Published in final edited form as:

Stat Atlases Comput Models Heart. 2010 ; 6364: 85–94. doi:10.1007/978-3-642-15835-3_9.

Robust Atlas-Based Segmentation of Highly Variable Anatomy: Left Atrium Segmentation

Michal Depa[1], **Mert R. Sabuncu**[2], **Godtfred Holmvang**[3], **Reza Nezafat**[4], **Ehud J. Schmidt**[5], and **Polina Golland**[1]

[1]Computer Science and Artificial Intelligence Lab, MIT, Cambridge, MA, USA

[2]Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, MA, USA

[3]Cardiac MRI Unit, Massachusetts General Hospital, Boston, MA, USA

[4]Department of Medicine (Cardiovascular Division), Beth Israel Deaconess Medical Center, Boston, MA, USA

[5]Department of Radiology, Brigham & Women's Hospital, Boston, MA, USA

Abstract

Automatic segmentation of the heart's left atrium offers great benefits for planning and outcome evaluation of atrial ablation procedures. However, the high anatomical variability of the left atrium presents significant challenges for atlas-guided segmentation. In this paper, we demonstrate an automatic method for left atrium segmentation using weighted voting label fusion and a variant of the demons registration algorithm adapted to handle images with different intensity distributions. We achieve accurate automatic segmentation that is robust to the high anatomical variations in the shape of the left atrium in a clinical dataset of MRA images.

Keywords

Atlas-based segmentation; left atrium segmentation; cardiac segmentation; label fusion; non-rigid registration

1 Introduction

The high anatomical variability of the heart's left atrium makes its segmentation a particularly difficult problem. Specifically, the shape of the left atrium cavity, as well as the number and locations of the pulmonary veins connecting to it, vary substantially across subjects (Fig. 1). In this paper, we propose and demonstrate a robust atlas-based method for automatic segmentation of the left atrium in contrast-enhanced magnetic resonance angiography (MRA) images.

Clinically, left atrium segmentation is a highly relevant problem. Atrial fibrillation is known to be one of the most common heart conditions. It manifests itself by causing irregular contractions of the heart's atria and can have serious consequences such as stroke and heart failure [1, 2]. Catheter-based radio-frequency ablation has recently emerged as a treatment for this condition. It involves burning the cardiac tissue that is responsible for the re-entry electrical currents that cause fibrillation. The high anatomical variability of the left atrium

shape and the pulmonary veins that enter it presents significant difficulties for cardiac ablation since it is commonly performed at the junction of the atrial body and pulmonary veins. Consequently, accurate visualization of the patient's left atrium promises to substantially improve intervention planning. The knowledge of the shape of the left atrium can also aid in the subsequent segmentation of the resulting ablation scars and thus in the evaluation of the outcome of the procedure [3].

One approach to segment the left atrium is whole heart segmentation, where all of the heart chambers, and sometimes other structures, are included in a single model and segmented simultaneously. Unfortunately, most whole heart segmentation methods do not model the pulmonary veins of the left atrium [4, 5]. An exception is [6], where the geometrical model of the heart constructed from CT images includes the pulmonary veins. However, the approach involves building a mean shape model that will face considerable challenges in the presence of topological differences in anatomy.

An alternative approach is to focus on segmentation of the left atrium by first extracting the whole blood pool by intensity thresholding and then separating it into different heart chambers by making cuts at narrowings [7]. This work was extended to allow tracking of centerlines of the pulmonary veins entering the atrium [8, 9]. The method however suffers from requiring several thresholds to be set manually because of varying intensity distributions and anatomies of the left atrium across patients.

In this work, we perform the segmentation via a label fusion algorithm [10, 11] that uses a training set of MRA images of different patients with corresponding manual segmentations. We first align the training images to the test subject image to be segmented and apply the resulting deformations to the corresponding manual segmentation label maps to yield a set of left atrium segmentations in the coordinate space of the test subject. These form a non-parametric subject-specific statistical atlas. We then use a weighted voting algorithm to assign every voxel to the left atrium or to the background. A similar approach was demonstrated in [12] for segmentation of the aorta and heart extent in CT images. In contrast, we aim to delineate the considerably more complex structure of the left atrium. This requires more powerful label fusion and registration algorithms. Notably, we use a weighted label fusion scheme that assigns higher weights to voxels in training segmentations that are located deeper within the structure of interest and that have similar intensities in training and test images [11]. We also handle varying intensity distributions between images by incorporating iterative intensity equalization in a variant of the demons registration algorithm [13, 14] used for the registration of the training images to the novel test image.

