

# Dress Anyone: A Method & Dataset for 3D Garment Retargeting

## Supplementary Material

### 1. Evaluation Metrics

Given a 3D garment mesh  $\mathcal{G}$  to be retargeted and the corresponding GT garment mesh  $\mathcal{G}_{GT}$  (where  $v_i \in \text{vertices}(\mathcal{G})$  and  $\hat{v}_i \in \text{vertices}(\mathcal{G}_{GT})$ ), we use the following standard metrics for evaluation:

**Chamfer Distance (CD):** Given two sets of points  $S_1$  and  $S_2$ , Chamfer distance measures the discrepancy between them as follows:

$$CD = \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2^2 \quad (1)$$

In our case,  $S_1 = \text{vertices}(\mathcal{G})$  and  $S_2 = \text{vertices}(\mathcal{G}_{GT})$ .

**Point-to-Surface (P2S) Distance:** P2S measures the average L2 distance between each vertex of the garment mesh and the nearest point to it on the target body surface.

**Interpenetration Ratio(IR):** It is computed as the ratio of the area of garment faces inside the body to the overall area of the garment faces; hence lower values are desired to ensure the least amount of penetration of the garment mesh with the target body mesh.

### 2. Discussion

#### 2.1. Isomap Embeddings for 3DBiCar

In order to enable coarse retargeting of a given 3D garment onto a biped cartoon character mesh from 3DBiCar[] dataset, we first need to estimate Isomap Embeddings for the character mesh. Since the body proportions of 3DBiCar samples are drastically different from human body, it is difficult to perform SMPL registration for Isomap Embedding extrapolation. However, 3DBiCar offers a parametric representation for all the characters and a common mesh template. Therefore, if we can estimate Isomap Embeddings for the common template, same Isomap embeddings can be applied to all other characters, as the common template is deformed according to the shape parameters, retaining the common meshing.

Due to lack one-to-one correspondences between SMPL template mesh and 3DBiCar template mesh, we first designate anchors on both the templates. As shown in [Figure 1](#), we carefully choose corresponding anchor vertices on both the meshes, considering the scale and proportions of both the templates. Though this process is manual, it is a one-time effort that needs to be carried only for the 3DBiCar template mesh. Once sparse salient correspondences are established, we transfer Isomap Embeddings from the SMPL

anchor vertices to 3DBiCar anchor vertices. For each of the remaining non-anchor vertex of the 3DBiCar template, we perform a weighted interpolation of all the neighboring anchor vertices, where the weight is a RBF kernel over the geodesic distances between that non-anchor vertex and neighboring anchor vertices.

The aforementioned approach can be applied to any parametric body representation to retarget any arbitrary 3D garment.

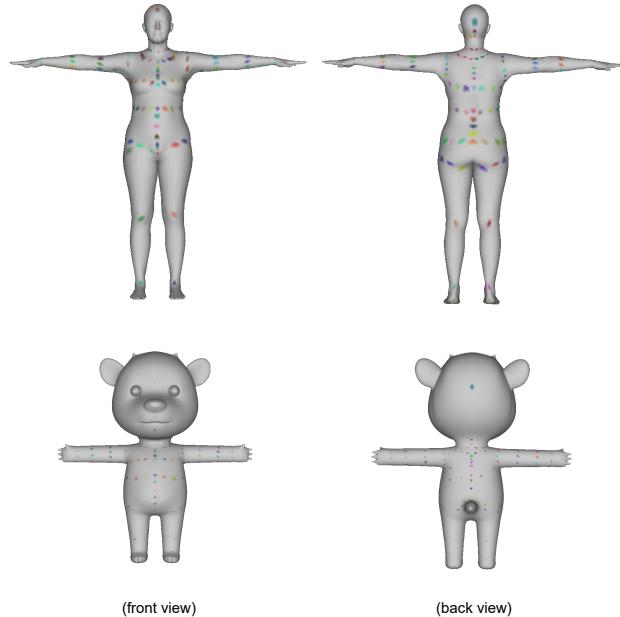


Figure 1. Anchor vertices on SMPL template (top row) and 3DBiCar template (bottom row) for Isomap Embedding transfer.

