MANIKIN: Biomechanically Accurate Neural Inverse Kinematics for Human Motion Estimation

ETH zürich — sensing, interaction & perception lab

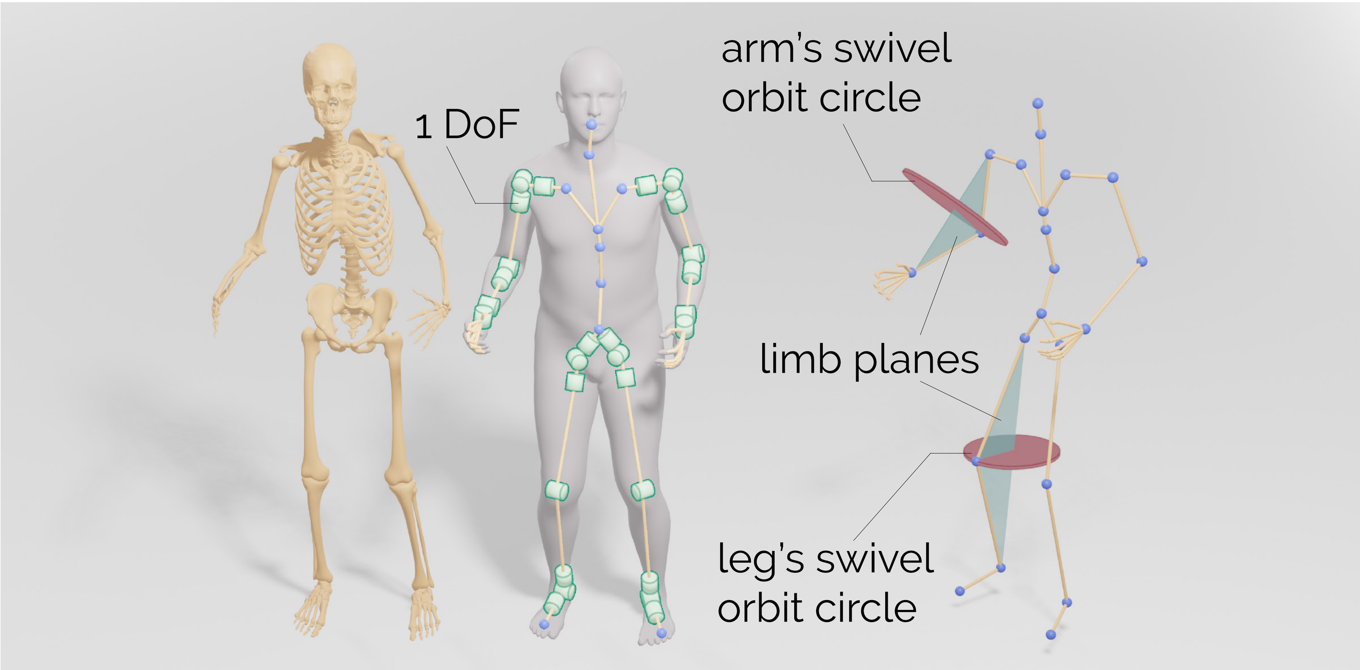
Jiaxi Jiang, Paul Streli, Xuejing Luo, Christoph Gebhardt, Christian Holz