We demonstrate fully automatic, accurate segmentations of both the atrial body and pulmonary veins connected to it on a set of 16 clinical MRA images. Our method captures all of the pulmonary veins in all patients in our dataset. Comparison to traditional atlas-based segmentation and majority voting non-parametric segmentation demonstrates the advantage of the proposed method for this problem.

2 Methods

In this section we describe the registration and segmentation algorithms we employ in this work. We let $\{I_i\}$ be the set of N training images, $\{L_i\}$ be the set of corresponding expert manual segmentations and $\{\Phi_i\}$ be the warps from the training images $\{I_i\}$ to the test image I. Our goal is to estimate the label map L of the test image I.

2.1 Diffeomorphic Demons Registration with Intensity Equalization

We perform pairwise registrations by first aligning the images affinely using a mutual information metric [15], then using a diffeomorphic variant of the demons registration algorithm [16]. The method represents warps Φ with a smooth and stationary velocity field υ using a one-parameter subgroup of diffeomorphisms [17]. In this formulation, $\Phi(x) = \exp(\upsilon)(x)$, i.e., the flow of the velocity field at time one is equal to its equivalent deformation. In addition to guaranteeing diffeomorphic registration, this parametrization is computationally efficient and offers convenient access to the inverse deformation $\Phi^{-1}(x) = \exp(-\upsilon)(x)$. At each iteration, the incremental update velocity field u is found by minimizing the energy function [13]:

$$E(I_F, I_M, \Phi, u) = \|I_F - I_M \circ \Phi \circ \exp(u)\|^2 + \|u\|^2, \quad (1)$$

where I_F and I_M are the fixed and moving images respectively, and Φ is the warp at the current iteration. The new updated velocity field is then smoothed to optimize a regularization constraint.

One disadvantage of demons registration algorithms is that they are driven by intensity differences between images I_F and I_M. Although the MRA images we work with are of the same modality, the intensity distribution varies from one image to the next. To address this problem, we introduce an intensity transformation:

$$\tilde{I}_M(x) = \sum_{k=1}^{K} \theta_k b_k(I_M(x)) = B(I_M(x))\theta, \quad (2)$$

where $\{b_1(\cdot) \ldots b_K(\cdot)\}$ is the set of basis functions and $\theta = \{\theta_1 \ldots \theta_K\}$ is the vector of corresponding coefficients. This transformation effectively modifies the energy function we are optimizing:

$$E(I_F, I_M, \Phi, u) = \|I_F - B[I_M \circ \Phi \circ \exp(u)]\theta\|^2 + \|u\|^2. \quad (3)$$

Similar to [18], we use polynomial basis functions up to degree K. For a fixed velocity field u, Eq. (3) reduces to a standard linear least squares problem. We thus alternate between estimating coefficients $\{\theta_i\}$ from corresponding voxel pairs in $I_M \circ \Phi$ and I_F (using robust least squares with outlier detection) and performing the standard demons iteration.

2.2 Label Fusion Segmentation

Rather than summarize the training set through average statistics, label fusion algorithms keep the atlas in the form of the original training images with their expert manual segmentations. After registering the training images $\{I_i\}$ to the test image I, we obtain a non-parametric subject-specific atlas composed of N warped images and corresponding label maps.

