White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/160603/

Version: Accepted Version

Proceedings Paper:
Attar, R, Pereañez, M, Bowles, C et al. (2019) 3D Cardiac Shape Prediction with Deep Neural Networks: Simultaneous Use of Images and Patient Metadata. In: Lecture Notes in Computer Science. MICCAI 2019: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, 13-17 Oct 2019, Shenzhen, China. Springer Verlag, pp. 586-594. ISBN: 9783030322441. ISSN: 0302-9743. EISSN: 1611-3349.

https://doi.org/10.1007/978-3-030-32245-8_65

© Springer Nature Switzerland AG 2019. This is an author produced version of a conference paper published in Lecture Notes in Computer Science. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

3D Cardiac Shape Prediction with Deep Neural Networks: Simultaneous Use of Images and Patient Metadata

No Author Given

No Institute Given

Abstract. Large prospective epidemiological studies acquire cardiovascular magnetic resonance (CMR) images for pre-symptomatic populations, and follow these over time. To support this approach, fully automatic large-scale 3D analysis is essential. In this work, we propose a novel deep neural network using both CMR images and patient metadata to directly predict cardiac shape parameters. The proposed method uses the promising ability of statistical shape models to simplify shape complexity and variability together with the advantages of convolutional neural networks for the extraction of solid visual features. To the best of our knowledge, this is the first work that uses such an approach for 3D cardiac shape prediction. We validated our proposed CMR analytics method against a reference cohort containing 500 3D shapes of the cardiac ventricles. Our results show broadly significant agreement with the reference shapes in terms of the estimated volume of the cardiac ventricles, myocardial mass, 3D Dice, and mean and Hausdorff distance.

1 Introduction

Cardiovascular disease (CVD) is the most prevalent cause of death worldwide [1]. Early quantitative assessment of cardiac function and structure allow for proper preventive care, and early cardiovascular treatment. To support such an approach, analysis and interpretation of large-scale population-based cardiovascular magnetic resonance (CMR) imaging studies are of high importance in the medical image analysis community. This helps to identify patterns and trends across population groups, and accordingly, reveal insights into key risk factors before CVDs fully develop.

We believe that true 3D analysis is essential for the structural assessment of global and regional cardiac function. We propose a new approach that ensures the global coherence of the cardiac anatomy and naturally lends itself to further analysis in which full 3D anatomy is necessary; for example, in mechanical and flow simulations, or modelling the relationship between cardiac morphology

and patient information such as: socio-demographic, lifestyle and environmental, family history, genetic, and omics data.

Though fully automatic 3D segmentation is required for further analysis, the complexity of anatomical structures and their local intensity variation across a population cohort make it challenging. Statistical 3D shape model-based approaches such as [1] have been successfully used for automatically segmenting cardiac structures and generating associated function indexes. This is mainly attributed to the inclusion of prior knowledge of the cardiac shape into the segmentation method. These segmentation approaches typically use very simple features such as gradients on intensity profiles to fit a 3D model. This is an iterative process in which the goal is to minimise the Mahalanobis distance between an intensity profile sampled at a candidate position and its corresponding intensity appearance model by deforming the shape within its range of normal variation to match the image data. On the other hand, in the last decade, fully convolutional networks (FCN) have shown great potential in image-based pattern recognition in a variety of tasks, including cardiac segmentation. However, their output results are, by nature, 2D segmentations masks for every short axis (SAX) and long axis (LAX) CMR slices. Although these 2D masks are sometimes extended via a further step of non-rigid registration to a 3D atlas to produce a 3D cardiac shape [2], this is not efficient for learning topological shape information.

In this paper, we propose to exploit image features obtained using deep FCNs trained on both SAX and LAX views, along with the rich shape priors learned using statistical shape models, to jointly and simultaneously predict the parameters of 3D cardiac shapes, instead of a pixel-wise classification across each 2D slice. Another significant aspect of this work is the integration of patient metadata into the process of shape prediction using a Multilayer Perceptron (MLP). This information, which is currently ignored in cardiac segmentation or shape generation, has been shown in different clinical studies to have an impact on cardiac morphology and structure [3]. We hope this work inspires other researchers to exploit the priors offered by patient metadata in other applications for potentially more accurate and patient-specific models.

The contributions of this paper are three-fold: we propose 1) an innovative end-to-end deep neural network that directly predicts 3D shape parameters derived from a Principal Component Analysis (PCA) space; 2) a novel approach using two CMR image views and patient metadata simultaneously to predict cardiac shape; 3) a creative loss function defined in the domain of 3D shape parameters which weights each PCA mode of variation independently, prioritising the more significant modes and leading to more accurate shape prediction.

