Spatio-temporal neural distance fields for conditional generative modeling of the heart

Kristine Sørensen[1], Paula Diez[1], Jan Margeta[2], Yasmin El Youssef[3], Michael Pham[3], Jonas Jalili Pedersen[3], Tobias Kühl[3,4], Ole de Backer[3], Klaus Kofoed[3], Oscar Camara[5], and Rasmus Paulsen[1]

[1] DTU Compute, Technical University of Denmark, Kongens Lyngby, Denmark
[2] KardioMe, Research & Development, Nova Dubnica, Slovakia
[3] Heart center, Rigshospitalet, Copenhagen, Denmark
[4] Dep. of Cardiology, Zealand University Hospital, Denmark
[5] Physense, BCN MedTech, Universitat Pompeu Fabra, Barcelona, Spain
kajul@dtu.dk

Abstract. The rhythmic pumping motion of the heart stands as a cornerstone in life, as it circulates blood to the entire human body through a series of carefully timed contractions of the individual chambers. Changes in the size, shape and movement of the chambers can be important markers for cardiac disease and modeling this in relation to clinical demography or disease is therefore of interest. Existing methods for spatio-temporal modeling of the human heart require shape correspondence over time or suffer from large memory requirements, making it difficult to use for complex anatomies. We introduce a novel conditional generative model, where the shape and movement is modeled implicitly in the form of a spatio-temporal neural distance field and conditioned on clinical demography. The model is based on an auto-decoder architecture and aims to disentangle the individual variations from that related to the clinical demography. It is tested on the left atrium (including the left atrial appendage), where it outperforms current state-of-the-art methods for anatomical sequence completion and generates synthetic sequences that realistically mimics the shape and motion of the real left atrium. In practice, this means we can infer functional measurements from a static image, generate synthetic populations with specified demography or disease and investigate how non-imaging clinical data effect the shape and motion of cardiac anatomies.

Keywords: Neural Implicit Functions · Spatio-Temporal Representations · Multi-Modal Inputs · Cardiac Anatomy

1 Introduction

During a heart beat, the heart chambers undergo a series of complex 3D deformations and the morphology as well as the motion of contraction and relaxation are critical functions to pump blood to all human organs. Such shape and motion can be captured by Computed Tomography (CT), where Cardiac Functional

Analysis (CFA) has enabled fast acquisition with 20 (or more) time frames per heartbeat. The cardiac anatomy and motion vary widely between individuals and it is therefore of interest to disentangle the individual characteristics from that of clinical factors such as gender, age or disease. Traditional shape modeling based on shape correspondence and point distribution models [9] can be used for generative modeling, but the integration of clinical data in such models is currently an unsolved problem. Modeling the spatio-temporal characteristics of the heart has been achieved through 4D registration methods [21] or by building spatio-temporal atlases [13,16]. More recently, deep learning based methods have been used to learn the distribution over plausible cardiac shapes based on dense point clouds [20,5], meshes [6] and voxel volumes [7]. Modeling the temporal movement is usually approached by mapping the anatomy at one time frame to another time frame (ie. end-diastole (ED) to end-systole (ES) or vice versa) and has been handled using i.e. mesh U-nets [4], spatio-temporal graph convolutions [15] or adversarial methods in the image domain [17]. Generative frameworks have incorporated non-imaging clinical data with methods for generating cardiac ultrasound images with specified functional properties [24] and Magnetic Resonance Images (MRI) with specified anatomical characteristics [2], pathology [10] or biophysical parameters [22]. Generation of spatio-temporal cardiac anatomies based on clinical demography has been approached by [23], who learned a temporal latent space based on a recurrent neural network and generated cardiac anatomies represented by voxelized labelmaps. The explicit voxelmap representation however suffer from large memory requirements and are not ideal for representing the smooth and often highly detailed cardiac anatomy. Implicit representations (such as distance fields) have shown to be an effective representation of complex shapes for medical image segmentation [27,1,28] as well as shape generation in 3D [18,8] and 4D [11]. In contrast to other representations, the continuous nature of neural distance fields allows for modeling highly complex structures without requiring correspondence between samples or have memory requirements growing cubically with the image resolution.