To perform the segmentation, we use a weighted voting scheme at each voxel, taking into account not only the number of occurrences of each label, but also their locations in the manually segmented structures and the similarity between the intensities of corresponding voxels in the training and test images, similar to [11]. Formally, we compute the maximum a posteriori (MAP) estimate of the label map:

$$\hat{L} = \arg\max_{L} p(L|I, \{L_i, I_i, \Phi_i\}) = \arg\max_{L} p(L, I|\{L_i, I_i, \Phi_i\}). \quad (4)$$

We make a simplifying assumption that each voxel is generated from the training set independently from all other voxels. Furthermore, we assume that each training image is equally likely to generate any particular voxel a priori. The MAP estimation then reduces to an independent decision at each voxel:

$$\hat{L}(x) = \arg\max_{l \in 1,\ldots,\mathscr{L}} \sum_{i=1}^{N} p(L(x)=l, I(x)|L_i, I_i, \Phi_i) \quad (5)$$

$$= \arg\max_{l \in 1,\ldots,\mathscr{L}} \sum_{i=1}^{N} p(L(x)=l|L_i, \Phi_i) p(I(x)|I_i, \Phi_i), \quad (6)$$

where $\mathscr{L}$ is the total number of possible labels ($\mathscr{L} = 2$ in our case). Eq. (6) assumes that the label and intensity values at each voxel of the test image are conditionally independent given the warp Φ_i and the fact that they were generated from training subject i. This decision rule can be viewed as weighted soft voting with $p(L(x) = l|L_i, \Phi_i)$ providing the vote and $p(I(x)|I_i, \Phi_i)$ serving as a weight. We set weights using a Gaussian image likelihood:

$$p(I(x)|I_i, \Phi_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\left(I(x) - \tilde{I}_i(\Phi_i(x))\right)^2}, \quad (7)$$

where $\tilde{I}_i(\Phi_i(\cdot))$ is the training image I_i, registered to the test image I and intensity equalized by applying the intensity transformation estimated during the registration step. The weight is higher when the two corresponding voxels in the aligned images have similar intensities. We define the votes through the label likelihood term:

$$p(L(x)=l|L_i, \Phi_i) \propto e^{\rho D_i^l(\Phi_i(x))}, \quad (8)$$

where $D_i^l(\Phi_i(\cdot))$ is the signed Euclidean distance map of the manual segmentation of the training subject *i* in the coordinate space of the test subject and ρ is the slope parameter. Voxels that are inside the structure and farther from the boundary are assigned higher votes.

3 Results

We validate our method on a set of 16 electro-cardiogram gated Gadolinium-DTPA (0.2 mmol/kg) contrast-enhanced MRA images (CIDA sequence, TR= 4.3ms, TE=2.0ms, θ=40°, in-plane resolution varying from 0.51mm to 0.68mm, slice thickness varying from 1.2mm to 1.7mm, ±80 kHz bandwidth, atrial diastolic ECG timing to counteract considerable volume changes of the left atrium). We perform leave-one-out experiments by treating one subject as the test image and the remaining 15 as the training set, and repeating for each subject in the dataset. We use the Dice overlap score [19] between the automatic and expert manual segmentations as a quantitative measure of segmentation quality. Dice scores vary from 0 to 1, with 1 corresponding to perfect overlap.

In the label fusion segmentation algorithm, we set $\sigma = 100$ and $\rho = 1.5$. We explored the parameter space by varying σ between 50 and 500, and ρ between 0.3 and 2.5. During this process, we confirmed that our method is in fact robust to the choice of the parameters. The difference between the best and the worst Dice scores obtained for each subject while varying the parameters was 0.05 ± 0.03. We also explored different values for the polynomial degree of the intensity transformation in the registration algorithm. We varied the degree from 1 to 5 and found that it had similarly little effect on the results, with a 0.008 ± 0.007 difference between the best and worst overlap scores for each subject. We chose a degree of 3 because it provided the highest overall Dice scores.

We compare our method of weighed voting (WV) label fusion to three alternative atlas-based approaches: majority voting (MV) label fusion, parametric atlas thresholding (AT) and atlas-based EM-segmentation (EM). The majority voting label fusion is similar to weighted voting, except it assigns each voxel to the label that occurs most frequently in the registered training set at this voxel [10, 20]. We also construct a parametric atlas that summarizes all 16 subjects in a single template image and a probabilistic label map by performing group-wise registration to an average space. After registering this new atlas to the test subject, we segment the left atrium using two different approaches. In atlas thresholding, we simply threshold the warped probabilistic label map at 0.5 to obtain the segmentation. We also use this parametric atlas as a spatial prior in a traditional model-based EM-segmentation [21]. Note that this construction favors the baseline algorithms as it includes the test image in the registration of all subjects into a single coordinate frame.

