



Deep White Matter Analysis: Fast, Consistent Tractography Segmentation Across Populations and dMRI Acquisitions

Fan Zhang^(✉), Nico Hoffmann, Suheyla Cetin Karayumak, Yogesh Rathi, Alexandra J. Golby, and Lauren J. O'Donnell

Harvard Medical School, Brigham and Women's Hospital, Boston, USA
fzhang@bwh.harvard.edu

Abstract. We present a deep learning tractography segmentation method that allows fast and consistent white matter fiber tract identification across healthy and disease populations and across multiple diffusion MRI (dMRI) acquisitions. We create a large-scale training tractography dataset of 1 million labeled fiber samples (54 anatomical tracts are included). To discriminate between fibers from different tracts, we propose a novel 2D multi-channel feature descriptor (FiberMap) that encodes spatial coordinates of points along each fiber. We learn a CNN tract classification model based on FiberMap and obtain a high tract classification accuracy of 90.99%. The method is evaluated on a test dataset of 374 dMRI scans from three independently acquired populations across health conditions (healthy control, neuropsychiatric disorders, and brain tumor patients). We perform comparisons with two state-of-the-art white matter tract segmentation methods. Experimental results show that our method obtains a highly consistent segmentation result, where over 99% of the fiber tracts are successfully detected across all subjects under study, most importantly, including patients with space occupying brain tumors. The proposed method leverages deep learning techniques and provides a much faster and more efficient tool for large data analysis than methods using traditional machine learning techniques.

This work is supported by the following NIH grants: P41 EB015902, P41 EB015898, R01 MH074794, R01 MH097979, R01 MH119222, and U01 CA199459.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-32248-9_67) contains supplementary material, which is available to authorized users.

1 Introduction

Diffusion MRI (dMRI) allows the estimation of white matter fiber tracts in the brain via a process called tractography [1]. White matter tract segmentation, i.e. identifying tractography fibers (streamline trajectories) belonging to anatomically meaningful fiber tracts, is an essential step to enable tract quantification and visualization. Automated tract segmentation can enable the analysis of new, large dMRI datasets that are being acquired to study neural systems [7].

Recent studies have applied deep learning techniques to perform automated white matter tract segmentation (refer to [13] for a detailed review), which can be divided into voxel-based strategies and fiber-based strategies based on the input data provided to the network. The voxel-based strategies use a fiber orientation volume, where each voxel in this volume indicates the direction and the presence of a fiber tract. Simultaneous tractography and tract segmentation is performed by fiber tracking following the fiber directions in the predicted tract orientation volume [14, 16, 17]. For example, in the TractSeg method, fiber tracking is performed within a mask containing fiber orientation distribution function peaks [16, 17]. On the other hand, the fiber-based strategies directly segment tractography fibers by classifying them using neural networks trained on fiber feature descriptors [4–6, 8, 18]. Previously proposed fiber feature descriptors include curvature, torsion and distances to anatomical landmarks [8], and spatial coordinates of fiber points [4–6, 18].

Deep learning tract segmentation methods have achieved good performance, but challenges remain, particularly in generalization to dMRI data from different sources (e.g. across different acquisitions and across healthy and disease populations). Voxel-based methods leverage an intermediate tract orientation volume, which is robust to different acquisitions, so a trained tract segmentation model can be applied to data from multiple acquisitions [17]. However, for data from subjects with high anatomical variability, e.g., brain tumor patients, a voxel-based method could be problematic due to displacement of tracts by mass effect from the tumor as well as effects of surrounding edema. Fiber-based strategies based on fiber clustering techniques can handle this large across-subject anatomical variability using traditional machine learning approaches [3, 10, 15, 19]. But, such methods usually require multiple time-consuming processing steps.

