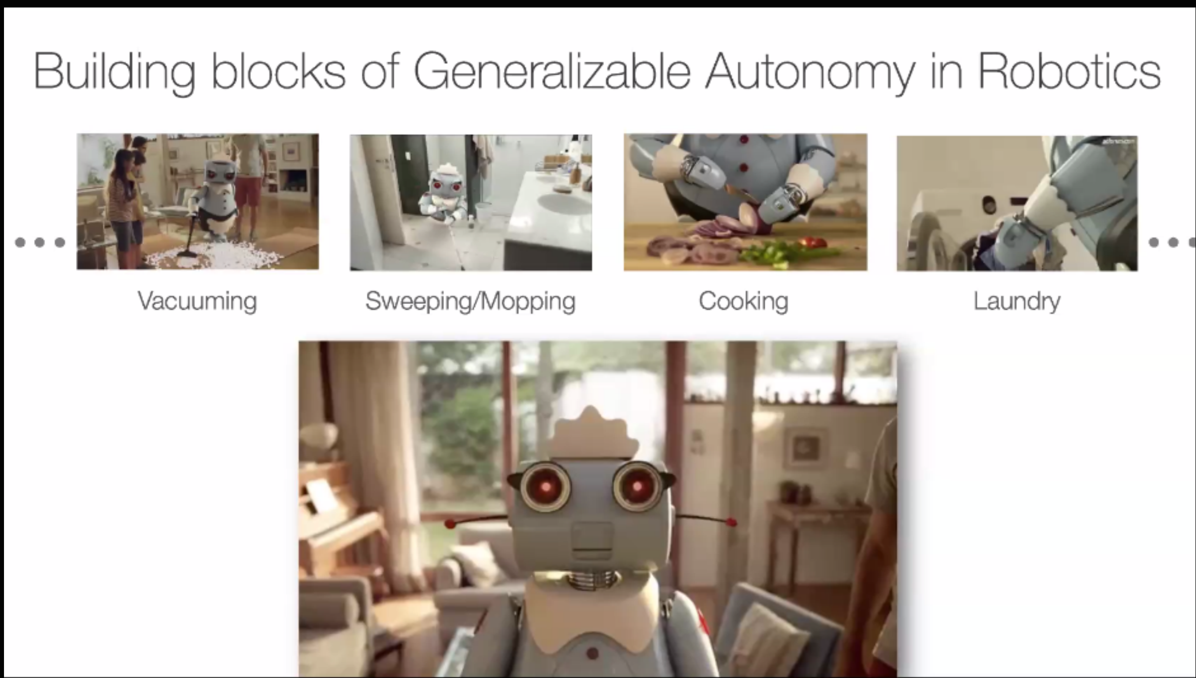
Day-15

Speaker: Prof. Animesh Garg, University of Toronto

Title: Building Blocks of Generalizable Autonomy



Building blocks of Generalizable Autonomy in Robotics

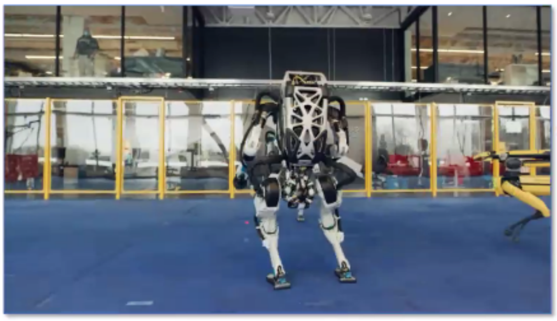
Vacuuming | Sweeping/Mopping | Cooking | Laundry

Diversity: New Scenes, Tools,…

Building blocks of Generalizable Autonomy in Robotics

How to achieve Algorithmic Generalization

Objects, Scenes, Semantics, Task Goals



Known Structured Environment

* Engineered Robots are for particular task

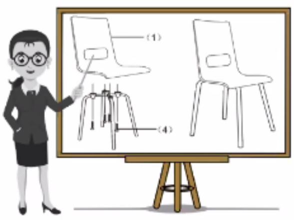
Unstructured/Unknown New Environment

Boston Dynamics

Building blocks of Generalizable Autonomy in Robotics

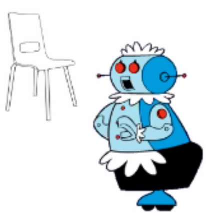
Vision: Build Intelligent Robotic Companions towards Human Enrichment and Augmentation

Supervision



Input

Task Imitation



Task Performance in Data Scarce Set-up

Generalization

New Task Variations in Novel Environments

Building blocks of Generalizable Autonomy in Robotics

Imitation: But at which level? What should I copy?

Imitative Babbling
Movement Skills

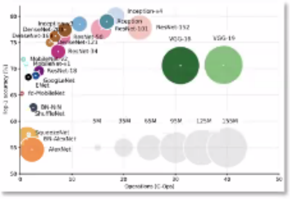
Dexterity
Skill Sequencing

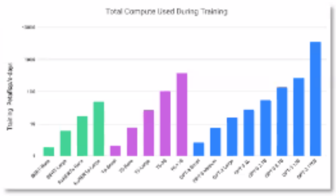
Causality
Semantic Purpose

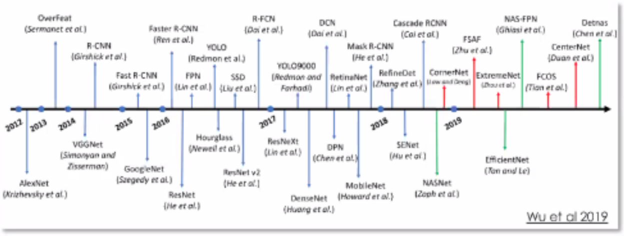
Task Specification
{Language, Video,
Kinesthetic}

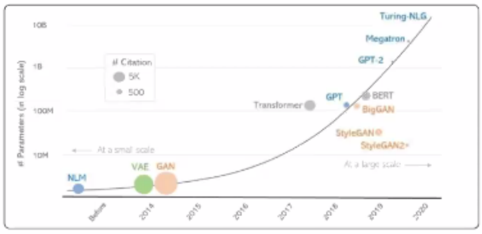
Cheng et al. Sci.Rob. 2019

Generalization in Vision & Language









Computer Vision

Language Understanding

Generalization in Vision & Language

Ingredients of Modern Machine Learning & Applications

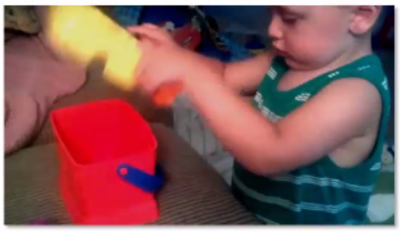
Vision   Language

Large Structured Models
- Over-parameterized
- Structured Biases

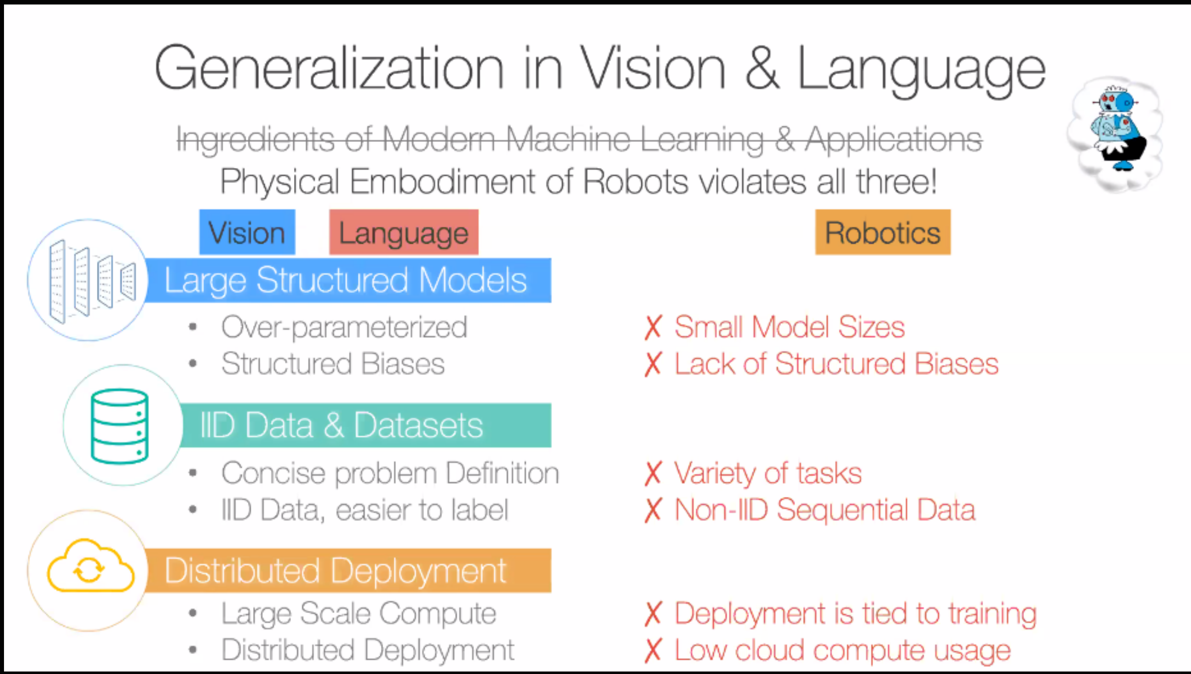
IID Data & Datasets
- Concise problem Definition
- IID Data, easier to label

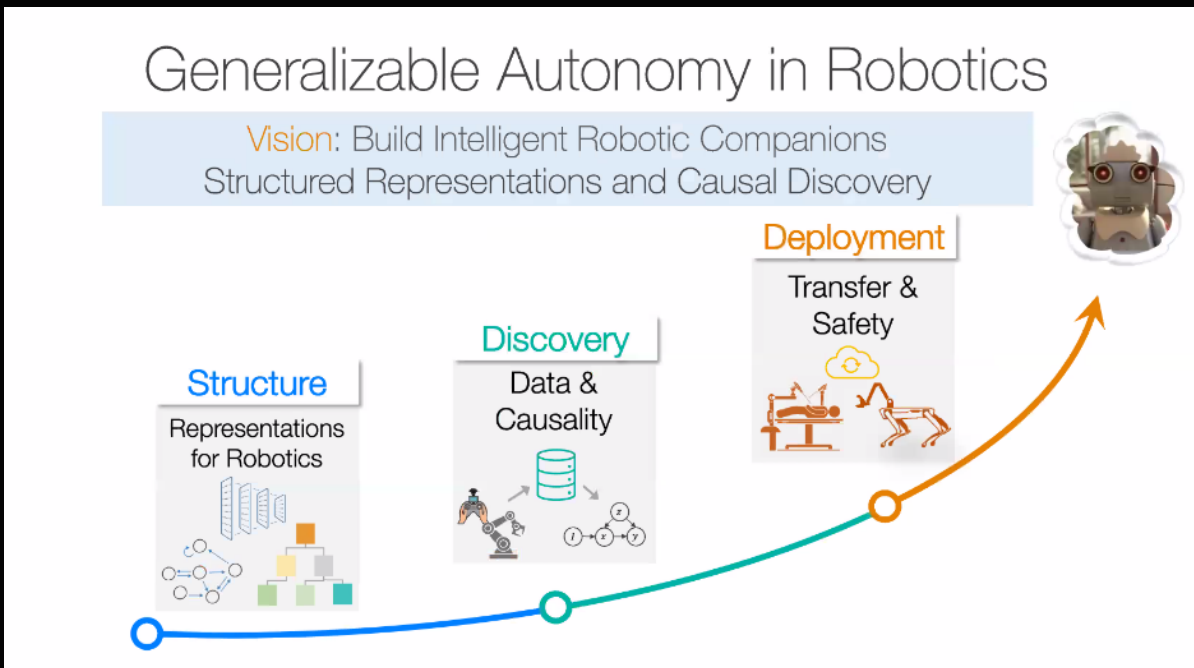
Distributed Deployment
- Large Scale Compute
- Distributed Deployment

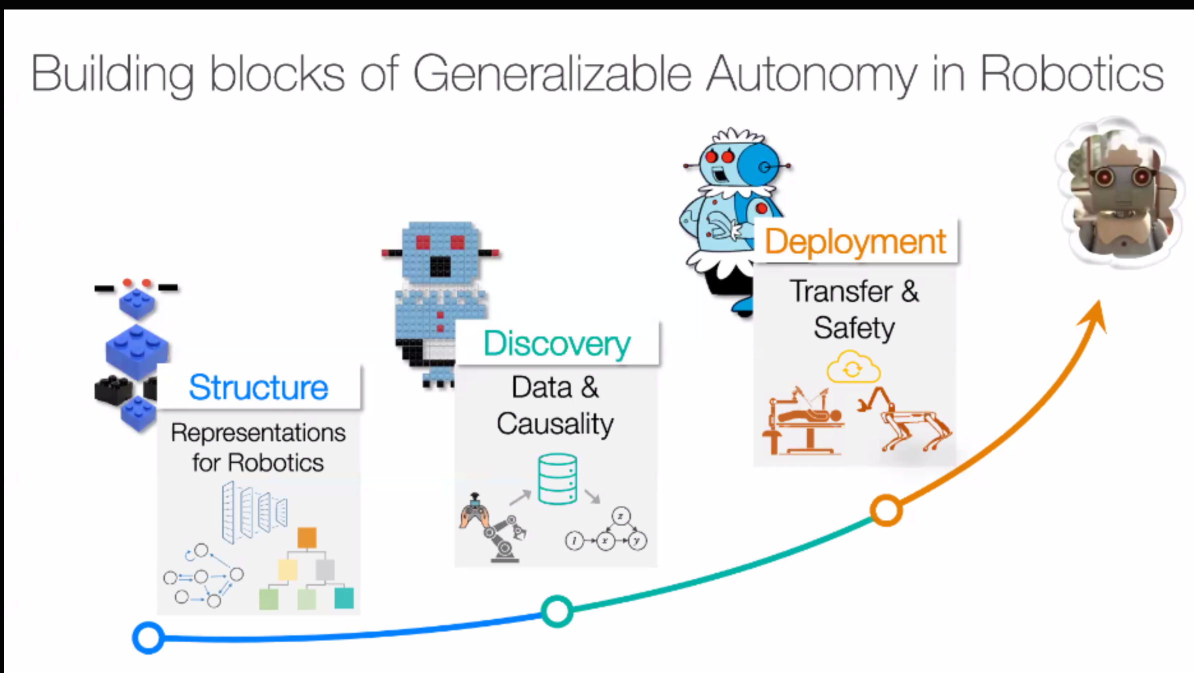
Robotics

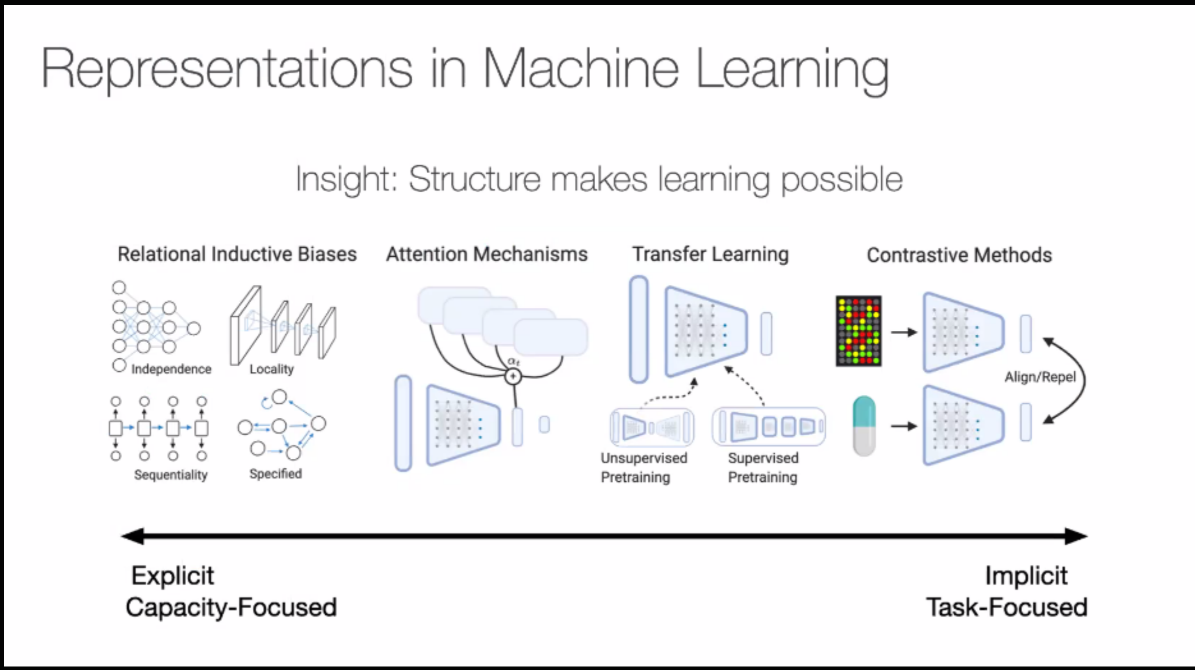


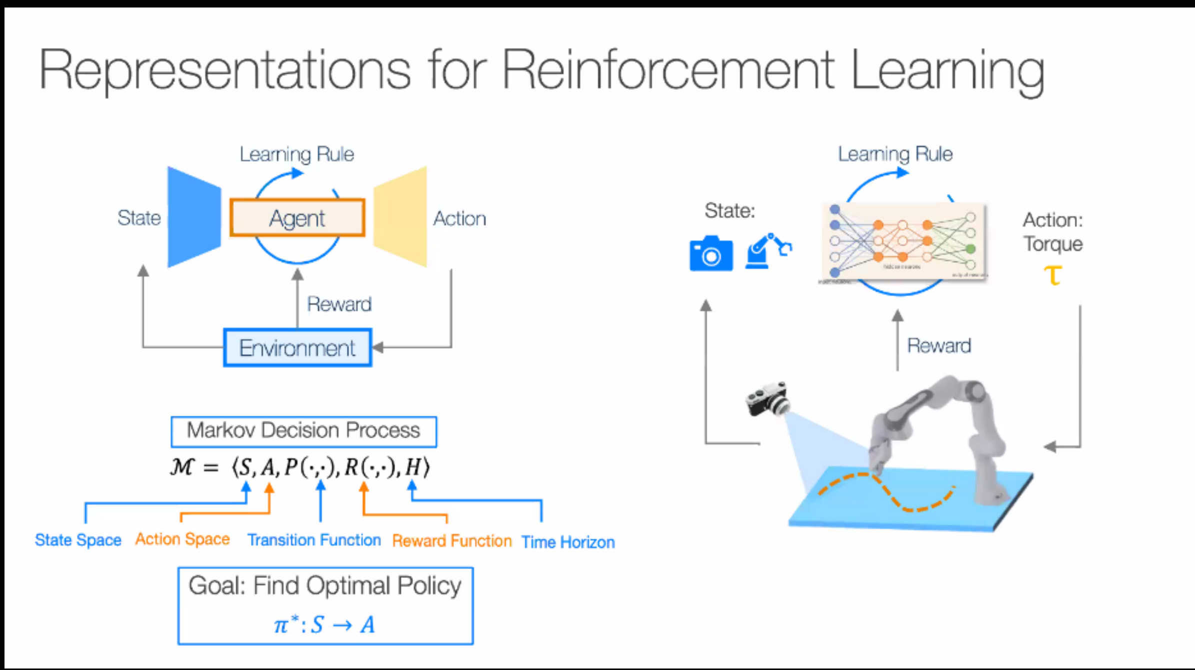
Embodied Interaction
Creative solutions
Dexterous manipulation
Breaks assumptions!

