HHS Public Access

Author manuscript

Med Image Comput Comput Assist Interv. Author manuscript; available in PMC 2018 July 20.

Published in final edited form as:

Med Image Comput Comput Assist Interv. 2015 October ; 9351: 132–139.

Brain Tissue Segmentation Based on Diffusion MRI Using ℓ_0 Sparse-Group Representation Classification

Pew-Thian Yap[†], **Yong Zhang**[‡], and **Dinggang Shen**[†]

[‡]Department of Radiology and Biomedical Research Imaging Center, The University of North Carolina at Chapel Hill, U.S.A.

[†]Department of Psychiatry & Behavioral Sciences, Stanford University, U.S.A.

Abstract

We present a method for automated brain tissue segmentation based on diffusion MRI. This provides information that is complementary to structural MRI and facilitates fusion of information between the two imaging modalities. Unlike existing segmentation approaches that are based on diffusion tensor imaging (DTI), our method explicitly models the coexistence of various diffusion compartments within each voxel owing to different tissue types and different fiber orientations. This results in improved segmentation in regions with white matter crossings and in regions susceptible to partial volume effects. For each voxel, we tease apart possible signal contributions from white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) with the help of diffusion exemplars, which are representative signals associated with each tissue type. Each voxel is then classified by determining which of the WM, GM, or CSF diffusion exemplar groups explains the signal better with the least fitting residual. Fitting is performed using ℓ_0 sparse-group approximation, circumventing various reported limitations of ℓ_1 fitting. In addition, to promote spatial regularity, we introduce a smoothing technique that is based on ℓ_0 gradient minimization, which can be viewed as the ℓ_0 version of total variation (TV) smoothing. Compared with the latter, our smoothing technique, which also incorporates multi-channel WM, GM, and CSF concurrent smoothing, yields marked improvement in preserving boundary contrast and consequently reduces segmentation bias caused by smoothing at tissue boundaries. The results produced by our method are in good agreement with segmentation based on T_1-weighted images.

1 Introduction

Brain tissue segmentation is most commonly performed using T_1-weighted images, which are typically rich with anatomical details thanks to their higher spatial resolution ($1 \times 1 \times 1$ mm^3). However, the recent availability of high spatial resolution ($1.25 \times 1.25 \times 1.25$ mm^3) diffusion MRI data from the Human Connectome Project[1] begs the following questions: 1) Can tissue segmentation be performed equally well solely based on diffusion data, therefore making it possible to avoid the technical difficulties involved in transferring segmentation information from T_1-weighted images, such as geometric distortion and cross-modality

Correspondence to: Pew-Thian Yap.

[1]http://www.humanconnectome.org/

registration? 2) Can diffusion data, acquired based on a totally different contrast mechanism, provide information complementary to T_1-weighted images for further improving segmentation?

In this paper, we attempt to address these questions by introducing a segmentation method that works directly with diffusion MRI data. In contrast to existing segmentation methods that are based on diffusion tensor imaging (DTI) [1–3], our method explicitly models the coexistence of various diffusion compartments within each voxel owing to different tissue types and different fiber orientations. This improves segmentation in regions with white matter crossings and in regions susceptible to partial volume effects. For each voxel, we tease apart possible signal contributions from white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) with the help of diffusion exemplars, which are representative signals associated with each tissue type. More specifically, the WM diffusion exemplars are sampled from diffusion tensors oriented in different directions with different axial and radial diffusivities; GM from isotopic tensors of low diffusivities; and CSF from isotropic tensors of high diffusivities. Each voxel is then classified by determining which of the WM, GM, or CSF diffusion exemplars explain the signal better with the least fitting residual.

Fitting is performed using ℓ_0 sparse-group approximation, circumventing various reported limitations of ℓ_1 fitting. The use of ℓ_0 penalization is motivated by the observations reported in [4], where the authors have shown that the commonly used ℓ_1-norm penalization [5, 6] conflicts with the unit sum requirement of the volume fractions and hence results in suboptimal solutions. To overcome this problem, the authors propose to employ the reweighted ℓ_1 minimization approached described by Candès et al. [7] to obtain solutions with enhanced sparsity, approximating solutions given by ℓ_0 minimization. However, despite giving improved results, this approach is still reliant on the suboptimal solution of the unweighted ℓ_1 minimization problem that has to be solved in the first iteration of the reweighted minimization scheme. In the current work, we will employ an algorithm that is based directly on ℓ_0 minimization.

