SUPERVISED LEARNING FOR BRAIN MR SEGMENTATION VIA FUSION OF PARTIALLY LABELED MULTIPLE ATLASES

Yan Deng, Anand Rangarajan, and Baba C. Vemuri

Department of CISE, University of Florida, Gainesville, FL 32611, USA

ABSTRACT

Fully labeled manual segmentation—a cornerstone of neuroanatomical structure segmentation, is known to be a tedious, time-consuming and error-prone task even for trained experts. In this paper, we propose a novel partially labeled multiple atlas-based segmentation algorithm which can simultaneously segment multiple structures from a given image. Intra- and Inter- structural constraints are imposed to preserve spatial relationships and to propagate the segmentation from the labeled regions to the unlabeled regions. We present several experiments on real data sets which show that our approach yields accurate segmentations of the test data even in the absence of a large percentage of the atlas labels. Further, our approach has the ability to refine the given partially labeled atlases via a supervised learning stage.

Index Terms— Partially Labeled Multiple Atlases, MRI Segmentation, Structural Constraints

1. INTRODUCTION

Neuroimage analysis and its associated application to the diagnosis and treatment of brain disorders has attracted immense attention in the past two decades. Automated segmentation of neuroanatomical structures from 3D MR scans is one of the key steps in Neuroimage analysis tasks. In the context of brain MRI segmentation, there is ample literature on segmenting structures such as the hippocampus, amygdala, caudate, nucleus, ventricles, etc. Multi-atlas based segmentation is the most popular approach used to achieve the aforementioned tasks. Global, semi-global or local weights are learned to propagate the voxel-wise labels from the warped atlas to the subject, as described in [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11].

Most of the multi-atlas based segmentation algorithms require fully (manually) labeled atlases as input. However, the manual segmentation of neuroanatomical structures is known to be an error-prone, tedious, and time-consuming procedure. A partially labeled atlas based segmentation algorithm could greatly help to reduce this excessive manual workload. There are a few partially labeled multi-atlas based segmentation algorithms. In [12], authors use partially labeled atlases as a prior in an expectation maximization (EM) based estimator. More recently, in [13], Koch *et al.* proposed an algorithm called "GLa", where they pose the multi-atlas based segmentation problem in a graph-cuts framework, where unlabeled nodes are disconnected to the rest of graph. In this paper, we propose a novel partially labeled multi-atlas based segmentation (PL-MAS) algorithm which is different from *all* of the aforementioned methods. In the following, we present a brief summary. We develop a supervised learning framework to learn the optimal voxel-wise weighted combination of warped, partially labeled, atlas segmenters. In the training stage, the optimal weights are learned through the minimization of the error between the weighted combination of warped atlas segmentations and the ground-truth segmentation within the labeled region. Further, we impose intra- and inter- structural constraints to regularize the weights within the labeled region and infer the weights of the unlabeled region. The resulting optimal weighted combination of the warped atlas segmentations for each training data set automatically leads to a refinement of the given partially labeled ground-truth segmentation. In the testing phase, we use the warped optimal weights associated with the most relevant data from the training set to achieve the segmentation of a test image.

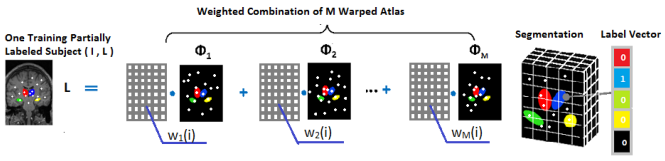

Fig. 1. Framework for voxel-wise combination of M warped atlas segmentations and label vectors

The rest of the paper is organized as follows: In section 2, we present details of the segmentation methodology of our partially labeled multi-atlas based segmentation algorithm. The algorithms are validated experimentally in section 3 and conclusions are presented in section 4.

2. METHODOLOGY

Given the partially labeled atlas L_{atlas} and its corresponding intensity image I_{atlas}, a general nonrigid-deformation based segmentation method can be used to create a weak segmenter $\Phi = L_{\text{atlas}} \circ f_{I_{\text{atlas}} \to I}$ of a given intensity image, which we will henceforth refer to as a "warped atlas segmenter" or a "weak segmentation". Here, f is the nonrigid deformation

This research is in part supported by the grant NIH NS066340 to BC Vemuri.

(symmetric diffeomorphic transformation) from the intensity image of the atlas to the image to be segmented.

