How can spherical-CNNs improve ML-based diffusion MRI parameter estimation?

Tobias Goodwin-Allcock[1] TOBIAS.GOODWIN-ALLCOCK@UCL.AC.UK
Jason McEwen [2]
Robert Gray [3]
Parashkev Nachev [3]
Hui Zhang[1]
[1] *Department of Computer Science and Centre for Medical Image Computing, UCL, London, United Kingdom*
[2] *Kagenova Limited, Guildford, United Kingdom*
[3] *Department of Brain Repair & Rehabilitation, Institute of Neurology, UCL, London, United Kingdom*

Editors: Under Review for MIDL 2022

Abstract

This paper demonstrates spherical convolutional neural networks (spherical CNN) offer distinct advantages over conventional fully connected networks (FCN) at estimating rotation-invariant indices of tissue microstructure from diffusion MRI (dMRI). Such microstructure indices are valuable for identifying pathology and quantifying its extent. However, current clinical practise commonly acquires dMRI data consisting of only 6 diffusion weighted images (DWIs), limiting the accuracy and precision of estimated microstructure indices.

Machine learning (ML) has been proposed to address this challenge. However, existing ML-based methods are not robust to differing dMRI sampling schemes, nor are they rotation equivariant. Lack of robustness to sampling schemes requires a new network to be trained for each scheme, complicating the analysis of data from multiple sources. The lack of rotational equivariance potentially prevents these methods from estimating the same microstructure viewed from different angles consistently. Here, we show spherical CNNs represent a compelling alternative that is robust to new gradient schemes as well as offering rotational equivariance. We show the latter can be leveraged to decrease the number of training datapoints required.

1. Introduction

This work aims to demonstrate the advantages of spherical convolutional neural networks (S-CNNs) over standard fully-connected network (FCN) methods for estimating microstructure indices from diffusion MRI data. Diffusion MRI (dMRI) is a non-invasive method of measuring diffusion of water in tissue. This methodology has led to discoveries in neuroscience (Scholz et al., 2009; Sagi et al., 2012) as well as providing methods to identify and quantify pathology in neural tissue.

Diffusion MRI acquires data in the form of diffusion-weighted images (DWI). Each DWI voxel contains a measurement sensitised to diffusion along a direction of choice known as the

diffusion gradient direction. This measurement depends on the angle between the diffusion gradient direction and the underlying tissue microstructure we probe. For a full picture of microstructure within a voxel, multiple DWIs are acquired from multiple directions which collectively are known as a gradient scheme. The measurements drawn from the corresponding voxel of these DWIs can be fitted to mathematical models, such as the diffusion tensor (DT) (Basser et al., 1994), to capture the diffusion occurring at the voxel in all directions. This is typically in the form of the apparent diffusion coefficient (ADC) profile, a spherical function. From these models, orientation-independent indices can be derived to describe the microstructure; example DT indices are mean diffusivity and fractional anisotropy (FA) which characterise the size and shape of diffusion respectively. The computation of these indices from DWIs is known as dMRI parameter estimation. However, signal-noise limits estimation fidelity for conventional models, requiring more DWI acquisitions than are routinely acquired clinically (Jones, 2013).

As in many other fields, dMRI parameter estimation has recently been revolutionised by exploiting deep learning, yielding greatly increased accuracy in comparison with conventional methods when estimating with a limited number of DWIs (Golkov et al., 2016; Aliotta et al., 2019). However, current deep-learning models, which are based on FCNs, are ignorant of the gradient scheme adopted by a given DWI acquisition, rendering these methods inflexible to new sampling schemes. This complicates the application of a model across data acquired from multiple sources. Moreover, these networks do not exhibit rotational equivariance and therefore lacking generalisation between different views of the same microstructure.

Previous attempts to include the relationship between the gradient schemes and image acquisitions (Chen et al., 2020; Park et al., 2021) do not utilize the topological features of the sphere inherent in the fundamental structure of the modelled signal. S-CNNs, recently proposed as an alternative to FCNs (Sedlar et al., 2021), provide a more natural solution to this problem. However, currently there exists no direct evidence of the theoretical benefits of S-CNNs, such as rotational equivariance and robustness to different gradient schemes. Here we aim to provide the very first empirical evidence of these advantages in the context of estimating rotation-invariant diffusion parameters.

