NEURAL INFORMATION PROCESSING SYSTEMS

DEAL: Diffusion Evolution Adversarial Learning for Sim-to-Real Transfer

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

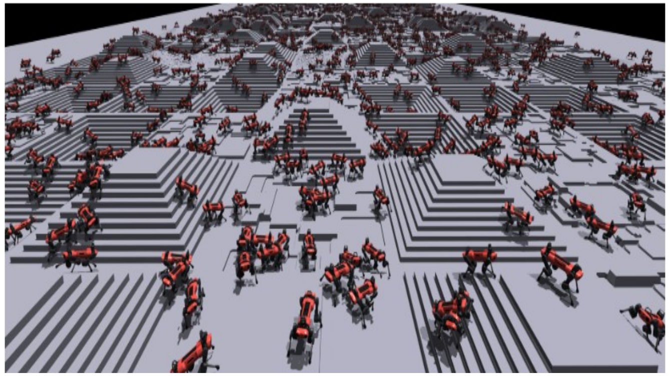
Nanjing University, China

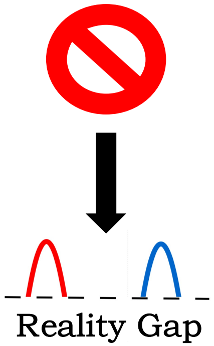
Robotics & Reinforcement Learning Control

Motivation

Why Sim-to-Real is Hard?

- Training in simulation is efficient/safe, but policies often degrade in reality due to the reality gap.

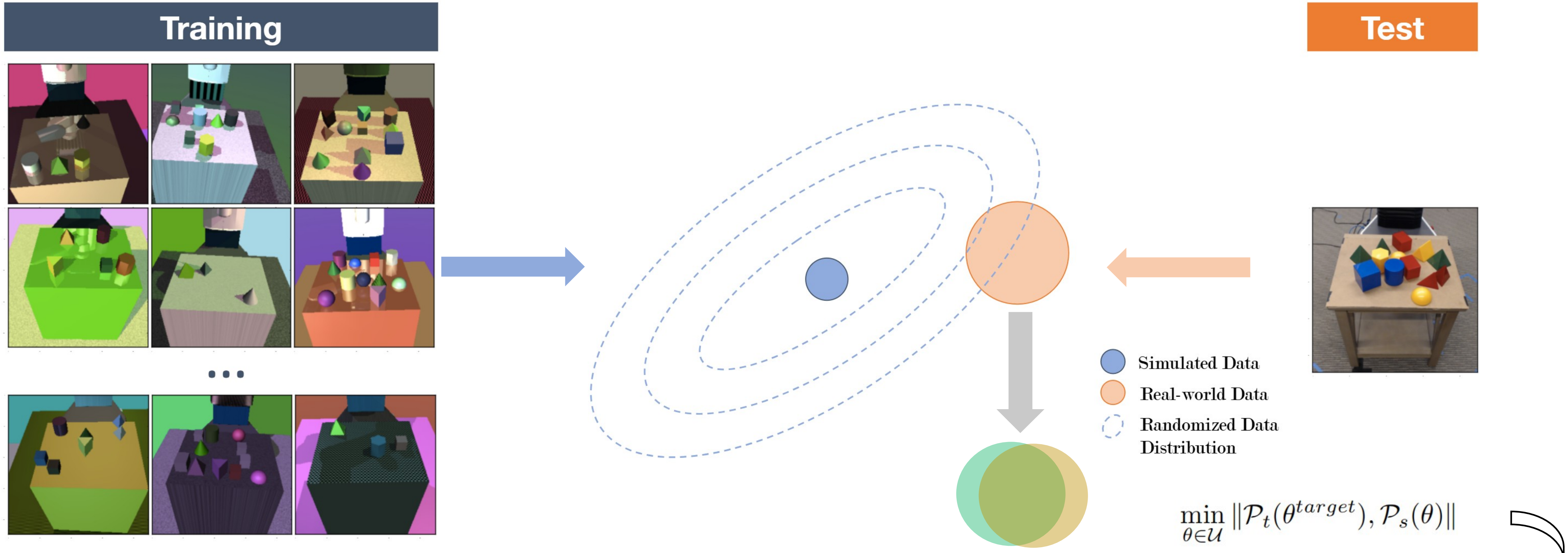




- DR improves robustness but needs expert priors and can be conservative/unstable.

Motivation

- Prior Sys-ID struggles with high-dimensional system identification collapse, low identification accuracy, unstable convergence dynamics.



