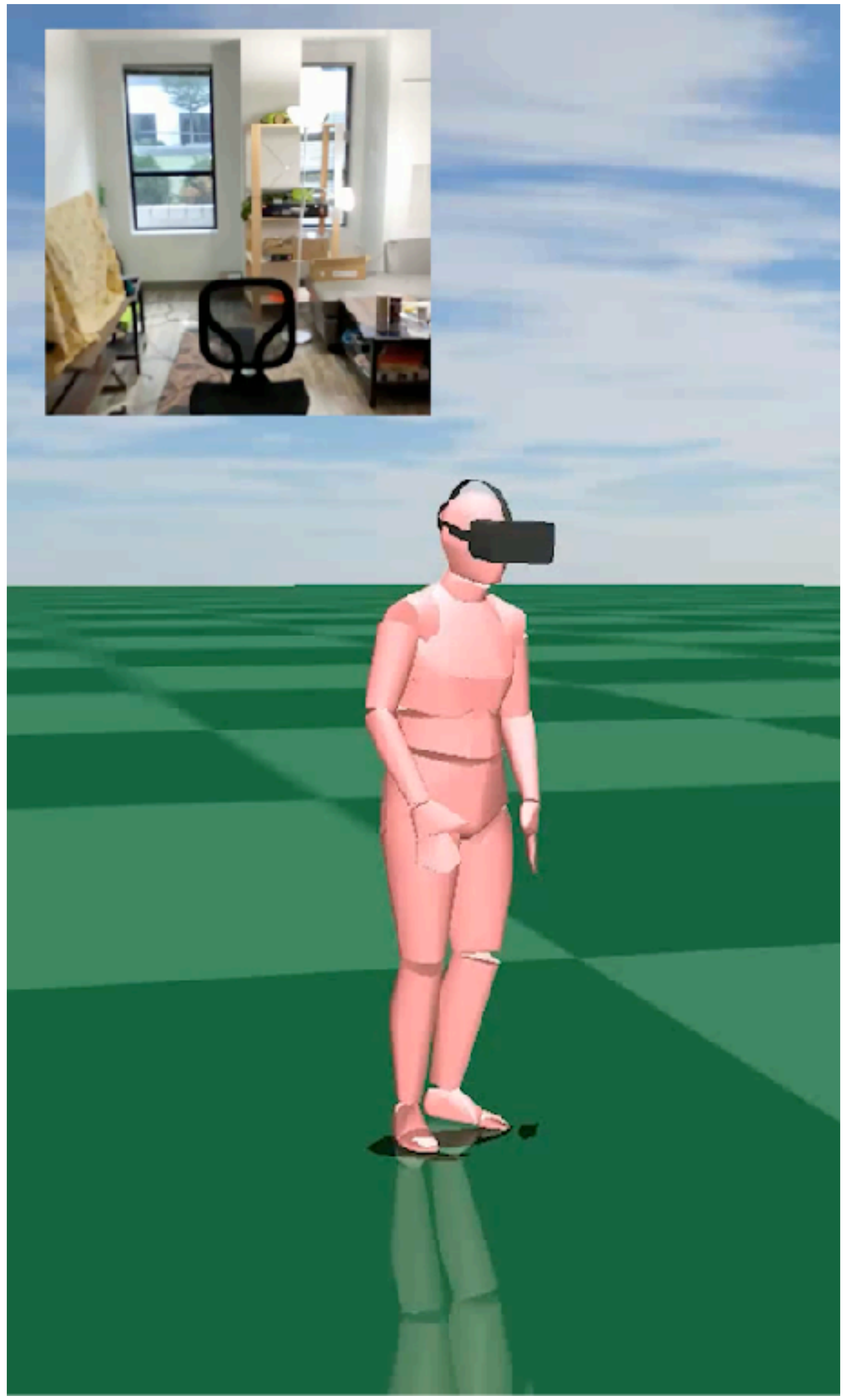
Dynamics-Regulated Kinematic Policy for Egocentric Pose Estimation

Inferring physically valid human pose and human-object interactions from wearable headsets

https://zhengyiluo.github.io/projects/kin_poly/

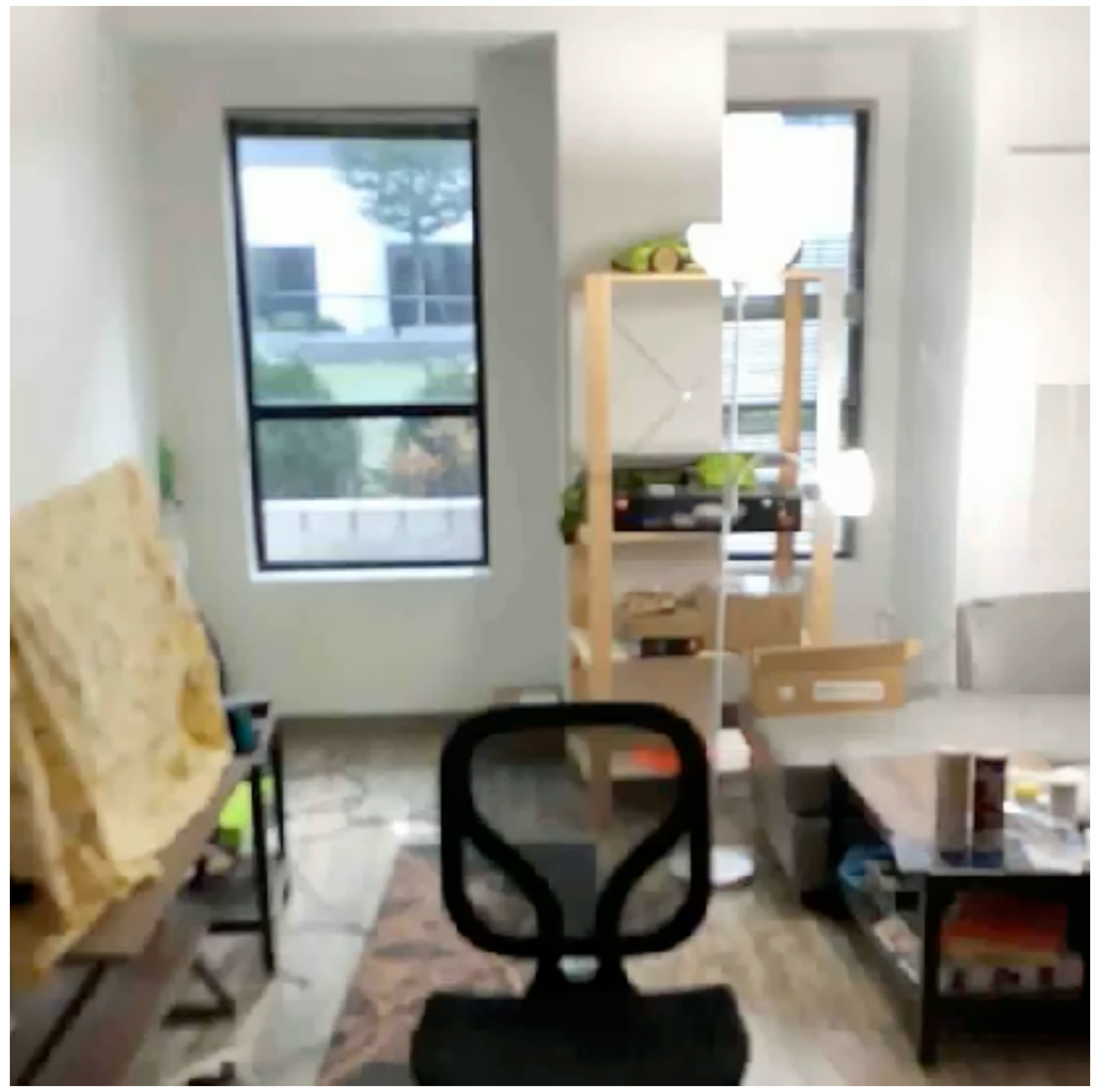
The Robotics Institute, Carnegie Mellon University

