Supplementary Material:
DeepDraper: Fast and Accurate 3D Garment Draping over a 3D Human Body

Lokender Tiwari   Brojeshwar Bhowmick

TCS Research, India

lokender.tiwari@tcs.com

In this supplementary material, we present the additional insights, results of our DeepDraper methods and additional comparisons against Tailornet [3]. In particular, the results consists of the following:

- Intermediate multi-view rendering results as training progress.

- Visual comparison results with TailorNet [3] using unseen body shape, pose, and garment style.

- Visual comparison with Tailornet using AMASS [2] dataset.

- Visual results of dressing human from a Youtube video using our DeepDraper method.

- Front-views and back-views visual results of our DeepDraper method with varying body height and weight.

- Explanation of multi-view rendering setup with camera and light configuration.

- We also provide the animated results in the supplementary video.

S1. Intermediate rendering results as training progress

We show front and back view rendering of normal mapped skinned t-shirt in Fig. S1 as training progress. The first row in the Fig. S1 shows the ground-truth vertices and normals of a t-shirt. The intermediate results of our predicted garment and normals during the training is shown from the row 2 to row 4 of Fig. S1. We observe that as training finishes the DeepDraper is able to retain most of the wrinkles and the folds.

S2. Visual comparison results with TailorNet [3] using unseen body shape, pose, and garment style

As discussed in the introduction of the paper (Fig. 2) and experimental result section of the paper (Fig. 6), Tailornet [3] fails to generalize beyond the range of body shape and garment style parameters used in their training. We provide additional results here with more unseen garment styles (not seen during training) and varying body poses. Fig. S2 shows the results of a fixed unseen style (style parameter [2.75, 2.75,0,0]) drapped over different body poses. It is evident that while Tailornet [3] fails to produce realistic wrinkles and folds according to the pose, our DeepDraper method renders visually plausible wrinkles and folds. We further experiment on different unseen styles of the garments keeping the body pose fixed. Fig. S3 shows the comparisons with different style parameters mentioned along with the figures. DeepDraper produces significantly better results here compared to the Tailornet.

S3. Visual comparison with Tailornet using AMASS [2] dataset

We use the AMASS [2] dataset to evaluate the generalization of our method on unseen poses and shapes which is never seen during the training. Fig. S4 shows that our method produces visually plausible rendering of the t-shirt deformations according to the poses and shapes. Also, it can be seen that our rendering is more realistic compared to the Tailornet [3] (see the red boxes for the comparisons).

S4. Visual results of dressing human from a Youtube video using our DeepDraper method

We further study the efficacy of our method using the poses and shapes of human obtained from a video. Any virtual try-on application requires such a pipeline to drape a t-shirt over the body parameters computed from a video. To

this end, we select a video from Youtube and use the method VIBE [1] to compute the body pose and shape from an image to SMPL. In addition to the results shown in Fig. 1 and Fig. 8 in the paper, here we show the result on a different pose. Fig. S5 shows the output of our DeepDraper method. It is evident from the Fig. S5 that our method produces visually plausible rendering compared to the Tailornet (see the red boxes for the bulges). We present the output of our DeepDraper method on this sequence in our supplementary video.

S5. Front-views and back-views visual results of our DeepDraper method with varying body height and overall fatness

We evaluate our method on different shape and size of the body keeping the pose fixed. While Fig. 7 in the paper shows the output on this experiment for the front-views, here in Fig. S6 we show both front-views and back-views of the t-shirt rendering. This also shows that our method produces good fitment of garment over different body heights and fatness.

S6. Explanation of multi-view rendering setup with camera and light configuration

The multi-view rendering setup we use in our work is shown in Fig. S7. We normalize the garment vertices by subtracting the mean before rendering. In Fig. S7, for visualization purpose, we have shown draped t-shirt. However, for computing the loss, we only render the garments. All the lights are at distance of 3 units from the garment center and facing towards the garment center, specifically we place the lights at locations $[0.0, 0.0, 3.0], [3.0, 0.0, 0.0], [0.0, 0.0, -3.0]$, $[-3.0, 0.0, 0.0], [0.0, 3.0, 0.0]$. We have tested our method with different light configurations, and there could be multiple such configurations that will give the same results. Similarly we studied the effect of different camera locations. We have fixed the camera distance to be 0.8 units from the garment center for all the views.