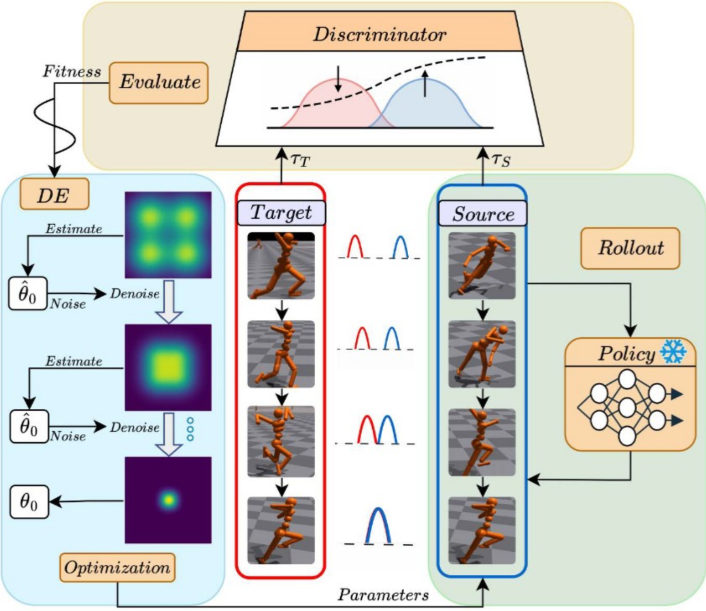
We need a sim2real method that is **data-efficient / stable and accurate in high dimensions / scalable to real robots.**

Method

Diffusion Evolution Adversarial Learning (**DEAL**), it couples a discriminator that evaluates transition similarity with a diffusion-evolution denoising process that refines parameters. The two are co-optimized, pushing the simulator toward reality.

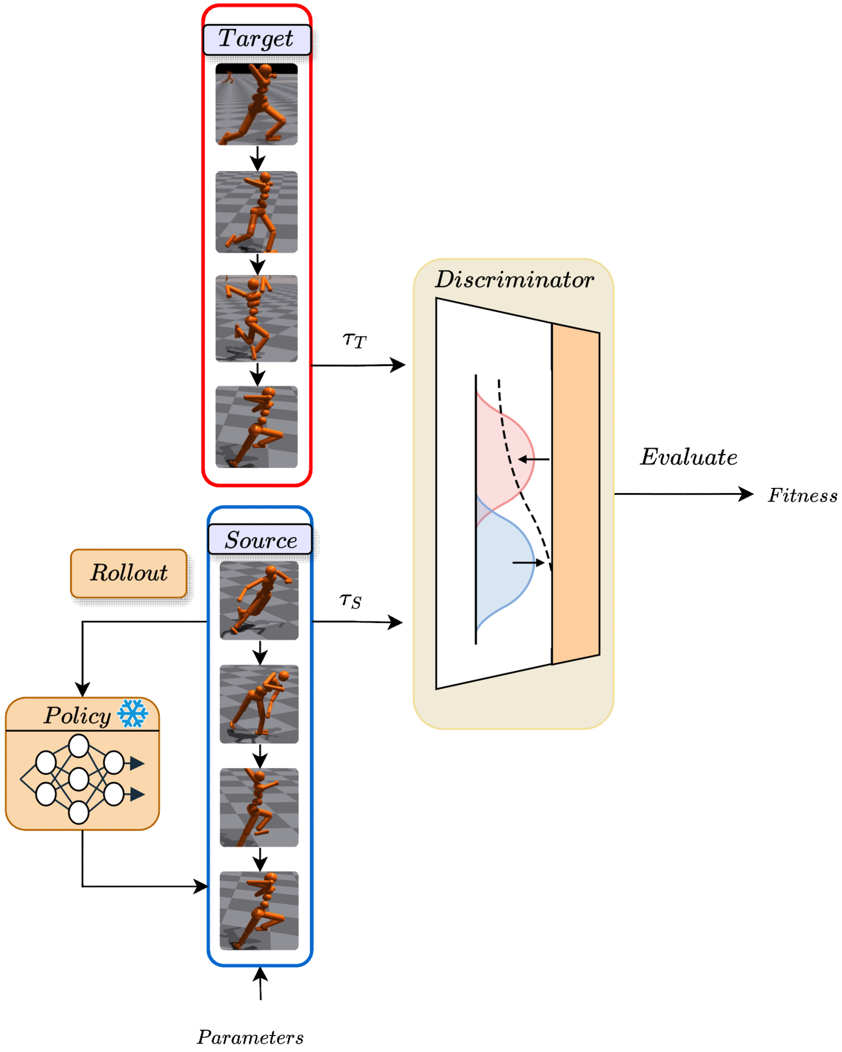
- The discriminator evaluates the similarity of state transitions sampled in source domain trajectories and target domain trajectories as fitness

- The DE estimates the optimal parameter based on the fitness probabilities, adaptively updates noise predictions and performs denoising to optimize parameter distributions until convergence.