Zhengyi Luo, Ryo Hachiuma, Ye Yuan, Kris Kitani



Estimating 3D Human Pose & Human-Object Interaction from Egocentric Videos

From a video captured by a single head-mounted wearable camera (i.e. smart-glasses, action camera, body camera), we want to **infer and simulate** the wearer's 3D pose and interaction with objects in the scene



Input: egocentric video

Output: 3D human pose

Reference 3rd person pose

$$\boldsymbol{I}_{1:T}(\text{Input video frames}) \rightarrow \widehat{\boldsymbol{q}}_{1:T}(\text{3D human pose})$$

Model overview

Kinematics-based approach

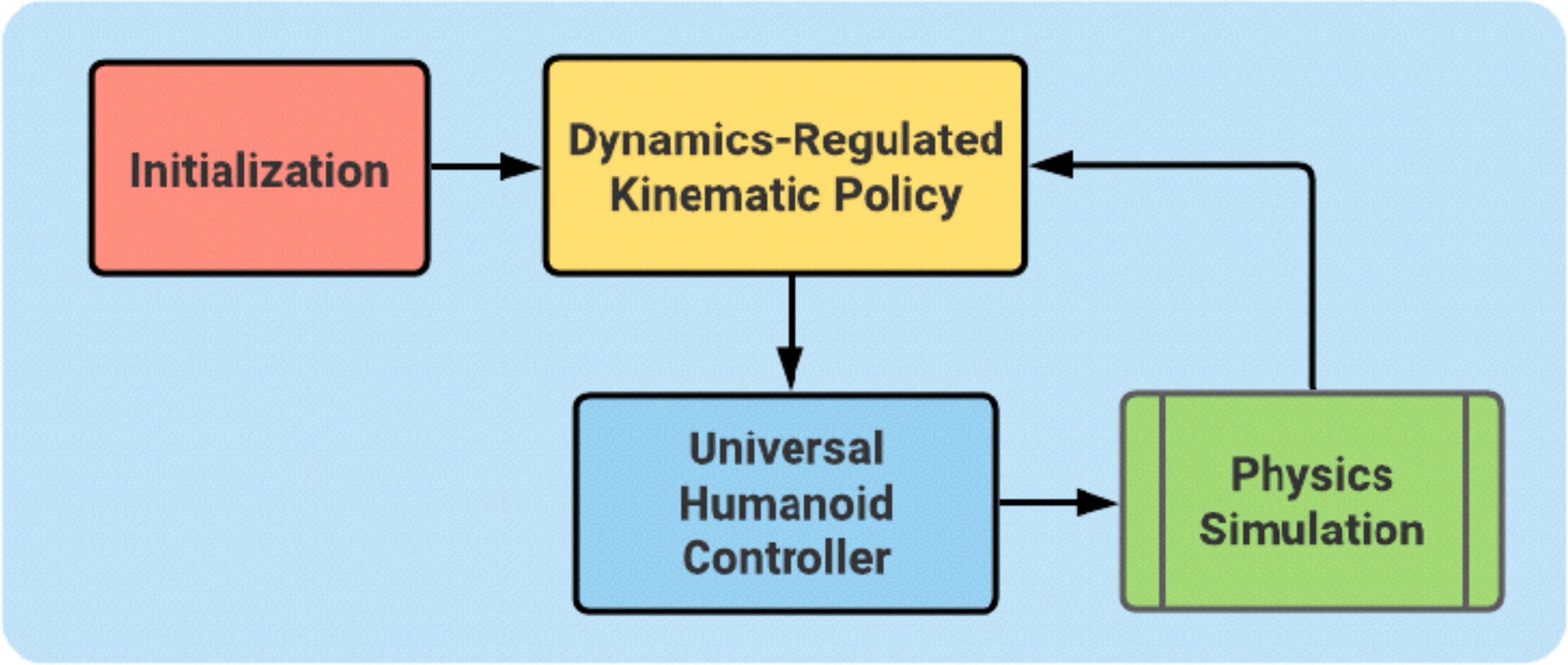
Study motion without regard to forces that cause it

Dynamics-based approach

Study motion that result from forces

- **Universal humanoid Controller**: **task agnostic** physics-based humanoid controller
- **Dynamics-regulated kinematic policy** inside a physics simulation (Mujoco)
  - Initialization module for estimating object pose and image features
  - Per-step policy for causal physics-based pose estimation