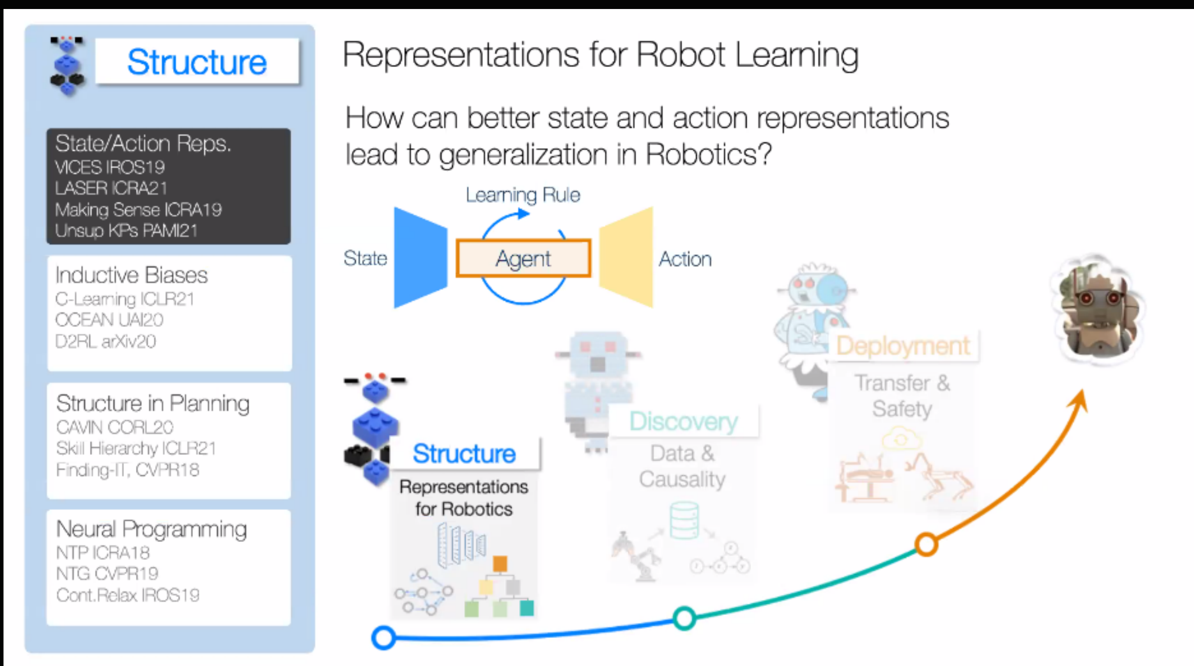

Generalization in Vision & Language
~~Ingredients of Modern Machine Learning & Applications~~
Physical Embodiment of Robots violates all three!
Vision
Language
Robotics
Large Structured Models
• Over-parameterized
• Structured Biases
✗ Small Model Sizes
✗ Lack of Structured Biases
IID Data & Datasets
• Concise problem Definition
• IID Data, easier to label
✗ Variety of tasks
✗ Non-IID Sequential Data
Distributed Deployment
• Large Scale Compute
• Distributed Deployment
✗ Deployment is tied to training
✗ Low cloud compute usage


Generalizable Autonomy in Robotics
Vision: Build Intelligent Robotic Companions
Structured Representations and Causal Discovery
Deployment
Transfer & Safety
Discovery
Data & Causality
Structure
Representations for Robotics


Building blocks of Generalizable Autonomy in Robotics
Deployment
Transfer & Safety
Discovery
Data & Causality
Structure
Representations for Robotics


Representations in Machine Learning
Insight: Structure makes learning possible
Relational Inductive Biases
Independence
Locality
Sequentiality
Specified
Attention Mechanisms
Transfer Learning
Unsupervised Pretraining
Supervised Pretraining
Contrastive Methods
Align/Repel
Explicit
Capacity-Focused
Implicit
Task-Focused


Representations for Reinforcement Learning
Learning Rule
State
Agent
Action
Reward
Environment
Markov Decision Process
$\mathcal{M} = \langle S, A, P(\cdot,\cdot), R(\cdot,\cdot), H \rangle$
State Space
Action Space
Transition Function
Reward Function
Time Horizon
Goal: Find Optimal Policy
$\pi^*: S \rightarrow A$
Learning Rule
State:
Action:
Torque
τ
Reward


Structure
State/Action Reps.
VICES IROS19
LASER ICRA21
Making Sense ICRA19
Unsup KPs PAMI21
Inductive Biases
C-Learning ICLR21
OCEAN UAI20
D2RL arXiv20
Structure in Planning
CAVIN CORL20
Skill Hierarchy ICLR21
Finding-IT, CVPR18
Neural Programming
NTP ICRA18
NTG CVPR19
Cont.Relax IROS19
Representations for Robot Learning
How can better state and action representations
lead to generalization in Robotics?
Learning Rule
State
Agent
Action
Deployment
Transfer &
Safety
Discovery
Data &
Causality
Structure
Representations
for Robotics

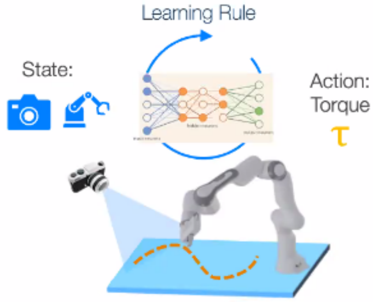
Representations for RL

Task Spaces

Latent (low-dim) manifold for maximum task-progress (reward)!

People Opening a Door for Friends
No vertical/sideways yanking (!)

Robot Trying to Open a Door
Applies forces in all direction (!)

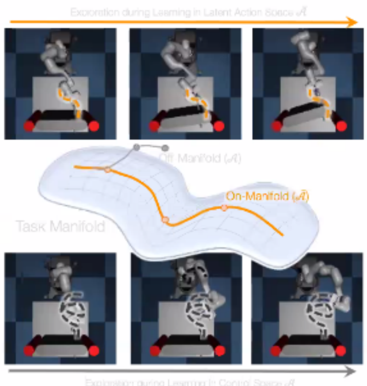


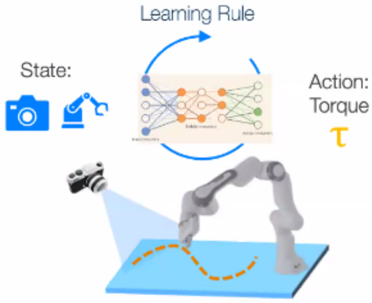
Representations for RL

Task Spaces

Latent (low-dim) manifold for maximum task-progress (reward)!

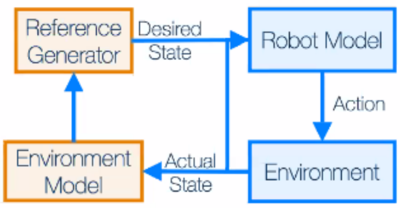
Multiple valid-task spaces, may exist, but only some are useful!

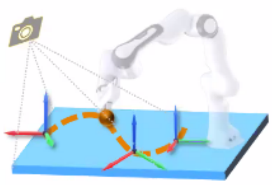




Representations RL: Task Spaces

RL with Variable Impedance Task-Space

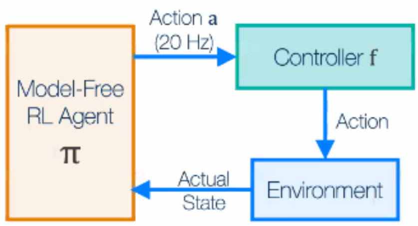




Variable Impedance Control in End-Effector Space, IROS 2019

Representations RL: Task Spaces

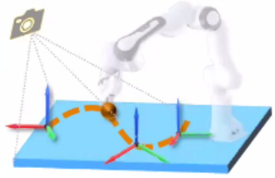
RL with Variable Impedance Task-Space



Reference Generator (learned)

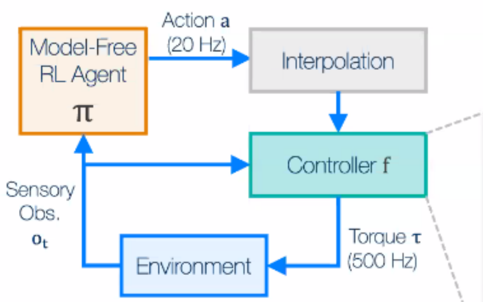
$$\tau = f(\pi(o_t))$$

Robot Model (Deterministic)



Representations RL: Task Spaces

RL with Variable Impedance Task-Space



$$\tau = f(\pi(o_t))$$

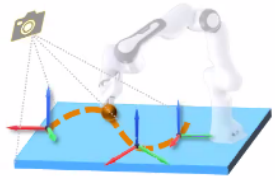
$$\pi(o_t) = a: [\boxed{x_d, \dot{x}_d}, \boxed{K_p, K_v}]$$

Pose and Velocity

Impedance Gains

$$\tau = \boxed{f}(x_d, \dot{x}_d, K_p, K_v)$$

Deterministic Position-Velocity Control Jacobian J and Inertia M



Representations RL: Task Spaces

Surface Wiping

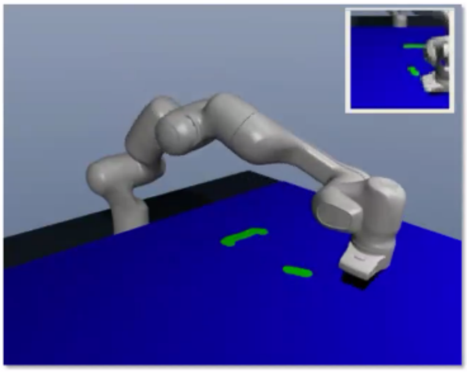
Input: Image (48x48)

Minimize the number of Dirty Tiles

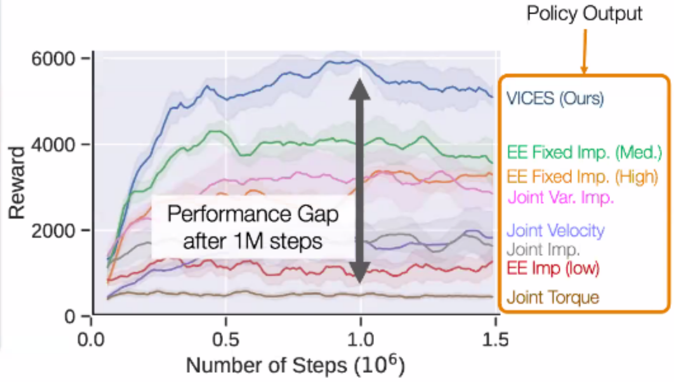
Maintain Contact with the Table

Don't push with more than Robot Payload

Reward: $\lambda_1 \sum(\text{dirt\_on\_table}) + \lambda_2(\text{distance\_to\_table}) - \lambda_3 \mathbb{I}\,(F \geq 40N)$

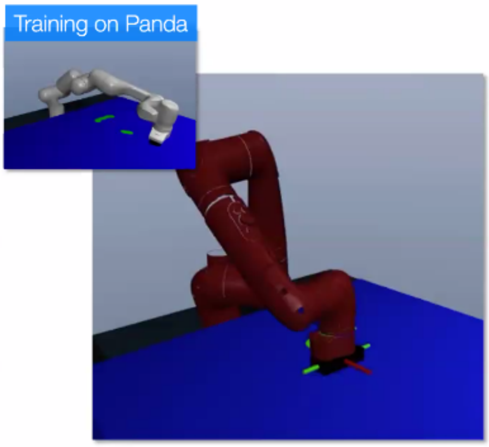


Trained Policy Rollout (Ours)

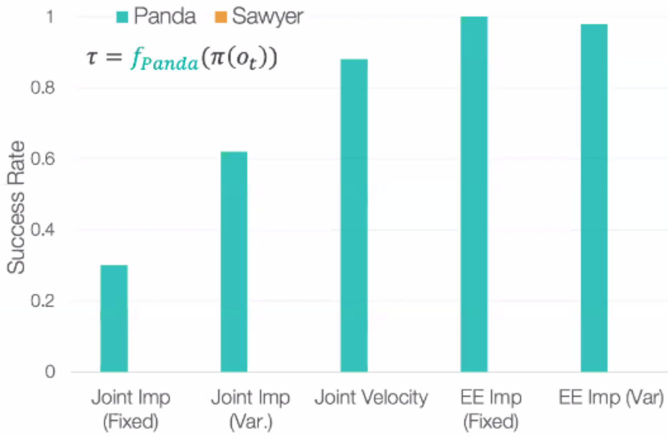


Evaluation During Training (PPO)

Representations RL: Task Spaces

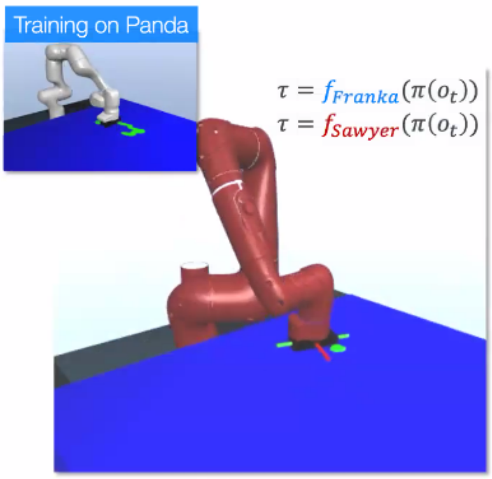


Evaluation on Sawyer

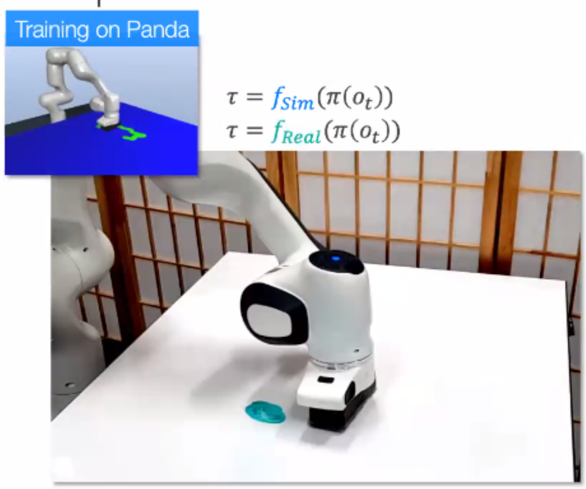


Variable Impedance Control in End-Effector Space, IROS 2019

Representations RL: Task Spaces



Evaluation on Sawyer



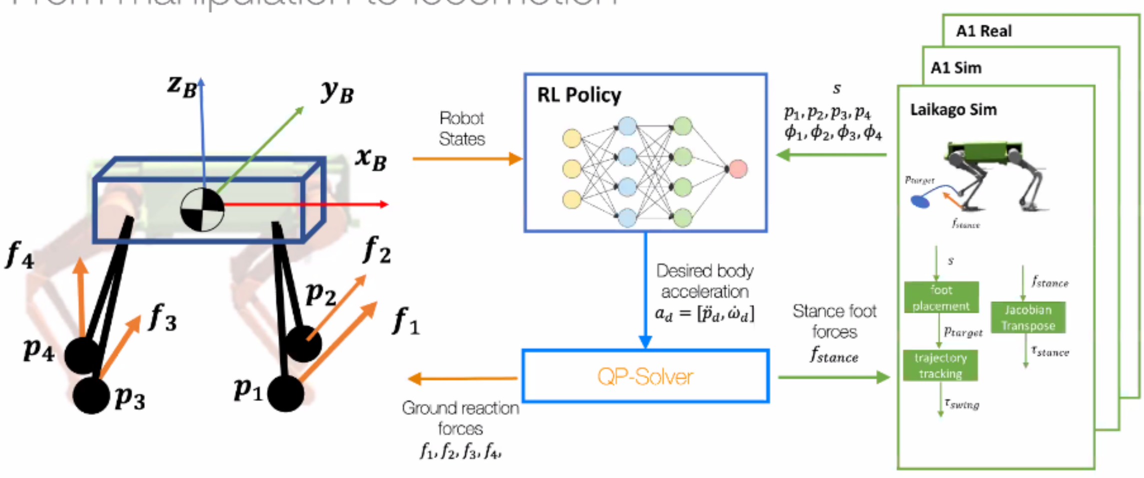
Evaluation on Real Robot

Success 80% (10 Trials)

Variable Impedance Control in End-Effector Space, IROS 2019

Representations RL: Task Spaces

From manipulation to locomotion



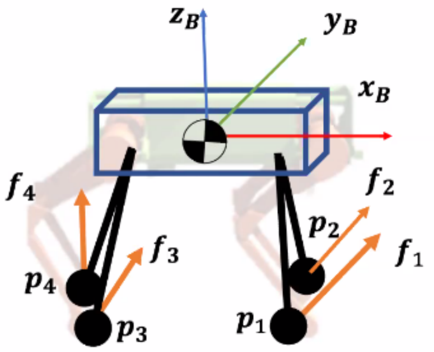
Centroidal Task Space

Full Robot Model

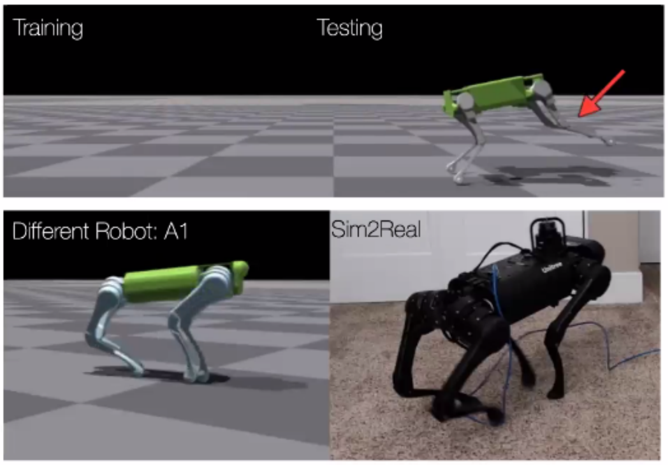
GLIDE: Generalizable Quadrupedal Locomotion, under review 2021

Representations RL: Task Spaces

From manipulation to locomotion



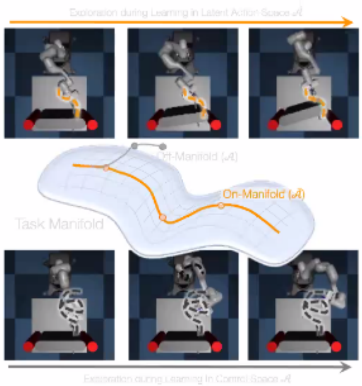
Centroidal Task Space



GLiDE: Generalizable Quadrupedal Locomotion, under review 2021

Representations RL: LASER

Learning the action space



$\mathcal{A}$ True action space

$\bar{\mathcal{A}}$ Latent action space

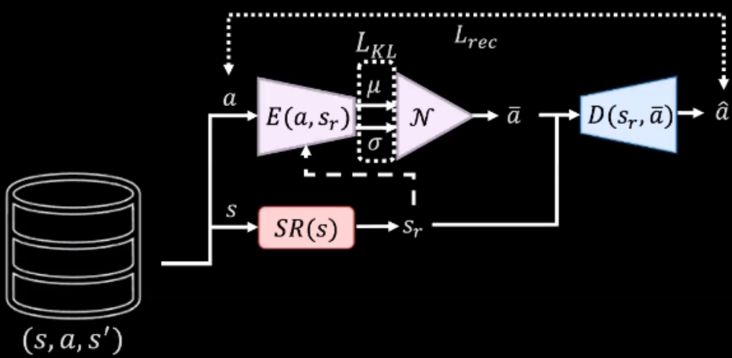
Conditioned	Mapping depends on State	$\bar{a} = f_{enc}(s, a)\ \forall\, s \in \mathcal{M}$
Controllable	Reachability on Manifold	$s' \leftarrow \mathcal{T}(s, a),\ \forall\, s \in \mathcal{M}$ $\exists\, \bar{a}, \text{s.t.}\, s' \leftarrow \bar{\mathcal{T}}(s, \bar{a})$
Consistent	Translational invariance	$\bar{\mathcal{T}}(f(s), \bar{a}) \approx f^{-1}(\bar{\mathcal{T}}(s, \bar{a}))$ $f: s \rightarrow s'\ \in \mathcal{M}$

