Enhancing Tactile-based Reinforcement Learning for Robotic Control

THE UNIVERSITY of EDINBURGH

Elle Miller, Trevor McInroe, David Abel, Oisin Mac Aodha, Sethu Vijayakumar

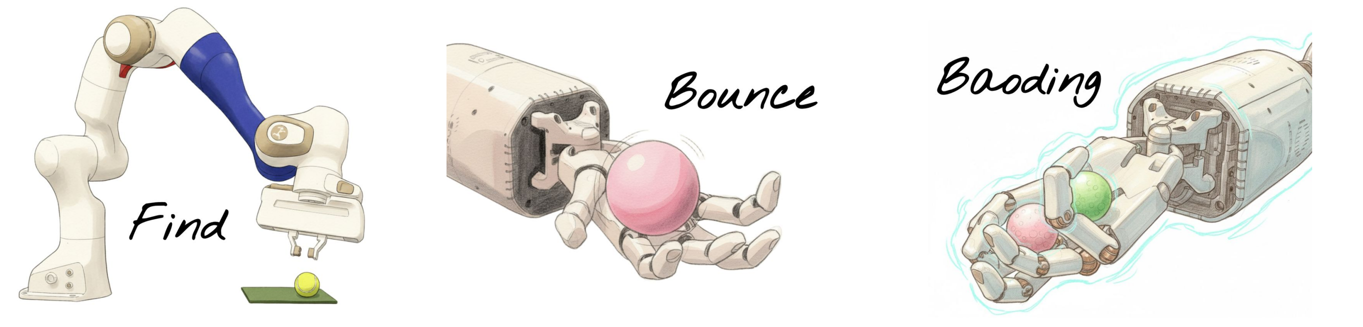
- ★ We propose **4x self-supervised learning objectives** to support the observation representation within the RL optimisation, finding multi-step latent forward dynamics to be most effective for our tactile-based agents
- ★ We demonstrate **blind superhuman dexterity** in simulation with only **proprioception + 17 binary contacts**
- ★ We show that **tactile sensing provides distinct benefits** over proprioceptive histories
- ★ We release new environments in **roto**, a benchmark to standardise and promote research in tactile RL

Open problem: how to effectively integrate tactile sensing into RL

- Robots must be able to **feel** the world to be truly useful and safely interact with humans
- The field of tactile learning faces **many challenges**:
  - → tactile data is sparse, discontinuous, and complex to interpret
  - → integrating tactile data with RL yields conflicting results
  - → the need for tactile sensing in robotics is hotly debated
- First, we designed some new tasks that covered a **wide range of tactile interactions** (sparse, intermittent, sustained) and released them as:

Robot Tactile Olympiad

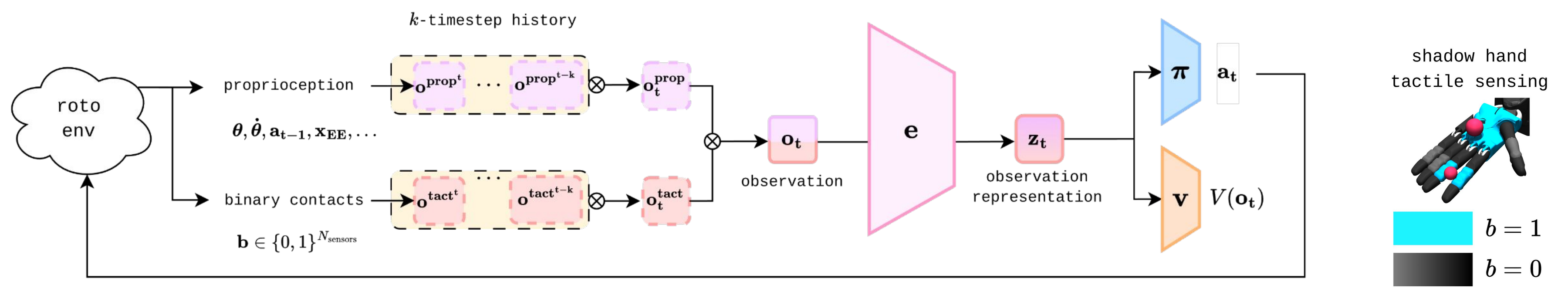
roto



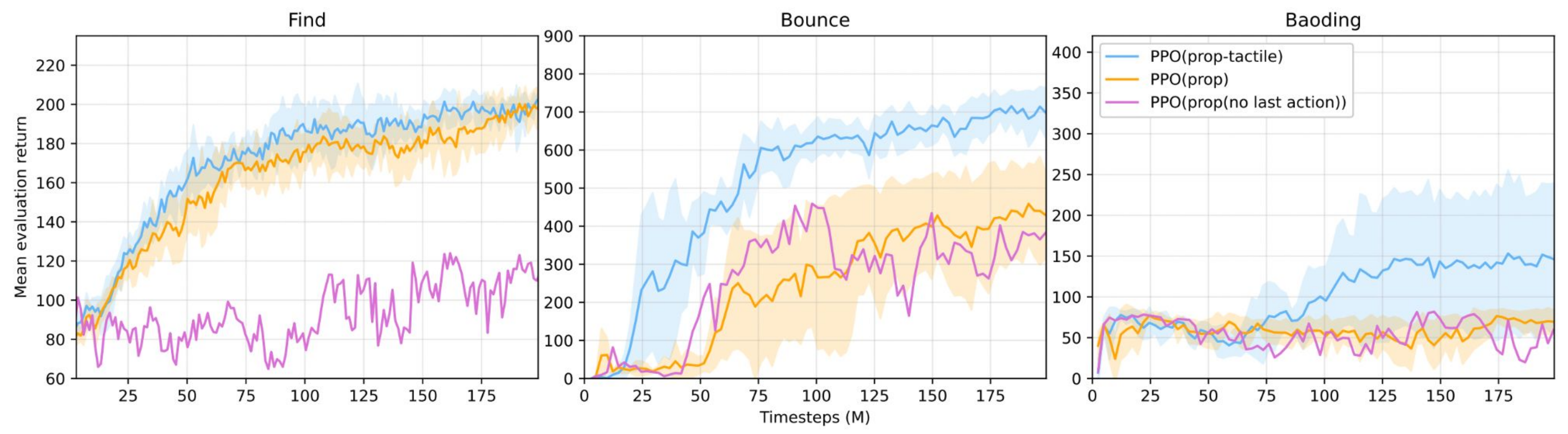
Locate a ball on a 20 x 20 cm plate — *Bounce a ball as fast as possible* — *Rotate the Baoding balls as fast as possible*

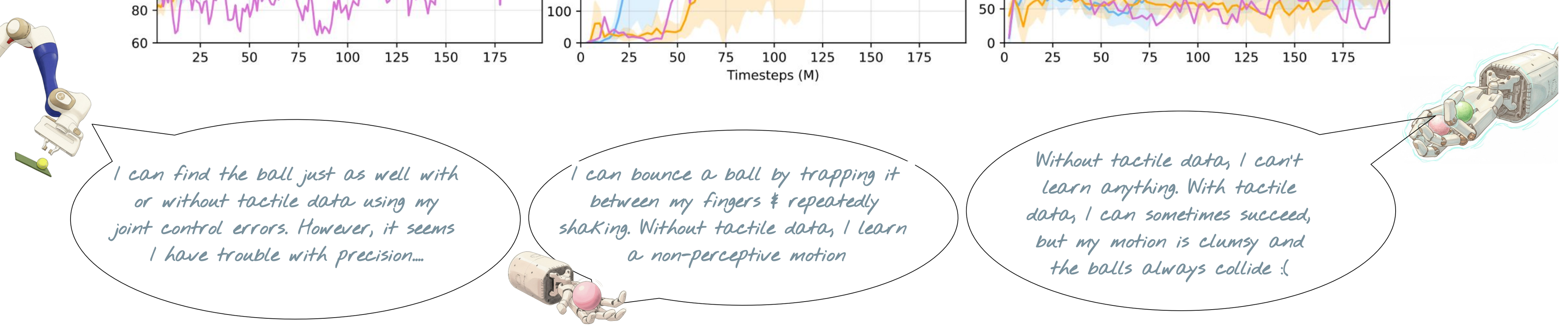
Initial investigation: what can blind tactile agents learn with RL?

