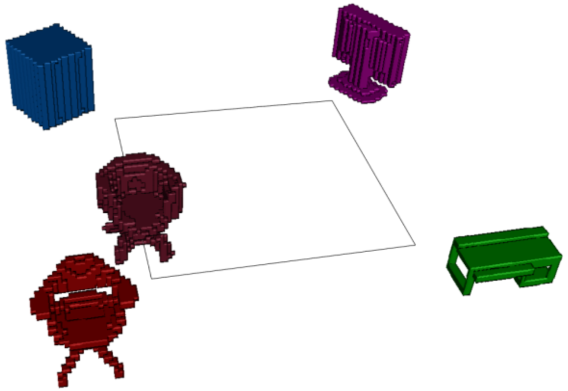
Semantics-Guided Latent Space Exploration for Shape Generation

Tansin Jahan   Yanran Guan   Oliver van Kaick

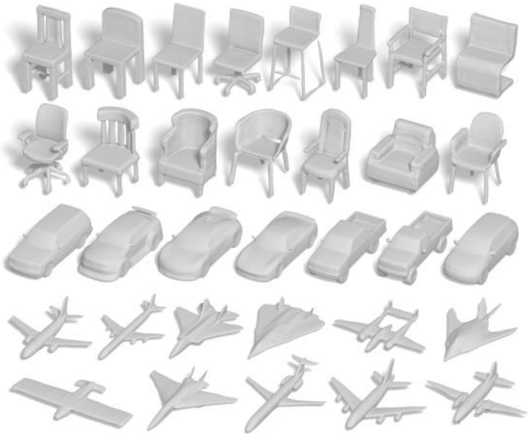
School of Computer Science, Carleton University, Canada

Deep Generative Models for 3D Shapes

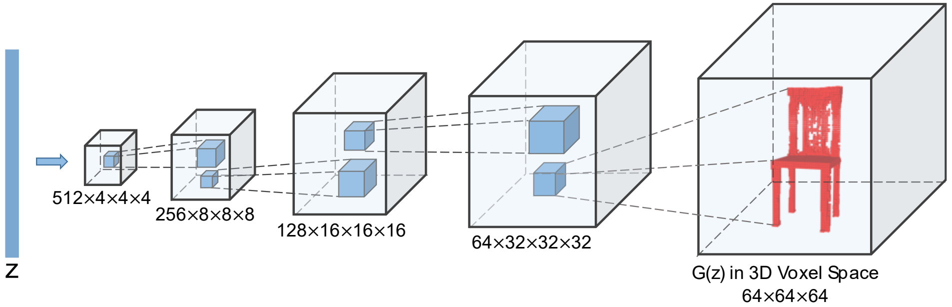


Volumetric VAE [BLRW16]

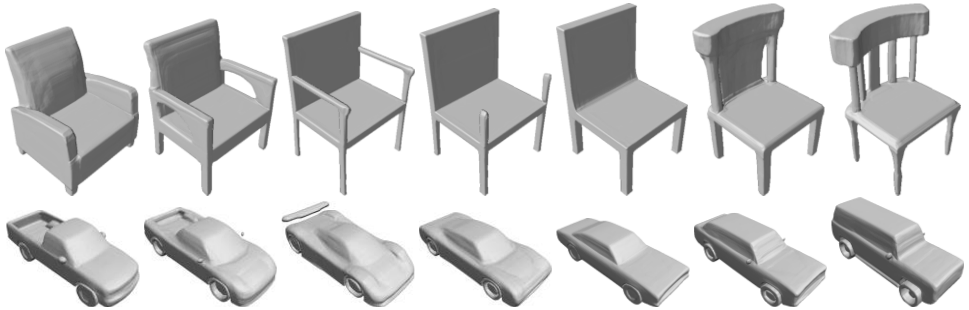
Point cloud AE [ADMG18]



Implicit decoder [CZ19]

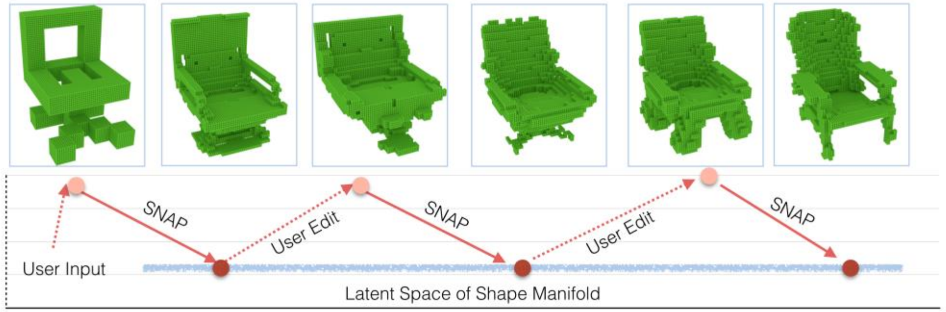


Volumetric GAN [WZX*16]

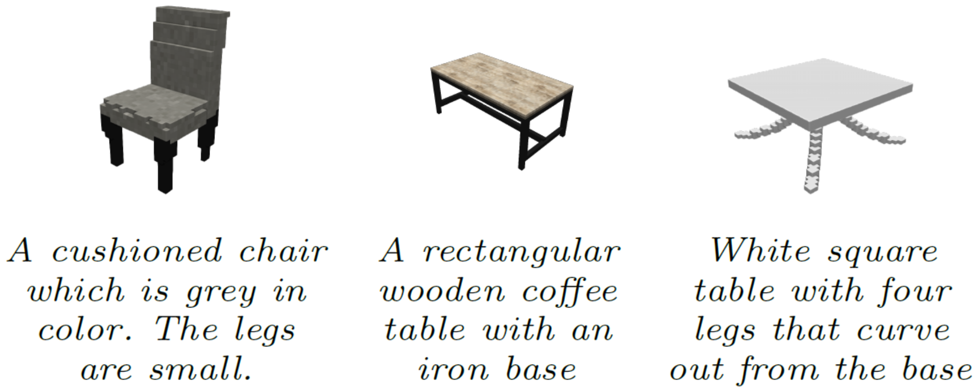


DeepSDF [PFS*19]

Controlling Shape Generation

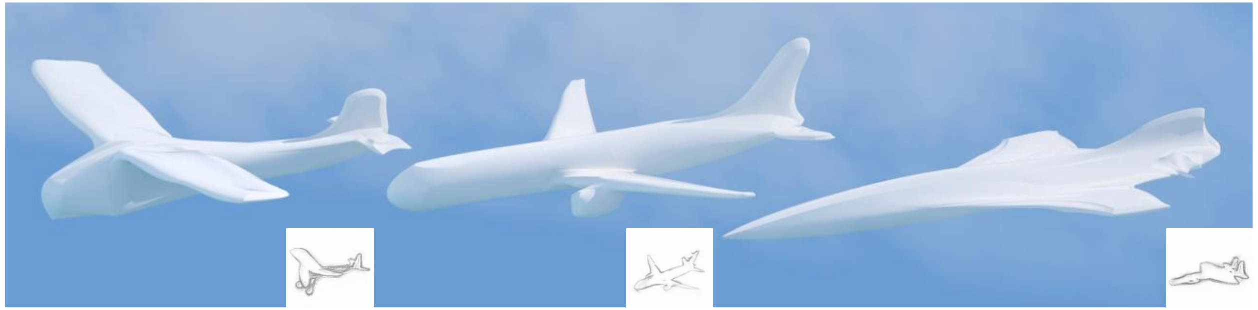


“Snapping” mechanism [LYF17]



