Reinforcement Learning for Robot Control

Jens Kober

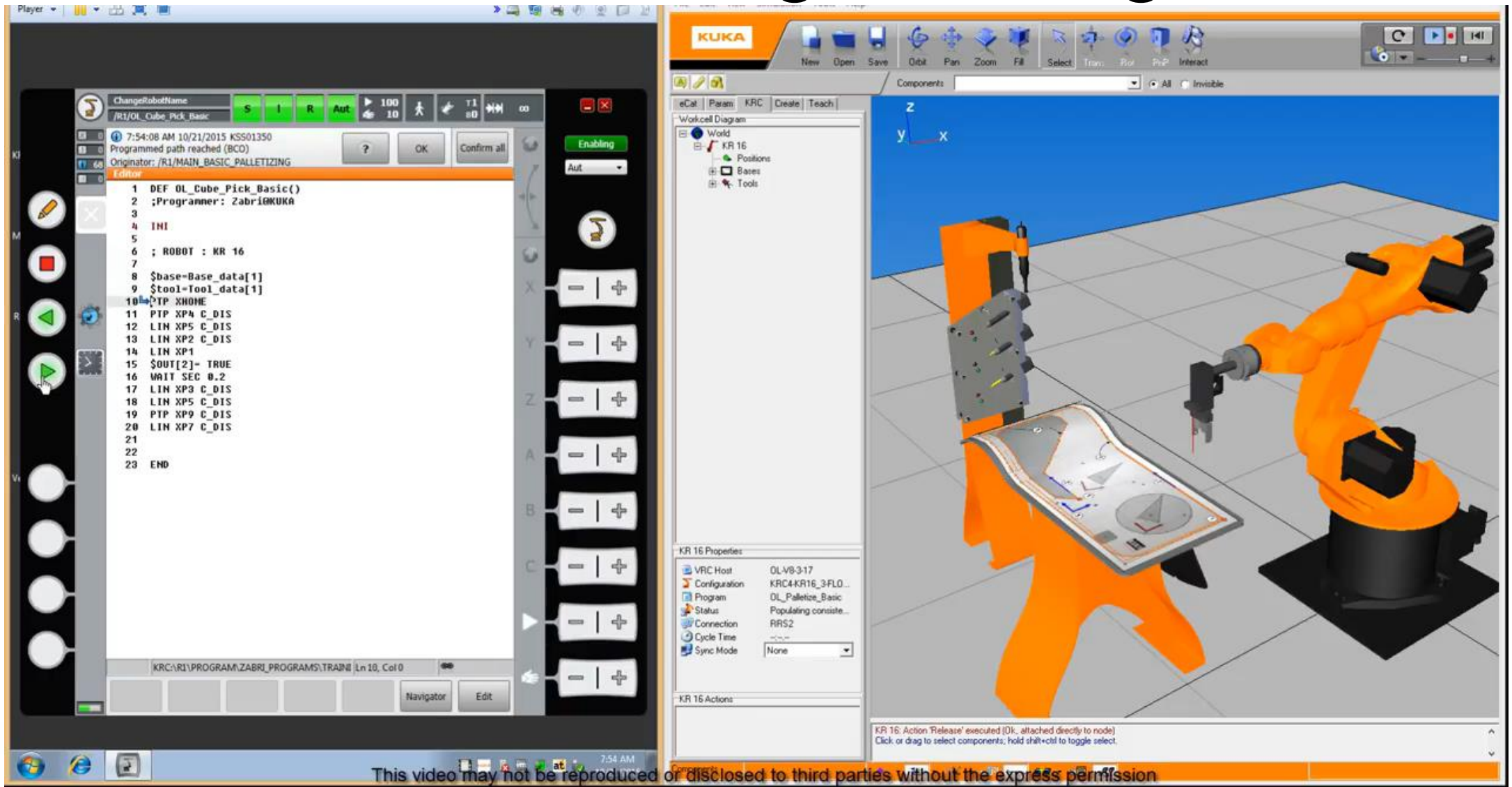
March 24, 2025

Workshop on Themes across Control and Reinforcement Learning

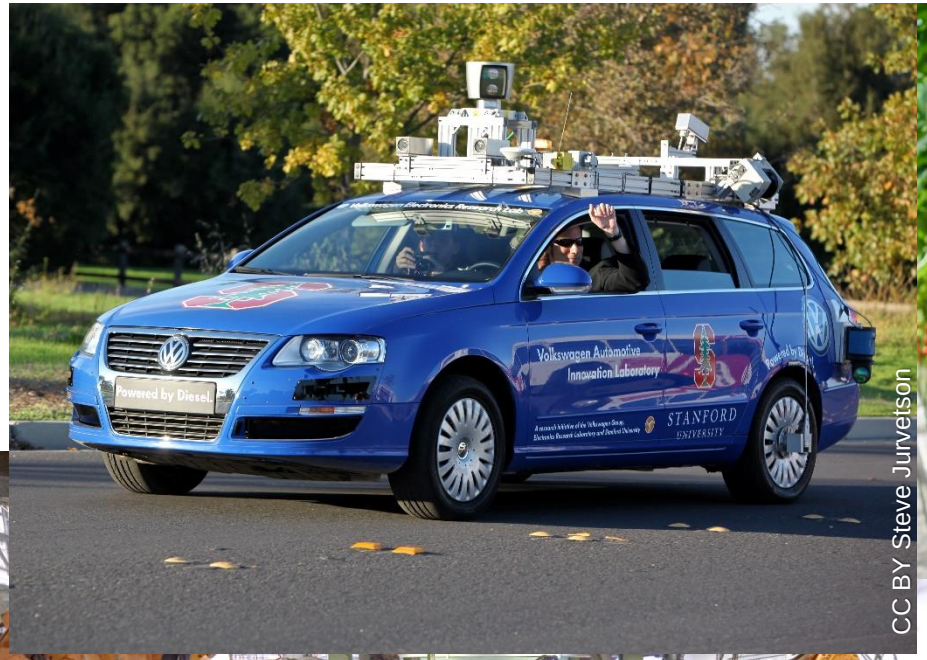
Robot Learning

Zabri Zakaria

Traditional Robot Programming



What's the problem?



CC BY Steve Jurvetson

CC BY-SA Mixabest

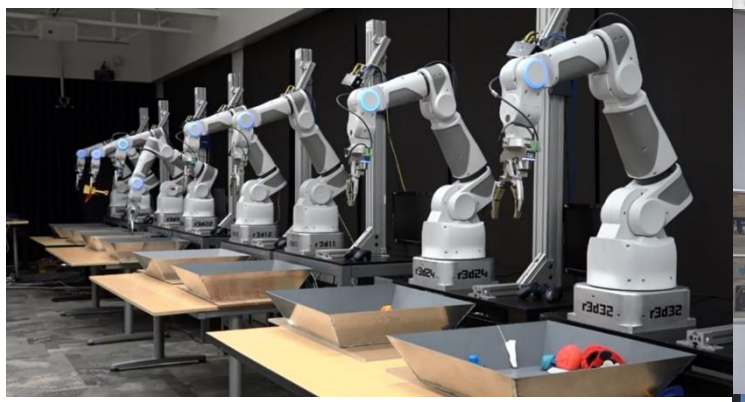
CC BY-NC-SA Brandon Heyer

What's the problem?

- Uncertainties and variations
  - Objects
  - Environment
  - Tasks
  - Human behavior
- Occurring in all robot domains!

(Reinforcement) Learning in Robotics

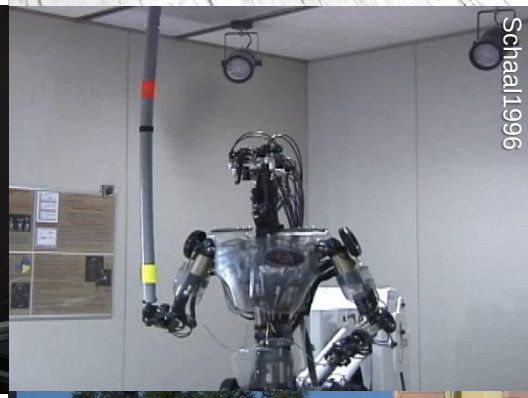
- Safe with real robots



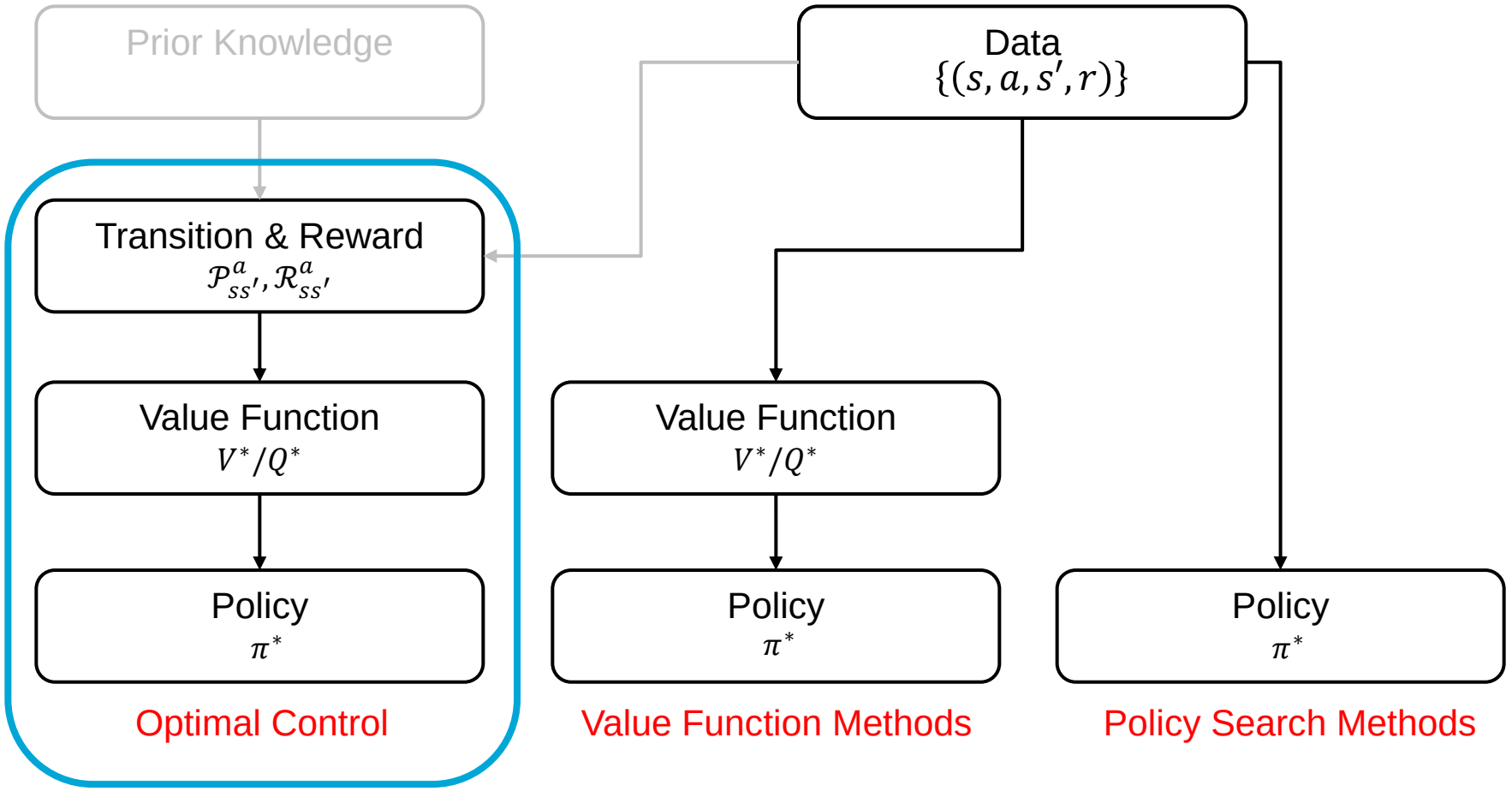
Google

- Fast learning
  - Sample efficient – “small” data
  - Incorporate prior knowledge
  - Few open parameters