The goal of this study is to propose a fiber-based deep learning method (Fig. 1) for fast and consistent white matter tract segmentation across healthy and disease populations, as well as different dMRI acquisitions. This paper has three main contributions. First, we create a large-scale training dataset of a million labeled tractography fibers from 100 subjects, including fibers from anatomical fiber tracts and those from false positive tracking. Second, we propose a novel 2D multi-channel fiber feature descriptor (FiberMap) that is insensitive to the order of points along fibers and is robust to fiber local region differences (e.g. due to effects of tumors and edema). Third, we demonstrate successful tract segmentation on a large test dataset (374 subjects). We believe the proposed method is the first fiber-based deep learning tract segmentation method that can gen-

eralize to dMRI data with different acquisition parameters and from different populations, including brain tumor patients.

2 Methods

2.1 Datasets

Training Dataset: We created a large tractography training dataset of 1 million fiber samples labeled into 54 anatomical fiber tracts (see Supplementary Table S1 for the full tract list), as well as a tract category of “other fibers” including, most importantly, those from false positive tracking. To do this, we leveraged the ORG-800FC-100HCP white matter tractography atlas [19] (Fig. 1(a)). This atlas includes clustered and neuroanatomically expert labeled tractography data (computed using a two-tensor unscented Kalman filter (UKF) method [9]) from 100 healthy Human Connectome Project (HCP) [2] subjects. In the present study, we used training fiber samples from a total of 54 tracts of interest, such as arcuate fasciculus (AF) and corticospinal tract (CST), for a total of 273379 fibers. We grouped the fibers from all other clusters and the rejected false positive fibers into the category of “other fibers” (a total of 726621 fibers). In total, we had 55 tract classes in the training dataset. We note that false positive fibers were not provided in the ORG atlas. Therefore, we computed these fibers from whole-brain UKF tractography from each of the 100 training HCP subjects by applying atlas-based clustering followed by outlier fiber identification [19].

Three Independently Acquired Test Datasets (374 Subjects): (1) HCP test dataset: dMRI data from 100 HCP subjects (different from the ones included

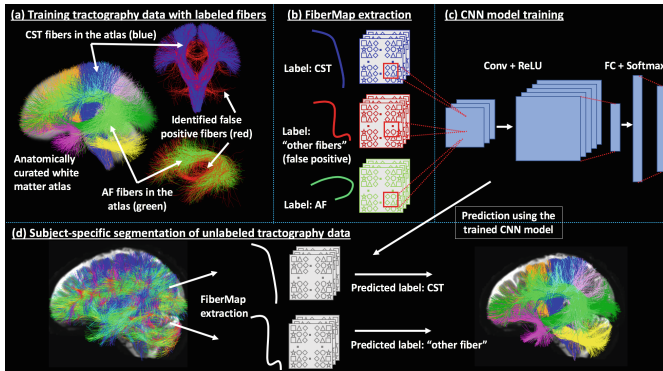


Fig. 1. Method overview. A training tractography dataset (a) is created, where each fiber is associated with an anatomical tract label (e.g. AF and CST) or a category of “other fibers,” including false positive fibers (in red). A 2D multi-channel feature descriptor (FiberMap) (b) is extracted for each fiber. A CNN tract classification model (c) is trained based on the FiberMap descriptors. Subject-specific tract segmentation (d) is performed by FiberMap feature extraction of each fiber, followed by tract label prediction using the trained CNN model. (Color figure online)

in the training dataset) with 18 $b = 0$ and 90 $b = 3000$ images, TE/TR = 89/5520 ms, resolution = $1.25 \times 1.25 \times 1.25 \text{ mm}^3$. (2) Consortium for Neuropsychiatric Phenomics (CNP) test dataset: dMRI data from 41 ADHD patients, 49 bipolar disorder patients, 50 schizophrenia patients and 125 healthy controls from the CNP [12], with 1 $b = 0$ and 63 $b = 1000$ images, TE/TR = 93/9000 ms, resolution = $2 \times 2 \times 2 \text{ mm}^3$. (3) BTP test dataset: dMRI data from 39 brain tumor patients (BTPs) acquired at Brigham and Women’s Hospital, Boston, with 1 $b = 0$ and 30 $b = 2000$ images, TE/TR = 98/12700 ms, resolution = $2.3 \times 2.3 \times 2.3 \text{ mm}^3$. We computed whole brain tractography (Fig. 1(d)) including about 1 million fibers per subject, using the same two-tensor UKF method as in generating the training data. (See Supplementary Fig. S1 for additional tract segmentation results from tractography data generated using other fiber tracking methods)

