



Steerable Anatomical Shape Synthesis with Implicit Neural Representations

Bram de Wilde^{1,3,*}, Max T. Rietberg^{2,*}, Guillaume Lajoinie¹, and Jelmer M. Wolterink³

¹ Physics of Fluids Group, Technical Medical (TechMed) Centre, University of Twente, Enschede, The Netherlands

² Multi-Modality Medical Imaging, Technical Medical (TechMed) Centre, University of Twente, Enschede, The Netherlands

³ Department of Applied Mathematics, Technical Medical (TechMed) Centre, University of Twente, Enschede, The Netherlands
contact@bramdewilde.com

Abstract. Generative modeling of anatomical structures plays a crucial role in virtual imaging trials, which allow researchers to perform studies without the costs and constraints inherent to *in vivo* and phantom studies. For clinical relevance, generative models should allow targeted control to simulate specific patient populations rather than relying on purely random sampling. In this work, we propose a steerable generative model based on implicit neural representations. Implicit neural representations naturally support topology changes, making them well-suited for anatomical structures with varying topology, such as the thyroid. Our model learns a disentangled latent representation, enabling fine-grained control over shape variations. Evaluation includes reconstruction accuracy and anatomical plausibility. Our results demonstrate that the proposed model achieves high-quality shape generation while enabling targeted anatomical modifications.

Keywords: Shape Synthesis · Implicit Neural Representations · Latent Space Disentanglement

1 Introduction

Evaluating the performance of novel medical imaging modalities through patient trials is expensive in terms of patient risk, time, and cost [1]. Simulating *in vivo* trials with tissue-mimicking phantoms is common practice, but since recreating human tissue is complex, experiments with high mimicking accuracy remain costly [2]. Virtual imaging trials, where both the patient and the imaging modality are simulated, provide a compelling alternative: they allow for inexpensive, rapid and flexible testing while closely matching patient tissue [3]. However, these trials are only meaningful if the underlying anatomical model of the patients is accurate, but also covers the variations seen in the population.

* These authors contributed equally to this work.

Generative models have the potential to synthesize anatomical structures [4], enhancing existing (open source) anatomical datasets, such as the Gold Atlas project [5] and MedShapeNet [6]. However, in medical imaging, factors not directly related to the clinical question can strongly influence the image and its interpretation. For example, imaging might be affected by anatomical variation and tissue composition, which purely random generative approaches cannot disentangle. There is therefore a need for models that provide control over such factors, allowing virtual cohorts to be generated in line with study objectives.

In recent years, implicit neural representations (INRs) have emerged as a powerful and flexible platform for shape modeling [7,8]. To describe shapes, INRs can represent surfaces as the zero-level set of their signed distance function (SDF), modeled in a neural network. Unlike more traditional template-based approaches, INRs are amenable to various conditioning mechanisms and naturally allow for modeling topological changes in a population [9,10]. Conditioning INRs on relevant shape features introduces an additional level of control for shape synthesis, which has previously been demonstrated outside [11,12], and inside the medical domain [13,14,15]. Furthermore, the ability of INRs to represent organic shapes with varying or changing topology has been previously demonstrated in biomedical applications [16]. Topological changes in anatomy can, for instance, occur due to certain diseases like tissue adhesion, renal fusion, or osteosarcoma, or due to surgical procedures like a laryngectomy.

In this paper, we investigate steerable anatomical shape synthesis in a population with topological variations. As a concrete use case we opt to model the thyroid gland, which is diverse in its shape and bilateral symmetry, and not topologically consistent across the patient population. The thyroid consists of two lobes, connected by a bridge called the *isthmus*. However, up to 33% of all patients show agenesis of the isthmus, having two separate lobes instead of a single connected thyroid [17]. We show that INRs are capable of synthesizing anatomically feasible thyroids. Additionally, we condition INRs on three key anatomical features (volume, isthmus cross-sectional area and symmetry) to generate and edit thyroids in a steerable way. Finally, we experiment with a simple correlation loss term to promote feature disentanglement.

