

# Deep Learning for Robotics - Pieter Abbeel - NIPS 2017 Keynotes

- Yunqiu Xu
  - Video: <https://www.youtube.com/watch?v=MBQTu8P1N4M>
- 

## 1. Current challenges for AI

### Many Pieces to the AI Robotics Puzzle

---

- Fast Reinforcement Learning
- ***Long Horizon Reasoning***
- Taskability
- Lifelong Learning
- Leverage Simulation
- Maximize Signal Extracted from Real World Experience

## 2. RL vs real environment

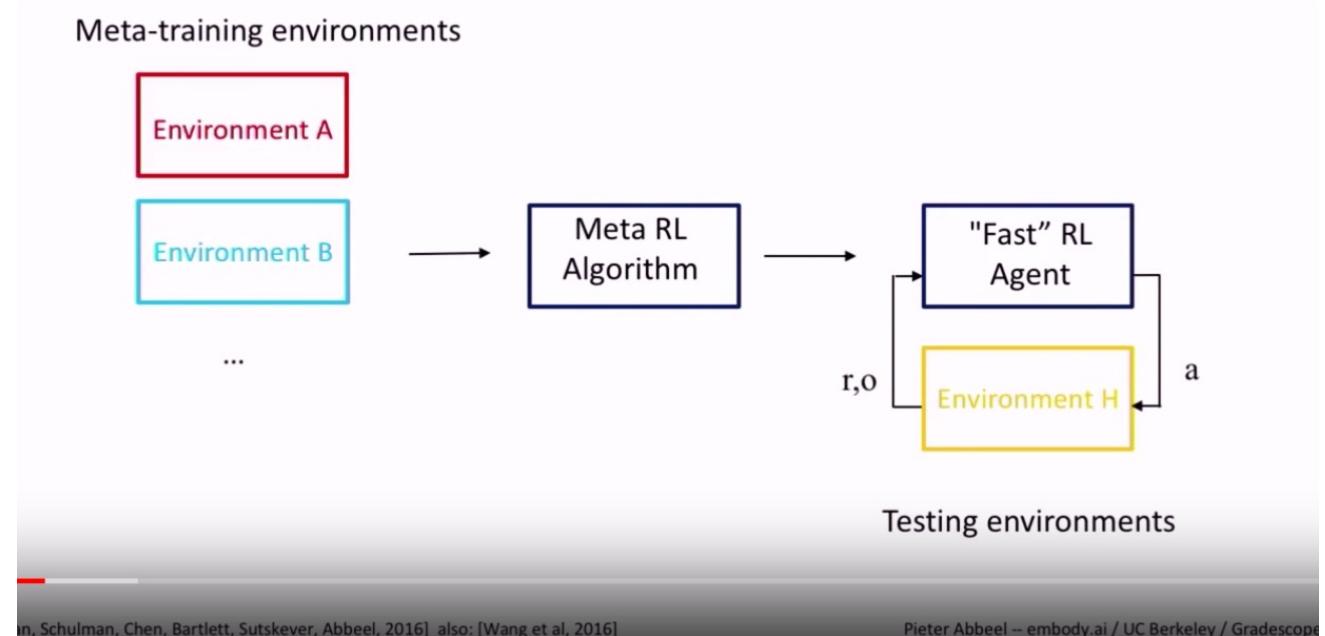
- TRPO, DQN, A3C, DDPG, PPO, Rainbow, ... are fully general RL algorithms
  - i.e., for any environment that can be mathematically defined, these algorithms are equally applicable
- Environments encountered in real world
  - = tiny, tiny subset of all environments that could be defined (e.g. they all satisfy our universe's physics)

**Can we develop “fast” RL algorithms that take advantage of this?**

### 3. Fast RL : Meta learning

- Meta learning: learning to learn

## Meta-Reinforcement Learning



- Meta learning formalize

$$\max_{\theta} \mathbb{E}_M \mathbb{E}_{\tau_M^{(k)}} \left[ \sum_{k=1}^K R(\tau_M^{(k)}) \mid \text{RLagent}_{\theta} \right]$$

