UV-based reconstruction of 3D garments from a single RGB image

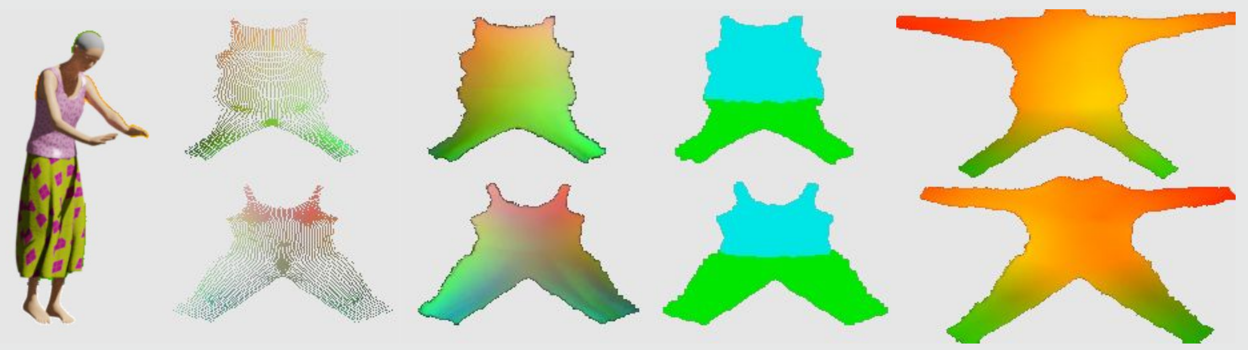
Albert Rial-Farras, Meysam Madadi, Sergio Escalera

Introduction

- Inferring 3D shapes from a single viewpoint is an essential human vision feature extremely difficult for computer vision machines.
- Despite the advances in the field of 3D human reconstruction, most research has concentrated only on unclothed bodies and faces, but modelling and recovering garments have remained notoriously tricky.
- We are interested in UV maps compared to other 3D surface representations such as meshes, point clouds or voxels, which are the ones commonly used in other 3D deep learning models.
- In this paper we adapt the LGGAN [1] architecture to predict garment UV map from the input image. We also introduce 3D loss functions to improve the surface quality.