We propose a novel conditional generative model architecture, that integrates clinical demographic data into the generation of spatio-temporal signed distance fields (SDFs) of dynamic anatomies. The model is tested for modeling the left atrium (LA) and its complex extrusion; the left atrial appendage (LAA). The shape and motion of the LA and LAA are related to the risk of thrombus formation and stroke [12] and a detailed representation of the spatio-temporal dynamics are therefore of interest. Capturing the small details of the LAA anatomy would require a very large voxel map and obtaining point correspondence across the widely varying samples are unattainable [26]. The use of a neural SDF representation allows for modeling complex surfaces jointly in space and time, while the auto-decoder formulation enables the derivation of two separate latent spaces that disentangles the patterns related to the clinical demography from those related to individual variation. The proposed method are used to complete the full cardiac cycle based on a single time frame and to generate realistic synthetic anatomical sequences with specified clinical demography.

2 Materials and methods

Overall framework[6]

A spatio-temporal neural distance field is a neural function that maps an arbitrary space-time coordinate $\mathbf{x} = (\mathbf{p}, t)$, consisting of the 3D coordinate $\mathbf{p}$ and the time index t, to the signed distance $\hat{d} = f_\theta(\mathbf{x})$. The surface represented by the neural SDF can thus be considered the decision boundary separating SDF<0 from SDF>0. To learn such decision boundary for a single sequence, the network is presented with a set X consisting of space-time coordinates $\mathbf{x}$ and corresponding SDF-values d, such that $X := \{(\mathbf{x}, d) : \mathrm{SDF}(\mathbf{x}) = d\}$. The network parameters θ are optimized using the clamped L1-loss between the true and predicted SDF-values in this set:

$$\mathcal{L}(f_\theta(\mathbf{x}), d) = |\mathrm{clamp}(f_\theta(\mathbf{x}), \delta) - \mathrm{clamp}(d, \delta)| \tag{1}$$

where the clamping parameter δ controls the distance from the surface, over which we expect to learn an accurate distance field. The continuous spatio-temporal SDF can then be approximated as $SDF(\mathbf{S}) \approx f_\theta(\mathbf{S})$, where $\mathbf{S}$ refers to all possible space-time coordinates.

To model multiple anatomical sequences with the same neural network, we introduce a latent vector to represent each of the anatomical sequences indexed by $n = \{1, ..., N\}$. This latent vector is concatenated with the space-time coordinate and given as input to f_θ, where it enables the neural function to create decision boundaries unique to each anatomical sequence. We propose to learn this latent vector as two separate parts: the clinical demographic latent vector $\mathbf{z}_c$ representing an embedding of the clinical demography (gender, age and SBP) and a residual latent vector $\mathbf{z}_r$ representing the individual information that cannot be described by the clinical demography. $\mathbf{z}_c$ is embedded from the clinical demography $\mathbf{c}$ with a neural network such that $\mathbf{z}_c = g_\phi(\mathbf{c})$. The residual latent vector $\mathbf{z}_r$ is a learnable embedding unique to each of the sequences in the training set $(\mathbf{z}_{r,1}, \mathbf{z}_{r,2}, ..., \mathbf{z}_{r,N})$ and is optimized jointly with f_θ and g_ϕ using an auto-decoder formulation. The use of an auto-decoder circumvents the need for designing and training a 3D data encoder, but requires test time optimization which is slightly more time consuming and may risk converging to a local solution. To optimize the autodecoder parameters, we sample $K = 110.000$ space-time coordinates for the 20 time frames ($t = \{0\%, 5\%, ..., 95\%\}$ of the cardiac cycle) in each of the $N = 290$ anatomical sequences in the training set and denote each of these samples $\mathbf{s}_{n,k,t}$ and the corresponding SDF value $d_{n,k,t}$. The loss over the predicted and measured SDF-values across all N · T · K samples are computed as:

$$\underset{\theta, \phi, \{\mathbf{z}_{r,n}\}_{n=1}^{N}}{\arg\min} \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{t=0\%}^{95\%} \mathcal{L}(f_\theta(g_\phi(\mathbf{c}_n) \oplus \mathbf{z}_{r,n} \oplus \mathbf{s}_{n,k,t}), d_{n,k,t}) + \frac{1}{\sigma^2} ||\mathbf{z}_{r,n}||_2^2, \tag{2}$$

