Multi-hypotheses Conditioned Point Cloud Diffusion for 3D Human Reconstruction from Occluded Images

KAIST Computer Vision and Learning Lab (KCVL)
KAIST

Donghwan Kim

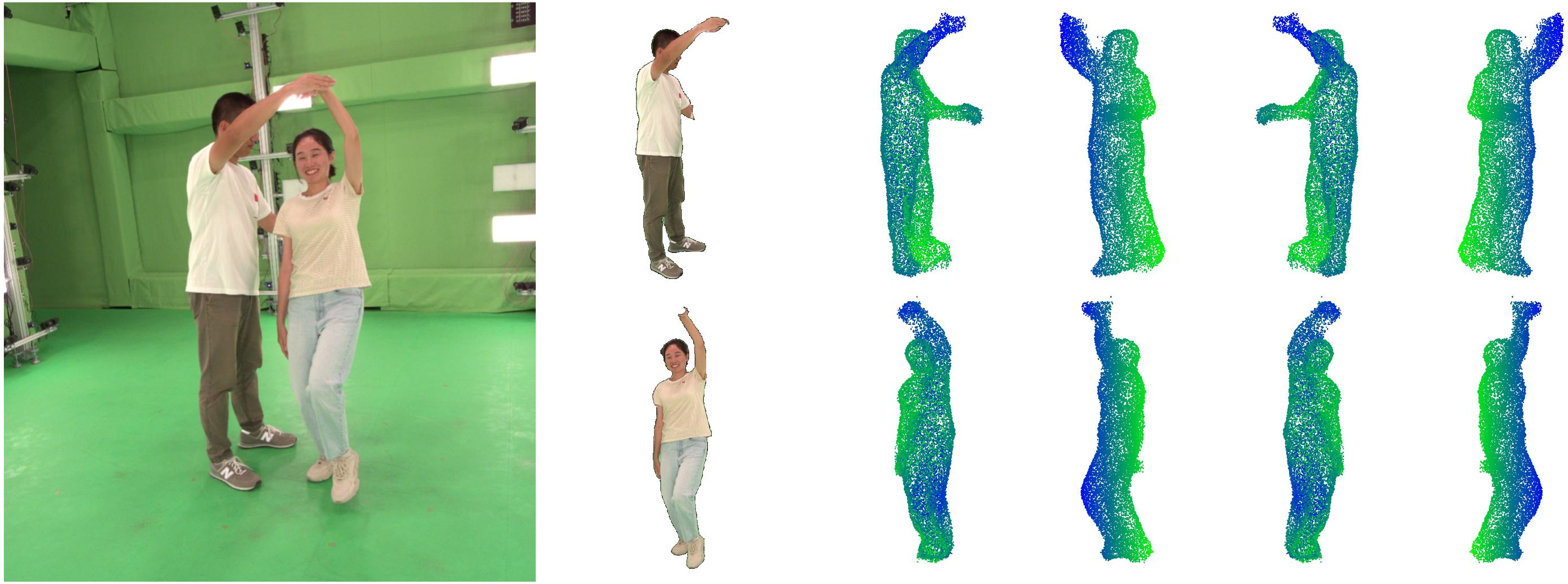
Tae-Kyun (T-K) Kim

NEURAL INFORMATION PROCESSING SYSTEMS

KAIST

MHCDIFF

predicts the *pixel-aligned* 3D shapes of humans *robustly* from occluded images.



