EXPLAINING ANATOMICAL SHAPE VARIABILITY: SUPERVISED DISENTANGLING WITH A VARIATIONAL GRAPH AUTOENCODER

Johannes Kiechle[1,2], *Dylan Miller*[1], *Jordan Slessor*[3], *Matthew Pietrosanu*[3], *Linglong Kong*[3], *Christian Beaulieu*[4], *Dana Cobzas*[1,5]

[1]Department of Computing Science, University of Alberta, Edmonton, Canada
[2]Department of Electrical and Computer Engineering, Technical University of Munich, Munich, Germany
[3]Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada
[4]Department of Biomedical Engineering, University of Alberta, Edmonton, Canada
[5]Department of Computer Science, MacEwan University, Edmonton, Canada

ABSTRACT

This work proposes a modular geometric deep learning framework that isolates shape variability associated with a given scalar factor (e.g., age) within a population (e.g., healthy individuals). Our approach leverages a novel graph convolution operator in a variational autoencoder to process 3D mesh data and learn a meaningful, low-dimensional shape descriptor. A supervised disentanglement strategy aligns a single component of this descriptor with the factor of interest during training. On a toy synthetic dataset and a high-resolution diffusion tensor imaging (DTI) dataset, the proposed model is better able to disentangle the learned latent space with a simulated factor and patient age, respectively, relative to other state-of-the-art methods. The relationship between age and shape estimated in the DTI analysis is consistent with existing neuroimaging literature.

Index Terms— Anatomical shape analysis, graph convolution, hippocampus, latent space disentanglement

1. INTRODUCTION

Shape analysis [1] has grown increasingly relevant in medical research for its potential to delineate morphological variability between and within populations, e.g., between healthy and abnormal structures. Technological advancements in medical imaging have led to an abundance of neurological shape data and a greater need for analytic methods able to identify and explain sources of inter-subject and inter-group variability.

This work is motivated primarily by the question of how the shape of the human hippocampus changes with age in a healthy population. This investigation is a crucial first step in understanding the impact of neurodegenerative diseases such as Alzheimer's disease on brain structure and, ultimately, cognitive function. Meaningful comparisons between healthy and non-normative populations are only possible once shape variation in the former is understood.

Shape analysis for medical imaging has traditionally been conducted with statistical shape models, which aim to describe a population of shapes [2] and typically use principal component analysis (PCA) as a workhorse. Despite this, the linear decomposition inherent to PCA might not adequately capture high-order, nonlinear shape variability. In addition, large datasets pose computational challenges that can limit the application of traditional methods for shape analysis. Despite the power of deep learning tools in computer vision, natural language processing, audio analysis, and many other domains, their use in shape analysis is still limited [3, 4].

In this work, we propose a graph convolution network (GCN) based on spatial convolutions [5]. This type of model is better able to represent 3D mesh data [6, 7] relative to other GCN-based methods (e.g., spectral approaches [8, 9]). We extend the standard GCN by incorporating a disentanglement strategy that isolates a given factor of interest (e.g., age) in the latent space. This approach permits control over the data-generative process on the basis of the given factor. The learned decoder captures high-order relationships between latent shape representations and 3D shapes and permits further examination of how the latter varies with the factor of interest. Through a comparison with ShapeWorks, a state-of-the-art statistical shape modeling tool, we demonstrate the superior potential of our model for disentangled shape modeling [10].

This work contributes a GCN based on spatial convolutions in a modular deep learning framework that can

1. describe global and local anatomical shape variability,
2. disentangle latent representations of 3D mesh data with a categorical or continuous factor of interest, and
3. generate anatomical 3D meshes and examine shape variability with respect to the given factor.

