CS 188: Artificial Intelligence

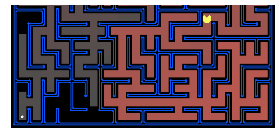
Advanced Applications: Robotics**

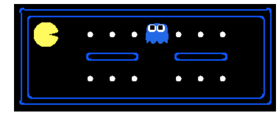
Instructors: Pieter Abbeel & Dan Klein --- University of California, Berkeley

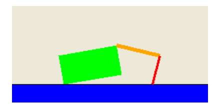
These slides were created by Dan Klein, Pieter Abbeel and Anca Dragan for CS188 Intro to AI at UC Berkeley.

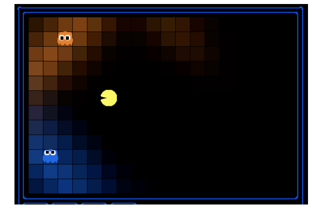
All CS188 materials are available at http://ai.berkeley.edu

So Far: Foundational Methods

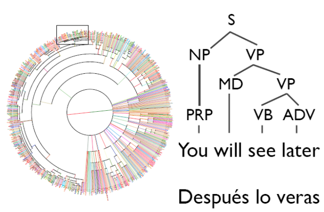




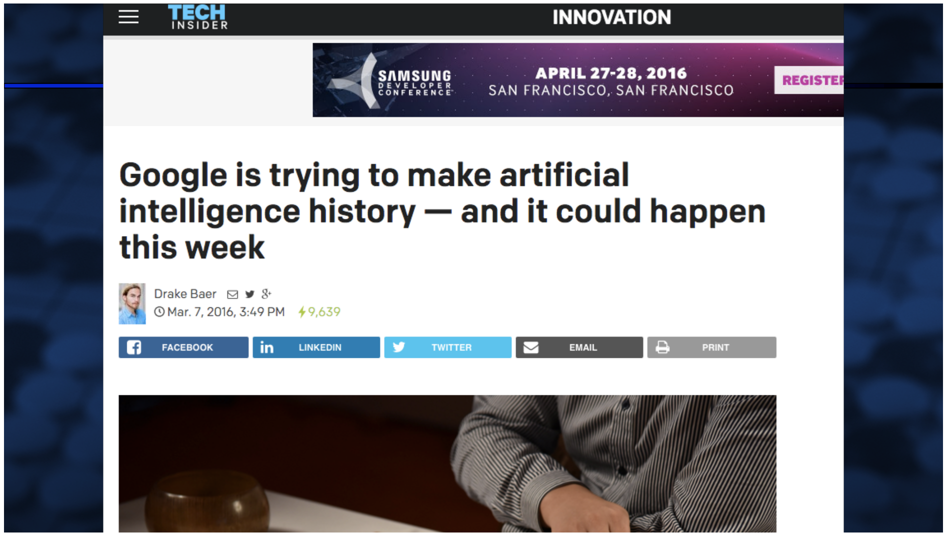


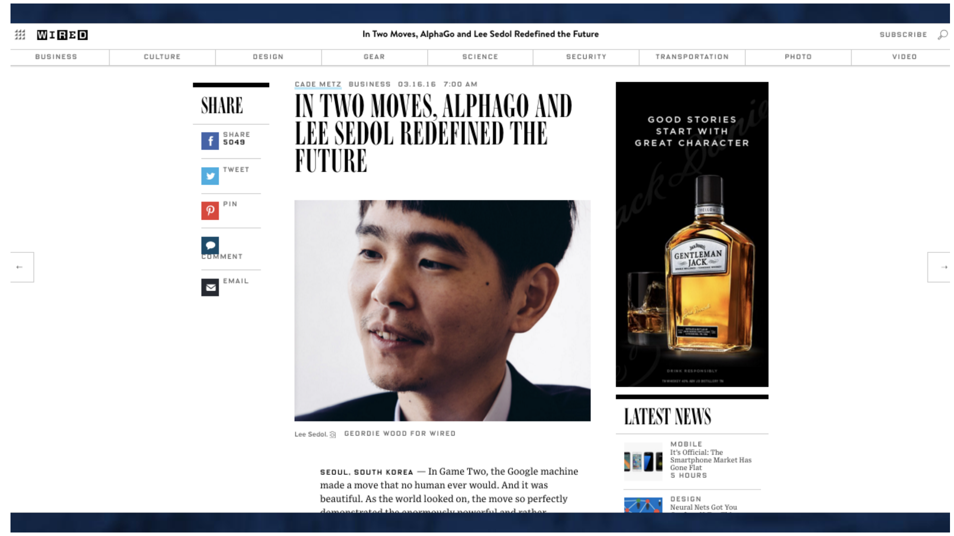


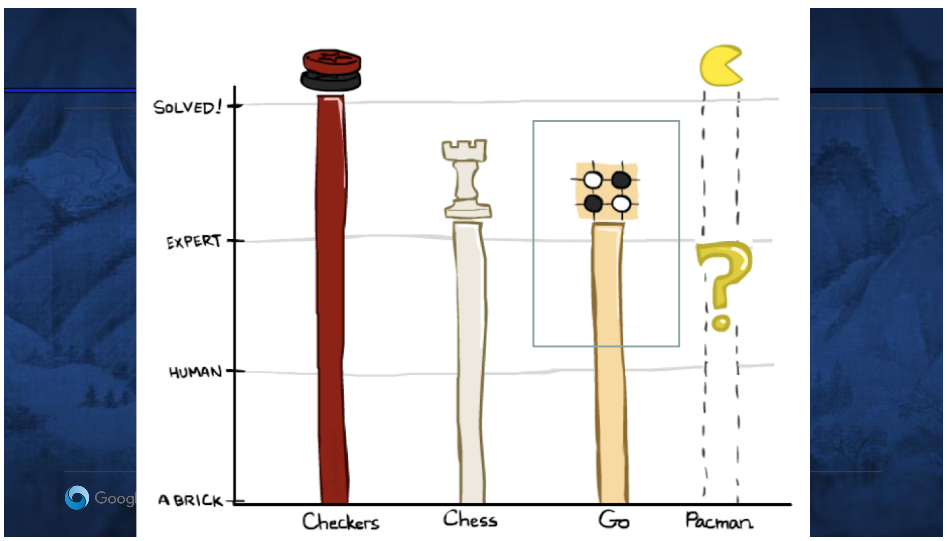
Now: Advanced Applications



AlphaGo

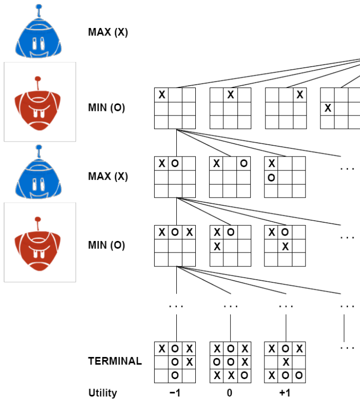


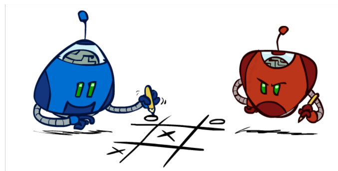




How would you make an AI for Go?

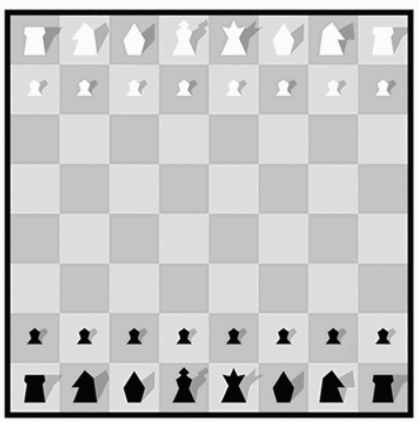
MiniMax!

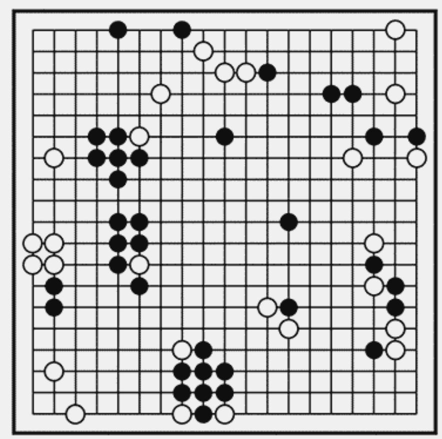




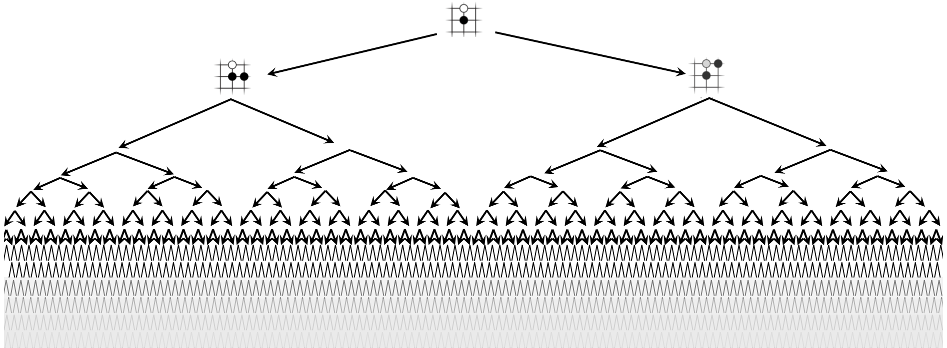
Why is it hard?

In particular, why is it harder than chess?