To promote spatial regularity, we introduce a smoothing technique that is based on ℓ_0 gradient minimization [8]. This can be viewed as the ℓ_0 version of total variation (TV) smoothing. Compared with the latter, our smoothing technique yields marked improvement in the preservation of boundary contrast. In addition, our method smooths the probability maps of WM, GM, and CSF concurrently. This is achieved by an ℓ_0 adaptation of a multi-channel smoothing algorithm [9], solved using alternating direction method of multipliers (ADMM) [10].

2 Approach

Our approach to tissue segmentation is inspired by the face recognition work of Wright et al. [11]. However, instead of the ℓ_1 sparse approximation used in [11], we use a sparse-group ℓ_0 minimization approach that circumvents the problems mentioned in [4]. To promote spatial regularity, we also propose a multi-channel gradient minimization algorithm for smoothing of the tissue probability maps, producing edge-preserving effect better than smoothing based on TV regularization [12].

Linear Subspaces

We assume that the signal from each class of tissue lies in a linear subspace. The subspace is spanned by diffusion exemplars, which are hypothetical signal vectors generated using the tensor model, $S(b, \hat{\mathbf{g}}) = S_0 \exp(-b\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{D}\hat{\mathbf{g}})$, with varying diffusion parameters. Here, $\hat{\mathbf{g}}$ is a unit vector representing the gradient direction, S_0 is the baseline signal with no diffusion weighting, and $\mathbf{D}$ is the diffusion tensor. The WM subspace is spanned by multiple groups of diffusion exemplars. Each WM diffusion exemplar group consists of signal vectors sampled from a set of unidirectional axial-symmetric diffusion tensor models with a range of typical axial and radial diffusivities. Multiple groups of WM diffusion exemplars are generated by tensors with principal directions uniformly covering the unit sphere. The GM and CSF diffusion exemplar groups consist of signal vectors sampled from isotropic tensors with GM diffusivities set lower than CSF diffusivities, consistent with what was reported in [1, 13]. For each class $c \in \mathscr{C} = \{\mathrm{WM}, \mathrm{GM}, \mathrm{CSF}\}$, we arrange the n_c signal vectors of the diffusion exemplars as columns of a matrix $\mathbf{A}_c = [\mathbf{s}_{c,1}, \mathbf{s}_{c,2}, \ldots, \mathbf{s}_{c,n_c}]$. We then concatenate the exemplar matrices of all tissue classes into a matrix $\mathbf{A} = [\mathbf{A}_{\mathrm{WM}}|\mathbf{A}_{\mathrm{GM}}|\mathbf{A}_{\mathrm{CSF}}]$, where $\mathbf{A}_{\mathrm{WM}} = [\mathbf{A}_{\mathrm{WM}_1}|\ldots|\mathbf{A}_{\mathrm{WM}_k}|\ldots|\mathbf{A}_{\mathrm{WM}_{N_{\mathrm{WM}}}}]$ and each numerical subscript k of the WM exemplar matrix $\mathbf{A}_{\mathrm{WM}_k}$ denotes the index corresponding to a WM direction.

