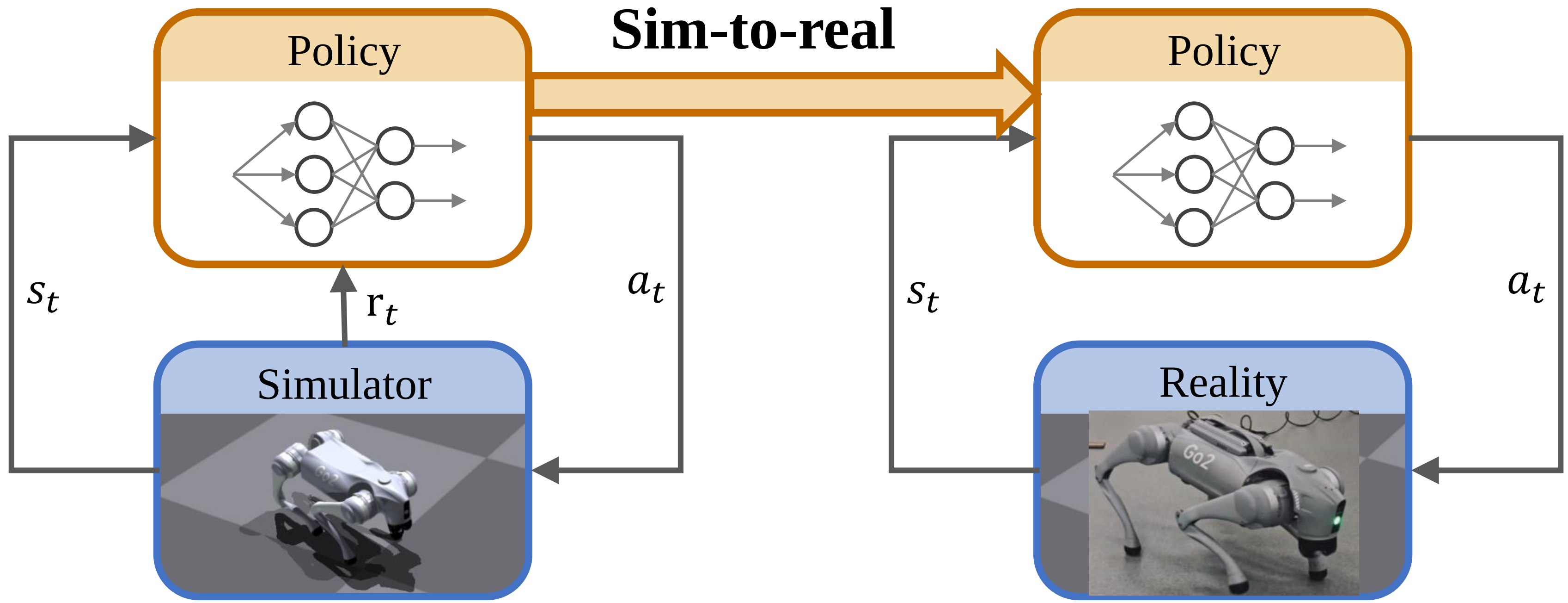
EASI: Evolutionary Adversarial Simulator Identification for Sim-to-Real Transfer

Haoyu Dong, Huiqiao Fu, Wentao Xu, Zhehao Zhou, Chunlin Chen

Nanjing University, China

Motivation

Transferring **simulated-trained policies to real world** in a stable and cost-effective manner has long been a goal for sim-to-real transfers.



Train in simulation

- Fast, safe,
- Low training cost.

Deploy in reality

- Reality gap lead to performance degradation.

Motivation

DR (Domain Randomization)

Requires specific domain prior knowledge and hand-engineering to determine the simulator parameter distribution.

Motivation

DR (Domain Randomization)

Requires specific domain prior knowledge and hand-engineering to determine the simulator parameter distribution.

