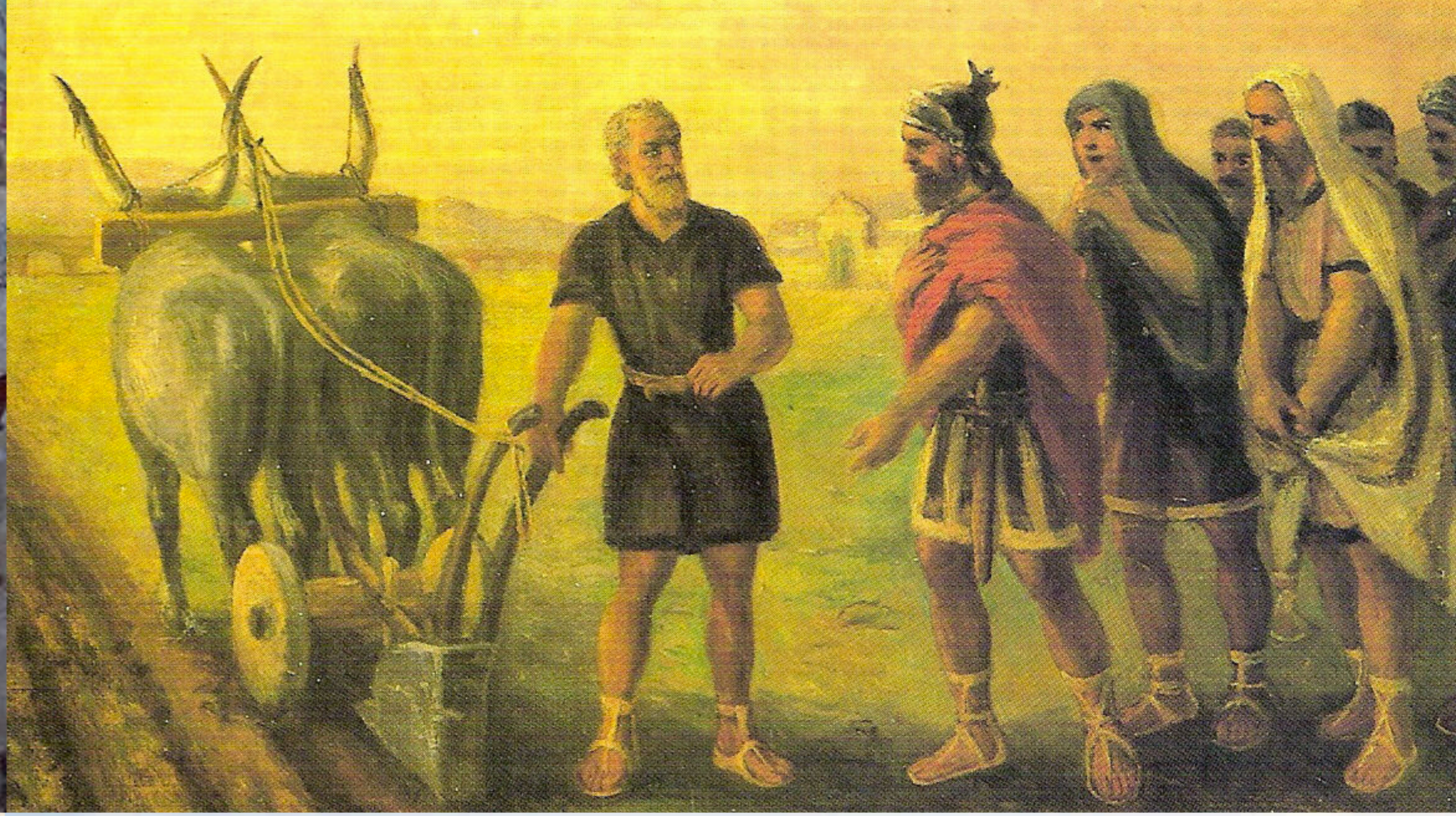
Closing the Perception-Action Loop

Towards General-Purpose Robot Autonomy

Yuke Zhu

Stanford University

STANFORD VISION & LEARNING


CLAAS
PHOTOGRAPHY

70
302331
302331
302159
DHL
microdrones.com
PACKSET
M

Traditional form of **automation**

Credit: WIRED

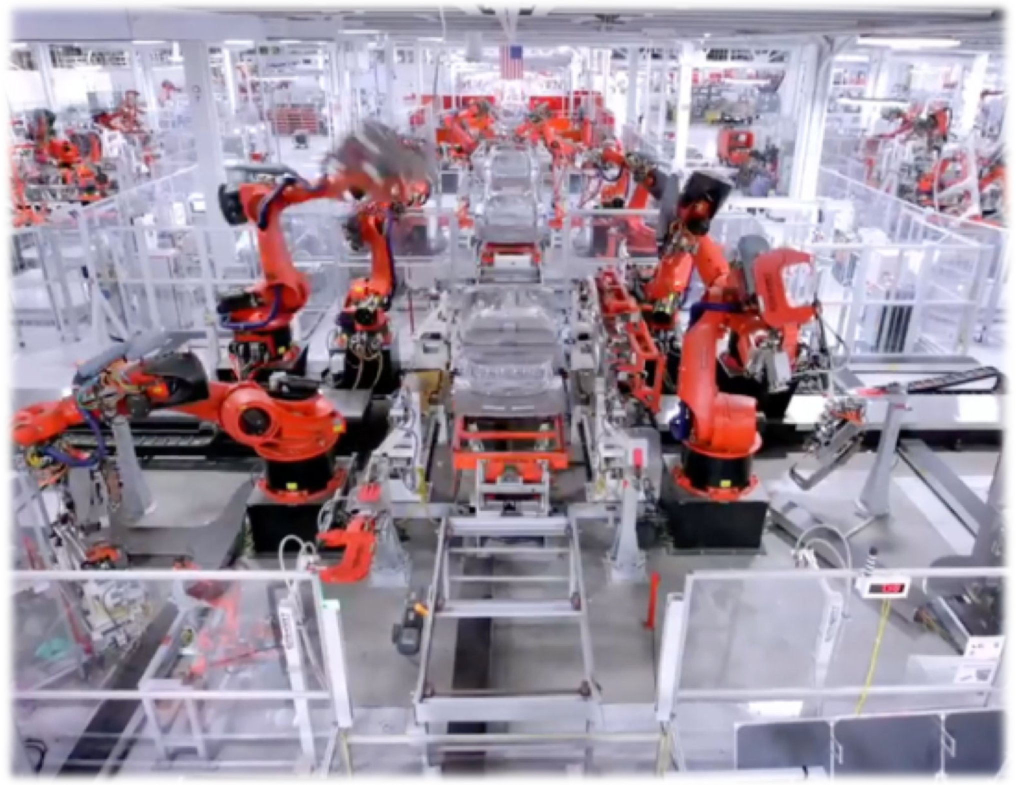
General-purpose robot hardware

Credit: Kinova Robotics

My research goal:

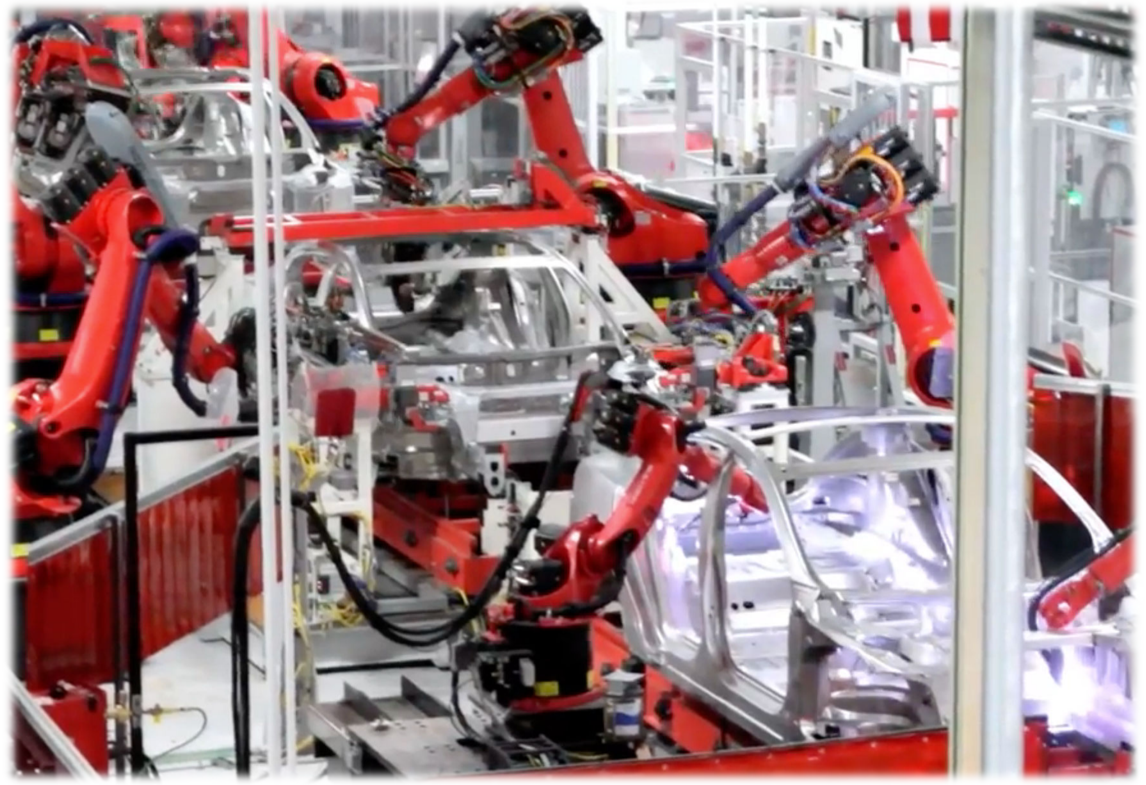
building robot intelligence to enrich human intelligence

Traditional form of **automation**



custom-built
robots

human expert
programming

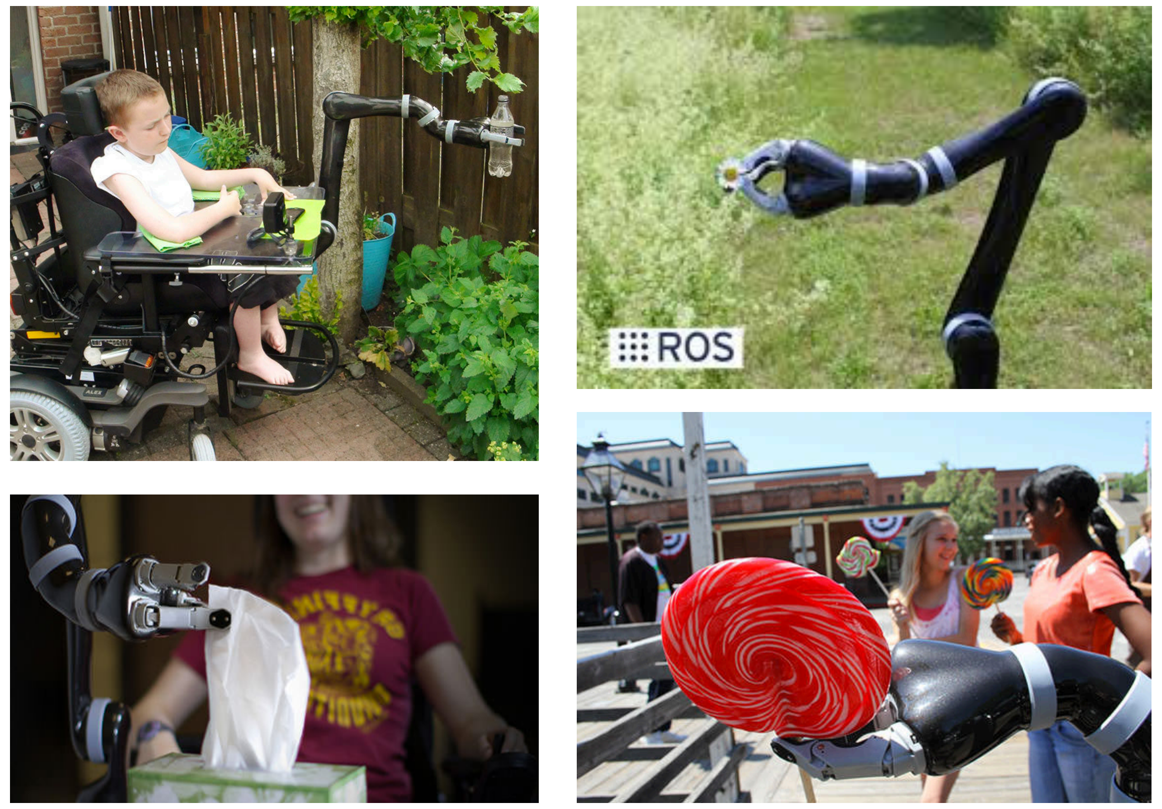


special-purpose
behaviors

New form of **automation**

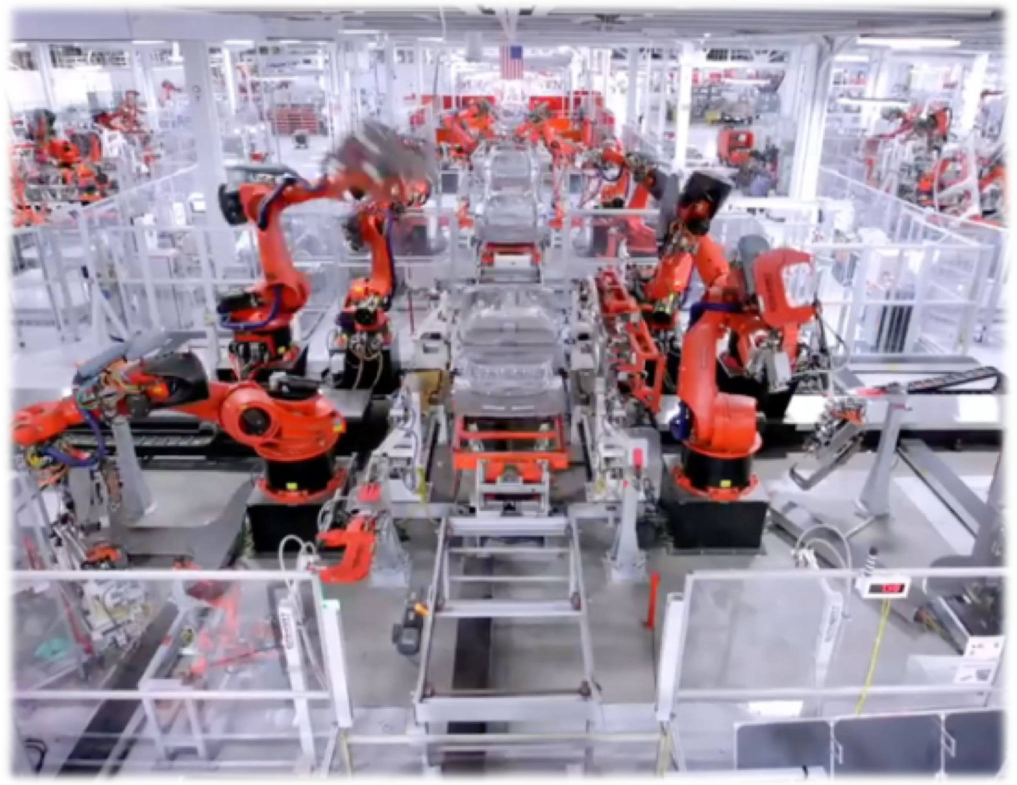
general-purpose
robots

?



general-purpose
behaviors

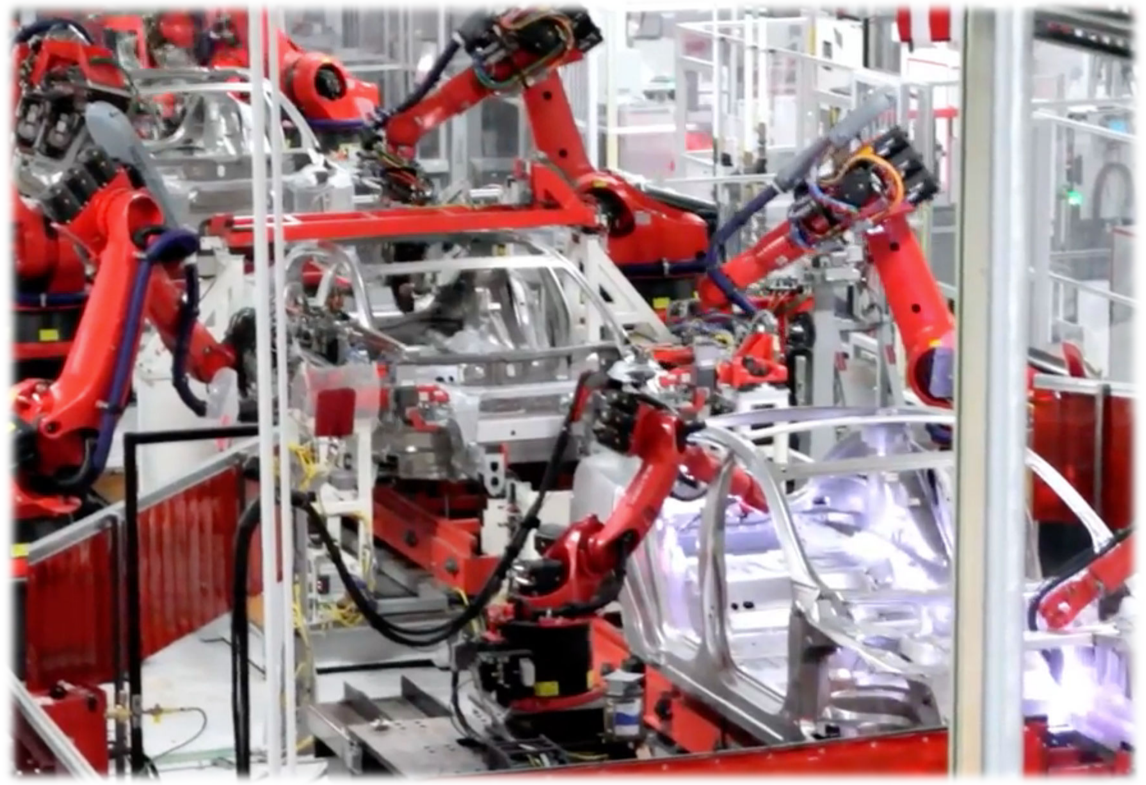
Traditional form of **automation**



custom-built
robots

structured environment

human expert
programming

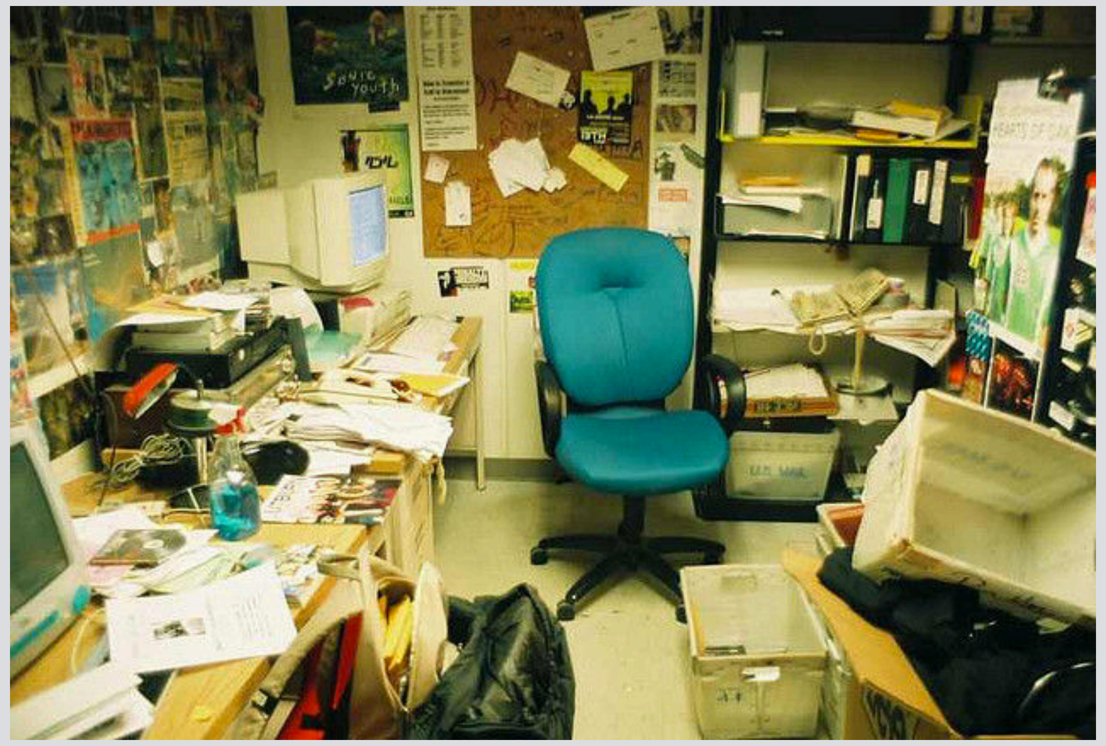


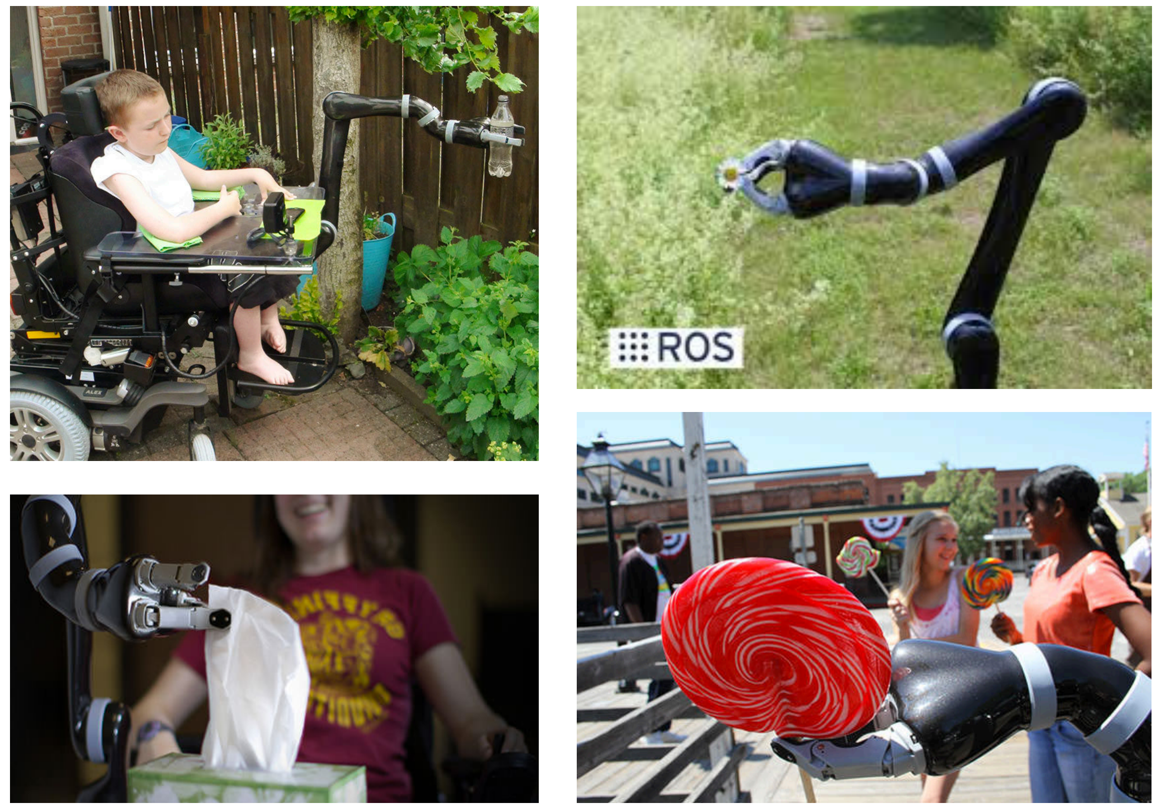
special-purpose
behaviors

New form of **automation**

general-purpose
robots

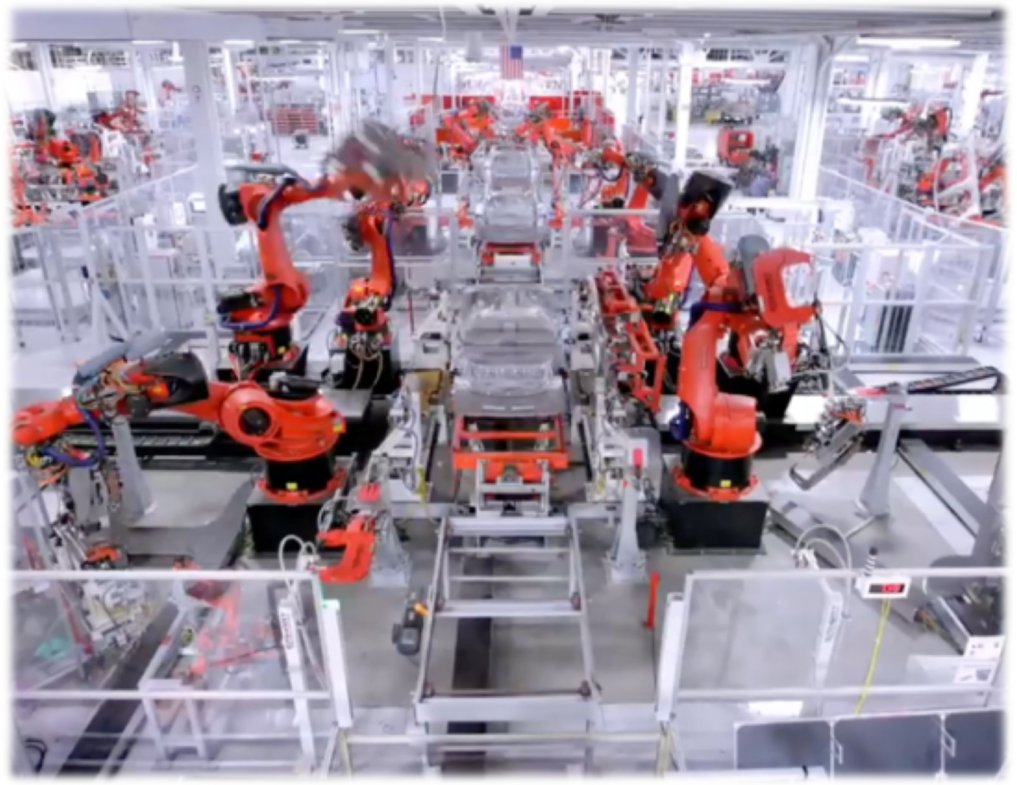
unstructured environment





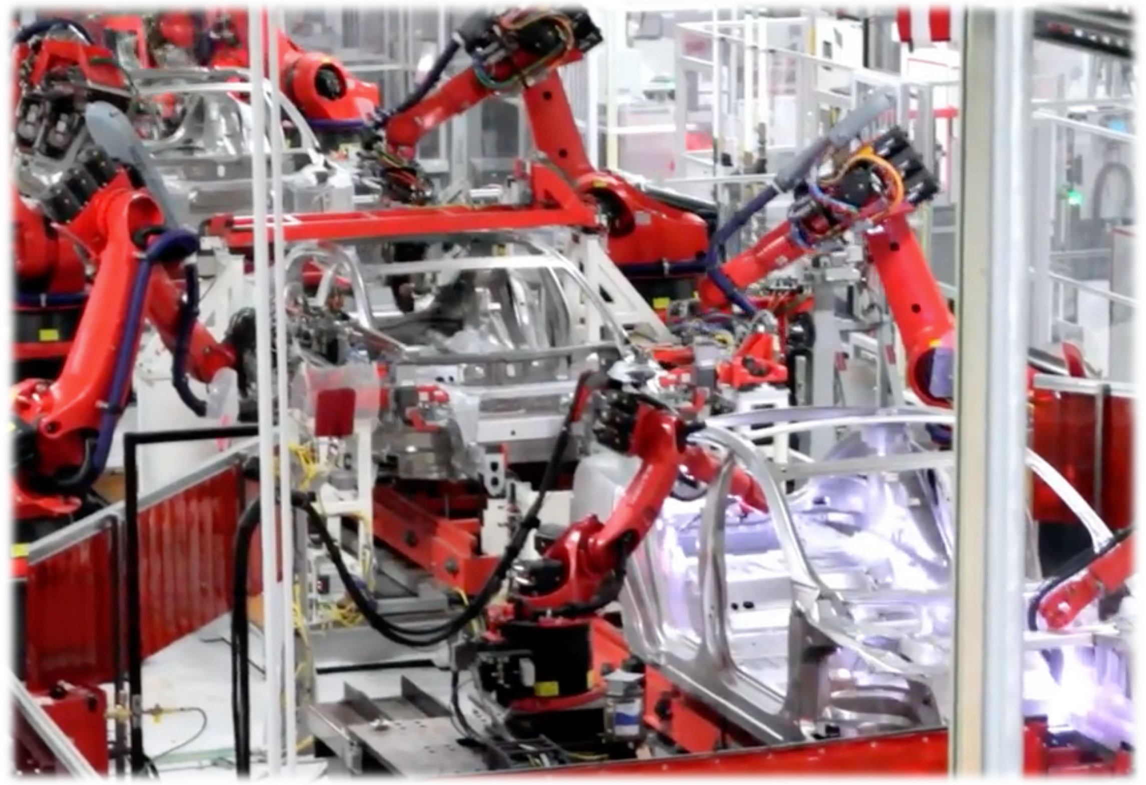
general-purpose
behaviors

Traditional form of **automation**



custom-built
robots

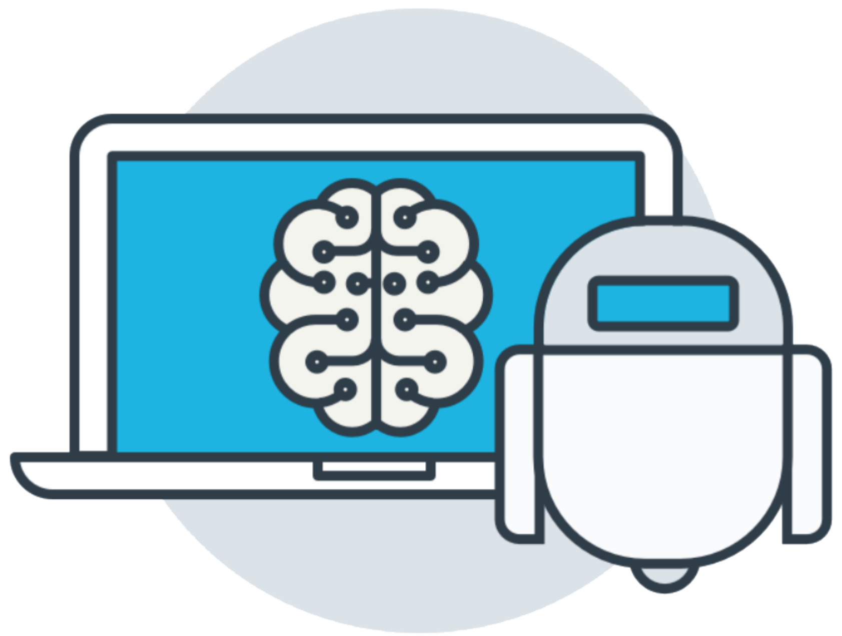
human expert
programming



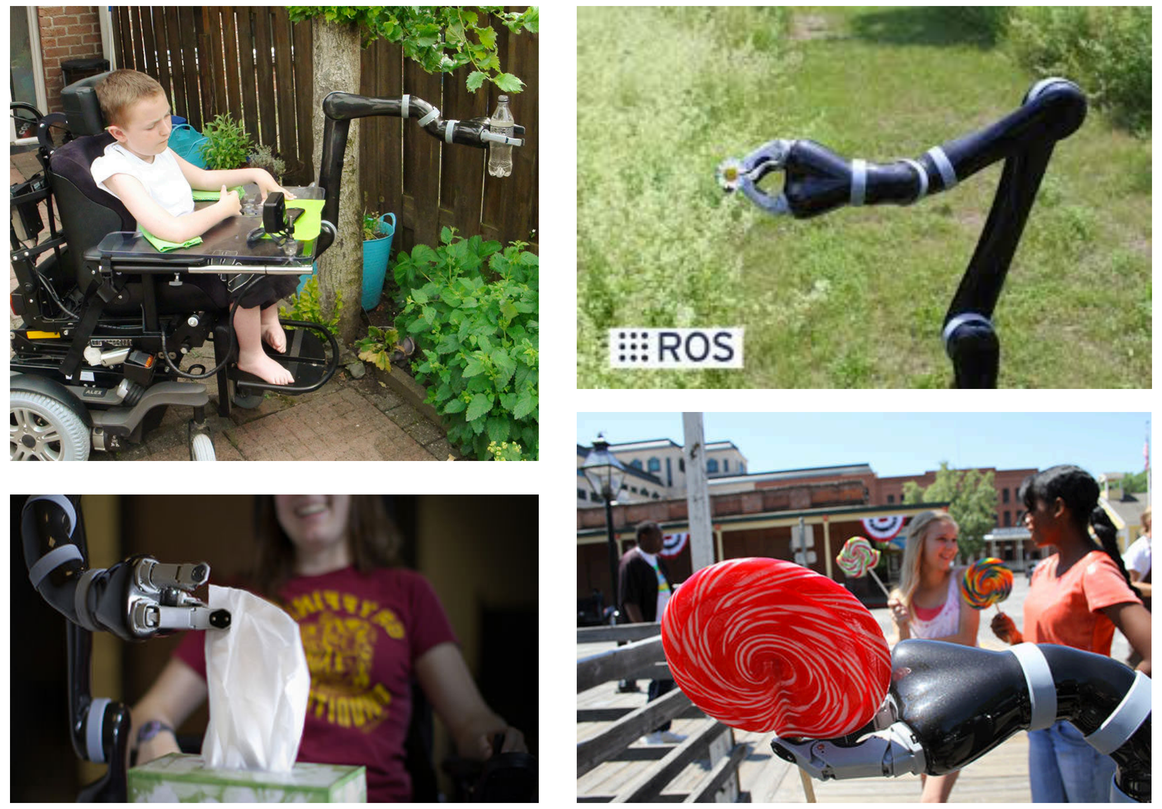
special-purpose
behaviors

New form of **automation**

general-purpose
robots



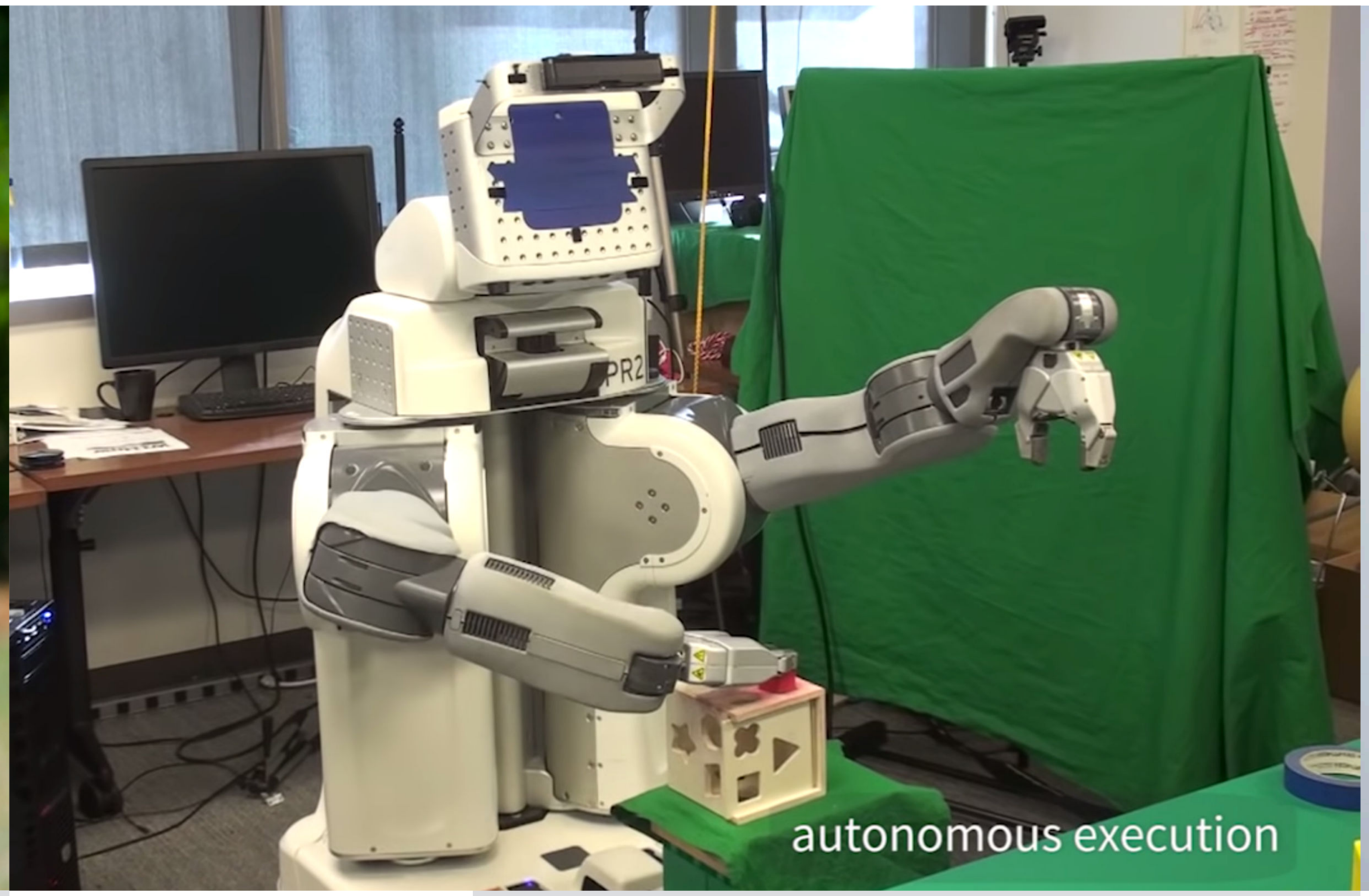
machine learning
& perception



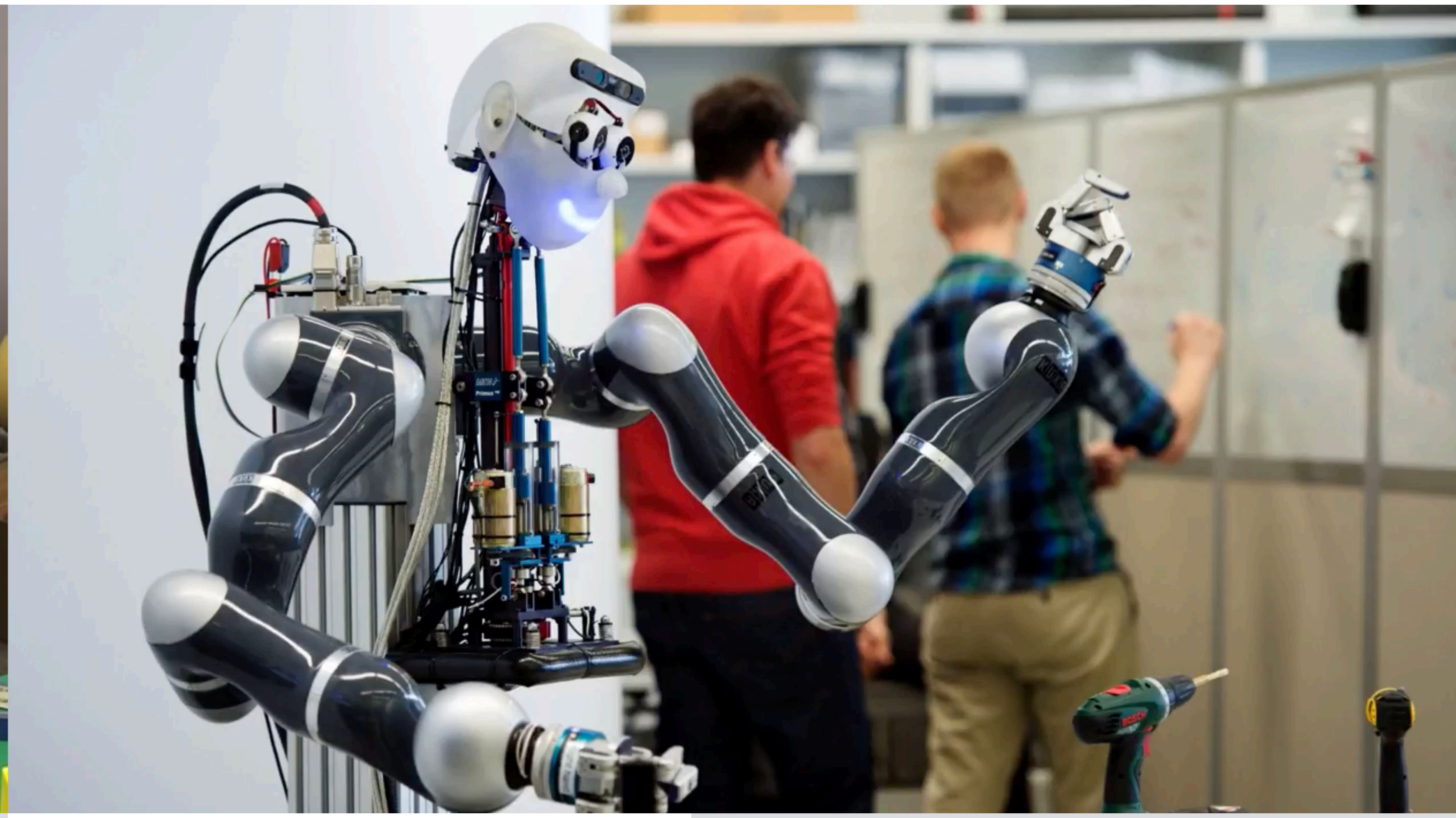
general-purpose
behaviors

closing the perception-action loop

[Sa et al. IROS 2014]



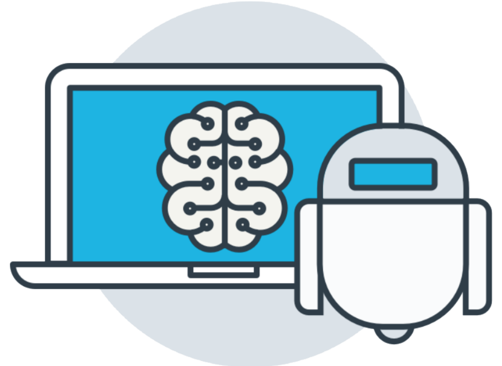
[Levine et al. JMLR 2016]



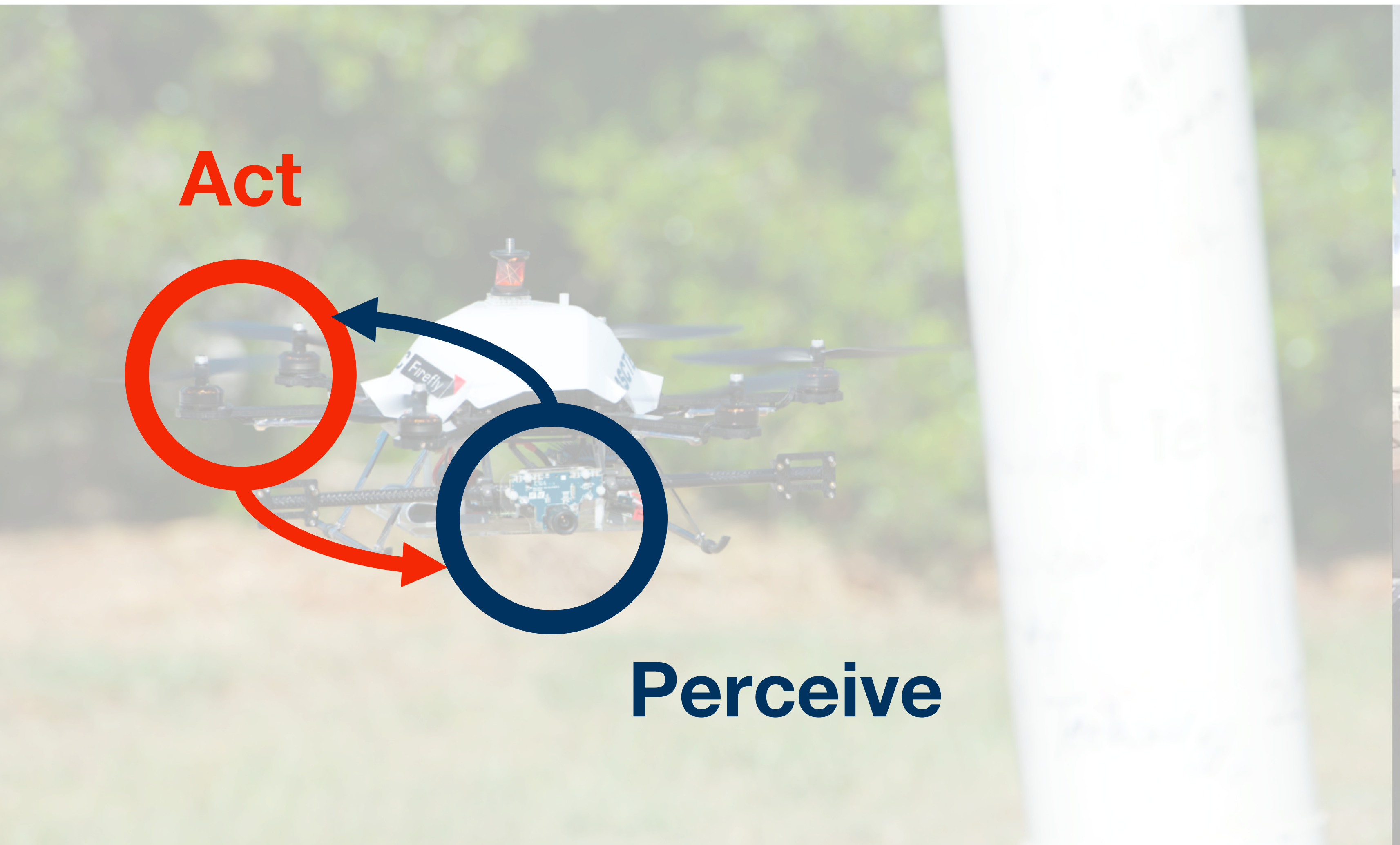
[Bohg et al. ICRA 2018]

[Gibson 1979; Bajcsy 1988; Ballard 1991; Espiau et al. 1992; Hutchinson et al. 1996; Hamel & Mahony 2002; Kragic & Christensen 2002; Jonschkowski & Brock 2015; Levine et al. 2016; Agrawal et al. 2016; Bojarski et al. 2016; Finn & Levine 2017; Florence et al. 2018]

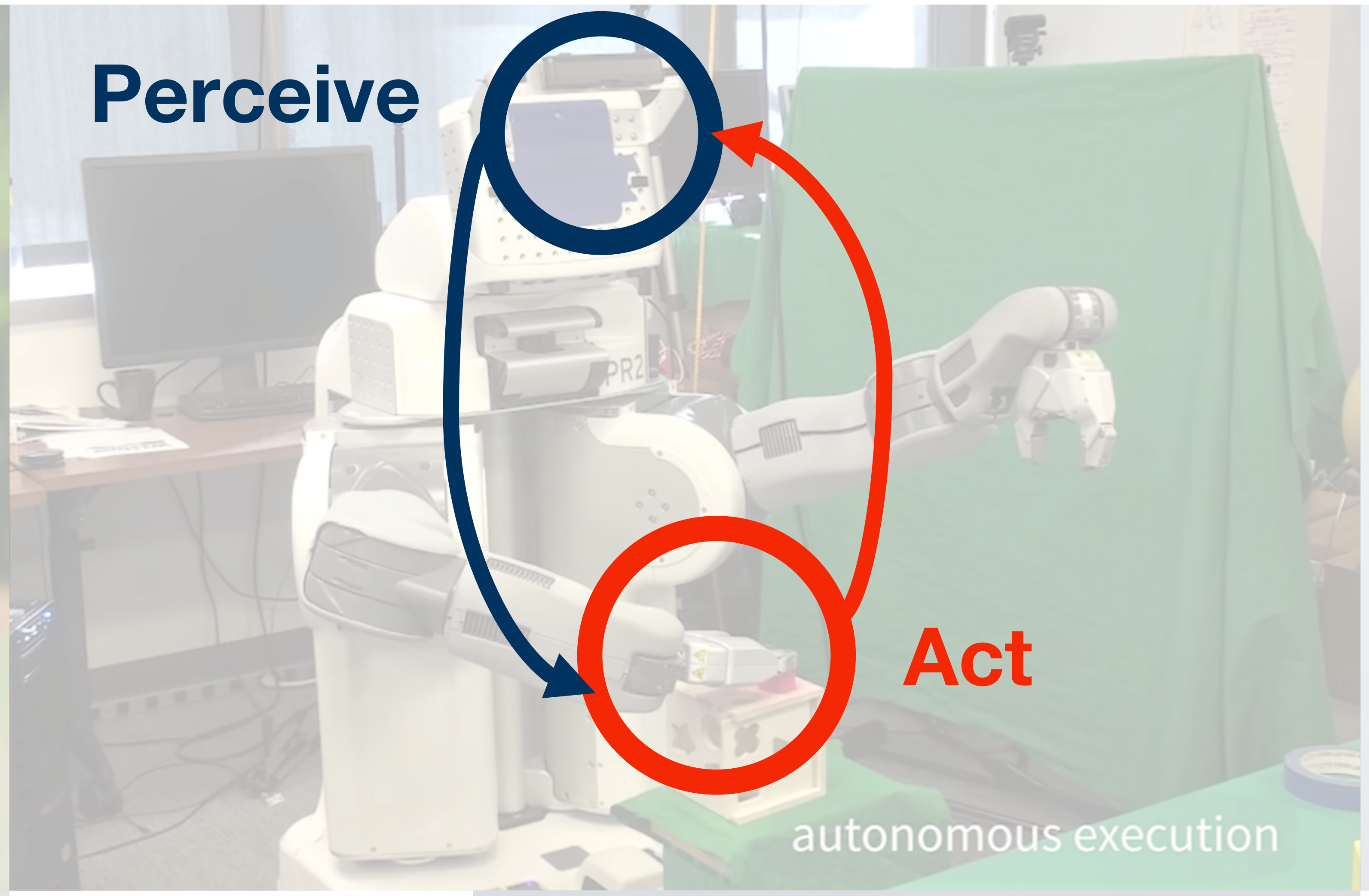
closing the perception-action loop



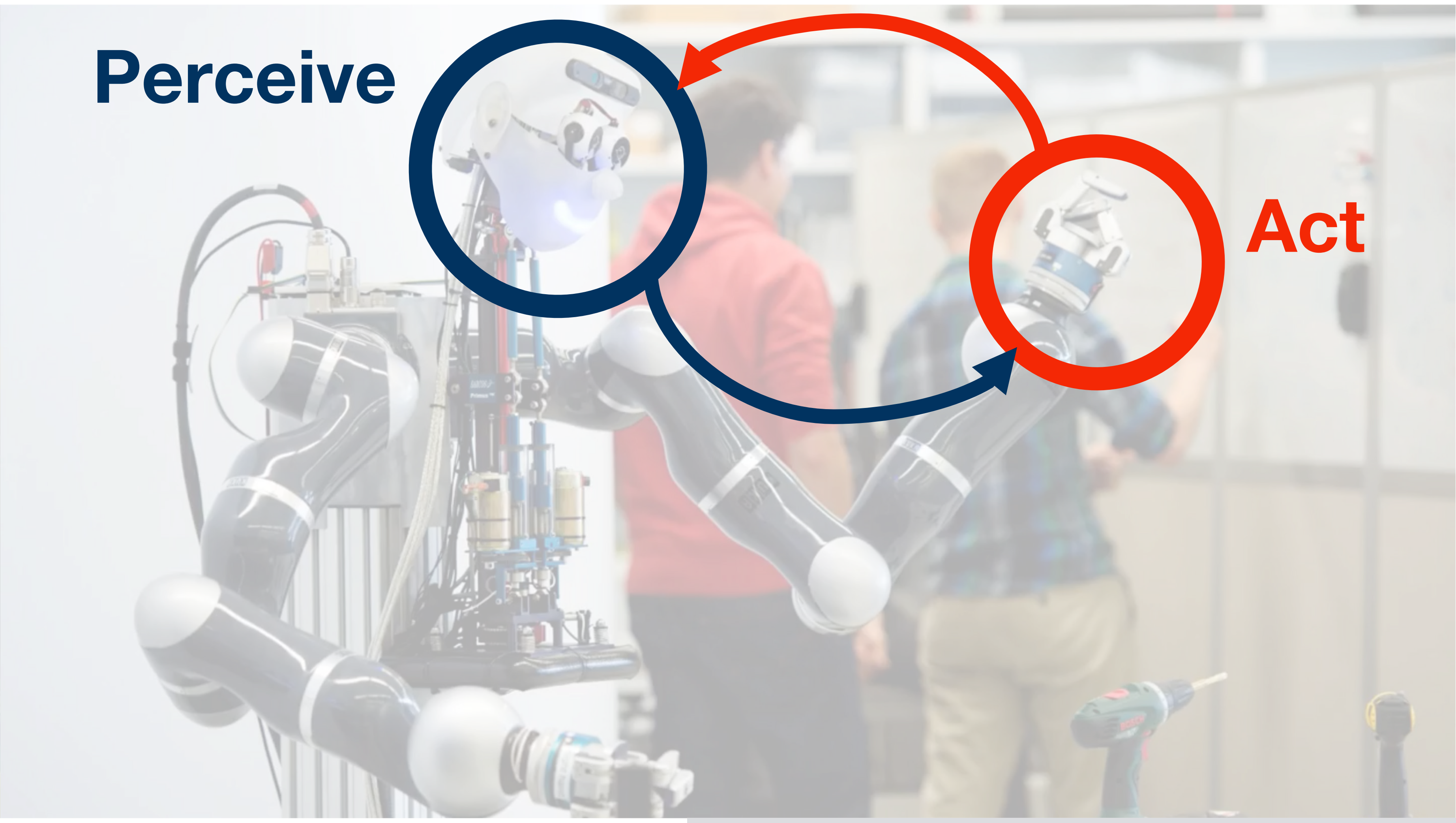
with machine learning & perception in robotics



[Sa et al. IROS 2014]



[Levine et al. JMLR 2016]



[Bohg et al. ICRA 2018]

[Gibson 1979; Bajcsy 1988; Ballard 1991; Espiau et al. 1992; Hutchinson et al. 1996; Hamel & Mahony 2002; Kragic & Christensen 2002; Jonschkowski & Brock 2015; Levine et al. 2016; Agrawal et al. 2016; Bojarski et al. 2016; Finn & Levine 2017; Florence et al. 2018]

closing the perception-action loop

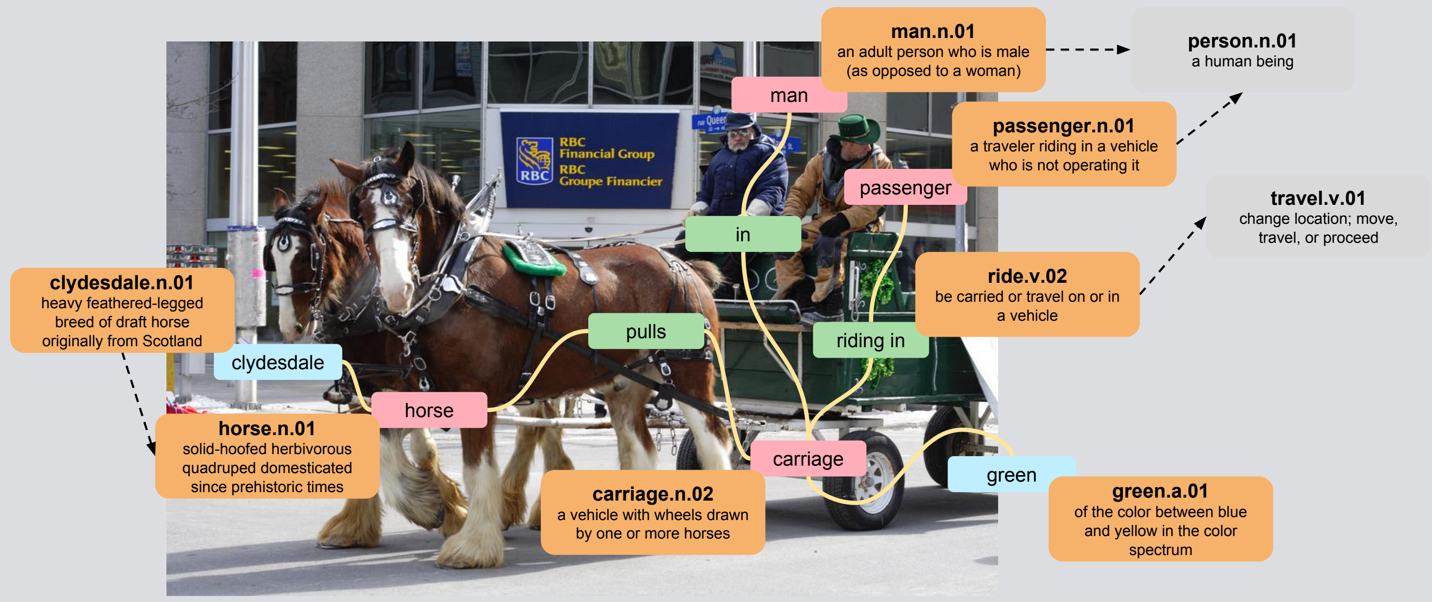
Perception is a weak link in the loop.

