A Conditional Flow Variational Autoencoder for Controllable Synthesis of Virtual Populations of Anatomy

Haoran Dou[1], Nishant Ravikumar[1], and Alejandro F. Frangi[1,2,3,4(✉)]

[1] Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), University of Leeds, Leeds, UK
n.ravikumar@leeds.ac.uk
[2] Division of Informatics, Imaging and Data Science, Schools of Computer Science and Health Sciences, University of Manchester, Manchester, UK
alejandro.frangi@manchester.ac.uk
[3] Medical Imaging Research Center (MIRC), Electrical Engineering and Cardiovascular Sciences Departments, KU Leuven, Leuven, Belgium
[4] Alan Turing Institute, London, UK

Abstract. The generation of virtual populations (VPs) of anatomy is essential for conducting in silico trials of medical devices. Typically, the generated VP should capture sufficient variability while remaining plausible and should reflect the specific characteristics and demographics of the patients observed in real populations. In several applications, it is desirable to synthesise virtual populations in a *controlled* manner, where relevant covariates are used to conditionally synthesise virtual populations that fit a specific target population/characteristics. We propose to equip a conditional variational autoencoder (cVAE) with normalising flows to boost the flexibility and complexity of the approximate posterior learnt, leading to enhanced flexibility for controllable synthesis of VPs of anatomical structures. We demonstrate the performance of our conditional flow VAE using a data set of cardiac left ventricles acquired from 2360 patients, with associated demographic information and clinical measurements (used as covariates/conditional information). The results obtained indicate the superiority of the proposed method for conditional synthesis of virtual populations of cardiac left ventricles relative to a cVAE. Conditional synthesis performance was evaluated in terms of generalisation and specificity errors and in terms of the ability to preserve clinically relevant biomarkers in synthesised VPs, that is, the left ventricular blood pool and myocardial volume, relative to the real observed population.

Keywords: Virtual Population · Generative Model · Normalizing Flow

N. Ravikumar and A. F. Frangi—Joint last authors.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43990-2_14.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14226, pp. 143–152, 2023.
https://doi.org/10.1007/978-3-031-43990-2_14

1 Introduction

In-silico trials (ISTs) use computational modelling and simulation techniques with virtual twin or patient models of anatomy and physiology to evaluate the safety and efficacy of medical devices virtually [22]. Virtual patient populations (VPs), distinct from virtual twin populations, comprise plausible instances of anatomy and physiology that do not represent any specific real patient's data (as in the case of the latter, viz. virtual twins). In other words, VPs comprise synthetic data that help expand/enrich the diversity of anatomical and physiological characteristics that can be investigated within an IST for a given medical device. A key aspect of patient recruitment in real clinical trials used to assess device performance and generate regulatory evidence for device approval is the clear definition of inclusion and exclusion criteria for the trial. These criteria define the target patient population considered appropriate/safe to assess the performance of the device of interest. Consequently, it is desirable to enable the *controlled* synthesis of VPs that may be used for device ISTs, in a manner that emulates the imposition of trial inclusion and exclusion criteria.

Virtual populations can be considered to be parametric representations of the anatomy sampled from a generative model. Traditional statistical shape models (SSMs), based on methods such as principal component analysis (PCA), have been widely explored in the past decade [8,9,15]. Recent studies focus on deep learning-based generative models due to their automatic and powerful hierarchical feature extraction [3,7]. For instance, Bonazzola *et al.* [3] used a graph convolutional variational auto-encoder (gcVAE) to learn latent representations of 3D left ventricular meshes and used the learnt representations as surrogates for cardiac phenotypes in genome-wide association studies. Dou *et al.* [7] proposed learning the shape representations of multiple cardiovascular anatomies using gcVAE independently and then assembling them into complete whole-heart anatomies termed virtual heart chimaeras. Other studies have investigated conditional-generative models for synthesis of VPs of anatomies. For example, Beetz *et al.* [1] employed a conditional VAE (cVAE), conditioned on gender and cardiac phase, to allow the synthesis of VPs from biventricular anatomies. In subsequent work [2,12], they extended their method to a multidomain VAE to model biventricular anatomies at multiple times (across the cardiac cycle), using patient-specific electrocardiogram (ECG) signals as additional conditioning information (in addition to patient demographic data and standard clinical measurements) to guide the synthesis. All aforementioned methods model the latent space in the VAEs/cVAEs as a multivariate Gaussian distribution with a diagonal covariance matrix. This limits the flexibility afforded to the cVAE, as the Gaussian distribution, being unimodal, is a poor approximation to multimodal latent posterior distributions. This in turn limits the overall variability in anatomical shape that can be captured by standard VAEs and cVAEs.