2. Methods

The study aims to assess two potential advantages of S-CNNs over FCNs for estimating rotation-invariant microstructure indices. First, are S-CNNs more robust to differing gradient schemes than FCNs? And second, does rotational equivariance provide benefit for estimation of microstructure indices? This will be assessed in experiment 1 and 2 respectively.

2.1. Common experimental settings

All of the networks evaluated in this paper are supervised neural networks, therefore, they require a high quality ground truth microstructure index to be estimated by the network coupled with the corresponding DWI inputs. FA is chosen as the microstructure index due to its potential sensitivity to differing gradient schemes and it's estimation should be

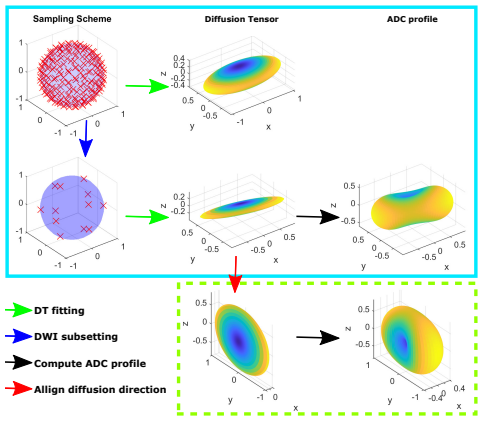

Figure 1: Here we show generation of GT DT estimates along with generating 6 DWI subsets and the noisy ADC profile for both training and testing in both experiments.

independent on the primary fibre orientation. Furthermore, we choose the input to the estimation algorithm to use 6 DWIs to mimic routine clinical acquisitions. For the high quality ground truth output, a dataset is required that contains a sufficiently large number of DWIs to provide accurate estimation of FA. For this reason, we have chosen the dMRI data from the Human Connectome Project (HCP), which includes 90 DWIs at b=1000$\frac{s}{mm^2}$. Ground-truth (GT) FA maps are computed from the complete set of DWIs. Subsets of 6-directional DWIs are sampled from the 90 DWI to be in maximal agreement with the chosen gradient schemes. Training was performed with one participant and, to show generalisation, data from 12 unseen subjects were used for testing.

Table 1: Table describing all of the different network architectures used in this paper.

Network	Input	Network architecture
Conventional FCN (C-FCN)	6 directional DWI measurements at each voxel, i.e. 6 scalar values	3 hidden layers with sizes [100,100,10] and Relu activation
Spherical FCN (S-FCN)	Interpolated noisy ADC profile, sampled with a 33x64 matrix	Identical to C-FCN
Spherical CNN (S-CNN)	Identical to S-FCN	Hybrid spherical CNN model (see appendix A)

2.2. Networks

All of the network architectures used across both experiments are described in table 1, all of these networks are designed to estimate FA voxel-by-voxel. The S-CNN architecture we use is known as the hybrid spherical CNN architecture (Cobb et al., 2021), this has been shown to estimate parameters with high fidelity for a variety of tasks. We employ a version of this model tailored for regression to estimate FA. Training parameters for FCNs were chosen in accordance with the literature: three hidden layers all consisting of [100,100,10] units with ReLU activation function. All networks were trained with the Adam optimizer

for 50 epochs with learning rate set to 0.001, the batch size 32 and the loss metric MSE. These values are consistent with the literature.