#### 2.2. Trade-off between Collision & Plausibility

We employ physics-based losses to achieve plausible pose-specific deformations of the coarsely retargeted garment. One of such loss is `e_hCollisionLoss` which is used to constrain the interaction of the garment with the underlying body, in order to avoid interpenetration of the garment mesh with the underlying body. However, the differentiable, soft nature of these losses doesn't provide a hard constraint over the body-garment collision. We observed that giving more weightage to collision loss results in implausible deformations of the garment over the body, increasing the convergence time for other constraints. Therefore, we empirically choose to downweight the collision term from the loss equation. Quantitatively stating, this results in a slight increase in Interpenetration Ratio(IR) values, however, the overall

067 draping is physically plausible. In order to get rid of small  
 068 interpenetrations, similar to [1, 2], we apply a interpenetration  
 069 solver as a post-processing step. Please note that we do  
 070 not apply interpenetration solver while performing quantita-  
 071 tive evaluation or comparisons.

### 072 2.3. Regarding comparison with DrapeNet

073 We observe that for Drapenet, good initial skinning of  
 074 canonical garment to the target pose is important, and any  
 075 noise in skinning results in outlier vertices which contribute  
 076 towards larger values of CD for Drapenet. Though we have  
 077 reported the quantitative comparison of our method with  
 078 Drapenet in the main paper, for a fair evaluation we also  
 079 compute the quantitative metrics for Drapenet on top 20  
 080 samples with lowest CD value & top 20 samples with high-  
 081 est CD values, and compare with the values on the same  
 082 samples using our method in [Table 2](#) and [Table 3](#) respec-  
 083 tively. For samples with lower CD values, Drapenet per-  
 084 forms better, however, in case of samples with higher CD  
 085 values, we significantly outperforms Drapenet.

## 086 3. Additional Experiments

### 087 3.1. Analysis of Isomap Embeddings

088 We propose a novel strategy that allows establishing cor-  
 089 respondences between different human scans, garments, or  
 090 anything that resembles human body structure. SMPL be-  
 091 ing a parametric human body model, acts as a reasonable  
 092 medium to establish correspondences across different body  
 093 shapes, poses, and appearances. As explained in the main  
 094 draft, once both the garment and the target body (paramet-  
 095 ric or non-parametric) are registered with SMPL, where the  
 096 target body can be an SMPL mesh itself, we compute 128-  
 097 dimensional isomap embeddings for each vertex of the gar-  
 098 ment and target body. Then, dense correspondences can  
 099 be established between the two by matching similar 128-  
 100 dimensional extrapolated features.

101 We arrive at this choice of feature modeling after care-  
 102 fully studying existing representations for dense correspon-  
 103 dence matching for humans. This problem is specifically  
 104 tough as humans are deformable objects and tend to un-  
 105 dergo non-rigid motion. Continuous Surface Embeddings  
 106 (CSE)[8] propose a learnable image-based representation  
 107 of dense correspondences and a model which predicts, for  
 108 each pixel in a 2D image, an embedding vector of the cor-  
 109 responding vertex in the object mesh, therefore establishing  
 110 dense correspondences between image pixels and 3D object  
 111 geometry. The authors show remarkable results in match-  
 112 ing correspondences across RGB human images via 16-  
 113 dimensional representation vectors. Recently, BodyMap[4]  
 114 proposed to extend this approach by extrapolating the CSE  
 115 embeddings of SMPLs registered with high-quality human  
 116 scans in UV space. We started with BodyMap representa-

Table 1. Analysis of choice of representations for correspondence estimation.  $\mathcal{R}_{score}$  takes values between 0 & 1, where lower values are preferred.

