Sample efficient DRL for embodied AI

Leaders and Facilitator

Vidhi Jain
MS student at CMU
Co-leader

Simin Liu
PhD student at CMU
Co-leader

Ganesh Iyer
Applied Scientist at Amazon Lab126
Facilitator

2-minute breakout room introductions

- Name
- Position
- What do you want from this session?

Session format

Presentation: ~30 min

Intro + 4 topics

- Post questions to chat!
- 1-2 clarifying questions after each topic

Discussion: ~30 min

2 sets of questions

- Discuss in breakout rooms, reconvene to share

Why Embodied AI?

Enhancing Intelligence

Source: Francis Vachon, Time laps of Charles-Edward, 9 month old son

Learning in real environments through explorative physical interaction

Why Embodied AI?

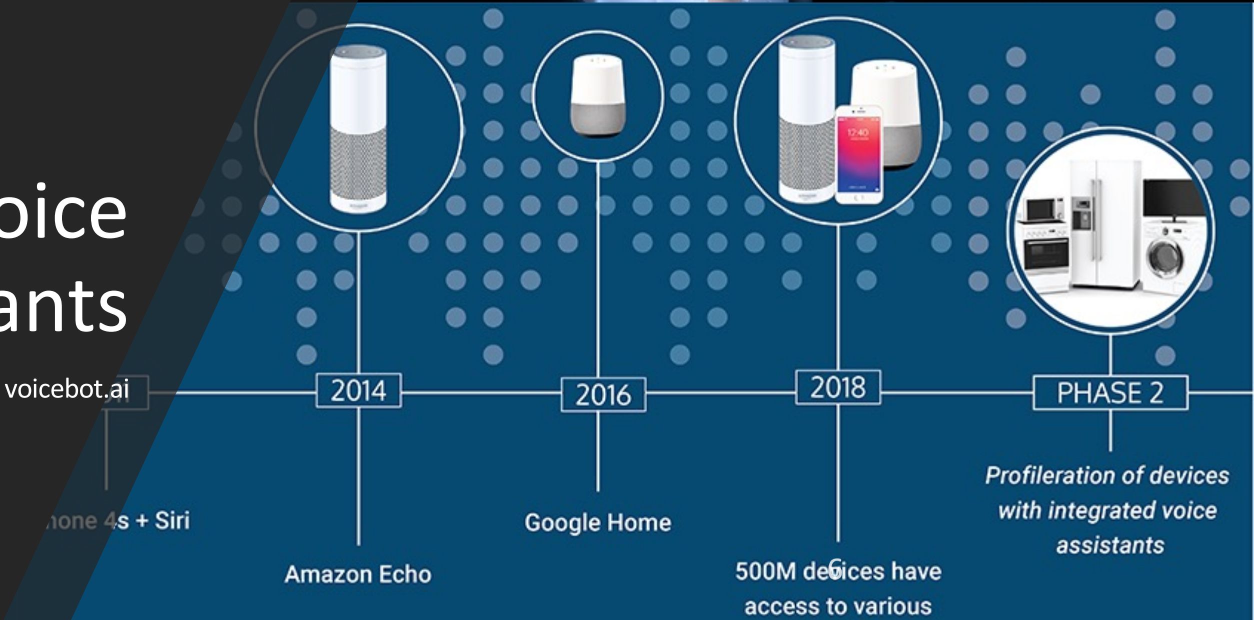
AI tools

Self-driving cars

Source: roboticsbusinessreview

Voice assistants

Source: voicebot.ai



DRL has had great success in simulation!

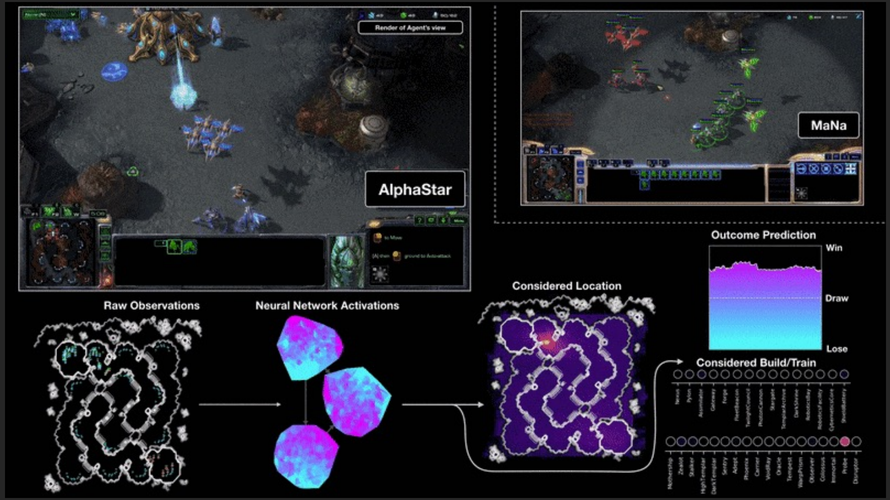
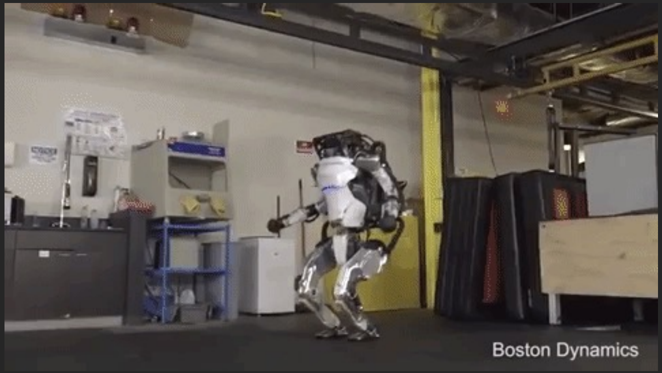


Image credits: Business Insider, DeepMind

But it’s been much harder applying it to real world platforms…

iRobot Roomba

Waymo AV



Boston Dynamics Atlas

What’s hindering us?

One main reason is **data efficiency**:

Not a big issue in simulation

Big issue for real platforms!

Two perspectives for solutions:

1. Improvise algorithmically
2. Scale up data collection

Session goals

Share opinions on the comparative merit of each method/perspective

Identify the “gaps” in the current research

Outline

- Careful choice of paradigm
- Using knowledge from other domains
- Human demonstrations and feedback
- Scaling data collection

Model based or model free?

MB: learn an explicit model of the transition function $p(s_{t+1}|s_t, a_t)$

MF: learn value function (i.e. $V(s)$, $Q(s,a)$, $A(s,a)$) or directly learn a policy

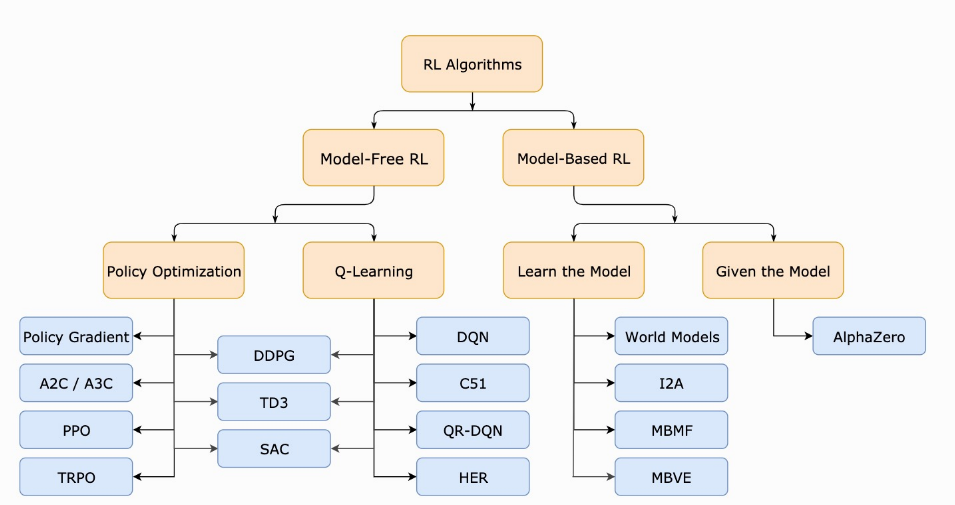


Image from https://spinningup.openai.com/en/latest/spinningup/rl_intro2.html

Model based or model free?

