清華大學
Tsinghua University

ODRL: A Benchmark for Off-Dynamics Reinforcement Learning

Jiafei Lyu[1], Kang Xu[2], Jiacheng Xu[3], Mengbei Yan[1], Jing-Wen Yang[2], Zongzhang Zhang[3], Chenjia Bai[4], Zongqing Lu[5], Xiu Li[1]

[1]Tsinghua Shenzhen International Graduate School, Tsinghua University

[2]Tencent

[3]National Key Laboratory for Novel Software Technology, Nanjing University

[4]Institute of Artificial Intelligence (TeleAI), China Telecom

[5]School of Computer Science, Peking Unveracity

NeurIPS 2024 Dataset and Benchmark Track

Introduction

- Human beings are able to transfer the policies swiftly to a structurally similar task
- This ability is also expected in decision-making agents, especially embodied AI
- In practice, we may train the robot in a simulated environment (i.e., source domain) and deploy the learned policy in real-world tasks (i.e., target domain), where the dynamics gap may pertain between them. The robot is expected to adapt to real-world dynamics quickly
- Such a setting is referred to as **off-dynamics RL** in previous work
- Existing researches realize policy adaptation under dynamics mismatch via system identification, domain randomization, etc.
- Despite that this is an active field, unfortunately, this field lacks a standard and unified benchmark
- Upon checking the latest off-dynamics RL methods, we found that they often manually construct their customized environments with dynamics shifts and conduct experiments on them. The results can be unreliable due to a lack of unified testbed

清華大學
Tsinghua University

Contribution

- We give a formal definition of the general off-dynamics RL setting
- We propose the first off-dynamics RL benchmark, which offers different experimental settings, diverse task categories, and dynamics shift types under a unified framework
- We isolate algorithm implementations into single files to facilitate a straightforward understanding of the key algorithmic designs
- We conduct extensive experiments to investigate the performance of existing methods under different dynamics shifts and experimental settings, and conclude some key observations and insights

Definition 1 (Off-dynamics RL setting)**.** *The agent has access to sufficient data from the source domain $\mathcal{M}_{\mathrm{src}}$ and a limited budget of data from the target domain $\mathcal{M}_{\mathrm{tar}}$, where there exist dynamics shifts between $\mathcal{M}_{\mathrm{src}}$ and $\mathcal{M}_{\mathrm{tar}}$. The agent aims at getting better performance in the target domain $\mathcal{M}_{\mathrm{tar}}$ by leveraging data from both domains.*

清華大學
Tsinghua University

Benchmark overview

Table 1: **A comparison between ODRL and other RL benchmarks.**

Benchmark	Offline Datasets	Diverse Domains	Multi-task	Single-task Dynamics Shift
D4RL [19]	✔	✔	✗	✗
DMC suite [65]	✗	✔	✗	✗
Meta-World [81]	✗	✗	✔	✗
RLBench [32]	✔	✗	✔	✗
CARL [5]	✗	✔	✗	✔
Gym-extensions [29]	✗	✗	✔	✔
Continual World [72]	✗	✗	✔	✗
ODRL (this work)	✔	✔	✗	✔

