Learning a Conditional Generative Model for Anatomical Shape Analysis

Benjamín Gutiérrez-Becker and Christian Wachinger

Artificial Intelligence in Medical Imaging (AI-Med), Department of Child and Adolescent Psychiatry, University Hospital, LMU Munich

Abstract. We introduce a novel conditional generative model for unsupervised learning of anatomical shapes based on a conditional variational autoencoder (CVAE). Our model is specifically designed to learn latent, low-dimensional shape embeddings from point clouds of large datasets. By using a conditional framework, we are able to introduce side information to the model, leading to accurate reconstructions and providing a mechanism to control the generative process. Our network design provides invariance to similarity transformations and avoids the need to identify point correspondences between shapes. Contrary to previous discriminative approaches based on deep learning, our generative method does not only allow to produce shape descriptors from a point cloud, but also to reconstruct shapes from the embedding. We demonstrate the advantages of this approach by: (i) learning low-dimensional representations of the hippocampus and showing low reconstruction errors when projecting them back to the shape space, and (ii) demonstrating that synthetic point clouds generated by our model capture morphological differences associated to Alzheimer's disease, to the point that they can be used to train a discriminative model for disease classification.

1 Introduction

Over the last decades, a variety of approaches for shape analysis have been developed for modeling the human anatomy from medical images [17]. These approaches have become a mainstay in medical image analysis, not only because of their utility in providing priors for segmentation, but also because of their value in quantifying shape changes between subjects and populations. Shape analysis helps in localizing anatomical changes, which can yield a better understanding of morphological changes due to aging and disease [7, 24].

Given that the morphology of organs across a population is highly heterogeneous, modeling and quantifying these shape variations is a challenging task. Thanks to the growing availability of large-scale medical imaging datasets, we have now the possibility to model these underlying shape variations in the population more accurately. Unfortunately, working on large sample sizes comes with computational challenges, which can limit the practical application of traditional methods for shape analysis [17]. In addition, imaging datasets usually come with valuable phenotypic information of the patient. This large amount of available

data, paired with recent advances in machine learning, calls for the development of a data-driven and learning-based shape analysis framework that can benefit from the large amount of image data and provides a mechanism to include prior information in the analysis.

Many fields in medical image analysis have recently been revolutionized by the introduction of deep neural networks [15]. These approaches have the ability to learn complex, hierarchical feature representations that have proven to outperform hand-crafted features in a variety of applications. One of the reasons for the superior performance is their ability to model complex non-linear relationships between variables. Medical shape analysis has not been untouched by this wave and deep neural networks for disease prediction have been proposed [10, 21]. Although these approaches have demonstrated the benefit of learning shape representations optimal for a given task with deep neural networks, the generation of new shapes based on low-dimensional representations has not yet been explored.

In this paper, we propose a conditional generative model for learning shape representations, which is based on a conditional variational autoencoder operating directly on unordered point clouds. Our model offers the following advantages: 1) our framework is invariant to similarity transformations, avoiding the need to pre-align the shapes to be analyzed; 2) our network operates on point clouds, which present a raw, simple and lightweight representation that is trivial to obtain from a segmented surface; 3) our method is invariant to the ordering of the elements in the point cloud, meaning that computing correspondences between points across shapes is not necessary; 4) our method does not impose any constraints on the topology of the shapes, providing high flexibility; 4) the conditional nature of our network gives us the possibility to introduce prior knowledge in a simple manner; 5) the model scales to analyzing large shape datasets; 6) the neural network learns modes of variation that capture complex shape changes, yielding a compact representation and the generation of realistic samples.

