Fiber Orientation Estimation Guided by a Deep Network

Chuyang Ye[1] and **Jerry L. Prince**[2]

[1]National Laboratory of Pattern Recognition & Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China

[2]Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA

Abstract

Diffusion magnetic resonance imaging (dMRI) is currently the only tool for noninvasively imaging the brain's white matter tracts. The *fiber orientation* (FO) is a key feature computed from dMRI for tract reconstruction. Because the number of FOs in a voxel is usually small, dictionary-based sparse reconstruction has been used to estimate FOs. However, accurate estimation of complex FO configurations in the presence of noise can still be challenging. In this work we explore the use of a deep network for FO estimation in a dictionary-based framework and propose an algorithm named *Fiber Orientation Reconstruction guided by a Deep Network* (FORDN). FORDN consists of two steps. First, we use a smaller dictionary encoding coarse basis FOs to represent diffusion signals. To estimate the mixture fractions of the dictionary atoms, a deep network is designed to solve the sparse reconstruction problem. Second, the coarse FOs inform the final FO estimation, where a larger dictionary encoding a dense basis of FOs is used and a weighted ℓ_1-norm regularized least squares problem is solved to encourage FOs that are consistent with the network output. FORDN was evaluated and compared with state-of-the-art algorithms that estimate FOs using sparse reconstruction on simulated and typical clinical dMRI data. The results demonstrate the benefit of using a deep network for FO estimation.

Keywords

diffusion MRI; fiber orientation estimation; deep network; sparse reconstruction

1 Introduction

Diffusion magnetic resonance imaging (dMRI) is currently the only tool that enables reconstruction of *in vivo* white matter tracts [5]. By capturing the anisotropic water diffusion in tissue, dMRI infers information about fiber orientations (FOs), which are crucial features in white matter tract reconstruction [5].

Various methods have been proposed to estimate FOs. In particular, methods based on sparse reconstruction have shown efficacy in reliable FO estimation with a reduced number of gradient directions [8]. The sparsity assumption is typically combined with multi-tensor models [3,4,8], which leads to dictionary-based sparse reconstruction of FOs. However,

accurate estimation of complex FO configurations in the presence of noise can still be challenging.

In this work, we explore the use of a deep network to improve dictionary-based sparse reconstruction of FOs. We model the diffusion signals using a dictionary, the atoms of which encode a set of basis FOs. Then, FO estimation can be formulated as a sparse reconstruction problem and we seek to solve it with the aid of a deep network. The proposed method is named *Fiber Orientation Reconstruction guided by a Deep Network* (FORDN), which consists of two steps. First, a deep network that unfolds the conventional iterative estimation process is constructed and its weights are learned from synthesized training samples. To reduce the computational burden of training, this step uses a smaller dictionary that encodes a coarse set of basis FOs, and thus gives approximate estimates of FOs. Second, the final sparse reconstruction of FOs is guided by the FOs produced by the deep network. A larger dictionary encoding dense basis FOs is used, and a weighted ℓ_1-norm regularized least squares problem is solved to encourage FOs that are consistent with the network output. Experiments were performed on simulated and typical clinical brain dMRI data, where promising results were observed compared with competing algorithms.

2 Methods

2.1 Background: FO Estimation by Sparse Reconstruction

Diffusion signals can be modeled with a set of fixed prolate tensors, each representing a possible FO by its primary eigenvector (PEV) [4,8]. Suppose the set of the basis tensors is $\{\mathbf{D}_i\}_{i=1}^N$ and their PEVs are $\{v_i\}_{i=1}^N$, where N is the number of the basis tensors. In this work we use $N = 289$ [14], which results from tessellating an octahedron. The eigenvalues of the basis tensors can be determined by examining the diffusion tensors in regions occupied by single tracts [8].

