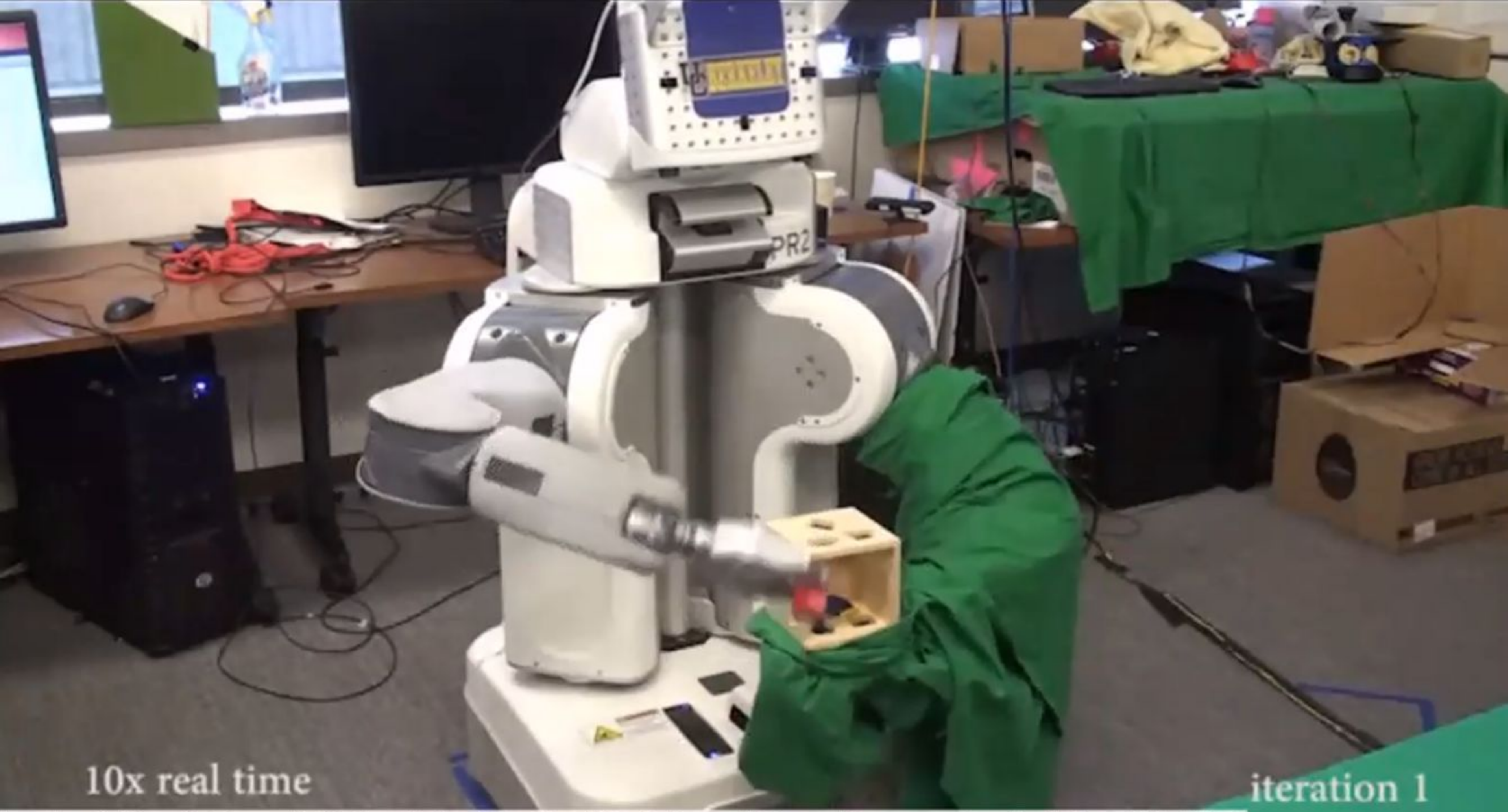
Self-Supervised Robot Learning

Chelsea Finn

Stanford   Google Brain



Levine*, Finn*, Darrell, Abbeel. *End-to-End Training of Deep Visuomotor Policies*. JMLR '16

Robot reinforcement learning

Finn et al. '16

Chebotar et al. '17

Yahya et al. '17

Ghadirzadeh et al. '17

Reinforcement learning

Atari

locomotion

We have a <u>big</u> problem.

Learn **one task** in **one environment**, ***starting from scratch***

Can we build more general robot learning systems?

People accumulate & learn from broad experiences

Source: https://youtu.be/8vNxjwt2AqY

Simple, yet **general**, manipulation skills are beyond the scope of current methods.

How is this the case?

It turns out — the **simpler** and **broader** capabilities are **really hard**.
(Moravec's Paradox)

This lecture: self supervised robot learning.

object classification

supervised learning

iid data

large labeled, curated dataset

well-defined notions of success

object manipulation

sequential decision making

action affects next state

how to collect data?
what are the labels?

what does success mean?