Department of Computer Science, ETH Zürich, Switzerland

project page

ECCV — EUROPEAN CONFERENCE ON COMPUTER VISION — MILANO 2024

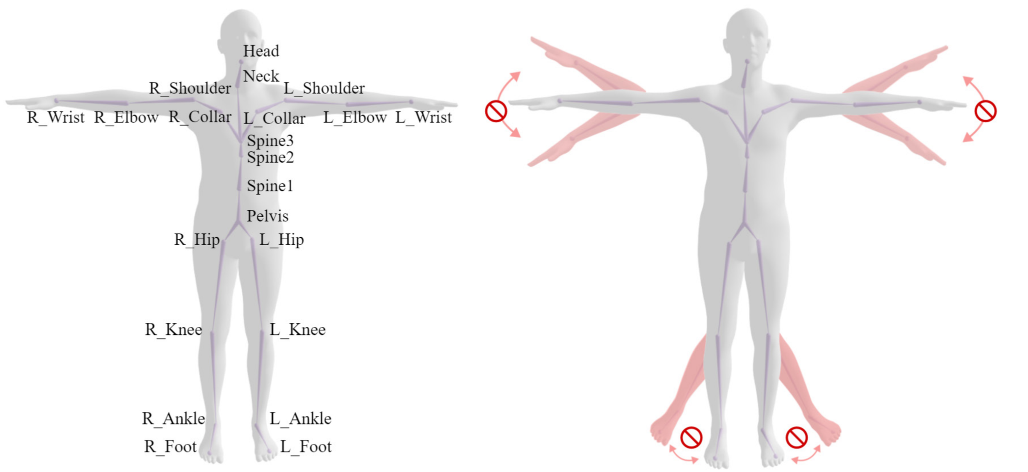
introduction



We introduce **MANIKIN**, a bio**m**echanically **a**ccurate **n**eural **i**nverse **kin**ematics solver for human motion estimation. Unlike previous methods, which often produce unrealistic human motions and suffer from significant accumulated position errors at the end effectors, our differentiable neural IK framework, grounded in realistic human motion modeling and swivel angle prediction, effectively addresses these issues while maintaining fast inference speed.

human motion modeling

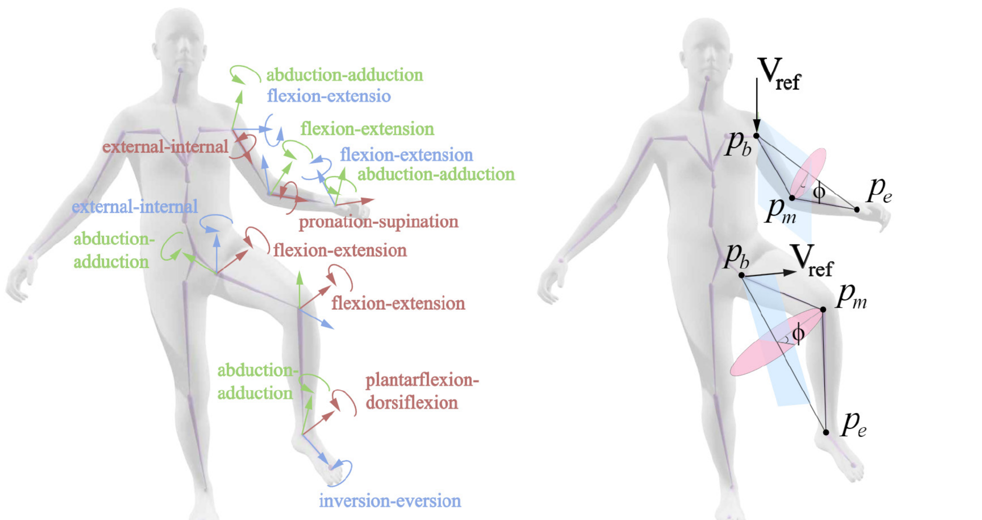
1. The limitations of SMPL models



(a) SMPL model (b) Unrealistic joint motions in AMASS.

SMPL model (a) is widely used in human pose and shape estimation for its realistic skin representation and bone length consistency. However, its 3-DoF assumption does not fully account for the inherent constraints governing realistic human motion. SMPL-based datasets (e.g., AMASS) usually contain instances of unrealistic SMPL joint configurations (Fig. b).

