Reinforcement Learning with Prototypical Representations

Denis Yarats
NYU & FAIR
@denisyarats

Rob Fergus
NYU
@rob_fergus

Alessandro Lazaric
FAIR
@alelazaric

Lerrel Pinto
NYU
@lerrelpinto

ICML 2021

Image-based RL vs State-based RL

- State-based D4PG (blue) vs Image-based D4PG (green).

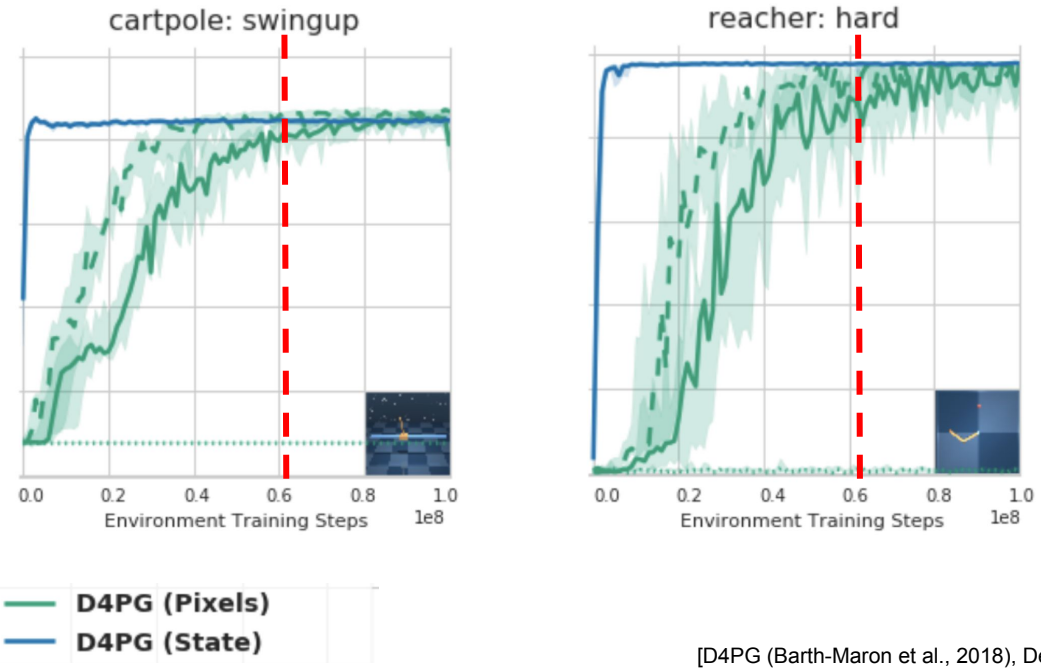


Image-based needs > 50M more training steps than state-based to solve same tasks

[D4PG (Barth-Maron et al., 2018), DeepMind Control Suite (Tassa et al., 2018), Slide credit: Pieter Abbeel]

Recent developments in Image-based RL

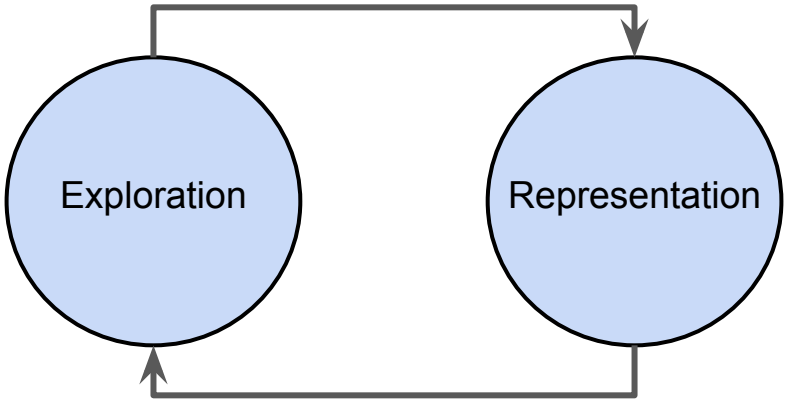
- Model-based: PlaNet (Hafner et al., 2018), Dreamer (Hafner et al., 2019).
- Auxiliary tasks: SAC-AE (Yarats et a., 2019), CURL (Srinivas et al., 2020), ATC (Stooke et al., 2020).
- Data-augmentation: DrQ (Yarats et al., 2020), RAD (Laskin et al., 2020).

What is missing?

- Learning a good representation still requires reward signal for a task.
- This means that learned representations are task-dependent and do not transfer well!
- Can we get closer to the pre-training + fine-tuning paradigm in CV/NLP?

Decoupling Representation and Exploration

- In CV/NLP, we start with a good dataset. Empirically, a better dataset means better performance.
- In RL, what should the dataset be?
  - Hint: Unsupervised exploration. But good exploration that has high coverage needs good representations to distinguish novel states from already visited.



Proto-RL

Proto-RL pre-training:

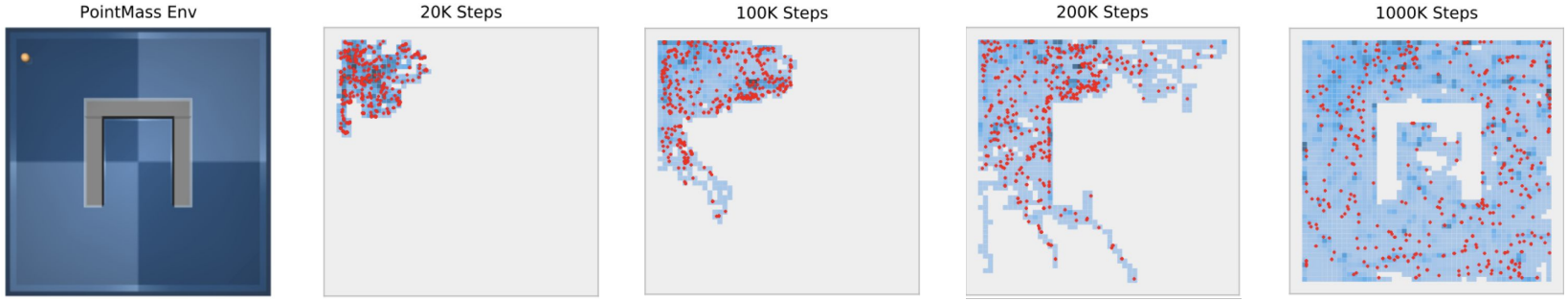
- Learns Prototypical representations via SSL on the collected dataset.
- Explores environment using MaxEnt intrinsic reward based on prototypes to collect the dataset.

Proto-RL fine-tuning:

- Learns a task-specific policy on pre-trained representations to cast image-based RL to state-based RL.

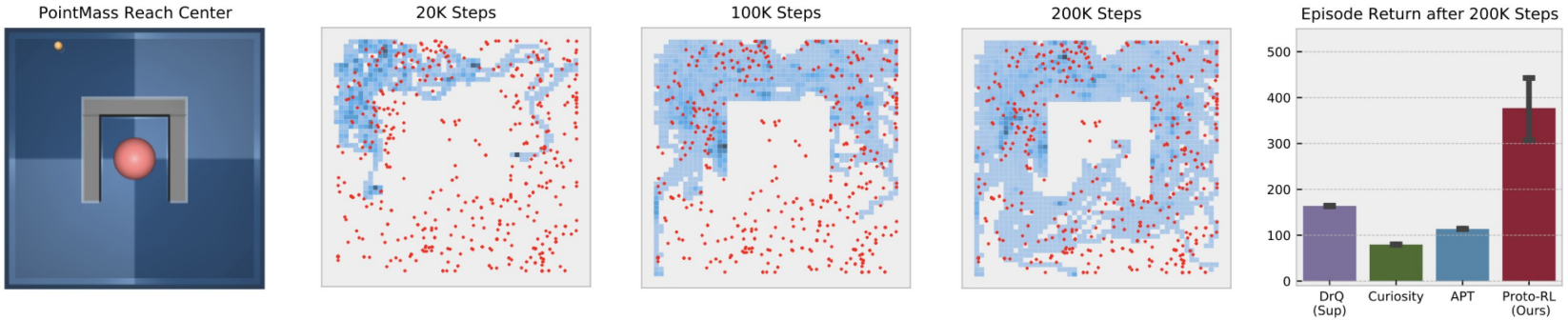
Proto-RL on a Toy Example

- Image-based continuous control environment.
- No reward is given, need to unsupervised exploration.
- Phase 1: task-agnostic pretraining.



Proto-RL on a Toy Example

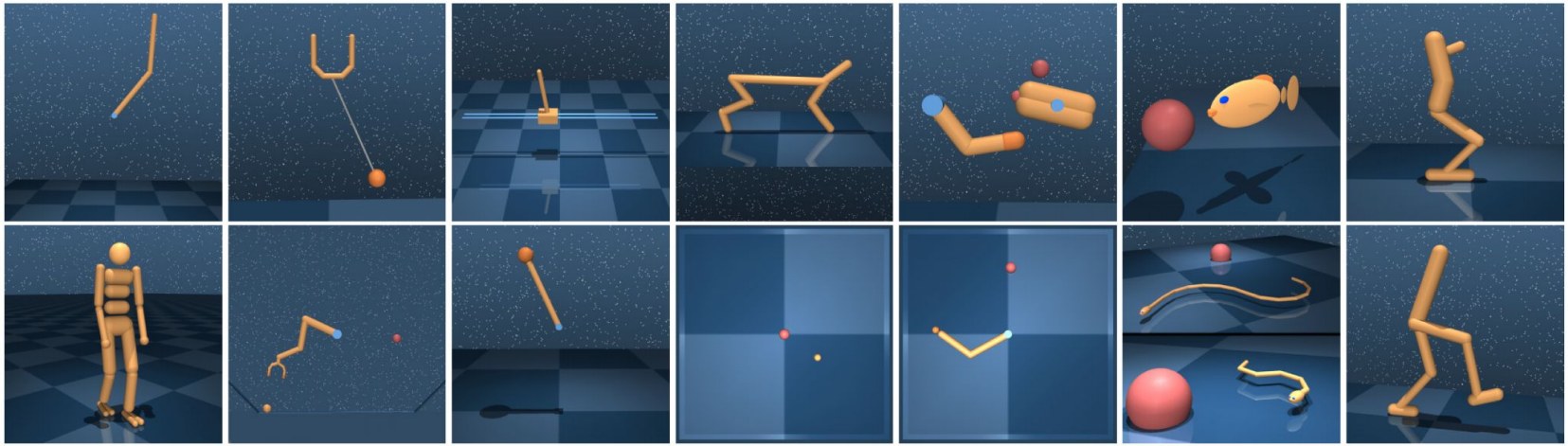
- Downstream objective is revealed (reach center).
- Can the pre-trained encoder and prototypes be useful?
- Phase 2: downstream fine-tuning.



[DrQ (Yarats et al., 2020), Curiosity (Pathak et al., 2019), APT (Hao et al., 2020)]

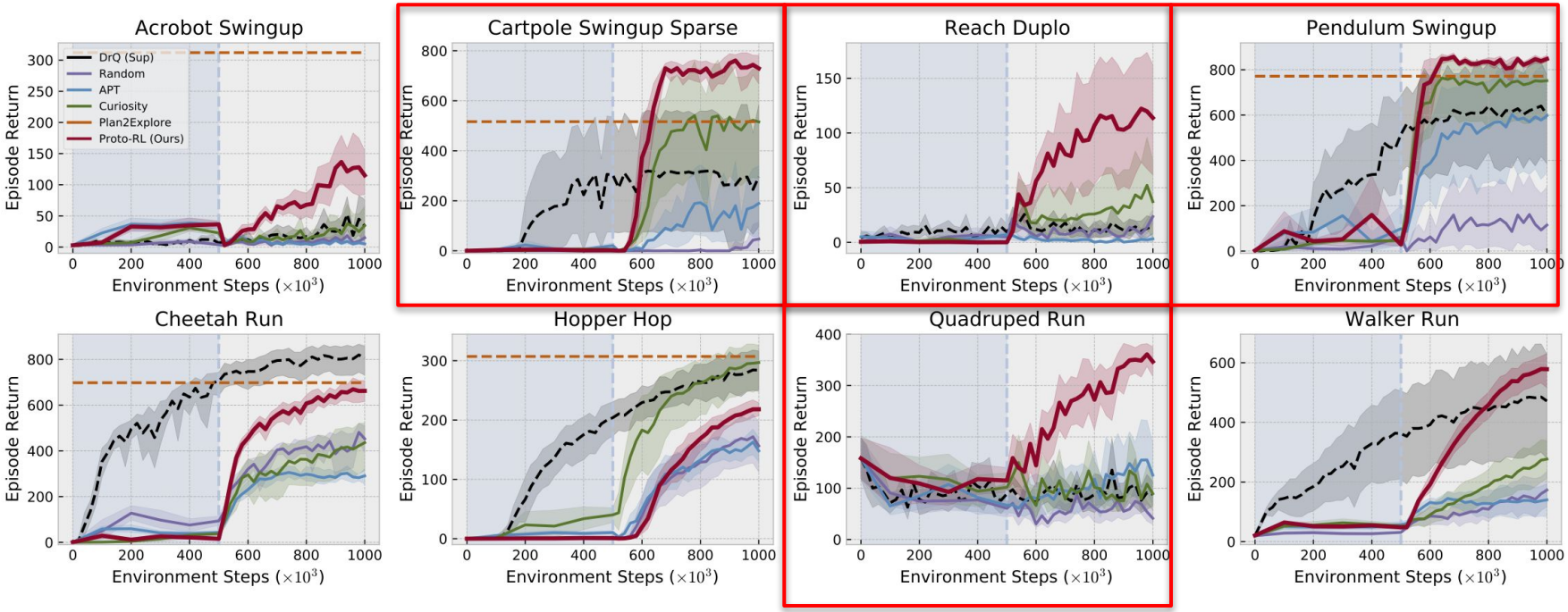
Experimental Setup and Comparisons

- DeepMind Control Suite (Tassa et al., 2018).
- Two phases: task-agnostic pretraining and downstream fine-tuning.



Large Scale Task-Agnostic Pretraining

- Does task-agnostic pre-training improve downstream RL?



[DrQ (Yarats et al., 2020), Curiosity (Pathak et al., 2019), APT (Hao et al., 2020), Plan2Explore (Sekar et al., 2020)]

Thank you for attention!