Text2Shape [CCS*18]

ShapeGlot [AFH*19]



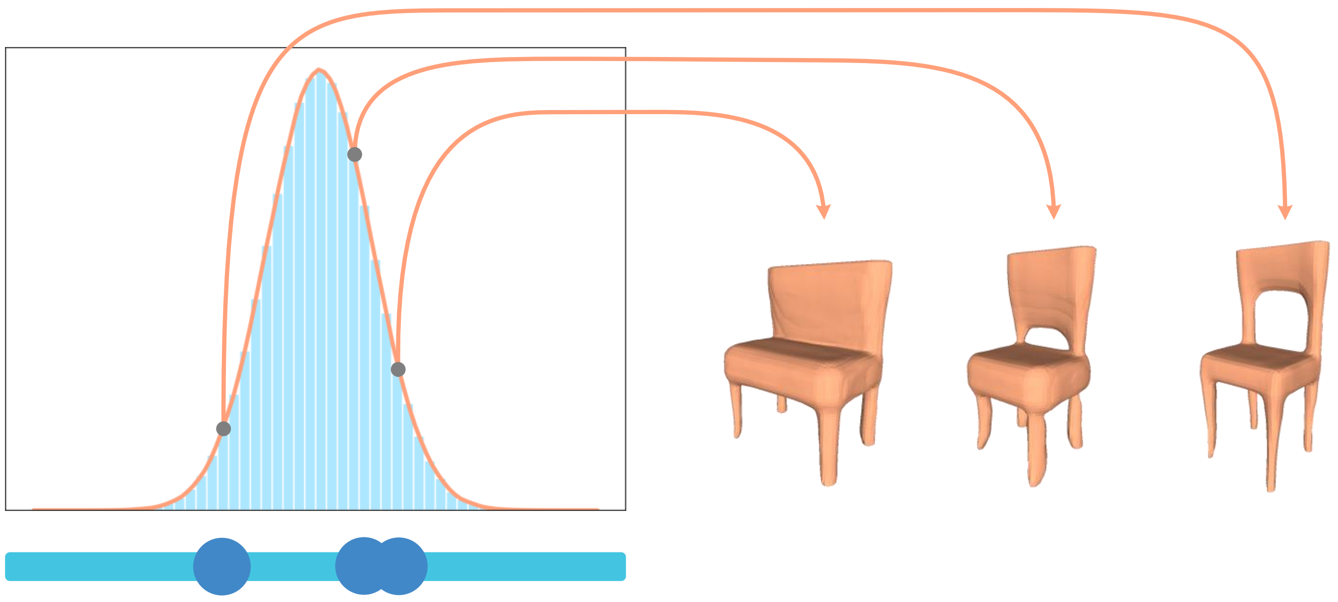
Sketch-based modeling [SBS19]

Our Method

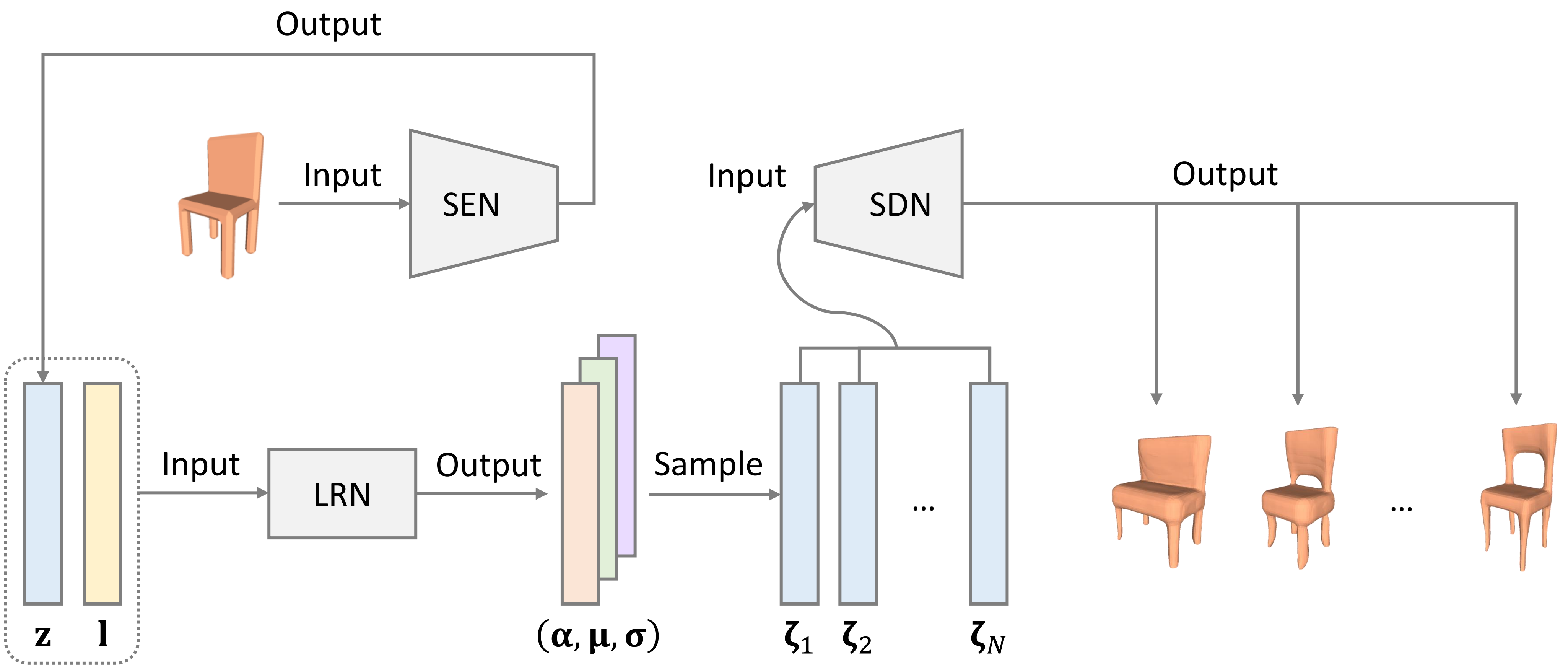
- Facilitates the exploration of the learned shape spaces through keywords
- Maps the keywords to distributions of the latent dimensions

“straight square back”