For each diffusion gradient direction g_k ($k = 1, \ldots, K$) associated with a b-value b_k, the diffusion weighted signal at each voxel can be represented as [8]

$$S(g_k) = S(\mathbf{0}) \sum_{i=1}^N f_i e^{-b_k g_k^T \mathbf{D}_i g_k} + n(g_k),\tag{1}$$

where $S(\mathbf{0})$ is the baseline signal without diffusion weighting, f_i is the unknown nonnegative *mixture fraction* (MF) for $\mathbf{D}_i$ ($\sum_{i=1}^N f_i = 1$), and $n(g_k)$ represents image noise. By defining $y(g_k) = S(g_k)/S(\mathbf{0})$ and $\eta(g_k) = n(g_k)/S(\mathbf{0})$, we have

$$y = \mathbf{G}f + \eta,\tag{2}$$

where $y = (y(g_1), \ldots, y(g_K))^T, f = (f_1, \ldots, f_N)^T, \eta = (\eta(g_1), \ldots, \eta(g_K))^T$, and $\mathbf{G} \in \mathbb{R}^{K \times N}$ is a dictionary matrix with $G_{ki} = e^{-b_k g_k^T \mathbf{D}_i g_k}$.

Because the number of FOs at a voxel is small compared with that of gradient directions, FOs can be estimated by solving a sparse reconstruction problem

$$\hat{f} = \underset{f \geq 0, \|f\|_1 = 1}{\arg\min} \|Gf - y\|_2^2 + \beta \|f\|_0. \tag{3}$$

To solve Eq. (3), the constraint of $\|f\|_1 = 1$ is usually removed [4, 8]. Then, the problem can be solved by using iterative reweighted ℓ_1-norm minimization [4] or by approximating the ℓ_0-norm with the ℓ_1-norm [8]. The solution is finally normalized so that the MFs sum to one and basis directions associated with MFs larger than a threshold are set to be FOs [8].

2.2 FO Estimation Using a Deep Network

Consider the general sparse reconstruction problem

$$\hat{f} = \underset{f}{\arg\min} \|Gf - y\|_2^2 + \beta \|f\|_0. \tag{4}$$

Using methods such as *iterative hard thresholding* (IHT) [2], Eq. (4) or its ℓ_1-norm relaxed version can be solved by iteratively updating f. At iteration $t + 1$,

$$f^{t+1} = h_\lambda(Wy + Sf^t), \tag{5}$$

where $W = G^T$, $S = I - G^T G$, and $h_\lambda(\cdot)$ is a thresholding operator with a parameter $\lambda \geq 0$. Motivated by this iterative process, previous works have explored the use of a deep network for solving sparse reconstruction problems. By unfolding and truncating the process in Eq. (5), feed-forward deep network structures can be constructed for sparse reconstruction, where W and S are learned from training data instead of predetermined by G [11, 12]. As demonstrated by [12], these learned layer-wise fixed weights could guarantee successful reconstruction across a wider range of *restricted isometry property* (RIP) conditions than conventional methods such as IHT.

In this work, to solve Eq. (3) we construct a deep network as shown in Fig. 1, where the input is the diffusion signal y at a voxel and the output is the MF f. The layers $L = 1, 2, \ldots, 8$ correspond to the unfolded and truncated iterative process in Eq. (5) (assuming $f^0 = 0$), where W and S (shared among layers) are learned. The thresholded *rectified linear unit* (ReLU) [7] is given by $[h_\lambda(a)]_i = a_i \mathbf{1}_{a_i \geq \lambda}$, where $\mathbf{1}$ is an indicator function, and is used in each of these layers. The thresholded ReLU yields the thresholding operator in IHT [2] (Eq. (5)). We empirically set $\lambda = 0.01$. Note that because of the nonnegative constraint on f in Eq. (3), $[h_\lambda(a)]_i$ is always zero when $a_i < 0$. A normalization layer is added before the output to enforce that the entries of f sum to one. To ensure numerical stability, we use $f \leftarrow (f + \tau \mathbf{1})/\|f + \tau \mathbf{1}\|_1$ for the normalization, where $\tau = 10^{-10}$. The network is implemented using the Keras library (http://keras.io/). We use the mean squared error as the loss function and the Adam

algorithm [6] as the optimizer; the learning rate is 0.001, the batch size is 64, and the number of epochs is 8, which achieves stable training loss in practice.