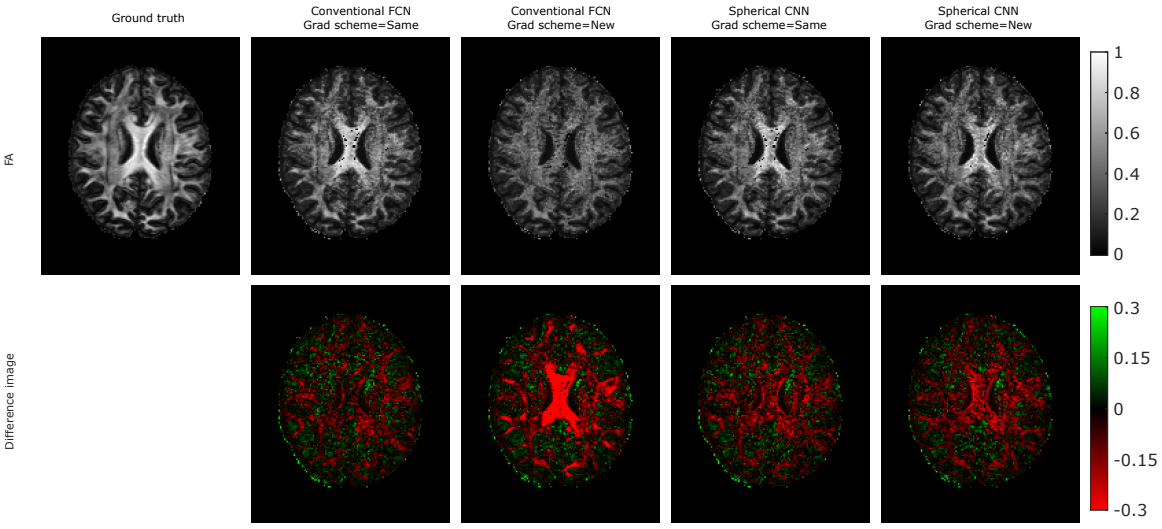

Figure 2: Results of experiment 1 are visualised on an subject unseen during training. Conventional fully-connected networks are compared against the SCNN model at estimating FA with either the same gradient scheme as used in training or a new sampling scheme.

2.3. Generating densely sampled spherical signals from 6 directional DWI at each voxel with sampled ADC profile

Densely sampled spherical signals are required as input for both S-CNNs and the S-FCN. The densely sampled spherical signals are generated for each voxel by exploiting the 1-to-1 mapping between the six-directional DWI measurements and diffusion tensor. Calculation of a diffusion tensor from just 6 DWIs will inevitably provide a noisy estimate of the DT, as the number of measurements is equal to the number of unknowns. The noisy DT is used to generate a ADC profile that is then highly sampled with a 33x64 matrix. This sampled ADC profile corresponds exactly to the original 6-direction DWI input. This process is visually described in Figure 1.

2.4. Experiment 1

This experiment addresses the first question. Gradient scheme dependence is evaluated by training using a subset of the DWI generated by one gradient scheme and testing on a subset of the DWIs generated with this same gradient scheme as well as a different scheme. We use two common gradient schemes, Skare (Skare et al., 2000) and Jones (Jones et al., 1999) to extract 6-directional DWIs from the 90 DWIs. FCN follows the standard practice of feeding each DWI signal to a different node so we call this model conventional FCN

(C-FCN) . This is compared against the S-CNN, whose input is the highly sampled ADC profile. The information for both of these networks is derived from one voxel at a time.

2.5. Experiment 2

The second question is evaluated with experiment 2. The benefit of rotation equivariance is explored by training both S-CNN and FCN on data with every primary fibre orientation aligned along a single axis. This is achieved by adapting the dense ADC sampling algorithm to produce a dataset with a single primary fibre orientation. In the dense ADC sampling algorithm, after the noisy DT is estimated, the shape of the noisy DT is extracted by eigendecomposition. Following this the primary and secondary directions of diffusion are set to the anterior-posterior and superior-inferior axes, respectively; see visual description in the light blue dashed box in Figure 1. The sampled noisy ADC profile aligned along the anterior-posterior axis is as input for both the S-CNNs and the FCN. For consistency with S-CNN, the input of the FCN must be adapted to be compatible with the spherical signal spherical FCN (S-FCN). At test time tissue with primary fibre orientation in all directions will be estimated: networks exhibiting rotational equivariance are expected to perform equally well here. To show further benefits of rotational equivariance we test to see if the rotational equivariance allows S-CNNs to estimate with high fidelity when starved of training data points. For this, a S-CNN is trained with only 10 % of the total training datapoints. The gradient scheme is kept consistent between training and testing, both using Skare.

2.6. Evaluation

In both experiments, quantitative measurements of the estimation error are calculated with root mean square error (RMSE) over specific FA ranges over the whole image. Statistical significance between distributions is quantified with paired t-tests. Qualitative assessment is made with maps of estimates and errors relative to the ground truth. In the second experiment the distribution of estimation error over the primary fibre orientation is evaluated.

3. Results and discussion

Results of experiment 1 are shown in figures 2 and 3. Figure 2 compares the C-FCN and the S-CNN models,when the gradient scheme is either consistent, or inconsistent, between training and testing. Note minimal differences between the C-FCN model and the S-CNN model when the gradient schemes at test time are consistent with training. However, when applied to the new sampling scheme, where the DWI directions do not correspond to the training set, the C-FCN model performs far worse. Note the error is structured, maximal in regions of high FA, whereas, the S-CNN model remains faithful to the ground truth image, with only a small increase in error.

It is expected that the C-FCN model is unable to accurately estimate regions of high FA when testing and training gradient schemes do not correspond. In highly anisotropic microstructure, the extent of signal attenuation greatly depends on the direction in which it is measured. This explains why areas of high FA, an example being the corpus callosum

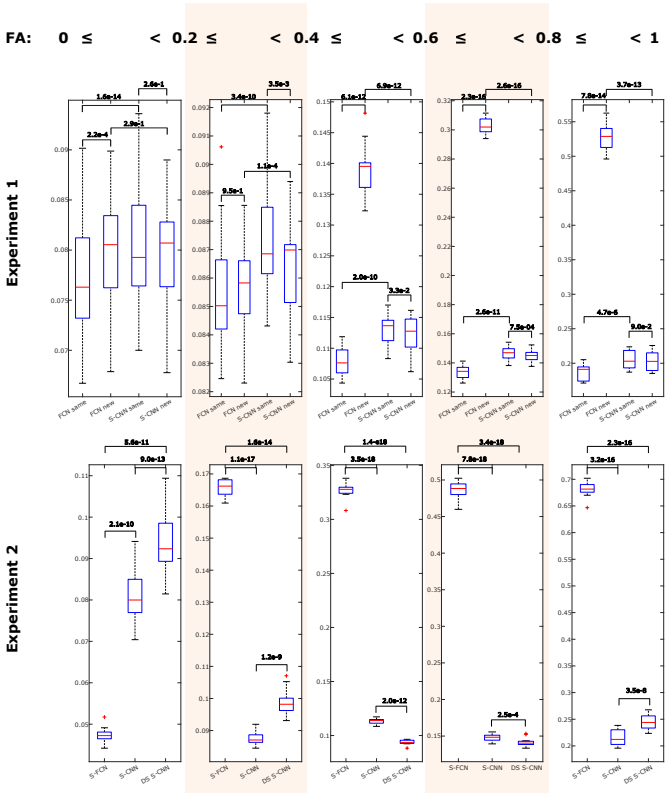

Figure 3: Boxplots of the RMSE over the 12 testing subjects for both experiments, with p-values between distributions from paired t-test shown.

at the centre of these brain images, are poorly estimated by the C-FCN model. Whereas, for regions of isotropic tissue microstructure the signal is less dependent on the gradient direction and therefore estimations on these voxels will be less affected.

Figure 3 substantiates this claim quantitatively with measures of the RMSE over the 12 subjects. We see for anisotropic signals (FA > 0.4) the conventional FCN performance is significantly worse ($p < 6e\text{-}12$) when applied to a new gradient scheme than on the training scheme, whereas, the S-CNN models estimation fidelity does not decrease when applied to a scheme it hasn't been trained on.

