

# FreeCloth: Free-form Generation Enhances Challenging Clothed Human Modeling

## Supplementary Material

In Sec. A, we elaborate on the implementation details of our proposed method and the experimental setups. We provide additional results and extended discussions in Sec. B.

### A. Implementation Details

#### A.1. Model Architecture

In the implementation of our pose encoder network  $\mathcal{E}_d$ , the PointNet++ [10] abstracts the point features for  $L = 4$  levels, and the numbers of the abstracted points are 2048, 512, 128, and 32 at each level, respectively. Both the local and global pose codes share a feature dimensionality of  $M_p = 256$ , whereas the garment code is represented by a  $M_g = 64$ -dimensional learnable parameter.

The structure-aware pose encoder  $\mathcal{E}_g$  for extracting pose feature embedding for the free-form generation module possesses a similar architecture with  $\mathcal{E}_d$ . Given our focus on modeling skirts and long dresses, we selectively sample posed points from  $K_b = 4$  local parts situated on the legs, including the left upper leg, left lower leg, right upper leg, and right lower leg. Although the short skirt doesn't directly cover the lower legs, their pose still indirectly affects the skirt's movement. Specifically, we uniformly sample 2048 points from each part, which are then inputted into  $\mathcal{E}_g$  to derive part-aware local features. A final global max-pooling layer is prepended to extract the global pose features.

As for the free-form generator  $\mathcal{G}$ , We modify a simple yet effective style-based point generator, SpareNet [13]. SpareNet employs point morphing techniques to map a unit square  $[0, 1]^2$  onto a 3D surface. Specifically, we utilize  $K$  surface elements (8 in our experiments) to construct the loose garment. For simplicity, we omit the refiner module and adversarial rendering. Empirically, we observe that refinement following the hybrid modeling of the garment doesn't yield performance improvements. The number of generated points, denoted as  $N_g$ , is manually configured to either 32768 for long dresses or 16384 for skirts.

#### A.2. Garment-specific Clothing-cut Map

Here we provide comprehensive details on computing the garment-specific clothing-cut maps, as outlined in the main paper. Following the methodology in POP [7], all baseline approaches [5, 8, 14] uniformly sample point sets from the UV map at a resolution of  $256 \times 256$ . Specifically,  $N_d = 47911$  points are sampled. We start by segmenting the unclothed regions, including the head, hands, and feet, which contain up to  $N_u = 13240$  points.

Then we apply the off-the-shelf image segmentation model, SAM [4], to automatically identify the loose region. Specifically, we select the frame that closely resembles the canonical pose in the training sequence and render the front and back view normal maps to cover all body points. These normal maps are fed into SAM to locate loose clothing including skirts and dresses. The segmented results are shown in Fig. A1. We back-project the detected pixel coordinates into 3D space and employ nearest neighbor search to assign each point on the UV map to the full scan, filtering the corresponding loose parts on the body surface. The extracted clothing-cut maps for all 5 subjects from the ReSynth [7] dataset are visualized in Fig. A2.

To ensure fair comparisons, we merge points from three sources, *i.e.* combining  $N_u$ ,  $N_d$ , and  $N_g$  points, and employ farthest point sampling (FPS) to obtain the final full point set with  $N = 47911$  points to match the baselines [7, 8].

#### A.3. Training

We train our network for 1000 epochs on the ReSynth [7] dataset, using the Adam [3] optimizer with a batch size of 8 and a learning rate of  $3.0 \times 10^{-4}$ . The loss weights are set to  $\lambda_p = 1 \times 10^4$ ,  $\lambda_n = 1.0$ ,  $\lambda_{rd} = 2 \times 10^3$ ,  $\lambda_{rg} = 1$  and  $\lambda_{col} = 2 \times 10^{-2}$  to balance loss terms. Following previous works [7, 14], we only activate the normal loss from the 400<sup>th</sup> epoch. The training procedure takes about 20 hours on a single RTX 3090 GPU. Given limited 3D training data, we enhance the robustness of our free-form generator to out-of-distribution poses by balancing the pose distribution. Specifically, we apply random horizontal flips along the  $x$ -axis, leveraging the symmetry of the human body.

#### A.4. Baselines

For POP [7] and FITE [5], we directly utilize the official model weight provided for inference. As for SkiRT [8], we train the model using the official code and successfully reproduce the results reported in the original paper. We perform inference using the trained model weight.

#### A.5. Details on Perceptual Study

We follow the official rendering scripts including camera and lighting configurations for baseline methods [5, 7, 8] and ours, where the output point cloud is rendered using a surfel-based renderer in Open3D [15] with a point size of 5. To assess the geometric visual quality, we render the front and the back views at a high resolution of  $1024 \times 1024$ . The deployed baseline models are discussed above in Sec. A.4.