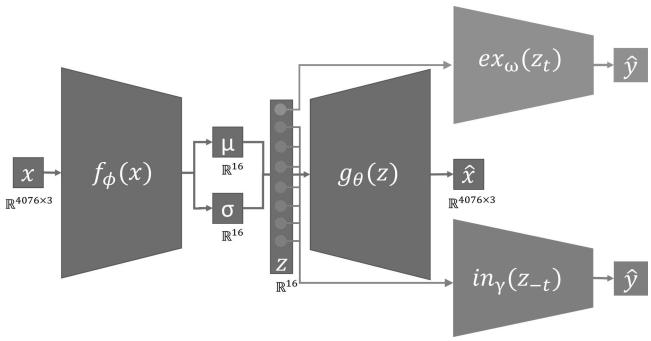

Fig. 1. The proposed SpiralNet variational graph autoencoder for supervised disentangled shape modeling.

2. METHODS

2.1. Neural network architecture

We propose a novel modular neural network framework, which we call a guided spiral β-VAE, that can efficiently identify shape variability associated with a given factor of interest y. Specifically, our network uses a deep variational autoencoder (VAE) [11] with spiral graph convolution layers [5, 7] to learn a meaningful low-dimensional shape descriptor. Our model further includes excitation and inhibition mechanisms that disentangle the latent space [12]. Figure 1 depicts our proposed modular neural network architecture.

The proposed deep graph convolution autoencoder takes, as an input, 3D mesh vertices $X = [x_0, x_1, \ldots, x_{N-1}]^\top \in \mathbb{R}^{N\times F}$, where F is the feature dimension and N is the total number of vertices per mesh. For 3D mesh data, $F = 3$ and X gives the coordinates of each vertex.

The network's encoder follows the SpiralNet structure [5]: all vertices of an input mesh are connected by a spiral trajectory starting at a random vertex. Spiral convolution operations are performed in the following manner. First, mesh vertices along the vertex trajectory within a fixed distance are concatenated (i.e., neighborhood aggregation). Second, the concatenated vertices are fed through a multilayer perceptron (MLP) (i.e., weight sharing). To disentangle the latent space with respect to a specified factor of interest y, our framework incorporates adversarial excitation and inhibition mechanisms, as proposed by Ding et al. [12], via separate feed-forward neural networks (Figure 1). Our approach to disentanglement extends the evidence lower bound,

$$\begin{aligned}&\mathrm{ELBO}(\theta,\phi) = \\ &\quad \mathbb{E}_{z\sim Q_\phi(\cdot\,|\,x)}[\log P_\theta(x \mid z)] - \beta\,\mathrm{KL}(Q_\phi(\cdot \mid x)\|P),\end{aligned} \tag{1}$$

which can be maximized over θ and ϕ to train a standard β-VAE. In Equation (1), the likelihood P_θ and the variational distribution Q_ϕ are functions parameterized by θ and ϕ in the decoder and encoder, respectively, and P denotes a standard normal prior.

Our extension adds two subtasks. The first, excitation, forces one latent variable (here the tth, denoted by z_t, for a prespecified t) to be discriminative. Mathematically, we formulate excitation loss as

$$\mathcal{L}_{\mathrm{Ex}}(\phi, t) = \max_{\omega} \mathbb{E}_{z_t\sim q_\phi(\cdot|x)}[\log p_\omega(y \mid z_t)], \tag{2}$$

where ω parameterizes the excitation network. This subtask encourages the latent variable z_t to correspond to the factor of interest y (i.e., to minimize classification or regression error, depending on whether y is categorical or numeric). The second subtask, inhibition, encourages the remaining latent variables to be adversarially discriminative. We formulate inhibition loss as

$$\mathcal{L}_{\mathrm{In}}(\phi, t) = \max_{\gamma} \mathbb{E}_{z_{-t}\sim q_\phi(\cdot|x)}[\log p_\gamma(y \mid z_{-t})], \tag{3}$$

where γ parameterizes the inhibition network. This subtask will ultimately, in (4), encourage the remaining latent variables (i.e., z_{-t}) to be as independent of y as possible. Since y is scaled to $[0, 1]$, we maximize the minimal inhibition error by choosing a label value of 0.5, as in Ding et al. [12].

