Ego-Body Pose Estimation via Ego-Head Pose Estimation

Jiaman Li, C. Karen Liu†, Jiajun Wu†

(†indicates equal contribution)

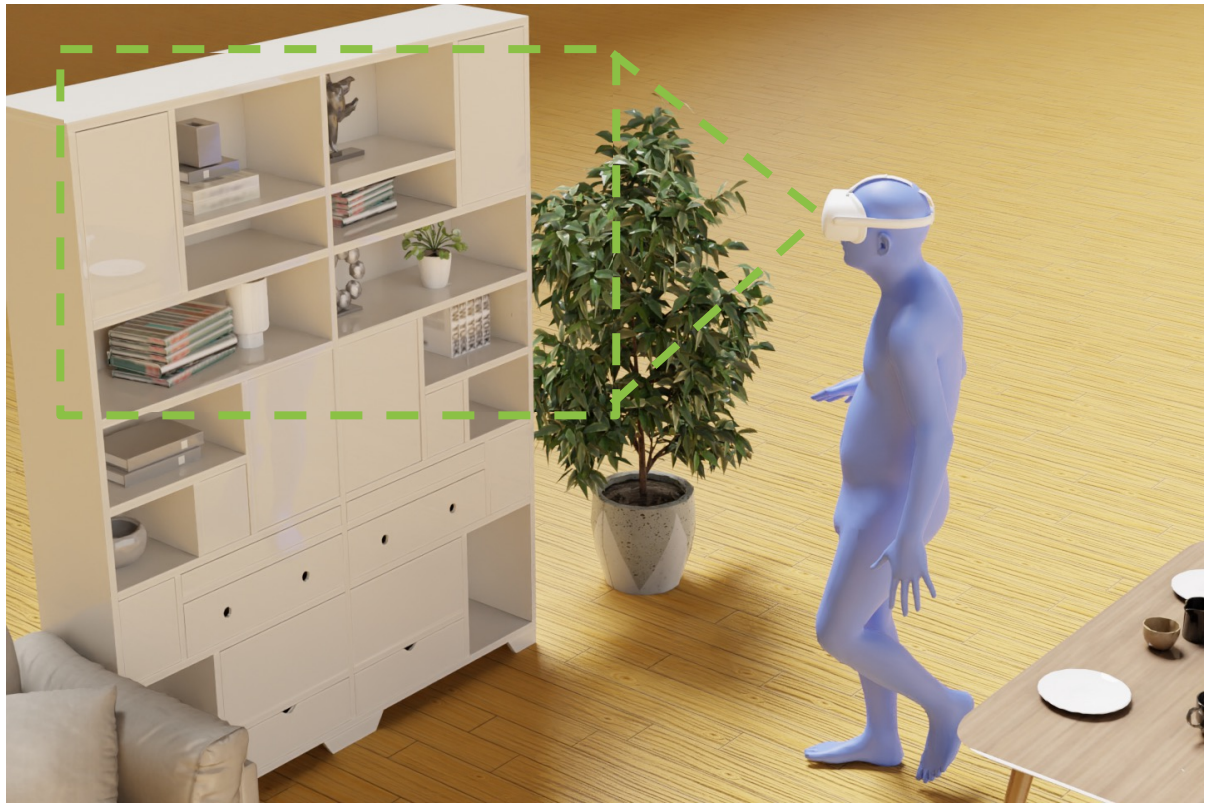
Stanford University

THU-AM-063

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Background

- Human motion estimation from egocentric video is critical to VR/AR applications.
- Enable human motion reconstruction everywhere with a portable device.

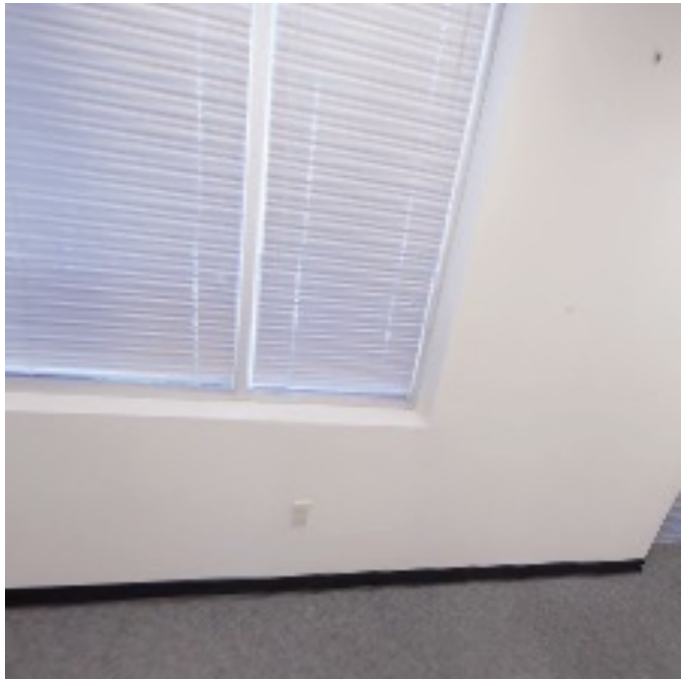


Egocentric View

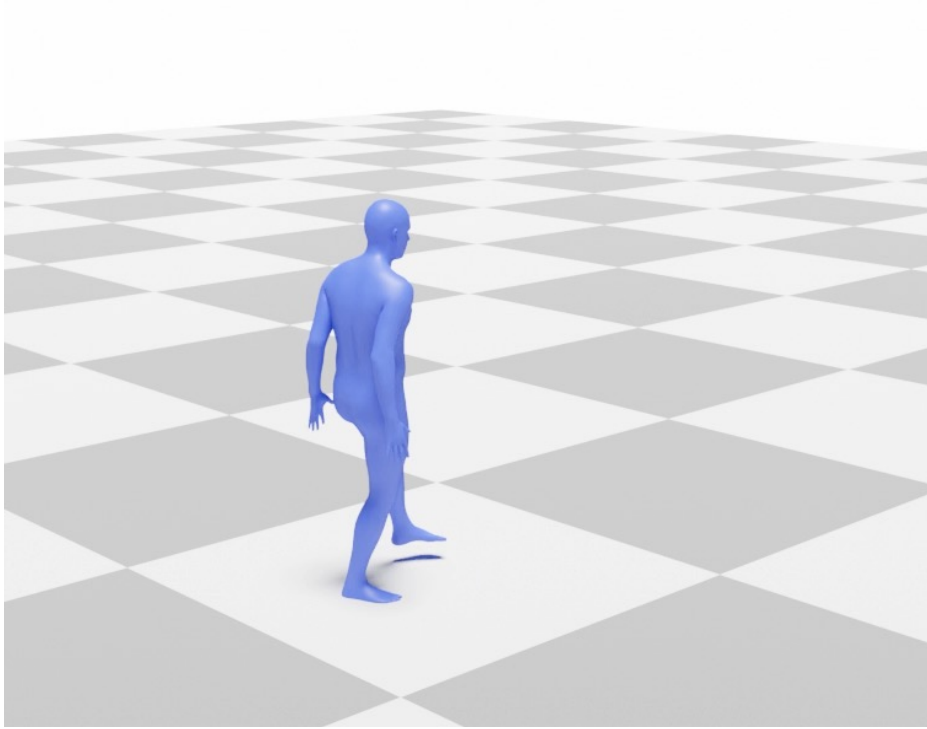
Problem Statement

Input: Egocentric video captured by a front-facing head-mounted camera.