Schematic overview of DEAL.

Method



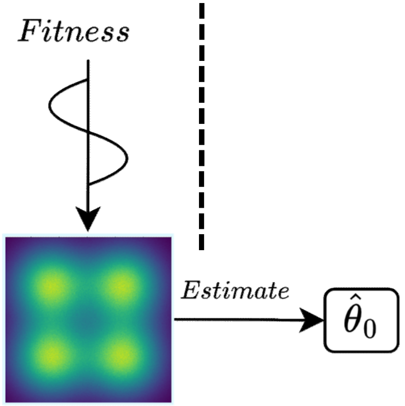
- The discriminator evaluates the fitness

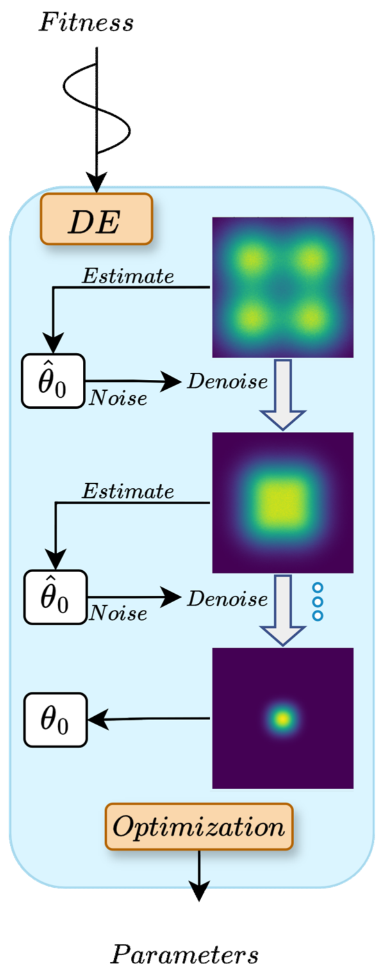
$$\max_{D} \mathbb{E}_{(\mathbf{s},\mathbf{a},\mathbf{s}')\sim d^{\mathcal{T}}(\theta^{target},\pi_0)}[D(\mathbf{s},\mathbf{a},\mathbf{s}')] - \mathbb{E}_{(\mathbf{s},\mathbf{a},\mathbf{s}')\sim d^{\mathcal{S}}(\theta,\pi_0)}[D(\mathbf{s},\mathbf{a},\mathbf{s}')]$$

Method

- The DE performs denoising to optimize parameter

$$\hat{\theta}_0 \leftarrow \frac{1}{Z}\sum_{i=1}^{N} p_i \cdot \mathcal{N}(\theta_t^{(j)}; \sqrt{\alpha_t}\theta_t^{(i)}, 1-\alpha_t)\theta_t^{(i)}$$



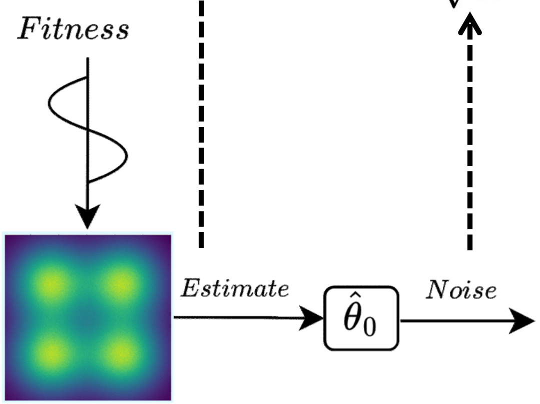


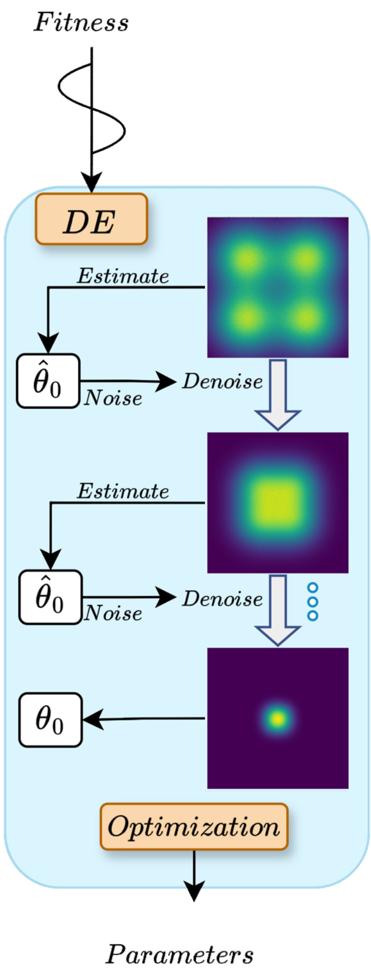
Method

- The DE performs denoising to optimize parameter

$$\hat{\theta}_0 \leftarrow \frac{1}{Z}\sum_{i=1}^{N} p_i \cdot \mathcal{N}(\theta_t^{(j)}; \sqrt{\alpha_t}\theta_t^{(i)}, 1-\alpha_t)\theta_t^{(i)}$$

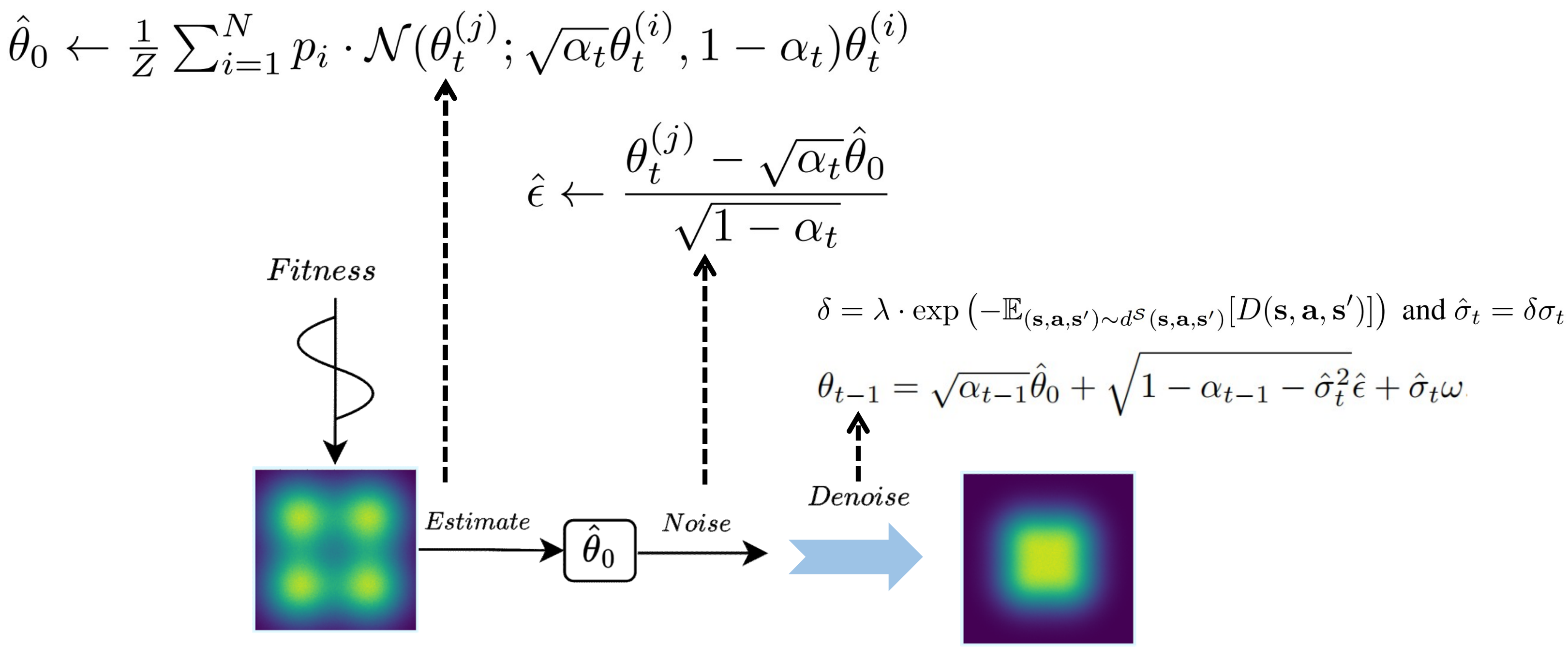
$$\hat{\epsilon} \leftarrow \frac{\theta_t^{(j)} - \sqrt{\alpha_t}\hat{\theta}_0}{\sqrt{1-\alpha_t}}$$

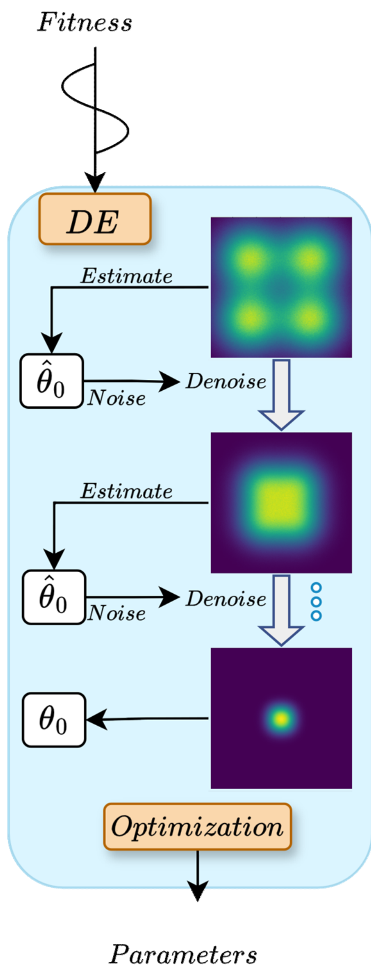




Method

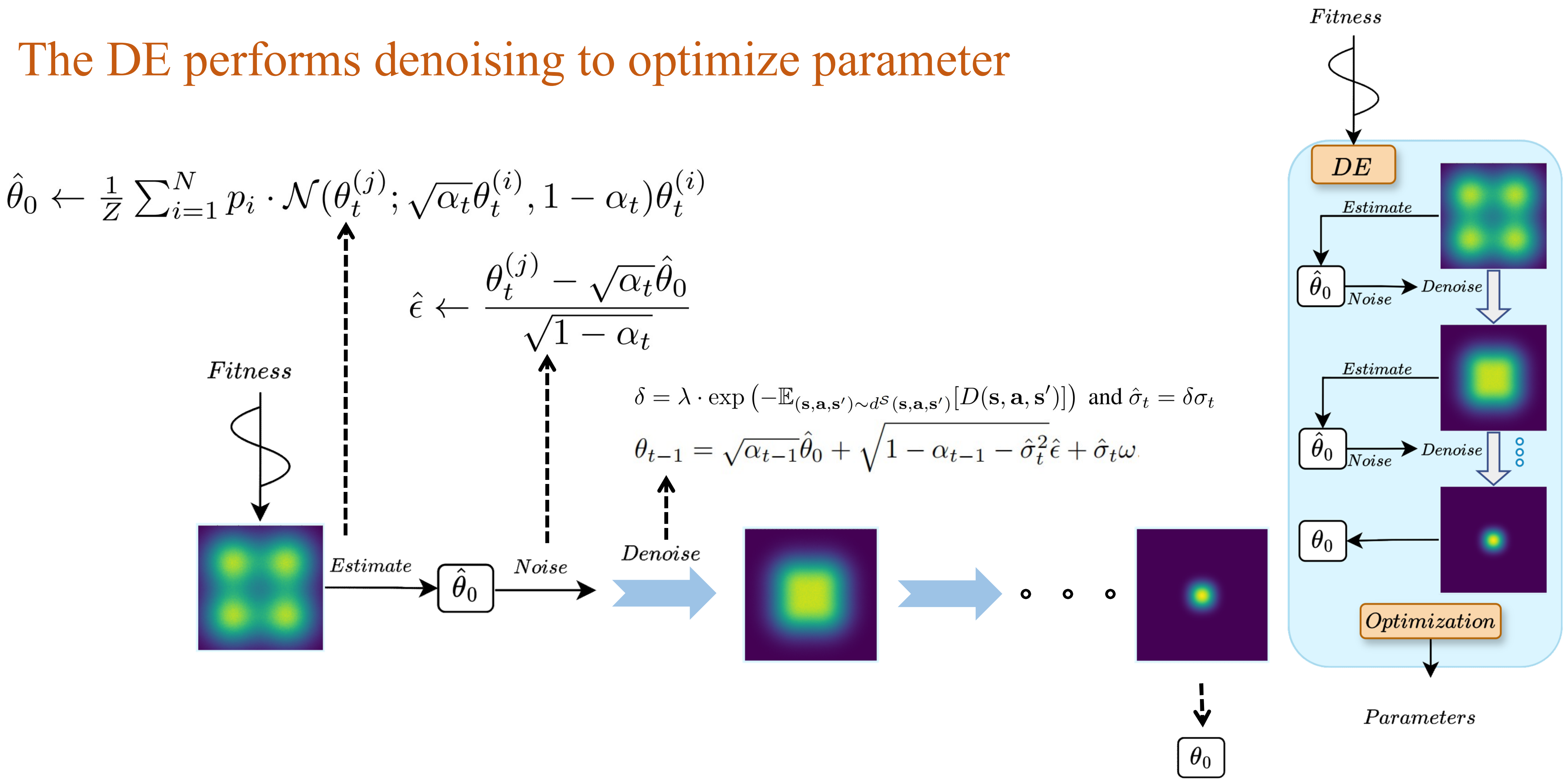
- The DE performs denoising to optimize parameter





