Medical Image Analysis 55 (2019) 49–64

Contents lists available at ScienceDirect

Medical Image Analysis

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

A deep network for tissue microstructure estimation using modified LSTM units

Chuyang Ye[a], Xiuli Li[b,c], Jingnan Chen[d,*]

[a] School of Information and Electronics, Beijing Institute of Technology, Beijing, China
[b] Deepwise AI Lab, Beijing, China
[c] Peng Cheng Laboratory, Shenzhen, China
[d] School of Economics and Management, Beihang University, Beijing, 37 Xueyuan Road, 100191, China

ARTICLE INFO

Article history:
Received 8 August 2018
Revised 15 March 2019
Accepted 17 April 2019
Available online 18 April 2019

Keywords:
Tissue microstructure
Deep network
Sparse coding
LSTM

ABSTRACT

Diffusion magnetic resonance imaging (dMRI) offers a unique tool for noninvasively assessing tissue microstructure. However, accurate estimation of tissue microstructure described by complicated signal models can be challenging when a reduced number of diffusion gradients are used. Deep learning based microstructure estimation has recently been developed and achieved promising results. In particular, optimization-based learning, where deep network structures are constructed by unfolding the iterative processes performed for solving optimization problems, has demonstrated great potential in accurate microstructure estimation with a reduced number of diffusion gradients. In this work, using the optimization-based learning strategy, we propose a deep network structure that is motivated by the use of historical information in iterative optimization for tissue microstructure estimation, and such incorporation of historical information has not been previously explored in the design of deep networks for microstructure estimation. We assume that (1) diffusion signals can be sparsely represented by a dictionary and its coefficients jointly in the spatial and angular domain, and (2) tissue microstructure can be computed from the sparse representation. Following these assumptions, our network comprises two cascaded stages. The first stage takes image patches as input and computes the spatial-angular sparse representation of the input with learned weights. Specifically, the network structure in the first stage is constructed by unfolding an iterative process for solving sparse reconstruction problems, where historical information is incorporated. The components in this network can be shown to correspond to modified *long short-term memory* (LSTM) units. In the second stage, fully connected layers are added to compute the mapping from the sparse representation to tissue microstructure. The weights in the two stages are learned jointly by minimizing the mean squared error of microstructure estimation. Experiments were performed on dMRI scans with a reduced number of diffusion gradients. For demonstration, we evaluated the estimation of tissue microstructure described by three signal models: the *neurite orientation dispersion and density imaging* (NODDI) model, the *spherical mean technique* (SMT) model, and the *ensemble average propagator* (EAP) model. The results indicate that the proposed approach outperforms competing methods.

1. Introduction

Neurite morphology offers valuable information for understanding brain structures and functions (Conel, 1939; Jacobs et al., 1997; Fiala et al., 2002; Evangelou et al., 2000). To noninvasively quantify neurite morphology, *diffusion magnetic resonance imaging* (dMRI) has been developed as a probe of tissue microstructure. It captures the water diffusion in the neuronal tissue and is thus sensitive to tissue microstructure (Johansen-Berg and Behrens, 2013). By carefully designing signal models that relate tissue microstructure to diffusion signals, the organization of the neuronal tissue can be inferred from the observed measurements by fitting the signal models (Alexander et al., 2017a). For example, *diffusion tensor imaging* (DTI) (Basser et al., 1994) is a classic signal model for tissue microstructure quantification, and it has been successfully used to identify abnormalities of brain tissue (Barnea-Goraly et al., 2004; Werring et al., 1999; Skranes et al., 2007; Lipton et al., 2009).

Although DTI is sensitive to subtle alterations of brain tissue, it is oversimplified and the DTI measures—such as

* Corresponding author.
E-mail addresses: chuyang.ye@bit.edu.cn (C. Ye), jchen@buaa.edu.cn (J. Chen).

https://doi.org/10.1016/j.media.2019.04.006
1361-8415/

fractional anisotropy and mean diffusivity—are biologically non-specific (Pasternak et al., 2018). For example, different biological processes can result in similar changes in DTI measures (Pasternak et al., 2018). To improve the specificity of dMRI, more advanced signal models have been proposed, such as the *neurite orientation dispersion and density imaging* (NODDI) model (Zhang et al., 2012), the *spherical mean technique* (SMT) model (Kaden et al., 2016b), and the *ensemble average propagator* (EAP) model (Özarslan et al., 2013; Merlet and Deriche, 2013; Ning et al., 2015). These models not only increase the sensitivity of dMRI to brain tissue changes, but also provide measures that are specific markers of tissue microstructure. In particular, the rotationally invariant scalar measures given by the NODDI model have been applied to quite a number of neuroscientific studies (Kamagata et al., 2016; Colgan et al., 2016; Crombe et al., 2018; Winston et al., 2014; Lemkaddem et al., 2014; By et al., 2017; Genc et al., 2017; Genç et al., 2018; Batalle et al., 2017, 2019; Kelly et al., 2016; Jelescu et al., 2015; Kunz et al., 2014; Wu et al., 2018; Ocklenburg et al., 2018; Parker et al., 2018), where the importance of these scalar measures of tissue microstructure is demonstrated.

Due to the complicated model design, the advanced signal models beyond DTI usually require prolonged imaging protocols with a large number of diffusion gradients. For example, 90 diffusion gradients are recommended for the NODDI model (Zhang et al., 2012), and over 100 diffusion gradients have been used for the SMT model (Kaden et al., 2016a, 2016b). When a reduced number of diffusion gradients are used, e.g., in clinical dMRI scans, accurate estimation of tissue microstructure described by the advanced signal models can be challenging (Pasternak et al., 2018; Jones et al., 2018; McKinnon et al., 2018). Therefore, the development of reliable tissue microstructure estimation algorithms for dMRI scans with a reduced number of diffusion gradients is desired for these signal models, and it could further boost the use of the specific tissue biomarkers in neuroscientific studies.

To improve the quality of microstructure estimation on dMRI scans acquired with a reduced number of diffusion gradients, deep learning based approaches have been proposed (Golkov et al., 2016; Ye, 2017a, 2017b), where the observed diffusion signals are mapped directly to scalar tissue microstructure measures. Since diffusion signals correspond to measurements in the q-space, these methods are also referred to as q-space learning. In q-space learning, high-quality microstructure estimation is obtained from training diffusion signals densely sampled in the q-space, i.e., acquired with a large number of diffusion gradients, and the mapping from observed diffusion signals undersampled in the q-space to high-quality microstructure estimation is learned. The learned mapping can then be applied to estimate tissue microstructure on test dMRI scans. In Golkov et al. (2016), a *multiple layer perceptron* (MLP) is trained to estimate tissue microstructure, and the method is tested for the NODDI model (Zhang et al., 2012) and the kurtosis model (Hui et al., 2008). Ye (2017a) proposes a deep network specifically for the estimation of NODDI microstructure using an optimization-based learning strategy (Domke, 2012; Hershey et al., 2014). In optimization-based learning, deep networks are constructed by unfolding conventional iterative optimization processes, and such a strategy has been successfully applied for solving regularized least squares problems (Gregor and LeCun, 2010; Wang et al., 2016a; Xin et al., 2016; Sprechmann et al., 2015; Yang et al., 2016; Wang et al., 2016b; Zhang and Ghanem, 2018). Specifically, the network structure in Ye (2017a) is motivated by the linearization framework in Daducci et al. (2015), where diffusion signals in the q-space—also referred to as the angular domain (Schwab et al., 2016, 2018b)—can be sparsely represented with a dictionary. The deep network first uses a stage that unfolds the *iterative hard thresholding* (IHT) process (Blumensath and Davies, 2008) to compute the sparse representation with learned weights; then, a second stage motivated by the computation in Daducci et al. (2015) is used to map the sparse representation to NODDI microstructure with learned linear weighting. The method in Ye (2017a) performs end-to-end NODDI microstructure estimation, and it has achieved better accuracy than the MLP in Golkov et al. (2016). Ye (2017b) improves the network in Ye (2017a) by considering spatial information, where the improved network takes image patches as input and inserts a learned smoothing stage before the network structure proposed in Ye (2017a). The works in Ye (2017a) and Ye (2017b) demonstrate the great potential of optimization-based learning in accurate microstructure estimation when a reduced number of diffusion gradients are applied.

In this work, we continue to explore the design of deep networks for microstructure estimation using the optimization-based learning strategy. In the unfolded IHT process in MEDN and MEDN+, the current estimates of sparse codes are computed from the estimates in the previous iteration with a fixed schedule. In iterative optimization, strategies with fixed updates based on the output of the previous iteration can be improved by adaptively incorporating historical information, which refers to the information before the output of the previous iteration. For example, stochastic gradient descent can be improved by exploiting the information before the output of the previous iteration (Qian, 1999), and this information can be used adaptively so that different parameters can have different update rates (Duchi et al., 2011). The adaptive use of historical information can also be integrated into optimization-based learning. For example, in Zhou et al. (2018) a deep network that adaptively updates the sparse codes with historical information has been developed to improve learning-based sparse reconstruction. Therefore, motivated by these works, especially Zhou et al. (2018), we seek to adaptively incorporate historical information into the deep network for microstructure estimation, which has not been previously explored.

Previous deep networks for tissue microstructure estimation have only explored the sparsity of diffusion signals in the angular domain (Ye, 2017a, 2017b). Since diffusion signals can be sparsely represented jointly in the spatial and angular domain, where the signal redundancy in the two domains can be better exploited (Schwab et al., 2016, 2018b), our network design explores the sparse coding of diffusion signals in both the spatial and angular domain. Also, since it is common to compute tissue microstructure from the sparse representation of diffusion signals (Merlet and Deriche, 2013; Daducci et al., 2015), we assume that there exists a mapping from the spatial-angular sparse representation to tissue microstructure.

We focus on the estimation of scalar microstructure measures like in previous works (Golkov et al., 2016; Ye, 2017a, 2017b), and we aim at designing a generic deep network structure that is not limited to a specific signal model—e.g., the NODDI model like in Ye (2017a) and Ye (2017b). The proposed network is named MESC-Net[1], where MESC stands for Microstructure Estimation with Sparse Coding.

MESC-Net comprises two cascaded stages. The first stage takes image patches as input and estimates the spatial-angular sparse representation of diffusion signals. The second stage maps the sparse representation to tissue microstructure. Unlike in Ye (2017a) and Ye (2017b), the first stage in our network is inspired by the adaptive use of historical information in the iterative update for solving optimization problems. In our design, an adaptive updating process that incorporates historical information into IHT (Blumensath and Davies, 2008) is unfolded, and the

[1] A demo Python script of MESC-Net is provided at https://github.com/PkuClosed/MESC-Net.

component of the network can be shown to correspond to a variant of the *long short-term memory* (LSTM) unit (Hochreiter and Schmidhuber, 1997). In the second stage, since we seek to develop a method for generic microstructure estimation, unlike in Ye (2017a) and Ye (2017b), the form of the mapping from the sparse representation to tissue microstructure may be unknown. Thus, by analogy to convolutional neural networks, fully connected layers—which are capable of approximating continuous functions on compact subsets of $\mathbb{R}^n$ (Hornik, 1991; Sonoda and Murata, 2017)—are used to compute the final microstructure from the output of the first stage. The weights in the two stages are learned jointly from training data. To evaluate MESC-Net, experiments were performed on dMRI scans with a reduced number of diffusion gradients. For demonstration, we evaluated the estimation of tissue microstructure described by three typical signal models, which are the NODDI model (Zhang et al., 2012), the SMT model (Kaden et al., 2016b), and the EAP model using the SHORE basis (Merlet and Deriche, 2013).

The rest of the paper is organized as follows. In Section 2, we describe the design of the proposed network for microstructure estimation, as well as its training and evaluation strategy. In Section 3, experiments on brain dMRI scans are presented. In Section 4, discussion on the results and future work is given. Section 5 concludes the paper.

2. Methods

In this section, we first briefly review the background of sparse coding of diffusion signals. Then, we present our deep network for microstructure estimation. Finally, we describe the training and evaluation strategy.

2.1. Background: Sparse coding of diffusion signals

By applying a set of diffusion gradients, we can acquire a vector of diffusion signals at each voxel. Each entry in the signal vector is associated with a diffusion gradient. For convenience, the diffusion signals in this work refer to the diffusion signals that are normalized by the baseline $b0$ signal, which is acquired without diffusion weighting. With advanced signal models relating specific markers of tissue organization with diffusion signals, information about tissue microstructure can be inferred from the observed signals by fitting the models (Zhang et al., 2012; Özarslan et al., 2013; Kaden et al., 2016b). Because these advanced signal models are complicated and/or involve quite a number of free parameters, the model fitting may be sensitive to noise (McKinnon et al., 2018). Thus, a large number of diffusion gradients are usually applied for these models, and reliable tissue microstructure estimation on dMRI scans acquired with a reduced number of diffusion gradients—e.g., clinical scans—can be challenging (Pasternak et al., 2018).