Combining the evidence lower bound in Equation (1) with the two subtasks above yields the problem

$$\max_{\theta,\phi}\{\mathrm{ELBO}(\theta,\phi) + \mathcal{L}_{\mathrm{Ex}}(\phi, t) - \mathcal{L}_{\mathrm{In}}(\phi, t)\}. \tag{4}$$

Here, the negative in front of the inhibition term encourages z_{-t} to be as uninformative as possible with respect to the factor of interest y, by pushing the best possible inhibition regressor/classifier to be the least accurate [12].

2.2. Implementation details

The architecture of our underlying variational graph convolution autoencoder is based on that of SpiralNet++ [7]. The encoder module consists of four spiral convolution layers with output channel sizes of $[8, 8, 8, 8]$ and a latent channel size of 16 (i.e., $z \in \mathbb{R}^{16}$). The decoder module mirrors the transformations in the encoder module. We set $\beta = 0.3$. As in the image domain, dilated spiral convolution, through subsampling, improves overall performance without increasing the size of the spiral. We use a dilation factor of 2 and a fixed spiral sequence length of 9. In numerical experiments, we employ an $80/10/10$ split for training/validation/testing and use an ADAM optimizer with a batch size of 32, an initial learning rate of 10^{-4}, and a training horizon of 250 epochs. We employ a scheduler that decays the primal learning rate by a factor of 0.99 every epoch. The models train to completion in approximately five minutes on an Nvidia Tesla P100 GPU in both experiments.

Model	# Latent Channels	Hippocampus ($n = 51$) ED ↓	SAP ↑	PCC ↕	Synthetic Data ($n = 50$) ED ↓	SAP ↑	PCC ↕
Proposed: Guided spiral β-VAE	16	1.61	0.48	0.71	0.14	0.89	0.97
Spiral β-VAE	16	1.62	0.04	0.36	0.45	0.01	0.24
ShapeWorks + PCA	16	17.62*	0.16	0.51	3.03*	0.14	-0.32

Table 1. Quantitative results on the testing set in the hippocampus and synthetic data analyses. ED is the Euclidean distance (in mm) between the original and reconstructed meshes (averaged over all vertices). For the ShapeWorks model, ED is computed from particles. SAP, the separated attribute predictability score, ranges between 0 and 1 and describes the compactness of the disentanglement: an SAP of 1 indicates that the factor of interest y is completely captured by exactly one latent variable. For the proposed model, PCC is the Pearson correlation coefficient between the disentangled latent variable z_t and the factor of interest y; for the other models, PCC is the absolutely largest correlation between y and any latent variable. The number of latent channels indicates the dimension of the embedding space for both VAE models and PCA. For reference, ↓ means that lower values are better, ↑ that higher values are better, and ↕ that values further from 0 are better.

3. EXPERIMENTS & RESULTS

3.1. Data & preprocessing

Our main focus is a neuroimaging dataset composed of diffusion tensor imaging (DTI) scans. See Solar et al. [13] for details on the acquisition protocol. High-resolution data (1 mm isotropic voxel size, 3 T, volume $220 \times 216 \times 20$ mm^3) were acquired for 511 healthy subjects (age 5–90 years, 286 females). The hippocampus in each scan was automatically segmented [14] and subsequently preprocessed as follows. First, the volumetric (i.e., voxel-based) representations were converted into 3D mesh representations with a marching cubes algorithm [15]. Second, Laplacian surface smoothing and rigid alignment via an iterative closest point algorithm were used to remove rotational artefacts. Third, because spiral graph convolution requires the same topology across instances, Deformetrica [16] was used to establish a point correspondence across the meshes. The result is a set of diffeomorphic deformation maps between a computed mean atlas and the subject meshes. Each mesh has 4076 vertices and 8144 faces.