Benchmark overview

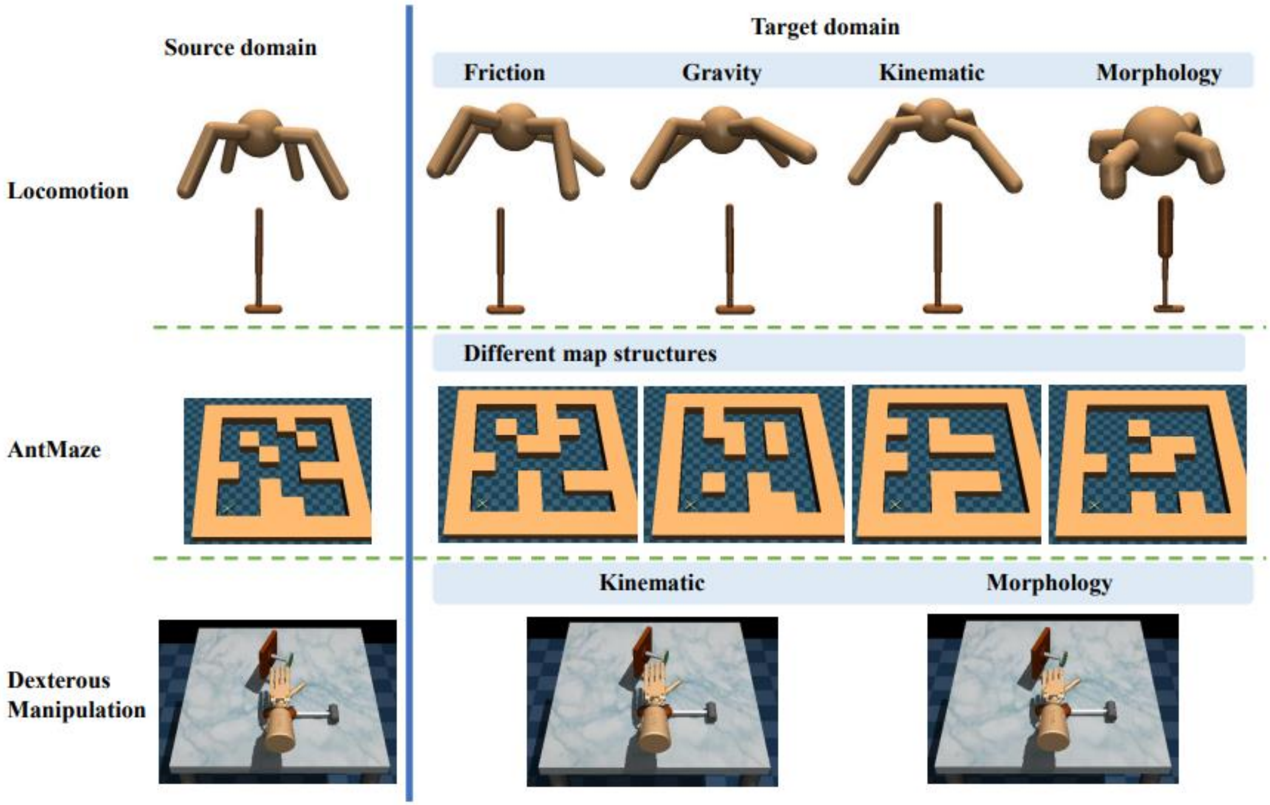
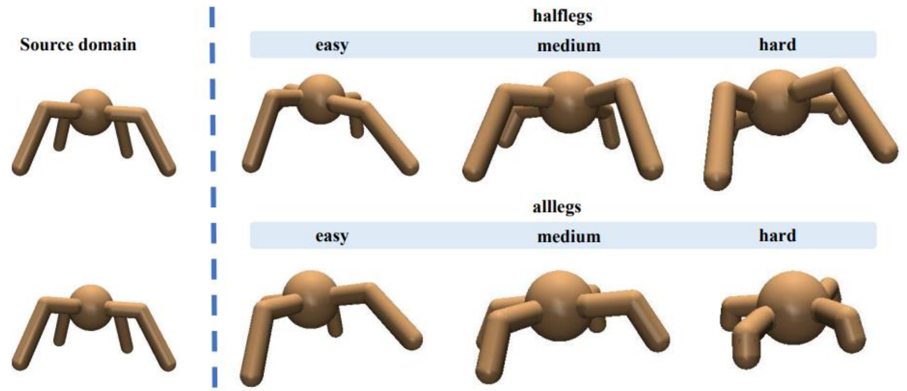
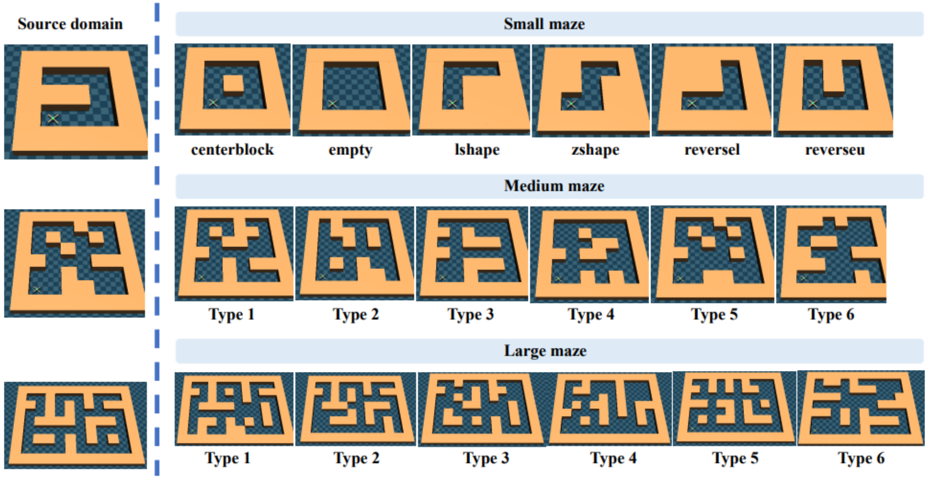