[6] Code available at https://github.com/kristineaajuhl/spatio_temporal_generative_cardiac_model.git

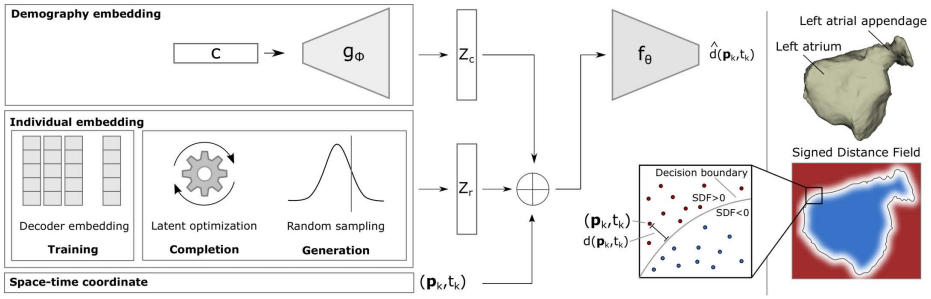

Fig. 1. The signed distance field approximates the surface as the decision boundary separating space-time coordinates $(\mathbf{p}_k, t_k)$ that are inside and outside the surface. For each $\mathbf{p}_k, t_k$ the signed distance $\hat{d}$ to the surface is predicted with the network f_θ based on a concatenation ($\oplus$) of the clinical demography latent vector $\mathbf{z}_c$, the individual latent vector $\mathbf{z}_r$ and the coordinate. $\mathbf{z}_c$ is embedded from the clinical demography encoder g_ϕ, whereas the source of $\mathbf{z}_r$ depends on the task. **Training:** Each training sample is assigned a learnable embedding $\mathbf{z}_r$ which is optimized jointly with g_ϕ and f_θ. **Reconstruction:** The individual embedding is learned by locking the parameters of f_θ and g_ϕ and optimize for $\mathbf{z}_r$. **Generation:** A new $\mathbf{z}_r$ is generated by sampling from a multivariate Gaussian distribution.

where the loss $\mathcal{L}$ is the clamped L1-loss from Equation 1, whereas $||\mathbf{z}_{r,n}||_2^2$ is a regularization term that encourages compact latent spaces, and σ^2 balances the L1-loss and the regularization. Figure 1 shows an overview of how we propose to integrate neural SDFs in a conditional generative framework and visualizes an example of how the left atrial shape is encoded as the decision boundary of the neural network based on point-distance samples.

Sequence completion — The proposed model can complete an anatomical sequence based on the clinical demography $\mathbf{c}$ and the static anatomy at a given time frame t_{given}. The clinical demography are mapped to the clinical demographic latent space as $\mathbf{z}_c = g_\phi(\mathbf{c})$, whereas the individual latent vector $\mathbf{z}_r$ is found using test-time optimization. We sample $K = 110.000$ space coordinates $(\mathbf{s}_{k|t_{\text{given}}})$, measure the distances to the static anatomy at t_{given} and optimize over the values in $\mathbf{z}_r$ while locking the parameters of f_θ and g_ϕ:

$$\hat{\mathbf{z}}_r = \arg\min_{\mathbf{z}_r} \sum_{k=1}^{K} \mathcal{L}(f_\theta(g_\phi(\mathbf{c}) \oplus \mathbf{z}_r \oplus \mathbf{s}_{k|t_{\text{given}}}), d_{k|t_{\text{given}}}) + \frac{1}{\sigma^2}||\mathbf{z}_r||_2^2, \qquad (3)$$

The full anatomical sequence can be reconstructed by evaluating the SDF at all points on a uniform grid for the desired time frames and extract the zero-level isosurface using Marching Cubes [14].