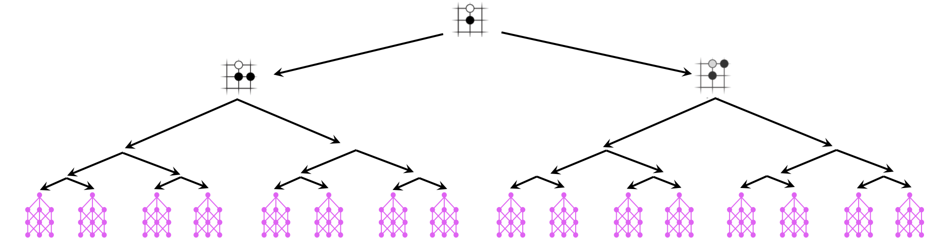




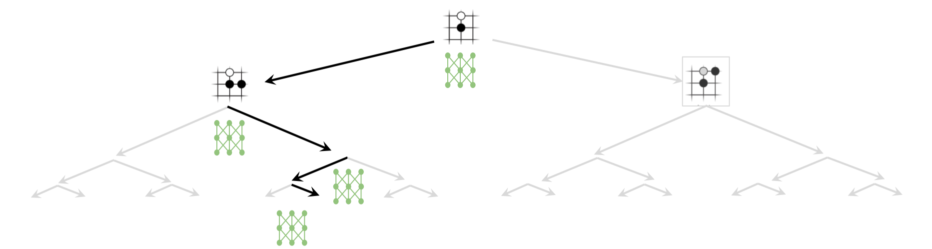
Exhaustive search



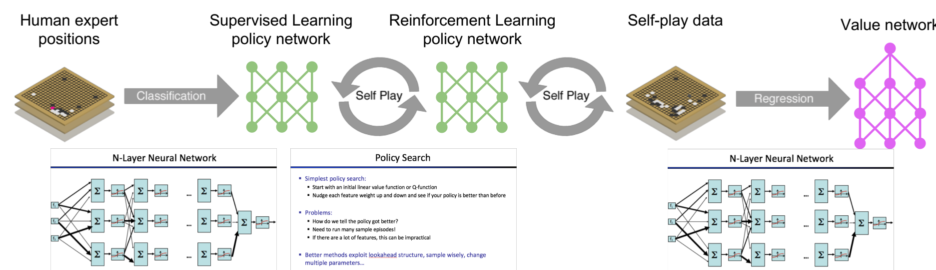
Reducing depth with value network



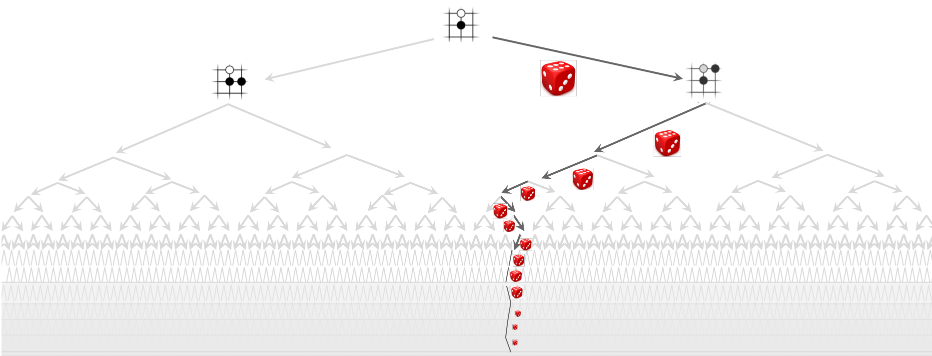
Reducing breadth with policy network

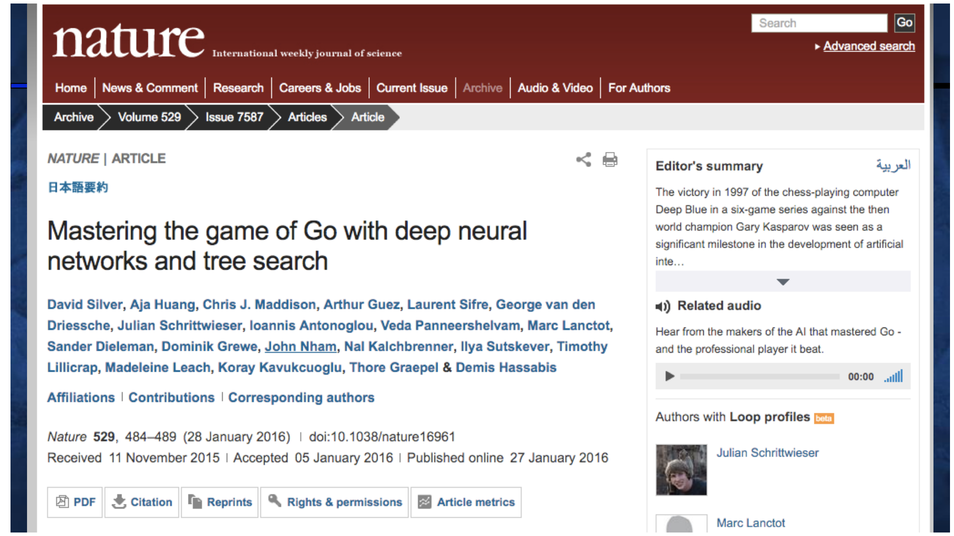


Neural network training pipeline



One more thing: Monte-Carlo rollouts





nature International weekly journal of science

Search Go
Advanced search

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 529 > Issue 7587 > Articles > Article

NATURE | ARTICLE

日本語要約

Mastering the game of Go with deep neural networks and tree search

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel & Demis Hassabis

Affiliations | Contributions | Corresponding authors

Nature **529**, 484–489 (28 January 2016) | doi:10.1038/nature16961
Received 11 November 2015 | Accepted 05 January 2016 | Published online 27 January 2016

PDF | Citation | Reprints | Rights & permissions | Article metrics

Editor's summary العربية

The victory in 1997 of the chess-playing computer Deep Blue in a six-game series against the then world champion Gary Kasparov was seen as a significant milestone in the development of artificial inte…

Related audio

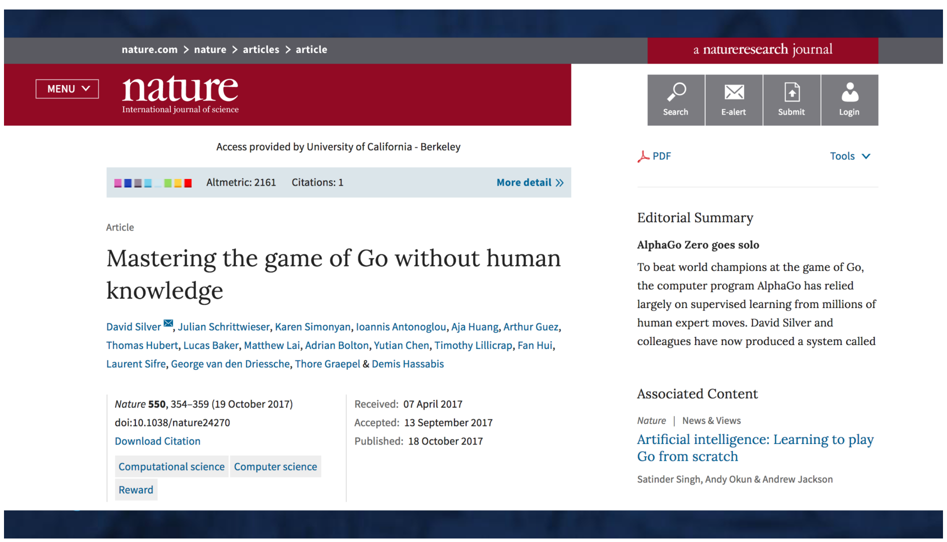
Hear from the makers of the AI that mastered Go - and the professional player it beat.

