Corpus callosum segmentation in MS studies using normal atlases and optimal hybridization of extrinsic and intrinsic image cues

Lisa Y. W. Tang, Ghassan Hamarneh*, Anthony Traboulsee, and Roger Tam

University of British Columbia, Vancouver, British Columbia, Canada
*Medical Image Analysis Lab., Simon Fraser University, Burnaby, Canada

Abstract. The corpus callosum (CC) is a key brain structure and change in its size and shape is a focal point in the study of neurodegenerative diseases like multiple sclerosis (MS). A number of automatic methods have been proposed for CC segmentation in magnetic resonance images (MRIs) that can be broadly classified as intensity-based and template-based. Imaging artifacts and signal changes due to pathology often cause errors in intensity-based methods. Template-based methods have been proposed to alleviate these problems. However, registration inaccuracies (local mismatch) can occur when the template image has large intensity and morphological differences from the scan to be segmented, such as when using publicly available normal templates for a diseased population. Accordingly, we propose a novel hybrid segmentation framework that performs optimal, spatially variant fusion of multi-atlas-based and intensity-based priors. Our novel *coupled* graph-labeling formulation effectively optimizes, on a per-voxel basis, the weights that govern the choice of priors so that intensity priors derived from the subject image are emphasized when spatial priors derived from the registered templates are deemed less trustworthy. This stands in contrast to existing hybrid methods that either ignore local registration errors or alternate between the optimization of fusion weights and segmentation results in an expectation-maximization fashion. We evaluated our method using a public dataset and two large in-house MS datasets and found that it gave more accurate results than those achieved by existing methods for CC segmentation.

1 Introduction

The corpus callosum (CC) is the largest white matter structure in the brain and plays the crucial role of relaying communication signals between the cerebral hemispheres. A growing body of recent literature [1–3] has shown that the change in its size as measured in structural MRIs[1] is a sensitive measure of regional brain atrophy that is effective for the monitoring of multiple sclerosis (MS) progression. However, in all of the aforementioned clinical studies [1–3], the CC structures were manually segmented by clinical experts. While various methods [4–6] have been proposed to segment the CC, these methods either perform segmentation in 2D [4–8], which is confounded by the selection of the appropriate plane of measurement [9] and hinders 3D shape analyses [9], or require human intervention ([4]), which makes large-scale analyses infeasible.

[1] We focus on structural MRIs as it is a more common modality than the others.

Existing automatic methods for 2D segmentation of the CC in structural MRIs may generally be classified [4] into two categories. In *intensity-based* methods [4–6, 8], segmentation is mostly driven by intrinsic data from the subject image. Both [5, 6] require a midsagittal plane extraction tool. Other methods [4, 8] require tissue segmentation, which makes them susceptible to partial volume effects and errors in segmentation. For example, it was noted in [4] that the method should not be used for MR images of patients with demyelinating diseases (e.g. MS) because significant alterations in image intensities tend to occur in these images. In *template-based* methods (e.g. [4, 10] and references therein), labels of pre-segmented template images are propagated and fused, in the case of multiple templates, to obtain the final segmentation. This approach, more generally known as multi-atlas segmentation (MAS) [11], is more robust than intensity-based methods due to the spatial constraints implicitly imposed by registration. However, segmentation accuracy not only depends on registration, but also on the choice of the template images ([4]). When there exists large intensity and morphological variability between the template(s) and the subject image, registration accuracy suffers, which in turn limits segmentation accuracy. Hence, one should use templates drawn from the population of the target image, which is often not publicly available (e.g. in case of pathology such as MS) nor easy to create without introducing biases (i.e. one with templates that span the whole population without favoring any subset).

In addition to the above problems, two issues particular to MS datasets bring further challenges: 1) retrospective studies can span over a decade, and older data tends to have low resolution, poor image contrast, and low signal-to-noise ratio (SNR); and 2) the possible use of gadolinium contrast in T1-weighted scans results in hyper-intensities in the CC periphery. These factors render the publicly available algorithms [4, 6] largely inapplicable in our two MS datasets.

To overcome the aforementioned problems, a hybrid approach may be adopted where MAS results are incorporated into an intensity-based segmentation framework in the form of spatial priors and the segmentation problem is then solved via energy minimization, as done in [12]. This approach was applied to the cerebellum and other structures but, to the best of our knowledge, has never been applied to the CC. More importantly, most hybrid methods, e.g. [12], employ spatial priors without accounting for *local* mismatches between the registered images. Hence, registration errors can be large in some regions, yielding spatial prior information that is misleading at these regions.