Egocentric Input

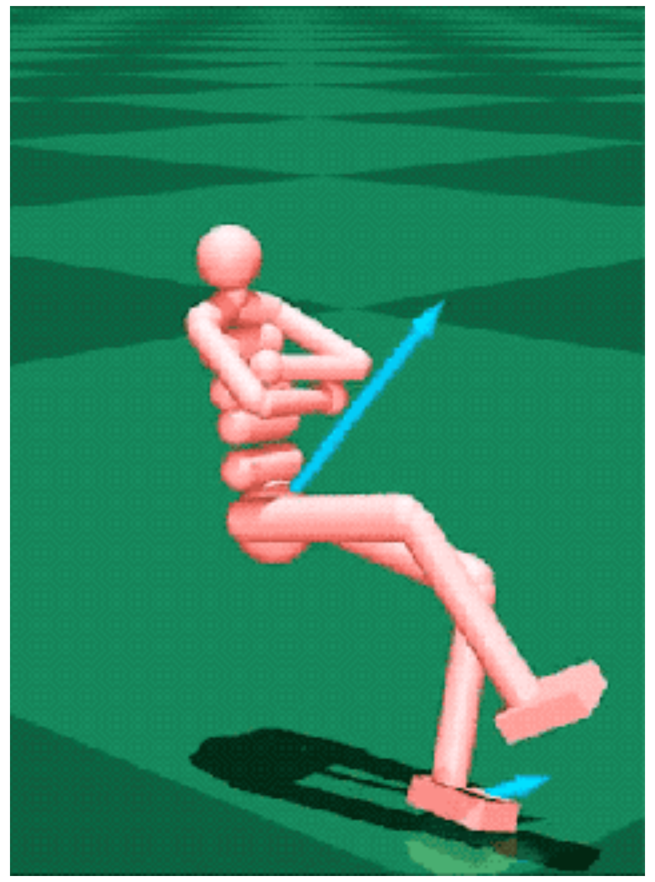


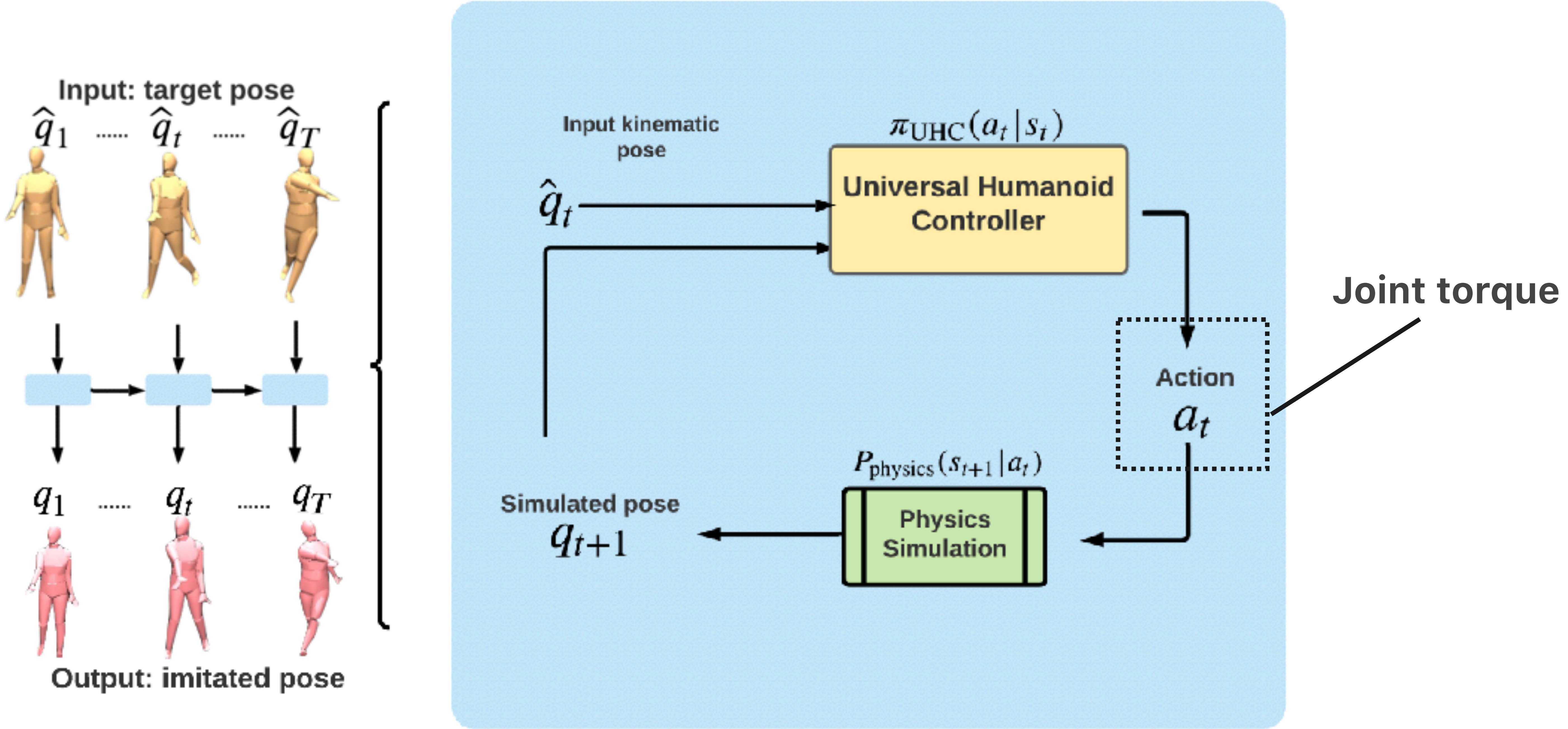
Pose Estimation Result

Universal Humanoid Controller

- **Task agnostic** physics-based humanoid controller
- Input: next-frame target pose; output: control signals
- Trained using Reinforcement Learning
- Compatible with SMPL:
  - Able to perform **97% of sequences** from the AMASS dataset

Residual Force Control
[Yuan et al, NeruIPS 2020]