00:00

Authors with Loop profiles beta

Julian Schrittwieser

Marc Lanctot



nature.com > nature > articles > article

a natureresearch journal

MENU

nature International journal of science

Search | E-alert | Submit | Login

Access provided by University of California - Berkeley

Altmetric: 2161 Citations: 1 More detail

PDF Tools

Article

Mastering the game of Go without human knowledge

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel & Demis Hassabis

Nature **550**, 354–359 (19 October 2017)
doi:10.1038/nature24270
Download Citation

Received: 07 April 2017
Accepted: 13 September 2017
Published: 18 October 2017

Computational science | Computer science | Reward

Editorial Summary

AlphaGo Zero goes solo

To beat world champions at the game of Go, the computer program AlphaGo has relied largely on supervised learning from millions of human expert moves. David Silver and colleagues have now produced a system called

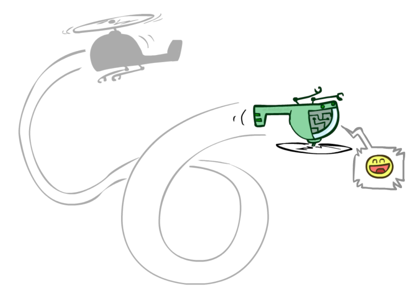
Associated Content

Nature | News & Views

Artificial intelligence: Learning to play Go from scratch

Satinder Singh, Andy Okun & Andrew Jackson

Robotic Helicopters



Motivating Example

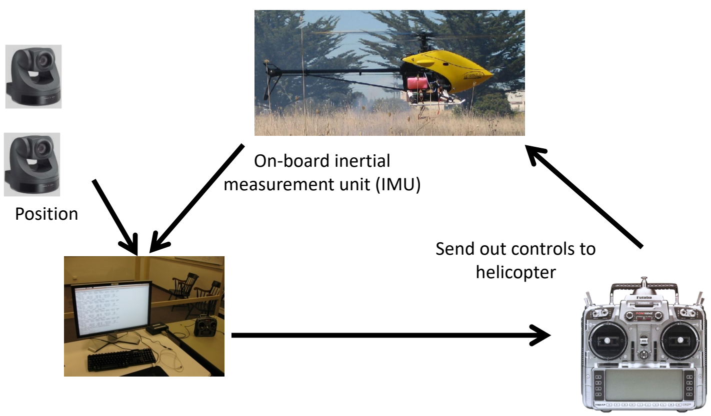
- How do we execute a task like this?

[VIDEO: tictoc_results.wmv]

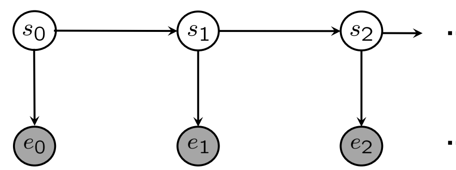
Autonomous Helicopter Flight

- Key challenges:
  - Track helicopter position and orientation during flight
  - Decide on control inputs to send to helicopter

Autonomous Helicopter Setup



HMM for Tracking the Helicopter



- State: $s = (x, y, z, \phi, \theta, \psi, \dot{x}, \dot{y}, \dot{z}, \dot{\phi}, \dot{\theta}, \dot{\psi})$
- Measurements: [observation update]
  - 3-D coordinates from vision, 3-axis magnetometer, 3-axis gyro, 3-axis accelerometer
- Transitions (dynamics): [time elapse update]
  - $s_{t+1} = f(s_t, a_t) + w_t$ f: encodes helicopter dynamics, w: noise

Helicopter MDP

- State: $s = (x, y, z, \phi, \theta, \psi, \dot{x}, \dot{y}, \dot{z}, \dot{\phi}, \dot{\theta}, \dot{\psi})$
- Actions (control inputs):
  - a_{lon} : Main rotor longitudinal cyclic pitch control (affects pitch rate)
  - a_{lat} : Main rotor latitudinal cyclic pitch control (affects roll rate)
  - a_{coll} : Main rotor collective pitch (affects main rotor thrust)
  - a_{rud} : Tail rotor collective pitch (affects tail rotor thrust)

- Transitions (dynamics):
  - $s_{t+1} = f(s_t, a_t) + w_t$

  [f encodes helicopter dynamics]

  [w is a probabilistic noise model]

- Can we solve the MDP yet?

Problem: What's the Reward?

- Reward for hovering:

$$\begin{aligned} R(s) \quad = \quad & -\alpha_x (x - x^*)^2 \\ & -\alpha_y (y - y^*)^2 \\ & -\alpha_z (z - z^*)^2 \\ & -\alpha_{\dot{x}} \dot{x}^2 \\ & -\alpha_{\dot{y}} \dot{y}^2 \\ & -\alpha_{\dot{z}} \dot{z}^2 \end{aligned}$$

RL: Helicopter Flight

[Andrew Ng]

[Video: HELICOPTER]

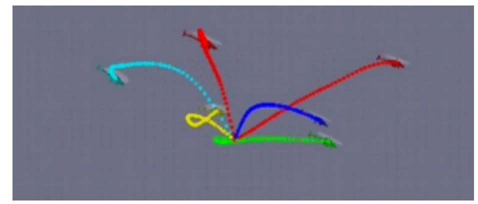
Problem for More General Case: What's the Reward?

- Rewards for "Flip"?
  - Problem: what's the target trajectory?
  - Just write it down by hand?

Flips (?)

[VIDEO: 20061204---bad.wmv]

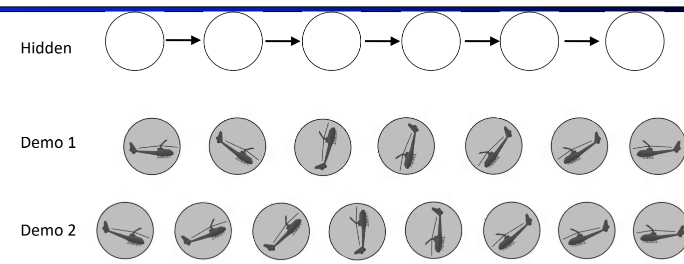
Helicopter Apprenticeship?



[VIDEO: airshow_unaligned.wmv]

Demonstrations

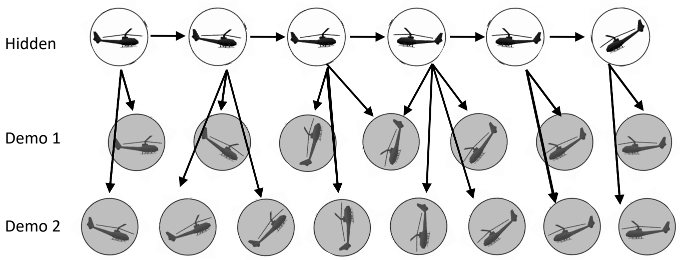
Learning a Trajectory



- HMM-like generative model
  - Dynamics model used as HMM transition model
  - Demos are observations of hidden trajectory
- Problem: how do we align observations to hidden trajectory?

