Expressive Whole-Body Control for Humanoid Robots

Xuxin Cheng* , Yandong Ji* , Junming Chen, Ruihan Yang, Ge Yang, Xiaolong Wang UC San Diego (2024)

Contribution

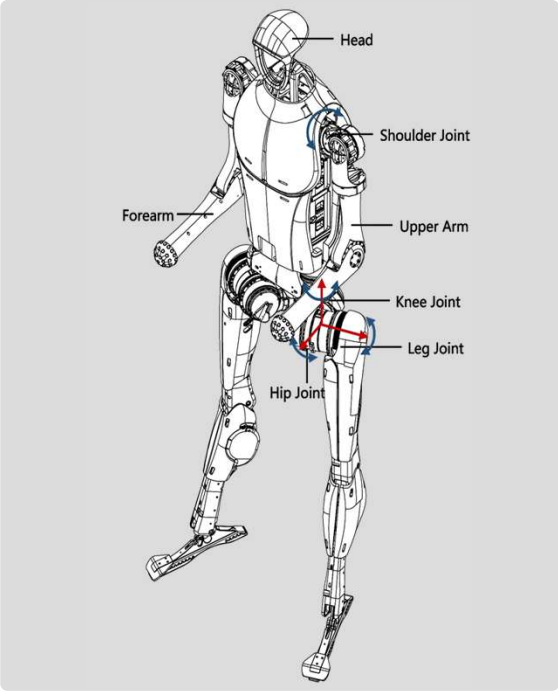
Traditional control approaches focus on task objective

- Tends to produce stiff and mechanical motion

Expressive Whole-Body Control (ExBody)

- Match humans in expressivity and richness
- Wide variety of motions

Executive Summary



- Humanoid - Unitree H1 robot
- Joint angle control
- Proprioceptive
- Reinforcement Learning


OF LA SALET

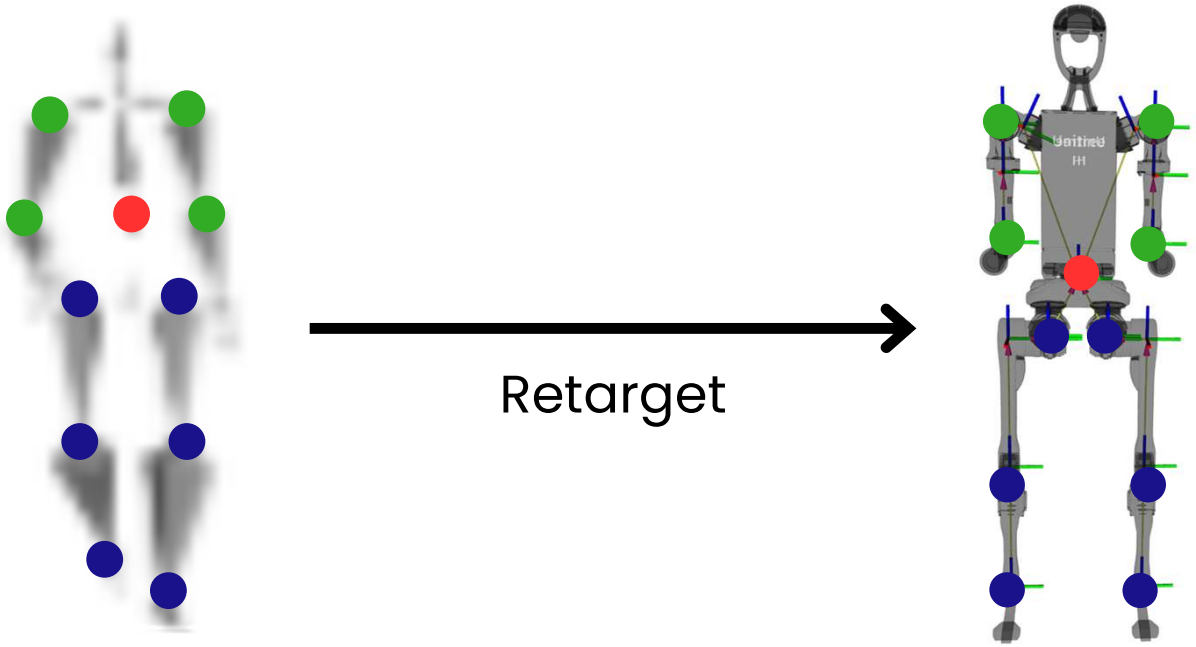
Approach

Demonstration Motion:

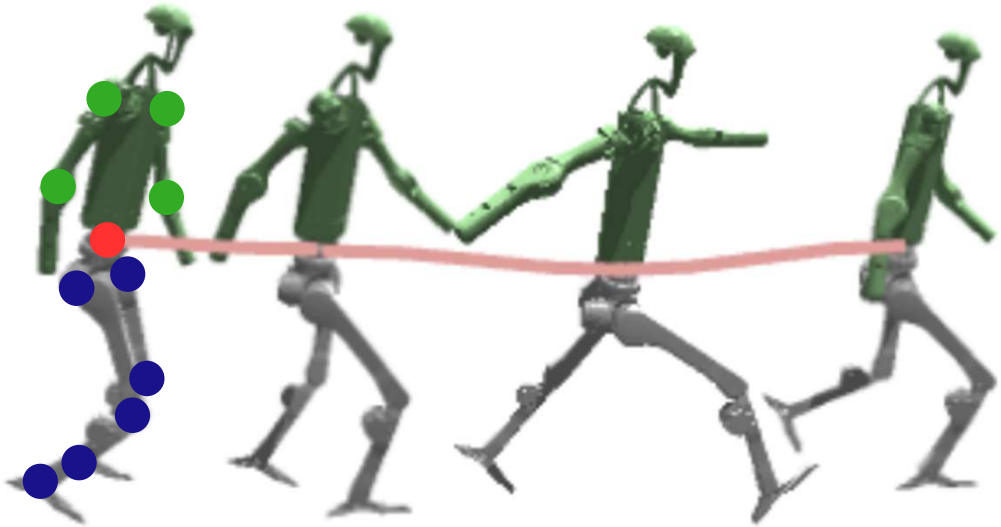
780 reference motion clips

CMU Motion Capture (MoCap) Dataset

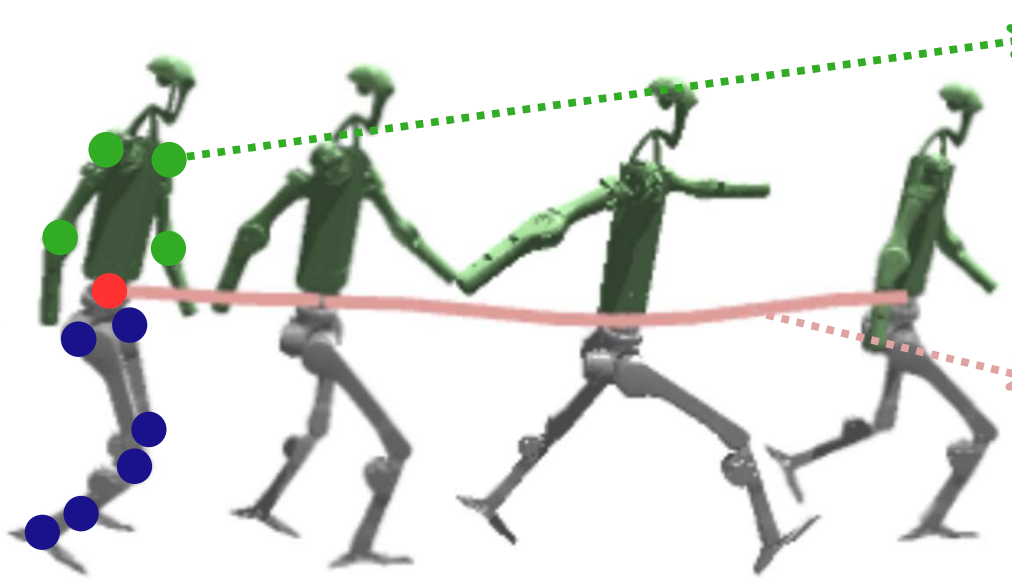
Approach