If the training data are generated by conventional algorithms, such as IHT, then the network only learns a strategy that approximates these suboptimal solutions [12]. Thus, we adopt the strategy of synthesizing observations [12] according to given FO configurations. However, synthesis of diffusion signals for all combinations is prohibitive. For example, for the cases

of three crossing fibers, the total number of FO combinations is $\binom{N}{3} \approx 4 \times 10^6$ and each combination requires sufficient training instances with noise sampling and different combinations of MFs. Thus, training the deep network using the full set of basis directions can be computationally intensive. Therefore, we use a two-step strategy to estimate FOs: 1) by using a smaller set of basis FOs, coarse FOs are estimated using the proposed deep network; 2) the final FO estimation is guided by these coarse FOs by solving a weighted ℓ_1-norm regularized least squares problem. Details of these two steps are described below.

Coarse FO Estimation Using a Deep Network—A smaller set of basis tensors $\{\tilde{\mathbf{D}}_{i'}\}_{i'=1}^{N'}$ ($N'=73$) are considered for coarse FO estimation using the deep network. As discussed and assumed in the literature, we consider cases with three or fewer FOs in synthesizing the training data [4]. The cases of FO configurations can be given by applying an existing FO estimation method to the subject of interest. In this work we use CFARI [8] which estimates FOs using sparse reconstruction. Note that such a method does not need to provide accurate FO configurations at every voxel. Instead, it provides a good estimate of the set of FO configurations in the brain or a region of interest.

Because the original CFARI method can give multiple close FOs to represent a single FO that is not collinear with a basis direction, which unnecessarily increases the number of FOs, and that these FOs may not be collinear with the smaller set of basis directions considered in the deep network, we post-process the CFARI FOs by selecting the peak directions in terms of MFs and then map them to their closest basis directions in the coarse basis set.

The refined CFARI FOs in the brain or a brain region provide a set of training FO configurations. For each FO configuration with a single or multiple basis directions, diffusion signals were synthesized with a single-tensor or multi-tensor model using the corresponding basis tensors, respectively. For a single basis direction, its MF was set to one; for multiple basis directions, different combinations of their MFs from 0.1 to 0.9 in increments of 0.1 were used for synthesis (note that they should sum to one). Rician noise was added to the synthesized signals, and the *signal-to-noise ratio* (SNR) can be obtained, for example, by placing bounding boxes in background and white matter areas [13]. For each MF combination, 500 samples were generated for training.

We further reduce the computational cost of training by parcellating the brain into different regions, each containing a small number of FO configurations. This is achieved by registering the EVE template [10] to the subject using the *fractional anisotropy* (FA) map and the SyN algorithm [1]. A deep network is then constructed for each region using all the

FO configurations in that region, and thus each network requires a much smaller number of training samples.

In the test phase, the trained networks estimate the MFs in their corresponding parcellated brain regions. Like [8], the basis directions with MFs larger than a threshold of 0.1 are set to be the FOs. The FOs predicted by the deep networks are denoted by $\mathcal{U} = \{\boldsymbol{u}_p\}_{p=1}^{U}$, where U is the cardinality of $\mathcal{U}$.

FO Estimation Guided by the Deep Network—The coarse FOs given by the deep networks provide only approximate FO estimates due to the low angular resolution of the coarse basis; however, they can guide the final sparse FO reconstruction that uses the larger set of basis directions. Specifically, at each voxel we solve the following weighted ℓ_1-norm regularized least squares problem [13] that allows incorporation of prior knowledge of FOs,

$$\hat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f} \geq \boldsymbol{0}} \|\mathbf{G}\boldsymbol{f} - \boldsymbol{y}\|_2^2 + \beta \|\mathbf{C}\boldsymbol{f}\|_1. \tag{6}$$