Learn a manifold of actions for a task within the full original action space

LASER: Learning a Latent Action Space for Efficient RL, ICRA 2021

Representations RL: LASER

Learning the action space

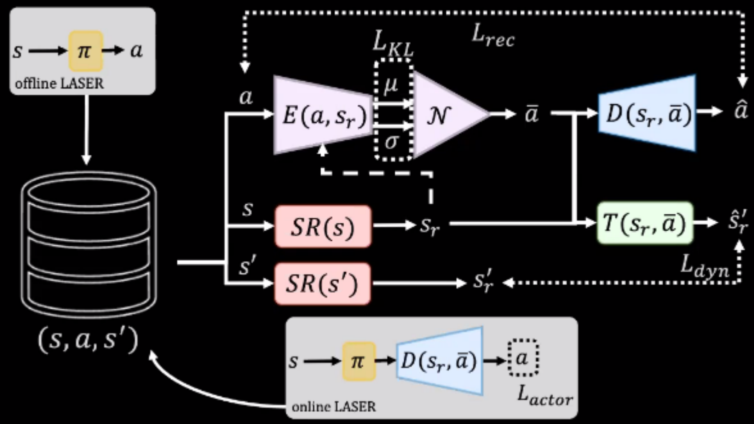


Variational auto encoder architecture with dynamic consistency

LASER: Learning a Latent Action Space for Efficient RL, ICRA 2021

Representations RL: LASER

Learning the action space

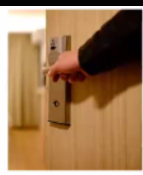


Can Use Batch Data as well as Replay Buffer concurrent with Policy

LASER: Learning a Latent Action Space for Efficient RL, ICRA 2021

LASER: Learning the action space

Evaluation



SAC Baseline
(Same Task)

Offline LASER
(Same Task)

Train Action Space
then Train Policy

LASER: Learning a Latent Action Space for Efficient RL, ICRA 2021

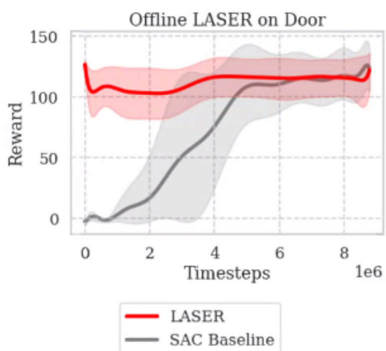
LASER: Learning the action space

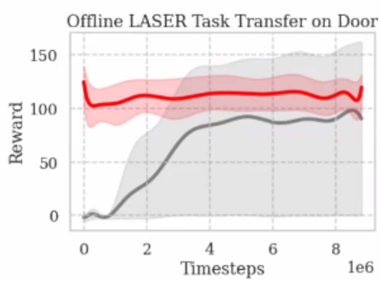
Evaluation

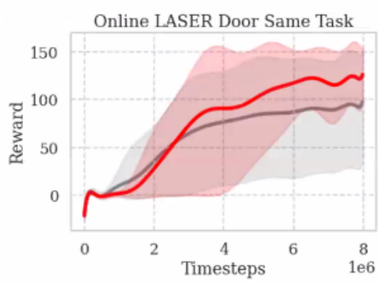
Collect Task Demos,
Learn Action Space
Does this work?

It must have memorized the
solutions (data from experts)!
Does it Generalize?

Gotcha! But policy already
known (good demos)!
No expert(s), what then?

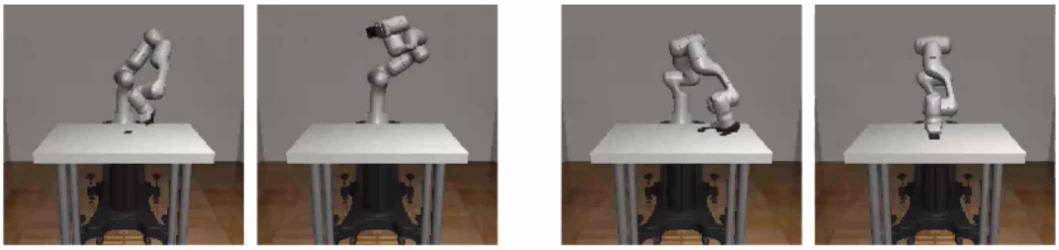






LASER: Learning a Latent Action Space for Efficient RL, ICRA 2021

LASER: Learning the action space
Evaluation



SAC Baseline (Same Task)

Offline LASER (Same Task)
Train Action Space then Train Policy

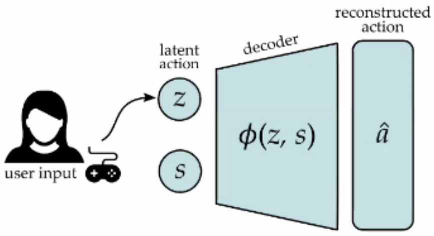
SAC Baseline (Transfer Task)

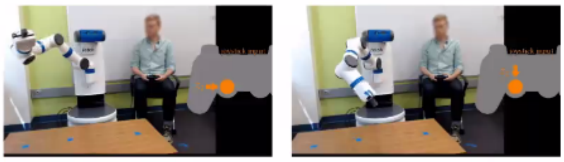
Offline LASER (Transfer Task)
Train Action Space then Train Policy

LASER: Learning a Latent Action Space for Efficient RL, ICRA 2021

Representation: Action Spaces
Application: Assistive Teleoperation

Offline, personal demonstrations of high-dimensional motions



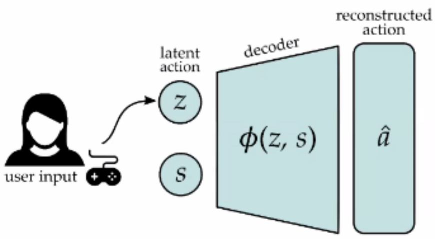


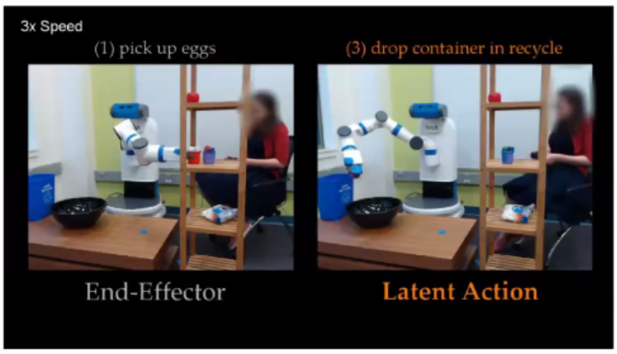
Learn low-dimensional latent representations for online control

Easier to control high-dimensional robots by embedding the robot's actions into a low-dimensional latent space

Controlling Assistive Robots with Learned Latent Actions, ICRA 2020

Representation: Action Spaces
Application: Assistive Teleoperation



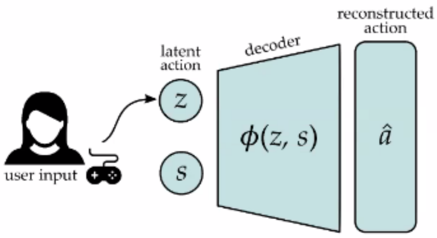


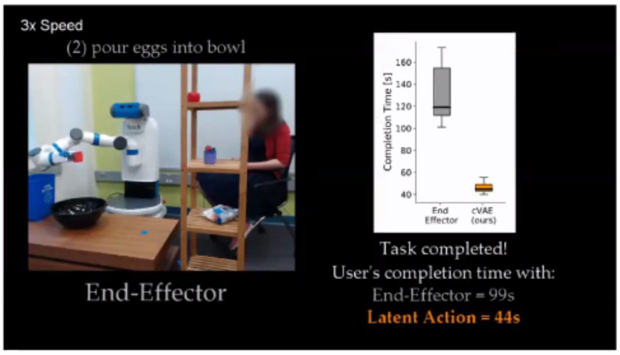
Easier to control high-dimensional robots by embedding the robot's actions into a low-dimensional latent space

Controlling Assistive Robots with Learned Latent Actions, ICRA 2020

Representation: Action Spaces

Application: Assistive Teleoperation





Easier to control high-dimensional robots by embedding the robot's actions into a low-dimensional latent space

Controlling Assistive Robots with Learned Latent Actions, ICRA 2020

Structure

State/Action Reps.
VICES IROS19
LASER ICRA21
Making Sense ICRA19
Unsup KPs PAMI21

Inductive Biases
C-Learning ICLR21
OCEAN UAI20
D2RL arXiv20

Structure in Planning
CAVIN CORL20
Skill Hierarchy ICLR21
Finding-IT, CVPR18

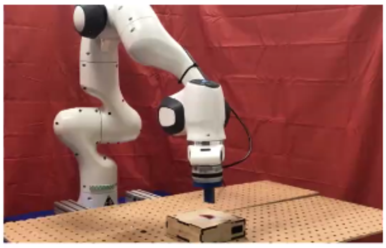
Neural Programming
NTP ICRA18
NTG CVPR19
Cont.Relax IROS19

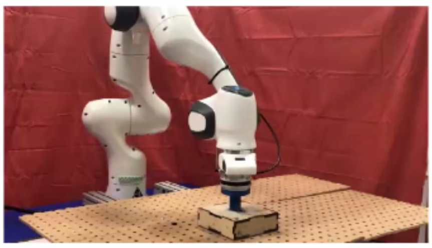
Representations for Robot Learning

How can better state representations capture multimodal data?

Generalizable Multi-modal State Representations

- Learn a joint Visuo-Tactile representation for Peg Transfer
- Representation transfers to new task, while Policy doesn't





Making Sense of Vision and Touch, ICRA 2019 T2020

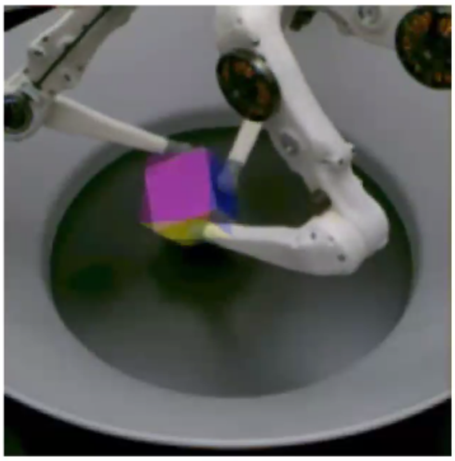
Representation: State Spaces

Application: Multi-finger Manip

Real Robot Challenge: Trifinger platforms

Task: repose in 6-DoF
(position + orientation)

Development done remotely in simulation using Isaac Gym, no physical robot access

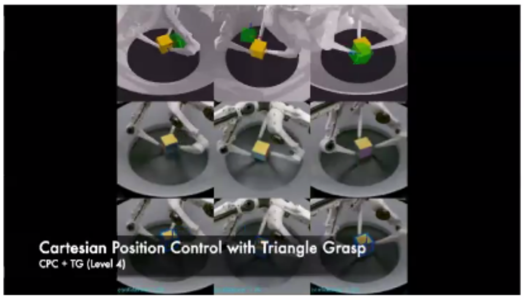


Easier to control high-dimensional robots with better action space and a efficient physics sim

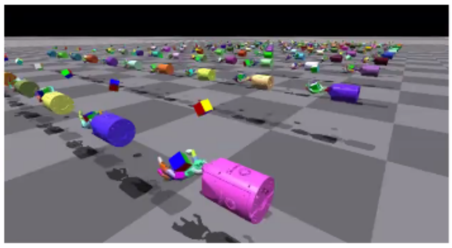
CORL2021

Background

Dextrous Manipulation via Simulation
OpenAI, 2018



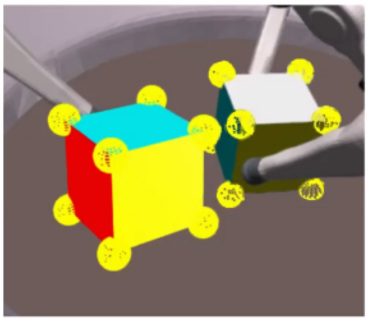
Real Robot Challenge
Structured Policies

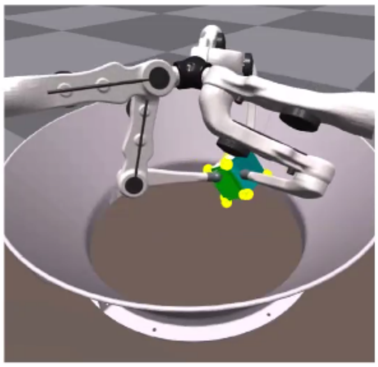


GPU-Simulated Manipulation
Isaac Gym

Reward and Pose Representation

- Traditional reward & observation performed poorly
- A better representation than position + quaternion in {observation, reward}?
- Allow for 6-DoF reposing

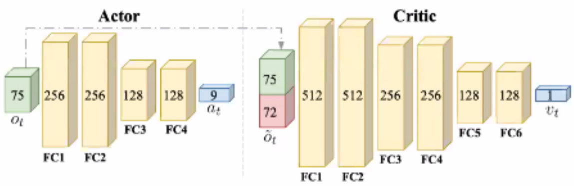




Representation: Action Spaces

Application: Multi-finger Manip

Asymmetric Actor-Critic



Available in sim or real

- Policy Observations:
  - Joint Position
  - Joint Velocity
  - Object Pose

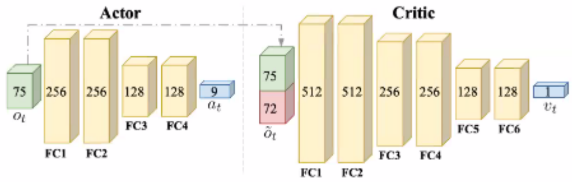
Sim-only

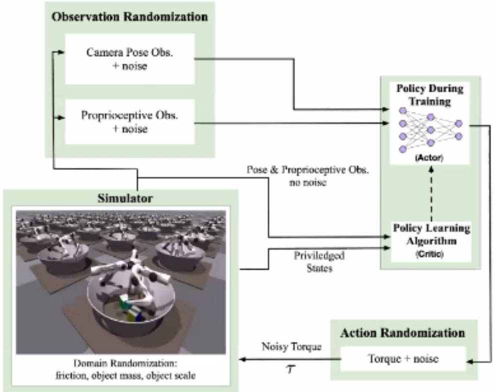
- Value function Observations:
  - Policy Observations
  - Joint Torques
  - Fingertip Positions
  - Fingertip Forces
  - Object Velocity

Easier to control high-dimensional robots by embedding the robot's actions into a low-dimensional latent space

Representation: Action Spaces
Application: Multi-finger Manip

Asymmetric Actor-Critic

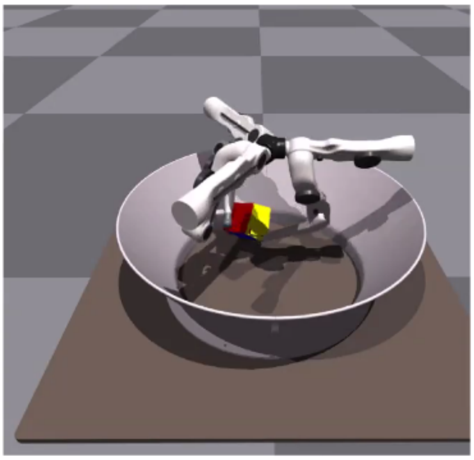


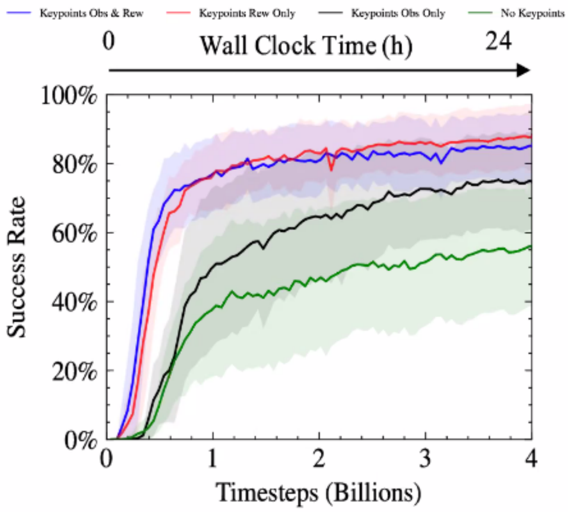


Easier to control high-dimensional robots by embedding the robot's actions into a low-dimensional latent space

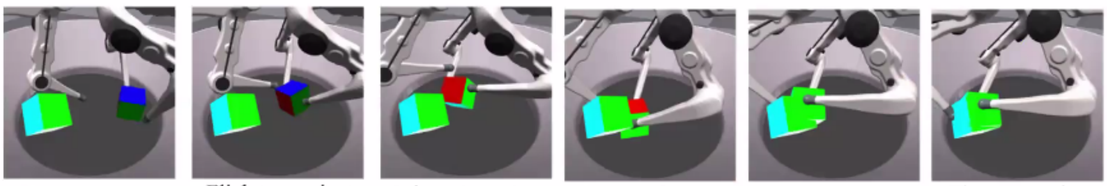
Representation: Action Spaces
Application: Multi-finger Manip

- Able to train in <24h on 1 GPU





Representation: Action Spaces
Application: Multi-finger Manip



Initial Grasp | Flick to reorient | 2nd reorientation | Drop & Regrasp | Lift + in-hand reorientation | Fine correction

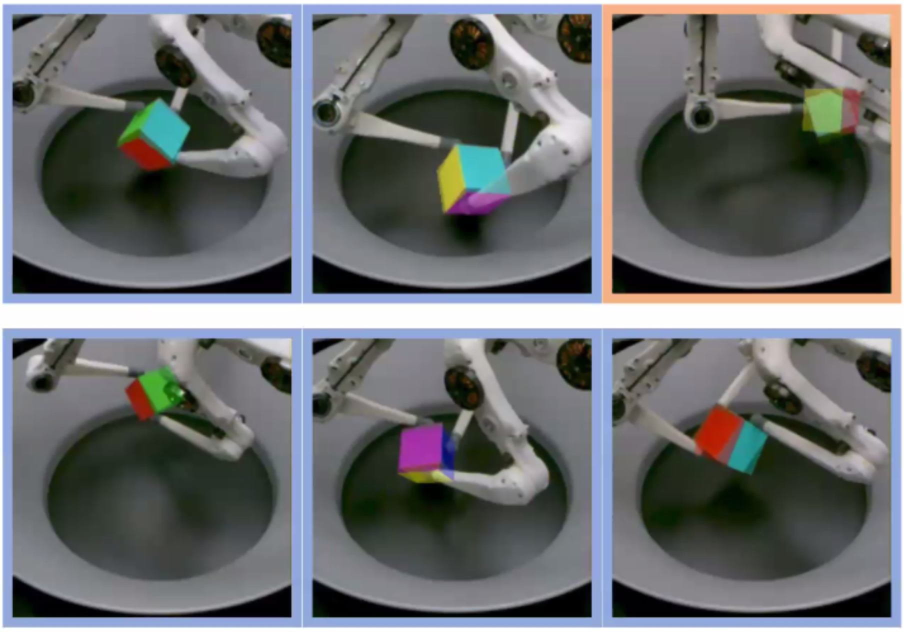
Sim2Real
Results

Robotics as a Service

No physical robot access

83% Success

(Real Time Videos)



Structure

State/Action Reps.
VICES IROS19
LASER ICRA21
Making Sense ICRA19
Unsup KPs PAMI21

Inductive Biases
C-Learning ICLR21
OCEAN UAI20
D2RL arXiv20

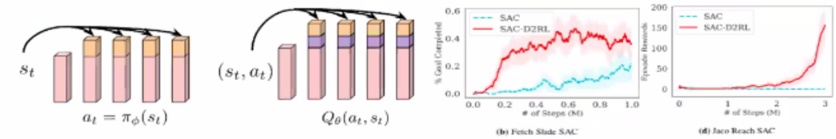
Structure in Planning
CAVIN CORL20
Skill Hierarchy ICLR21
Finding-IT, CVPR18

Neural Programming
NTP ICRA18
NTG CVPR19
Cont.Relax IROS19

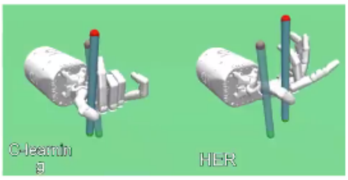
Representations for Robot Learning

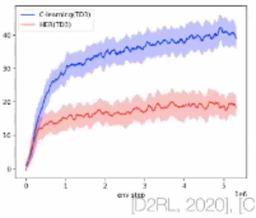
How can better architectures and learning rules lead to generalization in Robotics?