Accordingly, we propose a novel graph-based formulation that explicitly accounts for registration errors and spatially adapts fusion of priors accordingly by coupling the tasks of optimal locally adaptive label fusion and segmentation. This stands in contrast with existing MAS methods [11–13] in that we optimize the fusion of all data priors based on available information (as opposed to criteria derived from additional learning with training data [13] or "labeling unanimity" [11]) so that we can handle local mismatches that may occur in *all* templates due to inherent incompatibility between the diseased and normal images. In summary, unlike existing CC segmentation algorithms, our method 1) performs the segmentation in 3D; 2) requires no tissue segmentation nor training data from diseased-subjects; 3) is more robust to registration errors than existing methods, as we show in our results involving two large MS datasets; and 4) is robust even when a small template set is used, unlike existing methods.

2 Methods

Preliminaries. Let there be a set of N template images with corresponding expert segmentations of the CC. For a given subject image $I : \Omega \subset \mathbb{R}^d \mapsto \mathbb{R}$ which we aim to find the segmentation label field $\mathbf{S}$ of the CC, we register each template image to I and apply the resolved spatial transformation to its corresponding label field, resulting in pairs of template images and corresponding labels, denoted as $\mathcal{T}=\{T\}_{n=1}^{N}$ and $\mathcal{S}=\{S\}_{n=1}^{N}$ respectively, all approximately[2] aligned to I via spatial transforms $\tau=\{\tau\}_{n=1}^{N}$. These can then be used to generate a rough fuzzy segmentation $\hat{S}$ for I in the form of majority voting (MV) [14]; i.e. $\hat{\mathbf{S}}(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N} S_n(\mathbf{x})$, followed by thresholding.

In weighted majority voting (WMV), labels of the aligned templates are fused in a locally adaptive manner, where the normalized weight of template T_n at voxel $\mathbf{x}$ may be determined based on a patch-based image similarity measure Θ, i.e.

$$w(\mathbf{x}, n) = \frac{1}{\eta}\Theta(\mathbf{x}; I, T_n), \tag{1}$$

where $\eta=\sum_{n=1}^{N}\Theta(\mathbf{x}; I, T_n)$. Spatial regularization on w (e.g. via smoothing as done in [15]) may further be imposed to encourage smoothness of $\hat{\mathbf{S}}$, leading to a more general approach that we denote here as optimized label fusion (OLF). A fuzzy label assignment is then computed as $\Phi(\mathbf{x})=\sum_{n=1}^{N} w(\mathbf{x}, n)S_n(\mathbf{x})$, which is then binarized so that the final segmentation is computed as $\hat{\mathbf{S}}(\mathbf{x})=1$, if $\Phi(\mathbf{x})>0.5$, or $\hat{\mathbf{S}}(\mathbf{x})=0$ otherwise.

Rather than calculating w as a post-processing step that is independent of calculating $\mathbf{S}$, or as part of an expectation-maximization framework [11] that alternates between calculations of the two, we herein propose to optimize w *jointly* with $\mathbf{S}$ so that the segmentation process explicitly accounts for registration errors and performs spatially variant prior-selection accordingly. In doing so, we formulate CC segmentation as a graph-labeling task with a *coupled* label set $\mathcal{L}$, consisting of a segmentation label field for I and labels indexing elements of $\mathcal{T}$, i.e. $\mathcal{L}=\{0, 1\} \bigcup \mathcal{L}_{corr}$, where $\mathcal{L}_{corr}=\{1, 2, \ldots, N\}$ defines a set of labels indexing the n-th template that best corresponds to I at $\mathbf{x}$. As we explain in the next section, by employing the random walker (RW) formulation [16] that generates fuzzy label assignments, we perform segmentation and optimal label fusion simultaneously, with a byproduct of an optimized spatially variant fusion of priors.

Joint optimal label fusion and segmentation via RW. Coupling the two problems allows us to take the new perspective of viewing all available images simply as an augmented set of information *sources* $\mathcal{I}:=\{T_1, \cdots T_N, I\}= \{\mathcal{I}\}_{k=1}^{K}$, where K=N+1, such that *intrinsic* information from I and *extrinsic* information from T together guide the labeling of I. Furthermore, each source shall be accompanied by a function f_k that outputs an estimate of the class-likelihood given the observations made on that source. In this paper, for the intrinsic source, its corresponding function, f_K, is a naive Bayes classifier [17] while those of the extrinsic sources f_k are given by S_k ($k < K$). With this perspective, our task thus becomes optimizing the fusion of the outputs of f_k that have high confidence in casting label predictions correctly so that we can tune weights on the priors accordingly.

