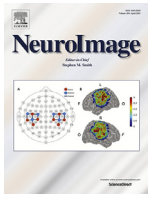
NeuroImage 269 (2023) 119900

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage



dtiRIM: A generalisable deep learning method for diffusion tensor imaging

E.R. Sabidussi [a,*], S. Klein [a], B. Jeurissen [b,c], D.H.J. Poot [a]

[a] Erasmus MC University Medical Center, Department of Radiology and Nuclear Medicine, Rotterdam, the Netherlands
[b] imec-Vision Lab, Department of Physics, University of Antwerp, Antwerp, Belgium
[c] Lab for Equilibrium Investigations and Aerospace, Department of Physics, University of Antwerp, Antwerp, Belgium.

ARTICLE INFO

Keywords:
Diffusion-weighted MRI
Diffusion tensor imaging
Deep learning
Recurrent inference machines
Generalisable

ABSTRACT

Diffusion weighted MRI is an indispensable tool for routine patient screening and diagnostics of pathology. Recently, several deep learning methods have been proposed to quantify diffusion parameters, but poor generalisation to new data prevents broader use of these methods, as they require retraining of the neural network for each new scan protocol. In this work, we present the dtiRIM, a new deep learning method for Diffusion Tensor Imaging (DTI) based on the Recurrent Inference Machines. Thanks to its ability to learn how to solve inverse problems and to use the diffusion tensor model to promote data consistency, the dtiRIM can generalise to variations in the acquisition settings. This enables a single trained network to produce high quality tensor estimates for a variety of cases. We performed extensive validation of our method using simulation and in vivo data, and compared it to the Iterated Weighted Linear Least Squares (IWLLS), the approach of the state-of-the-art MRTrix3 software, and to an implementation of the Maximum Likelihood Estimator (MLE). Our results show that dtiRIM predictions present low dependency on tissue properties, anatomy and scanning parameters, with results comparable to or better than both IWLLS and MLE. Further, we demonstrate that a single dtiRIM model can be used for a diversity of data sets without significant loss in quality, representing, to our knowledge, the first generalisable deep learning based solver for DTI.

1. Introduction

Diffusion Weighted MRI (DW-MRI) allows us to infer the micro structure of biological tissues by analysing how water diffuses within and between these structures. DW-MRI became an indispensable diagnostic tool for stroke, ischemia and multiple sclerosis, amongst others (Drake-Pérez et al., 2018) and is now part of routine scans in clinical practice.

Although its popularity enabled extensive research to improve and speed up DW-MRI scans, it also led to a large number of data sets with different scan protocols (e.g. echo and repetition times, signal-to-noise ratio, diffusion-weighting strength, and directions). While conventional, statistical methods to estimate diffusion parameters are guided by the signal model (which encodes scan settings), most deep learning (DL) methods are known to generalise poorly to out-of-training-distribution samples (i.e. when the distribution of testing data differs from that of the training data). Consequently, each different dataset requires a new (trained) model.

In Ye (2017); Ye et al. (2019), and Gibbons et al. (2018), deep neural networks were applied to estimate neurite parameters from the NODDI model, with results rivaling current state-of-the-art. However, either specific models had to be trained for different scan protocols, or the training data encoded a large variability of settings, which were identical to settings used to acquire the testing data sets. Further, in Sedlar et al. (2021) and Lin et al. (2019) DL was used to estimate the fiber orientation distribution function, and robustness to variation in the number of samples in q-space was reported, but other scanning parameters of the testing data were kept within the training distribution. The authors of Grussu et al. (2020); Pirkl et al. (2020) and de Almeida Martins et al. (2021) used DL to predict parameters of a joint diffusion-relaxometry model, but, similarly, generalisation to out-of-training-distribution data sets was not reported. Finally, Aliotta et al. (2020) used neural networks to estimate Fractional Anisotropy (FA) and Mean Diffusivity (MD) maps, but, with focus on reducing the number of weighted volumes required for fitting, only robustness to the number of q-space samples was demonstrated.

While these methods showcase important advantages of deep learning over conventional statistical methods, such as higher robustness to the angular resolution of the diffusion gradients and more precise estimates thanks to learned spatial priors, they are still highly restricted by the distribution of the training data. In this work, we present the dtiRIM, a new deep learning framework based on the Recurrent Inference Machines (RIMs) (Putzky and Welling, 2017) for estimation of Diffusion Tensor Imaging (DTI) parameters. RIMs are recurrent neural networks that learn a regularized solution to a model-based inverse problem. This unrolled (Gregor and LeCun, 2010) network is a hybrid method that uses the diffusion signal model to enforce data consistency (via a like-

* Corresponding author.
E-mail address: emanoel.sabidussi@gmail.com (E.R. Sabidussi).

https://doi.org/10.1016/j.neuroimage.2023.119900.
Received 7 July 2022; Received in revised form 19 January 2023; Accepted 21 January 2023
Available online 24 January 2023.

lihood function), in combination with a supervised learning approach, to learn to predict tissue parameters from DW images. Because RIMs learn an iterative optimization function and scanning parameters can be encoded in the signal model, we hypothesized it can generalise well to out-of-training-distribution testing data. Additionally, the strong data generalisation abilities of the RIM permit training with simulation data, consequently allowing for broader distribution of training samples than previous DL methods for DW-MRI.

We performed extensive evaluation of the proposed method in simulation and in vivo experiments with varying gradient schemes, physiology, noise levels, and noise distribution. We compared our method to the state-of-the-art IWLLS method (Veraart et al., 2013) from MRTrix3 (Tournier et al., 2019), and to a Maximum Likelihood Estimator (MLE) method, in terms of the bias and standard deviation of the Fractional Anistropy (FA) and Mean Diffusivity (MD) values. Further, a critical factor to consider when using DL for medical imaging is the correct representation of the underlying data anatomy. Previous research (Bhadra et al., 2021; Kelkar et al., 2021) showed that deep generative models might hallucinate tissue structures due to instability of learning based reconstruction methods (Antun et al., 2020). To assess this aspect, we also evaluated the structural integrity of the estimated maps and underlying fiber structure in simulation and in vivo experiments. The dtiRIM code, pre-trained model, testing datasets, and routines to simulate the training data are available online.[1]

2. dtiRIM – diffusion tensor imaging with recurrent inference machines

The Recurrent Inference Machine (RIM) framework (Putzky and Welling, 2017) is an example of an unrolled network that learns an iterative optimization algorithm to estimate parameters of a model. Akin to conventional iterative gradient-based methods (e.g. gradient descent), the RIM uses the current estimates and the gradients of a likelihood function to predict incremental updates that improve the estimates over time. The likelihood function promotes data consistency and allows the framework to adapt to variations in the acquisition schemes, lowering the risk of anatomical hallucinations and leading to better generalisation to out-of-training-distribution samples.

2.1. Diffusion tensor model

In a DTI experiment, the noiseless magnitude signal in a single voxel is given by the Stejskal–Tanner Equation (Stejskal and Tanner, 1965):

$$S = S_0 e^{-b\boldsymbol{g}^T \boldsymbol{D} \boldsymbol{g}} \tag{1}$$

where the measured signal S depends on the diffusion-encoding gradient vector $\boldsymbol{g}$, of unit length, the diffusion-weighting strength b, the non-diffusion weighted signal S_0 and the symmetric diffusion tensor $\boldsymbol{D}$. Further, we model the acquisition noise following Gudbjartsson and Patz, (1995), where acquired magnitude signals are Rice distributed, given independent Gaussian noise corrupting the real and imaginary terms of the acquired data. Let $\tilde{S} \sim Rice(S;\sigma)$ be a noisy magnitude signal, with σ the standard deviation of the (independent) complex Gaussian noise.

2.2. RIM framework

Let $\theta = [\boldsymbol{S}_0, \boldsymbol{D}_{11}, \boldsymbol{D}_{12}, \boldsymbol{D}_{13}, \boldsymbol{D}_{22}, \boldsymbol{D}_{23}, \boldsymbol{D}_{33}]$ be a vector of parameters containing a non-diffusion weighted image $\boldsymbol{S}_0$ and the unique elements of the diffusion tensor for all voxels of a scan. Additionally, denote $\boldsymbol{\zeta}$ as the collection of gradient vectors $\boldsymbol{g}$ and diffusion-weighting strength b that defines the acquisition scheme used to acquire N images in a DW-MRI experiment. The dtiRIM framework is defined as $\hat{\theta} = H_{\boldsymbol{\gamma}}(\tilde{\boldsymbol{S}}, \boldsymbol{\zeta}, \hat{\sigma})$ (Fig. 1, Algorithm 1), where $\hat{\theta}$ is the estimated parameter vector and

Algorithm 1 dtiRIM framework $H_{\boldsymbol{\gamma}}(\tilde{\boldsymbol{S}}, \boldsymbol{\zeta}, \hat{\sigma})$.

Input: $\tilde{\boldsymbol{S}}, \boldsymbol{\zeta}, \hat{\sigma}, \hat{\boldsymbol{\theta}}_0, \boldsymbol{h}_0$
Rescale input signal: $\check{\boldsymbol{S}} = \tilde{\boldsymbol{S}}/\hat{\sigma}$
for $j = 1$ **to** J **do**
  $\nabla_\theta = \frac{\partial}{\partial \boldsymbol{\theta}} L(\hat{\boldsymbol{\theta}}_{j-1}; \check{\boldsymbol{S}}, \boldsymbol{\zeta})$
  $\{\Delta\boldsymbol{\theta}_j, \boldsymbol{h}_j\} = g_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\theta}}_{j-1}, \nabla_\theta, \boldsymbol{h}_{j-1})$
  $\hat{\boldsymbol{\theta}}_j = \hat{\boldsymbol{\theta}}_{j-1} + \Delta\boldsymbol{\theta}_j$
end for
Rescale non-DW image: $\check{\boldsymbol{S}}_0 = \hat{\sigma}\hat{\boldsymbol{S}}_0$
Return $\check{\boldsymbol{\theta}}$

$H_{\boldsymbol{\gamma}}$ is a model based unrolled Recurrent Neural Network (RNN), with weights $\boldsymbol{\gamma}$, $\tilde{\boldsymbol{S}}$ are noisy images (including $\tilde{\boldsymbol{S}}_0$), and $\hat{\sigma}$ is an estimate of the noise level.