“Clean” kitchen for state-of-the-art robotics

“Messy” kitchen in the real world

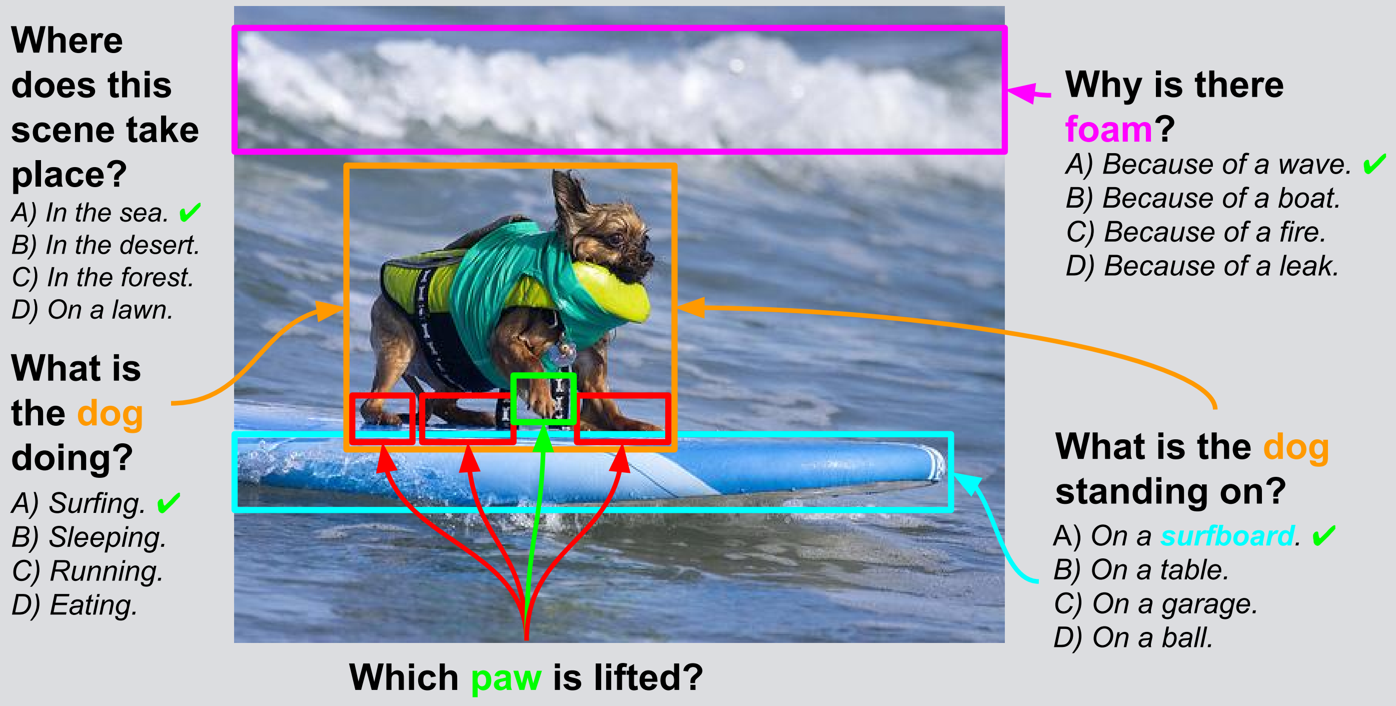
closing the perception-action loop

scene understanding



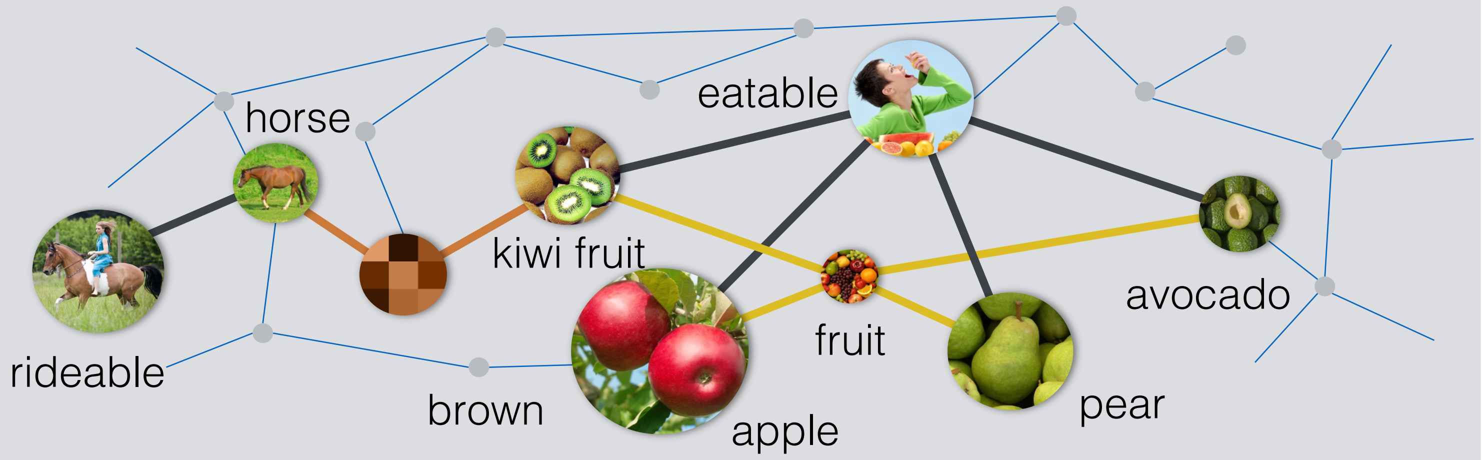
ICCV’15, CVPR’17a, CVPR’19

vision & language



CVPR’16, IJCV’17, CVPR’17b

affordance reasoning



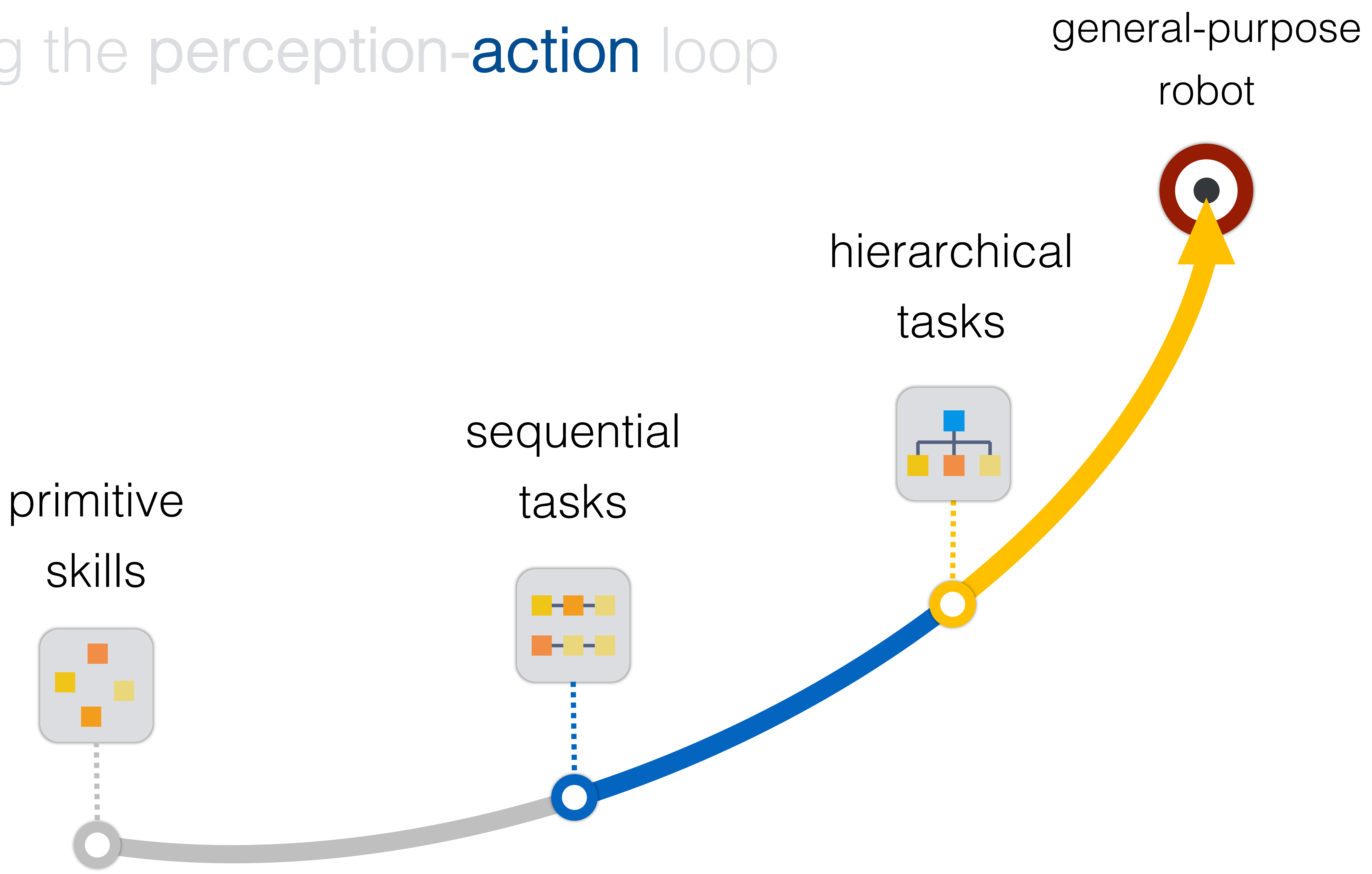
ECCV’14, arXiv’15

Johnson et al. CVPR'16; Krishna, **Zhu**, et al. IJCV'17

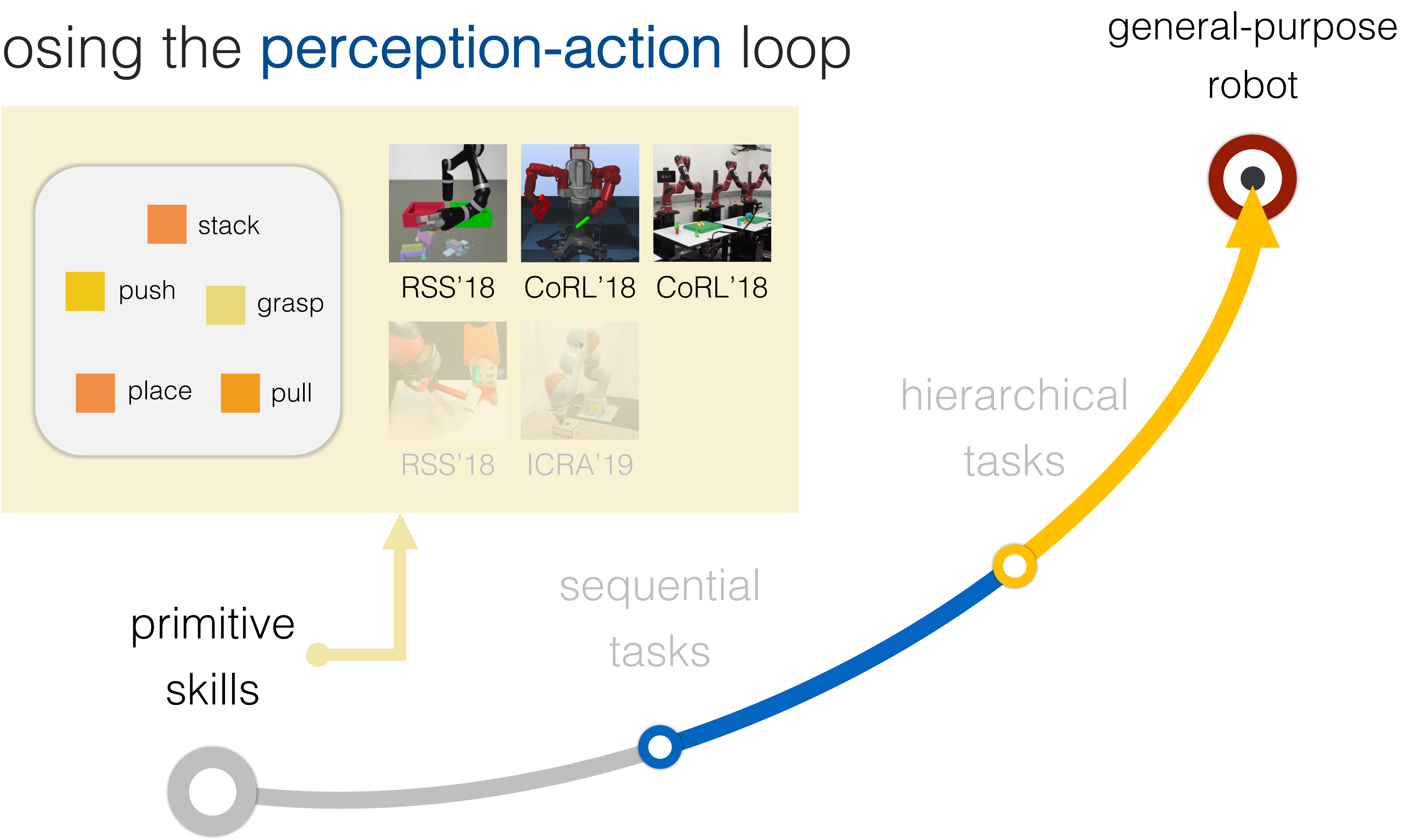
Zhu et al. CVPR’16, **Zhu** et al. CVPR’17

Xu, **Zhu**, Choy, Fei-Fei, CVPR'17

closing the perception-action loop



closing the perception-action loop



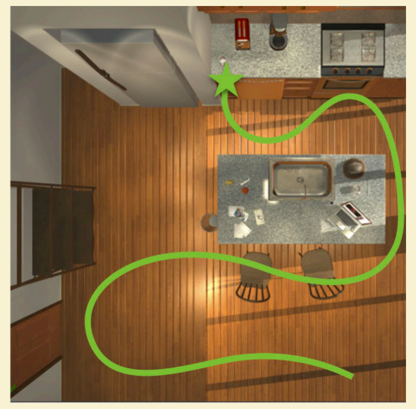
closing the perception-action loop

general-purpose robot

"get apple from fridge"

"put bowl in the sink"

ICRA'17



ICCV'17

CVPR'19

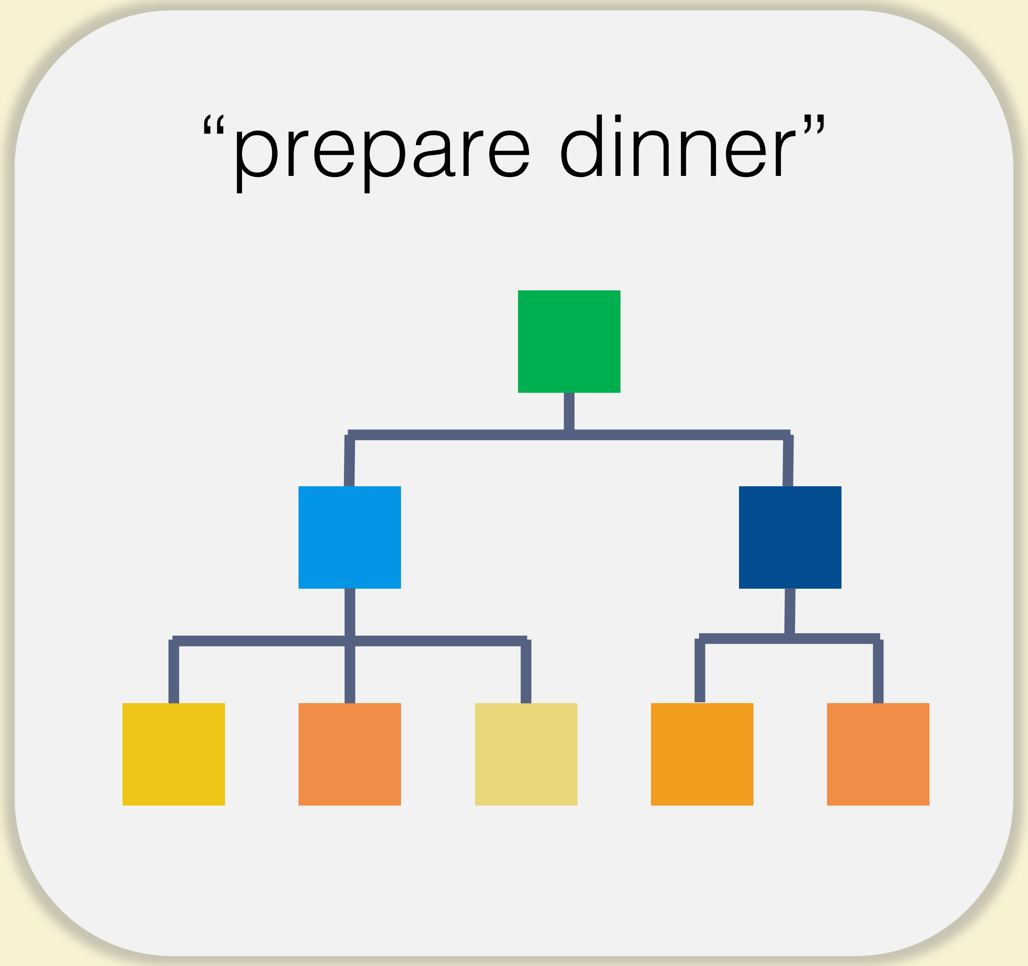
hierarchical tasks

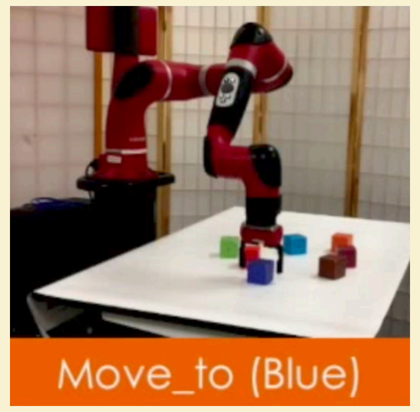
sequential tasks

primitive skills

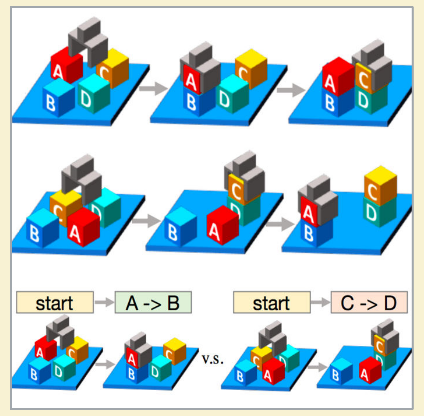
closing the perception-action loop

general-purpose
robot

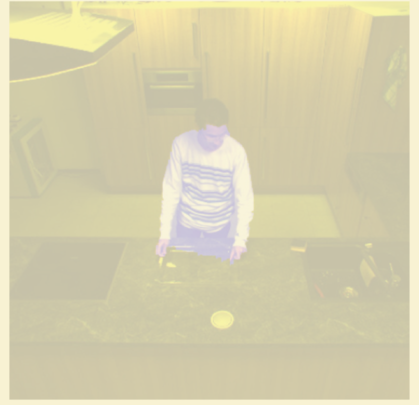




ICRA'18



CVPR'19

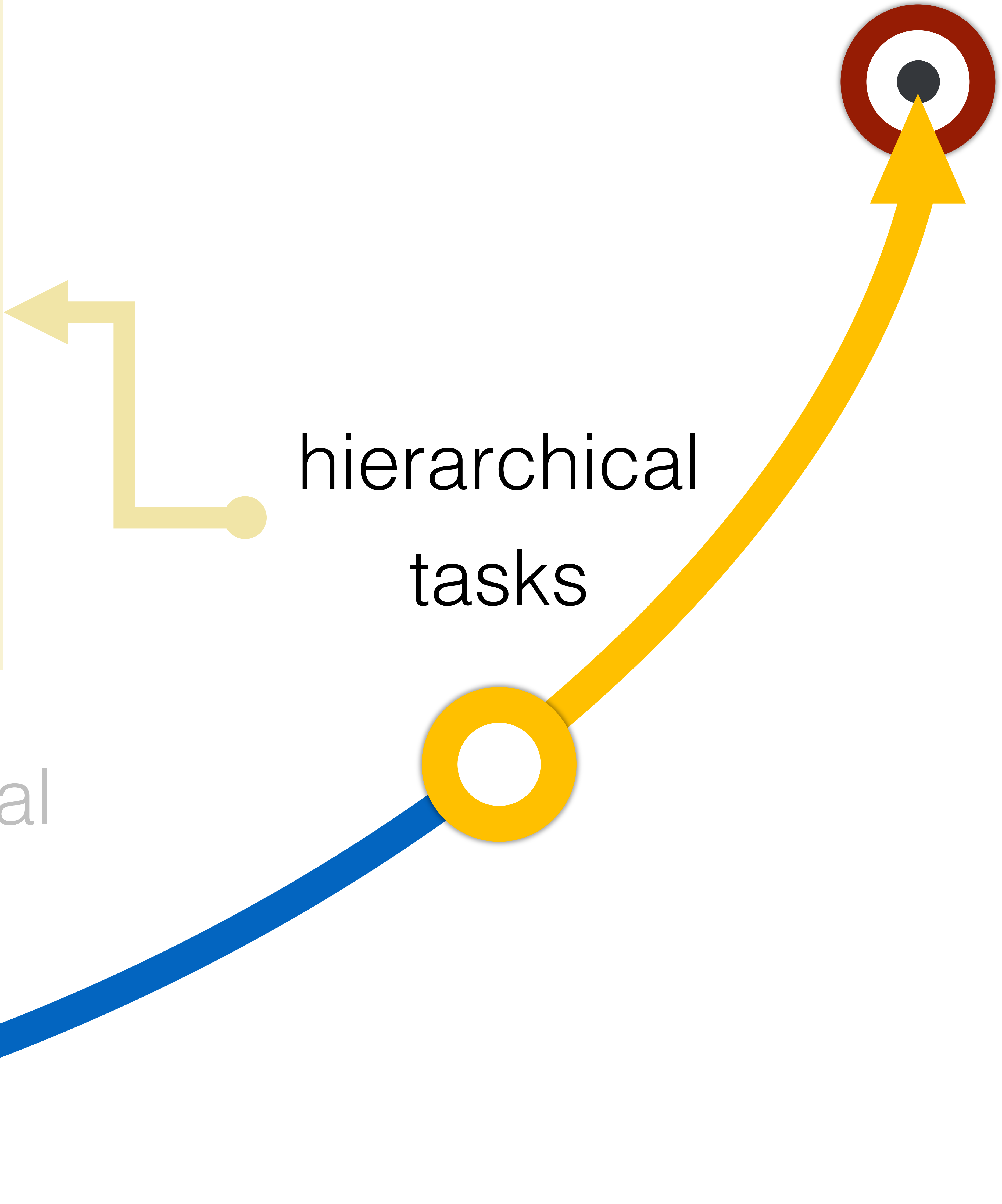


ICCV'15

hierarchical
tasks

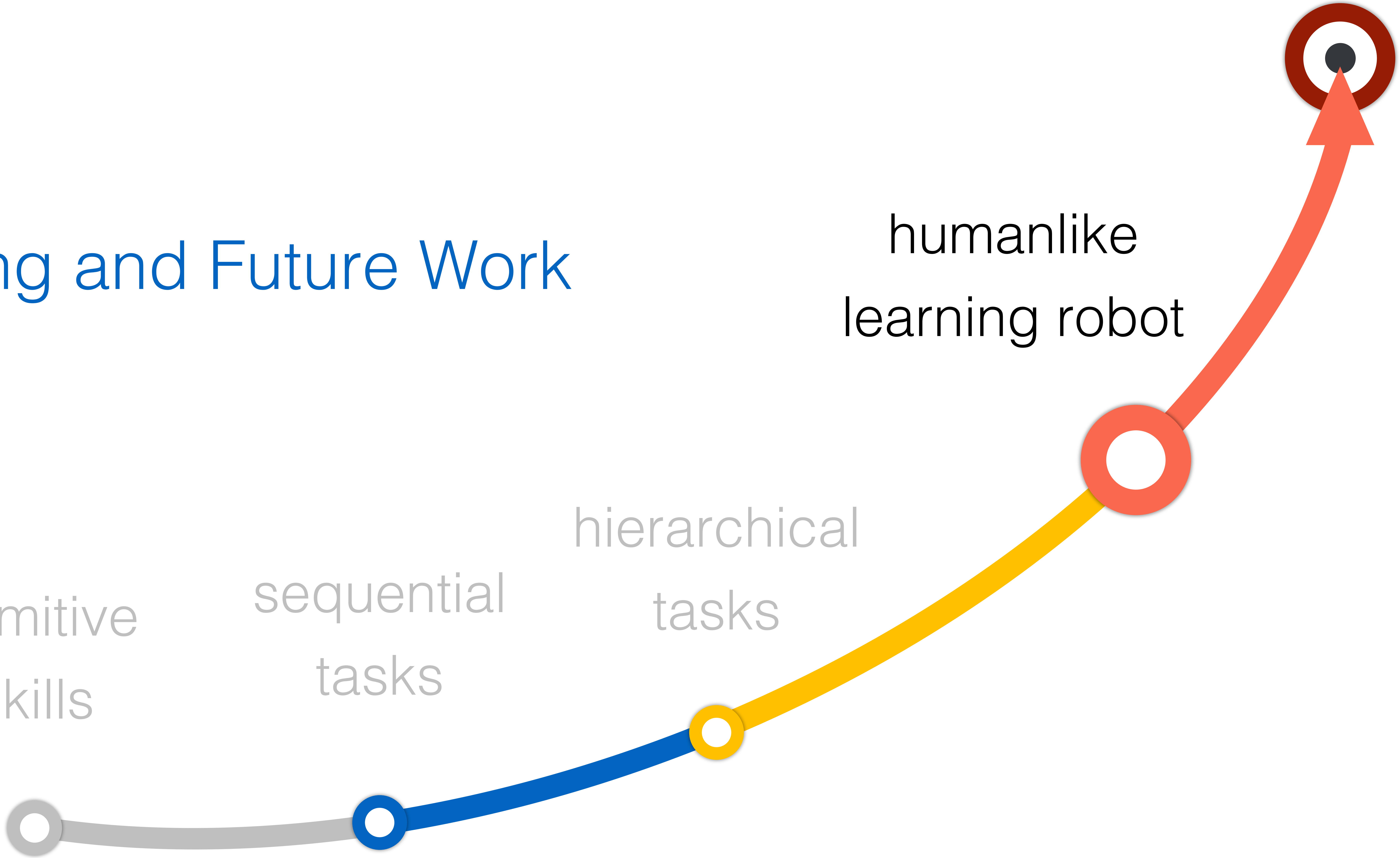
sequential
tasks

primitive
skills



closing the **perception-action** loop

Ongoing and Future Work



Part I: Primitive Skills

[**Zhu** et al., RSS 2018] Reinforcement and Imitation Learning […]

[Mandlekar, **Zhu** et al., CoRL 2018] RoboTurk: A Crowdsourcing Platform for Robotic Skill Learning […]

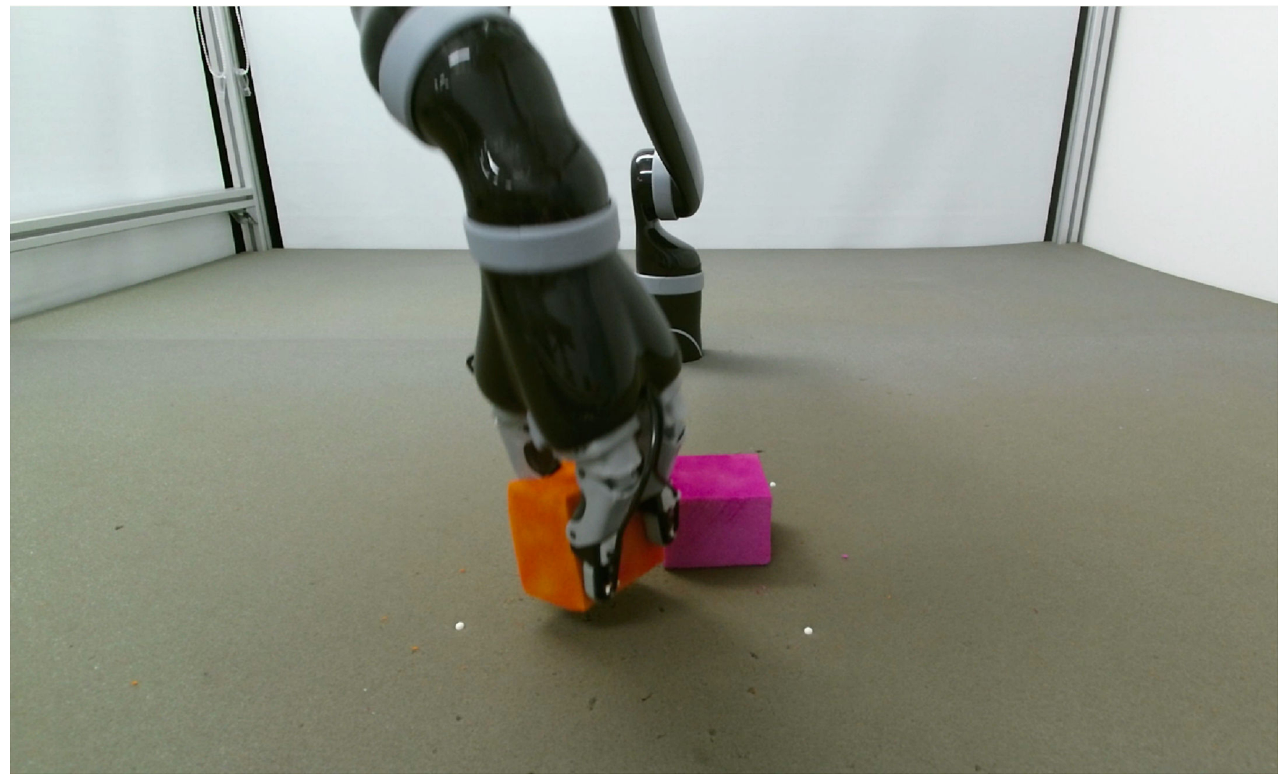
[**Zhu***, Fan* et al., CoRL 2018] SURREAL: Open-Source Reinforcement Learning Framework […]

Primitive Skills (Reaching, Grasping, Stacking, etc.)

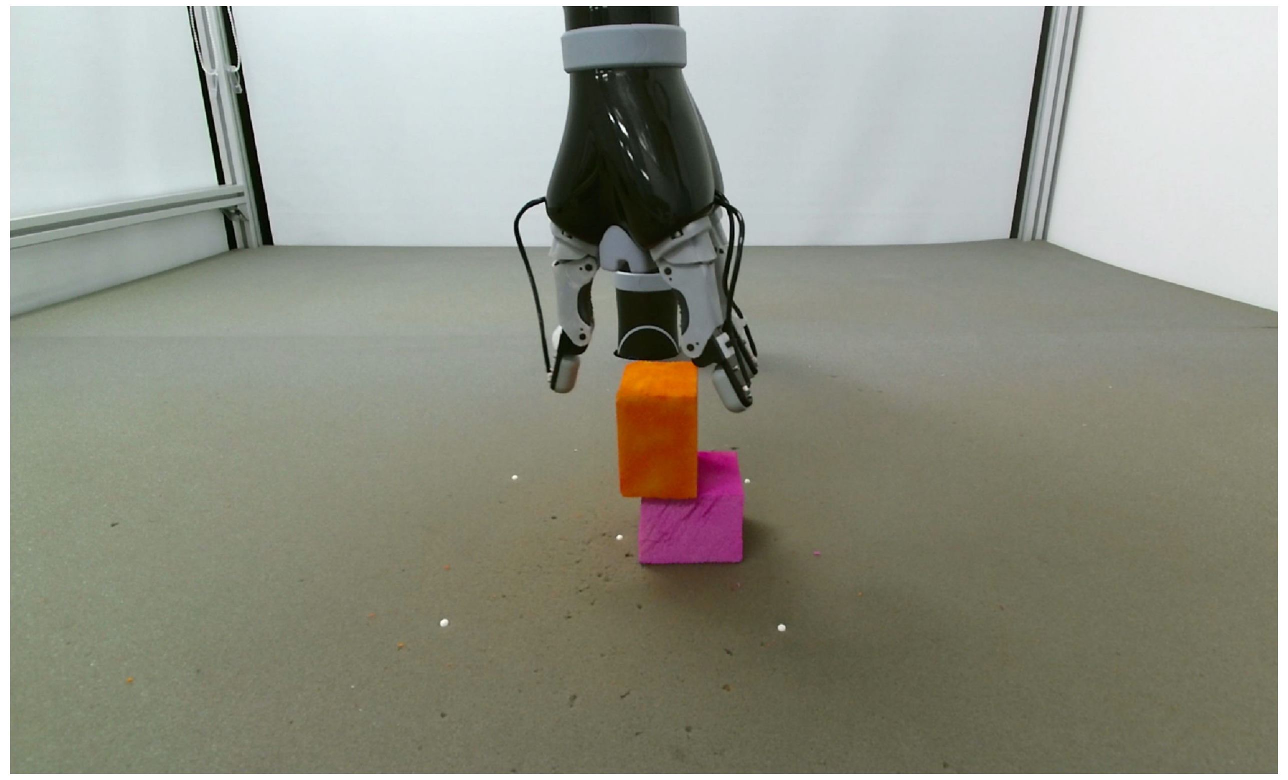
perception

action

Credit: BBC Earth Lab



sensory data

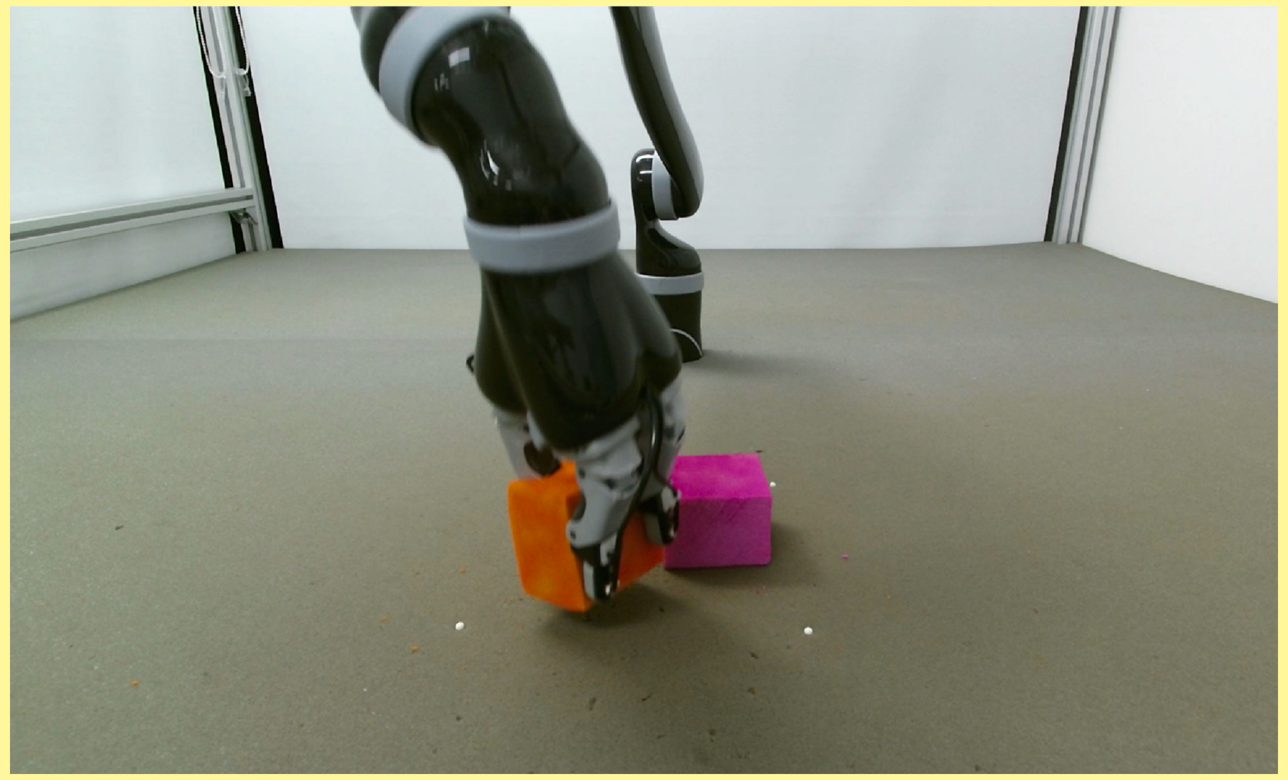


motor command

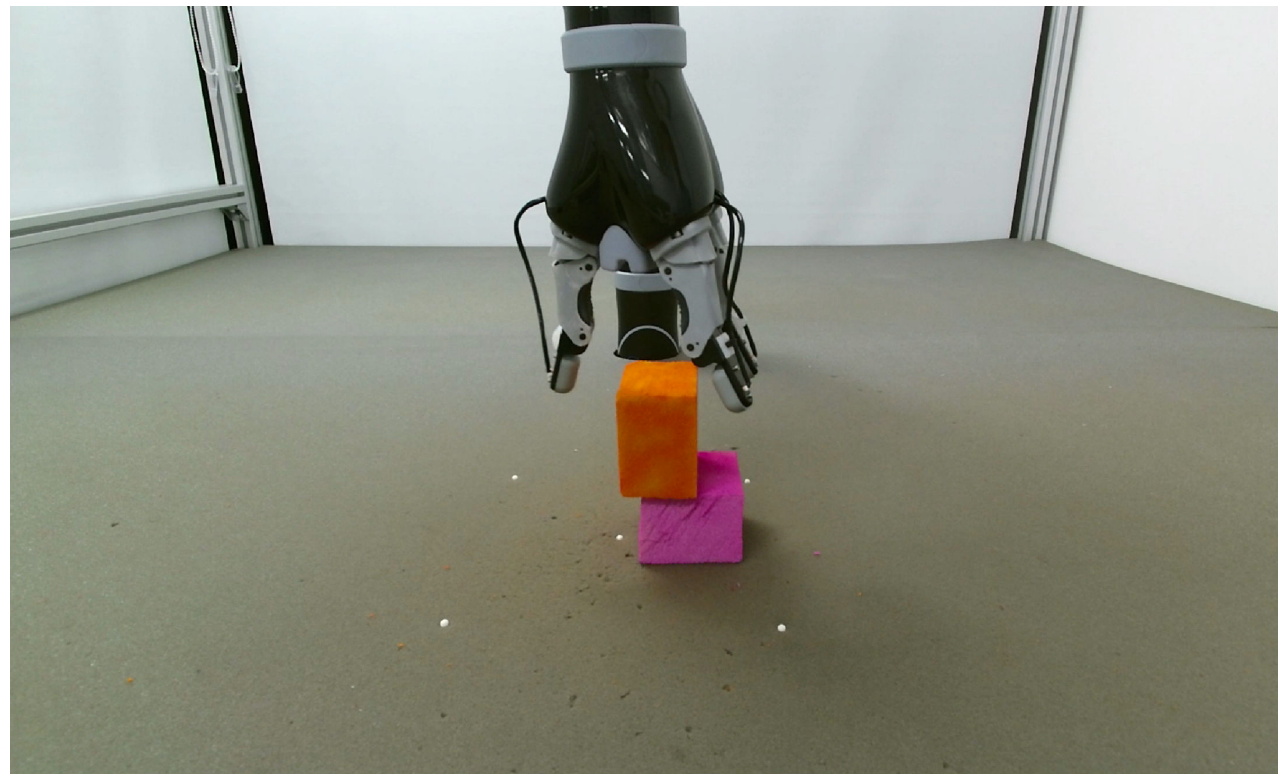
Primitive Skills (Reaching, Grasping, Stacking, etc.)

Challenge: Raw sensory data are high-dimensional, noisy, and multimodal.

