COMBINED GENERATION OF ELECTROCARDIOGRAM AND CARDIAC ANATOMY MODELS USING MULTI-MODAL VARIATIONAL AUTOENCODERS

Marcel Beetz[⋆] *Abhirup Banerjee*[†⋆] *Yuling Sang*[⋆] *Vicente Grau*[⋆]

[⋆] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK
[†] Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, UK

ABSTRACT

Understanding population-wide variability of the human heart is crucial to detect abnormalities and improve the assessment of both cardiac anatomy and function. While many computational modeling approaches have been developed to capture this variability separately for either cardiac anatomy or physiology, their complex interconnections have rarely been explored together. In this work, we propose a novel multi-modal variational autoencoder (VAE) capable of processing combined physiology and bitemporal anatomy information in the form of electrocardiograms (ECG) and 3D biventricular point clouds. Our method achieves high reconstruction accuracy on a UK Biobank dataset with Chamfer distances between predicted and input anatomies below the underlying image resolution and the ECG reconstructions outperforming a state-of-the-art benchmark approach specialized in ECG generation. We also evaluate its generative ability and find comparable populations of generated and gold standard anatomies, ECGs, and combined anatomy-ECG data in terms of common clinical metrics and maximum mean discrepancies.

***Index Terms*—** Multi-Modal VAE, 3D Cardiac Anatomy Generation, ECG Synthesis, Combined Anatomy and Electrophysiology Modeling, Geometric Deep Learning.

1. INTRODUCTION

Human cardiac function comprises of a complex interplay of mechanical deformation and electrophysiology, which often varies considerably between individuals. Accounting for this variability is crucial for an accurate understanding of healthy and pathological cardiac physiology and has therefore been the goal of numerous computational modeling approaches. While principal component analysis has traditionally been used to model such variability in anatomical shapes or electrocardiograms (ECG), deep learning techniques have more recently been proposed for this task [1, 2, 3]. Such models have a variety of benefits and clinical use cases, including dimensionality reduction, data augmentation for machine learning algorithms, interpretable clinical outcome prediction, and numerical simulations of cardiac mechanics and electrophysiology. However, most of the previous works have focused on uni-modal approaches to model either anatomy or physiology separately, ignoring their intricate interactions and limiting a more comprehensive understanding of cardiac function. In this work, we aim to develop the first multi-modal generative deep learning model capable of encoding, reconstructing, and synthesizing combined electrophysiology and anatomy information in the form of electrocardiograms and bitemporal 3D biventricular anatomy point clouds, respectively. To this end, we propose a novel multi-branch variational autoencoder (VAE) [4] design with each branch architecture specifically tailored to its respective modality. Efficient point cloud representations of the multi-class 3D anatomies at both the end-diastolic (ED) and end-systolic (ES) phases of the cardiac cycle are selected as network inputs to represent 3D temporal changes of three biventricular substructures (left ventricular (LV) cavity, LV myocardium, and right ventricular (RV) cavity). All VAE branches share a common latent space, which enables an effective inter-modal information exchange and the successful processing of combined ECG and anatomy data.

2. METHODS

2.1. Dataset

Our dataset consists of 1000 randomly selected cases with paired cardiac cine magnetic resonance (MR) images and ECGs from the UK Biobank imaging study [5]. We use the multi-step pipeline described in [6] to reconstruct 3D point clouds of the biventricular anatomy at both ED and ES from the corresponding cine MR image frames and select the lead II ECG signals averaged across multiple heartbeats for each case. Our data is randomly split into train, validation, and test sets of sizes $\sim$800, $\sim$50, and $\sim$150, respectively.

2.2. Multi-modal variational autoencoder architecture

The overall architecture of our method consists of a multi-modal beta-VAE [7] with three branches that share a common latent space for cross-modal information exchange (Fig. 1).

