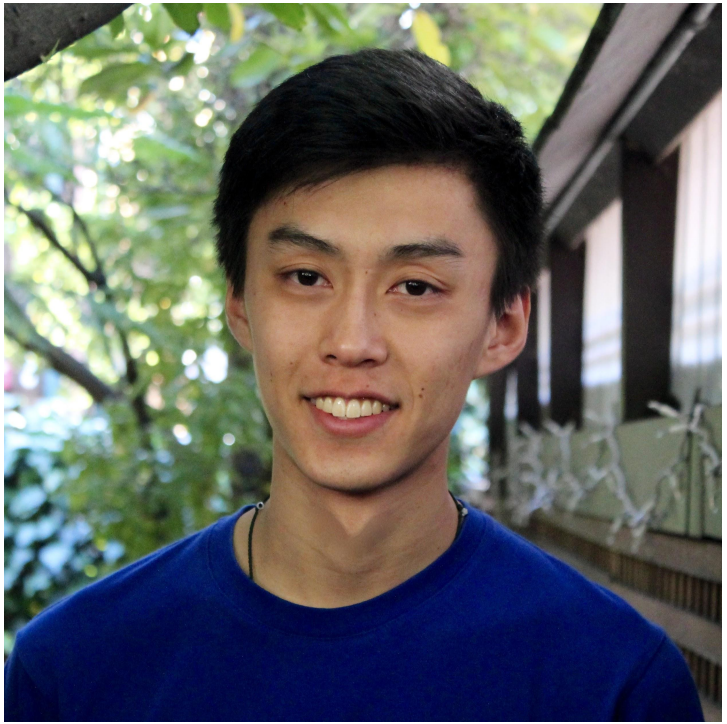
Marvin Zhang

- Second year PhD student at UC Berkeley, working with Prof. Pieter Abbeel and Prof. Sergey Levine
- Interests: model-based RL, representation learning
- Today I'll talk about combining model-based and model-free RL [1]



[1] Chebotar*, Hausman*, Zhang*, Sukhatme, Schaal, Levine. "Combining Model-Based and Model-Free Updates for Trajectory-Centric Reinforcement Learning." ICML 2017.

Motivation

Model-based reinforcement learning

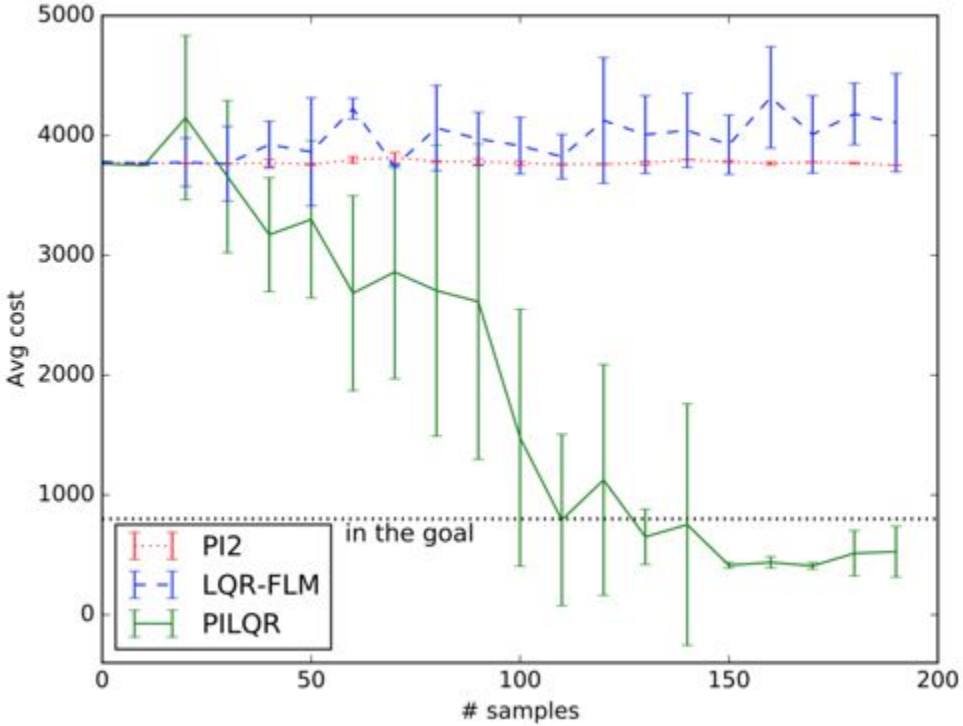
+ Sample efficiency: learn from scratch with a small number of trials
– Modeling bias: complex dynamics and costs can cause learning to fail

Model-free reinforcement learning

+ Can handle systems with arbitrarily complex dynamics and costs
– Significantly less sample-efficient

Can we integrate the two approaches?

Experiments: Hockey on a Real Robot



Sandy Huang

Fifth year Ph.D. student at UC Berkeley

Working with Pieter Abbeel and Anca Dragan

Interested in developing methods that shed light on what a model or agent has learned

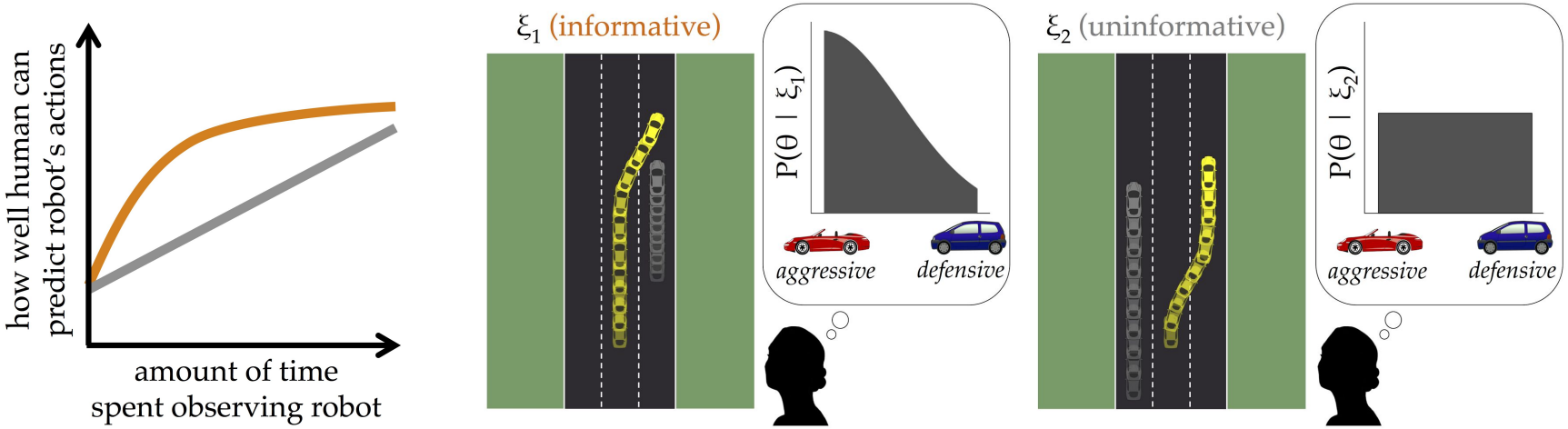
Essential for:

trust, human-robot interaction, safety

Teaching humans about how a robot acts

Goal: Help users quickly learn to anticipate a robot's behavior in novel situations