2 Methods

2.1 Model

To model three-dimensional shapes, we use coordinate-based multilayer perceptrons (MLPs) to encode the SDF. Similar to [7], we use a single MLP to represent multiple shape instances by conditioning the MLP on a latent code \mathbf{z} , which is concatenated to the input coordinates. We use MLPs with 3 hidden layers of 256 nodes and ReLU activations. For every target shape i , we sample the SDF value $s \in \mathbb{R}$ for a set of coordinates $\mathbf{x} \in \mathbb{R}^3$:

$$X_i = \{(\mathbf{x}, s) : SDF_i(\mathbf{x}) = s\} \quad (1)$$

The parameters θ of the MLP f_θ are optimized such that the model approximates the SDF for each shape i when conditioned on its latent code \mathbf{z}_i :

$$f_\theta(\mathbf{x}, \mathbf{z}_i) \approx SDF_i(\mathbf{x}) \quad (2)$$

2.2 Training

Both the parameters θ and the latent codes \mathbf{z}_i are optimized using the mean squared error loss on the predicted SDF values. Additionally, we apply L_2 regularization to the latent codes as we assume Gaussian noise on the SDF values, leading to the following total loss \mathcal{L} :

$$\mathcal{L}(f_\theta, X, \mathbf{z}) = \sum_{\mathbf{x} \in X} \|f_\theta(\mathbf{x}, \mathbf{z}) - SDF(\mathbf{x})\|_2^2 + \lambda \|\mathbf{z}\|_2^2 \quad (3)$$

2.3 Disentanglement

Training a model as described in Section 2.1 does not impose any structure on the latent space. Hence, randomly sampled codes provide novel samples, but there is no easy way to control anatomical aspects, such as the volume of the generated shape. To promote steerability of output shape characteristics, we split the latent code \mathbf{z}_i for each shape into a fixed part $\mathbf{z}_{i, \text{fixed}}$, which is not updated during training, and a trainable part $\mathbf{z}_{i, \text{trainable}}$. By letting $\mathbf{z}_{i, \text{fixed}}$ represent anatomical features, the model is directly conditioned on them, such that we can investigate disentanglement of $\mathbf{z}_{i, \text{fixed}}$ with respect to the trainable features. If $\mathbf{z}_{i, \text{fixed}}$ is, for instance, set to the volume of each shape i , then $\mathbf{z}_{i, \text{trainable}}$ will ideally model all shape changes *but* the volume. Concretely, we investigate the two following disentanglement strategies:

Fixed conditioning The anatomical feature(s) of interest are added as fixed latent code features and training is done as described in Section 2.1. This is a baseline approach to investigate how much the model disentangles the features by itself without any special strategies.

Correlation loss In addition to the **Fixed conditioning** strategy, we calculate a correlation loss term at the end of each epoch to update the latent codes. The loss term promotes disentanglement of the fixed features from the trainable features. Specifically, the loss term calculates the mean Pearson correlation coefficient between the fixed and trainable latent features:

$$\mathcal{L}_{\text{corr}} = \frac{1}{N} \sum_j \left| \frac{\text{Cov}(\mathbf{z}_{\text{fixed}}, \mathbf{z}_{\text{trainable}, j})}{\sigma_{\text{fixed}} \sigma_{\text{trainable}}} \right| \quad (4)$$



Fig. 1. Examples from the dataset illustrating the variety in thyroid anatomy. From left to right: (1) a typical connected thyroid, (2) a typical split thyroid, (3) a very large thyroid, (4) a very small thyroid, (5) a highly asymmetric thyroid.

Here N is the number of trainable latent features, j is the latent feature index. If a model is perfectly disentangled, the fixed feature should have no correlation to any of the trainable dimensions, and hence have a low loss value.