- Real time computations

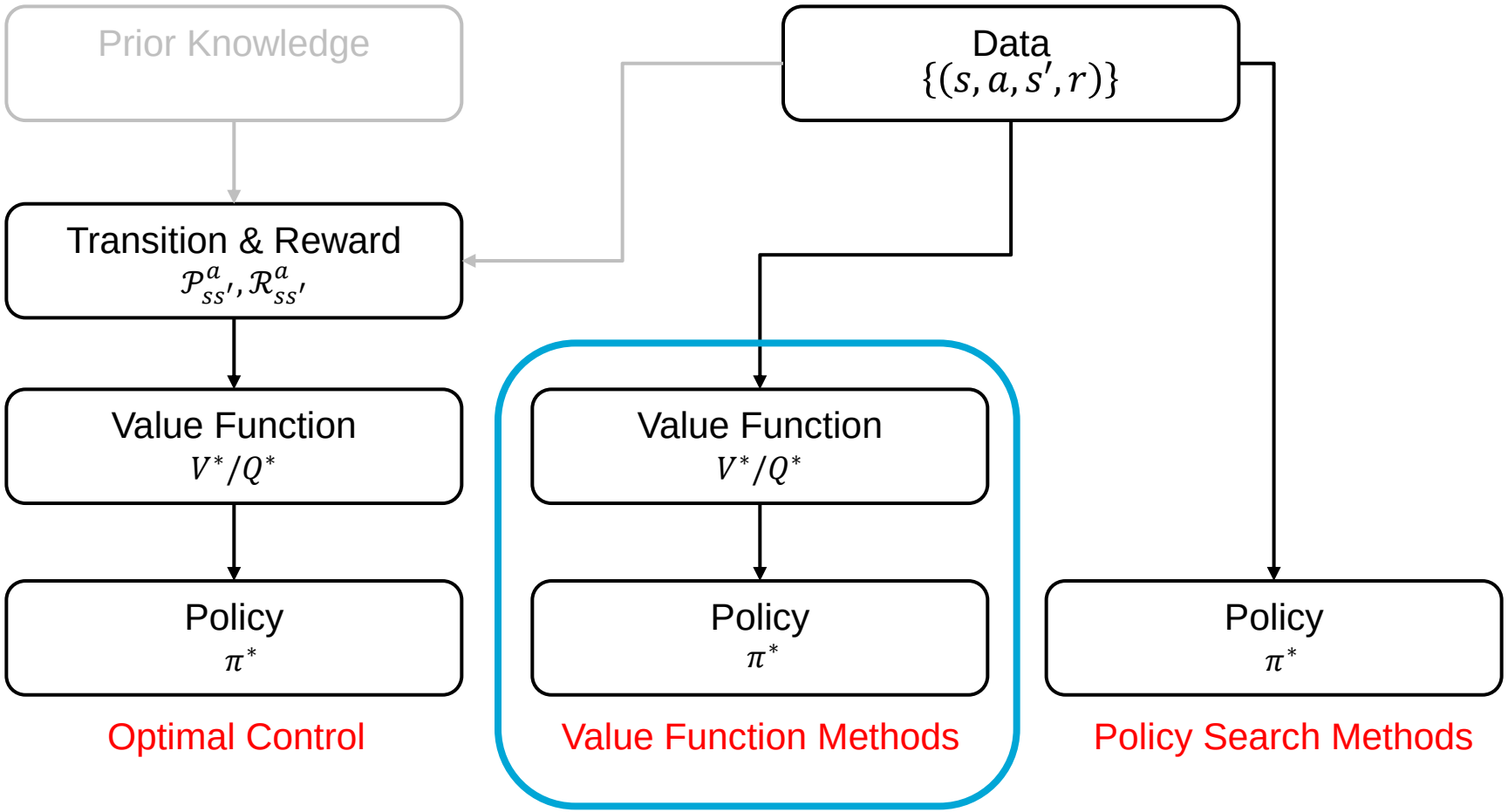


Types of RL Algorithms



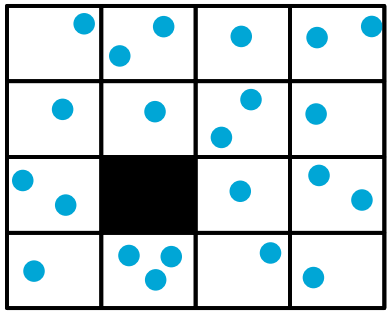
https://youtu.be/VCdxqn0fcnE

Types of RL Algorithms

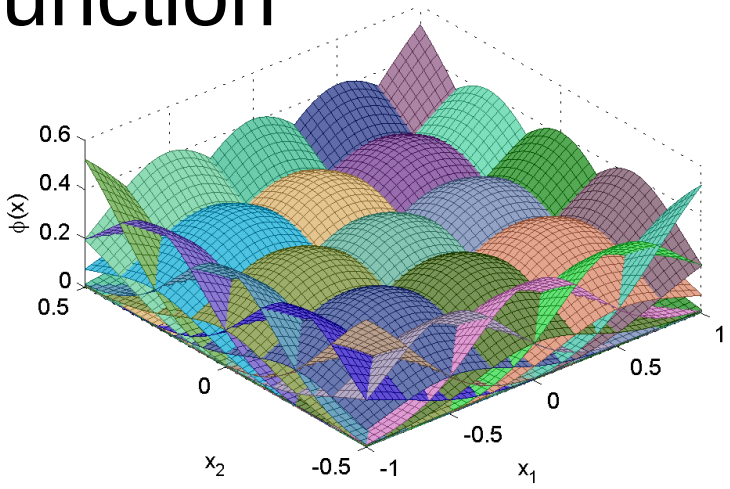


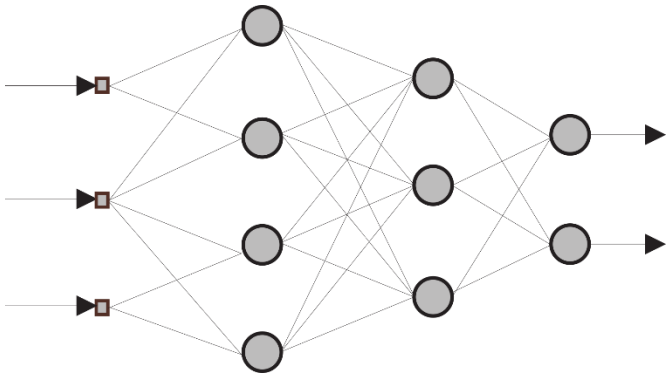
Dealing with Continuous States/Actions

- Discretization



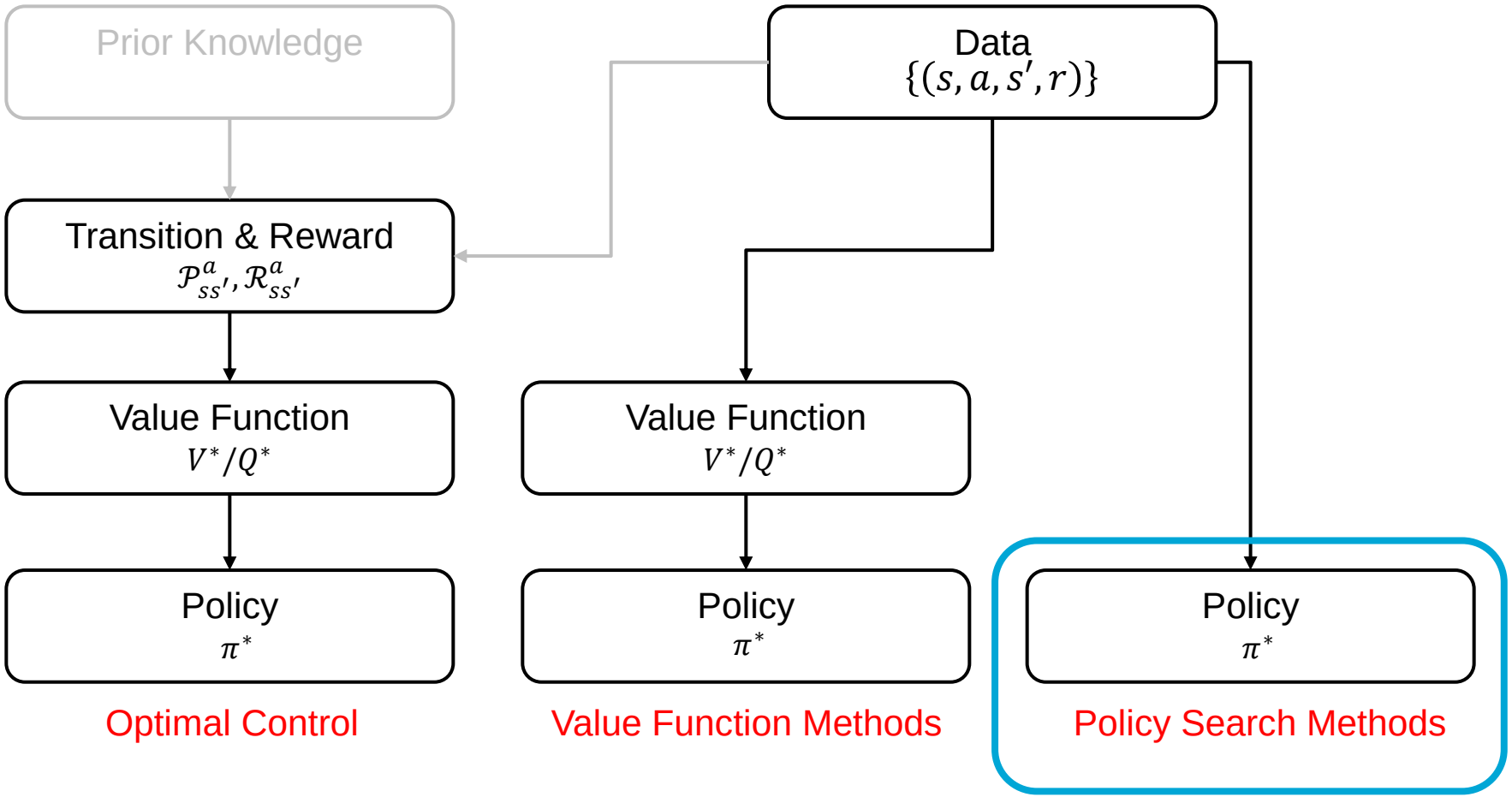
- Function approximation
  - Value function
  - Policy



