Multi-Garment Net: Learning to Dress 3D People from Images
Supplementary

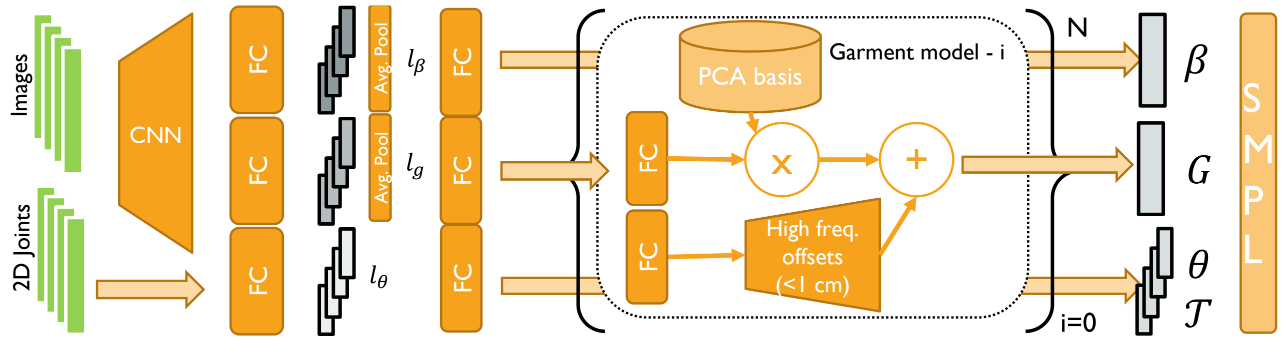

Figure 1:
Detailed architecture of MGN. We use CNNs to encode image and 2D joint information into latent codes. MGN incorporates explicit garment models in the prediction stream and maps the latent codes from CNN into the garment parameters directly. Our garment sub-networks essentially decode the garment parameters and add high frequency offsets on top of the prediction.

Figure 2: Robust segmentation is essential to register garments without artifacts. Here we can clearly see that incorrect segmentation leads to artifacts in registration. From left to right: scan, erroneous segmentation of feet, distorted pant registration, registration with corrected segmentation.

1. Additional results

In this section we show further results that highlight practical applications of our proposed approach. We show transfer of texture over a fixed topology and transfer of topology itself across subjects. We show more qualitative comparison with [1] in Fig. 8. Figs. 2 and 3 clearly show the significance of our segmentation pipeline and laplacian initialization respectively.

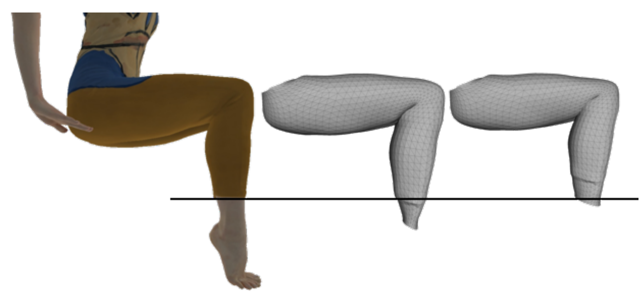

Figure 3: Our lapacian initialization is important to register garments with very different topology and geometry. From right to left: snapshot of the source to be registered, without laplacian initialization the optimization often gets stuck in local minima and the pant template is not pulled up at the boundary, our laplacian initialization smoothly matches the template boundary to the scan garment boundary.

Inferring 3D garments and underlying body In this experiment we use 8 images of each unseen test subject to generate 3D outputs using MGN. In Fig. 5, we show MGN predictions on the test set. We also study the effect of using a single image as input at inference time and report the following results (mean vertex-to-surface error): 15 frames: 5.43mm, 10 frames: 5.48mm, 8 frames: 5.78 mm, 6 frames: 5.60mm, 5 frames: 5.66mm , 3 frames: 5.95mm, 2 frames: 6.38mm.

	Alldieck *et al.*[1]	Ours	
	GT Pose	**GT Pose**	**Full pred.**
Pants	**5.44**	5.57	10.16
Short-Pants	8.23	**5.97**	10.00
T-Shirt	5.80	**5.63**	11.97
Shirt	**5.71**	6.33	9.05
Coat	5.85	**5.66**	9.09

Table 1: Quantitative comparison with [1]. We compute per garment mean vertex-to-surface error. We also report the performance of our approach with GT and predicted poses.

Table 1 shows further comparison between [1] and our approach. This experiment also details per-garment performance of MGN.
We also ablate the reconstruction accuracy of MGN, with and without using the high frequency displacement field on top of the PCA based reconstruction. We report mean vertex-to-surface error of 6.44mm without using the high frequency term as opposed to 5.78mm while using it.

Texture Transfer Our approach models garments as geometric deformations on a canonical garment mesh. Unlike single mesh methods, this preserves the semantic meaning of individual vertices across different geometries over a fixed topology. We leverage this consistency to map garment textures across different garments. Fig.4 shows the transfer of texture over fixed topology (shirts, coats and pants) but varying geometries.

