

Corpus Callosum Segmentation in Brain MRIs via Robust Target-Localization and Joint Supervised Feature Extraction and Prediction

Lisa Y.W. Tang^(✉), Tom Brosch, XingTong Liu, Youngjin Yoo,
Anthony Traboulsee, David Li, and Roger Tam

MS/MRI Research Group, University of British Columbia, Vancouver, BC, Canada
lisa.yw.tang@gmail.com

Abstract. Accurate segmentation of the mid-sagittal corpus callosum as captured in magnetic resonance images is an important step in many clinical research studies for various neurological disorders. This task can be challenging, however, especially more so in clinical studies, like those acquired of multiple sclerosis patients, whose brain structures may have undergone significant changes, rendering accurate registrations and hence, (multi-) atlas-based segmentation algorithms inapplicable. Furthermore, the MRI scans to be segmented often vary significantly in terms of image quality, rendering many generic unsupervised segmentation methods insufficient, as demonstrated in a recent work. In this paper, we hypothesize that adopting a supervised approach to the segmentation task may bring a break-through to performance. By employing a discriminative learning framework, our method automatically learns a set of latent features useful for identifying the target structure that proved to generalize well across various datasets, as our experiments demonstrate. Our evaluations, as conducted on four large datasets collected from different sources, totaling 2,033 scans, demonstrates that our method achieves an average Dice similarity score of 0.93 on test sets, when the models were trained on at most 300 images, while the top-performing unsupervised method could only achieve an average Dice score of 0.77.

1 Introduction

Accurate and robust segmentation of the corpus callosum in brain magnetic resonance images (MRI) is a crucial step in many automatic image analysis tasks, with clinical applications like computer-aided diagnosis and prognosis of several neurological diseases [1, 6, 10, 16].

The segmentation task is often challenging, however, especially when dealing with some clinical studies of chronic diseases, like those of multiple sclerosis patients, whose brain structures may have undergone significant changes over time, thereby rendering registrations of subject images to pre-segmented brain templates inaccurate [12, 18]. Consequently, the standard approach of employing (multi-) atlas-based segmentation algorithms may fail, as reported in recent

publications [15, 18]. Furthermore, the MRI scans to be segmented often vary significantly in terms of image quality, rendering many unsupervised segmentation algorithms that often require delicate image-dependent parameter-tuning impractical when a large dataset of images must be segmented.

As demonstrated in [18], a multi-atlas-based algorithm custom-designed for the segmentation of the corpus callosum achieved a segmentation accuracy of 0.77 in Dice similarity score (DSC), as evaluated on a challenging clinical dataset, while those of other generic multi-atlas-based methods, like weighted majority voting and STEPS [8], achieved an average DSC of 0.65 to 0.70.

In view of the availability of several public image datasets [12, 14], where some are accompanied with expert segmentations, we thus hypothesize that adopting a supervised approach [5, 7] to the segmentation task may provide a better alternative to existing unsupervised methods.

More specifically, we propose a two-stage process to improve the efficiency (of model training) and accuracy of the final segmentation. In the first stage, we employ multi-atlas based segmentation to localize the CC structure. Because neighbouring structures of the CC provide adequate contextual information, our localization step is shown to be tolerant to imaging artifacts and shape variations caused by significant brain atrophy, and is sufficient for CC localization, which requires less precision than the actual CC segmentation task. In the second stage, we apt a discriminative learning approach where we employ a subset of the available expert segmentations from a single dataset to train a neural network architecture called convolutional encoder network (CEN) [7] that performs feature extraction and model parameter training in a joint manner.

Our two-stage pipeline is motivated as follows. Rather than training a CEN that would extract features from the entire image slice, we propose to examine image features only within the bounding box of the target structure. By first performing target-localization to identify the region-of-interest (ROI) of the target structure, our two-stage approach explicitly constrains the network to learn and extract features only around the periphery of the CC, thereby yielding features that are more relevant to differentiating the CC structure from its neighbouring vessels, which, as similarly reported in [18], often brought confusion to existing unsupervised segmentation algorithms.

To this end, our main contributions are as follows. We (1) propose a novel pipeline for a fully automatic CC segmentation algorithm; (2) performed comparative analysis with existing unsupervised methods; and (3) conducted extensive evaluation of our method using 4 datasets involving over 2,000 scans with different image characteristics.

