Deep Learning for End-to-End Atrial Fibrillation Recurrence Estimation

Riddhish Bhalodia[1], Anupama Goparaju[1], Tim Sodergren[1], Alan Morris[2], Evgueni Kholmovski[2,3],
Nassir Marrouche[2], Joshua Cates[2], Ross Whitaker[1], Shireen Elhabian[1]

[1] Scientific Computing and Imaging Institute, School of Computing,
University of Utah, Salt Lake City, Utah, USA
[2] Comprehensive Arrhythmia Research and Management Center, Division of Cardiovascular
Medicine, School of Medicine, University of Utah, Salt Lake City, Utah, USA
[3] Department of Radiology and Imaging Sciences, School of Medicine,
University of Utah, Salt Lake City, Utah, USA

Abstract

Left atrium shape has been shown to be an independent predictor of recurrence after atrial fibrillation (AF) ablation. Shape-based representation is imperative to such an estimation process, where correspondence-based representation offers the most flexibility and ease-of-computation for population-level shape statistics. Nonetheless, population-level shape representations in the form of image segmentation and correspondence models derived from cardiac MRI require significant human resources with sufficient anatomy-specific expertise. In this paper, we propose a machine learning approach that uses deep networks to estimate AF recurrence by predicting shape descriptors directly from MRI images, with NO image pre-processing involved. We also propose a novel data augmentation scheme to effectively train a deep network in a limited training data setting. We compare this new method of estimating shape descriptors from images with the state-of-the-art correspondence-based shape modeling that requires image segmentation and correspondence optimization. Results show that the proposed method and the current state-of-the-art produce statistically similar outcomes on AF recurrence, eliminating the need for expensive pre-processing pipelines and associated human labor.

1. Introduction

Left atrium (LA) shape has been shown to be associated with atrial fibrillation (AF) [1, 2]. In the past three decades, catheter ablation has evolved to a safe, effective, and clinically acceptable therapeutic option in patients with AF, and is recognized as the first-line therapy in patients with symptomatic paroxysmal and persistent AF [3], but with considerable post-ablation recurrence leading to high rates of redo ablations that reach to 40% [3,4]. Recent studies demonstrate the efficacy of using LA shape as

a predictor for AF recurrence after catheter ablation [5–7]. However, such studies rely heavily on patient-specific LA shape representation that entails a time-consuming, expert-driven, expensive, irreproducible, and error-prone workflow of segmenting patient's LA from cardiac MRI volumes followed by a processing pipeline of shape registration and dense placement of corresponding landmarks, which requires significant amount of human resources and anatomy-specific expertise.

Landmarks are the most popular choice as a lightweight and general shape representation suitable for statistical analysis and visual communication of the results [8]. Landmarks should be defined consistently within a population to refer to the same, i.e., *corresponding*, anatomical position on every shape instance, resulting in correspondence-based models —aka point distribution models (PDMs). Such models and derived population-level statistics play an important role in ablation guidance, fibrosis quantification, and biophysical modeling in AF patients [1, 3, 9]. Furthermore, these models are useful in deriving shape-driven LA segmentation methods to account for misleading imaging information [10]. Nonetheless, PDM generation state-of-the-art methods (e.g., [11–13]) rely on manual segmentations and a subsequent heavy pre-processing of shape registration and correspondence optimization (and associated algorithmic parameters tuning).

Recently, deep networks have found many uses in medical image analysis. Their natural ability to learn complex functions makes them ideal for more complex computer-aided tasks such as automatic segmentation [14] and landmark detection [15]. In this paper, we propose an automated approach that rely on deep networks to generate a patient-specific landmark-based anatomy representation directly from 3D cardiac MRI, hence, negating any need for manual pre-processing and segmentation. However, a deep network cannot be viably trained with lim-

ited training samples —a typical situation in this application and in many similar medical imaging applications. To mitigate this problem, we propose a novel data augmentation scheme that generates more statistically feasible data, and hence enable training the deep network while reducing the associated risk of overfitting. We compare the efficacy of the proposed method with the state-of-the-art correspondence-based shape modeling, which requires segmentation and correspondence optimization, in terms of correspondence-level error and difference in AF recurrence prediction. Results show statistical equivalence. As the proposed method is based on learning shape descriptors from images, hence, the method has been leveraged for automatic LA segmentation with promising results.

2. Methods

Figure 2 shows the workflow of the proposed method, as compared to the standard workflow, including model training and the usage of the trained network for new images.

