Interactive Whole-Heart Segmentation in Congenital Heart Disease

Interactive Whole-Heart Segmentation in Congenital Heart Disease

Danielle F. Pace[1], Adrian V. Dalca[1], Tal Geva[2,3], Andrew J. Powell[2,3], Mehdi H. Moghari[2,3], and Polina Golland[1]

[1] Computer Science and Artificial Intelligence Lab, MIT, Cambridge, MA, USA
[2] Department of Cardiology, Boston Children's Hospital, Boston, MA, USA
[3] Department of Pediatrics, Harvard Medical School, Boston, MA, USA
dfpace@mit.edu

Abstract. We present an interactive algorithm to segment the heart chambers and epicardial surfaces, including the great vessel walls, in pediatric cardiac MRI of congenital heart disease. Accurate whole-heart segmentation is necessary to create patient-specific 3D heart models for surgical planning in the presence of complex heart defects. Anatomical variability due to congenital defects precludes fully automatic atlas-based segmentation. Our interactive segmentation method exploits expert segmentations of a small set of short-axis slice regions to automatically delineate the remaining volume using patch-based segmentation. We also investigate the potential of active learning to automatically solicit user input in areas where segmentation error is likely to be high. Validation is performed on four subjects with double outlet right ventricle, a severe congenital heart defect. We show that strategies asking the user to manually segment regions of interest within short-axis slices yield higher accuracy with less user input than those querying entire short-axis slices.

1 Introduction

Whole-heart segmentation in pediatric cardiac MRI has great potential to improve surgical planning in children with congenital heart defects by enabling creation of patient-specific 3D heart models. In particular, 3D-printed heart models promise to provide surgeons with an anatomically faithful, tactile experience [7]. Building such models requires delineating all of the cardiac structures in a patient's MRI, including the entire blood pool, epicardial surface and the great vessels. Clinically available tools often require 4-8 hours of user interaction to manually segment 100-200 slices covering the entire heart and the great vessels [4, 11], which precludes routine clinical use of 3D heart models.

Whole-heart segmentation is challenging even in normal subjects. Previously demonstrated methods typically employ atlas-based segmentation or fit deformable models to the image to be segmented [15]. The substantial changes in heart topology and high anatomical variability in congenital heart disease (CHD) render such model-based methods infeasible without an extremely large database of previously annotated scans. For example, in a subclass of CHD called

double outlet right ventricle (DORV), the aorta arises from the right ventricle (instead of the left) and a ventricular septal defect forms a hole in the septum between the two ventricles. Specialized segmentation algorithms have been developed for hearts with infarcts, left ventricular hypertropy, or pulmonary hypertension [1, 9]. These methods use probabilistic atlases or point distribution models built for normal subjects by finding some transformation between the abnormal heart and the normal model. These algorithms were designed for segmenting the ventricles, and it is unlikely that such methods will perform well for whole-heart segmentation in CHD. Automatic segmentation in cardiac MRI is also complicated by intensity inhomogeneities, low contrast, and thin heart walls near the atria, valves and great vessels which are barely visible. At the same time, high accuracy is required for creating useful heart models for surgical planning.

We therefore focus on developing efficient interactive segmentation methods. Interactive segmentation fits well into clinical workflows since physicians must validate any segmentation used for decision making and correct the errors that are inevitable in automatic segmentation. We present a patch-based [3, 6] interactive segmentation method that provides accurate whole-heart segmentation in CHD. The method uses a small set of manually labeled slices to segment the remaining volume, thus circumventing the challenges of anatomical variability.

Moreover, we examine active learning methods [8] to further reduce the number of interactions. At each step of an active learning session, the algorithm directs the user to manually label part of the data deemed most informative. These methods promise better accuracy with fewer user interactions compared to systems in which the user decides where to provide input. Most active learning methods for interactive medical image segmentation rely on uncertainty sampling with a batch selection query strategy [2, 5, 10, 12–14]. In uncertainty sampling, the active learner selects the voxels in which it is least confident. Confidence can be measured using image-based metrics [10, 14] or label probabilities [5, 13]. Ensemble methods assess the disagreement among votes [2], while SVM classifiers choose data based on distance to the margin [12]. Batch queries ask the user to label multiple voxels in each interaction step. A query can involve annotating sets of the most informative voxels [2, 5, 12], segmenting entire slice planes [10, 14] or deciding whether or not to include an entire hypothesized object [13].

