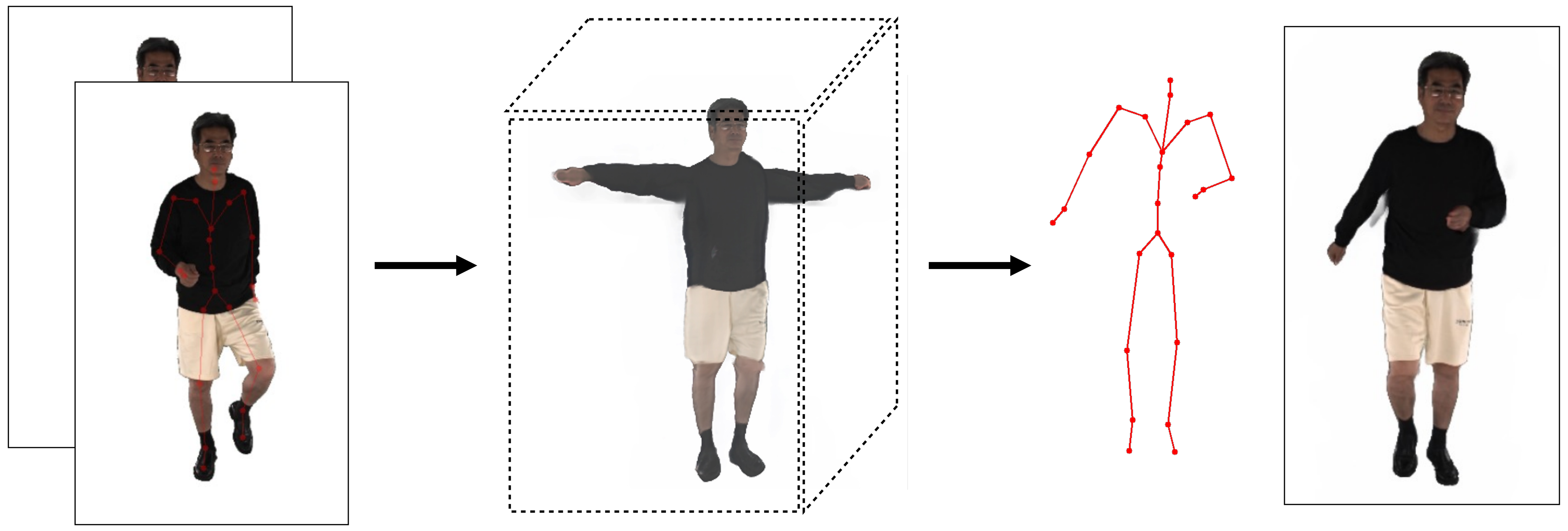
HumMorph: Generalized Dynamic Human Neural Fields from Few Views

Jakub Zadrożny, Hakan Bilen | University of Edinburgh

jakubzadrozny.github.io/hummorph

Introduction

Task: Given a few observed frames of a subject in motion, synthesize their dynamic 3D model that can be animated to a given target pose and rendered from an arbitrary viewpoint.