Collect **diverse** data by “playing”

In contrast to task learning: no notions of **progress** or **success**!

Collect data

Learn to predict

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

Contrast to:

Models capture **general purpose** knowledge about the world

Use **all** of the available supervision signal.

Also: No assumptions about task **representations**.

Prediction is a **supervised learning** problem.

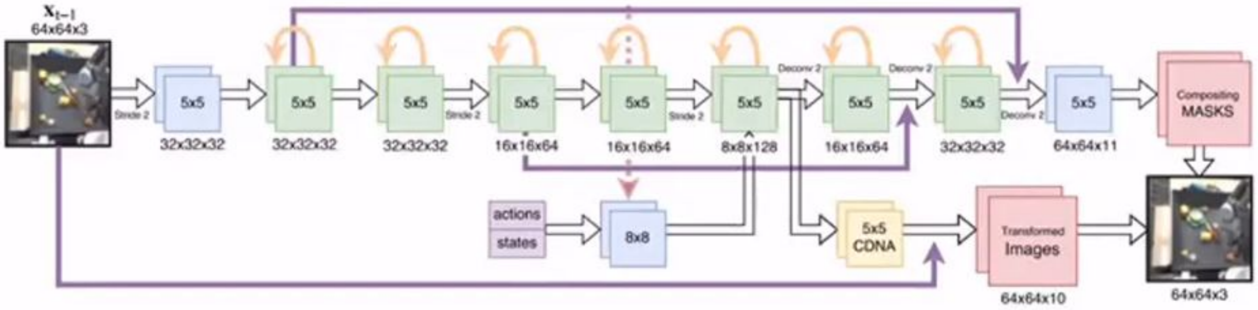
Larger models -> better predictions.

Villegas et al. NeurIPS '19

How to predict video?

deep neural network trained with supervised learning

o_t image a_t action $\rightarrow$ o_{t+1} next image



Finn, Goodfellow, Levine NIPS '16
Finn & Levine ICRA '17

Collect data

Learn to predict

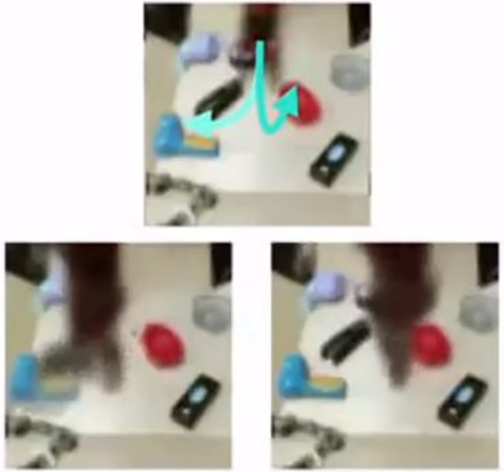
$I_t, a_{t:t+H} \rightarrow I_{t:t+H}$

Plan using model

Planning to accomplish goals

1. Consider **potential action sequences**
2. **Predict the future** for **each action sequence**
3. **Execute** the **best plan**
4. **Iteratively replan**

"visual model-predictive control" (visual MPC)



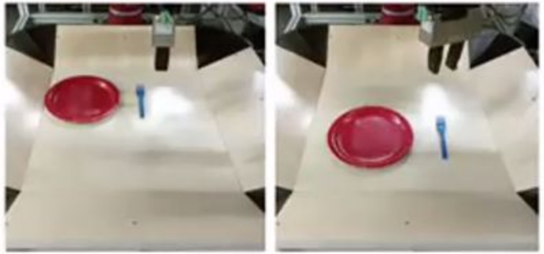
Finn & Levine. ***Deep Visual Foresight for Planning Robot Motion***. ICRA '17

Which future is the best one?

Human specifies a goal by:

Selecting where pixels should move.

Providing an image of the goal.



Providing a few examples of success.

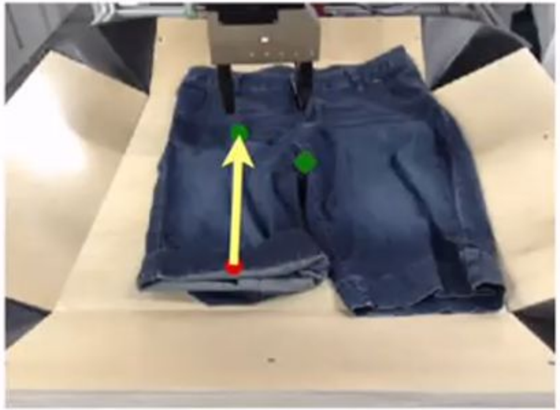
Finn & Levine ICRA '17

Ebert, Lee, Levine, Finn CoRL '18

Xie, Singh, Levine, Finn CoRL '18

How it works

Specify goal



Visual MPC
w.r.t. goal

Visual MPC execution

Ebert*, Finn*, Dasari, Xie, Lee, Levine. Visual Foresight: Model-Based Deep Reinforcement Learning for Vision-Based Robotic Control.

