

Jannis Hagenah*, Mohamad Mehdi, and Floris Ernst

Fully Data-Driven Pseudohealthy Synthesis for Planning Valve-Sparing Aortic Root Reconstruction using Conditional Variational Autoencoders

<https://doi.org/10.1515/cdbme-2020-3072>

Abstract: Aortic root aneurysm is treated by replacing the dilated root by a grafted prosthesis which mimics the native root morphology of the individual patient. The challenge in predicting the optimal prosthesis size rises from the highly patient-specific geometry as well as the absence of the original information on the healthy root. Therefore, the estimation is only possible based on the available pathological data. In this paper, we show that representation learning with Conditional Variational Autoencoders is capable of turning the distorted geometry of the aortic root into smoother shapes while the information on the individual anatomy is preserved. We evaluated this method using ultrasound images of the porcine aortic root alongside their labels. The observed results show highly realistic resemblance in shape and size to the ground truth images. Furthermore, the similarity index has noticeably improved compared to the pathological images. This provides a promising technique in planning individual aortic root replacement.

Keywords: aortic root reconstruction, representation learning, deep learning, conditional variational autoencoder

1 Introduction

Valve-sparing aortic root reconstruction surgery presents a promising treatment option for patients suffering from aortic root aneurysms [1]. During the surgery, the native aortic valve is preserved for patients with near-normal cusps in order to maintain smooth outflow and avoid anticoagulation [2]. The size of the implanted prosthesis has a significant influence on the blood flow in the aortic annulus as well as the long-term stability of the aortic valve [3]. However, the estimation of the optimal prosthesis size is still an intricate task. Patient prosthesis mismatch could lead to further complications as the aortic root shape and geometry is highly patient-specific [4]. Addi-

tionally, the available data on the patient root is only in the dilated state. Our aim is to assist the surgeons in the decision-making process by developing a preoperative planning tool. For this tool, we need an estimation of the healthy shape only based on its dilated state.

Previous works based on classic Machine Learning were conducted for personalized prosthesis size prediction. In the early steps, geometrical features were extracted from the ultrasound images as pairs of healthy and dilated features. These features were thereafter used to train a machine-learning algorithm, such as Support Vector Regression (SVR), to learn a mapping between the dilated and healthy states. Then, the SVR-model can be utilized to predict individual healthy features based on dilated ones [6]. A recent study has introduced an alternative solution for reconstructing the healthy state of the aortic root. Instead of handcrafting features from the ultrasound images, a new method, based on representation learning with neural networks, has been developed to learn the features from the input images at the hidden layers of the network. Variational Autoencoders (VAE) provide the possibility to convert the input into a compressed low-dimensional representation, also known as the latent space. After having the latent space for all images extracted, a translation vector is manually defined between the two classes to map between the dilated and the healthy images [5].

Variational Autoencoders have shown considerable performance as generative models, however, with standard VAE, the latent space is randomly sampled. This means that all latent points lie in one distribution. Thus, there is no control over the generated images. In this work, we propose an alternative approach for generating the healthy aortic root without direct interaction with the latent space. By including a condition on the image class to produce, it is possible to direct the autoencoder to output new reasonable images with specific attributes. In this paper, we present the concept of Conditional Variational Autoencoders (CVAE) for planning valve-sparing aortic root reconstruction. Furthermore, we perform a proof-of-concept study and compare our method to previously published methods.

*Corresponding author: Jannis Hagenah, Mohamad Mehdi, Floris Ernst, Institute for Robotics and Cognitive Systems, University of Luebeck, Ratzeburger Allee 160, 23562 Luebeck, Germany, e-mail: hagenah@rob.uni-luebeck.de. J. Hagenah and M. Mehdi contributed equally.