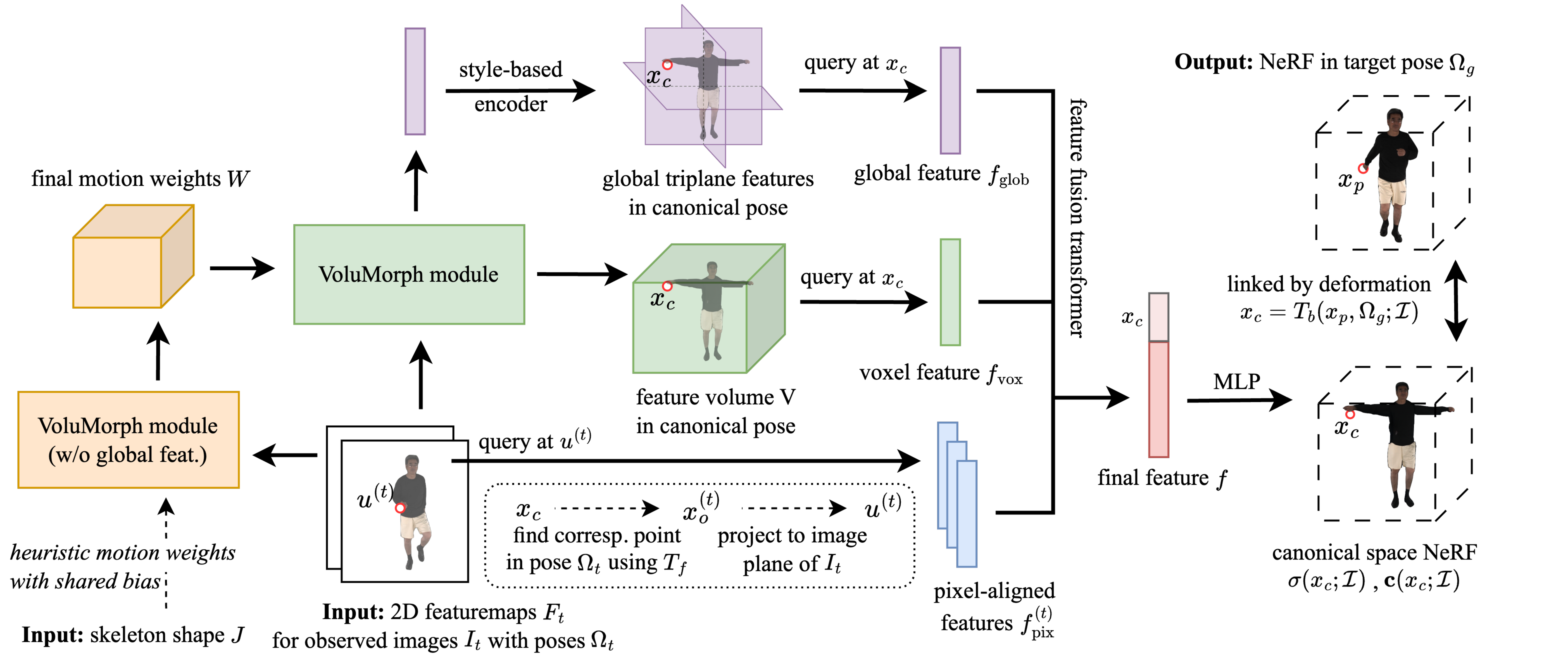
Input: a few monocular views with pose parameters

Estimate: subject NeRF model in canonical pose

Input: target pose parameters

Output: render in target pose

Method Overview



Why *HumMorph*?

Subject-specific approaches

- Require test-time optimization
- Need extensive observations (typically ca. 30 frames)

