Generative Models of Part-Structured 3D Objects

Charlie Nash and Chris Williams, University of Edinburgh, United Kingdom
charlie.nash@ed.ac.uk, ckiw@inf.ed.ac.uk

EPSRC Engineering and Physical Sciences Research Council

Introduction

We introduce two generative models of part-segmented 3D objects:

- The shape variational auto-encoder (ShapeVAE)
- The shape factor analyzer (ShapeFA)

These models describe a distribution over the co-existence of object parts, as well as over the continuous variability of the object surface, leveraging the part structure of 3D objects in their architecture.

Object Representation

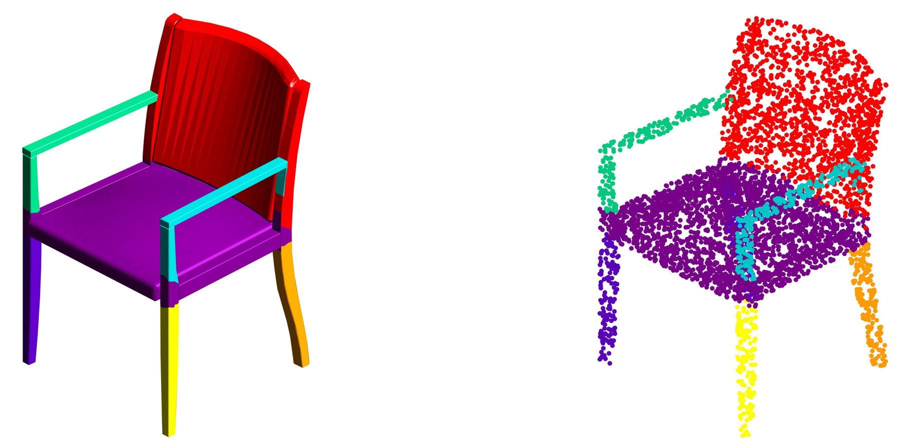


Figure 1: Part segmented 3D chair and surface point representation.

Dataset: We use a dataset from Huang *et al.*, [1] consisting of:

- Aligned 3D objects from the same object class
- 3D point cloud representation with correspondences
- Points labelled with part ID

Notation: We represent objects with D keypoints and K parts as:

- Surface points $\mathbf{x} \in (\mathbb{R} \cup m)^{3D}$
- Symbol m indicates that a variable is missing
- Part existence vector $\mathbf{e} \in \{0, 1\}^K$

Model Structure

Categorical part existence distribution:

$$p(\mathbf{e}|\boldsymbol{\phi}) = \mathrm{Cat}(\mathbf{e}|\boldsymbol{\phi})$$

Conditional surface point distribution with latent variables:

$$p(\mathbf{x}_m|\mathbf{e}, \boldsymbol{\theta}) = \mathbb{I}[\mathbf{x}_m = [m, \dots, m]]$$

$$p(\mathbf{x}_v|\mathbf{e}, \boldsymbol{\theta}) = \int_{\mathbf{z}} \mathcal{N}(\mathbf{x}_v|\boldsymbol{\mu}(\mathbf{z}, \mathbf{e}, \boldsymbol{\theta}), \boldsymbol{\Sigma}(\mathbf{z}, \mathbf{e}, \boldsymbol{\theta}))\mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})d\mathbf{z}$$

