



# Geometric deep learning for diffusion MRI signal reconstruction with continuous samplings (DISCUS)

Christian Ewert<sup>a,\*</sup>, David Kügler<sup>a,\*</sup>, Rüdiger Stirnberg<sup>a</sup>, Alexandra Koch<sup>a</sup>, Anastasia Yendiki<sup>b,c</sup>, Martin Reuter<sup>a,b,c</sup>

<sup>a</sup>German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

<sup>b</sup>A. A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, MA, United States

<sup>c</sup>Department of Radiology, Harvard Medical School, Boston, MA, United States

\*C.E. and D.K. are joint first authors

Corresponding Author: Martin Reuter ([martin.reuter@dzne.de](mailto:martin.reuter@dzne.de))

## ABSTRACT

Diffusion-weighted magnetic resonance imaging (dMRI) permits a detailed in-vivo analysis of neuroanatomical microstructure, invaluable for clinical and population studies. However, many measurements with different diffusion-encoding directions and possibly  $b$ -values are necessary to infer the underlying tissue microstructure within different imaging voxels accurately. Two challenges particularly limit the utility of dMRI: *long acquisition times* limit feasible scans to only a few directional measurements, and the *heterogeneity of acquisition schemes across studies* makes it difficult to combine datasets. Left unaddressed by previous learning-based methods that only accept dMRI data adhering to the specific acquisition scheme used for training, there is a need for methods that accept and predict signals for arbitrary diffusion encodings. Addressing these challenges, we describe the first geometric deep learning method for *continuous* dMRI signal reconstruction for arbitrary diffusion sampling schemes for both the input and output. Our method combines the reconstruction accuracy and robustness of previous learning-based methods with the flexibility of model-based methods, for example, spherical harmonics or SHORE. We demonstrate that our method outperforms model-based methods and performs on par with *discrete* learning-based methods on single-, multi-shell, and grid-based diffusion MRI datasets. Relevant for dMRI-derived analyses, we show that our reconstruction translates to higher-quality estimates of frequently used microstructure models compared to other reconstruction methods, enabling high-quality analyses even from very short dMRI acquisitions.

**Keywords:** deep learning, geometric deep learning, denoising, diffusion MRI, signal reconstruction,  $q$ -space

## 1. INTRODUCTION

Magnetic Resonance Imaging (MRI) has long been a key technology for the in-vivo study of the human brain in health and disease. Diffusion-weighted MRI (dMRI), in particular, provides “invaluable information” (Tuch, 2004) on tissue microstructure and macro-anatomy. For example, dMRI acquisitions enabled the study of neuroinflammatory response and structural breakdown in Alzheimer’s disease (Fick et al., 2017) as well as structural remodeling

after stroke (Galazzo et al., 2018). Beyond the study of local microstructure, high-quality diffusion acquisitions yield reconstructions of the brain’s neural pathways facilitating the study of brain architecture and connectivity.

However, constraints related to acquisition time and signal-to-noise ratio impose trade-offs in image-space and  $q$ -space (the space representing the direction and magnitude of diffusion weighting). Furthermore, different approaches to analyzing dMRI data (Wedeen et al., 2005;

Received: 25 May 2023 Revision: 9 January 2024 Accepted: 30 January 2024 Available Online: 13 March 2024



© 2024 Massachusetts Institute of Technology.  
Published under a Creative Commons Attribution 4.0  
International (CC BY 4.0) license.

Imaging Neuroscience, Volume 2, 2024  
[https://doi.org/10.1162/imag\\_a\\_00121](https://doi.org/10.1162/imag_a_00121)

H. Zhang et al., 2012) recommend different acquisition schemes. As a result, different studies have different acquisition protocols (single-, multi-shell, or grid-based sampling), where  $q$ -space is sampled at different sets of directions and  $b$ -values. In this paper, we focus on  $q$ -space super-resolution, that is, from a limited series of *obtained measurements* (**observations**), we predict the *unknown signals* for additional diffusion vectors in continuous  $q$ -space (**query vectors**). We present a method allowing for flexible input from and flexible query for any  $q$ -space acquisition scheme. This method reliably infers additional signals and is a prerequisite for data harmonization, for example, by resampling datasets acquired with different  $q$ -space samplings onto an arbitrary common sampling scheme, addressing both of the challenges mentioned above.

Learning-based methods have shown impressive success in  $q$ -space super-resolution and, despite only utilizing relatively simple neural networks (Gibbons et al., 2019; Golkov et al., 2016; Li et al., 2021; Lin et al., 2019; Nath et al., 2019, 2020; Ye, 2017), have outperformed prior model-based methods (Descoteaux et al., 2007; Özarslan et al., 2009, 2013). Generally, we distinguish between discrete and continuous reconstruction models. *Output-continuous* methods permit the flexible signal reconstruction for arbitrary  $q$ -vectors whereas *output-discrete* methods are restricted to an a priori fixed choice of  $q$ -vectors. Similarly, *input-continuous* methods can accept any  $q$ -space acquisition scheme, whereas *input-discrete* methods are inherently tied to a fixed acquisition scheme and do not apply to other acquisition schemes. The denoising and super-resolution capabilities of learning-based models are proven across medical imaging settings (Gondara, 2016). However, so far, proposed learning-based methods for dMRI signal reconstruction are from one of four categories: 1. both input- and output-discrete (Gibbons et al., 2019; Golkov et al., 2016; Li et al., 2021), 2. input-discrete (with Graph-CNN) (Chen et al., 2020, 2022), 3. achieve continuity through fitting a continuous model representation (Lin et al., 2019; Nath et al., 2019, 2020), and 4. are only output-continuous (Ren et al., 2021). In contrast to learning-based methods, model-based methods rely on assumptions about the physics of the dMRI signal and were proposed for both discrete (Paquette et al., 2015) and continuous signal reconstruction (Descoteaux et al., 2007; Fick et al., 2016; Özarslan et al., 2009, 2013), but can lack robustness when only few, noise-afflicted measurements (from a short acquisition) are available.

*Diffusion MRI Signal Reconstruction with Continuous Samplings* (DISCUS) combines the advantages of the continuous reconstruction of model-based methods with the improved prediction and denoising capabilities of

deep learning models. Instead of relying on a generative model of the signal, or decomposing the signal in terms of continuous basis functions, we interpret the reconstruction task as the continuous completion of a finite point cloud and learn to perform this task in an entirely data-driven fashion building on geometric deep learning methods for point clouds. Our contributions are:

- We provide the first learning-based *input-* and *output-continuous* method for continuous reconstruction of  $q$ -space signals that can be initialized with any (sufficiently dense)  $q$ -space sampling scheme.
- We demonstrate that our method outperforms model-based methods and performs on par with discrete learning-based methods on single-, multi-shell, and grid-based dMRI datasets while adding sampling flexibility.
- We show that our reconstructed signals yield better estimates of signal-derived microstructure models than those reconstructed with other methods, for example, for *Diffusion Tensor Imaging* (DTI), *Diffusion Kurtosis Imaging* (DKI), *Neurite Orientation Dispersion and Density Imaging* (NODDI), and *Mean Apparent Propagator MRI* (MAP-MRI).

## 2. PROBLEM STATEMENT AND RELATED WORK

Diffusion-weighted MRI reconstruction aims to infer additional signals from a set of acquired measurements. For clarity, we use the following definitions and concepts: A *dMRI acquisition* consists of a paired series of *diffusion vectors* and *diffusion-weighted images* (DWI). Each DWI is a 3D MR image and the diffusion vector is an a priori chosen parameter vector (updated in acquisition) determining the direction and weighting of the diffusion measurement. This vector is also referred to as the  $q$ -vector drawn from  $q$ -space  $\subset \mathbb{R}^3$ . At a fixed spatial position, the dMRI acquisition is a paired series  $(q_i, s_i)$  of  $q$ -vectors  $q_i$  and scalar diffusion signals  $s_i$ . In the reconstruction context, we refer to this paired series as the *observation set*  $x_{obs}$ . Based on the observation set, we aim to predict the missing signal  $\hat{s}$  corresponding to the *query vector*  $q$ , that is, obtain  $\hat{s} = \mathcal{M}(q | x_{obs})$  with the reconstruction model  $\mathcal{M}$ . Generally,  $q$  and  $\hat{s}$  may also be lists of arbitrary size.

The  $q$ -space super-resolution problem, that is, the prediction of signals for additional  $q$ -vectors, stands in contrast to image super-resolution (Albay et al., 2018; Chatterjee et al., 2021), where additional signals for new spatial locations are predicted. As previously introduced, we differentiate between discrete (Chen et al., 2020, 2022; Gibbons et al., 2019; Golkov et al., 2016; Li et al., 2021; Paquette et al., 2015) and continuous (Cheng et al., 2013; Descoteaux et al., 2007; Fick et al., 2016; Jha et al.,

2022; Lin et al., 2019; Mani et al., 2022; Michailovich et al., 2011; Nath et al., 2020, 2019; Özarslan et al., 2009, 2013; Rathi et al., 2014; Sedlar et al., 2021; Ye, 2017)  $q$ -space super-resolution methods. *Input-discrete* methods can only be applied to the specific acquisition scheme of the training data. Similarly, *output-discrete* methods can only predict signals for specific  $q$ -vectors used during training. Because of this, the data from multiple studies cannot be mixed in training to improve generalization and offset the cost of training data acquisition for input- or output-discrete methods. In fact, input-discrete models trained on data with a particular sampling scheme cannot be applied to datasets acquired with different sampling schemes. On the other hand, *input-continuous* models are applicable to any acquisition scheme and *output-continuous* models permit the flexible signal prediction for arbitrary  $q$ -vectors given the information of the observation set.

Reconstruction methods follow either a *learning-* or a *model-based* paradigm. Learning-based methods require training on a collection of data across voxels while model-based methods rely on a diffusion signal model fit to the observation set of the particular voxel. Model-based methods can lack robustness when measurements of only a few  $q$ -space points are available. In contrast, learning-based methods have proved their denoising capabilities in various settings (Fadnavis et al., 2020; Gondara, 2016; Koonjoo et al., 2021). An example of an entirely discrete, model-based method is compressed sensing for diffusion spectrum imaging (DSI) (Paquette et al., 2015), which uses the Fourier relationship between  $q$ -space and the propagator space for reconstruction. Entirely continuous (input- and output-continuous) model-based methods decompose the signal in terms of continuous basis functions such as spherical harmonics or spherical ridgelets for spherical signals (Descoteaux et al., 2007; Michailovich et al., 2011) and SHORE, MAP-MRI, spherical ridgelets with radial decay, and other approaches for 3D signals (Cheng et al., 2013; Özarslan et al., 2009, 2013; Rathi et al., 2014).

Entirely discrete, learning-based methods predict diffusion signals or derived quantities based on the observation set for individual voxels, for example, FA and MD from DTI or NDI and ODI from NODDI. They employ linear (Varadarajan & Haldar, 2018) or non-linear models, that is, multi-layer perceptrons (Golkov et al., 2016; Grussu et al., 2020), or predict whole images using convolutional neural networks (CNN) (Gibbons et al., 2019; Li et al., 2021).

So far, almost all proposed learning-based methods achieve entirely continuous reconstruction only indirectly by the workaround of incorporating established models, thus inheriting their limitations. Instead of learning reconstruction end-to-end, these methods first decompose the

signal in terms of basis functions like spherical harmonics (Lin et al., 2019; Jha et al., 2022; Nath et al., 2020; Mani et al., 2022) or SHORE functions (Nath et al., 2019; Ye, 2017), and train inherently discrete networks to predict coefficients of basis functions in the target domain (for signal reconstruction) or diffusion signal-based microstructure indices from coefficient vectors in the source domain. These methods are limited by a loss of information during the conversion to basis coefficients in the source and target domains. Furthermore, an optimal order truncating the basis of these models may depend on the tissue type, and assumptions about the signal properties may yield sub-optimal results if only a few noisy measurements are available, for example, the exponential signal decay in the SHORE model. In addition, learned coefficients may only be optimal for the number of observation and query vectors used during training as no work to date has trained a method with observation and query sets containing a wide range of different numbers of vectors. Graph-convolutional approaches interpret the  $q$ -space sampling as a graph (Chen et al., 2020, 2022), but have thus far only been applied to a priori chosen acquisition schemes and not addressed the challenges imposed by arbitrary acquisition schemes. Ren et al. (2021) proposed an output-continuous method to reconstruct DWIs in *image space* using a CNN from other MRI contrasts, that is, a stack of non-diffusion-weighted, T1-, and T2-weighted images, whereas we reconstruct DWIs in *voxel space* from other DWIs.