2 Methods

2.1 Reference 3D Shapes of Cardiac Ventricles

We generated a reference cohort of 3D shapes through the non-rigid registration of a 3D biventricular model to a set of 3D points obtained from manual

delineations using the Coherent Point Drift (CPD) method [4]. The 3D model is comprised of two structures; the Left Ventricle (LV) and the Right Ventricle (RV). The LV is a closed water-tight mesh comprising both endo and epicardial walls. The RV is an open mesh representing only the RV endocardium. The RV has two openings, the atrioventricular valve opening, and the pulmonary valve opening. Figure 1 shows a sample of manual 2D contours and its corresponding 3D shape obtained from the CPD method.

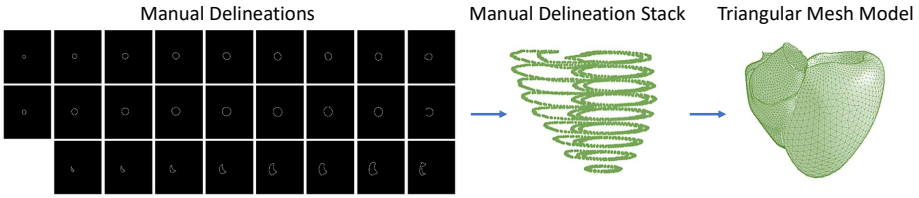

Figure 1. An example 3D shape of cardiac ventricles constructed from a stack of 2D manual contours on SAX view slices.

2.2 Point Distribution Mode (PDM)

The PDM encodes the mean and variance of the 3D cardiac shapes. The PDM is constructed during training using PCA on a set of generalised Procrustes-aligned shapes. Assume a training set of M shapes, each described by N points in $\mathcal{R}^3$, i.e., $\mathbf{x}_j^i = (\mathbf{x}_j^i, \mathbf{y}_j^i, \mathbf{z}_j^i)$ with $i = 1, ..., M$ and $j = 1, ..., N$. Further, let $\mathbf{s}_i = (\mathbf{x}_1^i, \mathbf{y}_1^i, \mathbf{z}_1^i, ..., \mathbf{x}_N^i, \mathbf{y}_N^i, \mathbf{z}_N^i)^T$ be the i-th vector representing the i-th shape. Finally, let $\mathbf{S} = [\mathbf{s}^1, ..., \mathbf{s}^M]$ be the set of all training shapes in matrix form. The shape class mean and covariance of $\mathbf{S}$ is calculated as follows:

$$\bar{\mathbf{s}} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{s}_i \qquad \text{and} \qquad \mathbf{C} = \frac{1}{M-1}\sum_{i=1}^{M}(\mathbf{s}_i - \bar{\mathbf{s}})(\mathbf{s}_i - \bar{\mathbf{s}})^T \tag{1}$$

The shape covariance is represented in a low-dimensional PCA space. This provides l eigenvectors $\mathbf{\Phi} = [\varphi_1\varphi_2...\varphi_l]$, and corresponding eigenvalues $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, ..., \lambda_l)$ computed through the Singular Value Decomposition of the covariance matrix . Hence, assuming the shape class follows a multi-dimensional Gaussian probability distribution, any shape in the shape class can be approximated from the following linear generative model:

$$\mathbf{s} \approx \bar{\mathbf{s}} + \mathbf{\Phi}\mathbf{b} \tag{2}$$

where $\mathbf{b}$ are shape parameters restricted to $|\mathbf{b}_i| \leq \beta\sqrt{\lambda_i}$; we typically set $\beta = 3$ to capture 99.7% of shape variability. The shape parameters of $\mathbf{s}$ can then be estimated as follows:

$$\mathbf{b} = \mathbf{\Phi}_l^T(\mathbf{s} - \bar{\mathbf{s}}). \tag{3}$$

Here, the entries of $\mathbf{b}$ are the projection coefficients of mean-centred shapes $(\mathbf{s} - \bar{\mathbf{s}})$ along the columns of $\mathbf{\Phi}$.

2.3 Images and Metadata

Each CMR image volume was pre-processed as follows. Each 9-slice SAX stack was intensity normalised by saturating the top 0.2% of intensities and scaling between 0 and 1, and spatially normalised by aligning a fixed point, defined as the average point of intersection between the three LAX views and each SAX slice, and angle, defined as the angle of the 4-chamber LAX, to a standard location and angle respectively. A 64×64 px region of interest (ROI) was then sampled from each slice at a 2 mm isotropic resolution. The corresponding 4-chamber LAX views were similarly intensity normalised, with a 80×60 px ROI sampled at the same 2 mm resolution around the point of intersection between the three LAX views and the basel SAX slice, and zero-padded to 80×80 px. Table 1 shows the summary of the metadata available for every image volume including both continuous and categorical variables. All variables were scaled to the range [0, 1], including categorical variables (viz. $sex/alcohol \in (0, 1)$, $smoking \in (0, 0.5, 1)$).

