Laboratory for Artificial Intelligence in Design
人工智能設計研究所

School of Fashion & Textiles
時裝及紡織學院

The Hong Kong Polytechnic University
香港理工大學

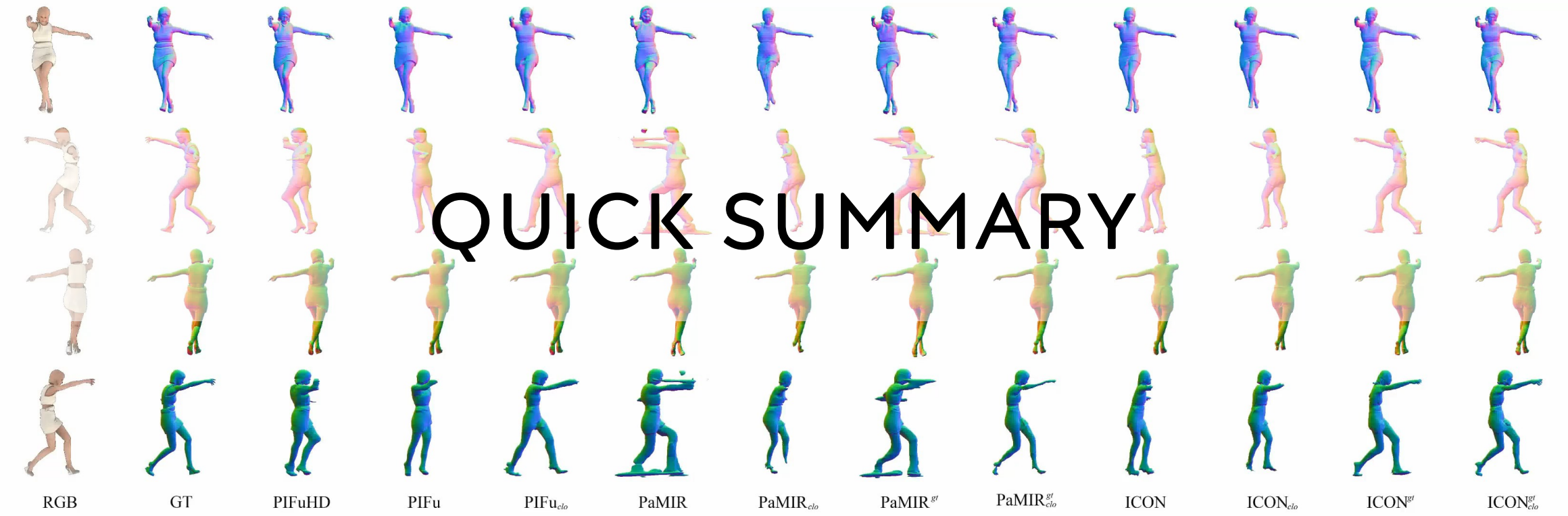
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

WED-PM-047

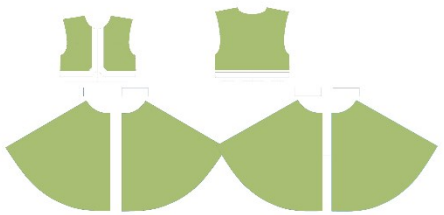
CLOTH4D: A Dataset for Clothed Human Reconstruction

Xingxing Zou[1,3] ⋆, Xintong Han[2] ⋆, Waikeung Wong[3,1]*
[1]Laboratory for Artificial Intelligence in Design, The Hong Kong Polytechnic University, [2]Huya Inc.,
[3]School of Fashion and Textiles, The Hong Kong Polytechnic University