Output: Full-body human motion.

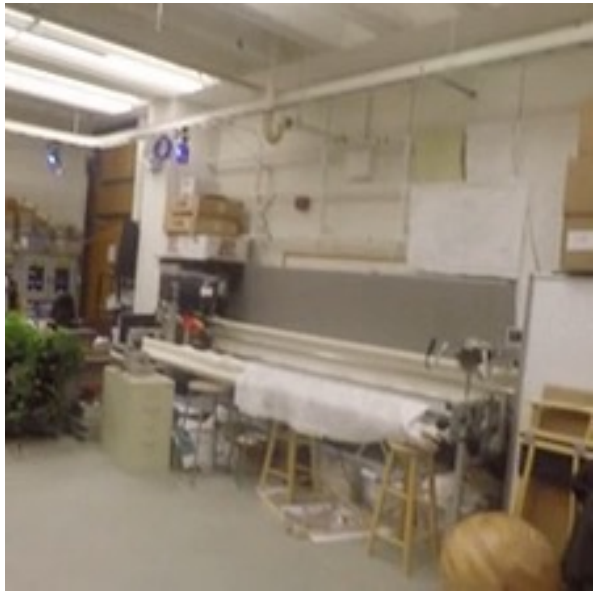


Egocentric Video



Human Motion Sequence

Challenges



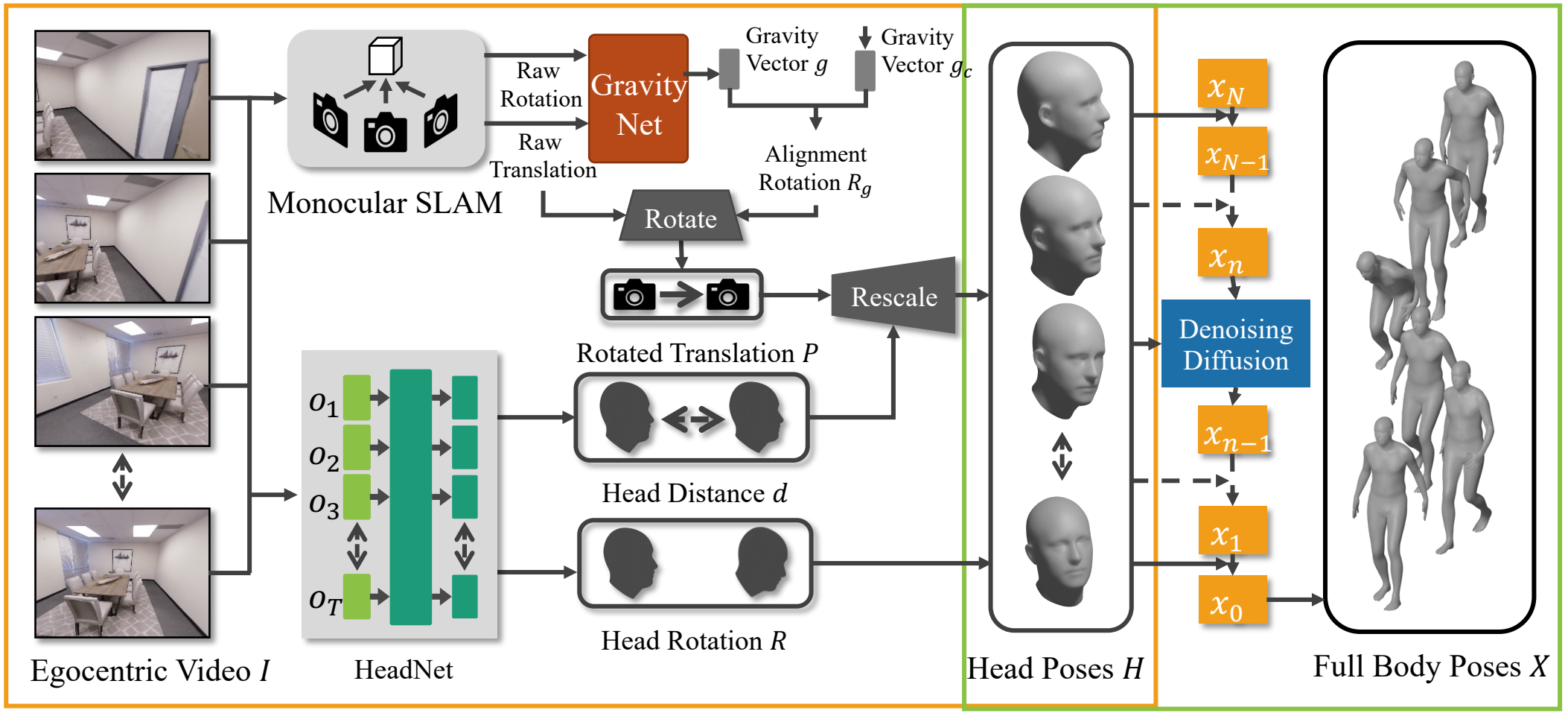
Existing Paired Dataset Limited in Scale and Scene Diversity (the video is from Kinpoly[1])

Full-Body Unobserved
(the video is from GIMO[2])

[1] Luo et al. Dynamics-Regulated Kinematic Policy for Egocentric Pose Estimation. NeurIPS 2021.

[2] Zheng et al. GIMO: Gaze-Informed Human Motion Prediction in Context. ECCV 2022.