Data augmentation: We use a population of 207 MRI scans of LA of AF patients (*original data*), and these are not enough to train a deep network with, given the high dimensionality of such images (and their respective LA shapes). We propose the following data augmentation scheme. We first compute the PDM of the original data using *ShapeWorks* [11] software. Then, we perform principal components analysis (PCA) on the original data, which reduces the dimension of each sample from thousands of 3D points to 15 PCA loadings (dominant modes that capture 95% of shape variability). This PCA subspace represents a multivariate Gaussian that describes the shape variations in the given population. Further, this subspace parameterizes the data distribution, enabling generating thousands *shape* samples that respect the population-level statistics. To obtain the corresponding MRI *image* of a generated shape sample, we find the nearest neighbor to the generated sample from the original data and use the correspondences as landmarks, to perform thin plate spline (TPS) warp [16]. The LA structure exhibits a natural clustering in the *shape space* due to the variability of pulmonary veins arrangements [17]. To account for such variability, we model LA shapes as a multi-modal Gaussian distribution in the PCA subspace, with 3 components yielding the best Bayesian information criterion (BIC). The method is pictorially depicted in Figure 1. In our experiments, we used 175 samples out of 207 to use for data augmentation, and the rest is set aside for deep network testing (as *unseen* samples).

Network architecture and training: To predict/estimate the shape descriptor in the form of 15 PCA loadings directly from the 3D MRI scans, we use a deep convolution neural network (CNN) [18] that uses the L2 loss function for optimization. We train the network for 240 epochs using Adagrad optimizer in Tensorflow. We use 5000 train-
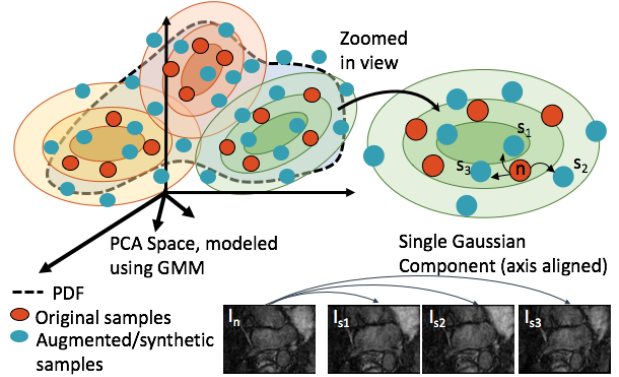

Figure 1. Pictorial representation of the data augmentation scheme, we use GMM to find pdf in PCA subspace and then use each Gaussian component to generate new samples.

ing samples generated from the three mixture components as described in the data augmentation scheme.

AF prediction: AF recurrence has been hypothesized to be dependent on shape descriptors such PCA loadings [1]. We hypothesize that the PCA loadings predicted using the proposed deep network are equivalent to the ones from the state-of-the-art pipeline with regards to AF recurrence prediction. Hence, we train a multi-layer perceptron (MLP) using the PCA loadings of the original 175 data derived from *ShapeWorks* correspondences, and their AF recurrence data (binary variable). The trained MLP predicts the AF recurrence probability that can be compared using the PCA loadings computed using the deep network and through the standard state-of-the-art PDM model.
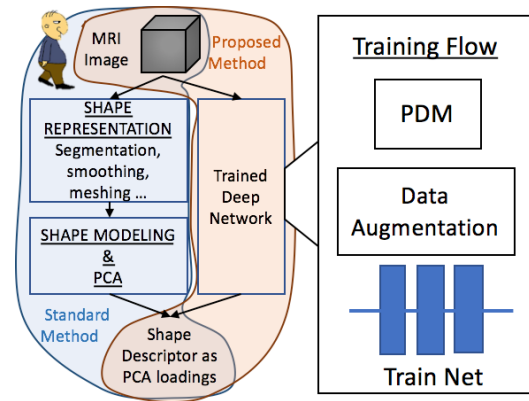

Figure 2. A pictorial representation of the existing standard shape modeling pipeline and the proposed pipeline. The blue-shaded pipeline requires manual pre-processing and segmentation for every new data to be processed. In contrast, once the network is trained, the proposed method (shaded in orange) is fully automatic.

3. Results

We use the PDM on 207 MRI's as our base data, of which 175 are used for data augmentation and hence, in-

fluenced the network training. The remaining 32 are called the *cross-validation* data that is used for several validation scenarios as detailed in the following.

PCA loadings and correspondence estimation: To serve as an automated surrogate to the standard work-flow, the proposed method needs to produce valid patient-specific correspondence model. We validate this assessment using three different aspects. The network produces the output in form of PCA loadings, hence, we first compare these estimated PCA loadings with those given by the PDM, i.e., the ground truth. As these are multivariate data, we use Hotelling T^2 statistic to identify if the differences were statistically significant. We obtain the statistic to be 11.1 with 78% confidence (97% for 175 data used for training and 74% for the data used for cross-validation).

