GarNet: A Two-stream Network for Fast and Accurate 3D Cloth Draping
Supplementary Material

In this document, we provide details about our dataset and our network architectures, as well as qualitative results on the dataset of [8]. We also detail our condition in the penetration term of our loss function.

A. Creating our Dataset

Using the same notation as in the paper, our PBS ground truth was obtained as follows: We use the SMPL framework [4] to generate the body meshes. We also use the SMPL skeleton information and Blender to rig the mesh. For the garment mesh, we rig it with the same skeleton as that of the body and compute the rigging weights using Blender.

Given the template garment mesh $\mathcal{M}^0$, and the body mesh $\mathcal{B}$, we obtain the ground-truth draping result $\mathcal{G}^G$ of $\mathcal{M}^0$ onto $\mathcal{B}$ via physics simulation. The physics engine is based on NvCloth [5] and incorporates a number of physical properties and forces acting on the garment, such as stretching, bending, gravity and dragging, while preventing garment-body interpenetration. Each garment type has a different set of values for these parameters depending on their material properties.

As $\mathcal{M}^0$ is usually modeled in the A-pose, shown in Fig. 1, while $\mathcal{B}$ can have an arbitrary pose, we first warp $\mathcal{B}$ to the same pose as $\mathcal{M}^0$. This pre-processing step is necessary for a good initialization of the physics engine. The warping process is achieved via Dual Quaternion Skinning (DQS) [3], which requires the skeleton and the blending weights of the body meshes. We automatically compute the skeletons and blending weights for each body mesh $\mathcal{B}$ and template garment mesh $\mathcal{M}^0$ using the Pinocchio Auto Rig framework [1].

After warping $\mathcal{B}$ to match the pose of $\mathcal{M}^0$, we gradually unwarp $\mathcal{B}$ to the original body pose. The garment is deformed accordingly during this process. Depending on the difference between the poses, this process can take from 100 to 200 iterations. To speed up the simulation, we use the GPU solver of NvCloth, which can only handle 500 triangles. Therefore, we first simulate the garment on a decimated 500-facet version of the body mesh on the GPU and then improve the draping results at full mesh resolution on the CPU. It should also be noted that the topology of the

Figure 1: Template garment $\mathcal{M}^0$ in the A-pose.

$\mathcal{M}^0$, $\mathcal{M}$ and $\mathcal{G}^G$ are the same, making it possible to use point-wise operations and convolutions, such as 1×1 convolution, mesh convolution and shared MLP layers.

For the training, validation and test sets, each garment type has 500, 20 and 80 bodies, respectively. We used the CMU mocap [2] motions, and sampled a total of 40, 31 and 23 poses for the T-shirt, the jeans and the sweater, respectively. Figs. 2, 3 and 4 depict example PBS results of the T-shirt, jeans and sweater datasets, respectively, on different body shapes and poses.

B. Network Architectures

Here, we give additional details about the network architectures, the Spatial Transformer Network (STN) and mesh convolution block used in the main paper. In all of the figures, grey colored boxes are shared MLP or 1×1 convolution layers [7], where the input and output tensors are shown on the left and right, respectively. The consecutive tensors are concatenated to act as input to the layer on their right. The numbers between parentheses specify the output feature dimensionality of the layers. Arrows indicate either inputs, outputs or skip connections. We use the mesh convolution operator of [7] to construct our mesh convolution subnetwork as shown in Fig. 9b. This subnetwork is employed in the **GarNet-Global** and **GarNet-Local** architectures. Fig. 9a depicts the Spatial Transformer Network employed in all of the architectures; and Figs. 5, 6 and 7 show the **GarNet-Local**, **GarNet-Global** and **GarNet-Naive** network architectures, respectively.

1

Figure 2: PBS samples of the T-shirt from our dataset.

C. Qualitative Results on the Dataset of [8]

In this section, we qualitatively show our results on the dataset of [8], which includes 400 test samples in a single pose. To adapt our framework to this dataset, we modified **GarNet-Local** as shown in Fig. 8. As can be seen in Figs. 10, 11 and 12, our adapted framework yields results that are close to the overall PBS shapes for different garment sewing patterns, and it can produce details with low frequency. Notably, the garment sewing patterns in the test set do not exist in the training set. This shows that the network can generalize well for changing sewing patterns, such as the thickness and length of the arm of the garment.