Sparse coding provides a useful tool for improving the estimation of tissue organization when the number of diffusion gradients—i.e., the length of the diffusion signal vector—is reduced (Daducci et al., 2015; Merlet and Deriche, 2013). By designing a proper dictionary $\mathbf{G}$, at each voxel the diffusion signals $\boldsymbol{y}$ can be represented in a linear form as

$$\boldsymbol{y} = \mathbf{G}\boldsymbol{f}, \tag{1}$$

where $\boldsymbol{f}$ is a sparse vector of dictionary coefficients. Then, $\boldsymbol{f}$ can be estimated by solving least squares problems with sparse regularization. According to the dictionary design, the computation of microstructure from the dictionary coefficients is also determined (Daducci et al., 2015; Merlet and Deriche, 2013; Özarslan et al., 2013). For convenience, we follow the terminology in Schwab et al. (2016) and refer to Eq. (1) as the angular sparse representation of diffusion signals.

Eq. (1)models the diffusion signals at each voxel separately. It is shown in Schwab et al. (2016) that diffusion signals in the spatial domain can also be sparsely coded, which reduces the redundancy in both the spatial and angular domain. For example, suppose the vector of concatenated diffusion signals in an image or a patch is denoted by $\tilde{\boldsymbol{y}}$; then, with a properly designed dictionary $\boldsymbol{\Phi}$ that encodes the information in the spatial and angular domain, it is possible to represent $\tilde{\boldsymbol{y}}$ with a sparse vector $\boldsymbol{x}$ of dictionary coefficients:

$$\tilde{\boldsymbol{y}} = \boldsymbol{\Phi}\boldsymbol{x}. \tag{2}$$

In Schwab et al. (2016), a dictionary $\boldsymbol{\Phi}$ is designed for the estimation of orientation distribution functions, where the problem is solved via a Kronecker dual alternating direction method of multipliers algorithm. However, the design of the dictionary $\boldsymbol{\Phi}$ for tissue microstructure estimation and the subsequent problem solving are not available or straightforward. In addition, for tissue microstructure described by different signal models, different dictionaries may be needed, which increases the difficulty in dictionary design.

2.2. A generic deep network for microstructure estimation using modified LSTM units

In this work, we exploit the spatial-angular sparse coding in Eq. (2), and propose a deep learning based approach to tissue microstructure estimation without the need of hand-crafted dictionary designs. Like in previous works (Golkov et al., 2016; Ye, 2017a, 2017b), we focus on the estimation of scalar tissue microstructure measures, which have provided informative biomarkers in a variety of neuroscientific studies (Genc et al., 2017; Genç et al., 2018; Batalle et al., 2017; 2019; Wu et al., 2018). Note that unlike Ye (2017a) and Ye (2017b), we aim at developing a deep network structure for generic microstructure estimation that is not limited to a specific signal model.

We design the deep network using the optimization-based learning strategy (Domke, 2012; Hershey et al., 2014), which has achieved promising results for NODDI microstructure estimation in Ye (2017a) and Ye (2017b). Specifically, iterative processes for estimating sparse vectors can be designed and unfolded for constructing the network, to which additional components are concatenated to compute tissue microstructure. For example, Ye (2017a) and Ye (2017b) construct the deep networks by unfolding the IHT process (Blumensath and Davies, 2008), which is a nonadaptive updating strategy for solving sparse reconstruction problems, and thus in these deep networks the output of each layer is updated according to a fixed schedule. Adaptive incorporation of historical information in iterative optimization processes can improve converge performance compared with optimization strategies using fixed update rules (Duchi et al., 2011), and it has been shown that a deep network unfolding the iterative process that adaptively incorporates historical information can lead to lower errors of sparse reconstruction (Zhou et al., 2018). Therefore, we seek to construct the deep network for microstructure estimation with adaptive incorporation of historical information, and we will show that such incorporation corresponds to modified LSTM units.

Like in Ye (2017b), we consider the diffusion signals $\tilde{\boldsymbol{y}}$ in a $3 \times 3 \times 3$ image patch and predict the tissue microstructure $\boldsymbol{z}$ at the center voxel. Note that $\boldsymbol{z}$ is a vector containing multiple scalar microstructure quantities described by a signal model. For example, $\boldsymbol{z}$ includes the intra-cellular volume fraction, cerebrospinal fluid volume fraction, and orientation dispersion for the NODDI model. Similar to Ye (2017a), the proposed network—i.e.,

MESC-Net—comprises two cascaded stages, where the first stage computes the spatial-angular sparse representation of diffusion signals and the second stage maps the sparse representation to tissue microstructure. Note that here the separation of the two stages is only functional, and the parameters in the two stages are jointly learned (as later described in Section 2.3). The network architecture is described in detail as follows.

2.2.1. Stage one: spatial-angular sparse representation

We assume that the diffusion signals in a patch can be sparsely represented jointly in the spatial and angular domain. In other words, for diffusion signals $\tilde{\boldsymbol{y}}$, there exists a dictionary $\boldsymbol{\Phi}$ and dictionary coefficients $\boldsymbol{x}$ that satisfy Eq. (2). Conventionally, with a known $\boldsymbol{\Phi}$, the coefficients $\boldsymbol{x}$ in Eq. (2) can be estimated by solving an ℓ_1-norm regularized least squares problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} ||\boldsymbol{\Phi}\boldsymbol{x} - \tilde{\boldsymbol{y}}||_2^2 + \beta||\boldsymbol{x}||_1, \tag{3}$$

where β is a tunable parameter adjusting the sparsity level of $\boldsymbol{x}$. Eq. (3) can be solved iteratively using IHT (Blumensath and Davies, 2008), where $\boldsymbol{x}$ is updated as follows

$$\boldsymbol{x}^t = h_\lambda(\mathbf{W}\tilde{\boldsymbol{y}} + \mathbf{S}\boldsymbol{x}^{t-1}). \tag{4}$$

Here, t is the iteration index, $\mathbf{W}$ and $\mathbf{S}$ are matrices determined by the dictionary $\boldsymbol{\Phi}$, and $h_\lambda(\cdot)$ is a thresholding function with a parameter λ (related to β)

$$[h_\lambda(\boldsymbol{a})]_i = \begin{cases} 0 & \text{if } |a_i| < \lambda \\ a_i & \text{if } |a_i| \geq \lambda \end{cases}, \tag{5}$$

where $\boldsymbol{a}$ denotes the input of $h_\lambda(\cdot)$. Note that λ implicitly determines the sparsity of $\boldsymbol{x}^t$, where greater λ leads to sparser $\boldsymbol{x}^t$.

The iterative process in Eq. (4) can be unfolded to construct a deep network for computing the sparse representation, where $\boldsymbol{W}$ and $\boldsymbol{S}$ are learned instead of predetermined and are shared among layers, and t becomes the layer index. In Ye (2017a) and Ye (2017b), the angular sparse representation of diffusion signals is computed using such a strategy. It has been demonstrated that, compared with the conventional IHT approach, these learned layer-wise fixed weights could guarantee successful reconstruction across a wider range of *restricted isometry property* (RIP) conditions (Xin et al., 2016).

The IHT process in Eq. (4) is a nonadaptive updating strategy, where the estimates are updated based on the output of the previous iteration according to a fixed schedule. Consequently, the network structure motivated by IHT also performs nonadaptive updates of the estimates using the output of the previous layer. Since it has been shown that adaptive incorporation of historical information in the updates of estimates could facilitate optimization-based learning for sparse reconstruction problems (Zhou et al., 2018), we incorporate historical information adaptively into the IHT update process in our network design. Specifically, at layer t we introduce intermediate terms $\boldsymbol{c}$ and $\tilde{\boldsymbol{c}}$ in the update as follows (Zhou et al., 2018)

$$\tilde{\boldsymbol{c}}^t = \mathbf{W}\tilde{\boldsymbol{y}} + \mathbf{S}\boldsymbol{x}^{t-1}, \tag{6}$$

$$\boldsymbol{c}^t = \boldsymbol{f}^t \circ \boldsymbol{c}^{t-1} + \boldsymbol{i}^t \circ \tilde{\boldsymbol{c}}^t, \tag{7}$$

$$\boldsymbol{x}^t = h_\lambda(\boldsymbol{c}^t), \tag{8}$$

where $\boldsymbol{f}^t$ and $\boldsymbol{i}^t$ are weight vectors for the information in the previous and current layer, respectively, and $\circ$ is an operator representing the Hadamard product (element-wise multiplication). In this way, the information $\boldsymbol{c}^{t-1}$, which encodes the outputs of the layers before $t-1$, and the information $\tilde{\boldsymbol{c}}^t$ in the current layer t are linearly combined to allow the incorporation of historical information. Such historical information can also be interpreted as momentum terms (Zhou et al., 2018). Since we seek to adaptively incorporate historical information in the update process, $\boldsymbol{f}^t$ and $\boldsymbol{i}^t$ are designed such that they are dependent on the input signal $\tilde{\boldsymbol{y}}$ and the sparse code $\boldsymbol{x}^{t-1}$ estimated by the previous layer, and this dependency is learned from data instead of hand-crafted. Specifically, motivated by the similarity between the formulation in Eqs. (6)–(8) and LSTM units, we let (Zhou et al., 2018)

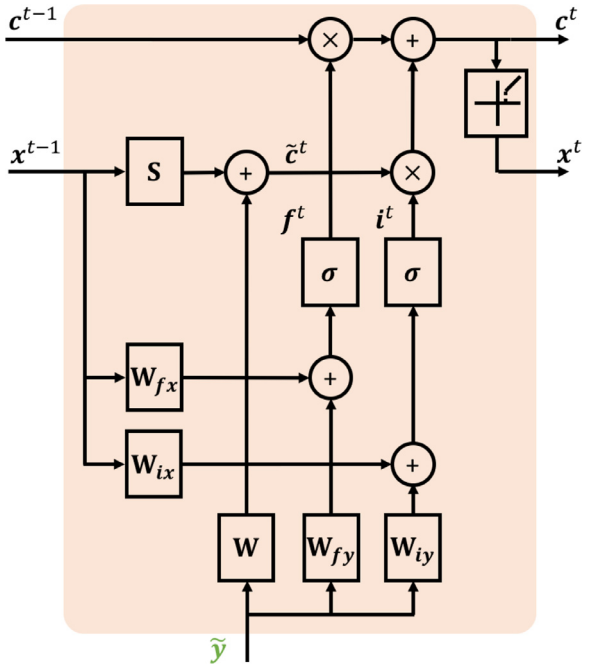

Fig. 1. The M-LSTM unit used for sparse reconstruction. The input signal $\tilde{\boldsymbol{y}}$ is indicated by the green color. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\boldsymbol{f}^t = \sigma(\mathbf{W}_{\boldsymbol{fx}}\boldsymbol{x}^{t-1} + \mathbf{W}_{\boldsymbol{fy}}\tilde{\boldsymbol{y}}), \tag{9}$$

$$\boldsymbol{i}^t = \sigma(\mathbf{W}_{\boldsymbol{ix}}\boldsymbol{x}^{t-1} + \mathbf{W}_{\boldsymbol{iy}}\tilde{\boldsymbol{y}}), \tag{10}$$

where $[\sigma(\boldsymbol{a})]_i = \frac{1}{1+e^{-a_i}}$ is a sigmoid function, $\mathbf{W}_{\boldsymbol{fx}}$ and $\mathbf{W}_{\boldsymbol{fy}}$ are matrices to be learned for computing $\boldsymbol{f}^t$, and $\mathbf{W}_{\boldsymbol{ix}}$ and $\mathbf{W}_{\boldsymbol{iy}}$ are matrices to be learned for computing $\boldsymbol{i}^t$.

One iteration of the update in Eqs. (6)–(8) with $\boldsymbol{f}^t$ and $\boldsymbol{i}^t$ defined in Eqs. (9) and (10) is equivalent to a modified LSTM unit. $\boldsymbol{i}^t$ and $\boldsymbol{f}^t$ can be viewed as an input gate and a forget gate, respectively; and unlike the standard LSTM unit, no output gate is needed. The modified LSTM (M-LSTM) unit is visualized in Fig. 1. Note that here we add a constraint $\boldsymbol{x} \geq 0$ as in Ye (2017b), and the activation $h_\lambda(\cdot)$ thus becomes a thresholded *rectified linear unit* (ReLU) (Konda et al., 2014). The threshold λ in the thresholded ReLU is set to 0.01 based on the previous observation that greater λ could lead to unstable training (Ye, 2017b). The matrices $\mathbf{W}$, $\mathbf{S}$, $\mathbf{W}_{\boldsymbol{fx}}$, $\mathbf{W}_{\boldsymbol{fy}}$, $\mathbf{W}_{\boldsymbol{ix}}$, and $\mathbf{W}_{\boldsymbol{iy}}$ are all learned from training data without requiring an explicit design of the dictionary, which will be explained later in Section 2.3. Suppose the length of $\tilde{\boldsymbol{y}}$ is M and the length of $\boldsymbol{x}$—i.e., the number of dictionary atoms—is N; we have $\mathbf{W}, \mathbf{W}_{\boldsymbol{fy}}, \mathbf{W}_{\boldsymbol{iy}} \in \mathbb{R}^{N\times M}$ and $\mathbf{S}, \mathbf{W}_{\boldsymbol{fx}}, \mathbf{W}_{\boldsymbol{ix}} \in \mathbb{R}^{N\times N}$. Here, the size of the dictionary is set according to Ye (2017a) and Ye (2017b), where $N = 301$, and M is determined by the input data.