Input image | Segmented images | 3D reconstruction as point cloud

Motivation

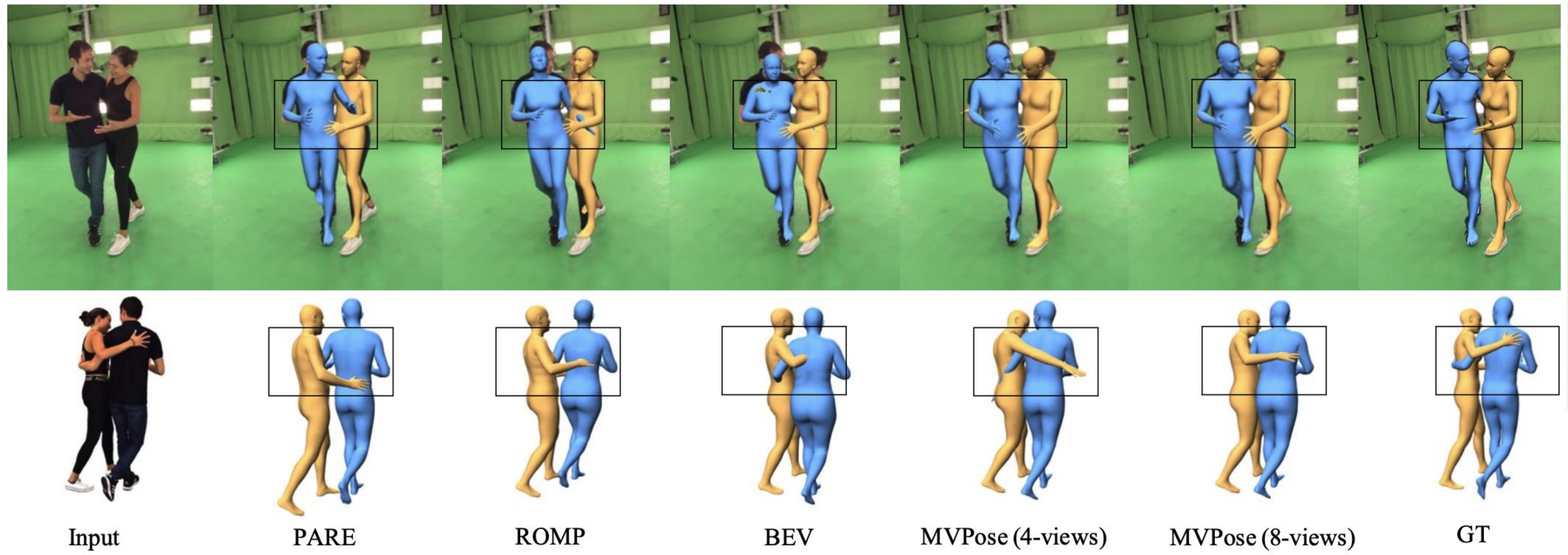
3D human reconstruction from 2D images play a significant role in metaverse.

Social interaction (human-object, human-human) make it more challenging due to occlusions.

3D Human Reconstruction

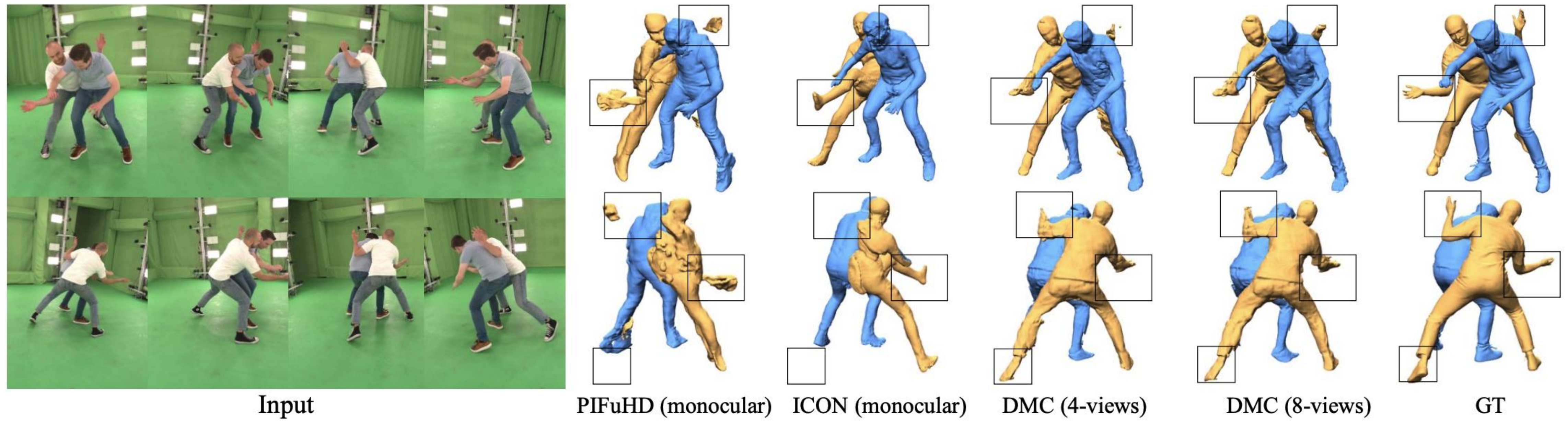
Parametric body models (SMPL/SMPL-X)

well regularized with human body priors.
robust to occlusion.
lack geometric details like clothing and hair.