Approach



Approach



Goal, G

G_e **Expression goal**

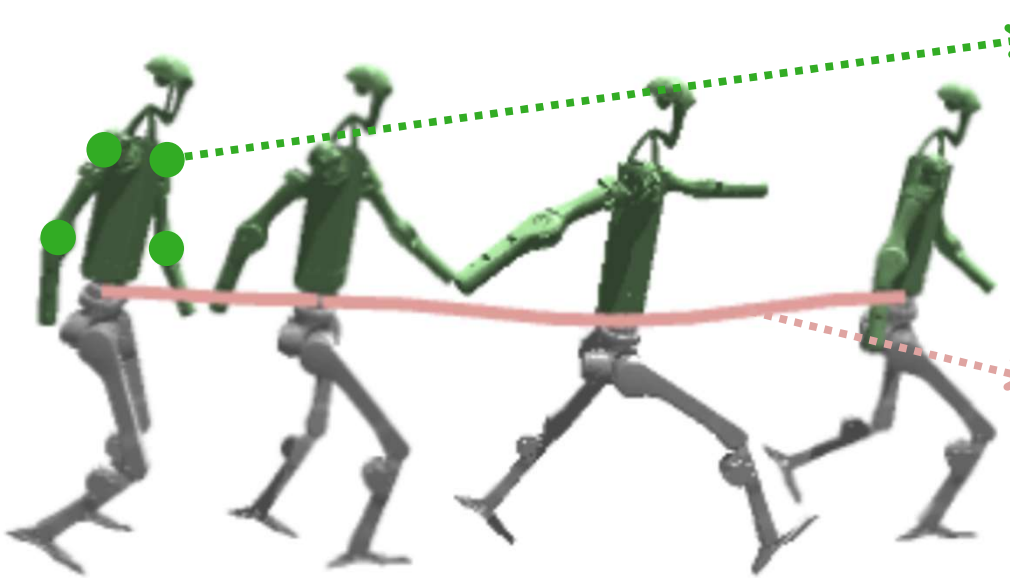
- Joint angles
- 3D locations of key points of the body