In this study, we address the limitations of the state-of-the-art conditional generative models used to synthesise VPs of anatomical structures. In particular, we propose a method to relax the constraint on modelling the latent distribution as a unimodal multivariate Gaussian, to boost the flexibility of the generative

model, and to enable conditional synthesis of diverse and plausible VPs generation. Recent advances in normalising flows [14,16,21] introduce a new solution for this limitation by leveraging a series of invertible parameterized functions to transform the unimodal distribution to a multimodal one. Motivated by this technique, we propose the first conditional flow VAE (parameterised as a graph-convolutional network) for the task of *controllable* synthesis of VPs of anatomy. The contributions are as follows: (i) we introduce normalising flows to learn a multimodal latent posterior distribution by transforming the latent variables from a simple unimodal distribution. This helps the generative model capture greater anatomical variability from the observed real population, leading to the synthesis of more diverse VPs; (ii) we condition the flow-based VAE on patient demographic data and clinical measurements. This enables conditional synthesis of plausible VPs (given relevant covariates/conditioning information as inputs), which reflect the observed correlations between nonimaging patient information and anatomical characteristics in the real population.

2 Methodology

In this study, we propose a cVAE model equipped with normalising flows for controllable synthesis of VPs of cardiovascular anatomy. A schematic of the proposed conditional flow VAE network architecture is shown in Fig. 1. We employ normalising flows in the latent space of the cVAE to transform the initial Gaussian posterior to a complex multimodal distribution.

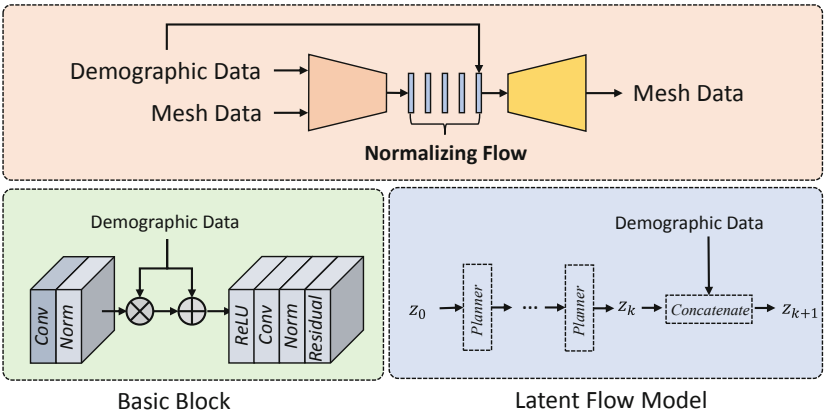

Fig. 1. Schematic illustration of our proposed conditional flow VAE

Conditional Variational Autoencoder: A VAE is a probabilistic generative model/network [11] that comprises an encoder and a decoder network branch. The encoder learns a mapping from the input data to a low-dimensional latent

space that abstracts the semantic representations from the observations, and the decoder reconstructs the original data from the low-dimensional latent representation. The latent space from which the observed data is generated is given by approximating the posterior distribution of the latent variables using variational inference. The VAE network is trained by maximising the evidence lower bound (ELBO), which is a summation of the expected log-likelihood of the data and the Kullback-Leibler divergence between the approximate posterior and some assumed prior distribution over the latent variables (typically a multivariate Gaussian distribution). Despite its effectiveness in capturing some of the observed variability in the training population (e.g. of anatomical shapes or images), VAEs do not provide any control over the generation process and hence cannot guarantee that the generated population anatomical shapes are representative of target patient populations with specific inclusion/exclusion criteria. Controllable synthesis of anatomical VPs is essential for constructing meaningful cohorts for use in ISTs. Conditional VAE [18] is a VAE-variant that uses additional covariates/conditioning information in addition to the input data (e.g. anatomical shapes) to learn a conditional latent posterior distribution (conditioned on the covariates), enabling controllable synthesis of VPs during inference (given relevant covariates/conditioning information as input).

