Unsupervised Representation Learning of Cingulate Cortical Folding Patterns

Joël Chavas[*1] JOEL.CHAVAS@CEA.FR
Louise Guillon[*1] LOUISE.GUILLON@CEA.FR
Marco Pascucci[1] MARCO.PASCUCCI@CEA.FR
Benoit Dufumier[1,2] BENOIT.DUFUMIER@CEA.FR
Denis Rivière [1] DENIS.RIVIERE@CEA.FR
Jean-François Mangin[1] JFMANGIN@GMAIL.COM
[1] *Université Paris-Saclay, CEA, CNRS, NeuroSpin, Baobab, Gif-sur-Yvette, France*
[2] *LTCI, Télécom Paris, IPParis, France*

Editors: Under Review for MIDL 2022

Abstract

The human cerebral cortex is folded, making sulci and gyri over the whole cortical surface. Folding presents a very high inter-subject variability, and some neurodevelopmental disorders are correlated to local folding structures, named folding patterns. However, it is tough to characterize these patterns manually or semi-automatically using geometric distances. Here, we propose a new methodology to represent and group individuals having similar folding patterns. We focus on the cingulate region, known to have a clinical interest, using so-called skeletons (3D representation of folding patterns). We compare two models, beta-VAE and SimCLR, in an unsupervised setting to learn a relevant representation of these patterns. Specifically, we leverage the data augmentations used in SimCLR to propose a novel kind of augmentations based on folding topology. Best clustering with Affinity Propagation has a silhouette score of 0.42. Comparison of cluster averages and interpolation in the latent space reveal new pattern structures, and test with the other half of the dataset demonstrates that the representation is stable. This structured representation shows that unsupervised learning can help in the discovery of still unknown patterns. We will gain further insights into folding patterns by using new priors in the unsupervised algorithms and integrating other brain data modalities. Code and experiments are available at github.com/neurospin-projects/2021_jchavas_lguillon_deepcingulate.

Keywords: beta-VAE, SimCLR, contrastive learning, folding pattern, cortex.

1. Introduction

The human cortex is convoluted, made of folds, called gyri, separated by grooves, the sulci. Contrary to macaque, whose cortical folding follows a systematic scheme, human cortex folding is highly variable, making it a fingerprint of each individual. Although this diversity seems, first, intractable, neuroanatomists have succeeded in defining a partially reproducible scheme, which has led to the nomenclature of sulci used in neuroscience (Ono et al., 1990). But each sulcus can have a large number of patterns, which hinders its reliable identification (Fig. 1B). Deep learning could be a real lever to deal with this tremendous inter-individual variability.

[*] Contributed equally

Past methods to map the various folding patterns rely on supervised learning to perform sulcus recognition and manifold learning based on pairwise geometric distances (Sun et al., 2012). They suffer from weaknesses of the sulcus recognition systems regarding unusual folding patterns and they do not scale up to the massive datasets available today.

This study aims to pave the way for unsupervised deep learning to systematize the detection of folding patterns across the cortex in the future. We aim to compare two unsupervised deep learning models in the task of obtaining a latent space structured enough to bring out folding patterns. To achieve this goal, we developed a deep learning pipeline that focuses on the folding pattern of predefined regions. We tested the pipeline on the cingulate region, as it is sufficiently variable to justify the use of our methods, and it has a clinical interest for psychiatric disorders (Yücel et al., 2003; Provost et al., 2003; Borst et al., 2014). Then, we chose, adapted and compared two powerful and standard unsupervised methods, namely a contrastive learning model, SimCLR (Chen et al., 2020), and a generative model, β-VAE (Higgins et al., 2017). We ensured that the learned latent representations are consistent and stable. Last, we proposed ways to analyze the results which are new and challenging with respect to classical deep learning literature as the input sample topologies are very different from classical 3D images.

2. Related works

2.1. Cortical folding patterns characterization

Studying the folding patterns can be done with various approaches. First, from the sulci, we can derive features, such as their depth, width, and length. Statistical analyzes can then reveal correlations with diverse pathologies. However, these "state" features are not stable throughout life (Cachia et al., 2016). Second, there are "trait" features (Cachia et al., 2016) that remain during lifespan. One such feature consists in the shape of the folding patterns. Some works tried to decipher folding patterns and identify the most common shapes. Historically, this was done visually (White et al., 1997), enabling to define central sulcus knob and the omega-shape of the mid-fusiform sulcus in particular (Yousry et al., 1997; Weiner et al., 2014). However, manually finding relevant geometrical shapes is very hard due to the high diversity of folding patterns.

Thus, some studies automated the characterization of folding patterns. A method to ease the process is first to use supervised learning to identify a specific sulcus in a dataset, then to use pairwise geometric similarity measures across sulci to perform the learning of a low dimensional manifold (Sun et al., 2012). This manifold summarizes the principal variability of the sulcus shape, from which the main patterns emerge and can correlate with pathology or behavior. In (Meng et al., 2018), the similarity measure relies on the sulcal pits, namely the locally deepest points of the cortical surface and they apply a clustering to identify the most representative patterns. (Duan et al., 2019) presented a similar method but applied on gyri rather than sulci. More recently, (Roy et al., 2020) trained neural network classifiers to map geometric shapes to folding patterns applied to the broken-H shape pattern in the orbitofrontal region. However, characterizing the full diversity of folding patterns remains out of reach for these automatic geometric methods. Unsupervised deep learning methods is a natural next step: they have been used for learning infant and neonatal cortical surface atlas (Cheng et al., 2020; Zhao et al., 2021) and for detecting

anomalies in folding patterns (Guillon et al., 2021), but they have not been used yet to characterize the normal inter-individual variability of folding patterns.

2.2. Unsupervised learning for features representation and clustering

Numerous approaches try to tackle unsupervised representation learning problems. On the one hand, auto-encoders (AE) are generative models that build a latent space comprising much fewer dimensions than the input, suggesting that the representations could be more easily understood, leading eventually to pattern discovery. For example, (Guillon et al., 2021) showed that β-VAE are promising to detect anomalies of folding patterns.