2.2 FiberMap Tractography Fiber Feature Descriptor

A fiber feature descriptor should discriminate between fibers from different tracts, while handling the following challenges. First, a good descriptor should capture the similarity between fibers regardless of the order of points along the fibers. For example, a fiber in CST could be tracked either from the cortex to the brainstem or from the brainstem to the cortex. Second, anatomical variability across subjects can result in local fiber differences within a tract. A descriptor should properly capture the similarity of such fibers to enable a tract segmentation method to generalize to data from different populations. Considering these two challenges, we propose a new feature descriptor for tractography fibers, which we refer to as FiberMap. FiberMap represents a fiber streamline as a 2D feature map with 3 channels that encode the spatial coordinates of points along the fiber. The FiberMap is computed by repeating and flipping the coordinates of the points along a fiber (as described in detail in the caption of Fig. 2). Due to this repetition, FiberMap is relatively insensitive to the order of points along a fiber (Fig. 2(b)) and robust to local fiber differences (Fig. 2(c)). In addition, the FiberMap descriptor is analogous to a 2D RGB image (Fig. 2(d)) for easy input to CNNs. In our experiments, we extracted $n = 15$ points per fiber, a reasonable number for fiber representation [11], leading to a $30 \times 30 \times 3$ dimensional FiberMap feature descriptor for each fiber (analogous to a square RGB image).

2.3 CNN Tract Segmentation Model Training

After extracting the FiberMap descriptor of each fiber in the training tractography data, we train a CNN model for tract segmentation. We tested multiple kernel sizes and layers and here we present the most successful architecture. This network contains 4 convolutional layers (32, 64, 128 and 256 filters, respectively) with kernel size 3, where each convolutional layer is followed by a ReLU activation layer. A max pooling layer of size 2 and a dropout layer of 2.5 are used to prevent overfitting. The last convolutional layer is followed by 3 fully connected layers of size 128, 256 and 512, and then a softmax layer with 55 outputs for

the 54 anatomical tracts and the category of “other fibers.” RMSprop is used for optimization with a loss function of categorical cross-entropy. The CNN is implemented using TensorFlow (1.12.0). In our experiments, we split the training dataset, where 80% of the fibers were used for training and 20% for validation.

2.4 Tract Segmentation of Unlabeled Tractography Data

First, subject-specific tractography is transferred to the atlas space by performing an affine registration between the baseline ($b = 0$) image of the subject (moving image) and the atlas population mean T2 image (reference image) using 3D Slicer (<https://www.slicer.org>). The obtained transform is applied to the subject-specific tractography data. Next, FiberMap feature extraction is performed and each fiber is classified based on the trained CNN model (Fig. 2(d)). Source code and the trained CNN model will be available online at <https://github.com/SlicerDMRI/DeepWMA>.