The framework employs the recursive functions

$$\{\Delta\boldsymbol{\theta}_j, \boldsymbol{h}_j\} = g_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\theta}}_{j-1}, \nabla_\theta, \boldsymbol{h}_{j-1}) \tag{2}$$

$$\hat{\boldsymbol{\theta}}_j = \hat{\boldsymbol{\theta}}_{j-1} + \Delta\boldsymbol{\theta}_j \tag{3}$$

to update the parameter estimates at each iteration $j = 1, \ldots, J$. The output of the RNN-cell $g_{\boldsymbol{\gamma}}$ contains the direction and amplitude of the update (which also implicitly encodes the optimization step size) and depends on the current estimates $\hat{\boldsymbol{\theta}}_{j-1}$, the gradients of a negative log-likelihood $\nabla_\theta = \frac{\partial}{\partial \boldsymbol{\theta}} L(\hat{\boldsymbol{\theta}}_{j-1}; \tilde{\boldsymbol{S}}, \boldsymbol{\zeta})$ and a vector of hidden states $\boldsymbol{h}_{j-1}$ it can use to regularize the updates over iterations.

The likelihood $L(\hat{\boldsymbol{\theta}}_{j-1}; \tilde{\boldsymbol{S}}, \boldsymbol{\zeta})$ was set as the sum of the negative log of the Rice PDF over all images and voxels. This summation causes the gradient of the likelihood to depend on the number of images N. Despite this, initial experiments suggested that training time and generalization improved over using the mean of the likelihood across images, which would remove the dependency on N.

2.2.1. Signal scaling

In practice, the range of signal amplitudes of $\tilde{\boldsymbol{S}}$ is largely dependent on the scan protocol, vendor and reconstruction algorithms, and might vary widely. Consequently, because the dtiRIM framework learns the step size implicitly from the data, the amplitude of the updates will be wrong if the signal amplitude is much higher (or lower) than what was observed during training. To avoid this issue, we standardize the input data with $\check{\boldsymbol{S}} = \tilde{\boldsymbol{S}}/\hat{\sigma}$. After inference, the estimate of the non-diffusion weighted image $\hat{\boldsymbol{S}}_0$ is remapped with $\check{\boldsymbol{S}}_0 = \hat{\sigma}\hat{\boldsymbol{S}}_0$ to recover the original amplitude range of the signal. We denote $\check{\boldsymbol{\theta}}$ as the set of parameter estimates that include $\check{\boldsymbol{S}}_0$ (instead of $\hat{\boldsymbol{S}}_0$).

2.2.2. Training loss

The dtiRIM was trained through supervised learning, with loss given by

$$C = \frac{1}{\hat{\sigma}^2 \boldsymbol{J}} \sum_{j=1}^{J} \left\| \boldsymbol{S}^{gt} - \check{\boldsymbol{S}}_j \right\|_2^2, \tag{4}$$

where summations over weighted images and voxels were omitted for simplicity of notation, and j iterates over the dtiRIM inference steps. First, DW images $\check{\boldsymbol{S}}_j$ were computed for every optimization step with Eq. (1) and $\check{\boldsymbol{\theta}}_j$. Subsequently, the squared residual between noiseless images $\boldsymbol{S}^{gt}$ and the (rescaled) estimated DW images was computed. Further, to prevent samples with high amplitude signal from (unfairly) having more importance than samples with low amplitude while training the network, we multiply the loss by $1/\hat{\sigma}^2$. We train the dtiRIM by minimizing C over all training samples.

Note that, unlike $\tilde{\boldsymbol{S}}$, the ground-truth images $\boldsymbol{S}^{gt}$ and predicted images $\check{\boldsymbol{S}}_j$ are not (directly) corrupted by (Rician) noise. Specifically, regardless of any noise-induced bias present in $\tilde{\boldsymbol{S}}$, the dtiRIM should be trained as an unbiased estimator, with $\check{\boldsymbol{S}}_j$ as close as possible to the noise free $\boldsymbol{S}^{gt}$, motivating the choice of loss function in Eq. (4).

[1] https://gitlab.com/radiology/quantitative-mri/emcqmri_dti.

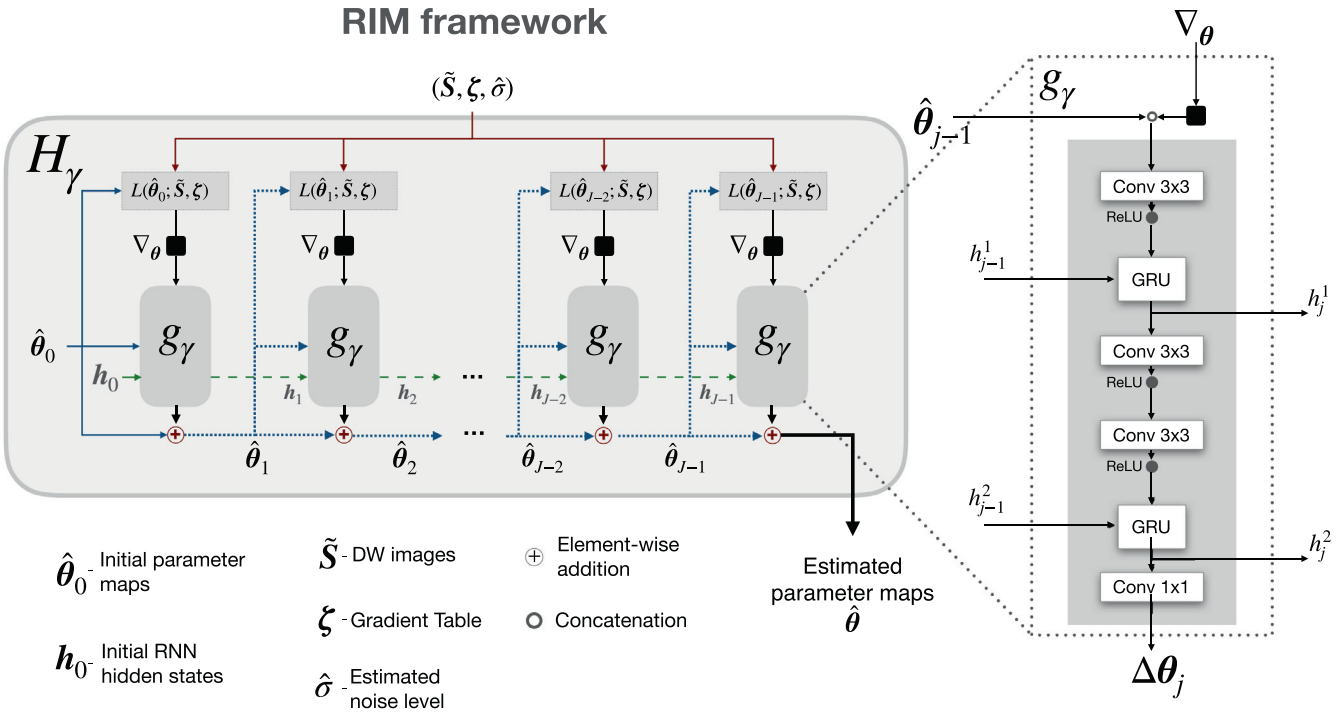


Fig. 1. RIM framework ($\hat{\theta} = H_\gamma(\tilde{S}; \zeta, \hat{\sigma})$). At a given optimization step $j \in \{1, \ldots, J\}$, the RNN-cell g_γ receives as input the current estimate of the signal model parameters, $\hat{\theta}_{j-1}$, the gradient of the negative log-likelihood with respect to θ, ∇_θ, and a vector of memory states $\boldsymbol{h}_{j-1}$ the RIM can use to keep track of optimization progress and perform more efficient updates. The network outputs an update to the current estimate ($\Delta\theta_j$) and the memory state to be used in the next iteration.

3. Materials and methods

3.1. Training dataset

3.1.1. Anatomical structure

The training data was generated with the Numerical Fiber Generator (Close et al., 2009) (NFG). The generator creates numerical fiber structures aiming to simulate the complex spatial configurations present in the human brain. The spatial distribution of the fibers is randomly generated and improved through a sequence of optimization steps, designed to approximate the complexity of real fiber bundles. Further, to simulate a DW signal that includes partial volume effects, the NFG divides each voxel into sub-voxels where a single fiber bundle is assumed to be present. The DW signal is simulated using Eq. (1) and the signal of each sub-voxel is added to produce the noiseless intensity for the voxel.

To reduce the time required to generate the training dataset, and knowing the RIM generalises across anatomies (Sabidussi et al., 2021), we opted to use a single configuration of fibers. We used the NFG to create one 3D sphere (70^3 voxels) containing regions of anisotropic diffusion (fiber bundles), isotropic diffusion (free fluid) and background. To further accelerate the creation of the dataset, we modified the NFG to allow the simulation of 2D images. The configuration files used in this simulation and modified codes are included in the open-source repository.[2]

3.1.2. Generation of training q-spaces

We selected a single gradient table (ζ_{NFG}, $N = 128$ – provided with the NFG) from which all training q-spaces were generated. First, we extracted subsets of sizes ranging from $N = 7$ to $N = 68$ (in steps of 6) from ζ_{NFG}. Following Jones et al., (1999)) we set the number of $b = 0$ images as $N_{b=0} = \lceil N/8 \rceil$, where $\lceil . \rceil$ is the ceiling function. To ensure q-spaces with approximately uniform distribution of directions, we randomly sampled 1000 sets of gradient directions of length N from the original NFG q-space, and selected the 10 (non-identical) sets with the lowest electrostatic repulsion (Cook et al., 2007) per N.

To further increase the dataset variability, we applied random rotations to the selected q-spaces. To construct the random rotation matrices, we generated and normalised 4D vectors with uniformly distributed values, then converted their quaternion representation to orthogonal (rotation) matrices (Kuipers, 2002). Fifteen rotated variations were created per q-space, resulting in a total of 1500 different training q-spaces (10 values of $N \times 10$ q-spaces $\times$ 15 rotations). We specify $b = 1000$ s/mm^2 for all samples and denote the set of training q-spaces by ζ_{Train}.