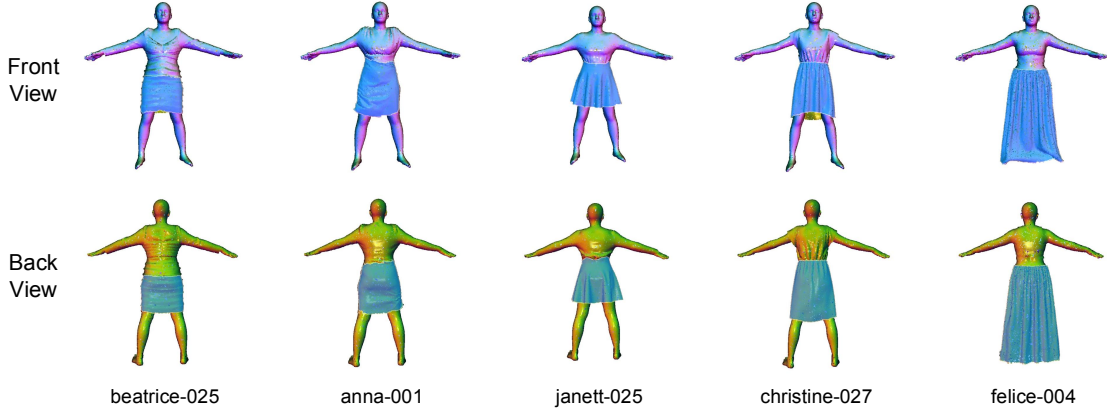


Figure A1. **The segmented loose regions of each cloth in the ReSynth [7] dataset.** We identify the loose regions in the front and back view normal maps utilizing the segmentation model SAM [4].

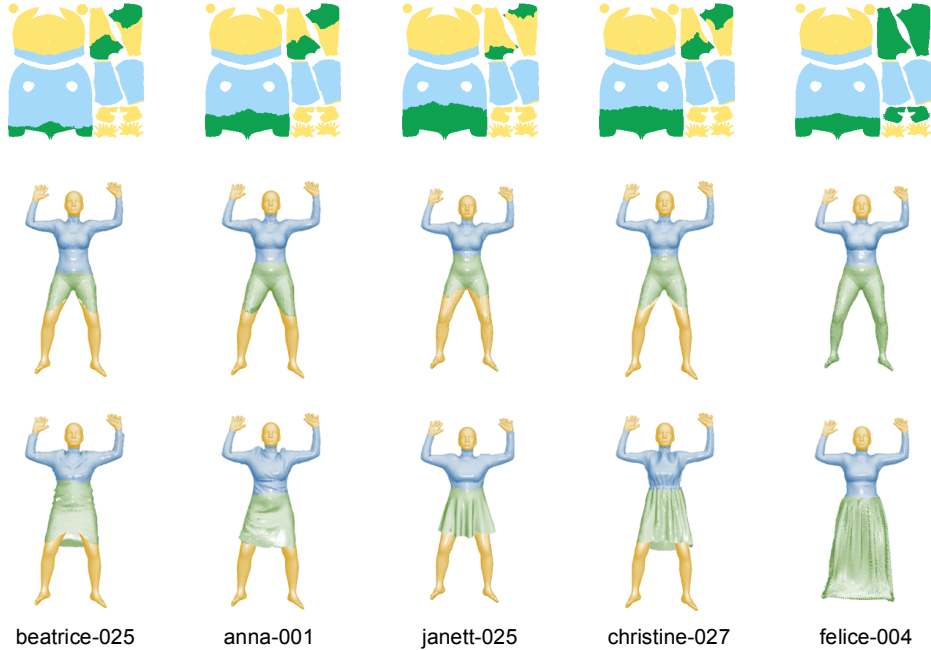


Figure A2. **The clothing-cut maps for five subjects in the ReSynth [7] dataset.** The first row depicts the clothing-cut maps distinguished by three different colors, while the second row illustrates the corresponding segmented regions. Specifically, the **yellow** color denotes the masked region, the **blue** indicates the body parts requiring deformation, and the **green** marks the loose parts to be modeled utilizing free-form generation. Finally, the last row displays the complete predictions generated by our model.

50 participants are presented with a set of 25 examples consisting of different subjects and poses, **randomly** sampled from the ReSynth [7] dataset results without cherry-picking. In each example, the GT reference is always put in the leftmost column, and we randomize the ordering of the results of different methods on the right. Fig. A3 shows an example. For each example, the participants are asked to select the most preferred single option based on the following two criteria: (1) realism, wrinkle details, smoothness, uniformity, and the presence of artifacts contribute to the overall visual quality of the clothing shape; (2) the similarity to

the reference GT result. Due to the inherent randomness in the generated results, an exact match with the reference effect may not be necessary. Therefore, priority should be given to the first point, which is the overall visual quality.

