Design2Cloth: 3D Cloth Generation from 2D Masks
Supplementary Material

Jiali Zheng, Rolandos Alexandros Potamias, Stefanos Zafeiriou
Imperial College London
{jiali.zheng, r.potamias, s.zafeirou}@imperial.ac.uk

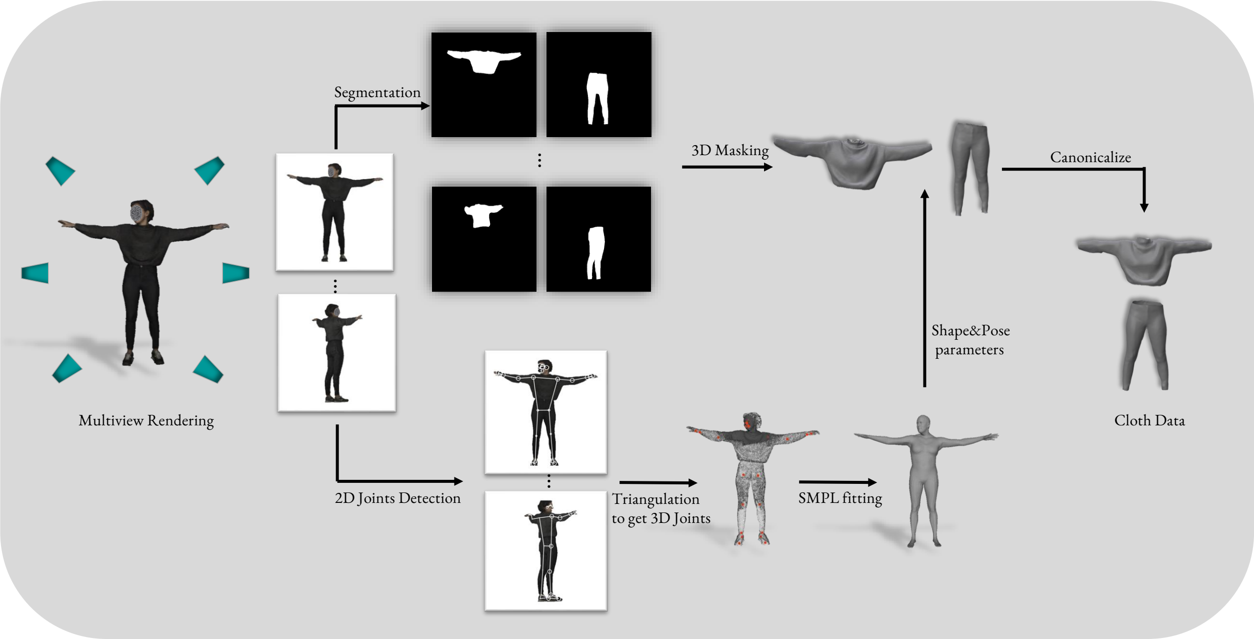

Figure 1. Pipeline for Cloth Dataset generation.

1. Dataset Curation

As mentioned in the main paper, the proposed dataset curation pipeline is composed of two main branches. The first part is related to the 3D cloth extraction whereas the second one tackles the alignment of a parametric body model to the raw data. Prior to cloth extraction, we render, using the same camera parameters, the 3D scan from multiple views. We empirically set the number of views to 40, striking a balance between runtime and triangulation results. For each one of the rendered views, we utilize SAM [5] to extract cloth segmentations. To automatically segment upper and lower garments, we use a voting scheme between the detected joints of the upper and lower body, respectively. Specifically, we utilized Mediapipe [2] to acquire 2D joint positions for view rendering. Using the obtained 2D joint positions, we can easily locate each mask. The mask which contains shoulders/elbows/spine indicates a mask for the top, while the mask that contains hips/knees/ankles is considered the bottom mask. Following this, using all different views, we lifted the 2D landmarks to 3D by linear triangulation.

