Handy: Towards a high fidelity 3D hand shape and appearance model

Rolandos-Alexandros Potamias, Stylianos Ploumpis, Stylianos Moschoglou, Vasilios Triantafyllou, Stefanos Zafeiriou

Imperial College London

Poster 6877
TUE-PM-051

Handy: A shape and appearance model

- The largest shape/appearance hand model trained with **over 1200 subjects**.
- Diverse subject distribution from **different age (1-81 y.o.)** and **ethnicity groups**
- **Highly detailed template** (7K vertices)
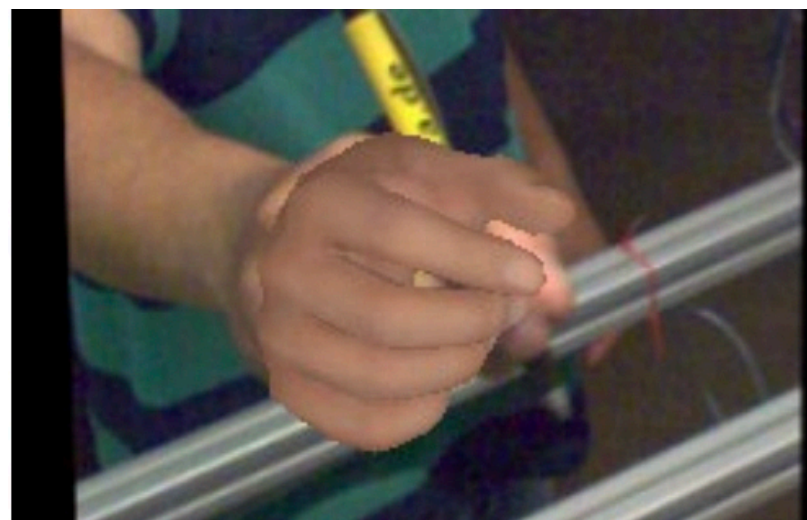- **High resolution texture model** based on StyleGAN3

Imperial College London

MANO [1]
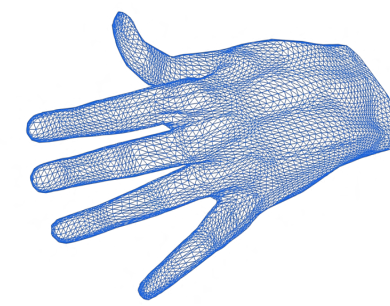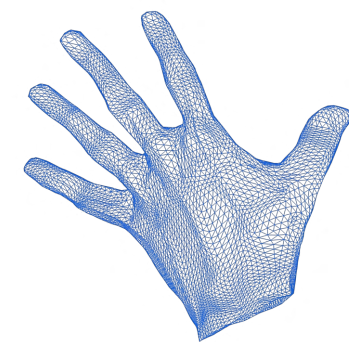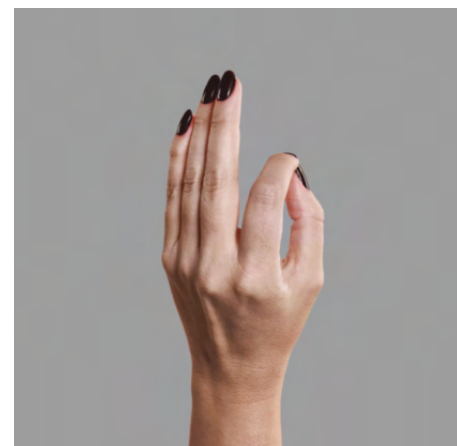
Current state-of-the art 3D hand models:

- Have **low polygon count** ⇒ Lacking high resolution.

- Trained on a **few subjects (up to 50)** ⇒ Do not match the population distribution.

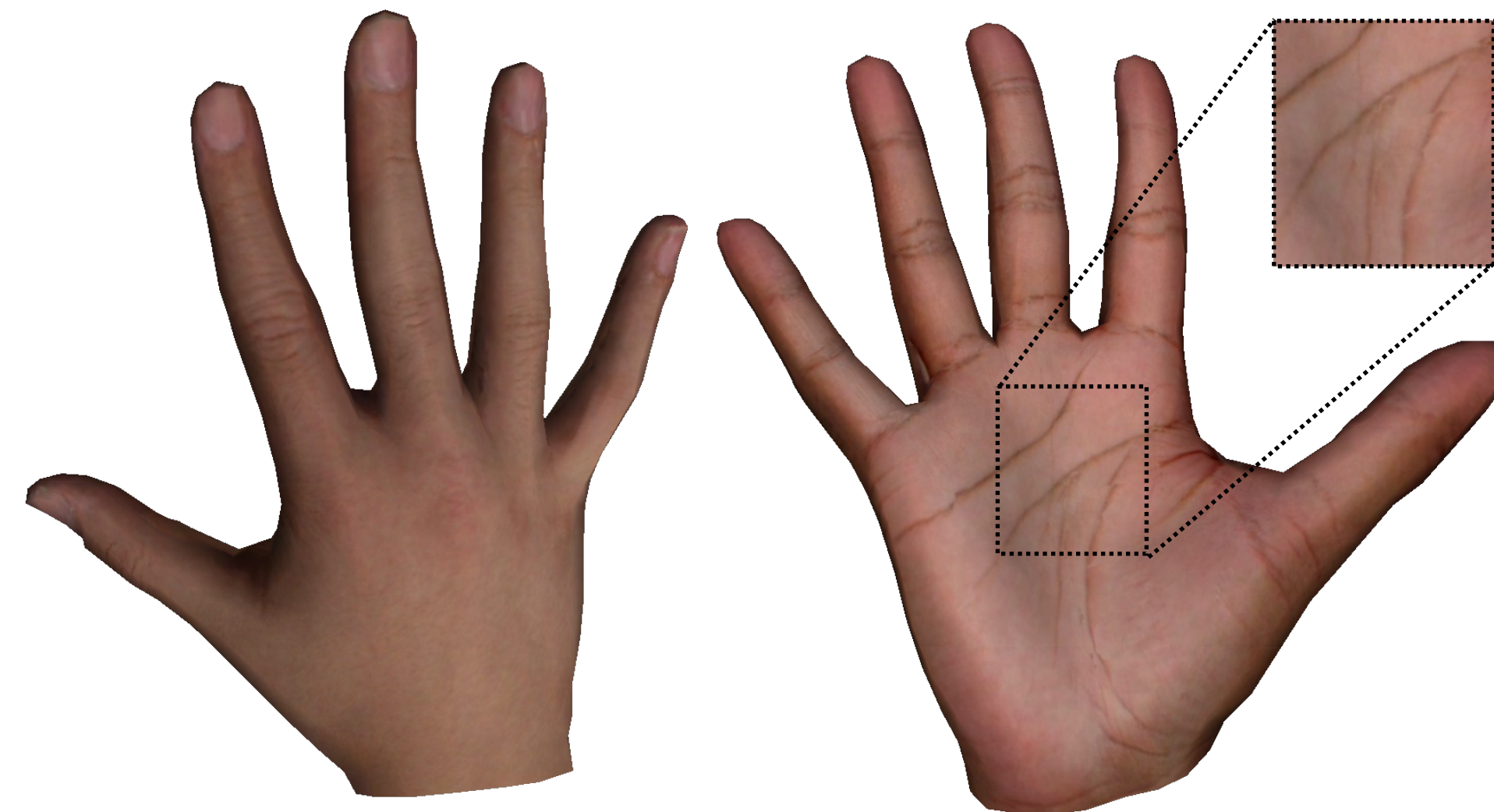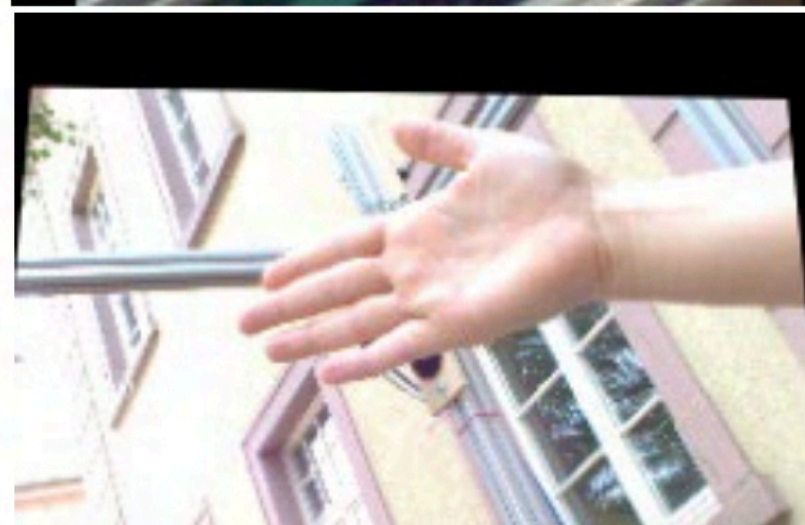- **Lack high resolution** texture ⇒ Limits reconstruction / generation of realistic hands
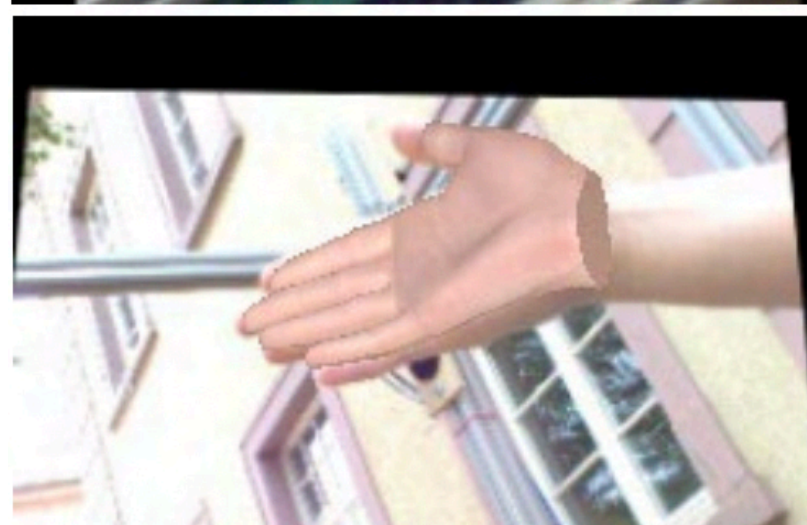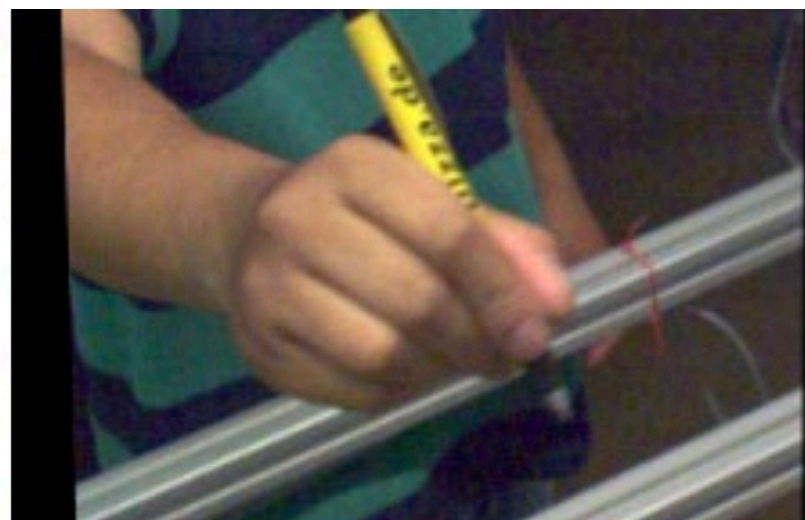
HTML [2]   Ground Truth