Approach: Select the most *informative* examples of the robot's behavior to show



Sandy Huang, David Held, Pieter Abbeel, Anca Dragan. Enabling robots to communicate their objectives. RSS 2017.

Black-box & dormant adversarial attacks on policies

black-box attack

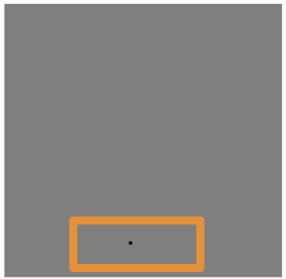
Chopper Command

original input

action taken: **down-left**
avoids collision

+

perturbation



$\operatorname{argmax}_i \nabla_x J(\theta, x, y)_i$

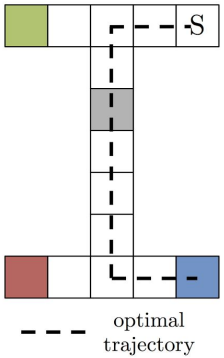
=

adversarial input

action taken: **left**
collides with enemy

dormant attack

ViZDoom I-Maze

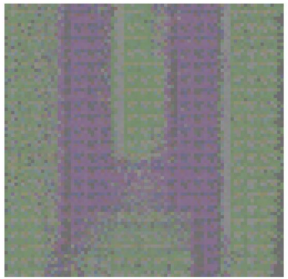


original input

action at end of hall: **left**
reaches correct goal (blue)

+ .1 ×

perturbation



=

adversarial input

action at end of hall: **right**
reaches wrong goal (red)

Sandy Huang, Nicolas Papernot, Ian Goodfellow, Yan Duan, Pieter Abbeel. Adversarial attacks on neural network policies. Under submission.

Rohin Shah

Fourth year Ph.D. student at UC Berkeley

Interested in CS education

Applying techniques from Programming Languages in order to speed up inference for probabilistic programming models

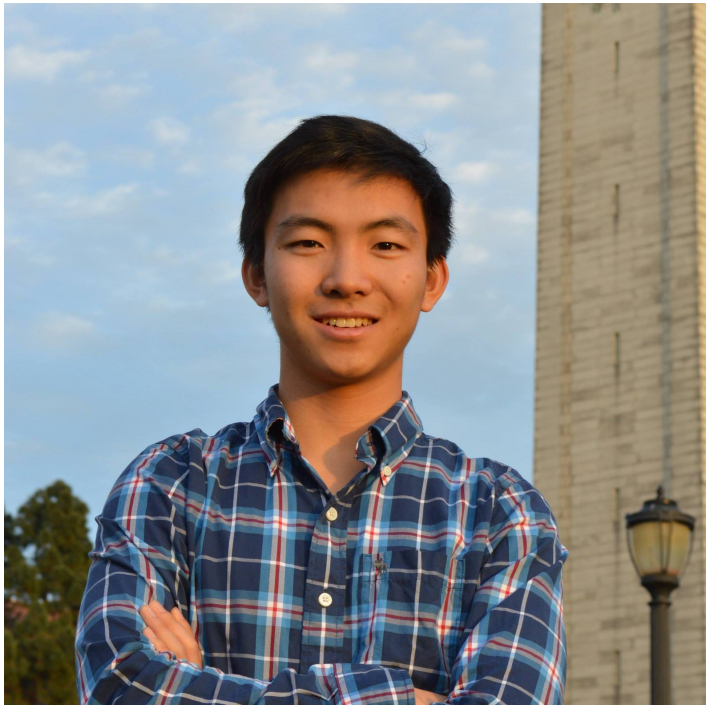
Looking to switch into AI/ML

Andrew Liu

First year Masters student at UC Berkeley

Learning robot behaviors from human video demonstrations [1]

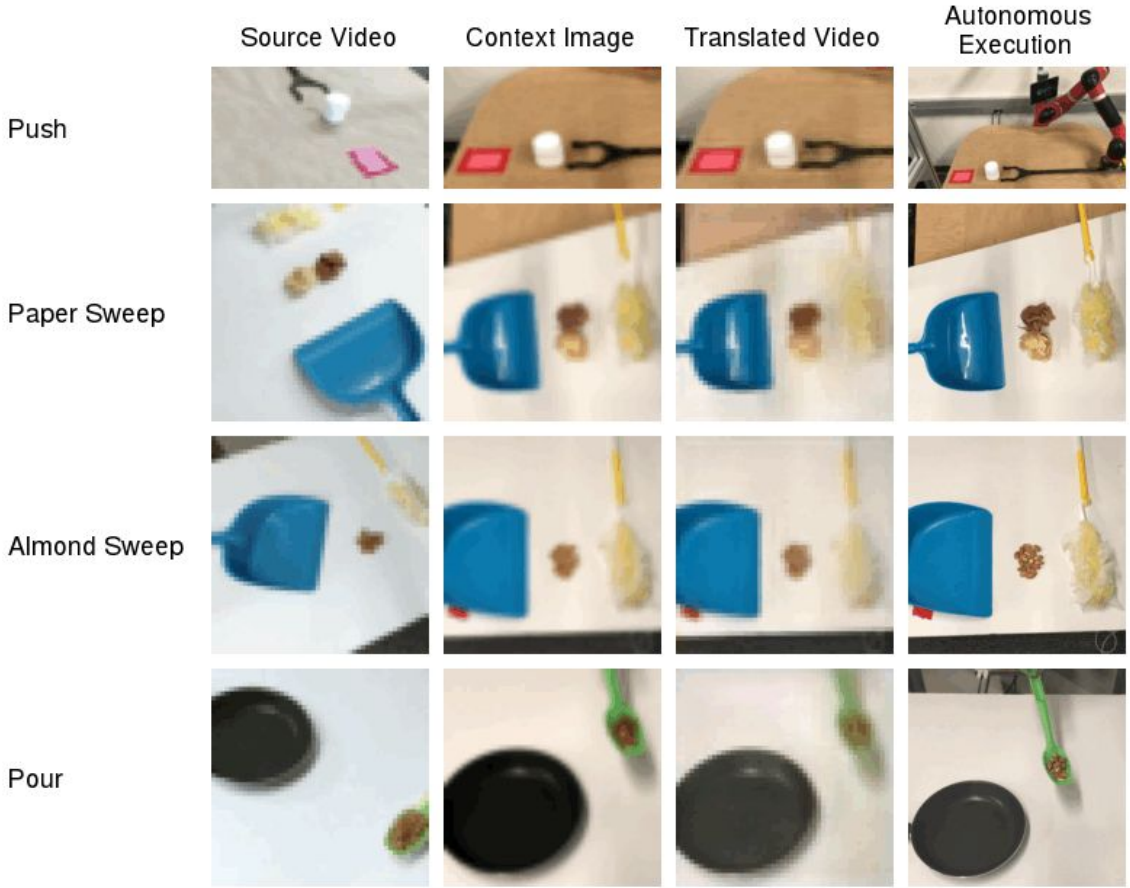
Transferring policies between robots with different physical configurations [2]



