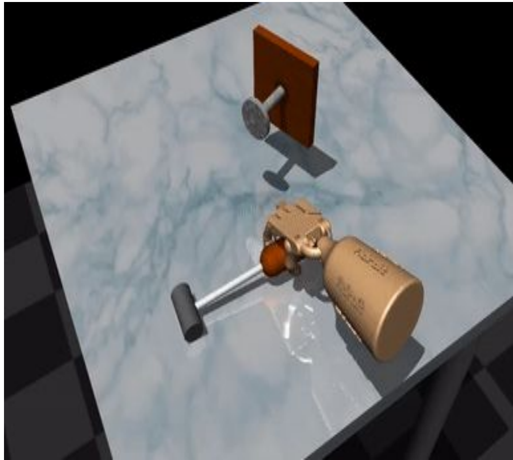
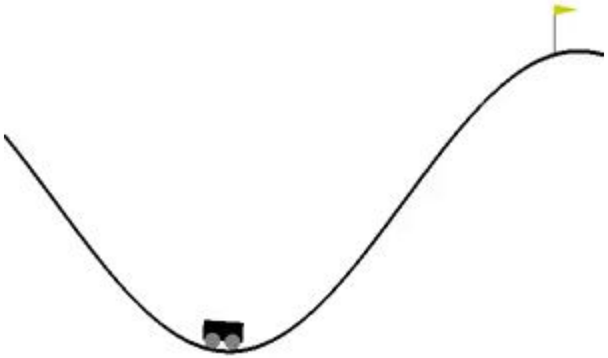
Reinforcement Learning in Robotics

Boyuan Chen
PhD student, Computer Science
Creative Machines Lab
http://www.cs.columbia.edu/~bchen/



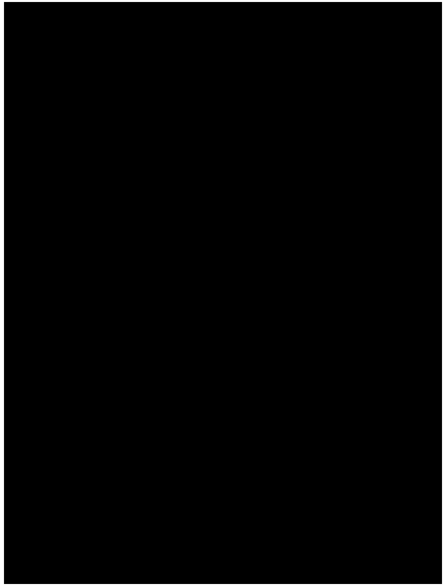
A great time for RL ...

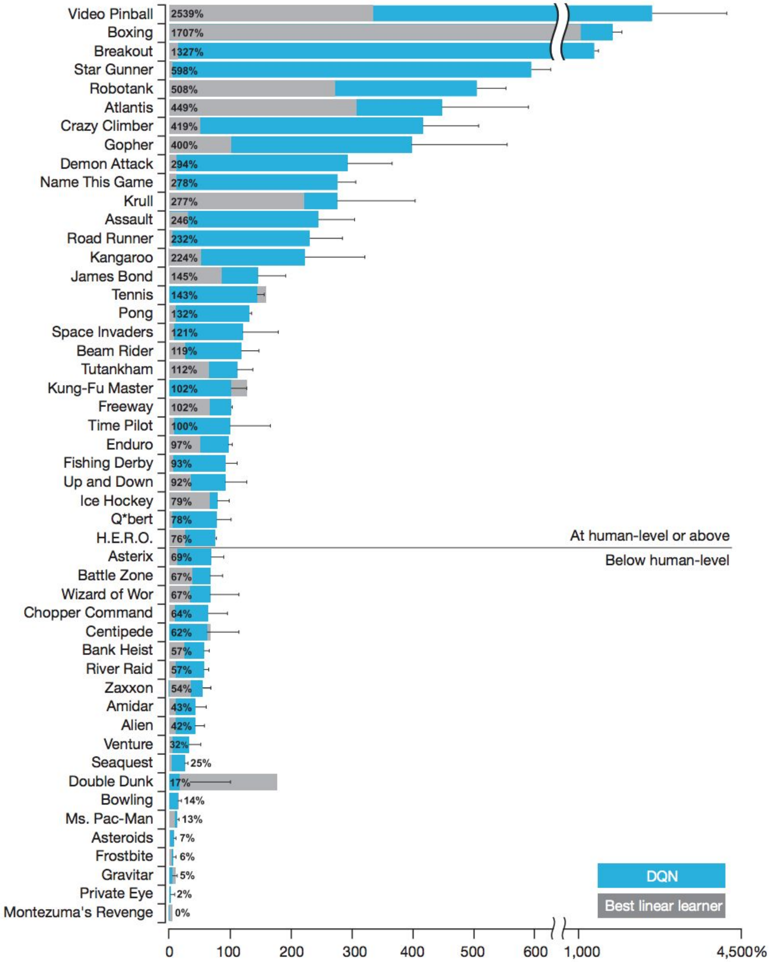
Classic Control:



A great time for RL ...

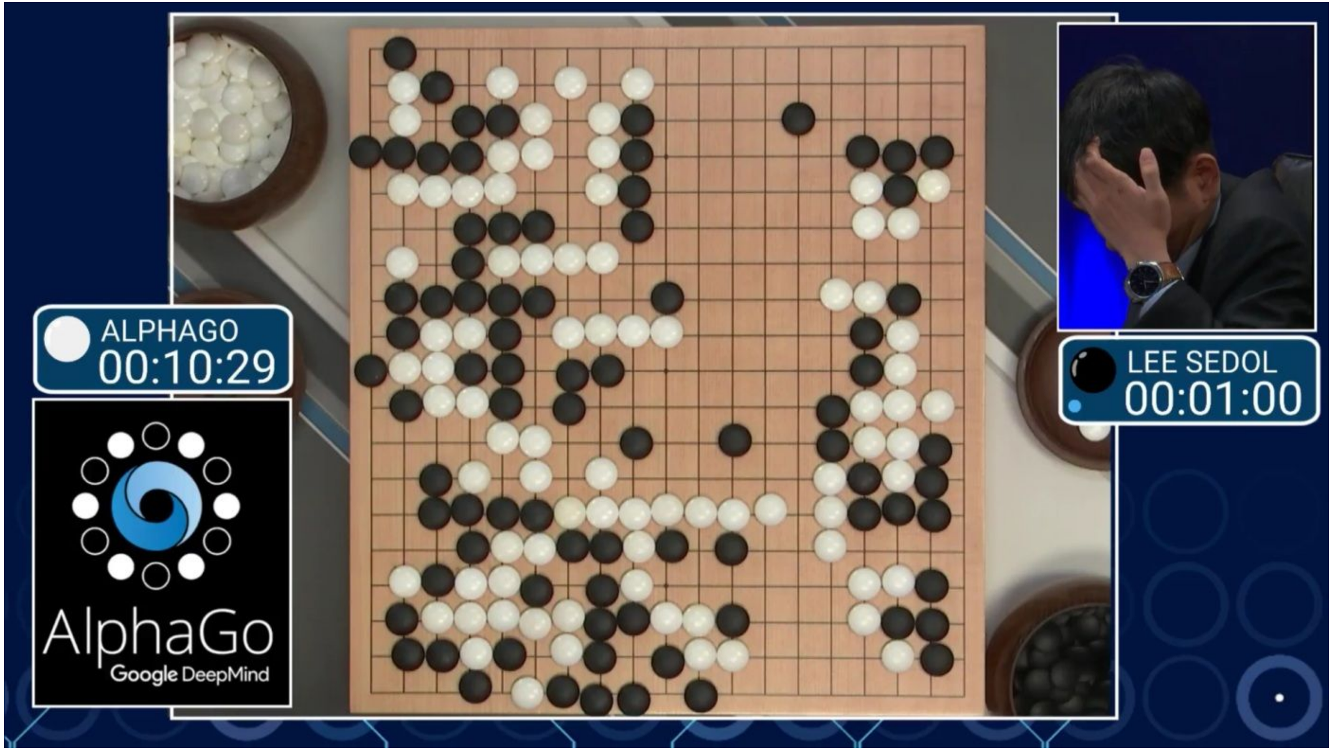
Atari Games:





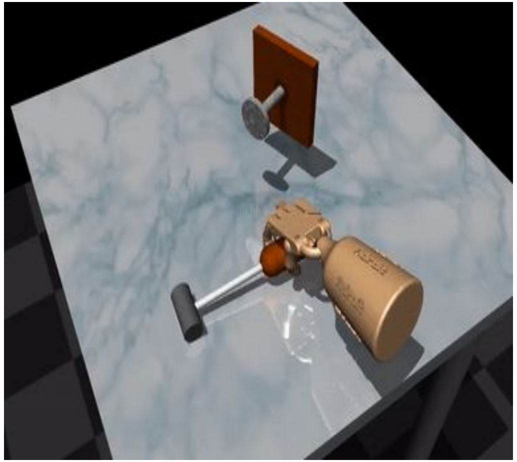
A great time for RL ...

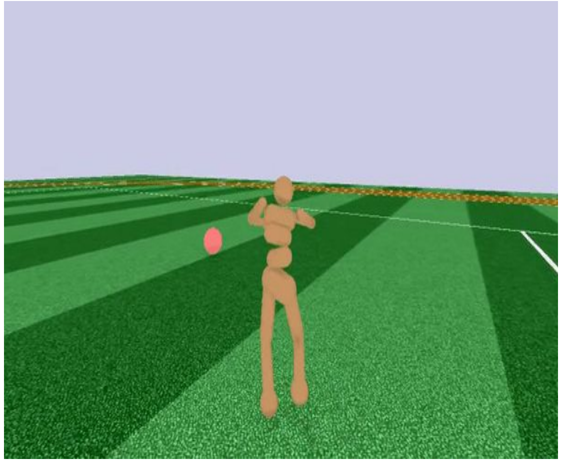
Go:



A great time for RL ...

Robotics

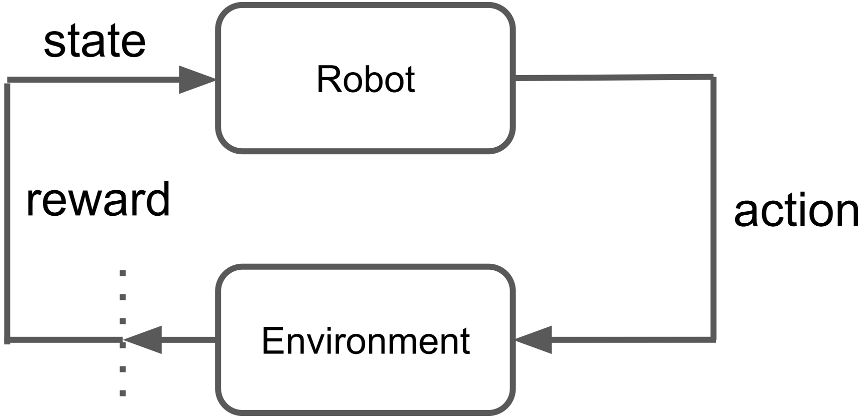




Outline

How to set up RL problems in Robotics? → Important topics → How to get started?

RL in Robotics

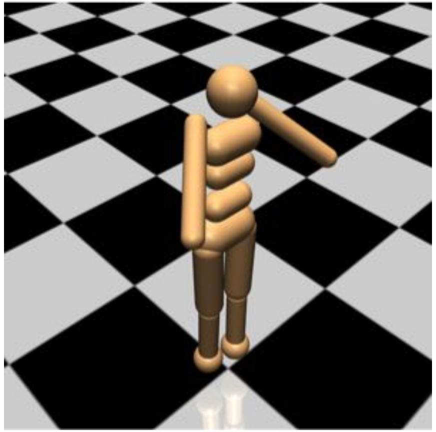


- No labelled data;
- No access to real model;
- No fixed rule
- Continuous space
- Complex transition dynamics

Problem Setting

- MDP process defined by: $\mathcal{S}, \mathcal{A}, P, r, \rho_0, \gamma, T$
- Policy: $\pi_\theta : \mathcal{S} \times \mathcal{A} \to \mathbb{R}_{\geqslant 0}$
- Expected Reward: $\mathbb{E}_\tau \left[\sum_{t=0}^{T} \gamma^t r(s_t, a_t) \right]$
- Trajectories: $\tau = (s_0, a_0, \ldots)$

$$\begin{tikzcd} s_t \arrow[r, "\pi_\theta"] \arrow[dr] & a_t \arrow[d] \\ & s_{t+1} \end{tikzcd}$$



Real robot?

Topics: Manipulation

Deep Reinforcement Learning for Robotics Manipulation with Asynchronous Off-Policy Updates: Gu et al, 2016.

Topics: Manipulation

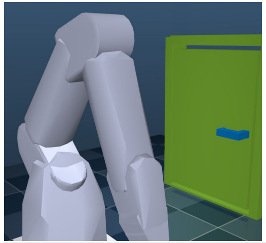
Contribution:

- DeepRL

Real complex robot system

Complex task

Asynchronous data collection



- Safety Exploration

Topics: Manipulation

Off-policy Deep Q-function based algorithms:

$$Q^{\pi_n}(\boldsymbol{x}_t, \boldsymbol{u}_t) = \mathbb{E}_{r_{i \geq t}, \boldsymbol{x}_{i>t} \sim E, \boldsymbol{u}_{i>t} \sim \pi_n}[R_t | \boldsymbol{x}_t, \boldsymbol{u}_t]$$

$$\boldsymbol{\mu}_{n+1}(\boldsymbol{x}_t) = \arg\max_{\boldsymbol{u}} Q^{\pi_n}(\boldsymbol{x}_t, \boldsymbol{u}_t)$$

- DDPG (Deep Deterministic Policy Gradient) Actor-Critic
- NAF (Normalized Advantage Function)

$$Q(\boldsymbol{x}, \boldsymbol{u}|\theta^Q) = A(\boldsymbol{x}, \boldsymbol{u}|\theta^A) + V(\boldsymbol{x}|\theta^V)$$

$$A(\boldsymbol{x}, \boldsymbol{u}|\theta^A) = -\frac{1}{2}(\boldsymbol{u} - \boldsymbol{\mu}(\boldsymbol{x}|\theta^\mu))^T \boldsymbol{P}(\boldsymbol{x}|\theta^P)(\boldsymbol{u} - \boldsymbol{\mu}(\boldsymbol{x}|\theta^\mu))$$