[2] We will examine the impact of registration accuracy on segmentation results in Sec. 3.

In this work, we propose to quantify the confidence of each source f_k in a spatially adaptive manner via the use of a function $\mathbf{C} : \Omega \mapsto \Delta^K$, where Δ^K refers to a K-dimensional unit simplex. In absence of training data that would allow us to estimate the confidence of each f_k based on some performance criteria (e.g. [13]), we infer confidence by quantifying its strength of belief on the CC class label. As similarly done in [17], for the intrinsic source, we infer confidence of its predictive function f_K by estimating the data-likelihood at $\mathbf{x}$ as:

$$\mathbf{C}_K(\mathbf{x}) = Pr(\mathbf{S}(\mathbf{x}) = 1|I(\mathbf{x})) = \frac{1}{a}\exp\left(\frac{-||I(\mathbf{x}) - \mu||^2}{\sigma^2}\right), \tag{2}$$

where μ and σ respectively represent the mean and standard deviation of the intensity values of the CC estimated using $\hat{\mathbf{S}}$, and a rescales the data-likelihood to [0,1]. For $k < K$, the confidence of f_k that uses extrinsic source T_k depends on how well T_k and I are registered at $\mathbf{x}$ and thus is estimated from information available from the registration results (e.g. quality of alignment as estimated by the regularity of τ_n and/or Θ measured between the registered image pair). Based on preliminary experiments, we found that examining only Θ was sufficient (i.e. required least computation without compromising accuracy). Hence, we estimated the confidence of each f_k ($k < K$) as:

$$\mathbf{C}_k(\mathbf{x}) = \frac{1}{b}\Theta(I(\mathbf{x}), T_k(\mathbf{x})), \tag{3}$$

where $b{=}\sum_{k=1}^{N}\Theta(I(\mathbf{x}), T_k(\mathbf{x}))$ is a normalization constant and Θ is based on the Modality-Independent Neighbourhood Descriptor (MIND) [18] (with its default parameters) to ensure robustness to variations in image contrast.

Without other *a priori* knowledge, $\mathbf{C}$ may be used directly for fuzzy label assignment (as in WMV noted above), i.e.:

$$\Phi(\mathbf{x}) = \alpha\mathbf{C}_K(\mathbf{x})\mathbf{S}_{init}(\mathbf{x}) + (1-\alpha)\sum_{k=1}^{N}\mathbf{C}_k(\mathbf{x})S_k(\mathbf{x}) \tag{4}$$

where α is a constant governing global preference for intensity-based priors and $\mathbf{S}_{init}(\mathbf{x})$=1 if $\hat{\mathbf{S}}(\mathbf{x}) > 0.5$, or
$\mathbf{S}_{init}(\mathbf{x})$=0 otherwise. However, as shown in the literature [11], the desired spatial smoothness of the segmentation is directly influenced by smoothness of the weights. We thus search for a function $\mathbf{W}$ surrogate of $\mathbf{C}$ that is spatially smooth but remains similar to $\mathbf{C}$, thus leading to this energy minimization:

$$\mathbf{W}^* = \arg\min_{\mathbf{W}}\int_{\Omega}||\mathbf{W}(\mathbf{x}) - \mathbf{C}(\mathbf{x})||d\mathbf{x} + \beta\sum_{k=1}^{K}\int_{\Omega}(\nabla\mathbf{W}_k(\mathbf{x}))^T\nabla\mathbf{W}_k(\mathbf{x})d\mathbf{x}, \tag{5}$$

where β is a constant governing the strength of diffusion-based regularization [16]; $\mathbf{W}_k$ and $\mathbf{C}_k$ denote the k-th components of $\mathbf{W}$ and $\mathbf{C}$, respectively.