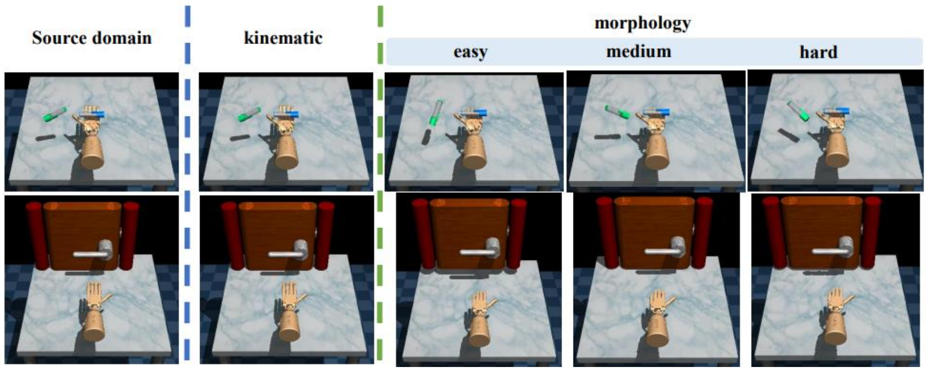
Figure 1: **An overview of selected benchmark tasks.** ODRL includes multiple domains with various types of dynamics shifts, making it a reliable platform for evaluating policy adaptation ability.

Benchmark tasks



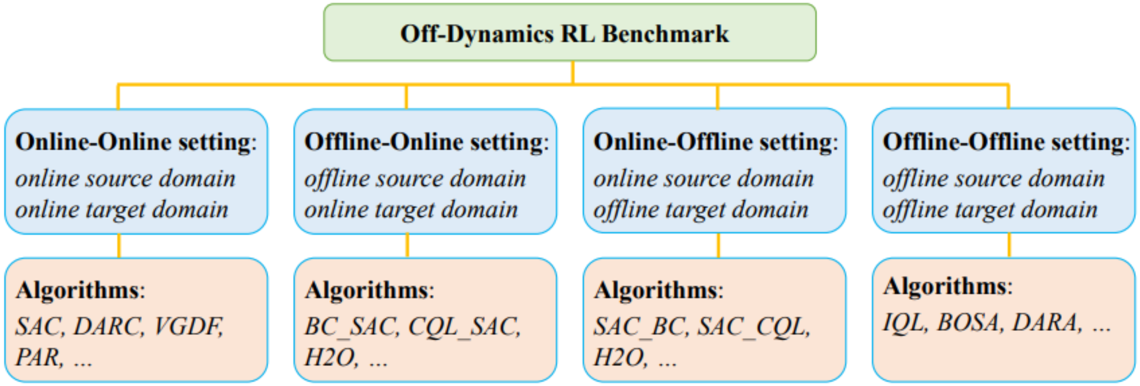
- Locomotion (HalfCheetah, Hopper, Walker2d, Ant)
  - Friction shifts: 0.1/0.5/2.0/5.0
  - Gravity shifts: 0.1/0.5/2.0/5.0
  - Kinematic shifts: easy/medium/hard
  - Morphology shifts: easy/medium/hard
- AntMaze
  - Map layout shifts under different map sizes
- Dexterous manipulation
  - Kinematic shifts: easy/medium/hard
  - Morphology shifts: easy/medium/hard