MB is more sample efficient…but there’s a caveat: poor asymptotic performance.

Examples of MBRL in the real world

Self-driving

Wayve used world models with BPTT (backprop through time)
https://wayve.ai/blog/dreaming-about-driving-imagination-rl

Millirobot path following

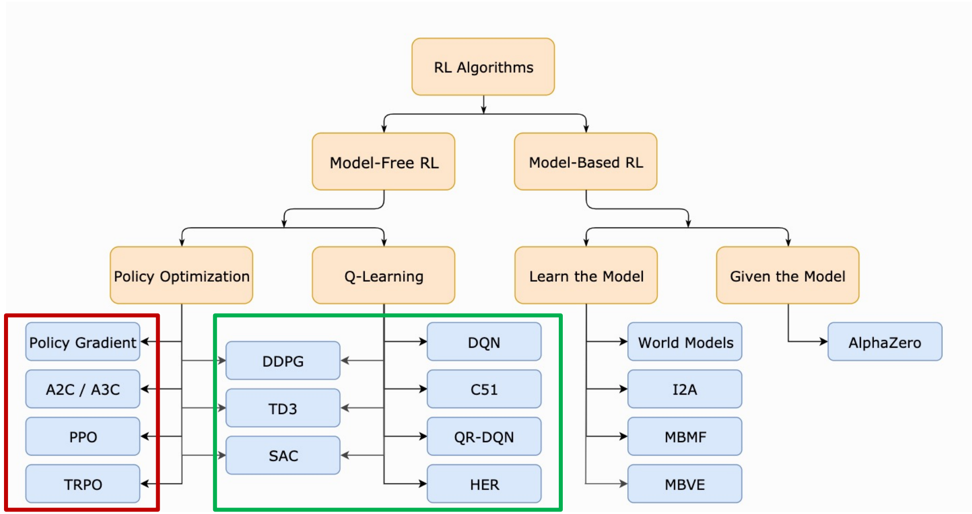
[Learning Image-Conditioned Dynamics Models for Control of Under-actuated Legged Millirobots](): **Anusha Nagabandi**, Guangzhao Yang, Thomas Asmar, Ravi Pandya, Gregory Kahn, Sergey Levine, Ronald S. Fearing

Off-policy or on-policy? (MF)

Off-policy: can use samples generated by any policy.
I.e. Q-learning

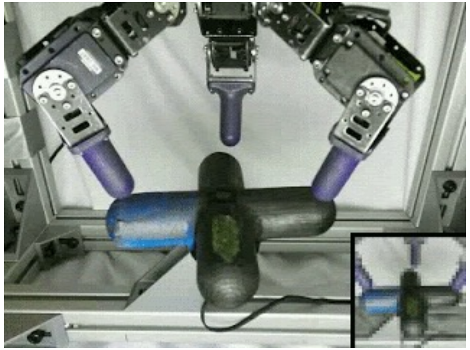
On-policy: can only use samples generated by current policy.
I.e. policy gradient

> Off-policy is more sample efficient



Examples of off-policy in the real world

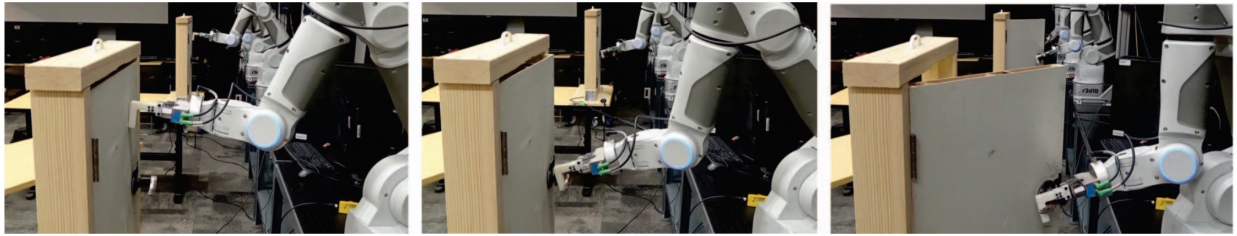
Soft Actor-Critic



Dexterous manipulation
(goal is put blue knob on the right);

Minitaur walking robot

Asynchronous
Q-learning



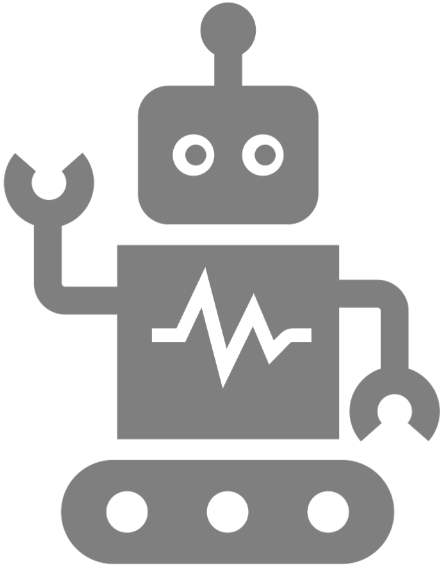
[Deep Reinforcement Learning for Robotic Manipulation with Asynchronous Off-Policy Updates]
ShiXiang Gu and Ethan Holly and Timothy Lillicrap and Sergey Levine

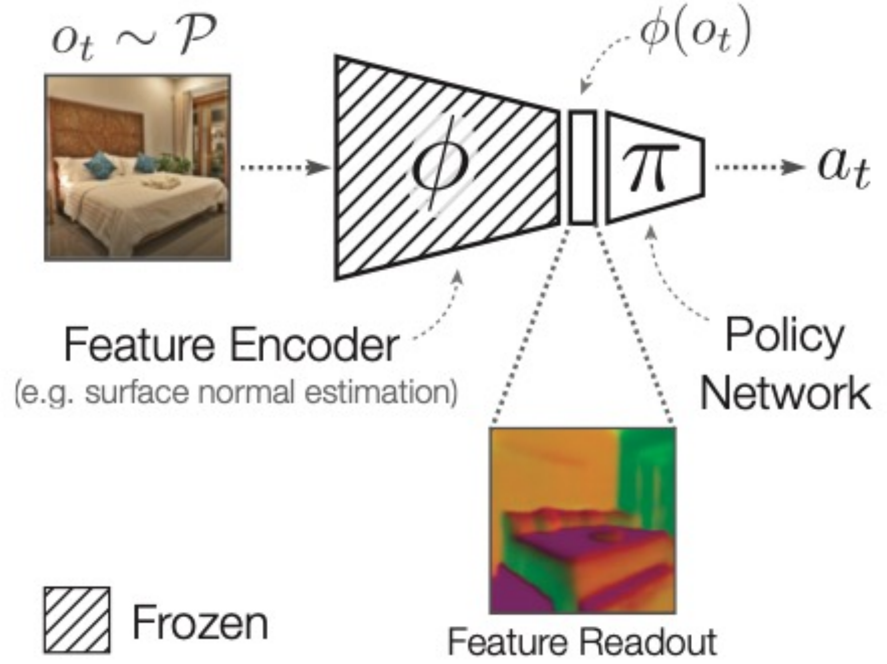
Outline

- Careful choice of paradigm
- Using knowledge from other domains
- Human demonstrations and feedback
- Scaling data collection