[1]: J. Romero et al., Embodied Hands: Modeling and Capturing Hands and Bodies Together, ACM ToG, 2017
[2]: N. Qian el al., HTML: A Parametric Hand Texture Model for 3D Hand Reconstruction and Personalization, ECCV, 2020

Imperial College London

MANO [1]

Handy

- **High quality template** that imitates human hand shape.
- **Diverse human hand shape** model.
- **High resolution texture** model.

➡ Reconstruction of 'in-the-wild' hands.

➡ Photorealistic synthetic datasets.
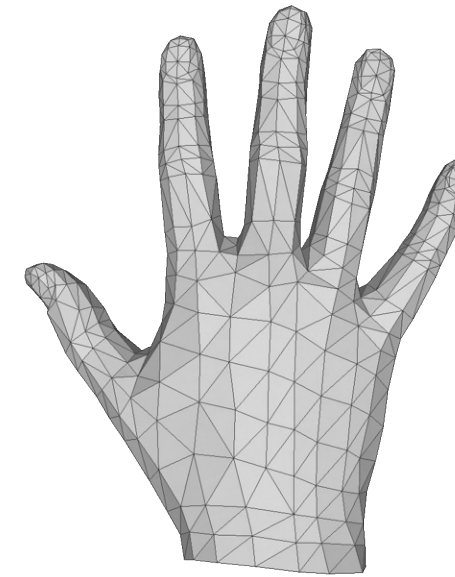
HTML [2]

Handy

[1]: J. Romero et al., Embodied Hands: Modeling and Capturing Hands and Bodies Together, ACM ToG, 2017