To evaluate the efficacy of the network to provide loadings that correctly reconstruct the patient-specific correspondence model, we reconstruct the correspondences produced through the network and compare it with the ground truth correspondences. Boxplots in Figure 3 show the per-point per-shape error in millimeters (*mm*) for all training, validation, testing and cross-validation data (called "unseen" in the figure). The errors for all but the unseen data are less than the voxel spacing of the images (2 *mm*) in average, and thereby achieving sub-voxel accuracy.

To evaluate the use of the proposed method as a possible automatic segmentation method or as an approximate driver towards LA segmentation, we use the meshes generated from the reconstructed correspondences, and we compare them with the original ground truth segmentations. For the validation metric, we use surface-to-surface distance between the reconstructed meshes and the original ones. In Figure 4, we observe that the largest surface-to-surface distances are concentrated in the pulmonary veins, but the piece-wise continuous regions of the LA anatomy are reconstruction with relatively minimal error. We also observe the degrading image quality from top to bottom in Figure 4 and the increase in the associated error. It is worth emphasizing that no pre-processing is performed when feeding these images to train the deep network.

AF recurrence prediction: The deep network should estimate shape parameters that subsequently can predict accurate outcome analysis (AF recurrence prediction in our case). To assess this aspect, we test for equivalence between the AF recurrence probabilities predicted by the PCA loadings from the PDM and that from the deep network using the two one-sided test (TOST) [19]. We find the recurrence predicted by deep network and ground truth PCA loadings to be equivalent with a confidence of 90% with the mean difference bounds of ± 0.06. We also report the error between both predictions in a boxplot as shown in Figure 3.
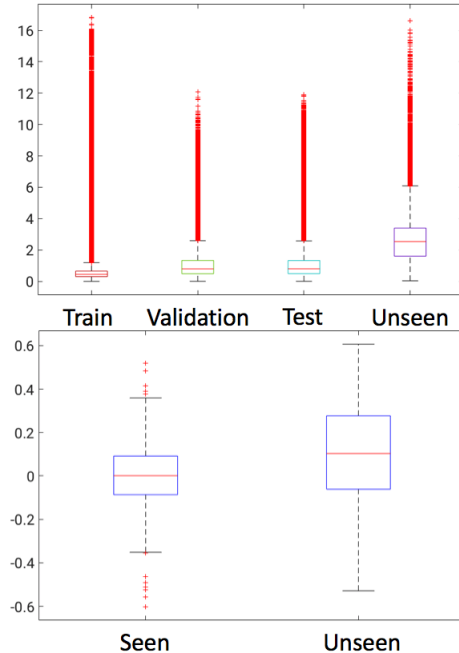

Figure 3. (Top) Box plots of per-point per-shape error in *millimeters* between the reconstructed correspondences produced through deep network, and, the ground truth correspondences. (Bottom) Difference between the recurrence probabilities from ground truth PCA loadings and the network produced PCA loadings, "seen" data is the original data used for PDM generation and "unseen" is again the cross-validation data.

4. Conclusion

We proposed a pre-processing-free method for LA shape analysis and demonstrated its usability in predicting AF recurrence. We compared this method with the standard state-of-the-art shape analysis workflow, which requires significant human supervision and anatomy-specific expertise, and found the results to be statistically comparable. The proposed method directly uses images and it is also shown to be a viable option for automatic LA segmentation. Errors of the method can be attributed to the huge variability in image intensities arising from different scanners and acquisition protocols. We believe that using an improved data augmentation method that takes into account the intensity variability and shape statistics of the data will improve the network generalization.

References

[1] NF M, D W, G H, et al. Association of atrial tissue fibrosis identified by delayed enhancement mri and atrial fibrillation catheter ablation: The decaaf study. JAMA 2014; 311(5):498–506.
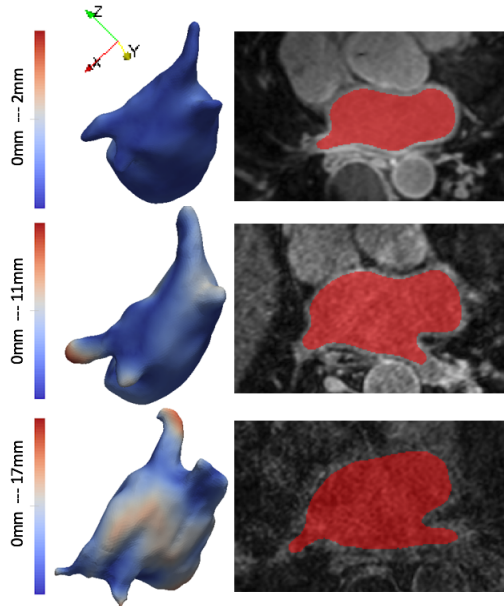
Figure 4.   This figure shows three different scans from the cross-validation data. Each row has the image and it's ground truth segmentation, and it also has a mesh surface with a surface to surface error heatmap (between ground truth segmentations and the reconstructed meshes using the network) overlay. From top to bottom we see images of degrading quality as well as the drastic shape differences in the LA surface meshes.