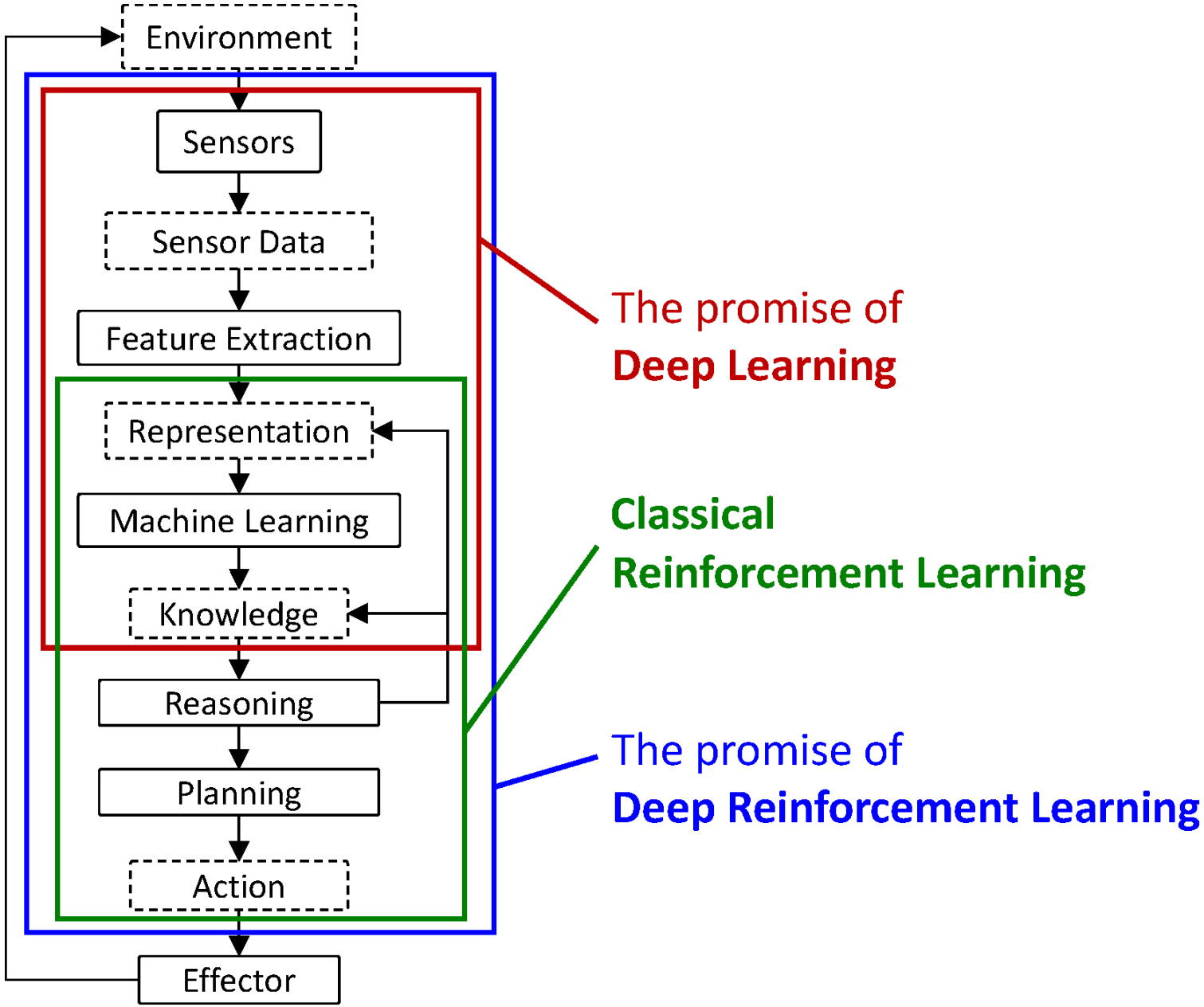



successful Neuro Dribbling at
RoboCup 2007 Atlanta, USA

Types of RL Algorithms

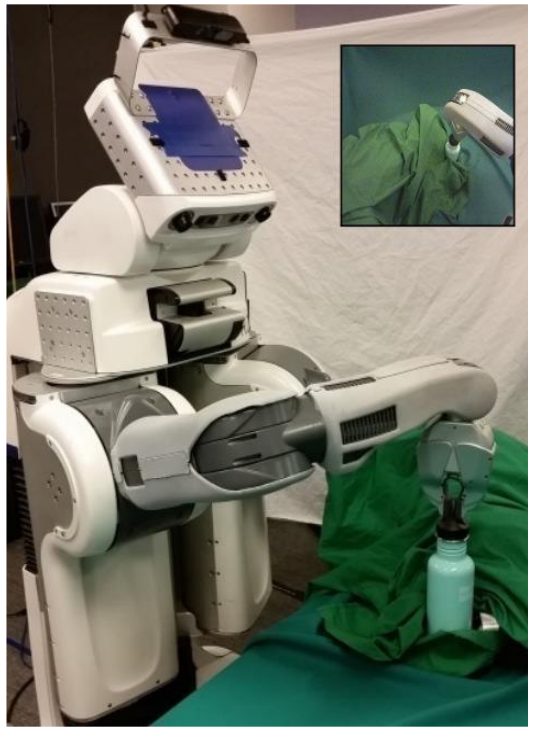


Kober & Peters, 2008 https://youtu.be/cNyoMVZQdYM

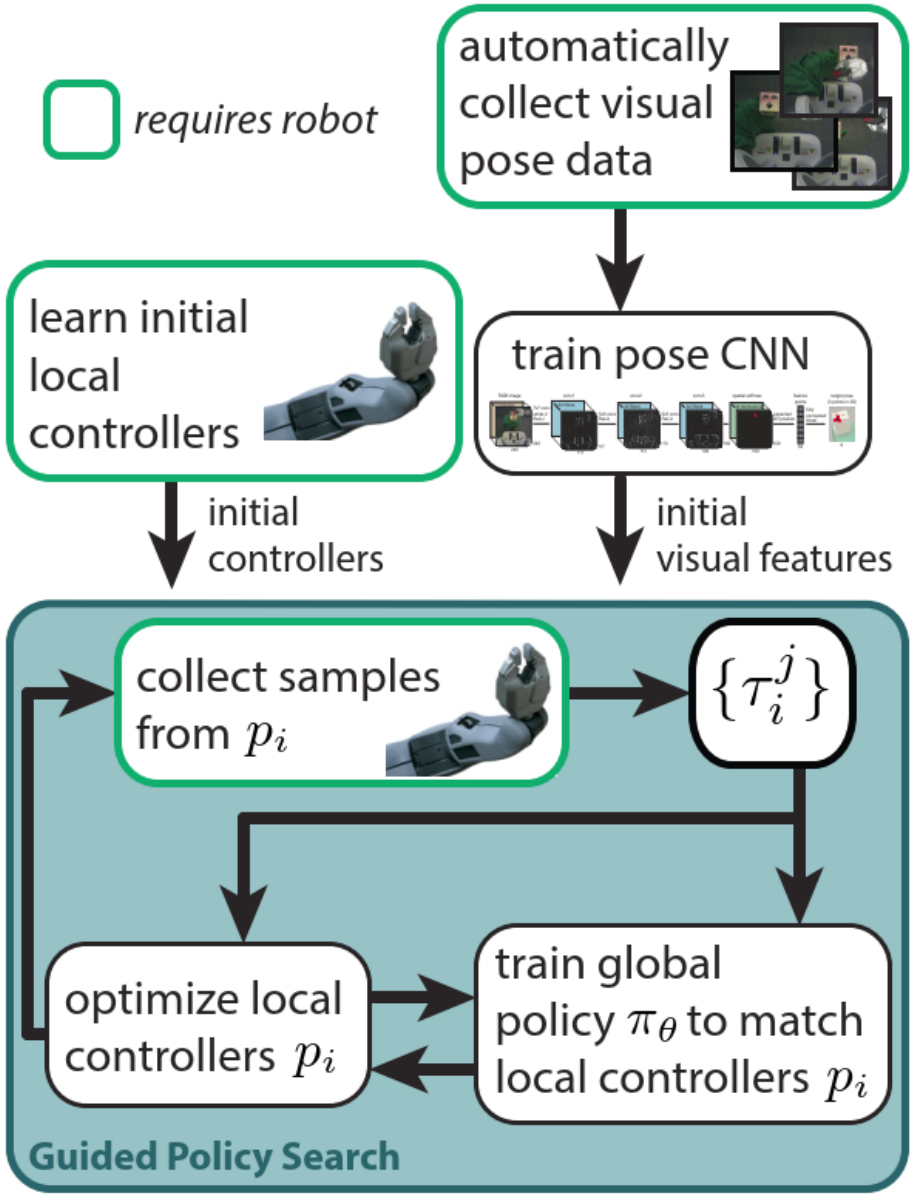

Environment
Sensors
Sensor Data
Feature Extraction
Representation
Machine Learning
Knowledge
Reasoning
Planning
Action
Effector
The promise of
Deep Learning
Classical
Reinforcement Learning
The promise of
Deep Reinforcement Learning

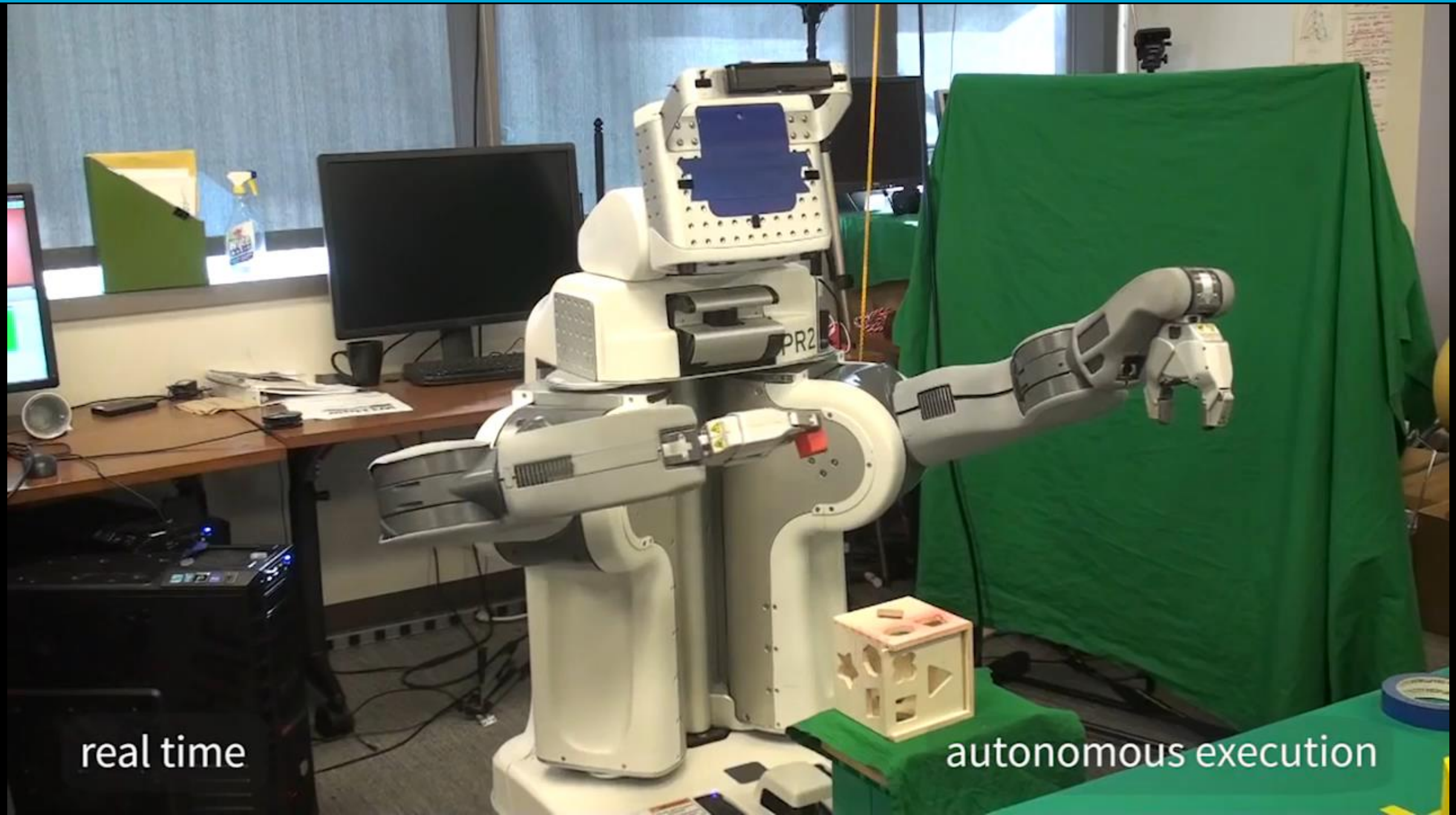
Guided Policy Search

- Real world
- Instrumented training



Levine et al. 2016




real time
autonomous execution

Sim2Real

Sim2Real

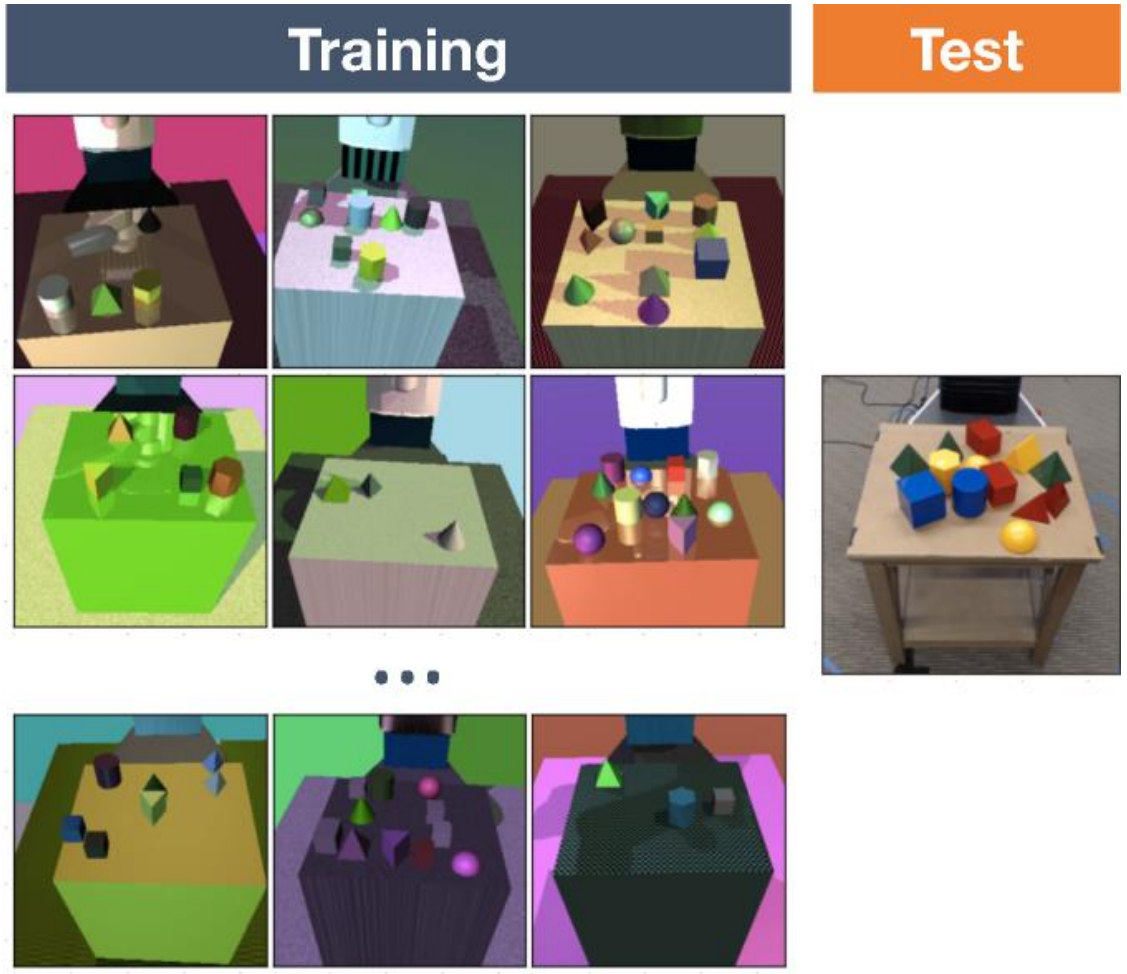
- Learn in Simulation, Transfer to Real-World

- Advantages:
  - Faster than real-time training
  - Simulated learning is inherently safe
  - Simulators can reset to any arbitrary state
  - Evaluating different task specifications is easier in simulation

Tobin et al. (2017)

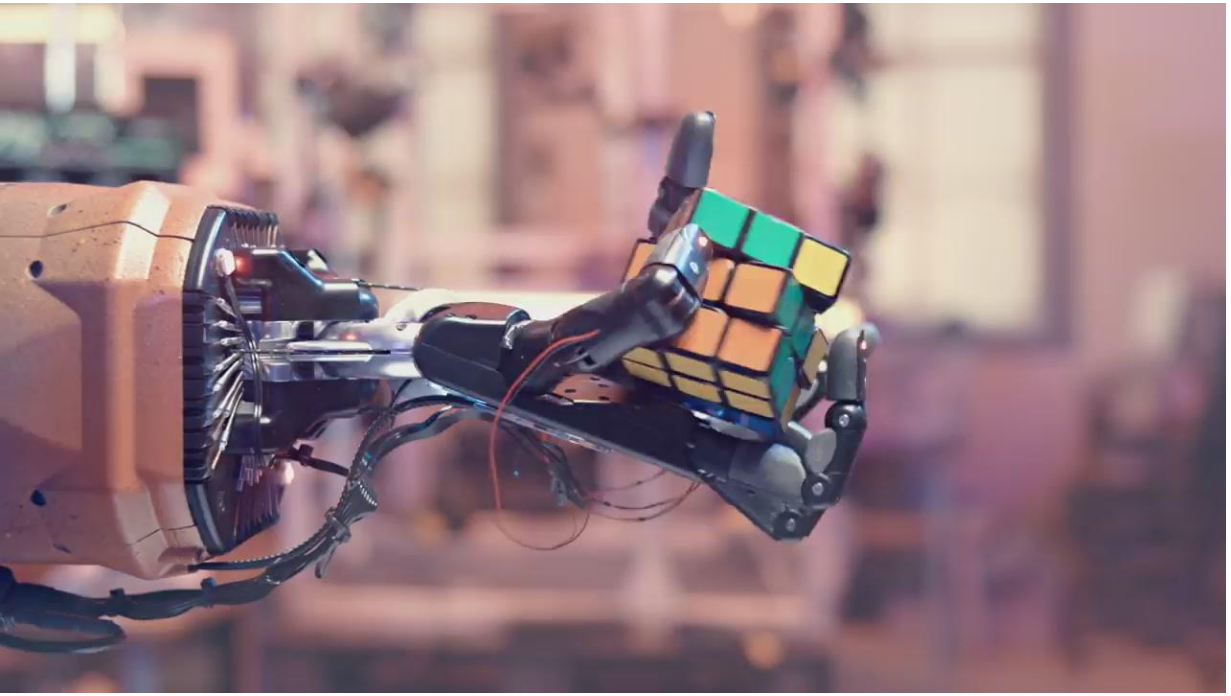
Training on Simulated Data

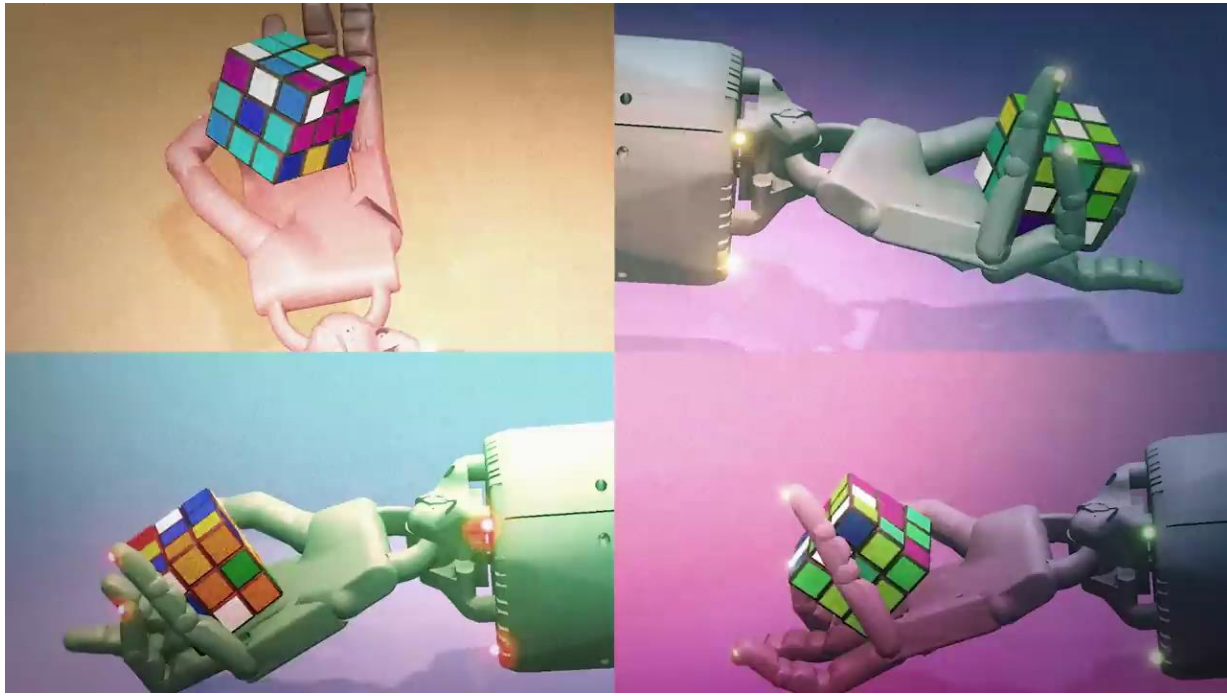
- Pre-train and finetune
- Direct Transfer
  - Make simulation very realistic
  - Make behavior robust