Method Overview



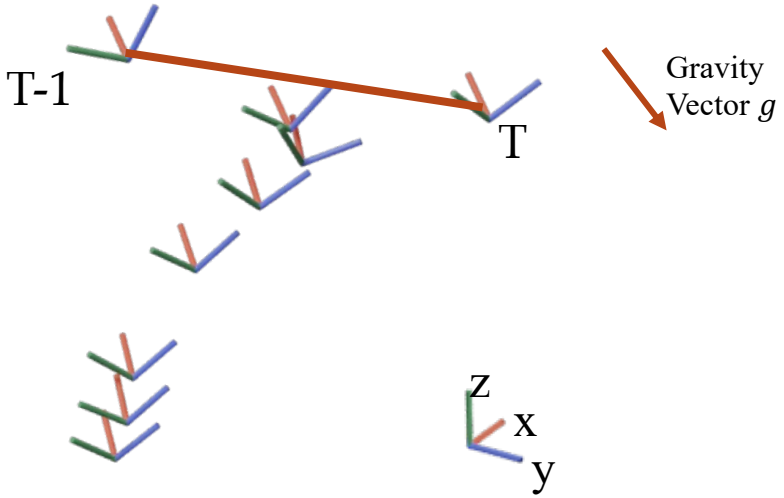
Stage 1: Head Pose Estimation

Stage 2: Full-Body Pose Prediction from Head Pose

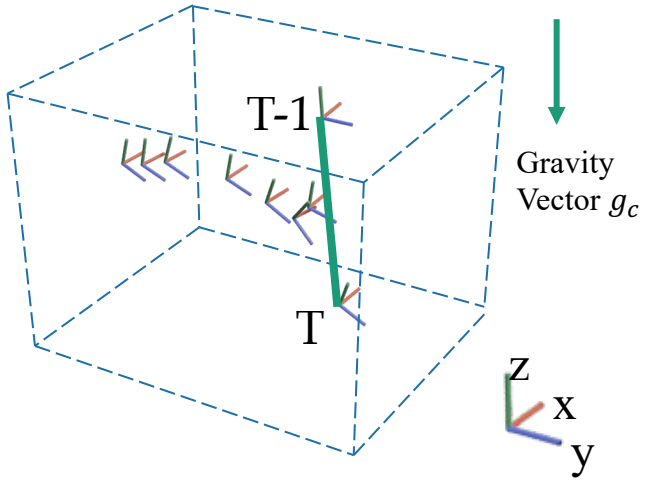
Head Pose Estimation: Monocular SLAM

Problems:

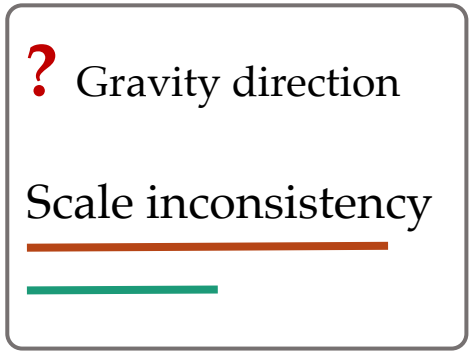
- The gravity direction is unknown.
- The scale of head translation is different from the scale of the real world.



Extracted Head Pose



Desired Head Pose

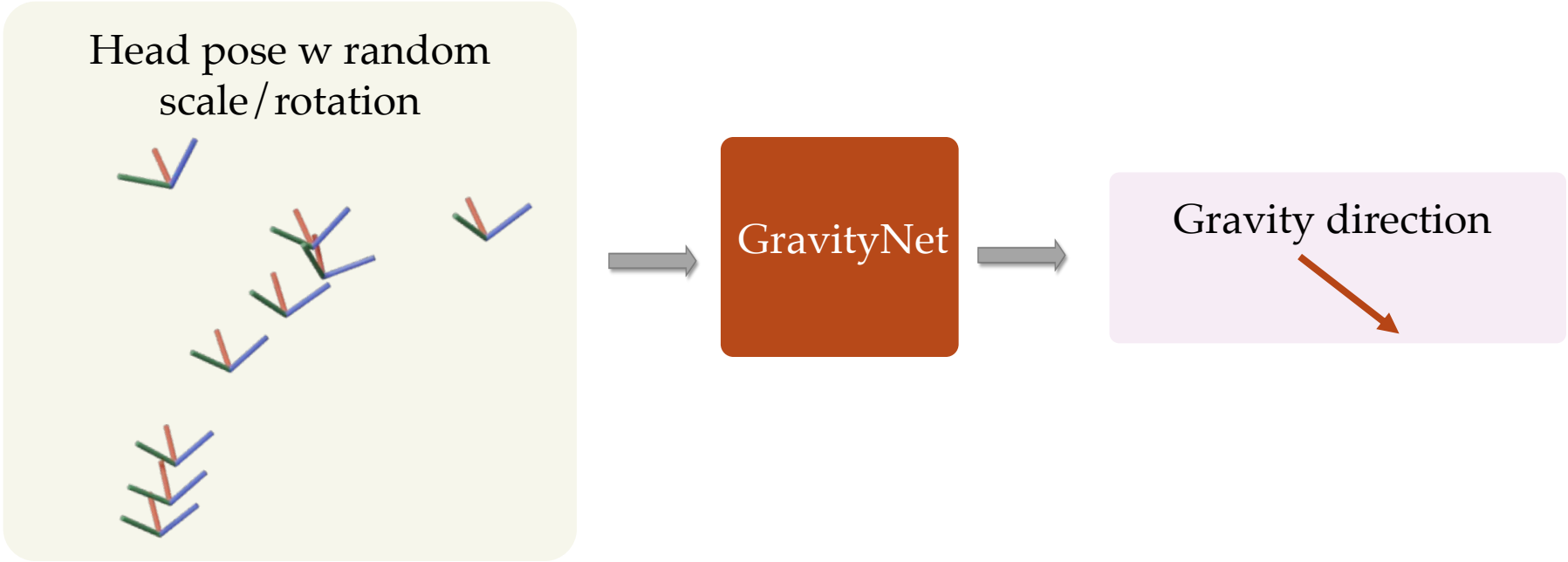


Head Pose Estimation: GravityNet

Input: Head translation and rotation from monocular SLAM.

Output: Gravity direction vector.

Training Dataset: AMASS, apply random scale and rotation to the head pose.