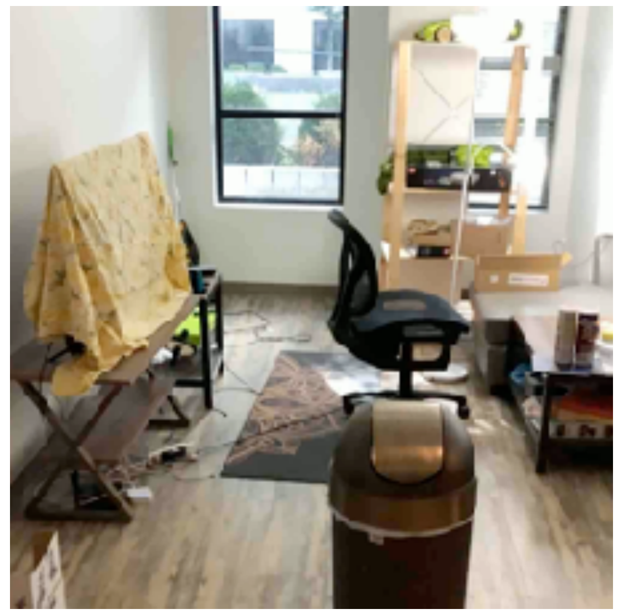




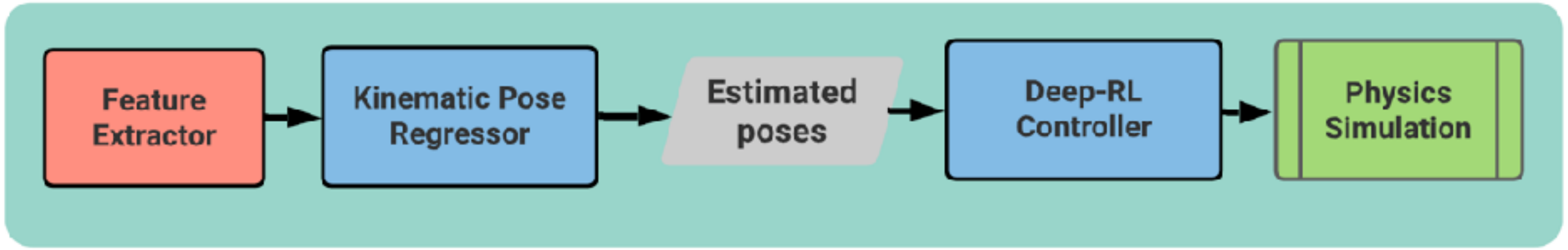

Given an input sequence of target motion

Dynamics-regulated Kinematic Policy

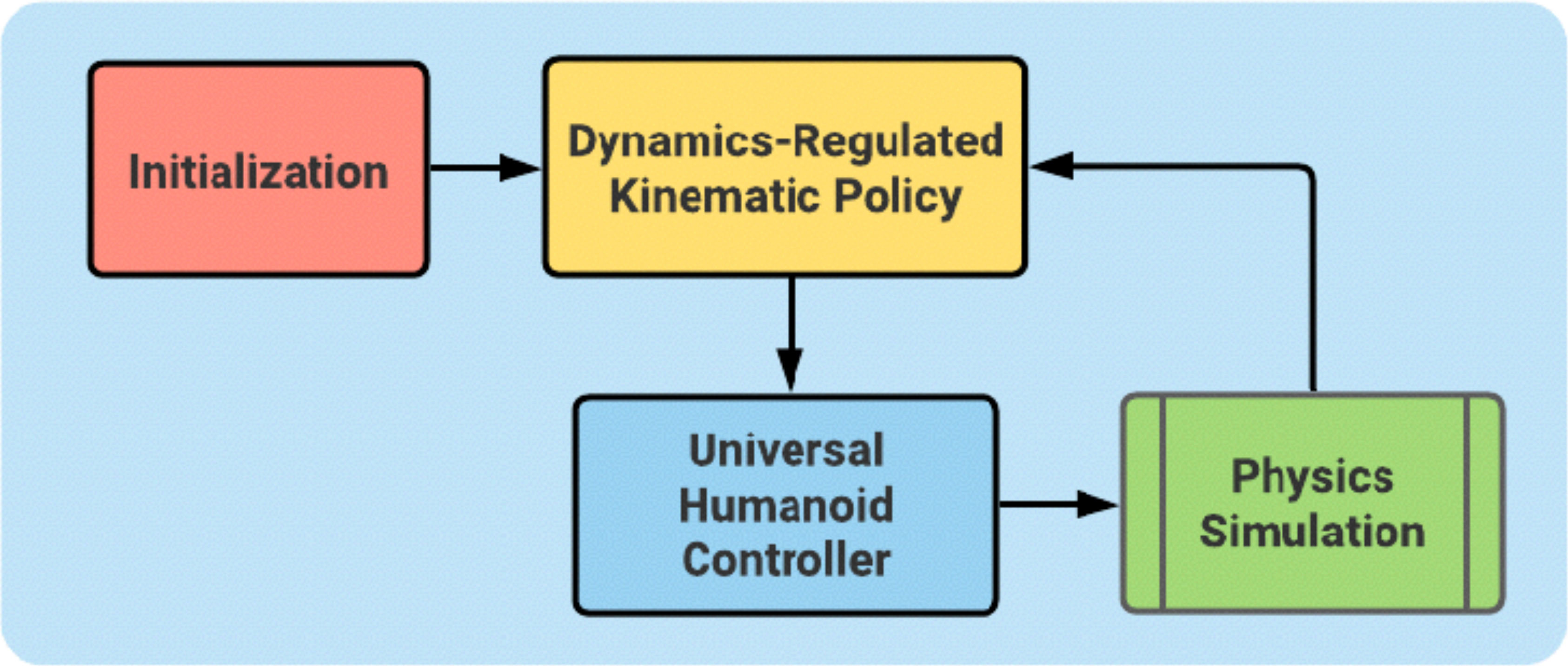
- **Utilizing the humanoid controller:**
  - Naive Approach: physics simulation as post processing



Kinematics and dynamics stays disjoint!