Using knowledge from other domains

- Transfer learning
- Multi-task learning
- Meta-learning
- Modular components



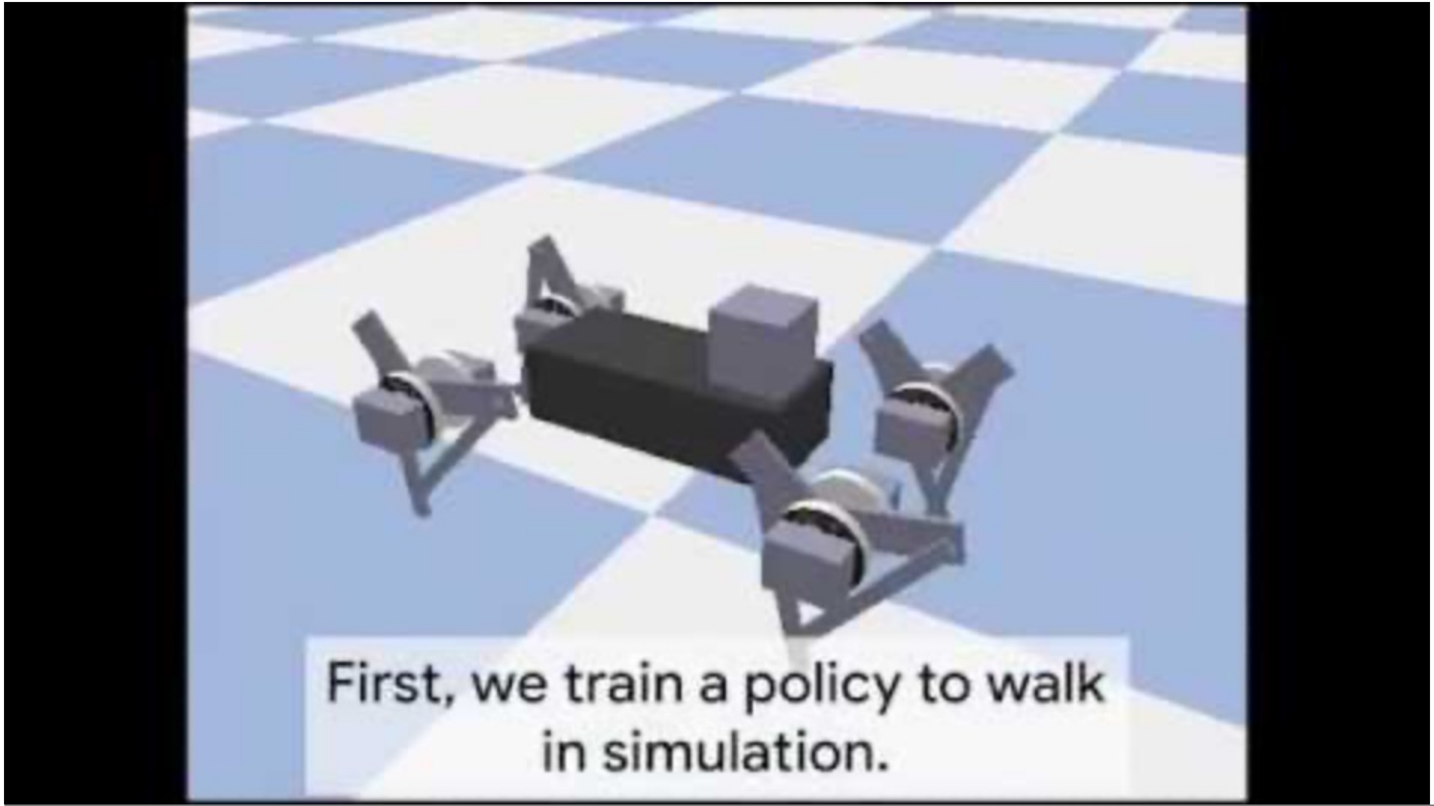


Mid level feature representations

Transfer learning

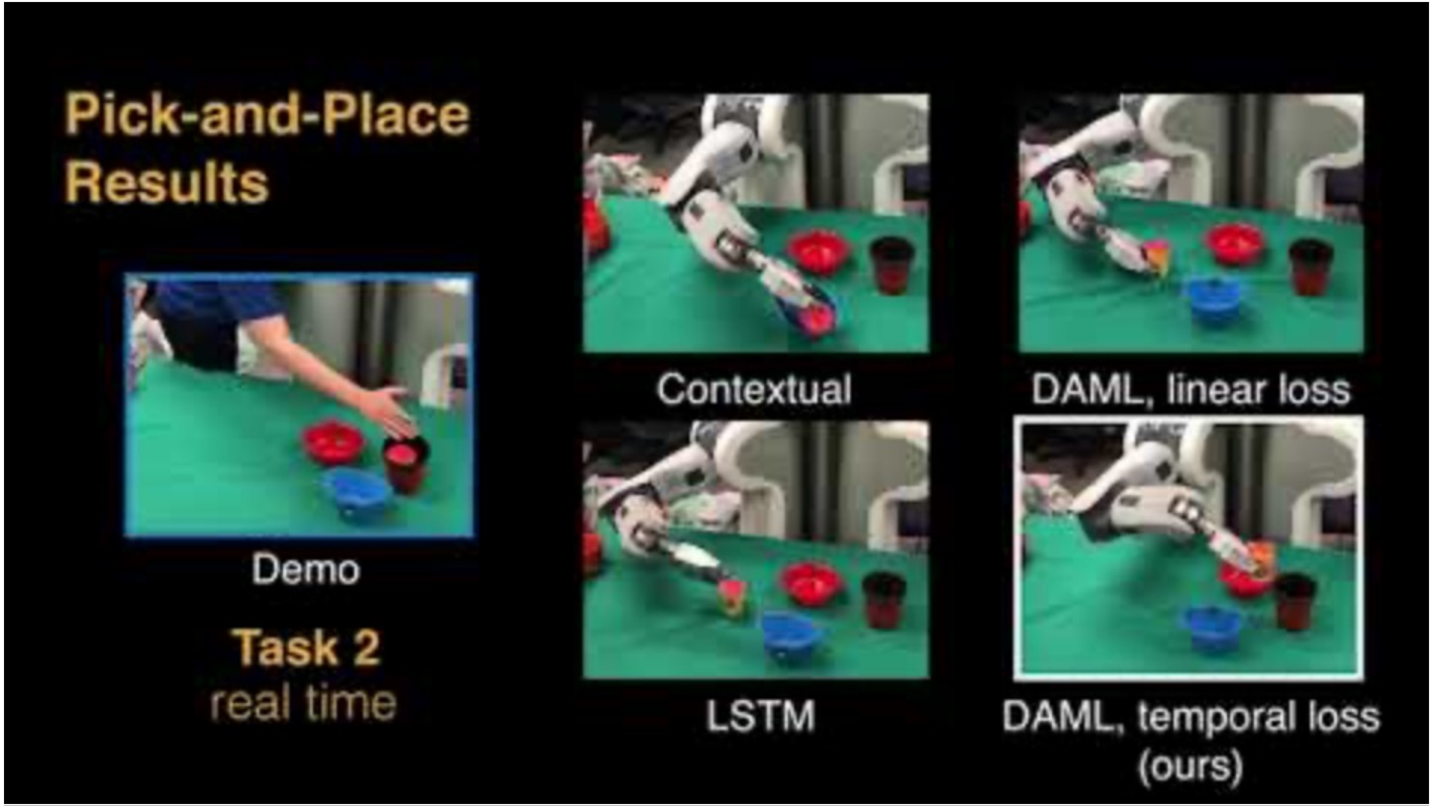
UNREAL: Unsupervised Auxiliary Task for RL Agent

Multi-task learning



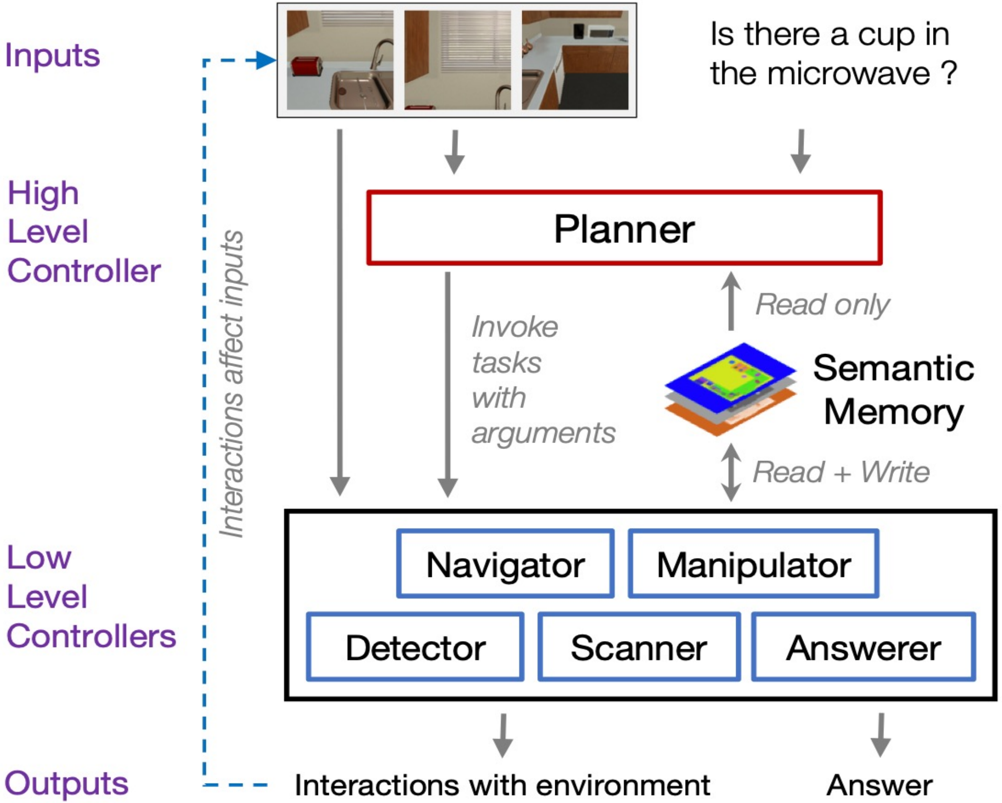
Evolutionary meta learning for adaptability in Legged Robots

