CSE599G: Deep Reinforcement Learning

University of Washington
Aravind Rajeswaran and Kendall Lowrey
Spring 2018

some slides in this lecture are borrowed from Sergey Levine

Course Logistics

- Lectures by Aravind Rajeswaran and Kendall Lowrey
- 2-3 guest lectures on related current research topics

- Course webpage: http://deeprl.cs.washington.edu/
- Slack: https://uwcse599g1.slack.com

- Class meets: Mon and Wed 2:00pm to 3:20pm at *THO134*
- Office hours: 10-11 Thursday and 11-12 Fridays
- We may use a different room/time for guest lectures

- Grading: 2 homeworks for total of 30%. Final project for 60%.
  10% for class participation, discussion etc.

Course Logistics

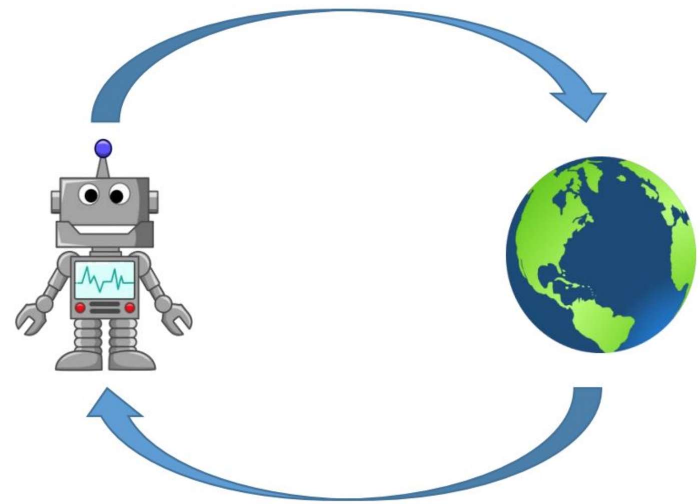
- **Prerequisites**
  - linear algebra and probability
  - machine learning, optimization
  - programming in python

- **Notes**
  - Attend recitation to setup the code infrastructure
  - A good laptop is sufficient for assignments
  - Talk to instructors if concerned about compute for project
  - Attend office hours and ask questions if something is too hard!

- Course treatment is intended to be fairly generic, but maybe a bit geared towards robotics for examples and assignments

What is Reinforcement Learning?

decisions (actions)



consequences
observations
rewards

Some examples

Actions: muscle contractions
Observations: sight, smell
Rewards: food

Actions: motor current or torque
Observations: camera images
Rewards: task success measure
(e.g., running speed)

Actions: what to purchase
Observations: inventory levels
Rewards: profit

What is Reinforcement Learning?

(modified version of Satinder Singh's description)

RL is about building algorithms that enable agents to make near-optimal decisions in complex, dynamic, uncertain, and potentially unknown environments.

- algorithms: provide the capability to solve problems as opposed to solutions
- near-optimal: notion of quality, cast problem as some form of optimization
- dynamic: decisions have long lived consequences; the world changes
- uncertain: exact state of the world is unknown; need to perform inference
- unknown: there is something unknown about the world; needs learning

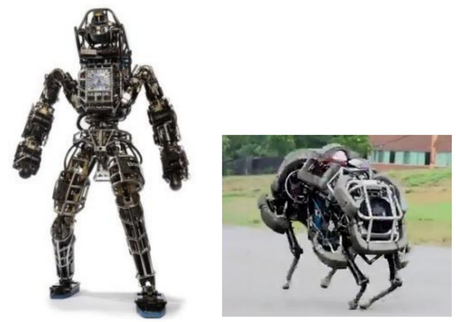
Basically, RL is AI complete, except maybe for game theory.

Why study Reinforcement Learning?

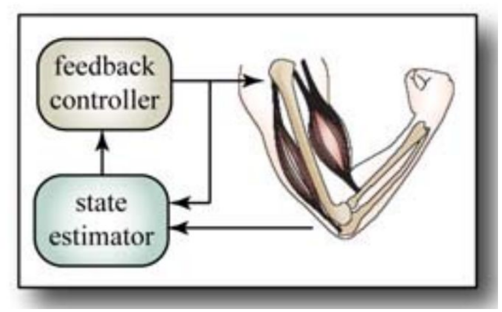
- **Algorithmic motivation:**
  Automatic iterative optimization based approach for finding near-optimal decision making rules

- **Scientific:**
  Understand how the brain makes decisions,
  help understand and address behavioral disorders