UV Maps

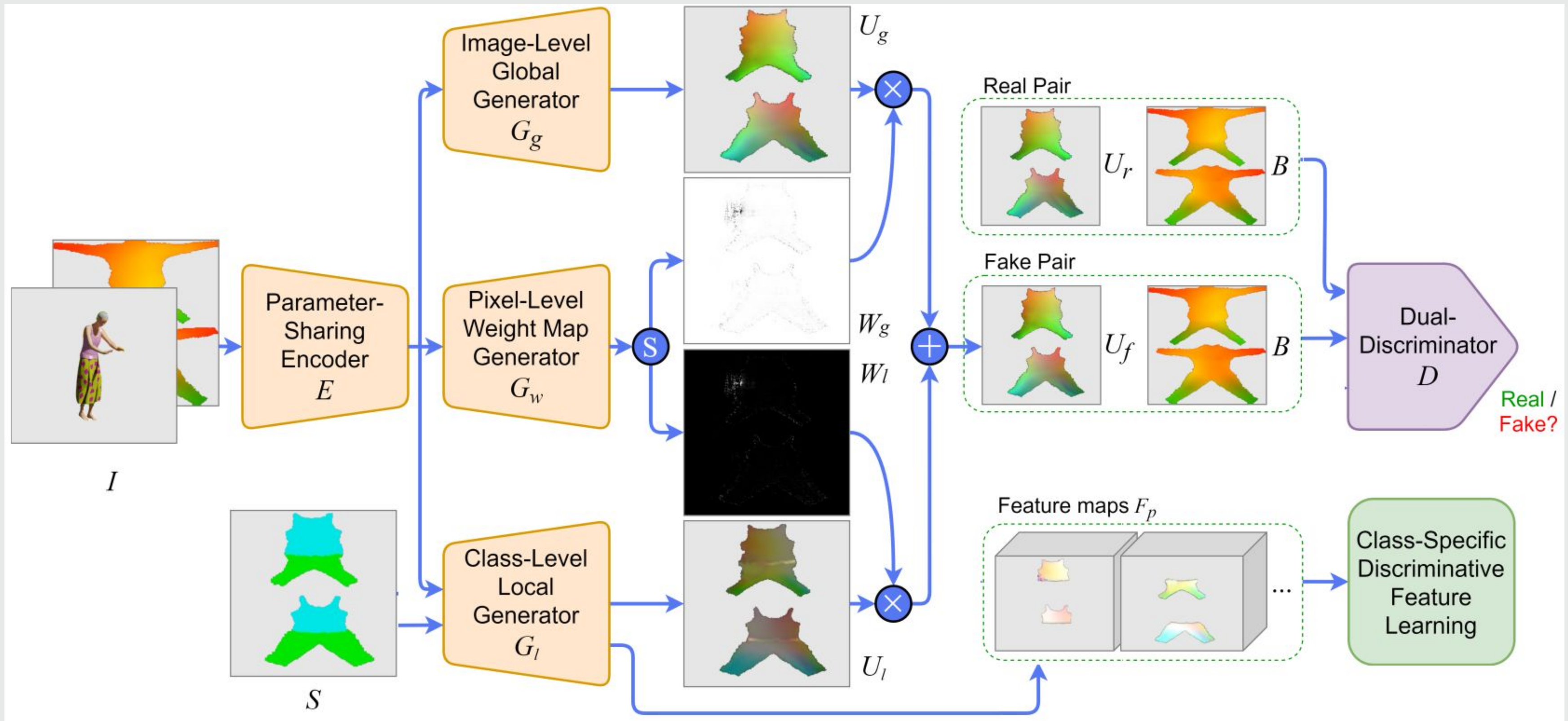
- Garments are registered on top of SMPL [2] mesh to have homogeneous topology at both training and inference time.
- The UV coordinates are discrete points, thus UV maps have empty gaps between vertices. we use image inpainting techniques to estimate the values of the empty spaces.
- We use displacement UV maps that store garment vertices as an offset over the estimated SMPL body vertices.
- We create UV maps for the garment mesh, garment semantic segments and body mesh.



Biography

[1] H. Tang, D. Xu, Y. Yan, P. H. S. Torr, and N. Sebe. Local class-specific and global image-level generative adversarial networks for semantic-guided scene generation, CVPR, 2020.
[2] 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.
[3] M. Madadi, H. Bertiche, W. Bouzouita, I. Guyon, and S. Escalera. Learning cloth dynamics: 3d + texture garment reconstruction benchmark. In Proceedings of the NeurIPS 2020 Competition and Demonstration Track, PMLR, volume 133, pages 57–76, 2021.
[4] E. Corona, A. Pumarola, G. Alenya, G. Pons-Moll, and F. Moreno-Noguer. Smplicit: Topology-aware generative model for clothed people, CVPR, 2021.