- We chose to study a setup with **no vision**. Instead, the agent can only **observe a history of proprioception and binary contacts:**



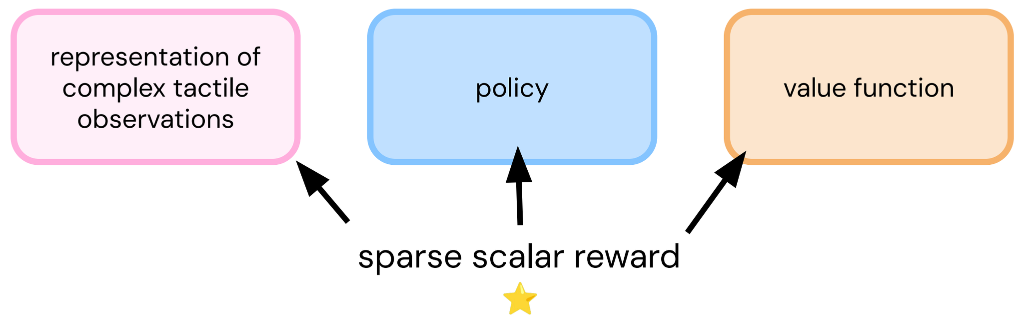
- Under this setup, we trained PPO agents **with** & **without** tactile sensing to evaluate its contribution. The dependence on tactile data varied across tasks, but overall the agents did not exhibit sophisticated control capabilites:



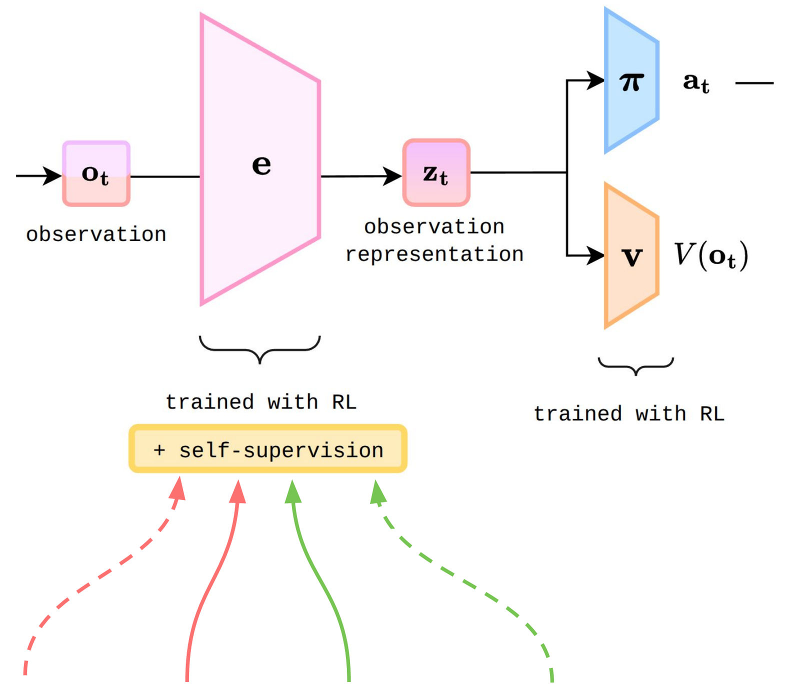


Idea: can self-supervision can help tactile agents?

