

Augmented Dreams: Data Augmentation & Self-supervised learning in Model-Based Reinforcement Learning

Karim Farid, Diego Fernandez and Sai Prasanna R.

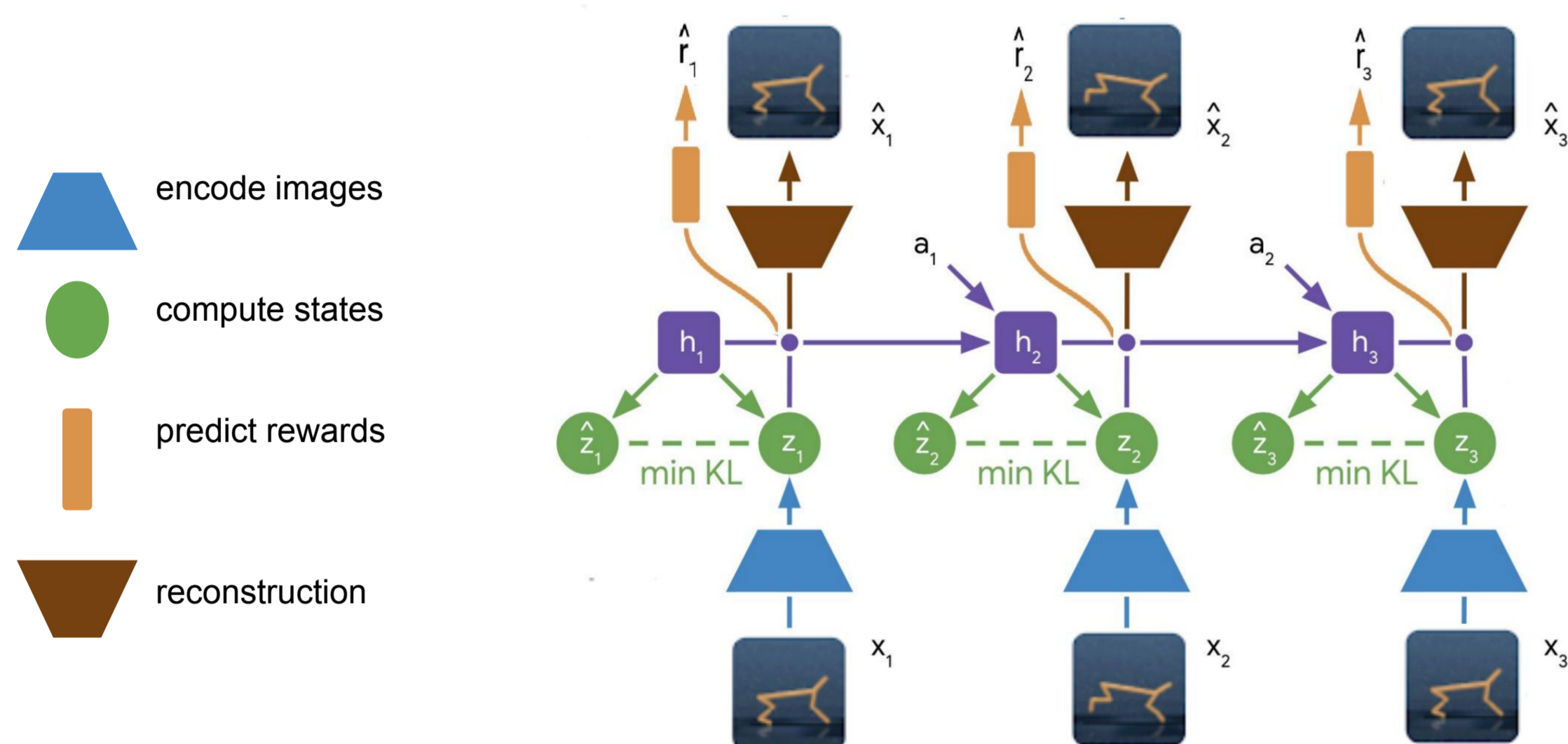


Motivation

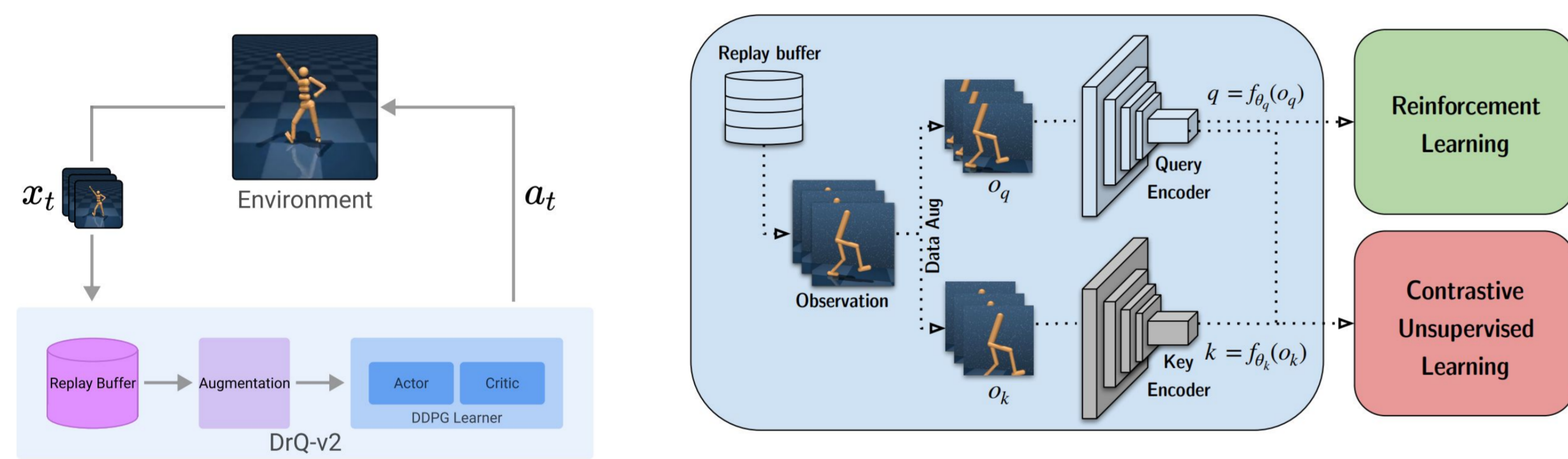
- **Sample efficiency** for high-dimensional inputs is an important challenge in reinforcement learning.
- **Model-based reinforcement learning** is one approach to improve sample efficiency.
- State of the art **model-free methods** for pixel inputs use **data augmentations** and **self-supervised learning**.
- **Idea**
 - Apply data augmentation and self-supervised learning over **Dreamer** (v2), a model-based reinforcement learning algorithm.
 - Test the efficacy of data augmentations for generalization to distracting backgrounds.