2 Methods

2.1 ROI Localization

In localizing the CC structure of each image, we employ a multi-atlas based segmentation approach. This step is done both at training and at test time. Our localization process involves performing registrations of 5 segmented atlases

from the public dataset of Heckemann et al. [12]. This allows us to transfer the CC segmentation label of each atlas onto the target image to be segmented. After registrations, we perform weighted majority voting on the transferred atlas labels and threshold the casted votes to define a rough outline of the CC in each image. To give us a confident estimation of the CC region, we employ a relatively high threshold value of 0.7 and set the weight of the vote of each atlas to be proportional to the average local correlation between the registered atlas and the subject image being segmented.

For registrations, we employ a coarse-to-fine, multi-stage (rigid to affine) registration pipeline as provided through the ANTS registration toolkit [2]. Initial experiments demonstrated that the localization results are insensitive to the choice of similarity measure used. We chose to use the mutual information measure due to its low computational and time requirements (as opposed to the use of local cross-correlation measure). In calculating the similarity measure, we also employ stochastic sampling to further increase computational efficiency.

2.2 Joint Feature-Learning and Model Training

Let there be a training set of N images $\mathcal{I} = \{I\}_{n=1}^N$ and corresponding CC segmentations $\mathcal{S} = \{S\}_{n=1}^N$. Following the supervised approach of [7], our goal is to seek a segmentation function f that maps input image I_n to its corresponding CC segmentation S_n :

$$\hat{f} = \arg \min_{f \in \mathcal{F}} \sum_n E(S_n, f(I_n)), \quad (1)$$

where \mathcal{F} is the set of possible segmentation functions, and E is an error function that measures the sum of squared differences between S_n and the segmentation predicted by f .

As done in previous work [7], \mathcal{F} is modeled by a convolutional encoder network (CEN) with shortcut connections, whose model architecture is divided into the *convolutional* pathway and the *deconvolutional* [5] pathways. The convolutional pathway is designed to automatically learn a set of features from \mathcal{I} and is constructed with alternating convolutional and pooling layers, where the former convolves its input signal with convolutional filter kernels, and the latter performs averaging over blocks of data from the convolved signal. Conversely, the deconvolutional pathway is designed to reconstruct the segmentation mask using inputs from the convolutional pathway and consists of alternating deconvolutional and unpooling layers. In short, the convolutional pathway learns a hierarchical set of low-level to high-level image features while the deconvolutional pathway learns to predict segmentations using the learned features extracted from the convolutional pathway.

As detailed in [3, 7], the use of low-level and high-level image features respectively provides the means to increase localization precision and provide contextual information to guide segmentation. Hence, to integrate these hierarchical

features, the two pathways in the CEN model [7] are further linked with “short-cut connections”, via which the activations of the convolutional pathway and those of the deconvolutional pathway become connected [7].

2.3 Details on Implementation and Model-Training

We trained a 2-layer CEN where, in each layer, we employ 32 filters of size of 9×9 and 2×2 respectively for the convolutional and pooling layers. For the hidden units, we employ an improved version of the rectified linear units [7]. We stopped training when the training error converged, which was usually around 800 epochs. To avoid overfitting [7], we also employ the dropout technique with dropout rates of 0.5–0.75.

Further, recent research [7, 17] has shown that pre-training of the CEN model parameters with adequate fine-tuning improves accuracy across a variety of application domains. Accordingly, in a similar manner as done in [7], we performed pre-training on the input images layer by layer using 3 layers of convolutional restricted Boltzmann machines (convRBMs) [13]. More specifically, the first layer is trained on the input images and each subsequent layer is trained on the hidden activations of its previous layer. The model parameters of our CENs are then initialized using the weights obtained from this pre-training step.

In training the convRBMs, we employ the AdaDelta algorithm and explored various weight-initializations for the convRBMs model parameters, i.e. with mean of zero and standard deviation of $\{0.1, 0.03, 0.01, 0.003, 0.001\}$, which was determined using a logarithmic scale for the condition that the images have been normalized to have mean and standard deviation of 0 and 1, respectively. We also explored different settings for the hyper-parameter in AdaDelta and optimized this parameter empirically (i.e. $\epsilon = \{1e-8, 1e-9, 1e-10, 1e-11\}$).