Method

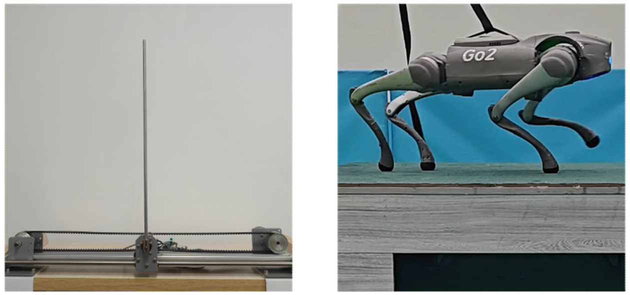
- The DE performs denoising to optimize parameter



Experiments

We test DEAL in 5 sim-to-sim tasks and 2 sim-to-real tasks.

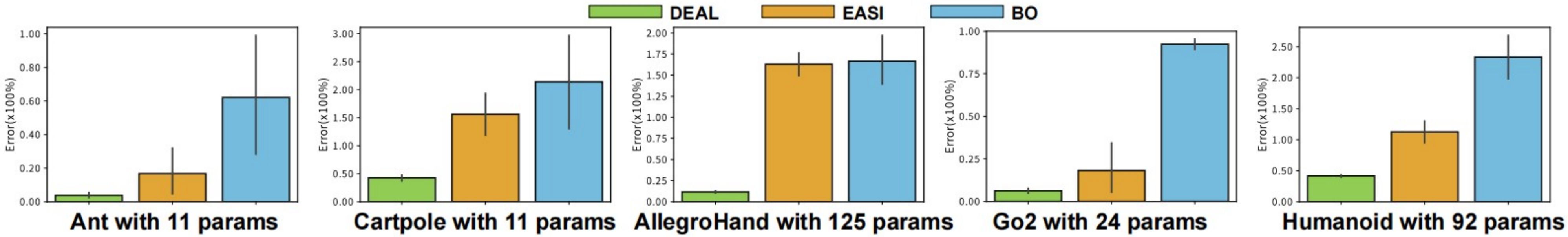
Experiment tasks in simulation.



Experiment tasks in reality.