D. Interpenetration

We now provide some more detail regarding our interpenetration loss term in Eq.(4) of the paper. Fig. 13a shows a case where a triangle-to-triangle intersection occurs be-

tween the body and the predicted garment mesh. For the illustration to be clear, we show this case in 2D, where the line between G_1^P and G_2^P is the projection of a facet in 2D. Although the vertices G_1^P and G_2^P are individually out of the body, there exists a triangle-to-triangle intersection between the body facets and garment facet that contains G_1^P, G_2^P and the third vertex not visible in this 2D projection. However, our loss term for interpenetration will penalize this case since G_1^P and G_2^P are not in the half-space S_i of B_i, provided that the garment vertices belonging to the intersecting facet are matched to the same body vertex. If the garment vertices of the intersecting facet are matched to different body vertices, then the loss term might not penalize this case, as shown in Fig. 13b, where G_1^P is matched to B_k rather than B_i. The situation in Fig. 13b is unlikely in our case since we use higher resolution garment meshes than that of the body. Hence, a third garment vertex G_3^P, which is matched to B_i, is highly likely to intervene between G_1^P

Figure 3: PBS samples of the jeans from our dataset.

and G_2^P, resulting in two facets and preventing the triangle-to-triangle intersection, as described in Fig. 13c.

In our extensive experiments with many body poses and shapes, we have not encountered a severe interpenetration, such as the one described above. This is because: (i) the provided rare cases, such as the one in Fig. 13b, tend to occur when the body surface is not smooth, and we do not have high-curvature body surfaces in the regions of garment-body interaction; (ii) the garment mesh resolution is higher than that of the body in all of our datasets, i.e., we have 10K garment vertices and 6K body ones; hence, each body vertex in the upper (for T-shirt and sweater) or lower (for jeans) part of the body matches to 3 garment vertices on average; (iii) we incrementally inflate the body mesh to ensure that these cases are well penalized; (iv) most extreme cases which would not happen in the PBS are also penalized by the terms such as L_{vert} and L_{bend} in Eq.(2) of the paper.
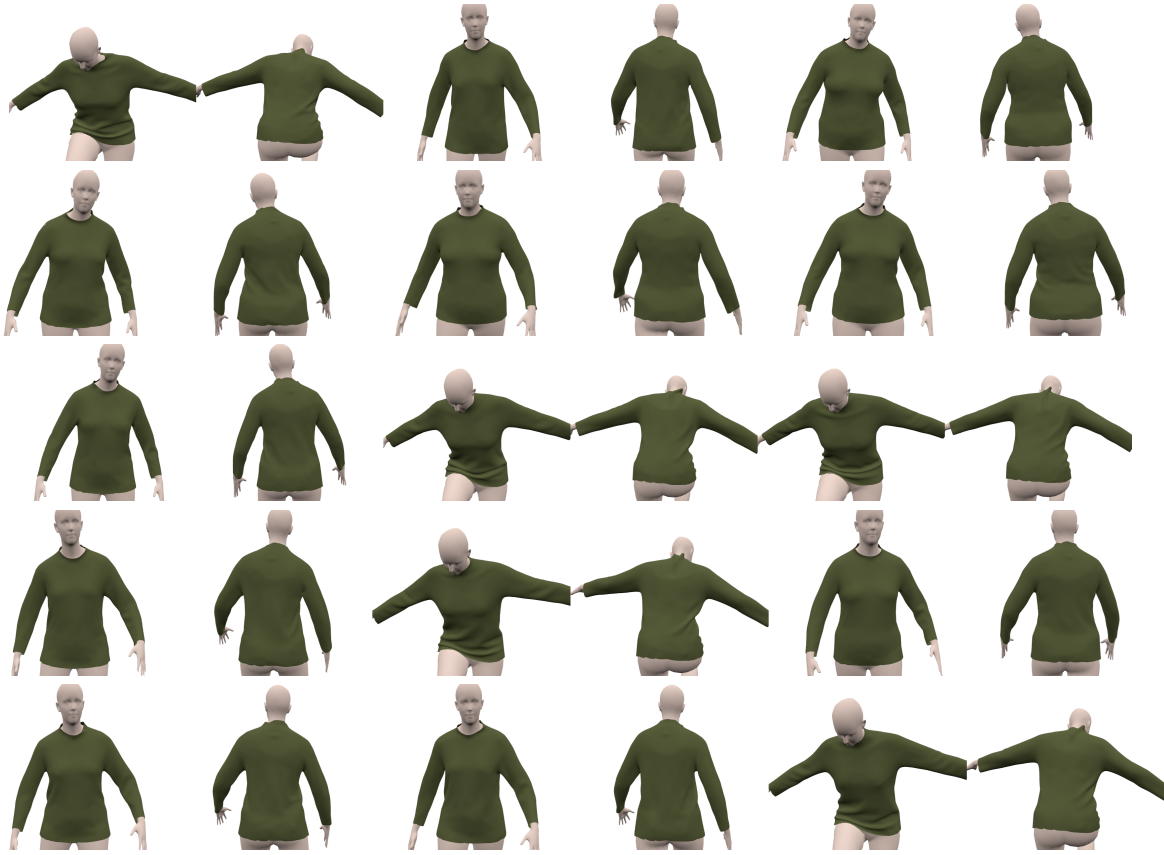
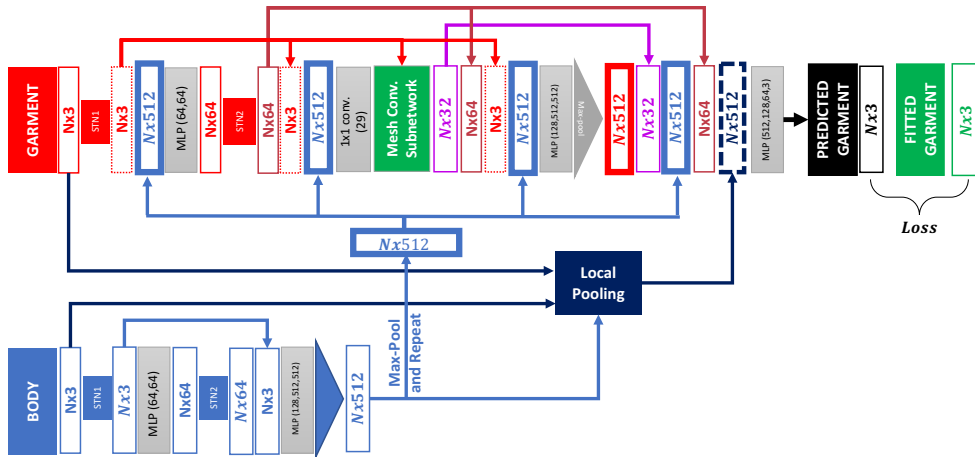Figure 4: PBS samples of the sweater from our dataset.

