Generating Subpopulation-Specific Biventricular Anatomy Models Using Conditional Point Cloud Variational Autoencoders

Marcel Beetz[1(✉)], Abhirup Banerjee[1,2], and Vicente Grau[1]

[1] Institute of Biomedical Engineering, Department of Engineering Science,
University of Oxford, Oxford OX3 7DQ, UK
marcel.beetz@eng.ox.ac.uk
[2] Division of Cardiovascular Medicine, Radcliffe Department of Medicine,
University of Oxford, Oxford OX3 9DU, UK
abhirup.banerjee@cardiov.ox.ac.uk

Abstract. Generative statistical models have a wide variety of applications in modelling of cardiac anatomy and function, including disease diagnosis and prediction, personalized shape analysis, and generation of population cohorts for electrophysiological and mechanical computer simulations. In this work, we propose a novel geometric deep learning method based on the variational autoencoder (VAE) framework capable of accurately encoding, reconstructing, and synthesizing 3D surface models of the biventricular anatomy. Our non-linear approach works directly with memory-efficient point clouds and is able to process multiple substructures of the cardiac anatomy at the same time in a multi-class setting. Furthermore, we introduce subpopulation-specific characteristics as additional conditional inputs to allow the generation of new personalized anatomies. Our method achieves high reconstruction quality on a dataset derived from the UK Biobank study with average Chamfer distances between reconstructed and gold standard point clouds below the underlying image pixel resolution, for all anatomical substructures and combinations of conditional inputs. We investigate our method's generative capabilities and show that it is able to synthesize virtual populations of realistic hearts with volumetric measurements in line with established clinical precedent. We also analyse the effects of variations in the latent space of the autoencoder on the generated anatomies and find interpretable changes in cardiac shapes and sizes.

Keywords: Cardiac anatomy synthesis · Point cloud generation · Beta-VAE · Cardiac anatomy reconstruction · Conditional generative models · Geometric deep learning · Cardiac MRI

1 Introduction

The human heart exhibits considerable inter-person variability both in terms of its shape and function, which significantly impacts the effectiveness of cardiac

disease prevention, diagnosis, and treatment. The ability to capture this variability with data-driven methods is highly beneficial for clinical practice and therefore a key objective of the cardiac image analysis community, as it allows population-specific shape analysis, disease and outcome prediction, dimensionality reduction, and computer modelling of cardiac function [14]. While traditional statistical models such as principal component analysis (PCA) have been widely used for this purpose [1,11,14], recent research efforts focus increasingly on deep learning methods [5,6,9,13]. In this paper, we propose a novel variational autoencoder (VAE) [8] architecture acting directly on memory-efficient point clouds to generate subpopulation-specifc 3D biventricular anatomy models. To the best of our knowledge, this is the first geometric deep learning approach for cardiac anatomy generation. Our point cloud surface representations avoid the sparsity issues of 3D voxelgrids leading to quick execution and high resolution. Compared to PCA and other traditional shape modelling techniques, our method can capture non-linear relations in the data and does not require any prior landmark detection or registration, making its application significantly simpler and less error-prone. The choice of VAE framework enables stable training and a compact but also interpretable latent space representation of population datasets. By additionally introducing multiple conditional inputs, we can generate arbitrarily large subpopulation-specific cohorts of artificial hearts, which allows us to visualize and better understand the effects of combinations of different subject characteristics on biventricular anatomy and function.

2   Methods

We first briefly describe the dataset used for method development, followed by the network architecture and training procedure.

2.1   Dataset

Our point cloud dataset is based on 3D reconstructions of cine MRI acquisitions obtained from volunteers of the UK Biobank study [10]. We randomly select $\sim$500 female and $\sim$500 male subjects and extract the end-diastolic (ED) and end-systolic (ES) slices from the temporal sequence for each case [2], allowing us to condition our method on two binary metadata variables (sex and cardiac phase). We follow the pipeline described in [3] to create the 3D point cloud reconstructions from each acquisition and split our dataset into $\sim$1700 and $\sim$300 point clouds for training and testing respectively with equal representation of all conditions.

2.2   Network Architecture

Our proposed model architecture consists of a point cloud-based geometric deep learning network embedded in a conditional β-VAE [7,8] framework (Fig. 1).