Solving a Rubik's Cube

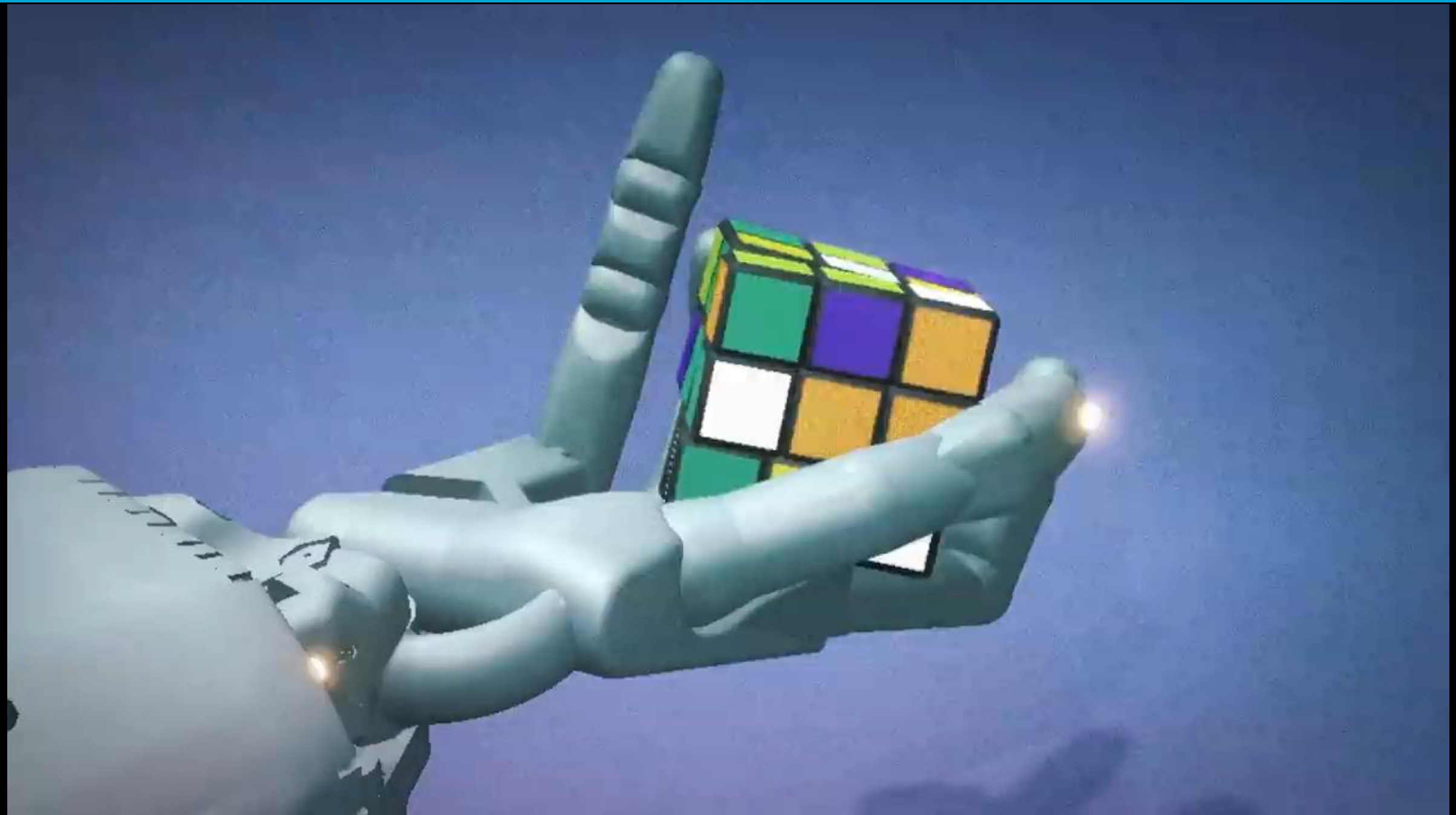
- Deep RL in simulation



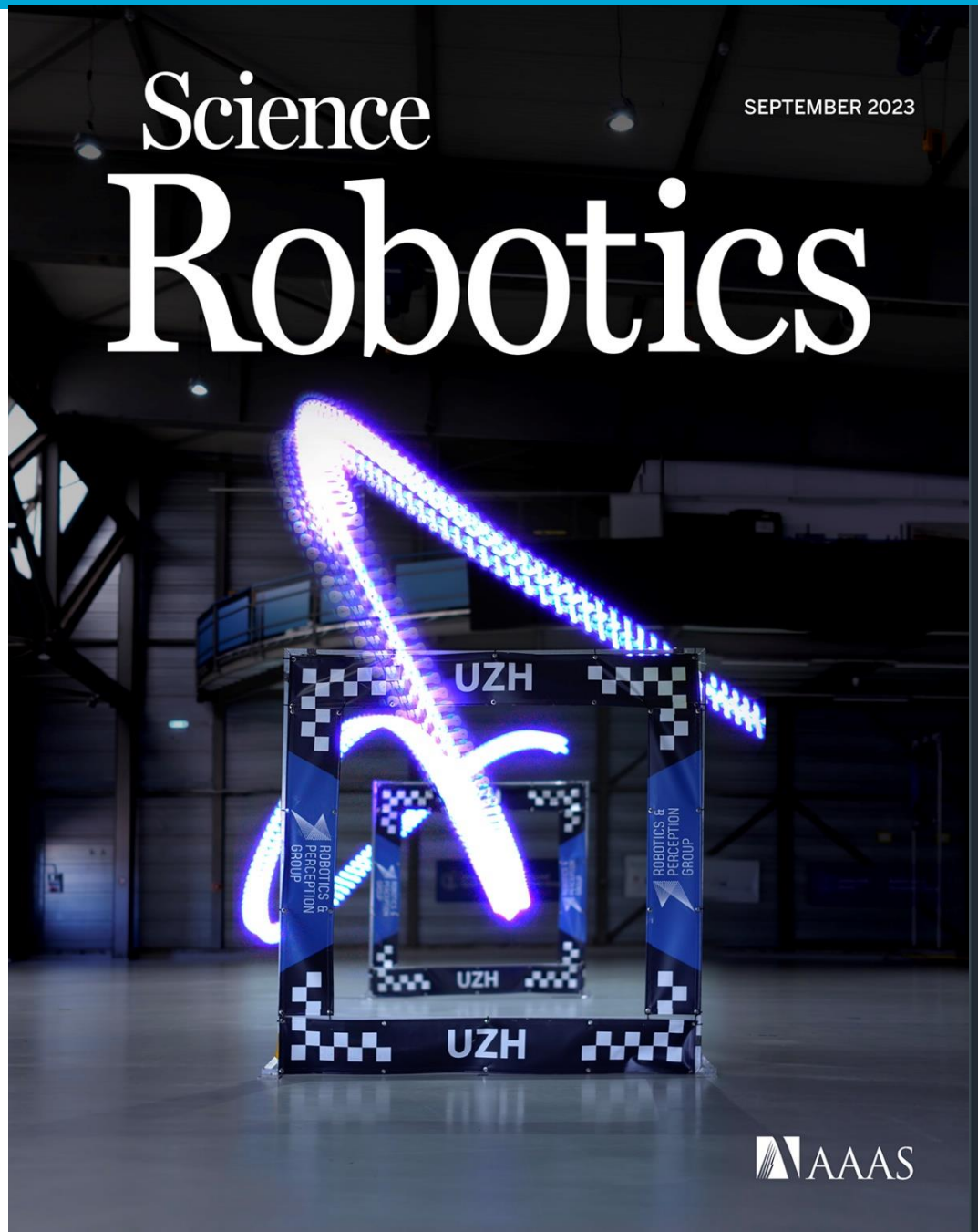


Akkaya et al. 2019

https://youtu.be/x4O8pojMF0w

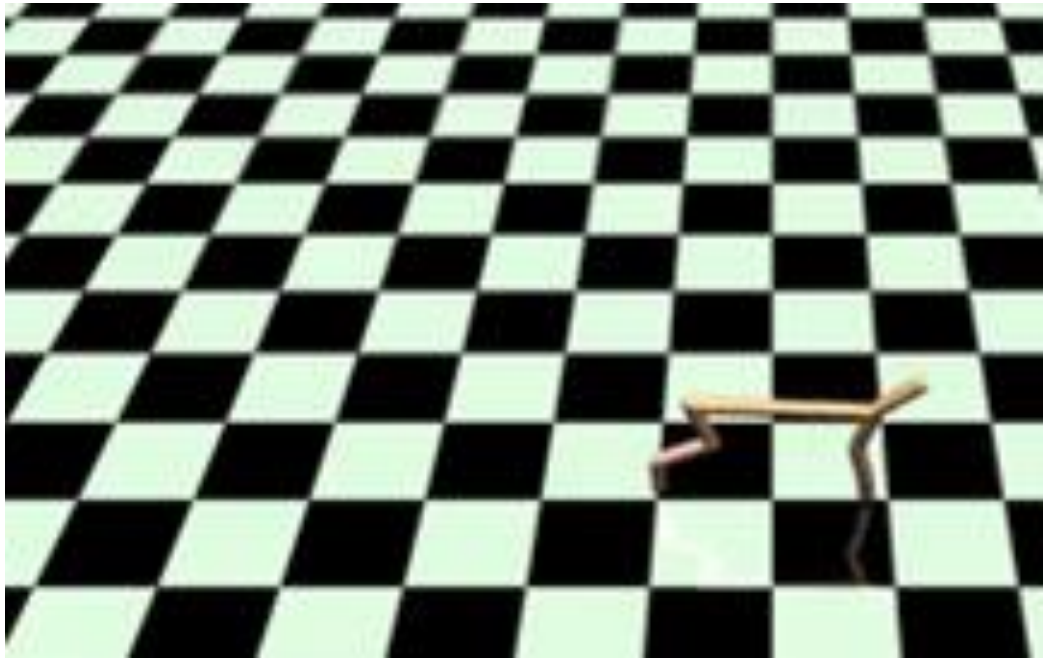


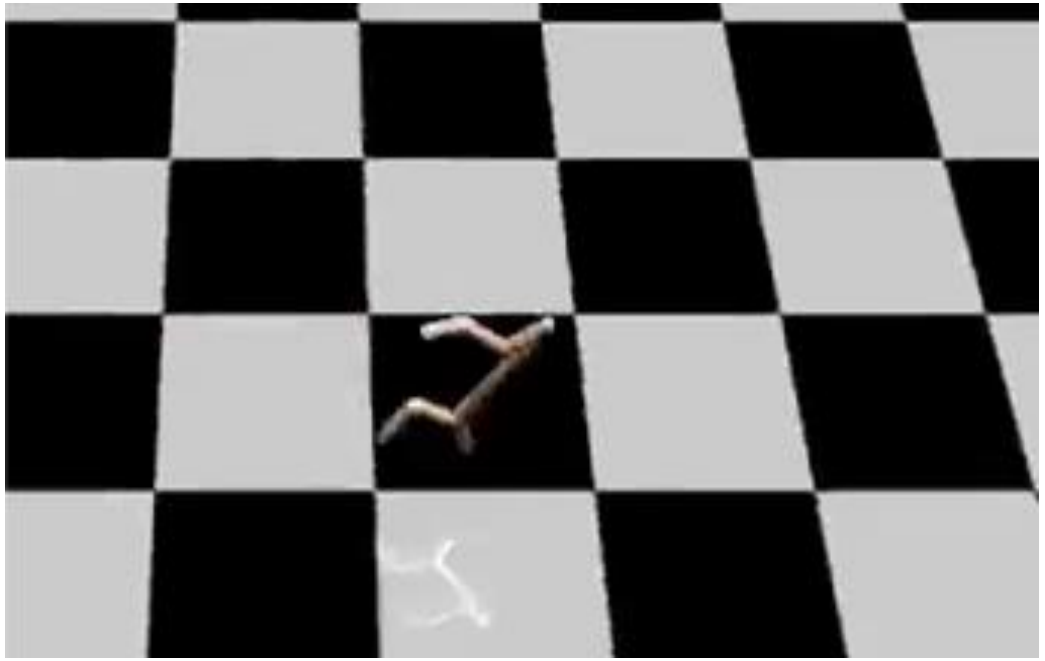

The international journal of science / 10 February 2022
nature
RACE TOGETHER 2021
DRIVING FORCE
AI algorithm outcompetes human champions in *Gran Turismo* racing game


Science
Robotics
SEPTEMBER 2023
UZH
UZH
UZH
ROBOTICS & PERCEPTION GROUP
ROBOTICS & PERCEPTION GROUP
AAAS


The international journal of science / 31 August 2023
nature
DRONE RACER
AI pilot beats human champions in aerial contest
Offset agreement
Overhaul pricing of carbon credits to help fund climate projects
Dining companions
Corals devour algal partners when food supplies run low
Sentence synthesis
Brain implants show promise in rendering speech from thoughts

Will it transfer?





