for diffusion MRI in early development of the human brain: resolution enhancement and model estimation

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB\_B248E402983E1

### **Droits d'auteur**

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

### **Copyright**

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



**UNIL** | Université de Lausanne

Faculté de biologie  
et de médecine

Département de Radiologie,

Centre Hospitalier Universitaire Vaudois (CHUV)

**Deep learning methods for diffusion MRI in early  
development of the human brain: resolution  
enhancement and model estimation**

**Thèse de doctorat ès sciences de la vie (PhD)**

présentée à la

Faculté de biologie et de médecine  
de l'Université de Lausanne

par

**Hamza KEBIRI**

Master de l'École Polytechnique Fédérale de Lausanne  
Master de l'Universität Osnabrück

**Jury**

Prof. Dr. Med. E. Mahmut ÖZSAHIN, Président  
Dre. Meritxell Bach Cuadra, Directrice de thèse  
Prof. Dr. Med. Roland Wiest, Expert  
Prof. Alessandro Daducci, Expert  
Prof. Daniel Rueckert, Expert

Lausanne  
(2023)



# Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

<b>Président·e</b>	Monsieur	Prof.	Esat Mahmut	<b>Özsahin</b>
<b>Directeur·trice de thèse</b>	Madame	Dre	Meritxell	<b>Bach Cuadra</b>
<b>Expert·e·s</b>	Monsieur	Prof.	Roland	<b>Wiest</b>
	Monsieur	Prof.	Alessandro	<b>Daducci</b>
	Monsieur	Prof.	Daniel	<b>Rueckert</b>

le Conseil de Faculté autorise l'impression de la thèse de

**Hamza Mohamed Kebiri**

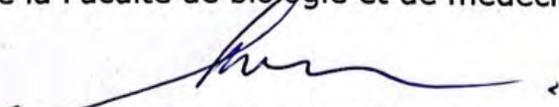
Master in Cognitive science, Universität Osnabrück, Allemagne

intitulée

**Deep learning methods for diffusion MRI in early  
development of the human brain: resolution  
enhancement and model estimation**

Lausanne, le 20 juillet 2023

pour le Doyen  
de la Faculté de biologie et de médecine



Prof. Esat Mahmut Özsahin



## ABSTRACT

Diffusion magnetic resonance imaging has emerged as the gold standard tool for studying the brain white matter both *in vivo* and non-invasively, offering valuable insights into underlying tissue microstructure and brain connectivity. However, applying this technique to investigate the human developing brain, such as in fetuses and newborns, poses unique challenges. In this sensitive population, the scanning time is limited for unpredictable motion risk minimization and for ethical reasons. Additionally, images have a low signal-to-noise ratio and a low spatial resolution. Moreover, the developing brain undergoes rapidly changing microstructural properties during the last months of pregnancy and early months of birth. The application of current diffusion magnetic resonance imaging methods to developing brains is severely constrained by all these aspects, necessitating the development of tailored approaches. This thesis tackles this specific problem by proposing two *deep learning* based methods that leverage high quality research datasets to improve constrained clinical acquisitions. First, we have developed a method to enhance the through-plane resolution using a deep *autoencoder*. We show its performance over conventional image interpolation methods of the raw signal and in estimated *diffusion tensor* scalar maps. Second, we designed a model to predict accurate orientation distribution functions from a low number of diffusion measurements that are typically available in clinical settings. We extensively demonstrate its performance on newborn subjects compared to state-of-the-art methods (such as constrained spherical deconvolution) that need significantly more diffusion directions. We additionally show the out-of-domain generalizability of the method on clinical cohorts of newborns and fetuses. Finally, aiming at deriving optimal schemes for fetal sequences, we have conducted a quantitative validation study on a *phantom* with crossing-fibers, to quantify the time trade-off that is imposed by the clinical constraints, between the number of gradient directions and the number of acquired volumes. Overall, we believe that the aforementioned methods that harness the capabilities of deep neural networks to extract transferable knowledge from large datasets, possess the potential to offer significant insights into the complex mechanisms underlying the early development of the human brain.



# Contents

1	INTRODUCTION	I
1.1	Aims and organization of this PhD thesis . . . . .	4
1.2	The developing brain . . . . .	5
1.3	Diffusion MRI . . . . .	11
1.4	The <i>in vivo</i> developing brain with diffusion MRI . . . . .	18
1.5	Clinical MRI applications in developing brains . . . . .	22
1.6	Machine learning and diffusion MRI . . . . .	23
1.7	Contributions of this PhD thesis . . . . .	31
2	SUMMARY OF RESULTS	34
2.1	Through-plane super-resolution with autoencoders in signal domain . . . . .	35
2.2	Through-plane super-resolution with autoencoders in spherical harmonics domain . . . . .	35
2.3	Fiber orientation distribution function estimation with deep learning . . . . .	37
2.4	A crossing-fibers phantom study . . . . .	38
3	DISCUSSION	40
3.1	Main contributions . . . . .	41
3.2	Limitations and future steps . . . . .	45
3.3	Conclusion . . . . .	48
	REFERENCES	50
4	MAIN PAPERS	83
5	APPENDIX	161
5.1	Angular super-resolution with deep neural networks . . . . .	162
5.2	Automated fetal brain segmentation of 2D magnetic resonance images: transfer learning and 3D topology correction . . . . .	168
5.3	Direct segmentation of brain white matter tracts in diffusion MRI . . . . .	171

5.4	<b>Spatio-temporal motion correction and iterative reconstruction of in-utero fetal fMRI . . . . .</b>	<b>171</b>
-----	--	------------



# Listing of figures

1.1	The different scales across which the brain is studied. The figure is adapted from <i>small-pocketlibrary.com</i> . . . . .	2
1.2	Typical field of views of macro, meso and micro scales in analysing the brain. (Adapted figure <sup>19,198</sup> ). . . . .	3
1.3	(a) Illustration of the different components of a neuron (b) White matter fiber bundles ensheathed by myelin. Combined figure from <i>slidemodel.com</i> and Avram et al. <sup>13</sup> . . . . .	6
1.4	The rates of volumetric change for each tissue volume over the lifespan, estimated using the first derivatives of the median volumetric trajectories (adapted figure <sup>21</sup> ). . . . .	7
1.5	Cortical surface reconstructions of pre-term newborns at varying post-menstrual ages (left value) and corresponding whole-brain sulcation index (right value). The cortical surfaces are color-coded to represent variations in surface curvature. Adapted figure <sup>42</sup> . . . . .	8
1.6	(a) Isotropic and (b) anisotropic illustration of Gaussian diffusion. Figure from Hagmann et al. <sup>80</sup> . . . . .	12
1.7	Diffusion tensor fractional anisotropy (FA) maps (top row) and color FA maps (bottom row) across gestational weeks (adapted figure <sup>117</sup> ). . . . .	15
1.8	Crossing fibers region in a biological phantom reconstructed by the DTI model, diffusion ODF and fiber ODF models. While the tensor fails at reconstructing the crossing area, ODF models can achieve this at different levels. Figure from Descoteaux et al. <sup>50</sup> . . . . .	17
1.9	Diffusion MRI fetal brain protocols across different centers. We can see the variability of acquisition schemes, the small b-values and the low number of directions (except for the research oriented dataset of the developing human connectome project, dHCP <sup>37</sup> ). . . . .	20
1.10	The ventral stream of the visual system where features are hierarchically reconstructed in the brain from primary visual areas to high-level visual areas such as the fusiform face area (FFA) in the Inferior Temporal cortex (IT). Eye processing is also displayed in the bottom figure. Convolutional neural networks (CNNs) bear several similarities to information processing of biological tissues. The figure is adapted from <i>neur-diness.wordpress.com</i> . . . . .	26

2.1	Qualitative comparison of colored FA in the one slice removed configuration for the best baseline interpolation method, i.e. Linear-1 (left), autoencoder enhancement AE-1 (middle) and GT (right). The red frame area highlight to a region where the linear interpolation shows a more accurate result <sup>110</sup> . . . . .	36
2.2	Colored FA and FA (top row) illustration of autoencoder enhancement between two original adjacent fetal slices in a still subject (35 GW). The bottom row shows a similar illustration of MD and FA for a moving subject (24 GW) <sup>110</sup> . Image generated with the autoencoder of section 2.1. . . . .	37
2.3	The deep learning method compared to CSD in different brain regions for 2 newborn subjects (left) and 2 fetal subjects (right) of 25 (top) and 29.4 (bottom) weeks of gestation. FODs are superimposed to the first SH coefficient of the method used. The DL first SH coefficient is shown in full-size too <sup>113</sup> . . . . .	38
5.1	Validation loss of the trained models with different patch sizes, i.e. $10^2$ , $32^2$ and $64^2$ voxels. Image generated with Tensorboard <sup>1</sup> . . . . .	164
5.2	Root mean squared error (RMSE) of four out of the six folds, for the 2D- and 3D-RCNN trained on different patch sizes. Spherical harmonics interpolation baseline is also shown on the first raw. Best and worst two scores are shown in blue and red, respectively . . . . .	165
5.3	On the left, 100 random points extracted from brain voxels where red refers to ground truth values and blue to network prediction values by the 2D-RCNN trained with $32 \times 32$ patches. On the right we can see the corresponding slices of the GT and network predictions. . . . .	166
5.4	Root mean squared error (RMSE) of validation and training subjects of spherical harmonics interpolation and RCNN trained on fetal data. . . . .	166
5.5	Validation loss on multiple models using Transformers with different parameters and by combining it with the LSTM module. Image generated with Tensorboard <sup>1</sup> . . . . .	167



”TRANSIRE SUUM PECTUS MUNDOQUE POTIRI” MANILIUS.



# Acknowledgments

First and foremost, I would like to extend my heartfelt thanks to my parents, my mother Malika Aboutalib, and my father Ali Kebiri for their unconditional love, unwavering support, and for giving me the freedom to follow my interests in this world. Your encouragement and belief in me have been a constant source of strength.

I want to extend my deepest gratitude to my academic advisor, Dr. Meritxell Bach Cuadra, for her invaluable support, kindness and guidance throughout my research journey. Her mentorship has been instrumental in the successful completion of this thesis.

I am also deeply thankful to the Swiss National Science Foundation for the financial support, which allowed me to pursue my research and acquire an international experience through a Mobility grant. In this regard, my sincere thanks go to Prof. Ali Gholipour and Prof. Davood Karimi, during my stay in Boston Children's Hospital and Harvard Medical School. Their knowledge and values have been significant in expanding my horizons and nurturing my passion for this field.

I also want to thank Dr. Erick Canales-Rodriguez, Dr. Yasser Aleman Gomez and Dr. Gabriel Girard for their extended assistance and help in the realm of diffusion MRI. To my uncle Dr. med. Hakim El Hamzaoui and my mentors, Prof. El Mahdi El Mhamdi and Prof. Rachid Guerraoui, before and during the PhD, thank you for accompanying me on this journey and helping me develop and discover my love for science. I also want to thank Dr. Alessandro Motta for his generous supervision during my stay in Max Planck Institute for Brain Research in Frankfurt, as well as Prof. Moritz Helmstaedter, Dr. Sahil Loomba and Dr. Anjali Gour. My genuine thanks also go to Dr. Farnaz Delavari for her valuable help on this thesis and for reviewing this manuscript.

I am grateful to the jury members, Prof. dr. med. Roland Wiest, Prof. Alessandro Daducci and Prof. Daniel Rueckert, for their valuable time, consideration, and insightful inputs during the evaluation of my thesis. I also want to thank Rizhong Lin for his dedication during his internship. I am indebted to Kinderspital Zurich and Boston Children's Hospital for providing access to their valuable data, which played a significant role in the execution of my research. Additionally, I extend my thanks to the developing Human Connectome Project (dHCP) consortium for providing open-source data, which has been instrumental in training the *deep learning* models.

In conclusion, the completion of this thesis would not have been possible without the unwavering support, encouragement, and guidance from all the individuals and organizations mentioned above and all my Medical Image Analysis Laboratory (MIAL) colleagues. Their contributions have left an indelible mark on my academic journey, and for that, I am sincerely grateful.



*even when we have explained the performance of all the cognitive and behavioral functions in the vicinity of experience—perceptual discrimination, categorization, internal access, verbal report—there may still remain a further unanswered question: Why is the performance of these functions accompanied by experience?*

David Chalmers

# 1

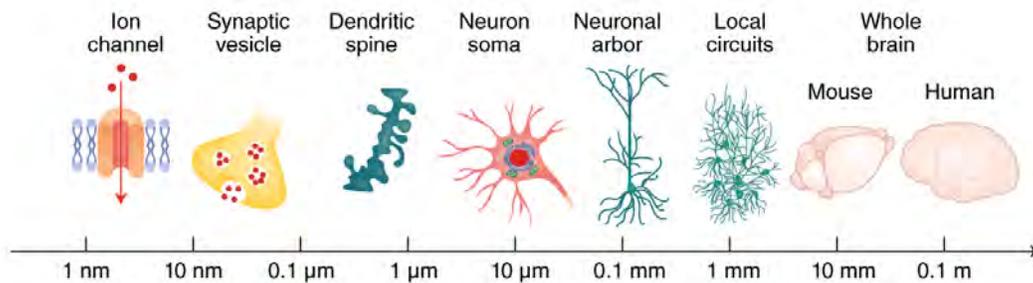
## Introduction

Perhaps the main mystery for which science is substantially missing formalism tools in the 21<sup>st</sup> century is the notion of *consciousness*. It can be simply defined as the continuous stream of subjective experience that switch off when we dreamless sleep and comes back the next morning<sup>211</sup>. In the past, consciousness was an exclusive topic of philosophy but with a growing understanding of the human brain, neural correlates of consciousness are more and more studied<sup>156,182,167,172,212,119</sup>. Yet, the brain-mind gap is still to be filled<sup>30,210,58,102</sup>, to hope to address the so called *hard problem of*

*consciousness*<sup>30</sup> that is referred to in the quotation above. This would shed light on whether consciousness is an emerging property of complex systems such as the nervous system, or a fundamental property of the universe.

A highly likely necessary condition of understanding human consciousness is the computational understanding of the central nervous system and specifically the brain, for which different configurations give rise to different mental states.

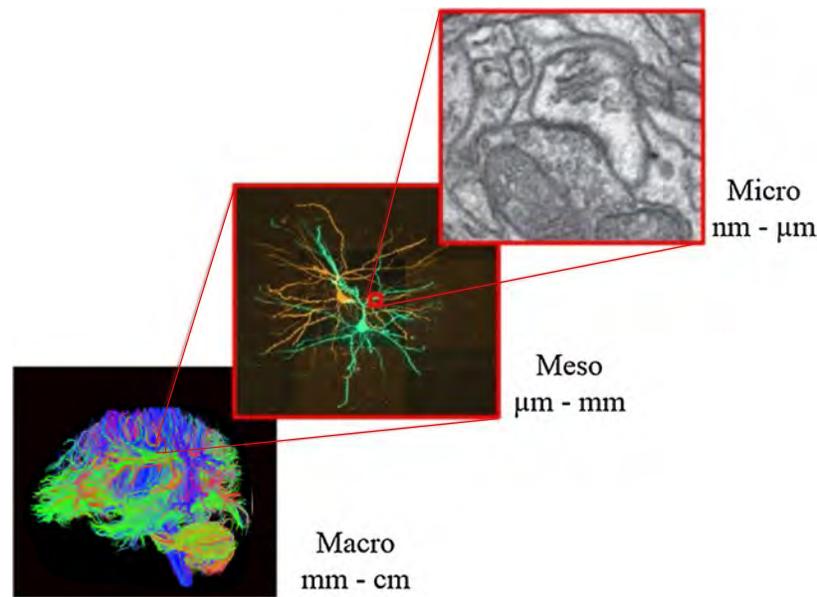
The brain is an organ of fascination that deserves to be studied for several direct practical reasons too. While their origin remains vastly unknown, neuro-psychiatric conditions such as depression, anxiety, bipolar disorder, psychosis or Alzheimer's disease have been shown to be associated with different brain structural and functional states from early development to adulthood, and at different scales<sup>146,234,202,148,200,136,27,45,75,44,24</sup>. In fact, the brain can be analysed at different levels, from the nanometric molecular scale, passing by the micrometric neuronal scale to the millimeter and centimeter scale of cortical circuits, brain lobes and tracts (Figure 1.1 and 1.2).



**Figure 1.1:** The different scales across which the brain is studied. The figure is adapted from *smallpocketlibrary.com*.

Functionally, at the micro-scale, electrophysiology using for example patch-clamp<sup>81</sup>, allows the analysis of single neuronal firings. Structurally, Electron Microscopy<sup>47</sup> can offer images of neuronal circuits that can be reconstructed at individual neurons and single synapse level<sup>158,194</sup>. Although this scale is highly accurate, it suffers from a low field of view which may hamper discoveries of patterns occurring at a broader level. Additionally, data analysis is lagging behind data acquisition be-

cause of the challenging task of analyzing petabytes of data<sup>159</sup>, despite the use of *machine learning* for automation. In fact, the largest reconstructed human *connectome*<sup>194</sup> (i.e. connectivity map) is around  $1mm^3$ . This is the typical unit, i.e. voxel size at the macro-scale level (Figure 1.2).



**Figure 1.2:** Typical field of views of macro, meso and micro scales in analysing the brain. (Adapted figure<sup>19,198</sup>).

At the macro-scale, which is the scale of interest in this work, a whole-brain field of view can be scanned, non-invasively, using magnetic resonance imaging (MRI). MRI is very unique in the sense that depending on which physical property we tune the machine to, the image can reveal substantially different tissue properties. Functionally, the blood oxygen level dependent (BOLD) signal, reflects changes in deoxyhemoglobin saturation, driven by variations in brain blood flow and blood oxygenation, which are correlated to neuronal activity. Hence, functional MRI (fMRI) can provide co-activation maps between distant brain regions using the BOLD signal as a proxy of macro-scale neuronal firing. Structurally, diffusion MRI (dMRI), by tracking brain water molecules, is the tool of reference to map connectivity between brain regions.

## 1.1 AIMS AND ORGANIZATION OF THIS PHD THESIS

These aforementioned MRI techniques have been extensively developed and applied to adult brains. However for developing brains, especially fetal brains, several challenges such as low signal-to-noise ratio (SNR), limited scanning time, and unpredictable motion hampers proper analyses of MR images. This is particularly the objective of this thesis: **developing advanced methods to better uncover developing brains from dMRI.**

In fact, a low SNR constrains the resolution of acquired images, that can hardly be used to extract valid anatomical information. We aimed at addressing this issue by employing *super-resolution* techniques with machine learning (a branch of Artificial Intelligence that we will detail further in Section 1.6). Additionally, restricted scanning time only allows the acquisition of a limited number of diffusion images, which provides inaccurate information about the diffusion of water molecules and hence the underlying brain structure. We aimed at **copied with this problem using machine learning by obtaining a mapping between reduced acquisitions and dMRI-derived metrics reconstructed with high quality research based acquisitions.** Lastly, since limited scanning time imposes a trade-off between several acquisition choices, we aimed at quantifying this trade-off between the number of dMR images and the number of repeated acquisitions. For that aim, we have scanned a *phantom*, i.e. a physical object that aims to simulate the brain tissue with different materials.

This manuscript is organized as follows: in section 1.2, we describe the most important studies relevant to the developing brain in terms of morphometry, myelination and white matter development. In section 1.3, we introduce the concept of MRI and dMRI. In section 1.4, we describe developing brain studies using dMRI, followed by a brief section 1.5 on clinical applications of fetal (diffusion) MRI. Next, we give an overview of machine learning and describe its applications in diffusion MRI (section 1.6). **Following the format of “thesis with articles” recommended by**

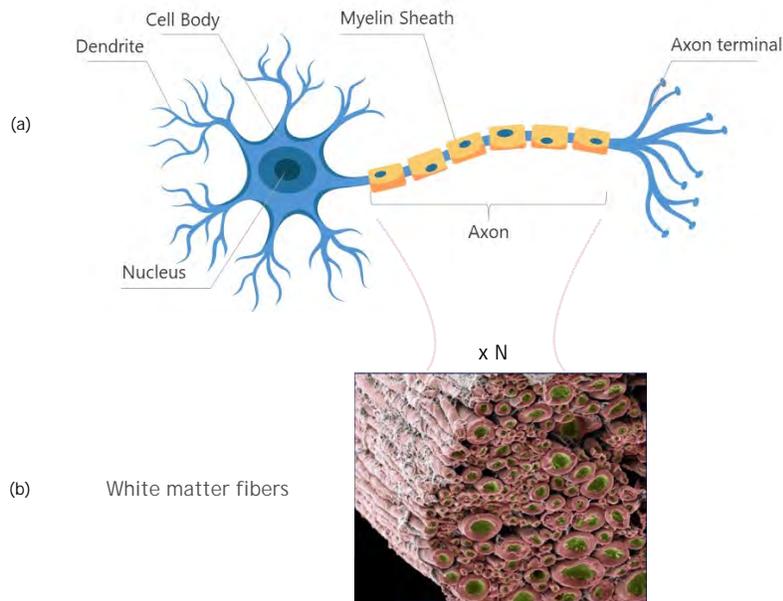
the University of Lausanne, we present in Chapter 2 a summary of the results for the 5 most important manuscripts written throughout the PhD (3 accepted<sup>115,110,111</sup>, 1 submitted<sup>114</sup> and 1 ready for submission<sup>113</sup>). In Chapter 3, we discuss and contextualize the results, then we provide an outlook on future directions, and finally we draw the conclusions of the thesis. The 5 main manuscripts are then attached. In the appendix chapter 5, we present a work that did not reach our goals but inspired our subsequent studies<sup>113,114</sup>, a second work on fetal brain segmentation on structural MRI, a third work<sup>112</sup> that uses similar methodologies<sup>113,114</sup> on children and adults rather than newborns and fetuses and a last work on the reconstruction of functional MRI images of fetal brains<sup>206</sup>.

## 1.2 THE DEVELOPING BRAIN

The brain is a highly interconnected network of nerve cells, known as neurons<sup>66</sup>. These neurons are composed of cell bodies that contain the nucleus, dendrites that are the electrochemical signal receiver component of the neuron, and axons that are the signal emitter component (Figure 1.3 - a).

One way of anatomically categorizing the brain at the macro-scale is by splitting it into gray matter, white matter and cerebrospinal fluid. Gray matter roughly corresponds to dendrites, cell bodies, unmyelinated axons and other structures such as glial cells, whereas white matter corresponds to myelinated axons that gather together in bundles (Figure 1.3 - b) and serve as connectors between neurons in different brain regions. The white matter is named for its light appearance due to the lipids contained in the myelin sheath that surrounds axons, which is needed to enhance the transmission speed of action potentials. Indeed, up to a 60 fold speed increase can be reached between for instance unmyelinated axons in the brain and myelinated axons in the peripheral nervous system<sup>54</sup>.

Both gray matter and white matter undergo substantial changes during development. In fact, the highest rate of volumetric growth across the lifespan occur in the fetal period, for both gray and white matter, as was shown in a meta-study comprising 123,984 MRI scans (Figure 1.4). During the

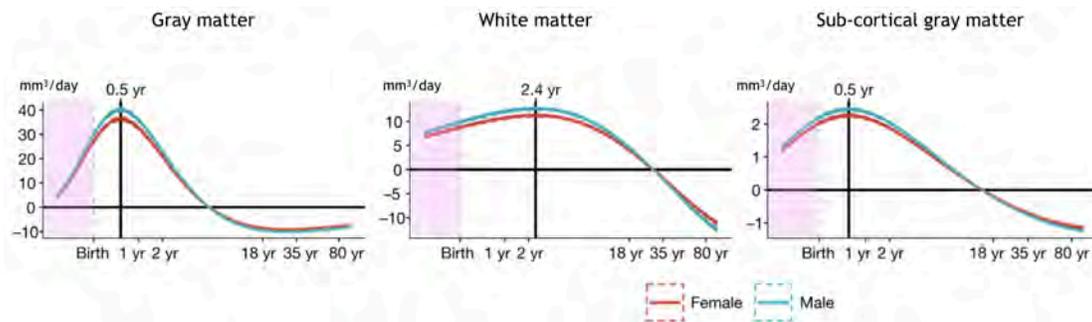


**Figure 1.3:** (a) Illustration of the different components of a neuron (b) White matter fiber bundles ensheathed by myelin. Combined figure from *slidemodel.com* and Avram et al. <sup>13</sup>.

late weeks of gestation and early post-natal months, there are notable cerebral alterations, which can be inferred from the non-linear increase of the cranial perimeter. This perimeter grows by approximately 14 cm within the first two post-natal years, followed by a subsequent increase of only 7 cm until adulthood <sup>54</sup>. Maturation and growth occur asynchronously; some regions, such as the sensory areas, experience early and rapid development, while other areas, such as the frontal associative regions, undergo slower development and continue to mature until the end of adolescence <sup>169</sup>.

### 1.2.1 MORPHOMETRY

Studying fetal brain development can provide valuable information to investigate *in vivo* brain maturation changes across time. Morphometric changes such as tissue volume or cortical folding can be studied from structural MRI. Significant changes in brain anatomy occur during the fetal period. For instance, the brain goes from a completely smooth organ at 22 weeks of gestation to the devel-



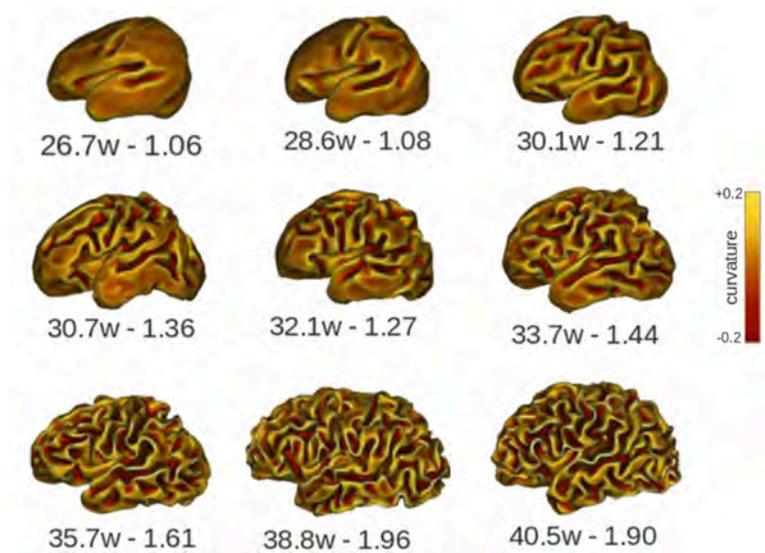
**Figure 1.4:** The rates of volumetric change for each tissue volume over the lifespan, estimated using the first derivatives of the median volumetric trajectories (adapted figure<sup>21</sup>).

opment of the major sulci and gyri at the 32<sup>nd</sup> gestational week. The identification of sulci in the developing brain can pose significant challenges due to their relatively smooth morphology during early developmental stages<sup>42</sup>.

The initial investigations of the formation of cerebral folds were conducted using postmortem fetal brains. In a pioneering study<sup>36</sup>, a total of 207 normal looking brains were analyzed and the chronological appearance of major fissures and sulci was documented. The results indicated that three successive folding phases are involved, with primary folds emerging at 20 GW, secondary folds at 32 GW, and tertiary folds at 38 GW (Figure 1.5).

By the time of full-term birth (~40 GW), all the primary and secondary sulci are developed, and the majority of tertiary sulci have also emerged. The resulting sulcal configuration in neonates is similar to that of adults in terms of shape and folding variability, with the exception of certain regions, such as the anterior cingulate, anterior temporal cortices, and mid-temporal sulcus. In these areas, neonates display greater variability in sulcal depth than adults<sup>85</sup>.

In the literature, prematurity is suggested to be a potential causative factor for changes in the dynamics of sulcal development due to differences between the intra-uterine and extra-uterine environments. Specifically, even in the absence of any obvious brain anomalies, the brains of pre-term infants exhibit more prominent folds, less cerebrospinal fluid and a more compact overall shape



**Figure 1.5:** Cortical surface reconstructions of pre-term newborns at varying post-menstrual ages (left value) and corresponding whole-brain sulcation index (right value). The cortical surfaces are color-coded to represent variations in surface curvature. Adapted figure<sup>42</sup>.

when compared to fetuses of the same age<sup>133</sup>. Moreover, the intensity and sharpness of folding, as well as the shape of sulci, were found to be altered in pre-term infants, and these effects were influenced by gestational age at birth.

### 1.2.2 MYELINATION

Prior to the advent of non-invasive brain imaging methods, human brain development was primarily investigated through rare post-mortem examinations. Myelin staining was used to determine the presence or absence of myelin in white matter region at a given age, but this information lacked specificity to individual bundles and could therefore be misleading at bundle crossings. Accurate measurements of myelin quantity remain elusive, hampering quantitative comparisons between white matter regions<sup>54</sup>.

Myelination is another highly important process that reaches its peak during the first post-natal

year. However, what is referred to as "pre-myelination" is when the glial cells responsible for myelination (oligodendrocytes) proliferate and migrate along the axons<sup>18</sup>, which happens much before birth. Subsequently, the process of myelination that induces significant changes in water molecule content, entails the ensheathment of oligodendroglial processes around the axons, along with the chemical maturation of the myelin sheath and an increase in macromolecular content<sup>175,54</sup>.

Post-mortem studies have provided detailed information about how myelination progression varies across cerebral regions, following caudo-rostral and central-peripheral directions. Microscopic observation of myelin has been reported as early as 20 gestational weeks (GW) in the bulb and pons. Mature myelin has been detected between 37 and 40 GW in the cerebellum and internal capsule. During the period between the first and third post-natal months, myelination occurs in the posterior limb of the internal capsule (PLIC), the optic radiations (OR), and the splenium of the corpus callosum. Mature myelin can be found in the anterior limb of the internal capsule and the genu of the CC at 6 months, in the occipital pole at 15 months, and in the frontal and temporal lobes at 23 months<sup>18</sup>.

### 1.2.3 POST-MORTEM STUDIES OF WHITE MATTER

The organization of white matter (WM) connections has been reported through post-mortem studies. After neurons migrate to their final position, they proceed to form connections with other neurons through the development of dendritic trees within the gray matter and the formation of axons that traverses the WM. The initial wave of migrating neurons reside in the subplate (located beneath the to-be cortex). This process of neural wiring occurs primarily during the second half of gestation and involves a cascade of events, such as neuronal and synaptic overproduction, cellular apoptosis regulating cell death programming, axonal retraction, and synaptic pruning. Synaptic pruning in particular is a very important process that serves to eliminate redundant or anomalous circuits.

Structural MRI has been informative in morphometric studies, but also in WM development

in early works. In fact, correlation studies<sup>103,179</sup> between histology and high-resolution structural MRI scans of post-mortem fetuses have been performed. Their findings revealed the presence of three identifiable fiber systems in fetuses as young as 12 GW. These systems include the corpus callosum (CC), the fornix, and the hemispheric stalk. The hemispheric stalk is a massive connection between the telencephalon (cerebrum) and the diencephalon (areas between the cerebrum and the brainstem) that contains all the projection fibers of the developing internal capsule.

Between 17 and 24 GW (mid-fetal period), there is a notable development of major cerebral fiber systems in intermediate zone (the to-be white matter). The fiber architectonics of the fetal cerebrum exhibits a tangential axon strata in the fronto-polar and occipito-polar regions. The fornix is well developed below the corpus callosum (CC), while the CC, internal capsule, and external capsule are growing. In the central white matter, the "periventricular crossroads" are the intersections of the major fiber systems, which include the callosal fibers in the transverse direction, associative fibers in the sagittal direction, and thalamo-cortical/cortico-fugal fibers in the radial direction<sup>54</sup>.

Between 24 and 32 GW, there is a notable development of the corona radiata, which arises from the transformation of the tangential fetal fiber-architectonic stratification. During this period, all major segments of the cerebral white matter (WM) can be identified, including the corpus callosum (CC), corona radiata, centrum semiovale (CS), and gyral white matter. The fibers continue to grow at the periventricular crossroads and the ventricular part of the CC, and by term birth (~40GW), all major fiber systems should be fully developed and in place.

After the process of maturation, fibers can be clustered according to their connection patterns into three categories: (1) *Commissural fibers* connecting the two brain hemispheres, mainly the corpus callosum; (2) *Projection fibers* connecting bidirectionally the cortex and thalamus, and the cortex with the brain stem and the spinal cord; (3) *Association fibers* encompassing cortico-cortical connections such as the arcuate fasciculus that connects the frontal and the temporal lobes.

While post-mortem studies are very valuable, they do not completely reflect the *in vivo* normal

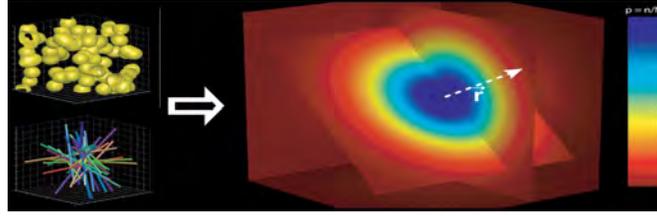
development. Thus, other techniques are needed to study early stages of brain maturation both *in vivo* and non-invasively.

### 1.3 DIFFUSION MRI

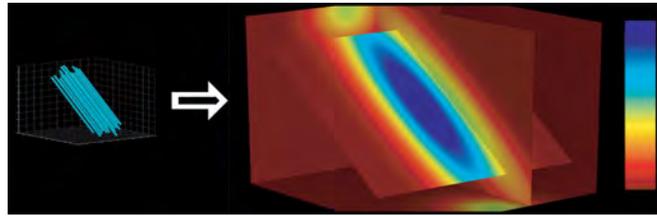
Before the progressive development of magnetic resonance imaging (MRI) as we know it today, **Nuclear Magnetic Resonance (NMR)**<sup>23,177</sup> was used to measure the chemical structure of substances. MRI relies on hydrogen nuclei that are exclusively composed of hydrogen atoms. After magnetic stimulation, they create a signal that is transformed to an image by the application of the Fourier transform. Since the signal received comes from all the tissue in the scanner, the typical way to incorporate localization is to introduce magnetic gradients, where the magnetization properties will depend on the location. This is performed by applying varying gradients that are weaker than the main magnetic field. Three gradients are necessary to produce a voxel (i.e. 3-dimensional) signal, namely slice, frequency and phase encoding gradients.

T<sub>1</sub>-weighted and T<sub>2</sub>-weighted MRI images are the most commonly employed imaging techniques in MRI. T<sub>1</sub>-weighted images rely on the longitudinal relaxation time of protons in tissues, which reflects the time required for spinning protons to realign with the main magnetic field. In contrast, T<sub>2</sub>-weighted images are more sensitive to variations in the decay rate of protons and reflect the time required for spinning protons to lose phase coherence among the nuclei spinning perpendicular to the main magnetic field, known as the transverse relaxation time.

**Diffusion MRI (dMRI)**<sup>131</sup>, with its unique capabilities of imaging the white matter both *in vivo* and non-invasively, relies on the Brownian motion of water molecules. First described by Robert Brown in 1827, after examining pollen grains suspended in water under a microscope, he observed a continuous jittery motion. In 1905, Albert Einstein modelled the motion of these pollen particles in one of his first major scientific contributions. The average particle displacement  $r$  (Equation 1.1) de-



(a) Isotropic diffusion



(b) Anisotropic diffusion

**Figure 1.6:** (a) Isotropic and (b) anisotropic illustration of Gaussian diffusion. Figure from Hagmann et al.<sup>80</sup>.

depends on the displacement time  $\Delta$ , on the particle size as well as the fluid temperature and viscosity ( $D$ )

$$\langle r^2 \rangle = 2\Delta D \quad (1.1)$$

In a homogeneous medium, the diffusion distribution, also called the displacement distribution is Gaussian. Depending on the shape of the medium, the diffusion can be either isotropic, when its geometry is independent of directions such as in the brain cerebro spinal fluid (CSF), or oppositely anisotropic when the diffusion occurs more along one specific direction, such as in coherent white matter fiber tracts. Figure 1.6 shows the probability density function of both configurations.

The equations of diffusion NMR were derived from Bloch equations<sup>23</sup> in 1964 by Stejskal and Tanner<sup>201</sup> (Pulse Gradient Spin Echo, PGSE) before the invention of the 3D spatial localization of the NMR signal using the three gradients<sup>128,143</sup>. Because of the complexity of the interaction

of these three gradients with a fourth gradient, i.e. the diffusion gradient along a given direction, Le Bihan et al. <sup>129</sup> suggested the *b-value* term to pack exclusively the terms that only depend on the sequence parameters (Equation 1.2).

$$\frac{S_g}{S_0} = \exp \left[ -\gamma^2 G^2 \delta^2 \left( \Delta - \frac{\delta}{3} \right) D \right], \quad b = \gamma^2 G^2 \delta^2 \left( \Delta - \frac{\delta}{3} \right) \quad (1.2)$$

$S_g$  is the diffusion weighted signal along a gradient direction  $g$ ,  $S_0$  is a non-diffusion weighted signal (commonly denoted as  $b_0$ ).  $\delta$  is the gyromagnetic ratio that depends on the isotope used (e.g. 42.58 MHz  $T^{-1}$  for hydrogen proton) and  $G$ ,  $\delta$  and  $\Delta$  are the PGSE sequence parameters, respectively the magnitude of the gradient pulses, the time of each of the two gradient pulses and the time between the two gradient pulses.  $D$  is the diffusion coefficient that can be extracted by fitting a model to the measured data. The principle of the the PGSE sequence is to create a gradient along the direction we aim to measure the diffusion using a radio frequency pulse. Hence the spins that are present along that direction will resonate at different frequencies. After a certain time (i.e.  $\Delta$ ), an inverse gradient is applied and the spins that did not diffuse will have a zero net magnetization compared to the moving spins that will experience a phase shift. This phase shift will translate to a signal loss or a low intensity voxel in the case of images.

It's only in 1985-1986, i.e. 11-12 years after the elaboration of the diffusion NMR equations, that the first MR images of the *in vivo* human brain were acquired using a whole-body sequence along one gradient direction <sup>129,130</sup>. Few years later, diffusion was performed along more directions in a cat brain <sup>157</sup>. In fact it was known that diffusion should be different between gradient directions because of its non-restriction in the direction along the fibers compared to the orthogonal direction. **Diffusion MRI captures the *microscopic* displacement of water molecules.** A typical isotropic voxel of dMRI in the adult brain ( $2mm^3$ ) contains an order of magnitude of  $10^{23}$  water molecules and hence captures a statistical displacement.

The most common model in dMRI is **diffusion tensor imaging (DTI)**<sup>17</sup>, that makes the assumption that the water diffusion is Gaussian. The displacements can then be characterized by the covariance matrix of the Gaussian function (Equation 1.3), namely the diffusion tensor, encompassing information about diffusion along the three axis. Because diffusion is supposed to be symmetric, the diffusion tensor contains 6 unique parameters and hence the need for at least 6 diffusion acquisitions to solve the system of equations. In practice, more than 6 measurements are needed because of the stochastic noise.

$$p(r) \triangleq e^{-r^T \Sigma^{-1} r}, \text{ with } \Sigma = 2\Delta D \quad (1.3)$$

On the image scale, the diffusion tensor is a 6D object, i.e. a 3D tensor for each voxel, posing a problem of visualization. Hence the reason to extract 3D scalar maps from it. For that aim, the tensor is diagonalized to extract the eigenvalues that represents the diffusion strength along each axis on the reference frame of the eigenvectors. For instance, if the eigenvalues are equal, we fall in the case of isotropic diffusion, otherwise the diffusion is anisotropic (Figure 1.6). Using these eigenvalues, one can extract several maps such as the **fractional anisotropy (FA)**, the **mean diffusivity (MD)**, the axial diffusivity (AD) or the radial diffusivity (RD).

FA represents a normalized measure of the variance of the eigenvalues, informing us about how different the diffusion is between the three main axes. MD provides an average of the diffusivity, i.e. independently of the directions. AD ignores the two weakest diffusion directions and provides information about the main diffusion direction as opposite to RD, which quantifies the diffusion perpendicular to the main direction. This is formulated in the equations below where  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  are the three eigenvalues of the diffusion tensor, where  $\lambda_1$  and  $\lambda_3$  are respectively the largest and the smallest.

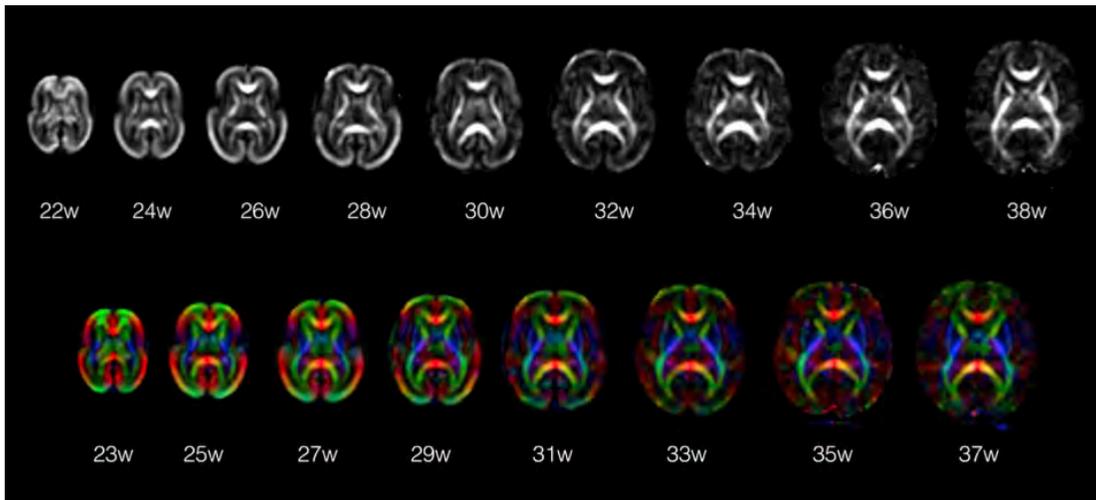
$$FA = \sqrt{\frac{1}{2} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}} \quad (1.4)$$

$$\text{MD} = \frac{1}{3}(\lambda_1 + \lambda_2 + \lambda_3) \quad (1.5)$$

$$\text{RD} = \frac{1}{2}(\lambda_2 + \lambda_3) \quad (1.6)$$

$$\text{AD} = \lambda_1 \quad (1.7)$$

Another way to illustrate a diffusion tensor is to represent each voxel by a color code that corresponds to the main diffusion direction. This map can be modulated by the FA values (Figure 1.7).



**Figure 1.7:** Diffusion tensor fractional anisotropy (FA) maps (top row) and color FA maps (bottom row) across gestational weeks (adapted figure<sup>117</sup>).

The assumption that the diffusion is Gaussian is easily violated in practice, meaning that DTI provides a limited and gross reconstruction of the anatomy. In fact, **DTI fails at reconstructing different fiber configurations such as fiber crossings**, occurring in 30% to 90% of the voxels<sup>245,101</sup>. Furthermore, under certain circumstances, the gaussian assumption does not hold for

complex tissues with different compartmental characteristics such as glial cells, that vary in shape and in sub-cellular structures. The DTI hypothesis holds less in high b-value acquisitions ( $b \geq 1500s/mm^2$ ). In that case, water molecules have a higher probability to collide with the different structures of the tissue and the diffusion will be along different axes and with different strength. This scenario cannot be captured by a Gaussian model because of its limited degrees of freedom<sup>49</sup>. Empirically, the linearity of the  $\log(S_g/S_0)$  (Equation 1.2) is not holding and a quadratic term in b has to be added to correctly model the diffusion signal. Similarly to DTI maps, diffusion kurtosis imaging (DKI) maps can be extracted by acquiring several diffusion weighted images and computing a diffusion kurtosis tensor.

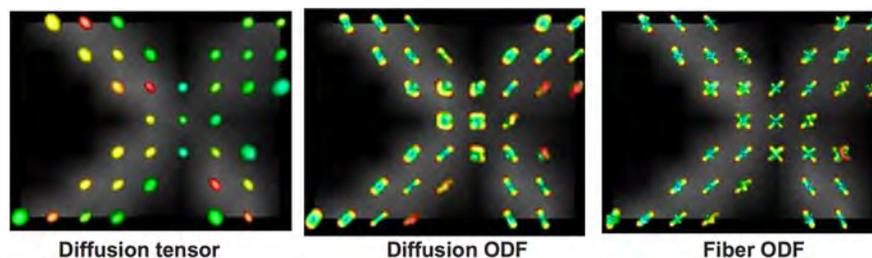
Other models that aim to overcome the major limitations of DTI, i.e. its inability to correctly represent fiber crossings or high-curvature areas, rely on the **Orientation Distribution Function (ODF)**. ODFs are mathematical formulations of the probability distribution of fiber orientations. Two predominant types of ODFs are used in diffusion MRI. The first is the diffusion ODF (dODFs)<sup>220</sup> that can be intuitively comprehended as a deformed sphere where the radius at each point is proportional to the diffusion in the direction between the center of the sphere and that point. It is defined in Equation 1.8 where  $\mathbf{u}$  is the direction along which we want to estimate diffusion, but since we do not know the probability density function  $P$ , the Funk-Radon transform<sup>218</sup> is used to estimate it.

$$\Psi(\mathbf{u}) = \int_0^\infty P(\alpha\mathbf{u})d\alpha \quad (1.8)$$

Diffusion ODFs can correctly reconstruct a broad range of fiber crossing angles, approximately between  $70^\circ$  and  $90^\circ$ , however it does not have enough degrees of freedom to disentangle angles between  $40^\circ$  to  $60^\circ$ . **Fiber ODFs (fODFs or FODs) can resolve these crossings using more degrees of freedom.** The underlying models are Spherical Deconvolutions (SD)<sup>217,100,28</sup>, that suppose that

the observed DWI signal is the result of convolving the true fiber distribution with a *response function*, where the latter can be estimated from a region containing one single fiber population. Hence to approximate the true fiber distribution, one has to perform the inverse operation, i.e. deconvolution, that is constrained on the positivity of the ODF. Fiber ODFs aim to directly model the fibers orientation as opposite to diffusion ODFs that aim to model the diffusion of water molecules. Figure 1.8 shows the reconstruction of a crossing area of a biological phantom using the three models. In the recent years, multi-shell multi-tissue constrained spherical deconvolution (MSMT-CSD)<sup>100</sup> became the most widely used model as it generates more accurate fODFs. In fact, the model takes into account the *response functions* of three different tissues: white matter as in CSD<sup>217</sup> but also gray matter and CSF.

Among the different models, DTI requires the lowest number of measurements: theoretically a minimum of 6 but practically 15-30 to cope with low SNR. ODF based models using SD typically require at least 30-60 gradient directions. But these numbers can vary depending on several factors including the b-value, the minimum angle of the crossing fibers and the SNR<sup>46</sup>.



**Figure 1.8:** Crossing fibers region in a biological phantom reconstructed by the DTI model, diffusion ODF and fiber ODF models. While the tensor fails at reconstructing the crossing area, ODF models can achieve this at different levels. Figure from Descoteaux et al.<sup>50</sup>.

## 1.4 THE *IN VIVO* DEVELOPING BRAIN WITH DIFFUSION MRI

Diffusion MRI (dMRI) is the tool of reference to shed light on white matter development *in vivo*. Diffusion acquisitions and models were mainly developed for adult brains. However, the structure of adult brains differs considerably from immature newborn and fetal brains.

### 1.4.1 CHALLENGES AND PROTOCOLS OF DEVELOPING BRAINS WITH dMRI

The developing brain has a high water content compared to the adult brain, which greatly affects diffusion imaging. Because of the increased diffusivity, some studies<sup>38,92,164</sup> have suggested a b-value for developing brain imaging to be around 700-800  $s/mm^2$ . Many research studies<sup>122,64</sup> have incorporated these b-values into their acquisition protocol. However, a single low b-value is insufficient to model complex anatomical configurations<sup>46</sup>. High b-values similar to adults protocols (1000-2000  $s/mm^2$ ) are used when applying the neurite orientation dispersion and density imaging (NODDI) model<sup>239</sup> to neonatal images<sup>99,107</sup>. Indeed, ideal protocol parameters may vary depending on the study and factors like age range and imaging goals. For instance, the Baby Connectome Project<sup>91</sup> has opted for a protocol that is optimal across the wide age-range studied longitudinally. Finally, most clinical studies stick with conventional protocols with high b-values (1000  $s/mm^2$ ) and 2 $mm^3$  resolutions<sup>141,219</sup>. Research protocols such as the one of the developing Human Connectome Project (dHCP) has optimized diffusion space sampling, allowing for acquisitions of a 300-sample multi-shell diffusion dataset in 20 minutes with a final resolution of  $1.17 \times 1.17 \times 1.5 mm^3$ . However, this research protocol can hardly be replicated in clinical settings.

Additionally, both neonates and fetuses are prone to involuntary movements such as spontaneous head and limb motions, which corrupt the acquired images. Limiting the acquisition time is crucial in mitigating motion-related challenges in developing brain MRI studies to ensure accurate interpretation and analysis of the imaging data. However, the limited time also translates to a small

number of acquired directions, hampering accurate estimation of fiber orientations using existing methods<sup>216,100</sup>.

**These constraints are even more pronounced in fetuses because of the unpredictability of the movements, the small fetal brain, the bigger distance of the fetal brain to the coils, and the presence of maternal tissue.** As a result, images with a low signal-to-noise ratio (SNR) are typically acquired. Hence the need to acquire thick slices (high through-plane resolution, 3-5 mm) to mitigate this problem. Some groups acquire several orthogonal acquisitions to improve the SNR and increase the probability of acquiring images along different directions because of fetal motion. Figure 1.9 summarizes typical fetal acquisitions in the different research teams.

While motion is aimed to be frozen by the echo planar imaging (EPI) acquisition within one slice, it can still occur within one volume (i.e. along one gradient direction) and across different volumes. Motion compensation strategies such as outlier rejection and registration are therefore more important in developing brains. Moreover, non-linear distortions typically occur because of the fast switching on and off of the diffusion gradients.

Finally, the contrast between gray and white matter in the fetal brain, especially in early gestation, is substantially blurred compared to adult brains, and hence methods such as state-of-the-art MSMT-CSD<sup>100</sup> may not be applicable in a straightforward manner.

#### 1.4.2 ***IN VIVO* STUDIES OF DEVELOPING BRAINS WITH DIFFUSION MRI**

Data obtained *in vivo* from pre-term infants and fetuses have corroborated findings from post-mortem investigations. The use of dMRI has allowed for the delineation of the early laminar organization of the cerebrum, which includes the cortical plate, subplate zone, intermediate zone, subventricular and periventricular zones, and germinal matrix, in newborns born at 25-27 gestational week (GW)<sup>138</sup>. Additionally, imaging studies of fetuses *in utero* have demonstrated the early depiction of the pyramidal tract, as well as the splenium and genu of the corpus callosum (CC), which

	Field strength & brand	#Directions	B-values (s/mm <sup>2</sup> )	Orthogonal stacks	Resolution (mm <sup>3</sup> )	Acquisition time in minutes:seconds	Main sequence parameters (TE/TR in ms)	Year
Jakab et al., Neuroimage: Clinical 2017	1.5T/3T GE	15	700	3 axial and 1-2 coronal	1x1x3-5	1:21 per 4D volume	TE = min(65) TR = 2200	2020
Deprez et al., IEEE Transactions on Medical Imaging 2019	3T Phillips	32	750	?	2x2x3.5	~5:00	TE = 75 TR = 7500	2020
Christiaens et al., ISMRM 2019 (DHCP)	3T Phillips	80	400 1000	?	2x2x2	~15:00 (whole dataset)	TE = 75 TR = 6100	2019
Marami et al., Neuroimage 2017	3T Siemens	12	500	2-8 axial/coronal	2x2x2-4	0:50-1:30 per 4D volume	TE = 60 TR = 3000/4000	2017
Jakab et al., Neuroimage 2015	1.5T Phillips	15	700	Axial	0.94x0.94x3.3	?	TE = 90 TR = 1745	2015
Fogtmann et al., IEEE Transactions on Medical Imaging 2014	1.5T Siemens	10	500 600 700	2 axial 2 sagittal 0-2 coronal	~ 2x2x2	00:10 per 4D volume	TE = 96 TR = 4000-6000	2014

**Figure 1.9:** Diffusion MRI fetal brain protocols across different centers. We can see the variability of acquisition schemes, the small b-values and the low number of directions (except for the research oriented dataset of the developing human connectome project, dHCP<sup>37</sup>).

can be reconstructed using tractography algorithms between 18 and 37 GW<sup>26,109,176</sup>. Furthermore, the Probst bundles can be identified in cases of CC agenesis<sup>108</sup>. Association tracts and subcortical projection tracts have also been identified in pre-term infants<sup>166,55</sup>.

In the last two decades, several studies have investigated the developmental changes in diffusion tensor imaging (DTI) measures. In fact, the focus was mostly on tracking fractional anisotropy (FA) and mean diffusivity (MD). Those studies have been then trying to explain these variations by suggesting links to microscopic changes that happen at large scale such as dendritic arborization, synaptic formation, neuronal differentiation or myelination.

Many studies reported a linear increase in FA and a decrease in MD over gestation<sup>183,26,191,88,95</sup>. However, non-linear changes in DTI measures or an initial increase in MD in some white matter (WM) areas have also been observed<sup>92,190,87,140,232</sup>. Schneider et al.<sup>190</sup> examined 78 fetuses with GA

ranging from 23 to 38 GW and found an initial increase in MD values in upper parts of the brain, followed by a decrease after 30 GW. In another work<sup>191</sup>, they also conducted a longitudinal analysis on 28 fetuses (21-34 GW) and reported a insignificant increase in MD in frontal WM. Similar MD patterns were also observed in another study<sup>87</sup>.

Because of limited sample sizes, insufficient gestational age range, and differences in data processing methods, apparent inconsistencies are observed across the above-mentioned studies. In addition, DTI metrics are not quantitative and depend on acquisition parameters such as the b-value.

Diffusion-weighted imaging (DWI) can give insights into the WM organization. For instance, Khan et al<sup>117</sup> reconstructed major fiber tracts across gestational ages (Figure 1.7), albeit with various limitations. Jakab et al.<sup>96</sup> compared neurotypical brains with corpus callosal brains and characterized their differences by tractographic and connectomic analyses. Furthermore, Machavos-Riva et al.<sup>139</sup> analysed and characterized the tracts that connect the cerebellum to the brain stem, called the cerebellar penduncles of 81 fetal brains. Other fiber regions such as the corticospinal tract, the forceps major and the inferior fronto-occipital fasciculus were studied<sup>140</sup> across 16 gestational weeks (23-38).

Apart from individual subjects analyses, **three fetal diffusion atlases exist today**. The first that was released is a DTI atlas<sup>71,79,117</sup> that used 67 dMRI scans of neurotypical fetuses between 22 and 38 GW. Two recent atlases<sup>34,232</sup> use more diffusion measurements and higher b-values and aim at reconstructing atlases relying on more advanced models than DTI. One of them is composed of 113 subjects from 22 to 37 GW<sup>232</sup> and the other one relies on 89 scans for subjects in the range 24-38 GW<sup>34</sup>. The former study aimed at characterizing diffusion metrics of five main fiber bundles, namely genu and splenium of corpus callosum, optic radiation, inferior longitudinal fasciculus and cortico-spinal tract. Based on diffusivity trajectories, the latter<sup>34</sup> speculated that the inflection point of the diffusivity trajectory may be aligned with pre-myelination.

## 1.5 CLINICAL MRI APPLICATIONS IN DEVELOPING BRAINS

Fetal brain MRI is mostly dominated by T2-weighted imaging, compared to diffusion weighted imaging (DWI) studies, and has its applications ranging from neuroscientific research as discussed above to pure clinical purposes.

**Clinically, fetal brain imaging is routinely performed with ultrasound** and used the follow-up of pregnancies and diagnosis of potential congenital pathologies. **Alternatively, MRI, that presents a very high sensitivity to depict soft tissue contrasts, is also indicated, if further investigation is needed to rule out, confirm or complement a suspected abnormality detected by ultrasound.**

In fact, Griffiths et al.<sup>77</sup> have conducted a prospective and multicentric study on 570 fetus, where they found a significant increase in accuracy diagnosis from 68% when only using ultrasound to 93% when combining it with MRI. This difference in accuracy was higher for high gestational ages.

MRI has been extensively supporting ultrasound in ventriculomegaly cases, i.e. where the ventricles are larger than normal. It can provide additional information<sup>76,127,189</sup> such as its cause or other misdiagnosed associated pathologies. Other brain abnormalities to which MRI can contribute in the diagnosis are posterior fossa malformations, corpus callosum agenesis or dysgenesis, and cortical malformations<sup>11</sup>.

Additionally, **DWI can be used in the detection of cytotoxic or vasogenic edema**, for example in hypoxia-ischemia, congenital infection, inherited metabolic diseases, or pregnancies at risk of fetal brain damage<sup>25,69,79</sup>. The diffusion signal intensity depends on the underlying condition. For instance diffusion is restricted in the case of cytotoxic edema, in ischemia for example, due to swelling of cells when extracellular water goes into cells (intracellular edema). On the other hand diffusion is increased in vasogenic edema, when the blood brain barrier is injured and fluid leaks from capillaries into white matter (extracellular edema).

## 1.6 MACHINE LEARNING AND DIFFUSION MRI

Prior to detailing the application of machine learning to diffusion magnetic resonance imaging, a concise overview of the principles of machine learning and artificial intelligence will be provided.

### 1.6.1 MACHINE LEARNING OVERVIEW

**Artificial intelligence (AI) can be defined as the ability of a machine to perform a task that was believed to be exclusively in the realm of humans.** It can be categorized into symbolic and non-symbolic AI. Non-symbolic AI and particularly *connectionism* aims at mapping an input to an output, for example an image to a label or a wave sound like speech to its written text, by finding correlations between the input and output, without necessarily being able to (humanly) explain the decision process. On the other hand, symbolic AI, such as rule-based systems, operates in an *if-else* paradigm and hence can trace back the decision process. Today, most successful AI systems fall into the non-symbolic category and are typically referred to as *machine learning* algorithms. Machine learning can be further categorized into supervised, unsupervised and reinforcement learning. Supervised learning needs labels to be trained, such as the label tumorous/non-tumorous in a medical image, whereas unsupervised learning does not rely on labels and aims to find patterns in the data using for instance clustering algorithms such as K-means<sup>82</sup>. Self-supervised learning was recently introduced and is somehow a hybrid version of the two paradigms, aiming at training an algorithm to predict a part of the input data from other parts of the same input data. Reinforcement learning, mostly used in robotics, cognitive architectures and games such as chess, aims at inferring a function from a reward signal. Both supervised and unsupervised algorithms were employed in this thesis as opposite to reinforcement learning that is rarely employed in medical imaging and thus will not be further discussed.

Example of popular machine learning algorithms are support vector machines (SVM)<sup>224</sup>, Ran-

dom Forest<sup>86</sup> and **artificial neural networks (ANN)**<sup>147</sup>. The latter is of particular interest and was named after neuronal networks of the nervous system because of their local basic computations, i.e. firing when action potential threshold is reached, and their global capabilities when working in synchronicity generating high-level capabilities such as learning or detection. *Deep learning (DL)*, a sub-field of machine learning that is the most successful today because of the maturity of the methods developed in the last few decades such as the back-propagation algorithm<sup>186</sup> and the availability of sufficient computer power (including Graphical Processing Units, GPUs) and large datasets (the *big data* era). A consequence of efficiently training these algorithms is the automatic *feature learning* from the input data that is not anymore performed by a human operator as in classical machine learning. The algorithm automatically extracts relevant features that depend on the current task, by efficiently finding the ones that discriminate the best the output, in the case of supervised learning for example.

Deep learning is the field of deep *artificial neural networks*, i.e. ANNs with several layers, as opposed to shallow ANNs that have one middle layer between inputs and outputs. In fact, the success of deep learning greatly resides in the high number of neurons distributed over several layers, i.e. over-parametrized networks<sup>196,6,237</sup>. Although *over-fitting*, i.e. overly fitting the training data distribution including the noise, may happen when the number of data points is low. There are several theoretical frameworks that try to rigorously formulate the generalization problem, i.e. being able to generalize to unseen data, such as the Vapnik-Chervonenkis (VC) dimension<sup>225</sup> or Rademacher complexity<sup>16</sup>, but their applications to deep neural networks remains unclear. In fact, it is not straightforward because of several factors such as the architecture used, the hyperparameters, information about the diversity within the dataset and fundamentally the absence of a general theoretical framework for deep learning. Several strategies however were suggested to tackle this problem such as dropout<sup>84,199</sup> or data augmentation that were both employed in this thesis.

**The cost of the great success of deep learning is the opacity of its models, often referred**

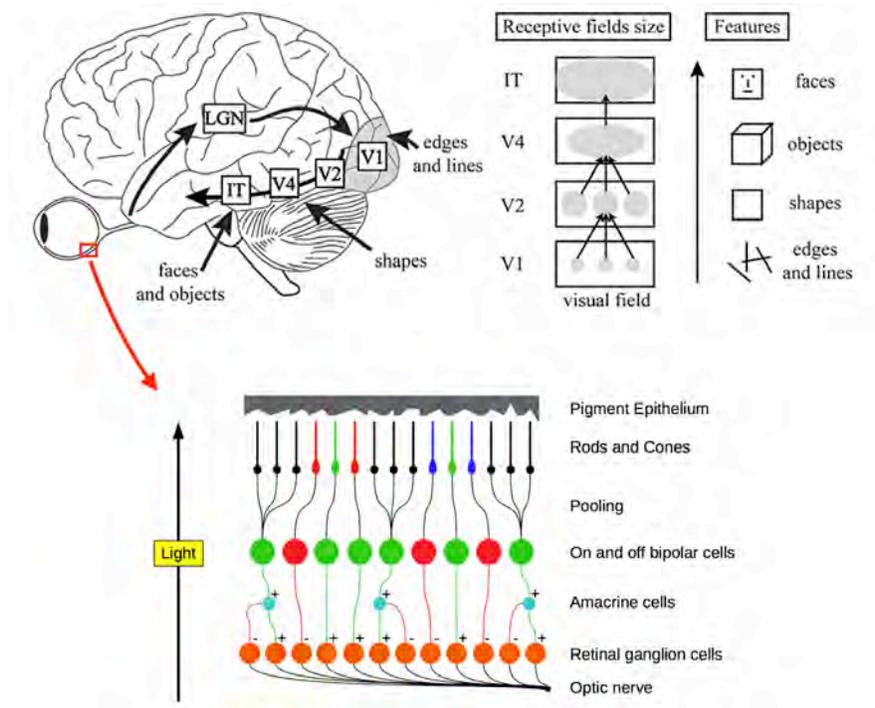
to as *black boxes*. Unlike symbolic AI, a deep learning network can hardly be broken down to be (humanly) understood because of the high complexity of non-linear correlations that the network learns between the input and the output. Several groups working in theoretical deep learning have been trying for instance to understand regimes that govern high-dimensional learning with back-propagation and gradient-descent <sup>12,51,228</sup>. Other groups have been trying to unify both schools (i.e. symbolic AI and deep learning) into one learning paradigm using cognitive models that are inspired from the brain that incorporates *inductive biases* <sup>20,73</sup>. Other groups in a field known as *explainability* <sup>14,195,155,193,203,10</sup> have been aiming at analyzing neural networks by studying the relevance of internal representations to predict the output.

Deep learning is essentially successful in two areas: (1) natural language processing (NLP) with applications such as machine translation <sup>241</sup> or text generation such as GPT <sup>178</sup>; and (2) computer vision which is the subject of this thesis, particularly medical imaging.

With the advent of deep learning, **computer vision was revolutionized by convolutional neural networks (CNNs or ConvNets)** <sup>132</sup>. This learning paradigm was in fact partly inspired by the neurobiology of the visual cortex. Figure 1.10 - top, shows the ventral stream of the visual system where a sensory input enters the eye, it is transferred by a thalamic structure called lateral geniculate nucleus (LGN) to primary visual cortex (V1) that contains neurons firing for low level features such as edges and lines. The signal is then gradually aggregated in high-level visual areas (V2, V4) to form shapes, objects or colors until reaching the final output such as a face in the fusiform face area (FFA) in inferior temporal cortex (IT), the brain area that is responsible for face processing.

CNNs share several properties with the visual system: (1) processing is divided into several consecutive layers (2) neurons process only a small subpart of the available information (3) neurons in each layer combine their inputs to higher order features (4), many neurons perform the same task for different parts of the input. Convolutions in CNNs are usually followed by pooling layers to reduce the data size and hence reduce parameters. Similar patterns can be found in the processing

system of the eye as can be seen in Figure 1.10 - bottom.



**Figure 1.10:** The ventral stream of the visual system where features are hierarchically reconstructed in the brain from primary visual areas to high-level visual areas such as the fusiform face area (FFA) in the Inferior Temporal cortex (IT). Eye processing is also displayed in the bottom figure. Convolutional neural networks (CNNs) bear several similarities to information processing of biological tissues. The figure is adapted from *neurdivness.wordpress.com*.

CNNs have proven to be successful in a wide range of computer vision tasks such as object recognition, object detection, image classification or image segmentation. While most methods are developed in the context of natural images, medical imaging with MRI is no exception to the success of CNNs. One of the first studies<sup>142</sup> in brain MRI using CNNs aimed at detecting multiple sclerosis using CNNs in the first layers and a feed-forward neural network in the last layer for binary classification.

Deep CNNs are widely used today in a variety of MRI tasks, from the segmentation of the brain<sup>241,154,118,3,188</sup>, or the segmentation of brain structures such as white matter, gray matter or the

cerebellum<sup>233,8,39,22</sup> or segmentation, detection and classification of lesions and tumors<sup>123,78,15,7,94,9</sup>. Segmentation with deep learning underwent a considerable boost after the advent of the U-Net architecture<sup>185</sup>. Its high performance is commonly explained by the joint learning of high-context low-resolution features and high-resolution low-context features thank to its skip connections. Registration and motion correction are also performed with deep learning nowadays<sup>43,56,29,197,89,235</sup>.

### 1.6.2 APPLICATIONS OF MACHINE LEARNING IN dMRI

Machine learning is also widely employed in diffusion MRI today. We will start by detailing classical machine learning application to dMRI, followed by deep learning applications in dMRI and finally machine learning studies in the specific population of newborns and fetuses.

### 1.6.3 CLASSICAL MACHINE LEARNING IN dMRI

Denoising dMRI images with a simple linear regression model in a self-supervised framework<sup>62</sup> has been successful in a variety of datasets. In this patch-based method, regressors are solely exposed to the signal contained within volumes apart from the target volume. By assuming noise independence across different volumes, i.e. predicting the noise of a target volume do not depend on the input volumes, the regressors acquire the ability to attenuate noise while retaining the signal of interest.

Other studies have employed Random Forest for several purposes such as Image Quality Transfer<sup>5</sup> to improve the quality of datasets that are acquired using limited acquisition times using the fine structural details of high-quality data sets with longer acquisition times. They were also used in mapping the DWI signal to fiber tractography<sup>163</sup> or for ischemic stroke identification<sup>152</sup>.

Support vector machine (SVM) is another machine learning algorithm that was utilized in diffusion MRI. In a study<sup>192</sup>, SVM was used to classify six distinct categories based on rotational invariant features extracted from spherical harmonic decomposition of a high angular resolution diffusion

imaging (HARDI) signal. The six classes were parallel and crossing neuronal fiber bundles in white matter, grey matter, partial volume between white and grey matter, cerebrospinal fluid and background noise. SVM was also used in the prediction of the survival rates in patients with brain cancer using diffusion and perfusion MR images<sup>61</sup>. Several other studies<sup>150,63,149</sup> of diffusion MRI using classical machine learning can be found in a review by Ravi et al.<sup>181</sup>.