The results of experiment 2 are shown in figures 4,5 and 3. Figure 4 shows the effect of estimating the full brain volume using networks trained only on tissue microstructure aligned with the anterior-posterior axis. For the S-FCN model the FA is consistently underestimated in regions where the underlying tissue microstructure does not align with anterior-posterior direction (e.g. the corpus callosum which consists of left-right white matter tracts). In contrast, the S-CNN models estimate FA with high accuracy for tensors in all directions, and the noise is far less structured than that of the S-FCN. This is mirrored

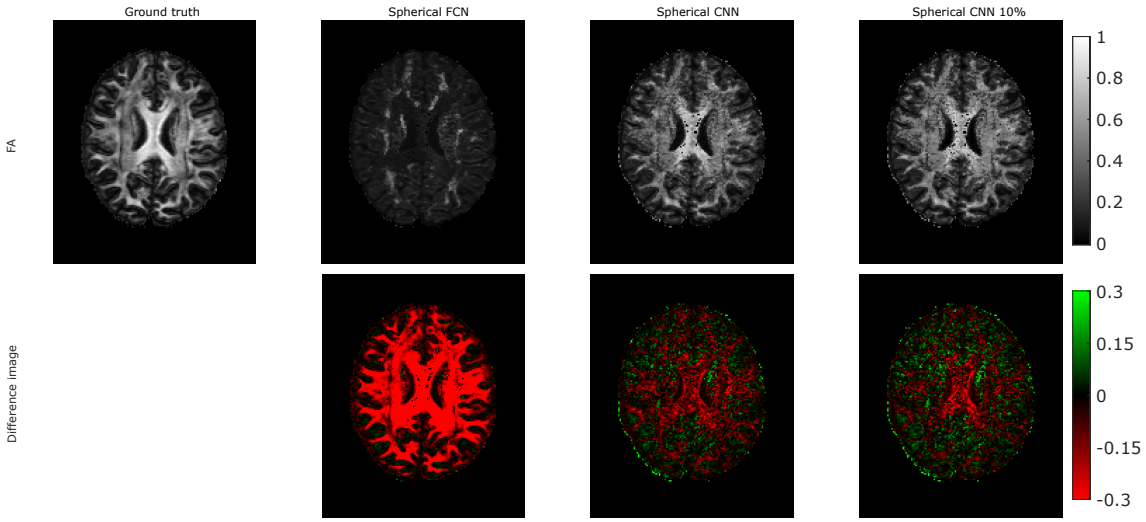

Figure 4: Comparison between the FCN model with spherical input and the S-CNN model at estimating the FA when trained only with microstrucure oriented along the anterior-posterior axis. These are also compared against a training data starved S-CNN.

in the quantitative measurements over the 12 testing subjects, shown in figure 3. The S-FCN model only well estimates isotropic signals (FA$<$ 0.2), notice that the maximum difference between the median RMSE for different methods is 0.04. When the signal becomes isotropic, the S-FCN model estimates the signal significantly (p=1.1e-17) worse than the S-CNN models with a difference of 0.08 between the median RMSEs. As the signals get more anisotropic the S-FCN model performs even worse until in the top FA bracket ($0.8 \leq$FA<1) the median RMSE between the S-FCN and the S-CNN models is 0.5, 10 times the difference in error of the isotropic signals.

These results are supported by Figure 5, where the distribution of the absolute error over the full range of primary fibre orientations is shown. The S-FCN model well estimates tissue microstructure aligned with the anterior-posterior axis, seen during training. However the error quickly grows as the primary fibre orientation deviates from this axis. This adverse feature is not exhibited by the S-CNN model as the estimation error is low and independent of training dataset distribution of the primary fibre orientation . The lack of rotational equivariance in FCNs hinders estimation performance when generalising to microstructure with primary fibre orientation not sampled in the training distribution. This has potential for orientation bias in the training dataset to lead to poor estimation for under-sampled directions.

In figures 4 and 3, another advantage of rotational equivariance is shown. Performance of the S-CNN network is not greatly changed when only a tenth of the training dataset is

used. This is shown in the maps of FA estimated by the data starved S-CNN model in figure 4 and the RMSE over the 12 testing subjects shown in figure 3. The robustness to a lack of training data is due to S-CNNs being able to generalise from microstructure across the full range of primary fibre orientation. Therefore, during training only a diverse distribution of the microstructure shape is required and not their orientations as well. This reduces the number of training datapoints required for good estimation. This is a real benefit when training with simulated data, as the microstructure orientation does not need to be sampled the shape parameters may be more densely sampled.