Abbeel, Coates, Ng, IJRR 2010

Probabilistic Alignment using a Bayes' Net



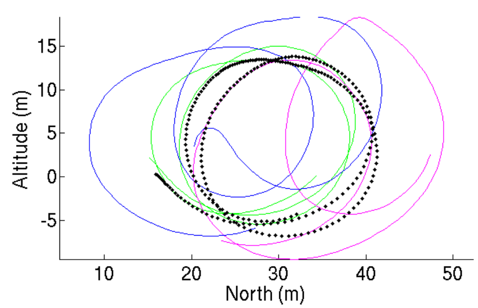
- Dynamic Time Warping
  (Needleman&Wunsch 1970, Sakoe&Chiba, 1978)
- Extended Kalman filter / smoother

Abbeel, Coates, Ng, IJRR 2010

Aligned Demonstrations

[VIDEO: airshow_unaligned.wmv]

Alignment of Samples



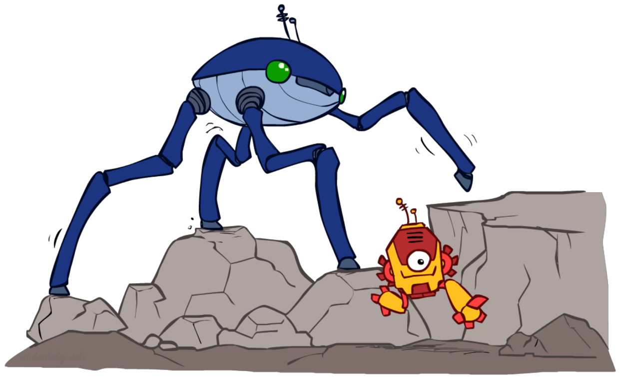
- Result: inferred sequence is much cleaner!

Learned Behavior

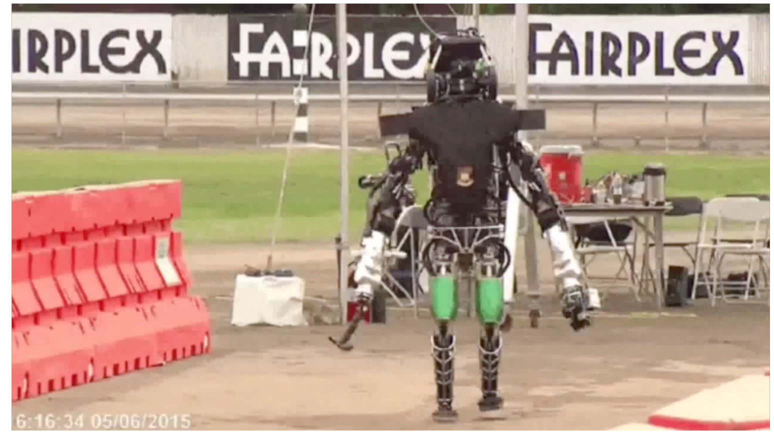
[VIDEO: airshow_trimmed.wmv]

[Abbeel, Coates, Quigley, Ng, 2010]

Legged Locomotion

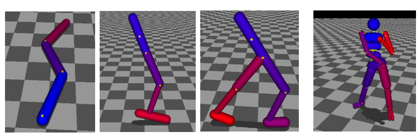


For Perspective: Darpa Robotics Challenge (2015)



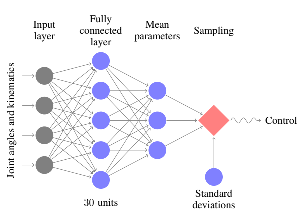
How About Continuous Control, e.g., Locomotion?

Robot models in physics simulator
(MuJoCo, from Emo Todorov)

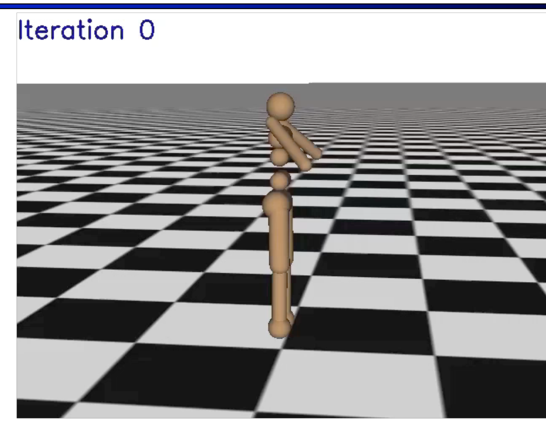


Input: joint angles and velocities
Output: joint torques

Neural network architecture:



Learning Locomotion



[Schulman, Moritz, Levine, Jordan, Abbeel, 2015]

Deep RL: Virtual Stuntman

[Peng, Abbeel, Levine, van de Panne, 2018]

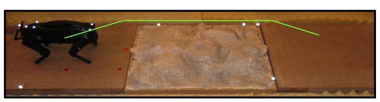
Quadruped

- Low-level control problem: moving a foot into a new location
  → search with successor function ~ moving the motors
- High-level control problem: where should we place the feet?
  - Reward function R(x) = w . f(s)   [25 features]

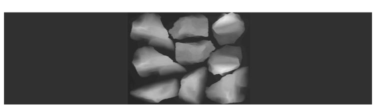
[Kolter, Abbeel & Ng, 2008]

Reward Learning + Reinforcement Learning

- Demonstrate path across the "training terrain"



- Learn the reward function
- Receive "testing terrain"---height map.