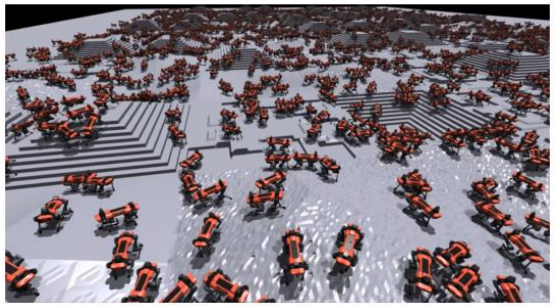
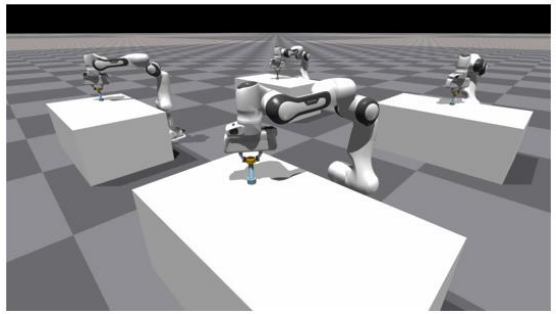
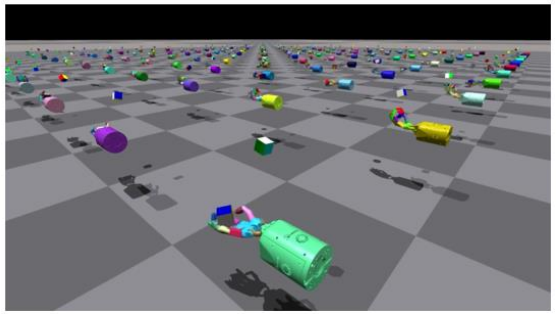
Credits: Nathan Lambert (@natolambert)

Guidelines for Sim2Real

- Minimize model mismatch
- Acknowledge all models are wrong
- Mitigate relevant mismatches

(None-Open) Tools

- Isaac Lab https://developer.nvidia.com/isaac/lab

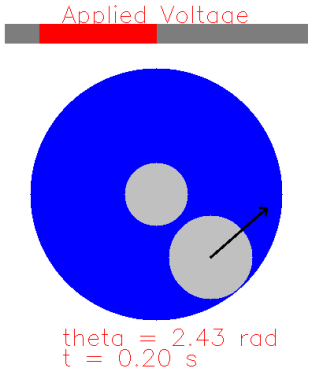
- Unity https://unity.com/solutions/automotive-transportation-manufacturing/robotics

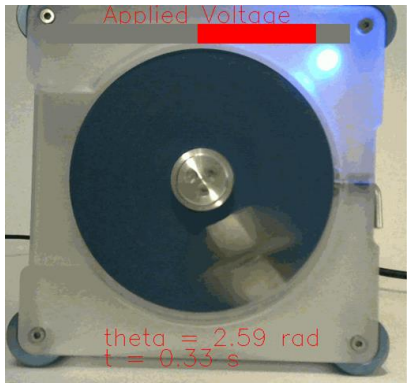
- Many more

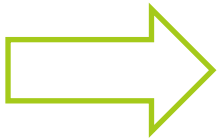
Common Pitfalls

Inaccuracies in the Simulated Dynamics

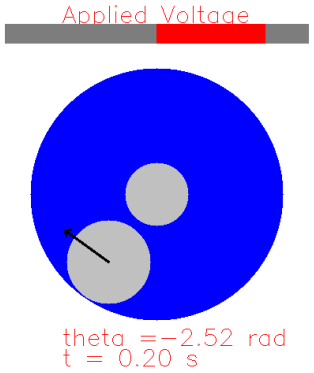
Inaccurate Model

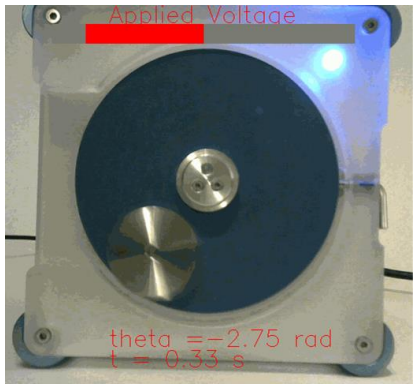


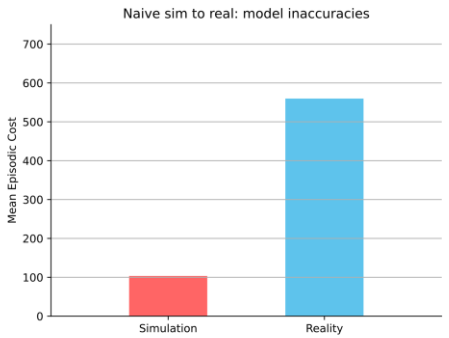


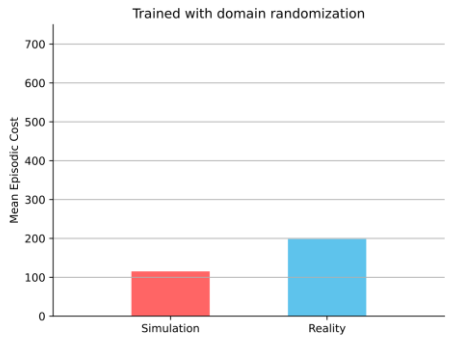


Domain Randomization





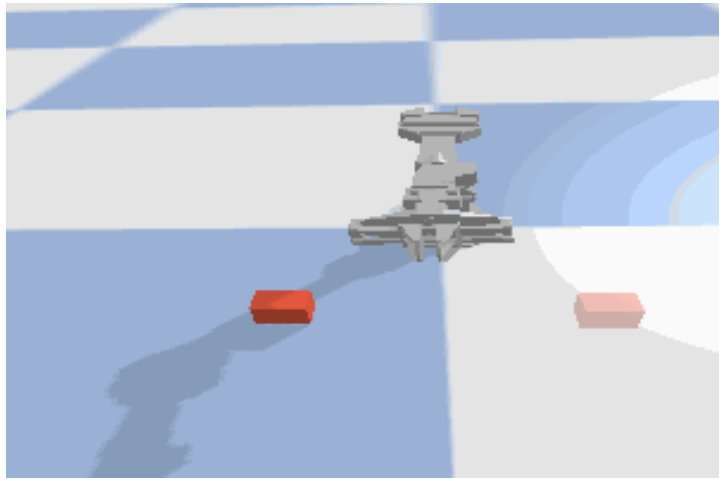


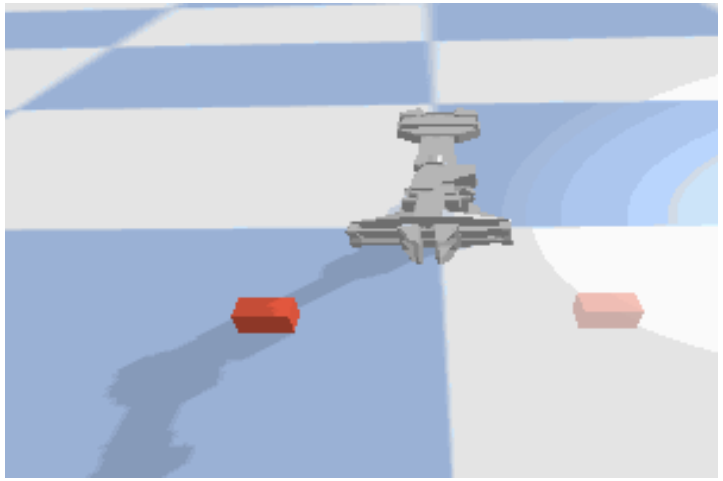


Unknowingly Changing the Markov-Decision Process

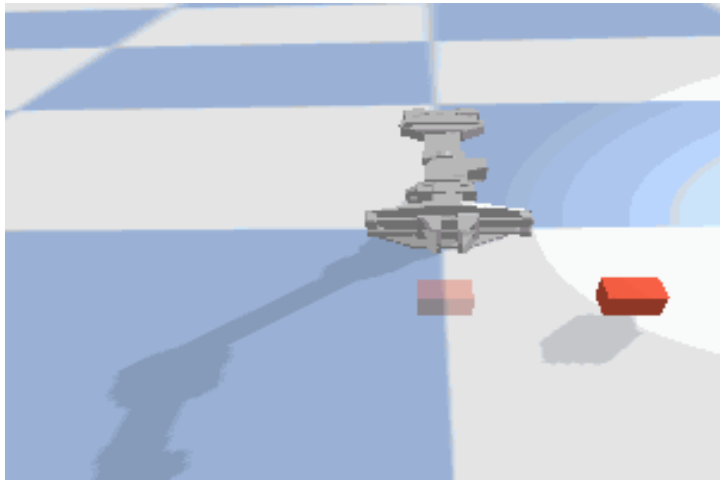
Training **Evaluation**

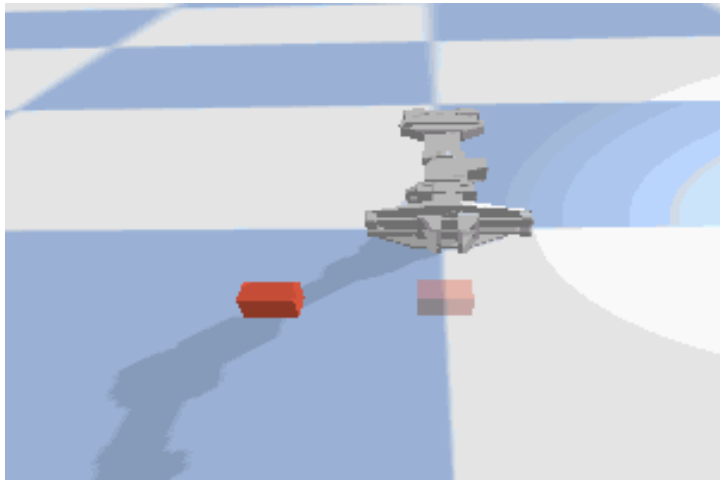
Trained **without** evaluated **with** collision avoidance.





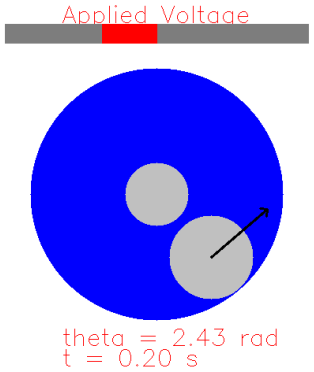
Trained **with** evaluated **with** collision avoidance.