ℓ_0 Sparse-Group Representation

Given the signal vector $\mathbf{s}$ of a voxel that we wish to classify, we first compute its sparse-representation coefficient vector $\mathbf{f}$ by solving the follow ℓ_0 sparse-group approximation problem:

$$\min_{\mathbf{f} \geq 0} \left\{ \phi(\mathbf{f}) = \|\mathbf{A}\mathbf{f} - \mathbf{s}\|_2^2 + \gamma \left[\alpha \|\mathbf{f}\|_0 + (1-\alpha) \sum_{g \in \mathscr{G}} \mathscr{I}(\|\mathbf{f}_g\|_2) \right] \right\}, \quad (1)$$

where $\mathscr{I}(z)$ is an indicator function returning 1 if $z \neq 0$ or 0 if otherwise. The ℓ_0-"norm" gives the cardinality of the support, i.e., $\|\mathbf{f}\|_0 = |\,\mathrm{supp}(\mathbf{f})| = |\{k : f_k \neq 0\}|$. Parameters $\alpha \in [0, 1]$ and $\gamma > 0$ are for penalty tuning, analogous to those used in the sparse-group LASSO [14]. Note that $\alpha = 1$ gives the ℓ_0 fit, whereas $\alpha = 0$ gives the group ℓ_0 fit. $\mathbf{f}_g$ denotes the subvector containing the elements associated with group $g \in \mathscr{G} = \{\mathrm{WM}_1, \ldots, \mathrm{WM}_{N_{\mathrm{WM}}}, \mathrm{GM}, \mathrm{CSF}\}$. We solve this problem using an algorithm called non-monotone iterative hard thresholding (NIHT) [15], inspired by [16, 17]. Proof of convergence can be obtained by modifying the results shown in [17].

Tissue Classification

Each voxel is classified as the class with diffusion exemplars that best explain the signal. This is achieved, based on [11], by determining the class that gives the least reconstruction residual:

$$\min_c \{r(\mathbf{s}|c) = \|\mathbf{A}\boldsymbol{\delta}_c(\mathbf{f}) - \mathbf{s}\|_2\}, \quad (2)$$

where $\boldsymbol{\delta}_c(\mathbf{f})$ is a new vector whose only nonzero entries are the entries in $\mathbf{f}$ that are associated with class c. We modify the above problem to become a maximum a posteriori (MAP) estimation problem:

$$\max_c \{p(c|\mathbf{s}) \propto p(\mathbf{s}|c)p(c)\}, \quad (3)$$

where $p(c)$ is the prior probability and $p(\mathbf{s}|c)$ is the likelihood function defined as

$$p(\mathbf{s}|c) = \frac{1}{\sigma_c\sqrt{2\pi}} \exp\left[-\frac{r^2(\mathbf{s}|c)}{2\sigma_c^2}\right]. \quad (4)$$

The scale σ can be determined from the data via $\sigma_c^2 = \frac{1}{|\Omega_c|}\sum_{i\in\Omega_c} r^2(\mathbf{s}_i|c)$, where $\Omega_c \subset \Omega = \{1, \ldots, N\}$ is the subset of indices of voxels with class c giving the least residuals. N is total number of voxels. This alternative formulation allows us to visualize the posterior probability maps $\{p(c|\mathbf{s}_i)|i \in \Omega, c \in \mathscr{C}\}$ (disregarding constant scaling) for qualitative assessment of tissue segmentation. The prior probabilities can be set according to a pre-computed probabilistic atlas for guided segmentation.

Multi-Channel Gradient Minimization

Tissue classification as discussed in the previous section can be improved in terms of robustness by imposing spatial regularity. To achieve this, we smooth the posterior probability maps of WM, GM, and CSF concurrently prior to MAP estimation. In contrast to the commonly used TV-regularized smoothing, which is essentially an ℓ_1 gradient minimization (L1-GM) algorithm, we will use here ℓ_0 gradient minimization (L0-GM), which has been shown in [8] to be more effective than L1-GM in preserving edges. Moreover, L0-GM is more suitable in our case due to the piecewise constant nature of the segmentation maps. Here, we describe a multi-channel version of L0-GM.