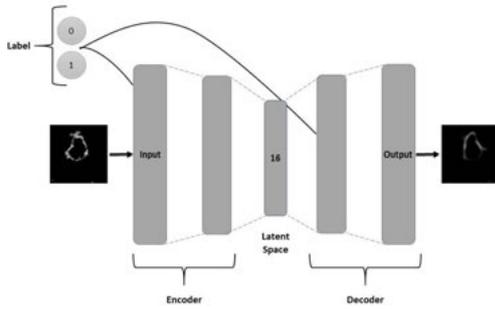


Fig. 1: Schematic illustration of the proposed architecture.

2 Material and Methods

2.1 Dataset

In this study, a dataset of 48 2D ultrasound images (greyscale, 100×100) of porcine valves has been used [6]. In this study, we only used the 2D images showing the commissure plane of the root. The dataset contains the healthy and dilated states of 24 aortic roots in their native, healthy state as well as a manually simulated pathological one so that the ground truth is known.

2.2 Model Architecture

In this experiment, the Conditional Variational Autoencoder (CVAE) has a similar architecture of traditional VAE. The encoder consists of five convolutional layers (kernel size 3×3) followed by average pooling (kernel size 2×2). The decoder is obtained by mirroring the architecture of the encoder. Furthermore, a new layer that represents the condition using one-hot encoding is imposed on the encoder and decoder inputs [7]. Figure 1 illustrates the architecture of the proposed model.

We used *ReLU* activation throughout the network except the last layer, where we applied *Sigmoid* activation. The size of the latent space is 16. This experiment is carried out using Neural Networks built with the help of Keras library with Tensorflow backend.

2.3 Model Training

The objective of training the VAE is to maximize the variational lower bound [8]. By introducing the condition c , the latent variable is now distributed under $P(z|c)$. This means that for each condition c , we obtain a distinct distribution $P(z)$ [7]. So, the loss function can be written as follows:

$$\mathcal{L} = \mathbb{E}[\log P(X|z, c)] - D_{KL}[Q(z|X, c)||P(z|c)] \quad (1)$$

Adam optimizer is chosen as an optimization algorithm. Due to the limited number of data points, we applied the following data augmentation techniques: rotation (max. ± 10), translation (0.1 height and width shift), zoom (0.1), and horizontal flip.

After training the model for 200 epochs, the model was used in two separate stages. We first used the model to reconstruct the encoded representation of the healthy aortic root images to evaluate the representational capability of the architecture for the given problem. The recovered images were then compared to the original input. In the second stage, we utilized the model for producing the healthy geometry of the aortic root based on the pathological root images. The dilated root images were passed to the encoder to compute their latent spaces. Then, the manipulation is done by forcing the decoder to reconstruct the desired geometry by switching the condition on the decoder to healthy. Thereby, the latent space is now reconstructed in the healthy root distribution.

2.4 Evaluation Methodology

To test the reliability of the model when encountering new unseen images, we applied 6-Fold cross-validation over the 24 valves, i.e. the pairs of healthy and dilated images.

The similarity between the generated images and the original input is measured with Mean Squared Error (MSE) and Structural Similarity Index (SSIM). However, using these metrics requires performing image registration as the generated and ground truth (healthy) images are not necessarily aligned. As we only care for the shape reconstruction and not the rotation of the shape, we performed intensity-based rigid image registration before comparing the images.

Furthermore, we evaluated other architectures that have different network depth. We also compared the model accuracy in dependency of the latent space dimensionality.

3 Results and Discussion

To evaluate the performance of our model, the experiments are conducted as follows. First, we examined the learning process of the CVAE to represent certain input in a lower-dimensional space so that the model is able to reconstruct the output from the reduced representation. Then, we generated aortic root images based on the described method. Afterwards, we checked the effect of changing the depth as well as the latent space size on the accuracy of the results. Finally, the results of this study were compared with those from other approaches.

Tab. 1: Similarity results between the reconstructed and input images based on MSE and SSIM metrics.