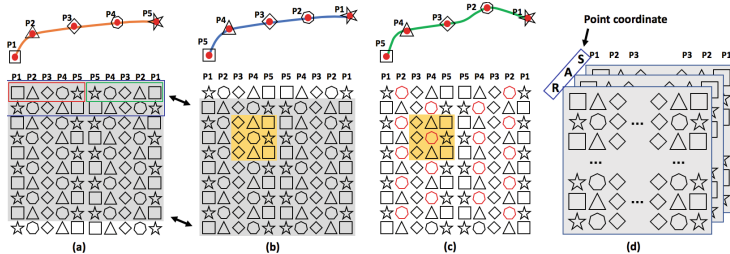


Fig. 2. (a) illustrates the proposed FiberMap feature descriptor. Given a fiber composed of a sequence of points (P1 to P5; red rectangle), a flipped copy (green rectangle) is created by reversing the order of the points. This flipped copy is repeated twice, once to the right and once below, and the original sequence is repeated once below to the right. These steps result in a two-row sequence of points (blue rectangle), which is further repeated to generate a 2D map, i.e. the FiberMap descriptor. The repetition is performed n times (i.e. the number of points per fiber, $n = 5$ in this example), which results in a square feature map. (b) shows an example of the insensitivity of FiberMap to the order of points along a fiber. If the original sequence of points along the fiber in (a) is reversed to that in (b), the FiberMap descriptors of these two fibers are the same except for a difference of one row (the part that is the same between the two FiberMap descriptors is highlighted in gray). (c) shows an example of the robustness of FiberMap to local fiber differences. The fiber in (c) is similar to that in (b), but with a local difference at point P2. Because the neighboring points of P2 on FiberMap (highlighted in yellow in (b) and (c)) are similar, the difference at P2 will have a small influence on the computation of fiber similarity. (d) illustrates that FiberMap includes three channels, analogous to a 2D RGB image. The channels encode the spatial Right-Anterior-Superior (RAS) coordinates of the fiber points. (Color figure online)

2.5 Experimental Evaluation

Comparison of Fiber Feature Descriptors: We compared the following fiber feature descriptors: a 1D-RAS descriptor that concatenated the RAS coordinates of all points along each fiber into a 1D vector (size: $3n \times 1$), a 2D-RAS descriptor that concatenated the RAS coordinates of all points along each fiber into a 2D matrix (size: $n \times 3$) [18], a CurTor descriptor that concatenated curvature and torsion at each point along a fiber (size: $n \times 2$) [8], a 2D-RAS+CurTor descriptor that concatenated the CurTor and Orig-RAS descriptors (size: $n \times 5$) [18], and the proposed FiberMap descriptor (size: $2n \times 2n \times 3$). $n = 15$ points per fiber was used across all descriptors. For each compared descriptor, a CNN classification model was used and the parameters were well tuned. The overall fiber classification accuracy (i.e. the percentage of fibers that were correctly classified into their ground truth tract category), as well as the mean recall and the mean precision across all tract categories, were compared across the methods using the same training and validation splits of the training tractography dataset.

Comparison of State-of-the-art Methods: We evaluated our proposed method with comparison to a fiber-based method using spectral clustering (WMA) [11, 19] and a voxel-based method using CNN (TractSeg) [16, 17]. We conducted a quantitative evaluation by measuring the percentage of tracts that were successfully detected for each test dataset under study. We note that WMA performed tractography segmentation by applying the same anatomically curated atlas (see [19] for details) as we used for generating our training data, while TractSeg was built upon different training data that defined different sets of white matter tracts. Therefore, we used 34 tracts that were commonly defined across all three methods for this evaluation. A tract was considered to be successfully detected if there were at least 20 fibers (as proposed in [19]). This metric of

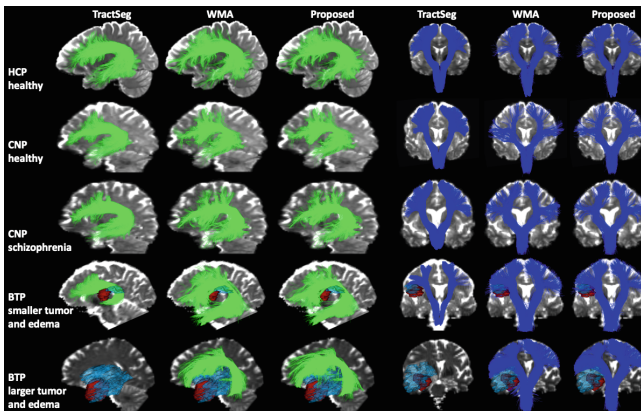


Fig. 3. Visual comparison of two example tracts (AF and CST) obtained using TractSeg, WMA and the proposed method. Results from five subjects from the HCP, the CNP and the BTP datasets were selected for visualization.

successful detection could be straightforwardly applied to all methods and was not affected by any small differences in the shape or location of anatomical tracts as defined by different methods. In addition, we performed a visual inspection of tracts obtained across the compared methods.

3 Experimental Results

Comparison of Fiber Feature Descriptors (Table 1): The 1D-RAS and CurTor descriptors had low fiber classification accuracies (around 50%), while the 2D-RAS and 2D-RAS+CurTor descriptors obtained higher accuracies (around 87%). The FiberMap descriptor obtained the highest accuracy (about 91%), as well as the highest mean recall and precision (85.67% and 88.47%, respectively).