To our knowledge, no learning-based, input- and output-continuous method offering full flexibility for reconstruction has been proposed to date. While some methods incorporate larger spatial contexts to make predictions for a particular voxel, that is, incorporating adjacent voxel or even image slice information, model-based methods such as SHORE and spherical harmonics take  $q$ -space data only from the voxel at hand into account. To provide a fair comparison with these methods, we use the same strategy, that is, we do not incorporate neighborhood information of any kind.

### 3. METHOD

Our *DISCUS* method has the flexibility of learning from and predicting signals for any  $q$ -space sampling scheme. We first introduce a list of aims for our reconstruction model, then explain the architecture and augmentation scheme, followed by the loss function used for training.

First, we pose the following six aims for our method:

- 1a) *Flexibility to input any number of observations* and
- 1b) *Invariance to the choice of observation vectors*: dMRI acquisitions often define unique  $q$ -space sampling schemes containing from single-digit to several hundred

measurements. 2) *Invariance to the ordering of observation vectors*: while dMRI measurements are obtained sequentially, the ordering is arbitrary and has no impact on reconstruction. 3) *Invariance to the sign of observation and query vectors*: dMRI signals are antipodally symmetric, that is,  $s(q) = s(-q)$ . 4) *Invariance to rotations of the  $q$ -space basis*: Since the  $q$ -space signal is reconstructed independently at each voxel (see Section 2), reconstruction relies only on the relative position of observation vectors to each other and the query vectors, not a specific choice of reference frame. 5) *Invariance to T2-shinethrough*: Each diffusion measurement has a T2-weighted and a diffusion-weighted signal component. As both components are independent per spatial voxel, the T2-weighted component has no impact on the diffusion signal reconstruction.

We address these aims with specific design choices in the architecture (requirements 1a, 2, and 3), with data augmentation (requirements 1b and 4), and preprocessing of the data (requirement 5). To isolate the diffusion signal from the T2-weighted component, that is, the signal at the origin of  $q$ -space, we simply normalize the diffusion-weighted signal by the mean  $q \equiv 0$  image, since the signal solely consists of the T2-weighted component there.

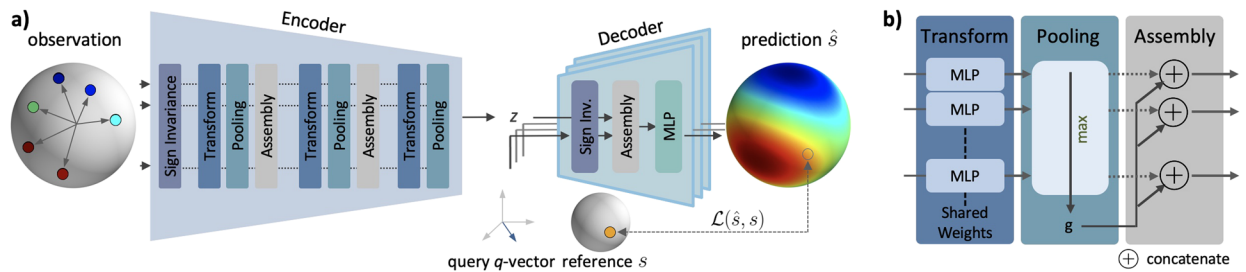
### 3.1. Architecture

Our architecture features an encoder and a decoder (see Fig. 1a). Per voxel, the encoder  $E: \mathbb{R}^{4 \times N} \rightarrow \mathbb{R}^L$  receives the observation set  $x_{obs}$ , that is,  $N$  diffusion measurements ( $q$ -vectors and acquired signals), as input and outputs a latent vector  $z = E(x_{obs})$  summarizing the information across measurements. The purpose of the

encoder is to obtain a representation characterizing the diffusion profile that is independent of the order, number, and specific choice of measurements. The decoder  $D: \mathbb{R}^{3 \times M} \times \mathbb{R}^L \rightarrow \mathbb{R}^M$  predicts the signals  $\hat{s} = D(q|z)$  for the query vectors  $q$  given the latent vector  $z$ . The combination of encoder and decoder yields our reconstruction model  $\hat{s} = D(q|z) = D(q|E(x_{obs})) = \mathcal{M}(q|x_{obs})$ . Here, we provide an overview of the components of our method as well as their properties. A detailed schematic of the DISCUS architecture can be found in Appendix Figure A.

#### 3.1.1. Encoder

The encoder’s task is to extract the relevant information for reconstruction from the  $N$  input measurements while keeping the number  $N$  flexible (aim 1a) and to implement aims 2 and 3. The  $N$  observation vectors are processed along  $N$  processing paths—one path is dynamically allocated per input observation vector. Each module works along these paths or across these paths. Along-the-paths modules apply a single function independently to each path. In contrast, the across-the-paths modules merge information between these paths (Qi et al., 2017). The observation vectors are processed by the sign-invariance module followed by a sequence of alternating transform and pooling modules with interleaved assembly modules (see Fig. 1a, “Encoder”). The sign-invariance module operates along the paths and obtains sign-invariant features by applying a single MLP independently to each  $q$ -vector implementing a mapping  $\mathbb{R}^3 \rightarrow \mathbb{R}^{128}$ . The corresponding signals bypass the sign-invariance module and are concatenated with the sign-invariant features. The transform modules (Fig. 1b, left) also operate along the paths applying a single module-specific MLP independently



**Fig. 1.** DISCUS architecture: Our method is designed to learn a continuous signal from a finite list of diffusion measurements end-to-end. (a) The DISCUS architecture consists of an encoder and a decoder. The encoder input is the paired series of  $q$ -vectors and observed signals  $s$  (the observation set  $x_{obs} = [q, s]$ ). It summarizes the information into a latent representation  $z$  that is independent of the order and the specific measurements. The decoder predicts signals for arbitrary query  $q$ -vectors given the latent vector  $z$  from the encoder. The encoder and decoder start with the extraction of features invariant to the sign of the input vectors. The discrepancy between predicted signals  $\hat{s}$  for query  $q$ -vectors and (withheld) corresponding reference signals  $s$  is employed as loss  $\mathcal{L}(\hat{s}, s)$  for training. (b) The interplay of the transform and pooling modules implements permutation invariance. The transform module transforms features individually while the pooling module aggregates information across features. Between pooling and transform modules, the assembly module concatenates features for further processing.



to each input. For example, the first transform module implements a mapping  $\mathbb{R}^{128+1} \rightarrow \mathbb{R}^{128}$  which is applied to each of the  $N$  sign-invariant features and signals. In contrast, the pooling module (Fig. 1b, middle) operates across the paths, has no learnable weights, and aggregates information via across-the-paths max-pooling, for example,  $\mathbb{R}^{N \times 128} \rightarrow \mathbb{R}^{128}$  for the first pooling module. The assembly module (Fig. 1b, right) concatenates along-the-paths features from the transform module with the across-the-paths features of the pooling module independently per path, for example, implementing the mapping  $\mathbb{R}^{128} \times \mathbb{R}^{128} \rightarrow \mathbb{R}^{256}$  in the first assembly module. Because each module accepts an arbitrary number of inputs and each pooling module’s “global features” are invariant to permutations of the paths, these properties extend to the entire encoder. The encoder has a single output: the latent vector  $z \in \mathbb{R}^L$  representing a summary of the  $N$  measurements that served as input.

### 3.1.2. Decoder

The decoder’s task is to predict signals for the  $M$  query vectors given the latent summary  $z$  of the encoder while keeping the number  $M$  flexible (aim 1b) and to implement aim 3. Like the along-the-paths processing in the encoder, the decoder dynamically allocates one processing path per query vector and independently predicts the signal for that vector (see Fig. 1a, “Decoder”). More specifically, for each query vector and independently along each path, the decoder first obtains sign-invariant features with the sign-invariance module implementing a mapping  $\mathbb{R}^3 \rightarrow \mathbb{R}^{128}$ . Then, the assembly module concatenates the sign-invariant features per path with the latent vector  $z$ :  $\mathbb{R}^{128} \times \mathbb{R}^L \rightarrow \mathbb{R}^{128+L}$  per path. In a final step and independently per path, the decoder’s MLP predicts the signal from the concatenated tensor:  $\mathbb{R}^{128+L} \rightarrow \mathbb{R}$ .

### 3.1.3. Sign invariance module (SI)

To implement antipodal symmetry, we introduce a sign-invariance module at the start of both the encoder and decoder for the observation and the query vector, respectively. With  $SI(q) = F(q) + F(-q)$  guaranteeing the invariance to the sign of  $q$  generally, our sign-invariance module implements a multi-layer perceptron (MLP) for  $F$  via  $SI(q) = \text{MLP}_{SI}(q) + \text{MLP}_{SI}(-q) = SI(-q)$  as suggested by Lim et al. (2023). In practice, we share the weights of the  $\text{MLP}_{SI}$  across all instances in the encoder and decoder.

### 3.1.4. Details

Each of the MLP layers in the transform modules is composed of a fully connected linear layer, a batch-

normalization (Ioffe & Szegedy, 2015), and a hyperbolic tangent activation. The MLP layers in the sign-invariance module and the decoder consist of a fully connected linear layer and a ReLU activation. The last layer of the decoder is only a linear layer. The MLPs in the sign-invariance module contain three hidden and one output layer whereas the remaining MLPs contain two hidden and one output layer. With two exceptions, all layers have a dimension of 128. The size of the output layer of the final transform module in the encoder, that is, the length of the latent vector, is determined via ablation in Section 5. The output layer of the decoder MLP has dimension 1, yielding one scalar signal prediction per query vector.

## 3.2. Augmentations

To achieve invariance with respect to rotations in  $q$ -space, we sample random rotations uniformly (Shoemaker, 1992) and apply them jointly to observation and query vectors during training. To learn the reconstruction from an arbitrary choice of observation vectors, we input a random choice of at least five observation vectors. To implement this “per-voxel sub-sampling” in fixed-shaped tensors, we input all observation vectors jointly with a binary mask indicating which vectors should be used for reconstruction (see also Appendix Fig. A). This mask is applied within each pooling operation in the encoder. At inference time, we input only the sub-sampled observation set. Both augmentations are applied per training sample providing a diversity of  $q$ -space sampling and rotations within each batch.

Since we have normalized the diffusion signal, we can assume that the signal at  $q \equiv 0$  is 1. Therefore, we also query  $q \equiv 0$  during training and obtain a loss for the discrepancy between prediction and target value equal to 1.

## 3.3. Loss function

To guide network training, we use the mean squared error between the signal predictions  $\hat{s}$  and their corresponding reference signals  $s$ . Additionally, we partition signals depending on whether they are part of the observation set (denoted  $\mathcal{O}$ ) or not (denoted  $\mathcal{Q} \setminus \mathcal{O}$ ) and weigh their relative error contributions with the parameter  $\lambda$ , which we determine in the ablation section.

$$\mathcal{L} = (1 - \lambda) \sum_{s \in \mathcal{O}} (\hat{s} - s)^2 + \lambda \sum_{s \in \mathcal{Q} \setminus \mathcal{O}} (\hat{s} - s)^2$$

## 4. EXPERIMENTS

This section details the datasets, the training, and the reference methods for the method comparison in the

Results section. Finally, we introduce the quality metrics for the diffusion signals and signal-derived measures evaluations.