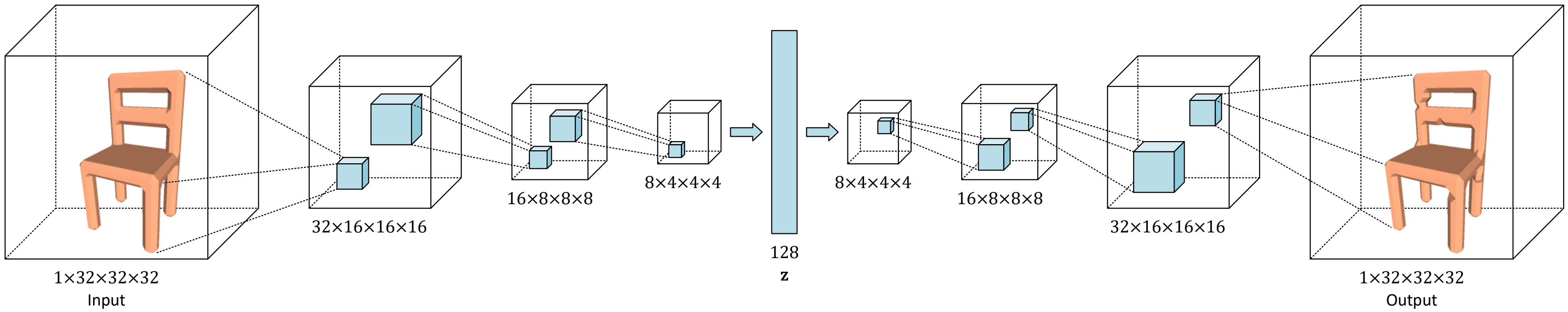
“four straight short legs”



Method Overview

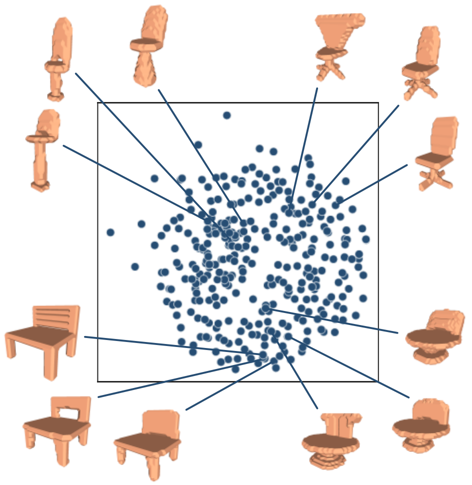


Shape Encoder Network (SEN)



Architecture of 3D-GAE [GJvK20]

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



Shape Decoder Network (SDN)

Decodes latent vectors into shapes

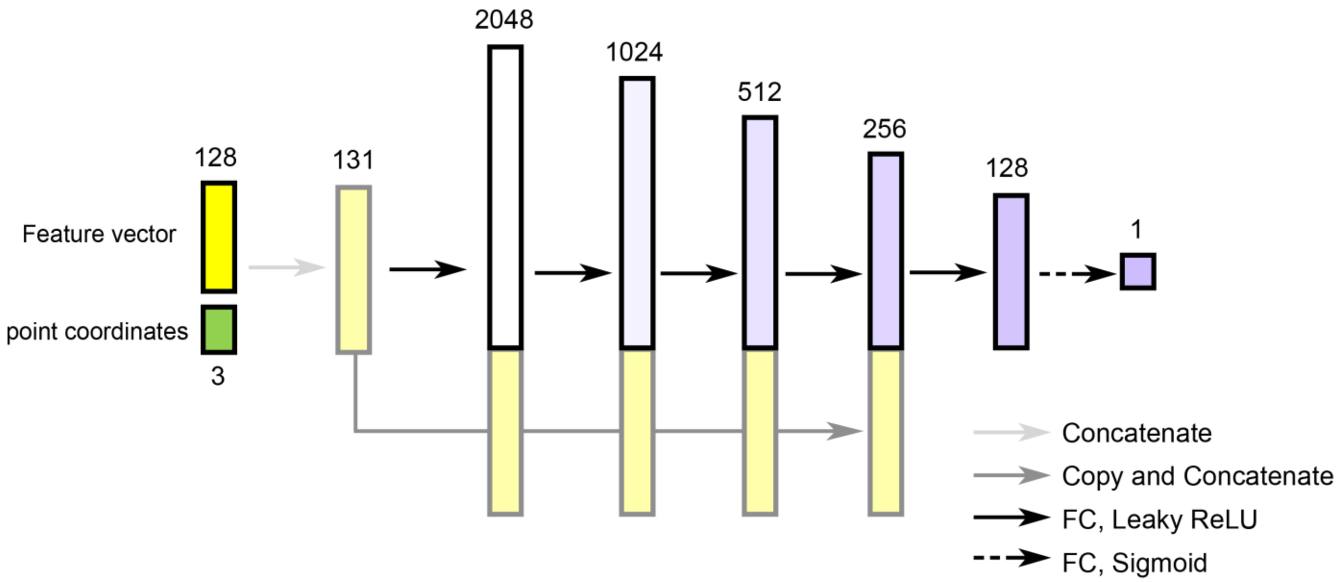
Outside the shape:

$$F(\mathbf{p}) = 0.$$

Inside the shape:

$$F(\mathbf{p}) = 1.$$

$$\mathcal{L}_{\mathrm{SDN}} = \frac{1}{|\mathcal{P}|} \sum_{\mathbf{p} \in \mathcal{P}} \| F(\mathbf{p}) - f(\mathbf{z}, \mathbf{p}) \|^2$$



Architecture of the implicit decoder [CZ19]

Label Regression Network (LRN)

Maps keywords to the regions of the latent space

The combination of probability density is

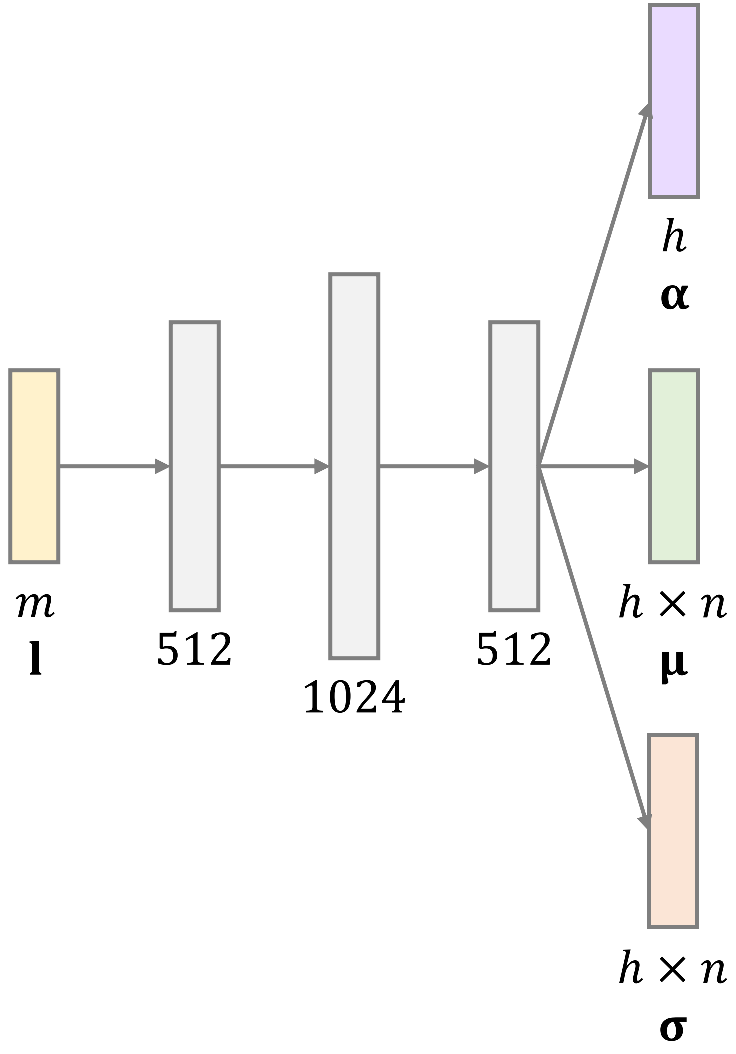
$$P(z_i|\mathbf{l}) = \sum_{w=1}^{h} \alpha_w \phi_{w_i}(z_i|\mathbf{l}),$$

with the probability density of each being

$$\phi_{w_i}(z_i|\mathbf{l}) = \frac{1}{\sigma_{w_i}\sqrt{2\pi}} \exp\left(-\frac{\left(z_i - \mu_{w_i}\right)^2}{2\sigma_{w_i}}\right).$$