As the hippocampus dataset does not provide a ground truth regarding the relationship between shape and age, we also consider a synthetic dataset with a clear relationship between object shape and y (Figure 2).

3.2. Results

For each dataset, we consider three models: the proposed guided spiral β-VAE [7, 11, 12], a spiral β-VAE [7, 11], and the particle-based statistical shape model in ShapeWorks [10] together with principle component analysis (PCA). In this section, we compare the quality of reconstructed meshes and generative power (based on latent disentanglement) across the models.

As shown in Table 1, the proposed model achieves the lowest reconstruction error (in terms of mean Euclidean distance) in both analyses. While the spiral β-VAE follows closely behind, there is a clear gap in performance relative to the ShapeWorks model. The top panel of Figure 2 confirms the faithfulness of the reconstructions obtained from the proposed model in the synthetic data analysis.

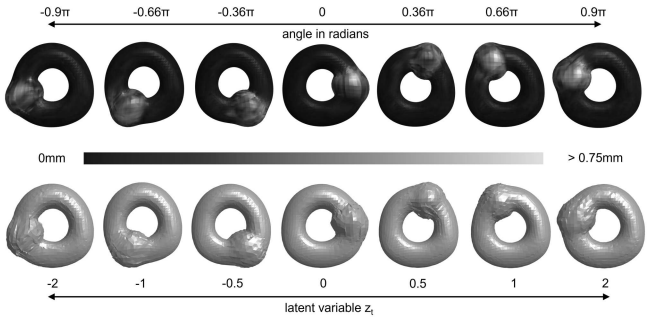

Fig. 2. Top: Reconstructions of the meshes in the synthetic dataset from the proposed model. A dark color indicates little deviation between the reconstruction and the original mesh. The number line indicates the true location (i.e., angle) of the "bump" on the torus. Bottom: The decoder's output while varying the disentangled latent variable z_t and holding the other latent variables constant at zero (i.e., $z_{-t} = 0$).

The bottom panel of Figure 2 highlights the generative power of the proposed model on the basis of the disentangled latent variable: the value of z_t clearly corresponds to the location of the "bump" in the upper panel. Table 1 further examines performance in this regard: in both analyses, the proposed model presents the highest separated attribute predictability score, while the other two models both yield a score close to 0. This result suggests that only the proposed model was successful in associating exactly one latent variable with y. Furthermore, the proposed model establishes a much stronger correlation between z_t and y in both analyses.

3.3. External validation

To externally validate the disentangled relationship between age and shape learned by the proposed model, we turn to ex-

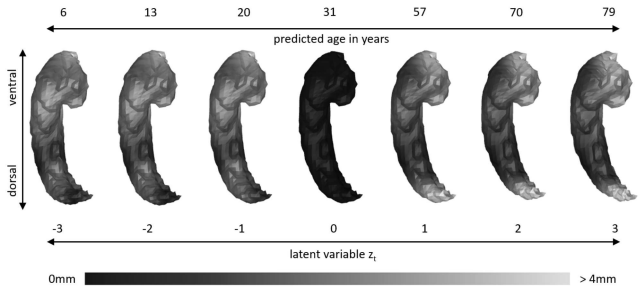

Fig. 3. Output from the decoder in the hippocampus analysis for different z_t when all other latent variables (i.e., z_{-t}) are held constant at 0. Dark colors indicate small differences relative to the shape with $z = 0$. The number lines indicate the learned correspondence between age and z_t.