On the other hand, self-supervised methods, particularly contrastive learning models, have proved to be very powerful. The foundation contrastive model, SimCLR (Chen et al., 2020), learns to bring together in the output space views from the same input image (the positive pairs) and move them away from all other views of the same batch (the negative pairs). This method permits structuring the obtained latent space without using any labels. Its strength lies in the possibility to integrate prior information either by choosing the adapted random augmentations (to build the views) or by integrating into the loss function similarity information from other modalities.

Most of the self-supervised literature lies in computer vision domain but many works start to apply such framework to biomedical imaging. Thus, (Taleb et al., 2020) proposed 3D versions of several self-supervised tasks on various objectives including brain tumor segmentation. Self-supervised methods offer the opportunity to leverage additional prior information from medical data. For instance, (Dufumier et al., 2021b) applied contrastive learning to brain MRI and took advantage of available meta-data such as age and sex. (Zhang et al., 2017) used 2D slices ordering of 3D CT images as a pretext task. This accelerating research on nearby fields shows that it is the right moment to apply self-supervised learning to the folding pattern characterization problem.

3. Methods

3.1. Pre-processing

From brain MRI images, we used skeletons whose concept was first introduced in (Mangin et al., 1995). They consist in 3D images of the cortical folds obtained with BrainVISA/Morphologist preprocessing pipeline (`https://brainvisa.info/`). Figure 1A represents crops of such skeletons. Skeletons' voxels are divided into background and folds. Fold voxels can hold several values depending on their topological meaning (fold bottom, fold junction, etc.). Using this input enables to focus on the folding geometry and eliminates some biases such as age or site. Dealing only with the geometry avoids taking into account the classical sulci widening that comes with aging for instance.

We focus our study on the cingulate region of the right hemisphere (Fig. 1A). We learned a mask of the cingulate and paracingulate sulci over a database where the folds were manually labelled (Borne et al., 2020). In short, labeled subjects were first affinely normalized to a standard brain referential (ICBMc2009); then, each subject voxel belonging to the sulci of interest increments a sulcus-specific mask. We combined and dilated these two resulting masks to get a simple Region of Interest (ROI). We then applied this final

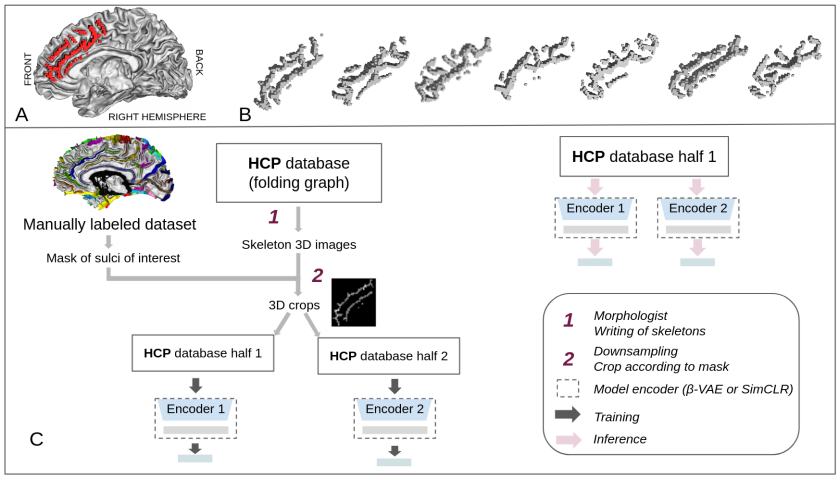

Figure 1: *Pipeline to study skeletons in the cingulate region of the HCP dataset.* **A)** Sample crop of the cingulate region represented as voxel sets superimposed with the white matter mesh. **B)** Samples of the studied crops, given as inputs to the unsupervised algorithms. **C)** Whole pipeline. Left: we generate crops of the cingulate region based on a manually labeled dataset. We divide the HCP dataset in half and train both models (β-VAE and SimCLR) on the two halves HCP_1 and HCP_2. Right: we infer and perform downstream analysis of all four models on HCP_1.

mask to skeleton images of any unlabeled brain. Our final input is a 2-mm resolution 3D crop of dimension 20x40x40 (Fig. 1B) with integer values representing local topologies.

3.2. Learning cingulate region representations

We compared two unsupervised deep learning models : an autoencoder-based model and a contrastive learning framework.

β-VAE. AE-based models are commonly used to learn representations and to model the inter-subject variability. With an encoder θ, they enable to project data from input space $\mathcal{X}$ onto a latent space $\mathcal{Z}$ comprising much fewer dimensions. The latent code is then reconstructed thanks to a decoder ϕ. β-VAE (Higgins et al., 2017), an extension of VAE (Kingma and Welling, 2014), is particularly interesting as the latent space is constrained to follow a prior distribution and input data are encoded as a distribution. The objective function is a combination of the reconstruction error and the matching of two distributions using the Kullback-Liebler (KL) divergence. β-VAE is trained to maximize:

$$\mathcal{L}(\theta, \phi; \mathbf{x}, \mathbf{z}, \beta) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log\, p_\theta(\mathbf{x}|\mathbf{z})] - \beta \mathcal{D}_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z})) \tag{1}$$

where $p(\mathbf{z})$ corresponds to the prior distribution (here, a reduced centered Gaussian distribution) and $q_\phi(\mathbf{z}|\mathbf{x})$, the posterior distribution. We ran the model on binarized skeletons.

SimCLR. SimCLR is an instance discrimination contrastive model. For each sample x of the batch of size N, we generate at each epoch two views x_i and x_j, whose model

outputs are respectively z_i and z_j. The model trains to bring together views from the same image, that is to minimize $\sum_{i=1}^{N} \ell_{i,j=pos(i)} + \sum_{j=1}^{N} \ell_{j,i=pos(i)}$, $\ell_{i,j}$ being the loss function for a positive pair of examples (τ is a temperature parameter) :

$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1,k\neq i}^{2N} \exp(\text{sim}(z_i, z_k)/\tau)}, \tag{2}$$

View generations are the algorithms specific to our problem: they use the discrete topology of the fold skeleton. For each fold, the bottom line voxels can be distinguished from the inner part of the fold surface because they do not split the skeleton background into two different local connected components (Mangin et al., 1995), see also Appendix A. Then, the bottom line tag permits to define a topology-based augmentation, which conserves the bottom lines in all views but removes the inner part of some folds. The first view combines random [-10,10]° rotations over all axes and a 60% rolling cutout with only bottom lines kept inside the cutout volume. The second view combines random [-10,10]° rotations over all axes followed by a 60% rolling cutout with the whole skeleton conserved inside the cutout whereas only bottom values are kept outside the cutout volume. All views are then binarized. This topology-based augmentation forces the model to learn the sheet-based structure of the fold-based skeleton.