The problem can be described as follows. Given M warped atlas (partially labeled) based segmentations $\{L_{\text{atlas}^t} \circ f_{I_{\text{atlas}^t} \to I}, t = 1, \ldots, M\}$ for a test image I, we seek to form a final strong segmentation from these weak segmentations. As shown in Figure 1, we use a voxel-wise weighted combination of the weak segmentations to achieve the goal. Here, we assume that all the data have already been affinely registered to a reference image so that they are approximately aligned to the same coordinate system. Prior to delving into the details of the training and testing stages, we first introduce the label representation used in this paper.

It is of great clinical importance to develop an algorithm for the simultaneous segmentation of multiple structures within a shape complex—a collection of neighboring neuro-anatomical structures—since many neurological disorders are diagnosed by structural abnormalities that may be present in several brain structures rather than just a single one. Therefore, our label field representation needs to be able to describe multi-class segmentation. Furthermore, since the given atlases are only partially segmented/labeled, our label field representation is also required to depict unlabeled regions. Assume that there are K structures within a shape complex. We use a $K+1$ dimensional vector to store the label information where the $(K+1)^{th}$ structure refers to the background. A standard 1-of-K encoding approach is used to represent the label field. More specifically, a $(K+1)$-dimensional label vector is associated with each voxel and if the voxel is labeled as belonging to the k^{th} structure, the k^{th} entry of the vector will be set to one with the rest of the elements set to zero. If a voxel is unlabeled, the associated label vector is set to the zero vector. As shown in Figure 1, the voxel is labeled as belonging to the 2^{nd} (blue) structure, so the vector associated with that voxel is $[0, 1, 0, 0, 0]^t$. This label representation facilitates in formalizing the supervised learning objective function as a least squares problem, which we will show subsequently. To simplify the presentation, we make use of some notation which is introduced now:

- $\mathcal{C} = \{1, \ldots, K+1\}$ denotes the label class set.
- The subscript $_t$ denotes the index to atlases.
- $\{I^1, \ldots, I^N\}$ denotes the training subjects. Superscript ^{n} denotes the index to the training subjects.
- $\Phi_t = \{L_t \circ f_{I_t \to I}\}$ represents the t^{th} warped atlas segmentation of the image to be segmented. $\boldsymbol{\Phi}_t(i)$ is the label vector of the i^{th} voxel. Notice that the i^{th} voxel is unlabeled if $\boldsymbol{\Phi}_t(i)$ is a zero vector.
- $\boldsymbol{\Theta}(i) = [\boldsymbol{\Phi}_1(i), \boldsymbol{\Phi}_2(i), \ldots, \boldsymbol{\Phi}_M(i)]_{(K+1)\times M}$ represents the M atlas warped segmentation results to be combined for the i^{th} voxel. $\boldsymbol{\Theta}(i,k), k \in \mathcal{C}$ is the M warped atlas segmentation results for the k^{th} class.
- $\boldsymbol{w}(i) = [w_1(i), \ldots, w_M(i)]$ is the weight vector used to combine the warped atlas segmentation for the i^{th} voxel.
- $\boldsymbol{S}(i) = \boldsymbol{\Theta}(i)\boldsymbol{w}(i)$ is the weighted combination of multi-atlas warped segmentation for the i^{th} voxel. $\boldsymbol{S}(i,k)$ is the k^{th} entry of the label vector.
- Ω_{labeled} represents a labeled region, $\Omega_{\text{unlabeled}}$ represents an unlabeled region, $\Omega_{\text{labeled}} \cup \Omega_{\text{unlabeled}} = \Omega$.

2.1. Supervised Learning Stage

In the training stage, for each given partially labeled training subject $\{L, I\}$, we want to learn the optimal linear weighted combination of warped atlas segmentations $\{L_t \circ f_{I_t \to I}, t = 1, \ldots, M\}$ which yields the best approximation to its ground-truth segmentation. With the above given notation, we can represent the voxel-wise linear combination of the warped atlas segmentations by $\boldsymbol{\Theta}(i)\boldsymbol{w}(i), i \in \Omega$. For the labeled region, since the ground-truth segmentation is known, it is natural to expect the error $\sum_{i \in \Omega_{\text{labeled}}} \|\boldsymbol{\Theta}(i)\boldsymbol{w}(i) - L(i)\|^2$ to be minimized. However, the optimal weights resulting from this simple regression model are limited to the labeled regions. Besides, the assumption of independent voxel weights leads to loss of spatial relationships within and across the structures. Therefore, we introduce *Intra- and Inter- Spatial Constraints* to preserve the within/across structure/s spatial relationships and propagate the weights from the labeled regions to the adjacent unlabeled regions.