End-to-end learning of robust representations for control



sensory data



motor command

Learning Primitive Skills in Robotics

reinforcement learning (RL)

learning through trial-and-error

requiring a lot of training data

[Mahadevan & Connell 1992; Gullapalli et al. 1994; Atkeson 1994; Schaal 1996; Bagnell & Schneider 2001; Smart & Kaelbling 2002; Kohl & Stone 2004; Kober & Peters 2009; Theodorou et al. 2010; Levine et al. 2015; Gu et al. 2016; Peng et al. 2017; Kalashnikov et al. 2018]

Learning Primitive Skills in Robotics

reinforcement learning (RL)

imitation learning (IL)

learning through trial-and-error

learning from demonstrations

requiring a lot of training data

limited by a suboptimal teacher

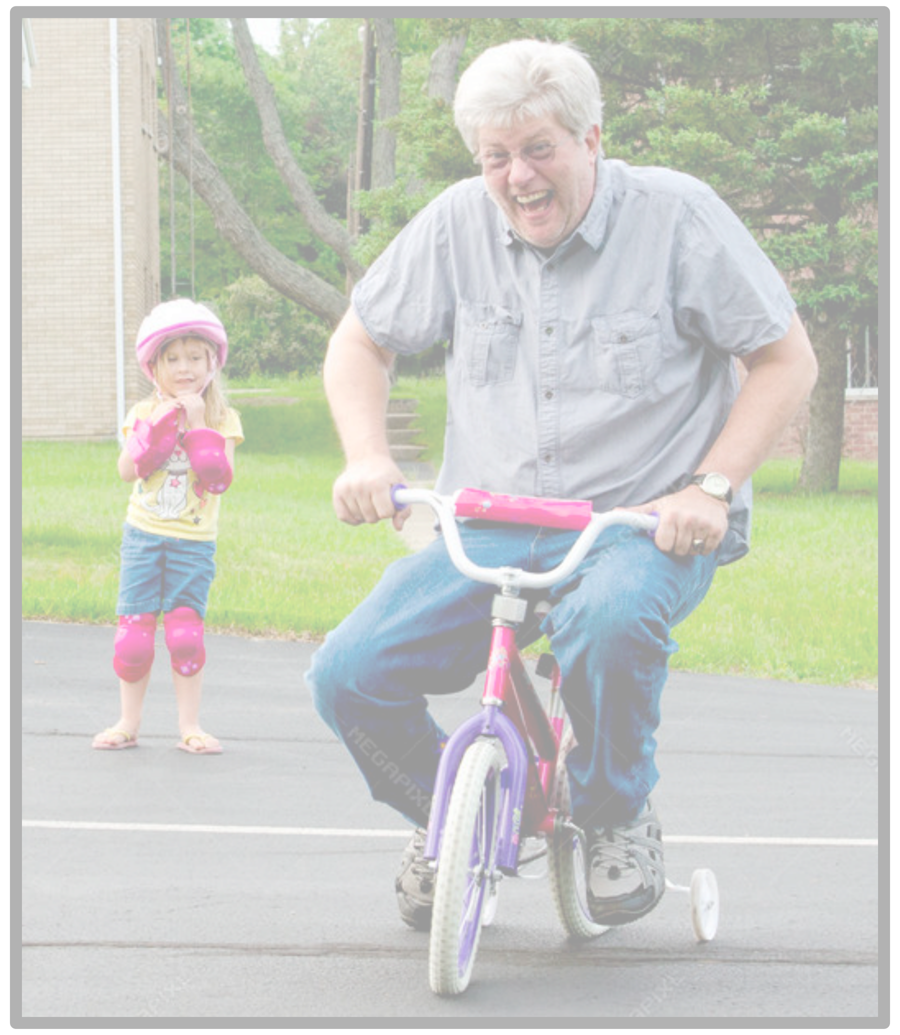
[Schaal 1996; Pollard & Hodgins 2002; Abbeel & Ng 2004; Brian et al. 2008; Pastor et al. 2009; Ross et al. 2011; Akgun et al. 2012; Bagnell 2015; Finn et al., 2016; Rahmatizadeh et al., 2017; James et al., 2017; Sermanet et al. 2017; Menda et al. 2017; Le Paine et al. 2018]

Learning Primitive Skills in Robotics

reinforcement learning (RL)

learning through trial-and-error

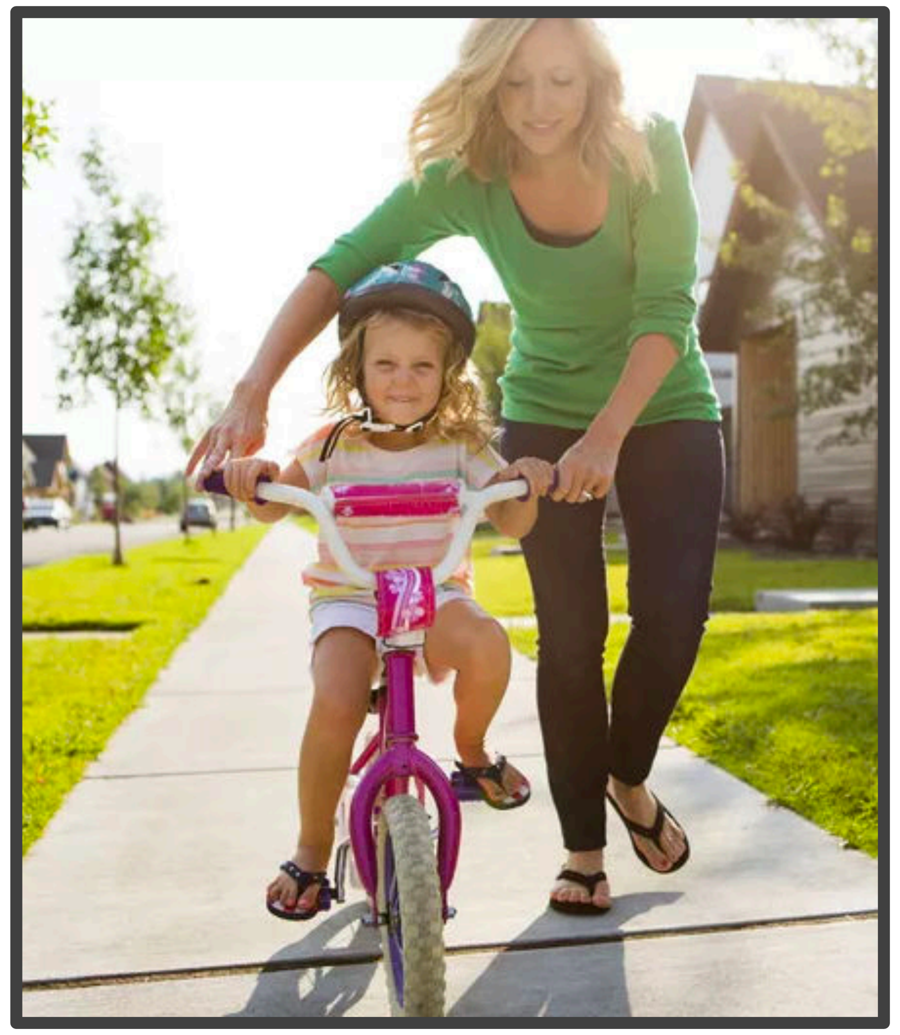
requiring a lot of training data



imitation learning (IL)

learning from demonstrations

limited by a suboptimal teacher



combined (RL+IL)

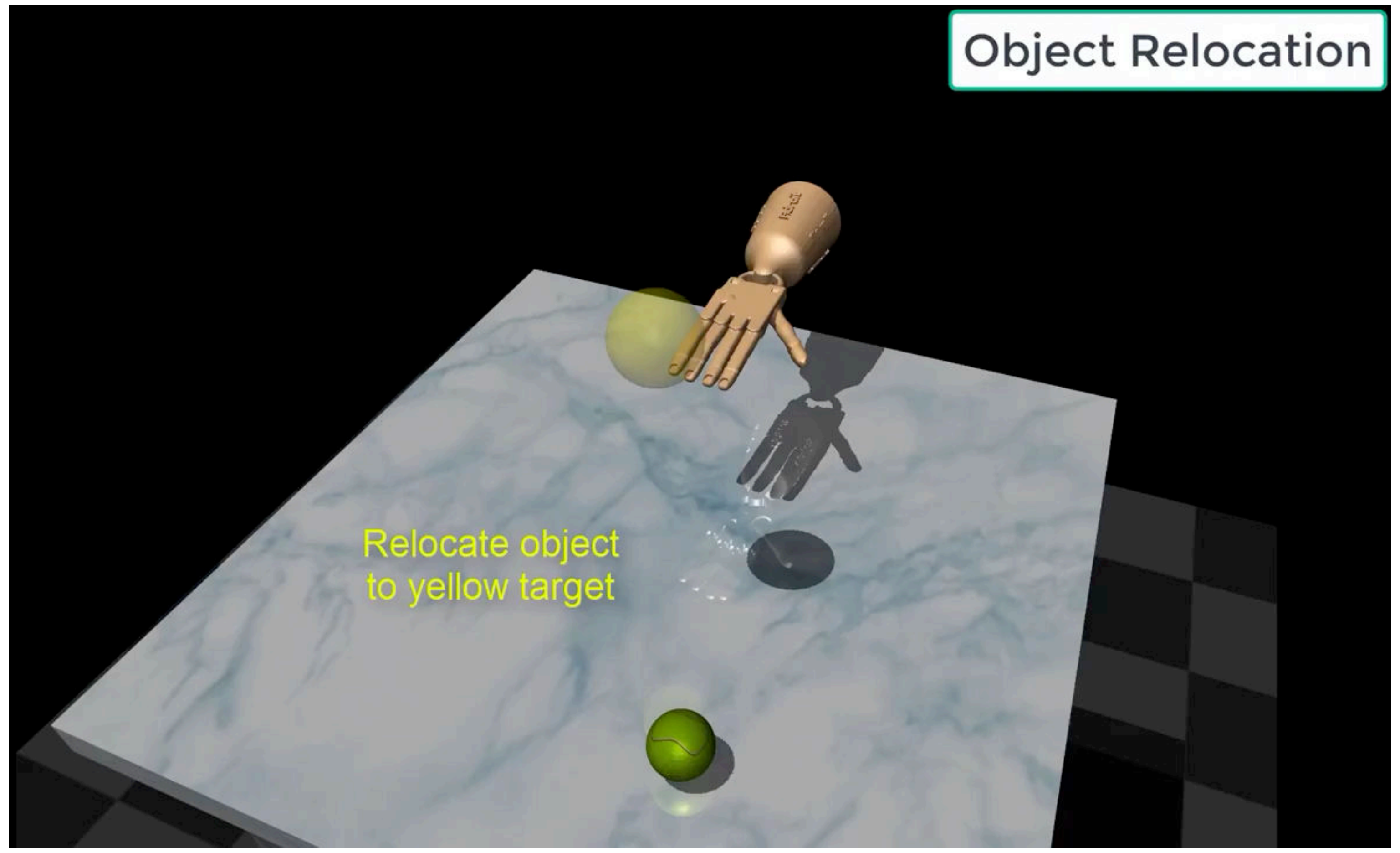
demonstrations offer guidance

better performance by trial-and-error

[Price & Boutilier 2003; Bentivegna et al. 2004; Latzke et al. 2007; Conn & Peters 2007; Ross & Bagnell 2014; Kumar et al. 2016; Sun et al. 2017; Vecerik et al. 2007; Rajeswaran et al. 2018; Cheng et al. 2018; Nair et al. 2018; Pfeiffer et al. 2018; Codevilla et al. 2018]

Reinforcement and Imitation Learning

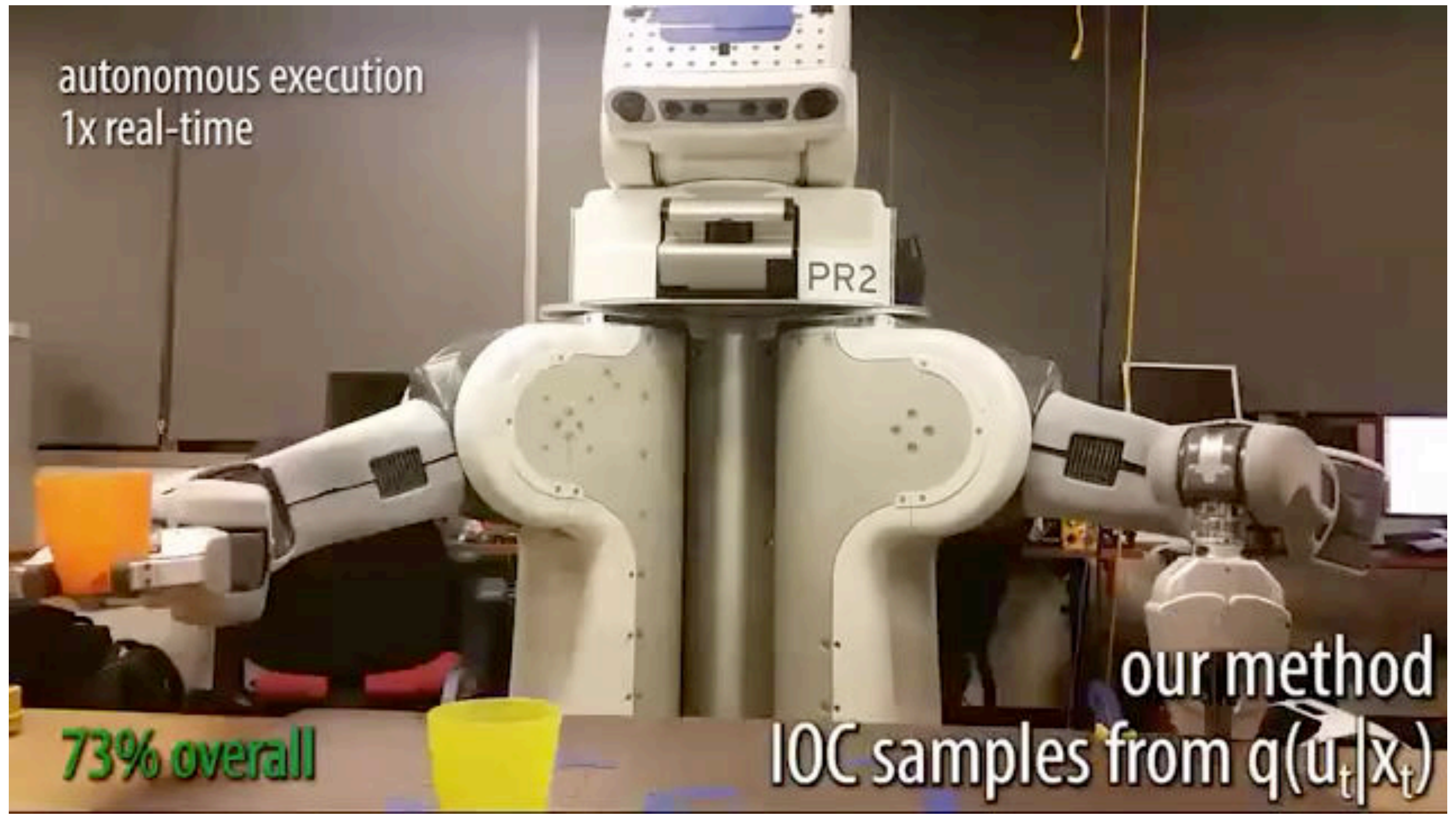
behavioral cloning warm start



[Rajeswaran et al. RSS 2018]

Ground-truth states

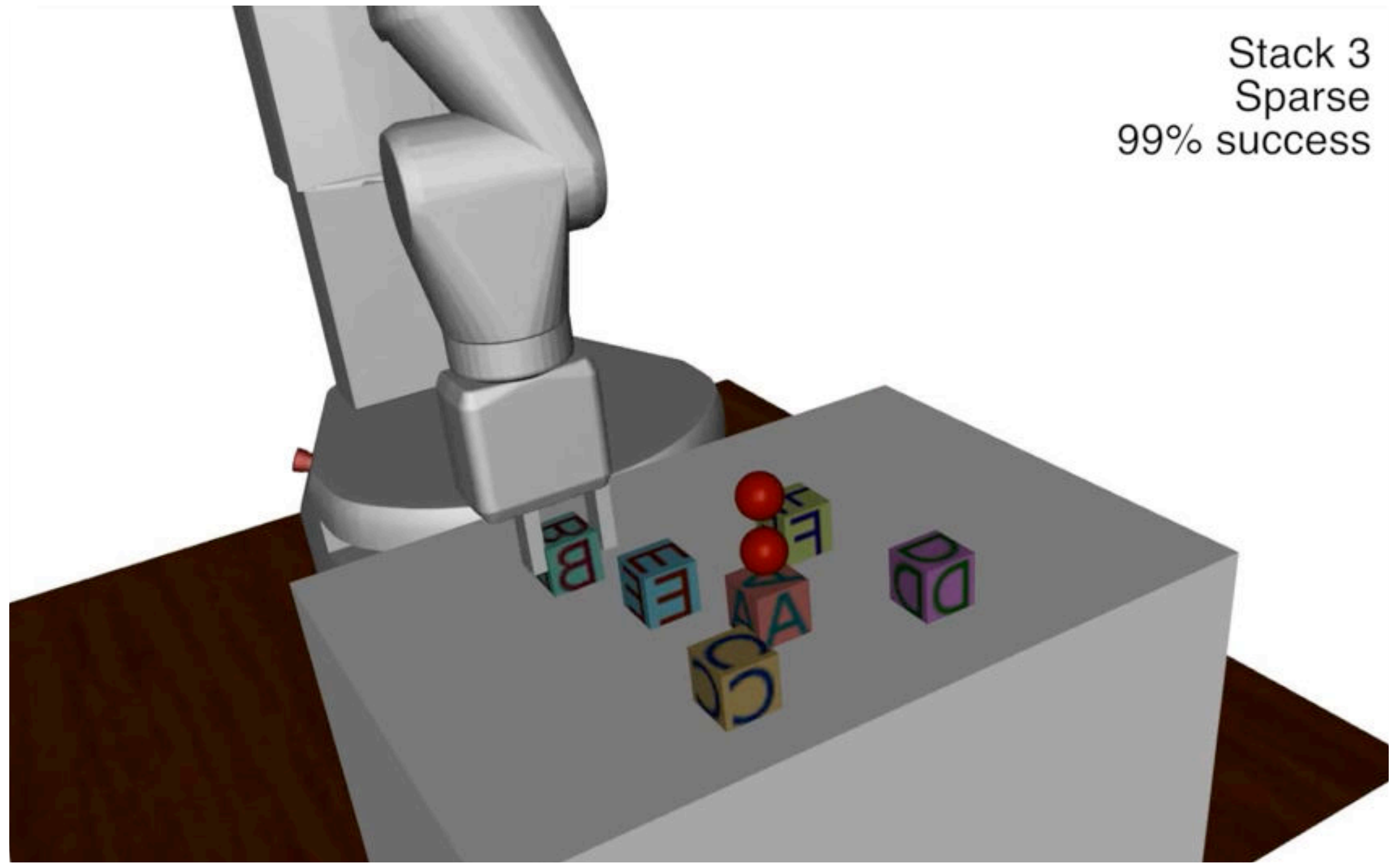
inverse reinforcement learning



[Finn et al. ICML 2016]

Short-horizon tasks

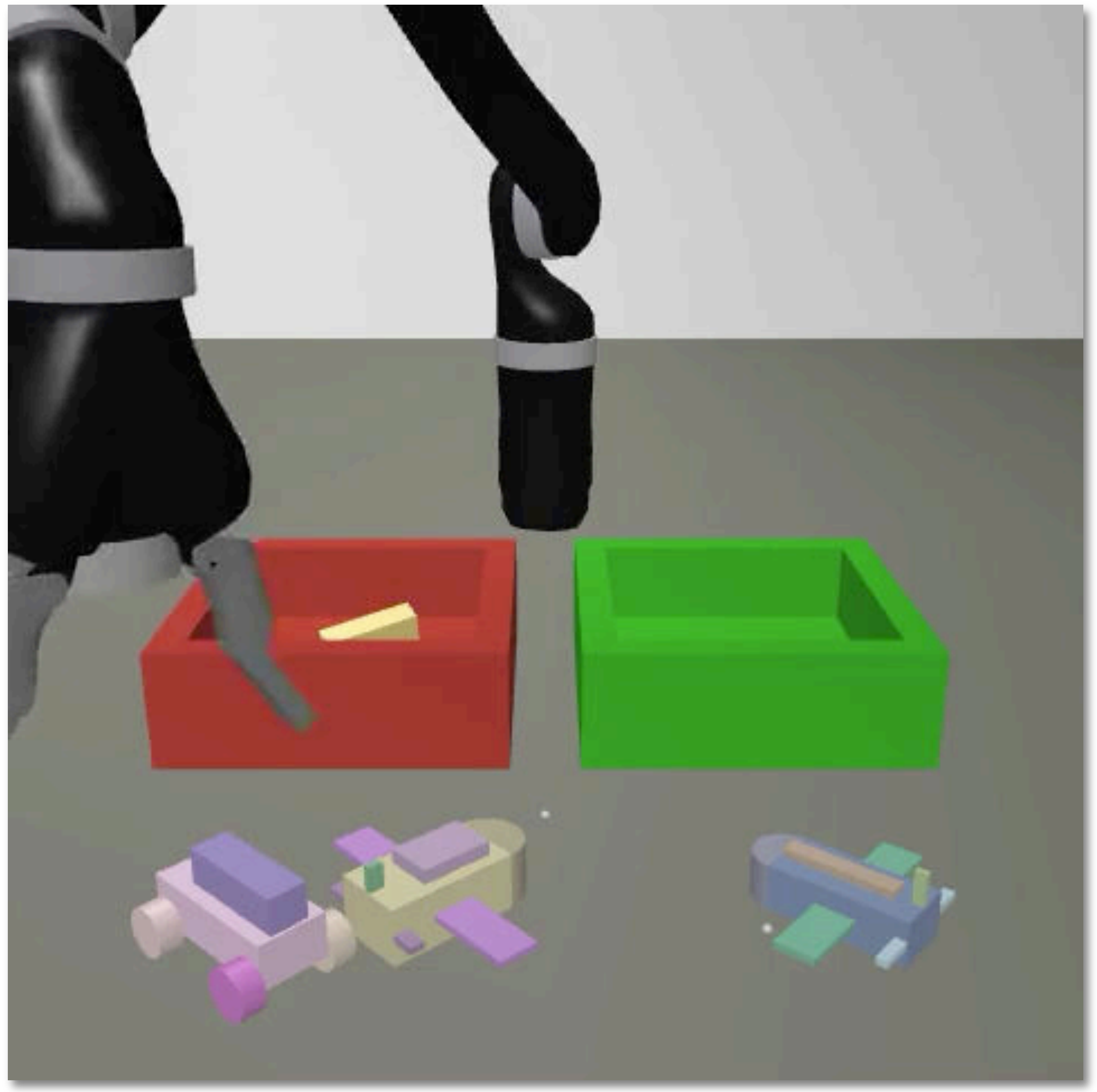
data augmentation



[Nair et al. ICRA 2018]

Fixed objects

Reinforcement and Imitation Learning



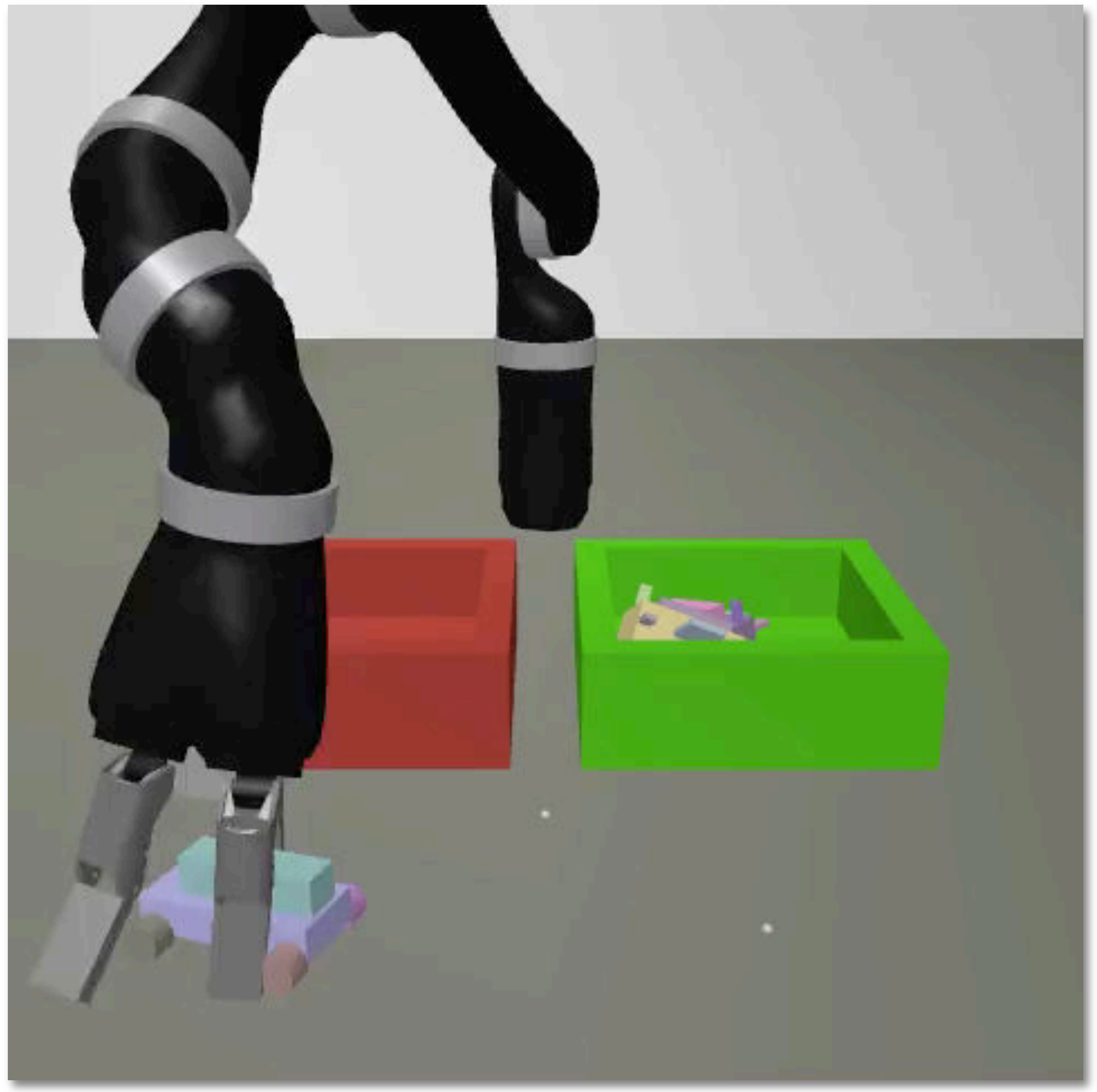
Task: Car → Red & Plane → Green

Our trained model

Ours	Prior work
Raw pixel inputs (RGB camera)	Ground-truth states
Long-horizon tasks (each task takes ~1min)	Short-horizon tasks
Object variation (procedural generation)	Fixed objects

Zhu, Wang, Merel, Rusu, Erez, Cabi, Tunyasuvunakool, Kramár, Hadsell, de Freitas, Heess, *RSS* 2018

Reinforcement and Imitation Learning



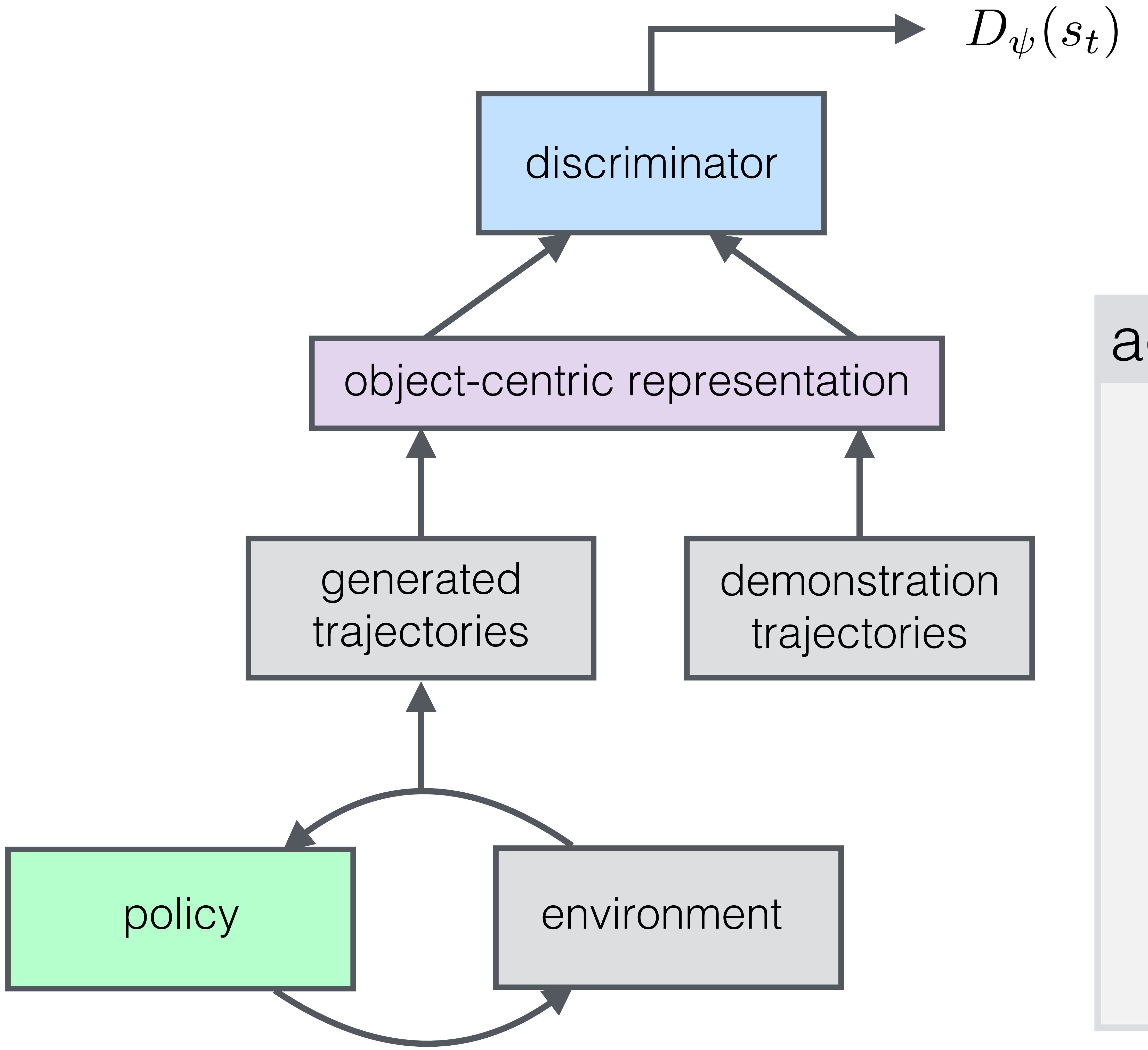
Task: Car → Red & Plane → Green

Our trained model

Effective RL + IL = Algorithm + Data

Zhu, Wang, Merel, Rusu, Erez, Cabi, Tunyasuvunakool, Kramár, Hadsell, de Freitas, Heess, *RSS* 2018

Reinforcement and Imitation Learning: Algorithm – Adversarial Learning[1]



<u>discriminator objective</u>

D_ψ predicts **0** if policy and **1** if demo

adversarial learning objective

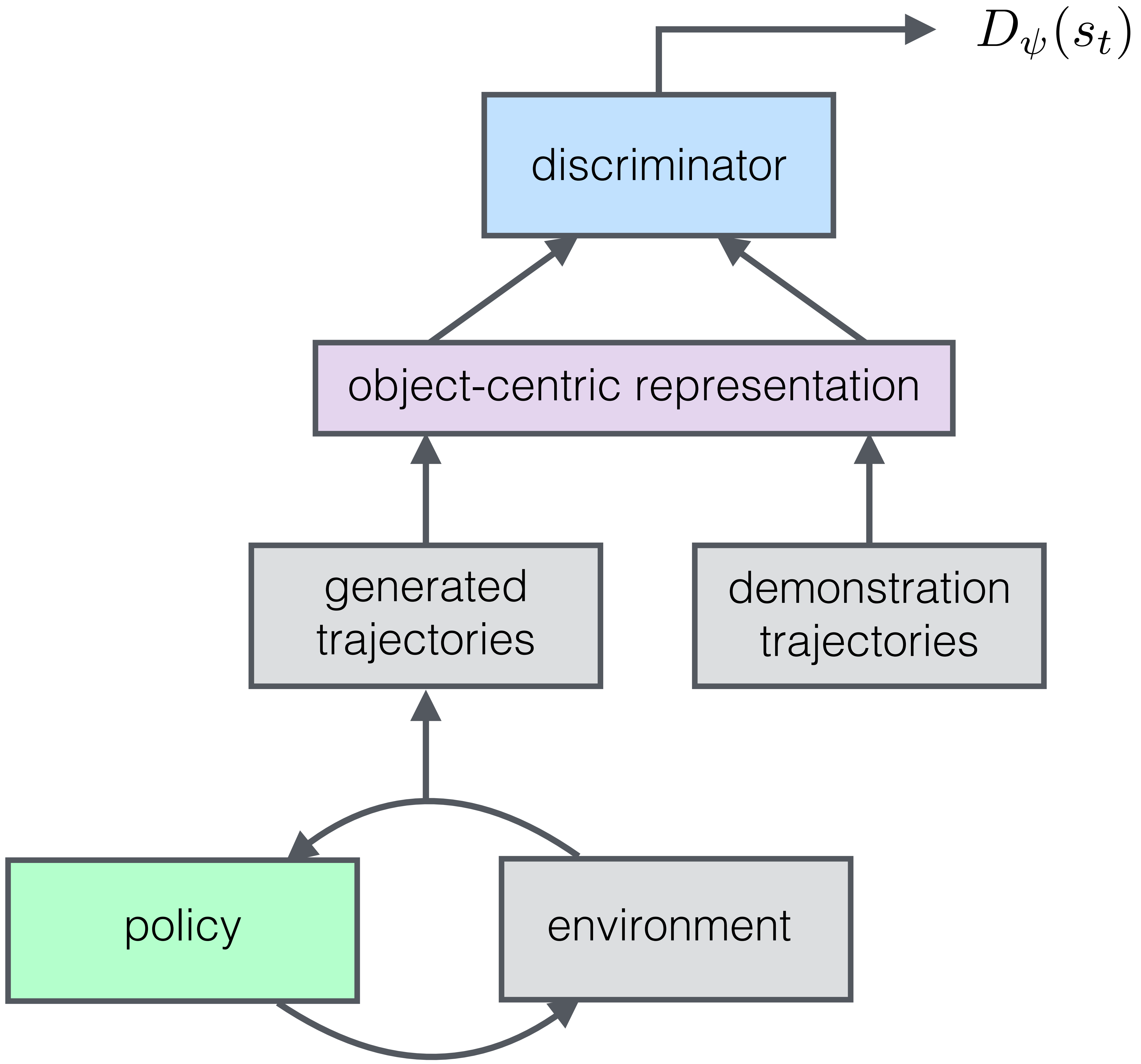
object-centric representation

$$\overbrace{\min_{\theta} \underbrace{\max_{\psi} \mathbb{E}_{\pi_E}[\log D_\psi(s)] + \mathbb{E}_{\pi_\theta}[\log(1 - D_\psi(s))]}_{\text{discriminator } D_\psi\text{: to tell policy apart from demonstration}}}^{\text{policy } \pi_\theta\text{: to fool the discriminator}}$$

[1]Goodfellow et al. 2014; Ho & Ermon, 2016

Zhu, Wang, Merel, Rusu, Erez, Cabi, Tunyasuvunakool, Kramár, Hadsell, de Freitas, Heess, *RSS* 2018

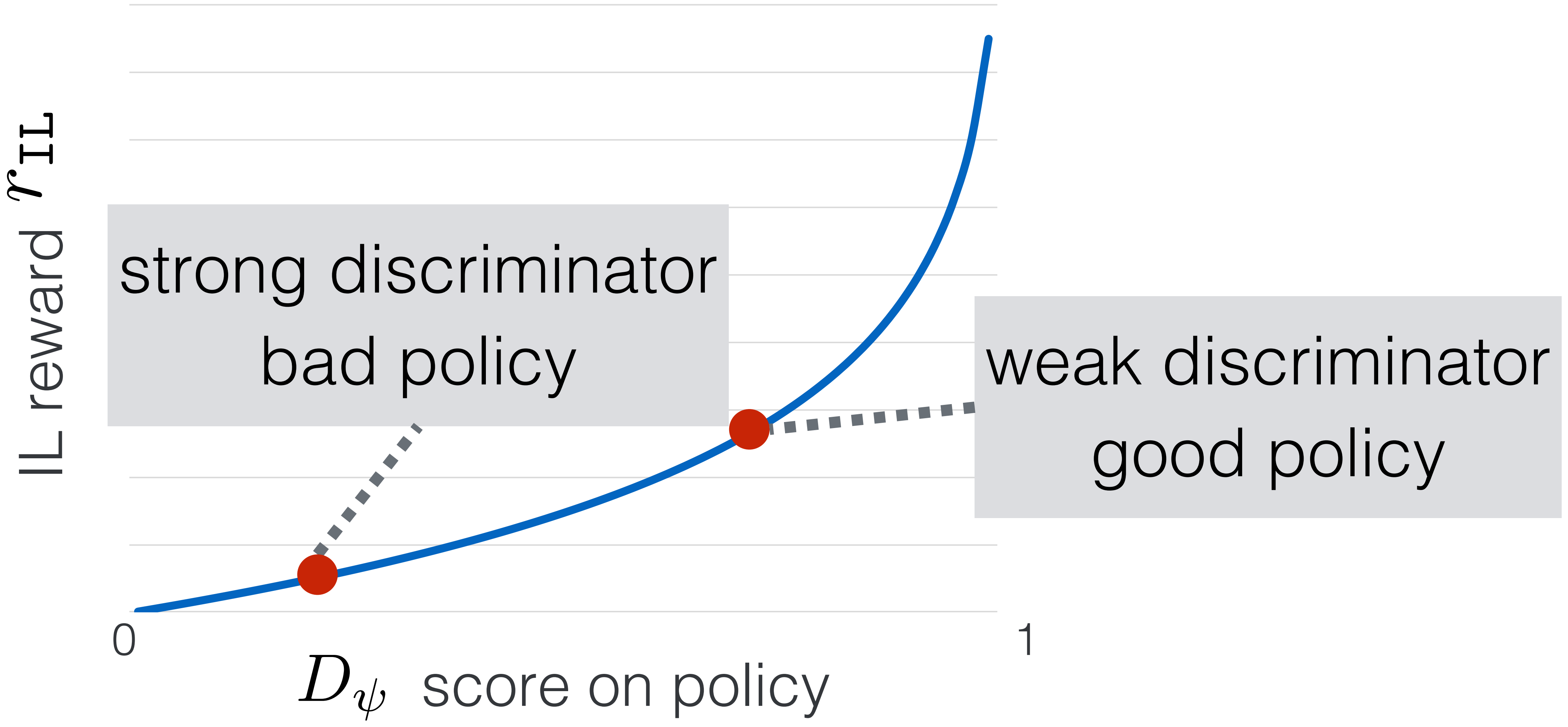
Reinforcement and Imitation Learning: Algorithm – Adversarial Learning[1]



discriminator objective

D_ψ predicts **0** if policy and **1** if demo

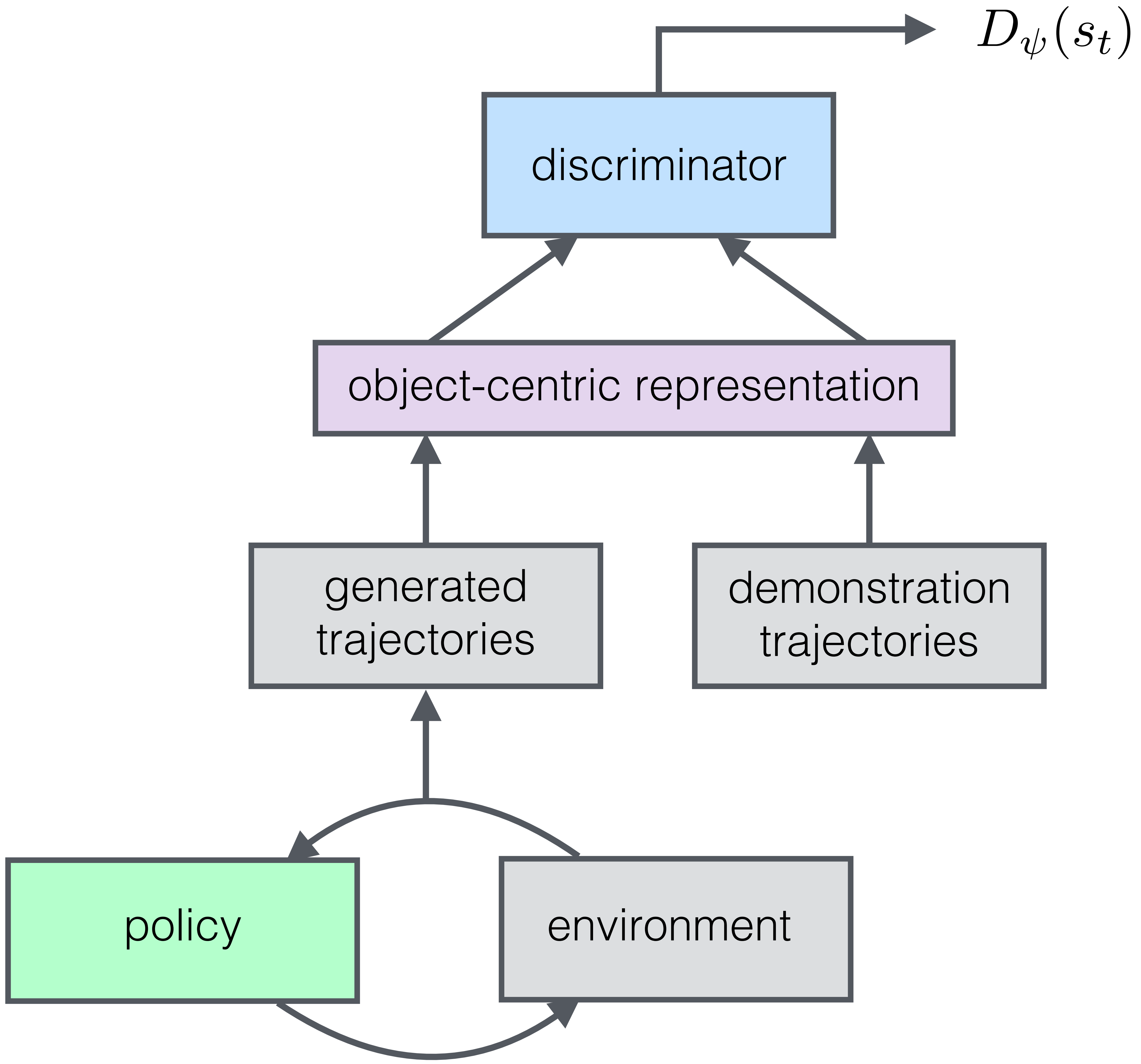
IL reward: $r_{\text{IL}}(s_t, a_t) = -\log(1 - D_\psi(s_t))$



[1]Goodfellow et al. 2014; Ho & Ermon, 2016

Zhu, Wang, Merel, Rusu, Erez, Cabi, Tunyasuvunakool, Kramár, Hadsell, de Freitas, Heess, *RSS* 2018

Reinforcement and Imitation Learning: Algorithm – Adversarial Learning[1]



<u>discriminator objective</u>

D_ψ predicts **0** if policy and **1** if demo

IL reward: $r_{\texttt{IL}}(s_t, a_t) = -\log(1 - D_\psi(s_t))$

RL reward: $r_{\texttt{RL}}(s_t, a_t)$

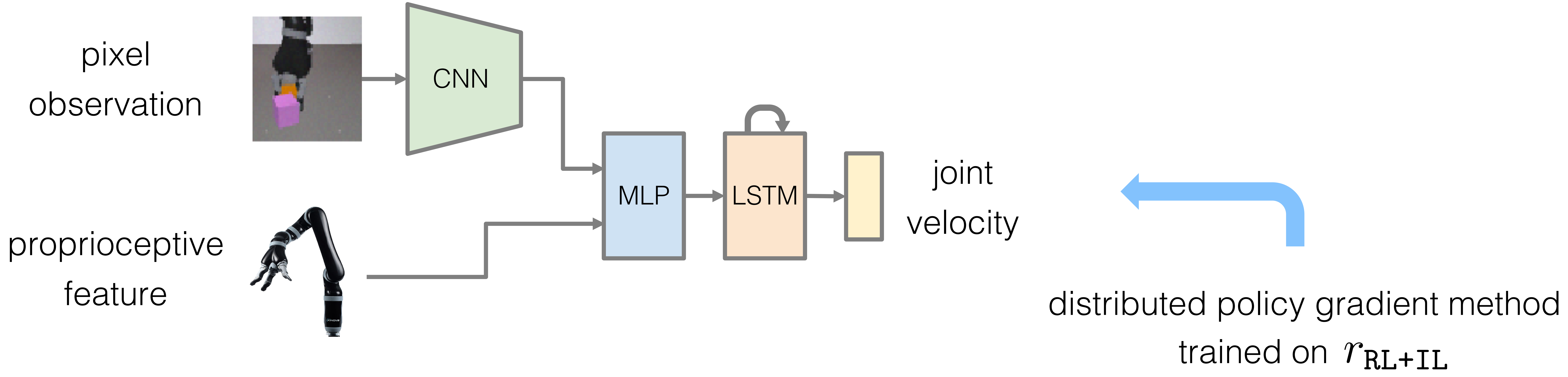
RL + IL reward:

$$r_{\texttt{RL+IL}} = \underbrace{\lambda r_{\texttt{IL}}(s_t, a_t)}_{\text{imitation}} + \underbrace{(1-\lambda) r_{\texttt{RL}}(s_t, a_t)}_{\text{reinforcement}}$$

[1]Goodfellow et al. 2014; Ho & Ermon, 2016

Zhu, Wang, Merel, Rusu, Erez, Cabi, Tunyasuvunakool, Kramár, Hadsell, de Freitas, Heess, *RSS* 2018

Reinforcement and Imitation Learning: Algorithm



Input 64 x 64 RGB pixel observation

positions and velocities of arm joints and grippers

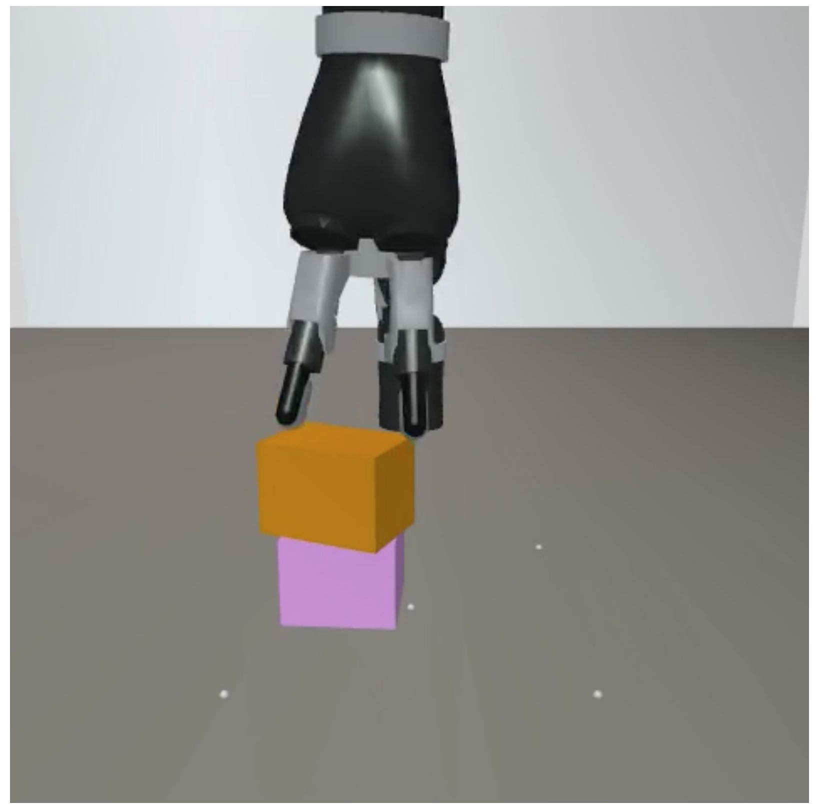
Output 9-DoF joint velocities at 20Hz

Zhu, Wang, Merel, Rusu, Erez, Cabi, Tunyasuvunakool, Kramár, Hadsell, de Freitas, Heess, *RSS* 2018

3D motion controller

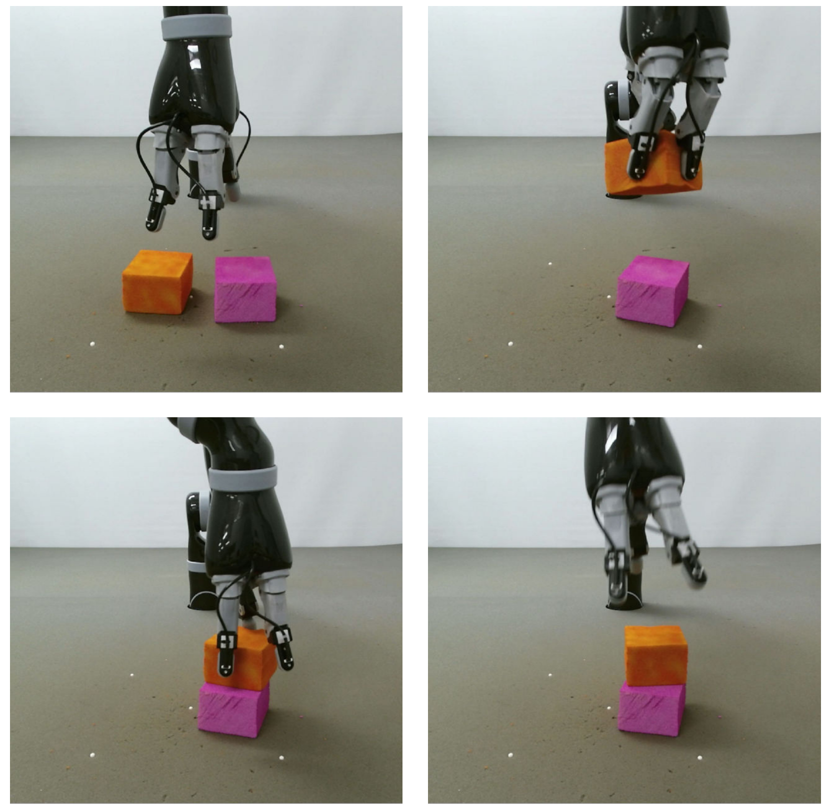
imitation

physics engine



sim2real

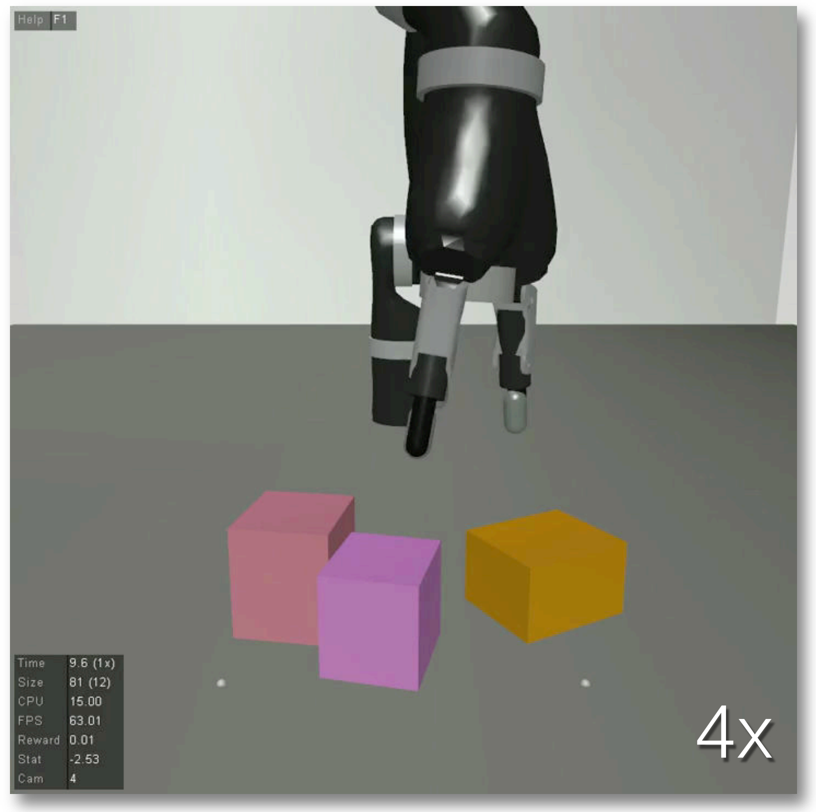
real environment

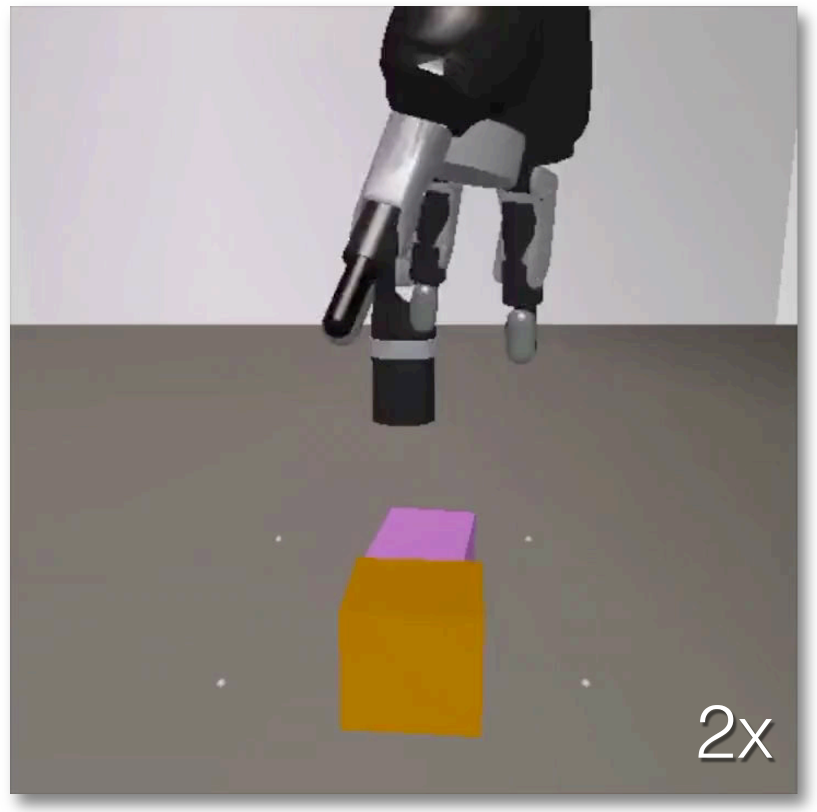


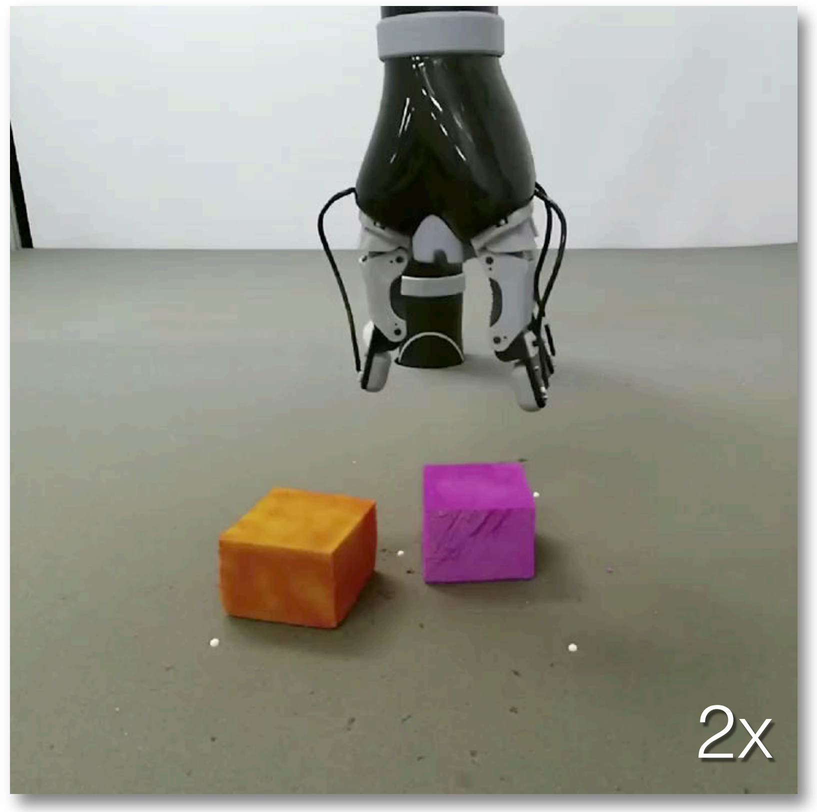
Collecting human demonstrations

Training in simulation

Running on real robot

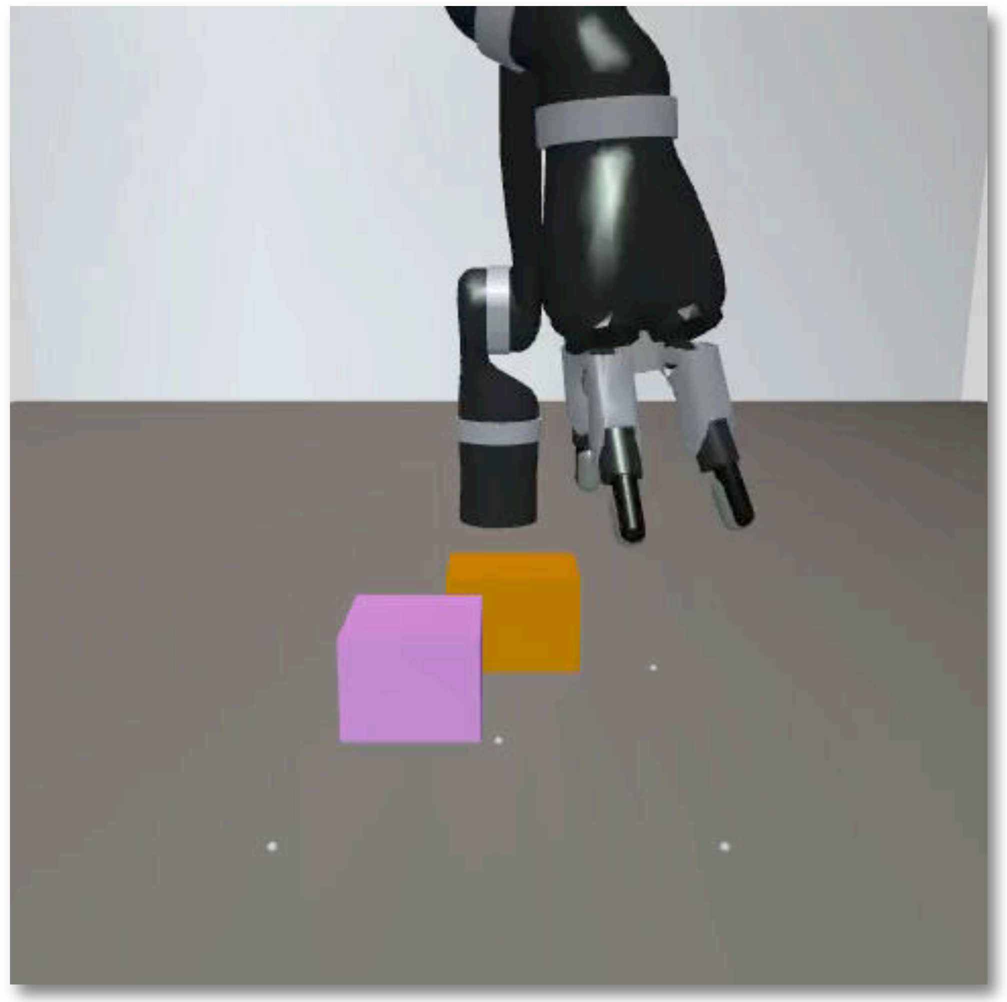




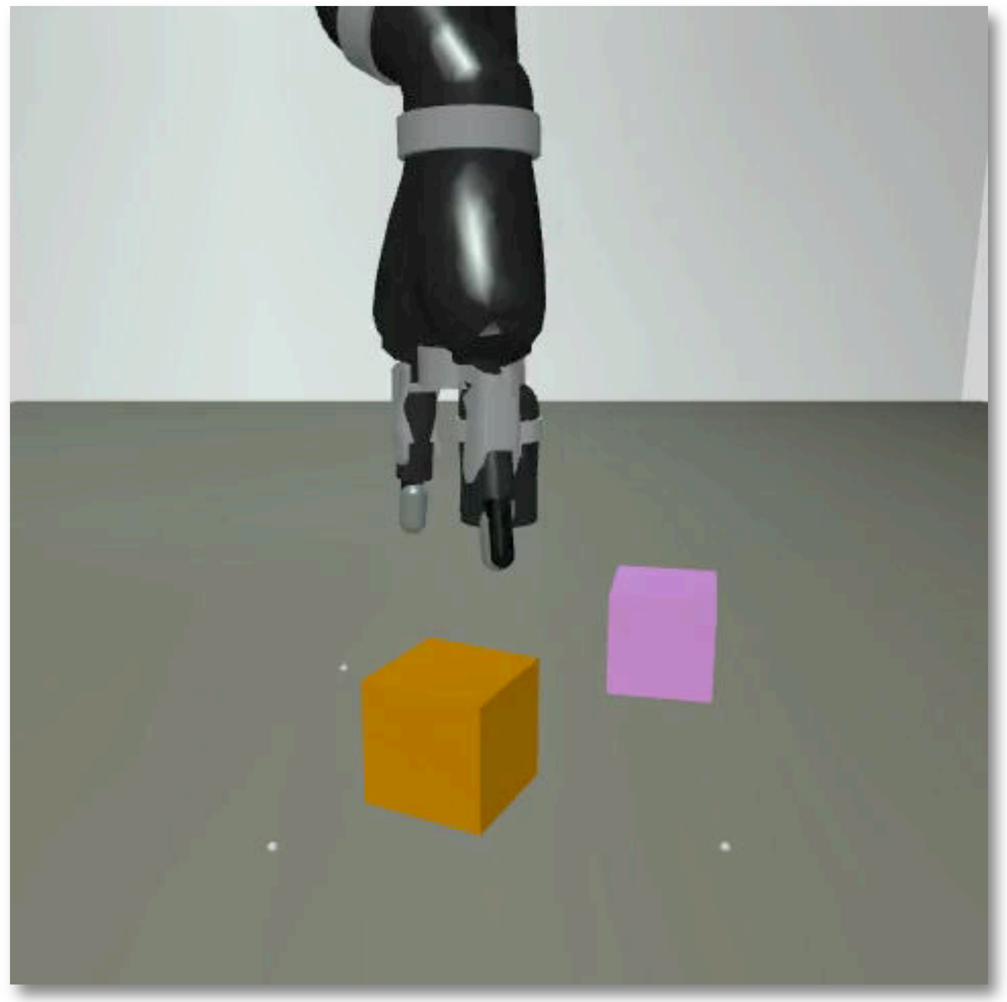


Zhu, Wang, Merel, Rusu, Erez, Cabi, Tunyasuvunakool, Kramár, Hadsell, de Freitas, Heess, *RSS* 2018