By unfolding the iterative process in Eqs. (6)–(8), we can construct a deep network shown as Stage One in Fig. 2. Like in Ye (2017b), eight layers ($t = 1, \ldots, 8$) are used in Stage One. Note

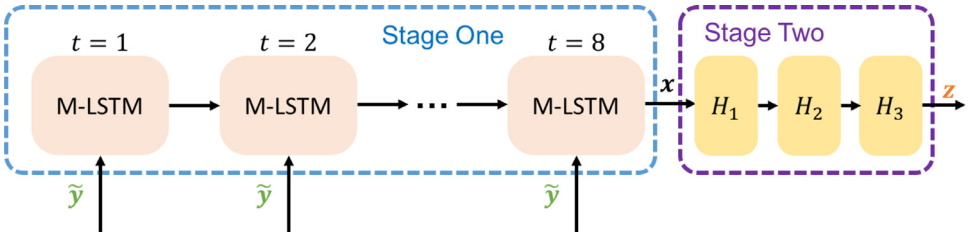

Fig. 2. The proposed deep network structure for generic tissue microstructure estimation using the M-LSTM units shown in Fig. 1 and fully connected layers (H_1, H_2, and H_3). The network input and output are indicated by the green and orange colors, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that the matrices $\mathbf{W}$, $\mathbf{S}$, $\mathbf{W}_{fx}$, $\mathbf{W}_{fy}$, $\mathbf{W}_{ix}$, and $\mathbf{W}_{iy}$ are shared among layers, and we initialize $\boldsymbol{c}^0 = \mathbf{0}$ and $\boldsymbol{x}^0 = \mathbf{0}$.

2.2.2. Stage two: mapping from sparse representation to tissue microstructure

Then, the mapping from dictionary coefficients $\boldsymbol{x}$ to tissue microstructure $\boldsymbol{z}$ should be determined. Because the dictionary is not explicitly designed in Stage One, the form of the mapping is unknown. By analogy to convolutional neural networks, where fully connected layers map the features extracted by convolutional layers to the final prediction, we use three fully connected layers, H_1, H_2, and H_3, for the estimation of microstructure (see Stage Two in Fig. 2). Each H_i $(i = 1, 2, 3)$ computes the output from its input $\boldsymbol{u}_i$ as follows

$$H_i(\boldsymbol{u}_i) = \sigma_H(\mathbf{W}_i\boldsymbol{u}_i + \boldsymbol{b}_i), \tag{11}$$

where $\mathbf{W}_i$ and $\boldsymbol{b}_i$ are the weights and bias terms to be learned for H_i, respectively, and $\sigma_H(\cdot)$ is a nonlinear ReLU activation function (Nair and Hinton, 2010), which is the same for each H_i. These layers are capable of approximating continuous functions on compact subsets of $\mathbb{R}^n$ (Hornik, 1991; Sonoda and Murata, 2017). At each of these layers 75 hidden units are used. The complete network for tissue microstructure estimation is now available as shown in Fig. 2, where the input and output are indicated by the green and orange colors, respectively.

2.3. Training and evaluation

The weights in the two stages are jointly learned by minimizing the mean squared estimation error of the microstructure vector $\boldsymbol{z}$. In this way, the dictionary can be implicitly determined from the training microstructure without an explicit dictionary design. Note that $\boldsymbol{z}$ is a vector containing multiple microstructure quantities of interest, and its entries can have different units and thus different magnitude scales. Therefore, before training we first rescale each entry z_i in $\boldsymbol{z}$: $z_i \rightarrow z_i/10^{\lfloor \log_{10} \bar{z}_i \rfloor}$, where $\bar{z}_i$ is the mean z_i value of the training data and $\lfloor \cdot \rfloor$ is a floor operator. In this way, $\boldsymbol{z}$ is rescaled so that the mean of each z_i is between 1 and 10. All $\bar{z}_i$'s are recorded, and after applying the trained network to test samples, $\bar{z}_i$'s are used to scale back the microstructure estimates for the test samples. Like in Ye (2017b), we used the Adam algorithm (Kingma and Ba, 2014) for optimization, where the learning rate was 0.0001, the batch size was 128, and 10 epochs were used. Also, 10% of the training samples were used as a validation set to prevent overfitting (Golkov et al., 2016; Ye, 2017b). The network is implemented using Keras[2] with a Theano[3] backend.

For each dMRI dataset where scans share a set $\mathcal{G}$ of diffusion gradients, one deep network should be trained for predicting the microstructure described by a particular signal model (Ye, 2017b). High-quality training dMRI scans should be acquired with a set $\tilde{\mathcal{G}}$ $(\mathcal{G} \subset \tilde{\mathcal{G}})$ of diffusion gradients that densely sample the q-space. From these training scans, training microstructure can be computed using conventional approaches that are based on model fitting (Golkov et al., 2016; Ye, 2017b). Then, the patches of undersampled diffusion signals corresponding to $\mathcal{G}$ from the training scans and the training microstructure at the center voxel of the patches can be used to train the deep network.

To quantitatively evaluate the estimation performance, we computed gold standard microstructure in a way similar to the generation of training samples. For each test dMRI scan with diffusion gradients $\mathcal{G}$, diffusion gradients $\tilde{\mathcal{G}}$ densely sampling the q-space were also applied. From the signals associated with $\tilde{\mathcal{G}}$, gold standard tissue microstructure was computed using conventional approaches. The mean absolute difference between the estimated microstructure and the gold standard in the brain was computed to measure the estimation accuracy.

3. Results

In this section, we first describe the data used for evaluation, and then present experimental results on the evaluation of the proposed MESC-Net. In the evaluation, MESC-Net was first applied to the estimation of microstructure described by the NODDI model (Zhang et al., 2012), which is a biophysical model that has been widely used for brain studies (Pasternak et al., 2018). Then, it was applied for additional signal models, including the SMT model (Kaden et al., 2016b) and the EAP model using the SHORE basis (Merlet and Deriche, 2013). In the experiments, MESC-Net was evaluated qualitatively and quantitatively, and compared with conventional and deep learning based methods. All experiments were performed on a 24-core Linux machine with an NVIDIA TITAN Xp GPU.

3.1. Data description

We selected 25 subjects from the *Human Connectome Project* (HCP) dataset (Van Essen et al., 2013), which is publicly available. The dMRI scans were acquired with three shells ($b = 1000$, 2000, and 3000 s/mm^2), and each shell has 90 gradient directions. The image resolution is 1.25 mm isotropic. For each dMRI scan, 60 fixed diffusion gradients were selected as the diffusion gradient set $\mathcal{G}$ that undersamples the q-space. These diffusion gradients consist of 30 gradient directions on each of the shells $b = 1000, 2000$ s/mm^2, and resemble diffusion gradients that are clinically feasible. The 30 gradient directions on each shell were selected using the following strategy. We selected a clinical imaging protocol that includes 30 gradient directions, and these 30 gradient directions were used as reference directions. Then, for each reference direction, the gradient direction that is closest to it was selected from the HCP data on each of the shells $b = 1000, 2000$ s/mm^2, and these gradient directions were

[2] http://keras.io/.
[3] http://deeplearning.net/software/theano/.

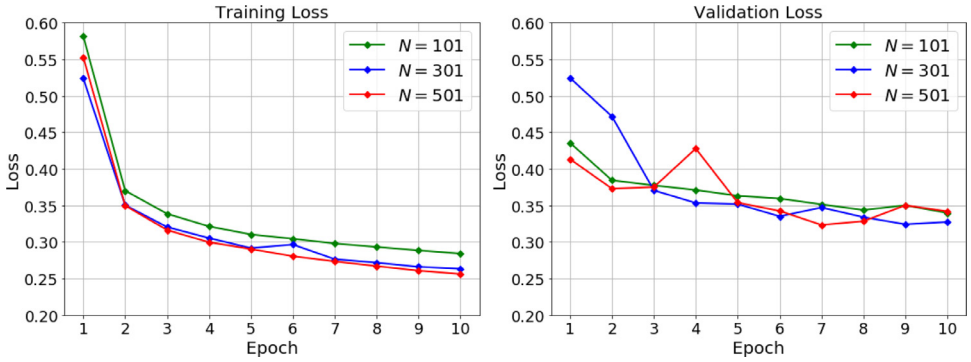

Fig. 3. The training and validation losses of MESC-Net for the NODDI model with different dictionary sizes N.

Table 1
The numbers of parameters for the learning-based methods in the NODDI microstructure estimation.

MLP	MEDN	MEDN+	m-MEDN+	MESC-Net
1.64×10^5	1.10×10^5	2.07×10^5	1.78×10^6	1.77×10^6

included in the final 60 diffusion gradients. This strategy gives approximately evenly distributed gradient directions.

The dMRI scans of five subjects were used as training data, and the rest 20 dMRI scans were used for testing. The training or gold standard microstructure was computed with the full set $\tilde{\mathcal{G}}$ of all 270 diffusion gradients as described in Section 2.3. In the training or test phase, the proposed network took the patches of diffusion signals associated with $\mathcal{G}$ as input. The patches of which the center voxel is inside the brain were included in training MESC-Net. This resulted in a total number of about 3.4×10^6 training samples and 3.8×10^5 validation samples.

3.2. Application to the NODDI model

MESC-Net was first applied to estimate the scalar NODDI microstructure measures, including the intra-cellular volume fraction v_{ic}, *cerebrospinal fluid* (CSF) volume fraction v_{iso}, and *orientation dispersion* (OD) (Zhang et al., 2012). Because compared with the original NODDI fitting (Zhang et al., 2012), the AMICO algorithm (Daducci et al., 2015) reduces the computational time by two orders of magnitude, like in Ye (2017b), the training microstructure and gold standard microstructure for evaluation were computed by AMICO using the full set of diffusion gradients.

MESC-Net was compared qualitatively and quantitatively with AMICO (Daducci et al., 2015), which represents the conventional microstructure estimation algorithm, and deep learning based approaches, including the MLP used in Golkov et al. (2016), and the deep network structures MEDN (Ye, 2017a) and MEDN+ (Ye, 2017b). MLP and MEDN perform voxelwise estimation, where microstructure estimation at each voxel is based on the diffusion signals at that voxel. MLP consists of three hidden layers each with 150 units (Golkov et al., 2016), and MEDN uses a deep network structure motivated by unfolding the IHT process (Blumensath and Davies, 2008). MEDN+ extends MEDN by taking image patches as input and inserting a learned smoothing stage before the MEDN structure (Ye, 2017b). Note that a separate MLP was used for the estimation of each microstructure measure as suggested in Golkov et al. (2016).

The numbers of parameters for the learning-based methods are listed in Table 1. Note that since MESC-Net has considerably more parameters than MLP, MEDN, and MEDN+, we have also added a modified version of MEDN+ for comparison, and investigated whether the improved performance of the proposed approach is due to our network design or the increase in the number of parameters. Here, MEDN+ was selected for this modification because both MEDN+ and MSEC-Net use spatial information. In the modified MEDN+, which is referred to as m-MEDN+, its dictionary length is increased to 1266 so that its number of parameters is comparable to (slightly more than) that of MESC-Net.

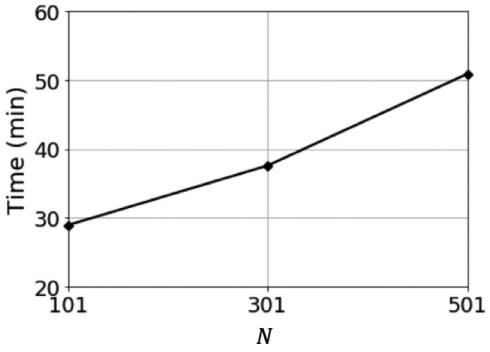

Fig. 4. The training times of MESC-Net for the NODDI model with different dictionary sizes N.

For the voxelwise learning-based methods MLP and MEDN, the diffusion signals and the training microstructure at each voxel that is inside the brain constitute a training sample; like in training MESC-Net, patches of which the center voxel is inside the brain were included in training the patch-based methods MEDN+ and m-MEDN+. We also used 10% of the samples in the training subjects as a validation set for these learning-based methods. Therefore, each learning-based method—including MESC-Net—has the same number of training and validation samples.

3.2.1. Training phase and the selection of N

We first investigated the training phase of MESC-Net, and like in Ye (2017b), we examined how the size N of the dictionary—i.e., the length of $\boldsymbol{x}$—affected training. Different values of $N \in \{101, 301, 501\}$ (Ye, 2017b) were used, and the corresponding training and validation losses after each epoch are plotted in Fig. 3. We can see that with 10 epochs the training loss becomes stable and the validation loss starts to oscillate for these cases. The training and validation losses for $N = 101$ are both higher than those for $N = 301$. Although increasing N from 301 to 501 slightly decreases the training loss, it leads to a higher validation loss after 10 epochs. The training times for these N's are shown in Fig. 4, where a larger N is associated with a longer training time. These results justify our use of $N = 301$, because further increasing N requires a longer training time and could lead to larger errors on the validation set. The parameter settings for the NODDI model will be also used for the additional signal models in Section 3.3.