[2]: N. Qian el al., HTML: A Parametric Hand Texture Model for 3D Hand Reconstruction and Personalization, ECCV, 2020
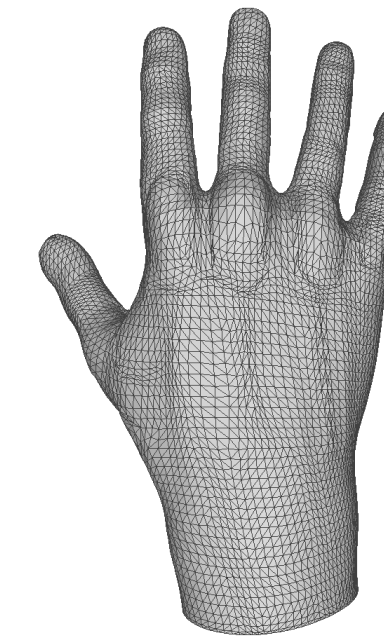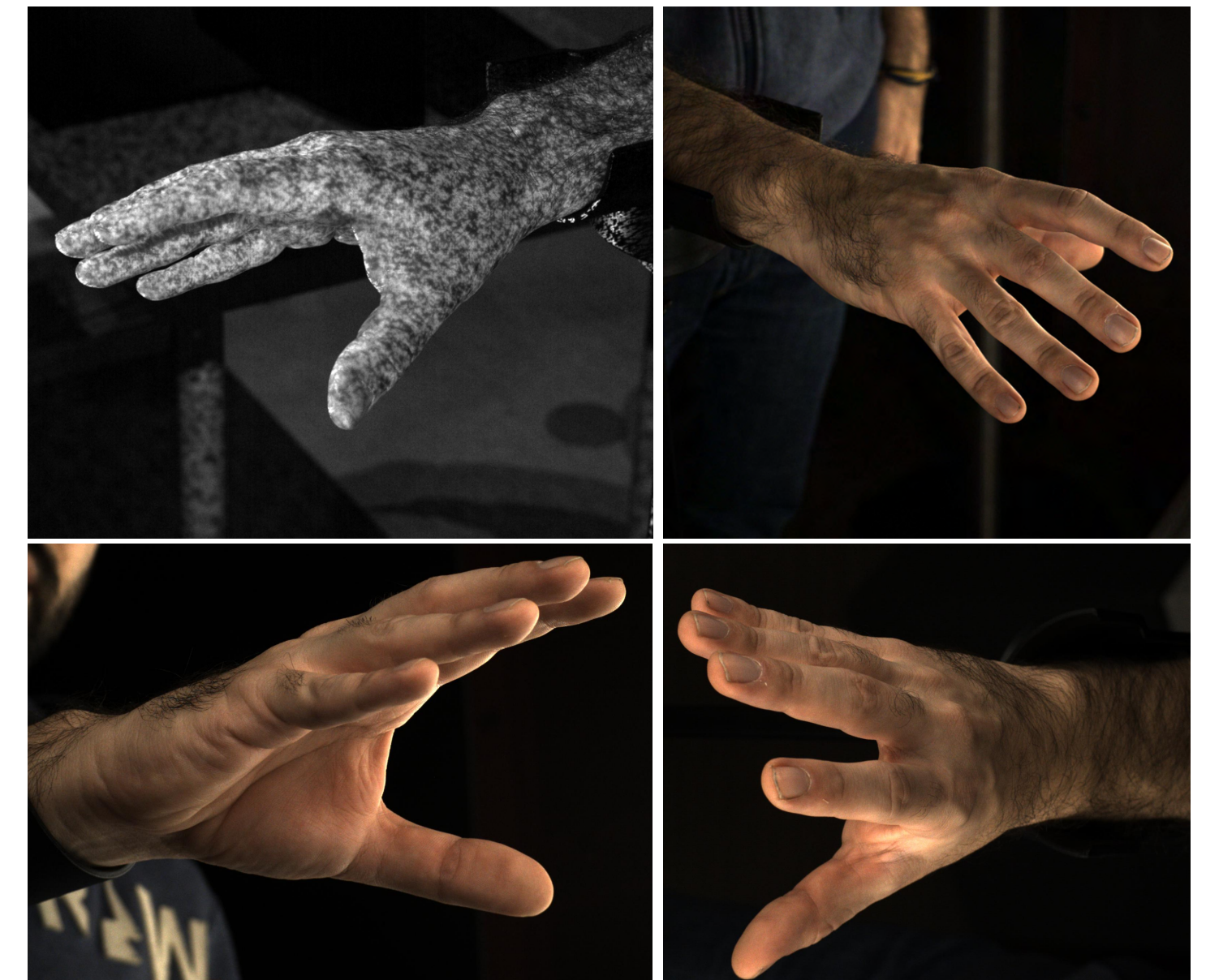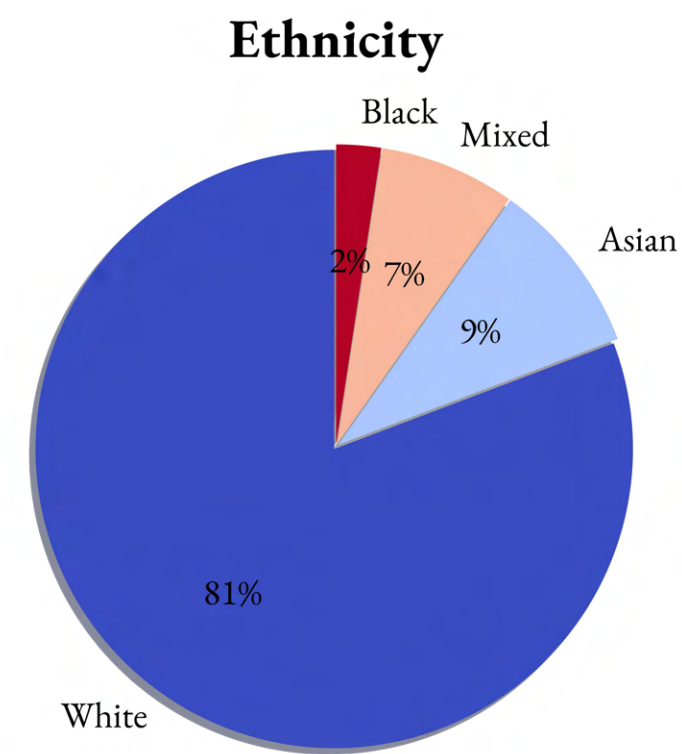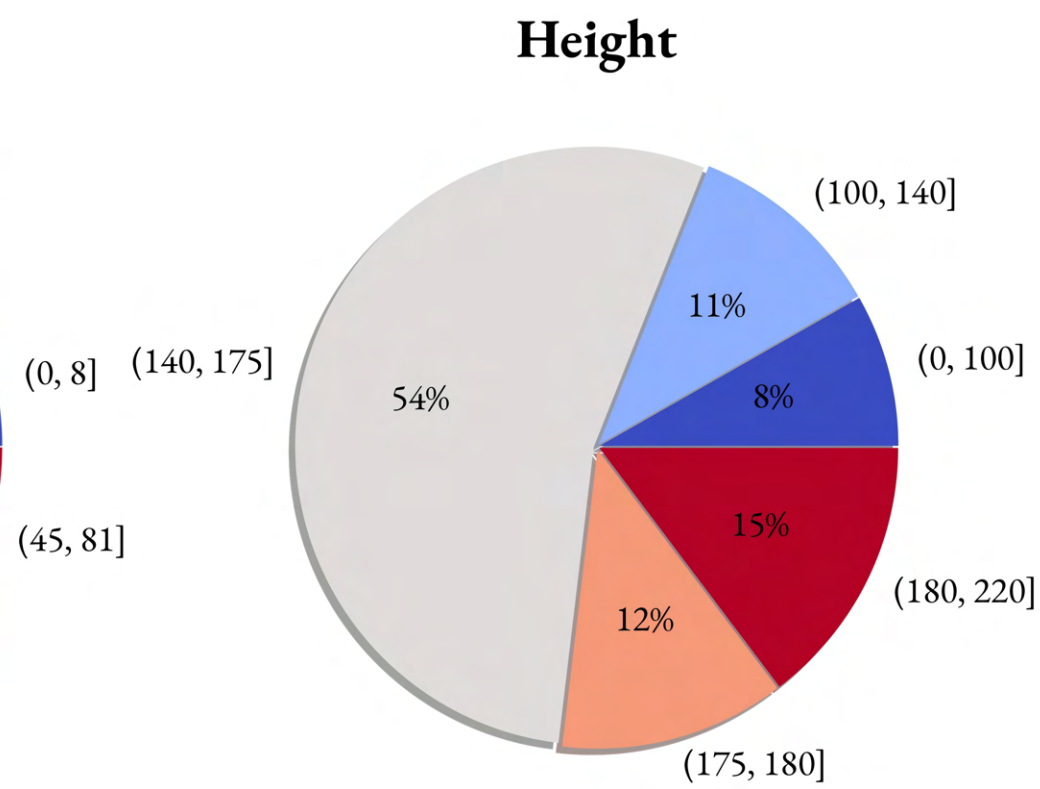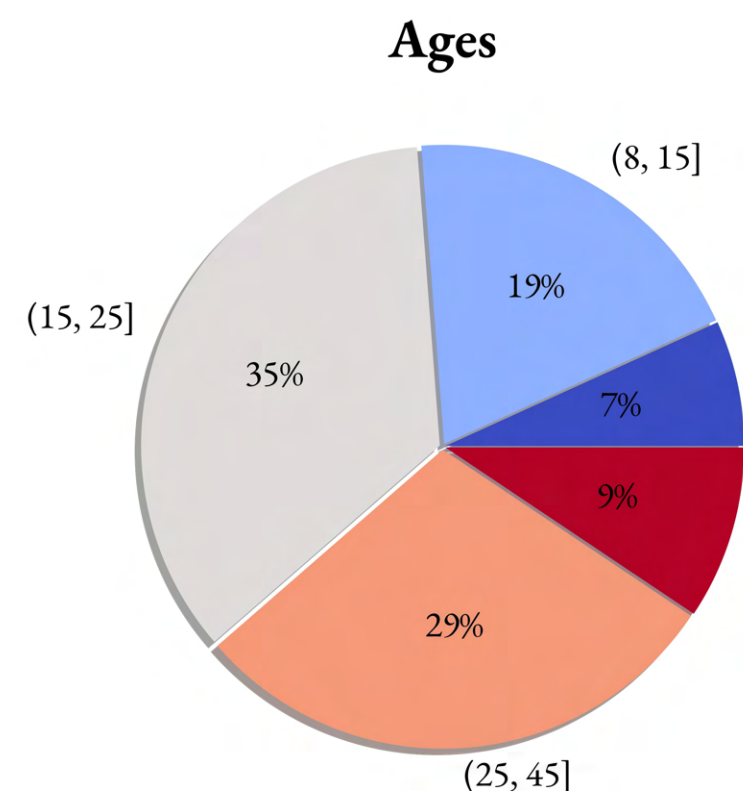
Imperial College London

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

- 3dMD stereo scanning device with structured light.

- High quality raw scans (approximately 30,000 vertices)

- High resolution textures.

- **Largest available hand dataset** composed by **1208** subjects.