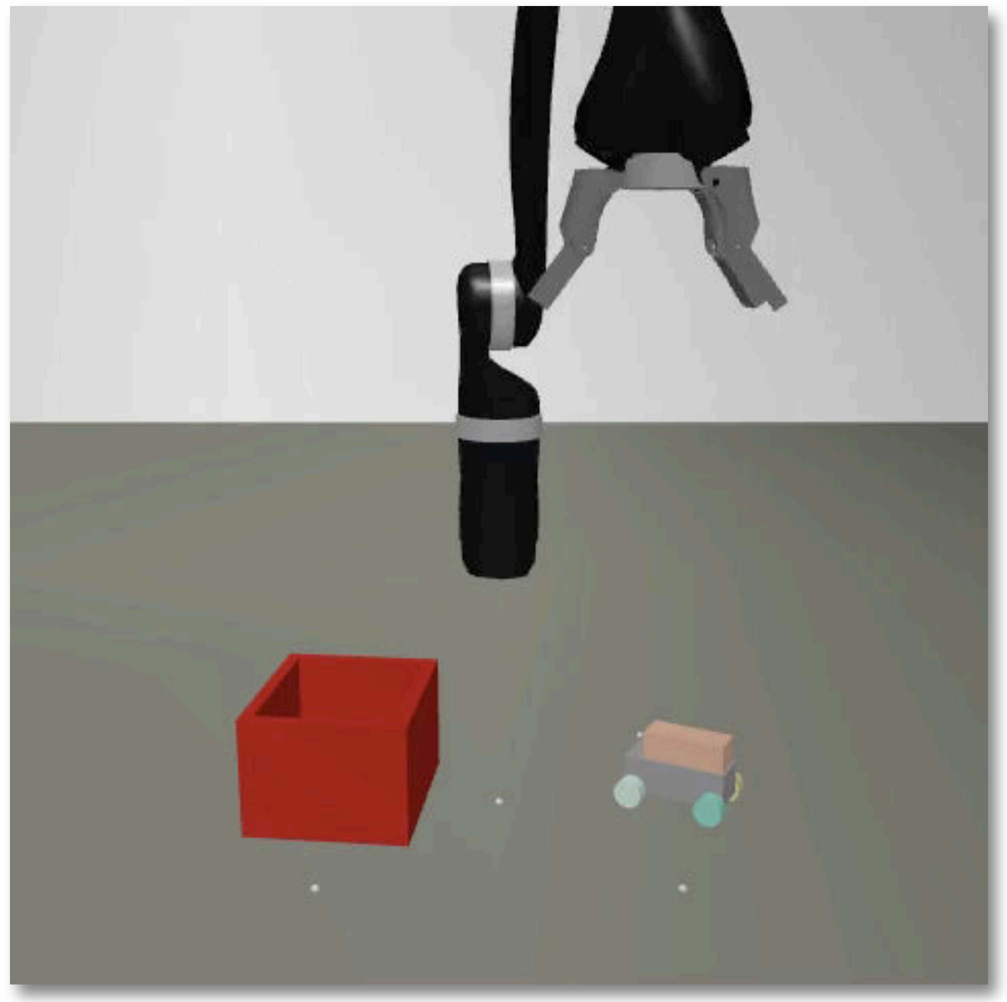
block lifting



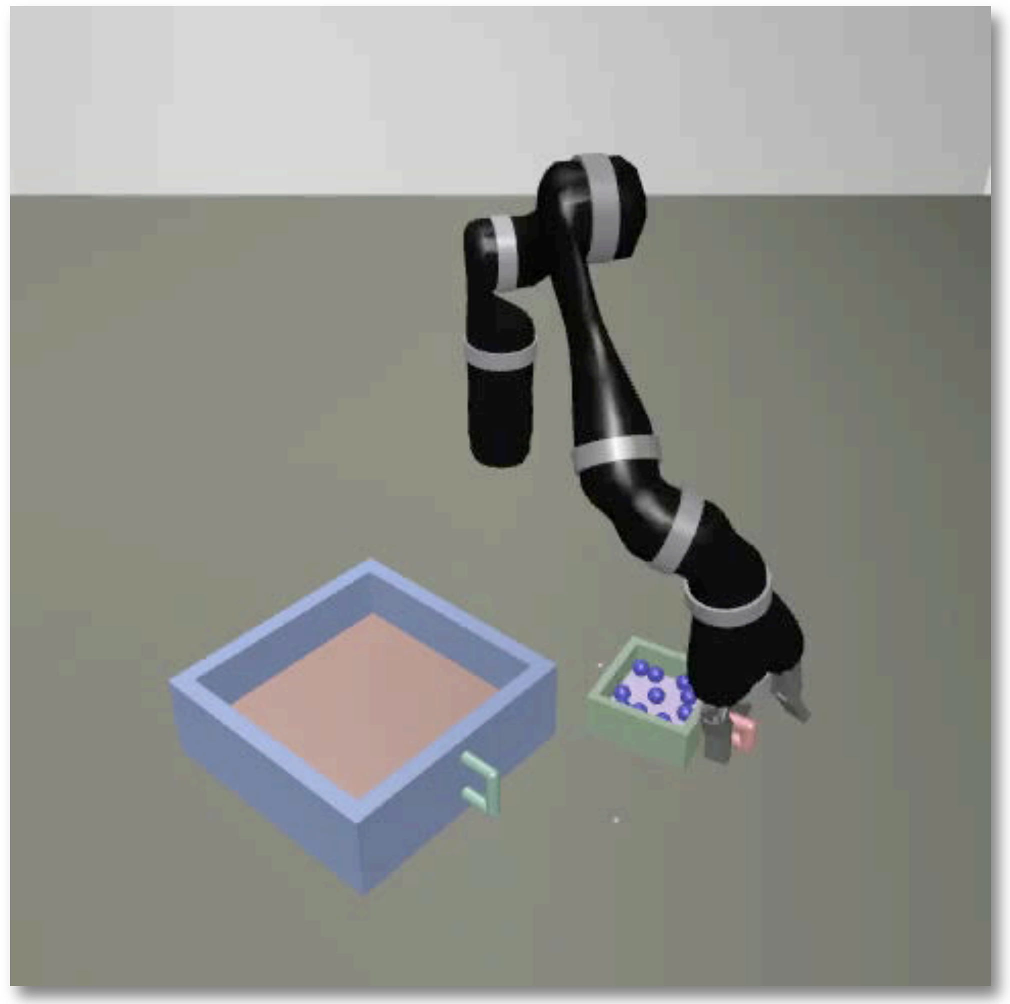
block stacking



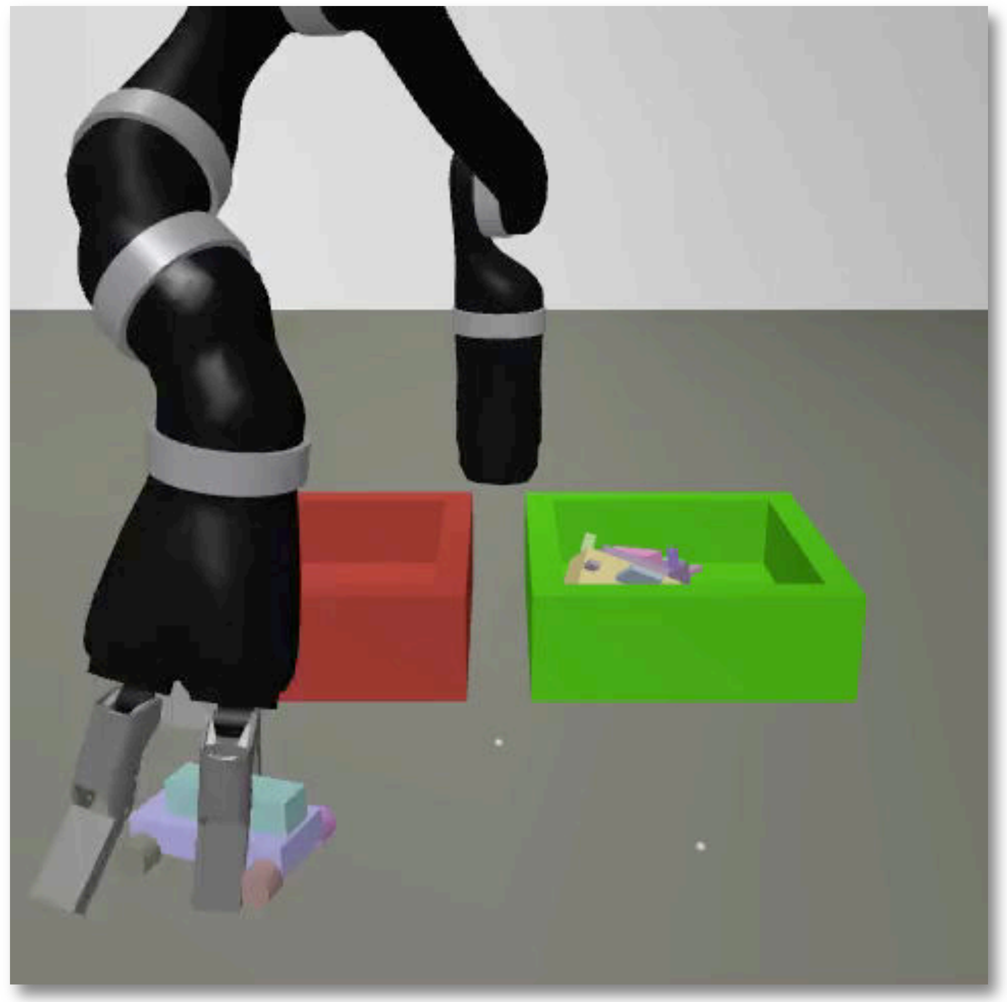
clearing table
with blocks



clearing table
with a box



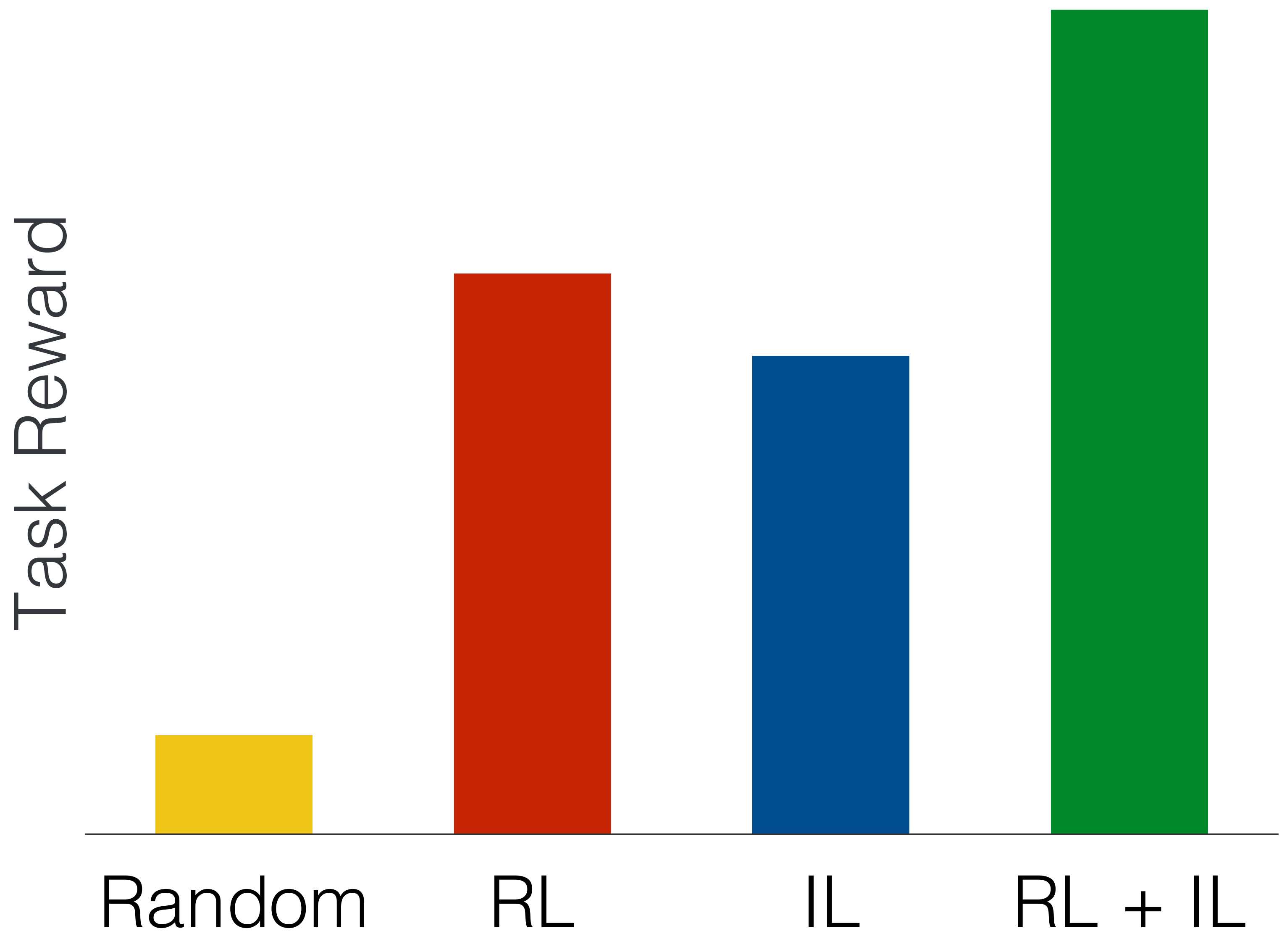
pouring liquid

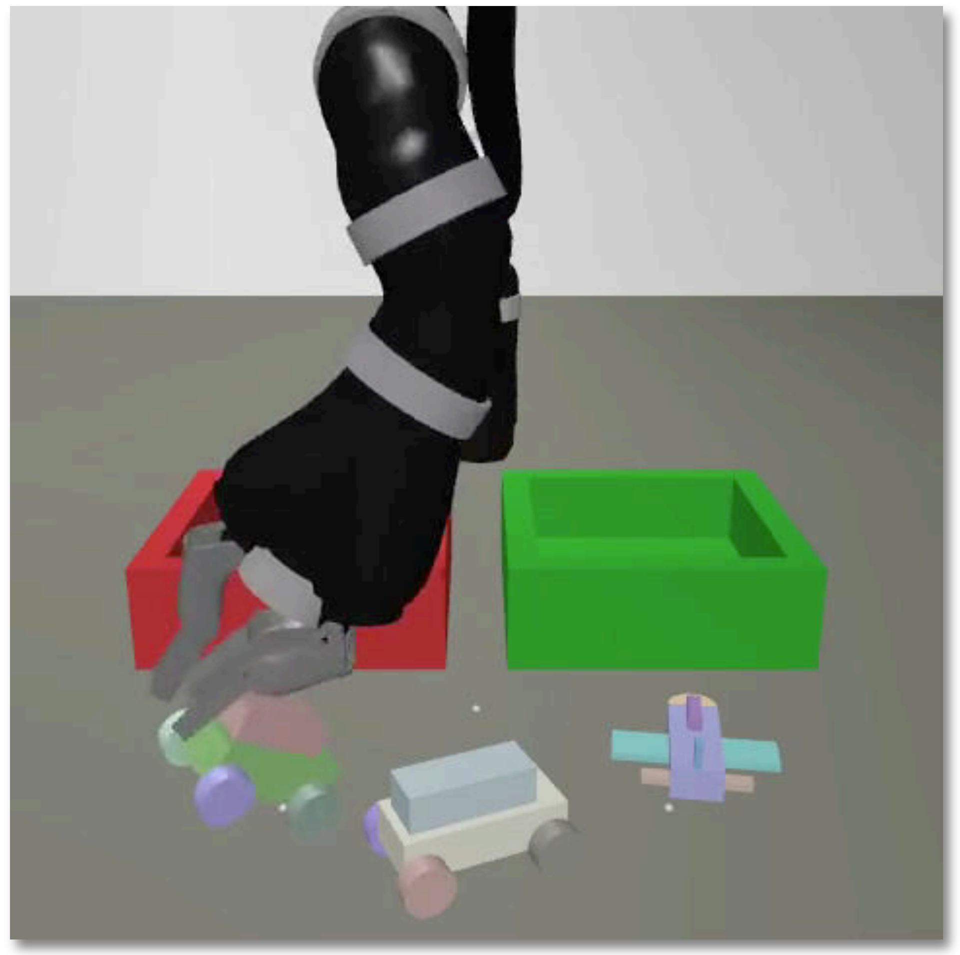


order fulfillment

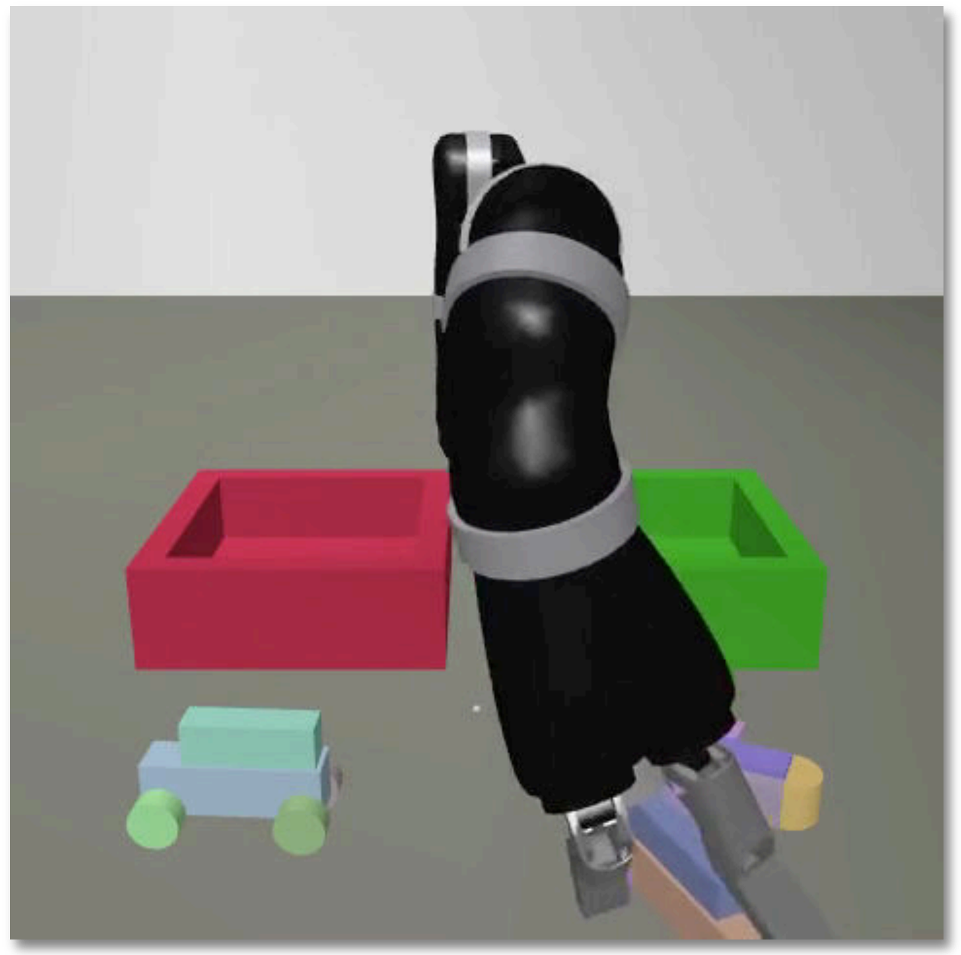
Zhu, Wang, Merel, Rusu, Erez, Cabi, Tunyasuvunakool, Kramár, Hadsell, de Freitas, Heess, *RSS* 2018

Reinforcement and Imitation Learning: Algorithm

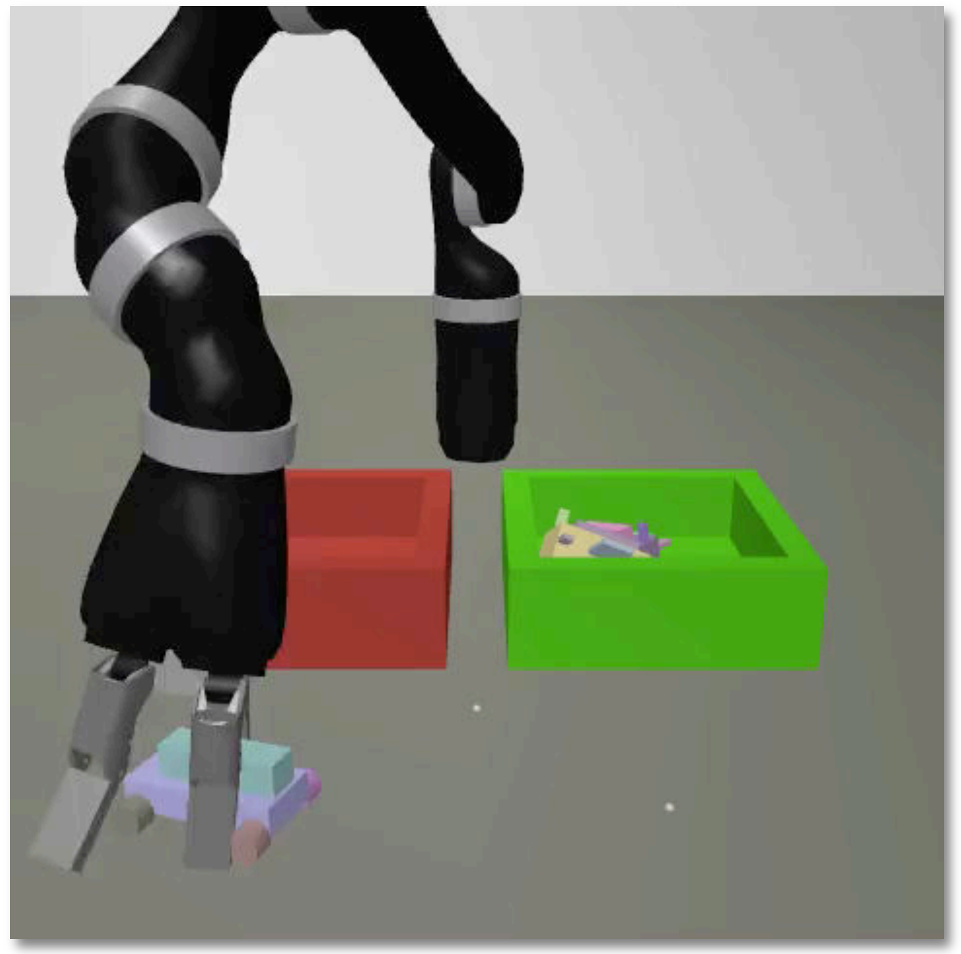




Reinforcement
(RL)



Imitation
(IL)

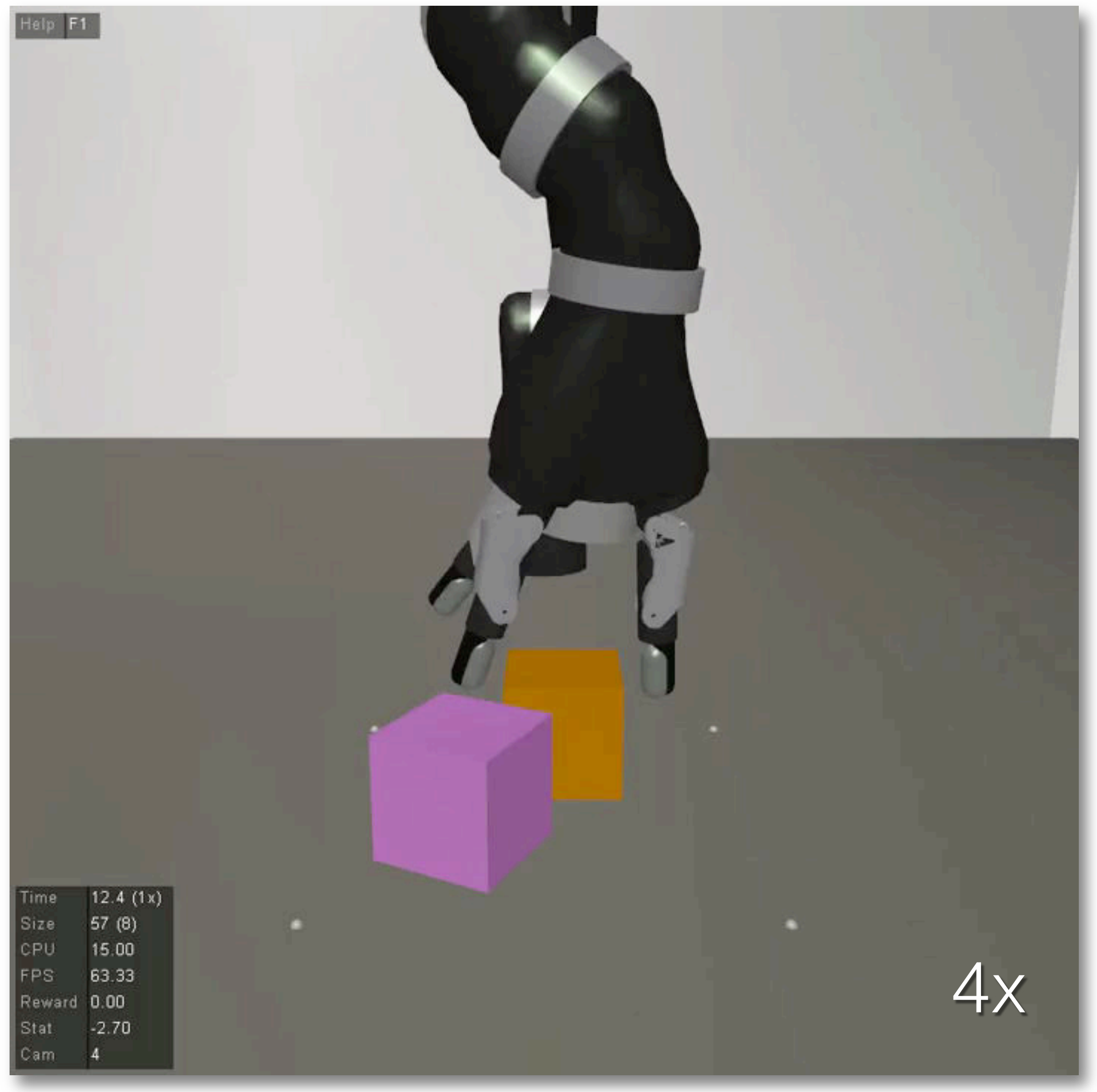


Ours
(RL + IL)

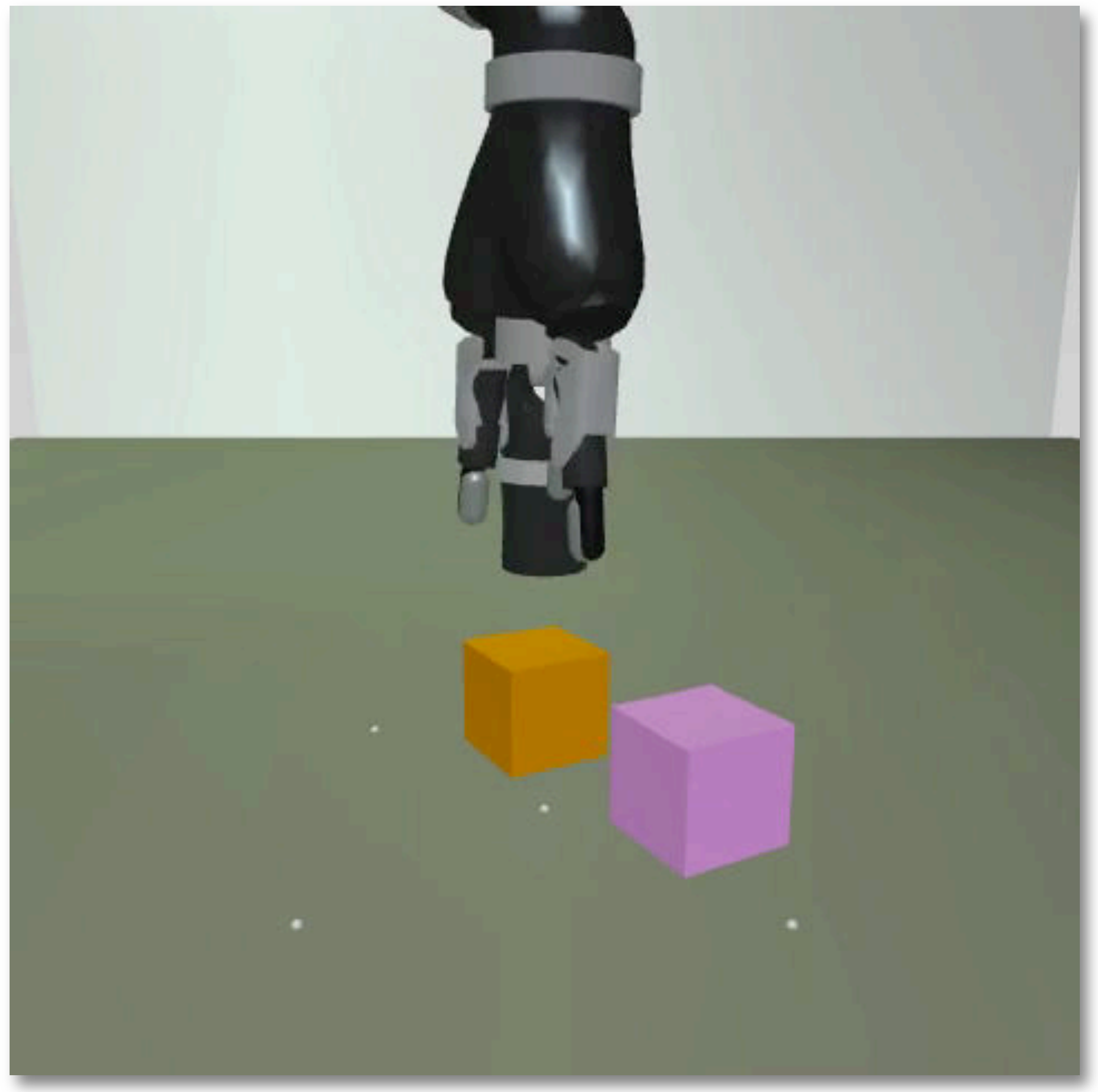
Zhu, Wang, Merel, Rusu, Erez, Cabi, Tunyasuvunakool, Kramár, Hadsell, de Freitas, Heess, *RSS* 2018

Reinforcement and Imitation Learning: Algorithm

Result: Emergent strategies from trial-and-error

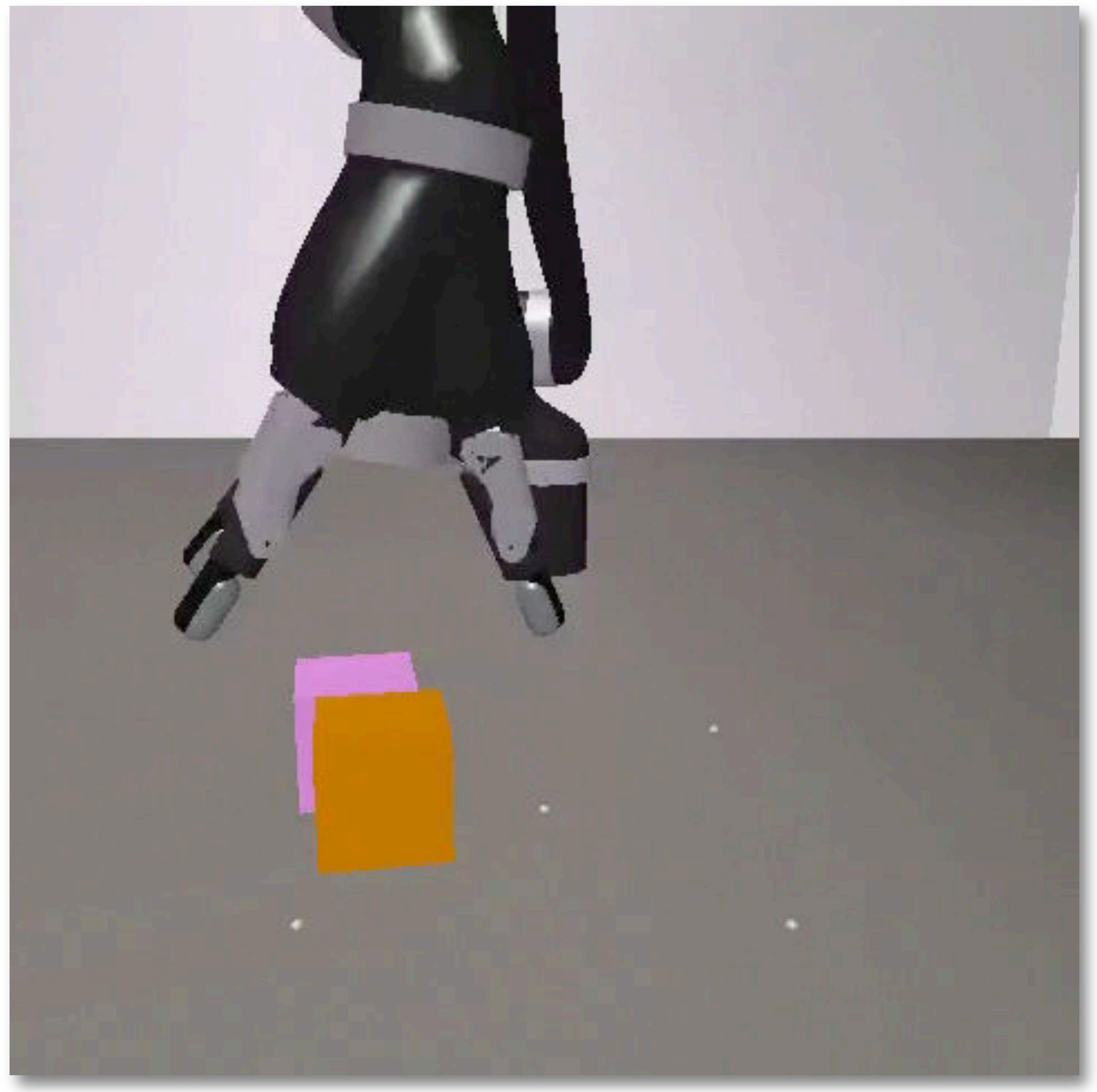


demonstrated solution



novel solution #1

“grasp two blocks from the top”



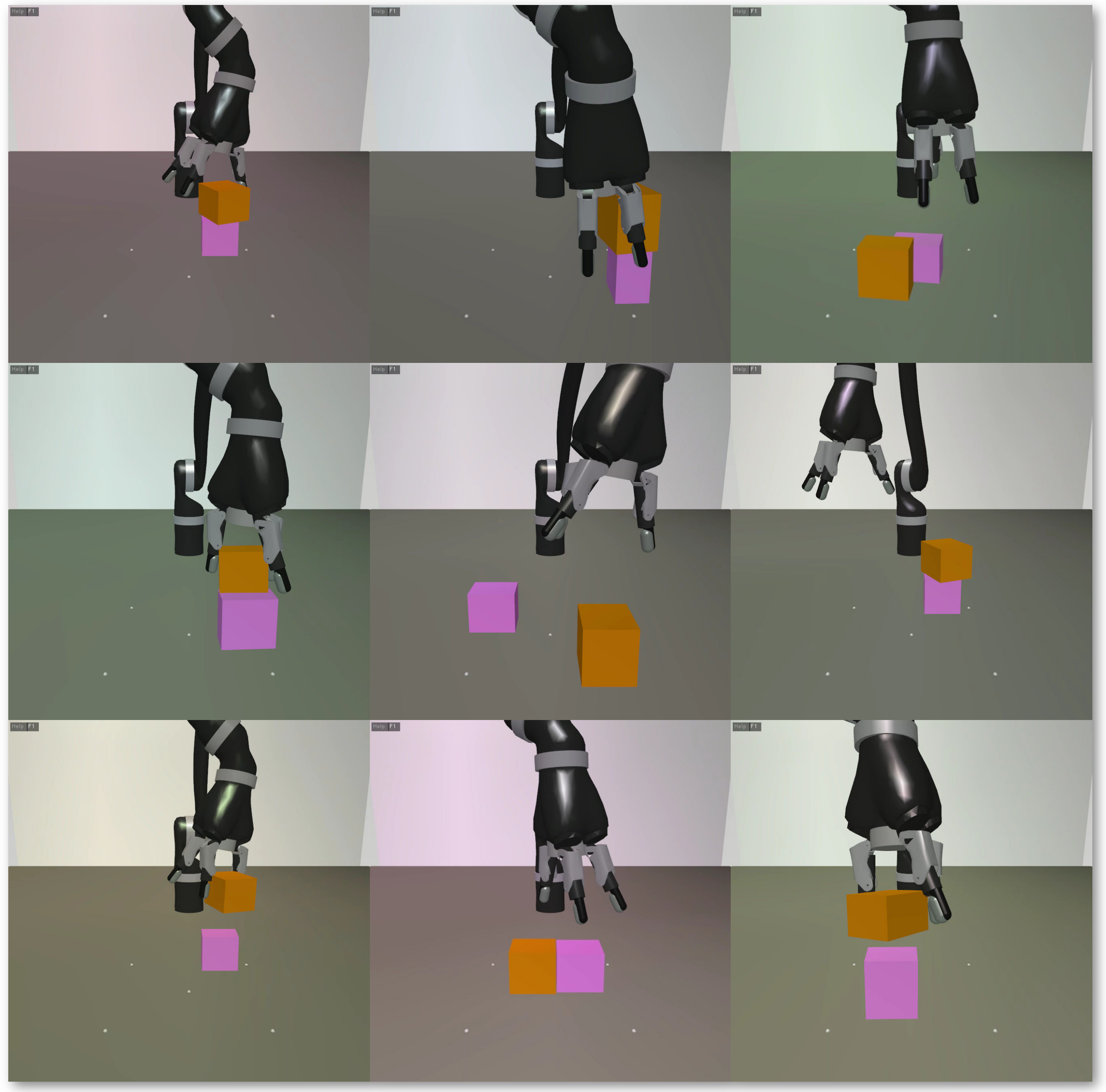
novel solution #2

“lift both blocks together with three fingers”

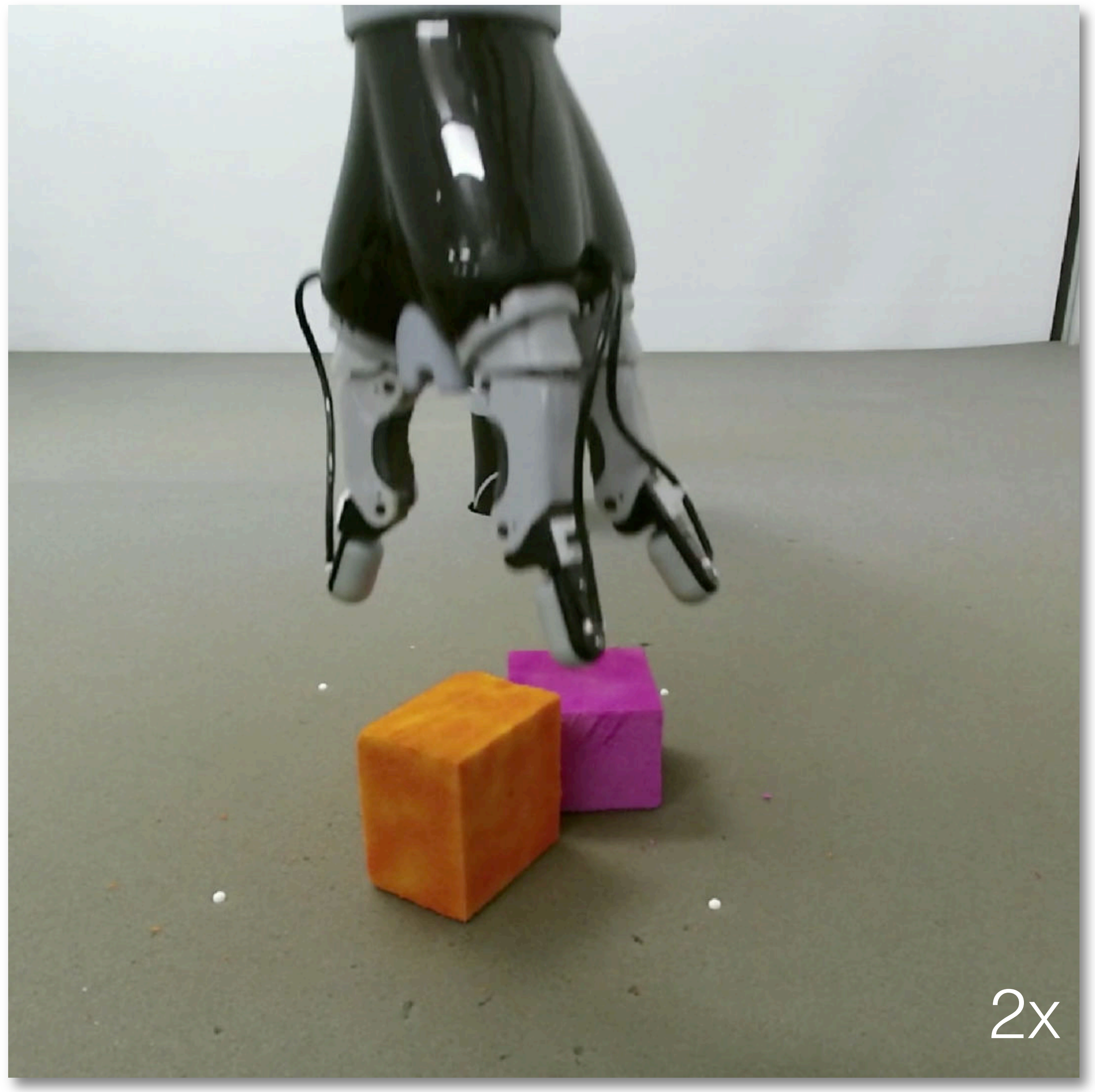
Zhu, Wang, Merel, Rusu, Erez, Cabi, Tunyasuvunakool, Kramár, Hadsell, de Freitas, Heess, *RSS* 2018

Reinforcement and Imitation Learning: Algorithm

Result: Zero-shot sim-to-real policy transfer



visual & dynamics randomization



real-robot deployment

Zhu, Wang, Merel, Rusu, Erez, Cabi, Tunyasuvunakool, Kramár, Hadsell, de Freitas, Heess, *RSS* 2018

Reinforcement and Imitation Learning

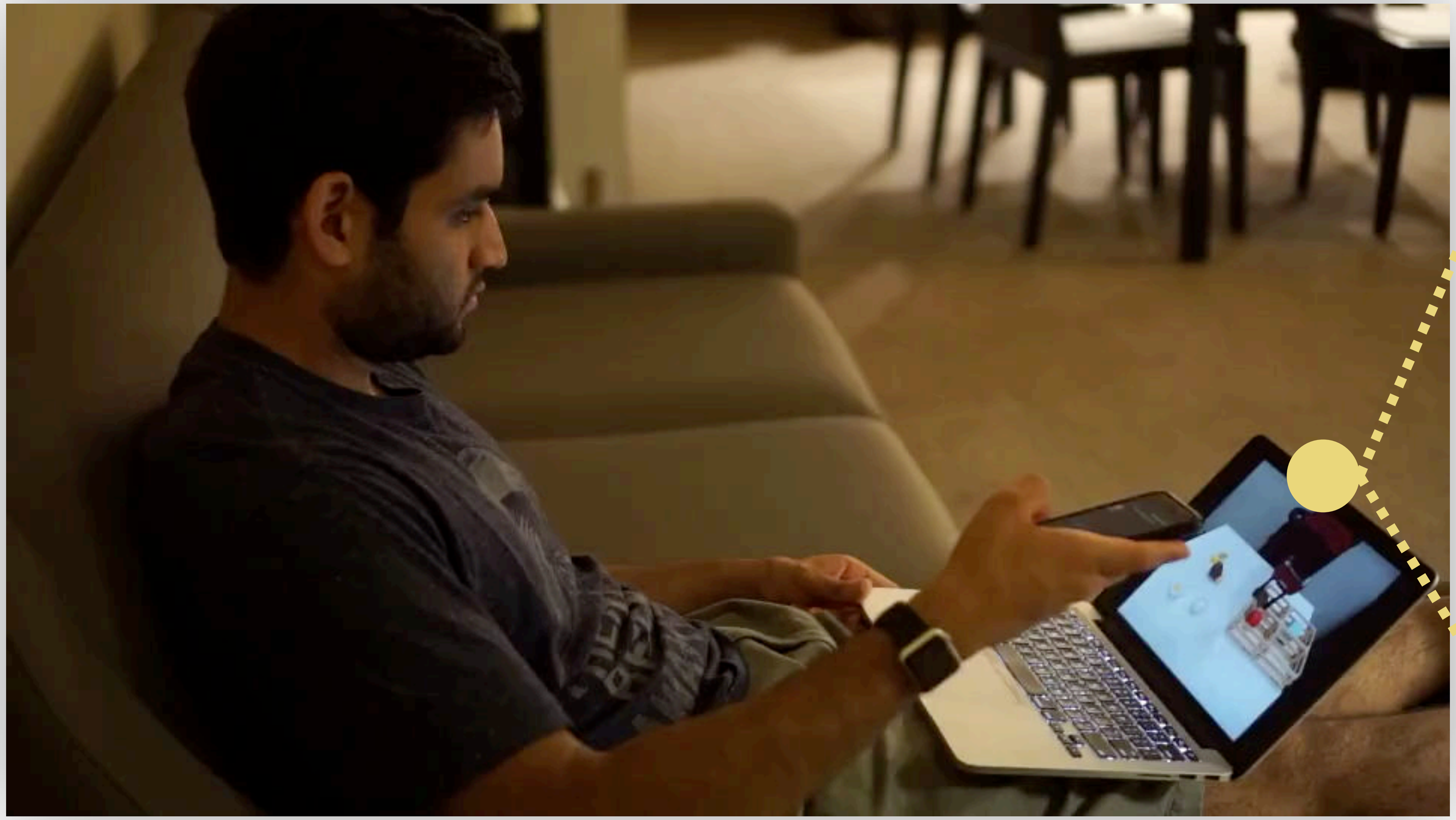
New algorithm that learns dexterous primitive skills

Effective RL + IL = Algorithm + Data

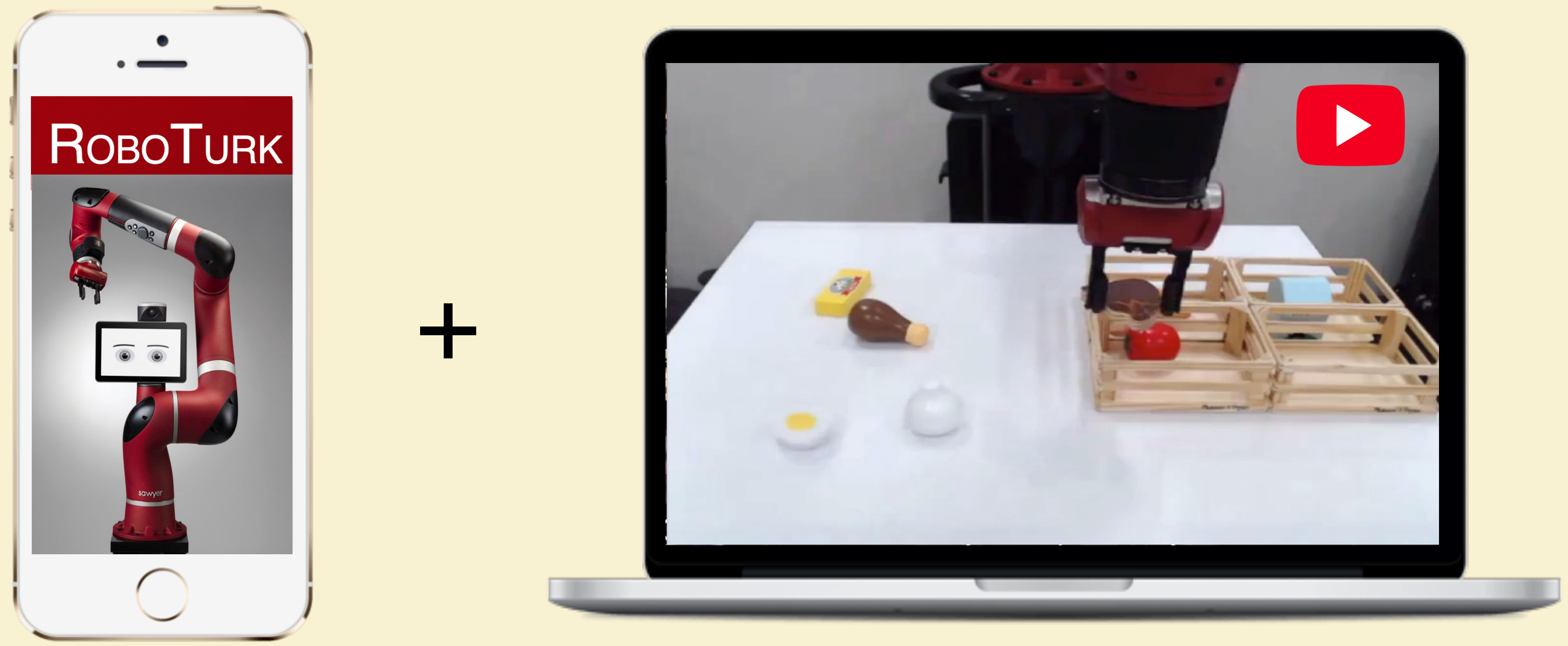
How can we collect demonstrations for diverse skills?