3 Experiments and Results

Materials. The following datasets of MR T1w were retrospectively collected for this work. Due to the retrospective nature, two protocols were used to generate segmentations of the CC for both training and testing purposes: manual vs. semi-automatic.

MS (in-house): 348 patients with secondary progressive multiple sclerosis (SPMS) were enrolled in a clinical trial. Each subject was scanned at 4 different timepoints: at screening, baseline (~ 4 weeks after screening), year one, and year two. This resulted in a total of $280 + 270 + 280 + 300 = 1130$ scans. Registration of each scan to the template space and automatic extraction of the mid-sagittal slice from each volume was performed. Then, the area of the CC was manually outlined by a trained medical student who was blinded to the chronological sequence of the four timepoints.

CIS (in-house): 140 subjects with clinical isolated syndrome (a prodromal phase of MS) were enrolled in another clinical trial that was completed approximately

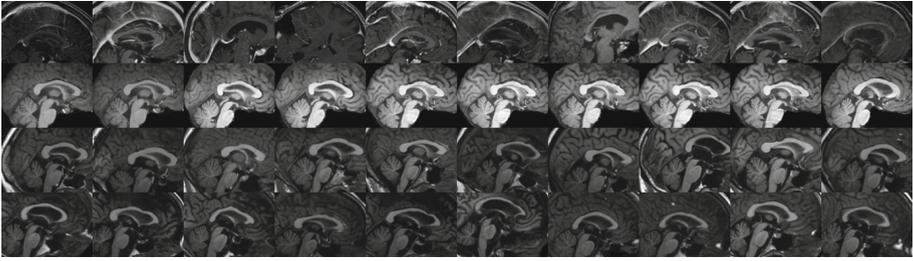


Fig. 1. Randomly selected images from our four test datasets (rows from top to bottom): MS#1 (first row), CIS, OAS-NC, and OAS-AD datasets.

Table 1. Characteristics of the datasets: size (mm), signal-to-noise-ratio (SNR_{mean} as given in [9]), original pixel spacing (mm) and image dimensions.

Dataset	Size of dataset	SNR	Mean CC area	Orig. spacing	Orig. dim
MS#1-4	280,270,280,300	2.06	731	$1 \times 1 \times 1$	$256 \times 256 \times 1$
CIS	140	2.54	661	$1 \times 1 \times 1$	$512 \times 512 \times 1$
OAS-AD	100	3.10	586	$0.5 \times 0.5 \times 1$	$512 \times 512 \times 1$
OAS-NC	316	2.85	573	$0.5 \times 0.5 \times 1$	$512 \times 512 \times 1$

in 2013. Each subject was scanned at up to 5 different timepoints: at screening, baseline (~ 4 weeks after screening), 3 months, 6 months, year one, and year two. This resulted in a total of $280 + 270 + 280 + 300 = 1130$ scans. Registration and automatic extraction of the mid-sagittal slice from each volume were performed. Then, the Yuki tool [6] was applied to segment the CC from each mid-sagittal slice. Finally, each segmentation was manually edited by a blinded operator using a graphical user interface.

OAS-NC (public): normal subjects collected from the OASIS dataset¹ were downloaded from NITRC². The CC segmentations were obtained using the semi-automatic CC extraction method of [6] with 20% of the scans subsequently edited by a blinded operator using the ITK-SNAP software, as noted in [6].

OAS-AD (public): subjects with Alzheimer’s disease also from the OASIS dataset were downloaded from NITRC. The CC segmentations were obtained using the same procedure as those used for the previous set.

Figure 1 shows a collection of the images from each dataset while Table 1 summarizes the characteristics of these datasets. All scans were registered to the template space and resampled to 256×256 with a pixel size of 1×1 mm prior to analysis. We chose the images acquired at the fourth timepoint of the MS dataset as the training set ($N = 300$). We denote this training set as MS#4.

¹ <http://www.oasis-brains.org>.

² https://www.nitrc.org/frs/?group_id=90.

Table 2. DSC obtained by different existing unsupervised methods and our novel supervised approach as averaged over all the available test data. “Ours” denotes using the proposed two-stage pipeline to learn and extract features only within the bounding box of the CC structure as localized from stage 1. “CEN” denotes applying the approach of [7] on the entire image slice.

