The Power of Points for Modeling Humans in Clothing

Qianli Ma[1,2], Jinlong Yang[1], Siyu Tang[2], Michael J. Black[1]

[1]Max Planck Institute for Intelligent Systems, Tübingen   [2]ETH Zürich

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS   ETH zürich   qianlim.github.io/POP
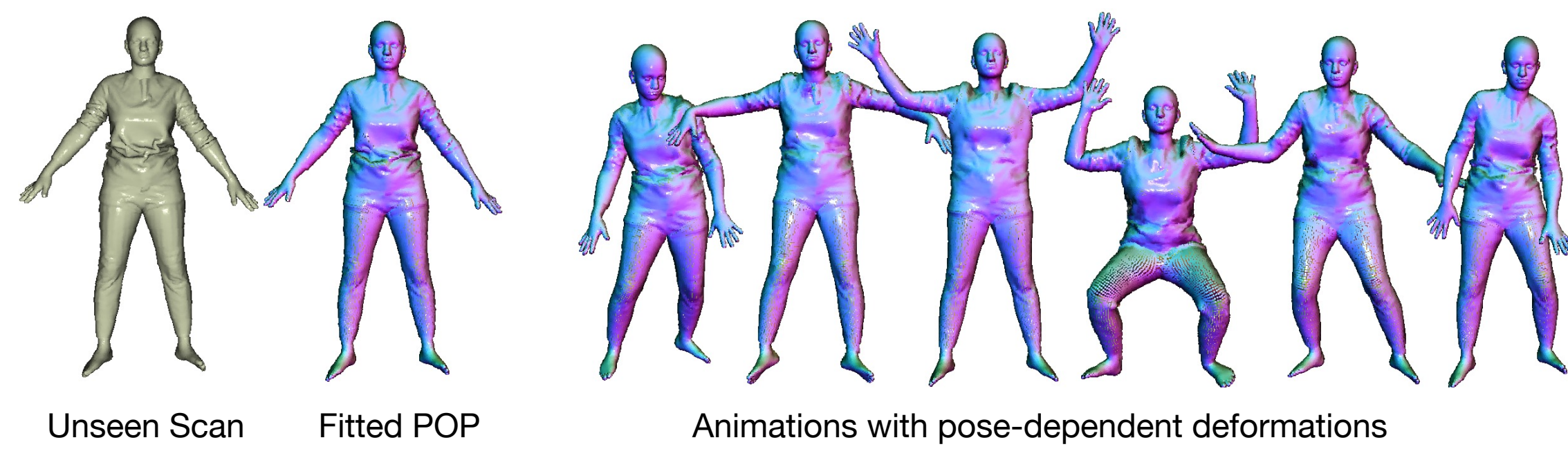
Goal

A pose-dependent shape model of clothed humans that generalizes across multiple subjects and outfits, which can create an animatable avatar from a single static 3D scan.

Unseen Scan   Fitted POP   Animations with pose-dependent deformations
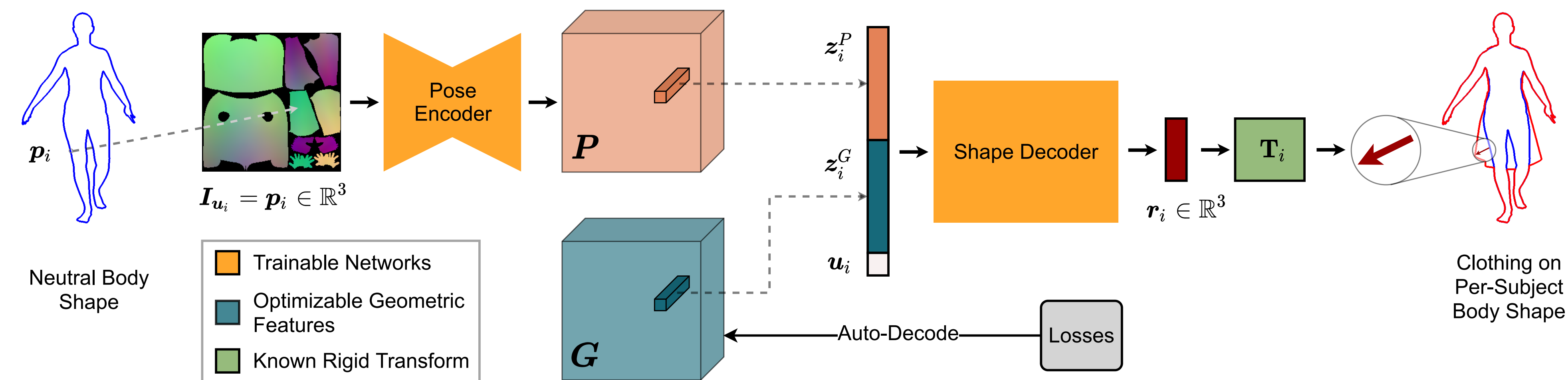
Problem

