



Full-Body Articulated Human-Object Interaction

Nan Jiang^{1,2*}, Tengyu Liu^{2,*}, Zhexiong Cao³, Jieming Cui^{1,2}, Zhiyuan Zhang^{2,3}, Yixin Chen², He Wang⁴, Yixin Zhu⁵✉, Siyuan Huang²✉

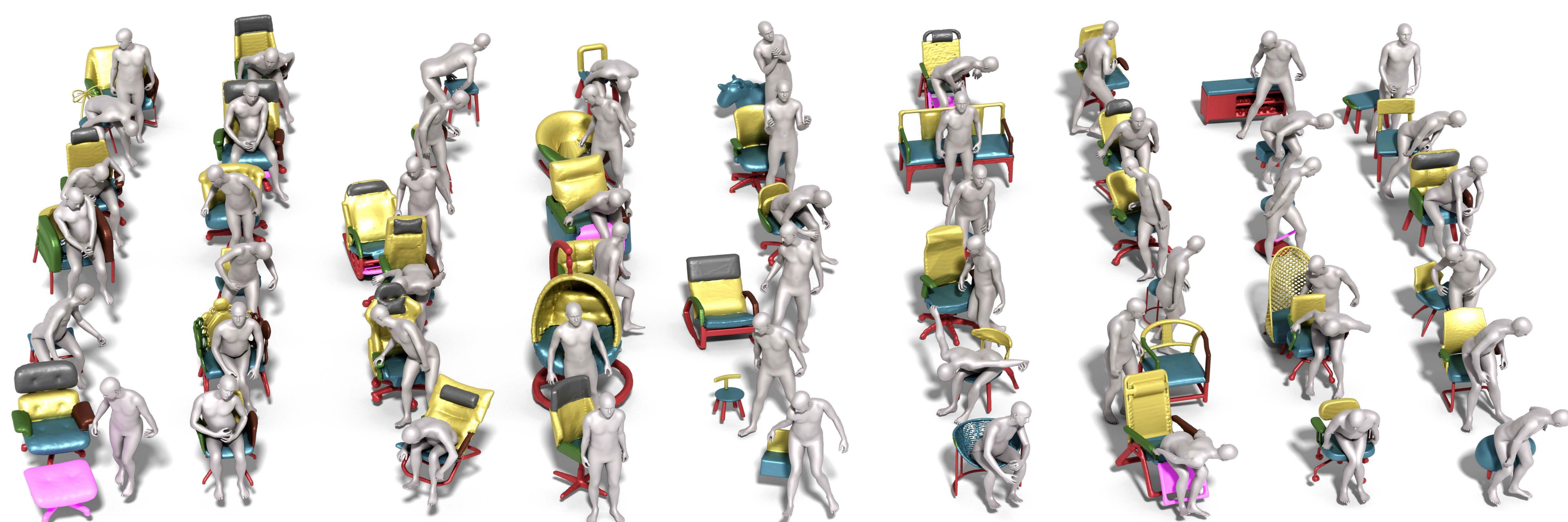
¹ School of Intelligence Science and Technology, Peking University ² Beijing Institute of General Artificial Intelligence (BIGAI) ³ Department of Automation, Tsinghua University ⁴ Center on Frontiers of Computing Studies, Peking University ⁵ Institute for Artificial Intelligence, Peking University