G_m **Root movement goal**

- Root pose
- Root velocity

Approach

Goal, G



Expression goal

- Joint angles
- 3D locations of key points of the body

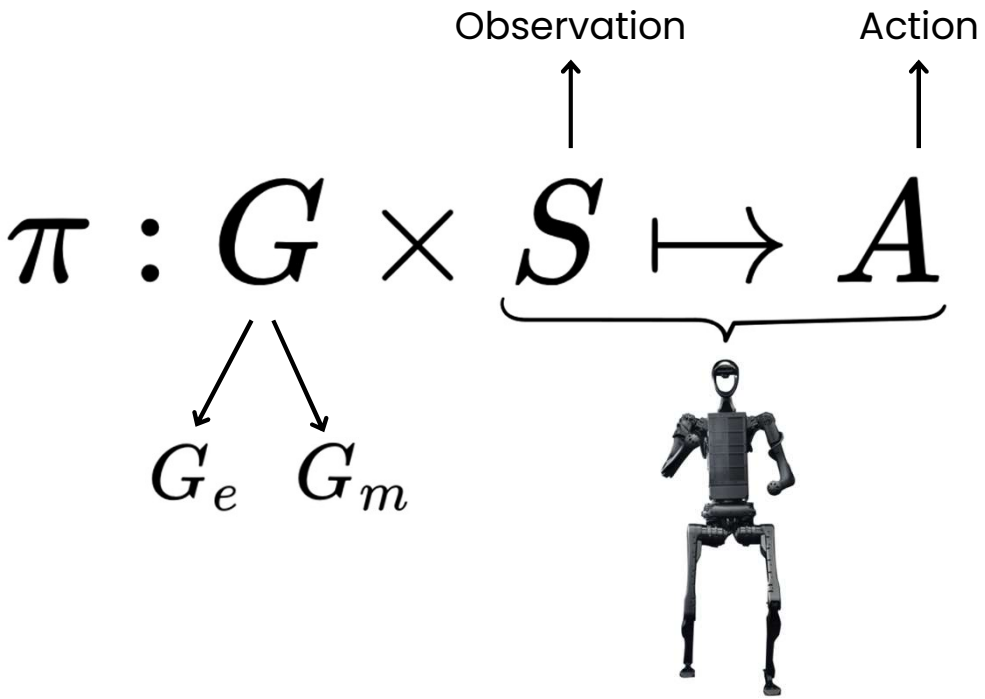
Root movement goal

- Root pose
- Root velocity

Used to satisfy G_m

Approach

ExBody Policy:



Approach

$$\pi : G \times S \mapsto A$$

Massively Parallel Simulation in Isaac Gym Environment

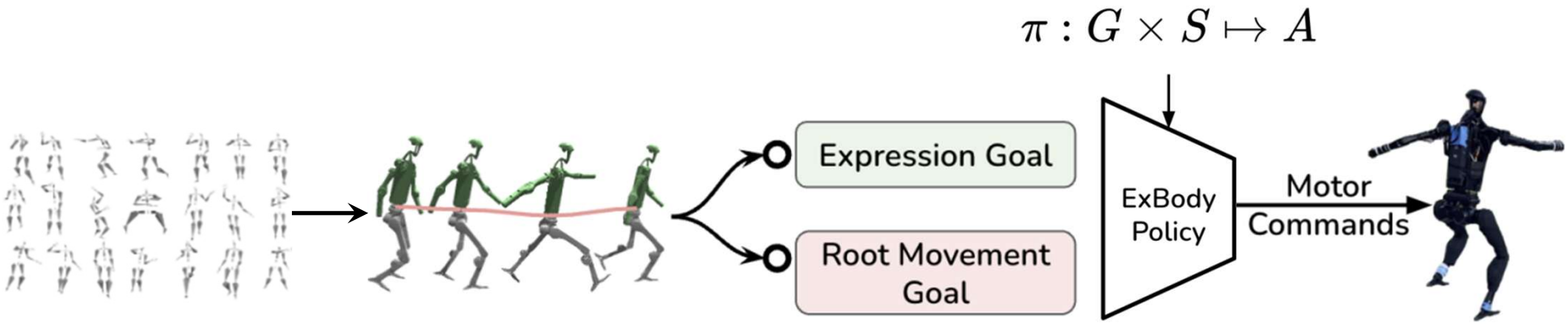
Rewards:

Term	Expression	Weight
Expression Goal G^e		
DoF Position	$\exp(-0.7\lvert\mathbf{q}_{\text{ref}} - \mathbf{q}\rvert)$	3.0
Keypoint Position	$\exp(-\lvert\mathbf{p}_{\text{ref}} - \mathbf{p}\rvert$	2.0
Root Movement Goal G^m		
Linear Velocity	$\exp(-4.0\lvert\mathbf{v}_{\text{ref}} - \mathbf{v}\rvert)$	6.0
Roll & Pitch	$\exp(-\lvert\mathbf{\Omega}^{\phi\theta}_{\text{ref}} - \mathbf{\Omega}^{\phi\theta}\rvert)$	1.0
Yaw	$\exp(-\lvert\Delta y\rvert)$	1.0

+

Regularization rewards

Approach - Framework



Results

Answering 4 questions:

How well does ExBody perform on tracking g_m?

How well does ExBody perform on tracking g_e?

Why should we learn from large data?

Why does ExBody not do full DoF tracking?

Results

How well does ExBody perform on tracking g_m?