I. Using Dense connections in Policy/Value improves sample efficiency



II. Learning Cumulative Accessibility $C(s, a, h)$ is better than $Q(s, a)$
Can represent multimodal horizon aware solutions as well as reachability





[D2RL, 2020], [C-Learning, ICLR 2021]

Structure

State/Action Reps.
VICES IROS19
LASER ICRA21
Making Sense ICRA19
Unsup KPs PAMI21

Inductive Biases
C-Learning ICLR21
OCEAN UAI20
D2RL arXiv20

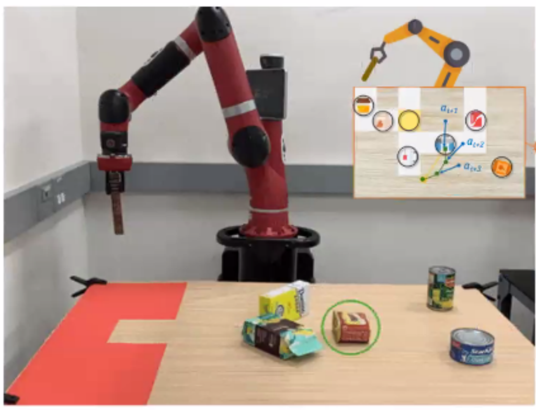
Structure in Planning
CAVIN CORL20
Skill Hierarchy ICLR21
Finding-IT, CVPR18

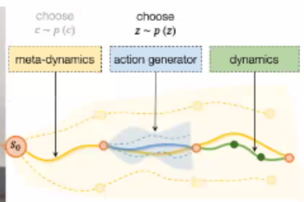
Neural Programming
NTP ICRA18
NTG CVPR19
Cont.Relax IROS19

Representations for Planning

What model structure enables longer term planning?

Hierarchical structure helps in long term discrete and continuous planning





CAVIN, CoRL 2019

Structure

State/Action Reps.
VICES IROS19
LASER ICRA21
Making Sense ICRA19
Unsup KPs PAMI21

Inductive Biases
C-Learning ICLR21
OCEAN UAI20
D2RL arXiv20

Structure in Planning
CAVIN CORL20
Skill Hierarchy ICLR21
Finding-IT, CVPR18

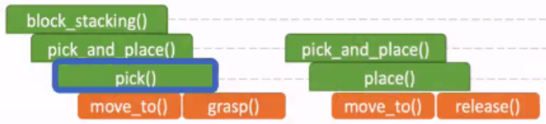
Neural Programming
NTP ICRA18
NTG CVPR19
Cont.Relax IROS19

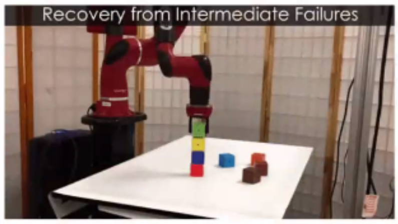
Representations for Planning

Hierarchy

What model structure enables longer term planning?

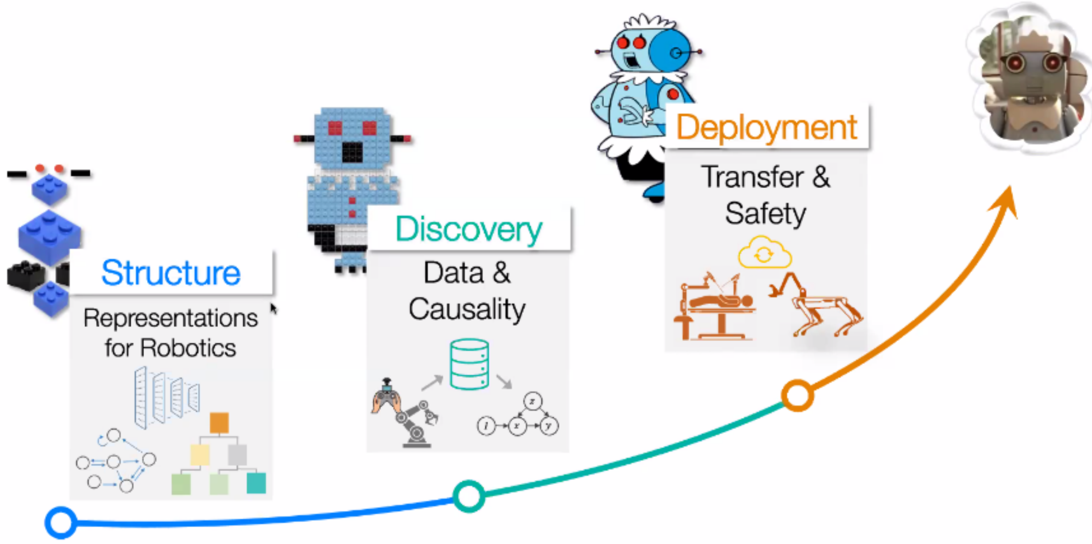
Program Induction provides a very efficient model of compositional generalization





NTP, ICRA 2018, NTG, CVPR 2019

Building blocks of Generalizable Autonomy in Robotics



Data in Robotics

✗ Short-Horizon skills
✗ Skill Specific learning

✗ Platform dependent data
✗ Scaling to other skills ?

✗ Small datasets (minutes)
✗ Low diversity ?

Manipulation

Mason & Salisbury 1985
Srinivasa et al 2010
Berenson 2013
Odhner1 et al 2014
Chavan-Dafle et al 2014
Yamaguchi, et. al, 2015
...

Li , Allen et al. 2015
Yahya et al, 2016
Schenck et al. 2017
Mar et al. 2017
Laskey et al 2017
Quispe et al 2018
...

Grasping

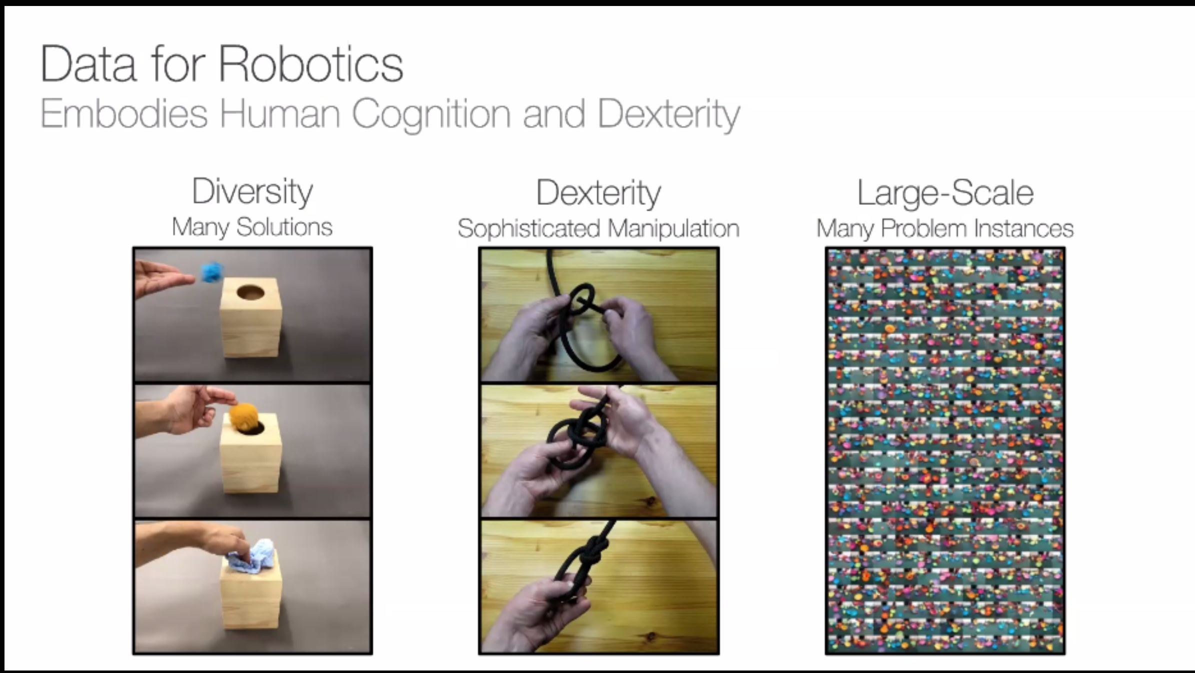
Mishra et al 1987
Ferrari & Canny, 1992
Ciocarlie & Allen, 2009
Dogar & Srinivasa, 2011
Rodriguez et al. 2012
Bohg et al 2014

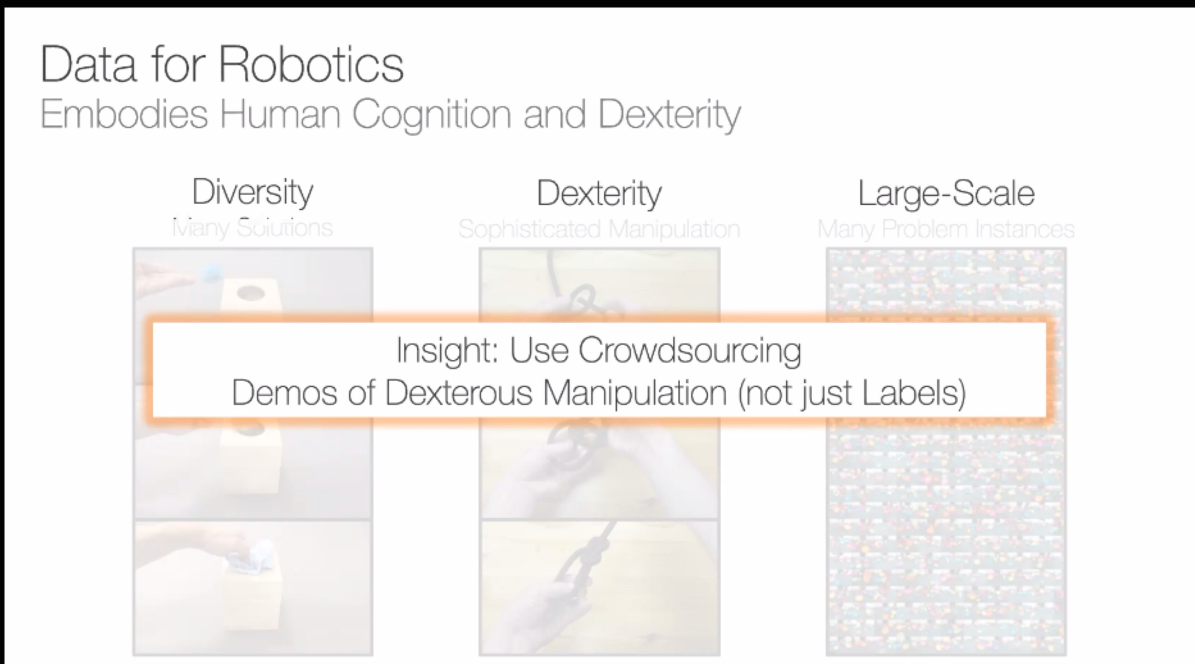
Pinto & Gupta, 2016
Levine et al 2016
Mahler et al 2017
Jang et al 2017
Viereck et al 2017
...

Imitation

Abbeel et al, 2004
Ratliff et al 2006,
Ziebart et al, 2009
Argall et al, 2009,
Boularias et al., 2011
Montfort et al 2015,
Wulfmeier et al 2015,

Krishnan et al 2017
Finn et al. 2017
Vecerik et al. 2017
Rajeswaran et al 2018
Zhu et al 2018
Ravichandar et al 2020...

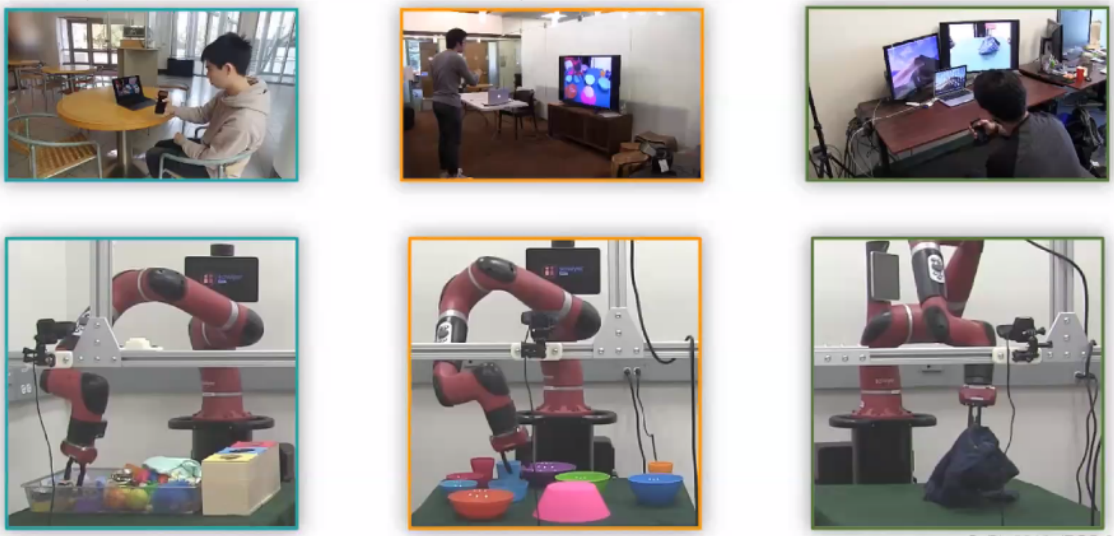

Data for Robotics
Embodies Human Cognition and Dexterity
Diversity
Many Solutions
Dexterity
Sophisticated Manipulation
Large-Scale
Many Problem Instances


Data for Robotics
Embodies Human Cognition and Dexterity
Diversity
Many Solutions
Dexterity
Sophisticated Manipulation
Large-Scale
Many Problem Instances
Insight: Use Crowdsourcing
Demos of Dexterous Manipulation (not just Labels)


RoboTurk: Dexterous Data Collection
Operator specifies full 6-DoF motion of the arm by moving their phone.

Roboturk Scalability

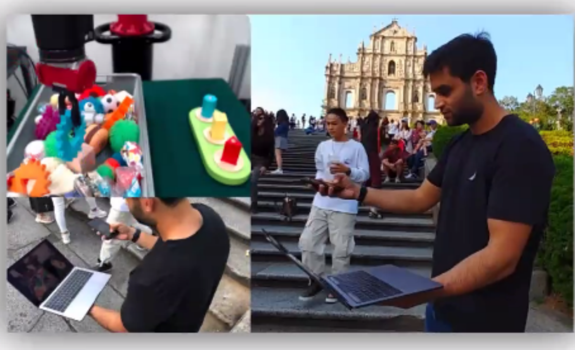
Multiple Simultaneous Teleoperation Connections



CoRL 2018, IROS 2019

Teleoperation across the world with RoboTurk

RoboTurk in the Swiss Alps
at ~7000 ft



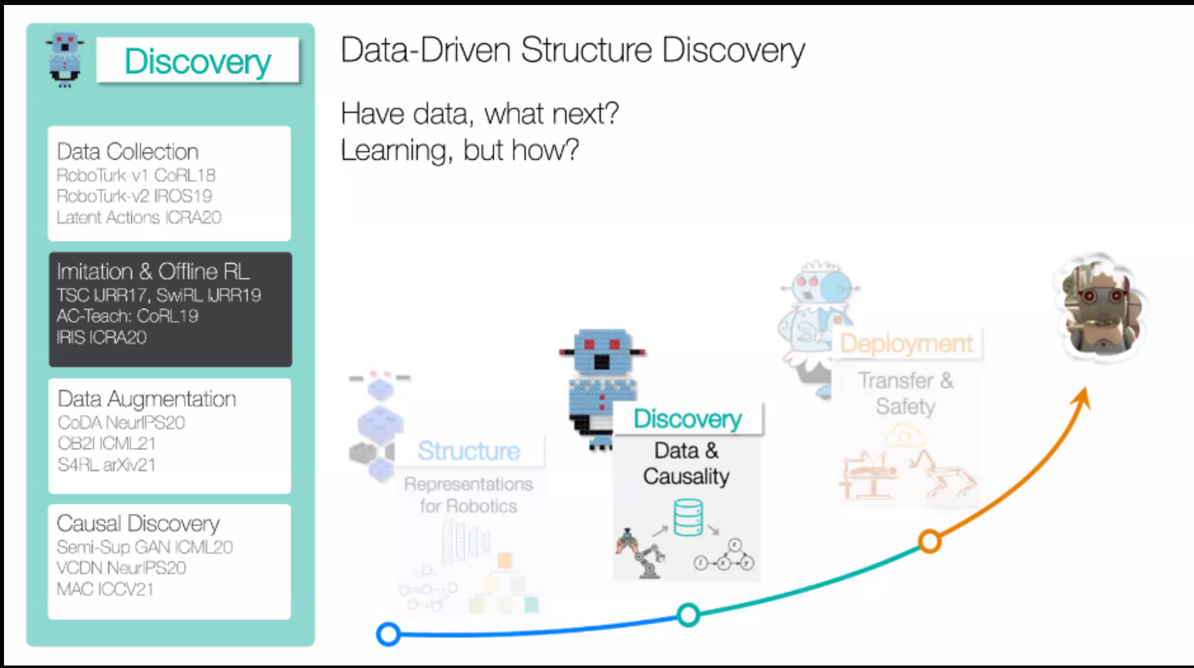
Real-Time Teleoperation from Macau to
Stanford – over 11,000 km!