1.1 Related Work

A large volume of work in medical shape analysis is based on point distribution models (PDMs) [3], which represent surfaces of objects as point clouds. A statistical model is built by finding correspondences between points of different shapes and by obtaining the principal modes of variation via principal component analysis. PDMs have been widely been used due to their simplicity and due to their application to segmentation through active shape models. One common drawback of PDMs is, however, that point correspondences have to be found between all shapes in a dataset. This usually involves a registration step, which is not only challenging but also computationally expensive for large databases. Moreover, homologous features may not exist when comparing shapes that are subject to strong variations, e.g., over the course of brain development. While our method is also based on point clouds, we do not require correspondences between shapes. Next to point clouds, other popular representations for shape

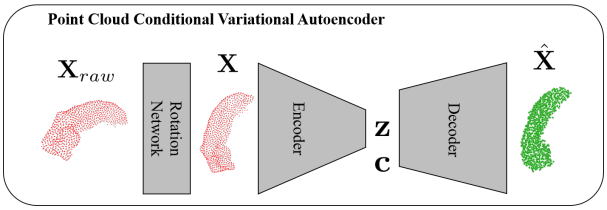

Fig. 1: Overview of our network architecture. Our framework is based on three main components: 1) a rotation network bringing the input point cloud to a canonical space, 2) an encoder approximating the posterior distribution $P_E(\mathbf{z}|\mathbf{X})$, and 3) a decoder reconstructing the point cloud by approximating the mapping $[\mathbf{z}, \mathbf{c}] \mapsto \mathbf{X}$.

analysis are skeletal models [18], spectral signatures [23], spherical harmonics [7], and deformations [4, 16].

Conditional variational autoencoders [14, 22] are an extension of the generative model in variational autoencoders by introducing a condition vector, which allows to include prior information in the autoencoder. A CVAE has recently been used in medical imaging for 3D fetal skull reconstruction from 2D ultrasound [2]. Conditional generative models have also recently become popular in the context of generative adversarial networks [8]. A conditional adversarial networks was proposed as a general-purpose solution to image-to-image translation problems [11]. In contrast to those previous work, we are proposing a conditional generative model for shape analysis on point cloud representations.

2 Method

An overview of our generative model is shown in Fig. 1. Our approach is based on a CVAE that encodes a point cloud $\mathbf{X} = \{\mathbf{p}_0, \mathbf{p}_1, ..., \mathbf{p}_n\}$ with $\mathbf{p}_i = [x_i, y_i, z_i]$ into a set of k-dimensional latent variables $\mathbf{z} \in \mathbf{\Omega}_z \subset \mathbb{R}^k$ and then decodes this embedding to reconstruct a point cloud $\hat{\mathbf{X}}$. Our network architecture consists of three main elements: 1) the rotation network aligning input point clouds to a canonical space, 2) the encoder aiming at finding the posterior distribution $P_E(\mathbf{z}|\mathbf{X})$, and 3) the decoder approximating the mapping $[\mathbf{z}, \mathbf{c}] \mapsto \mathbf{X}$, where $\mathbf{c} \in \mathbb{R}^m$ is a condition vector of dimension m. Our network is trained in an end-to-end fashion using a loss function, which jointly minimizes the alignment error with respect to a reference shape, the reconstruction error and the latent loss of the variational autoencoder.

Our generative model can be employed in two different ways: First, to obtain a low-dimensional embedding $\mathbf{z}$ given an input $\mathbf{X}$, which in turn can be used to perform basic operations between shapes and to compute shape statistics. Second, to generate synthetic point clouds $\hat{\mathbf{X}}$ from the learned embedding space by sampling $\mathbf{z}$ from a multivariate Gaussian and setting a condition vector $\mathbf{c}$.

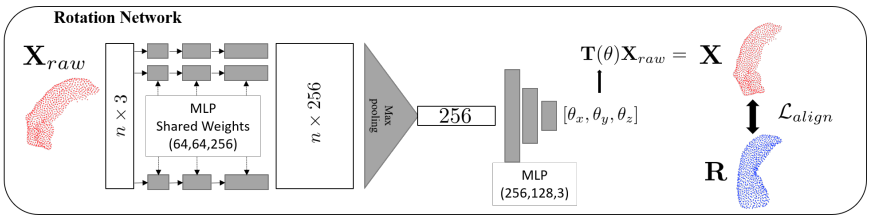

Fig. 2: Rotation network transforming the input point cloud $\mathbf{X}_{raw}$ to bring it into alignment with the reference template $\mathbf{R}$. The quality of the alignment is measured by the loss function $\mathcal{L}_{align}$. Numbers between the parenthesis are the dimensions of the layers.