QUICK SUMMARY



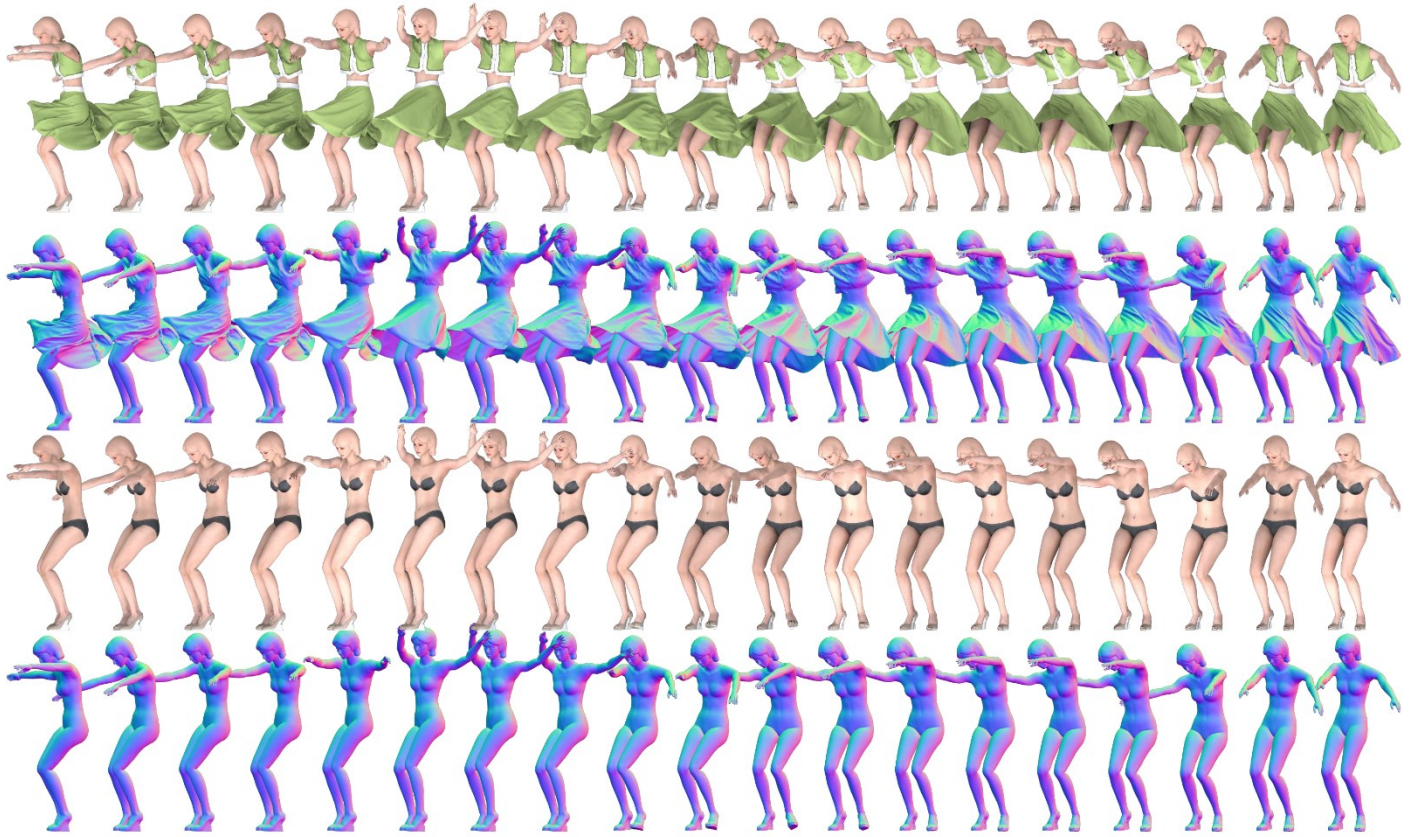
Motivation



(a) 2D patterns

(b) varied poses

(c) multi views



(d) series of paired dynamic textured meshes

We introduce **CLOTH4D** to fill the gap in large-scale and high-quality 4D clothing data.

Characteristics

1) Accurate and detailed clothing textured meshes
2) Separated textured clothing and under-clothing body meshes
3) Clothed human motion sequences covering fundamental and complicated dynamics
4) **300,000** meshes with pared naked model | **1,052** source files of 3D outfits releasing now

Reality Check to SOTAs

Quantitative Observations

1) GT-SMPL-guided + pixel-aligned > pure pixel-aligned > HMR-SMPL-guided + pixel-aligned.
2) With GT SMPL meshes, ICON^{gt} and PaMIR^{gt} are significantly better than using estimated SMPL (i.e., ICON and PaMIR). But, GT SMPL is unavailable at test time.
3) We propose temporal metrics and find pure pixel-aligned methods have higher reconstruction stability since the off-the-shelf HMR methods are jittery and unstable.
4) ICON outperforms PaMIR as it models mesh-based local features while PaMIR depends more on global information and its volumetric representation is of limited resolution.