Our conditional flow VAE (cVAE-NF) is a graph-convolutional network which takes as input a triangular surface mesh representation of an anatomical structure of interest, i.e., the Left Ventricle (LV) in this study, and its associated covariates/conditioning variables, i.e., the patient demographic data and clinical measurements, such as gender, age, weight, blood cholesterol, etc., and outputs the reconstructed surface mesh. Each mesh is represented by a list of 3D spatial coordinates of its vertices and an adjacency matrix defining vertex connectivity (i.e. edges of mesh triangles). The encoder and decoder contain five residual graph-convolutional blocks, respectively. Each block comprises two Chebyshev graph convolutions, each of which is followed by batch normalisation and ELU activation. A residual connection is added between the input and the output of each graph-convolutional block. Hierarchical mesh down/up-sampling operations proposed in CoMA [13] are adopted after each block to capture the global and local shape context. The VAE model is conditioned on covariates by scaling the hidden representations in the encoder similar to adaptive instance normalization [10] given the covariates as input to generate the scaling factor, and by concatenating the covariates with the latent variables before decoding.

Flexible Posterior Using Normalizing Flow: Vanilla cVAEs model the approximate posterior distribution using Gaussian distributions with a diagonal covariance matrix. However, such a unimodal distribution is a poor approximation of the complex true latent posterior distribution in most real-world applications (e.g. for shapes of the LV observed across a population), limiting the anatomical variability captured by the model. In this study, we introduce normalising flows to construct a flexible multi-modal latent posterior distribution by applying a series of differentiable, invertible/diffeomorphic transformations

iteratively to the initial simple unimodal latent distribution. As shown in Fig. 2, a two-dimensional Gaussian distribution can be transformed into a multi-modal distribution by applying several normalising flow steps to the former.

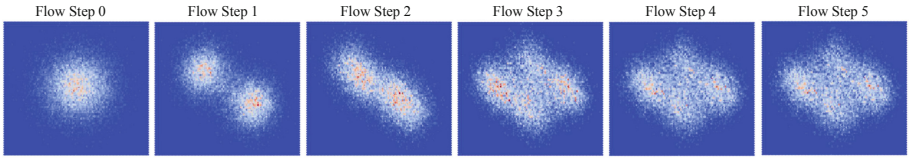

Fig. 2. Effect of normalising flow on Gaussian distribution. Step 0 is the initial two-dimensional Gaussian distribution, and step 1–5 represents the distribution of latent variables transformed by the normalising flow layers (i.e., planar flow).

Consider an invertible and smooth mapping function $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ with inverse $f^{-1} = g$, and a random variable $\mathbf{z}$ with distribution $q(\mathbf{z})$. The transformed variable $\mathbf{z}' = f(\mathbf{z})$ follows a distribution given by:

$$q(\mathbf{z}') = q(\mathbf{z}) \left| \det \frac{\partial f}{\partial \mathbf{z}} \right|^{-1} \tag{1}$$

where the $\det \frac{\partial f}{\partial \mathbf{z}}$ is the Jacobian determinant of f. Therefore, we can obtain a complex multi-modal density by composing multiple invertible mappings to transform the initial, simple and tractable density sequentially, as follows,

$$\mathbf{z}_i = f_i \circ \cdots \circ f_2 \circ f_1(\mathbf{z}_0) \tag{2}$$

$$\ln q_i(\mathbf{z}_i) = \ln q_0(\mathbf{z}_0) - \sum^{i} \ln \left| \det \frac{\partial f_i}{\partial \mathbf{z}_{i-1}} \right| \tag{3}$$