In our application, correctly segmenting all of the pulmonary veins of the left atrium is crucial. Therefore it is important to visually inspect the resulting segmentations to fully evaluate them. Fig. 2 shows segmentation outlines of expert manual segmentations and the four methods we compare on corresponding slices of four different subjects. In the first row, majority voting and atlas thresholding miss a pulmonary vein that is correctly identified by our approach. EM-segmentation segments that vein only partially while at the same time producing false positives in the aorta and atrial body. The second and third rows show

similar situations. In the last row, all methods correctly segment the pulmonary veins, but our method produces the most accurate outlines. Detailed analysis of all subjects shows that our method does not miss a single pulmonary vein in the whole dataset, in spite of the high anatomical variability.

Fig. 3 reports the segmentation accuracy for each method, as measured by the volume overlap Dice scores. We also report the differences in segmentation accuracy between our method and the benchmark algorithms. To compute the difference between two methods, we subtract the Dice score of the second method from the score of the first for each subject. Our approach clearly outperforms other algorithms (WV vs. MV: $p < 10^{-9}$, WV vs. AT: $p < 0.002$, WV vs. EM: $p < 0.003$; single-sided paired t-test). To focus the evaluation on the critical part of the structure, we manually isolate the pulmonary veins in each of the manual and automatic segmentations, and compare the Dice scores for these limited label maps. Again, we observe consistent improvements offered by our approach (WV vs. MV: $p < 10^{-7}$, WV vs. AT: $p < 10^{-7}$, WV vs. EM: $p < 0.03$; single-sided paired t-test). Since atlas-based EM-segmentation is an intensity based method, it performs relatively well in segmenting pulmonary veins, but suffers from numerous false positives in other areas, which lower its overall Dice scores.

In Table 1, we present the computational cost for the different methods. The computation time consists of the time needed to perform the registrations and the time required by the segmentation step. We use an ITK implementation of the diffeomorphic demons registration algorithm [14] and implement the segmentation step in MATLAB. The weighted voting and majority voting label fusion methods register all of the training images (15 in our case) to the test subject. Each registration takes on average 8 minutes. The parametric atlas can be computed without any knowledge about the test image. Therefore, the parametric atlas thresholding and the atlas-based EM-segmentation require only a single registration of the atlas to the test subject.

4 Discussion and Conclusions

We demonstrated a non-parametric atlas-based method for automatic left atrium segmentation. This label fusion style approach first registers the whole training set to the test subject and then combines weighted votes from training subjects to make decisions. These votes are computed independently at each voxel and depend on the intensity similarity between the training and test images, as well as the voxel's location in the structure of interest. To handle global shifts in the intensity distribution across images, we modified the diffeomorphic demons registration algorithm to perform iterative intensity equalization during registration.

Experimental results illustrate the capacity of our method to handle high anatomical variability, yielding accurate segmentation and detecting all pulmonary veins in all subjects. By explicitly modeling the anatomical variability represented in the label maps and the corresponding training images, the proposed method outperforms traditional atlas-based segmentation algorithms and a simple label fusion benchmark.

This increased accuracy however comes at the cost of additional computation time since the whole training set needs to be registered to every test subject that is being segmented. Although the weighted voting label fusion approach is more computationally expensive than the other methods, this requirement does not pose a problem in our application because the left atrium segmentation does not need to be produced in real-time. The computation time can be substantially reduced by parallelizing the registration step since the registrations are independent from each other. Moreover, clustering training images, similar to the approach in [22], and using cluster centers as training templates can further reduce the number of necessary registrations. The registration algorithm itself also clearly affects the overall segmentation results and a careful study will be necessary to inform future development of the method.