## B. Extended Results and Discussions

### B.1. Discussions on the Evaluation Metric

As pointed out by DPF [9] and FITE [5], we emphasize that conventional metrics used in the previous works [7, 8, 11], Chamfer distance (CD) and  $\mathcal{L}_1$  normal discrepancy (NML)

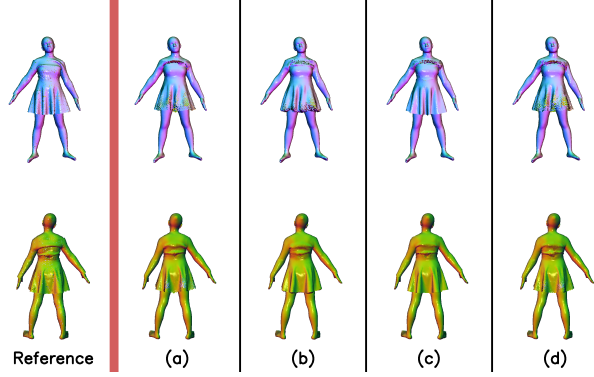


Figure A3. **Example of perceptual study image.** We randomize the ordering of the results of different methods per example. We always put the GT result in the leftmost column.

implicitly assume a one-to-one mapping from body pose to the clothing shape. However, in reality, the clothing shape possesses diversity and randomness which can be influenced by many other factors such as the motion speed and the history [9]. Consequently, given a similar or same pose, multiple clothing statuses can be reasonable, as illustrated in Fig. B4. Our model generates plausibly-looking results that, may not conform strictly to the ground truth, hence obtaining high CD errors.

To further highlight the limitations of the CD metric, we examine a case involving generated points for a long dress. When reducing the number of points generated by the free-form generator  $N_g$  from 32768 to 4096, the CD error substantially decreases (from 14.06 to 6.57, a 53.3% reduction), as shown in Fig. B5. However, this reduction comes at the cost of point density uniformity and detail, such as wrinkles. This observation has been proven in previous works that the CD metric lacks awareness of the point density distribution [2, 6, 12]. This phenomenon also helps explain why methods like POP [7] and SkIRT [8] achieve lower CD errors despite significantly lower point densities on loose skirts and dresses.

Building on the limitations discussed, we employ a generation-based metric, FID, which compares distributions and relaxes the strict one-to-one mapping constraint, alongside the reconstruction-based MSE loss to assess our model’s quality holistically. This evaluation approach better aligns with the model’s objective.

## B.2. Quantitative Results

In addition to the user study, we also employ the GPT-4o model [1] to select the best result across all methods. The testing prompt is as follows: *"Select the most preferred option based on the following two criteria: (1) realism, wrinkle details, smoothness, uniformity, and the presence of artifacts, which contribute to the overall visual quality of the clothing shape; (2) similarity to the reference ground*

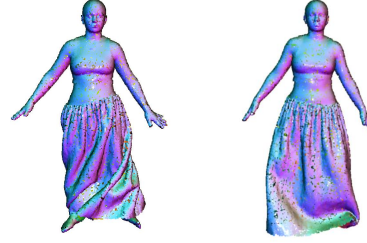


Figure B4. **An example illustrating the stochasticity of clothing shape with two similar given poses.**

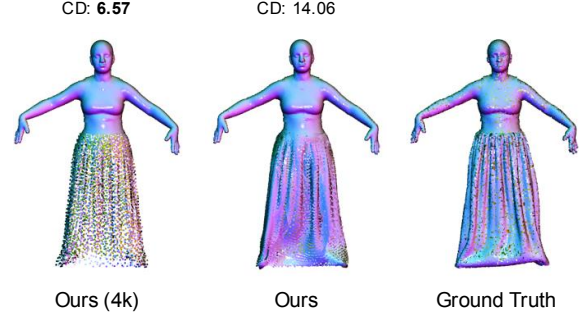


Figure B5. **Illustration of the paradox of lower CD error with worse visual quality.** Reducing the number of generated points significantly decreases the CD error, yet results in visually unsatisfactory outcomes with non-uniform point density.

truth (GT) result. Priority should be given to the first criterion, which emphasizes the overall visual quality." The per-subject preference rates of human users and GPT-4o are presented in Tab. B1, denoted as PR-H and PR-G, respectively. As shown, the results are generally consistent between the two, with our method demonstrating significant advantages in handling challenging cases. For the two most difficult skirts, the preference rates from GPT-4o reach 80%, while human users show a preference rate exceeding 85%, confirming the effectiveness of our hybrid design in modeling loose clothing. For tighter skirts, our model performs on par with FITE [5] and SkIRT [8], both of which rely purely on LBS but still generate promising results.

For reference, we also follow previous works [5, 7, 8, 14] to evaluate the Chamfer Distance (CD) and the  $\mathcal{L}_1$  normal discrepancy (NML), as specified by the formulas in the main paper. The default units for reporting CD and NML are  $\times 10^{-4}m^2$  and  $\times 10^{-1}$ , respectively. Tab. B2 presents the quantitative errors on the ReSynth [7] dataset. Notably, while FITE [5] exhibits the highest visual quality among the baselines, it also results in significantly larger quantitative errors. This further verifies that the CD metric may not accurately reflect performance, as discussed in Sec. B.1. Our model shows comparable performance to FITE in CD errors while significantly reducing the normal discrepancy, which corroborates our observation that FITE generates un-

Table B1. **Perceptual study results on the ReSynth [7] dataset for each subject.** We report the preference rates (PR-H) obtained from a perceptual study involving 50 participants, alongside the preference rates (PR-G) voted by the GPT-4o [1] model. The final scores are generally consistent with those of the human participants. The best results are highlighted in **bold**, and the second best are underlined. The subject IDs are listed in descending order based on the looseness of the clothing.