$M$  : sample MDP

$\tau_M^{(k)}$  :  $k$ 'th trajectory in MDP  $M$

Meta-train:

$$\max_{\theta} \sum_{M \in M_{\text{train}}} \mathbb{E}_{\tau_M^{(k)}} \left[ \sum_{k=1}^K R(\tau_M^{(k)}) \mid \text{RLagent}_{\theta} \right]$$

$$\max_{\theta} \sum_{M \in M_{\text{train}}} \mathbb{E}_{\tau_M^{(k)}} \left[ \sum_{k=1}^K R(\tau_M^{(k)}) \mid \text{RLagent}_{\theta} \right]$$

- RLagent = RNN = generic computation architecture
  - different weights in the RNN means different RL algorithm and prior
  - different activations in the RNN means different current policy
  - meta-train objective can be optimized with an existing (slow) RL algorithm

43

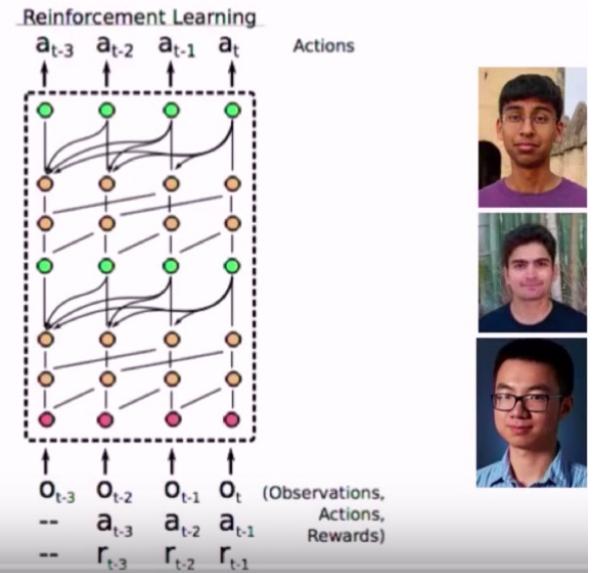
- Simple neural attentive meta learner

# Simple Neural Attentive Meta-Learner

- Like RL2 but:
  - replace the LSTM with dilated temporal convolution (like wavenet)
  - + attention

[Wavenet: van den Oord et al, 2016]

[Attention-is-all-you-need: Vaswani et al, 2017]



- MAML

## Model-Agnostic Meta-Learning (MAML)

- Starting observation:
  - Computer vision practice:
    - Train on ImageNet
    - Fine-tune on actual task
  - works really well!

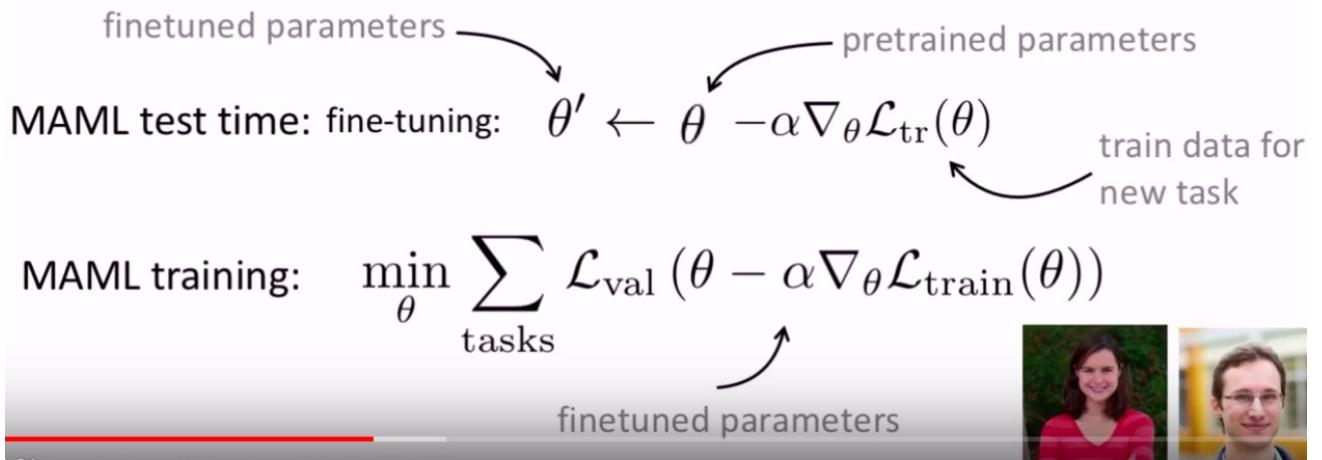
[Deng et al. '09]