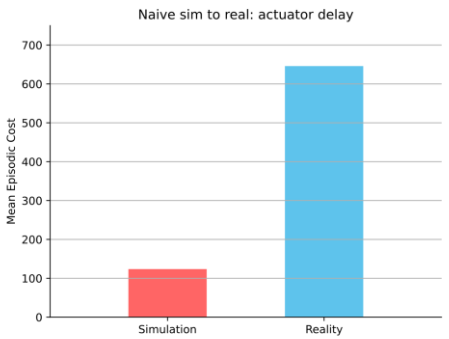




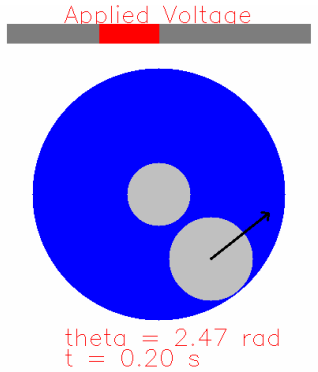
Delays

Ignored



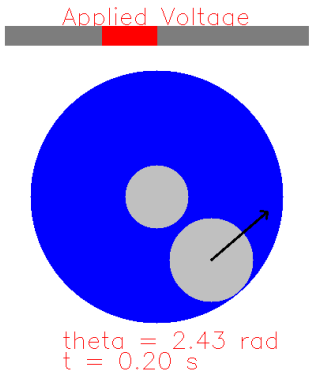


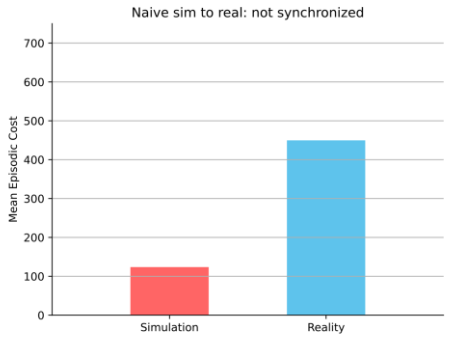
Included in Simulation



Asynchronous Frameworks

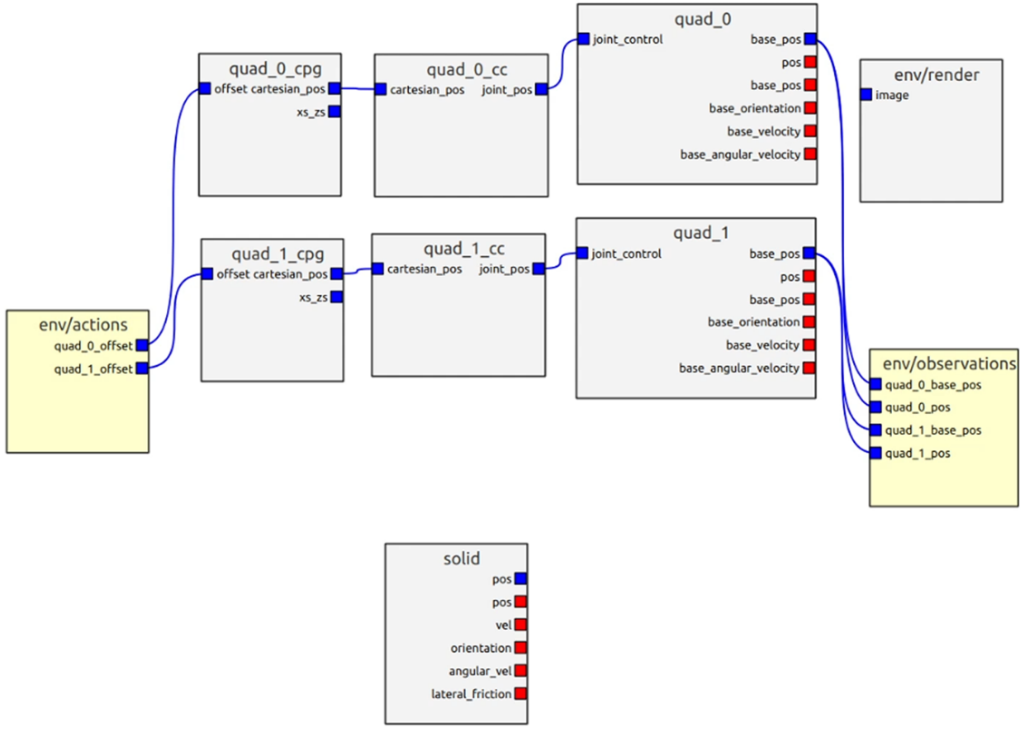
Naive



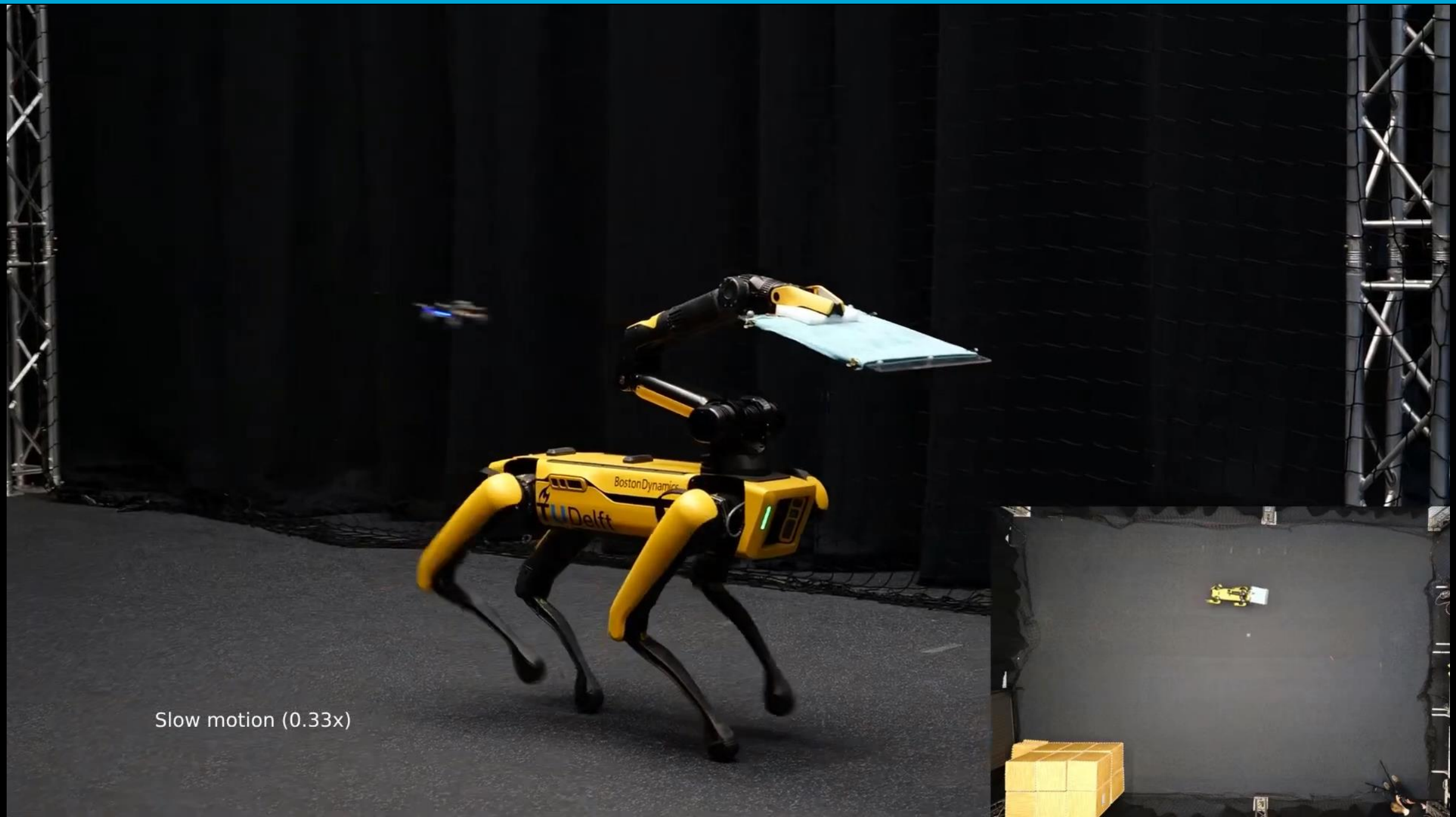


EAGERx (Engine Agnostic Graph Environments for Robotics)

"A Python3 framework that lets you easily define OpenAI gym compatible environments that work both in simulation and the real-world."



Van der Heijden, Luijkx, Ferranti, Kober, & Babuška, RAM 2024


Boston Dynamics
TU Delft
Slow motion (0.33x)

Path following with R=1.0 meter (see fig. 9a)

delay sim.

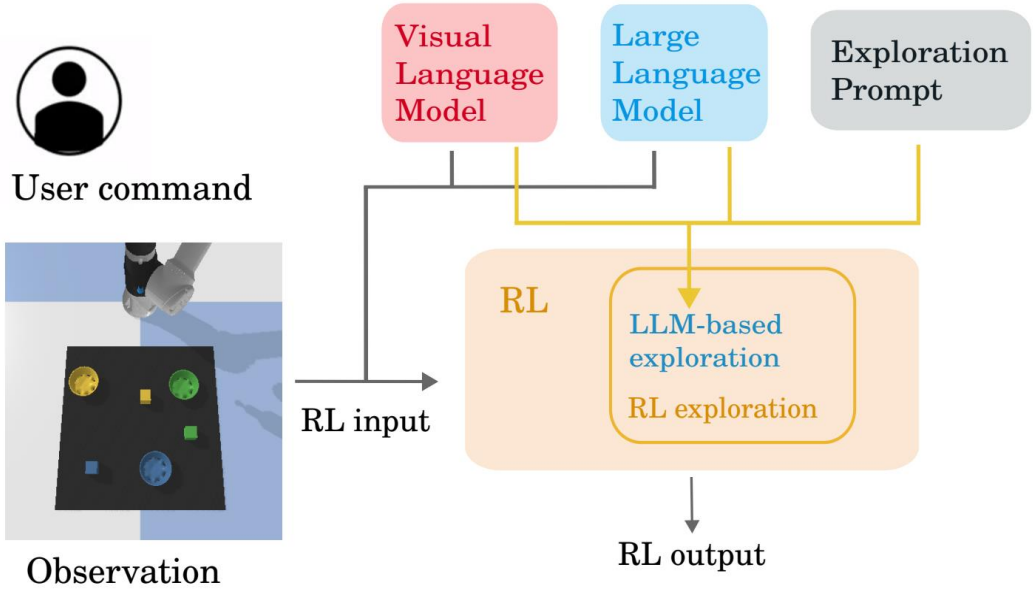
no delay sim.

Background lights turned off for better visibility

Other Ideas

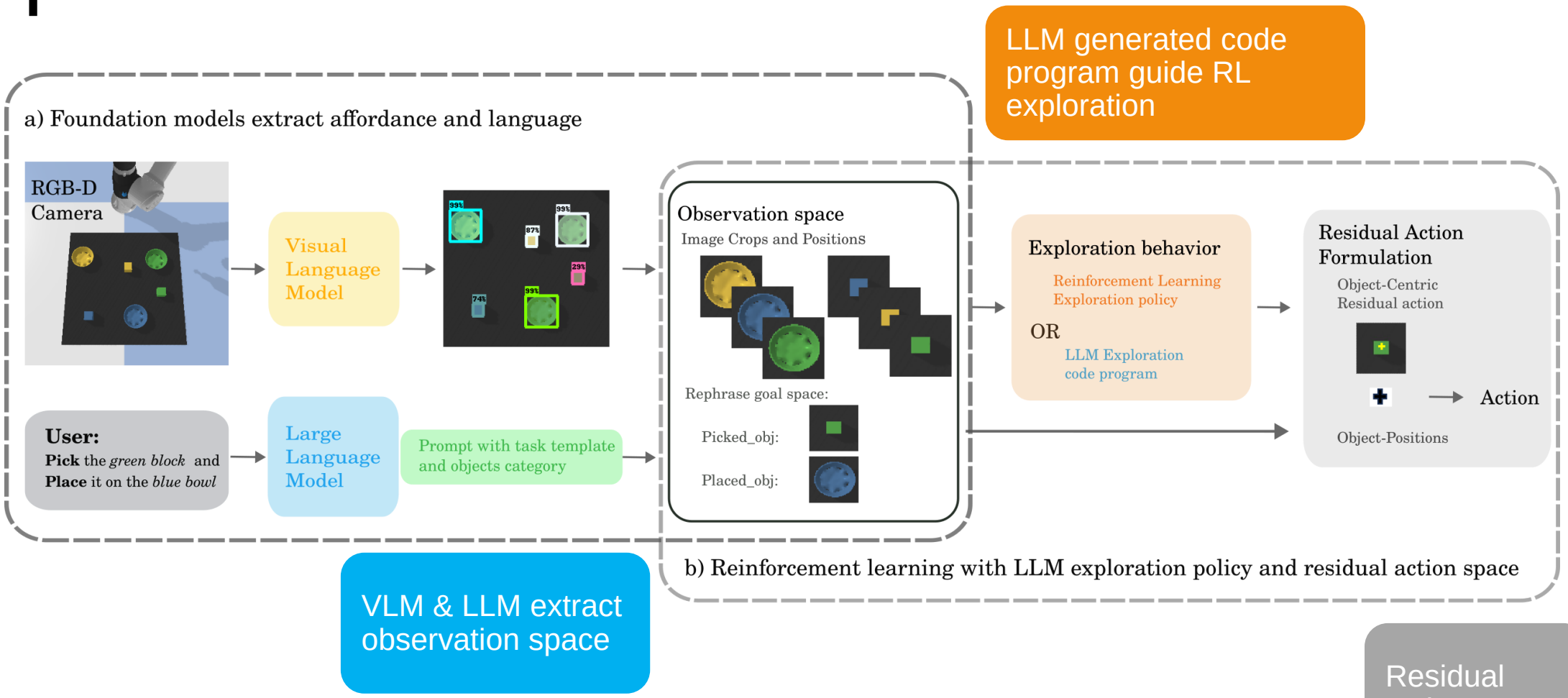
Language conditioned robot policy learning

- Method: Foundation Models + RL
- Utilizing the commonsense knowledge in foundation models to help end-to-end reinforcement learning
- Foundation models guide exploration for reinforcement learning
- Foundation models extract observation space



Ma, Luijkx, Ajanović, & Kober, ICRA 2025

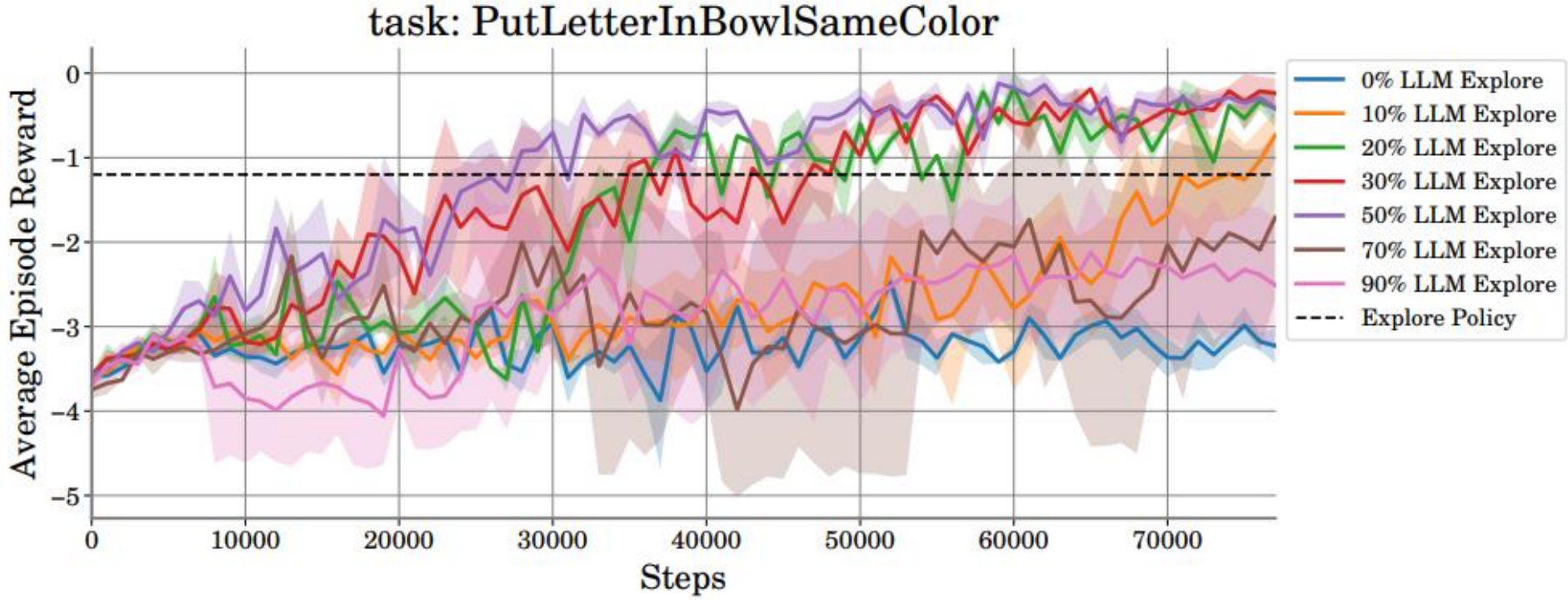
ExploRLLM



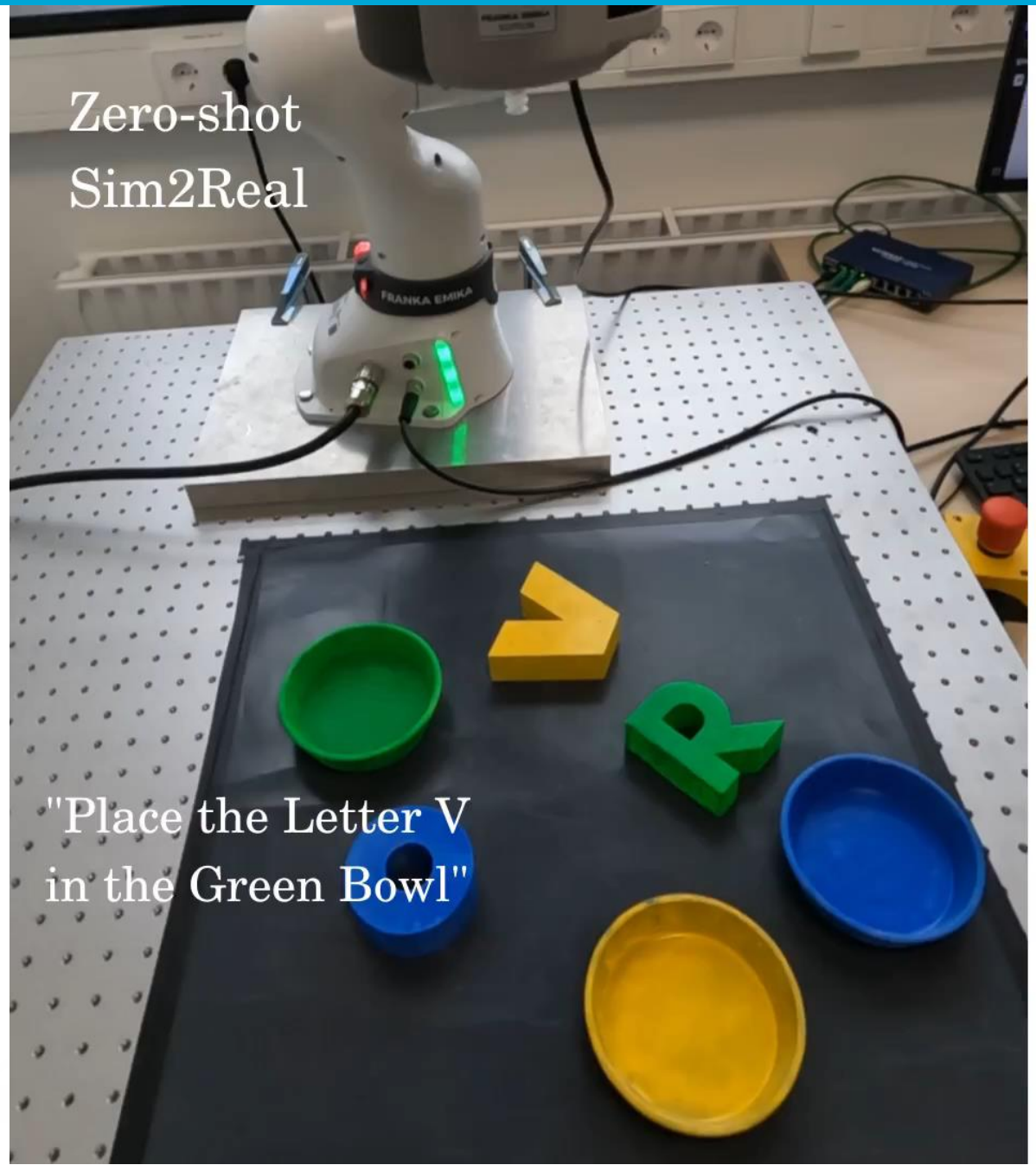
Ma, Luijkx, Ajanović, & Kober, ICRA 2025