S7. Supplementary video

We show the animated effect on wrinkles and folds as a function human body height and overall body fatness. We also show animated draping results of t-shirt and pants using our method on a sample Youtube video and unseen pose sequence from AMASS dataset.

References

[1] Muhammed Kocabas, Nikos Athanasiou, and Michael J Black. Vibe: Video inference for human body pose and shape estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5253–5263, 2020. 2

[2] Naureen Mahmood, Nima Ghorbani, Nikolaus F Troje, Gerard Pons-Moll, and Michael J Black. Amass: Archive of motion capture as surface shapes. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5442–5451, 2019. 1

[3] Chaitanya Patel, Zhouyingcheng Liao, and Gerard Pons-Moll. Tailornet: Predicting clothing in 3d as a function of human pose, shape and garment style. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7365–7375, 2020. 1

Ground-Truth (front) Ground-Truth (back)

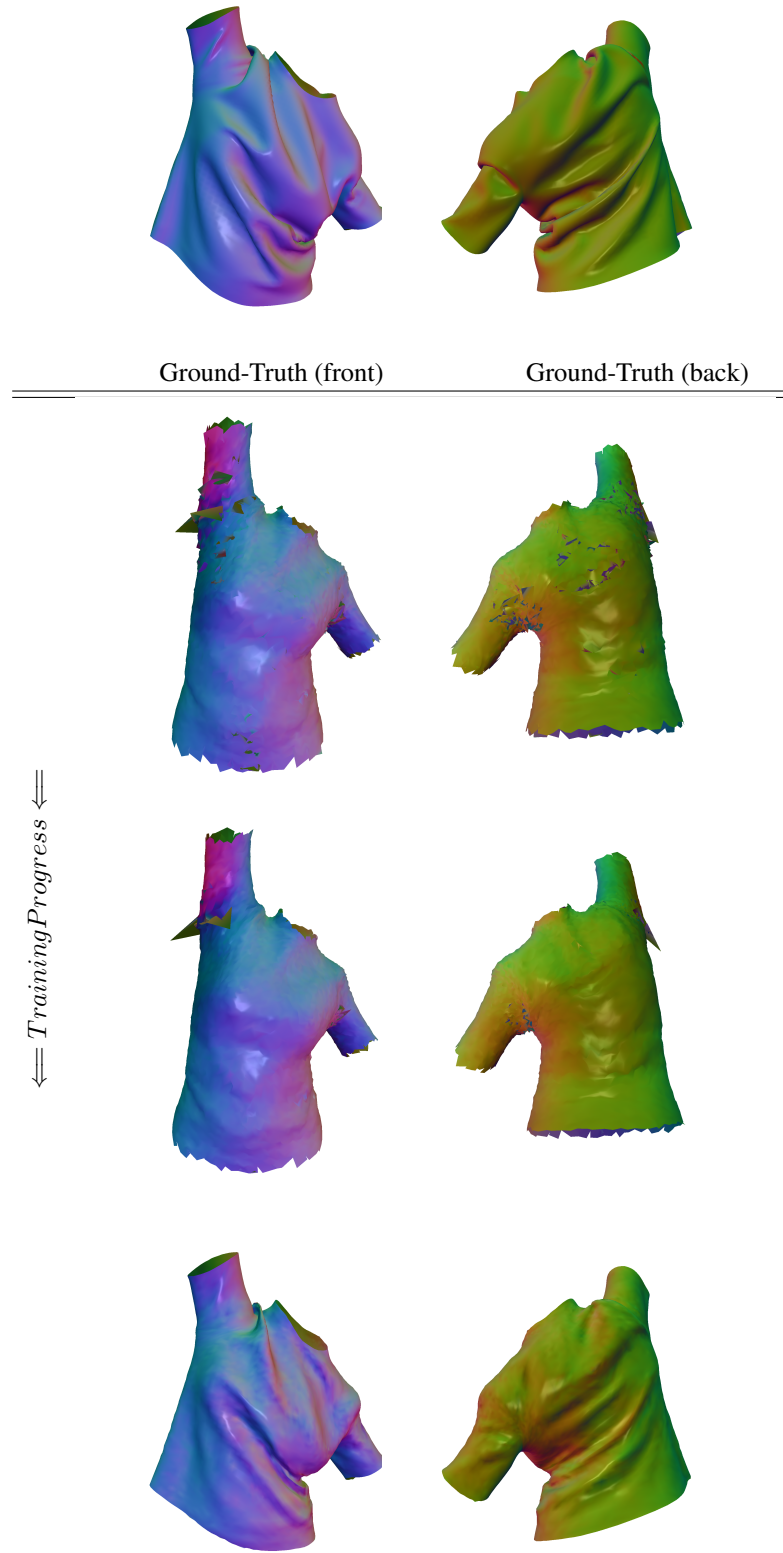$\Longleftarrow Training Progress \Longleftarrow$