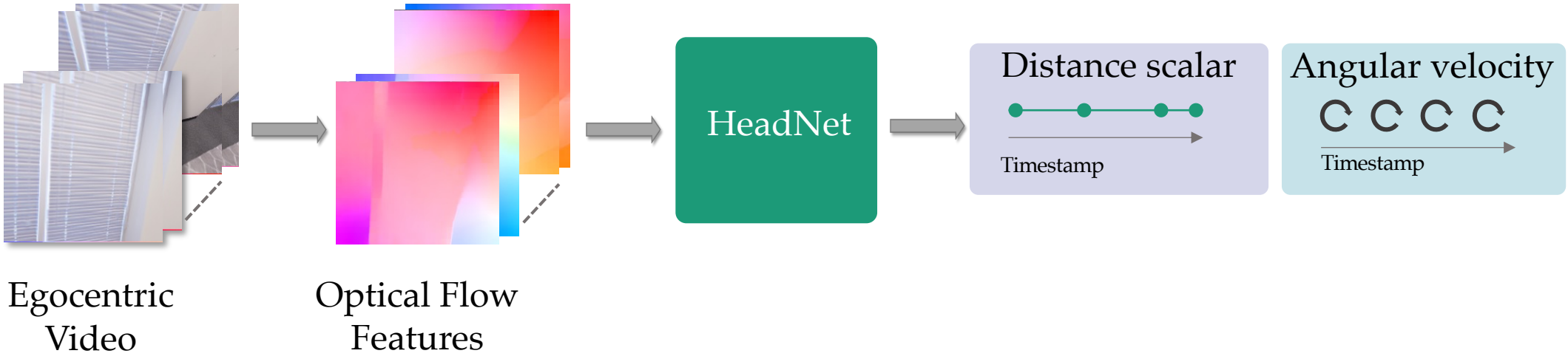


Head Pose Estimation: HeadNet

Input: Optical flow features.

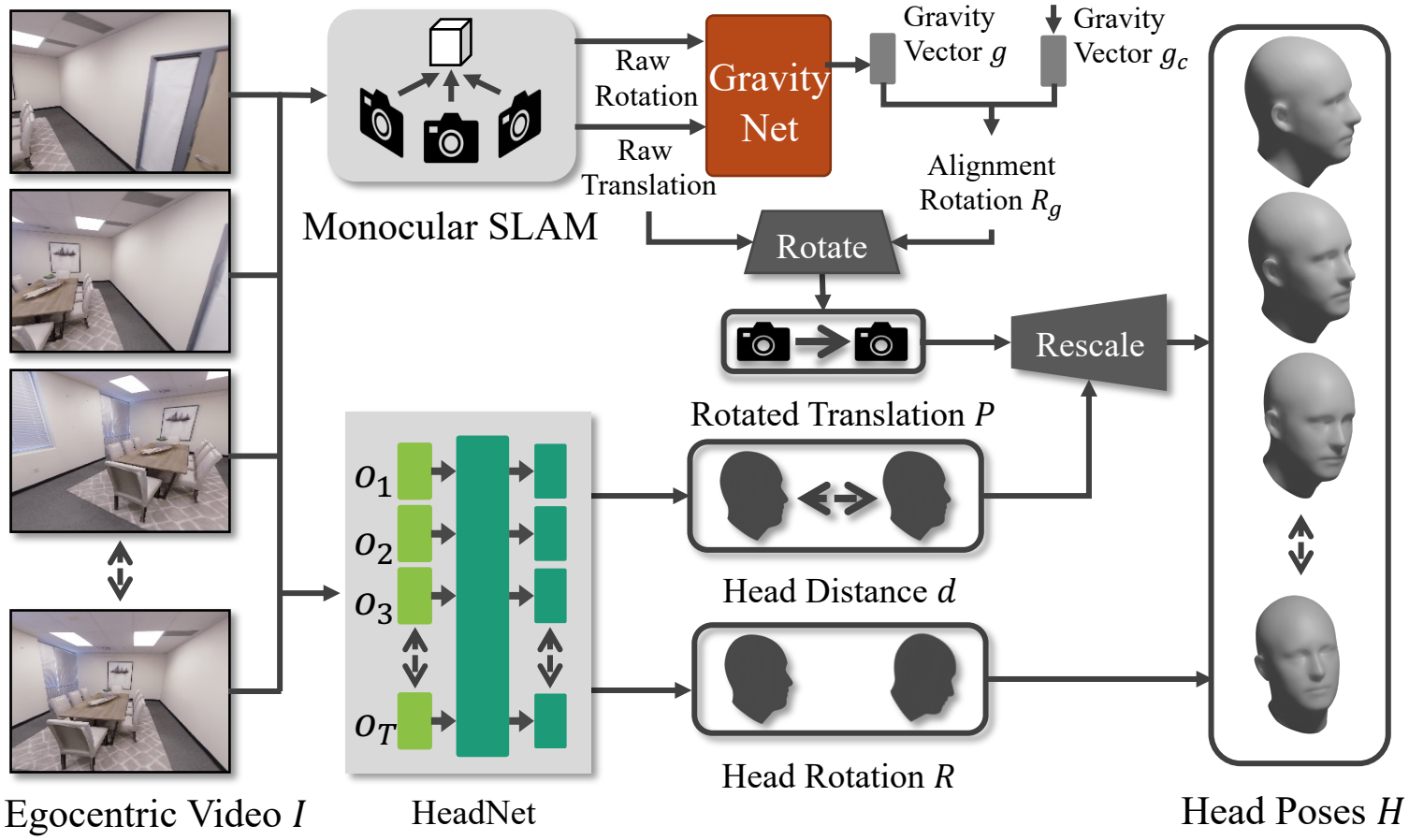
Output: Distance scalar value between every two steps, and angular velocity.

Training dataset: paired egocentric video and head pose.



Head Pose Estimation: A Hybrid Approach

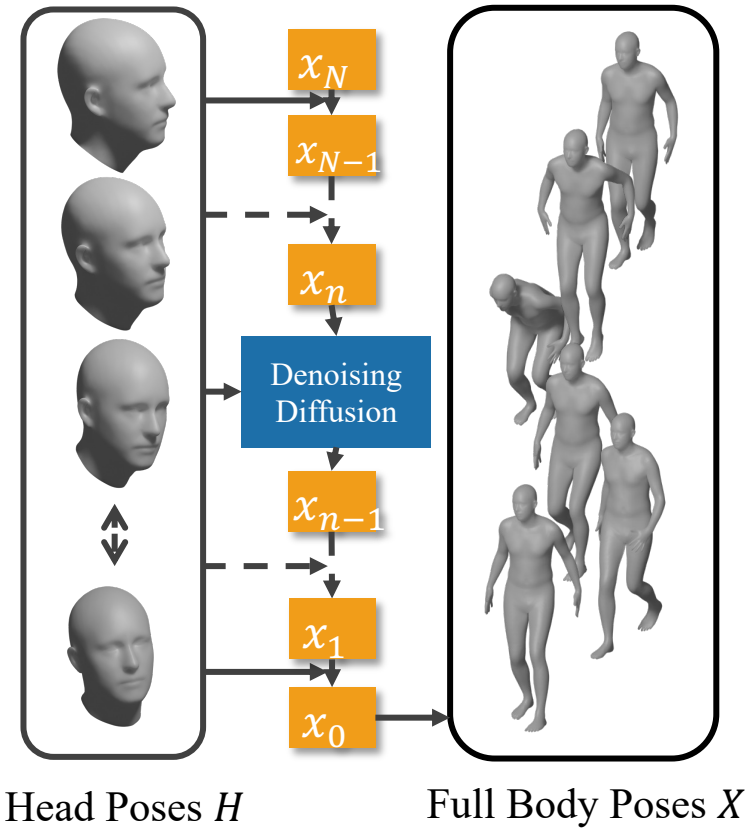
- Our hybrid solution: Monocular SLAM + GravityNet + HeadNet.