In the second stage of our pipeline, we used the detected 3D joints $\mathbf{J}$ to fit the parametric SMPL body model [6]. To do so

Method	CD ($\times 10^{-2}$) ↓	NC ↑
w/o Triplane	0.47	0.92
w/o Dual Discriminator	0.28	0.96
Proposed	**0.12**	**0.98**

Table 1. Quantitative comparison of the reconstruction performance for ablation study.

we devised a multi-term loss function $\mathcal{L}$:

$$\mathcal{L} = \mathcal{L}_J + \mathcal{L}_{CD} + \lambda_{outter}\mathcal{L}_{outter} + \lambda_{reg}\mathcal{L}_{reg} \tag{1}$$

where $\mathcal{L}_J$ denotes an L_1 loss between the detected and the SMPL 3D joints, $\mathcal{L}_{CD}$ the Chamfer distance loss between the SMPL body and the raw scan, $\mathcal{L}_{outter}$ a loss that penalizes SMPL body to lie inside the raw scan, $\mathcal{L}_{reg}$ a regularization and $\lambda_{outter}, \lambda_{reg}$ weighting terms to scale the losses. More specifically, to ensure the SMPL fitting remains within the inner surface of the scan, we integrated a loss function that directs the vertices of the SMPL body, which fall outside the scan, towards the inner surface of the scan. To identify the vertices of SMPL that are outside the scan surface, we find for every vertex of SMPL body $\mathbf{v}_b$ its closest vertex on the scan $\mathbf{v}_s$ and measure the angle difference between the vector connecting the two vertices $\hat{\mathbf{v}}_{b\rightarrow s}$ and the normal vector $\mathbf{n}_b$ of vertex $\mathbf{v}_b$. If the angle of the two vectors

$$\alpha = cos^{-1}(\frac{\hat{\mathbf{v}}_{b\rightarrow s} \cdot \mathbf{n}_b}{||\hat{\mathbf{v}}_{b\rightarrow s}||||\mathbf{n}_b||}) \tag{2}$$

is greater than 90^o, we apply an L_1 contact loss to the vertex $\mathbf{v}_b$ to match the scan vertex $\mathbf{v}_s$. Finally, similar to [6], we apply a regularization to the SMPL shape parameters $\boldsymbol{\beta}$ to prevent irregular body shapes:

$$\mathcal{L}_{reg} = ||\boldsymbol{\beta}||_2 \tag{3}$$

The two branches are aggregated to normalize the cropped 3D garments to the canonical pose. In particular, for every point in the garment surface, we find its closest point in the SMPL mesh and using the linear blend-skinning of SMPL, we canonicalize the garment to the zero pose. An overview of the pipeline is depicted in Fig. 1.

2. Ablation Study

We conducted an ablation study to illustrate the functionality of the core components of the proposed model. In particular, we ablated two key components of the proposed model: a) the *Triplane* generator and b) the *Dual Discriminator*. Both network variants were trained on the same training/test set with the proposed method.

Tri-plane Generator. We built the proposed model on a tri-plane generator, which can effectively encode high frequency details and increase the capacity of the network, without affecting the memory requirements. As can be observed in Fig. 2, the proposed model trained without the tri-plane generator, despite maintaining an overall appropriate garment style, fails to encode high frequency details and learns approximations of the wrinkles that result in non-smooth surfaces. This can be quantitatively validated in Tab. 1, where the model with the ablated tri-plane generator exhibits a significant performance drop on the normal consistency (NC) metric. Intuitively, the tri-plane generator simplifies the task of the decoder module, by learning an implicit-explicit representation of the grid.

Dual Discriminator. The second component of the proposed network is the Dual Discriminator that enforces the generator module to produce realistic clothes. Motivated by the lack of high frequency details on the generated clothes, we built a dual branch discriminator that can not only guide the generation of realistic styles, but also enforce high frequency details such as wrinkles and creases on the generated garments. The discriminator takes as input a dual representation of the cloth considering both the global and the local details of the cloth, which enforces the generation of high frequency details. As we showed in Fig. 3, sampling points based on their curvature results in points that span in the detailed regions of the cloth which improves the generation of high frequency details.

As can be easily observed on Fig. 2, the exclusion of the dual discriminator leads to smooth surface garment generation, lacking intricate details such as wrinkles and creases. Quantitatively, this results in a performance drop on both the Chamfer distance (CD) and the normal consistency (NC) metrics, as depicted in Tab. 1.