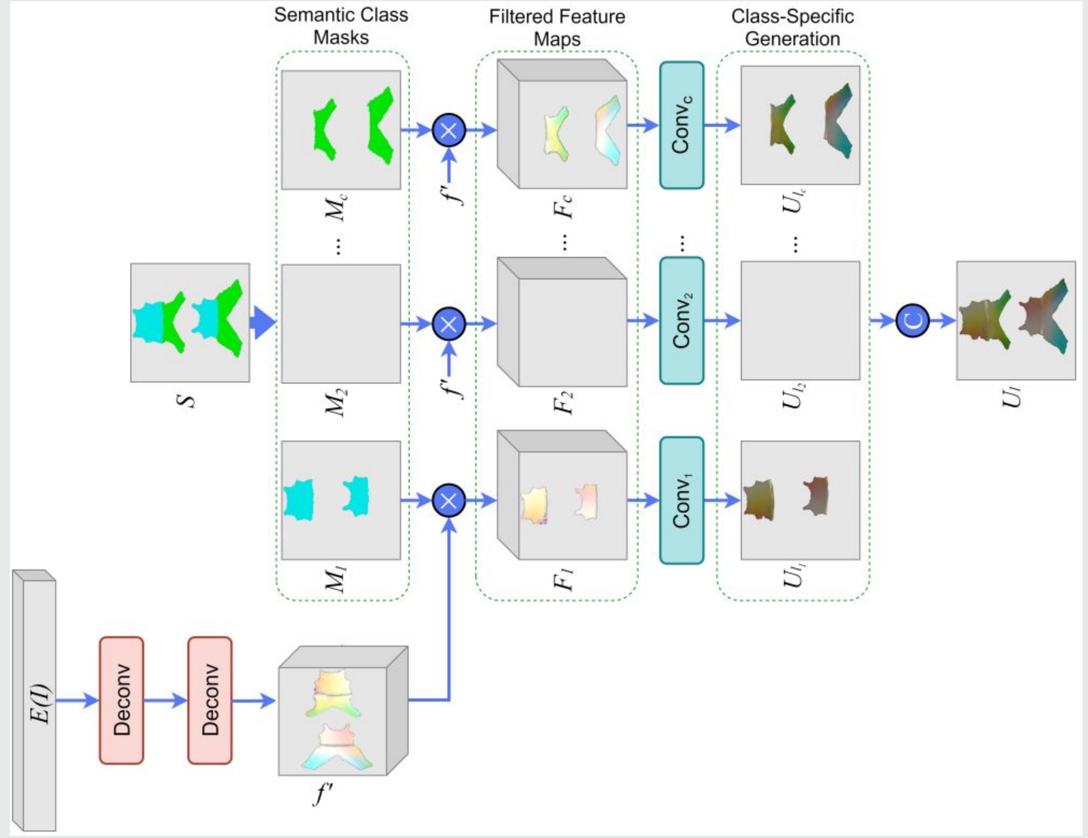
RGB to UV Map Translation - LGGAN Architecture



- LGGAN is conditioned on body UV map and garment UV semantic segments,
- Global decoder G_g is responsible to predict **low** frequency details (overall shape),
- Local decoder G_l is responsible to learn specific dynamics w.r.t. each garment class,
- The amount of aggregation of global and local information is controlled by weight decoder G_w through softmax,
- Class-specific discriminative feature learning branch classifies masked features to the target garment class,
- Discriminator D receives the conditioning image (body UV map) and predicted garment UV map to distinguish real data from fake.

Class-Specific Local Generation Network

- Shared latent code is upconvolved to form per-pixel features,
- Features are masked into specific branches by garment semantic labels,
- garment-specific convolutions map the features to relative 3D coordinates,
- All branches are aggregated to form the final UV map.



Generator Loss

- L1 loss on the reconstructed UV map,
- Smooth L1 loss on the reconstructed mesh,
- L2 loss on the surface normals based on nearest neighbor matching,
- Laplacian smoothing regularization,
- Edge length regularizan.