## 4.1. Datasets

### 4.1.1. Human Connectome project (HCP)

This dataset features 270 DWIs acquired at diffusion weightings ( $b$ -values) of 1,000, 2,000, and 3,000 s/mm<sup>2</sup> with 90 DWIs per  $b$ -value and participant on a modified 3 Tesla Siemens MAGNETOM Skyra scanner (Feinberg et al., 2010; Moeller et al., 2010; Setsompop et al., 2012; Sotiropoulos et al., 2013; van Essen et al., 2012; Xu et al., 2012). Further parameters were: 1.25 mm isotropic spatial resolution, slice multiband factor 3, TR = 5,520 ms, TE = 89.5 ms,  $\Delta$  = 43.1 ms, and  $\delta$  = 10.6 ms. We build gender-balanced training-, validation-, and test-sets featuring data from 8, 2, and 20 participants, respectively. From this dataset, we form a multi-shell dataset and a single-shell dataset (featuring signals on the  $b = 1,000$  s/mm<sup>2</sup> shell).

### 4.1.2. Rhineland Study (RLS)

We utilize data with a grid sampling acquired in the Rhineland Study (Breteler et al., 2014; St  cker, 2016). This acquisition features DWIs for 112  $q$ -vectors sparsely sampled (Tobisch et al., 2018) from an 11×11×11 Cartesian grid in  $q$ -space with  $b$ -values in the range of 270 to 6,800 s/mm<sup>2</sup>. All images were obtained on a 3 Tesla Siemens MAGNETOM Prisma scanner at an isotropic spatial resolution of 1.5 mm, a multiband factor of 3, and with scan parameters TR = 5,500 ms, TE = 105 ms,  $\Delta$  = 51.3 ms, and  $\delta$  = 20.1 ms. We build gender- and age-balanced training-, validation-, and test-sets featuring data from 16, 2, and 20 participants.

### 4.1.3. High-SNR dataset

For one healthy young male participant and RLS parameters, we acquired DWIs for the RLS  $q$ -space sampling and additional low-noise DWIs (20 times repeated measurement of 20  $q$ -vectors, repeated measurements are averaged). Six  $b$ -vectors uniformly cover the shell at 1,000 s/mm<sup>2</sup>, while 15 vectors are contained in the Rhineland Study protocol and have been chosen to uniformly cover the  $b$ -value range up to 6,800 s/mm<sup>2</sup>. These images were acquired on a 3 Tesla Siemens MAGNETOM Prisma scanner with approval from the ethics committee of the Medical Faculty, University of Bonn, Germany, and with written, informed consent of the participant.

### 4.1.4. Dataset processing

While the HCP dataset is distributed as preprocessed data, we account for off-resonance fields, eddy currents, and head motion in the Rhineland Study data using *FSL topup* (Andersson et al., 2003; Smith et al., 2004) and *eddy* (Andersson & Sotiropoulos, 2016). Since *eddy* expects the  $q$ -vectors to be located on shells, we approximate each  $q$ -vector by its nearest neighbor on one of 20 shells as outlined by Tobisch et al. (2018). Furthermore, to address T2-shinethrough, we normalize DWIs by dividing them with the mean  $q \equiv 0$  image of the respective participant. We constrain the scope of training and evaluation to white and grey matter voxels based on tissue segmentations obtained with *FSL FAST* (Smith et al., 2004; Y. Zhang et al., 2001). Cerebrospinal fluid, air, and voxels with a mean  $q \equiv 0$  intensity of zero are excluded.

### 4.1.5. Observation and query sets for evaluation

For each dataset, we first obtain one set of 30 query vectors and then multiple sets of observation vectors, which do not overlap with the query set. For a meaningful evaluation, we require each observation vector set and the query vector set to cover the respective domain (spheres or 3D space) as uniformly as possible. We ensure the best uniform coverage for the query set by drawing it first from all available vectors. From the vectors not contained in the query set, we draw a nested list of observation sets containing  $N=10, 20, 30, 50, (80, 160)$  vectors starting with the largest set. For the shell-based datasets, we obtain these sets via *DMRITool* (Cheng et al., 2018) yielding sets with uniform radial (constrained to shells) and angular coverage (also across shells). Aiming to maintain a consistent angular and radial distribution of vectors for the grid-based dataset, we employ the following approach: First, we group vectors by  $b$ -value range (using bins of size 1,000 s/mm<sup>2</sup>). Second, we sort the vectors in each group by choosing a random starting vector and then successively adding the vector with maximum spherical distance to the already chosen set of vectors until no vectors are left. Finally, we draw  $N$  samples evenly spaced from the obtained list.

## 4.2. Training implementation

Our method is trained and evaluated on an NVIDIA V100 GPU with 32 GB memory and a batch size in the range between 4,000 and 10,000 depending on the dataset. The learning rate (initially 0.01) is reduced (i.e., multiplied by 0.1), whenever the loss on the validation set has plateaued for 20 epochs. Early stopping is applied after a minimum of 40 epochs when the validation error has not

decreased for 40 successive epochs. As described in Section 3.2, we sample independent observation sets by random undersampling during training. The upper ceiling of the observation vectors is given by the measurements available in each dataset, that is, 270 for the HCP multi-shell dataset, 90 for the HCP single-shell dataset, and 112 for the grid-based dataset (see Section 4.1).

### 4.3. Reference methods

Inspired by reconstruction challenges (Ning et al., 2015, 2019), we compare our method against the following state-of-the-art methods: For the reconstruction of signals on a spherical domain, that is, for single-shell acquisitions, we choose the regularized least squares fit of spherical harmonics functions (Descoteaux et al., 2007) using the *Dipy* implementation (Garyfallidis et al., 2014) with an order of 8 and refer to this as *SH*. For reconstruction of signals on a continuous 3D space, that is, for multi-shell or grid-based acquisition, we use the *Dipy* implementations of *SHORE* (Özarslan et al., 2009) and *MAP-MRI* (Özarslan et al., 2013) (radial order 4 consistent with the setting in Ning et al. (2015)). We also compare with a compressed sensing approach (Paquette et al., 2015; Tobisch et al., 2018, 2019) (CS) on the grid-based dataset. Here, the optimization problem is solved via the iterative shrinkage and thresholding algorithm (Gong et al., 2013). On the validation set, we determine the optimal values for the regularization parameter  $\lambda_{CS}$  via grid-search (between 1.0 and  $10^{-9}$ ):  $\lambda_{CS} = 5 \cdot 10^{-6}, 5 \cdot 10^{-6}, 1 \cdot 10^{-6}, 1 \cdot 10^{-6}, 5 \cdot 10^{-7}$  for 10, 20, 30, 50, and 80 observation vectors, respectively.

In addition, we compare with discrete learning-based methods (see Section 2) for all domains. All discrete learning-based methods are based on MLPs predicting some signal-derived metric. For a signal reconstruction comparison, we adapt the network by Golkov et al. (2016) adjusting the output layer of the MLP to predict all signals available at training time. To ensure a fair comparison, we improve and ablate the best-performing MLP architecture under the condition of a similar number of parameters as *DISCUS* ( $\approx 170k$  parameters). The ablation, for example, establishes the optimal allocation of parameters to network depth (number of layers) and breadth (units per layer). We chose the best-performing approach on the validation set—an MLP having one hidden layer with 566 units—followed by a ReLU activation for this reference method and refer to it as *MLP*. We trained one instance per observation set for each dataset (a total of 15 networks) using stochastic gradient descent with a momentum of 0.9. The learning rate scheduling and early stopping are applied analogously to our method.

We also compare with a learning-based approach that learns SH-coefficients providing similar query flexibility for single-shell sampling as *DISCUS* (*MLP-SH*). This method predicts SH-coefficients from a signal vector. Based on work by Ye (2017), observations, and our ablation, we choose an end-to-end model driven by a signal reconstruction loss instead of a coefficient reconstruction loss. This end-to-end training also minimizes the previously mentioned information-loss limitation. More specifically, we reconstruct signals based on SH-coefficients during training and compute the  $\ell_2$ -loss between predicted and reference signal vectors. The network is the same as the aforementioned *MLP* but has a fixed number of 45 output neurons representing the SH-coefficients with order 8 and an initial learning rate of 0.01. As the number of input neurons is fixed and tied to the observation set, separate instances, that is, one per observation set, are trained analogously to the *MLP* method.

### 4.4. Objects of analysis and evaluation metrics

We compare different methods with respect to the quality of their respective signal reconstructions as well as the outputs of some typical analyses of these signals. As each voxel is an independent sample, we compare distributions of voxel-wise error metrics across the voxels of the test participants. These distributions are represented by boxplots, where the box is bounded by the 25th and 75th percentile; the marker inside the box denotes the median; the whisker's extension from the box denotes minimum (maximum) of 1.5 times the interquartile range (height of the box) and the maximum (minimum) error. For signal predictions  $\hat{s}_{ij}$  and reference signals  $s_{ij}$  at voxel  $i$  and for  $q$ -vector  $j = 1, \dots, J$ , we quantify the reconstruction error via

$$Error_{\text{signal}}(\hat{s}_i, s_i) = \frac{1}{J} \sum_{j=1}^J \frac{(\hat{s}_{ij} - s_{ij})^2}{s_{ij}^2}.$$

We have evaluated our method on various error metrics (squared, normalized squared, absolute error) and different ways to aggregate errors per imaging voxel, that is, along diffusion signals per voxel (mean, median, percentiles), but both the ranking and the overall statistics are consistent across different error metrics.

Furthermore, we compare reconstruction methods on downstream analyses of the signal by reconstructing DWIs from the observation sets and fitting *Diffusion Tensor Imaging* (DTI) (Basser et al., 1994; Westin et al., 1997), *Diffusion Kurtosis Imaging* (DKI) (Jensen et al., 2005), *Neurite Orientation Dispersion and Density Imaging* (NODDI) (H. Zhang et al., 2012), and *Mean Apparent Propagator MRI* (MAP-MRI) (Özarslan et al., 2013) models. As

an additional baseline, we also compare to the model-fits on the observation sets, that is, without intermediate reconstruction of withheld  $q$ -space signals.

We fit the DTI model with *FSL dtifit* (Smith et al., 2004), the DKI and MAP-MRI models with *Dipy* (Garyfallidis et al., 2014), and the NODDI model with the *Microstructure Diffusion Toolbox* (Harms & Roebroeck, 2018; Harms et al., 2017). We quantify the error between predicted metrics  $\hat{m}_i$  with metrics obtained from model-fits on the full reference data  $m_i$  at voxel  $i$  via

$$\text{Error}_{\text{metric}}(\hat{m}_i, m_i) = \frac{(\hat{m}_i - m_i)^2}{m_i^2}.$$

For the tensor model, we obtain separate reference data per observation set, that is, 10, 20, 30, etc. by fitting the tensor model only to those measurements not contained in the observation set, that is, the remaining 80, 70, 60, etc. to make input observation and output tensor fit independent and to avoid rewarding overfits to noise in the observation set.

Finally, we compare angular errors of peaks of fiber orientation distribution functions (fODFs) obtained via constrained spherical deconvolution on the single-shell data of the HCP dataset with the *MRtrix* package (Tournier et al., 2019). More specifically, we obtain response functions with *dwi2response* (Tournier et al., 2013), ODFs with *dwi2fod* (Tournier et al., 2007), and extract three peaks of the ODFs via *sh2peaks* (Jeurissen et al., 2013). We obtain masks for three characteristic fiber configurations: single fiber, crossings of two and three fibers (see Appendix Section A.3 for details) and report the distribution of average angular errors over voxels for each configuration and different observation set sizes.

## 5. ABLATION

To explore the impact of specific method components on performance on the data of the validation participants, we run these ablations by varying a single method component of interest and keeping everything else fixed. While we tested various components of our method, including different encoder and decoder designs, loss functions, and hyper-parameters, we present only five ablations in this section: latent space dimension and number of training participants on the HCP multi-shell dataset, augmentation schemes on the HCP single-shell dataset, and the ablation of loss function components on both datasets.

### 5.1. Latent space

The latent space is the interface between the encoder and the decoder. Its dimension determines the amount of

information that can be passed between the two components. To compare the role of the latent space dimension on performance, we train separate instances of our method with latent space dimensions of 4, 8, 16, 32, and 128 illustrated in Figure 2a. We find that the performance is best for a latent space of dimension 16 which is close to the optimal value for the compression of the diffusion signal with autoencoders determined by Zucchelli et al. (2021). In contrast, for small latent spaces, performance is substantially worse, suggesting that a dimension of 16 is large enough for sufficient information to be passed between the encoder and the decoder. Unsurprisingly, the full, rich information in dMRI cannot be compressed into a very small latent space. The performance decline for larger latent spaces is typical for autoencoder designs as compression is often a critical feature to regularize the reconstruction.