Meta Learning



One-Shot Imitation from Watching Videos

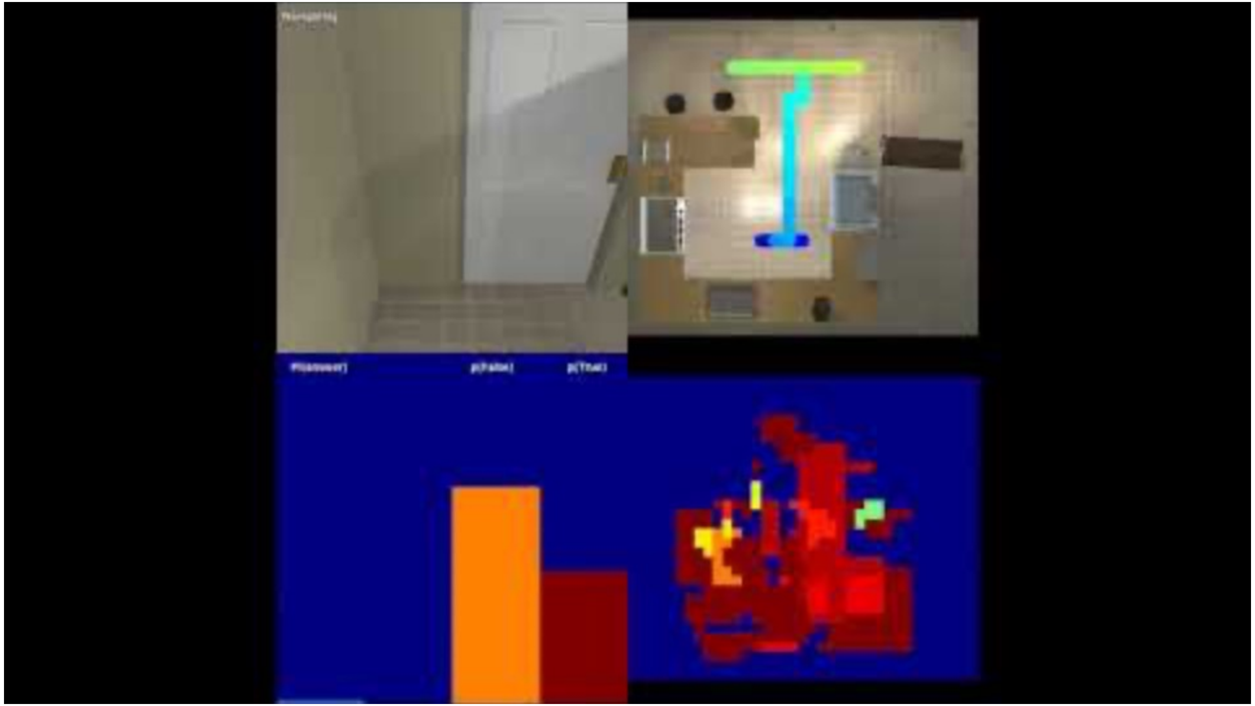
Meta Learning



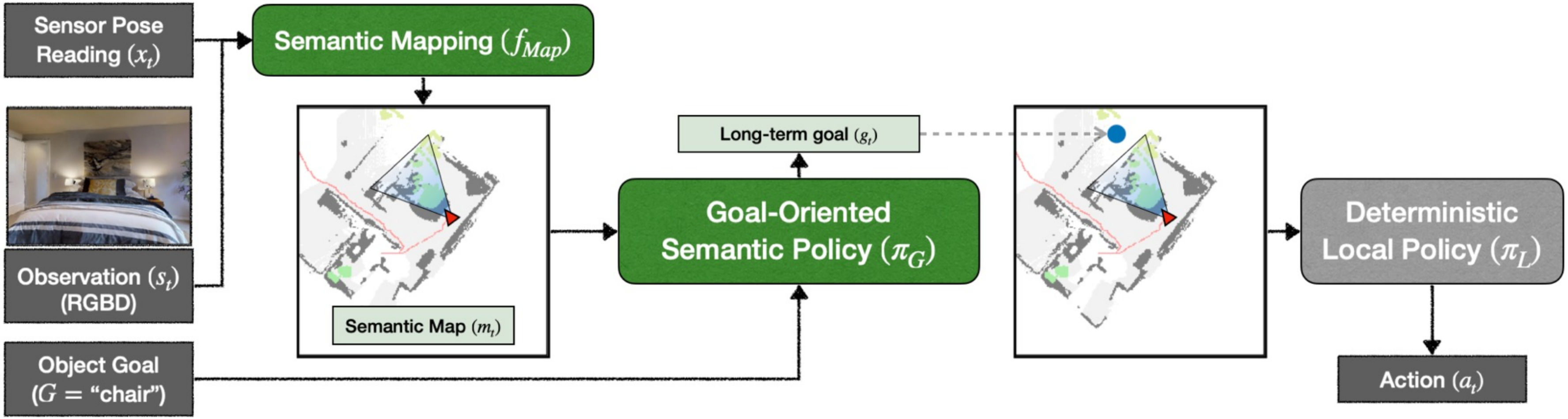
Hierarchical Interactive Memory Network (IQA)

Modular components for Interactive QA

Hierarchical Interactive Memory Network (IQA)



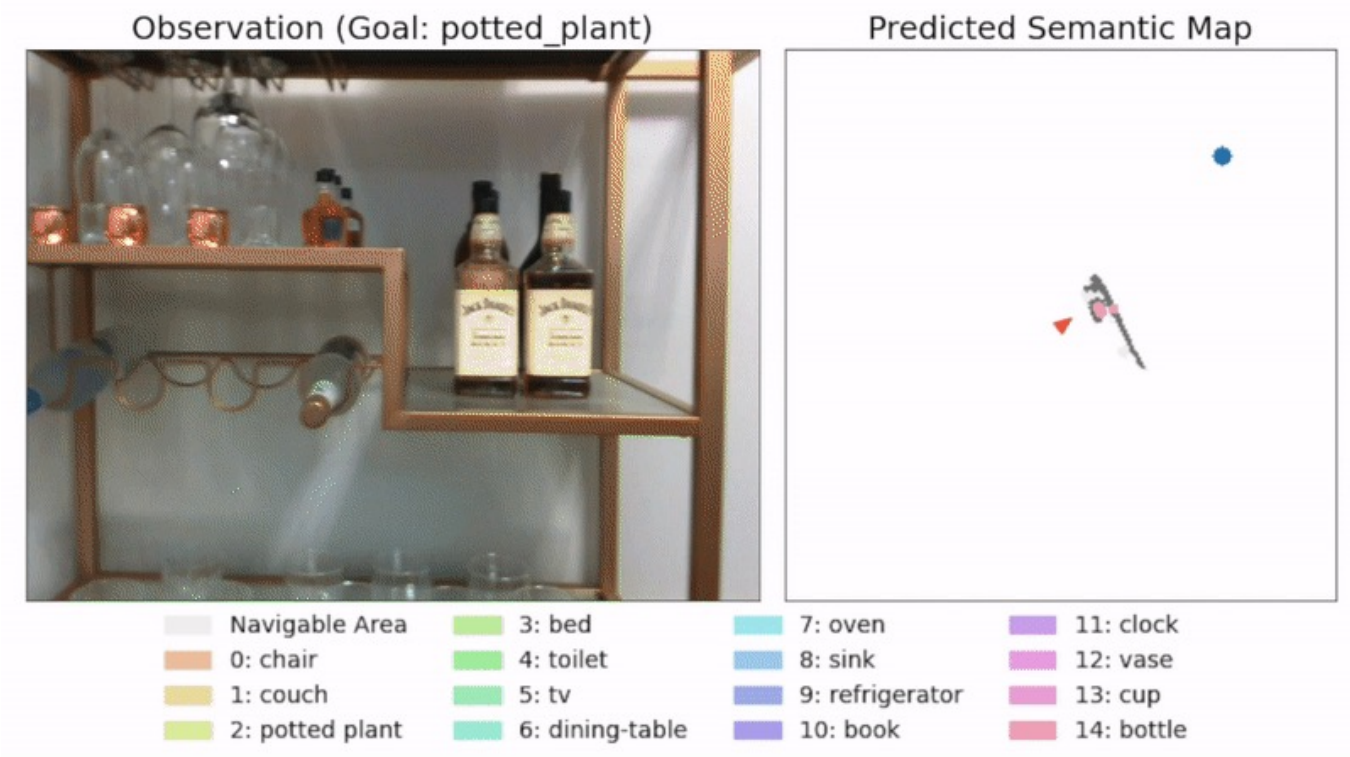
https://www.youtube.com/watch?v=pXd3C-1jr98&t=2s