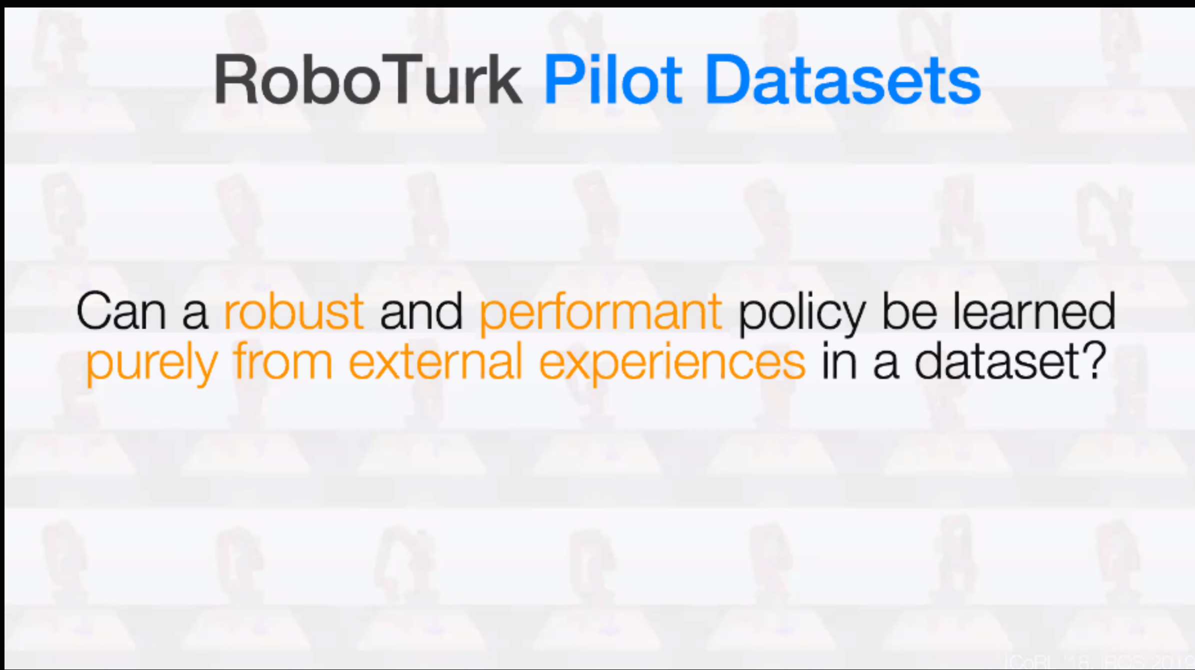
CoRL 2018, IROS 2019

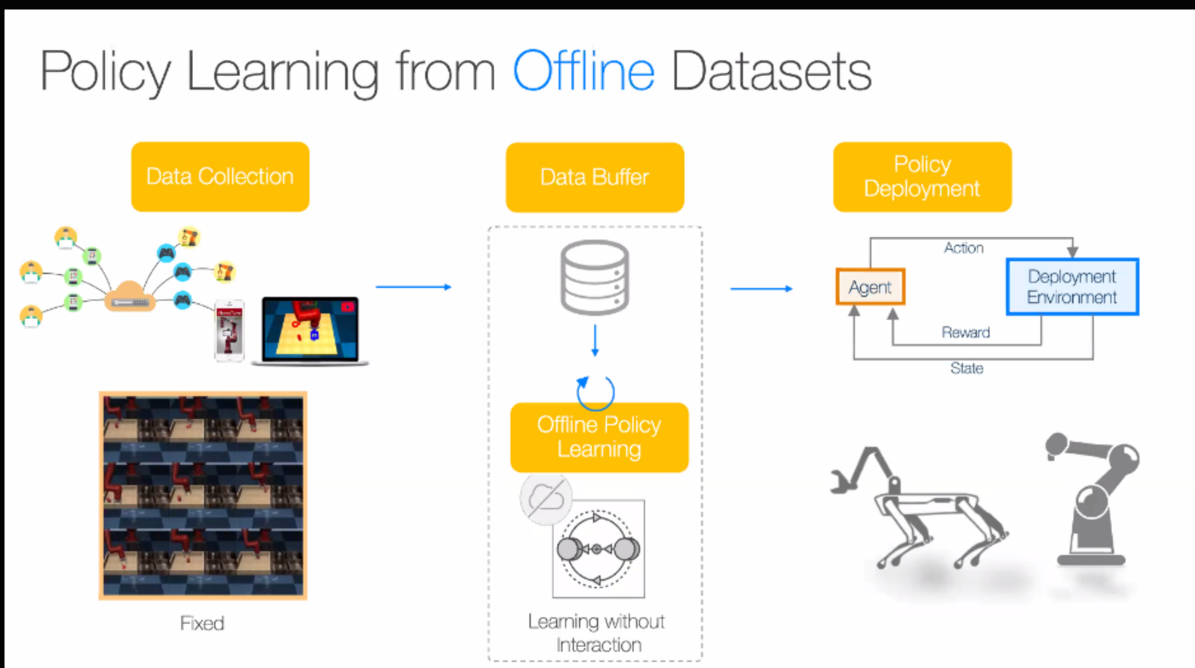
RoboTurk Pilot Datasets

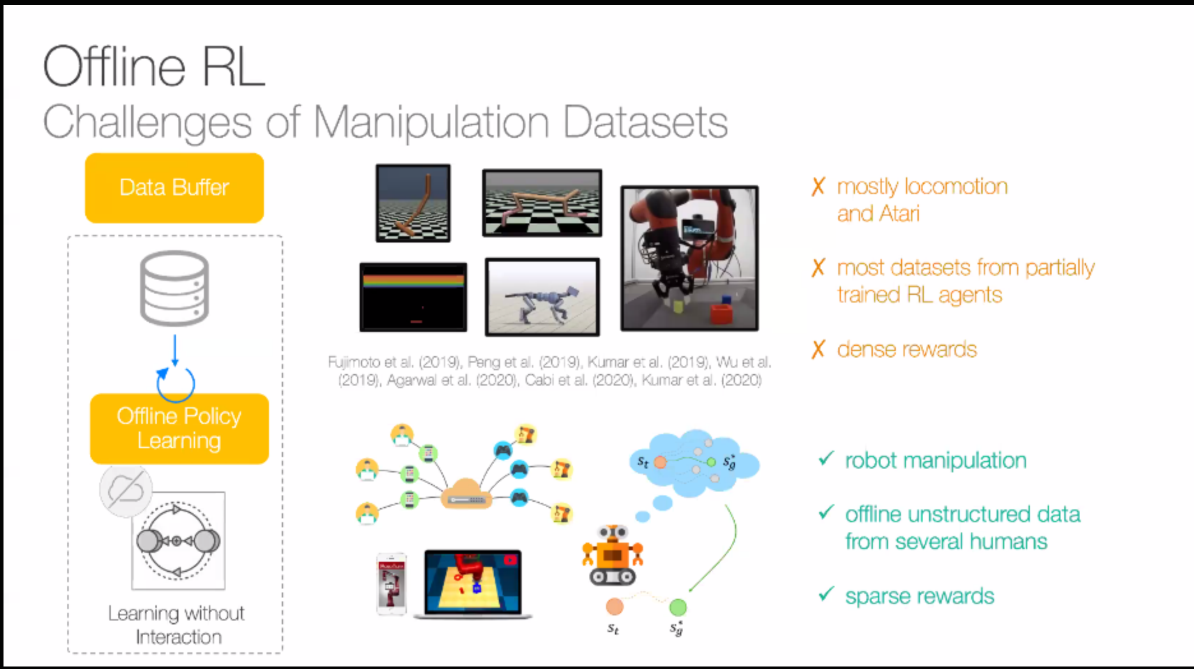
Simulated Data	Real Robot Data
137.5 hours of demonstrations	111 hours of robot demos
22 hours of total platform usage	1 week of data collection
3 dexterous manipulation tasks	3 dexterous manipulation tasks
3224 total attempted demos	2144 total demonstrations
15 novice, remote users	54 non-expert users

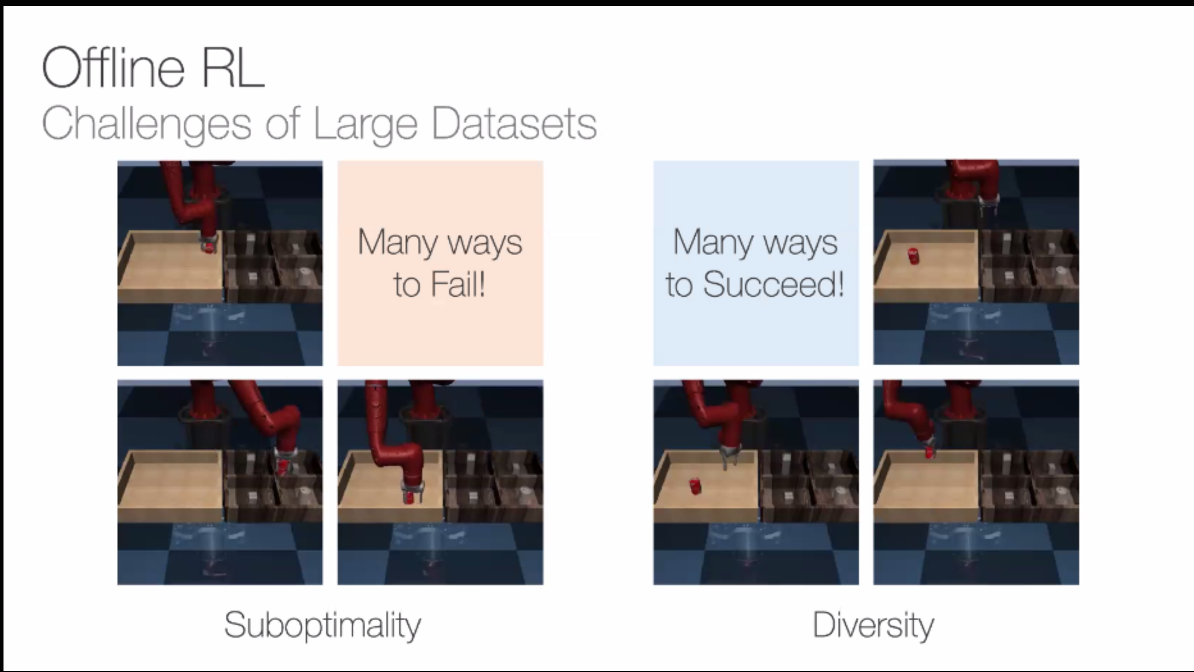
CoRL 2018, IROS 2019


Discovery
Data Collection
RoboTurk v1 CoRL18
RoboTurk-v2 IROS19
Latent Actions ICRA20
Imitation & Offline RL
TSC IJRR17, SwiRL IJRR19
AC-Teach: CoRL19
IRIS ICRA20
Data Augmentation
CoDA NeurIPS20
CB2I ICML21
S4RL arXiv21
Causal Discovery
Semi-Sup GAN ICML20
VCDN NeurIPS20
MAC ICCV21
Data-Driven Structure Discovery
Have data, what next?
Learning, but how?
Structure
Representations for Robotics
Discovery
Data & Causality
Deployment
Transfer & Safety


RoboTurk Pilot Datasets
Can a robust and performant policy be learned purely from external experiences in a dataset?

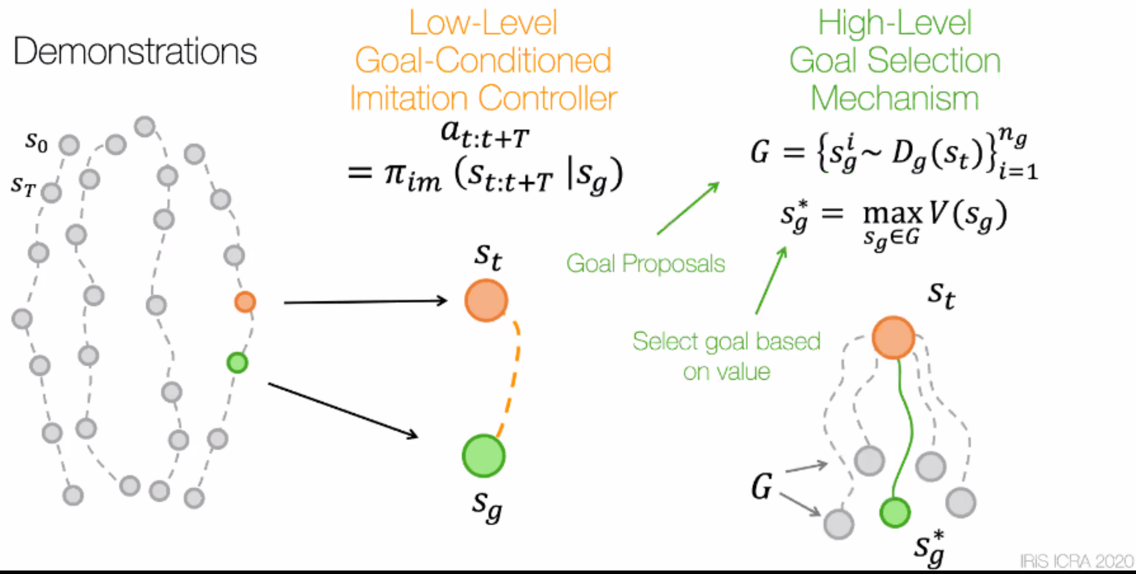

Policy Learning from Offline Datasets
Data Collection
Data Buffer
Policy Deployment
Offline Policy Learning
Learning without Interaction
Action
Agent
Deployment Environment
Reward
State
Fixed


Offline RL
Challenges of Manipulation Datasets
Data Buffer
Offline Policy Learning
Learning without Interaction
Fujimoto et al. (2019), Peng et al. (2019), Kumar et al. (2019), Wu et al. (2019), Agarwal et al. (2020), Cabi et al. (2020), Kumar et al. (2020)
✗ mostly locomotion and Atari
✗ most datasets from partially trained RL agents
✗ dense rewards
✓ robot manipulation
✓ offline unstructured data from several humans
✓ sparse rewards


Offline RL
Challenges of Large Datasets
Many ways to Fail!
Many ways to Succeed!
Suboptimality
Diversity


Offline RL
Challenges of Large Datasets
Many ways to Fail!
Many ways to Succeed!
Hence vanilla supervised learning does not work
Suboptimality
Diversity

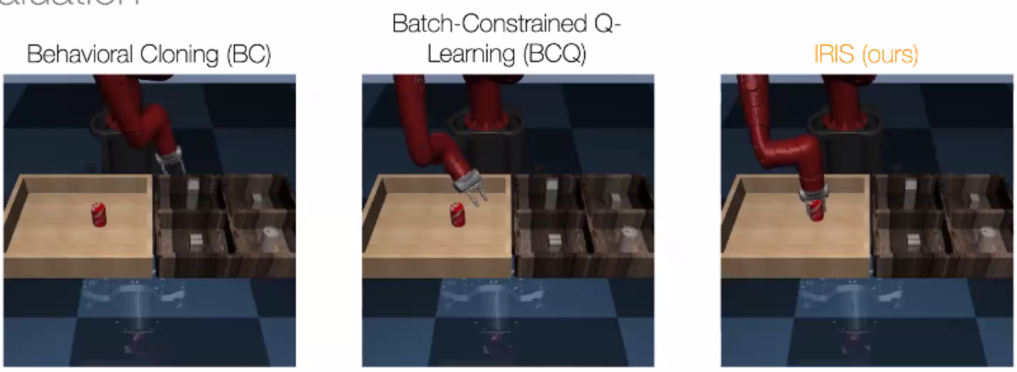
Offline RL: IRIS



IRIS ICRA 2020

Offline RL: IRIS

Evaluation



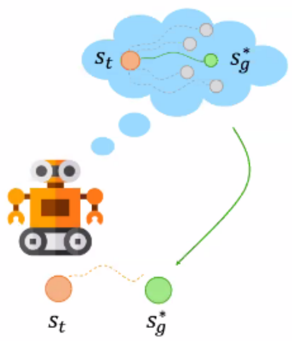
Unable to imitate reasonable behavior from dataset since models do not account for diverse data.

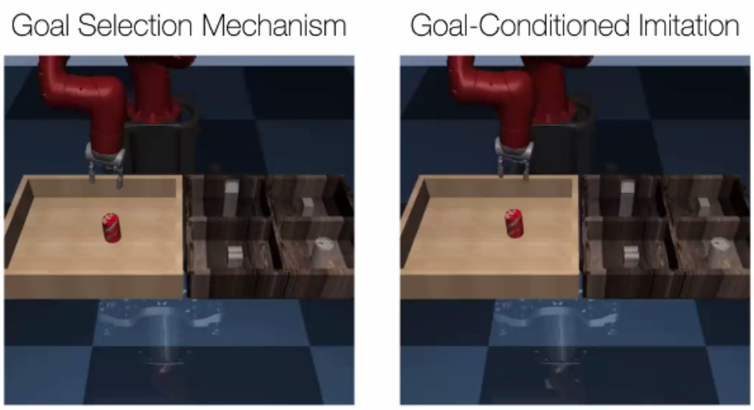
Goal-conditioned imitation is critical for learning from diverse dataset.

IRIS ICRA 2020

Offline RL: IRIS

Evaluation



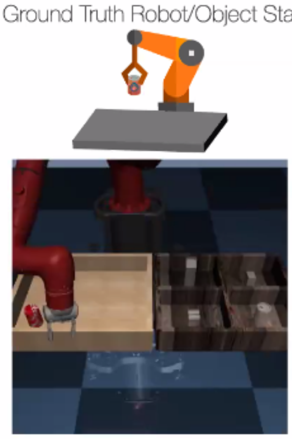


The low-level controller reaches the goals set by the high-level to accomplish the task.

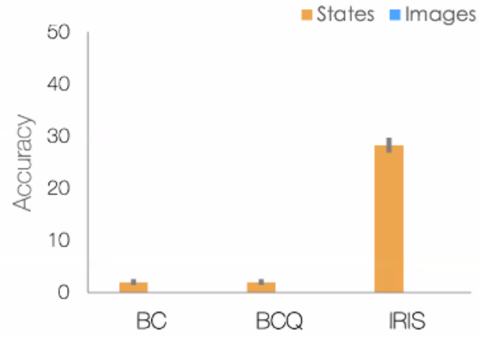
IRIS ICRA 2020

Offline RL: IRIS
Evaluation

Ground Truth Robot/Object States



IRIS with States



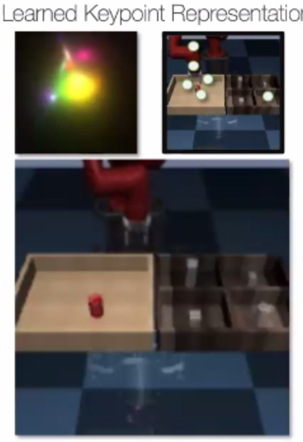
IRIS ICRA 2020

Offline RL: IRIS
Evaluation

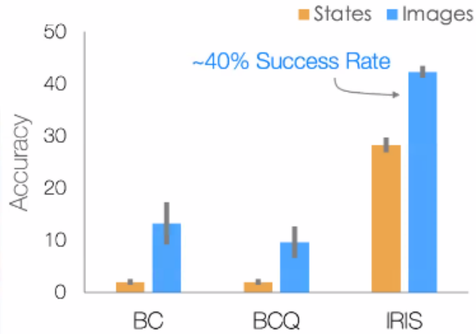
Ground Truth Robot/Object States

IRIS with States

Learned Keypoint Representations



IRIS with Images



IRIS ICRA 2020

Discovery

Data Collection
RoboTurk-v1 CoRL18
RoboTurk-v2 IROS19
Latent Actions ICRA20

Imitation & Offline RL
TSC IJRR17, SWIRL IJRR19
AC-Teach: CoRL19
IRIS ICRA20

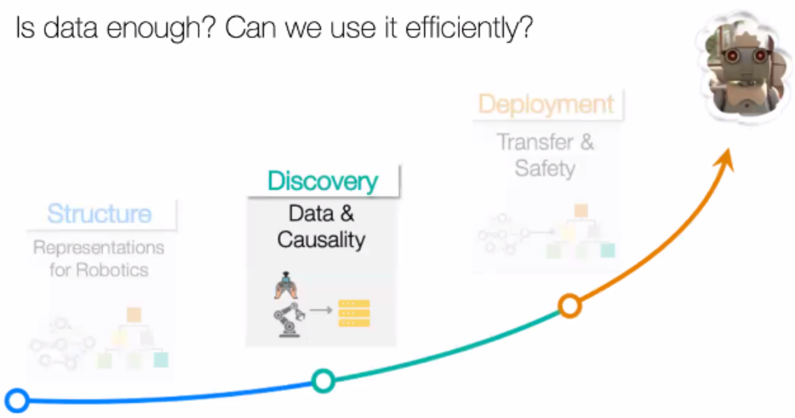
Data Augmentation
CoDA NeurIPS20
OB2I ICML21
S4RL arXiv21

Causal Discovery
Semi-Sup GAN ICML20
VCDN NeurIPS20
MAC ICCV21

Data-Driven Structure Discovery

Collect Structured Dexterous Data
Learn with Offline RL

Is data enough? Can we use it efficiently?