Reality Check to SOTAs

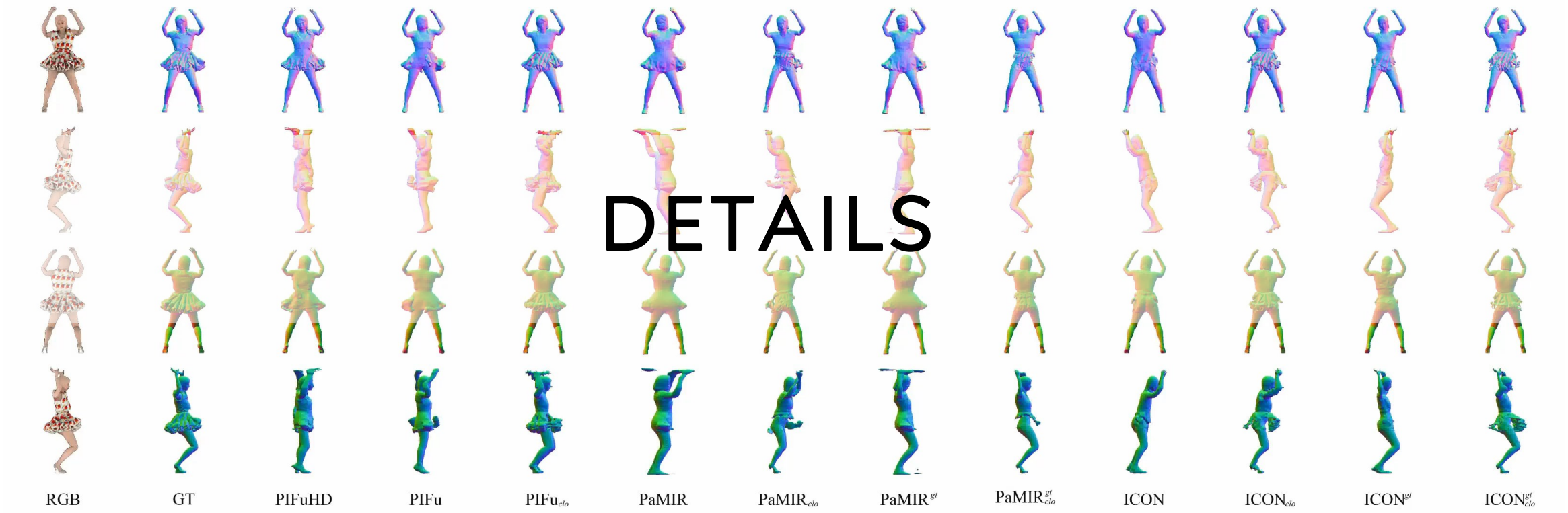
Qualitative Observations

1) All baselines reconstruct the overall body shape. PIFuHD presents the finest details by focusing on local features, but it faces overfitting and poor generalization to complicated clothing and large motions.
2) ICON^{gt} and ICON robustly recover the challenging poses, while others may produce artifacts like broken limbs.
3) The clothing prints can affect the surface reconstruction due to the ambiguity of geometry and appearance. ICON and PIFuHD show higher robustness.
4) HMR-SMPL-guided methods (PaMIR/ICON) generate more temporally inconsistent meshes than pure pixel-aligned methods (PIFu/PIFuHD).

Reality Check to SOTAs

Train on CLOTH4D

1) Training on CLOTH4D generally outperform the original ones trained on scan data and generate more details.
2) We input the normal image to all methods, the influence of garment print to PIFu and PaMIR are slightly relieved, validating that intermediate representations can reduce the ambiguity of geometry and appearance.
3) SOTAs fail to model layered and thin clothing structures where holes and tattered pieces are generated. The original PIFu and PaMIR can roughly generate loose clothing but fail when trained on CLOTH4D.


