Supplementary Material for "UDiFF: Generating Conditional Unsigned Distance Fields with Optimal Wavelet Diffusion"

Junsheng Zhou[1*], Weiqi Zhang[1*], Baorui Ma[1,2†], Kanle Shi[3], Yu-Shen Liu[1†], Zhizhong Han[4]
School of Software, Tsinghua University, Beijing, China[1]
BAAI, Beijing, China[2], Kuaishou Technology, Beijing, China[3]
Department of Computer Science, Wayne State University, Detroit, USA[4]
zhoujs21@mails.tsinghua.edu.cn, zwq23@mails.tsinghua.edu.cn, brma@baai.ac.cn
shikanle@kuaishou.com, liuyushen@tsinghua.edu.cn, h312h@wayne.edu

1. More Visualizations

In this section, we provide more qualitative illustrations of the generation results produced by UDiFF.

Category-conditional generation on DeepFashion3D. We provide extra open-surface shape generations achieved by the UDiFF model trained under DeepFashion3D [8] dataset with the cloth categories as the conditions. Specifically, we generate 8 categories of cloth shapes, including *"long sleeve dress"*, *"long sleeve upper"*, *"pants"*, *"no sleeve dress"*, *"no sleeve upper"*, *"dress"*, *"shot sleeve dress"* and *"shot sleeve upper"*. The visualizations are shown in Fig. 2 and 3, where UDiFF generates diverse and novel shapes correctly corresponds to the text conditions.

Unconditional generation on ShapeNet. We further provide more unconditional shape generation results achieved by the UDiFF model trained under single categories of ShapeNet [1] dataset. We generate shapes of the *"chair"* and *"airpane"* categories. The visualizations are shown in Fig. 4, where UDiFF generates visual-appealing shapes.

2. Analysis on Meshing and Texturing

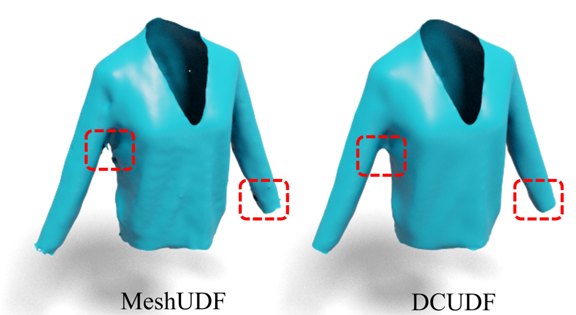

Figure 1. Mesh extraction comparisons between MeshUDF and DCUDF.

Meshing. Different from SDFs, UDFs fail to extract surfaces by the marching cubes [5] since UDFs cannot perform inside/outside tests on 3D grids. Recent works [3, 7] leverage the gradients at UDF grids as the signals to mesh UDFs. However, for the generated UDFs, the approximated gradients may not be stable enough at the zero-level set, which leads to errors and holes. The approximated gradient at a grid point q is defined as the direction from q to the neighbour grid q_n where the UDF from q to q_n increases rapidly the most. We adopt DCUDF [4] with double covering to mesh the generated UDF of UDiFF, which results in more continuous surfaces. We make an adaption to DCUDF on the double covering operation to replace the time-consuming optimizations with an explicit vertices refinement strategy. We move each vertices against the surface normals with a stride of unsigned distances to reach the zero-level sets, and then leverage the min-cut algorithm to achieve the final model. We show the comparison of meshing the generated UDF with MeshUDF [3] and DCUDF [4] in Fig. 1.

Texturing. We leverage Text2Tex [2] to generate textures for the extracted meshes. This is achieved with a progressive texture generation process and a texture refinement process. Specifically, we first render the texture-less initial mesh from the preset viewpoint and generate the appearance according to the text prompt with the depth-guided stable-diffusion [6]. We then adjust to the next preset viewpoint and repeat the appearance generation process until the last preset viewpoint where the whole mesh is textured. Finally, we optimize the textures with automatically selected viewpoints for refinement.