Object Goal Navigation using Goal-Oriented Semantic Exploration

Modular components for object navigation

Real Transfer: Goal-Oriented Semantic Exploration



Outline

- Careful choice of paradigm
- Using knowledge from other domains
- Human demonstrations and feedback
- Scaling data collection

Expert demonstrations and human feedback

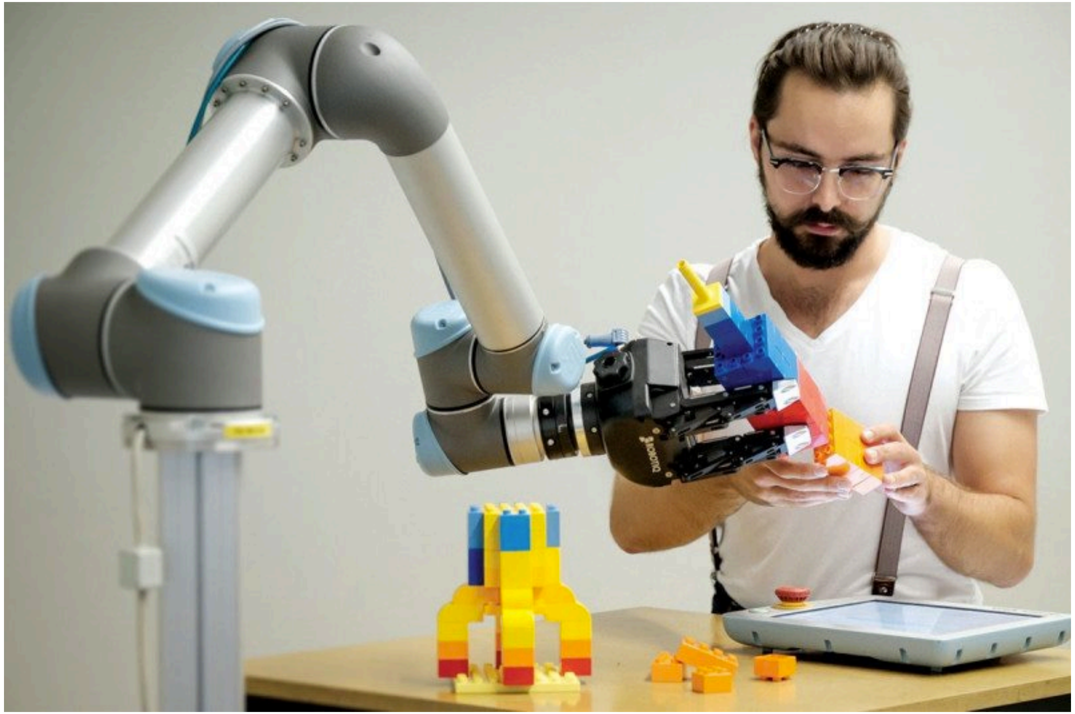


Image credit: http://sainslaboratmarit.blogspot.com/2016/10/robot-learns-to-play-with-lego-by.html

Imitation learning: copying experts

Algorithm:

1. Collect expert demonstrations (trajectories τ^*)
2. Treat demos as i.i.d. state-action pairs and split into dataset: $(s_0^*, a_0^*), (s_1^*, a_1^*), \dots$
3. Learn policy via supervised learning: minimize $L\left(a^*, \pi_\theta(s)\right)$