3.3. Identifying folding patterns

Characterizing folding shapes. To identify folding patterns, data are encoded to the latent space of both models and reduced to a 2-dimensions space with t-SNE algorithm. The reduction to two dimensions enables to get more hints of the learned representations and to analyze subjects groups more easily. A clustering is then performed with hierarchical affinity propagation (AP) algorithm (Frey and Dueck, 2007). One advantage of AP is that the number of expected clusters does not have to be precised. However it may output a very large number of clusters, making it difficult to understand from an anatomical point of view. Hence, following the method used in (Meng et al., 2018), we applied the algorithm in an iterative way until a maximum number of five clusters is found.

The analysis of the main anatomical characteristics of the clusters can be done on the latent codes or on the input space based on cluster labels. The first method is specific to the β-VAE and enables to understand the encoded characteristics in the latent dimensions. We generated images corresponding to clusters' centroids from their latent codes which are next decoded. Then we travelled between clusters through the latent space to analyse variations across dimensions. The second method computes the local *per*-cluster averaging pattern in the input space (Sun et al., 2012).

Ensuring representation generalization. To test the generalization ability of our latent representations, we trained both models (β-VAE and SimCLR) twice, on two different datasets leading to two encoders per model. We refer to them as E1 and E2. We then encoded the first dataset with the two encoders (Fig. 1C). The different embeddings are reduced to a lower dimension space with t-SNE. E1 embeddings are clustered and we report the labels to E2 embeddings. We assess the generalization ability both visually and quantitatively: visually, we assess if the first visualization (E1) remains localized in the

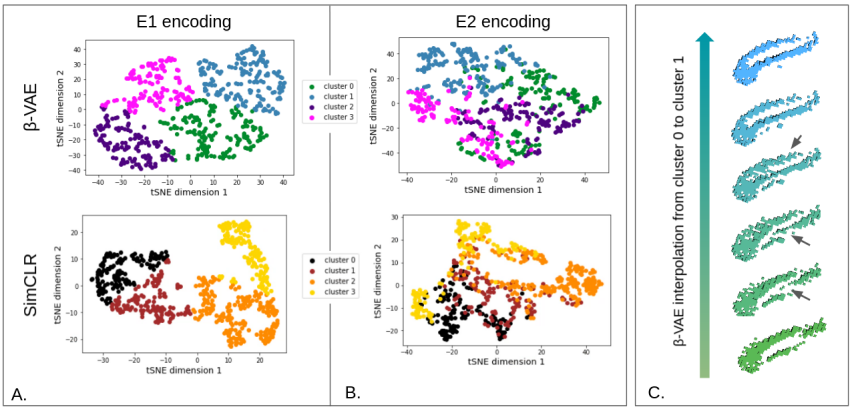

Figure 2: *β-VAE and SimCLR latent spaces analysis.* For both models, training is made on HCP_1 (model 1) and HCP_2 (model 2). The inference on HCP_1 leads to the encodings E1 and E2, encodings of the latent space obtained respectively with model 1 and model 2. To visualize the stability of the model and of its analysis, the t-SNE representation of model 1 is clustered (**A.**) and we report the labels on the t-SNE representation of model 2 (**B.**). Lowest and uppermost patterns of (**C.**) are respectively the decoded latent code of cluster 0 and 1 centroids of the β-VAE model. Intermediate patterns are obtained traveling through the latent space and then decoded.

second visualization (E2), meaning that the learned representations are not dependent on the training data but have captured some general features; quantitatively we measure the adjusted mutual information score (AMI) (Vinh et al., 2010).

4. Experiments and results

Datasets. We use HCP database[1] in which MRI images were obtained with a Siemens Skyra Connectom scanner with isotropic resolution of 0.7 mm. We focused only on the right hemisphere of the 1101 subjects. As presented on Fig. 1C, we divided equally HCP into two subsets HCP_1 and HCP_2. In both subsets, 80% of subjects were used for training and the remaining 20% were used for validation.

Model implementation. Our β-VAE comprises a fully convolutional encoder and decoder of symmetrical architectures with three convolutional blocks and two fully connected layers. The backbone of our SimCLR model is the DenseNet (Huang et al., 2017), followed by two fully connected projection overheads based on (Dufumier et al., 2021a) benchmark on 3D MRI images. To adapt to our smaller input, we reduced the size of the DenseNet network down to two dense blocks. We call latent space, the representation space of the SimCLR model, which has a better representation quality than output space (Chen et al., 2020). Hyperparameter choice method and implementation details are presented in Appendix F.

1. https://www.humanconnectome.org/

Grid search gave a latent space size of 4 for both models.

Latent space structure. Fig. 2A presents clustering results obtained on HCP_1 using embeddings E1. The silhouette score with AP on the latent space is 0.17 and 0.42, respectively for β-VAE and SimCLR. It becomes 0.43 and 0.44 when applied to the t-SNE space, indicating a tendency towards a clustered distribution with close clusters. This range of score is common when dealing with complex data such as neuroimaging modalities (Lebenberg et al., 2019). For both models, four clusters were identified but the organization of the latent space is different: β-VAE latent space seems to distinguish four groups of subjects, separated only with a thin boundary whereas SimCLR latent space is more structured and could be interpreted as a manifold, consistent with the biological reality of folding patterns. The comparison of the two clusterings (Fig. 2A and B) shows that some clusters remain more stable than others. With β-VAE embeddings, subjects of the blue cluster are still grouped. With SimCLR, the orange cluster remains, whereas the yellow cluster is split in two. But overall, clusters' subjects remain close. These observations are confirmed by the AMI score, which is 0.37 for the $\beta - VAE$ and 0.31 for SimCLR.