2. Biomechanically accurate human motion modeling



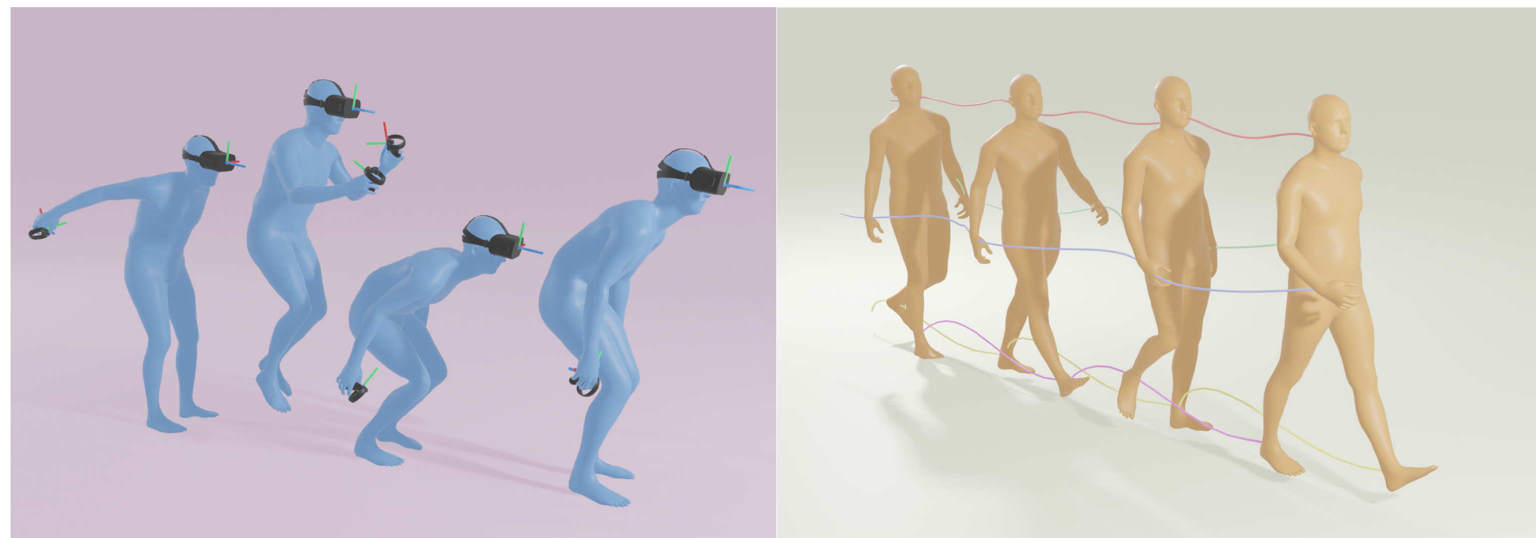
(c) Our MANIKIN body model (d) Limb swivel angles

In key literature of biomechanics, the prevalent model for describing the human arms and legs constitutes a total of seven degrees of freedom. Informed by this, we correct the DoFs in SMPL model's joints to more accurately mirror the natural motion constraints of human limbs (Fig. c).

3. Swivel angle of human limbs

As shown in Fig. d, we base motion modeling of arms and legs on the swivel angle, which represents the extent to which the respective mid joint (elbow or knee) is rotated around the end-base axis (wrist-shoulder or ankle-hip). The swivel angle can parameterize the extra degree of freedom to give a unique solution.

task descriptions



(a) full-body moion tracking using MR devices (b) motion synthesize from 3D trajectories

In this paper, we showcase two key applications of our designed neural analytic IK algorithm based on realistic biomechanical constraints:

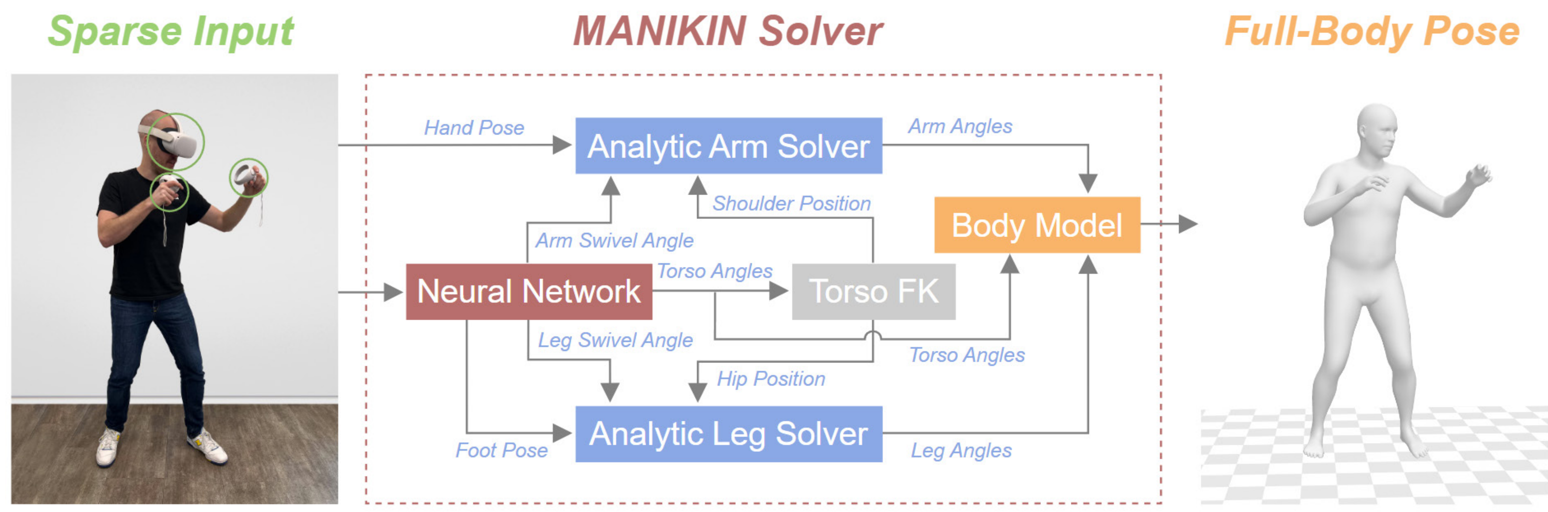
(a) Estimating full-body motion using only 6 DoF head and hand poses within a Mixed Reality environment.