#### 1.6.4 DEEP LEARNING IN dMRI

Deep learning was used for several purposes in dMRI such as the development of customized microstructure-sensitive loss functions in order to predict multishell data and enhance angular resolution<sup>33</sup>. Quality control<sup>74</sup>, data harmonization across scanners<sup>160</sup>, brain tissue segmentation<sup>238</sup>, detecting abnormalities in multiple sclerosis patients<sup>226</sup> and predicting motor outcomes from dMRI pre-term data<sup>187</sup> was also performed using deep neural networks.

**Deep learning in diffusion MRI have been first used in a pioneer work by Golkov et al.<sup>72</sup>** with the goal of (1) reducing acquisition time and (2) decreasing the error accumulation of sub-optimal modeling steps that are routinely employed in dMRI. In fact, these processing steps can be thought of as handcrafted feature extraction. Golkov et al.<sup>72</sup> suggest to directly learn the object of interest, in particular diffusion kurtosis maps and neurite orientation dispersion and density imaging (NODDI)<sup>239</sup> maps by learning a mapping using an artificial neural network between a low number of DWI measurements and the latter maps reconstructed using all available measurements. That way, the error is fully back-propagated and minimized in an end-to-end paradigm. A 12-fold gain in the number of samples required and hence in acquisition time was obtained. A similar approach will be used in this thesis in the case of fiber tracts estimation.

Another work used deep learning and spherical harmonics decomposition to estimate data of a non-acquired shell from existing shells<sup>120</sup>. Another work<sup>209</sup> have trained a deep CNN to predict six high quality diffusion images along with one bo derived from the tensor reconstructed from all

available measurements, using as input structural images (T<sub>1</sub>- and T<sub>2</sub>-weighted) and the original six diffusion measurements and one bo. Their results showed that the predicted volumes showed comparable performance as a fully sampled scan in terms of diffusion tensor and tractography. In contrast, Li et al.<sup>134</sup>, have directly predicted tensor maps with six diffusion measurements with CNNs. These maps did not show a significant loss in accuracy compared to maps computed with 60-90 directions, even in pathological subjects, where lesions were still visible.

Other works have been trying to estimate fiber Orientation Distribution Functions (ODFs) using deep learning. A study from Koppers et al.<sup>121</sup> aimed to estimate the number of fibers and the fiber orientation of a fiber ODF by discretizing the angles into 250 ranges using CNNs and spherical harmonics basis. Nath et al.<sup>162</sup> used the ex-vivo dMRI signal to predict fiber ODFs reconstructed from histology ground truth. Another work<sup>135</sup> deployed a 3D CNN to predict them using a small neighborhood of the diffusion signal. Hosseini et al.<sup>90</sup> used a two-stages Transformer-CNN to first map 200 directions to 60 directions and the latter to fiber ODFs.

Instead of predicting fiber ODFs, another pioneer work in tractography is named TracSeg<sup>230</sup>. In fact, it directly aims to perform tractography, using a segmentation paradigm. It learns a mapping using a 2.5D CNN (a CNN that takes 2D images from the three planes) in the shape of a U-Net<sup>185</sup>, between the peaks extracted from applying constrained spherical deconvolution (CSD) to the DWI images, and the corresponding segmented tract, with a total of 72 tracts. The authors argue that they did not aim at using the original DWI images as input data to the network, to be more sequence agnostic. However, this happens at the cost of applying a model (i.e. CSD) prior to the learning process.

Another study have been trying to learn non-acquired gradient directions from acquired ones, instead of directly learning end-goal quantities (tracts, fiber ODFs, tensors or scalar maps). While spherical harmonics decomposition can provide an interpolation basis for that, deep learning was also used by Lyon et al.<sup>137</sup> using a recurrent CNN in an encoder-decoder architecture. Their results

show that the smaller number of input directions, the more the model outperforms SH interpolation.

Autoencoders were trained to learn a regularizer used to reformulate CSD and achieved competitive results with low b-values and a low number of diffusion directions<sup>168</sup>. Other works have tried to improve spatial resolution using deep learning in diffusion such Tanno et al.<sup>205</sup> who learned a mapping with CNNs between low resolution and high resolution images. Their network additionally outputs an uncertainty map that is shown to be correlated with prediction errors. Super-resolution was also performed in other works with CNNs<sup>60</sup>, a customized U-Net<sup>32,185</sup> and generative adversarial networks (GANs)<sup>4</sup>.

Recently, models<sup>161,59</sup> that take into account the spherical properties of the dMRI data, i.e. each gradient volume come from a points in the sphere, are in the process of development. The signal in the diffusion signal can also come from several b-values, i.e. different shells, and mixing shells in training a network can be sub-optimal also if these properties are not taken into account. Deep learning of diffusion MRI will likely undergo a significant boost when these models will become well established, especially if they can incorporate the physics of dMRI in their learning framework.

### 1.6.5 MACHINE LEARNING IN DMRI OF DEVELOPING BRAINS

Most machine learning studies on developing brains relate to newborns. Perhaps one of the first machine learning works applied to neonates is the work of Ziv et al.<sup>244</sup> who analyzed brain connectivity differences between a control group and a group with encephalopathy using a supervised (SVM) and unsupervised models (Principal component analysis, PCA). Another study<sup>231</sup> used SVM and Random Forest to detect lesions and predict outcomes for neonatal hypoxic ischemic encephalopathy patients. Vassar et al.,<sup>227</sup> used logistic regression using metrics extracted from structural images and diffusion tensor maps find associations between these neuroimaging data obtained in the near-term and subsequent language function in a cohort of extremely pre-term infants. A similar

study<sup>221</sup> have aimed at predicting language outcomes of pre-term neonates using Random Forest on diffusion tensor metrics.

Deep learning has also been used in newborns dMRI in a diverse group of applications. First works related to DTI scalar maps harmonization across different scanning sites has also been performed using a GAN<sup>243</sup>. Similarly as previously done in adult dMRI, Karimi et al.<sup>105</sup> proposed to predict the fiber ODFs from the diffusion signal with a feed-forward neural network. The same group of researchers also showed deep learning promises for the generation of biomarkers from atlases with a high accuracy<sup>106</sup>. Finally, most recent works tackled a prediction task, where a ResNet<sup>83</sup> with spatial attention modules was trained to predict the acute bilirubin encephalopathy from multimodal data including MD of the diffusion tensor<sup>240</sup>.

Deep learning use in fetal diffusion MRI is however barely explored. To the best of our knowledge, only one study<sup>104</sup> have used deep learning on fetal brains. The authors study use pre-term data to (1) synthesize fetal data from the diffusion tensor (color FA) of the pre-terms, to which the noise of the fetal data is added and (2) predicts this color FA with this synthesized data.

## 1.7 CONTRIBUTIONS OF THIS PHD THESIS

This work has focused on improving acquired diffusion MRI (dMRI) images of developing brains using computer vision techniques and particularly deep learning. Specifically, we have aimed at enhancing spatial and angular resolution of dMRI images of newborns and fetuses. In this thesis we have used *in vitro* phantom acquisitions and real dMRI data. We mainly used a high quality research-dedicated dataset (the developing Human Connectome Project, dHCP<sup>93</sup>) to train neural networks on (pre-term) newborns and further adapt them to clinical acquisitions of fetuses from two different hospitals.

Our first contribution relates to the **spatial resolution enhancement**. Anisotropic images are

generally acquired in fetal and newborn (diffusion) MRI to freeze in-plane motion and increase signal-to-noise ratio (SNR). This is typically circumvented using conventional interpolation methods, before further downstream analyses (diffusion tensor or Orientation Distribution Function (ODF) estimation for example). In this thesis we proposed a convolutional autoencoder<sup>110</sup> to artificially increase the through-plane spatial resolution using a weighted average on the latent space representation. The same autoencoder, trained on pre-term newborns, was applied on fetuses. We have also explored the same idea on the spherical harmonics domain of the signal<sup>111</sup>.

Our second contribution focused on the **angular resolution enhancement to improve reconstruction model (ODFs)**. Our first approach, a deep learning framework<sup>137</sup> to infer new gradient directions (see Appendix 5) did not succeed. We proposed then to directly learn ODFs using a modified U-Net<sup>185</sup>. Our method trained on ODFs reconstructed with a high number of diffusion measurements (300) is able then to infer ODFs from very few samples (6-12)<sup>113,114</sup>. Such approach is very appealing for dMRI in developing brains as in that context very few directions are often available in clinical practice mainly due to limited acquisition times. The network trained on pre-term newborns has been tested on out of domain fetal brains and clinical newborns datasets.

Finally, the third contribution of this PhD thesis explored a more practical problem of **optimal acquisition schemes**<sup>115</sup>. We setup an *in vitro* study using a phantom with fetal-like diffusion properties (mimicking expected low FA values in white matter of fetuses) to explore different acquisition schemes. We conducted the study at 1.5 and 3 Tesla, and used these acquisitions (publicly available<sup>116</sup>) to quantify the limited time trade-off between the number of diffusion gradient directions and orthogonal volumes.

Beyond these methodological contributions on the topic of this PhD, I have contributed to other research projects related to fetal brain image analysis. I have developed a deep learning based brain extraction tool for T<sub>2</sub>-weighted images that is used in-house (see Appendix 5) and implemented in a publicly available software<sup>213</sup>. I have also collaborated in the development of T<sub>2</sub>-weighted brain

tissue segmentation methods<sup>39,126,41,170,40,171</sup>. I have also supported the study of super-resolution reconstruction of functional MRI of the fetal brain<sup>206</sup> (see Appendix 5). Last but not least, I have actively contributed to a software development project of MIALSRTK<sup>214,215</sup>, a BIDS (Brain Imaging Data Structure) application<sup>213</sup> which aims at making MIALSRTK, a super-resolution pipeline for fetal T<sub>2</sub>-weighted images, compatible with datasets formatted in the BIDS standards.

# 2

## Summary of Results

In this section, we summarize the main results of the papers included in the thesis. Namely, (1) single volume through-plane super-resolution with autoencoders in signal domain <sup>110</sup> and (2) in spherical harmonics domain <sup>111</sup>; (3) fiber orientation distribution function estimation with deep learning <sup>113,114</sup>; (4) a crossing-fibers phantom study <sup>115,116</sup>.

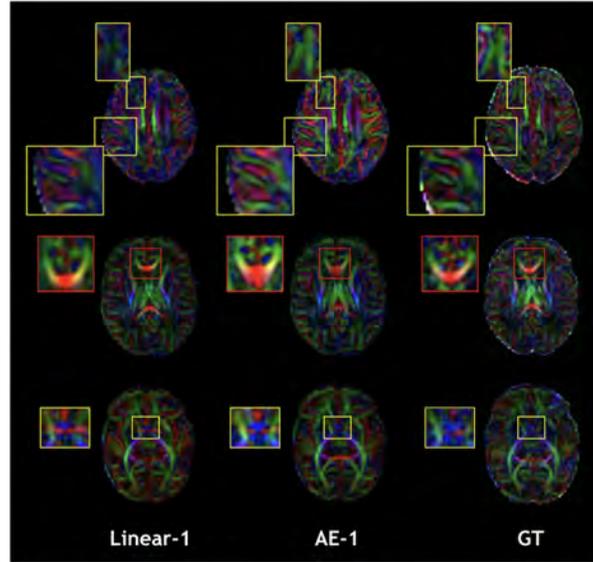
## 2.1 THROUGH-PLANE SUPER-RESOLUTION WITH AUTOENCODERS IN SIGNAL DOMAIN

This work<sup>110</sup> used the latent space of an autoencoder to enhance the through-plane resolution of diffusion magnetic resonance imaging (dMRI) data of pre-term newborns from the developing Human Connectome Project (dHCP) dataset. Quantitative and qualitative comparison was performed against typically employed interpolations, in terms of the original signal, both  $b_0$  ( $b = 0 \text{ s/mm}^2$ ) and  $b = 1000 \text{ s/mm}^2$ , and in terms of diffusion tensor maps.

The autoencoder was able to generalize to  $b = 1000 \text{ s/mm}^2$  despite being exclusively trained on  $b_0$  images. The second observation was the expected higher error in the configuration where two middle slices were removed and estimated/predicted, independently of the method used. Regarding the original signal ( $b_0$  and  $b = 1000 \text{ s/mm}^2$ ), the more slices were removed, the higher the error gap between the baseline interpolation methods and the autoencoder enhancement. The autoencoder had a significantly lower mean squared error (MSE) when compared to the baseline methods. The general superior performance of the autoencoder is also evident in the DTI maps compared to the optimal baseline technique (i.e., linear interpolation), particularly when two slices were removed. Notably, the colored FA plot revealed that the autoencoder produced tracts consistent with the ground truth (Figure 2.1). Moreover, the same network generalized to fetal data, where in a proof-of-concept, we qualitatively show the validity of the DTI maps. The fetuses used were acquired in a different scanner and with a different protocol than the network trained in the pre-term newborn data (Figure 2.2).

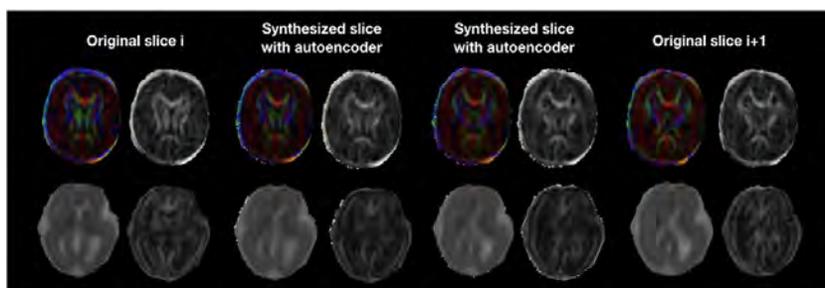
## 2.2 THROUGH-PLANE SUPER-RESOLUTION WITH AUTOENCODERS IN SPHERICAL HARMONICS DOMAIN

This study<sup>111</sup> is an extension of the previous work<sup>110</sup> where we explored the ability of our network to enhance the through-plane resolution in the spherical harmonics (SH) representation of the



**Figure 2.1:** Qualitative comparison of colored FA in the one slice removed configuration for the best baseline interpolation method, i.e. Linear-1 (left), autoencoder enhancement AE-1 (middle) and GT (right). The red frame area highlight to a region where the linear interpolation shows a more accurate result<sup>110</sup>.

signal. Additionally, we have compared the results (original signal and DTI maps) with the signal domain network<sup>111</sup> and with interpolations both in signal and SH representations, in the case of removing and estimating one or two slices. Results indicate that SH proved to be more accurate in reconstructing DTI-FA maps, while the raw signal was more suitable for estimating raw data. We hypothesized that this could be due to a global bias introduced to the back-projected raw signal by the SH-trained autoencoder. However, the SH autoencoder better preserved orientation information, resulting in better depiction of the FA, which is scale invariant. Finally, we successfully applied our method, trained on newborn data, to enhance the through-plane resolution of anisotropic fetal data, acquired in a different scanner with lower b-value and fewer gradient directions.



**Figure 2.2:** Colored FA and FA (top row) illustration of autoencoder enhancement between two original adjacent fetal slices in a still subject (35 GW). The bottom row shows a similar illustration of MD and FA for a moving subject (24 GW)<sup>110</sup>. Image generated with the autoencoder of section 2.1.

### 2.3 FIBER ORIENTATION DISTRIBUTION FUNCTION ESTIMATION WITH DEEP LEARNING

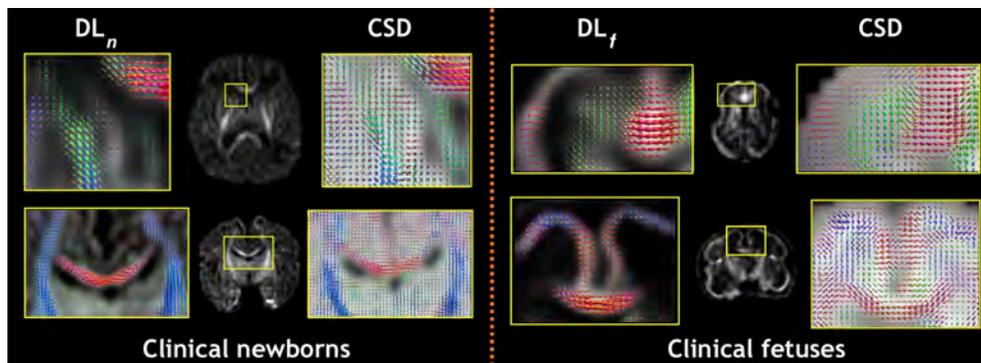
We propose in this work<sup>113,114</sup> a solution to address the challenge of estimating fiber orientation distribution functions (FODs) from a limited number of diffusion measurements in newborns and fetuses. Our approach involves using a deep learning method to map 6-12 diffusion gradient volumes to the target FOD reconstructed with 300 measurements. We conducted extensive quantitative evaluation, which demonstrated that our deep learning method using significantly fewer measurements can achieve comparable or superior results to standard methods such as Constrained Spherical Deconvolution<sup>216</sup>, Constant Solid Angle<sup>2</sup> or the Sparse Fascicle Model<sup>184</sup>. Comparison was performed for the number of predicted fibers, their orientations and the apparent fiber density (Table 2.1).

Furthermore, we demonstrated the out-of-domain generalization of the model to clinical datasets of fetal and newborn subjects (Figure 2.3), despite differences in acquisition parameters and anatomy relative to the training dataset. Domain transfer from research-oriented datasets to clinical datasets can be a promising approach for developing cohorts such as preterm and fetal populations that cannot afford long acquisition times. The proximity of the two acquisition b-values used in our

Method	b-values ( $s/mm^2$ )	$N_m$	Angular error (Agreement rate in %)			AFD error
			Single fibers	Two fibers	Three fibers	
$DL_n$	{0, 1000}	7	12.6°(78.4%)	24.2°(15.8%)	33.3°(3.8%)	0.27 ( $\pm 0.03$ )
CSD	{0, 2600}	148	7.5°(11.7%)	16.5°(16.5%)	27.2°(5.9%)	1.31 ( $\pm 0.23$ )
CSA	{0, 400, 1000, 2600}	300	47.0°(27.7%)	41.4°(14.8%)	36.1°(7.9%)	3.46 ( $\pm 0.46$ )
SFM	{0, 400, 1000, 2600}	300	51.4°(83%)	40.7°(10%)	35.4°(3.5%)	0.80 ( $\pm 0.55$ )
$\Delta GS$	{0, 400, 1000, 2600}	150	13.8°(80.4%)	29.1°(30.3%)	35.4°(27.9%)	0.2 ( $\pm 0.025$ )

**Table 2.1:** Mean angular error, agreement rate on number of peaks and Apparent Fiber Density (AFD) error between GT (MSMT-CSD) and the different methods.  $\Delta GS$  refers to the agreement within the ground truth. The number of measurements ( $N_m$ ) and the b-values used are also reported. All results were statistically significant compared to  $\Delta GS$  ( $p \leq 9e^{-10}$  for angular error, except SFM three fibers, and  $p \leq 4.5e^{-3}$  for AFD error)<sup>113</sup>.

training data ( $b = 400 s/mm^2$  and  $b = 500 s/mm^2$ , respectively) likely contributed to the successful generalization of the model from preterm to fetal datasets.



**Figure 2.3:** The deep learning method compared to CSD in different brain regions for 2 newborn subjects (left) and 2 fetal subjects (right) of 25 (top) and 29.4 (bottom) weeks of gestation. FODs are superimposed to the first SH coefficient of the method used. The DL first SH coefficient is shown in full-size too<sup>113</sup>.

## 2.4 A CROSSING-FIBERS PHANTOM STUDY

In this study<sup>115</sup>, we aimed at quantifying the acquisition trade-off between orthogonal volumes and the diffusion gradient volumes, given the limited time of acquisition in clinical settings for the developing cohort. For that aim, we have extensively scanned a fetal customized crossing phantom<sup>116</sup>,

with a high-resolution (angular and spatial) that we consider as a pseudo ground-truth and low resolutions fetal-like acquisitions. We observed that diffusion properties exhibited variability, which was effectively reduced through scattered data interpolation of multiple volumes. Interestingly, we found that increasing the number of directions did not consistently reduce error metrics such as FA, MD and fiber orientations.

# 3

## Discussion

This section will discuss the contributions and results reported in this PhD thesis, placing them into a broader context. Additionally, we will outline the forthcoming actions to address the limitations. Finally, we will conclude the thesis by sharing some reflections and comments on the potential impact of the research.

### 3.1 MAIN CONTRIBUTIONS

The goal of this PhD work was to investigate different strategies for super-resolution in diffusion MR images to support early human brain development studies. Given the various caveats and limitations linked to this population, namely the low signal-to-noise ratio (SNR) and the limited scanning time, we have tackled both spatial and angular resolution enhancement in the context of *in utero* fetal-imaging and also newborns. We have first explored the enhancement of the raw diffusion signal in the spatial and then in the angular domain. We have then bypassed the later step, because of the complexity of the problem, by directly learning a model, namely orientation distribution functions estimated with multi-shell multi-tissue constrained spherical deconvolution<sup>100</sup> from few available measurements. Specifically, the main methodological contributions of this thesis are the following:

1. Proposing an unsupervised learning model using autoencoders to increase the spatial through-plane resolution outperforming conventional interpolations methods
2. Developing a supervised learning model to improve microstructure estimation using a low angular resolution
3. Quantifying the trade-off between the number of gradient directions and the number of orthogonal volumes in fetal acquisitions using a crossing phantom

In the following paragraphs, we will further discuss these contributions and highlight their impact in the medical imaging community.

The first research project focus of this PhD thesis was to tackle the enhancement of spatial resolution in diffusion MRI (dMRI), primarily in the context of anisotropic dMRI volumes or in the context of slice outlier recovery of in newborns and fetuses. Traditionally, conventional interpolations such as B-spline or linear are used for that aim. However, these methods do not exploit semantic

contextual information. To address this problem, we proposed the use of **autoencoders to increase the through-plane resolution**. Through-plane resolution in developing brains is typically higher than in adults. In fetal dMRI, the values typically range between 2 to 5  $mm^3$ <sup>97,48,37,144,67,96</sup> and acquisitions are mostly anisotropic (Figure 1.9). In comparison, for newborns, isotropic acquisitions (2-3  $mm^3$ ) are more common<sup>122,219,64</sup>. However, still major datasets including the developing human connectome project (dHCP)<sup>93</sup> encompasses data with anisotropic resolution. In circumstances where sequence optimization cannot be achieved, post processing algorithms are essential for more accurate downstream analyses. In that regards, several methods including the work presented in this thesis, have been aiming to address this issue in (diffusion) MRI with the help of machine learning. Unsupervised methods such as the one presented in this thesis are very valuable in the high variability that is presented in fetal and newborn brain scans. In fact, besides variations in (through-plane) resolution, acquisitions also vary in the b-value (400, 500, 600, 700  $s/mm^2$ ) and gradient directions<sup>97,48,37,144,67,96,122,219,64</sup>, as there is no consensus on optimal acquisition parameters for the population with developing brains.

As most deep-learning based super-resolution (SR) methods are supervised-learning methods<sup>35,173,60,52,32</sup> (except few<sup>242,124</sup> that do not require external data), their applicability for a broad range of acquired dMRI protocols is limited. On the contrary, the method presented in this PhD thesis is agnostic to the magnitude of the through-plane resolution, the b-value and the gradient directions and hence is very valuable in the context of developing brains.

The second contribution of this thesis aimed at addressing a direct consequence of limited acquisition time that is the low number of diffusion measurements. This caveat is more pronounced in fetal brains for which clinical acquisitions inevitably do not go beyond few minutes<sup>97,48,144,67</sup>. Hence, only simple models such as the diffusion tensor can be afforded. However, current state-of-the-art reconstruction models<sup>216,100</sup> require a significant number of directions (and high or multiple b-values)<sup>46</sup> in order to estimate accurate diffusion metrics such as the number of fibers, fiber

orientations or the apparent fiber density<sup>180</sup>.

To tackle this problem, we have proposed a **deep neural network to accurately learn a mapping between a small number of diffusion directions and the orientation distribution function (ODF) estimated with densely-sampled multiple-shell data** that is impossible to acquire in a clinical environment. Extensive validation on 320 dHCP newborn subjects has shown that our model performs levels on par as state-of-the-art methods that require significantly more data. Additionally, as a proof of concept, we demonstrate the generalizability of the network on 26 clinical subjects. Among which, 11 fetuses that were validated by an expert neuroanatomist and compared to histological data of equivalent gestational age.

Such mappings can save more than an order of magnitude in acquisition time, without compromising accuracy. In fact, frameworks<sup>209,134</sup> with similar goal of mapping few measurements to high quality diffusion metrics have been developed for diffusion tensor (DTI) derived metrics. These works are very valuable but they come with the known limitations of DTI.

Reducing acquisition time by different means has always been at the core of MRI research. Sequences such as Fast Spin Echo (FSE) or Echo Planar Imaging (EPI), parallel imaging algorithms or compressed sensing are different attempts aiming at reducing scanning time. Deep learning, particularly through the utilization of large-scale datasets, is definitely one of those techniques that are and will greatly contribute to that aim. This will particularly benefit sensitive cohorts such as fetuses and newborns.

An outcome of the first two contributions of this thesis is **the use of large datasets (namely the dHCP dataset) acquired with high standards protocols, to feed machine learning models and learn patterns from a reduced version of the data**. This under-sampled data, whether spatial or angular, can represent clinically constrained datasets. In fact, in both works<sup>110,111,113,114</sup>, we have shown that the models that were exclusively trained on these high quality datasets, could generalize to clinical datasets that are reduced in both spatial and angular resolutions. This strategy of leverag-

ing high quality data has also been explored in another work<sup>104</sup> for fetal color FA prediction.

The third contribution of this PhD thesis relates to the quantification of dMRI accuracy versus acquisition time. As previously discussed, time is a limiting factor in dMRI acquisitions and in practice, a trade-off has to be made between the angular and the spatial resolution. Moreover, there is no general guidelines or consensus on optimal sequence schemes. To provide more insights on this aspect, we proposed **an *in vitro* quantification study of the time trade-off between the q-space angular dimension and the number of orthogonal acquisitions**. To that aim, we have scanned a crossing phantom, that was customized to have a low fractional anisotropy ( $\sim 0.6$ ) in single fiber areas to simulate the low anisotropy of developing brains. Pulsed gradient spin-echo (PGSE) sequences of low resolution fetal acquisitions (3-6 orthogonal series of  $1 \times 1 \times 4 \text{ mm}^3$ ; 9, 16 and 25 directions) at 1.5T were compared to a ground truth high-resolution acquisition ( $1.5 \text{ mm}^3$ ; 61 directions) at 3T. The results showed more stability with more orthogonal acquisitions, especially when they were combined with scattered data interpolation, and no additional benefit was observed with high angular schemes.

Since a small number of directions can be sufficient, deep learning was leveraged to compensate the missing directions, by directly estimating diffusion properties that would come from high resolutions images.

Few works<sup>236,229</sup> have tried to address the spatial-angular trade-off, but not in the context of developing brains nor in the case of multiple volumes acquisitions. Investigating this trade-off<sup>229</sup> has suggested while maintaining an acceptable level of spatial resolution, higher angular resolution has a more substantial impact on improving tract reconstructions. As resolutions higher than  $2 \text{ mm}^3$  benefit multi-fiber tractography methods only if angular resolution is not compromised. Another study<sup>236</sup> suggests that in deterministic tractography, a decrease in angular resolution had a more pronounced negative impact on tractography results. However, lower spatial resolution datasets (e.g.,  $2.5 \text{ mm}^3$ ) have exhibited a more drastic diminished quality in probabilistic tract reconstruc-

tions. In addition, the choice of the trade-off is highly dependent on the complexity of fiber configuration. By simulating crossing, brushing, and kissing fibers, the authors<sup>236</sup> suggest high spatial resolution is beneficial for resolving kissing fibers while, high angular resolution favors crossing fiber. Therefore, how a region is constituted by different fiber orientation may alter the optimal trade-off for that region.

## 3.2 LIMITATIONS AND FUTURE STEPS

### 3.2.1 NETWORK HALLUCINATIONS

Both methodological contributions of this thesis, the **spatial SR enhancement** and the **fiber ODF estimation using few measurements** make use of deep neural networks that are known to be black boxes. In the absence of a clear framework on how they internally operate, there is no guarantee that these models will not generate new non-existing structures or will not remove existing ones (also known as *hallucinations*, in analogy with the phenomenon of hallucination in human psychology). Hallucinations<sup>204,165</sup> occur when the network is trained on a small and not diversified dataset and hence becomes biased and too specialised. For instance, in the **spatial SR enhancement** work<sup>110,111</sup>, it is not impossible that a small lesion present in adjacent slices will be wiped out by the generative autoencoder.

This is a general problem in digital medicine in the deep learning era, and as of today, there is no trivial solution to it. There exist however some strategies that aim at mitigating the risk of hallucinations. For instance, if available, the inclusion of abnormal/pathological data in the training phase while guaranteeing an adequate ratio with normal cases. Another strategy to alleviate this problem would be the use of patch-based methods, as more local information with a reduced field of view (FOV) is used in this context.

In a recent work<sup>65</sup>, the problem of trusting learning-based systems in terms of the quality of the

network output, was tackled in the context of semantic segmentation. The authors suggest a fail-safe mechanism where prior expert knowledge is incorporated and merged a posteriori with the network. This problem bears similarities with the problem of not trusting the network predictions in the case of hallucinations, where a rule-based system can be combined as a post-processing step with the network prediction.

### 3.2.2 DOMAIN SHIFTS

Hallucinations can also be seen as a consequence of domain shifts, these domains can be for instance normal and pathological, different scanners, different acquisition protocols or different anatomies such as fetal and (pre-term) newborns.

While in our work **fiber ODF estimation using few measurements**, we did not explore pathological subjects, we have faced out-of-domain generalization problems for data of different scanners, protocols and anatomies. Normalization by  $b_0$  seemed to be a robust solution to reduce the dependence on the b-value (i.e. the protocol), however further steps can be explored for inter-scanner variability such as *data harmonization* techniques<sup>151,160,174</sup> that do not necessarily need paired subjects.

To overcome the limited availability of (high quality) fetal dMRI brain data, in this thesis we have leveraged data from pre-term neonates of the dHCP to train our neural networks. An implicit assumption that is made in both projects, is the hypothetical global **similarity between the pre-term and the fetal brain** and/or the ability of the network to generalize beyond the anatomy that it was trained on. In fact, for both works, the network trained on pre-term newborns has been shown to generate plausible fetal diffusion properties. Patch-based methods such as in our ODF estimation is again a strategy that can address this generalization problem. Moreover, data augmentation techniques, namely random noise injection and slight rotations, were crucial for the network to generate plausible FODs in the fetal data on the fiber ODF estimation project, and hence the network

generalization for a different anatomy. In future work, we aim to take advantage of artificial data generators such as those developed<sup>126</sup> for fetal T2-weighted images. In fact, they can be adapted to generate realistic fetal diffusion images that can be used as training data. We additionally aim at directly training the neural networks on forthcoming releases of high-quality fetal data, such as the dCHP fetal project<sup>37</sup> for high quality FOD prediction from a small number of diffusion measurements.

### 3.2.3 MOTION

Motion, as one of the major problems in non-cooperative cohorts such as fetuses and newborns was not the direct focus of our work. It was however addressed prior to our through-plane super-resolution model, where intra-volume and inter-volumes motion was corrected using registration tools<sup>153</sup>. The correction was then applied to the corresponding gradient directions. The autoencoder can however be used in the case of severe signal drops by recovering the corrupted slice using the neighboring clean ones.

Regarding the **fiber ODF estimation** work, motion is also an issue that was not taken into account in the current framework. In fact, motion correction will create "new" directions and hence we will be confronted with a sparse q-space grid. Our approach projects the diffusion samples to a spherical harmonics representation, and hence is less dependant on gradient directions and potential motion compensation.

Another potential extension of this work is the use of existing multiple volume super-resolution diffusion pipelines<sup>145,144,48</sup> or adapting recent state-of-the-art pipelines that were developed for fetal T2-weighted images<sup>235</sup>. In fact, since we aimed for a minimal acquisition scheme with few measurements, we have only used a single 4D volume at a time. However, in several centers, more than one and sometimes orthogonal volumes are available in fetal brains. Merging them using super-resolution techniques will certainly increase the SNR resulting in a better FOD estimation, which is

in line with the results of our phantom study.

Concerning the **crossing phantom study**, one way to make such findings more realistic in the context of fetal brains is to add - during acquisition - different levels of motion (translation, rotation) using motion simulators. Then we would be able to correct this motion and compute the diffusion properties for different configurations to be able to better quantify the acquisition trade-off between the number of gradient directions and orthogonal volumes in a more clinical setting. However, the interpretation of the results must to be limited in the context of the study. In order to generalize the findings, additional work with diverse protocols, i.e. more ground truth acquisitions and potentially with *in-vivo* animal data, need to be done to draw more generalizable conclusions, which can be translated to dMRI acquisition protocols of fetal imaging.

### 3.3 CONCLUSION

The outcome of methodological researches that aim at improving early brain imaging hold significant implications in understanding the developmental mechanisms and may pave the way for new research directions or improved diagnostic and clinical applications in pediatric neuroimaging. In fact, by reducing the burden of long scanning times through the implementation of deep neural networks, different centers can be encouraged to incorporate diffusion MRI protocols into their clinical routine. This, will in turn increase the size and diversity of early brain datasets in the community, which will further improve the prediction of models.

Anatomical information acquired from structural and diffusion MRI could be combined with functional MRI - which is also a growing field in fetal brain imaging<sup>98,222,207</sup> - to shed light on early brain anatomically constrained functional patterns. Such studies can be very valuable to answer fundamental questions related to fetal self-awareness. For example, neural correlates of consciousness<sup>31</sup> could be further explored in early development<sup>125</sup>. Furthermore, these studies provide an avenue to

address pertinent practical questions, including the determination of if and when a fetus can have a subjective experience, such as pain or joy<sup>208</sup>.



## References

- [1] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: a system for large-scale machine learning. In *Osdi*, volume 16 (pp. 265–283): Savannah, GA, USA.
- [2] Aganj, I. et al. (2010). Reconstruction of the orientation distribution function in single- and multiple-shell q-ball imaging within constant solid angle. *Magnetic resonance in medicine*, 64(2), 554–566.
- [3] Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., & Erickson, B. J. (2017). Deep learning for brain mri segmentation: state of the art and future directions. *Journal of digital imaging*, 30, 449–459.
- [4] Albay, E., Demir, U., & Unal, G. (2018). Diffusion mri spatial super-resolution using generative adversarial networks. In *PRedictive Intelligence in MEdicine: First International Workshop, PRIME 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 1* (pp. 155–163): Springer.

- [5] Alexander, D. C., Zikic, D., Zhang, J., Zhang, H., & Criminisi, A. (2014). Image quality transfer via random forest regression: applications in diffusion mri. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part III 17* (pp. 225–232).: Springer.
- [6] Allen-Zhu, Z., Li, Y., & Liang, Y. (2019). Learning and generalization in overparameterized neural networks, going beyond two layers. *Advances in neural information processing systems*, 32.
- [7] Angulakshmi, M. & Deepa, M. (2021). A review on deep learning architecture and methods for mri brain tumour segmentation. *Current Medical Imaging*, 17(6), 695–706.
- [8] Anwar, S. M., Irmakci, I., Torigian, D. A., Jambawalikar, S., Papadakis, G. Z., Akgun, C., Ellermann, J., Akcakaya, M., & Bagci, U. (2020). Semi-supervised deep learning for multi-tissue segmentation from multi-contrast mri. *Journal of Signal Processing Systems*, (pp. 1–14).
- [9] Arabahmadi, M., Farahbakhsh, R., & Rezazadeh, J. (2022). Deep learning for smart healthcare—a survey on brain tumor detection from medical imaging. *Sensors*, 22(5), 1960.
- [10] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58, 82–115.
- [11] Arroyo, M. S., Hopkin, R. J., Nagaraj, U. D., Kline-Fath, B., & Venkatesan, C. (2019). Fetal brain mri findings and neonatal outcome of common diagnosis at a tertiary care center. *Journal of Perinatology*, 39(8), 1072–1077.

- [12] Aubin, B., Krzakala, F., Lu, Y., & Zdeborová, L. (2020). Generalization error in high-dimensional perceptrons: Approaching bayes error with convex optimization. *Advances in Neural Information Processing Systems*, 33, 12199–12210.
- [13] Avram, A. V. (2011). *Diffusion Tensor Imaging of Myelin Water*. PhD thesis.
- [14] Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one*, 10(7), e0130140.
- [15] Barquero, G., La Rosa, F., Kebiri, H., Lu, P.-J., Rahmanzadeh, R., Weigel, M., Fartaria, M. J., Kober, T., Théaudin, M., Du Pasquier, R., et al. (2020). Rimnet: A deep 3d multimodal mri architecture for paramagnetic rim lesion assessment in multiple sclerosis. *NeuroImage: Clinical*, 28, 102412.
- [16] Bartlett, P. L. & Mendelson, S. (2002). Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov), 463–482.
- [17] Basser, P. J., Mattiello, J., & LeBihan, D. (1994). Estimation of the effective self-diffusion tensor from the nmr spin echo. *Journal of Magnetic Resonance, Series B*, 103(3), 247–254.
- [18] Baumann, N. & Pham-Dinh, D. (2001). Biology of oligodendrocyte and myelin in the mammalian central nervous system. *Physiological reviews*, 81(2), 871–927.
- [19] Bennett, S. H., Kirby, A. J., & Finnerty, G. T. (2018). Rewiring the connectome: evidence and effects. *Neuroscience & Biobehavioral Reviews*, 88, 51–62.
- [20] Besold, T. R., Garcez, A. d., Bader, S., Bowman, H., Domingos, P., Hitzler, P., Kühnberger, K.-U., Lamb, L. C., Lowd, D., Lima, P. M. V., et al. (2017). Neural-symbolic learning and reasoning: A survey and interpretation. *arXiv preprint arXiv:1711.03902*.

- [21] Bethlehem, R. A., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., et al. (2022). Brain charts for the human lifespan. *Nature*, 604(7906), 525–533.
- [22] Billot, B., Greve, D. N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A. V., Iglesias, J. E., et al. (2023). Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. *Medical Image Analysis*, 86, 102789.
- [23] Bloch, F. (1946). Nuclear induction. *Physical review*, 70(7-8), 460.
- [24] Bortolin, K., Delavari, F., Preti, M. G., Sandini, C., Mancini, V., Mullier, E., Van De Ville, D., & Eliez, S. (2022). Neural substrates of psychosis revealed by altered dependencies between brain activity and white-matter architecture in individuals with 22q11 deletion syndrome. *NeuroImage: Clinical*, 35, 103075.
- [25] Brunel, H., Girard, N., Confort-Gouny, S., Viola, A., Chaumoitre, K., D’ercole, C., Figarella-Branger, D., Raybaud, C., Cozzone, P., & Panuel, M. (2004). Fetal brain injury. *Journal of neuroradiology*, 31(2), 123–137.
- [26] Bui, T., Daire, J.-L., Chalard, F., Zaccaria, I., Alberti, C., Elmaleh, M., Garel, C., Luton, D., Blanc, N., & Sebag, G. (2006). Microstructural development of human brain assessed in utero by diffusion tensor imaging. *Pediatric radiology*, 36, 1133–1140.
- [27] Camchong, J., MacDonald III, A. W., Bell, C., Mueller, B. A., & Lim, K. O. (2011). Altered functional and anatomical connectivity in schizophrenia. *Schizophrenia bulletin*, 37(3), 640–650.
- [28] Canales-Rodríguez, E. J., Legarreta, J. H., Pizzolato, M., Rensonnet, G., Girard, G., Patino, J. R., et al. (2019). Sparse wars: A survey and comparative study of spherical deconvolution algorithms for diffusion MRI. *NeuroImage*, 184, 140–160.

- [29] Cao, X., Yang, J., Wang, L., Xue, Z., Wang, Q., & Shen, D. (2018). Deep learning based inter-modality image registration supervised by intra-modality similarity. In *Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 9* (pp. 55–63).: Springer.
- [30] Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of consciousness studies*, 2(3), 200–219.
- [31] Chalmers, D. J. (2000). What is a neural correlate of consciousness. *Neural correlates of consciousness: Empirical and conceptual questions*, (pp. 17–39).
- [32] Chatterjee, S., Sciarra, A., Dünnwald, M., Mushunuri, R. V., Podishetti, R., Rao, R. N., Gopinath, G. D., Oeltze-Jafra, S., Speck, O., & Nürnberger, A. (2021). Shuffleunet: Super resolution of diffusion-weighted mris using deep learning. In *2021 29th European Signal Processing Conference (EUSIPCO)* (pp. 940–944).: IEEE.
- [33] Chen, G., Hong, Y., Huynh, K. M., & Yap, P.-T. (2023). Deep learning prediction of diffusion mri data with microstructure-sensitive loss functions. *Medical Image Analysis*, (pp. 102742).
- [34] Chen, R., Sun, C., Liu, T., Liao, Y., Wang, J., Sun, Y., Zhang, Y., Wang, G., & Wu, D. (2022). Deciphering the developmental order and microstructural patterns of early white matter pathways in a diffusion mri based fetal brain atlas. *NeuroImage*, 264, 119700.
- [35] Chen, Y., Xie, Y., Zhou, Z., Shi, F., Christodoulou, A. G., & Li, D. (2018). Brain mri super resolution using 3d deep densely connected neural networks. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)* (pp. 739–742).: IEEE.

- [36] Chi, J. G., Dooling, E. C., & Gilles, F. H. (1977). Gyral development of the human brain. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 1(1), 86–93.
- [37] Christiaens, D., Cordero-Grande, L., Price, A., Hutter, J., Hughes, E., Counsell, S., Tournier, J., & Hajnal, J. (2019). Fetal diffusion mri acquisition and analysis in the developing human connectome project. In *Proceedings of the annual meeting of the International Society of Magnetic Resonance in Medicine (ISMRM)*.
- [38] Conturo, T. E., McKinstry, R. C., Aronovitz, J. A., & Neil, J. J. (1995). Diffusion mri: precision, accuracy and flow effects. *NMR in Biomedicine*, 8(7), 307–332.
- [39] de Dumast, P., Kebiri, H., Atat, C., Dunet, V., Koob, M., & Cuadra, M. B. (2021). Segmentation of the cortical plate in fetal brain mri with a topological loss. In *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis: 3rd International Workshop, UNSURE 2021, and 6th International Workshop, PIPPI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 3* (pp. 200–209).: Springer.
- [40] de Dumast, P., Kebiri, H., Dunet, V., Koob, M., & Cuadra, M. B. (2022a). Multi-dimensional topological loss for cortical plate segmentation in fetal brain mri. *arXiv preprint arXiv:2208.07566*.
- [41] de Dumast, P., Kebiri, H., Payette, K., Jakab, A., Lajous, H., & Cuadra, M. B. (2022b). Synthetic magnetic resonance images for domain adaptation: Application to fetal brain tissue segmentation. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)* (pp. 1–5).: IEEE.

- [42] de Vareilles, H., Rivière, D., Mangin, J., & Dubois, J. (2023). Development of cortical folds in the human brain: an attempt to review biological hypotheses, early neuroimaging investigations and functional correlates. *Developmental Cognitive Neuroscience*, (pp. 101249).
- [43] De Vos, B. D., Berendsen, F. F., Viergever, M. A., Staring, M., & Išgum, I. (2017). End-to-end unsupervised deformable image registration with a convolutional neural network. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3* (pp. 204–212).: Springer.
- [44] Delavari, F., Sandini, C., Zöllner, D., Mancini, V., Bortolin, K., Schneider, M., Van De Ville, D., & Eliez, S. (2021). Dysmaturation observed as altered hippocampal functional connectivity at rest is associated with the emergence of positive psychotic symptoms in patients with 22q11 deletion syndrome. *Biological psychiatry*, 90(1), 58–68.
- [45] DelEtoile, J. & Adeli, H. (2017). Graph theory and brain connectivity in alzheimer’s disease. *The Neuroscientist*, 23(6), 616–626.
- [46] Dell’Acqua, F. & Tournier, J.-D. (2019). Modelling white matter with spherical deconvolution: How and why? *NMR in Biomedicine*, 32(4), e3945.
- [47] Denk, W. & Horstmann, H. (2004). Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS biology*, 2(11), e329.
- [48] Deprez, M., Price, A., Christiaens, D., Estrin, G. L., Cordero-Grande, L., Hutter, J., Dadducci, A., Tournier, J.-D., Rutherford, M., Counsell, S. J., et al. (2019). Higher order spherical harmonics reconstruction of fetal diffusion mri with intensity correction. *IEEE transactions on medical imaging*, 39(4), 1104–1113.

- [49] Descoteaux, M. (2008). *High angular resolution diffusion MRI: from local estimation to segmentation and tractography*. PhD thesis, Université Nice Sophia Antipolis.
- [50] Descoteaux, M. & Deriche, R. (2015). From local q-ball estimation to fibre crossing tractography. In *Handbook of Biomedical Imaging* (pp. 455–473). Springer.
- [51] Dhifallah, O. & Lu, Y. M. (2021). Phase transitions in transfer learning for high-dimensional perceptrons. *Entropy*, 23(4), 400.
- [52] Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2), 295–307.
- [53] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- [54] Dubois, J., Dehaene-Lambertz, G., Kulikova, S., Poupon, C., Hüppi, P. S., & Hertz-Pannier, L. (2014). The early development of brain white matter: a review of imaging studies in fetuses, newborns and infants. *neuroscience*, 276, 48–71.
- [55] Dudink, J., Lequin, M., van Pul, C., Buijs, J., Conneman, N., van Goudoever, J., & Govaert, P. (2007). Fractional anisotropy in white matter tracts of very-low-birth-weight infants. *Pediatric radiology*, 37, 1216–1223.
- [56] Duffy, B. A., Zhang, W., Tang, H., Zhao, L., Law, M., Toga, A. W., & Kim, H. (2018). Retrospective correction of motion artifact affected structural mri images using deep learning of simulated motion. In *Medical Imaging with Deep Learning*.

- [57] d’Ascoli, S., Touvron, H., Leavitt, M. L., Morcos, A. S., Biroli, G., & Sagun, L. (2021). Convit: Improving vision transformers with soft convolutional inductive biases. In *International Conference on Machine Learning* (pp. 2286–2296).: PMLR.
- [58] Edelman, G. M. & Tononi, G. (2008). *A universe of consciousness: How matter becomes imagination*. Hachette UK.
- [59] Elaldi, A., Gerig, G., & Dey, N. (2023).  $e(3)\times so(3)$ -equivariant networks for spherical deconvolution in diffusion mri. *arXiv preprint arXiv : 2304.06103*.
- [60] Elsaid, N. M. & Wu, Y.-C. (2019). Super-resolution diffusion tensor imaging using srcnn: a feasibility study. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 2830–2834).: IEEE.
- [61] Emblem, K. E., Due-Tonnessen, P., Hald, J. K., Bjornerud, A., Pinho, M. C., Scheie, D., Schad, L. R., Meling, T. R., & Zoellner, F. G. (2014). Machine learning in preoperative glioma mri: survival associations by perfusion-based support vector machine outperforms traditional mri. *Journal of magnetic resonance imaging*, 40(1), 47–54.
- [62] Fadnavis, S., Batson, J., & Garyfallidis, E. (2020). Patch2self: Denoising diffusion mri with self-supervised learning. *arXiv preprint arXiv:2011.01355*.
- [63] Feis, D.-L., Brodersen, K. H., von Cramon, D. Y., Luders, E., & Tittgemeyer, M. (2013). Decoding gender dimorphism of the human brain using multimodal anatomical and diffusion mri data. *Neuroimage*, 70, 250–257.
- [64] Feng, K., Rowell, A. C., Andres, A., Bellando, B. J., Lou, X., Glasier, C. M., Ramakrishnaiah, R. H., Badger, T. M., & Ou, X. (2019). Diffusion tensor mri of white matter of healthy full-term newborns: relationship to neurodevelopmental outcomes. *Radiology*, 292(1), 179–187.

- [65] Fidon, L., Aertsen, M., Kofler, F., Bink, A., David, A. L., Deprest, T., Emam, D., Guffens, F., Jakab, A., Kasprian, G., et al. (2022). A dempster-shafer approach to trustworthy ai with application to fetal brain mri segmentation. *arXiv preprint arXiv:2204.02779*.
- [66] Finger, S. (2001). *Origins of neuroscience: a history of explorations into brain function*. Oxford University Press, USA.
- [67] Fogtmann, M., Seshamani, S., Kroenke, C., Cheng, X., Chapman, T., Wilm, J., Rousseau, F., & Studholme, C. (2013). A unified approach to diffusion direction sensitive slice registration and 3-d dti reconstruction from moving fetal brain anatomy. *IEEE transactions on medical imaging*, 33(2), 272–289.
- [68] Frank, L. R. (2002). Characterization of anisotropy in high angular resolution diffusion-weighted mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 47(6), 1083–1099.
- [69] Garel, C. & Garel, C. (2004). *MRI of the Fetal Brain*. Springer.
- [70] Gholipour, A., Limperopoulos, C., Clancy, S., Clouchoux, C., Akhondi-Asl, A., Estroff, J. A., & Warfield, S. K. (2014). Construction of a deformable spatiotemporal mri atlas of the fetal brain: evaluation of similarity metrics and deformation models. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part II 17* (pp. 292–299): Springer.
- [71] Gholipour, A., Rollins, C. K., Velasco-Annis, C., Ouaalam, A., Akhondi-Asl, A., Afacan, O., Ortinau, C. M., Clancy, S., Limperopoulos, C., Yang, E., et al. (2017). A normative spatiotemporal mri atlas of the fetal brain for automatic segmentation and analysis of early brain growth. *Scientific reports*, 7(1), 476.

- [72] Golkov, V. et al. (2016). Q-space deep learning: twelve-fold shorter and model-free diffusion mri scans. *IEEE TMI*, 35(5), 1344–1351.
- [73] Goyal, A. & Bengio, Y. (2022). Inductive biases for deep learning of higher-level cognition. *Proceedings of the Royal Society A*, 478(2266), 20210068.
- [74] Graham, M. S., Drobnjak, I., & Zhang, H. (2018). A supervised learning approach for diffusion mri quality control with minimal training data. *NeuroImage*, 178, 668–676.
- [75] Griffa, A., Baumann, P. S., Ferrari, C., Do, K. Q., Conus, P., Thiran, J.-P., & Hagmann, P. (2015). Characterizing the connectome in schizophrenia with diffusion spectrum imaging. *Human brain mapping*, 36(1), 354–366.
- [76] Griffiths, P., Reeves, M., Morris, J., Mason, G., Russell, S., Paley, M., & Whitby, E. (2010). A prospective study of fetuses with isolated ventriculomegaly investigated by antenatal sonography and in utero mr imaging. *American journal of neuroradiology*, 31(1), 106–111.
- [77] Griffiths, P. D., Bradburn, M., Campbell, M. J., Cooper, C. L., Embleton, N., Graham, R., Hart, A. R., Jarvis, D., Kilby, M. D., Lie, M., et al. (2019). Mri in the diagnosis of fetal developmental brain abnormalities: the meridian diagnostic accuracy study. *Health Technology Assessment (Winchester, England)*, 23(49), 1–144.
- [78] Grøvik, E., Yi, D., Iv, M., Tong, E., Rubin, D., & Zaharchuk, G. (2020). Deep learning enables automatic detection and segmentation of brain metastases on multisequence mri. *Journal of Magnetic Resonance Imaging*, 51(1), 175–182.
- [79] Guimiot, F., Garel, C., Fallet-Bianco, C., Menez, F., Khung-Savatovsky, S., Oury, J.-F., Sebag, G., & Delezoide, A.-L. (2008). Contribution of diffusion-weighted imaging in the evaluation of diffuse white matter ischemic lesions in fetuses: correlations with fetopathologic findings. *American journal of neuroradiology*, 29(1), 110–115.

- [80] Hagmann, P., Jonasson, L., Maeder, P., Thiran, J.-P., Wedeen, V. J., & Meuli, R. (2006). Understanding diffusion mr imaging techniques: from scalar diffusion-weighted imaging to diffusion tensor imaging and beyond. *Radiographics*, 26(suppl\_1), S205–S223.
- [81] Hamill, O. P., Marty, A., Neher, E., Sakmann, B., & Sigworth, F. J. (1981). Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Archiv*, 391, 85–100.
- [82] Hartigan, J. A. & Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1), 100–108.
- [83] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
- [84] Hertz, J., Krogh, A., Palmer, R. G., & Horner, H. (1991). Introduction to the theory of neural computation. *Physics Today*, 44(12), 70.
- [85] Hill, J., Inder, T., Neil, J., Dierker, D., Harwell, J., & Van Essen, D. (2010). Similar patterns of cortical expansion during human development and evolution. *Proceedings of the National Academy of Sciences*, 107(29), 13135–13140.
- [86] Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1 (pp. 278–282): IEEE.
- [87] Hoffmann, C., Weisz, B., Lipitz, S., Yaniv, G., Katorza, E., Bergman, D., & Biegon, A. (2014). Regional apparent diffusion coefficient values in 3rd trimester fetal brain. *Neuroradiology*, 56, 561–567.
- [88] Hooker, J. D., Khan, M. A., Farkas, A. B., Lirette, S. T., Joyner, D. A., Gordy, D. P., Storrs, J. M., Roda, M. S., Bofill, J. A., Smith, A. D., et al. (2020). Third-trimester in utero fetal brain diffusion

tensor imaging fiber tractography: a prospective longitudinal characterization of normal white matter tract development. *Pediatric radiology*, 50, 973–983.

- [89] Hossbach, J., Splitthoff, D. N., Cauley, S., Clifford, B., Polak, D., Lo, W.-C., Meyer, H., & Maier, A. (2022). Deep learning-based motion quantification from k-space for fast model-based mri motion correction. *Medical Physics*.
- [90] Hosseini, S. et al. (2022). Cttrack: A cnn+ transformer-based framework for fiber orientation estimation & tractography. *Neuroscience Informatics*, 2(4), 100099.
- [91] Howell, B. R., Styner, M. A., Gao, W., Yap, P.-T., Wang, L., Baluyot, K., Yacoub, E., Chen, G., Potts, T., Salzwedel, A., et al. (2019). The unc/umn baby connectome project (bcp): An overview of the study design and protocol development. *NeuroImage*, 185, 891–905.
- [92] Hüppi, P. S., Maier, S. E., Peled, S., Zientara, G. P., Barnes, P. D., Jolesz, F. A., & Volpe, J. J. (1998). Microstructural development of human newborn cerebral white matter assessed in vivo by diffusion tensor magnetic resonance imaging. *Pediatric research*, 44(4), 584–590.
- [93] Hutter, J. et al. (2018). Time-efficient and flexible design of optimized multishell hardi diffusion. *Magnetic resonance in medicine*, 79(3), 1276–1292.
- [94] Ismael, S. A. A., Mohammed, A., & Hefny, H. (2020). An enhanced deep learning approach for brain cancer mri images classification using residual networks. *Artificial intelligence in medicine*, 102, 101779.
- [95] Jaimes, C., Machado-Rivas, F., Afacan, O., Khan, S., Marami, B., Ortinau, C. M., Rollins, C. K., Velasco-Annis, C., Warfield, S. K., & Gholipour, A. (2020). In vivo characterization of emerging white matter microstructure in the fetal brain in the third trimester. *Human Brain Mapping*, 41(12), 3177–3185.

- [96] Jakab, A., Kasprian, G., Schwartz, E., Gruber, G. M., Mitter, C., Prayer, D., Schöpf, V., & Langs, G. (2015). Disrupted developmental organization of the structural connectome in fetuses with corpus callosum agenesis. *Neuroimage*, 111, 277–288.
- [97] Jakab, A., Tuura, R., Kellenberger, C., & Scheer, I. (2017). In utero diffusion tensor imaging of the fetal brain: a reproducibility study. *NeuroImage: Clinical*, 15.
- [98] Jardri, R., Pins, D., Houfflin-Debarge, V., Chaffiotte, C., Rocourt, N., Pruvo, J.-P., Steinling, M., Delion, P., & Thomas, P. (2008). Fetal cortical activation to sound at 33 weeks of gestation: a functional mri study. *Neuroimage*, 42(1), 10–18.
- [99] Jelescu, I. O., Veraart, J., Adisetiyo, V., Milla, S. S., Novikov, D. S., & Fieremans, E. (2015). One diffusion acquisition and different white matter models: how does microstructure change in human early development based on wmti and noddif? *Neuroimage*, 107, 242–256.
- [100] Jeurissen, B. et al. (2014). Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion mri data. *Neuroimage*, 103, 411–426.
- [101] Jeurissen, B., Leemans, A., Tournier, J.-D., Jones, D. K., & Sijbers, J. (2013). Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Human brain mapping*, 34(11), 2747–2766.
- [102] Ji, X., Elmoznino, E., Deane, G., Constant, A., Dumas, G., Lajoie, G., Simon, J., & Bengio, Y. (2023). Sources of richness and ineffability for phenomenally conscious states. *arXiv preprint arXiv:2302.06403*.
- [103] Judaš, M., Radoš, M., Jovanov-Milošević, N., Hrabac, P., Kostović, I., et al. (2005). Structural, immunocytochemical, and mr imaging properties of periventricular crossroads of growing cortical pathways in preterm infants. *American journal of neuroradiology*, 26(10), 2671–2684.

- [104] Karimi, D. et al. (2021a). Deep learning-based parameter estimation in fetal diffusion-weighted mri. *Neuroimage*, 243, 118482.
- [105] Karimi, D. et al. (2021b). Learning to estimate the fiber orientation distribution function from diffusion-weighted mri. *Neuroimage*, 239, 118316.
- [106] Karimi, D. & Gholipour, A. (2022). Atlas-powered deep learning (adl)-application to diffusion weighted mri. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part I* (pp. 123–132).: Springer.
- [107] Karmacharya, S., Gagoski, B., Ning, L., Vyas, R., Cheng, H., Soul, J., Newberger, J., Shenton, M., Rathi, Y., & Grant, P. (2018). Advanced diffusion imaging for assessing normal white matter development in neonates and characterizing aberrant development in congenital heart disease. *NeuroImage: Clinical*, 19, 360–373.
- [108] Kasprian, G., Brugger, P. C., Schöpf, V., Mitter, C., Weber, M., Hainfellner, J. A., & Prayer, D. (2013). Assessing prenatal white matter connectivity in commissural agenesis. *Brain*, 136(1), 168–179.
- [109] Kasprian, G., Brugger, P. C., Weber, M., Krssák, M., Krampfl, E., Herold, C., & Prayer, D. (2008). In utero tractography of fetal white matter development. *Neuroimage*, 43(2), 213–224.
- [110] Kebiri, H., Canales-Rodríguez, E. J., Lajous, H., de Dumast, P., Girard, G., Alemán-Gómez, Y., Koob, M., Jakab, A., & Bach Cuadra, M. (2022a). Through-plane super-resolution with autoencoders in diffusion magnetic resonance imaging of the developing human brain. *Frontiers in Neurology*, 13, 765.
- [111] Kebiri, H. et al. (2022b). Slice estimation in diffusion mri of neonatal and fetal brains in image and spherical harmonics domains using autoencoders. *arXiv preprint arXiv:2208.13328*.

- [112] Kebiri, H., Gholipour, A., Cuadra, M. B., & Karimi, D. (2023a). Direct segmentation of brain white matter tracts in diffusion mri. *arXiv preprint arXiv:2307.02223*.
- [113] Kebiri, H., Gholipour, A., Karimi\*, D., & Bach Cuadra\*, M. (2023b). Robust estimation of the microstructure of the early developing brain using deep learning. *International conference on medical image computing and computer-assisted intervention (MICCAI) - under review*.
- [114] Kebiri, H., Gholipour, A., Vasung, L., Krsnik, Ž., Karimi, D., & Cuadra, M. B. (2023c). Deep learning microstructure estimation of developing brains from diffusion mri: a newborn and fetal study. *bioRxiv*.
- [115] Kebiri, H., Lajous, H., Alemán-Gómez, Y., Girard, G., Rodríguez, E. C., Tourbier, S., Pizzolato, M., Ledoux, J.-B., Fornari, E., Jakab, A., et al. (2021a). Quantitative evaluation of enhanced multi-plane clinical fetal diffusion mri with a crossing-fiber phantom. In *Computational Diffusion MRI: 12th International Workshop, CDMRI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings* (pp. 12–22).: Springer.
- [116] Kebiri, H., Lajous, H., Aleman-Gomez, Y., Ledoux, J., Fornari, E., & Jakab, A., B. C. M. (2021b). Dataset: Quantitative evaluation of enhanced multi-plane clinical fetal diffusion mri with a crossing-fiber phantom. *Zenodo*.
- [117] Khan, S., Vasung, L., Marami, B., Rollins, C. K., Afacan, O., Ortinau, C. M., Yang, E., Warfield, S. K., & Gholipour, A. (2019). Fetal brain growth portrayed by a spatiotemporal diffusion tensor mri atlas computed from in utero images. *NeuroImage*, 185, 593–608.
- [118] Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M., & Biller, A. (2016). Deep mri brain extraction: A 3d convolutional neural network for skull stripping. *NeuroImage*, 129, 460–469.

- [119] Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience*, 17(5), 307–321.
- [120] Koppers, S., Haarbuerger, C., & Merhof, D. (2017). Diffusion mri signal augmentation: from single shell to multi shell with deep learning. In *Computational Diffusion MRI: MICCAI Workshop, Athens, Greece, October 2016 19* (pp. 61–70).: Springer.
- [121] Koppers, S. & Merhof, D. (2016). Direct estimation of fiber orientations using deep learning in diffusion imaging. In *International Workshop on Machine Learning in Medical Imaging* (pp. 53–60).: Springer.
- [122] Kunz, N., Zhang, H., Vasung, L., O’Brien, K. R., Assaf, Y., Lazeyras, F., Alexander, D. C., & Hüppi, P. S. (2014). Assessing white matter microstructure of the newborn with multi-shell diffusion mri and biophysical compartment models. *Neuroimage*, 96, 288–299.
- [123] La Rosa, F., Abdulkadir, A., Fartaria, M. J., Rahmzadeh, R., Lu, P.-J., Galbusera, R., Barakovic, M., Thiran, J.-P., Granziera, C., & Cuadra, M. B. (2020). Multiple sclerosis cortical and wm lesion segmentation at 3t mri: a deep learning method based on flair and mp2rage. *NeuroImage: Clinical*, 27, 102335.
- [124] Lächler, K., Lajous, H., Unser, M., Cuadra, M. B., & Pla, P. d. A. (2022). Self-supervised isotropic superresolution fetal brain mri. *arXiv preprint arXiv:2211.06502*.
- [125] Lagercrantz, H. & Changeux, J.-P. (2009). The emergence of human consciousness: from fetal to neonatal life. *Pediatric research*, 65(3), 255–260.
- [126] Lajous, H., Roy, C. W., Hilbert, T., de Dumast, P., Tourbier, S., Alemán-Gómez, Y., Yerly, J., Yu, T., Kebiri, H., Payette, K., et al. (2021). Fabian: A fetal brain magnetic resonance acquisition numerical phantom. *arXiv preprint arXiv:2109.03624*.

- [127] Launay, S., Robert, Y., Valat, A., Thomas, D., Devisme, L., Rocourt, N., & Vaast, P. (2002). Cerebral fetal mri and ventriculomegaly. *Journal de radiologie*, 83(6 Pt 1), 723–730.
- [128] Lauterbur, P. C. (1973). Image formation by induced local interactions: examples employing nuclear magnetic resonance. *nature*, 242(5394), 190–191.
- [129] Le Bihan, D. & Breton, E. (1985). Imagerie de diffusion in vivo par résonance magnétique nucléaire. *Comptes rendus de l'Académie des sciences. Série 2, Mécanique, Physique, Chimie, Sciences de l'univers, Sciences de la Terre*, 301(15), 1109–1112.
- [130] Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., & Laval-Jeantet, M. (1986). Mr imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2), 401–407.
- [131] Le Bihan, D., Mangin, J.-F., Poupon, C., Clark, C. A., Pappata, S., Molko, N., & Chabriat, H. (2001). Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 13(4), 534–546.
- [132] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4), 541–551.
- [133] Lefèvre, J., Germanaud, D., Dubois, J., Rousseau, F., de Macedo Santos, I., Angleys, H., Mangin, J.-F., Hüppi, P. S., Girard, N., & De Guio, F. (2015). Are developmental trajectories of cortical folding comparable between cross-sectional datasets of fetuses and preterm newborns? *Cerebral cortex*, 26(7), 3023–3035.
- [134] Li, H. et al. (2021). Superdti: Ultrafast dti and fiber tractography with deep learning. *Magnetic resonance in medicine*, 86(6), 3334–3347.

- [135] Lin, Z. et al. (2019). Fast learning of fiber orientation distribution function for mr tractography using convolutional neural network. *Medical physics*, 46(7), 3101–3116.
- [136] Lynall, M.-E., Bassett, D. S., Kerwin, R., McKenna, P. J., Kitzbichler, M., Muller, U., & Bullmore, E. (2010). Functional connectivity and brain networks in schizophrenia. *Journal of Neuroscience*, 30(28), 9477–9487.
- [137] Lyon, M., Armitage, P., & Álvarez, M. A. (2022). Angular super-resolution in diffusion mri with a 3d recurrent convolutional autoencoder. In *International Conference on Medical Imaging with Deep Learning* (pp. 834–846): PMLR.
- [138] Maas, L. C., Mukherjee, P., Carballido-Gamio, J., Veeraraghavan, S., Miller, S. P., Partridge, S. C., Henry, R. G., Barkovich, A. J., & Vigneron, D. B. (2004). Early laminar organization of the human cerebrum demonstrated with diffusion tensor imaging in extremely premature infants. *Neuroimage*, 22(3), 1134–1140.
- [139] Machado-Rivas, F., Afacan, O., Khan, S., Marami, B., Rollins, C., Ortinau, C., Velasco-Annis, C., Warfield, S., Gholipour, A., & Jaimes, C. (2021a). Tractography of the cerebellar peduncles in second-and third-trimester fetuses. *American Journal of Neuroradiology*, 42(1), 194–200.
- [140] Machado-Rivas, F., Afacan, O., Khan, S., Marami, B., Velasco-Annis, C., Lidov, H., Warfield, S., Gholipour, A., & Jaimes, C. (2021b). Spatiotemporal changes in diffusivity and anisotropy in fetal brain tractography.
- [141] Mader, I., Schöning, M., Klose, U., & Küker, W. (2002). Neonatal cerebral infarction diagnosed by diffusion-weighted mri: pseudonormalization occurs early. *Stroke*, 33(4), 1142–1145.
- [142] Maleki, M., Teshnehlab, M., & Nabavi, M. (2012). Diagnosis of multiple sclerosis (ms) using convolutional neural network (cnn) from mris. *Global Journal of Medicinal Plant Research*, 1(1), 50–54.

- [143] Mansfield, P. (1977). Multi-planar image formation using nmr spin echoes. *Journal of Physics C: Solid State Physics*, 10(3), L55.
- [144] Marami, B., Salehi, S. S. M., Afacan, O., Scherrer, B., Rollins, C. K., Yang, E., Estroff, J. A., Warfield, S. K., & Gholipour, A. (2017). Temporal slice registration and robust diffusion-tensor reconstruction for improved fetal brain structural connectivity analysis. *NeuroImage*, 156, 475–488.
- [145] Marami, B., Scherrer, B., Afacan, O., Erem, B., Warfield, S. K., & Gholipour, A. (2016). Motion-robust diffusion-weighted brain mri reconstruction through slice-level registration-based motion tracking. *IEEE transactions on medical imaging*, 35(10), 2258–2269.
- [146] Masters, C. L., Simms, G., Weinman, N. A., Multhaup, G., McDonald, B. L., & Beyreuther, K. (1985). Amyloid plaque core protein in alzheimer disease and down syndrome. *Proceedings of the National Academy of Sciences*, 82(12), 4245–4249.
- [147] McCulloch, W. S. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115–133.
- [148] McDonald, C., Bullmore, E., Sham, P., Chitnis, X., Suckling, J., MacCabe, J., Walshe, M., & Murray, R. M. (2005). Regional volume deviations of brain structure in schizophrenia and psychotic bipolar disorder: computational morphometry study. *The British Journal of Psychiatry*, 186(5), 369–377.
- [149] McWhinney, S. R., Tremblay, A., Chevalier, T. M., Lim, V. K., & Newman, A. J. (2016). Using cforest to analyze diffusion tensor imaging data: A study of white matter integrity in healthy aging. *Brain connectivity*, 6(10), 747–758.
- [150] Mesaros, S., Rocca, M., Kacar, K., Kostic, J., Copetti, M., Stosic-Opincal, T., Preziosa, P., Sala, S., Riccitelli, G., Horsfield, M., et al. (2012). Diffusion tensor mri tractography and cognitive impairment in multiple sclerosis. *Neurology*, 78(13), 969–975.