Table 1. Summary of the subject metadata used in this study.

Type	Metadata	Range
Continues	Age (years)	61 ± 7
	Weight (kg)	76 ± 15
	Height (cm)	170 ± 9
	Body mass index (kg/m2)	27 ± 4
	Body surface area (m2)	1.8 ± 0.2
	Heart rate (bpm)	68 ± 11
	Diastolic blood pressure (mmHg)	79 ± 11
	Systolic blood pressure (mmHg)	139 ± 19
Categorical	Sex	male/female
	Smoking status	never/previous/current
	Alcohol consumed	yes/no

2.4 Network Architecture and Loss Function

Fig. 2 shows a diagram of the proposed method. The network has three inputs: SAX view images, LAX view image, and metadata. The output is the predicted shape parameters $b^P = \{b_i^P | i = 1, ..., k\}$. To train the proposed architecture, we introduce the following loss function for training:

$$\mathbf{E}(\theta) = \sum_{i=1}^{k} f(b_i^P(\theta),\ b_i^R)\ .\ w(i,k) \quad \text{where} \quad w(i,k) = \sqrt{\frac{k-i+1}{k}} \tag{4}$$

where k is the number of shape parameters, θ denotes the network parameters, $f(.)$ denotes the absolute error of the difference between the reference value (b_i^R) and the value ($b_i^P(\theta)$) predicted by the network. $w(.)$ denotes a weighting function depending on the importance of the $i-th$ mode of variation on shape prediction, i.e. it assigns a higher weight to first modes of variation in the shape's PCA space.

As illustrated in Fig. 3, the FCN used in this work has been adapted from the down-sampling path of a U-Net [5] architecture with an encoder depth of 2 for the LAX and 4 for the SAX images. The last layer of the FCNs have an extra convolutional layer with the kernel size of the current feature map dimensions

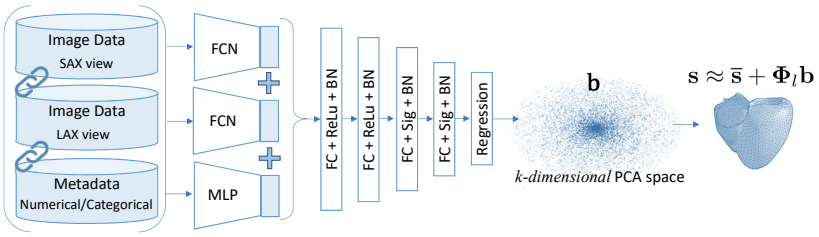

Figure 2. The proposed method extracts a high-level representation of the image from SAX and LAX views using two FCNs, and concatenates the image features together along with the output of an MLP network applied to the metadata. Four fully connected layers with ReLU or Sigmoid activation functions and batch normalisation then produce the k parameters in PCA space which describe the 3D shape of the cardiac ventricles.

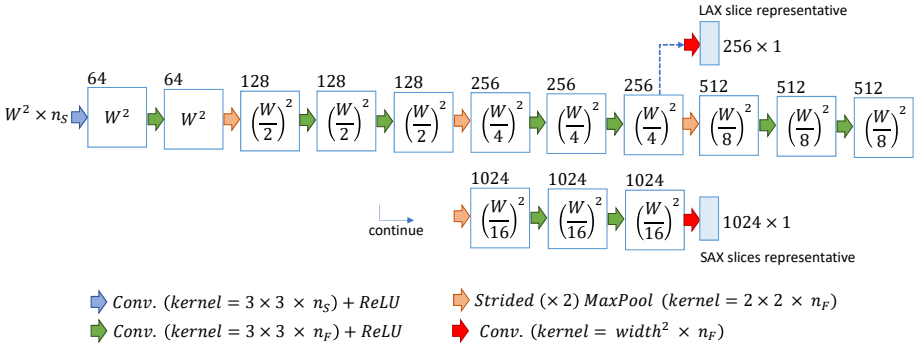

Figure 3. The architecture of the two FCNs used in this study to obtain a vector of features representing the image information derived from the LAX and SAX views. A separate network is used for each view, with the LAX and SAX networks containing 9 and 15 layers respectively. The FCNs are composed of convolutional layers, with ReLUs and max-pooling. W is the image width and height, n_S and n_F are the number of slices in each image volume and the number of activation maps respectively.

to produce a vector of features - 1024 for SAX and 256 for LAX. The MLP has 11 inputs (size of metadata feature vector), 3 hidden layers (with 16, 32, and 64 neurons), and an output layer (with 128 neurons). ReLU is used in hidden and output layers.