- **Applications:**
  Robotics, NLP, operations research, ad placement



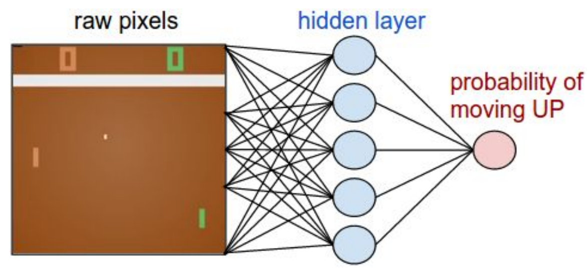
Dynamic walking robots



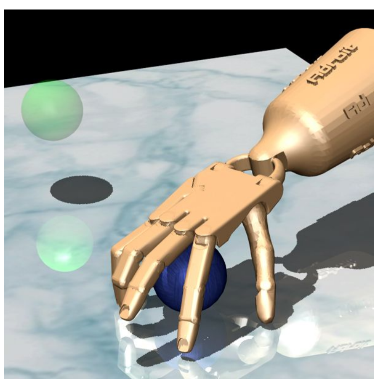
Optimal control in biology

Some deep RL success stories

AlphaGo:
Supervised learning + policy gradients + value functions + Monte Carlo tree search



Playing Atari Games:
Q-learning with convolutional networks

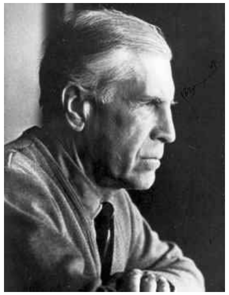
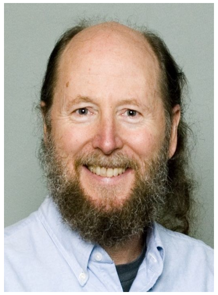


Dexterous manipulation:
Policy gradients + human demonstrations

Control from vision:
Guided Policy Search

Which Reinforcement Learning?

- Reinforcement learning the "problem statement" is same as optimal control and dynamic programming. Sequential decision making under uncertainty.
- Reinforcement learning the "method" is used synonymously with "model-free" RL. Techniques that work with only sampling access.
- Model-based RL is basically System ID + optimal control (in a loop)

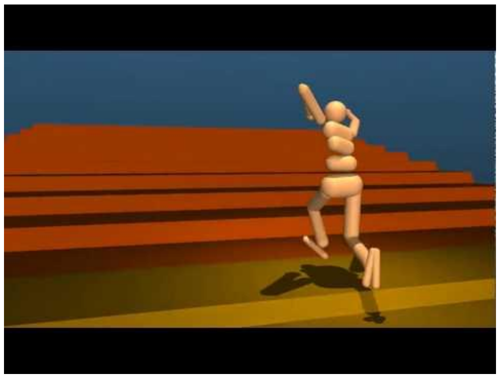
   

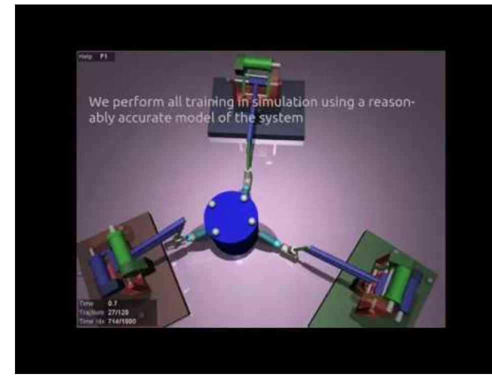
Course Plan

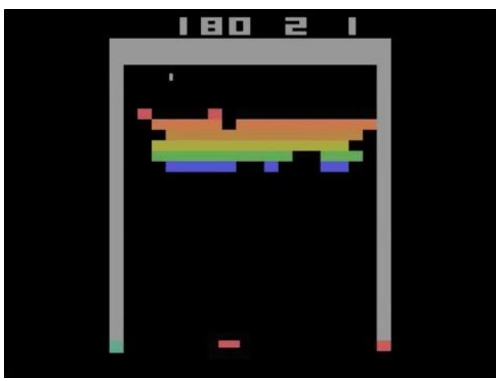
Broad perspective on Reinforcement Learning (as a problem statement/setting)

- Mathematical formalism through MDPs
- Value and policy iteration in the tabular case (dynamic programming)
- Model-free RL with function approx: evolution, policy gradient, Q-learning
- Model-based RL: MCTS, trajectory optimization, guided policy search
- Case study: AlphaGo and AlphaZero
- Other topics (survey): imitation, hierarchy, transfer, sim2real
- Invited talks from experts

