7th International Conference on Computational and Mathematical Biomedical Engineering - CMBE2021
27–29 June 2022, Italy
P. Nithiarasu and C. Vergara (Eds.)

GENERATION OF VIRTUAL PATIENT'S AORTA ANATOMY USING CONVOLUTIONAL NEURAL NETWORKS

Álvaro López[1], Alfonso Domínguez[1], Pau Romero[2], Miguel Lozano[2], Pablo Lamata[3], and Ignacio García-Fernández[2]

[1]Facultat de Matemàtiques, Universitat de València, Spain, {loloal5,aldodela}@alumni.uv.es
[2]CoMMLab, Dept. d'Informàtica, ETSE - Universitat de València, Spain,
{pau.romero,miguel.lozano,ignacio.garcia}@uv.es
[3]Department of Biomedical Engineering, School of Biomedical Engineering and Imaging Sciences, Kings College London, UK, pablo.lamata@kcl.ac.uk

SUMMARY

The methodology to generate virtual patient cohorts can strongly influence the efficiency of the process and the quality of the results. The use of low dimensional representations of anatomy is a common strategy, but by representing a case in a latent space we can lose relevant spatial information. We explore the utility of Convolutional Neural Networks, combined with an appropriate geometry representation, to exploit the spatial information when generating synthetic thoracic aorta anatomies.

Key words: *Virtual patient cohorts, anatomy, Convolutional Neural Networks, Generative Adversarial Networks, digital twin.*

1 INTRODUCTION

The development of computational models, together with the use of personalised clinical data, has become a powerful aid in cardiology. When applied to cohorts of patients it enables the conduction of *in-silico* trials and population analyses. The combination of the computational models of the cardiovascular system together with machine learning (ML) techniques brings access to predicting relevant information, such as therapy outcomes or risks assessment, from the available clinical data [1, 2]. However, the access to these clinical data such as Computer Tomography Scans or Magnetic Resonance Images is often limited due to the inherent complexity of the process. This, together with the need of large data sets to train most ML models, hinders their application in many real clinical contexts.

In order to overcome this problem different authors have proposed the use of cohorts of synthesised virtual patients [3, 4, 5]. This approach consists in generating virtual patients by sampling an observed distribution of real cases to define the anatomy and the biophysical properties of the synthetic digital twins [3]. I the case of the anatomy, the generation of the patient's phenotype is typically done using Statistical Shape Modelling [4, 6, 7], although ML methods such as Generative Adversarial Networks are gaining popularity in the last few years [5, 8].

In this work we continue the previous research by Romero et al. [5] in which we built large cohorts of synthetic thoracic aorta anatomies using, among other methods, Generative Adversarial Networks (GAN). In that research, the shape of the aortas was projected onto a low dimensional space by means of a Principal Component Analysis (PCA), before training the GAN. In this projection, based on a linear transformation, deformation modes tend to affect the whole shape of the organ, possibly losing the local nature of some relevant anatomical features. Our hypothesis is that, by using Convolutional Neural Networks (CNN), these local anatomical features can be identified non linearly during the training process. In this paper we will use the wall and centerline information, represented as a texture, to train the network using input convolutional layers, and present preliminary comparisons with respect to a PCA based training.

2 MATERIAL AND METHODS

The training dataset is formed by a retrospective population of 34 thoracic aortas scanned with CT from a cohort of patients in ages between 78 and 89 years old with ascending aortic aneurysm. The input data used in this study was a set anonymised triangular surface meshes representing the walls of the aortas.

2.1 Preprocessing of the aorta wall

To use a common representation of all the aortas we follow the method proposed by Romero et al. [9]. For each wall mesh the centerline is extracted and approximated as a differentiable parametric curve $\alpha(s)$. The aorta wall is represented as a differentiable surface that can be described by a generalised set of cylindrical coordinates: the parameter s of the closes point of the centerline, an angle θ with respect to a reference direction (orthogonal to the centerline at s, and a radius $r(s, \theta)$, which represents the distance of the corresponding wall point from the centerline. Thus, given an aorta wall, our representation of the aorta can be described by means of the centerline curve $\alpha(s)$ and the map of wall distances to the centerline $r(s, \theta)$.