References

[1] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-

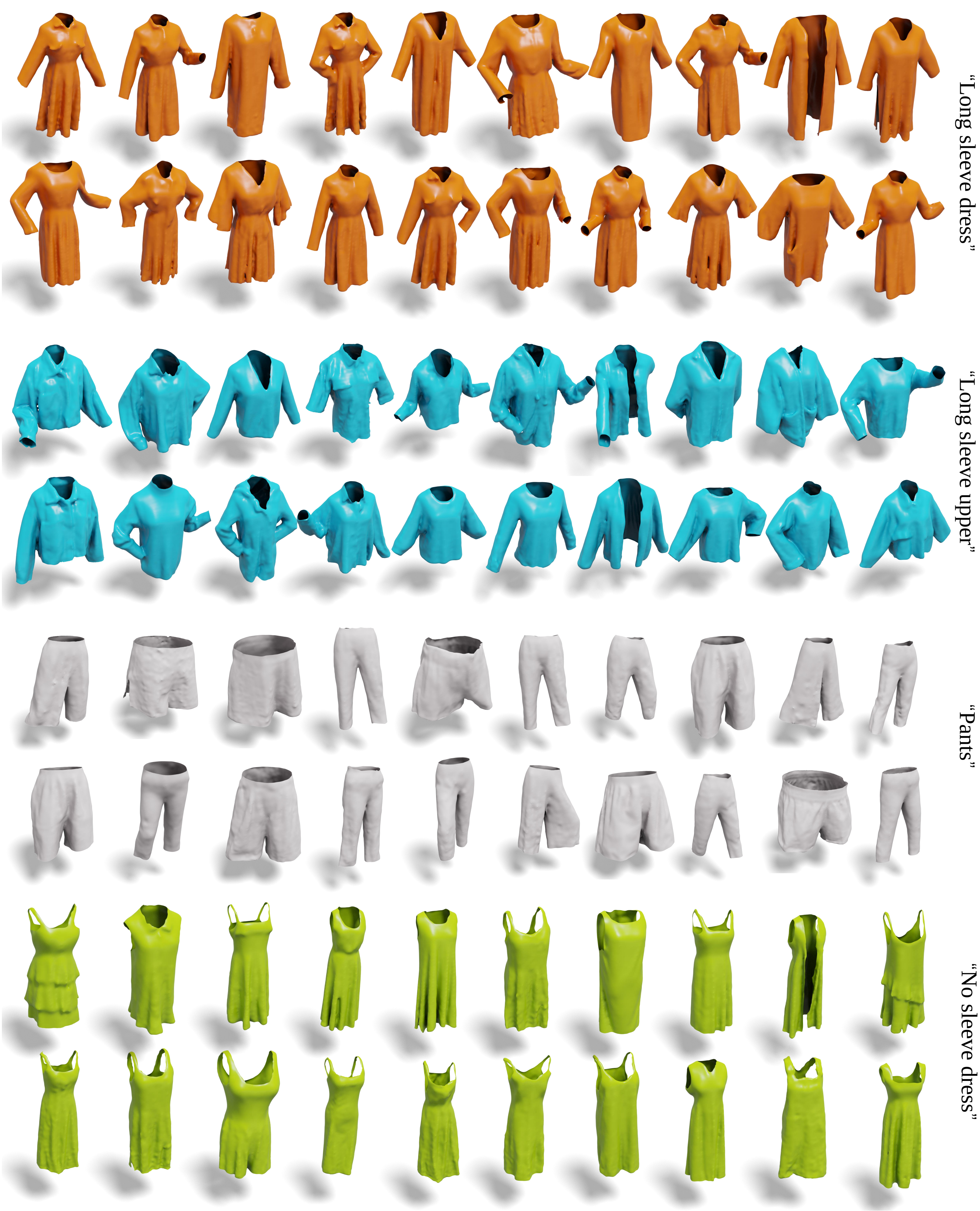

Figure 2. Category conditional generations under DeepFashion3D dataset. Here, we visualize the conditional generations of categories *"long sleeve dress"*, *"long sleeve upper"*, *"pants"* and *"no sleeve dress"*