Execution of nearly identical movements

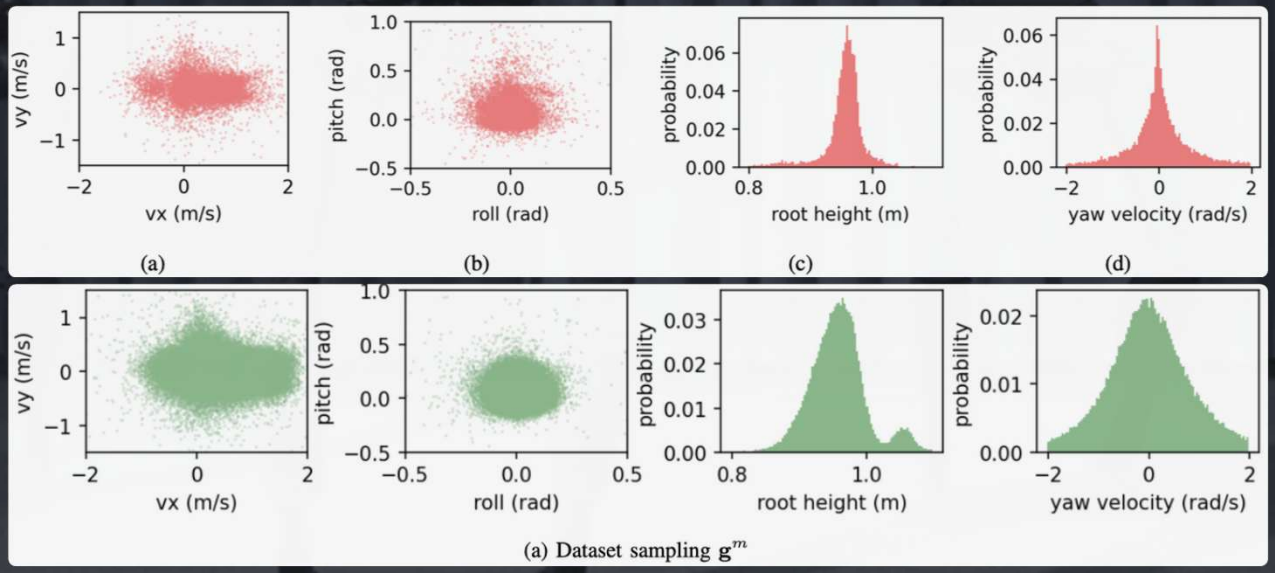


Fig 3: CMU dataset

Fig 5a: Trained ExBody

Results

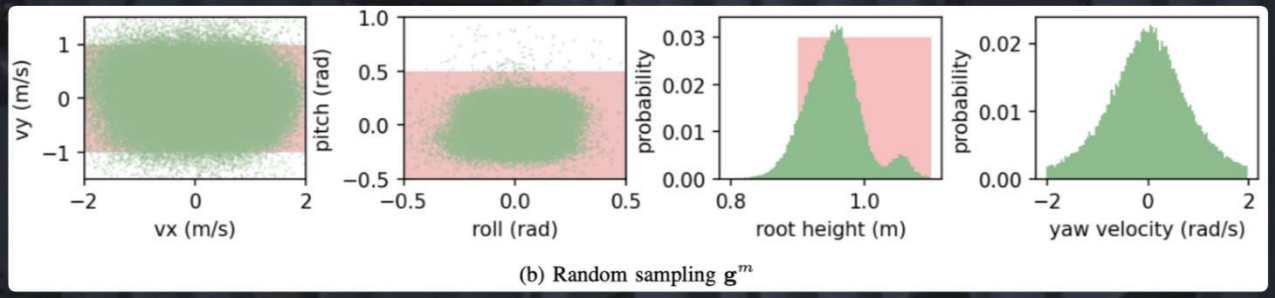
How well does ExBody perform on tracking g_m?