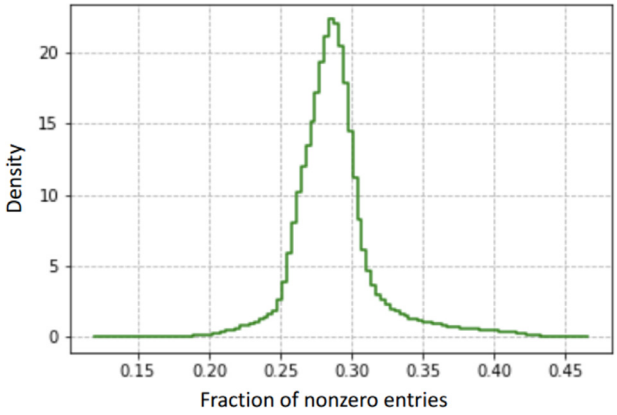

Fig. 5. The distribution of the fraction of nonzero entries in the spatial-angular sparse representation $\boldsymbol{x}$ given by MESC-Net for a representative test subject.

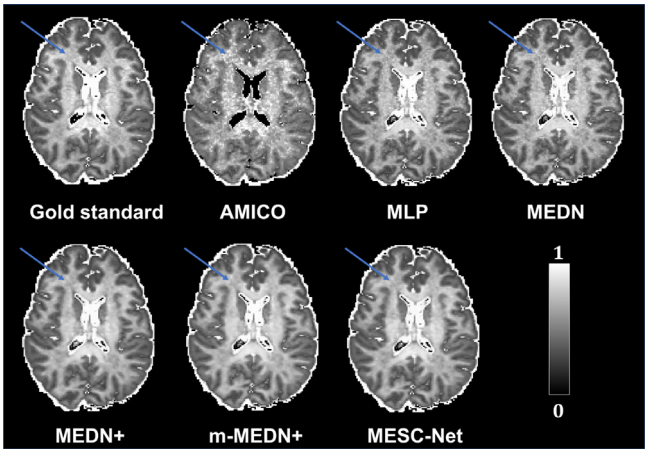

Fig. 6. A representative cross-sectional view of the v_{ic} map estimated by MESC-Net. The result of MESC-Net is shown together with the gold standard and the results of the competing methods. Note the regions pointed by arrows for comparison.

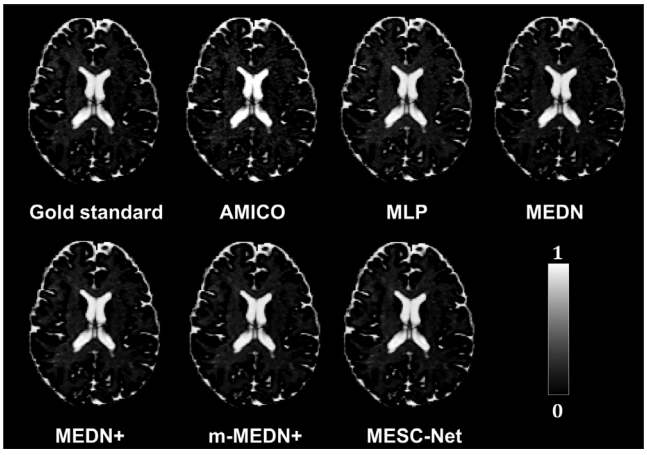

Fig. 7. A representative cross-sectional view of the v_{iso} map estimated by MESC-Net. The result of MESC-Net is shown together with the gold standard and the results of the competing methods.

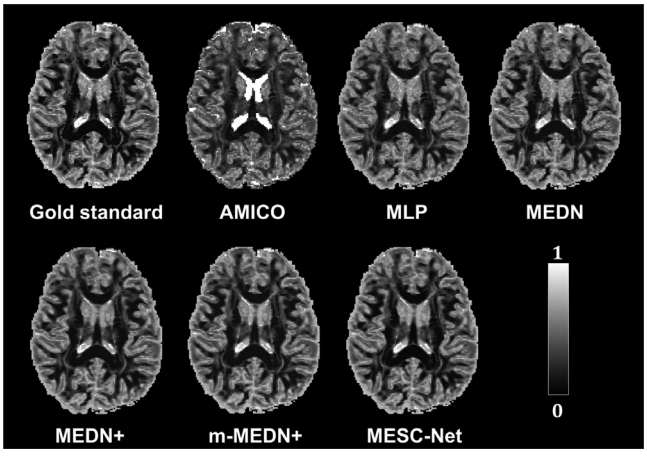

Fig. 8. A representative cross-sectional view of the OD map estimated by MESC-Net. The result of MESC-Net is shown together with the gold standard and the results of the competing methods.

With the selected $N = 301$, we also report the training cost in time and memory usage for MESC-Net. About 47GB RAM and about 330MB GPU memory were used in training MESC-Net, and the training took about 35 min. For comparison, about 4GB RAM was used for training MLP and MEDN, which perform estimation using information in a single voxel; about 47GB RAM was used for training the patch-based methods MEDN+ and m-MEDN+. All competing learning-based methods used more than 300MB and less than 400MB GPU memory. The training time was about 20 min for MLP and 10 min for MEDN; for the patch-based methods MEDN+ and m-MEDN+, the training took about 30 and 35 min, respectively.

3.2.2. Evaluation of estimation quality

MESC-Net was then applied to the test subjects for evaluation. The computation took about 3 min for each subject.

To demonstrate that MESC-Net indeed finds a sparse representation of diffusion signals, for the test subjects we computed the fraction of nonzero entries in the intermediate vector $\boldsymbol{x}$, which is the output of Stage One of MESC-Net. The distribution of the fraction is shown for a representative test subject in Fig. 5. We can see that for most patches, only about 0.3 of the entries in $\boldsymbol{x}$ are nonzero, indicating that $\boldsymbol{x}$ is a sparse representation of signals.

Then, we qualitatively examined the estimation results of MESC-Net. Cross-sectional views of the estimated microstructure maps of a representative test subject are shown in Figs. 6–8, together with the gold standard and the results of the competing methods. In the MESC-Net result, v_{ic} is higher in the white matter and smaller in the gray matter, OD is smaller in the white matter, and v_{iso} is larger in the CSF. These observations are consistent with the appearances of the gold standard maps. In addition, learning-based methods produced less noisier results than AMICO; and among the learning-based methods, patch-based ones (MEDN+, m-MEDN+, and MESC-Net) produced smoother results than voxelwise estimation (MLP and MEDN). For example, see the regions highlighted by arrows in Fig. 6. Note that in the CSF, the estimation of v_{ic} and OD is degenerate.[4] Thus, the CSF will not be included in the quantitative evaluation shown later.

Next, the accuracy of the proposed approach was evaluated quantitatively. The average estimation errors of MESC-Net in the brains of the test subjects (S01–S20) are shown and compared with those of AMICO, MLP, MEDN, MEDN+, and m-MEDN+ in Fig. 9. MESC-Net consistently achieves the smallest errors. The means and standard deviations of the average estimation errors shown in Fig. 9 are also summarized in Fig. 10, and the errors are compared between MESC-Net and the competing methods using paired Student's t-tests. Compared with the competing methods, the errors of MESC-Net are all highly significantly ($p < 0.001$) smaller, which is indicated in Fig. 10 by the asterisks. Note that m-MEDN+ has

[4] If $v_{iso} = 1$, then $1 - v_{iso} = 0$, and according to the NODDI model v_{ic} and OD can take arbitrary values without affecting the signal value.

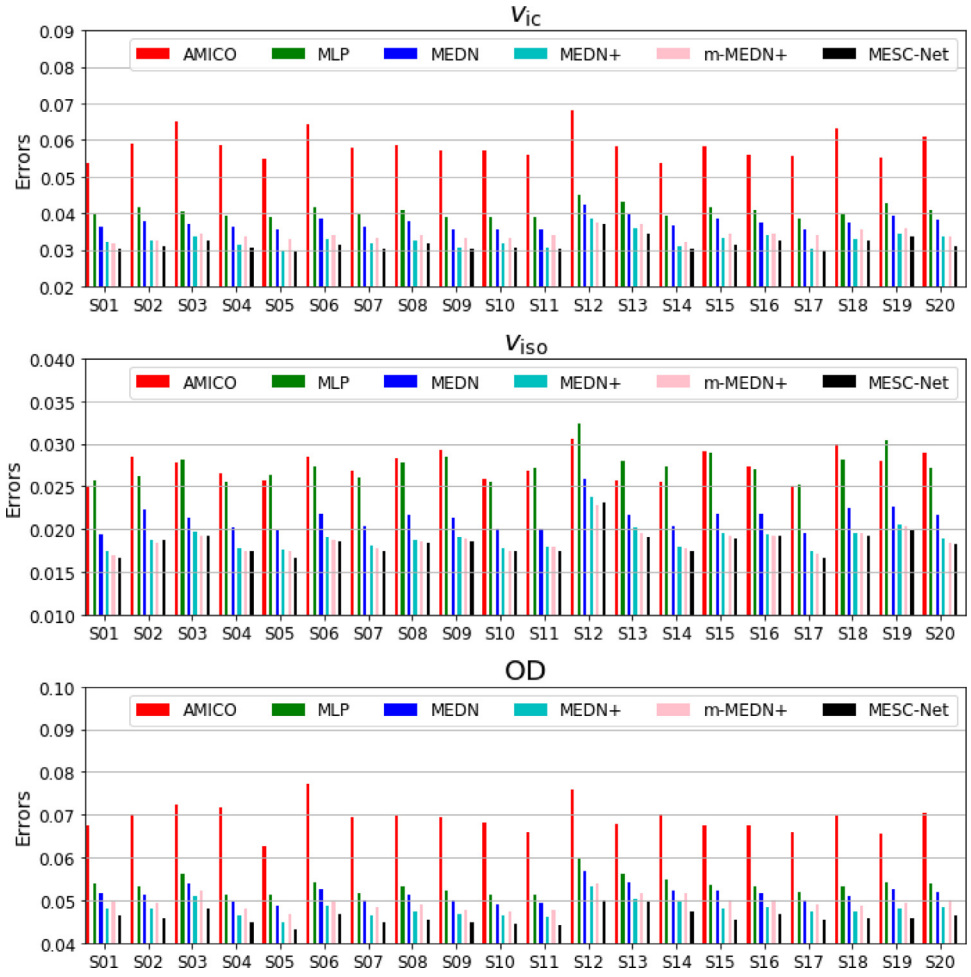

Fig. 9. Average microstructure estimation errors in the brains of the test subjects for the NODDI model.

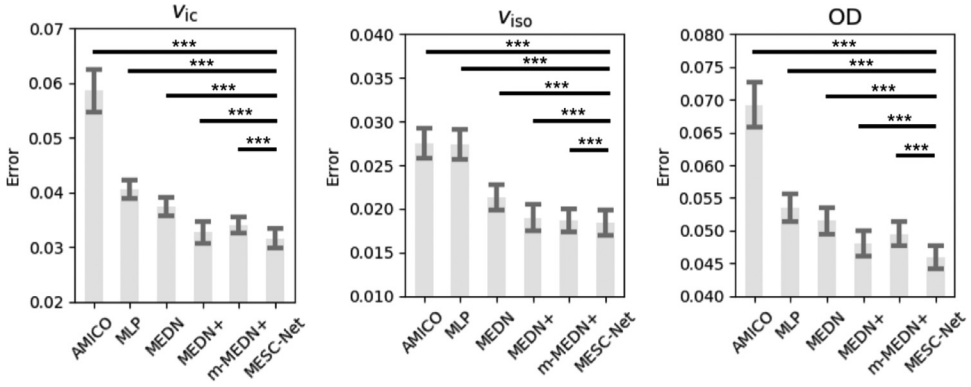

Fig. 10. Means and standard deviations of the average microstructure estimation errors shown in Fig. 9 for the NODDI model. Asterisks (***) indicate that the difference between methods is highly significant ($p < 0.001$) using a paired Student's t-test.

greater v_{ic} and OD errors than the original MEDN+, indicating that simply increasing the number of parameters does not necessarily lead to improved performance. Therefore, m-MEDN+ will not be considered in the remaining experiments.

3.2.3. Individual effects of LSTM units and spatial-angular sparse coding

We further investigated the individual effects of the incorporation of historical information using the LSTM units and the spatial-angular sparse coding in MESC-Net. Specifically, two networks, MESC-Net without spatial-angular sparse coding (MESC-Net w/o SASC) and MESC-Net without LSTM units (MESC-Net w/o LSTM), were constructed. In MESC-Net w/o SASC, the spatial-angular sparse coding is replaced with the smoothing stage in MEDN+; in MESC-Net w/o LSTM, the LSTM units are replaced by the update in MEDN+. Illustrations of the two networks are shown in Fig. 11.

We first report the training cost in time and memory for the two networks. The training settings for MESC-Net were applied to these two networks. Each of them used about 47GB RAM and 320MB GPU memory. The training took about 30 min for each network.