2.1 Rotation Network

According to one of its most popular definitions, shape is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object [13]. Thus when our network receives as input a raw point cloud $\mathbf{X}_{raw}$, we must first ensure that its output is invariant to similarity transformations (scaling, translation, and rotations).

Invariance to scaling and translation can be enforced by first centering an input shape $\mathbf{X}_{raw}$ around its center of mass and normalizing the point coordinates in the $[0, 1]$ range. To guarantee invariance to rotation, we introduce a rotation network (Fig. 2) that learns the mapping $f(\mathbf{X_{raw}}) \mapsto \theta$, such that $\mathbf{X} = \mathbf{T}(\theta)\mathbf{X}_{raw}$ is in spatial alignment with a reference point cloud $\mathbf{R}$. The rotation matrix $\mathbf{T}(\theta)$ is parameterized by the rotation vector $\theta = [\theta_x, \theta_y, \theta_z]^T$. An important challenge when working with point cloud representations is that point clouds are in mathematical terms an order-less set. Traditional statistical shape models solve this challenge by first finding point correspondences between point clouds, therefore inducing an order to the set. Instead, we propose to use a network architecture, which is invariant to point ordering. The architecture of the rotation network and the encoder are based on PointNet [19], which operates directly on orderless point clouds. The basic operation of the rotation network is to first pass each individual point of the network through a multilayer perceptron (MLP), with shared weights among all points, projecting each 3D point to a higher dimensional representation. These representations are aggregated using the max pooling operator across all points. Max pooling is a symmetric operation, and therefore invariant to point ordering. Third, the output of the max pooling layer is fed into a MLP, which predicts the rotation parameters θ. Our transformation network therefore has the form:

$$f(\mathbf{X_{raw}}) = [\theta_x, \theta_y, \theta_z] = \text{MLP}\Big(\max_{\mathbf{p} \in \mathbf{X_{raw}}} h(\mathbf{p})\Big), \tag{1}$$

where h corresponds to the operations of the MLP with shared weights. Note that the separate convolution of each point and the following aggregation guarantee the invariance to point ordering. Our mechanism to measure the quality of our

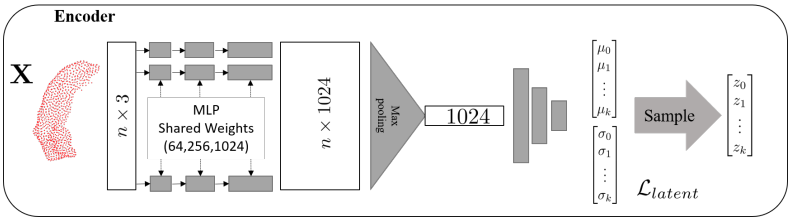

Fig. 3: Encoder network approximates the distribution $P_E(\mathbf{z}|\mathbf{X})$. The input of the network is the aligned point cloud $\mathbf{X}$ and the output is the embedding $\mathbf{z}$ generated by the normal distribution $\mathcal{N}_z(\mu, \Sigma)$. Numbers inside the parenthesis are layer sizes.

alignment to the reference template is to measure a distance between $\mathbf{R}$ and $\mathbf{X}$. Since our framework operates on unordered point clouds, we require a metric which is permutation invariant. We use the 1-Wasserstein distance, also known as earth mover's distance (EMD) [20], defined as:

$$\mathcal{L}_{align}(\mathbf{X}, \mathbf{R}) = EMD(\mathbf{X}, \mathbf{R}) = \min_{\phi:\mathbf{X}\to\hat{\mathbf{R}}} \sum_{\mathbf{p}\in\mathbf{X}} ||\mathbf{p} - \phi(\mathbf{p})||_1, \tag{2}$$

where $\phi(\mathbf{p})$ is a bijection, which maps a point $\mathbf{p} \in \mathbf{X}$ to its closest point $\mathbf{r} \in \mathbf{R}$.