### 5.2. Rotation augmentation

As outlined in Section 3.2, we apply random rotations to the diffusion-encoding vectors during training. To show the effect of this augmentation type, we train two instances with and without rotation augmentation and evaluate each instance with and without a random rotation at test time. In Figure 2b, we compare the four instances. Comparing the model instances with rotation augmentation during training, we see that the rotation at test time does not impact performance at all, as expected. Performance improves a little when a model trained without rotations is tested on rotated data. Omitting the rotation augmentation during training worsens performance by about 17% while a bigger performance loss would be expected. This is likely because diffusion-encoding vectors of the same acquisition vary very little between participants and the coverage of  $q$ -vectors on the sphere is dense (90 vectors and mirrored versions, with inter-participant variance).

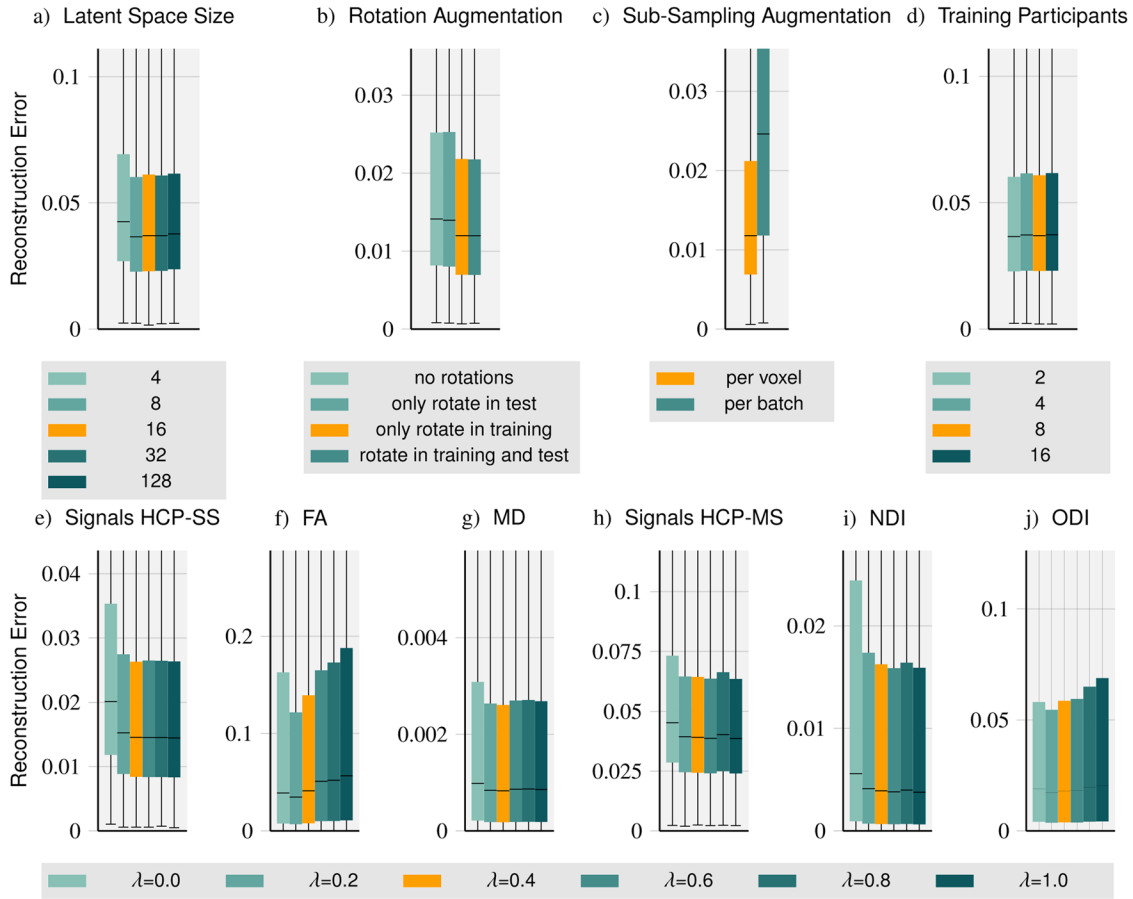
### 5.3. Sub-sampling augmentation

In addition to diverse rotations, we sub-sample and dynamically mask the observation set to establish diverse observation sets within each batch (“per voxel”, see also Section 3.2). As an alternative approach, we compare to a single random sub-sampling “per batch”. Figure 2c shows dynamic, diverse, “per voxel” sub-sampling significantly outperforming “per batch” sub-sampling underlining the impact of signal sampling diversity even within a batch.

### 5.4. Training participants

We compare performance when using data from different numbers of training participants. We evaluate separate





**Fig. 2.** Method ablations: We compare the impact of the latent space dimension (a), rotation augmentation (b), sub-sampling augmentation (c), and the number of training participants (d) on method performance. Furthermore, we vary the loss function parameter  $\lambda$  and evaluate signal reconstructions for single- and multi-shell (HCP-SS (e) and HCP-MS (h), respectively) as well as tensor-derived FA (f) and MD (g) based on HCP-SS data and NODDI-derived NDI (i) and ODI (j) based on HCP-MS data.

method instances trained with data from 2, 4, 8, and 16 participants, respectively. As shown in Figure 2d, the number of training participants surprisingly has no substantial impact on method performance. Since the reconstruction in each voxel is considered an independent reconstruction task, we attribute the performance to the number of training samples per participant (approximately 400,000 voxels are contained in the white and grey matter mask per HCP participant).

### 5.5. Loss weighing $\lambda$

In a final ablation, we explore the impact of the parameter  $\lambda$  weighing the contributions of observation and query components in the loss function on performance. In contrast to previous results, here, we analyze the errors for signal prediction and signal-derived microstructure models (DTI, NODDI). For signal prediction in single- and multi-shell settings, larger values of  $\lambda$ , that is, higher emphasis on the prediction of (unseen) signals (not contained in the observation set) improve performance. We

observe similar behavior for *Mean Diffusivity* (MD) and the *Neurite Density Index* (NDI). In contrast, for *Fractional Anisotropy* (FA) and the *Orientation Dispersion Index* (ODI), we observe an almost opposite characteristic showing a performance decrease for higher values of  $\lambda$ . Small values of  $\lambda$  incentivize the reconstruction of (noisy) inputs only which are possibly just passed through the network while the training of signal reconstruction for unseen query vectors is very limited. For larger values of  $\lambda$ , the incentive is shifted towards the signal prediction on uniform samples of  $q$ -space where extreme values essential for accurate FA estimation are rarely attained. Taking the two patterns into account, we choose  $\lambda = 0.4$  yielding a mixture of observation and query components in our method's loss function.

## 6. RESULTS

This section contains two parts, focusing on different aspects of reconstruction methods. First, we compare our *DISCUS* method against established state-of-the-art

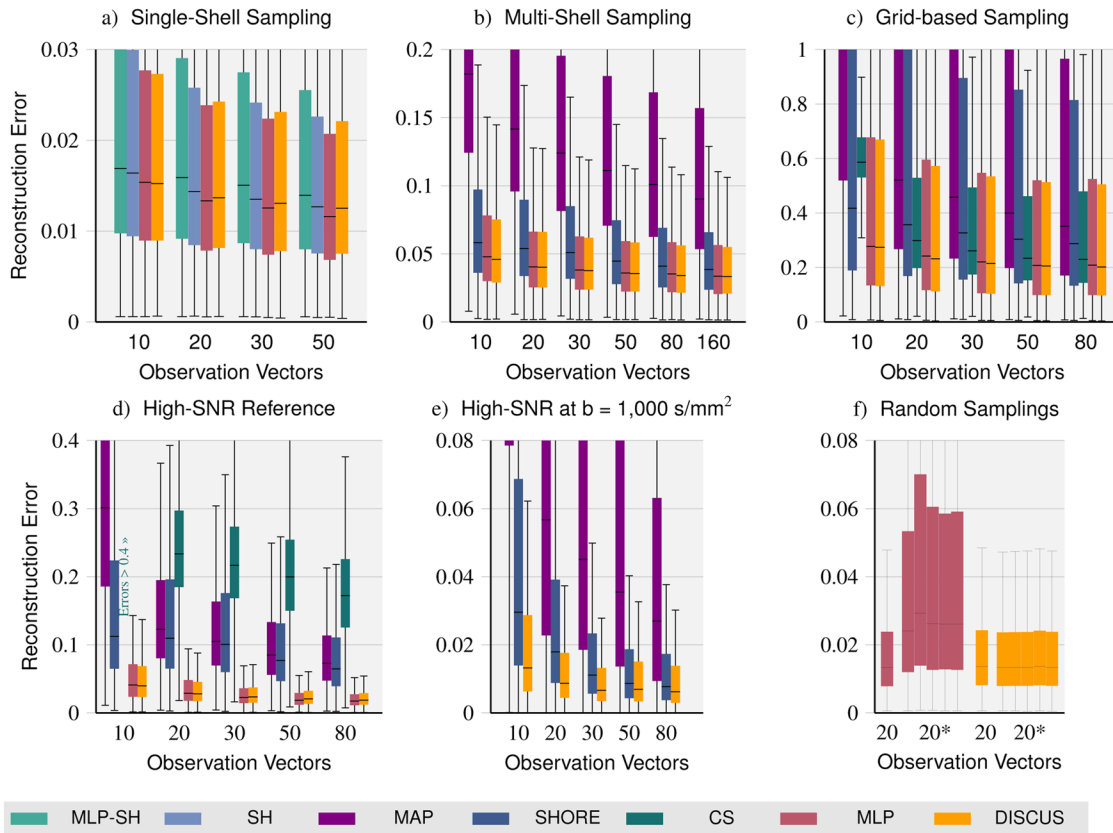
methods and compare the signal reconstruction quality on HCP and RLS datasets with different  $q$ -space samplings. Then, we compare reconstruction methods on signal-derived microstructure models frequently used for the analysis of dMRI signals.

### 6.1. Signal reconstruction

Here, we compare *DISCUS* with state-of-the-art reference methods (see Section 4.3) on datasets with single-shell, multi-shell, and grid-sampling of  $q$ -space (see Section 4.1). On the single-shell dataset (see Fig. 3a), we

compare our method with the *MLP-SH*, *SH*, and *MLP* methods. The *SH* method is fit independently to each set of observation vectors to provide signal predictions for the corresponding query vectors. The *MLP* requires separate training and a dedicated model per set of observation vectors, that is, 4, 6, and 5 models for single-shell, multi-shell, and grid, respectively. Analogously, four separate *MLP-SH* instances are trained for the single-shell dataset. In contrast, *DISCUS* only uses one model per sampling paradigm.

For all methods, we observe decreasing errors with larger observation sets indicating the increasing benefit



**Fig. 3.** Quantitative signal reconstruction comparison: On datasets featuring a single-shell (a), a multi-shell (b), and a grid sampling of  $q$ -space (c), we compare our *DISCUS* method (yellow) with state-of-the-art reconstruction methods: an MLP learning spherical harmonics coefficients (*MLP-SH*), spherical harmonics (*SH*) (on single-shell data), MAP-MRI (*MAP*), *SHORE* on multi-shell and grid-based data, and compressed sensing DSI (*CS*) on grid-based data as well as the *MLP*, which is separately trained per dataset and sub-sampling. For all methods, we obtain decreasing reconstruction errors for larger observation sets as more information for reconstruction is available. The learning-based methods *MLP* and *DISCUS* seem to outperform *SH*, *MLP-SH*, *CS*, and *SHORE* slightly and considerably so for *MAP*. Both the *MLP* and *DISCUS* yield almost equal error distributions. Other aspects of the signal reconstruction method comparison: We compare the reconstruction quality on two high-SNR reference datasets: The first set (d) contains high SNR signals for  $q$ -vectors contained in the sampling scheme of the Rhineland Study. The second set (e) contains six high-SNR signals for  $q$ -vectors at a  $b$ -value of 1,000 s/mm<sup>2</sup> which are not contained in the Rhineland Study sampling (training data). Unlike the *MLP* and *CS*, *SHORE*, *MAP*, and our *DISCUS* method support the prediction of signals for  $q$ -vectors not contained in the training data. Subfigure (f) compares the *MLP* and *DISCUS* networks trained on the single-shell dataset on different subsets of 20 observation vectors: First, on the predefined set of Subfigure (a), and then additionally on five unseen, randomly sampled observation sets (20\*). None of the observation vectors are contained in the query set. As the random observation sets do not match the training data, performance drops for the *MLP* method whereas our *DISCUS* method maintains its performance. Parts (e) and (f) emphasize the capabilities *DISCUS* has over the *MLP* method.

methods can gain from the increased, available information for reconstruction. For larger observation sets, the reconstruction quality of *SH* is equivalent to our *DISCUS* method. For smaller observation sets, we note a gap where our method outperforms *SH*. The extremely similar error distributions observed for *MLP* and *DISCUS* across observation sets illustrate that the error is increasingly dominated by the acquisition noise. For larger observation sets, we note a gap where the *MLP* has slightly lower errors compared to *DISCUS*. All methods outperform *MLP-SH* on all observation sets.

