

OPTIMAL PARAMETER MAP ESTIMATION FOR SHAPE REPRESENTATION: A GENERATIVE APPROACH

Shireen Y. Elhabian, Praful Agrawal and Ross T. Whitaker

Scientific Computing and Imaging Institute, University of Utah, USA

ABSTRACT

Probabilistic label maps are a useful tool for important medical image analysis tasks such as segmentation, shape analysis, and atlas building. Existing methods typically rely on blurred signed distance maps or smoothed label maps to model uncertainties and shape variabilities, which do not conform to any generative model or estimation process, and are therefore suboptimal. In this paper, we propose to learn probabilistic label maps using a generative model on given set of binary label maps. The proposed approach generalizes well on unseen data while simultaneously capturing the variability in the training samples. Efficiency of the proposed approach is demonstrated for consensus generation and shape-based clustering using synthetic datasets as well as left atrial segmentations from late-gadolinium enhancement MRI.

Index Terms— parameter map, probabilistic labeling, generative models, consensus generation, shape representation

1. INTRODUCTION

Uncertainty in boundary of an anatomical shape is common in medical imaging applications involving soft tissue imaging, including neurology, cardiology, and oncology. Heterogeneous pixel intensities, imaging artifacts, partial volume effects and ill-defined boundaries usually induce a level of disagreement among human raters as well as (semi) automated segmentation algorithms in defining voxel-wise *labeling*. Likewise, differences in anatomy between subjects, patients, or specimens introduces variability in what one might expect to find in an image. Quantifying these uncertainties can benefit a variety of medical applications, such as label fusion in multi-atlas segmentation [1, 2], deformable atlas building [3, 4, 5], segmentation [5, 6, 7, 8], tractography [9] and longitudinal anatomical studies [10, 11].

Uncertainties associated with a voxel-wise label assignment can be encoded using a *probabilistic label* which reflects the likeliness of assigning a specific label to a voxel. It is typically represented as a vector of $L - 1$ non-negative fractions that sum up to one where L denotes the number of objects including the background. A probabilistic label map defined on the spatial domain of an image can thus be considered as a *parametric* distribution over the space of label maps.

A *label map* \mathbf{f} defines a labeling function which maps each image voxel $\mathbf{x} \in \mathbb{R}^d$ ($d = 2, 3$) to a single label from a set of labels $\mathcal{L} = \{0, 1, \dots, L\}$. Without loss of generality, we will focus on a single object scenario where $L = 2$. Consider a raster defined over spatial domain $\Omega \subset \mathbb{R}^d$ containing M pixels. The object of interest (i.e. foreground) $\omega \subset \Omega$ is represented by a label map $\mathbf{f} \in \{0, 1\}^M$ where $\mathbf{f}(\mathbf{x}) = 1$, iff $\mathbf{x} \in \omega$.

The authors would like to acknowledge Sam Preston for fruitful discussions. Josh Cates, the CARMA Center, CIBC by NIH Grant P41 GM103545-14 and NAMIC through NIH Grant U54 EB005149, for providing Utah fibrosis data.

In a *generative* sense, a label map \mathbf{f} is considered as a realization of a field of M independent Bernoulli random variables (multinomial in case of $L > 2$) defined on the domain Ω with a voxel-wise *parameter* $\mathbf{q}(\mathbf{x}) \in [0, 1]$. This parameter quantifies the probability of a voxel \mathbf{x} to be labeled as the foreground object, i.e. $\mathbf{q}(\mathbf{x}) = p(\mathbf{x} \in \omega)$. The *parameter map* $\mathbf{q} : \Omega \rightarrow [0, 1]^M$ defines a probability distribution over the observed label maps and models, and describes the underlying uncertainty about the label assignment. The probability of a label map generated from such a distribution, assuming i.i.d. voxels, can be factored out in terms of voxel-wise probabilities given by,

$$p(\mathbf{f}(\mathbf{x})|\mathbf{q}(\mathbf{x})) = [\mathbf{q}(\mathbf{x})]^{\mathbf{f}(\mathbf{x})} [1 - \mathbf{q}(\mathbf{x})]^{1-\mathbf{f}(\mathbf{x})}. \quad (1)$$