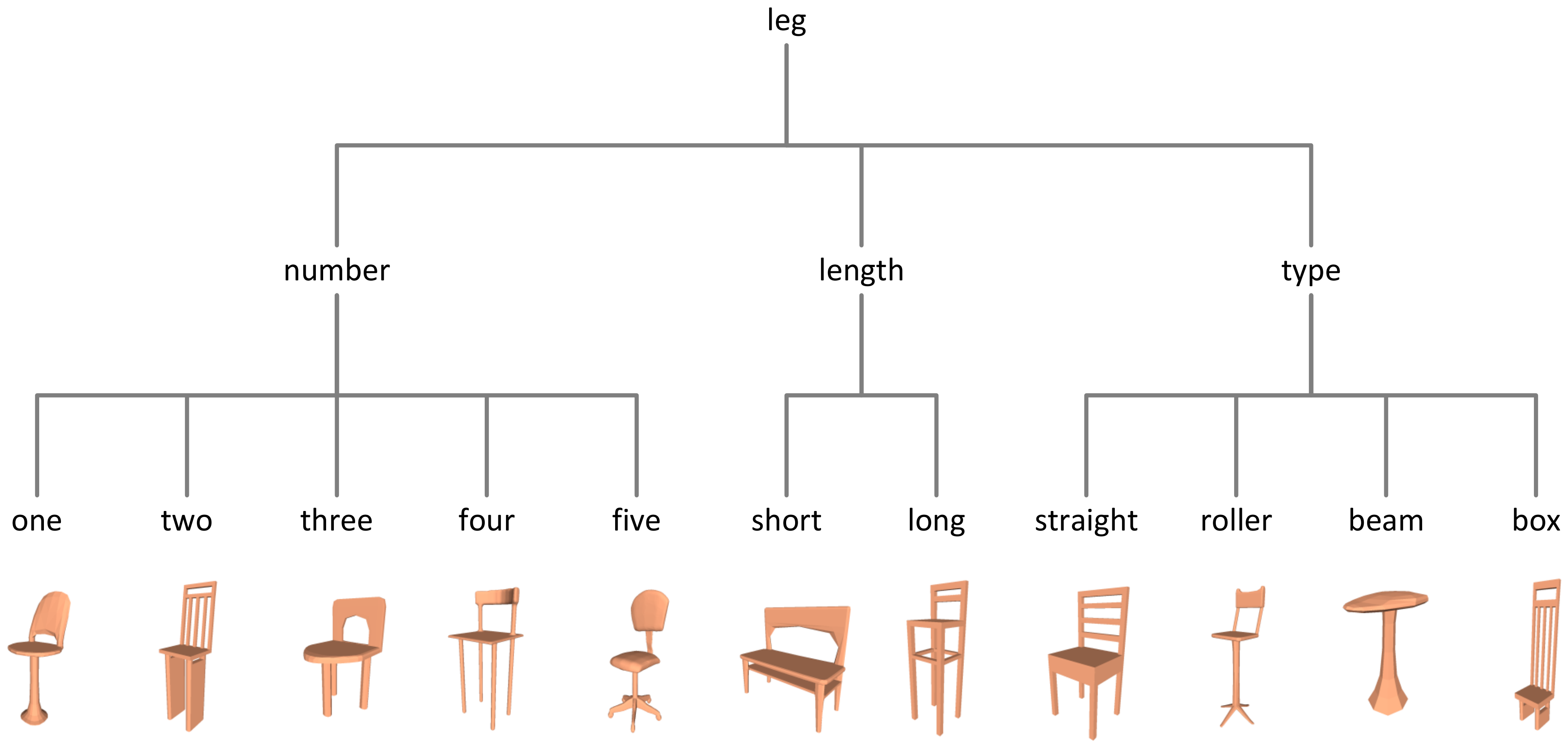
So, the final loss of LRN is

$$\mathcal{L}_{\mathrm{LRN}} = -\log\big(P(z_i|\mathbf{l})\big).$$



Architecture of LRN

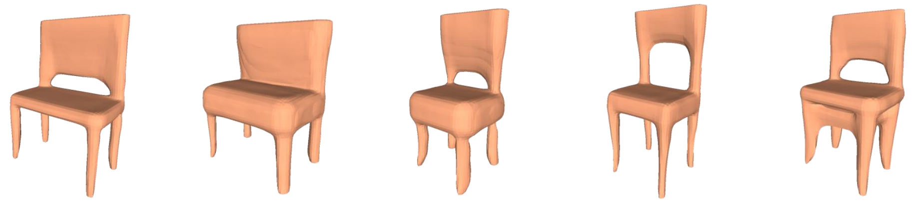
Datasets: Shapes and Labels



Results

Back:	size --- full
	side view --- straight
Seat:	shape --- square
Leg:	number --- four
	length --- short
	type --- straight