- [151] Mirzaalian, H., Ning, L., Savadjiev, P., Pasternak, O., Bouix, S., Michailovich, O., Grant, G., Marx, C. E., Morey, R. A., Flashman, L. A., et al. (2016). Inter-site and inter-scanner diffusion mri data harmonization. *NeuroImage*, 135, 311–323.
- [152] Mitra, J., Bourgeat, P., Fripp, J., Ghose, S., Rose, S., Salvado, O., Connelly, A., Campbell, B., Palmer, S., Sharma, G., et al. (2014). Lesion segmentation from multimodal mri using random forest following ischemic stroke. *NeuroImage*, 98, 324–335.
- [153] Modat, M., Cash, D. M., Daga, P., Winston, G. P., Duncan, J. S., & Ourselin, S. (2014). Global image registration using a symmetric block-matching approach. *Journal of medical imaging*, 1(2), 024003–024003.
- [154] Moeskops, P., Viergever, M. A., Mendrik, A. M., De Vries, L. S., Benders, M. J., & Išgum, I. (2016). Automatic segmentation of mr brain images with a convolutional neural network. *IEEE transactions on medical imaging*, 35(5), 1252–1261.
- [155] Montavon, G., Lapuschkin, S., Binder, A., Samek, W., & Müller, K.-R. (2017). Explaining nonlinear classification decisions with deep taylor decomposition. *Pattern recognition*, 65, 211–222.
- [156] Moruzzi, G. & Magoun, H. W. (1949). Brain stem reticular formation and activation of the eeg. *Electroencephalography and clinical neurophysiology*, 1(1-4), 455–473.
- [157] Moseley, M., Cohen, Y., Mintorovitch, J., Kucharczyk, J., Tsuruda, J., Weinstein, P., & Norman, D. (1990). Evidence of anisotropic self-diffusion. *Radiology*, 176, 439–445.
- [158] Motta, A., Berning, M., Boergens, K. M., Staffler, B., Beining, M., Loomba, S., Hennig, P., Wissler, H., & Helmstaedter, M. (2019a). Dense connectomic reconstruction in layer 4 of the somatosensory cortex. *Science*, 366(6469), eaay3134.

- [159] Motta, A., Schurr, M., Staffler, B., & Helmstaedter, M. (2019b). Big data in nanoscale connectomics, and the greed for training labels. *Current opinion in neurobiology*, 55, 180–187.
- [160] Moyer, D., Ver Steeg, G., Tax, C. M., & Thompson, P. M. (2020). Scanner invariant representations for diffusion mri harmonization. *Magnetic resonance in medicine*, 84(4), 2174–2189.
- [161] Müller, P., Golkov, V., Tomassini, V., & Cremers, D. (2021). Rotation-equivariant deep learning for diffusion mri. *arXiv preprint arXiv:2102.06942*.
- [162] Nath, V. et al. (2019). Deep learning reveals untapped information for local white-matter fiber reconstruction in diffusion-weighted mri. *Magnetic resonance imaging*, 62, 220–227.
- [163] Neher, P. F., Côté, M.-A., Houde, J.-C., Descoteaux, M., & Maier-Hein, K. H. (2017). Fiber tractography using machine learning. *Neuroimage*, 158, 417–429.
- [164] Neil, J. J. & Smyser, C. D. (2021). Diffusion magnetic resonance imaging of infants. *Magnetic Resonance Imaging Clinics*, 29(2), 185–193.
- [165] Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 427–436).
- [166] Partridge, S. C., Mukherjee, P., Henry, R. G., Miller, S. P., Berman, J. I., Jin, H., Lu, Y., Glenn, O. A., Ferriero, D. M., Barkovich, A. J., et al. (2004). Diffusion tensor imaging: serial quantitation of white matter tract maturity in premature newborns. *Neuroimage*, 22(3), 1302–1314.
- [167] Parvizi, J. & Damasio, A. R. (2003). Neuroanatomical correlates of brainstem coma. *Brain*, 126(7), 1524–1536.
- [168] Patel, K. et al. (2018). Better fiber odfs from suboptimal data with autoencoder based regularization. In *MICCAI* (pp. 55–62).: Springer.

- [169] Paus, T., Collins, D., Evans, A., Leonard, G., Pike, B., & Zijdenbos, A. (2001). Maturation of white matter in the human brain: a review of magnetic resonance studies. *Brain research bulletin*, 54(3), 255–266.
- [170] Payette, K., de Dumast, P., Kebiri, H., Ezhov, I., Paetzold, J. C., Shit, S., Iqbal, A., Khan, R., Kottke, R., Grehten, P., et al. (2021). An automatic multi-tissue human fetal brain segmentation benchmark using the fetal tissue annotation dataset. *Scientific Data*, 8(1), 167.
- [171] Payette, K., Li, H., de Dumast, P., Licandro, R., Ji, H., Siddiquee, M. M. R., Xu, D., Myronenko, A., Liu, H., Pei, Y., et al. (2023). Fetal brain tissue annotation and segmentation challenge results. *Medical Image Analysis*, (pp. 102833).
- [172] Perry, E. K. & Perry, R. H. (2004). Neurochemistry of consciousness: cholinergic pathologies in the human brain. *Progress in Brain Research*, 145, 287–299.
- [173] Pham, C.-H., Tor-Díez, C., Meunier, H., Bednarek, N., Fablet, R., Passat, N., & Rousseau, F. (2019). Multiscale brain mri super-resolution using deep 3d convolutional networks. *Computerized Medical Imaging and Graphics*, 77, 101647.
- [174] Pinto, M. S., Paoletta, R., Billiet, T., Van Dyck, P., Guns, P.-J., Jeurissen, B., Ribbens, A., den Dekker, A. J., & Sijbers, J. (2020). Harmonization of brain diffusion mri: Concepts and methods. *Frontiers in Neuroscience*, 14, 396.
- [175] Poduslo, S. E. & Jang, Y. (1984). Myelin development in infant brain. *Neurochemical research*, 9, 1615–1626.
- [176] Pontabry, J., Rousseau, F., Oubel, E., Studholme, C., Koob, M., & Dietemann, J.-L. (2013). Probabilistic tractography using q-ball imaging and particle filtering: application to adult and in-utero fetal brain studies. *Medical image analysis*, 17(3), 297–310.

- [177] Purcell, E. M., Torrey, H. C., & Pound, R. V. (1946). Resonance absorption by nuclear magnetic moments in a solid. *Physical review*, 69(1-2), 37.
- [178] Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al. (2018). Improving language understanding by generative pre-training.
- [179] Radoš, M., Judaš, M., & Kostović, I. (2006). In vitro mri of brain development. *European journal of radiology*, 57(2), 187–198.
- [180] Raffelt, D. et al. (2012). Apparent fibre density: a novel measure for the analysis of diffusion-weighted magnetic resonance images. *Neuroimage*, 59(4), 3976–3994.
- [181] Ravi, D., Ghavami, N., Alexander, D. C., & Ianus, A. (2019). Current applications and future promises of machine learning in diffusion mri. In *Computational Diffusion MRI: International MICCAI Workshop, Granada, Spain, September 2018 22* (pp. 105–121).: Springer.
- [182] Rees, G., Kreiman, G., & Koch, C. (2002). Neural correlates of consciousness in humans. *Nature Reviews Neuroscience*, 3(4), 261–270.
- [183] Righini, A., Bianchini, E., Parazzini, C., Gementi, P., Ramenghi, L., Baldoli, C., Nicolini, U., Mosca, F., & Triulzi, F. (2003). Apparent diffusion coefficient determination in normal fetal brain: a prenatal mr imaging study. *American journal of neuroradiology*, 24(5), 799–804.
- [184] Rokem, A. et al. (2015). Evaluating the accuracy of diffusion mri models in white matter. *PloS one*, 10(4), e0123272.
- [185] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234–241).: Springer.

- [186] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088), 533–536.
- [187] Saha, S., Pagnozzi, A., Bourgeat, P., George, J. M., Bradford, D., Colditz, P. B., Boyd, R. N., Rose, S. E., Fripp, J., & Pannek, K. (2020). Predicting motor outcome in preterm infants from very early brain diffusion mri using a deep learning convolutional neural network (cnn) model. *Neuroimage*, 215, 116807.
- [188] Salehi, S. S. M., Hashemi, S. R., Velasco-Annis, C., Oualam, A., Estroff, J. A., Erdogmus, D., Warfield, S. K., & Gholipour, A. (2018). Real-time automatic fetal brain extraction in fetal mri by deep learning. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 720–724): IEEE.
- [189] Salomon, L., Ouahba, J., Delezoide, A.-L., Vuillard, E., Oury, J.-F., Sebag, G., & Garel, C. (2006). Third-trimester fetal mri in isolated 10-to 12-mm ventriculomegaly: is it worth it? *BJOG: An International Journal of Obstetrics & Gynaecology*, 113(8), 942–947.
- [190] Schneider, J., Confort-Gouny, S., Le Fur, Y., Viout, P., Bennathan, M., Chapon, F., Fogliarini, C., Cozzone, P., & Girard, N. (2007). Diffusion-weighted imaging in normal fetal brain maturation. *European radiology*, 17, 2422–2429.
- [191] Schneider, M. M., Berman, J., Baumer, F., Glass, H., Jeng, S., Jeremy, R., Esch, M., Biran, V., Barkovich, A., Studholme, C., et al. (2009). Normative apparent diffusion coefficient values in the developing fetal brain. *American Journal of Neuroradiology*, 30(9), 1799–1803.
- [192] Schnell, S., Saur, D., Kreher, B., Hennig, J., Burkhardt, H., & Kiselev, V. G. (2009). Fully automated classification of hardi in vivo data using a support vector machine. *NeuroImage*, 46(3), 642–651.

- [193] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618–626).
- [194] Shapson-Coe, A., Januszewski, M., Berger, D. R., Pope, A., Wu, Y., Blakely, T., Schalek, R. L., Li, P. H., Wang, S., Maitin-Shepard, J., et al. (2021). A connectomic study of a petascale fragment of human cerebral cortex. *BioRxiv*, (pp. 2021–05).
- [195] Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning important features through propagating activation differences. In *International conference on machine learning* (pp. 3145–3153).: PMLR.
- [196] Soltanolkotabi, M., Javanmard, A., & Lee, J. D. (2018). Theoretical insights into the optimization landscape of over-parameterized shallow neural networks. *IEEE Transactions on Information Theory*, 65(2), 742–769.
- [197] Sommer, K., Saalbach, A., Brosch, T., Hall, C., Cross, N., & Andre, J. (2020). Correction of motion artifacts using a multiscale fully convolutional neural network. *American Journal of Neuroradiology*, 41(3), 416–423.
- [198] Soni, N., Mehrotra, A., Behari, S., Kumar, S., & Gupta, N. (2017). Diffusion-tensor imaging and tractography application in pre-operative planning of intra-axial brain lesions. *Cureus*, 9(10).
- [199] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929–1958.
- [200] Stam, C. J., Jones, B., Nolte, G., Breakspear, M., & Scheltens, P. (2007). Small-world networks and functional connectivity in alzheimer’s disease. *Cerebral cortex*, 17(1), 92–99.

- [201] Stejskal, E. O. & Tanner, J. E. (1965). Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The journal of chemical physics*, 42(1), 288–292.
- [202] Strakowski, S. M., DelBello, M. P., Sax, K. W., Zimmerman, M. E., Shear, P. K., Hawkins, J. M., & Larson, E. R. (1999). Brain magnetic resonance imaging of structural abnormalities in bipolar disorder. *Archives of general psychiatry*, 56(3), 254–260.
- [203] Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *International conference on machine learning* (pp. 3319–3328): PMLR.
- [204] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.
- [205] Tanno, R., Worrall, D. E., Kaden, E., Ghosh, A., Grussu, F., Bizzi, A., Sotiropoulos, S. N., Criminisi, A., & Alexander, D. C. (2021). Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion mri. *NeuroImage*, 225, 117366.
- [206] Taymourtash, A., Kebiri, H., Schwartz, E., Nenning, K.-H., Tourbier, S., Kasprian, G., Prayer, D., Bach Cuadra, M., & Langs, G. (2022). Spatio-temporal motion correction and iterative reconstruction of in-utero fetal fmri. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI* (pp. 603–612): Springer.
- [207] Taymourtash, A., Schwartz, E., Nenning, K.-H., Sobotka, D., Licandro, R., Glatter, S., Diogo, M. C., Golland, P., Grant, E., Prayer, D., et al. (2023). Fetal development of functional thalamo-cortical and cortico–cortical connectivity. *Cerebral Cortex*, 33(9), 5613–5624.
- [208] Thill, B. (2023). The fetal pain paradox. *Frontiers in Pain Research*, 4.

- [209] Tian, Q., Bilgic, B., Fan, Q., Liao, C., Ngamsombat, C., Hu, Y., Witzel, T., Setsompop, K., Polimeni, J. R., & Huang, S. Y. (2020). Deepditi: High-fidelity six-direction diffusion tensor imaging using deep learning. *NeuroImage*, 219, 117017.
- [210] Tononi, G. (2004). An information integration theory of consciousness. *BMC neuroscience*, 5, 1–22.
- [211] Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7), 450–461.
- [212] Tononi, G. & Koch, C. (2008). The neural correlates of consciousness: an update. *Annals of the New York Academy of Sciences*, 1124(1), 239–261.
- [213] Tourbier, S., Bresson, X., Hagmann, P., Meuli, R., & Bach Cuadra, M. (2019). sebastientourbier/mialsuperresolutiontoolkit: MIAL Super-Resolution Toolkit v1.0. *Zenodo*.
- [214] Tourbier, S., Bresson, X., Hagmann, P., Thiran, J.-P., Meuli, R., & Cuadra, M. B. (2015). An efficient total variation algorithm for super-resolution in fetal brain mri with adaptive regularization. *NeuroImage*, 118, 584–597.
- [215] Tourbier, S., Velasco-Annis, C., Taimouri, V., Hagmann, P., Meuli, R., Warfield, S. K., Cuadra, M. B., & Gholipour, A. (2017). Automated template-based brain localization and extraction for fetal brain mri reconstruction. *NeuroImage*, 155, 460–472.
- [216] Tournier, J.-D. et al. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted mri data using spherical deconvolution. *Neuroimage*, 23(3), 1176–1185.
- [217] Tournier, J.-D., Yeh, C.-H., Calamante, F., Cho, K.-H., Connelly, A., & Lin, C.-P. (2008). Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. *Neuroimage*, 42(2), 617–625.

- [218] Tristan-Vega, A., Westin, C.-F., & Aja-Fernandez, S. (2010). A new methodology for the estimation of fiber populations in the white matter of the brain with the funk–radon transform. *NeuroImage*, 49(2), 1301–1315.
- [219] Trivedi, S. B., Vesoulis, Z. A., Rao, R., Liao, S. M., Shimony, J. S., McKinstry, R. C., & Mathur, A. M. (2017). A validated clinical mri injury scoring system in neonatal hypoxic-ischemic encephalopathy. *Pediatric radiology*, 47, 1491–1499.
- [220] Tuch, D. S. (2004). Q-ball imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 52(6).
- [221] Valavani, E., Blesa, M., Galdi, P., Sullivan, G., Dean, B., Cruickshank, H., Sitko-Rudnicka, M., Bastin, M. E., Chin, R. F., MacIntyre, D. J., et al. (2022). Language function following preterm birth: prediction using machine learning. *Pediatric Research*, 92(2), 480–489.
- [222] Van den Heuvel, M. I. & Thomason, M. E. (2016). Functional connectivity of the human brain in utero. *Trends in cognitive sciences*, 20(12), 931–939.
- [223] Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., Consortium, W.-M. H., et al. (2013). The wu-minn human connectome project: an overview. *Neuroimage*, 80, 62–79.
- [224] Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.
- [225] Vapnik, V. N. (1998). Adaptive and learning systems for signal processing communications, and control. *Statistical learning theory*.
- [226] Vasilev, A., Golkov, V., Meissner, M., Lipp, I., Sgarlata, E., Tomassini, V., Jones, D. K., & Cremers, D. (2020). q-space novelty detection with variational autoencoders. In *Computational Diffusion MRI: MICCAI Workshop, Shenzhen, China, October 2019* (pp. 113–124).: Springer.

- [227] Vassar, R., Schadl, K., Cahill-Rowley, K., Yeom, K., Stevenson, D., & Rose, J. (2020). Neonatal brain microstructure and machine-learning-based prediction of early language development in children born very preterm. *Pediatric neurology*, 108, 86–92.
- [228] Veiga, R., Stephan, L., Loureiro, B., Krzakala, F., & Zdeborová, L. (2022). Phase diagram of stochastic gradient descent in high-dimensional two-layer neural networks. *arXiv preprint arXiv:2202.00293*.
- [229] Vos, S. B., Aksoy, M., Han, Z., Holdsworth, S. J., Maclaren, J., Viergever, M. A., Leemans, A., & Bammer, R. (2016). Trade-off between angular and spatial resolutions in in vivo fiber tractography. *Neuroimage*, 129, 117–132.
- [230] Wasserthal, J., Neher, P., & Maier-Hein, K. H. (2018). Tractseg-fast and accurate white matter tract segmentation. *NeuroImage*, 183, 239–253.
- [231] Weiss, R. J., Bates, S. V., Song, Y., Zhang, Y., Herzberg, E. M., Chen, Y.-C., Gong, M., Chien, I., Zhang, L., Murphy, S. N., et al. (2019). Mining multi-site clinical data to develop machine learning mri biomarkers: application to neonatal hypoxic ischemic encephalopathy. *Journal of translational medicine*, 17(1), 1–16.
- [232] Wilson, S., Pietsch, M., Cordero-Grande, L., Price, A. N., Hutter, J., Xiao, J., McCabe, L., Rutherford, M. A., Hughes, E. J., Counsell, S. J., et al. (2021). Development of human white matter pathways in utero over the second and third trimester. *Proceedings of the National Academy of Sciences*, 118(20), e2023598118.
- [233] Wu, J., Zhang, Y., Wang, K., & Tang, X. (2019). Skip connection u-net for white matter hyperintensities segmentation from mri. *IEEE Access*, 7, 155194–155202.
- [234] Wurtman, R. J. & Wurtman, J. J. (1995). Brain serotonin, carbohydrate-craving, obesity and depression. *Obesity research*, 3(S4), 477S–480S.

- [235] Xu, J., Moyer, D., Gagoski, B., Iglesias, J. E., Grant, P. E., Golland, P., & Adalsteinsson, E. (2023). Nesvor: Implicit neural representation for slice-to-volume reconstruction in mri. *IEEE Transactions on Medical Imaging*.
- [236] Zhan, L., Leow, A. D., Jahanshad, N., Chiang, M.-C., Barysheva, M., Lee, A. D., Toga, A. W., McMahon, K. L., De Zubicaray, G. I., Wright, M. J., et al. (2010). How does angular resolution affect diffusion imaging measures? *neuroimage*, 49(2), 1357–1371.
- [237] Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021a). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3), 107–115.
- [238] Zhang, F., Breger, A., Cho, K. I. K., Ning, L., Westin, C.-F., O’Donnell, L. J., & Pasternak, O. (2021b). Deep learning based segmentation of brain tissue from diffusion mri. *NeuroImage*, 233, 117934.
- [239] Zhang, H., Schneider, T., Wheeler-Kingshott, C. A., & Alexander, D. C. (2012). Noddi: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*, 61(4), 1000–1016.
- [240] Zhang, H., Zhuang, Y., Xia, S., & Jiang, H. (2023). Deep learning network with spatial attention module for detecting acute bilirubin encephalopathy in newborns based on multimodal mri. *Diagnostics*, 13(9), 1577.
- [241] Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., & Shen, D. (2015). Deep convolutional neural networks for multi-modality iso-intense infant brain image segmentation. *NeuroImage*, 108, 214–224.
- [242] Zhao, C., Dewey, B. E., Pham, D. L., Calabresi, P. A., Reich, D. S., & Prince, J. L. (2020). Smore: a self-supervised anti-aliasing and super-resolution algorithm for mri using deep learning. *IEEE transactions on medical imaging*, 40(3), 805–817.

- [243] Zhong, J., Wang, Y., Li, J., Xue, X., Liu, S., Wang, M., Gao, X., Wang, Q., Yang, J., & Li, X. (2020). Inter-site harmonization based on dual generative adversarial networks for diffusion tensor imaging: application to neonatal white matter development. *Biomedical engineering online*, 19(1), 1–18.
- [244] Ziv, E., Tymofiyeva, O., Ferriero, D. M., Barkovich, A. J., Hess, C. P., & Xu, D. (2013). A machine learning approach to automated structural network analysis: application to neonatal encephalopathy. *PloS one*, 8(11), e78824.
- [245] Zuo, N., Cheng, J., & Jiang, T. (2012). Diffusion magnetic resonance imaging for brainnetome: A critical review. *Neuroscience bulletin*, 28(4), 375–388.

# 4

Main papers



# Through-Plane Super-Resolution With Autoencoders in Diffusion Magnetic Resonance Imaging of the Developing Human Brain

Hamza Kebiri<sup>1,2\*</sup>, Erick J. Canales-Rodríguez<sup>3</sup>, Hélène Lajous<sup>1,2</sup>, Priscille de Dumast<sup>1,2</sup>, Gabriel Girard<sup>1,2,3</sup>, Yasser Alemán-Gómez<sup>1</sup>, Mériam Koob<sup>1</sup>, András Jakab<sup>4,5</sup> and Meritxell Bach Cuadra<sup>1,2,3</sup>

<sup>1</sup> Department of Radiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland, <sup>2</sup> CIBM Center for Biomedical Imaging, Lausanne, Switzerland, <sup>3</sup> Signal Processing Laboratory 5 (LTS5), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, <sup>4</sup> Center for MR Research University Children's Hospital Zurich, Zurich, Switzerland, <sup>5</sup> Neuroscience Center Zurich, University of Zurich, Zurich, Switzerland

## OPEN ACCESS

### Edited by:

Daan Christiaens,  
KU Leuven, Belgium

### Reviewed by:

Maryam Afzali,  
Cardiff University, United Kingdom  
Arghya Pal,  
Harvard Medical School,  
United States

### \*Correspondence:

Hamza Kebiri  
hamza.kebiri@unil.ch

### Specialty section:

This article was submitted to  
Applied Neuroimaging,  
a section of the journal  
Frontiers in Neurology

Received: 02 December 2021

Accepted: 28 March 2022

Published: 02 May 2022

### Citation:

Kebiri H, Canales-Rodríguez EJ, Lajous H, de Dumast P, Girard G, Alemán-Gómez Y, Koob M, Jakab A and Bach Cuadra M (2022) Through-Plane Super-Resolution With Autoencoders in Diffusion Magnetic Resonance Imaging of the Developing Human Brain. *Front. Neurol.* 13:827816. doi: 10.3389/fneur.2022.827816

Fetal brain diffusion magnetic resonance images (MRI) are often acquired with a lower through-plane than in-plane resolution. This anisotropy is often overcome by classical upsampling methods such as linear or cubic interpolation. In this work, we employ an unsupervised learning algorithm using an autoencoder neural network for single-image through-plane super-resolution by leveraging a large amount of data. Our framework, which can also be used for slice outliers replacement, outperformed conventional interpolations quantitatively and qualitatively on pre-term newborns of the developing Human Connectome Project. The evaluation was performed on both the original diffusion-weighted signal and the estimated diffusion tensor maps. A byproduct of our autoencoder was its ability to act as a denoiser. The network was able to generalize fetal data with different levels of motions and we qualitatively showed its consistency, hence supporting the relevance of pre-term datasets to improve the processing of fetal brain images.

**Keywords:** unsupervised learning, autoencoders, super-resolution, diffusion-weighted imaging, magnetic resonance imaging (MRI), pre-term neonates, fetuses, brain

## 1. INTRODUCTION

The formation and maturation of white matter are at their highest rate during the fetal stage of human brain development. To have more insight into this critical period, *in utero* brain imaging techniques offer a unique opportunity. Diffusion weighted-magnetic resonance imaging (DW-MRI) is a well-established tool to reconstruct *in vivo* and non-invasively the white matter tracts in the brain (1, 2). Fetal DW-MRI, in particular, could characterize early developmental trajectories in brain connections and microstructure (3–6). Hence, fetal DW-MRI has been of significant interest for the past years where studies (7–9) have provided analysis using diffusion tensor imaging (DTI) by computing diffusion scalar maps such as fractional anisotropy (FA) or mean diffusivity (MD), using a limited number of gradient directions. A recent study focused on reconstructing fiber Orientation Distribution Functions (fODF) (10) using higher quality datasets and rich information including several gradient directions (32 and 80), higher b-values (750 and

1,000  $s/mm^2$ ), and signal-to-noise ratio (SNR) (3 Tesla magnetic field strength). Additionally, the datasets were acquired in a controlled and uniform research setting with healthy volunteers, which can hardly be reproduced in the clinical environment.

Albeit promising results, acquiring high-quality data remains the main obstacle in the field of fetal brain imaging. First, unpredictable and uncontrollable fetal motion is a major challenge. To overcome this problem, fast echo-planar imaging (EPI) sequences are typically used to freeze intra-slice motion. However, intra- and inter-volume motion still have to be addressed in the post-processing steps using sophisticated slice-to-volume registration (SVR) (11–13). Moreover, EPI sequences generate severe non-linear distortions that need adapted distortion correction algorithms (14). Additionally, the resulting images display low SNR due to at least three factors: the inherently small size of the fetal brain, the surrounding maternal structures and amniotic fluid, and the increased distance to the coils. In order to compensate for the low SNR in EPI sequences, series with thick voxels (i.e., low through-plane resolution) are often acquired. Finally, to shorten the acquisition time, small  $b$ -values ( $b = 400 - 700s/mm^2$ ) and a low number of gradient directions (10–15) (8, 9) are commonly used in fetal imaging, which in turn will result in a low angular resolution.

Clinical protocols typically acquire several anisotropic orthogonal series of 2D thick slices to cope with high motion and low SNR. Then, super-resolution reconstruction techniques that have been originally developed for structural T2-weighted images (15–20) by combining different 3D low resolution volumes have also been successfully applied in 4D fetal functional (21) and diffusion MRI contexts (10, 12). Still, despite these two pioneer works, super-resolution DW-MRI from multiple volumes has been barely explored *in vivo*. In fact, the limited scanning time to minimize maternal discomfort hampers the acquisition of several orthogonal series, resulting in a trade-off between the number of gradient directions and orthogonal series. Thus, DW-MRI fetal brain protocols are not standardized from one center to another (**Supplementary Table S1**) and more experiments have to be conducted in this area to design optimal sequences (22, 23). Sequence-based super-resolution methods that were applied in adult brains (24–27) could also be explored and adapted to fetal brains such as in Ning et al. (24) that acquire same orientation shifted low-resolution images in the slice encoding direction and in a non-overlapping gradient scheme to reconstruct one high-resolution volume using compressed sensing. The term *super-resolution* is used by both the image processing and the MR sequence development communities, though in a slightly different way. While the former works mainly on image space and the latter works on  $k$ -space, both aim at increasing the image resolution at different stages either using multiple volumes or single volumes.

In fact, fetal DW-MRI resolution enhancement could also benefit from *single image super-resolution* approaches, i.e., either within each DW-MRI 3D volume separately or using the whole 4D volume including all diffusion measurements. It has indeed been demonstrated that a linear or cubic interpolation of the raw signal enhances the resulting scalar maps and tractography (28). In practice, this is typically performed either at the signal

level or at DTI scalar maps (29). We believe that single volume and multiple volumes super-resolution can also be performed together, i.e., where the output of the former is given as the input of the latter. This aggregation could potentially lead to a better motion correction and hence to a more accurate final high-resolution volume.

Several studies have proposed single image super-resolution enhancement methods for DW-MRI but, to the best of our knowledge, none of them has been applied neither to anisotropic datasets nor to the developing brain. In Coupé et al. (30), the authors utilized a non-local patch-based approach in an inverse problem paradigm to improve the resolution of adult brain DW-MRI volumes using a non diffusion weighted image ( $b = 0s/mm^2$ ) prior. Although this approach yielded competitive results, it was built upon a sophisticated pipeline which made it not extensively used. The first machine learning study (31) have used shallow learning algorithms to learn the mapping between diffusion tensor maps of a downsampled high-resolution image and the maps of the original image. Recently, deep learning models which can implicitly learn relevant features from training data were used to perform single image super-resolution with a convolutional neural network (32, 33) and a customized U-Net (34, 35). Both approaches produced promising results in a *supervised* learning scheme. Supervision needs however large high quality datasets that are scarce for the perinatal brain for the reasons enumerated above.

The specific challenge of fetal DW-MRI is 3–5 mm acquired slice thickness, with only a few repetitions available. Hence, our main objective is to focus on through-plane DW-MRI resolution enhancement. This would be valuable not only for native anisotropic volumes but also for outlier slice recovery. In fact, motion-corrupted slices in DW-MRI is either discarded, which results in a loss of information, or replaced using interpolation (36–38). We approached this problem from an image synthesis point of view using *unsupervised* learning networks such as autoencoders (AEs), as demonstrated in cardiac T2-weighted MRI (39) and recent works in DW-MRI (40). Here, we present a framework with autoencoders that are neural networks learning in an *unsupervised* way to encode efficient data representations and can behave as generative models if this representation is structured enough. By accurately encoding DW-MRI slices in a low-dimensional latent space, we were able to successfully generate new slices that accurately correspond to in-between “missing” slices. In contrast to the above referred *supervised* learning approaches, this method is scale agnostic, i.e., the enhancement scale factor can be set *a posteriori* to the network training.

Realistically enhancing the through-plane resolution would potentially help the clinicians to better assess whether the anterior and posterior commissures are present in cases with complete agenesis of the corpus callosum (6). It can reduce partial volume effects and thus contribute to the depiction of more accurate white matter properties in the developing brain.

In this work, we present the first unsupervised through-plane resolution enhancement for perinatal brain DW-MRI. We leverage the high-quality dataset of the developing Human Connectome Project (dHCP) where we train and quantitatively

**TABLE 1** | Pre-term newborns and fetal attributes of the processed data that were used in our experiments.

	GW	Resolution ( $mm^3$ )	#directions	$b$ -value ( $s/mm^2$ )	Scanner	#subjects
Pre-terms	[29.3,37.0]	1.17x1.17x1.5	88	0 and 1,000	Philips 3T	31
<b>Fetal</b>						
Sub-1	35	1x1x5	15	0 and 700	GE 1.5T	–
Sub-2	29	1x1x4	15	0 and 700	GE 1.5T	–
Sub-3	24	1x1x5	15	0 and 700	GE 1.5T	–
Sub-4	23	1x1x4	15	0 and 700	GE 1.5T	–
Sub-5	24	1x1x5	15	0 and 700	GE 1.5T	–
Sub-6	27	1x1x5	15	0 and 700	GE 1.5T	–

The distribution of gestational ages is shown in **Supplementary Figure S1**.

validate pre-term newborns that are anatomically close to fetal subjects. We finally demonstrate the performance of our approach in fetal brains.

## 2. MATERIALS AND METHODS

### 2.1. Data

#### 2.1.1. Pre-term dHCP Data

We selected all the 31 pre-term newborns of 37 gestational weeks (GW) or less at the time of scan (range: [29.3,37.0], mean: 35.5, median: 35.7) from the dHCP dataset (41) (subject IDs in **Supplementary Table S2**). Acquisitions were performed using a 3T Philips Achieva scanner (32-channel neonatal head-coil and 70 mT/m gradients) with a monopolar spin-echo EPI Stejskal-Tanner sequence ( $\Delta = 42.5$  ms,  $\delta = 14$  ms, TR = 3,800 ms, TE = 90ms, echo spacing = 0.81ms, EPI factor = 83) and a multiband factor of 4, resulting in an acquisition time of 19:20 min. In a field of view of  $150 \times 150 \times 102$   $mm^3$ , 64 interleaved slices were acquired with an in-plane resolution of 1.5 mm, a slice thickness of 3 mm, and a slice overlap of 1.5 mm. An isotropic volume of  $1.5$   $mm^3$  was obtained after super-resolution. The dataset was acquired with a multi-shell sequence using four  $b$ -values ( $b \in \{0, 400, 1, 000, \text{and } 2, 600\}$   $s/mm^2$ ) with 300 volumes but we have only extracted the 88 volumes corresponding to  $b = 1, 000$   $s/mm^2$  (b1000) as a compromise of high contrast-to-noise ratio (CNR), i.e., b1000 has a higher CNR than b400 and b2600 (42), and proximity to the  $b = 700$   $s/mm^2$  that is typically used in clinical settings for fetal DW-MRI. The main attributes of the pre-term data are summarized in **Table 1**. Brain masks and region/tissue labels segmented using a pipeline based on the Draw-EM algorithm (43, 44) were available in the corresponding anatomical dataset. All the images were already corrected (42) for inter-slice motion and distortion (susceptibility, eddy currents and motion). After pre-processing, the final image resolution and FOV were, respectively,  $1.17 \times 1.17 \times 1.5$   $mm^3$  and  $128 \times 128 \times 64$   $mm^3$ .

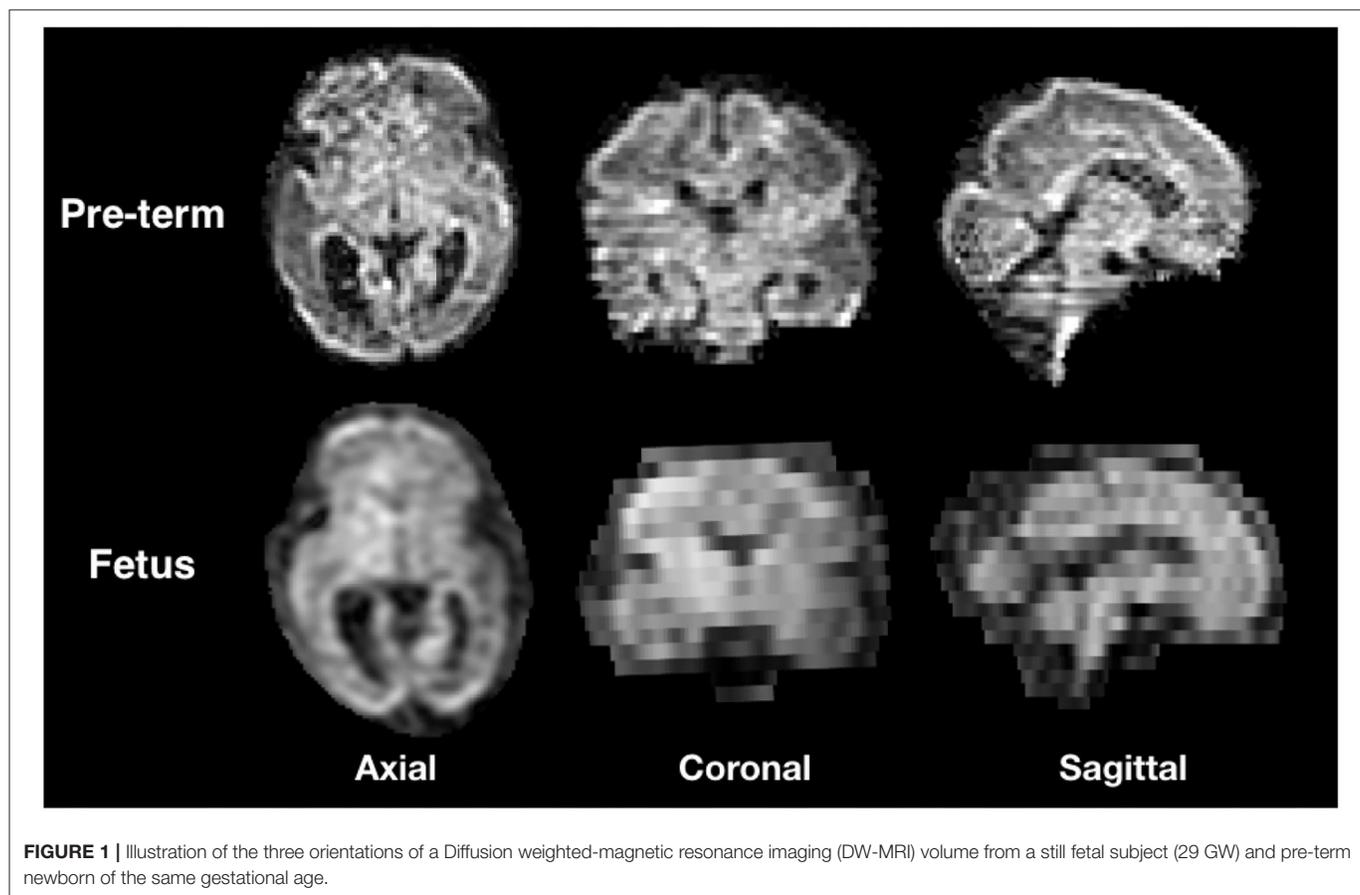
#### 2.1.2. Fetal Data

Fetal acquisitions were performed at 1.5T (MR450, GE Healthcare, Milwaukee, WI, USA) in the University Children's Hospital Zürich (KISPI) using a single-shot EPI sequence (TE = 63 ms, TR = 2200 ms) and 15 gradient directions at  $b = 700$   $s/mm^2$  (b700). The acquisition time was approximately 1.3

min per 4D volume. The in-plane resolution was  $1 \times 1$   $mm^2$ , the slice thickness was 4–5 mm, and the field of view  $256 \times 256 \times 14$  – 22 voxels. Three axial series and a coronal one were acquired for each subject. Brain masks were manually generated for the  $b_0$  ( $b = 0$   $s/mm^2$ ) of each acquisition and automatically propagated to the diffusion-weighted volumes. Between 8 and 18, T2-weighted images were also acquired for each subject where corresponding brain masks were automatically generated using an in-house deep learning based method using transfer learning from Salehi et al. (45). Manual refinements were needed for a few cases at the brain boundaries.

#### 2.1.3. Fetal Data Processing

We selected three subjects with high quality imaging and without motion artifacts (24, 29, and 35 GW) and three subjects with a varying degree of motion (23, 24, and 27 GW). **Supplementary Figure S1** shows the distribution of gestational age of both 31 pre-term newborns and the 6 fetal subjects used in this study. A DW-MRI volume of a motion-free case (Sub-2, 29 GW) and a pre-term of equivalent age are illustrated in **Figure 1**. By performing quality control, we discarded highly corrupted volumes due to motion resulting in severe signal drops in two moving subjects and very low SNR volumes in one motion-free subject. **Table 2** presents the different characteristics of each subject as well as its corresponding discarded volumes. The coronal volume was not used to avoid any interpolation confounding factor while co-registering different orientations. All the subjects were pre-processed for noise, bias field inhomogeneities, and distortions using the Nipype framework (46). The denoising was performed using a Principal Component Analysis based method (47), followed by an N4 bias-field inhomogeneity correction (48). Distortion was corrected using an in-house implementation of a state-of-the-art algorithm for the fetal brain (14) consisting in rigid registration (49) of a structural T2-weighted image to the  $b_0$  image, followed by a non-linear registration (49) in the phase-encoding direction of the  $b_0$  to the same T2-weighted image. The transformation was then applied to the diffusion-weighted volumes. A block matching algorithm for symmetric global registration was also performed for two subjects (sub-4, sub-6) with motion [NiftyReg, (50)]. The  $b_0$  image of the first axial series was selected as a reference to which we subsequently registered the remaining volumes, i.e.,



**TABLE 2** | Fetal motion level and discarded directions for each 4D volume.

	Motion level	Discarded			
		Axial-1	Axial-2	Axial-3	Coronal
Sub-1	No	–	–	–	All
Sub-2	No	–	–	–	All
Sub-3	No	All except b0	–	All except b0	All
Sub-4	High	Vol 2,3,4,5,6,7	Vol 11,14	Vol 14	All
Sub-5	Average	–	Vol 6,7,15	Vol 11	All
Sub-6	Low	Vol 11	Vol 11	Vol 11,14	All

the non b0 images from the first axial and all volumes from the two others. Gradient directions were rotated accordingly. **Supplementary Figure S2** shows an example of a DWI volume (from sub-4) of original, pre-processing, and motion correction.

## 2.2. Model

### 2.2.1. Architecture

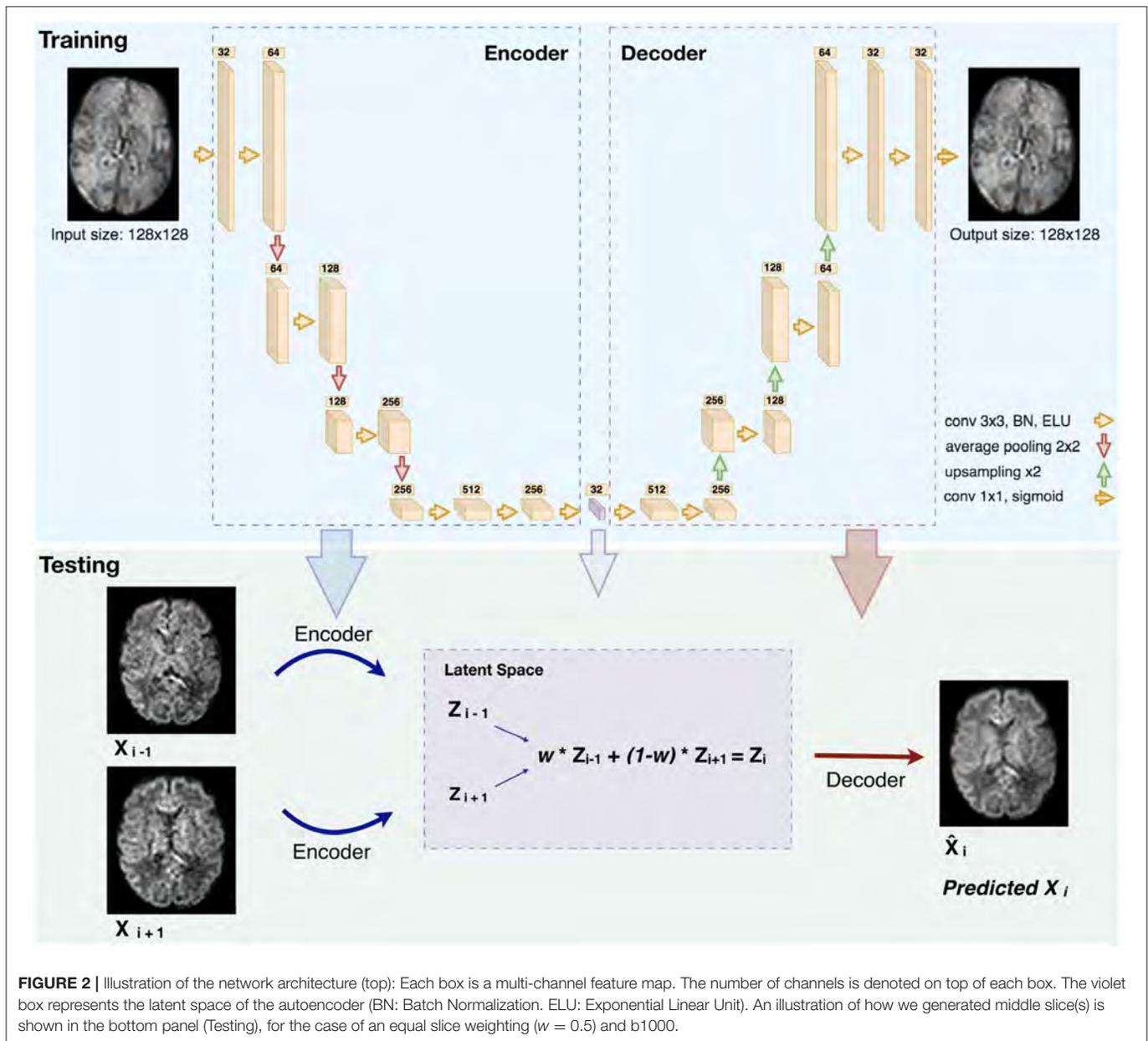
Our network architecture, similarly to Sander et al. (39), is composed of four blocks in the encoder and four in the decoder (**Figure 2**). Each block in the encoder consists of two layers made of 3 x 3 convolutions followed by a batch normalization (51) and an Exponential Linear Unit non-linearity.

The number of feature maps is doubled from 32 after each layer and the resulting feature maps are average-pooled. We

further added two layers of two 3 x 3 convolutions in which the feature maps of the last layer were used as the latent-space of the autoencoder. The decoder uses the same architecture as the encoder but by conversely halving the number of feature maps and upsampling after each block using nearest-neighbor interpolation. At the final layer a 1 x 1 convolution using the sigmoid function is applied to output the predicted image. The number of network parameters is 6,098,689.

### 2.2.2. Training and Optimization

We have trained our network solely on b0 images (15 per subject), using an 8-fold nested cross validation where we trained and validated on 27 subjects and tested on four. The proportion of the validation data was set to 15% of the training set. The



training/validation set contains 25,920 slices of a 128 x 128 field of view, totaling 424,673,280 voxels. Our network was trained in an unsupervised manner by feeding normalized 2D axial slices that are encoded as feature maps in the latent space. The number of feature maps, and hence the dimensionality of the latent space, was optimized (optimal value to 32) using Keras-Tuner (52). The batch size and the learning rate were additionally optimized and set to 32 and  $5e-5$ , respectively. The network that was initialized using (53) was trained for 200 epochs to minimize the mean squared error loss between the predicted and the ground truth image. We have utilized for this aim the Adam optimizer (54) with the default parameters  $\beta_1 = 0.5$ ,  $\beta_2 = 0.999$ , and the network corresponding to the epoch with the minimal validation loss was then selected. The implementation was performed in

the framework of TensorFlow 2.4.1 (55) and an Nvidia GeForce RTX 2080 GPU was deployed for training. Network code and checkpoint examples can be found in our Github repository<sup>1</sup>.

### 2.2.3. Inference

The network trained on b0 images was used for the inference of b0 and b1000 volumes. Two slices were encoded in the latent space and their  $N$  “in-between” slice(s) ( $N = 1, 2$  in our experiments) were predicted using weighted averages of the latent codes of the two slices. The weights for  $N = 1$  and  $N = 2$  were set proportionally to their distance to the neighboring

<sup>1</sup>[www.github.com/Medical-Image-Analysis-Laboratory/Perinatal\\_SR\\_Auto\\_encoder](http://www.github.com/Medical-Image-Analysis-Laboratory/Perinatal_SR_Auto_encoder)

original slices [as performed in Sander et al. (39)], i.e., an equal weighting for  $N = 1$  and  $\{\frac{1}{3}, \frac{2}{3}\}$ ,  $\{\frac{2}{3}, \frac{1}{3}\}$  for  $N = 2$ . Performing a grid search on ten weights (0.1–0.9 with a step of 0.1) confirmed the optimality of the previous choice. An example of pre-term b1000 data for a weight of 0.5 is shown in **Figure 2** (Testing). Similarly, the same b0 network was also used to enhance the through-plane resolution of fetal b0 and b700 volumes. Finally, since the network outputs were normalized between 0 and 1, histogram normalization to the weighted average of the input images was performed.

## 2.3. Experiments and Evaluation

### 2.3.1. Pre-term Newborns

Our network was separately tested on b0 images and the 88 volumes of b1000 using an 8-fold cross validation where 7-folds contain four subjects and 1-fold contains three subjects. We removed  $N$  intermediate slices ( $N = [1,2]$ ) from the testing set volumes in alternating order and used the (weighted) average latent space feature maps of the to-be adjacent slices to encode the  $N$  missing slice(s) using the autoencoder (**Figure 2**, Testing). The resulting latent representation was then decoded to predict the  $N$  slices in the voxel space, which were compared to the previously removed  $N$  slices, i.e., the ground truth (GT). The same  $N$  slices were also generated using three baseline approaches: trilinear, tricubic, and B-spline of 5<sup>th</sup> order interpolations [using Tournier et al. (56) and Avants et al. (49)] for comparison. We denote them, respectively, for removing one or two slices: Linear-1, Cubic-1, Spline-1 and Linear-2, Cubic-2, Spline-2.

*Latent space exploratory analysis* - In order to have an intuitive idea of the latent space representation, we have compared the latent space representations between different gradient directions of all possible pairs from the 88 volumes of the b1000 4D volume. As two volumes with closely aligned gradient directions are more similar than two volumes with orthogonal directions, we aimed to check whether this property is globally preserved in the latent encoding of our input images.

*Robustness to noise* - We have added different low levels of Rician noise (57) to the original signal as follows: for each pixel with a current intensity  $S_{clean}$ , the new intensity  $S_{noisy} = \sqrt{(S_{clean} + GN_1)^2 + GN_2^2}$ , where  $GN_1$  and  $GN_2$  are random numbers sampled from a Gaussian distribution with zero mean and a SD of  $S_{clean}(b = 0)/SNR_{out}$  and  $SNR_{out}$  is the desired SNR we aim to simulate. Three SNRs of {27, 25, 23} and {20, 16, 13} were simulated for b0 and b1000, respectively. We have used higher noise levels for b1000 to better simulate the inherently lower SNR in this configuration.

*Scalar maps* - By merging the b0 and b1000 using the autoencoder enhancement, we reconstructed FA, MD, axial diffusivity (AD), and radial diffusivity (RD) from DTI using Dipy (58) separately for AE-1 or AE-2, i.e., where we, respectively, remove one ( $N = 1$ ) or two slices ( $N = 2$ ). We further subdivided the computation in specific brain regions (cortical gray matter, white matter, corpus callosum, and brainstem as provided by the dHCP). Region labels were upsampled and manually refined to match the super-resoluted/interpolated volumes. We performed similar computation of the diffusion

maps generated using the trilinear, tricubic, and B-spline interpolated signals.

### 2.3.2. Fetal

For each subject and each 3D volume (b0 or DW-MRI), we generated one or two middle slices using the autoencoder, hence synthetically enhancing the resolution from  $1 \times 1 \times 4\text{--}5 \text{ mm}^3$  to a simulated resolution of  $1 \times 1 \times 2\text{--}2.5 \text{ mm}^3$  and  $1 \times 1 \times 1.33\text{--}1.67 \text{ mm}^3$ , respectively. We then generated whole-brain DTI maps (FA, MD, AD, and RD) and showed the colored FA. Splenium and genu structures of the corpus callosum were additionally segmented on FA maps for subjects in which these structures were visible. The mean FA and MD were reported for these regions for original and autoencoder enhanced volumes.

### 2.3.3. Quantitative Evaluation

*Raw diffusion signal* - We computed the voxel-wise error between the raw signal synthesized by the autoencoder and the GT using the mean squared error (MSE) and the peak SNR (PSNR). We compared the autoencoder performance with the three baseline approaches: trilinear, tricubic, and B-spline of 5<sup>th</sup> order interpolations.

*Latent space exploratory analysis* - We have computed the average squared Euclidian voxel-wise distance between slices of all 3D b1000 volume pairs. This was performed both at the input space and at the latent space representation. The images were flattened from 2D to one-dimensional vectors and compared as follows:

$$d(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \dots + (u_n - v_n)^2} \quad (1)$$

Where  $\vec{u}$  and  $\vec{v}$  are the vectors to be compared for all the  $n$  corresponding pixels. The final distance between each two 3D volumes is the average distance of all 2D distance computed in 1.

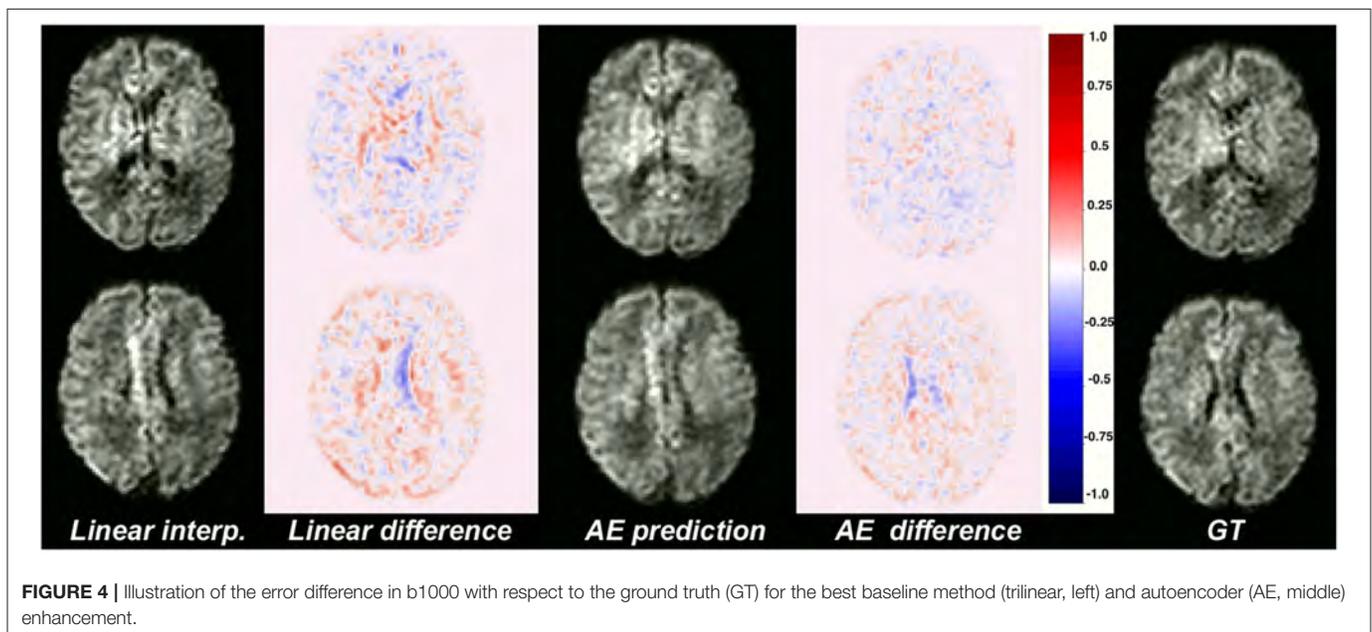
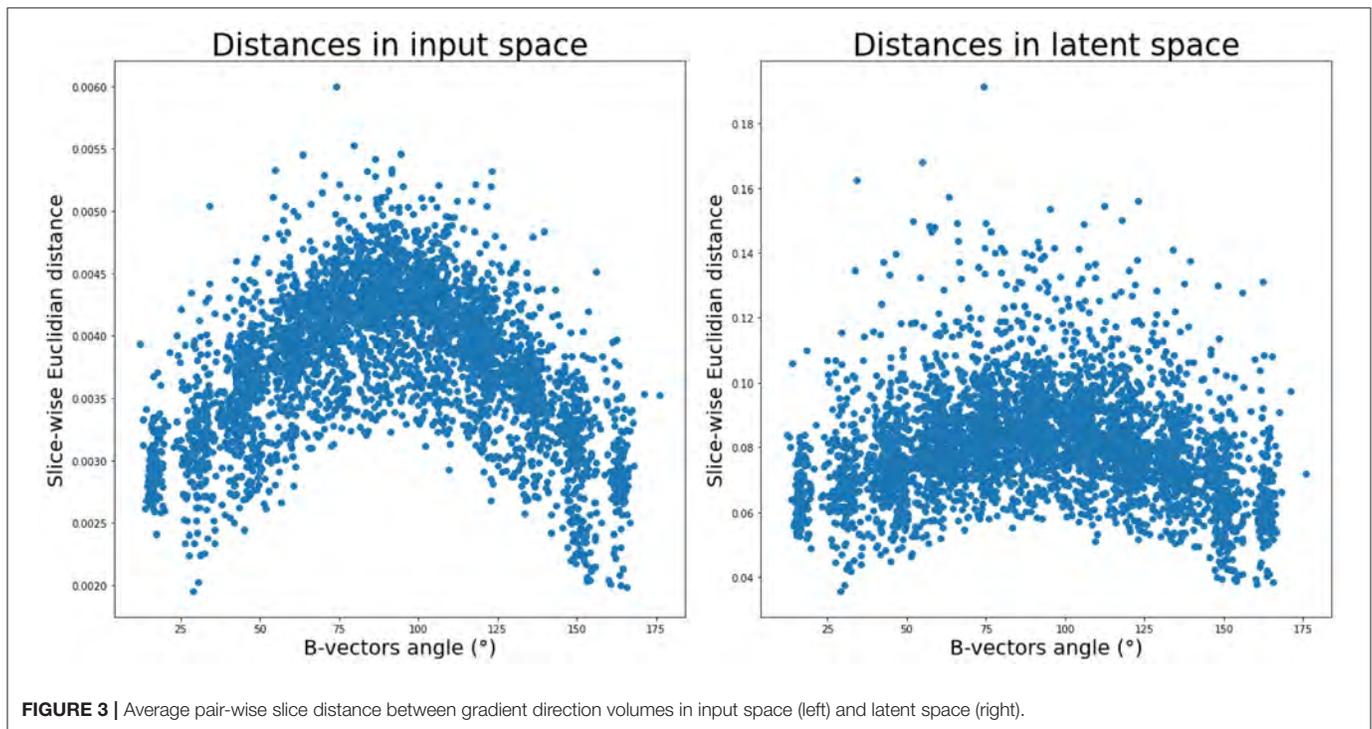
*Robustness to noise* - We computed with respect to the GT signal, the error of the signal with noise, and the output of the autoencoder using the signal with noise as input. We compared the results using MSE separately for b0 and b1000.

*Scalar maps* - We computed the voxel-wise error between the diffusion tensor maps reconstructed with the GT and the one by merging the b0 and b1000 using the autoencoder enhancement. We computed the error separately using either AE-1 or AE-2. We used the MSE and the PSNR as metrics and the same diffusion maps generated using the trilinear, tricubic, and B-spline interpolated signal as a baseline. Moreover, we qualitatively compare colored FA generated using the best baseline method, autoencoder, and the GT.

## 3. RESULTS

### 3.1. Pre-term Newborns

First, we inspected the latent space and how the 88 DW-MRI volumes are encoded with respect to each other. We can notice in **Figure 3** (right panel) that as two b-vectors' angle approaches orthogonality ( $90^\circ$ ), the difference between the latent representations of their corresponding volumes increases. On



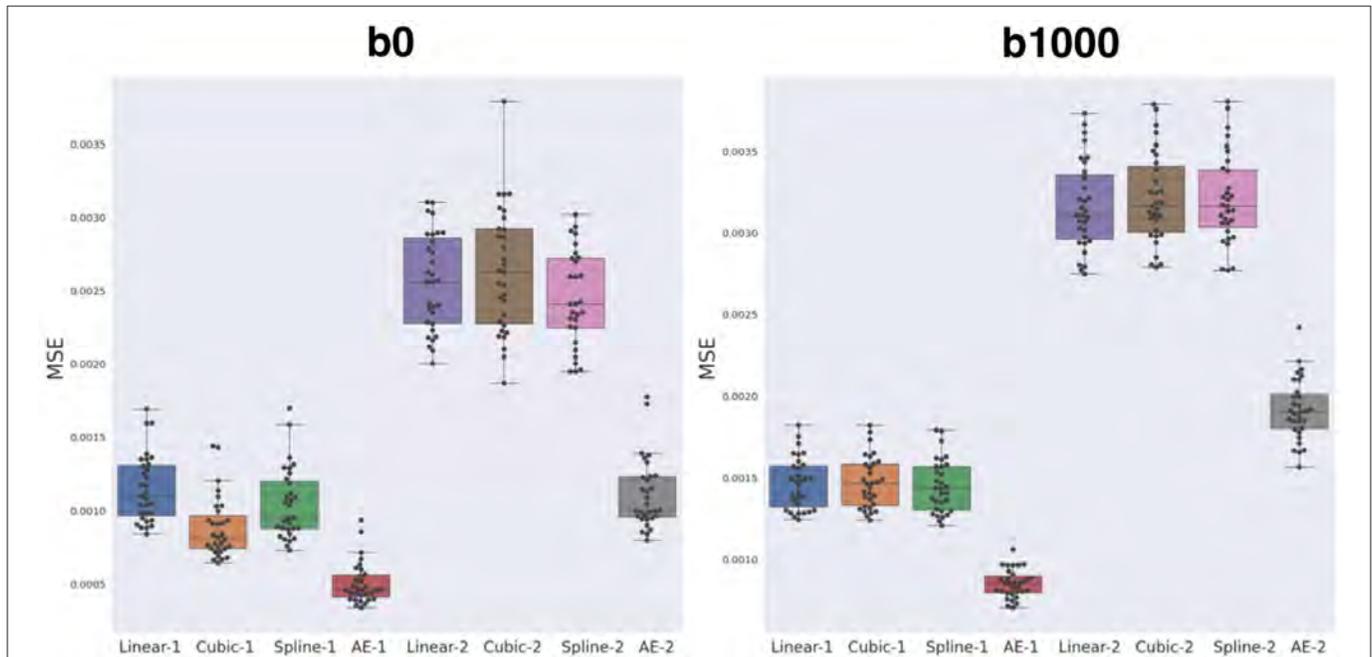
the contrary, the difference decreases the more the angle tends toward  $0^\circ$  or  $180^\circ$ . Although the pattern is more pronounced in the input space (**Figure 3**, left panel), this trend is a fulfilled necessary condition to the generation for coherent representations of the input data by our network.

Moreover, our network that was exclusively trained on b0 images was able to generalize to b1000. In fact, the signal similarity between b0 and DW images was also used in Coupé et al. (30) in an inverse problem paradigm in which a b0 prior was incorporated to reconstruct b700 volumes.

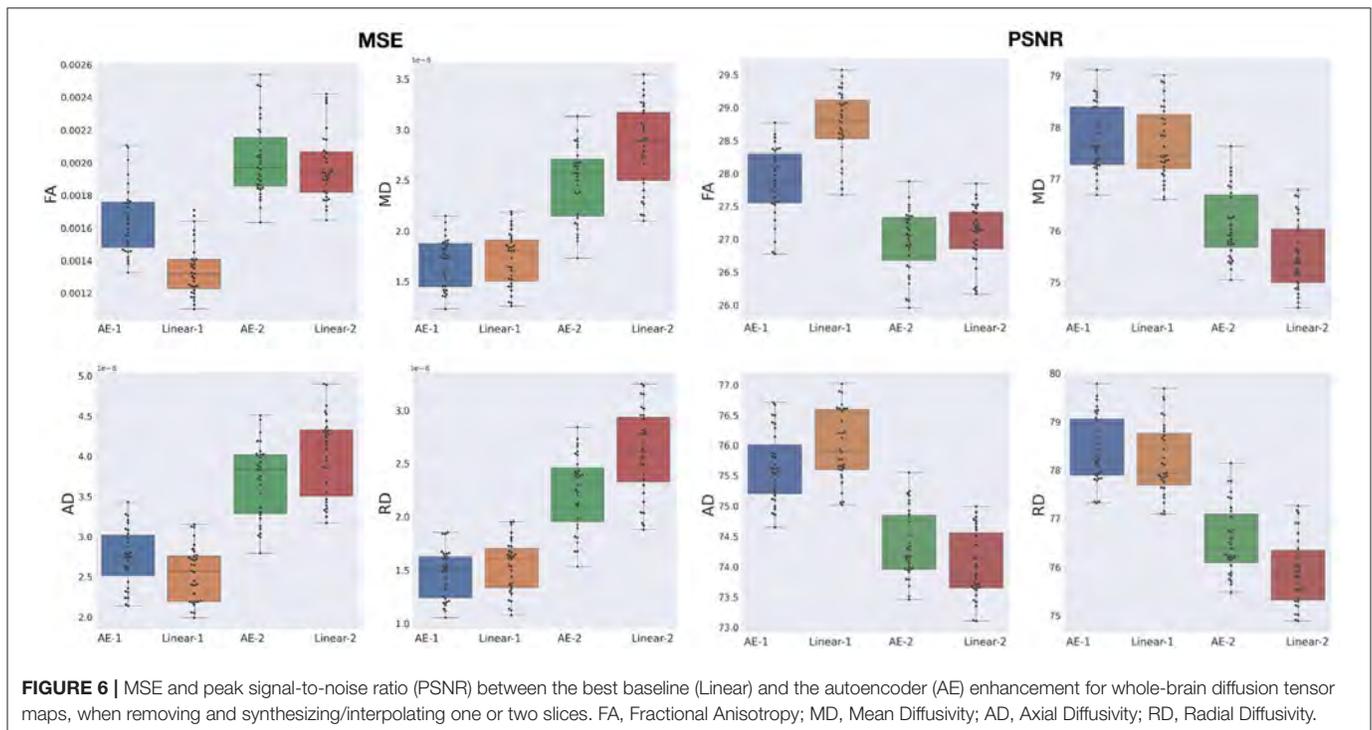
**Figure 4** illustrates qualitative results and absolute errors for  $N = 1$  with respect to the GT (right) between the best interpolation baseline (trilinear, left) and the autoencoder enhancement (middle) for b1000. We overall saw from these representative examples, higher absolute intensities in the Linear-1 configuration than in the AE-1. However, ventricles are less visible when using an autoencoder. We hypothesize this is because of their higher intensity in b0 images on which the network was trained.

The average MSE with respect to the original DW-MRI signal within the whole brain is shown in **Figure 5** for both the autoencoder enhanced volume and the baseline methods

(trilinear, tricubic, and B-spline), for the configurations where one (Method-1) or two (Method-2) slices were removed. The first observation was the expected higher error for the configuration



**FIGURE 5 |** Mean squared error (MSE) between the three baseline methods (linear, cubic, and B-spline 5<sup>th</sup> order) and autoencoder (AE) enhancement both for b0 (left) and b1000 (right). Two configurations were assessed: either  $N = 1$ , i.e., removing one slice and interpolating/synthesizing it (Linear-1, Cubic-1, Spline-1, AE-1) or  $N = 2$ , i.e., the same approach with two slices (Linear-2, Cubic-2, Spline-2, and AE-2). The autoencoder has a significantly lower MSE when compared to each respective best baseline method (paired Wilcoxon signed-rank test  $p < 1.24e-09$ ).