Comparison of State-of-the-art Methods (Table 2, Fig. 3): Table 2 shows the percentage of successfully detected tracts for each method. For the HCP and CNP datasets, all three methods performed well, where over 98.8% of the tracts could be successfully detected. For the BTP dataset, the proposed method performed better than the other two methods, by detecting over 99% of the tracts. Additionally, Fig. 3 gives a visual comparison of example tracts obtained for each method. The visual/qualitative performance of all methods was reasonable for AF and CST tracts in the healthy HCP subject, the healthy CNP subject, the CNP subject with schizophrenia, and the brain tumor patient with relatively small tumor and edema. However, unlike the two fiber-based approaches, the voxel-based TractSeg method did not detect the AF and CST tracts in the patient with larger tumor and edema.

Table 1. Comparison across different fiber feature descriptors.

	1D-RAS	2D-RAS	CurTor	2D-RAS+CurTor	FiberMap
Accuracy	47.26%	87.18%	48.98%	7.06%	90.99%
Recall	22.54±15.46%	75.57±10.8%	34.16±17.67%	75.22±10.65%	85.67±5.72%
Precision	1.70±0.92%	87.53±4.99%	1.7±0.63%	86.38±7.38%	88.47±5.19%

4 Discussion

We demonstrated that the proposed FiberMap feature descriptor improved tract segmentation performance compared to several other descriptors from the literature. The 1D-RAS descriptor was not successful, while the 2D-RAS descriptor improved classification performance. This was most likely because the CNN filters could access the 3 RAS coordinates of a point and those of its neighbors together when using the 2D-RAS descriptor. The CurTor descriptor obtained very low classification accuracy, and combining it with the 2D-RAS descriptor did not achieve improvement compared to using 2D-RAS only. This could

be because fibers from different tracts could have highly similar curvature and torsion values (e.g. different corpus callosum tracts), causing the fiber feature descriptors to be less discriminative, which is in line with the finding in [18]. Using the proposed FiberMap descriptor, an accuracy of over 90% was obtained, which was the highest across all compared descriptors.

We performed a visual inspection of the misclassified fibers for each tract in our method and found that most of them were at the boundary between the annotated anatomical fibers in the atlas and the false positive fibers. Such fibers are difficult, even impossible, to be definitively delineated in tractography because of a lack of ground truth. Therefore, we believe that the obtained 90.99% accuracy represents a good whole brain tractography segmentation performance. An additional preprocessing step could be applied during fiber tracking to improve tractography, e.g., filtering out false positive streamlines as applied in [18], and then applying this data as input to our tract segmentation model. However, such preprocessing would need additional anatomical information, e.g., from a T1-weighted image, as well as inter-modality image registration.

In this work, we demonstrated the first fiber-based deep learning method that allowed consistent tract segmentation across multiple dMRI acquisitions and across different populations including patients with space occupying brain tumors. (See supplementary Figs. S2 and S3 for additional results of all 54 tracts of one example HCP subject and the left AF tracts of all 39 brain tumor patients.) In related work, only one voxel-based deep learning method (i.e. TractSeg) was demonstrated to be able to generalize to datasets acquired with different scanners and with multiple pathologies [16,17]. Our results agreed with this finding in the HCP and CNP datasets. (Additional results to show that the proposed method has a good spatial coverage relative to TractSeg are provided in Supplementary Table S2.) Nevertheless, improved performance was obtained on the BTP dataset using our fiber-based deep learning approach by successfully detecting fiber tracts that were affected by tumors and edema. One possible reason could be that our fiber-based method worked directly on tractography data obtained using a two-tensor fiber tracking method that has been demonstrated to be sensitive in tracking through peritumoral edema [19].

The proposed method provides a fast and efficient tool for large-scale dMRI analysis. While achieving similar results to the WMA method in terms of the segmentation consistency, the proposed method performed much faster by leveraging GPU computation. For segmentation of whole brain tractography data that contained a million fibers, the WMA method (including tractography registration, fiber clustering, and tract identification) took about 1.5 h using 4 CPU cores, while the proposed method (including volume registration, FiberMap extraction, and CNN classification) required a shorter time of 41 min on the same 4-core CPU, or 8 min using the 4-core CPU plus 1 GPU.