Model-free Reinforcement Learning







Heess et al. Emergence of Locomotion Behaviours in Rich Environments, 2017

Rajeswaran et al. Learning Complex Dexterous Manipulation …, 2017

Lowrey et al. Reinforcement learning for non-prehensile manipulation…,, 2018

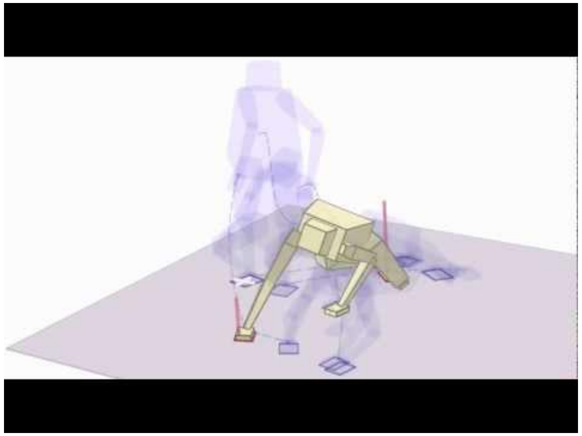
Mnih et al. Playing Atari with Deep Reinforcement Learning, 2013

Model-based Optimization

Tassa et al. IROS 2012

Erez et al. IROS 2012

Mordatch et al. SIGGRAPH 2012



Kumar et al. ICRA 2016

Imitation learning

Previous approaches require a reward function to shape behavior. Can we simply demonstrate the behavior we want and have the agent mimic this?

Maximum Margin Planning, Ratliff et al. 2006

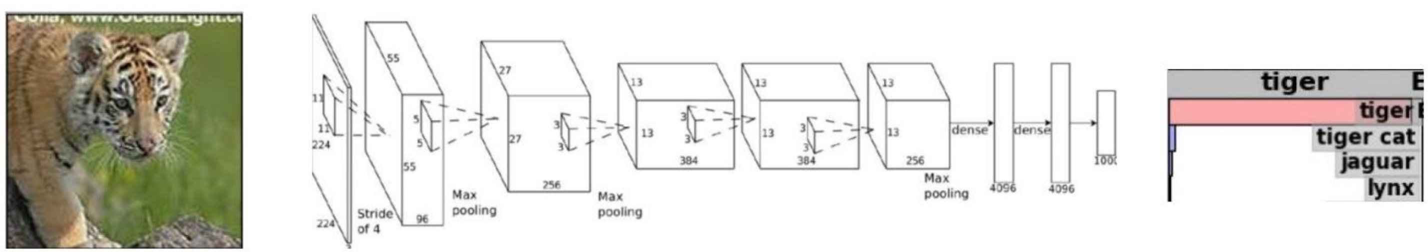
Ratliff et al. Maximum margin planning, 2006

Note: Not very universal, often times humans may not be able to demonstrate the desired behavior, and often times the goal is to exceed human capabilities.

Why “deep” RL?

Traditional paradigm is a modular, feature-based approach familiar in engineering

Deep Learning is an end-to-end approach



Why "deep" RL?

End-to-end approaches are clearly working better for tasks associated with perception.

A large part of the brain is involved in even extremely simple movements, suggesting a tight coupling between perception, internal model of the world, and motor control.

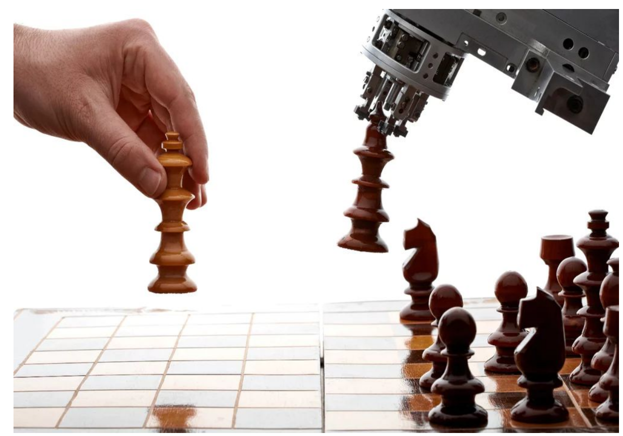
Complex animals learn most of their behavioral repertoire (instead of using genetically defined control strategies), thus their brain needs to have not only control machinery, but also machinery that builds control machinery (chocolate box vs chocolate factory).

It would be nice (in practice, understanding, and pedagogy) to have a **single** learning algorithm to replicate the functionality of the brain.

Why “deep” RL?

