

# Faithful Contouring: Near-Lossless 3D Voxel Representation Free from Iso-surface

## Supplementary Material

### 1. Face Reduction Performance

To assess the decimation robustness and geometric stability of the reconstructed surfaces, we apply the Fast Quadric Mesh Reduction (FQMR) algorithm (Garland and Heckbert [1]) to each mesh. Under a strict maximum geometric error tolerance of  $10^{-6}$ , our reconstructed meshes still achieve an average **91.32% face-reduction efficiency** while preserving visually indistinguishable geometry. Even after such aggressive simplification, the reconstructed surfaces continue to exhibit **remarkable planar smoothness and well-preserved edge sharpness**, indicating that the proposed method produces surfaces with high geometric consistency, smoothness, and excellent decimability under strict error constraints. Examples are presented in Fig. 1.

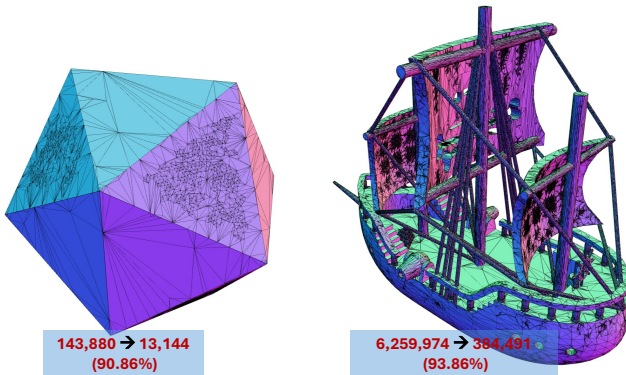


Figure 1. Applying FQMR decimation under a strict geometric error tolerance of  $10^{-6}$ , Faithful contouring maintains extremely high face-reduction efficiency (87.04%–95.11%) while preserving the global structure and fine geometric details. Even after aggressive simplification, the reconstructed surfaces retain planar smoothness, sharp edges, and clean topology, demonstrating the strong geometric stability and decimability of our reconstruction pipeline.

### 2. High-Fidelity Detail Preservation

Fig. 2 provides a visual comparison of detailed surface reconstruction at resolutions of  $1024^3$  and  $2048^3$ . While SparC3D at  $1024^3$  loses high-frequency structures and introduces voxel-lattice artifacts, our **Faithful Contour** representation preserves sharp edges, fine ornamental geometry, and consistent surface smoothness even at the same resolution. At  $2048^3$ , Faithful Contour captures micro-structures with near-lossless accuracy, demonstrating strong

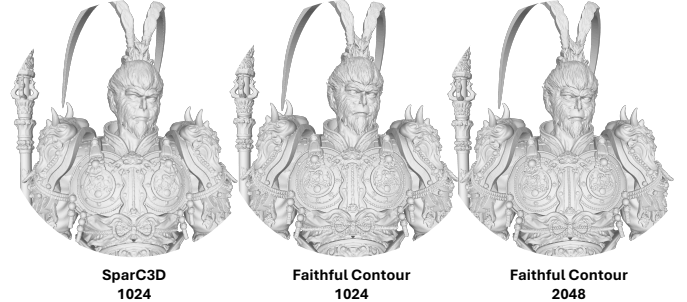


Figure 2. **Visual comparison of high-frequency detail preservation.** From left to right: SparC3D ( $1024^3$ ), Faithful Contour ( $1024^3$ ), and Faithful Contour ( $2048^3$ ). Our method preserves sharp edges, crisp relief patterns, and smooth surfaces at the same resolution, and recovers nearly all micro-structures at  $2048^3$ .

scalability and superior detail retention compared with existing voxelized methods.

### 3. Efficiency & Computational Cost

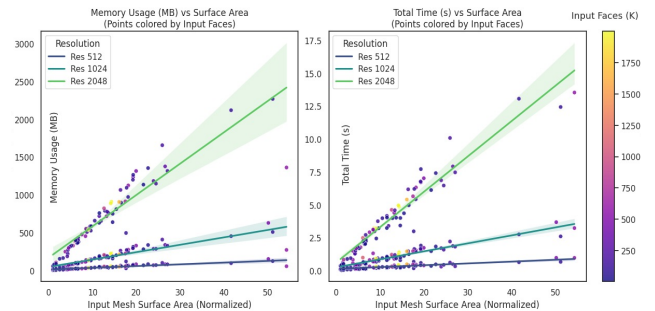


Figure 3. **Scalability of FC.** Memory usage and total runtime (encoding + decoding) as functions of input surface area, showing approximately linear scaling across resolutions 512–2048.