[1] YuXuan Liu*, Abhishek Gupta*, Pieter Abbeel, Sergey Levine. Imitation from Observation: Learning to Imitate Behaviors from Raw Video via Context Translation.

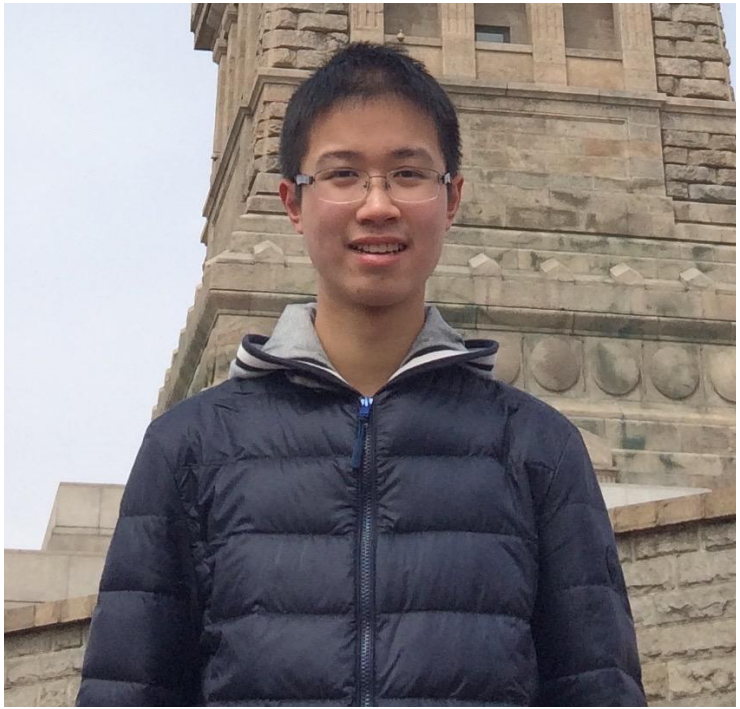
[2] Abhishek Gupta*, Coline Devin*, YuXuan Liu, Pieter Abbeel, Sergey Levine. Learning Invariant Feature Spaces to Transfer Skills with Reinforcement Learning.

Robots Learning from Human Videos



Tianhe (Kevin) Yu

- Fourth-year Undergraduate Student at UC Berkeley, working with Prof. Pieter Abbeel and Prof. Sergey Levine
- Research Interest: inverse RL, model-Based RL, imitation learning and meta-learning
- Main Project: One-Shot Visual Imitation Learning via Meta-Learning (MIL) (in submission)



Overview of the Project

- Problem statement: Learn to imitate from a single visual demonstration using meta-learning
- Prior work: Model-Agnostic Meta-Learning (MAML) (Finn et al.'17) and One-Shot Imitation Learning (Duan et al.'17)
- Proposed (new) angle for solution:
  - Combine MAML with behavioral cloning.
  - Use meta-learned context variable for low-D state input tasks.
  - Learn to learn from imperfect and noisy demos. Learn to learn from raw videos only.

MAML Meta-Objective:

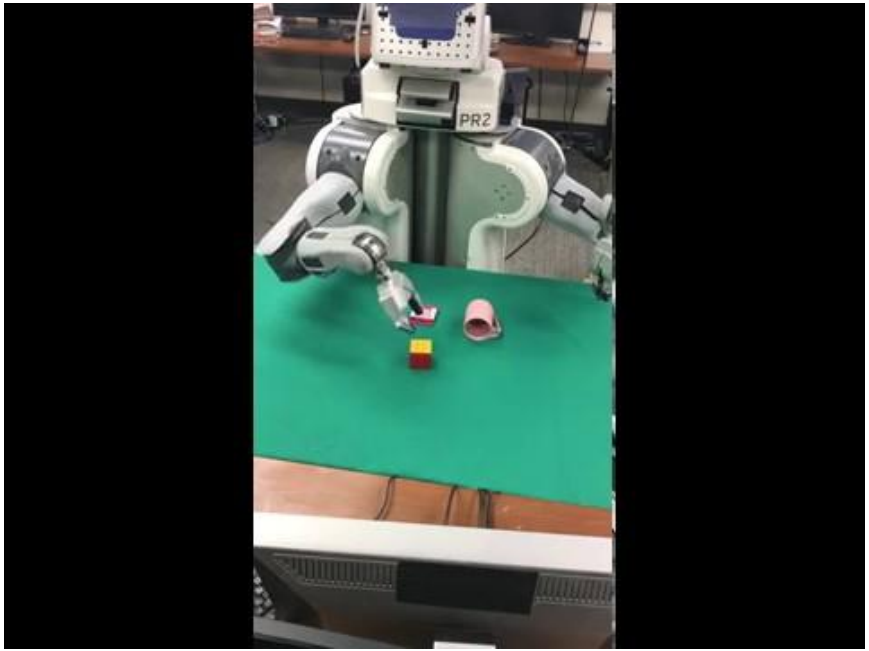
$$\min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'}) = \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)})$$

where

$$\mathcal{L}_{\mathcal{T}_i}(f_\phi) = \sum_{\tau^{(j)} \sim \mathcal{T}_i} \sum_t \| f_\phi(\mathbf{x}_t^{(j)}) - \mathbf{y}_t^{(j)} \|_2^2,$$