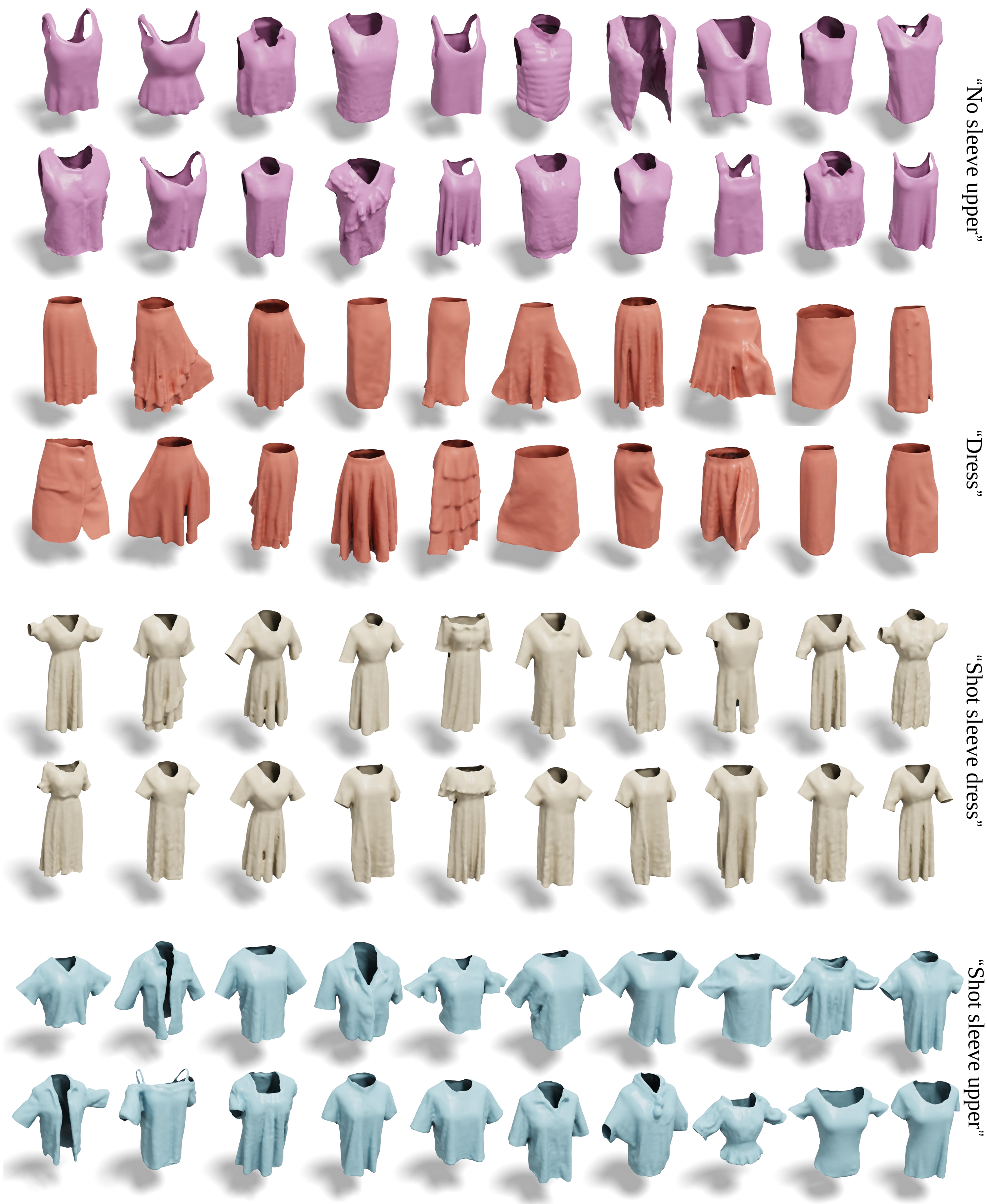

Figure 3. Category conditional generations under DeepFashion3D dataset. Here, we visualize the conditional generations of categories *"no sleeve upper"*, *"dress"*, *"shot sleeve dress"* and *"shot sleeve upper"*

