Final Project: Reinforcement Learning from Images

Principles of Artificial Intelligence
Fall 2022 (JCCX0021)
Shanghai Jiao Tong University

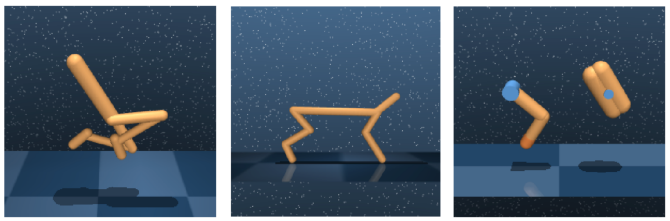
1 Introduction

The camera is a convenient and inexpensive way to acquire information, especially in complex, non-stationary, and unstructured environments, such as autonomous driving. Thus, effective reinforcement learning (RL) approaches that can leverage images as input have received widespread attention and have been employed for a wide range of real-world applications. The existing methods can be roughly grouped into two categories, that is, model-free methods [4, 7] and model-based methods [3, 2, 8, 6]. In this project, your goal is to implement the RL algorithm (model-based RL or model-free RL) in the following environment to achieve good performance.

In these visual control tasks, the agents learn the action policy directly from high-dimensional observations. We formulate visual control as a partially observable Markov decision process (POMDP) with a tuple (S, A, T, R, O), where S is the state space, A is the action space, O is the observation space, $R(s_t, a_t)$ is the reward function, and $T(s_{t+1} \mid s_t, a_t)$ is the state-transition distribution. In this setting, the agent interacting with the environment doesn't have access to the actual states in S, but to some partial information in the form of observations. Therefore, the input of your model can only be the observations in O. At each timestep t, the agent takes an action $a_t \in A$ to interact with the environment and receives a reward $r_t = R(s_t, a_t)$. We consider episodic environments with the length fixed to T. The goal of standard RL is to learn a policy that maximizes the expected cumulative reward $E_p[\sum_{\tau=t+1}^{T} r_\tau]$.

2 Environment

The DeepMind Control Suite (DMC) is a set of simulated continuous control tasks with a standardized structure and interpretable rewards, intended to serve as performance benchmarks for reinforcement learning agents. The tasks are written in Python and powered by the MuJoCo physics engine, making them easy to use and modify. The Control Suite is publicly available at `https://github.com/deepmind/dm_control`. Please refer to the starter code in the attachments for more details about how to set up this environment.



(a) Walker walk (b) Cheetah run (c) Finger spin

Figure 1: Three tasks from DMC environment used in this project.

Please train and evaluate your models on the three tasks of DMC environment, *i.e.*, Walker walk, Cheetah run, and Finger spin, as shown in Figure 1. Table 1 gives the comparison of some current methods. In all three tasks, SAC [1]:pixel (an agent that learns from pixels) is significantly outperformed by SAC:state (an agent that learns from states), which shows that learning from pixels is a more challenging task. PlaNet [3] and SLAC [5] are the model-based methods. We may play with

Table 1: A comparison of current methods

TASK	NUMBER OF EPISODES	SAC:PIXEL	PLANET	SLAC	SAC:STATE
WALKER WALK	1000	33 ± 2	949 ± 9	864 ± 35	$\mathbf{974 \pm 1}$
CHEETAH RUN	3000	366 ± 68	701 ± 6	$\mathbf{830 \pm 32}$	$\mathbf{836 \pm 105}$
FINGER SPIN	1000	645 ± 37	659 ± 45	$\mathbf{900 \pm 39}$	$\mathbf{945 \pm 19}$

with these baseline models (most of them are open sourced) and are expected to improve them for better control results (in terms of rewards) or sample efficiency (by a faster training convergence).

3 Submissions

In this project, you need to submit:

- **A final report** that is expected to cover the following sections: the background description, literature review, proposed method, technical details, experimental results, and conclusions. You may use the NeurIPS template provided in the attachments.
- **The source code and the pre-trained models.** Note that "readme.md" is also required, in which you need to describe how to train and evaluate your model. Typically, we encourage you to provide the additional materials, *i.e.*, video demos, proofs, and experimental results of other environments (such as Atari) to support your model, and give a certain bonus.
- **A poster** that is expected to comprehensively show the key idea and main results of the project. You may use the template (in PPT format) we have provided, or you may also find the LaTeX templates on Overleaf.
- **Slides** for the 12min oral presentation in the class.

All the above materials should be zipped to the single file, named after "studentID+name", and submitted to the Canvas.

4 Scoring

The score (ONLY for the technical report) includes the following five parts:

- Background knowledge and literature review. (20%)
- Correctness of the method. (20%)
- Technical novelty. (5%)
- Reproducibility of experimental results, and rich quantitative and qualitative results compared to the existing methods. Note that since there are three groups working on the same project, we'll give higher scores to the better-performed models. (35%)
- Writing. (20%)

References

[1] Tuomas Haarnoja, Aurick Zhou, Kristian Hartikainen, George Tucker, Sehoon Ha, Jie Tan, Vikash Kumar, Henry Zhu, Abhishek Gupta, Pieter Abbeel, et al. Soft actor-critic algorithms and applications. *arXiv preprint arXiv:1812.05905*, 2018.

[2] Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *ICLR*, 2020.

[3] Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *ICML*, pages 2555–2565. PMLR, 2019.

[4] Michael Laskin, Aravind Srinivas, and Pieter Abbeel. CURL: Contrastive unsupervised representations for reinforcement learning. In *ICML*, pages 5639–5650. PMLR, 2020.

[5] A. X. Lee, A. Nagabandi, P. Abbeel, and S. Levine. Stochastic latent actor-critic: Deep reinforcement learning with a latent variable model. In *NeurIPS*, 2020.

[6] Minting Pan, Xiangming Zhu, Yunbo Wang, and Xiaokang Yang. Iso-dream: Isolating noncontrollable visual dynamics in world models. In *NeurIPS*, 2022.

[7] Denis Yarats, Amy Zhang, Ilya Kostrikov, Brandon Amos, Joelle Pineau, and Rob Fergus. Improving sample efficiency in model-free reinforcement learning from images. In *AAAI*, pages 10674–10681, 2021.

[8] Amy Zhang, Rowan McAllister, Roberto Calandra, Yarin Gal, and Sergey Levine. Learning invariant representations for reinforcement learning without reconstruction. In *ICLR*, 2021.