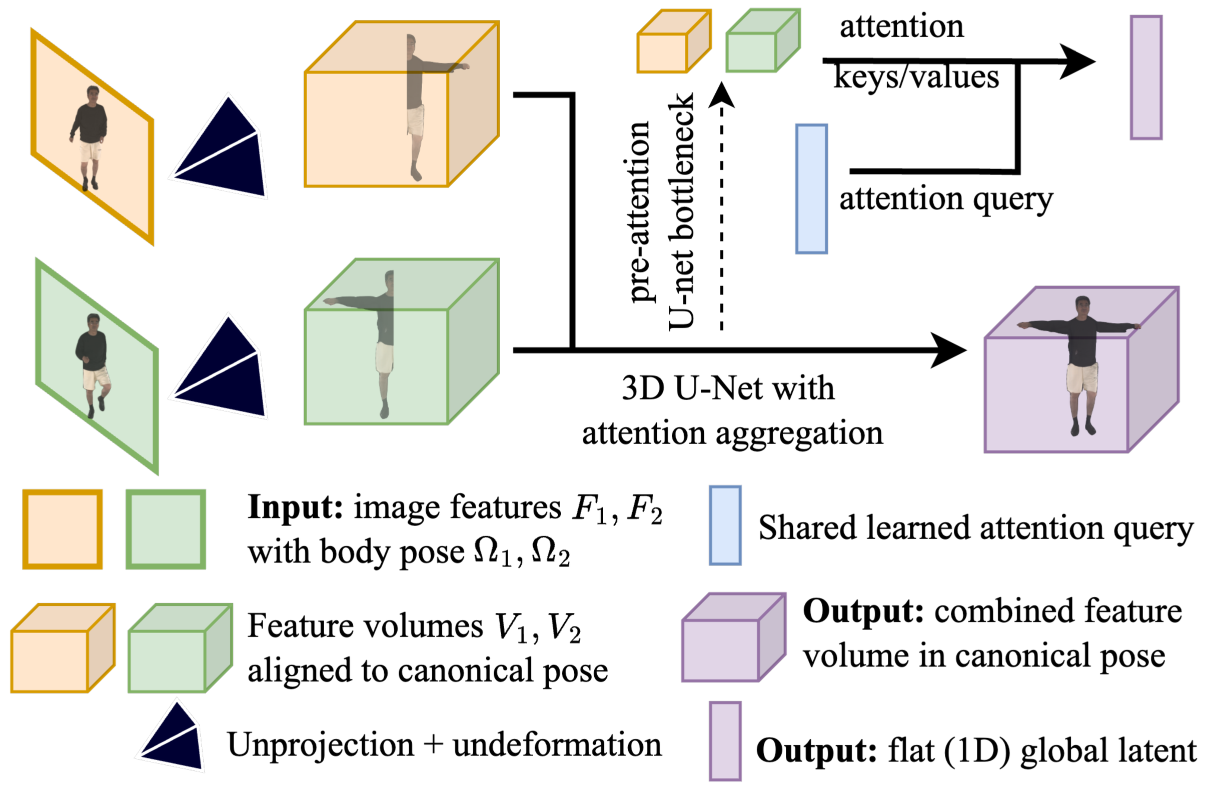
Other generalized approaches

- Assume accurate body shape and pose parameters (impractical)

***HumMorph* (ours)**

- Uses only feed-forward passes during inference
- Requires less observed views (1 - 4)
- Learns a prior, inpaints unobserved details
- Significantly more robust to errors in the noisy parameters