Within our patch-based interactive segmentation framework for high-quality segmentation in CHD, we explore the potential benefits of active learning with batch queries based on uncertainty sampling. We show that methods that select entire slices for manual delineation fail to perform significantly better than a simple strategy based on a uniform distribution of the input slices. In contrast, active learning queries that asks the user to segment regions of interest (ROIs) within short-axis planes are more accurate with less user interaction.

2 Patch-based interactive segmentation

In this section we describe our interactive patch-based segmentation algorithm that incorporates user annotations. The method also provides a baseline for our

study of active learning strategies for cardiac MRI segmentation.

Given input image $I : \Omega_I \in \mathbb{R}^3 \to \mathbb{R}$, we seek a label map $L : \Omega_I \to \{l_b, l_m, l_k\}$ that parcellates image I into blood pool, myocardium and background. For the purpose of creating 3D heart models, the myocardium class includes the papillary muscles and the great vessel walls. We let $\mathcal{R}_I = \{r_i : \Omega_i \in \mathbb{R}^2 \to \{l_b, l_m, l_k\}\}$ be a set of manually labeled reference regions, where each subdomain $\Omega_i \in \Omega_I$ is defined on a short-axis plane. We focus on short-axis slices because clinicians are already accustomed to segmenting short-axis views for making cardiac function measurements such as ejection fraction. In the simplest case, a reference domain Ω_i is an entire short-axis slice plane, but it may represent a region of interest within a slice. In our baseline algorithm, the expert segments entire short-axis slices that are uniformly distributed in the MRI volume.

At each step of the interactive segmentation procedure, the user manually segments a provided short-axis region and a patch-based method is used to update the segmentation volume [3, 6]. The manually segmented regions provide patient-specific information on the heart's shape and the local appearance of the blood pool, myocardium and surrounding organs, which is exploited by the algorithm to infer labels in the remaining image.

For every target slice t to be segmented, a library of intensity patches with corresponding labels is constructed using the target's set of relevant reference regions $\mathcal{R}_t \in \mathcal{R}_I$. If every reference domain Ω_i is an entire short-axis slice, each remaining target slice in the volume is segmented using the closest reference slice above and below. If $\{\Omega_i\}$ contains smaller ROIs, each target slice is segmented using patches from the two closest entire reference slices plus all of the ROIs between them. An ROI segmentation "shadows" the region behind it so that only the physically closest information is used for each voxel in the target slice.

To segment patch $p_t(x^i)$ centered at voxel position x^i in target slice t, we find the k most similar patches in $\mathcal{R}_t$. We use $x = [x_1, x_2, x_3]$ to denote the three coordinates of position x, where x_1 and x_2 are in-plane (short-axis) coordinates and x_3 is the out-of-plane coordinate. Given a patch $p_r(x^j)$ centered at voxel position x^j in a reference $r \in \mathcal{R}_t$ with domain Ω_r, the distance between patch $p_t(x^i)$ and patch $p_r(x^j)$ depends on the patch intensities, gradients and positions:

$$\begin{aligned} d(p_t(x^i), p_r(x^j)) = \alpha \|p_t(x^i) - p_r(x^j)\|^2 + \beta \|\nabla p_t(x^i) - \nabla p_r(x^j)\|^2 \\ + \delta_{\parallel}(x^i, \Omega_r)\Big[(x_1^i - x_1^j)^2 + (x_2^i - x_2^j)^2\Big] + \delta_{\perp}(x^i, \mathcal{R}_t)(x_3^i - x_3^j)^2. \end{aligned} \tag{1}$$

Here, α and β are weights on the relative importance of the intensity and gradient terms, respectively, and $\delta_{\parallel}(x^i, \Omega_r)$ and $\delta_{\perp}(x^i, \mathcal{R}_t)$ are spatially-varying weights on the in-plane and out-of-plane components of the positions, respectively.