Gaussian 1
$\alpha_1 = 0.52$



Gaussian 2
$\alpha_2 = 0.22$

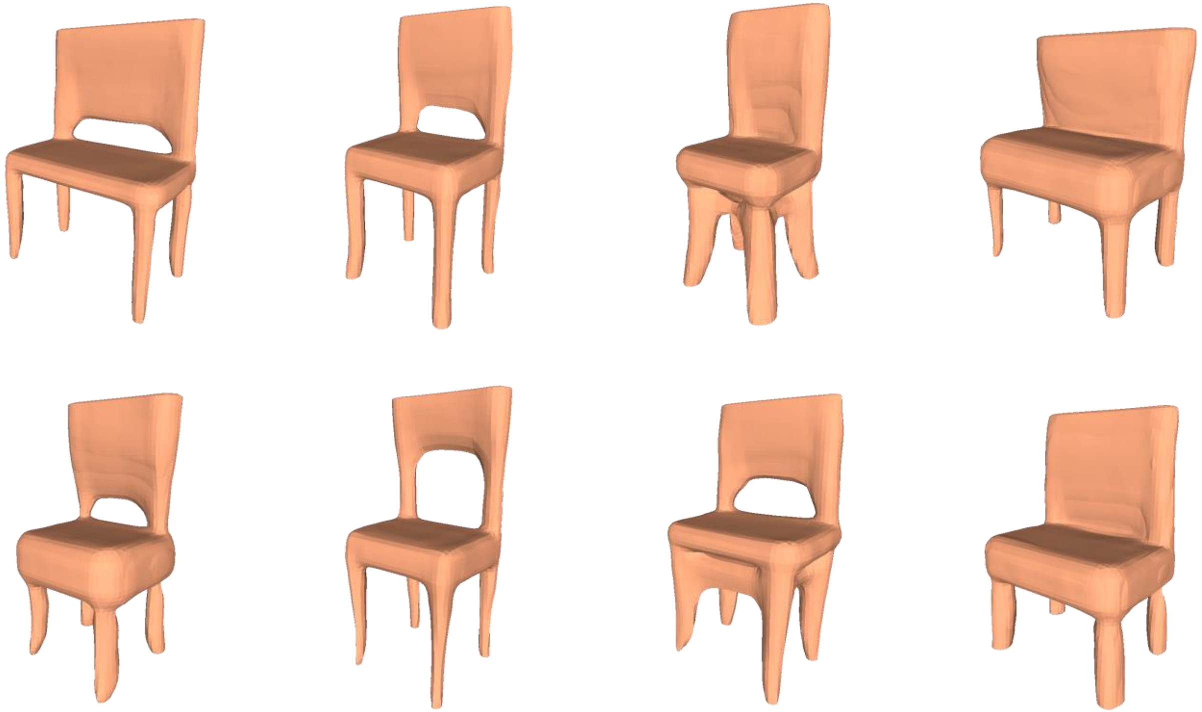
Gaussian 3
$\alpha_3 = 0.18$

Gaussian 4
$\alpha_4 = 0.04$

Gaussian 5
$\alpha_5 = 0.04$

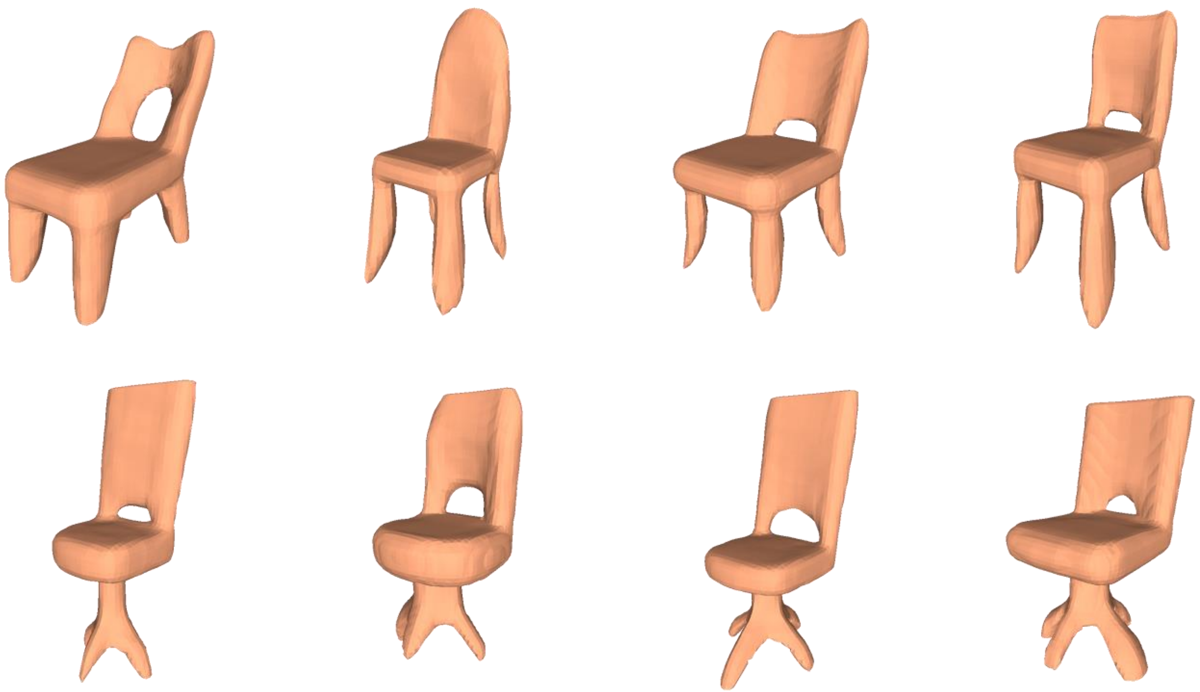
Results

Back:	size --- full
	side view --- straight
Seat:	shape --- square
Leg:	number --- four
	length --- short
	type --- straight



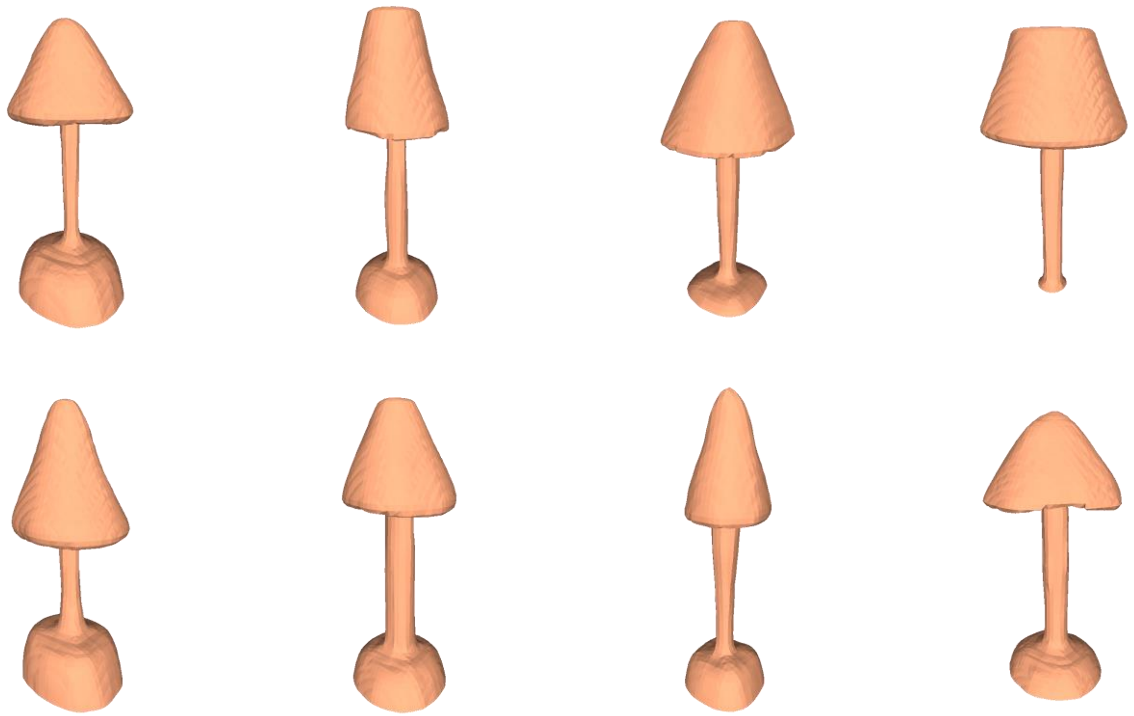
Results

Back:	size --- full
	fill --- vertical ladder
	fill --- hole(s)
	side view --- bent
	front view --- square
Seat:	shape --- circular
Leg:	length --- short