- **Large diversity** on ethnicity, age and heights.

Ages
- (8, 15] — 19%
- (15, 25] — 35%
- (0, 8] — 7%
- (45, 81] — 9%
- (25, 45] — 29%

Height
- (100, 140] — 11%
- (140, 175] — 54%
- (0, 100] — 8%
- (180, 220] — 15%
- (175, 180] — 12%

Ethnicity
- Black — 2%
- Mixed — 7%
- Asian — 9%
- White — 81%

Imperial College London

- **Bring scans into dense correspondence.**
  - ➡ Highly detailed artist-curated template (7K vertices)
  - ➡ Render scans from multiple views and detect 2D landmarks using Mediapipe [3].
  - ➡ Lift 2D landmarks to 3D, using linear triangulation.
  - ➡ Non-rigid ICP between the template and the raw scan.
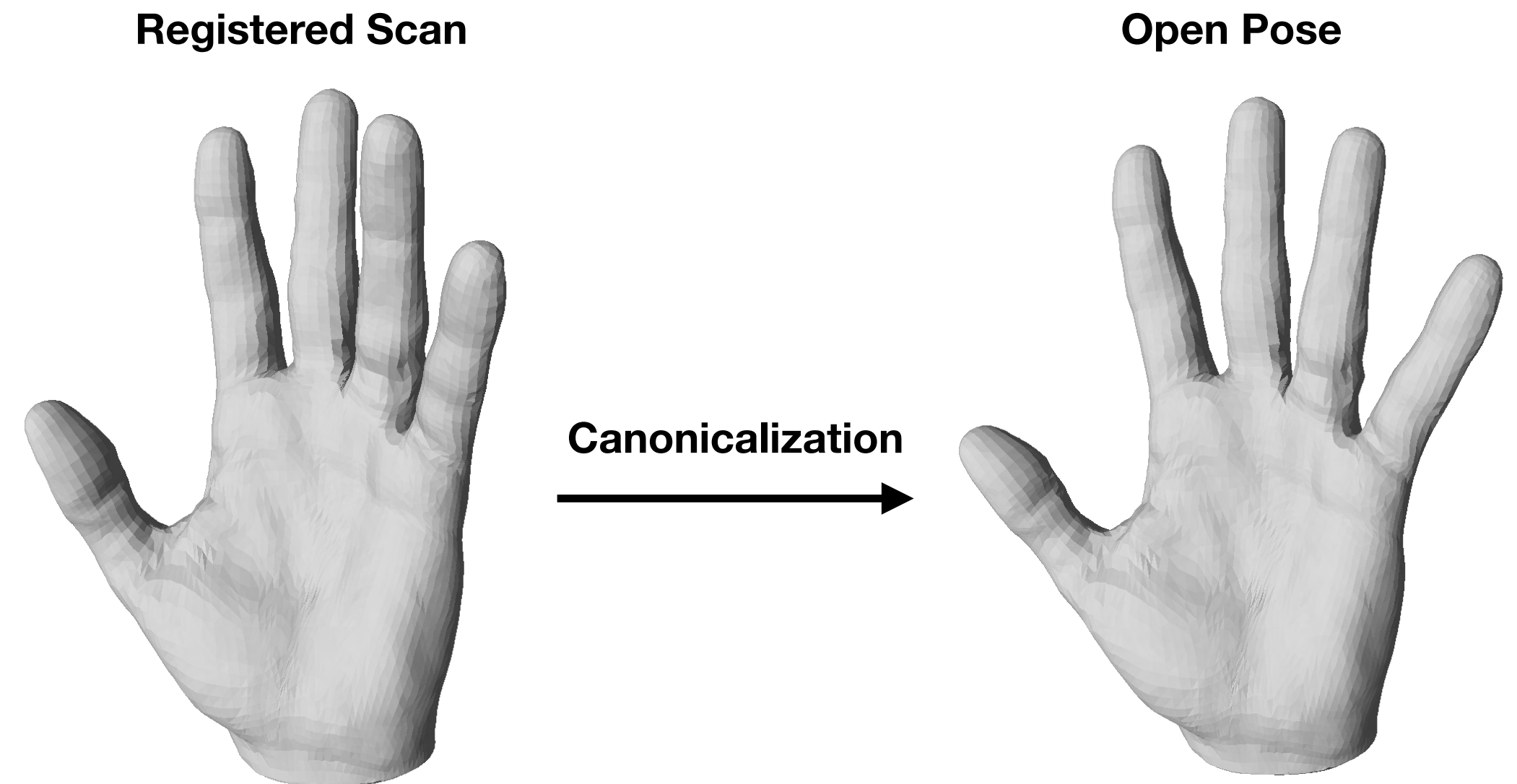 - ➡ Transfer raw scan textures to the template UV map.