The specific mathematical formulation of the normalising flow function is important and must be chosen with care to allow for efficient gradient computation during training, scalable inference, and efficiency in computing the determinant of the Jacobian. In this study, we leverage the planar flow in [16] as a basic unit of our latent normalising flow net. Specifically, each transformation unit is given by,

$$f(\mathbf{z}) = \mathbf{z} + \mathbf{u} h(w^\top \mathbf{z} + b) \tag{4}$$

where $\mathbf{w} \in \mathbb{R}^d$, $\mathbf{u} \in \mathbb{R}^d$ and $b \in \mathbb{R}$ are learnable parameters; $h(\cdot)$ is a smooth element-wise non-linear function with derivative $h'(\cdot)$ (we use tanh in our study) and $\mathbf{z}$ denotes the latent variables sampled from the posterior distribution. Therefore, we could compute the log determinant of the Jacobian term in $O(D)$ time as follows:

$$\phi(\mathbf{z}) = h'(\mathbf{w}^\top \mathbf{z} + b)\mathbf{w} \tag{5}$$

$$\left| \det \frac{\partial f_i}{\partial \mathbf{z}_{i-1}} \right| = \left| \det(\mathbf{I} + \mathbf{u}\phi(\mathrm{z})^\top) \right| = \left| 1 + \mathbf{u}^\top \phi(\mathbf{z}) \right| \tag{6}$$

Finally, the network is trained by optimizing the modified ELBO based on Eq. 3:

$$\ln p(\mathbf{x}|\mathbf{c}) \geq \mathbb{E}_{q(\mathbf{z}_0|\mathbf{x},\mathbf{c})}\left[\ln p(\mathbf{x}|\mathbf{z}_i,\mathbf{c}) + \sum^{i} \ln \left|\det \frac{\partial f_i}{\partial \mathbf{z}_{i-1}}\right|\right] - \mathrm{KL}(q(\mathbf{z}_0|\mathbf{x},\mathbf{c})\|p(\mathbf{z}_i)) \quad (7)$$

where, $\ln p(\mathbf{x}|\mathbf{c})$ is the marginal log-likelihood of the observed data $\mathbf{x}$ (i.e. here $\mathbf{x}$ represents an LV graph/mesh), conditioned on the covariates of interest (i.e. patient demographics and clinical measurements) $\mathbf{c}$; i is the steps of the normalizing flows. $p(\mathbf{x}|\mathbf{z}_i,\mathbf{c})$ is the likelihood of data parameterised by the decoder network, which reconstructs/predicts $\mathbf{x}$ given the latent variables $\mathbf{z}_i$, transformed by latent (planar) normalising flows, and the conditioning variables $\mathbf{c}$; $\mathrm{KL}(q(\mathbf{z}_0|\mathbf{x})\|p(\mathbf{z}_i))$ is the Kullback-Leibler divergence of the approximate posterior initial $q(\mathbf{z}_0|\mathbf{x},\mathbf{c})$ from the prior, $p(z) = \mathcal{N}(z \mid 0, I)$.

3 Experimental Setup and Results

Data: In this study, we created a cohort of 2360 triangular meshes of the left ventricle (LV) based on a subset of cardiac cine-MR imaging data available from the UK Biobank (UKBB) by registering a cardiac LV atlas mesh [17] in manual contours (as described in [23]). We randomly split the data set into 422/59/1879 for training, validation, and testing, respectively. All meshes have the same and fixed graph topology, sharing the same edges and faces but differing in the position of vertices; i.e. there is pointwise correspondence across all shapes. We used 14 covariates available for the same subjects in UKBB as conditioning variables for our model, including, gender, age, height, weight, pulse, alcohol drinker status, smoking status, HbA1c, cholesterol, C-reactive protein, glucose, high-density lipoprotein cholesterol (HDL), insulin-like growth factor 1 (IGF-1), and low-density lipoprotein (LDL) cholesterol. These covariates were chosen because they are known cardiovascular risk factors.

Implementation Details: The framework was implemented using PyTorch on a standard PC with a NVIDIA RTX 2080Ti GPU. We trained our model using the AdamW optimizer with an initial learning rate of 1e–3 and batch size of 16 for 1000 epochs. The feature number for each graph convolutional block in the encoder was 16, 32, 32, 64, 64, and in reverse order in the decoder. The latent dimension was set at 16. The down/up-sampling factor was four, and we used a warm-up strategy [19] to the weight of the KL loss to prevent model collapse.