Table 2. Percentage of tracts that were successfully detected in each method.

Dataset	TractSeg	WMA	DeepWMA
HCP	100.0%	100.0%	100.0%
CNP	98.87%	99.35%	99.66%
BTP	94.42%	98.27%	99.17%

5 Conclusion

We present a deep learning tractography segmentation method that allows fast and consistent tract segmentation across dMRI acquisitions and different populations. Future work could include an investigation of more advanced network architectures. In addition, a more comprehensive training dataset could be created, e.g., further subdividing “other fibers” into anatomically meaningful tracts (e.g. fornix, anterior commissure, and superficial fiber tracts).

References

1. Basser, P.J., et al.: In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.* **44**(4), 625–632 (2000)
2. Essen, D.C.V., et al.: The WU-Minn human connectome project: an overview. *NeuroImage* **80**, 62–79 (2013)
3. Garyfallidis, E., et al.: Recognition of white matter bundles using local and global streamline-based registration and clustering. *NeuroImage* **170**, 283–295 (2018)
4. Gupta, T., et al.: BrainSegNet: a segmentation network for human brain fiber tractography data into anatomically meaningful clusters. [arXiv:1710.05158](https://arxiv.org/abs/1710.05158) (2017)
5. Gupta, V., Thomopoulos, S.I., Rashid, F.M., Thompson, P.M.: FiberNET: an ensemble deep learning framework for clustering white matter fibers. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 548–555. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_63
6. Gupta, V., et al.: FiberNet 2.0: an automatic neural network based tool for clustering white matter fibers in the brain. In: ISBI, pp. 708–711 (2018)
7. Harms, M.P., et al.: Extending the human connectome project across ages: imaging protocols for the lifespan development and aging projects. *NeuroImage* **183**, 972–984 (2018)
8. Lam, P.D.N., et al.: TRAFIC: fiber tract classification using deep learning. In: *Processing SPIE Medical Imaging*, p. 1057412 (2018)
9. Malcolm, J., et al.: Filtered multitensor tractography. *IEEE Trans. Med. Imaging* **29**(9), 1664–1675 (2010)
10. O’Donnell, L.J., et al.: Automated white matter fiber tract identification in patients with brain tumors. *NeuroImage Clin.* **13**, 138–153 (2017)
11. O’Donnell, L.J., Westin, C.F.: Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Trans. Med. Imaging* **26**(11), 1562–1575 (2007)
12. Poldrack, R., et al.: A phenome-wide examination of neural and cognitive function. *Sci. Data* **3**, 160110 (2016)
13. Poulin, P., et al.: Tractography and machine learning: Current state and open challenges. *Magnetic Resonance Imaging* **2**, 160110 (2019)
14. Reisert, M., et al.: HAMLET: hierarchical harmonic filters for learning tracts from diffusion MRI. [arXiv:1807.01068](https://arxiv.org/abs/1807.01068) (2018)
15. Tunc, B., et al.: Individualized map of white matter pathways: connectivity-based paradigm for neurosurgical planning. *Neurosurgery* **79**(4), 568–577 (2015)
16. Wasserthal, J., et al.: TractSeg - fast and accurate white matter tract segmentation. *NeuroImage* **183**, 239–253 (2018)

17. Wasserthal, J., et al.: Combined tract segmentation and orientation mapping for bundle-specific tractography. [arXiv:1901.10271](https://arxiv.org/abs/1901.10271) (2019)
18. Xu, H., et al.: Objective detection of eloquent axonal pathways to minimize post-operative deficits in pediatric epilepsy surgery using diffusion tractography and convolutional neural networks. *IEEE Trans. Med. Imaging* **38**(8), 1910–1922 (2019)
19. Zhang, F., et al.: An anatomically curated fiber clustering white matter atlas for consistent white matter tract parcellation across the lifespan. *NeuroImage* **179**, 429–447 (2018)