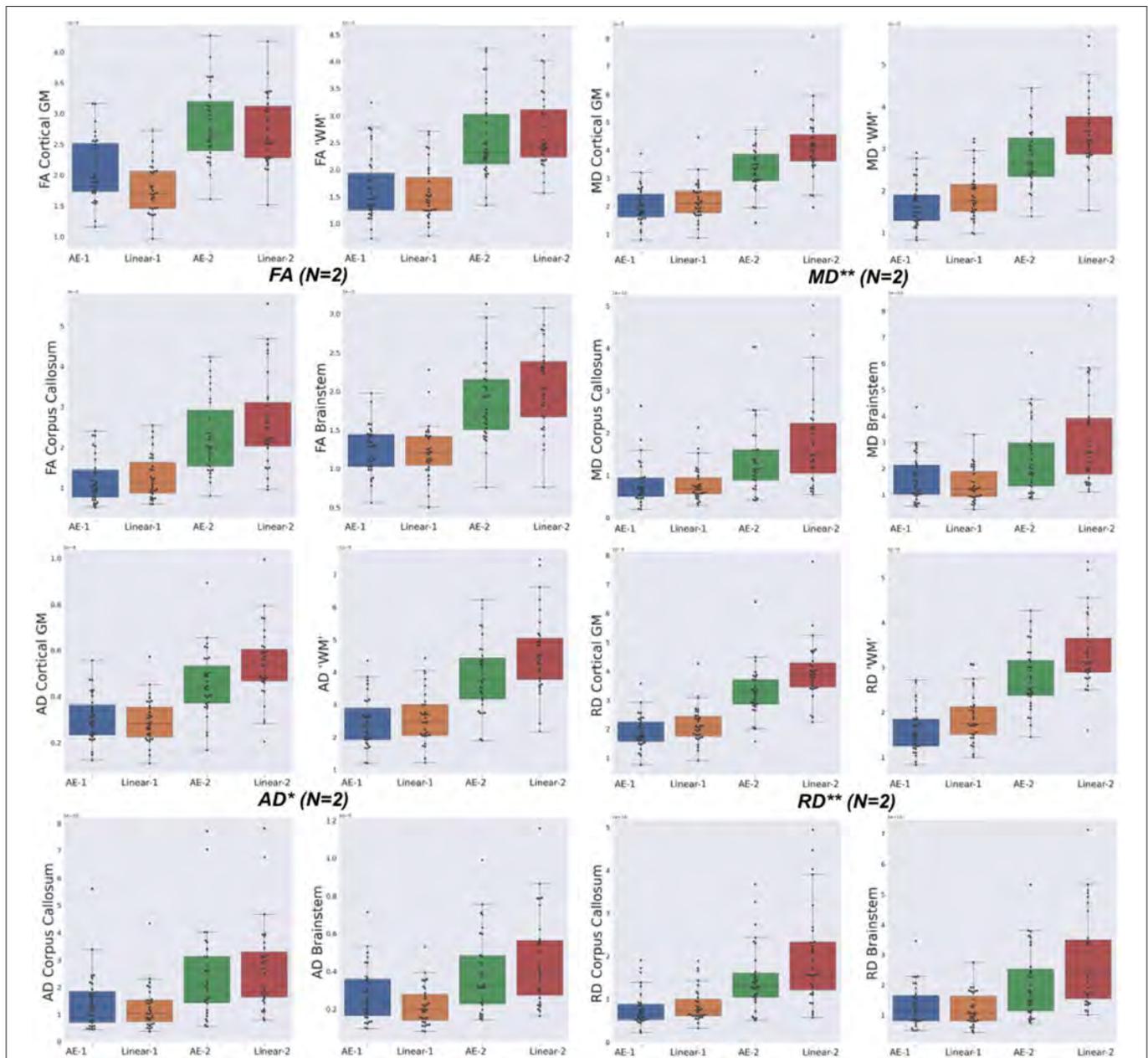


**FIGURE 6 |** MSE and peak signal-to-noise ratio (PSNR) between the best baseline (Linear) and the autoencoder (AE) enhancement for whole-brain diffusion tensor maps, when removing and synthesizing/interpolating one or two slices. FA, Fractional Anisotropy; MD, Mean Diffusivity; AD, Axial Diffusivity; RD, Radial Diffusivity.

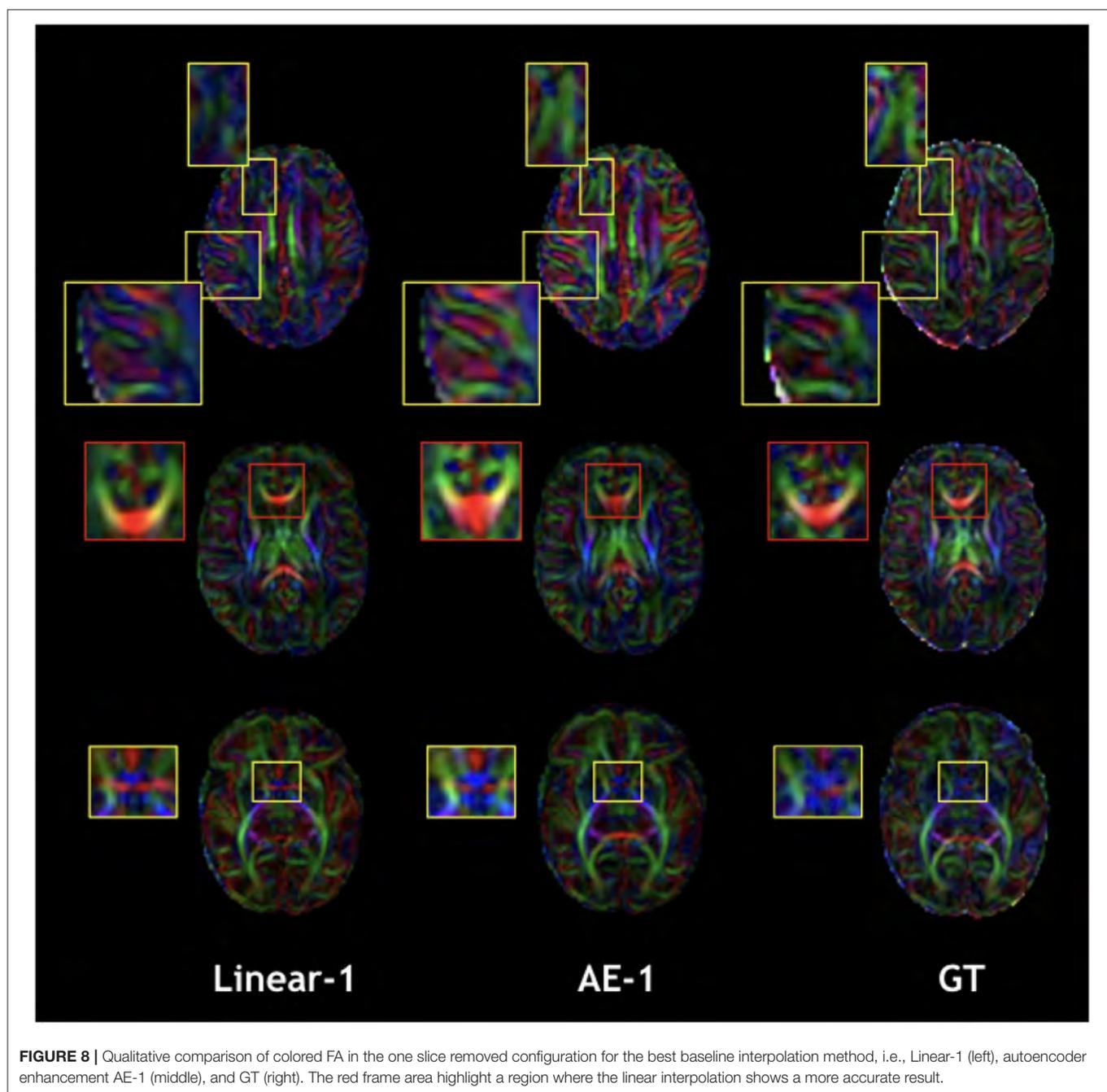
where two slices are removed ( $N = 2$ ), independently of the method used. Additionally, the autoencoder enhancement clearly outperformed the baseline methods in all configurations (paired Wilcoxon signed-rank test  $p < 1.24e-09$ ). Particularly, the more slices we remove, the higher the gap between the baseline interpolation methods and the autoencoder enhancement. For  $b_0$ , the MSE gain was around 0.0005 for  $N = 1$  and 0.0015 for  $N = 2$  between the autoencoder and the average baseline method (Spline-1 v.s. AE-1 and Linear-2 v.s. AE-2). For  $b_{1000}$ , the gain

between AE-1 and Cubic-1 was 0.0007 and 0.0015 between AE-2 and Cubic-2.

The overperformance of the autoencoder is also shown overall in the DTI maps, where MD, AD, and RD were better approximated when compared to the best baseline method (linear interpolation), particularly in the configuration where two slices were removed (Figure 6). However, the FA showed the opposite trend, especially for the configuration, where one slice was removed (AE-1 v.s. Linear-1). However, FA

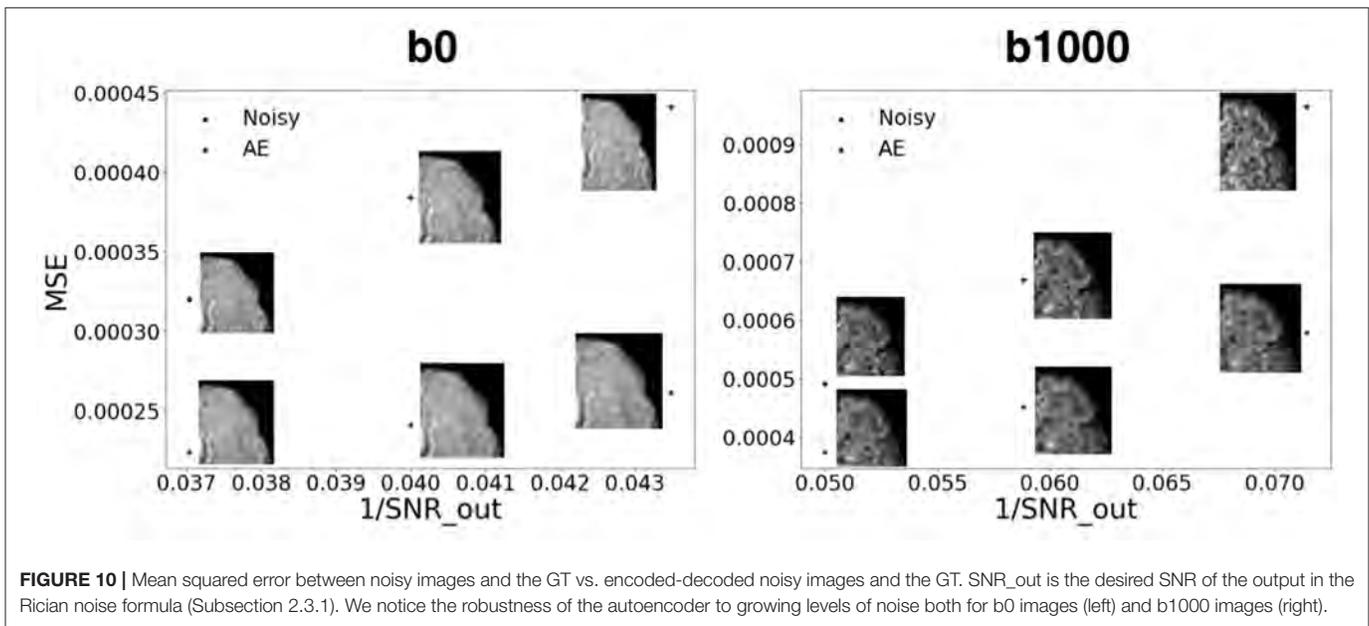
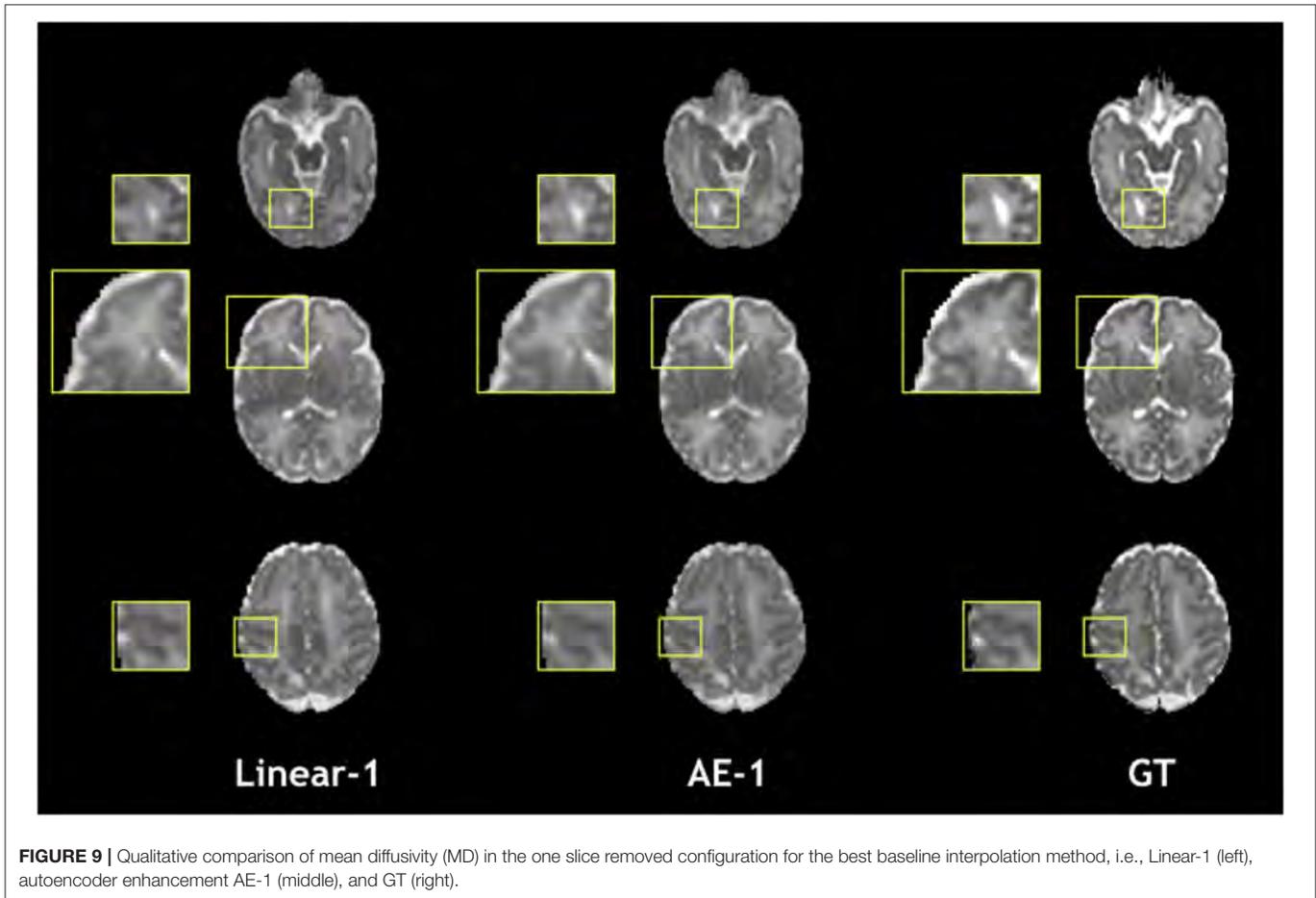


**FIGURE 7 |** Mean squared error (MSE) with respect to the GT of the best baseline method (Linear) and the autoencoder (AE) enhancement in the different brain structures [Cortical Gray Matter (GM), White Matter (WM), Brainstem, and Corpus Callosum] for each diffusion tensor map (FA, MD, AD, and RD) for one slice removal ( $N = 1$ ) and two slices removal ( $N = 2$ ). Comparing the DTI maps of the merged brain region labels, we found that the AE-2 significantly outperforms other conventional methods for MD, RD, and AD. (Paired Wilcoxon signed-rank test: \*\* $p < 0.0018$  and \* $p < 0.017$ ).

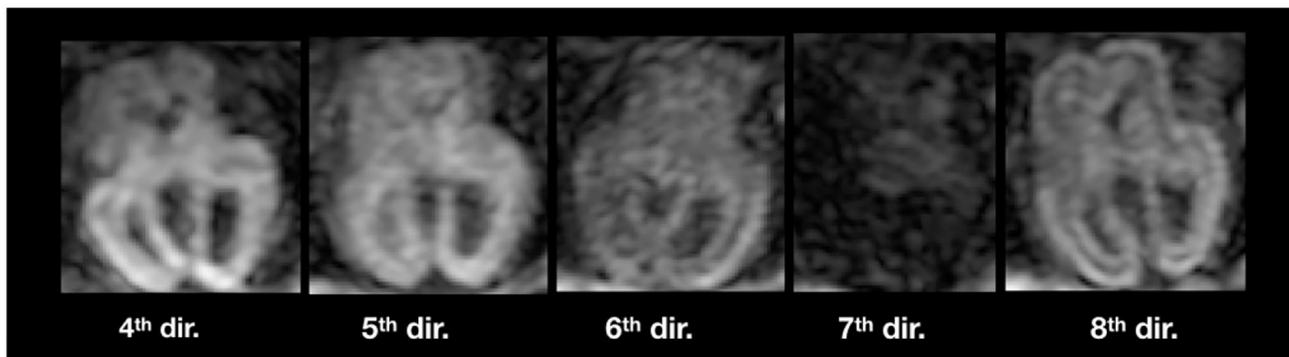


for white matter-like structures (“WM”, corpus callosum, and brainstem) showed higher performance with the autoencoder as depicted for each structure in **Figure 7**. In fact, by plotting colored FA for these two configurations, we observed that the autoencoder generates tracts that were consistent with the GT. For instance, autoencoder enhancement showed higher frequency details around the superficial WM area (**Figure 8**, top row) and removed artifacts between the internal capsules better than the linear method (**Figure 8**, bottom row). However, in some cases, the baseline method better depicted tracts such as in the corpus callosum (**Figure 8**, middle row).

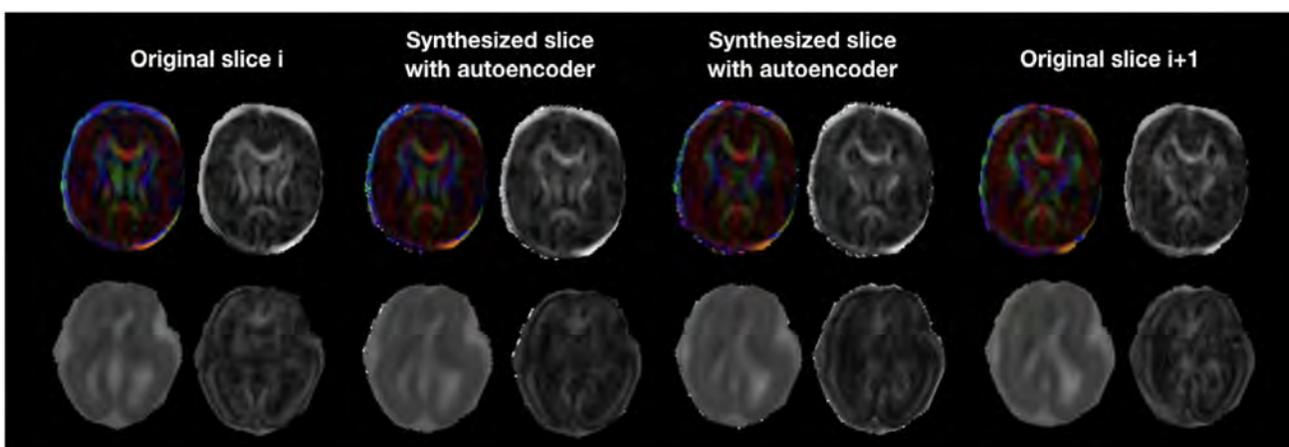
ODFs generated using spherical harmonics order 8 are also depicted in **Supplementary Figure S3** where the autoencoder enhanced data show little qualitative differences with the GT ODFs. **Figure 9** shows similar comparisons for MD in different brain regions between the baseline method (Linear), the autoencoder, and the GT. Overall, quantitatively, for structures in the case where two slices were removed, the autoencoder enhancement outperformed the best baseline method in 15 out of 16 configurations (**Figure 7**). However, it is not always the case when one slice is removed, such as in the AD of the brainstem.



**Figure 10** shows how our autoencoder was robust to reasonable amounts of noise. In fact, simply encoding and decoding the noisy input generates a slice that was closer to the GT than the noisy slice, as depicted for different levels of noise for both b0 and b1000.



**FIGURE 11** | Illustration of inter-volume motion in five different gradient directions of sub-4 (Table 2). Note the severe signal drop in the seventh direction because of motion.



**FIGURE 12** | Colored FA and FA (top row) illustration of autoencoder enhancement between two original adjacent fetal slices in a still subject (sub-1, 35 GW). The bottom row shows a similar illustration of MD and FA for a moving subject (sub-4, 23 GW).

### 3.2. Fetuses

Figure 11 illustrates inter-volume motion between five diffusion-weighted volumes where we also notice a severe signal drop in the seventh direction (sub-4, 23 GW).

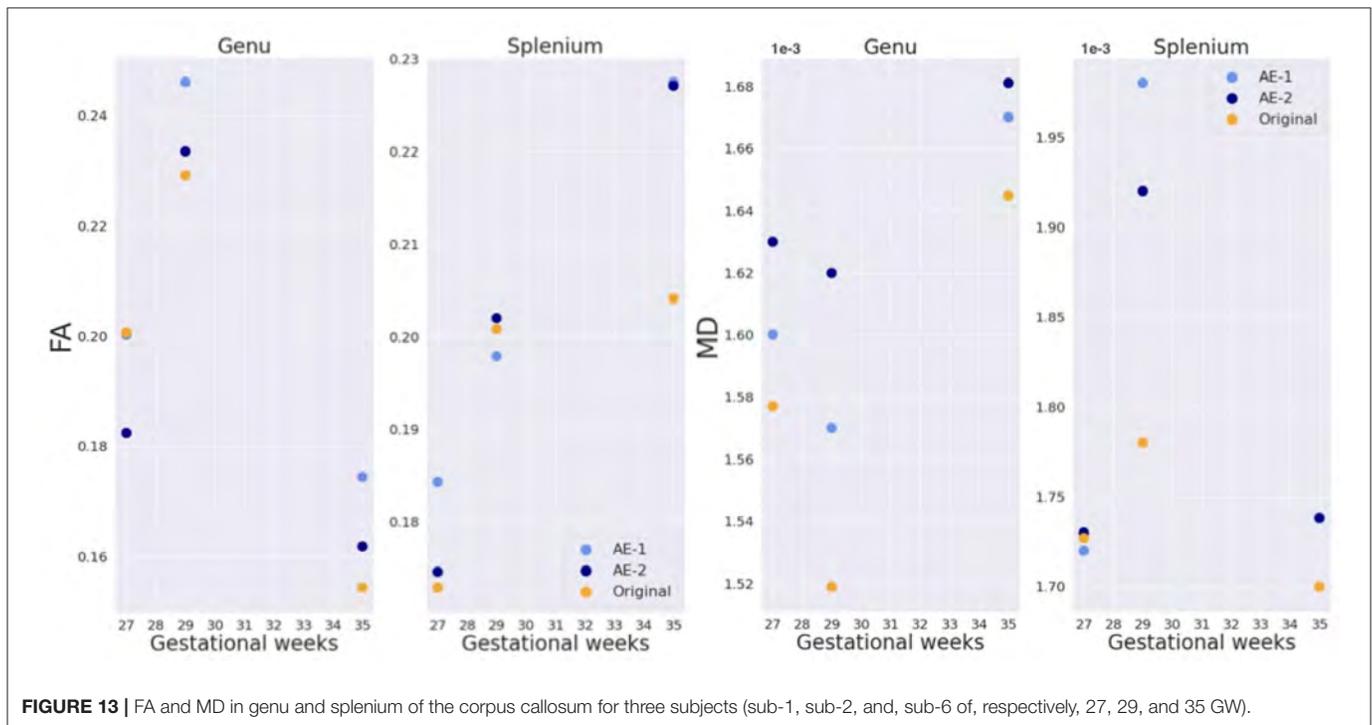
The autoencoder trained on pre-term b0 images was able to coherently enhance fetal acquisitions both at b0 and DW-MRI volumes at b700. The network was able to learn low-level features that could generalize over anatomy, contrast, and b-values. Corresponding FA and colored FA for a still subject (sub-1, 35 GW) are illustrated in Figure 12 (top) where we clearly see the coherence of the two synthesized images as we go from one original slice to the next one. In fact, both the corpus callosum and the internal/external capsules follow a smooth transition between the two slices. Similarly, Figure 12 (bottom) exhibits MD and FA for a moving subject (sub-4, 23 GW) where we also notice, particularly for the MD, the smooth transition between the originally adjacent slices. FA and MD for the remaining subjects are shown in Supplementary Figure S4. Tractography on a fetal subject (sub-1, 35 GW) using both the original and autoencoder enhancement AE-1 DW-MRI is shown in Supplementary Figure S5.

The splenium and genu of the corpus callosum were only sufficiently visible in the three late GW subjects (sub-1, sub-2, and sub-6). Figure 13 shows quantitative results for FA and MD in the two structures. Both maps fall into the range of reported values in the literature (59) for the respective gestational age, for original and autoencoder enhanced volumes.

### 4. DISCUSSION

In this work, we have shown that (1) autoencoders can be used for through-plane super-resolution in diffusion MRI, (2) training on b0 images can generalize to gradient diffusion volumes of different contrasts, and (3) as a proof of concept, training on pre-term anatomy can generalize to fetal images.

In fact, we have demonstrated how autoencoders can realistically enhance the resolution of DW-MRI perinatal brain images. We have compared it to conventionally used methods such as trilinear, tricubic, and B-spline interpolations both qualitatively and quantitatively for pre-term newborns of the dHCP database. Resolution enhancement was performed at the



**FIGURE 13** | FA and MD in genu and splenium of the corpus callosum for three subjects (sub-1, sub-2, and, sub-6 of, respectively, 27, 29, and 35 GW).

diffusion signal level and the downstream benefits propagated to the DTI maps.

Additionally, our network that was solely trained on non-diffusion weighted images ( $b_0$ ) was able to generalize to a  $b1000$  contrast. In fact, the most intuitive approach is to infer  $b1000$  images using a network trained on  $b1000$ . We have indeed tried but the network did not converge for the majority of the folds. This might be due to the high variability of  $b1000$  images across directions and their inherently low SNR. However, in the 1-fold that the network converged, it slightly underperformed the network that was trained on  $b_0$  only, on both  $b1000$  pre-term and  $b700$  fetal images. Moreover, being  $b$ -value independent is a desirable property since different  $b$ -values are used in different centers, in particular for clinical fetal imaging (400, 500, 600, 700  $s/mm^2$ ) (6, 10, 12, 29, 60). In fact, the same  $b_0$  network trained on pre-term data was generalized to  $b700$  fetal images where we qualitatively show its advantage, hence supporting the utility of pre-term data for fetal imaging, such as in Karimi et al. (61), where they have used pre-term colored FA and DW-MRI fetal scans to successfully predict fetal colored FA using a convolutional neural network. Furthermore, FA and MD of the corpus callosum, which were generated using the autoencoder enhanced volumes, are in the range of values provided by a recent study (59). This is a necessary but non sufficient condition for the validity of our framework in fetal data.

Notably, our trained network was able to reduce the noise from the data by learning the main features across images for different noise levels. This can be explained by two points. First, our autoencoder was exposed to different low levels of noise (as the dHCP data was already denoised) and hence the encoded features of the latent space are ought to be noise independent. Second, generative autoencoders intrinsically yield high SNR

outputs due to the desired smoothness property of the latent space (62).

The proposed framework could be applied to correct for anisotropic voxel sizes and can be used for slice outliers recovery in case of extreme motion artifacts for example. In fact, the artificially removed middle slices in our experiments can represent corrupted slices that may need to be discarded or replaced using interpolation (36–38). Our autoencoder can hence be used to recover these damaged slices using neighboring ones.

The power of our method compared to conventional interpolations resides in two points. First, the amount of data used to predict/interpolate the middle slice. While only two slices will be used in traditional interpolation approaches, our method will in addition take advantage of the thousands of slices to which the network has been exposed and from which the important features have been learned (without any supervision) in the training phase. Second, based on the manifold hypothesis, our method performs interpolations in the learned encoding space, which is closer to the intrinsic dimensionality of the data (63), and hence all samples from that space will be closer to the true distribution of the data compared to a naive interpolation in the pixel/voxel space.

Although our network performed quantitatively better than conventional interpolation methods in pre-term subjects, its output is usually smoother and hence exhibits lesser details. This is a well-known limitation of generative autoencoders, such as variational autoencoders, and the consequence of the desirable property of making the latent space smooth (62). Generative Adversarial Networks (64) can be an interesting alternative to overcome this issue. However, they have other drawbacks as being more unstable and less straightforward to train (65)

than autoencoders. But if trained properly, they can achieve competitive results.

In this work, qualitative results only were provided on fetal DW-MRI. We are limited by the lack of ground truth in this domain, hence our results are a proof of concept. The future release of the fetal dHCP dataset will be very valuable to further develop our framework and proceed to its quantitative assessment for fetal DW-MRI.

In future work, we want to add random Rician noise in the training phase to increase the network robustness and predictive power. We also want to extend the autoencoding to the angular domain by using spherical harmonics decomposition for each 4D voxel and hence enhancing both spatial and angular resolutions (66).

Although *unsupervised* learning *via* autoencoders has been recently used in DW-MRI to cluster individuals based on their microstructural properties (67), this is to the best of our knowledge, the first *unsupervised* learning study for super-resolution enhancement in DW-MRI using autoencoders.

As diffusion fetal imaging suffers from low through-plane resolution, super-resolution using autoencoders is an appealing method to artificially but realistically overcome this caveat. This can help depict more precise diffusion properties through different models, such as DTI or ODFs, and potentially increase the detectability of fiber tracts that are relevant for the assessment of certain neurodevelopmental disorders (29).

## DATA AVAILABILITY STATEMENT

Part of the analyzed datasets were publicly available. This data can be found here: <http://www.developingconnectome.org/data-release/data-release-user-guide/>.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Cantonal Ethical Committee, Zürich. Written

informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

HK performed the technical analysis, wrote the manuscript, provided original idea, and integrated all revisions. EC-R and HK contributed to the conceptualization of the research project. EC-R, PD, GG, AJ, and HL revised the manuscript. HL helped in the data generation. PD helped in the technical analysis. GG, MK, and YA-G helped in the processing of the fetal data. YA-G and MK acknowledged the manuscript. AJ provided the fetal data. MB conceptualized, designed and supervised the research project, contributed to the manuscript and to the final revision, and provided funding. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Swiss National Science Foundation (regular project 205321-182602, Ambizione grant PZ00P2\_185814, and NCCR SYNAPSY 185897). AJ was supported by the Prof. Max Cloetta Foundation, the EMDO Foundation, the Vontobel Foundation, the Novartis Foundation for Medical-Biological Research, and the URPP Adaptive Brain Circuits in Development and Learning (AdaBD) of the University of Zürich.

## ACKNOWLEDGMENTS

We thank Athena Taymourash for the help in the acceleration of the technical analysis and Samuel Lamon (MD) for the help in the segmentation of the corpus callosum.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2022.827816/full#supplementary-material>

## REFERENCES

- Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J Magn Reson.* (2011) 213:560–70. doi: 10.1016/j.jmr.2011.09.022
- Johansen-Berg H, Behrens TE. *Diffusion MRI: From Quantitative Measurement to in vivo Neuroanatomy.* Cambridge, MA: Academic Press (2013).
- Bui T, Daire JL, Chalard F, Zaccaria I, Alberti C, Elmaleh M, et al. Microstructural development of human brain assessed in utero by diffusion tensor imaging. *Pediatr Radiol.* (2006) 36:1133–40. doi: 10.1007/s00247-006-0266-3
- Huang H, Xue R, Zhang J, Ren T, Richards LJ, Yarowsky P, et al. Anatomical characterization of human fetal brain development with diffusion tensor magnetic resonance imaging. *J Neurosci.* (2009) 29:4263–73. doi: 10.1523/JNEUROSCI.2769-08.2009
- Huang H, Vasung L. Gaining insight of fetal brain development with diffusion MRI and histology. *Int J Dev Neurosci.* (2014) 32:11–22. doi: 10.1016/j.ijdevneu.2013.06.005
- Jakab A, Kasprian G, Schwartz E, Gruber GM, Mitter C, Prayer D, et al. Disrupted developmental organization of the structural connectome in fetuses with corpus callosum agenesis. *Neuroimage.* (2015) 111:277–88. doi: 10.1016/j.neuroimage.2015.02.038
- Righini A, Bianchini E, Parazzini C, Gementi P, Ramenghi L, Baldoli C, et al. Apparent diffusion coefficient determination in normal fetal brain: a prenatal MR imaging study. *Am J Neuroradiol.* (2003) 24:799–804.
- Kasprian G, Brugger PC, Weber M, Krssák M, Krampfl E, Herold C, et al. In utero tractography of fetal white matter development. *Neuroimage.* (2008) 43:213–24. doi: 10.1016/j.neuroimage.2008.07.026
- Khan S, Vasung L, Marami B, Rollins CK, Afacan O, Ortinau CM, et al. Fetal brain growth portrayed by a spatiotemporal diffusion tensor MRI

- atlas computed from in utero images. *Neuroimage*. (2019) 185:593–608. doi: 10.1016/j.neuroimage.2018.08.030
10. Deprez M, Price A, Christiaens D, Estrin GL, Cordero-Grande L, et al. Higher order spherical harmonics reconstruction of fetal diffusion MRI with intensity correction. *IEEE Trans Med Imaging*. (2019) 39:1104–13. doi: 10.1109/TMI.2019.2943565
  11. Kim K, Habas PA, Rousseau F, Glenn OA, Barkovich AJ, Koob M, et al. Reconstruction of a geometrically correct diffusion tensor image of a moving human fetal brain. In: *Medical Imaging 2010: Image Processing*. Vol. 7623. *International Society for Optics and Photonics*. San Diego, CA (2010). p. 76231I.
  12. Fogtman M, Seshamani S, Kroenke C, Cheng X, Chapman T, Wilm J, et al. A unified approach to diffusion direction sensitive slice registration and 3-D DTI reconstruction from moving fetal brain anatomy. *IEEE Trans Med Imaging*. (2013) 33:272–89. doi: 10.1109/TMI.2013.2284014
  13. Marami B, Scherrer B, Afacan O, Erem B, Warfield SK, Gholipour A. Motion-robust diffusion-weighted brain MRI reconstruction through slice-level registration-based motion tracking. *IEEE Trans Med Imaging*. (2016) 35:2258–69. doi: 10.1109/TMI.2016.2555244
  14. Kuklisova-Murgasova M, Estrin GL, Nunes RG, Malik SJ, Rutherford MA, Rueckert D, et al. Distortion correction in fetal EPI using non-rigid registration with a Laplacian constraint. *IEEE Trans Med Imaging*. (2017) 37:12–9. doi: 10.1109/TMI.2017.2667227
  15. Rousseau F, Glenn OA, Iordanova B, Rodriguez-Carranza C, Vigneron DB, Barkovich JA, et al. Registration-based approach for reconstruction of high-resolution in utero fetal MR brain images. *Acad Radiol*. (2006) 13:1072–81. doi: 10.1016/j.acra.2006.05.003
  16. Gholipour A, Estroff JA, Warfield SK. Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain MRI. *IEEE Trans Med Imaging*. (2010) 29:1739–58. doi: 10.1109/TMI.2010.2051680
  17. Kuklisova-Murgasova M, Quaghebeur G, Rutherford MA, Hajnal JV, Schnabel JA. Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. *Med Image Anal*. (2012) 16:1550–64. doi: 10.1016/j.media.2012.07.004
  18. Tourbier S, Bresson X, Hagmann P, Thiran JP, Meuli R, Cuadra MB. An efficient total variation algorithm for super-resolution in fetal brain MRI with adaptive regularization. *Neuroimage*. (2015) 118:584–97. doi: 10.1016/j.neuroimage.2015.06.018
  19. Kainz B, Steinberger M, Wein W, Kuklisova-Murgasova M, Malamateniou C, Keraudren K, et al. Fast volume reconstruction from motion corrupted stacks of 2D slices. *IEEE Trans Med Imaging*. (2015) 34:1901–13. doi: 10.1109/TMI.2015.2415453
  20. Ebner M, Wang G, Li W, Aertsen M, Patel PA, Aughwane R, et al. An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI. *Neuroimage*. (2020) 206:116324. doi: 10.1016/j.neuroimage.2019.116324
  21. Taymourtash A, Kebiri H, Tourbier S, Schwartz E, Nenning K-H, Licandro R, et al. 4D iterative reconstruction of brain fMRI in the moving fetus. *arXiv [Preprint]*. (2021). arXiv: 2111.11394. Available online at: <https://arxiv.org/pdf/2111.11394.pdf>
  22. Kebiri H, Lajous H, Alemán-Gómez Y, Girard G, Rodríguez EC, Tourbier S, et al. Quantitative evaluation of enhanced multi-plane clinical fetal diffusion MRI with a crossing-fiber phantom. In: *International Workshop on Computational Diffusion MRI*. Strasbourg: Springer (2021). p. 12–22.
  23. Kebiri H, Lajous H, Alemán-Gómez Y, Ledoux JB, Fornari E, Jakab A, et al. *Dataset: Quantitative Evaluation of Enhanced Multi-Plane Clinical Fetal Diffusion MRI With a Crossing-Fiber Phantom*. (2021). Available online at: [https://zenodo.org/record/5153507#.YZ-uw\\_xKjRZ](https://zenodo.org/record/5153507#.YZ-uw_xKjRZ)
  24. Ning L, Setsompop K, Michailovich O, Makris N, Shenton ME, Westin CF, et al. A joint compressed-sensing and super-resolution approach for very high-resolution diffusion imaging. *Neuroimage*. (2016) 125:386–400. doi: 10.1016/j.neuroimage.2015.10.061
  25. Liao C, Chen Y, Cao X, Chen S, He H, Mani M, et al. Efficient parallel reconstruction for high resolution multishot spiral diffusion data with low rank constraint. *Magn Reson Med*. (2017) 77:1359–66. doi: 10.1002/mrm.26199
  26. Mani M, Jacob M, Kelley D, Magnotta V. Multi-shot sensitivity-encoded diffusion data recovery using structured low-rank matrix completion (MUSSELS). *Magn Reson Med*. (2017) 78:494–507. doi: 10.1002/mrm.26382
  27. Ramos-Llorden G, Ning L, Liao C, Mukhometzianov R, Michailovich O, Setsompop K, et al. High-fidelity, accelerated whole-brain submillimeter in vivo diffusion MRI using gSlider-spherical ridgelets (gSlider-SR). *Magn Reson Med*. (2020) 84:1781–95. doi: 10.1002/mrm.28232
  28. Dyrby TB, Lundell H, Burke MW, Reislev NL, Paulson OB, Pfito M, et al. Interpolation of diffusion weighted imaging datasets. *Neuroimage*. (2014) 103:202–13. doi: 10.1016/j.neuroimage.2014.09.005
  29. Jakab A, Tuura R, Kellenberger C, Scheer I. In utero diffusion tensor imaging of the fetal brain: a reproducibility study. *Neuroimage Clin*. (2017) 15:601–12. doi: 10.1016/132704
  30. Coupé P, Manjón JV, Chamberland M, Descoteaux M, Hiba B. Collaborative patch-based super-resolution for diffusion-weighted images. *Neuroimage*. (2013) 83:245–61. doi: 10.1016/j.neuroimage.2013.06.030
  31. Alexander DC, Zikic D, Ghosh A, Tanno R, Wottschel V, Zhang J, et al. Image quality transfer and applications in diffusion MRI. *Neuroimage*. (2017) 152:283–98. doi: 10.1016/j.neuroimage.2017.02.089
  32. Elsaid NM, Wu YC. Super-resolution diffusion tensor imaging using SRCNN: a feasibility study. *Annu Int Conf IEEE Eng Med Biol Soc*. (2019) 2019:2830–4. doi: 10.1109/EMBC.2019.8857125
  33. Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell*. (2015) 38:295–307. doi: 10.1109/TPAMI.2015.2439281
  34. Chatterjee S, Sciarra A, Dünwald M, Mushunuri RV, Podishetti R, Rao RN, et al. ShuffleUNet: super resolution of diffusion-weighted MRIs using deep learning. *arXiv [Preprint]*. arXiv:210212898. (2021) doi: 10.23919/EUSIPCO54536.2021.9615963
  35. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich: Springer (2015). p. 234–41.
  36. Niethammer M, Bouix S, Aja-Fernández S, Westin CF, Shenton ME. Outlier rejection for diffusion weighted imaging. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Brisbane, QLD: Springer (2007). p. 161–8.
  37. Chang LC, Jones DK, Pierpaoli C. RESTORE: robust estimation of tensors by outlier rejection. *Magn Reson Med*. (2005) 53:1088–95. doi: 10.1002/mrm.20426
  38. Andersson JL, Graham MS, Zsoldos E, Sotiropoulos SN. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *Neuroimage*. (2016) 141:556–72. doi: 10.1016/j.neuroimage.2016.06.058
  39. Sander J, de Vos BD, Išgum I. Unsupervised super-resolution: creating high-resolution medical images from low-resolution anisotropic examples. In: *Medical Imaging 2021: Image Processing*. vol. 11596. *International Society for Optics Photonics*. San Diego, CA (2021). p. 115960E.
  40. Chung H, Kim J, Yoon JH, Lee JM, Ye JC. Simultaneous super-resolution and motion artifact removal in diffusion-weighted MRI using unsupervised deep learning. *arXiv [Preprint]*. arXiv:210500240. (2021). doi: 10.48550/arXiv.2105.00240
  41. Hutter J, Tourbier JD, Price AN, Cordero-Grande L, Hughes EJ, Malik S, et al. Time-efficient and flexible design of optimized multishell HARDI diffusion. *Magn Reson Med*. (2018) 79:1276–92. doi: 10.1002/mrm.26765
  42. Bastiani M, Andersson JL, Cordero-Grande L, Murgasova M, Hutter J, Price AN, et al. Automated processing pipeline for neonatal diffusion MRI in the developing Human Connectome Project. *Neuroimage*. (2019) 185:750–63. doi: 10.1016/j.neuroimage.2018.05.064
  43. Makropoulos A, Robinson EC, Schuh A, Wright R, Fitzgibbon S, Bozek J, et al. The developing human connectome project: a minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage*. (2018) 173:88–112. doi: 10.1016/j.neuroimage.2018.01.054
  44. Makropoulos A, Gousias IS, Ledig C, Aljabar P, Serag A, Hajnal JV, et al. Automatic whole brain MRI segmentation of the developing neonatal brain. *IEEE Trans Med Imaging*. (2014) 33:1818–31. doi: 10.1109/TMI.2014.2322280
  45. Salehi SSM, Hashemi SR, Velasco-Annis C, Oualam A, Estroff JA, Erdogmus D, et al. Real-time automatic fetal brain extraction in fetal MRI by deep

- learning. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. Washington, DC: IEEE (2018). p. 720–4.
46. Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front Neuroinform.* (2011) 5:13. doi: 10.3389/fninf.2011.00013
  47. Veraart J, Fieremans E, Novikov DS. Diffusion MRI noise mapping using random matrix theory. *Magn Reson Med.* (2016) 76:1582–93. doi: 10.1002/mrm.26059
  48. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging.* (2010) 29:1310–20. doi: 10.1109/TMI.2010.2046908
  49. Avants BB, Tustison N, Song G. et al. Advanced normalization tools (ANTS). *Insight J.* (2009) 2:1–35. doi: 10.54294/uvnhn
  50. Modat M, Cash DM, Daga P, Winston GP, Duncan JS, Ourselin S. Global image registration using a symmetric block-matching approach. *J Med Imaging.* (2014) 1:024003. doi: 10.1117/1.JMI.1.2.024003
  51. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *International Conference on Machine Learning*. Lille: PMLR (2015). p. 448–56.
  52. Chollet F, et al. keras. (2015).
  53. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings*. Sardinia (2010). p. 249–56.
  54. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv [Preprint]* arXiv:1412.6980. (2014). doi: 10.48550/arXiv.1412.6980
  55. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: a system for large-scale machine learning. In: *12th USENIX Symposium on Operating Systems Design and Implementation (SOSDI 16)*. Savannah, GA (2016). p. 265–83.
  56. Tournier JD, Calamante F, Connelly A. MRtrix: diffusion tractography in crossing fiber regions. *Int J Imaging Syst Technol.* (2012) 22:53–66. doi: 10.1002/ima.22005
  57. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. *Magn Reson Med.* (1995) 34:910–4. doi: 10.1002/mrm.1910340618
  58. Garyfallidis E, Brett M, Amirbekian B, Rokem A, Van Der Walt S, Descoteaux M, et al. Dipy, a library for the analysis of diffusion MRI data. *Front Neuroinform.* (2014) 8:8. doi: 10.3389/fninf.2014.00008
  59. Wilson S, Pietsch M, Cordero-Grande L, Price AN, Hutter J, Xiao J, et al. Development of human white matter pathways in utero over the second and third trimester. *Proc Natl Acad Sci USA.* (2021) 118:e2023598118. doi: 10.1073/pnas.2023598118
  60. Marami B, Salehi SSM, Afacan O, Scherrer B, Rollins CK, Yang E, et al. Temporal slice registration and robust diffusion-tensor reconstruction for improved fetal brain structural connectivity analysis. *Neuroimage.* (2017) 156:475–88. doi: 10.1016/j.neuroimage.2017.04.033
  61. Karimi D, Jaimes C, Machado-Rivas F, Vasung L, Khan S, Warfield SK, et al. Deep learning-based parameter estimation in fetal diffusion-weighted MRI. *Neuroimage.* (2021) 243:118482. doi: 10.1016/j.neuroimage.2021.118482
  62. Berthelot D, Raffel C, Roy A, Goodfellow I. Understanding and improving interpolation in autoencoders via an adversarial regularizer. *arXiv [preprint]*. arXiv:180707543. (2018). doi: 10.48550/arXiv.1807.07543
  63. Chollet F. *Deep learning With Python*. Simon and Schuster (2017).
  64. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. *Adv Neural Inf Process Syst.* (2014) 27.
  65. Mescheder L, Geiger A, Nowozin S. Which training methods for GANs do actually converge? In: *International Conference on Machine Learning*. Stockholm: PMLR (2018). p. 3481–90.
  66. Ma J, Cui H. Hybrid graph convolutional neural networks for super resolution of DW images. In: *Computational Diffusion MRI*. Strasbourg: Springer (2021). p. 201–12.
  67. Chamberland M, Genc S, Tax CM, Shastin D, Koller K, Raven EP, et al. Detecting microstructural deviations in individuals with deep diffusion MRI tractometry. *Nat Computat Sci.* (2021) 1:598–606. doi: 10.1038/s43588-021-00126-8

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Kebiri, Canales-Rodríguez, Lajous, de Dumast, Girard, Alemán-Gómez, Koob, Jakab and Bach Cuadra. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Supplementary Material

## 1 SUPPLEMENTARY TABLES AND FIGURES

### REFERENCES

- Christiaens, D., Cordero-Grande, L., Price, A. N., Hutter, J., Hughes, E., Counsell, S. J., et al. (2019). Fetal diffusion mri acquisition and analysis in the developing human connectome project. In *ISMRM Annual Meeting & Exhibition, Date: 2019/05/11-2019/05/16, Location: Montréal, QC, Canada*. vol. 27
- Deprez, M., Price, A., Christiaens, D., Estrin, G. L., Cordero-Grande, L., et al. (2019). Higher order spherical harmonics reconstruction of fetal diffusion mri with intensity correction. *IEEE Transactions on Medical Imaging* 39, 1104–1113
- Fogtmann, M., Seshamani, S., Kroenke, C., Cheng, X., Chapman, T., Wilm, J., et al. (2013). A unified approach to diffusion direction sensitive slice registration and 3-d dti reconstruction from moving fetal brain anatomy. *IEEE transactions on medical imaging* 33, 272–289
- Jakab, A., Kasprian, G., Schwartz, E., Gruber, G. M., Mitter, C., Prayer, D., et al. (2015). Disrupted developmental organization of the structural connectome in fetuses with corpus callosum agenesis. *Neuroimage* 111, 277–288
- Jakab, A., Tuura, R., Kellenberger, C., and Scheer, I. (2017). In utero diffusion tensor imaging of the fetal brain: a reproducibility study. *NeuroImage: Clinical* 15
- Marami, B., Salehi, S. S. M., Afacan, O., Scherrer, B., Rollins, C. K., Yang, E., et al. (2017). Temporal slice registration and robust diffusion-tensor reconstruction for improved fetal brain structural connectivity

	Field strength & brand	#Directions	B-values (s/mm <sup>2</sup> )	Orthogonal stacks	Resolution (mm <sup>3</sup> )	Acquisition time in minutes:seconds	Main sequence parameters (TE/TR in ms)	Year
Jakab et al., <i>Neuroimage: Clinical</i> 2017	1.5T/3T GE	15	700	3 axial and 1-2 coronal	1x1x3-5	1:21 per 4D volume	TE = min(65) TR = 2200	2020
Deprez et al., <i>IEEE Transactions on Medical Imaging</i> 2019	3T Phillips	32	750	?	2x2x3.5	~5:00	TE = 75 TR = 7500	2020
Christiaens et al., <i>ISMRM 2019 (DHCP)</i>	3T Phillips	80	400 1000	?	2x2x2	~15:00 (whole dataset)	TE = 75 TR = 6100	2019
Marami et al., <i>Neuroimage</i> 2017	3T Siemens	12	500	2-8 axial/coronal	2x2x2-4	0:50-1:30 per 4D volume	TE = 60 TR = 3000/4000	2017
Jakab et al., <i>Neuroimage</i> 2015	1.5T Phillips	15	700	Axial	0.94x0.94x3.3	?	TE = 90 TR = 1745	2015
Fogtmann et al., <i>IEEE Transactions on Medical Imaging</i> 2014	1.5T Siemens	10	500 600 700	2 axial 2 sagittal 0-2 coronal	~ 2x2x2	00:10 per 4D volume	TE = 96 TR = 4000-6000	2014

**Table S1.** Main settings of acquisition protocols in fetal brain diffusion MRI (Jakab et al., 2017; Deprez et al., 2019; Christiaens et al., 2019; Marami et al., 2017; Jakab et al., 2015; Fogtmann et al., 2013)

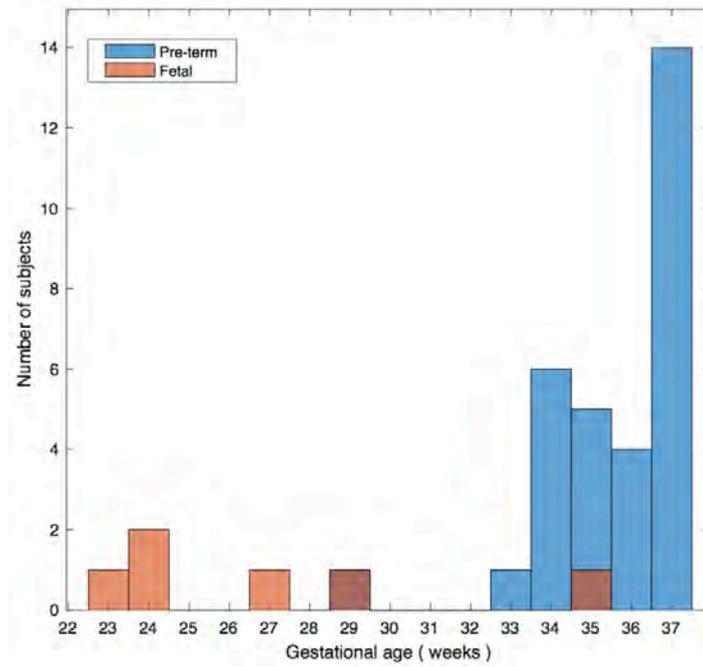


Figure S1: Histogram of the gestational age of pre-term and fetal subjects

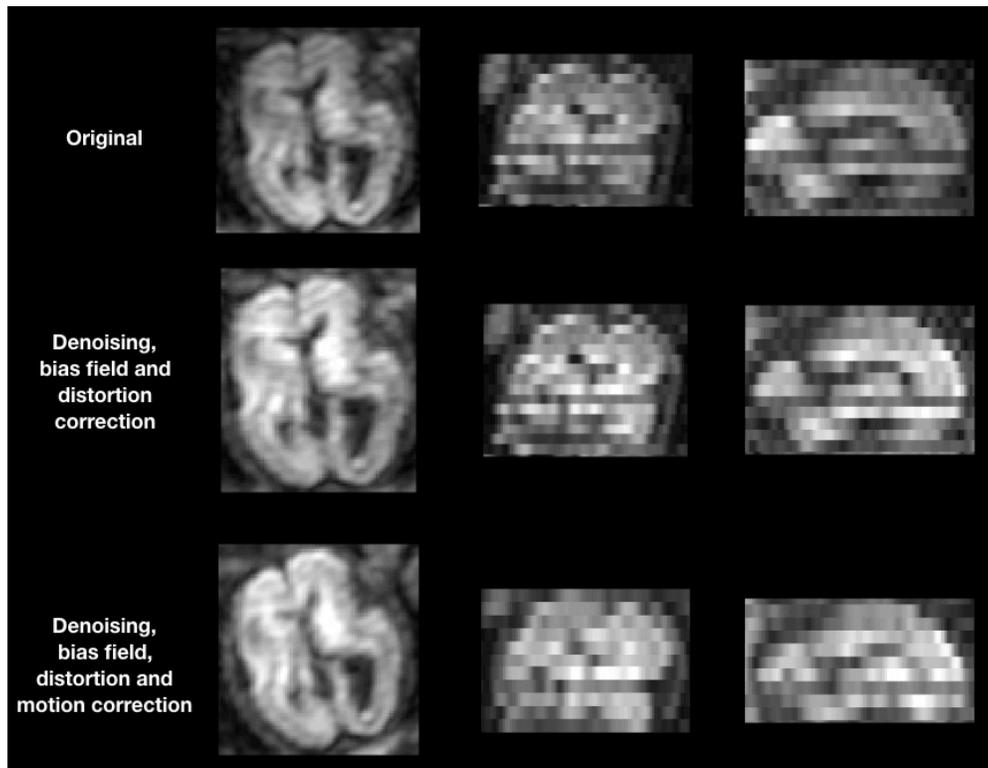


Figure S2: DWI volume of original, pre-processed and motion corrected subject (sub-4, 23 GW)

---

Subject ID

---

sub-CC00788XX22  
sub-CC00764BN14  
sub-CC00764AN14  
sub-CC00760XX10  
sub-CC00735XX18  
sub-CC00703XX10  
sub-CC00670XX11  
sub-CC00627XX17  
sub-CC00571AN11  
sub-CC00570XX10  
sub-CC00563XX11  
sub-CC00492BN15  
sub-CC00492AN15  
sub-CC00423XX11  
sub-CC00351XX05  
sub-CC00293BN14  
sub-CC00293AN14  
sub-CC00281AN10  
sub-CC00238BN16  
sub-CC00238AN16  
sub-CC00231XX09  
sub-CC00216AN10  
sub-CC00177XX13  
sub-CC00161XX05  
sub-CC00147XX16  
sub-CC00132XX09  
sub-CC00129BN14  
sub-CC00129AN14  
sub-CC00124XX09  
sub-CC00087BN14  
sub-CC00063AN06

**Table S2.** IDs of the pre-term subjects used from the developing Human Connectome Project (dHCP) dataset.

analysis. *NeuroImage* 156, 475–488

Tournier, J. D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., et al. (2019). MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage* 202, 116137. doi:10.1016/j.neuroimage.2019.116137

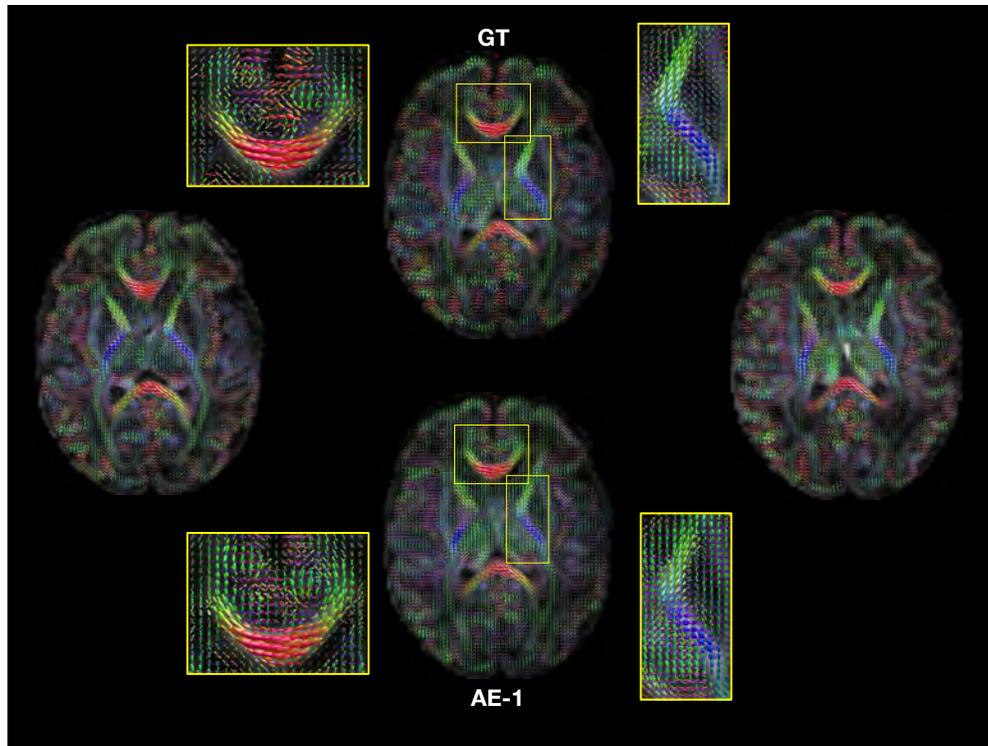


Figure S3: Orientation Distribution Functions (ODFs) of the GT and the autoencoder generated slice (AE-1) overlaid on top of Fractional Anisotropy (FA) map where we can see the right internal capsule and the corpus callosum in zoomed panels. Little qualitative difference can be observed. The left and right slices represent the two adjacent axial slices.

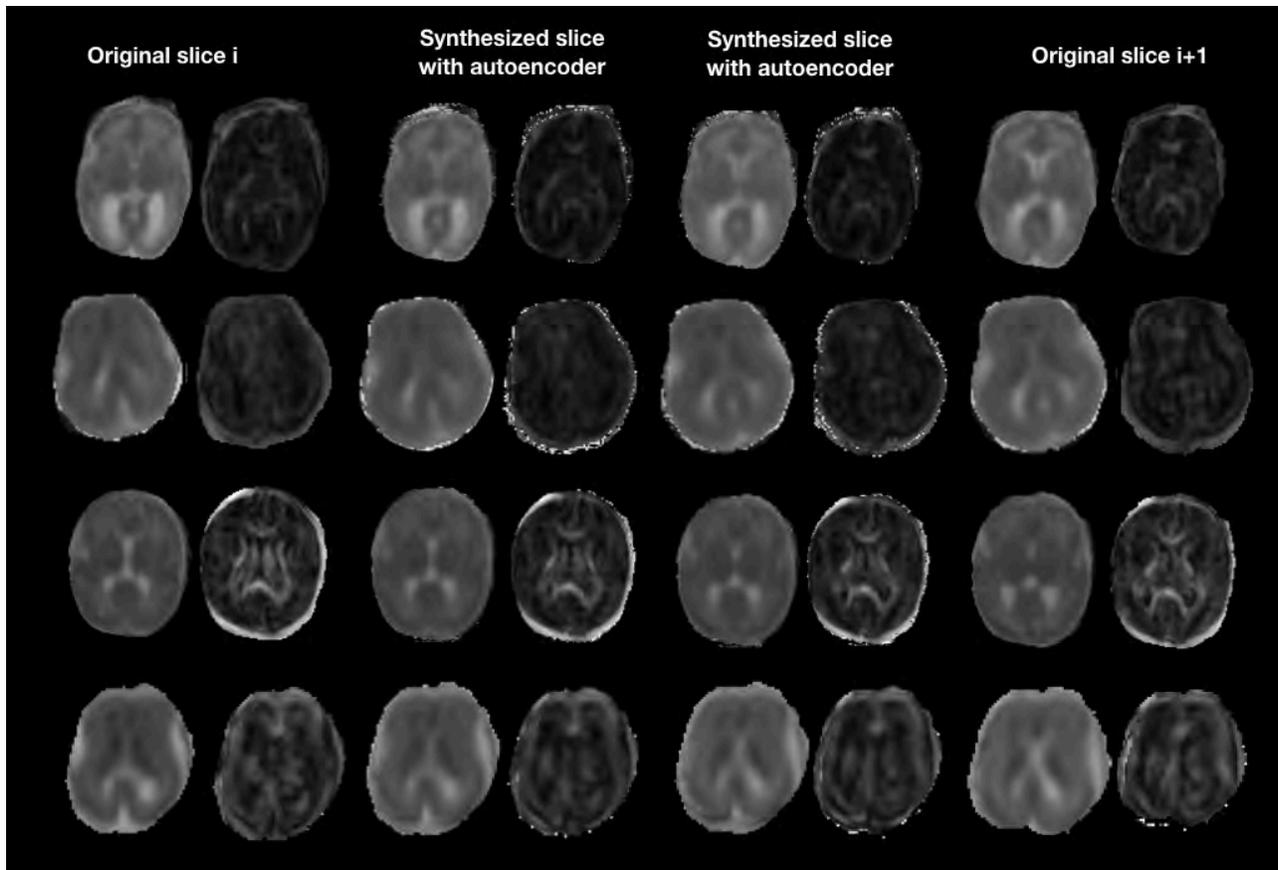


Figure S4: MD and FA illustration of AE enhancement between two adjacent fetal slices in two moving subject (top rows: sub-6, 27 GW, sub-5, 24 GW) and two still subjects (bottom rows: sub-1, 35 GW, sub-3, 24 GW).

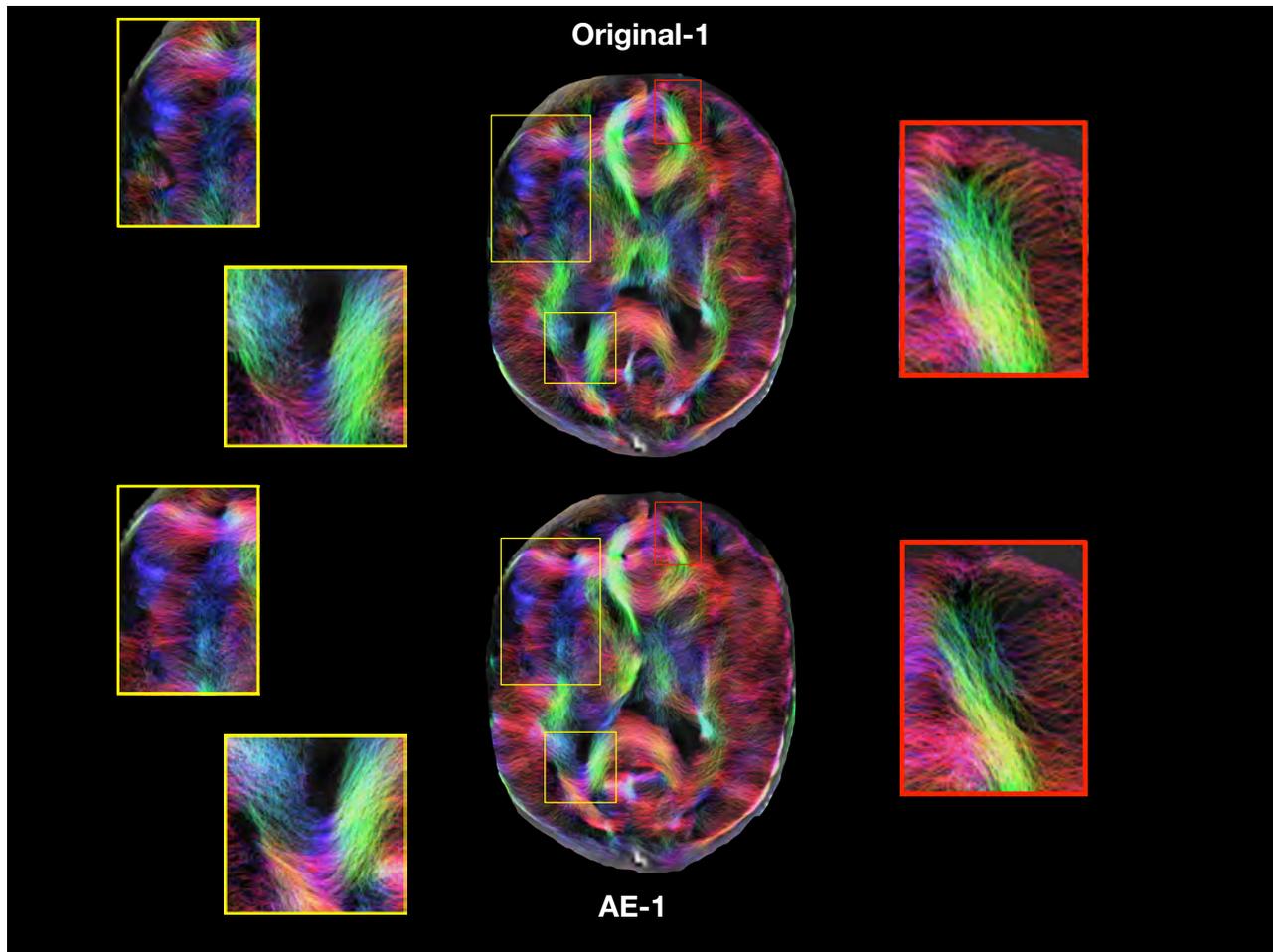


Figure S5: Tractography of a fetus (sub-1, 35 GW) generated using ODFs with spherical harmonics order 4, overlaid over the FA map. The streamlines were obtained using the MRtrix3 iFOD2 algorithm (Tournier et al., 2019) for identical tractography parameters for both the original and autoencoder enhanced data. The figure shows the streamline segments (2.5 mm) intersecting the slice using the red (left-right), green (anterior-posterior) and blue (superior-inferior) coloring. The bottom part depicts the segment for an autoencoder generated slice (AE-1) and the top part its equivalent in original stack. Yellow boxes show white matter pathways with improved density for AE-1, while the red box show region with reduced density. Overall, tractography using AE-1 recovers more locally coherent white matter structures.

# Slice estimation in diffusion MRI of neonatal and fetal brains in image and spherical harmonics domains using autoencoders

Hamza Kebiri<sup>1,2</sup>, Gabriel Girard<sup>1,2,3</sup>, Yasser Alemán-Gómez<sup>1</sup>, Thomas Yu<sup>3,4</sup>,  
András Jakab<sup>5,6</sup>, Erick Jorge Canales-Rodríguez<sup>3</sup>, and Meritxell Bach  
Cuadra<sup>2,1</sup>

<sup>1</sup> Department of Radiology, Lausanne University Hospital and University of  
Lausanne, Lausanne, Switzerland

<sup>2</sup> CIBM Center for Biomedical Imaging, Switzerland

<sup>3</sup> Signal Processing Laboratory 5 (LTS5), École Polytechnique Fédérale de Lausanne  
(EPFL), Lausanne, Switzerland

<sup>4</sup> Advanced Clinical Imaging Technology, Siemens Healthineers International AG,  
Lausanne, Switzerland

<sup>5</sup> Center for MR Research, University Children’s Hospital Zurich, Zurich, Switzerland

<sup>6</sup> Neuroscience Center Zurich, University of Zurich, Zurich, Switzerland

**Abstract.** Diffusion MRI (dMRI) of the developing brain can provide valuable insights into the white matter development. However, slice thickness in fetal dMRI is typically high (i.e., 3-5 mm) to freeze the in-plane motion, which reduces the sensitivity of the dMRI signal to the underlying anatomy. In this study, we aim at overcoming this problem by using autoencoders to learn unsupervised efficient representations of brain slices in a latent space, using raw dMRI signals and their spherical harmonics (SH) representation. We first learn and quantitatively validate the autoencoders on the developing Human Connectome Project pre-term newborn data, and further test the method on fetal data. Our results show that the autoencoder in the signal domain better synthesized the raw signal. Interestingly, the fractional anisotropy and, to a lesser extent, the mean diffusivity, are best recovered in missing slices by using the autoencoder trained with SH coefficients. A comparison was performed with the same maps reconstructed using an autoencoder trained with raw signals, as well as conventional interpolation methods of raw signals and SH coefficients. From these results, we conclude that the recovery of missing/corrupted slices should be performed in the signal domain if the raw signal is aimed to be recovered, and in the SH domain if diffusion tensor properties (i.e., fractional anisotropy) are targeted. Notably, the trained autoencoders were able to generalize to fetal dMRI data acquired using a much smaller number of diffusion gradients and a lower b-value, where we qualitatively show the consistency of the estimated diffusion tensor maps.