DETAILS
RGB
GT
PIFuHD
PIFu
PIFu$_{clo}$
PaMIR
PaMIR$_{clo}$
PaMIR^{gt}
PaMIR$^{gt}_{clo}$
ICON
ICON$_{clo}$
ICON^{gt}
ICON$^{gt}_{clo}$

Background

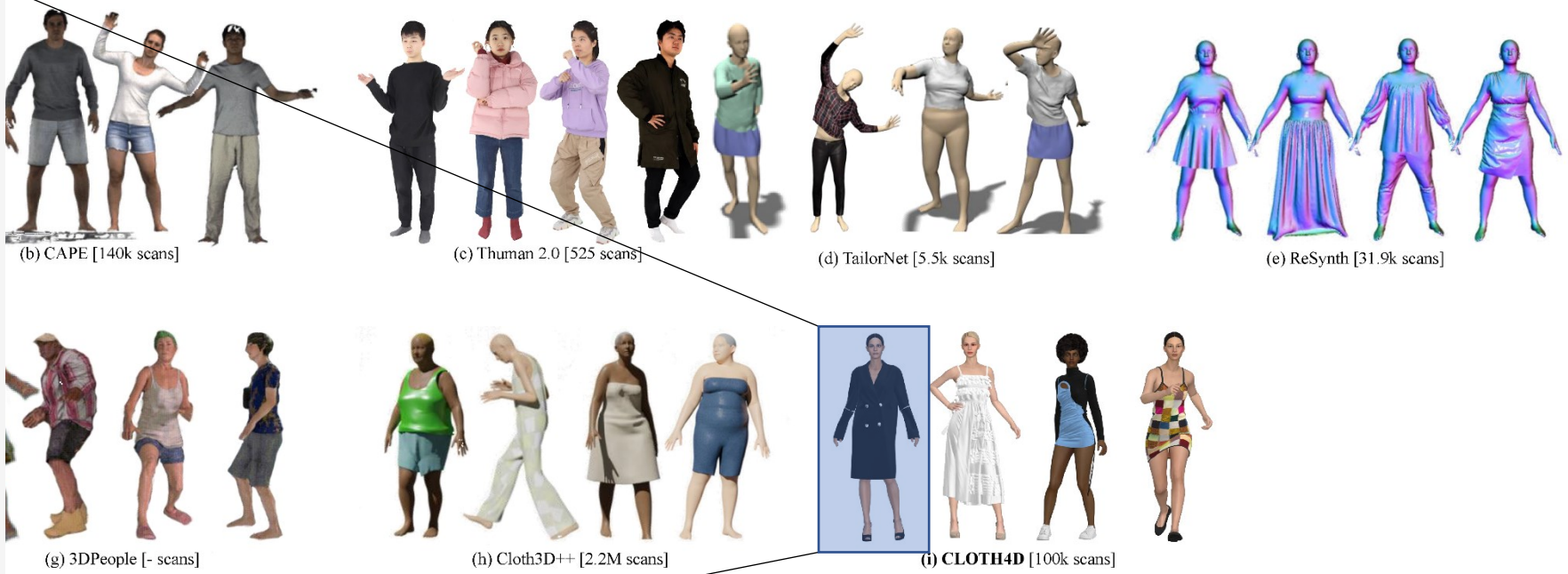
Existing Datasets for Clothed Human Reconstruction

Dataset	#Subjects	#Action	#Scan	2D Pattern	TexCloth	TexHuman	w/ SMPL	Public	Photorealistic
BUFF [52]	6	-	13.6k	-	✓	-	✓	✓	✓
RenderPeople [1]	-	-	825	-	✓	-	✓	-	✓
DeepWrinkles [30]	2	2	9.2k	-	✓	-	-	-	✓
CAPE [35]	15	600	140k	-	-	-	✓	✓	✓
THuman2.0 [51]	200	-	525	-	✓	-	✓	✓	✓
DRAPE [16]	7	23	24.5k	-	-	-	-	-	-
Wang et al. [47]	-	-	24k	✓	✓	-	✓	✓	-
3DPeople [40]	80	72	-	-	✓	-	-	-	✓
DCA [43]	-	56	7.1k	-	-	-	✓	-	-
GarNet [17]	600	-	18.8k	-	-	-	✓	✓	-
TailorNet [38]	9	-	5.5k	-	✓	-	✓	✓	-
Cloth3D [8]	8.5k	7.9k	2.1M	-	✓	-	✓	✓	-
Cloth3D++ [36]	9.7k	8k	2.2M	✓	✓	✓	✓	✓	-
CLOTH4D	1k	289	100k	✓	✓	✓	✓	✓	✓