Reinforcement and Imitation Learning: Data

RoboTurk: Crowdsourcing Platform for Large-Scale Demonstration Collection



RoboTurk in action



6-DoF controller

real-time streaming from remote robot

roboturk.stanford.edu

Mandlekar, **Zhu**, Garg, Booher, Spero, Tung, Gao, Emmons, Gupta, Orbay, Savarese, Fei-Fei, CoRL 2018

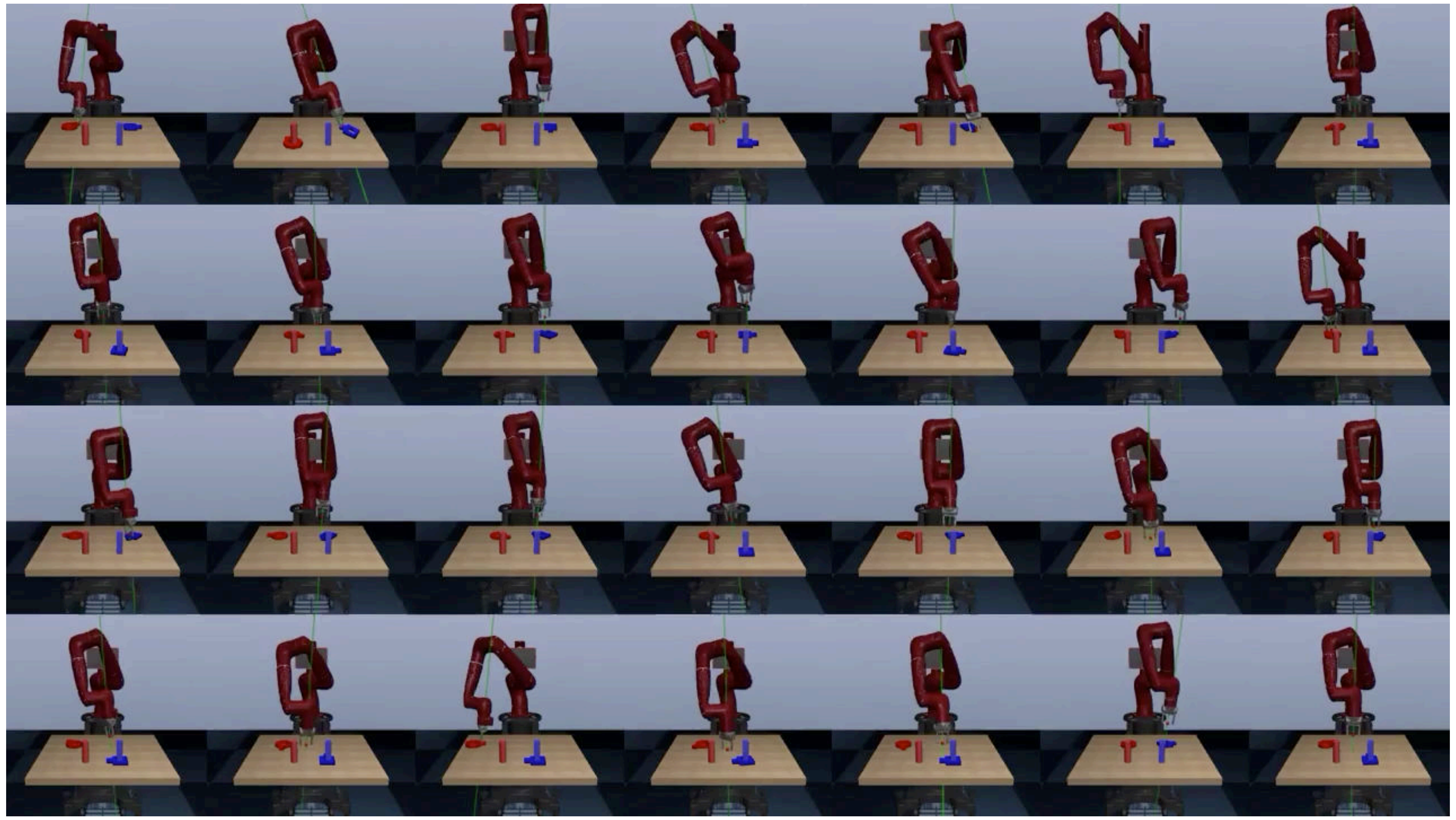
Reinforcement and Imitation Learning: Data

RoboTurk Pilot Dataset

137.5 hours of demonstrations

22 hours of total platform usage

2218 successful demonstrations



teleoperated demonstrations

surreal.stanford.edu

roboturk.stanford.edu

Zhu*, Fan*, Zhu, Liu, Zeng, Gupta, Creus-Costa, Savarese, Fei-Fei, CoRL 2018

Mandlekar, Zhu, Garg, Booher, Spero, Tung, Gao, Emmons, Gupta, Orbay, Savarese, Fei-Fei, CoRL 2018

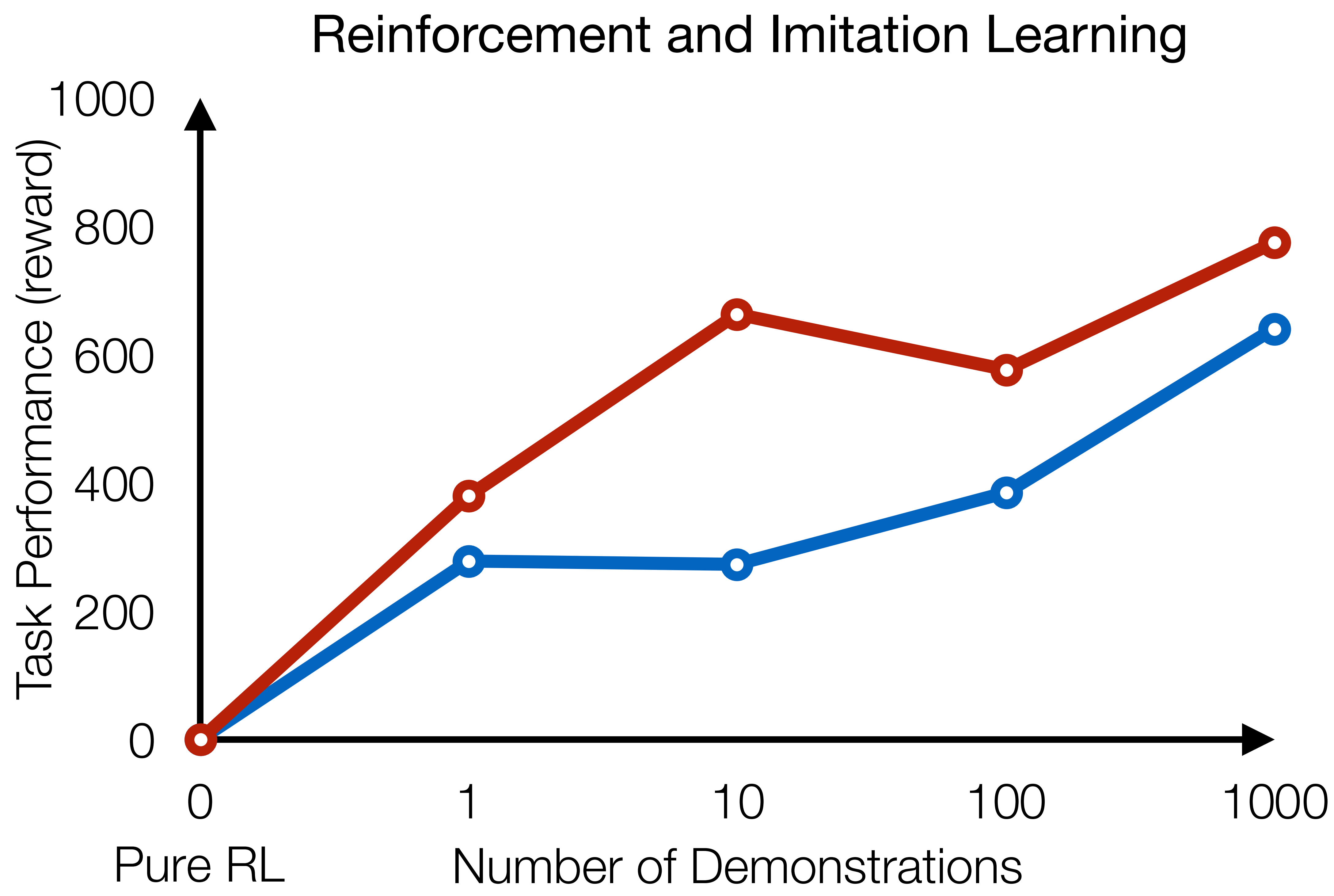
Reinforcement and Imitation Learning: Data

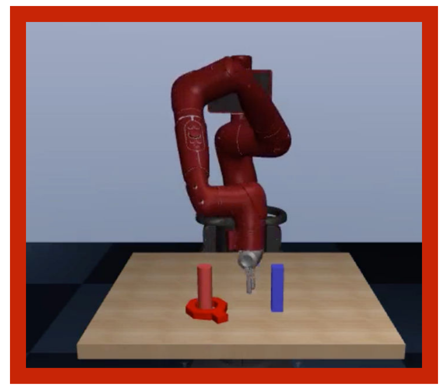
RoboTurk Pilot Dataset

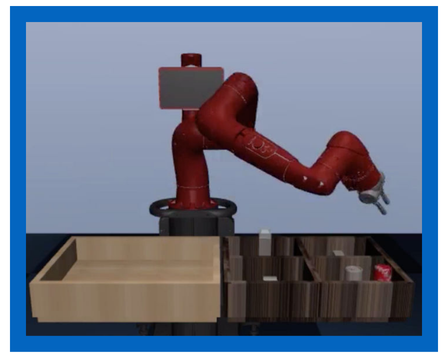
137.5 hours of demonstrations

22 hours of total platform usage

2218 successful demonstrations







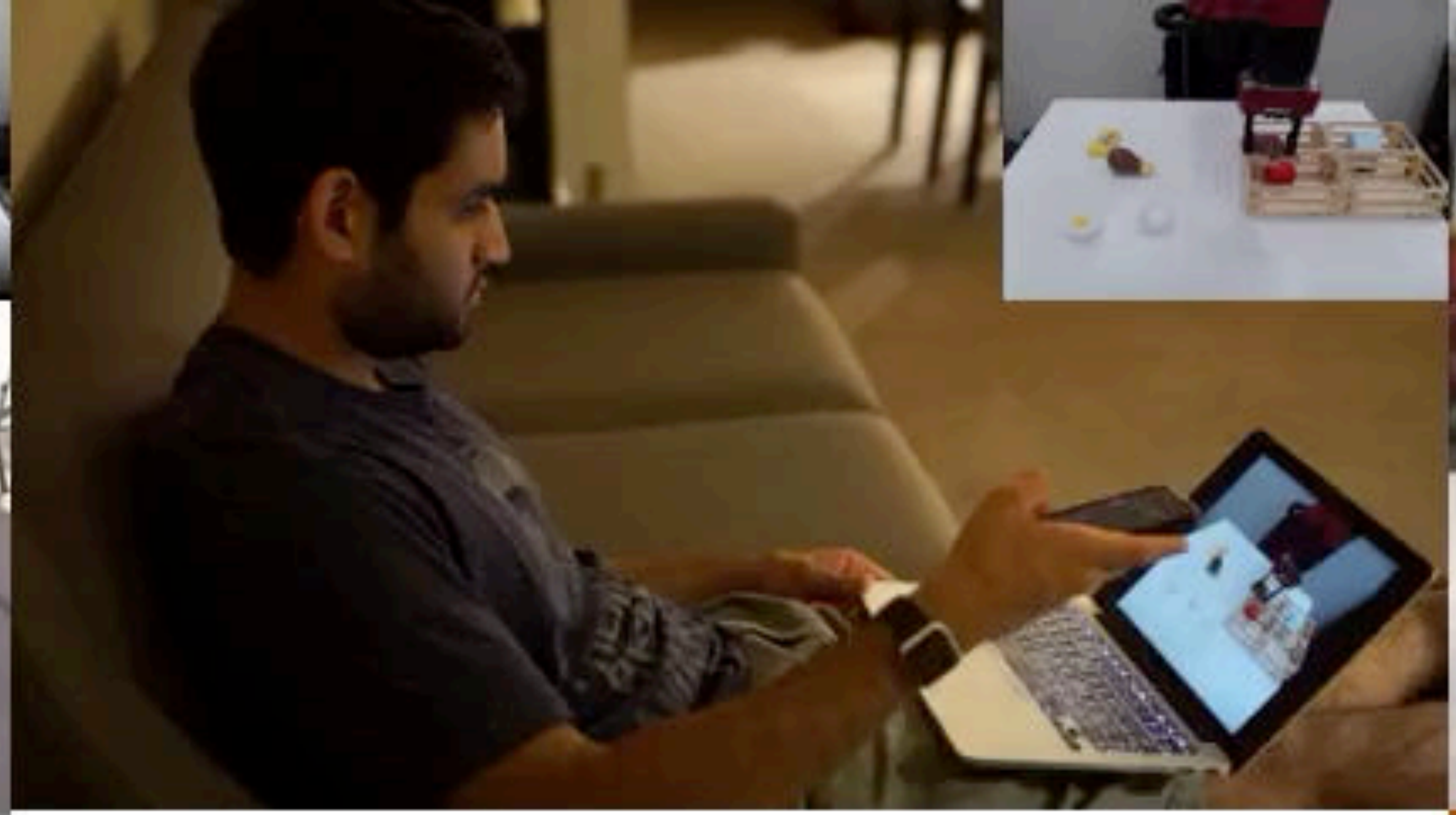
surreal.stanford.edu

roboturk.stanford.edu

Zhu*, Fan*, Zhu, Liu, Zeng, Gupta, Creus-Costa, Savarese, Fei-Fei, CoRL 2018

Mandlekar, Zhu, Garg, Booher, Spero, Tung, Gao, Emmons, Gupta, Orbay, Savarese, Fei-Fei, CoRL 2018

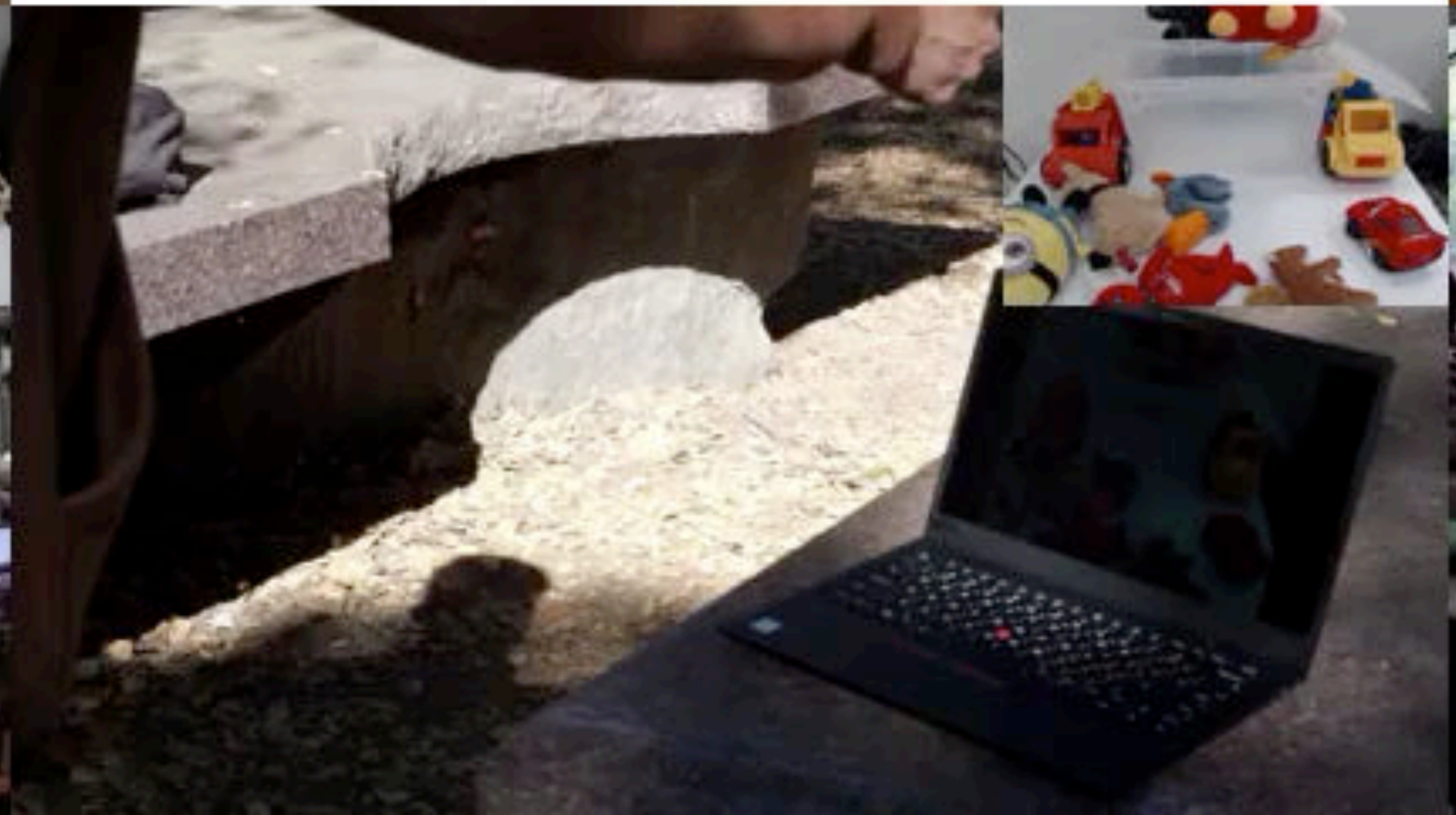
RoboTurk on Physical Robots

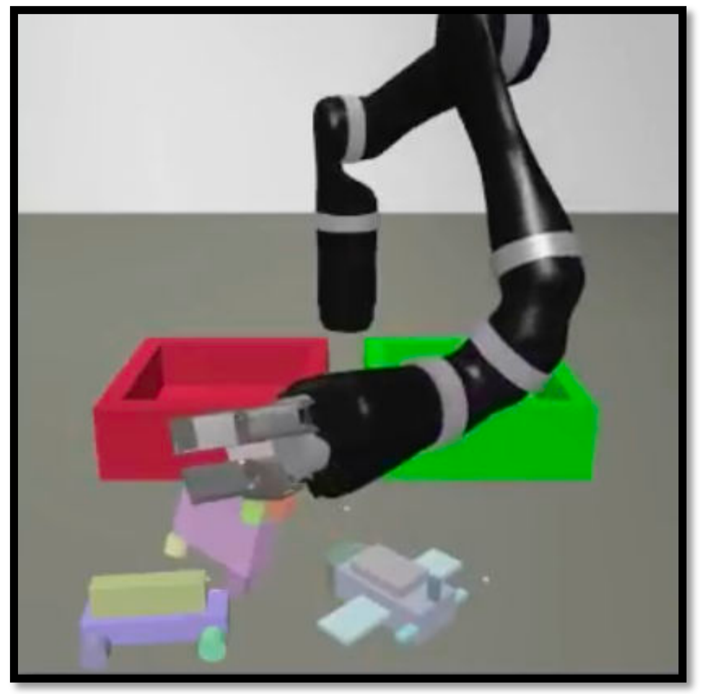
RoboTurk for

everyone, everywhere

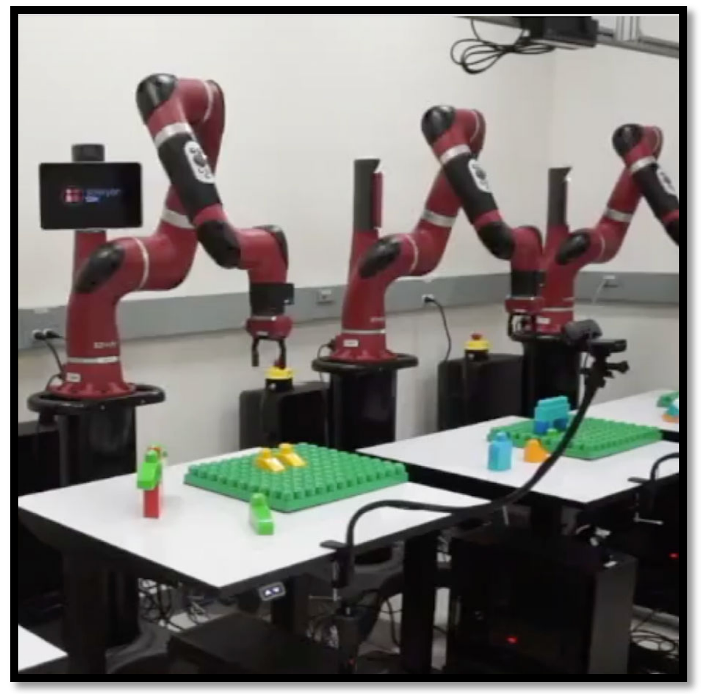





Summary - Part I



Combining **reinforcement and imitation learning** to learn **primitive skills** from **raw sensory inputs**



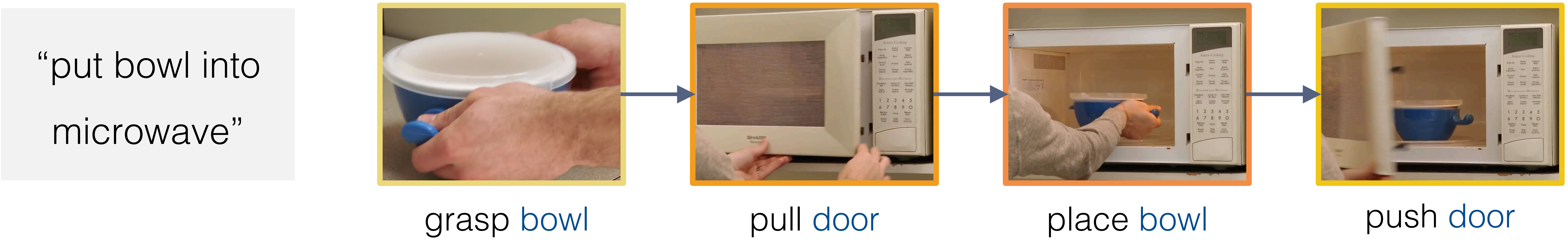
Scaling up demonstration collection with **teleoperated crowdsourcing** using the RoboTurk platform

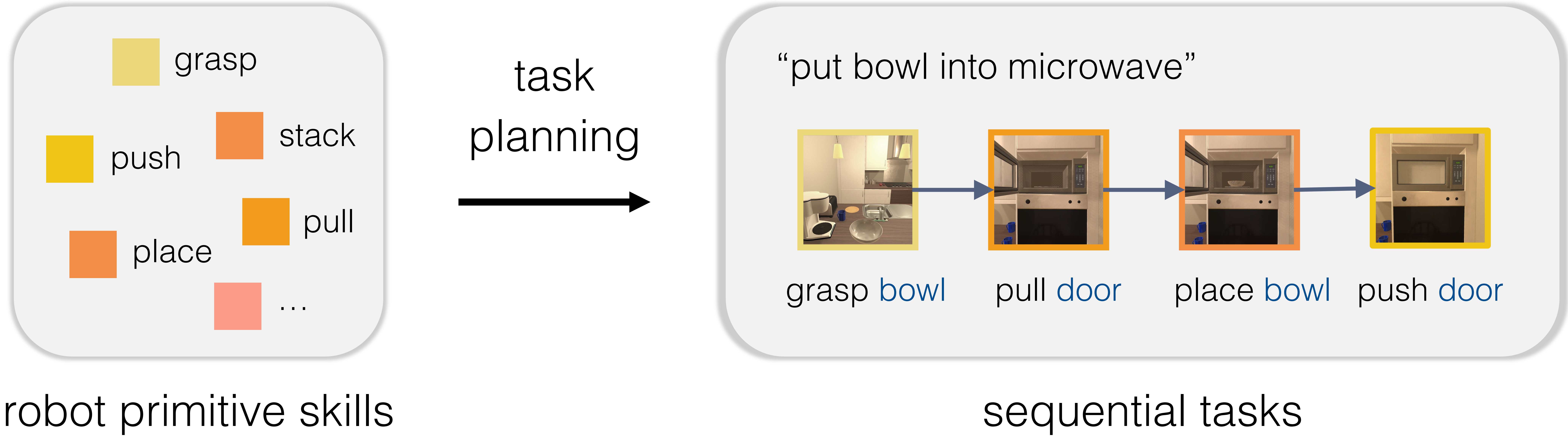
Part II: Sequential Tasks

[**Zhu** et al., ICRA 2017] Target-driven Visual Navigation […]

[**Zhu** et al., ICCV 2017] Visual Semantic Planning […]

Sequential Tasks

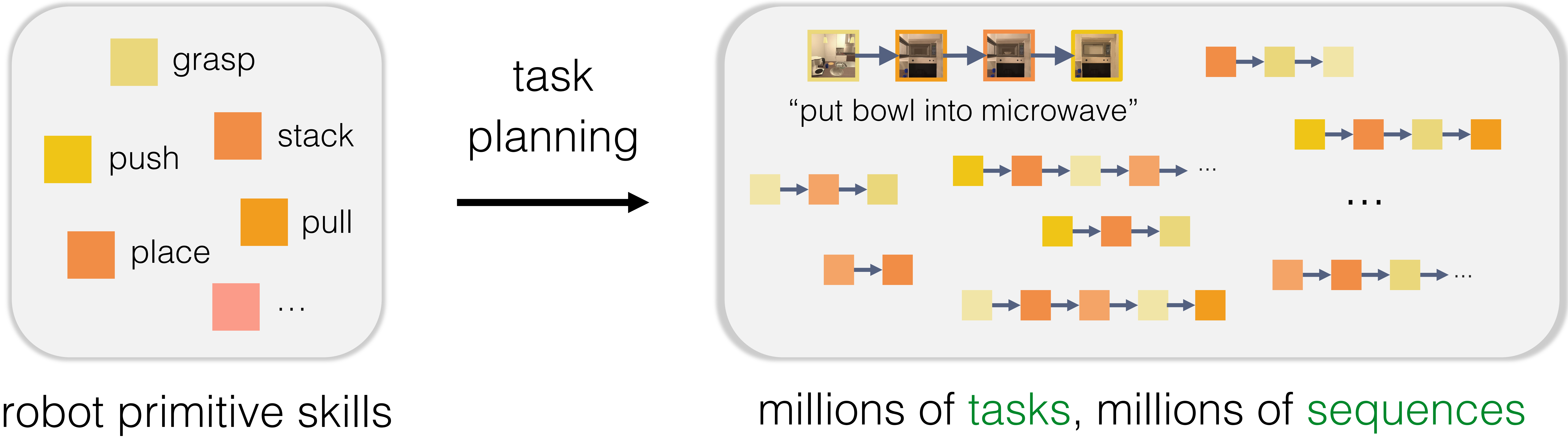




Sequential Tasks

Challenge: very large space of sequential tasks

Sharing and transferring knowledge across tasks



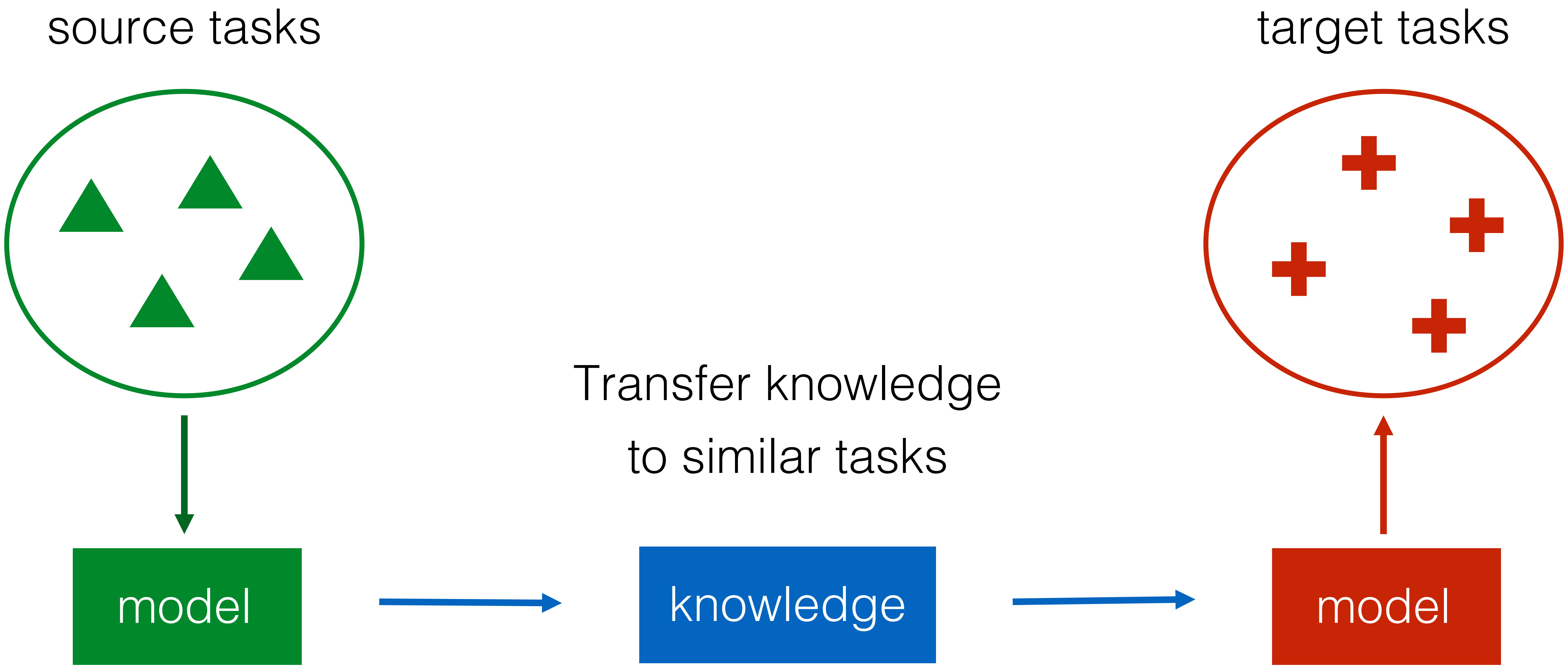
Prior knowledge is used to learn new tasks faster.

prior knowledge

similar new task

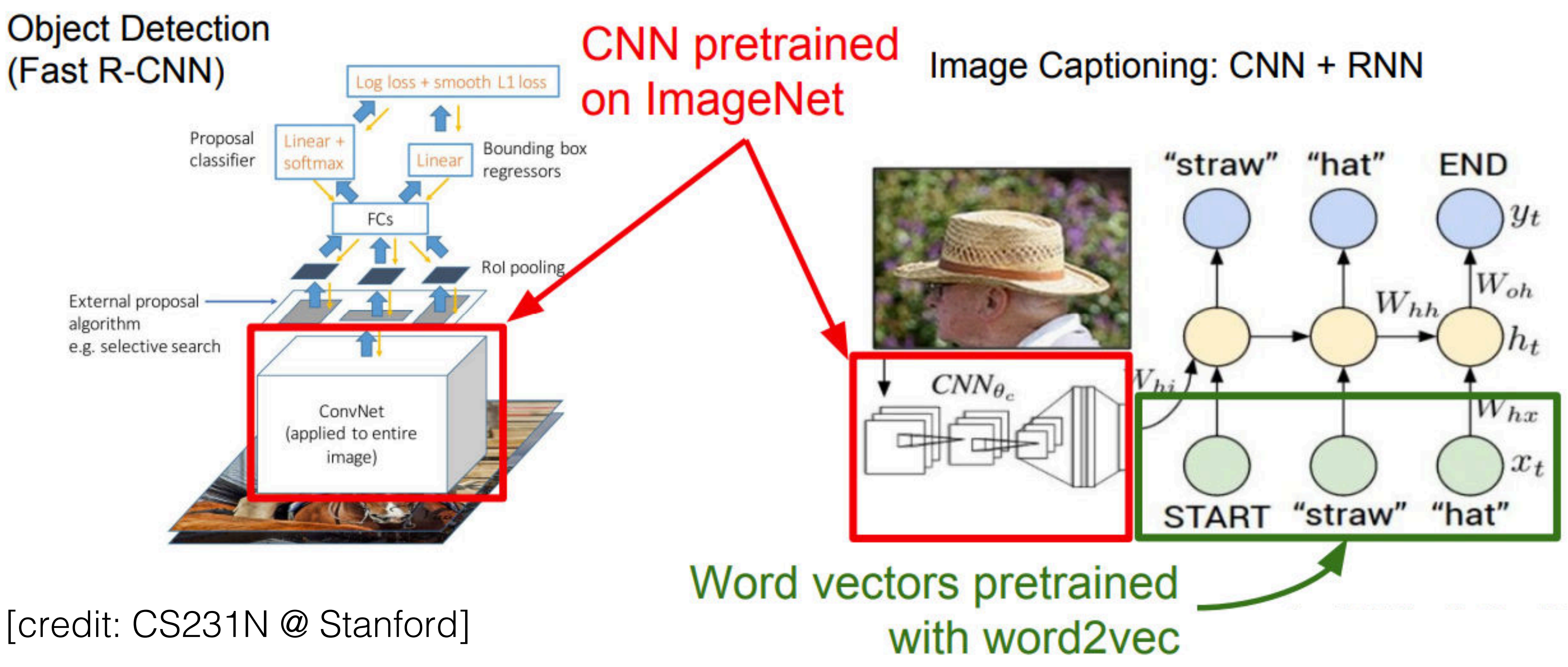
Piaget 1977; Bernstein et al. 1988; Adolph & Berger, 2006; Tuckman & Monetti, 2011; Lake et al., 2016; Tsividis et al., 2017; Dubey et al., 2018

Transfer Learning



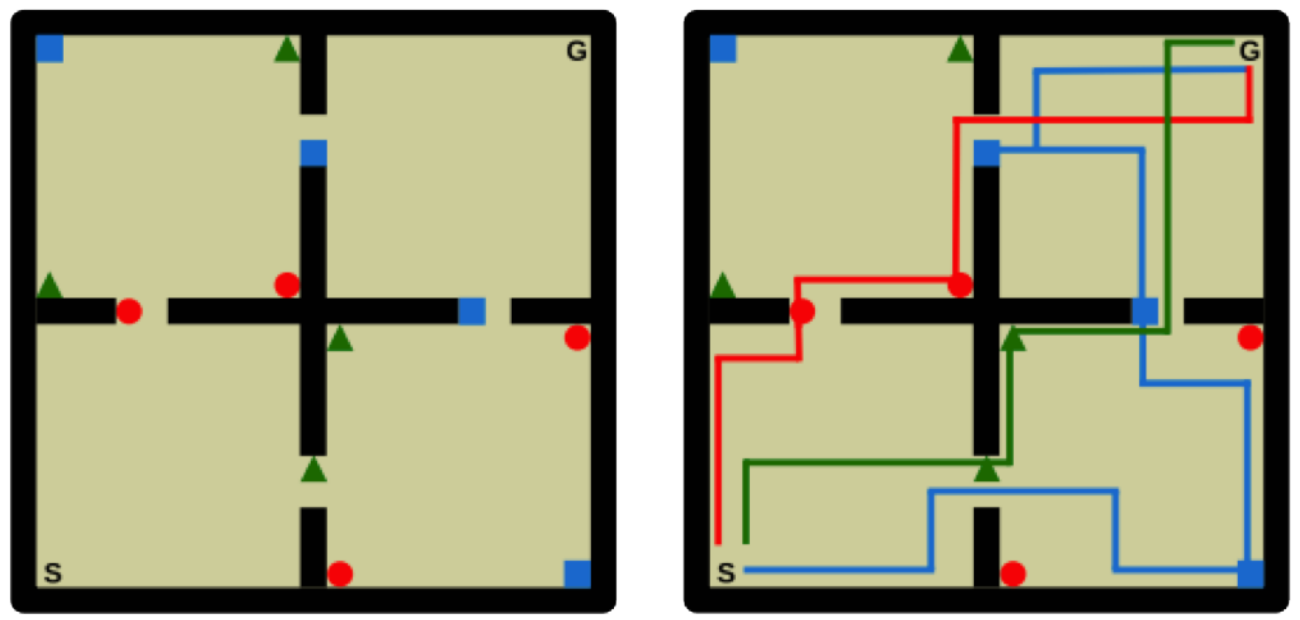
Daumé III *ACL* 2007; Bingel & Søgaard *EACL* 2017; Bousmalis et al. *NIPS* 2016; Ganin & Lempitsky *ICML* 2015; Ganin et al. *JMLR* 2016; Kirkpatrick et al. *PNAS* 2017; Li & Hoiem *PAMI* 2017; Luo et al. *NIPS* 2017
Caruana *ICML* 1993; Baxter *ML* 1997; Duong et al. *ACL* 2015; Schaul et al. *ICML* 2015; Parisotto et al. *ICLR* 2016; Long et al. *NIPS* 2017; Hashimoto et al. *EMNLP* 2017; Rusu et al. *CoRL* 2017; Lu et al. *CVPR* 2017

Transfer Learning



[credit: CS231N @ Stanford]

Atari games



grid worlds

Transfer Learning in Computer Vision

[Zeiler & Fergus 2014; Mahendran & Vedaldi 2015; Huh et al. 2016]

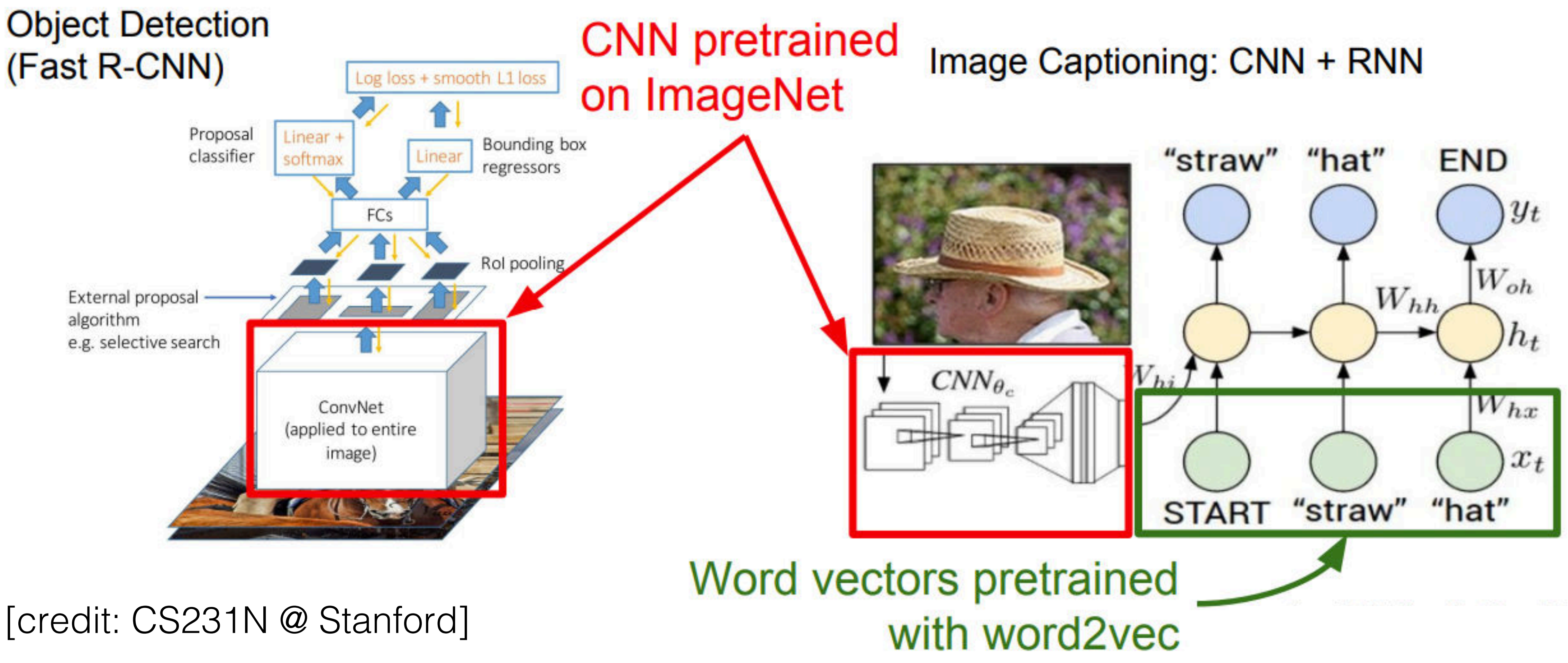
pervasive

Transfer Learning in Interactive Tasks

[Rusu et al. 2016; Parisotto et al. 2016; Oh et al. 2017; Barreto et al. 2018]

limited

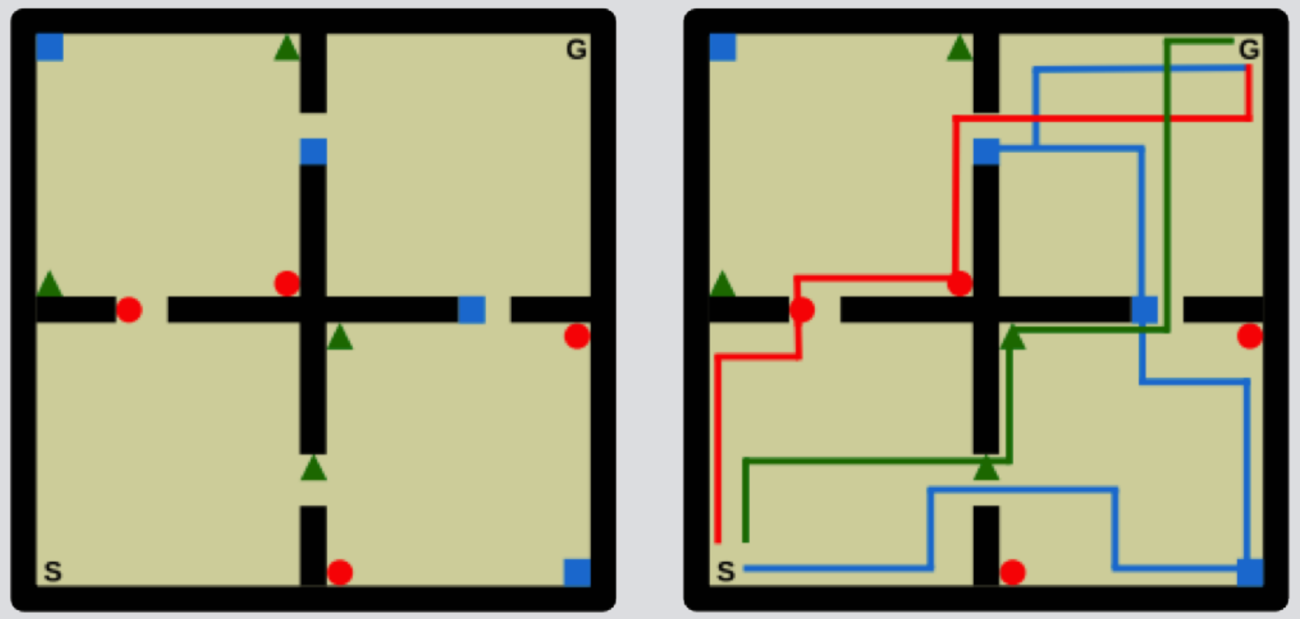
Transfer Learning



[credit: CS231N @ Stanford]

We need a new **platform**!