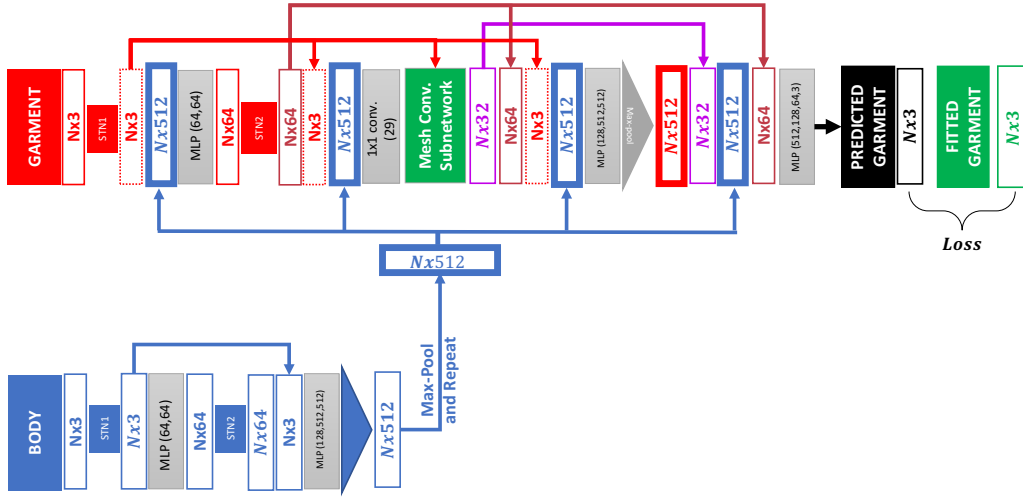Figure 5: Our full **GarNet-Local** network architecture.

Figure 6: Our architecture without the local max-pooling of body features in the middle stream of Fig. 5. This architecture is referred to as **GarNet-Global** in the main paper.
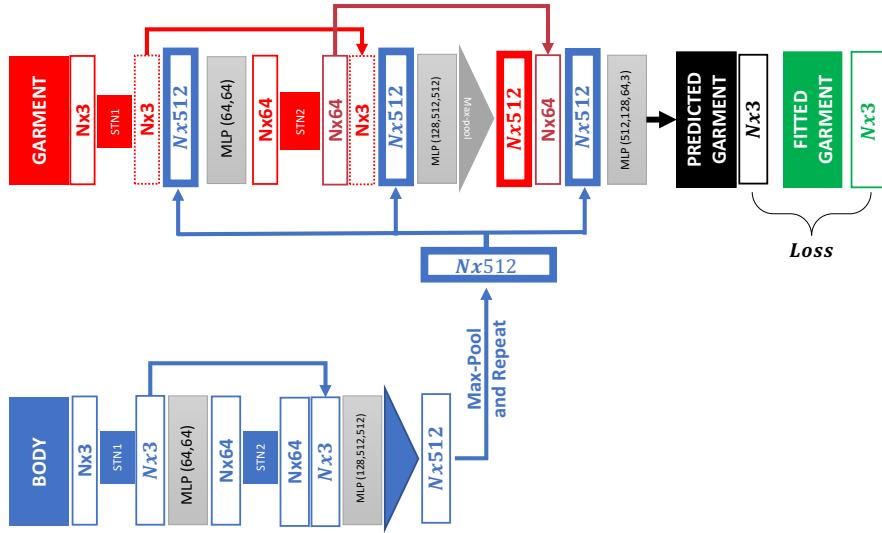