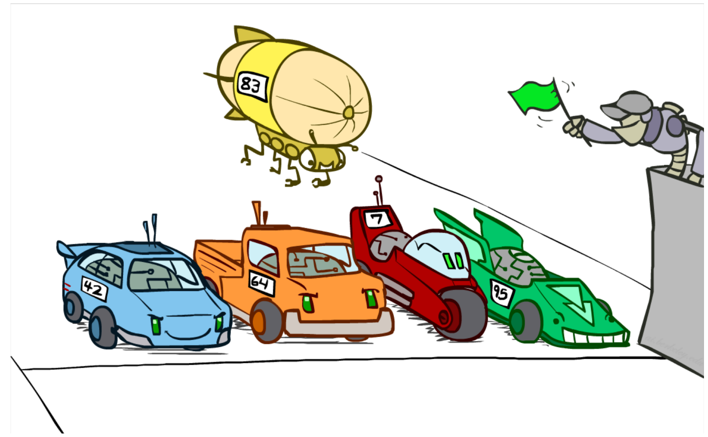
- Find the optimal policy with respect to the *learned reward function* for crossing the testing terrain.

[Kolter, Abbeel & Ng, 2008]

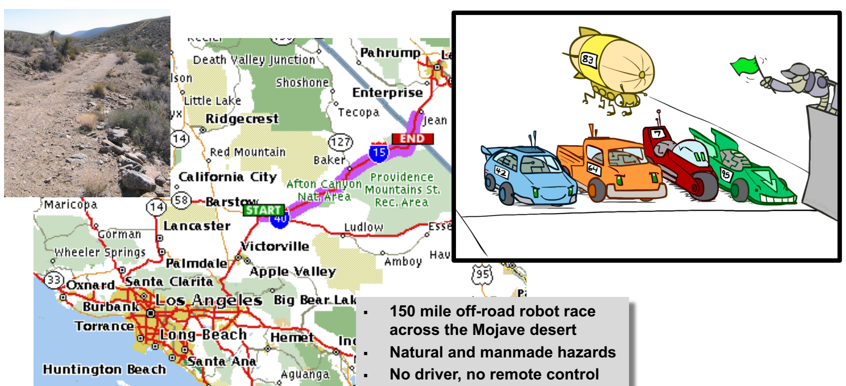
Without reward learning

With reward learning

Autonomous Driving



Grand Challenge 2005: Barstow, CA, to Primm, NV



- 150 mile off-road robot race across the Mojave desert
- Natural and manmade hazards
- No driver, no remote control
- No dynamic passing

Autonomous Vehicles

Autonomous vehicle slides adapted from Sebastian Thrun

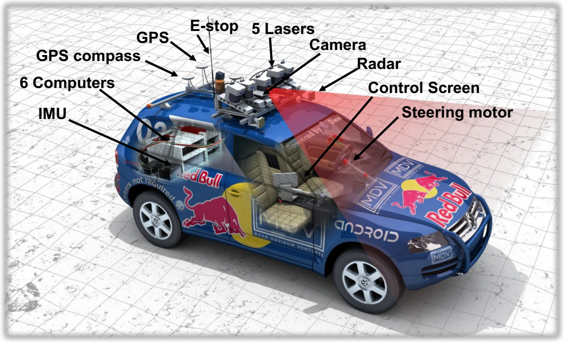
Grand Challenge 2005 Nova Video

[VIDEO: nova-race-supershort.mp4]

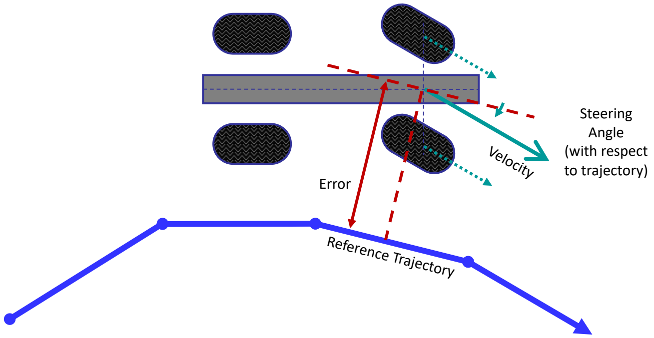
Grand Challenge 2005 – Bad

[VIDEO: grand challenge – bad.wmv]