Topics: Manipulation

- State Representation

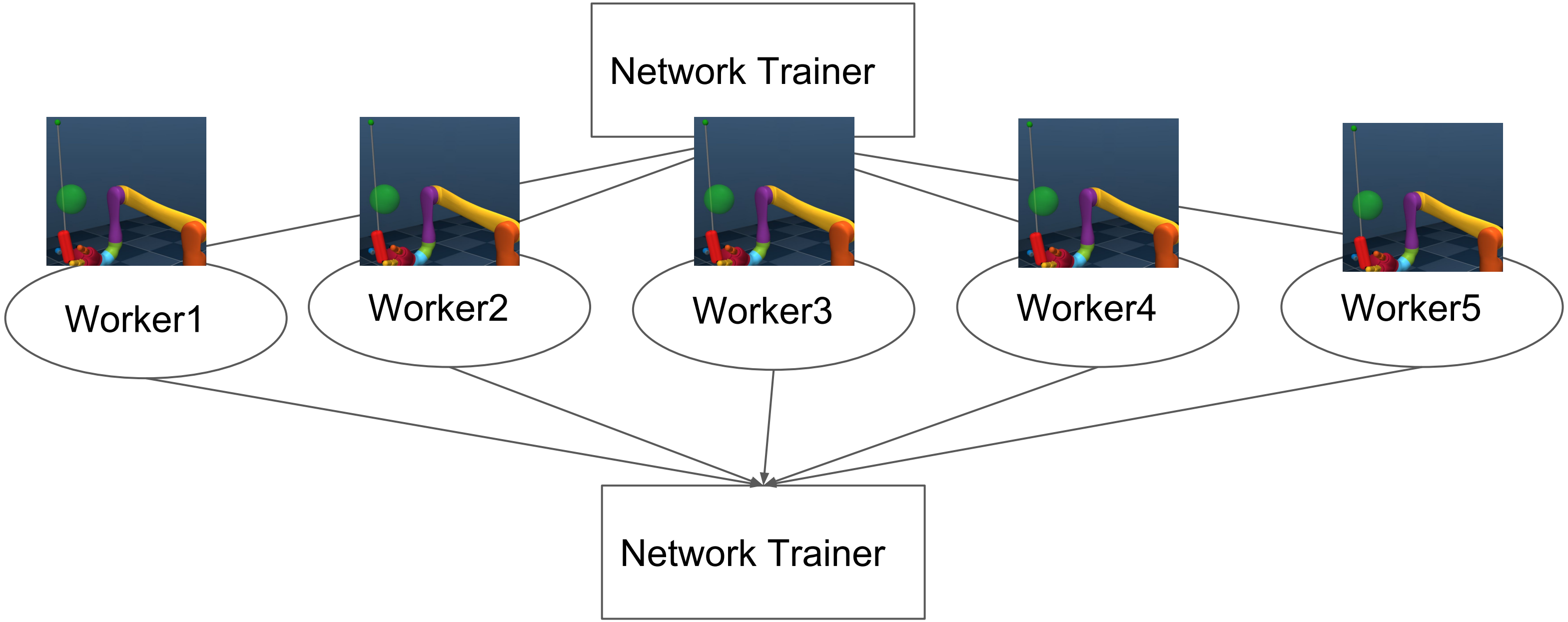
Joint Angles	End-effector Positions	Time Derivatives	Target Position to the State

Distance from target location

Handle position when door is closed + Quaternion of the door

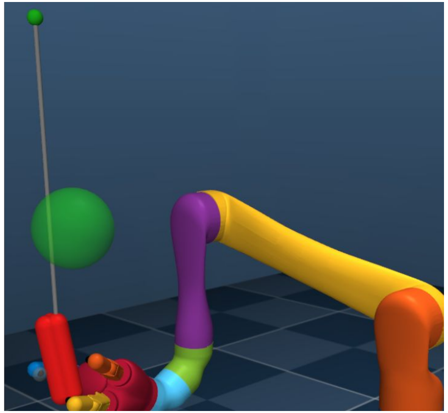
Topics: Manipulation

- Asynchronous NAF



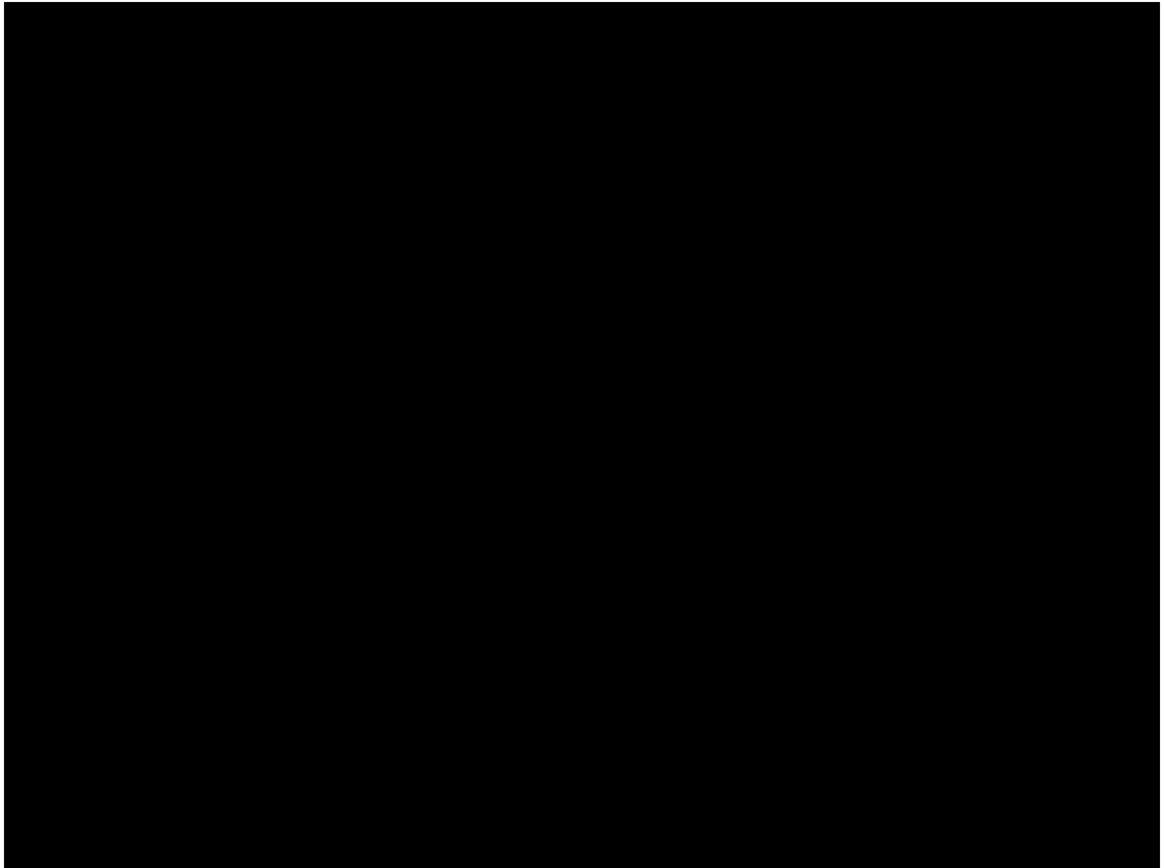
Topics: Manipulation

- Safety Constraints
  - Joint Position Limits:
    - Maximum commanded velocity allowed per joint
    - Strict position limits for each joint
  - Bounding sphere for end-effector position



Very important for training from scratch on real systems!

Topics: Manipulation



Topics: Manipulation

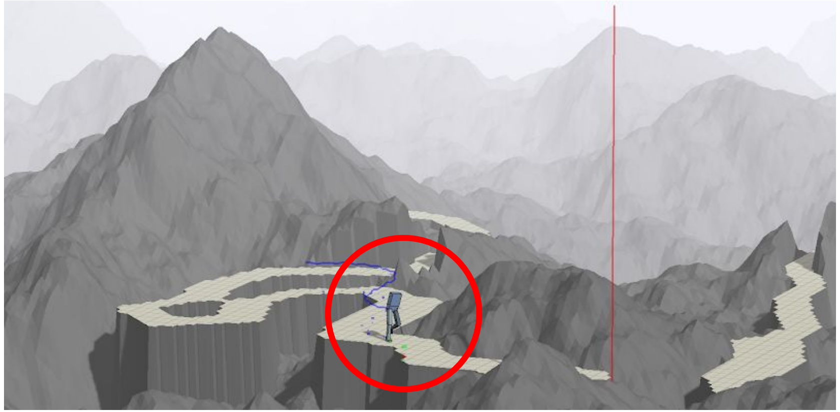
What have we learned from this?