We first define for the i-th voxel a probability vector $\mathbf{p}_i \equiv \mathbf{p}(\mathbf{s}_i) = [p(\text{WM}|\mathbf{s}_i), p(\text{GM}|\mathbf{s}_i), p(\text{CSF}|\mathbf{s}_i)]^{\text{T}}$. We then solve for a smoothed version of the probability map $\{\mathbf{p}_i \in \mathbb{R}^{|\mathscr{C}|}, i \in \Omega\}$, i.e., $\mathbf{u} = \{\mathbf{u}_i \in \mathbb{R}^{|\mathscr{C}|}, i \in \Omega\}$ via the following problem:

$$\min_{\mathbf{u}} \left\{\psi(\mathbf{u}) = \sum_i \|\mathbf{u}_i - \mathbf{p}_i\|_2^2 + \beta \sum_i \left\| \sqrt{\sum_d \|D_{i,d}\mathbf{u}\|_2^2} \right\|_0 \right\}. \quad (5)$$

We let $D_{i,d}\mathbf{u} \in \mathbb{R}^{1\times|\mathscr{C}|}$, where $\mathbf{u} \in \mathbb{R}^{N\times|\mathscr{C}|}$, be a row vector concatenating the finite difference values of all channels of $\mathbf{u}$ in the d-th spatial dimension. Note that $D_{i,d} \in \mathbb{R}^{1\times N}$ is the finite difference matrix. The first term in (5) maintains data fidelity and the second term penalizes small edges in a multi-channel image. If we replace the ℓ_0-"norm" in the second term with ℓ_1-norm, the above problem become a TV-regularized smoothing problem. Note that the above optimization problem is known to be computationally intractable. We thus implement an

approximate solution using ADMM [18] by introducing a number of auxiliary variables. The ADMM formulation amounts to repeatedly performing hard thresholding and spatial convolution/deconvolution [10].

3 Experiments

3.1 Data

Diffusion weighted (DW) datasets from the Human Connectome Project (HCP) [19] were used. DW images with $1.25 \times 1.25 \times 1.25$ mm^3 resolution were acquired with diffusion weightings $b = 1000$, 2000, and 3000 s/mm^2, each applied in 90 directions. 18 baseline images with low diffusion weighting $b = 5$ s/mm^2 were also acquired. The DW datasets were acquired with reversed phase encoding to correct for EPI distortion. T_1-weighted anatomical images were acquired as anatomical references.

3.2 Diffusion Parameters

The parameters of the tensors used to generate the diffusion exemplars were set to cover the typical values of the diffusivities of the WM, GM, and CSF voxels in the above dataset: $\lambda_{\parallel}^{\mathrm{WM}} = 1 \times 10^{-3}\ \mathrm{mm}^2/s, \lambda_{\perp}^{\mathrm{WM}} = [0.1:0.1:0.3] \times 10^{-3}\ \mathrm{mm}^2/s, \lambda^{\mathrm{GM}} = [0.00:0.01:0.80] \times 10^{-3}\ \mathrm{mm}^2/s$, and $\lambda^{\mathrm{CSF}} = [1.0:0.1:3.0] \times 10^{-3}\ \mathrm{mm}^2/s$. The notation $[a:s:b]$ denotes values from a to b, inclusive, with step s. Note that in practice, these ranges do not have to be exact but should however cover possible parameter values. The direction of each group of the WM diffusion exemplars corresponds to one of the 321 points evenly distributed on a hemisphere, generated by the subdivision of the faces of an icosahedron three times.

3.3 Comparison Methods

We compared the proposed method (**L200**) with the following methods:

- **L211**: Sparse-group LASSO [14] using diffusion exemplars identical to the proposed method. Similar to [4] and according to [7], we executed sparse-group LASSO multiple times, each time reweighing the ℓ_{21}-norm and the ℓ_1-norm so they eventually approximate their ℓ_0 counterparts.
- **L0**: ℓ_0 minimization using a single diffusion exemplar each for WM, GM, and CSF [13]. Similar to [13], WM-GM-CSF segmentation was used to help determine the parameters for the diffusion exemplars. The axial and radial diffusivities of the WM diffusion exemplars were determined based on WM voxels with fractional anisotropy (FA) greater than 0.7. The diffusivity of the isotropic GM/CSF diffusion exemplar was determined based on GM/CSF voxels with FA less than 0.2.