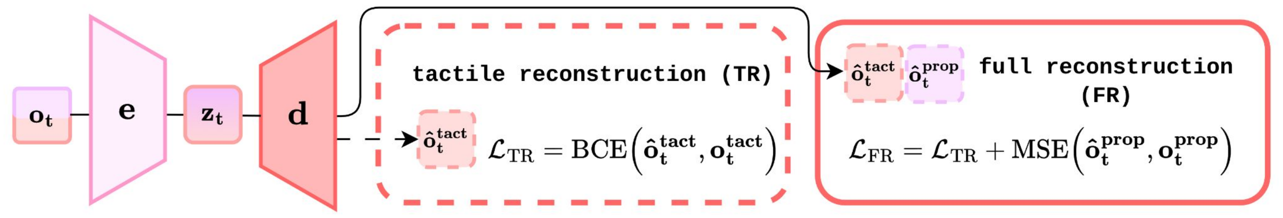
- We wondered: are tactile-based agents are struggling to simultaneously learn the observation encoding, policy, and value function from a single, scalar reward?
- **Chicken and egg problem:**
  - → agents need a good observation representation to learn a good policy,
  - → agents need a good policy to visit useful states to learn a good observation representation



- What if we jointly train the observation encoder with **self-supervised learning** (SSL)?
- This would help the agent transform complex sensory input into representations useful for control
- We propose and analyse **4x SSL objectives**

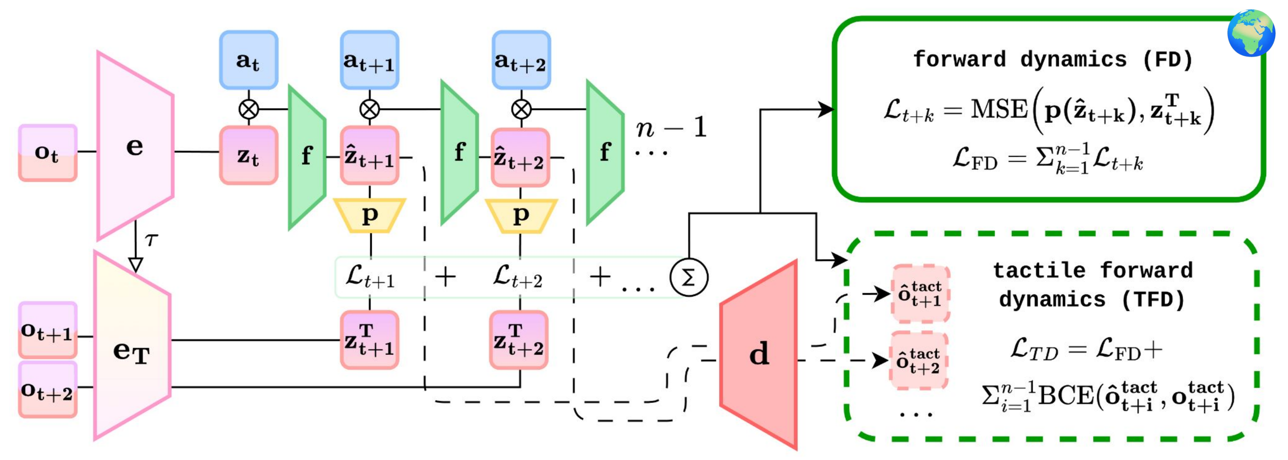


"Extract the right information from the observation such that the encoder can..."



"reconstruct the tactile observation"

"reconstruct the full observation, like an autoencoder"