Representation	$\mathcal{R}_{score} \downarrow$
BodyMap[4]	0.955
16-dim. Isomap Embeddings	0.491
32-dim. Isomap Embeddings	0.473
64-dim. Isomap Embeddings	0.437
128-dim. Isomap Embeddings	0.426
256-dim. Isomap Embeddings	0.424

tion but later found it to produce a lot of false matching, and  
 117 we decided to analyze the behavior quantitatively.  
 118

119 The representation for correspondence estimation should  
 120 be rich and varied enough to avoid repetitions in the feature  
 121 space when extrapolated, otherwise, different body parts  
 122 would map nearby in the embedding space. More specifi-  
 123 cally, geodesically far-apart vertices should map far apart in  
 124 the embedding space and vice-versa. Based on this ideation,  
 125 we design an evaluation metric, **Richness Score**( $\mathcal{R}_{score}$ ) for  
 126 each vertex  $v_i$  of SMPL mesh, which is calculated as fol-  
 127 lows:

$$\mathcal{R}_{score_i} = (\mathcal{R}_{near_i} + \mathcal{R}_{far_i})/2 \quad (2)$$

$$\mathcal{R}_{near_i} = \frac{1}{k^2} \sum_{i=1}^k \min(|\mathcal{N}_{geo}^{rank} - \mathcal{N}_{emb}^{rank}|, k) \quad (3)$$

$$\mathcal{R}_{far_i} = \frac{1}{k^2} \sum_{i=1}^k \min(|\mathcal{F}_{geo}^{rank} - \mathcal{F}_{emb}^{rank}|, k) \quad (4)$$

133 where,  $\mathcal{N}_{geo}^{rank}$  &  $\mathcal{N}_{emb}^{rank}$  denotes the ranks of k-nearest  
 134 neighbors of  $v_i$  in both geodesic and embedding space, and  
 135 similarly,  $\mathcal{F}_{geo}^{rank}$  &  $\mathcal{F}_{emb}^{rank}$  denotes the ranks of k-farthest  
 136 neighbors of  $v_i$  in both geodesic and embedding space.  
 137 Thus,  $\mathcal{R}_{score}$  penalizes if the rank of neighbors (k-nearest  
 138 and k-farthest) in geodesic and embedding space doesn't  
 139 match. We report the values in Table.??, where it can be  
 140 seen that extrapolating isoembedding values in Euclidean  
 141 space has a better effect than BodyMap[4]. The remaining  
 142 values show that high dimensionality is preferred. However,  
 143 empirically, values are saturated once a significant dimen-  
 144 sionality is reached.

### 145 3.2. Qualitative Comparison with M3DVTON

146 In context of virtual tryon for garments, several 2D VTON  
 147 methods [3, 6, 7, 9, 10, 12] exist which employ generative  
 148 networks for synthesizing 2D garments over 2D human im-  
 149 ages. However, 2D VTON methods have limited ability in

Table 2. Top-20 samples with **lowest** CD values

Module	Garment Type	CD ↓	P2S ↓	IR Ratio % ↓
DrapeNet	Top	<b>0.000007</b>	<b>0.00017</b>	<b>0.0</b>
	Bottom	0.00123	<b>0.00251</b>	<b>0.0001</b>
Ours	Top	0.00031	0.00806	0.796
	Bottom	<b>0.00017</b>	0.00588	01.6146

Table 3. Top-20 samples with **highest** CD values

Module	Garment Type	CD ↓	P2S ↓	IR Ratio % ↓
DrapeNet	Top	1.08879	0.032882	1.463
	Bottom	1.15114	0.05091	4.636
Ours	Top	<b>0.0315</b>	<b>0.02338</b>	<b>0.659</b>
	Bottom	<b>0.00117</b>	<b>0.01317</b>	<b>2.667</b>