Subject ID	felice-004		janett-025		christine-027		anna-001		beatrice-025		Average	
Method	PR-H $\uparrow$	PR-G $\uparrow$	PR-H $\uparrow$	PR-G $\uparrow$	PR-H $\uparrow$	PR-G $\uparrow$	PR-H $\uparrow$	PR-G $\uparrow$	PR-H $\uparrow$	PR-G $\uparrow$	PR-H $\uparrow$	PR-G $\uparrow$
POP [7]	0.4%	0.0%	0.8%	0.0%	0.8%	0.0%	1.2%	0.0%	0.8%	0.0%	0.8%	0.0%
SkiRT [8]	3.2%	0.0%	0.0%	0.0%	3.2%	0.0%	12.8%	<u>20.0%</u>	10.4%	<b>60.0%</b>	5.9%	16.0%
FITE [5]	<u>28.0%</u>	<u>40.0%</u>	<u>6.0%</u>	<u>20.0%</u>	<u>10.8%</u>	<u>20.0%</u>	<b>54.0%</b>	<b>40.0%</b>	<b>50.8%</b>	<u>20.0%</u>	<u>29.9%</u>	<u>28.0%</u>
Ours	<b>68.4%</b>	<b>60.0%</b>	<b>93.2%</b>	<b>80.0%</b>	<b>85.2%</b>	<b>80.0%</b>	<u>32.0%</u>	<b>40.0%</b>	<u>38.0%</u>	<u>20.0%</u>	<b>63.4%</b>	<b>56.0%</b>

Table B2. **Additional quantitative comparison of different methods on the ReSynth [7] dataset for each subject.**

Subject ID	felice-004		christine-027		janett-025		anna-001		beatrice-025	
Method	CD $\downarrow$	NML $\downarrow$	CD $\downarrow$	NML $\downarrow$	CD $\downarrow$	NML $\downarrow$	CD $\downarrow$	NML $\downarrow$	CD $\downarrow$	NML $\downarrow$
POP [7]	7.34	1.24	1.72	0.97	1.24	0.89	0.62	0.82	0.34	0.75
SkiRT [8]	6.45	1.25	1.54	0.99	1.10	0.82	0.58	0.81	0.31	0.77
FITE [5]	11.27	2.38	2.16	1.15	1.52	1.05	0.74	0.91	0.46	0.85
Ours	10.61	1.78	2.18	1.01	1.59	0.94	0.81	0.84	0.48	0.74

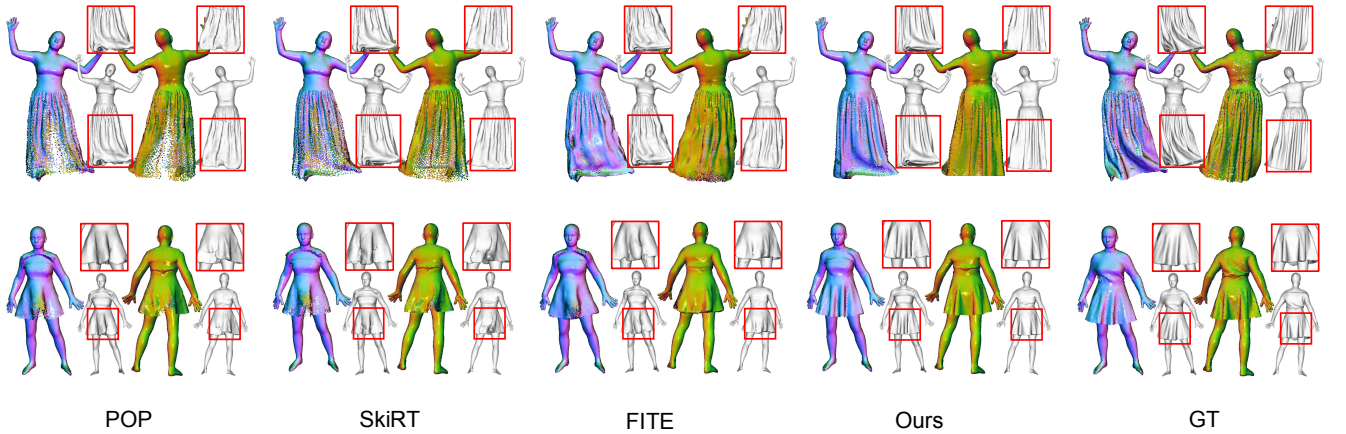


Figure B6. **Qualitative comparison between baselines and our model for modeling loose clothing, with highlighted details.** Subject IDs from top to bottom: “felice-004” and “janett-025”. Best viewed zoomed-in on a color screen.

natural and excessively bent wrinkles, whereas our model effectively captures complex local details.