We found that there was no clear relationship between our method's performance on a specific subject and the number of similar examples in the training set. For example, one subject in our dataset had a pulmonary vein that was not present in any of the other patients. Our method still produced an accurate segmentation of that vein, even with no similar left atrium anatomy in the training set. A more detailed analysis of the effects of sub-populations in the training set on the quality of the resulting segmentations is an interesting future research topic.

In addition to the benefits automatic segmentation offers for the planning stages of cardiac ablation, our method can also assist in the evaluation of the procedure outcome. Segmentation of the ablation scars in post-procedure images is a clinically relevant but difficult problem. Using left atrium segmentation as a prior for scar location is a promising future direction of research we will pursue.

Acknowledgments

This research was supported in part by NAMIC (NIH NIBIB NAMIC U54-EB005149), the NAC (NIH NCRR NAC P41-RR13218), the NIH NINDS R01-NS051826 grant, the NSF CAREER 0642971 grant, and the NIH R01EB008743-01A2 grant. Michal Depa was supported by the Irwin Mark Jacobs and Joan Klein Jacobs Presidential Fellowship and the Julie Payette NSERC Research Scholarship.

References

1. Flegel KM, Shipley MJ, Rose G. Risk of stroke in non-rheumatic atrial fibrillation. Lancet. 1987; 1:526–529. [PubMed: 2881082]
2. Cha YM, Redfield MM, Shen WK, Gersh BJ. Atrial fibrillation and ventricular dysfunction: a vicious electromechanical cycle. Circulation. 2004; 109:2839–2843. [PubMed: 15197156]
3. Reddy VY, Schmidt EJ, Holmvang G, Fung M. Arrhythmia recurrence after atrial fibrillation ablation: Can magnetic resonance imaging identify gaps in atrial ablation lines? Journal of Cardiovascular Electrophysiology. 2007:434–437. [PubMed: 18179530]
4. Ecabert, O.; Peters, J.; Weese, J. Proc of SPIE Medical Imaging: Image Processing. 2006. Modeling shape variability for full heart segmentation in cardiac computed-tomography images.
5. Zheng Y, Barbu A, Georgescu B, Scheuering M, Comaniciu D. Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. IEEE Transactions on Medical Imaging. 2008; 27(11):1668–1681. [PubMed: 18955181]
6. Lorenz C, von Berg J. A comprehensive shape model of the heart. Medical Image Analysis. 2006; 10(4):657–670. [PubMed: 16709463]

7. John M, Rahn N. Automatic left atrium segmentation by cutting the blood pool at narrowings. Proc of MICCAI: Medical Image Computing and Computer-Assisted Intervention. 2005:798–805.
8. Karim, R.; Mohiaddin, R.; Rueckert, D. Proc of SPIE Medical Imaging: Visualization, Image-guided Procedures, and Modeling. 2008. Left atrium segmentation for atrial fibrillation ablation.
9. Karim, R.; Mohiaddin, R.; Rueckert, D. Proc of SPIE Medical Imaging: Visualization, Image-guided Procedures, and Modeling. 2009. Left atrium pulmonary veins: Segmentation and quantification for planning atrial fibrillation ablation.
10. Heckemann RA, Hajnal JV, Aljabar P, Rueckert D, Hammers A. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage. 2006; 33(1):115–126. [PubMed: 16860573]
11. Sabuncu MR, Yeo BT, Van Leemput K, Fischl B, Golland P. Nonparametric mixture models for supervised image parcellation. Proc of PMMIA 2009: MICCAI Workshop on Probabilistic Models for Medical Image Analysis. 2009:301–313.
12. Išgum I, Staring M, Rutten A, Prokop M, Viergever MA, van Ginneken B. Multi-atlas-based segmentation with local decision fusion – application to cardiac and aortic segmentation in CT scans. IEEE Transactions on Medical Imaging. 2009; 28(7):1000–1009. [PubMed: 19131298]
13. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage. 2009; 45(1):S61–S72. [PubMed: 19041946]
14. Dru F, Vercauteren T. An ITK implementation of the symmetric log-domain diffeomorphic demons algorithm. Insight Journal. 2009
15. Wells WM, Viola P, Atsumi H, Nakajima S, Kikinis R. Multi-modal volume registration by maximization of mutual information. Medical Image Analysis. 1996; 1(1):35–51. [PubMed: 9873920]
16. Thirion JP. Image matching as a diffusion process: an analogy with Maxwell's demons. Medical Image Analysis. 1998; 2(3):243–260. [PubMed: 9873902]
17. Arsigny V, Commowick O, Pennec X, Ayache N. A log-euclidean framework for statistics on diffeomorphisms. Proc of MICCAI: Medical Image Computing and Computer-Assisted Intervention. 2006:924–931.
18. Guimond A, Roche A, Ayache N, Meunier J. Three-dimensional multimodal brain warping using the demons algorithm and adaptive intensity corrections. IEEE Transactions on Medical Imaging. 2001; 20(1):58–69. [PubMed: 11293692]
19. Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945; 26(3): 297–302.
20. Rohlfing T, Brandt R, Menzel R, Maurer CR. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage. 2004; 21(4):1428–1442. [PubMed: 15050568]
21. Van Leemput K, Maes F, Vandermeulen D, Seutens P. Automated model-based tissue classification of MR images of the brain. IEEE Transactions on Medical Imaging. 1999; 18(10): 897–908. [PubMed: 10628949]
22. Sabuncu MR, Balci SK, Shenton ME, Golland P. Image-driven population analysis through mixture modeling. IEEE Transactions on Medical Imaging. 2009; 28(9):1473–1487. [PubMed: 19336293]