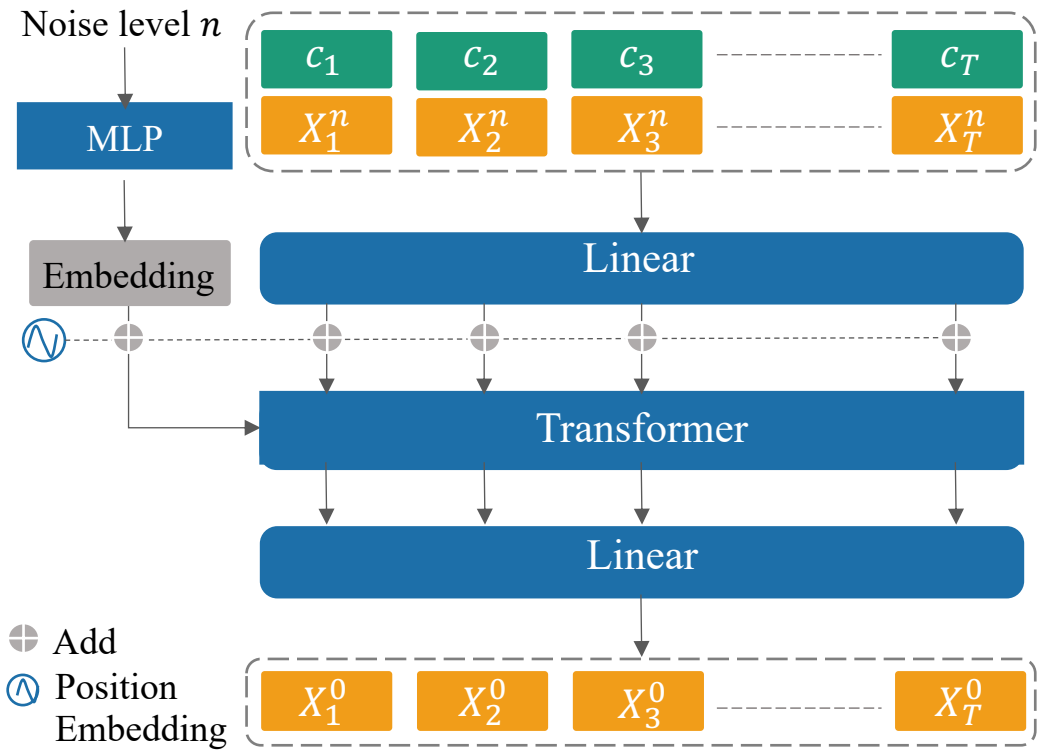
Full-Body Motion Estimation: Conditional Diffusion Model

Input: Predicted head pose from stage 1.

Output: Full-body human motion.