### B.3. More Qualitative Results

In this section, we present additional visualization comparisons that extend the results discussed in the main paper. Fig. B6 illustrates the details of the generated loose clothing, which are highlighted within the red box. Furthermore, we showcase three testing examples for each of the five subjects from the ReSynth [7] dataset, as illustrated in Figs. B15 to B19. We recommend zooming in to observe finer details, particularly the wrinkles in skirts and dresses. Please refer to the visualization demo in our supplementary

materials, which includes sequences of testing data to better demonstrate the high-quality performance of our method.

As discussed in the main paper, the advantages of our hybrid modeling approach become particularly evident with loose skirts or dresses, as illustrated by the examples in Figs. B15 and B16. For tighter skirts, LBS-based models like FITE [5] already perform well since the clothing adheres closely to the body. As shown in Figs. B18 and B19, FITE [5] generates nearly perfect outputs that closely resemble the ground truth, and our model produces results comparable to those of FITE [5]. However, it is noteworthy that FITE [5] still fails to fully eliminate redundant points on the open surface of tighter skirts (see Fig. B7).



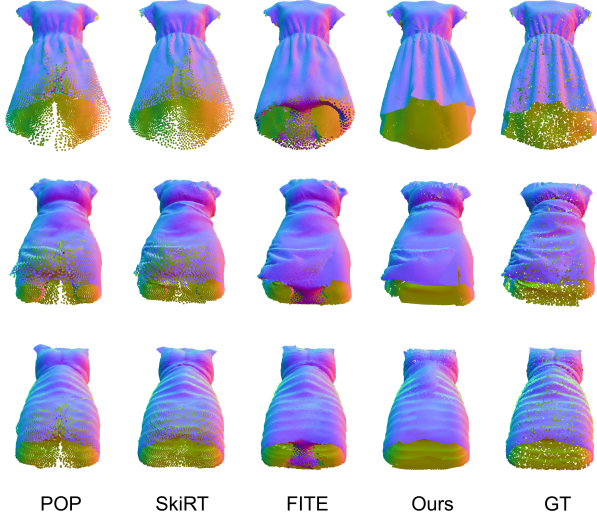


Figure B7. **Visualization results of loose clothing.** As shown, while FITE [5] successfully captures intricate details like wrinkles in tighter skirts, it still faces the “open-surface” challenge. In contrast, our model generates more accurate geometry and achieves superior visual quality.

Above all, our approach stands out in its ability to operate without subject-specific templates coupled with LBS fields, allowing for more flexible, multi-subject modeling. This opens up new possibilities for avatar modeling while maintaining high performance.

#### B.4. Multi-Subject Experiments

In this study, we explore the potential of hybrid modeling for loose clothing and significantly improve the performance under a single-subject setting. Nevertheless, our hybrid paradigm can be naturally extended to modeling multiple garments, conditioned on various global garment codes. Experimental results show that our unified, multi-subject model demonstrates promising performance in modeling various types of skirts and long dresses, confirming the expressive power of our free-form generator.

To explore the learned latent space of garment codes, we perform interpolation experiments focusing on two crucial attributes: length and tightness. As shown in Fig. B8, our model allows effective control over garment length through manipulation of the garment code. Furthermore, when varying the tightness, the generated skirts smoothly transition from tight to loose. In summary, our model successfully disentangles pose-related effects from garment-specific features, providing controllable and realistic generation results.

#### B.5. Fitting Non-skirt Clothing

Although the main focus of this paper is to investigate the hybrid modeling of loose garments such as skirts and long dresses, we also conduct experiments to handle non-skirt

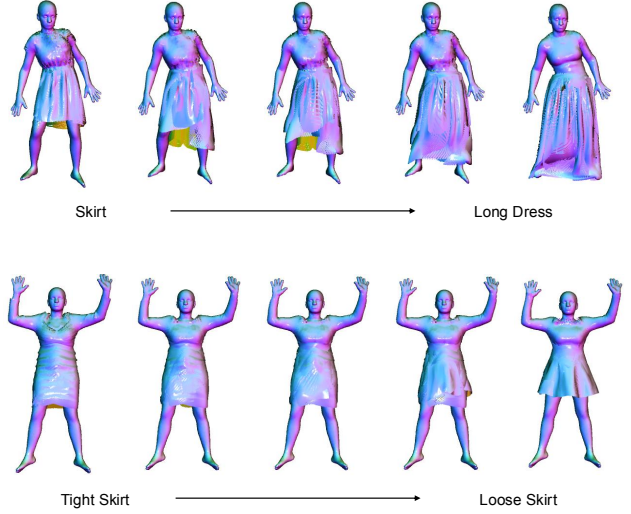


Figure B8. **Interpolation results when varying the length and the tightness of the skirt.**

clothing, *e.g.* suits. Note that the global pose feature is extracted from the PointNet++ [10] without part-aware local feature learning. Visualization results (Fig. B9) illustrate the generator’s capacity to autonomously learn and represent loose components, such as collars. This demonstrates the flexibility and the promising expressiveness of the proposed free-form generator.