Evaluation Metrics: We compared our model (cVAE-NF) with a traditional PCA-based SSM, two generative models without conditioning information including a vanilla VAE and a VAE with normalising flow (VAE-NF) and the vanilla cVAE. Comparison of the vanilla cVAE can also validate the performance of existing approaches [1,2] because they are built on the cVAE with different covariates and basic units in the network. We evaluated the performance of all

methods using three different metrics: 1) the reconstruction error, which evaluates the generalisability of the trained model to reconstruct/represent unseen shapes, using the distance between the reconstructed mesh with the ground truth/original mesh; 2) the specificity error, which measures the anatomical plausibility of the virtual cohorts synthesised, using the distance between the generated meshes and its nearest neighbour in the unseen real population [6]; and 3) the variability in the left ventricular volume in the synthesised cohorts, to assess the diversity of the instances generated in terms of a clinically relevant cardiac index. The variability in LV volume was quantified as the standard deviation of the volumes of LV blood pools (BPVols). The Euclidean distance was used to evaluate all three metrics. Additionally, we measured the activity of the latent dimension using the statistic $A = \mathrm{Cov}_{\mathbf{x}}(\mathbb{E}_{\mathbf{z}\sim q(\mathbf{z}|\mathbf{x})}[z])$ of the observations $\mathbf{x}$ [4]. A higher activity score indicates that a given latent dimension can capture greater population-wide shape variability.

Table 1. The quantitative results of the investigated methods in a hold-out test dataset. The bold values represent the results are significantly better than those of other methods.

Methods	Reconstruction Error↓	Specificity Error↓	Volume Variability↑
PCA	$\mathbf{0.82 \pm 0.16}$	1.48 ± 0.26	**32.74**
VAE	1.29 ± 0.21	$\mathbf{1.39 \pm 0.98}$	3.00
VAE-NF	0.90 ± 1.76	1.60 ± 0.34	16.03
cVAE	1.43 ± 0.26	$\mathbf{1.32 \pm 0.21}$	28.39
Ours	$\mathbf{1.23 \pm 0.23}$	1.38 ± 0.20	**29.91**

The results of our method are presented in Table 1. Our model outperforms the cVAE in terms of reconstruction error and the amount of volume variability captured in the synthesised VP (the reference volume variability for the real UKBB population was 33.38 mm^3). However, the cVAE achieved lower specificity errors than our model. This indicates that our method is better at capturing the population's shape variability, but it also creates some instances that are further away from the real population, resulting in higher specificity errors. We attribute this to the normalising flow's ability to learn a more flexible approximate posterior latent distribution of the observed shapes than the cVAE. This is also seen when comparing the performance of VAE and VAE-NF, where the latter can synthesise significantly more diverse VPs (e.g. it improves the volume variability from 3.00 to 16.03). Figure 3 shows the variability captured in each latent dimension. We observe that VAE-NF has higher activity scores in all latent dimensions compared to vanilla cVAE. The normalising flow allows for the approximation of multimodal latent distributions in the generative model, resulting in greater shape variability. Although PCA outperforms our method in terms of generalisation error and volume variability captured, it does not allow for controllable synthesis of VPs based on relevant patient demographic information and clinical

measurements, making it less useful for our application of synthesising VPs for use in ISTs.

It is essential to capture the distribution of clinically relevant biomarkers (e.g. BPVol) in the synthesised virtual populations (VPs) based on the specified covariates/conditioning information available for real patients, in order to effectively replicate the inclusion/exclusion criteria used during trial design in ISTs. For example, the BPVol of women is known to be lower than that of men [20]. To verify this, we generated VPs using cVAE and our method, conditioned on real patient data (covariates) from the UK Biobank. Figure 3 summarises the BPVol distribution for both genders in the synthesised VPs and the real UKBB population, and the former accurately reflects the known trend of women having lower BPVol than men. Compared to cVAE, our model generates a VP that more closely matches the distribution of the volume of the LV blood pool observed in the real population. We also visualised the effect of manipulating individual attributes on two real patients in Fig. 4. We selected two representative attributes that are significantly associated with BPVol and myocardial volume (MyoVol): weight and age. We observe that BPVol and MyoVol of the