3. Additional Results

In this section we present additional qualitative results of the proposed method, aligned with the main paper.

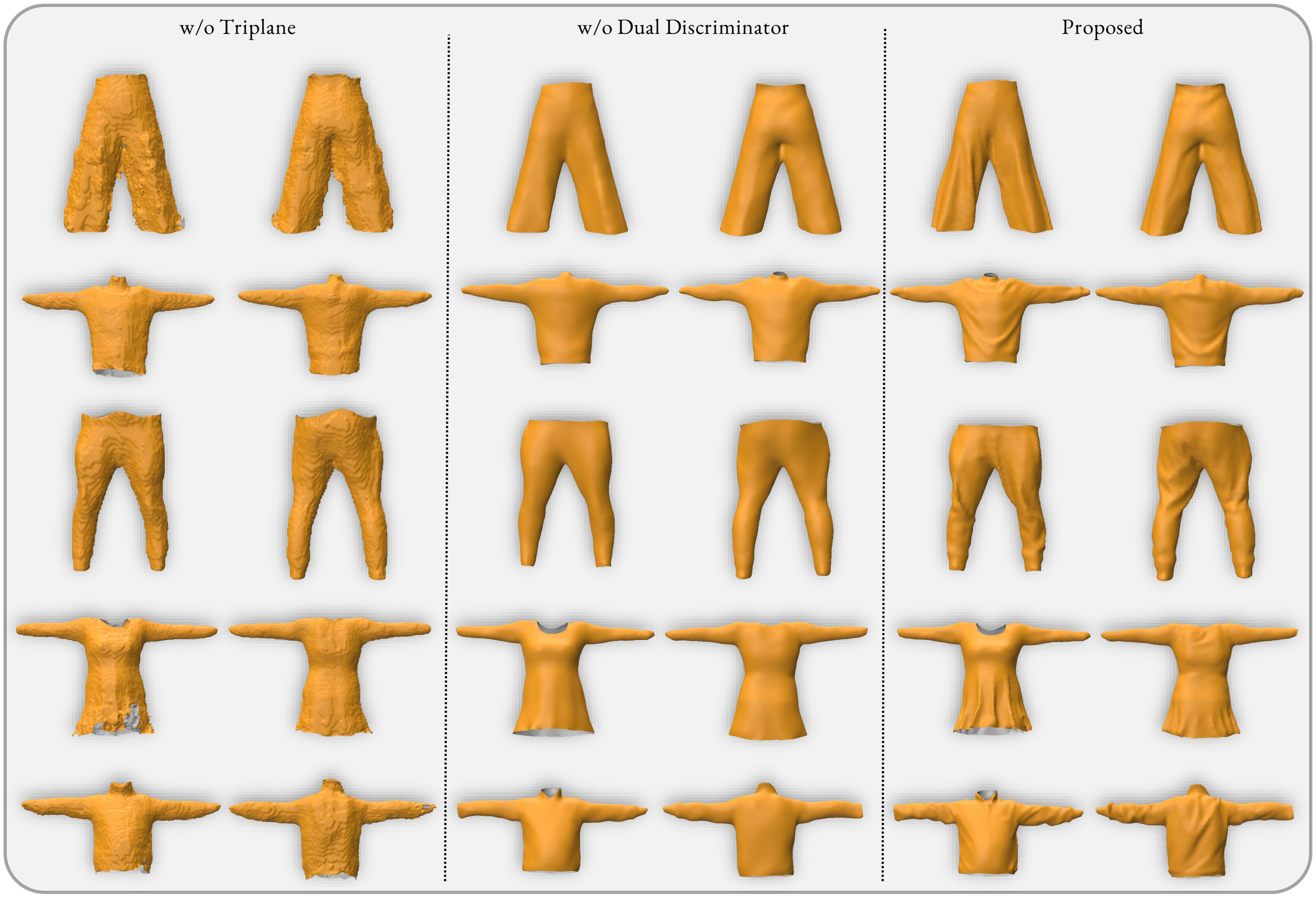

Figure 2. **Ablation Study.** Qualitative evaluation of the main components of the proposed model.