As presented in [16], we can obtain a unique, globally optimal solution to (5) by adopting a graph-based approach. Specifically, if we let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be a graph representing the subject image I, where each graph node $p \in \mathcal{V}$ represents a spatial coordinate

$\mathbf{x}_p \in \Omega$, and let L be the Laplacian matrix encoding $\mathcal{E}$, then (5) is equivalent [16] to:

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} \sum_{j=1, j\neq k}^{K} \mathbf{W}_k^T \Lambda_j \mathbf{W}_k + (1-\mathbf{W}_k)^T \Lambda_k (1-\mathbf{W}_k) + \beta (\mathbf{W}_k)^T L \mathbf{W}_k, \quad (6)$$

where Λ_k is a diagonal matrix with entries $[\mathbf{C}_{1k}, \cdots, \mathbf{C}_{Vk}]$ where $\mathbf{C}_{pk}$ denotes the normalized[3] confidence value of source k as estimated at $\mathbf{x}_p$. The minimum of the energy (6) is obtained when $\mathbf{W}_k$ is the solution to a combinatorial Laplace equation, which can be solved in closed-form as shown in [16]. Once (6) is solved, we compute $\mathbf{S}$ by using (4), with $\mathbf{C}$ now replaced by $\mathbf{W}^*$.

3 Evaluation Results

Materials. Our template set ($\mathcal{T}$) is derived from the publicly available MRI dataset of normal subjects from [19]; this set is hereafter denoted as HAMMERS. Various registration settings were explored for finding the registration solutions (τ); these include affine and deformable registration using free-form-deformation (FFD) and SyN [20], all of which used cross-correlation (CrCo) as the image similarity measure, as well as the method of [21] that computed the image similarity measure based on MIND [18]. Except when varying N to determine its effect, we set N=6 based on experiments in [10], which showed that N=6 using STAPLE yielded high accuracy. Note that minimizing N is advantageous by requiring fewer registrations and thus reducing computations. As our experiments below show, our method is relatively insensitive to the choice of N when compared to other MAS-based methods.

To assess the performance of our method on images acquired from subjects with MS, we further collected T1-weighted brain MRIs originally acquired using different imaging protocols for two independent MS clinical trials. We denote these as MS Dataset1 and MS Dataset2 (sizes of 85 and 187, respectively). Automatic CC segmentation of these MS images is much more challenging due to the reasons highlighted in Sec. 1. In addition, these images were acquired from multiple imaging centers, leading to large intensity and anatomical variations within and across these image sets. The latter is the most challenging of all datasets examined in this paper due to the additional use of contrast agent during image acquisition. In both MS datasets, only 2D segmentations are available. These were manually prepared by a clinical expert using the ITK-SNAP® software, subsequent to 2D midsagittal plane extraction (see [2] for further details).

Experiment I: Effect of combined intensity and spatial priors on accuracy. As conventional baselines, we compared our proposed method with MV and WMV, which do not employ intrinsic information from I, i.e. intensity-based priors. To specifically determine the effect of spatially adaptive, optimized fusion of priors that our proposed method uses, we also tested random walker segmentation [16] using constant weight (α) between the intensity and spatial priors constructed from $\mathcal{T}$ which we denote as RW. For reference, we also examined the case of constructing spatial priors optimized

[3] $\mathbf{C}$ is normalized to rows with unity sum.

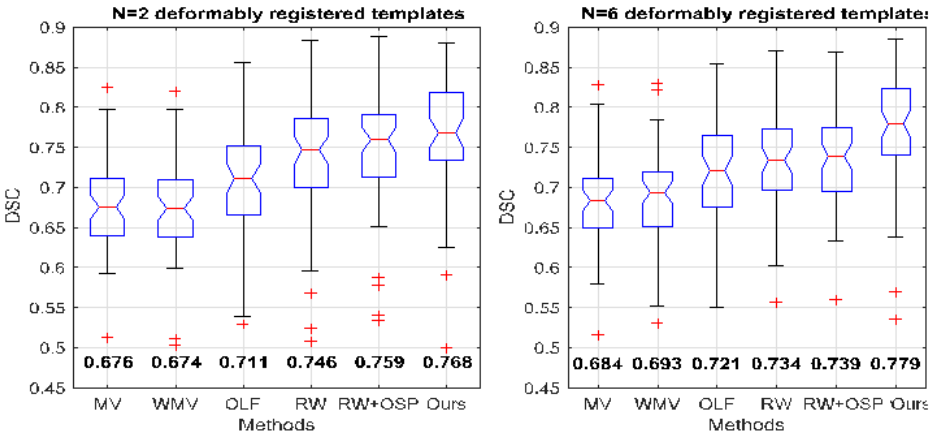

Fig. 1: Accuracy as measured with DSC of various methods compared (bolded numbers indicate the medians of DSC). Note that even for the case when N=2, our method achieved results better than (weighted) majority voting (MV/WMV), optimal label fusion (OLF), and hybrid random walker (RW) with/without optimized spatial priors (OSP).