Fig. 3 shows the cost of FC at resolutions from 512 to 2048. Using GPU-accelerated local operations without global optimization or iterative solvers, FC stores data only at surface-intersecting voxels. Thus, memory and runtime scale linearly with mesh surface area, not cubic resolution or input mesh face count. In practice, the total processing memory/time grows from 43 MB / 0.33 s (512) to 711 MB / 4.14 s (2048) for meshes up to 34.7M faces on a single H100 GPU.

### 4. VAE Implementation Details

Our VAE implementation adopts a structure similar to those in SparseFlex [2] and Sparc3D [3]. The encoder takes ge-

ometry features as input and performs three downsampling stages using sparse convolutional blocks with residual connections. Following this, four local attention blocks [4] are applied to further aggregate information in the latent space.

For the decoder, we follow the SparseFlex design but replace the swin-attention blocks [5] with local attention blocks. The number of upsampling blocks is increased to three to restore the original input grid resolution. Each upsampling block is augmented with an additional linear layer to predict voxel occupancy and prune redundant voxels.

In the case of point inputs, a lightweight local attention module is incorporated to aggregate point-level features within their corresponding voxels into voxel-level representations.

During the scenerio training higher resolution (e.g.  $1024^3$ ), we employ a chunk-wise upsampling strategy with voxel-level supervision to reduce GPU memory usage. Specifically, we first randomly select one of the eight corners from the dense volume, then expand the coordinates along all three axes to extract the bounding box w.r.t. voxel chunks for upsampling. The expansion ratio is adaptively determined by a predefined threshold based on the number of voxels in each object, ensuring efficient utilization of GPU memory.

The loss weights used in the final optimization objective are set as follows:  $\lambda_x = 0.5$ ,  $\lambda_n = 0.01$ ,  $\lambda_{axis} = 0.5$ ,  $\lambda_{mask} = 1.0$ ,  $\lambda_{occ} = 0.2$ ,  $\lambda_{KL} = 10^{-5}$ .

## 5. FC in 3D Generative Task

For validating the effectiveness of Faithful Contouring in generative tasks, we build a diffusion framework upon the introduced VAE. Here we demonstrate the qualitative results of the Image-to-3D generation results in Fig. 4. From which it is able to see that FC can faithfully preserve sharp edges, details and inner structures of given images.

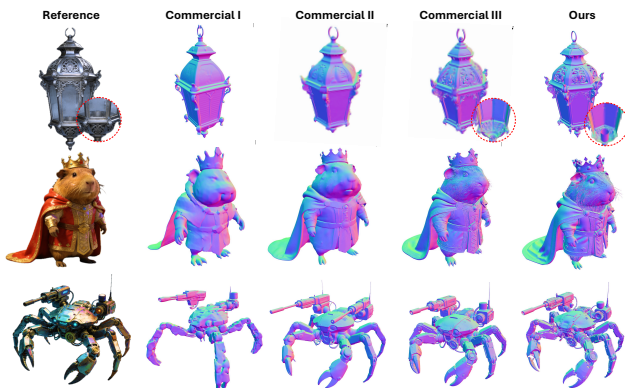


Figure 4. **Generative results of Diffusion-based model with FC.** The first row specifically shows the generation of internal structures. Others show the high fidelity of detail generations.

## References

- [1] Michael Garland and Paul S Heckbert. Surface simplification using quadric error metrics. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 209–216, 1997.
- [2] Xianglong He, Zi-Xin Zou, Chia-Hao Chen, Yuan-Chen Guo, Ding Liang, Chun Yuan, Wanli Ouyang, Yan-Pei Cao, and Yangguang Li. Sparseflex: High-resolution and arbitrary-topology 3d shape modeling. *arXiv preprint arXiv:2503.21732*, 2025.
- [3] Zhihao Li, Yufei Wang, Heliang Zheng, Yihao Luo, and Bihan Wen. Sparc3d: Sparse representation and construction for high-resolution 3d shapes modeling. *arXiv preprint arXiv:2505.14521*, 2025.
- [4] Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xi-hui Liu, Yu Qiao, Wanli Ouyang, Tong He, and Hengshuang Zhao. Point transformer v3: Simpler faster stronger. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4840–4851, 2024.
- [5] Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jialong Yang. Structured 3d latents for scalable and versatile 3d generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 21469–21480, 2025.