Deciphering the patterns. Figure 2C shows generations based on the centroids' codes. Generated sulcus patterns differ in the two clusters: cluster 0 shows a split cingulate sulcus, whereas cluster 1 seems to present a cingulate sulcus and a long paracingulate. Both patterns are described in the literature (Meng et al., 2018). Interpolating from one cluster to another shows that the latent space is continuous and regular, and we can progressively see the change of patterns as indicated by the arrows. More detailed and complete interpolations are presented in Appendix B and C.

Figure 3 presents the cluster average of the folding pattern based on the input space. For both models, we observe different shapes depending on the clusters. The study of the shapes informs on what could be encoded in the latent space. For β-VAE, we can identify specificities for each one of the averages. Blue average seems to be the simplest pattern, a long cingulate without paracingulate or vertical branches. Indigo presents a sulcus parallel to the cingulate divided into several pieces that may look like a sketch of the paracingulate sulcus. In return, the pink average includes several branches, vertical to the cingulate, that could not merge to create a parallel sulcus. Lastly, the green shows a pattern defined in the literature (Meng et al., 2018), a split anterior cingulate sulcus.

It is interesting to link these observations to the patterns generated thanks to the $\beta - VAE$ reconstructions. We find a similar shape for the green average (cluster 0): cingulate split in two. Conversely, for the blue cluster, based only on the average pattern, we interpreted a simple cingulate, but in the light of the reconstructions, the swollen anterior part could represent a paracingulate. Cingulate and paracingulate could be merged in the average representation due to positional variations among subjects.

The structure of SimCLR latent space makes the interpretation also interesting as it is organized as a manifold with continuous evolution. Here, the black average could correspond to a simple anterior cingulate. Subjects of the brown cluster could have a sketch of the paracingulate sulcus, which increases in length in the orange average to present two long parallel sulci. Finally, the yellow average also includes a sketch of a sulcus parallel to

Figure 3: *Representative patterns as cluster averages.* **A)** Description of typical folding structures in our ROI using the icbm152 average template. **B), C)** Local average sulci obtained for each cluster with β-VAE and SimCLR encodings respectively. Colors match cluster colors of Fig. 2. Averages are done on HCP_1.

the anterior cingulate, but in the left part of the ROI, where it is not usually called a paracingulate sulcus in anatomical literature.

5. Discussion and conclusion

Our work proposes several method contributions that can be useful for the community. We introduced topology-based augmentations in the SimCLR setting, which is directly applicable for studies working on skeletons or similar inputs (Harrison et al., 2021; Rao et al., 2021). Moreover, we used for the first time local average folding patterns (Sun et al., 2012) in a Deep learning pipeline. Last, we proposed a preprocessing based on a mask, enabling to focus on the region of interest, while avoiding the disadvantages of parallelepipedic bounding boxes used in (Guillon et al., 2021).

Our work also finds a structured latent space for the cingulate region with both models, β-VAE and SimCLR. The organization obtained with SimCLR seems more consistent with anatomical reality of folding patterns and can be linked to folding manifolds (Mangin et al., 2016). In return, the generative aspect of the $\beta - VAE$ is a real lever to understand the learned representations and ease the analysis of this complex region. To encourage a structured and well separated latent space, we wish to introduce in future cluster objectives in the learning phase both for generative models (Danks and Yau, 2021), and for contrastive models (Caron et al., 2018, 2021; Li et al., 2021).

Another line of research will be to adapt our model further to the folding topology by developing other topology-based augmentations and by introducing other specific priors such as the geometry-based similarity measure between input samples (Sun et al., 2012).

Finally, we found cluster averages to be similar to known cingulate patterns that correlate with executive functions and psychiatric disorders. This similarity makes us firmly believe that such latent space structures could correlate with medically relevant parameters. This correlation will be the object of a follow-up study.

Acknowledgments

We thank colleagues from the Deep learning journal club for thorough discussions. We thank Zaccharie Ramzi for helping set up rotation augmentations. This research received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 945539 (HBP SGA3), from the FRM DIC20161236445, the ANR-19-CE45-0022-01 IFOPASUBA, the ANR-20-CHIA-0027-01 FOLDDICO. Data were provided in part by the Human Connectome Project funded by the NIH.

References

Léonie Borne, Denis Rivière, Martial Mancip, and Jean-François Mangin. Automatic labeling of cortical sulci using patch- or CNN-based segmentation techniques combined with bottom-up geometric constraints. *Medical Image Analysis*, 62:101651, May 2020. ISSN 1361-8415.

G. Borst, A. Cachia, J. Vidal, Grégory Simon, A. Pineau, C. Fischer, N. Poirel, J.-F. Mangin, and O. Houdé. Folding of the anterior cingulate cortex partially explains inhibitory control during childhood: A longitudinal study. *Developmental Cognitive Neuroscience*, 9:126–135, July 2014. Publisher: Elsevier.

A. Cachia, G. Borst, C. Tissier, C. Fisher, M. Plaze, O. Gay, D. Rivière, N. Gogtay, J. Giedd, J. F. Mangin, O. Houdé, and A. Raznahan. Longitudinal stability of the folding pattern of the anterior cingulate cortex during development. *Developmental Cognitive Neuroscience*, 19:122–127, June 2016. ISSN 1878-9293.

Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep Clustering for Unsupervised Learning of Visual Features. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, Lecture Notes in Computer Science, pages 139–156, Cham, 2018. Springer International Publishing. ISBN 978-3-030-01264-9.

Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. *arXiv:2006.09882 [cs]*, January 2021. URL http://arxiv.org/abs/2006.09882. arXiv: 2006.09882.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations. In *Proceedings of the 37th International Conference on Machine Learning*, pages 1597–1607. PMLR, November 2020. ISSN: 2640-3498.