Here, $\mathbf{C}$ is a diagonal matrix encoding the guiding FOs predicted by the deep network, and basis directions closer to the guiding FOs are encouraged. The diagonal weights are specified as [14]

$$C_i \frac{1 - \alpha \max_{p=1,\ldots,U} |\boldsymbol{v}_i \cdot \boldsymbol{u}_p|}{\min_{q=1,\ldots,N} \left(1 - \alpha \max_{p=1,\ldots,U} |\boldsymbol{v}_q \cdot \boldsymbol{u}_p|\right)}, \quad i = 1, \ldots, N. \tag{7}$$

When $\boldsymbol{v}_i$ is close to the guiding FOs, C_i is small and therefore f_i is encouraged to be large and $\boldsymbol{v}_i$ is encouraged to be selected as an FO. Eq. (6) can be solved using the strategy in [13]. We set $\alpha = 0.8$ as in [14], and selected $\beta = 0.25$ because the number of diffusion gradients used in this work is about half of that used in [14]. The MFs are normalized so that they sum to one, and the FOs are determined as the basis directions associated with MFs larger than 0.1 [8].

3 Results

3.1 3D Digital Crossing Phantom

A 3D digital crossing phantom was created, where the tract geometries and diffusion parameters in [14] were used. The phantom consists of regions of single tracts, two crossing tracts, and three crossing tracts. Thirty gradient directions were applied with $b = 1000$ s/mm^2. Rician noise (SNR = 20 on the b0 image) was added to the *diffusion weighted images* (DWIs).

We quantitatively evaluated the accuracy of FORDN using the error measure in [14]. We compared FORDN with two sparse reconstruction based methods, CFARI [8] and L2L0 [4]. Note that in [14] CFARI has already been compared with techniques that are not based on

sparse reconstruction, where they achieve similar estimation performance. We also compared the final FORDN results with the intermediate output from the deep network (DN). The errors in the entire phantom and in each region containing noncrossing, two crossing, or three crossing tracts are shown in Fig. 2(a). In all cases, FORDN achieves more accurate FO reconstruction. In addition, the intermediate DN results already improves FO estimation in regions with crossing tracts compared with CFARI and L2L0.

The FO errors in Fig. 2(a) were also compared between FORDN and CFARI, L2L0, and DN using a paired Student's *t*-test. In all cases, the FORDN errors are significantly smaller ($p <$ 0.001), and the effect sizes (Cohen's *d*) are shown in Fig. 2(b). The effect sizes are larger in regions with three crossing tracts, indicating greater improvement in regions with more complex fiber structures.

3.2 Brain dMRI

FORDN was next applied to a dMRI scan of a random subject from the Kirby21 dataset [9].[1] DWIs were acquired on a 3T MR scanner (Achieva, Philips, Best, Netherlands). Thirty-two gradient directions ($b = 700$ s/mm^2) were used. The in-plane resolution is 2.2 mm isotropic and was upsampled by the scanner to 0.828 mm isotropic. The slice thickness is 2.2 mm. We resampled the DWIs so that the resolution is 2.2 mm isotropic. The SNR is about 22 on the b0 image.

FOs in a region where the CC and SLF cross are shown in Fig. 3 and compared with CFARI and L2L0. FORDN better reconstructs the transverse CC FOs and the anterior–posterior SLF FOs than CFARI and L2L0 (see the high-lighted region). Fiber tracking was then performed using the strategy in [15], where seeds were placed in the noncrossing CC (see Fig. 4). The FA threshold is 0.2, the turning angle threshold is 45°, and the step size is 1 mm. The results are shown in Fig. 4, and each segment is color-coded using the standard color scheme—red: left–right; green: front–back; and blue: up–down. FORDN FOs do not produce the false (green) streamlines going in the anterior–posterior direction as in the CFARI and L2L0 results (see the zoomed region). Note that the streamlines tracked by FORDN propagate through multiple regions parcellated by the EVE atlas, which indicates that consistency of the streamlines is preserved although each region is associated with a different deep network.