Triangulation

Non-Rigid ICP

[1]: F. Zhang et al., MediaPipe Hands: On-device Real-time Hand Tracking, CVPRw, 2020
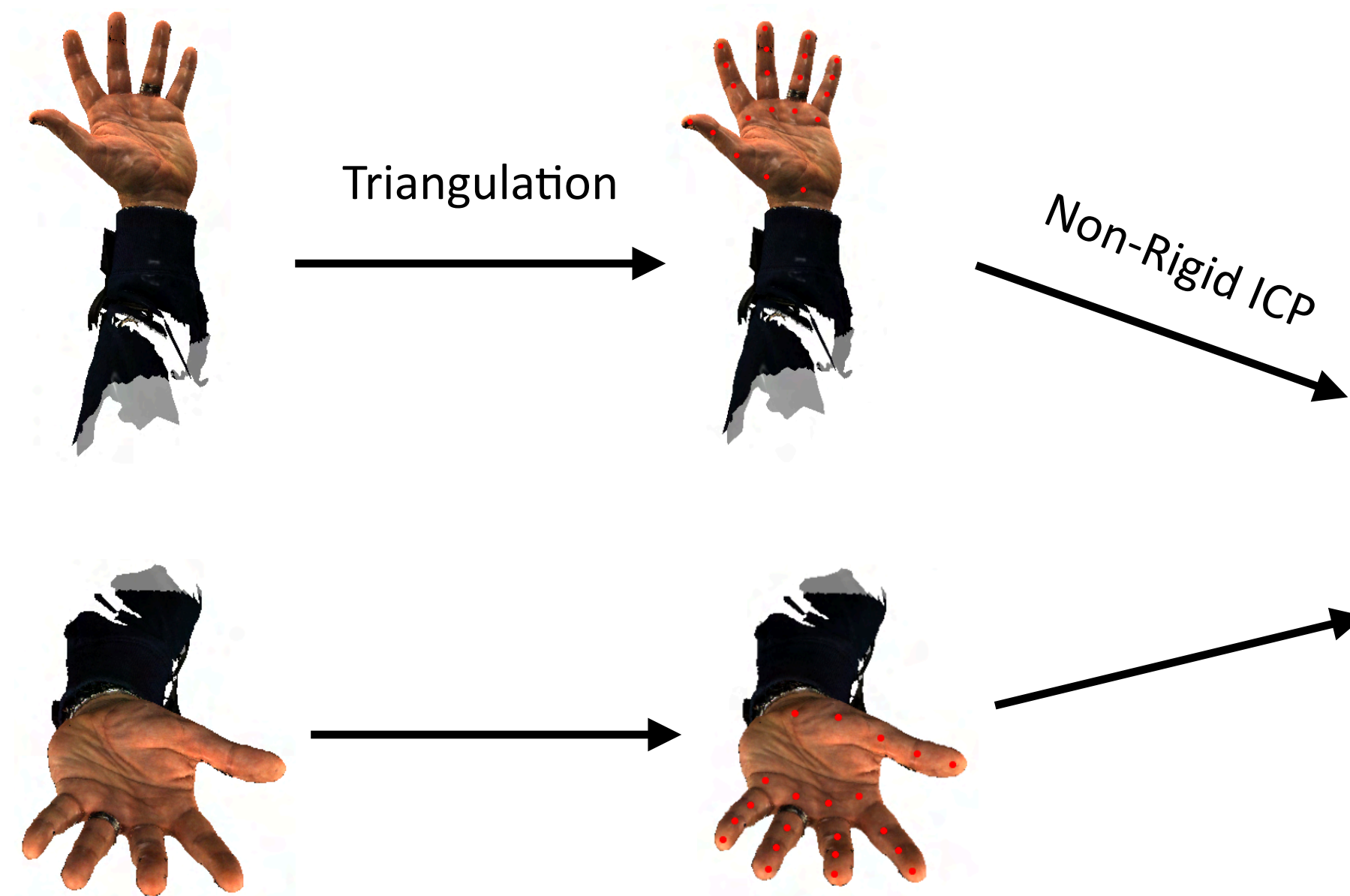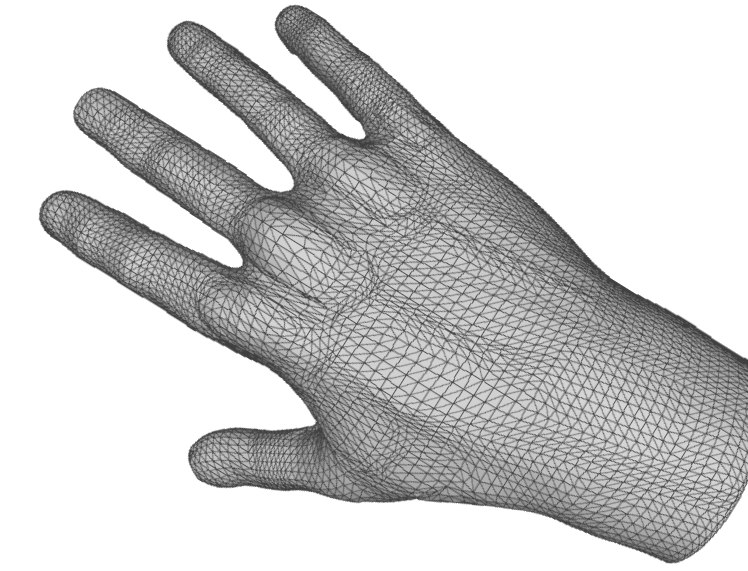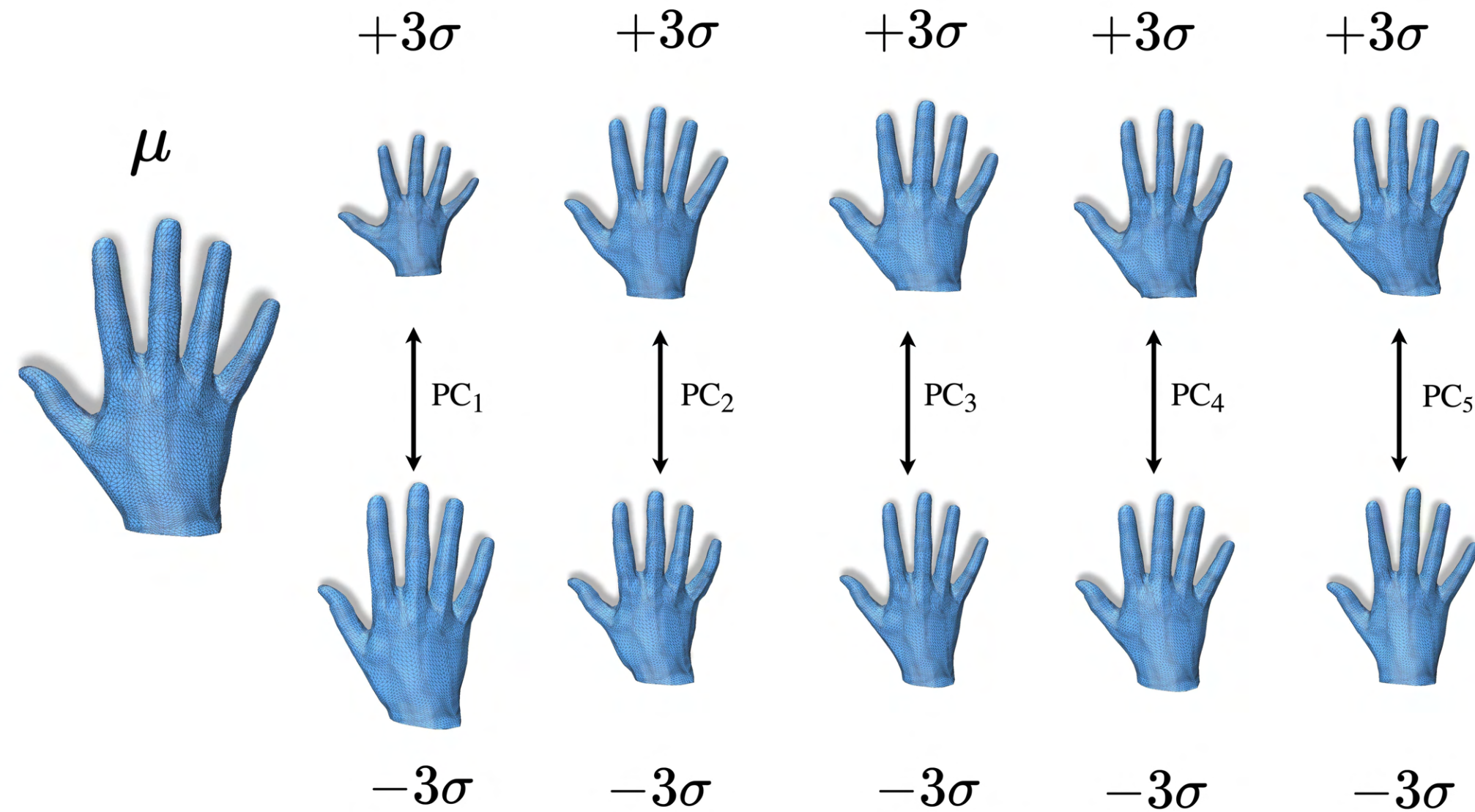
- **Shape Model**
  - ➡ Normalize all shapes to canonical open-palm pose.
  - ➡ Deformable hand model described as a linear basis of shapes:

$$B_s(\boldsymbol{\beta}) = \mathbf{T} + \sum_{i=0}^{n_c} \mathbf{U_i}\beta_\mathbf{i} \in \mathbb{R}^{3N}$$
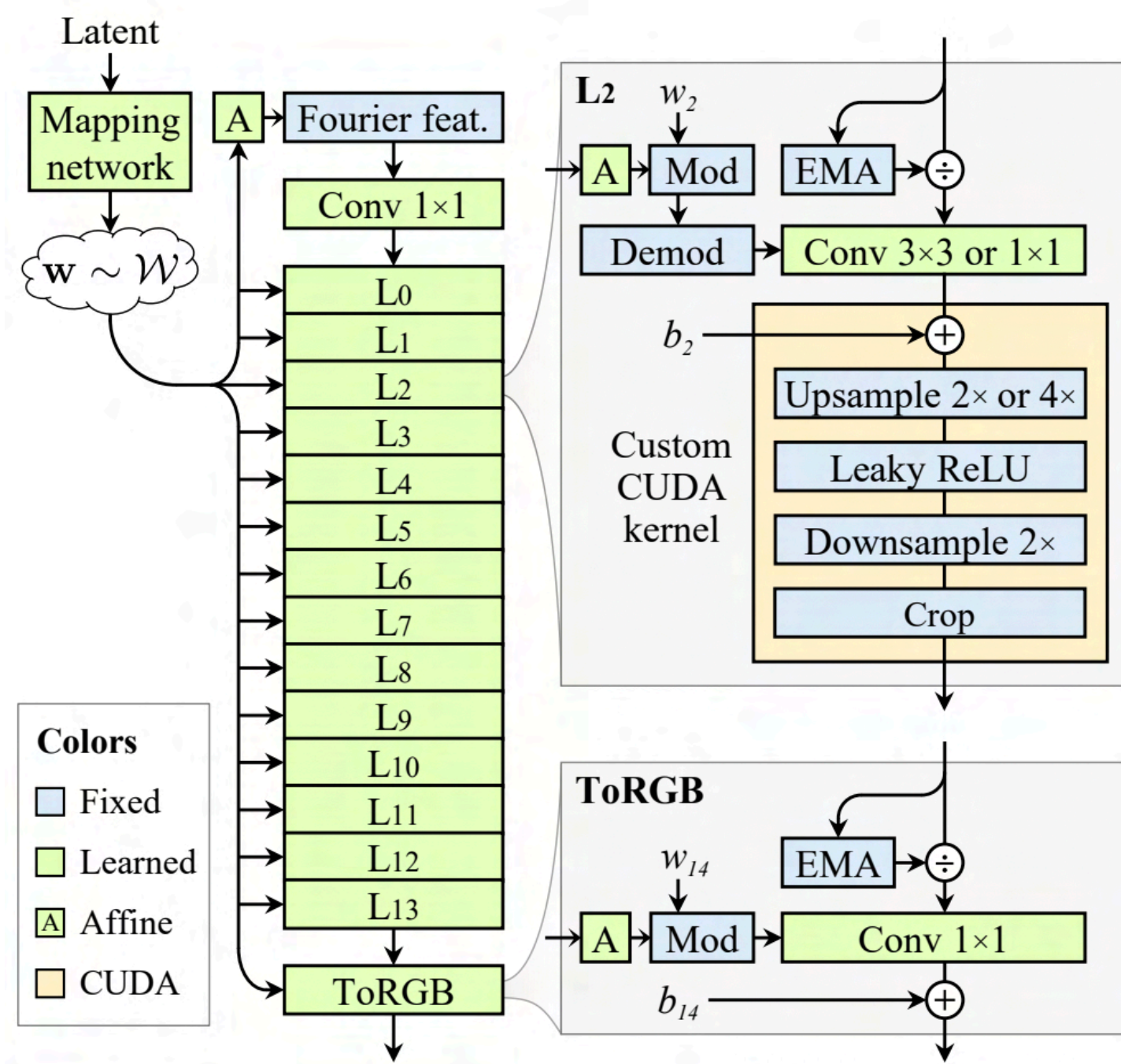
  - ➡ Retain only the first n_c components of orthonormal bases $\mathsf{U} \in \mathbb{R}^{3N \times n_c}$