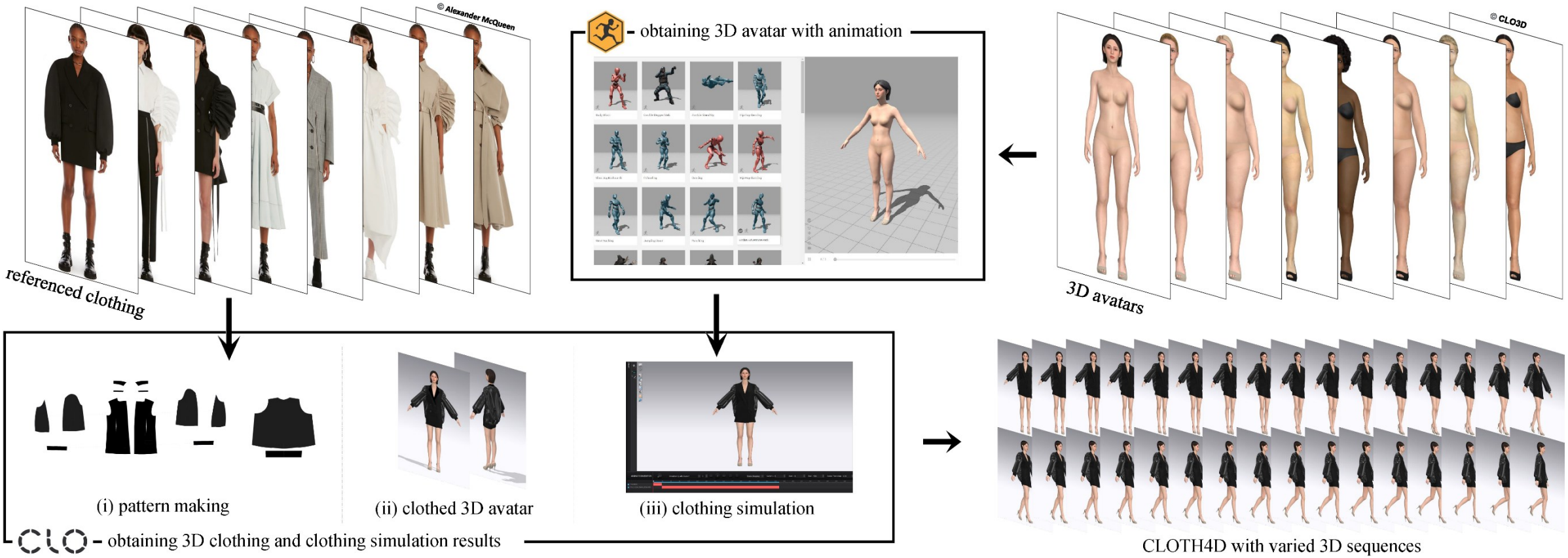
Background

Existing Datasets for Clothed Human Reconstruction

©Alexander McQueen



Data Creation



Baseline Evaluation

Table 2. Quantitative evaluation on CLOTH4D. PaMIR^{gt} and ICON^{gt} denote that the fitted ground truth SMPL is used during the inference. Gray color

Method											ICON$^{gt}_{clo}$
Normals ↓											0.103
P2S ↓											2.068
Chamfer ↓											1.367
Normals$_{ddt}$ ↓											0.035
Normals$_{dtd}$ ↓											0.033
P2S$_{ddt}$ ↓											0.367
P2S$_{dtd}$ ↓											0.526
Chamfer$_{ddt}$ ↓											0.369
Chamfer$_{dtd}$ ↓											0.525

Method											ICON$^{gt}_{clo}$
Normals ↓											0.077
P2S ↓											1.193
Chamfer ↓	4.204	3.927	4.258	1.654	3.962	**1.038**	4.381	4.080	1.427	3.544	1.397