(b) Synthesizing human movements based on 3D trajectories of the head, hands, and feet.

Our primary focus is on (a), as it addresses a widely recognized and well-established challenge.

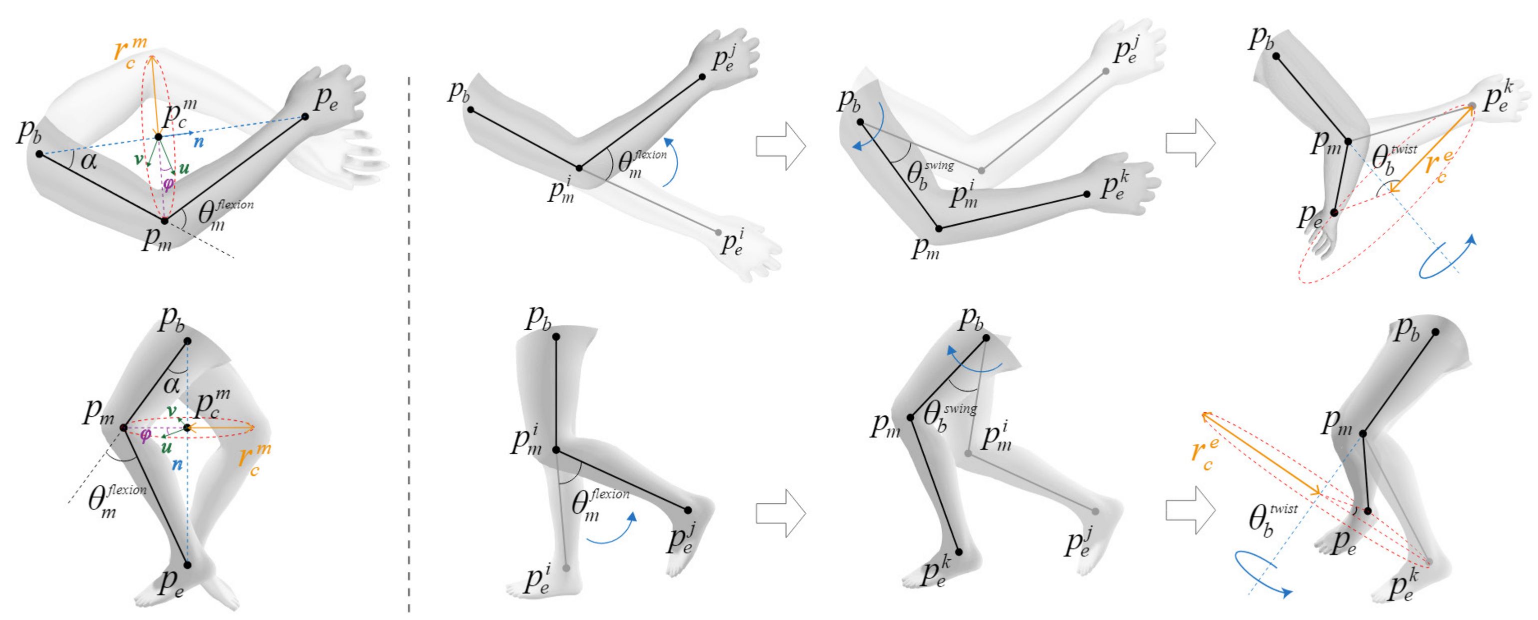
neural analytic IK

1. Method overview



Given the 6D poses of the head and hands as input, the Neural Network predicts the body's global orientation, local poses of the joints on the torso, the foot pose, and the swivel angle of arms and legs. Using torso angles, forward kinematics is applied to attain shoulder and hip positions (Torso FK). Our biomechanical constraints allow us to analytically compute limb angles via inverse kinematics on the respective swivel angle and base joint position (Analytic Arm/Leg Solver). Finally, we render the full-body model.