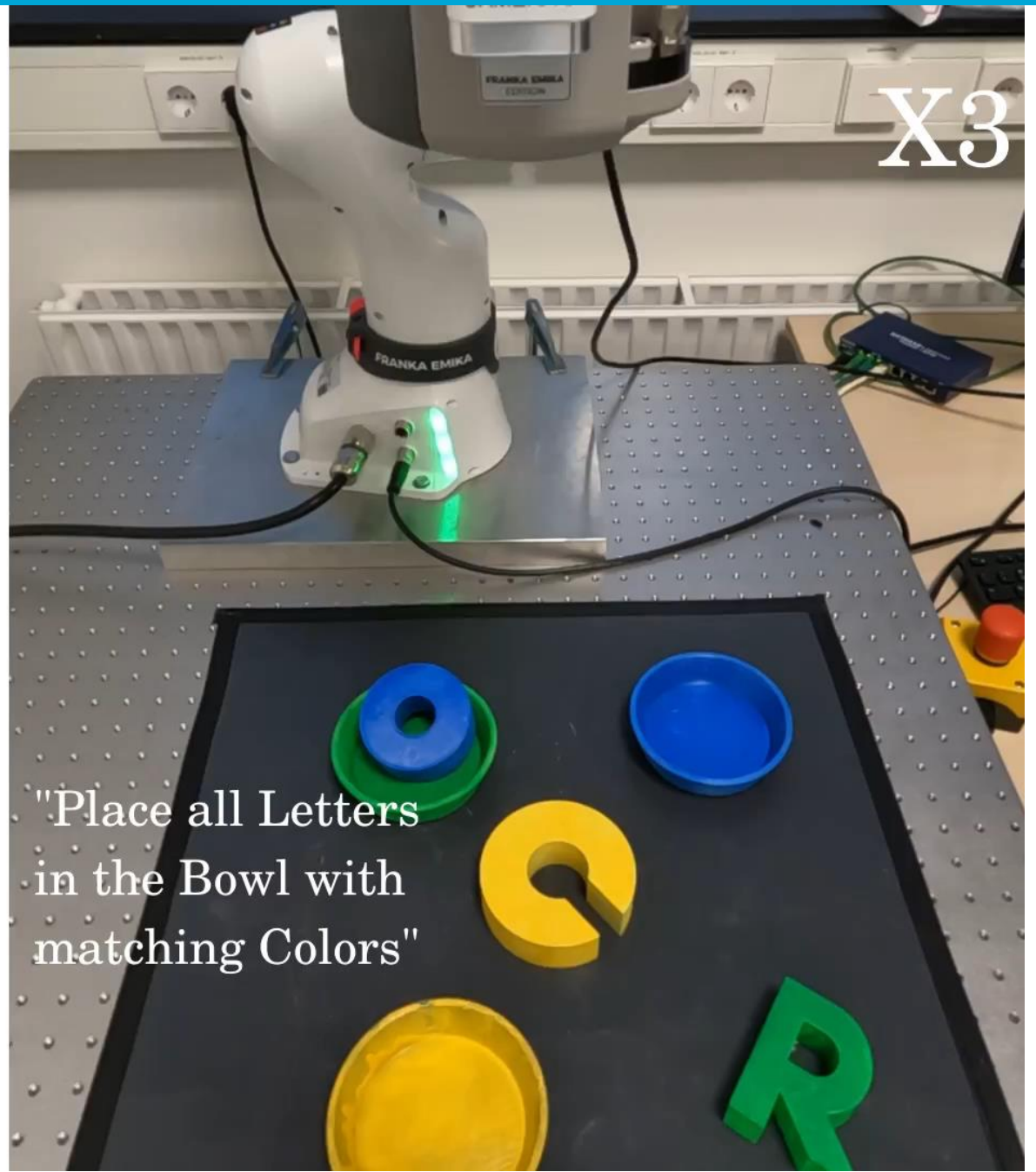
Experiment

Analyze influence of LLM exploration frequency



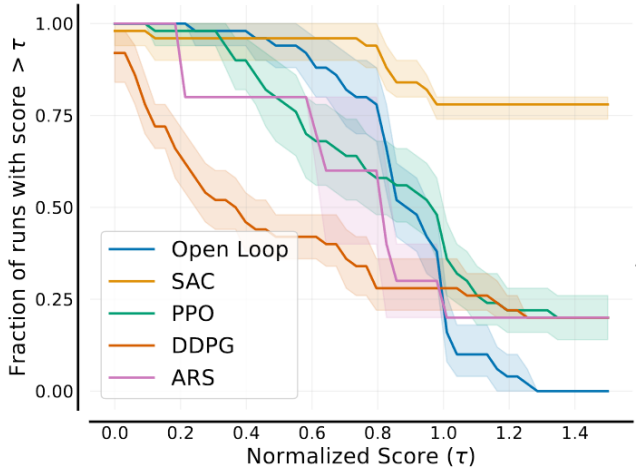
Ma, Luijkx, Ajanović, & Kober, ICRA 2025

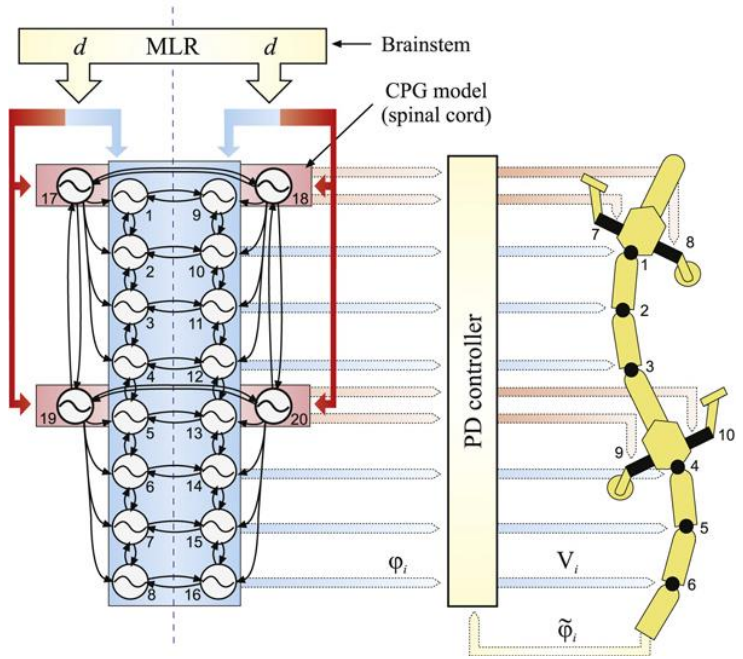

Zero-shot
Sim2Real
"Place the Letter V
in the Green Bowl"


X3
"Place all Letters
in the Bowl with
matching Colors"

Central Pattern Generators

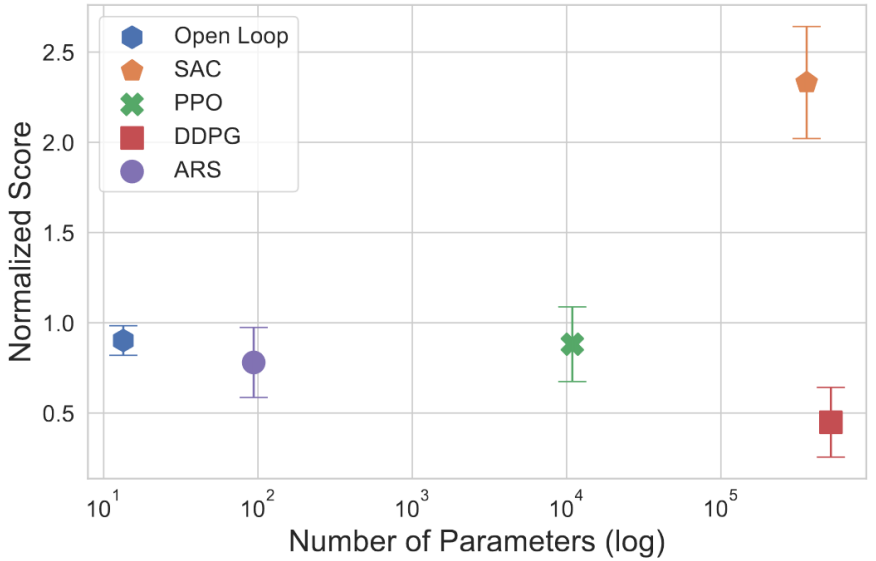
- Perform as well





- But are a lot more efficient

	SAC		PPO		DDPG		ARS		Open-Loop	
	CPU	GPU	CPU	GPU	CPU	GPU	CPU	GPU	CPU	GPU
Runtime (in min.)	80	30	10	14	60	25	5	N/A	**2**	N/A