ICCV23
PARIS



<https://jnnan.github.io/project/chairs/>

Dataset



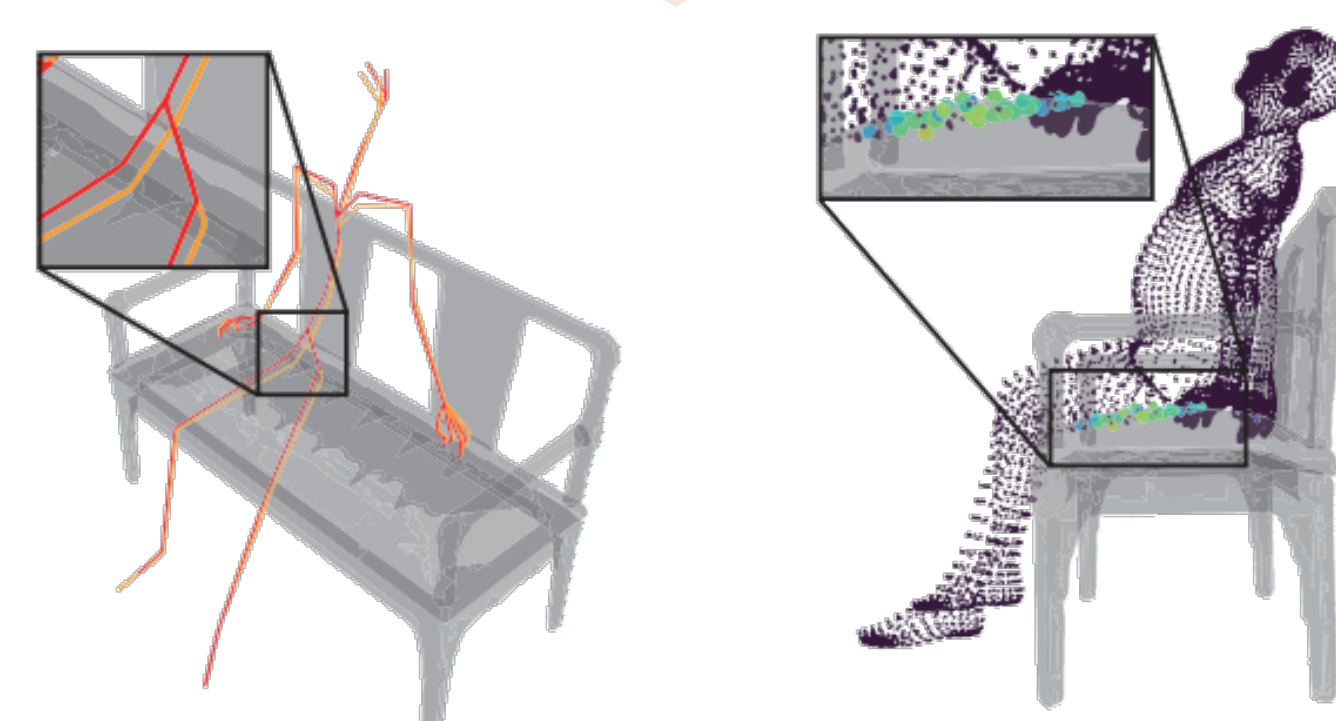
- ✓ **CHAIRS**, an extensive motion-captured **full-body articulated** HOI dataset.
- ✓ Comprises **17.3 hours** of diverse **interaction sequences**.
- ✓ Involves **46 participants** and **81 objects** with various movable parts.



- Data collection setup consisted of
- Four front-facing **RGB-D cameras**
 - **Hybrid trackers** on movable parts of objects
 - **17 IMUs** distributed on the participants.

Penetration removal process:

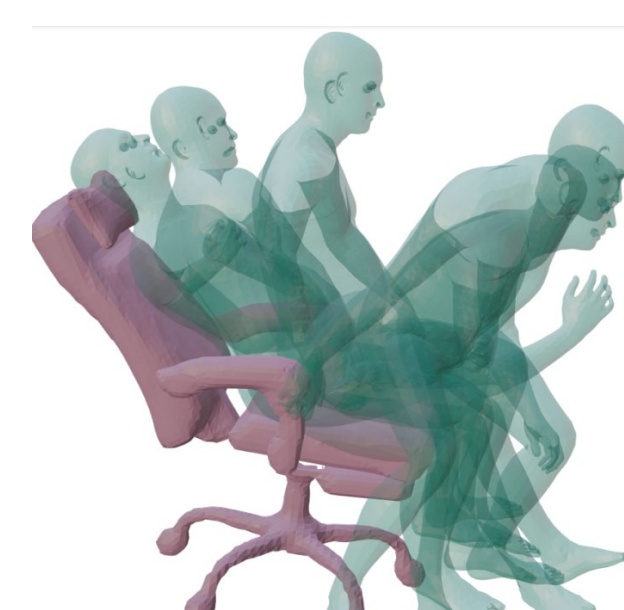
- Adjust positions of joints responsible for penetration
- Adjust skeleton by IK