Figure S1: Visual Results as Training Progress. First row shows the ground-truth normal map of front and back of the t-shirt. Rows 2-4 show the intermediate visual results of our training.
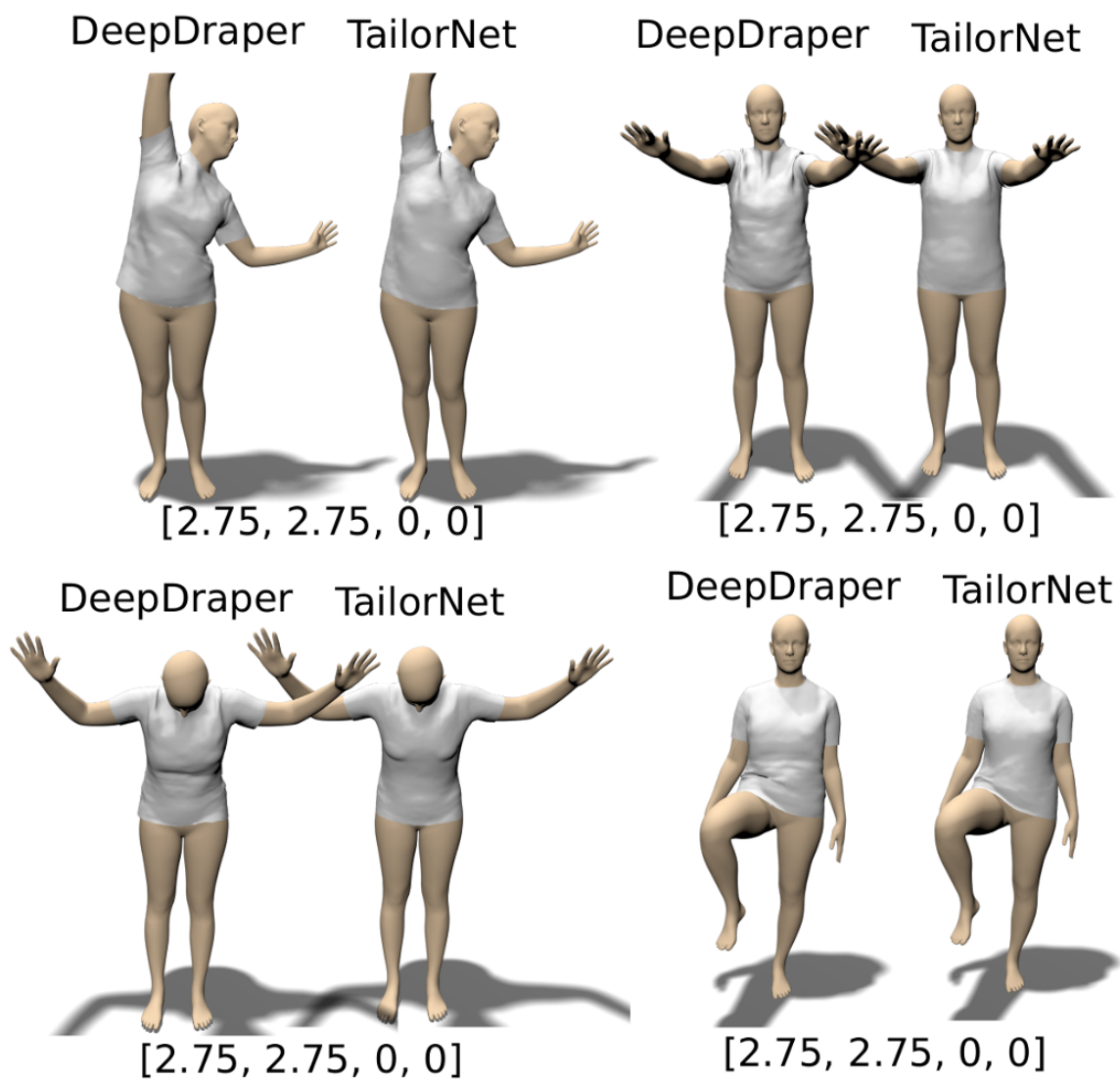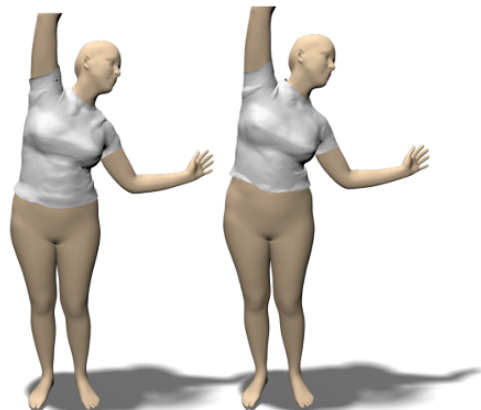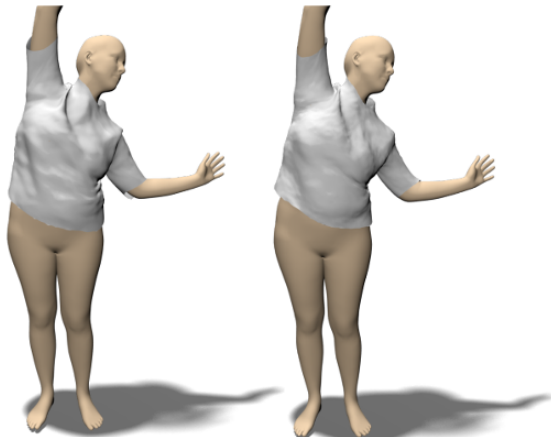
DeepDraper TailorNet