Sequence generation — The model can generate new plausible anatomical sequences by sampling randomly in the individual latent space $\mathbf{z}_r$ and concatenate it with the embedding of the clinical demography. In the derivation of the auto-decoder formulation of neural distance fields (see [18]) the prior distribution

over the latent variables $\mathbf{z}_r$ is assumed to come from a zero-mean multivariate Gaussian distribution. To generate new sequences with individual shape and motion characteristics, we sample $\mathbf{z}_r$ from such distribution.

Implementation details

The model was implemented with PyTorch [19] and trained on a NVIDIA RTX A4000 (16GB) for 9 hours. The clinical demographic input is passed as a vector with one-hot encoded gender (male/female) and age-group (<50,50-59,60-69 and >69) and uses the continuous SBP normalized to $[0; 1]$. Both f_θ and g_ϕ are implemented as a multi-layer perceptron (MLP). f_θ follows the architecture from [18], whereas g_ϕ consist of two hidden layers with each 128 neurons. We use 64 dimensions for both $\mathbf{z}_c$ and $\mathbf{z}_r$. During training we randomly dropout $\mathbf{z}_r$ in 20% of the training steps, which was found necessary to force the network to make use of the clinical demography. All surfaces are extracted from an SDF evaluated on an uniform grid sized 128^3.

Data and preprocessing

A dataset of 4D geometries was extracted from CFA scans of 667 randomly selected participants from the Copenhagen General Population Study. All subjects are asymptomatic individuals from the general population, with 301/366 male/female participants, aged 41-89 years and with known systolic blood pressure (SBP). Participation was conducted following the declaration of Helsinki and approved by the ethical committee (H-KF-01-144/01). Each CFA series consists of a cardiac computed tomography angiography (CCTA) image at 20 equally spaced time steps during one heart beat ($t = 0\%, 5\%, ..., 95\%$). Each image was segmented using an automatic deep learning based segmentation method specifically developed for LAA segmentation from CT images [27]. We aligned all anatomies at $t = 0\%$ by matching the center-of-mass, fine-tuned with a rigid transformation based on iterative closest point (ICP)[3] and applied the found transformation to all time frames. Finally, each anatomy was scaled with a fixed factor such that all surfaces lie within the unit-sphere. We follow the sampling strategy from [27] and sample 10.000 random samples within the unit sphere and 100.000 samples in the vicinity of the surface based on the shape diameter at each vertex for each time frame. All transformations and point-to-surface distances were obtained using VTK [25]. We split the dataset randomly with 290/10/367 anatomical sequences for training/validation/testing.

3 Experiments and results

We evaluate the proposed methods on two different tasks - *anatomical sequence completion* based on a single static anatomy as well as the clinical demography and *anatomical sequence generation* based only on the clinical demography.

Sequence completion — Evaluating the sequence completion abilities of a generative model assure us that the model has learned a descriptive distribution of both shape and movement from the sequences in the training data. We report results for completion based on the static anatomy at $t = 0\%$ as this allows for comparison to [23]. We evaluate the completion quality using symmet-

ric Chamfer Distance (CD) and Hausdorff Distance (HD) between the predicted and true surfaces at each time step. The ability to capture the dynamic movement is measured by comparing the maximum volume ($V_{\max}$), Fractional Change ($FC = (V_{\max} - V_{\min})/V_{\max}$)) and Cyclic Change $CC = V_{\max} - V_{\min}$ between the true and completed sequences. We have compared against the default CHeart implementation [23] trained on our dataset as well as investigated the effect of removing the clinical demography from our model, by setting the clinical demography vector $\mathbf{c}$ to a zero-vector for all samples during training and testing. The results are seen in Table 1. We observe that the proposed method significantly outperforms CHeart [23], which is attributed the joint spatial and temporal modeling and the increased representational power from the SDF. An ablation study on the clinical demography, demonstrate a positive effect on estimating the functional parameters (FC, CC and V_{max}) even though the reconstruction errors do not show an improvement.

Figure 2 shows the completion results from three different test cases. The blue volume curve is an example of a normal atrial function consisting of a relaxation phase ($t = 0\%$ to first peak), a passive emptying phase (first peak to plateau at $t \approx 75\%$) and an active emptying phase (plateau to $t = 95\%$). The red curve is an example of an abnormal atrial motion dominated by passive emptying. We observe that the proposed model correctly completes both of these sequences, which indicates that the model are able to learn individual motion patterns.