- Questions:
  - How to generalize this to behavior learning?
  - And can we explicitly train end-to-end for being maximally ready for efficient fine-tuning?

# Model-Agnostic Meta-Learning (MAML)

**Key idea:** End-to-end learning of parameter vector  $\theta$  that is good init for fine-tuning for many tasks



- Related work

- Classification:

## Meta Learning for Classification

**Task distribution:** different classification datasets (input: images, output: class labels)

- Hochreiter et al., (2001) Learning to learn using gradient descent
- Younger et al., (2001), Meta learning with back propagation
- Koch et al., (2015) Siamese neural networks for one-shot image recognition
- Santoro et al., (2016) Meta-learning with memory-augmented neural networks
- Vinyals et al., (2016) Matching networks for one shot learning
- Edwards et al., (2016) Towards a Neural Statistician
- Ravi et al., (2017) Optimization as a model for few-shot learning
- Munkhdalai et al., (2017) Meta Networks
- Snell et al., (2017) Prototypical Networks for Few-shot Learning
- Shyam et al., (2017) Attentive Recurrent Comparators
- Finn et al., (2017) Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks
- Mehrotra et al., (2017) Generative Adversarial Residual Pairwise Networks for One-Shot Learning
- Mishra et al., (2017) Meta-Learning with Temporal Convolutions
- Li et al., (2017) Meta-SGD: Learning to Learn Quickly for Few Shot Learning
- Finn and Levine, (2017) Meta-Learning and Universality: Deep Representations and Gradient Descent can Approximate any Learning Algorithm

21 ■ Anon@OpenReview, (2017) Recasting Gradient-Based Meta-Learning as Hierarchical Bayes

Pieter Abbeel – embodysai / UIC Berkeley / GradeScope

- Optimization:



# Meta Learning for Optimization

**Task distribution: different neural networks, weight initializations, and/or different loss functions**

- Bengio et al., (1990) Learning a synaptic learning rule
- Naik et al., (1992) Meta-neural networks that learn by learning
- Hochreiter et al., (2001) Learning to learn using gradient descent
- Younger et al., (2001), Meta learning with back propagation
- Andrychowicz et al., (2016) Learning to learn by gradient descent by gradient descent
- Chen et al., (2016) Learning to Learn for Global Optimization of Black Box Functions
- Wichrowska et al., (2017) Learned Optimizers that Scale and Generalize
- Ke et al., (2017) Learning to Optimize Neural Nets

21 ■ Wu et al., (2017) Understanding Short-Horizon Bias in Stochastic Meta-Optimization  
Pieter Abbeel – embody.ai / UC Berkeley / Gradescope

- RL:

## Meta Learning for RL

**Task distribution: different environments**

- Schmidhuber. Evolutionary principles in self-referential learning. (1987)
- Wiering, Schmidhuber. Solving POMDPs with Levin search and EIRA. (1996)
- Schmidhuber, Zhao, Wiering. Shifting inductive bias with success-story algorithm, adaptive Levin search, and incremental self-improvement. (MLJ 1997)
- Schmidhuber, Zhao, Schraudolph. Reinforcement learning with self-modifying policies (1998)
- Zhao, Schmidhuber. Solving a complex prisoner's dilemma with self-modifying policies. (1998)
- Schmidhuber. A general method for incremental self-improvement and multiagent learning. (1999)
- Singh, Lewis, Barto. Where do rewards come from? (2009)
- Singh, Lewis, Barto. Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective (2010)
- Niekum, Spector, Barto. Evolution of reward functions for reinforcement learning (2011)
- Duan et al., (2016) RL2: Fast Reinforcement Learning via Slow Reinforcement Learning
- Wang et al., (2016) Learning to Reinforcement Learn
- Finn et al., (2017) Model-Agnostic Meta-Learning
- Mishra, Rohinjenad et al., (2017) Simple Neural Attentive meta-Learner

21 ■ Frans et al., (2017) Meta-Learning Shared Hierarchies  
Pieter Abbeel – embody.ai / UC Berkeley / Gradescope

## 4. Long Horizon Reasoning : HRL

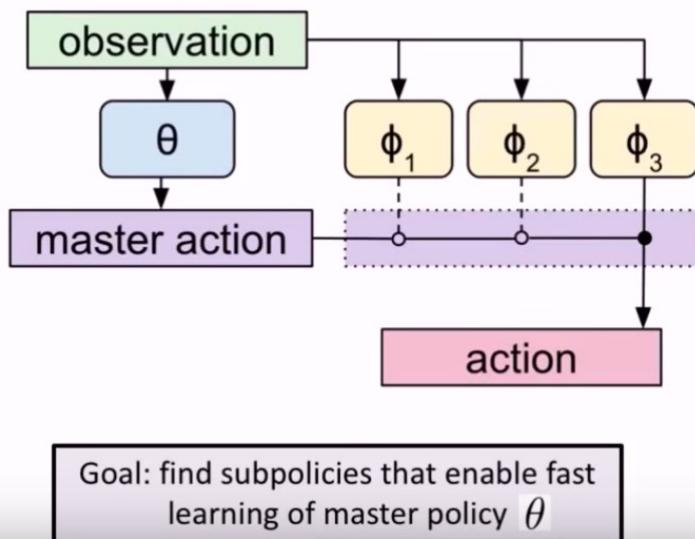
- Related work

# Hierarchical Reinforcement Learning

- Dayan and Hinton, 1993 Feudal RL
- Parr and Russell, 1998 RL with Hierarchies of Machines
- Sutton, Precup, Singh, 1999 Options
- Precup, 2000 Temporal abstraction in reinforcement learning
- Dietterich et al, 2000 MaxQ
- Fox, Moshkovitz, Tishby, 2016 Principled Option Learning
- Heess et al, 2016 Learning and transfer of modulated locomotor controllers
- Vezhnevets et al 2017 Feudal Networks for HRL
- Bacon, Harb, Precup, 2017 Option-Critic
- Florensa, Duan, Abbeel, 2017 SNNs for HRL
- Andreas, Klein, Levine, 2017 Policy Sketches

- MLSH

## Meta-Learning Shared Hierarchies



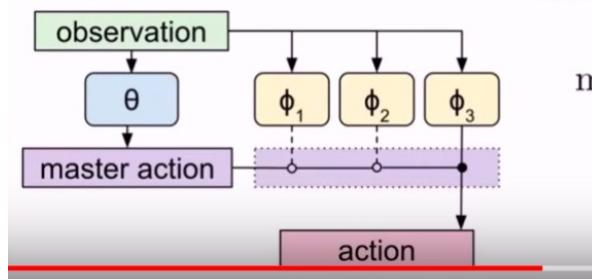
[2] Frans, Ho, Chen, Abbeel, Schulman, 2017]