DeepDraper TailorNet

[2.75, 2.75, 0, 0]

[2.75, 2.75, 0, 0]

DeepDraper TailorNet

DeepDraper TailorNet

[2.75, 2.75, 0, 0]

[2.75, 2.75, 0, 0]

Figure S2: Qualitative Comparison of DeepDraper with TailorNet on varying varying body pose with a fixed unseen garment style (the style parameter is [2.75, 2.75,0,0]).
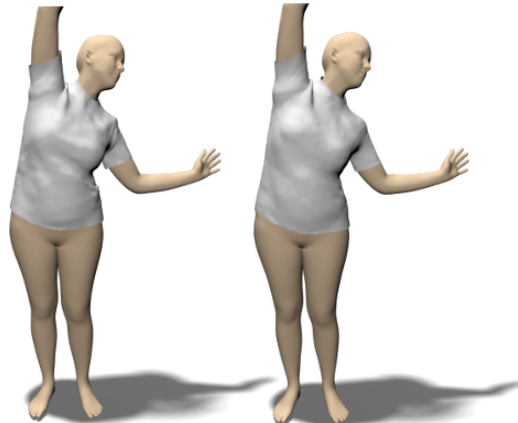
Figure S3: Qualitative Comparison of DeepDraper with TailorNet on varying unseen garment styles with fixed body pose. The garment style parameters are shown along with each of these figures.