Experiments

Real-World Reaching
(demo)



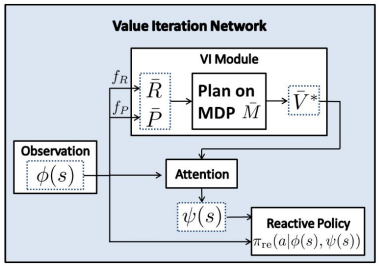
Real-World Reaching
(ours)

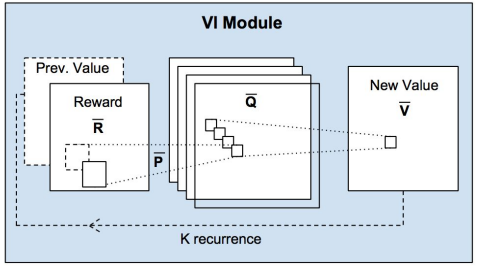
Garrett Thomas

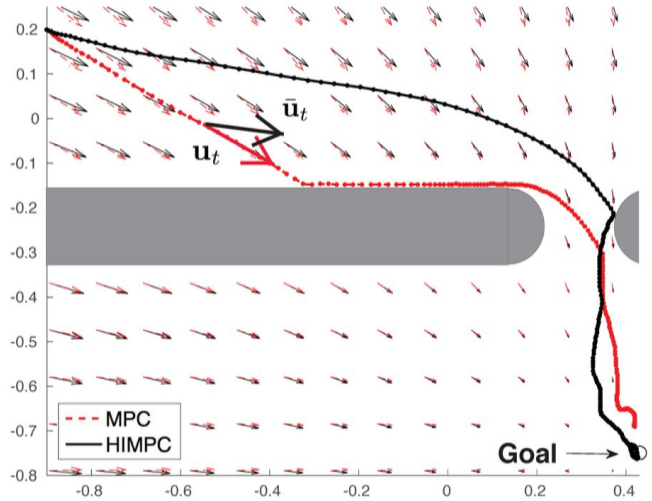
- Fourth year undergraduate (CS, Math) here at Berkeley
- Work in Professor Abbeel's group doing deep RL and applications in robotics
  - Supervised by Aviv Tamar 
- Common thread: policies that plan, especially their generalization properties

Research

- *Value iteration networks* (NIPS 2016): embed value iteration into network via standard CNN operations (convolution, max-pooling), learn to plan
  => generalizes to unseen test instances
- *Hindsight MPC* (ICRA 2017): learn a cost-shaping for short horizon model predictive control that encourages it to mimic longer horizon plan
- Currently: combining traditional robot motion planning with learning via guided policy search





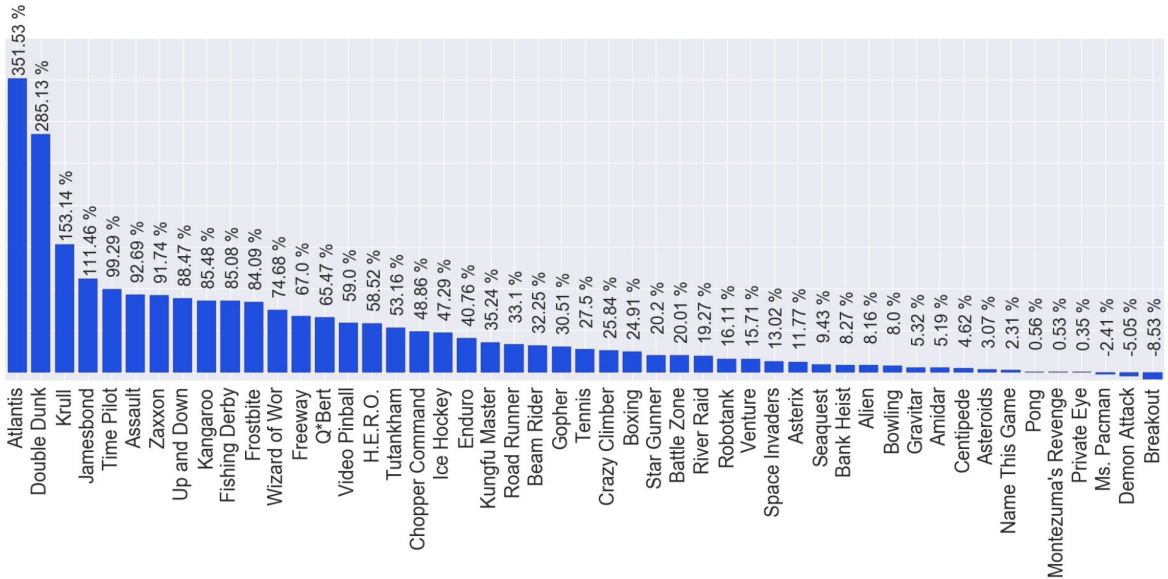


Yang Liu

- Third year Ph.D. student at UIUC and a research intern at OpenAI
- Current Interests : Exploration in RL, Bayesian RL

Research

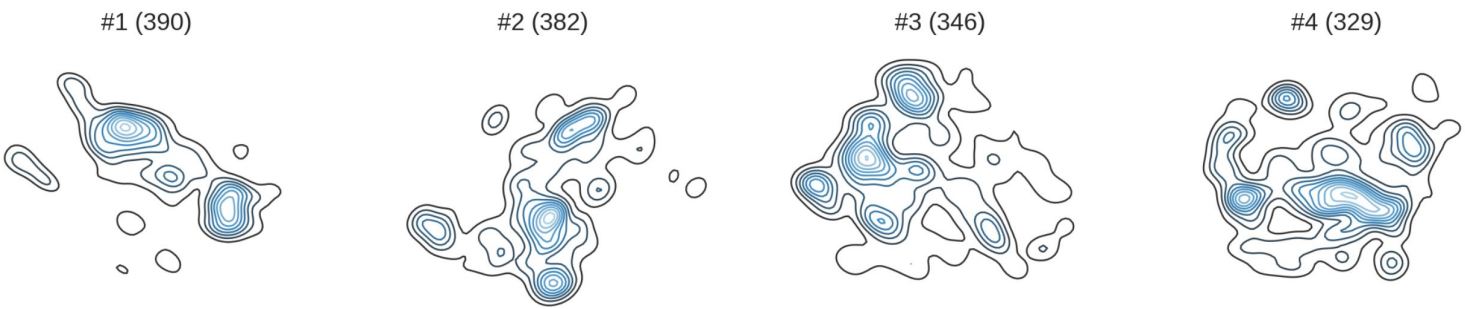
- Improve DQN with optimality tightening (ICLR 2017):
  - Faster reward propagation using a constrained optimization.
  - Achieved faster convergence and higher score on Atari Benchmark



He, F.S., Liu, Y., Schwing, A.G. and Peng, J., 2016. Learning to play in a day: Faster deep reinforcement learning by optimality tightening.

