Deep Whole-Body Control

Learning a Unified Policy for Manipulation and Locomotion

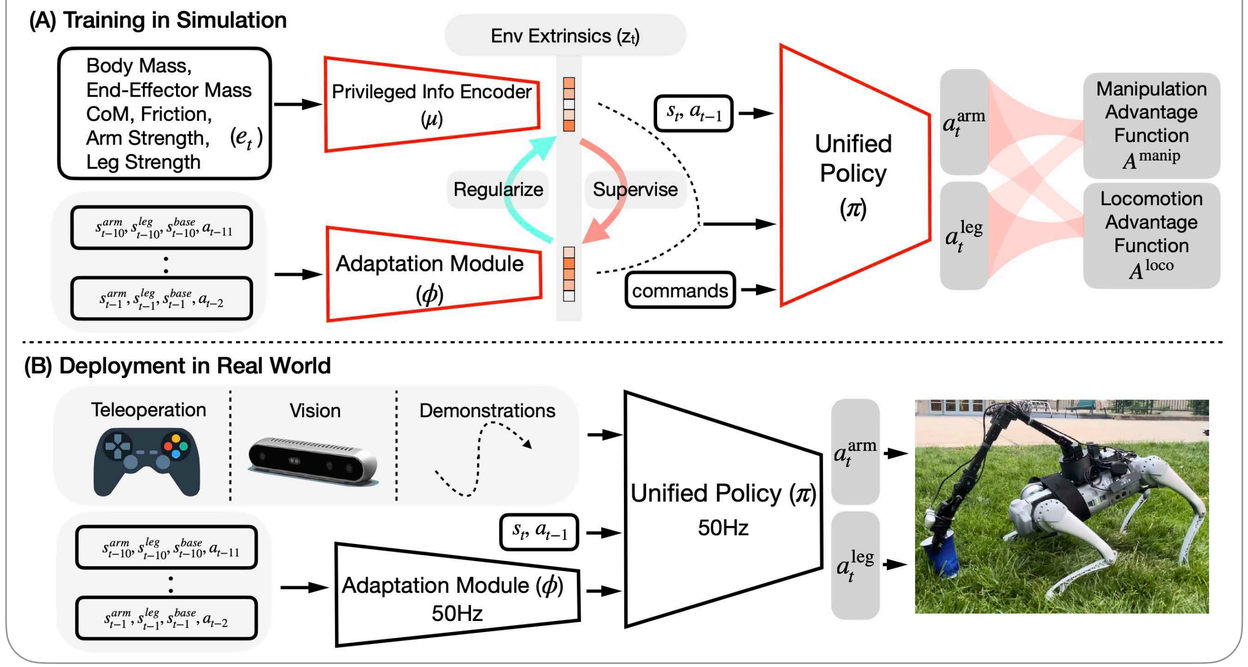
Zipeng Fu* Xuxin Cheng* Deepak Pathak

Carnegie Mellon University — The Robotics Institute

Videos & Code

https://maniploco.github.io

DeepWBC Pipeline



Motivation: whole-body >> modular, low-cost >> expensive hardware

TL;DR: learning an **end-to-end** unified policy for whole-body control of a custom-built **low-cost** quadruped **mobile manipulator**

Regularized Online Adaptation for Sim-to-Real Transfer

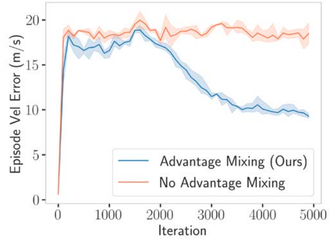
(λ, β follow simple linear curricula)

$$L(\theta_\pi, \theta_\mu, \theta_\phi) = \underbrace{-J(\theta_\pi, \theta_\mu)}_{\text{RL Loss}} + \underbrace{\lambda||z^\mu - \text{sg}[z^\phi]||_2}_{\text{Regularization}} + \underbrace{||\text{sg}[z^\mu] - z^\phi||_2}_{\text{Adaptation}}$$

Advantage Mixing for Policy Learning

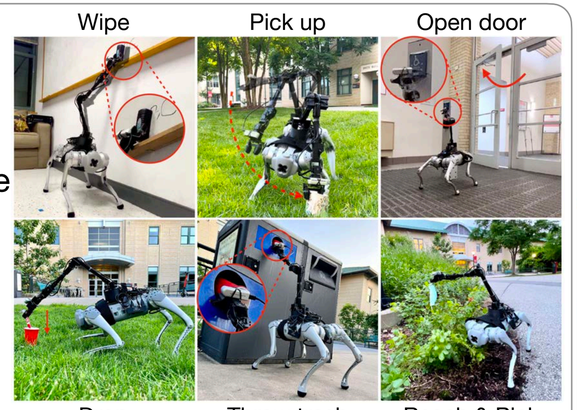
$$J(\theta_\pi) = \frac{1}{|\mathcal{D}|} \sum_{(s_t, a_t) \in \mathcal{D}} \log \pi(a_t^{\text{arm}} \mid s_t) \left(A^{\text{manip}} + \beta A^{\text{loco}}\right) + \log \pi(a_t^{\text{leg}} \mid s_t) \left(\beta A^{\text{manip}} + A^{\text{loco}}\right)$$

	Realizability Gap $\|z^\mu - z^\phi\|_2$ ↓	Survival ↑	Vel Error ↓	EE Error ↓
Domain Randomization	-	95.8 ± 0.2	0.46 ± 0.00	0.40 ± 0.00
RMA [22]	0.31 ± 0.01	95.2 ± 0.2	0.44 ± 0.00	0.26 ± 0.04
Regularized Online Adapt (Ours)	**2e-4** ± 0.00	**97.4** ± 0.1	**0.39** ± 0.01	**0.21** ± 0.00
Expert w/ Reg.	-	97.8 ± 0.2	0.40 ± 0.01	0.21 ± 0.00
Expert w/o Reg.	-	98.3 ± 0.2	0.39 ± 0.00	0.21 ± 0.00



Large Whole-Body Workspace

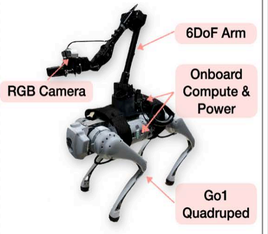
Diverse Task Set



Hardware Setup Comparison

Spot with Arm
- > $100K
- No low-level control APIs

Anymal B + Kinova
- Bulky (~40kg)
- Expensive



Ours (Go1 + WidowX 250s)
- $6.3K hardware cost
- Fully onboard compute + power