Shape samples

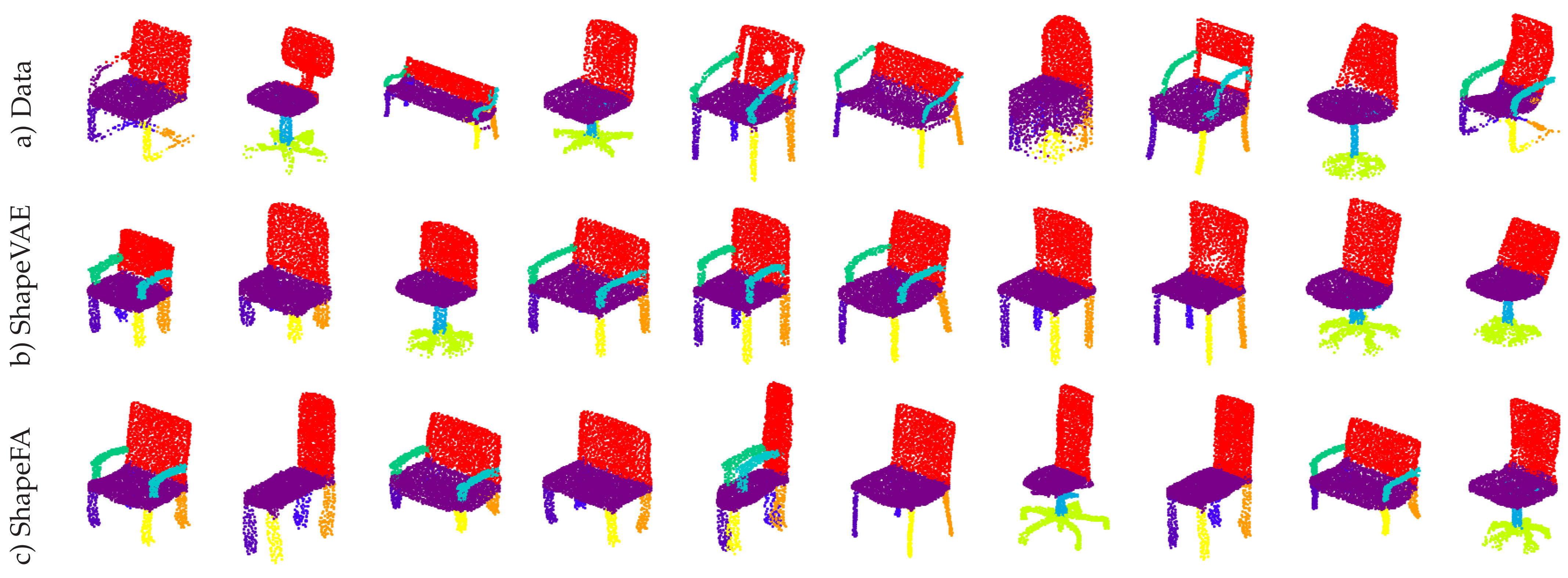
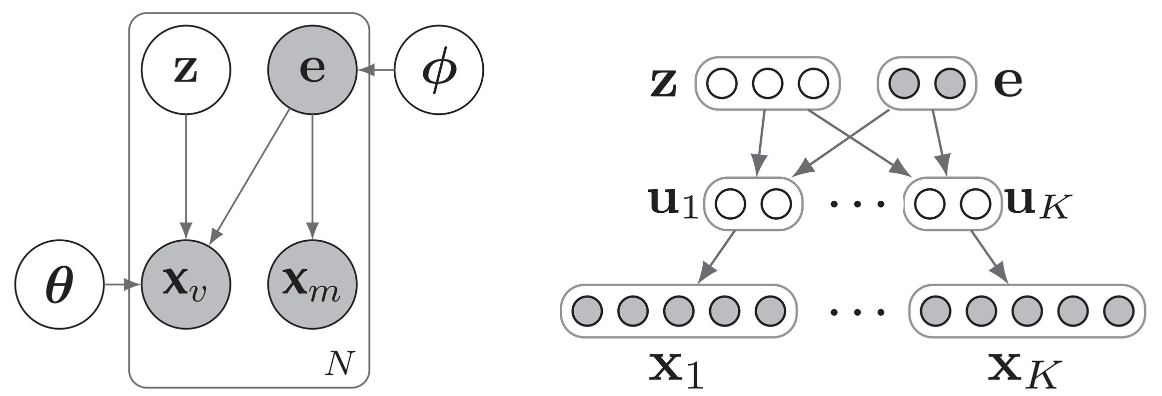


Figure 2: **Shape samples**

- The ShapeFA produces samples that are mainly plausible, however there are examples of unusually stretched parts
- The samples produces by the ShapeVAE are realistic and do not suffer from the same stretching issues as the ShapeFA samples

Shape Factor Analysis



a) Graphical model b) ShapeFA

Figure 3: **Model structure**

In this variant of the keypoint distribution we use a hierarchical factor analysis model:

- Factor analysis for each part: $p(\mathbf{x}_k|\mathbf{u}_k, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}_k|\mathbf{W}_k^{(1)}\mathbf{u}_k + \boldsymbol{\mu}_k^{(1)}, \boldsymbol{\Psi}_k^{(1)})$
- Structural factor analysis for part latent variables: $p(\mathbf{u}_v|\mathbf{z}, \mathbf{e}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{u}_v|\mathbf{W}_\mathbf{e}^{(2)}\mathbf{z} + \boldsymbol{\mu}_\mathbf{e}^{(2)}, \boldsymbol{\Psi}_\mathbf{e}^{(2)})$

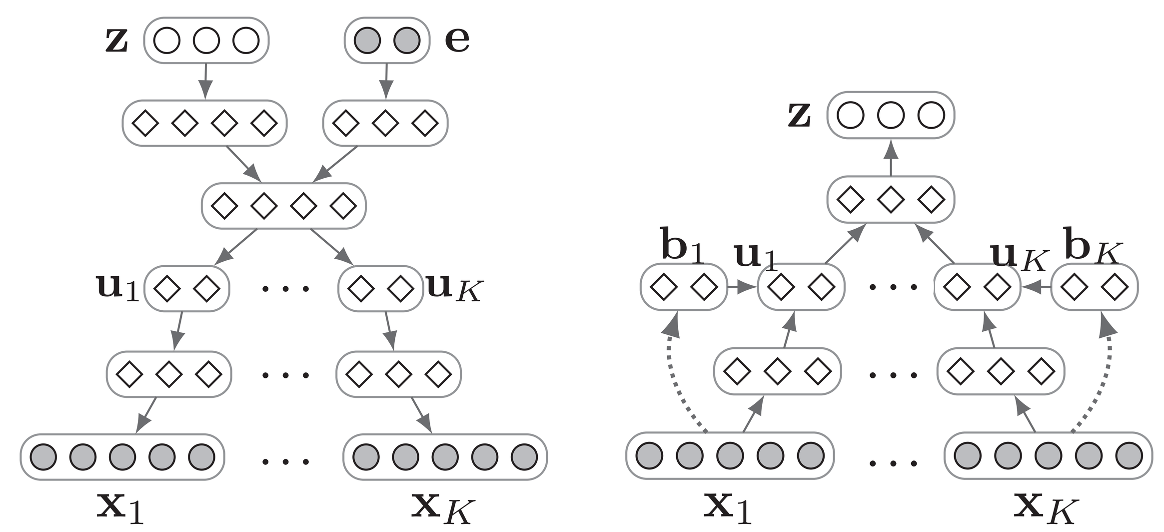
By integrating out the part latent variables we obtain a shallow factor analysis model:

$$\boldsymbol{\mu}(\mathbf{z}, \mathbf{e}, \boldsymbol{\theta}) = \mathbf{W}_v^{(1)}(\mathbf{W}_\mathbf{e}^{(2)}\mathbf{z} + \boldsymbol{\mu}_\mathbf{e}^{(2)}) + \boldsymbol{\mu}_v^{(1)}$$

$$\boldsymbol{\Sigma}(\mathbf{z}, \mathbf{e}, \boldsymbol{\theta}) = \boldsymbol{\Psi}_v^{(1)} + \mathbf{W}_v^{(1)}\boldsymbol{\Psi}_\mathbf{e}^{(2)}\mathbf{W}_v^{(1)\intercal}$$

We train the model using the greedy layer-wise procedure described by Tang *et al.* [2].

Shape Variational Auto-Encoder



a) ShapeVAE Decoder b) ShapeVAE Encoder

Figure 6: **ShapeVAE**

Decoder: Map from from latent variables and existences to part representation $\mathbf{u}(\mathbf{z}, \mathbf{e}) = \mathtt{MLP}(\mathbf{z}, \mathbf{e})$ and then split into parts $\mathbf{u} = [\mathbf{u}_1, \dots, \mathbf{u}_K]$.

Output parameters are obtained using $\boldsymbol{\mu}_k(\mathbf{z}, \mathbf{e}) = \mathtt{Linear}(\mathbf{h}_k(\mathbf{z}, \mathbf{e}))$, and $\boldsymbol{\sigma}_k^2(\mathbf{z}, \mathbf{e}) = \exp(\mathtt{Linear}(\mathbf{h}_k(\mathbf{z}, \mathbf{e})))$.