Sequence generation — The proposed method can be used to generate unique sequences with specified clinical demography. Figure 4(left) shows four synthetically created samples based on the same clinical demography. We observe that all generated anatomies are continuous and smooth surfaces, with volumes indicating normal atrial motion. It is evident that multiple plausible LA and LAA geometries and motions exist despite being generated from the same clinical demography. This diversity is attributed to factors not included in the model (ie. height, weight, smoking status, etc.), but also individual traits that cannot be directly related to demography or disease. To evaluate the models ability to generate realistic cohorts, we generate a synthetic counterpart to the test set, where a synthetic anatomical sequence is generated based on the clinical demography of each person in the test set. Since multiple plausible sequences can be generated based on the same clinical demography, we do not compare the

Table 1. Sequence completion evaluation measuring the Chamfer distance (CD), Hausdorff distance (HD), Maximum volume ($V_{\max}$), Fractional Change (FC) and Cyclic Change (CC). For each test sample we collect the time instance with the minimum, average and maximum CD and HD and report the average across all test samples for each of these. Bold indicates best evaluation.

Method	CD [mm]			HD [mm]			Difference [%]		
	Min	Mean	Max	Min	Mean	Max	$V_{\max}$	FC	CC
CHeart [23]	2.80	4.13	6.23	9.59	16.79	54.16	4.41%	8.26%	9.65%
Ours (÷ demography)	**1.71**	**2.88**	4.26	**6.75**	11.64	17.08	3.90%	5.64%	9.28%
Ours	1.93	2.89	**4.08**	6.94	**11.49**	**16.56**	**3.32**%	**4.59**%	**7.83**%

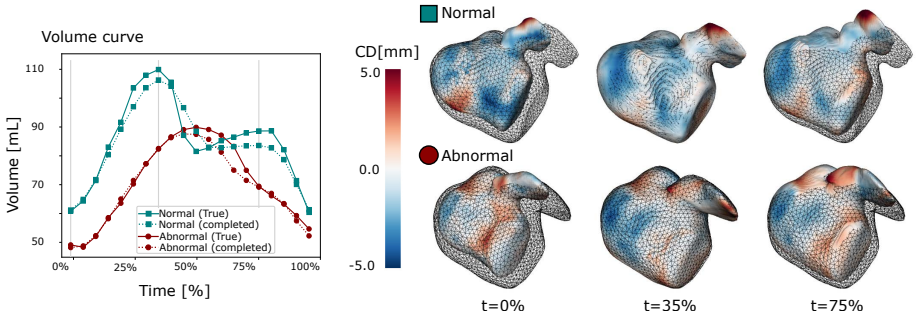

Fig. 2. Two examples of completed sequences illustrated with volume curves and the chamfer distance (CD) between the predicted and true surface at time frame 0%,35% and 75% (coloured surfaces) as well as surface with maximum volume (wireframe). The blue (□) correspond to the $25th$ percentile evaluated on average CD, whereas the red (◯) shows an abnormal atrial motion without an active emptying phase.

generated surfaces directly to those from the test set. Instead, Figure 3 shows the distribution of measured FC in both the real and synthetic data in eight subgroups split on age and gender. Both the real and the generated data exhibit similar trends; the FC decreases with age and are higher in women compared to men. The longer tails in the real data are attributed outliers errors made by the automatic image segmentation, where the volume is over- or underestimated in one or more time frames.

Demography manipulation — We sample a single individual vector $\mathbf{z}_r$ and investigate the effect of changing the gender and age in the clinical demography input. Figure 4(right) shows how changing the demography alters the anatomical sequence. It can be noted that the shared $\mathbf{z}_r$ ensures a relatively