Research

- Stein Variational Policy Gradient (UAI 2017)
  - Formulated policy optimization as a bayesian inference problem.
  - Applied Stein Variational Gradient Descent to solve the bayesian inference efficiently.
  - Faster exploration and diverse policies



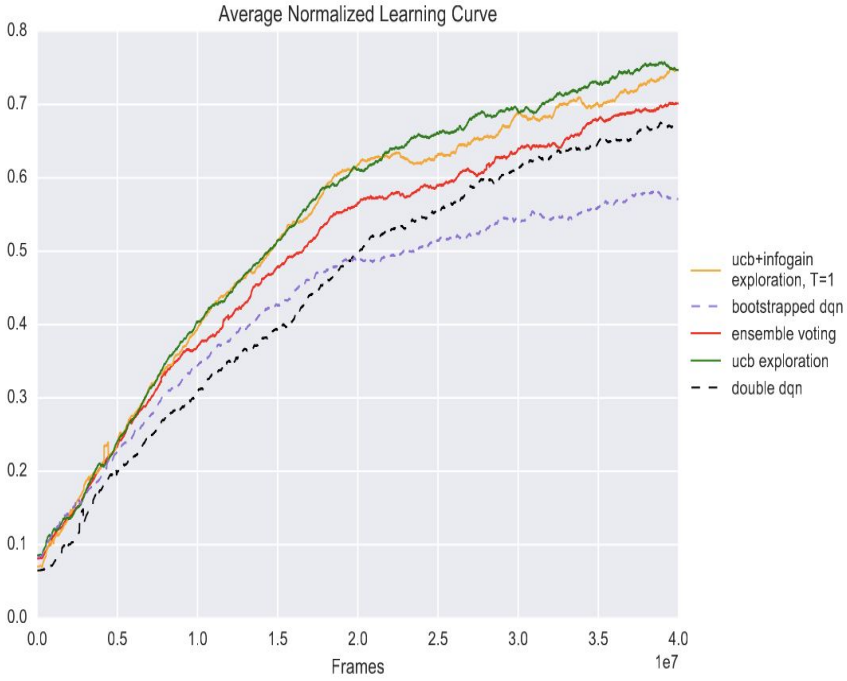
Liu, Y., Ramachandran, P., Liu, Q. and Peng, J., 2017. Stein Variational Policy Gradient.

Richard Chen

- Research Scientist at OpenAI
- Current interest:
  - Exploration in RL
  - Multi-agent RL

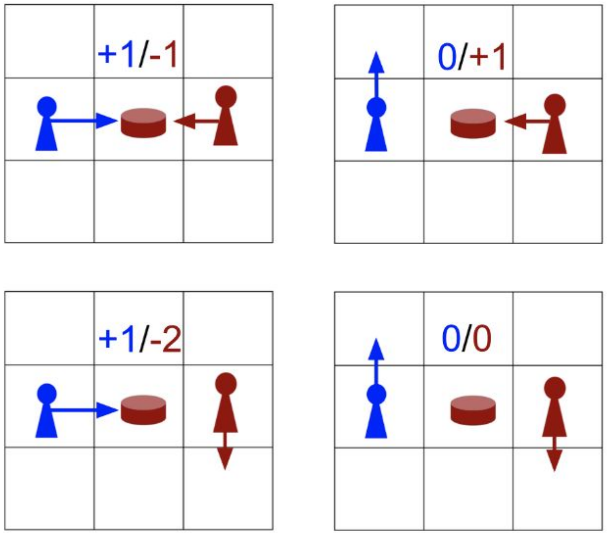
Research

- Improve DQN with Ensemble Q-learning
  - Ensemble Q functions
    - approximate sampling from the Q posterior
    - provide uncertainty estimate
  - Agent takes action by balancing
    - exploration to reduce uncertainty
    - exploitation to gain high reward



Research

- Learning with Opponent Learning Awareness (LOLA)
  - In a multi-agent setting, independent learners can lead to suboptimal outcomes
  - By explicitly modeling the learning of the opponent, collaboration emerge



- Independent agents pick up any coin
- LOLA agents pick up coins of the agents' color.

Haoran Tang

Third year PhD student at UC Berkeley

Working with Prof. Pieter Abbeel and Prof. Sergey Levine

Interests: model-free RL, model-based RL, meta-learning

Research

#Exploration: A Study of Count-Based Exploration for Deep Reinforcement Learning[1]

$$\tilde{r}(s,a) = r(s,a) + \frac{\beta}{\sqrt{N(\phi(s))}}$$

About SOTA on Freeway, Frostbite, and Solaris (Atari games)

Exploration in Frostbite

[1]Haoran Tang, Rein Houthooft, Davis Foote, Adam Stooke, Xi Chen, Yan Duan, John Schulman, Filip De Turck, Pieter Abbeel. arXiv preprint arXiv:1611.04717.

Research

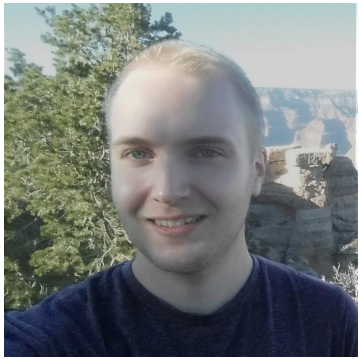
Reinforcement Learning with Deep Energy-Based Policies[2]

a.k.a **Soft Q-learning**

Tuomas will talk about it

[2]Haarnoja, T., Tang, H., Abbeel, P., & Levine, S. (2017). ICML 2017.