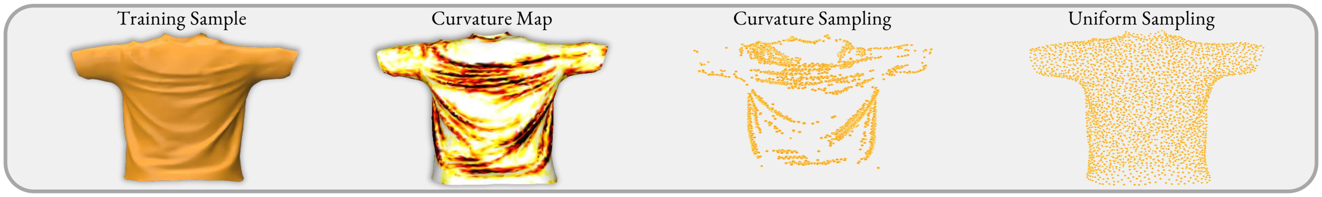

Figure 3. **Dual Discriminator**: Illustration of points sampled by their curvature.

3.1. Mask to Cloth Generation

To showcase the generative power of the proposed model we selected as input a set of diverse 2D cloth visibility masks. In particular, as shown in Fig. 4, we evaluated the model in generating complex and challenging garments, like pleated skirts and bell-sleeve tops which develop simple everyday clothing styles, such as T-shirts and trousers. Additionally, we explore the generation of garments with asymmetric sleeve and leg lengths, as shown in the last two samples of Fig. 4. The results demonstrate the capability of the model in producing realistic outputs, that are aligned with the condition mask, which can undoubtedly provide a powerful tool that aids fashion technology. Furthermore, the model exhibits a commendable ability to interpolate and generate realistic results for garments with asymmetric features. This underscores the ability of the model in cloth design through the simple input of a 2D mask.

