A Framework for Generating 3D Shape Counterfactuals

Rajat Rasal[1] RRR2417@IMPERIAL.AC.UK
Daniel C. Castro[1,2] DACOELH@MICROSOFT.COM
Nick Pawlowski[1,2] NICK.PAWLOWSKI@MICROSOFT.COM
Ben Glocker[1] B.GLOCKER@IMPERIAL.AC.UK
[1] *Department of Computing, Imperial College London, UK*
[2] *Microsoft Research, Cambridge, UK*

Editors: Jelmer Wolterink, Angelica I. Aviles-Rivero, Erik Bekkers

Abstract

Many important problems in medical imaging require analysing the causal effect of genetic, environmental, or lifestyle factors on the normal and pathological variation of anatomical phenotypes. There is, however, a lack of computational tooling to enable causal reasoning about morphological variations of 3D surface meshes. To tackle this problem, we present the framework of deep structural causal shape models (CSMs) using a database of subcortical brain meshes. CSMs enable subject-specific prognoses through counterfactual mesh generation, by utilising high-quality mesh generation techniques, from geometric deep learning, within the expressive framework of deep structural causal models (DSCM).

Keywords: Causality, 3D Shape Models, Counterfactuals, Medical Imaging

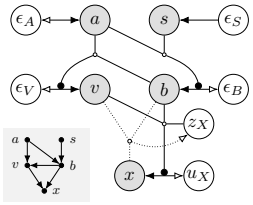
1. Introduction

The causal modelling of non-Euclidean structures is a problem which machine learning research has yet to tackle[1]. Thus far, state-of-the-art deep causal structure learning frameworks have been employed to model the data generation process of 2D images (Pawlowski et al., 2020; Vlontzos et al., 2022; Sauer and Geiger, 2021). To the best of our knowledge, none have been applied to non-Euclidean data such as 3D surface meshes, an important data structure for medical imaging applications. In this work, we present the first causally grounded 3D shape model, called a deep structural causal shape model (CSM), utilising geometric deep learning components (Bronstein et al., 2017). We use our CSM to model the causal data generation process for 3D brain stem meshes assumed in Fig. 1. Namely, we illustrate the CSM's ability to robustly answer subject-specific, retrospective, hypothetical questions, also known as counterfactuals i.e. "How would *this* patient's brain structure change if they had been ten years older or from the opposite (biological) sex?" (Fig. 2), by the process of deep counterfactual mesh inference.

2. Deep Structural Causal Shape Models

Applying DSCM framework. Our deep structural causal shape model (CSM) utilises the DSCM framework (Pawlowski et al., 2020) to implement the data generating functions in Fig. 1 using approximately invertible neural networks. Due to the correspondence between data generating functions and probabilistic graphical models (Pearl, 2009),

1. Refer to Rasal et al. (2022) for more results and details.



$$
\begin{aligned}
a &:= f_A(\epsilon_A) \\
s &:= f_S(\epsilon_S) \\
b &:= f_B(\epsilon_B; a, s) \\
v &:= f_V(\epsilon_V; a, b) \\
x &:= f_X(\epsilon_X; v, b), \quad \epsilon_X = \text{Concat}(u_X, z_X)
\end{aligned}
$$

Figure 1: Computational graph of our CSM (**left**), graphical model of the joint $p(x, v, b, a, s)$ (**bottom-left**) and its corresponding functional form (**right**). Variables are brain stem mesh (x), age (a), sex (s), and total brain (b) and brain stem (v) volumes, and all exogenous variable $\epsilon \sim \mathcal{N}(0, I)$. z_X and u_X encode a mesh's shape and scale. Reproduced with permission from Pawlowski et al. (2020).

generating medical data from our CSM amounts to sampling an estimator of the joint $p(x, v, b, a, s)$. This is clear when considering the conditional factorisation $p(x, v, b, a, s) = p(x|v, b) \cdot p(v|a, b) \cdot p(b|s, a) \cdot p(a) \cdot p(s)$, in which each distribution clearly corresponds to a function in our data generating process, e.g. $x = f_X(\epsilon_X; v, b)$ is equivalent to $x \sim p(x|v, b)$, $v = f_V(\epsilon_V; a, b)$ is equivalent to $v \sim p(v|a, b)$, and so on. This can be rewritten using the chain rule of probability as $p(x, v, b, a, s) = p(x|v, b) \cdot p(v, b, a, s)$.