2.2 Encoder Network

The encoder seeks an approximation to the posterior distribution $P_E(\mathbf{z}|\mathbf{X})$. The architecture of the encoder is illustrated in Fig. 3. The encoder and the rotation network have a very similar architecture, since both take unordered point clouds as input and predict a vector of parameters. Two main differences exist between the rotation network and the encoder: first, the dimensions of the MLP layers of the encoder are larger, to give additional descriptive power to the encoding task; second, while the rotation network estimates rotation parameters, the encoder estimates vectors $\mu = [\mu_0, \mu_1, ..., \mu_k]$ and $\Sigma = \text{diag}[\sigma_0, \sigma_1, ...\sigma_k]$. These vectors are the parameters of a normal distribution $\mathcal{N}_z(\mu, \Sigma)$, which approximates the posterior $P_E(\mathbf{z}|X)$. This means that during training, given an input $\mathbf{X}$, the low-dimensional embedding $\mathbf{z}$ is obtained by drawing a sample at random from $\mathcal{N}_z$. At this stage, we introduce a latent loss for the variational autoencoder $\mathcal{L}_{latent}$ given by the Kullback-Leibler divergence between $\mathcal{N}_z$ and a Gaussian prior $\mathcal{N}(0, \mathbf{I})$. Since Σ is a diagonal matrix, the Kullback-Leibler divergence between these distributions is:

$$\mathcal{L}_{latent} = \sum_{i=1}^{k} \sigma_i + \mu_i - \log(\sigma_i) - 1. \tag{3}$$

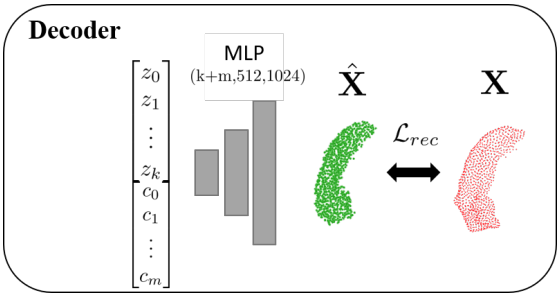

Fig. 4: Decoder network approximates the mapping $[\mathbf{z}, \mathbf{c}] \mapsto \mathbf{X}$. The input of the decoder are the embedding $\mathbf{z}$ and the condition vector $\mathbf{c}$. The output is the reconstructed point cloud $\hat{\mathbf{X}}$. The accuracy of the reconstruction is measured using the reconstruction loss $\mathcal{L}_{rec}$. Numbers inside the parenthesis are layer sizes.

2.3 Decoder Network

The last part of our framework is the decoder network (Fig. 4). The decoder maps the embedding to a reconstructed point cloud by approximating the mapping $[\mathbf{z}, \mathbf{c}] \mapsto \hat{\mathbf{X}}$. Similar to previous approaches based on CVAEs, the decoder is a fully connected MLP with 3 layers, which maps the low-dimensional representation back to a reconstruction $\hat{\mathbf{X}}$. The decoder also takes as input the vector $\mathbf{c} \in \mathbb{R}^m$, which allows our network to include conditions to the reconstruction of $\hat{\mathbf{X}}$. The quality of the reconstruction is evaluated by a reconstruction loss $L_{rec} = EMD(\mathbf{X}, \hat{\mathbf{X}})$, which measures the EMD between the input shape and its reconstruction.

The full network is trained in an end-to-end fashion using stochastic gradient descent by optimizing the loss function:

$$\mathcal{L} = \mathcal{L}_{align} + \mathcal{L}_{latent} + \mathcal{L}_{rec}. \tag{4}$$

3 Experiments

3.1 Conditional Shape Model of 3D Digits

As a first experiment, we train a generative shape model using a 3D point cloud version of the MNIST database[1] and successively sample point clouds from the low-dimensional embedding. This dataset consists of 5000 3D point clouds of handwritten digits from 0 to 9. For this experiment, we trained two separate generative models. For the first one, we set the the dimension of the embedding $\mathbf{z}$ to $k = 2$, and we use a 10-dimensional one hot encoding of the class of each digit as the condition vector $\mathbf{c}$. The second model is trained under the same