Figure B9. **Our hybrid model can also handle non-skirt clothing such as suits.** As shown on the left-hand side, the free-form generation module can model loose regions such as collars.

#### B.6. More Ablation Studies

**Effects of the Hybrid Paradigm.** To evaluate the efficacy of the proposed free-form generator, we quantitatively assess the deformation-only variant using the ReSynth [7] dataset. As shown in Tab. B3, this variant achieves an average FID of 56.23 and MSE of 2.74, comparable to SKiRT [8]. In contrast, our full model substantially improves these metrics, demonstrating the effectiveness of the free-form generation module in capturing the dynamics of loose clothing. Additionally, we examine a generation-only variant that discards LBS-based deformation and synthesizes full-body clothing points from a global pose feature,

Table B3. **Ablation study of the free-form generation module on the ReSynth [7] dataset.** In the setting *Ours\**, the free-form generator is removed, relying solely on body point deformation to model loose clothing.

Subject	All		felice-004		janett-025		christine-027		anna-001		beatrice-025	
Metric	FID ↓	MSE ↓	FID ↓	MSE ↓	FID ↓	MSE ↓	FID ↓	MSE ↓	FID ↓	MSE ↓	FID ↓	MSE ↓
<b>Ours*</b>	56.23	2.73	63.12	5.72	52.10	2.06	59.29	2.41	51.68	<b>1.84</b>	54.96	<b>1.62</b>
<b>Ours</b>	<b>37.75</b>	<b>2.61</b>	<b>42.41</b>	<b>5.24</b>	<b>27.95</b>	<b>1.92</b>	<b>37.43</b>	<b>2.35</b>	<b>39.63</b>	1.89	<b>41.24</b>	1.68

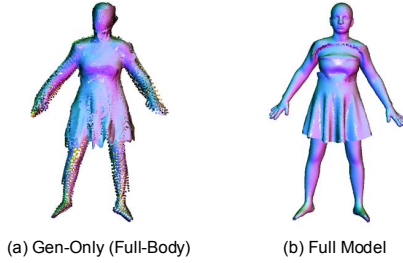


Figure B10. **Ablation study of the full-body free-form generation.** (a) Completely discarding LBS deformation results in a significant performance drop when compared to (b) our full model.

as illustrated in Fig. B10. The resulting noisy surfaces in articulated regions and overly coarse details further emphasize the necessity of the hybrid paradigm.

**Number of Patches.** We conduct ablation studies to manipulate the number of patches  $K$  utilized in the free-form generator. This experiment also serves to illustrate the inherent complexity involved in modeling loose garments. As a case study, we select the long dress, which features intricate details such as wrinkles. The results, shown in Fig. B11, are compared across different patch sizes:  $K = 2, 4, 8, 16, 32, 64$ .

Visualization (a) shows that  $K = 2$  fails to capture the intricate details of the long dress, resulting in over-smoothed outcomes. In addition, empirical findings indicate that the setting  $K = 8$  is sufficient to generate high-quality details, producing superior visual results. When  $K$  surpasses 8 or is set to 4, the generated surface exhibits increased noise, leading to a loss of clarity in fine-grained details. Notably, features such as clothing wrinkles become less distinct and sharply defined. This observation suggests that modeling the ostensibly complex long dress may be less daunting than anticipated. Furthermore, it verifies the remarkable expressiveness of our hybrid framework. Unless otherwise stated,  $K = 8$  is selected for our experiments.

To better investigate the properties of the free-form generator, we visualize  $K = 8$  patches of two loose skirts and a long dress, each rendered in different colors. As illustrated in Fig. B12, our free-form generator successfully recovers authentic fine-grained details while preserving good locality

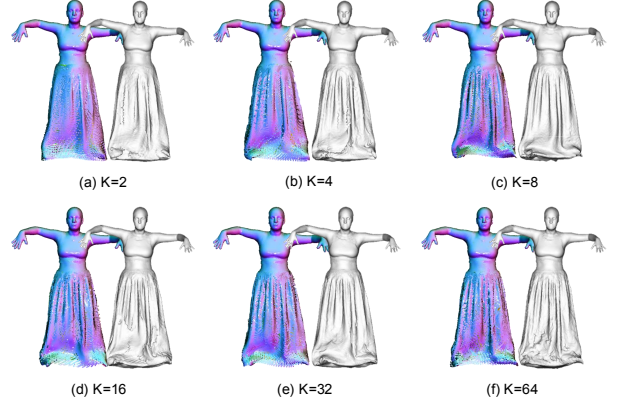


Figure B11. **Ablation studies of the number of patches  $K$  used in the free-form generator.** The visualizations reveal that opting for  $K = 2$  leads to smoothed results without details. Empirically, selecting  $K = 8$  yields the best visual outcomes. In other cases, the surface becomes progressively noisier, compromising the clarity of fine-grained elements like clothing wrinkles.

within each patch. Generally, the points within each patch are arranged in a vertical direction, and different patches seamlessly integrate to form a complete surface. These results, derived from data-driven learning, suggest that decomposing a loose garment into several “vertical” patches is a plausible approach for detailed modeling. Additionally, we visualize the convergence process of the free-form generator throughout the training phase in the demo.