	MSE	SSIM
Training set	249 ± 14	0.881 ± 0.002
Test set	475 ± 231	0.83 ± 0.02

3.1 Representation Learning

Table 1 lists the similarity results between the healthy aortic root images propagated through the network and their recovered output. The given results are the average values over each set.

The choice of data representation plays a significant role in the success of the performance of machine learning algorithms. In other words, learning proper representations facilitates solving the task by extracting the most relevant and useful feature points to train the desired algorithm. Moreover, representation learning is heavily reliant on the training data. This can be seen from the drop in similarity values when predicting on the test set.

3.2 Aortic Root Reconstruction

Figure 2 shows three reconstructed aortic root images. For qualitative analysis, the images are compared with the ground truth and the pathological (dilated) states.

The changes in shape and size attributes can be clearly observed in the reconstructed images. The distorted shape caused by the dilation has disappeared. Instead, the images show smoother geometries without losing the anatomical properties of the aortic root. Despite splitting the dataset, the model still shows potentials in generating meaningful images when new unseen images are introduced. The similarity results between

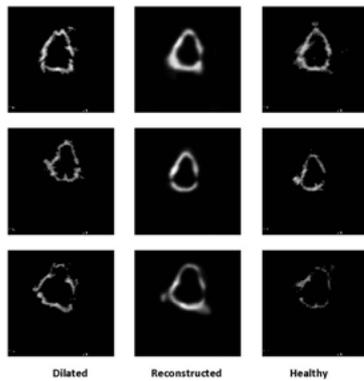


Fig. 2: Visualization of three reconstructed images compared with the dilated and healthy states. The first row is obtained from the training set, and the rest is from the test set.

Tab. 2: Quantitative results for the similarity between the generated aortic root and the ground truth based on MSE and SSIM metrics.

	MSE	SSIM
Pathological vs. Healthy	1566	0.766
Reconstruction vs. Healthy (Training set)	1039 ± 129	0.801 ± 0.009
Reconstruction vs. Healthy (Test set)	915 ± 448	0.79 ± 0.03

the predicted aortic root images and the ground truth healthy valves images are recorded in table 2.

The achieved results are far from the optimum. However, based on the results between the training and testing sets, the model is still able to maintain a good generalization when provided with new images.

3.3 Evaluation of Various Architectures

Figure 3 demonstrates the relationship between the depth and structural similarity index for different latent space sizes. Four architectures have been used in this comparison. A1 represents a network with 3 convolutional layers in the encoder, A2 has 4 convolutional layers, A3 (the proposed architecture) has 5 convolutional layers, and A4 has 6 convolutional layers. The decoder is the mirrored architecture of each relevant encoder. Increasing the network depth does not seem to improve the accuracy of reconstructing the aortic root. With a small dataset size, taking into account the increase in the number of nodes with depth, adding more layers would only contribute to a higher number of hyperparameters; consequently, the model is more likely to overfit. On the other hand, using shallow architectures, for example, three convolutional layers in the encoder and similarly for the decoder, resulted in the lowest similarity index values compared to other depth choices. In terms of latent space size, using eight nodes was not sufficient to capture

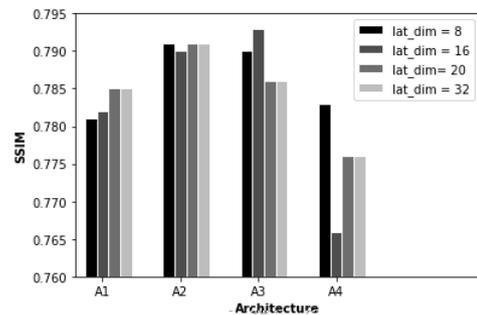


Fig. 3: Comparison between various architectures with different latent space sizes

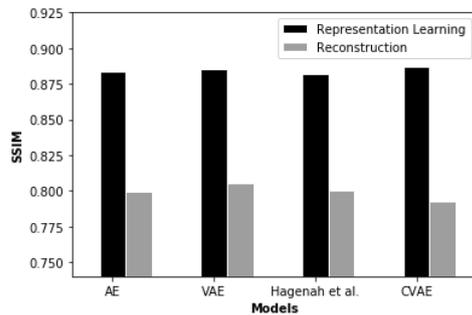


Fig. 4: Bar graph depicting the SSIM values of four different models based on their performance in representation learning and reconstructing the healthy aortic root.

the features necessary to provide optimal input retrieval. By fixing the latent space size to 16 neurons as in the proposed architecture, the model provides the highest similarity among other options.