Tuomas Haarnoja



- PhD student working with Pieter Abbeel and Sergey Levine.
- Research interests: model-free RL.
- Currently focusing on **soft Q-learning**.

$$Q(\mathbf{s}_t, \mathbf{a}_t) \leftarrow \mathbb{E}_{\mathbf{s}_{t+1}}\left[r(\mathbf{s}_t, \mathbf{a}_t, \mathbf{s}_{t+1}) + \gamma \max_{\mathbf{a}'} Q(\mathbf{s}_{t+1}, \mathbf{a}')\right]$$

$$\log \int_{\mathcal{A}} \exp Q(\mathbf{s}_{t+1}, \mathbf{a}') d\mathbf{a}'$$

"softmax"

Paper: Tuomas Haarnoja*, Haoran Tang*, Pieter Abbeel, Sergey Levine, “Reinforcement Learning with Deep Energy-Based Policies,” ICML 2017

Tianhao Zhang

Second year Ph.D. student at UC Berkeley

Advised by Prof. Pieter Abbeel

Interests in developing learning algorithms useful for controlling real robotic systems

Important aspects: sample efficiency, using raw sensory inputs, robustness, engineering efforts

MPC-Guided Policy Search

Goal: robustness + sample efficient, using raw sensory inputs

Method: replace offline method in GPS with MPC for generating training data

With 0.05 kg mass error

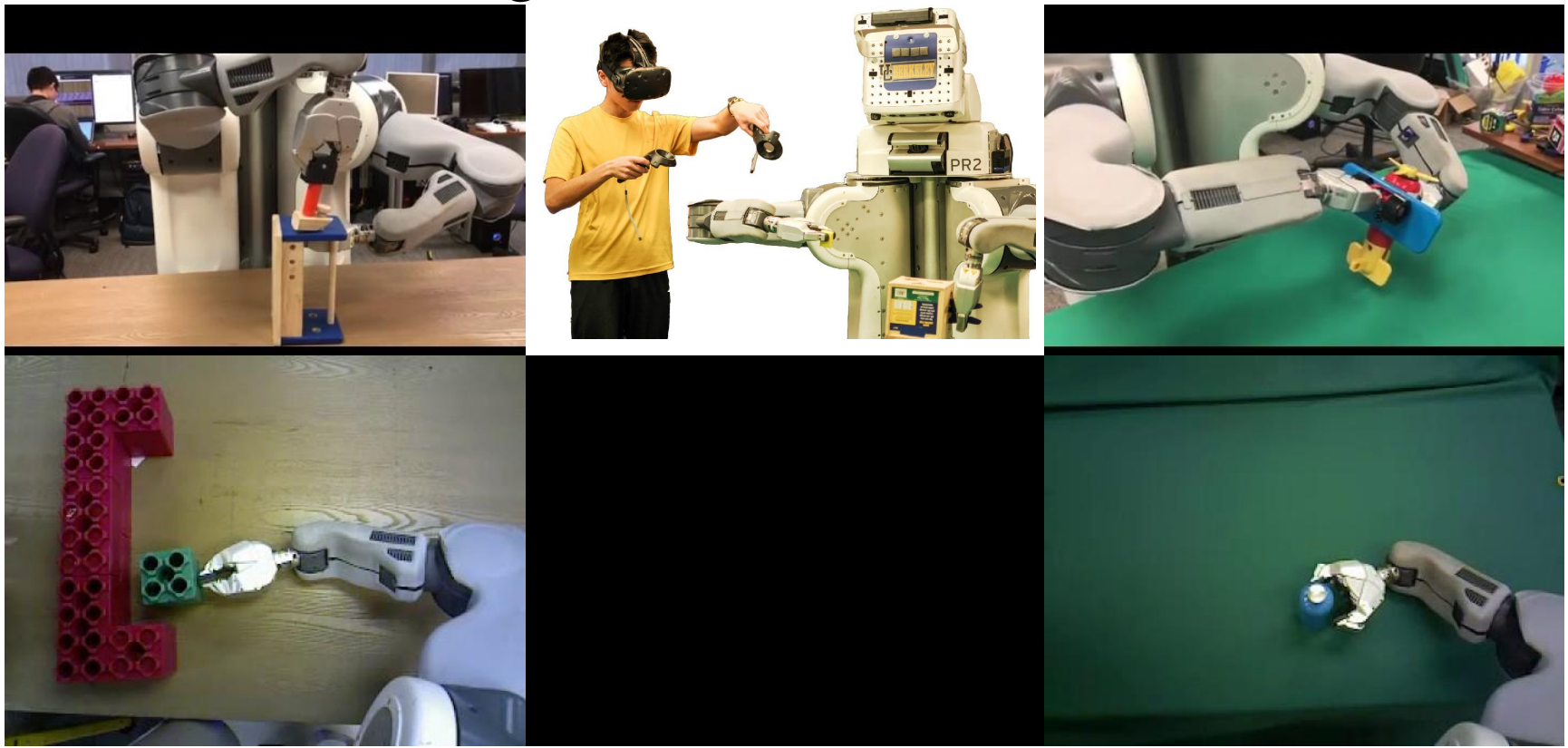
Baseline (training scene) | MPC-GPS (training scene) | MPC-GPS (test scene)

Training scene

Test scene

Zhang, Kahn, Levine, Abbeel. Learning Deep Control Policies for Autonomous Aerial Vehicles with MPC-Guided Policy Search. ICRA 2016.

Imitation learning from human demonstrations



McCarthy*, Zhang*, Jow, Lee, Goldberg, Abbeel. Deep Imitation Learning for Complex Manipulation Tasks from Virtual Reality Teleoperation. (working paper)