The outputs of the three sub-networks are concatenated to construct one feature vector (with the size of 1024+256+128=1408 neurons) that contains the behavioural, phenotypic, and demographic information derived from the metadata in addition to visual information from the imaging data. This information is fed into four fully connected layers, with ReLU (first two layers) and Sigmoid (last two layers) activation functions and batch normalisation, so that, by minimising $E(\theta)$ from Eq. 4, they produce the first k parameters in PCA space which describe the 3D shape of the cardiac ventricles. To capture 99.7% of shape variability in the training dataset we set $k = 28$ and regress only those parameters from randomly initialised weights.

3 Experiments and Results

3.1 Data and Annotations

We performed experiments on 3500 CMR image volumes from the UK Biobank (UKB) using both end-diastolic and end-systolic time points. In terms of population sample size, experimental setup, and quality control, the most reliable reference annotations of cardiovascular structure and function found in the literature are those reported by [6], in which CMR scans were manually delineated and analysed by a team of eight expert observers. These delineations were used to generate the reference 3D shapes, as explained in Sec. 2.1. The dataset was randomly split into a training (3000) and test set (500). The performance is reported on the test set with mean $\pm$ standard deviation.

3.2 Implementation and Training

The method was implemented using Python and Tensorflow. The network was trained using Adam for optimising the loss function (Eq. 4) with the learning rate of 0.001 and iteration number of 50,000 with a batch size of 10 subjects, all of which were determined empirically. There was no data augmentation. Training took $\sim$10 hours on Nvidia Tesla V100 GPUs hosted by Amazon Web Service. At test time, it took less than a second to predict the shape parameters.

3.3 Accuracy of Predicted Shapes

Fig. 4 shows some samples of ventricular shapes generated by our proposed method (in purple) overlaid with the corresponding reference shapes (in grey). It confirms that the network is capable of predicting accurate shape parameters to generate shapes very similar to the reference shapes obtained by manual delineations. To quantify the amount of similarity, we evaluated the performance of the proposed method by computing the Dice index ($\mathcal{D}$), and the mean ($\mathcal{M}$) and Hausdorff distance ($\mathcal{H}$) between reference and predicted shapes. Since this method outputs the parameters of a shape in the space, we first align the two shapes by removing their orientation and translation before computing the aforementioned metrics. $\mathcal{D}$ is between 0 and 1, with a higher $\mathcal{D}$ indicating a better match between the two shapes. $\mathcal{M}$ and $\mathcal{H}$ measure the mean and maximum distance, respectively, between the two surfaces, with a lower value indicating a better the agreement. Moreover, we report the effect of including the metadata in Table 2. As expected, the use of the metadata alongside the image information improves the network, leading to a more accurate prediction in all cardiac substructures. In addition to comparing against reference measurements, we also compare against one baseline method proposed by Attar et al. [1] in which the authors carried out 3D analysis of the UKB CMR images using a shape model-based approach where the model is fitted during an iterative process using traditional intensity profiles.

As shown in Table 2, $\mathcal{D}$ values show excellent agreement between reference and predicted shapes (≥ 0.90). $\mathcal{M}$ values are comparable to the in-plane pixel

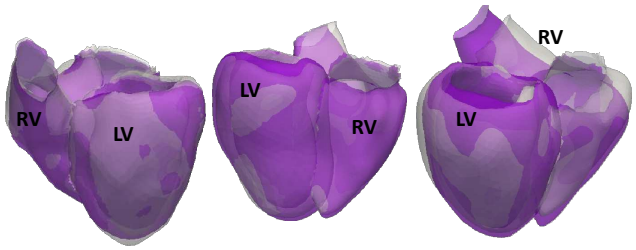

Figure 4. Three samples of the generated 3D shapes of LV and RV. The gray shape is the reference whereas the purple shape is the predicted.

Table 2. Comparison of shape prediction accuracy using only images (IMG) or images with metadata (IMG+MTDT) in terms of $\mathcal{D}$(%), $\mathcal{M}$(mm) and $\mathcal{H}$(mm) for LV endo-/epicardium and RV endocardium. **Bold** indicates best performing method.