Front   Back

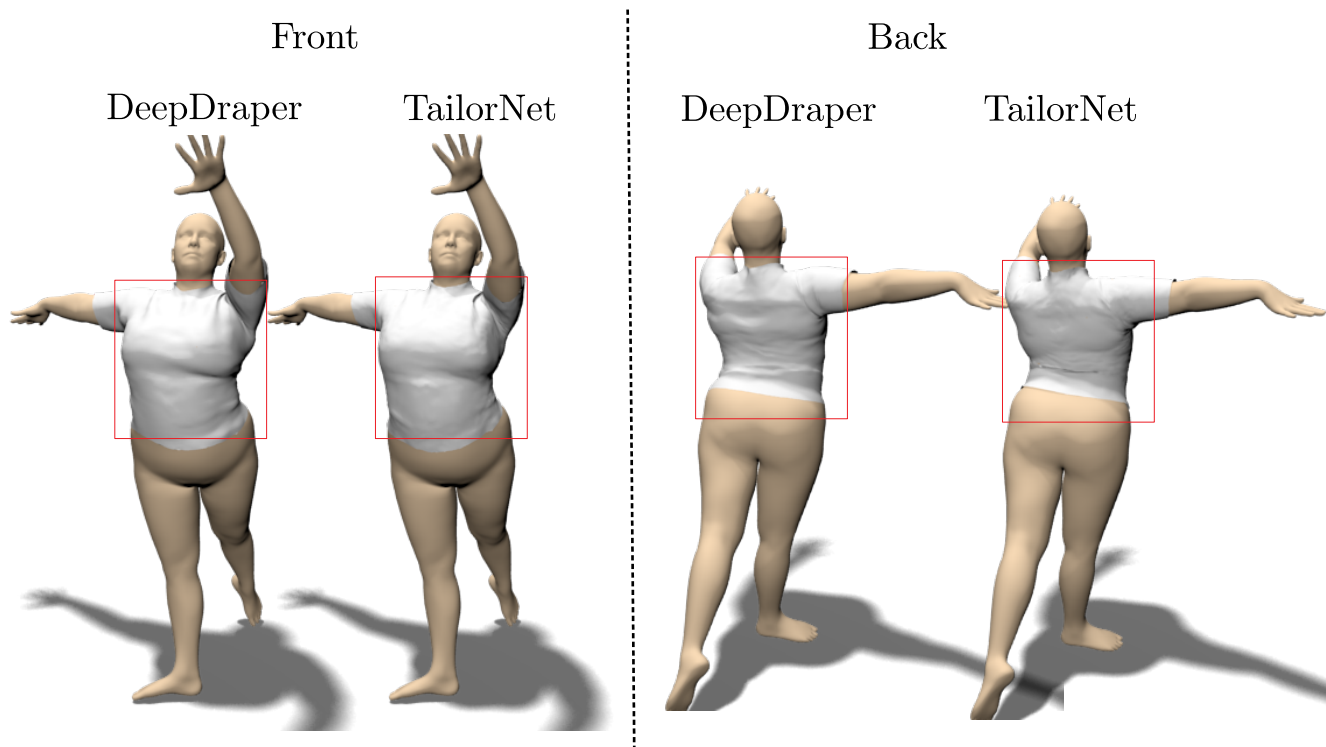DeepDraper   TailorNet   DeepDraper   TailorNet

Figure S4: Sample qualitative results on pose sequence of AMASS dataset. We highlight the differences using a red bounding box. Our method predicts accurate fitment with folds and wrinkles.

