

Ground-truth effects in learning-based fiber orientation distribution estimation in neonatal brains

Rizhong Lin^{1,2,3,*}, Hamza Kebiri^{4,2,*}, Ali Gholipour^{5,6}, Yufei Chen³, Jean-Philippe Thiran^{1,2,4}, Davood Karimi⁵, and Meritxell Bach Cuadra^{4,2}

¹ Signal Processing Laboratory (LTS5), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

² Department of Radiology, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland
hamza.kebiri@unil.ch

³ College of Electronic and Information Engineering, Tongji University, Shanghai, China

⁴ CIBM Center for Biomedical Imaging, Switzerland

⁵ Computational Radiology Laboratory, Department of Radiology, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA

⁶ Department of Radiological Sciences, University of California Irvine, CA, USA

Abstract. Diffusion Magnetic Resonance Imaging (dMRI) is a non-invasive method for depicting brain microstructure *in vivo*. Fiber orientation distributions (FODs) are mathematical representations extensively used to map white matter fiber configurations. Recently, FOD estimation with deep neural networks has seen growing success, in particular, those of neonates estimated with fewer diffusion measurements. These methods are mostly trained on target FODs reconstructed with *multi-shell multi-tissue constrained spherical deconvolution* (MSMT-CSD), which might not be the ideal ground truth for developing brains. Here, we investigate this hypothesis by training a state-of-the-art model based on the U-Net architecture on both MSMT-CSD and *single-shell three-tissue constrained spherical deconvolution* (SS3T-CSD). Our results suggest that SS3T-CSD might be more suited for neonatal brains, given that the ratio between single and multiple fiber-estimated voxels with SS3T-CSD is more realistic compared to MSMT-CSD. Additionally, increasing the number of input gradient directions significantly improves performance with SS3T-CSD over MSMT-CSD. Finally, in an age domain-shift setting, SS3T-CSD maintains robust performance across age groups, indicating its potential for more accurate neonatal brain imaging.

Keywords: FOD estimation · Neonatal brain · Deep learning · SS3T-CSD · MSMT-CSD · Age domain shift

* R. Lin and H. Kebiri—Equal contribution.

1 Introduction

Diffusion Magnetic Resonance Imaging (dMRI) is a crucial tool for analyzing *in vivo* the microstructural organization of white matter fibers in the brain. This technique measures the random motion of water molecules, providing unique insights into brain connectivity and revealing abnormalities undetectable by other modalities both *in vivo* and non-invasively. The accurate depiction of white matter (WM) fibers using dMRI is particularly important in early brain development stages, where WM is still in maturation. Advanced methods, such as constrained spherical deconvolution (CSD) [30] and multi-shell multi-tissue CSD (MSMT-CSD) [12], are used to reconstruct orientation distribution functions (FODs), enabling both local and global quantitative analyses [4].

Supervised deep learning on high-quality datasets have enabled a growing number of methods [20, 21, 15, 11, 16, 14, 28, 26] to predict FODs from the raw diffusion signal [15, 11, 28] or its spherical harmonics (SH) representation [20, 21, 14, 16], or from a spherical deconvolution model [26]. Some of these methods [15, 17] have been successfully applied to developing brains for which acquisition times are more constrained, and hence a lower number of measurements is available. In particular, [17] employed a U-Net architecture to predict large patches of FODs from the newborns of the Developing Human Connectome Project (dHCP), using as few as six diffusion measurements and one b_0 as input, making it suitable for clinical acquisitions; [19] investigated age- and site-related domain shifts for fiber estimation within rapidly developing populations. The vast majority of these methods either focus on innovating at the architecture level, the learning choices and parameters, or the input format to the network.

The target ground truth (GT) used to be learned from is not often discussed. The majority of the works [20, 15, 11, 16, 14, 28, 26] use MSMT-CSD. However, its use for developing brains has been questioned [23] given the similarity of the response function profiles for each b -value and each tissue at this maturation stage. In fact, the mean signal across b -values of the gray matter (GM) lies within the WM intensities, as opposed to adult data where it lies below, allowing a clear distinction between tissues [23]. This distinction between tissues is also age-dependent, making the problem of estimating accurate FODs even more challenging for neonates. Recently, [7] has shown that employing the b_{1000} shell in the case of newborn (dHCP) data with a method relying on a single non-zero b -value measurement and a b_0 ⁷, named single-shell three-tissue CSD (SS3T-CSD) [5, 6], offers superior detection of crossing fibers compared to MSMT-CSD using all the available shells (b_{400} , b_{1000} and b_{2600}).