Pieter Abbeel -- embody.ai / UC Berkeley / Gradescope

### RL2 Meta-Learning Objective:

$$\max_{\theta} \mathbb{E}_M \mathbb{E}_{\tau_M^{(k)}} \left[ \sum_{k=1}^K R(\tau_M^{(k)}) \mid \text{RLagent}_{\theta} \right]$$

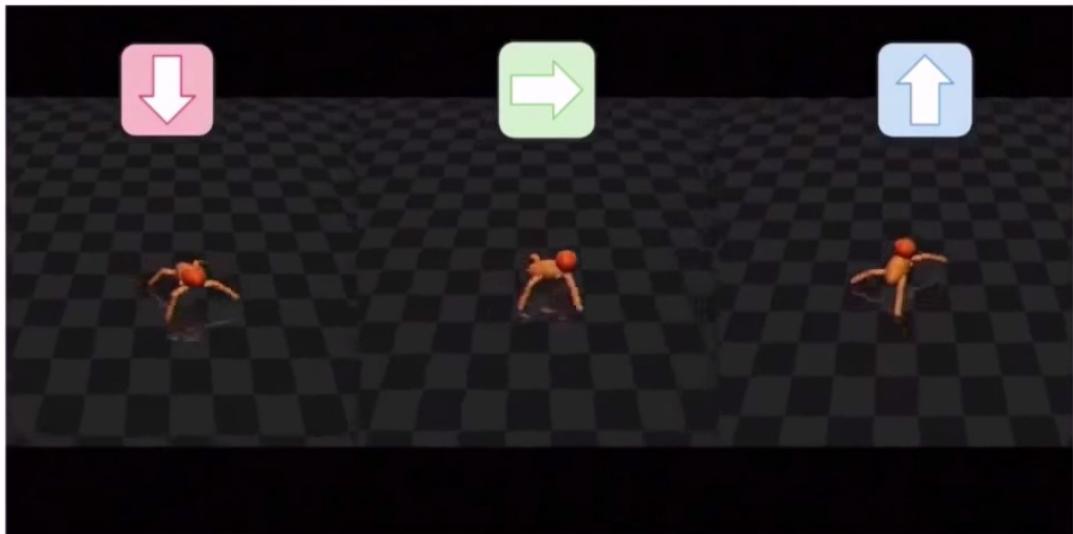
### MLSH Meta-Learning Objective:



$$\max_{\phi} \mathbb{E}_{\theta_0} \mathbb{E}_M \mathbb{E}_{\tau_M^{(k)}} \left[ \sum_{k=1}^K R(\tau_M^{(k)}) \mid \phi, \text{RLagent}_{\theta_0} \right]$$

