Learning Representations for Versatile Behavior

Chelsea Finn
If you gave the parts to a 9 month old…

... vs. older child or adult

Humans reuse prior experience.
Robot learning

**paradigm:** train/test on 1 task in 1 environment, starting from scratch
How can robots use past experience?

1. Learn about the physical world

2. Learn to learn
How can robots use past experience?

1. Learn about the physical world
   - Finn et al. NIPS ’16, Finn & Levine ICRA ’17
   - Ebert et al. ’17 (under review)

2. Learn to learn
   - Finn et al. ICML ’17
   - Finn*, Yu* et al. ’17 (under review)
How can robots use past experience?

1. Learn about the physical world

2. Learn to learn

Finn et al. NIPS ’16, Finn & Levine ICRA ’17
Ebert et al. ’17 (under review)

Finn et al. ICML ’17
Finn*, Yu* et al. ’17 (under review)
Learning about the Physical World

*self-supervised setting*

- use **raw sensory inputs** (i.e. vision)
- scale up data collection to **learn without supervision**
- instead of learning single-purpose policy, learn **reusable model**
Data collection - 50k sequences (1M+ frames)

Finn, Goodfellow, Levine, NIPS ’16
Ebert, Finn, Lee, Levine ’17 (under sub.)
Train predictive model

\[ I_t, a_{t:t+H} \rightarrow I_{t:t+H} \]

\[ \text{<— varying actions —>} \]

Finn, Goodfellow, Levine, NIPS ’16
Ebert, Finn, Lee, Levine ’17 (under sub.)
Planning with Visual Foresight

1. Consider potential action sequences
2. Predict the future for each action sequence
3. Pick best future & execute corresponding action
4. Repeat 1-3 to replan in real time

visual model-predictive control (MPC)
How to predict video?

**Goal:** Design a prediction model that’s good for control
- action-conditioned
- multi-frame prediction
- explicitly model motion

**Key idea:** output how pixels will move, rather than pixels themselves

Finn, Goodfellow, Levine, NIPS’16
Stacked ConvLSTM

convolution with untied weights
(locally connected)

action $a_t$

$t_t$

$t_{t+1}$

$P$

$t_0$

$a_0$

$t_1$

$a_1$

$t_2$

$a_2$

$t_3$
Use both first frame and most recent frame for prediction
Use both first frame and most recent frame for prediction

Designated Pixel

most recent frame only

both

Ebert, Finn, Lee, Levine, ‘17
How it works

user specifies goal
How does this approach do?

- evaluation on maneuvering seen & novel objects
- model trained on 8 days of unlabeled robot data

*Completely self-supervised:* Only human involvement during training is: programming initial motions and providing objects

Finn & Levine ICRA ‘17
Ebert, Finn, Lee, Levine, ‘17
Planning with raw sensory inputs

**Pros:** learn for a *variety of tasks*, entirely *self-supervised*

**Cons:** can’t [yet] learn *complex skills*, compute intensive at test time

*Next steps for this approach:*
- stochastic model for *handling uncertainty* in the long term
- collect data using model to achieve *more complex goals*
- *long term planning* in more abstract spaces
How can robots use past experience?

1. Learn about the physical world
   - *Finn* et al. NIPS ‘16, *Finn* & Levine ICRA ‘17
   - Ebert et al. ‘17 (under review)

2. Learn to learn
   - *Finn* et al. ICML ‘17
   - *Finn*, *Yu* et al. ‘17 (under review)
**Few-shot learning:** incorporate prior knowledge from other tasks for fast learning

**Image Recognition**

<table>
<thead>
<tr>
<th>Training data</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Meta-training</strong></td>
<td><img src="image1.png" alt="Training data images" /></td>
</tr>
<tr>
<td><strong>Meta-testing</strong></td>
<td><img src="image3.png" alt="Training classes" /></td>
</tr>
</tbody>
</table>

**What is few-shot learning for behavior?**

*Diagram adapted from Ravi & Larochelle ’17*
Few Shot Learning via Recurrence

(Santoro et al. ’16, Duan et al. ’17, Wang et al. ’17)
Learning Few-Shot Adaptation

**Transfer learning:** finetune from ImageNet-trained features (Deng et al. ’09, Donahue et al. ’14)
+ simple, works well, same learning rule  
- no ImageNet for behavior…

How can we get transferable features for behavior?
Learning Few-Shot Adaptation

**Transfer learning:** finetune from ImageNet-trained features (Deng et al. ’09, Donahue et al. ’14)

+ simple, works well, same learning rule  
- no ImageNet for behavior…

How can we get transferable features for behavior?