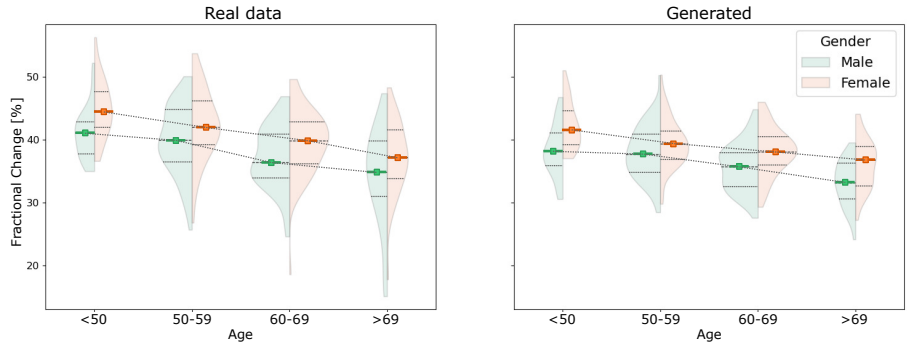

Fig. 3. Distribution of left atrial fractional change (FC) across the subgroups based on gender (male/female) and age (< 50,50-59,60-69 and > 69) in the population from the test set (left) and a synthetic population generated with the same clinical demography as the population in the test set (right).

stable overall anatomy across all samples, but that associations are captured such that changing the gender from male to female for example results in a smaller LA volume. The blue/pink opaque circles in the Principle Component Analysis (PCA) plots in Figure 4 show the embeddings of the training data for male/female participants. We observe a clear gender separation in clinical demographic space, whereas the individual latent space shows no obvious gender separation. This supports our idea of $\mathbf{z}_r$ as a residual latent vector and that the two spaces can be sampled independently.

4 Discussion and conclusion

We have presented a conditional generative model, that is capable of creating plausible temporal sequences of cardiac anatomy while taking patient demography into account. The neural SDF models the cardiac anatomy and movement jointly, and allows for generating smooth and detailed anatomical sequences with close resemblance to real anatomies. We showed that the proposed model outperforms a current state-of-the-art method for anatomical sequence completion and that it can be used to estimate functional parameters from the large pool of CT protocols, where only one or two best-phase images are acquired. In contrast to previous work, the versatility of the neural SDF representation allows for

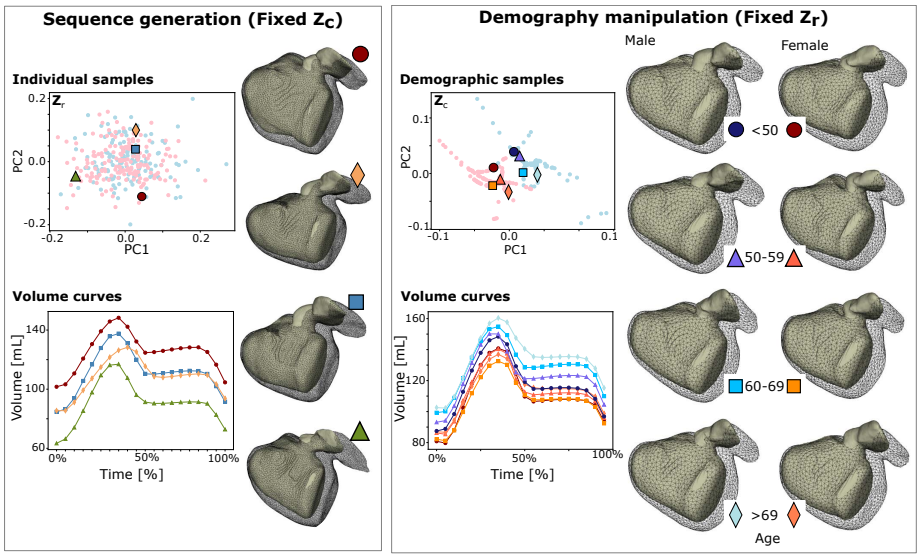

Fig. 4. Synthetically created samples. Left: fixed clinical demography (50-59 years old, Male, systolic blood pressure equal to 130 mmHg) and sampled $\mathbf{z}_r$. Right: fixed individual latent and varying $\mathbf{z}_c$. The figure show the first two principle components (PC) of the latent spaces, the volume curves for all samples as well as the generated anatomies at $t = \%0$ (surface) and at $V_{\max}$ (wireframe).

basing the sequence completion on any single time frame or even multiple scans (i.e. end-systolic and end-diastolic). We demonstrated that the model was able to learn abstract associations between the clinical demography and atrial shape and motion. We expect that the model can be extended with a larger collection of diverse non-imaging data, which will allow for generating synthetic populations with specific demography, disease status or biophysical constraints. Being able to generate realistic anatomical sequences of dynamically moving anatomies based on such conditions is a valuable tool for methods such as fluid simulation, operation planning and disease detection not only in the cardiac domain, but any domain where spatio-temporal models are of interest.