Our Approach

Dreamer World Model Architecture



Data Augmentation and SSL

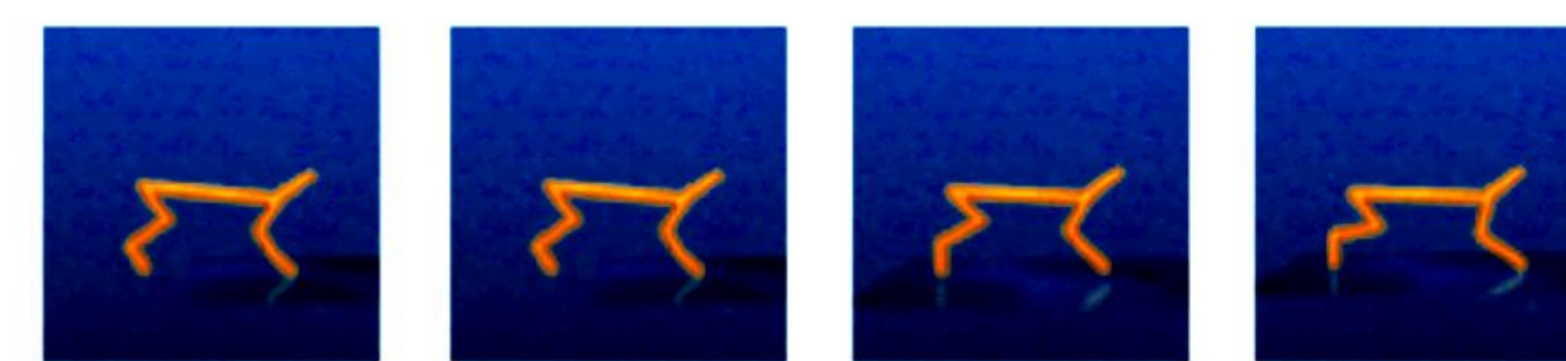


- **Data augmentation only:** Model-free methods include DrQv2, RAD.
- **Data Augmentation with self-supervised learning auxiliary loss:** CURL

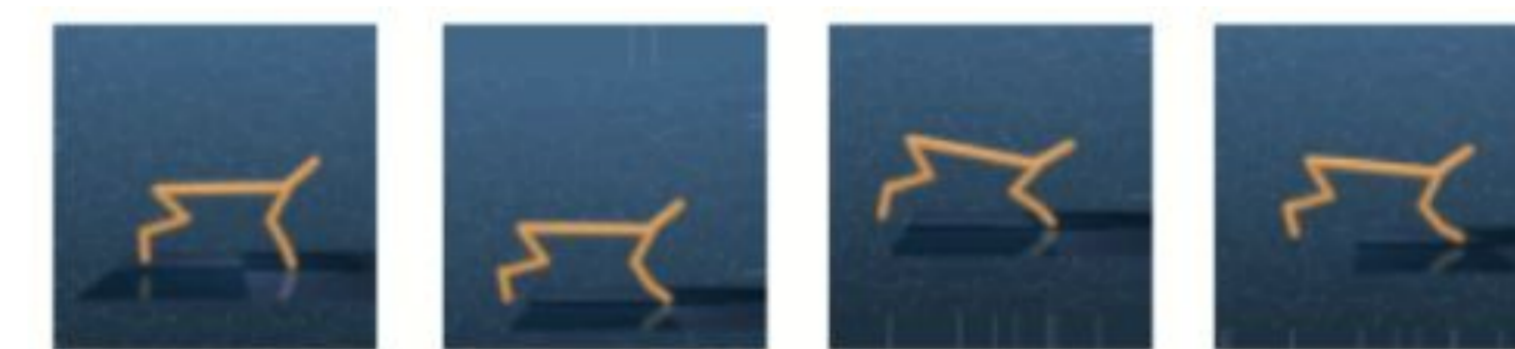
Data Augmentation approaches

We augment all the observations in an episode **consistently**.

- **Color Augment:** Augmenting the brightness, contrast, adding color jitter, etc.



- **RandShift Augment:** Translate the image by 4 pixels with padding using pixels at the border



- **CURL:** Random crop 64x64 pixel from 84x84 pixel images

SSL on world-model states

- Apply contrastive learning on augmented trajectories for the same episode encoded into states by the world-model.
- Allows us to **remove** the costly **reconstruction loss**.
- We tried **triplet loss**, **barlow twins** loss and, **contrastive predictive coding**.
- Requires more experimentation with hyperparameter tuning to evaluate the approaches thoroughly.