Figure 7: Our architecture without the mesh convolution layers and the local max-pooling of body features in the middle stream of Fig. 5. This architecture is referred to as **GarNet-Naive** in the main paper.
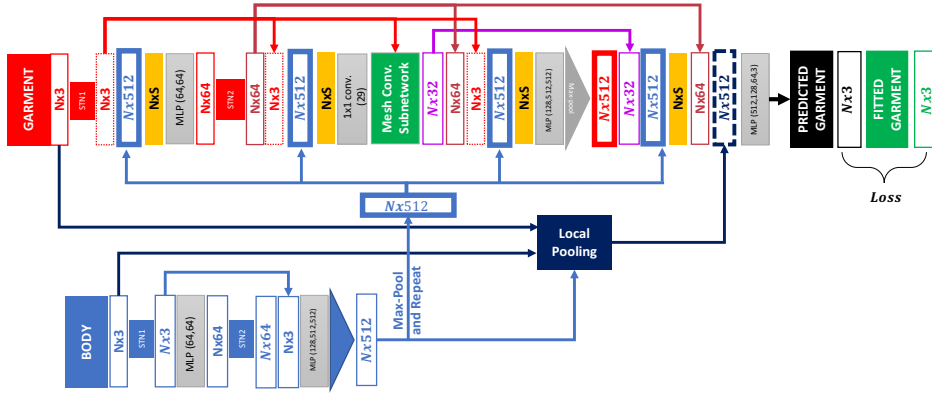
Figure 8: Our modified network architecture for the publicly available dataset of [8]. We repeat the S garment sewing parameters ($N \times S$ yellow box) for each garment vertex and further concatenate them in other layers.
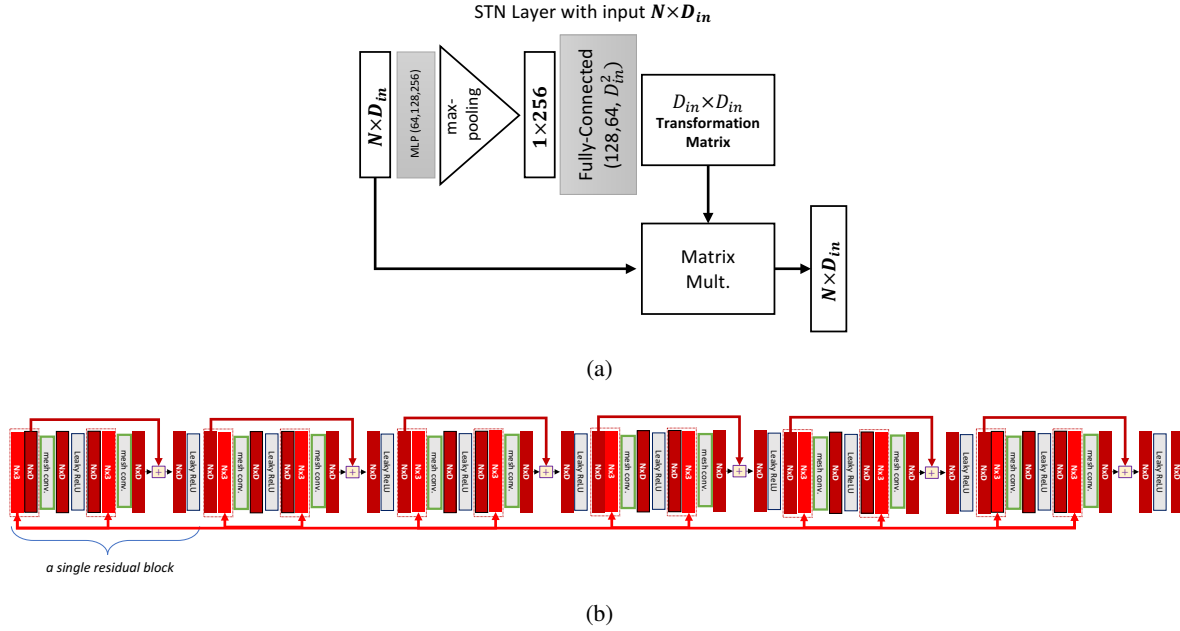

(a)

(b)

Figure 9: (a) Spatial Transformer Network (STN) portion of all the models used in the paper. It is similar to the architecture used in [6]. For STN1 and STN2, D_{in} is set to 3 and 64, respectively. (b) Mesh convolution subnetwork of the **GarNet-Global** and **GarNet-Local** architectures.

