Representation Learning for RL using CURL

Edwin R. Ranjitkumar (s140035)
02456 – Deep Learning
DTU Compute, Technical University of Denmark

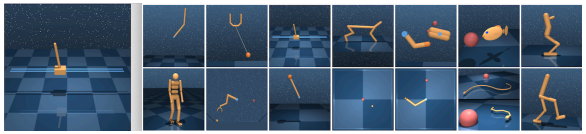
Introduction

Deep neural networks have achieved unprecedented progress in many tasks, such as image classification and object detection. Most of the progress is driven by the supervised learning paradigm, where good performance largely relies on a large number of human-annotated labels (e.g. ImageNet). However, collecting manual labels is expensive and very difficult to scale up.

However, recently contrastive unsupervised learning has emerged popular choice of method for training models without using labels, going toe to toe with pure supervised approaches to achieve state of the art results. So, the goal of the project is to work within this framework (commonly referred to as contrastive self-supervised learning) to explore its capabilities.

Challenge

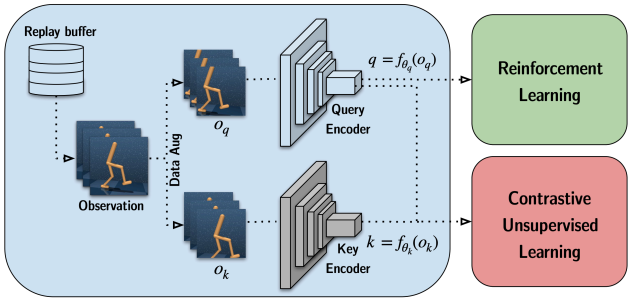
- Key task: Cartpole benchmark in DeepMind control suite (DMC).



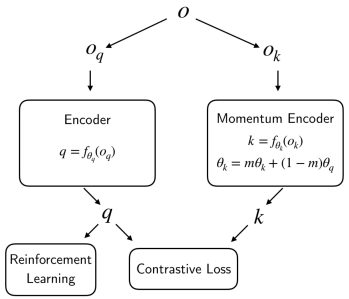
- Hypothesis: A RL agent should be more data efficient if it learns a useful representation from high dimensional observation.

Method

- Concept of CURL: A contrastive self-supervised task is jointly optimized together with the RL objective.



Model



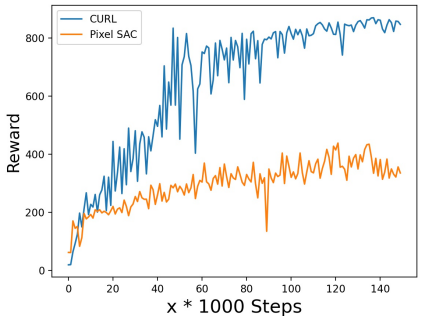
Loss function

$$\mathcal{L}_q = \log \frac{\exp(q^T W k_+)}{\exp(q^T W k_+) + \sum_{i=0}^{K-1} \exp(q^T W k_i)}$$

Experiment

Results: Effect on data efficiency

- CURL was tested on the Cartpole benchmark in comparison to a pixel based SAF



Results: Model comparisons

- Further evaluation on the CURL model was conducted with on the Walker and Cheetah benchmark in comparisons two other models.
- CURL illustrates very promising results, achieving state of the art rewards on several benchmarks.
- Although seems to have issues in the more complex benchmark environment

500k step scores	CURL	Pixel SAC	State SAC
Cartpole, Swingup	850	419 40	848 21
Walker, Walk	912	42 12	795 30
Cheetah, Run	505	197 15	948 54

Results: Ablation studies

In hope of further improvement of CURL's results, critical components of the model was tested. Studies on:

- Decoupling Representation Learning from Reinforcement Learning
- Removing Data Augmentation for the Actor Critic

Conclusion & Outlook / future work

- We described an unsupervised model based on CURL, which contrastive self-supervised learning objective for model-free RL agents.
- The model was trained to extract useful representation to solve pixel based continuous control task in DMC suite.
- CURL illustrates that it is possible to archive amazing data -efficiency results on image based RL task with unsupervised representation learning.
- The ablations test illustrates question in key aspects of the model, which leaves open questions for both in this project and in the literature.
- One direction to look into for further improvement is data augmentation.

References

[1] Aravind Srinivas, Michael Laskin, Pieter Abbeel. CURL: Contrastive Unsupervised Representations for Reinforcement Learning (2020)
[2] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, Ross Girshick. Momentum Contrast for Unsupervised Visual Representation Learning (2020)