3.4 Comparison to State-of-the-Art Models

The results of this study have also been compared with previous related works. Figure 4 displays a bar graph comparing the SSIM values of four different models when performing the above tasks. The first model uses a standard autoencoder AE with convolutional layers. The second one represents a standard variational autoencoder VAE both with a manually designed translation vector in the latent space. The next model is the one proposed by Hagenah et al. [5]. The last model is our proposed model CVAE.

The primary downside with standard autoencoders is the noncontinuous distribution of the latent space, which makes it challenging to generate new realistic images when drawing new samples from the latent space. Nevertheless, in variational autoencoder, new images are sampled from a Gaussian distribution over the latent space, therefore, resulting in smooth images. Despite that CVAE does not show ideal performance in reconstructing the healthy aortic root, the important point is that the reached accuracy is comparable to the previously published method without any kind of manual deformation modeling. These results indicate that a CVAE is capable of learning the deformation completely data-driven and no human bias has to be introduced.

4 Conclusion

In this study, we proposed a novel, end-to-end automatized approach for reconstructing the healthy aortic root using a Conditional Variational Autoencoder. The proposed model showed strong performance in generating new images based on changing the label of the latent space of the dilated images. Further experiments could employ transfer learning to address the problem with the small dataset.

Author Statement

Research funding: The author state no funding involved. **Conflict of interest:** Authors state no conflict of interest. **Ethical approval:** The only dataset used in this study was collected and published previously, so no ethical conflicts are present.

References

- [1] V. L. Gott, et al., *Replacement of the Aortic Root in Patients with Marfan's Syndrome*. vol. 340, no. 17, pp. 1307–1313, 1999.
- [2] K. J. Zehr, M. J. Thubrikar, G. G. Gong, J. R. Headrick, and F. Robicsek, *Clinical introduction of a novel prosthesis for valve-preserving aortic root reconstruction for annuloaortic ectasia*. *J. Thorac. Cardiovasc. Surg.*, vol. 120, no. 4, pp. 692–698, 2000.
- [3] T. Kuniyama et al., *Preoperative aortic root geometry and postoperative cusp configuration primarily determine long-term outcome after valve-preserving aortic root repair*. *J. Thorac. Cardiovasc. Surg.*, vol. 143, no. 6, pp. 1389–1395.e1, 2012.
- [4] V. Paruchuri et al., *Aortic Size Distribution in the General Population: Explaining the Size Paradox in Aortic Dissection*. *Cardiol.*, vol. 131, no. 4, pp. 265–272, 2015.
- [5] J. Hagenah, M. Mehdi, F. Ernst, *Generating Healthy Aortic Root Geometries from Ultrasound Images of the Individual Pathological Morphology using Deep Convolutional Autoencoders*. In *computing in cardiology*, vol. 46, pp. 1–4, 2019.
- [6] J. Hagenah, E. Werrmann, M. Scharfschwerdt, F. Ernst, and C. Metzner, *Prediction of individual prosthesis size for valve-sparing aortic root reconstruction based on geometric features*. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2016–Octob, pp. 3273–3276, 2016.
- [7] K. Sohn, X. Yan, and H. Lee, *Learning structured output representation using deep conditional generative models*. In *Advances in Neural Information Processing Systems*, vol. 2015–Janua, pp. 3483–3491, 2015.
- [8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press 2016.