2.1.1. Intra- Spatial Constraint

Strong spatial dependencies exist within regions in most real images. It is therefore well-justified to assume that the weights are smooth within each structure. A **Local Neighborhood Graph** $\mathcal{G}$ is constructed from each partially labeled ground truth training dataset to capture the spatial dependency information within each structure. For each voxel i, $\mathcal{G}(i,j), j \in \mathcal{O}(i)$ is defined as follows:

$$\mathcal{G}(i,j) \stackrel{\text{def}}{=} \begin{cases} \boldsymbol{L}(i) \cdot \boldsymbol{L}(j), & if \quad i,j \in \Omega_{\text{labeled}} \\ H(\delta - |\boldsymbol{v}_{j \to i} \cdot \boldsymbol{n}(i)|), & else. \end{cases} \tag{1}$$

where $\boldsymbol{v}_{j \to i}$ is the normalized vector from voxel j to voxel i, $\boldsymbol{n}(i)$ is the unit surface normal at the i^{th} voxel and H is the Heaviside (step) function. δ is a threshold which is set to be 0.5. In 3D, the neighborhood of the i^{th} voxel $\mathcal{O}(i)$ used is $6, 18$ or 26 reflecting standard nearest neighbor relationships. This neighborhood graph is used to regularize the distances between weight vectors of adjacent voxels. When the i^{th} and the j^{th} voxels are both labeled, the graph entry $\mathcal{G}(i,j) = 1$, if the two voxel are located in the same structure, otherwise $\mathcal{G}(i,j) = 0$. When either of the two voxels' labels is missing, we can infer the spatial dependency from the geometric properties. If the i^{th} voxel is a point on the surface (constructed by the partially labeled ground-truth segmentation), $\mathcal{G}(i,j) = 1$, when the angle between $\boldsymbol{v}_{j \to i}$ and the normal is larger than $across(\delta)$, else $\mathcal{G}(i,j) = 0$. If the i^{th} voxel is a point within the structure where $n(i) = 0$, $\mathcal{G}(i,j) = 1$, since all of its neighbors are within the structure. It is natural to expect the weights corresponding to the neighboring voxels within the same structure or residing along the tangent direction to be the similar. To this end, we use the regularization term $\mathcal{G}(i,j)\|w_i - w_j\|^2$ to preserve intra-structural smoothness.

2.1.2. Inter- Spatial Constraint

In addition to intra-structural smoothness, we also propose a novel spatial constraint to avoid the overlapping of neighboring structures in the final segmentation. Given a shape complex with K structures in it, we first construct a **Global Neighborhood Graph** $\mathcal{N}$ containing spatial information about the neighboring structures. Here, $\mathcal{N}$ is a $(K+1) \times (K+1)$ matrix.

$$\mathcal{N}(k,l) \stackrel{\text{def}}{=} \begin{cases} 1, & if\ k \neq l, k, l \in \{\mathcal{C} \backslash (K+1)\} \\ 0, & else. \end{cases} \quad (2)$$

We expect that each structure does not intersect any of the other structures with the exception of the background which can intersect any other structure. More specifically, for each voxel, we expect the linear combination $S(i)$ to satisfy the condition $\mathcal{N}(k,l)S(i,k)S(i,l) = 0$ to prevent structure overlap. *While there are many different ways in which one can enforce competition between different structures attempting to claim each voxel as their own, we adopt an approach with scalability in mind. That is, we would prefer a sparse linear solver for classification due to its attractive scalability properties.* Scalability is important in this context due to the presence of multiple weight vectors—with variation over both the voxels and the training set. It is due to this variability that we seek a sparse linear system solution.

2.1.3. Objective Function

In order to obtain a classifier with a sparse linear solver, we construct an objective function which is quadratic in the weight vectors with the spatial constraints enforced. Based on these considerations, the overall optimization problem is given by,

$$\begin{aligned} &w^\star = \arg\min_w \textstyle\sum_{i \in \Omega_{\text{labeled}}} \|\boldsymbol{\Theta}(i)\boldsymbol{w}(i) - L(i)\|^2 + \\ &\lambda_1 \textstyle\sum_{i \in \Omega} \sum_{j \in \mathcal{O}(i)} \mathcal{G}(i,j) \|\boldsymbol{w}(i) - \boldsymbol{w}(j)\|^2 + \\ &\lambda_2 \textstyle\sum_{k,l \in \mathcal{C}} \sum_{i \in \Omega} \mathcal{N}(k,l)(\boldsymbol{w}(i) \cdot \boldsymbol{\Theta}(i,k))(\boldsymbol{w}(i) \cdot \boldsymbol{\Theta}(i,l)). \end{aligned} \quad (3)$$