We choose the PointNet++ [12] and the Point Completion Network [15] as the baseline architectures of our encoder and decoder, respectively. We adapt
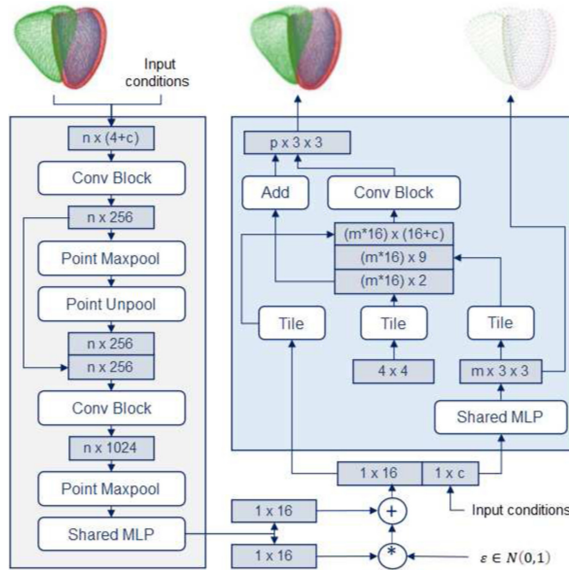
Fig. 1. Architecture of the proposed conditional Point Cloud β-VAE. The input (top left) is an unstructured point cloud with n points. Each point is represented by a 4-dimensional vector with three coordinate values (x,y,z) and a class label. A conditional vector c which contains additional information about the subject is concatenated to each input point vector as well as the latent space vector. The output consists of a coarse (top right) and a dense (top centre) point cloud generated from a random sample of the latent space distribution. Separate 3D coordinate values are used for each of the three classes in both output point clouds.

them to our multi-class setting by adding class information about the cardiac substructures (left ventricular (LV) endocardium, LV epicardium, right ventricular (RV) endocardium) to the encoder input and adjust the decoder architecture to output separate point clouds for each class. We enable conditional point cloud generation by concatenating our global input conditions to both encoder and decoder inputs. In order to effectively process high-density surface data and cope with the difficulty of latent space sampling, we also insert multiple fully connected layers to facilitate the exchange of spatial, class, and condition information. The standard reparameterization approach [8] is applied in the network's latent space. We choose a latent space size of 16, which we found to be sufficiently large to capture almost all of the variability in cardiac shapes and maintain good disentanglement.

2.3 Loss Function

Our loss function follows the design of the β-VAE [7] with a reconstruction loss and a latent space loss balanced by a weighting parameter β. We use a β value of 0.2, chosen empirically as a good trade-off between low reconstruction error

and high latent space quality. The Kullback-Leibler divergence between the prior
and posterior distributions of the latent space is used as a second loss term [8].
We split the reconstruction loss into a coarse and a dense loss term [15], which
respectively compare the low-density and high-density point cloud predictions
of our network to the gold standard point clouds for all $C = 3$ classes in the
biventricular anatomy:

$$L_{recon} = \sum_{i=1}^{C} \left(L_{coarse,i} + \alpha * L_{dense,i} \right). \tag{1}$$

The weighting parameter α allows to dynamically adjust the importance of
each reconstruction loss term during training. Initially, it is set to a low value of
0.01 to allow the network to focus on accurate reconstruction of global shapes,
and is then gradually increased during training until it reaches the value 5.0 to
put more emphasis on local structures in the high density output while main-
taining a good overall shape. Due to its approximation of a surface-to-surface
distance and its ability to process point cloud data, we propose the Chamfer dis-
tance (CD) between the predicted point cloud P_1 and the gold standard input
point cloud P_2 as a metric for both terms of the reconstruction loss:

$$CD(P_1, P_2) = \frac{1}{2} \left(\frac{1}{|P_1|} \sum_{x \in P_1} \min_{y \in P_2} \|x - y\|_2 + \frac{1}{|P_2|} \sum_{y \in P_2} \min_{x \in P_1} \|y - x\|_2 \right). \tag{2}$$

3 Experiments

We evaluate our method in terms of both its point cloud reconstruction and
generation performance. We also analyze its ability to correctly incorporate con-
ditional inputs into the generation process and calculate commonly used clinical
metrics over the generated heart shapes.

3.1 Reconstruction Quality

In order to assess the VAE's reconstruction ability, we select the point clouds
of the unseen test dataset as our gold standard, input them into the network,
and compare these inputs to the network's reconstructions using the Chamfer
distance. We report the results separated by class and subpopulation in Table 1.