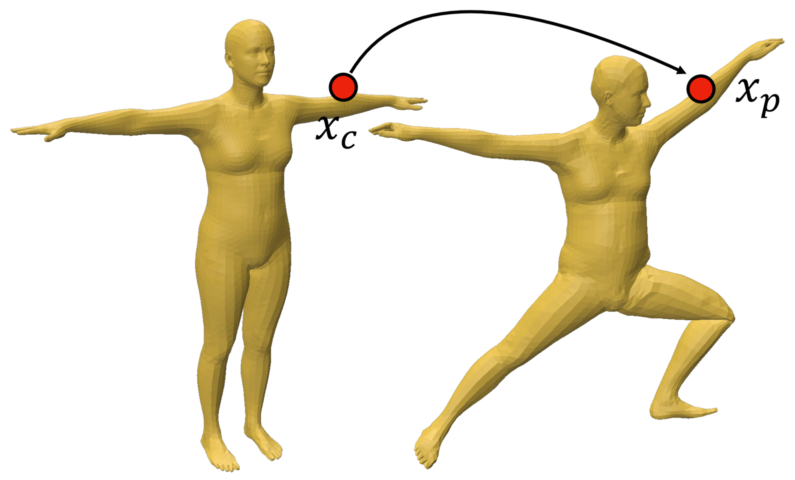
The *VoluMorph* Module



Deformations

We use **linear blend skinning** for body deformations T_f, T_b

$$T_f(x_c, \Omega) = x_p \qquad T_b(x_p, \Omega) \approx x_c$$

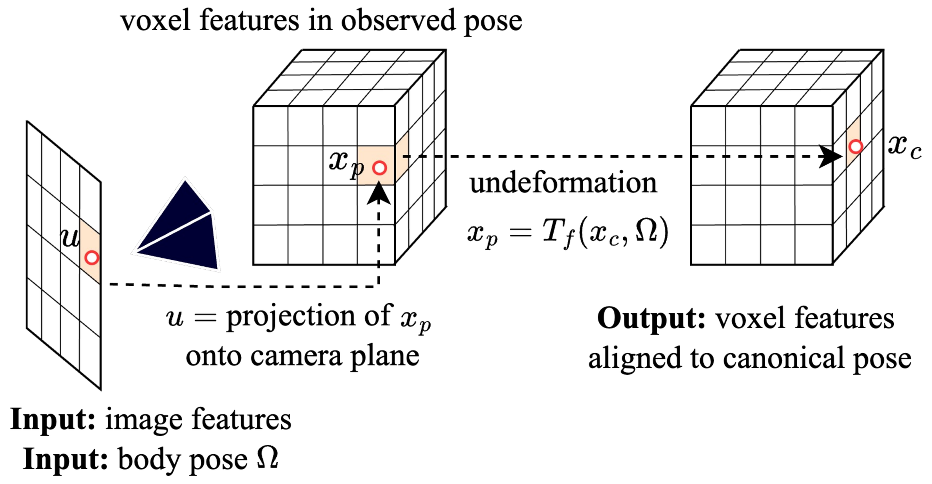


Unprojection + Undeformation

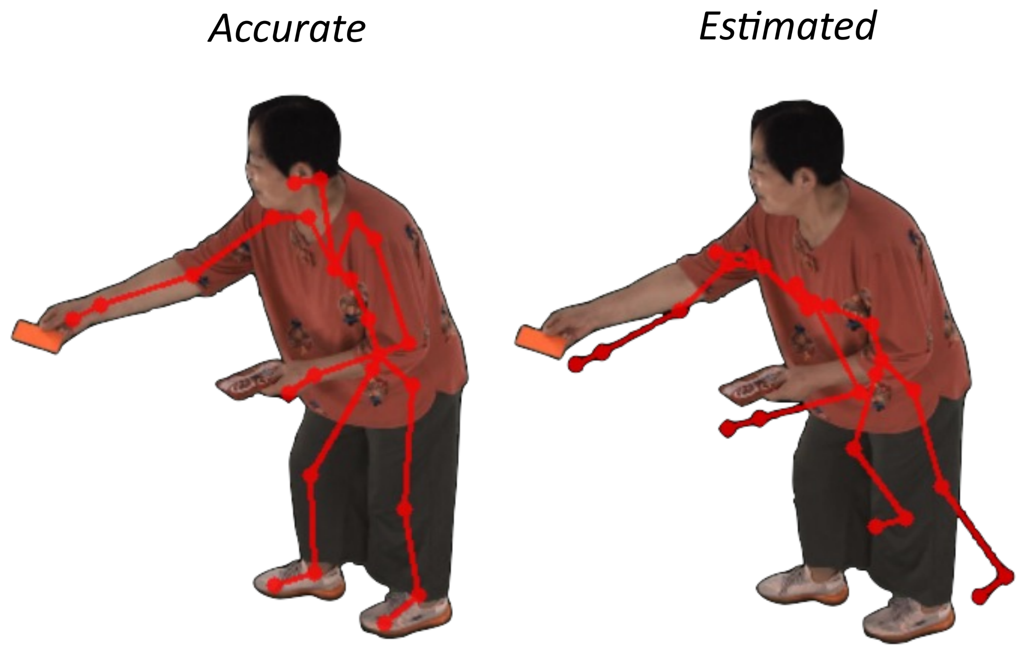
The aligned feature volume V at voxel position x_c is

$$V(x_c) = F\big[\pi\big(T_f(x_c, \Omega), \mathcal{K}, \mathcal{E}\big)\big]$$

image features, projection, camera params.