Where, $w = \{\boldsymbol{w}(i)\}$, λ_1 and λ_2 are regularization parameters. The optimization problem can be posed as the solution to a sparse linear system that results in a closed form solution after some algebra which we omit here due to lack of space. As shown in the objective function, weights of the unlabeled regions will be propagated from the connecting labeled regions. The non-overlap constraint forces the weighted combination (segmentation estimation) of unlabeled regions to be discriminative.

2.2. Testing Stage

When a test image $\tilde{I}$ is required to be labeled, we first obtain its warped atlas based weak segmentations $\{\tilde{\boldsymbol{\Phi}}_t(i), t = 1, \ldots, M, i \in \Omega\}$. During testing, we assume that images with similar features result in similar segmentations, which is also assumed in many label fusion methods. In other words, images with similar features result in similar generative models. In order to utilize w resulting from the training stage, we first warp the training intensity images and the associated w to the test image, then use the "deformed" w from "similar training images" to combine the warped multi-atlas segmentations of the test image. Therefore, the final strong segmentation $\hat{\boldsymbol{S}}(i)$ for the test image can be constructed as follows,

$$\hat{S}(i) = \sum_{n=1}^{N} \sum_{t=1}^{T} \gamma^n \tilde{\boldsymbol{\Phi}}_t(i)(\boldsymbol{w}_t^n(i) \circ f_{I^n \to \tilde{I}}(i)). \quad (4)$$

where, $f_{I^n \to \tilde{I}}$ is the deformation from the training intensity image I^n to the test image. w_t^n is the learned weight matrix associated with the n^{th} training image for the t^{th} weak segmenter. γ^n performs the role of weighting the trained weight matrices. For each voxel i, the label is assigned based on the structure with the largest value (probability) in the strong segmentation $\hat{\boldsymbol{S}}(i)$ with the exception that when $\hat{\boldsymbol{S}}(i) = 0$, we classify the voxel as background. *To avoid overfitting, in the testing stage, not all the warped training results are used.* Several techniques can be used to find the "similar" training images related to an incoming test image. A popular approach to achieve this is to find the K Nearest Neighbors. Formally, let $\{Y^n, n = 1, \ldots, N\}$ be the feature vector for each warped training image. (We need to warp each training image to the test image's coordinate system in order to evaluate the similarity.) Let $\tilde{Y}$ be the features for the given test image. We use the normalized variation of information to measure the dissimilarity between two feature vectors, so that γ_n can be defined as $\gamma^n = \frac{exp(-\alpha VI(Y^n, \tilde{Y}))}{\sum_{n=1}^{N} exp(-\alpha VI(Y^n, \tilde{Y}))}$. By tuning the value of α, we can automatically control the number of the nearest neighbors. In this paper, we use the vectorized intensity image as the feature.

3. EXPERIMENTAL RESULTS

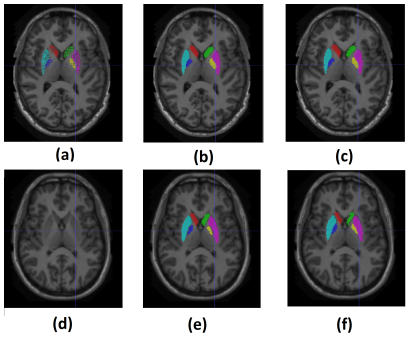

Fig. 2. Segmentation Example: (a) Partially labeled atlas, (b) Intermediate result: refined atlas (c) Fully labeled atlas (reference), (d) The test subject, (e) Segmentation using our algorithm PL-MAS, (f) Ground-truth segmentation of the test subject.