We find mean distance values to be consistently below the pixel resolution
of the underlying MR images ($1.8 \times 1.8 \times 8.0$ mm) [10] and standard deviations
all in the range of 0.19 mm to 0.32 mm.

For a qualitative evaluation of our method's reconstructions, we visualize
the network input and output point clouds of five sample cases in Fig. 2. We
observe that our method is able to reconstruct anatomical surfaces with high
accuracy on both a global and local level for all biventricular substructures and
can successfully cope with considerable variations.

Table 1. Reconstruction results of the proposed method on the test dataset.

Condition		Chamfer distance (mm)		
Sex	Phase	LV endocardium	LV epicardium	RV endocardium
Female	ED	1.06 (±0.23)	1.17 (±0.32)	1.37 (±0.29)
	ES	0.88 (±0.23)	1.00 (±0.21)	1.11 (±0.25)
Male	ED	1.16 (±0.23)	1.25 (±0.21)	1.38 (±0.24)
	ES	0.91 (±0.19)	1.08 (±0.21)	1.22 (±0.23)

Values represent mean (± standard deviation) in all cases.
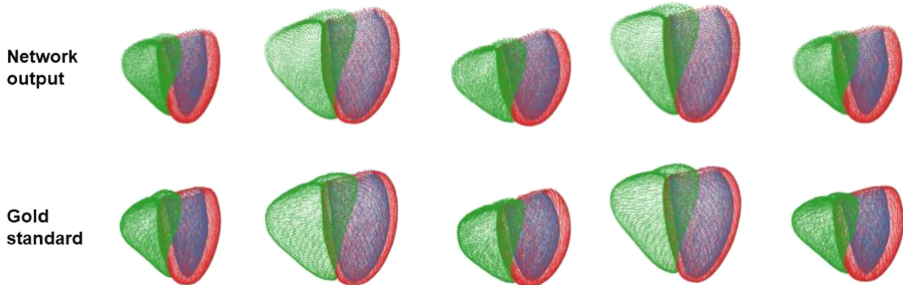

Fig. 2. Qualitative reconstruction results of our method on five sample cases.

3.2 Conditional Point Cloud Generation

In order to evaluate the generative performance of our method, we randomly sample from the latent space probability distribution and add either a 'male' or a 'female' label as well as either an 'ED' or an 'ES' label as conditional inputs to assess the ability of the method to generate specific subpopulations. We then pass the samples through the trained decoder part of our network. Figure 3 shows the generated point clouds from two such samples.

Comparing the point clouds in Fig. 3, we observe noticeable differences in sizes and shapes, indicating the decoder's ability to generate diverse point clouds. The effects of changing conditional inputs of each latent space vector on the reconstructed anatomy are also easily visible in a column-wise comparison and match well-known clinical expectations. For example, male hearts exhibit a larger size in both ED and ES phases than their female counterparts.

Next, we randomly sample 500 latent space vectors and use our trained decoder to generate random subpopulations for each combination of conditional inputs (ED female, ES female, ED male, ES male). We then convert both generated and test set point clouds into meshes using the Ball Pivoting algorithm [4]. This allows us to calculate common clinical metrics for each mesh and thereby quantify the clinical accuracy of our generated subpopulations compared to the meshes of the test dataset, that we consider to be our gold standard (Table 2).
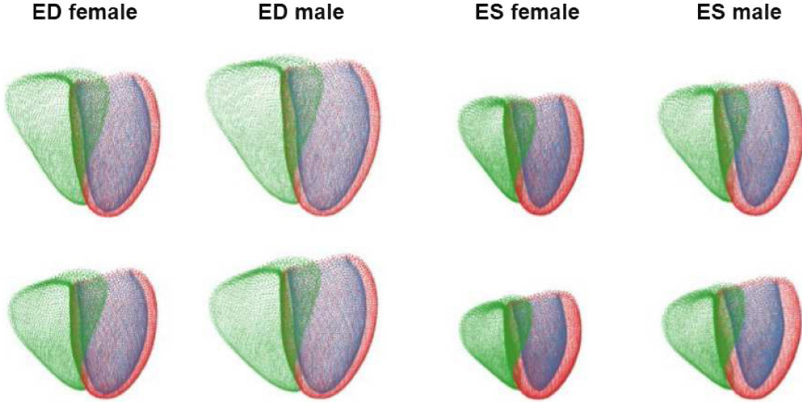
Fig. 3. Generated point clouds from two randomly sampled latent space vectors (rows) for each combination of input conditions (columns).