Jieyu Cheng, Adrian V. Dalca, and Lilla Zöllei. Unbiased Atlas Construction for Neonatal Cortical Surfaces via Unsupervised Learning. In Yipeng Hu, Roxane Licandro, J. Alison Noble, Jana Hutter, Stephen Aylward, Andrew Melbourne, Esra Abaci Turk, and Jordina Torrents Barrena, editors, *Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis*, Lecture Notes in Computer Science, pages 334–342, Cham, 2020. Springer International Publishing. ISBN 978-3-030-60334-2. doi: 10.1007/978-3-030-60334-2_33.

Dominic Danks and Christopher Yau. BasisDeVAE: Interpretable Simultaneous Dimensionality Reduction and Feature-Level Clustering with Derivative-Based Variational Autoencoders. pages 2410–2420. PMLR, 2021. URL https://proceedings.mlr.press/v139/danks21a.html.

Dingna Duan, Shunren Xia, Islem Rekik, Yu Meng, Zhengwang Wu, Li Wang, Weili Lin, John H. Gilmore, Dinggang Shen, and Gang Li. Exploring folding patterns of infant cerebral cortex based on multi-view curvature features: Methods and applications. *NeuroImage*, 185:575–592, January 2019. ISSN 1053-8119.

Benoit Dufumier, Pietro Gori, Ilaria Battaglia, Julie Victor, Antoine Grigis, and Edouard Duchesnay. Benchmarking CNN on 3D Anatomical Brain MRI: Architectures, Data Augmentation and Deep Ensemble Learning. *arXiv:2106.01132 [cs, eess]*, June 2021a. URL http://arxiv.org/abs/2106.01132. arXiv: 2106.01132.

Benoit Dufumier, Pietro Gori, Julie Victor, Antoine Grigis, Michele Wessa, Paolo Brambilla, Pauline Favre, Mircea Polosan, Colm McDonald, Camille Marie Piguet, Mary Phillips, Lisa Eyler, and Edouard Duchesnay. Contrastive Learning with Continuous Proxy Meta-Data For 3D MRI Classification. In *MICCAI 2021 - Accepted Papers and Reviews*, Strasbourg (virtuel), France, September 2021b.

Brendan J. Frey and Delbert Dueck. Clustering by Passing Messages Between Data Points. *Science*, 315(5814):972–976, February 2007. ISSN 0036-8075, 1095-9203.

Louise Guillon, Bastien Cagna, Benoit Dufumier, Joël Chavas, Denis Rivière, and Jean-François Mangin. Detection of Abnormal Folding Patterns with Unsupervised Deep Generative Models. In Ahmed Abdulkadir, Seyed Mostafa Kia, Mohamad Habes, Vinod Kumar, Jane Maryam Rondina, Chantal Tax, and Thomas Wolfers, editors, *Machine Learning in Clinical Neuroimaging*, Lecture Notes in Computer Science, pages 63–72, Cham, 2021. Springer International Publishing. ISBN 978-3-030-87586-2.

Josquin Harrison, Marco Lorenzi, Benoit Legghe, Xavier Iriart, Hubert Cochet, and Maxime Sermesant. Phase-Independent Latent Representation for Cardiac Shape Analysis. In Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, Lecture Notes in Computer Science, pages 537–546, Cham, 2021. Springer International Publishing. ISBN 978-3-030-87231-1. doi: 10.1007/978-3-030-87231-1_52.

I. Higgins, L. Matthey, A. Pal, Christopher P. Burgess, Xavier Glorot, M. Botvinick, S. Mohamed, and Alexander Lerchner. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. In *ICLR*, 2017.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, July 2017. ISSN: 1063-6919.

Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. *ICLR*, May 2014. URL http://arxiv.org/abs/1312.6114. arXiv: 1312.6114.

J. Lebenberg, J. F. Mangin, B. Thirion, C. Poupon, L. Hertz-Pannier, F. Leroy, P. Adibpour, G. Dehaene-Lambertz, and J. Dubois. Mapping the asynchrony of cortical maturation in the infant brain: A MRI multi-parametric clustering approach. *NeuroImage*, 185: 641–653, January 2019. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2018.07.022. URL https://www.sciencedirect.com/science/article/pii/S1053811918306372.

Yunfan Li, Peng Hu, Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng. Contrastive Clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(10):8547–8555, May 2021. ISSN 2374-3468. Number: 10.

J.-F. Mangin, J. Lebenberg, S. Lefranc, N. Labra, G. Auzias, M. Labit, M. Guevara, H. Mohlberg, P. Roca, P. Guevara, J. Dubois, F. Leroy, G. Dehaene-Lambertz, A. Cachia, T. Dickscheid, O. Coulon, C. Poupon, D. Rivière, K. Amunts, and Z.Y. Sun. Spatial normalization of brain images and beyond. *Medical Image Analysis*, 33:127–133, October 2016. ISSN 13618415.

Jean-Francois Mangin, Vincent Frouin, Isabelle Bloch, Jean Rigis, and Jaime Lopez-Krahe. From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations. *Journal of Mathematical imaging and Vision*, 5(4):297–318, 1995.

Yu Meng, Gang Li, Li Wang, Weili Lin, John H Gilmore, and Dinggang Shen. Discovering cortical sulcal folding patterns in neonates using large-scale dataset. *Human Brain Mapping*, 39(9):3625–3635, April 2018. ISSN 1065-9471.

Michio Ono, Stefan Kubik, and Chad D. Abarnathey. *Atlas of the Cerebral Sulci*. Thieme-Stratton Corp, Stuttgart : New York, 1er édition edition, January 1990. ISBN 978-0-86577-362-2.

Jean-Bernard Le Provost, David Bartrés-Faz, Marie-Laure Paillère-Martinot, Eric Artiges, Sabina Pappata, Christophe Recasens, Mercedes Pérez-Gómez, Miquel Bernardo, Imma Baeza, Frank Bayle, and Jean-Luc Martinot. Paracingulate sulcus morphology in men with early-onset schizophrenia. *British Journal of Psychiatry*, 182(3):228–232, March 2003. ISSN 0007-1250, 1472-1465.