How it works

Given 5 examples of success

infer goal classifier

visual MPC w.r.t. goal classifier

Visual MPC with learned objective

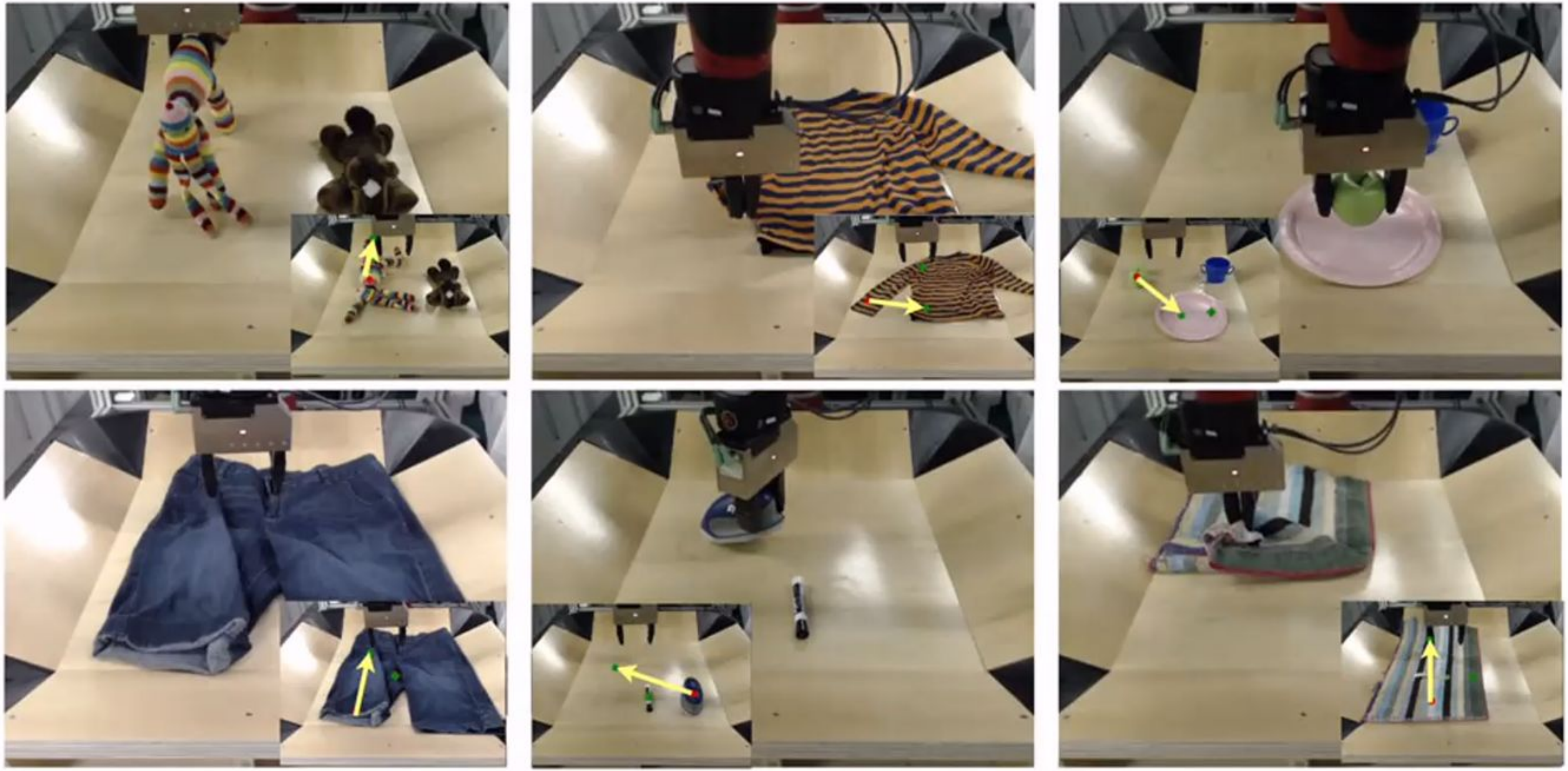
Xie, Singh, Levine, Finn. Few-Shot Goal Inference for Visuomotor Learning and Planning, CoRL 2018

Algorithm Recap

1. run base policy $\pi_0(\mathbf{a}_t|\mathbf{o}_t)$ (e.g., random policy) to collect $\mathcal{D} = \{(\mathbf{o}, \mathbf{a}, \mathbf{o}')_i\}$
2. learn model $f_\phi(\mathbf{o}, \mathbf{a})$ to minimize $\sum_i ||f_\phi(\mathbf{o}_i, \mathbf{a}_i) - \mathbf{o}'_i||^2$
3. use model $f_\phi(\mathbf{o}, \mathbf{a})$ to optimize action sequence
4. execute the first planned action, observe resulting state $\mathbf{o}'$
5. append $(\mathbf{o}, \mathbf{a}, \mathbf{o}')$ to dataset $\mathcal{D}$

Planning with a **single, self-supervised model** for many tasks

Video speed: 2x



Some remaining questions

Why is the robot so slow?