2.4 Inference

Shapes are synthesized by conditioning the model on a latent code \mathbf{z} and sampling the predicted SDF values $f_\theta(\mathbf{x}, \mathbf{z})$ for $\mathbf{x} \in [0, 1]^3$, where \mathbf{x} is sampled on a 64^3 grid. Meshes are obtained with marching cubes. Novel samples are synthesized by sampling randomly from the latent space. For each trainable latent dimension, we fit and draw random values from a normal distribution. For fixed latent dimensions in conditioned models, we sample directly from the distribution seen in the dataset, such that the synthesized population follows the training population.

2.5 Data

We collect thyroid shapes from the TotalSegmentator training dataset [18]. The dataset consists of 1228 CT scans on which 117 structures, including the thyroid gland, have been annotated. Of these 1228 scans, there are 415 scans where the thyroid gland is fully in the field of view and is completely annotated. We additionally collect the trachea shape for each thyroid to center the meshes. Due to the diverse nature of the dataset, not all cases are annotated consistently. After a check for annotation quality, which includes checking for a number of connected components and watertightness, we discard 62 cases. Finally, we train all models on a set of 353 thyroids. Representative examples from the dataset are shown in Figure 1.

We choose to center all meshes on the trachea, instead of on thyroid center of mass, because asymmetric thyroids would give off-center results. To allow for centering, we first convert all binary voxel grid thyroid masks to meshes with marching cubes, and then center them on the center of mass of the trachea. The meshes are finally normalized to the largest extent in the dataset in each dimension, such that all shapes fall within a common unit cube, but size variations are preserved.

For each thyroid mesh, we sample the SDF values for 50,000 coordinates for training. 40,000 coordinates are randomly sampled on the mesh surface and

hence have an SDF of 0. The remaining 10,000 points are a random sample of the 40,000 surface points plus a randomly sampled displacement in each direction from a Gaussian distribution with a standard deviation of 0.1. We publish the processed meshes, pre-sampled SDF values and our source code online.⁴

2.6 Validation

To validate that the baseline model is able to capture the anatomical variations in the dataset, we evaluate reconstruction accuracy with the Chamfer distance between reconstructed and reference meshes. To validate that the baseline model can also *generate* meaningful novel samples, we validate that they are anatomically realistic with respect to three key anatomical features, of which we compare training to generated population. Specifically, the anatomical features we consider are thyroid **volume**, **isthmus area**, and **symmetry**.

Volume is a natural measure for size variations. Furthermore, isthmus area allows for a continuous description from a split thyroid (area=0) to a connected one. It is calculated as the thyroid area in the midsagittal plane. Finally, to characterize the large variation in symmetry in the dataset, we flip one half of the thyroid onto the other by mirroring at the midsagittal plane, and calculating the intersection over union between both halves. In this measure, a perfectly symmetric thyroid has a symmetry score of 1, whereas a completely asymmetric thyroid has a score of 0.

To validate steerability of the conditioned models (*fixed* and *correlation*), we randomly generate 1000 thyroids and evaluate the correlation between the conditioned anatomical features and the actual generated features with the Pearson correlation coefficient (PCC).

3 Results

All models are trained with the Adam optimizer with a learning rate of $3 \cdot 10^{-4}$. Each model is trained for 10,000 epochs, where each epoch consists of 1000 coordinate-SDF pairs for each shape, sampled randomly from X_i . Models are trained with a trainable latent code of size $N = 64$, with their values initialized randomly from $\mathcal{N}(0, 0.01^2)$ and where applicable, features added as 3-vectors. Throughout the results we compare three different models:

- *Baseline* A model conditioned only on a trainable latent code.
- *Fixed* The baseline model, but additionally conditioned on volume, isthmus area, and symmetry. This model is trained with the default loss (3).
- *Correlation* The fixed model, with the correlation loss (4) added.