Objective. We then derive a lower bound on $\log p(x, v, b, a, s)$, in Appendix A, as

$$
\log p(x, v, b, a, s) \geq \alpha + \underbrace{\mathbb{E}_{q(z_X|x,v,b)}[\log p(x|z_X, v, b)] - D_{KL}[q(z_X|x, v, b)\|p(z_X)]}_{\beta}, \tag{1}
$$

which, when maximised, jointly learns all data generating functions in the CSM. Here, $\alpha = \log p(v, b, a, s)$, $q(z_X|x, v, b)$ is a variational posterior and $p(z_X) = \mathcal{N}(0, I)$. β learns a mesh conditional variational autoencoder (Sohn et al., 2015) within the CSM structure, whose encoder $q(z_X|x, v, b)$ and decoder $p(x|z_X, v, b)$ are parametrised by the neural networks (Appendix B). These are implemented using sequences of spectral graph convolutions with quadric subsampling, akin to Ranjan et al. (2018).

Deep Counterfactual Mesh Inference. For an individual m, a counterfactual mesh $x_{m,\text{cf}}$ can be generated from the observation $(x_m, v_m, b_m, a_m, s_m)$ as follows:

1. **Abduction:** Calculate subject-specific mesh features $\epsilon_{X,m}$ by roughly performing the inversion $\epsilon_{X,m} = (u_{X,m}, z_{X,m}) = f_m^{-1}(x_m; v_m, b_m)$; first sample $z_{X,m} \sim q(z_{X,m}|x_m, v_m, b_m)$, then find $u_{X,m}$ by inverting the function $u_X \to x$ in the computational graph. We also calculate $\epsilon_{V,m} = f_V^{-1}(v_m; a_m, b_m)$ and $\epsilon_{B,m} = f_B^{-1}(b_m; a_m, s_m)$.

2. **Action:** Simulate an action by fixing the value of a data generating function. For example, the counterfactual question "what if the person m, aged a_m, had been 10 years older" involves fixing $f_A(\cdot)$ to $a_m + 10$. This intervention is denoted $do(a_m + 10)$.

3. **Prediction:** Generate a 3D surface mesh $x_{m,\text{cf}}$ as the answer to the counterfactual question. We input values calculated in the abduction and action steps into the data generating functions, e.g. **1)** $b_{m,\text{cf}} = f_B(\epsilon_{B,m}; a_m + 10, s_m)$; **2)** $v_{m,\text{cf}} = f_V(\epsilon_{V,m}; a_m + 10, b_{m,\text{cf}})$; **3)** $x_{m,\text{cf}} = f_X(u_m, z_m; v_{m,\text{cf}}, b_{m,\text{cf}})$, thereby sampling the counterfactual distribution $x_{m,\text{cf}} \sim p(x_{m,\text{cf}}|v_{m,\text{cf}}, b_{m,\text{cf}})$, and generating the counterfactual mesh.

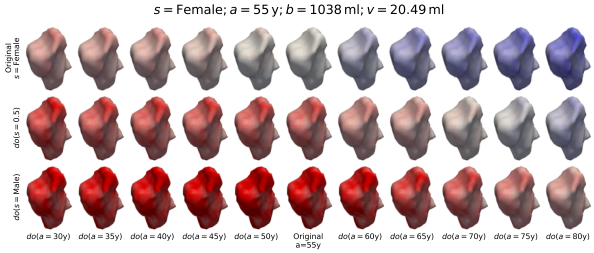


Figure 2: Counterfactual meshes for an individual under $do(a)$ and $do(s)$ – *"What would this person's brain stem look like if they were older/younger or male?"*. Colours show the vertex Euclidean distance between observed and counterfactual meshes – ■ = +5mm to ■ = -5mm.

3. Experiments & Results

In Fig. 2, subject-specific traits, encoded by ϵ_X, are successfully preserved in counterfactual meshes under the full range of interventions, whilst trends from the training data are also present. Namely, volume decreases as age increases, a male brain stem is larger than its female counterpart by the same scale factor for each age, and counterfactuals under $do(s = 0.5)$ are *half way* between the male and female meshes. Further, the CSM generates realistic meshes under the out-of-sample interventions $do(s = 0.5)$, $do(a > 70\text{y})$ and $do(a < 40\text{y})$.