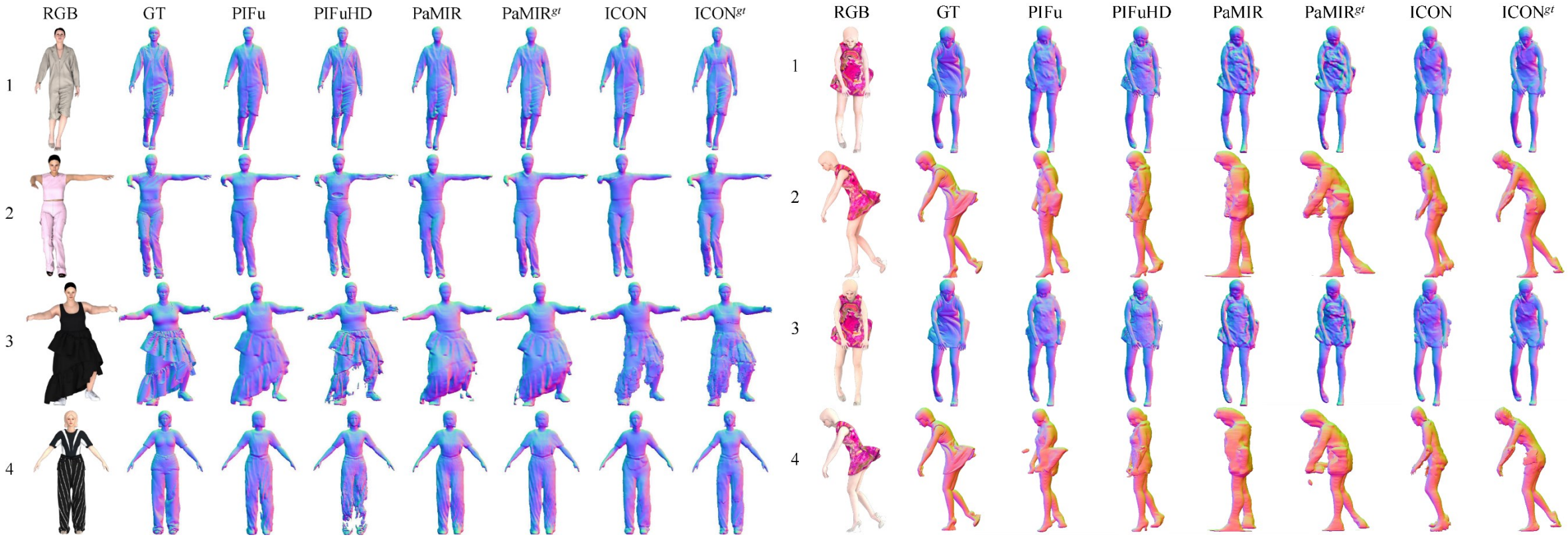
$$\text{Normals}_{ddt} = \frac{1}{T}\sum_{t}\left|\left(\mathcal{N}_{t}^{pr} - \mathcal{N}_{t}^{gt}\right)^{2} - \left(\mathcal{N}_{t+1}^{pr} - \mathcal{N}_{t+1}^{gt}\right)^{2}\right|, \quad (1)$$

$$\text{Normals}_{dtd} = \frac{1}{T}\sum_{t}\left|\left(\mathcal{N}_{t}^{pr} - \mathcal{N}_{t+1}^{pr}\right)^{2} - \left(\mathcal{N}_{t}^{gt} - \mathcal{N}_{t+1}^{gt}\right)^{2}\right|, \quad (2)$$

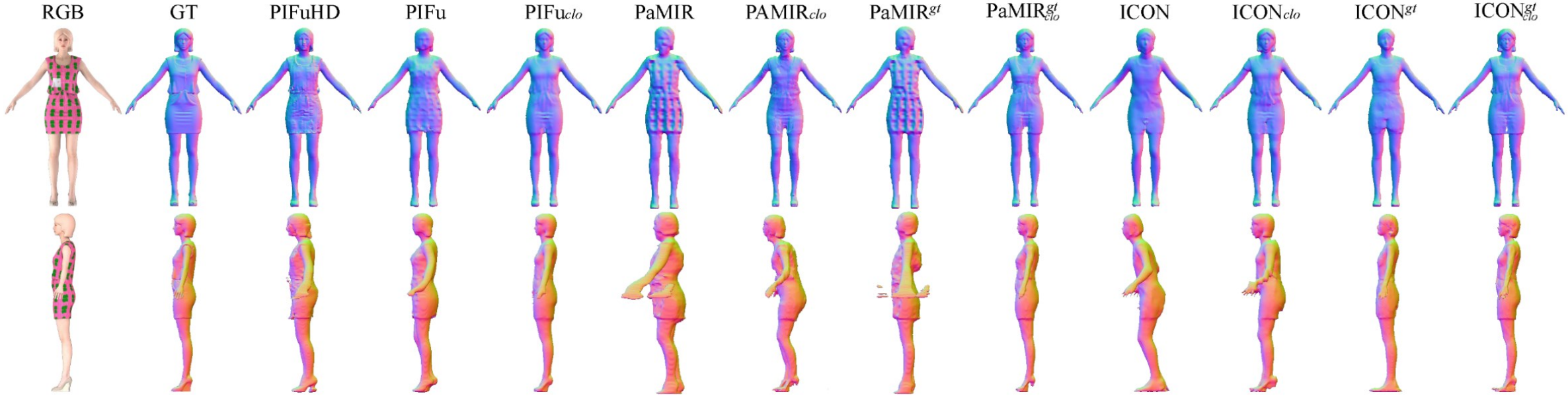
where T is the length of the sequence. N_t^{pr} and N_t^{gt} denote the rendered normal images from the predicted mesh and the ground truth mesh at time step t, respectively.