In this work, we explored the suitability of SS3T-CSD for deep learning-based FOD estimation in early development stages. We began by evaluating the consistency of GTs generated by MSMT-CSD, commonly used in this context, and SS3T-CSD, which we investigated here. Next, we extensively tested the state-of-the-art deep learning FOD estimation method on 207 newborn subjects, assessing the impact of training with SS3T-CSD compared to MSMT-CSD GT

⁷ b_x denotes diffusion weighting, where x is the b -value represented in s/mm^2 .

for the first time. Finally, we examined potential domain shift effects due to age by comparing the performance of models trained and tested on cohorts of similar and different ages using both SS3T-CSD and MSMT-CSD GTs.

2 Methodology

2.1 Datasets and processing

We used neonatal brain MRI data from the 3rd release of the Developing Human Connectome Project (dHCP) [8], which consists of newborn subjects scanned at the post-menstrual age (PMA) of 26 to 45 weeks using a 3T Philips Achieva scanner. The dMRI data release, having a multi-shell sequence with b -values 0, 400, 1000 and 2600 s/mm² and 20, 64, 88 and 128 measurements respectively, had been preprocessed using SHARD [3]. The resulting resolution is 1.5 × 1.5 × 1.5 mm³, covering a field of view of 100 × 100 × 64 voxels.

Following the approach in [16], we generated a WM mask by integrating the *White Matter* and *Brainstem* labels from the provided parcellation of the T2-weighted image of each subject, aligning it to the space of the respective dMRI image using ITK-SNAP [31]. Voxels with Fractional Anisotropy (FA) values (computed using MRtrix [29]) greater than 0.25 were also included.

For our experiments (detailed in Section 2.3), we selected two subsets, denoted as S_1 and S_2 , totaling 207 unique subjects:

- **Subset S_1** : 100 subjects (PMA at scan: [35.57, 44.29], median: 40.86, mean: 40.11, SD: 2.38).
- **Subset S_2** divided into two age groups:
 - *Early-stage group (S_{2e})*: 65 subjects (PMA at scan: [33.29, 37.86], median: 35.57, mean: 35.69, SD: 1.41).
 - *Late-stage group (S_{2l})*: 65 subjects (PMA at scan: [41.0, 45.14], median: 42.29, mean: 42.40, SD: 1.07).

Each subject’s data preparation for the deep learning pipeline (described in Section 2.2) involves sampling single-shell b1000 dMRI images from the full sequence as proposed in [27]. These images are normalized by a single b0 image to reduce b-value dependency and then projected onto the corresponding SH spaces to enhance acquisition independence.

2.2 FOD estimation learning-based model

The method [17] is based on learning a mapping between the SH representation of the raw diffusion signal and the FOD (Fig. 1, top).

Ground-truth models. MSMT-CSD extends (single-shell single-tissue) CSD [30] to accommodate multiple shells and multiple tissues; it aims to solve a constrained linear least squares problem using convex quadratic programming [12], in order to find the optimal FOD coefficients for each tissue, requiring acquisitions with at least 3 b -values. On the other hand, SS3T-CSD addresses

the optimization problem of finding the optimal FOD coefficients differently: initially, it fixes prior WM coefficients and estimates GM and cerebrospinal fluid (CSF) coefficients; then, it fixes the CSF coefficients and estimates WM and GM coefficients—this two-step iteration is repeated until convergence [5], allowing the method to rely on only two b-value samples (non-b0 and b0).