Atari games



grid worlds

Transfer Learning in Computer Vision

[Zeiler & Fergus 2014; Mahendran & Vedaldi 2015; Huh et al. 2016]

pervasive

Transfer Learning in Interactive Tasks

[Rusu et al. 2016; Parisotto et al. 2016; Oh et al. 2017; Barreto et al. 2018]

limited

AI2-THOR: A New Platform for Visual AI

ai2thor.allenai.org

AI2-THOR: A New Platform for Visual AI

ai2thor.allenai.org

* developed in Unity 3D game engine

- changing viewpoints
- walking and jumping
- applying forces
- picking & placing
- opening & closing

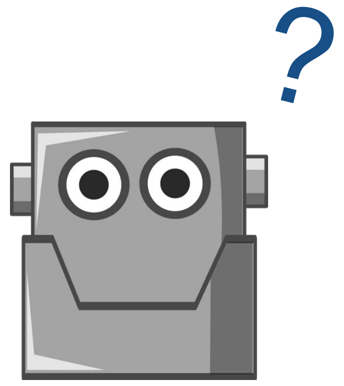
Zhu, Mottaghi, Kolve, Lim, Gupta, Fei-Fei, Farhadi, *ICRA*, 2017

Visual Task Planning

Interactive Visual Environment

Input

agent's view



Task

Putting bowl into microwave

Output a sequence of high-level commands

<start>

→

navigate to table

→

pick up bowl

…

open microwave

→

put bowl

Zhu*, Gordon*, Kolve, Fox, Fei-Fei, Gupta, Mottaghi, Farhadi, *ICCV* 2017

Target-driven Visual Navigation

Input

visual observation

target

Output

target-driven navigation policy*

* Domain adaption with model trained in AI2-THOR

Zhu, Mottaghi, Kolve, Lim, Gupta, Fei-Fei, Farhadi, *ICRA*, 2017

Visual Task Planning

Goal-directed policy learning

Putting bowl into microwave $\xrightarrow{\pi}$ Navigate to Table

$$\pi(\underbrace{a}_{\text{action}}|\underbrace{s}_{\text{state}}, \underbrace{g}_{\text{goal}}) = \arg\max_a Q(s, a, g)$$

Q: expected sum of future rewards approximated by **neural network**

The goal-conditional Bellman equation

$$Q^*(s, a, g) = \mathbb{E}_\pi[\underbrace{r_g}_{\text{immediate reward of task g}}(s, a) + \gamma \max_{a'} Q^*(\underbrace{s'}_{\text{next state}}, \underbrace{a'}_{\text{next action}}, g)]$$

Zhu*, Gordon*, Kolve, Fox, Fei-Fei, Gupta, Mottaghi, Farhadi, *ICCV* 2017

Visual Task Planning

Key idea: decoupling environment dynamics and goal specification

goal-directed Q-value

successor feature[1]

$$Q(s,a,g) \approx \psi(s,a)^\top \times \mathbf{w}_g$$

$\mathbf{w}_g$ goal embedding

[1]Dayan 1993; Kulkarni et al., 2016

Zhu*, Gordon*, Kolve, Fox, Fei-Fei, Gupta, Mottaghi, Farhadi, *ICCV* 2017

Visual Task Planning

Key idea: decoupling environment dynamics and goal specification

$$\underbrace{\boxed{r_g(s,a)}}_{\text{reward}} = \underbrace{\boxed{\phi(s,a)}}_{\text{state-action feature}}^{\top} \underbrace{\boxed{\mathbf{w}_g}}_{\text{goal embedding}}$$

$$Q(s_i, a_i, g) = \mathbb{E}[\sum_{i=t}^{\infty} \gamma^{i-t} r_g(s_i, a_i)]$$

successor feature[1]

$$= \mathbb{E}[\sum_{i=t}^{\infty} \gamma^{i-t} \phi(s_i, a_i)^{\top} \mathbf{w}_g] = \mathbb{E}[\sum_{i=t}^{\infty} \gamma^{i-t} \phi(s_i, a_i)^{\top}] \mathbf{w}_g = \psi(s_i, a_i)^{\top} \mathbf{w}_g$$

[1]Dayan 1993; Kulkarni et al., 2016

Zhu*, Gordon*, Kolve, Fox, Fei-Fei, Gupta, Mottaghi, Farhadi, *ICCV* 2017

Visual Task Planning

Key idea: decoupling environment dynamics and goal specification

$$\underbrace{\boxed{r_g(s,a)}}_{\text{reward}} = \underbrace{\boxed{\phi(s,a)}}_{\text{state-action feature}}^{\top} \underbrace{\boxed{\mathbf{w}_g}}_{\text{goal embedding}}$$

$$\underbrace{\boxed{Q(s,a,g)}}_{\text{Q-value}} = \underbrace{\boxed{\psi(s,a)}}_{\text{successor feature[1]}}^{\top} \boxed{\mathbf{w}_g}$$

[1]Dayan 1993; Kulkarni et al., 2016

Zhu*, Gordon*, Kolve, Fox, Fei-Fei, Gupta, Mottaghi, Farhadi, *ICCV* 2017

Visual Task Planning

action goal

$$\pi(a|s,g) = \arg\max_a Q(s,a,g)$$

state

goal-directed
Q-value

$$Q(s,a,g)$$

$\approx$

goal-independent

environment
dynamics

$$\psi(s,a)^\top$$

$\times$

$\mathbf{w}_g$

goal-specific

goal
specification

Zhu*, Gordon*, Kolve, Fox, Fei-Fei, Gupta, Mottaghi, Farhadi, *ICCV* 2017

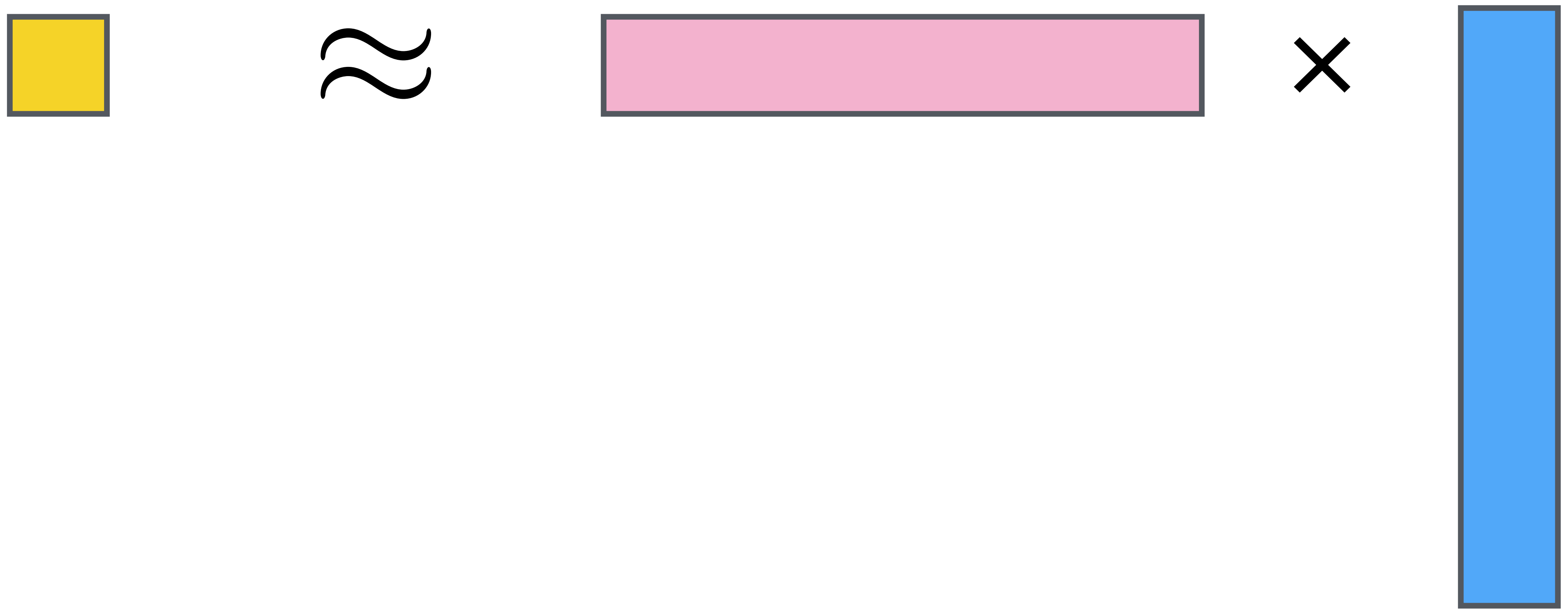
Visual Task Planning

$$\pi(\underbrace{a}_{\text{action}}|\underbrace{s}_{\text{state}},\underbrace{g}_{\text{goal}}) = \arg\max_a Q(s,a,g)$$

searching for apple

shared across tasks

$$Q(s,a,g_0) \approx \psi(s,a)^\top \times \mathbf{w}_{g_0}$$



searching for apple

Zhu*, Gordon*, Kolve, Fox, Fei-Fei, Gupta, Mottaghi, Farhadi, *ICCV* 2017

Visual Task Planning

action goal

$$\pi(a|s, g) = \arg\max_a Q(s, a, g)$$

state

shared across tasks

searching for banana

$$Q(s, a, g_1) \approx \psi(s, a)^\top \times \mathbf{w}_{g_1}$$

searching for banana

Zhu*, Gordon*, Kolve, Fox, Fei-Fei, Gupta, Mottaghi, Farhadi, *ICCV* 2017

Visual Task Planning

action goal

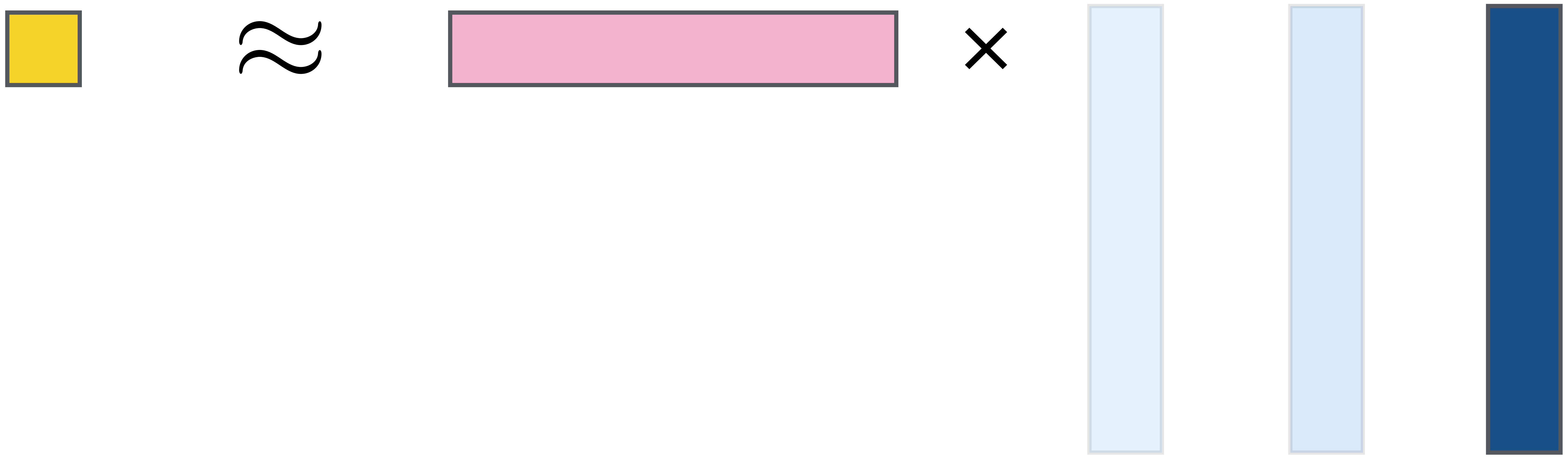
$$\pi(a|s, g) = \arg\max_a Q(s, a, g)$$

state

putting apple to fridge

shared across tasks

$$Q(s, a, g_2) \approx \psi(s, a)^\top \times \mathbf{w}_{g_2}$$

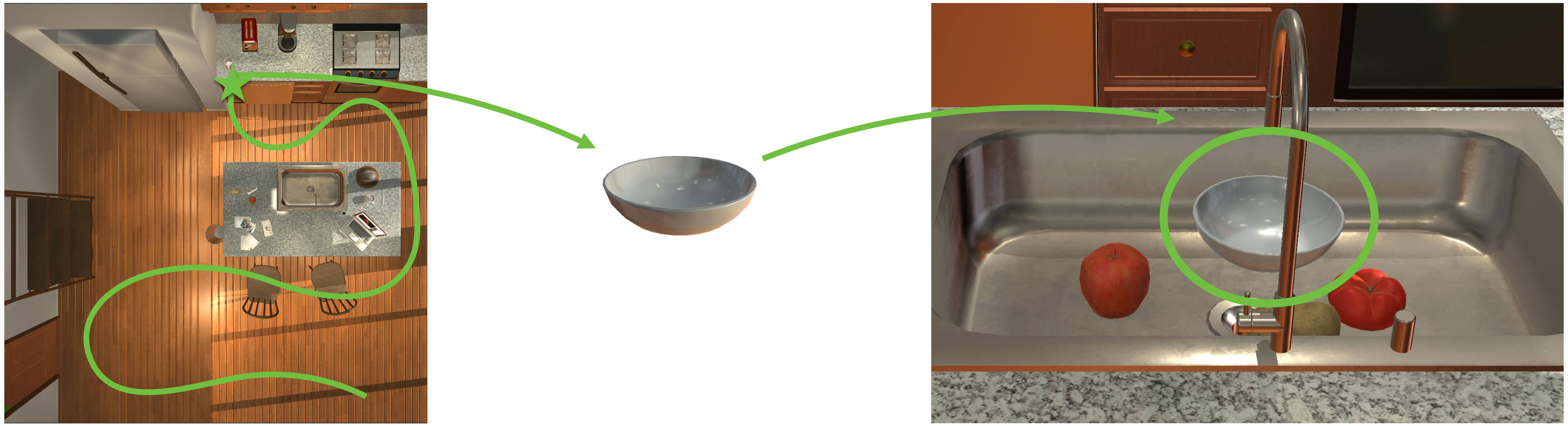


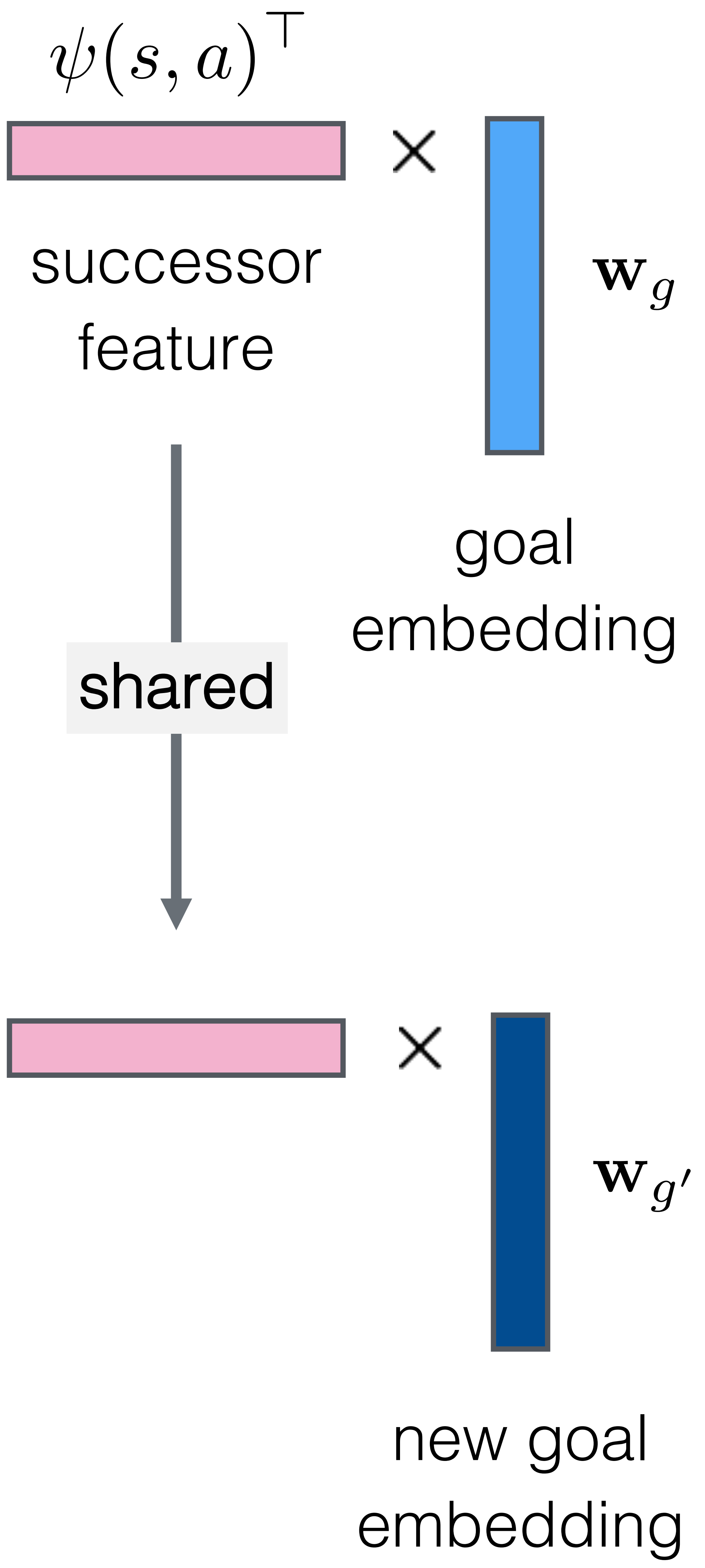
putting apple to fridge

Zhu*, Gordon*, Kolve, Fox, Fei-Fei, Gupta, Mottaghi, Farhadi, *ICCV* 2017

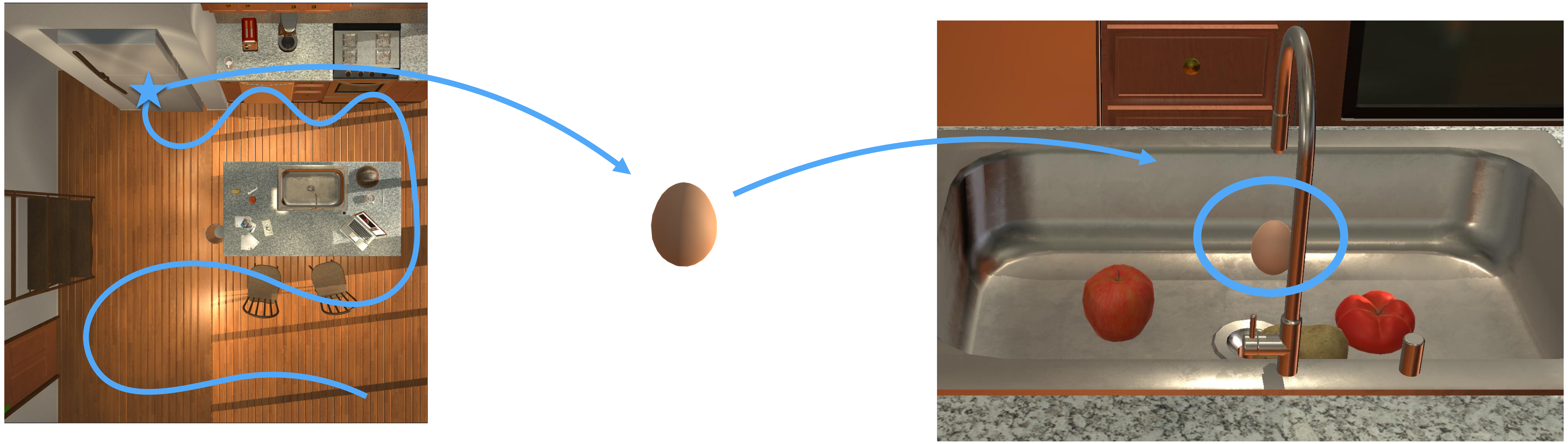
Trained Task: searching for a bowl

Scene Object Goal





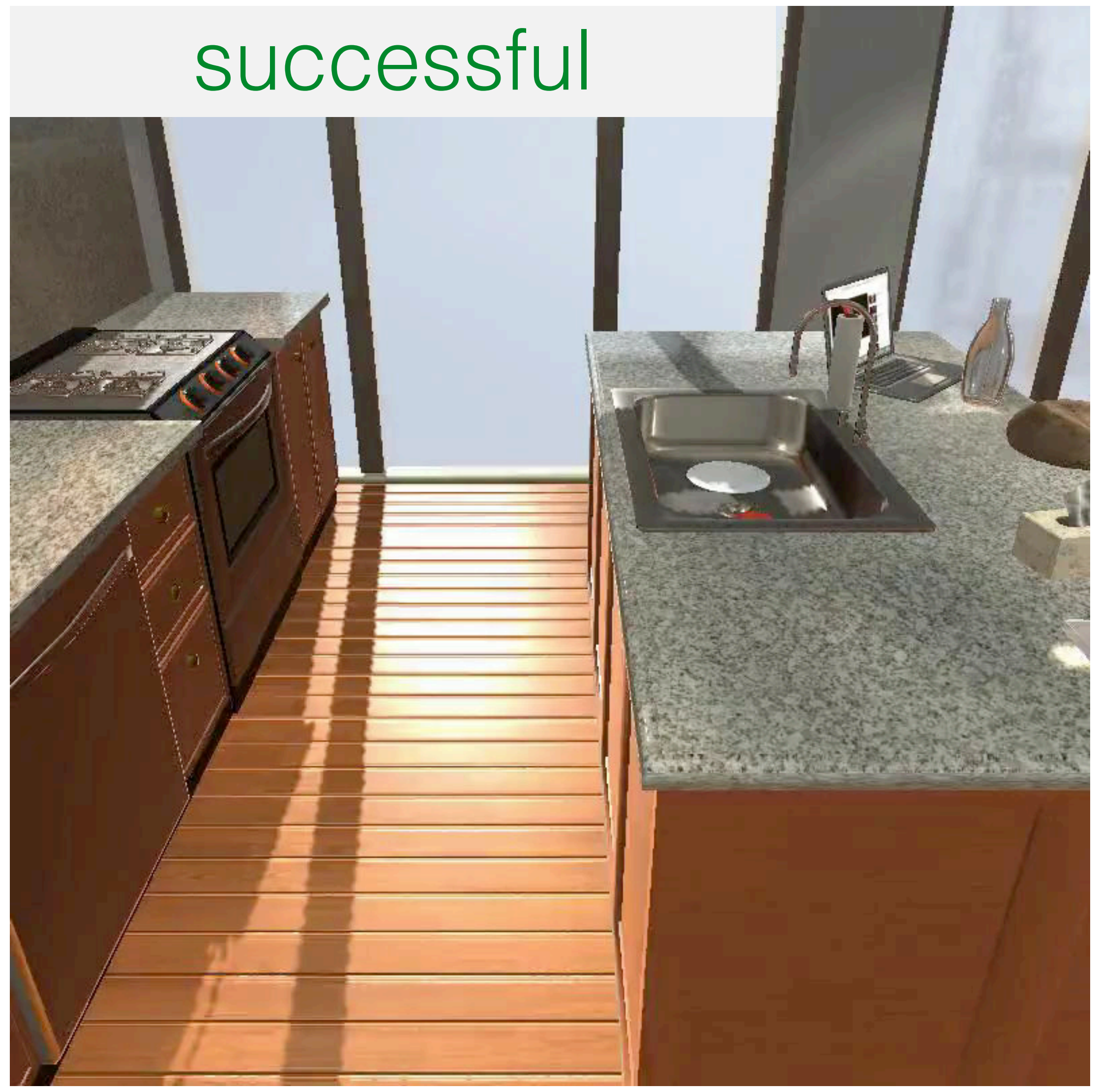
New task: searching for an egg



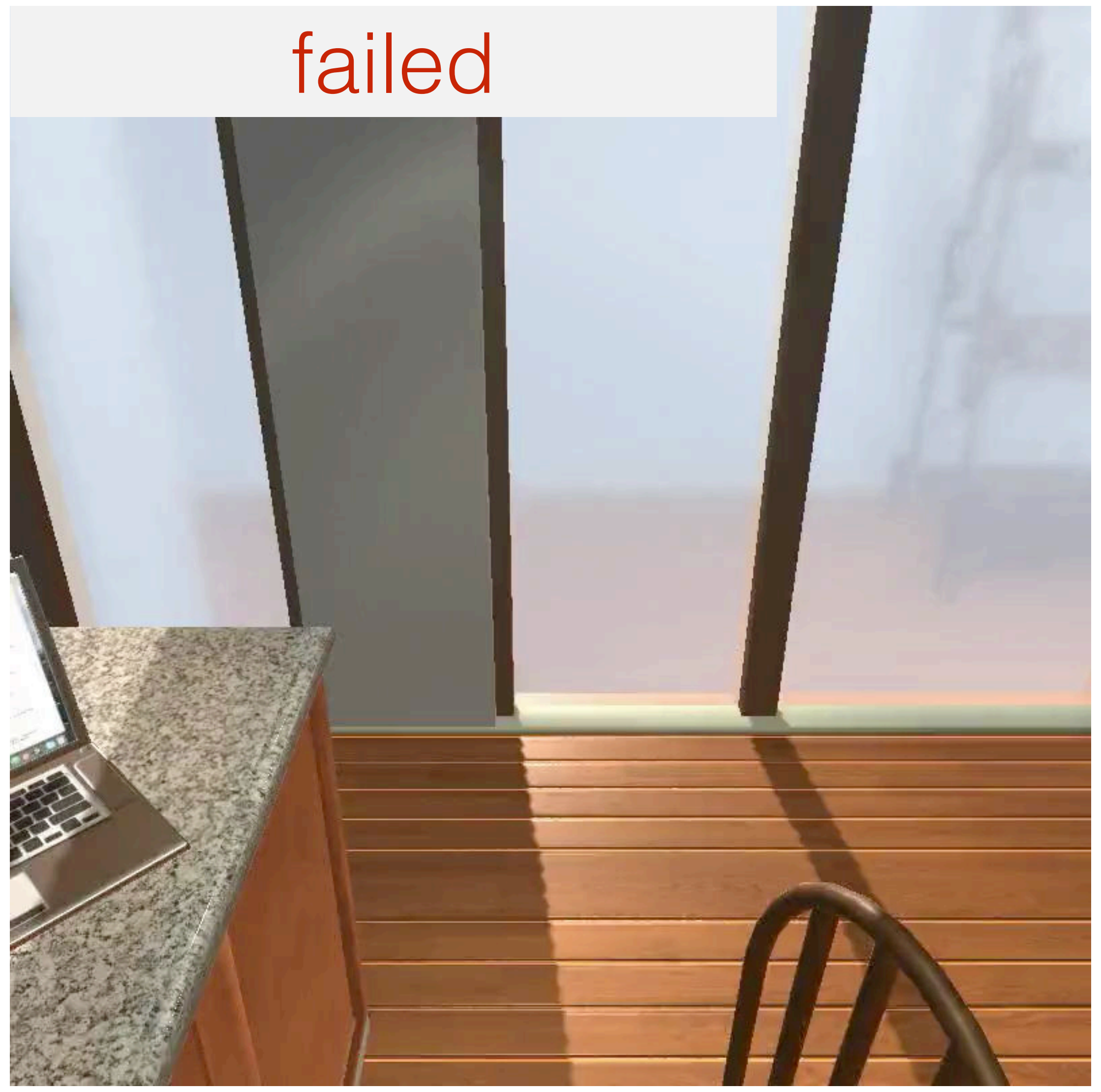
Zhu*, Gordon*, Kolve, Fox, Fei-Fei, Gupta, Mottaghi, Farhadi, *ICCV* 2017

Fast Policy Transfer with New Goal Embedding

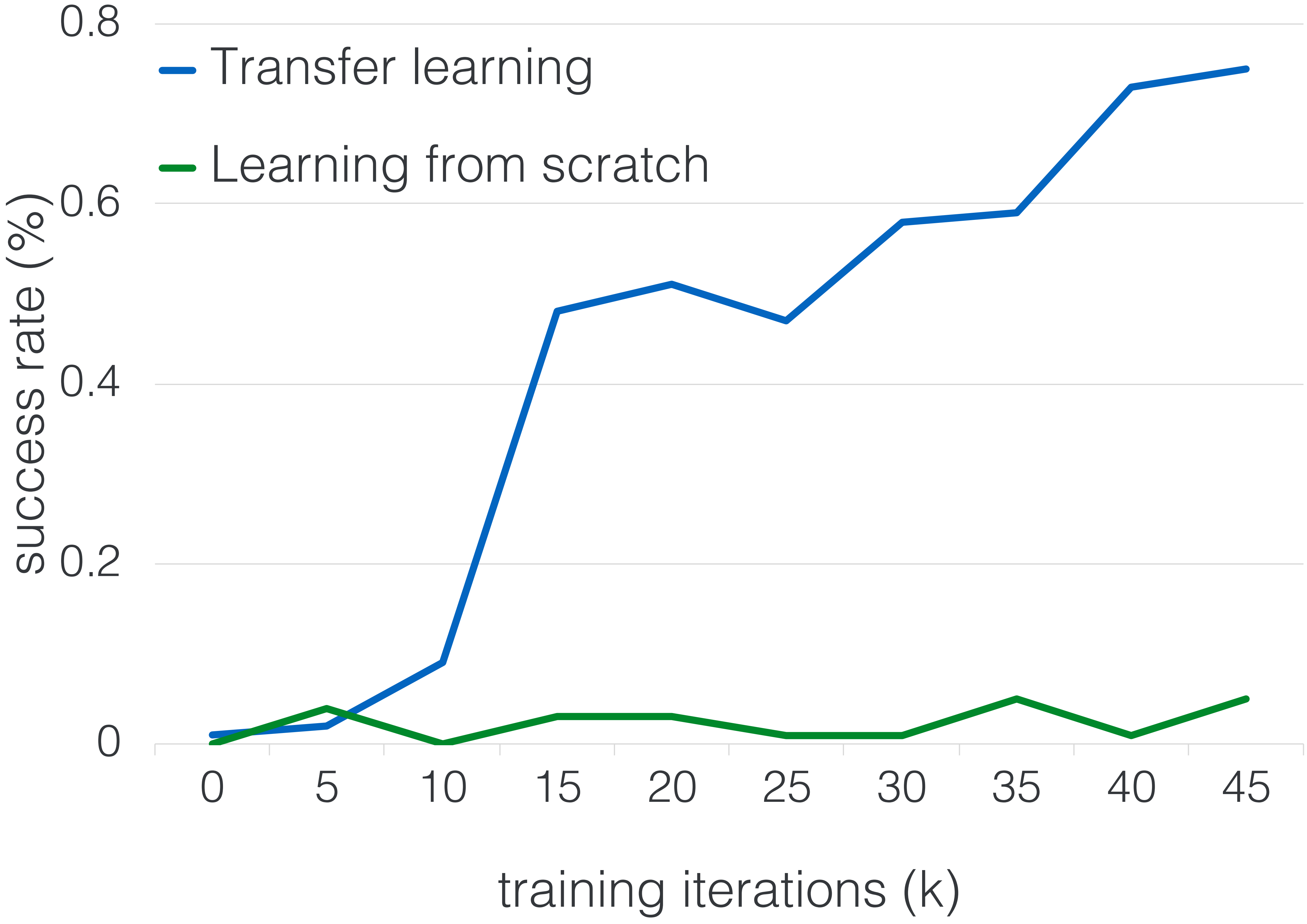
Task: Search for an egg and put it into the sink



Transfer Learning



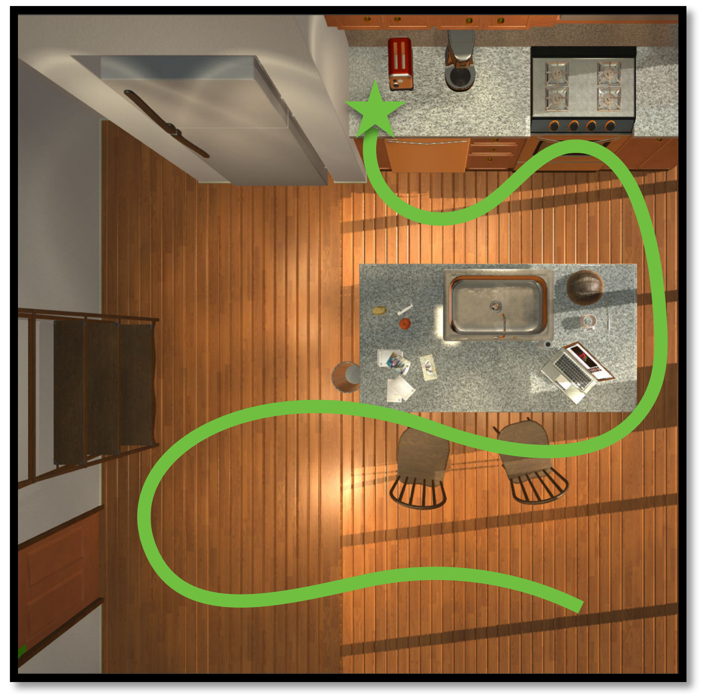
Learning from scratch



Zhu*, Gordon*, Kolve, Fox, Fei-Fei, Gupta, Mottaghi, Farhadi, *ICCV* 2017

Summary - Part II

Developed a 3D virtual world (AI2-THOR) to study embodied agents in **interactive visual environments**



Transfer learning between **sequential tasks** through the decoupling of environment dynamics and goal specification

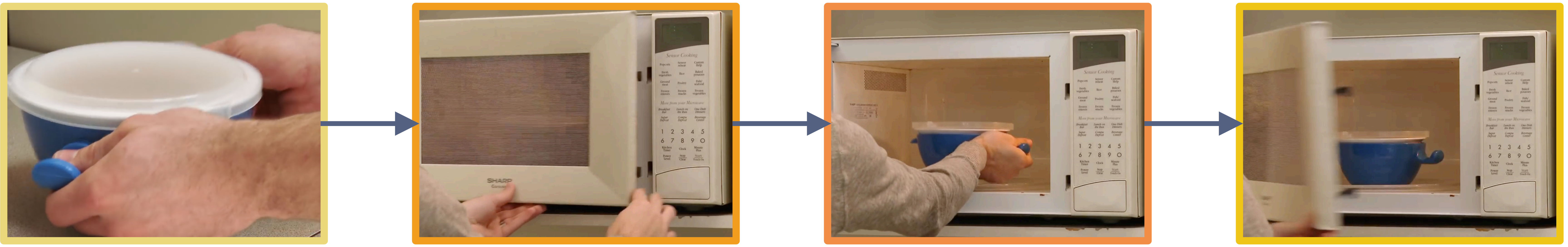
Part III: Hierarchical Tasks

[Xu*, Nair*, **Zhu**, et al. ICRA 2018] Neural Task Programming (NTP) […]

[Huang*, Nair*, Xu*, **Zhu** et al. CVPR 2019] Neural Task Graphs (NTG) […]

Hierarchical Tasks

“put bowl into microwave”



grasp bowl   pull door   place bowl

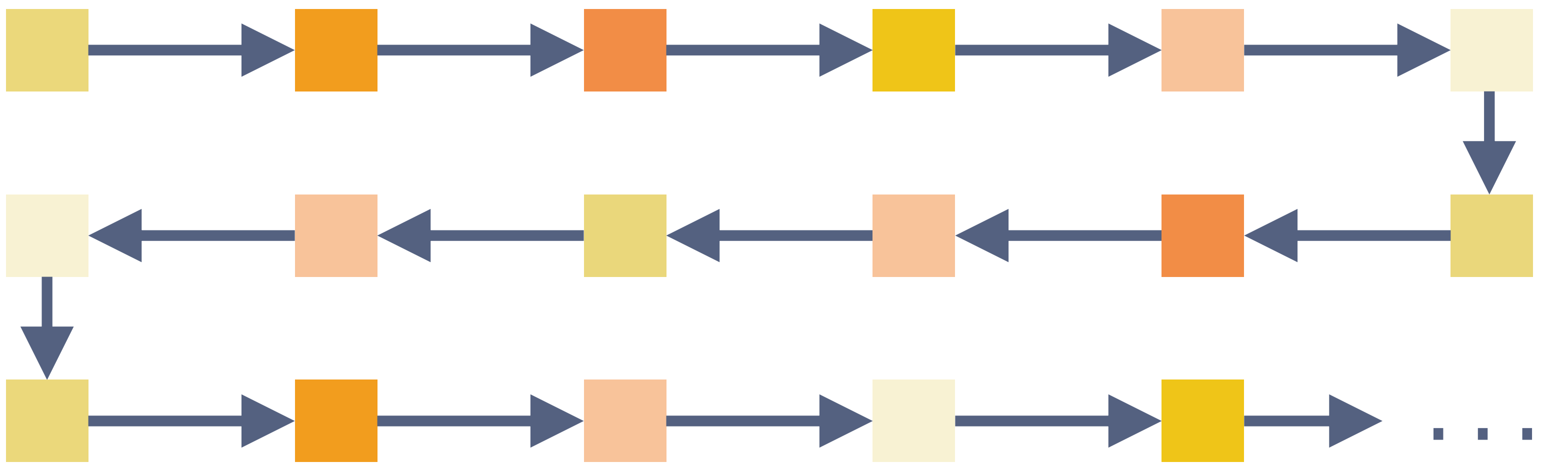
“prepare dinner”

Complex real-world tasks are hierarchical.

Hierarchical Tasks

Challenge: Task complexity grows exponentially.

"prepare dinner"



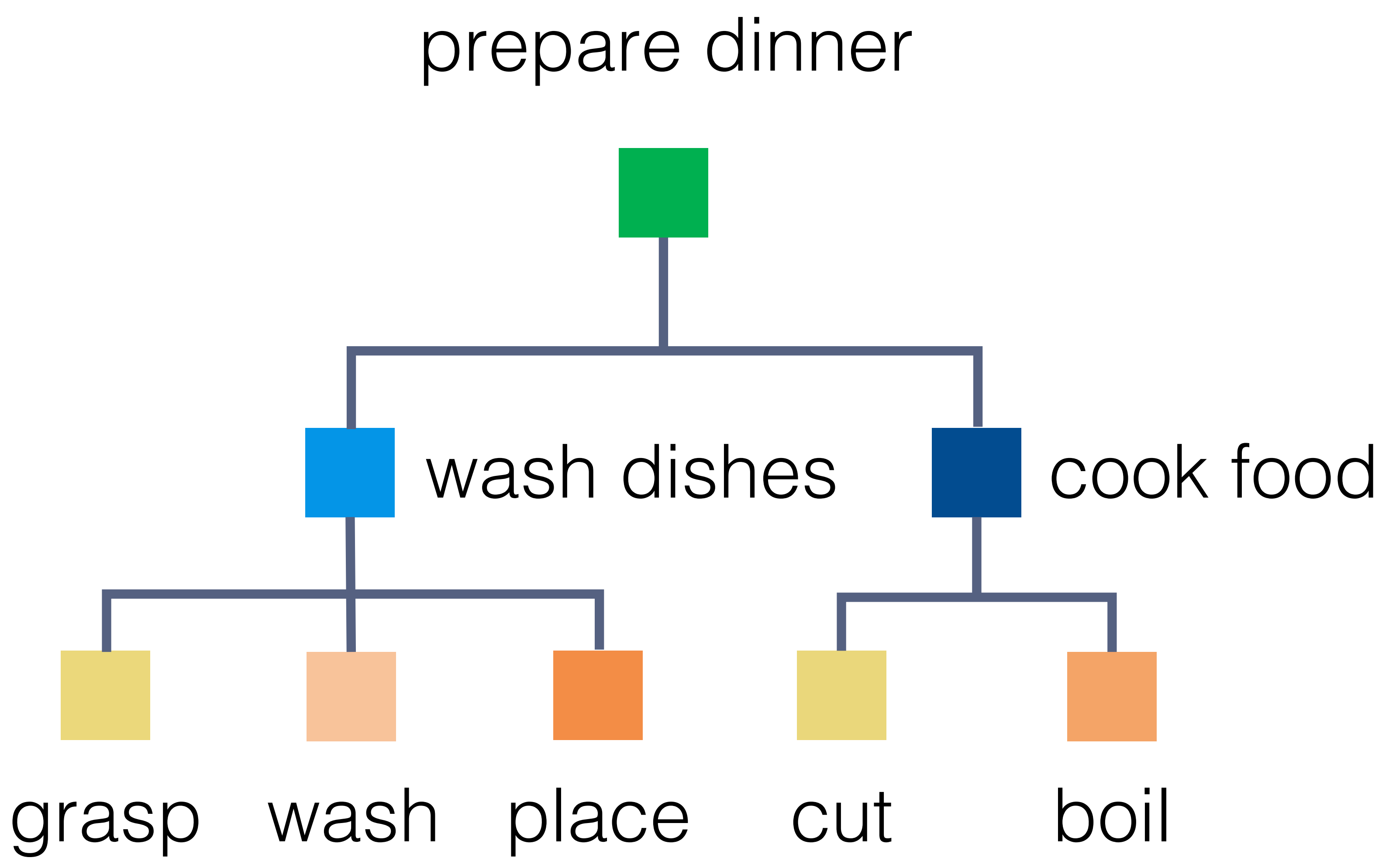
Intractable!

Hierarchical Tasks

Challenge: Task complexity grows exponentially.

Leveraging the compositionality of hierarchical tasks

“prepare dinner”



Videos supply massive knowledge to solve new tasks.

THE VERGE TECH SCIENCE CULTURE CARS REVIEWS LONGFORM VIDEO MORE

TECH YOUTUBE CULTURE

Half of YouTube viewers use it to learn how to do things they've never done

10

Some of us are on there just to pass the time, though

By Patricia Hernandez | @xpatriciah | Nov 7, 2018, 12:36pm EST

SHARE

how to

how to make slime
how to tie a tie
how to draw
how to basic
how to get boogie down dance
how to cake it
how to train your dragon 3
how to get the galaxy skin in fortnite
how to make slime without glue
how to solve a rubik's cube

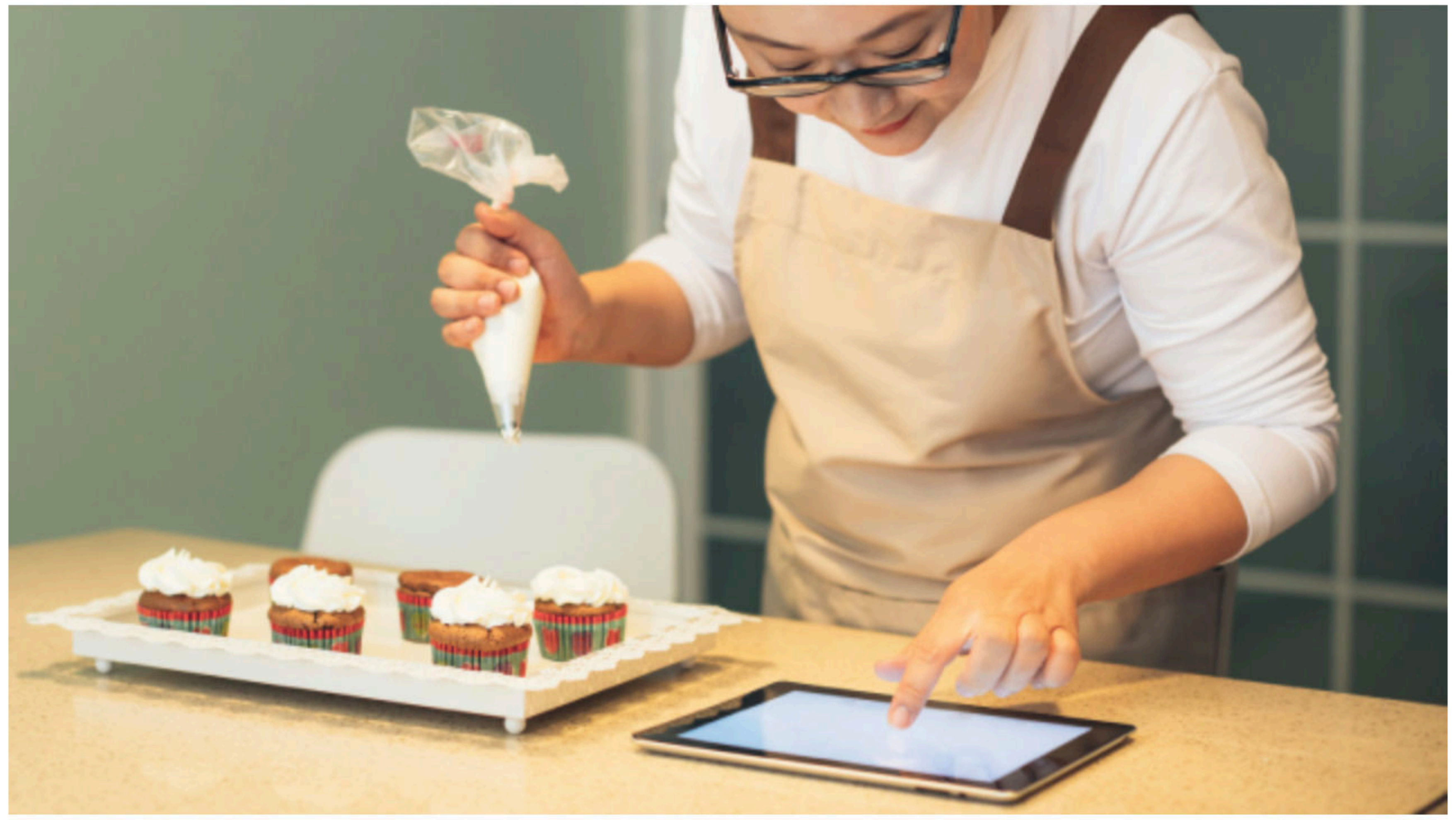
Report search predictions

NOVEMBER 7, 2018

Many Turn to YouTube for Children's Content, News, How-To Lessons

An analysis of videos suggested by the site's recommendation engine finds that users are directed toward progressively longer and more popular content

BY AARON SMITH, SKYE TOOR AND PATRICK VAN KESSEL



(MaaHoo Studio/Getty Images)

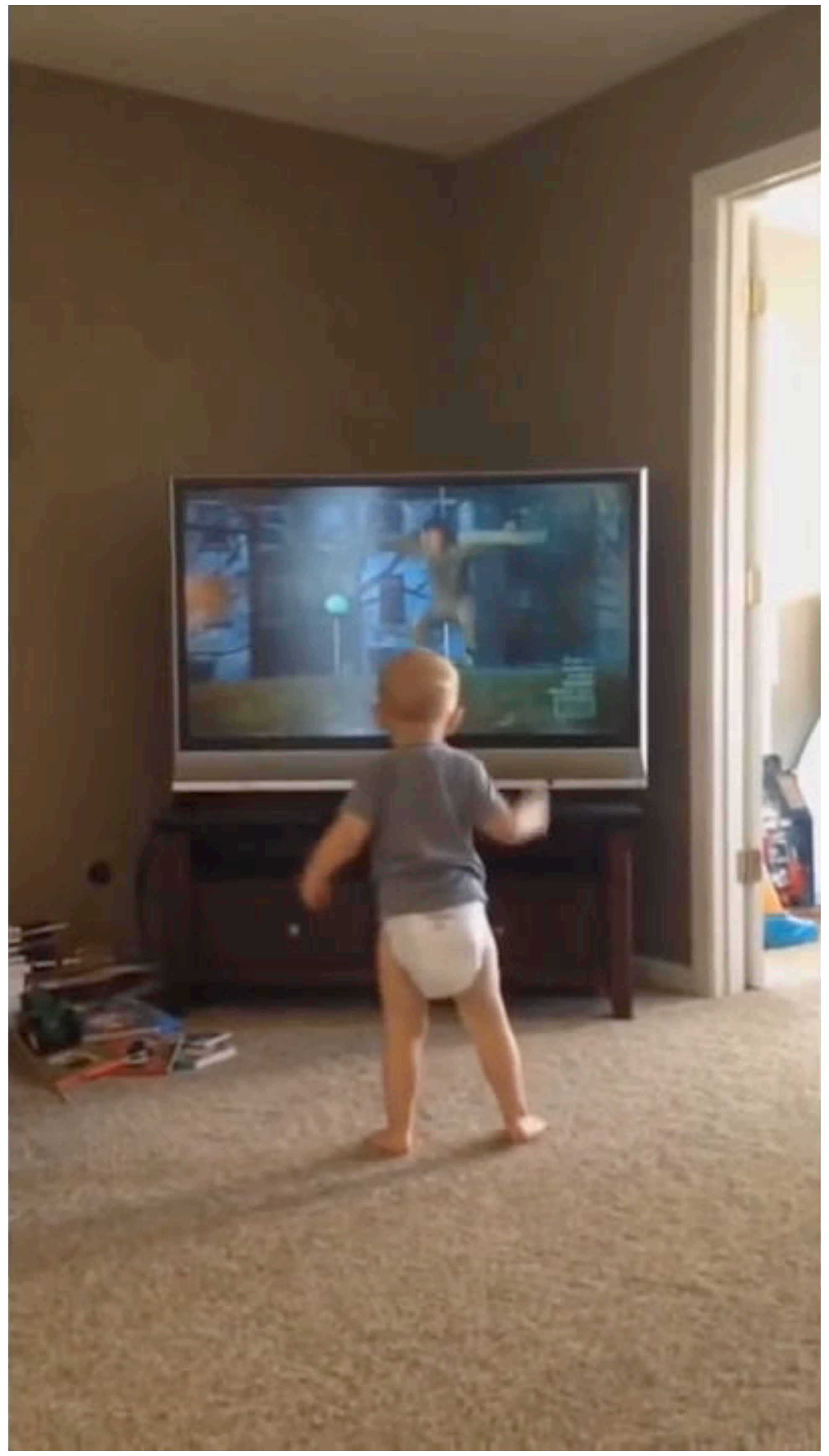
Source: The Verge, Pew Research Center

Humans learn efficiently from **video demonstrations**.

Imitation of Televised Models by Infants

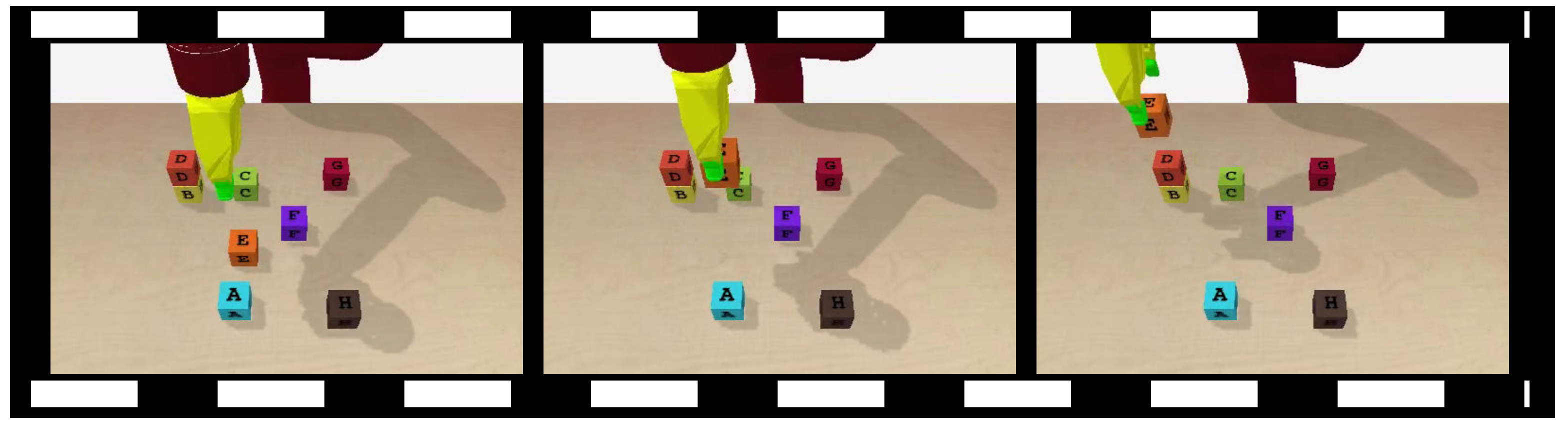
Andrew N. Meltzoff, *Child Development* 1988

Babies (14-24 months) can learn by imitating demonstrations from the TV screen.

