Supplementary Materials

1. Implementation details

1.1. StructureNet losses

We train LSD-StructureNet with the loss

$$\mathcal{L} = \mathcal{L}_{var} + \mathcal{L}_{recon} + \mathcal{L}_{sc} \quad (1)$$

where $\mathcal{L}_{var}$ is the variational loss defined in section 3.3. We proceed to briefly summarize the meanings of losses $\mathcal{L}_{recon}$ and $\mathcal{L}_{sc}$ that were originally introduced in [1].

The reconstruction loss $\mathcal{L}_{recon}$ seeks to evaluate the best possible correspondence between input shape $\mathcal{S}$ and output $\mathcal{S}'$. This is achieved by computing a linear assignment between parts of the input and output. Parts are matched via comparison of the geometries of the parts in each shape hierarchy. This assignment is then evaluated based on the geometry of the resulting pairs of parts, which are compared on the one hand via chamfer distance and via bounding box normals when the geometry is represented as bounding boxes. The existence of parts $\mathbf{P}$, edges $\mathbf{R}$, leaf nodes in $\mathbf{H}$ and semantic labels are also compared via cross-entropy.

The structure consistency loss $\mathcal{L}_{sc}$ seeks to enforce consistency between the geometric relationships $\mathbf{R}$ between siblings of a parent node in $\mathbf{H}$ (recall that these can be of types adjacency or rotational/translational/reflective symmetry) and the geometry $\mathbf{P}$ of output shapes $\mathcal{S}$. Symmetries are evaluated by computing an affine transformation between the bounding boxes fitted to the point clouds corresponding to the parts connected by each edge in $\mathbf{R}$, that transforms one part such that the symmetry is satisfied. The chamfer distance between the transformed and non-transformed point clouds are summed over $\mathbf{R}$ to constitute the final symmetry loss. For edges in $\mathbf{R}$ that are categorized as adjacencies, the loss is simply the minimum distance between geometries summed over $\mathbf{R}$.

1.2. Training details

We encode shape part geometry (bounding boxes or point clouds) into 256-dim feature vectors. We dimension the encoding and decoding LSTMs such that they respectively output and input 512-dim feature vectors respectively. Aside from the input dimensions of the graph decoder g_{dec} being 512 instead of 256, it and graph encoder g_{enc} are identical to those of StructureNet. In terms of hyperparameters, we train LSD-StructureNet with the same weights attributed to each of the losses described in the previous paragraph, optimizer, learning rate, weight and learning rate decay, etc. as vanilla StructureNet.

Category	Method	Coverage↓	Quality↓	FPD↓
Chair	PQ-Net [1]	**8.9**	116.5	**28.9**
	LSD-SNet	25.4	**47.2**	39.3
Lamp	PQ-Net [1]	7.10	110.3	**51.2**
	LSD-SNet	**5.9**	**67.8**	141.3

Table 1. Quality, coverage and FPD of a set of 1000 sampled shapes characterized by bounding box geometry sampled from StructureNet, LSD-StructureNet and PQ-Net.

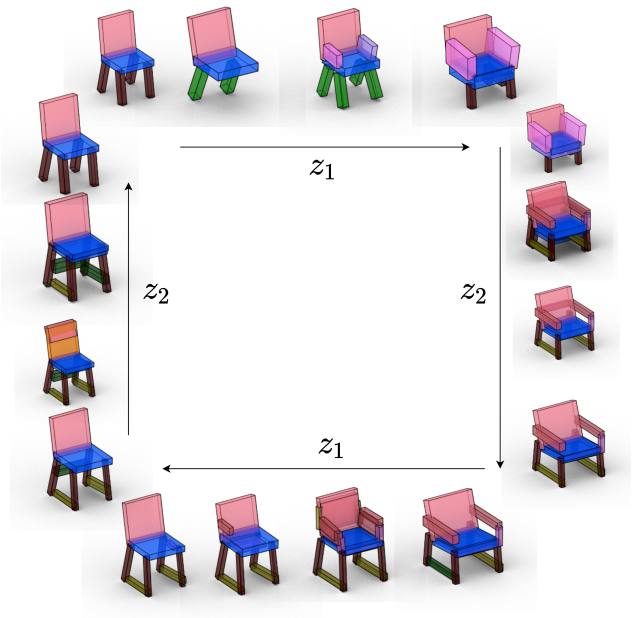

Figure 1. From a starting point at the top-left of the square decoded from a sequence of latent vectors $\mathbf{z}$, we interpolate between both z_1 and another latent vector z_1', and z_2 and another latent vector z_2' and decode the resulting sequences.

Figure 2. Conditional outputs. For 5 shapes each for StructureNet and LSD-StructureNet (1st columns), we generate 9 other conditioning shapes (other columns) that only differ with respect to the subhierarchies of a given node (LSD-StructureNet) or are as similar as rejection sampling allows (StructureNet). We provide results for each penultimate node of Chair, Table and Cabinet/Storage hierarchies. Best viewed zoomed in.