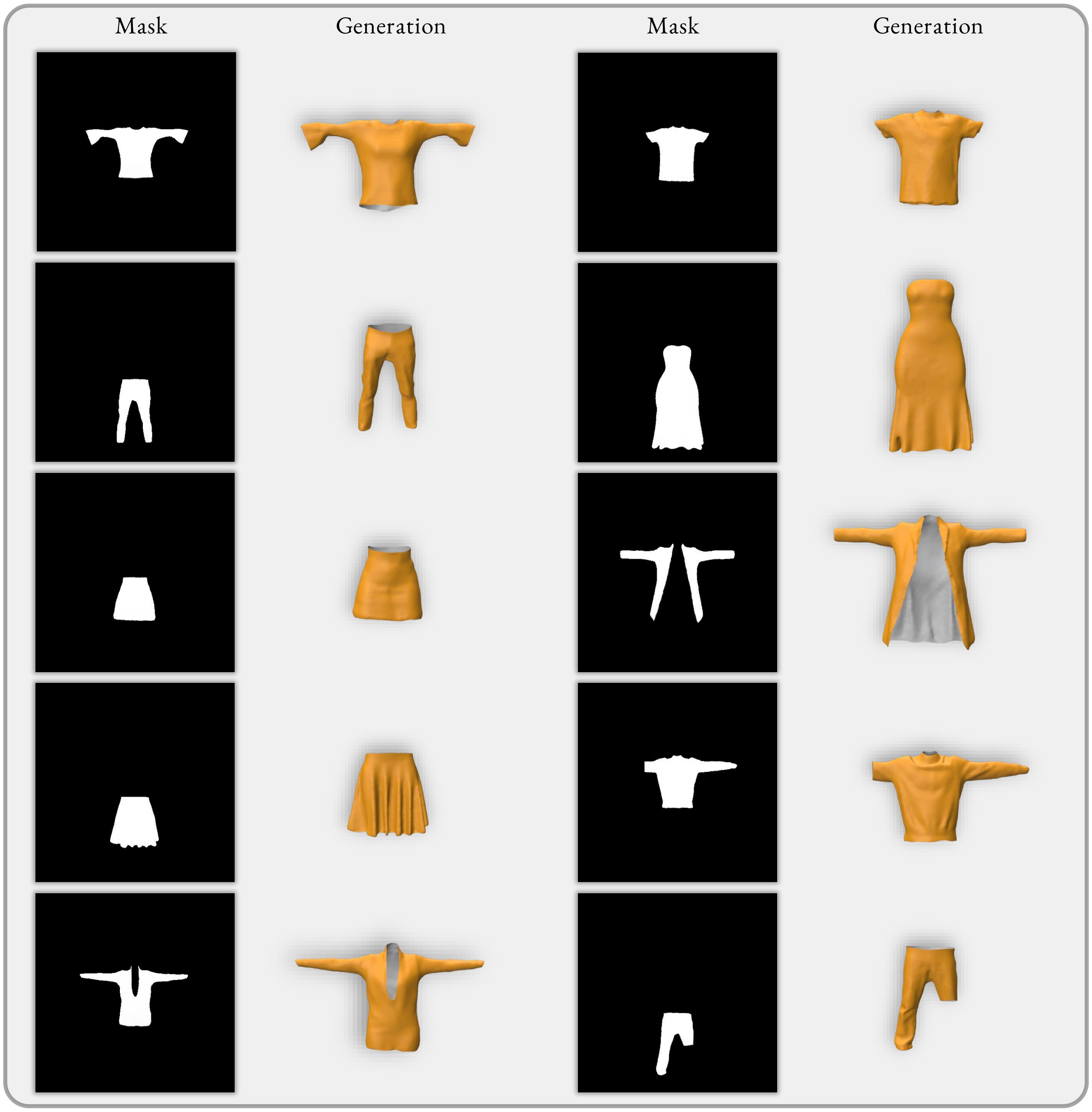

Figure 4. **Additional Results:** Garment Generations from 2D masks.

3.2. Garment Reconstruction Comparison

In Fig. 5 further examples of cloth mesh reconstruction are presented. To qualitatively evaluate the reconstruction performance of the proposed method we selected several everyday clothing types. Additionally, we report the reconstruction performance of our model on two different cloth datasets, namely Cloth3D [3] and ClothesNet [8]. As can be seen in Fig. 5, it is evident that the proposed model adeptly reconstructs clothing styles, preserving natural wrinkle details. In contrast, DrapeNet [4] tends to generate excessively smooth meshes that lack details and creases. In the case of Cloth3D [3] garments, our model excels in replicating the overall mesh shape compared to Drapenet. Interestingly, Design2Cloth achieves to successfully reproduces the overall shape of the dress from ClothesNet [8] dataset, while DrapeNet falls short in achieving a cloth-like result.