⁴ <https://github.com/MIAGroupUT/steerable-shape-synthesis>

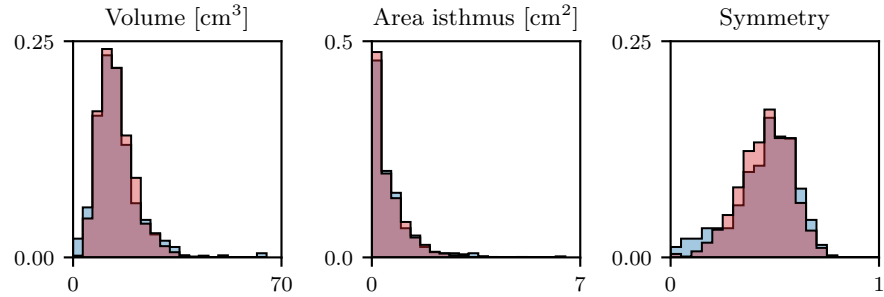


Fig. 2. Comparison between volume, isthmus area and symmetry for the training data (blue) and 1000 randomly generated meshes using the *baseline* model (red). The dark red denotes the overlap of both distribution

3.1 Reconstruction quality

The mean Chamfer distance between training meshes and their corresponding reconstructed meshes were 1.56 ± 0.38 mm (std) for the *baseline* model, 1.63 ± 0.12 mm for the *fixed* model, and 1.60 ± 0.06 mm for the *correlation* model. This shows that the baseline model is able to fit all shapes in the dataset well, and that adding extra conditioning to the latent code does not impact reconstruction quality, as the dimensions of the shapes in the training set are $44.75 \pm 9.79 \times 29.32 \pm 5.91 \times 52.18 \pm 7.03$ mm.

3.2 Generation quality

To demonstrate the generative performance of the *baseline* model, 1000 meshes were randomly generated. The generated distributions of volume, isthmus area and symmetry compared to the training distributions are shown in Figure 2. These results indicate that our baseline approach is able to generate thyroids which cover the entire training distribution. Moreover, the learned latent space follows the anatomical distribution of the training data. Furthermore, to verify that we are generating new shapes and are not simply replicating shapes in the training set, we computed the pairwise Chamfer distance between all shapes in the training set. This showed that the average Chamfer distance to the closest shape was 4.13 ± 1.33 (std). For the 1000 randomly generated shapes, the average distance to the closest shape in the training set was 3.67 ± 0.51 , 3.72 ± 0.95 , and 3.69 ± 0.76 for the *baseline*, *fixed* and *correlation* versions, respectively. Hence, while shapes are samples from the same distribution, synthetic shapes have no match in the training set.

3.3 Steerability

To evaluate to which extent latent conditioning on volume, isthmus area and symmetry works, we generate 1000 random meshes for both the *fixed* and the

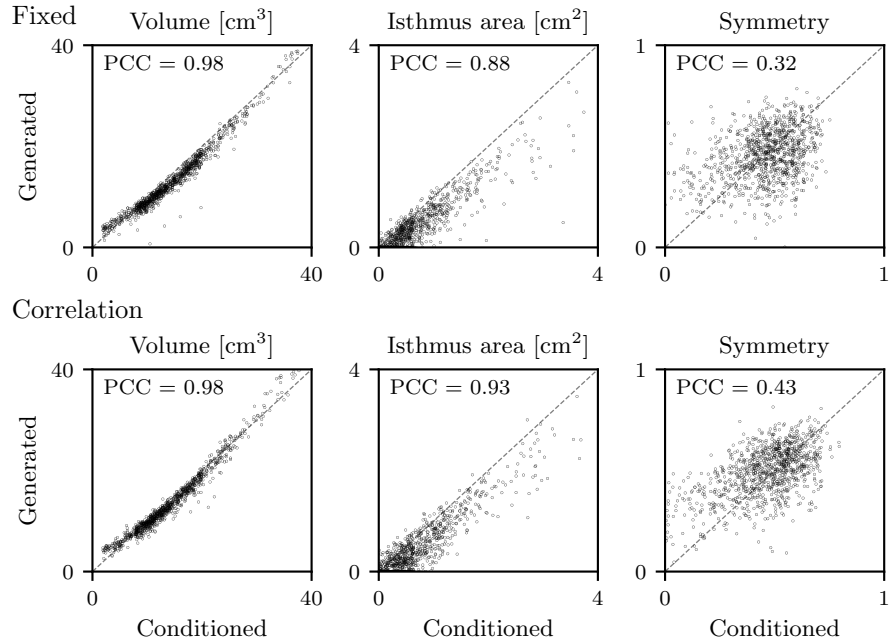


Fig. 3. Correlation between conditioned and generated anatomical features for the *fixed* model (top row) and the *correlation* model (bottom row). The inset shows the Pearson correlation coefficient.