Table 1
Simulation parameters to create the training dataset. $\mathcal{U}(\min, \max)$ denotes a uniform distribution with samples in between min and max. The diffusion tensor $\boldsymbol{D}$ at each voxel was computed from the FA and MD values and three principal fiber directions. MD_{iso} represents the diffusivity values in isotropic regions.

Dataset	NFG
Non-DW intensity (S_0)	$\sim \mathcal{U}(0.1, 3.0)$ a.u.
Fractional Anisotropy (FA)	$\sim \mathcal{U}(0.05, 1.0)$ a.u.
Mean diffusivity (MD)	$\sim \mathcal{U}(0.05, 1.0)$ μm^2/ms
Isotropic mean diffusivity (MD_{iso})	$\sim \mathcal{U}(1.0, 3.0)$ μm^2/ms

3.1.3. Simulation of DW images

Noiseless 2D DW images $\boldsymbol{S}^{gt}$ were simulated with tissue parameters drawn from the uniform distributions shown in Table 1, q-space $\zeta_{sample} \in \zeta_{Train}$, and Eq. (1). The noisy images $\tilde{\boldsymbol{S}}$ were simulated by corrupting $\boldsymbol{S}^{gt}$ with Rician noise, with σ drawn from a log-uniform distribution with values in the range $[0.002, 1.0]$, corresponding to SNR levels in the range of 50 to 3, respectively (SNR calculated as the ratio between the mean $\boldsymbol{S}_0$ intensity in fiber regions and σ). The training dataset was further augmented by creating 100 independent noise realizations per label, resulting in a total of $100 \times 1500 = 150,000$ training samples.

[2] https://gitlab.com/radiology/quantitative-mri/nfg_1.1.dti-rim.

3.2. RIM architecture and implementation details

The architecture of the RNN-cell (Fig. 1) consisted of one input convolutional layer that receives 14 features (corresponding to the diffusion parameters and their gradients), and produces 120 features, followed by a ReLU activation function and a Gated Recurrent Unit (Cho et al., 2014) (GRU), used to process one of the dtiRIM hidden states. This unit is followed by two hidden convolutional layers, with ReLU activations placed after each, and a second GRU. All hidden layers (including GRUs) have 120 features. The last convolutional layer outputs 7 feature maps, corresponding to $\Delta\boldsymbol{\theta}_j$. Except in the last layer (which has a 1×1 convolution kernel), all convolutions use a 3×3 kernel and are zero-padded to retain the original image size. The number of optimization steps was set to $J = 12$. The hyper-parameters values (including number of steps) were selected so as to minimize the mean training loss over the last 5000 samples. The initial diffusion tensor for all voxels was chosen to be isotropic, with $\hat{\boldsymbol{D}}$ an identity matrix, the initial non-diffusion weighted image $\hat{\boldsymbol{S}}_0$ set as $\check{\boldsymbol{S}}_0$, and $\boldsymbol{h}_0 = \boldsymbol{0}$.

Experiments were performed on an INTEL i7-8650U CPU. The framework was trained on a NVIDIA P100 GPU, using the Adam optimizer with a learning rate of 2.5×10^{-4}. The model was implemented with PyTorch 1.9.0. The network was trained for 48 h, in a single epoch, going through all training samples in mini-batches of 2. The initial weights were set by the Kaiming initialization (He et al., 2015).

3.3. Estimation of noise level $\hat{\sigma}$

The specification of the noise level $\hat{\sigma}$ of the data is an important (and required) step in the dtiRIM framework. In all in vivo experiments, where the true noise level is unknown, we estimated $\hat{\sigma}$ using the technique by St-Jean et al. (2020), where the noise variance in multi-coil experiments is automatically computed via the method of moments. We used the default settings. Note that, for training and in simulation experiments, $\hat{\sigma}$ is known.

3.4. Reference methods

We compared the dtiRIM to two reference methods that are commonly used to estimate diffusion tensors. The first is the Iterated Weighted Linear Least Squares estimator (IWLLS) (Veraart et al., 2013) from MRTrix3 (Tournier et al., 2019), configured with default settings. The second reference method was built using the Maximum Likelihood Estimator framework (MLE), where the diffusion parameters were estimated by minimizing the sum of the negative log of the Rician likelihood. For the (brute force) MLE, we used the ADAM optimizer, with initial learning rate of 0.1, and set the number of iterations to 2000, as it was sufficient for convergence of the loss.

3.5. Experiments

3.5.1. Robustness to anatomy

We simulated testing datasets with varying FA and MD to assess how well the dtiRIM estimates tissues with different properties. First, we evaluated our method in homogeneous regions, where the FA and MD values are the same for all pixels of a 100×100 image. In the **experiment FA**, we varied the fractional anisotropy from 0.05 to 1 in steps of 0.05 while keeping the MD at 0.8 $\mu m^2/ms$. In the **experiment MD**, we varied the mean diffusivity from 0.05 $\mu m^2/ms$ to 1.0 $\mu m^2/ms$ in steps of 0.05 $\mu m^2/ms$, while the FA was kept at 0.8. For both experiments, we set the SNR to 30, the b-value to 1000 s/mm^2, and q-space $\boldsymbol{\zeta} \in \boldsymbol{\zeta}_{Train}$ ($N = 68, N_{b=0} = 9$). We specified $\boldsymbol{S}_0 = \boldsymbol{1}$ and chose the orientation of the main eigenvector in all pixels to be parallel to the x-axis ($\boldsymbol{\lambda} = [1, 0, 0]$).

Then, we performed two simulation experiments in heterogeneous regions, where (small) structures with different anatomy from the background are present. In the **experiment Anatomy**, we varied the FA and MD values of small structures (1×1, 2×2, and 3×3 pixels) placed at the center of an otherwise homogeneous 100×100 pixels region. We simulated the surrounding with constant FA $= 0.5$ and MD $= 0.5$ $\mu m^2/ms$, while simultaneously varying FA and MD for the central structure – FA from 0.05 to 1 in steps of 0.05 and MD from 0.05 $\mu m^2/ms$ to 1.0 $\mu m^2/ms$ in steps of 0.05 $\mu m^2/ms$. We set the SNR, b-value, $\boldsymbol{\zeta}$, $\boldsymbol{S}_0$ and orientation of the main eigenvector as in the previous experiment. The mean predicted FA and MD values in the central pixel are reported, alongside the standard deviation over 100 realizations.

In the **experiment Fiber Orientation**, we estimated the orientation of the main eigenvector of a single pixel in which the ground-truth fiber direction differed from the homogeneous surrounding. We applied the method described in Section 3.1.2 to create 100 random 3D rotations that were used to reorient the eigenvector. We simulated $R = 100$ DW image sets per rotation. We specified the FA and MD of all pixels as 0.8 and 0.8 $\mu m^2/ms$, respectively, and the fibers in surrounding pixels aligned with the x-axis. We kept the other acquisition parameters as in the previous experiments. We report the mean predicted fiber orientation, calculated per rotation, as the orientation of $\hat{\lambda}_\Lambda$, the largest eigenvector of the mean dyadic tensor $\hat{\boldsymbol{\Lambda}}$, with

$$\hat{\boldsymbol{\Lambda}} = \frac{1}{R}\sum_{i=1}^{R} \hat{\lambda}_i \hat{\lambda}_i^T \tag{5}$$

where $\hat{\lambda}_i$ is the estimated largest eigenvector for the ith sample and $\hat{\lambda}_i^T$ its transpose. Additionally, we report the standard deviation of the estimated fiber orientations (per rotation) as

$$\sigma_\lambda = \sqrt{\frac{1}{R}\sum_{i=1}^{R}\left(\arccos\frac{\hat{\lambda}_i \cdot \hat{\lambda}_\Lambda}{|\hat{\lambda}_i||\hat{\lambda}_\Lambda|}\right)^2} \tag{6}$$

where $\hat{\lambda}_i \cdot \hat{\lambda}_\Lambda$ is the dot product and $|.|$ is the vector norm.

3.5.2. Generalisation to acquisition protocols

We performed five experiments with simulation data to evaluate the dtiRIM's robustness to variations in the acquisition settings. In each experiment, a single setting was varied from its default value and 100 DW sets were generated. We report the mean and standard deviation of FA and MD estimates for the central pixel, and compare it to both reference methods. The default $\boldsymbol{S}_0$ signal was set to $\boldsymbol{1}$ for all experiments. In the list of experiments below, the experiment's name refers to the acquisition parameter being varied and the default value refers to the value used in all other experiments:

- **Experiment SNR**: Signal-to-noise ratio varied from 4 to 30 in steps of 2. The input parameter $\hat{\sigma}$ was adjusted to match the noise level of the sample; Default SNR: 30;
- **Experiment $\boldsymbol{N}$**: The number of DW images varied from 6 to 126, in steps of 6, with 1 $b = 0$ image for every 8 DW images; Default N: 68;
- **Experiment b-value**: The diffusion-weighting strength varied from 400 s/mm^2 to 1200 s/mm^2 in steps of 100 s/mm^2; Default b: 1000 s/mm^2;
- **Experiment $\boldsymbol{\zeta}$**: four datasets generated with q-spaces not used during training, denoted by $\boldsymbol{\zeta}_{t1}$ ($N = 36$), $\boldsymbol{\zeta}_{t2}$ ($N = 69$), $\boldsymbol{\zeta}_{t3}$ ($N = 71$), and $\boldsymbol{\zeta}_{t4}$ ($N = 23$); Default q-space: $\boldsymbol{\zeta}_s$ ($N = 68$), which was one of the q-spaces used during training;
- **Experiment $\boldsymbol{\alpha}_\zeta$**: To evaluate the dependence of the estimated fiber direction on the orientation of the gradients, we applied $R = 100$ random rotations (created as in Section 3,1.2) to the default q-space to simulate experiments with rotated gradient directions. The ground-truth fiber orientation was kept constant, with $\lambda^{gt} = [1, 0, 0]$. We compute the average error in the predicted orientation as the angle between $\hat{\lambda}_\Lambda$ and λ^{gt}, and the standard deviation with Eq. (6).