Vanilla imitation learning

NVIDIA AV

Combining IL + meta-learning

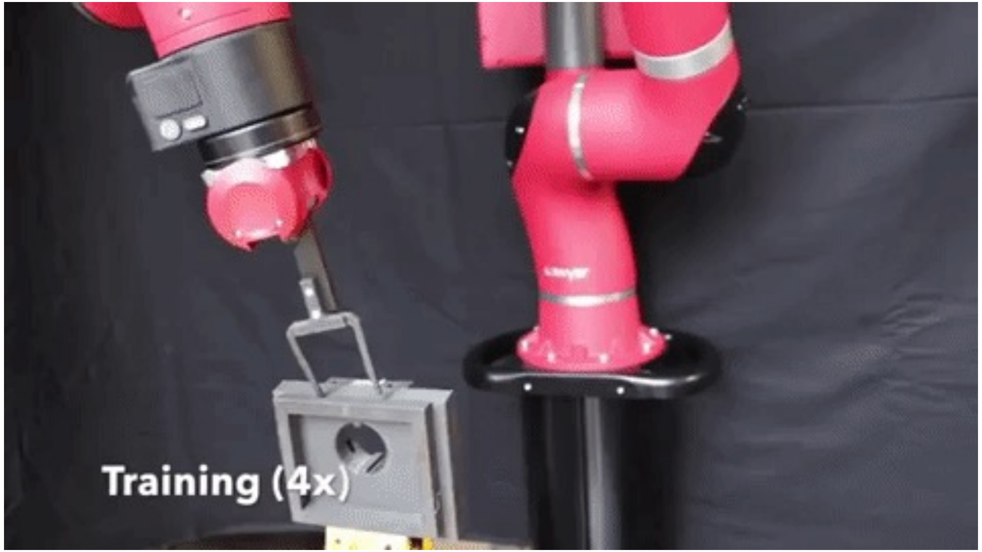
One-shot visual imitation learning

Using demos in an RL fashion

1. Split expert trajectories into (s_t, a_t, r_t, s_{t+1}) tuples
2. Insert into off-policy algorithm's data buffer

Demos for deep Q-learning

Clip insertion task

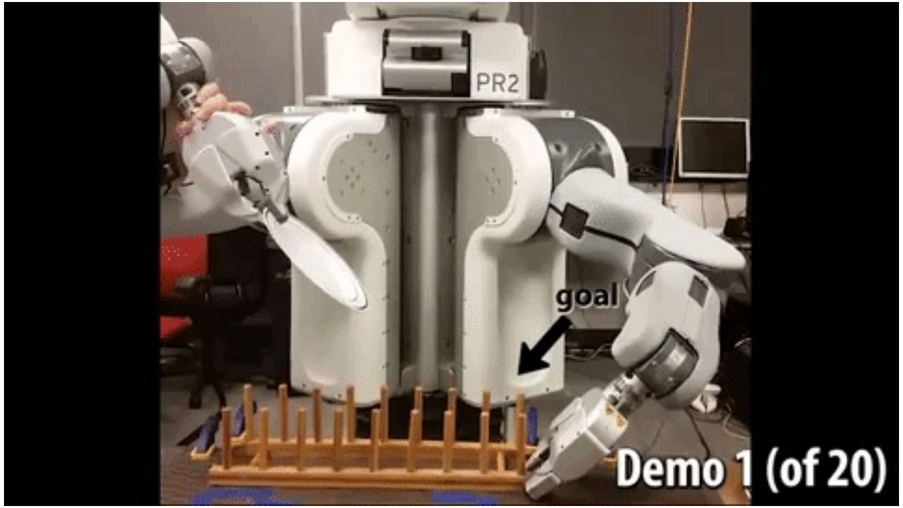


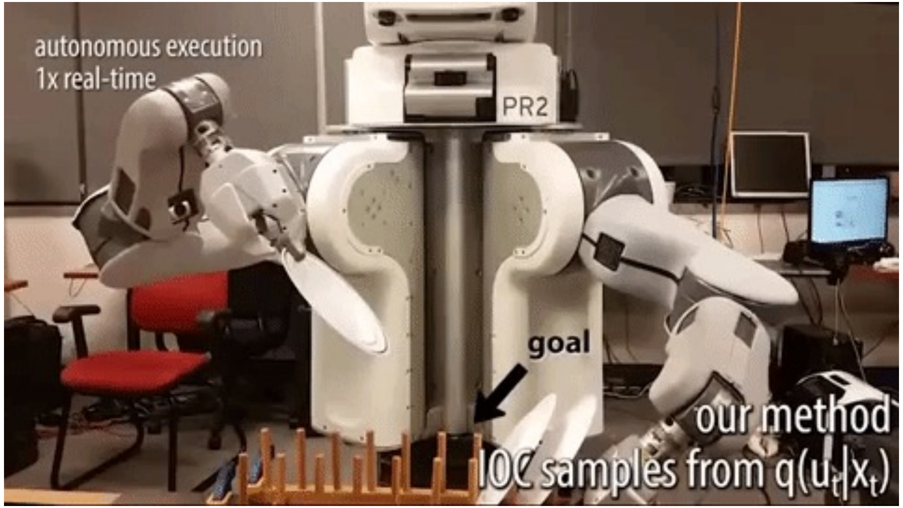
Inverse reinforcement learning

Given (s_t, a_t, s_{t+1}) from expert, assume expert optimality and find $r(s_t, a_t)$

Sample-based maxent IRL

Guided cost learning

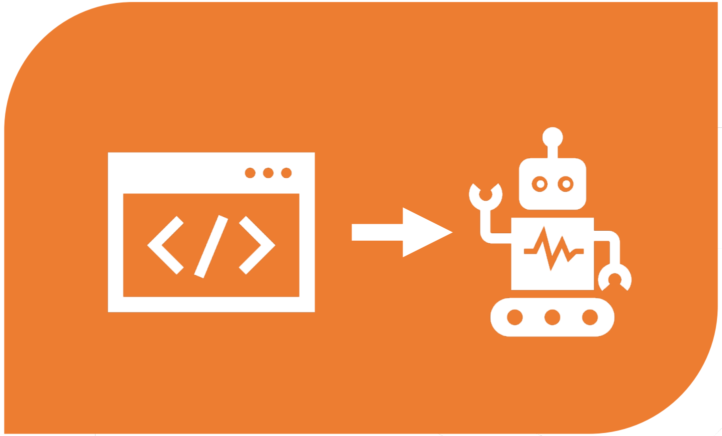




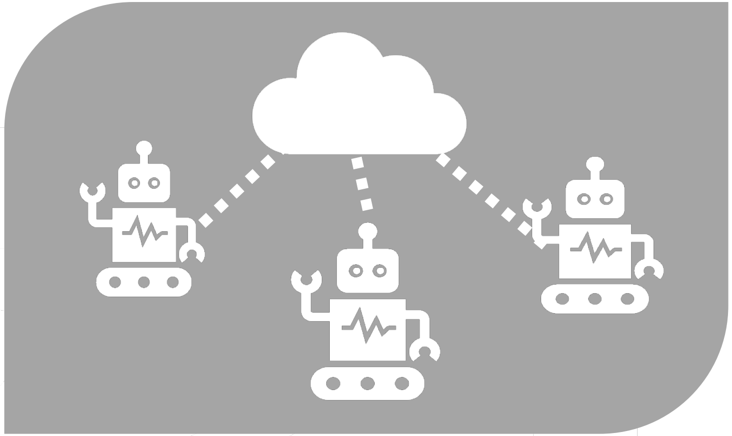
Outline

- Careful choice of paradigm
- Using knowledge from other domains
- Human demonstrations and feedback
- Scaling data collection

Gather data at scale



SIM-2-REAL TRANSFER



PARALLELIZED METHODS

Outline

- Careful choice of paradigm
- Using knowledge from other domains
- Human demonstrations and feedback
- Scaling data collection
    - Sim2real
    - Parallelized methods

Sim2Real

What is sim2real?

Benefits for training sim2real

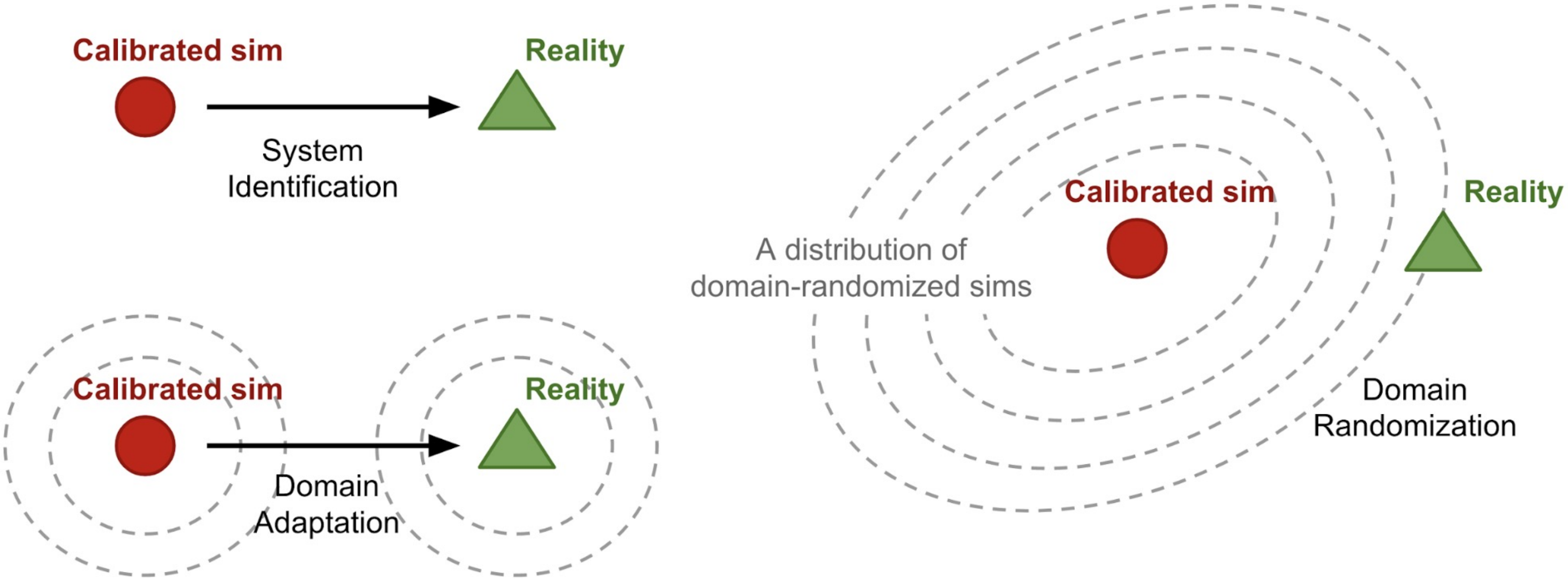
Sim2Real gap

Train in simulation, transfer policy to real world

- cheap data
- safe to learn and explore
- effortless to scale

visual and physical differences between simulation and reality

Ways of sim-2-real transfer



Source: https://lilianweng.github.io/lil-log/2019/05/05/domain-randomization.html