To discretize this representation we first fit the centerline by an order 3 B-Spline curve and the distance map by an order 3 bivariate B-Spline surface. For a given number of B-Spline knots, fixed beforehand, we get $3 \times n_c$ coefficients for the centerline and $n_{r\theta} \times n_{rs}$ coefficients for the surface. Note that both n_{rs} and n_c represent the degrees of freedom of our representation in the longitudinal direction of the aorta. If we take $n_c = n_{rs}$, then we can store the complete set of coefficients of the aorta approximation in a matrix A of dimensions $n_{r\theta} + 3 \times n_{rs}$. The set of all matrices is normalised, so that all the values are in the range $[0, 1]$. Since the centerline and the wall values are in a very different range, they are normalised separately. Figure 1 shows the different steps of this procedure. In the resulting 2D image, the aortic root is placed at the left hand side. The first three rows correspond to the centerline coefficients, and the rest of the rows correspond to the coefficients of the wall distance to the centre, in the angular coordinate.

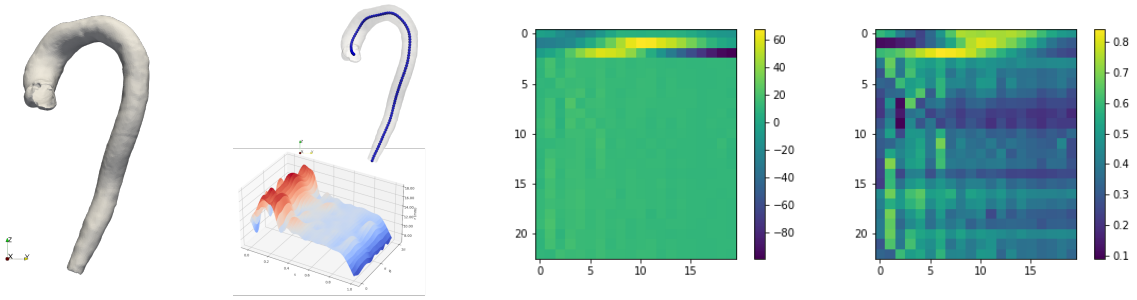

Figure 1: All the aortas are transformed to a 2D image. From left to right, the wall is decomposed onto a centerline curve and a distance function from the wall to the centerline; the curve and the function are approximated by cubic B-Splines, and the coefficients are arranged as a matrix that can be interpreted as an image; the data set is finally normalised.

2.2 Network training

The dataset formed by the 34 normalised textures is used to train a GAN. In our experiments, we have used polynomials with 20×20 coefficients for the wall, leading to images of size 23×20. The generator has a dense layer with 256 neurons followed by a combination of RelU and convolutional layers to make a total of 5 hidden layers. The output layer has the desired 23×20 size. The hidden layers of the discriminator are also a combination of convolutional and RelU layers, to output with a single neuron dense layer to make a total of 4 hidden layers. The size of the convolution kernel is 5×5. We have used Binary Crossentropy validation with random splitting of the dataset in two groups (batch size of 17) and $100,000$ training epochs.

Once the network was trained, we have generated 1000 images that represent 1000 synthetic aortas. After reconstructing the wall we have measured several anatomic features and compared the distribution of the resulting virtual population with that of the real aortas cohort. We also compare the results to a population that was generated using a GAN after applying a PCA to the set of polynomial coefficients in [5]. The measured anatomic features are the radius of the sinuses of Valsalva (SoV), the radius of the aorta at the sinotubular junction (PA), the width of the aortic arch (W) and the height of the aortic arch (h). The detailed definition of such biomarkers is the one described in [5].