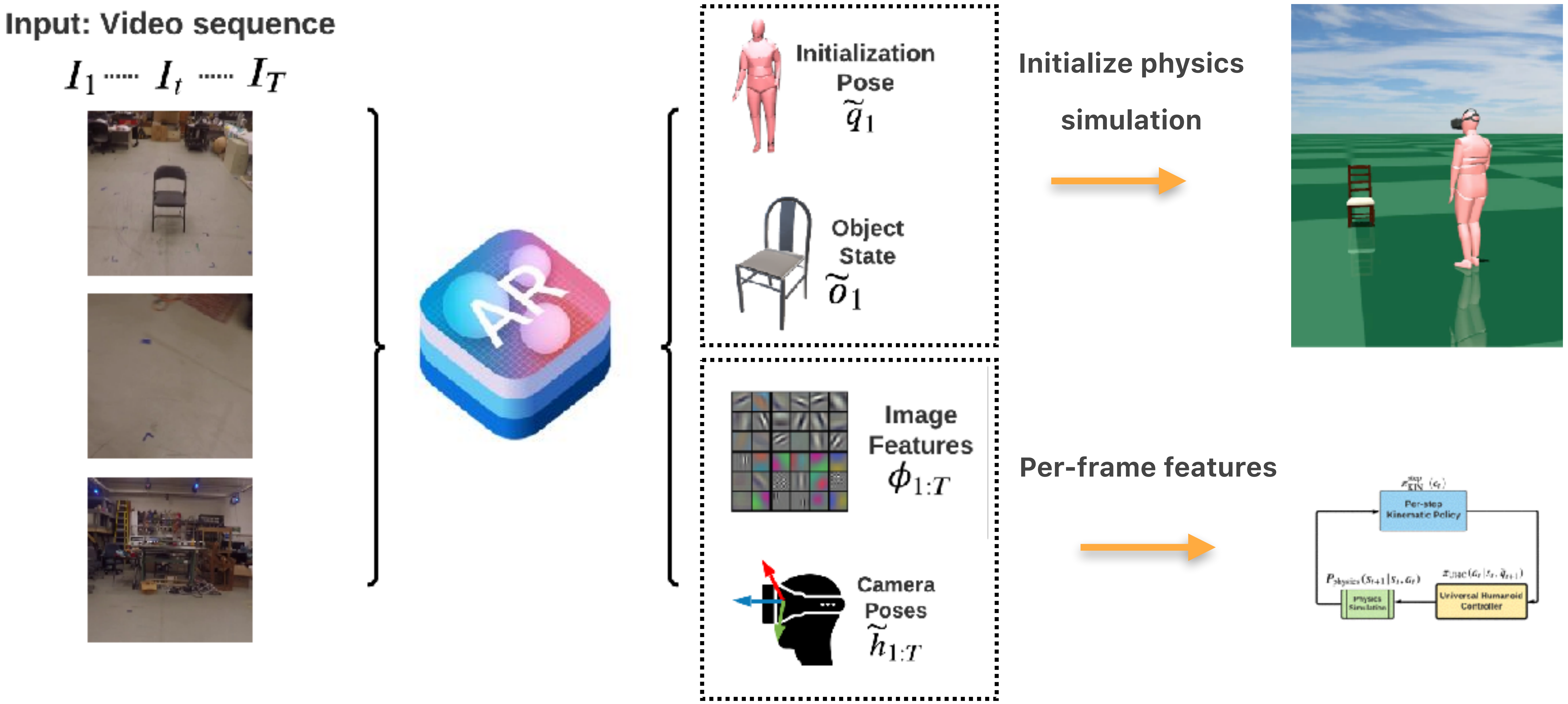


- Our proposal: synergize kinematics and dynamics through dynamics-regulated kinematic policy



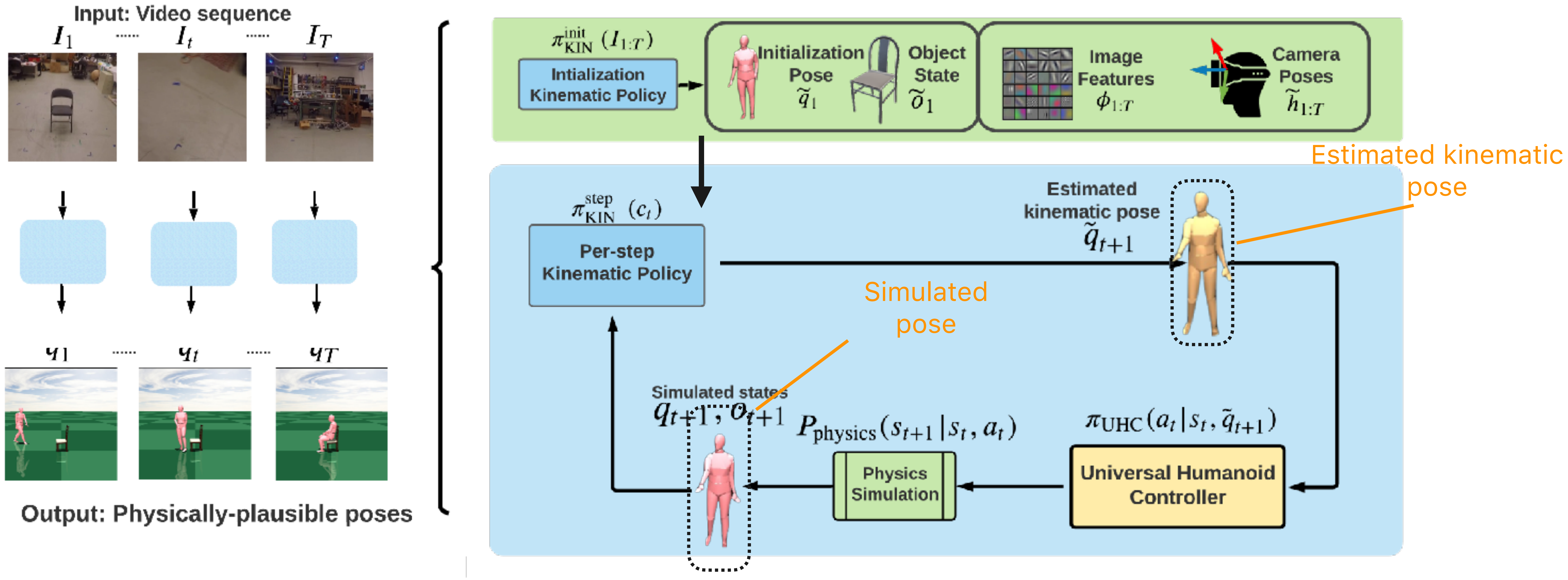
Dynamics-regulated Kinematic Policy

- **Initialization module:**
  - Off-the-shelf camera pose and object pose extractor
  - Optical-flow image feature extractor
  - Computes first-frame humanoid pose