(a)	HAMMERS					(b)	MS Dataset1			MS Dataset2			
Spatial priors	MV	WMV	STAPLE [10]	RW+OSP	Ours	MV	WMV	STAPLE [10]	Ours	MV	WMV	STAPLE [10]	Ours
Affine+CrCo [20]	.586	.699	.499	.824	**.825**	.708	.719	.710	**.812**	.617	.630	.617	**.658**
Affine/FFD+CrCo [20]	.662	.712	.513	.832	**.837**	.871	.878	.824	**.901**	.689	.690	.614	**.713**
Affine/FFD [21]+MIND	.831	.831	.822	.842	**.865**	.838	.838	.796	**.860**	.729	.729	.702	**.763**
Affine/SyN+CrCo [20]	.740	.733	.556	.833	**.852**	.898	.898	.863	**.926**	.718	.724	.666	**.751**

Table 1: Evaluation on (a) HAMMERS and (b) two MS datasets using N=6 templates aligned by different registration algorithms. MS Dataset2 contains images with contrast agent and thus presents more challenging cases.

for I where results from OLF were used in RW. We denote this approach as RW with optimized spatial priors (RW+OSP).

Fig. 1 compares the results of segmentation accuracy of each method on a subset of the MS Dataset2 (60 randomly selected images) when evaluated in a leave-one-out cross validation experiment, where we optimized the hyper-parameters for each method by performing a grid-search over $\alpha = \{.1, .3, .4, .5, .6, .7, .9\}$ and $\beta = \{0.01, 0.1, 1\}$. Firstly, from the figure, where accuracy was measured with the Dice coefficient (DSC), we see that optimal label fusion (OLF) achieved better results than both MV and WMV, which supports the use of spatially regularized fusion weights as suggested in the literature [11]. Secondly, we see the advantage of jointly examining both spatial and intensity-based priors in RW, which performed better than MV, WMV, and OLF, which disregard intrinsic information. However, due to registration errors, the constructed spatial priors are not trustworthy everywhere in the image. With our proposed spatially variant prior-selection, our method was able to outperform RW with non-adaptive priors. Note also that because RW+OSP does not perform prior-selection in a spatially adaptive manner, it did not produce an improvement similar to ours.

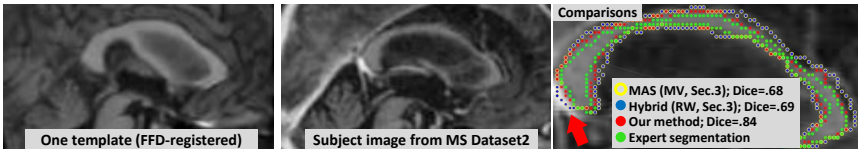

Fig. 2: An example segmentation for an image from the MS Dataset2. As noted by the red arrow in the zoomed-up view (right subfigure), due to local misregistration between the subject and template images (as contrast enhanced vessels in the subject are not visible in any of the templates used), non-adaptive methods cannot segment the noted region accurately. Our hybrid method with optimized spatially adaptive fusion of priors placed greater weights on intensity-based priors in this region, and thus could segment this region properly.

Experiment II: 3D validation using the HAMMERS dataset [19]. We next validated our method using 14 images in the HAMMERS dataset (exclusive of the aforementioned $N = 6$ randomly chosen template images). Table 1a reports the performance (median of DSC) of our proposed method with respect to MV, WMV, RW+OSP, and STAPLE [10]. We also examined the effects of registration accuracy (and thus the quality of the spatial priors) on segmentation accuracy. As shown in the table, our method achieved the best accuracy regardless of how the templates were registered.

Experiment III: Accuracy evaluated on two MS-specific datasets. We also compared our method with several published methods examined in [10] (MW/WMV/STAPLE), using all of the MS images not used in Experiment I. Table 1b summarizes the results. Dataset2 consists of MRIs with contrast agent, so that segmentation of these images was especially challenging (mismatch between the templates and I around the CC periphery occurred more often. As illustrated in Fig. 2, our proposed use of spatially adaptive optimized fusion of priors enables our method to handle these problem regions better than the other methods. Overall, our method achieved the best accuracy in all cases tested. Lastly, performing a comparison to the state-of-art algorithm STEPS [22] using MS Dataset2 showed that our proposed method remains highly competitive: using N=20 atlases randomly selected from the HAMMERS dataset, our method achieved a mean Dice of 0.798, while that of STEPS [22] is 0.775. When N=15, our method still maintained a mean Dice of 0.798, while that of STEPS [22] dropped to 0.751.