Tsinghua University

Implemented algorithms

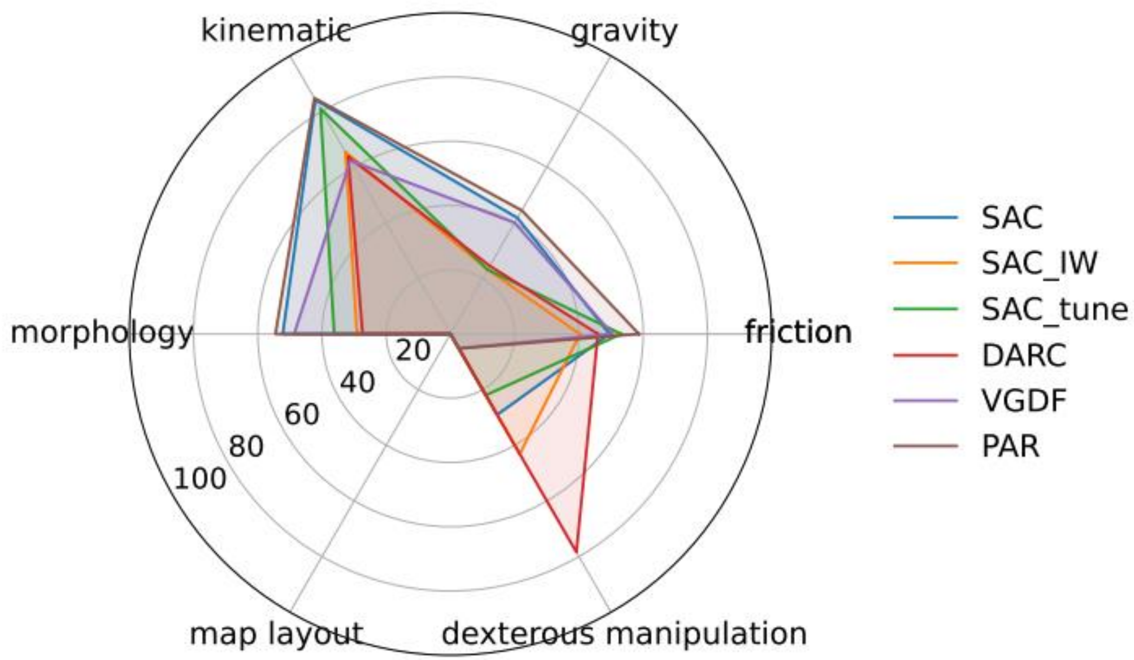


- Four different kinds of settings: Online-Online, Online-Offline, Offline-Online, Offline-Offline
- We provide offline datasets for every target domain task! The offline datasets have limited budget according to the definition of off-dynamics RL (limited budget in the target domain)
- We provide single-file implementation of the following methods:
  - Online-Online: DARC, VGDF, PAR, SAC, SAC_tune, SAC_IW
  - Offline-Online: H2O, BC_VGDF, BC_PAR, BC_SAC, MCQ_SAC, CQL_SAC, RLPD
  - Online-Offline: H2O, PAR_BC, SAC-BC, SAC_CQL, SAC_MCQ
  - Offline-Offline: DARA, BOSA, IQL, TD3BC
- Evaluation protocol:
  - We use two metrics: return and normalized score