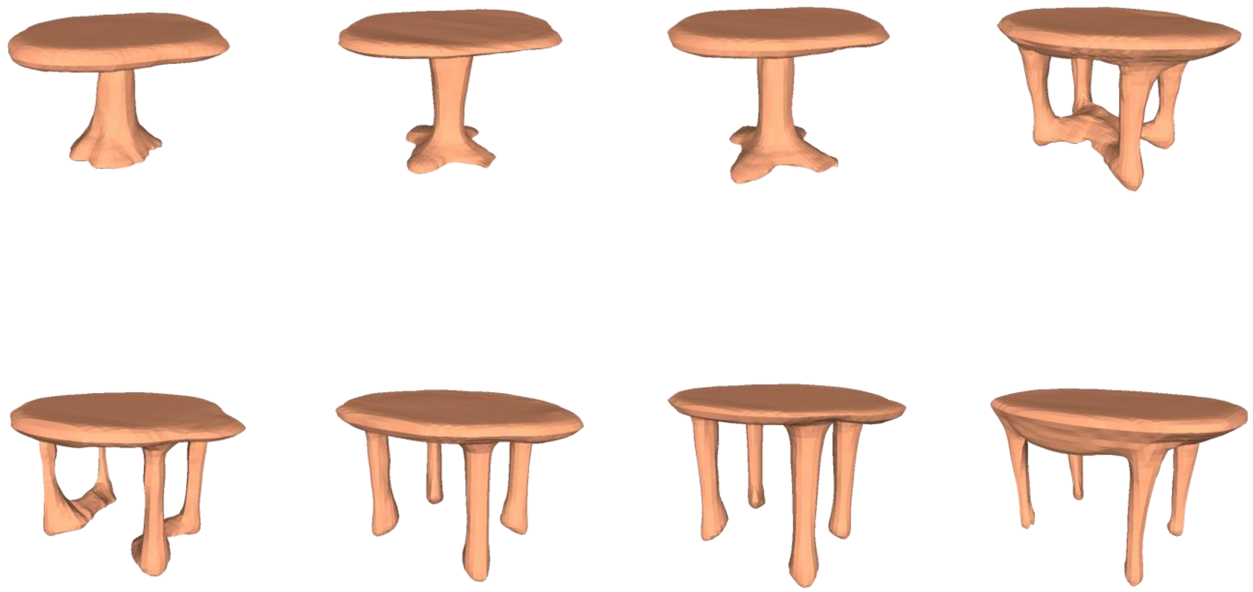
Results

Shade: front view --- bell
fitting --- empty
Body: type --- pipe
Base: connection --- untangled

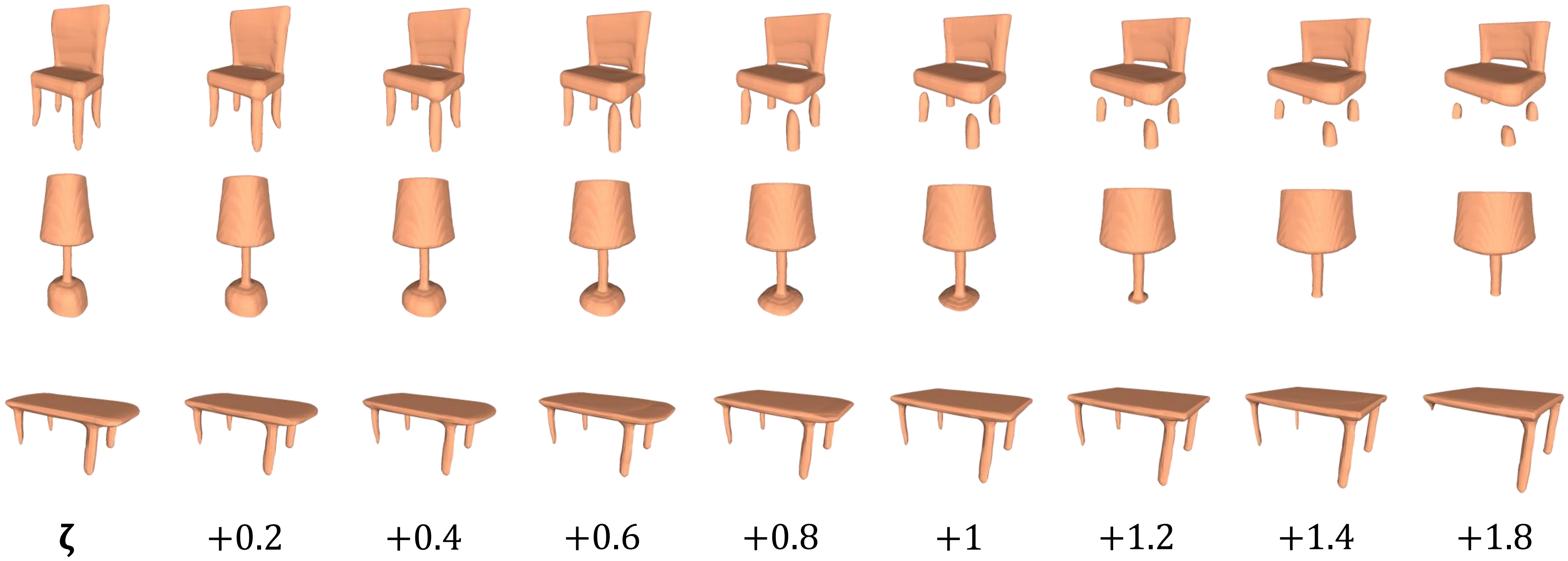


Results

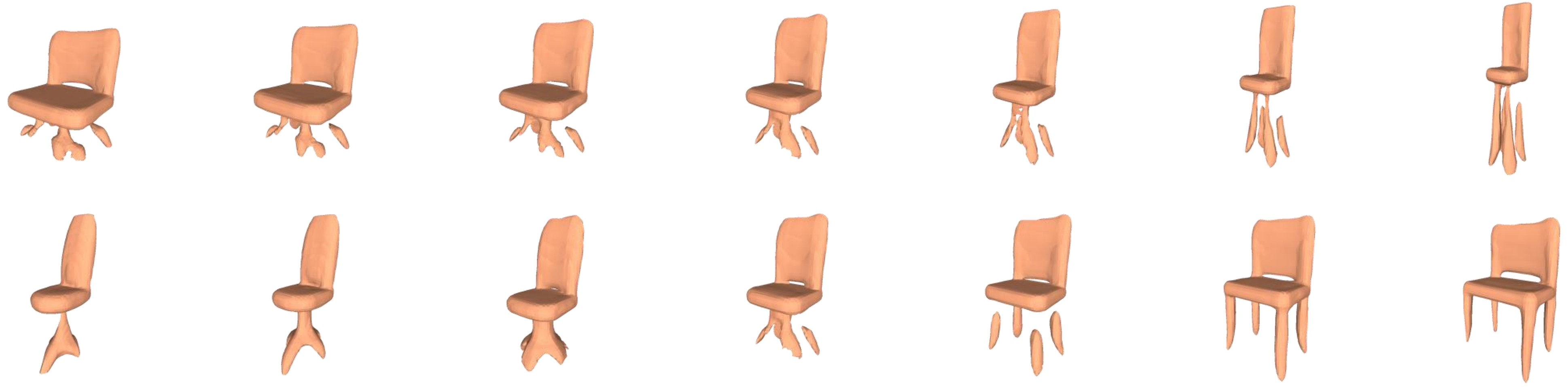
Top: type --- single
shape --- round
Leg: length --- medium