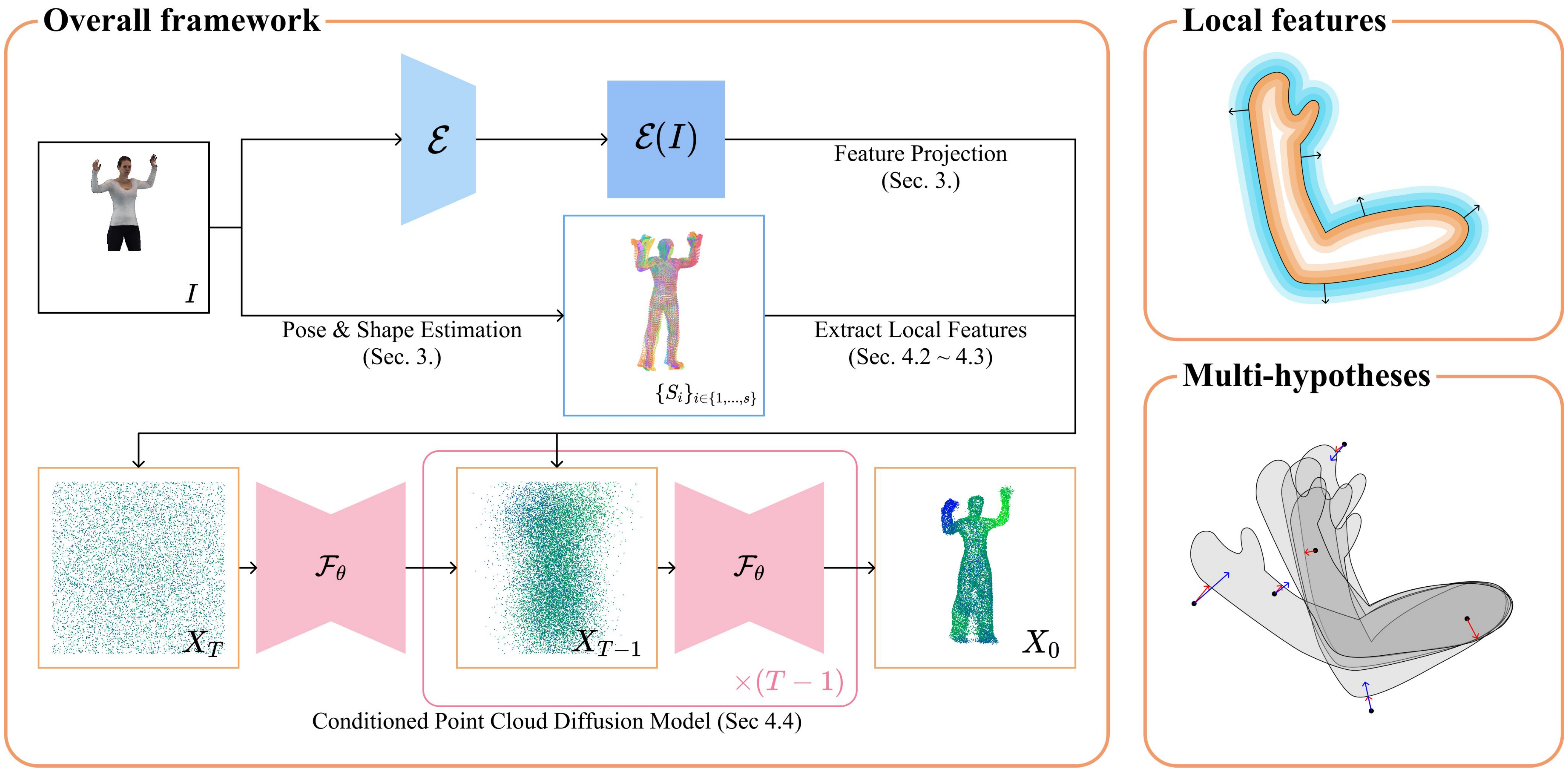
Implicit-function-based models

conditioned on SMPL estimation.
predict the pixel-aligned 3D shapes of humans.
sensitive to occlusion, cannot inpaint the invisible regions.



Yin et al. **Hi4D: 4D Instance Segmentation of Close Human Interaction**. In *CVPR*, 2023.