3.5.3. Mischaracterization of the data distribution

We performed two experiments with simulation data to evaluate the quality of estimates when the data distribution is misspecified, or mischaracterized, leading to *out-of-training-distribution* likelihood values.

- **Experiment** $\hat{\sigma}$: As a key input to the dtiRIM, $\hat{\sigma}$ influences the amplitude of the likelihood gradients, which, as consequence, might affect the amplitude of the updates produced by the dtiRIM. In this experiment, we evaluated the robustness of each method to misspecification of the estimated noise level $\hat{\sigma}$. We set $\hat{\sigma} = \{\frac{1}{40}, \frac{1}{38}, \frac{1}{36}, \ldots, \frac{1}{20}\}$ (i.e. assumed SNR between 40 and 20), and used the default dataset from the previous experiment, with SNR = 30 ($\sigma = \frac{1}{30}$), $N = 68$, $b = 1000$ s/mm^2, and $\boldsymbol{\zeta}_s$.
- **Experiment** nc-χ: In this experiment we evaluated the robustness of the dtiRIM when the distribution of the input signal departs from the Rice PDF, which might occur when acquisitions are performed in multi-coil systems and images reconstructed with GRAPPA, or data pre-processing is applied. We simulated the noiseless testing data as in **Experiment SNR** and corrupted the images with a non-central χ distributed noise, as proposed by Aja-Fernández et al. (2010): We assumed that acquisitions were fully-sampled and that the composite magnitude image can be obtained with the root-mean-sum-of-squares (SoS) method using multiple coil images. These images were created by multiplying each DW image by a normalised set of 8 coil sensitivity maps (available from unrelated phantom experiments performed in-house), followed by adding complex value noise with standard deviation σ. We set $\sigma = \{\frac{1}{4}, \frac{1}{6}, \frac{1}{8}, \ldots, \frac{1}{30}\}$ and measured the bias and standard deviation of the FA and MD estimates over 100 realisations per σ value. The input parameter $\hat{\sigma}$ was chosen to match the noise level of the sample.

3.5.4. *In vivo experiments*

Repeatability In vivo data was used to evaluate the repeatability of our method with real scans and anatomy. We selected one set of DW images from one subject from the publicly available dataset AOMIC (Snoek et al., 2021). This DW set contains 66 images (including 6 $b = 0$) and was acquired with a q-space not used in training, with $b = 1000$ s/mm^2. The scans were performed with a 32-channel head coil at a 3 T scanner and images were reconstructed using SENSE. The noise level $\hat{\sigma}$ was estimated at 20.76, with SNR = 28 measured as the ratio between the average $\boldsymbol{S}_0$ signal (in tissue regions) and $\hat{\sigma}$.

Although this dataset includes repeated scans, quantification of pixel-level error is not possible since the inter-experiment DWI sets are not aligned. Image registration is possible, but interpolation and small registration errors would still affect the estimates. Therefore, we chose to emulate repeated scans with sub-sampled q-spaces from this set. Similar as described in Section 3.1.2, to extract sub-sampled q-spaces (ζ_N, with $N \in \{13, 19, 25, 31\}$) we randomly sampled 200 sets of gradient directions of length N from the fully-sampled q-space, and selected the 20 non-identical sets with the lowest electrostatic repulsion per N.

First, we present a visual comparison of the FA and MD estimates of each method for the fully-sampled dataset ($N = 66$). Then, we present the Root Mean Square Error (RMSE) maps over the 20 repetitions in each sub-sampled dataset, with ground-truth set independently per method as the FA and MD predictions from the fully-sampled dataset. Finally, we show maps with the tensor ellipsoid representation, color encoded by direction, for a single sample of each dataset.

White matter hyperintensities A second in vivo dataset was used to assess the estimates when tissue abnormalities, such as white matter hyperintensities (WMH), are present. We used a DW set from one subject of the Rotterdam Study dataset (Ikram et al., 2011), containing 28 images, including 3 $b = 0$, and acquired with a q-space not used in training and $b = 1000$ s/mm^2. The scan was performed with a 8-channel head coil, at 1.5 T, and images were reconstructed using SENSE. Prior to reconstruction, the scans, originally acquired on a grid of 64×96 voxels, were upsampled to 256×256 voxels by zero-padding the k-space. Further, the images were pre-processed to compensate patient motion, to correct Eddy current distortions, and to reduce Gibbs ringing. The noise level $\hat{\sigma}$ was estimated at 19.82, with SNR = 33. We present a visual comparison of the FA and MD maps of each method.

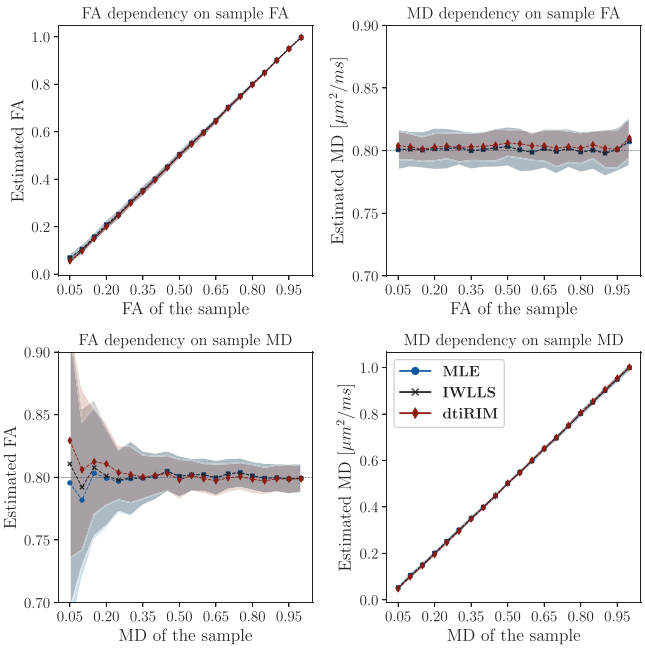


Fig. 2. Top row: Experiment FA. Estimates of Fractional Anistropy (FA) and Mean Diffusivity (MD) as a function of the FA value of the sample. Bottom row: **Experiment MD**. Estimates of FA and MD as a function of the MD value of the sample. Solid lines represent the mean predicted value over 100 samples. Shaded region represents ± the standard deviation around the mean. Ground-truth values are indicated by the black dotted line.

4. Results

4.1. *Robustness to anatomy*

The top row of Fig. 2 shows the results of **experiment FA**. The dtiRIM estimates of FA were similar to both MLE and IWLLS for the entire range. The MD predictions from dtiRIM showed a small bias compared to the ground-truth, but lower standard deviation than the other methods for all MD values. The bottom row of Fig. 2 shows the results of **experiment MD**. The dtiRIM showed lower standard deviation for FA predictions in the entire MD range, but elevated FA bias for samples with low MD. The dtiRIM estimates of MD had similar bias to and lower standard deviation than the references for all ground-truth MD values.

Figure 3 presents the results for **experiment Anatomy**. We observe that all methods overestimated FA values when the FA and MD of the central pixel are low. For structures of size 1×1 and 2×2 pixels, the dtiRIM overestimated MD when the central pixel has FA and MD lower than its surrounding, and underestimated it when FA and MD are higher, indicating blur. The bias of dtiRIM decreased for the 3×3 pixels structure, approximating the baseline. The estimates from the MLE and IWLLS were similar.

Figure 4 presents the results for **experiment Fiber Orientation**. The dtiRIM estimates had an average bias of 0.25° (over all orientations and realisations) and average standard deviation of 1.04°, compared to 0.13° bias and 1.07° standard deviation from the MLE, and 0.12° and 1.06° from IWLLS, respectively. Note that, for one orientation, the MLE produced outliers that increased the standard deviation of the predictions. We see no other dependencies on the fiber direction.

4.2. *Generalisation to acquisition protocols*

Experiment SNR Figure 5 shows the FA and MD estimates for various values of noise level (SNR). For SNR > 8, all methods predicted FA and MD with comparable bias and standard deviation. For noisier samples, the dtiRIM had lower standard deviation than both references.

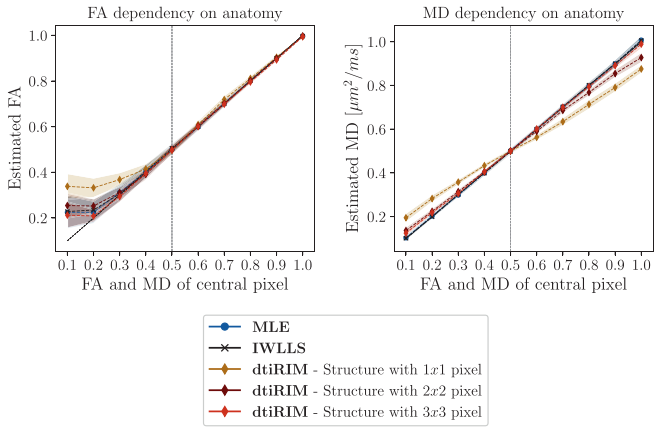


Fig. 3. Experiment Anatomy. Estimates of FA and MD as a function of FA and MD values of the central structure. Three structure sizes were evaluated: 1×1, 2×2, and 3×3 pixels. The FA and MD values of surrounding pixels are indicated by the dotted vertical line. Solid lines represent the mean predicted value over 100 samples. Shaded region represents ± the standard deviation around the mean. Ground-truth values are indicated by the black dotted line.