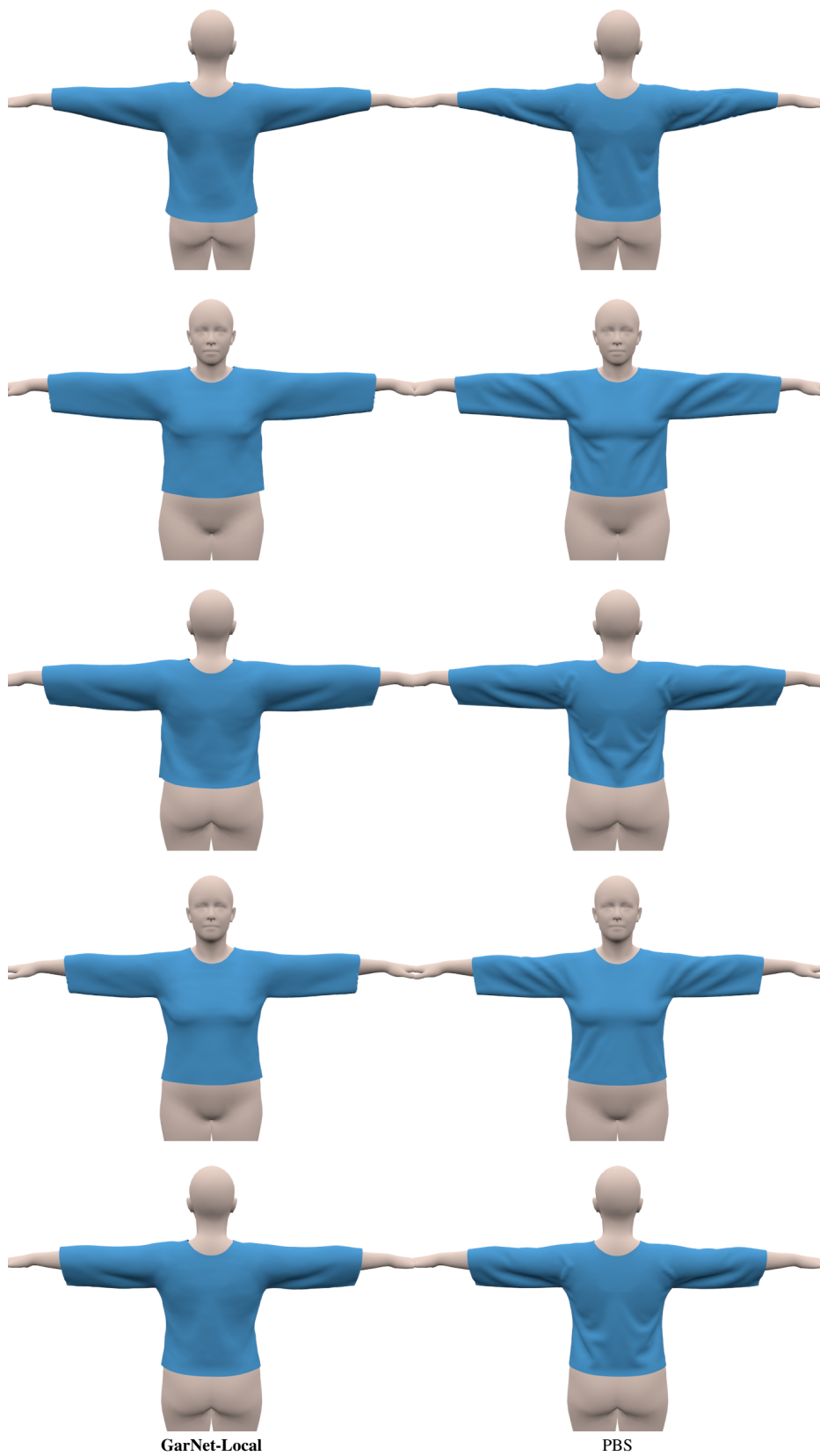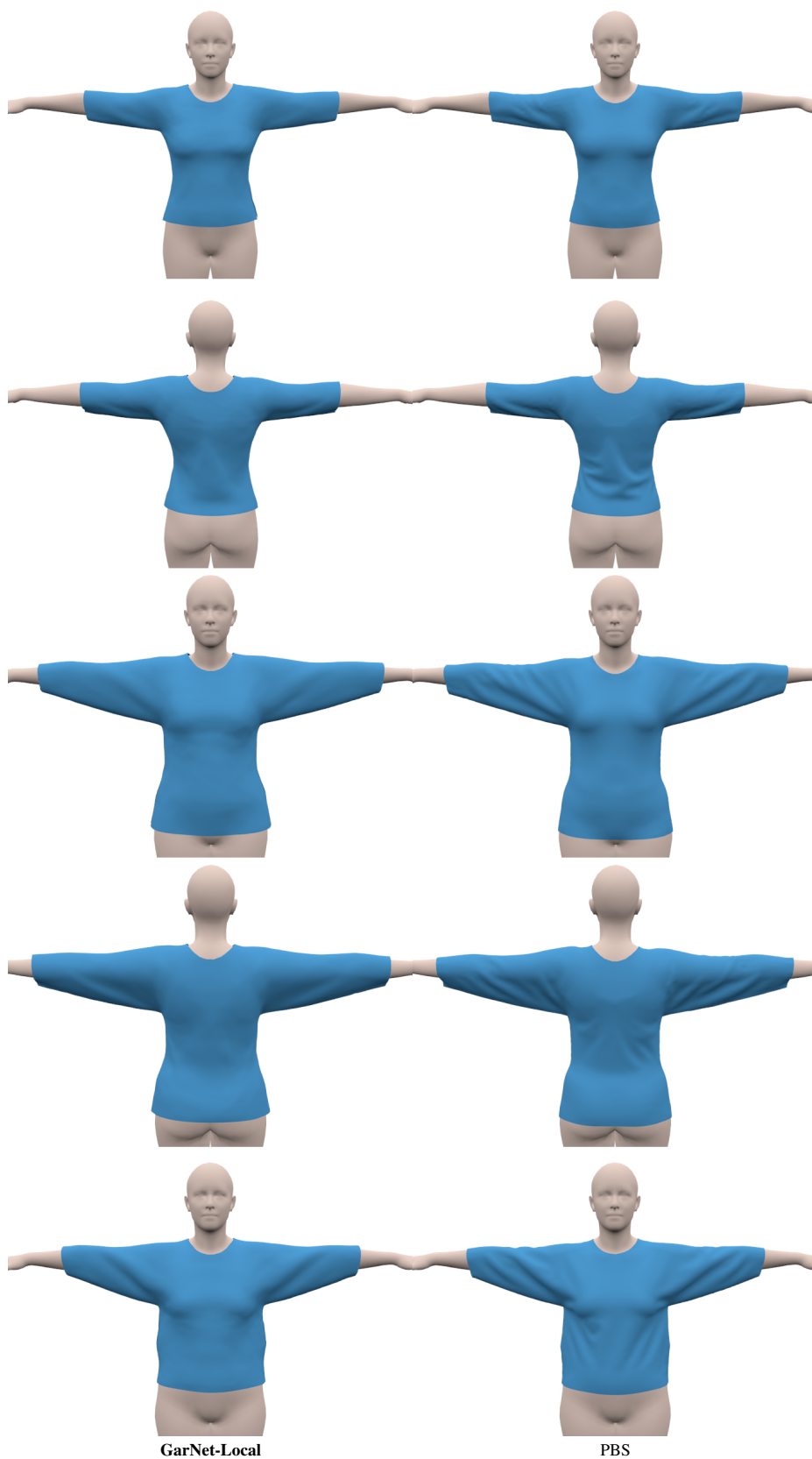**GarNet-Local**   **PBS**

