Generalized Autoencoder for Volumetric Shape Generation

Yanran Guan Tansin Jahan Oliver van Kaick
{yanran.guan, tansin.jahan, oliver.vankaick}@carleton.ca

INTRODUCTION

What constitutes a good 3D generative model? A good generative model should not only be able to reconstruct well the training shapes, but should also be able to generalize, which implies that the model should at least enable a meaningful interpolation between shapes.

We introduce a 3D generative model based on the generalized autoencoder (GAE) [1], which allows one to control the latent manifold learned by the model. We guide the construction of a latent manifold of 3D shapes with data similarities computed via the Chamfer distance [2], and train the model with a loss that is the combination of the traditional autoencoder (AE) and GAE losses. We show that this model leads to more meaningful manifold structures and better interpolations between shapes when compared to previous approaches.

METHOD

Our 3D generalized autoencoder (3D-GAE) uses the GAE loss [1] that takes each shape $\mathbf{x}_i$ to reconstruct a set of shapes $\Omega_{\mathbf{x}_i}$, which consists of the k-nearest neighbors of $\mathbf{x}_i$ given by the Chamfer distance. We combine the GAE loss with the reconstruction error of each batch B, so that our model can better converge to a global optimum:

$$L(B, B') = \|B - B'\|^2 + \sum_{\mathbf{x}_i \in B} \sum_{\mathbf{x}_j \in \Omega_{\mathbf{x}_i}} s_{i,j} \left\| \mathbf{x}_j - \mathbf{x}_i{}' \right\|^2,$$

where $s_{i,j}$ is the weight given by:

$$s_{i,j} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}\right),$$

where we set $t = 200$ in our implementation.

The network of our 3D-GAE follows a symmetric architecture (Figure 1) that reconstructs input shapes represented by $32 \times 32 \times 32$ voxels.

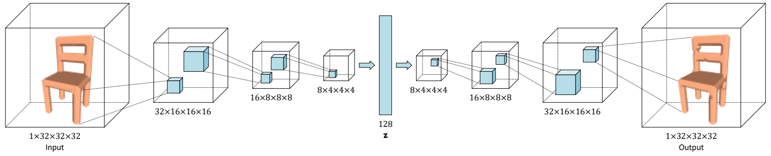

Figure 1. Network architecture.

SHAPE SYNTHESIS

We linearly interpolate reference shapes to synthesize new shapes. A collection of shapes interpolated by the 3D-GAE is shown in Figure 2. The 3D-GAE generates more meaningful interpolations than the volumetric autoencoder (3D-AE) and the volumetric variational autoencoder (3D-VAE) [3], with less spurious parts, see Figure 3. Moreover, we apply arithmetic operations to the latent vectors learned by the 3D-GAE to enable shape extrapolation, see Figure 4.

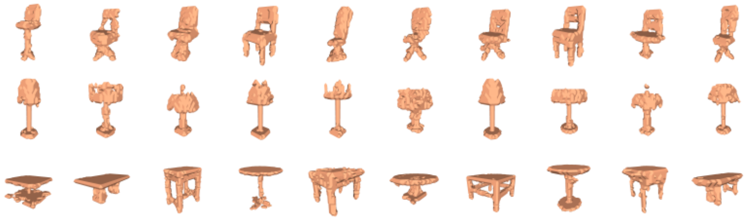

Figure 2. A collection of shapes synthesized via linear interpolation.

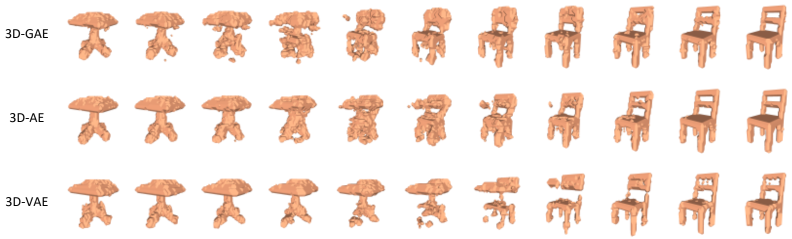

Figure 3. Linear interpolations obtained with different models.

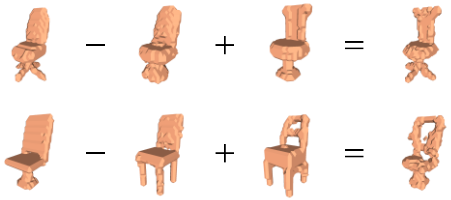

Figure 4. Examples of extrapolation via shape arithmetic.

EVALUATION

For a quantitative evaluation of the quality of the shapes from different categories interpolated with the three methods, we compute their inception scores [4], see Table 1. Also, we can see that the 3D-GAE learns a meaningful latent space from the multi-dimensional scaling (MDS) diagram of the dissimilarity among the latent vectors, as shown in Figure 5. We can observe how shapes with similar structure are grouped closely together in the same region of the diagram.

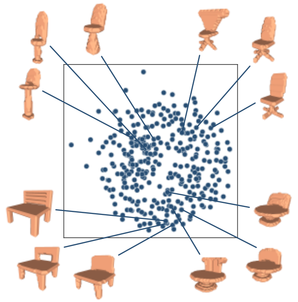

Figure 5. MDS diagram.

Dataset	3D-GAE	3D-AE	3D-VAE
Chair	3.09 ± 0.36	2.98 ± 0.41	2.92 ± 0.38
Lamp	1.57 ± 0.24	1.33 ± 0.31	1.19 ± 0.25
Table	2.45 ± 0.35	2.19 ± 0.28	1.78 ± 0.32

Table 1. Inception scores of interpolated shapes.

REFERENCES

[1] W. Wang, Y. Huang, Y. Wang, and L. Wang. Generalized autoencoder: A neural network framework for dimensionality reduction. In *CVPRW*, 2014.

[2] H. Fan, H. Su, and L. Guibas. A point set generation network for 3D object reconstruction from a single image. In *CVPR*, 2017.

[3] A. Brock, T. Lim, J. M. Ritchie, and N. Weston. Generative and discriminative voxel modeling with convolutional neural networks. *CoRR*, abs/1608.04236, 2016.

[4] J. Xie, Z. Zheng, R. Gao, W. Wang, S.-C. Zhu, and Y. N. Wu. Learning descriptor networks for 3D shape synthesis and analysis. In *CVPR*, 2018.