[1] https://github.com/Harry-Zhi/3DMNIST

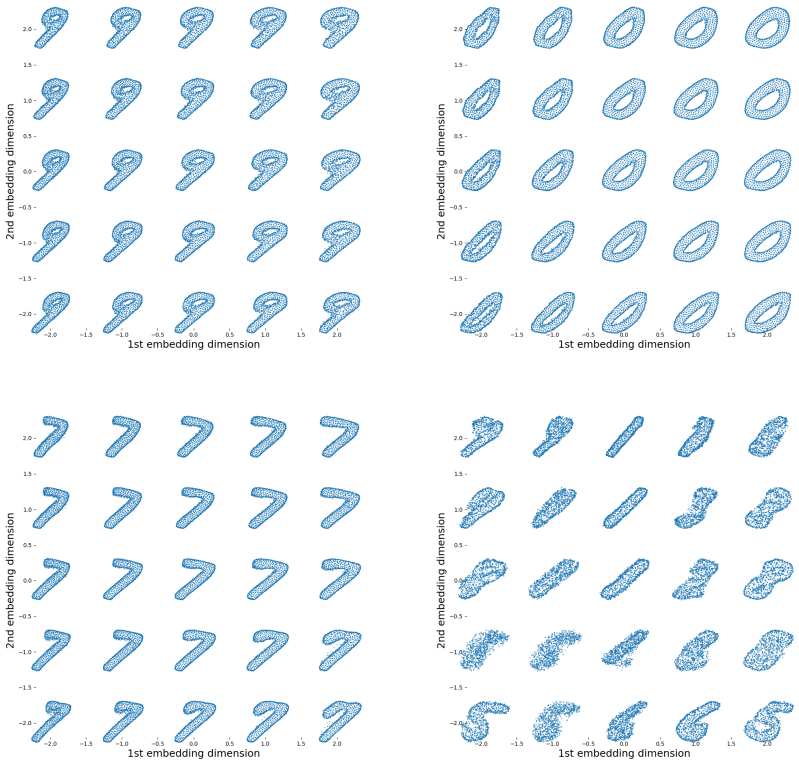

Fig. 5: Point clouds sampled from the 2D embedding space generated by training our model using the 3D MNIST dataset. On the bottom right we show 3D point clouds generated by setting the conditional vector $\mathbf{c}$ to zero. For the other figures, $\mathbf{c}$ is set to generate point clouds of the digits $9, 0$ and 7.

settings but with the condition vector $\mathbf{c}$ set to all zeros. This means that both models are essentially identical, with the important difference that the first one is equipped with a condition vector, which allows us to give information to the network about the digit to be encoded and reconstructed. In Fig. 5, we present artificial point clouds generated by these two models. At the bottom right of Fig. 5, we show point clouds generated without the use of the condition vector $\mathbf{c}$. Although the model is able to generate some realistically looking digits (like the 1s in the center column), the reconstructed point clouds are generally not as sharp as those generated by the conditional model. In contrast, by setting the condition vector to generate a specific digit, we are able to obtain sharp point clouds while at the same time capturing complex non-linear deformations for each digit. The digits in Fig. 5 present a very similar orientation (tilted to the right and aligned with respect to the x, y plane). This is the result of

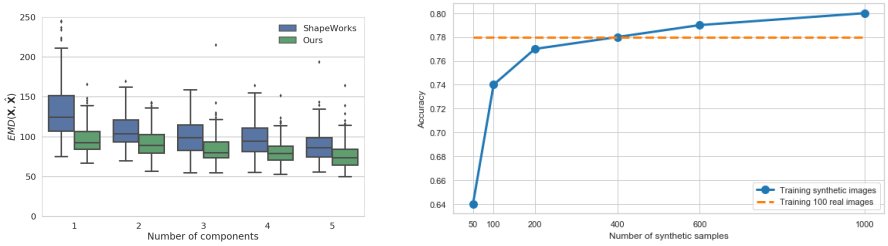

Fig. 6: Left: reconstruction error of shapes generated using either ShapeWorks or our CVAE framework with respect to the input point cloud. Right: HC vs AD Classification accuracy for a PointNet model trained either using point clouds obtained from real segmentations obtained from the ADNI database or synthetic hippocampus point clouds generated by our model.

aligning the point clouds to a reference template using the rotation network. An important observation is that all digits are sampled from the same shape space $\mathbf{\Omega}_z$, and only the condition vector $\mathbf{c}$ changes. This means that the encoding $\mathbf{z}$ is able to encode common shape characteristics between all digits. For example, the 1st embedding dimension in Fig. 5 captures the width of the digits. It is also worth mentioning that for many typical statistical shape models, training a shape model consisting of 5000 point clouds would be impractical due to memory limitations and to the computationally expensive task of finding corresponding points between all these shapes.