In Fig. 3, subject-specific traits for both persons are preserved under the interventions also. Trends in the true distribution, seen in the background contours, are visible in the counterfactual trajectories with some subject-specific variations. For example, $do(s = 0)$ shifts the observed v and b values to the female region of the distribution for person A, resulting in a non-uniform shrinkage of the brain stem from a complex, non-linear transformation learned by $f_X(\cdot)$.

4. Conclusion & Discussion

We have successfully applied the DSCM framework to perform counterfactual inference of 3D meshes in a biomedical scenario. Our CSM generates novel counterfactual meshes under out-of-sample interventions and preserves subject-specific traits. The modularity of our approach enables the integration of SOTA 3D morphable models to improve our results. Alternative CSMs could also be built to model disease prognosis or treatment outcomes.

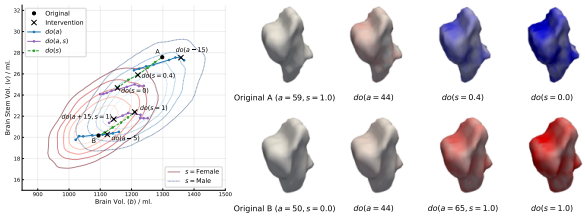


Figure 3: Counterfactual trajectories of b_{cf} and v_{cf} under interventions $do(a \pm T)$, $do(s = S)$, and $do(a \pm T, s = S')$, where $T \in \{5, 10, 15, 20\}$, $S \in \{0, 0.2, 0.4, 0.6, 1\}$ and S' is the opposite of the observed sex. Counterfactual meshes, corresponding to points marked ($\times$) on the trajectories, are visualised on the right. Contour plots depict the true density $p(v, b|s)$. Colours show the vertex Euclidean distance between the observed and counterfactual meshes – ■ = +5mm to ■ = -5mm.

Acknowledgments

This project has received funding from the European Research Council (ERC under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 757173, Project MIRA). The UK Biobank data is accessed under Application 12579. We thank Oliver Lewis for his support in the completion of this project. We also thank Daniel Clarke and William Turner for engaging discussions and suggesting improvements on the drafts of this paper.

References

Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Nick Pawlowski, Daniel C. Castro, and Ben Glocker. Deep structural causal models for tractable counterfactual inference. In *Advances in Neural Information Processing Systems*, volume 33, pages 857–869, 2020.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009. ISBN 9780521895606.

Anurag Ranjan, Timo Bolkart, Soubhik Sanyal, and Michael J. Black. Generating 3D faces using convolutional mesh autoencoders. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 704–720, 2018.

Rajat Rasal, Daniel C. Castro, Nick Pawlowski, and Ben Glocker. Deep structural causal shape models. In *ECCV 2022 Workshop on Causality in Vision*, 2022. URL https://arxiv.org/abs/2208.10950.

Axel Sauer and Andreas Geiger. Counterfactual generative networks. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=BXewfAYMmJw.

Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Athanasios Vlontzos, Daniel Rueckert, and Bernhard Kainz. A review of causality for learning algorithms in medical image analysis. *arXiv preprint arXiv:2206.05498*, 2022.

Appendix A. Proof of Objective

From the causal, graphical model in Fig. 1, we can write down the factorisation

$$p(x, v, b, a, s) = p(x|v, b) \cdot p(v|a, b) \cdot p(b|s, a) \cdot p(a) \cdot p(s) = p(x|v, b) \cdot p(v, b, a, s). \tag{2}$$

The joint including the independent exogenous variable z_X can be factorised as $p(x, z_X, v, b, a, s) = p(x|z_X, v, b) \cdot p(z_X) \cdot p(v, b, a, s)$ where z_X can be marginalised as

$$p(x, v, b, a, s) = \int p(x|z_X, v, b) \cdot p(z_X) \cdot p(v, b, a, s)\ dz_X \tag{3}$$

$$\iff \log p(x, v, b, a, s) = \log \int p(x|z_X, v, b) \cdot p(z_X) \cdot p(v, b, a, s)\ dz_X \tag{4}$$

$$= \alpha + \log \int p(x|z_X, v, b) \cdot p(z_X)\ dz_X, \tag{5}$$

where $\alpha = \log p(v, b, a, s)$. Since the marginalisation over z_X is intractable, we introduce the variational distribution $q(z_X|x, v, b) \approx p(z_X|x, v, b)$,

$$= \alpha + \log \int p(x|z_X, v, b) \cdot p(z_X) \cdot \frac{q(z_X|x, v, b)}{q(z_X|x, v, b)}\ dz_X, \tag{6}$$

then, by Jensen's inequality, arrive at a formulation for the evidence lower bound (ELBO),

$$\log p(x, v, b, a, s) \geq \alpha + \mathbb{E}_{q(z_X|x,v,b)} \left[\log \left(p(x|z_X, v, b) \cdot \frac{p(z_X)}{q(z_X|x, v, b)} \right) \right]. \tag{7}$$