To relax the voxel-wise independency assumption, contextual information can be incorporated using spatial priors such as Markov random fields (MRF). In this case the *generative model* for binary label maps becomes:

$$p(\mathbf{f}|\mathbf{q}) = \prod_{\mathbf{x} \in \Omega} p(\mathbf{f}(\mathbf{x})|\mathbf{q}(\mathbf{x})) \times \frac{1}{Z} \exp\left(-\frac{1}{T}U(\mathbf{f})\right) \quad (2)$$

where $U(\mathbf{f})$ are clique potentials that favor spatially coherent label maps. Z is a normalization constant and T is a constant called the temperature.

Existing methods define the parameter map \mathbf{q} through heuristics that do not typically have a statistical foundation. For example, commonly used representations of parameter map include sigmoid of signed distance maps (SDMs) [1, 2, 4, 5, 8]; where distance of a voxel to the shape's boundary does not correlate well with the underlying probability distribution over label maps. Another popular representation is smoothed average of label maps [3, 5, 6, 8]. However blurring of label maps blindly smooths out shape features irrespective of the degree of uncertainty along shape boundary. Therefore, optimized weighted average of label maps are often used as an alternative representation [12, 13, 14]; nonetheless such an approach has been shown to overfit the limited training data [15].

In this paper, we propose to estimate a parameter map from a set of image segmentations, using a generative model a maximum-a-posteriori (MAP) formulation. This formulation gives an optimal parametric representation that encodes uncertainties in the observed label maps while being generalizable to unseen data. We further demonstrate efficiency of the proposed approach on medical imaging applications: consensus generation and shape-based clustering using synthetic datasets as well as real datasets of left atrial segmentations from late-gadolinium enhancement MRI.

2. METHODS

Consider a set of label maps $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_N\}$ independently drawn from a population. We seek to estimate the optimal parameter map $\mathbf{q}^* : \Omega \rightarrow [0, 1]^M$ which corresponds to the generative model described by voxel-wise Bernoulli distributions and MRF Gibbs potential. Since \mathbf{q} is a constrained field where $\mathbf{q}(\mathbf{x}) \in [0, 1]$, formulating the parameter estimation problem w.r.t. \mathbf{q} would require either

imposing a conjugate prior, i.e. Beta distribution, that would pull estimates of \mathbf{q} to $[0, 1]$, or solving a constrained optimization problem subject to inequality constraints. It is worth noting that imposing a conjugate prior would introduce spatially varying hyperparameters in order to define the voxel-wise prior distribution. This would further require a hyperprior to favor smooth spatial changes of these hyperparameters. Instead, we opt to solve an unconstrained optimization problem in terms of the log-likelihood ratio ϕ (a.k.a. log-odds [5]). The merit of log-odds lies in being an unconstrained, real-valued field which corresponds directly to the parameter map \mathbf{q} ,

$$\phi(\mathbf{x}) = \log \frac{p(\mathbf{x} \in \omega)}{p(\mathbf{x} \in \Omega \setminus \omega)}, \quad \mathbf{q}(\mathbf{x}) = \frac{1}{1 + e^{-\phi(\mathbf{x})}}. \quad (3)$$

$$= \log \frac{\mathbf{q}(\mathbf{x})}{1 - \mathbf{q}(\mathbf{x})} \quad (4)$$

The Bayesian formulation of our estimation problem amounts to finding the optimal log-ratio transform ϕ^* that maximizes log-posterior probability, i.e. $\log p(\phi|\mathcal{F})$. Assuming independent label maps, the posterior can be written as,

$$p(\phi|\mathcal{F}) \propto p(\mathbf{f}_1, \dots, \mathbf{f}_N|\phi)p(\phi) = \left[\prod_{n=1}^N p(\mathbf{f}_n|\phi) \right] p(\phi). \quad (5)$$