MHCDIFF: Multi-hypotheses Conditioned Point Cloud Diffusion



MHCDIFF: Multi-hypotheses Conditioned Point Cloud Diffusion

4.2 Local features from SMPL

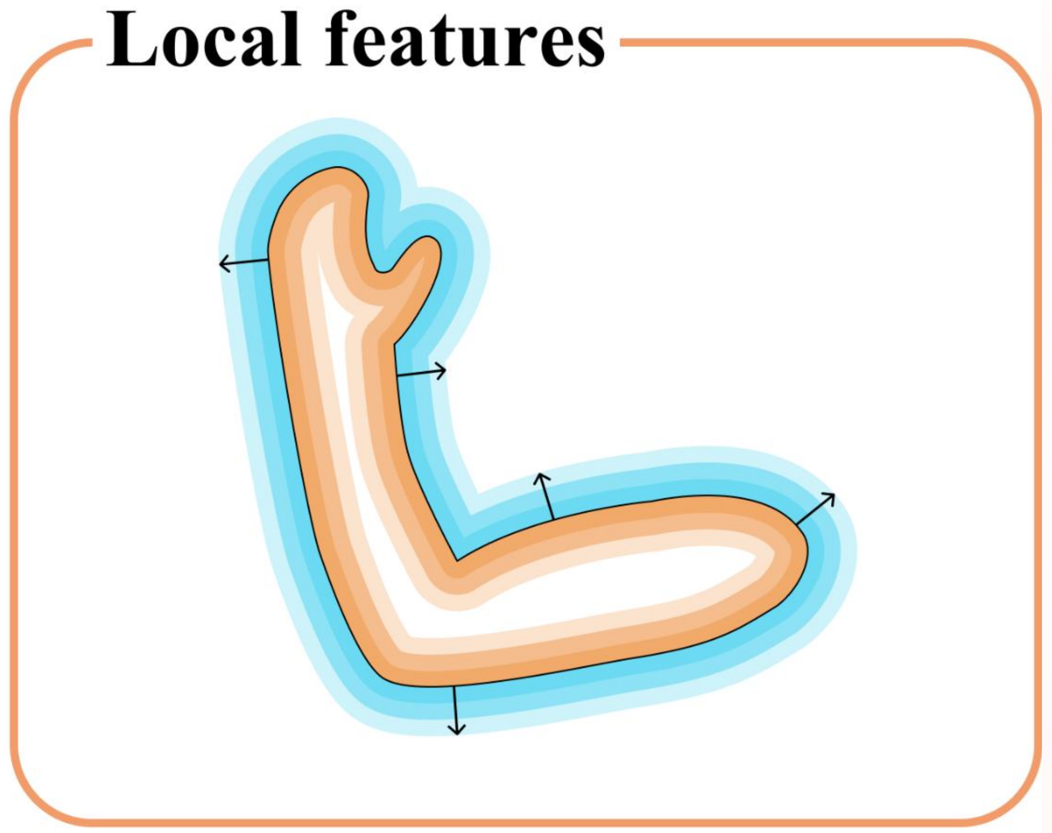
signed distance and normal obtained from the closest surface of SMPL mesh
independent of global pose.
generalize well in diverse SMPL estimation due to occlusion.

