SparseFlex: High-Resolution and Arbitrary-Topology 3D Shape Modeling

Supplementary Material

A. VAE Training Losses

For training our SparseFlex VAE, we use the losses described below:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{render}} + \lambda_2 \mathcal{L}_{\text{prune}} + \lambda_3 \mathcal{L}_{\text{KL}} + \lambda_4 \mathcal{L}_{\text{flex}}, \quad (1)$$

where $\lambda_1 = 1.0$, $\lambda_2 = 0.2$, $\lambda_3 = 0.001$, and $\lambda_4 = 1.0$.

We use $\mathcal{L}_{\text{render}}$ to supervise the rendered depth maps, normal maps and mask maps, which is defined as:

$$\mathcal{L}_{\text{render}} = \lambda_d \mathcal{L}_d + \lambda_n \mathcal{L}_n + \lambda_m \mathcal{L}_m + \lambda_{ss} \mathcal{L}_{ss} + \lambda_{lp} \mathcal{L}_{lp}, \quad (2)$$

where $\lambda_d = 10.0$, $\lambda_n = 4.0$, $\lambda_m = 1.0$, $\lambda_{ss} = \lambda_{lp} = 0.5$. $\mathcal{L}_{ss}$ and $\mathcal{L}_{lp}$ denote SSIM loss and LPIPS loss, and are only applied on normal maps. L1 supervision is applied for the rest of losses.

For the structure loss $\mathcal{L}_{\text{prune}}$, we use it after two upsampled blocks, which leads to cross-level self-pruning operations, enhancing more accurate reconstruction results for local details.

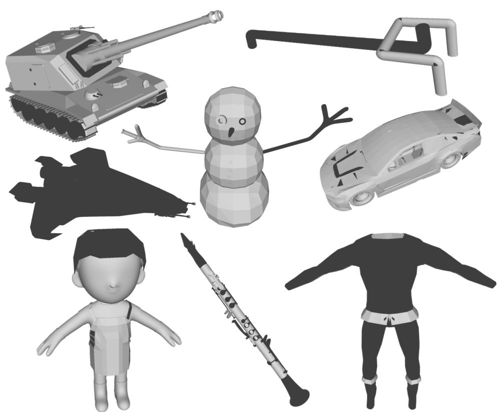

Figure 1. **Examples with inconsistent normal orientations in the large-scale dataset.**

B. Degradation in Watertight Conversion

Approaches based on SDF/occupancy field often require a time-consuming watertight conversion for constructing 3D ground-truth supervision. Most extract the double-side mesh from the UDF field computed from ground-truth mesh, retaining the maximum connected component [3] or removing the interior by calculating visibility [4]. This watertight conversion pipeline requires applying Marching Cubes [2] to extract the surface with a small iso-value, which introduces further inaccuracies and artifacts. Furthermore, the dilation

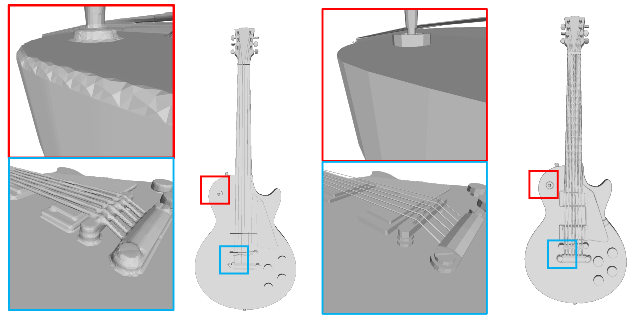

Figure 2. **Comparison of mesh processed with watertight conversion (right) and raw mesh (left).**

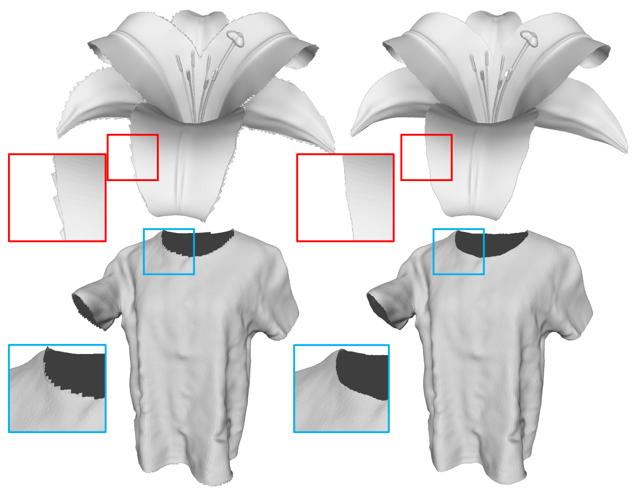

Figure 3. **Effects of self-pruning for open surface shapes.** It exhibits noticeable jagged artifacts without applying self-pruning in upsampling module.

introduced by Marching Cubes applied to UDF makes it challenging to preserve the sharp features of raw data. Fig. 2 demonstrates the comparison between the mesh converted by code scripts from Dora [1] and raw mesh, it exhibits significant degradation of details in watertight mesh (right).

C. Self-Pruning for Open Surface

After dense upsampling, voxels near the open boundaries are often redundant and cause noticeable jagged artifacts for the open surface, leading to degraded perceptual quality. In that case, we incorporate self-pruning for SparseFlex VAE. Figure 3 shows the comparisons with and without the apply self-pruning. The visualization demonstrates that self-pruning effectively reduces the artifacts around open boundaries.

References

[1] Rui Chen, Jianfeng Zhang, Yixun Liang, Guan Luo, Weiyu Li, Jiarui Liu, Xiu Li, Xiaoxiao Long, Jiashi Feng, and Ping Tan. Dora: Sampling and benchmarking for 3d shape variational auto-encoders. *arXiv preprint arXiv:2412.17808*, 2024. 1

[2] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *SIGGRAPH*, pages 163–169, 1987. 1

[3] Peng-Shuai Wang, Yang Liu, and Xin Tong. Dual octree graph networks for learning adaptive volumetric shape representations. *ACM Transactions on Graphics (TOG)*, 41(4):1–15, 2022. 1

[4] Longwen Zhang, Ziyu Wang, Qixuan Zhang, Qiwei Qiu, Anqi Pang, Haoran Jiang, Wei Yang, Lan Xu, and Jingyi Yu. Clay: A controllable large-scale generative model for creating high-quality 3d assets. *ACM TOG*, 43(4):1–20, 2024. 1