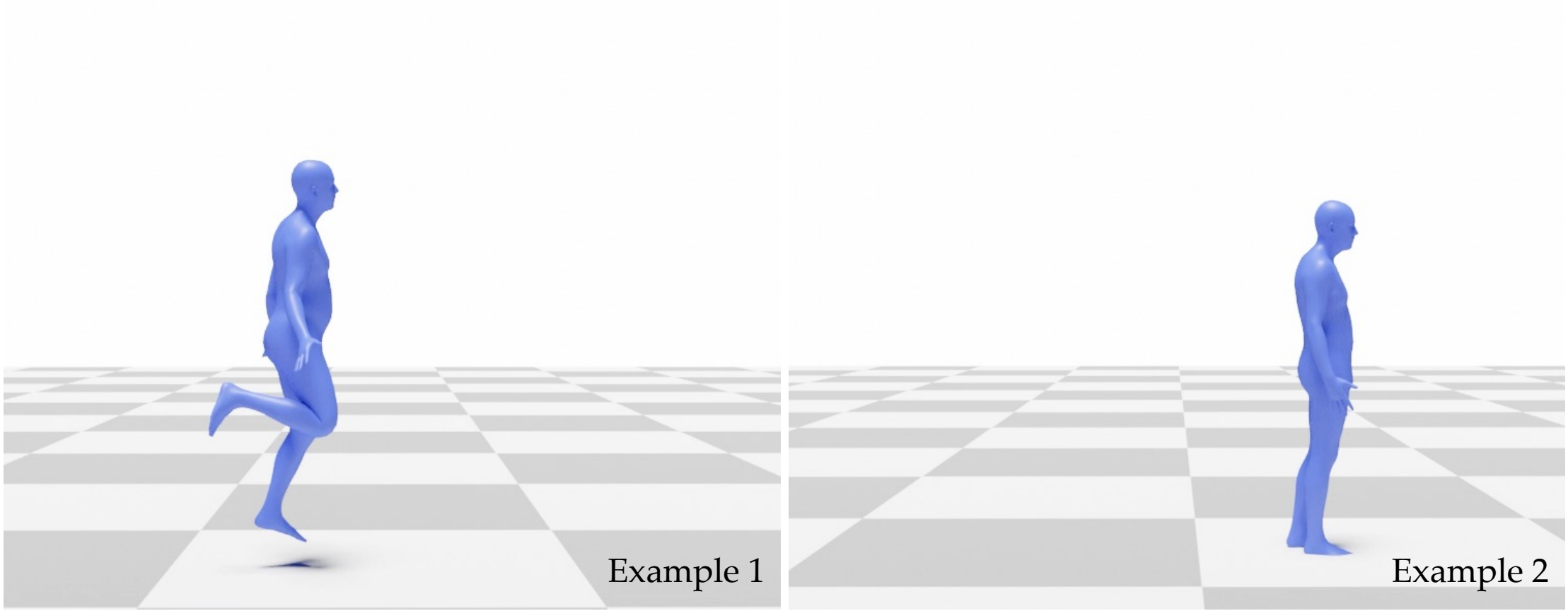
Our Conditional Diffusion Model



Model Architecture of the Denoising Network

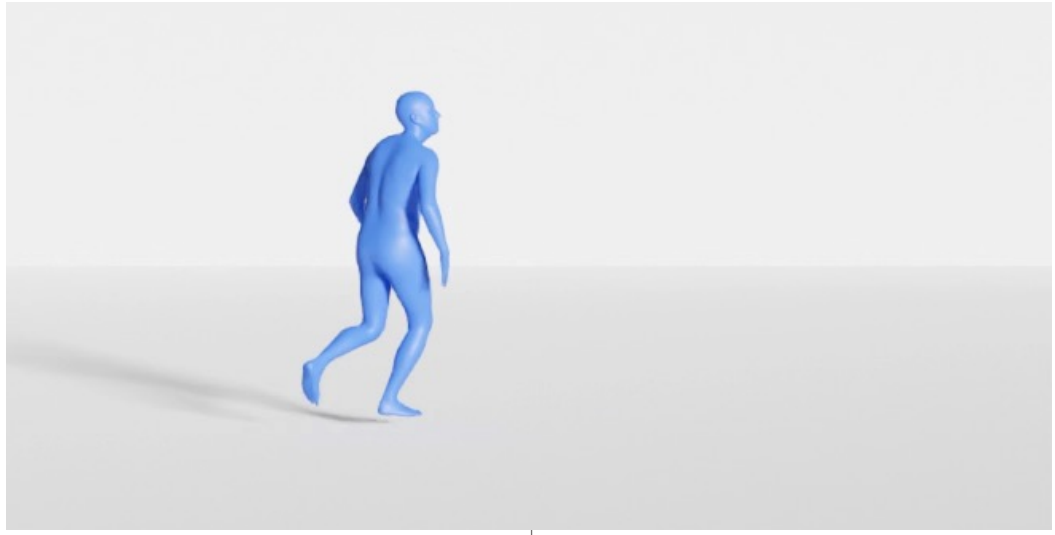
Full-Body Motion Estimation: Conditional Diffusion Model

Results of our full-body motion generation given ground truth head pose as input.

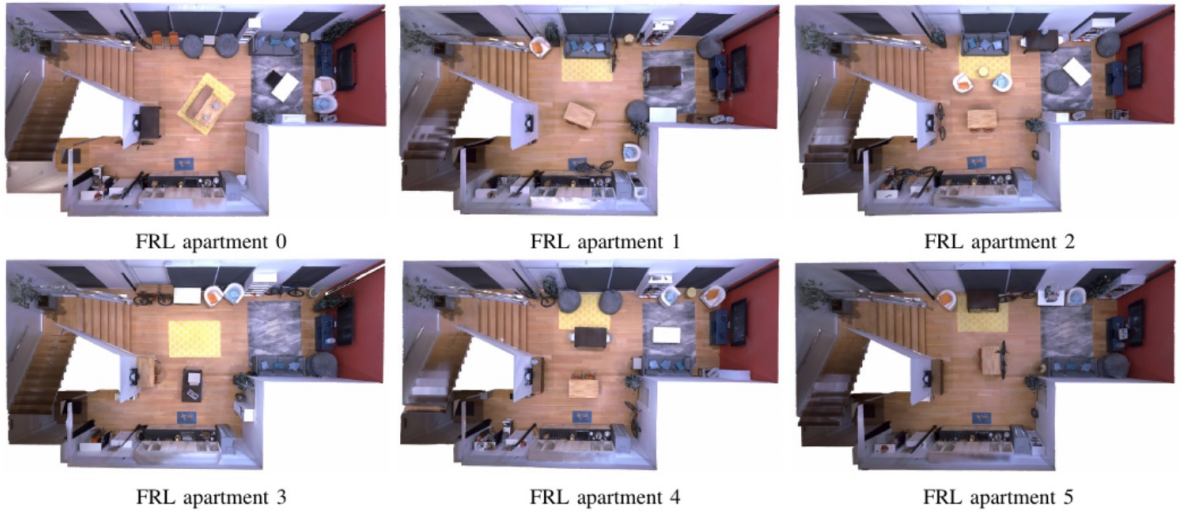


Synthetic Dataset Generation

Human Motion Dataset: AMASS



3D Scene Dataset: Replica



Detect Penetration

Habitat

AMASS-Replica-Ego-Syn Dataset (ARES)

Paired egocentric video and full-body motion.

Example from ARES

Ego-View

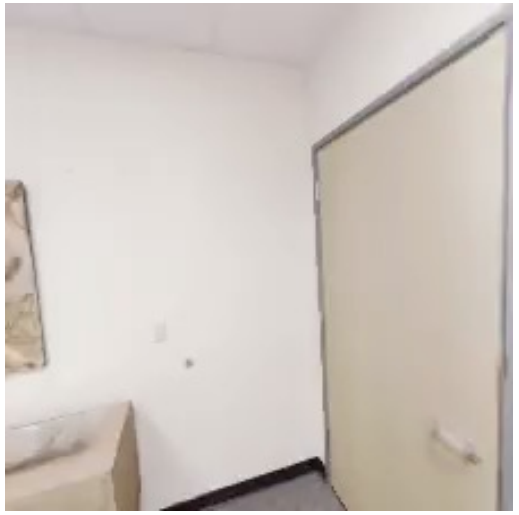
Ground Truth Motion

Example from ARES

Ego-View

Ground Truth Motion

More Examples of Egocentric Video





Quantitative Evaluation of Full-Body Pose Estimation from Egocentric Video

- Evaluation on ARES.

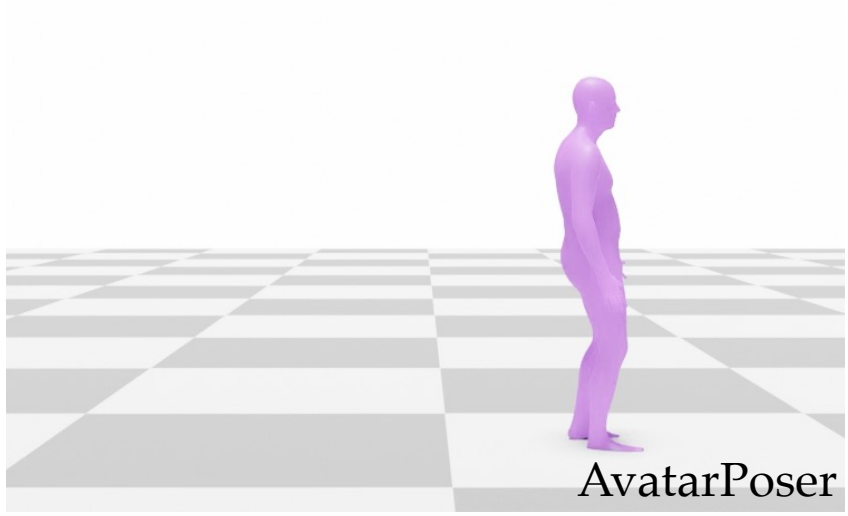
	O_{head}	T_{head}	MPJPE	Accel	FS
PoseReg[1]	0.77	354.7	147.7	127.6	87.1
Kinpoly-OF[2]	0.62	323.4	141.6	7.3	4.2
EgoEgo	**0.20**	**148.0**	**121.1**	**6.2**	**2.7**