Network specifications. The network receives a $16 \times 16 \times 16$ patch of SH representations from n_{sig} single-shell diffusion measurements (b1000) as input and outputs the corresponding patch of FODs represented in the same SH basis ($\text{SH-}L_{\max}$ order 8). We train two models: one targeting FODs estimated from all 300 multi-shell measurements using MSMT-CSD [12], and the other from 88 b1000 and 20 b0 measurements using SS3T-CSD [5], as depicted upper in Fig. 1.

Training. For each experimental setting, training was conducted using the Adam optimizer [18] to minimize the ℓ_2 norm loss of the predicted 45 SH coefficients against the GT FOD SH coefficients, which were generated with MRtrix [29] or one of its forks, MRtrix3Tissue (<https://3Tissue.github.io>).

During each epoch, 128 patches were randomly extracted from non-empty volumes of each training subject, with each batch containing all 128 patches at a batch size of 1. The training regimen featured an initial learning rate of 1×10^{-4} , a dropout rate of 0.1, and used an early stopping strategy. A sliding window technique was employed during inference to process all patches consecutively.

Implementation and code availability. Training was performed on an NVIDIA RTX 2080 Ti GPU, each session lasting approximately 24 hours, using PyTorch [22], Lightning [9], and MONAI [2]. The code is available at https://github.com/Medical-Image-Analysis-Laboratory/dl_fiber_domain_shift.git.

2.3 Experiments

Experiments testing the efficacy of deep learning models in predicting FODs from MSMT- and SS3T-CSD using dMRI data are outlined in Fig. 1.

(i) Ground-truth consistency. Following the methodology described in [17], we assess GT consistency using MSMT-CSD and SS3T-CSD algorithms on two distinct half subsets of dMRI measurements: 150 measurements each for MSMT-CSD (10 b0's, 32 b400's, 44 b1000's, and 64 b2600's) and 54 measurements each for SS3T-CSD (10 b0's and 44 b1000's). This method evaluates the reliability of the GT algorithms by comparing FODs generated from equivalent but separate subsets, using 80 subjects randomly sampled from data subset S_1 .

(ii) Model assessment and ablation study on the number of input directions. Performance between models trained on SS3T-CSD and those trained

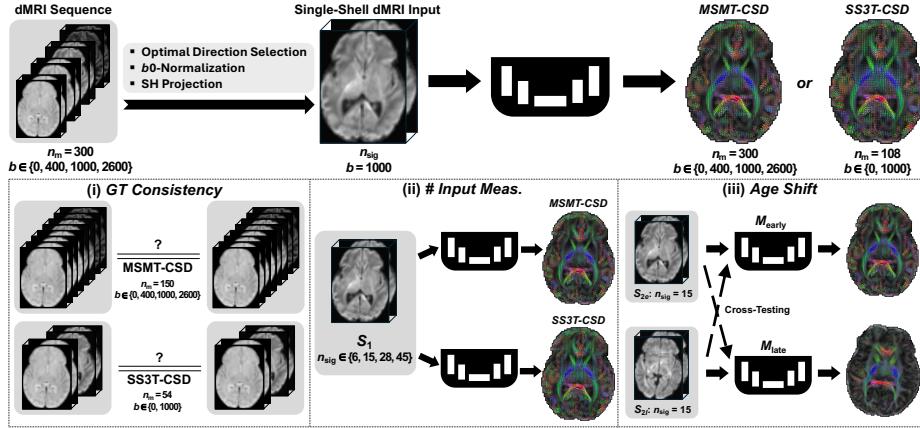


Fig. 1. Workflow overview. Single-shell b1000 dMRI images are sampled from the full series, normalized by b_0 and projected into SH space. These images are inputs to a deep learning model predicting GT FODs, using either MSMT-CSD (300 meas.) or SS3T-CSD (108 meas.). The experiments assess: (i) consistency of GT algorithms, (ii) impact of input quantity on model performance, and (iii) model effectiveness across different neonatal developmental stages. Illustrative brain images are from dHCP.

on MSMT-CSD was compared. This comparison was also performed across a different number of input directions to the network. The network input has n_{sig} measurements ($n_{sig} \in \{6, 15, 28, 45\}$) from the b1000 shell, each projected onto an SH basis with $L_{max} \in \{2, 4, 6, 8\}$. In this phase, the data subset S_1 was divided into groups of 70 for training, 15 for validation, and 15 for testing.