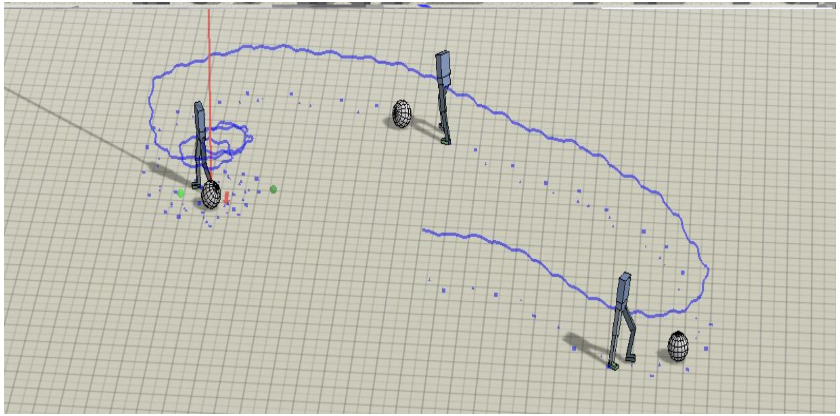
- Efficiency is important for real robots
- Safety is important for real robots
- It is possible to apply DeepRL on real robots to accomplish complex tasks

Complex Task?

Topics: Locomotion

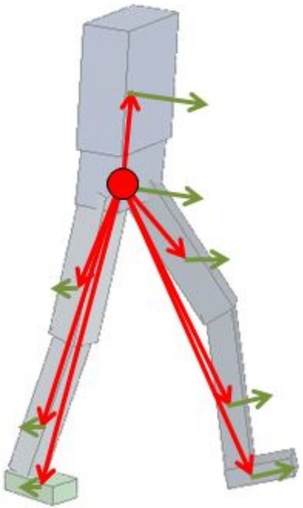
DeepLoco: Dynamic Locomotion Skills Using Hierarchical Deep Reinforcement Learning: Peng et al, 2017

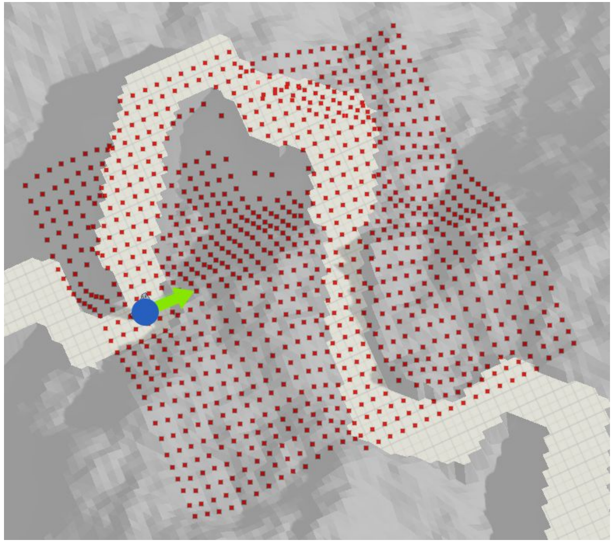




Topics: Locomotion

DeepLoco: Dynamic Locomotion Skills Using Hierarchical Deep Reinforcement Learning: Peng et al, 2017



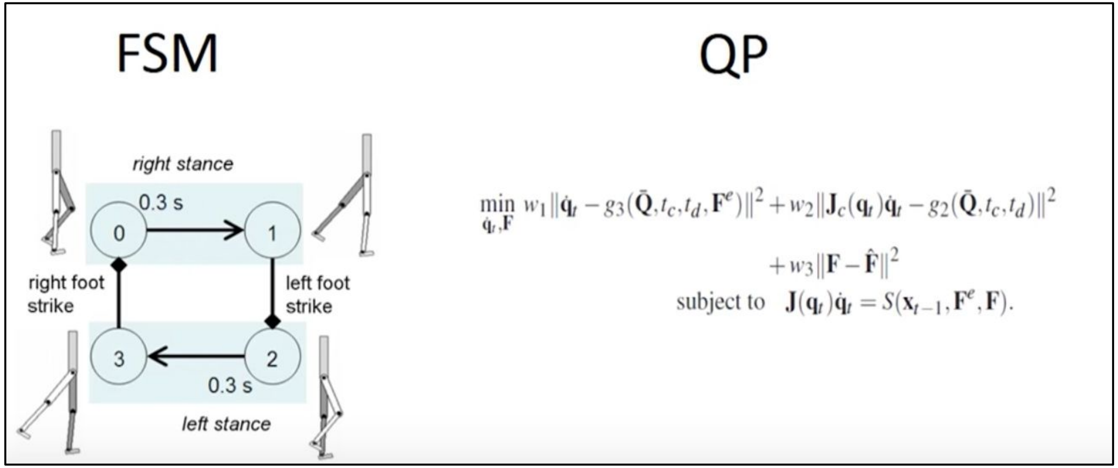


Topics: Locomotion

- Highlight:

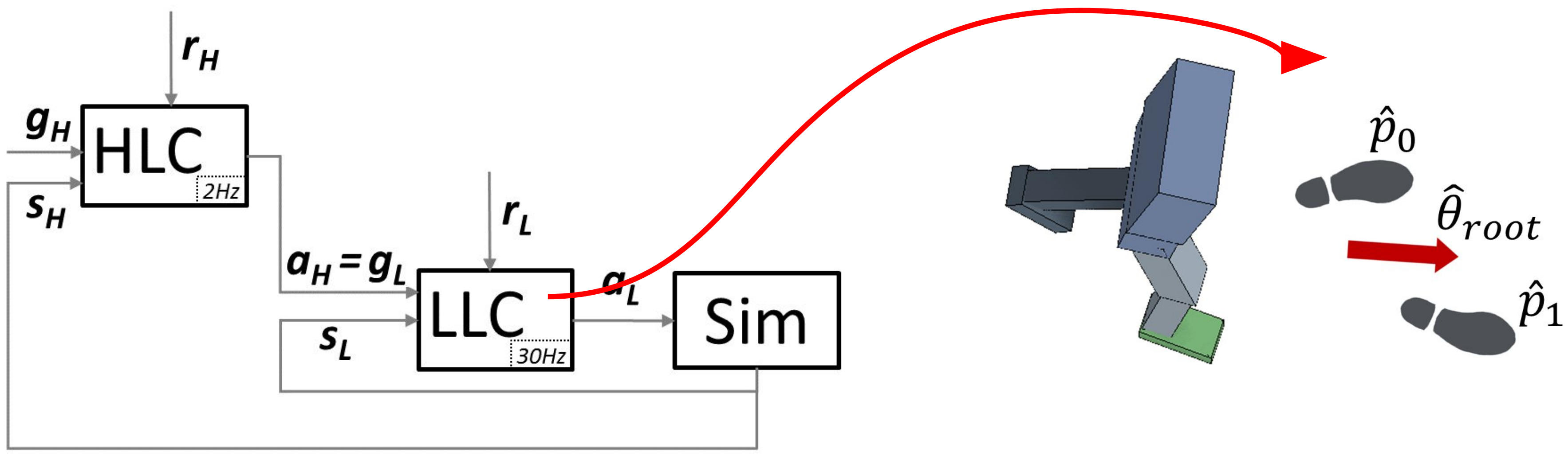
  Less prior knowledge

  Hierarchical RL



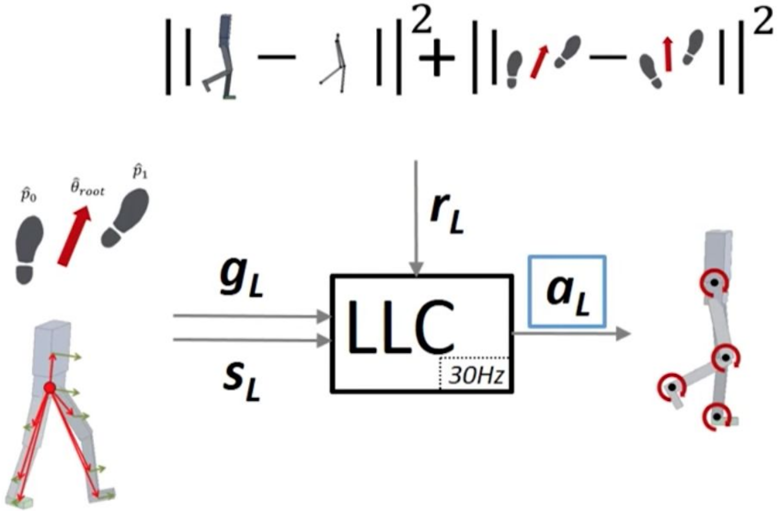
Topics: Locomotion

DeepLoco: Dynamic Locomotion Skills Using Hierarchical Deep Reinforcement Learning: Peng et al.



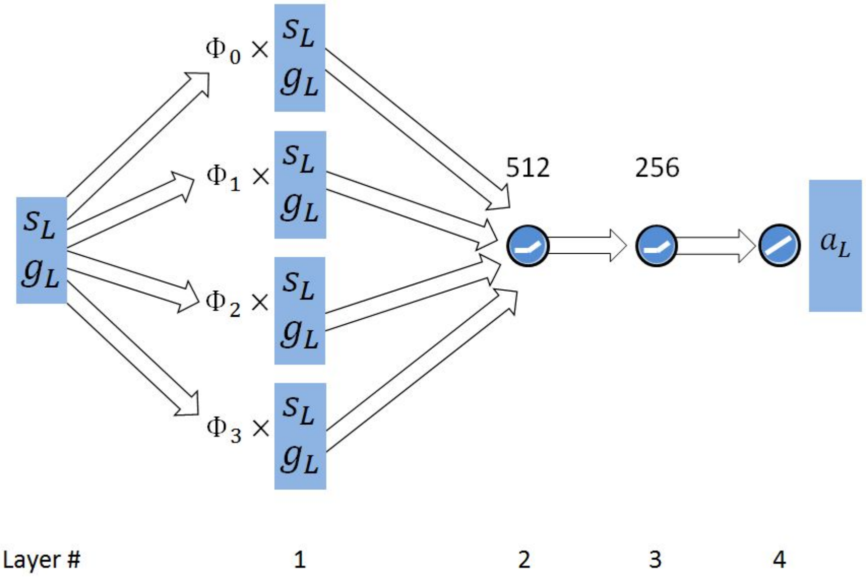
Topics: Locomotion

DeepLoco: Dynamic Locomotion Skills Using Hierarchical Deep Reinforcement Learning: Peng et al.



Topics: Locomotion

LLC (Low Level Controller):

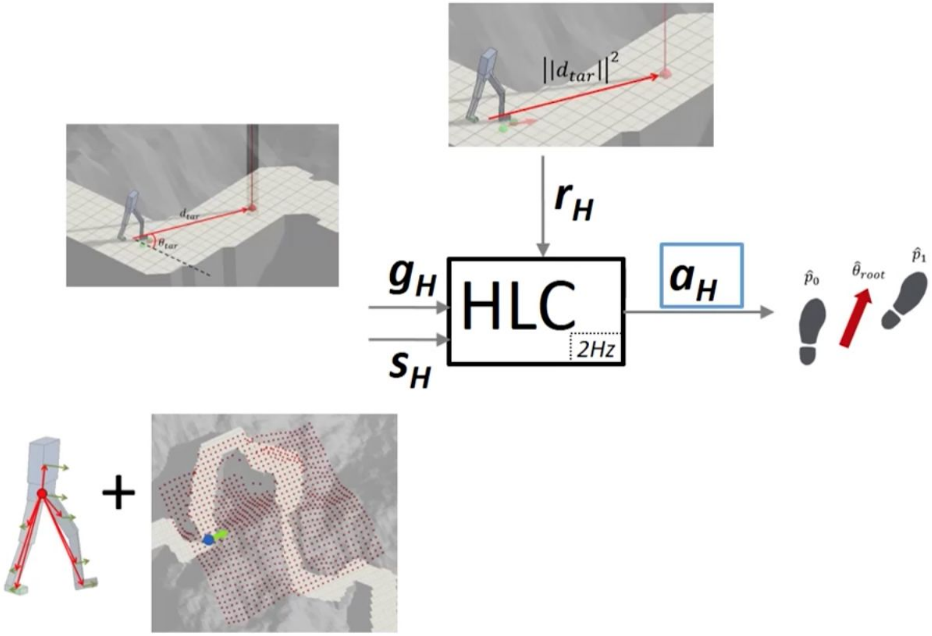


- State
- Action
- Goal
- Reward

$$r_L = w_{pose}r_{pose} + w_{vel}r_{vel} + w_{root}r_{root}$$
$$+w_{com}r_{com} + w_{end}r_{end} + w_{heading}r_{heading}$$

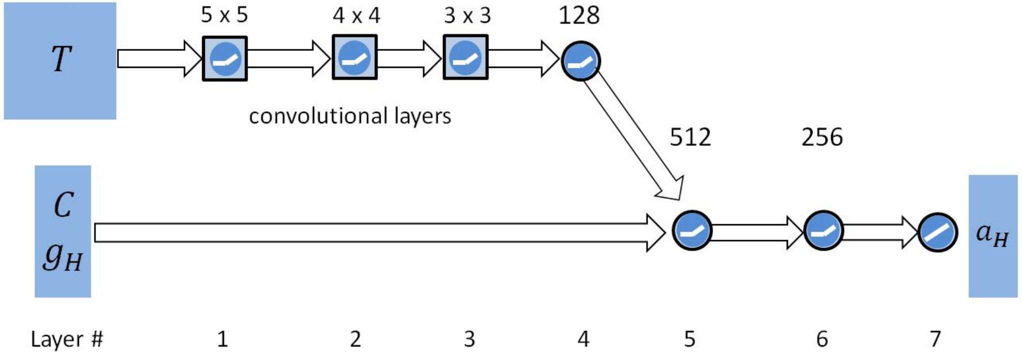
Topics: Locomotion

HLC (High Level Controller):