correlation model. Figure 3 shows the correlation between features the model is conditioned on and the actual generated features for the both models. Generated volume correlates well with the conditioned volume with a PCC of 0.98 for both models, indicating that volume is a relatively easy feature to disentangle. Isthmus area and especially symmetry appear more complex features to disentangle, and here the use of the correlation loss improves the PCC in both cases: from 0.88 to 0.93 for the isthmus area, and from 0.32 to 0.43 for symmetry.

To investigate the ability of the *correlation* model to independently vary specific features, we show the editing of a particular training mesh in Figure 4. While varying a specific feature, the other features show almost no difference, with exception of the isthmus area when increasing the symmetry value. We hypothesize this might be because the cross-sectional area increases when you tilt a thyroid, i.e., make it less symmetrical. This could be a reason why it is harder for the model to disentangle symmetry from the isthmus area than from the volume.

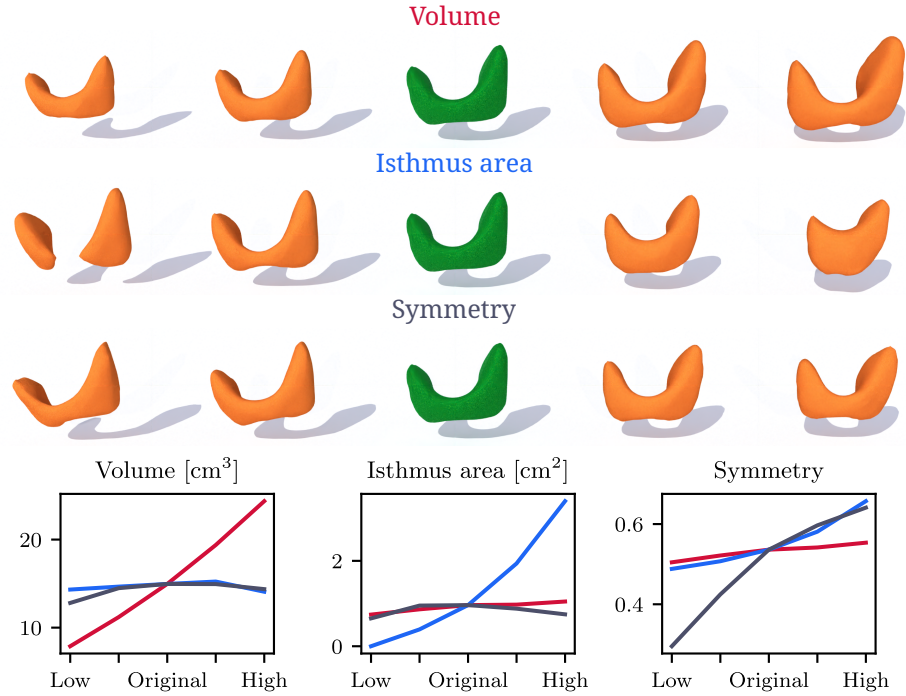


Fig. 4. Editing a training mesh (middle column, green) by independently varying volume (red), isthmus area (blue) and symmetry (grey). The plots show each anatomical feature for each of the rows, demonstrating that features can be independently varied.

4 Discussion

We show that INRs are capable of synthesizing anatomically accurate thyroid glands, including the ability to model topological changes across a patient population. By conditioning the models on volume, isthmus area, and symmetry, it is possible to synthesize thyroids in a controlled manner. While prior work by Sørensen et al. [14] has shown that anatomy synthesis with INRs can be conditioned on explicit patient characteristics, we here show that explicitly minimizing the correlation between fixed and trainable features enables disentanglement of these features in the latent space. The resulting model can be used to generate patient cohorts with specific anatomical characteristics and allows editing anatomical features of existing thyroids with a high degree of independence.