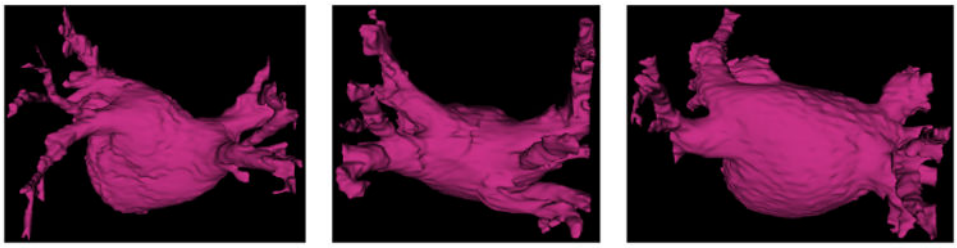

Fig 1.

Manual segmentations of the left atrium in three different subjects, illustrating the variability of the anatomy.

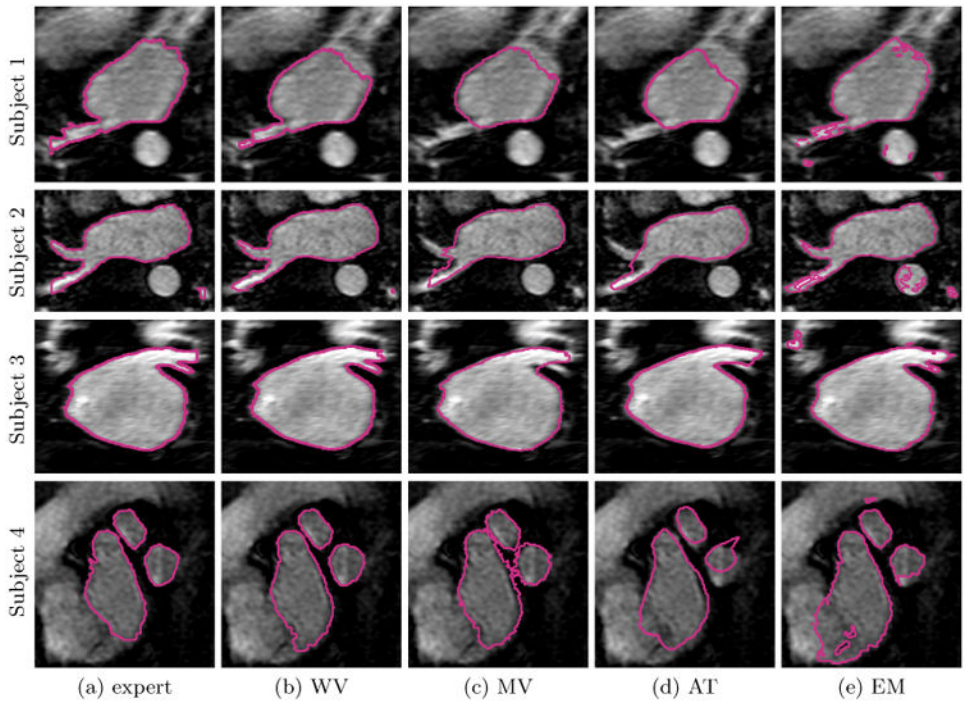

Fig 2.