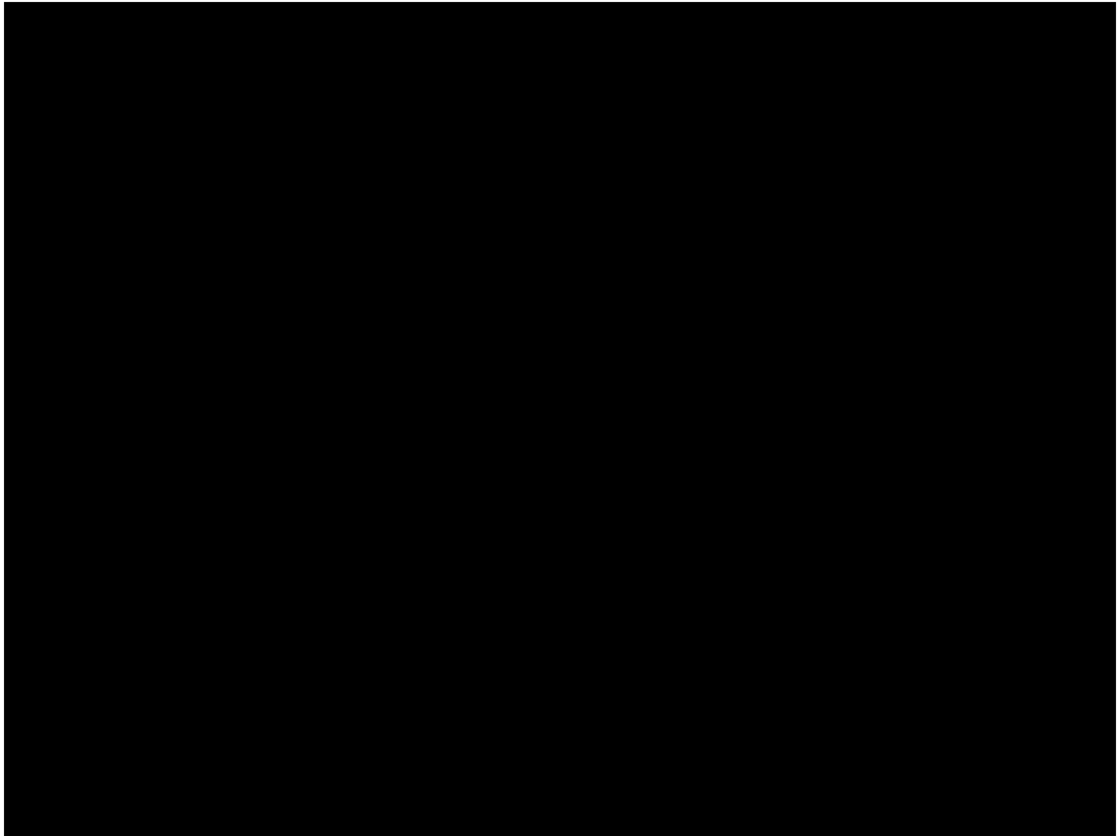
Topics: Locomotion

HLC (High Level Controller):



- State
- Training

Topics: Locomotion



Topics: Locomotion

What have we learned from this?

- Identify hierarchical structure is important.
- State representation for different hierarchies are important.

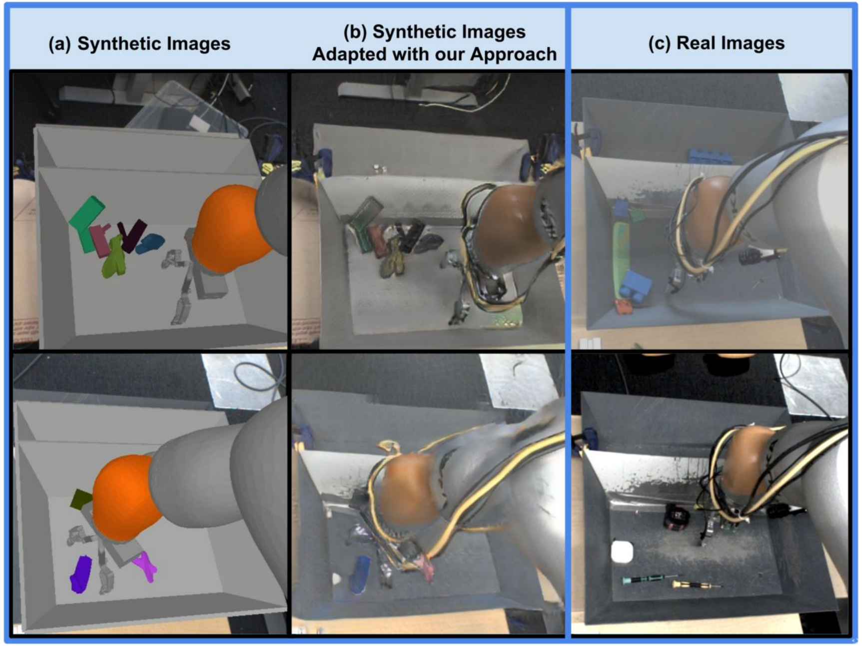
Can we do better?

- Current topic: identify the internal hierarchical structure automatically.

Are we ready?

Topics: Sim2Real

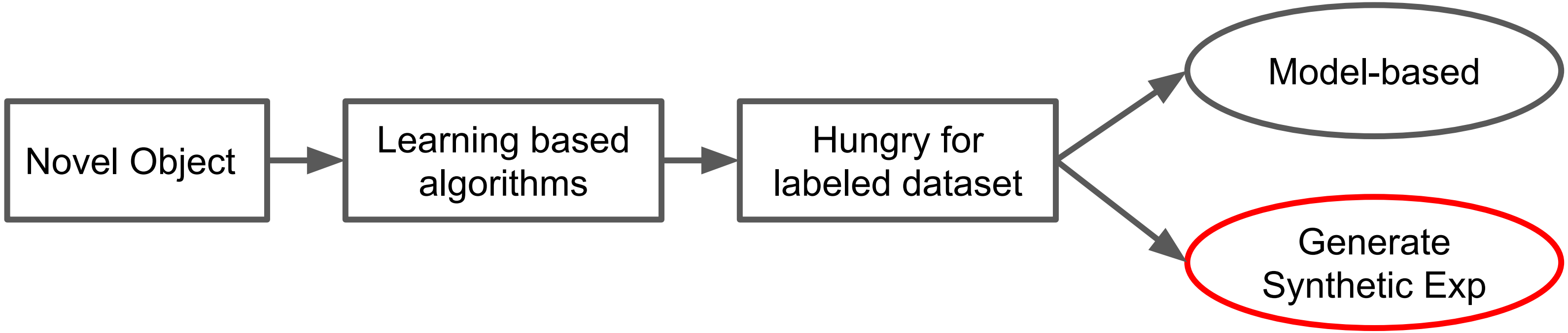
Using Simulation and Domain Adaptation to Improve Efficiency of Deep Robotic Grasping: Bousmalis et al, 2017



Topics: Sim2Real

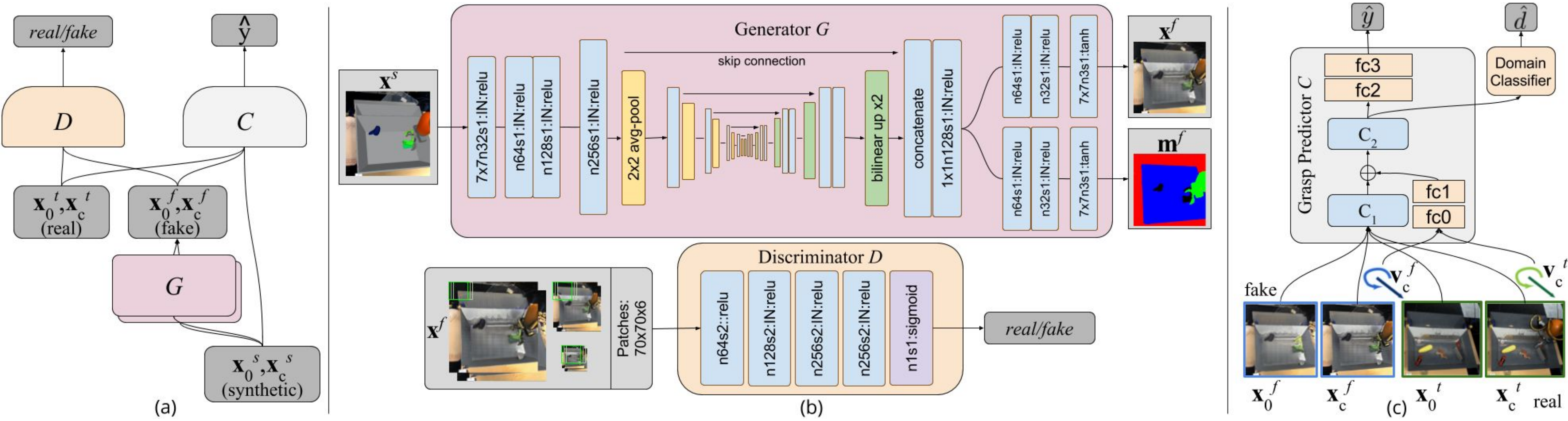
Why this is important?