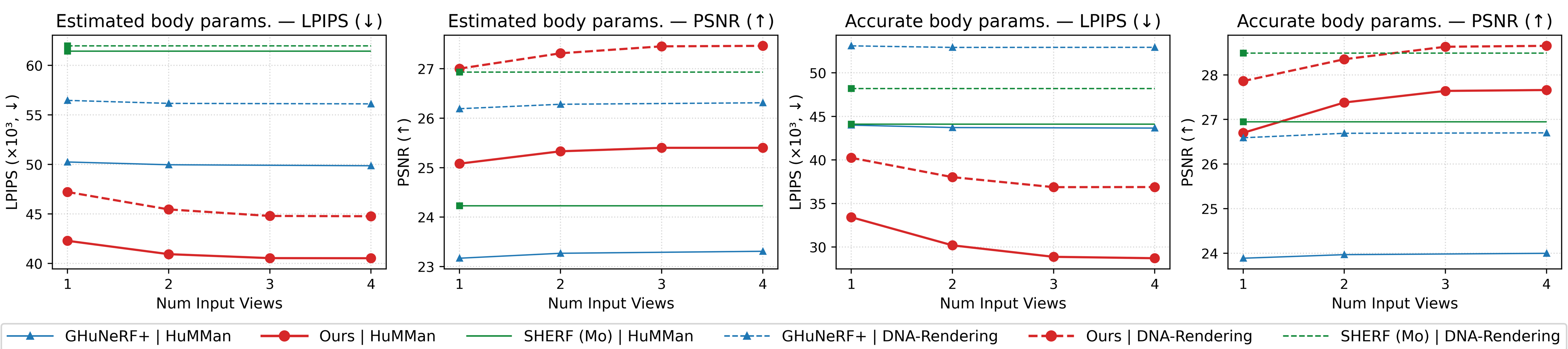


Estimated Body Shape and Pose

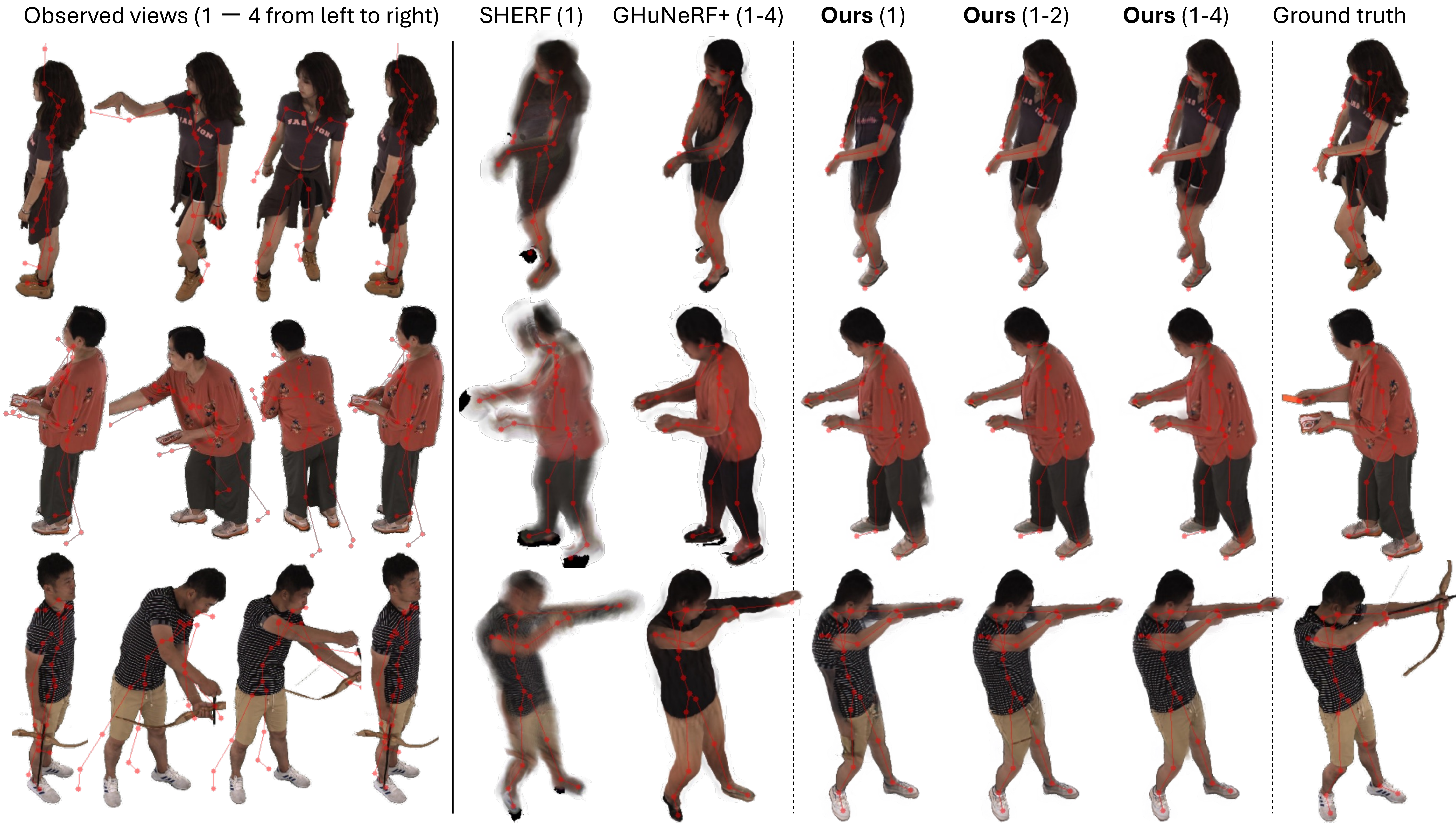


- The *accurate* body shape and pose parameters are typically estimated from multi-view camera setups.
- **They should be directly estimated from the input views instead.**
- Fig. on the left: frames with skeleton annotated in red using *accurate* (left) and *estimated* (right) body shape and pose parameters. Parameters estimated using HybrIK (Li et al., CVPR '21).