(iii) Age-related domain shifts. To explore age-related variations within each dataset, we conducted age-specific training and testing within and across these ages [19]. As mentioned in Section 2.1, in the dHCP dataset, subjects in S_2 were selected and divided into two age groups, denoted as *early* and *late*, respectively. Each age group consists of 65 subjects, further partitioned into splits of 40 for training, 10 for validation, and 15 for testing. We fixed the number of input measurements to 15 (corresponding to SH order 4) and trained separate models on the *early* and *late* age groups. Each model was then tested on its own age group (self-testing) as well as on the other age group (cross-age testing).

2.4 Evaluation metrics

Quantitative validation was performed based on the agreement rate (AR) [17] between the number of fibers estimated by the number of peaks extracted using Dipy [10] (with a mean separation angle of 45° , a maximum number of 3 peaks and relative peak threshold of 0.5), the angular error (AE) between those peaks [17], and the apparent fiber density (AFD) [24] error. For the AFD error, we

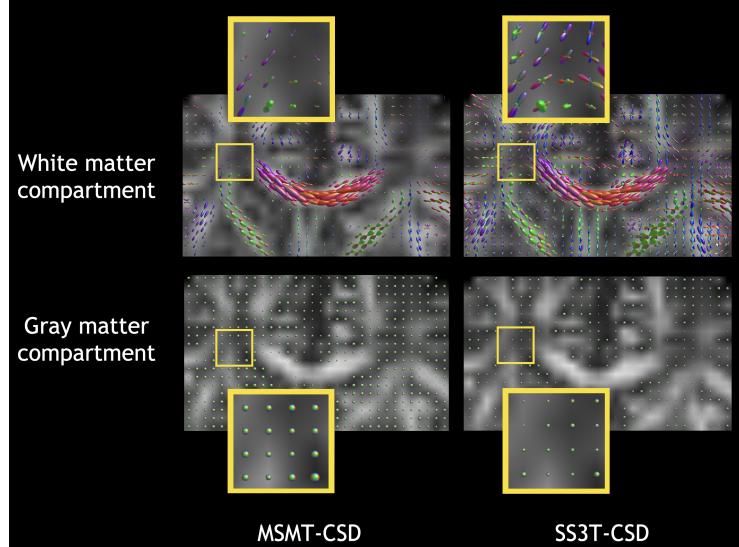


Fig. 2. Qualitative comparison of coronal slices of the FODs used as GT for the model, reconstructed with MSMT-CSD and SS3T-CSD, respectively, from the dMRI scan of a subject at 40 weeks PMA. Both gray matter and white matter compartments are displayed on the fractional anisotropy (FA) image computed from dMRI data. FOD estimation and visualization were performed with MRtrix [29].

report the mean absolute percentage error (MAPE) as the magnitude of the SS3T-CSD-derived FODs is higher than that of MSMT-CSD.

3 Results

Firstly, we show a qualitative comparison between the FOD reconstructed using the two GT methods for a neonate with 40 weeks PMA (Fig. 2), where we clearly see missing crossings in MSMT-CSD, likely due to the overestimation of the gray matter compartment, compared to SS3T-CSD.

3.1 Ground-truth consistency

Table 1 provides a comparison between the two GT methods. MSMT-CSD shows a higher agreement rate for single fibers but similar angular errors when compared to SS3T-CSD. In contrast, in the 2-fiber configuration, SS3T-CSD outperforms MSMT-CSD in agreement rate and exhibits slightly better angular errors. It is fair to mention that SS3T-CSD employs the minimum number of directions given the SH-order of the FOD (44 directions for 45 coefficients), compared to 140 directions by MSMT-CSD. Hence, the SS3T-CSD consistency errors reported here represent the upper bounds of expected errors. This is particularly

Table 1. Quantitative comparison of the consistency of GT FODs generated with MSMT-CSD and SS3T-CSD. Agreement Rate (AR, %) and Angular Error (AE, $^{\circ}$) for each fiber number configuration, and AFD MAPE (ΔAFD , %) are listed, together with the number of measurements (n_m) used and the corresponding b -values. AE is computed among fibers with agreed peak predictions only.