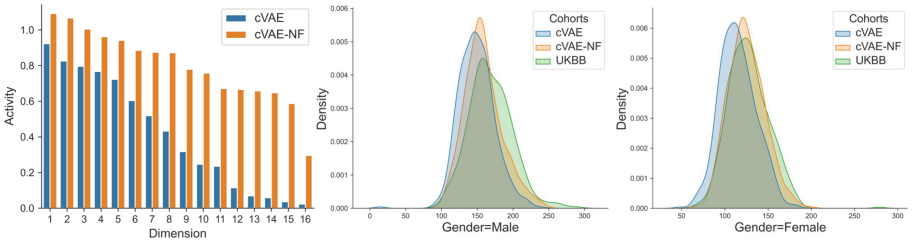

Fig. 3. Left: Comparison of the activity scores in different latent dimensions between the cVAE and cVAE-NF; right: Kernel density plots for BPVol from the VPs generated by cVAE and cVAE-NF and the real patient population (UKBB).

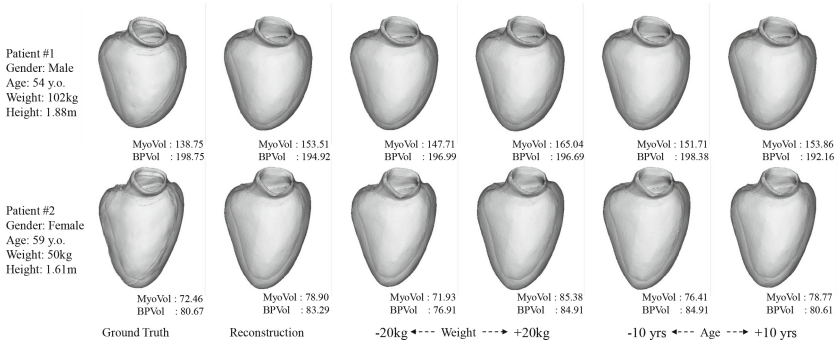

Fig. 4. Two representative examples of the reconstructed shapes and their variations through manipulation over two demographic attributes, i.e., weight and Age. MyoVol and BPVol are shown in the bottom right corner.

LV are positively correlated with the weight of the patients (as expected). On the other hand, increasing the individual's age results in a smaller BPVol, but an increased MyoVol (as visualised in Fig. 4), which is known to be due to cardiac hypertrophy caused by aging [5].

4 Conclusion

We proposed a conditional flow VAE model for the controllable synthesis of VPs of anatomy. Our approach was demonstrated to increase the flexibility of the learnt latent distribution, resulting in VPs that captured greater variability in the LV shape than the vanilla cVAE. Furthermore, our model was able to model the relationship between covariates/conditional variables and the shape of the LV, and synthesise target VPs that fit the desired criteria (in terms of demographics of the patient and clinical measurements) and closely matched the real population in terms of a clinically relevant biomarker (LV BPVol). These results suggest that our approach has potential for the controllable synthesis of diverse, yet plausible, VPs of anatomy. Future work will focus on modelling the whole heart and exploring the impact of individual covariates on VP synthesis in more detail.

Acknowledgement. This research was carried out using data from the UK Biobank (access application 11350). This work was supported by the Royal Academy of Engineering (INSILEX CiET1819/19), Engineering and Physical Sciences Research Council (EPSRC) UKRI Frontier Research Guarantee Programmes (INSILICO, EP/Y030494/1), and the Royal Society Exchange Programme CROSSLINK IES\NSFC\201380.