Data Augmentation in RL
Do more with the same data

Left Arm Pick and Place

Right Arm Pick and Place

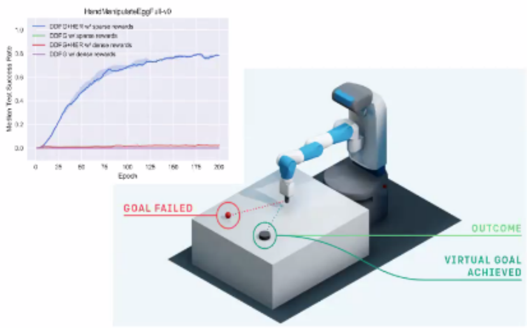
Which of the following is possible (only based on observed data)

Independent
Compositional Generalization

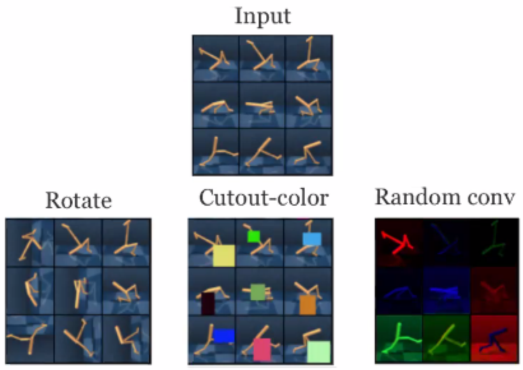
Not-Independent (!)
Hence need evidence of possibility

Data Augmentation in RL
How to do this Algorithmically

- Substantial performance boosts!



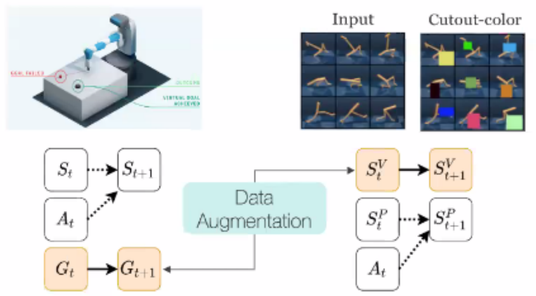
Goal relabeling (e.g, HER])



Visual Input relabeling (e.g, RAD])

RAD: Laskin et al 2020, DrQ: Kostrikov et al 2020, CURL: Laskin et al. 2020, S4RL: Sinha et al. 2021, HER: Andrychowicz et al 2018

Data Augmentation in RL
Unified View

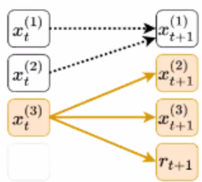


Goal is independent of State/Action Dynamics

Visual characteristics (e.g., crop) are independent of physical dynamics

Counterfactual reasoning to generate new, causally valid (counterfactual) data!

Generic CoDA



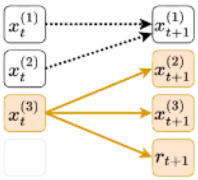
Given two independent mechanisms, Relabel one (conditional independence!)

RAD: Laskin et al 2020, DrQ: Kostrikov et al 2020, CURL: Laskin et al. 2020, S4RL: Sinha et al. 2021, HER: Andrychowicz et al 2018

Counterfactual Data Augmentation

Counterfactual reasoning to generate new, causally valid (counterfactual) data!

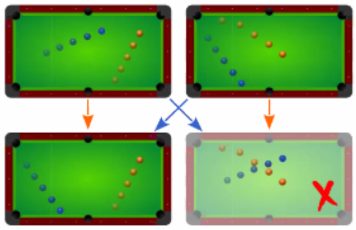
Generic CoDA



Given two independent mechanisms, Relabel one (conditional independence!)

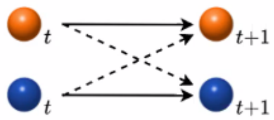
✓ Model-Free relabelling

✗ But Causal Independence is not Global

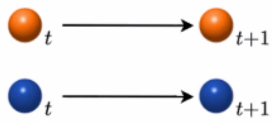


- For the most part, balls behave independently, and we can use CoDA
- But balls are not always independent, so this can also produce nonsense

Counterfactual Data Augmentation
Local Causal Model



Global Model



Local Model

$$\mathcal{M}_t = \langle V_t, U_t, \mathcal{F} \rangle \xrightarrow{\text{Condition on } (s_t, a_t) \in \mathcal{L}} \mathcal{M}_t^{\mathcal{L}} = \langle V_t^{\mathcal{L}}, U_t^{\mathcal{L}}, \mathcal{F}^{\mathcal{L}} \rangle$$

Structural Causal Model (SCM) that marginalizes across all possible transitions

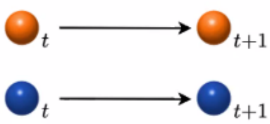
Local Causal Model (LCM) that behaves like the global SCM in local subspace $\mathcal{L}$

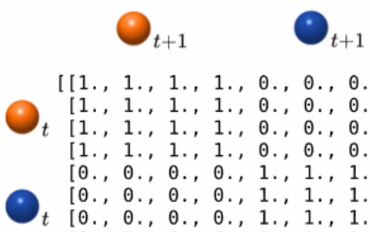
Counterfactual Data Augmentation
Learning Local Causal Model

- Input: 2 balls, each with 4 features: $[x, y, \dot{x}, \dot{y}]$

```
[[ 1.23, -0.73,  1.31,  1.07],
 [-0.6 ,  2.51, -1.51, -0.89]]
```

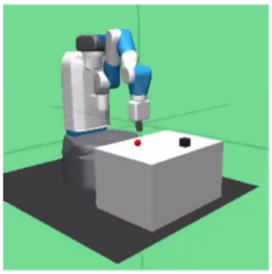
- Output: Adjacency matrix M of the causal graph (between x_t and x_{t+1})
- (intuition) M: the input-output Jacobian is non-zero

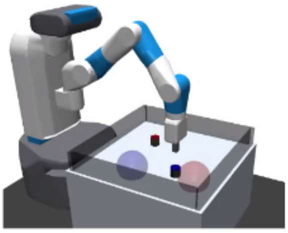


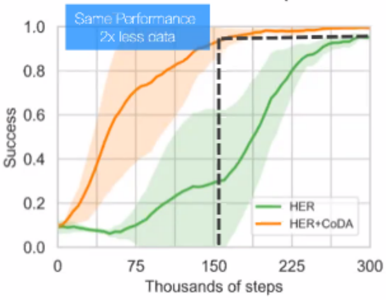


```
[[1., 1., 1., 1., 0., 0., 0., 0.],
 [1., 1., 1., 1., 0., 0., 0., 0.],
 [1., 1., 1., 1., 0., 0., 0., 0.],
 [1., 1., 1., 1., 0., 0., 0., 0.],
 [0., 0., 0., 0., 1., 1., 1., 1.],
 [0., 0., 0., 0., 1., 1., 1., 1.],
 [0., 0., 0., 0., 1., 1., 1., 1.],
 [0., 0., 0., 0., 1., 1., 1., 1.]]
```

CoDA: Goal-Conditioned (Online) RL

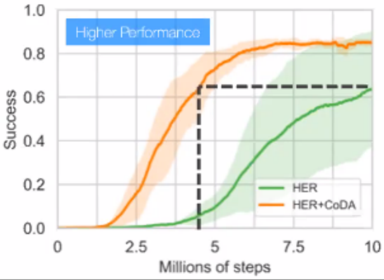






Fetch-Push-v1
state space: [Robot and 1 object]

CoDA Heuristic for independence:
Obj ⫫ Gripper | if $|x_g - x_o| \geq 10\ cm$



Fetch-Slide2
state space: [Robot and 2 objects]
Harder task (30x more samples)!

CoDA NeurIPS 2020

Discovery

Data Collection
RoboTurk-v1 CoRL18
RoboTurk-v2 IROS19
Latent Actions ICRA20

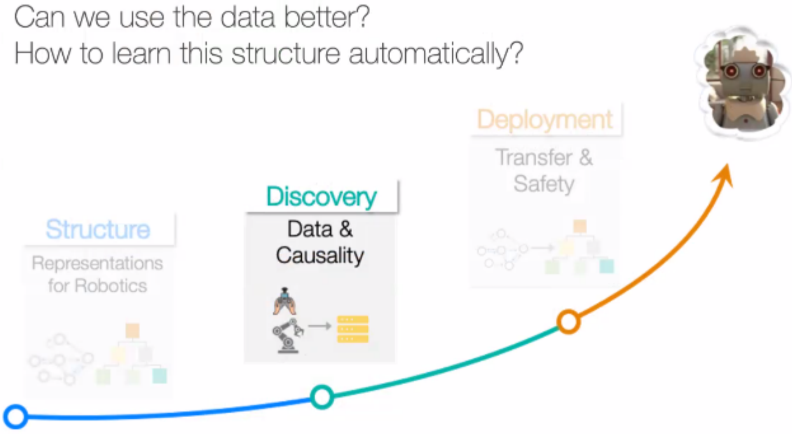
Imitation & Offline RL
TSC IJRR17, SwiRL IJRR19
AC-Teach: CoRL19
IRIS ICRA20

Data Augmentation
CoDA NeurIPS20
OB2I ICML21
S4RL arXiv21

Causal Discovery
Semi-Sup GAN ICML20
VCDN NeurIPS20
MAC ICCV21

Data for Robotics
How to collect large-scale structured supervision and then learn from it?

Can we use the data better?
How to learn this structure automatically?



Causal Discovery

Guess Model

1-Ball bouncing on Flat surface with no damping

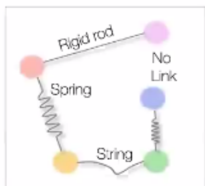
3-Balls (different) in a box with no damping

5-Balls (different) in a box with no damping

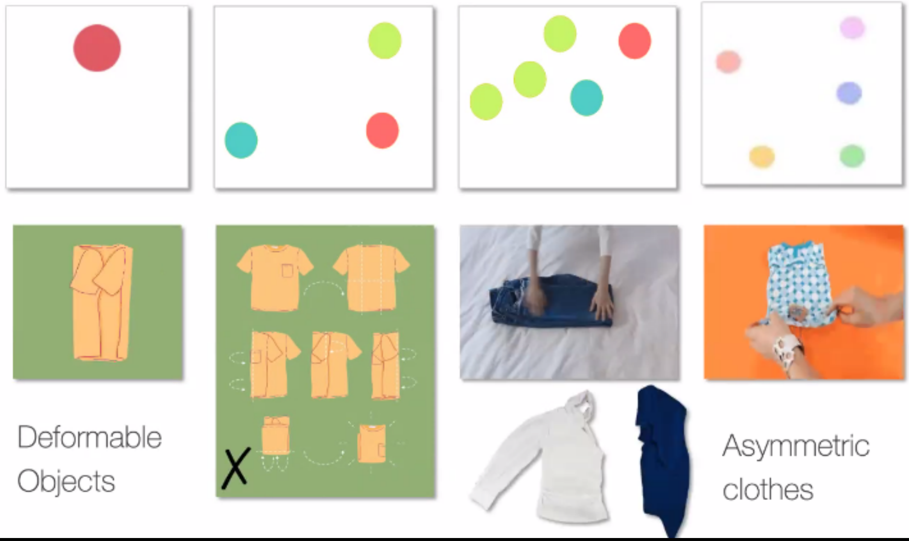
5-Balls (different) in a box with unknown connections

Model is tricky to Guess
- Latent Mechanisms impose constraints

Can't Memorize, need to Generalize to new latent graphs
- Each Test case may be a different graph



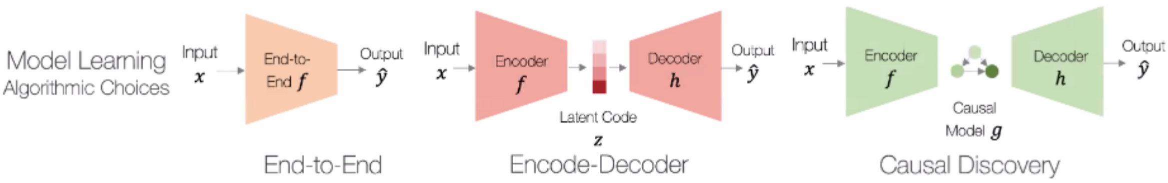
Causal Discovery



Causal Discovery of Dynamics

Learning Latent Generative Dynamics Model

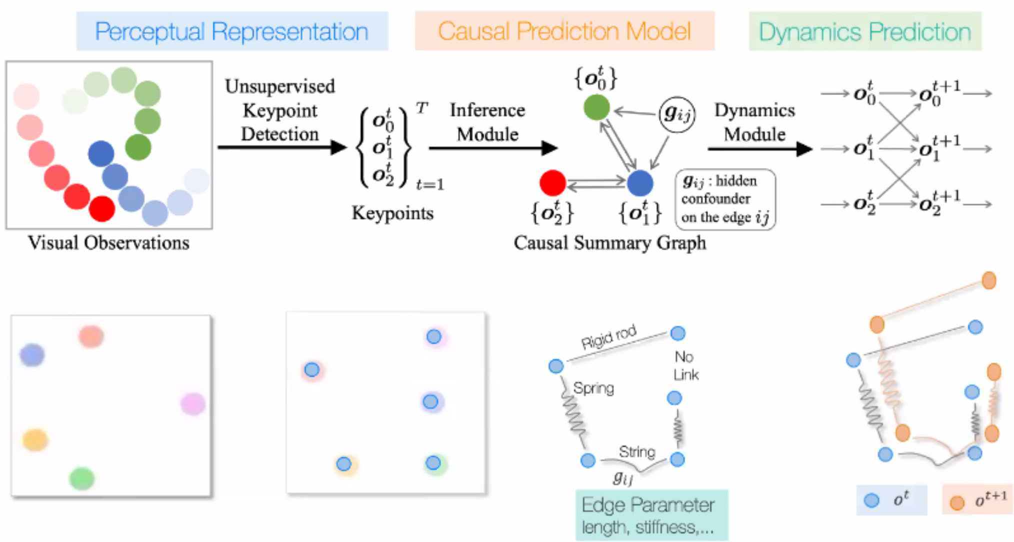
Output: Single model to multiple latent generative graphs



	End-to-End	Encode-Decoder	Causal Discovery
Likelihood	✓	✓	✓
Disentanglement	✗	✓	✓
Generalization	✗	✓/✗	✓
Learning to Learn	✗	✗	✓

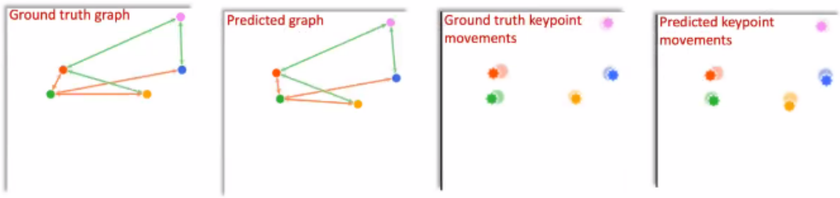
VCDN NeurIPS 2020

Visual Causal Discovery Network

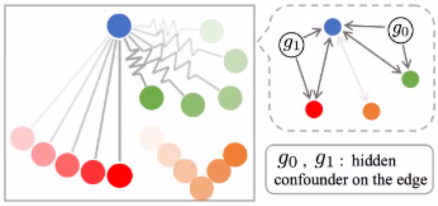


VCDN NeurIPS 2020

Visual Causal Discovery Network
Evaluation



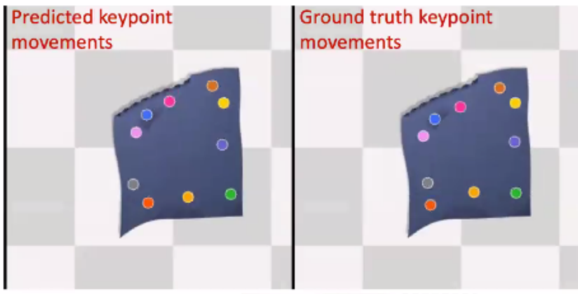
Balls Environment
- N-balls with different edge types {spring, rigid, null}
- Each edge has random parameters
- Dataset: 4800 Training Episodes with random latent generative model



Visual Causal Discovery Network
Evaluation

Predicted graph

No Absolute Comparable Ground Truth

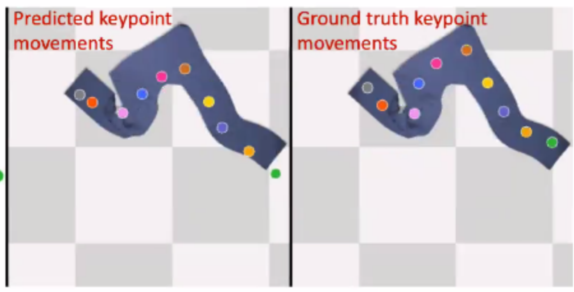


Fabrics Environment
- 3 Types: Shirts, Towels, Pants
- Each instance has variable geometry
- Dataset: 1800 Training Episodes with random latent generative model

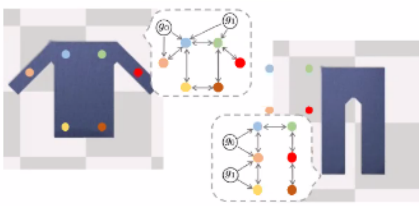
Visual Causal Discovery Network
Evaluation

Predicted graph

No Absolute Comparable Ground Truth

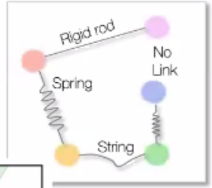


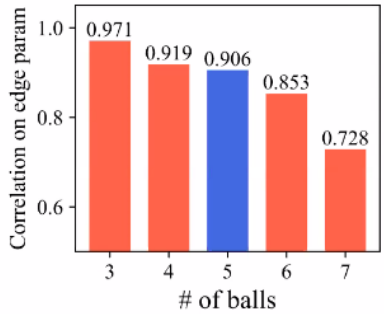
Fabrics Environment
- 3 Types: Shirts, Towels, Pants
- Each instance has variable geometry
- Dataset: 1800 Training Episodes with random latent generative model



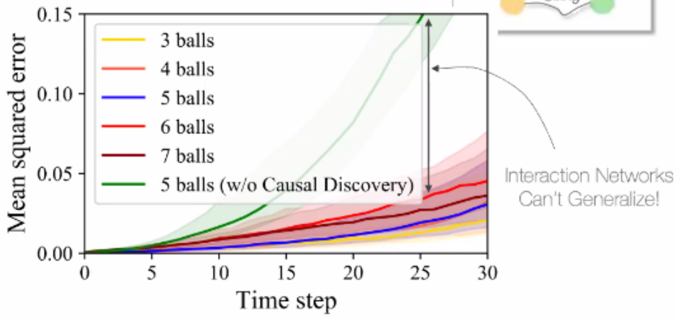
Visual Causal Discovery Network

Extrapolation Generalization





Correlation on the rest length of the spring relation
(No Ground Truth for training)