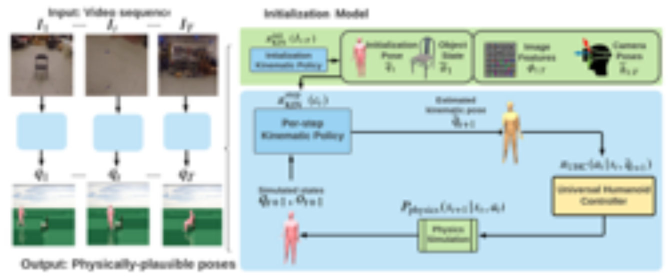
Dynamics-regulated Kinematic Policy

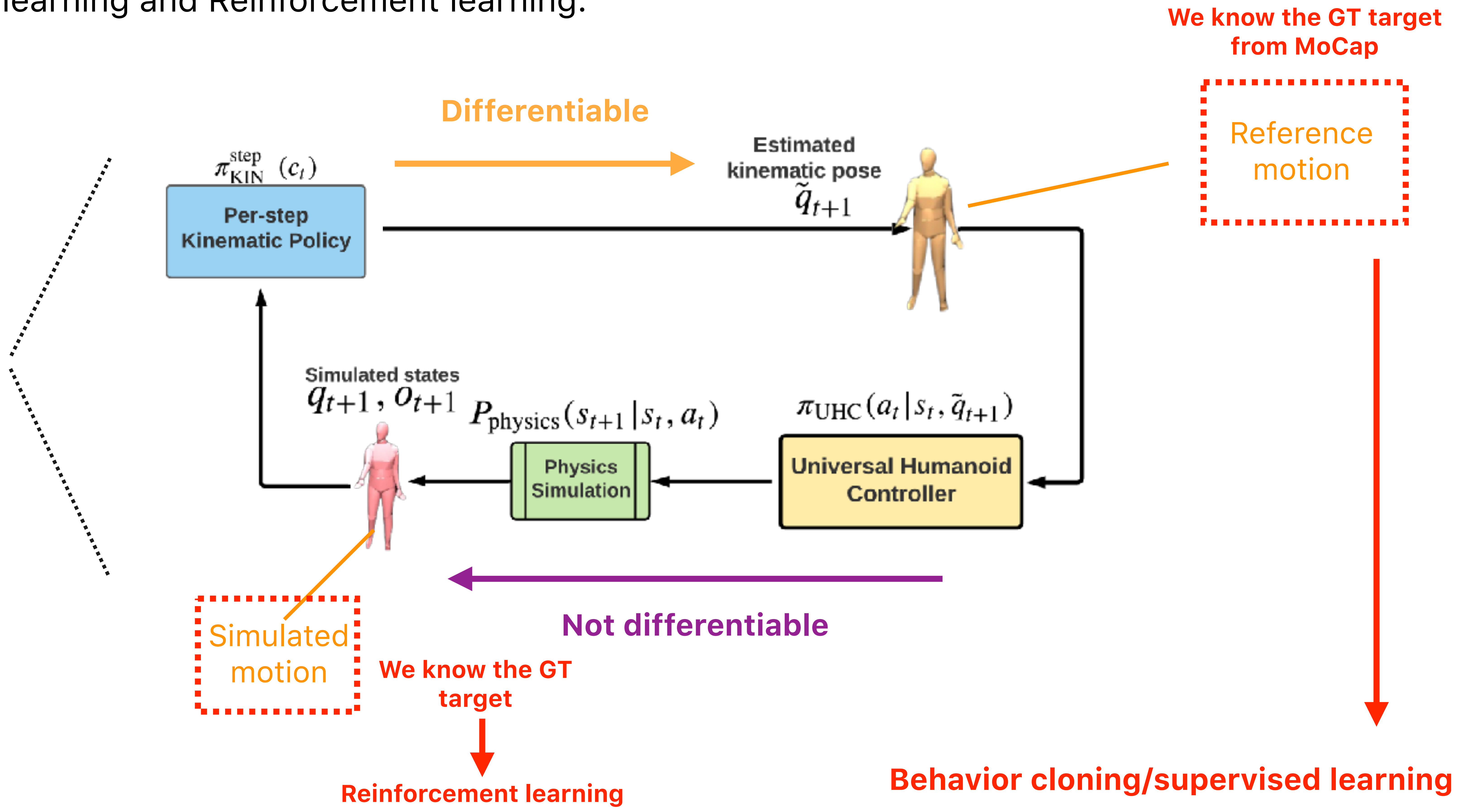
- **Per-step model**
  - Input: current humanoid pose from **physics simulation** and image features
  - Output: per-step target pose for universal humanoid controller
  - Closed loop system with pose estimation and control



Dynamics-regulated Kinematic Policy

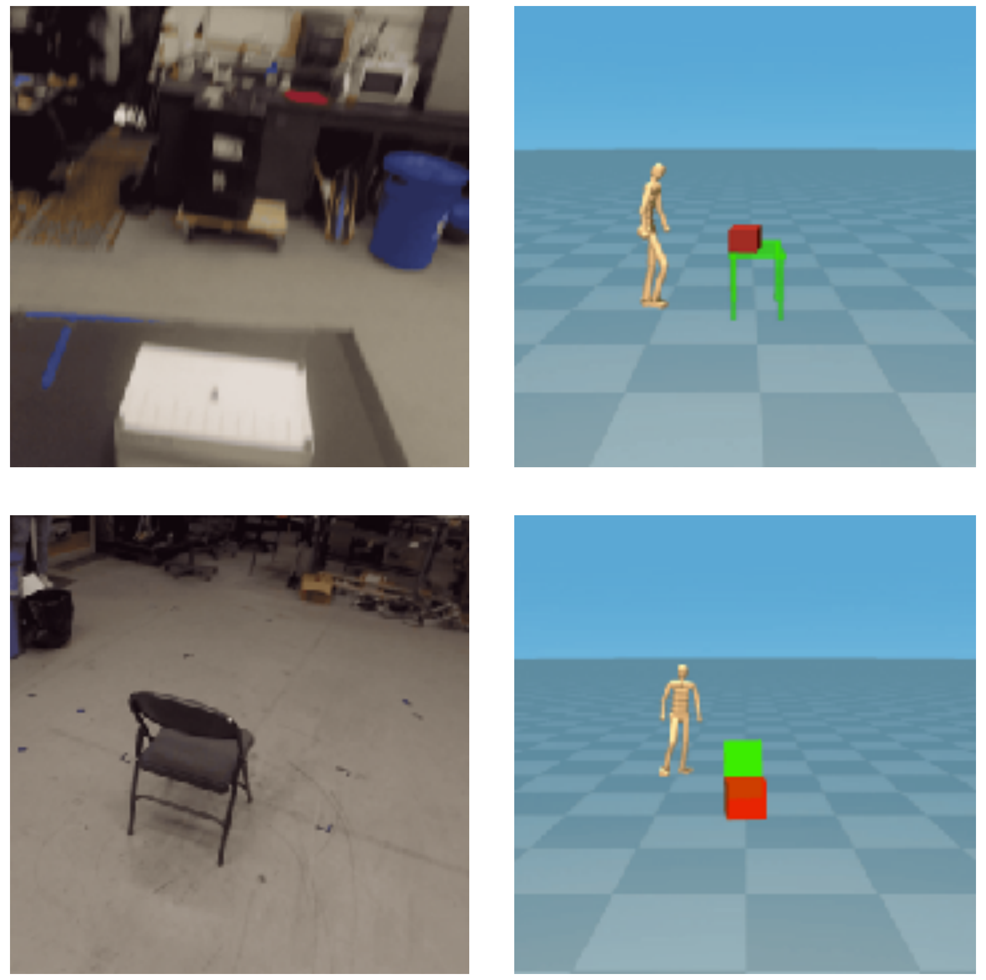
- **Optimization: dynamics regulated training**
  - Mixture of Supervised learning and Reinforcement learning.