References

1. Beetz, M., Banerjee, A., Grau, V.: Generating subpopulation-specific biventricular anatomy models using conditional point cloud variational autoencoders. In: Puyol Antón, E., et al. (eds.) STACOM 2021. LNCS, vol. 13131, pp. 75–83. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-93722-5_9
2. Beetz, M., Banerjee, A., Grau, V.: Multi-domain variational autoencoders for combined modeling of mri-based biventricular anatomy and ecg-based cardiac electrophysiology. Front. Physiol. **991** (2022)
3. Bonazzola, R., Ravikumar, N., Attar, R., Ferrante, E., Syeda-Mahmood, T., Frangi, A.F.: Image-derived phenotype extraction for genetic discovery via unsupervised deep learning in CMR images. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 699–708. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_67
4. Burda, Y., Grosse, R., Salakhutdinov, R.: Importance weighted autoencoders. arXiv preprint arXiv:1509.00519 (2015)
5. Chiao, Y.A., Rabinovitch, P.S.: The aging heart. Cold Spring Harbor Perspect. Med. **5**(9), a025148 (2015)

6. Davies, R.H., Twining, C.J., Cootes, T.F., Taylor, C.J.: Building 3-d statistical shape models by direct optimization. IEEE Trans. Med. Imaging **29**(4), 961–981 (2009)
7. Dou, H., Virtanen, S., Ravikumar, N., Frangi, A.F.: A generative shape compositional framework: towards representative populations of virtual heart chimaeras. arXiv preprint arXiv:2210.01607 (2022)
8. Frangi, A.F., Rueckert, D., Schnabel, J.A., Niessen, W.J.: Automatic construction of multiple-object three-dimensional statistical shape models: application to cardiac modeling. IEEE Trans. Med. Imaging **21**(9), 1151–1166 (2002)
9. Gooya, A., Davatzikos, C., Frangi, A.F.: A bayesian approach to sparse model selection in statistical shape models. SIAM J. Imaging Sci. **8**(2), 858–887 (2015)
10. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1501–1510 (2017)
11. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
12. Li, L., Camps, J., Banerjee, A., Beetz, M., Rodriguez, B., Grau, V.: Deep computational model for the inference of ventricular activation properties. In: Camara, O., et al. (eds.) Statistical Atlases and Computational Models of the Heart. Regular and CMRxMotion Challenge Papers: 13th International Workshop, STACOM 2022, Held in Conjunction with MICCAI 2022, Singapore, 18 September 2022, Revised Selected Papers, pp. 369–380. Springer, Heidelberg (2023). https://doi.org/10.1007/978-3-031-23443-9_34
13. Ranjan, A., Bolkart, T., Sanyal, S., Black, M.J.: Generating 3d faces using convolutional mesh autoencoders. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 704–720 (2018)
14. Rasal, R., Castro, D.C., Pawlowski, N., Glocker, B.: Deep structural causal shape models. In: European Conference on Computer Vision, pp. 400–432. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-25075-0_28
15. Ravikumar, N., Gooya, A., Çimen, S., Frangi, A.F., Taylor, Z.A.: Group-wise similarity registration of point sets using student's t-mixture model for statistical shape models. Med. Image Anal. **44**, 156–176 (2018)
16. Rezende, D., Mohamed, S.: Variational inference with normalizing flows. In: International Conference on Machine Learning, pp. 1530–1538. PMLR (2015)
17. Rodero, C., et al.: Linking statistical shape models and simulated function in the healthy adult human heart. PLoS Comput. Biol. **17**(4), e1008851 (2021)
18. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. Adv. Neural Inf. Process. Syst. **28**, 1–9 (2015)
19. Sønderby, C.K., Raiko, T., Maaløe, L., Sønderby, S.K., Winther, O.: Ladder variational autoencoders. Adv. Neural Inf. Process. Syst. **29**, 1–9 (2016)
20. St Pierre, S.R., Peirlinck, M., Kuhl, E.: Sex matters: a comprehensive comparison of female and male hearts. Front. Physiol. **13**, 303 (2022)
21. Tomczak, J.M., Welling, M.: Improving variational auto-encoders using householder flow. arXiv preprint arXiv:1611.09630 (2016)
22. Viceconti, M., Pappalardo, F., Rodriguez, B., Horner, M., Bischoff, J., Tshinanu, F.M.: In silico trials: verification, validation and uncertainty quantification of predictive models used in the regulatory evaluation of biomedical products. Methods **185**, 120–127 (2021)
23. Xia, Y., et al.: Automatic 3d+ t four-chamber CMR quantification of the UK biobank: integrating imaging and non-imaging data priors at scale. Med. Image Anal. **80**, 102498 (2022)