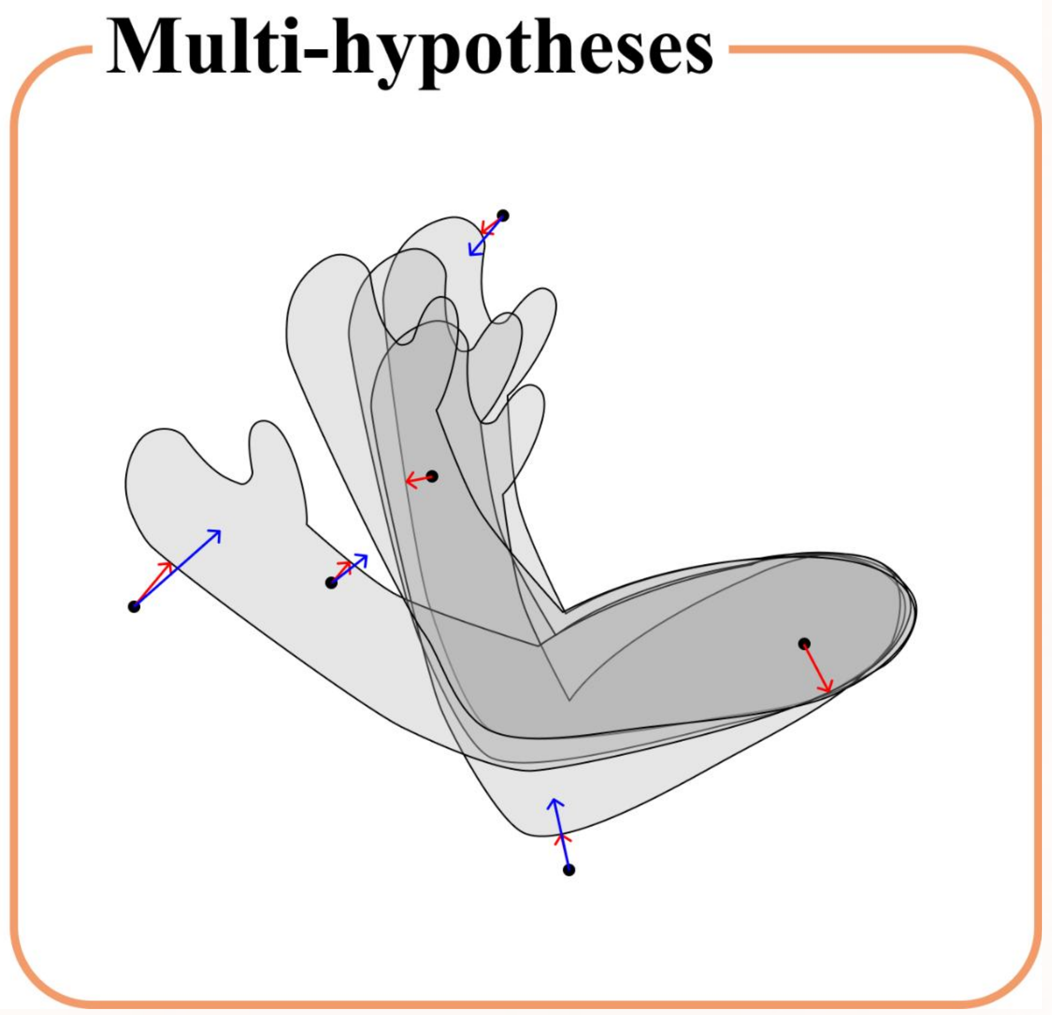
$$X_t^{SMPL} = [\gamma(d(X_t|S)), \mathbf{n}(X_t|S)]$$

4.3 Multi-hypotheses condition

effectively captures the distribution of multiple plausible SMPL meshes.
robust to the noise of each SMPL estimation due to the occlusion.

$$X_t^{SMPL} = [\frac{1}{s}\sum_{i=1}^{s}\gamma(o(X_t|S_i)), \gamma(d(X_t|S_{\bar{i}})), \mathbf{n}(X_t|S_{\bar{i}})]$$



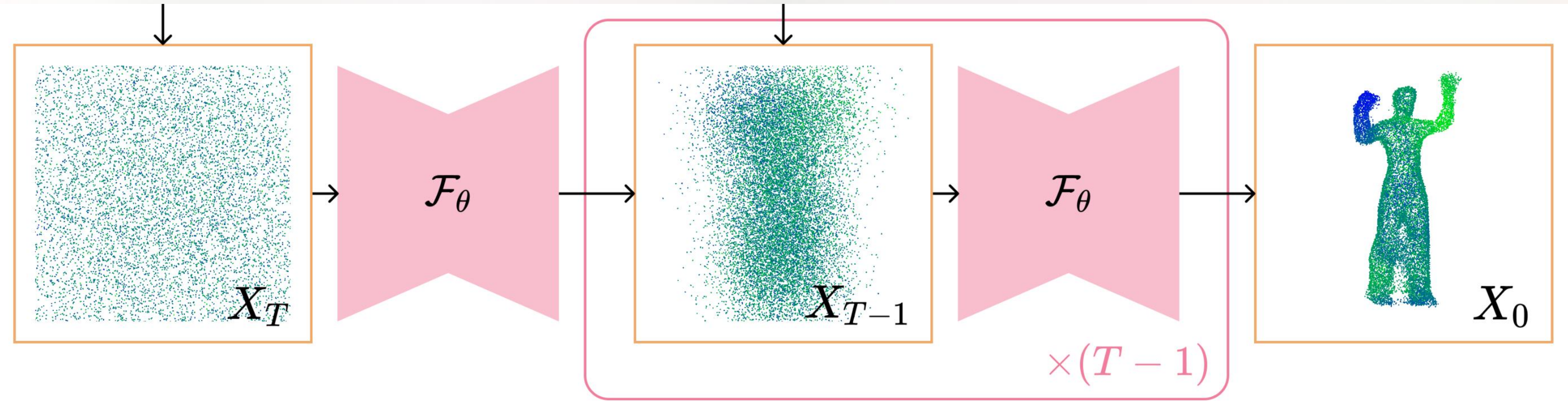


MHCDIFF: Multi-hypotheses Conditioned Point Cloud Diffusion

4.4 Conditioned point cloud diffusion model

capture the global consistent features and generate the invisible parts.
correct the misaligned SMPL estimation during the denoising process.

$$\mathcal{F}_\theta(\cdot) : \mathbb{R}^{(3+c+4L+3)N} \rightarrow \mathbb{R}^{3N}$$



Conditioned Point Cloud Diffusion Model (Sec 4.4)

Experiments

Quantitative evaluation

MHCDIFF outperforms prior implicit-function-based methods and SMPL/SMPL-X estimation methods on occluded images, and shows comparable performance on full-body images.