- **Texture Model**
  - ➡ Utilized a powerful style-based GAN to model the UV textures.
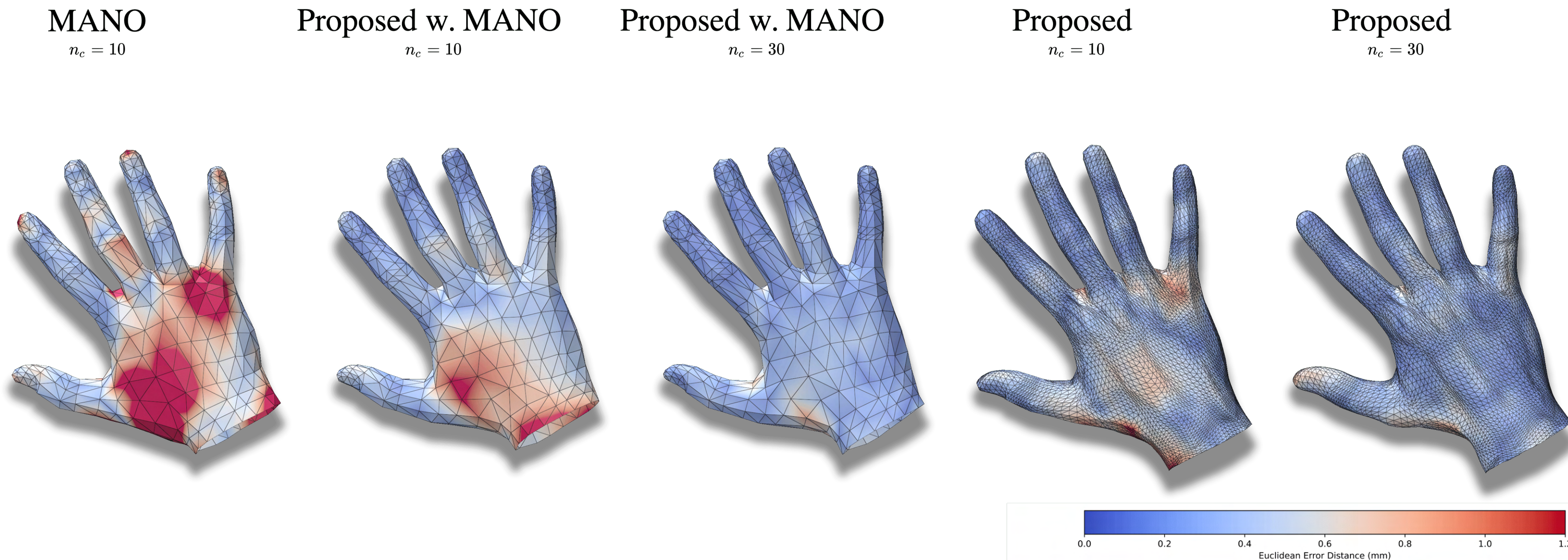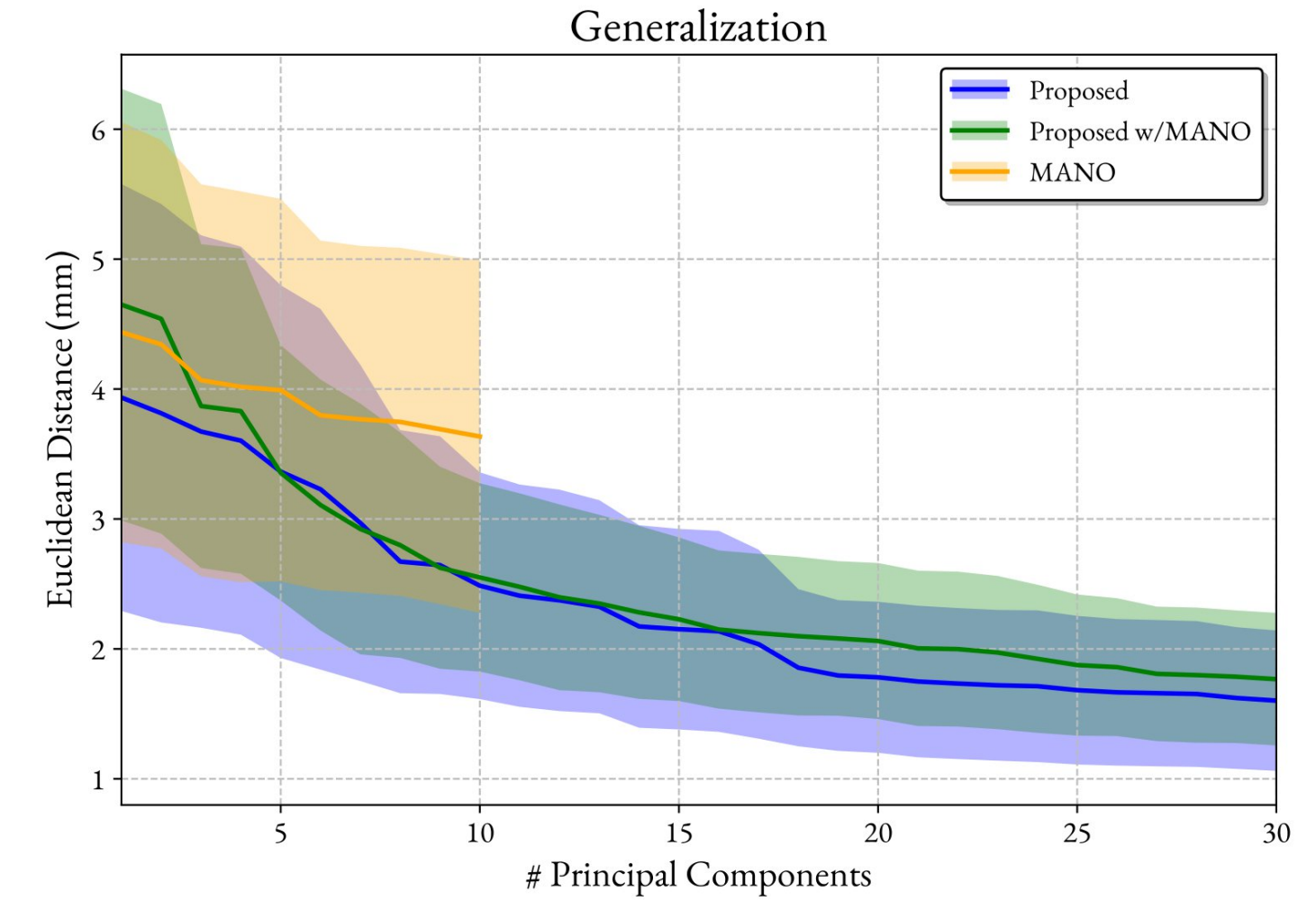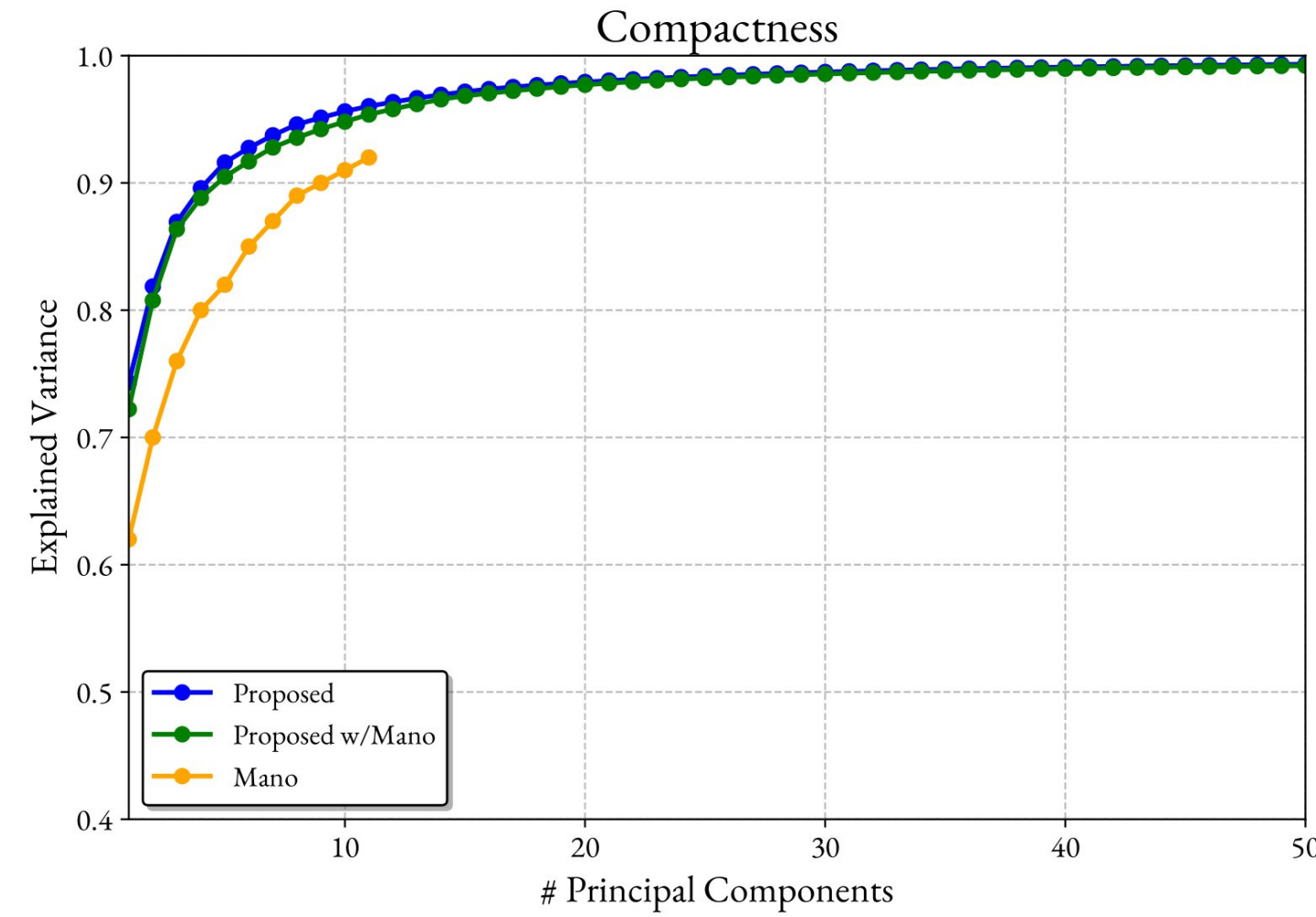
StyleGAN3