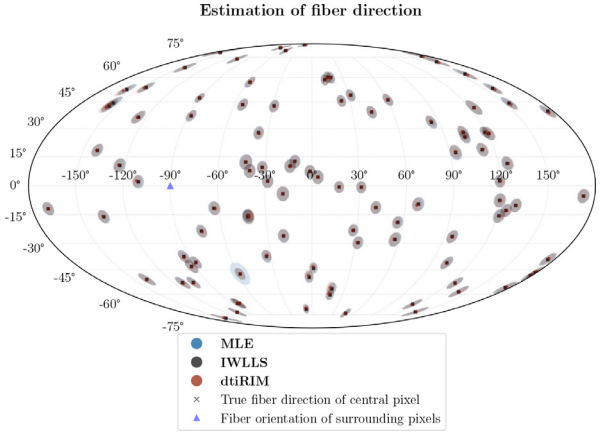


Fig. 4. Experiment Fiber Orientation. Estimated orientation of the main eigenvector in the central pixel of the sample. Orientations shown in the Mollweide projection. The blue **x** indicates the ground-truth orientation of the eigenvector. The radius of the shaded areas shows 3 times the standard deviation over 100 repeated experiments per orientation, with center given by the mean fiber orientation. The fibers in the background were aligned with the x axis (blue triangle) in all experiments. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

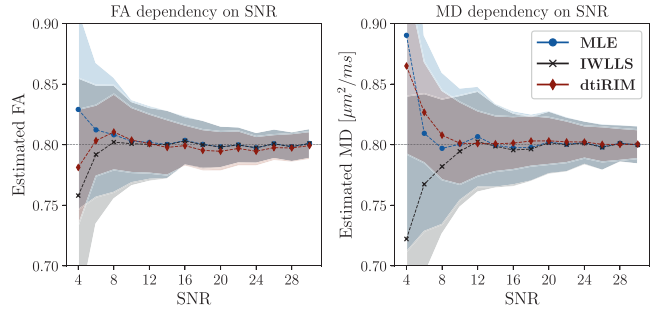


Fig. 5. Experiment SNR. Estimates of FA and MD as a function of the SNR of the sample. Solid lines represent the mean predicted value over 100 samples. Shaded region represents ± the standard deviation around the mean. Ground-truth values are indicated by the black dotted line.

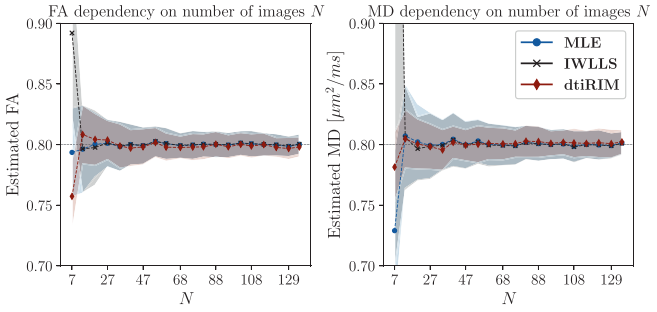


Fig. 6. Experiment ***N***. Estimates of FA and MD as a function of the number of Diffusion Weighted Images. Solid lines represent the mean predicted value over 100 samples. Shaded region represents ± the standard deviation around the mean. Ground-truth values are indicated by the black dotted line.

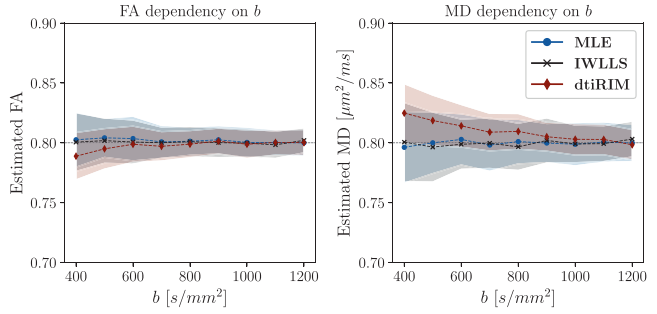


Fig. 7. Experiment ***b***. Estimates of FA and MD as a function of the diffusion-weighting strength. Solid lines represent the mean predicted value over 100 samples. Shaded region represents ± the standard deviation around the mean. Ground-truth values are indicated by the black dotted line.

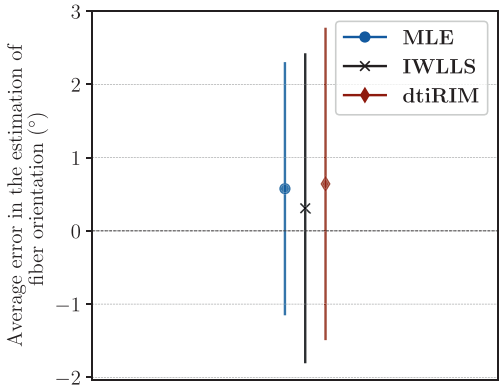


Fig. 8. Experiment α_ζ. Crosses represent the mean predicted value over 100 random rotations applied to the acquisition gradients. Vertical lines extend ± the standard deviation around the mean.

Experiment ***N*** Figure 6 shows the FA and MD estimates for datasets created with varying number of samples in q-space. All methods predicted FA and MD values with similar bias and standard deviation for most N. For $N = 7$, both dtiRIM and MLE produced estimates with lower error than IWLLS.

Experiment ***b*** Figure 7 shows the FA and MD estimates for datasets created with different values of diffusion-weighting strength b. The dtiRIM had lower standard deviation than the reference methods for most of the b-values, for both FA and MD estimates, but the bias depended slightly on the b-value (bias < 0.012 for FA and < 0.025 $\mu m^2/ms$ for MD on $b = 400$ s/mm^2). Neither MLE's or IWLLS's estimates were dependent on the diffusion-weighting strength.

Experiment α_ζ Figure 8 presents the bias and standard deviation of the estimated fiber orientation across 100 random rotations applied to

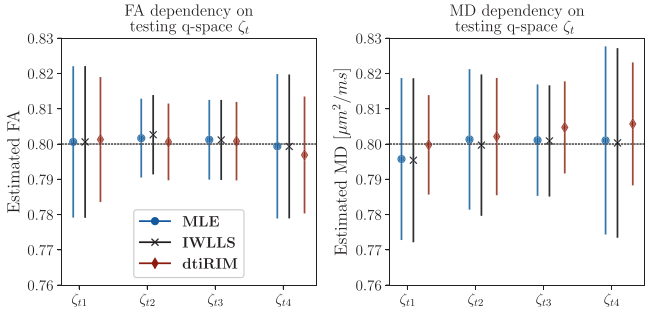


Fig. 9. Experiment ζ. Estimates of FA and MD for datasets created with q-spaces not used during training of the dtiRIM. Crosses represent the mean predicted value over 100 samples. Vertical lines extend ± the standard deviation around the mean. Ground-truth values are indicated by the black dotted line.

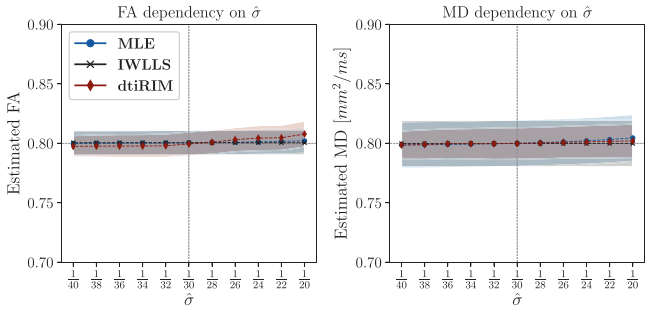


Fig. 10. Experiment $\hat{\sigma}$. Estimates of FA and MD as a function of the input noise level $\hat{\sigma}$. Solid lines represent the mean predicted value over 100 samples. Shaded region represents ± the standard deviation around the mean. Ground-truth values are indicated by the horizontal black dotted line, and the correct $\hat{\sigma}$ value for this dataset ($\hat{\sigma} = 1/30$) is indicated by the vertical black dotted line.

the simulation q-space. The dtiRIM had bias of 0.64° and standard deviation of 2.13°, the MLE had 0.58° of bias and 1.73° of standard deviation, while IWLLS's were 0.31° and 2.12°, respectively.

Experiment ζ Figure 9 shows the FA and MD estimates for 4 datasets simulated with gradient directions not used during training. The dtiRIM produced estimates with low bias (< 0.003 for FA, < 0.006 μm^2/ms for MD), comparable to the reference results. The dtiRIM standard deviation was lower than both references for all datasets in this experiment.

4.3. Mischaracterization of the data distribution

Experiment $\hat{\sigma}$ Figure 10 shows the results of the experiment that evaluated the robustness of each method to misspecification of the noise level. We observe that predictions from both dtiRIM and MLE were dependent on the specified noise level, notably when $\hat{\sigma}$ is assumed to be larger than the real noise level of the sample. For FA, the dtiRIM estimates were susceptible to misspecification of $\hat{\sigma}$, but the bias was smaller than 1% for a $\hat{\sigma}$ 50% larger than the correct value. For MD, the dtiRIM estimates were more accurate than MLE's for the entire range. Note that the IWLLS doesn't depend on the estimate of the noise level, hence the reported mean and standard deviation of the predictions are the same for all $\hat{\sigma}$. The dtiRIM had lower standard deviation than both references.

Experiment nc-χ Figure 11 shows the FA and MD estimates for a dataset corrupted by non-central χ distributed noise. All methods produced biased FA and MD estimates in the entire σ range. For FA, the dtiRIM predictions had lower bias and standard deviation than the references, independently of the noise level. For MD, the dtiRIM had larger bias than references for $\sigma > \frac{1}{12}$, but comparable estimates for lower noise levels. The predictions of all methods tended to the ground-truth FA/MD values for low noise cases.

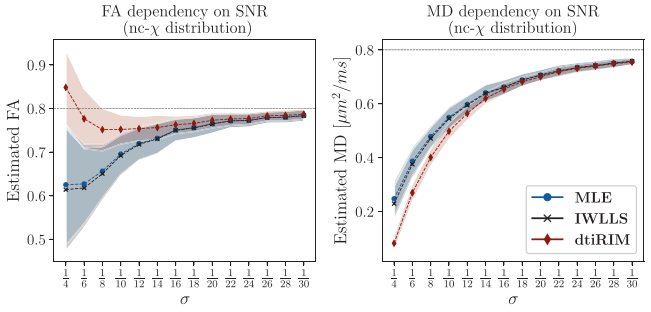


Fig. 11. Experiment nc-χ. Estimates of FA and MD as a function of the SNR for non-central χ distributed data. Solid lines represent the mean predicted value over 100 samples. Shaded region represents ± the standard deviation around the mean. Ground-truth values are indicated by the black dotted line.