**Keywords:** Super-resolution · Autoencoders · Spherical Harmonics · Diffusion Tensor Imaging · Pre-term · Fetal · Brain · MRI

## 1 Introduction

Neonatal and fetal brain development involves complex cerebral growth and maturation both for gray and white matter [4,10]. Diffusion MRI (dMRI) has been widely employed to study this developmental process *in vivo*, including neonates and fetuses [16,18,28]. As the diffusion weighted signal is sensitive to the displacement of water molecules, several models have been proposed for estimating the underlying anatomy such as diffusion tensor imaging (DTI) or spherical deconvolution methods [2,32,6]. The accuracy of these models is dependant on the angular and spatial resolution of the acquisitions that is typically limited for the neonate and fetal subjects [19,22]. Stochastic motion and low signal-to-noise ratio (SNR) due to the small size of the developing brain often translate to degraded images with low spatial resolution. Additionally, slice thickness in fetal dMRI is typically high, varying between 3-5 mm, to freeze the in-plane motion, and hence reduces the sensitivity of the dMRI signal to the underlying anatomy. This highlights the need for methods to interpolate or synthesize new slices that were either (1) corrupted because of motion or (2) acquired using anisotropic voxel sizes. Interpolation is often performed either at scanner level or in post-processing [19], and has been demonstrated to be relevant for raw signal recovery and for subsequent analysis such as tractography [11]. Similarly, super-resolution (SR) methods that aim at increasing dMRI resolution can be applied at the acquisition-reconstruction level [27,29] or at post-processing [5,12,7]. The latter used *supervised learning* methods, which require high resolution training data that is often unavailable for the developing brain. Additionally, these methods focus on enhancing the resolution homogeneously over all dimensions and were not assessed for anisotropic voxels, commonly acquired for fetuses and neonates [19,22]. Additionally to the raw dMRI signal interpolation, other representations such as Spherical Harmonics (SH) could be of interest. SH are a combination of smooth orthogonal basis functions defined on the surface of a sphere able to represent spherical signals, such as the dMRI signal acquired using uniformly distributed gradient directions [13,15]. Previous work used deep learning methods to map the SH coefficients from one shell to another [24,20]. However, no prior work, to the best of our knowledge, relies on the SH decomposition to enhance the spatial image resolution.

In this study, we have used *unsupervised learning* to extend the application of autoencoders for through-plane super-resolution [30,21] in the image domain to spherical harmonics domain where we synthesize SH coefficients of *missing* slices. As such, our network has access to both angular and spatial information. In contrast to training with non-DWI volumes [21], we have additionally trained a second network on spherical averaged dMRI images to complement and compare its performance in relation to the SH trained network. Moreover, we have compared both methods to conventional interpolation methods both using raw dMRI signals and their SH representation. The comparison was performed both on the raw dMRI signal; and on fractional anisotropy (FA) and mean diffusivity (MD) maps derived from the estimated diffusion tensors. Finally, we verified

that the SH networks trained on pre-term data successfully generalized to fetal images, where we present the coherence of the synthesized slices.

## 2 Methodology

### 2.1 Materials

**Neonatal data** - The developing Human Connectome Project (dHCP) data<sup>1</sup> were acquired in a 3T Philips Achieva scanner in a multi-shell scheme ( $b \in \{0, 400, 1000, 2600\}$  s/mm<sup>2</sup>). Details on acquisition parameters can be found in [17]. The data was denoised, motion and distortion corrected [3] and has a final resolution of  $1.17 \times 1.17 \times 1.5$  mm<sup>3</sup> in a FOV of  $128 \times 128 \times 64$  mm<sup>3</sup>. In addition to  $b = 0$  s/mm<sup>2</sup> images (b0), we have selected the corresponding 88 volumes with  $b = 1000$  s/mm<sup>2</sup> (b1000) from all pre-term subjects (31) defined with less than 37 gestational weeks (GW) ([29.3, 37.0], mean=35.5). In the anatomical dataset, brain tissue labels and masks [26] were provided.

**Fetal data** - The fetal data were acquired with the approval of the ethics committee. Acquisitions were performed at 1.5T (GE Healthcare) with a single shot echo planar imaging sequence (TE=63 ms, TR=2200 ms) using  $b = 700$  s/mm<sup>2</sup> (b700) and 15 directions. The acquisition FOV was  $256 \times 256 \times 14 - 22$  mm<sup>3</sup> for a resolution of  $1 \times 1 \times 4 - 5$  mm<sup>3</sup>. Three axial and one coronal acquisitions were performed for each subject. Four subjects were used in our study: two of 35 and 29 GW where three axial volumes were used, and two young subjects of 24 GW where one axial volume was used. We have only used axial acquisitions to avoid any confounding factor due to interpolation in the registration that would be needed between the orthogonal orientations. Volumes were corrected for noise [34], bias-field inhomogeneities [33] and distortions [25,1] and did not require any motion correction.

### 2.2 Model

**Network architecture** - Our network is composed of four blocks in the encoder and four blocks in the decoder, where each block consists of two layers of  $3 \times 3$  convolutions, a batch normalization and an Exponential Linear Unit (ELU) activation function [9]. After each block of the encoder, a  $2 \times 2$  average pooling operation was performed and the number of feature maps was doubled after each layer. Hence starting from 32 feature maps to 256 while three additional  $3 \times 3$  convolutions were added in the last block with 512, 256 and  $M$  feature maps respectively,  $M \in \{16, 32, 64, 128\}$ . The last  $M$  feature maps were considered as the latent space of our autoencoder. The decoder goes back to original input dimensions by means of either  $3 \times 3$  transposed convolutions with strides of 2 or by  $2 \times 2$  nearest neighbor interpolations (mutually exclusive), where the number of feature maps decreases by two after each layer from 512 to 32. A last

<sup>1</sup> <http://www.developingconnectome.org/data-release/second-data-release/>

1x1 convolution with sigmoid activation function was performed to generate the predicted image.

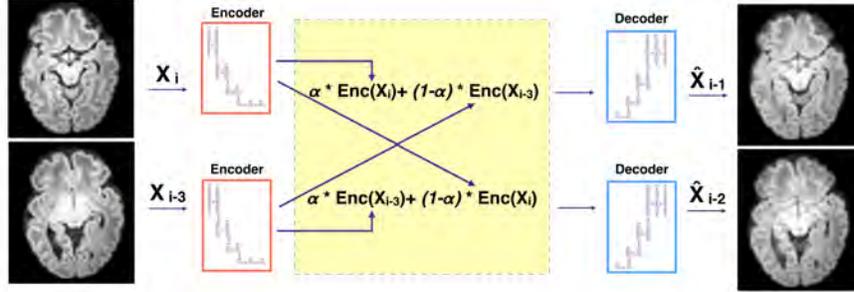
**Training** - Using the same architecture, we have trained three networks, with different inputs: b0 images (*b0-net*), average b1000 (*Avg-b1000-net*) (see *Raw signal networks* subsection) and a maximum SH order ( $L_{max}$ ) of 4 (*SH4-net*) (see *Spherical harmonics networks* subsection). Input images were first normalized to the range  $[0, 1]$  by  $x = \frac{x - x_{min}}{x_{max} - x_{min}}$  where  $x_{min}$  and  $x_{max}$  are the minimum and maximum intensities respectively in a given slice. All networks were trained using an Nvidia GeForce RTX 3090 GPU in the TensorFlow framework (version 2.4.1) with Adam optimizer [23] for 200 epochs using mean squared error loss function, a batch size of 32 and a learning rate of  $5 \times 10^{-5}$ . The validation was performed on 15% of the training data. The number of feature maps of the latent space was optimized using Keras-tuner [8] and the checkpoint with the minimal validation loss was finally selected for inference.

**Raw signal networks** - While *b0-net* was trained using b0 images, *Avg-b1000-net* was trained on average b1000 images, as training directly on individual b1000 images did not consistently converge [21]. We have thus trained *Avg-b1000-net* on average b1000 images with the aim of increasing the SNR and reducing variability. The average was computed over  $n$  randomly selected volumes,  $n \in \{3, 6, 15, 30, 40\}$ . Empirically, higher  $n$  means a lower risk of network divergence, at the cost of increased smoothness/risk of losing image detail. Therefore  $n$  must be tuned. In the end, *b0-net* was used to infer b0 images whereas *Avg-b1000-net* was used to infer b1000 volumes.

**Spherical harmonics network** - We have fit SH representations by using  $L_{max}=4$  to the dMRI signal using Dipy [14] and fed the resulting 15 SH coefficients, slice by slice, to *SH4-net*. Let us note that we preliminary computed the mean squared error difference with respect to the ground truth data when estimating SH and projecting back to original grid from SH bases of  $L_{max} \in \{4, 6, 8\}$ . As differences were relatively low between them (9.80, 8.64 and 9.95 for  $L_{max} \in \{4, 6, 8\}$  respectively, scale  $\times 10^{-4}$ ) and we aim at further testing on fetal data (where only 15 DWI are available) we selected to stick in what follows to  $L_{max}=4$ .

**Inference in neonates** - For all networks (*b0-net*, *Avg-b1000-net* and *SH4-net*), nested cross validation was performed where the 31 subjects were split into 8 folds. For each subject and each volume in the testing set, we removed  $N$  intermediate slices,  $N \in \{1, 2\}$  that were considered as the ground truth we aim to predict. Using the two adjacent slices, we input each separately to the encoder part of the network to get the  $M$  latent feature maps. These feature maps were averaged using an equal weighting for  $N = 1$  and a  $\{\frac{1}{3}, \frac{2}{3}\}$ ,  $\{\frac{2}{3}, \frac{1}{3}\}$  weighting for  $N = 2$  (Figure 1). The missing slices were then recovered by using the decoder part from the resulting latent feature maps. The output of the network was then mapped back to the range of input intensities. This was performed using histogram matching (using cumulative probability distributions) between the network output as a source image and the (weighted) average of the two adjacent input slices as a reference image. Finally, the histogram matched output of *SH4-*

*net* was projected back to the original grid of 88 directions to recover the dMRI signal in the image domain.



**Fig. 1.** Inference for two adjacent slices of the first coefficient of  $SH\text{-}L_{max}$  order 4 illustrated for the case of  $N = 2$  where  $\alpha = \frac{2}{3}$ .

**Evaluation in neonates** - The inferred slices of *Avg-b1000-net* were compared to conventional interpolations, namely trilinear, tricubic and B-spline of 5<sup>th</sup> order [31,1]. The comparison was performed separately for one and two missing slices ( $N \in \{1, 2\}$ ) using the mean squared error (MSE). As all interpolation baselines produce similar results with a slight overperformance for the linear method (for  $N = 2$ , MSE of 0.003164, 0.003204 and 0.003211 for linear, cubic and B-spline respectively), the former was chosen for further comparison with autoencoders. The two networks were additionally compared for FA and MD maps that were extracted from the diffusion tensors, as estimated in Dipy [14]. The DTI fit used the synthesized  $b_0$  by *b0-net*. The linear baseline was further compared with *SH4-net* and with the signal recovered from the same interpolation of the SH coefficients. The comparison was also extended for DTI maps (FA, MD). To compute them, DTI fit of *SH4-net* relied on the  $b_0$  as synthesized by *b0-net*, and the linear SH4 used corresponding linear interpolated  $b_0$ . All comparisons were done using MSE for FA and MD maps in white matter, cortical gray matter, and corpus callosum. Moreover, we have fit SH representations of the ground truth signal by using  $L_{max} = 4$  which were compared after projecting back to the original grid of 88 gradient unit vectors to the original DWI signal, separately for ( $N \in \{1, 2\}$ ). This was considered as the lower bound error of *SH4-net*.

**Application to fetal DWI** - After fitting the SH coefficients with  $L_{max}=4$  to the fetal data. We have used *SH4-net*, i.e., trained on pre-term neonates to infer SH coefficients of middle ( $N \in \{1, 2\}$ ) slices of fetal subjects. The inference was performed in a similar manner as for neonates (Figure 1). Cropping of fetal images to  $128 \times 128$  voxels was necessary before feeding them to the encoder. Then, we generated the diffusion tensor based on this new DWI signal and  $b_0$  using *b0-net*, and visually assessed the consistency of the new slices in MD and

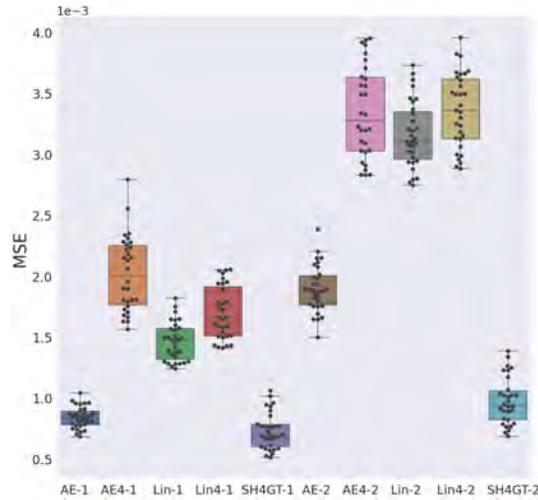
FA maps for the four subjects. Only qualitative evaluation was performed for fetal enhancement because of the lack of ground truth.

### 3 Results

Based on the validation loss, the optimal number of feature maps in the latent space was found to be 32 for *b0-net* and *Avg-b1000-net*, and 64 for *SH4-net*. For *Avg-b1000-net*, averaging  $n = 15$  DWI was also found to be optimal. Moreover, the transposed convolution in the decoder did not reduce the validation loss as compared to performing a nearest neighbor interpolation. Hence all networks used the latter in the decoder part to avoid unnecessary overparameterization of the network.

#### 3.1 DWI assessment

Autoencoder average b1000 trained network (*Avg-b1000-net*) produces superior results compared to linear interpolation (Figure 2). The difference is higher for the case of two slices removed ( $N = 2$ ).



**Fig. 2.** Mean squared error (MSE) on dMRI images of autoencoder enhanced using *Avg-b1000-net* slices (AE-1, AE-2 for  $N = 1, 2$  respectively) and for the baseline interpolation (linear on raw signal: Lin-1, Lin-2) and for *SH4-net* and SH linearly interpolated (Lin4-1, Lin4-2 for  $N = 1, 2$  respectively). The lower bounds for the SH errors (SH4GT) were also included as a reference. (Method-1, Method-2 for synthesizing/interpolating  $N = 1$  and  $N = 2$  slices, respectively)

Comparing raw and SH domain enhancement (Figure 2), we first observe that independently of the method (autoencoder or linear), working directly on

the raw signal outperforms working on SH and projecting back to signal. In fact, autoencoder *Avg-b1000-net* outperforms linear interpolation, and for  $N = 1$  it is closely comparable to the SH encoding (SH4GT-1 in Figure 2). While the SH autoencoder enhancement underperforms the classical SH linear interpolation for  $N = 1$ , *SH4-net* slightly outperforms linear-SH for  $N = 2$ . This gap between  $N = 1$  and  $N = 2$  for SH linear and autoencoder can be explained by the rich information that the autoencoder was exposed to in the training phase from similar images compared to the interpolation that has solely access to local information.

### 3.2 FA and MD in newborns

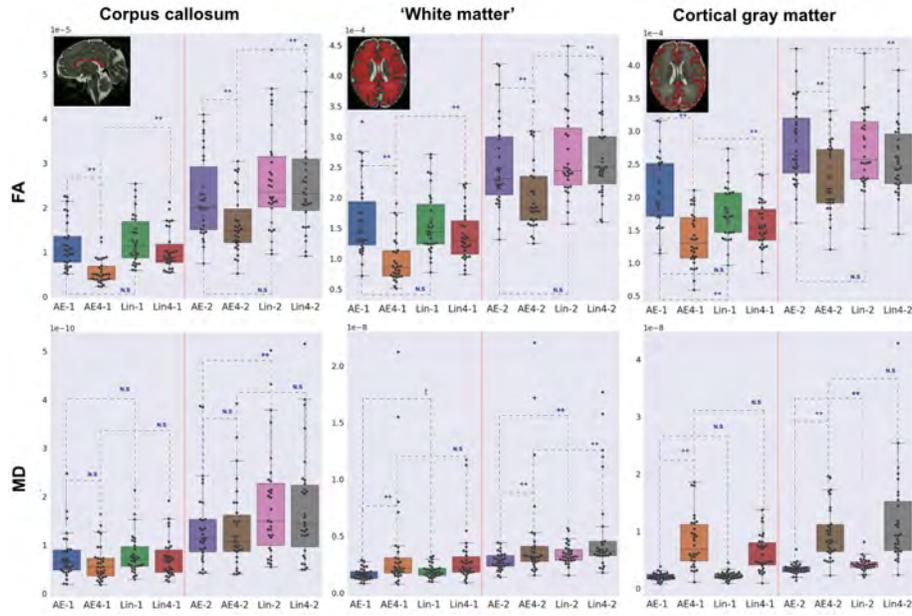
Comparing DTI scalar maps (Figure 3) for the same previous configurations (see Figure 2), we notice that the autoencoder enhancement outperforms the linear interpolation in all brain regions (except MD for cortical gray matter when removing one slice, i.e.  $N = 1$ ) regardless of whether raw signal or SH was used. This outperformance is significant (paired Wilcoxon signed-rank test) for FA in all SH configurations, and for MD in one third of all configurations. The difference is typically more pronounced when we remove two slices ( $N = 2$ ). Let us note that, opposite of what we observed at the DWI signal level, *SH4-net* outperforms linearly interpolated SH. Furthermore, for the FA map, *SH4-net* obtains the lowest mean squared errors, thus it is more suitable than autoencoder *Avg-b1000-net* or the linear interpolation. The opposite trend, i.e. *Avg-b1000-net* outperforming *SH4-net* with statistical significance, can be noticed for MD, with exception of the corpus callosum.

### 3.3 Qualitative results of FA and MD in fetuses

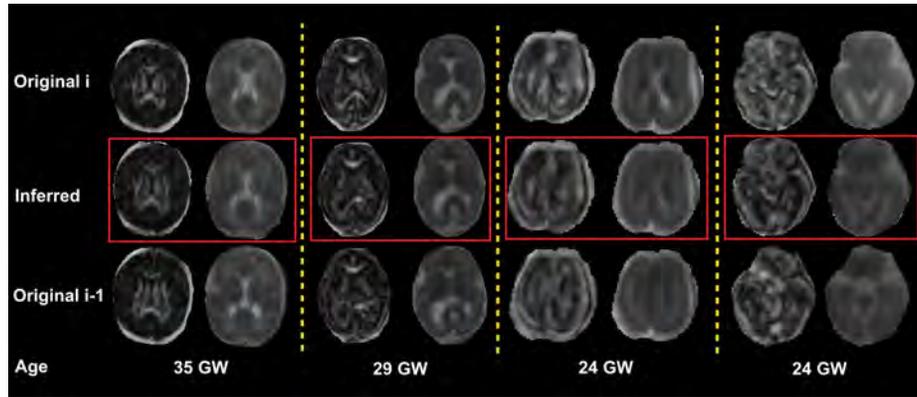
The DWIs synthesized by *SH4-net* using the latent space were visually consistent as they smoothly vary between the adjacent slices. Figure 4 displays the corresponding FA and MD maps for four subjects. We can clearly delineate the smooth transition between the two adjacent slices, especially in late gestational weeks fetuses in which the structures are more visible. For instance, the corpus callosum and the internal capsules of the synthesized slices displayed in FA maps are coherent with respect to their neighbouring slices.

## 4 Conclusion

We have proposed autoencoders for dMRI through-plane slice inference in early brain development. The assessment was performed in both raw signal and spherical harmonics (SH) domains, where the latter proved to be more accurate for DTI-FA maps reconstruction and the former for raw data estimation. We hypothesize that this could be explained by some global bias introduced to the back projected raw signal by the SH trained autoencoder. However, the orientation information (i.e., signal’s shape) was better preserved and hence, FA which



**Fig. 3.** Mean squared error of fractional anisotropy (FA) and mean diffusivity (MD) for different methods in three brain regions. See caption Figure 2 for methods description. (Paired Wilcoxon signed-rank test: \*: significant,  $p < 0.028$  - t: trending,  $p = 0.06$  - N.S.: non significant:  $p > 0.06$ )



**Fig. 4.** Fractional anisotropy (FA) and mean diffusivity (MD) for four fetal subjects of respectively, from left to right, 4, 5, 4 and 4 mm of slice thickness. The middle row (red frames) illustrates synthesized slices corresponding to the diffusion tensor reconstructed with inferred DWI volumes with *SH4-net* and  $b_0$  with *b0-net*, using the two neighboring original slices (top and bottom rows).

is scale invariant, was clearly better depicted by SH autoencoder estimation. Lastly, we have successfully applied our method trained on newborn data to enhance the through-plane resolution of fetal data acquired in a different scanner, with a lower b-value and fewer gradient directions. Inferring missing slices or realistically increasing the through-plane resolution has the potential to translate to more accurate diffusion properties and hence a better uncovering of the underlying brain structure. In future work, we aim to increase the angular resolution in fetal images by using *supervised learning* to map spherical harmonics coefficients of order 4 (i.e., the maximal order that can be fit with clinical fetal images) to higher orders (6 or 8) using pre-term data.

## 5 Acknowledgments

This work was supported by the Swiss National Science Foundation (project 205321-182602, grant No 185897: NCCR-SYNAPSY- "The synaptic bases of mental diseases" and the Ambizione grant PZ00P2\_185814). We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), École polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE), Geneva University Hospitals (HUG) and the Leenaards and Jeantet Foundations.

## References

1. Avants, B.B., Tustison, N., Song, G., et al.: Advanced normalization tools (ants). *Insight j* **2**(365), 1–35 (2009)
2. Basser, P.J., Mattiello, J., LeBihan, D.: Mr diffusion tensor spectroscopy and imaging. *Biophysical journal* **66**(1), 259–267 (1994)
3. Bastiani, M., Andersson, J.L., Cordero-Grande, L., Murgasova, M., Hutter, J., Price, A.N., , et al.: Automated processing pipeline for neonatal diffusion mri in the developing human connectome project. *NeuroImage* **185**, 750–763 (2019)
4. Batalle, D., Edwards, A.D., O’Muircheartaigh, J.: Annual research review: not just a small adult brain: understanding later neurodevelopment through imaging the neonatal brain. *Journal of Child Psychology and Psychiatry* **59**(4), 350–371 (2018)
5. Blumberg, S.B., Tanno, R., Kokkinos, I., Alexander, D.C.: Deeper image quality transfer: Training low-memory neural networks for 3d images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 118–125. Springer (2018)
6. Canales-Rodríguez, E.J., Legarreta, J.H., Pizzolato, M., Rensonnet, G., Girard, G., Patino, J.R., et al.: Sparse wars: A survey and comparative study of spherical deconvolution algorithms for diffusion MRI. *NeuroImage* **184**, 140–160 (2019)
7. Chatterjee, S., Sciarra, A., Dünnwald, M., Mushunuri, R.V., Podishetti, R., Rao, R.N., et al.: Shuffleunet: Super resolution of diffusion-weighted mris using deep learning. *arXiv preprint arXiv:2102.12898* (2021)
8. Chollet, F., et al.: *keras* (2015)
9. Clevert, D.A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289* (2015)
10. Dubois, J., Alison, M., Counsell, S.J., Hertz-Pannier, L., Hüppi, P.S., et al.: Mri of the neonatal brain: a review of methodological challenges and neuroscientific advances. *Journal of Magnetic Resonance Imaging* **53**(5), 1318–1343 (2021)
11. Dyrby, T.B., Lundell, H., Burke, M.W., Reislev, N.L., Paulson, O.B., Ptito, M., Siebner, H.R.: Interpolation of diffusion weighted imaging datasets. *NeuroImage* **103**, 202–213 (2014)
12. Elsaid, N.M., Wu, Y.C.: Super-resolution diffusion tensor imaging using srcnn: A feasibility study. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. pp. 2830–2834. IEEE (2019)
13. Frank, L.R.: Characterization of anisotropy in high angular resolution diffusion-weighted mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **47**(6), 1083–1099 (2002)
14. Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M., Nimmo-Smith, I.: Dipy, a library for the analysis of diffusion mri data. *Frontiers in neuroinformatics* **8** (2014)
15. Hess, C.P., Mukherjee, P., Han, E.T., Xu, D., Vigneron, D.B.: Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **56**(1), 104–117 (2006)
16. Hüppi, P.S., Dubois, J.: Diffusion tensor imaging of brain development. In: *Seminars in Fetal and Neonatal Medicine*. vol. 11, pp. 489–497. Elsevier (2006)
17. Hutter, J., Tournier, J.D., Price, A.N., Cordero-Grande, L., Hughes, E.J., Malik, S., Steinweg, J., et al.: Time-efficient and flexible design of optimized multishell hardi diffusion. *Magnetic resonance in medicine* **79**(3), 1276–1292 (2018)

18. Jakab, A., Kasprian, G., Schwartz, E., Gruber, G.M., Mitter, C., Prayer, D., Schöpf, V., Langs, G.: Disrupted developmental organization of the structural connectome in fetuses with corpus callosum agenesis. *Neuroimage* **111** (2015)
19. Jakab, A., Tuura, R., Kellenberger, C., Scheer, I.: In utero diffusion tensor imaging of the fetal brain: a reproducibility study. *NeuroImage: Clinical* **15** (2017)
20. Jha, R.R., Nigam, A., Bhavsar, A., Pathak, S.K., et al.: Multi-shell d-mri reconstruction via residual learning utilizing encoder-decoder network with attention (msr-net). In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 1709–1713. IEEE (2020)
21. Kebiri, H., Canales-Rodríguez, E.J., Lajous, H., de Dumast, P., Girard, G., Alemán-Gómez, Y., Koob, M., Jakab, A., Bach Cuadra, M.: Through-plane super-resolution with autoencoders in diffusion magnetic resonance imaging of the developing human brain. *Frontiers in Neurology* **13** (2022)
22. Kimpton, J., Batalle, D., Barnett, M., Hughes, E., Chew, A., Falconer, S., et al.: Diffusion magnetic resonance imaging assessment of regional white matter maturation in preterm neonates. *Neuroradiology* **63**(4), 573–583 (2021)
23. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
24. Koppers, S., et al.: Diffusion mri signal augmentation: from single shell to multi shell with deep learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 61–70. Springer (2016)
25. Kuklisova-Murgasova, M., Estrin, G.L., Nunes, R.G., Malik, S.J., Rutherford, M.A., et al.: Distortion correction in fetal epi using non-rigid registration with a laplacian constraint. *IEEE Transactions on Medical Imaging* **37**(1) (2017)
26. Makropoulos, A., Gousias, I.S., Ledig, C., Aljabar, P., Serag, A., Hajnal, J.V., et al.: Automatic whole brain mri segmentation of the developing neonatal brain. *IEEE transactions on medical imaging* **33**(9), 1818–1831 (2014)
27. Ning, L., Setsompop, K., Michailovich, O., Makris, N., Shenton, M.E., Westin, C.F., Rathi, Y.: A joint compressed-sensing and super-resolution approach for very high-resolution diffusion imaging. *NeuroImage* **125**, 386–400 (2016)
28. Ouyang, M., Dubois, J., Yu, Q., Mukherjee, P., Huang, H.: Delineation of early brain development from fetuses to infants with diffusion mri and beyond. *Neuroimage* **185**, 836–850 (2019)
29. Ramos-Llordén, G., Ning, L., Liao, C., Mukhometzianov, R., Michailovich, O., Setsompop, K., Rathi, Y.: High-fidelity, accelerated whole-brain submillimeter in vivo diffusion mri using gslider-spherical ridgelets (gslider-sr). *Magnetic resonance in medicine* **84**(4), 1781–1795 (2020)
30. Sander, J., de Vos, B.D., Išgum, I.: Unsupervised super-resolution: creating high-resolution medical images from low-resolution anisotropic examples. In: *Medical Imaging 2021: Image Processing*. vol. 11596, p. 115960E. International Society for Optics and Photonics (2021)
31. Tournier, J.D., Calamante, F., Connelly, A.: Mrtrix: diffusion tractography in crossing fiber regions. *International journal of imaging systems and technology* **22**(1), 53–66 (2012)
32. Tournier, J.D., Yeh, C.H., Calamante, F., Cho, K.H., Connelly, A., Lin, C.P.: Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. *Neuroimage* **42**(2), 617–625 (2008)
33. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., et al.: N4itk: improved n3 bias correction. *IEEE transactions on medical imaging* **29**(6), 1310–1320 (2010)
34. Veraart, J., Fieremans, E., Novikov, D.S.: Diffusion mri noise mapping using random matrix theory. *Magnetic resonance in medicine* **76**(5), 1582–1593 (2016)

# Deep learning microstructure estimation of developing brains from diffusion MRI: a newborn and fetal study

Hamza Kebiri<sup>1,2,3</sup>, Ali Gholipour<sup>3</sup>, Lana Vasung<sup>4</sup>, Željka Krsnik<sup>5</sup>, Davood Karimi<sup>3,\*</sup>, and Meritxell Bach Cuadra<sup>1,2,\*</sup>

<sup>1</sup>CIBM Center for Biomedical Imaging, Switzerland

<sup>2</sup>Department of Radiology, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland

<sup>3</sup>Computational Radiology Laboratory, Department of Radiology, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts, USA

<sup>4</sup>Department of Pediatrics, Boston Children's Hospital, and Harvard Medical School, Boston, Massachusetts, USA

<sup>5</sup>Croatian Institute for Brain Research, School of Medicine, University of Zagreb, Zagreb, Croatia

\*These authors contributed equally to this work.

## ABSTRACT

Diffusion-weighted magnetic resonance imaging (dMRI) is widely used to assess the brain white matter. Fiber orientation distribution functions (FODs) are a common way of representing the orientation and density of white matter fibers. However, with standard FOD computation methods, accurate estimation of FODs requires a large number of measurements that usually cannot be obtained for newborns and fetuses. We propose to overcome this limitation by using a deep learning method to map as few as six diffusion measurements to the target FOD. To train the model, we use the FODs computed using High Angular Resolution Diffusion Imaging (HARDI) measurements as target. Extensive quantitative evaluations show that the new deep learning method, using significantly fewer measurements, achieves comparable or superior results to standard methods such as Constrained Spherical Deconvolution. Furthermore, we demonstrate the generalizability of the new method across scanners, acquisition protocols, and anatomy on two clinical datasets of newborns and fetuses. Finally, we compute agreement metrics within the HARDI newborn dataset and advocate for the need for methods that are more tailored to the developing brain.

## Introduction

Early brain growth is characterized by rapid and complex structural and functional developments that are vulnerable to various genetic and environmental factors. The influence of early brain development and disorders on the brain health later in life has received growing interest<sup>1-5</sup>. Magnetic Resonance Imaging (MRI) is a non-invasive method for assessing these developments *in vivo*. Diffusion MRI (dMRI), specifically, offers a means to assess the micro-structure of the white matter using the diffusion of water molecules as a proxy measure<sup>6,7</sup>. However, application of dMRI to study the developing brain has been limited due to motion, limited scan time, and low signal-to-noise ratio (SNR)<sup>8-10</sup>. Despite these limitations, prior works have shown the potential of dMRI to probe the early brain development. For instance, several studies<sup>11-13</sup> have used spatiotemporal changes in Fractional Anisotropy (FA), Mean Diffusivity (MD) and different cortical morphology indices to characterize normal brain development. Recent availability of large high-quality datasets such as those collected under the developing Human Connectome Project (dHCP)<sup>14,15</sup> present a unique opportunity to expedite our understanding of the developing brain. These datasets are of highest quality, including measurements at different b-values and gradient directions. As such, derived dMRI quantities can be considered as reference values or ground truths to which derived metrics from more constrained clinical datasets, which usually do not exceed 15 diffusion measurements with a single low b-value ( $500 - 750s/mm^2$ ), can be compared to.

The prevailing way of extracting diffusion properties from the diffusion signal involves a model, typically a diffusion tensor imaging (DTI) model<sup>16</sup>. More complex models such as multi-shell multi-tissue constrained spherical deconvolution (MSMT-CSD), aiming to reconstruct Fiber Orientation Distribution Functions (FODs)<sup>17,18</sup> that allow depiction of more intricate white matter configurations such as fiber crossings require densely sampled (multiple-shell) data. The output of these models can be studied directly, i.e. by computing metrics such as FA or MD from the diffusion tensor or the apparent fiber density<sup>19</sup> from the FOD. Alternatively, they can be further processed globally to generate fiber tracts<sup>20,21</sup> that are responsible

for transmitting action potentials between different regions of the brain.

In general, mapping the acquired diffusion signal to an interpretable and informative diffusion metric requires a prior model. Conventional model computation does not provide feedback of which parts of the raw signal are important and which are not in estimating the given model. Differently, deep neural networks can treat the problem as a single learnable task that can be optimized via *back-propagation*, by directly learning a mapping between the diffusion signal and the target diffusion quantity. Hence, bypassing the sub-optimal model fitting step that can be sensitive to noise. Golkov et al.<sup>22</sup> proposed the first deep learning (DL) model that directly estimated diffusion kurtosis<sup>23</sup> and neurite orientation dispersion and density measures<sup>24</sup> from a small number of diffusion measurements in adult brains. They showed a drastic decrease in scanning time with limited loss in accuracy. Since then few other works have explored DL methods in adult brains as to directly estimate diffusion scalars or model reconstruction (FODs). For instance, with superDTI<sup>25</sup>, accurate predictions of tensor maps using a neural network was achieved using only six diffusion measurements. Their model was robust to various noise levels and could depict lesions present in the dataset. Kopper et al.<sup>26</sup>, employed 2D convolutional neural network (CNN) in a classification approach to predict the orientation of fibers, while Lin et al.<sup>27</sup> utilized a 3D CNN to predict FODs based on a small neighborhood of the diffusion signal. Karimi et al.<sup>28</sup> used a multi-layer perceptron to predict FODs. However, this approach did not leverage the correlations between neighboring voxels. A two-stage Transformer-CNN was used to map 200 measurements to 60 measurements by Hosseini et al.<sup>29</sup>, followed by predicting FODs. Acquiring such a large number of measurements is infeasible for noncooperative cohorts, such as neonates or fetuses.

To the best of our knowledge, these learning-based FOD estimation methods have not yet been critically evaluated for fetal populations and in non-research protocols of newborns. In this study, we demonstrate that a deep convolutional neural network with a large field of view (FOV) can accurately estimate FODs using only 6-12 diffusion-weighted measurements. Firstly, we show, on N=465 subjects from the dHCP dataset, that a deep learning approach can achieve a level of accuracy that is comparable to the accuracy of the state-of-the-art methods, while reducing the required number of measurements by a factor of  $\sim 21-43$ . Secondly, we present evidence of a low agreement between state-of-the-art methods and acquisitions in terms of various metrics. Finally, we show the generalizability of deep learning methods on two out-of-domain clinical datasets of 26 *in vivo* fetuses and neonates that were scanned with different scanners and acquisition protocols.

## Results

### Research dMRI acquisitions of neonates

We trained our deep learning model ( $DL_n$ ) on dMRI data from neonates. FODs estimated with the MSMT-CSD method using 280 diffusion-weighted and 20 b0 measurements are used as estimation target. We refer to MSMT-CSD estimations as ground truth (GT). The input to  $DL_n$  consists of 6 diffusion-weighted measurements normalized with one b0 measurement. After training, the network was applied on independent test data.

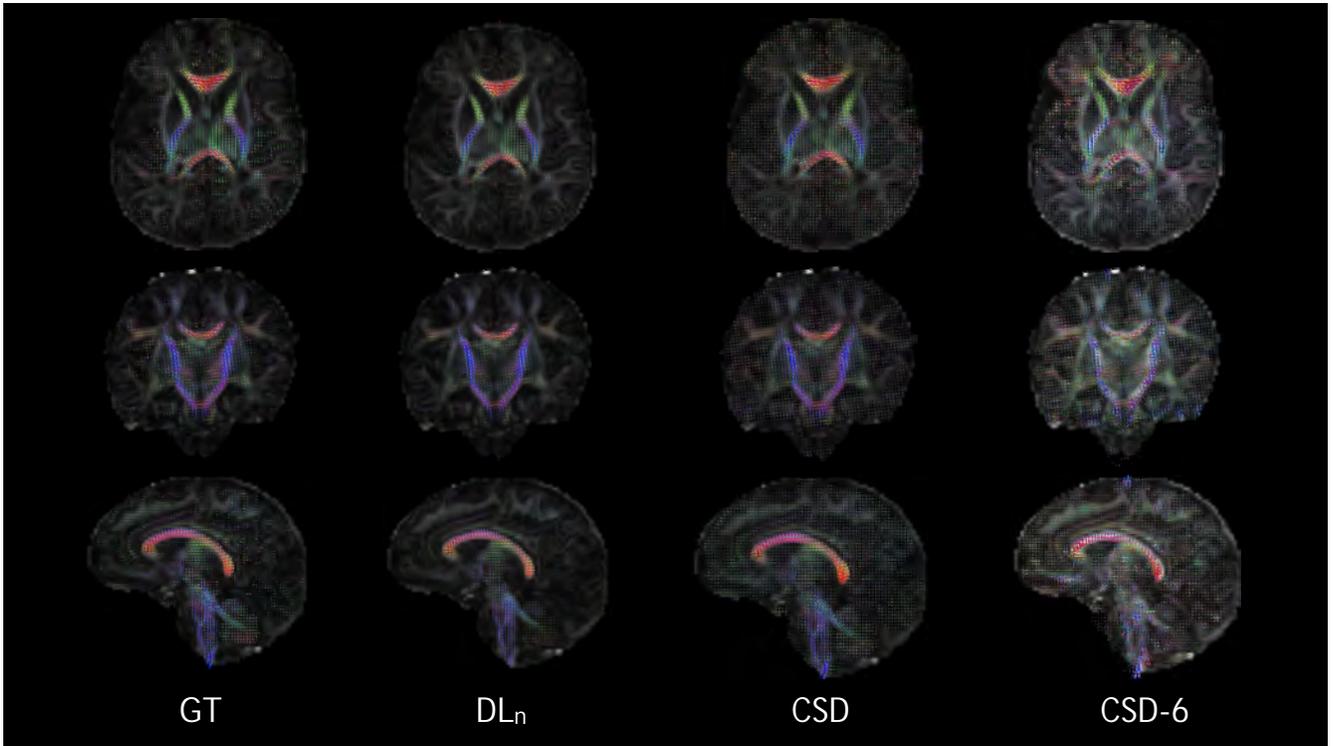
Qualitatively, the FODs estimated by  $DL_n$  were very similar to those estimated by MSMT-CSD using 300 measurements in 3 shells (Figure 1). We also compared our results with those of CSD using 148 measurements. CSD overestimates the number of peaks in the non-white matter regions. Although estimating the 45 FOD coefficients using 6 measurements is an under-determined problem, for the sake of comparison we present the CSD estimated FODs with the six measurements as those used for  $DL_n$  (CSD-6 in Figure 1). CSD-6 results show significant errors even in the location of major white matter tracts such as the corpus callosum.

### Quantitative assessment

On N=320 independent test subjects from the dHCP dataset,  $DL_n$  showed low estimation error (with respect to the MSMT-CSD GT) in terms of several metrics compared to the various standard estimation methods. We assessed the reproducibility of the GT by applying MSMT-CSD on subsets of the measurements from the same subject. Specifically, we split the 300 measurements into two disjoint subsets of 150 multi-shell measurements and applied MSMT-CSD on each subsets to compute two independent FOD estimations, which we denote with GS1 and GS2. GS1 and GS2 can be viewed as two high-quality scans of the same subject, conducted with a similar protocol.

In particular, the  $DL_n$  model has lower error rates on apparent fiber density<sup>19</sup> of 0.178 ( $\pm 0.083$ ) when compared to the GT, which is in close proximity to the corresponding gold standard difference of 0.064 ( $\pm 0.034$ ), in terms of the mean and standard deviation. In contrast, the other methods, namely Constrained Spherical Deconvolution (CSD)<sup>17</sup> using 148 measurements Constant Solid Angle<sup>30</sup> (CSA) and the Sparse Fascicle Model (SFM) using all 300 measurements, display elevated error rates when compared to the  $DL_n$  model as can be depicted from Figure 2. It is noteworthy that statistical significance of all pairs of models is observed, in particular  $p \leq 4.8^{-11}$  paired t-test corrected for multiple comparisons with Bonferroni between the  $DL_n$  and the other methods.

Peaks count and orientations were also estimated from the FOD generated by all methods. We first identified a low level of agreement rate (AR) for multiple fibers within the MSMT-CSD GT ( $\Delta GS$ ) as can be shown in Figure 3 (a). The AR was

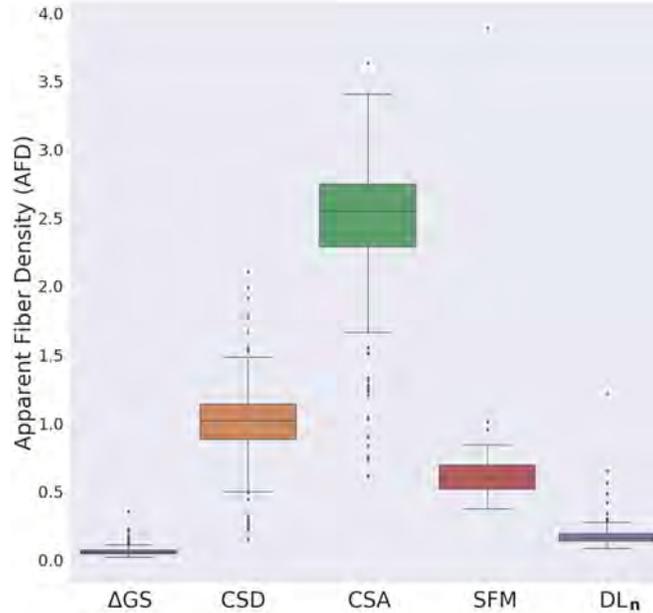


**Figure 1.** Qualitative high level comparison between, from left to right, the MSMT-CSD GT using the 300 multi-shell samples, the deep learning method  $DL_n$  using six  $b = 1000 \text{ s/mm}^2$  measurements and one  $b_0$ , and CSD using 128 measurements of  $b = 2600 \text{ s/mm}^2$  and 20  $b_0$  images. Also shown on the right CSD-6, i.e. CSD with the same measurements that the DL method used. Axial, coronal and sagittal views are shown from top to bottom and the background images corresponds to fractional anisotropy (FA) extracted from the diffusion tensor estimated with all  $b = 1000 \text{ s/mm}^2$  measurements.

extracted from the confusion matrix of the estimated number of peaks (Table 1, details in Methods section). For instance, the 1-peaks AR was 88.2%, while 46.7% and 47.2% were observed for 2-peaks and 3-peaks, respectively. Our proposed method,  $DL_n$ , achieved an agreement of 77.5%, 22.2%, and 8% for 1-peaks, 2-peaks, and 3-peaks, respectively, which was globally the closest to the agreement between the gold standards when compared to other methods. Although the single-fiber model (SFM) produced a relatively high level of agreement for 1-peaks with the ground truth (GT) at 84.6%, the agreement decreased to 4.6% and 2.5% for 2-peaks and 3-peaks, respectively. In contrast, the constrained spherical deconvolution (CSD) model achieved the lowest 1-peaks AR at 21.7%. This model showed a bias towards the estimation of multiple peaks, with 78% of the voxels modeled as having two or three peaks, which could be explained by the high b-value ( $b = 2600 \text{ s/mm}^2$ ) that contains high levels of noise.

The relatively low agreement observed for voxels with multiple intravoxel fiber orientations might be attributed to their incongruence across the GT, resulting in the absence of a consistent pattern to be learned by the neural network. In fact, this is supported by the modest agreement between the two gold standards ( $\Delta GS$ ) where both the subjects and the number of measurements are the same, with only the gradient directions varying and already resulting in a drop of more than 50% in multiple fibers depiction. It is noteworthy that the agreement between different methods such as CSD versus CSA, SFM versus CSA, CSD versus  $DL_n$ , among others, was also low. The confusion matrices for  $\Delta GS$  agreement and the different methods can be found in Table 1.

Our analysis, presented in the table of Figure 3 (b), quantifies the angular error for different FOD methods. Our proposed learning model achieves an error rate that is comparable to GS1 and GS2. However, SFM and CSA methods demonstrate a higher error rate for single and two-fiber voxels, whereas CSD outperforms the other techniques by achieving the lowest error rate. This could be attributed to the low AR observed for CSD, which reduces the error computation to a smaller subset of common voxels between the ground truth and CSD, as indicated in Figure 3 (a). The table of Figure 3 (b) also displays the number of measurements and the b-values that each method used. Notably, the angular error exhibits a nearly linear increase for voxels containing one, two, or three fibers. It is worth noting that training a network with 15 directions instead of 6 did not lead to a noticeable improvement in the results.



**Figure 2.** Apparent fiber density error with respect to the MSMT-GT for the different methods, along with the agreement between the two gold standard datasets ( $\Delta GS$ ) that is shown as an upper bound error. The different baseline methods used are Constrained Spherical Deconvolution (CSD)<sup>17</sup>, using 128 gradient directions (b-value of  $2600 \text{ s/mm}^2$ ) and 20  $b0$  images; Constant Solid Angle<sup>30</sup> (CSA) and the Sparse Fascicle Model (SFM) model<sup>31</sup> using all available 300 measurements.  $DL_n$  method, with less than an order of magnitude in the number of samples (six  $b=2600 \text{ s/mm}^2$  samples) and one  $b0$  image) achieves the lowest error by a high margin.

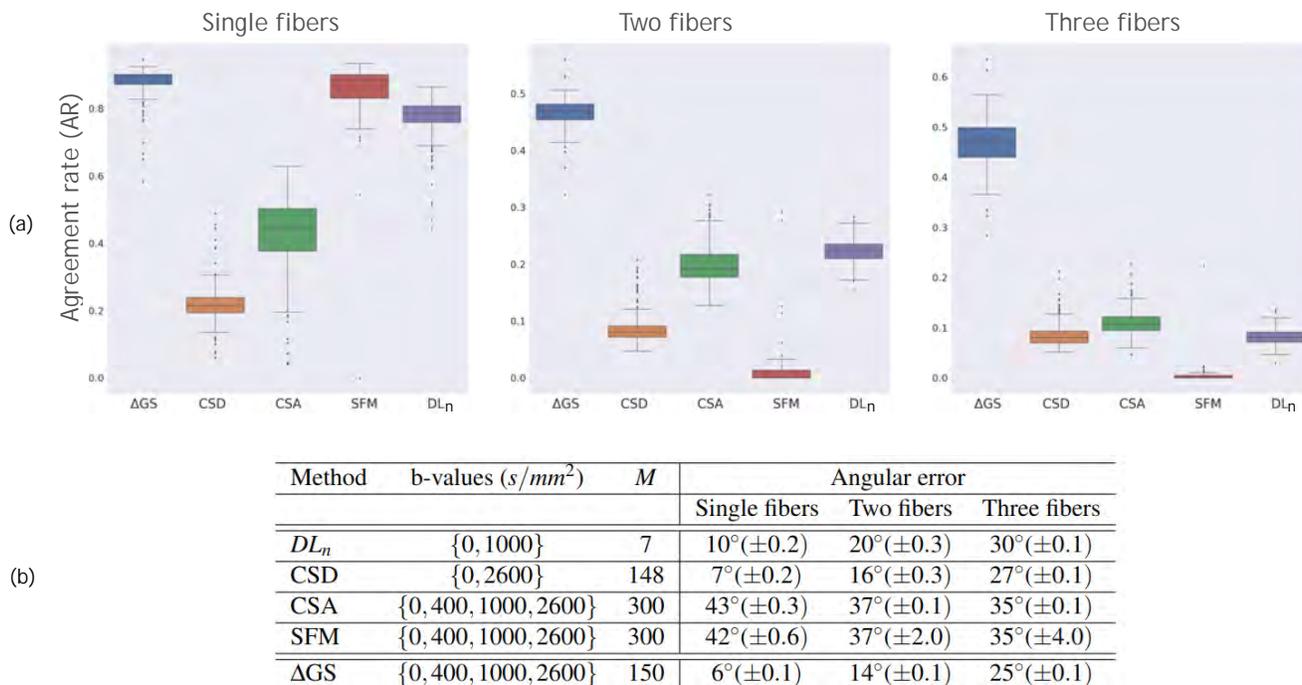
	$\Delta GS$			CSD			CSA			SFM			$DL_n$		
#Fibers	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	72.3	4.79	0.31	16.7	11.7	47.8	34.7	27.7	13.9	83.7	0.45	1.82	70.47	5.16	0.59
2	4.27	11.0	1.87	0.51	2.55	14.4	2.28	9.46	5.76	10.2	0.53	0.32	11.4	5.57	5.15
3	0.26	1.57	3.60	0.07	0.37	5.80	0.85	2.57	2.83	2.70	0.09	0.12	3.27	2.38	0.59

**Table 1.** Confusion matrices for number of peaks agreement (in %), normalized over all population. From left to right: gold standards GS1 vs. GS2, followed by the different methods CSA, CSD, SFM and  $DL_n$  compared to the GT MSMT-CSD. Each confusion matrix reports the average result for 320 test subjects (except SFM with 56 subjects).

Finally, we explored the correlation with the quality control (QC) metrics provided by dHCP and the error metrics for  $\Delta GS$ ,  $DL_n$ , and CSD. The different error measures showed no correlation to QC metrics, i.e. SNR, outlier-ratio, nor scan age (Figure 10 in Supplementary Material for  $DL_n$ ) except for the motion as estimated by the SHARD<sup>32</sup> pipeline. The more motion was estimated the higher the correlation to a lower agreement rate and a higher AFD error across subjects for the intra-agreement metrics of the MSMT-CSD GT ( $\Delta GS$ ) and both methods ( $DL_n$  and CSD) as can be shown in Figure 4. Statistical interaction analysis did not show however any significant difference in the way  $DL_n$  and CSD are influenced by motion ( $p = 0.8$ ,  $p = 0.18$  and  $p = 0.11$  for single, two and three fibers respectively). Given that the motion was compensated<sup>32</sup>, we hypothesize that subjects with strong initial motion have still an increased residual motion after correction.

### Uncertainty

Using wild bootstrap ( $N_{WBS}=60$ ) on the six input directions of the 88 volumes of  $b = 1000 \text{ s/mm}^2$  volumes, we have computed uncertainty maps using normalized standard deviation (please see Methods section). Figure 5 shows these maps compared to FA, where both images were applied a white matter mask. We can appreciate low uncertain regions in the highly anisotropic regions such as corpus callosum (body, splenium and genu) and internal capsules. In fact, this is in line with the results (Figure 3) where the  $DL_n$  model is less prone to errors in single coherent fiber populations as was previously studied in diffusion tensor imaging<sup>33-35</sup>. Hence, since these uncertainty maps do not need any ground truth and can express an increased correlation with erroneous predictions<sup>36</sup>, they can be used as an informative proxy to error detection, in case enough gradient directions are available for bootstrap.



**Figure 3.** (a) Agreement rates, extracted from confusion matrices as defined in the Methods section, for different methods compared to the MSMT-CSD GT and for the agreement between the gold standard subsets. From right to left, the deep learning method using six measurements and  $b_0$ , SFM and CSA using 300 multishell samples, CSD using 148 measurements and the agreement between the two gold standard ( $\Delta GS$ ) mutually exclusive subsets using each 150 samples. (b) Mean and standard deviation of angular error between GT (MSMT-CSD) and the different methods.  $\Delta GS$  refers to GS1 and GS2 agreements. The number of measurements ( $M$ ) and the b-values used are also reported. All results were statistically significant compared to  $\Delta GS$  ( $p \leq 9e^{-10}$ ). Our method achieves results comparable to the agreement rate  $\Delta GS$  while using six measurements. It is worth noting that CSD is achieving slightly lower error because it misses more than 3 times GT-true single fiber voxels and more than two times GT-true two-fiber voxels, as can be seen in its low agreement rate in (a).

## Generalizability to clinical datasets

### Neonates dMRI

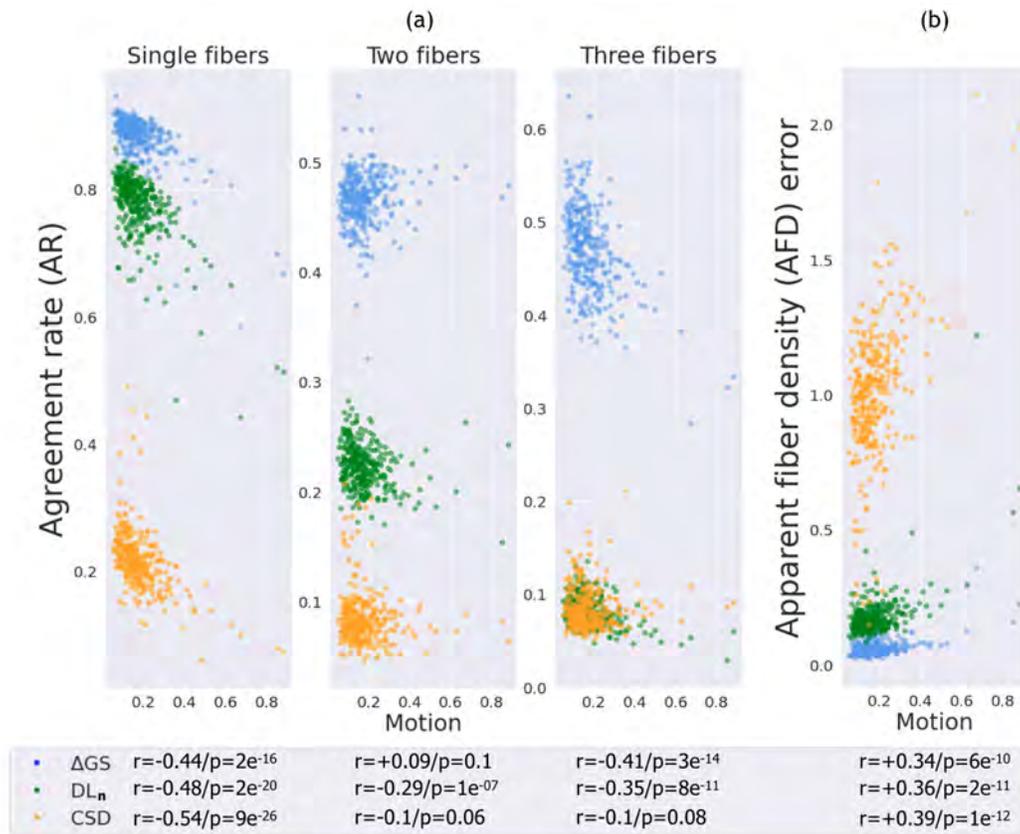
The network  $DL_n$  trained on dHCP neonates was tested on 15 clinical newborns using six  $b_0$ -normalized input volumes of  $b = 1000 s/mm^2$  as can be seen in Figure 6. As for the dHCP newborns, we can see the absence of high magnitude FODs in non-white matter regions, as opposite to noise-sensitive CSD (estimated using all 30  $b = 1000 s/mm^2$  diffusion measurements and 5  $b_0$  volumes) that displays several false positive crossing fibers. These false crossings can also be noticed in some known single fiber areas such as the internal capsules as can be depicted in subject 1 of Figure 6.

### In-utero fetal dMRI

We tested the proposed deep learning model,  $DL_f$ , on 11 fetuses and compared it to CSD. In the absence of dMRI ground truth, we qualitatively evaluate the results. In Table 2, we summarized the analyzed anatomical regions. We point out to the frequency on which  $DL_f$  or CSD was depicted as better or in which they seem equivalent. The evaluation was conducted by an experienced developmental neuroanatomist and was based on former knowledge from histology.

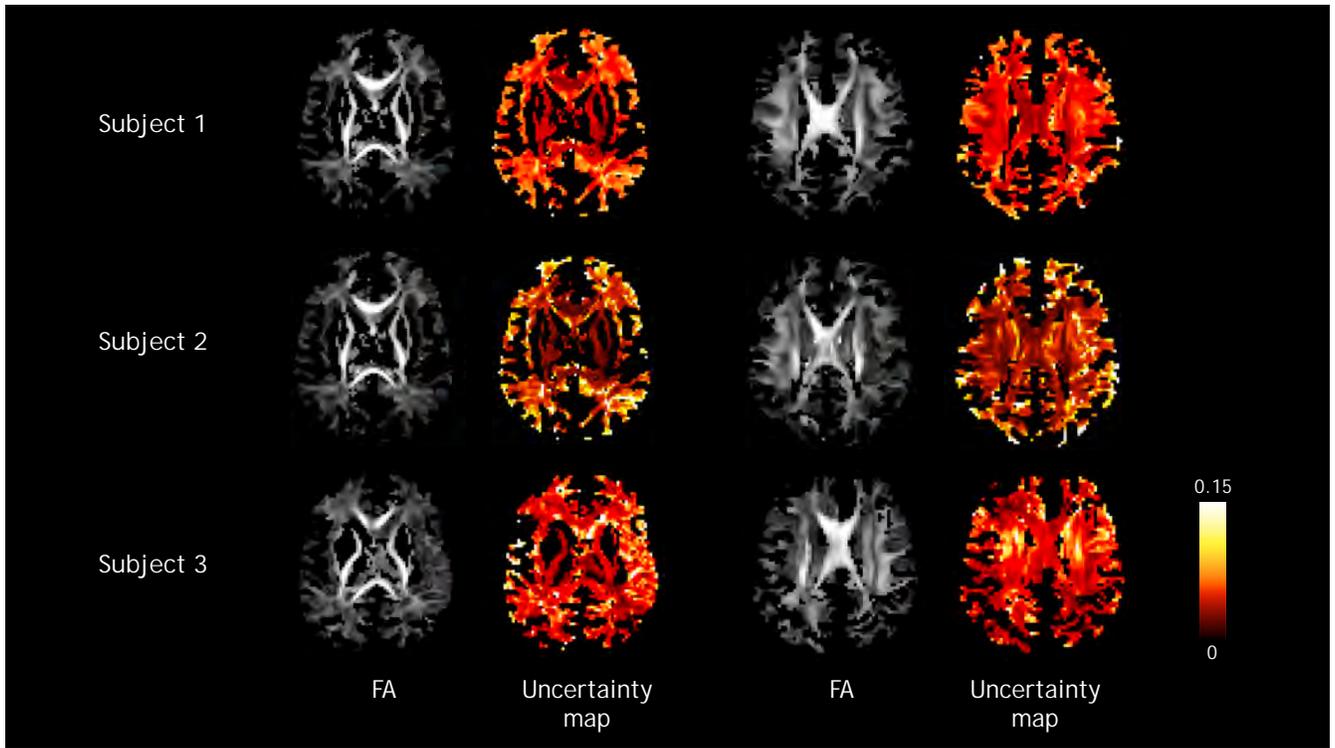
This qualitative assessment relied on visually inspecting FOD maps in ROIs. We selected ROIs within regions whose tissue components are relatively known during early development. Specifically, these ROIs included: i) regions in the proximity of the frontal crossroad C2, ii) corpus callosum, iii) cortical plate (in the insula, superior temporal gyrus, prefrontal cortex), iv) subplate (precentral gyrus and sulcus and prefrontal cortex) v) internal capsule, vi) cerebral peduncles, and vii) intermediate zone (regions containing geniculocortical fibers, regions containing callosal fibers). Figure 7 depicts some of the aforementioned ROIs within two example subjects across the two methods on the corresponding FOD maps.

Overall,  $DL_f$  performed better than CSD in predicting fiber orientation across most regions. Specifically, upon visual inspection, the regions surrounding the frontal crossroad region C2, genu of corpus callosum, intermediate zone containing



**Figure 4.** Agreement rate for voxels containing one, two, and three fibers, and apparent fiber density (AFD) error for the inter agreement between the gold standard datasets ( $\Delta GS$ ), deep learning method ( $DL_n$ ) and CSD, as a function of motion parameters (average translation and rotation parameters) for N=320 subjects. A negative correlation is generally observed with agreement rate (Spearman's rank correlation coefficients shown in the figure with corresponding p-value). Similarly, a positive correlation with AFD error can be seen on the (b) panel. Other quality control (QC) metrics (Outlier ratio, Signal-to-noise ratio) and scan age didn't exhibit any trend with the prediction accuracy (See Figure Sx in Supplementary Materials?). Interaction analysis showed that CSD and  $DL_n$  were not significantly affected by motion ( $p \geq 0.11$ ).

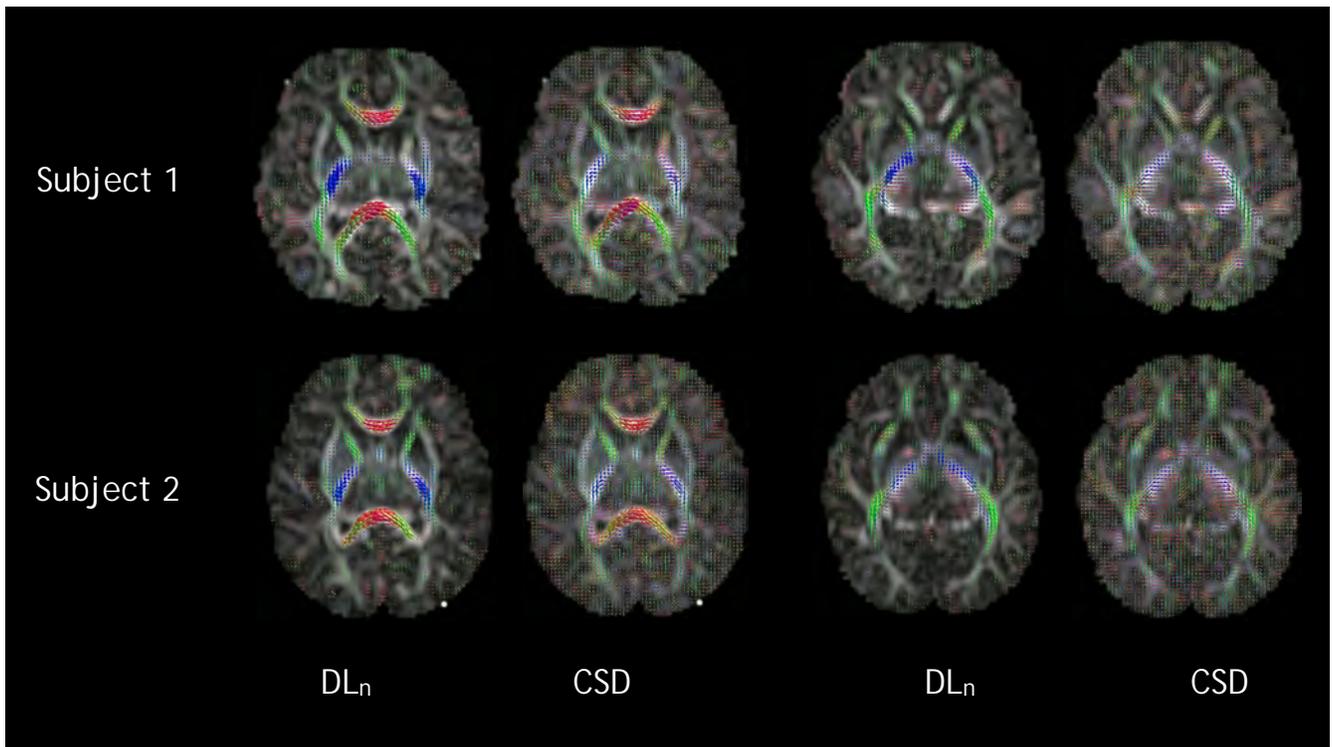
callosal fibers, and prefrontal cortical plate were better defined using  $DL_f$ . It is worth noting CSD systematically outperformed  $DL_f$  for cerebral peduncles and internal capsules on coronal sections. For the purpose of the current article, we added two slides to Figure 8 showing corresponding histology slices stained with GFAP (stains glial fibrillary acidic protein) or SMI 312 (stains highly phosphorylated axonal epitopes of neurofilaments)<sup>37</sup>. The criteria for evaluation included orientation, magnitude, and coherence of FODs. Specifically, in regions of corpus callosum<sup>38</sup>, cerebral peduncles, intermediate zone containing geniculocortical or callosal fibers, and internal capsule we expected high coherence with high magnitude, and orientation along or perpendicular to the main brain axes<sup>38,39</sup>. In contrast, within ROIs in the proximity of the frontal C2 crossroad<sup>40</sup>, we expected decreased coherence, with low magnitude and ambiguous orientation with FOD maps. Finally, the magnitudes, orientation, and coherence within the cortical plate and subplate ROIs were based on diffusion<sup>39,41</sup> and histological descriptions of underlying microstructure<sup>42</sup>. As depicted in Figure 8,  $DL_f$  successfully defines the mediolateral orientation below the sulcus and rostrocaudal orientation of fibers in the gyrus.



**Figure 5.** From top to bottom: three dHCP test subjects of respectively 43, 42 and 40 weeks. Uncertainty maps, computed using coefficient-normalized standard deviation of 60 bootstrapped gradient directions as described in the Methods section, are shown on the right. On the left, corresponding FA maps calculated from the diffusion tensor, that highlights regions of high anisotropy. Low uncertainty can be seen in such regions as the corpus callosum or the cortico-spinal tract, where the network has lower prediction errors. A white matter mask was applied to all images.

Fetal brain region	$DL_f$	Tied	CSD
Frontal crossroad region	10	0	1
Genu of corpus callosum	11	0	0
Cortex of Insula	2	5	4
Posterior limb of internal capsule	7	2	2
Cortex of superior temporal gyrus	6	2	3
Subplate of the precentral gyrus	4	0	7
Internal capsule	3	0	8
Cerebral peduncles	1	1	9
Intermediate zone, geniculocortical	4	3	4
Intermediate zones, callosal	10	0	1
Prefrontal subplate	6	4	1
Prefrontal cortical plate	8	2	1
<b>Count per ROI</b>	<b>7</b>	<b>1</b>	<b>3</b>
<b>Count per subject</b>	<b>9</b>	<b>0</b>	<b>2</b>

**Table 2.** Comparison between the preferred method ( $DL_f$ , CSD, or tied) for different regions of interest (ROI) in assessing the validity of the fibers in neurotypical fetal brains.



**Figure 6.** Qualitative comparison for two clinical newborn subjects (subject 1 and subject 2 of 41.8 and 38.1 weeks respectively) between the deep learning method  $DL_n$  (trained on dHCP dataset) using six  $b = 1000 \text{ s/mm}^2$  measurements and one  $b0$ , and CSD using 30 measurements and 5  $b0$  images. The background images are the corresponding fractional anisotropy (FA) maps.