	LV endocardium			**LV epicardium**			**RV endocardium**		
	[1]	IMG	IMG+MTDT	[1]	IMG	IMG+MTDT	[1]	IMG	IMG+MTDT
$\mathcal{D}$	**0.91±0.05**	0.82±0.09	0.90±0.04	0.92±0.05	0.83±0.08	**0.93±0.05**	0.88±0.08	0.79±0.09	**0.90±0.08**
$\mathcal{M}$	1.85±0.75	3.45±0.94	**1.81±0.70**	**1.80±0.62**	3.02±0.84	1.82±0.66	2.02±0.72	3.00±0.91	**2.00±0.70**
$\mathcal{H}$	3.76±1.52	8.78±1.96	**3.11±1.49**	**3.32±1.38**	7.69±1.77	3.55±1.49	8.32±3.12	12.11±5.21	**7.05±3.03**

spacing range of 1.8 mm to 2.3 mm found in the UKB. Although $\mathcal{H}$ is larger, it is still within an acceptable range when compared with the distance range seen in [1] or [2]. Note that the performance of the proposed method on RV is consistently better than the other approaches. Furthermore, we report the absolute and relative difference of the main cardiac function indexes (viz. LV and RV volume (ml) and myocardium mass (g)) derived from the predicted and the reference shapes in Table 3. The proposed method achieved significantly lower error in volume and mass estimation, with $p<0.001$ in paired t-tests.

Table 3. Comparison of the absolute and relative difference between the reference and predicted shapes. **Bold** indicates best performing method.

	Absolute difference			**Relative difference (%)**		
	[1]	IMG	IMG+MTDT	[1]	IMG	IMG+MTDT
LV Volume	7.51±5.42	9.80±6.33	**6.01±4.98**	9.50±8.80	10.31±9.45	**8.03±5.05**
LV Mass	8.42±5.22	10.11±8.14	**7.11±5.14**	9.10±8.01	12.03±9.22	**8.12±7.54**
RV Volume	10.59±7.16	12.62±10.14	**9.24±5.20**	11.36±8.11	14.55±9.89	**10.03±7.00**

Overall, the proposed method (IMG+MTDT) has superior accuracy to reference shapes than [1], while being on average ∼30 times faster during test time. This can be attributed to the combined use of image and patient metadata within a single network to directly predict shape parameters. The introduction of the metadata yielded a substantial positive impact on shape prediction with a ∼15% average improvement across all metrics. We believe that including this information provides the network with a variable prior by allowing it to learn the likely distributions of shape parameters across different populations.

4 Conclusion

In this study, we presented a fully automatic framework capable of producing 3D cardiac shapes via the simultaneous use of images and patient metadata. We validated our workflow on a reference cohort of 500 subjects for which ground truth shapes exist with promising results. In particular, we showed a significant positive impact from including the metadata. As future work, in addition to investigating the effect of other clinical variables on shape prediction, we would like to increase the robustness of our pipeline to locate the shape in the image space, and handle severe pathological morphology, variable image quality, and alternative modalities. We also plan to explore the use of patient metadata in other deep learning applications.

Acknowledgements To be added ...

References

1. R. Attar, M. Pereañez, A. Gooya, X. Albà, L. Zhang, S. K. Piechnik, S. Neubauer, S. E. Petersen, and A. F. Frangi, "High Throughput Computation of Reference Ranges of Biventricular Cardiac Function on the UK Biobank Population Cohort," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 114–121, Springer, 2018.
2. J. Duan, G. Bello, J. Schlemper, W. Bai, T. J. Dawes, C. Biffi, A. de Marvao, G. Doumou, D. P. O'Regan, and D. Rueckert, "Automatic 3D Bi-ventricular Segmentation of Cardiac Images by a Shape-refined Multi-task Deep Learning Approach," *IEEE transactions on medical imaging*, 2019.
3. K. Gilbert, W. Bai, C. Mauger, P. Medrano-Gracia, A. Suinesiaputra, A. M. Lee, M. M. Sanghvi, N. Aung, S. K. Piechnik, *et al.*, "Independent Left Ventricular Morphometric Atlases Show Consistent Relationships with Cardiovascular Risk Factors: A UK Biobank Study," *Scientific reports*, vol. 9, no. 1, p. 1130, 2019.
4. A. Myronenko and X. Song, "Point Set Registration: Coherent Point Drift," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.
5. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
6. S. E. Petersen, N. Aung, M. M. Sanghvi, F. Zemrak, K. Fung, J. M. Paiva, J. M. Francis, M. Y. Khanji, E. Lukaschuk, A. M. Lee, *et al.*, "Reference Ranges for Cardiac Structure and Function Using Cardiovascular Magnetic Resonance (CMR) in Caucasians from the UK Biobank Population Cohort," *Journal of Cardiovascular Magnetic Resonance*, vol. 19, no. 1, p. 18, 2017.