The tuning parameter γ was set to 1×10^{-4} for all methods. In addition, we set $\alpha = 0.05$, $\beta = 0.001$, $p(\mathrm{WM}) = 0.35$, $p(\mathrm{GM}) = 0.50$, and $p(\mathrm{CSF}) = 0.15$ for the proposed method.

3.4 Results

Qualitative—Figure 1 indicates that the segmentation result of the proposed method, L200, resembles very closely to that produced using the T_1-weighted image with the FSL FAST algorithm [20]. L211 produces WM segmentation result that is similar to L200, but underestimates GM. Note that these two methods are able to separate the deep GM structures, such as caudate and putamen, from the surrounding WM. The segmentation of the thalamus is more challenging because it is a mixture of GM and WM (see likelihood maps in the bottom row of Fig. 1).

Quantitative—Figure 2 shows the Dice scores for WM-GM-CSF segmentation of 5 subjects from the HCP data repository, confirming again that the proposed method produces segmentation results that agree most with segmentation based on T_1-weighted images. The average Dice scores for L200/L211/L0 are 0.8603/0.8581/0.8019 (WM), 0.8105/0.7177/0.6844 (GM), and 0.7204/0.5941/0.6985 (CSF).

Smoothing—Figure 3 shows the effects of smoothing with different strengths using L1-GM and L0-GM. The results confirm that despite the increased smoothing strength, L0-GM can still preserve edges effectively. On the other hand, L1-GM blurs the edges when the smoothing strength is increased.

4 Conclusion

In this paper, we have presented a tissue segmentation method that works directly with diffusion MRI data. We demonstrated that the proposed method is able to produce segmentation results that are in good agreement with the more conventional T_1-based segmentation. We also showed that diffusion MRI provides additional information for segmentation of deep gray matter structures, complementary to T_1-weighted imaging, where image contrast in this region is typically low. Future research will be directed to further improving the segmentation of deep gray matter.

Acknowledgments

This work was supported in part by a UNC BRIC-Radiology startup fund and NIH grants (EB006733, EB009634, AG041721, MH100217, AA010723, and 1UL1TR001111).

References

1Liu T, Li H, Wong K, Tarok A, Guo L, Wong ST. Brain tissue segmentation based on DTI data. NeuroImage. 2007; 38(1):114–123. [PubMed: 17804258]

2Awate SP, , Zhang H, , Simon TJ, , Gee JC. Multivariate segmentation of brain tissues by fusion of MRI and DTI data; IEEE International Symposium on Biomedical Imaging (ISBI); 2008 213216

3Lenglet C, , Rousson M, , Deriche R. A statistical framework for DTI segmentation; International Symposium on Biomedical Imaging; Apr, 2006 794797

4Daducci A, Ville DVD, 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; 18:820–833. [PubMed: 24593935]

5Yap PT, Shen D. Spatial transformation of DWI data using non-negative sparse representation. IEEE Transactions on Medical Imaging. 2012; 31(11):2035–2049. [PubMed: 22711770]

6Ramirez-Manzanares A, Rivera M, Vemuri BC, Carney P, Mareci T. Diffusion basis functions decomposition for estimating white matter intra-voxel fiber geometry. IEEE Transactions on Medical Imaging. 2007; 26(8):1091–1102. [PubMed: 17695129]

7Candès EJ, Wakin MB, Boyd SP. Enhancing sparsity by reweighted ℓ_1 minimization. Journal of Fourier Analysis and Applications. 2008; 14(5):877–905.

8Xu L, Lu C, Xu Y, Jia J. Image smoothing via ℓ_0 gradient minimization. ACM Transactions on Graphics. 2011; 30(5)

9Yang J, Yin W, Zhang Y, Wang Y. A fast algorithm for edge-preserving variational multichannel image restoration. SIAM Journal on Imaging Sciences. 2009; 2(2):569–592.

10Tao M, , Yang J. Technical report Department of Mathematics, Nanjing University; 2009Alternating direction algorithms for total variation deconvolution in image reconstruction.

11Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009; 31(2):1–18.

12Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Physica D. 1992; 60:259–268.