Each of the three branches has an encoder-decoder structure and is responsible for processing either the ED anatomy

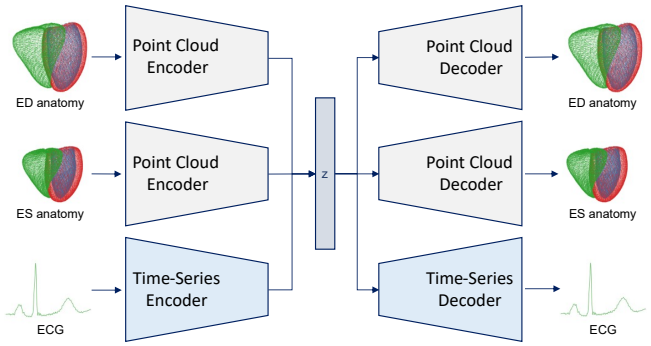

Fig. 1. Architecture of our proposed multi-modal VAE. The input consists of two point clouds which represent the multi-class biventricular anatomy at the ED and ES phases of the cardiac cycle and the corresponding ECG signal. Each of the three inputs is passed through separate encoder-decoder branches which share a common latent space z for cross-modal information exchange. The architecture of the individual branches is specifically tailored to the respective modality to allow efficient data processing.

point clouds, the ES anatomy point clouds, or the ECGs. The encoder outputs of the three branches are tasked with predicting the mean and standard deviation vectors of the multivariate Gaussian distribution that makes up the latent space z of the VAE. A 12-dimensional vector is sampled from this distribution and passed into each decoder of the three branches for the respective reconstruction. The standard reparameterization trick [4] is applied during training. Each of the two anatomy branches consists of an extended PointNet++ architecture [8] as the encoder and an expanded version of the Point Completion Network upsampling mechanism [9] as the decoder [1]. The encoder of the ECG branch starts with two convolutional blocks followed by an average pooling layer and two fully connected layers, while its decoder passes the sampled latent space vector through two fully connected layers and two transpose convolutions.

2.3. Loss function and training

Following the formulation of the beta-VAE [7] framework, our loss function is composed of the sum of a reconstruction loss term and the Kullback-Leibler divergence term weighted by the parameter β. We choose a β value of 0.25 which we have empirically found to provide a good balance between overall reconstruction quality and latent space quality. Our reconstruction loss L_{recon} consists of three loss terms, one for each of the three branches in the multi-modal VAE:

$$L_{recon} = L_{ED} + L_{ES} + \gamma * L_{ECG} \quad (1)$$

We introduce a parameter γ to control the importance of the ECG reconstruction during training and select the mean squared error as our ECG loss term. Each of the two anatomy loss terms L_{ED} and L_{ES} consists of the sum of a coarse and a dense loss term weighted by a parameter α [1]. It is used during training to initially prioritize the coarse loss, maintaining a good global structure of the prediction, and then gradually focus on a higher local accuracy in the dense point cloud prediction. Both the coarse and dense loss terms are calculated using the Chamfer distance [1].

3. EXPERIMENTS

3.1. Reconstruction quality

We first evaluate our network's ability to accurately reconstruct both the two input point clouds and the corresponding input electrocardiograms. To this end, we pass the ED point cloud, ES point cloud, and the ECG time series for each case of the test dataset through the trained network and compare the network's predicted outputs with the respective inputs. Figure 2 shows input and prediction data of three such sample cases. We observe good global and local alignment between inputs and predictions of both time series and point cloud data including all three anatomical substructures.

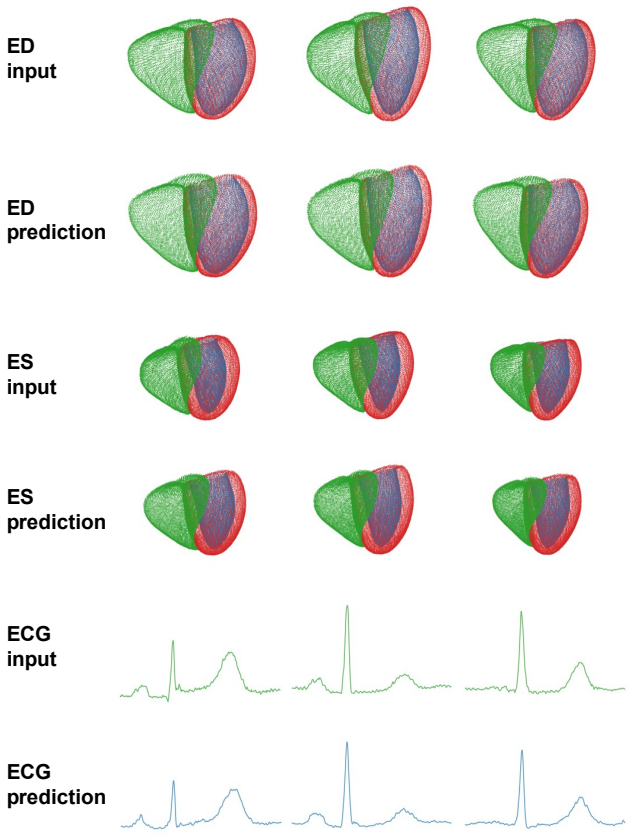

Fig. 2. Visual reconstruction results for three sample cases.

Next, we quantify our method's reconstruction ability on the test dataset using separate metrics for each modality. Sim-

ilar to [3], we select the percentage root mean squared distance (PRD) and root mean squared error (RMSE) metrics to determine our method's ECG reconstruction error (Table 1). We find that our method achieves lower reconstruction errors than the state-of-the-art BiLSTM-CNN generative adversarial network (GAN) in [3] for both metrics. However, this should be interpreted as only an approximate comparison, since they focused exclusively on ECG data without incorporating any anatomical information and also used a different ECG dataset.

Table 1. ECG reconstruction results of two different methods.

Method	Dataset	PRD	RMSE
BiLSTM-CNN GAN [3]*	MIT-BIH	51.80	0.22
Proposed	UK Biobank	27.45	0.17

*Values obtained directly from [3].

In order to quantify the reconstruction ability of our method for the ED and ES point clouds, we calculate the Chamfer distance between input and reconstructed point clouds of the test dataset. We find low distance values of 1.36 ± 0.33 mm and 1.23 ± 0.46 mm for the ED and ES anatomies respectively, both of which are smaller than the pixel sizes of the underlying images (1.8×1.8 mm^2).

3.2. Generative ability

In order to assess our network's ability to generate sufficiently diverse populations of realistic anatomies and ECGs, we randomly sample from the latent space distribution and pass the resulting vectors through the three branches of the decoder. The resulting decoder outputs are depicted for four sample cases in Fig. 3. We observe that all generated outputs follow realistic shapes and sizes while maintaining a good amount of diversity between different cases.

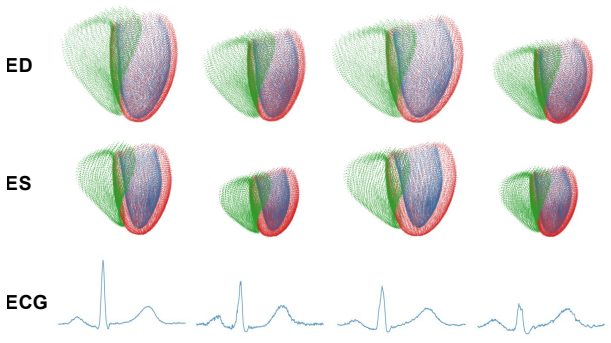

Fig. 3. Four randomly generated sample outputs.

Next, we quantify the VAE's generative capabilities by synthesizing 500 virtual anatomies and ECGs from randomly sampled latent space vectors. We calculate the Maximum Mean Discrepancy (MMD) [10] between the synthesized ECGs and the ECGs in our test dataset, as well as the MMD between two randomly selected subsets of the test dataset as our gold standard (Table 2). We observe that our method achieves MMD values closer to the gold standard than the GAN proposed in [2] albeit using a different dataset.

Table 2. ECG generation results of two different methods.

Method	Dataset	MMD
4CNN GAN [2]*	MIT-BIH	1.03×10^{-3}
Gold standard (test dataset)	UK Biobank	1.40×10^{-4}
Proposed	UK Biobank	3.54×10^{-5}

*Values obtained directly from [2].

We assess the quality of the 500 generated ED and ES anatomies by calculating common clinical metrics of both the synthesized point clouds and the gold standard test set point clouds (Table 3). All clinical metrics show high degrees of similarity between generated and gold standard point cloud populations for both ED and ES phases.

Table 3. Clinical metrics of ED and ES anatomy point clouds.

Phase	Clinical Metric	Gold Standard	Proposed
ED	LV volume (ml)	141 ($\pm$30)	139 ($\pm$31)
	RV volume (ml)	170 ($\pm$34)	176 ($\pm$37)
ES	LV volume (ml)	59 ($\pm$15)	58 ($\pm$16)
	RV volume (ml)	78 ($\pm$20)	80 ($\pm$24)
ED/ES	LV mass (g)	102 ($\pm$28)	99 ($\pm$29)

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

In order to ascertain that our method does indeed capture combined ECG/point cloud sets, rather than independent models for each, we first obtain a low-dimensional representation of the point cloud data by calculating 9 commonly used cardiac anatomy (presented in Table 3) and function metrics (LV and RV stroke volume, LV and RV ejection fraction) for each subject. We then select a weighted combination of these clinical metrics and the respective ECG signals as a unified representation of anatomy and ECG for each case. We use these combined representations to calculate the MMD between the generated and test datasets and the gold standard MMD between two random subsets of the test set. Our method obtains an MMD value of 5.04×10^{-2} close to the gold standard of 5.14×10^{-2}, suggesting a good degree of coupling between the generated anatomy and ECG outputs.