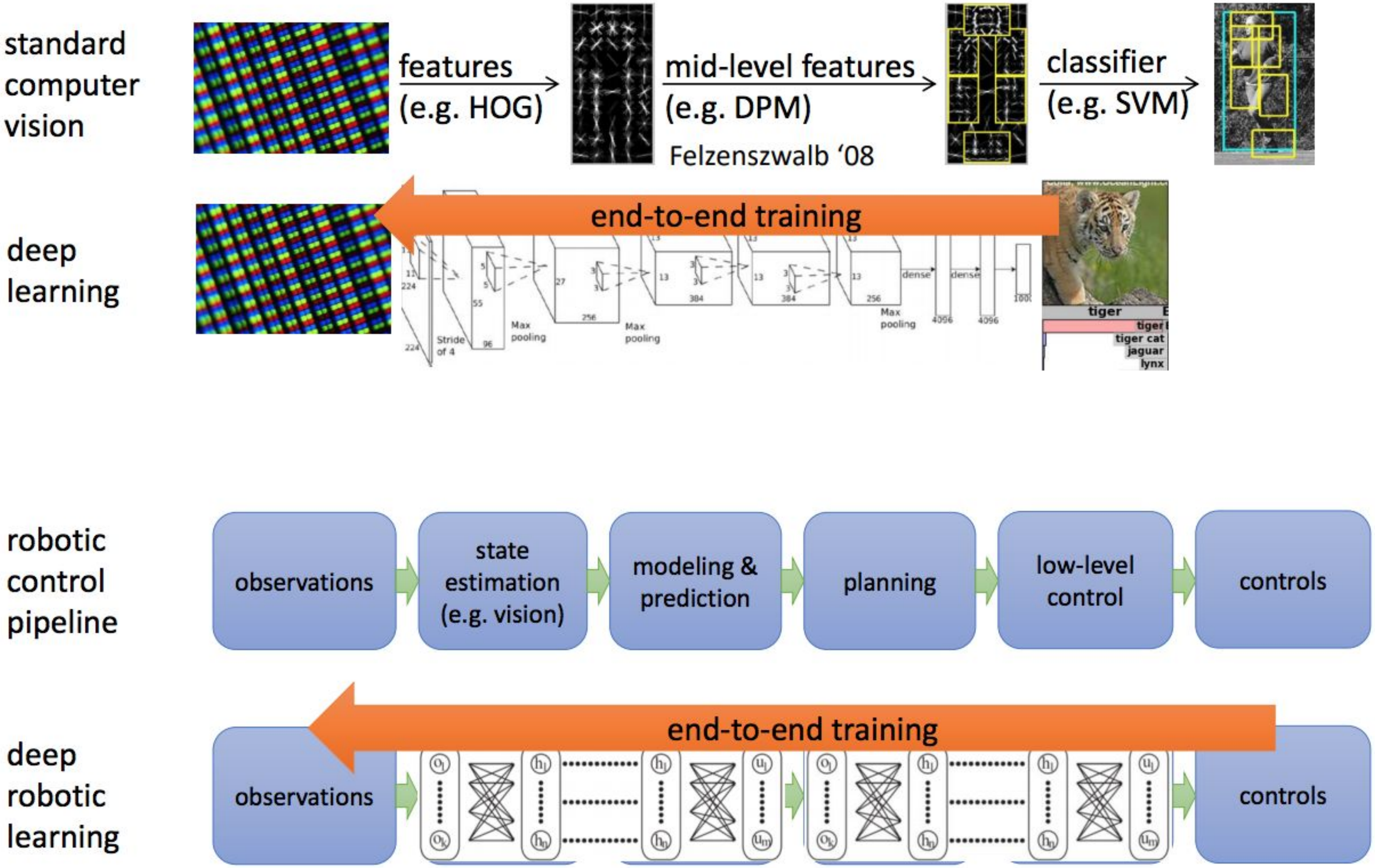
What must a single learning algorithm be capable of?

- Interpret rich sensory inputs (vision, speech, tactile etc.)
- Be capable of making complex decisions : from low level muscle movements (pull muscles for walking) to high level plans (moves in a chess game)



Combining deep learning with reinforcement learning is a step towards this direction

What does deep RL look like?



Assignments & Project

- **Homework 1 :** Policy gradients on MuJoCo (robotics simulator)
- **Homework 2 :** RL with convolutional networks for Atari games
- **Project :** Good research project related to RL (max group size is 3).
  Proposal : Apr 16 Milestone : May 7 Final report : June 6
  Start early and talk to us regularly!
- **Course outcomes:**
  a. Be proficient enough to understand RL and do basic research.
  b. Good chance that some projects will become conference papers.
  c. Conference on Robot Learning (CoRL) deadline is **June 15th.**

Word of caution

Deep RL (the method) is very appealing for a number of reasons, but it is not magic!

- Model-free RL is generic, but generality comes at the price of efficiency.
- There may be better domain specific approaches to solve a problem.
- “Deep” representations may not always be needed.
- A very interesting research area with a lot of activity, but important to be grounded and apply a discount factor :)