Randomly masked CAPE

	Methods	Chamfer Distance (cm)	Point-to-Surface (cm)
A	PaMIR [106]	12.912	12.619
	ICON [92]	2.896	2.789
	ICON (PIXIE estimation)	3.329	3.212
	SIFU [104]	14.397	14.087
	HiLo [96]	13.711	13.405
B	PIXIE (SMPL-X) [15]	2.705	2.662
	ProPose (SMPL) [14]	<u>2.370</u>	<u>2.307</u>
Ours	MHCDIFF	**1.872**	**1.810**

MultiHuman

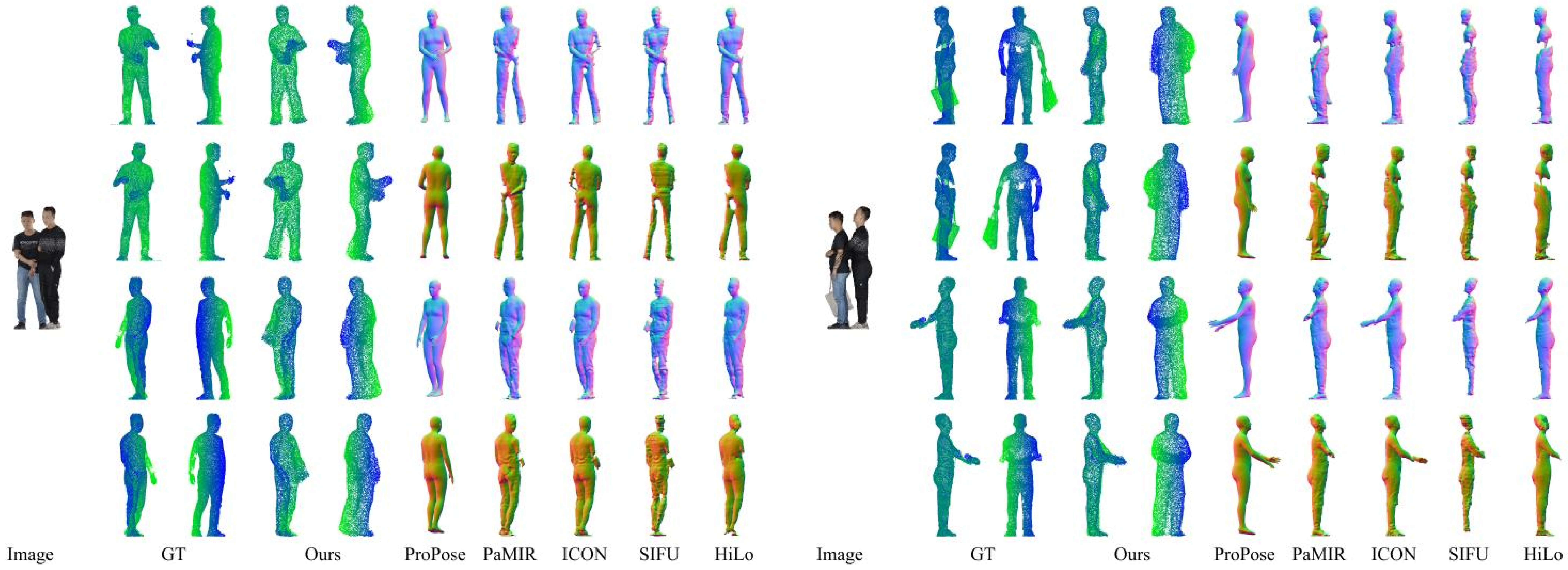
	Methods	single	occluded single	two natural-inter	two closely-inter	three
A	PaMIR [106]	0.690	2.349	5.154	3.752	4.714
	ICON [92]	**0.555**	<u>0.549</u>	<u>0.563</u>	0.786	**0.669**
	SIFU [104]	0.644	3.335	4.796	3.503	3.264
	HiLo [96]	0.606	2.808	4.139	3.346	4.398
B	PIXIE (SMPL-X) [15]	0.868	0.813	0.755	0.951	0.809
	ProPose (SMPL) [14]	0.675	0.567	0.574	<u>0.766</u>	0.688
Ours	MHCDIFF	<u>0.591</u>	**0.491**	**0.536**	**0.703**	<u>0.673</u>