Execution of nearly identical movements

Policy coverage

sampled commands

Fig 5b: Policy Coverage



(b) Random sampling $\mathbf{g}^m$

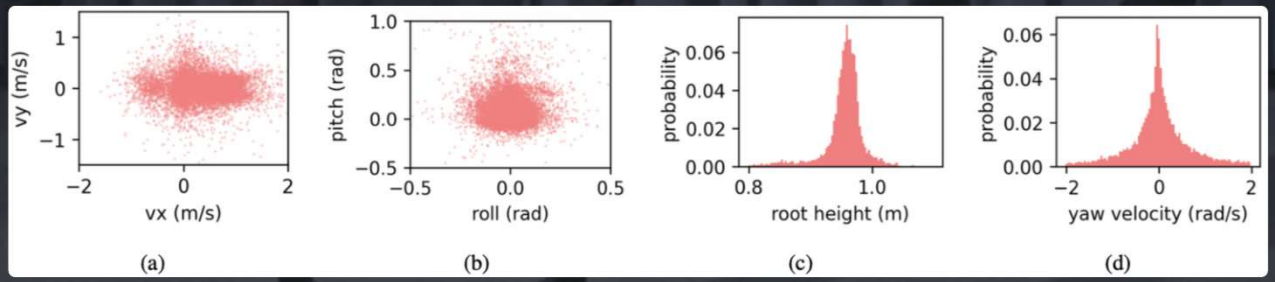
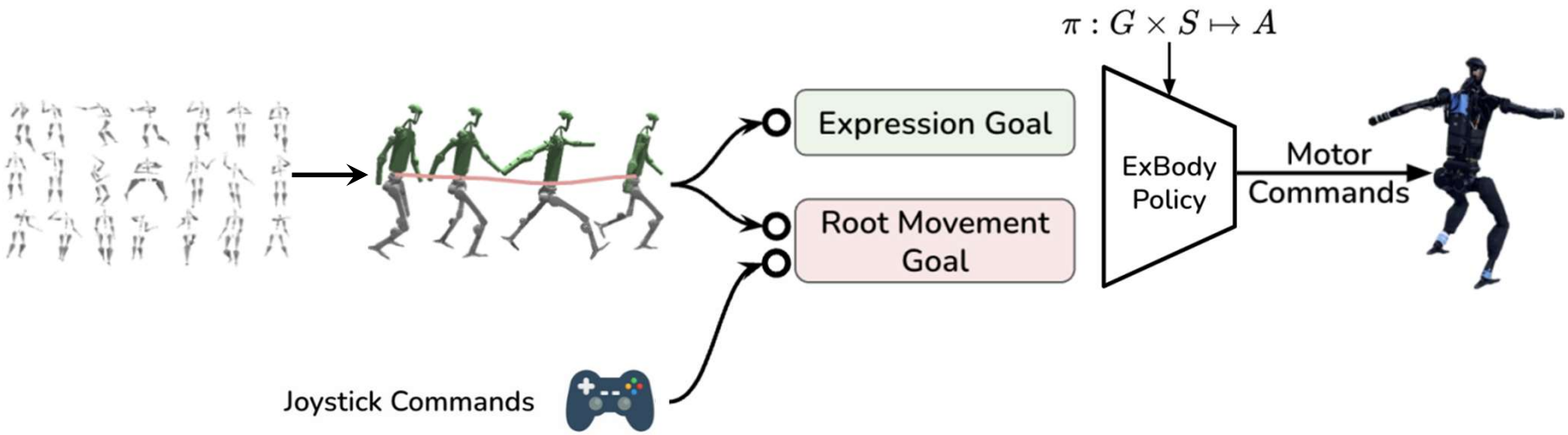


Fig 3: Training Data Set

Results

How well does ExBody perform on tracking g_m?