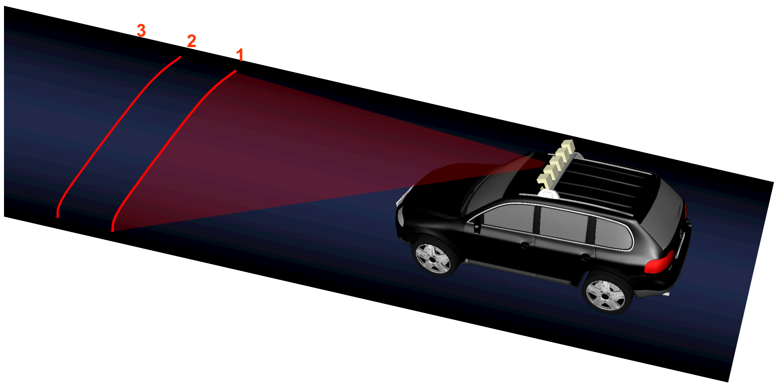
An Autonomous Car



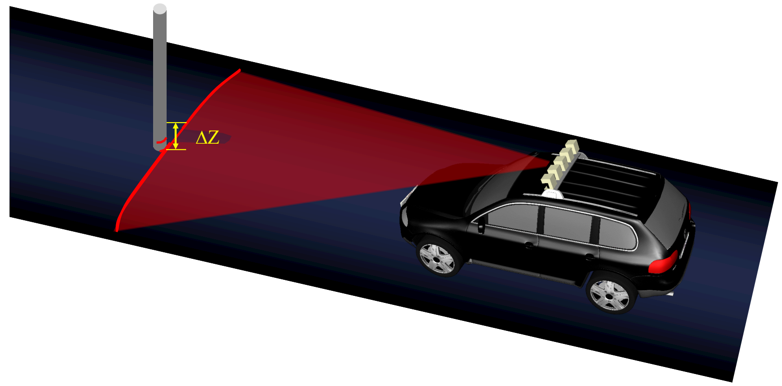
Actions: Steering Control



Laser Readings for Flat / Empty Road

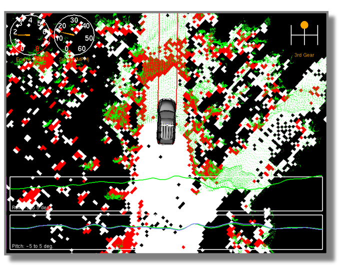


Laser Readings for Road with Obstacle



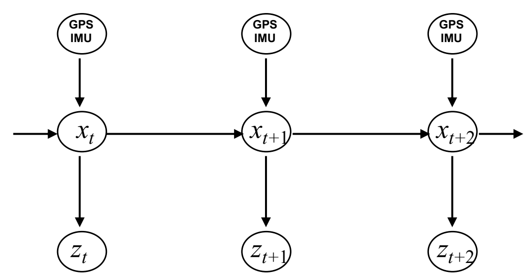
Obstacle Detection

Trigger if $|Z^i - Z^j| > 15\text{cm}$ for nearby z^i, z^j

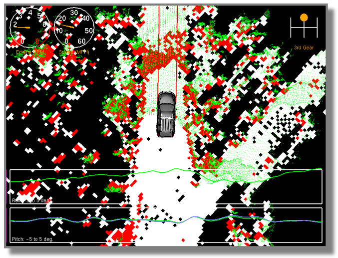


Raw Measurements: 12.6% false positives

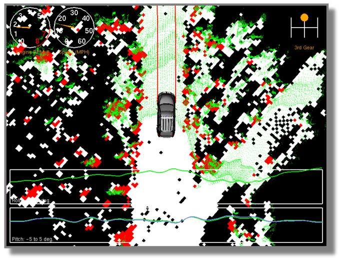
Probabilistic Error Model



HMMs for Detection

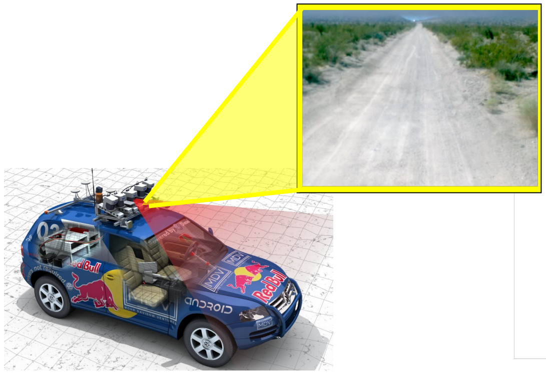


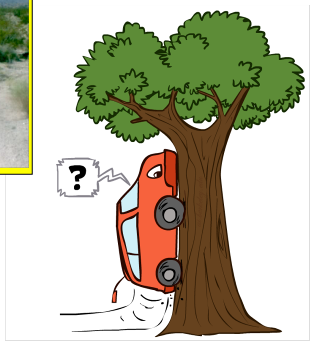
Raw Measurements: 12.6% false positives



HMM Inference: 0.02% false positives

Sensors: Camera





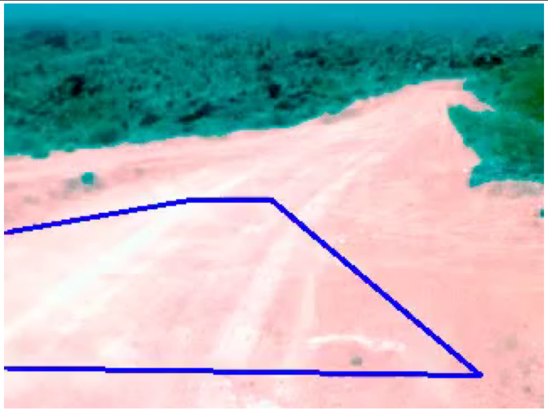
Vision for a Car

[VIDEO: lidar vision for a car]

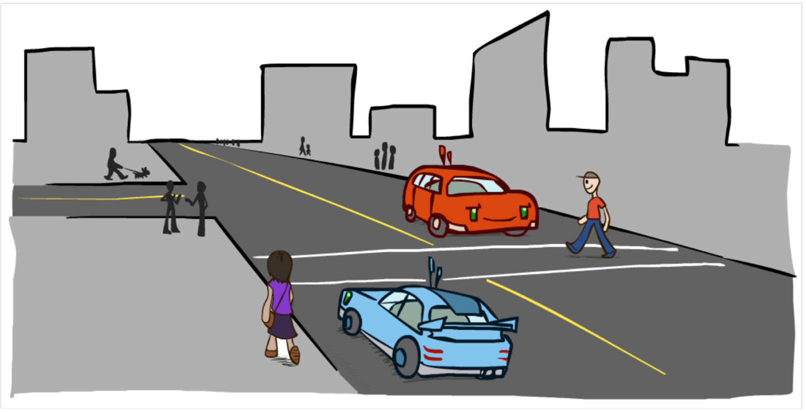
Vision for a Car

[VIDEO: self-supervised vision]

Self-Supervised Vision



Urban Environments



[VIDEO: ROBOTICS – gcar.m4v]

Google Self-Driving Car (2013)

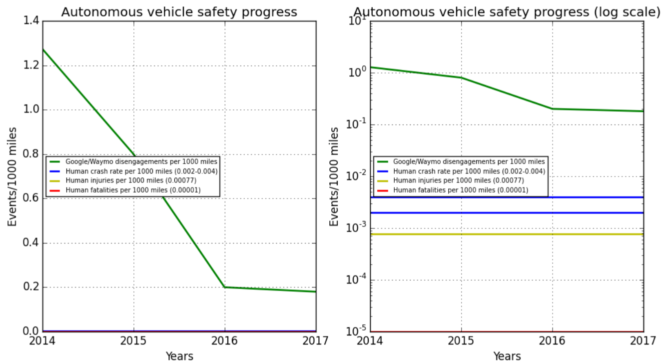
(mostly lidar)