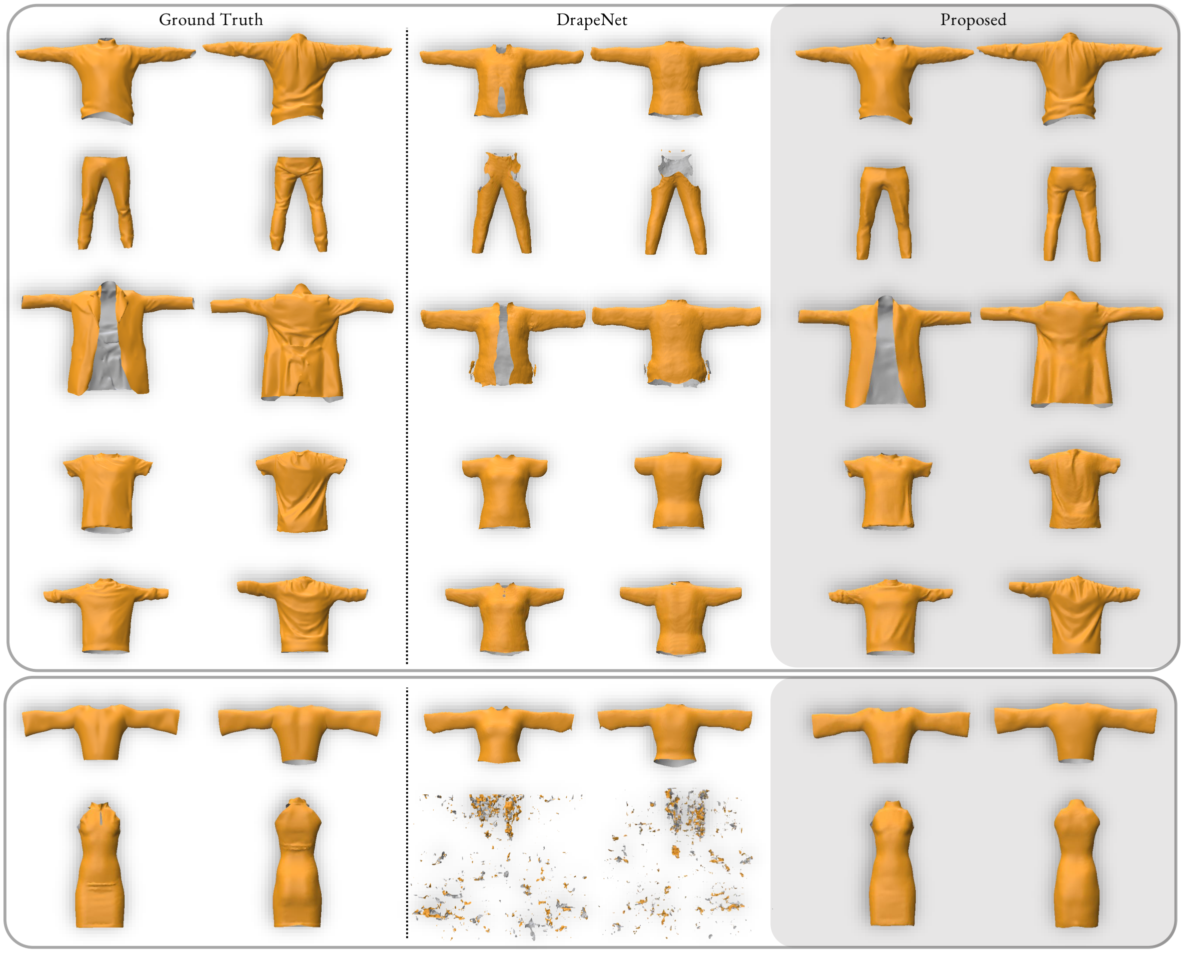

Figure 5. **Additional Results:** Garment reconstruction comparison. Top: DigitalMe data; Bottom: Cloth3D [3] and ClothesNet [8] data.

3.3. Interpolation

Beyond the interpolation outcomes detailed in the main paper, we showcase further interpolations, on both style and shape latent components. Fig. 6 illustrates the ability of the proposed method to generate smooth interpolations between diverse cloth styles. For better visual evaluation please see also the supplementary video.

3.4. Generation Diversity

To showcase the diversity of our model, we measured the similarity between training and generated distributions. We presented the result of measuring Jensen-Shannon Divergence (JSD), Coverage (COV), Minimum matching distance (MMD)[1], and 1-nearest neighbor accuracy (1-NNA)[7] for point clouds in the training set and generated set in Tab. 2.

	JSD ($\times 10^3$) ↓	MMD($\times 10^3$) ↓	COV(%,↑)	1-NNA (%,↓)
DrapeNet	9.036	6.952	44.73	76.14
Proposed	**2.753**	**1.638**	**51.04**	**62.21**

Table 2. Generation results.

3.5. Generation Detail

We showed the generalization capabilities of the network in Fig. 7 where the generated samples have different high frequency details compared to their closest training sample in the latent space.