This can be written using the Kullback–Leibler ($D_{KL}(\cdot)$) divergence,

$$\text{RHS Eq. (7)} = \alpha + \mathbb{E}_{q(z_X|x,v,b)}[\log p(x|z_X, v, b)] - \mathbb{E}_{q(z_X|x,v,b)} \left[\log \frac{q(z_X|x, v, b)}{p(z_X)} \right] \tag{8}$$

$$= \alpha + \underbrace{\mathbb{E}_{q(z_X|x,v,b)}[\log p(x|z_X, v, b)] - D_{KL}[q(z_X|x, v, b) \| p(z_X)]}_{\beta}, \tag{9}$$

which clearly demonstrates that β learns a mesh CVAE within the CSM structure.

Appendix B. Mesh CVAE Architecture

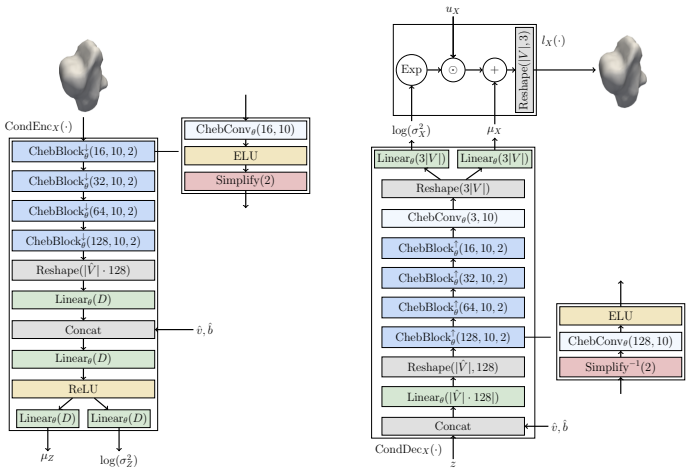


Figure 4: Network architectures for CondEnc$_X(\cdot)$ and CondDec$_X(\cdot)$ to implement $f_X(\cdot)$, utilising Chebyshev polynomial based spectral graph convolutions Defferrard et al. (2016). Reshape(t) reshapes the input to the shape given by tuple t. Linear$_\theta(M)$ is a fully connected layer with M output features. ReLU$(\cdot)$ refers to a rectified linear unit. ELU$(\cdot)$ refers to an exponential linear unit. $|\hat{V}|$ are the number of vertices output after ChebBlock$^{\downarrow}_{\theta}(128, 10, 2)$.

Appendix C. Preliminary Results on Hippocampus Meshes

We present preliminary results for a CSM of hippocampus meshes for the same set of individuals as the brain stem experiments. We assume the causal, graphical model in Fig. 1, where v is now the volume of the hippocampus and x is the hippocampus mesh. The architecture in Fig. 4 is reconfigured to account for the topological difference between hippocampus and brain stem meshes, whilst all other hyperparameters are kept the same. Results are seen in Fig. 5; subject-specific traits are preserved over the range of interventions, and the expected shape and volumetric trends are also present. Since hippocampus meshes are more topologically unsmooth than brain stem meshes, counterfactual meshes display surface deformities. This could be overcome by topology-specific hyperparamter tuning, utilising spatial graph convolutions or introducing mesh smoothing constraints to the objective.

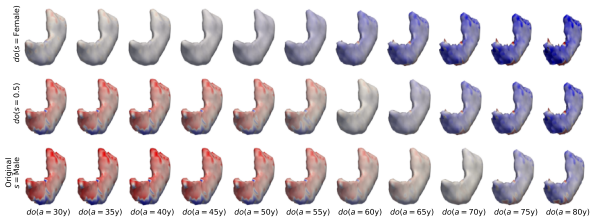


Figure 5: Counterfactual meshes for an individual under $do(a)$ and $do(s)$ – *"What would this person's hippocampus look like if they were older/younger or male?"*. Colours show the vertex Euclidean distance between observed and counterfactual meshes – ■ = +5mm to ■ = -5mm.