Example segmentations of four different subjects: (a) expert manual segmentation, (b) weighted voting label fusion (WV), (c) majority voting label fusion (MV), (d) parametric atlas thresholding (AT) and (e) EM-segmentation using the parametric atlas as a spatial prior (EM).

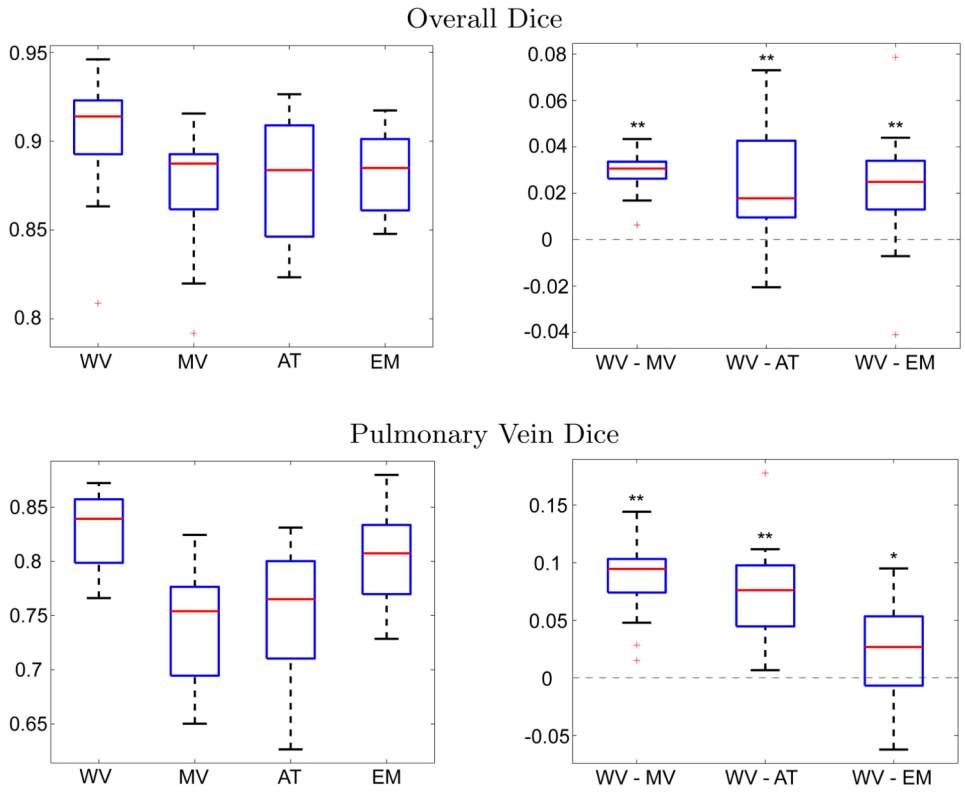

Fig 3.

Dice scores of results for weighted voting label fusion (WV), majority voting label fusion (MV), parametric atlas thresholding (AT) and atlas-based EM-segmentation (EM). For each box plot, the central red line indicates the median, the boxes extend to the 25th and 75th percentiles, and the whiskers extend to the most extreme values not considered outliers, which are plotted as red crosses. Stars indicate that the weighted label fusion method achieves significantly more accurate segmentation than the baseline method (single-sided paired t-test, *: $p < 0.05$, **: $p < 0.01$).

Table 1

Computation times for different methods.

Method	Registration	Segmentation	Total
WV	8 min × 15	5 min	125 min
MV	8 min × 15	0.5 min	120.5 min
AT	8 min	0.1 min	8.1 min
EM	8 min	15 min	23 min