**Clothing-cut Map.** As discussed in the main paper, directly employing the generation module to fill in the loose regions in the previous LBS-based framework can cause split-up artifacts. Here we present additional cases demonstrating that this issue occurs under various poses, as illustrated in Fig. B13. This issue becomes particularly evident when the underlying leg approaches the surface of the dress. In such instances, deformed points from the leg and the generated region become disjoint, causing the dress to appear torn. Additionally, a partial shape of the underlying leg can be observed through the cracks in the broken dress.

## B.7. Efficiency Analysis

We evaluate the inference speed of our approach and other SOTA methods on an RTX 3090 GPU, with a batch size of

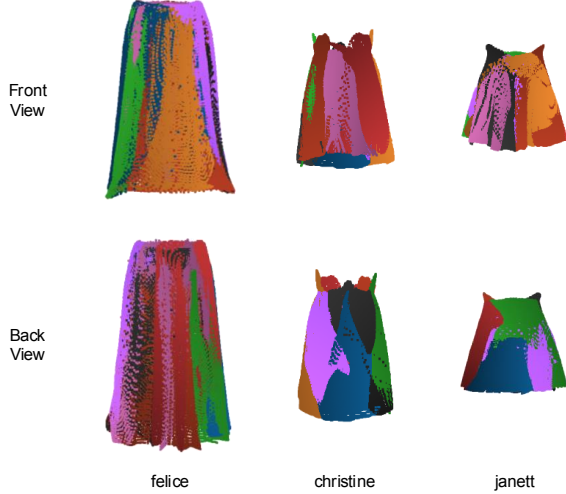


Figure B12. **Visualization of the generated  $K = 8$  patches which comprise the loose dress and skirts.** We show the results on the three subjects.

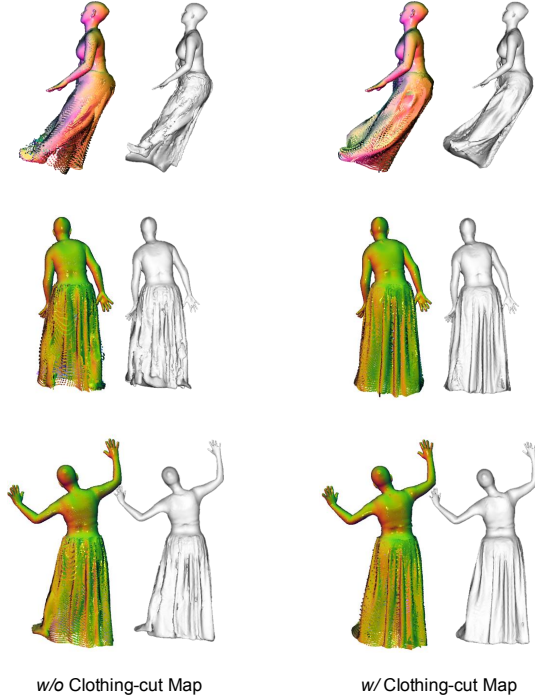


Figure B13. **More ablation study results of employing clothing-cut map.** Split-up artifacts between two modules can result in disjointed, noisy areas and torn appearances. The use of the clothing-cut map notably mitigates this issue.

1. Additionally, we report the FLOPs and parameter counts to quantify computational resource requirements. As shown in Tab. B4, our model has the smallest number of parameters and achieves real-time inference speed at 64.1 FPS. Notably, our model significantly outperforms SOTAs without introducing extra computational overhead.

Table B4. **Efficiency analysis of our method with other works.**

Method	FID ↓	FPS ↑	FLOPS (G)	Params. (M)
POP [7]	57.87	69.9	128.81	11.33
SkiRT [8]	53.32	79.9	77.12	11.13
FITE [5]	39.02	31.5	68.87	11.02
Ours	37.75	64.1	78.82	10.83

## B.8. Limitations and Failure Cases

Currently, our method focuses on single-frame modeling of clothed humans and does not consider the temporal cues that could provide constraints for clothing deformation due to motion. Consequently, discontinuities may appear in transitions between frames. Future work could explore incorporating temporal information to achieve smoother and more realistic modeling results.

Despite employing pose augmentation, our model remains susceptible to failure when confronted with extremely challenging poses, resulting in clothing penetration artifacts, as depicted in Fig. B14 (a). This issue is particularly noticeable when the skirt becomes tighter. Training our free-form generator on a larger dataset could enhance its robustness to out-of-distribution poses and reduce such artifacts. Additionally, while our pipeline employs different strategies to handle deformed and generated areas, we cannot guarantee the perfect blending of point clouds from two modules. As illustrated in Fig. B14 (b), “seams” at the boundary regions are occasionally observed.