Perturbation of Latent Vectors



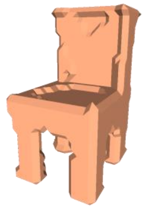
Comparison to the PCA-Based Exploration



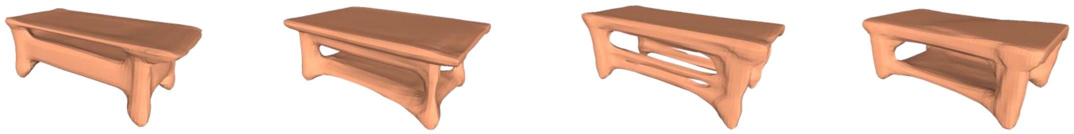
$\mu - 2\sigma$ $\mu - 1.33\sigma$ $\mu - 0.67\sigma$ μ $\mu + 0.67\sigma$ $\mu + 1.33\sigma$ $\mu + 2\sigma$

Comparison to Text2Shape

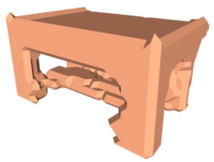
Back: size --- full
front view --- square
Seat: shape --- square



A chair with a high back and a square shape



Top: shape --- rectangular
Leg: number --- four
Side: connection --- closed



A rectangular table with four linked legs

Conclusion

- Mapping of semantic labels to distributions in the latent space enables users to **explore subspaces** of shapes constrained by the labels
- The user **can generate a variety of shapes** with the specified attributes
- Solution combines:
  - **Label regression network** that learns distributions of latent vectors conditioned on the labels
  - **Generative network** which translates sampled latent vectors into 3D shapes

Limitations

- Our method requires **labeled datasets**
- Thus, our current datasets are **small**
- Exploration requires an **experimental threshold** to select relevant dimensions to inspect

Future Work

- **Crowdsourcing** could provide labeling for more and larger datasets
- Could explore the mapping of labels to **other types of distributions** in the latent space
- Could **combine keyword guidance** with **PCA-based exploration** and **“snapping”** latent vectors to the manifold

References

[ADMG18] ACHLIOPTAS P., DIAMANTI O., MITLIAGKAS I., GUIBAS L.: Learning representations and generative models for 3D point clouds. In *ICML* (2018), pp. 40–49.

[AFH*19] ACHLIOPTAS P., FAN J., HAWKINS R., GOODMAN N., GUIBAS L.: ShapeGlot: Learning language for shape differentiation. In *ICCV* (2019), pp. 8937–8946.

[BLRW16] BROCK A., LIM T., RITCHIE J. M., WESTON N.: Generative and discriminative voxel modeling with convolutional neural networks. *CoRR abs/1608.04236* (2016).

[CCS*18] CHEN K., CHOY C. B., SAVVA M., CHANG A. X., FUNKHOUSER T., SAVARESE S.: Text2Shape: Generating shapes from natural language by learning joint embeddings. In *ACCV* (2018), pp. 100–116.

[CZ19] CHEN Z., ZHANG H.: Learning implicit fields for generative shape modeling. In *CVPR* (2019), pp. 5932–5941.

[GJvK20] GUAN Y., JAHAN T., VAN KAICK O.: Generalized autoencoder for volumetric shape generation. In *CVPRW* (2020), pp. 1082–1088.

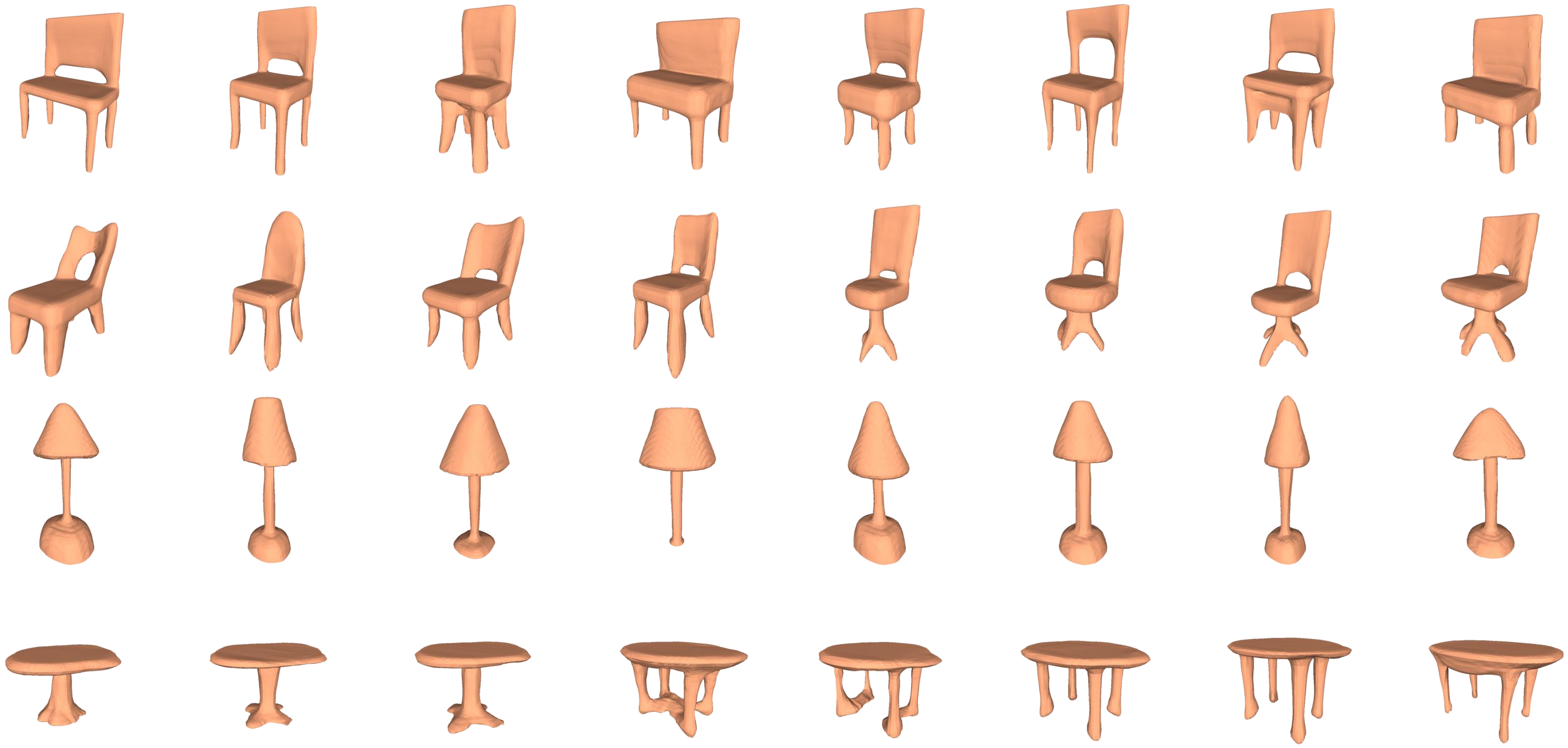
[LYF17] LIU J., YU F., FUNKHOUSER T.: Interactive 3D modeling with a generative adversarial network. In *3DV* (2017), pp. 126–134.

[PFS*19] PARK J. J., FLORENCE P., STRAUB J., NEWCOMBE R., LOVEGROVE S.: DeepSDF: Learning continuous signed distance functions for shape representation. In *CVPR* (2019), pp. 165–174.

[SBS19] SMIRNOV D., BESSMELTSEV M., SOLOMON J.: Learning manifold patch-based representations of man-made shapes. *CoRR abs/1906.12337* (2019).