Evaluation

Egocentric video GT MoCap



Mocap Data

- 3 Subjects, paired egocentric videos, human pose, and object pose
- 266 takes in total, 6-10 seconds
- Actions:
  - Sitting down/Standing up, avoiding obstacles, pushing a box, stepping on a box.
- Captured in a mocap studio
- 8:2 split for training and testing

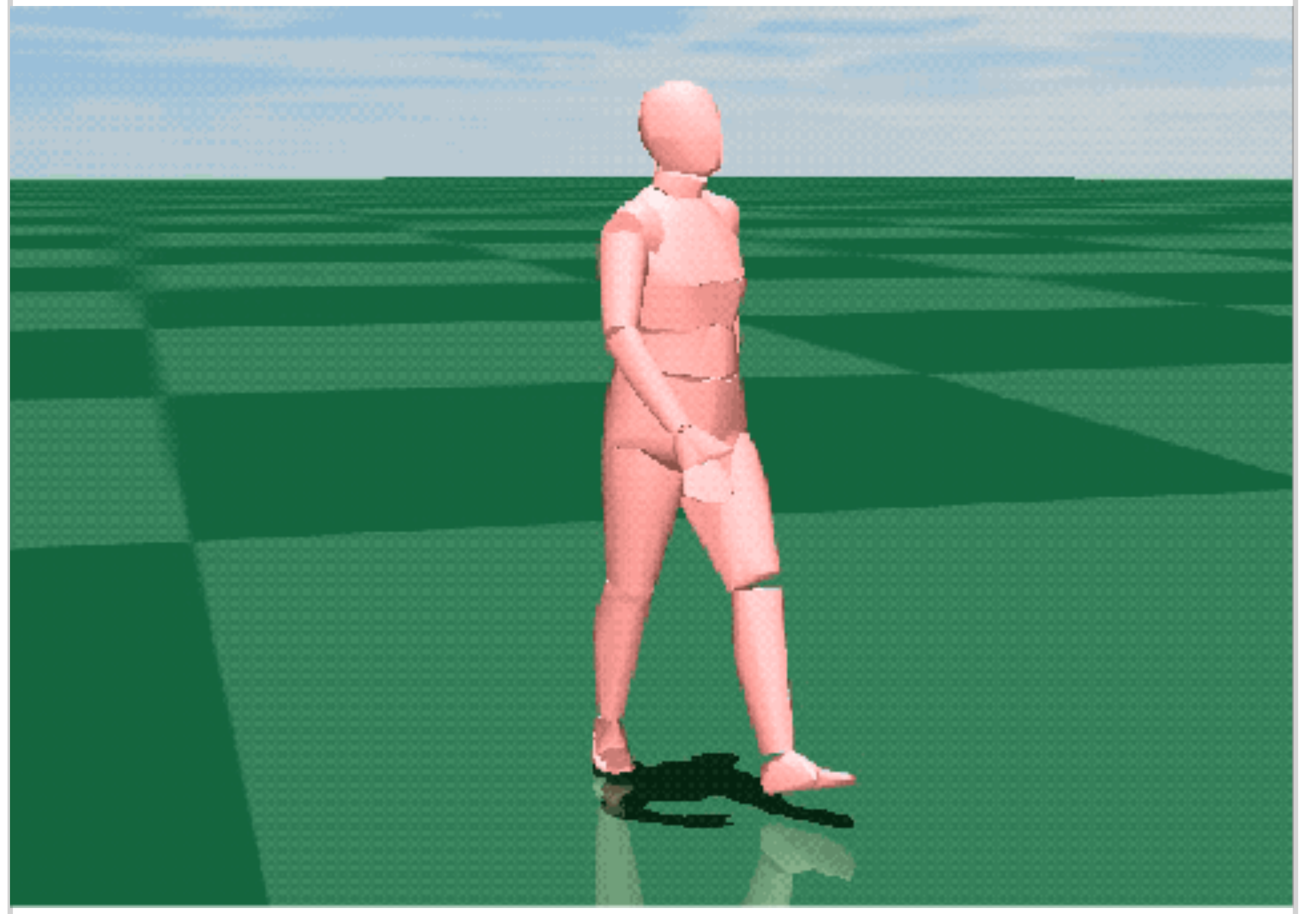
Egocentric video Data capture mount

Real world data

- 1 Subject, egocentric videos
- Object pose from Apple ARKit
- VIO Camera trajectory from Apple ARKit
- 183 takes in total, 6-10 seconds
- Actions:
  - Sitting down/Standing up, avoiding obstacles, pushing a box
- Captured in a living room
- All used for testing