Both Figure 3 and 4 show that the three anatomical features we include are not equally complex for the model. Volume is fitted well by both models, but isthmus area is already more difficult, which might be due to fact that it is a more local feature. Symmetry is clearly the most complex, although Figure 2 and 4 show that the models can model the feature to some extent. In order for the

model to disentangle symmetry, it arguably needs to learn some representation of that feature, which might be difficult for our relatively small MLP models.

While the thyroid is a complex use-case due to variations in topology and symmetry, the shapes themselves are still relatively smooth. Whether our approach generalizes to organs with sharper features that require a representation containing high frequencies is an interesting direction for future research. In this case, investigating alternative model architectures such as Siren, which has been shown to have a stronger capacity for representing sharp features, is probably a fruitful direction [19].

Another interesting avenue for future investigation is extending our approach to a multi-organ setting. In this setting, a model could be conditioned on a unique latent code for each output organ. The generative process can then be steered on an organ-level basis, while still providing a coherent set of generated organs, respecting the anatomical hierarchy. Using the proposed correlation loss could help models to disentangle organ representations in latent space, keeping their representation independent, and allowing for imposing variations in single organs.

Acknowledgments. All authors acknowledge the funding from KWF, TKI-Life Sciences and Health and Seno Medical Instruments in project THYNAS+. Additionally, Guillaume Lajoinie acknowledges funding from the HORIZON.1.1 program of the European Research Council for the Super-FALCON project, grant agreement ID: 101076844.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Fogel, D.B.: Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemporary Clinical Trials Communications* **11**, 156–164 (2018). <https://doi.org/https://doi.org/10.1016/j.conctc.2018.08.001>, <https://www.sciencedirect.com/science/article/pii/S2451865418300693>
2. McGarry, C.K., Grattan, L.J., Ivory, A.M., Leek, F., Liney, G.P., Liu, Y., Miloro, P., Rai, R., Robinson, A.P., Shih, A.J., Zeqiri, B., Clark, C.H.: Tissue mimicking materials for imaging and therapy phantoms: a review. *Physics in Medicine & Biology* **65**(23), 23TR01 (dec 2020). <https://doi.org/10.1088/1361-6560/abbd17>, <https://dx.doi.org/10.1088/1361-6560/abbd17>
3. Abadi, E., Segars, W.P., Tsui, B.M.W., Kinahan, P.E., Bottenus, N., Frangi, A.F., Maidment, A., Lo, J., Samei, E.: Virtual clinical trials in medical imaging: a review. *Journal of Medical Imaging* **7**(4), 042805 (2020). <https://doi.org/10.1117/1.JMI.7.4.042805>, <https://doi.org/10.1117/1.JMI.7.4.042805>
4. Liu, Y., Dwivedi, G., Boussaid, F., Bennamoun, M.: 3d brain and heart volume generative models: A survey. *ACM Computing Surveys* **56**(6), 1–37 (Jan 2024). <https://doi.org/10.1145/3638044>, <http://dx.doi.org/10.1145/3638044>
5. Nyholm, T., Svensson, S., Andersson, S., Jonsson, J., Sohlén, M., Gustafsson, C., Kjellén, E., Söderström, K., Albertsson, P., Blomqvist, L., Zackrisson, B., Olsson,