3.2 Conditional Shape Model of the Hippocampus

In our second experiment, we build a shape model of the left hippocampus. Our goal is to assess shape differences between healthy controls (HC) and subjects diagnosed with Alzheimer's disease (AD). Several previous studies have established strong morphological changes in the hippocampus associated to the progression of dementia [6, 7]. Magnetic resonance images of 200 subjects were randomly selected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [12] and processed with Freesurfer [5] to obtain segmentations. For comparison, we build a statistical shape model of the hippocampus using the ShapeWorks framework [1]. ShapeWorks is a statistical shape model tool, which achieved the best performance in several shape analysis tasks in a recent comparison [9]. For our evaluation, we split the images into a training and testing set (100/100 split) and we build a statistical shape model of the left hippocampus using the training set (50 HC and 50 AD). Segmentations are pre-processed using the grooming operations included in ShapeWorks to obtain smooth hippocampi surfaces, and models of 1024 points are trained. As a condition vector $\mathbf{c}$, we use a one hot encoding of the diagnosis of the patient ($[0, 1]$ for HC, $[1, 0]$ for AD). We limit our analysis to this relatively small number of samples to be able to perform a fair comparison with ShapeWorks, which is limited in the number of samples to

be analyzed due to memory constraints. It is also worth mentioning that training the ShapeWorks model for 100 images took 5h, compared to the 2h training time for our model.

Reconstruction error. We first evaluate the ability of our model to obtain an accurate and compact representation of the hippocampus shape. To this end, we measure the reconstruction error between the reconstructed shapes $\hat{\mathbf{X}}$ and the input shapes $\mathbf{X}$ by evaluating $EMD(\mathbf{X}, \hat{\mathbf{X}})$. We train 5 different models with embedding dimensions ranging from $k = 1$ to $k = 5$. As a comparison, we quantify the reconstruction error of synthetic hippocampus shapes generated by ShapeWorks. The lower reconstruction errors of our method in Fig. 6 indicate that it captures the complex deformations of the hippocampi and therefore allows for a compact shape representation with few modes.

Effect of conditioning the shape model using a diagnostic label. One of the main contributions that separates our framework from previous approaches for shape analysis is the introduction of the conditional vector $\mathbf{c}$. We have observed in our experiment on the MNIST dataset that our method is able to generate realistic shapes of digits given different condition vectors $\mathbf{c}$. To evaluate the effects of the condition vector in the model of the hippocampus shapes, we use the model trained on the previous experiment (for embedding dimension $k = 2$) and generate a set of synthetic point clouds by sampling values of $\mathbf{z}$ and assigning either $\mathbf{c} = [1, 0]$ or $\mathbf{c} = [0, 1]$ to generate synthetic hippocampus shapes corresponding to morphological characteristics associated to either HC or AD. In Fig. 7 we can observe some of the synthetic shapes generated by our model, corresponding to the mean shape (center) and shapes generated by moving across the first embedding dimension z_0. Notice that shapes in the same column correspond were generated using the same embedding $\mathbf{z}$, with different condition vector $\mathbf{c}$. In Fig.7 we can observe that by moving across z_0 our model captures shape differences which are common between the HC and AD cases. For example, we observe that the left most example for both cases has a large deformation on the top part of the hippocampus. On the bottom row, we show differences between the point clouds of the top two rows, which correspond to the shape variations that our model associates to the presence of AD. These shape variations correspond to large variations in the lateral part of the hippocampus body, roughly around the CA1 subfield. These observations are in line with previous findings on shape differences of the left hippocampus associated to AD diagnosis [6, 7].