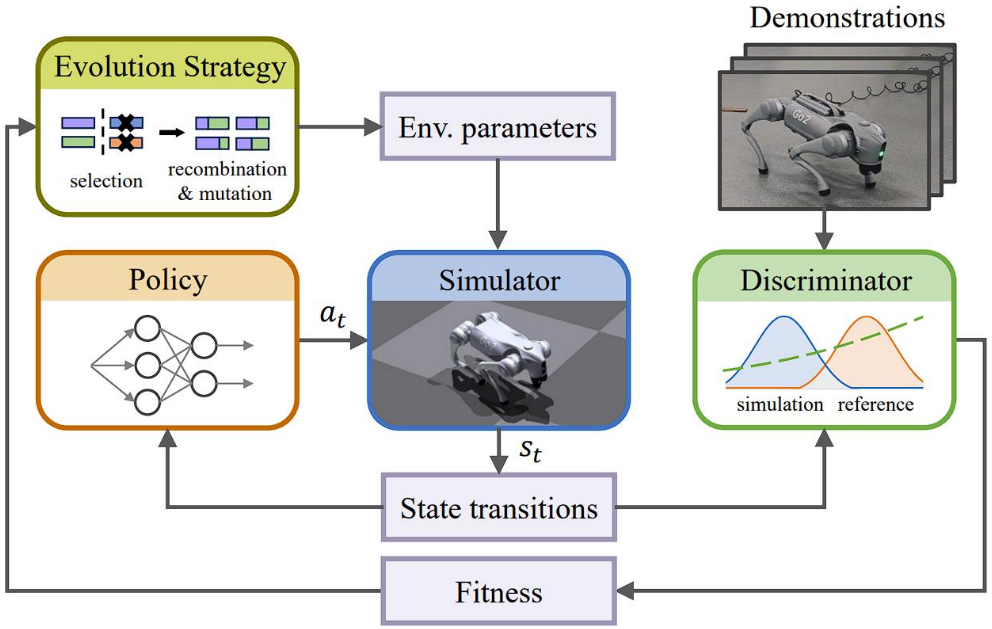
We aim to find a method for adjusting the simulator's parameters that helps us obtain **high-fidelity, low-cost** simulators, which can provide a **real-world-like environment** for RL training.

Method

Evolutionary Adversarial Simulator Identification (EASI), aiming to find physical parameter distributions that make the **state transitions** between simulation and reality **as similar as possible**.

ES acts as a generator **in adversarial competition** with a neural network discriminator, distinguishing between simulation and reality state transitions.

- Imitate State Transitions
- Evolution Strategy as the generator



Schematic overview of EASI.

Method

- Imitate State Transitions

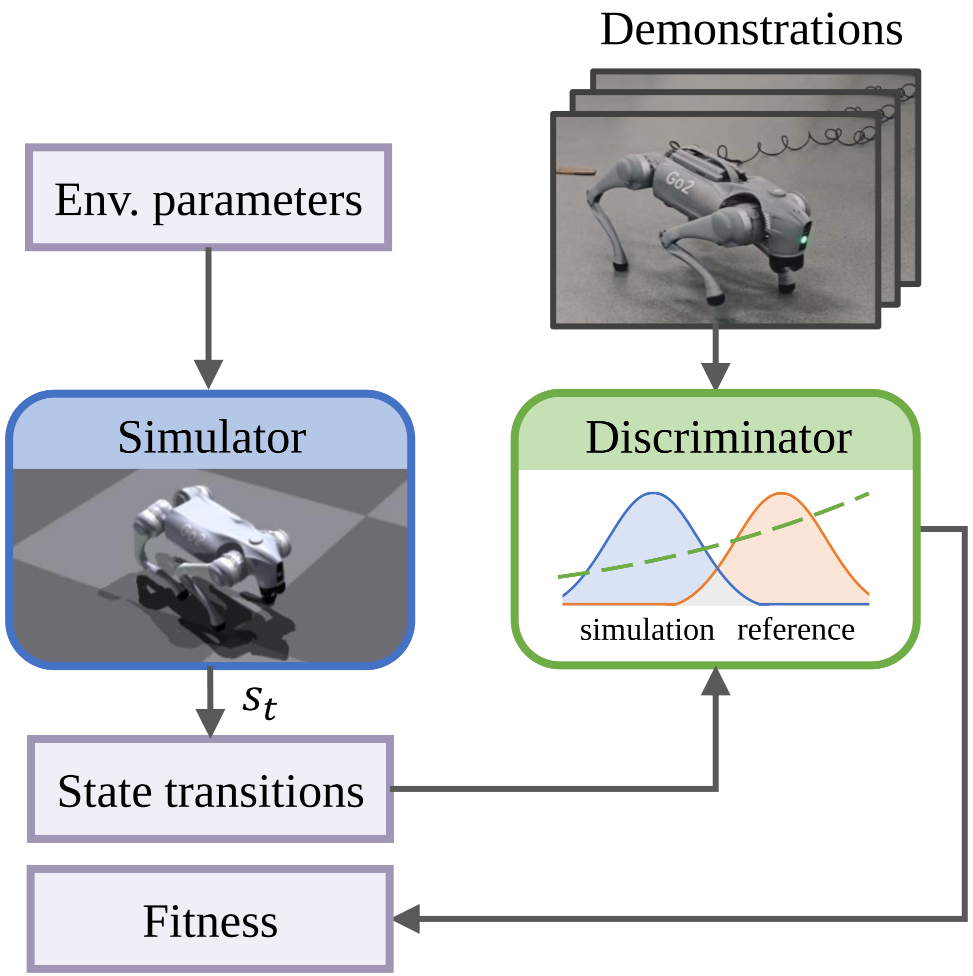
EASI uses discriminator to distinguish **whether state transitions come from simulation or reality**

$$\max_{D} \mathbb{E}_{d^{\mathcal{M}}(\mathbf{s},\mathbf{a},\mathbf{s}')}[log(D(\mathbf{s},\mathbf{a},\mathbf{s}'))] + \mathbb{E}_{d^{\mathcal{B}}(\mathbf{s},\mathbf{a},\mathbf{s}')}[log(1 - D(\mathbf{s},\mathbf{a},\mathbf{s}'))]$$

WGAN-style discriminator is utilized to mitigate the issue of **gradient vanishing**

$$\max_{D} \mathbb{E}_{d^{\mathcal{M}}(\mathbf{s},\mathbf{a},\mathbf{s}')}[D(\mathbf{s},\mathbf{a},\mathbf{s}')] - \mathbb{E}_{d^{\mathcal{B}}(\mathbf{s},\mathbf{a},\mathbf{s}')}[D(\mathbf{s},\mathbf{a},\mathbf{s}')]$$

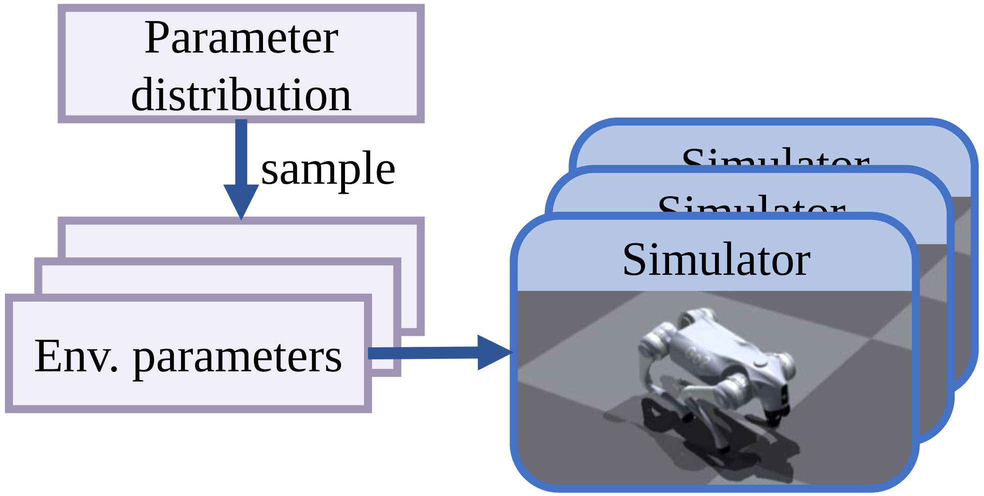
Which is an efficient approximation to the Earth-Mover distance between state transition distribution in simulation and reality.



Method

- Evolution Strategy as the generator

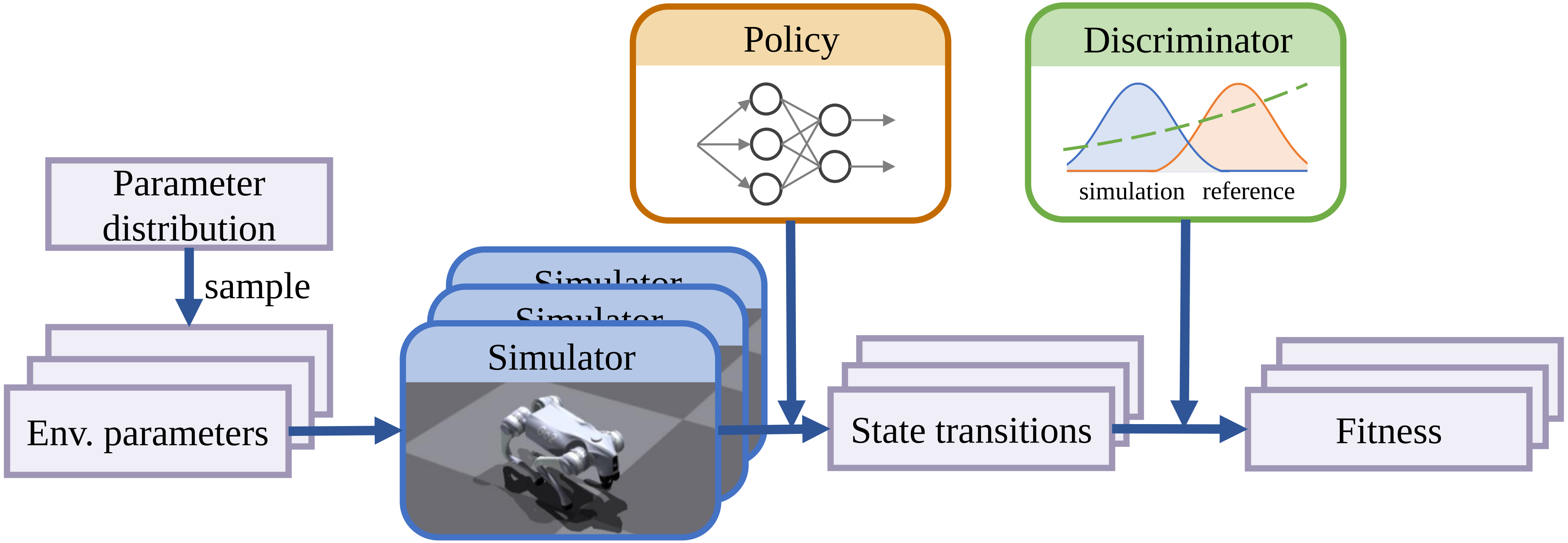
$$\Xi = \arg\min_{\xi_{sim} \in \Xi} \|\mathcal{P}_r(\xi_{real}), \mathcal{P}_s(\xi_{sim})\|$$



Method

- Evolution Strategy as the generator

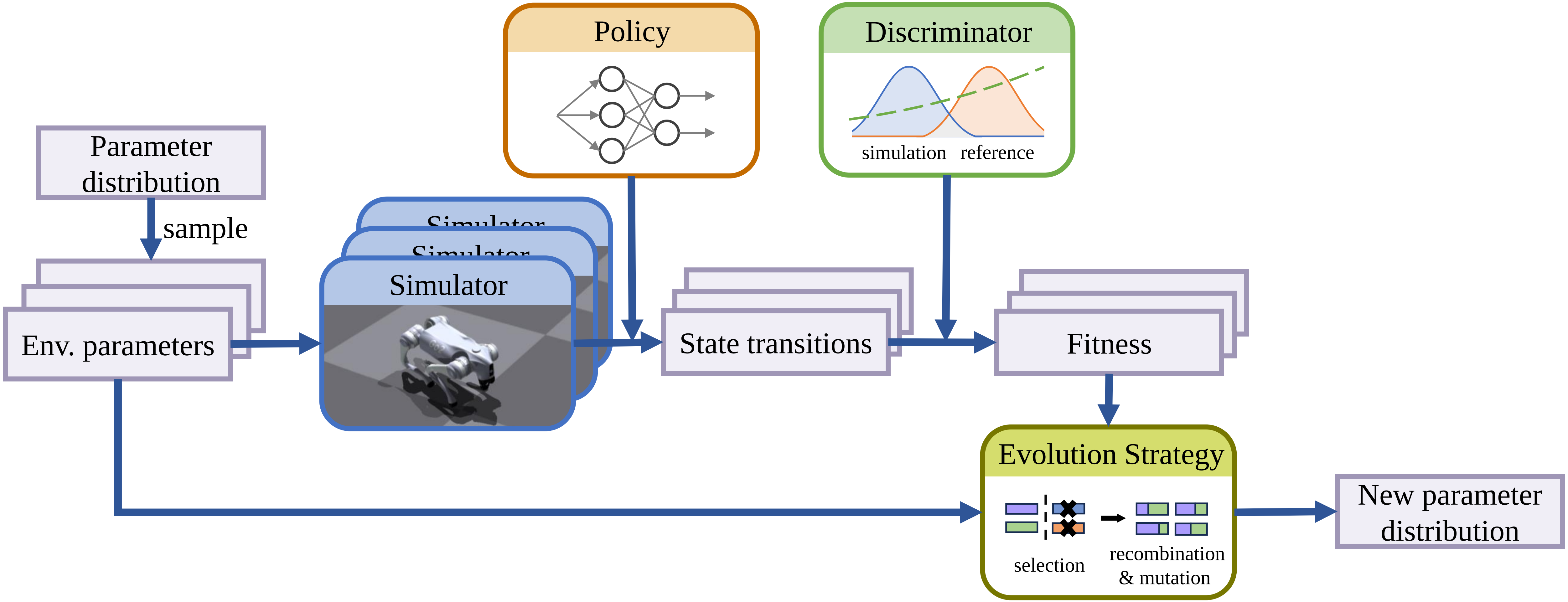
$$\Xi = \arg\min_{\xi_{sim} \in \Xi} \|\mathcal{P}_r(\xi_{real}), \mathcal{P}_s(\xi_{sim})\|$$



Method

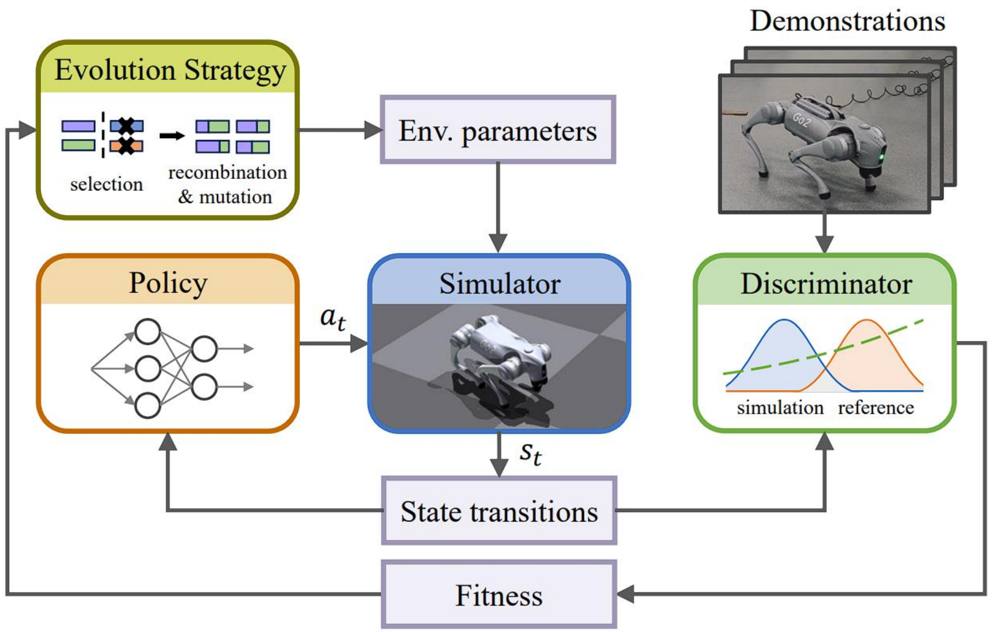
- Evolution Strategy as the generator

$$\Xi = \arg\min_{\xi_{sim} \in \Xi} \|\mathcal{P}_r(\xi_{real}), \mathcal{P}_s(\xi_{sim})\|$$



Method

- Evolution Strategy as the generator



$$\Xi^* = \arg\min_{\Xi} \max_{D} \mathbb{E}_{d^{\mathcal{M}}(\mathbf{s},\mathbf{a},\mathbf{s}')}[D(\mathbf{s},\mathbf{a},\mathbf{s}')] - \mathbb{E}_{d^{\mathcal{B}}(\mathbf{s},\mathbf{a},\mathbf{s}')}[D(\mathbf{s},\mathbf{a},\mathbf{s}')]$$

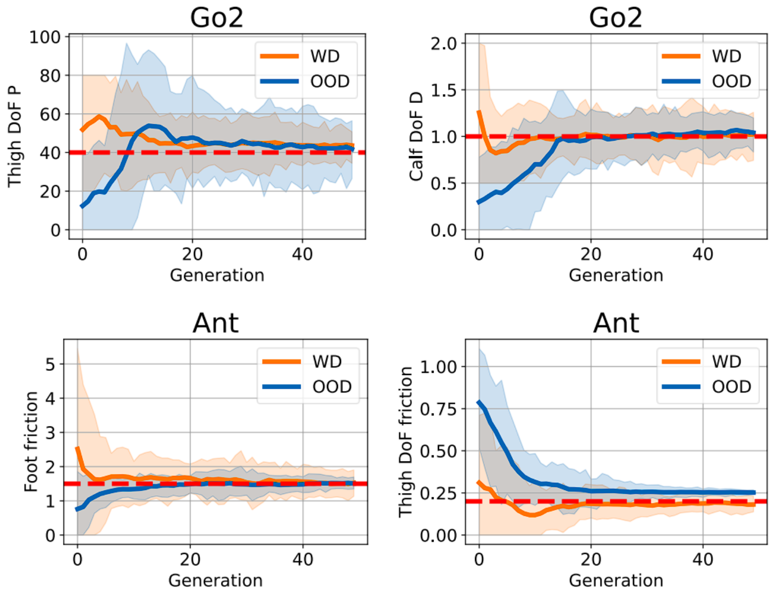
Experiments

We test EASI in 4 sim-to-sim tasks and 2 sim-to-real tasks.

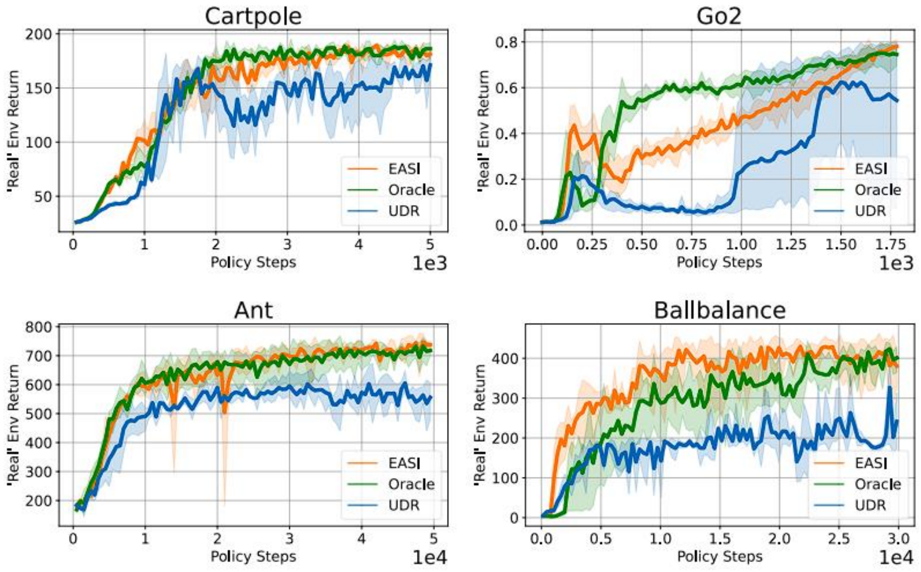
Experiment tasks in simulation.

Experiment tasks in reality.

Experiments



Evolution process of parameters



Policy's performance in the pseudo-real environment throughout the training process.

- ## Parameter Evolution

As the parameter evolution progresses, the parameter distribution quickly adjusts to the vicinity of the target parameters.

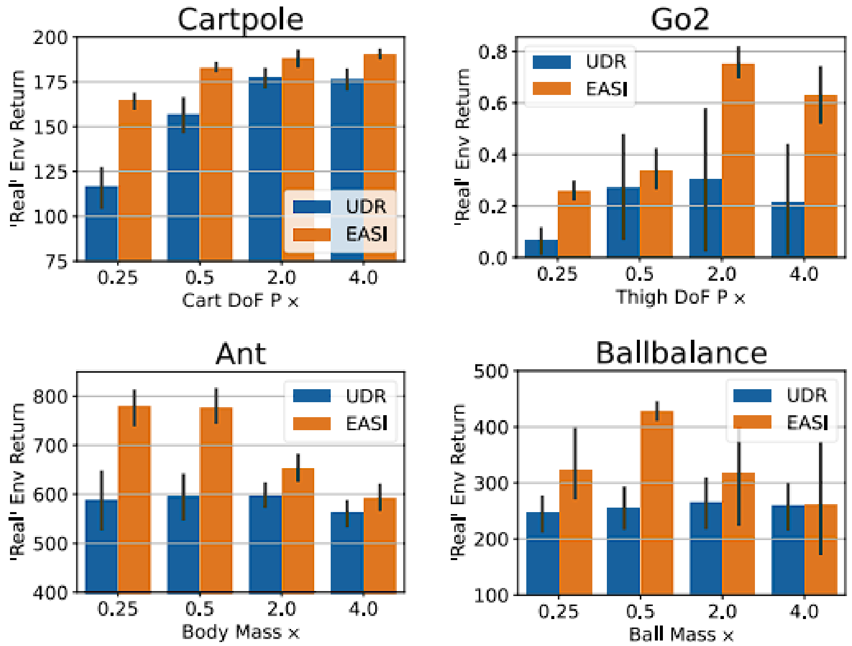
- ## Training with EASI

Training with EASI-optimized parameters, training process become faster and get better final performance.

Experiments

- ## Different target environments

Simulators optimized by EASI are more similar to the target environment, thus policies trained in these simulations are more adaptable to target environments and achieve higher performance in the target domain.



Policy performance in target environments with different parameters.

- ## Data budget

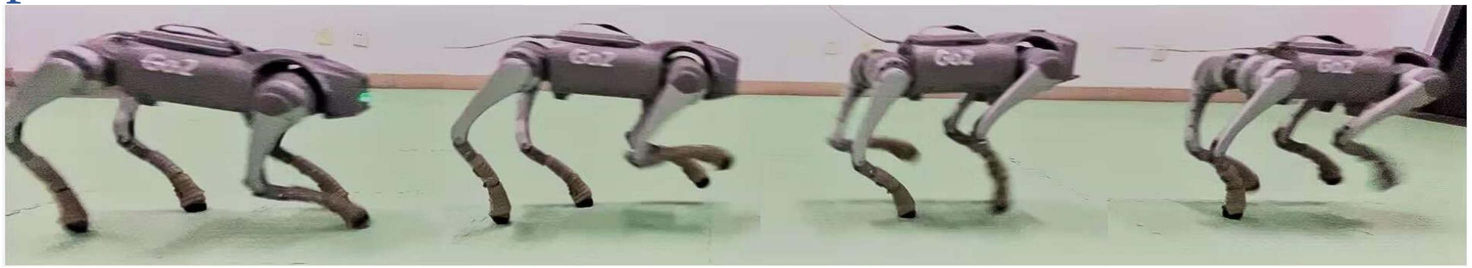
Only a small amount of real-world data is needed for EASI to identify parameters.

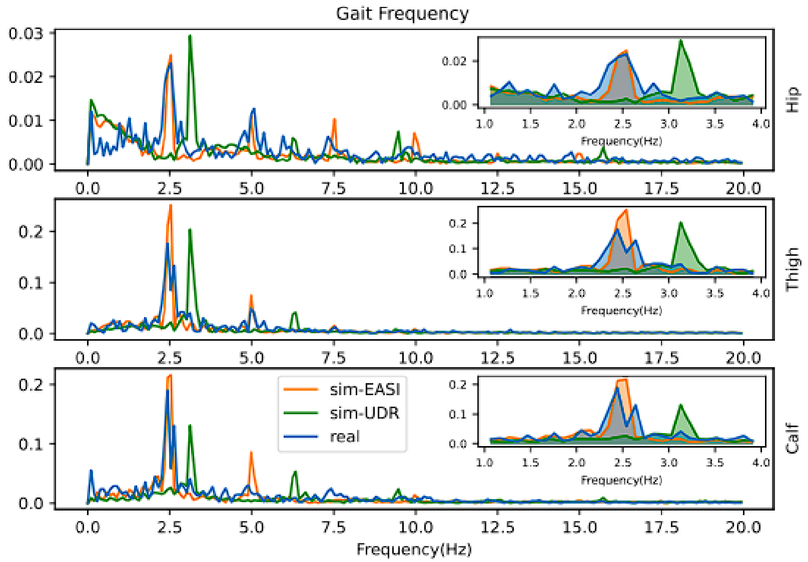
Trajectories	UDR	1	50	100	200
Cartpole	161.3±15.3	185.1±3.1	182.7±2.7	181.1±4.4	182.9±4.3
Ant	557.4±55.6	724.2±7.5	747.0±25.0	703.1±63.1	724.0±31.5
Ballbalance	226.9±61.7	360.1±53.8	421.3±26.1	428.5±16.9	393.1± 60.5
Go2	0.62±0.25	0.44±0.28	0.80±0.06	0.72±0.05	0.78±0.02

Performance of policies in target environments, policies trained with EASI-optimized parameters using varying numbers of reference trajectories.

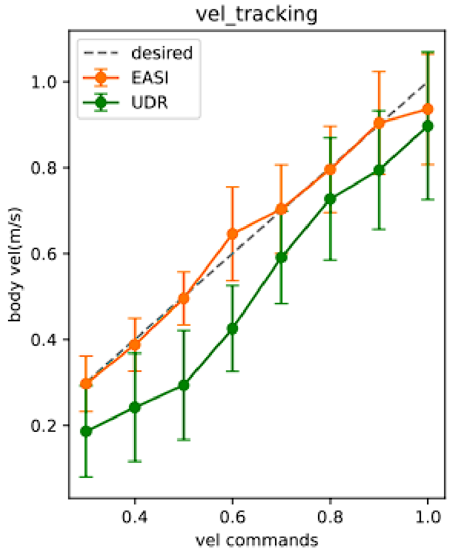
Experiments

- Real Go2 Experiment





(a) Movement spectrum of the joint.



(b) Velocity tracking performance.

(a) More Realistic Simulator,

The motion spectrum of the Go2's joints in the simulation becomes closer to that in the real environment.

(b) Improved Performance

Speed tracking performance improved.

Experiments

The experimental results demonstrate that:

- EASI enhances the simulator's similarity to real-world.
- EASI improves performance in sim-to-real transfer tasks.
- EASI requires only a minimal amount of real-world data.

Thank You

Contact: haoyudong@smail.nju.edu.cn

Our page at: https://blackvegetab1e.github.io/evolutionary_adversarial/

南京大學 NANJING UNIVERSITY 1902

Robotics & Reinforcement Learning Control