Figure 10: Visual comparison of **GarNet-Local** predictions and PBS on the dataset of [8].

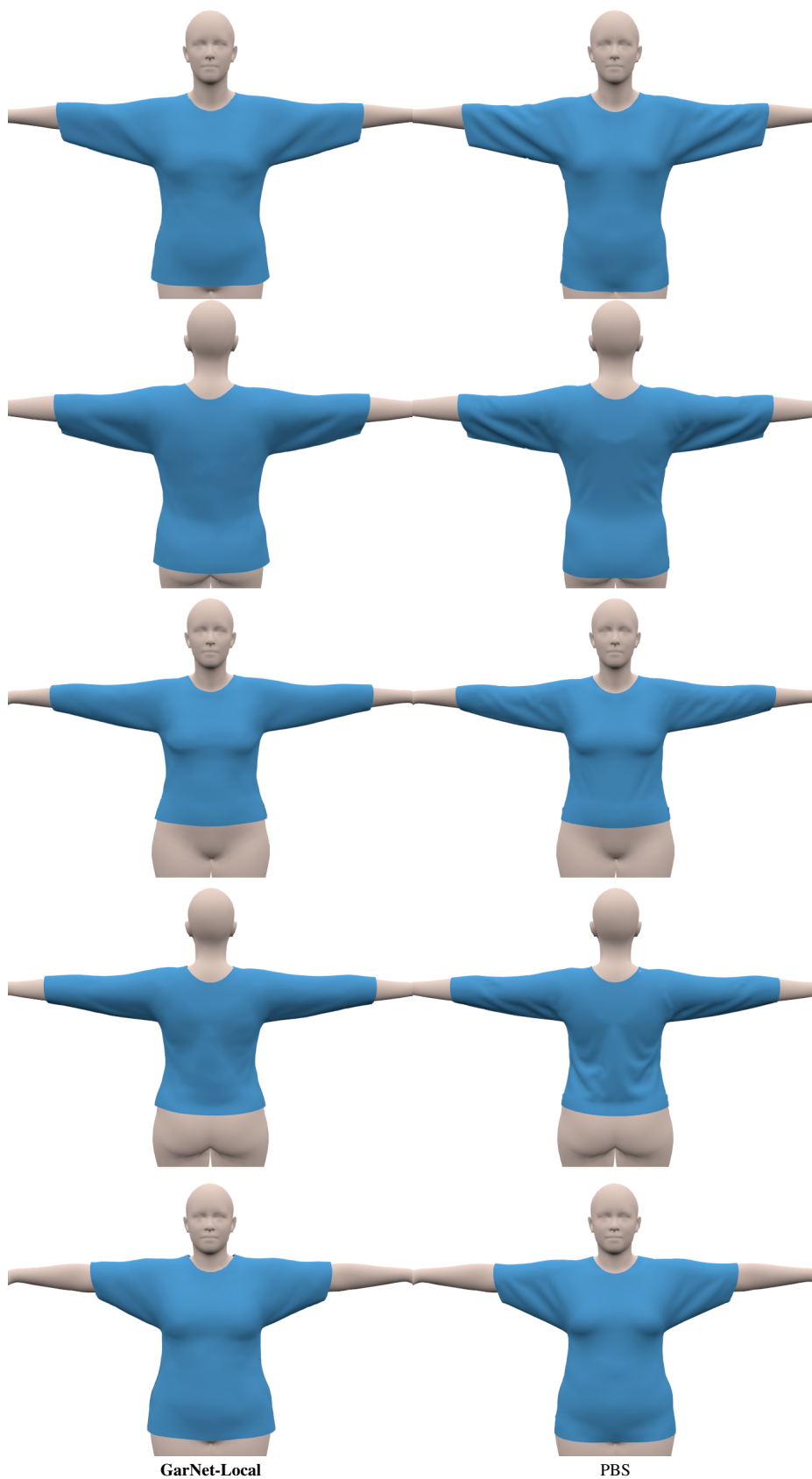