- Existing models for clothed humans are mostly subject-specific and cannot generalize to unseen outfits.
- Existing 3D shape representations cannot satisfy the need for high-quality cross-outfit modeling:
  - Meshes: fixed topology;
  - Implicit surfaces: slow inference, incompatibility with thin cloth structures;
  - Surface patches: discontinuity between patches.

References

[1] Groueix et al. 3D-CODED: 3D Correspondences by Deep Deformation. ECCV 2018.

[2] Ma et al. SCALE: Modeling Clothed Humans with a Surface Codec of Articulated Local Elements. CVPR 2021.

[3] Ma et al. Learning to Dress 3D People in Generative Clothing. CVPR 2020.

[4] Deng et al. Neural Articulated Shape Approximation. ECCV 2020.

Our Approach: POP

- Represent 3D clothed humans as *dense* point clouds, decoded from *continuous bilinear local features*.
- Model the common properties of pose-dependent deformation with a *cross-outfit shape decoder*.
- Decouple the intrinsic, pose-*in*dependent clothed body shape using a *geometric feature tensor*.

Neutral Body Shape $I_{u_i} = p_i \in \mathbb{R}^3$

- Trainable Networks
- Optimizable Geometric Features
- Known Rigid Transform

Pose Encoder P z_i^P Shape Decoder z_i^G $r_i \in \mathbb{R}^3$ T_i Clothing on Per-Subject Body Shape

u_i G Auto-Decode   Losses

Continuous Local Point Features

Point cloud decoded from:

Single global feature [1]   Per-patch features [2]   Continuous per-point features (ours)

Lack Details   **"Patchy" Artifacts**   **Clean overall shape, Sharp local details**
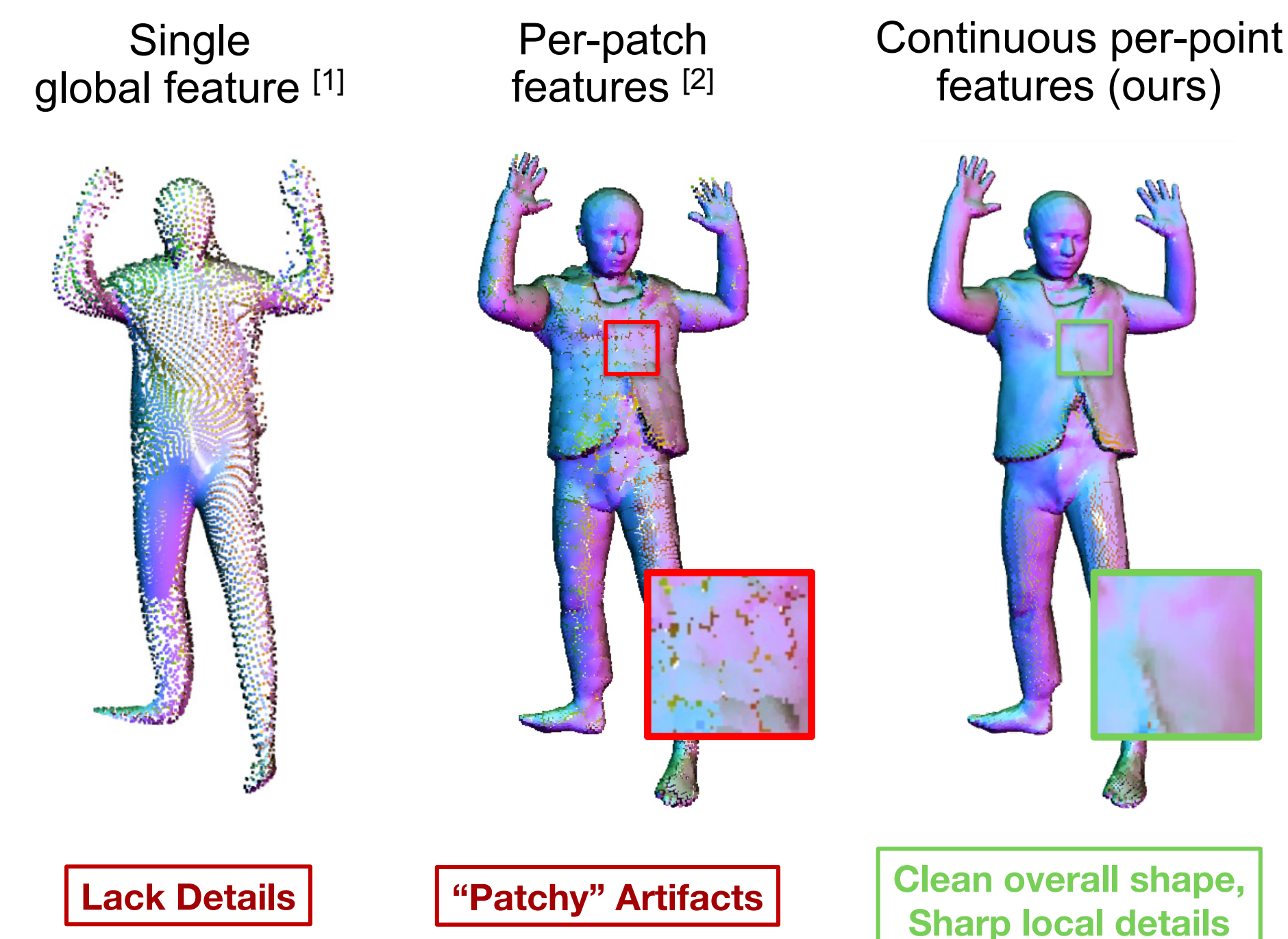
Illustration: per-patch features [2] vs bilinear features (ours)