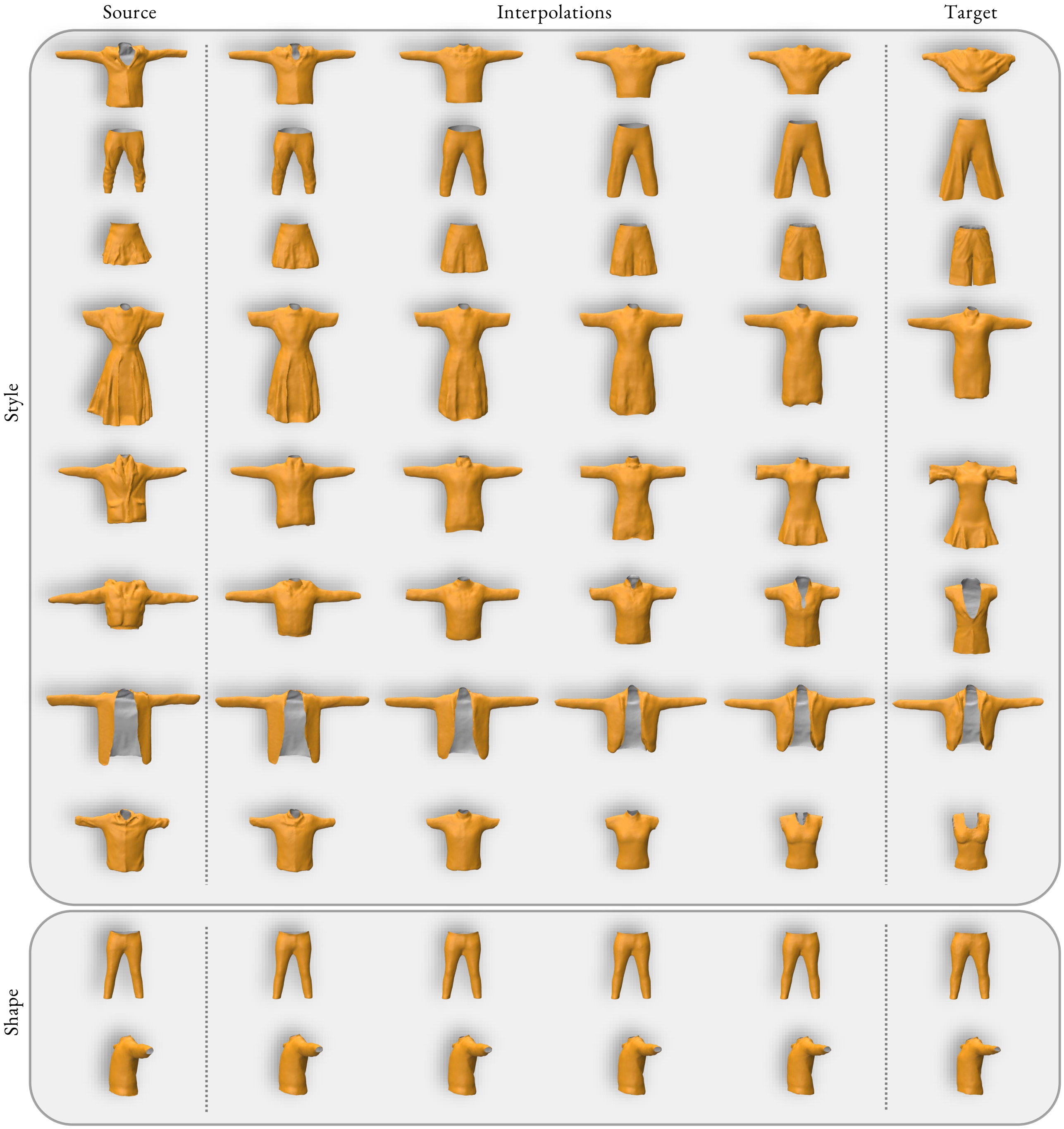

Figure 6. **Additional Results:** Interpolation results between diverse garment styles and shapes.