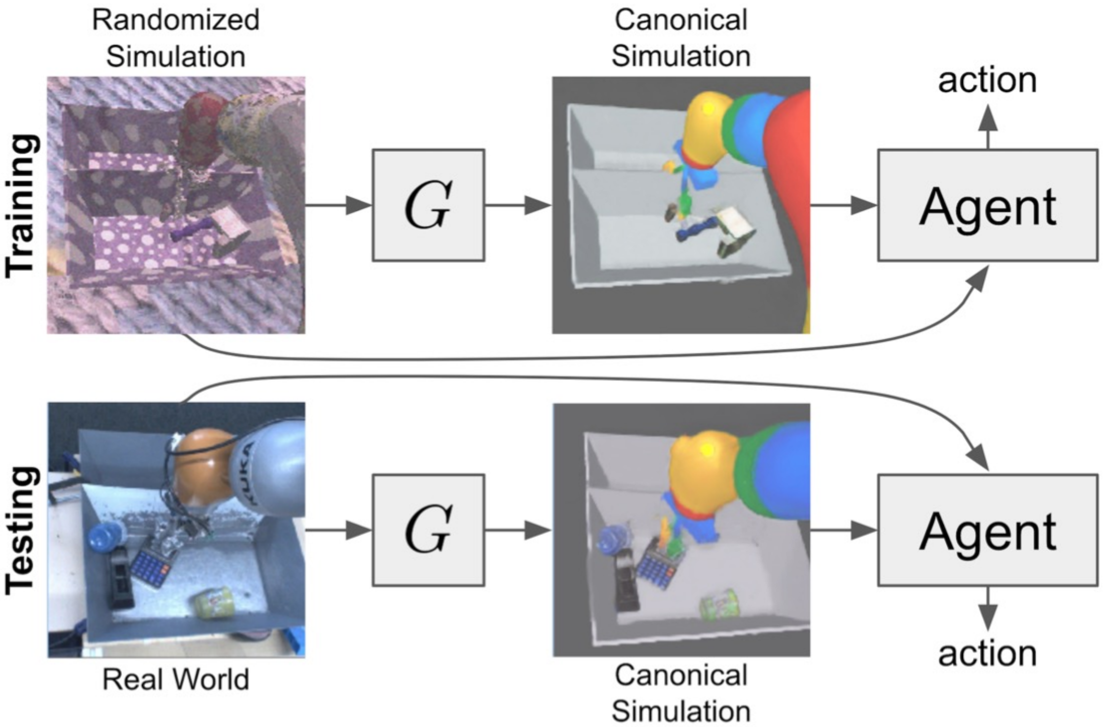
Simulator realism: What kind of realism is desirable?

VISUAL REALISM: MESHES

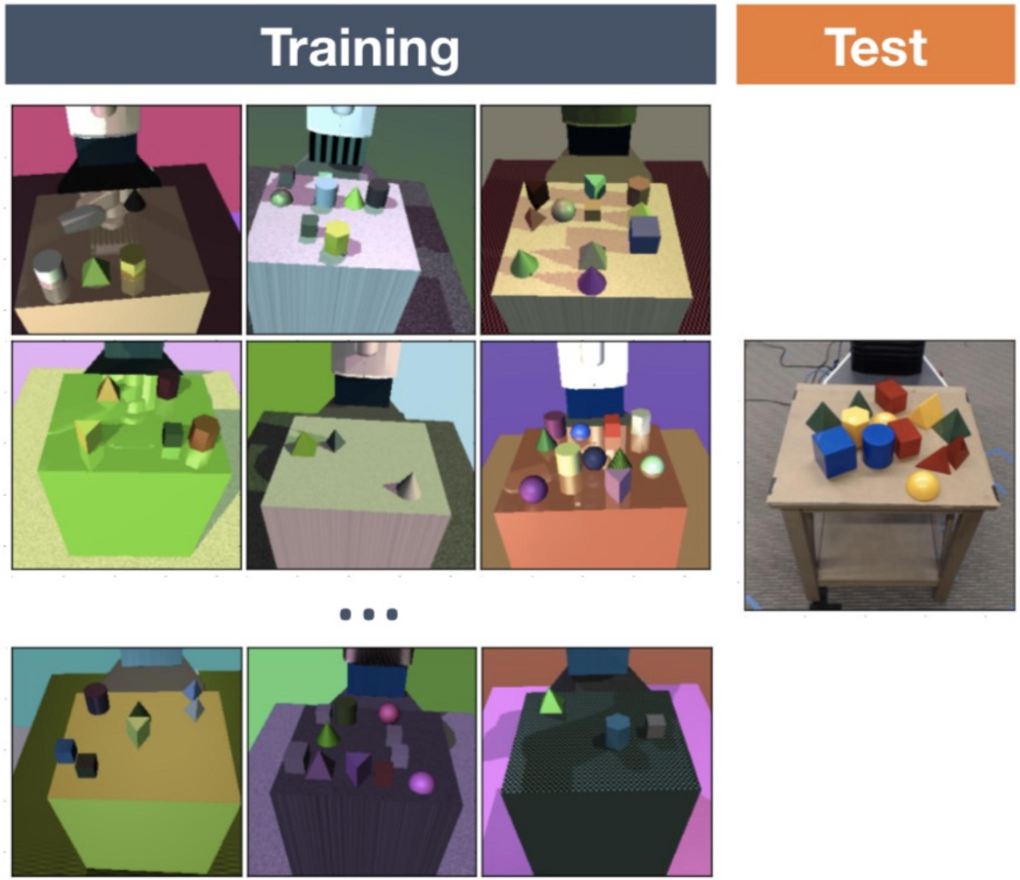
PHYSICAL REALISM: GAME ENGINES, CAD + PHYSICS MODELS

Combining both aspects of realism?

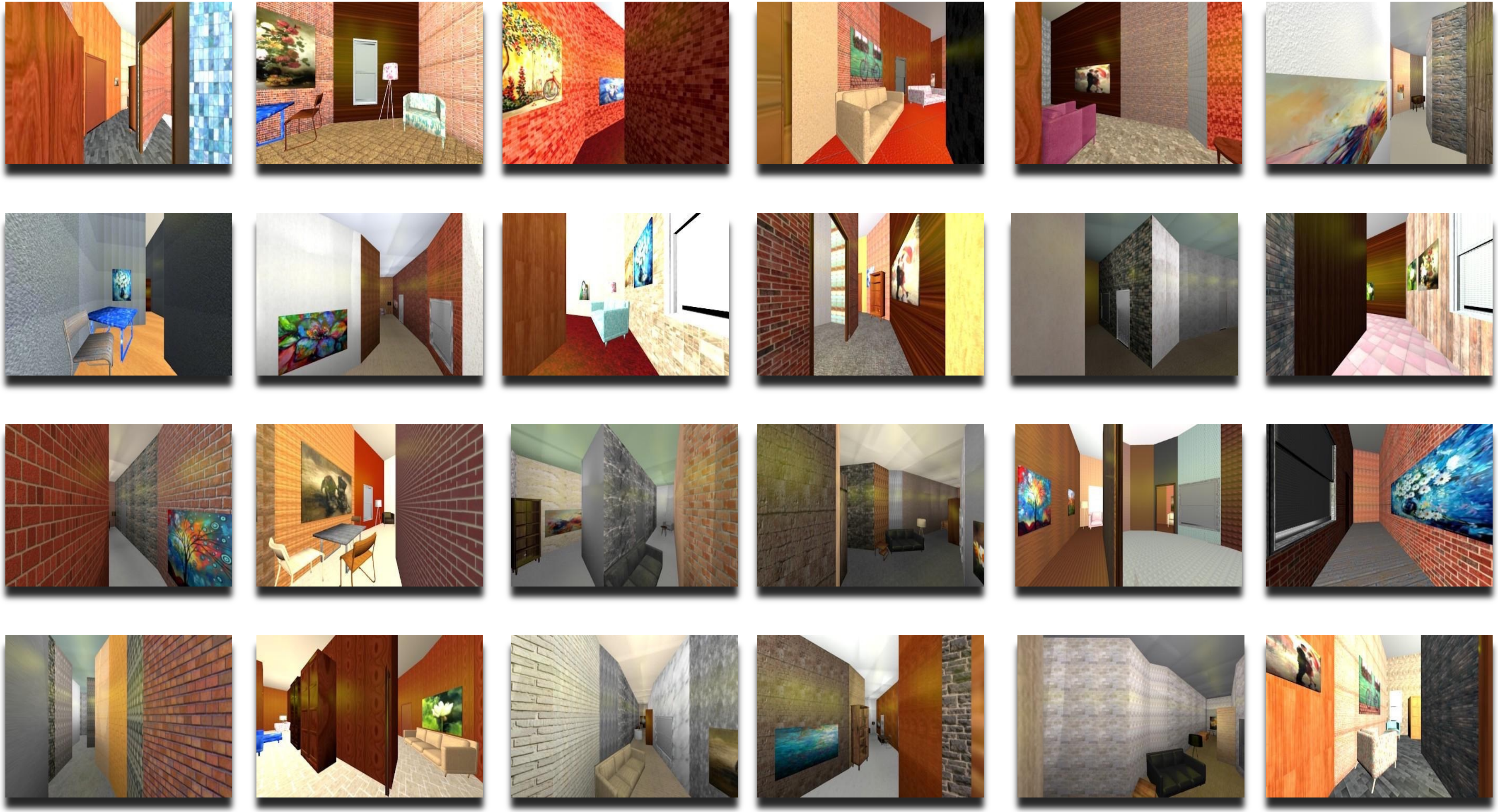
Source: iGibson, http://svl.stanford.edu/igibson/