By construction, the likelihood of a label map $p(\mathbf{f}_n|\phi)$ can be defined as $p(\mathbf{f}_n|\mathbf{q})$. It should be clear that the choice of MRF prior on label maps in (2) does not affect ϕ -optimization. Using (3), it can be shown that the log-likelihood can be written as follows, with $\bar{\mathbf{f}}$ being the average label map of \mathcal{F} ,

$$\log p(\mathbf{f}_1, \dots, \mathbf{f}_N|\phi) = N \left([\bar{\mathbf{f}} - 1]^T \phi - \log(1 + e^{-\phi}) \right). \quad (6)$$

The spatial coherency of ϕ is promoted through a smoothness prior $p(\phi)$ that encourages smooth transitions in local neighborhoods. MRF prior on ϕ can be written as a Gibbs distribution, $p(\phi) \propto \exp\{-\lambda U(\phi)\}$; $\lambda \geq 0$. λ is a hyperparameter which governs the impact of smoothness prior relative to the data fidelity term (6). The Gibbs energy $U(\phi)$ acts as an edge detector which penalizes violation of the smoothness assumption. Hence, we use laplacian-square energy, i.e. $U(\phi) = \|\Delta\phi\|^2$. We coin the term *ShapeOdds* to refer to the estimated log-odds map due to its spatial coherency structure. Its MAP estimate can be written as,

$$\phi^*(\lambda, \mathcal{F}) = \arg \min_{\phi} -\log p(\phi|\mathcal{F}, \lambda) = \arg \min_{\phi} E(\phi|\mathcal{F}, \lambda) \quad (7)$$

where prior energy in the objective function E can be written in terms of the Laplacian operator L such that,

$$E(\phi|\mathcal{F}, \lambda) = N \log(1 + e^{-\phi}) - N [\bar{\mathbf{f}} - 1]^T \phi + \frac{\lambda}{2} \|L\phi\|^2 \quad (8)$$

The first variation of (8) reads as,

$$\frac{dE}{d\phi} = N [\mathbf{q} - \bar{\mathbf{f}}] + \lambda L^\dagger L \phi \quad (9)$$

where $L^\dagger L$ denotes the bilaplacian (a.k.a. biharmonic) operator. It is worth mentioning that the objective function E is *convex* w.r.t. ϕ due to the positive definiteness of bilaplacian operator and the constrained values attained by \mathbf{q} which would result in a positive definite Hessian matrix.

We can solve for a global optimum with a gradient descent scheme. In order to enable large time steps Δt while maintaining stable updates, we use a *semi-implicit scheme* with finite-forward time marching to define an iterative update for ϕ ,

$$\phi^{(t+1)} = \left(\frac{1}{1 + \lambda \Delta t L^\dagger L} \right) \otimes \left(\phi^{(t)} - \Delta t N [\mathbf{q} - \bar{\mathbf{f}}] \right) \quad (10)$$

where spatial convolution \otimes can be efficiently performed as multiplication in Fourier domain. Note that (10) forms a data-driven smoothing operator which respects the underlying uncertainty along shape boundary. Further, we can use a voting-based initialization for the parameter map \mathbf{q} . Initial probabilities $\mathbf{q}(\mathbf{x}) = p(\mathbf{x} \in \omega)$ can be computed based on the frequency of $\mathbf{x} \in \omega$ in the given set \mathcal{F} . Our experiments demonstrate the insensitivity of ShapeOdds estimation to the initial solution using voting-based, all zero, or random initializations. To avoid log-ratios blowing up at voxels that have unanimous values in the training data, we relax the input infinitesimally, so that $\mathbf{f}_n(\mathbf{x}) = 1 - \epsilon$ if $x \in \omega$ and $\mathbf{f}_n(\mathbf{x}) = \epsilon$ otherwise, for $\epsilon > 0$. We have found that the ϕ -optimization is not sensitive to such setting, and we use $\epsilon = 1e - 6$ for the experiments in this paper.