3 RESULTS AND CONCLUSIONS

Figure 2 shows four synthetic aortas generated with the trained GAN. We observe that the overall generated shape corresponds to a plausible aorta anatomy. There are, however, some samples that show a rather extreme shape, such as the rightmost aorta in the figure. When generating a synthetic database this is not an issue, provided that we have criteria to disregard the unfeasible shapes [3, 5], but the frequency of unfeasible samples will, however, determine the efficiency of the process [5]. For an objective quantification of the results, we have compared the geometric features of the reconstructed anatomies in the synthetic population to the values observed in the original population. Figure 3 shows an approximation of the probability density functions fitted from the data.

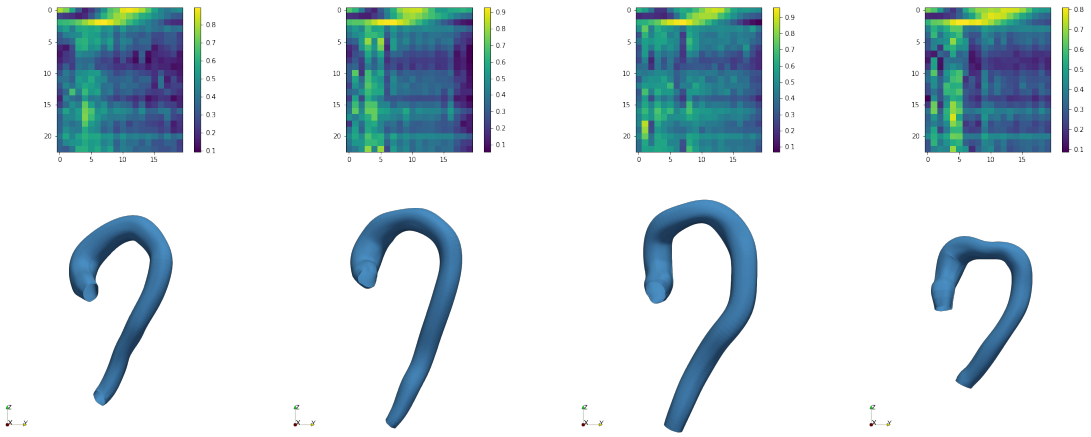

Figure 2: Four synthetic aortas generated using a GAN. For each sample the generated image is shown in the top row and the corresponding reconstructed wall is presented below it.

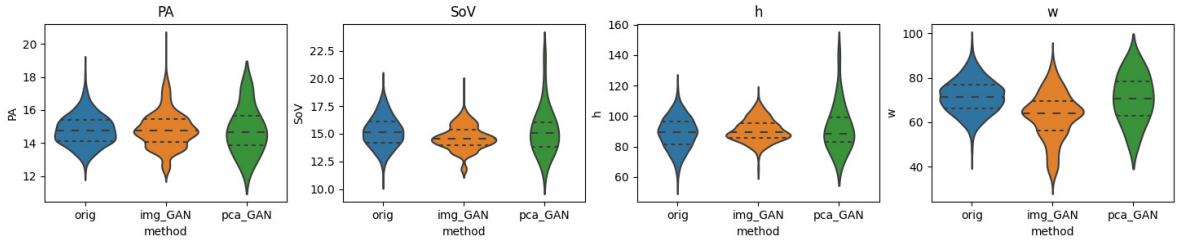

Figure 3: Distribution of some geometric measurements on the generated aortas. The measured features are the radius of the sinuses of Valsalva (SoV), the radius of the aorta at the sinotubular junction (PA), the width of the aortic arch (w) and the height of the aortic arch (h). For each feature, a violin plot is presented for the observed distribution of the training set (orig), the trained GAN (img_GAN) and the GAN trained after a PCA (pca_GAN) from [5]. Distances are in millimetres.