Method	b -values	n_m	1-Fiber		2-Fiber		3-Fiber		ΔAFD
			AR	AE	AR	AE	AR	AE	
MSMT-CSD	{0, 400, 1000, 2600}	150	88.8	6.92	45.6	14.91	52.8	25.96	2.87
SS3T-CSD	{0, 1000}	54	63.9	6.76	57.7	12.32	42.1	23.38	0.43

Table 2. Confusion matrices in the number of peaks agreement, normalized over all populations (in %) for MSMT-CSD and SS3T-CSD ground truth consistency.

MSMT-CSD				SS3T-CSD			
# Fibers	1	2	3	# Fibers	1	2	3
1	0.715	0.0446	0.0052	1	0.2955	0.0775	0.0062
2	0.0362	0.1013	0.021	2	0.0774	0.3517	0.0494
3	0.0035	0.0188	0.0544	3	0.0056	0.0532	0.0835

relevant for 3-fiber configurations, which are more susceptible to noise and thus less reliable. Importantly, closer examination of the confusion matrices (Table 2) reveals that the proportion of multiple fibers estimated by MSMT-CSD is around 23%, significantly lower than the literature-reported values of over 60% [13, 25], whereas SS3T-CSD estimates this proportion to be 61%.

3.2 Model assessment and number of input directions effect

Fig. 3 illustrates the comparative results in terms of agreement rate, angular error, and AFD error across varying numbers of input gradient directions, using MSMT-CSD and SS3T-CSD GTs, respectively.

We observe that SS3T-CSD consistently shows higher agreement rates and lower angular errors for multiple fibers compared to MSMT-CSD across all input directions, except for 6 directions where they are comparable. Increasing the number of input directions improves the network’s accuracy in predicting multiple fibers more significantly with SS3T-CSD than MSMT-CSD. Specifically, SS3T-CSD achieves an agreement rate of around 70%–76% for single and 2-fiber populations with 28–45 input directions, while MSMT-CSD’s 2-fiber population agreement rate remains around 30% compared to 88% for single fibers, even with 45 input directions. However, for single fiber populations, MSMT-CSD produces more accurate predictions. For AFD, SS3T-CSD shows lower errors across all input levels compared to MSMT-CSD, with the former exhibiting lower error variance. This can be attributed to the fact that AFD is computed using the same b -value as the input b -value measurement to the network for SS3T-CSD, whereas it is computed with all three shells in MSMT-CSD.

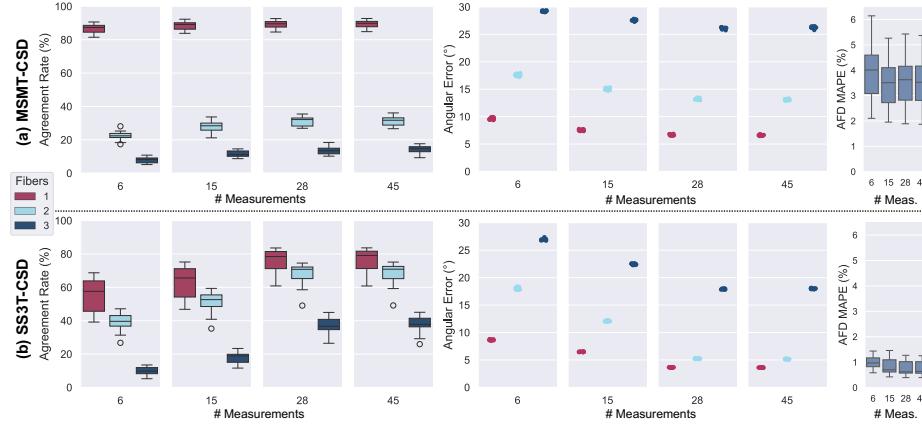


Fig. 3. Comparison of performance of the models based on incrementing numbers of input directions, using GT: (a) MSMT-CSD; (b) SS3T-CSD. AR and AE under different fiber number configurations and AFD Error are depicted.