Estimation of the optimal ShapeOdds ϕ^* requires knowledge of the hyperparameter λ which results in smooth parameter maps without overfitting the given set of label maps. In the limit of infinite sample size, i.e. $N \rightarrow \infty$, ShapeOdds converges to the average label map $\bar{\mathbf{f}}$ which precisely models the label maps distribution. As such, the role of smoothness prior asymptotically diminishes with $\lambda \rightarrow 0$. With the limited sample size, common in practice, λ is estimated from training samples while maximizing the expected generalization performance. Hence, estimating λ can be considered as a *model selection* problem. This motivates a K -fold cross-validation approach for tuning λ for a training set \mathcal{F} . We can formulate our model selection criterion S as the cross-validation estimate of negative log-likelihood of unseen (held-out) label maps.

$$S(\lambda, \mathcal{F}, K) = \frac{1}{K} \sum_{k=1}^K \frac{1}{|\mathcal{F}_k|} \sum_{\mathbf{f}_n \in \mathcal{F}_k} -\log p(\mathbf{f}_n|\phi^*(\lambda, \mathcal{F}_{-k})) \quad (11)$$

where \mathcal{F}_k is the set of label maps contained in the k -th fold while \mathcal{F}_{-k} contains the remaining label maps to be used for training. A grid search can be performed to exhaustively sweep the hyperparameter space, nonetheless we have shown that S is convex w.r.t. λ , which enables smarter/faster solutions by sweeping through a hierarchical (i.e. coarse-to-fine) grid search. The best hyperparameter $\lambda^*(\mathcal{F}, K)$ is then used to estimate the optimal map $\phi^*(\lambda^*, \mathcal{F})$ for entire training set.

3. EXPERIMENTAL RESULTS

The proposed parameter map estimation can be used for a broad class of applications which require building a probability distribution over the label map space while being generalizable to unseen samples. For instance, it can be used to model shape priors for Bayesian image segmentation. It can also serve as a shape class prototype in applications that involve shape matching such as multi-atlas segmentation. In this paper, we evaluate our model w.r.t. two medical applications: consensus generation and shape clustering. We start off with a proof-of-concept in order to assess the performance of proposed parameter map estimation compared to existing representations.

3.1. Proof-of-Concept Experiments

In order to simulate uncertainties associated with a segmentation process, consider a 2D a rectangular shape where its imaging resulted in weak boundaries along its right side. A set of 100 hypothetical segmentations (i.e. label maps) were generated using a rectangle undergoing random spatial deformations where each deformation field was convolved with a spatially varying Gaussian kernel to reflect increased degrees of disagreement on the right side. Training subsets of $N = [4, 32]$ in increments of 4 were randomly drawn 100 times. For each experiment, λ was first estimated using the model

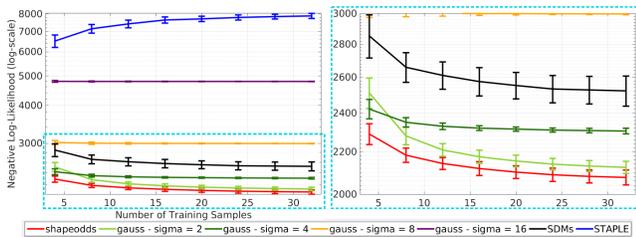


Fig. 1: Generalization performance (negative log-likelihood — lower is better) of different parameter maps as a function of training sample size. The right side is a zoom-in on the cyan part of the graph in the left side. Image size: 150×150 .

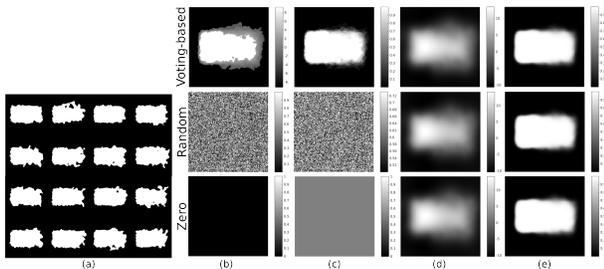


Fig. 2: Effect of initialization: (a) Training label maps, (b) ϕ^0 , (c) q^0 , (d) ϕ^* and (e) q^* . Second row shows voting-based initialization, third row shows random initialization and fourth row shows zero initialization. Mean absolute difference of estimated parameter maps using different initialization is 9.893×10^{-5} .

selection criterion defined in (11) and then used to estimate the optimal parameter map for training set of label maps. Parameter maps based on sigmoid of an average SDM, Gaussian smoothed average label map (with different kernel sizes) and probabilistic labeling resulting from STAPLE were computed for comparison. The negative log-likelihood (6) over the held-out/testing label maps was used as the generalization error for an estimated parameter map.