- Grasping:



Topics: Sim2Real

Approach



How to get started?

Algorithm

Environment

Algorithms

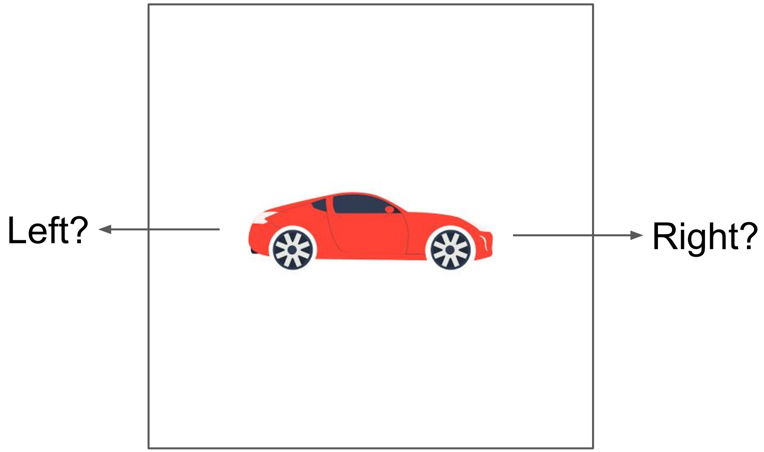
- Libraries:
    - Baselines (OpenAI): https://github.com/openai/baselines
    - Rllab (OpenAI): https://github.com/rll/rllab
    - Coach (Intel): https://github.com/NervanaSystems/coach
    - ...
- Implement your own:
    - Go back to the original paper
    - Use open-source code as reference
    - Start testing from toy examples

Algorithms

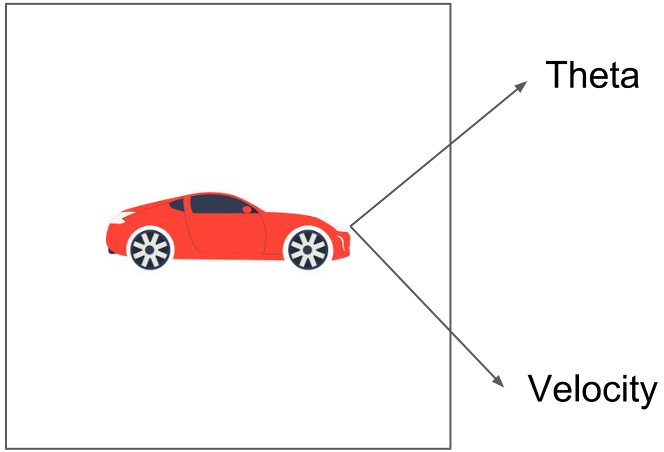
How to select an algorithm for your problem?

- Action space: continuous? Discrete?

More often in Robotics



Discrete

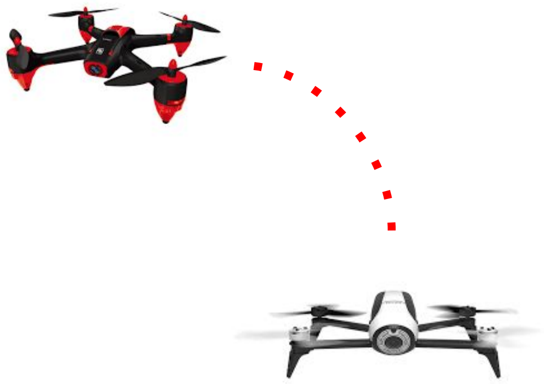


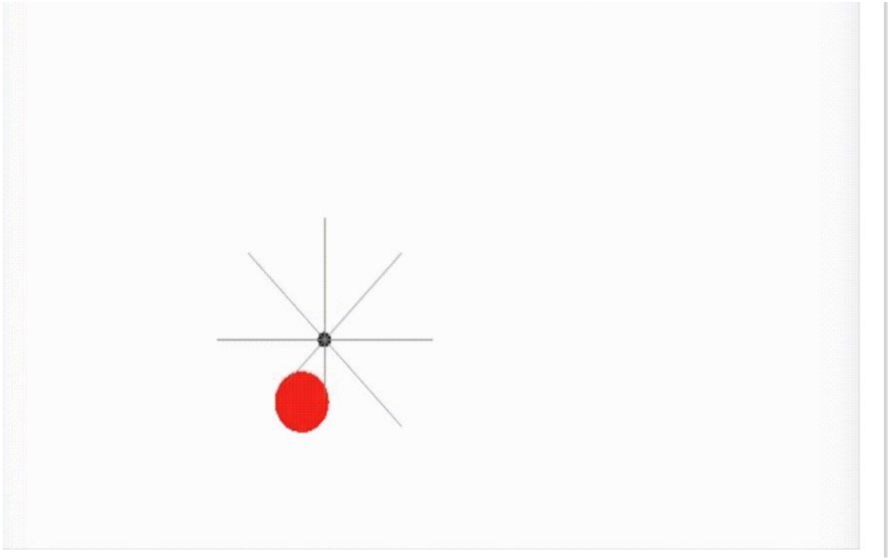
Continuous

Algorithms

How to select an algorithm for your problem?

- Reduce your problem to toy example

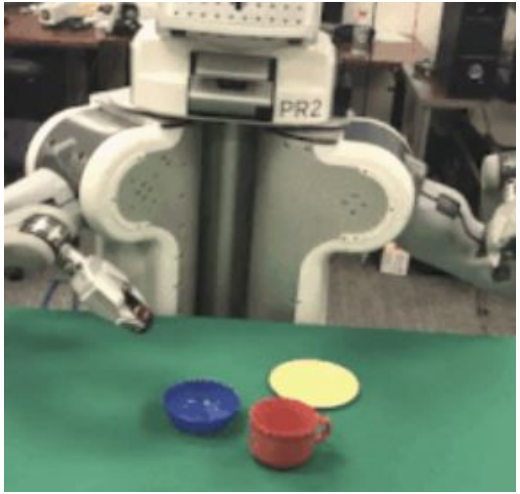




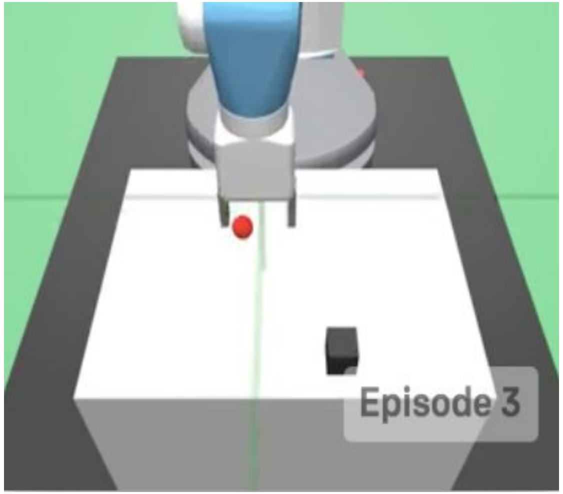
Algorithms

How to select an algorithm for your problem?