The position weighting is higher when a target patch is close to one of its references, to encourage matching to the close reference since it likely contains the same structures. To this end, the out-of-plane position weight $\delta_{\perp}(x^i, \mathcal{R}_t)$ is defined as the distance from x^i to the closest point within any of its references:

$$\begin{aligned} \delta_{\perp}(x^i, \mathcal{R}_t) &= \gamma_1 \exp^{-\gamma_2 \cdot D_{\perp}(x^i, \mathcal{R}_t)} + \gamma_3, \\ D_{\perp}(x^i, \mathcal{R}_t) &= \min_{r \in \mathcal{R}_t} \min_{x^j \in \Omega_r} \|x^i - x^j\|. \end{aligned} \tag{2}$$

Second, we allow a larger effective search area within distant reference slices, which enables matching of structures that might change shape substantially across neighboring slices. In contrast, if the reference slice is very close then the matching structure is probably at a similar position. The in-plane position weight $\delta_{\|}(x^i, \Omega_r)$ is therefore different for each reference r, and is defined as the distance from x^i to the closest point in the reference domain Ω_r:

$$\begin{aligned} \delta_{\|}(x^i, \Omega_r) &= \lambda_1 \exp^{-\lambda_2 \cdot D_{\|}(x^i, \Omega_r)} + \lambda_3, \\ D_{\|}(x^i, \Omega_r) &= \min_{x^j \in \Omega_r} \|x^i - x^j\|. \end{aligned} \tag{3}$$

Given the k most similar reference patches, the labeling for the target patch is determined through majority voting. The reference patch segmentations contribute overlapping votes (i.e., they vote for all voxels in a patch), which smooths the segmentation. This variant removes the need for additional smoothness constraints that could potentially eliminate small walls in the heart and great vessels.

3 Empirical study: active learning for reference selection

Here we investigate batch query strategies for automatically choosing the subdomains $\{\Omega_i\}$ to be segmented by the expert. First, we compare interactive workflows that select entire short-axis slices versus those selecting smaller ROIs of a fixed size. To decouple the effect of the reference domain size from that of a specific uncertainty sampling method, we use a gold-standard manual segmentation to identify the next region to be segmented. At each step, the region with the highest segmentation error over its domain spanning $\pm h$ slices is selected for manual segmentation. We refer to this iterative setup as **oracle** uncertainty estimation. We emphasize that our goal is to investigate the effect of the interaction strategy, as this approach is clearly infeasible for segmentation of novel images. In practice, uncertainty can be measured using metrics that locally estimate segmentation accuracy through the entropy of the patch vote distributions, alignment of label boundaries with image gradients, and intensity homogeneity within small regions with the same label [10, 14].

We compare the two active learning methods (entire slices vs. smaller ROIs) with several baseline approaches in which the user segments entire slices. First, we test against our baseline algorithm using **uniform** slice distribution. We also compare against **random** slice selection, a common baseline in active learning [10, 14]. These two slice selection schemes are not iterative. Finally, we implemented a strategy that exhaustively tries all possible new reference slices at each step and uses the gold-standard segmentation to add the slice that maximally reduces the segmentation error. This represents an iterative **optimal greedy** error reduction strategy with respect to maximizing improvement at each step.

4 Results

Data: Validation was performed using four pediatric cardiac MRI images from patients with DORV. A high resolution isotropic whole-heart image was acquired

for each patient as part of surgical planning. The scans were performed without contrast agents on a clinical 1.5T scanner (Philips Achieva), using a free-breathing steady-state free precession (SSFP) pulse sequence with ECG and respiratory nagivator gating (TR = 3.4ms/TE =1.7ms/$\alpha = 60°$). All images were cropped to a tight region around the heart, rotated into a short-axis orientation and smoothed slightly using anisotropic diffusion. The final image size was around $120 \times 150 \times 200$ and varied across patients. Voxel size was around 0.9mm^3 and also varied slightly across subjects. A gold-standard manual segmentation of the entire image volume was created in approximately eight hours per scan, and is used to simulate user input in interactive segmentation.