Experiments

Qualitative evaluation (CAPE)

Experiments

Qualitative evaluation (MultiHuman)

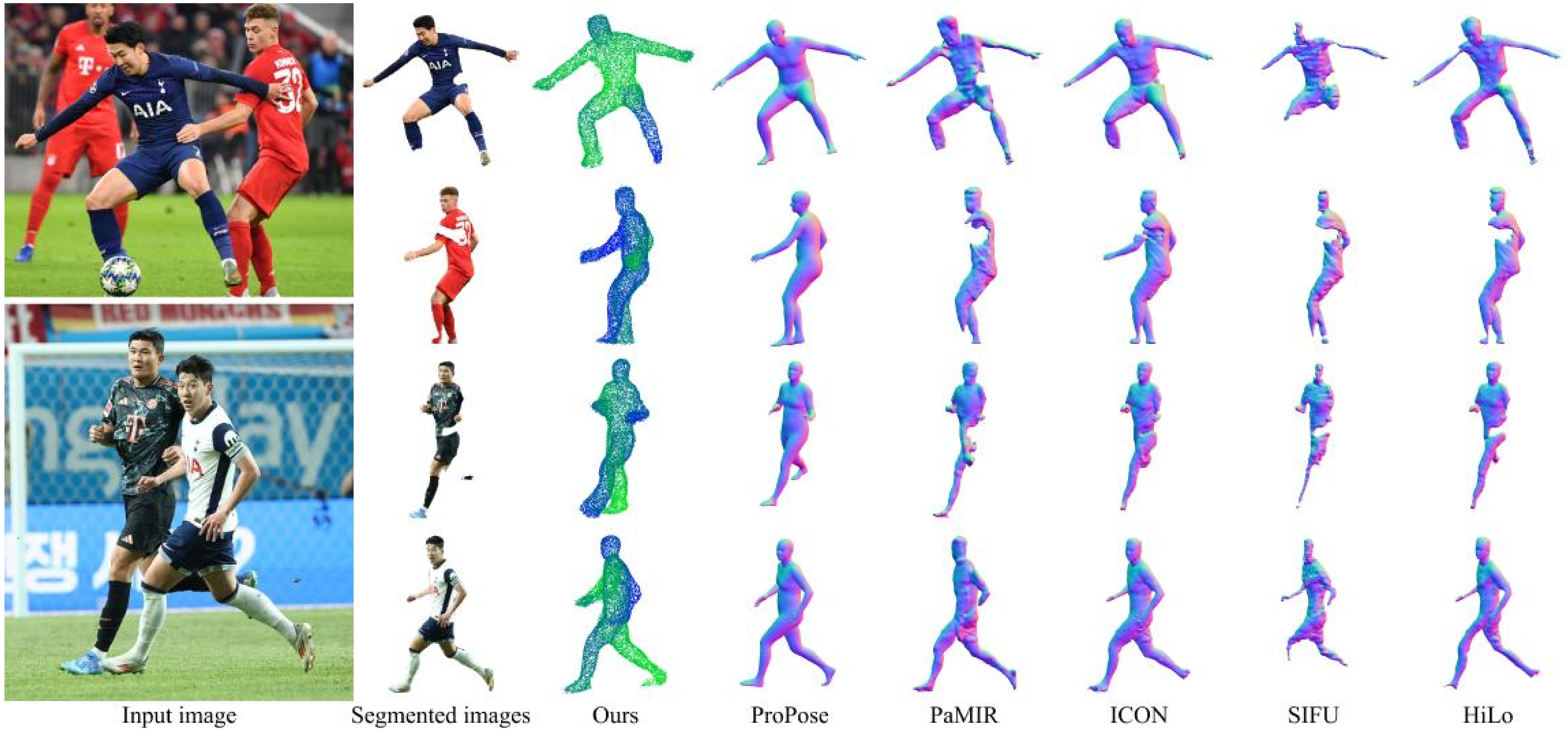


Experiments

Qualitative evaluation (Hi4D)