4.4. In vivo *experiments*

4.4.1. *Repeatability*

Figure 12 presents the predicted FA and MD maps for the fully-sampled ($N = 66$ directions) *repeatability* dataset. We observe that the dtiRIM's estimate contained less outliers than both MLE and IWLLS predictions, while retaining similar anatomy and FA values for all structures. Similar behavior was observed in the MD estimates, where the dtiRIM prediction was less noisy than the references, with visible differences around the ventricles.

Figure 13 shows the results of the FA and MD repeatability experiment. The first three rows show the RMSE maps of the FA predictions for all methods and sub-sampled datasets ζ_{31}, ζ_{25}, ζ_{19}, and ζ_{13}. The dtiRIM presented lower RMSE compared to both references for all cases, notably for dataset ζ_{13}. The three bottom rows show the RMSE maps of the MD predictions for the same datasets. The dtiRIM estimates had lower RMSE for all cases, most notably within the ventricles and for dataset ζ_{13}.

Figure 14 shows the tensor ellipsoid representation of one DW set of each sub-sampled dataset, color encoded by direction, overlaid on the predicted FA map. For clarity, only a portion of each map is shown. The direction of the tensors were largely in agreement between methods for dataset ζ_{31}, with less outliers in the dtiRIM estimates. For datasets with lower number of images, it is clear that dtiRIM produced more consistent predictions than the reference methods.

4.4.2. *White matter hyperintensities*

Figure 15 shows the FA and MD maps of each method for the dataset containing WMH regions, alongside a FLAIR scan of the same slice. We observe that all methods predicted maps with similar FA and MD values and anatomical structure, with hyperintense areas visible in all predictions.

5. Discussion

Variations in the dataset (number of weighted images, gradient directions, b-value, etc.) hinder the broader use of deep learning in DW-MRI, since current neural networks for estimation of diffusion parameters do not generalise well to out-of-training-distribution samples, and new network models have to be trained if the testing dataset changes significantly from the training data.

In this work we presented the dtiRIM, a new method to estimate diffusion tensor parameters from DW images. We demonstrated that the diffusion tensor estimates from dtiRIM have low dependency on the physiology, anatomy and acquisition settings (e.g. direction and strength of diffusion gradients), with similar bias to and lower standard deviation than a Maximum Likelihood Estimator (MLE), and IWLLS (MRTrix3), a widely used method for DTI. Our results indicate that a single dtiRIM model can predict the diffusion parameters from various

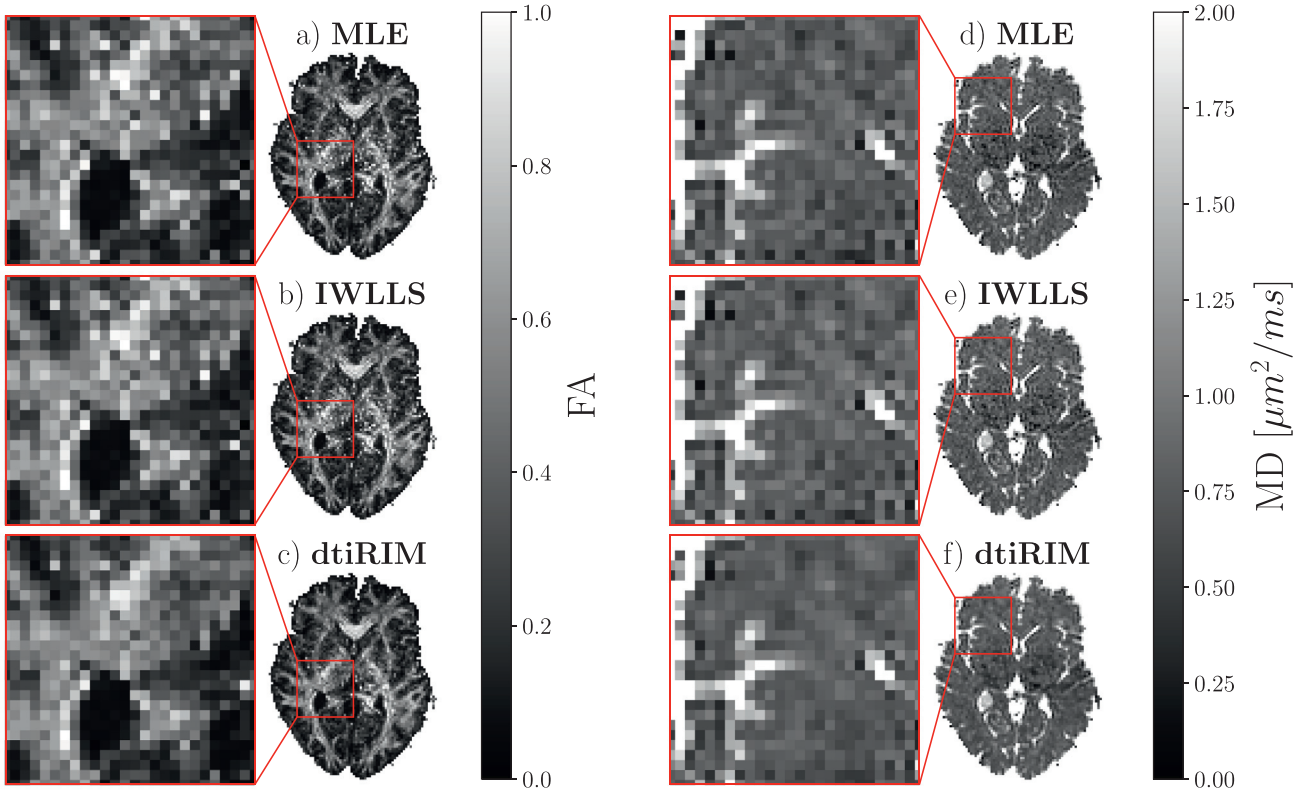


Fig. 12. In vivo experiment. Scan from the AOMIC dataset. The SNR was measured at 26 and the dataset contained 66 images ($N = 66$). Figures *a*), *b*) and *c*) show the FA maps produced with the MLE, IWLLS and dtiRIM, respectively. A zoomed detail is shown in the red box. Figures *d*), *e*) and *f*) show the MD maps from each method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

datasets, without need to retrain the network. To our knowledge, the dtiRIM is the first generalisable deep learning based estimator for DTI.

We showed in simulation experiments that the dtiRIM is robust to variations in the anatomy and physiology. Our method was able to infer a wide range of FA and MD values normally seen in brain tissues. In homogeneous regions (that is, regions with uniform FA and MD), the dtiRIM had lower standard deviation than and comparable bias to other methods, suggesting an effective spatial regularization of the estimates. Moreover, the test data had radically different structures than the training data, indicating that dtiRIM generalises well to unseen anatomies.

Experiments with non-homogeneous images (when one or more pixels of the testing image has parameters different than the surrounding) showed that the proposed method introduces blur to very small structures (1 or 2 pixels wide) in MD maps. This effect, however, largely disappeared for bigger structures. Further, the blurring of small structures did not seem to affect the estimates of the fiber orientation, whose bias and standard deviation were comparable to both the MLE and IWLLS. Additionally, we observed no dependence on the true orientation of the fiber, as the dtiRIM correctly predicted the fiber orientation for a range of directions, in structures as small as 1 pixel. These results indicate that the dtiRIM is robust to anatomical variations in the data.

We also demonstrated that the dtiRIM has low dependency on acquisition settings. We evaluated our method with variations in the SNR, number of diffusion-weighted images, the strength of the diffusion gradients, and random rotations of the q-space. Our results evidenced that the FA and MD estimates from dtiRIM had low standard deviation and low bias, independently of the dataset used during testing. Additionally, as most of the scenarios evaluated fell outside the training distribution (e.g. only $b = 1000\,\mathrm{s/mm^2}$ was used during training), we showed that the proposed network also generalises well to out-of-training-distribution data, and can, therefore, be applied to datasets with different characteristics than the training data. Further, results with rotated gradients indicated that the dtiRIM is insensitive to the orientation of the diffusion gradients. Note, however, that the dataset used in this experiment had a high number of images ($N = 68$) and results might differ for low N due to the reduced angular resolution of the gradient set.

Further, we showed that the misspecification of the data distribution (i.e. when the noise distribution of the testing data does not follow that of the training data) has limited impact on the dtiRIM performance. Our framework is robust to over/underestimation of the noise level, with bias lower than 1% for all cases studied. Thus indicating that the margin of error for noise level estimates is sufficiently broad for practical use, and that the dtiRIM generalises to variations in amplitude of the likelihood gradients. In addition, we demonstrated that the dtiRIM estimates are comparable to the MLE's and IWLLS's even when the data does not follow a Rice distribution, producing estimates comparable to the reference methods. The large bias present on all methods is, in part, caused by the substantially higher noise floor of the non-central χ distribution compared to the Rice PDF. This demonstrates the importance of assuming the correct noise distribution, and how estimates of widely used DTI methods can be affected by the image reconstruction algorithm. In contrast to the SENSE reconstruction, in which weighted images remain Rice distributed (Aja-Fernández et al., 2014), GRAPPA alters the noise distribution of the data. In these cases, the modification of the likelihood function within the MLE or dtiRIM methods (e.g. using the non-central χ PDF) is required to reduce estimation bias. We don't recommend the use of IWLLS with GRAPPA-reconstructed data.