Acknowledgments. This work was supported by a PhD grant from the Technical University of Denmark - Department of Applied Mathematics and Computer Science (DTU Compute) and d by the Spanish Ministry of Science and Innovation under the Programme "Proyectos de Generación de Conocimiento 2022" (PID2022-143239OB-I00).

Disclosure of Interests. Kristine Sørensen now works for Novo Nordisk A/S. Rasmus Paulsen and Klaus Kofoed has received a research grant from Novo Nordisk A/S.

References

1. Amiranashvili, T., Lüdke, D., Li, H.B., Menze, B., Zachow, S.: Learning shape reconstruction from sparse measurements with neural implicit functions. Medical Imaging with Deep Learning (MIDL) **172**, 22–34 (2022)
2. Amirrajab, S., Khalil, Y.A., Lorenz, C., Weese, J., Pluim, J., Breeuwer, M.: A framework for simulating cardiac mr images with varying anatomy and contrast. IEEE Transactions on Medical Imaging **42**(3), 726–738 (2023)
3. Arin, K., Huan, T., Blostein, S.: Least-squares fitting of 2 3-d point sets. IEEE Transactions on Pattern Analysis and Machine Intelligence **9**(5), 699–700 (1987)
4. Beetz, M., Acero, J.C., Banerjee, A., Eitel, I., Zacur, E., Lange, T., Stiermaier, T., Evertz, R., Backhaus, S.J., Thiele, H., Bueno-Orovio, A., Lamata, P., Schuster, A., Grau, V.: Mesh u-nets for 3d cardiac deformation modeling. Statistical Atlases and Computational Modeling of the Heart (STACOM) pp. 245–257 (2022)
5. Beetz, M., Banerjee, A., Grau, V.: Generating subpopulation-specific biventricular anatomy models using conditional point cloud variational autoencoders. Statistical Atlases and Computational Modeling of the Heart (STACOM) pp. 75–83 (2022)
6. Beetz, M., Corral Acero, J., Banerjee, A., Eitel, I., Zacur, E., Lange, T., Stiermaier, T., Evertz, R., Backhaus, S.J., Thiele, H., Bueno-Orovio, A., Lamata, P., Schuster, A., Grau, V.: Interpretable cardiac anatomy modeling using variational mesh autoencoders. Frontiers in Cardiovascular Medicine **9** (2022)
7. Biffi, C., Cerrolaza, J.J., Tarroni, G., Bai, W., De Marvao, A., Oktay, O., Ledig, C., Le Folgoc, L., Kamnitsas, K., Doumou, G., Duan, J., Prasad, S.K., Cook, S.A., O'Regan, D.P., Rueckert, D.: Explainable anatomical shape analysis through deep hierarchical generative models. IEEE Transactions on Medical Imaging **39**(6), 2088–2099 (2020)
8. Chou, G., Bahat, Y., Heide, F.: Diffusion-sdf: Conditional generative modeling of signed distance functions. International Conference on Computer Vision (ICCV) pp. 2262–2272 (2023)

9. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Computer vision and image understanding **61**(1), 38–59 (1995)
10. Duchateau, N., Sermesant, M., Delingette, H., Ayache, N.: Model-based generation of large databases of cardiac images: Synthesis of pathological cine mr sequences from real healthy cases. IEEE Transactions on Medical Imaging **37**(3), 755–766 (2018)
11. Erkoc, Z., Ma, F., Shan, Q., Niesner, M., Dai, A.: Hyperdiffusion: Generating implicit neural fields with weight-space diffusion. International Conference on Computer Vision (ICCV) pp. 14254–14264 (2023)
12. Glikson, M., Wolff, R., Hindricks, G., Mandrola, J., Camm, A.J., Lip, G.Y., Fauchier, L., Betts, T.R., Lewalter, T., Saw, J., Tzikas, A., Sternik, L., Nietlispach, F., Berti, S., Sievert, H., Bertog, S., Meier, B.: EHRA/EAPCI expert consensus statement on catheter-based left atrial appendage occlusion - An update. EuroIntervention **15**(13), 1133–1180 (2020)
13. Hoogendoorn, C., Sukno, F.M., Ordás, S., Frangi, A.F.: Bilinear models for spatio-temporal point distribution analysis :application to extrapolation of left ventricular, biventricular and whole heart cardiac dynamics. International Journal of Computer Vision **85**(3), 237–252 (2009)
14. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. ACM siggraph computer graphics **21**(4), 163–169 (1987)
15. Lu, P., Bai, W., Rueckert, D., Noble, J.A.: Multiscale graph convolutional networks for cardiac motion analysis. Statistical Atlases and Computational Modeling of the Heart (STACOM) pp. 264–272 (2021)
16. Medrano-Gracia, P., Cowan, B.R., Suinesiaputra, A., Young, A.A.: Atlas-based anatomical modeling and analysis of heart disease. Drug Discovery Today: Disease Models **14**, 33–39 (2014)
17. Ossenberg-Engels, J., Grau, V.: Conditional generative adversarial networks for the prediction of cardiac contraction from individual frames. Statistical Atlases and Computational Modeling of the Heart (STACOM) pp. 109–118 (2020)
18. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. Computer Vision and Pattern Recognition Conference (CVPR) pp. 165–174 (2019)
19. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems **32** (2019)
20. Peng, J., Beetz, M., Banerjee, A., Chen, M., Grau, V.: Generating virtual populations of 3d cardiac anatomies with snowflake-net. STACOM) pp. 163–173 (2024)
21. Peyrat, J.M., Delingette, H., Sermesant, M., Xu, C., Ayache, N.: Registration of 4d cardiac ct sequences under trajectory constraints with multichannel diffeomorphic demons. IEEE Transactions on Medical Imaging **29**(7), 1351–1368 (2010)
22. Prakosa, A., Sermesant, M., Delingette, H., Marchesseau, S., Saloux, E., Allain, P., Villain, N., Ayache, N.: Generation of synthetic but visually realistic time series of cardiac images combining a biophysical model and clinical images. IEEE Transactions on Medical Imaging **32** (2013)
23. Qiao, M., Wang, S., Qiu, H., De Marvao, A., O'Regan, D.P., Rueckert, D., Bai, W.: Cheart: A conditional spatio-temporal generative model for cardiac anatomy. IEEE Transactions on Medical Imaging (2023)

24. Reynaud, H., Vlontzos, A., Dombrowski, M., Gilligan Lee, C., Beqiri, A., Leeson, P., Kainz, B.: D'artagnan: Counterfactual video generation. Medical Image Computing and Computer-Assisted Intervention (MICCAI) pp. 599–609 (2022)
25. Schroeder, W., Martin, K., Lorensen, B.: The Visualization Toolkit–An Object-Oriented Approach To 3D Graphics. Kitware, Inc., fourth edn. (2006)
26. Slipsager, J.M., Juhl, K.A., Sigvardsen, P.E., Kofoed, K.F., De Backer, O., Olivares, A.L., Camara, O., Paulsen, R.R.: Statistical shape clustering of left atrial appendages. In: Statistical Atlases and Computational Modeling of the Heart (STACOM). pp. 32–39 (2018)
27. Sorensen, K., Camara, O., Backer, O.D., Kofoed, K.F., Paulsen, R.R.: Nudf - neural unsigned distance fields for high resolution 3d medical image segmentation. IEEE International Symposium on Biomedical Imaging (ISBI) (2022)
28. Stolt-Ansó, N., McGinnis, J., Pan, J., Hammernik, K., Rueckert, D.: Nisf: Neural implicit segmentation functions. Medical Image Computing and Computer-Assisted Intervention (MICCAI) pp. 734–744 (2023)