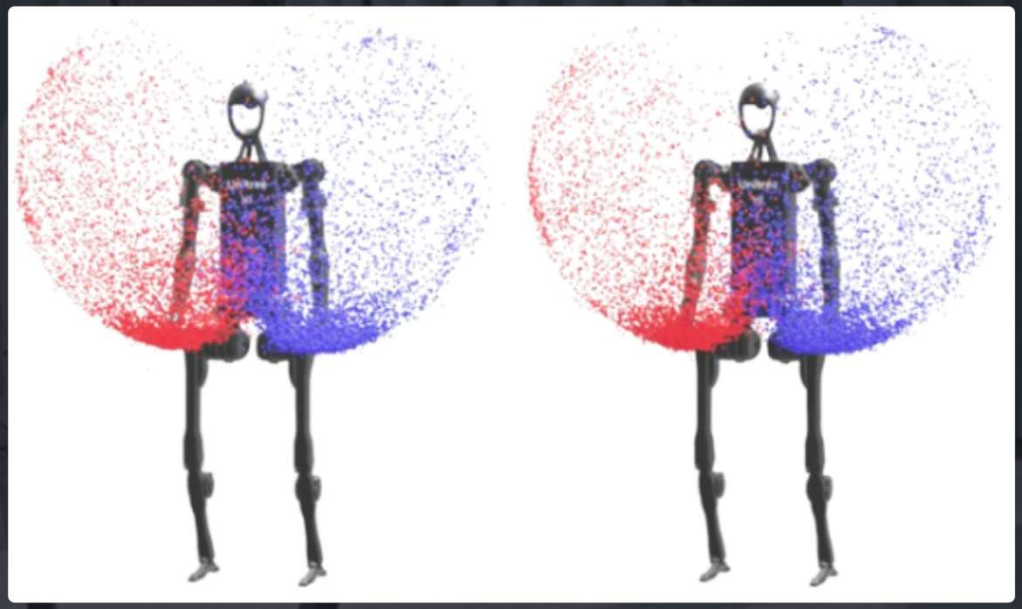


Results

How well does ExBody perform on tracking g_e?

The samples in Fig. 6 show a nearly identical distribution

Figure 6: Distribution of hand positions



Retargeted motion dataset

Learned ExBody policy rollouts

Results

Why should we learn from large data?

- Random State Initialization (RSI) encourages safe exploration
- Without RSI, the policy tends to self terminate

Baselines	Motion Sample				Random Sample			
	MEL↑	MELV↑	MERP↑	MEK↑	MEL	MELV	MERP	MEK
ExBody (Ours)	16.87	**318.67**	754.92	**659.78**	13.51	**132.14**	523.79	483.67
ExBody + AMP	**17.28**	205.60	**765.85**	635.51	15.59	95.11	583.82	**544.59**
ExBody + AMP NoReg	16.16	87.83	714.74	561.56	15.40	36.76	584.23	515.53
No RSI	0.23	0.63	10.09	7.25	0.22	0.10	7.41	7.15
Random Sample	16.50	181.85	704.73	326.66	16.37	38.51	**586.83**	324.10
Full Body Tracking	13.28	246.11	584.40	397.25	10.76	76.46	407.88	284.69

Table 4: Comparisons with baselines

Results

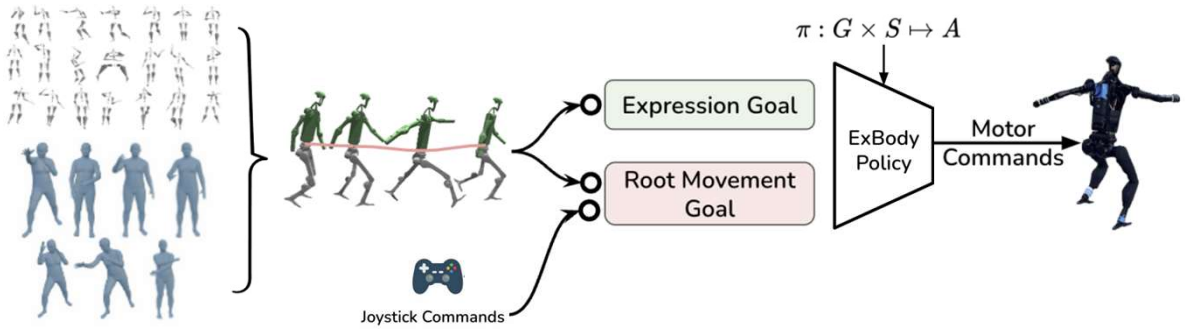
Why does not ExBody do full DoF tracking?