Parameter selection: For interactive patch-based segmentation, we use 5×5 patches and retrieve $k = 10$ nearest neighbor patches for each target patch. The nearest neighbor search was limited to a 101×101 in-plane bounding box. The weights governing the relative influence of the terms used to calculate patch similarity in Eqs. (1)-(3) were determined empirically using the four datasets: for all patients, $\alpha = 1, \beta = 1, \gamma = [8.49, 0.02, 0.0375]$ and $\lambda = [1.62, 0.2, 1.25]$. In both oracle strategies, we evaluate errors on regions spanning $\pm h = 2$ slices. In the oracle ROI strategy, we use ROIs of size 39×39. Finally, the oracle and optimal greedy strategies are initialized using three uniformly distributed slices.

Findings: Fig. 1 shows example heart models and segmentations created

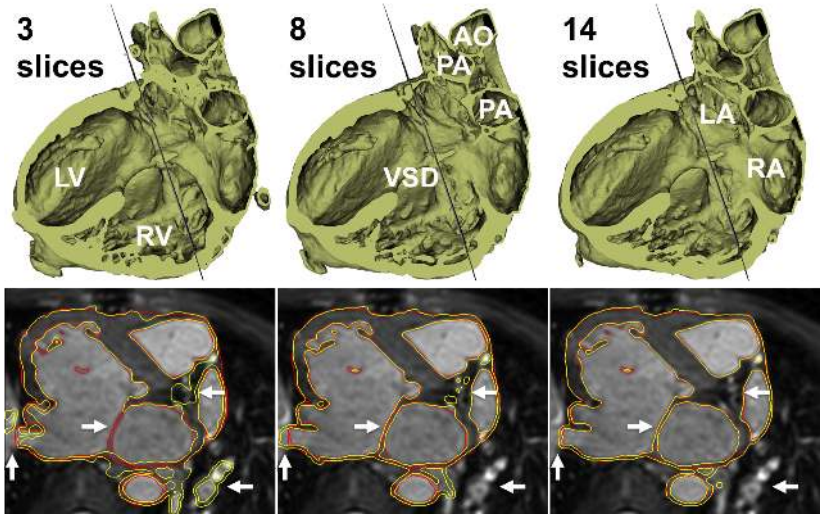

Fig. 1: Example 3D heart models (cut in half to visualize the interior) and segmentation results for a subject with DORV, from patch-based interactive segmentation instantiated with 3, 8 and 14 uniformly distributed reference slices, respectively. Interactive segmentation (yellow) and gold-standard segmentation (red) are shown. LV/RV/LA/RA = left/right ventricle/atrium; VSD = ventricular septal defect; AO = aorta; PA = pulmonary artery. Arrows indicate segmentation errors that are corrected by including more reference slices.

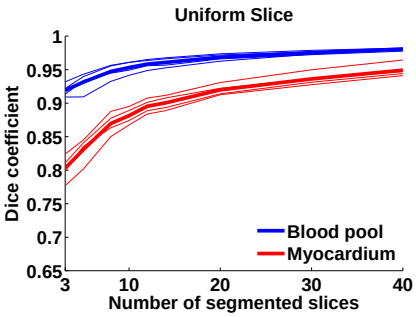

Fig. 2: Accuracy of patch-based interactive segmentation as a function of the number of uniformly distributed reference slices. Thin lines represent each subject and the thick line corresponds to the mean.

using interactive patch-based segmentation instantiated with 3, 8 and 14 uniformly distributed reference slices, respectively. The accuracy improvement when more input is provided is clear. A high quality model can be created using only 14 reference short-axis segmentations (out of ∼200 slices). Even the model instantiated with 3 reference slices shows roughly correct global structure. Fig. 2 shows segmentation accuracy for uniform slice selection measured as Dice volume overlap for the blood pool and myocardium. The patch-based segmentation method achieves good accuracy using relatively few segmented slices, especially considering the difficulty of whole-heart segmentation in CHD.