Generated Texture Samples

➡ More compact than MANO.

➡ Handy can generalize better in unseen shapes.

The case of child hand reconstruction.

➡ MANO fails to model out-of-distribution children hands.

➡ Handy achieves state-of-the-art performance.

MANO
$n_c = 10$

Proposed w. MANO
$n_c = 10$

Proposed w. MANO
$n_c = 30$

Proposed
$n_c = 10$

Proposed
$n_c = 30$

Generate a synthetic dataset for 3D hand reconstruction
- ➡ Diverse hand shapes and textures from Handy model
- ➡ Different poses from GraspIT and objects from Obman.
- ➡ Render under different lights and illuminations.
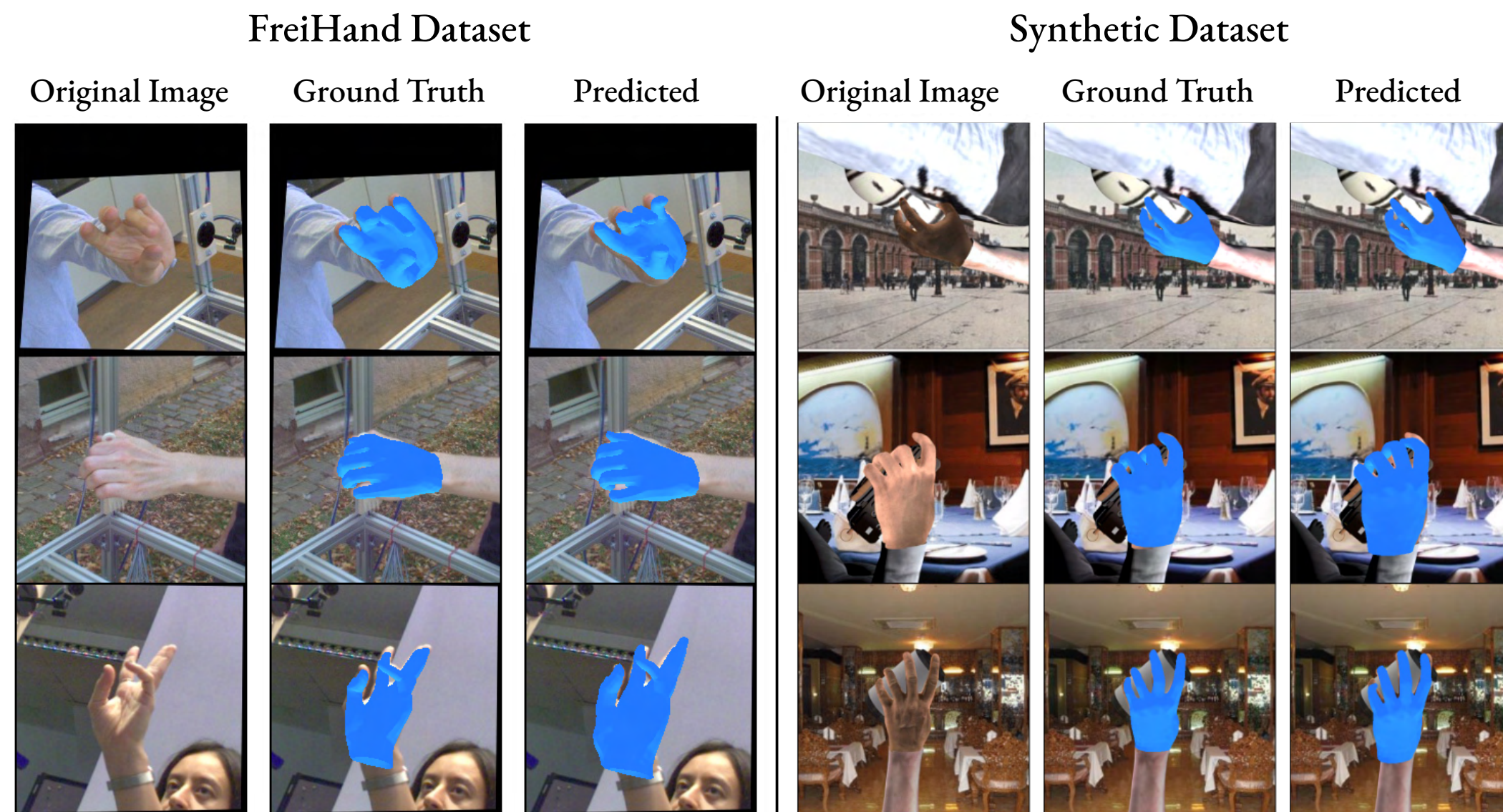
Train a hand reconstruction method that:
- ➡ Regresses pose and shape parameters of Handy.
- ➡ Regresses intermediate latent space $\mathcal{W}$ of Texture Model.