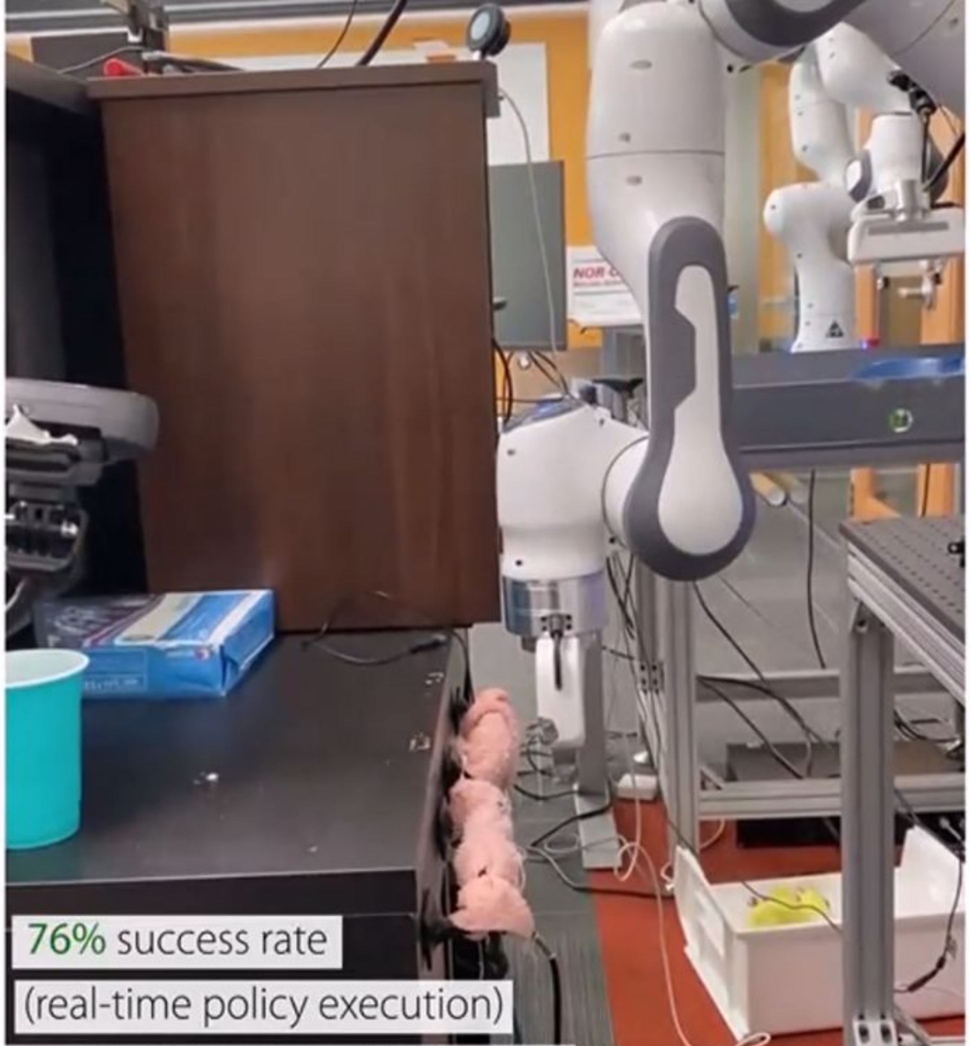
Can we scale these models to broader data?

How can humans help?

Optimizing a policy $\pi(a \mid s)$ instead of actions.

+

Using a more targeted dataset



Rafailov*, Yu*, Rajeswaran, Finn. ***Offline Reinforcement Learning from Images with Latent Space Models***, L4DC '21

Some remaining questions

Why is the robot so slow?

Can we scale these models to broader data?

How can humans help?

What is the bottleneck for handling large, diverse datasets?

Ground truth RoboNet videos

Predictions from SVG' model (Villegas et al. NeurIPS '19)

Bottleneck: *underfitting*

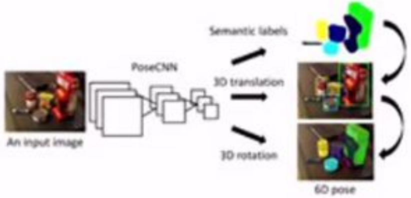
Ruben Villegas et al. *High Fidelity Video Prediction with Large Stochastic Recurrent Neural Networks*. NeurIPS '19.

How can we scale dynamics models?

Thought 1: Can we learn to model **only what matters**?

(1a) learn a representation using task supervision

e.g. standard state estimation



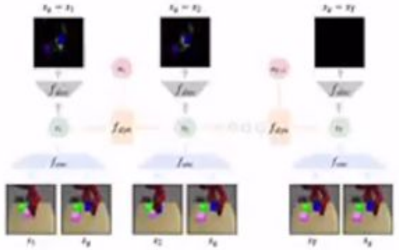
Xiang, Schmidt, Narayanan, Fox. RSS '18

e.g. Hafner et al. ***Learning Latent Dynamics for Planning from Pixels.*** ICML '19

+ works well if you can get **a lot** of supervision

- requires significant supervision/ engineering per task

(1b) learn a representation tailored to provided goal image



Nair, Savarese, Finn. ***Goal-Aware Prediction: Learning to Model What Matters***. ICML '20

+ no supervision required

+ can successfully redistribute model errors

- more research needed :)

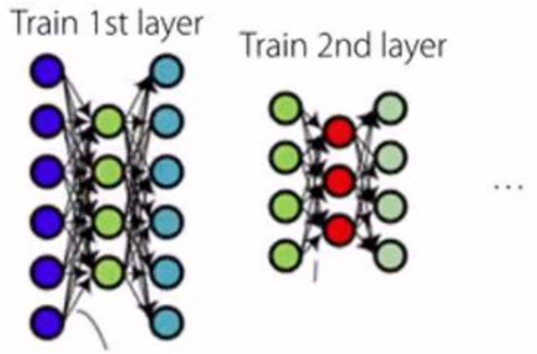
How can we scale dynamics models?

One practical bottleneck: GPU memory

Old trend: Layer-wise training of neural networks

Hinton, Osindero, Teh. ***A fast learning algorithm for deep belief nets***. Neural Computation '06.

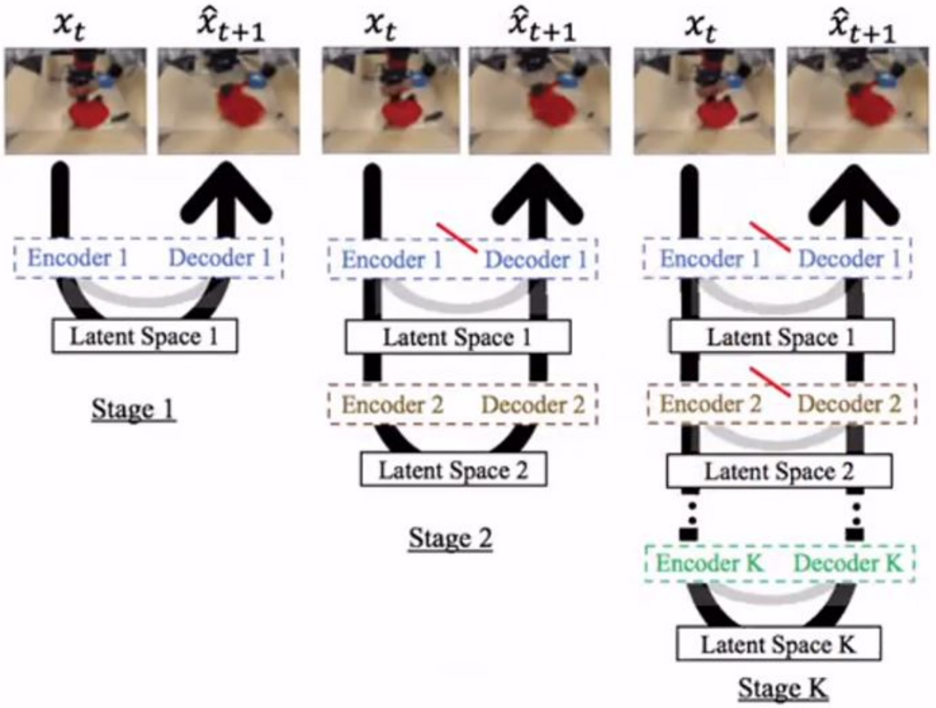
Vincent, Larochelle, Lajoie, Bengio, Manzagol, Bottou. ***Stacked denoising autoencoders***. JMLR '10



Greedy training reduces memory costs!

How can we scale dynamics models?

Greedy training of video prediction models

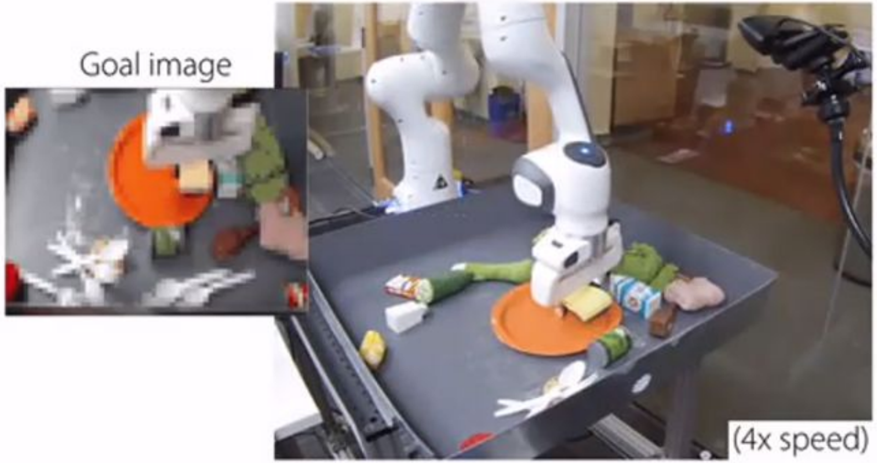


High-level bits:

- train modules **sequentially**
- **optimize** each w.r.t. **evidence lower bound**
- encoder **compresses spatial dimensions** to prevent correlations in latent variable

Wu, Nair, Martin Martín-Martín, Fei-Fei, Finn. ***Greedy Hierarchical Variational Autoencoders for Large-Scale Video Prediction***, CVPR '21

Do these models lead to **better downstream planning**?



Success rate on tasks with **unseen** objects

Method	Test Task Success Rate	
	Pick&Wipe	Pick&Sweep
GHVAEs	**90.0%**	**87.5%**
SVG	50.0%	50.0%

Wu, Nair, Martin Martín-Martín, Fei-Fei, Finn. ***Greedy Hierarchical Variational Autoencoders for Large-Scale Video Prediction***, CVPR '21

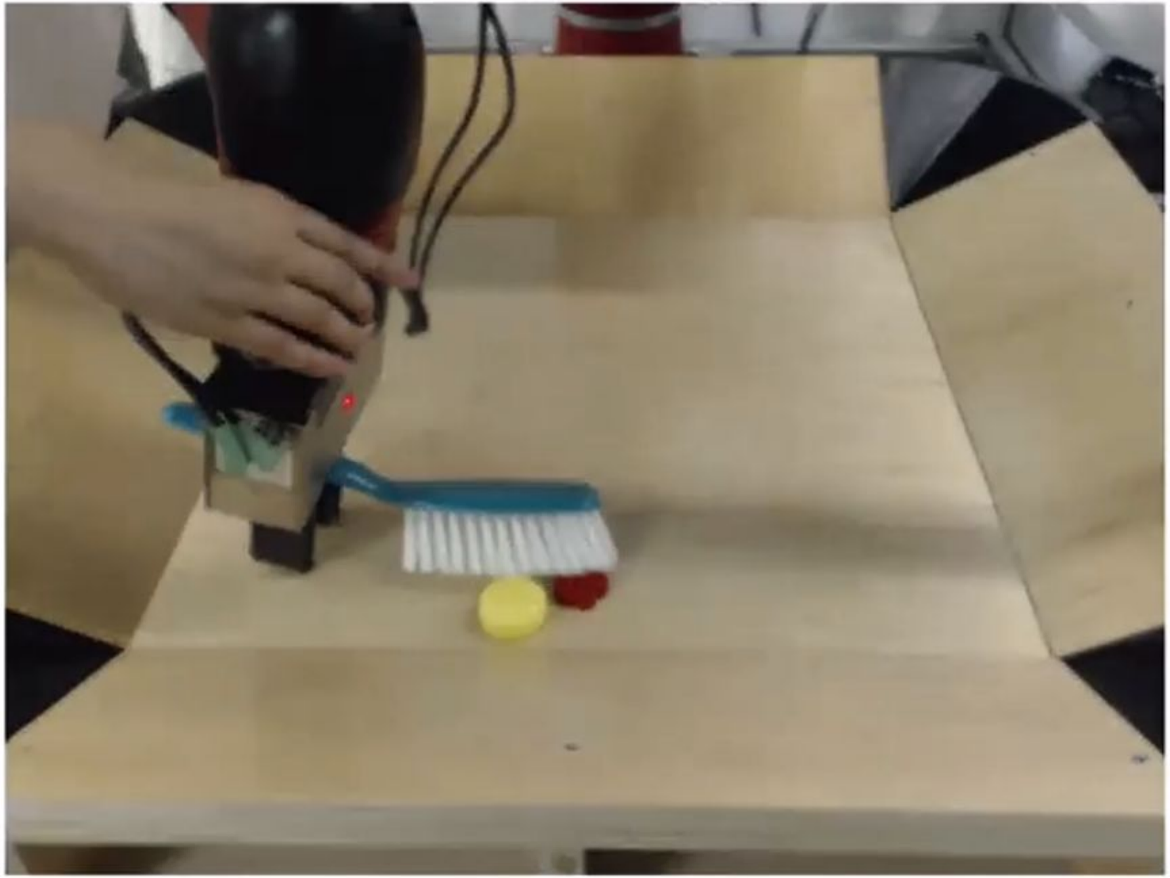
Some remaining questions

Why is the robot so slow?