= find a set of subpolicies that enable fast learning of the master policy

## Discovered Three Gaits

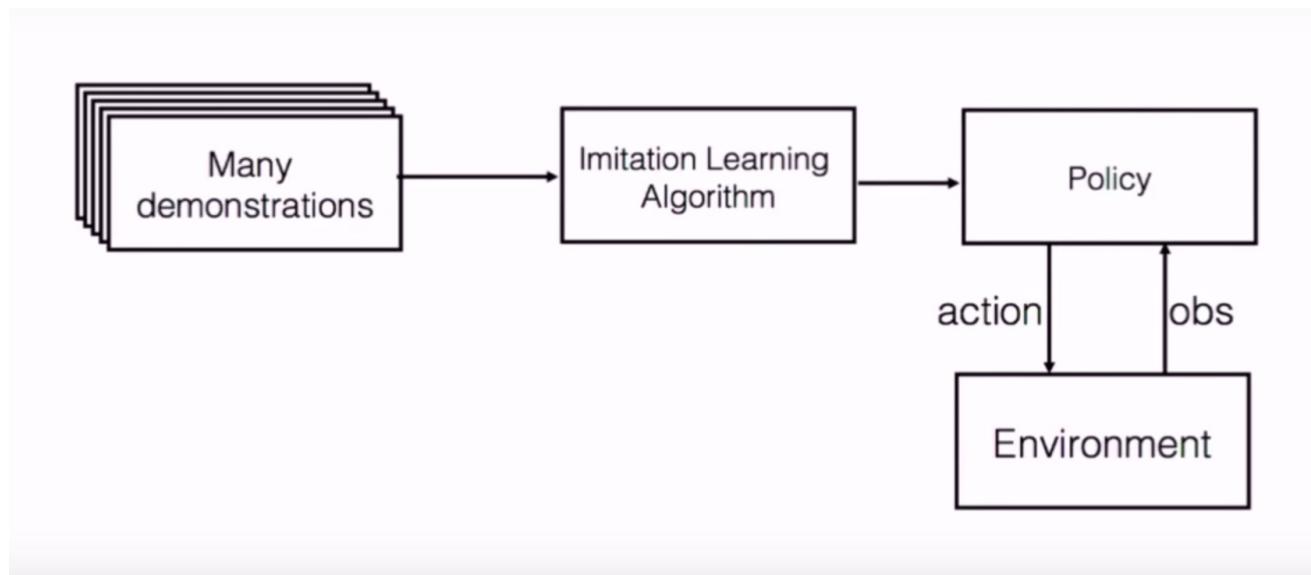


MLSH agent was trained on nine separate mazes.

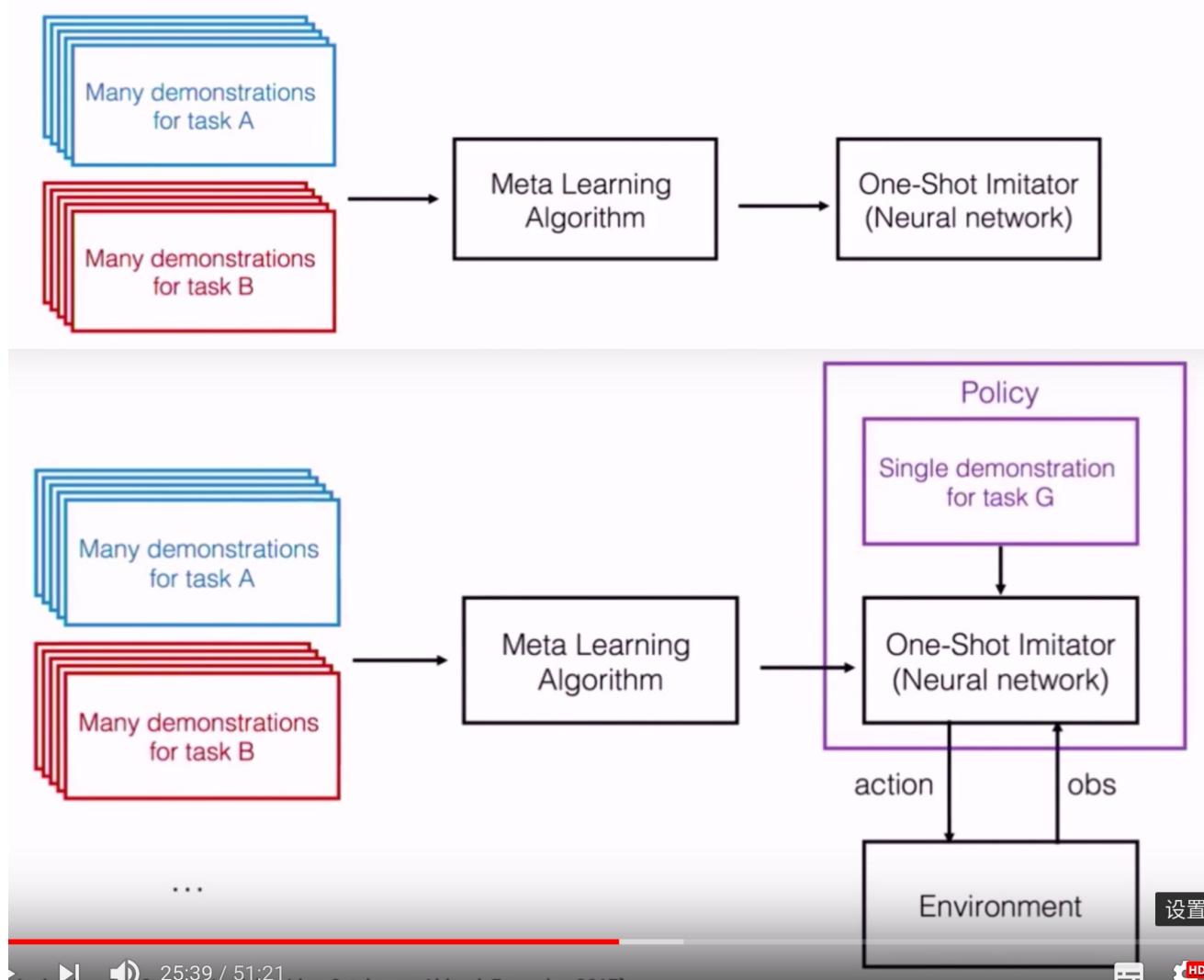
~~It discovered sub policies for upwards, rightwards, and downwards movement.~~