Table 2. Clinical metrics of meshed point clouds generated by our method with specific input conditions.

| Sex	Phase	Clinical metric	Gold standard	Ours
Female	ED	LV volume (ml)	124 ($\pm$24)	122 ($\pm$24)
		RV volume (ml)	147 ($\pm$26)	146 ($\pm$27)
	ES	LV volume (ml)	50 ($\pm$13)	53 ($\pm$13)
		RV volume (ml)	64 ($\pm$15)	69 ($\pm$14)
	ED/ES	LV mass (g)	85 ($\pm$19)	85 ($\pm$21)
Male	ED	LV volume (ml)	153 ($\pm$28)	154 ($\pm$29)
		RV volume (ml)	190 ($\pm$27)	192 ($\pm$32)
	ES	LV volume (ml)	65 ($\pm$13)	73 ($\pm$15)
		RV volume (ml)	91 ($\pm$16)	98 ($\pm$16)
	ED/ES	LV mass (g)	121 ($\pm$23)	122 ($\pm$27)

Values represent mean ($\pm$ standard deviation) in all cases.

We find comparable values across all clinical metrics and subpopulations in terms of both means and standard deviations. Slightly better scores are achieved for female hearts and the ED phase than for male hearts and the ES phase.

3.3   Latent Space Analysis

The quality of the latent space distribution plays an important role in the VAE's ability to synthesize artificial populations of realistic hearts that are also sufficiently diverse. We analyze the contributions of each part of the latent space to the generated point clouds by varying individual latent space components, while keeping the remaining latent space constant, and passing the resulting vectors

through the decoder to obtain the respective outputs. Figure 4 shows the synthesized point clouds corresponding to variations in three sample latent space dimensions, similar to the most important modes of variation in a PCA analysis.
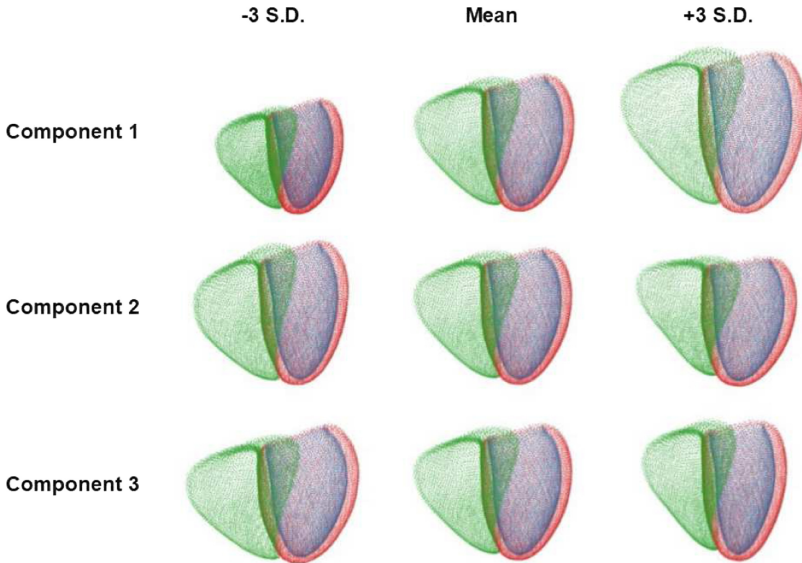
Fig. 4. Effect of different latent space components on point cloud reconstructions.

We observe gradual interpretable changes to the biventricular shapes and sizes without loss of a realistic appearance, while individual components encode different aspects of the biventricular anatomy. Among other things, component 1 is responsible for the overall heart size, component 2 changes the orientation angle of the basal plane of the heart, while component 3 transforms thin hearts with small mid-ventricular short-axis diameters into thicker ones.