**Axial Slices**

**DL<sub>f</sub>**

**CSD**

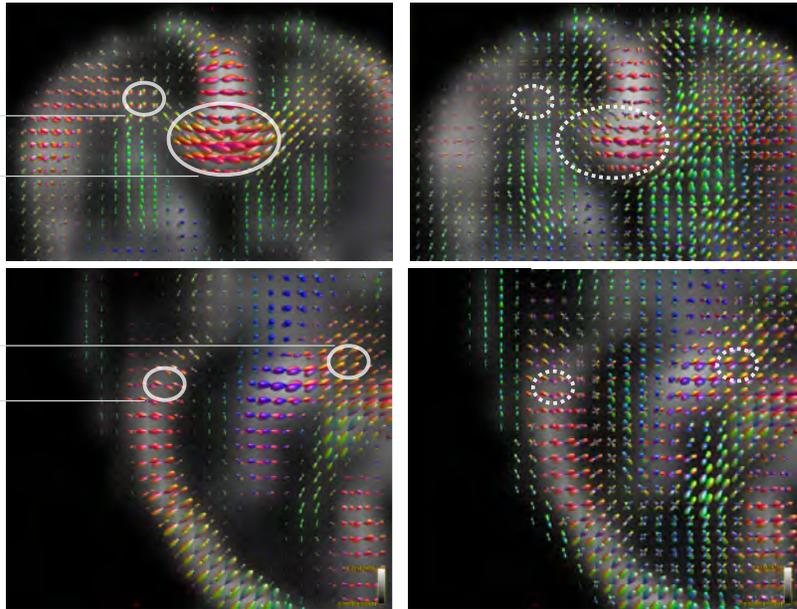
Example subject 1

Frontal Crossroad Area C2

Genu of Corpus Callosum

Posterior Limb of Internal Capsule

Cortical Plate of Superior Temporal Gyrus



Example subject 2

Frontal Crossroad Area C2

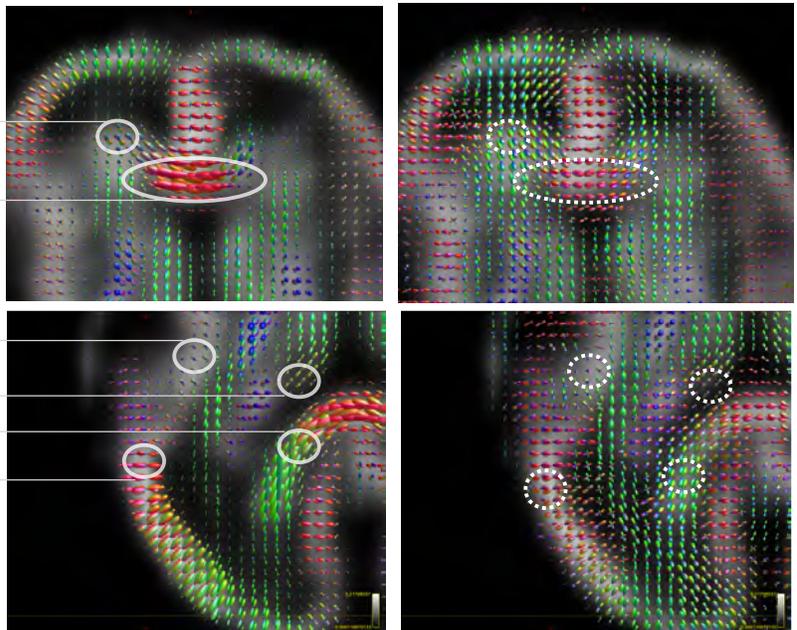
Genu of Corpus Callosum

Cortical Plate of Insula

Posterior Limb of Internal Capsule

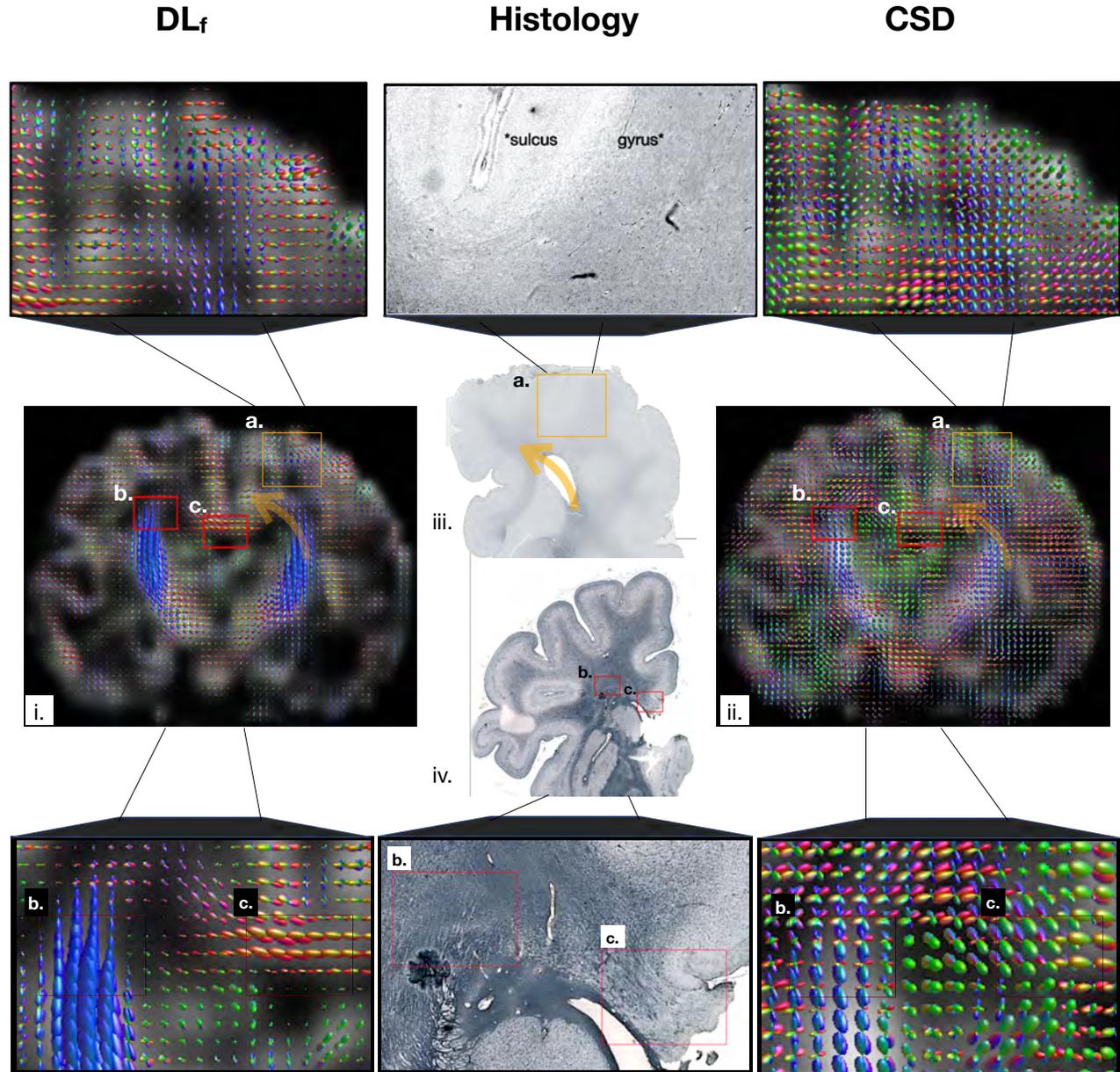
Intermediate Zone Splenium of Callosum

Cortical Plate of Superior Temporal Gyrus



**Figure 7.** Qualitative assessment: Visual inspection of ROIs within FOD maps for two example subjects computed with  $DL_f$  and CSD. ROIs were selected based on the knowledge of microstructure from histology and immunohistochemistry.

## Coronal Slices



**Figure 8.** Qualitative assessment: Panel (iii.) shows a slice of a 40 GW fetal brain stained with SMI 312 directed against highly phosphorylated axonal epitopes of neurofilaments<sup>37</sup>, with rostral ROIs marked with an orange rectangle. This section is an example of coronal sections which were taken into consideration for the assessment of accuracy for our (i)  $DL_f$  and (ii) CSD method. Note the compactness of stained regions (marked with asterisks (\*)) in the magnified panels above figure iii. suggesting the mediolateral orientation of axonal fibers below the sulcus and rostrocaudal orientation with fanning of fibers within the gyrus. Corresponding regions are marked as (a.) within the FOD maps of both methods. Panel (iv.) shows another example of the coronal sections (40 GW fetal brain stained with GFAP) with 2 ROIs marked with red rectangles (e.g., the proximity of frontal crossroad area C2 (b.) and corpus callosum (c.)) that were taken into consideration for the assessment of accuracy for our (i)  $DL_f$  and (ii) CSD method. Note the compactness of GFAP-stained regions in red rectangles suggesting the orientation of axons in these regions in the magnified panels below.

## Discussion

In this study, we showed the effectiveness of deep neural networks in microstructure estimation of the developing human brain. Quantitative evaluation was performed on the highly controlled research oriented dHCP dataset where we show using several metrics that with six uniform measurements<sup>43</sup>, a carefully trained network can achieve performance level on par with the standard methods such as CSD. In particular, apparent fiber density<sup>19</sup>, a measure that is sensitive to fiber partial volume fraction was best estimated with the deep learning model.

We have additionally shown the out-of-domain generalization of the model to fetal and newborn clinical datasets, despite acquisition and anatomical gaps with respect to the training dataset. In fact, transferring knowledge from rich research oriented datasets (multiple b-values, multiple gradient directions, high magnetic field strength)<sup>14,44</sup> to clinical datasets can be a winning strategy for developing cohorts such as from pre-terms to fetuses<sup>28,45</sup> that cannot afford prolonged acquisition times because of increased motion, maternal discomfort and the sensitivity of this population.

Another important aspect of this study is the low agreement in estimating multiple intra-voxel fibers within the ground truth multi-shell multi-tissue CSD, hampering deep learning methods from learning consistent crossing fibers across different subjects. This highlights the need for acquisitions and reconstruction methods that are physically and anatomically informed, and tailored to the developing brain<sup>46</sup>. The data preprocessing has also an impact on the quality of the reconstruction. We have found an intra-agreement difference between two pipelines<sup>32,47</sup> that reached 16% and 23% for the *agreement rate*, and 15° and 10°, for two and three fibers estimation<sup>18</sup> respectively. Hence our simple and efficient strategy of splitting HARDI datasets into two independent subsets and computing their agreement for different metrics can provide a method to assess consistency of preprocessing pipelines. Lastly, we have shown that uncertainty maps computed using wild bootstrap can be a proxy for voxel-wise error detection.

Recently, fiber orientation distribution function prediction using deep learning has shown a growing interest spanning several goals that go beyond the objective of directly learning FODs from raw data or its spherical harmonics representation<sup>27–29</sup>. For instance two recent studies, one aiming at mapping 3T dMRI data to 7T FODs<sup>48</sup> and one trying to simultaneously learn FODs from all radial combinations of multiple shells data using spherical convolutions<sup>49</sup> have been conducted. This last study can be particularly interesting to increase the generalizability of our work to multiple b-values. In fact, our model generalization from pre-term to fetuses is also due to the proximity of the two acquired b-values for training and evaluation ( $b = 400 \text{ s/mm}^2$  and  $b = 500 \text{ s/mm}^2$ , respectively). In contrast, our network has failed predicting coherent FODs of a different fetal dataset acquired at  $b = 700 \text{ s/mm}^2$ , likely because of the lower SNR and contrast difference, despite training our network on *b0*-normalized images. Hence, data harmonization<sup>50</sup> and other strategies enhancing generalization<sup>49</sup> can be adapted for future work.

Another limitation of this study is the absence of pathological datasets, which we aim to include in future work. We also intend to incorporate in the neural network, convolutions and deconvolutions that take into account the spherical property of the diffusion signal (angular dimension) such as roto-translation equivariant convolutions<sup>51</sup>. Moreover, as there is no consensus on the diffusion protocol of fetal brain diffusion imaging<sup>44,52–55</sup>, we want to explore optimal gradient tables that recover the most accurate white matter representation.

## Methods

### Data

#### **Research dMRI acquisition protocol in neonates**

We used the data from the third release of the publicly available dHCP dataset<sup>1</sup>. Scans were performed on a 3T Philips Achieva system with a customized 32-channel neonatal head coil. The protocol employed a TE of 90ms, TR of 3800ms, a multiband factor of 4, a SENSE factor of 1.2, a Partial Fourier factor of 0.855, a 1.5mm in-plane resolution, and 3mm slice thickness with 1.5mm slice overlap<sup>14</sup>. The diffusion gradient scheme used four shells  $\{0, 400, 1000, 2600\} \text{ s/mm}^2$  with 20, 64, 88, and 128 samples, respectively. The slice order was interleaved with a factor of 3 and a shift of 2. Data were processed and reconstructed with the SHARD<sup>32,56</sup> pipeline that included denoising, Gibbs ringing suppression, distortion correction and motion correction. The resolution of the processed data is  $1.5 \text{ mm}^3$  isotropic with a field of view of  $100 \times 100 \times 64$  voxels.

Two subsets were extracted from the SHARD-processed dataset, (i) 465 subjects with age range [26.71, 45.14] weeks (mean $\pm$ std =  $39.75 \pm 3.05$  weeks), and (ii) a group of 77 pre-term subjects with ages ranging from 26.71 to 38.0 weeks (mean $\pm$ std =  $34.79 \pm 2.52$  weeks). We have generated a white matter mask by combining the *White Matter* and the *Brainstem* labels provided by the dHCP with the voxels where Fractional Anisotropy (FA) was greater than 0.25. Finally, the dHCP labels were resampled from T2-w resolution ( $0.5 \text{ mm}^3$  isotropic) to  $1.5 \text{ mm}^3$  resolution.

<sup>1</sup><https://www.developingconnectome.org/data-release/third-data-release/>

### **Clinical dMRI acquisitions in neonates**

We retrospectively used data from 15 newborns ([38.14, 48] weeks,  $\text{mean} \pm \text{std} = 41.25 \pm 2.34$  weeks), while they were in natural sleep, using Siemens Trio and Skyra machines at 3T. The imaging protocol included acquiring 5  $b_0$  images and 30 diffusion-weighted images with  $b = 1000 \text{ s/mm}^2$ . The TR-TE values used were 3700-104 ms, and the voxel size was 2 mm isotropic. Images were resampled 1.5 mm<sup>3</sup> resolution.

### **Clinical dMRI acquisitions in fetuses**

A total of 11 motion-free fetuses at different gestational weeks ([24, 38.71] gestational weeks (GW),  $\text{mean} \pm \text{std} = 28.89 \pm 4.6$  GW) were included in this study. All subjects were scanned using a 3T Siemens Skyra MRI scanner, with one  $b_0$  and 12 diffusion-sensitized images at  $b = 500 \text{ s/mm}^2$ , with a TR of 3000–4000 ms and a TE of 60 ms. Preprocessing of the data was performed to correct for noise<sup>57</sup> and bias field inhomogeneities<sup>58</sup>. Registration of the images to a T2 atlas<sup>59</sup> was carried out using rigid transformation, and b-vectors were subsequently rotated accordingly. The resulting images were upsampled from  $2 \times 2 \times 3 - 4 \text{ mm}^3$  to  $1.5 \text{ mm}^3$ . Ethical approval for both clinical newborn and fetuses was granted by the institutional review board committee.

### **Histological post-mortem fetuses**

Neonatal post-mortem human brain specimens without evident pathological changes are part the Zagreb Collection of Human Brains<sup>60</sup>. Tissue was obtained during regular autopsies either after spontaneous abortions, or after the death of prematurely born infants at the clinical hospitals associated to the University of Zagreb, School of Medicine. After fixation in 4% paraformaldehyde (PFA), tissue was embedded in paraffin. Sections were cut in coronal plane and proceeded with routine immunohistochemistry protocol. In brief, after deparaffinization and 0.3% hydrogen peroxide treatment, sections were incubated in blocking solution: 3% bovine serum albumin BSA and 0.5% Triton x-100 (Sigma, St. Louis, MO) in 0.1M PBS. Next, sections were incubated with primary antibodies (anti-GFAP, Dako, z-0334, 1:1000; anti-SMI-312 [panaxonal anti-neurofilament marker], Biologend, 837904, 1:1000) at room temperature overnight. Following washes, sections were incubated with secondary, biotinylated antibodies according to manufacturer's protocol (Vectastain ABC kit, Vector Laboratories, Burlingame, CA). Staining was developed using 3,3-diaminobenzidine (DAB) with enhancer (Sigma, St. Louis, MO) and slides were coverslipped (Histomount mounting medium, National Diagnostics, Charlotte, NC). Finally, staining was visualised by a high-resolution digital slide scanner NanoZoomer 2.0RS (Hamamatsu, Japan). Tissue sampling was performed in agreement with the Declaration of Helsinki, 2000, previously approved by the Institutional Review Board of the Ethical Committee, University of Zagreb, School of Medicine.

### **Model**

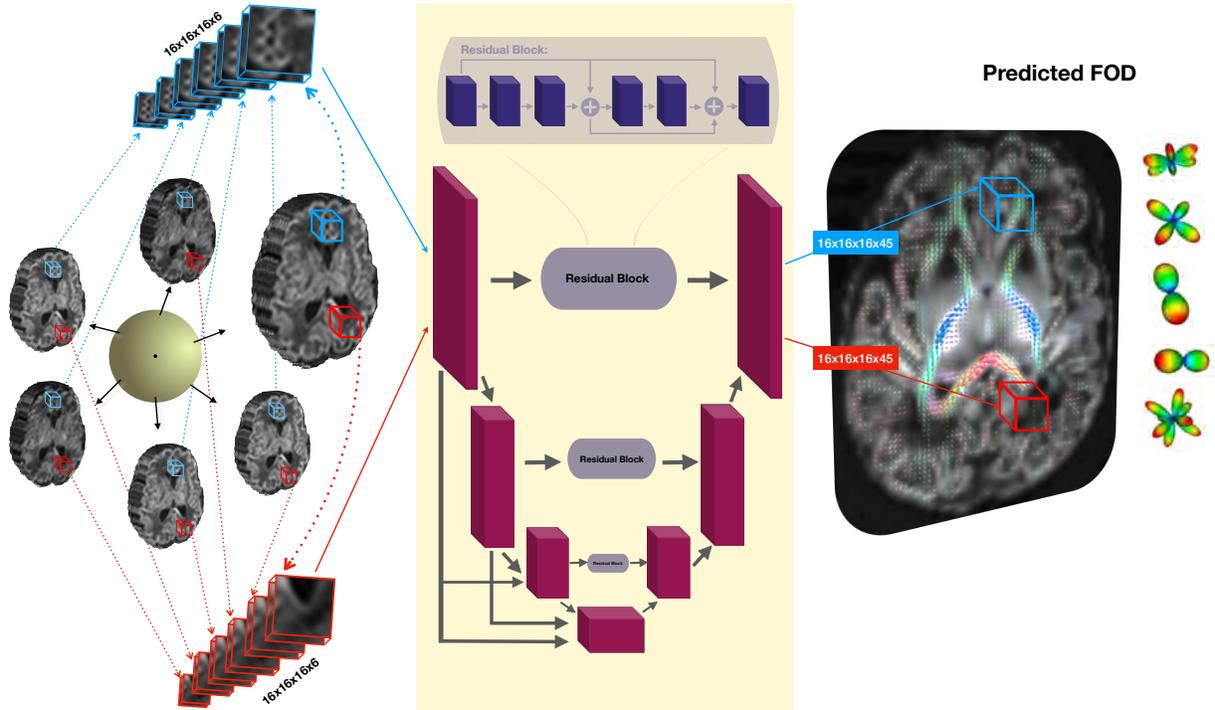
Our study employed two different neural networks, one for inference on neonates ( $DL_n$ ) and another one on fetuses ( $DL_f$ ).  $DL_n$  was trained on newborn subjects using six single-shell ( $b = 1000 \text{ s/mm}^2$ ) and  $DL_f$  was trained on pre-term subjects using twelve single-shell ( $b = 400 \text{ s/mm}^2$ ) measurements. To make the model independent on gradient directions, we projected the signal onto spherical harmonics basis (SH) with SH- $L_{max}$  order 2 and 4 for  $DL_n$  and  $DL_f$ , respectively, to predict the fiber orientation distribution (FOD) represented in the SH basis with SH- $L_{max}$  order 8. The latter is composed of 45 coefficients (45 channels for the network) and was generated using 300 multi-shell measurements obtained using MSMT-CSD<sup>18</sup>. These measurements were distributed over three shells with b-values of 400, 1000, 2600  $\text{s/mm}^2$  and had 64, 88, and 128 samples, respectively, along with 20  $b_0$  ( $b = 0 \text{ s/mm}^2$ ) images. The input measurements for the model were based on the scheme proposed by Skare et al.<sup>43</sup>, which minimized the *condition number* of the diffusion tensor reconstruction matrix.

### **Network architecture**

The deep convolutional neural network can be seen in the yellow box of Figure 9. Its architecture resembles that of U-Net<sup>61</sup> with two main modifications. Firstly, the network has extensive short and long-range residual connections, which provide more context to subsequent layers. This design choice is particularly important given the low dimensionality of our input (6 channels) compared to the output (45 channels). Secondly, the conventional max-pooling operations in the contracting path have been replaced with stride-2 convolutions to enable downsampling as a learnable step that is specific to each layer. The first block is set to 36 feature maps that are doubled after each contracting block. Each layer is composed of convolutions that are followed by Rectified Linear Unit (ReLU)<sup>62</sup> activation functions, followed by a dropout<sup>63</sup> layer. For the output layer, no ReLU nor dropout were applied.

### **Training strategies**

The input data was first normalized by  $b_0$  to improve network convergence and to reduce b-value dependency. Data were split into training, validation and test sets: for  $DL_n$ , 109, 36 and 320 subjects, respectively; for  $DL_f$ , 58, 19, pre-terms and 11 fetal subjects, respectively. To ensure balanced patch selection per batch, the number of FOD peaks (extracted using Dipy<sup>64</sup>) was



**Figure 9.** Schematic illustration of the proposed deep learning framework for predicting the Fiber Orientation Distribution (FOD). The input to the network consists of 3D patches derived from 6 diffusion measurements, which are normalized with  $b0$ . The network predicts the spherical harmonic coefficients (of order  $SH-L_{max} = 8$ ) of the FOD for the input patch. Two example patches are shown in blue and red. The network trained on pre-term newborns takes 12 instead of 6 measurements.

used as a criterion. The central voxel of each patch was restricted to be in the generated white matter mask and to have one peak in  $\frac{2}{3}$  of the batch and more than one peak in  $\frac{1}{3}$  of the batch. This condition ensured that empty patches were not selected. The patch size was empirically varied between in  $\{8^3, 16^3, 36^3, 48^3\}$  voxels (no performance increase was observed for patches bigger than  $16^3$ , so all networks use  $16^3$  voxels). At testing, the method employed a sliding window technique to sequentially process all non-empty patches.

We have used Adam optimizer<sup>65</sup> to minimize the  $\ell_2$  norm loss function between the predicted 45 spherical harmonic (SH) coefficients and the ground truth fiber orientation distribution (FOD) SH coefficients. Since the order of magnitude of the coefficients depends on which SH-order the coefficient belongs to, we have used pre-defined weights to penalize small coefficients. These weights were inversely proportional to the order of magnitude of the coefficient in the GT. Namely, these weights were proportional to the first SH coefficient and were around 2.5, 4, 7.5 and 20 for coefficients of SH order 2, 4, 6 and 8, respectively. However, no gain was observed with this scheme so all coefficient weights were set to 1.

The batch size was set to 27 for  $DL_n$  and 9 for  $DL_f$ , and the initial learning rate was set to  $10^{-4}$ . The learning rate was decreased by a factor of 0.9 whenever the validation loss did not improve after one epoch. The total number of training epochs was 10000, and a dropout rate of 0.1 was used in all layers to reduce overfitting and improve generalization. In  $DL_f$ , Gaussian noise ( $\mu = 0$ ,  $\sigma = 0.025$ ) was injected to input data to be robust to fetal noise. Moreover, small rotations (uniformly from  $[-5^\circ, +5^\circ]$ ) were applied to improve the robustness of the model to minor uncorrected movements due to small differences in scanning field of view and fetal head motion.

### Implementation details

All models were implemented using TensorFlow (1.6) and run on an NVIDIA GeForce GTX A6000 on a Linux machine with 125 GB of memory and 20 CPU cores. Convergence of each model took approximately 40 hours. Testing takes less than 1 minute per subject on the same machine. Code will be made publicly available.

### Evaluation

Quantitative evaluation has been carried out for  $DL_n$  predictions compared to the GT MSMT-CSD. Moreover, three state-of-the-art techniques were computed as baseline models, namely: Constrained Spherical Deconvolution (CSD) method<sup>17</sup>, using 128

gradient directions from the highest shell (b-value of 2600  $s/mm^2$ ) and 20  $b0$  images; Constant Solid Angle model<sup>30</sup> which is referred to as CSA; and the Sparse Fascicle Model (SFM)<sup>31</sup> for which the default regularization parameters were employed. The latter was only applied on 57 dHCP subjects as it takes a significant time to be run (around 7 hours per subject).

$DL_n$ , using 6  $b0$ -normalized diffusion measurements, was also tested on the clinical neonate dataset and compared against CSD using all measurements (35). Similarly,  $DL_f$  using 12  $b0$ -normalized diffusion samples was tested and evaluated against CSD. Because clinical datasets do not have densely sampled (multiple-shell) measurements that can be considered as high quality ground truth, only qualitative evaluation was performed.

### **Agreement within ground truth ( $\Delta GS$ )**

We evaluated the concordance between two distinct and mutually exclusive datasets derived from the 300 measurements of the dHCP dataset (GS1 and GS2, denoting gold standards 1 and 2 respectively, and their metrics difference  $\Delta GS$ ). Each subgroup consists of 150 samples ( $b \in 0, 400, 1000, 2600 s/mm^2$ ), with 10, 32, 44, and 64 scans respectively (half of the total scans of the GT). Both GS1 and GS2 can be regarded as independent high-quality scans of the same subject with a similar protocol (i.e. the same b-values and the same number but different gradient directions). Hence, any discrepancies between them with respect to diffusion metrics can be considered as an error upper bound of errors between the different methods and the full GT.

### **Quantitative performance on dHCP dataset**

We conducted a quantitative validation by evaluating the performance of different fiber orientation distribution (FOD) estimation methods. The validation was based on three metrics, namely, the **number of peaks, angular error, and the apparent fiber density (AFD)**<sup>19</sup>. The number of peaks was computed for the FODs predicted by the network and those estimated by various methods (GT, GS1, GS2, CSD, CSA, and SFM). We set up a maximum number of 3 peaks, a mean separation angle of 45°, and a relative peak threshold of 0.5. The choice of these parameters was guided by the work of Schilling et al.<sup>66</sup>, which demonstrated the limitations of current diffusion MRI models in correctly estimating multiple fiber populations and low angular crossing fibers. We compared the different models based on confusion matrices and the **agreement rate (AR)** that is extracted from the latter. AR was defined for each number of peaks  $p$  as:

$$AR = \frac{A_p}{\Sigma D_p} \quad (1)$$

where  $A_p$  represents the percentage of voxels where both methods agree on  $p$  number of peaks and  $D_p$  denotes the percentage of voxels where at least one of the two methods predicts  $p$  and the other predicts  $p'$  ( $p \neq p'$ ). This metric hence captures intuitively the rate of concordance between two methods.

**Mean angular error** was also computed for voxels containing the same number of estimated peaks. For voxels with multiple fibers, we extracted the corresponding peaks between the selected method and the GT (or the agreement between GS1 and GS) by computing the minimum angle between all configurations, namely 4 configurations for 2 peaks and 9 for 3 peaks. We subsequently eliminated those peaks and applied the same algorithm recursively until all peaks are matched. We also compared the error related to the apparent fiber density (FOD amplitude) along with the agreement between GS1 and GS2. We performed a statistical validation using paired t-test corrected for multiple comparisons with Bonferroni method to compare the errors of the different methods with respect to GT and the difference between GS1 and GS2.

The different error measures were correlated to quality control (QC) metrics provided by the SHARD pipeline of dHCP<sup>32</sup>. Namely, Signal-to-Noise Ratio that is calculated from denoising residuals; (2) Motion metrics, i.e. translation and rotation quantifying subject activity during scan and (3) Outlier ratio, as detected in slice-to-volume reconstruction<sup>32</sup>. We averaged both translation and rotation metrics to have one metric that we label as *motion*. We have also added the age of scan to the different QC metrics to check for any potential correlation. We have performed this analysis for the GT MSMT-CSD to assess the consistency of the dataset across the QC metrics.

### **Qualitative assessment of clinical datasets**

Detailed assessment was performed for the FODs generated on the clinical fetal dataset by an expert fetal neuroanatomist (LV). The images were blinded and the method used to reconstruct the maps was masked for the reader. The 12 ROIs were selected based on the anatomical knowledge (previously reported in Kunz et al. 2014<sup>67</sup>). Next, the corresponding slices of volumes reconstructed with both methods were placed side by side and FODs were inspected in each ROIs using MRView<sup>68</sup>. Based on the visual inspection and taking into the consideration coherence, orientation, and magnitudes, the ROIs were marked as 'better with  $DL_f$ ', 'better with CSD', or equal. After examining all the brains, we generated the table and counted ROIs and subjects where  $DL_f$  outperformed CSD, CSD outperformed  $DL_f$  or tied.

### **Uncertainty estimation**

A metric that expresses an increased likelihood on erroneous predictions can be very valuable in the absence of ground truth. Uncertainty in that sense can be used for that aim. Post-hoc uncertainty using wild bootstrap that has been used in diffusion tensor imaging<sup>69-71</sup> was the method of choice that was most suited to our study. We randomly selected 6 gradient directions ( $N_{WBS}=60$ ) from the 88 samples of the  $b = 1000 \text{ s/mm}^2$  shell from the dHCP data. The 6 directions were constrained to have a *condition number*<sup>43</sup> of at most 2 to guarantee that the b-vectors are minimally uniformly distributed. We then computed for each voxel, the standard deviation of the predicted FODs of the  $N_{WBS}$  bootstrapped volumes (Equation 2). Given that FOD coefficients have different orders of magnitude, this standard deviation was normalized by the norm of the FOD (Equation 3). Specifically, for each voxel we calculate  $\sigma_{norm}$  that we define as our uncertainty measure from:

$$\sigma = \frac{1}{N_{WBS}} \sum_{i=1}^{N_{WBS}} \|FOD_i - \mu\|_2 \quad \text{where } \mu_c = \frac{1}{N_{WBS}} \sum_{i=1}^{N_{WBS}} FOD_{i,c} \quad \text{and } c \in \{1, 45\} \quad (2)$$

$$\sigma_{norm} = \frac{\sigma}{m_j} \quad \text{where } m_j = \frac{1}{45} \sum_{c=1}^{45} \|FOD_{j,c}\| \quad \text{and } j \in \{1, N_{WBS}\} \quad (3)$$

## References

1. Volpe, J. Brain injury in premature infants: A complex amalgam of destructive and developmental disturbances. *Lancet Neurol.* **8**, 110–124, DOI: [10.1016/S1474-4422\(08\)70294-1](https://doi.org/10.1016/S1474-4422(08)70294-1) (2009).
2. Bhat, S., Acharya, U., Adeli, H., Bairy, G. & Adeli, A. Autism: Cause factors, early diagnosis and therapies. *Rev. Neurosci.* **25**, 841–850, DOI: [10.1515/revneuro-2014-0056](https://doi.org/10.1515/revneuro-2014-0056) (2014).
3. Kwon, E. & Kim, Y. What is fetal programming?: A lifetime health is under the control of in utero health. *Obstet. & Gynecol. Sci.* **60**, 506–519, DOI: [10.5468/ogs.2017.60.6.506](https://doi.org/10.5468/ogs.2017.60.6.506) (2017).
4. O'Donnell, K. & Meaney, M. Fetal origins of mental health: The developmental origins of health and disease hypothesis. *The Am. J. Psychiatry* **174**, 319–328, DOI: [10.1176/appi.ajp.2016.16020138](https://doi.org/10.1176/appi.ajp.2016.16020138) (2017).
5. Bilder, D. *et al.* Early second trimester maternal serum steroid-related biomarkers associated with autism spectrum disorder. *J. autism developmental disorders* **49**, 4572–4583, DOI: [10.1007/s10803-019-04162-2](https://doi.org/10.1007/s10803-019-04162-2) (2019).
6. Stejskal, E. O. & Tanner, J. E. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The journal chemical physics* **42**, 288–292 (1965).
7. Le Bihan, D. *et al.* Mr imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology* **161**, 401–407 (1986).
8. Dubois, J. *et al.* The early development of brain white matter: a review of imaging studies in fetuses, newborns and infants. *neuroscience* **276**, 48–71 (2014).
9. Christiaens, D. *et al.* In utero diffusion mri: challenges, advances, and applications. *Top. Magn. Reson. Imaging* **28**, 255–264 (2019).
10. Jakab, A. *et al.* Fetal cerebral magnetic resonance imaging beyond morphology. In *Seminars in Ultrasound, CT and MRI*, vol. 36, 465–475 (Elsevier, 2015).
11. Wilson, S. *et al.* Development of human white matter pathways in utero over the second and third trimester. *Proc. Natl. Acad. Sci.* **118**, e2023598118 (2021).
12. Xu, X. *et al.* Spatiotemporal atlas of the fetal brain depicts cortical developmental gradient. *J. Neurosci.* **42**, 9435–9449 (2022).
13. Calixto, C. *et al.* Detailed anatomic segmentations of a fetal brain diffusion tensor imaging atlas between 23 and 30 weeks of gestation. *Hum. Brain Mapp.* **44**, 1593–1602 (2023).
14. Hutter, J. *et al.* Time-efficient and flexible design of optimized multishell hardi diffusion. *Magn. resonance medicine* **79**, 1276–1292 (2018).
15. Tournier, J.-D. *et al.* A data-driven approach to optimising the encoding for multi-shell diffusion mri with application to neonatal imaging. *NMR Biomed.* **33**, e4348 (2020).
16. Basser, P. J., Mattiello, J. & LeBihan, D. Mr diffusion tensor spectroscopy and imaging. *Biophys. journal* **66**, 259–267 (1994).
17. Tournier, J.-D. *et al.* Direct estimation of the fiber orientation density function from diffusion-weighted mri data using spherical deconvolution. *Neuroimage* **23**, 1176–1185 (2004).
18. Jeurissen, B. *et al.* Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion mri data. *Neuroimage* **103**, 411–426 (2014).
19. Raffelt, D. *et al.* Apparent fibre density: a novel measure for the analysis of diffusion-weighted magnetic resonance images. *Neuroimage* **59**, 3976–3994 (2012).
20. Jbabdi, S. & Johansen-Berg, H. Tractography: where do we go from here? *Brain connectivity* **1**, 169–183 (2011).
21. Jeurissen, B. *et al.* Diffusion mri fiber tractography of the brain. *NMR Biomed.* **32**, e3785 (2019).
22. Golkov, V. *et al.* Q-space deep learning: twelve-fold shorter and model-free diffusion mri scans. *IEEE TMI* **35**, 1344–1351 (2016).
23. Lu, H., Jensen, J. H., Ramani, A. & Helpert, J. A. Three-dimensional characterization of non-gaussian water diffusion in humans using diffusion kurtosis imaging. *NMR Biomed. An Int. J. Devoted to Dev. Appl. Magn. Reson. In vivo* **19**, 236–247 (2006).
24. Zhang, H., Schneider, T., Wheeler-Kingshott, C. A. & Alexander, D. C. Noddi: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage* **61**, 1000–1016 (2012).

25. Li, H. *et al.* Superdti: Ultrafast dti and fiber tractography with deep learning. *Magn. resonance medicine* **86**, 3334–3347 (2021).
26. Koppers, S. & Merhof, D. Direct estimation of fiber orientations using deep learning in diffusion imaging. In *International Workshop on Machine Learning in Medical Imaging*, 53–60 (Springer, 2016).
27. Lin, Z. *et al.* Fast learning of fiber orientation distribution function for mr tractography using convolutional neural network. *Med. physics* **46**, 3101–3116 (2019).
28. Karimi, D. *et al.* Deep learning-based parameter estimation in fetal diffusion-weighted mri. *Neuroimage* **243**, 118482 (2021).
29. Hosseini, S. *et al.* Ctrack: A cnn+ transformer-based framework for fiber orientation estimation & tractography. *Neurosci. Informatics* **2**, 100099 (2022).
30. Aganj, I. *et al.* Reconstruction of the orientation distribution function in single-and multiple-shell q-ball imaging within constant solid angle. *Magn. resonance medicine* **64**, 554–566 (2010).
31. Rokem, A. *et al.* Evaluating the accuracy of diffusion mri models in white matter. *PloS one* **10**, e0123272 (2015).
32. Christiaens, D. *et al.* Scattered slice shard reconstruction for motion correction in multi-shell diffusion mri. *Neuroimage* **225**, 117437 (2021).
33. Jones, D. K. Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor mri. *Magn. Reson. Medicine: An Off. J. Int. Soc. for Magn. Reson. Medicine* **49**, 7–12 (2003).
34. Yuan, Y., Zhu, H., Ibrahim, J. G., Lin, W. & Peterson, B. S. A note on the validity of statistical bootstrapping for estimating the uncertainty of tensor parameters in diffusion tensor images. *IEEE transactions on medical imaging* **27**, 1506–1514 (2008).
35. Jeong, H.-K. & Anderson, A. W. Characterizing fiber directional uncertainty in diffusion tensor mri. *Magn. Reson. Medicine: An Off. J. Int. Soc. for Magn. Reson. Medicine* **60**, 1408–1421 (2008).
36. Karimi, D., Warfield, S. K. & Gholipour, A. Calibrated diffusion tensor estimation. *arXiv preprint arXiv:2111.10847* (2021).
37. Ulfig, N., Nickel, J. & Bohl, J. Monoclonal antibodies smi 311 and smi 312 as tools to investigate the maturation of nerve cells and axonal patterns in human fetal brain. *Cell tissue research* **291**, 433–443 (1998).
38. Vasung, L., Raguz, M., Kostovic, I. & Takahashi, E. Spatiotemporal relationship of brain pathways during human fetal development using high-angular resolution diffusion mr imaging and histology. *Front. neuroscience* **11**, 348 (2017).
39. Huang, H. *et al.* White and gray matter development in human fetal, newborn and pediatric brains. *Neuroimage* **33**, 27–38 (2006).
40. Judaš, M. *et al.* Structural, immunocytochemical, and mr imaging properties of periventricular crossroads of growing cortical pathways in preterm infants. *Am. journal neuroradiology* **26**, 2671–2684 (2005).
41. Khan, S. *et al.* Fetal brain growth portrayed by a spatiotemporal diffusion tensor mri atlas computed from in utero images. *Neuroimage* **185**, 593–608 (2019).
42. Kostović, I., Išasegi, I. Ž. & Krsnik, Ž. Sublaminar organization of the human subplate: developmental changes in the distribution of neurons, glia, growing axons and extracellular matrix. *J. Anat.* **235**, 481–506 (2019).
43. Skare, S. *et al.* Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with mri. *J. magnetic resonance* **147**, 340–352 (2000).
44. Deprez, M. *et al.* Higher order spherical harmonics reconstruction of fetal diffusion mri with intensity correction. *IEEE transactions on medical imaging* **39**, 1104–1113 (2019).
45. Kebiri, H. *et al.* Through-plane super-resolution with autoencoders in diffusion magnetic resonance imaging of the developing human brain. *Front. Neurol.* **13**, 765 (2022).
46. Pietsch, M. *et al.* A framework for multi-component analysis of diffusion mri data over the neonatal period. *Neuroimage* **186**, 321–337 (2019).
47. Bastiani, M. *et al.* Automated processing pipeline for neonatal diffusion mri in the developing human connectome project. *Neuroimage* **185**, 750–763 (2019).
48. Jha, R. R., Kumar, B. R., Pathak, S. K., Bhavsar, A. & Nigam, A. Trganet: Transforming 3t to 7t dmri using trapezoidal rule and graph based attention modules. *Med. Image Analysis* 102806 (2023).

49. Yao, T. *et al.* A unified single-stage learning model for estimating fiber orientation distribution functions on heterogeneous multi-shell diffusion-weighted mri. *arXiv preprint arXiv:2303.16376* (2023).
50. Pinto, M. S. *et al.* Harmonization of brain diffusion mri: Concepts and methods. *Front. Neurosci.* **14**, 396 (2020).
51. Elaldi, A., Dey, N. & Gerig, G.  $e(3) \times so(3)$ -equivariant networks for spherical deconvolution in diffusion mri. In *Medical Imaging with Deep Learning*.
52. Fogtman, M. *et al.* A unified approach to diffusion direction sensitive slice registration and 3-d dti reconstruction from moving fetal brain anatomy. *IEEE transactions on medical imaging* **33**, 272–289 (2013).
53. Jakab, A. *et al.* Disrupted developmental organization of the structural connectome in fetuses with corpus callosum agenesis. *Neuroimage* **111**, 277–288 (2015).
54. Marami, B. *et al.* Temporal slice registration and robust diffusion-tensor reconstruction for improved fetal brain structural connectivity analysis. *NeuroImage* **156**, 475–488 (2017).
55. Jakab, A., Tuura, R., Kellenberger, C. & Scheer, I. In utero diffusion tensor imaging of the fetal brain: a reproducibility study. *NeuroImage: Clin.* **15**, 601–612 (2017).
56. Pietsch, M., Christiaens, D., Hajnal, J. V. & Tournier, J.-D. dstripe: slice artefact correction in diffusion mri via constrained neural network. *Med. Image Analysis* **74**, 102255 (2021).
57. Veraart, J., Fieremans, E. & Novikov, D. S. Diffusion mri noise mapping using random matrix theory. *Magn. resonance medicine* **76**, 1582–1593 (2016).
58. Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y. *et al.* N4itk: improved n3 bias correction. *IEEE transactions on medical imaging* **29**, 1310–1320 (2010).
59. Gholipour, A. *et al.* A normative spatiotemporal mri atlas of the fetal brain for automatic segmentation and analysis of early brain growth. *Sci. reports* **7**, 476 (2017).
60. Hrabač, P. *et al.* The zagreb collection of human brains: entering the virtual world. *Croat. Med. J.* **59**, 283 (2018).
61. Ronneberger, O. *et al.* U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 234–241 (Springer, 2015).
62. Agarap, A. F. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375* (2018).
63. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal machine learning research* **15**, 1929–1958 (2014).
64. Garyfallidis, E. *et al.* Dipy, a library for the analysis of diffusion mri data. *Front. neuroinformatics* **8** (2014).
65. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
66. Schilling, K. G. *et al.* Histological validation of diffusion mri fiber orientation distributions and dispersion. *Neuroimage* **165**, 200–221 (2018).
67. Kunz, N. *et al.* Assessing white matter microstructure of the newborn with multi-shell diffusion mri and biophysical compartment models. *Neuroimage* **96**, 288–299 (2014).
68. Tournier, J.-D., Calamante, F. & Connelly, A. Mrtrix: diffusion tractography in crossing fiber regions. *Int. journal imaging systems technology* **22**, 53–66 (2012).
69. Chung, S., Lu, Y. & Henry, R. G. Comparison of bootstrap approaches for estimation of uncertainties of dti parameters. *NeuroImage* **33**, 531–541 (2006).
70. Whitcher, B., Tuch, D. S., Wisco, J. J., Sorensen, A. G. & Wang, L. Using the wild bootstrap to quantify uncertainty in diffusion tensor imaging. *Hum. brain mapping* **29**, 346–362 (2008).
71. Zhu, T., Liu, X., Connelly, P. R. & Zhong, J. An optimized wild bootstrap method for evaluation of measurement uncertainties of dti-derived parameters in human brain. *Neuroimage* **40**, 1144–1156 (2008).

## **Acknowledgements**

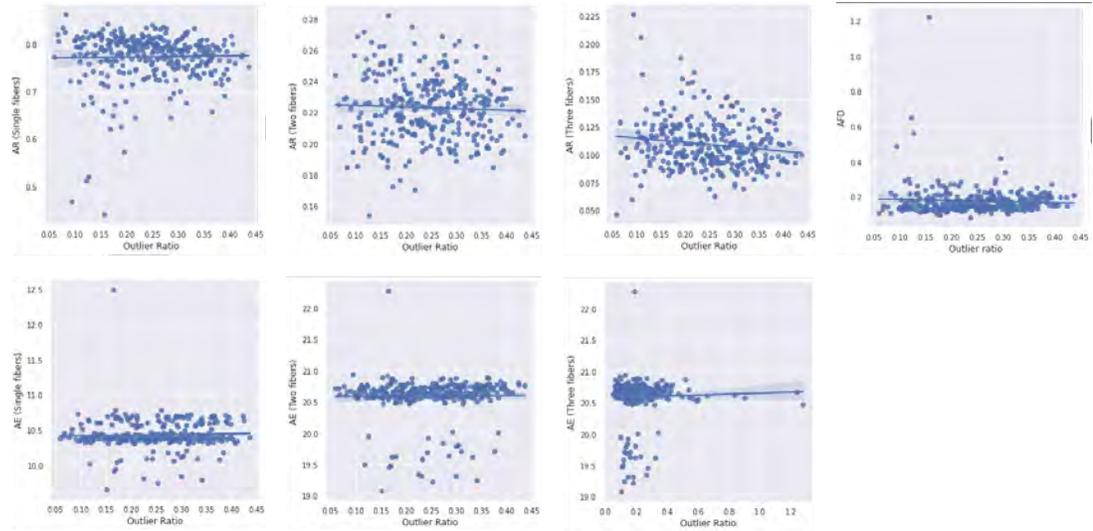
This work was supported by the Swiss National Science Foundation (project 205321-182602). We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), Ecole polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE) and Geneva University Hospitals (HUG).

This research was supported in part by the National Institute of Biomedical Imaging and Bioengineering, the National Institute of Neurological Disorders and Stroke, and Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (NIH) under award numbers R01HD110772, R01NS106030, R01EB031849, R01EB032366, and R01HD109395; in part by the Office of the Director of the NIH under award number S10OD025111; and in part by the National Science Foundation (NSF) under grant number 212306. This research was also partly supported by NVIDIA Corporation and utilized NVIDIA RTX A6000 and RTX A5000 GPUs. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, NSF, or NVIDIA.

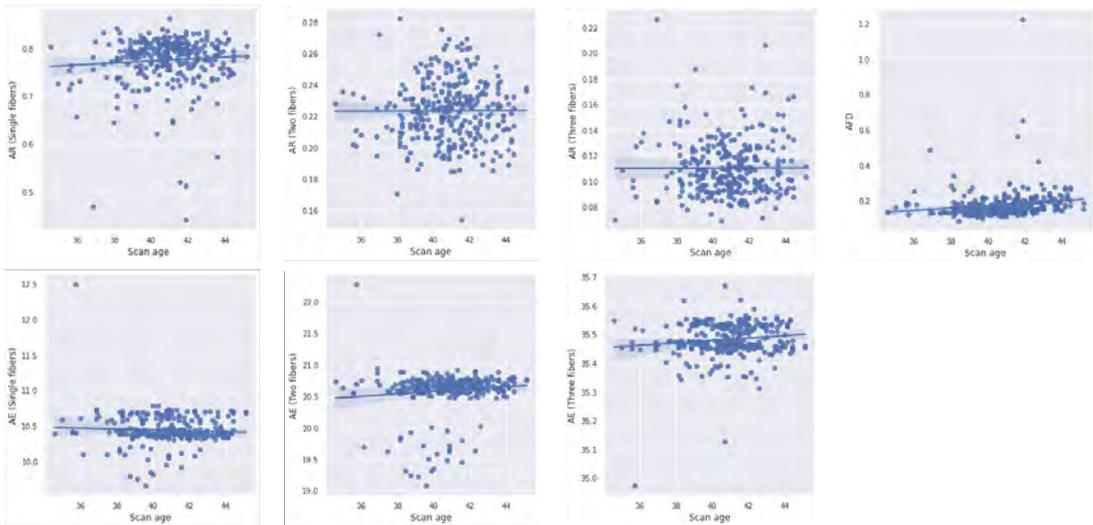
We would like to thank Dr. Maria Deprez and Dr. Daan Christiaens from King's College London (London, United Kingdom) for their support in answering questions about the dHCP dataset. We would also like to thank Dr. Farnaz Delavari for her help in the design of the figures and statistical analysis.

## **Supplementary Material**

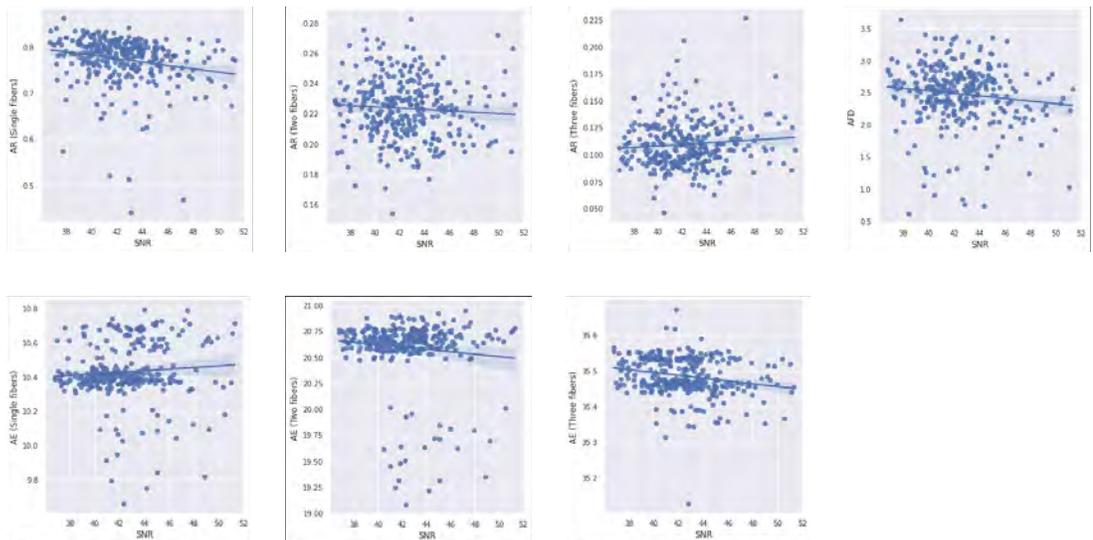
Outlier Ratio



Scan Age



SNR



**Figure 10.** Agreement rate (AR) and angular error (AE) for single, two and three fibers, and apparent fiber density (AFD) error for the deep learning method ( $DL_n$ ), as a function of quality control (QC) metrics (Outlier ratio, Signal-to-noise ratio) and scan age for  $N=320$  subjects. No correlation is generally observed between the QC metrics and the error rates.

# Robust estimation of the microstructure of the early developing brain using deep learning

Hamza Kebiri<sup>1,2,3</sup>, Ali Gholipour<sup>3</sup>, Rizhong Lin<sup>2,4</sup>, Lana Vasung<sup>5</sup>, and Davood Karimi<sup>3,\*</sup>, Meritxell Bach Cuadra<sup>1,2,\*</sup>

<sup>1</sup> CIBM Center for Biomedical Imaging, Switzerland

<sup>2</sup> Department of Radiology, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland

<sup>3</sup> Computational Radiology Laboratory, Department of Radiology, Boston Children’s Hospital and Harvard Medical School, Boston, Massachusetts, USA

<sup>4</sup> Signal Processing Laboratory 5 (LTS5), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

<sup>5</sup> Department of Pediatrics, Boston Children’s Hospital, and Harvard Medical School, Boston, Massachusetts, USA

**Abstract.** Diffusion Magnetic Resonance Imaging (dMRI) is a powerful non-invasive method for studying white matter tracts of the brain. However, accurate microstructure estimation with fiber orientation distribution (FOD) using existing computational methods requires a large number of diffusion measurements. In clinical settings, this is often not possible for neonates and fetuses because of increased acquisition times and subject movements. Therefore, methods that can estimate the FOD from reduced measurements are of high practical utility. Here, we exploited deep learning and trained a neural network to directly map dMRI data acquired with as low as six diffusion directions to FODs for neonates and fetuses. We trained the method using target FODs generated from densely-sampled multiple-shell data with the *multi-shell multi-tissue constrained spherical deconvolution* (MSMT-CSD). Detailed evaluations on independent newborns’ test data show that our method achieved estimation accuracy levels on par with the state-of-the-art methods while reducing the number of required measurements by more than an order of magnitude. Qualitative assessments on two out-of-distribution clinical datasets of fetuses and newborns show the consistency of the estimated FODs and hence the cross-site generalizability of the method.

## 1 Introduction

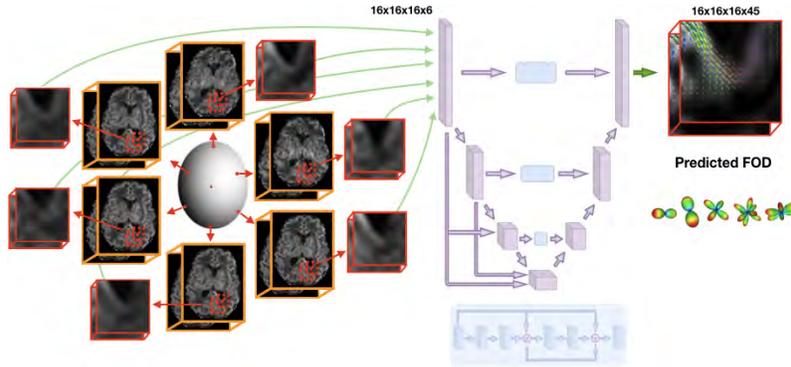
Depiction of white matter fiber tracts is of paramount importance for brain characterization in health and disease. Diffusion-weighted magnetic resonance imaging (dMRI) is the method of choice to study axon bundles that connect different brain regions. Several models have been proposed to map the 4-dimensional diffusion signal to objects such as tensors or fiber orientation distribution functions (FODs) [23,32], which can be further processed to compute metrics such

---

\* Both authors contributed equally.

as tract orientation and apparent fiber density [16,26]. Model-based FODs are the mathematical frameworks of choice for microstructure estimation. In fact, FODs accurately describe the underlying microstructure by a deformed sphere in which different radii correspond to different intra-voxel fibers. Moreover, without FODs, tracking stops prematurely and favors shorter fiber tracts [8]. Standard FOD estimation methods [23,32,1,15,27] process voxels individually and thus do not exploit correlations between neighboring voxels. As a result, these methods demand dMRI measurements with multiple b-values and a high number of gradient directions to account for the response function of each tissue type [15]. These acquisitions require prolonged scans that are not affordable for newborn and fetal subjects because of the sensitivity of these cohorts and the increased risk of motion. Acquisitions have to be fast to freeze in-plane motion; yet data dropout rates are high in these cohorts because of motion artifacts. Reconstructing FODs in developing brains has been performed [30,5,7,4] using high-quality datasets and rich information including several gradient directions, higher and/or multiple b-values, and high signal-to-noise ratio (3 Tesla magnetic field strength). Additionally, the datasets were acquired in a controlled and uniform research setting with healthy volunteers, which can hardly be reproduced in the clinical environment. Moreover, and in contrast to adult brains, anisotropy increases in white matter fibers of developing brains because of increased water volume and poor alignment of the fibers [6]. Gray matter on the other hand, during early gestational weeks, is highly anisotropic because of the complexity of the formation of cell bodies, glial cells, and the different neuronal structures [6]. This dynamic period for microstructure [9,2] makes FOD estimation a more challenging task. Adaptive learning-based methods can be leveraged to learn from high-quality datasets and exploit this knowledge in clinical routine acquisitions. Deep learning models, first suggested in [12], promise to overcome the error accumulation of suboptimal processing steps that are characteristic of standard estimation techniques. This end-to-end learning paradigm has been then applied in dMRI for several purposes [24,22,25,18,13]. The authors in [24] have accurately predicted tensor maps with six diffusion measurements. In [22], a 2D convolutional neural network (CNN) was used to predict the orientation of the fibers in a classification approach whereas [25] deployed a 3D CNN to predict FODs using a small neighborhood of the diffusion signal. In [18], a feedforward neural network was used to predict the FODs and found that 44 directions can be sufficient. However, the network does not exploit neighboring voxels correlations. A more recent work [13] used a Transformer-CNN block to first map 200 to 60 directions and the latter to FODs. However, for uncooperative cohorts such as neonates or fetuses, this number of measurements is unrealistic to acquire. To the best of our knowledge, no learning-based method to predict FODs has been reported for newborn and fetal brains. In this work, we demonstrate that a deep convolutional neural network with a large field of view (FOV) can accurately estimate FODs using only 6-12 diffusion-weighted measurements. Our contribution is three-fold. We first show that a deep learning method can achieve an accuracy level that is comparable with the agreement between the state-of-

the-art methods, while drastically reducing the number of measurements, for developing brains. We then show a low agreement between state-of-the-art methods in terms of different metrics using data from a highly controlled setting, namely the developing Human Connectome Project (dHCP). This addresses the need to build reproducible and reliable pipelines for white matter characterization [29], particularly for the developing brain. Finally, we demonstrate the generalizability of our method on two clinical datasets of fetuses and newborns that were acquired in completely different settings than those used in the training data.



**Fig. 1.** The proposed framework to predict Fiber Orientation Distribution (FOD) functions in the spherical harmonics domain (SH- $L_{max}$  order 8). The network takes 3D input patches from 6 diffusion measurements and outputs patches of SH coefficients. Network architecture details can be found in Figure S1 of Supplementary materials and code at<sup>1</sup>.

## 2 Methodology

### 2.1 Paradigm

The method is based on directly learning a mapping between the raw diffusion signal and the FOD in a supervised manner (Figure 1). Our model inputs are respectively 6 single-shell ( $b = 1000 \text{ s/mm}^2$ ) measurements for the neonate network ( $DL_n$ ) and 12 single-shell ( $b = 400 \text{ s/mm}^2$ ) measurements for the fetal inference network ( $DL_f$ ), trained on pre-term subjects (as in [17,19]). To be independent on gradient directions, projection of the signal onto spherical harmonics basis (SH) (SH- $L_{max}$  order 2 and 4 respectively for the two networks) was performed to predict the FOD represented in the SH basis (SH- $L_{max}$  order 8). To train the model, these target coefficients are estimated from 300 multi-shell measurements using MSMT-CSD [15]. These measurements are distributed

<sup>1</sup> [https://github.com/Medical-Image-Analysis-Laboratory/Perinatal\\_fODF\\_DL\\_estimation](https://github.com/Medical-Image-Analysis-Laboratory/Perinatal_fODF_DL_estimation)

over 3 shells of  $\{400, 1000, 2600\}$   $s/mm^2$  with 64, 88 and 128 samples, respectively, and 20  $b_0$  ( $b = 0$   $s/mm^2$ ) images. The few measurements used as input to the model were based on the scheme in [31], whereby the gradient directions minimized the condition number of the diffusion tensor reconstruction matrix.

## 2.2 Data processing

**dHCP Newborns** - We have selected two subsets from the developing Human Connectome Project (dHCP) dataset (1) 100 subjects (weeks: [32.1, 44.7], mean: 40, standard deviation: 2.4) and (2) 68 pre-term subjects (weeks: [29.3, 37.0], mean: 34.9, standard deviation: 1.8). The data was acquired with a 3T Philips Achieva scanner in a multi-shell scheme ( $b \in \{0, 400, 1000, 2600\}$   $s/mm^2$ ) [14] and was denoised, motion and distortion corrected [3]. It has a final resolution of  $1.17 \times 1.17 \times 1.5$   $mm^3$  in a field of view of  $128 \times 128 \times 64$  voxels. We have upsampled the data to 1 mm isotropic resolution to account for network isotropic 3D patches. We have additionally normalized the input data by  $b_0$ . A white matter mask was generated using the union of the *White Matter* and the *Brainstem* labels provided by the dHCP, and the voxels where Fractional Anisotropy (FA) was higher than 0.25. A resampling of the dHCP labels from T2-w resolution (0.5  $mm^3$  isotropic) to 1  $mm^3$  resolution was performed.

**Clinical newborns & fetuses** - Acquisitions of 8 neonates ([38.1, 39.4, 40.1, 40.4, 40.7, 40.9, 41.8, 42] weeks), were performed during natural sleep at 3T (Siemens Trio and Skyra). Five  $b_0$  images and 30  $b = 1000$   $s/mm^2$  were acquired. The TR-TE were 3700-104 ms and voxel size was 2 mm isotropic. Eight fetal subjects ([24, 25, 26.3, 26.6, 26.7, 26.9, 29.4, 38.7] gestational weeks, GW) were scanned using a 3T Siemens Skyra MRI scanner (TR=3000–4000 ms, TE=60 ms) with one  $b_0$  and 12 diffusion-sensitized images at  $b = 500$   $s/mm^2$ . All subjects were processed for noise [34] and bias field inhomogeneities [33]. Rigid registration to a T2 atlas [11] was performed and b-vectors were rotated accordingly for fetal data. The different volumes were upsampled to 1  $mm^3$  and normalized by  $b_0$ . The studies were approved by the institutional review board committee.

## 2.3 Training

Two networks,  $DL_n$  and  $DL_f$  (see subsection 2.1 above), were trained using Adam optimizer [21] to minimize the  $\ell_2$  norm loss function between the predicted 45 SH coefficients and the ground truth FOD SH coefficients generated using the 300 directions and the 4 b-values ( $\{0, 400, 1000, 2600\}$   $s/mm^2$ ), i.e.

$$\text{minimize } \sum_{i=1}^{45} \left\| FOD_i^{pred} - FOD_i^{GT} \right\|^2$$

We used 70% of the subjects for training, 15% for validation, and 15% for testing. We used the number of FOD peaks (extracted from Dipy [10]) to balance patch selection per batch. The central voxel of each patch was constrained to be in the generated white matter mask and to be 1 peak in  $\frac{2}{3}$  of the batch and more than one peak in  $\frac{1}{3}$  of the batch. This condition implicitly guarantees the non

selection of empty patches. The patch size was empirically set to  $16^3$  voxels. The batch size was set to 27 for  $DL_n$  and 9 for  $DL_f$ , and the initial learning rate to  $10^{-4}$  and was decreased by 0.9 whenever the validation loss did not improve after one epoch. The total number of training epochs was 10000 and a dropout rate of 0.1 was used in all layers to reduce overfitting and improve generalization. In  $DL_f$ , Gaussian noise injection (mean=0, sigma=0.025) was applied as well as small rotations (uniformly from  $[-5^\circ, +5^\circ]$ ) to make the model robust to minor uncorrected movements due to small differences in FOV and fetal head motion.

## 2.4 Evaluation of dHCP newborns

**Comparison with state-of-the-art methods** - In addition to comparing our network ( $DL_n$ ) prediction with FODs estimated using MSMT-CSD of 300 directions (considered as ground truth, GT), we have assessed the agreement between two mutually exclusive subsets extracted from the ground truth (gold standards 1 and 2, respectively GS1 and GS2). Each subset contains 150 directions ( $b \in \{0, 400, 1000, 2600\}$  s/mm<sup>2</sup>) with respectively 10, 32, 44, and 64 measurements (half measurements of GT data). GS1 and GS2 subsets can be considered as independent high-quality scans, and differences in terms of subsequent metrics can be considered as an upper bound error for the different methods deployed. Furthermore, we have computed three state-of-the-art methods: (1) Constrained spherical deconvolution (CSD) [32] using the 128 gradient directions of the highest shell, i.e.  $b = 2600$  s/mm<sup>2</sup> and 20 b0 images; (2) Constant Solid Angle ODF (Q-Ball) model [1] that we refer to as CSA and (3) the Sparse Fascicle Model (SFM) [27] model for which we have used the default regularization parameters. We also compared our method with the multilayer perceptron (MLP) in [18], which has been shown to outperform the method of [25].

**Error metrics** - Quantitative validation was performed based on the number of peaks, the angular error and the apparent fiber density (AFD) [26]. The number of peaks was generated from the FOD predicted by the network and the ones estimated by the different methods (GT, GS1, GS2, CSD, CSA and SFM) using the same parameters (mean separation angle of  $45^\circ$ , a maximum number of 3 peaks and relative peak threshold of 0.5). The conservative choice of these parameters was guided by [28] which shows the limitations of current dMRI models at depicting multiple number of peaks and low angular crossing fibers. We have compared these models in terms of confusion matrices, and the *agreement rate* (AR). AR is defined for each number of peaks  $p$  as:  $AR = \frac{A_p}{\Sigma D_p}$  where  $A_p$  is the percentage of voxels on which both methods agree on  $p$  number of peaks and  $D_p$  the percentage of voxels where at least one of the two methods predicts  $p$  and the other  $p'$  where  $p \neq p'$ . For the voxels containing the same number of peaks, we have computed the angular error with respect to the GT, as well as between GS1 and GS2. For voxels with multiple fibers, we have first extracted corresponding peaks between the two methods by computing the minimum angle between all configurations (4 for 2 peaks and 9 for 3 peaks); we then removed these peaks and recursively apply the same algorithm. We have also compared AFD, that is defined as the FOD amplitude. AFD was extensively demonstrated

as a biologically plausible measure that is not only sensitive to the fiber partial volume fraction but also to fiber density or membrane permeability [26]. Statistical validation using paired t-test corrected for multiple comparisons with Bonferroni method was performed between the errors of the different methods with respect to GT and the difference between GS1 and GS2.

## 2.5 Evaluation of clinical datasets

$DL_f$  was tested on fetal volumes whereas  $DL_n$  was tested on the clinical newborn dataset. Due to the lack of ground truth for both clinical datasets, we qualitatively assess the network predictions with 12 and 6 measurements, as compared to CSD using all available measurements (SH- $L_{max}$  order 4 and 8), respectively for fetuses and newborns.

## 3 Results

The networks consistently learned a mapping between the six/twelve diffusion measurements and the ground truth FOD constructed with 300 measurements across 4 b-values, as evaluated on the independent test data (Figures 2 and 3).

### 3.1 In-domain quantitative evaluation in newborns dHCP

**Number of peaks** - We first observe a low agreement (AR) between the two gold standard acquisitions (GS1 vs. GS2), that is more pronounced for multiple fibers voxels. For instance, 1-peaks AR is 80.4%, 30.3% for 2-peaks and 27.9% for 3-peaks. SFM achieves a relatively high 1-peaks agreement with the GT of 83% and the lowest with multiple fibers voxels (10% and 3.5% for 2- and 3-peaks, respectively). In contrast, CSD estimates a high number of multiple fibers (16.5% and 5.9% for 1- and 2-peaks respectively) and achieves the lowest 1-peaks AR with 11.7%. In fact, the latter is biased towards multiple peaks estimation with more than 90% of the voxels modeled as either two or three peaks. This might be explained by the high b-value ( $b = 2600 \text{ s/mm}^2$ ) containing high levels of noise. Our method,  $DL_n$ , achieves an agreement for 1-, 2- and 3-peaks of respectively 79%, 16% and 3% that is globally the closest to the agreement between the gold standards when compared to other methods. We believe that the relatively low agreement for multiple intravoxel fiber orientations is due to their incongruence across GT subjects, and hence the absence of a consistent pattern to be learned by the neural network. In fact, this is supported by the modest agreement between the two gold standards ( $\Delta$ GS), in which both the subjects and the number of measurements are the same, only the gradient directions vary and already result in a drop of 70% in multiple fibers depiction. It is worth noting that the agreement between the different methods (CSD vs. CSA, SFM vs. CSA, CSD vs.  $DL_n$ , etc) was also low. The confusion matrices for  $\Delta$ GS agreement and the different methods can be found in Table S1 and the comparison with [18] in Section 3 of Supplementary materials.