Haocong Rao, Siqi Wang, Xiping Hu, Mingkui Tan, Yi Guo, Jun Cheng, Xinwang Liu, and Bin Hu. A Self-Supervised Gait Encoding Approach with Locality-Awareness for 3D Skeleton Based Person Re-Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2021. ISSN 0162-8828, 2160-9292, 1939-3539. doi: 10.1109/TPAMI.2021.3092833. URL http://arxiv.org/abs/2009.03671. arXiv: 2009.03671.

Arnab Roy, Tyler McMillen, Donielle L. Beiler, William Snyder, Marisa Patti, and Vanessa Troiani. A pipeline to characterize local cortical folds by mapping them to human-interpretable shapes. Technical report, November 2020. URL https://www.biorxiv.org/content/10.1101/2020.11.25.388785v1. Company: Cold Spring Harbor Laboratory Distributor: Cold Spring Harbor Laboratory Label: Cold Spring Harbor Laboratory Section: New Results Type: article.

Zhong Yi Sun, Stefan Klöppel, Denis Rivière, Matthieu Perrot, Richard Frackowiak, Hartwig Siebner, and Jean-François Mangin. The effect of handedness on the shape of the central sulcus. *NeuroImage*, 60(1):332–339, March 2012. ISSN 1053-8119.

Aiham Taleb, Winfried Loetzsch, Noel Danz, Julius Severin, Thomas Gaertner, Benjamin Bergner, and Christoph Lippert. 3D Self-Supervised Methods for Medical Imaging. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 18158–18172. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/d2dc6368837861b42020ee72b0896182-Paper.pdf.

Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance. *Journal of Machine Learning Research*, 11(95):2837–2854, 2010. ISSN 1533-7928. URL http://jmlr.org/papers/v11/vinh10a.html.

Kevin S. Weiner, Golijeh Golarai, Julian Caspers, Miguel R. Chuapoco, Hartmut Mohlberg, Karl Zilles, Katrin Amunts, and Kalanit Grill-Spector. The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *NeuroImage*, 84:453–465, January 2014. ISSN 1095-9572.

L. E. White, T. J. Andrews, C. Hulette, A. Richards, M. Groelle, J. Paydarfar, and D. Purves. Structure of the human sensorimotor system. I: Morphology and cytoarchitecture of the central sulcus. *Cerebral Cortex (New York, N.Y.: 1991)*, 7(1):18–30, February 1997. ISSN 1047-3211.

T A Yousry, U D Schmid, H Alkadhi, D Schmidt, A Peraud, A Buettner, and P Winkler. Localization of the motor hand area to a knob on the precentral gyrus. A new landmark. *Brain*, 120(1):141–157, January 1997. ISSN 0006-8950.

Murat Yücel, Stephen J. Wood, Lisa J. Phillips, Geoffrey W. Stuart, Deidre J. Smith, Alison Yung, Dennis Velakoulis, Patrick D. Mcgorry, and Christos Pantelis. Morphology of the anterior cingulate cortex in young men at ultra-high risk of developing a psychotic illness. *The British Journal of Psychiatry*, 182(6):518–524, June 2003. ISSN 0007-1250, 1472-1465. Publisher: Cambridge University Press.

Pengyue Zhang, Fusheng Wang, and Yefeng Zheng. Self supervised deep representation learning for fine-grained body part recognition. *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017.

Fenqiang Zhao, Zhengwang Wu, Li Wang, Weili Lin, Shunren Xia, and Gang Li. Learning 4D Infant Cortical Surface Atlas with Unsupervised Spherical Networks. In Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, Lecture Notes in Computer Science, pages 262–272, Cham, 2021. Springer International Publishing. ISBN 978-3-030-87196-3. doi: 10.1007/978-3-030-87196-3_25.

Appendix A. Illustration of skeleton topologies

Figure 4 illustrates what represents the skeleton and the corresponding bottom lines. Skeleton lies between gyri. The corresponding bottom line follows the sheet structure of the skeleton. We can see from this figure that the bottom line can be a good proxy of the folding shape.

In the topology-based augmentations, only the orange part is deleted in different parts of the volume.

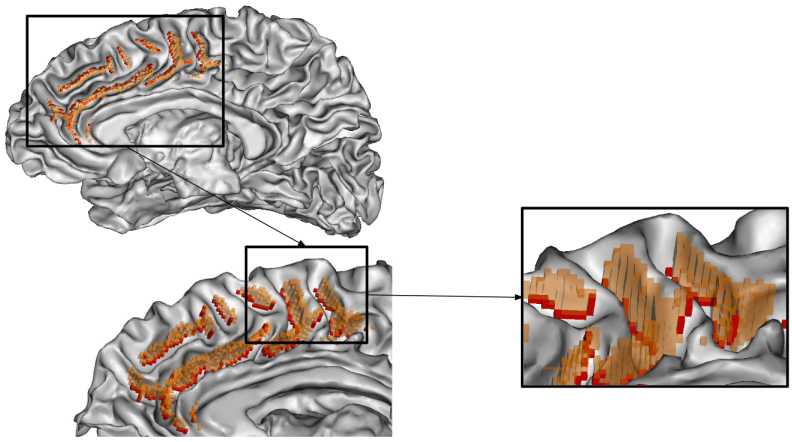

Figure 4: *Understanding bottom lines of skeletons.* Orange represents the skeleton of the ROI. Red represents the corresponding bottom line. Grey is the white matter. Skeleton is represented here on 1 mm resolution. Note that the skeleton includes its bottom line, we use two colors to facilitate understanding.

Appendix B. Visualization of β-VAE learned latent representation

Figure 5 presents an analysis and visualization of β-VAE learned latent representation. Between the blue cluster (C0) and the green one (C1), the intermediate pattern shows a split of the cingulate sulcus. It demonstrates that the latent space is continuous and that the four dimensions of the latent space have encoded anatomically relevant features. More detailed variations interpolations are presented in figure 6 of Appendix C.