Experimental results for active learning are reported in Fig. 3. We observed that the slices selected by active learning are sampled more densely near the base of the heart and less so near the apex, which correlates with the relative difficulty of segmenting these areas. The active learning strategy that selects entire short-axis slices does not achieve a meaningful improvement compared to uniform slice selection. Even the optimal greedy strategy only shows a modest improvement compared to uniform distribution, indicating that there is not much scope for improvement for active learning methods that iteratively choose entire short-axis slices. All methods substantially outperform random slice selection. This suggests that random selection is not necessarily the most appropriate baseline when evaluating new active learning methods, although it is widely used.

Oracle ROI active learning shows substantial improvement (∼5 Dice points for the myocardium) using less user input. Having the user label ROIs targets areas of concentrated errors, leading to more efficient interactive segmentation.

Manual delineation of 15 short-axis slices requires less than one hour of an expert's time, vs 4-8 hours for the entire volume. The runtime of our current implementation of patch-based segmentation takes roughly one hour per scan. The computation time associated with adding a new reference region is proportional to the number of affected target slices. Future work will focus on improving the algorithm runtime, e.g., through the use of approximate nearest neighbors.

Our experiments raised an interesting question of what should be used as a surrogate measure for the amount of user interaction. The results reported in Fig. 3 employ the ROI area as such a measure. However, several alternative metrics are plausible, depending on whether the user paints the labels or draws curves outlining the different tissue types. We also examined accuracy as a function of the number of edge pixels in the reference label maps. When evaluated this way, the differences between slice and ROI active learning disappear, and simply uniformly distributing the reference slices may be the best choice. In this case,

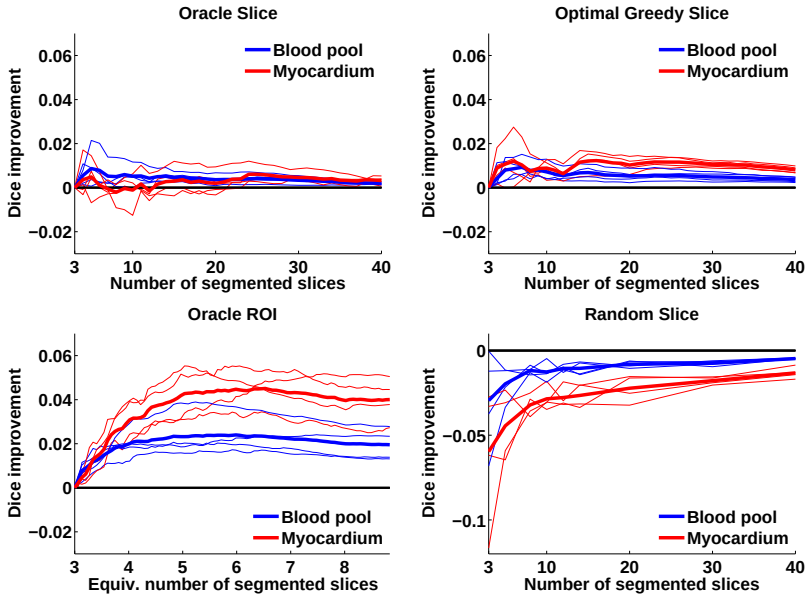

Fig. 3: Segmentation accuracy of alternative reference selection methods, reported as the improvement over uniform slice selection. Negative values indicate that uniform slice selection outperforms the method. For oracle ROI active learning, results are reported as a function of the cumulative area segmented by the user. Random slice selection scores are averages over five trials per subject.

even the optimal greedy strategy, which does not normalize by the effort required to segment each slice, can be worse than uniform slice selection. A user study evaluating the time required to manually segment slices versus ROIs is the best way to determine the most appropriate proxy for interaction time.

5 Conclusions

We presented an accurate interactive method for whole-heart segmentation in congenital heart disease. To the best of our knowledge, this is one of the first demonstrations towards clinically practical image segmentation to enable routine use of 3D heart models for surgical planning in CHD. We also show that active learning approaches in which the user annotates uncertain ROIs have potential to further reduce segmentation time. Future work includes active learning for arbitrarily oriented slices or ROIs. More sophisticated active learning methods that probabilistically model the expected error reduction given a candidate ROI could also yield improved segmentation results with minimal user effort.
Acknowledgements: NSERC CGS-D, the Wistron Corporation, NIH NIBIB NAMIC U54-EB005149, the Boston Children's Hospital Translational Research Program Fellowship and Office of Faculty Development, and Harvard Catalyst.