13Jeurissen B, Tournier JD, Dhollander T, Connelly A, Sijbers J. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. NeuroImage. 2014

14Simon N, Friedman J, Hastie T, Tibshirani R. A sparse-group lasso. Journal of Computational and Graphical Statistics. 2013; 22(2):231–245.

15Yap PT, Zhang Y, Shen D. Diffusion compartmentalization using response function groups with cardinality penalization. Medical Image Computing and Computer-Assisted Intervention. 2015

16Blumensath T, Davies ME. Iterative thresholding for sparse approximations. Journal of Fourier Analysis and Applications. 2008; 14(5–6):629–654.

17Lu Z. Iterative hard thresholding methods for ℓ_0 regularized convex cone programming. Mathematical Programming. 2013:1–30.

18Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trend in Machine Learning. 2010; 3(1):1–122.

19Essen DCV, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K. The WU-Minn human connectome project: An overview. NeuroImage. 2013; 80:62–79. [PubMed: 23684880]

20Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Transactions on Medical Imaging. 2001; 20(1):45–57. [PubMed: 11293691]

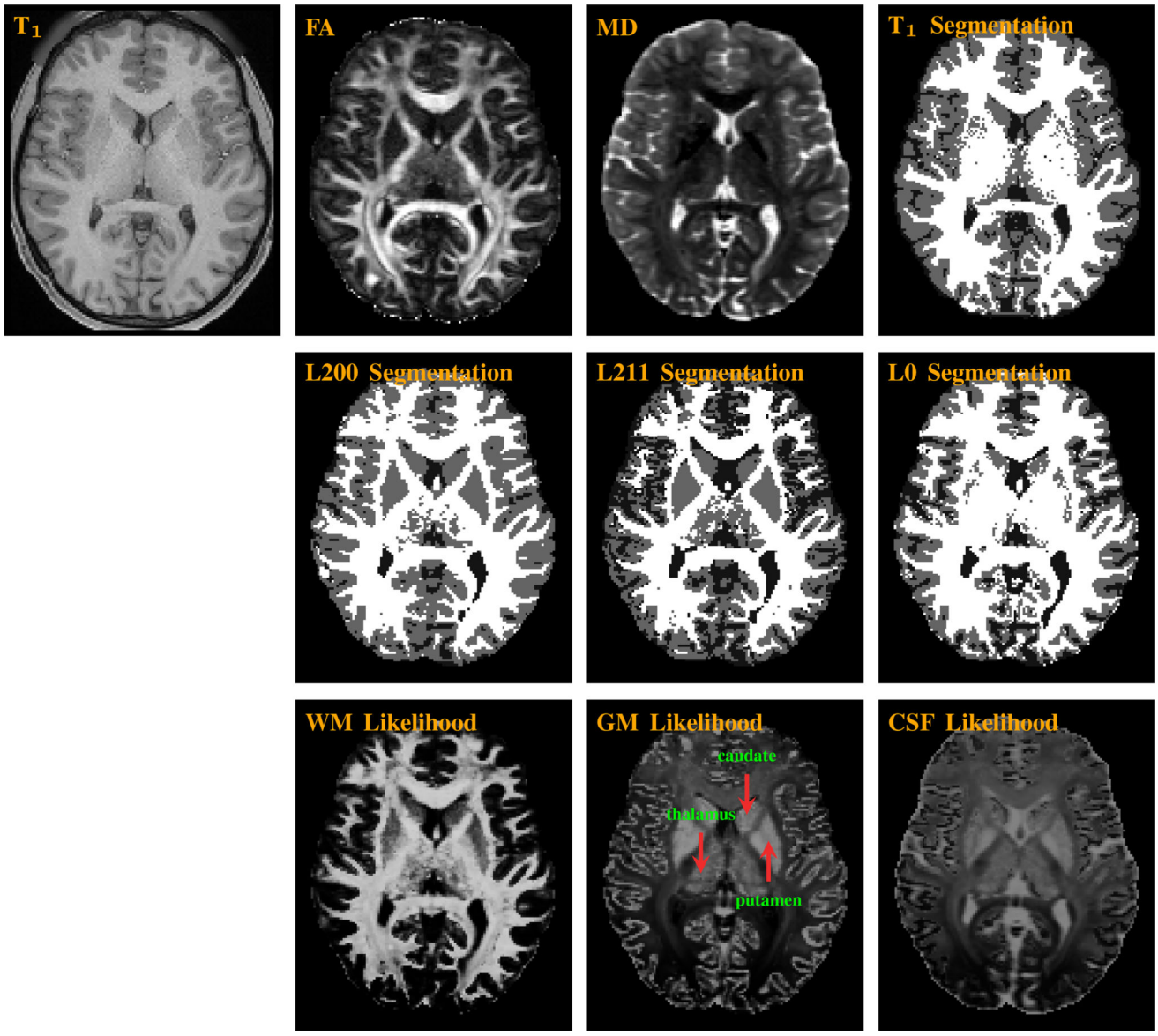

Fig. 1.

(Top) T_1-weighted, fractional anisotropy (FA), mean diffusivity (MD), and T_1 segmentation images. **(Middle)** Segmentation maps given by L200 (proposed), L211, and L0. **(Bottom)** Likelihood maps for WM, GM, and CSF given by L200.

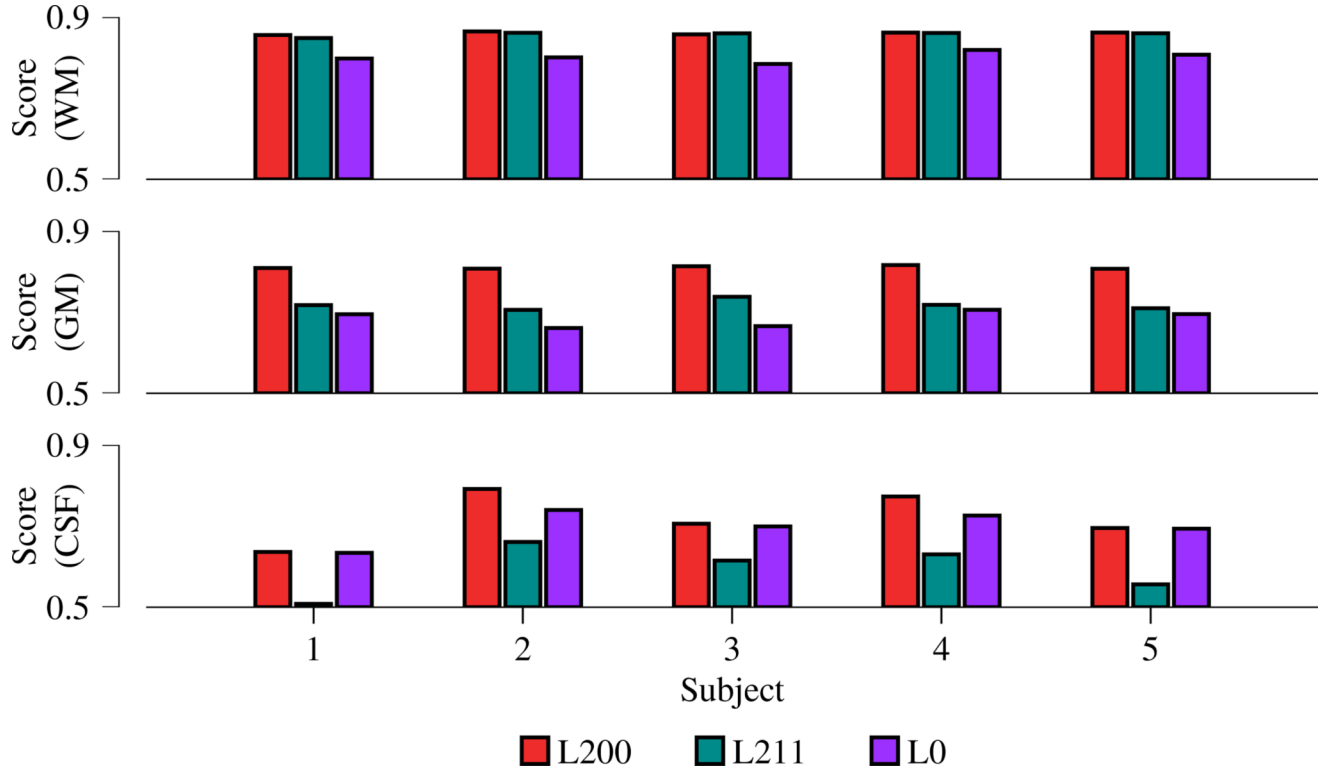

Fig. 2.