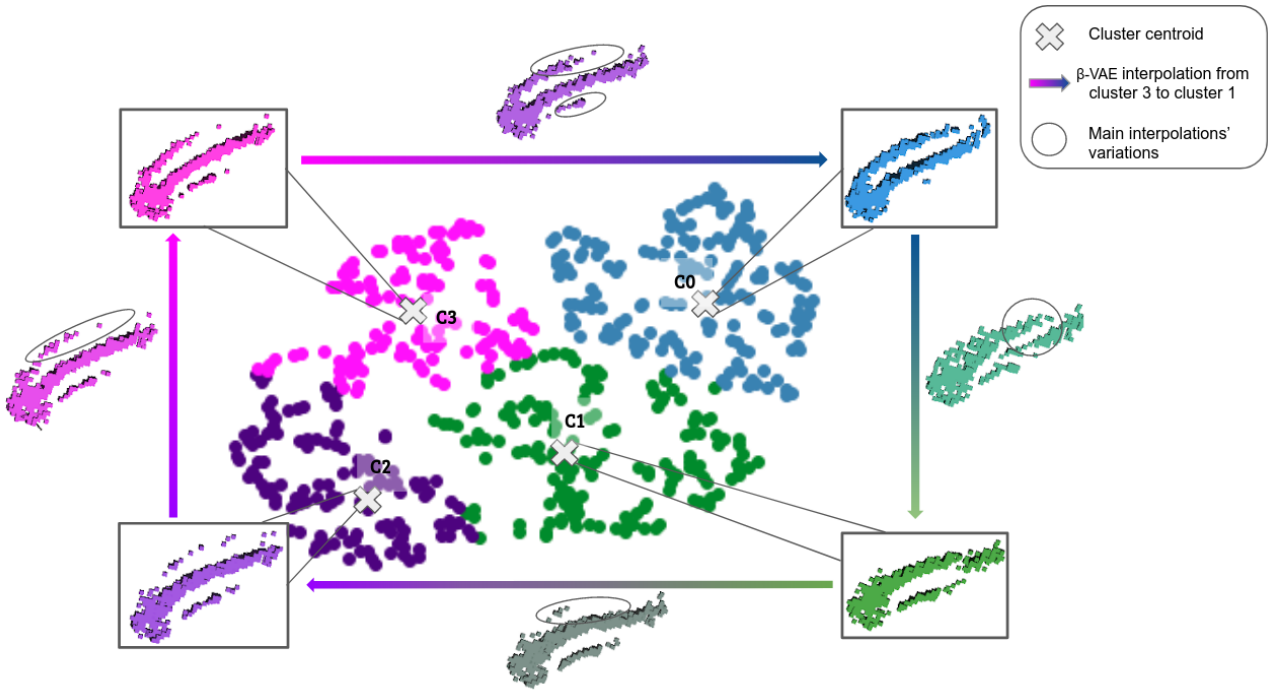

Figure 5: *Visualisation of β-VAE learned clusters' representations.* Framed patterns correspond to generations of centroids' latent representation thanks to the β-VAE decoder part. Between two centroids C_i and C_j, an intermediate point is sampled and decoded in order to highlight variations between clusters.

Appendix C. Detailed β-VAE interpolation

Figure 6 illustrates in details the interpolations between each pair of cluster travelling through $\beta - VAE$ latent space.

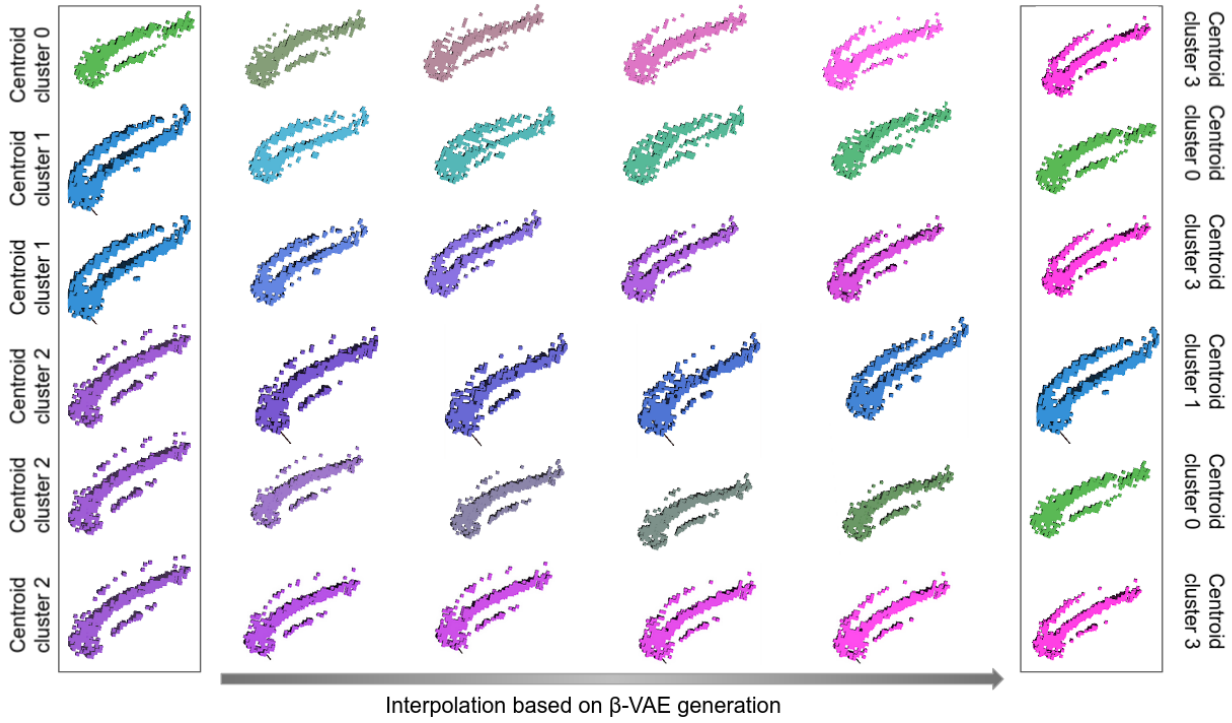

Figure 6: *Travelling through β-VAE latent space from one cluster to another.* The extreme left and right columns represent patterns generated from the centroids latent codes, the colors refers to Fig.2. Between centroids patterns, new sampled obtained travelling through $\beta - VAE$ latent space illustrate the variations from one cluster's particularities to another.