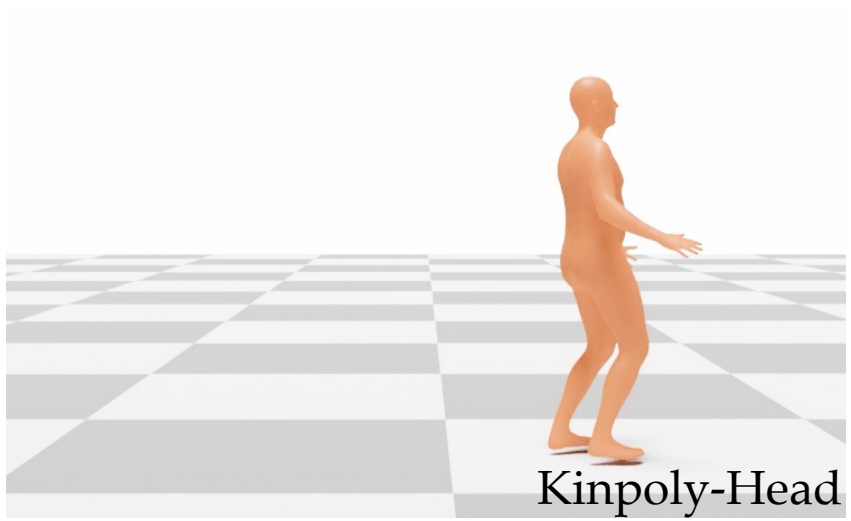
- Evaluation on real-world datasets.

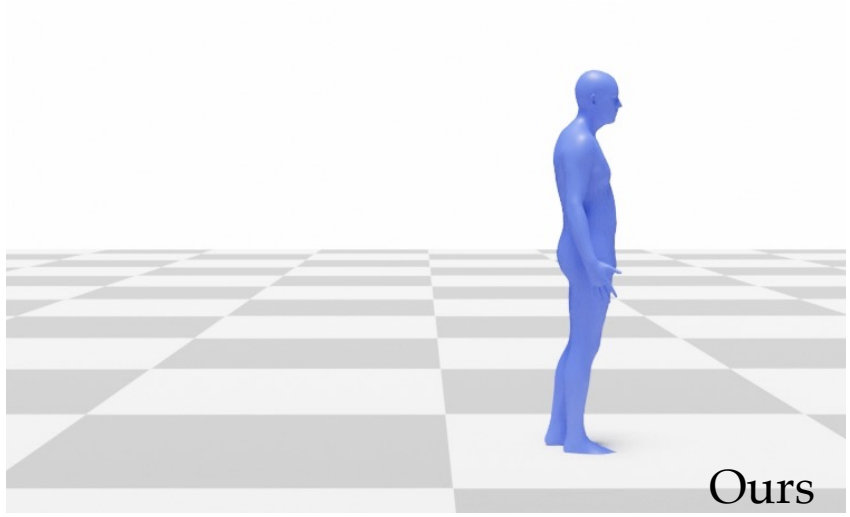
	Kinpoly-MoCap[2]					**GIMO[3]**				
	O_{head}	T_{head}	MPJPE	Accel	FS	O_{head}	T_{head}	MPJPE	Accel	FS
PoseReg[1]	1.05	1943.9	160.4	61.8	10.8	1.51	1528.6	189.3	71.5	14.2
Kinpoly-OF[2]	1.33	2475.5	230.5	16.4	15.8	1.52	1739.3	404.2	21.9	14.4
EgoEgo	**0.58**	**505.1**	**125.9**	**8.0**	**1.6**	**0.67**	**356.8**	**152.1**	**10.4**	**1.9**

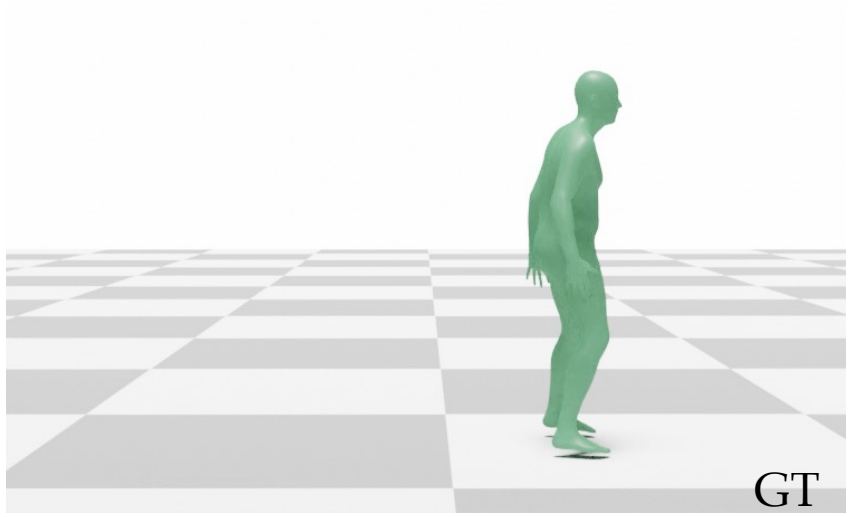
[1] Yuan et al. Ego-Pose Estimation and Forecasting as Real-Time PD Control. ICCV 2019.
[2] Luo et al. Dynamics-Regulated Kinematic Policy for Egocentric Pose Estimation. NeurIPS 2021.
[3] Zheng et al. GIMO: Gaze-Informed Human Motion Prediction in Context. ECCV 2022.