Synthesizing training data. A critical question to answer is whether our synthetically generated point clouds capture shape differences that are specific to AD. We assess this by generating synthetic point clouds associated to HC or AD, and training a PointNet classifier [19, 10] to discriminate between hippocampi belonging either to HC or AD subjects. We experiment with synthetic datasets

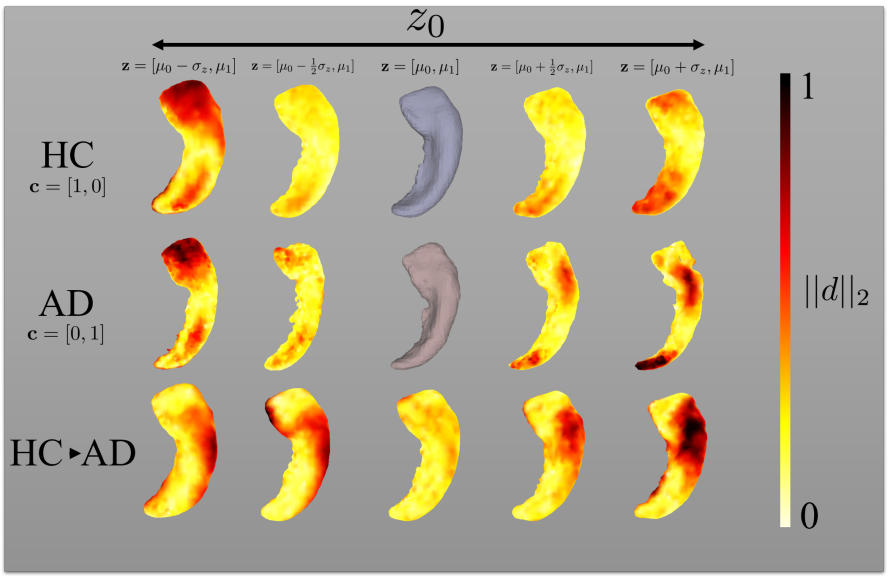

Fig. 7: Hippocampus surfaces generated using point clouds sampled from our model trained on segmented images obtained from the ADNI database. The top row corresponds to point clouds generated by setting the condition vector to generate HC shapes, and the middle row corresponds to AD. Point clouds are generated by moving along the first embedding dimension. For the top two rows, the color coding shows the deformation (measured as the absolute distance between corresponding points) between the mean and the generated point cloud. In the bottom row, the deformation between HC and AD shapes generated using the same shape embedding $\mathbf{z}$ are shown.

generated by our model of sizes: 50, 100, 200, 400, 600, and 1000. For each dataset, a separate PointNet classifier is trained. We compare the classification accuracy of our model with a PointNet classifier trained using the 100 samples of the training set directly, without the use of the generative model. The results in Fig. 6 show that our generated samples are realistic enough to train a classifier relying solely on the synthetic images. Interestingly, our generator allow us to sample an arbitrary number of samples, giving us the possibility to boost the accuracy of the classifier by increasing artificially the size of the dataset.

4 Conclusions

In this work, we have presented a conditional generative model to model anatomical shapes. This model is able to generate low-dimensional shape representations

taking as input unordered point clouds, without the need of finding point correspondences between them. We have demonstrated that our model can be used to encode complex shape variations using a low-dimensional embedding and we have shown that by introducing a conditional vector, we are able to obtain more accurate reconstructions. We have demonstrated the properties of our generative model by creating realistically looking synthetic shapes, which can even be used to train deep learning based models. This has the potential to enable the use of powerful models in scenarios where the amount of annotated data is limited. On the hippocampus experiments, we operated on relatively small sample sizes to ensure a fair comparison to previous approaches, but on the MNIST data we demonstrated that our network scales to datasets with thousands of shapes. Our network facilitates processing of large datasets, since we do not require expensive operations for finding point correspondences between samples. We believe that our framework can be used to analyze other anatomical structures and more importantly the use of different condition vectors, which include diverse phenotypic information.

5 Acknowledgments.

This work was supported in part by DFG and the Bavarian State Ministry of Education, Science and the Arts in the framework of the Centre Digitalisation.Bavaria (ZD.B).

References

1. Cates, J., Fletcher, P.T., Styner, M., Hazlett, H.C., Whitaker, R.: Particle-based shape analysis of multi-object complexes. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 477–485. Springer (2008)
2. Cerrolaza, J.J., Li, Y., Biffi, C., Gomez, A., Sinclair, M., Matthew, J., Knight, C., Kainz, B., Rueckert, D.: 3d fetal skull reconstruction from 2dus via deep conditional generative networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 383–391. Springer (2018)
3. 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)
4. Durrleman, S., Prastawa, M., Charon, N., Korenberg, J.R., Joshi, S., Gerig, G., Trouvé, A.: Morphometry of anatomical shape complexes with dense deformations and sparse parameters. NeuroImage 101, 35–49 (2014)
5. Fischl, B.: Freesurfer. Neuroimage 62(2), 774–781 (2012)
6. Frisoni, G.B., Ganzola, R., Canu, E., Rüb, U., Pizzini, F.B., Alessandrini, F., Zoccatelli, G., Beltramello, A., Caltagirone, C., Thompson, P.M.: Mapping local hippocampal changes in alzheimer's disease and normal ageing with mri at 3 tesla. Brain 131(12), 3266–3276 (2008)
7. Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.S., Niethammer, M., Dubois, B., Lehéricy, S., Garnero, L., et al.: Multidimensional classification of hippocampal shape features discriminates alzheimer's disease and mild cognitive impairment from normal aging. Neuroimage 47(4), 1476–1486 (2009)

8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
9. Goparaju, A., Csecs, I., Morris, A., Kholmovski, E., Marrouche, N., Whitaker, R., Elhabian, S.: On the evaluation and validation of off-the-shelf statistical shape modeling tools: A clinical application. In: International Workshop on Shape in Medical Imaging. pp. 14–27. Springer (2018)
10. Gutierrez, B., Wachinger, C.: Deep multi-structural shape analysis: Application to neuroanatomy. International Conference on Medical Image Computing and Computer-Assisted Intervention (2018)
11. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5967–5976. IEEE (2017)
12. Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L Whitwell, J., Ward, C., et al.: The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of magnetic resonance imaging 27(4), 685–691 (2008)
13. Kendall, D.G.: A survey of the statistical theory of shape. Statistical Science pp. 87–99 (1989)
14. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
15. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis 42, 60–88 (2017)
16. Miller, M.I., Younes, L., Trouvé, A.: Diffeomorphometry and geodesic positioning systems for human anatomy. Technology 2(01), 36–43 (2014)
17. Ng, B., Toews, M., Durrleman, S., Shi, Y.: Shape analysis for brain structures. In: Shape Analysis in Medical Image Analysis, pp. 3–49. Springer (2014)
18. Pizer, S.M., Jung, S., Goswami, D., Vicory, J., Zhao, X., Chaudhuri, R., Damon, J.N., Huckemann, S., Marron, J.: Nested sphere statistics of skeletal models. In: Innovations for Shape Analysis, pp. 93–115. Springer (2013)
19. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. Proc. Computer Vision and Pattern Recognition (CVPR), IEEE 1(2), 4 (2017)
20. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. International journal of computer vision 40(2), 99–121 (2000)
21. Shakeri, M., Lombaert, H., Tripathi, S., Kadoury, S., Initiative, A.D.N., et al.: Deep spectral-based shape features for alzheimer's disease classification. In: International Workshop on Spectral and Shape Analysis in Medical Imaging. pp. 15–24. Springer (2016)
22. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. In: Advances in Neural Information Processing Systems. pp. 3483–3491 (2015)
23. Wachinger, C., Golland, P., Kremen, W., Fischl, B., Reuter, M.: BrainPrint: A discriminative characterization of brain morphology. Neuroimage 109, 232–248 (2015)
24. Wachinger, C., Rieckmann, A., Reuter, M.: Latent processes governing neuroanatomical change in aging and dementia. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 30–37. Springer (2017)