4 Conclusions

We introduced a novel framework that performs automatic volumetric CC segmentation in MRIs without the use of disease-specific brain templates. Experiments on 3 datasets show that our method can perform robustly, even in presence of large intensity and morphological variability, thanks to our coupled framework that explicitly accounts for local registration errors and weighs down untrusted spatial priors in a spatially varying manner. For $N \leq 6$, our method took $\approx$5 minutes to solve (6) (as executed on a dual-core 2.4 GHz CPU) in addition to the time needed to register the templates. Future work will involve evaluating the proposed method on other brain structures, including thalamus and putamen, which also have relevance to MS as shown in recent clinical studies.

Acknowledgements. We wish to thank NSERC, Engineers-in-Scrubs of UBC, and the Milan and Maureen Ilich Foundation for partially funding this work.

References

1. Wang et al., "3D vs. 2D surface shape analysis of the corpus callosum in premature neonates," in *MICCAI: Workshop on Paediatric and Perinatal Imaging*, 2012.
2. Kang et al., "Corpus callosum atrophy in a large SPMS cohort and its correlation with PASAT as a cognitive marker," in *ECTRIMS 2013*, 2013.
3. Mitchell et al., "Reliable callosal measurement: population normative data confirm sex-related differences," *American Journal of Neuroradiology*, vol. 24, no. 3, pp. 410–418, 2003.
4. Herron et al., "Automated measurement of the human corpus callosum using MRI," *Frontiers in neuroinformatics*, vol. 6, 2012.
5. Ardekani et al., "Multi-atlas corpus callosum segmentation with adaptive atlas selection," *Proc Int Soc Magn Reson Med Melbourne, Australia, Abstract*, vol. 2564, 2012.
6. Adamson et al., "Software pipeline for midsagittal corpus callosum thickness profile processing," *Neuroinformatics*, pp. 595–614, 2014.
7. McIntosh and Hamarneh, "Medial-based deformable models in nonconvex shape-spaces for medical image segmentation," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 1, pp. 33–50, Jan 2012.
8. Vachet et al., "Automatic corpus callosum segmentation using a deformable active Fourier contour model," in *SPIE Medical Imaging*, 2012, pp. 831707–831707.
9. Changizi et al., "Extraction of the Plane of Minimal Cross-Sectional Area of the Corpus Callosum using Template-Driven Segmentation," in *MICCAI*, 2010, vol. 6363, pp. 17–24.
10. Meyer, "Multi-atlas Based Segmentation of Corpus Callosum on MRIs of Multiple Sclerosis Patients," in *Pattern Recognition: 36th German Conference*, pp. 729–735. 2014.
11. Wu et al., "A generative probability model of joint label fusion for multi-atlas based brain segmentation," *Medical Image Analysis*, pp. 881–890, 2014.
12. Ltjnen et al., "Fast and robust multi-atlas segmentation of brain magnetic resonance images," *NeuroImage*, pp. 2352 – 2365, 2010.
13. Sanroma et al., "Learning-based atlas selection for multiple-atlas segmentation," in *IEEE CVPR*, 2014, pp. 3111–3117.
14. Aljabar et al., "Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy," *NeuroImage*, vol. 46, pp. 726–738, 2009.
15. Wang et al., "Multi-atlas segmentation with joint label fusion," *TPAMI*, pp. 611–623, 2013.
16. L Grady, "Multilabel random walker image segmentation using prior models," in *CVPR*. IEEE, 2005, vol. 1, pp. 763–770.
17. Zhang et al., "A novel image matting approach based on naive Bayes classifier," in *Intelligent Computing Technology*, pp. 433–441. 2012.
18. Heinrich et al., "MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration," *Medical Image Analysis*, vol. 16, pp. 1423–1435, 2012.
19. Hammers et al., "Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe," *Human Brain Mapping*, vol. 19, pp. 224–247, 2003.
20. Avants et al., "A reproducible evaluation of ANTs similarity metric performance in brain image registration," *NeuroImage*, vol. 52, pp. 2033–2044, 2011.
21. Heinrich et al., "Globally optimal deformable registration on a minimum spanning tree using dense displacement sampling," in *MICCAI*, 2012, pp. 115–122.
22. Cardoso et al., "STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation," *Medical Image Analysis*, vol. 17, no. 6, pp. 671–684, 2013.