However, it is crucial to note that our experiments underscore the promise and versatility of our proposed hybrid approach. By transcending the limitations imposed by relying solely on LBS-based deformation, our method demonstrates notable expressive capabilities. We believe that with larger datasets, our approach has considerable potential for superior performance in future applications.

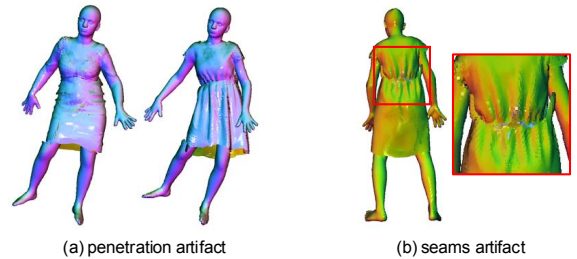


Figure B14. **Two typical failure modes.** (a) In challenging poses, the generated skirt or dress occasionally collides with the human body. (b) At the boundaries between deformed and generated regions, our model sometimes produces discontinuous “seams”.





Figure B15. **Additional Qualitative comparison between baselines and our model.** The subject ID is “felice-004” from the ReSynth [7] dataset. Best viewed zoomed-in on a color screen.





Figure B16. **Additional Qualitative comparison between baselines and our model.** The subject ID is “janett-025” from the ReSynth [7] dataset. Best viewed zoomed-in on a color screen.



Figure B17. **Additional Qualitative comparison between baselines and our model.** The subject ID is “christine-027” from the ReSynth [7] dataset. Best viewed zoomed-in on a color screen.



Figure B18. **Additional Qualitative comparison between baselines and our model.** The subject ID is “anna-001” from the ReSynth [7] dataset. Best viewed zoomed-in on a color screen.



Figure B19. **Additional Qualitative comparison between baselines and our model.** The subject ID is “beatrice-025” from the ReSynth [7] dataset. Best viewed zoomed-in on a color screen.



## References

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023. [3](#), [4](#)
- [2] Tianxin Huang, Xuemeng Yang, Jiangning Zhang, Jinhao Cui, Hao Zou, Jun Chen, Xiangrui Zhao, and Yong Liu. Learning to train a point cloud reconstruction network without matching. In *European Conference on Computer Vision*, pages 179–194. Springer, 2022. [3](#)
- [3] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. [1](#)
- [4] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023. [1](#), [2](#)
- [5] Siyou Lin, Hongwen Zhang, Zerong Zheng, Ruizhi Shao, and Yebin Liu. Learning implicit templates for point-based clothed human modeling. In *European Conference on Computer Vision*, pages 210–228. Springer, 2022. [1](#), [2](#), [3](#), [4](#), [5](#), [7](#)
- [6] Minghua Liu, Lu Sheng, Sheng Yang, Jing Shao, and Shi-Min Hu. Morphing and sampling network for dense point cloud completion. In *Proceedings of the AAAI conference on artificial intelligence*, pages 11596–11603, 2020. [3](#)
- [7] Qianli Ma, Jinlong Yang, Siyu Tang, and Michael J Black. The power of points for modeling humans in clothing. In *International Conference on Computer Vision (ICCV)*, pages 10974–10984, 2021. [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#), [8](#), [9](#), [10](#), [11](#), [12](#)
- [8] Qianli Ma, Jinlong Yang, Michael J Black, and Siyu Tang. Neural point-based shape modeling of humans in challenging clothing. In *2022 International Conference on 3D Vision (3DV)*, pages 679–689. IEEE, 2022. [1](#), [2](#), [3](#), [4](#), [5](#), [7](#)
- [9] Sergey Prokudin, Qianli Ma, Maxime Raafat, Julien Valentin, and Siyu Tang. Dynamic point fields. *arXiv preprint arXiv:2304.02626*, 2023. [2](#), [3](#)
- [10] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017. [1](#), [5](#)
- [11] Shunsuke Saito, Jinlong Yang, Qianli Ma, and Michael J Black. Scanimate: Weakly supervised learning of skinned clothed avatar networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2886–2897, 2021. [2](#)
- [12] Tong Wu, Liang Pan, Junzhe Zhang, Tai Wang, Ziwei Liu, and Dahua Lin. Density-aware chamfer distance as a comprehensive metric for point cloud completion. *arXiv preprint arXiv:2111.12702*, 2021. [3](#)
- [13] Chulin Xie, Chuxin Wang, Bo Zhang, Hao Yang, Dong Chen, and Fang Wen. Style-based point generator with adversarial rendering for point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4619–4628, 2021. [1](#)
- [14] Hongwen Zhang, Siyou Lin, Ruizhi Shao, Yuxiang Zhang, Zerong Zheng, Han Huang, Yandong Guo, and Yebin Liu. Closet: Modeling clothed humans on continuous surface with explicit template decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 501–511, 2023. [1](#), [3](#)
- [15] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3d: A modern library for 3d data processing. *arXiv preprint arXiv:1801.09847*, 2018. [1](#)