150 terms of multiview consistency, and the draping is hallucinated, leading to implausible shape-specific deformations.  
 151 M3DVTON[13] comes closest to enabling 3D VTON application,  
 152 by learning to synthesize 2.5D representation (front back 2D depth maps) to reconstruct 3D mesh from garment  
 153 and target images, while still performing tryon in 2D image  
 154 sapce. However, its generative nature leads to hallucinations  
 155 and there is no support for 3D garments/bodies.  
 156

157 **Figure 2** shows a comparison of M3DVTON[14] with  
 158 our framework on random internet images (as mentioned  
 159 earlier, we use off-the-shelf method [11] to extract 3D gar-  
 160 ments and target human body). It is evident from the figure  
 161 that since M3DVTON performs retargeting in 2D space, it  
 162 doesn't produce accurate geometric deformations. More-  
 163 over, since it uses a supervised keypoint detection method  
 164 for initial TPS-based draping, it suffers when the target sub-  
 165 ject's garment category doesn't match the source garment  
 166 category. However, our method doesn't suffer from such  
 167 limitations and can retarget arbitrary garments on arbitrary  
 168 targets.  
 169

## 170 4. Real-3DVTON Dataset

171 **Figure 3** and **Figure 4** shows samples of texture-3D gar-  
 172 ments from our proposed Real-3DVTON dataset.  
 173 We also show an example of retargeting real, arbitrarily  
 174 posed 3D garments from Real-3DVTON dataset on non-  
 175 parametric human scans from THUman2.0 dataset in **Fig-**  
 176 **ure 5.**

## 177 5. Additional Qualitative Results

178 **Figure 6** shows qualitative results of retargeting 3D gar-  
 179 ments onto 3D human meshes reconstructed from images  
 180 (using [11, 15]). This is yet another proof of a good gen-  
 181 eralization of our method on in-the-wild OOD samples (e.g.  
 182 yoga pose).



Figure 2. Qualitative Comparison with M3DVTON

183 **Figure 7** shows additional examples of draping on biped  
 184 character meshes from 3DBiCar[5] dataset.



Figure 3. *Topwear*: The figure shows visualization of our collected dataset, first three rows depict the geometry of our collected garment in different poses, while last three shows the textured rendering of the respective geometries.



;

Figure 4. *BottomWear*: The figure shows visualization of our collected dataset, first three rows depict the geometry of our collected garment in different poses, while last three shows the textured rendering of the respective geometries.

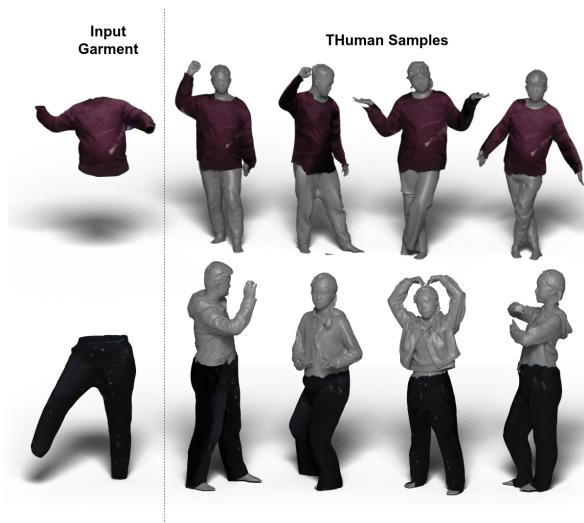


Figure 5. Retargeting garments from our proposed Real-3DVTON dataset

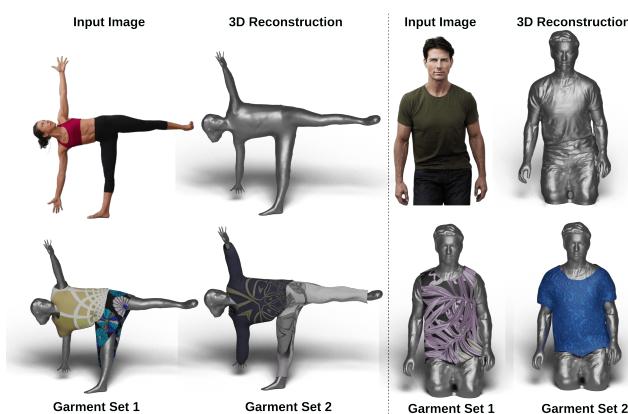


Figure 6. 3D Garment draping on Avatars reconstructed from internet images.