[2]   Bisbal F, Guiu E, Calvo N, Marin D, Berruezo A, Arbelo E, Ortiz-Prez J, Caralt TM, Tolosana JM, Borrs R, Sitges M, Brugada J, MONT L. Left atrial sphericity: A new method to assess atrial remodeling. impact on the outcome of atrial fibrillation ablation. Journal of Cardiovascular Electrophysiology 2013;24(7):752–759.

[3]   Calkins H, Hindricks G, Cappato R, Kim YH, Saad EB, Aguinaga L, Akar JG, Badhwar V, Brugada J, Camm J, et al. 2017 hrs/ehra/ecas/aphrs/solaece expert consensus statement on catheter and surgical ablation of atrial fibrillation. Heart Rhythm 2017;14(10):e275–e444.

[4]   Darby AE. Recurrent atrial fibrillation after catheter ablation: considerations for repeat ablation and strategies to optimize success. Journal of atrial fibrillation 2016;9(1).

[5]   Bieging ET, Morris A, Wilson BD, McGann CJ, Marrouche NF, Cates J. Left atrial shape predicts recurrence after atrial fibrillation catheter ablation. Journal of cardiovascular electrophysiology 2018;.

[6]   Cates J, Bieging E, Morris A, Gardner G, Akoum N, Kholmovski E, Marrouche N, McGann C, MacLeod RS. Computational shape models characterize shape change of the left atrium in atrial fibrillation. Clinical Medicine Insights Cardiology 2014;8:CMC–S15710.

[7]   Gardner G, Morris A, Higuchi K, MacLeod R, Cates J. A point-correspondence approach to describing the distribution of image features on anatomical surfaces, with application to atrial fibrillation. In 2013 IEEE 10th International Symposium on Biomedical Imaging. ISSN 1945-7928, April 2013; 226–229.

[8]   Sarkalkan N, Weinans H, Zadpoor AA. Statistical shape and appearance models of bones. Bone 2014;60:129–140.

[9]   Krueger MW, Dorn A, Keller DU, Holmqvist F, Carlson J, Platonov PG, Rhode KS, Razavi R, Seemann G, Dössel O. In-silico modeling of atrial repolarization in normal and atrial fibrillation remodeled state. Medical biological engineering computing 2013;51(10):1105–1119.

[10]   Tobon-Gomez C, Geers AJ, Peters J, Weese J, Pinto K, Karim R, Ammar M, Daoudi A, Margeta J, Sandoval Z, et al. Benchmark for algorithms segmenting the left atrium from 3d ct and mri datasets. IEEE transactions on medical imaging 2015;34(7):1460–1473.

[11]   Cates J, Elhabian S, Whitaker R. Shapeworks: Particle-based shape correspondence and visualization software. In Statistical Shape and Deformation Analysis. Elsevier, 2017; 257–298.

[12]   Durrleman S, Prastawa M, Charon N, Korenberg JR, Joshi S, Gerig G, Trouvé A. Morphometry of anatomical shape complexes with dense deformations and sparse parameters. NeuroImage 2014;101:35–49.

[13]   Styner M, Oguz I, Xu S, Brechbuehler C, Pantazis D, Levitt J, Shenton M, Gerig G. Framework for the statistical shape analysis of brain structures using spharm-pdm 07 2006;.

[14]   Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. CoRR 2015;abs/1505.04597. URL http://arxiv.org/abs/1505.04597.

[15]   Zheng Y, Liu D, Georgescu B, Nguyen H, Comaniciu D. 3d deep learning for efficient and robust landmark detection in volumetric data. In MICCAI 2015. Springer International Publishing. ISBN 978-3-319-24553-9, 2015; 565–572.

[16]   Bookstein FL. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on PAMI 1989;11(6):567–585.

[17]   Sohns C, Vollmann D, Luethje L, Dorenkamp M, Seegers J, Schmitto JD, Zabel M, Obenauer S. Mdct in the diagnostic algorithm in patients with symptomatic atrial fibrillation. World journal of radiology 2011;3(2):41.

[18]   Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE Nov 1998;86(11):2278–2324. ISSN 0018-9219.

[19]   Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. Journal of Pharmacokinetics and Biopharmaceutics 1987;15(6):657–680.

Address for correspondence:

Riddhish Bhalodia
72 Central Campus Drive, 3rd floor desk, Salt Lake City, Utah-84112, USA
riddhishb@sci.utah.edu