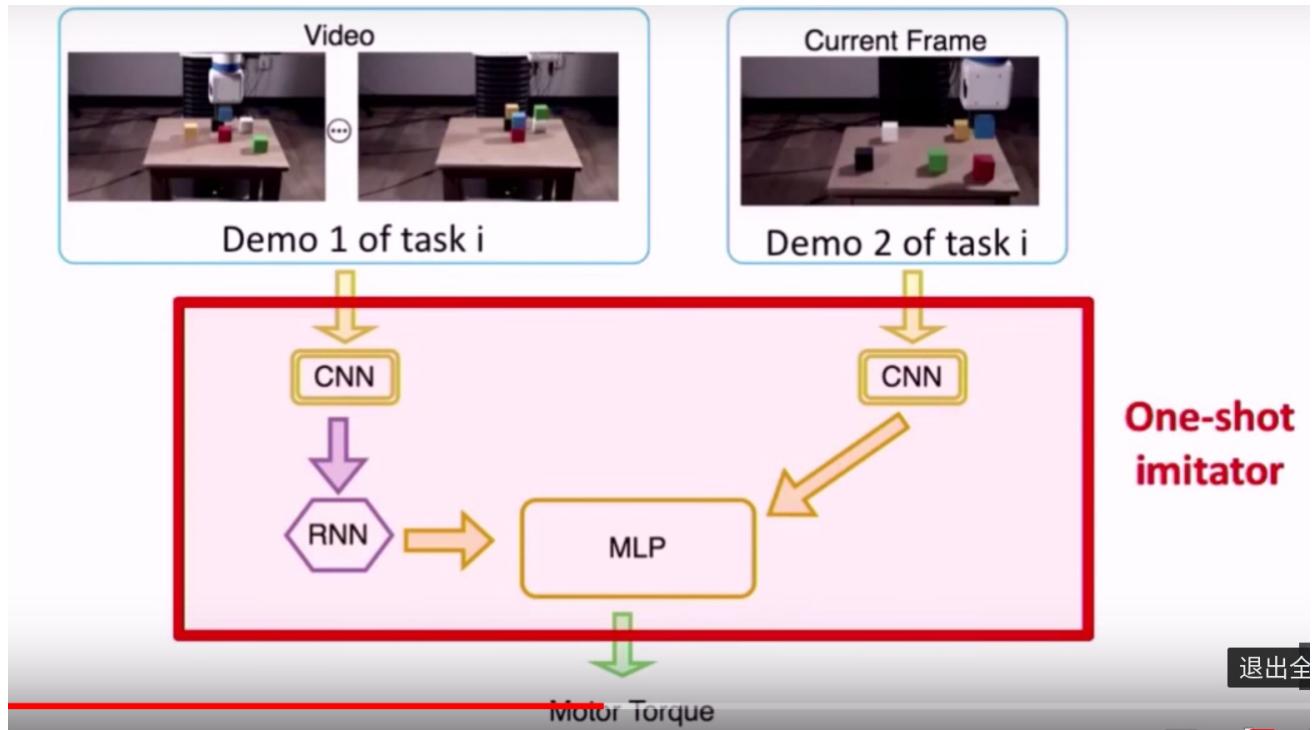
## 5. Imitation Learning

- Typical IL



- One shot imitation learning





- MIL

## Learning a One-Shot Imitator with MAML

- Meta-learning loss:

$$\min_{\theta} \sum_{\text{tasks}} \mathcal{L}_{\text{val}} (\theta - \alpha \nabla_{\theta} \mathcal{L}_{\text{train}}(\theta))$$



- Task loss = behavioral cloning loss: [Pomerleau'89, Sammut'92]

$$\mathcal{L}(\theta) = \sum_t \|\pi_{\theta}(o_t) - a_t^*\|^2$$



## Meta-training targets / objects



## ■ Meta-testing targets / objects



## 6. Lifelong Learning

- Related work

McCloskey and Cohen, Catastrophic inference in connectionist networks: the sequential learning problem (1989)

Thrun, Is learning the n-th thing any easier than learning the first? (NIPS 1996)

Thrun, Lifelong learning algorithms (1998)

Schmidhuber, Zhao, Schraudolph, Reinforcement learning with self-modifying policies (1998)

Bowling, Convergence and regret in multi-agent learning (2005)

Conitzer and Sandholm, Awesome (2007)

Hadsell et al, Learning Long-Range Vision for Autonomous Off-Road Driving (2009)

Silver, Yang, Li, Lifelong ML systems: beyond learning algorithms (2013)

Mitchell et al, Never ending learning (2015)

Rusu et al, Progressive NNs (2016)

Kirkpatrick et al, Overcoming catastrophic forgetting in NNs (2016)

Gradient Episodic Memory for Continual Learning, David Lopez-Paz, Marc'Aurelio Ranzato (2017)

Finn, Abbeel, Levine, Lifelong few-shot learning (2017)

退出全

Tessler et al, A Deep Hierarchical Approach to Lifelong Learning in Minecraft (2017)

- What is lifelong learning

- Current machine learning paradigm:
  - Step 1: Run machine learning
  - Step 2: Deploy
  - All learning happens ahead of time
- BUT: real-world deployment will face ever changing situations  
→ requires learning during deployment

## 7. Leverage Simulation

- Domain randomization



If the model sees enough simulated variation, the real world  
simulated variation even if none of it  
is realistic like ~~the next simulator~~

- Sim-to-real

Compared to the real world, simulated data collection is...

- Less expensive
  - Faster / more scalable
  - Less dangerous
  - Easier to label

to get labels because they're built into your simulator but the

## How can we learn useful real-

- Related work

## Approach 1 – Use Realistic Simulated Data



Carefully match the simulation to the world [1,2,3,4]

Augment simulated data with real data [5,6]

- [1] Stephen James, Edward Johns. *3d simulation for robot arm control with deep q-learning* (2016)
- [2] Johns, Leutenegger, Davision. *Deep learning a grasp function for grasping under gripper pose uncertainty* (2016)
- [3] Mahler et al, Dex-Net 3.0 (2017)

- [5] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. *Playing for data: Ground truth from computer games* (2016)
- [6] Bousmalis et al. *Using simulation and domain adaptation to improve efficiency of robotic grasping* (2017)

## 8. Review

- Meta-learning can be useful for other methods

— Future Developments in Learning

- Long Horizon Research
- Taskability
- Lifelong Learning
- Leveling the playing field
- Maximizing Signal-to-Noise

## Recurring Theme:

### Meta-Learning

Enables discovering algorithms that are

powered by data/experience (vs. just human ingenuity)

Requires more compute

Which is something we are continuing to get