On the multi-shell dataset (see Fig. 3b), we compare *DISCUS* with the *MLP*, *MAP*, and *SHORE* methods. Analogously to *SH*, *MAP* and *SHORE* are fit independently to each observation set, voxel, and participant. Here, we also observe decreasing errors for larger observation sets for all methods. Both *MAP* and *SHORE* are outperformed by the *MLP* and *DISCUS* across observation sets. The error distributions of *MLP* and *DISCUS* look very similar with slightly lower errors for *DISCUS*.

Finally, for the grid-based sampling, we compare our approach with *MAP*, *SHORE*, *CS*, and the *MLP* (see Fig. 3c) finding decreasing errors for all methods for larger observation sets. Similar to the multi-shell dataset, the learning-based methods *MLP* and *DISCUS* outperform both *MAP* and *SHORE* substantially. For the *CS* method, we note narrower error distributions. However, while its performance is very good for large observation sets, the performance declines slightly for smaller observation sets and drops substantially when only 10 observation vectors are available. As on the multi-shell dataset, *MLP* and *DISCUS* have very similar error distributions with a slight advantage for *DISCUS*.

As the acquisition noise is a substantial part of the errors, reconstruction methods are evaluated and compared on the high-SNR evaluation dataset (see Fig. 3d and e). Here, *MAP*, *SHORE*, *CS*, *MLP*, and *DISCUS* methods are fit to the same observation sets as in the Rhineland Study evaluation above. However, reconstruction quality is determined on the high-SNR signals based on 20 acquisitions.

In the first part of this evaluation, these high-SNR signals were obtained for *q*-vectors also contained in the Rhineland Study, that is, the *CS* and pre-trained *MLP* methods are applicable. We observe low performance of *CS*, performance in the middle range of *MAP* and *SHORE*, and high performance of the pre-trained learning methods *MLP* and *DISCUS* which have very similar error distributions again.

In the second part, we evaluate signals for *q*-vectors not contained in the sampling of the RLS, that is, unknown for the training dataset where the *CS* and *MLP* methods

cannot be applied. Figure 3e shows that *DISCUS* substantially outperforms *SHORE* and *MAP*.

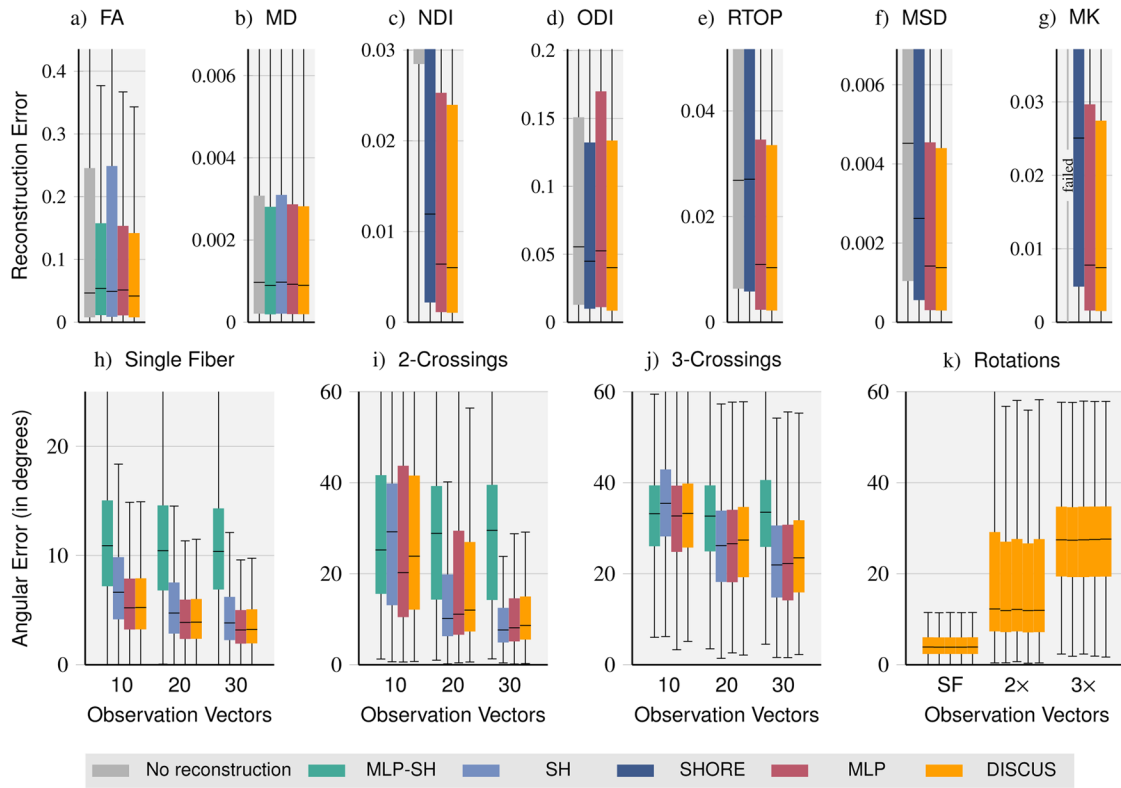
Finally, we test the effect of changes to the observation sets at test time and compare the resulting performance on the single-shell dataset in Figure 3f. As expected, the performance of the *MLP* suffers a lot as the choice and order of observations differ from the fixed pattern used for training. In contrast, our *DISCUS* method maintains its performance as it is permutation-invariant by design and has been trained to reconstruct based on arbitrary observation sets. Here, the remaining variability in performance between observation sets stems from the fact that different observation sets contain different acquisition noise and information for the reconstruction.

## 6.2. Downstream analysis models

Beyond signal reconstruction, we evaluate how signal predictions impact measures derived from the DWIs. In Figure 4, we compare the reconstruction methods on DTI-derived metrics for observation sets with 10 observation vectors of the single-shell data and DKI-, NODDI-, and MAP-MRI-derived metrics based on 10 observation vectors of the multi-shell data. Target reference values are computed from all DWIs not contained in the observation set, that is, 80 DWIs for the single-shell, but all 270 DWIs for the multi-shell data.

For the DTI-derived metrics *Fractional Anisotropy* (FA) and *Mean Diffusivity* (MD), we compare fits directly on the sub-sampled data (no reconstruction of missing signals) and 90 signals reconstructed with *MLP-SH*, *SH*, *MLP*, and *DISCUS*, respectively. For FA, the learning-based methods *MLP-SH*, *MLP* and *DISCUS* outperform *no reconstruction* and *SH*. *DISCUS* performs best. For MD, we obtain very similar error distributions overall and particularly for *MLP-SH*, *MLP*, and *DISCUS*.

On the multi-shell dataset, we compare methods on DKI, NODDI, and MAP-MRI models. Due to the bad signal reconstruction performance of *MAP* on the multi-shell dataset, we exclude *MAP* in this comparison. For the *Neurite Density Index* (NDI), *no reconstruction* performs substantially worse indicating that the reconstruction of additional signals is very beneficial for this metric. Again, the learning-based methods perform best. For the *Orientation Dispersion Index* (ODI), *SHORE* outperforms *no reconstruction* and *MLP*, while *DISCUS* outperforms *SHORE* slightly. Similarly, for the MAP-MRI-derived metrics *Return To Origin Probability* (RTOP) and *Mean Squared Displacement* (MSD) and DKI-derived *Mean Kurtosis* (MK), the learning-based methods outperform *SHORE* and *no reconstruction*. Without reconstruction, the diffusion kurtosis model failed, that is, returned only zeros when only 10 observation vectors were available.



**Fig. 4.** Method comparison on signal-derived metrics: We compare the quality of DTI-derived *Fractional Anisotropy* (FA), *Mean Diffusivity* (MD), NODDI-derived *Neurite Density Index* (NDI) and *Orientation Dispersion Index* (ODI), MAP-MRI-derived *Return To Origin Probability* (RTOP) and *Mean Squared Displacement* (MSD), and DKI-derived *Mean Kurtosis* (MK) metrics obtained with sub-sampled data (no reconstruction) and with data reconstructed with the *MLP-SH*, *SH*, *SHORE*, *MLP*, and *DISCUS* methods, respectively (a-g). In addition, we compare the angular errors of fiber orientation distribution function (fODF) peaks for different fiber configurations: single fiber (h), and crossings of two (i) and three fibers (j), respectively. Finally, we let *DISCUS* reconstruct signals for five random rotations in *q*-space, obtain fODF peaks, and report angular errors for single fiber (SF), two- and three-crossings (2 $\times$  and 3 $\times$ ) illustrating the rotation invariance of *DISCUS* (k).

On closer inspection, we saw a sharp performance decline, that is, more and more zeros were returned as we incrementally lowered the number of observation vectors. For most metrics, the results suggest that *DISCUS* performs best albeit just slightly better than the *MLP*.

Another important application of dMRI is the computation of *Fiber Orientation Distribution Functions* (fODFs), for example, to inform tractography. In Figure 4h-j, we evaluate the impact of signal reconstruction with different methods on the estimation of fODF peaks and report error distributions for each method, different degrees of sparsity (10, 20, and 30 observation vectors), as well as different fiber configurations, that is, voxels containing a single fiber and two- and three-crossings. *MLP-SH* is outperformed with very few exceptions. On single-fiber voxels, *MLP* and *DISCUS* perform equally well outperforming *SH*. On voxels with two- and three-crossings, *MLP* and *DISCUS* outperform *SH* for sparse observations (10 *q*-vectors) and *SH* slightly outperforms the two on larger observation sets (20 and 30 vectors). *MLP* outperforms *DISCUS* on the voxels featuring fiber crossings

albeit just slightly in most settings. The magnitude of angular error suggests that fiber directions can be reliably inferred in single-fiber voxels even from sparse acquisitions. In contrast, as the fiber configurations get more complex, more measurements seem necessary to accurately infer the peaks. For two-crossings, we observe lower errors starting at 30 observation vectors. For three-crossings, all methods yield errors above 20 degrees (median value), suggesting that more than 30 vectors are required to resolve three-crossings well. In Figure 4k, we obtain five random rotations of *q*-space, reconstruct signals with *DISCUS* for the rotated observation and query vectors, and determine fODF peaks. The very similar performance across rotations illustrates the rotation invariance of the *DISCUS* method.

## 7. CONCLUSION

In this paper, we have presented the first entirely continuous learning-based method for the reconstruction of dMRI signals. This method is input-continuous—it can be



applied to any (sufficiently dense) acquisition scheme—and output-continuous—it can infer signals for arbitrary  $q$ -vectors. *DISCUS* offers a methodological framework to address two challenges in dMRI: the sparsity of  $q$ -space sampling to keep scan times short and the diversity of sampling schemes between studies. Our approach permits the reliable inference of additional signals, especially from short acquisitions. Specifically, we have shown that *DISCUS* outperforms both continuous (*SHORE* and *MAP-MRI*) and discrete model-based methods (*CS*) as well as learning-based semi-continuous methods (*MLP-SH*). Compared to the strongest competitor, an optimized *MLP*, *DISCUS*'s and the *MLP*'s error distributions are very similar, but with the difference that *DISCUS* is entirely continuous offering support for arbitrary samplings—unlike the discrete *MLP*. Additionally, we have demonstrated that *DISCUS*- and *MLP*-derived reconstructions yield very similar errors in downstream evaluations consistently outperforming other reference methods. While technically within error margins, the *MLP* is slightly better in fODF-based angular errors and *DISCUS* is slightly better in estimates of microstructure models (DTI, DKI, NODDI, MAP-MRI). Our proposed *DISCUS* method may be used to accelerate acquisitions, to derive higher-quality signal-derived metrics retrospectively from existing acquisitions, or for dMRI data harmonization by resampling datasets with different  $q$ -space acquisition schemes onto a common scheme.