Sim-to-Real via Sim-to-Sim: Data-efficient Robotic Grasping via Randomized-to-Canonical Adaptation Networks



Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World



CAD2RL: Real Single-Image Flight Without a Single Real Image

Outline

- Careful choice of paradigm
- Using knowledge from other domains
- Human demonstrations and feedback
- Scaling data collection
    - Sim2real
    - Parallelized methods

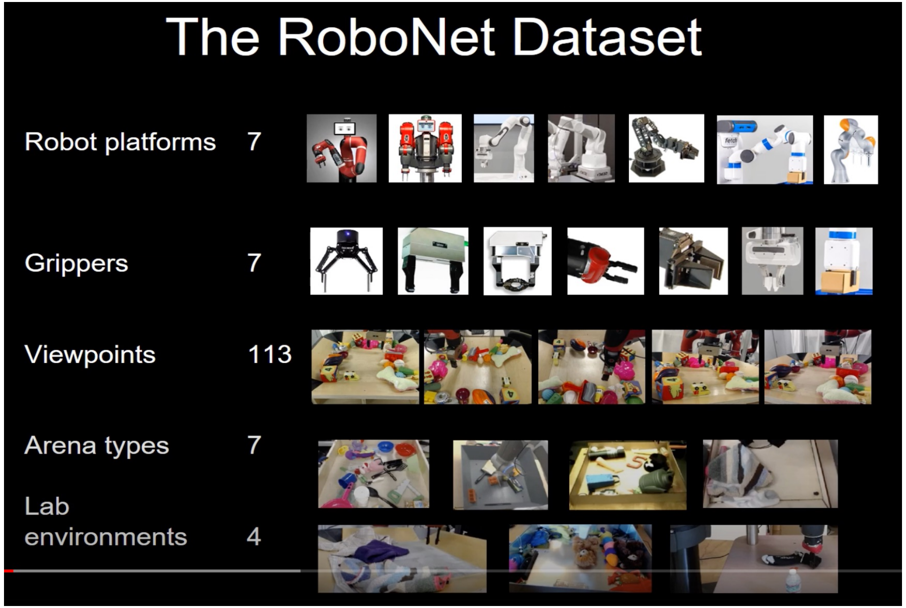
Parallelized methods with multiple devices

Parallelized, asynchronous data collection: edge workers merely send data to server

Federated learning: edge workers update personal models; asynchronously send model parameters to update the global features on server

Parallelized, asynchronous data collection

Distributed Q-learning algorithm with Google Arm Farm for grasping from vision



ROBONET: Scaling up data collection with multiple robots

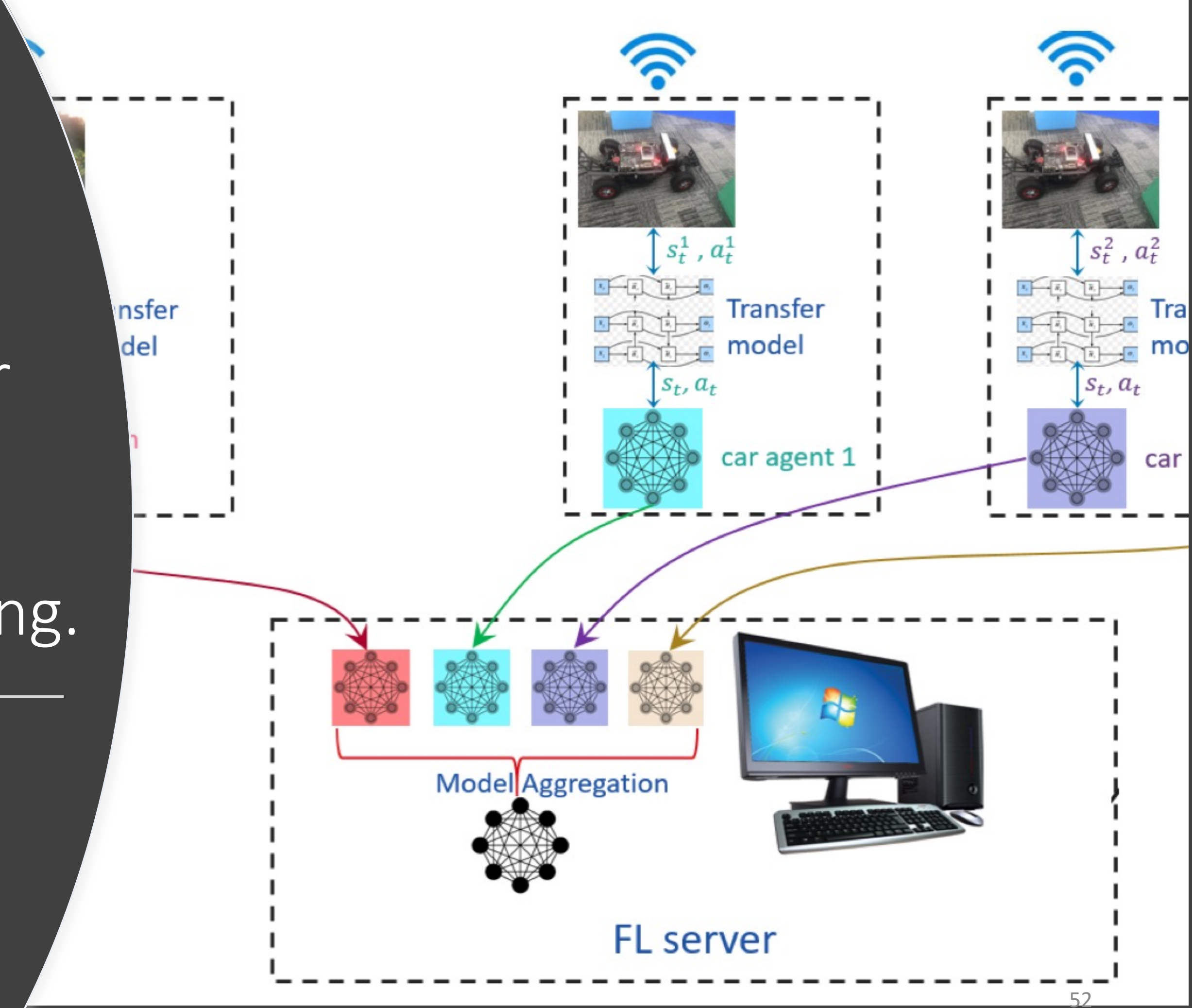
Federated learning

Local adaptation of robot

Global features in communication-efficient way

Privacy preserving way to leverage personal data

Federated Transfer Reinforcement Learning for Autonomous Driving.



Discussion section format

2 sets of questions. For each:

- Break-out room – 8 minutes
- Reconvene + share – 5 minutes

Questions part 1

1. When is DRL useful/necessary for embodied AI applications? (i.e. when do data-driven methods have an advantage over traditional planning & control methods?)

2. Is sample inefficiency a bottleneck in the progress of DRL for robotics?

3. Should our focus as a community be on circumventing sample efficiency issues (i.e. thru gathering data at scale) or addressing it head-on?

4. Does sim2real work? If sim2real works, can’t we just use any of our DRL algorithms, even if data inefficient?

Questions part 2

5. Are there any approaches that we missed?
6. Do you see ways in which these methods can be combined?
7. What’s wrong with the way we currently measure/quantify sample efficiency?
8. Federated learning has shown early promise in areas like query suggestions on mobile phones, smart speakers, etc. What other applications can you think of?

Wrapping up

- Slides will be posted to our WiML Slack
    - Channel name: #breakout_session_4-3
- You can contact us by private message on WiML Slack