Garment Re-targeting Our formulation of layered multi-mesh representation of clothed bodies allows us to dress SMPL using a digital wardrobe. In Fig.6 we show garment re-targeting using 3D data. More interestingly we can re-target garments just using images by inferring the 3D garments and body shape using MGN. We show the results in Fig.7. This experiment highlights the strength of our approach in correctly estimating 3D geometry of the garments and human body shape. Note that each garment class has a different mesh topology (along with different geometry). Our approach can seamlessly add and remove these disparate meshes. To the best of our knowledge this is the first approach that allows topology-agnostic garment re-targeting using only images.
We further experiment with garment fitting by extracting garments from the images of different subjects and dress diverse SMPL bodies with the same garment. Fig. 10 shows the re-targeting results on diverse body shapes. Fig. 9 shows the comparison between a naive and our body aware re-targeting.

Testing on real world data The input to our approach is semantic segmentation. This allows us to train our model entirely on scan data and test on real world images. Fig. 11 shows the garment predictions from our network on PeopleSnapshot [4] dataset. Note that network was never trained on real images.

2. Limitations and Future Work:

In this section (and Fig. 12) we discuss some of the research avenues that our approach opens up or shows unsatisfactory performance. We hope that this would simulate further research into the direction of modelling 3D garments, underlying body and their interactions.

- The proposed approach does not deal with pose dependent deformations.
- Skinning garments though convenient, often leads to artifacts while re-posing in case of extreme poses (Fig 12 a).
- Re-targeting relies heavily on segmentation. In case we wrongly segment part of skin as garment our approach incorrectly moves the skin along with the garment. (Fig 12 b)
- Current approach cannot impaint the skin texture underneath the garments. This creates artifacts when re-targeting short garments (eg: t-shirt) on a body which was previously wearing long garments (eg: coat) (Fig 12 c)
- In its current form MGN does not model hair.

References

[1] Thiemo Alldieck, Marcus Magnor, Bharat Lal Bhatnagar, Christian Theobalt, and Gerard Pons-Moll. Learning to reconstruct people in clothing from a single RGB camera. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2, 5

[2] Thiemo Alldieck, Marcus Magnor, Weipeng Xu, Christian Theobalt, and Gerard Pons-Moll. Video based reconstruction of 3D people models. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2018. 6

Figure 4: Texture transfer. We model each garment class as a mesh with fixed topology and surface parametrization. This enables us to transfer texture from any garment to any other registered instance of the same class. Rows correspond to shirt, coat and pant respectively. The first column shows the source garment mesh, while the subsequent images show original and transferred garment texture registrations.

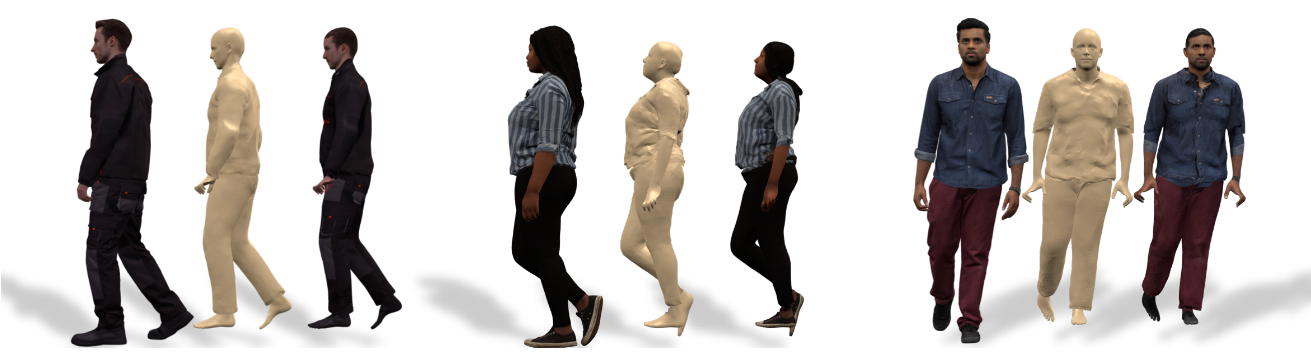

Figure 5: Figure shows the MGN predictions for three test subjects. Each set shows (left to right) the 2D image of the subject and corresponding MGN prediction. We show the predictions with and without texture for higher clarity

Figure 6: Re-targeting using registered 3D garments and body shapes. Each set contains (from left to right) a source subject, a target subject and the re-dressed target. In first row we show re-targeting two different source garment sets on the same target subject in different poses. In the second row we show re-targeting of the same source garment set to different targets. In the third row we some some more re-targeting examples.

Figure 7: Re-targeting using images. Each set contains (from left to right) a source subject, a target subject and the re-dressed target. In first row we show re-targeting two different source garment sets on the same target subject in different poses. In the second row we show re-targeting of the same source garment set to different targets. Note that re-targeting reflects the difference of underlying body shapes. In the third row we some some more re-targeting examples highlighting minimal distortions while re-dressing. We use MGN to infer 3D garments and body shape from the images of the source subject and re-target the garments to target subject.

Figure 8: More comparative results with Alldieck *et al.*[1] (left) and our approach (right). Notice that our approach has fewer distortions.