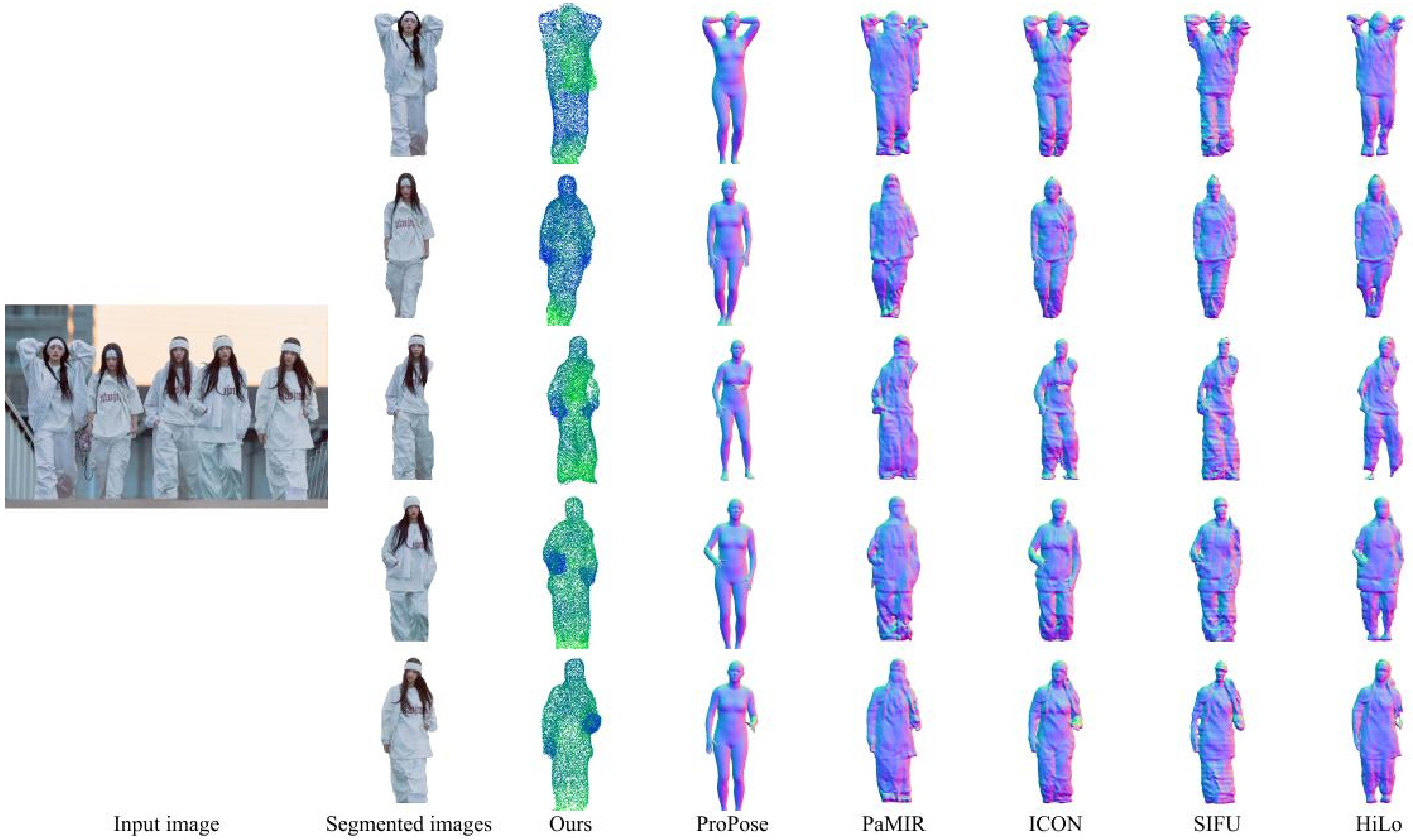
Experiments

Qualitative evaluation (in-the-wild)



Experiments

Qualitative evaluation (in-the-wild)



Conclusion

Contributions

Multi-hypotheses conditioning mechanism effectively captures the distribution of multiple plausible SMPL meshes.
Point cloud diffusion model captures the global consistent features and inpaints the invisible parts.

Limitations

Inference is slow due to iterative denoising procedures.
Point cloud may not be directly usable in real-world applications.

Thank you!

Website: https://donghwankim0101.github.io/projects/mhcdiff
Code: https://github.com/DonghwanKIM0101/MHCDIFF

Donghwan Kim

Tae-Kyun (T-K) Kim

NEURAL INFORMATION PROCESSING SYSTEMS

KAIST