## DATA AND CODE AVAILABILITY

The MRI data of the Human Connectome Project (HCP) are publicly available on the website <https://www.humanconnectome.org>. The MRI data of the Rhineland Study are not publicly available due to data protection regulations. Access can be provided to scientists in accordance with the Rhineland Study's Data Use and Access Policy. Requests to access the data should be directed to Dr. Monique Breteler at [RS-DUAC@dzne.de](mailto:RS-DUAC@dzne.de). The source code for the *DISCUS* method can be found on the website <https://github.com/Deep-MI/DISCUS>.

## AUTHOR CONTRIBUTIONS

Christian Ewert: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Writing—Original Draft, and Writing—Review & Editing; David Kügler: Conceptualization, Methodology, Formal Analysis, Writing—Original Draft, Writing—Review & Editing, and Supervision; Rüdiger Stirnberg: Resources, Writing—Original Draft, and Writing—Review & Editing; Alexandra Koch: Software, Data Curation, Writing—Original Draft, and Writing—Review & Editing; Anastasia Yendiki:

Writing—Original Draft, Writing—Review & Editing, and Funding Acquisition; Martin Reuter: Conceptualization, Resources, Writing—Original Draft, Writing—Review & Editing, Supervision, Project Administration, and Funding Acquisition

## FUNDING

This work is supported by DZNE institutional funds, the Federal Ministry of Education and Research of Germany (grant numbers 031L0206, 01GQ1801), the Chan-Zuckerberg Initiatives Essential Open Source Software for Science RFA (EOSS5 2022-252594), the Helmholtz-AI project DeGen (grant number ZT-I-PF-5-078), by the National Institute of Biomedical Imaging and Bioengineering (grant numbers R01-EB021265, U01-EB026996, and P41-EB030006), the National Institute of Neurological Disorders and Stroke (grant number R01-NS119911), the National Institute on Aging (grant number R01-AG064027), the National Institute of Mental Health (grant number R56-MH121426), and the National Library of Medicine (grant number R01-LM012719). The funding sources had no involvement in the study design, the collection, analysis, and interpretation of data, in the writing of the report, or the decision to submit the article for publication.

## DECLARATION OF COMPETING INTEREST

The authors have no conflicts of interest to declare.

## ACKNOWLEDGEMENTS

Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. We thank the Rhineland Study group at the DZNE (PI Monique M.B. Breteler) for supporting the data acquisition and management.

## REFERENCES

- Albay, E., Demir, U., & Ünal, G. B. (2018). Diffusion MRI spatial super-resolution using generative adversarial networks. In *MICCAI 2018, PRIME Workshop* (pp. 155–163). Springer. [https://doi.org/10.1007/978-3-030-00320-3\\_19](https://doi.org/10.1007/978-3-030-00320-3_19)
- Andersson, J., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, 20, 870–888. [https://doi.org/10.1016/s1053-8119\(03\)00336-7](https://doi.org/10.1016/s1053-8119(03)00336-7)
- Andersson, J., & Sotiropoulos, S. (2016). An integrated approach to correction for off-resonance effects and

- subject movement in diffusion MR imaging. *NeuroImage*, 125, 1063–1078. <https://doi.org/10.1016/j.neuroimage.2015.10.019>
- Basser, P. J., Mattiello, J., & LeBihan, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B*, 103(3), 247–254. <https://doi.org/10.1006/jmrb.1994.1037>
- Breteler, M., Stöcker, T., Pracht, E., Brenner, D., & Stirnberg, R. (2014). MRI in the Rhineland study: A novel protocol for population neuroimaging. *Alzheimer's & Dementia*, 10, 92. <https://doi.org/10.1016/j.jalz.2014.05.172>
- Chatterjee, S., Sciarra, A., Dünnwald, M., Mushunuri, R. V., Podishetti, R., Rao, R. N., Gopinath, G. D., Oeltze-Jafra, S., Speck, O., & Nürnberger, A. (2021). ShuffleUNet: Super resolution of diffusion-weighted MRIs using deep learning. In *EUSIPCO 2021* (pp. 940–944). IEEE. <https://doi.org/10.23919/eusipco54536.2021.9615963>
- Chen, G., Hong, Y., Zhang, Y., Kim, J., Huynh, K. M., Ma, J., Lin, W., Shen, D., & Yap, P. T. (2020). Estimating tissue microstructure with undersampled diffusion data via graph convolutional neural networks. In *MICCAI 2020* (pp. 280–290). Springer. [https://doi.org/10.1007/978-3-030-59728-3\\_28](https://doi.org/10.1007/978-3-030-59728-3_28)
- Chen, G., Jiang, H., Liu, J., Ma, J., Cui, H., Xia, Y., & Yap, P. T. (2022). Hybrid graph transformer for tissue microstructure estimation with undersampled diffusion MRI data. In *MICCAI 2022* (pp. 113–122). Springer. [https://doi.org/10.1007/978-3-031-16431-6\\_11](https://doi.org/10.1007/978-3-031-16431-6_11)
- Cheng, J., Jiang, T., Deriche, R., Shen, D., & Yap, P. T. (2013). Regularized spherical polar Fourier diffusion MRI with optimal dictionary learning. In *MICCAI 2013* (pp. 639–646). Springer. [https://doi.org/10.1007/978-3-642-40811-3\\_80](https://doi.org/10.1007/978-3-642-40811-3_80)
- Cheng, J., Shen, D., Yap, P., & Basser, P. (2018). Single- and multiple-shell uniform sampling schemes for diffusion MRI using spherical codes. *IEEE Transactions on Medical Imaging*, 37, 185–199. <https://doi.org/10.1109/tmi.2017.2756072>
- Descoteaux, M., Angelino, E., Fitzgibbons, S., & Deriche, R. (2007). Regularized, fast, and robust analytical Q-ball imaging. *Magnetic Resonance in Medicine*, 58(3), 497–510. <https://doi.org/10.1002/mrm.21277>
- Fadnavis, S., Batson, J. D., & Garyfallidis, E. (2020). Patch2Self: Denoising diffusion MRI with self-supervised learning. In *NeurIPS 2020* (pp. 16293–16303). [https://proceedings.neurips.cc/paper\\_files/paper/2020/file/bc047286b224b7bfa73d4cb02de1238d-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2020/file/bc047286b224b7bfa73d4cb02de1238d-Paper.pdf)
- Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Gunther, M., Glasser, M. F., Miller, K. L., Ugurbil, K., & Yacoub, E. (2010). Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. *PLoS One*, 5(12), e15710. <https://doi.org/10.1371/journal.pone.0015710>
- Fick, R., Daianu, M., Pizzolato, M., Wassermann, D., Jacobs, R. E., Thompson, P. M., Town, T. C., & Deriche, R. (2017). Comparison of biomarkers in transgenic Alzheimer rats using multi-shell diffusion MRI. In *MICCAI 2016* (pp. 187–199). Springer. [https://doi.org/10.1007/978-3-319-54130-3\\_16](https://doi.org/10.1007/978-3-319-54130-3_16)
- Fick, R., Wassermann, D., Caruyer, E., & Deriche, R. (2016). MAPL: Tissue microstructure estimation using Laplacian-regularized MAP-MRI and its application to HCP data. *NeuroImage*, 134, 365–385. <https://doi.org/10.1016/j.neuroimage.2016.03.046>
- Galazzo, I. B., Brusini, L., Obertino, S., Zucchelli, M., Granziera, C., & Menegaz, G. (2018). On the viability of diffusion MRI-based microstructural biomarkers in ischemic stroke. *Frontiers in Neuroscience*, 12. <https://doi.org/10.3389/fnins.2018.00092>
- Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Walt, S., Descoteaux, M., & Nimmo-Smith, I. (2014). Dipy, a library for the analysis of diffusion MRI data. *Frontiers in Neuroinformatics*, 8. <https://doi.org/10.3389/fninf.2014.00008>
- Gibbons, E. K., Hodgson, K. K., Chaudhari, A. S., Richards, L. G., Majersik, J., Adluru, G., & DiBella, E. V. R. (2019). Simultaneous NODDI and GFA parameter map generation from subsampled q-space imaging using deep learning. *Magnetic Resonance in Medicine*, 81, 2399–2411. <https://doi.org/10.1002/mrm.27568>
- Golkov, V., Dosovitskiy, A., Sperl, J. I., Menzel, M. I., Czisch, M., Sämann, P. G., Brox, T., & Cremers, D. (2016). q-space deep learning: Twelve-fold shorter and model-free diffusion MRI scans. *Transactions on Medical Imaging*, 35, 1344–1351. <https://doi.org/10.1109/tmi.2016.2551324>
- Gondara, L. (2016). Medical image denoising using convolutional denoising autoencoders. In *ICDMW 2016* (pp. 241–246). IEEE. <https://doi.org/10.1109/icdmw.2016.0041>
- Gong, P., Zhang, C., Lu, Z., Huang, J. Z., & Ye, J. (2013). A General iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In: *ICML 2013* (pp. 37–45). <https://proceedings.mlr.press/v28/gong13a.pdf>
- Grussu, F., Blumberg, S. B., Battiston, M., Ianus, A., Singh, S., Gong, F., Whitaker, H., Atkinson, D., Wheeler-Kingshott, C. A. M. G., Punwani, S., Panagiotaki, E., Mertzaniadou, T., & Alexander, D. C. (2020). SARU-Net: A new method for model-free, data-driven experiment design in quantitative MRI. In *ISMRM 2020* (p. 1035). <https://cds.ismrm.org/protected/20MProceedings/PDFfiles/1035.html>
- Harms, R. L., Fritz, F. J., Tobisch, A., Goebel, R., & Roebroeck, A. (2017). Robust and fast nonlinear optimization of diffusion MRI microstructure models. *NeuroImage*, 155, 82–96. <https://doi.org/10.1016/j.neuroimage.2017.04.064>
- Harms, R. L., & Roebroeck, A. (2018). Robust and fast Markov chain Monte Carlo sampling of diffusion MRI microstructure models. *Frontiers in Neuroinformatics*, 12. <https://doi.org/10.3389/fninf.2018.00097>
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML 2015* (pp. 448–456). PMLR. <http://proceedings.mlr.press/v37/loff15.html>
- Jensen, J. H., Helpert, J. A., Ramani, A., Lu, H., & Kaczynski, K. (2005). Diffusional kurtosis imaging: The quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine*, 53, 1432–1440. <https://doi.org/10.1002/mrm.20508>
- Jeurissen, B., Leemans, A., Tournier, J. D., Jones, D. K., & Sijbers, J. (2013). Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Human Brain Mapping*, 34(11), 2747–2766. <https://doi.org/10.1002/hbm.22099>
- Jha, R. R., Jaswal, G., Bhavsar, A. V., & Nigam, A. (2022). Single-shell to multi-shell dMRI transformation using spatial and volumetric multilevel hierarchical reconstruction framework. *Magnetic Resonance Imaging*, 87, 133–156. <https://doi.org/10.1016/j.mri.2021.12.011>
- Koonjoo, N., Zhu, B., Bagnall, G. C., Bhutto, D. F., & Rosen, M. S. (2021). Boosting the signal-to-noise of low-field MRI with deep learning image reconstruction. *Scientific Reports*, 11, 8248. <https://doi.org/10.1038/s41598-021-87482-7>