Finally, we validated our method with two in vivo experiments. We demonstrated that the dtiRIM generalises to real data, with higher quality estimates than MLE and IWLLS in repeated experiments. Thanks to efficient spatial regularization, the dtiRIM estimates had less outliers

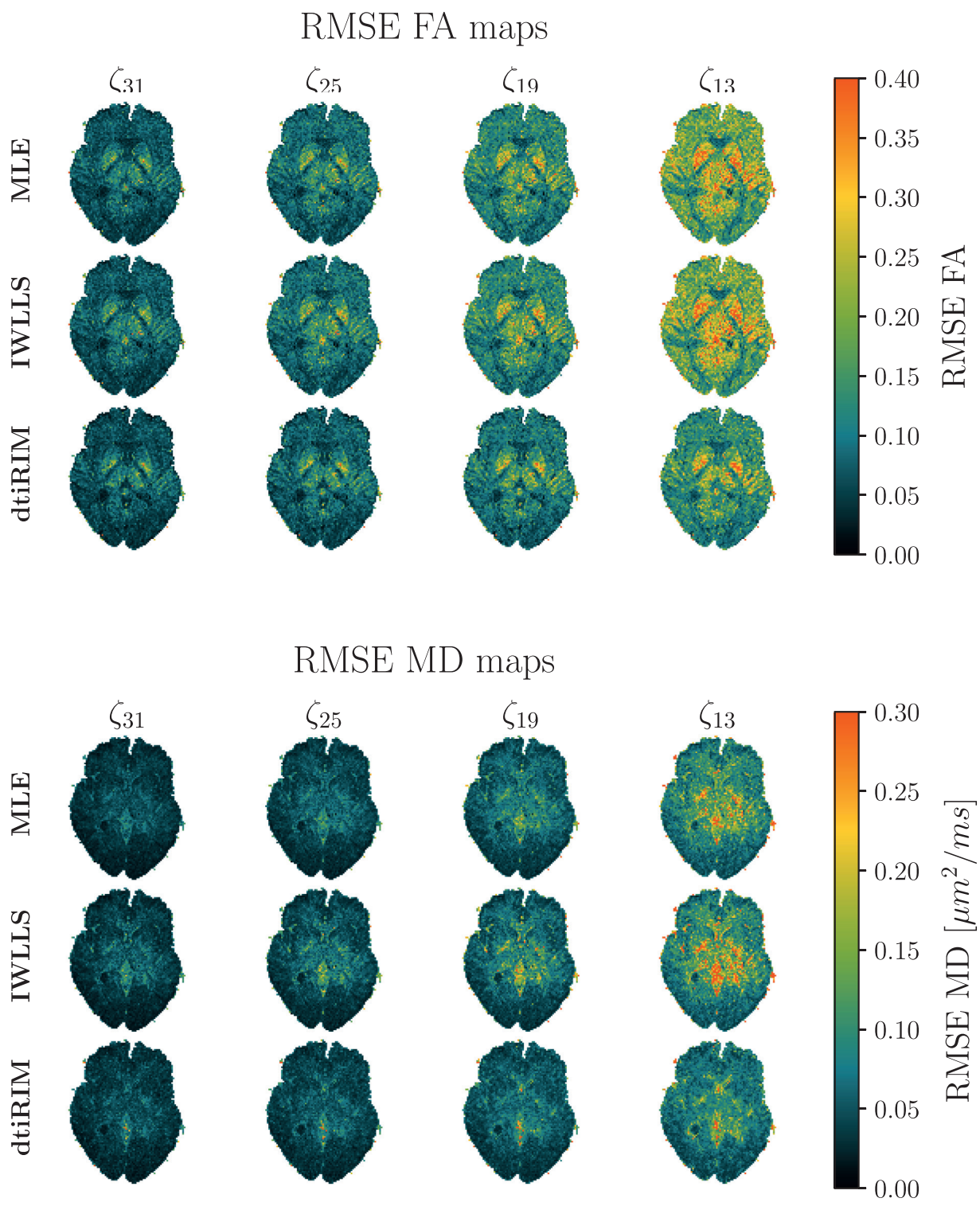


Fig. 13. Repeatability experiment with the in vivo dataset AOMIC. Repeated experiments were emulated by sub-sampling the gradient directions of the original data. Ground-truth was set as the FA or MD map computed by each method with the original DWI set ($N = 66$). The columns indicate the sub-sampled data sets ζ_{31} ($N = 31$), ζ_{25} ($N = 25$), ζ_{19} ($N = 19$), and ζ_{13} ($N = 13$). The root-mean-squared-error (RMSE) of FA maps of each method are shown in rows 1 to 3; the RMSE MD maps are shown in rows 4 to 6.

and were more precise, especially for data sets with low number of images. Moreover, we showed that the underlying fiber structure predicted by the dtiRIM was consistent with the references, and was more robust to low number of gradient directions. These results further suggest that the dtiRIM generalises well to unseen structures and acquisition settings. Furthermore, the dtiRIM correctly predicted the FA and MD values in abnormal tissues, with estimates similar to both MLE and IWLLS in regions with white matter hyperintensities. Notice that a number of pre-processing steps were applied to this dataset, such as upsampling and correction of Eddy current distortions. It is likely, therefore, that the noise distribution was altered, and the noise became spatially correlated. Even with this deviation from the assumptions during training,

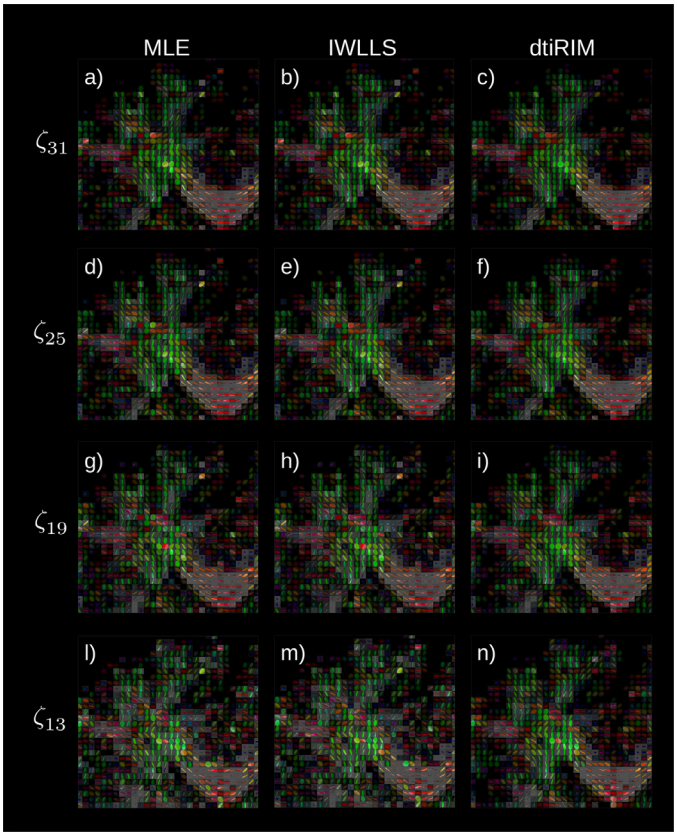


Fig. 14. Representation of the tensor ellipsoid of all voxels color encoded by direction. A single sample per dataset is shown. The ellipsoid shape is given by the eigenvectors and eigenvalues of the estimated tensor at each voxel. The patch is located in the anterior region of the brain, with the corpus callosum shown at the bottom right of each image. Rows indicate the sub-sampled data sets ζ_{31} ($N = 31$), ζ_{25} ($N = 25$), ζ_{19} ($N = 19$), and ζ_{13} ($N = 13$).

similar FA and MD predictions were obtained across methods, further attesting to the dtiRIM's robustness.

Throughout this paper, we demonstrated the robustness and consistency of the dtiRIM estimates in several experiments, with simulation and in vivo data sets. Our deep learning method is able to learn a generalisable model that only needs to be trained once, with simulated data. Despite the predictions of the dtiRIM being similar to both MLE and IWLLS, our method consistently produces parameter maps with lower standard deviation, with up to 15% improvement in some experiments, thanks to the efficient priors learned by the dtiRIM. This aspect is particularly interesting for future research in DWI-MRI, where fitting of more complex diffusion models (Alexander et al., 2017) could benefit from learned constrains and learned priors.

6. Conclusion

We proposed the dtiRIM, a new method for estimating the diffusion tensors from diffusion-weighted images. We demonstrated that our method has similar bias to and lower standard deviation than state-of-the-art DTI methods, independently of the physiology, anatomy, and acquisition schemes. With our approach, a single dtiRIM model is able to predict the diffusion tensors for datasets acquired with diverse settings, making it the first generalisable deep learning method for DTI.

Data and code availability statements

- The core code for the method presented in this paper is publicly available under the Apache 2.0 license, available at https://gitlab.com/radiology/quantitative-mri/emcqmri_dti.
- A modified version of the code of the Numerical Fiber Generator (Close et al., 2009) was used to generate training data. The modified version is made available in the open repository https://gitlab.com/radiology/quantitative-mri/nfg_1.1.dti-rim distributed under the GPL license.
- Data from the AOMIC: Amsterdam Open MRI Collection, dataset AOMIC-ID1000, version 1.2.1 (Snoek et al., 2021) is used in this paper. The full dataset and documentation can be downloaded from doi:10.18112/openneuro.ds003097.v1.2.1.