Appendix D. Closest and furthest examples in the output space of SimCLR

Figure 7 illustrates four closest pairs and four furthest pairs for SimCLR, using the similarity measure of SimCLR in the output space.

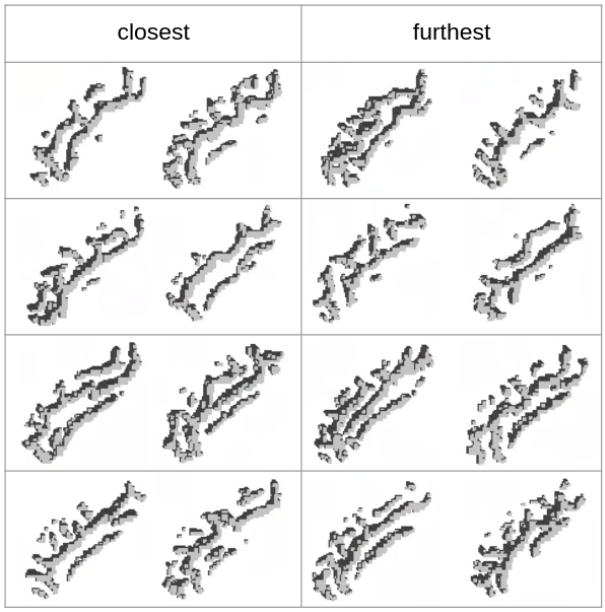

Figure 7: *Closest and furthest examples in the output of SimCLR model.* The left column and right column represent respectively four pairs of closest examples and four pairs of furthest examples using as distance the similarity between output encodings of SimCLR model.

Closest pairs demonstrate similarities such as the cingulate shape, the number of parallel sulci for the third instance for example, or a similar complexity of sulcus shapes. Furthest pairs seem indeed less alike than the closest.

Appendix E. Closest and furthest examples in the latent space of β-VAE and SimCLR

Figure 8 illustrates closest and furthest examples in the latent space as encoded by both models.

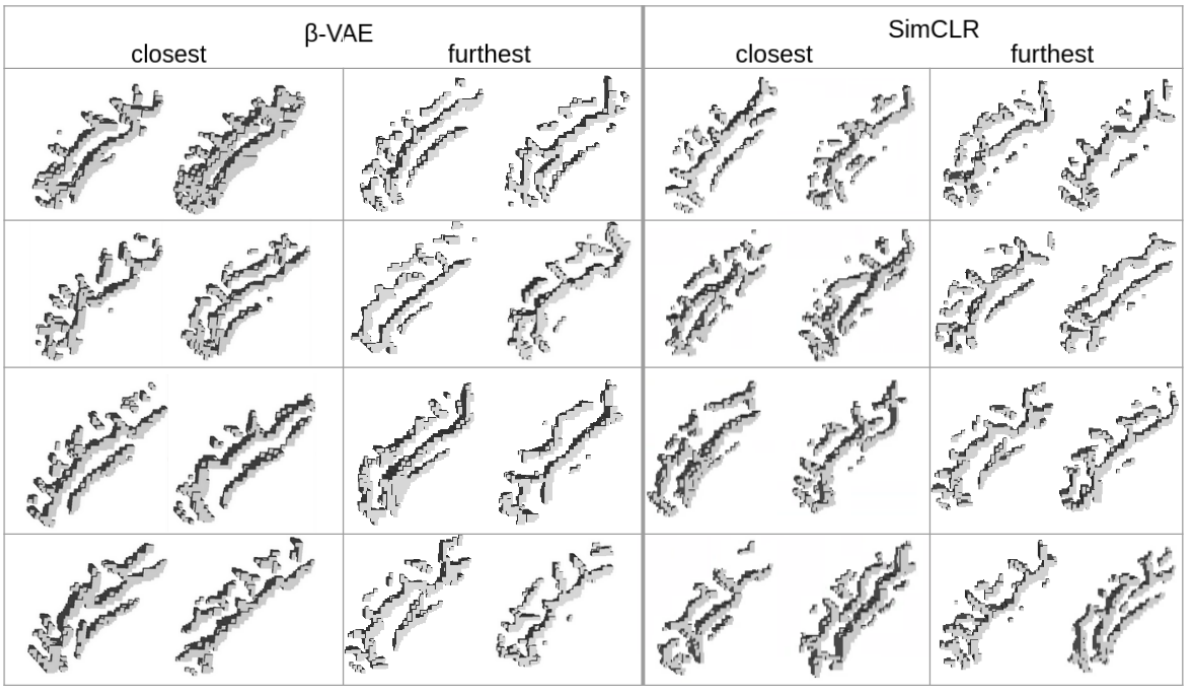

Figure 8: *Closest and furthest examples in latent space of both β-VAE and SimCLR.* The two left columns represent closest and furthest examples for the β-VAE model. Similarly, the two right columns represent closest ad furthest examples for the SimCLR model. In both models, the distance is the Euclidean distance in the latent space.

For both models, closest and furthest crops were determined based on their Euclidean distance in the latent space.

Though it is tough to characterize folding patterns visually, it seems that for β-VAE, the proximity of samples is based on the similarity of the cingulate shape and the likeness of vertical branches. For SimCLR, closest pairs demonstrate similarities such as the cingulate shape, the merging of the paracingulate in the cingulate for the second pair for instance. Note that the first pair was also classified as "closest" in the output space (Appendix D). In both cases, furthest pairs seem indeed less alike than the closest. However, these visualizations remind us of the challenge of interpreting such complex patterns visually.

Appendix F. Hyperparameter tuning and implementation details

To find the best hyperparameters (size of the latent space for both models, β value for β-VAE and temperature τ for SimCLR), we performed a gridsearch on HCP_1 where the best combination is chosen based on the loss value, the silhouette score on the latent space and the reconstruction abilities for β-VAE. We obtained β=2 and τ=0.1. For both models, we selected a latent size of 4, which enabled to balance between the model performance and the clustering quality. Training of 300 epochs lasted for approximately 1 hour and 2 hours for β-VAE and SimCLR respectively, on an Nvidia Quadro RTX5000 GPU.