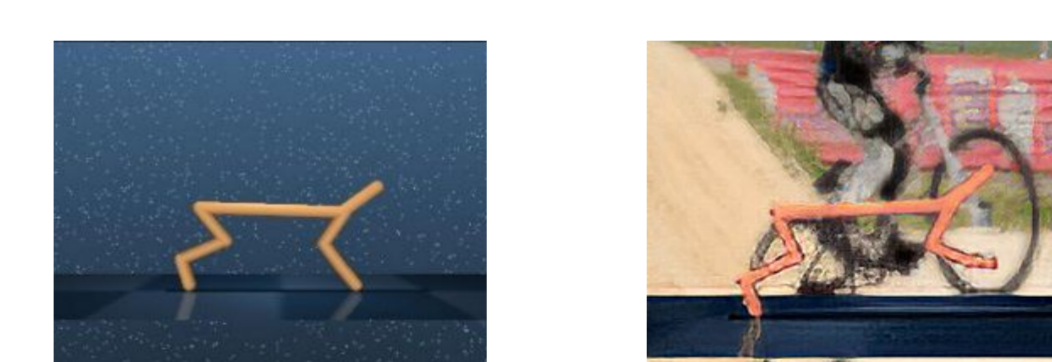
Experimental Results

- We **out-perform** model-free methods on **2 out of 3** tested environments in dmcontrol100k benchmarks.
- **Colour Augmentation** proves robust to changes in the background for distracting control suite. It also has good gains to sample efficiency on two normal setting.
- "Finger, Spin" environment is challenging for the world model in all the settings requiring further investigation.
- The **batch size** is an important hyperparameter for sample efficiency in dreamer with augmentation.

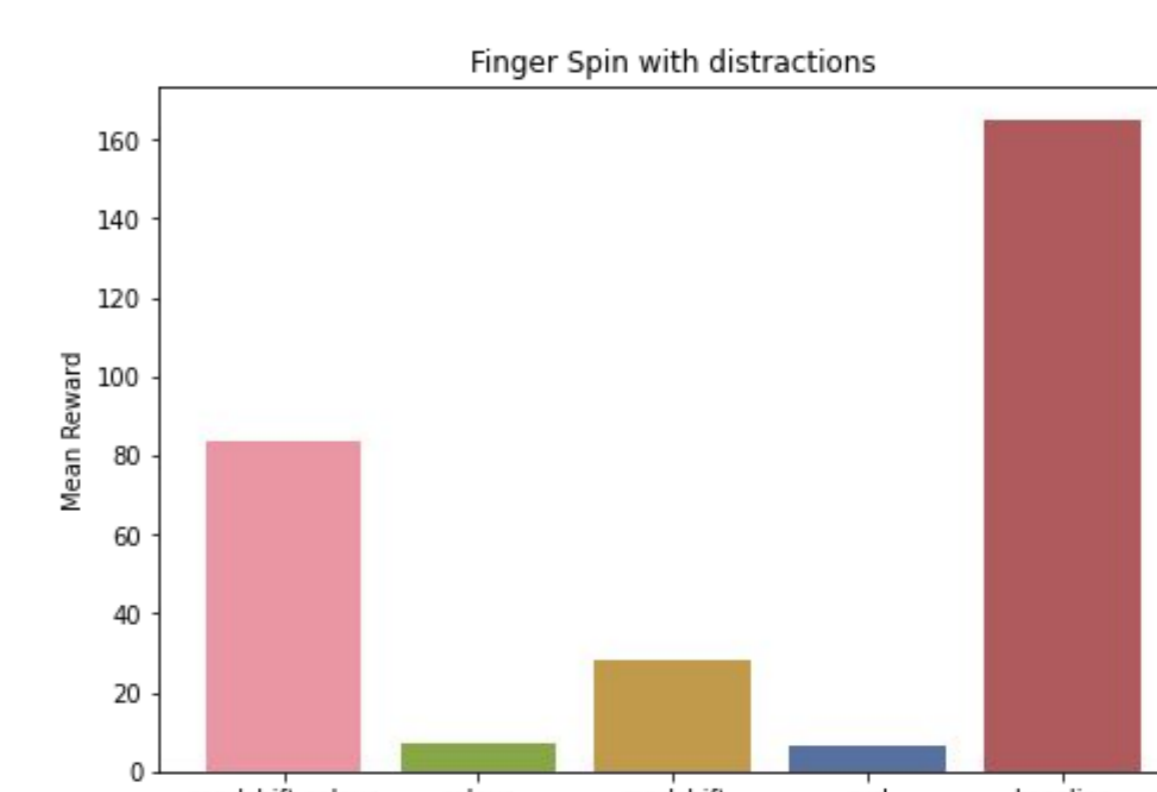
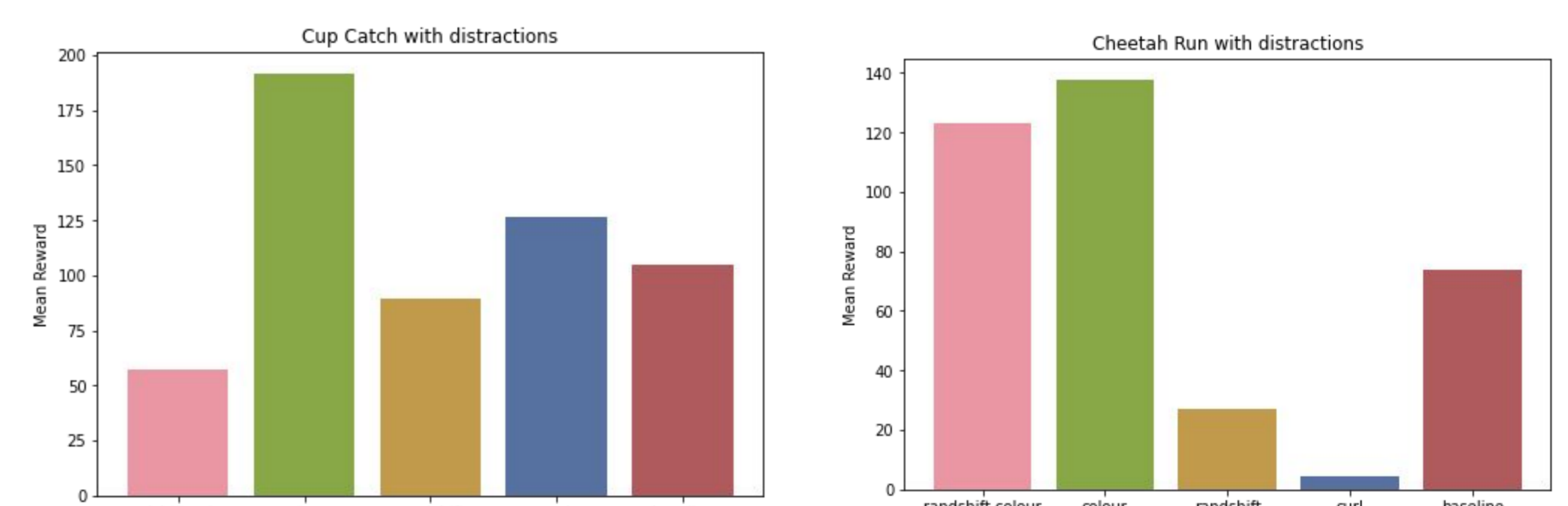
Environment	CURL	DrQv2	RAD	Dreamer 2	Dreamer 2 +Colour	Dreamer 2 +RandShift	Dreamer 2 +Colour +RandShift	Dreamer 2 +CURL
Cheetah, Run	299 ± 48	272 ± 129	447 ± 88	377 ± 69	470 ± 38	463 ± 3	248 ± 226	227 ± 153
Finger, Spin	767 ± 56	352 ± 310	856 ± 73	668 ± 460	270 ± 99	376 ± 288	528 ± 207	679 ± 282
Ball Cup, Catch	769 ± 43	359 ± 244	840 ± 179	654 ± 567	939 ± 65	584 ± 507	245 ± 424	976 ± 21

DMControl100k Continuous control with 100k environment steps

Distracting Control: DMControl with color and background distractors



Left: no distractors Right: background and color distractors



Code & Results



bit.ly/3BGen6c