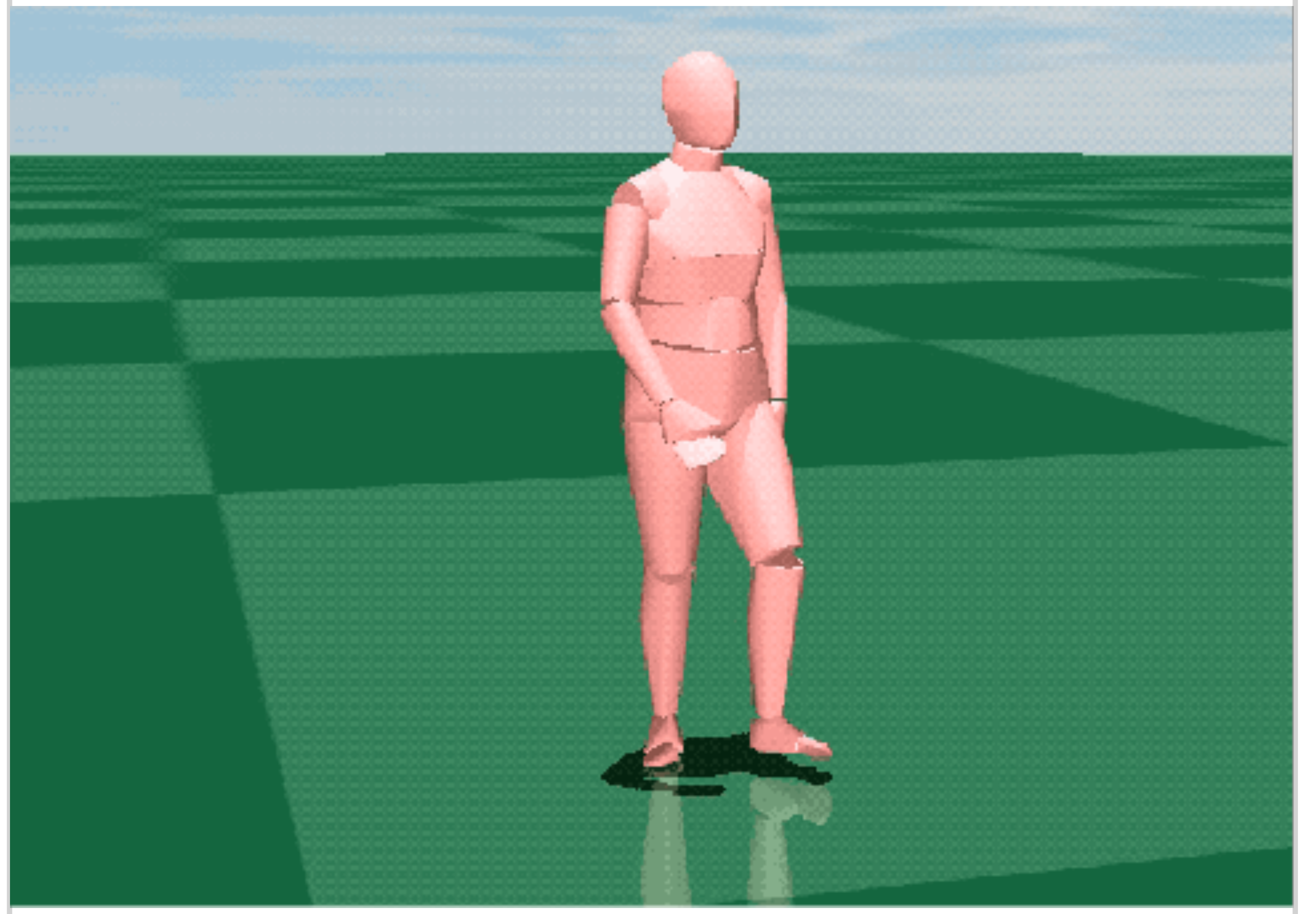
Evaluation: Baselines

PoseReg (kinematics-only)



[Yuan et al. ICCV 2019]

EgoPose (dynamics-only)



[Yuan et al. ICCV 2019]

THE ROBOTICS INSTITUTE

Evaluation: Results

S_{inter} : Human-object interaction success rate

E_{acc} : human joint acceleration error

E_{cam} : camera/head pose error

E_{root} : root pose error

FS : foot sliding

E_{mpjpe} : human joint position error

PT : penetration

MoCap dataset							
Method	Physics	S_{inter} ↑	E_{root} ↓	E_{mpjpe} ↓	E_{acc} ↓	FS ↓	PT ↓
PoseReg	✗	-	0.857	87.680	12.981	8.566	42.153
Kin_poly: supervised learning (ours)	✗	-	**0.176**	**33.149**	**6.257**	5.579	10.076
EgoPose	✓	48.4%	1.957	139.312	9.933	2.566	7.102
Kin_poly: dynamics-regulated (ours)	✓	**96.9%**	0.205	40.443	7.064	**2.474**	**0.686**

Real-world dataset						Per class success rate S_{inter} ↑			
Method	Physics	S_{inter} ↑	E_{cam} ↓	FS ↓	PT ↓	Sit	Push	Avoid	Step
PoseReg	✗	-	1.260	6.181	50.414				
Kin_poly: supervised learning (ours)	✗	-	0.491	5.051	34.930				
EgoPose	✓	9.3%	1.896	**2.700**	1.922	7.93%	6.81%	4.87%	0.2%
Kin_poly: dynamics-regulated (ours)	✓	**92.3%**	**0.476**	2.742	**1.229**	**98.4%**	**90.9%**	**100%**	**74.2%**

Comparision with the state-of-the-art
on the hold-out MoCap dataset

THE ROBOTICS INSTITUTE

Limitations and Future Work

Limitation & Failure Modes

- Humanoid can still lose balance on challenging poses.
- Kinematic policy is trained on a relatively small dataset, and **requires known action classes as a strong prior.**
- Does not handle head rotation well.

Future directions

- Factoring in hand and figure motion for egocentric human-object interaction.
- Incorporating universal humanoid controller to third person pose estimation.
- Full body egocentric pose estimation on large scenes.

Applications and Conclusions

- **Egocentric pose estimation:**
  - Inferring wearer's motion and interaction with the scene
  - Telepresence
- Universal humanoid controller and kinematic policy:
  - Physically-valid human motion estimation
  - Robotics/Manipulation

https://zhengyiluo.github.io/projects/kin_poly/