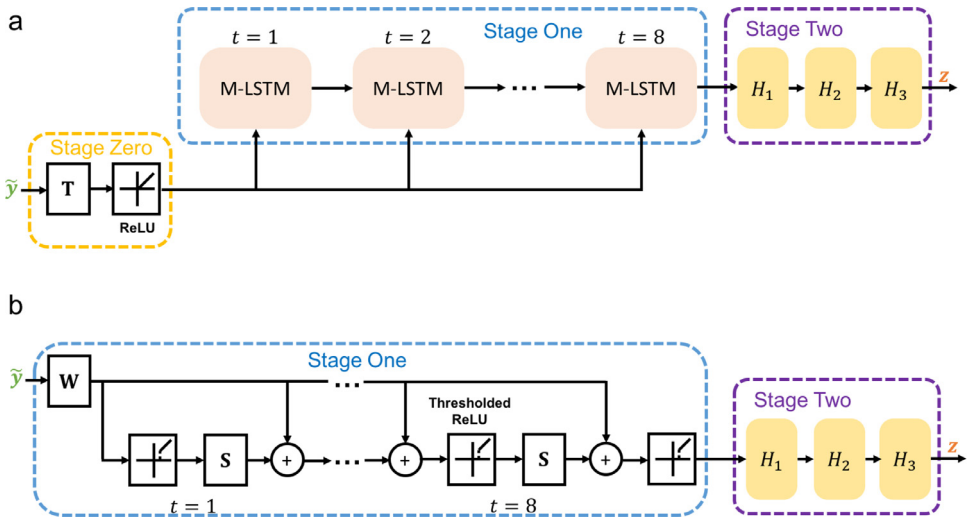

Fig. 11. The networks for investigating the individual effects of LSTM units and spatial-angular sparse coding: (a) MESC-Net without spatial-angular sparse coding (MESC-Net w/o SASC) and (b) MESC-Net without LSTM units (MESC-Net w/o LSTM).

Table 2
Means and standard deviations (in the format of mean ± standard deviation) of the average NODDI microstructure estimation errors in the brain for MESC-Net w/o SASC and MESC-Net w/o LSTM. The errors are shown together with those of MEDN+ and MESC-Net for reference.

	v_{ic}	v_{iso}	OD
MEDN+	$3.27 \times 10^{-2} \pm 0.20 \times 10^{-2}$	$1.90 \times 10^{-2} \pm 0.15 \times 10^{-2}$	$4.81 \times 10^{-2} \pm 0.20 \times 10^{-2}$
MESC-Net w/o SASC	$3.22 \times 10^{-2} \pm 0.19 \times 10^{-2}$	$1.87 \times 10^{-2} \pm 0.15 \times 10^{-2}$	$4.67 \times 10^{-2} \pm 0.19 \times 10^{-2}$
MESC-Net w/o LSTM	$3.21 \times 10^{-2} \pm 0.17 \times 10^{-2}$	$1.88 \times 10^{-2} \pm 0.14 \times 10^{-2}$	$4.80 \times 10^{-2} \pm 0.18 \times 10^{-2}$
MESC-Net	$3.16 \times 10^{-2} \pm 0.19 \times 10^{-2}$	$1.84 \times 10^{-2} \pm 0.15 \times 10^{-2}$	$4.60 \times 10^{-2} \pm 0.17 \times 10^{-2}$

The trained MESC-Net w/o SASC and MESC-Net w/o LSTM were then applied to the test subjects. The means and standard deviations of the average errors in the brain are shown in Table 2 for MESC-Net w/o SASC and MESC-Net w/o LSTM. For reference, the results of MEDN+ and MESC-Net are also shown. It can be seen that both MESC-Net w/o SASC and MESC-Net w/o LSTM perform better than MEDN+ and worse than MESC-Net. These results demonstrate the necessity of the incorporation of historical information using the LSTM units and the spatial-angular sparse coding, where both of them contribute to the improved performance in MESC-Net.

3.2.4. Impact of the number of diffusion gradients

We also studied the impact of the number of diffusion gradients by using two additional sets of undersampled diffusion gradients. Since in clinical scans, it is usually practical to use about 30 to 60 diffusion gradients, the two additional sets of diffusion gradients contain 36 diffusion gradients (18 on each of the shells $b = 1000, 2000$ s/mm^2) and 24 diffusion gradients (12 on each of the shells $b = 1000, 2000$ s/mm^2), respectively. The diffusion gradients were selected in a way that is similar to the selection of 60 diffusion gradients. Here, 18 and 12 gradient directions from two clinical protocols were used as reference directions for the two sets, respectively. All the settings used for the experiments with 60 diffusion gradients were applied to the experiments with 36 and 24 diffusion gradients.

We computed the estimation errors for MESC-Net and the competing methods, and the means and standard deviations of the average errors in the brain computed from the 20 test subjects are shown in Fig. 12. In all cases, MESC-Net achieves smaller errors than the competing methods, and the differences are highly significant ($p < 0.001$) using paired Student's t-tests. Together with the results in Section 3.2.2, it can also be observed that when a smaller number of diffusion gradients are used, the differences of errors between MESC-Net and the competing methods become larger. In addition, for each method the estimation error becomes larger when a smaller number of diffusion gradients are used.

3.3. Application to additional signal models

MESC-Net was next applied to the estimation of microstructure described by additional signal models with the parameters specified for the NODDI microstructure estimation. Two additional models were considered: the SMT model (Kaden et al., 2016b) and the EAP model using the SHORE basis (Merlet and Deriche, 2013).

3.3.1. The SMT model

In the SMT model, microstructure is quantified by the intra-neurite volume fraction v_{int} and the intrinsic diffusion coefficient λ. The training microstructure and gold standard microstructure for evaluation were computed using the SMT approach (Kaden et al., 2016b) in the DMIPY (Fick et al., 2018) software with the full set of diffusion gradients. Note that because DMIPY computes the extra-neurite volume fraction v_{ext} and λ, for convenience, we estimated v_{ext} and λ as well, and the estimation of v_{ext} is equivalent to the estimation of v_{int} because $v_{int} = 1 - v_{ext}$.

The training phase of MESC-Net was first investigated, and the training and validation losses are shown in Fig. 13. We can see that with 10 epochs the training loss becomes stable and the validation loss starts to oscillate.

Then, MESC-Net was applied to the test subjects. It was compared with the original SMT model fitting (Kaden et al., 2016b)—referred to as SMT—using the implementation given in DMIPY (Fick et al., 2018), which represents the conventional microstructure estimation algorithm, and the MLP in Golkov et al. (2016), which represents the deep learning based approach. A separate MLP was used for the estimation of each

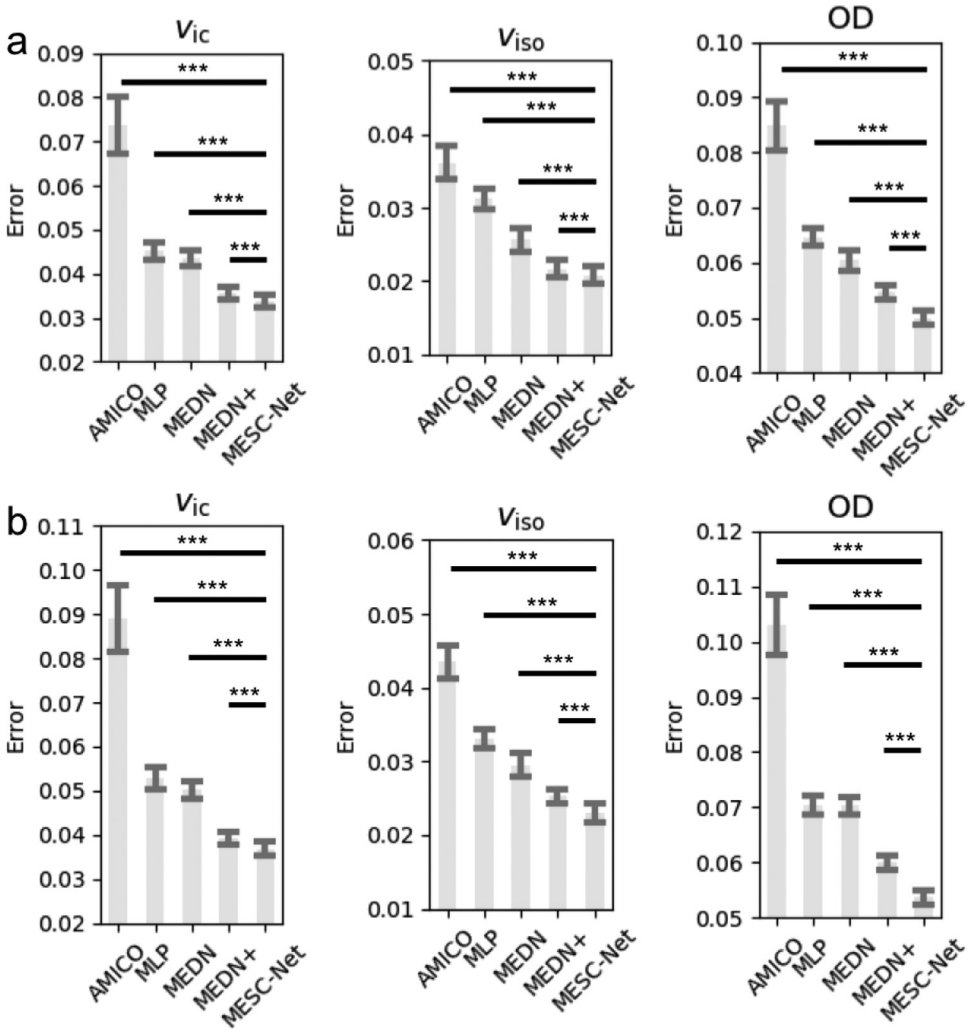

Fig. 12. Means and standard deviations of the average microstructure estimation errors in the brain for the NODDI model when (a) 36 and (b) 24 diffusion gradients were used. Asterisks (***) indicate that the difference between methods is highly significant ($p < 0.001$) using a paired Student's t-test.

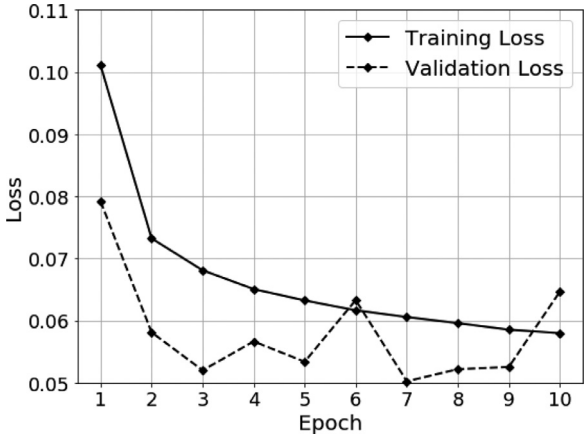

Fig. 13. The training and validation losses of MESC-Net for the SMT model.

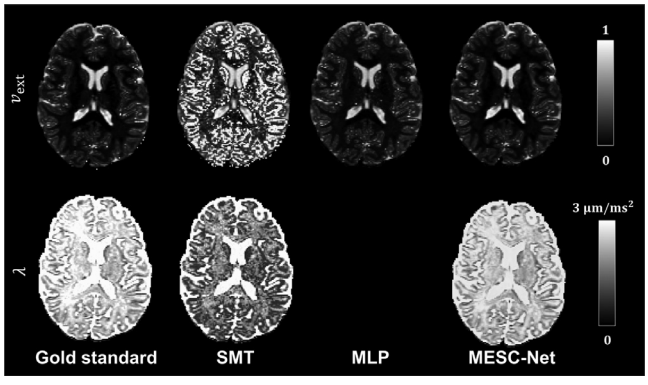

Fig. 14. Representative cross-sectional views of the SMT microstructure estimated by MESC-Net. The results of MESC-Net are shown together with the gold standard and the results of the competing methods. Note that MLP failed to estimate a sensible λ map and its result appears black with the specified grayscale mapping.

microstructure measure as suggested in Golkov et al. (2016). Note that MEDN and MEDN+ are designed specifically for the NODDI microstructure estimation and cannot be applied to the SMT model.