- Find similar task in “standard” problems



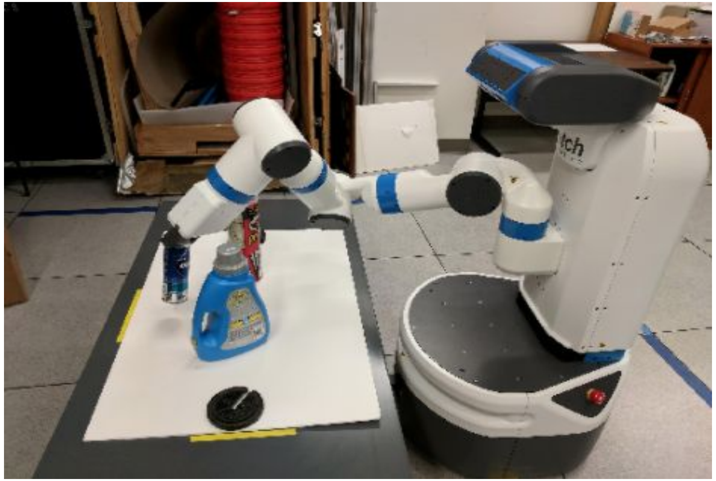
PR2 Robot (Huge robot, dual arm)



Fetch simulation in gym

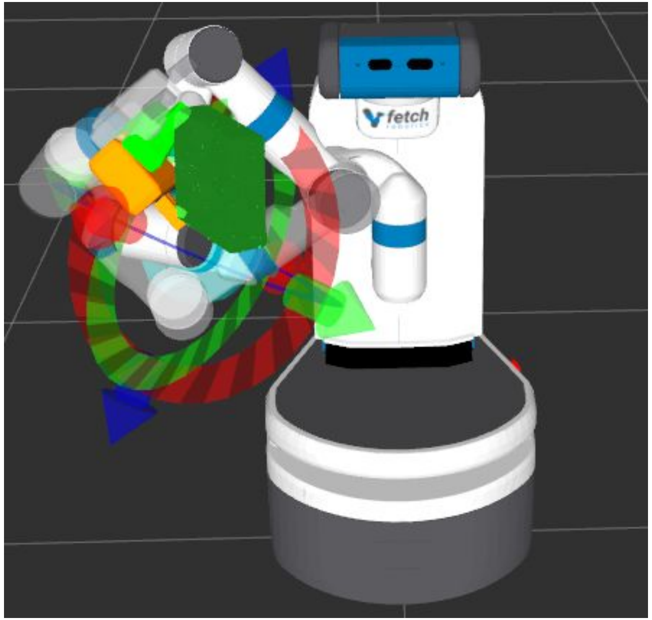
Environment

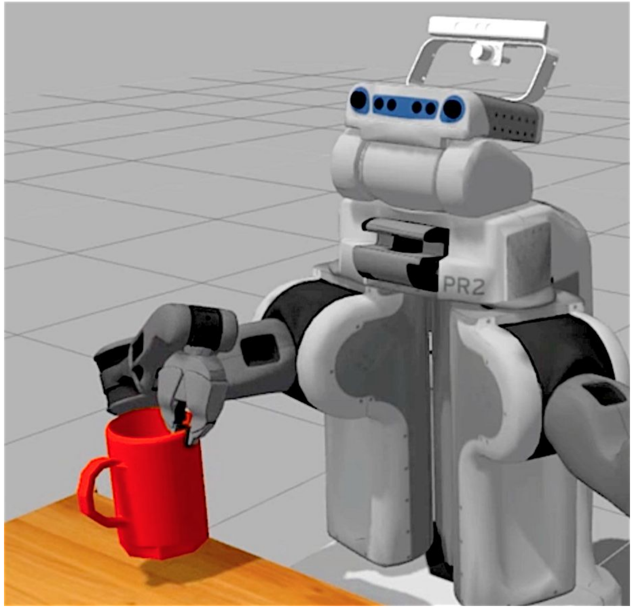
- Hardware robot
- Simulation



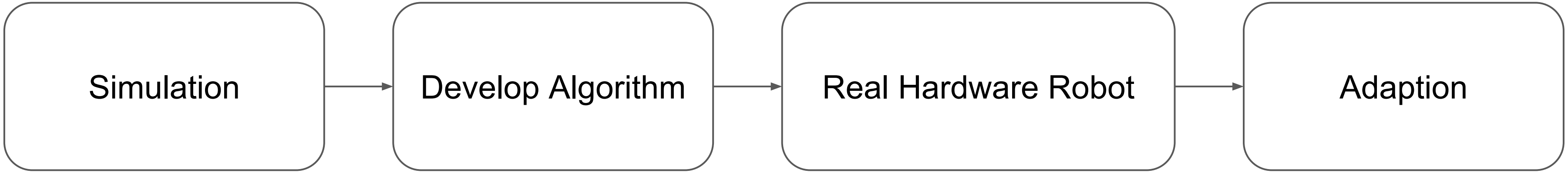
Environment

- Hardware robot
- Simulation





Environment



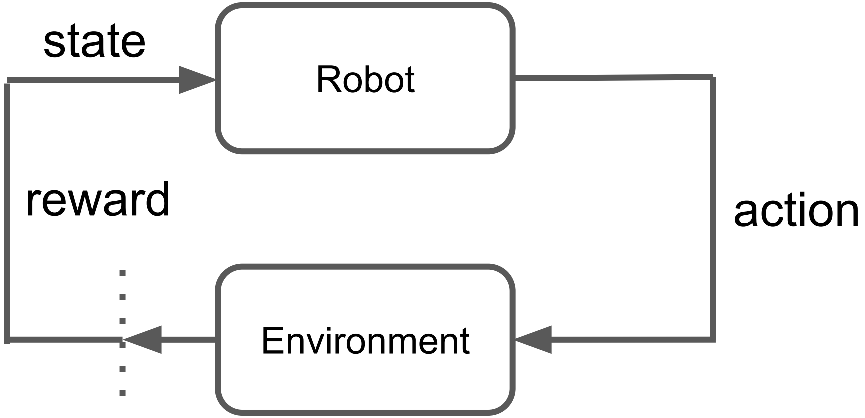
What is the problem of starting from real robot?

- **Expensive**
- **Safety**

Simulator

- Simulation Environment
  - OpenAI Gym
  - MuJoCo-py
  - PyBullet
  - Gazebo
  - V-rep
  - Roboschool
  - Dart
  - ....
- Dynamics Engine
  - Box2D
  - Bullet
  - ODE
  - ...

RL in Robotics: problems



- Reward?
- Structure?
- Exploration?
- Stability?

Future Directions

- Efficient RL
- Long Horizon Reasoning
- Hierarchical RL
- Meta-RL
- Reward Function
- Multi-model
- Lifelong Learning
- Simulation to Real
- ...

References

[1] https://gym.openai.com/

[2] Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." Nature 518.7540 (2015): 529.

[3] Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search." nature 529.7587 (2016): 484-489.

[4] Rajeswaran, Aravind, et al. "Learning complex dexterous manipulation with deep reinforcement learning and demonstrations." arXiv preprint arXiv:1709.10087 (2017).

[5] Roboschool: https://blog.openai.com/roboschool/

[6] Yan, Duan. “Meta Learning for Control.” PhD Thesis (2017).

[7] Gu, Shixiang, et al. "Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates." Robotics and Automation (ICRA), 2017 IEEE International Conference on. IEEE, 2017.

References

[8] Peng, Xue Bin, et al. "Deeploco: Dynamic locomotion skills using hierarchical deep reinforcement learning." ACM Transactions on Graphics (TOG) 36.4 (2017): 41.

[9] Bousmalis, Konstantinos, et al. "Using simulation and domain adaptation to improve efficiency of deep robotic grasping." arXiv preprint arXiv:1709.07857 (2017).

[10] Finn, Chelsea, et al. "One-shot visual imitation learning via meta-learning." arXiv preprint arXiv:1709.04905 (2017).

Thank you!