Figure 1 demonstrates the generalization performance of ShapeOdds compared to other parameter maps as a function of the available training sample size. One can observe the poor generalization of STAPLE-based probabilistic labels revealing its tendency to overfit the training dataset. SDM-based parameter maps, on the other hand, make use of more training samples for better generalization. Nonetheless, lesser performance indicate that they do not correlate well with the underlying generative process. Smoothed average label maps generalize better than SDMs and STAPLE while losing the ability to model the probability distribution of the given label maps due to the blind smoothing along shape boundaries irrespective of the underlying uncertainties. Figure 2 illustrates the effect of initialization on the final estimation of ShapeOdds and the corresponding parameter map where we considered voting-based, random and zero initializations for ϕ . One can notice the convergence to the same parameter map regardless of the initial solution. Figure 3 shows the mean and standard deviation of optimal hyperparameter λ^* , as a function of training sample size. Notice that it asymptotically approaches to zero with large number of training label maps, decreasing the effect of the smoothness prior on ϕ that enables ShapeOdds to generalize to unseen data. Further, Figure 3 plots the average of model selection criterion (11) as a function of λ for different training sizes. Notice the convexity of this function while the average of optimal λ decreases with increasing number of training samples.

3.2. Consensus Generation

With ill-defined boundaries and shape irregularities, accurately delineating the contour of left atrium (LA) wall from late-gadolinium

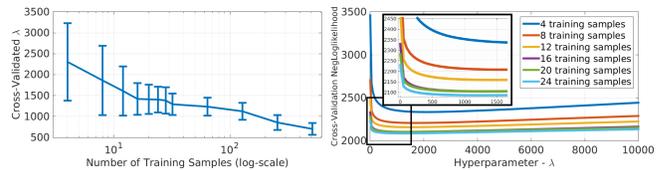


Fig. 3: Left: optimal hyperparameter λ found using coarse-to-fine grid search as function of training sample size. Right: average model selection criterion as function of λ for different training sizes. Image size: 150×150 .

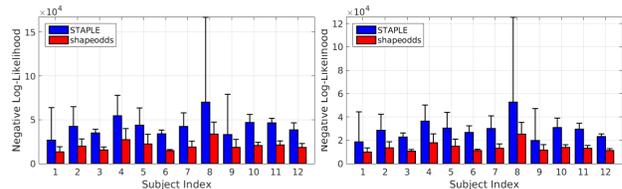


Fig. 4: The average and standard deviation of negative log-likelihood (lower is better) of held-out manual segmentations for epicardium (left) and endocardium (right) using ShapeOdds and STAPLE parameter maps. Image size: $188 \times 143 \times 149$.

enhancement cardiac (LGE) MRI is a challenging task for human raters as well as automatic segmentation algorithms. Manual segmentations of epicardium and endocardium typically exhibit variations among human experts reflecting uncertainties along shape boundaries, calling for consensus generation. A *good consensus* need to accommodate variations among raters as well as automated segmentation algorithms. In specific, estimating probability distribution over label maps enables modeling the random process associated with the segmentation process. The generality of such a distribution w.r.t. unseen samples, while being consistent to the given ones, is thus crucial to provide means to evaluate new manual/automated segmentations as being drawn from similar distribution. To that regard, we collected 12 LGE-MRI of LA from patients with atrial fibrillation (pre-ablation). Each scan was segmented by three human experts defining epi/endocardium regions. We estimated the ShapeOdds for epi/endocardium of each patient using leave-one-out strategy.