We first qualitatively examined the estimation results of MESC-Net, and cross-sectional views of the estimated microstructure maps of a representative test subject are shown in Fig. 14. Gold standard microstructure maps and the results of the competing methods are also shown in Fig. 14 for comparison. In the MESC-Net result, the v_{ext} and λ maps are consistent with the appearances of the gold standard maps. Compared with the gold standard, SMT tends to overestimate v_{ext}, and the λ map

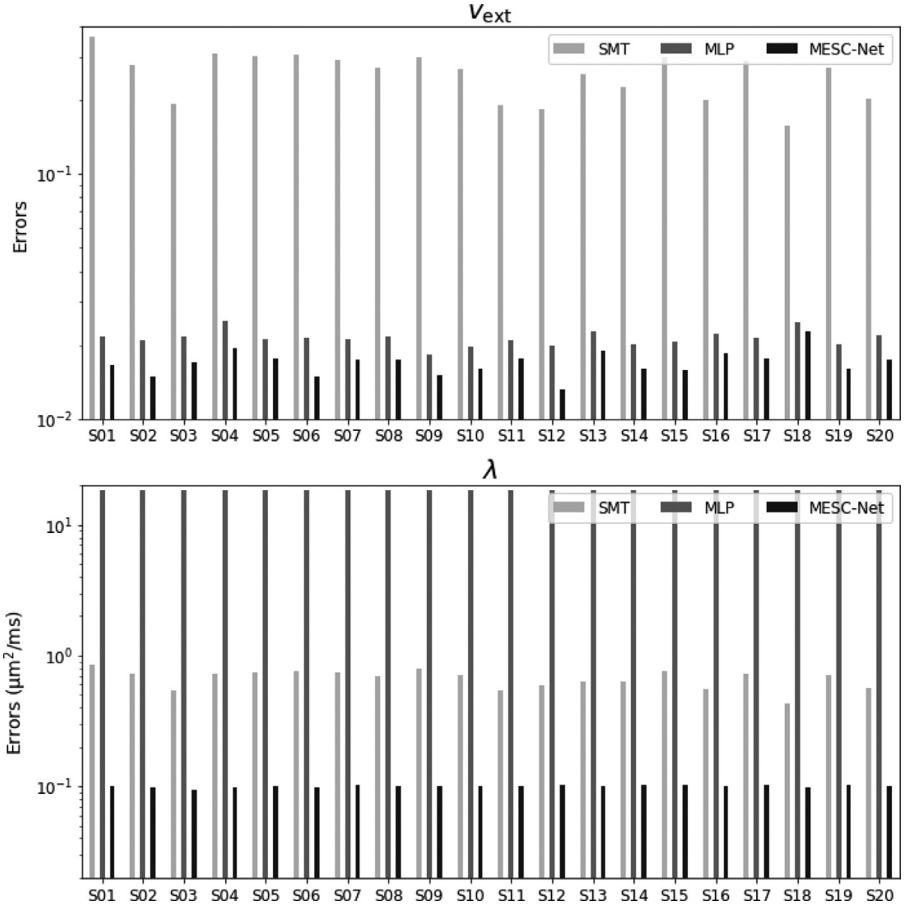

Fig. 15. Average microstructure estimation errors in the brains of the test subjects for the SMT model. Note that a log scale is used for the *y*-axis.

given by SMT has reduced intensities in the white matter. The MLP failed to estimate the λ map, where the result only has very small intensities and appears black with the given grayscale mapping.

MESC-Net was next evaluated quantitatively in terms of its accuracy. The average estimation errors of MESC-Net in the brains of the test subjects are shown and compared with those of the original SMT model fitting and MLP in Fig. 15. In all cases, MESC-Net achieves the smallest errors. The means and standard deviations of the average estimation errors shown in Fig. 15 are also summarized in Fig. 16, and the errors are compared between MESC-Net and the competing methods using paired Student's *t*-tests. Compared with the competing methods, the errors of MESC-Net are all highly significantly ($p < 0.001$) smaller, which is indicated in Fig. 16 by the asterisks. Note that the learning-based method MLP can have bigger estimation errors of λ than the conventional SMT model fitting.

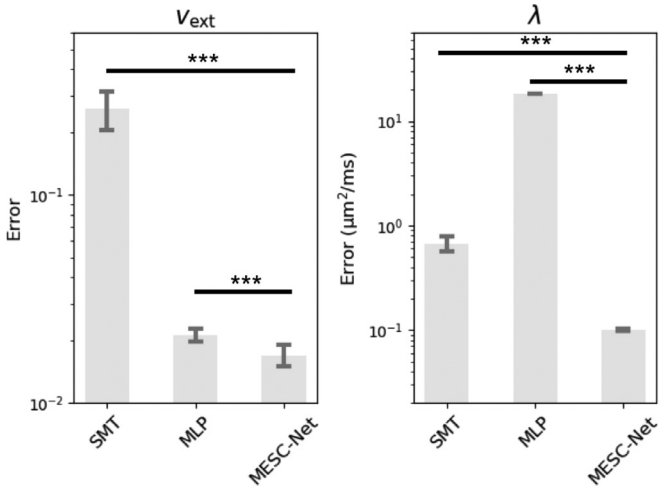

Fig. 16. Means and standard deviations of the average microstructure estimation errors shown in Fig. 15 for the SMT model. Asterisks (***) indicate that the difference between methods is highly significant ($p < 0.001$) using a paired Student's *t*-test. Note that a log scale is used for the *y*-axis.

3.3.2. The EAP model using the SHORE basis

In the EAP model using the SHORE basis, which is referred to as the SHORE EAP model for convenience in this section, we estimated the *mean squared displacement* (MSD) (Wu et al., 2008) and the *return-to-the-origin probability* (RTOP) (Özarslan et al., 2013). MSD is related to the mean diffusivity (Wu et al., 2008) and RTOP represents the reciprocal of the statistical mean pore volume (Özarslan et al., 2013). Note that in the evaluation, we evaluated the cubic root of RTOP (denoted by $\text{RTOP}^{1/3}$) like in Özarslan et al. (2013). The training microstructure and gold standard microstructure for evaluation were derived from the EAPs computed by the DIPY (Garyfallidis et al., 2014) SHORE implementation[5] using the full set of diffusion gradients.

The training phase of MESC-Net was first investigated, and the training and validation losses are shown in Fig. 17. We can see that

[5] With the default parameters $((N, \zeta, \lambda_N, \lambda_L) = (6, 700, 10^8, 10^8))$ provided at http://nipy.org/dipy/examplesbuilt/reconstshore.html.

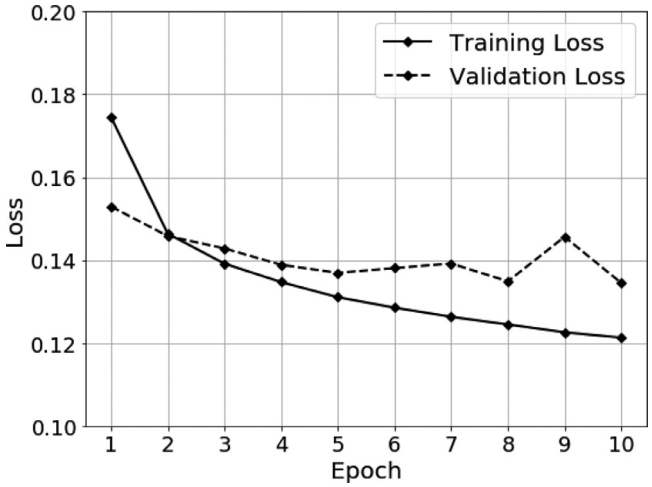

Fig. 17. The training and validation losses of MESC-Net for the SHORE EAP model.

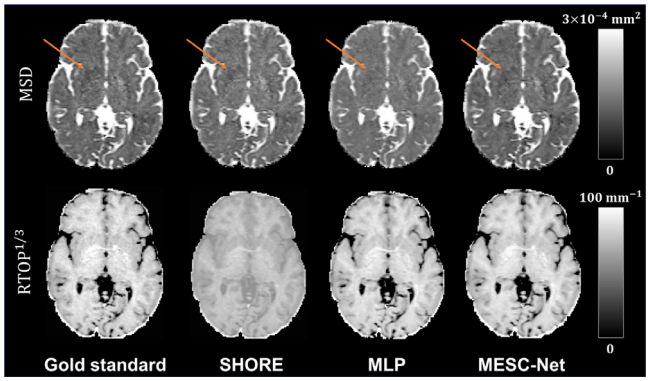

Fig. 18. Representative cross-sectional views of the microstructure estimated by MESC-Net for the SHORE EAP model. The results of MESC-Net are shown together with the gold standard and the results of the competing methods. Note the regions pointed by arrows for comparison.

with 10 epochs the training loss becomes stable and the validation loss starts to oscillate.

Then, the results estimated by MESC-Net were compared with those derived from the EAPs estimated by the conventional SHORE estimation (Merlet and Deriche, 2013)—referred to as SHORE—and those estimated by the MLP in Golkov et al. (2016). A separate MLP was used for the estimation of each microstructure measure as suggested in Golkov et al. (2016). Note that MEDN and MEDN+ are designed specifically for the NODDI microstructure estimation and cannot be applied to the SHORE EAP model.

We first qualitatively examined the estimation results of MESC-Net, and cross-sectional views of the estimated microstructure maps of a representative test subject are shown in Fig. 18. Gold standard microstructure maps and the results of the competing methods are also shown in Fig. 18 for comparison. In the MESC-Net result, the MSD and $\text{RTOP}^{1/3}$ maps are consistent with the appearances of the gold standard maps. The $\text{RTOP}^{1/3}$ map given by SHORE has a contrast that is different from the gold standard. MLP tends to produce MSD maps that are brighter than the gold standard (see the region highlighted by arrows in Fig. 18 for example).

MESC-Net was next evaluated quantitatively in terms of its accuracy. The average estimation errors of MESC-Net in the brains of the test subjects are shown and compared with those of the conventional SHORE estimation and MLP in Fig. 19. In all cases, MESC-Net achieves the smallest errors. The means and standard deviations of the average estimation errors shown in Fig. 19 are also summarized in Fig. 20, and the errors are compared between MESC-Net and the competing methods using paired Student's t-tests. Compared with the competing methods, the errors of MESC-Net are all highly significantly ($p < 0.001$) smaller, which is indicated in Fig. 20 by the asterisks. Note that the learning-based method MLP can have bigger estimation errors of MSD than the conventional SHORE model fitting.

4. Discussion

MESC-Net and the networks in Ye (2017a) and Ye (2017b) are motivated by the iterative optimization algorithms for solving sparse reconstruction problems. However, the proposed network is different from the networks in Ye (2017a) and Ye (2017b) in the following ways. First, our network incorporates historical information adaptively, which has not been explored by the previous networks in Ye (2017a) and Ye (2017b) for microstructure estimation. Adaptive incorporation of historical information has been used for improving optimization performance (Duchi et al., 2011), and such incorporation can be introduced into the network design to improve learning-based sparse reconstruction (Zhou et al., 2018). Note that the use of historical information is a generic technique for iterative optimization processes, and it is not necessarily restricted to sparse reconstruction. Second, the networks in Ye (2017a) and Ye (2017b) only consider sparse coding in the angular domain. In this work, we exploit the joint sparse coding in the spatial and angular domain, which could reduce the redundancy in diffusion signals (Schwab et al., 2016, 2018b). Third, unlike the networks in Ye (2017a) and Ye (2017b), which are designed specifically for the NODDI model, we aim to design a network for generic microstructure estimation, which is not necessarily limited to a particular signal model. Therefore, unlike Ye (2017a) and Ye (2017b) the second stage in our network uses fully connected layers to map the sparse representation of diffusion signals to scalar tissue microstructure measures.

As demonstrated in the experiments, the above improvements can lead to improved estimation performance. In particular, the individual benefits of the adaptive incorporation of historical information and spatial-angular sparse coding are confirmed by the investigation in Section 3.2.3. Also, MESC-Net was successfully applied to three different signal models, and this confirms that the proposed method can be used as a generic method for tissue microstructure estimation. We expect that MESC-Net can be used for signal models that are comparable to the signal models considered in the experiments in terms of model complexity.

The major hyperparameter of MESC-Net is the size N of the implicitly determined dictionary. We studied its impact on training for the NODDI model. Greater N leads to a smaller training loss but a longer training time. Also, the validation loss increases when N increases from 301 to 501. Therefore, we selected $N = 301$ for MESC-Net, which is the same dictionary size as used in Ye (2017b). Using this dictionary size, MESC-Net also performed well for the two additional signal models evaluated in the experiments, where it outperformed the competing methods. These results indicate the robustness of the proposed network structure and its hyperparameters. Note that although the size $N = 301$ is smaller than the length of the input vector of signal patches, which indicates that an undercomplete dictionary is used for the sparse representation, our task is to predict microstructure for the center voxel, where the sparse representation does not necessarily need to retain all the information in the input signal and could thus lie in a low-dimension space.