Method	b-values ( $s/mm^2$ )	$N_m$	Angular error (Agreement rate in %)			AFD error
			Single fibers	Two fibers	Three fibers	
$DL_n$	{0, 1000}	7	12.6°(78.4%)	24.2°(15.8%)	33.3°(3.8%)	0.27 ( $\pm 0.03$ )
CSD	{0, 2600}	148	7.5°(11.7%)	16.5°(16.5%)	27.2°(5.9%)	1.31 ( $\pm 0.23$ )
CSA	{0, 400, 1000, 2600}	300	47.0°(27.7%)	41.4°(14.8%)	36.1°(7.9%)	3.46 ( $\pm 0.46$ )
SFM	{0, 400, 1000, 2600}	300	51.4°(83%)	40.7°(10%)	35.4°(3.5%)	0.80 ( $\pm 0.55$ )
$\Delta GS$	{0, 400, 1000, 2600}	150	13.8°(80.4%)	29.1°(30.3%)	35.4°(27.9%)	0.2 ( $\pm 0.025$ )

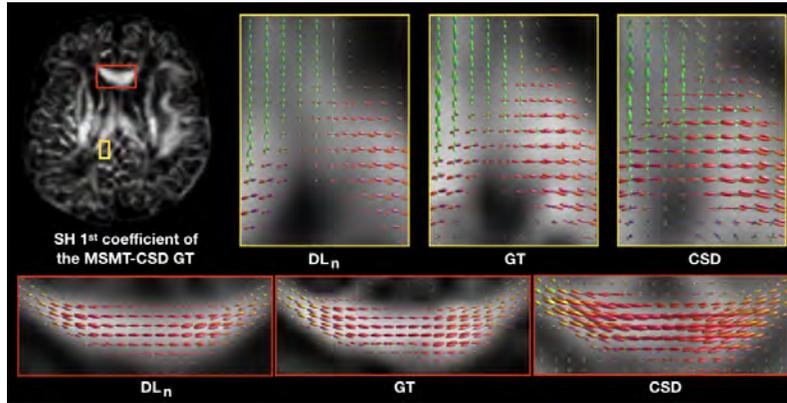
**Table 1.** Mean angular error, agreement rate on number of peaks and Apparent Fiber Density (AFD) error between GT (MSMT-CSD) and the different methods.  $\Delta GS$  refers to GS1 and GS2 agreements. The number of measurements ( $N_m$ ) and the b-values used are also reported. All results were statistically significant compared to  $\Delta GS$  ( $p \leq 9e^{-10}$  for angular error, except SFM three fibers, and  $p \leq 4.5e^{-3}$  for AFD error).

**Angular error** - The agreement in terms of the number of peaks does not guarantee that the fibers follow the same orientation. Table 1 shows the angular error and the agreement rate (AR) in numbers of peaks for the different configurations. In GS1 and GS2, the angular difference increases almost linearly for one, two and three fibers. Our learning model achieves an error rate that is comparable (although statistically different,  $p \leq 9e^{-10}$ ) to GS1 and GS2. SFM and CSA achieve a higher error rate for single and two fiber voxels, whereas CSD achieves the lowest. This is because of the low AR and hence the error is computed among a small subset of common voxels between the GT and CSD as shown in Table 1. It is worth mentioning that using 15 directions instead of 6 as input to the network did not improve the results; and in general, these angular errors are higher than those reported for adult data, such as the Human Connectome Project as in [13]. We hypothesize this can be due to immature and high variability of the developing brain anatomy.

**Apparent Fiber Density** - The last column in Table 1 shows the differences between AFD averaged over the 15 test subjects. Our model achieves the closest error rate of 0.27 ( $\pm 0.03$ ) to the GT compared with the gold standards difference of 0.2 ( $\pm 0.025$ ), in terms of mean and standard deviation. The other methods have an increased error rate compared to  $DL_n$  with factors of around 2.5, 4.5 and 9.5-fold for SFM, CSD and CSA respectively. Results were statistically significant ( $p \leq 4.5e^{-3}$ ) compared to the agreement between the gold standard models.

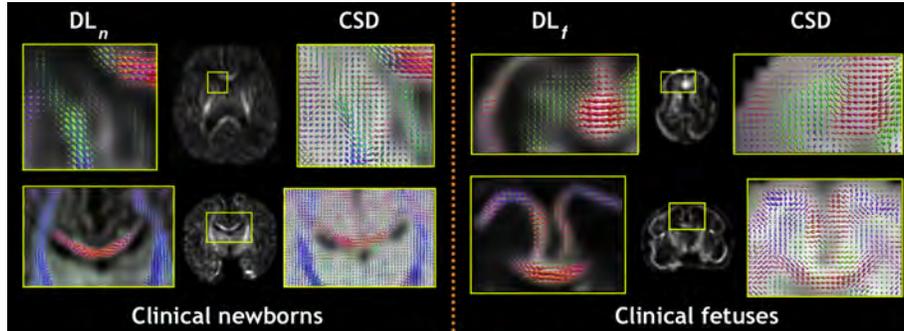
### 3.2 Generalizability to clinical acquisitions (newborns and fetuses)

$DL_f$  successfully generalized to fetal data as can be seen in Figure 3 (right) for two subjects. Callosal fibers are clearly delineated on the top and bottom subjects. The radial coherence of cortical plate at early gestation [20] is also highlighted on the same panels. Similarly,  $DL_n$  generalized to the new newborn dataset (Figure 3, left), despite differences in scanner and protocol. Both cortico-spinal tract and corpus callosum are shown in the bottom subject. As opposed



**Fig. 2.** Qualitative comparison between the deep learning method  $DL_n$ , the MSMT-CSD GT and CSD in two brain regions of a newborn dHCP subject.

to CSD that overestimated false positive crossing fibers, likely due to residual noise, the deep learning method trained on MSMT-CSD directly produced low amplitude FODs in isotropic or non-consistent regions. The results for six other subjects can be found in Figure S2 in Supplementary materials.



**Fig. 3.** The deep learning method compared to CSD in different brain regions for 2 newborn subjects (left) and 2 fetal subjects (right) of 25 (top) and 29.4 (bottom) weeks of gestation. FODs are superimposed to the first SH coefficient of the method used.

## 4 Conclusion

We have demonstrated how a deep neural network can successfully reconstruct high angular multi-shell FODs from a reduced number (6 to 12) of diffusion measurements. The substantially lower number of samples is compensated by learning from high-quality training data and by exploiting the spatial neighborhood

information. The network was quantitatively evaluated on the dHCP dataset which was acquired in a highly controlled setting that cannot be reproduced in clinical settings. We showed that our method relying on six measurements can be leveraged to reconstruct plausible FODs of clinical newborn and fetal brains. We compared our model to commonly used methods such as CSD and MSMT-CSD between two gold standard datasets. The results exhibit low agreements between the different methods, particularly for multiple fiber orientations, despite using high angular multi-shell data. This highlights the need to build robust and reproducible methods for microstructure estimation in developing brains.

## Acknowledgment

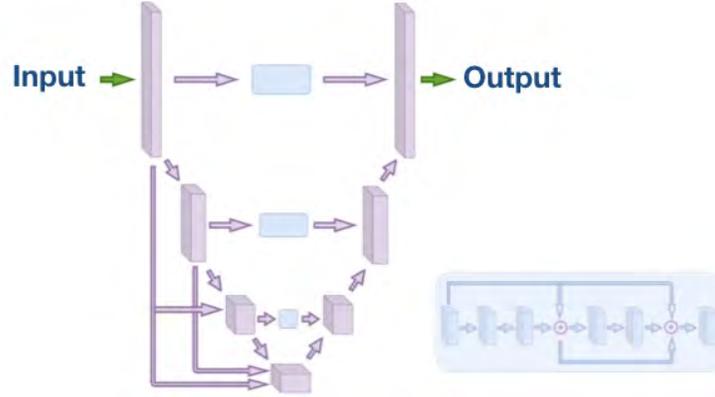
This work was supported by the Swiss National Science Foundation (project 205321-182602). We acknowledge the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by CHUV, UNIL, EPFL, UNIGE, HUG and the Leenaards and Jeantet Foundations. This research was also partly supported by the US National Institutes of Health (NIH) under awards R01NS106030 and R01EB032366; by the Office of the Director of the NIH under award S10OD0250111; and by NVIDIA Corporation; and utilized an NVIDIA RTX A6000 GPU.

## References

1. Aganj, I., et al.: Reconstruction of the orientation distribution function in single- and multiple-shell q-ball imaging within constant solid angle. *Magnetic resonance in medicine* **64**(2), 554–566 (2010)
2. Andescavage, N.N., et al.: Complex trajectories of brain development in the healthy human fetus. *Cerebral Cortex* **27**(11), 5274–5283 (2017)
3. Bastiani, M., et al.: Automated processing pipeline for neonatal diffusion mri in the developing human connectome project. *Neuroimage* **185**, 750–763 (2019)
4. Chen, R., et al.: Deciphering the developmental order and microstructural patterns of early white matter pathways in a diffusion mri based fetal brain atlas. *NeuroImage* **264**, 119700 (2022)
5. Christiaens, D., et al.: Multi-shell shard reconstruction from scattered slice diffusion mri data in the neonatal brain. *ISMRM (Paris)* (2018)
6. Counsell, S.J., Arichi, T., Arulkumaran, S., Rutherford, M.A.: Fetal and neonatal neuroimaging. *Handbook of clinical neurology* **162**, 67–103 (2019)
7. Deprez, M., et al.: Higher order spherical harmonics reconstruction of fetal diffusion mri with intensity correction. *IEEE transactions on medical imaging* **39**(4) (2019)
8. Descoteaux, M.: High angular resolution diffusion MRI: from local estimation to segmentation and tractography. Ph.D. thesis, Univ. Nice Sophia Antipolis (2008)
9. Dubois, J., et al.: The early development of brain white matter: a review of imaging studies in fetuses, newborns and infants. *neuroscience* **276**, 48–71 (2014)
10. Garyfallidis, E., et al.: Dipy, a library for the analysis of diffusion mri data. *Frontiers in neuroinformatics* **8** (2014)
11. Gholipour, A., et al.: A normative spatiotemporal mri atlas of the fetal brain for automatic segmentation and analysis of early brain growth. *Scientific reports* **7**(1), 476 (2017)

12. Golkov, V., et al.: Q-space deep learning: twelve-fold shorter and model-free diffusion mri scans. *IEEE TMI* **35**(5), 1344–1351 (2016)
13. Hosseini, S., et al.: Cttrack: A cnn+ transformer-based framework for fiber orientation estimation & tractography. *Neuroscience Informatics* **2**(4), 100099 (2022)
14. Hutter, J., et al.: Time-efficient and flexible design of optimized multishell hardi diffusion. *Magnetic resonance in medicine* **79**(3), 1276–1292 (2018)
15. Jeurissen, B., et al.: Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion mri data. *Neuroimage* **103**, 411–426 (2014)
16. Jeurissen, B., et al.: Diffusion mri fiber tractography of the brain. *NMR in Biomedicine* **32**(4), e3785 (2019)
17. Karimi, D., et al.: Deep learning-based parameter estimation in fetal diffusion-weighted mri. *Neuroimage* **243**, 118482 (2021)
18. Karimi, D., et al.: Learning to estimate the fiber orientation distribution function from diffusion-weighted mri. *Neuroimage* **239**, 118316 (2021)
19. Kebiri, H., et al.: Slice estimation in diffusion mri of neonatal and fetal brains in image and spherical harmonics domains using autoencoders. In: *International Workshop on Computational Diffusion MRI, MICCAI, Singapore*. Springer (2022)
20. Khan, S., et al.: Fetal brain growth portrayed by a spatiotemporal diffusion tensor mri atlas computed from in utero images. *Neuroimage* **185**, 593–608 (2019)
21. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
22. Koppers, S., Merhof, D.: Direct estimation of fiber orientations using deep learning in diffusion imaging. In: *International Workshop on Machine Learning in Medical Imaging*. pp. 53–60. Springer (2016)
23. Le Bihan, D., et al.: Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging* **13**(4), 534–546 (2001)
24. Li, H., et al.: Superdti: Ultrafast dti and fiber tractography with deep learning. *Magnetic resonance in medicine* **86**(6), 3334–3347 (2021)
25. Lin, Z., et al.: Fast learning of fiber orientation distribution function for mr tractography using convolutional neural network. *Medical physics* **46**(7) (2019)
26. Raffelt, D., et al.: Apparent fibre density: a novel measure for the analysis of diffusion-weighted magnetic resonance images. *Neuroimage* **59**(4) (2012)
27. Rokem, A., et al.: Evaluating the accuracy of diffusion mri models in white matter. *PloS one* **10**(4), e0123272 (2015)
28. Schilling, K.G., et al.: Histological validation of diffusion mri fiber orientation distributions and dispersion. *Neuroimage* **165**, 200–221 (2018)
29. Schilling, K.G., et al.: Prevalence of white matter pathways coming into a single white matter voxel orientation: The bottleneck issue in tractography. *Human brain mapping* **43**(4), 1196–1213 (2022)
30. Shen, K., et al.: A spatio-temporal atlas of neonatal diffusion mri based on kernel ridge regression. In: *International Symposium on Biomedical Imaging (ISBI)* (2017)
31. Skare, S., et al.: Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with mri. *Journal of magnetic resonance* **147**(2), 340–352 (2000)
32. Tournier, J.D., et al.: Direct estimation of the fiber orientation density function from diffusion-weighted mri data using spherical deconvolution. *Neuroimage* **23**(3), 1176–1185 (2004)
33. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., et al.: N4itk: improved n3 bias correction. *IEEE transactions on medical imaging* **29**(6), 1310–1320 (2010)
34. Veraart, J., Fieremans, E., Novikov, D.S.: Diffusion mri noise mapping using random matrix theory. *Magnetic resonance in medicine* **76**(5), 1582–1593 (2016)

## 1 Network architecture



**Fig. S1.** The 3D network architecture is inspired by U-Net [Ronneberger et al., 2015]. Two modifications have been performed to the conventional U-Net to adapt it to the low number of measurements supplied in the input: (1) extensive short and long-range residual connections to provide more context to subsequent layers, and (2) the substitution of max-pooling operations in the contracting path by stride-2 convolutions to make the downsampling a learnable step and hence layer-specific. Convolutions were followed by Rectified Linear Unit (ReLU) activation functions and the number of feature maps in the first block was set to 36.

## 2 More results: dHCP newborns

$\Delta$ GS			CSA			CSD			SFM			$DL_n$			
1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	
1	68.9	7.62	0.79	22.5	35	21.2	9.32	9.68	60.7	81.3	1.77	0.32	74.2	4.82	0.66
2	7.7	8.34	2.02	2.08	7.73	5.38	0.22	1.47	13.5	11.1	1.47	0.17	11.5	3.38	0.33
3	0.73	1.86	2.09	0.55	2.06	2.51	0.02	0.17	4.93	3.41	0.26	0.15	3.49	1.4	0.23

**Table S1.** Confusion matrices for number of peaks agreement (in %), normalized over all population. From left to right: gold standards GS1 vs. GS2, followed by the different methods CSA, CSD, SFM and  $DL_n$  compared to the GT MSMT-CSD. Each confusion matrix reports the average result for 15 test subjects.

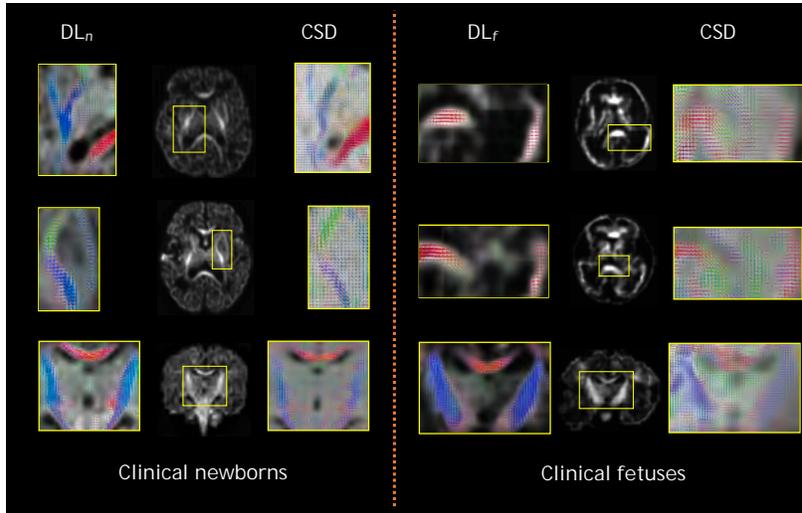
### 3 More results: comparison with a deep learning method

We have carried out a further comparison of our method against the multilayer perceptron (MLP) approach proposed by Karimi et al., 2021 on the same dHCP data. To avoid overfitting, given the smaller amount of data in our study, we have adjusted the original architecture empirically to  $[n_{sig}, 40, 40, 40, 50, 60, 70, n_{tar}]$ . For our implementation, the parameters  $n_{sig}$  and  $n_{tar}$  were set to 6 and 45, respectively. The results detailed in Table S2 indicate that the performance of the MLP method was lower than that achieved by our proposed method.

Method	b-values ( $s/mm^2$ )	$N_m$	Angular error (Agreement rate in %)			AFD error
			Single fibers	Two fibers	Three fibers	
$DL_n$	{0, 1000}	7	12.6°(78.4%)	24.2°(15.8%)	33.3°(3.8%)	0.27 ( $\pm 0.03$ )
MLP	{0, 1000}	7	49.9°(65%)	39.4°(11.8%)	37.1°(4.9%)	0.67 ( $\pm 0.12$ )

**Table S2.** Mean angular error, agreement rate on number of peaks and Apparent Fiber Density (AFD) error between our method ( $DL_n$ ) and the MLP.

### 4 More results: Clinical newborns and fetuses



**Fig. S2.** The deep learning method compared to CSD in different brain regions for 3 newborn subjects (left) and 3 fetal subjects (right) of 26.4, 26.6 and 38.9 gestational weeks of respectively top, middle and bottom rows. FODs are superimposed to the first SH coefficient of the method used. The DL first SH coefficient is shown in full-size too.

# Quantitative evaluation of enhanced multi-plane clinical fetal diffusion MRI with a crossing-fiber phantom

Hamza Kebiri<sup>1,2</sup>, H el ene Lajous<sup>1,2</sup>, Yasser Alem an-G omez<sup>1</sup>, Gabriel Girard<sup>3</sup>,  
Erick Canales Rodr iguez<sup>3</sup>, S ebastien Tourbier<sup>1</sup>, Marco Pizzolato<sup>3,4</sup>,  
Jean-Baptiste Ledoux<sup>1,2</sup>, Eleonora Fornari<sup>1,2</sup>, Andr as Jakab<sup>5,6</sup>, and Meritxell  
Bach Cuadra<sup>2,1,3</sup>

<sup>1</sup> Department of Radiology, Lausanne University Hospital and University of  
Lausanne, Lausanne, Switzerland

<sup>2</sup> CIBM Center for Biomedical Imaging, Switzerland

<sup>3</sup> Signal Processing Laboratory 5 (LTS5), Ecole Polytechnique F ed erale de Lausanne  
(EPFL), Lausanne, Switzerland

<sup>4</sup> Department of Applied Mathematics and Computer Science, Technical University  
of Denmark, Kgs. Lyngby, Denmark

<sup>5</sup> Center for MR Research, University Children’s Hospital Zurich, Zurich, Switzerland

<sup>6</sup> Neuroscience Center Zurich, University of Zurich/ETH Zurich, Zurich, Switzerland

**Abstract.** Diffusion Magnetic Resonance Imaging (dMRI) has become widely used to study *in vivo* white matter tissue properties non-invasively. However, fetal dMRI is greatly limited in Signal-to-Noise ratio and spatial resolution. Due to the uncontrollable fetal motion, echo planar imaging acquisitions often result in highly degraded images, hence the ability to depict precise diffusion MR properties remains unknown. To the best of our knowledge, this is the first study to evaluate diffusion properties in a fetal customized crossing-fiber phantom. We assessed the effect of scanning settings on diffusion quantities in a phantom specifically designed to mimic typical values in the fetal brain. Orthogonal acquisitions based on clinical fetal brain schemes were preprocessed for denoising, bias field inhomogeneity and distortion correction. We estimated the fractional anisotropy (FA) and mean diffusivity (MD) from the diffusion tensor, and the fiber orientations from the fiber orientation distribution function. Quantitative evaluation was carried out on the number of diffusion gradient directions, different orthogonal acquisitions, and enhanced 4D volumes from scattered data interpolation of multiple series. We found out that while MD does not vary with the number of diffusion gradient directions nor the number of orthogonal series, FA is slightly more accurate with more directions. Additionally, errors in all scalar diffusion maps are reduced by using enhanced 4D volumes. Moreover, reduced fiber orientation estimation errors were obtained when used enhanced 4D volumes, but not with more diffusion gradient directions. From these results, we conclude that using enhanced 4D volumes from multiple series should be preferred over using more diffusion gradient directions in clinical fetal dMRI.

**Keywords:** Fetal· MRI· Brain· Phantom· Diffusion Tensor Imaging· Orientation Distribution Function· Scattered Data Interpolation

## 1 Introduction

Diffusion Magnetic Resonance Imaging (dMRI) has been the mainstay of non-invasive white matter investigation *in vivo*. As the diffusion signal is sensitive to the displacement of water molecules in brain tissues, various biophysical models have been proposed for estimating the underlying tissue architecture. These models can either be Gaussian, e.g., diffusion tensor imaging (DTI) is the most simple and widely used model to characterize the diffusion process, or non-Gaussian, e.g., q-ball imaging [31], diffusion spectrum imaging [34,2], and spherical deconvolution [30,3], which estimate Orientation Distribution Functions (ODFs) for resolving multiple intravoxel fiber orientations. However, the unavailability of a ground truth makes the quantitative validation of these models an elusive goal. Monkey brains have been used for connectivity validation of dMRI when compared to histological connectivity obtained from viral tracer injections [1]. Nevertheless, a direct comparison of diffusion orientations at the voxel level is challenging using orientations derived from histological data [27].

On the other hand, phantoms provide an additional possibility for the quantitative evaluation because they offer more controlled, reproducible, and easily accessible experiments. Physical phantoms have been used in dMRI validation setups. For example, the reproducibility of MD measurements was assessed in [17], whereas the recovery of the Ensemble average propagator was validated in a crossing phantom in [23]. In the Fiber Cup [10,5] and ISBI 2018 [26] challenges, tractography reconstructions were compared to ground-truth fiber configurations from physical phantoms. Synthetic software-based phantoms also proved to be a valid alternative to physical phantoms for validation purposes, e.g., see [21,24] and references therein.

In fact, fetal subjects are a sensitive cohort, thus preventing from assessing different acquisition configurations. Hence, the evaluation of our technique on a quantitative dMRI phantom is crucial before applying it to *in vivo* data. However, designing a phantom that matches a fetal brain is extremely complex and challenging. In this work, we use a small size phantom with a customized fractional anisotropy (FA) in the single fiber population in the upper values reported in fetal brains. Indeed, in their atlas, Khan et al. [15] modelled the splenium of the corpus callosum (CC) of a fetus of 37 gestational weeks with an approximately close FA. Similar values were reported both for the genu and the splenium of the CC [8]. Therefore, our phantom is relevant to perform a benchmark analysis in fetuses in the 3rd trimester of gestation. Additionally, the dMRI signal obtained from physical phantoms is similar to *in vivo* data and is more realistic than the dMRI signal obtained from numerical simulations.

Fetal dMRI severely suffers from the unpredictable motion and artifacts caused by the small fetal brain structure that is surrounded by amniotic fluid and maternal organs. Scanning times are typically shorter than that of postna-

tal studies, limiting the possibility of long diffusion MRI acquisitions based on a large number of diffusion gradient directions and high b-values, which are required to disentangle complex fiber configurations. Furthermore, the use of fast Echo Planar Imaging acquisitions to freeze intra-slice motion leads to highly blurred and distorted images. These images also have a low Signal-to-Noise ratio (SNR), due to the tissue properties of the fetal white matter. Orthogonal scans of anisotropic resolution are usually acquired to overcome these pitfalls. In clinical practice, there is a strong constraint on scanning time, often below 10 minutes. This does not allow to acquire a high number of orthogonal volumes and a high number of diffusion directions at the same time. Additionally, clinical protocols are not consensual between sites. Typically, clinical fetal brain dMRI have an in-plane resolution of 1-2 mm, a slice thickness of 3-5 mm, the number of gradient directions ranges between 4 and 32, and unique b-values between 400 and  $1000s/mm^2$  are employed [12,19,20]. Conversely, in the pioneer research initiative of Developing Human Connectome Project protocol (DHCP) [4,6], up to 141 diffusion volumes can be acquired with multiple b-values (400 and  $1000s/mm^2$ ) and a scanning time of about 15 minutes per 4D volume.

Our study focuses on the quantitative evaluation of the accuracy of DTI and ODF reconstructions from *in vivo* fetal dMRI acquisitions to identify a good trade-off between the number of series and the number of diffusion gradient directions in a more clinically realistic scenario (summarized in Figure 1).

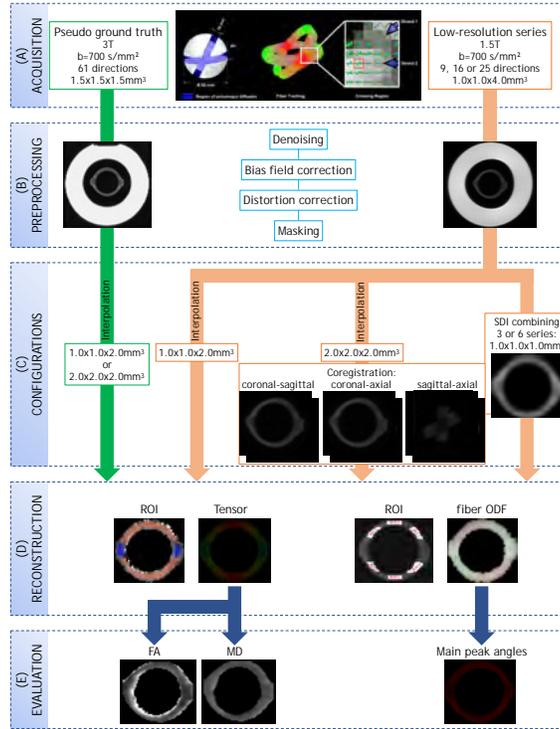
## 2 Methodology

### 2.1 Materials

**Fetal crossing phantom** - We used a customized fiber crossing phantom (diameter & height of  $150mm$ ) [22] made of two interleaved polyester fiber strands encapsulated in an aqueous solution. The fibers diameter is of  $15\mu m$ , the crossing angle between the two strands is approximately  $60^\circ$ , and a customized FA to mimic fetal values in the single fiber population of 0.6 was requested. These values were reported by the vendor who computed them from 128 diffusion-weighted images (DWI) and  $b = 1000s/mm^2$ . (Figure 1A).

**MRI acquisitions** - High-resolution (HR) (spatial and angular) images were acquired at 3T (MAGNETOM Prisma-Fit, Siemens Healthcare, Erlangen, Germany), with a 16-channel body array coil and a 32-channel spine coil using a pulsed gradient spin-echo (PGSE) sequence with four different b-values, 400, 700, 1000 and  $3000s/mm^2$ . The spatial resolution was  $1.5mm^3$  isotropic with a field of view of  $256x256x88mm^3$ , acquired with 61 directions. The echo time (TE) was 52 ms, the repetition time (TR) was 4200 ms and the flip angle was  $90^\circ$ . Only the  $b = 700s/mm^2$  acquisition was considered as the pseudo ground truth (pseudo-GT) in the validation framework.

Low-resolution (LR) acquisitions were performed at 1.5T (MAGNETOM Sola, Siemens Healthcare, Erlangen, Germany), with an 18-channel body array coil and a 32-channel spine coil, using a PGSE sequence (TE=82 ms, TR=2000 ms,



**Fig. 1.** Summary of our phantom evaluation framework for fetal dMRI acquisitions.

flip angle=90°). The acquisition time was approximately one minute per 4D volume. The in-plane resolution was 1x1 mm<sup>2</sup>, the slice thickness was 4mm and the field of view 207x207x69 mm<sup>3</sup>. We used  $b = 700$  s/mm<sup>2</sup> and either 9, 16 or 25 directions, uniformly distributed in the half-sphere. In order to correct non-linear distortions, we also acquired a 1mm isotropic T2-weighted (T2-w) image using a Sampling Perfection with Application optimized Contrasts using different flip angle Evolution (SPACE) sequence (TE=380 ms, TR=3200 ms). Both our data<sup>1</sup> and code<sup>2</sup> will be available to ensure reproducibility of the results.

## 2.2 Data processing

**Preprocessing** - Both the pseudo-GT and LR datasets were preprocessed as follows: a denoising step using a Principal Component Analysis based method [33], followed by an N4 bias-field inhomogeneity correction [32]. Distortion was corrected using a state-of-the-art algorithm for fetal brain [16]. We started by a rigid registration of the distortion free T2-w image to the b0 ( $b = 0$  s/mm<sup>2</sup>)

<sup>1</sup> [www.zenodo.org/record/5153507.YQgEA3UzbRY](http://www.zenodo.org/record/5153507.YQgEA3UzbRY)

<sup>2</sup> [www.github.com/Medical-Image-Analysis-Laboratory/FetalBrainDMRI\\_CrossingPhantom](https://www.github.com/Medical-Image-Analysis-Laboratory/FetalBrainDMRI_CrossingPhantom)

image followed by a non-linear registration in the phase-encoding direction of the b0 to the same T2-w image. The transformation was then applied to the diffusion-weighted images.

**Definition of regions-of-interest (ROI)** - Masks of the fiber endpoint regions (single fiber and crossing fiber ROIs) were obtained using mathematical morphology operations, intensity thresholding in the b0 image and manual refinement. Manual segmentation of each region was performed in the  $1 \times 1 \times 2 \text{ mm}^3$  resolution and propagated by nearest neighbor interpolation and manual refinement to other resolution volumes. Borders were not considered to avoid partial volume effect. The single fiber ROI was further subdivided in six ROIs: ROI 1 and ROI 2 in which the fibers are oriented horizontally and ROI 3-6 where they are oblique (Figure 1D).

**Interpolation** - Since the pseudo-GT and LR series have very different resolutions, they were both mapped to a middle ground resolution of  $1 \times 1 \times 2 \text{ mm}^3$  and  $2 \times 2 \times 2 \text{ mm}^3$  using trilinear interpolation. We chose these trade-off resolutions to avoid to significantly degrade the pseudo-GT by introducing artifacts and to enhance the LR volumes as it was demonstrated in [7]. Additionally, up-sampling LR DWI images by a factor of two is a common practice in clinical fetal dMRI [13]. The  $1 \times 1 \times 2 \text{ mm}^3$  resolution was used for unique volumes, i.e., either axial, coronal or sagittal and the  $2 \times 2 \times 2 \text{ mm}^3$  resolution for combined ones, i.e., axial-coronal, axial-sagittal or coronal-sagittal. For the combined volumes, we registered the b0 images of the coronal and the sagittal acquisitions to the b0 image of the axial one using landmarks [9]. This transformation was then applied to the DWI images. To reduce error propagation related to interpolation, we have performed the latter after the preprocessing. We have also computed the different metrics at the different resolutions to quantify variations linked to interpolating the data.

**Scattered data interpolation** - We generated a HR volume from a set of either three or six LR orthogonal series using Scattered Data Interpolation (SDI) reconstruction [25] as implemented in MIALSRTK (version 2.0.1) [28]. It was applied separately to each DWI image and each b0. This consisted in co-registering to an axial reference volume, resampled to isotropic high-resolution, all the series as a first step. Then, the intensity of each voxel in the HR volume grid was computed by averaging the intensities of the corresponding neighboring voxels in the LR volumes using a Gaussian kernel. To match the underlying point spread function of the data, the Gaussian kernel profile was set to be perpendicular to the slice plane with a zero mean and a Full Width at Half Maximum (FWHM of  $\sim 2.355$  standard deviation) equal to the voxel resolution.

**Reconstruction** - We reconstructed (1) the diffusion tensor from which we derived both the FA and mean diffusivity (MD) maps and (2) the fiber ODF using the constrained spherical deconvolution (CSD) method [30] from which the main peak (i.e., fiber orientation) was determined. The fiber ODF is represented in the Spherical Harmonics (SH) basis, where an order 4 (15 parameters) was used to best fit all directions (15, 25 for the LR volumes and 61 for the pseudo-GT) and be able to make a one-to-one comparison. In the CSD algorithm, we have

constrained the maximum number of peaks to two and the minimum separation angle to  $25^\circ$ . Dipy (version 1.3.0) [11] was used for reconstruction and visualization, and MRtrix3 [29] for fiber ODF visualization.

**Evaluation metrics** - To be able to fairly compare diffusion metrics, unbiased by different b-values, we only used as reference the HR data acquired with the same b-value i.e.  $b = 700s/mm^2$  (i.e., pseudo-GT) as the LR data. Scalar maps were evaluated by computing the relative difference between images, i.e., difference between the average LR and the average pseudo-GT map, divided by the average pseudo-GT map. The coefficient of variation (CV, i.e., standard deviation/mean) was also used to quantify the variability of scalar maps.

### 3 Results

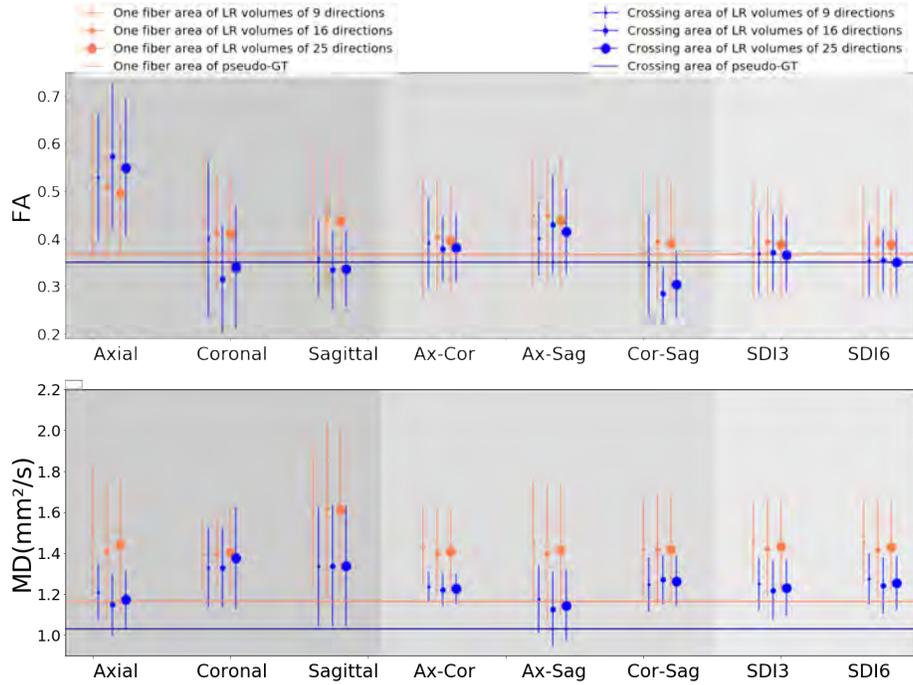
#### 3.1 Scalar maps

**Evaluation of pseudo-GT** - We first assessed the pseudo-GT compared to the diffusion properties given by the vendor. The estimated FA from the pseudo-GT was found to be equal to 0.367 (horizontal orange line in Figure 2) which did not correspond to the FA reported by the vendor (0.6) in the single fiber population. This is not surprising since FA strongly depends on the acquisition parameters, and in particular on the b-value. Indeed, the same observation was made in the Fiber Cup study [10], where an increase of 75% in the mean FA was reported between  $b = 650s/mm^2$  and  $b = 2000s/mm^2$ . The computed FA of 0.367 falls in the same range of FA reported in [14] (using  $b = 700s/mm^2$  & 32 directions) for various fetal brain structures. Conversely, the mean MD =  $1.165mm^2/s$  was more consistent with the value reported by the vendor (i.e.,  $\sim 1.2mm^2/s$ ).

Let us note that scalar maps did not show major differences across different pseudo-GT interpolations, with a CV of 0.5% for single fiber and up to 6.5% for crossing populations. This is lower than the CV of the FA (up to 22%) and MD (up to 12%) values within single and crossing fibers areas of each scalar maps.

**Assessment of enhanced acquisitions** - Figure 2 shows the results from the LR scalar maps for the different configurations compared to the pseudo-GT (two horizontal lines). The orange color refers to the single fiber population and the blue color to the crossing fiber populations, and the bigger the disk diameter the more diffusion directions are used in the reconstruction. For FA, SDI methods outperform the other configurations, especially when considering the single fiber population that shows a difference from the pseudo-GT of 6.1%. Single LR volumes and combinations of pairs are more sensitive to the number of diffusion directions (in these cases, the more directions, the smaller the error), whereas SDI does not show this influence.

The axial acquisition exhibits a singular behaviour compared to the two other single-volume acquisitions, depicting a higher FA in the crossing area compared to the single fiber area. By inspecting the scanner FA map, we found out an already high FA, particularly in the crossing area of 17% more than in the



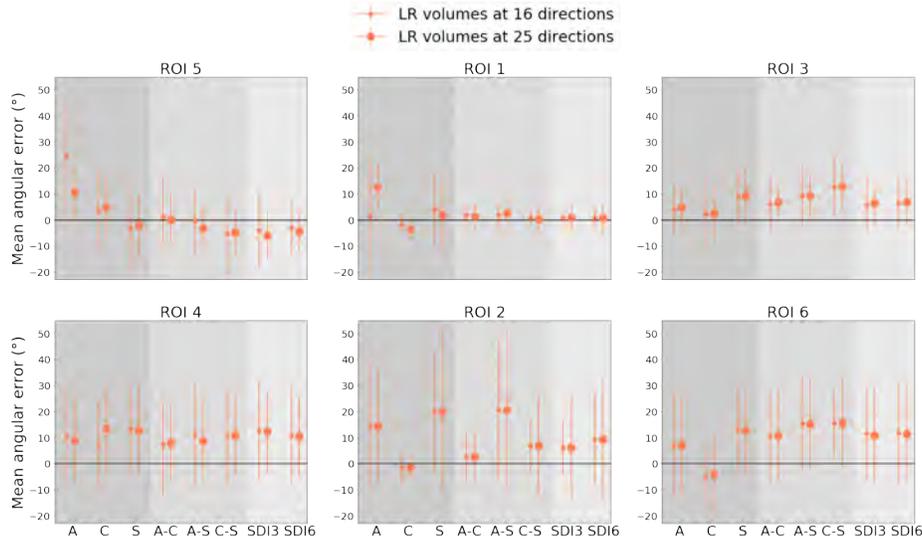
**Fig. 2.** Scalar maps estimation error compared to the pseudo-GT (horizontal lines, single fiber in orange, crossings in blue, see Figure 1D). Axial, Coronal, and Sagittal data correspond to single volumes with a resolution of  $1 \times 1 \times 2 \text{ mm}^3$ . Ax-Cor, Ax-Sag, and Cor-Sag denote combined volumes with a resolution of  $2 \text{ mm}$  isotropic. SDI3 and SDI6 are the interpolated scattered data by using three or six  $1 \text{ mm}$  isotropic volumes, respectively.

coronal and the sagittal maps. So merging orthogonal volumes can reduce any potential discrepancy between the different acquisitions (due to outliers and artifacts in the data or due to the anisotropy of the acquisitions capturing the non-symmetrical anatomy across planes of the fibers) and SDI provides the most robust solution.

Differently than FA, MD errors are not influenced by the number of directions neither for single, pairs nor SDI volumes. Both merging pairs of orthogonal volumes and SDI reconstructions help attenuate the high error rate of the sagittal volume by a difference from the average pseudo-GT of about 15% (single fiber population) and 20% (crossing area). The difference between the LR and pseudo-GT values can be explained by the magnetic field strength. Indeed, it was shown in [18], that MD was significantly different between 1.5T and 3T acquisitions.

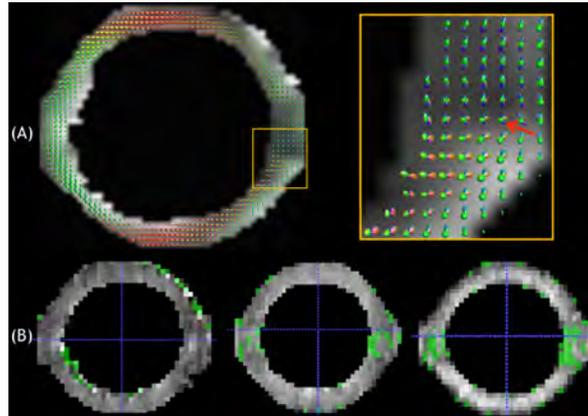
### 3.2 Fiber orientation errors

Fiber orientations estimation in the pseudo-GT across interpolations is stable in the different ROIs. The maximum standard deviation in ROIs 1-2, where the fibers are close to the x-axis coordinates, is  $1.6^\circ$ , whereas it reaches  $4.5^\circ$  in ROIs 3-6 where the fibers are rotated by around  $50^\circ$ . As depicted in Figure 3, the angular error of the LR estimated orientations doesn't correlate to the different orthogonal volumes configurations, except for SDI that always shows a lower angular error than, at least, the most under-performing single volume reconstruction. Furthermore, we can observe that the standard deviation of the angular error (vertical lines in Figure 3) strongly depends on the region of interest. For instance, ROI 1 angles are less variant and closer to the pseudo-GT whereas in ROI 2, the sagittal acquisition compromises the estimated angle of other reconstructions where it belongs. In contrast to ROI 4 and ROI 6, the errors in ROI 3 and ROI 5 are not dramatic as they are located below the mean separation angle of  $25^\circ$ . Importantly, the error difference between the LR and the pseudo-GT volumes is independent of the number of diffusion directions used to compute the main ODF peak.



**Fig. 3.** Mean angular error in different single fiber ROIs corresponding to Figure 1D for different configurations. Graphs for each ROI are positioned in the corresponding order of their locations on the phantom. A: Axial, C: Coronal, S: Sagittal.

Figure 4A shows fiber ODFs overlaid on the FA map of a LR volume. As can be noted, only very few crossing fibers can be detected at  $b = 700s/mm^2$ .



**Fig. 4.** (A) Fiber ODFs of a LR coronal image overlaid on the FA map. Red arrow: detected crossing. (B) Voxels detected as two peaks in the high resolution acquisition using b-values of 700, 1000 and  $3000s/mm^2$  (left to right, respectively).

Results shown in Figure 4B demonstrated that in the HR data, fiber crossings (i.e., two peaks) can only be significantly resolved at  $b = 3000s/mm^2$ . In the crossing region, a median inter-fiber angle of  $62^\circ$  close to that reported by the vendor (i.e.,  $60^\circ$ ) was detected by using a SH order of 8, although with a high standard deviation of  $29^\circ$ . For this reason, we did not perform fiber orientation analyses in the fiber crossing area of the LR data.

## 4 Conclusion and Discussion

We have demonstrated how reported diffusion properties of a fetal customized crossing phantom vary across orthogonal series and the number of diffusion directions, and how scattered data interpolation of multiple volumes can reduce this variability and so better approximate the pseudo ground truth. Increasing the number of directions did not consistently reduce error metrics (MD, FA, and fiber orientations) because of the low b-value and the relatively low number of directions employed, which only allow estimating a single fiber per voxel. The main limitation of this study is the absence of unpredictable motion which is one of the main challenges in fetal MRI. However, random motion could be a confounding factor to evaluate different acquisition schemes. Hence setting up a first ideal motion-free scenario to quantify the maximum expected variability of fetal dMRI measurements is a key starting point. Hence, these conclusions have to be taken as an upper bound that can be achieved. In future studies, we plan to extend this work by considering other acquisition protocols (such as the DHCP protocol), by using motion-induced acquisitions for testing different super-resolution reconstruction methods [6], and by implementing scan-rescan analyses in different scanners.

## 5 Acknowledgments

This work was supported by the Swiss National Science Foundation (project 205321-182602, grant No 185897: NCCR-SYNAPSY- "The synaptic bases of mental diseases" and the Ambizione grant PZ00P2\_185814), the Centre d'Imagerie BioMédicale (CIBM) of the University of Lausanne (UNIL), the Swiss Federal Institute of Technology Lausanne (EPFL), the University of Geneva (UniGe), the Centre Hospitalier Universitaire Vaudois (CHUV), the Hôpitaux Universitaires de Genève (HUG), and the Leenaards and Jeantet Foundations. This work was also supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754462.

## References

1. Aydogan, D.B., Jacobs, R., Dulawa, S., Thompson, S.L., Francois, M.C., Toga, A.W., Dong, H., Knowles, J.A., Shi, Y.: When tractography meets tracer injections: a systematic study of trends and variation sources of diffusion-based connectivity. *Brain Structure and Function* **223**(6), 2841–2858 (2018)
2. Canales-Rodríguez, E.J., Iturria-Medina, Y., Alemán-Gómez, Y., Melie-García, L.: Deconvolution in diffusion spectrum imaging. *Neuroimage* **50**(1), 136–149 (2010)
3. Canales-Rodríguez, E.J., Legarreta, J.H., Pizzolato, M., Rensonnet, G., Girard, G., Patino, J.R., et al.: Sparse wars: A survey and comparative study of spherical deconvolution algorithms for diffusion MRI. *NeuroImage* **184**, 140–160 (2019)
4. Christiaens, D., Cordero-Grande, L., Price, A., Hutter, J., Hughes, E., Counsell, S., Tournier, J., Hajnal, J.: Fetal diffusion mri acquisition and analysis in the developing human connectome project. In: *Proceedings of the annual meeting of the International Society of Magnetic Resonance in Medicine (ISMRM)* (2019)
5. Côté, M.A., Girard, G., Boré, A., Garyfallidis, E., Houde, J.C., Descoteaux, M.: Tractometer: Towards validation of tractography pipelines. *Medical Image Analysis* **17**(7) (2013)
6. Deprez, M., Price, A., Christiaens, D., Estrin, G.L., Cordero-Grande, L., et al.: Higher order spherical harmonics reconstruction of fetal diffusion mri with intensity correction. *IEEE Transactions on Medical Imaging* **39**(4), 1104–1113 (2019)
7. Dyrby, T.B., Lundell, H., Burke, M.W., Reislev, N.L., Paulson, O.B., Ptito, M., Siebner, H.R.: Interpolation of diffusion weighted imaging datasets. *NeuroImage* **103**, 202–213 (2014)
8. Estrin, G.L., Wu, Z., Deprez, M., Bertelsen, Á., Rutherford, M.A., Counsell, S.J., Hajnal, J.V.: White and grey matter development in utero assessed using motion-corrected diffusion tensor imaging and its comparison to ex utero measures. *Magnetic Resonance Materials in Physics, Biology and Medicine* **32**(4), 473–485 (2019)
9. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., et al.: 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging* **30**(9), 1323–1341 (2012)
10. Fillard, P., Descoteaux, M., Goh, A., Gouttard, S., Jeurissen, B., et al.: Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom. *NeuroImage* **56**(1), 220–234 (2011)
11. Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M., Nimmo-Smith, I.: Dipy, a library for the analysis of diffusion mri data. *Frontiers in neuroinformatics* **8** (2014)

12. Jakab, A., Kasprian, G., Schwartz, E., Gruber, G.M., Mitter, C., Prayer, D., Schöpf, V., Langs, G.: Disrupted developmental organization of the structural connectome in fetuses with corpus callosum agenesis. *Neuroimage* **111** (2015)
13. Jakab, A., Tuura, R., Kellenberger, C., Scheer, I.: In utero diffusion tensor imaging of the fetal brain: a reproducibility study. *NeuroImage: Clinical* **15** (2017)
14. Kasprian, G., Brugger, P.C., Weber, M., Krssák, M., Krampl, E., Herold, C., Prayer, D.: In utero tractography of fetal white matter development. *Neuroimage* **43**(2), 213–224 (2008)
15. Khan, S., Vasung, L., Marami, B., Rollins, C.K., Afacan, O., Ortinau, C.M., Yang, E., Warfield, S.K., Gholipour, A.: Fetal brain growth portrayed by a spatiotemporal diffusion tensor mri atlas computed from in utero images. *NeuroImage* **185**, 593–608 (2019)
16. Kuklisova-Murgasova, M., Estrin, G.L., Nunes, R.G., Malik, S.J., Rutherford, M.A., Rueckert, D., Hajnal, J.V.: Distortion correction in fetal epi using non-rigid registration with a laplacian constraint. *IEEE Transactions on Medical Imaging* **37**(1) (2017)
17. Lavdas, I., Behan, K.C., Papadaki, A., McRobbie, D.W., Aboagye, E.O.: A phantom for diffusion-weighted MRI (DW-MRI). *Journal of Magnetic Resonance Imaging* **38**(1) (2013)
18. Lavdas, I., Miquel, M.E., McRobbie, D.W., Aboagye, E.O.: Comparison between diffusion-weighted mri (dw-mri) at 1.5 and 3 tesla: A phantom study. *Journal of Magnetic Resonance Imaging* **40**(3), 682–690 (2014)
19. Machado-Rivas, F., Afacan, O., Khan, S., Marami, B., Rollins, C., Ortinau, C., Velasco-Annis, C., Warfield, S., Gholipour, A., Jaimes, C.: Tractography of the cerebellar peduncles in second-and third-trimester fetuses. *American Journal of Neuroradiology* **42**(1), 194–200 (2021)
20. Machado-Rivas, F., Afacan, O., Khan, S., Marami, B., Velasco-Annis, C., Lidov, H., Warfield, S., Gholipour, A., Jaimes, C.: Spatiotemporal changes in diffusivity and anisotropy in fetal brain tractography (2021)
21. Maier-Hein, K.H., Neher, P.F., Houde, J.C., Côté, M.A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.C., Lin, Y.C., Ji, Q., et al.: The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications* **8**(1) (2017)
22. Moussavi-Biugui, A., Stieltjes, B., et al.: Novel spherical phantoms for q-ball imaging under in vivo conditions. *Magnetic resonance in medicine* **65**(1), 190–194 (2011)
23. Ning, L., Laun, F., Gur, Y., DiBella, E.V., Deslauriers-Gauthier, S., Megherbi, T., et al.: Sparse Reconstruction Challenge for diffusion MRI: Validation on a physical phantom to determine which acquisition scheme and analysis method to use? *Medical Image Analysis* **26**(1), 316–331 (2015)
24. Rafael-Patino, J., Romascano, D., Ramirez-Manzanares, A., Canales-Rodríguez, E.J., Girard, G., Thiran, J.P.: Robust Monte-Carlo Simulations in Diffusion-MRI: Effect of the substrate complexity and parameter choice on the reproducibility of results (2019)
25. Rousseau, F., Glenn, O.A., Iordanova, B., Rodriguez-Carranza, C., Vigneron, D.B., Barkovich, J.A., Studholme, C.: Registration-based approach for reconstruction of high-resolution in utero fetal mr brain images. *Academic radiology* **13**(9) (2006)
26. Schilling, K.G., Gao, Y., Stepniewska, I., Janve, V., Landman, B.A., Anderson, A.W.: Anatomical accuracy of standard-practice tractography algorithms in the motor system - A histological validation in the squirrel monkey brain. *Magnetic Resonance Imaging* **55**, 7–25 (2019)

27. Schilling, K.G., Janve, V., Gao, Y., Stepniewska, I., Landman, B.A., Anderson, A.W.: Histological validation of diffusion MRI fiber orientation distributions and dispersion. *NeuroImage* **165**, 200–221 (2018)
28. Tourbier, S., Bresson, X., Hagmann, P., Thiran, J.P., Meuli, R., Cuadra, M.B.: An efficient total variation algorithm for super-resolution in fetal brain mri with adaptive regularization. *NeuroImage* **118**, 584–597 (2015)
29. Tournier, J.D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., et al.: MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage* **202**, 116137 (2019)
30. Tournier, J.D., Yeh, C.H., Calamante, F., Cho, K.H., Connelly, A., Lin, C.P.: Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. *Neuroimage* **42**(2), 617–625 (2008)
31. Tuch, D.S.: Q-ball imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **52**(6) (2004)
32. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., et al.: N4itk: improved n3 bias correction. *IEEE transactions on medical imaging* **29**(6), 1310–1320 (2010)
33. Veraart, J., Fieremans, E., Novikov, D.S.: Diffusion mri noise mapping using random matrix theory. *Magnetic resonance in medicine* **76**(5), 1582–1593 (2016)
34. Wedeen, V.J., Wang, R., Schmahmann, J.D., Benner, T., Tseng, W.Y.I., Dai, G., Pandya, D., Hagmann, P., D’Arceuil, H., de Crespigny, A.J.: Diffusion spectrum magnetic resonance imaging (dsi) tractography of crossing fibers. *Neuroimage* **41**(4), 1267–1277 (2008)

# 5

## Appendix

This section incorporates results related to four studies: (1) Angular super-resolution of the diffusion signal with convolutional LSTMs and Transformers; (2) Automated fetal brain segmentation of 2D magnetic resonance images: transfer learning and 3D topology correction; (3) Direct segmentation of brain white matter tracts in diffusion MRI<sup>112</sup> and (4) Spatio-temporal motion correction and iterative reconstruction of in-utero fetal fMRI<sup>206</sup>. The first study is related to our works of resolution enhancement, exploring whether the scheme in Lyon et al.<sup>137</sup> could also infer new diffu-

sion directions in developing brains. However, our study did not reach the expected improvements. The second project aimed at segmenting fetal brains from T2-weighted images and was accepted in the European Congress of Magnetic Resonance in Neuropediatrics (ECMRN20). The third study employs a similar methodological approach to our FOD estimation work<sup>114</sup> but applied for white matter tract segmentation to a population of 2-8 years and adults<sup>112</sup>, and not developing brains which is the focus of this thesis. The last project tackles the problem of motion correction and super-resolution reconstruction in fetal fMRI<sup>206</sup> and was accepted in MICCAI22.

## 5.1 ANGULAR SUPER-RESOLUTION WITH DEEP NEURAL NETWORKS

In this section, we will briefly detail methods and results about the angular super-resolution of the diffusion signal.

### 5.1.1 METHODOLOGY

This study aimed at applying the method developed by Lyon et al.<sup>137</sup> in adult brains (using the Human Connectome Project<sup>223</sup>) to developing brains of newborns and fetuses, and potentially enhance it by the use of Transformers<sup>53</sup>. We used two datasets: the second release of the newborn developing Human Connectome Project (dHCP) dataset<sup>93</sup> and a fetal clinical dataset.

In a nutshell, the RCNN (Recurrent Convolutional Neural Network) model takes as input  $q_{in}$  diffusion 3D patches along with their corresponding b-vector, repeated and concatenated in the channel dimension. A cascade of convolutions and distributed convolutions is then applied until combining the feature maps of each  $q_{in}$  feature using a convolutional LSTM block. In the decoder part, the hidden internal state of the LSTM is concatenated with the  $q_{out}$  target b-vectors, for which we aim to predict their corresponding diffusion volume. Convolutions blocks are then applied with residual connections from the target b-vectors until predicting the final  $q_{out}$  patches. Further details

can be found in the original paper<sup>137</sup>.

In our work, we have used the same pre-term subjects of the dHCP dataset that we employed in our through-plane super-resolution study<sup>110</sup>, i.e. 37 subjects of 88 directions from  $b=1000\text{ s/mm}^2$ . We have additionally resampled the data to isotropic resolution, i.e. from  $1.17 \times 1.17 \times 1.5\text{ mm}^3$  to  $1.25\text{ mm}^3$  to account for the 3D network. The fetal data is also the same data used in our previous work<sup>110</sup>, i.e. 6 subjects of 15 directions of  $b=700\text{ s/mm}^2$ , with a spatial resolution of  $1 \times 1 \times 3 - 5\text{ mm}^3$ .

Since our goal was ultimately to generate new gradient volumes in the fetal data, we have trained the RCNN network on the pre-term subjects that has 88 directions, first using the original 3D architecture, then by converting it to 2D because of the thick fetal slices, i.e. high anisotropy of the fetal data, and finally by incorporating Transformers at different stages of the network, mostly after the convolutional blocks as this is a good practice in training Transformers<sup>53,57</sup> in the case of small datasets such as in our medical imaging field. Similarly to Lyon et al.<sup>137</sup>, we have also used spherical harmonics (order 2) interpolation as a baseline method for comparison<sup>68</sup>.

We have used cross validation after splitting the pre-term training data into six folds of 4 subjects each. We have used  $q_{in}=10$  for training, as in the original model<sup>137</sup> and  $q_{out}=5$  (i.e.  $q_{in}+q_{out}$ = number of directions in fetal data). We have varied the patch size ( $PS \in \{10^2, 32^2, 64^2\}$  for 2D networks and  $PS \in \{10^3, 16^3\}$  for 3D networks) as it is an important parameter of the model. We have additionally trained a network on 10 carefully selected fetal subjects.

### 5.1.2 RESULTS

The patch size effect was clear and in favor of higher patches as can be seen in the case of 2D networks in Figure 5.1. The order of magnitude of errors of 3D networks with a patch size of  $16^3$  was similar to 2D networks of patch size  $32^2$ . Figure 5.2 however does not show a clear advantage for the RCNN method (2D or 3D) and the spherical harmonics interpolation baseline.



Figure 5.1: Validation loss of the trained models with different patch sizes, i.e.  $10^2$ ,  $32^2$  and  $64^2$  voxels. Image generated with Tensorboard<sup>4</sup>.

Regarding fetal data, the network generated plausible predictions as can be seen in Figure 5.3, on one of  $q_{out}=5$  gradient directions, for both network prediction and ground truth (GT). We can however notice smooth images generated by the RCNN. The left panel of the same Figure 5.3 shows for 100 random points in the corresponding slice on the right panel, how the prediction and the GT evolve. For the two top slices, the error is higher compared to the two bottom slices. In fact, it is not surprising that the error is higher around the extremities of the brain compared to slices in the middle of the brain. Quantitative results for the network trained on fetal data (Figure 5.4) show that a simple spherical harmonics interpolation generates better results than the RCNN.

	Fold 1	Fold 2	Fold 3	Fold 4
SH-2 Baseline	0.000997	0.001225	0.000915	0.001174
3D RCNN	0.000898	0.000748	0.001430	0.001553
10x10x10 3D RCNN	0.000789	0.001290	0.001421	0.001598
16x16x16 2D RCNN	0.001111	0.000999	0.000639	0.001331
32x32 2D RCNN	0.000780	0.001320	0.001128	0.000843
64x64				

**Figure 5.2:** Root mean squared error (RMSE) of four out of the six folds, for the 2D- and 3D-RCNN trained on different patch sizes. Spherical harmonics interpolation baseline is also shown on the first row. Best and worst two scores are shown in blue and red, respectively

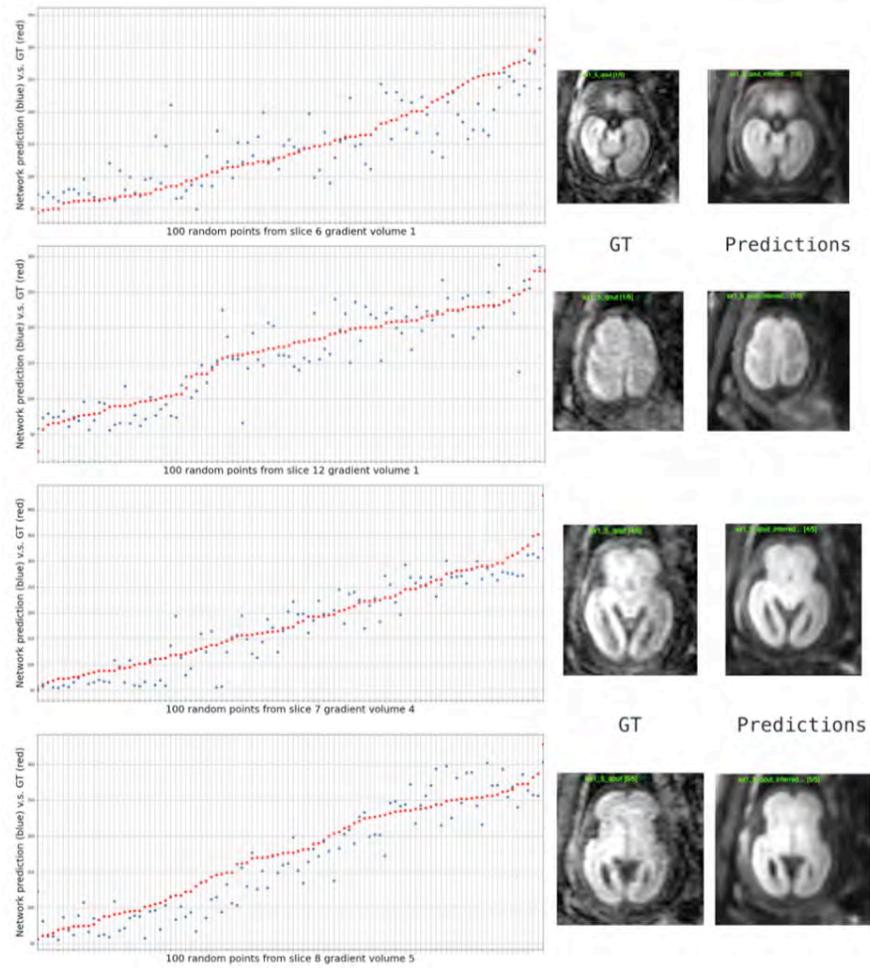


Figure 5.3: On the left, 100 random points extracted from brain voxels where red refers to ground truth values and blue to network prediction values by the 2D-RCNN trained with 32x32 patches. On the right we can see the corresponding slices of the GT and network predictions.

	Validation subject AX-1	Validation subject AX-2	Training subject AX-1	Training subject AX-2	Training subject AX-3
SH order 2	0.012434	0.013803	0.009477	0.012194	0.007469
RCNN	0.022064	0.022446	0.020715	0.020263	0.010924

Figure 5.4: Root mean squared error (RMSE) of validation and training subjects of spherical harmonics interpolation and RCNN trained on fetal data.

Following that, we aimed at incorporating Transformers in several parts of the network, by for instance replacing the LSTM part by a transformer or combining the LSTM with a transformer, for which we varied parameters such as the ( $D$ ) or the  $D_b$ , that sets the dimensionality of the query/key/value matrix<sup>53</sup>. Although achieving some incremental improvements by fine-tuning these parameters (Figure 5.5) and adding data augmentation to artificially increase our dataset, the results did not reach our expectations and we have hence aimed to directly predict FODs<sup>114</sup> instead of learning new non-acquired gradient directions. We hypothesize that the developing brain datasets (newborns and especially fetuses) do not show enough learnable patterns, at the level of the original raw signal, due to noise and high variability of the anatomy, and hence it is hard to train a neural network on this type of data.

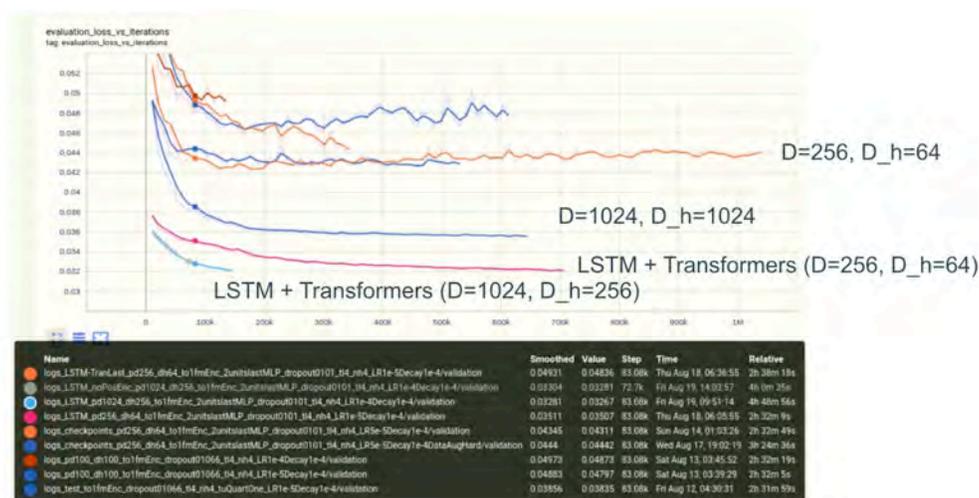


Figure 5.5: Validation loss on multiple models using Transformers with different parameters and by combining it with the LSTM module. Image generated with Tensorboard<sup>1</sup>.

## 5.2 **AUTOMATED FETAL BRAIN SEGMENTATION OF 2D MAGNETIC RESONANCE IMAGES: TRANSFER LEARNING AND 3D TOPOLOGY CORRECTION**

### 5.2.1 **BACKGROUND**

Fetal brain extraction in 2D thick series acquired by magnetic resonance imaging (MRI) is usually the first step for further processing such as super-resolution reconstruction, brain tissue segmentation, or in utero resting state fMRI analysis. Manual annotation of brain voxels is time consuming and hence inappropriate to automated analysis and large-scale studies. With the advent of deep supervised learning, fetal brain segmentation can be performed in real-time in 2D clinical scans with reasonably high accuracy. However, large labelled datasets are required in the training phase. Despite some solutions are publicly available, the generalization of such techniques to process unseen data (such as acquired with different scanners or sequence parameters) is not straightforward. In this work we present a transfer of learning, that is, how to adapt pre-existing deep learning solutions to fetal brain segmentation of a new clinical dataset acquired in a different hospital. Moreover, we added a post-processing 3D topology correction within the 2D brain mask series.

### 5.2.2 **MATERIALS AND METHODS**

Our dataset was acquired at the Lausanne University Hospital (CHUV). It consists of 227 series from 39 fetal brains, 20 pathological and 19 healthy subjects, ranging between 20 and 36 weeks of gestation. Repeated orthogonal T2-w Half-Fourier Acquisition Single-shot Turbo spin Echo (HASTE) sequences (slice thickness 3 mm) were performed on each subject at 1.5 Tesla.

We fine-tuned the well-established U-Net convolutional neural network for 2D brain segmentation from clinical images. We explored two different scenarios: 1) training this model from scratch on our dataset using an average of 4,767 slices (Random\_init\_U-Net) or 2) using pre-trained weights

from previous published work where an average of 13,000 slices were used for training, corresponding to 385 series of 41 fetal brains acquired at 3T from Boston Children's hospital. We refer to this approach as Pre-trained\_U-Net.

The network was trained on 2D slices because of fetal motion and hence did not use any 3D brain continuity information. The training lasted for about 200 epochs and used ADAM optimiser and a weighted cross-entropy loss function. Evaluation was performed through the dice score in a leave-four-out cross validation setup.

A second contribution of this work uses information from neighbouring slices within a series by using morphological operations such as closing, opening and connected components in order to refine the binary brain mask output by the 2D U-net. Wilcoxon rank sum test was performed with R software to compare Random\_init\_U-Net and Pre-trained\_U-Net configurations in both healthy and pathological datasets.

### 5.2.3 RESULTS

Directly testing the pre-trained weights does not generate plausible segmentations as the training data was very different from our acquisitions (3T v.s. 1.5T and different acquisition parameters). Accuracy obtained by training our network using the CHUV data only obtains an average dice score of 71%. We hypothesise as overall that the number of datapoints used is rather limited. In contrast, the pre-trained U-Net achieved improved results with an average dice of 0.85. It significantly outperforms ( $p < 0.05$ ) U-Net\_random\_init as can be shown in the quantitative figure of the poster below, in both healthy and pathological datasets. Some remaining errors were corrected by 3D brain continuity strategies, e.g. using the fact that the surface area of the brain cannot abruptly drop and increase within a one slice neighbourhood as demonstrated in the qualitative figure of the poster below.

#### 5.2.4 CONCLUSION

Directly testing the pre-trained weights does not generate plausible segmentations as the training data was very different from our acquisitions (3T v.s. 1.5T and different acquisition parameters). Accuracy obtained by training our network using the CHUV data only obtains an average dice score of 71%. We hypothesise as overall that the number of datapoints used is rather limited. In contrast, the pre-trained U-Net achieved improved results with an average dice of 0.85. It significantly outperforms ( $p < 0.05$ ) U-Net\_random\_init as can be shown in the quantitative figure of the poster below, in both healthy and pathological datasets. Some remaining errors were corrected by 3D brain continuity strategies, e.g. using the fact that the surface area of the brain cannot abruptly drop and increase within a one slice neighbourhood as demonstrated in the qualitative figure of the poster below.