□ Body basis point
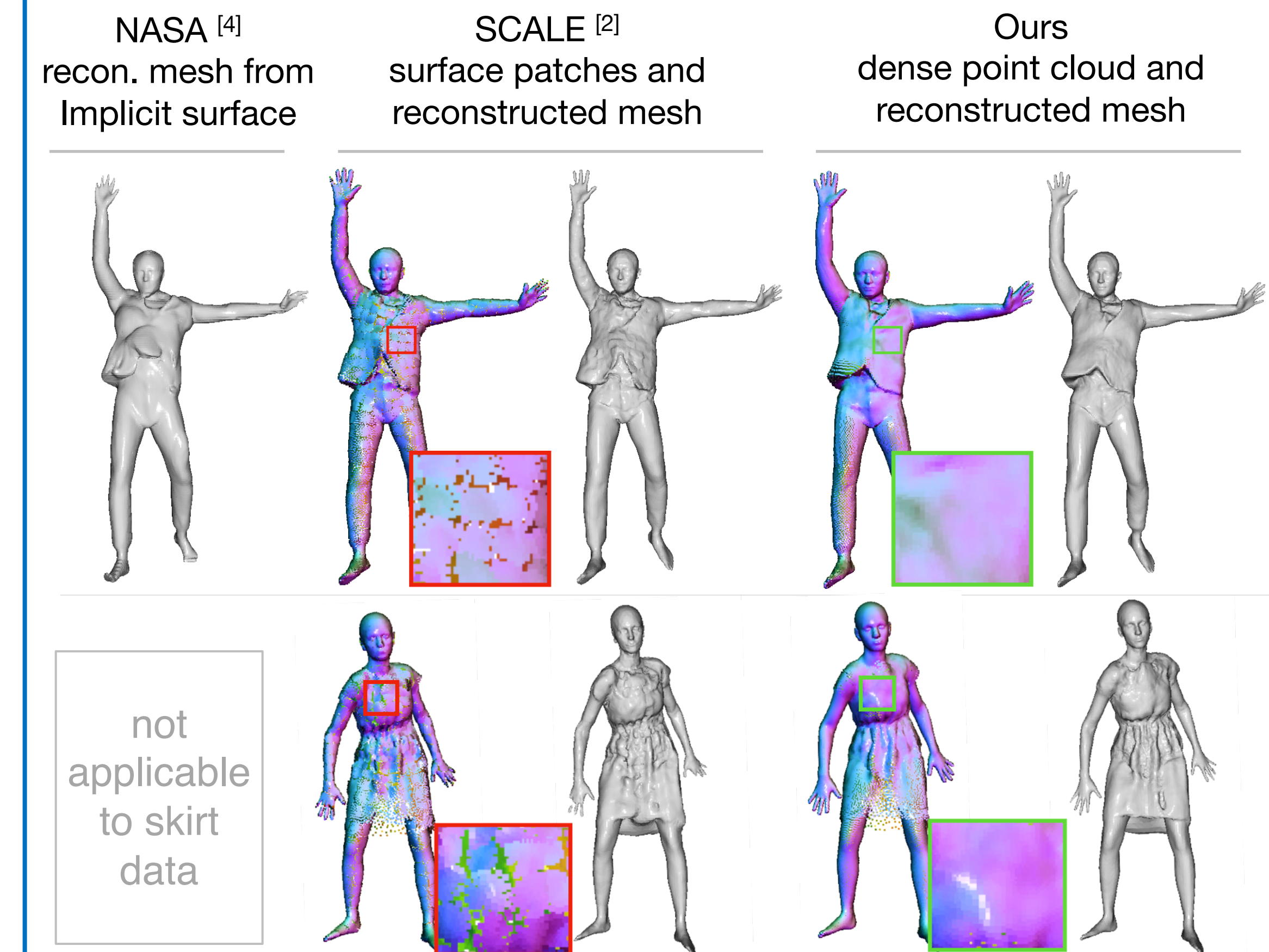○ Query point
→ Nearest assignment
⇢ Bilinear interpolation
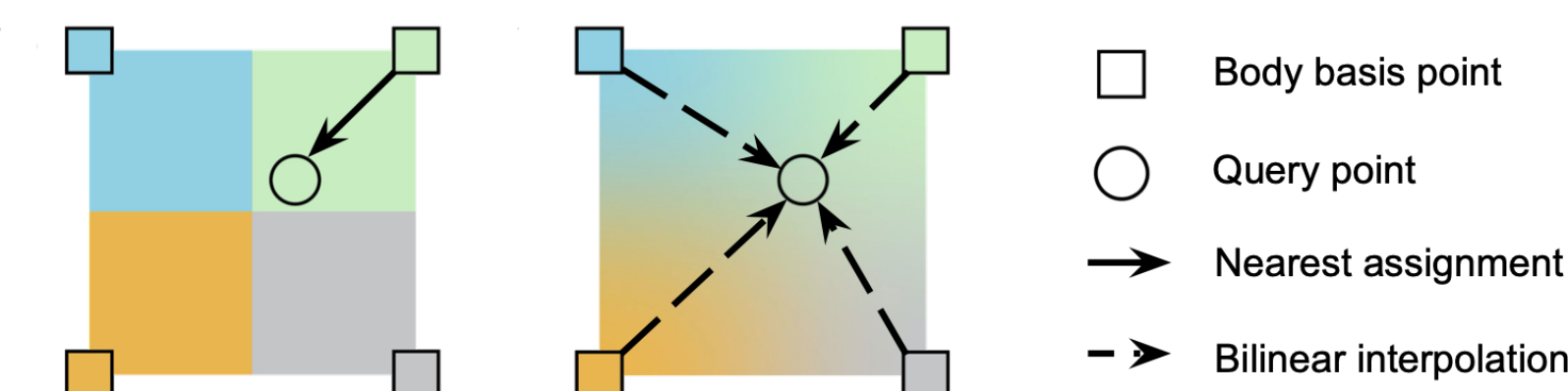
Geometric Feature Tensor

- Training: auto-decoding. Enforce a consistent instance for each outfit across all poses.
- Inference: optimize the tensor w. r. t. an unseen scan to get a shape representation of it.
- Serves as a "garment template" as in traditional mesh-based modeling but allows for varied topology.

Results

Improved representation power

NASA [4] recon. mesh from Implicit surface   SCALE [2] surface patches and reconstructed mesh   Ours dense point cloud and reconstructed mesh

not applicable to skirt data

Cross-outfit modeling: results on the CAPE [3] data

	NASA [4]	SCALE [2]	POP per-outfit	POP unified	POP unified, ½ data
Chamfer-$L2$ ($\times 10^{-4} m^2$) ↓	6.087	0.721	0.639	**0.592**	0.598
Normal Diff. ($\times 10^{-1}$) ↓	1.275	1.168	1.146	**1.115**	1.122

Single scan animation

Unseen subject   Fitted POP model   Animations with pose-dependent shapes