**Imperial College London**
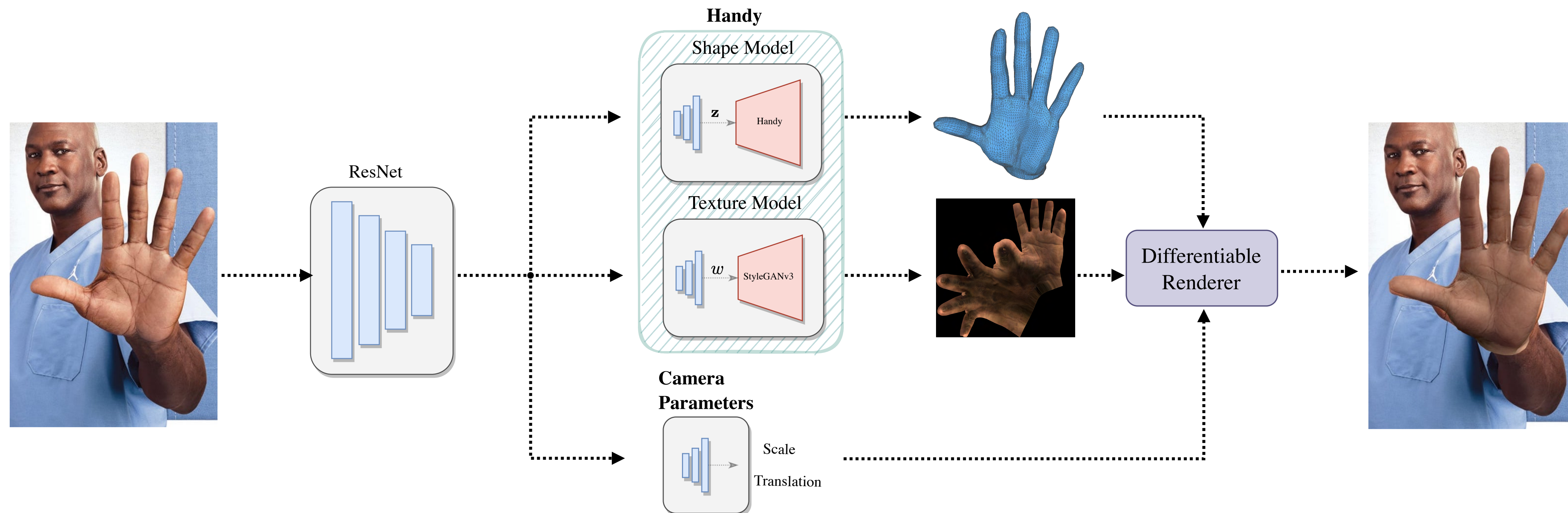
Evaluation on FreiHand dataset.

➡ Using Handy improves reconstruction compared to MANO-based reconstruction methods.

➡ Pre-training with the synthetic dataset further improves performance.

Method	MPVPE ↓	MPJPE ↓	F@5 mm ↑	F@15 mm ↑
Hasson *et al.* [16]	13.2	-	0.436	0.908
Boukhayma *et al.* [3]	13.	-	0.435	0.898
MANO CNN [55]	10.8		0.529	0.935
MANO FIT [55]	13.7	-	0.439	0.892
HTML [37]	11.1	11.0	0.508	0.930
S²Hand [6]	11.8	11.9	0.48	0.920
Ren *et al.* [38]	8.1	8.0	0.649	0.966
Proposed w/Obman	9.9	9.7	0.572	0.922
Proposed w/Synthetic	8.8	8.7	0.612	0.952
Proposed	**7.8**	**7.8**	**0.654**	**0.971**

FreiHand Dataset — Original Image | Ground Truth | Predicted

Synthetic Dataset — Original Image | Ground Truth | Predicted

Imperial College London

Reconstruction from in-the-wild images.

➡ Both shape and texture.

➡ Handy can reconstruct texture details such as veins and colour discontinuities.
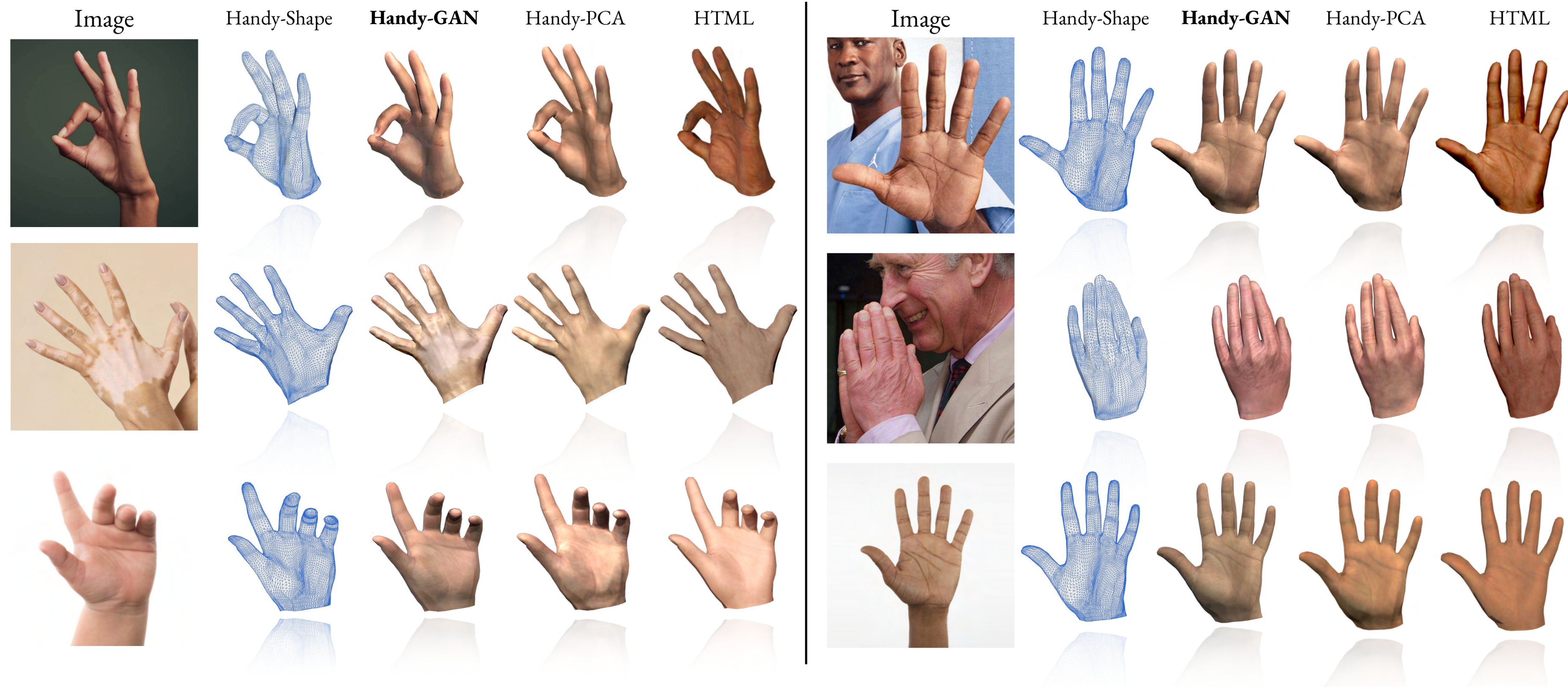
Thank you for your attention !

Models Available !!

Poster 6877
TUE-PM-051