Recent Progress: NN Semantic Scene Segmentation

~ neural net classifies every pixel

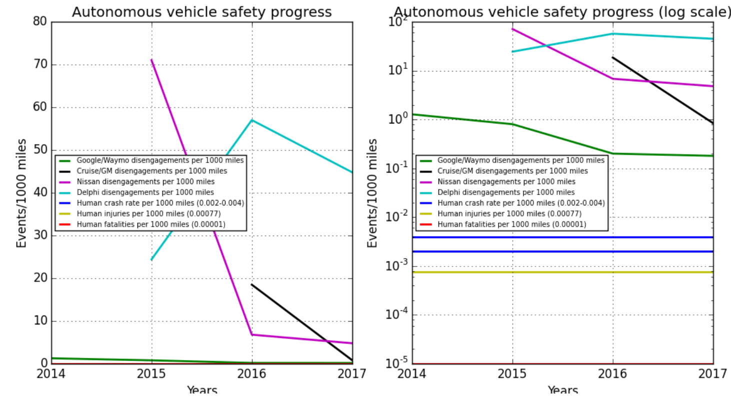
PSPNet50

Self-Driving Cars -- Stats



Pieter Abbeel -- UC Berkeley | Gradescope | Covariant.AI

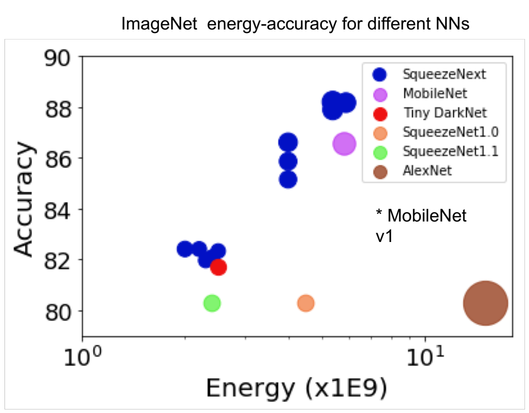
Self-Driving Cars -- Stats



Energy-Inference-Accuracy Landscape on the Squeezelator

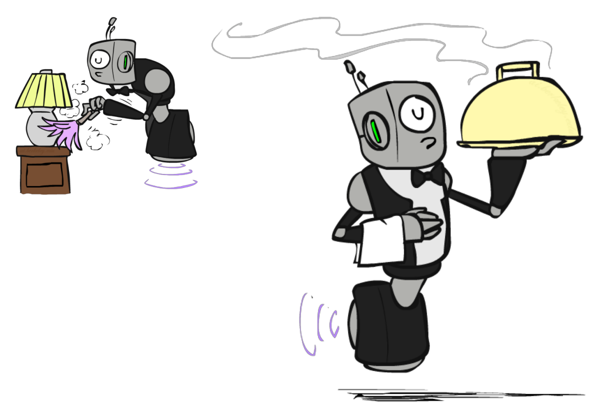
SqueezeNext vs SqueezeNet/AlexNet

- 8% more accurate
- 2.25x better than SqueezeNet
- 7.5x better than AlexNet



[slide credit: Kurt Keutzer]

Personal Robotics

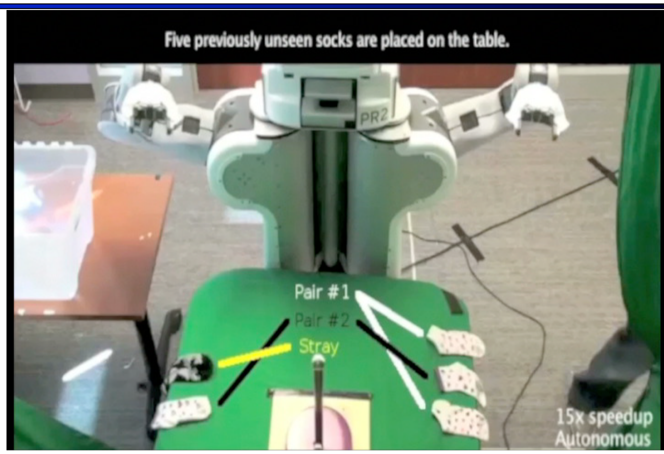


PR-1

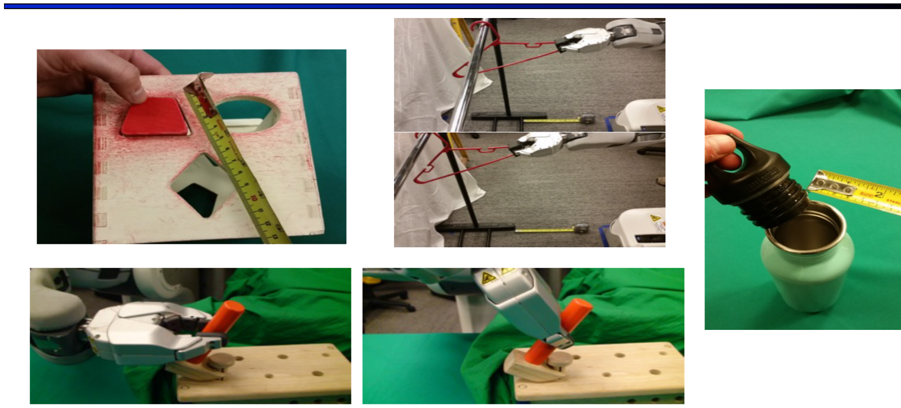
[Wyrobek, Berger, van der Loos, Salisbury, ICRA 2008]

Challenge Task: Robotic Laundry

Sock Sorting



How about a range of skills?



[Levine*, Finn*, Darrell, Abbeel, JMLR 2016]

Reinforcement Learning

[Levine*, Finn*, Darrell, Abbeel, JMLR 2016]

Learned Skills

Pieter Abbeel -- UC Berkeley / OpenAI / Gradescope

[Levine*, Finn*, Darrell, Abbeel, JMLR 2016]

[Levine et al, 2016]

Next Time:

- Natural language processing
- Final contest
- Course wrap-up
- Where to go next