- Li, H., Liang, Z., Zhang, C., Liu, R., Li, J., Zhang, W., Liang, D., Shen, B., Zhang, X., Ge, Y., Zhang, J., & Ying, L. (2021). SuperDTI: Ultrafast DTI and fiber tractography with deep learning. *Magnetic Resonance in Medicine*, 86(6), 3334–3347. <https://doi.org/10.1002/mrm.28937>
- Lim, D., Robinson, J., Zhao, L., Smidt, T. E., Sra, S., Maron, H., & Jegelka, S. (2023). Sign and basis invariant networks for spectral graph representation learning. In *ICLR 2023*. <https://openreview.net/pdf?id=Q-UHqMorzil>
- 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. *Medical Physics*, 46, 3101–3116. <https://doi.org/10.1002/mp.13555>
- Mani, M. P., Yang, B., Bathla, G., Magnotta, V. A., & Jacob, M. (2022). Multi-band- and in-plane-accelerated diffusion MRI enabled by model-based deep learning in q-space and its extension to learning in the spherical harmonic domain. *Magnetic Resonance in Medicine*, 87, 1799–1815. <https://doi.org/10.1002/mrm.29095>
- Michailovich, O. V., Rath, Y., & Dolui, S. (2011). Spatially regularized compressed sensing for high angular resolution diffusion imaging. *IEEE Transactions on Medical Imaging*, 30(5), 1100–1115. <https://doi.org/10.1109/tmi.2011.2142189>
- Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., & Ugurbil, K. (2010). Multiband multislice GE-EPI at 7 Tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic Resonance in Medicine*, 63(5), 1144–1153. <https://doi.org/10.1002/mrm.22361>
- Nath, V., Lyu, I., Schilling, K., Parvathaneni, P., Hansen, C., Huo, Y., Janve, V., Gao, Y., Stepniewska, I., Anderson, A., & Landman, B. (2019). Enabling multi-shell b-value generalizability of data-driven diffusion models with deep SHORE. In *MICCAI 2019* (pp. 573–581). Springer. [https://doi.org/10.1007/978-3-030-32248-9\\_64](https://doi.org/10.1007/978-3-030-32248-9_64)
- Nath, V., Pathak, S. K., Schilling, K. G., Schneider, W., & Landman, B. A. (2020). Deep learning estimation of multi-tissue constrained spherical deconvolution with limited single shell DW-MRI. In *SPIE 2020* (p. 113130S). <https://doi.org/10.1117/12.2549455>
- Ning, L., Bonet-Carne, E., Grussu, F., Sepehrband, F., Kaden, E., Veraart, J., Blumberg, S. B., Khoo, C. S., Palombo, M., Coll-Font, J., Scherrer, B., Warfield, S., Karayumak, S. C., Rath, Y., Koppers, S., Weninger, L., Ebert, J., Merhof, D., Moyer, D., ... Tax, C. M. (2019). Multi-shell diffusion MRI harmonisation and enhancement challenge (MUSHAC): Progress and results. In *MICCAI 2019, CDMRI Workshop* (pp. 217–224). Springer. [https://doi.org/10.1007/978-3-030-05831-9\\_18](https://doi.org/10.1007/978-3-030-05831-9_18)
- Ning, L., Laun, F., Gur, Y., Bella, E. D., Deslauriers-Gauthier, S., Megherbi, T., Ghosh, A., Zucchelli, M., Menegaz, G., Fick, R., St-Jean, S., Paquette, M., Aranda, R., Descoteaux, M., Deriche, R., O'Donnell, L., & Rath, Y. (2015). Sparse reconstruction challenge for diffusion MRI: Validation on a physical phantom to determine which acquisition scheme and analysis method to use? *Medical Image Analysis*, 26(1), 316–331. <https://doi.org/10.1016/j.media.2015.10.012>
- Özarslan, E., Koay, C. G., Shepherd, T., Blackband, S. J., & Basser, P. J. (2009). Simple harmonic oscillator based reconstruction and estimation for three-dimensional q-space MRI. In *ISMRM 2009*. <https://archive.ismrm.org/2009/1396.html>
- Özarslan, E., Koay, C. G., Shepherd, T. M., Komlos, M. E., Irfanoglu, M. O., Pierpaoli, C., & Basser, P. J. (2013). Mean apparent propagator (MAP) MRI: A novel diffusion imaging method for mapping tissue microstructure. *NeuroImage*, 78, 16–32. <https://doi.org/10.1016/j.neuroimage.2013.04.016>
- Paquette, M., Merlet, S., Gilbert, G., Deriche, R., & Descoteaux, M. (2015). Comparison of sampling strategies and sparsifying transforms to improve compressed sensing diffusion spectrum imaging. *Magnetic Resonance in Medicine*, 73(1), 401–416. <https://doi.org/10.1002/mrm.25093>
- Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). PointNet: Deep learning on point sets for 3D classification and segmentation. In *CVPR 2017* (pp. 77–85). IEEE. <https://doi.org/10.1109/cvpr.2017.16>
- Rathi, Y., Michailovich, O. V., Laun, F. B., Setsompop, K., Grant, P. E., & Westin, C. F. (2014). Multi-shell diffusion signal recovery from sparse measurements. *Medical Image Analysis*, 18(7), 1143–1156. <https://doi.org/10.1016/j.media.2014.06.003>
- Ren, M., Kim, H., Dey, N., & Gerig, G. (2021). Q-space conditioned translation networks for directional synthesis of diffusion weighted images from multi-modal structural MRI. In *MICCAI 2021 (Part VII)* (pp. 530–540). Springer. [https://doi.org/10.1007/978-3-030-87234-2\\_50](https://doi.org/10.1007/978-3-030-87234-2_50)
- Sedlar, S., Alimi, A., Papadopoulos, T., Deriche, R., & Deslauriers-Gauthier, S. (2021). A spherical convolutional neural network for white matter structure imaging via dMRI. In *MICCAI 2021 (Part III)* (pp. 529–539). Springer. [https://doi.org/10.1007/978-3-030-87199-4\\_50](https://doi.org/10.1007/978-3-030-87199-4_50)
- Setsompop, K., Gagoski, B. A., Polimeni, J. R., Witzel, T., van Wede, J., & Wald, L. L. (2012). Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic Resonance in Medicine*, 67(5), 1210–1224. <https://doi.org/10.1002/mrm.23097>
- Shoemaker, K. (1992). Uniform random rotations. In D. B. Kirk (Ed.), *Graphics Gems III (IBM Version)* (pp. 124–132). Academic Press. The Graphics Gems Series. <https://doi.org/10.1016/b978-0-08-050755-2.50036-1>
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23, s208–s219. <https://doi.org/10.1016/j.neuroimage.2004.07.051>
- Sotiropoulos, S. N., Moeller, S., Jbabdi, S., Xu, J., Andersson, J. L., Auerbach, E. J., Yacoub, E., Feinberg, D., Setsompop, K., Wald, L. L., Behrens, T. E., Ugurbil, K., & Lenglet, C. (2013). Effects of image reconstruction on fiber orientation mapping from multichannel diffusion MRI: Reducing the noise floor using SENSE. *Magnetic Resonance in Medicine*, 70(6), 1682–1689. <https://doi.org/10.1002/mrm.24623>
- Stöcker, T. (2016). Big Data: The Rhineland study. In *ISMRM 2016*. <https://cds.ismrm.org/protected/16MProceedings/PDFfiles/6865.html>
- Tobisch, A., Schultz, T., Stirnberg, R., Varela-Mattatall, G., Knutsson, H., Irarrázaval, P., & Stöcker, T. (2019). Comparison of basis functions and q-space sampling schemes for robust compressed sensing reconstruction accelerating diffusion spectrum imaging. *NMR in Biomedicine*, 32(3), e4055. <https://doi.org/10.1002/nbm.4055>
- Tobisch, A., Stirnberg, R., Harms, R., Schultz, T., Roebroek, A., Breteler, M., & Stöcker, T. (2018). Compressed sensing diffusion spectrum imaging for accelerated diffusion microstructure MRI in long-term



- population imaging. *Frontiers in Neuroscience*, 12. <https://doi.org/10.3389/fnins.2018.00650>
- Tournier, J. D., Calamante, F., & Connelly, A. (2007). Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage*, 35(4), 1459–1472. <https://doi.org/10.1016/j.neuroimage.2007.02.016>
- Tournier, J. D., Calamante, F., & Connelly, A. (2013). Determination of the appropriate b value and number of gradient directions for high-angular-resolution diffusion-weighted imaging. *NMR in Biomedicine*, 26(12), 1775–1786. <https://doi.org/10.1002/nbm.3017>
- Tournier, J. D., Smith, R. E., Raffelt, D. A., 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. <https://doi.org/10.1016/j.neuroimage.2019.116137>
- Tuch, D. S. (2004). Q-ball imaging. *Magnetic Resonance in Medicine*, 52(6), 1358–1372. <https://doi.org/10.1002/mrm.20279>
- van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S. W., Della Penna, S., Feinberg, D., Glasser, M. F., Harel, N., Heath, A. C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S. E., Prior, F., Schlaggar, B. L., Smith, S. M., Snyder, A. Z., Xu, J., & Yacoub, E. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage*, 62(4), 2222–2231. <https://doi.org/10.1016/j.neuroimage.2012.02.018>
- Varadarajan, D., & Haldar, J. P. (2018). Towards optimal linear estimation of orientation distribution functions with arbitrarily sampled diffusion MRI data. In *ISBI 2018* (pp. 743–746). IEEE. <https://doi.org/10.1109/isbi.2018.8363680>
- Wedeer, V. J., Hagmann, P., Tseng, W. Y. I., Reese, T. G., & Weisskoff, R. M. (2005). Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magnetic Resonance in Medicine*, 54, 1377–1386. <https://doi.org/10.1002/mrm.20642>
- Westin, C. F., Peled, S., Gudbjartsson, H., Kikinis, R., & Jolesz, F. A. (1997). Geometrical diffusion measures for MRI from tensor basis analysis. In *ISMRM 1997* (p. 1742).
- Xu, J., Moeller, S., Strupp, J., Auerbach, E., Feinberg, D. A., Ugurbil, K., & Yacoub, E. (2012). Highly accelerated whole brain imaging using aligned-blipped-controlled-aliasing multiband EPI. In *ISMRM 2012*. <https://archive.ismrm.org/2012/2306.html>
- Ye, C. (2017). Learning-based ensemble average propagator estimation. In *MICCAI 2017* (pp. 593–601). Springer. [https://doi.org/10.1007/978-3-319-66182-7\\_68](https://doi.org/10.1007/978-3-319-66182-7_68)
- Zhang, H., Schneider, T., Wheeler-Kingshott, C. A. M., & Alexander, D. C. (2012). NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *NeuroImage*, 61(4), 1000–1016. <https://doi.org/10.1016/j.neuroimage.2012.03.072>
- Zhang, Y., Brady, M., & Smith, S. M. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1), 45–57. <https://doi.org/10.1109/42.906424>
- Zucchelli, M., Deslauriers-Gauthier, S., & Deriche, R. (2021). Brain tissue microstructure characterization using dMRI based autoencoder neural-networks. In *MICCAI 2021, CDMRI Workshop* (pp. 48–57). Springer. [https://doi.org/10.1007/978-3-030-87615-9\\_5](https://doi.org/10.1007/978-3-030-87615-9_5)

## APPENDIX

### A.1. Difference maps

In [Appendix Figure C](#), we show an example diffusion-weighted image reconstructed with *DISCUS* from **only 10** observation vectors (for a *b*-vector that was not part of the observation set). Additionally, we show an FA map computed from 90 predicted DWIs (also predicted based on 10 observations). Along with these images, we plot the reference images and the normalized squared error maps with respect to reference signals and FA maps, respectively. The signal error map shows low errors along major white matter tracts and most of the grey matter. Errors are observed over CSF regions such as the ventricles and are noticeable in the subcortical grey matter. The error in FA is also low along major white matter tracts, pronounced in CSF regions, and substantial in subcortical grey matter, for example, thalamus. Because of the limited relevance of the CSF in dMRI, CSF voxels are not part of the training set and are ignored in the evaluation. Differences in magnitude between white and grey matter and CSF may help explain these error maps (as the error is normalized).