Comparisons of Full-Body Estimation from Head Pose



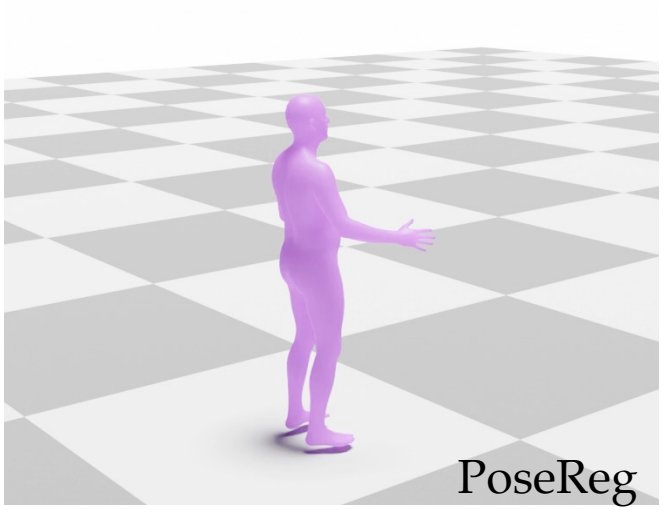


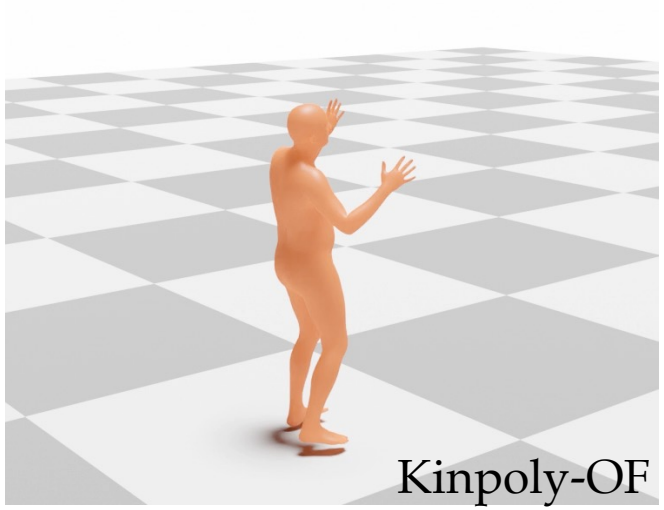


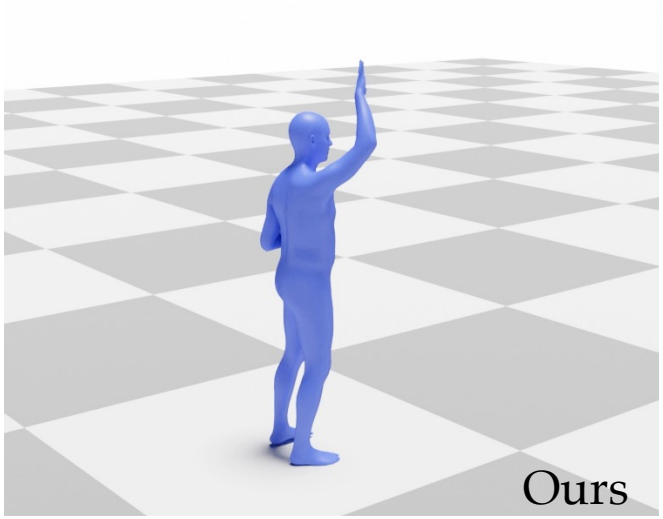


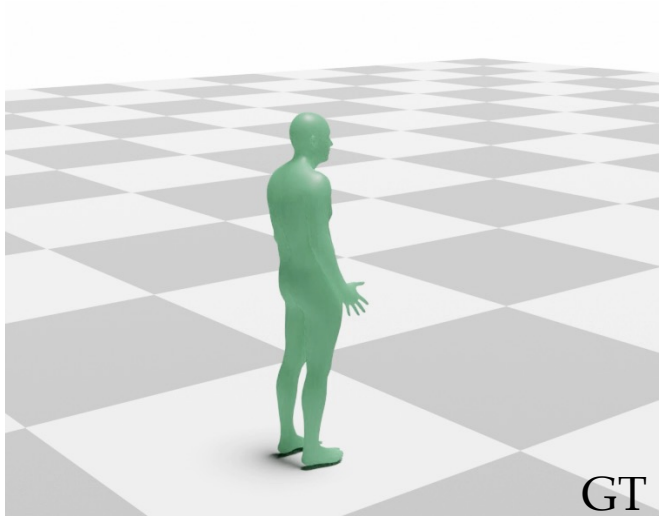
Comparisons of Full-Body Estimation from Egocentric Video

Egocentric Video Input







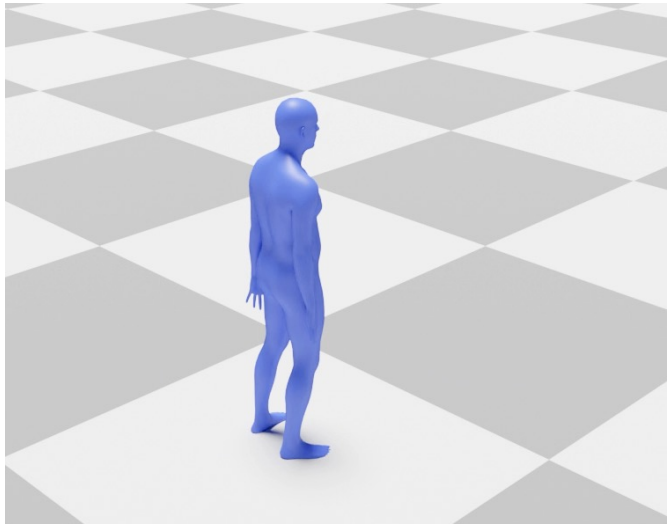


More Results

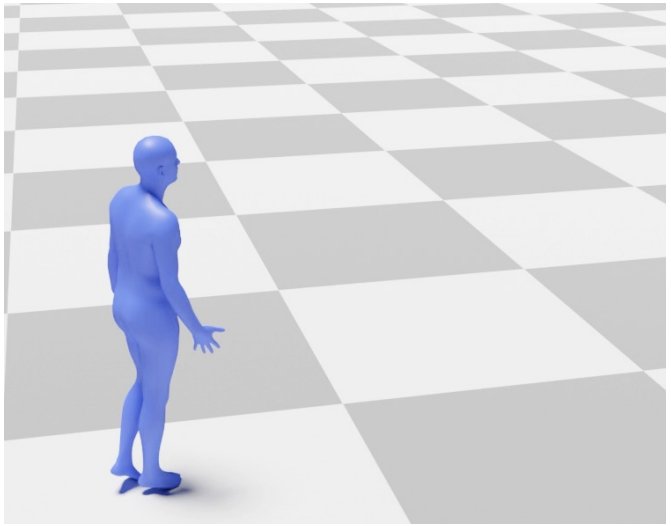
Input

Output

ARES
(our synthetic dataset)



GIMO
(real-world dataset)



Ego-Body Pose Estimation via Ego-Head Pose Estimation

Jiaman Li, C. Karen Liu†, Jiajun Wu†

(†indicates equal contribution)

Stanford University

THU-AM-063

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA