Accelerating Goal-Conditioned RL Algorithms and Research

Michał Bortkiewicz, Władysław Pałucki, Vivek Myers,
Tadeusz Dziarmaga, Tomasz Arczewski,
Łukasz Kuciński, Benjamin Eysenbach

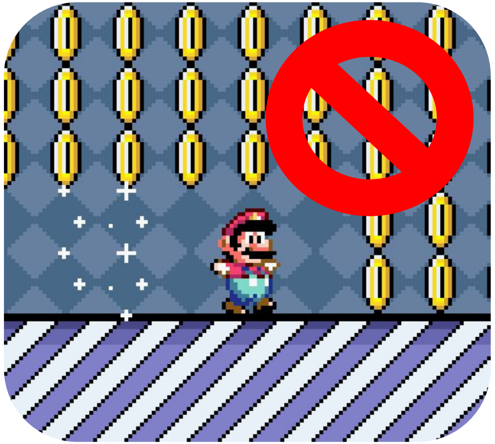
Outline

- What are the obstacles to scale up RL?
- Obstacle 1.
- JaxGCRL – our solution for speeding up and scaling up GCRL.
- Obstacle 2 and our results.

I have a dream… of Reinforcement Learning

New emergent behaviours

Unstructured Interaction

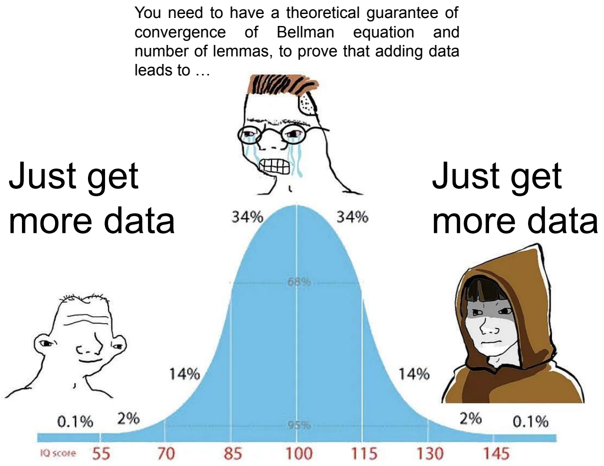


No reward shaping

Scaling up RL

Two main obstacles to scaling up RL:

- Need for massive amounts of data.
  - We use relatively small datasets of ~10^6 transitions.

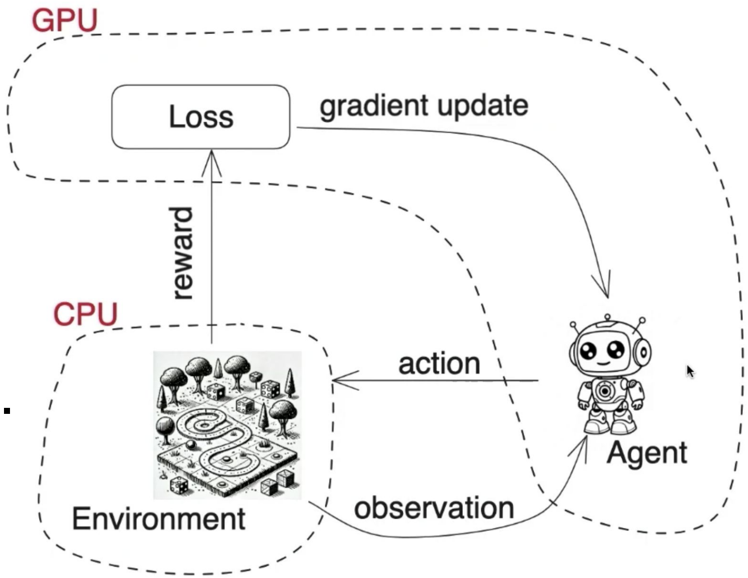


- Stable algorithms and architectures that utilize that data.
  - Data is not enough. We need proper algorithms.

Obstacle 1 - The data problem

- Time lost on transferring data between GPU and CPU.
- Small GPU utilisation.
- A lot of CPU threads to gather data.

Training takes ~ 4-8 hours for DMC environments.

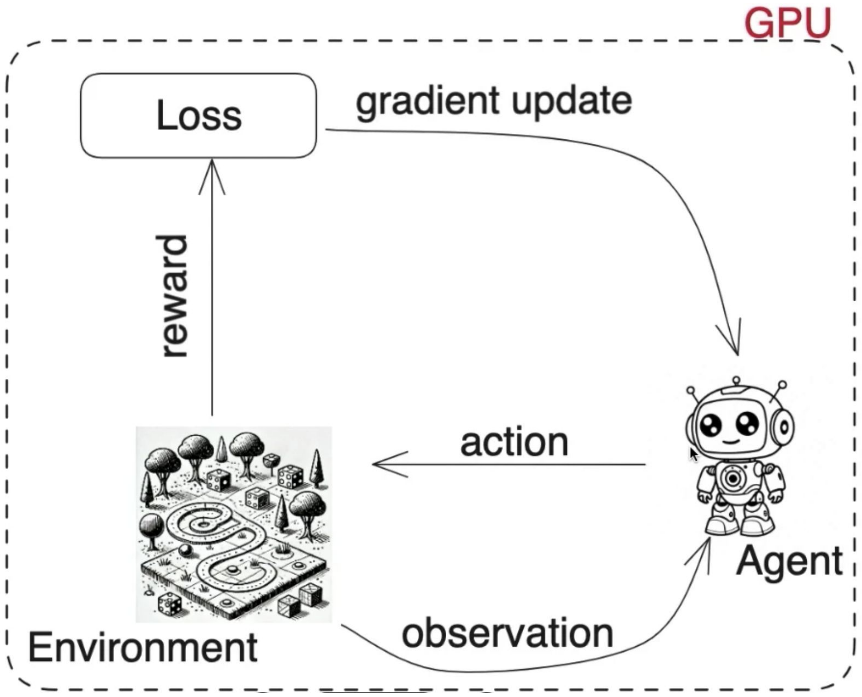


[Source: Jakob Foerster ICML 2024 talk - Reinforcement Learning at the Hyperscale]

The solution

- Use JAX vectorize environments
- Run everything on GPU
- JIT compile training loop.

1000 x

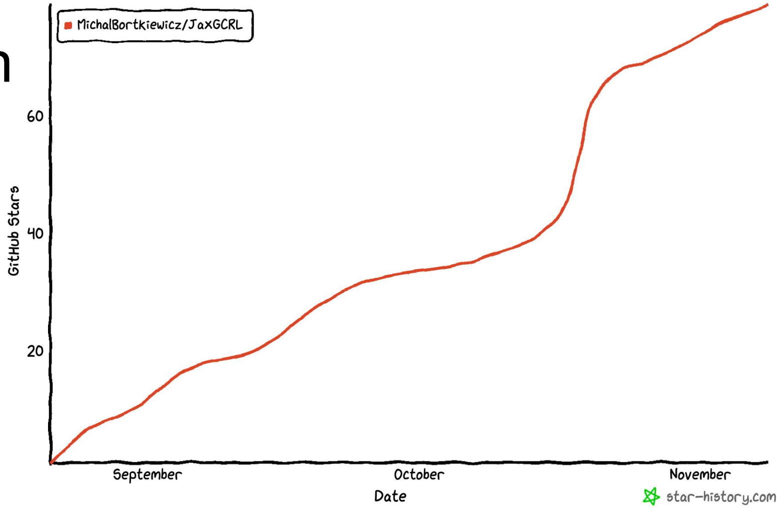


[Source: Jakob Foerster ICML 2024 talk - Reinforcement Learning at the Hyperscale]

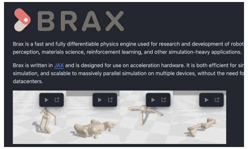
Accelerating Goal-Conditioned RL Algorithms and Research

JaxGCRL – benchmark + codebase

- 10+ GPU-accelerated BRAX/MJX environments.
- Fully JIT-compiled training.
  - `ant` training 10M steps < 10 min
- Easy to modify and extend.

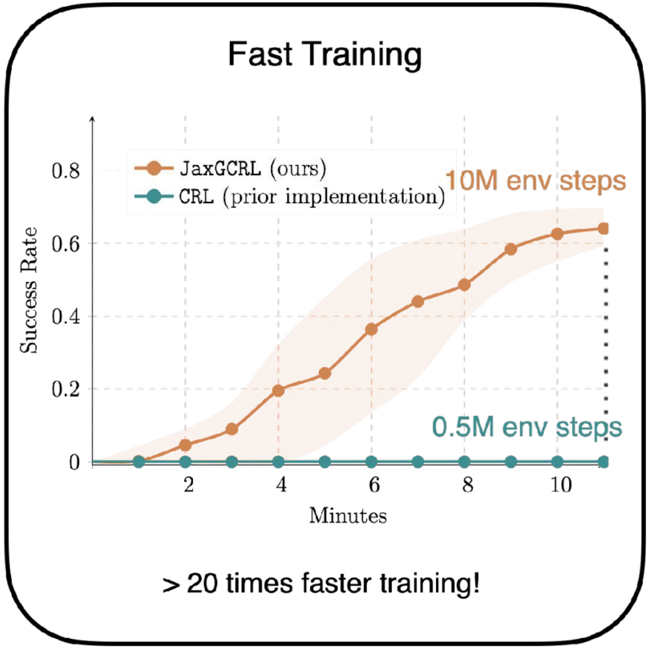


Open-source:
GPU-accelerated sim:

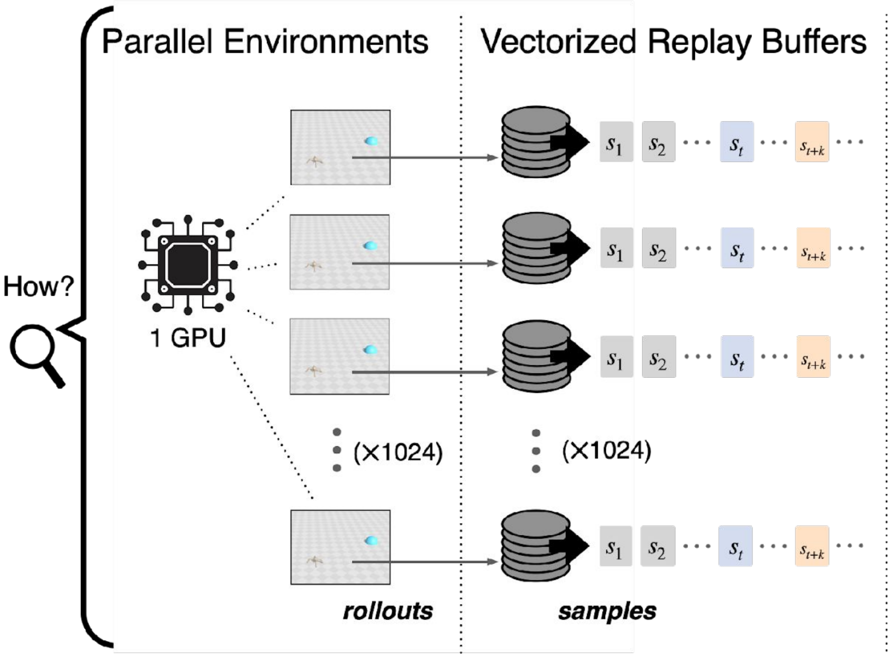


Brax, MuJoCo MJX

Training speedup in JaxGCRL



How?



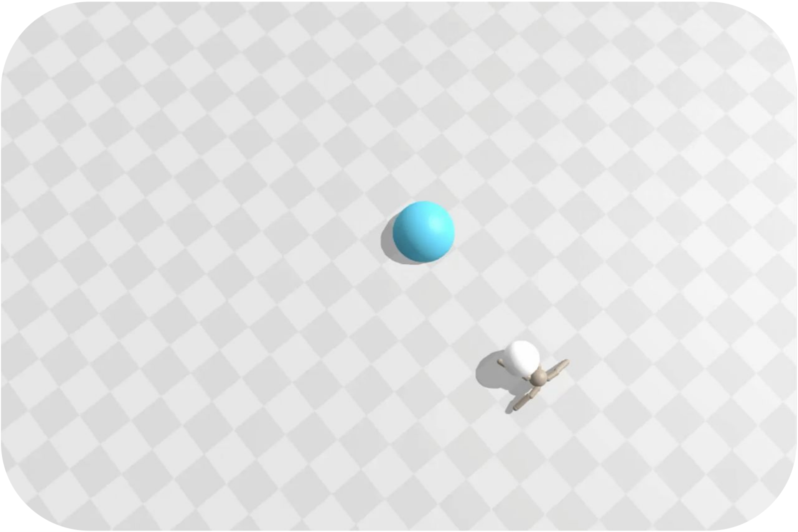
+ Training

Takeaway:

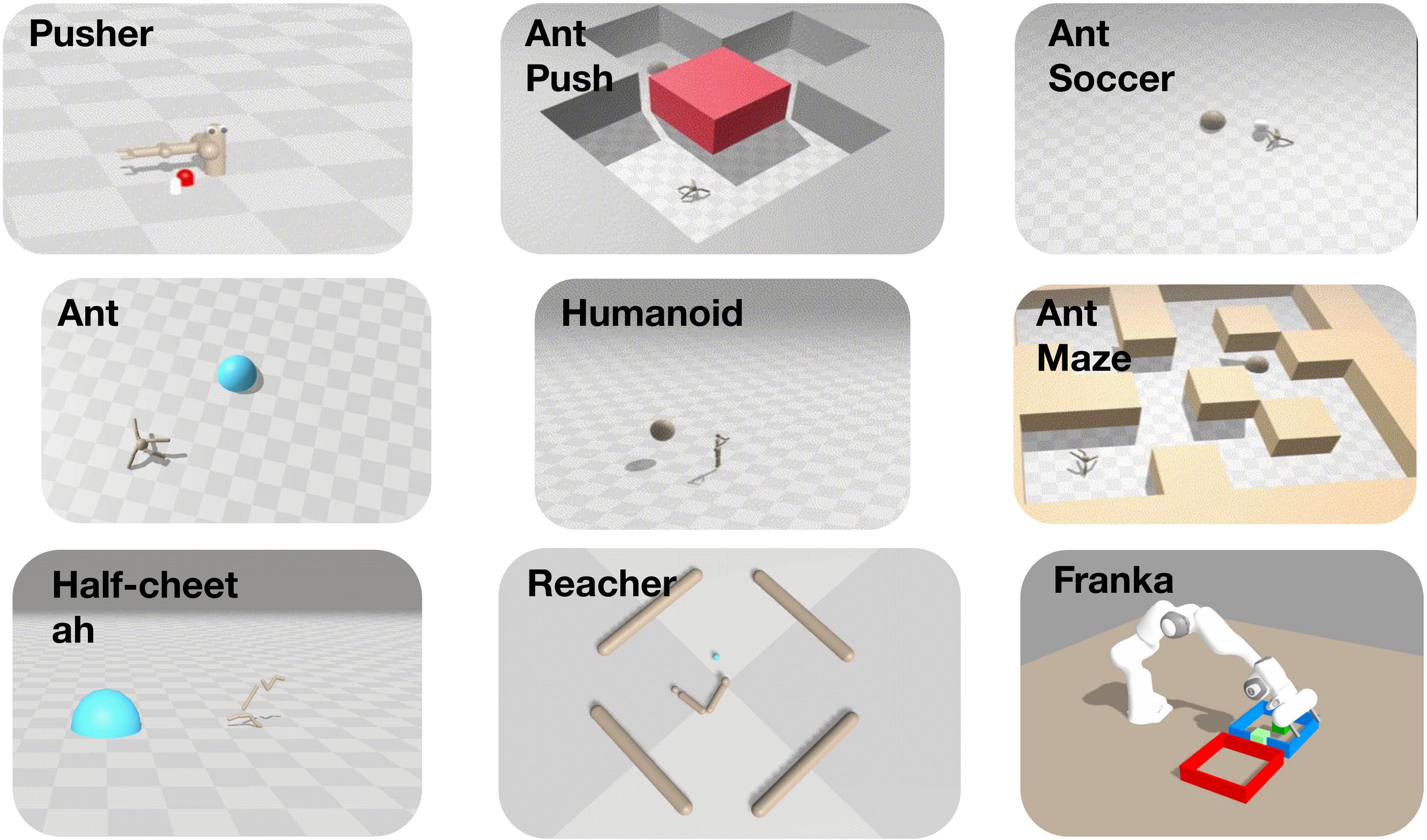
Training, data collection, and replay buffer are all run on a single GPU device.

What exactly does JaxGCRL enable?

- You can work with RL projects, like data science projects, with a quick feedback loop.
- Anyone with access to a single GPU can contribute to SOTA GCRL research.
- You can have a working policy like this one in ~10 minutes.



Environments



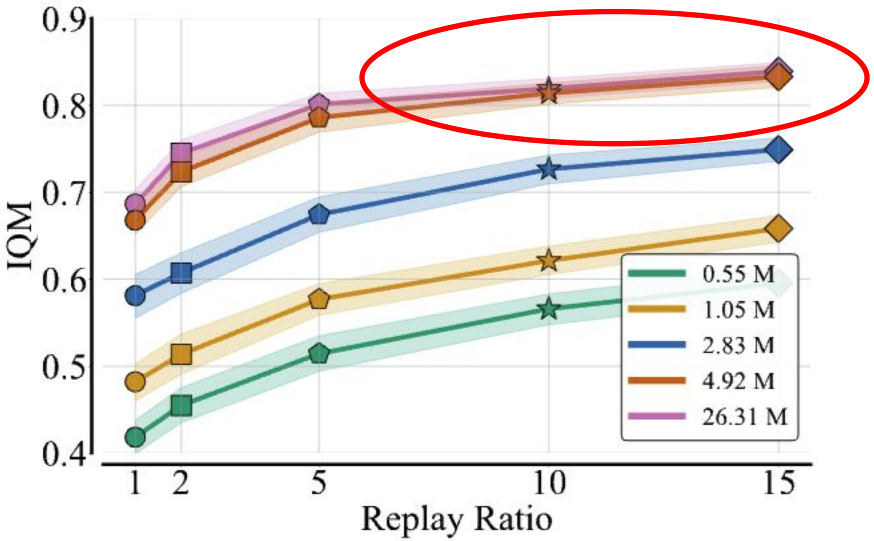
Obstacle 2 – algorithms and architectures that scale

Obstacle 2 - algorithms and architectures

- We use relatively small model architectures.
- The performance saturates quickly with model size.

BRO Performance

(Nauman et al., NeurIPS 2024 Spotlight)

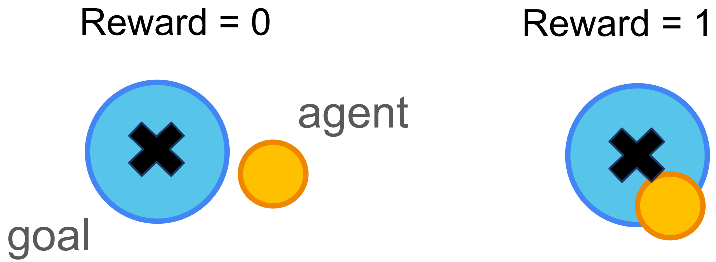


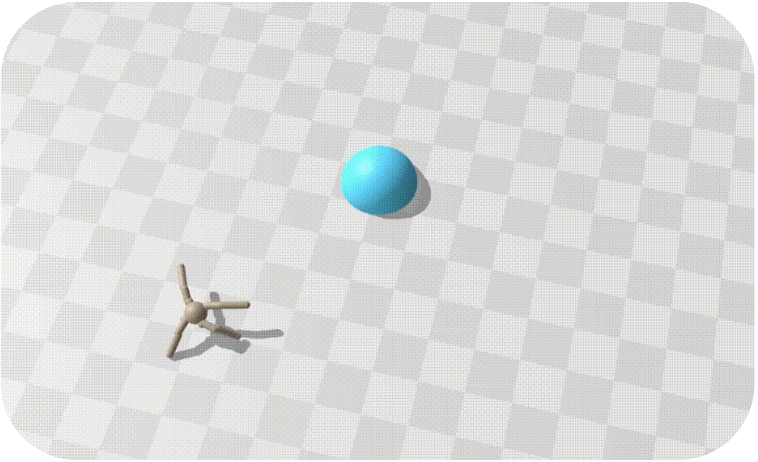
Goal-Conditioned Reinforcement Learning

- The objective is to reach the **goal** state.
- The **goal** can be defined as a subset of state space, i.e., just x and y coordinates.
- Often used in sparse reward settings.

The policy is conditioned on both state and goal:

$$\pi(a \mid s, g)$$

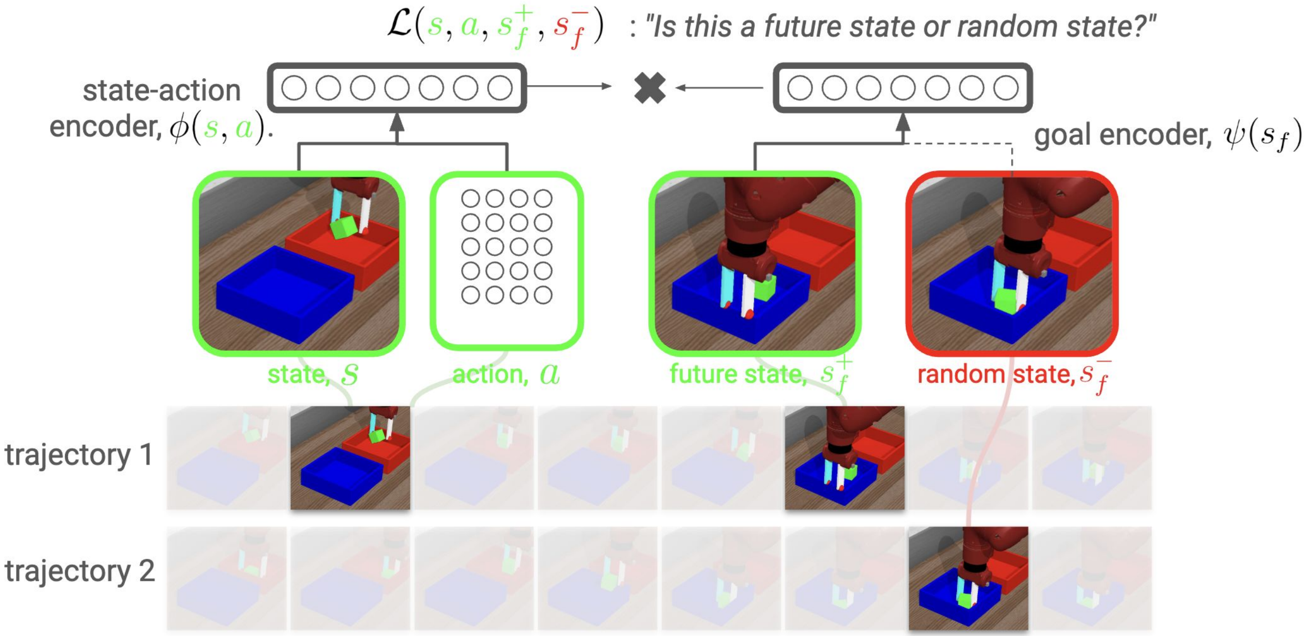




Contrastive Learning as Goal-Conditioned RL

(Eysenbach et al., NeurIPS 2022)

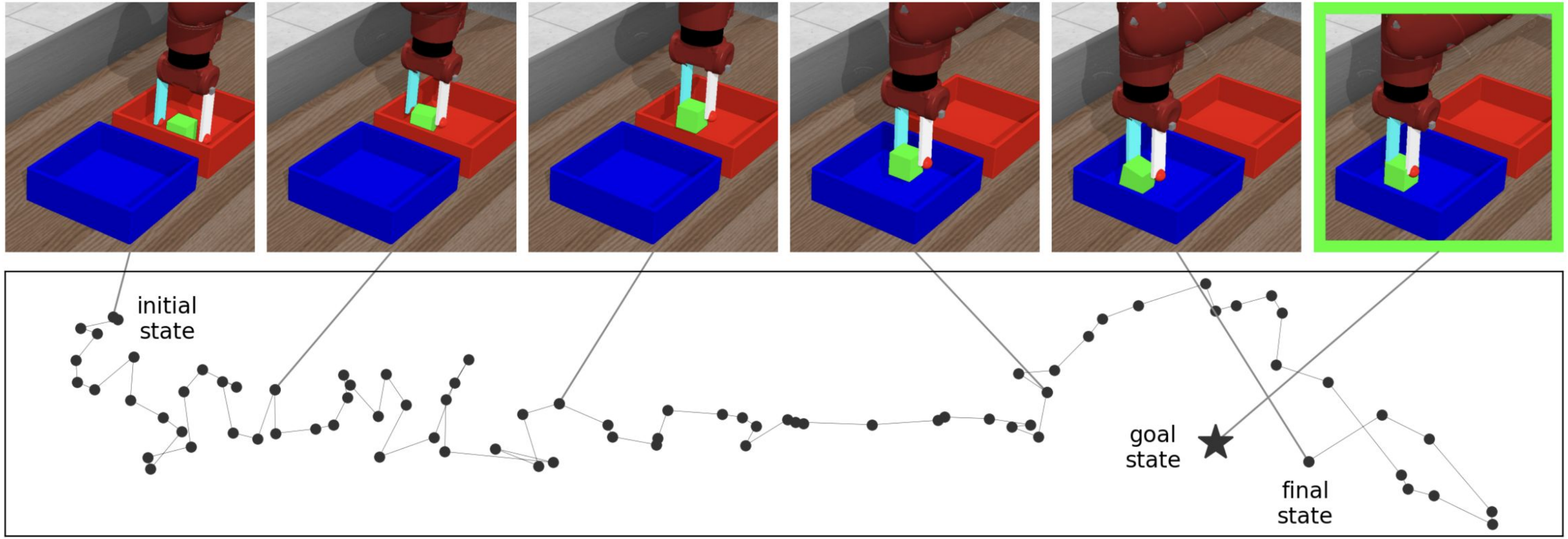
Objective: Discriminate future states from random states.



Source: Eysenbach et al., 2022

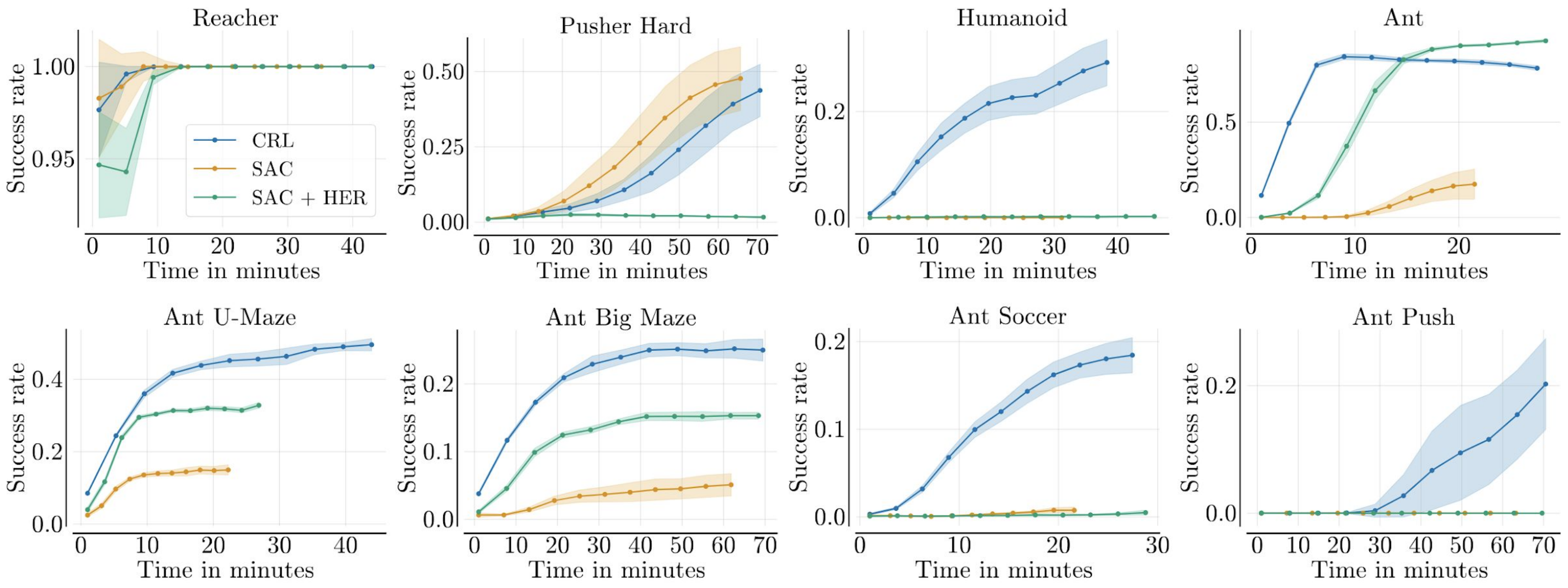
Why it works?