In this section, we experimentally validate our algorithm (PL-MAS) and compare it with the most recent work on segmentation using partially labeled atlases, called "GLa" [13]. We used the code provided by the authors of [13] for this comparison[1]. We

[1] https://github.com/lmkoch/multi-atlas-graph-labelling

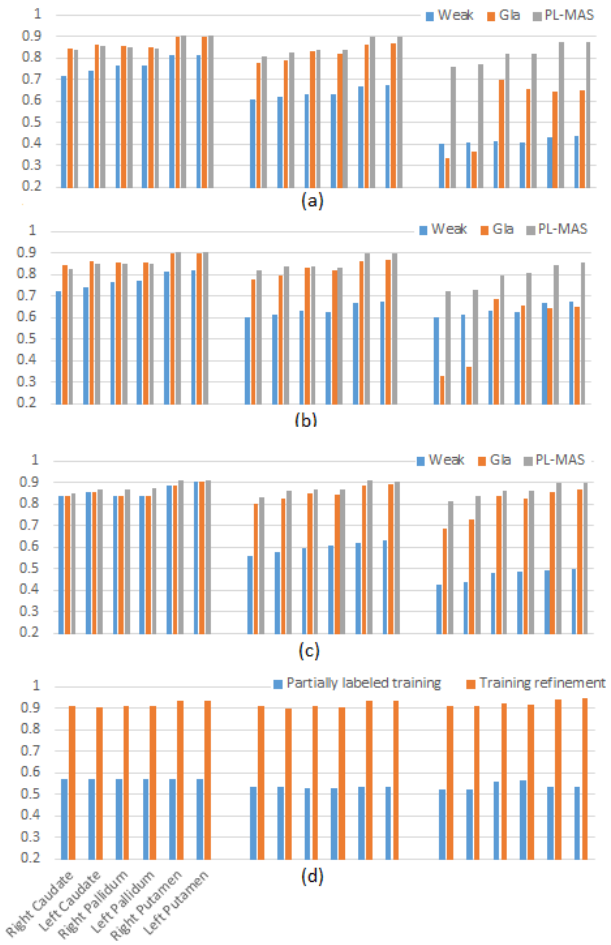

Fig. 3. Mean Dice Coefficient for testing when, (a) $20\%, 40\%, 60\%$ atlases' labels are locally missing, (b) $20\%, 40\%, 60\%$ atlases' labels are globally missing, (c) $20\%, 40\%, 60\%$ atlases' slices are missing. (d) Mean Dice Coefficient for refined atlases (intermediate results from PL-MAS) when 60% labels are (locally, globally, slice-wise) missing.

tested our algorithm on the MICCAI 2012 Grand Challenge dataset[2](Since the algorithms submitted for this challenge are not designed to segment MR image based on partially labeled atlases, we omit them here). There are 15 training and 20 test subjects respectively. We chose the shape complex composed of the Caudate, the Pallidum, and the Putamen, which are most affected in several neurological disorders. Since the shape complex is within a certain region of the brain, we therefore define a bounding box that approximately encloses the shape complex and only take this ROI from all the data sets as the input to our segmentation algorithm PL-MAS.

In "GLa", we use all of the 15 training subjects as the input multi-atlases. In our algorithm, for each training subject, we use the remaining 14 training subjects as the multi-atlases to learn the optimal weighted combination and the weights associated with this training subject are set to zero in the supervised learning stage. We created three different partially labeled multi-atlas settings to test our algorithm on. These settings are listed below.

- Missing Local Information: For each voxel location, we randomly remove $20\%, 40\%, 60\%$ atlases' labels.
- Missing Global Information: For each atlas, we randomly remove $20\%, 40\%, 60\%$ voxels' labels.
- Missing Slices: For each atlas, we randomly remove 20%, $40\%, 60\%$ slices.

We chose the parameter values $\lambda_1 = 100, \lambda_2 = 1, \alpha = 1$ by leave-one-out cross-validation solely on training data set.

Figure 2 shows an example where 40% labels are globally missing from the ground-truth atlas. It shows that PL-MAS gives a satisfactory segmentation estimate even when a large portion of the atlases' labels are missing. Aside from segmenting the novel image, our algorithm (the supervised learning stage) automatically produces a good refinement of the partially labeled atlas. We use the Dice's coefficient between the obtained segmentation result and fully labeled ground-truth segmentation to evaluate the performance. Figure 3 shows the mean Dice's coefficients for the 20 test cases based on the aforementioned partially labeled atlases. Here, the term "Weak" represents the warped atlas segmentation. When 20% atlases' labels are missing, "GLa" and PL-MAS give comparable results in all three settings. When the percentage of the missing labels increases, PL-MAS outperforms the competition. The mean Dice coefficient is approximately 0.8 for most structures even when 60% labels are missing. Meanwhile, the mean Dice of the 15 refined atlases obtained from PL-MAS is around 0.9 for all the structures when 60% labels are missing.