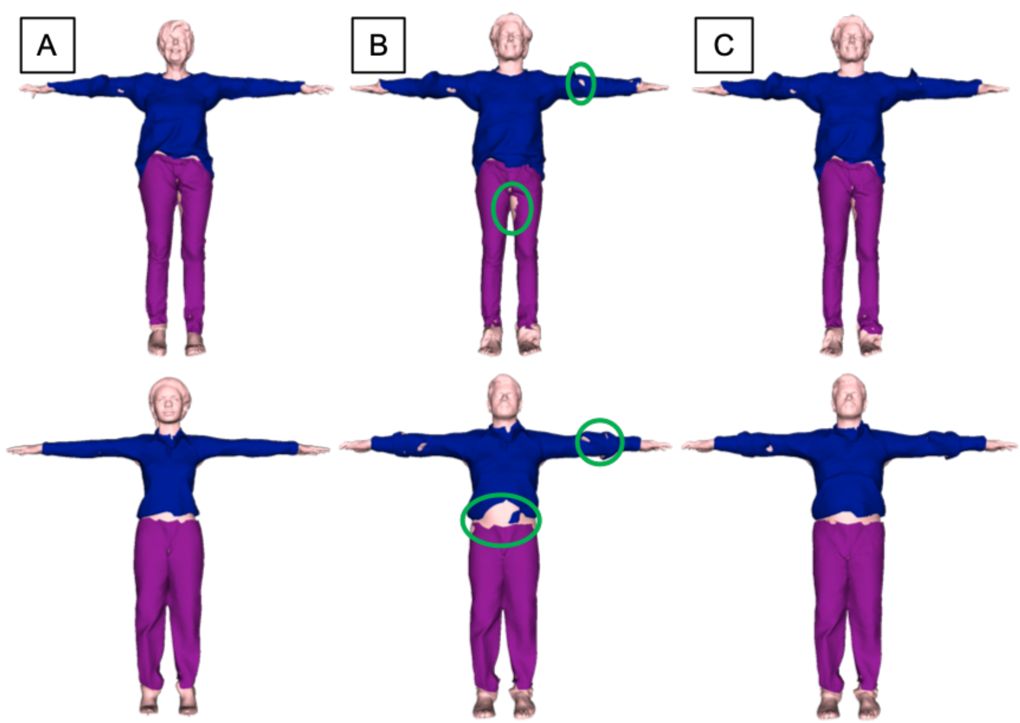

Figure 9: Comparison between naive and our body aware re-targeting. In each set, (left to right) we show the source, target with naive re-targeting and target with our body aware re-targeting. Notice the lower inter-penetrations in our body aware approach.

Figure 10: Garments inferred by MGN can easily be used to dress novel subjects. Rows correspond to female and male subjects respectively. In each row we show (from left to right) the images/ video from source subject, the undressed body shapes from SMPL shape subjects and dressed body shapes respectively.

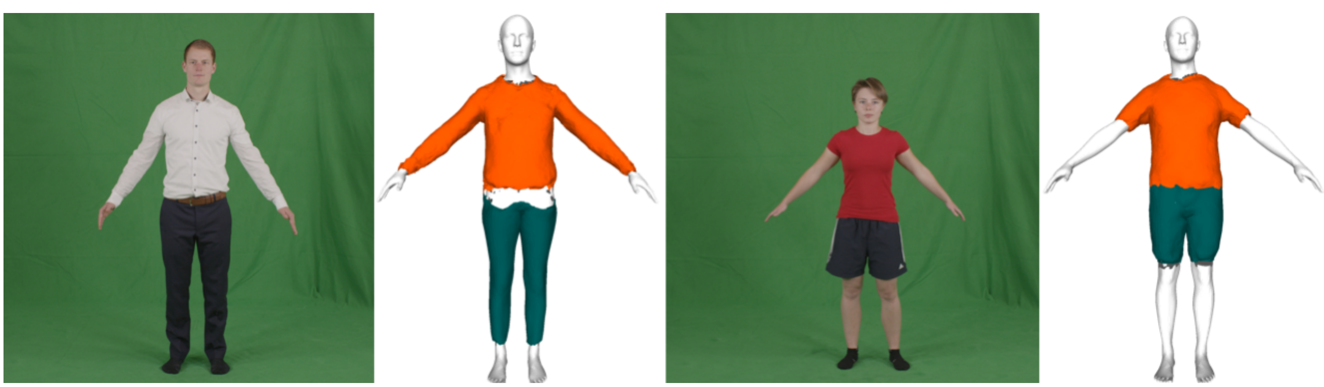

Figure 11: Garment prediction by MGN on real world dataset [2]. The network is trained entirely on scan data but testing is done on real images. This highlights the utility of appearance agnostic semantic segmentation as input.

Figure 12: Ours is the first approach to infer separable 3D garments from images. Though very promising, the proposed approach has shortcomings. In this figure we present some interesting challenges for future work. From left to right: A) Unposing artifacts due to skinning, B) part of source hair got moved along with the garments due to incorrect segmentation at the boundary, C) Current approach cannot impaint texture under clothing.