2. Comparison with PQ-Net

We supplement our comparison to vanilla StructureNet in Section 1 with a comparison to PQ-Net [2], a Seq2Seq model that can encode and decode sequences of directly observable part geometries of PartNet shapes (i.e. parts $\in \mathbf{P}$ at leaf nodes of part hierarchies $\mathbf{H}$ of shapes $\mathcal{S} = (\mathbf{P}, \mathbf{H}, \mathbf{R})$). We compare by sampling and decoding 1000 shapes

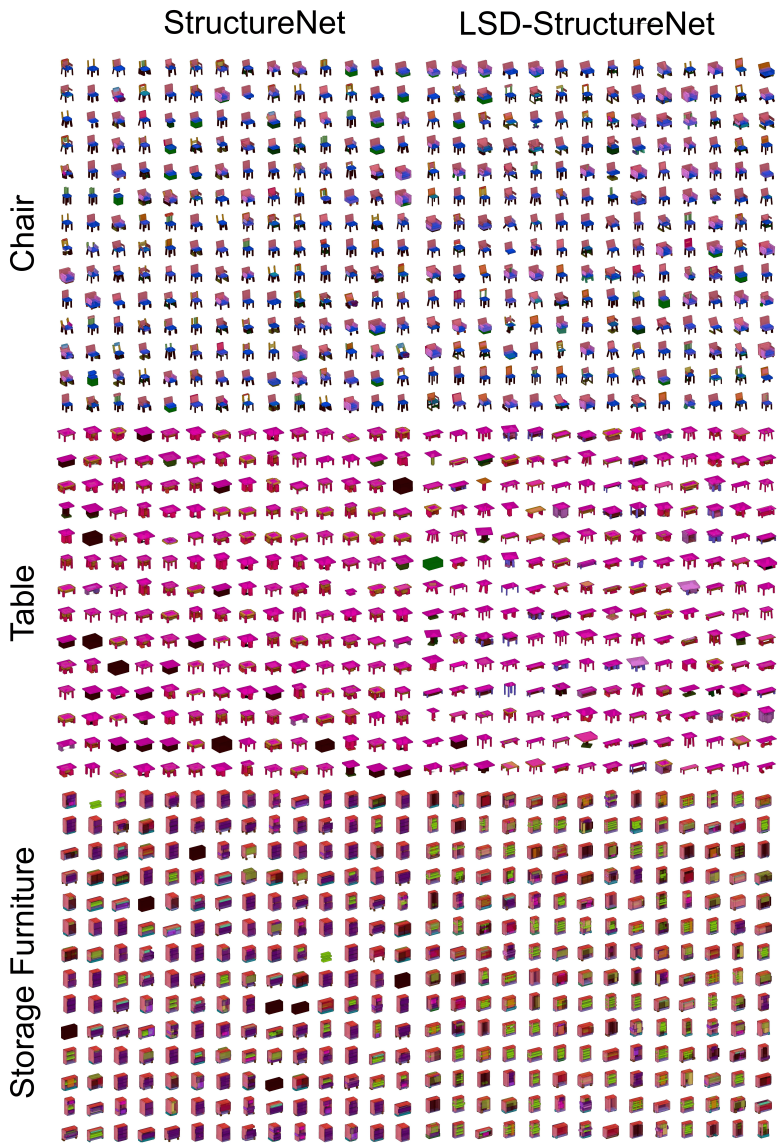

Figure 3. Unconditional outputs. We generate 100 Chairs, Tables and Cabinet/Storage shapes for both LSD-StructureNet and StructureNet. Best viewed zoomed in.

from each method and report quality, coverage and FPD for shapes in the Chair and Lamp categories as these were the only two categories with available pretrained models for PQ-Net.

PQ-Net generates and is trained on sequences of parts as opposed to hierarchies. As such, it is not possible to augment PQ-Net directly so it can model intermediary levels of structural detail as we do with LSD-StructureNet. Despite this, while we outperform PQ-Net in terms of quality and coverage, it exhibits far stronger FPD than LSD-StructureNet, that mirror their similarly strong performance against StructureNet on similar metrics [1]. This incentivizes potential future work consisting of consolidating the design choices of PQ-Net (obtention of latent space via Latent GAN instead of VAE and prediction of sequences, as opposed to graphs, of parts) while retaining the hierarchical structure of LSD- and vanilla StructureNet inputs and outputs.

3. Visualizing outputs

We decode several different $\mathbf{z}$, linearly interpolating the 1st and 2nd vectors in the sequence between 2 extremes and

visualizing the resulting outputs in Figure 1 to provide intuition as to the significance of the different latent spaces. Note that the parts of PartNet object hierarchies with semantic category *chair arm* fade in (top row) or out (bottom row) when varying z_1, as they are situated at depth 1 of PartNet object hierarchies. The children of *chair arm* parts are in this case leaf nodes with corresponding semantic categories *arm sofa style* (top-right corner) and *arm horizontal bar* (bottom-right corner) at depth 2, which is why varying z_2 produces interpolations between these two types (right column). In contrast, z_2 does not modify the structure at depth 1 and thus does not affect the presence of arms when varying it. We also qualitatively compare our outputs and those of StructureNet in Figures 2 and 3.

References

[1] Kaichun Mo, Paul Guerrero, Li Yi, Hao Su, Peter Wonka, Niloy J Mitra, and Leonidas Guibas. Structurenet: hierarchical graph networks for 3d shape generation. *TOG*, 38(6):1–19, 2019. 1, 3

[2] Rundi Wu, Yixin Zhuang, Kai Xu, Hao Zhang, and Baoquan Chen. Pq-net: A generative part seq2seq network for 3d shapes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2