Unsupervised methods					Supervised method	
CCTPA	WMV	STAPLE	STEPS	RW [18]	CEN only	Ours
0.734	0.776	0.713	0.790	0.798	0.929	0.947

Experiment I. Comparative analysis with unsupervised methods. We first performed comparative analysis using existing unsupervised methods; these include the intensity-based approach of Adamson et al. [1] (denoted as CCTPA), weighted majority voting (WMV), STAPLE [4], STEPS [8], and the random walker-based algorithm of [18]. Note that all multi-atlas based methods (except CCTPA) employed the same region-of-interest that we employed for our proposed 2-stage pipeline. For CCTPA, as its parameters depend on the input size, we did not use the same ROI. Table 2 reports the Dice similarity score (DSC) obtained by each method as averaged over the entire MS dataset, excluding the training set MS#4. Examining the numbers, our proposed supervised approach gave results superior to those derived from unsupervised methods.

We also examined the effect of training the CEN directly on the entire mid-sagittal slices, bypassing the CC localization step. From the same table, we can see that our two-stage ROI-specific CEN approach gave the highest DSC, thereby confirming the usefulness of the proposed two-stage pipeline.

Experiment II. Sensitivity analysis of the training sample size. Table 3 highlights the effect of training sample size (N); its top three rows report the segmentation accuracy obtained by our method for each dataset not used for training. Evidently, the more training samples used, the higher the accuracy can be obtained for all four datasets.

We again examined the effect of omitting the proposed two-stage pipeline, and trained the CEN directly on the entire image slices. Examining the bottom row of Table 3, we can observe again that our proposed two-stage approach also gave higher segmentation accuracy on all datasets except for MS#1. Hence, the proposed two-stage is not only more efficient (i.e. smaller image domain translates to fewer degrees of freedom being needed to model f), but is also capable of achieving better results.

Figure 2 shows example segmentation results that were obtained by thresholding the probabilistic predictions inferred from the trained CEN, i.e. with no other post-processing. Note that because we adopt a voxel-wise classification approach, and that the CENs that we employ do not explicitly enforce any form of shape and/or spatial regularization, some disconnected voxels can be mislabeled as CC, as highlighted in the figure in red. However, the produced segmentations can still be refined by fairly simple processing steps, such as connected

Table 3. Effect of using N training images randomly drawn from MS#4. Dice scores of the coloured cells indicate top performance achieved for each dataset. Asterisks denote significant improvement over second competing method ($p < 0.001$).

Method	N	MS Datasets			Other Datasets		
		MS#1	MS#2	MS#3	CIS	OAS-NC	OAS-AD
Ours	100	.890±.088	.899±.088	.877±.087	.824±.078	.833±.062	.780±.087
	200	.934±.063	.901±.055	.930±.075	.814±.092	.850±.057	.796±.082
	300	.910±.051	.924±.046*	.947±.044*	.860±.054*	.868±.058*	.868±.011
CEN	300	.906±.032	.902±.031	.932±.029	.720±.151	.791±.071	.858±.131

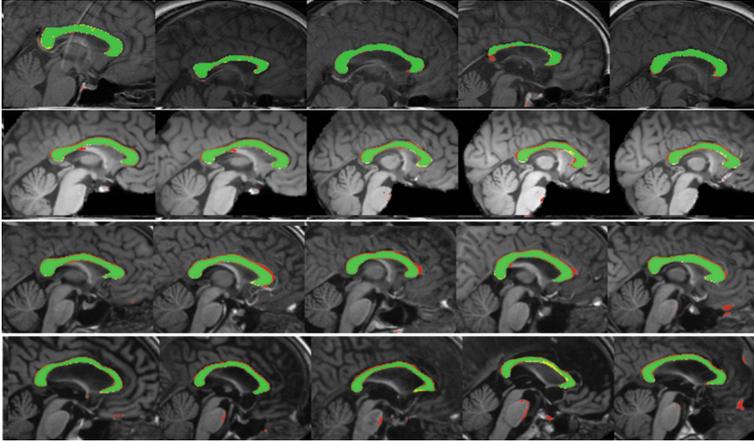


Fig. 2. Segmentation results randomly drawn from each of the four datasets (in order from top row to bottom: MS#1, CIS, OAS-NC, OAS-AD). The true positive, false positive, and false negative pixels are shown in green, red, and yellow, respectively.

component analysis, but was not done in this work as to avoid interference with the evaluation of the ROI-specific CEN approach.