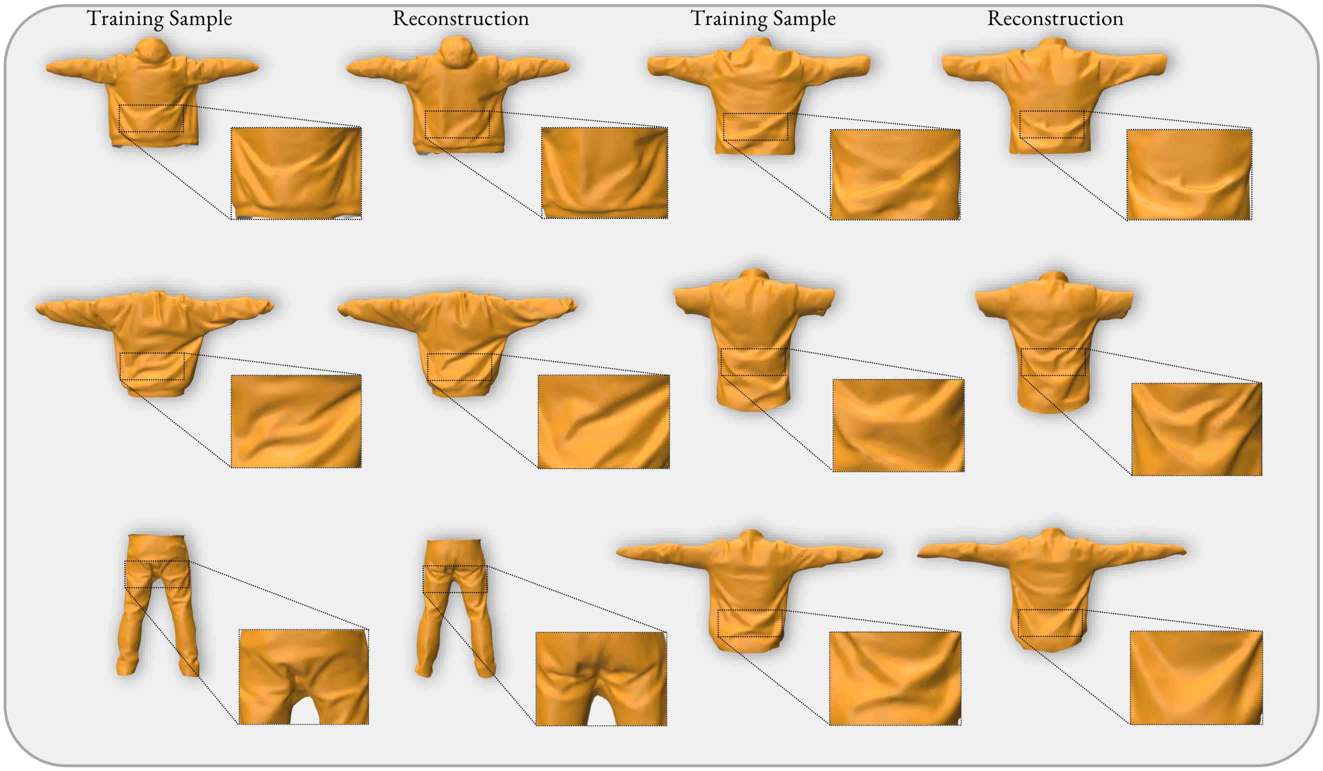

Figure 7. **Generation Detail**: Reconstructions and the closest training samples.

References

[1] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3d point clouds. In *International conference on machine learning*, pages 40–49. PMLR, 2018. 5

[2] Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann. Blazepose: On-device real-time body pose tracking, 2020. 1

[3] Hugo Bertiche, Meysam Madadi, and Sergio Escalera. Cloth3d: clothed 3d humans. In *European Conference on Computer Vision*, pages 344–359. Springer, 2020. 4, 5

[4] Luca De Luigi, Ren Li, Benoît Guillard, Mathieu Salzmann, and Pascal Fua. Drapenet: Garment generation and self-supervised draping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1451–1460, 2023. 4

[5] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. *arXiv:2304.02643*, 2023. 1

[6] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, 2015. 1, 2

[7] David Lopez-Paz and Maxime Oquab. Revisiting classifier two-sample tests. *arXiv preprint arXiv:1610.06545*, 2016. 5

[8] Bingyang Zhou, Haoyu Zhou, Tianhai Liang, Qiaojun Yu, Siheng Zhao, Yuwei Zeng, Jun Lv, Siyuan Luo, Qiancai Wang, Xinyuan Yu, et al. Clothesnet: An information-rich 3d garment model repository with simulated clothes environment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 20428–20438, 2023. 4, 5