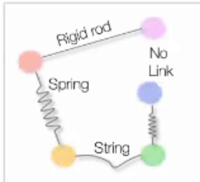


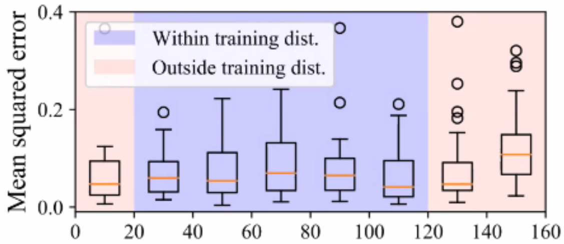
Mean squared error on future prediction

VCDN NeurIPS 2020

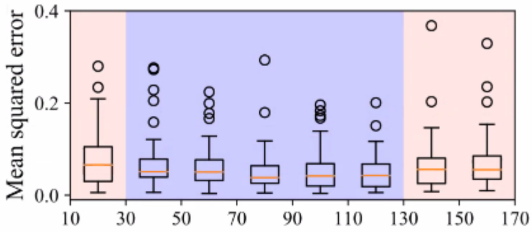
Visual Causal Discovery Network

Counterfactual Generalization





Intervention: Length of Rigid Rod



Intervention: Rest length of Spring

Change specific underlying parameters for given datapoints

VCDN NeurIPS 2020

Discovery

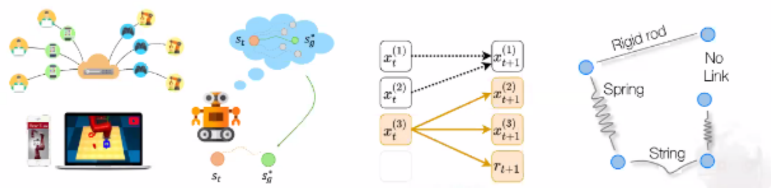
Data Collection
RoboTurk-v1 CoRL18
RoboTurk-v2 IROS19
Latent Actions ICRA20

Imitation & Offline RL
TSC IJRR17, TSC-DL ICRA17
SwiRL IJRR19
AC-Teach: CoRL19
IRIS ICRA20

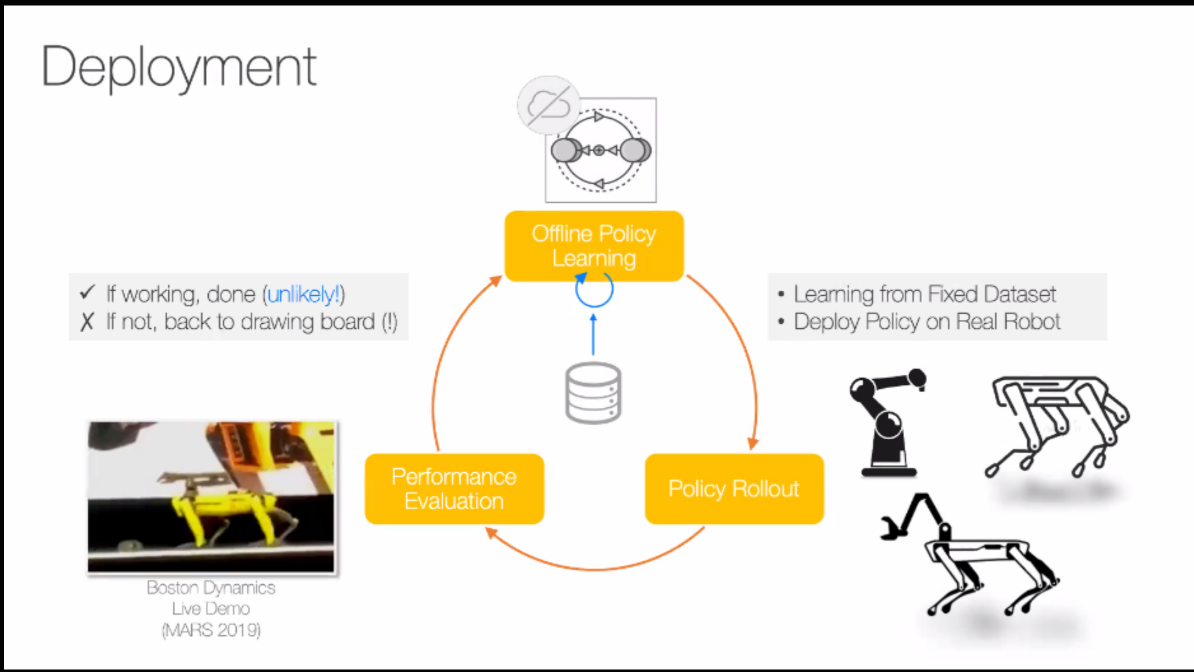
Data Augmentation
CoDA NeurIPS20
CB2I ICML21
S4RL arXiv21

Causal Discovery
Semi-Sup GAN ICML20
VCDN NeurIPS20
MAC ICCV21

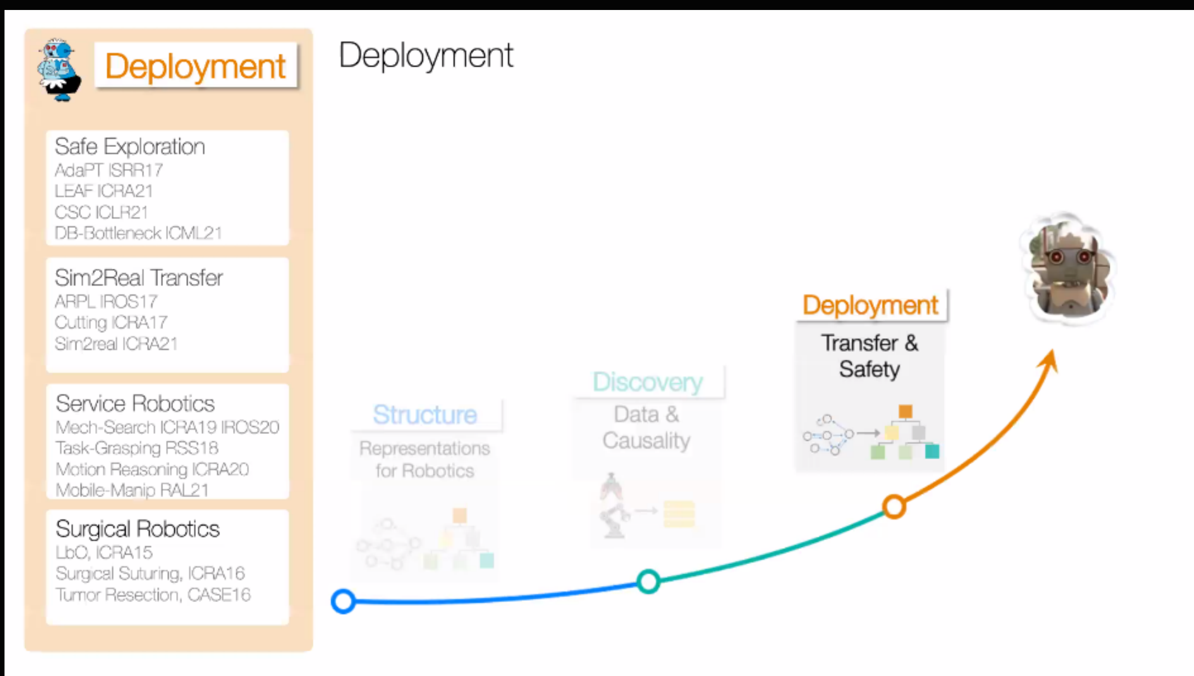
Data-Driven Structure Discovery



Structured-supervision through scalable crowdsourcing
Causal Discovery is vital for compositional generalization


Deployment
Offline Policy Learning
Performance Evaluation
Policy Rollout
✓ If working, done (unlikely!)
✗ If not, back to drawing board (!)
• Learning from Fixed Dataset
• Deploy Policy on Real Robot
Boston Dynamics
Live Demo
(MARS 2019)

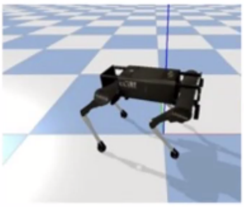

Deployment: Mission Critical Systems
Legged Locomotion
Autonomous Driving
Dexterous Manipulation
Robot Assisted Surgery
Space Robotics
Need to evaluate behavior before deployment and safety during deployment!


Deployment
Safe Exploration
AdaPT ISRR17
LEAF ICRA21
CSC ICLR21
DB-Bottleneck ICML21
Sim2Real Transfer
ARPL IROS17
Cutting ICRA17
Sim2real ICRA21
Service Robotics
Mech-Search ICRA19 IROS20
Task-Grasping RSS18
Motion Reasoning ICRA20
Mobile-Manip RAL21
Surgical Robotics
LbO, ICRA15
Surgical Suturing, ICRA16
Tumor Resection, CASE16
Deployment
Structure
Representations for Robotics
Discovery
Data & Causality
Deployment
Transfer & Safety

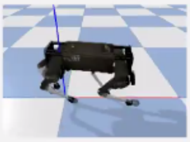
Deployment

Challenge: Safe Exploration

How should an agent learn without Catastrophic Failures during training?



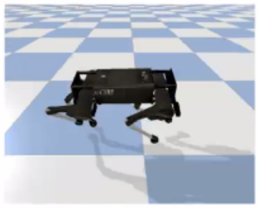
Initial State



Repeated Failures

✗ Costly

✗ Infeasible



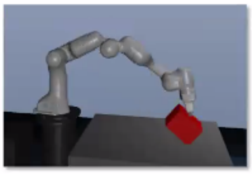
Goal: Run

Deployment

Challenge: Safe Exploration

How should an agent learn without Catastrophic Failures during training?

Respecting constraints of the form $\mathbb{P}(failure) \leq \chi$



Dexterous Manipulation

Objective: Learn to Push
Failure: Glass Topples

Legged Locomotion

Objective: Learn to Run
Failure: Fall Down

Deployment

Challenge: Safe Exploration

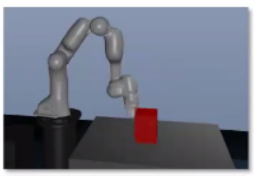
How should an agent learn without Catastrophic Failures during training?

Respecting constraints of the form $\mathbb{P}(failure) \leq \chi$

- Assume a binary constraint function $\mathbb{I}(fail)$ from Environment

$$C: \mathcal{S} \to \{0, 1\}, (C \equiv \mathbb{I}\{failure\})$$

- Analogous to sparse rewards {+1: failure occurs, 0: otherwise}



Dexterous Manipulation

Objective: Learn to Push
Failure: Glass Topples

$C = 1 \; if \; \mathbb{I}(topple)$

Legged Locomotion

Objective: Learn to Run
Failure: Fall Down

$C = 1 \; if \; \mathbb{I}(fall)$

Conservative Safety Critic

Insight: Learning Constraints as Critics

Constrained MDP Objective:

$$\pi^* = \underset{\pi \in \Pi_C}{\operatorname{argmax}}\, V_R^\pi(\mu)$$ Task Objective

$$s.t.\ \mathbb{P}(failure \mid \mu) \le \chi$$ Safety Constraint

$$\mathbb{P}(failure \mid \mu) = \mathbb{E}_{\tau \sim \pi}[\mathbb{I}\{failure\}]$$

$$= \mathbb{E}_{\tau \sim \pi}\left[\sum_{t=0}^{\infty} C(s_t)\right]$$ Expectation over observed Rollouts

$$= V_C^\pi(\mu)$$ Safety Value Function

Conservative Safety Critic

Insight: Learning Constraints as Critics

Constrained MDP Objective:

$$\pi^* = \underset{\pi_\phi}{\operatorname{argmax}}\, \mathbb{E}_{s\sim\rho_\phi, a\sim\pi_\phi}[Q_R^{\pi_\phi}(s,a)]$$

$$s.t.\ \mathbb{E}_{s\sim\rho_\phi, a\sim\pi_\phi}[Q_C(s,a)] \le \chi$$

Reformulation

$$\pi^* = \underset{\pi \in \Pi_C}{\operatorname{argmax}}\, V_R^\pi(\mu)$$ Task Objective

$$s.t.\ \mathbb{P}(failure \mid \mu) \le \chi$$ Safety Constraint

Original Problem

Lagrangian: $\max_{\pi_\phi} \min_{\lambda \ge 0} \mathbb{E}_{s\sim\rho_\phi, a\sim\pi_\phi}[Q_R^{\pi_\phi}(s,a) \ - \lambda(Q_C(s,a) - \chi\ )]$

✓ Solve with Primal-Dual Grad. Descent

✓ Off-policy Value estimates Task $(Q_R(s,a))$ & safety $(Q_C(s,a))$

Safety Critic Algorithm

✗ Incorrect Constraint Value Function $Q_C(s,a)$

Estimation Error

Conservative Safety Critic

Insight-2: Be more Cautious with Overestimation

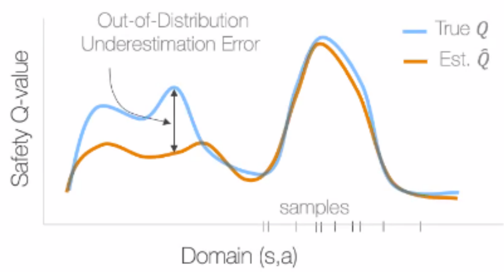
✗ How to fix Estimation error in $Q_C(s,a)$

Q-Learning (review)

$$Q_C(s,a) \leftarrow r(s,a) + \mathbb{E}_{a' \sim \pi_{new}}[Q_C(s',a')]$$ Q_{Target}

Distribution Mismatch

Objective: $\min_Q \mathbb{E}_{a \sim \pi_B}[Q_C(s',a') - Q_{Target}]$



Objective: $\min_Q \mathbb{E}_{(s,a)\sim Data}[Q_C(s',a') - Q_{Target}] - \alpha \mathbb{E}_{s\sim Data,\, a\sim\pi_\phi}[Q_C(s',a')]$

Regularizer (Push Q-values up)

Conservative Q-Learning

(CQL, Kumar et al 2020)

Provably $\hat{Q} \ge Q$

Conservative Safety Critic

Lagrangian: $\max_{\pi_\phi} \min_{\lambda \geq 0} \mathbb{E}_{s \sim \rho_\phi, a \sim \pi_\phi}[Q_R^{\pi_\phi}(s,a) - \lambda(Q_C(s,a) - \chi)]$

✓ Solve with Primal-Dual Grad. Descent

✓ Off-policy Value estimates Task $(Q_R(s,a))$ & safety $(Q_C(s,a))$

Safety Critic Algorithm

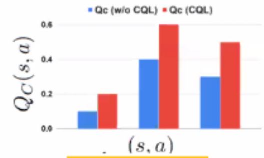
Lagrangian: $\max_{\pi_\phi} \min_{\lambda \geq 0} \mathbb{E}_{s \sim \rho_\phi, a \sim \pi_\phi}[Q_R^{\pi_\phi}(s,a) - \lambda(Q_C(s,a) - \chi)]$

$- \alpha \mathbb{E}_{s \sim Data,\, a \sim \pi_\phi}[Q_C(s', a')]$

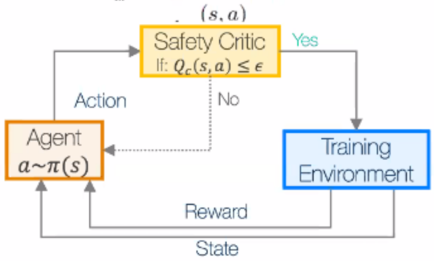
Conservative Safety Critic

Conservative Safety Critic

Inference Mechanism



Vanilla RL



Conservative Safety Critic

Lagrangian: $\max_{\pi_\phi} \min_{\lambda \geq 0} \mathbb{E}_{s \sim \rho_\phi, a \sim \pi_\phi}[Q_R^{\pi_\phi}(s,a) - \lambda(Q_C(s,a) - \chi)]$

$- \alpha \mathbb{E}_{s \sim Data,\, a \sim \pi_\phi}[Q_C(s', a')]$

Conservative Safety Critic

Guarantees

Safe Training: Constraint is satisfied during for each iterate

Theorem 1. *Consider policy updates that solve the constrained optimization problem defined in Equation 5. With high probability $\geq 1 - \omega$, we have the following upper bound on expected probability of failure $V_C^{\pi_{\phi_{new}}}(\mu)$ for $\pi_{\phi_{new}}$ during every policy update iteration:*

$$V_C^{\pi_{\phi_{new}}}(\mu) \leq \chi + \zeta - \frac{\Delta}{1-\gamma} + \frac{\sqrt{2\delta}\gamma\epsilon_C}{(1-\gamma)^2} \quad \text{where} \quad \zeta \leq \frac{C'\sqrt{\log(1/\omega)}}{|N|} \tag{7}$$

Safety not just at the end but all along training

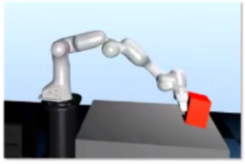
Sample Efficiency: Bounded effect on rate of convergence

Theorem 2 (Convergence rate for policy gradient updates with the safety constraint). *If we run the policy gradient updates through equation 5, for policy π_ϕ, with μ as the starting state distribution, with $\phi^{(0)} = 0$, learning rate $\eta > 0$, and choose α as mentioned in the discussion of Theorem 1, then for all policy update iterations $T > 0$ we have, with probability $\geq 1 - \omega$,*

$$V_R^*(\mu) - V_R^{(T)}(\mu) \leq \frac{\log|\mathcal{A}|}{\eta T} + \frac{1}{(1-\gamma)^2 T} + K\frac{\sum_{t=0}^{T-1}\lambda^{(t)}}{\eta T} \quad \text{where} \quad K \leq (1-\chi) + \frac{4\sqrt{2\delta}\gamma}{(1-\gamma)^2}$$

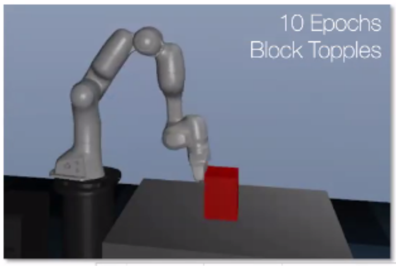
Adding constraint does not affect sample efficiency