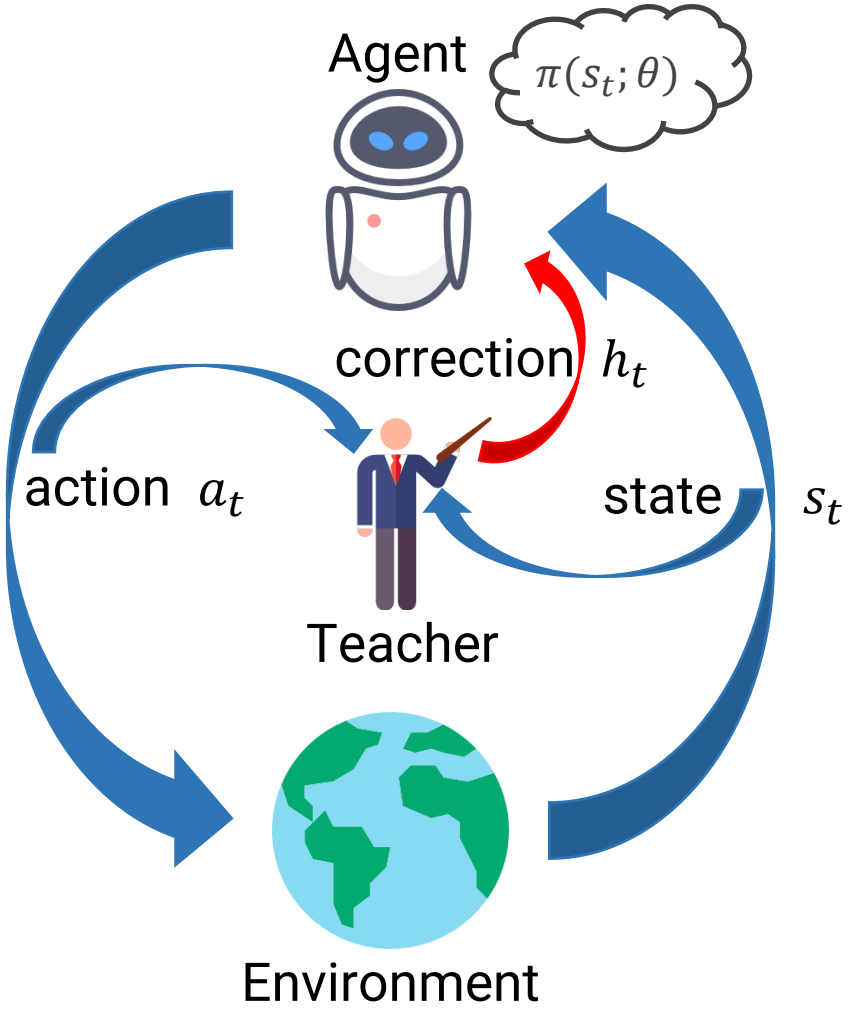
- And a lot more robust

Raffin et al., RLC 2024

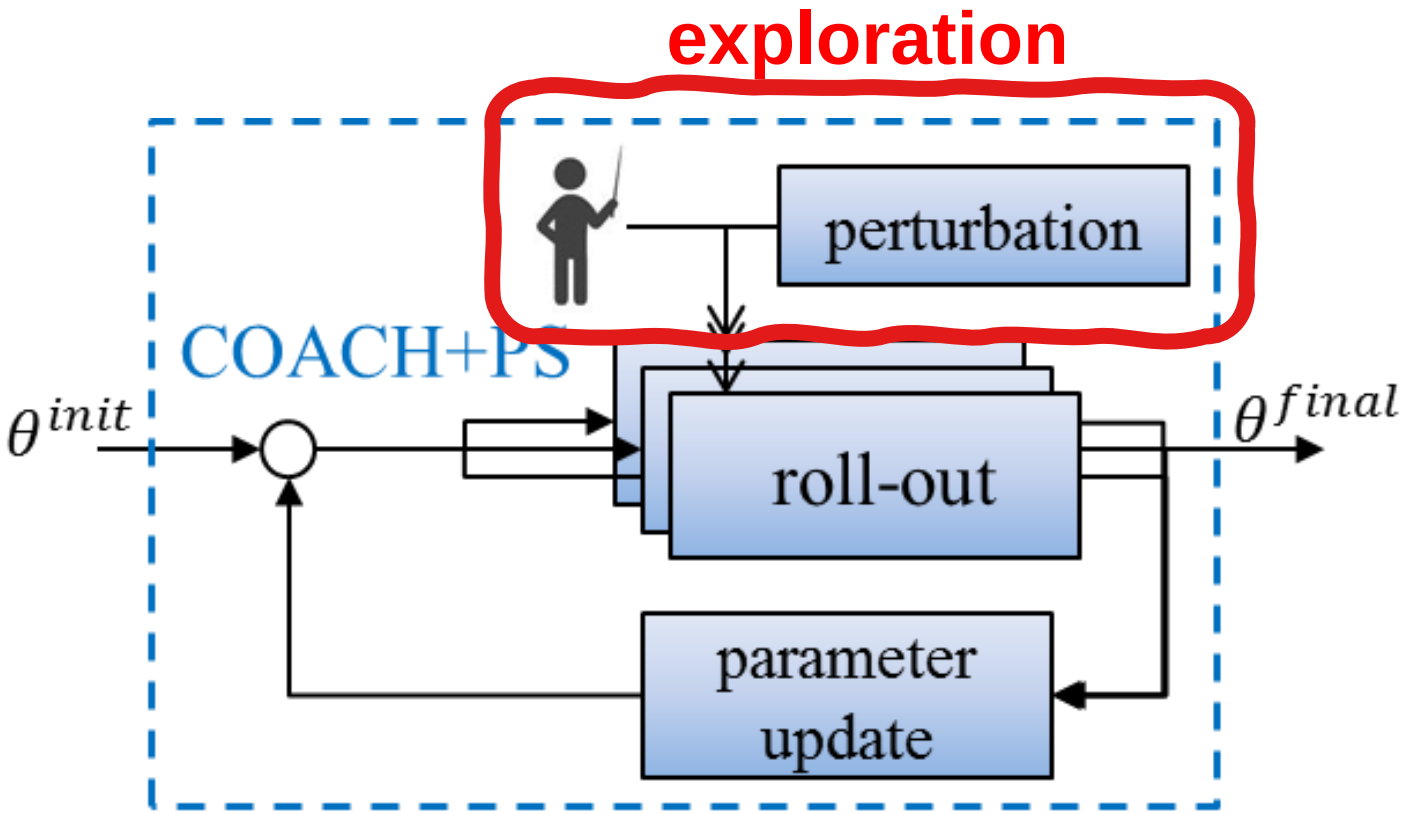
RL in Simulation

Interactive Reinforcement Learning

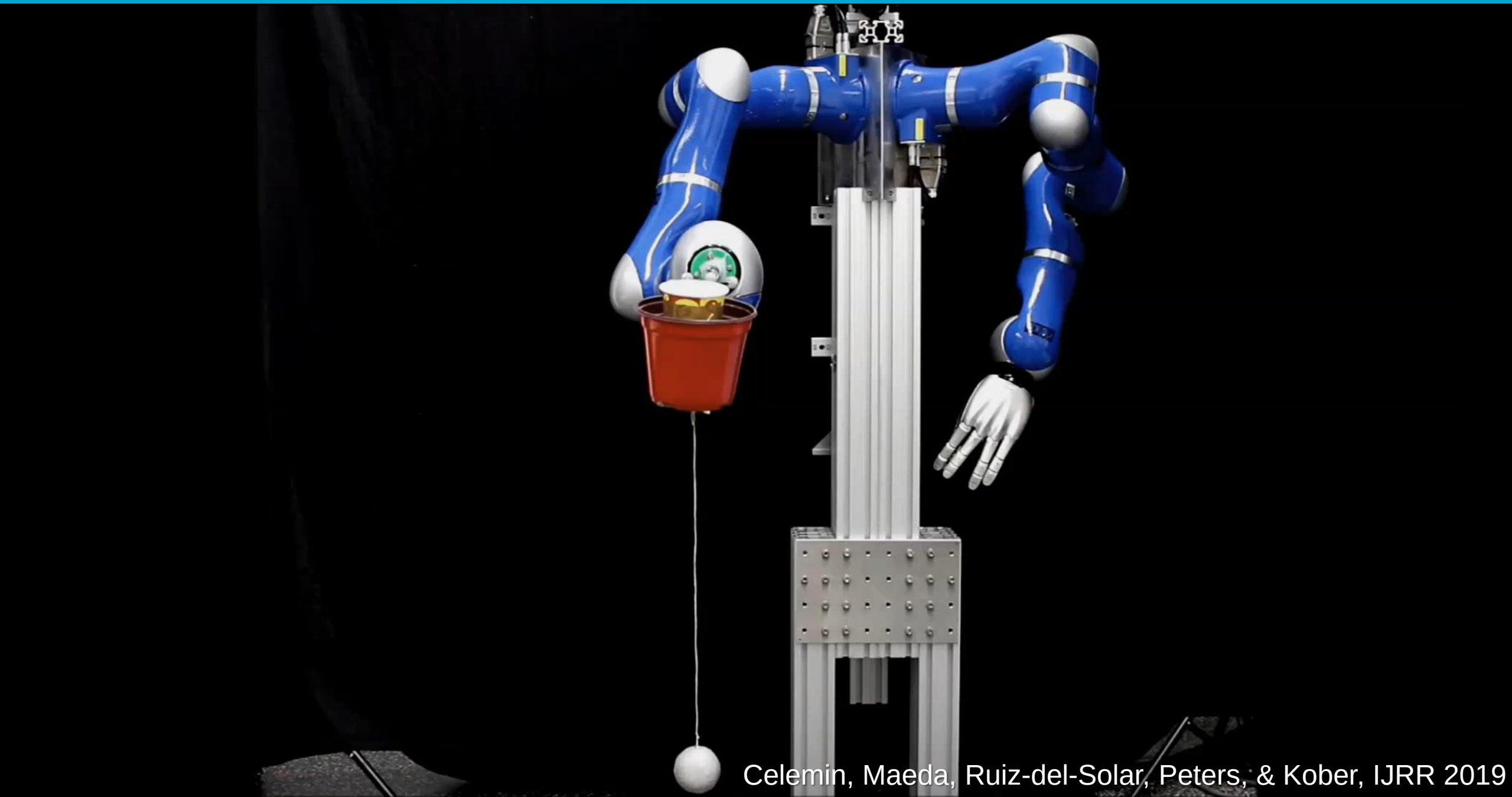
Interactive Learning



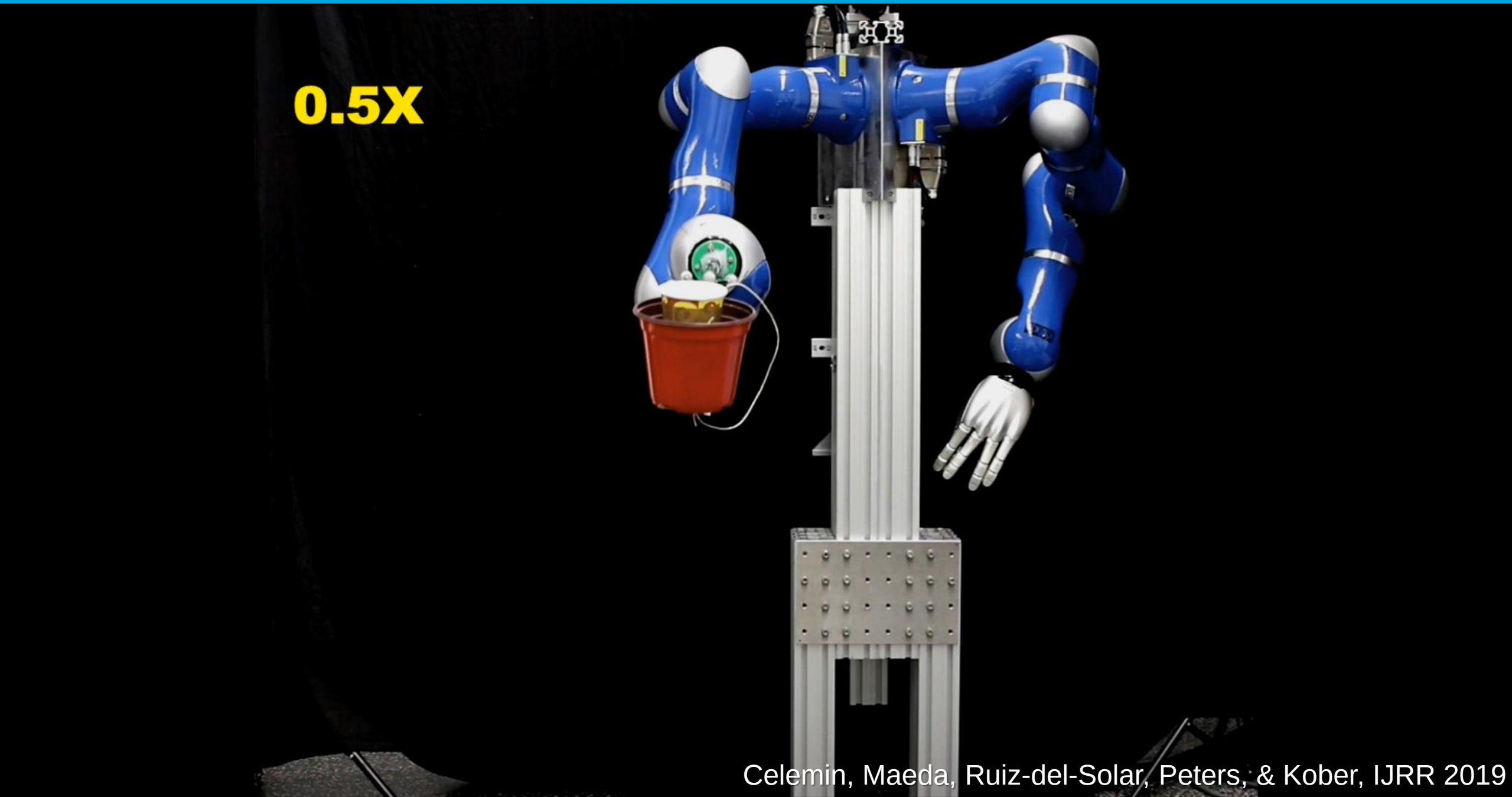
IL + RL: Simultaneous



Celemin, Maeda, Ruiz-del-Solar, Peters, & Kober, IJRR 2019

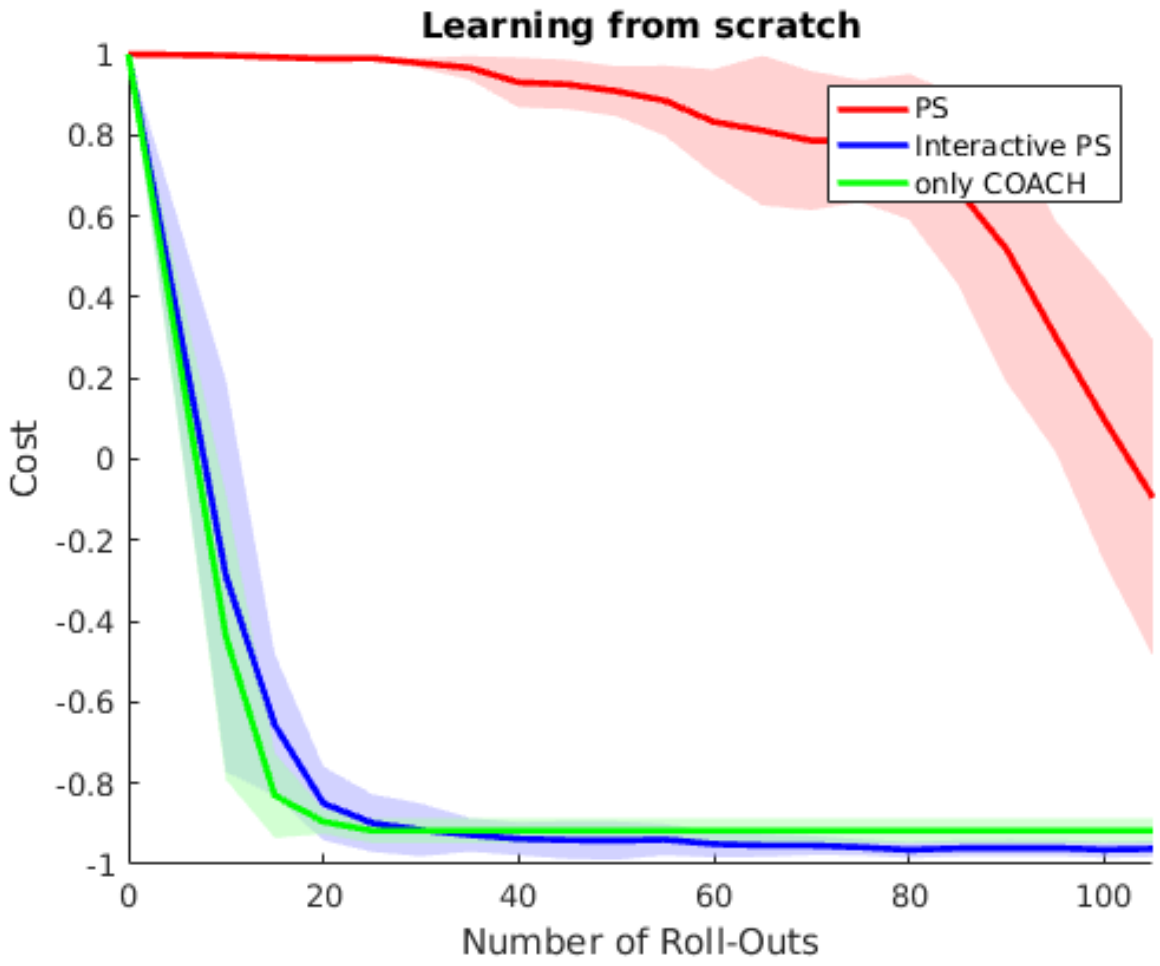


Celemin, Maeda, Ruiz-del-Solar, Peters, & Kober, IJRR 2019



Celemin, Maeda, Ruiz-del-Solar, Peters, & Kober, IJRR 2019

Results



Celemin, Maeda, Ruiz-del-Solar, Peters, & Kober, IJRR 2019

Summary

Summary

- Specific challenges
  - Small (real) data
  - Safety
- Tractability through prior knowledge
  - Policy structure
  - Demonstrations
  - Simulations
  - Foundation models
  - Etc.