Experiments

- **Parameter Search Capability**



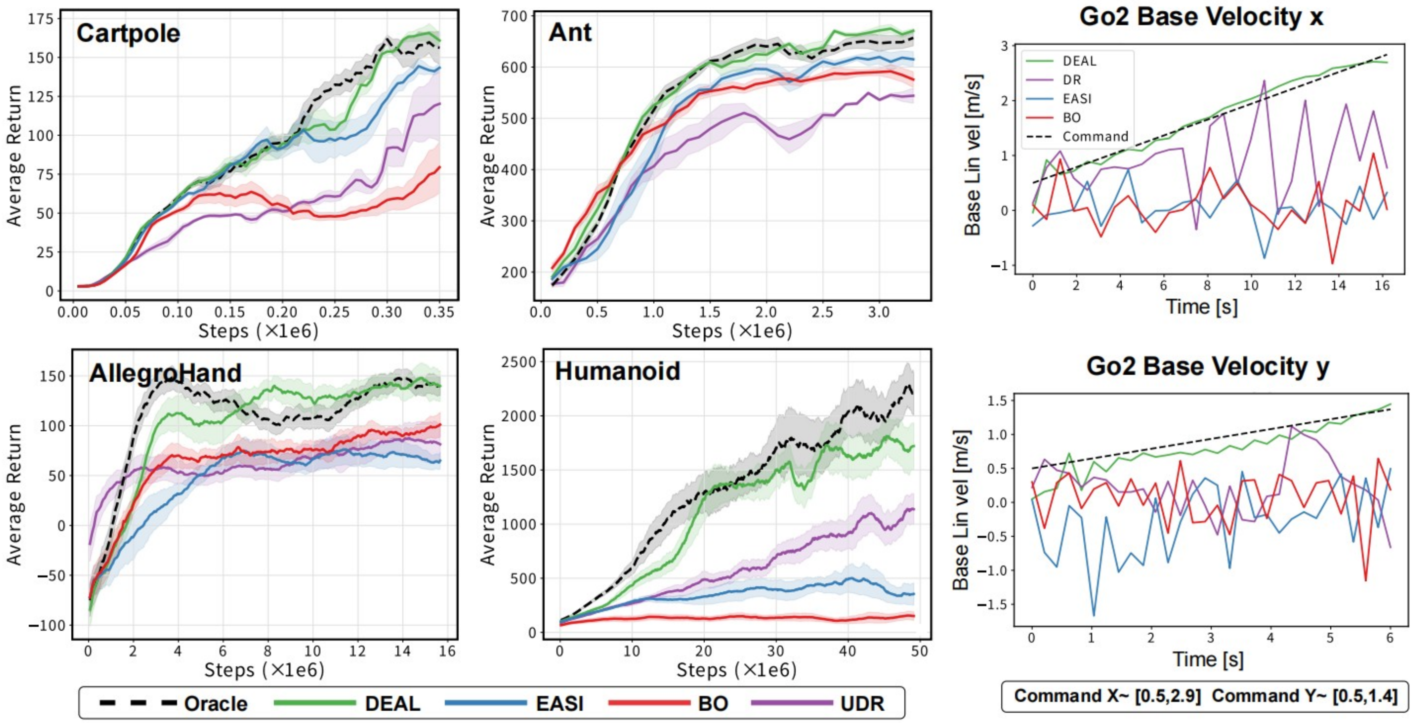
Average search errors for each method

Avg. Search Error (%)	CartPole
DEAL	**6.9±1.1**
Model-based EKF	37.4±8.8
Least-squares	53.0±19.6

Comparison with model-based methods

Experiments

- **Sim-to-Sim Transfer**



Left: The average return in the target environment for the policies trained with each method
Right: The speed tracking display of Go2 under training with each method

Experiments

- **Parameter Search Adaptability and Data Requirements**

Table 1: Average search error percentage ($\times 100\%$) (See Appendix A.8 for error bars).

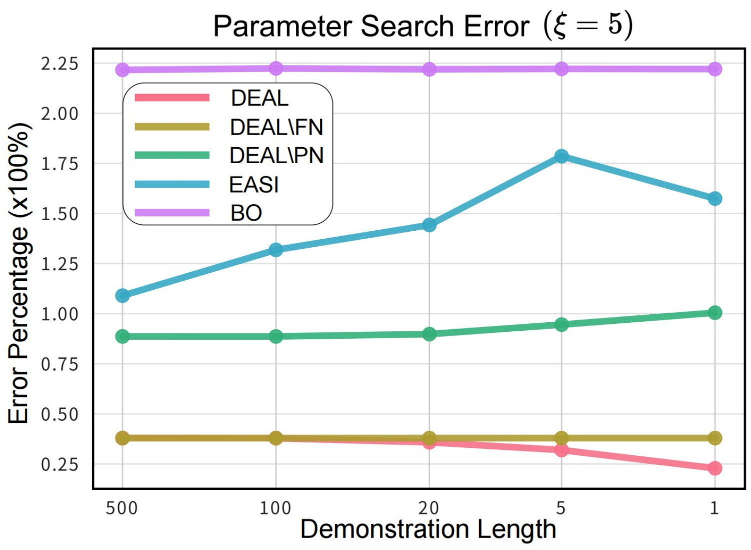
Method	Cartpole $\xi=10$	Cartpole $\xi=15$	Cartpole $\xi=20$	Humanoid $\xi=10$	Humanoid $\xi=15$	Humanoid $\xi=20$	AllegroHand $\xi=10$	AllegroHand $\xi=15$	AllegroHand $\xi=20$
DEAL	**0.85**	**1.74**	**2.63**	**0.81**	**1.65**	**2.50**	**0.83**	**1.69**	**2.57**
DEAL\PN	1.90	2.85	3.76	2.40	3.53	4.58	2.60	3.80	4.92
DEAL\FN	1.67	2.96	4.28	1.57	2.79	4.03	1.59	2.80	4.06
EASI	4.03	6.12	8.40	2.94	4.64	6.03	4.25	6.82	9.32
BO	5.20	8.52	11.78	4.67	7.29	10.06	3.75	6.32	9.16

Search for parameters on larger search scales

Table 5: Average search results at each checkpoint.