- L.E., Gunnlaugsson, A.: Mr and ct data with multiobserver delineations of organs in the pelvic area—part of the gold atlas project. *Medical Physics* **45**(3), 1295–1300 (2018). <https://doi.org/https://doi.org/10.1002/mp.12748>, <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.12748>
6. Li, J., Zhou, Z., Yang, J., Pepe, A., Gsaxner, C., Luijten, G., Qu, C., Zhang, T., Chen, X., Li, W., Wodzinski, M., Friedrich, P., Xie, K., Jin, Y., Ambigapathy, N., Nasca, E.: Medshapenet – a large-scale dataset of 3d medical shapes for computer vision. *Biomedical Engineering / Biomedizinische Technik* **70**(1), 71–90 (2025). <https://doi.org/doi:10.1515/bmt-2024-0396>, <https://doi.org/10.1515/bmt-2024-0396>
 7. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 165–174 (2019)
 8. Peng, S., Niemeyer, M., Mescheder, L., Pollefeys, M., Geiger, A.: Convolutional occupancy networks. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) *Computer Vision – ECCV 2020*, pp. 523–540. Springer International Publishing, Cham (2020)
 9. Dupont, E., Kim, H., Eslami, S.M.A., Rezende, D., Rosenbaum, D.: From data to functa: Your data point is a function and you can treat it like one. <https://doi.org/10.48550/arXiv.2201.12204>, <http://arxiv.org/abs/2201.12204>
 10. Liu, H.T.D., Williams, F., Jacobson, A., Fidler, S., Litany, O.: Learning smooth neural functions via lipschitz regularization. <https://doi.org/10.48550/arXiv.2202.08345>, <http://arxiv.org/abs/2202.08345>
 11. Mu, J., Qiu, W., Kortylewski, A., Yuille, A., Vasconcelos, N., Wang, X.: A-sdf: Learning disentangled signed distance functions for articulated shape representation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 13001–13011 (2021)
 12. Miao, J., Ikeda, T., Raytchev, B., Mizoguchi, R., Hiraoka, T., Nakashima, T., Shimizu, K., Higaki, T., Kaneda, K.: Manipulating vehicle 3d shapes through latent space editing. <https://doi.org/10.48550/arXiv.2410.23931>, <http://arxiv.org/abs/2410.23931>
 13. Kong, F., Stocker, S., Choi, P.S., Ma, M., Ennis, D.B., Marsden, A.L.: Sdf4chd: Generative modeling of cardiac anatomies with congenital heart defects. *Medical Image Analysis* **97**, 103293 (2024). <https://doi.org/https://doi.org/10.1016/j.media.2024.103293>, <https://www.sciencedirect.com/science/article/pii/S1361841524002184>
 14. Sørensen, K., Diez, P., Margeta, J., El Youssef, Y., Pham, M., Pedersen, J.J., Kühl, T., de Backer, O., Kofoed, K., Camara, O., Paulsen, R.: Spatio-temporal neural distance fields for conditional generative modeling of the heart. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*. pp. 422–432. Springer Nature Switzerland, Cham (2024)
 15. Dannecker, M., Kyriakopoulou, V., Cordero-Grande, L., Price, A.N., Hajnal, J.V., Rueckert, D.: CINA: Conditional Implicit Neural Atlas for Spatio-Temporal Representation of Fetal Brains . In: *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*. vol. LNCS 15009. Springer Nature Switzerland (October 2024)
 16. Wiesner, D., Suk, J., Dummer, S., Nečasová, T., Ulman, V., Svoboda, D., Wolterink, J.M.: Generative modeling of living cells with so (3)-equivariant implicit neural representations. *Medical image analysis* **91**, 102991 (2024)

17. Ranade, A.V., Rai, R., Pai, M., Nayak, S., Prakash, K.A., Krisnamurthy, A., Narayana, S.: Anatomical variations of the thyroid gland: possible surgical implications. *Singapore medical journal* **49**(10), 831 (2008)
18. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., Bach, M., Segeroth, M.: Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence* **5**(5), e230024 (2023). <https://doi.org/10.1148/ryai.230024>, <https://doi.org/10.1148/ryai.230024>
19. Sitzmann, V., Martel, J.N.P., Bergman, A.W., Lindell, D.B., Wetzstein, G.: Implicit neural representations with periodic activation functions. <https://doi.org/10.48550/arXiv.2006.09661>, <http://arxiv.org/abs/2006.09661>