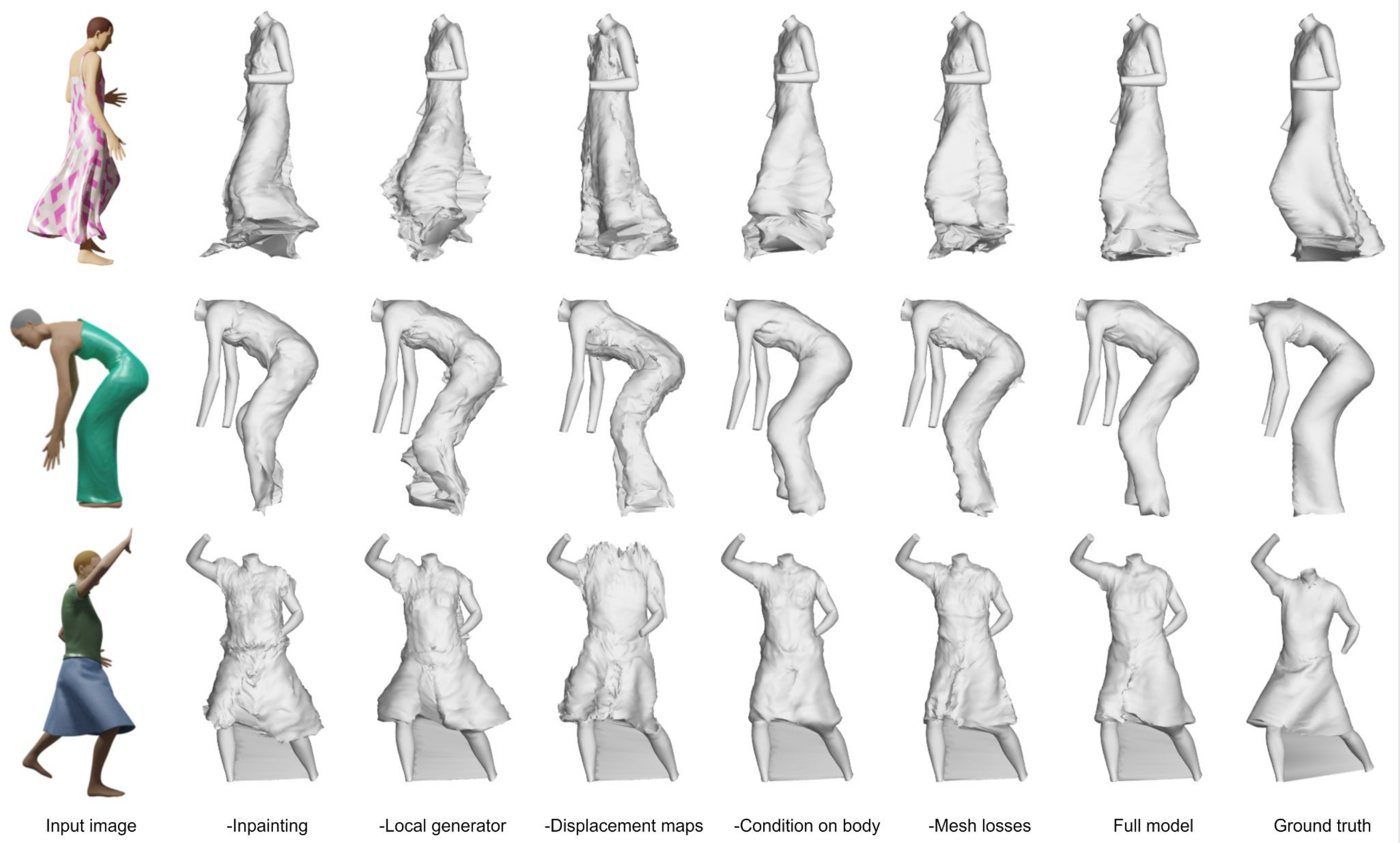
CLOTH3D++ Dataset [3]

- We use CLOTH3D++, the first large-scale synthetic dataset of 3D clothed human sequences.
- Garments are simulated on top of thousands of different human pose sequences and body shapes, generating realistic cloth dynamics.
- It has over 2 million 3D samples with a large variety of garment types, topology, shape, size, tightness and fabric.

Evaluation Metric

- We use surface-to-surface (S2S) error [3], an extension of the chamfer distance (CD), that computes the distance based on the nearest faces rather than nearest vertices.
- As S2S evaluates the results based on mesh format, during evaluation we unwrap UV map representations into 3D meshes.

Results



Model	S2S error (in mm)						
	Top	T-shirt	Trousers	Jumpsuit	Skirt	Dress	All
Full model	**18.1**	**21.6**	**15.3**	**18.3**	25.6	**21.1**	**19.1**
-Mesh losses	18.3	21.9	15.4	18.5	**25.1**	**21.1**	19.2
-Condition on body	18.9	22.3	16.3	**18.3**	26.4	21.6	19.5
-Displacement maps	45.6	44.5	42.1	41.4	47.8	46.5	43.7
-Local generator	24.5	28.3	23.2	26.3	44.9	32.1	27.8
-Inpainting	20.9	24.1	22.6	21.3	42.6	31.2	24.9

Comparison with SMPLicit [4]