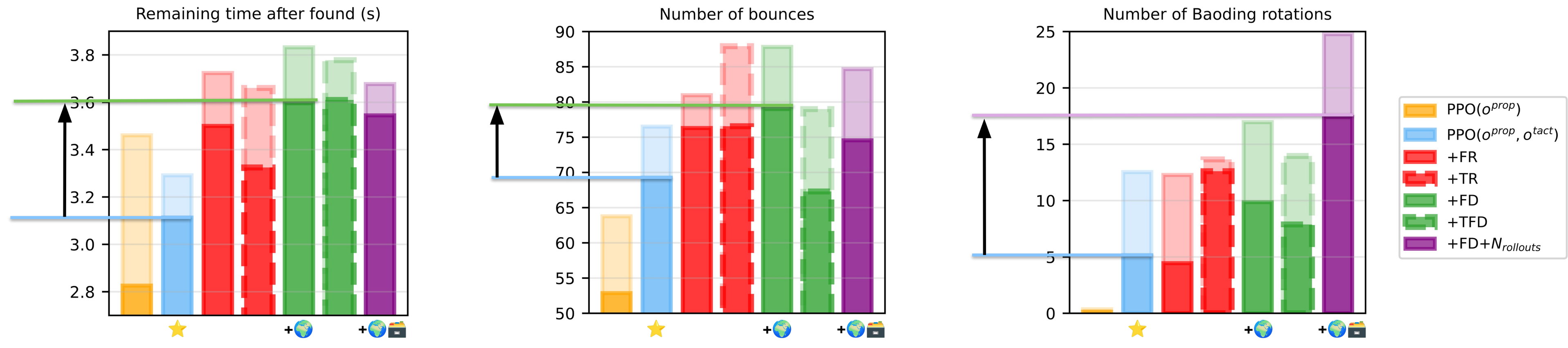


"given an initial observation and some actions, predict the subsequent latent states"

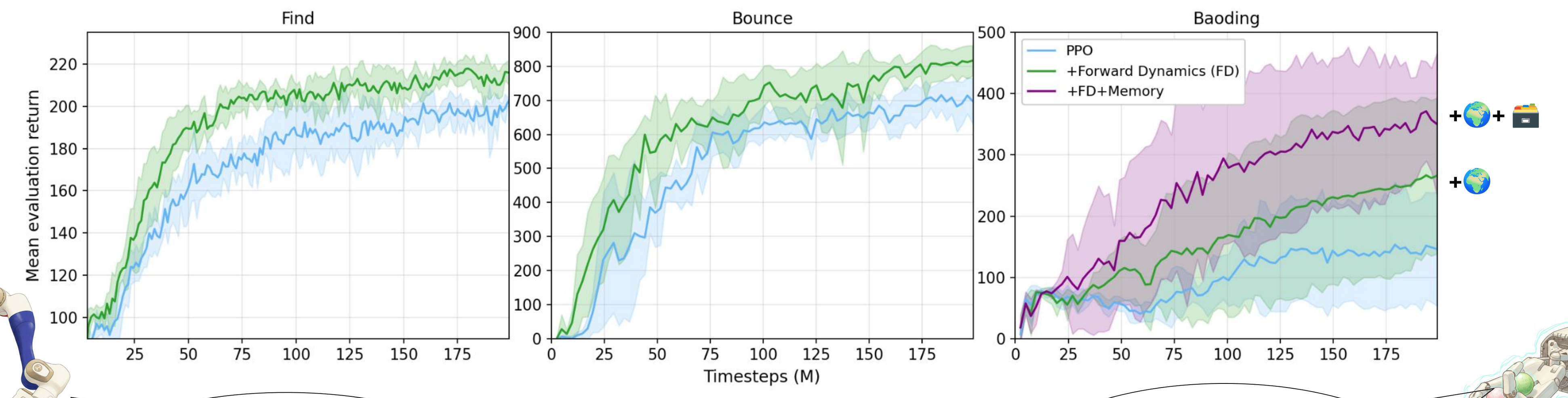
"do the above WHILE also decoding the future tactile observations from the latent state"

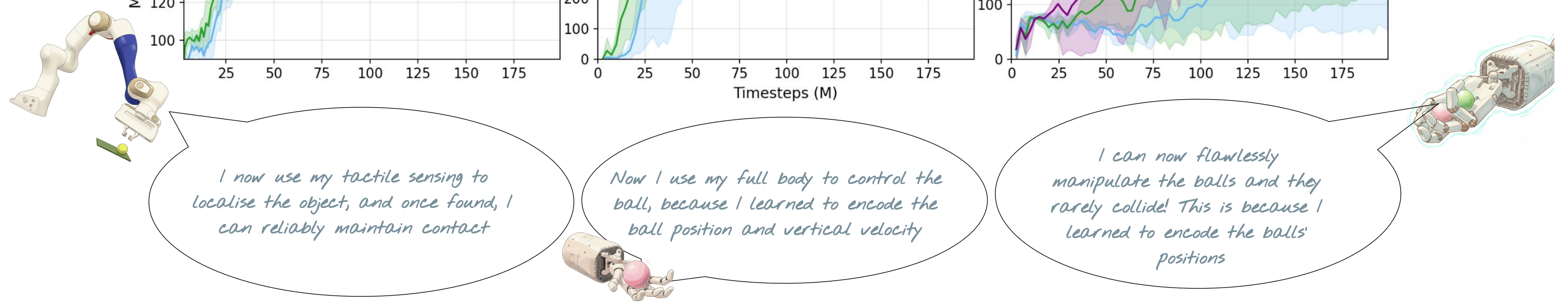
Result: self-supervision enables super-human tactile agents

- Best overall: **forward dynamics** (FD) objective (aka learn a world model)
- We also experiment with increasing the SSL memory → sometimes agents **can benefit from being trained on larger memories**
- Compared to **RL-only** agents (trained only with reward), the best SSL agents:



Find an object 36% faster (0.5s) — Bounce a ball 10 more times in 10s (79 vs 69) — Do 12 more Baoding rotations (17 vs 5) in 10s

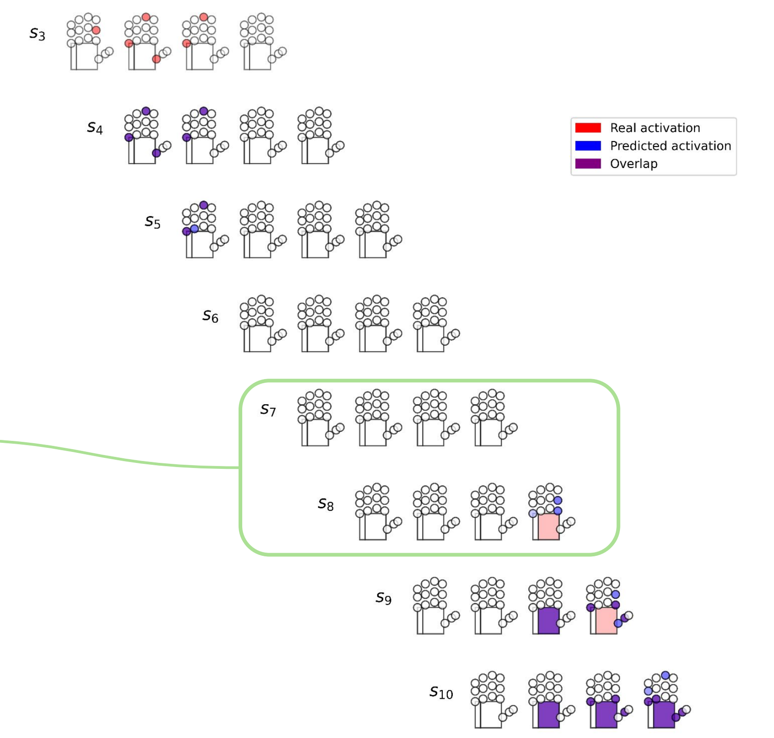




Why does it work?

- We estimated the **mutual information** between the learned observation representation and the underlying state variables → training with **forward dynamics** uncovered **the most state information**, e.g. ball positions and velocities

We also found that dynamics-based agents could predict future tactile states from a state of no contact, e.g. when and where a ball would land!



We hope our work is a small step towards robots that can not only see, but **feel** the world around them.