Baseline Evaluation

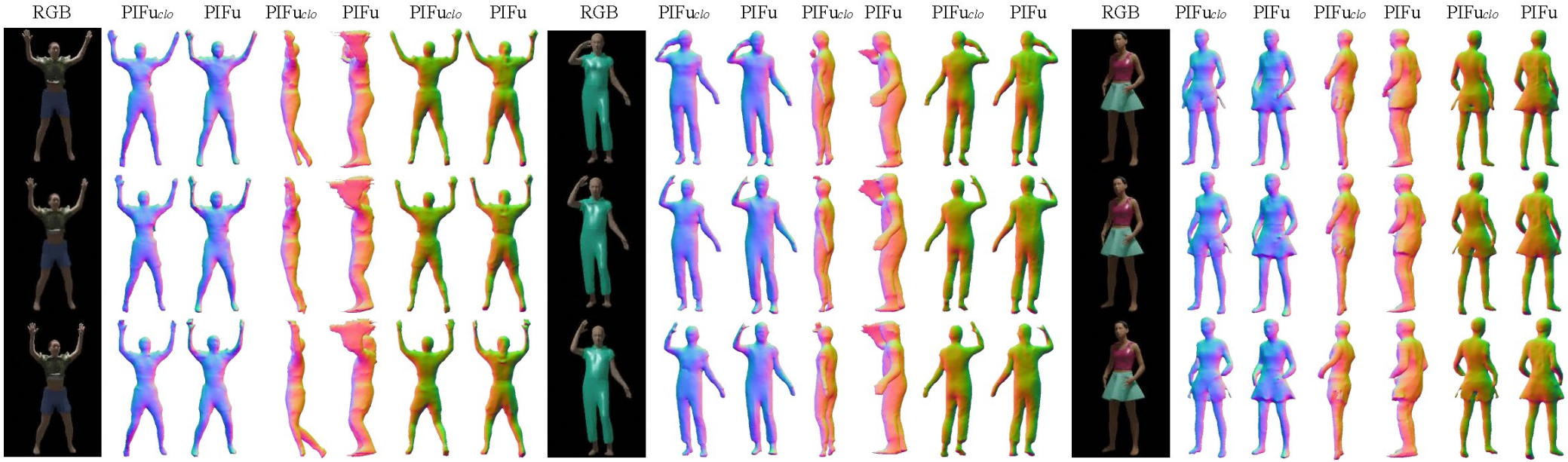
Qualitative results of baseline methods tested on CLOTH4D with front view and side view.