References

1. Albà, X., Lekadir, K., Hoogendoorn, C., Pereanez, M., Swift, A.J., et al.: Reusability of statistical shape models for the segmentation of severely abnormal hearts. In: Camara, O., Mansi, T., Pop, M., Rhode, K., Sermesant, M., Young, A. (eds.) Statistical Atlases and Computational Models of the Heart, LNCS, vol. 8896, pp. 257–264. Springer International Publishing (2015)
2. Chyzhyk, D., Dacosta-Aguayo, R., Matar, M., Graa, M.: An active learning approach for stroke lesion segmentation on multimodal MRI data. Neurocomputing 150(A), 26–36 (2015)
3. Coupé, P., Manjn, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. NeuroImage 54(2), 940–954 (2011)
4. Jacobs, S., Grunert, R., Mohr, F.W., Falk, V.: 3D-Imaging of cardiac structures using 3D heart models for planning in heart surgery: A preliminary study. Interact. Cardiovasc. Thorac. Surg. 7(1), 6–9 (2008)
5. Mahapatra, D., Schffler, P.J., Tielbeek, J.A.W., Vos, F.M., Buhmann, J.M.: Semi-supervised and active learning for automatic segmentation of Crohns Disease. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, LNCS, vol. 8150, pp. 214–221. Springer Berlin Heidelberg (2013)
6. Rousseau, F., Habas, P., Studholme, C.: A supervised patch-based approach for human brain labeling. IEEE Trans. Med. Imaging 30(10), 1852–1862 (2011)
7. Schmauss, D., Haeberle, S., Hagl, C., Sodian, R.: Three-dimensional printing in cardiac surgery and interventional cardiology: A single-centre experience. Eur. J. Cardio.-Thorac. (2014 [epub ahead of print])
8. Settles, B.: Active Learning. Morgan & Claypool Publishers, San Rafael (2012)
9. Shi, W., Zhuang, X., Wang, H., Duckett, S., Oregan, D., Edwards, P., Ourselin, S., Rueckert, D.: Automatic segmentation of different pathologies from cardiac cine MRI using registration and multiple component EM estimation. In: Metaxas, D., Axel, L. (eds.) FIMH 2011, LNCS, vol. 6666, pp. 163–170. Springer Berlin Heidelberg (2011)
10. Top, A., Hamarneh, G., Abugharbieh, R.: Active learning for interactive 3D image segmentation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, LNCS, vol. 6893, pp. 603–610. Springer Berlin Heidelberg (2011)
11. Valverde, I., Gomez, G., Gonzalez, A., Suarez-Mejias, C., Adsuar, A., Coserria, J.F., Uribe, S., Gomez-Cia, T., Hosseinpour, A.R.: Three-dimensional patient-specific cardiac model for surgical planning in Nikaidoh procedure. Cardiol. Young 25(4), 698–704 (2014)
12. Veeraraghavan, H., Miller, J.: Active learning guided interactions for consistent image segmentation with reduced user interactions. In: IEEE International Symposium on Biomedical Imaging. pp. 1645–1648. IEEE Press, New York (2011)
13. Wang, B., Liu, K.W., Prastawa, K.M., Irima, A., Vespa, P.M., et al.: 4D active cut: An interactive tool for pathological anatomy modeling. In: IEEE International Symposium on Biomedical Imaging. pp. 529–532. IEEE Press, New York (2014)
14. Yifrah, S., Zadicario, E., Ju, T., Cohen-Or, D.: An algorithm for suggesting delineation planes for interactive segmentation. In: IEEE International Symposium on Biomedical Imaging. pp. 361–364. IEEE Press, New York (2014)
15. Zhuang, X.: Challenges and methodologies of fully automatic whole heart segmentation: A review. J. Healthc. Eng. 4(3), 371–408 (2013)