- Fewer Degrees of Freedom and limited torques
- Worse performance with all metrics for full body tracking

Baselines	Motion Sample				Random Sample			
	MEL↑	MELV↑	MERP↑	MEK↑	MEL	MELV	MERP	MEK
ExBody (Ours)	16.87	**318.67**	754.92	**659.78**	13.51	**132.14**	523.79	483.67
ExBody + AMP	**17.28**	205.60	**765.85**	635.51	15.59	95.11	583.82	**544.59**
ExBody + AMP NoReg	16.16	87.83	714.74	561.56	15.40	36.76	584.23	515.53
No RSI	0.23	0.63	10.09	7.25	0.22	0.10	7.41	7.15
Random Sample	16.50	181.85	704.73	326.66	16.37	38.51	**586.83**	324.10
Full Body Tracking	13.28	246.11	584.40	397.25	10.76	76.46	407.88	284.69

Table 4: Comparisons with baselines

Results - Real World



Test on various expressive motions in sim and real world

	Category	Clips	Length (s)
Training	Walk	546	9076.6
	Dance	78	1552.3
	Basketball	36	766.1
	Punch	20	800.0
	Others	100	1188.0
	Total	780	13383.0
Real-World Test	Punch	1	18.9
	Wave Hello	1	5.0
	Mummy Walk	1	22.5
	Zombie Walk	1	13.0
	Walk, Exaggerated Stride	1	2.5
	High Five	1	3.3
	Basketball Signals	1	32.6
	Adjust Hair	1	9.6
	Drinking from Bottle	1	15.2
	Direct Traffic	1	39.3
	Hand Signal	1	32.2
	Russian Dance	1	8.2
	Total	11	202.3
Additional Realworld Test (Diffusion [54])	Boxing	1	4.0
	Hug	1	4.0
	Shake Hands	1	4.0

Variety of sources:

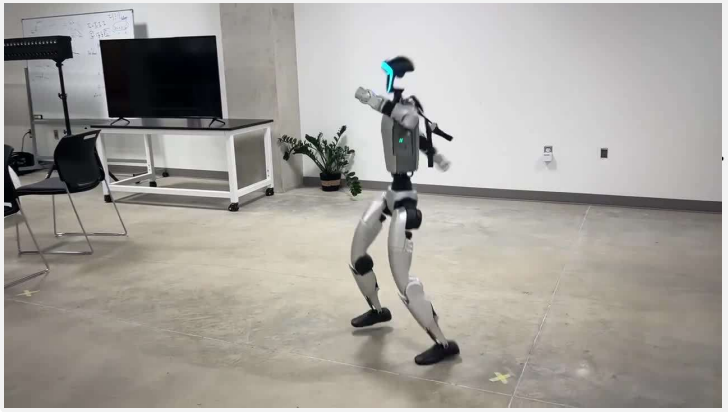
- static motion datasets
- diffusion models
- video-to-skeleton models

Citations

Direct follow-ups

ExBody2

- Larger, filtered dataset
- Teacher-student training setup
- Decoupled key point tracking and velocity control



Mobile-TeleVision

- Neural video-to-3D reconstruction and retargeting

Other Citings

156 – Google Scholar

147 – Research Gate

161 – Semantic Scholar

Analysis

Pros

- Framework to imitate human expression
- Reasonably accurate, smooth motion
- Disturbance robustness
- Strong Sim2Real transfer
- Generalizable framework

Cons

- Only upper-body expressivity
- Entangled upper-body and locomotion
- Loss of information in motion retargeting due to fewer DoFs

Thank you for listening!