Encoder: Map from keypoints $\mathbf{x}$ to a part representation $\mathbf{u} = [\mathbf{u}_1, \dots, \mathbf{u}_K]$. Parts that are missing simply generate a bias which is added in the appropriate position. The part representation is then passed through fully-connected layers to obtain the output parameters.

Results

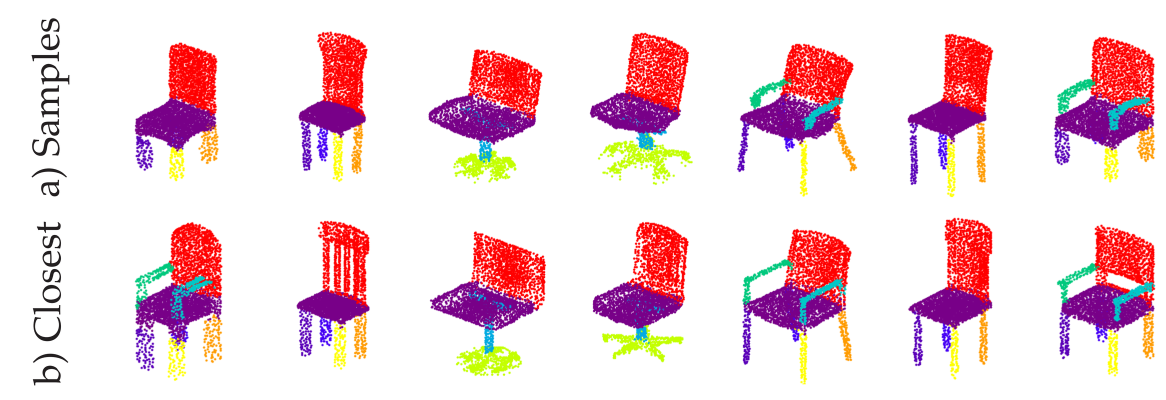
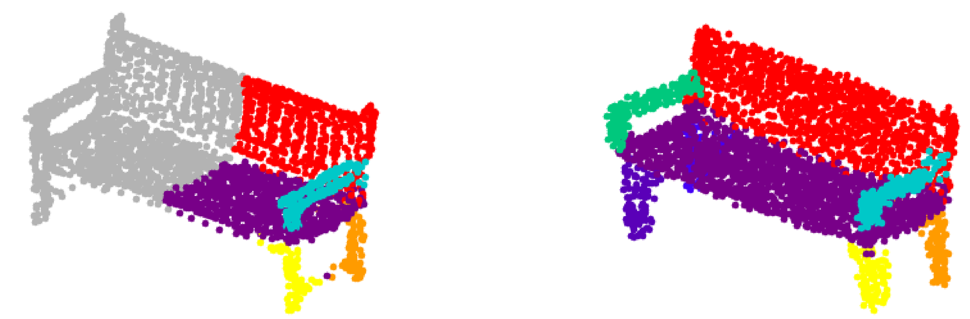


Figure 4: **Generalisation**

Generalisation: To demonstrate that the model is capable of producing samples that are distinct from the training set we plot a set of samples along with their closest data examples.



a) Partially obscured b) ShapeVAE completion

Figure 5: **Shape completion**

Shape completion: For the ShapeVAE the completion is obtained by optimizing the latent variables so as to match the observations.

Model	log-lik
Diag-Gaussian	0.37 ± 0.32
ShapeFA-8	$\mathbf{1.40 \pm 1.00}$
ShapeFA-32	1.35 ± 1.35
ShapeVAE-8	1.35 ± 0.51
ShapeVAE-32	1.35 ± 0.35

Table 1: **Test log-likelihood.**

Density estimation: The ShapeFA with 8 latent dimensions achieves the best performance.

References

[1] H. Huang, E. Kalogerakis, and B. M. Marlin. Analysis and synthesis of 3D shape families via deep-learned generative models of surfaces. *Computer Graphics Forum*, 34(5):25–38, 2015.

[2] Y. Tang, R. Salakhutdinov, and G. E. Hinton. Deep mixtures of factor analysers. In *ICML*. icml.cc / Omnipress, 2012.

Funding

The work of CW is supported in part by EPSRC grant EP/N510129/1. CN is supported by the Centre for Doctoral Training in Data Science, funded by EPSRC grant EP/L016427/1.