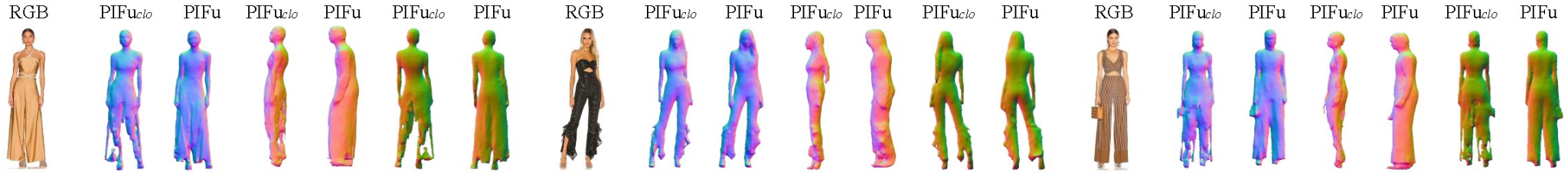
Baseline Enhancement



More Results

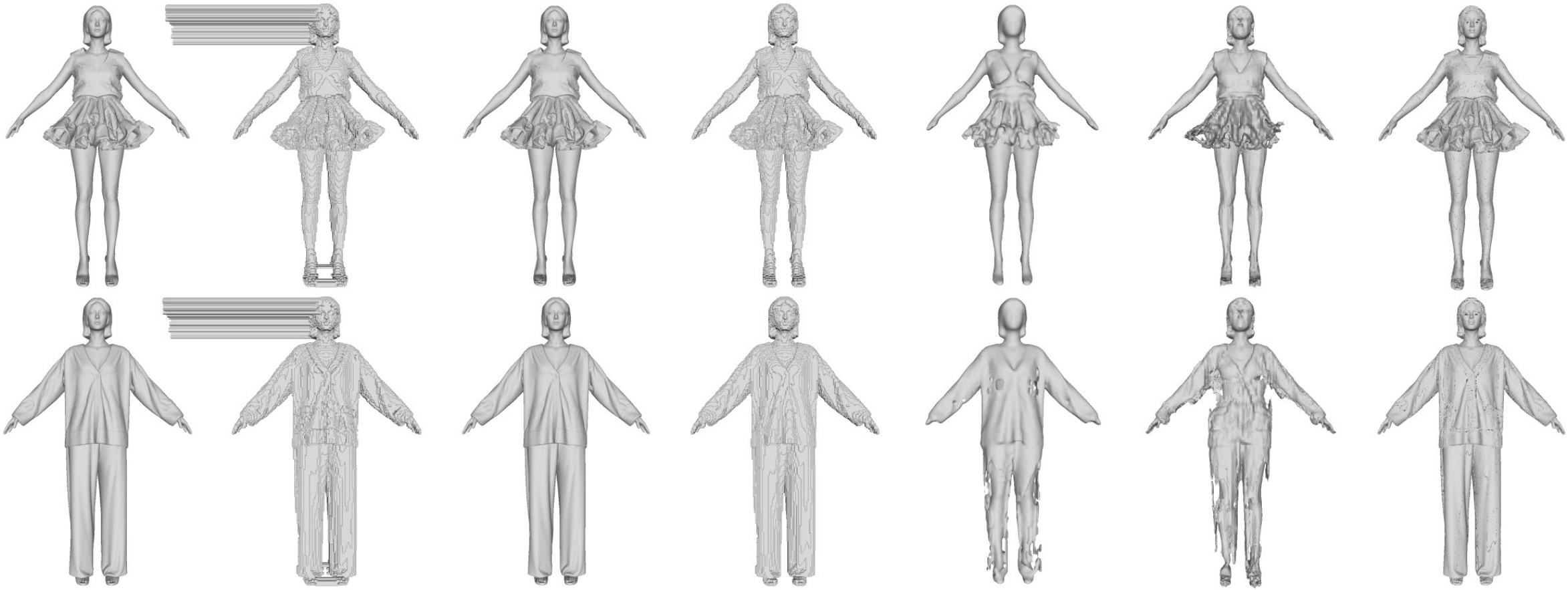


(b) Temporal qualitative results on Cloth3D++. Training on CLOTH4D results in higher temporal consistency



(c) Qualitative results on In-the-wild Images. Training on CLOTH4D reconstructs better body shapes, especially for side views

Our Trying



(a) original mesh | (b) MC of of original occ | (c) watertight mesh | (d) MC of watertight occ | (e) MC of smooth watertight occ | (f) ICON$^{gt}_{clo}$ prediction | (g) GT reconstruction of 3PSDF

Potential Directions

1) To explore better strategies for balancing the local and global reconstruction quality.
2) To jointly train or optimize body prior with mesh reconstruction.
3) To disentangle the geometry and appearance.
4) To predict more intermediate 2D/3D representations.
5) To investigate temporal modeling of implicit functions.
6) To seek implicit representations for reconstructing multi-layer thin structures.

Project Page

Data Preview

THANKS FOR WATCHING