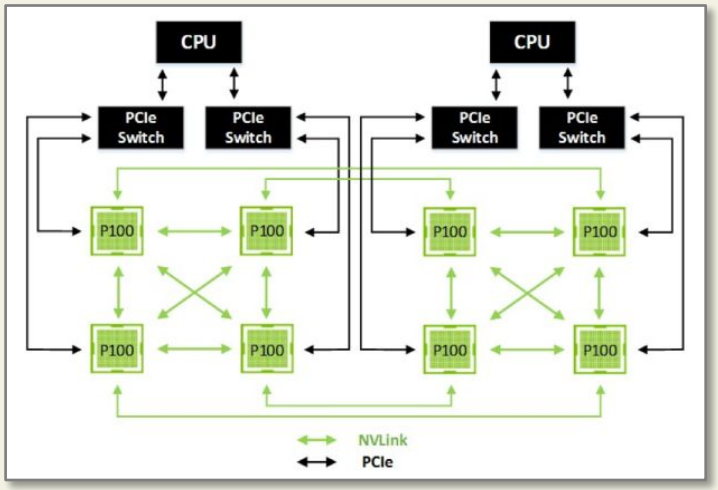
Adam Stooke

Third year PhD student at UC Berkeley

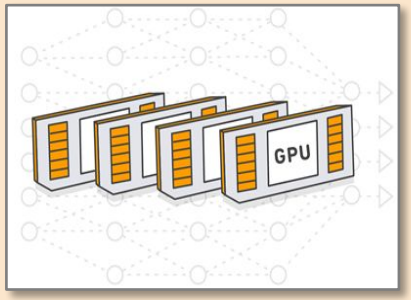
Working with Prof. Pieter Abbeel

Interests: scaling up RL (systems)

Modern Hardware: Multi-GPU



“DGX-1”: 8 GPUs



“P2” Instances: up to 16 GPUs

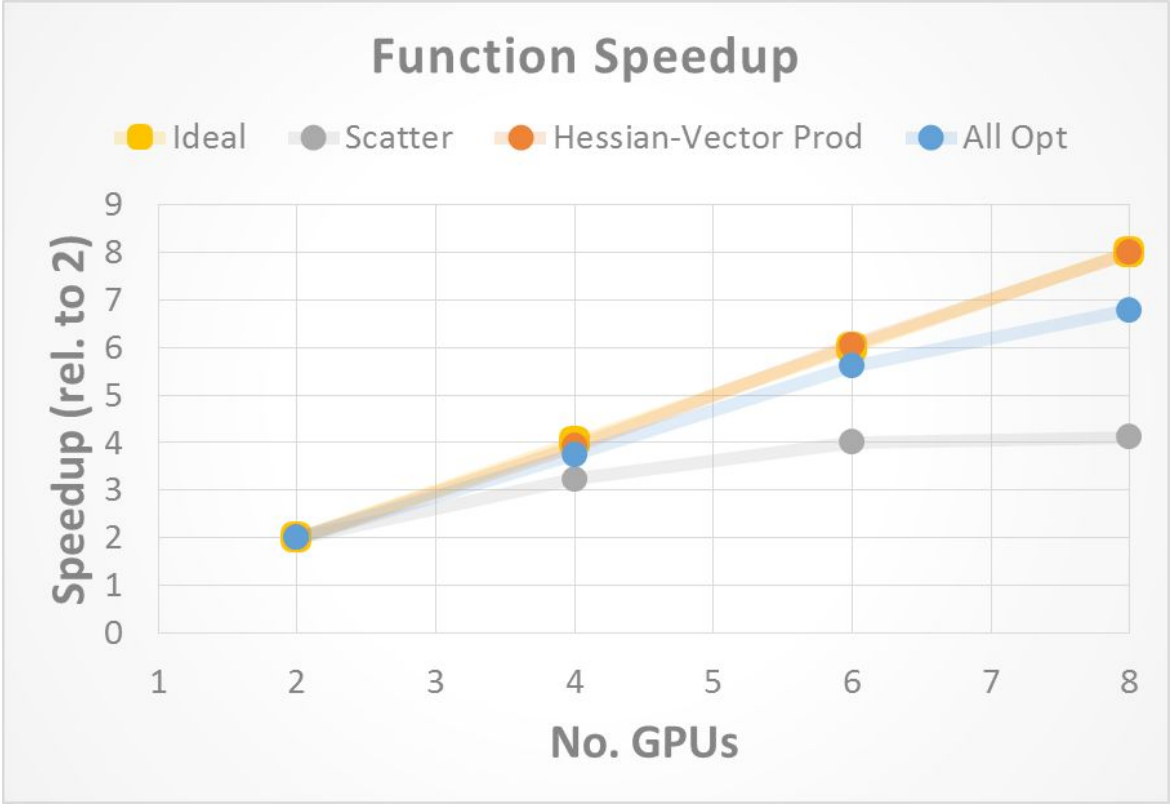
Scaling Performance

Algorithm: TRPO

Environment: Atari

Data Batch: **19** GB

Overall Speedup: 7x



Carlos Florensa

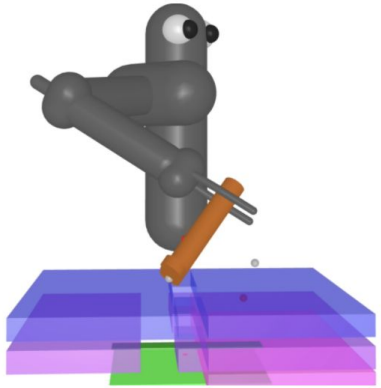
- 3rd year PhD student at UC Berkeley, working with prof. Pieter Abbeel
- Interested in training policies that **generalize to large sets of tasks**, under **weak supervision**:
  - Hierarchical RL (lear re-usable skills)
  - Curriculum Learning (improve efficiency on hard tasks)
  - Meta-learning from demonstrations (adapt to new tasks)
- Today I'll talk about "Reverse curriculum generation for RL"[1]

[1] *Carlos Florensa, David Held, Markus Wulfmeier, Pieter Abbeel. Reverse Curriculum for Reinforcement Learning. 2017,* **arXiv:1707.05300.**

Reverse curriculum generation

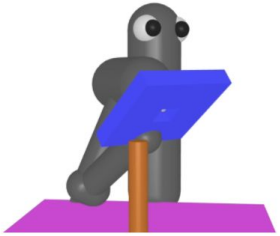
Motivation:

- Many robotics tasks require to **reach a desired configuration (goal) from everywhere.**
- Challenging for current RL: inherently **sparse rewards,** most start positions get 0 reward.