Following the existing works on q-space learning, we have computed the gold standard with conventional approaches using densely sampled diffusion gradients. Previous works have shown that the use of a larger number of diffusion gradients generally

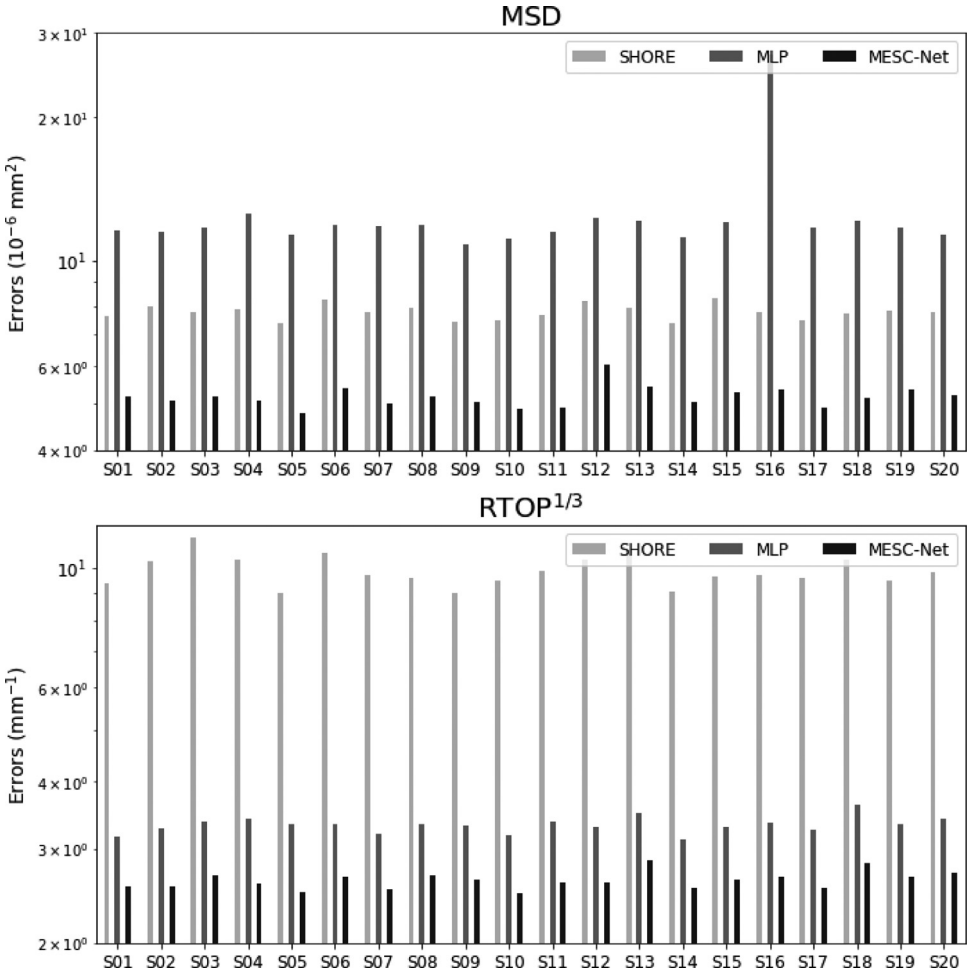

Fig. 19. Average microstructure estimation errors in the brains of the test subjects for the SHORE EAP model. Note that a log scale is used for the *y*-axis.

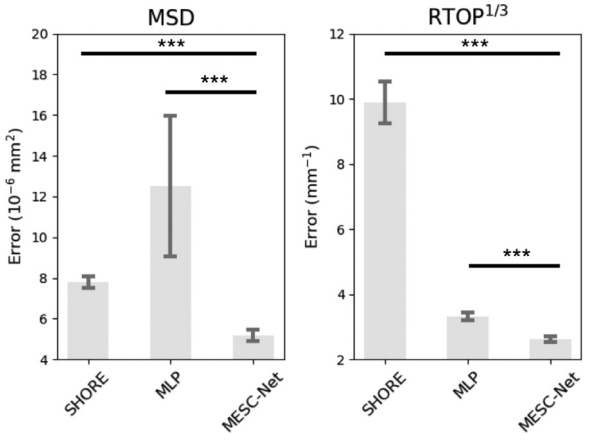

Fig. 20. Means and standard deviations of the average microstructure estimation errors shown in Fig. 19 for the SHORE EAP model. Asterisks (***) indicate that the difference between methods is highly significant ($p < 0.001$) using a paired Student's *t*-test.

leads to more accurate tissue microstructure estimation, and results estimated from a large number of diffusion gradients can be used as gold standard when results estimated from a fewer number of diffusion gradients are evaluated (Zhang et al., 2012). This observation is consistent with the qualitative comparison of the results of conventional approaches computed from undersampled and densely sampled diffusion signals in this work. Moreover, this observation is consistent with our quantitative results obtained with different numbers of diffusion gradients in Sections 3.2.2 and 3.2.4, where results computed from a smaller number of diffusion gradients have greater errors with respect to the gold standard.

Like in Ye (2017b), we use a $3\times3\times3$ patch as the input to MESC-Net and predict the center voxel. Such a strategy of predicting the center voxel or voxels is common for patch-based methods in other applications as well (Alexander et al., 2017b; Tanno et al., 2017; Kamnitsas et al., 2017). Another possible strategy is to predict microstructure for the whole patch and then average the results for overlapping voxels. However, we have empirically observed that (results not shown) such an alternative strategy led to worse estimation performance than the proposed strategy. This is possibly because the size N of the dictionary is sufficient for estimating microstructure for the center voxel but not for the whole patch. Although increasing the size N could allow better microstructure estimation for the whole patch, it would substantially increase the number of network parameters and a much larger number of training samples may be required. Since currently the training already requires about 47GB RAM, substantially increasing the number of training samples could lead to insufficient memory. Thus, we choose to perform microstructure prediction for the center voxel. It is also possible to use a larger input patch to better exploit the spatial information. However, increasing the patch size could also lead to much greater memory consumption in training, which is impractical due to memory overhead. Future work could explore the expansion of the input and output size with improved

training efficiency, so that the spatial-angular sparse coding can be better exploited.

The improvement to the proposed network structure could be explored in future work. In particular, the first stage for sparse reconstruction could be restructured. For example, in He et al. (2017) a Bayesian learning strategy has been incorporated into the design of networks for sparse reconstruction, leading to an LSTM architecture. It is possible to combine historical information with Bayesian learning to design an improved LSTM architecture. In addition, the computational cost could be reduced by using a separable formulation of spatial-angular sparse coding (Schwab et al., 2018a), and the iterative process developed in Schwab et al. (2018a) could inspire the design of a deep network with separable spatial-angular sparse coding.

Currently, the q-space learning techniques for tissue microstructure estimation (Golkov et al., 2016; Ye, 2017a, 2017b) require acquisitions of densely sampled diffusion signals to compute training microstructure. However, such training signals may not be available for existing datasets. To apply q-space learning to these datasets, it is possible to generate synthetic training data. Training with synthesized data has been successfully applied to computer vision tasks when real training data is not available (Ros et al., 2016). For q-space learning, we could synthesize densely sampled diffusion signals and ensure that their undersampled version has a distribution that is similar to the distribution of observed diffusion signals. The distribution mismatch could be measured by the maximum mean discrepancy (Gretton et al., 2012) or minimized with an adversarial learning strategy (Arjovsky et al., 2017; Goodfellow et al., 2014).

A similar situation is when the test data may not have a distribution that is close enough to that of the training data. For example, test scans could be acquired from scanners that are different from the one that scans the training subjects, although the training and test scans share the same set of undersampled diffusion gradients. The inter-scanner variability could lead to domain shift, where the training and test signals follow different distributions (Mirzaalian et al., 2016). We could either incorporate additional training images that can represent the distribution of test samples, or use transfer learning strategies to deal with the domain shift.

Some signal models—e.g., NODDI (Zhang et al., 2012)—assume a single fiber orientation in a voxel. However, there can be multiple fiber orientations. Therefore, orientation-specific microstructure could be estimated to give more detailed information about tissue organization. For example, the NODDI model could be extended to include compartments associated with different orientations (Farooq et al., 2016). The proposed network could be extended to estimate orientation-specific microstructure as well. For example, fiber orientations can be discretized (Landman et al., 2012; Ye et al., 2016) and the dictionary coefficients can be grouped, where each group corresponds to a discretized orientation. The coefficients in each group can then be mapped to compute the microstructure associated with the corresponding orientation. However, in this strategy the number of entries in the sparse representation substantially increases, which could lead to a much more challenging sparse reconstruction problem given the reduced number of diffusion gradients. Thus, investigation of a reliable solution to this problem may still be required.

The proposed network could also be applied to other dMRI processing tasks with proper modifications. For example, fiber orientation estimation can be casted as a sparse reconstruction problem (Landman et al., 2012; Ye et al., 2016), and networks can be designed by unfolding the IHT process to aid fiber orientation estimation (Ye and Prince, 2017). The first stage in MESC-Net can be readily used to learn the mapping from diffusion signals to fiber orientations. Similarly, EAP estimation could be solved via sparse reconstruction (Merlet and Deriche, 2013; Merlet et al., 2013), and the first stage in MESC-Net can be used for EAP estimation as well. But note that in these tasks, there are many more parameters in the output to estimate than in the task of tissue microstructure estimation. Thus, estimation of these quantities with a modified MESC-Net could be more challenging and requires further investigation.

5. Summary and conclusion

We have proposed a generic deep network, which is referred to as MESC-Net, for the estimation of scalar tissue microstructure measures. MESC-Net exploits the sparse representation of diffusion signals in the spatial-angular domain, and learns the sparse representation computation using a network architecture with modified LSTM units. These modified LSTM units are motivated by the adaptive incorporation of historical information into the iterative update for solving sparse reconstruction problems. The sparse representation is then mapped to scalar tissue microstructure measures, where the mapping is learned jointly with the sparse representation computation. The proposed method was applied to brain dMRI scans, where tissue microstructure described by various signal models was quantified. Results indicate that MESC-Net outperforms state-of-the-art conventional and deep learning based estimation approaches.

Conflicts of interest

None.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC 61601461), Beijing Natural Science Foundation (7192108), Beijing Institute of Technology Research Fund Program for Young Scholars, and Beihang Zhuobai Program (grant no. ZG216S1871). Data were provided 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.

References

Alexander, D.C., Dyrby, T.B., Nilsson, M., Zhang, H., 2017a. Imaging brain microstructure with diffusion MRI: practicality and applications. NMR Biomed.

Alexander, D.C., Zikic, D., Ghosh, A., Tanno, R., Wottschel, V., Zhang, J., Kaden, E., Dyrby, T.B., Sotiropoulos, S.N., Zhang, H., Criminisi, A., 2017b. Image quality transfer and applications in diffusion MRI. Neuroimage 152, 283–298.

Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein generative adversarial networks. In: International Conference on Machine Learning, pp. 214–223.

Barnea-Goraly, N., Kwon, H., Menon, V., Eliez, S., Lotspeich, L., Reiss, A.L., 2004. White matter structure in autism: preliminary evidence from diffusion tensor imaging. Biol. Psychiatry 55 (3), 323–326.

Basser, P.J., Mattiello, J., LeBihan, D., 1994. MR Diffusion tensor spectroscopy and imaging. Biophys. J. 66 (1), 259–267.

Batalle, D., Hughes, E.J., Zhang, H., Tournier, J.-D., Tusor, N., Aljabar, P., Wali, L., Alexander, D.C., Hajnal, J.V., Nosarti, C., Edwards, A.D., Counsell, S.J., 2017. Early development of structural networks and the impact of prematurity on brain connectivity. Neuroimage 149, 379–392.

Batalle, D., O'Muircheartaigh, J., Makropoulos, A., Kelly, C.J., Dimitrova, R., Hughes, E.J., Hajnal, J.V., Zhang, H., Alexander, D.C., Edwards, A.D., Counsell, S.J., 2019. Different patterns of cortical maturation before and after 38 weeks gestational age demonstrated by diffusion MRI in vivo. Neuroimage 185, 764–775.

Blumensath, T., Davies, M.E., 2008. Iterative thresholding for sparse approximations. J. Fourier Anal. Appl. 14 (5–6), 629–654.

By, S., Xu, J., Box, B.A., Bagnato, F.R., Smith, S.A., 2017. Application and evaluation of NODDI in the cervical spinal cord of multiple sclerosis patients. NeuroImage 15, 333–342.