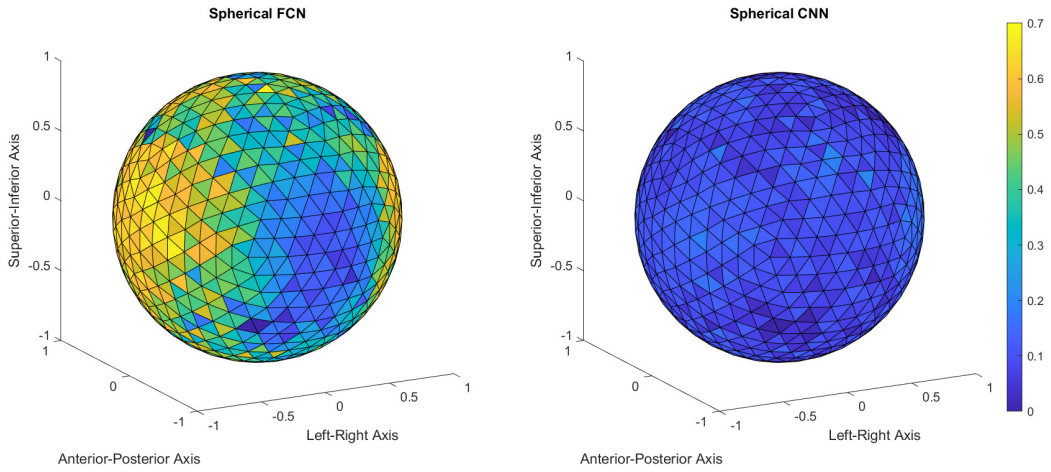

Figure 5: The distribution of the estimation error over the sphere is compared between the spherical fully-connected network and spherical-CNN.

4. Conclusion

In this work we explore the advantages of S-CNNs for dMRI parameter estimation over conventional fully-connected networks. Representing diffusion-weighted imaging as a spherical signal is here demonstrated to introduce robustness to the ordering of gradients absent from C-FCNs, at no cost to fidelity. This removes the need to retrain a new network for every gradient sampling scheme, a feature especially beneficial when combining data from multiple sites. S-CNN is shown to be superior to FCN methods because of its rotational equivariance property. This enables the network to encode information about the pattern of the signal irrespective of primary fibre orientation. This obviates sampling of diffusion primary fibre orientation, reducing the number of samples needed to cover the full parameter space. In further work we shall expand to other microstructure indices and evaluate if microstructure orientation bias in common dMRI datasets causes poor estimation.

References

Eric Aliotta, Hamidreza Nourzadeh, Jason Sanders, Donald Muller, and Daniel B. Ennis. Highly accelerated, model-free diffusion tensor MRI reconstruction using neural networks. *Medical Physics*, 46(4):1581–1591, 5 2019. ISSN 2473-4209. doi: 10.1002/mp.13400. URL https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13400http://files/329/Aliottaetal.-2019-Highlyaccelerated,model-freediffusiontensorMR.pdfhttp://files/130/mp.htmlhttps://onlinelibrary.wiley.com/doi/abs/10.1002/mp.13400.

P J Basser, J Mattiello, and D Lebihan. Estimation of the Effective Self-Diffusion Tensor from the NMR Spin Echo. *Journal of Magnetic Resonance, Series B*, 103(3):247–254, 6 1994. ISSN 1064-1866. doi: 10.1006/jmrb.1994.1037. URL http://www.sciencedirect.com/science/article/pii/S1064186684710375http://files/282/S1064186684710375.html.

Geng Chen, Yoonmi Hong, Yongqin Zhang, Jaeil Kim, Khoi Minh Huynh, Jiquan Ma, Weili Lin, Dinggang Shen, and Pew Thian Yap. Estimating Tissue Microstructure with Undersampled Diffusion Data via Graph Convolutional Neural Networks. In *MICCAI 2020*, volume 12267 LNCS, pages 280–290. Springer Science and Business Media Deutschland GmbH, 2020. ISBN 9783030597276. doi: 10.1007/978-3-030-59728-3{_}28.