4 Conclusions

We proposed a two-stage, fully automatic method for the segmentation of the CC structure in brain MRIs and presented thorough validation using four datasets collected from public and in-house sources. Our method was shown to be capable of performing results superior to those generated by existing unsupervised methods. Furthermore, our proposed two-stage pipeline that trains a ROI-specific CEN was shown to give accuracy higher than a non-ROI-specific CEN for most of the datasets tested. Future work involves implementing post-processing steps, e.g. using connected-component analysis to minimize the amount of false positive predictions, thereby further improving the segmentation accuracy of our proposed pipeline.

Acknowledgments. We thank the authors of [4, 7, 8, 18] for providing us with their software; NSERC, Engineers-in-Scrubs of UBC, and the Milan and Maureen Ilich Foundation for partially funding this work; and WestGrid for providing us with their computing resources.

References

1. Adamson, C., et al.: Software pipeline for midsagittal corpus callosum thickness profile processing. *Neuroinformatics* **12**(4), 595–614 (2014)
2. Avants, B., et al.: Advanced Normalization Tools, V1.0. In: *Insight Journal*, July 2009
3. Ronneberger, O., et al.: U-Net: convolutional networks for biomedical image segmentation. In: *MICCAI 2015*, p. 8 (2015)
4. Warfield, S.K., et al.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imag.* **23**(7), 903–921 (2004)
5. Zeiler, M.D., et al.: Adaptive deconvolutional networks for mid and high level feature learning. In: *2011 IEEE ICCV*, pp. 2018–2025. IEEE (2011)
6. Ardekani, B.A., et al.: Corpus callosum shape changes in early Alzheimer’s disease: an MRI study using the OASIS brain database. *Brain Struct. Funct.* **219**(1), 343–352 (2014)
7. Brosch, T., et al.: Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE Trans. Med. Imaging* **35**(5), 1229–1239 (2016)
8. Cardoso, M., et al.: STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcellation. *Med. Image Anal.* **17**(6), 671–684 (2013)
9. Dietrich, O., et al.: Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. *J. Magn. Reson. Imaging* **26**(2), 375–385 (2007)
10. Elahi, S., et al.: Corpus callosum atrophy rate in mild cognitive impairment and prodromal Alzheimer’s disease. *J. Alzheimer’s Disease JAD* **45**(3), 921–931 (2015)
11. Granberg, T., et al.: Corpus callosum atrophy is strongly associated with cognitive impairment in multiple sclerosis: results of a 17-year longitudinal study. *Multiple Sclerosis J.* **21**(9), 1152–1158 (2014)
12. Heckemann, R.A., et al.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage* **33**, 115–126 (2006)
13. Lee, H., et al.: Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: *AICML*, pp. 609–616. ACM (2009)
14. Marcus, S., et al.: Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.* **19**(9), 1498–1507 (2007)
15. Meyer, A.: Multi-atlas based segmentation of corpus callosum on MRIs of multiple sclerosis patients. In: Jiang, X., Hornegger, J., Koch, R. (eds.) *GCPR 2014*. LNCS, vol. 8753, pp. 729–735. Springer, Heidelberg (2014). doi:[10.1007/978-3-319-11752-2_61](https://doi.org/10.1007/978-3-319-11752-2_61)
16. Palmer, S.L., et al.: Decline in corpus callosum volume among pediatric patients with medulloblastoma: longitudinal MR imaging study. *Am. J. Neuroradiol.* **23**(7), 1088–1094 (2002)

17. Tajbakhsh, N., et al.: Convolutional neural networks for medical image analysis: fine tuning or full training? *IEEE Trans. Med. Imag.* **35**(99), 1299–1312 (2016)
18. Tang, L.Y.W., Hamarneh, G., Traboulsee, A., Li, D., Tam, R.: Corpus callosum segmentation in MS studies using normal atlases and optimal hybridization of extrinsic and intrinsic image cues. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015. LNCS*, vol. 9351, pp. 123–131. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-24574-4_15](https://doi.org/10.1007/978-3-319-24574-4_15)