Quantitative Results

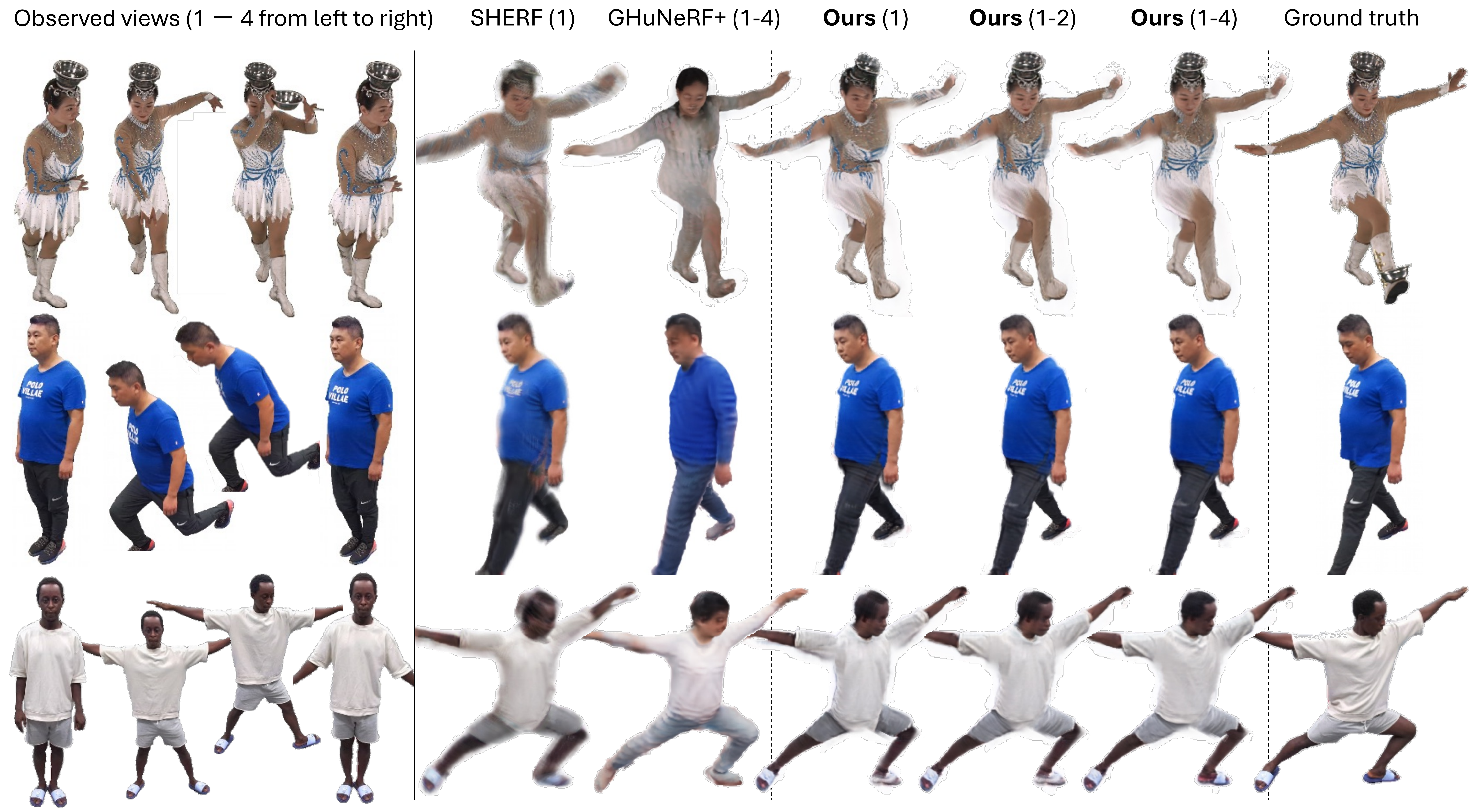


Results with **Estimated** Pose Parameters



Numbers in parentheses indicate the range of observed views. SHERF only accepts a single input view.
Poses estimated using HybrIK directly from input views (shown in red).

Results with **Accurate** Pose Parameters



Numbers in parentheses indicate the range of observed views. SHERF only accepts a single input view.
The accurate poses are provided by the datasets.