- Meaningful environment dynamics representations



Source: Eysenbach et al., 2022

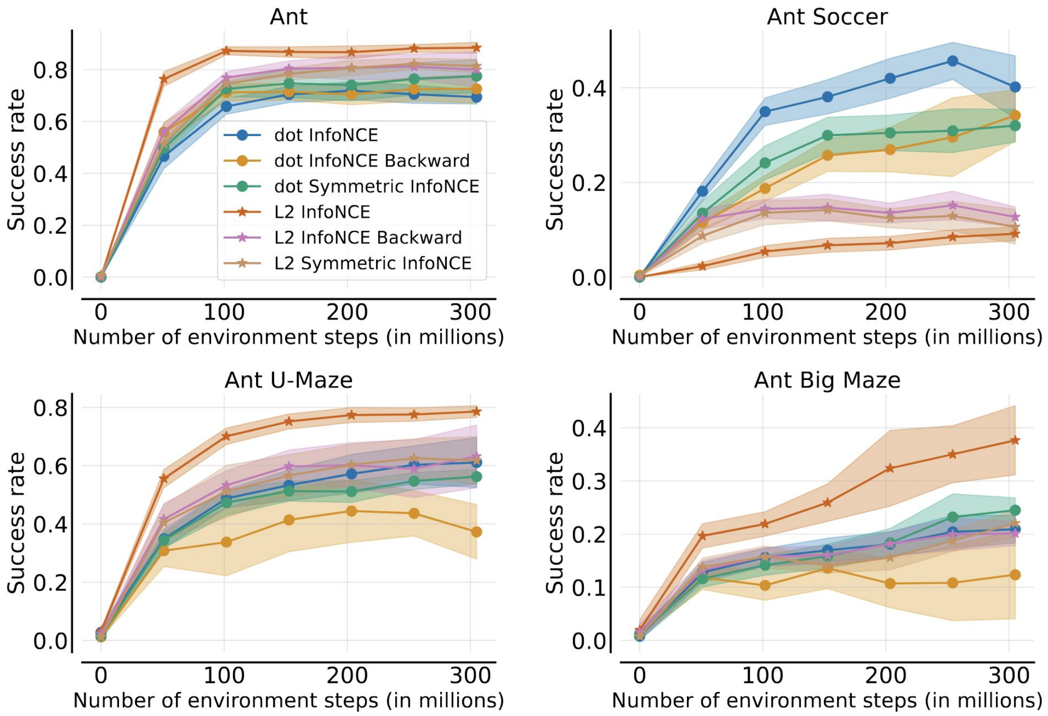
JaxGCRL Benchmark results



- Contrastive RL learns non-trivial policy in every environment.

What if we scale the number of env steps?

- Currently, CRL does not scale effectively with large amounts of data.



How important are energy and contrastive functions?

- Different contrastive energy functions and contrastive objectives based on InfoNCE (and DPO) perform on-pair.

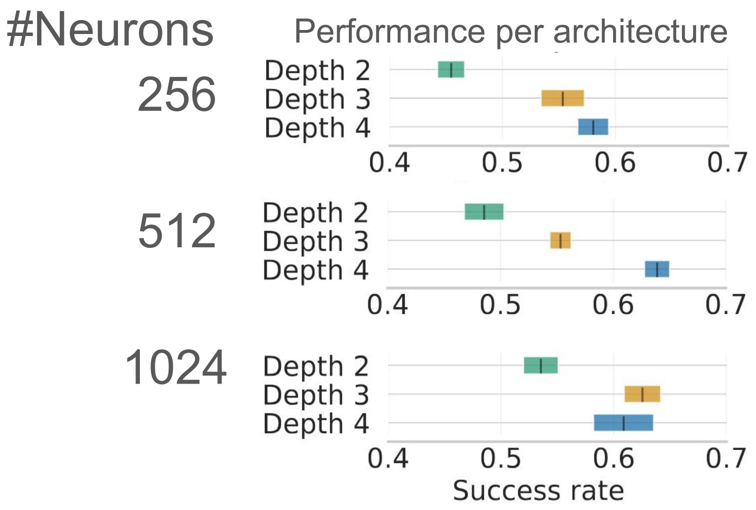
Where is the bottleneck, then?

Contrastive objectives

Energy functions

Is architecture scale a bottleneck?

- Architecture size helps but needs to be scaled up correctly.
- Layer Normalization is helping in large architectures.



Performance

Takeaways

- **Obstacles: Data and Algorithms/Architectures.**
- JaxGCRL addresses Obstacle 1 and speeds up research on Obstacle 2.
- Experiments with 10M steps can be completed in minutes, while those with billions of environment steps can be done in a few hours.
- *“We are experiencing another seismic shift in (RL) field”* – Jakob Foerster 2024

Thank you!

michalbortkiewicz8@gmail.com, wladek.palucki@gmail.com

github.com/MichalBortkiewicz/JaxGCRL