Tasks

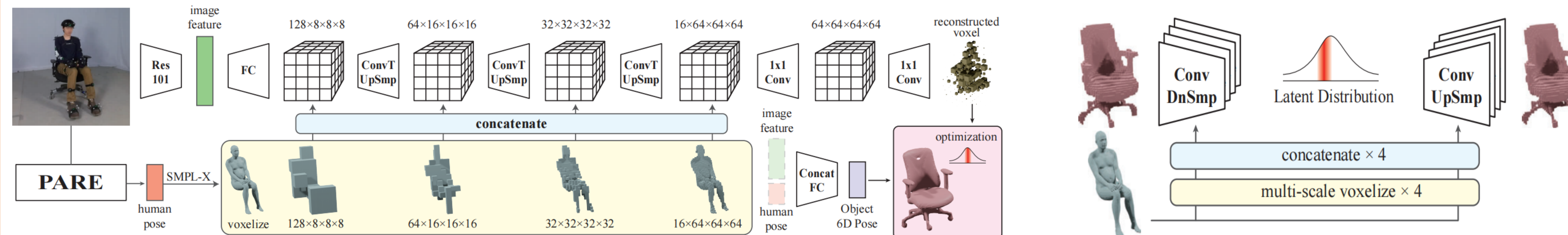


Task I: articulated object pose estimation from images



Task II: interacting human pose synthesis from articulated chairs

Method



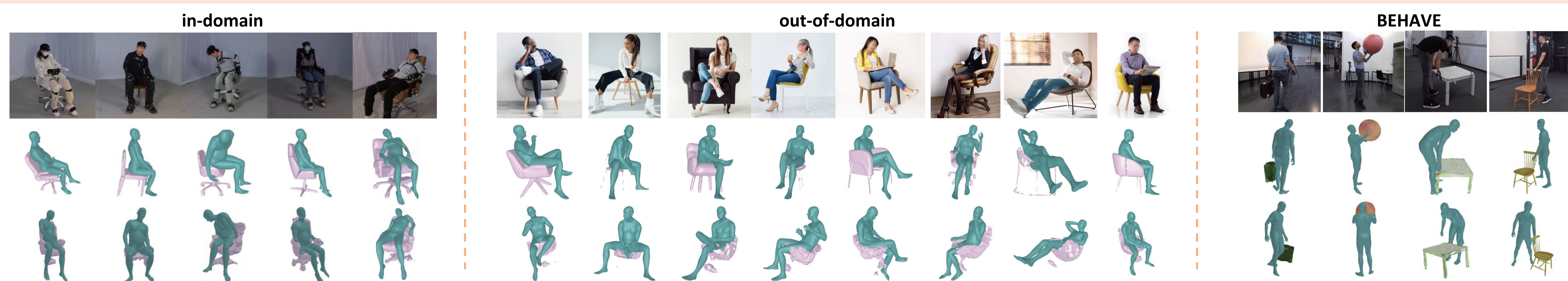
The **reconstruction model** uses the human pose for object pose estimation. We

1. regress the root 6D pose from human pose and image feature;
2. reconstruct object voxel from voxelized human and image feature; and
3. refine object pose and shape with reconstructed voxel and interaction prior.

The **interaction prior model**, a cVAE, is trained to generate object voxels conditioned on human poses. During inference, we minimize the norm of the latent code.

Results

Qualitative results on Task I: object pose estimation

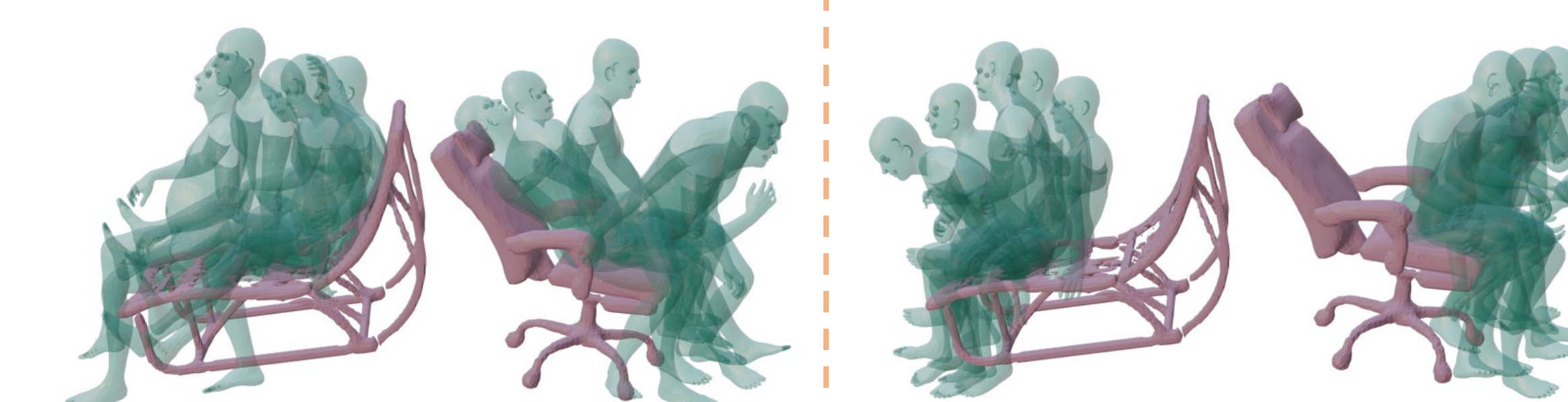


Quantitative results on task I: pose estimation

Method	Object					HOI	
	Rot.↓ (°)	Transl.↓ (mm)	CD↓ (mm)	IoU↑ (%)	Pene.↓ (mm)	Cont.↓ (mm)	
LASR [†] [59]	/	/	205.2	/	/	/	
Ours (w/o opt.) [†]	/	/	160.2	11.03	4.530	2.720	
ANCSH* [28]	/	/	90.36	/	/	/	
PHOSA* [65]	29.31	175.2	177.9	7.60	2.046	1.689	
D3D-HOI* [58]	27.31	119.2	126.9	16.60	7.472	1.163	
CHORE* [56]	21.82	87.58	95.40	16.44	1.050	1.742	
Ours (w/ opt.) [*]	19.35	66.23	72.30	21.57	1.143	1.562	

*: method requires knowledge of object structure and/or geometry; †: method does not rely on object-related knowledge.

Qualitative Results on Task II: human pose synthesis



Ours

Baseline