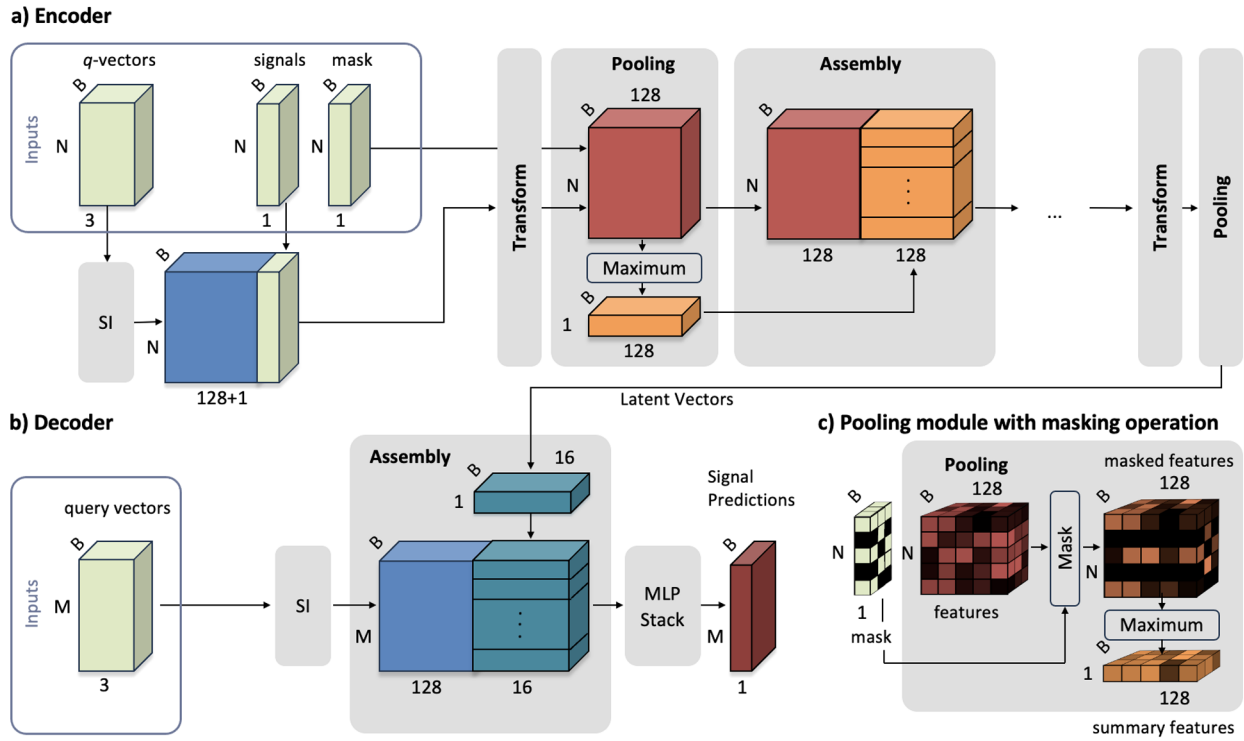
### A.2. Errors in grey and white matter

All reconstruction methods feature larger (normalized) errors in the grey matter compared to the white matter for microstructure model-derived evaluations, such as the diffusion tensor and NODDI fit (see [Appendix Fig. B](#)). However, *DISCUS* often outperforms the other reconstruction methods and the approach without reconstruction, illustrating that the difference between GM and WM errors is likely linked to a more complex microstructure in the GM, which requires more DWIs for accurate reconstruction.

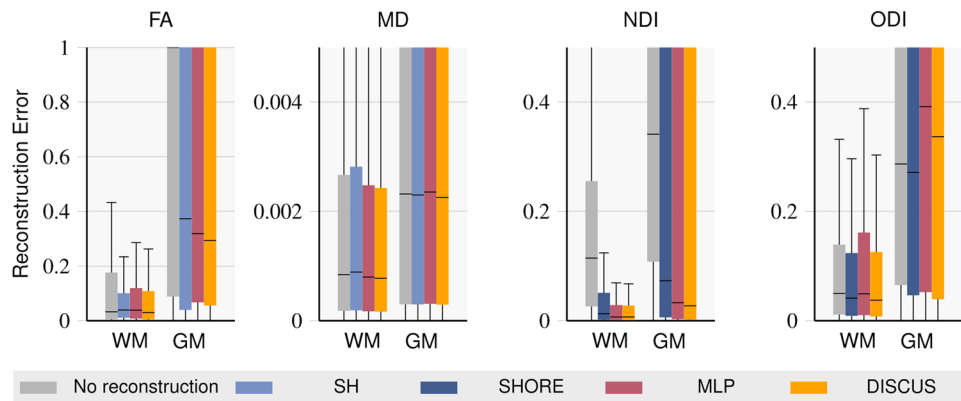
### A.3. fODF Evaluation

We evaluate angular errors between peaks of fiber orientation distribution functions (fODF) per fiber configuration, that is, in voxels containing a single fiber, two fibers crossing, and three fibers crossing, respectively. We determine these voxels on the reference data having obtained three (normalized) peaks  $p_1, p_2, p_3$  sorted by their corresponding decreasing magnitudes  $m_1, m_2, m_3$  with angular distances  $\angle(p_1, p_2), \angle(p_1, p_3), \angle(p_2, p_3)$ . We determine single fiber voxels as having a large peak

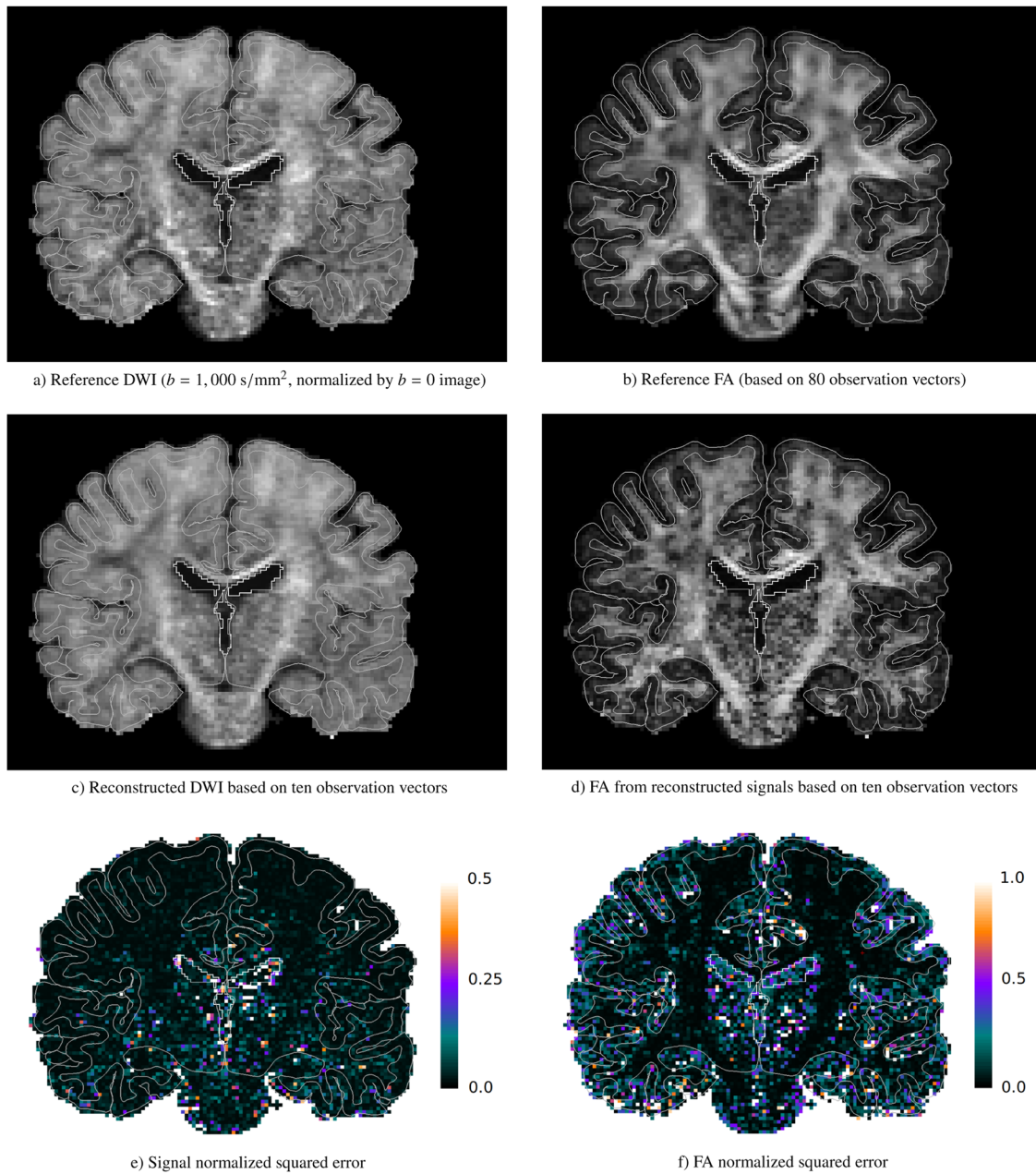




**Appendix Fig. A.** Details of the *DISCUS* architecture: (a) the encoder, (b) the decoder, and (c) the pooling module with masking operation during training. (a) The encoder receives  $q$ -vectors ( $B \times N \times 3$  with batch size  $B$  and number of observations  $N$ ), corresponding observed signals ( $B \times N \times 1$ ) and a mask indicating a random subsampling ( $B \times N \times 1$ ). At inference time, the mask may be omitted to deactivate the masking. After sign-invariant features (blue) are obtained from the  $q$ -vectors by the SI module, the signals are concatenated, and each feature vector of length  $128+1$  is processed independently by the MLP in the first transform module, resulting in the transform module output (red). Before pooling, the mask is applied to make sure only features for the desired subset of paths are summarized in the pooling operation (see c). Relevant features are summarized (orange), and per-path features (red) are concatenated with summary features (orange) by the assembly module. These are the inputs to the next transform module. The summary features of the final pooling module are the latent vectors (turquoise). (b) The decoder receives  $M$  query vectors ( $B \times M \times 3$ ) as input and passes them through the SI module obtaining sign-invariant features. An assembly module concatenates the sign-invariant features with the latent vectors to be processed by the MLP into signal predictions (dark red). (c) The masking operation reduces feature values, so masked values are always the smallest in the pooling direction of the proceeding maximum operation. They can thus never contribute to the summarized features.



**Appendix Fig. B.** White and grey matter normalized squared errors of the microstructure models derived from reconstructed or sub-sampled data, respectively. While we aggregate errors across grey (GM) and white matter (WM) in our main evaluations, this figure shows the errors per tissue for diffusion tensor-derived Fractional Anisotropy (FA) and Mean Diffusivity (MD) as well as NODDI-derived Neurite Density Index (NDI) and Orientation Dispersion Index (ODI).



**Appendix Fig. C.** The reference signal and FA maps are shown in (a) and (b). Signals reconstructed with *DISCUS* from only 10 observation vectors and Fractional Anisotropy (FA) based on the reconstructed signals are displayed in (c) and (d). Finally, (e) and (f) illustrate the error maps between reference and predicted maps. Overlaid are the white matter and pial surfaces (gray) as well as the outline of a merged CSF/ventricles region (white). All slices are from or based on the data of the Human Connectome Project (HCP).

( $m_1 > 0.5$ ) and two small peaks ( $m_2 < 0.2 \cdot m_1$ ). Voxels with two-crossings are determined as those with two large peaks ( $m_1 > 0.3$ ,  $m_2 > 0.3$ ) and a small third peak ( $m_3 < 0.1 \cdot m_2$ ) in addition to a large angular distance between the two large peaks ( $\angle(p_1, p_2) > 60^\circ$ ). Finally, we determine three-crossings as three similar peak

magnitudes ( $m_3 > 0.7 \cdot m_1$ ,  $m_3 > 0.225$ ) with pronounced angular distances between each pair of fibers ( $\angle(p_1, p_2) > 60^\circ$ ,  $\angle(p_1, p_3) > 60^\circ$ , and  $\angle(p_2, p_3) > 60^\circ$ ). We filter these sets of voxels and use only connected components of at least 20 voxels for the single fiber voxels and 10 voxels for two- and three-crossings, respectively.