Oliver J Cobb, Christopher G. R. Wallis, Augustine N Mavor-Parker, Augustin Marignier, Matthew A Price, Mayeul D'Avezac, and Jason D McEwen. Efficient Generalized Spherical CNNs. In *ICLR 2021*, 2021. URL http://arxiv.org/abs/2010.11661.

V Golkov, A Dosovitskiy, J I Sperl, M I Menzel, M Czisch, P Sämann, T Brox, and D Cremers. q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans. *IEEE Transactions on Medical Imaging*, 35(5):1344–1351, 2016. ISSN 0278-0062. doi: 10.1109/TMI.2016.2551324. URL http://files/203/Golkovetal.-2016-q-SpaceDeepLearningTwelve-FoldShorterandMod.pdfhttp://files/204/7448418.html.

D K Jones, M A Horsfield, and A Simmons. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine*, 42(3):515–525, 1999. ISSN 07403194. doi: 10.1002/(SICI)1522-2594(199909)42:3⟨515::AID-MRM14⟩3.0.CO;2-Q.

Derek K. Jones. Gaussian Modeling of the Diffusion Signal. In *Diffusion MRI: From Quantitative Measurement to In vivo Neuroanatomy: Second Edition*, pages 87–104. Elsevier Inc., 11 2013. ISBN 9780123964601. doi: 10.1016/B978-0-12-396460-1.00005-6.

Juhyung Park, Woojin Jung, Eun Jung Choi, Se Hong Oh, Jinhee Jang, Dongmyung Shin, Hongjun An, and Jongho Lee. DIFFnet: Diffusion parameter mapping network generalized for input diffusion gradient schemes and b-values. *IEEE Transactions on Medical Imaging*, 2021. ISSN 1558254X. doi: 10.1109/TMI.2021.3116298.

Yaniv Sagi, Ido Tavor, Shir Hofstetter, Shimrit Tzur-Moryosef, Tamar Blumenfeld-Katzir, and Yaniv Assaf. Learning in the Fast Lane: New Insights into Neuroplasticity. *Neuron*, 73(6):1195–1203, 3 2012. ISSN 0896-6273. doi: 10.1016/J.NEURON.2012.01.025.

Jan Scholz, Miriam C. Klein, Timothy E.J. Behrens, and Heidi Johansen-Berg. Training induces changes in white-matter architecture. *Nature Neuroscience 2009 12:11*, 12(11): 1370–1371, 10 2009. ISSN 1546-1726. doi: 10.1038/nn.2412. URL https://www.nature.com/articles/nn.2412.

Sara Sedlar, Abib Alimi, Théodore Papadopoulo, Rachid Deriche, and Samuel Deslauriers-Gauthier. A Spherical Convolutional Neural Network for White Matter Structure Imaging via dMRI. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12903 LNCS:529–539, 9 2021. ISSN 16113349. doi: 10.1007/978-3-030-87199-4{_}50. URL https://link.springer.com/chapter/10.1007/978-3-030-87199-4_50.

Stefan Skare, Maj Hedehus, Michael E. Moseley, and Tie Qiang Li. 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. ISSN 10907807. doi: 10.1006/jmre.2000.2209. URL https://pubmed.ncbi.nlm.nih.gov/11097823/.

Appendix A. Hybrid spherical CNN architecture

The S-CNN architecture used in this paper is an implementation of the hybrid spherical CNN architecture (Cobb et al., 2021). This consists of a spherical convolutional layer which maps the spherical image onto a function on SO(3), followed by a convolution layer on SO(3). This is followed by a restricted generalized convolution layer and then 3 channel-wise tensor-product activations and two restricted generalized convolutions, until the final restricted generalized convolution which produces a rotationally invariant representation of its input. Following this, there is a dropout layer (dropout rate of 0.5), a single fully connected layer (30 nodes) and ReLu activation followed by another dropout layer (dropout rate of 0.5). The final output layer of size one follows, transforming its input with a sigmoid activation function.