4   Discussion

In this work, we have developed an efficient and easy-to-use method for synthesizing 3D biventricular anatomies conditioned on subject metadata. The method does not require any registration or point correspondence while maintaining high accuracy and diversity in its generation task. It is also capable of efficiently working with high-dimensional 3D MRI-based surface data due to its usage of point clouds instead of highly-sparse and memory-intensive voxelgrids. We achieve mean Chamfer distances considerably below the pixel resolution of the underlying images, demonstrating good reconstruction quality, while the small standard deviation values indicate that our method is highly robust and can successfully

cope with a variety of different morphologies, both within and between subpopulations. Our approach is able to process multi-class point clouds which allows us to model different cardiac substructures with a single network. Despite no explicit constraint on the connectivity of the different substructures, we do not observe any sizeable disconnected or overlapping components between them. We therefore conclude that the low values in the general reconstruction loss were sufficient to implicitly impose correct inter-class connectivity. The closeness in mean clinical metrics of the synthesized subpopulation-specific distributions and the respective gold standard values show our method's good generative performance as well as its ability to accurately incorporate multiple conditional inputs into the generation process. In addition, the observed similarities in standard deviation values demonstrate that our method can produce a highly diverse set of point clouds that is representative of the real population. We find easily interpretable and gradual anatomical changes resulting from latent space variations of each component, which indicate that the latent space resembles a continuous unimodal probability distribution. This finding is also in line with other commonly used statistical approaches for population-based cardiac shape modelling, such as the effect of varying along the primary modes of variation in a PCA model. However, due to its non-linear design, our method is capable of capturing more complex relationships in the data while maintaining interpretability. Furthermore, we observe good latent space disentanglement with each component encoding different aspects of the biventricular anatomy. To this end, the weighting parameter β of the β-VAE framework was important for our high-dimensional dataset as it allowed for the right balance to be set between latent space and reconstruction quality.

5   Conclusion

In this work, we have presented an easy and efficient geometric deep learning method capable of generating arbitrarily-sized populations of realistic biventricular anatomies. We have shown how different subject metadata can be successfully incorporated into our approach to synthesize subpopulation-specific heart cohorts and how our method's compact latent space representation enables an interpretable shape analysis of cardiac anatomical variability.

References

1. Bai, W., et al.: A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion. Med. Image Anal. **26**(1), 133–145 (2015)
2. Banerjee, A., et al.: A completely automated pipeline for 3D reconstruction of human heart from 2D cine magnetic resonance slices. Philosoph. Trans. Royal Soc. A., p. 20200257 (2021)
3. Beetz, M., Banerjee, A., Grau, V.: Biventricular surface reconstruction from cine MRI contours using point completion networks. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 105–109 (2021)
4. Bernardini, F., Mittleman, J., Rushmeier, H., Silva, C., Taubin, G.: The ball-pivoting algorithm for surface reconstruction. IEEE Trans. Visual. Comput. Graphics **5**(4), 349–359 (1999)
5. Biffi, C., et al.: Explainable anatomical shape analysis through deep hierarchical generative models. IEEE Trans. Med. Imaging **39**(6), 2088–2099 (2020)
6. Gilbert, K., Mauger, C., Young, A.A., Suinesiaputra, A.: Artificial intelligence in cardiac imaging with statistical atlases of cardiac anatomy. Front. Cardiovasc. Med. **7**, 102 (2020)
7. Higgins, I., et al.: beta-VAE: learning basic visual concepts with a constrained variational framework. In: 5th International Conference on Learning Representations (ICLR), pp. 1–13 (2017)
8. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114 (2013)
9. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
10. Petersen, S.E., et al.: UK Biobank's cardiovascular magnetic resonance protocol. J. Cardiovasc. Magn. Reson. **18**(1), 1–7 (2015)
11. Piazzese, C., Carminati, M.C., Pepi, M., Caiani, E.G.: Statistical shape models of the heart: applications to cardiac imaging. In: Statistical Shape and Deformation Analysis, pp. 445–480. Elsevier (2017)
12. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: deep hierarchical feature learning on point sets in a metric space. In: Advances in neural information processing systems, pp. 5099–5108 (2017)
13. Rezaei, M.: Chapter 5 - Generative adversarial network for cardiovascular imaging. In: Al'Aref, S.J., Singh, G., Baskaran, L., Metaxas, D. (eds.) Machine Learning in Cardiovascular Medicine, pp. 95–121. Academic Press (2021)
14. Tavakoli, V., Amini, A.A.: A survey of shaped-based registration and segmentation techniques for cardiac images. Comput. Vision Image Understanding, **117**(9), 966–989 (2013)
15. Yuan, W., Khot, T., Held, D., Mertz, C., Hebert, M.: PCN: point completion network. In: 2018 International Conference on 3D Vision (3DV), pp. 728–737 (2018)