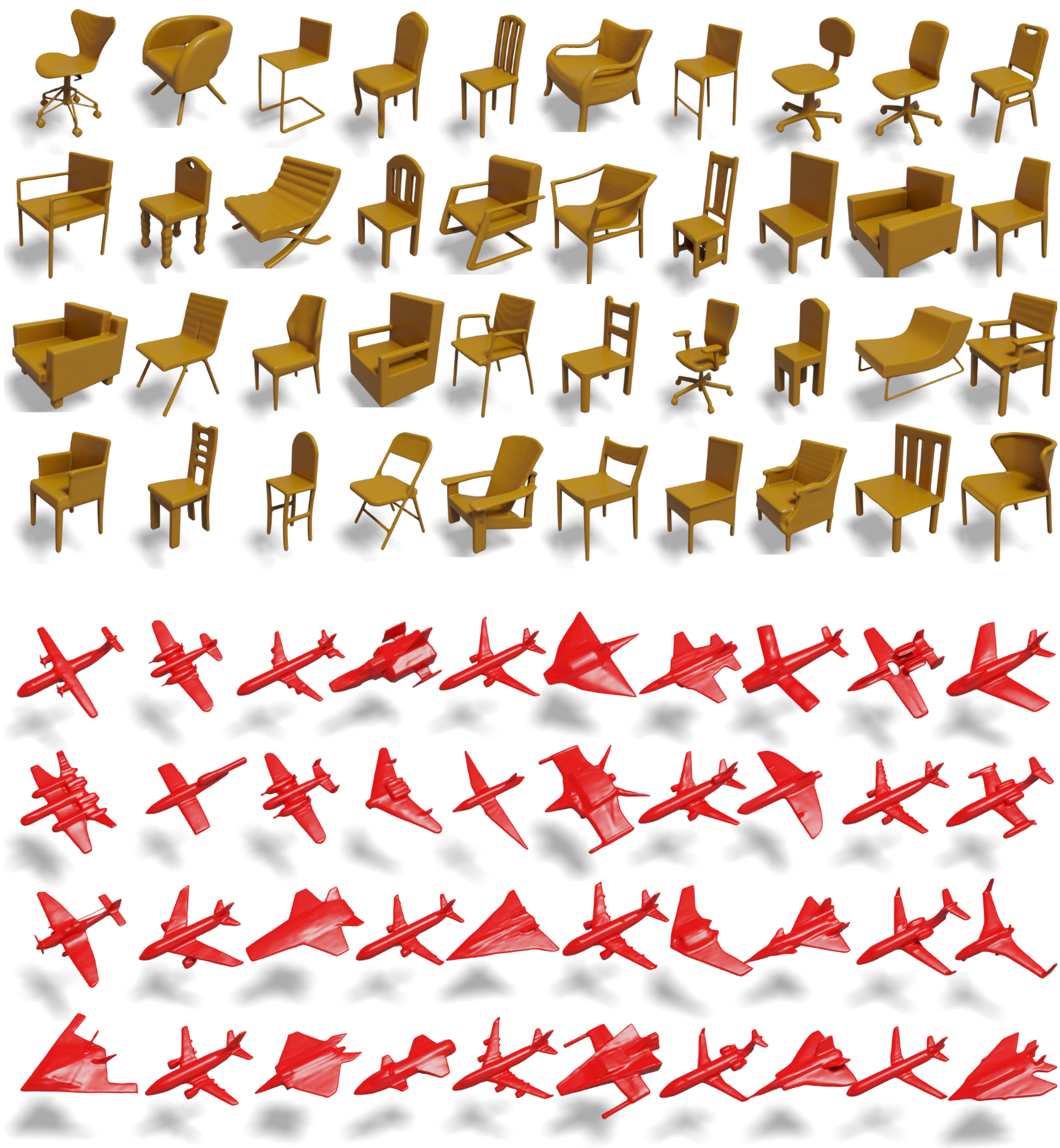

Figure 4. Unconditional generations under the *"chair"* and *"airplane"* categories of the ShapeNet dataset.

rich 3D model repository. *arXiv preprint arXiv:1512.03012*, 2015. 1

[2] Dave Zhenyu Chen, Yawar Siddiqui, Hsin-Ying Lee, Sergey Tulyakov, and Matthias Nießner. Text2tex: Text-driven texture synthesis via diffusion models. *arXiv preprint arXiv:2303.11396*, 2023. 1

[3] Benoit Guillard, Federico Stella, and Pascal Fua. MeshUDF: Fast and differentiable meshing of unsigned distance field networks. *European Conference on Computer Vision*, 2022. 1

[4] Fei Hou, Xuhui Chen, Wencheng Wang, Hong Qin, and

Ying He. Robust zero level-set extraction from unsigned distance fields based on double covering. *arXiv preprint arXiv:2310.03431*, 2023. 1

[5] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3D surface construction algorithm. *ACM Siggraph Computer Graphics*, 21(4):163–169, 1987. 1

[6] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 1

[7] Junsheng Zhou, Baorui Ma, Yu-Shen Liu, Yi Fang, and Zhizhong Han. Learning consistency-aware unsigned distance functions progressively from raw point clouds. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 1

[8] Heming Zhu, Yu Cao, Hang Jin, Weikai Chen, Dong Du, Zhangye Wang, Shuguang Cui, and Xiaoguang Han. Deep fashion3d: A dataset and benchmark for 3d garment reconstruction from single images. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 512–530. Springer, 2020. 1