$$NS = \frac{J_\pi - J_r}{J_e - J_r} \times 100$$

Experimental results

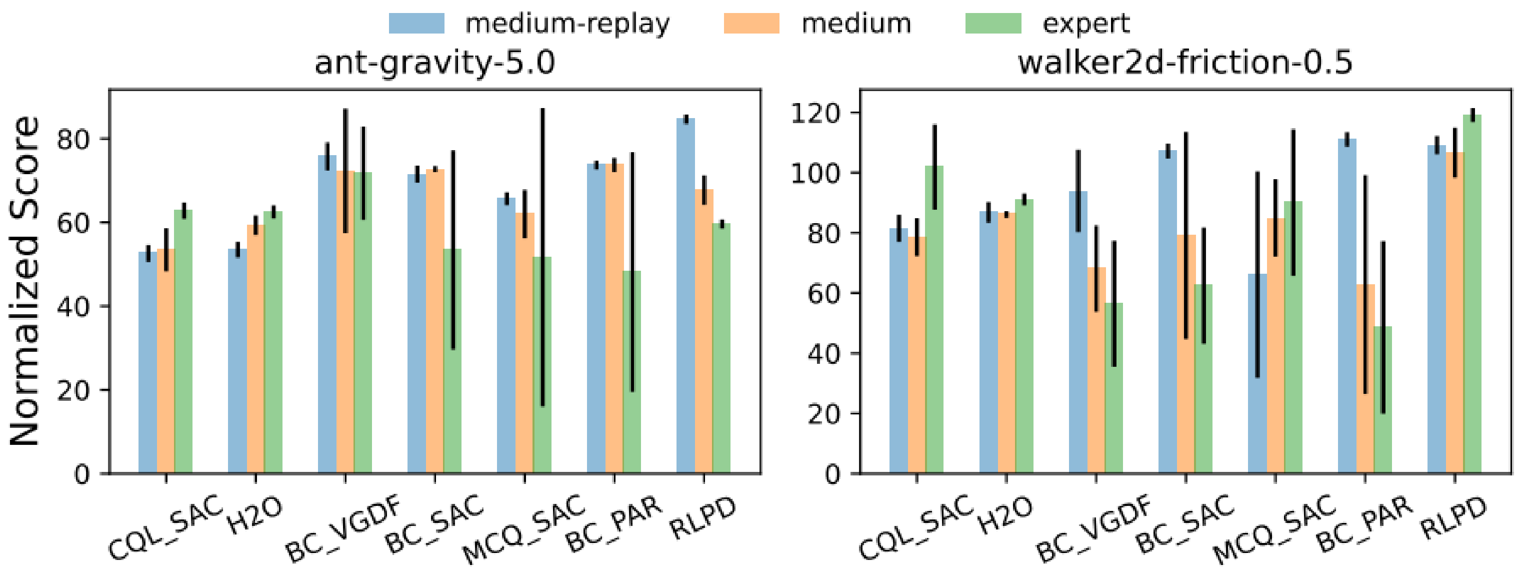
- We conduct experiments on some selected tasks. Online-Online setting
- **Obs 1.** *No single off-dynamics RL algorithm can exhibit advantages across all scenarios*.
- **Obs 2.** *PAR achieves the best performance on locomotion tasks but fails on the Antmaze domain and Adroit domain.*



Tsinghua University

Experimental results

- We conduct experiments on some selected tasks. Offline-Online setting
- **Obs 6.** *A higher quality of the source domain dataset does not necessarily imply better performance in the target domain, even when an expert source domain dataset is provided.*
- **Obs 7.** *Baseline methods that treat two domains as one mixed domain can achieve good performance on some tasks, sometimes even surpassing off-dynamics methods like BC_PAR, BC_VGDF, and H2O.*



清華大學
Tsinghua University

Experimental results

- If you have any problem, feel free to contact the authors via:

lvjf20@mails.tsinghua.edu.cn

- Please find codes in https://github.com/OffDynamicsRL/off-dynamics-rl
- Please find the paper in https://arxiv.org/pdf/2410.20750?