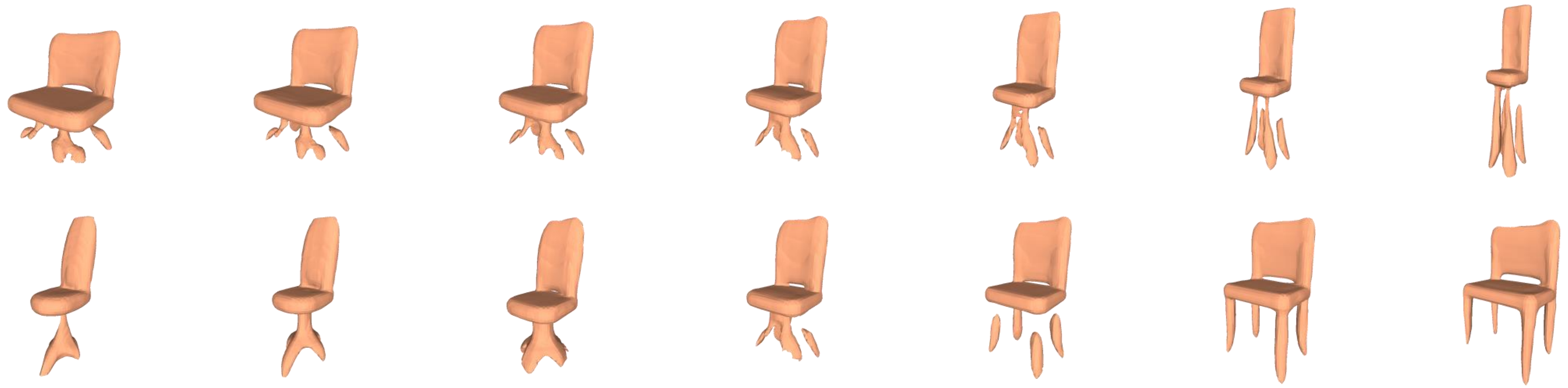
[WZX*16] WU J., ZHANG C., XUE T., FREEMAN W. T., TENENBAUM J. B.: Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In *NIPS* (2016), pp. 82–90.

Thank you!



Appendix A: Comparison to the PCA-Based Exploration

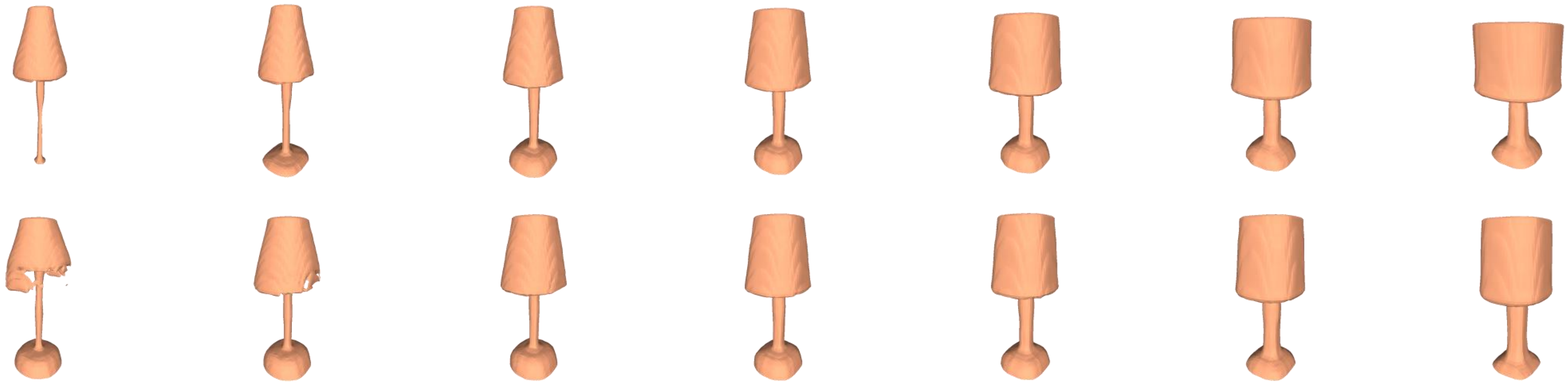
Chairs:



$\mu - 2\sigma$ | $\mu - 1.33\sigma$ | $\mu - 0.67\sigma$ | μ | $\mu + 0.67\sigma$ | $\mu + 1.33\sigma$ | $\mu + 2\sigma$

Appendix A: Comparison to the PCA-Based Exploration

Lamps:



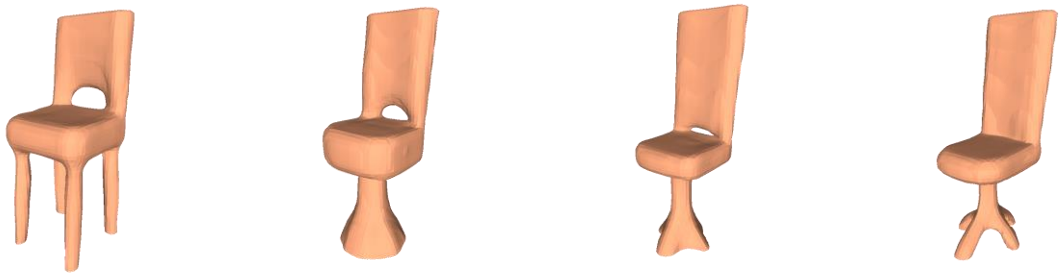
$\mu - 2\sigma$ | $\mu - 1.33\sigma$ | $\mu - 0.67\sigma$ | μ | $\mu + 0.67\sigma$ | $\mu + 1.33\sigma$ | $\mu + 2\sigma$

Appendix A: Comparison to the PCA-Based Exploration

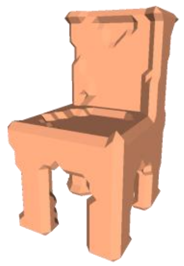
Tables:

$\mu - 2\sigma$ | $\mu - 1.33\sigma$ | $\mu - 0.67\sigma$ | μ | $\mu + 0.67\sigma$ | $\mu + 1.33\sigma$ | $\mu + 2\sigma$

Appendix B: Comparison to Text2Shape

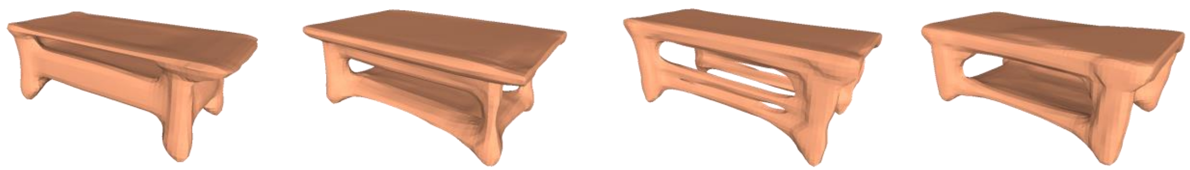


Back: size --- full
front view --- square
Seat: shape --- square

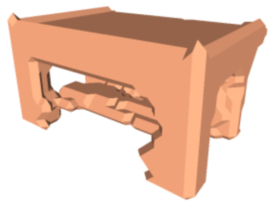


A chair with a high back
and a square shape

Appendix B: Comparison to Text2Shape



Top: shape --- rectangular
Leg: number --- four
Side: connection --- closed



A rectangular table with four linked legs