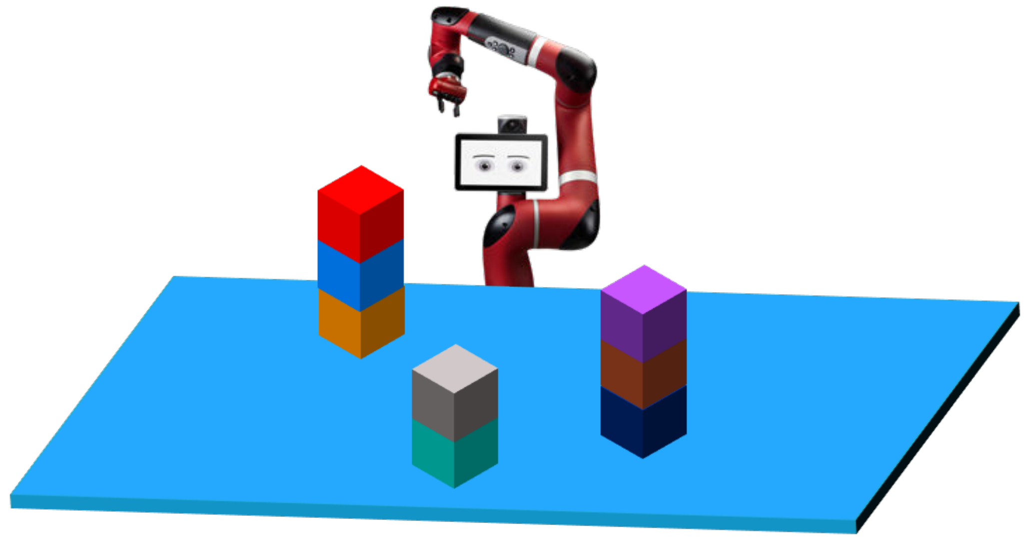


Meltzoff & Moore 1977; Meltzoff & Moore 1989, Meltzoff 1988

One-Shot Imitation Learning from Videos



meta-learning model

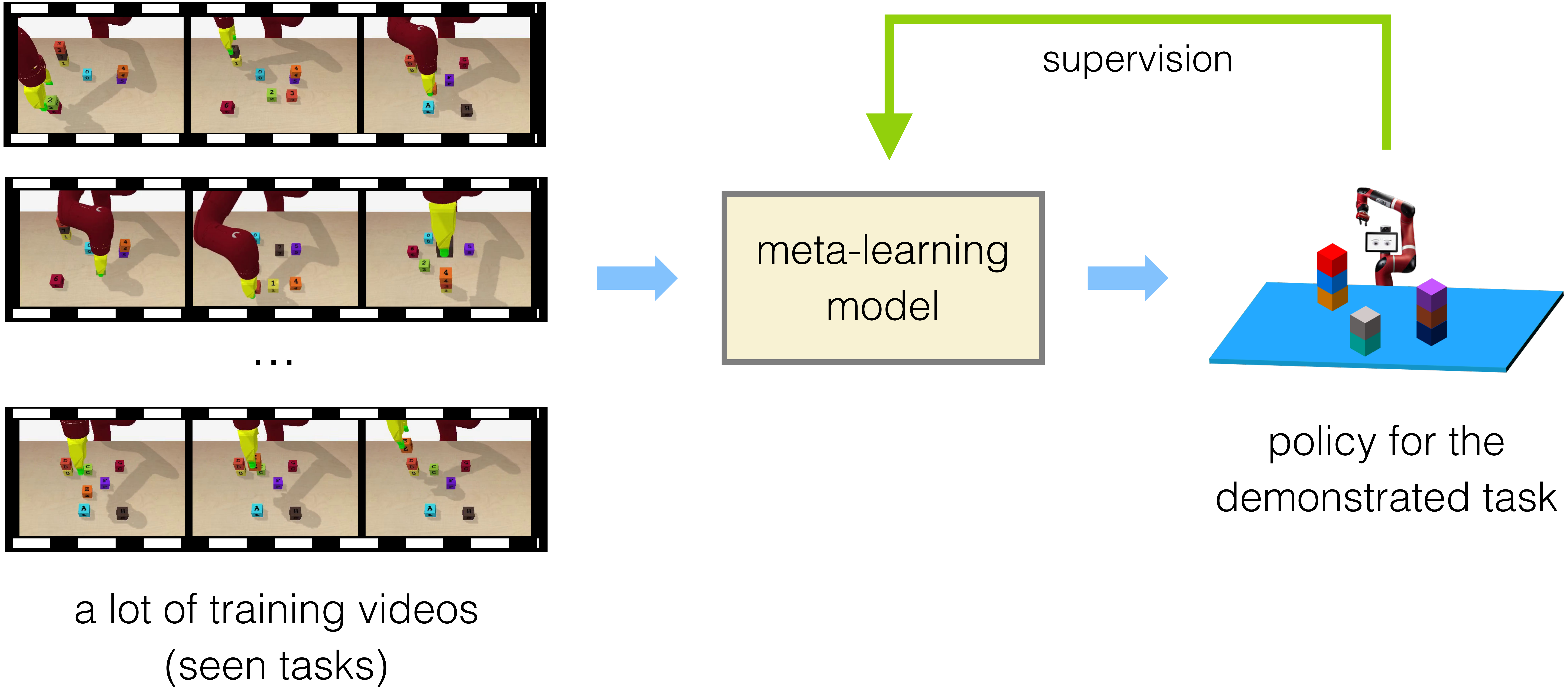


single video demonstration

policy for the demonstrated task

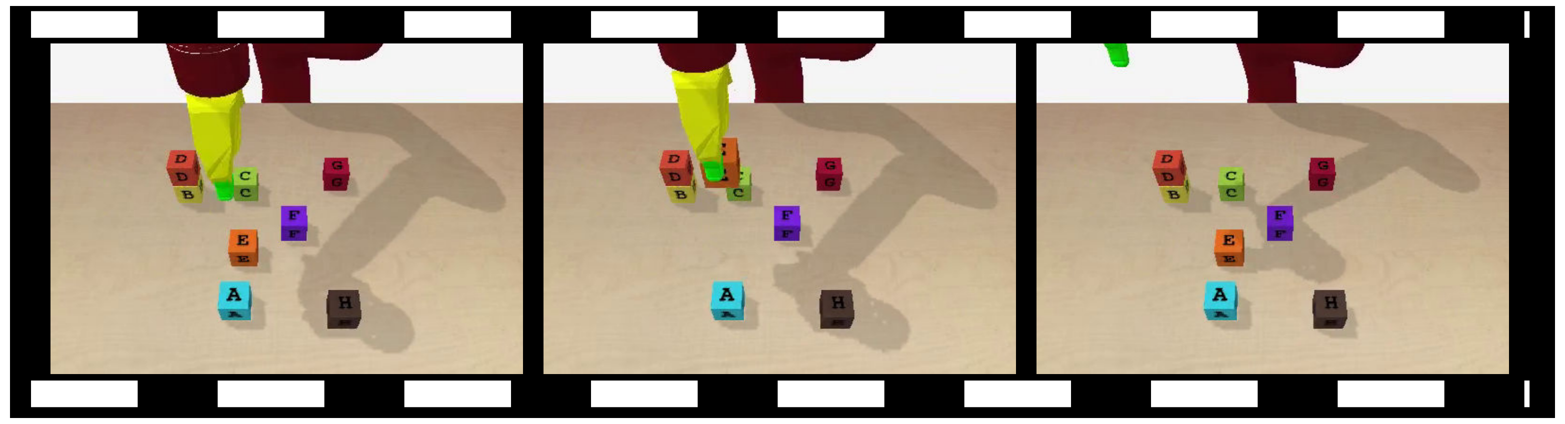
Xu*, Nair*, **Zhu**, Gao, Garg, Fei-Fei, Savarese. *ICRA 2018*

One-Shot Imitation Learning from Videos



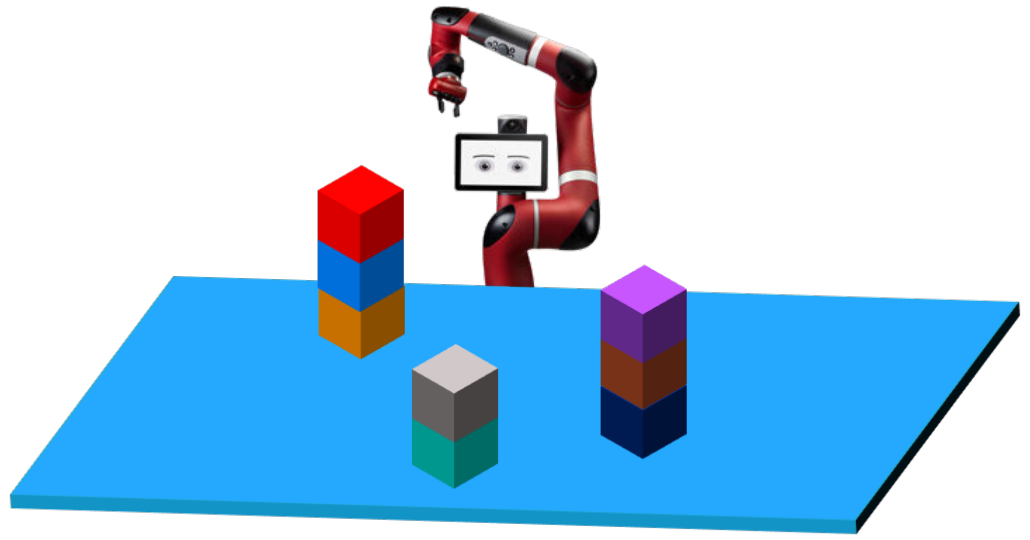
Xu*, Nair*, **Zhu**, Gao, Garg, Fei-Fei, Savarese. *ICRA 2018*

One-Shot Imitation Learning from Videos



single test video
(unseen task)

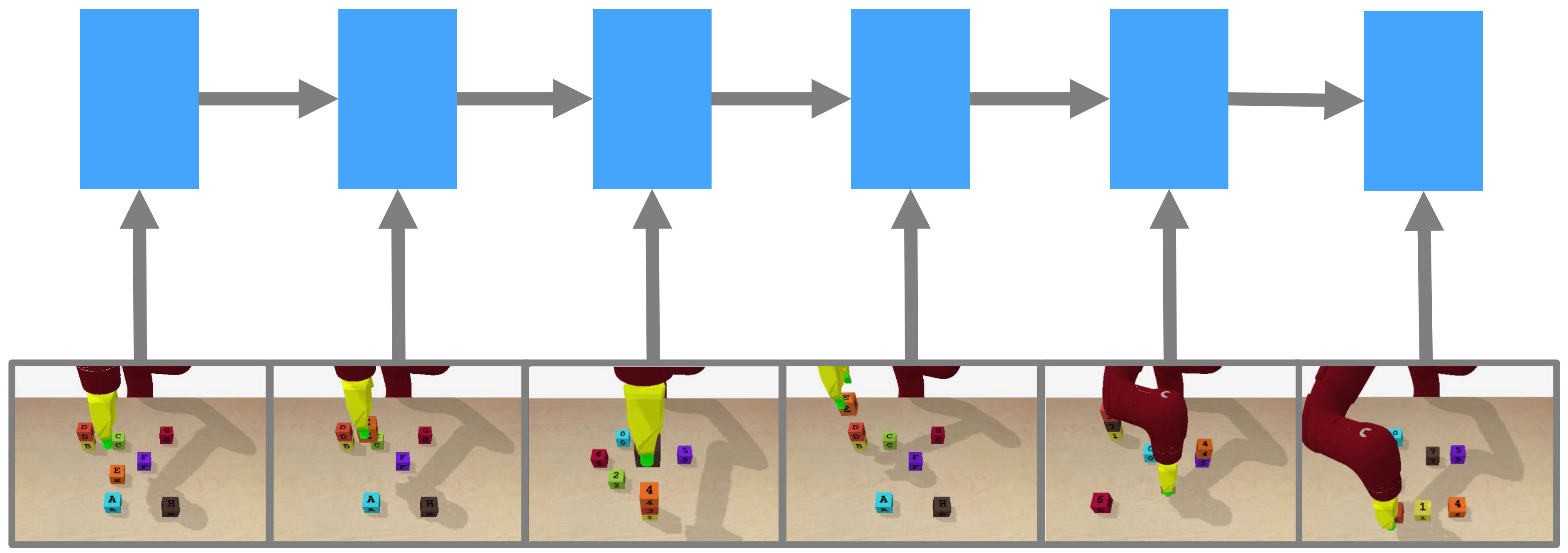
→ meta-learning model →



policy for the
demonstrated task

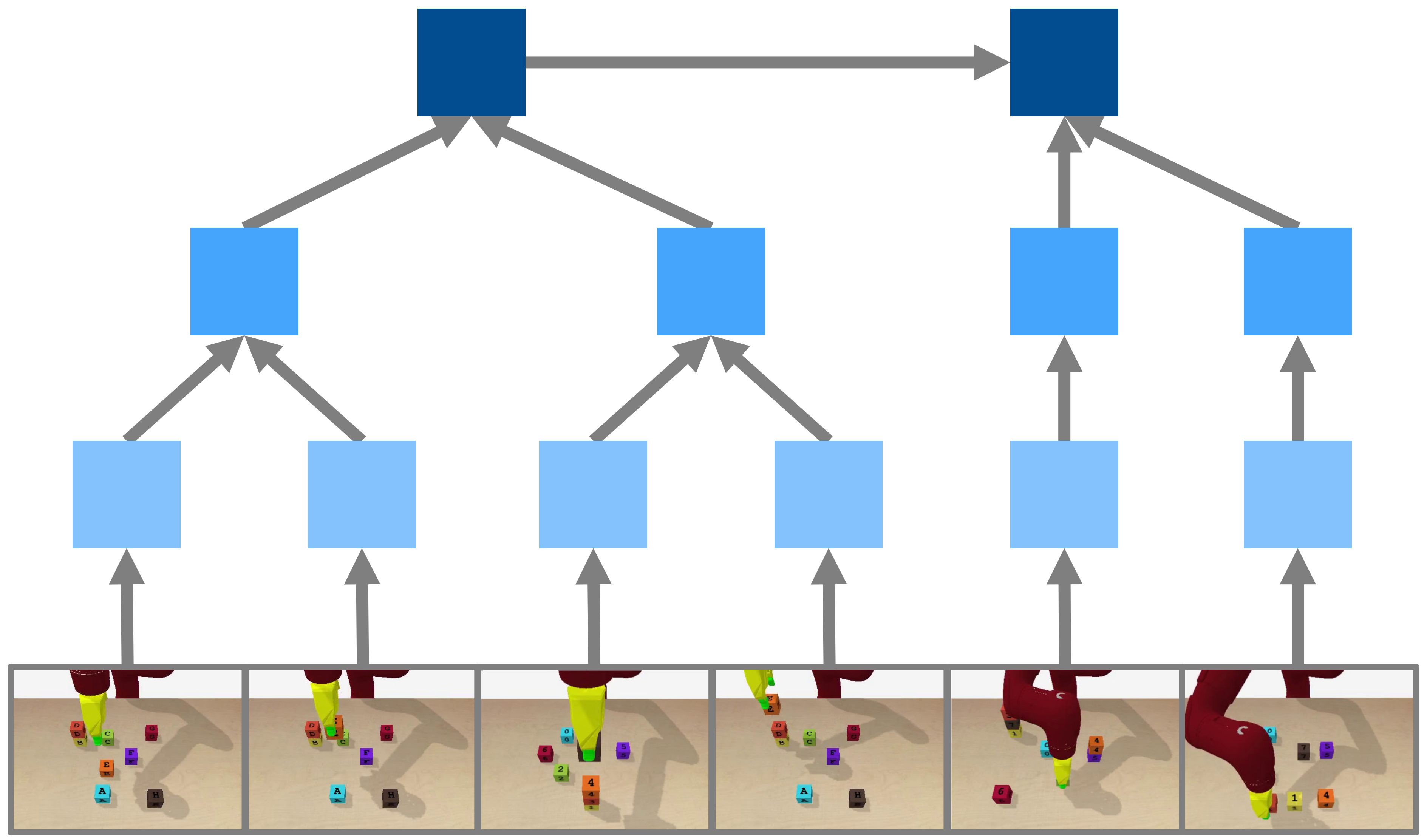
Xu*, Nair*, **Zhu**, Gao, Garg, Fei-Fei, Savarese. *ICRA 2018*

One-Shot Imitation Learning from Videos



[Duan et al. 17; Finn et al. 2017; Wang et al. 2017; Yu et al. 2018]

modeling demonstration
as a **flat sequence**

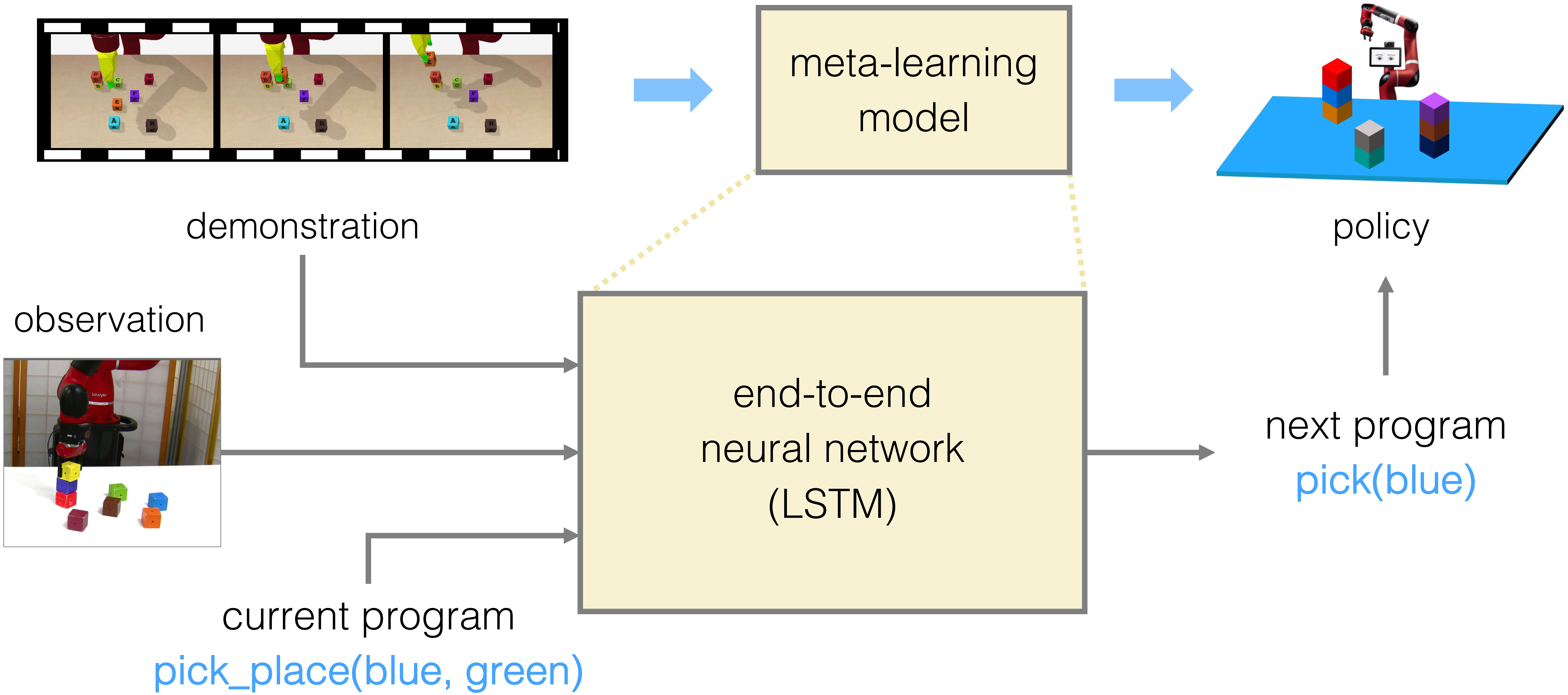


modeling demonstration
as a **compositional structure**

Xu*, Nair*, **Zhu**, Gao, Garg, Fei-Fei, Savarese. *ICRA 2018*

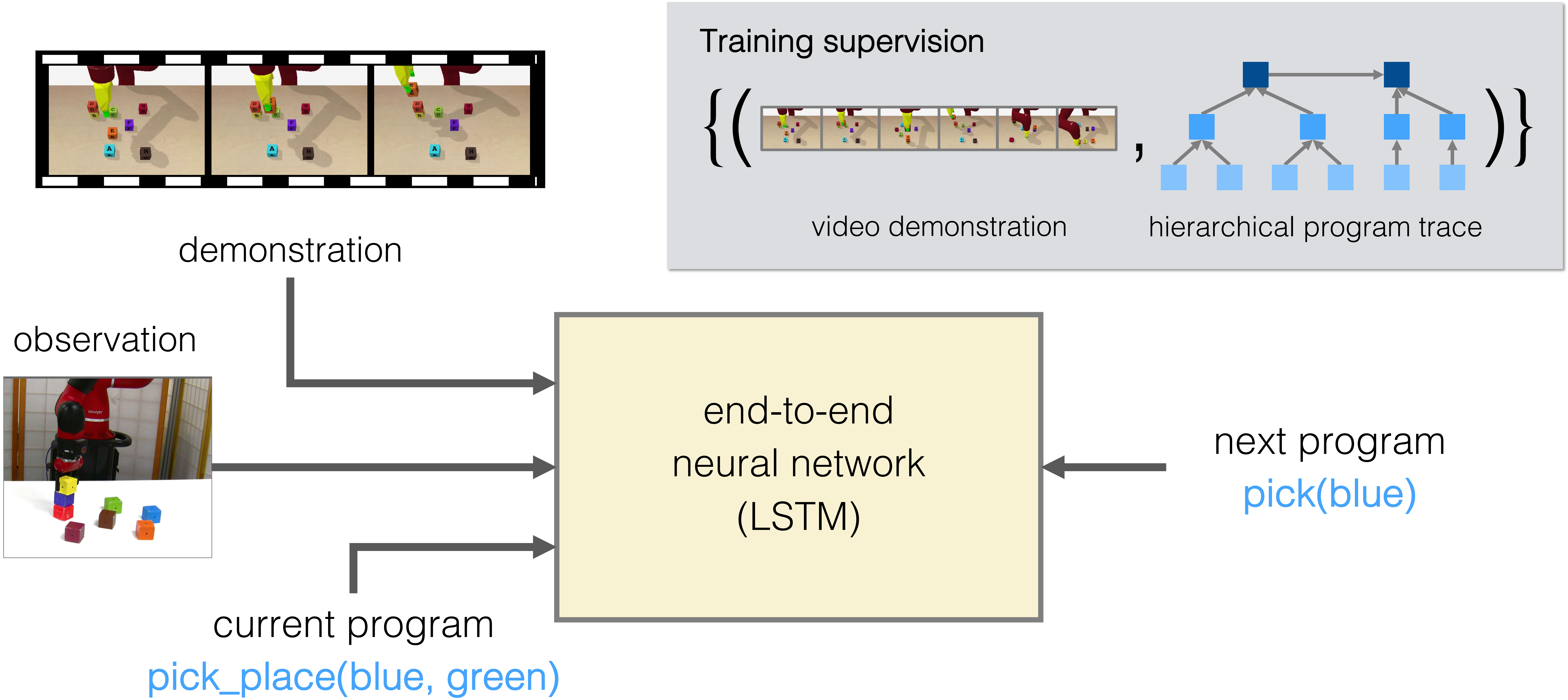
Neural Task Programming (NTP): Hierarchical Policy Learning as Neural Program Induction

One-Shot Imitation Learning from Videos: Neural Task Programming (NTP)



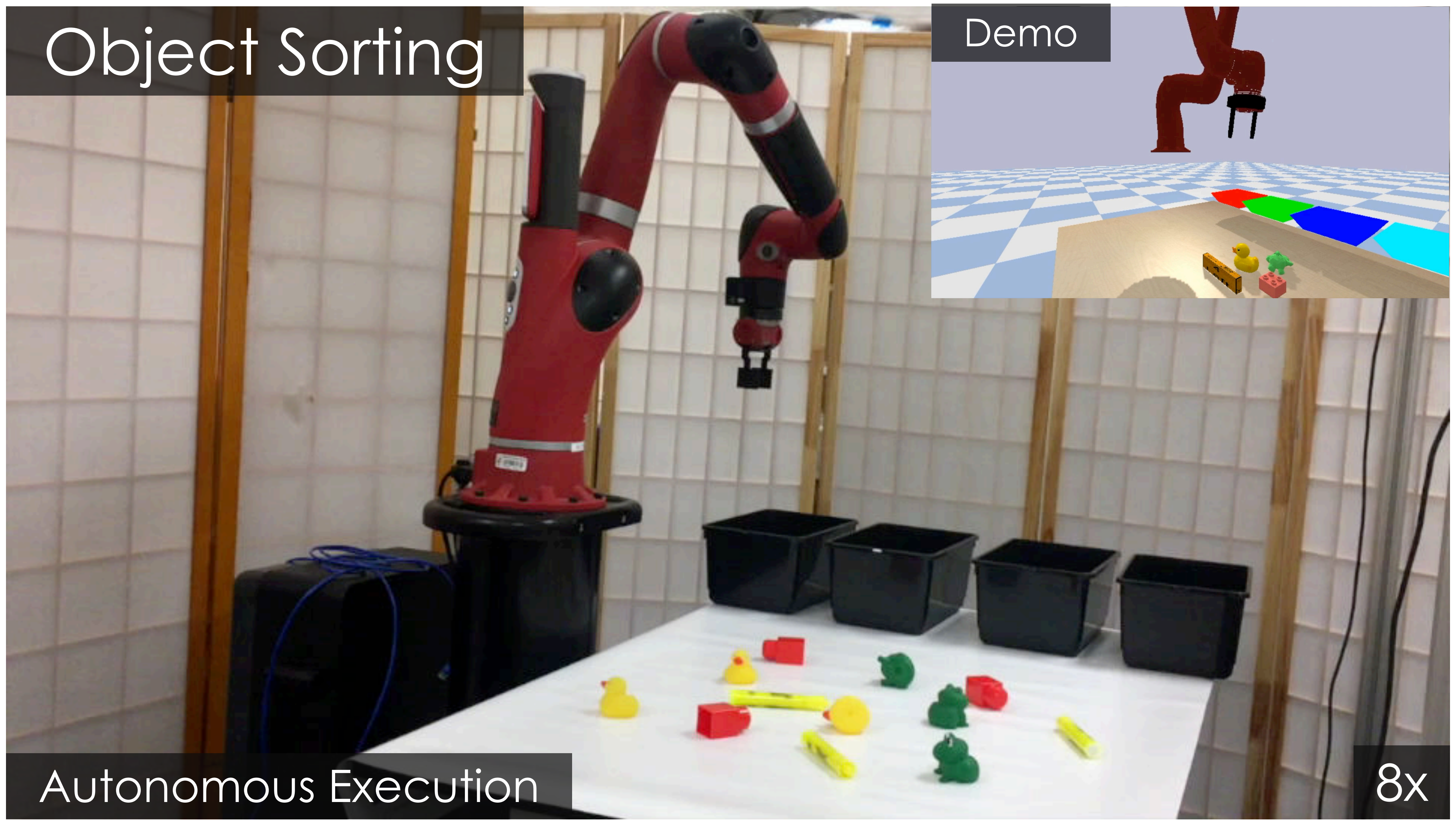
Xu*, Nair*, **Zhu**, Gao, Garg, Fei-Fei, Savarese. *ICRA 2018*

One-Shot Imitation Learning from Videos: Neural Task Programming (NTP)

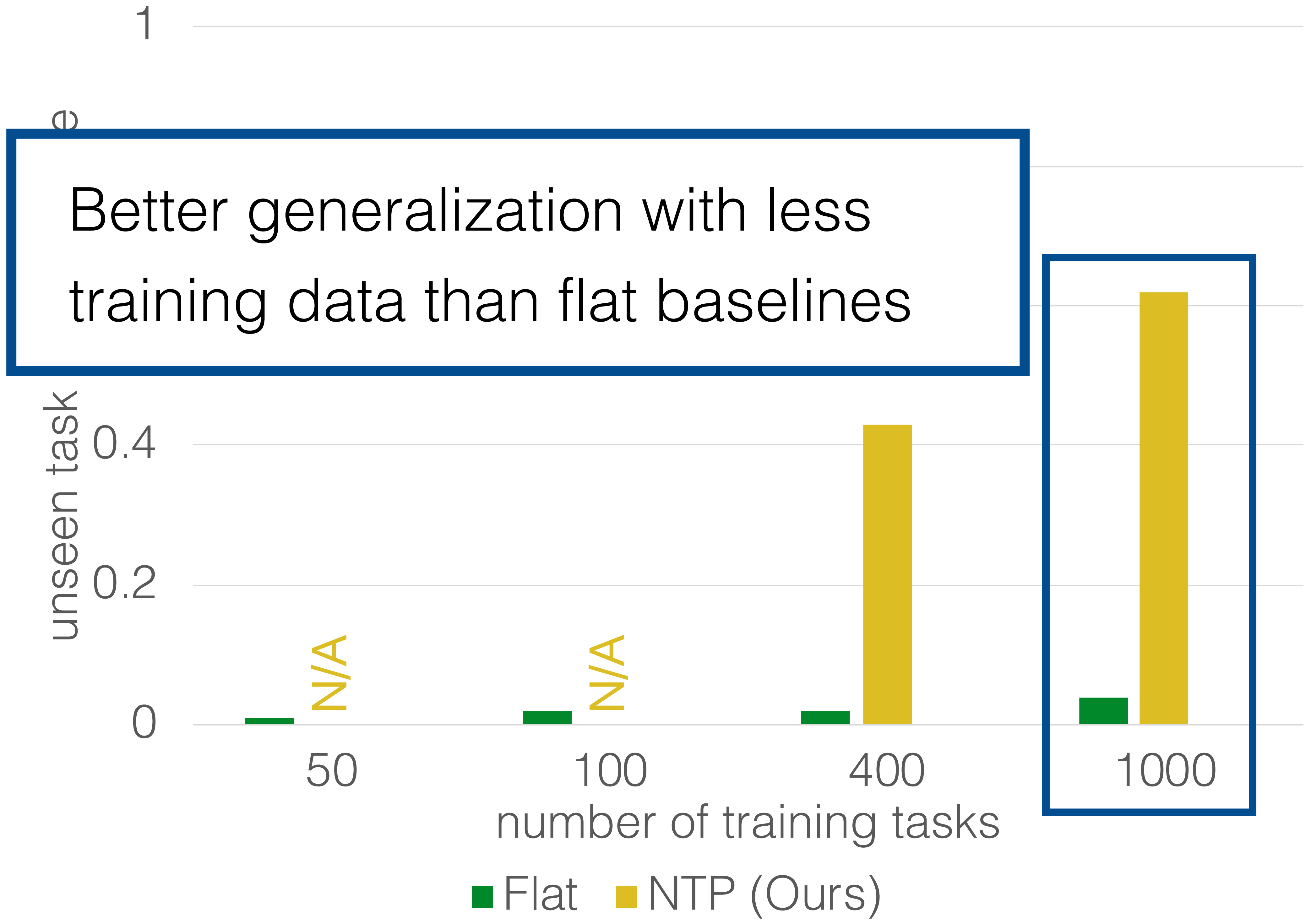


Xu*, Nair*, **Zhu**, Gao, Garg, Fei-Fei, Savarese. *ICRA 2018*

One-Shot Imitation Learning from Videos: Neural Task Programming (NTP)



Qualitative

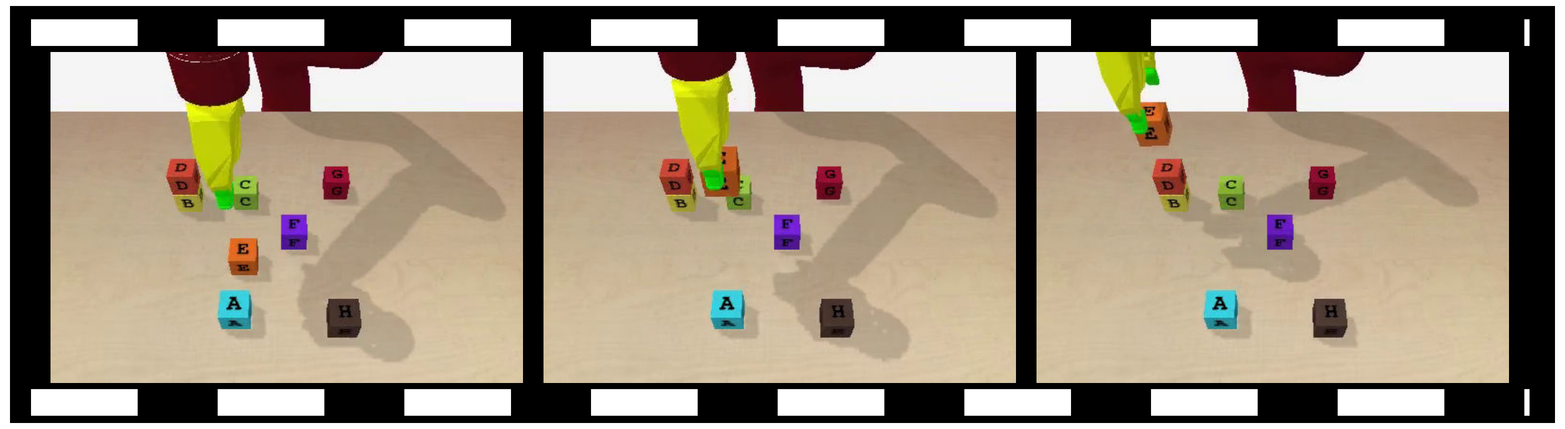


Quantitative

(the higher the better)

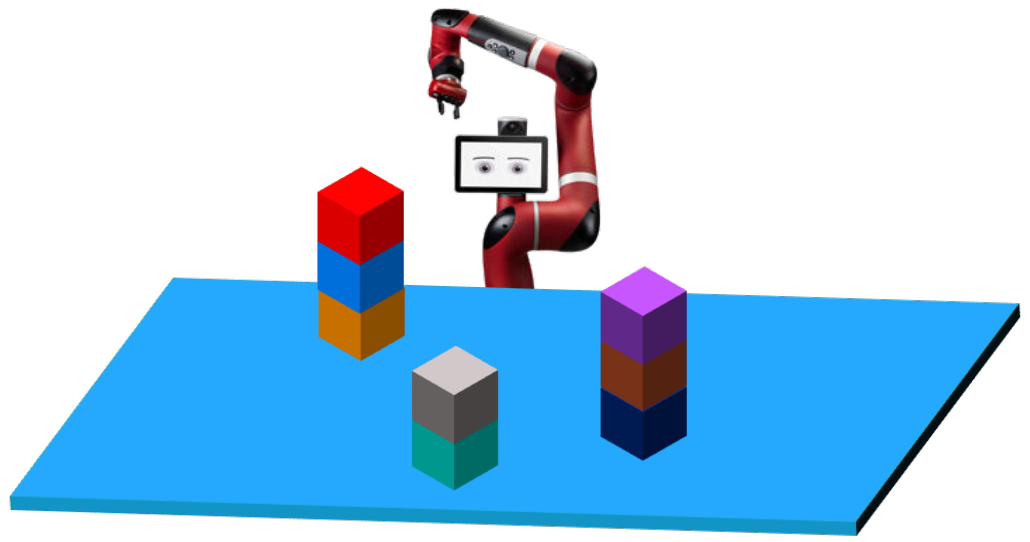
Xu*, Nair*, **Zhu**, Gao, Garg, Fei-Fei, Savarese. *ICRA 2018*

One-Shot Imitation Learning from Videos: Neural Task Programming (NTP)



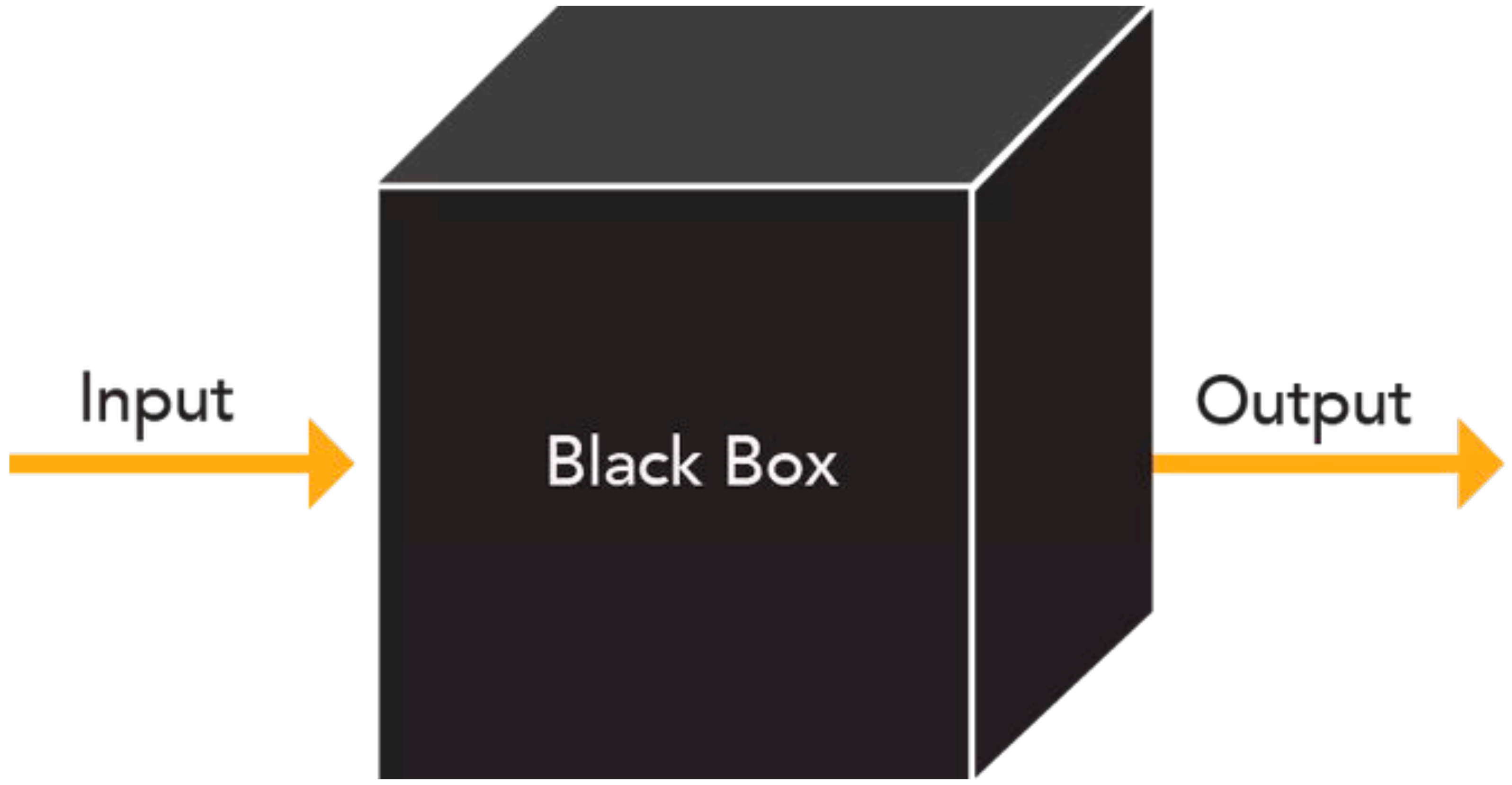
demonstration

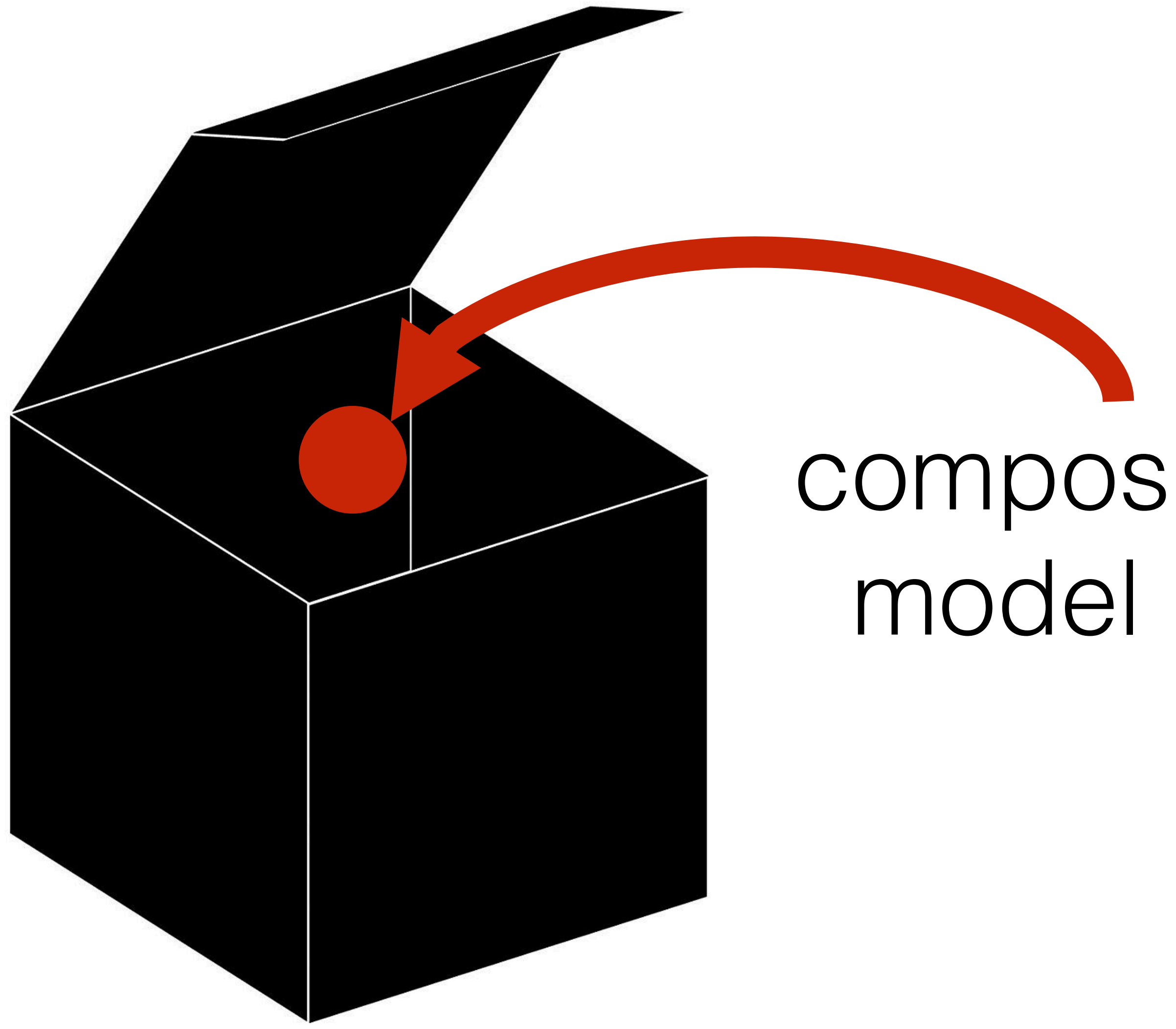
meta-learning model



policy

end-to-end neural network (LSTM)

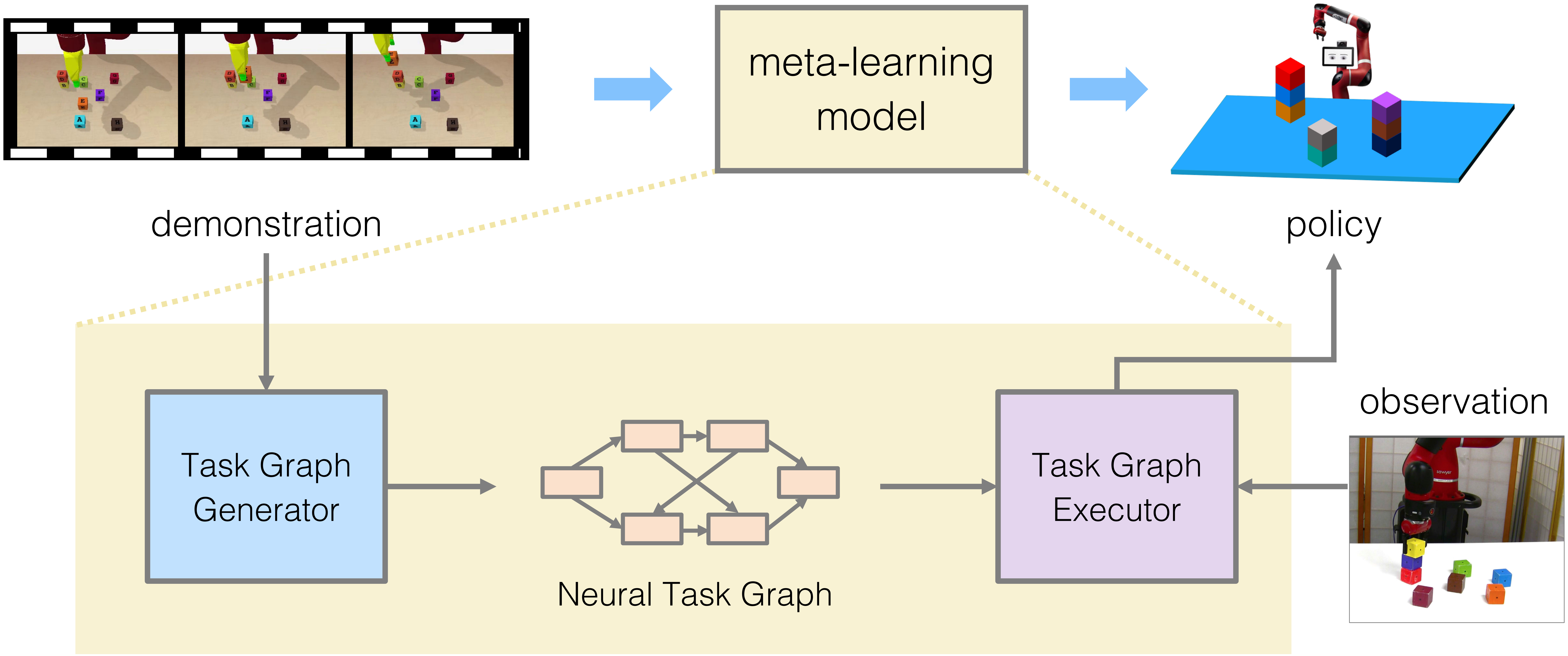




compositional model prior

Xu*, Nair*, **Zhu**, Gao, Garg, Fei-Fei, Savarese. *ICRA 2018*

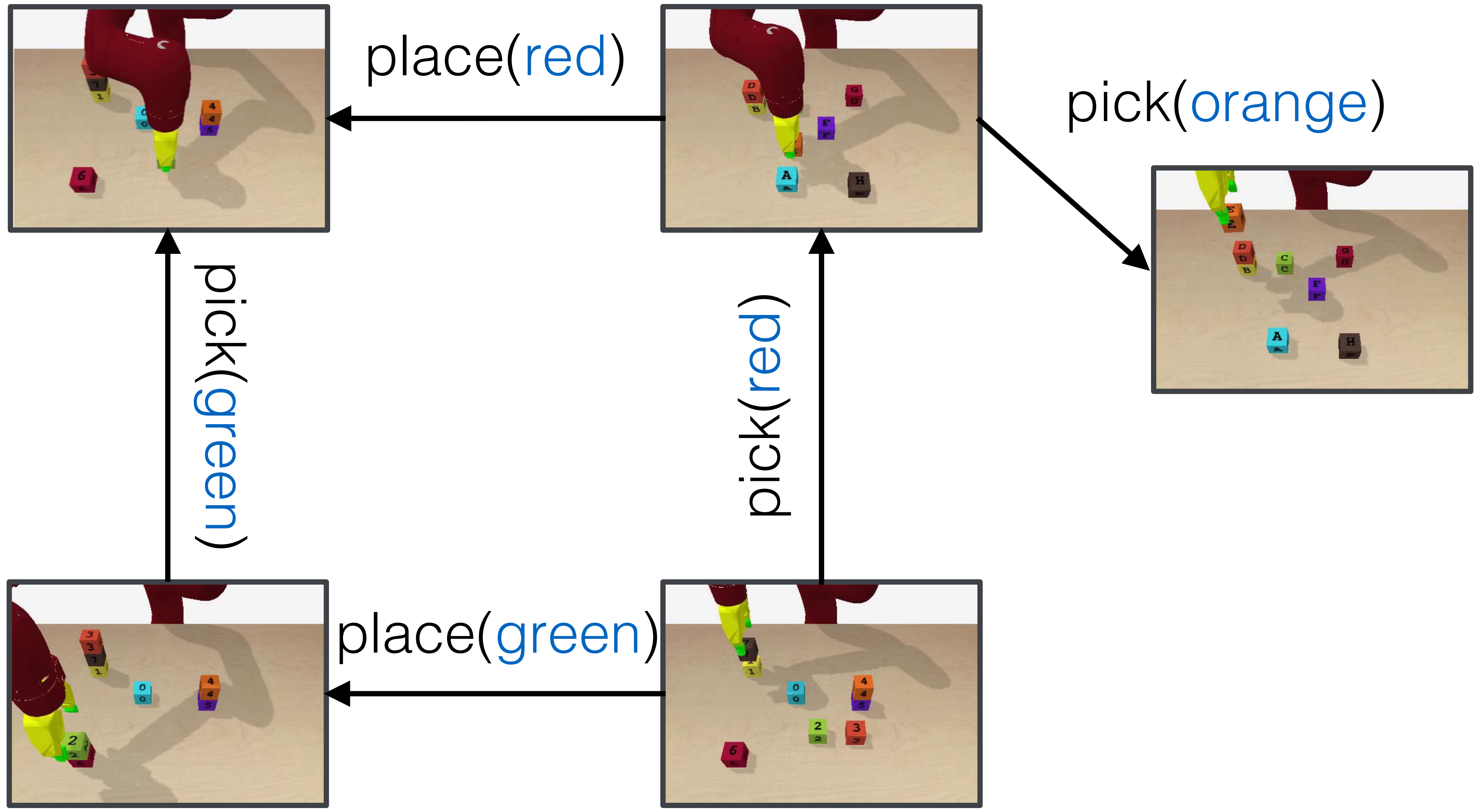
One-Shot Imitation Learning from Videos: Neural Task Graphs (NTG)



Huang*, Nair*, Xu*, **Zhu**, Garg, Fei-Fei, Savarese, Niebles. *CVPR 2019*

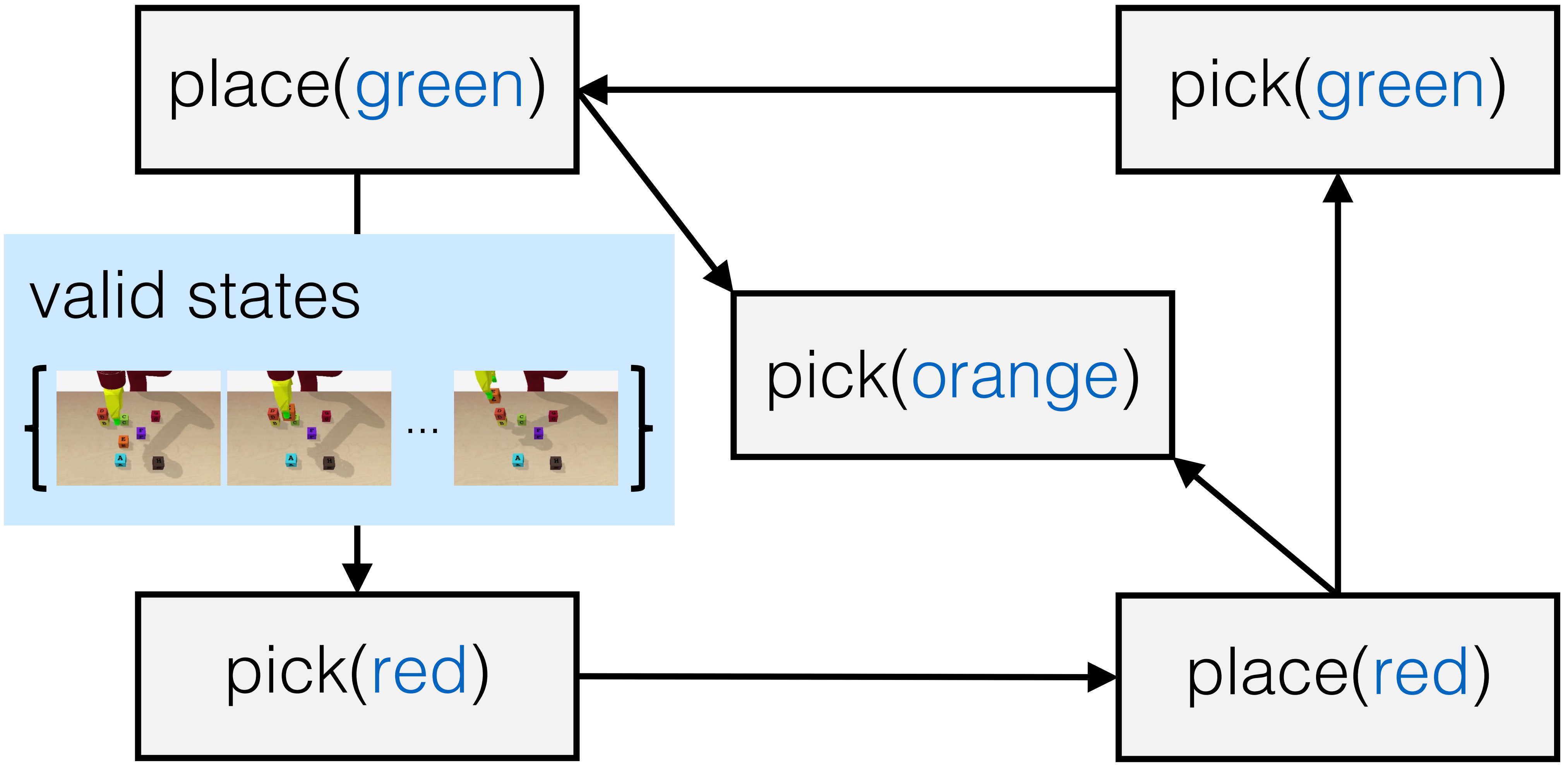
One-Shot Imitation Learning from Videos: Neural Task Graphs (NTG)

Task Graph



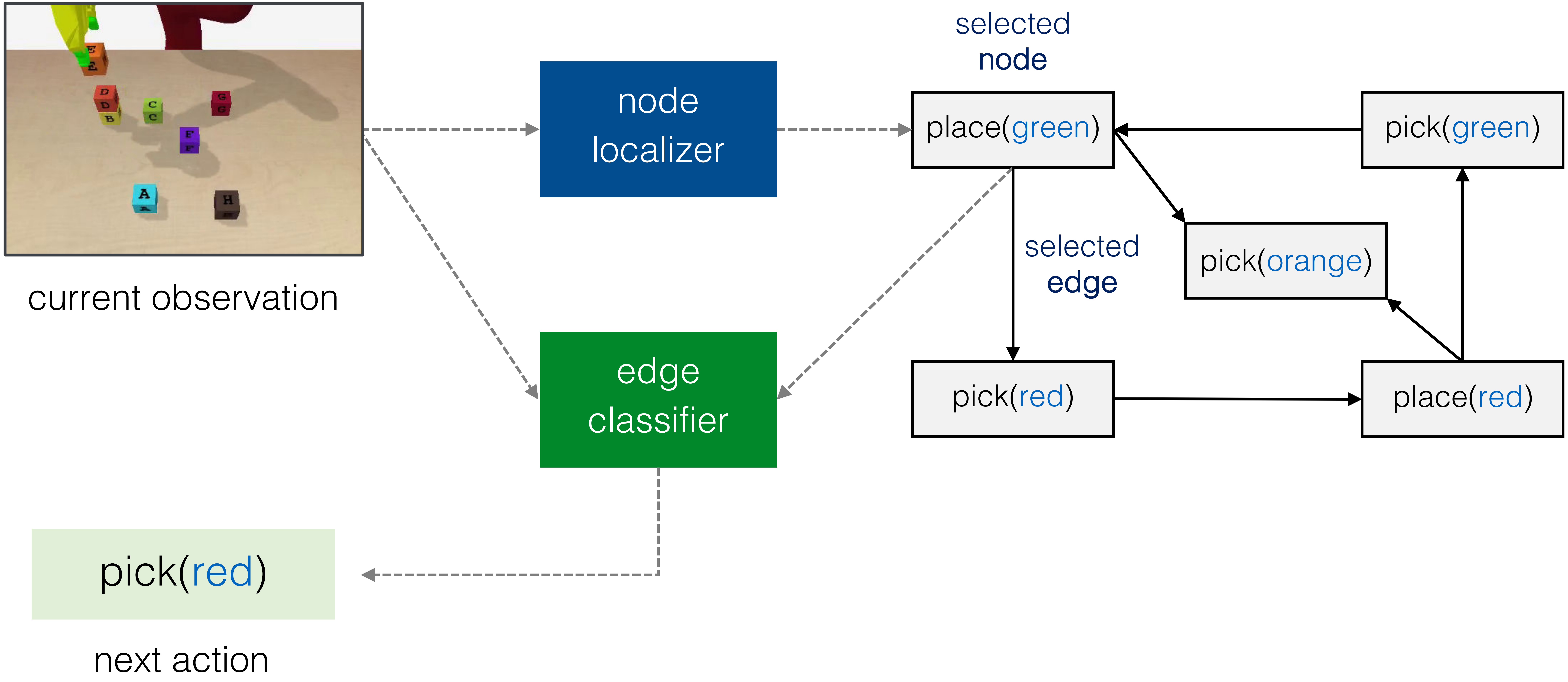
Nodes	States	infinite
Edges	Actions	

Conjugate Task Graph



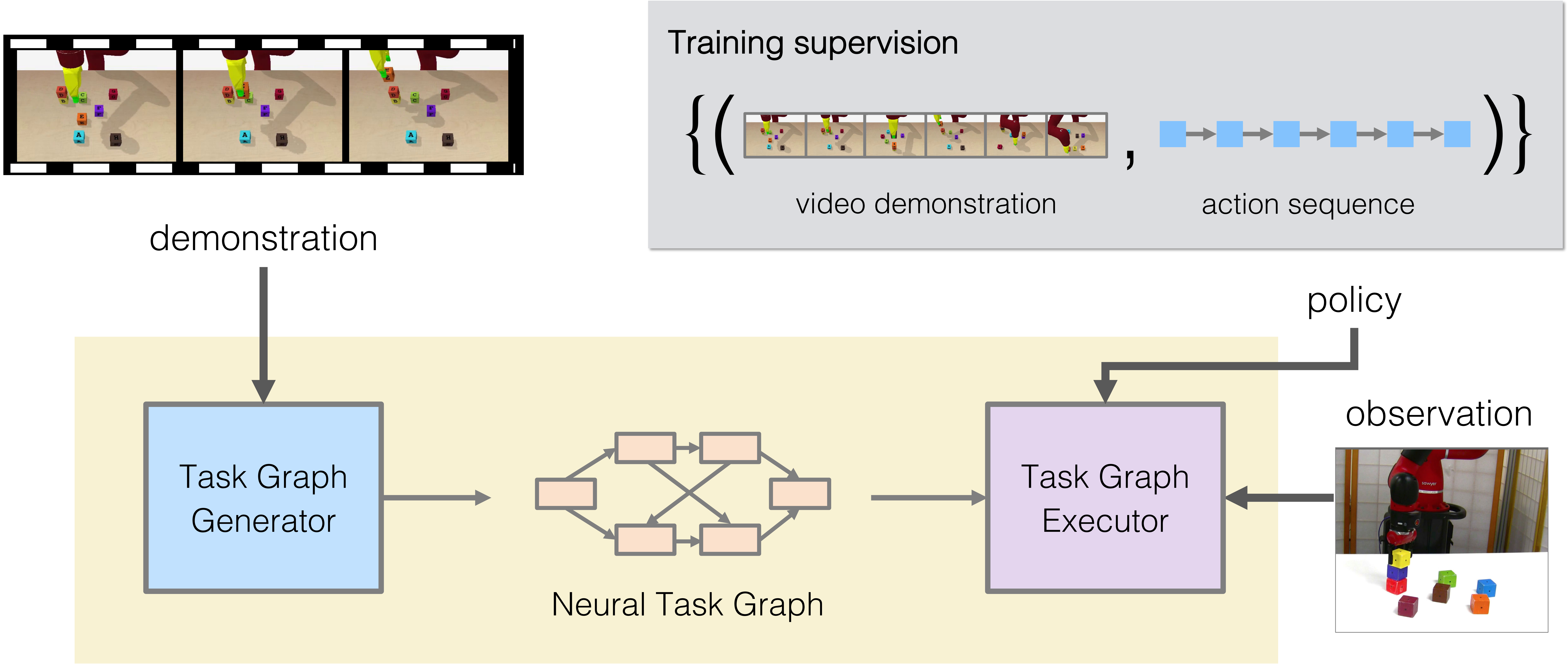
Nodes	Actions	finite
Edges	States (Preconditions)	