Accuracy of segmentation outcomes evaluated based on Dice score using T_1 segmentations as the ground truth.

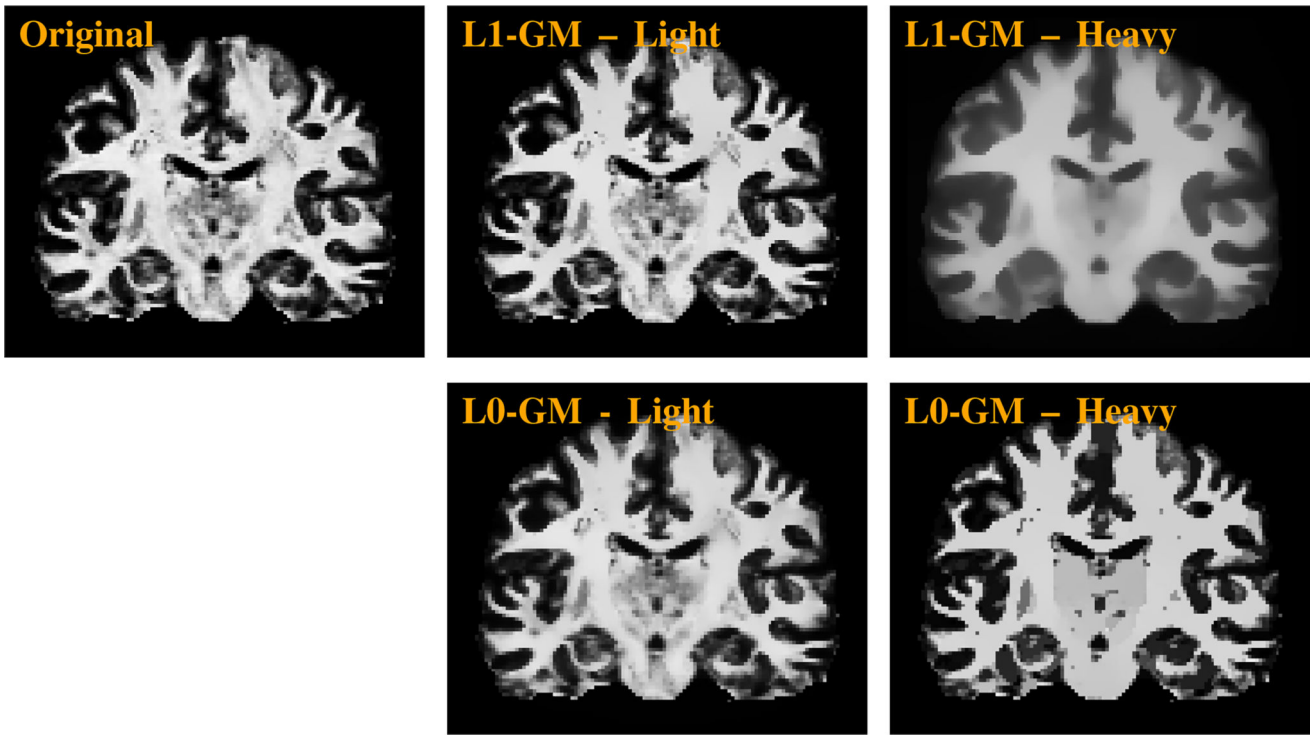

Fig. 3.

Effects of light and heavy smoothing using L1-GM and L0-GM. The WM, GM, CSF posterior probability maps are smoothed concurrently. However, due to space limitation, only the WM probability maps are shown here.