Colgan, N., Siow, B., O'Callaghan, J., Harrison, I., Wells, J., Holmes, H., Ismail, O., Richardson, S., Alexander, D., Collins, E., Fisher, E., Johnson, R., Schwarz, A., Ahmed, Z., O'Neill, M., Murray, T., Zhang, H., Lythgoe, M., 2016. Application of

neurite orientation dispersion and density imaging (NODDI) to a tau pathology model of Alzheimer's disease. Neuroimage 125, 739–744.
Conel, J.L.R., 1939. The Postnatal Development of the Human Cerebral Cortex. Harvard Univ. Press.
Crombe, A., Planche, V., Raffard, G., Bourel, J., Dubourdieu, N., Panatier, A., Fukutomi, H., Dousset, V., Oliet, S., Hiba, B., Tourdias, T., 2018. Deciphering the microstructure of hippocampal subfields with in vivo DTI and NODDI: applications to experimental multiple sclerosis. Neuroimage 172, 357–368.
Daducci, A., Canales-Rodríguez, E.J., Zhang, H., Dyrby, T.B., Alexander, D.C., Thiran, J.-P., 2015. Accelerated microstructure imaging via convex optimization (AMICO) from diffusion MRI data. Neuroimage 105, 32–44.
Domke, J., 2012. Generic methods for optimization-based modeling. In: AISTATS, vol. 22, pp. 318–326.
Duchi, J., Hazan, E., Singer, Y., 2011. Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res. 12 (Jul), 2121–2159.
Evangelou, N., Esiri, M.M., Smith, S., Palace, J., Matthews, P.M., 2000. Quantitative pathological evidence for axonal loss in normal appearing white matter in multiple sclerosis. Ann. Neurol. 47 (3), 391–395.
Farooq, H., Xu, J., Nam, J.W., Keefe, D.F., Yacoub, E., Georgiou, T., Lenglet, C., 2016. Microstructure imaging of crossing (MIX) white matter fibers from diffusion MRI. Sci. Rep. 6.
Fiala, J.C., Spacek, J., Harris, K.M., 2002. Dendritic spine pathology: cause or consequence of neurological disorders? Brain Res. Rev. 39 (1), 29–54.
Fick, R., Wassermann, D., Deriche, R., 2018. Mipy: An open-source framework to improve reproducibility in brain microstructure imaging. In: OHBM 2018-Human Brain Mapping, pp. 1–4.
Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M., Nimmo-Smith, I., Contributors, D., 2014. Dipy, a library for the analysis of diffusion MRI data. Front. Neuroinform. 8 (8), 1–17.
Genç, E., Fraenz, C., Schlüter, C., Friedrich, P., Hossiep, R., Voelkle, M.C., Ling, J.M., Güntürkün, O., Jung, R.E., 2018. Diffusion markers of dendritic density and arborization in gray matter predict differences in intelligence. Nat. Commun. 9 (1), 1905.
Genc, S., Malpas, C.B., Holland, S.K., Beare, R., Silk, T.J., 2017. Neurite density index is sensitive to age related differences in the developing brain. Neuroimage 148, 373–380.
Golkov, V., Dosovitskiy, A., Sperl, J.I., Menzel, M.I., Czisch, M., Sämann, P., Brox, T., Cremers, D., 2016. Q-Space deep learning: twelve-fold shorter and model-free diffusion MRI scans. IEEE Trans. Med. Imaging 35 (5), 1344–1351.
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680.
Gregor, K., LeCun, Y., 2010. Learning fast approximations of sparse coding. In: International Conference on Machine Learning, pp. 399–406.
Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A., 2012. A kernel two-sample test. J. Mach. Learn. Res. 13 (1), 723–773.
He, H., Xin, B., Ikehata, S., Wipf, D., 2017. From Bayesian sparsity to gated recurrent nets. In: Advances in Neural Information Processing Systems 30, pp. 5554–5564.
Hershey, J. R., Roux, J. L., Weninger, F., 2014. Deep unfolding: model-based inspiration of novel deep architectures. arXiv:1409.2574.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. Neural Netw. 4 (2), 251–257.
Hui, E.S., Cheung, M.M., Qi, L., Wu, E.X., 2008. Towards better MR characterization of neural tissues using directional diffusion kurtosis analysis. Neuroimage 42 (1), 122–134.
Jacobs, B., Driscoll, L., Schall, M., 1997. Life-span dendritic and spine changes in areas 10 and 18 of human cortex: a quantitative Golgi study. J. Comp. Neurol. 386 (4), 661–680.
Jelescu, I.O., Veraart, J., Adisetiyo, V., Milla, S.S., Novikov, D.S., Fieremans, E., 2015. One diffusion acquisition and different white matter models: how does microstructure change in human early development based on WMTI and NODDI? Neuroimage 107, 242–256.
Johansen-Berg, H., Behrens, T.E.J., 2013. Diffusion MRI: From quantitative measurement to in vivo neuroanatomy. Academic Press, Waltham.
Jones, D., Alexander, D., Bowtell, R., Cercignani, M., Dell'Acqua, F., McHugh, D., Miller, K., Palombo, M., Parker, G., Rudrapatna, U., Tax, C., 2018. Microstructural imaging of the human brain with a super-scanner: 10 key advantages of ultra--strong gradients for diffusion MRI. Neuroimage 182, 8–38.
Kaden, E., Kelm, N.D., Carson, R.P., Does, M.D., Alexander, D.C., 2016a. Multi-compartment microscopic diffusion imaging. Neuroimage 139, 346–359.
Kaden, E., Kruggel, F., Alexander, D.C., 2016b. Quantitative mapping of the per-axon diffusion coefficients in brain white matter. Magn. Reson. Med. 75 (4), 1752–1763.
Kamagata, K., Hatano, T., Okuzumi, A., Motoi, Y., Abe, O., Shimoji, K., Kamiya, K., Suzuki, M., Hori, M., Kumamaru, K.K., Hattori, N., Aoki, S., 2016. Neurite orientation dispersion and density imaging in the substantia nigra in idiopathic Parkinson disease. Eur. Radiol. 26 (8), 2567–2577.
Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78.
Kelly, C.E., Thompson, D.K., Chen, J., Leemans, A., Adamson, C.L., Inder, T.E., Cheong, J.L., Doyle, L.W., Anderson, P.J., 2016. Axon density and axon orientation dispersion in children born preterm. Hum. Brain Mapp. 37 (9), 3080–3102.
Kingma, D., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv:1412.6980.
Konda, K., Memisevic, R., Krueger, D., 2014. Zero-bias autoencoders and the benefits of co-adapting features. arXiv:1402.3337.
Kunz, N., Zhang, H., Vasung, L., O'brien, K.R., Assaf, Y., Lazeyras, F., Alexander, D.C., Hüppi, P.S., 2014. Assessing white matter microstructure of the newborn with multi-shell diffusion MRI and biophysical compartment models. Neuroimage 96, 288–299.
Landman, B.A., Bogovic, J.A., Wan, H., ElShahaby, F.E.Z., Bazin, P.-L., Prince, J.L., 2012. Resolution of crossing fibers with constrained compressed sensing using diffusion tensor MRI. Neuroimage 59 (3), 2175–2186.
Lemkaddem, A., Daducci, A., Kunz, N., Lazeyras, F., Seeck, M., Thiran, J.-P., Vulliémoz, S., 2014. Connectivity and tissue microstructural alterations in right and left temporal lobe epilepsy revealed by diffusion spectrum imaging. NeuroImage 5, 349–358.
Lipton, M.L., Gulko, E., Zimmerman, M.E., Friedman, B.W., Kim, M., Gellella, E., Gold, T., Shifteh, K., Ardekani, B.A., Branch, C.A., 2009. Diffusion-tensor imaging implicates prefrontal axonal injury in executive function impairment following very mild traumatic brain injury. Radiology 252 (3), 816–824.
McKinnon, E.T., Helpern, J.A., Jensen, J.H., 2018. Modeling white matter microstructure with fiber ball imaging. Neuroimage 176, 11–21.
Merlet, S., Caruyer, E., Ghosh, A., Deriche, R., 2013. A computational diffusion MRI and parametric dictionary learning framework for modeling the diffusion signal and its features. Med. Image Anal. 17 (7), 830–843.
Merlet, S.L., Deriche, R., 2013. Continuous diffusion signal, EAP and ODF estimation via compressive sensing in diffusion MRI. Med. Image Anal. 17 (5), 556–572.
Mirzaalian, H., Ning, L., Savadjiev, P., Pasternak, O., Bouix, S., Michailovich, O., Grant, G., Marx, C., Morey, R., Flashman, L., George, M., McAllister, T., Andaluz, N., Shutter, L., Coimbra, R., Zafonte, R., Coleman, M., Kubicki, M., Westin, C., Stein, M., Shenton, M., Rathi, Y., 2016. Inter-site and inter-scanner diffusion MRI data harmonization. Neuroimage 135, 311–323.
Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted Boltzmann machines. In: International Conference on Machine Learning, pp. 807–814.
Ning, L., Westin, C.-F., Rathi, Y., 2015. Estimating diffusion propagator and its moments using directional radial basis functions. IEEE Trans. Med. Imaging 34 (10), 2058–2078.
Ocklenburg, S., Friedrich, P., Fraenz, C., Schlüter, C., Beste, C., Güntürkün, O., Genç, E., 2018. Neurite architecture of the planum temporale predicts neurophysiological processing of auditory speech. Sci. Adv. 4 (7). doi:10.1126/sciadv.aar6830.
Özarslan, E., Koay, C.G., Shepherd, T.M., Komlosh, M.E., İrfanoğlu, 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.
Parker, T.D., Slattery, C.F., Zhang, J., Nicholas, J.M., Paterson, R.W., Foulkes, A.J., Malone, I.B., Thomas, D.L., Modat, M., Cash, D.M., Crutch, S.J., Alexander, D.C., Ourselin, S., Fox, N.C., Zhang, H., Schott, J.M., 2018. Cortical microstructure in young onset Alzheimer's disease using neurite orientation dispersion and density imaging. Hum. Brain Mapp. 39 (7), 3005–3017.
Pasternak, O., Kelly, S., Sydnor, V.J., Shenton, M.E., 2018. Advances in microstructural diffusion neuroimaging for psychiatric disorders. Neuroimage 182, 259–282.
Qian, N., 1999. On the momentum term in gradient descent learning algorithms. Neural Netw. 12 (1), 145–151.
Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A.M., 2016. The SYNTHIA dataset: a large collection of synthetic images for semantic segmentation of urban scenes. In: Computer Vision and Pattern Recognition, pp. 3234–3243.
Schwab, E., Haeffele, B., Charon, N., Vidal, R., 2018a. Separable dictionary learning with global optimality and applications to diffusion MRI. arXiv:1807.05595.
Schwab, E., Vidal, R., Charon, N., 2016. Spatial-angular sparse coding for HARDI. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 475–483.
Schwab, E., Vidal, R., Charon, N., 2018. Joint spatial-angular sparse coding for dMRI with separable dictionaries. Med. Image Anal. 48, 25–42.
Skranes, J., Vangberg, T., Kulseng, S., Indredavik, M., Evensen, K., Martinussen, M., Dale, A., Haraldseth, O., Brubakk, A.-M., 2007. Clinical findings and white matter abnormalities seen on diffusion tensor imaging in adolescents with very low birth weight. Brain 130 (3), 654–666.
Sonoda, S., Murata, N., 2017. Neural network with unbounded activation functions is universal approximator. Appl. Comput. Harmon Anal. 43 (2), 233–268.
Sprechmann, P., Bronstein, A.M., Sapiro, G., 2015. Learning efficient sparse and low rank models. IEEE Trans Pattern Anal Mach Intell 37 (9), 1821–1833.
Tanno, R., Worrall, D.E., Ghosh, A., Kaden, E., Sotiropoulos, S.N., Criminisi, A., Alexander, D.C., 2017. Bayesian image quality transfer with CNNs: Exploring uncertainty in dMRI super-resolution. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 611–619.
Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., 2013. The WU-Minn human connectome project: an overview. Neuroimage 80, 62–79.
Wang, Z., Ling, Q., Huang, T.S., 2016a. Learning deep ℓ_0 encoders. In: AAAI Conference on Artificial Intelligence, pp. 2194–2200.
Wang, Z., Yang, Y., Chang, S., Ling, Q., Huang, T.S., 2016b. Learning a deep ℓ_∞ encoder for hashing. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. AAAI Press, pp. 2174–2180.
Werring, D., Clark, C., Barker, G., Thompson, A., Miller, D., 1999. Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis. Neurology 52 (8). 1626–1626
Winston, G.P., Micallef, C., Symms, M.R., Alexander, D.C., Duncan, J.S., Zhang, H., 2014. Advanced diffusion imaging sequences could aid assessing patients with focal cortical dysplasia and epilepsy. Epilepsy Res. 108 (2), 336–339.

Wu, Y.-C., Field, A.S., Alexander, A.L., 2008. Computation of diffusion function measures in *q*-space using magnetic resonance hybrid diffusion imaging. IEEE Trans. Med. Imaging 27 (6), 858–865.

Wu, Y.-C., Mustafi, S.M., Harezlak, J., Kodiweera, C., Flashman, L.A., McAllister, T., 2018. Hybrid diffusion imaging in mild traumatic brain injury. J. Neurotrauma 35 (20), 2377–2390.

Xin, B., Wang, Y., Gao, W., Wipf, D., 2016. Maximal sparsity with deep networks? In: Advances In Neural Information Processing Systems, pp. 4340–4348.

Yang, Y., Sun, J., Li, H., Xu, Z., 2016. Deep ADMM-net for compressive sensing MRI. In: Advances in Neural Information Processing Systems, pp. 10–18.

Ye, C., 2017a. Estimation of tissue microstructure using a deep network inspired by a sparse reconstruction framework. In: International Conference on Information Processing in Medical Imaging 2017, pp. 466–477.

Ye, C., 2017b. Tissue microstructure estimation using a deep network inspired by a dictionary-based framework. Med. Image Anal. 42, 288–299.

Ye, C., Prince, J.L., 2017. Fiber orientation estimation guided by a deep network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 575–583.

Ye, C., Zhuo, J., Gullapalli, R.P., Prince, J.L., 2016. Estimation of fiber orientations using neighborhood information. Med. Image Anal. 32, 243–256. https://doi.org/10.1016/j.media.2016.05.008.

Zhang, H., Schneider, T., Wheeler-Kingshott, C.A., Alexander, D.C., 2012. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. Neuroimage 61 (4), 1000–1016.

Zhang, J., Ghanem, B., 2018. ISTA-Net: interpretable optimization-inspired deep network for image compressive sensing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1828–1837.

Zhou, J.T., Di, K., Du, J., Peng, X., Yang, H., Pan, S.J., Tsang, I., Liu, Y., Qin, Z., Goh, R.S.M., 2018. SC2Net: sparse LSTMs for sparse coding. In: AAAI Conference on Artificial Intelligence, pp. 4588–4595.