One-Shot Imitation Learning from Videos: Neural Task Graphs (NTG)



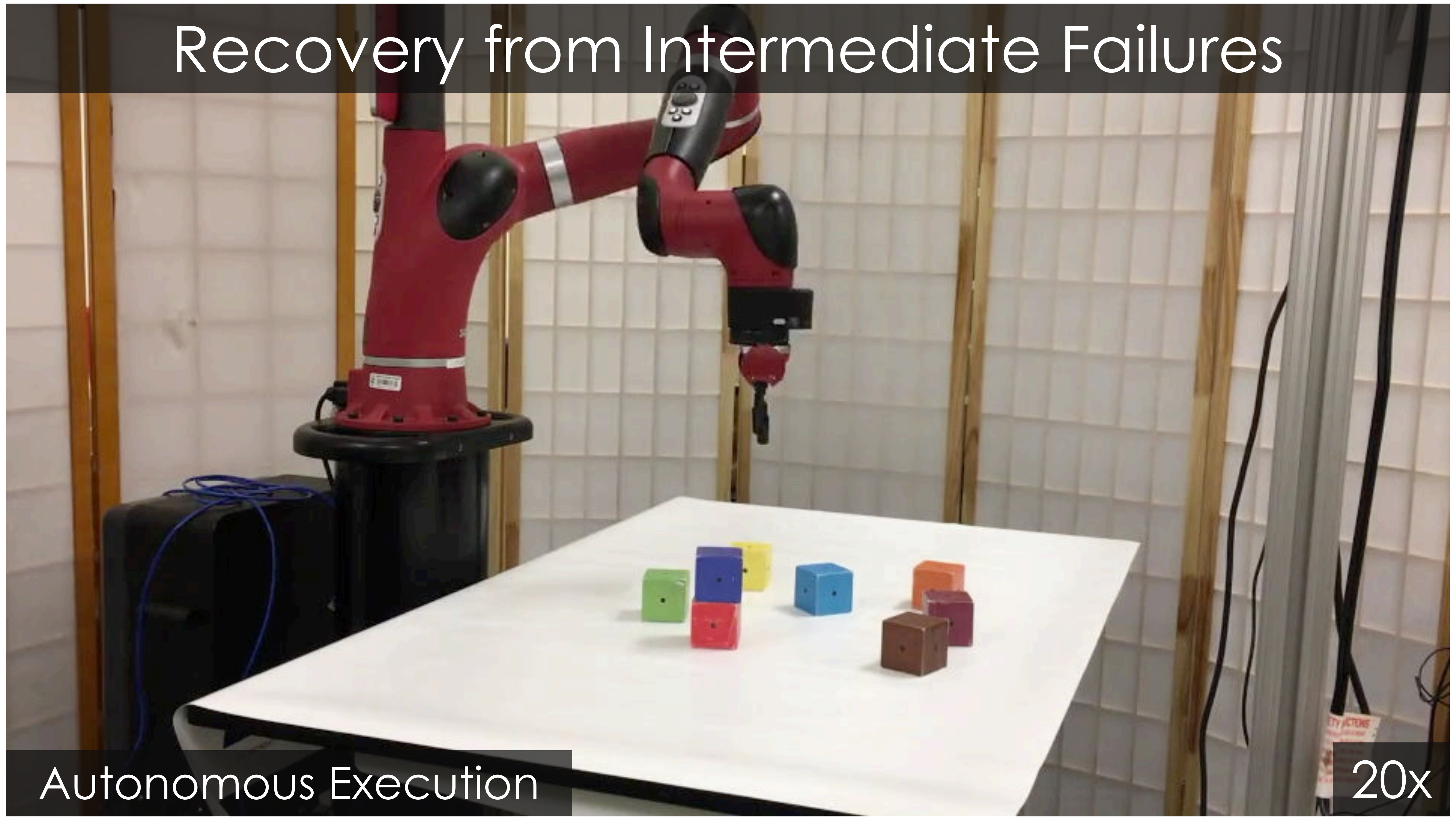
Huang*, Nair*, Xu*, **Zhu**, Garg, Fei-Fei, Savarese, Niebles. *CVPR 2019*

One-Shot Imitation Learning from Videos: Neural Task Graphs (NTG)

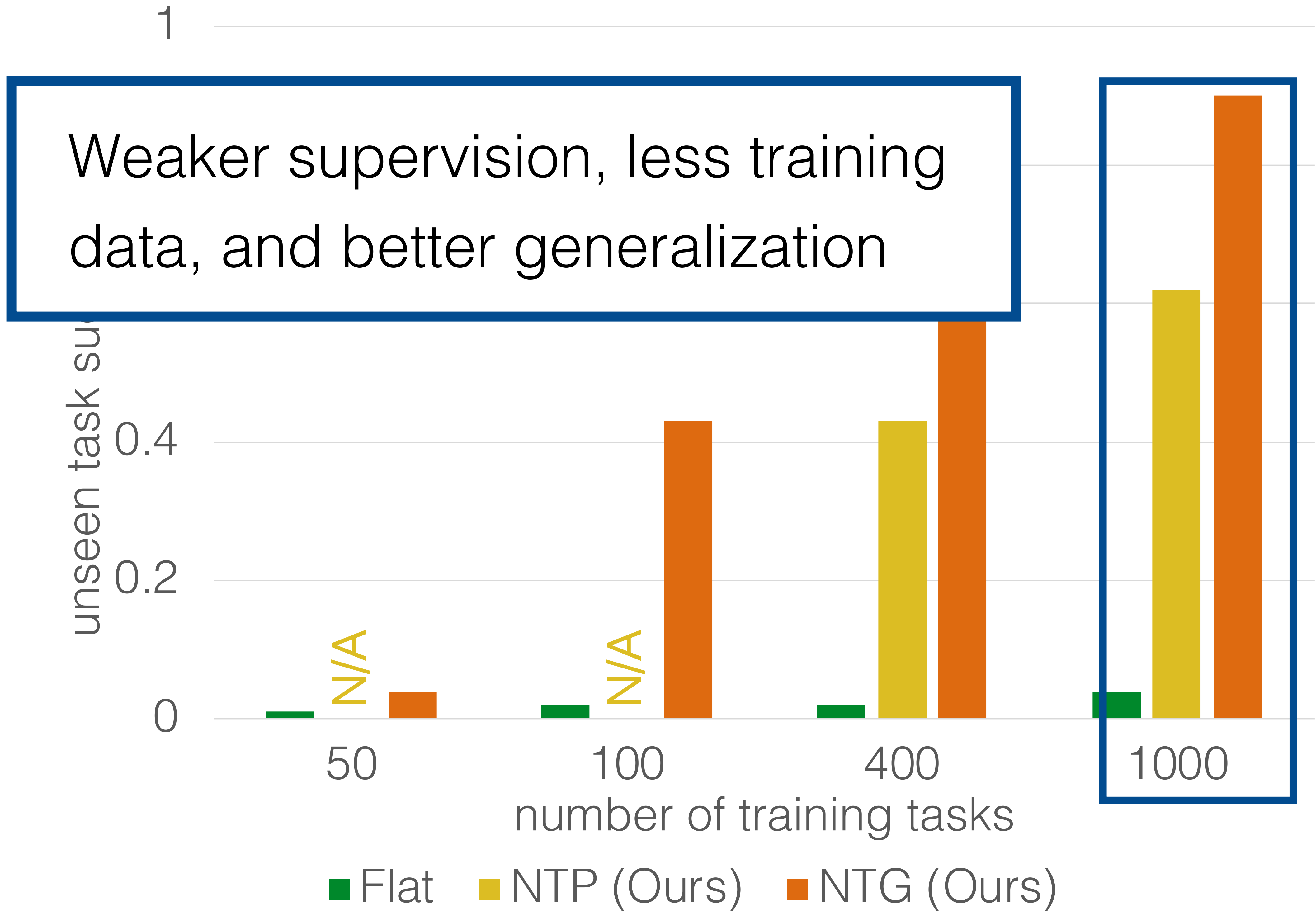


Huang*, Nair*, Xu*, **Zhu**, Garg, Fei-Fei, Savarese, Niebles. *CVPR 2019*

One-Shot Imitation Learning from Videos: Neural Task Graphs (NTG)



Qualitative

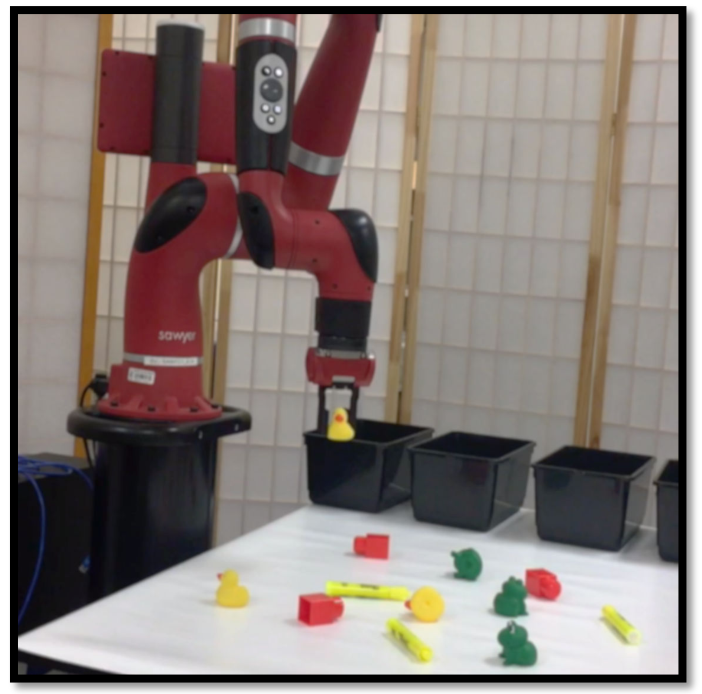


Quantitative

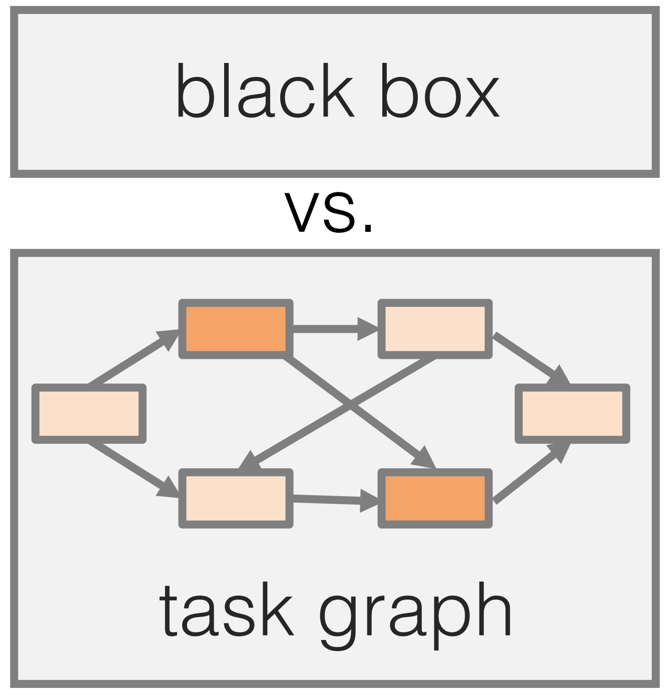
(the higher the better)

Huang*, Nair*, Xu*, **Zhu**, Garg, Fei-Fei, Savarese, Niebles. *CVPR 2019*

Summary - Part III

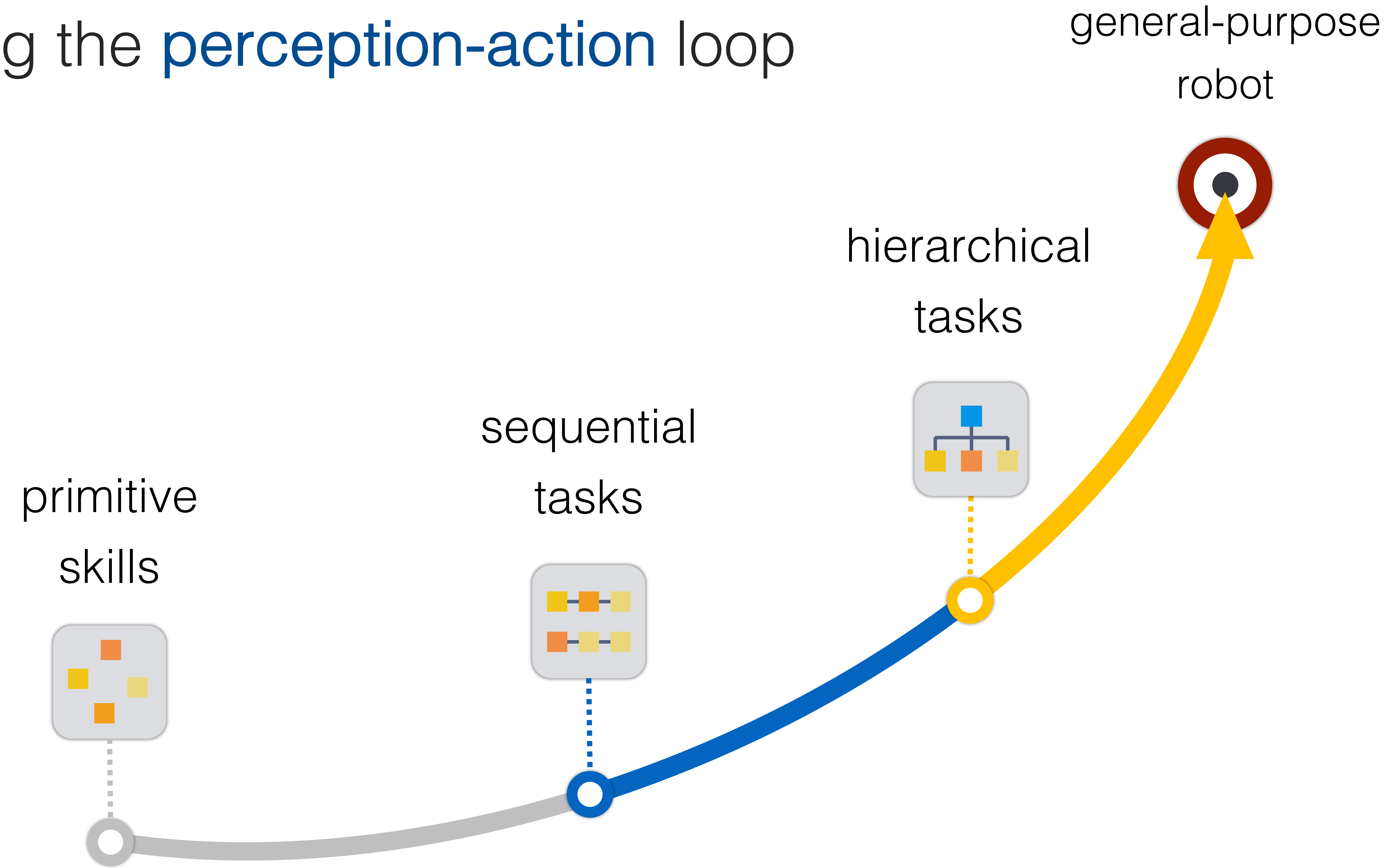


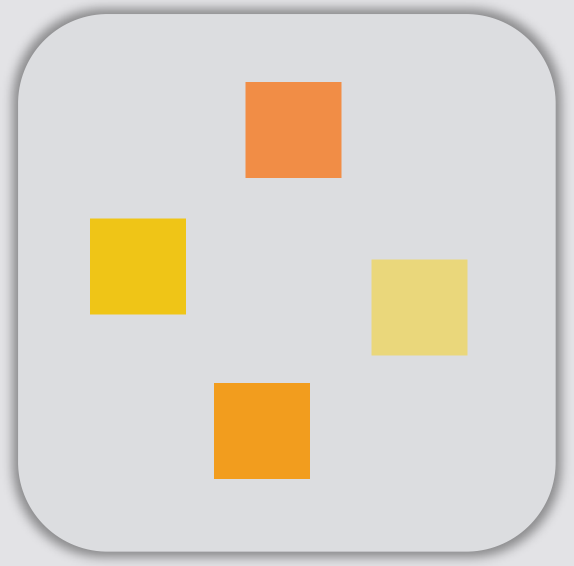
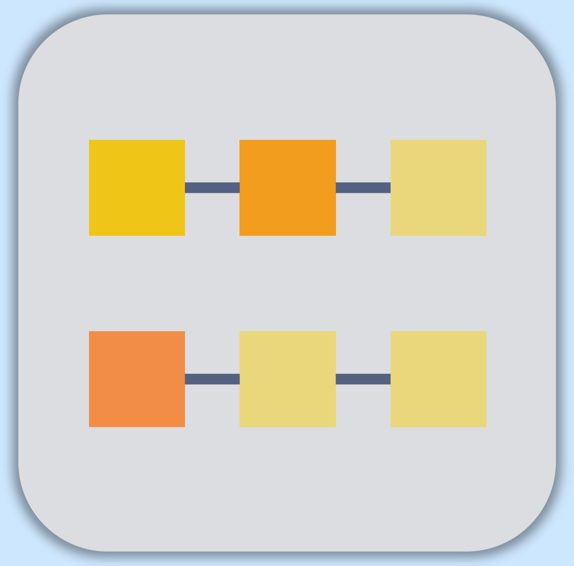
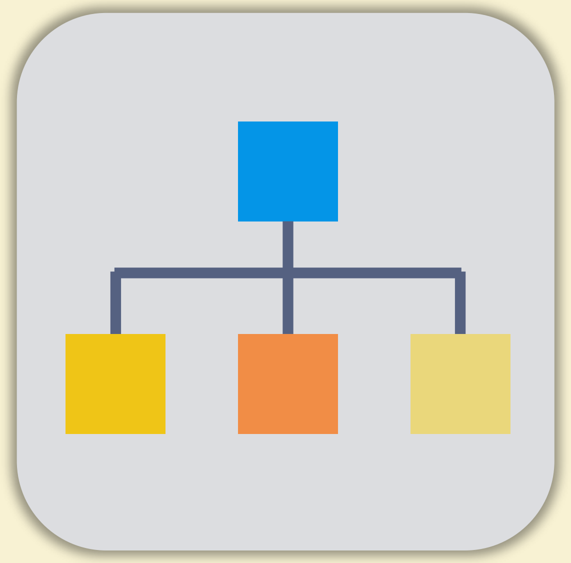
Extracting knowledge about the compositional structure of **hierarchical tasks** from **video demonstrations**



Meta-learning models with compositional priors generalize better than black-box models

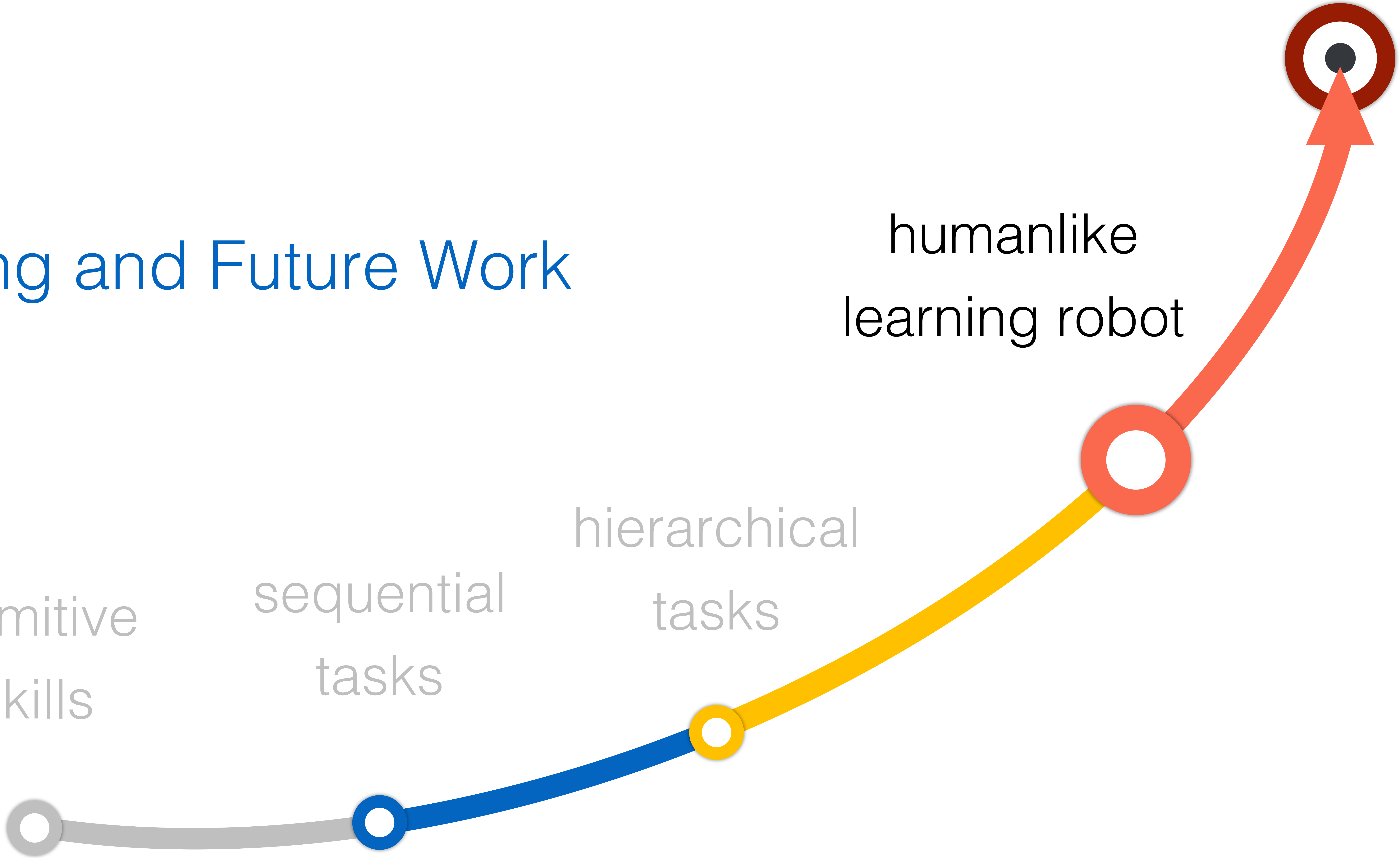
closing the perception-action loop



Closing the Perception-Action Loop	Perception Modality	Action Abstraction	Learning Method
 primitive skills [RSS'18, CoRL'18a, CoRL'18b]	real-world sensory data	joint torque	reinforcement & imitation learning
 sequential tasks [ICRA'17, ICCV'17]	interactive visual environment	high-level command	transfer learning
 hierarchical tasks [ICRA'18, CVPR'19]	unstructured video data	task structure	meta-learning

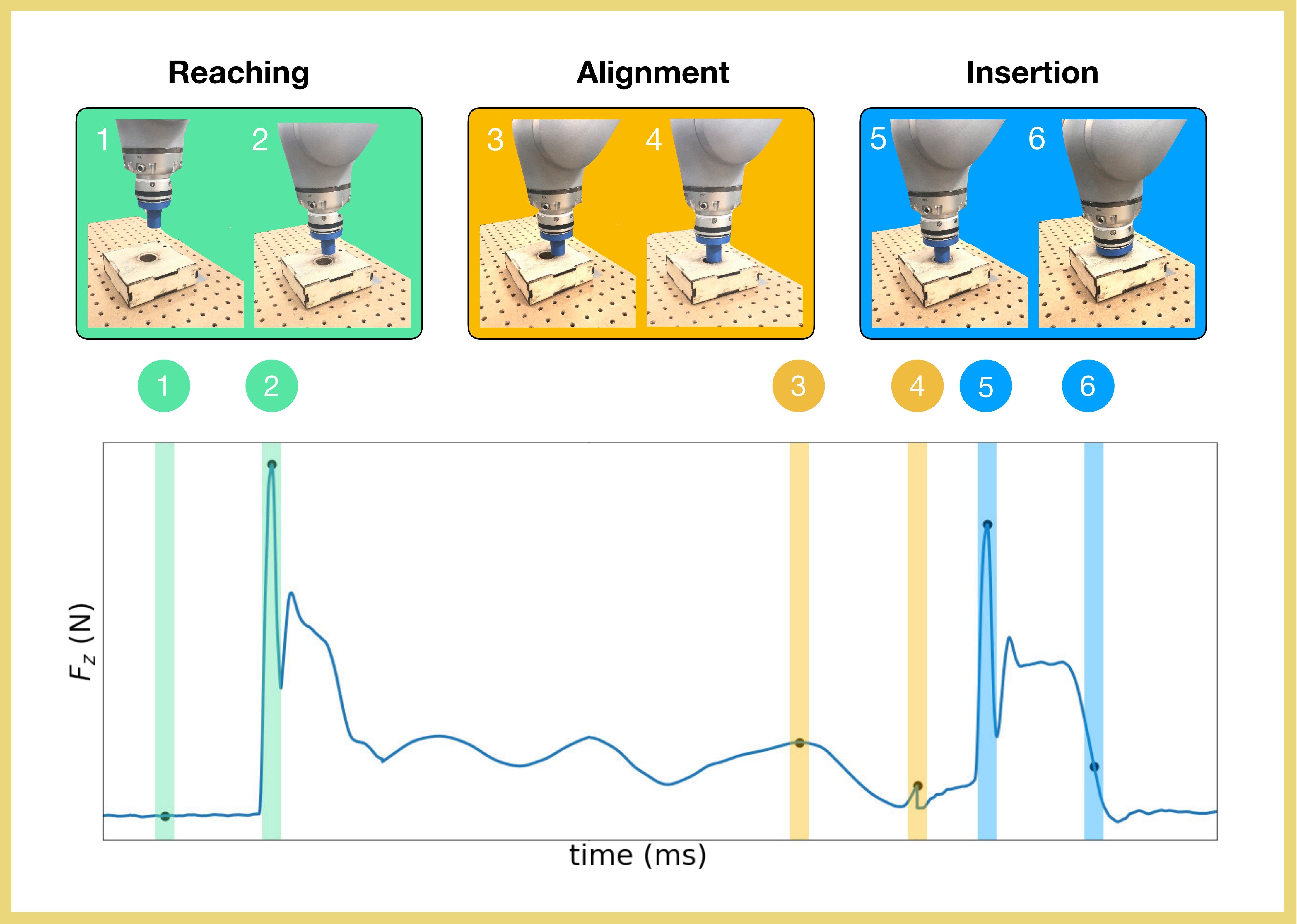
closing the perception-action loop

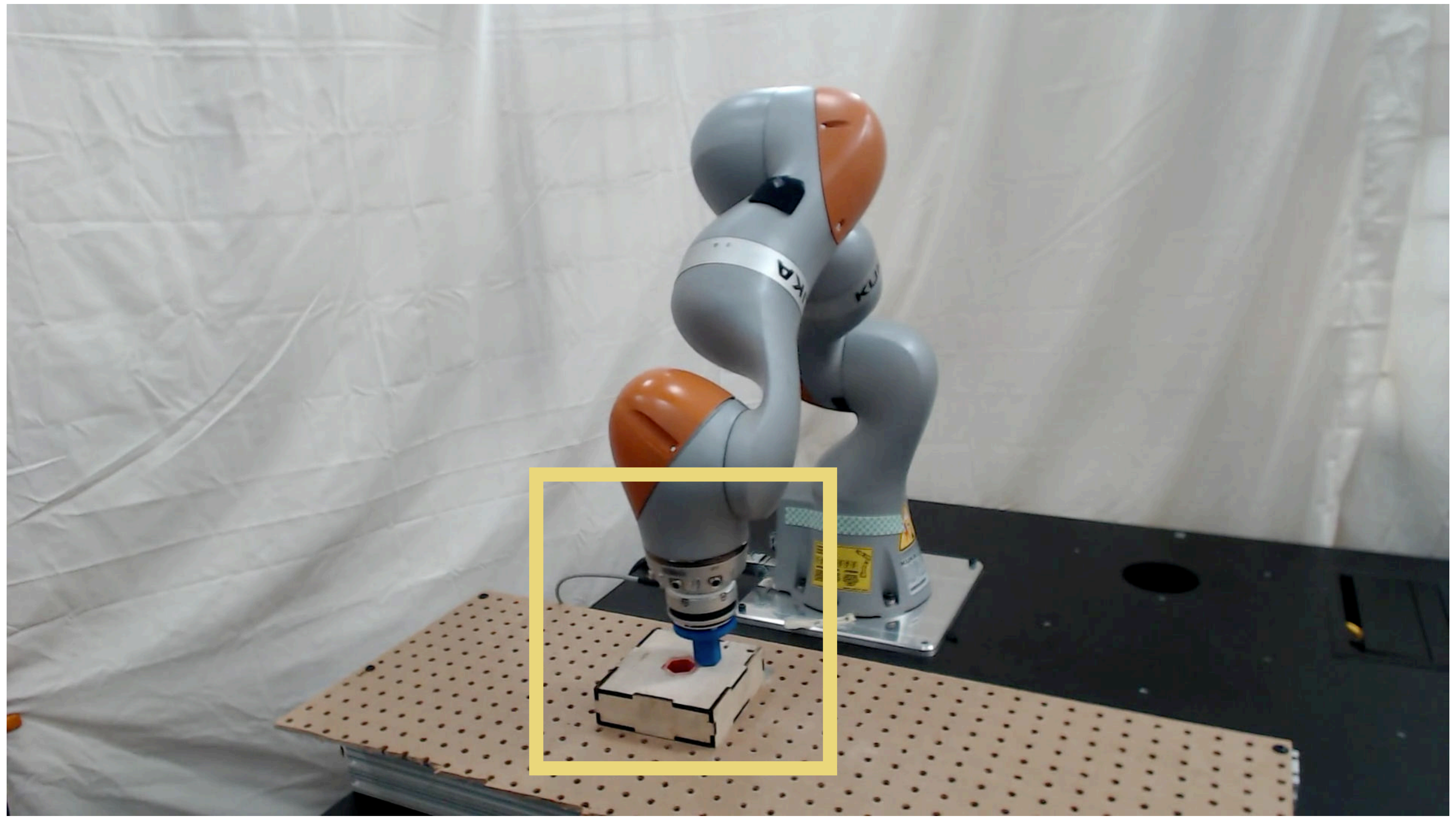
Ongoing and Future Work



Future Direction: Multimodal Perception Beyond Vision

Learning coherent representations of multimodal information for control





combining vision and force for manipulation

Lee*, Zhu*, Srinivasan, Shah, Savarese, Fei-Fei, Garg, Bohg. *ICRA 2019* (Best Conference Paper)

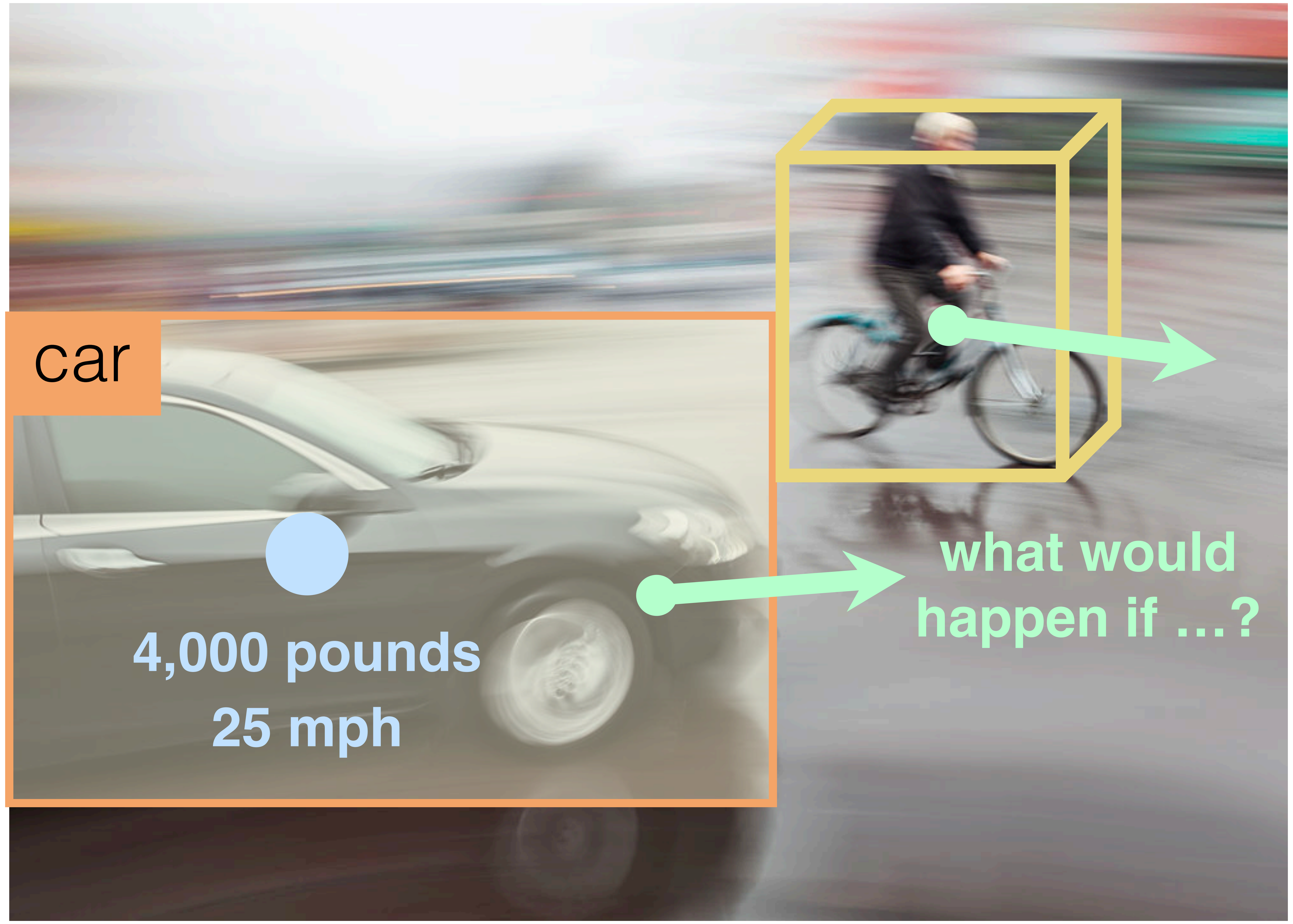
Future Direction: Learning Knowledge of the World from Interaction

semantic

[**Zhu** et al. CVPR'16; Krishna, **Zhu** et al. IJCV'17; Xu, **Zhu** et al. CVPR'17; **Zhu** et al. CVPR'17]

physical

[**Zhu** et al. ECCV'14; Fang, **Zhu** et al. RSS'18]



"To accelerate or to brake?"

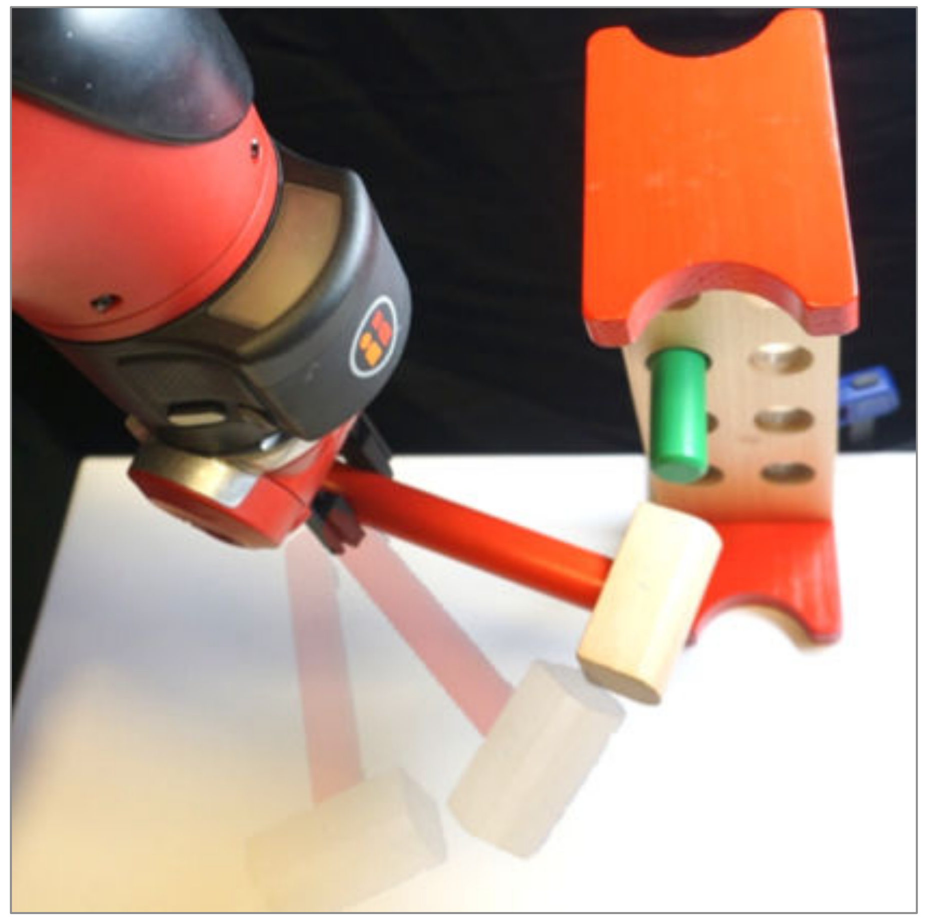
geometric

[Chen, Xu, **Zhu** et al. CVPR'19]

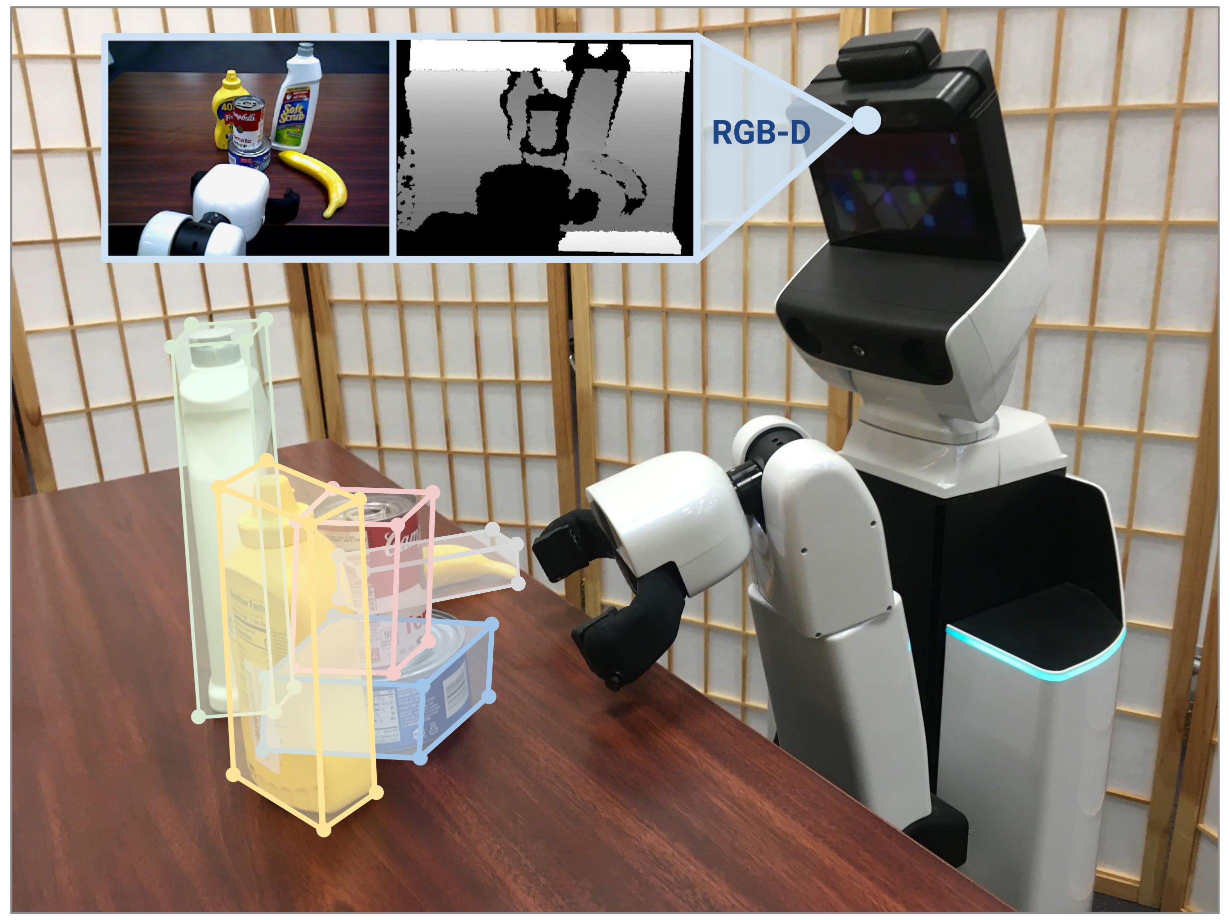
causal

[ongoing work]

Future Direction: Integrating Perception and Knowledge for Autonomy



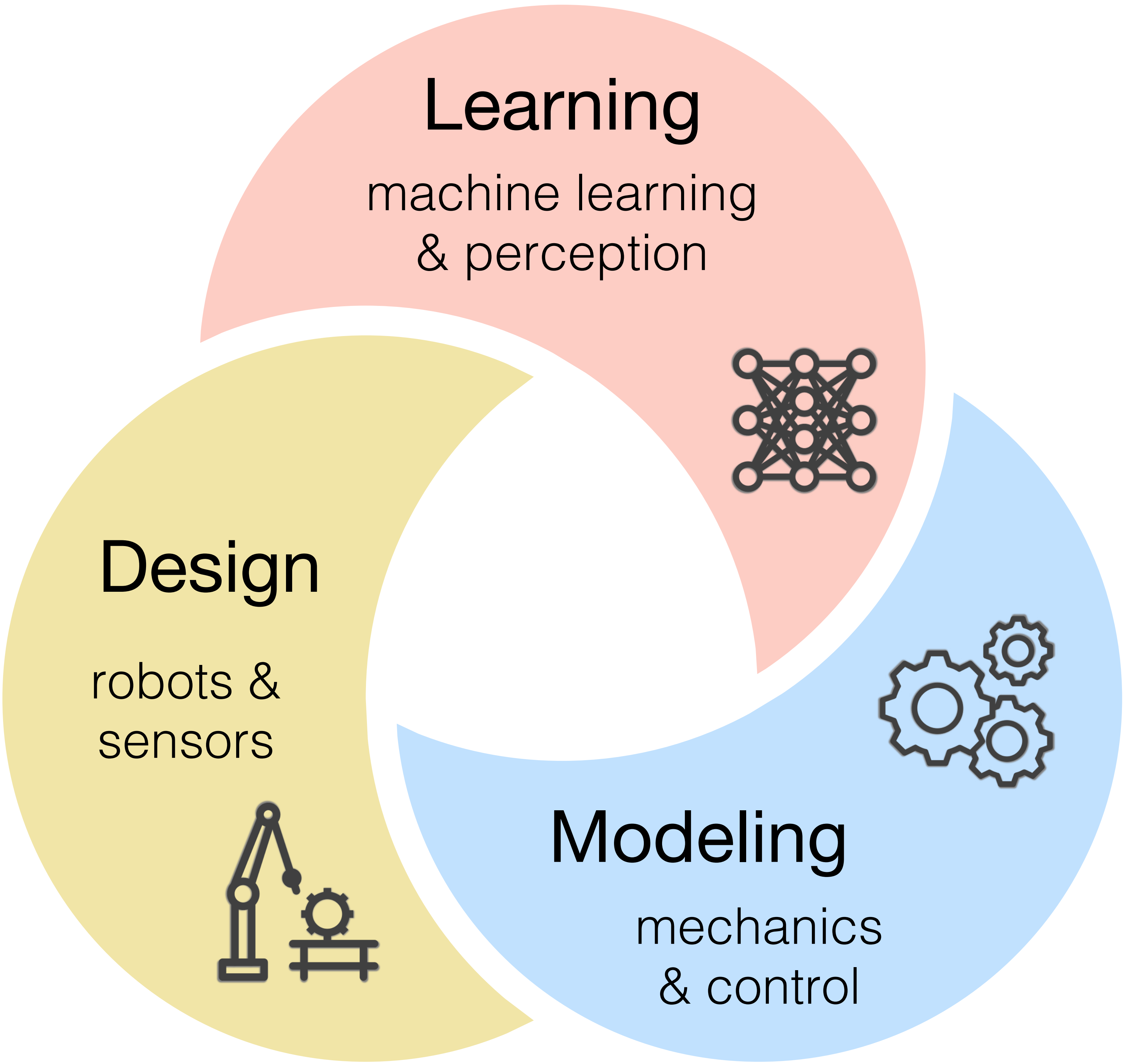
[Fang, **Zhu** et al. RSS’18]



[Wang, Xu, **Zhu** et al. CVPR’19]

Data-driven + Model-driven Methods

ongoing work



broader collaboration

Acknowledgements

Fei-Fei Li

Silvio Savarese

Jeannette Bohg

Nando de Freitas

Ali Farhadi

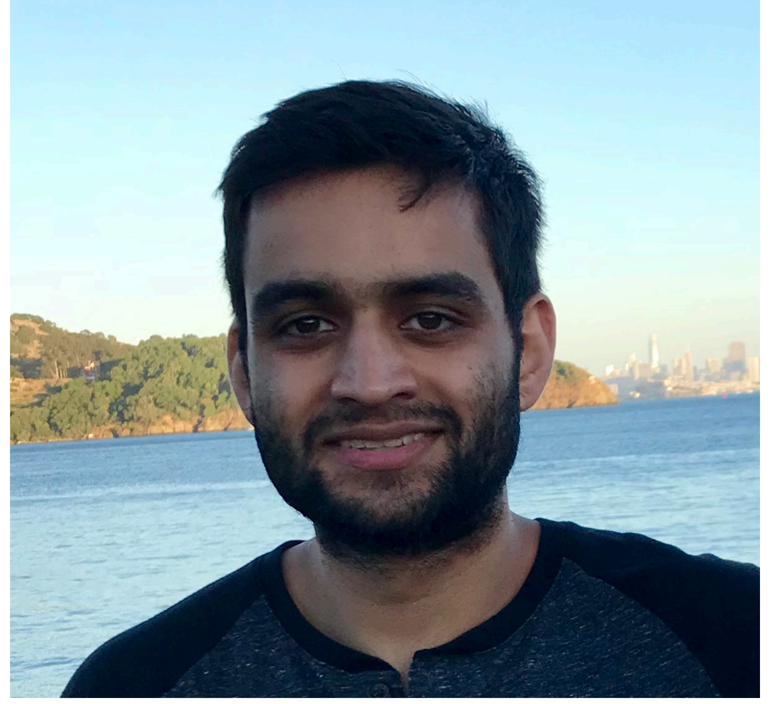
Abhinav Gupta

Animesh Garg

Joseph Lim

Raia Hadsell

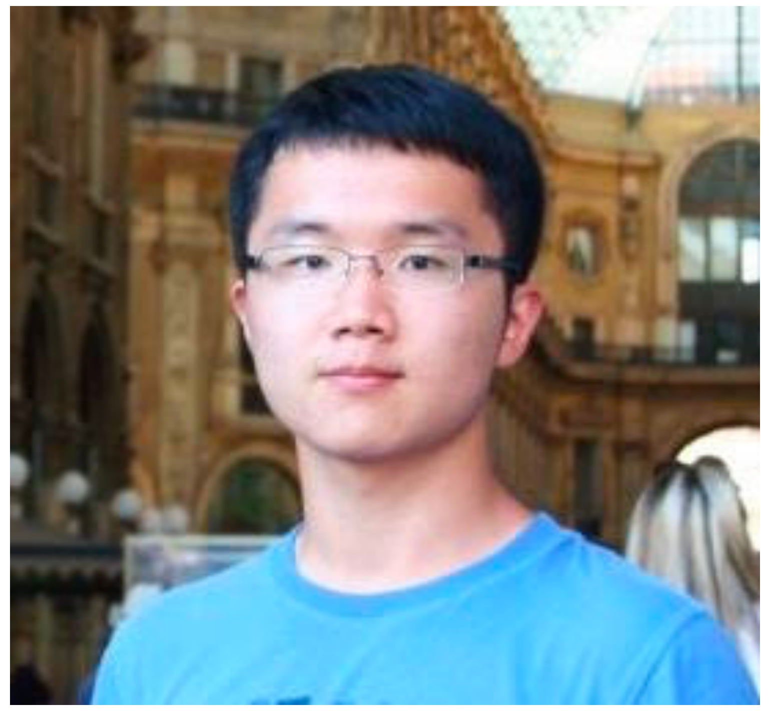
Danfei Xu



Ajay Mandlekar

De-An Huang

Michelle Lee



Kuan Fang

Suraj Nair

Jim Fan

TOYOTA RESEARCH INSTITUTE

DEPARTMENT OF THE NAVY ONR Science & Technology

Google

JD.COM