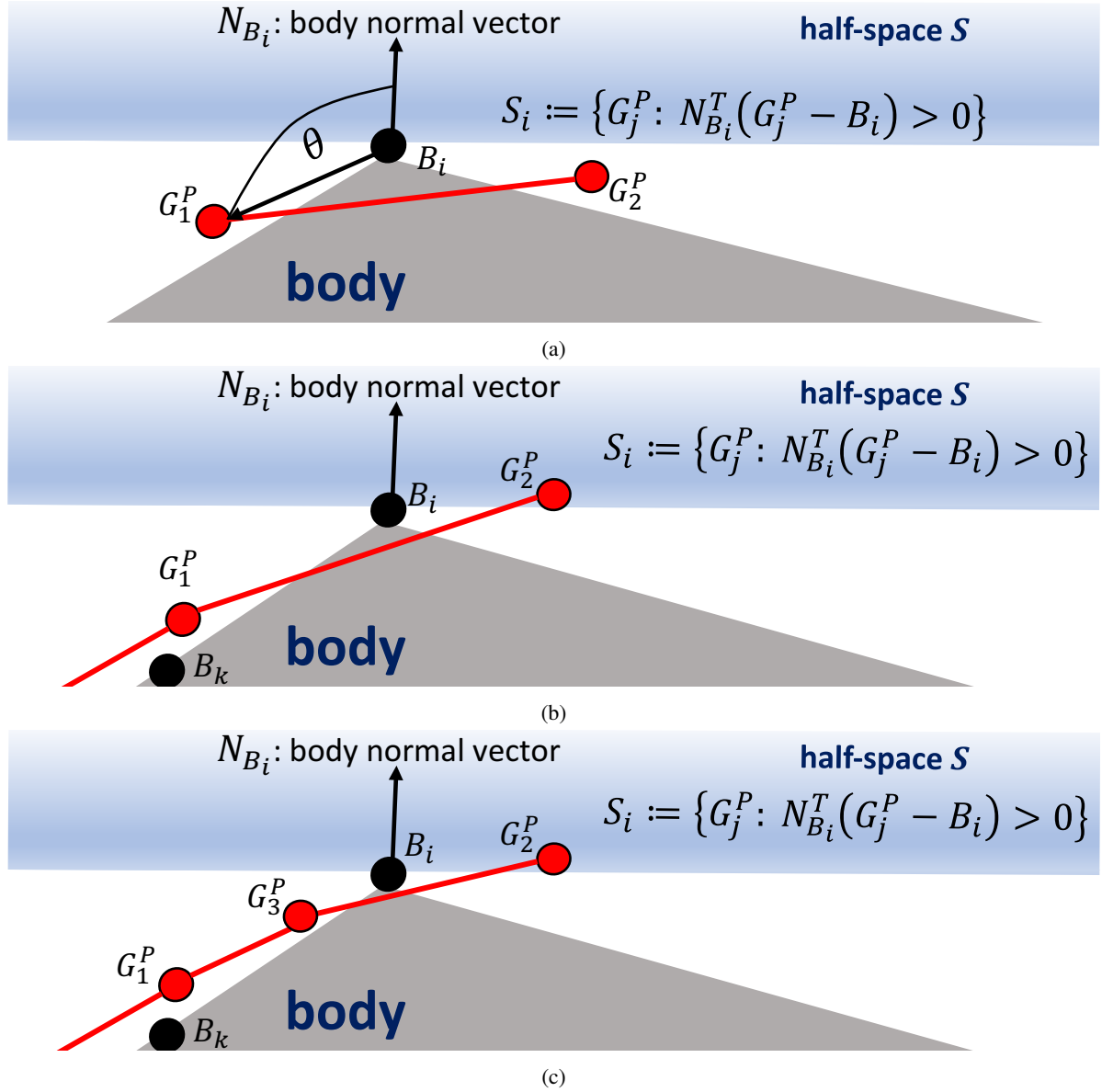Figure 11: Visual comparison of **GarNet-Local** predictions and PBS on the dataset of [8].