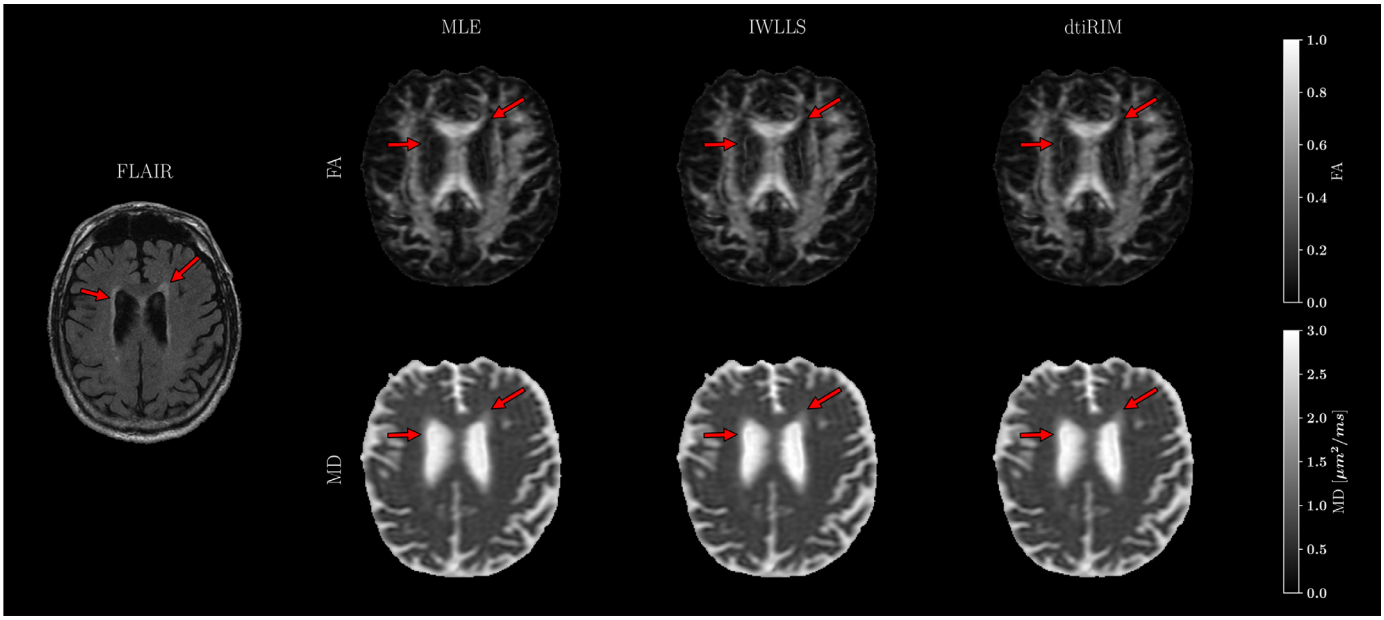


Fig. 15. FA and MD estimates using in vivo dataset with white matter hyperintensities (WMH). The left-most figure presents the FLAIR scan, with WMH indicated by the red arrows. On the right, the top row shows the FA maps for each method. Bottom row shows the MD maps. The red arrows indicate the same anatomical position in all figures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Data availability

Data will be made available on request.

Credit authorship contribution statement

E.R. Sabidussi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. **S. Klein:** Conceptualization, Methodology, Supervision, Validation, Writing – review & editing. **B. Jeurissen:** Conceptualization, Methodology, Validation, Writing – review & editing. **D.H.J. Poot:** Conceptualization, Investigation, Methodology, Supervision, Validation, Writing – review & editing.

Acknowledgments

We would like to thank Mr. Riwaj Byanju for providing the coil sensitivity maps, and Dr. Bo Li, Dr. Esther Bron, and the Rotterdam Study team for providing access to the in vivo testing data. This work is part of the project B-Q MINDED which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 764513. BJ is supported by FWO Vlaanderen, grant number G090020N and by the Belgian Science Policy Prodex (Grant ISLRA 2009–1062). BJ is a member of the μNEURO Research center of Excellence of the University of Antwerp.

References

Aja-Fernández, S., Tristán-Vega, A., Hoge, W.S., 2010. Statistical noise analysis in GRAPPA using a parametrized noncentral chi approximation model. Magn. Reson. Med. 65 (4), 1195–1206. doi:10.1002/mrm.22701.

Aja-Fernández, S., Vegas-Sánchez-Ferrero, G., Tristán-Vega, A., 2014. Noise estimation in parallel MRI: GRAPPA and SENSE. Magn. Reson. Imaging 32 (3), 281–290. doi:10.1016/j.mri.2013.12.001.

Alexander, D.C., Dyrby, T.B., Nilsson, M., Zhang, H., 2017. Imaging brain microstructure with diffusion MRI: practicality and applications. NMR Biomed. 32 (4). doi:10.1002/nbm.3841.

Aliotta, E., Nourzadeh, H., Patel, S.H., 2020. Extracting diffusion tensor fractional anisotropy and mean diffusivity from 3-direction DWI scans using deep learning. Magn. Reson. Med. 85 (2), 845–854. doi:10.1002/mrm.28470.

de Almeida Martins, J.P., Nilsson, M., Lampinen, B., Palombo, M., While, P.T., Westin, C.-F., Szczepankiewicz, F., 2021. Neural networks for parameter estimation in microstructural MRI: application to a diffusion-relaxation model of white matter. NeuroImage 244, 118601. doi:10.1016/j.neuroimage.2021.118601.

Antun, V., Renna, F., Poon, C., Adcock, B., Hansen, A.C., 2020. On instabilities of deep learning in image reconstruction and the potential costs of AI. Proc. Natl. Acad. Sci. USA 117 (48), 30088–30095. doi:10.1073/pnas.1907377117.

Bhadra, S., Kelkar, V. A., Brooks, F. J., Anastasio, M. A., 2021. On hallucinations in tomographic image reconstruction. arXiv:2012.00646

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNNencoder-decoder for statistical machine translation. doi:10.48550/ARXIV.1406.1078.

Close, T.G., Tournier, J.-D., Calamante, F., Johnston, L.A., Mareels, I., Connelly, A., 2009. A software tool to generate simulated white matter structures for the assessment of fibre-tracking algorithms. NeuroImage 47 (4), 1288–1300. doi:10.1016/j.neuroimage.2009.03.077.

Cook, P.A., Symms, M., Boulby, P.A., Alexander, D.C., 2007. Optimal acquisition orders of diffusion-weighted MRI measurements. J. Magn. Reson. Imaging 25 (5), 1051–1058. doi:10.1002/jmri.20905 .

Drake-Pérez, M., Boto, J., Fitsiori, A., Lovblad, K., Vargas, M.I., 2018. Clinical applications of diffusion weighted imaging in neuroradiology. Insights Imaging 9 (4), 535–547. doi:10.1007/s13244-018-0624-3.

Gibbons, E.K., Hodgson, K.K., Chaudhari, A.S., Richards, L.G., Majersik, J.J., Adluru, G., DiBella, E.V.R., 2018. Simultaneous NODDI and GFA parameter map generation from subsampled q-space imaging using deep learning. Magn. Reson. Med. 81 (4), 2399–2411. doi:10.1002/mrm.27568.

Gregor, K., LeCun, Y., 2010. Learning fast approximations of sparse coding. In: Proceedings of the 27th International Conference on International Conference on Machine Learning. Omnipress, Madison, WI, USA, pp. 399–406. doi:10.5555/3104322.3104374.

Grussu, F., Battiston, M., Palombo, M., Schneider, T., Gandini Wheeler-Kingshott, C. A. M., Alexander, D. C., 2020. Deep learning model fitting for diffusion-relaxometry: a comparative study. bioRxiv. doi:10.1101/2020.10.20.347625.

Gudbjartsson, H., Patz, S., 1995. The Rician distribution of noisy MRI data. Magn. Reson. Med. 34 (6), 910–914. doi:10.1002/mrm.1910340618.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. CoRR abs/1502.01852

Ikram, M.A., van der Lugt, A., Niessen, W.J., Krestin, G.P., Koudstaal, P.J., Hofman, A., Breteler, M.M.B., Vernooij, M.W., 2011. The rotterdam scan study: design and update up to 2012. Eur. J. Epidemiol. 26 (10), 811–824.

Jones, D.K., Horsfield, M.A., Simmons, A., 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magn. Reson. Med. 42 (3), 515–525. 10.1.1.462.9303

Kelkar, V.A., Bhadra, S., Anastasio, M.A., 2021. Compressible latent-space invertible networks for generative model-constrained image reconstruction. IEEE Trans. Comput. Imaging 7, 209–223. doi:10.1109/tci.2021.3049648.

Kuipers, J.B., 2002. Quaternions and Rotation Sequences. Princeton University Press, Princeton, NJ.

Lin, Z., Gong, T., Wang, K., Li, Z., He, H., Tong, Q., Yu, F., Zhong, J., 2019. Fast learning of fiber orientation distribution function for MR tractography using convolutional neural network. Med. Phys. 46 (7), 3101–3116. doi:10.1002/mp.13555.

Pirkl, C. M., Gmez, P. A., Lipp, I., Buonincontri, G., Molina-Romero, M., Sekuboyina, A., Waldmannstetter, D., Dannenberg, J., Endt, S., Merola, A., Whittaker, J. R., Tomassini, V., Tosetti, M., Jones, D. K., Menze, B. H., Menzel, M. I., 2020. Deep learning-based parameter mapping for joint relaxation and diffusion tensor mr fingerprinting. arXiv:2005.02020

Putzky, P., Welling, M., 2017. Recurrent inference machines for solving inverse problems. arXiv:1706.04008

Sabidussi, E.R., Klein, S., Caan, M.W.A., Bazrafkan, S., den Dekker, A.J., Sijbers, J., Niessen, W.J., Poot, D.H.J., 2021. Recurrent inference machines as inverse problem solvers for MR relaxometry. Med. Image Anal. 74, 102220. doi:10.1016/j.media.2021.102220.

Sedlar, S., Papadopoulo, T., Deriche, R., Deslauriers-Gauthier, S., 2021. Diffusion MRI fiber orientation distribution function estimation using voxel-wise spherical U-Net. In: Computational Diffusion MRI. Springer International Publishing, pp. 95–106. doi:10.1007/978-3-030-73018-5-8.

Snoek, L., van der Miesen, M., van der Leij, A., Beemsterboer, T., Eigenhuis, A., Scholte, S., 2021. "AOMIC-ID1000 dataset". 10.18112/openneuro.ds003097.v1.2.1

St-Jean, S., De Luca, A., Tax, C.M., Viergever, M.A., Leemans, A., 2020. Automated characterization of noise distributions in diffusion mri data. Med. Image Anal. 65, 101758. doi:10.1016/j.media.2020.101758.

Stejskal, E.O., Tanner, J.E., 1965. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. J. Chem. Phys. 42 (1), 288–292. doi:10.1063/1.1695690.

Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H., Connelly, A., 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.

Veraart, J., Sijbers, J., Sunaert, S., Leemans, A., Jeurissen, B., 2013. Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls. NeuroImage 81, 335–346. doi:10.1016/j.neuroimage.2013.05.028.

Ye, C., 2017. Tissue microstructure estimation using a deep network inspired by a dictionary-based framework. Med. Image Anal. 42, 288–299. doi:10.1016/j.media.2017.09.001.

Ye, C., Li, X., Chen, J., 2019. A deep network for tissue microstructure estimation using modified LSTM units. Med. Image Anal. 55, 49–64. doi:10.1016/j.media.2019.04.006.