4 Conclusion

We have proposed an algorithm of FO estimation guided by a deep network. The diffusion signals are modeled in a dictionary-based framework. A deep network designed for sparse reconstruction provides coarse FO estimation using a smaller set of the dictionary atoms, which then informs the final FO estimation using weighted ℓ_1-norm regularization. Results on simulated and clinical brain dMRI have demonstrated promising results compared with the competing methods.

---

[1]Additional results can be found at http://arxiv.org/abs/1705.06870.

Acknowledgments

References

1. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis. 2008; 12(1):26–41. [PubMed: 17659998]

2. Blumensath T, Davies ME. Iterative thresholding for sparse approximations. Journal of Fourier Analysis and Applications. 2008; 14(5–6):629–654.

3. Chen G, Zhang P, Li K, Wee CY, Wu Y, Shen D, Yap PT. Improving estimation of fiber orientations in diffusion MRI using inter-subject information sharing. Scientific Reports. 2016; 6:37847. [PubMed: 27892534]

4. Daducci A, Van De Ville D, Thiran JP, Wiaux Y. Sparse regularization for fiber ODF reconstruction: From the suboptimality of ℓ_2 and ℓ_1 priors to ℓ_0. Medical Image Analysis. 2014; 18(6):820–833. [PubMed: 24593935]

5. Johansen-Berg, H., Behrens, TEJ. Diffusion MRI: from quantitative measurement to in vivo neuroanatomy. Waltham: Academic Press; 2013.

6. Kingma D, Ba J. Adam: A method for stochastic optimization. 2014 arXiv preprint arXiv: 1412.6980.

7. Konda K, Memisevic R, Krueger D. Zero-bias autoencoders and the benefits of co-adapting features. 2014 arXiv preprint arXiv:1402.3337.

8. Landman BA, Bogovic JA, Wan H, ElShahaby FEZ, Bazin PL, Prince JL. Resolution of crossing fibers with constrained compressed sensing using diffusion tensor MRI. NeuroImage. 2012; 59(3): 2175–2186. [PubMed: 22019877]

9. Landman BA, Huang AJ, Gifford A, Vikram DS, Lim IAL, Farrell JA, Bogovic JA, Hua J, Chen M, Jarso S, Smith SA, Joel S, Mori S, Pekar JJ, Barker PB, Prince JL, van Zijl PC. Multi-parametric neuroimaging reproducibility: A 3-T resource study. NeuroImage. 2011; 54(4):2854–2866. [PubMed: 21094686]

10. Oishi K, Faria A, Jiang H, Li X, Akhter K, Zhang J, Hsu JT, Miller MI, van Zijl PC, Albert M, et al. Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and Alzheimer's disease participants. NeuroImage. 2009; 46(2):486–499. [PubMed: 19385016]

11. Wang, Z., Ling, Q., Huang, TS. Learning deep ℓ_0 encoders; AAAI Conference on Artificial Intelligence; 2016. p. 2194-2200.

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

13. Ye C, Murano E, Stone M, Prince JL. A Bayesian approach to distinguishing interdigitated tongue muscles from limited diffusion magnetic resonance imaging. Computerized Medical Imaging and Graphics. 2015; 45:63–74. [PubMed: 26296155]

14. Ye C, Zhuo J, Gullapalli RP, Prince JL. Estimation of fiber orientations using neighborhood information. Medical Image Analysis. 2016; 32:243–256. [PubMed: 27209007]