GarNet-Local   PBS

Figure 12: Visual comparison of **GarNet-Local** predictions and PBS on the dataset of [8].

N_{B_i}: body normal vector

half-space S

$S_i := \{ G_j^P : N_{B_i}^T (G_j^P - B_i) > 0 \}$

θ

B_i

G_1^P

G_2^P

body

(a)

N_{B_i}: body normal vector

half-space S

$S_i := \{ G_j^P : N_{B_i}^T (G_j^P - B_i) > 0 \}$

B_i

G_2^P

G_1^P

body

B_k

(b)

N_{B_i}: body normal vector

half-space S

$S_i := \{ G_j^P : N_{B_i}^T (G_j^P - B_i) > 0 \}$

B_i

G_2^P

G_3^P

G_1^P

body

B_k

(c)

Figure 13: Some examples for triangle-to-triangle intersection. The case in (a) is penalized by our loss term since G_1^P and G_2^P are matched to B_i. The case in (b) is not penalized by our loss term, however, it is more likely to occur when the garment resolution is less than the body one. When the garment resolution is higher as shown in the case (c), G_3^P is penalized for not being in the half-space of B_i, provided that G_3^P is matched to B_i. The situation in the case **b** could seldom occur in our smooth body meshes.

References

[1] I. Baran and J. Popovic. Automatic rigging and animation of 3d characters. *ACM Transactions on graphics (TOG)*, 26(3):72, 2007. 1

[2] CMU Graphics Lab Motion Capture Database, 2010. http://mocap.cs.cmu.edu. 1

[3] L. Kavan, S. Collins, J. Žára, and C. O'Sullivan. Skinning with Dual Quaternions. In *Proceedings of the 2007 Symposium on Interactive 3D Graphics and Games*, pages 39–46, 2007. 1

[4] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, Oct. 2015. 1

[5] Nvidia. NvCloth, 2018. https://docs.nvidia.com/gameworks/content/gameworkslibrary/physx/nvCloth/index.html. 1

[6] C. Qi, H. Su, K. Mo, and L. Guibas. Pointnet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *CVPR*, 2017. 6

[7] N. Verma, E. Boyer, and J. Verbeek. Feastnet: Feature-Steered Graph Convolutions for 3D Shape Analysis. In *CVPR*, 2018. 1

[8] T. Y. Wang, D. Ceylan, J. Popovic, and N. J. Mitra. Learning a Shared Shape Space for Multimodal Garment Design. In *ACM SIGGRAPH Asia*, 2018. 1, 2, 6, 7, 8, 9