3.3 Age domain shift

Fig. 4 presents the effects of age-specific training and testing on model performance using MSMT- and SS3T-CSD GTs. We first notice that when testing in the *early*-stage cohort, SS3T-CSD cross-age training is more robust to domain shift across all metrics, including single-fiber populations AR and AE, compared to MSMT-CSD. In fact, the drop for instance in 2-fibers agreement rate goes from around 28% to 18% for the latter, but only from 55% to 50% for the former. However, for the configuration when tested in the *late*-stage cohort, no significant difference can be observed between the two GTs, except for AFD where SS3T-CSD seems more robust to domain shift. In general, an increased variance is observed in cross-testing scenarios in for both SS3T-CSD- and MSMT-CSD-trained models, especially for AFD.

4 Conclusion

We have demonstrated the differences in the training of deep learning models using MSMT-CSD and SS3T-CSD for FOD estimation in neonatal brains. Compared to MSMT-CSD, SS3T-CSD improved accuracy and reliability in multiple fiber configurations, which is a major bottleneck in diffusion MRI and, in particular, for rapidly developing brains [1]. However, there is a drop in performance in single-fiber populations compared to MSMT-CSD, potentially due to higher values in gray matter compartments in GT MSMT-CSD that smoothens the overall estimated FODs, resulting in less predicted multi-fibers voxels, and vice-versa for SS3T-CSD and multiple fibers predictions. More b1000 measurements for SS3T-CSD can potentially lower this effect. Furthermore, SS3T-CSD showed

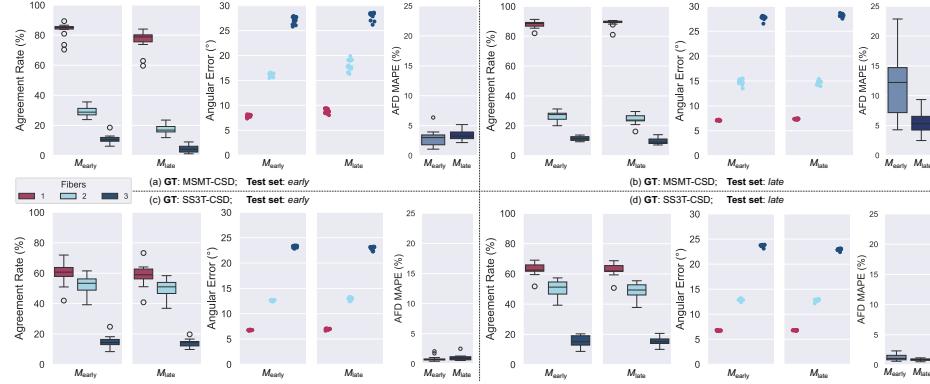


Fig. 4. Comparison of performance metrics for FOD estimation models trained and tested on *early* and *late* developmental stages using MSMT-CSD and SS3T-CSD ground truths. The agreement rate, angular error, and AFD error are depicted across models trained on *early* (M_{early}) and *late* (M_{late}) age groups.

robust performance across different age groups, reducing the performance drop caused by age domain shifts.

Future work will focus on evaluating SS3T-CSD in other pediatric populations, such as fetuses and babies, and exploring downstream performance in tractography. We will also investigate a variety of models beyond U-Net and different strategies to address domain shift challenges through data harmonization and domain adaptation techniques.

Ethics Statement. This study uses pre-approved dHCP data by UK Health Research Authority (REC reference: 14/LO/1169), requiring no additional ethical approval.

Acknowledgments. 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.

This research was partly supported by grants from the Swiss National Science Foundation (182602 and 215641); the US National Institutes of Health (NIH), including: the National Institute of Neurological Disorders and Stroke (R01NS106030 and R01NS128281), the National Institute of Biomedical Imaging and Bioengineering (R01EB032366), and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD110772); and the National Natural Science Foundation of China (62173252). The opinions expressed herein are solely those of the authors and do not necessarily reflect those of the funding agencies.

Disclosure of Interests. The authors declare no relevant competing interests regarding the content of this article.