2. Analytic inverse kinematics



(a) swivel angel (b) mid flexion (c) base swing (d) base twist

Illustrations of the triangular geometry of the human limbs. (a) shows the relationship between the swivel angle and mid-joint position. (b) to (d) show the procedure to rotate the limb from the T-pose to the desired positions.

numerical results

1. Comparisons with SoTA methods on the AMASS dataset

Methods	MPJPE	H-PE	GP	U-PE	L-PE	MPJVE	Jitter	FPS	Param.	FLOPs
AvatarPoser* [21]	4.20	1.98	2.36	2.01	7.85	29.32	16.64	612	4.12M	0.33G
AGRoL* [13]	3.86	1.42	2.24	1.60	7.13	50.78	48.30	17.8	7.48M	1.00G
MANIKIN-S (Ours)	**3.36**	**0.02**	**0.75**	**1.32**	**6.72**	**23.18**	**14.08**	580	4.12M	0.33G
AvatarPoser [21]	4.18	1.15	2.36	1.76	7.85	29.40	16.80	12.4	4.12M	0.33G
AGRoL [13]	3.71	1.31	2.21	1.55	6.84	18.59	7.26	0.24	7.48M	1.00G
AvatarJLM [74]	3.35	1.24	0.81	1.53	6.54	20.79	10.08	31.1	63.8M	4.64G
MANIKIN-L (Ours)	3.19	0.01	0.69	1.43	6.27	20.10	9.97	30.5	63.8M	4.64G
MANIKIN-LN (Ours)	**2.73**	**0.01**	**0.52**	**1.30**	**5.13**	**13.55**	**7.95**	0.81	63.8M	4.64G