15. Yeh FC, Wedeen VJ, Tseng WYI. Generalized q-sampling imaging. IEEE Transactions on Medical Imaging. 2010; 29(9):1626–1635. [PubMed: 20304721]
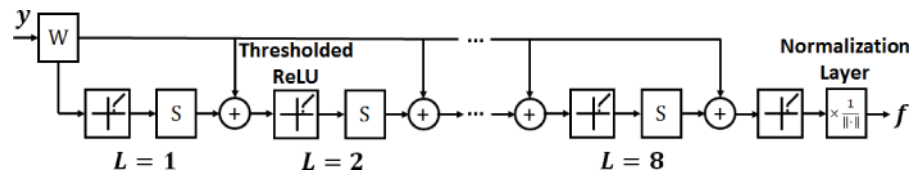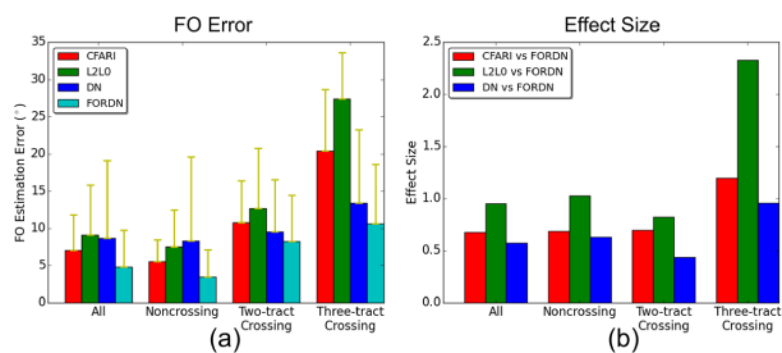
Fig. 1.
The structure of the deep network used in this work to guide FO estimation.

Fig. 2.
(a) Means and standard deviations of FO errors. (b) Effects sizes for the comparison between FORDN and other methods.
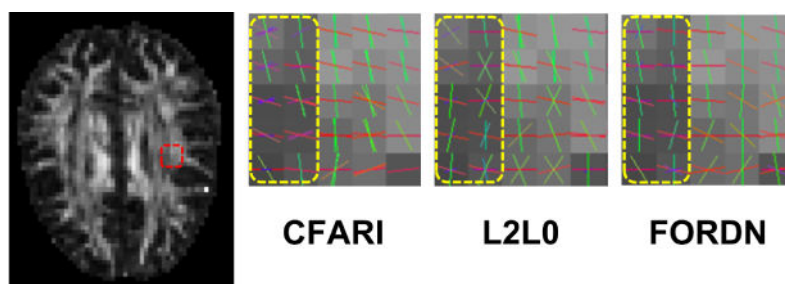
Fig. 3.
FO estimation results of brain dMRI overlaid on the FA map: an axial view of the crossing of the CC and SLF. Note the highlighted region for comparison.

Fig. 4.
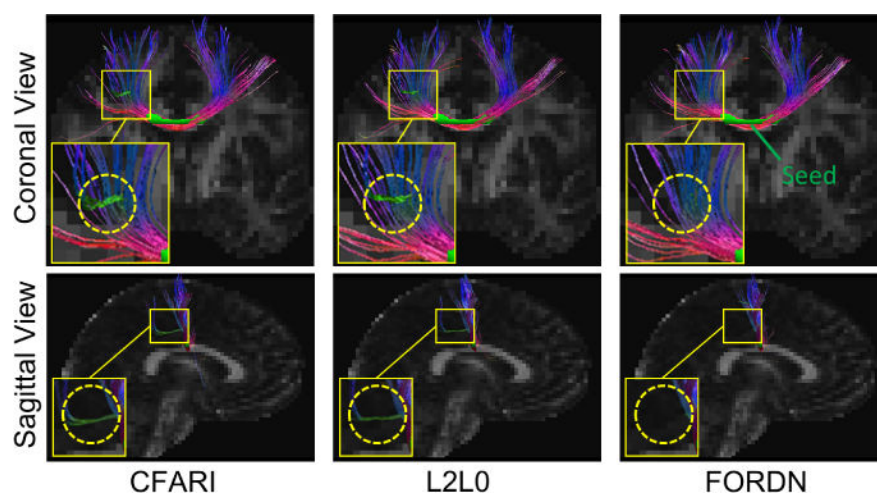Fiber tracking results seeded in the CC. Note the zoomed region for comparison.