References

1. Calixto, C. *et al.*: White matter tract crossing and bottleneck regions in the fetal brain. bioRxiv 2024.07.20.603804 (2024)
2. Cardoso, M.J. *et al.*: MONAI: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
3. Christiaens, D. *et al.*: Scattered slice SHARD reconstruction for motion correction in multi-shell diffusion MRI. NeuroImage **225**, 117437 (2021)
4. Descoteaux, M.: High angular resolution diffusion MRI: from local estimation to segmentation and tractography. PhD thesis, Université Nice Sophia Antipolis (2008).
5. Dhollander, T., Connelly, A.: A novel iterative approach to reap the benefits of multi-tissue CSD from just single-shell ($+b=0$) diffusion MRI data. In: ISMRM, p. 3010 (2016)
6. Dhollander, T., Raffelt, D., Connelly, A.: Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion MR data without a co-registered T1 image. In: ISMRM workshop on breaking the barriers of diffusion MRI (2016)
7. Dhollander, T. *et al.*: Improved white matter response function estimation for 3-tissue constrained spherical deconvolution. In: ISMRM, p. 0555 (2019)
8. Edwards, A.D. *et al.*: The Developing Human Connectome Project Neonatal Data Release. Frontiers in Neuroscience **16** (2022)
9. Falcon, W., The PyTorch Lightning team, PyTorch Lightning, (2024). <https://doi.org/10.5281/zenodo.11644096>.
10. Garyfallidis, E. *et al.*: Dipy, a library for the analysis of diffusion MRI data. Frontiers in Neuroinformatics **8** (2014)
11. Hosseini, S.M.H. *et al.*: CTtrack: A CNN+Transformer-based framework for fiber orientation estimation & tractography. Neuroscience Informatics **2**(4), 100099 (2022)
12. Jeurissen, B. *et al.*: Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. NeuroImage **103**, 411–426 (2014)
13. Jeurissen, B. *et al.*: Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. Human Brain Mapping **34**(11), 2747–2766 (2013)
14. Jha, R.R. *et al.*: Undersampled single-shell to MSMT fODF reconstruction using CNN-based ODE solver. Computer Methods and Programs in Biomedicine **230**, 107339 (2023)
15. Karimi, D. *et al.*: Learning to estimate the fiber orientation distribution function from diffusion-weighted MRI. NeuroImage **239**, 118316 (2021)
16. Kebiri, H. *et al.*: Deep learning microstructure estimation of developing brains from diffusion MRI: A newborn and fetal study. Medical Image Analysis **95**, 103186 (2024)
17. Kebiri, H. *et al.*: Robust Estimation of the Microstructure of the Early Developing Brain Using Deep Learning. In: MICCAI, pp. 293–303 (2023)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2015)
19. Lin, R. *et al.*: Cross-age and cross-site domain shift impacts on deep learning-based white matter fiber estimation in newborn and baby brains. In: IEEE ISBI (2024)
20. Lin, Z. *et al.*: Fast learning of fiber orientation distribution function for MR tractography using convolutional neural network. Medical Physics **46**(7), 3101–3116 (2019)

21. Nath, V. *et al.*: Deep learning reveals untapped information for local white-matter fiber reconstruction in diffusion-weighted MRI. *Magnetic Resonance Imaging* **62**, 220–227 (2019)
22. Paszke, A. *et al.*: PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc. (2019)
23. Pietsch, M. *et al.*: A framework for multi-component analysis of diffusion MRI data over the neonatal period. *NeuroImage* **186**, 321–337 (2019)
24. Raffelt, D. *et al.*: Apparent fibre density: a novel measure for the analysis of diffusion-weighted magnetic resonance images. *NeuroImage* **59**(4), 3976–3994 (2012)
25. 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)
26. da Silva, M.O. *et al.*: FOD-Swin-Net: angular super resolution of fiber orientation distribution using a transformer-based deep model. In: *IEEE ISBI* (2024)
27. 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)
28. Spears, T., Fletcher, P.T.: Learning spatially-continuous fiber orientation functions. In: *IEEE ISBI* (2024)
29. 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)
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. Yushkevich, P.A., Gao, Y., Gerig, G.: ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In: *IEEE EMBC*, pp. 3342–3345 (2016)