2. Results of cross-dataset evaluation

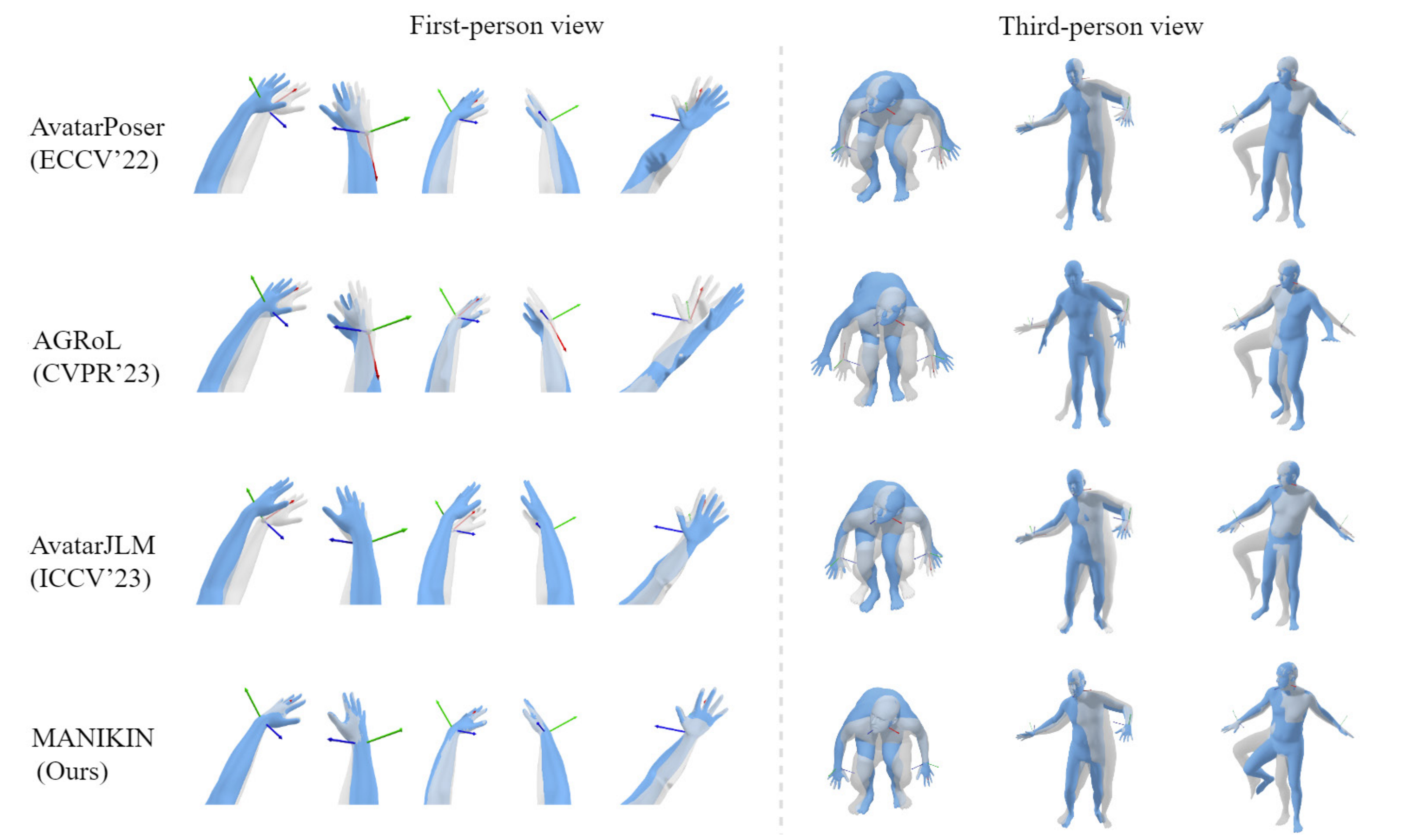
Methods	CMU				BML				MPI			
	MPJPE	H-PE	GP	MPJVE	MPJPE	H-PE	GP	MPJVE	MPJPE	H-PE	GP	MPJVE
AvatarPoser* [21]	8.40	4.72	2.06	35.88	7.08	3.37	2.85	43.85	8.07	5.61	2.64	30.92
AGRoL* [13]	8.81	3.41	2.98	85.85	7.34	2.01	2.21	56.21	8.18	2.46	3.00	103.01
MANIKIN-S (Ours)	**7.21**	**0.04**	**0.72**	**29.06**	**6.31**	**0.03**	**1.99**	**40.29**	**6.41**	**0.04**	**0.70**	**27.03**
AvatarPoser [21]	8.37	2.08	2.06	35.76	7.04	1.54	2.85	43.70	8.05	2.56	2.64	30.85
AGRoL [13]	8.87	3.19	2.20	28.02	7.29	1.97	2.97	34.29	7.91	2.34	2.86	26.09
AvatarJLM [74]	7.75	1.37	1.11	26.54	6.49	0.85	2.24	36.96	6.60	1.17	1.65	23.57
MANIKIN-L (Ours)	6.92	0.04	0.62	25.38	6.02	0.02	1.97	33.74	6.19	0.03	0.66	23.28
MANIKIN-LN (Ours)	**6.90**	**0.04**	**0.60**	**22.49**	**5.96**	**0.01**	**1.68**	**30.33**	**6.08**	**0.01**	**0.65**	**21.74**

3. Results of motion estimation from trajectories of the head, hands, and feet

Methods	MPJPE	MPJVE	U-PE	L-PE	H-PE	F-PE
Transformer-SMPL	5.34	22.74	4.37	6.73	5.65	6.48
Transformer-NeCoIK (Ours)	**3.74**	**14.98**	**3.31**	**4.36**	**0.11**	**0.12**
RNN-SMPL	7.93	43.84	5.64	11.24	7.94	12.86
RNN-NeCoIK (Ours)	**4.92**	**24.97**	**4.10**	**6.10**	**0.20**	**0.76**
MLP-SMPL	8.68	32.26	6.52	11.81	9.96	13.17
MLP-NeCoIK (Ours)	**5.37**	**20.10**	**4.52**	**6.60**	**0.24**	**0.78**

visual results

1. Visual comparisons with SoTA methods on the AMASS dataset



2. Visual comparisons on the real VR dataset