**Fine-tuning:**
\[
\theta \leftarrow \theta - \alpha \nabla_{\theta} \mathcal{L}_T(\theta)
\]

**Our method:**
\[
\min_{\theta} \sum_{\text{task } T} \mathcal{L}_T(\theta - \alpha \nabla_{\theta} \mathcal{L}_T(\theta))
\]

**Key idea:** Train over many tasks, to learn parameter vector $\theta$ that transfers
Learning Few-Shot Adaptation

\[
\min_{\theta} \sum_{\text{task } T} \mathcal{L}_T(\theta - \alpha \nabla_\theta \mathcal{L}_T(\theta))
\]

\(\theta\) parameter vector being meta-learned

\(\theta_i^*\) optimal parameter vector for task \(i\)

Model-Agnostic Meta-Learning
Few-Shot Learning Experiments

**Few-Shot Classification**

compare to prior methods

---

**One-Shot Imitation**

---

**Fast Adaptation in Reinforcement Learning**
Few-Shot Image Recognition

Omniglot

- siamese networks [1]
- matching networks [2]
- neural statistician [3]
- memory module [4]
- MAML

1200 training classes, 423 test classes

MinilImagenet

- matching networks
- LSTM optimizer
- MAML

64 training classes, 24 test classes

One-Shot Visual Imitation Learning

Vision-Based Manipulation Problems

- simulated reaching
- simulated pushing

Methods:

- MAML
- LSTM (Duan et al. ’17)
- contextual

supervised learning for all objectives
One-Shot Imitation: simulated reaching from pixels

input demonstration

MAML contextual

planar reaching (with vision)

success rate (%)

0 1000 2000 3000 4000 5000 6000

total number of demonstrations in training set

LSTM (test) LSTM (train) MIL (test) MIL (train) Contextual (test) MIL+ltv (test) Contextual (train) MIL+ltv (train)
One-Shot Imitation: simulated pushing from pixels

input demonstration

learned policy

115 random objects with random textures, masses, frictions, etc.

**Takeaway:** reuse experience across objects when learning to interact with new objects
Fast Adaptation in Reinforcement Learning

**Locomotion problems:**

1. run at goal velocity (continuous range of tasks)
2. run forward or backward (2 tasks)

**Methods:**

- MAML
- pretrain on all tasks
- random init

meta-learning and adaptation with policy gradients

all: 20 roll-outs for learner update
run backward or forward
walk/run at goal velocity
Main Takeaway: Robots can reuse prior experience for faster learning

1. Learn about the physical world

2. Learn to learn

Finn et al. NIPS ’16, Finn & Levine ICRA ’17
Ebert et al. ’17 (under review)

Finn et al. ICML ’17
Finn*, Yu* et al. ’17 (under review)
Future Directions

continual learning in the real world

learning about the world

learning to learn

infer reward from human demonstrations

what is the reward?

Finn, Levine, Abbeel, ICML '16
Collaborators

Sergey Levine  Pieter Abbeel  Frederik Ebert  Tianhe (Kevin) Yu

Ian Goodfellow  Alex Lee  Tianhao Zhang
Questions?

All data and code linked at: people.eecs.berkeley.edu/~cbfinn

cbfinn@eecs.berkeley.edu
Omniglot Few-Shot Classification

Omniglot dataset

Memory-Augmented Approach
Santoro et al. ‘16

Class Prediction

$(x_t, y_t)(x_{t+1}, y_{t+1})$

Episode

Omniglot Few-Shot Classification

Test error

5-way 1-shot

5-way 5-shot

non-convolutional network

memory-augmented neural network

MAML
Foresight quantitative comparison

![Bar charts showing distance to goal in pixels for different methods and conditions.](chart1)

![Bar charts showing average distance for multi-objective pushing benchmark.](chart2)