4. DISCUSSION AND CONCLUSION

In this work, we have developed a novel generative deep learning model to process combined cardiac anatomy and electrophysiology data in a multi-modal variational setting. We find high reconstruction accuracy of our method for both ECG and anatomy data showing that the modality-specific multi-branch VAE design is well suited to capture both intra-modal and inter-modal patterns. While the different datasets and number of modalities only allow for an approximate comparison of our method's ECG performance with previous work, its better scores do provide further validation of its good reconstruction ability. It also indicates that cardiac information from other domains might be beneficial for ECG synthesis. The similar mean and standard deviation values of the clinical metrics between synthetic and ground truth anatomy distributions show that the VAE can generate realistic shapes on both a local and global level and for both cardiac phases while exhibiting accurate levels of population diversity. This is achieved for multiple cardiac substructures represented as high resolution point cloud data, which is only made possible by the efficient point cloud-based deep learning design of the VAE's anatomy branches. Furthermore, the high similarity in MMD scores between generated and gold standard ECGs indicates that the ECG-specific branch of the VAE is capable of converting the random latent space inputs into realistic ECGs, both individually and on a population level. Compared to principal component analysis, our VAE can model more complex, non-linear relationships and does not require any point-to-point correspondence or registration of the input anatomies. Finally, we also find good MMD scores when combining representative anatomy biomarkers with ECG data, which suggests that the information shared via the common latent space enables each decoder branch to not only generate realistic data for its respective domain but at the same time take the overall cross-domain generation task into account.

5. COMPLIANCE WITH ETHICAL STANDARDS

This is a numerical simulation study for which no ethical approval was required.

6. ACKNOWLEDGMENTS

This research has been conducted using the UK Biobank Resource under Application Number '40161'. The authors express no conflict of interest. The work of M. Beetz was supported by the Stiftung der Deutschen Wirtschaft (Foundation of German Business). The work of A. Banerjee was supported by the British Heart Foundation (BHF) Project under Grant HSR01230. The work of V. Grau was supported by the CompBioMed 2 Centre of Excellence in Computational Biomedicine (European Commission Horizon 2020 research and innovation programme, grant agreement No. 823712).

7. REFERENCES

[1] M. Beetz, A. Banerjee, and V. Grau, "Generating subpopulation-specific biventricular anatomy models using conditional point cloud variational autoencoders," in *Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge*, 2022, pp. 75–83.

[2] A.M. Delaney, E. Brophy, and T.E. Ward, "Synthesis of realistic ECG using generative adversarial networks," *arXiv preprint arXiv:1909.09150*, 2019.

[3] F. Zhu, F. Ye, Y. Fu, Q. Liu, and B. Shen, "Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network," *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.

[4] D.P. Kingma and M. Welling, "Auto-encoding variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[5] S.E. Petersen, P.M. Matthews, F. Bamberg, et al., "Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK biobank - rationale, challenges and approaches," *Journal of Cardiovascular Magnetic Resonance*, vol. 15, no. 1, pp. 1–10, 2013.

[6] M. Beetz, A. Banerjee, and V. Grau, "Biventricular surface reconstruction from cine MRI contours using point completion networks," in *2021 IEEE 18th International Symposium on Biomedical Imaging*, 2021, pp. 105–109.

[7] I. Higgins, L. Matthey, A. Pal, et al., "beta-VAE: Learning basic visual concepts with a constrained variational framework," in *5th International Conference on Learning Representations*, 2017, pp. 1–13.

[8] C.R. Qi, L. Yi, H. Su, and L.J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," in *Advances in neural information processing systems*, 2017, pp. 5099–5108.

[9] W. Yuan, T. Khot, D. Held, C. Mertz, and M. Hebert, "PCN: Point completion network," in *2018 International Conference on 3D Vision*, 2018, pp. 728–737.

[10] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 723–773, 2012.