Iterations(CartPole)	Avg. Search Error (%)	Iterations(Humanoid)	Avg. Search Error (%)
50	8.84 ± 2.96	5e3	13.59 ± 5.37
100	7.87 ± 2.35	1e4	14.70 ± 4.80
250	7.19 ± 2.96	2e4	13.08 ± 5.00
1000	7.34 ± 2.88	2.5e4	10.24 ± 4.13
2500	5.10 ± 2.02	3e4	10.78 ± 3.40

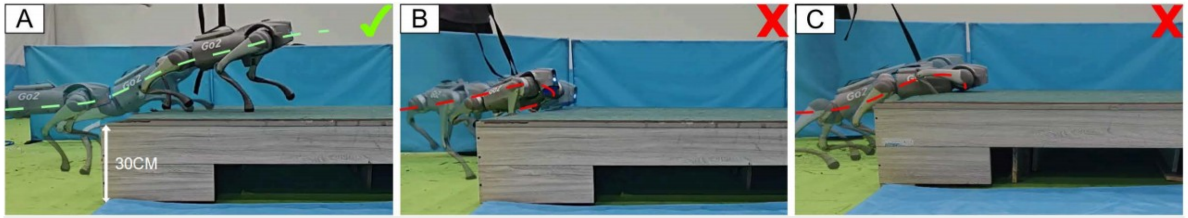
Impact of trajectories quality



Average search errors of DEAL and other baselines when given different demonstration lengths in Humanoid task

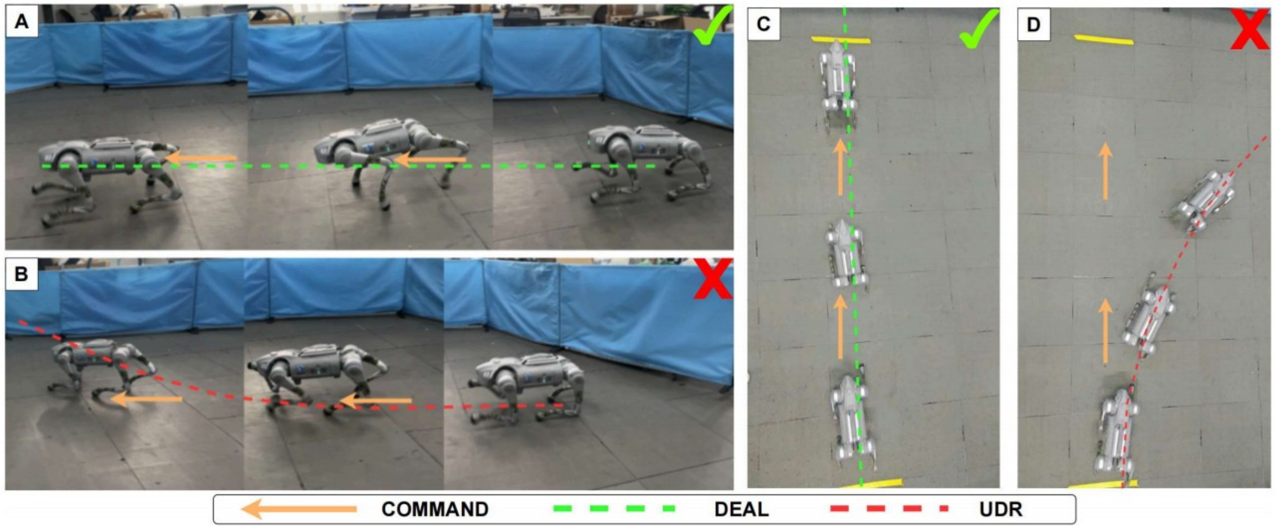
Experiments

- **Sim-to-Real Transfer**



Method	Angle Error $\times 10^{-2}$	Cart Vel $\times 10^{-1}$
UDR	3.655 ± 1.122	1.480 ± 0.367
DEAL	1.372 ± 0.382	1.214 ± 0.118

Cartpole sim-to-real performance



Go2 sim-to-real performance

Experiments

The experimental results demonstrate that:

- DEAL has demonstrated strong search capabilities in various environments.
- DEAL significantly improves the transfer performance in both simulation and real world.
- DEAL is data-efficient and robust to data quality.

Thank You

Contact: wentaoxu@smail.nju.edu.cn

Robotics & Reinforcement Learning Control