Figure 4 shows the average and standard deviation of negative log-likelihood of held-out samples for epicardium and endocardium. Notice that ShapeOdds parameter maps is able to generalize to unseen samples compared to those of STAPLE. Figure 5 shows a sample LGE-MRI slice along with the manual segmentations and the corresponding estimated parameters maps. One can observe the tendency of STAPLE to produce probabilistic labeling that overfits the training samples. In particular, STAPLE is estimating weights for the given label maps to produce an optimized weighted average. With two label maps for training, STAPLE produces a nonsmooth parameter map which limits its generalization aspect. One the other hand, ShapeOdds provide smooth parameter maps that are generalizable to unseen samples.

3.3. Shape Clustering

Proposed generative model can also be used for unsupervised clustering of label maps. This can be useful for multi-atlas segmentation algorithms which typically require a large database of atlases. In particular, shape-based clustering can benefit atlas selection algorithms to exclude irrelevant atlases that might misguide the segmentation process. Further, representative parameter maps can be used to model multi-modal shape prior distributions for image segmentation applications. Representative parameter maps of a set of training label maps can be learned in an expectation-maximization (EM) fash-

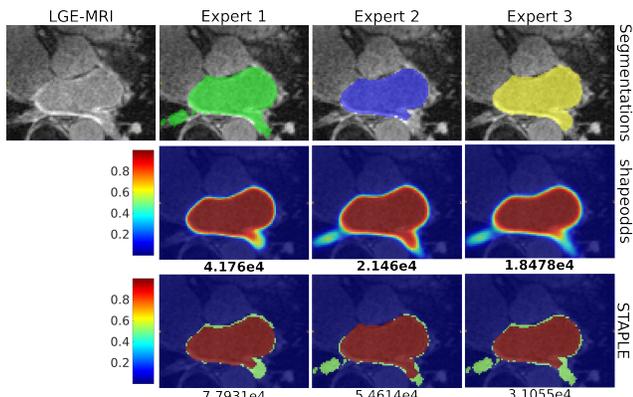


Fig. 5: Consensus of the epicardium of sample LGE-MRI slice. The first row shows the manual segmentations of three human experts who disagree on the accurate contour of epicardium. The second and third rows show the parameter maps estimated via ShapeOdds and STAPLE, respectively, by leaving the respective expert segmentation out of the training sample. Negative log-likelihood of the expert segmentation given the ShapeOdds/STAPLE parameter map is also reported.

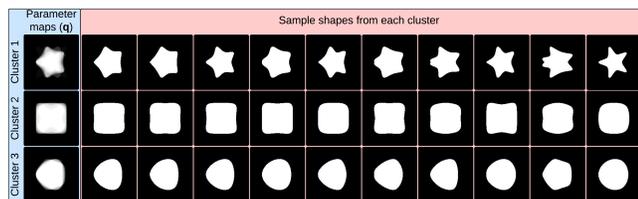


Fig. 6: Synthetic data clustering results. Samples for each cluster are presented in decreasing order of log-likelihood (left to right).

ion while simultaneously clustering similar shapes via maximizing the likelihood in 6. We demonstrate the results on synthetic dataset containing 90 supershapes [16] corresponding to three, four and five rotational symmetries as well as a real dataset of 62 LGE-MRI of LA from patients with atrial fibrillation (pre-ablation). Initial clusters are assigned based on maximum log-likelihood from multiple random assignments. Figures 6 and 7 demonstrate the parameter maps and sample shapes corresponding to final clusters. Results indicate that samples with similar shape characteristics are assigned to the same cluster.

4. CONCLUSION

This paper proposed a generative approach to model uncertainties in boundaries of anatomical structures as a parameter map, beneficial for medical applications such as segmentation, atlas building and consensus generation. Proposed approach performed better as compared to existing practices such as blurred signed distance maps,

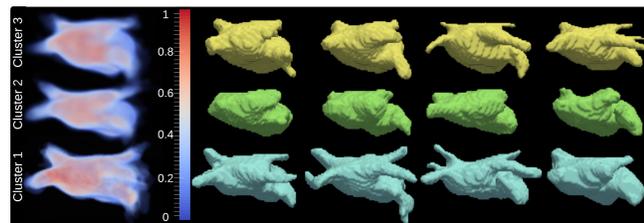


Fig. 7: Clustering results using left atrium dataset. First column shows the estimated representative parameter maps corresponding to final clusters. Further columns display samples for each cluster presented in decreasing order of log-likelihood (left to right).

smoothed average of label maps and weighted average of label maps. The parameter map obtained is generalizable even with few training samples.