5.3 **DIRECT SEGMENTATION OF BRAIN WHITE MATTER TRACTS IN DIFFUSION MRI**

5.4 **SPATIO-TEMPORAL MOTION CORRECTION AND ITERATIVE RECONSTRUCTION OF  
IN-UTERO FETAL fMRI**

# Direct segmentation of brain white matter tracts in diffusion MRI

Hamza Kebiri<sup>1,2</sup>, Ali Gholipour<sup>1</sup>, Meritxell Bach Cuadra<sup>2</sup>,  
Davood Karimi<sup>1</sup>

<sup>1</sup> Computational Radiology Laboratory (CRL), Department of Radiology, Boston Children’s Hospital, and Harvard Medical School, USA

<sup>2</sup> Department of Radiology, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland

**Abstract.** The brain white matter consists of a set of tracts that connect distinct regions of the brain. Segmentation of these tracts is often needed for clinical and research studies. Diffusion-weighted MRI offers unique contrast to delineate these tracts. However, existing segmentation methods rely on intermediate computations such as tractography or estimation of fiber orientation density. These intermediate computations, in turn, entail complex computations that can result in unnecessary errors. Moreover, these intermediate computations often require dense multi-shell measurements that are unavailable in many clinical and research applications. As a result, current methods suffer from low accuracy and poor generalizability. Here, we propose a new deep learning method that segments these tracts directly from the diffusion MRI data, thereby sidestepping the intermediate computation errors. Our experiments show that this method can achieve segmentation accuracy that is on par with the state of the art methods (mean Dice Similarity Coefficient of 0.826). Compared with the state of the art, our method offers far superior generalizability to undersampled data that are typical of clinical studies and to data obtained with different acquisition protocols. Moreover, we propose a new method for detecting inaccurate segmentations and show that it is more accurate than standard methods that are based on estimation uncertainty quantification. The new methods can serve many critically important clinical and scientific applications that require accurate and reliable non-invasive segmentation of white matter tracts.

**Keywords:** white matter tracts · segmentation · neuroimaging

## 1 Introduction

The brain white matter is organized into a set of distinct tracts. These tracts are bundles of myelinated axons that connect different brain regions such as the cerebral cortex and the deep gray matter. Although they are tightly packed and often cross one another, each tract has an entirely different function and connects different regions of the brain [32, 36]. Accurate segmentation of these tracts is needed in clinical studies and medical research. For example, in surgical

planning one needs to know the precise extent of the individual tracts in order to assess the risk of damage to specific neurocognitive functions that may result from surgical removal of brain tissue. As another prominent example, changes in the micro-structural properties of different tracts is commonly used in studying brain development and disorders.

Magnetic resonance imaging (MRI) is the modality of choice for non-invasive assessment of white matter tracts in vivo. Although some of the tracts may be identifiable on T1, T2, or FLAIR images [36], accurate segmentation of most tracts is only possible with diffusion MRI. Individual tracts may be extracted from whole-brain tractograms by specifying inclusion and exclusion regions of interest (ROIs). This process, which is usually referred to as “virtual dissection”, is time-consuming, subjective, and it has low reproducibility [24]. Some prior works have aimed at automating the virtual dissection process by learning to compute the inclusion/exclusion ROIs [28, 37]. It is also possible to extract the tracts from a whole-brain tractogram by grouping similar streamlines using a clustering approach. This can be done by comparing individual streamlines with a predefined set of tracts in an atlas [7, 13]. Some techniques additionally take into account the location of the streamlines relative to anatomical landmarks in the brain [25, 26]. Tractography-based methods are inherently limited by the errors in streamline tractography [14]. To avoid these errors, some methods segment the tracts on diffusion tensor or fiber orientation images, thereby avoiding the tractography. Some of the segmentation techniques that have been explored in the past include Markov Random Fields [3], k-nearest neighbors technique [19], template matching [6], and more recently deep learning [5, 34]. However, none of these intermediate parameters (e.g., the diffusion tensor) have an unambiguous biophysical meaning and their computation entails unavoidable estimation errors. Moreover, the intermediate computations for most existing methods assume availability of dense multi-shell diffusion MRI measurements, which are not acquired in many clinical and research applications. As a result, existing methods have low accuracy and limited generalizability when applied to typical clinical scans.

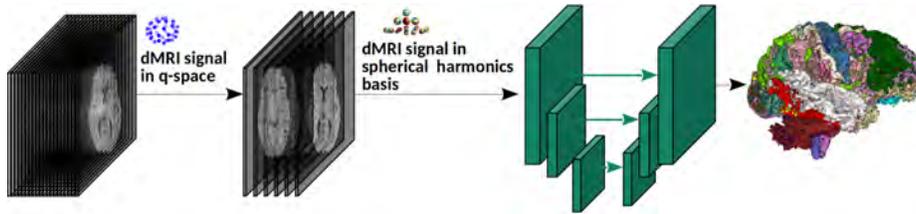
In this work, we develop and validate a new method that segments white matter tracts directly from the diffusion MRI data. The new method does not require tractography or computation of other intermediate parameters. Moreover, we present a simple but effective technique for detecting less accurate segmentations. We show that the new methods achieve superior accuracy and generalizability compared with the existing methods.

## 2 Materials and methods

### 2.1 Segmentation model

Our method, shown schematically in Figure 1, is based on a fully convolutional network (FCN). The network architecture is similar to nnU-Net (we refer to [9] for the details of the architecture). Our method predicts tract segmentations directly from the diffusion MRI data. To enhance the generalizability of the

method and to enable it to work with scans acquired using different gradient tables (i.e., different gradient strengths and/or different gradient directions): (i) We train the model with measurements that are typically acquired for diffusion tensor imaging (DTI). DTI-style scans include single-shell measurements at a b-value of around 750-1200 s/mm<sup>2</sup> [10]. They are the most common acquisition in clinical and research applications. We normalize these measurements by a non-weighted (b=0) measurement. (ii) We project the normalized data onto a fixed spherical harmonics (SH) basis. We use SH bases formulations of [30] with an order 2, which results in 6 SH coefficients regardless of the number of measurements. We use these 6 coefficient maps as input to the FCN.



**Fig. 1:** Overview of the proposed tract segmentation method.

Our approach of using the data as the model input has three advantages:

(1) It eliminates the need to compute intermediate parameters (e.g., fiber orientation distribution or tractogram), thereby avoiding the associated computational errors [23, 22]. If the goal is tract segmentation, there is no need to incur those errors by going through intermediate computations.

(2) It improves the generalizability of the method with respect to different acquisition schemes. If, for example, the input is the tractogram, the tract segmentation results can be significantly influenced by the tractography method that is used to compute the tractogram. Moreover, computation of intermediate parameters may demand especial measurement schemes that may be unavailable at test time. For example, methods that are based on fiber orientation distribution typically require high angular resolution measurements, which can result in a loss of accuracy if such measurements are not available [4, 34].

(3) It offers a highly effective data augmentation method during both training and test/inference. Data augmentation during training improves the training of large deep learning models with limited data. It is especially common in applications such as medical imaging where labeled data are costly to obtain. Test-time data augmentation, on the other hand, can be used to improve prediction accuracy and also to estimate prediction uncertainty [1, 16, 17]. Our train- and test-time data augmentation strategies are explained below.

Let us denote the set of b0-normalized measurements in a scan with  $\{x(q_i)\}_{i=1}^m$ , where  $q_i$  is the unit vector indicating the gradient direction for the  $i^{\text{th}}$  measurement. During training, in each iteration we select a subset of size 6-12 from the  $m$  measurements  $\{x(q_j)\}_{j \in S \subseteq \{1, \dots, m\}}$ , chosen uniformly at random without

replacement. We select these measurements such that the gradient directions for each pair of measurements are far apart in the  $q$  space, using an approach similar to [10, 27]. We use the selected measurement subset (after projecting onto the SH basis) as input to the model. This can act as a highly effective and computationally-efficient data augmentation strategy as it presents a different view of the input to the model in each training iteration.

During inference, we use  $n$  different measurement subsets, selected similarly as in training described above, to predict  $n$  different segmentations. Let us denote the segmentation probability map for a specific tract with each of these measurement subsets as  $\{y_k\}_{k=1}^n$ . We compute the voxel-wise average of these predictions to obtain a final segmentation prediction, which we denote with  $\bar{y}$ . Furthermore, we can compute a measure of disagreement between these  $n$  predictions to estimate segmentation uncertainty. Disagreement between segmentation predictions is usually quantified using metrics of volume overlap or surface distance [29]. Each of these metrics quantifies the segmentation error from a narrow perspective. Furthermore, these metrics discard the probability information by binarizing the segmentations. Recent segmentation uncertainty quantification methods have also followed a purely voxel-wise approach [15, 33], which ignores the spatial distribution of the segmentation probabilities. To characterize the disagreement in a way that accounts for the complete probability distribution of the predicted segmentations, we use a method based on the Wasserstein Distance, also known as earth mover’s distance (EMD) [21]. Given two probability distributions  $p$  and  $q$  defined on the same metric space, this distance is defined as  $\text{EMD}(p, q) = \inf_{\gamma \in \Gamma(p, q)} \mathbf{E}_{(x, y) \sim \gamma} d(x, y)$ , where  $d$  is a distance measure and  $\Gamma(p, q)$  is the set of joint probability distributions whose marginals are equal to  $p$  and  $q$ . Intuitively, if  $p$  and  $q$  are considered as two piles of earth, EMD is the cost of turning one into the other. Although EMD can be easily quantified for scalar variables, to the best of our knowledge there are no methods for computing EMD for probability distributions in  $\mathbb{R}^2$  or  $\mathbb{R}^3$ . Here, we adopt an approximation that was originally proposed in [35] for comparing multi-dimensional histograms. We demonstrate this computation for a simple  $3 \times 3$  histogram in Figure 2. Given a pair of multi-dimensional histograms (or probability distributions), the method first unfolds the histograms as shown in the example in Figure 2 and finds a minimum distance pairing between the two. The distance between the two histograms is defined as the sum of the pair-wise distances in the pairing.

$$\begin{aligned}
 f = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad g = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} & \xrightarrow{(1)} \quad \begin{aligned} \text{UF}(f) &= \{(1, 1), (2, 2), (2, 2), (3, 3)\} \\ \text{UF}(g) &= \{(1, 1), (1, 2), (2, 2), (2, 3)\} \end{aligned} \\
 & \xrightarrow{(2)} \quad \begin{aligned} \rho(f, g) &= \mu\{(1, 1), (1, 1)\} + \mu\{(2, 2), (2, 1)\} + \mu\{(2, 2), (2, 2)\} + \mu\{(3, 3), (2, 3)\} \\ &= 0 + 1 + 0 + 1 = 2 \end{aligned}
 \end{aligned}$$

**Fig. 2:** An illustration of the extension of the Wasserstein Distance to multi-dimensional signals, proposed in [35].

Based on this approximation, we compute the EMD between two segmentation probability maps in  $\mathbb{R}^3$  as  $\text{EMD}(p, q) = \sum_{t=0}^1 d(P(t), Q(t))$ , where  $P$  is the cumulative sum of unfolded  $p$  as shown in Figure 2 and the same for  $Q$ , and  $d$  computes the  $\ell_2$  distance between the paired  $P$  and  $Q$ . This computation requires that the two inputs have the same mass, which we satisfy by normalizing the segmentations to have a unit sum. Furthermore, to reduce the computation time, we reduce the size of the segmentation volumes by a factor of 4 in each dimension via cubic interpolation. Given the set of  $n$  segmentation predictions computed as explained above, we estimate the segmentation uncertainty as  $u = \frac{1}{n} \sum_k \text{EMD}(y_k, \bar{y})$ .

## 2.2 Implementation details

The segmentation network was implemented in TensorFlow 1.6 and run on an NVIDIA GeForce GTX 1080 GPU on a Linux machine with 64 GB of memory and 20 CPU cores. The network takes 3D patches of size  $96^3$  voxels as input and estimate the tract segmentation map for that patch. The network input has 6 channels as described above. The network output has 41 channels for the 41 tracts considered in this work. A complete description of these tracts can be found in [34]. We merged the left and right sections of bilateral tracts, such as arcuate fasciculus, into one label. We trained the network to maximize the Dice similarity coefficient (DSC) between the predicted and ground-truth segmentation of the tracts using Adam [12] with a batch size of 1 and a learning rate of  $10^{-4}$ , which was reduced by half if after a training epoch the validation loss did not decrease. We compare our method with TractSeg [34]. TractSeg was shown to be vastly superior to many tractography dissection methods [34]. Therefore, we do not compare with those methods.

## 3 Experimental Results and Discussion

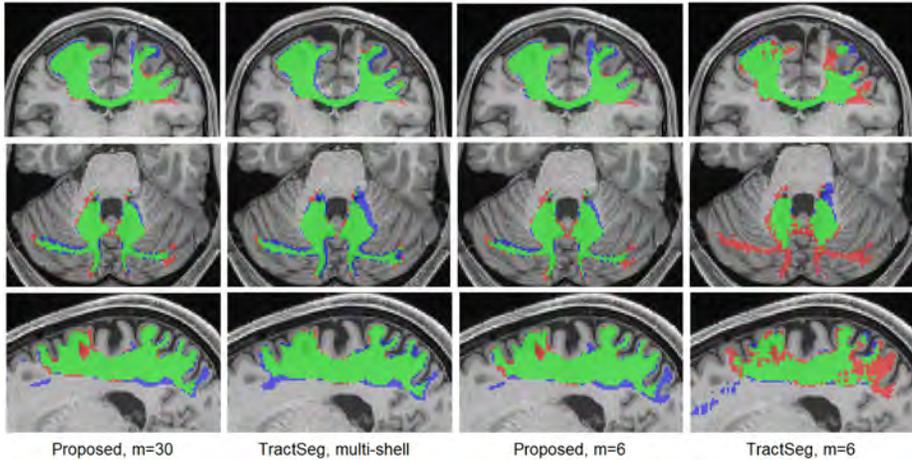
We applied the method on 105 subjects in the Human Connectome Project study [8, 31]. Manual segmentations of 72 tracts for these subjects are publicly available [34]. We followed a five-fold cross-validation approach, each time leaving 21 subjects for test and training on the remaining 84 subjects. Table 1 summarizes the performance of the proposed method and TractSeg. We report DSC, 95 percentile of the Hausdorff Distance (HD95), and average symmetric surface distance (ASSD). In addition to TractSeg, we compare our method with atlas-based segmentation (MAS), whereby 20 training images are registered to the test subject and the registration transforms are used to warp the segmentation labels from the training images to the test image. Voxel-wise averaging is then used to estimate the segmentations for the test image. We implemented this in two ways: MAS-FA, where we computed the registrations based on fractional anisotropy (FA) images using ANTS [2], and MAS-FOD, where we computed the registrations based on fiber orientation density images using mrregister [18].

**Table 1:** Segmentation performance of different methods. Asterisks denote significantly better results at  $p = 0.01$ .

Data	method	DSC	HD95 (mm)	ASSD (mm)
Multi-shell, m=270	TractSeg	$0.829 \pm 0.056$	$2.50 \pm 1.33$	$0.740 \pm 0.202$
b=1000 shell, m=90	TractSeg	$0.800 \pm 0.071$	$2.78 \pm 1.51$	$0.799 \pm 0.285$
	MAS-FA	$0.765 \pm 0.080$	$3.12 \pm 2.01$	$1.004 \pm 0.801$
	MAS-FOD	$0.792 \pm 0.076$	$2.76 \pm 1.55$	$0.815 \pm 0.289$
	Proposed	$0.826 \pm 0.056^*$	$2.48 \pm 1.28^*$	$0.746 \pm 0.201^*$
b=1000 shell, m=6	TractSeg	$0.687 \pm 0.155$	$5.29 \pm 6.51$	$1.471 \pm 1.427$
	MAS-FA	$0.760 \pm 0.089$	$3.30 \pm 2.27$	$1.124 \pm 1.038$
	MAS-FOD	$0.693 \pm 0.140$	$4.10 \pm 3.13$	$1.270 \pm 1.361$
	Proposed	$0.825 \pm 0.058^*$	$2.48 \pm 1.27^*$	$0.747 \pm 0.211^*$

Segmentation performance results are presented in Table 1. Figure 3 shows example tract segmentations predicted by our method and TractSeg. Our method using only the DTI measurements (b=1000) achieved segmentation accuracy that was very close to TractSeg using the multi-shell data with three times as many measurements. Paired t-tests did not show any significant differences (at  $p = 0.01$ ) between our method and TractSeg in terms of any of the three criteria. When TractSeg was applied on the b=1000 measurements, its performance was worse than our method in terms of all three criteria. To simulate under-sampled clinical scans, we selected 6 of the b=1000 measurements as proposed in [10, 27]. As shown in Table 1, the performance of our method remained almost unchanged, whereas the performance of TractSeg deteriorated significantly. Paired t-tests with a  $p = 0.01$  threshold showed that (1) the performance of our method did not change in terms of any of the three criteria on any of the 41 tracts when 6 measurements were used compared with 90 measurements. (2) Our method achieved significantly higher DSC and significantly lower HD95 and ASSD (all with  $p < 0.01$ ) with both 90 and six measurements compared with the other three methods. As shown in Figure 3, segmentations produced by our method are almost indistinguishable between 90 and 6 measurements. Although we cannot present the segmentation results for all tracts, Table 2 shows the mean DSC for six of the tracts, including anterior commissure and fornix which were the two most difficult tract to segment for our method and for TractSeg.

We further tested our method on scans of children between 2-8 years of age from an independent dataset [20]. Each scan in this dataset included 30 measurements in a single shell at b=750. We chose six measurements as input to our model as described above. We manually extracted 32 tracts from 12 different subjects on this dataset. Our method achieved DSC, HD95, and ASSD of  $0.786 \pm 0.076$ ,  $2.85 \pm 1.20$ , and  $1.017 \pm 0.291$ , respectively. Although this shows a drop in accuracy, it is a highly encouraging result given the fact that this was a completely independent test dataset that was different from our training dataset in two important ways: (1) subject age: young children (2-8 years) versus adults (21-36 years), and (2) measurement b-value of 750 versus 1000.



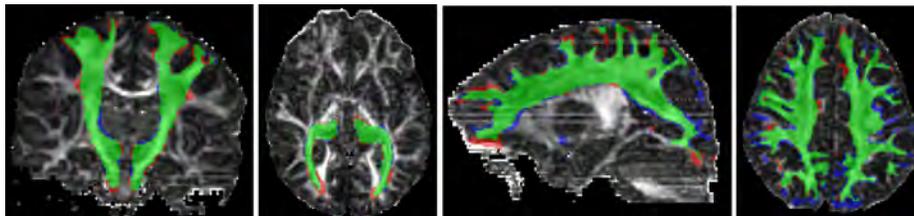
**Fig. 3:** Example segmentation results for the proposed method and TractSeg. Green indicates voxels with correct segmentation; red and blue indicate, respectively, false negatives and false positives.

**Table 2:** Mean DSC for six tracts. CCg: genu of corpus callosum; ThPr: thalamo-prefrontal; MCP: middle cerebellar peduncle; OpR: optic radiation; AC: anterior commissure; FNX: fornix. Asterisks denote significantl better results at  $p = 0.01$ .

Data	method	CCg	ThPr	MCP	OpR	AC	FNX
Multi-shell, m=280	TractSeg	0.867	0.883	0.871	0.827	0.696	0.689
Single shell, m=88	TractSeg	0.862	0.857	0.826	0.731	0.617	0.528
	Proposed	0.901*	0.897*	0.864*	0.810*	0.703*	0.675*
Single shell, m=6	TractSeg	0.772	0.783	0.740	0.704	0.366	0.436
	Proposed	0.903*	0.897*	0.857*	0.811*	0.680*	0.666*

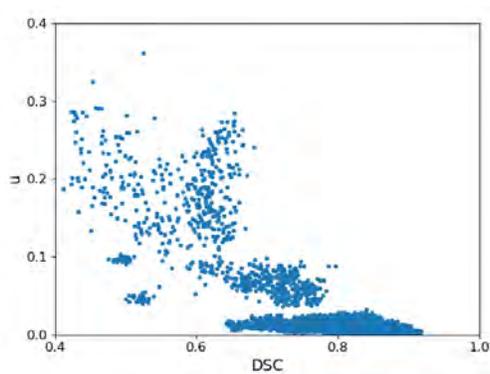
Compared with our method, TractSeg failed on this dataset, completely missing most of the tracts and achieving a mean DSC of 0.070. To further evaluate the reproducibility of our method on this dataset, we selected two disjoint subsets of six measurements from each scan and applied our method to segment the tracts. We computed the DSC between the tracts computed with the two measurement subsets. We did this for 100 scans, each from a different subject. The DSC for our method was  $0.867 \pm 0.041$ , whereas it was  $0.115 \pm 0.109$  for TractSeg. Example results for our method on this dataset are shown in Figure 4.

Figure 5 shows a plot of our proposed segmentation uncertainty,  $u$ , versus accuracy in terms of DSC. It shows that  $u$  is highly effective in identifying the less accurate segmentations. If we choose segmentations with a DSC of 0.70 and lower to be inaccurate, with a threshold of  $u = 0.30$  we can detect such segmentation with sensitivity=0.86, specificity=0.92, and accuracy=0.91. In Table 6 we compare method with the two standard methods based on estimation segmentation uncertainty: dropout, and ensemble methods. We refer to [15] for a



**Fig. 4:** Example segmentation results for our proposed method on an independent dataset. Green indicates voxels with correct segmentation; red and blue indicate, respectively, false negatives and false positives.

description of these methods. Our method achieves overall better results. Note that the ensemble method requires training of multiple models. We trained 10 models in this experiment, which increased the training time by a factor of 10.



**Fig. 5:** Plot of our proposed uncertainty  $u$  versus accuracy in terms of DSC.

Method	Acc	Sen	Spc
EMD	0.91	0.86	0.92
Drp	0.82	0.82	0.84
Ens	0.88	0.90	0.88

**Fig. 6:** Comparison of different methods for identifying inaccurate segmentations, defined as those with  $DSC < 0.70$ . (Drp: dropout; Ens: ensembles; Acc: accuracy; Sen: sensitivity; Spc: specificity.)

### 3.1 Computational time and other experiments

Training time for our method is approximately 24 hours. Our method segments a test image in 2.4 seconds. TractSeg requires approximately 60 seconds to segment an image. MAS methods require much longer time, approximately 3 minutes for MAS-FA and 12 minutes for MAS-FOD.

In recent years attention-based vision models have become very common in medical image segmentation. To experiment with one such model, we applied the model of [11], which has been developed specifically for 3D medical image segmentation. This model achieved a  $DSC$  of  $0.740 \pm 0.125$ , which was far lower segmentation performance those reported above.

## 4 Conclusions

Our method shows great promise in segmenting various white matter tracts. The appeal of our method is twofold: (1) Superior accuracy on under-sampled data that are typical of clinical scans, as clearly demonstrated by our results in Figure 3 and Tables 1 and 2. (2) Superior generalizability to multi-center data. This was clearly demonstrated in our experiment with an independent validation dataset, with some examples presented in Figure 4.

## Acknowledgment

This research was supported in part by the National Institute of Biomedical Imaging and Bioengineering, the National Institute of Neurological Disorders and Stroke, and Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (NIH) under award numbers R01HD110772, R01NS128281, R01NS106030, R01EB018988, R01EB031849, R01EB032366, and R01HD109395. This research was also partly supported by NVIDIA Corporation and utilized NVIDIA RTX A6000 and RTX A5000 GPUs. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NVIDIA. This work was also supported by the Swiss National Science Foundation (project 205321-182602). We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), Ecole polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE) and Geneva University Hospitals (HUG).

## References

1. Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., et al.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion* **76**, 243–297 (2021)
2. Avants, B.B., Tustison, N., Song, G., et al.: Advanced normalization tools (ants). *Insight j* **2**(365), 1–35 (2009)
3. Bazin, P.L., et al.: Direct segmentation of the major white matter tracts in diffusion tensor images. *Neuroimage* **58**(2), 458–468 (2011)
4. Descoteaux, M., Deriche, R.: High angular resolution diffusion mri segmentation using region-based statistical surface evolution. *Journal of Mathematical Imaging and Vision* **33**(2), 239–252 (2009)
5. Dong, X., et al.: Multimodality white matter tract segmentation using cnn. In: *Proceedings of the ACM Turing Celebration Conference-China*. pp. 1–8 (2019)
6. Eckstein, I., et al.: Active fibers: Matching deformable tract templates to diffusion tensor images. *Neuroimage* **47**, T82–T89 (2009)
7. Garyfallidis, E., et al.: Recognition of white matter bundles using local and global streamline-based registration and clustering. *NeuroImage* **170**, 283–295 (2018)
8. Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al.: The minimal preprocessing pipelines for the human connectome project. *Neuroimage* **80**, 105–124 (2013)
9. Isensee, F., et al.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods* **18**(2), 203–211 (2021)
10. Jones, D.K., Horsfield, M.A., Simmons, A.: Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **42**(3), 515–525 (1999)
11. Karimi, D., Vasylechko, S.D., Gholipour, A.: Convolution-free medical image segmentation using transformers. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 78–88. Springer (2021)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: *Proceedings of the 3rd International Conference on Learning Representations (ICLR)* (2014)
13. Labra, N., et al.: Fast automatic segmentation of white matter streamlines based on a multi-subject bundle atlas. *Neuroinformatics* **15**(1), 71–86 (2017)
14. Maier-Hein, K.H., et al.: The challenge of mapping the human connectome based on diffusion tractography. *Nature communications* **8**(1), 1–13 (2017)
15. Mehrtash, A., Wells, W.M., Tempany, C.M., Abolmaesumi, P., Kapur, T.: Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. *IEEE transactions on medical imaging* **39**(12), 3868–3878 (2020)
16. Moshkov, N., Mathe, B., Kertesz-Farkas, A., Hollandi, R., Horvath, P.: Test-time augmentation for deep learning-based cell segmentation on microscopy images. *Scientific reports* **10**(1), 1–7 (2020)
17. Nalepa, J., Myller, M., Kawulok, M.: Training-and test-time data augmentation for hyperspectral image segmentation. *IEEE Geoscience and Remote Sensing Letters* **17**(2), 292–296 (2019)
18. Raffelt, D., Tournier, J.D., Fripp, J., Crozier, S., Connelly, A., Salvado, O.: Symmetric diffeomorphic registration of fibre orientation distributions. *Neuroimage* **56**(3), 1171–1180 (2011)

19. Ratnarajah, N., Qiu, A.: Multi-label segmentation of white matter structures: application to neonatal brains. *NeuroImage* **102**, 913–922 (2014)
20. Reynolds, J.E., Long, X., Paniukov, D., Bagshawe, M., Lebel, C.: Calgary preschool magnetic resonance imaging (mri) dataset. *Data in brief* **29**, 105224 (2020)
21. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover’s distance as a metric for image retrieval. *International journal of computer vision* **40**(2), 99–121 (2000)
22. Schilling, K.G., Daducci, A., Maier-Hein, K., Poupon, C., Houde, J.C., Nath, V., Anderson, A.W., Landman, B.A., Descoteaux, M.: Challenges in diffusion mri tractography—lessons learned from international benchmark competitions. *Magnetic resonance imaging* **57**, 194–209 (2019)
23. Schilling, K.G., Nath, V., Hansen, C., Parvathaneni, P., Blaber, J., Gao, Y., Neher, P., Aydogan, D.B., Shi, Y., Ocampo-Pineda, M., et al.: Limits to anatomical accuracy of diffusion tractography using modern approaches. *NeuroImage* **185**, 1–11 (2019)
24. Schilling, K.G., et al.: Tractography dissection variability: What happens when 42 groups dissect 14 white matter bundles on the same dataset? *NeuroImage* **243**, 118502 (2021)
25. Siless, V., et al.: Anatomiccuts: Hierarchical clustering of tractography streamlines based on anatomical similarity. *NeuroImage* **166**, 32–45 (2018)
26. Siless, V., et al.: Registration-free analysis of diffusion mri tractography data across subjects through the human lifespan. *NeuroImage* **214**, 116703 (2020)
27. Skare, S., Hedehus, M., Moseley, M.E., Li, T.Q.: Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with mri. *Journal of magnetic resonance* **147**(2), 340–352 (2000)
28. Suarez, R.O., Commowick, O., Prabhu, S.P., Warfield, S.K.: Automated delineation of white matter fiber tracts with a multiple region-of-interest approach. *Neuroimage* **59**(4), 3690–3700 (2012)
29. Taha, A.A., Hanbury, A.: Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging* **15**(1), 1–28 (2015)
30. Tournier, J.D., Calamante, F., Connelly, A.: Robust determination of the fibre orientation distribution in diffusion mri: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage* **35**(4), 1459–1472 (2007)
31. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.M.H., et al.: The wu-minn human connectome project: an overview. *Neuroimage* **80**, 62–79 (2013)
32. Wakana, S., Jiang, H., Nagae-Poetscher, L.M., Van Zijl, P.C., Mori, S.: Fiber tract-based atlas of human white matter anatomy. *Radiology* **230**(1), 77–87 (2004)
33. Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S., Vercauteren, T.: Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing* **338**, 34–45 (2019)
34. Wasserthal, J., Neher, P., Maier-Hein, K.H.: Tractseg-fast and accurate white matter tract segmentation. *NeuroImage* **183**, 239–253 (2018)
35. Werman, M., Peleg, S., Rosenfeld, A.: A distance metric for multidimensional histograms. *Computer Vision, Graphics, and Image Processing* **32**(3), 328–336 (1985)
36. Wycoco, V., Shroff, M., Sudhakar, S., Lee, W.: White matter anatomy: what the radiologist needs to know. *Neuroimaging Clinics* **23**(2), 197–216 (2013)
37. Yendiki, A., et al.: Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. *Frontiers in neuroinformatics* **5**, 23 (2011)

# Spatio-temporal motion correction and iterative reconstruction of in-utero fetal fMRI

Athena Taymourtash<sup>1</sup>, Hamza Kebiri<sup>2,3</sup>, Ernst Schwartz<sup>1</sup>, Karl-Heinz Nanning<sup>1,4</sup>, Sébastien Tourbier<sup>2</sup>, Gregor Kasprian<sup>5</sup>, Daniela Prayer<sup>1,5</sup>, Meritxell Bach Cuadra<sup>2,3</sup>, and Georg Langs<sup>\*1</sup>

<sup>1</sup> Computational Imaging Research Lab, Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Austria.

<sup>2</sup> Medical Image Analysis Laboratory, Department of Radiology, Lausanne University Hospital and University of Lausanne, Switzerland.

<sup>3</sup> CIBM Center for Biomedical Imaging, Switzerland.

<sup>4</sup> Center for Biomedical Imaging and Neuromodulation, Nathan Kline Institute, Orangeburg, NY, United States.

<sup>5</sup> Division of Neuroradiology and Muskulo-skeletal Radiology, Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Austria.

georg.langs@meduniwien.ac.at

**Abstract.** Resting-state functional Magnetic Resonance Imaging (fMRI) is a powerful imaging technique for studying functional development of the brain *in utero*. However, unpredictable and excessive movement of fetuses have limited its clinical applicability. Previous studies have focused primarily on the accurate estimation of the motion parameters employing a single step 3D interpolation at each individual time frame to recover a motion-free 4D fMRI image. Using only information from a 3D spatial neighborhood neglects the temporal structure of fMRI and useful information from neighboring timepoints. Here, we propose a novel technique based on four dimensional iterative reconstruction of the motion scattered fMRI slices. Quantitative evaluation of the proposed method on a cohort of real clinical fetal fMRI data indicates improvement of reconstruction quality compared to the conventional 3D interpolation approaches.

**Keywords:** Fetal fMRI · image reconstruction · motion-compensated recovery · regularization.

## 1 Introduction

Functional magnetic resonance imaging (fMRI) offers a unique means of observing the functional brain architecture and its variation during development, aging, or disease. Despite the insights into network formation and functional growth of the brain, *in utero* fMRI of living human fetuses, and the developmental functional connectivity (FC), however, remain challenging. Since the fMRI acquisition

takes several minutes, unconstrained and potentially large movements of the fetuses, uterine contractions, and maternal respiration can cause severe artifacts such as in-plane blurring, slice cross-talk, and spin-history artifacts that likely vary over time. Without mitigation, motion artifacts can considerably affect the image quality, leading to a bias of subsequent conclusions about the FC of the developing brain.

Standard motion correction approaches, including frame-by-frame spatial realignment along with discarding parts of data with excessive motion, have been adopted so far to address motion artifacts of *in utero* fMRI [20, 10, 17]. More recently, cascaded slice-to-volume registration [13] combined with spin history correction [4], and framewise registration based on the 2<sup>nd</sup> order edge features instead of raw intensities [11] were suggested. These studies used 3D linear interpolation of motion scattered data at each volume independently to reconstruct the entire time series. Since *in utero* motion is unconstrained and complex, the regular grid of observed fMRI volumes becomes a set of irregularly motion scattered points possibly out of the field-of-view of the reconstruction grid, which might contain gaps in regions with no points in close proximity. Therefore interpolation in each 3D volume cannot recover the entire reconstruction grid.

Here we propose a new reconstruction method that takes advantage of the temporal structure of fMRI time series and rather than treating each frame independently, it takes both the spatial and the temporal domains into account to iteratively reconstruct a full 4D *in utero* fMRI image. The proposed method relies on super-resolution techniques that attracted increasing attention in structural fetal T2-weighted imaging, aiming to estimate a 3D high-resolution (HR) volume from multiple (semi-)orthogonal low resolution scans [5, 15, 3]. In case of fMRI, orthogonal acquisitions are not available, instead the reconstruction of a 4D image from a single sequence acquired over time is desired (An illustration of the problem is shown in Figure 1). Currently, existing single-image reconstruction methods are generally proposed for 3D structural MR images with isotropic voxels, while the effect of motion is implicitly modeled via blurring the desired HR image [14]. None of these methods have been tailored for 4D fMRI with high-levels of movement such as the fetal population.

Our contribution is threefold: (1) we develop a 4D optimization scheme based on low-rank and total variation regularization to reconstruct 4D fMRI data as a whole (2) we explicitly model the effect of motion in the image degradation process since it is the main source of gaps between interpolated slices; (3) we show the performance of our algorithm on the highly anisotropic *in utero* fMRI images. Experiments were performed on 20 real individuals, and the proposed method was compared to various interpolation methods.

## 2 Method

We first describe the fMRI image acquisition model and then its corresponding inverse problem formulation to recover a 4D artifact-free fMRI from a single scan of motion corrupted image, using low-rank and total variation regularizations.

## 2.1 The Reconstruction Problem

fMRI requires the acquisition of a number of volumes over time (fMRI time-series, bold signal) to probe the modulation of spontaneous (or task-related) neural activity. This activity is characterized by low frequency fluctuations ( $< 0.1Hz$ ) of bold signals and therefore temporal smoothing is often applied as a pre-processing step in fMRI analysis. We aim at estimating the motion-compensated reconstruction of fMRI time series ( $\mathcal{X} \in \mathbb{R}^{\hat{B} \times \hat{K} \times \hat{H} \times N}$ ) from observed motion-contaminated fMRI volumes ( $\mathcal{T} \in \mathbb{R}^{B \times K \times H \times N}$ ) that integrates temporal smoothing within a full 4D iterative framework. Both  $\mathcal{X}$  and  $\mathcal{T}$  are composed of  $N$  3D volumes  $\mathbf{X}_n, \mathbf{T}_n$  acquired over  $N$  timepoints. In MR image acquisition, a degradation process yields a low-resolution image from the latent high-resolution image:

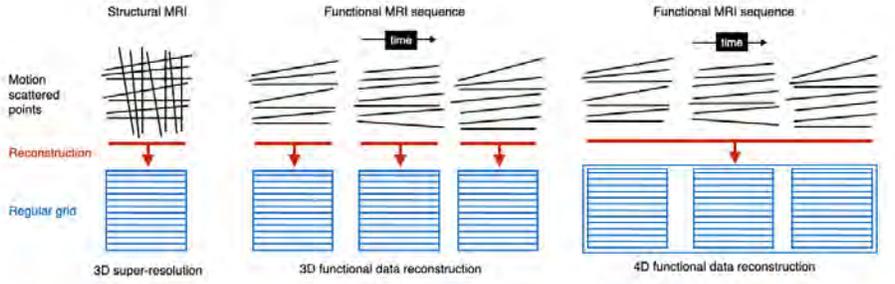
$$\mathbf{T}_n = DSM_n \mathbf{X}_n + z \quad (1)$$

where  $D$  is a 3D downsampling operator,  $S$  is a 3D blurring operator,  $M$  is the set of estimated motion parameters (three rotation and three translation parameters for each slice  $\mathbf{t}_{n,h} \in \mathbb{R}^{B \times K}$  of the volume  $\mathbf{T}_n$ , estimated prior to optimization (Sec. 3.1)), and  $z$  represents the observation noise. The application of  $M_n$  in the model here is equivalent to transforming each slice by the motion followed by resampling them on a 4D regular grid. Successful recovery of  $\mathcal{X}$  from the  $\mathcal{T}$  not only ensures the compensation of motion but also smoother bold signals due to the implicit temporal structure present in the data. However, since the Eq.(1) is ill-posed, direct recovery of  $\mathcal{X}$  is not possible without enforcing a prior. Hence, the reconstruction of the latent desired 4D image  $\mathcal{X}$  is achieved by minimizing the following cost function based on the inverse problem formulation:

$$\min_{\mathcal{X}} \sum_{n=1}^N \|DSM_n \mathbf{X}_n - \mathbf{T}_n\|^2 + \lambda \mathfrak{R}(\mathcal{X}) \quad (2)$$

where  $\mathfrak{R}(\mathcal{X})$  is a spatio-temporal regularization term, and  $\lambda$  balances the contributions of the data fidelity and regularization terms. We propose two regularization terms in this context, 4D low-rank for missing data recovery and total variation for preserving local spatial consistency.

**4D Low-Rank Regularization** Rank as a measure of nondegenerateness of the matrix, is defined by the maximum number of linearly independent rows or columns in the matrix. Since self-similarity is widely observed in fMRI images, low rank prior has been successfully used in matrix completion of censored fMRI time series [1]. Here we use low rank as a regularization term to help retrieve relevant information from all image regions. To compute the rank for a 4D image  $\mathcal{X}$ , we first unfold it into a 2D matrix along each dimension [7]. Specifically, suppose the size of  $\mathcal{X}$  is  $B \times K \times H \times N$ , we unfold it into four 2D matrices  $\{X_{(i)}, i = 1, 2, 3, 4\}$  with size of  $B \times (K \times H \times N)$ ,  $K \times (B \times H \times N)$ ,  $H \times (B \times K \times N)$ , and  $N \times (B \times K \times H)$  where  $X(i)$  means unfold  $\mathcal{X}$  along dimension  $i$ . Then we compute the sum of the singular values in each matrix for their



**Fig. 1.** Illustration of the image reconstruction using super-resolution technique. Over-sampling exists in case of 3D structural MRI (left panel), however, there is not enough data for separate reconstruction of each 3D fMRI volume (middle panel). Here we propose to reconstruct the whole 4D fMRI at once using both spatial and temporal data structure (right panel).

trace norms  $\|X_{(i)}\|_{tr}$ . Finally, the rank of  $\mathcal{X}$  is approximated as the combination of trace norms of all unfolded matrices [14]:

$$\mathfrak{R}_{rank}(\mathcal{X}) = \sum_{i=1}^4 \alpha_i \|X_{(i)}\|_{tr} \quad (3)$$

where  $\{\alpha_i\}$  are parameters satisfying  $\alpha_i \geq 0$ , and  $\sum_{i=1}^4 \alpha_i = 1$ . By minimizing this term, we obtain a low-rank approximation of  $\mathcal{X}$ . The low rank regularization is applied in the entire 4D data retrieving useful information for the reconstruction task from both spatial and temporal domains.

**Total Variation Regularization** Total variation (TV) is defined as integrals of absolute gradient of the signal. For a 4D functional image  $\mathcal{X}$ :

$$\mathfrak{R}_{tv}(\mathcal{X}) = \sum_{n=1}^N \int |\nabla \mathbf{X}_n| dbdkdh \quad (4)$$

where the gradient operator is performed in 3D spatial space. TV regularization has been largely adopted in image recovery because of its powerful ability in edge preservation [15, 14]. Here, we use TV in 3D space instead of 4D space based on the notion that primarily the spatial neighborhood exhibits consistency and thus TV in temporal domain may not be effective.

## 2.2 Optimization

The proposed 4D single acquisition reconstruction is thus formulated as below:

$$\min_{\mathcal{X}} \sum_{n=1}^N \|DSM_n \mathbf{X}_n - \mathbf{T}_n\|^2 + \lambda_{rank} \mathfrak{R}_{rank}(\mathcal{X}) + \lambda_{tv} \sum_{n=1}^N \mathfrak{R}_{tv}(\mathbf{X}_n) \quad (5)$$

**Algorithm 1** 4D motion-compensated reconstruction of fMRI time series**Input:** Single scan fMRI image  $\mathcal{T}$ , realignment parameters**Initialize:** The desired  $\mathcal{X}$  by resampling motion-transformed image  $\mathcal{T}$  with linear interpolation. Set auxiliary variable  $Y_i^{(0)} = 0, U_i^{(0)} = 0, i = 1, 2, 3, 4$ **while**  $\|\mathcal{X}^k - \mathcal{X}^{k-1}\| / \|\mathcal{T}\| > \varepsilon$  **do**    Update  $\mathcal{X}^k$  by using gradient descent:

$$\arg \min_{\mathcal{X}} \sum_{n=1}^N \left\| DSM_n \mathbf{X}_n^{(k-1)} - \mathbf{T}_n \right\|^2 + \sum_{i=1}^4 \frac{\rho}{2} \left\| \mathcal{X}^{(k-1)} - Y_i^{(k-1)} + U_i^{(k-1)} \right\|^2 + \lambda_{tv} \sum_{n=1}^N \int |\nabla \mathbf{X}_n^{(k-1)}| dbdkdh \quad (7)$$

    Update  $Y_i^{(k)}$  by using Singular Value Thresholding:

$$Y_i^{(k)} = \text{fold}_i \left[ SVT_{\lambda_{rank} \alpha_i / \rho} \left( \mathcal{X}_{(i)}^{(k)} + U_{i(i)}^{(k-1)} \right) \right] \quad (8)$$

    with  $\text{fold}_i(Y_{i(i)}) = Y_i$ 

$$\text{Update } U_i^{(k)} = U_i^{(k-1)} + \left( \mathcal{X}^{(k)} - Y_i^{(k)} \right) \quad (9)$$

**end while**

We employ the alternating direction method of multipliers (ADMM) algorithm to minimize the cost function in Eq.(5). ADMM has been proven efficient for solving optimization problems with multiple non-smooth terms [2]. Briefly, we first introduce redundant variables  $\{Y_i\}_{i=1}^4$  with equality constraints  $\mathcal{X}_{(i)} = Y_{i(i)}$ , and then use Lagrangian dual variables  $\{U_i\}_{i=1}^4$  to integrate the equality constraints into the cost function:

$$\min_{\mathcal{X}, \{Y_i\}_{i=1}^4, \{U_i\}_{i=1}^4} \sum_{n=1}^N \left\| DSM_n \mathbf{X}_n - \mathbf{T}_n \right\|^2 + \lambda_{rank} \sum_{i=1}^4 \alpha_i \left\| Y_{i(i)} \right\|_{tr} + \sum_{i=1}^4 \frac{\rho}{2} \left( \left\| \mathcal{X} - Y_i + U_i \right\|^2 - \left\| U_i \right\|^2 \right) + \lambda_{tv} \sum_{n=1}^N \int |\nabla \mathbf{X}_n| dbdkdh \quad (6)$$

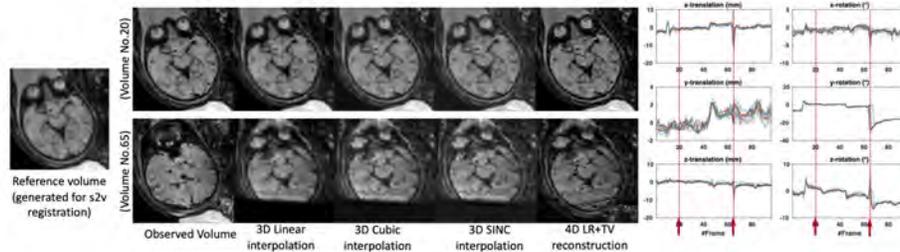
We break the cost function into subproblems for  $\mathcal{X}$ ,  $Y$ , and  $U$ , and iteratively update them. The optimization scheme is summarized in Algorithm 1.

### 3 Experiments and Results

#### 3.1 Data

**Data acquisition:** Experiments in this study were performed on 20 *in utero* fMRI sequences obtained from fetuses between 19 and 39 weeks of gestation. None of the cases showed any neurological pathology. Pregnant women were scanned on a 1.5T clinical scanner (Philips Medical Systems, Best, Netherlands) using single-shot echo-planar imaging (EPI), and a sensitivity encoding (SENSE) cardiac coil with five elements. Image matrix size was  $144 \times 144$ , with  $1.74 \times 1.74 \text{mm}^2$  in-plane resolution,  $3 \text{mm}$  slice thickness, a TR/TE of 1000/50 ms, and a flip angle of  $90^\circ$ . Each scan contains 96 volumes obtained in an interleaved slice order to minimize cross-talk between adjacent slices.

**Preprocessing:** For preprocessing, a binary brain mask was manually delineated on the average volume of each fetus and dilated to ensure it covered the fetal brain through all ranges of the motion. A four dimensional estimate of



**Fig. 2.** Reconstruction of *in-utero* fMRI for a typical fetus, and the estimated slice-wise realignment parameters. When motion is small (volume No.20) all interpolation methods recovered a motion compensated volume, and our approach resulted in a sharper image. In contrast, with strong motion relative to the reference volume (volume No.65), single step 3D interpolation methods are not able to recover the whole brain, and parts remain missing, whereas the proposed 4D iterative reconstruction did recover the entire brain.

the bias field for spatio-temporal signal non-uniformity correction in fMRI series was obtained using N4ITK algorithm [18] as suggested previously [12]. Intensity normalization was performed as implemented in mialSRTK toolkit [16]. Finally, motion parameters were estimated by performing a hierarchical slice-to-volume registration based on the interleaved factor of acquisition to a target volume created by automatically finding a set of consecutive volumes of fetal quiescence and averaging over them [13]. Image registration software package NiftyReg [8] was used for all motion correction steps in our approach. Demographic information of all 20 subjects as well as the maximum motion parameters estimated were reported in Supplementary Table S1.

### 3.2 Experimental Setting and Low-Rank Representation

We first evaluated to which extent *in utero* fMRI data can be characterized by its low-rank decomposition. The rapid decay of the singular values for a representative slice of our cohort is shown in Supplementary Figure S1. We used the top 30, 60, 90, and 120 singular values to reconstruct this slice and measured signal-to-noise ratio (SNR) to evaluate the reconstruction accuracy. The number of used singular values determines the rank of the reconstructed image. Using the top 90 or 120 singular values (out of 144), the reconstructed image does not show visual differences compared to the original image while it has a relatively high SNR (Figure S1).

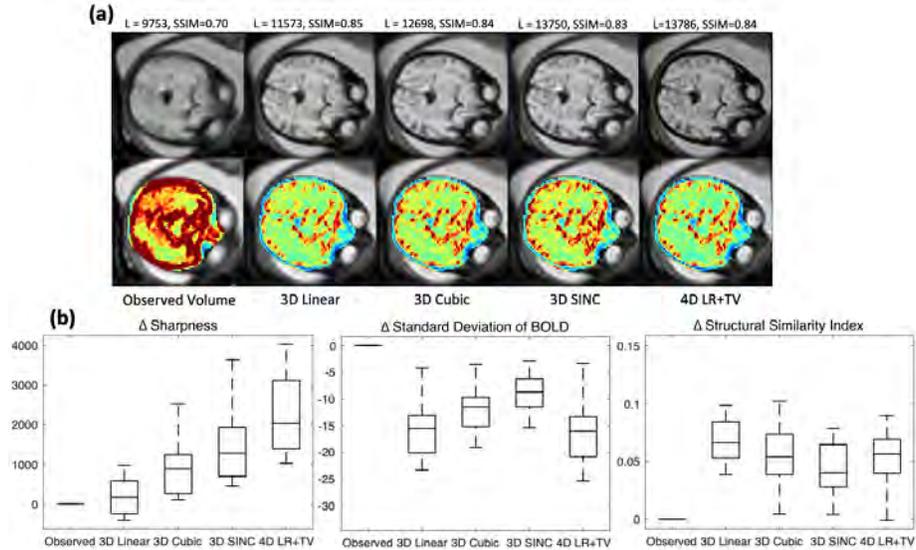
For the full 4D fMRI data of our cohort with the size of  $144 \times 144 \times 18 \times 96$ , four ranks, one for each unfolded matrix along one dimension is computed. Each is less than the largest image size 144. These ranks are relatively low in comparison to the total number of elements, implying *in utero* fMRI images could be represented using their low-rank approximations. We set  $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1/4$  as all dimensions are assumed to be equally important,  $\lambda_{\text{rank}} = 0.01$ ,  $\lambda_{\text{tv}} = 0.01$

were chosen empirically. The algorithm stopped when the difference in iterations was less than  $\varepsilon = 1e - 5$ .

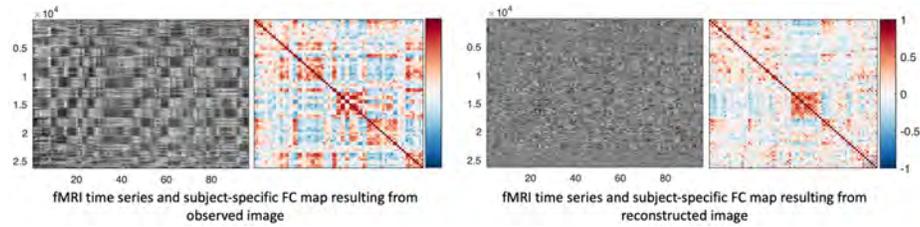
### 3.3 Evaluation of Image Reconstruction

A number of interpolation methods was employed to be compared with our reconstructed image including linear, cubic spline, and SINC interpolation. For each method, we applied the same realignment parameters as the ones used in our model, and in accordance with standard motion correction techniques, each 3D volumes of fMRI time series was interpolated separately. We quantified sharpness [9] of the average recovered image, standard deviation of bold signal fluctuations (SD) through-out the sequence, and the Structural Similarity Index (SSIM) which correlates with the quality of human visual perception [19]. Higher values of sharpness and SSIM, and lower values of SD are indicative of better recovery.

Figure 2 shows, from left to right, the reference volume, two corresponding slices in the observed image, and the results of different reconstruction methods. Volume No.20 exhibits minor motion, volume No.65 exhibits strong motion. The motion estimate plots on the right show their respective time points. The figure shows the recovered slices of these two volumes using 3D linear, cubic, SINC, and the proposed 4D LR+TV method, respectively. In the case of excessive complex



**Fig. 3.** Evaluation metrics for a typical fetus (a) and the whole cohort (b). Panel (a) shows an example slice in the average volume (top row) and voxel-wise standard deviation of the bold signal during fMRI acquisition. Higher Laplacian (sharpness) and SSIM, and lower standard deviation are indicative of better recovery. Panel (b) demonstrates these metrics in our fetal dataset.



**Fig. 4.** Carpet plot and functional connectivity maps achieved for an example subject using the observed fMRI time series and the time series recovered by 4D iterative reconstruction.

motion ( $30^\circ$  out of the plane combined with in-plane rotation and translation), the 3D interpolation methods cannot recover the whole slice as they utilize information only from the local spatial neighborhoods. The reconstructed slice by the proposed 4D iterative reconstruction approach recovers the image information, is sharper, and preserves more structural detail of the brain. Figure 3 shows a qualitative and quantitative comparison of reconstruction approaches. Figure 3(a) shows the average volume (top row), and the standard deviation of intensity changes over time (bottom row) for one subject. 4D reconstruction achieves sharper structural detail, and overall reduction of the standard deviation, which is primarily related to motion as described earlier. Although linear interpolation results in signals as smooth as the proposed method, severe blurring is observed in the obtained image by this approach. Figure 3 (b) provides the quantitative evaluation for the entire study population. The proposed method significantly ( $p < 0.01$ , paired-sample t-tests for each comparison) outperforms all comparison methods. The average gain of sharpness over the observed image is 2294 in our method compared to 1521 for 3D SINC, 959 for 3D Cubic, and 294 for 3D Linear, and the average reduction of SD relative to the observed image is -17 in our method compared to -9.34 for 3D SINC, -12.70 for 3D Cubic, and -16.50 for 3D Linear. The difference between linear interpolation and our approach did not reach the statistical significance level for SSIM ( $p = 0.28$ ). In summary, 4D iterative reconstruction reduces standard deviation over time, while increasing sharpness and recovered structure, which the 3D approaches failed to achieve.

### 3.4 Functional Connectivity Analysis

Figure 4 illustrates the impact of the accurate motion correction and reconstruction for the analysis of functional connectivity (FC) in the fetal population. The details of the pipeline employed for extracting subject-specific FC maps is explained in the supplementary material. When using the time series recovered by our proposed approach for FC analysis, the number of motion-corrupted correlations decreased significantly as visible in the *carpet plot* of signals, and the associated connectivity matrix.

## 4 Conclusion

In this work, we presented a novel spatio-temporal iterative 4D reconstruction approach for *in-utero* fMRI acquired while there is unconstrained motion of the head. The approach utilizes the self-similarity of fMRI data in the temporal domain as 4D low-rank regularisation together with total variation regularization based on spatial coherency of neighboring voxels. Comparative evaluations on 20 fetuses show that this approach yields a 4D signal with low motion induced standard deviation, and recovery of fine structural detail, outperforming various 3D reconstruction approaches.

## Acknowledgment

This work has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 765148, and partial funding from the Austrian Science Fund (FWF, P35189, I 3925-B27), in collaboration with the French National Research Agency (ANR), and the Vienna Science and Technology Fund (WWTF, LS20-065). This work was also supported by the Swiss National Science Foundation (project 205321-182602). We acknowledge access to the expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), Ecole polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE) and Geneva University Hospitals (HUG).

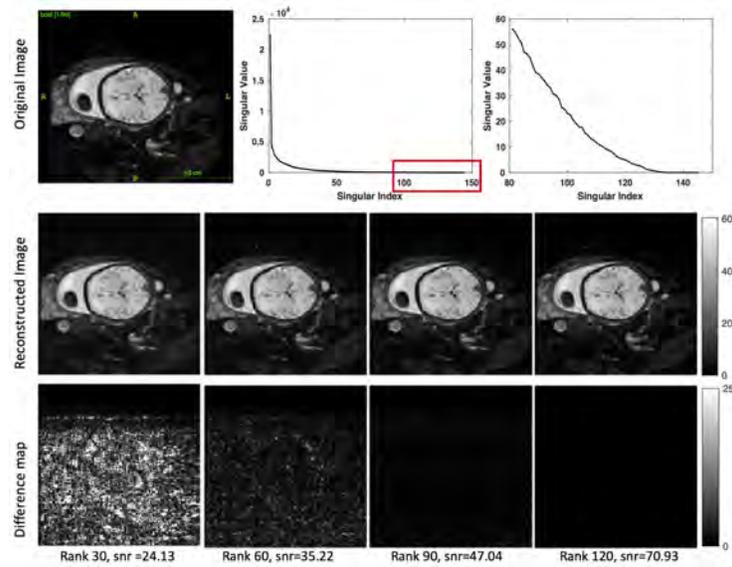
## References

1. Balachandrasekaran, A., Cohen, A.L., Afacan, O., Warfield, S.K., Gholipour, A.: Reducing the effects of motion artifacts in fmri: A structured matrix completion approach. *IEEE Transactions on Medical Imaging* **41**(1), 172–185 (2021)
2. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning* **3**(1), 1–122 (2011)
3. Ebner, M., Wang, G., Li, W., Aertsen, M., Patel, P.A., Aughwane, R., Melbourne, A., Doel, T., Dymarkowski, S., De Coppi, P., et al.: An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain mri. *NeuroImage* **206**, 116324 (2020)
4. Ferrazzi, G., Murgasova, M.K., Arichi, T., Malamateniou, C., Fox, M.J., Makropoulos, A., Allsop, J., Rutherford, M., Malik, S., Aljabar, P., et al.: Resting state fmri in the moving fetus: a robust framework for motion, bias field and spin history correction. *Neuroimage* **101**, 555–568 (2014)
5. Gholipour, A., Estroff, J.A., Warfield, S.K.: Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain mri. *IEEE transactions on medical imaging* **29**(10), 1739–1758 (2010)
6. Gholipour, A., Rollins, C.K., Velasco-Annis, C., Ouaalam, A., Akhondi-Asl, A., Afacan, O., Ortinau, C.M., Clancy, S., Limperopoulos, C., Yang, E., et al.: A normative spatiotemporal mri atlas of the fetal brain for automatic segmentation and analysis of early brain growth. *Scientific reports* **7**(1), 476 (2017)

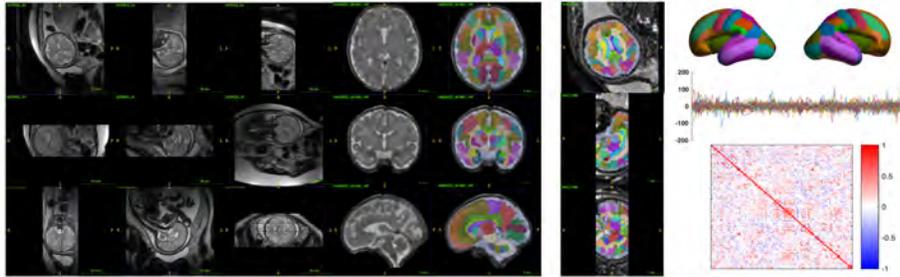
7. Liu, J., Musialski, P., Wonka, P., Ye, J.: Tensor completion for estimating missing values in visual data. *IEEE transactions on pattern analysis and machine intelligence* **35**(1), 208–220 (2012)
8. Modat, M., Cash, D.M., Daga, P., Winston, G.P., Duncan, J.S., Ourselin, S.: Global image registration using a symmetric block-matching approach. *Journal of medical imaging* **1**(2), 024003 (2014)
9. Pech-Pacheco, J.L., Cristóbal, G., Chamorro-Martinez, J., Fernández-Valdivia, J.: Diatom autofocusing in brightfield microscopy: a comparative study. In: *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*. vol. 3, pp. 314–317. IEEE (2000)
10. Rutherford, S., Sturmfels, P., Angstadt, M., Hect, J., Wiens, J., van den Heuvel, M.I., Scheinost, D., Thomason, M., Sripada, C.: Observing the origins of human brain development: Automated processing of fetal fmri. *bioRxiv* p. 525386 (2019)
11. Scheinost, D., Onofrey, J.A., Kwon, S.H., Cross, S.N., Sze, G., Ment, L.R., Papademetris, X.: A fetal fmri specific motion correction algorithm using 2 nd order edge features. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. pp. 1288–1292. IEEE (2018)
12. Seshamani, S., Cheng, X., Fogtmann, M., Thomason, M.E., Studholme, C.: A method for handling intensity inhomogenities in fmri sequences of moving anatomy of the early developing brain. *Medical image analysis* **18**(2), 285–300 (2014)
13. Seshamani, S., Fogtmann, M., Cheng, X., Thomason, M., Gatenby, C., Studholme, C.: Cascaded slice to volume registration for moving fetal fmri. In: *2013 IEEE 10th International Symposium on Biomedical Imaging*. pp. 796–799. IEEE (2013)
14. Shi, F., Cheng, J., Wang, L., Yap, P.T., Shen, D.: Lrtv: Mr image super-resolution with low-rank and total variation regularizations. *IEEE transactions on medical imaging* **34**(12), 2459–2466 (2015)
15. Tourbier, S., Bresson, X., Hagmann, P., Thiran, J.P., Meuli, R., Cuadra, M.B.: An efficient total variation algorithm for super-resolution in fetal brain mri with adaptive regularization. *NeuroImage* **118**, 584–597 (2015)
16. Tourbier, S., Velasco-Annis, C., Taimouri, V., Hagmann, P., Meuli, R., Warfield, S.K., Cuadra, M.B., Gholipour, A.: Automated template-based brain localization and extraction for fetal brain mri reconstruction. *NeuroImage* **155**, 460–472 (2017)
17. Turk, E.A., Luo, J., Gagoski, B., Pascau, J., Bibbo, C., Robinson, J.N., Grant, P.E., Adalsteinsson, E., Golland, P., Malpica, N.: Spatiotemporal alignment of in utero bold-mri series. *Journal of Magnetic Resonance Imaging* **46**(2), 403–412 (2017)
18. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: improved n3 bias correction. *IEEE transactions on medical imaging* **29**(6), 1310–1320 (2010)
19. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**(4), 600–612 (2004)
20. You, W., Evangelou, I.E., Zun, Z., Andescavage, N., Limperopoulos, C.: Robust preprocessing for stimulus-based functional mri of the moving fetus. *Journal of Medical Imaging* **3**(2), 026001 (2016)

**Table S1.** Gestational age, motion characteristics, and the achieved values of the evaluation metrics including sharpness (quantified by Laplacian), standard deviation of BOLD signal fluctuations (SD), and structural similarity index (SSIM) of all fetuses. The proposed method (LRTV) outperformed other reconstruction methods in terms of sharpness and SD significantly, and the difference between linear interpolation and our approach did not reach the statistical significance level for SSIM ( $p=0.28$ )

	GA (week+day)	max trans(mm)			max rot(°)			Sharpness					SD					SSIM				
		x	y	z	roll	pitch	yaw	Raw	Linear	Cubic	Sinc	LRTV	Raw	Linear	Cubic	Sinc	LRTV	Raw	Linear	Cubic	Sinc	LRTV
<b>S1</b>	19w5d	6.87	8.06	2.89	6.17	5.86	4.48	1159	1145	12504	13161	<b>13786</b>	38.29	<b>34.15</b>	34.76	35.40	34.94	0.9682	<b>0.9727</b>	0.9725	0.9721	0.9673
<b>S2</b>	22w5d	21.86	8.11	19.29	27.88	20.63	27.95	9383	9523	10187	10721	<b>11053</b>	77.43	59.11	64.36	68.70	<b>58.26</b>	0.8755	<b>0.9210</b>	0.9109	0.9031	0.9105
<b>S3</b>	28w3d	2.04	1.94	13.73	10.27	17.16	1.51	6610	7352	9094	10251	<b>10282</b>	59.85	46.05	50.09	53.63	<b>45.82</b>	0.8229	<b>0.8807</b>	0.8684	0.8577	0.8673
<b>S4</b>	29w2d	2.39	4.00	6.95	5.18	3.82	2.81	10636	10435	10776	11214	<b>11864</b>	59.38	<b>45.45</b>	50.31	54.59	45.95	0.8046	<b>0.8650</b>	0.8465	0.8323	0.8499
<b>S5</b>	27w1d	7.34	3.93	4.61	1.66	4.09	2.79	6533	9049	9059	9159	<b>9314</b>	59.22	46.85	49.51	52.00	<b>45.98</b>	0.8216	<b>0.8885</b>	0.8825	0.8772	0.8790
<b>S6</b>	25w5d	13.83	5.56	7.84	13.88	25.73	15.54	9134	9379	9977	10428	<b>10979</b>	77.49	56.74	61.79	65.98	<b>55.17</b>	0.8125	<b>0.8758</b>	0.8609	0.8503	0.8679
<b>S7</b>	26w3d	5.97	2.42	2.37	2.23	3.54	2.60	9168	8928	9462	9919	<b>10405</b>	57.61	49.71	52.14	54.56	<b>48.87</b>	0.8519	<b>0.8905</b>	0.8825	0.8742	0.8787
<b>S8</b>	30w2d	3.94	7.45	3.93	16.63	2.34	21.84	8942	8530	9156	9780	<b>10245</b>	75.52	<b>56.02</b>	59.96	63.29	56.57	0.8241	<b>0.8895</b>	0.8789	0.8703	0.8814
<b>S9</b>	29w5d	5.23	6.99	2.24	6.55	5.29	4.95	9753	11573	12698	113751	<b>13786</b>	76.71	<b>53.89</b>	58.10	61.90	53.97	0.7734	<b>0.8618</b>	0.8517	0.8432	0.8487
<b>S10</b>	31w6d	11.12	3.68	3.97	3.06	13.96	4.16	9223	9658	10368	11074	<b>11444</b>	63.67	48.41	53.08	57.19	<b>48.05</b>	0.7821	<b>0.8302</b>	0.8139	0.7999	0.8132
<b>S11</b>	32w4d	3.80	2.00	2.09	2.87	2.49	1.75	8715	8459	9205	9817	<b>9754</b>	69.73	53.88	57.32	60.43	<b>53.78</b>	0.7582	<b>0.8382</b>	0.8255	0.8149	0.8194
<b>S12</b>	29w5d	3.48	3.35	3.33	2.00	2.19	4.69	16344	16142	17332	18436	<b>19875</b>	72.10	55.71	61.49	66.50	<b>55.25</b>	0.7776	<b>0.8570</b>	0.8368	0.8205	0.8389
<b>S13</b>	36w1d	2.17	2.75	3.80	9.54	5.96	6.22	9862	9531	10088	10509	<b>11122</b>	87.33	54.95	58.48	61.65	<b>54.28</b>	0.7041	<b>0.8554</b>	0.8422	0.8302	0.8390
<b>S14</b>	34w5d	42.33	20.62	15.89	7.33	16.43	10.52	9230	9051	10164	11093	<b>12054</b>	64.13	45.16	49.59	53.43	<b>44.74</b>	0.6508	<b>0.7839</b>	0.7529	0.7293	0.7655
<b>S15</b>	23w6d	12.93	12.44	7.42	3.04	4.74	4.29	20016	20999	21363	21625	<b>23388</b>	70.93	55.90	60.42	64.48	<b>55.21</b>	0.8264	<b>0.8904</b>	0.8746	0.8615	0.8756
<b>S16</b>	29w4d	2.83	5.09	3.10	19.82	12.99	21.57	19293	19943	20304	20566	<b>22160</b>	77.39	54.03	58.29	61.99	<b>51.95</b>	0.7637	<b>0.8617</b>	0.8444	0.8314	0.8531
<b>S17</b>	29w3d	20.97	9.18	8.07	7.61	17.55	16.10	11628	11225	11735	12266	<b>13096</b>	79.02	57.64	64.22	69.69	<b>56.74</b>	0.7582	<b>0.8331</b>	0.8114	0.7955	0.8215
<b>S18</b>	24w4d	4.33	5.65	6.13	6.05	3.97	5.04	11628	11225	11735	12266	<b>13096</b>	64.57	53.39	56.06	58.36	<b>52.26</b>	0.8657	<b>0.9084</b>	0.9005	0.8942	0.8955
<b>S19</b>	34w3d	4.48	1.51	12.55	10.16	12.93	3.16	8454	8959	9404	9651	<b>10116</b>	88.02	72.88	74.73	76.55	<b>71.88</b>	0.7576	<b>0.8561</b>	0.8443	0.8344	0.8341
<b>S20</b>	39w2d	0.94	3.78	1.06	3.57	2.95	1.74	17439	17633	17725	17891	<b>21223</b>	70.36	<b>58.90</b>	60.18	61.68	58.92	0.7554	<b>0.8494</b>	0.8443	0.8367	0.8256



**Fig. S1.** Low rank approximation of *in-utero* fMRI of a fetus with gestational age of 34w+4d. Top row shows the original slice, singular-value plot, and zoomed singular-value plot of indices from 80 to 144. Bottom row shows the four reconstructed slices and their differences with the original image by using top 30, 60, 90, and 120 singular values, respectively. SNR values were reported at the bottom of each reconstructed slices.



**Fig. S2.** subject-specific functional connectivity analysis was performed in the native functional space. For this, cortical ROIs were first obtained using an automatic atlas-based segmentation of  $T_2$  scans acquired during the same session as the fMRI, using a publicly available atlas of fetal brain anatomy [6]. The resulting parcellation consists of 78 ROIs and was mapped to the motion corrected fMRI space using a rigid transformation. For each parcel, the average time series of all voxels was computed, and aCompCor nuisance regression and temporal filtering were performed subsequently. FC matrix was estimated by measuring Pearson's correlation between the average time series of parcels.