SMPL 3D   T-Shirt Style 1   T-Shirt Style 2
Human Body   DeepDraper   TailorNet   DeepDraper   TailorNet
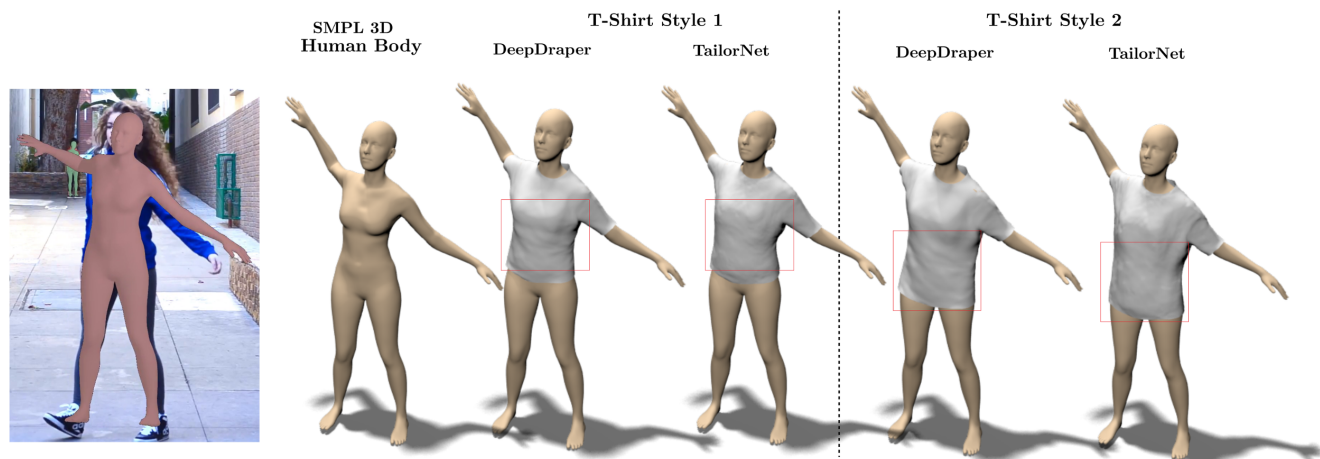
Figure S5: Sample qualitative results on a Youtube video. We highlight the differences using a red bounding box. TailorNet predicts a bulge around the stomach, while our method predicts accurate fold and wrinkles.