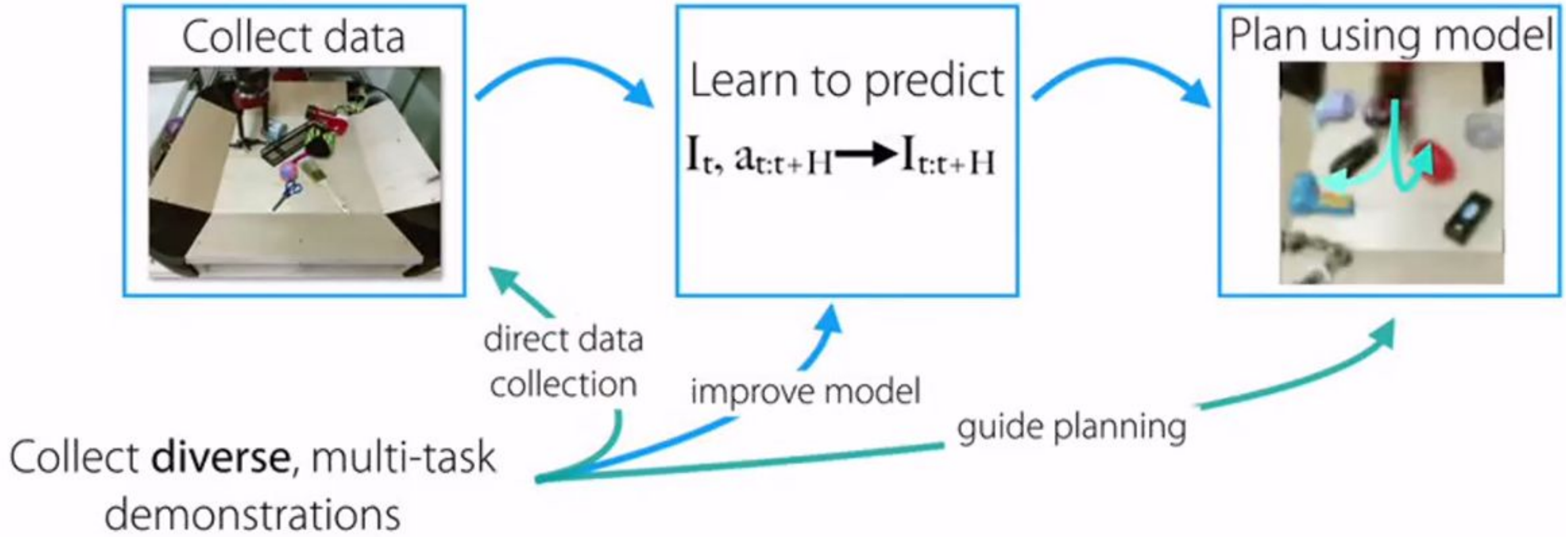
Can we scale these models to broader data?

How can humans help?

Example form of guidance: demonstrating to the robot how to use a tool.



Xie, Ebert, Levine, Finn. Improvisation through Physical Understanding, RSS '19



Fit model of behaviors $p(a_{t:t+H} \mid I_t)$ to the demonstration data.

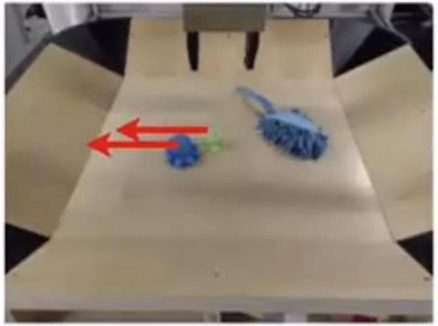
Example multi-task demonstrations:

Samples from behavior model:

Xie, Ebert, Levine, Finn. Improvisation through Physical Understanding, RSS '19

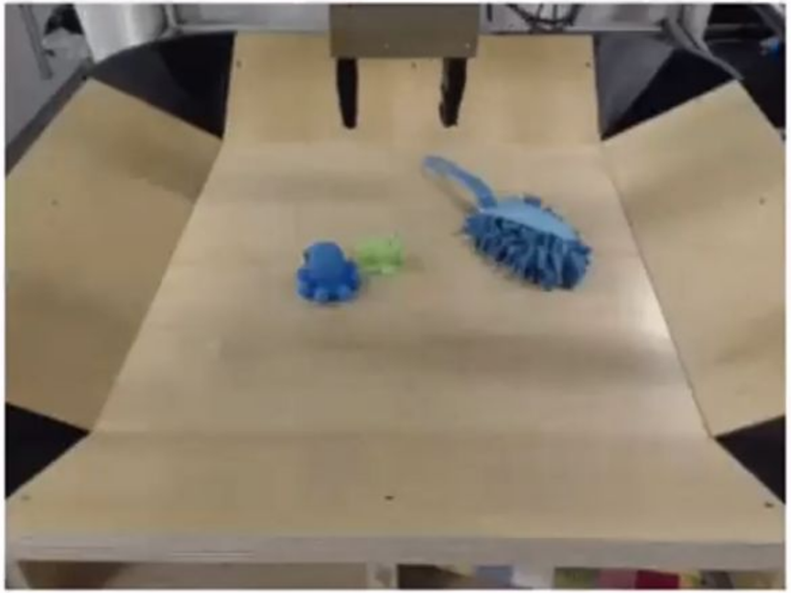
How it works

Specify goal



Guided visual planning w.r.t. goal

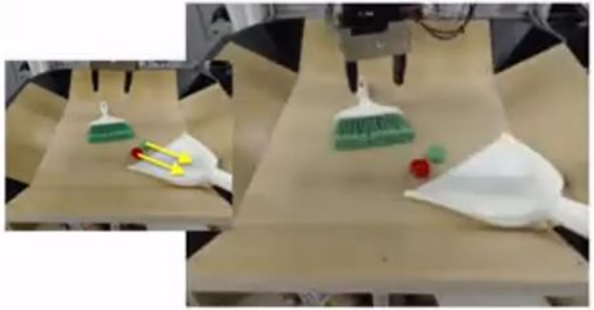
Executing actions



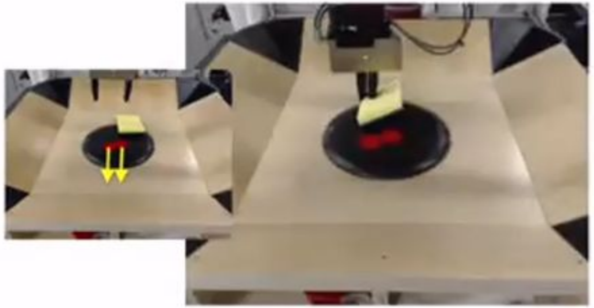
Xie, Ebert, Levine, Finn. Improvisation through Physical Understanding, RSS '19

Planning with a **single model** for many tasks

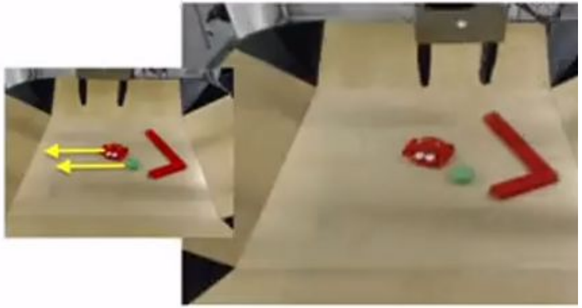
solve new tasks



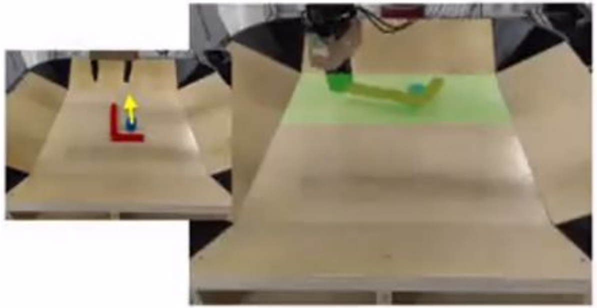
unseen tools



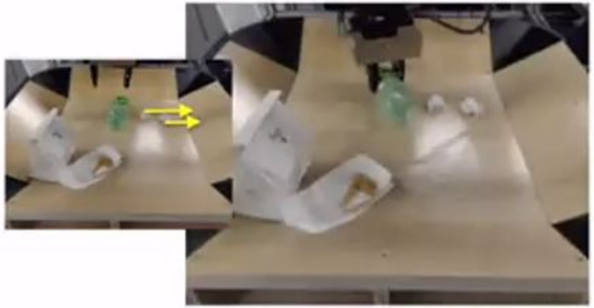
decide when to use a tool…



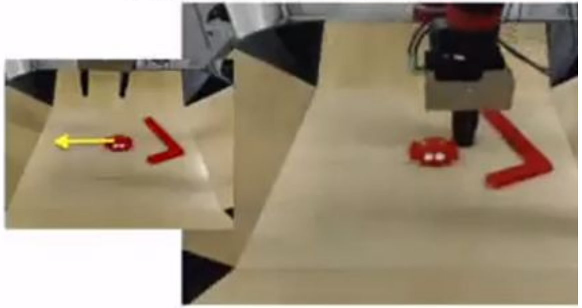
out-of-reach objects



unseen *unconventional* tools



…and when not to



Xie, Ebert, Levine, Finn. Improvisation through Physical Understanding, RSS '19

Takeaways

If we want robots to generalize broadly, train them with broad data.

—> self-supervised learning

How?

(1) train **generative models** of the data

(2) solve tasks via **planning** in this model

Frontiers:

- offline policy optimization
- scaling to larger models
- incorporating human guidance

Students

Annie Chen Suraj Nair Alex Nam Sudeep Dasari Frederik Ebert Sergey Levine Kostas Daniilidis

Annie Xie Bohan Wu Karl Schmeckpeper Oleh Rybkin Roberto Martin-Martin Fei-Fei Li