Figure 7. Dressing samples from 3DBiCar dataset

**References**

- [1] Ruochen Chen, Liming Chen, and Shaifali Parashar. Gaps: Geometry-aware, physics-based, self-supervised neural garment draping, 2024. 2
- [2] Luca De Luigi, Ren Li, Benoît Guillard, Mathieu Salzmann, and Pascal Fua. Drapenet: Generating garments and draping them with self-supervision, 2022. 2
- [3] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S Davis. Viton: An image-based virtual try-on network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7543–7552, 2018. 2
- [4] Anastasia Ianiina, Nikolaos Sarafianos, Yuanlu Xu, Ignacio Rocco, and Tony Tung. Bodymap: Learning full-body dense correspondence map. In *CVPR*, 2022. 2
- [5] Zhongjin Luo, Shengcai Cai, Jinguo Dong, Ruibo Ming, Liangdong Qiu, Xiaohang Zhan, and Xiaoguang Han. Rabbit: Parametric modeling of 3d biped cartoon characters with a topological-consistent dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 3
- [6] Davide Morelli, Alberto Baldrati, Giuseppe Cartella, Marcella Cornia, Marco Bertini, and Rita Cucchiara. LaDIFTON: Latent Diffusion Textual-Inversion Enhanced Virtual Try-On. In *Proceedings of the ACM International Conference on Multimedia*, 2023. 2
- [7] Assaf Neuberger, Eran Borenstein, Bar Hilleli, Eduard Oks, and Sharon Alpert. Image based virtual try-on network from unpaired data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5184–5193, 2020. 2
- [8] Natalia Neverova, David Novotný, Vasil Khalidov, Marc Szafraniec, Patrick Labatut, and Andrea Vedaldi. Continuous surface embeddings. *ArXiv*, abs/2011.12438, 2020. 2
- [9] Dan Song, Tianbao Li, Zhendong Mao, and An-An Liu. Spviton: shape-preserving image-based virtual try-on network. *Multimedia Tools and Applications*, 79(45):33757–33769, 2020. 2
- [10] Bochao Wang, Huabin Zheng, Xiaodan Liang, Yimin Chen, Liang Lin, and Meng Yang. Toward characteristic-preserving image-based virtual try-on network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 589–604, 2018. 2
- [11] Yuliang Xiu, Jinlong Yang, Xu Cao, Dimitrios Tzionas, and Michael J. Black. ECON: Explicit Clothed humans Obtained from Normals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 3
- [12] Ruiyun Yu, Xiaoqi Wang, and Xiaohui Xie. Vtnfp: An image-based virtual try-on network with body and clothing feature preservation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10511–10520, 2019. 2
- [13] Fuwei Zhao, Zhenyu Xie, Michael Kampffmeyer, Haoye Dong, Songfang Han, Tianxiang Zheng, Tao Zhang, and Xiaodan Liang. M3d-vton: A monocular-to-3d virtual try-on network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13239–13249, 2021. 3
- [14] Fuwei Zhao, Zhenyu Xie, Michael Kampffmeyer, Haoye Dong, Songfang Han, Tianxiang Zheng, Tao Zhang, and Xiaodan Liang. M3d-vton: A monocular-to-3d virtual try-on network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13239–13249, 2021. 3
- [15] Heming Zhu, Lingteng Qiu, Yuda Qiu, and Xiaoguang Han. Registering explicit to implicit: Towards high-fidelity garment mesh reconstruction from single images, 2022. 3