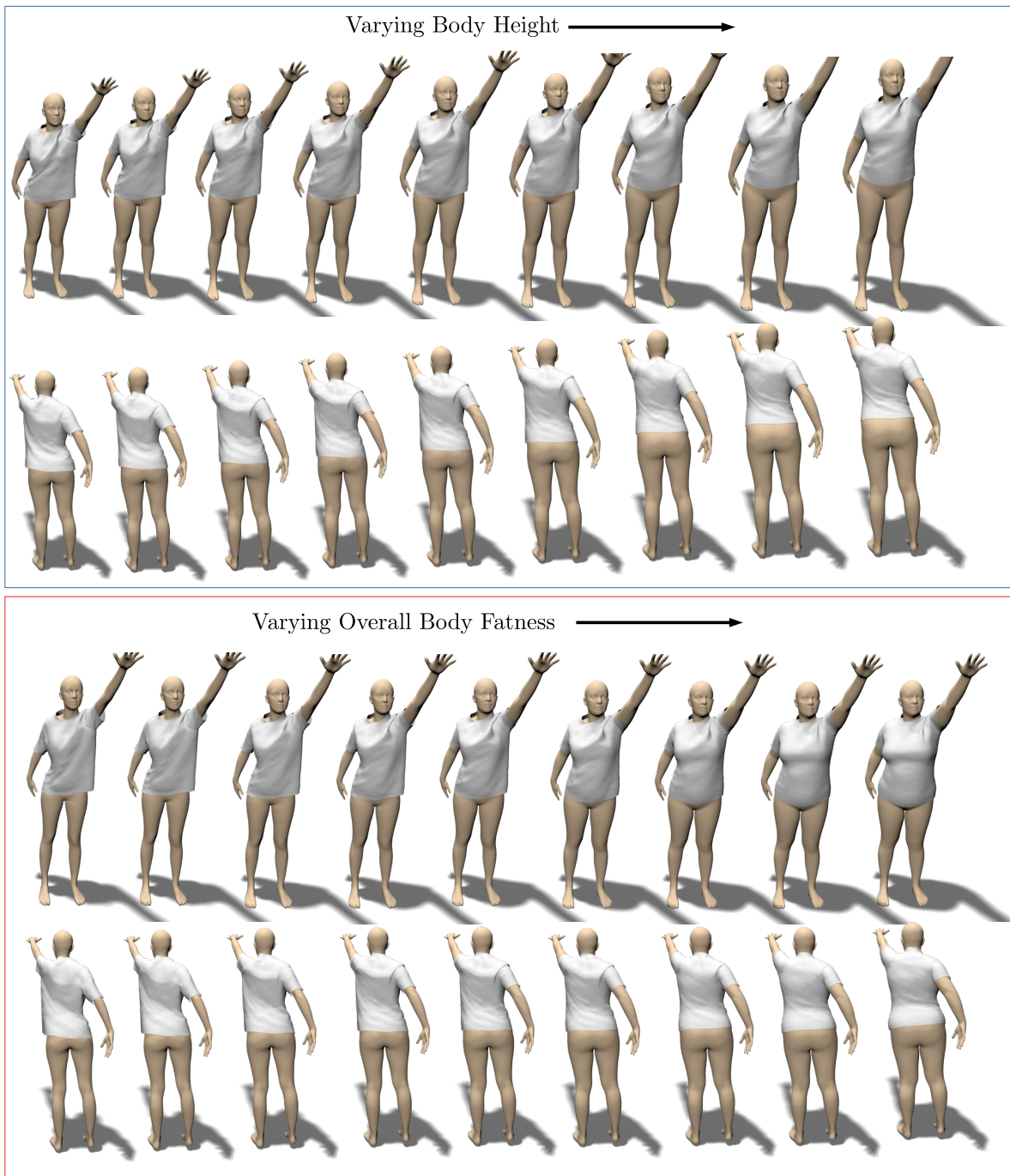
Figure S6: DeepDraper's sample qualitative results showing the effect on wrinkles and folds and overall fitment as the human body varies in height and fatness.
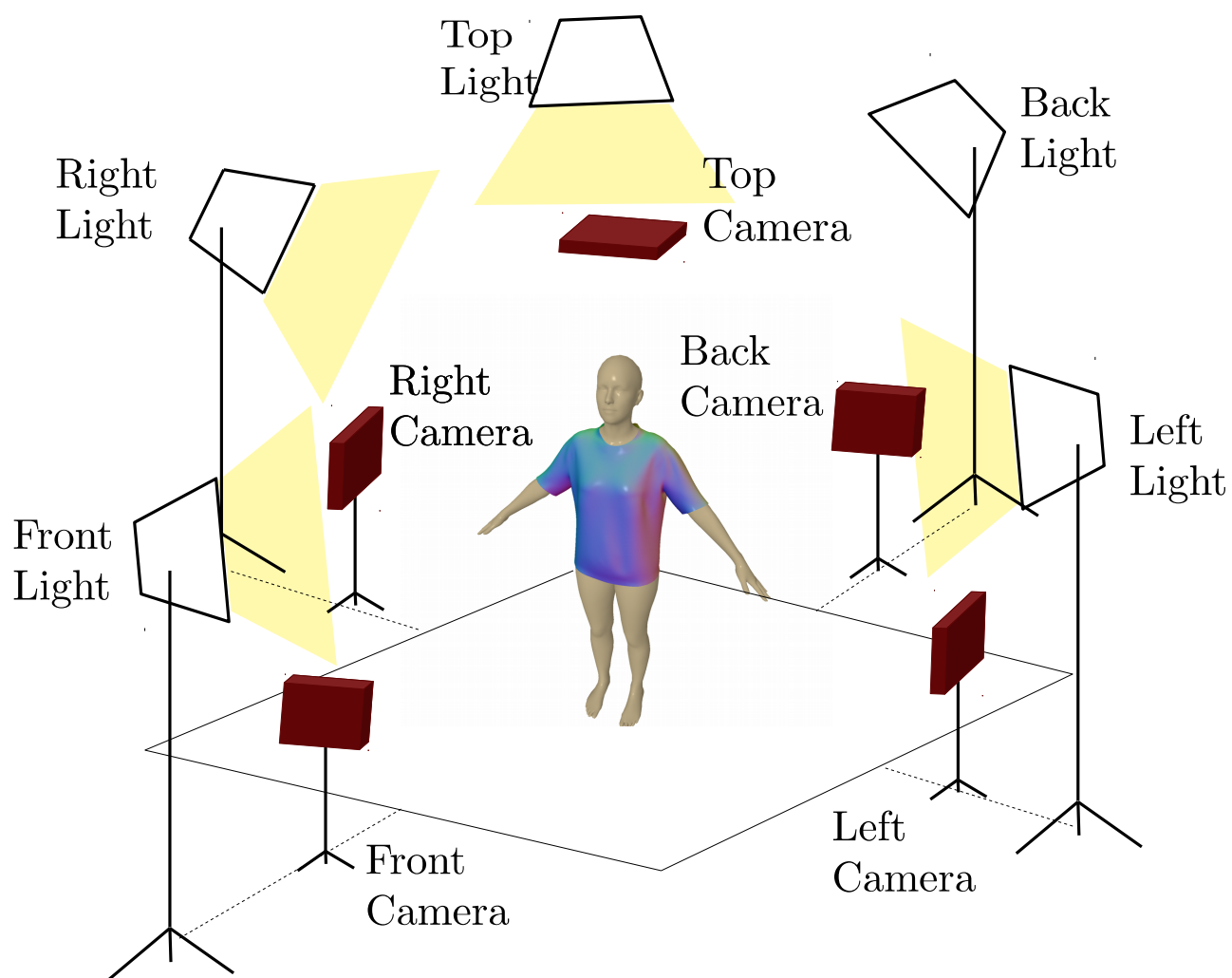
Figure S7: Mutli-view rendering setup.