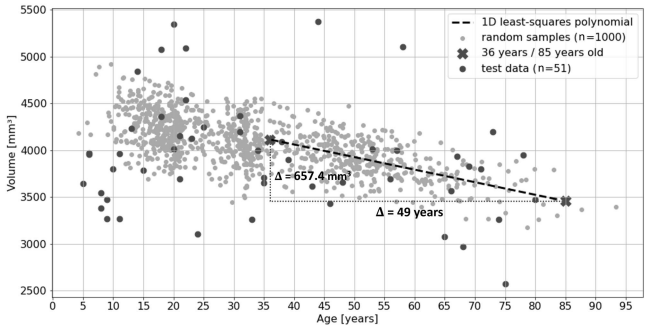

Fig. 4. Hippocampus volume versus age. Yellow points represent hippocampi decoded from randomly sampled latent codes, whereas green points represent the volume versus age information for the original test data. The dashed black line represents a least-squares linear fit to all randomly sampled hippocampi with a predicted age between 36 and 85 years.

isting research on hippocampus shape in normative populations. We consider both local (structural) and global (e.g., volume) shape changes in this section.

The local structural change in the tail (i.e., the dorsal subarea) of the hippocampus suggested by Figure 3 for larger ages is consistent with the findings of Solar et al. [13] and Bussy et al. [17]. Specifically, both works report significant linear decreases in volume in the tail of the hippocampus with age. We additionally observe structural changes in the body (i.e., the ventral subarea) of the hippocampus.

In investigating the relationship between hippocampus volume and age, Bussy et al. [17], Nobis et al. [18], Raz and Acker [19] found a generally negative association between volume and age, which coincides with our findings (Figure 4). Furthermore, Schuff et al. [20] reported that hippocampus volume diminishes by 20% between the ages of 36 and 85 (i.e., 14.6 mm^3 per year on average). Reconstructions from the proposed model are consistent with this finding and exhibit an approximate 16% (i.e., 13.4 mm^3 per year on average) decrease in volume over the same age range. While the proposed model satisfactorily captures the relationship between volume and age for adults, we acknowledge that performance degrades for infant and early-adolescent subjects. In our hippocampus dataset, less than 20% of all subjects were less than 13 years old, so we attribute this weak performance to a lack of information. Our future work will incorporate additional information such as the intracranial volume to improve the estimated age–shape relationship for young subjects.

4. CONCLUSION

The proposed modular geometric deep learning framework is able to isolate variability in shape associated with a given factor of interest. While this work is primarily motivated by the question of how hippocampus shape changes with age, our methods can also be applied to more-general tasks in shape analysis. Our evaluations on a synthetic dataset and real-world, high-resolution hippocampus data demonstrate that our model outperforms conventional statistical shape models [10]. The disentanglement performed by our model further provides new functionality in that the data-generative process can be controlled with respect to a factor of interest. The results of our neuroimaging analyses are consistent with existing literature and support the validity of our approach in applied contexts. To the best of our knowledge, there are no comparable methods applicable to biomedical applications with 3D mesh data. Future research can examine the impact of neurodegenerative diseases on brain structure (relative to healthy controls) and the study of compounding effects (e.g., age-by-sex effects).

5. COMPLIANCE WITH ETHICAL STANDARDS

No ethics approval was required for the synthetic data analysis. The neuroimaging study was approved by the Human Research Ethics Board at the University of Alberta.

6. ACKNOWLEDGEMENTS

We thankfully acknowledge the support of Cory Efird, who provided the synthetic “torus bump” data for our experiments. DC acknowledges student funding from the Natural Sciences and Engineering Research Council of Canada. JK acknowledges funding from the University of Alberta Research Experience Scholarship Program and from the German Academic Exchange Service IFI Program. We gratefully acknowledge the computational resources provided by the Digital Research Alliance of Canada.

7. REFERENCES

[1] K.V. Mardia and I.L. Dryden, "The statistical analysis of shape data," *Biometrika*, vol. 76, no. 2, pp. 271–281, 1989.

[2] A. Goparaju, K. Iyer, A. Bone, N. Hu, H.B. Henninger, A.E. Anderson, S. Durrleman, M. Jacxsens, A. Morris, I. Csecs, N. Marrouche, and S.Y. Elhabian, "Benchmarking off-the-shelf statistical shape modeling tools in clinical applications," *Medical Image Analysis*, 2022.