The results indicate that the synthetic population presents a reasonable variability variability in the observed biomarkers; the range of the sinotubular junction radius (PA) and of the arch width (w) in the synthetic population are comparable to those observed in the original cohort, and the other two, while narrower, still span over most of the original range. In the case of the two radii (sinuses of Valsalva (SoV) and sinotubular junction (PA)), the synthetic population could show a bimodal distribution which is not present in the other two approximations. This could be explained as a result

of a possible over-fitting, due to the reduced size of the training set, that would make the synthetic cohort concentrate around a reduced set of principal shapes. These preliminary results do not shed light yet over the relevance of using the convolutional layers. To assess the impact of over-fitting, we intend to repeat the training with a larger dataset that was generated in [5] using bootstrapping. This would lead to a more fair comparison, since the GAN trained therein used this dataset. In addition, a training with non-convolutional GAN has to be done on the current representation (without the PCA) in order to asses the importance of using convolutional layers. In [5] we also observed that the linear projection provided by the PCA did not help in problems such as classification or prediction of the anatomy phenotype. We expect that the non linear nature of CNN will make it easier to process the anatomical patterns present in the population, and to project them onto a latent space in which classification or regression are easier to resolve.

REFERENCES

[1] Liang Liang, Minliang Liu, Caitlin Martin, John A. Elefteriades, and Wei Sun. A machine learning approach to investigate the relationship between shape features and numerically predicted risk of ascending aortic aneurysm. *Biomechanics and Modeling in Mechanobiology*, 16(5):1519–1533, Oct 2017.

[2] Jorge Corral-Acero, Francesca Margara, Maciej Marciniak, Cristobal Rodero, Filip Loncaric, et al. The 'digital twin' to enable the vision of precision cardiology. *Eur Heart J*, 41(48):4556–4564, 12 2020.

[3] S. A. Niederer, Y. Aboelkassem, C. D. Cantwell, C. Corrado, S. Coveney, E. M. Cherry, T. Delhaas, F. H. Fenton, A. V. Panfilov, P. Pathmanathan, G. Plank, M. Riabiz, C. H. Roney, R. W. dos Santos, and L. Wang. Creation and application of virtual patient cohorts of heart models. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378(2173):20190558, 2020.

[4] B Thamsen, P Yevtushenko, L Gundelwein, A A A Setio, H Lamecker, M Kelm, M Schafstedde, T Heimann, T Kuehne, and L Goubergrits. Synthetic database of aortic morphometry and hemodynamics: Overcoming medical imaging data availability. *IEEE Trans Med Imaging*, 40(5):1438–1449, May 2021.

[5] Pau Romero, Miguel Lozano, Francisco Martínez-Gil, Dolors Serra, Rafael Sebastián, Pablo Lamata, and Ignacio García-Fernández. Clinically-driven virtual patient cohorts generation: An application to aorta. *Frontiers in Physiology*, 12, 2021.

[6] Liang Liang, Minliang Liu, Caitlin Martin, and Wei Sun. A deep learning approach to estimate stress distribution: a fast and accurate surrogate of finite-element analysis. *Journal of The Royal Society Interface*, 15(138):20170844, 2018.

[7] Cristobal Rodero, Marina Strocchi, Maciej Marciniak, Stefano Longobardi, John Whitaker, Mark D. O'Neill, Karli Gillette, Christoph Augustin, Gernot Plank, Edward J. Vigmond, Pablo Lamata, and Steven A. Niederer. Linking statistical shape models and simulated function in the healthy adult human heart. *PLOS Computational Biology*, 17(4):1–28, 04 2021.

[8] Andrew Gilbert, Maciej Marciniak, Cristobal Rodero, Pablo Lamata, Eigil Samset, and Kristin McLeod. Generating synthetic labeled data from existing anatomical models: An example with echocardiography segmentation. *IEEE Transactions on Medical Imaging*, pages 1–1, 2021.

[9] Pau Romero, Samuel Santos, Rafael Sebastian, Francisco Martinez-Gil, Dolors Serra, Pilar Calvillo, Alejandro Rodríguez, Rebeca Maldonado Puig, Luis Martí-Bonmatí, Angel Alberich-Bayarri, Miguel Lozano, and Ignacio García-Fernández. Reconstruction of the aorta geometry using canal surfaces. In *International Conference on Computational and Mathematical Biomedical Engineering, CMBE*, 2019.