Overall, the results show that the optimal weighted combination learning based on the geometric properties of partially labeled atlases leads to a better generative model compared with the combination directly learned from the intensity image. However, since the geometric properties are learned from partially labeled data, some false local geometries resulting from the missing labels may influence the testing phase.

4. CONCLUSION

In this paper, we proposed a novel partially labeled multi-atlas based segmentation algorithm, PL-MAS, which learns a weighted weak segmenter combination. During the training stage, PL-MAS incorporates both non-overlap and voxel-wise smoothness constraints to preserve the inter- and intra-structural spatial relationships of the subcortical structures. Furthermore , it propagates the weights from the labeled regions to adjacent unlabeled regions. The training stage is formulated as a least-squares optimization problem that leads to solving a sparse linear system. In the testing stage, multiple anatomical structures can be simultaneously segmented based on the partially labeled multi-atlas segmentations. The experimental results show that our algorithm successfully segments the data even when a large portion of labels are missing in the atlases. This suggests the ability of PL-MAS to significantly reduce the workload involved in manual labeling tasks.

[2]https://masi.vuse.vanderbilt.edu/workshop2012/index.php/Main_page

5. REFERENCES

[1] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation," *IEEE Trans. Med. Imag.*, vol. 23, no. 7, pp. 903–921, 2004.

[2] X. Artaechevarria, A. Munoz-Barrutia, and C. O. de Solorzano, "Combination strategies in multi-atlas image segmentation: Application to brain MR data," *IEEE Trans. Med. Imag.*, vol. 28, no. 8, pp. 1266–1277, 2009.

[3] M. Sdika, "Combining atlas based segmentation and intensity classification with nearest neighbor transform and accuracy weighted vote," *Medical Image Analysis*, vol. 14, no. 2, pp. 219–226, 2010.

[4] M. R. Sabuncu, B. T. T. Yeo, K. V. Leemput, P. Golland, and B. Fischl, "A generative model for image segmentation based on label fusion," *IEEE Trans. Med. Imag.*, vol. 29, no. 10, pp. 1714–1729, 2010.

[5] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. A. Viergever, and B. van Ginneken, "Multi-atlas-based segmentation with local decision fusion–application to cardiac and aortic segmentation in CT scans," *IEEE Trans. Med. Imaging*, vol. 28, no. 7, pp. 1000–1010, 2009.

[6] A. R. Khan, N. Cherbuin, W. Wen, K. J. Anstey, P. Sachdev, and M. F. Beg, "Weights for local multi-atlas fusion using supervised learning and dynamic information (SuperDyn): Validation on Hippocampus segmentation," *NeuroImage*, vol. 56, no. 1, pp. 126–139, 2011.

[7] T. Chen, B. C. Vemuri, A. Rangarajan, and S. J. Eisenschenk, "Mixture of segmenters with discriminative spatial regularization and sparse weight selection," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI(3),*, 2011, pp. 595–602.

[8] H. Wang, J. W. Suh, S. Das, J. Pluta, C. Craige, and P. Yushkevich, "Multi-atlas segmentation with joint label fusion," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 611–623, 2013.

[9] J. E. Iglesias, M. R. Sabuncu, and K. V. Leemput, "A unified framework for cross-modality multi-atlas segmentation of brain MRI," *Medical Image Analysis*, vol. 17, no. 8, pp. 1181–1191, 2013.

[10] A. Gooya, K. M. Pohl, M. Bilello, G. Biros, and C. Davatzikos, "Joint segmentation and deformable registration of brain scans guided by a tumor growth model," in *MICCAI (2)*, 2011, pp. 532–540.

[11] G. Wu, M. Kim, G. Sanroma, Q. Wang, B. C. Munsell, and D. Shen, "Hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition," *NeuroImage*, vol. 106, pp. 34–46, 2015.

[12] B. A. Landman, A. J. Asman, A. G. Scoggins, J. A. Bogovic, F. Xing, and J. L. Prince, "Robust statistical fusion of image labels," *IEEE Trans. Med. Imaging*, vol. 31, no. 2, pp. 512–522, 2012.

[13] L. M. Koch, M. Rajchl, T. Tong, J. Passerat-Palmbach, P. Aljabar, and D. Rueckert, "Multi-atlas segmentation as a graph labelling problem: Application to partially annotated atlas data," in *Information Processing in Medical Imaging - 24th International Conference, IPMI*, 2015, pp. 221–232.