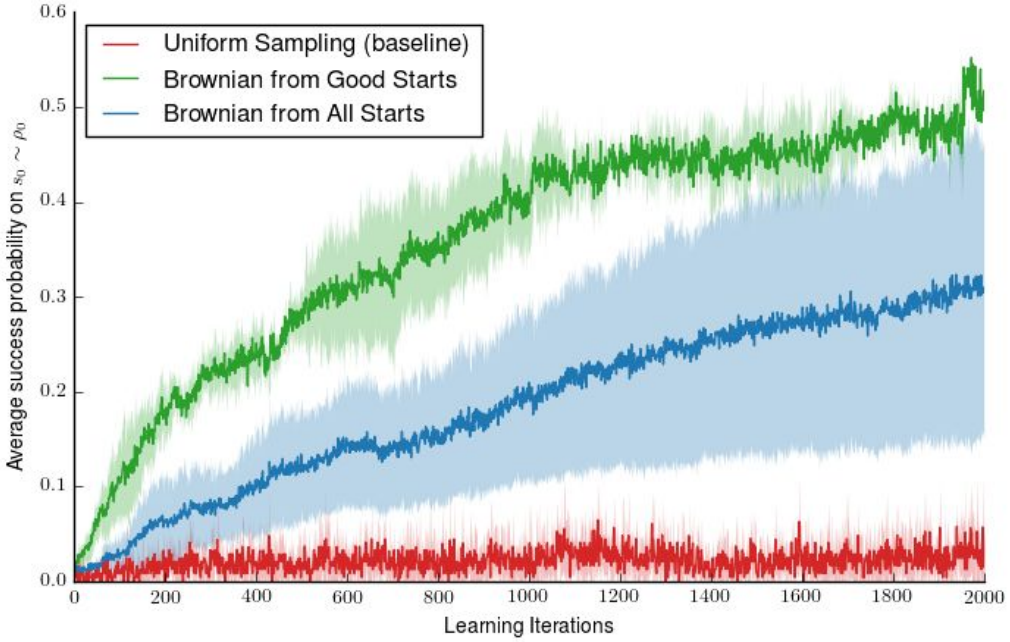


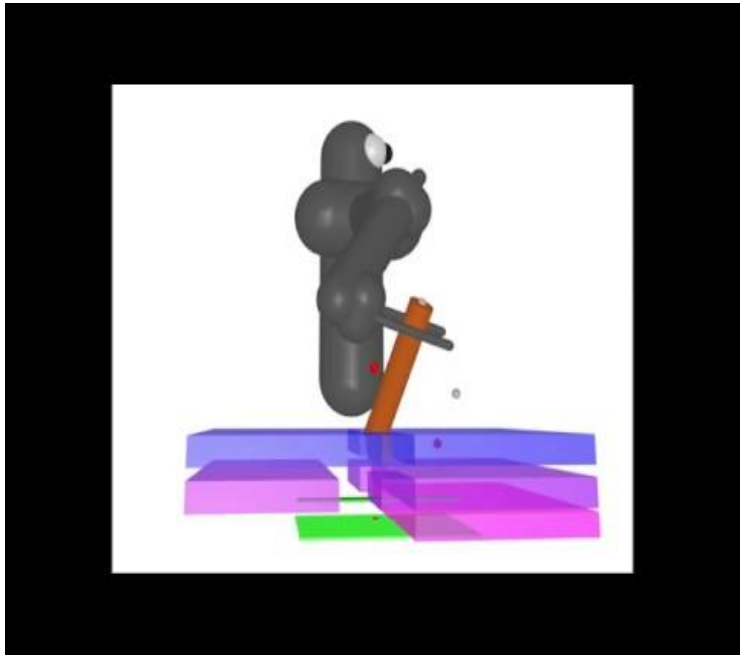
Approach:

- Solve the task in **reverse**, first training from positions **closer to the goal** and then **bootstrap** this knowledge to solve from further!
- **Sample more** start states from where you **succeed sometimes** but not always (best efficiency!).



Results





[1] *Carlos Florensa, David Held, Markus Wulfmeier, Pieter Abbeel. Reverse Curriculum for Reinforcement Learning. 2017, arXiv:1707.05300.*

Ignasi Clavera

- First Year PhD student at UC Berkeley, working with Prof. Pieter Abbeel

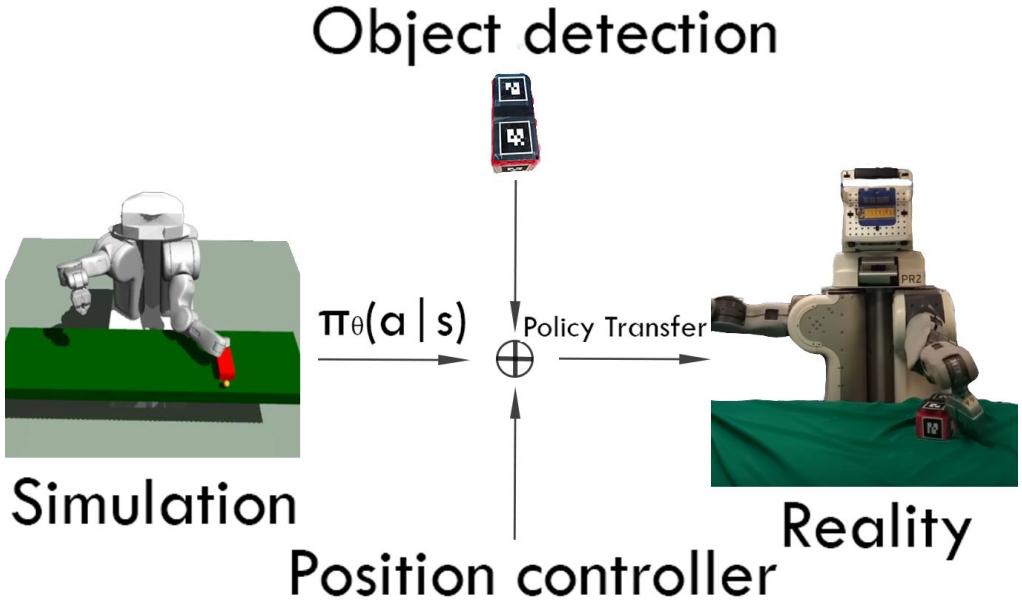
- Interests: Model-free DeepRL, Transfer Learning [1]

- I'll talk about Transfer Learning from sim to real

[1] Ignasi Clavera*, David Held*, Pieter Abbeel. Policy transfer via modularity and reward guiding.

Transfer Learning from Sim to Real

Goal: Transfer the policy from the simulator to the real world w/o further training in the real world using model-free methods.



Thanard Kurutach

Second-year PhD at UC Berkeley.

Advised by Prof. Pieter Abbeel and Prof. Stuart Russell.

Interested in model-based reinforcement learning.

Model-based RL in continuous control

Motivation: Sample efficiency and generalization across tasks.

Problems: **Model bias**, exploration, and long horizon.

Approach:

Learn multiple multiple NN dynamics models and train a NN policy to perform well against all of them.

Use TRPO on the models - free data.

Results