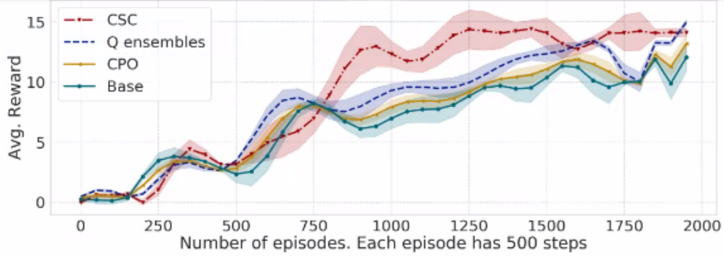
Conservative Safety Critic
Evaluation



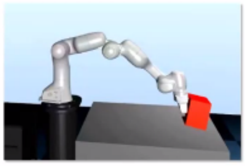
Dexterous Manipulation

Objective: Learn to Push
Failure: Glass Topples



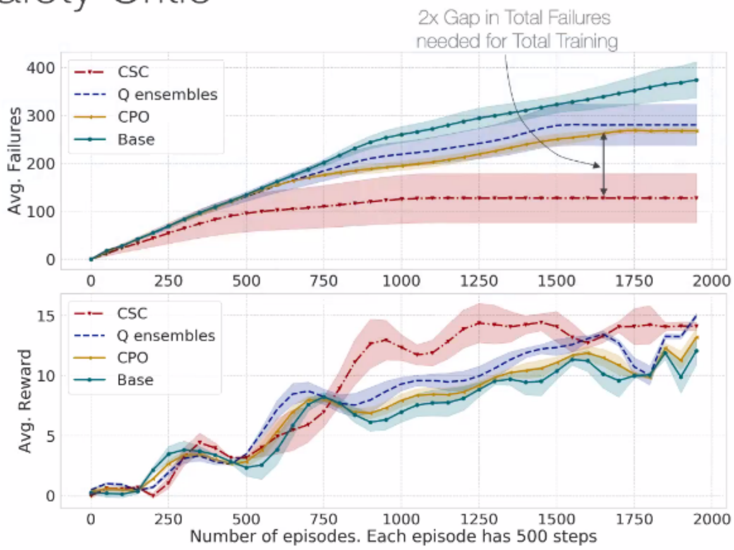


Conservative Safety Critic
Evaluation



Dexterous Manipulation

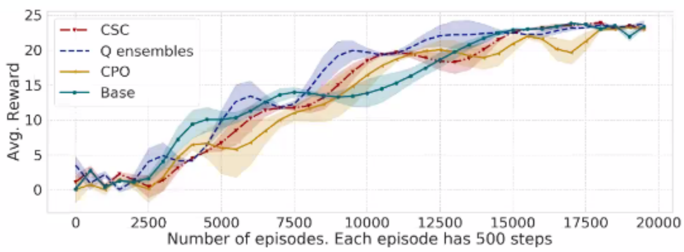
Objective: Learn to Push
Failure: Glass Topples

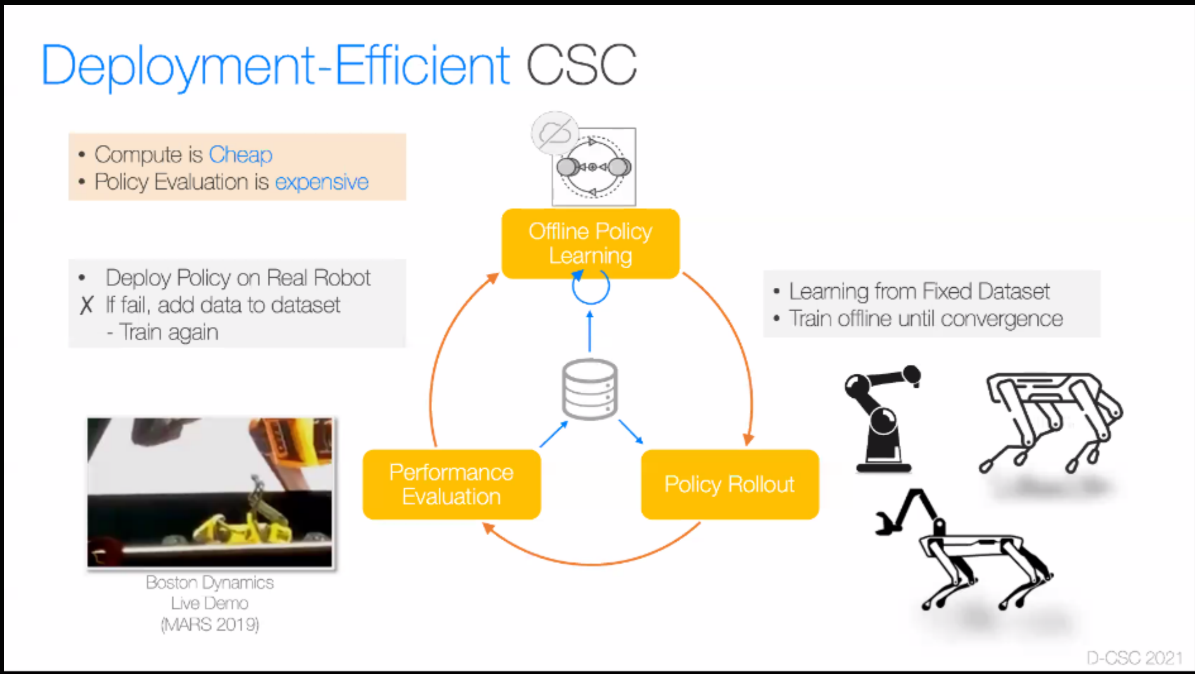


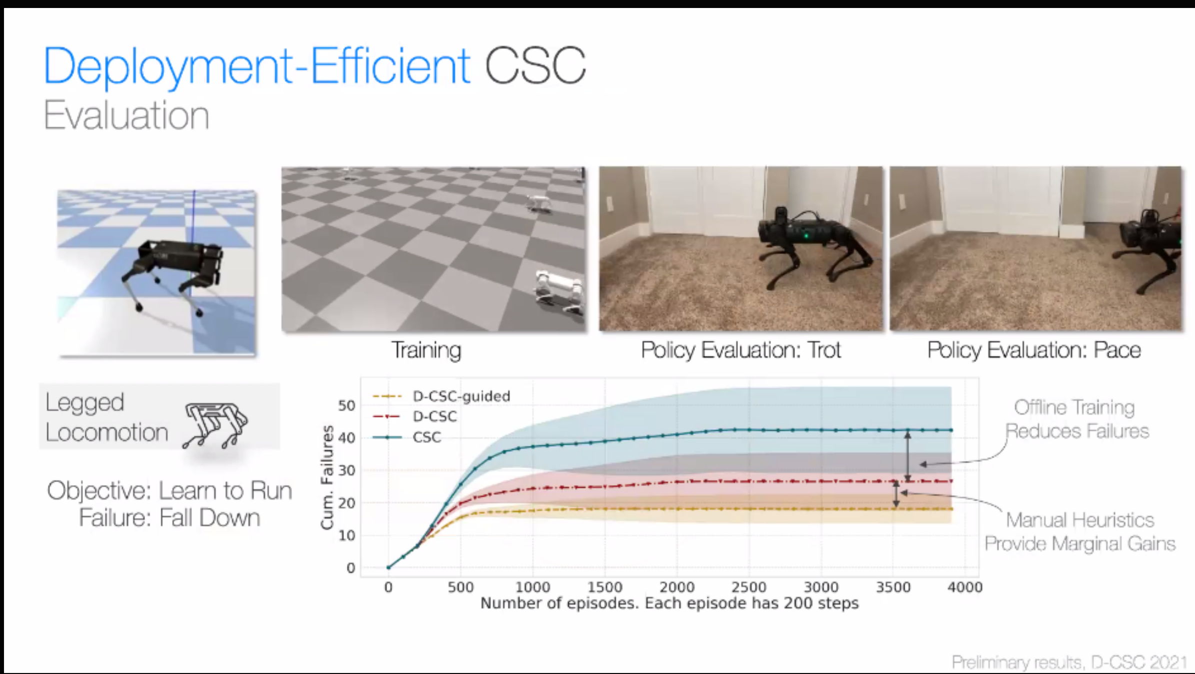
Conservative Safety Critic
Evaluation

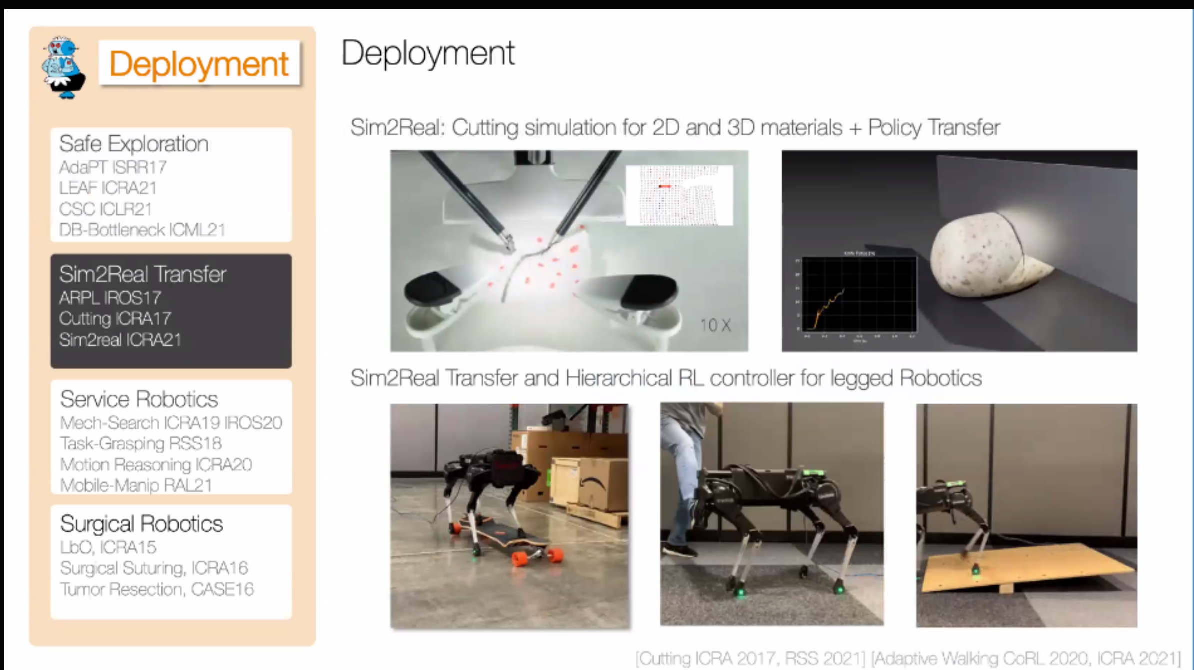
Legged Locomotion

Objective: Learn to Run
Failure: Fall Down




Deployment-Efficient CSC
• Compute is Cheap
• Policy Evaluation is expensive
• Deploy Policy on Real Robot
✗ If fail, add data to dataset
- Train again
Offline Policy Learning
Performance Evaluation
Policy Rollout
• Learning from Fixed Dataset
• Train offline until convergence
Boston Dynamics
Live Demo
(MARS 2019)
D-CSC 2021


Deployment-Efficient CSC
Evaluation
Training
Policy Evaluation: Trot
Policy Evaluation: Pace
Legged Locomotion
Objective: Learn to Run
Failure: Fall Down
D-CSC-guided
D-CSC
CSC
Cum. Failures
0
10
20
30
40
50
0
500
1000
1500
2000
2500
3000
3500
4000
Number of episodes. Each episode has 200 steps
Offline Training Reduces Failures
Manual Heuristics Provide Marginal Gains
Preliminary results, D-CSC 2021


Deployment
Safe Exploration
AdaPT ISRR17
LEAF ICRA21
CSC ICLR21
DB-Bottleneck ICML21
Sim2Real Transfer
ARPL IROS17
Cutting ICRA17
Sim2real ICRA21
Service Robotics
Mech-Search ICRA19 IROS20
Task-Grasping RSS18
Motion Reasoning ICRA20
Mobile-Manip RAL21
Surgical Robotics
LbC, ICRA15
Surgical Suturing, ICRA16
Tumor Resection, CASE16
Deployment
Sim2Real: Cutting simulation for 2D and 3D materials + Policy Transfer
10 X
Sim2Real Transfer and Hierarchical RL controller for legged Robotics
[Cutting ICRA 2017, RSS 2021] [Adaptive Walking CoRL 2020, ICRA 2021]

Deployment

Safe Exploration
AdaPT ISRR17
LEAF ICRA21
CSC ICLR21
DB-Bottleneck ICML21

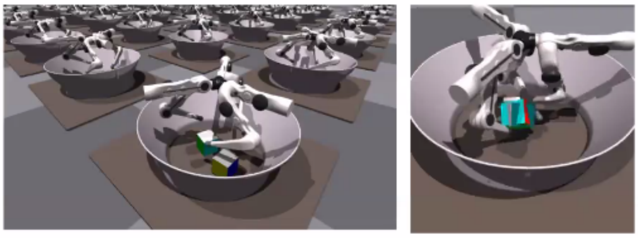
Sim2Real Transfer
ARPL IROS17
Cutting ICRA17
Sim2real ICRA21

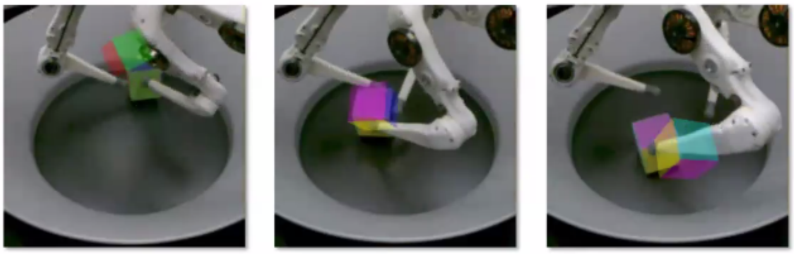
Service Robotics
Mech-Search ICRA19 IROS20
Task-Grasping RSS18
Motion Reasoning ICRA20
Mobile-Manip RAL21

Surgical Robotics
LbO, ICRA15
Surgical Suturing, ICRA16
Tumor Resection, CASE16

Deployment

Sim2Real: In-hand manipulation





[CORL 2021 (under review)]

Deployment

Safe Exploration
AdaPT ISRR17
LEAF ICRA21
CSC ICLR21
DB-Bottleneck ICML21

Sim2Real Transfer
ARPL IROS17
Cutting ICRA17
Sim2real ICRA21

Service Robotics
Mech-Search ICRA19 IROS20
Task-Grasping RSS18
Motion Reasoning ICRA20
Mobile-Manip RAL21

Surgical Robotics
LbO, ICRA15
Surgical Suturing, ICRA16
Tumor Resection, CASE16

Deployment

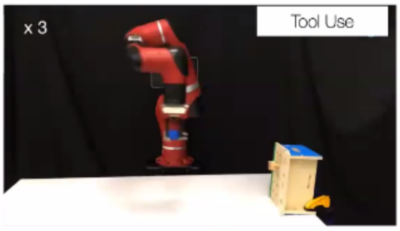
Learning Manipulation Skills

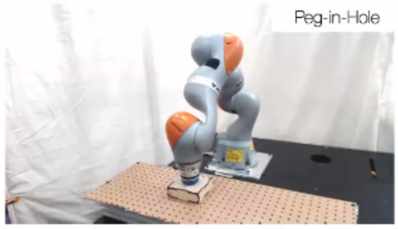
Object Sorting

Object Search

Goal-Conditioned Imitation

Robot Execution





[Sorting ICRA 2018], [Object Search: ICRA 2019], [Goal Imitation ICRA 2020], [Tool Use RSS2018, RSS 2021], [Peg-in Hole ICRA 2019]

Deployment

Safe Exploration
AdaPT ISRR17
LEAF ICRA21
CSC ICLR21
DB-Bottleneck ICML21

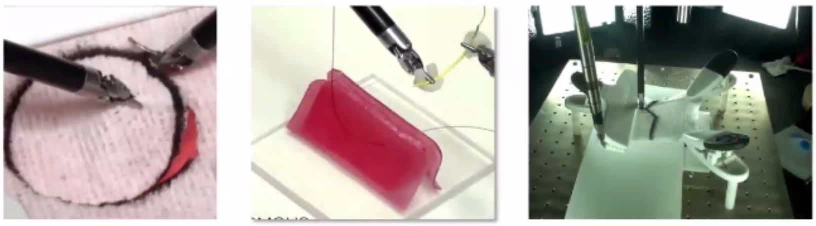
Sim2Real Transfer
ARPL IROS17
Cutting ICRA17
Sim2real ICRA21

Service Robotics
Mech-Search ICRA19 IROS20
Task-Grasping RSS18
Motion Reasoning ICRA20
Mobile-Manip RAL21

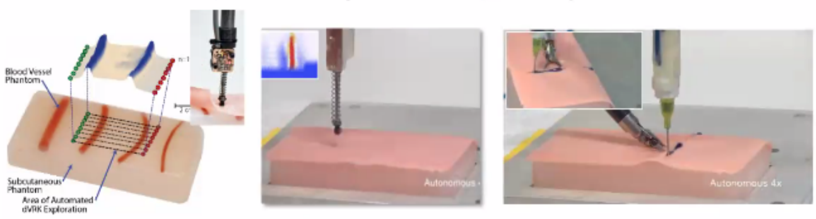
Surgical Robotics
LbO, ICRA15
Surgical Suturing, ICRA16
Tumor Resection, CASE16

Deployment

Surgical Subtask Automation: Imitation Learning for Structured Tasks



Tumor Localization and Resection: Bayesian Search with Noisy Sensors



[Subtask Automation ICRA 2015, ICRA 2016, IJRR 2019] [Tumor Localization CASE2016, Hamlyn 2015]

Deployment

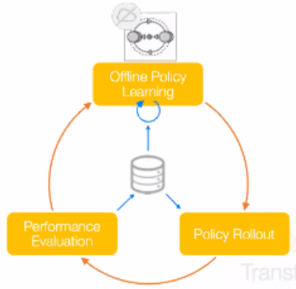
Safe Exploration
AdaPT ISRR17
LEAF ICRA21
CSC ICLR21
DB-Bottleneck ICML21

Sim2Real Transfer
ARPL IROS17
Cutting ICRA17
Sim2real ICRA21

Service Robotics
Mech-Search ICRA19 IROS20
Task-Grasping RSS18
Motion Reasoning ICRA20
Mobile-Manip RAL21

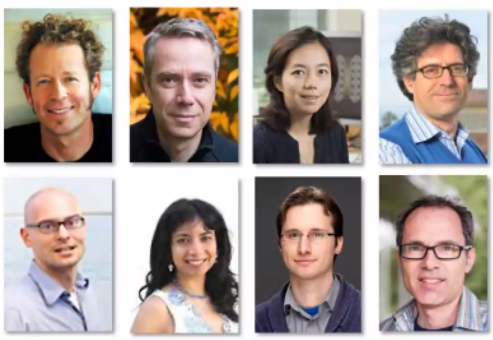
Surgical Robotics
LbO, ICRA15
Surgical Suturing, ICRA16
Tumor Resection, CASE16

Deployment



Practical challenges in Robot Deployment
Novel Algorithms + Systems' view of Robotics

Acknowledgements



Collaborators and Colleagues

Structure	Discovery	Deployment	Agenda
State/Action Reps. VICES IROS19 LASER ICRA21 Making Sense ICRA19 Unsup KPs PAMI21	**Data Collection** RoboTurk-v1 CoRL18 RoboTurk-v2 IROS19 Latent Actions ICRA20	**Safe Exploration** AdaPT ISRR17 LEAF ICRA21 CSC ICLR21 DB-Bottleneck ICML21	**Perception** Dense Action Understanding Concept Grounding Learning from Interaction
Inductive Biases C-Learning ICLR21 OCEAN UAI20 D2RL arXiv20	**Imitation & Offline RL** TSC IJRR17, SwIRL IJRR19 AC-Teach: CoRL19 IRIS ICRA20	**Sim2Real Transfer** ARPL IROS17 Cutting ICRA17 Sim2real ICRA21	**Planning & Control** Bayes Adaptive RL Causality via Intervention Neuro-Symbolic Planning
Structure in Planning CAVIN CORL20 Skill Hierarchy ICLR21 Finding-IT, CVPR18	**Data Augmentation** CoDA NeurIPS20 OB2I ICML21 S4RL arXiv21	**Service Robotics** Mech-Search ICRA19 IROS20 Task-Grasping RSS18 Motion Reasoning ICRA20 Mobile-Manip RAL21	**Data & Benchmarks** Multi-modal Datasets Taxonomy of Manipulation Platoon Autonomy
Neural Programming NTP ICRA18 NTG CVPR19 Cont.Relax IROS19	**Causal Discovery** Semi-Sup GAN ICML20 VCDN NeurIPS20 MAC ICCV21	**Surgical Robotics** LbO, ICRA15 Surgical Suturing, ICRA16 Tumur Resection, CASE16	**Applications** Personal Industrial Healthcare