5. REFERENCES

- [1] MR Sabuncu, BTT Yeo, KV Leemput, B Fischl, and P Golland, "A generative model for image segmentation based on label fusion," *IEEE TMI*, vol. 29, no. 10, pp. 1714–1729, 2010.
- [2] JE Iglesias, MR Sabuncu, and KV Leemput, "A unified framework for cross-modality multi-atlas segmentation of brain MRI," *MedIA*, vol. 17, no. 8, pp. 1181–1191, 2013.
- [3] J Ashburner and KJ Friston, "Computing average shaped tissue probability templates," *Neuroimage*, vol. 45, no. 2, pp. 333–341, 2009.
- [4] TR Raviv, KV Leemput, BH Menze, WM Wells, and P Golland, "Segmentation of image ensembles via latent atlases," *MedIA*, vol. 14, no. 5, pp. 654–665, 2010.
- [5] KM Pohl, J Fisher, S Bouix, M Shenton, RW McCarley, WEL Grimson, R Kikinis, and WM Wells, "Using the logarithm of odds to define a vector space on probabilistic atlases," *MedIA*, vol. 11, no. 5, pp. 465–477, 2007.
- [6] PL Bazin and DL Pham, "Homeomorphic brain image segmentation with topological and statistical atlases," *MedIA*, vol. 12, no. 5, pp. 616–625, 2008.
- [7] KM Pohl, R Kikinis, and WM Wells, "Active mean fields: Solving the mean field approximation in the level set framework," in *IPMI*, 2007, pp. 26–37.
- [8] KM Pohl, J Fisher, R Kikinis, WEL Grimson, and WM Wells, "Shape based segmentation of anatomical structures in magnetic resonance images," in *CVBIA*, pp. 489–498, 2005.
- [9] CJ Brown, BG Booth, and G Hamarneh, "K-confidence: Assessing uncertainty in tractography using K optimal paths," in *ISBI*, 2013, pp. 250–253.
- [10] PA Habas, K Kim, JM Corbett-Detig, F Rousseau, OA Glenn, AJ Barkovich, and C Studholme, "A spatiotemporal atlas of MR intensity, tissue probability and shape of the fetal brain with application to segmentation," *Neuroimage*, vol. 53, no. 2, pp. 460–470, 2010.
- [11] E Dittrich, TR Raviv, G Kasprian, R Donner, PC Brugger, D Prayer, and G Langs, "A spatio-temporal latent atlas for semi-supervised learning of fetal brain segmentations and morphological age estimation," *MedIA*, vol. 18, no. 1, pp. 9–21, 2014.
- [12] SK Warfield, KH Zou, and WM Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE TMI*, vol. 23, no. 7, pp. 903–921, 2004.
- [13] A Suinesiaputra, BR Cowan, AO Al-Agamy, MA Elattar, N Ayache, AS Fahmy, AM Khalifa, PM Gracia, MP Jolly, AH Kadish, DC Lee, J Margeta, SK Warfield, and AA Young, "A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images," *MedIA*, vol. 18, no. 1, pp. 50–62, 2014.
- [14] H Wang, JW Suh, SR Das, JB Pluta, C Craige, and P Yushkevich, "Multi-atlas segmentation with joint label fusion," *IEEE TPAMI*, vol. 35, no. 3, pp. 611–623, 2013.
- [15] WR Crum, O Camara, and DLG Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE TMI*, vol. 25, no. 11, pp. 1451–1461, 2006.
- [16] J Gielisz, "A generic geometric transformation that unifies a wide range of natural and abstract shapes," *American Journal of Botany*, vol. 90, no. 3, pp. 333–338, 2003.