[3] R. Bhalodia, S.Y. Elhabian, L. Kavan, and R.T. Whitaker, "DeepSSM: A deep learning framework for statistical shape modeling from raw images," in *International Workshop on Shape in Medical Imaging*, 2018, pp. 244–257.

[4] A. Raju, S. Miao, D. Jin, L. Lu, J. Huang, and A.P. Harrison, "Deep implicit statistical shape models for 3D medical image delineation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, vol. 36, pp. 2135–2143.

[5] I. Lim, A. Dielen, M. Campen, and L. Kobbelt, "A simple approach to intrinsic correspondence learning on unstructured 3d meshes," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, pp. 349–362.

[6] G. Bouritsas, S. Bokhnyak, S. Ploumpis, M. Bronstein, and S. Zafeiriou, "Neural 3D morphable models: Spiral convolutional networks for 3D shape representation learning and generation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7213–7222.

[7] S. Gong, L. Chen, M. Bronstein, and S. Zafeiriou, "SpiralNet++: A fast and highly efficient mesh convolution operator," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 4141–4148.

[8] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun, "Spectral networks and locally connected networks on graphs," in *2nd International Conference on Learning Representations*, 2014.

[9] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 3844–3852.

[10] J. Cates, S. Elhabian, and R. Whitaker, "Shapeworks: Particle-based shape correspondence and visualization software," in *Statistical Shape and Deformation Analysis*, pp. 257–298. Elsevier, 2017.

[11] Diederik P. Kingma and Max Welling, "Auto-encoding variational bayes," in *2nd International Conference on Learning Representations*, 2014.

[12] Z. Ding, Y. Xu, W. Xu, G. Parmar, Y. Yang, M. Welling, and Z. Tu, "Guided variational autoencoder for disentanglement learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7920–7929.

[13] K.G. Solar, S. Treit, and C. Beaulieu, "High resolution diffusion tensor imaging of the hippocampus across the healthy lifespan," *Hippocampus*, vol. 31, no. 12, pp. 1271–1284, 2021.

[14] C. Efird, S. Neumann, K.G. Solar, C. Beaulieu, and D. Cobzas, "Hippocampus segmentation on high resolution diffusion MRI," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2021, pp. 1369–1372.

[15] W.E. Lorensen and H.E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *ACM SIGGRAPH Computer Graphics*, vol. 21, no. 4, pp. 163–169, 1987.

[16] S. Durrleman, M. Prastawa, N. Charon, J.R. Korenberg, S. Joshi, G. Gerig, and A. Trouvé, "Morphometry of anatomical shape complexes with dense deformations and sparse parameters," *NeuroImage*, vol. 101, pp. 35–49, 2014.

[17] A. Bussy, R. Patel, E. Plitman, S. Tullo, A. Salaciak, S.A. Bedford, S. Farzin, M.-L. Béland, V. Valiquette, C. Kazazian, C.L. Tardif, G.A. Devenyi, and M.M. Chakravarty, "Hippocampal shape across the healthy lifespan and its relationship with cognition," *Neurobiology of Aging*, vol. 106, pp. 153–168, 2021.

[18] L. Nobis, S.G. Manohar, S.M. Smith, F. Alfaro-Almagro, M. Jenkinson, C.E. Mackay, and M. Husain, "Hippocampal volume across age: Nomograms derived from over 19,700 people in UK Biobank," *NeuroImage: Clinical*, vol. 23, pp. 101904, 2019.

[19] N. Raz and J.D. Acker, "Regional brain changes in aging healthy adults: General trends, individual differences and modifiers," *Cerebral Cortex*, vol. 15, no. 11, pp. 1676–1689, 2005.

[20] N. Schuff, D.L. Amend, R. Knowlton, D. Norman, G. Fein, and M.W. Weiner, "Age-related metabolite changes and volume loss in the hippocampus by magnetic resonance spectroscopy and imaging," *Neurobiology of Aging*, vol. 20, no. 3, pp. 279–285, 1999.