ICLR

Self-Supervised Policy Adaptation during Deployment

Nicklas Hansen, Rishabh Jangir, Yu Sun, Guillem Alenyà,
Pieter Abbeel, Alexei A. Efros, Lerrel Pinto, Xiaolong Wang

ICLR 2021

UC San Diego

DTU

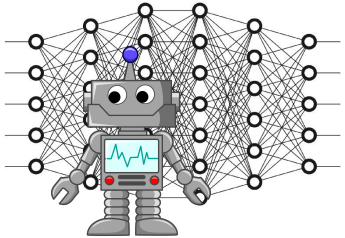
BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Berkeley
UNIVERSITY OF CALIFORNIA

Institut de Robòtica
i Informàtica Industrial

NYU

Observation

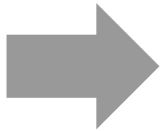


Learned policy

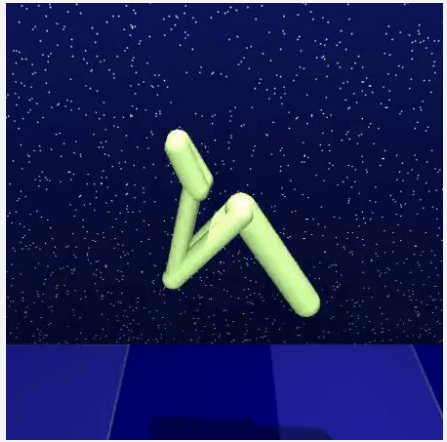
Environment interaction

Complex tasks can be solved by **end-to-end** policy learning

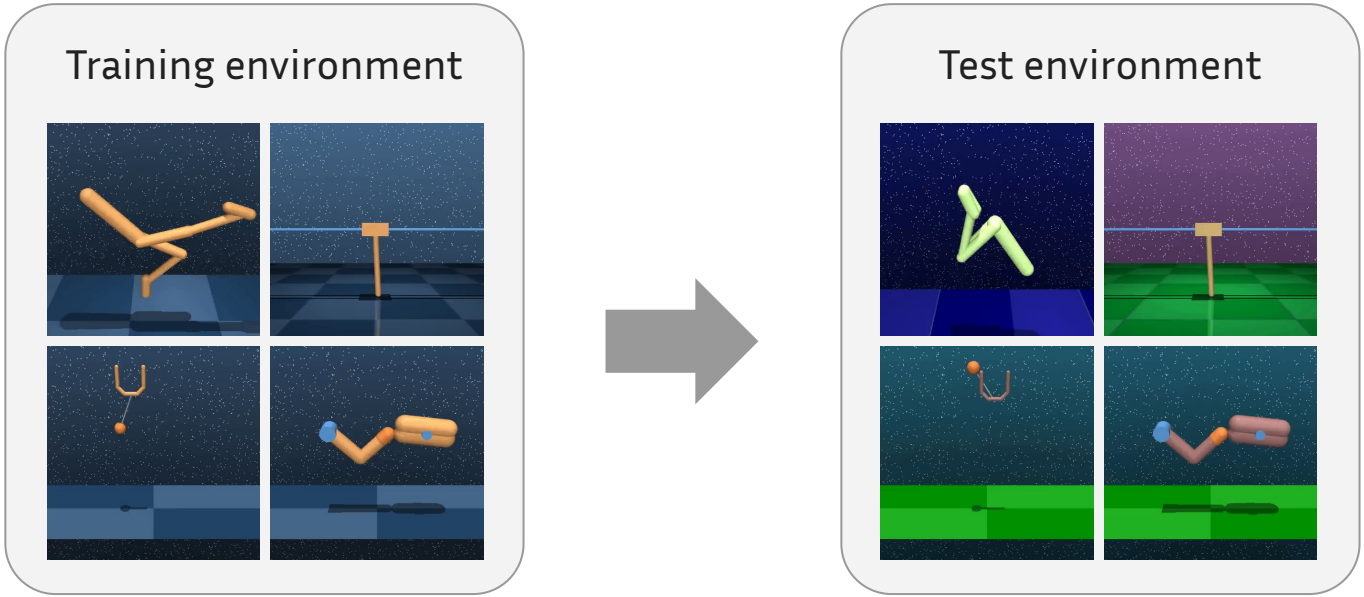
Training environment



Test environment



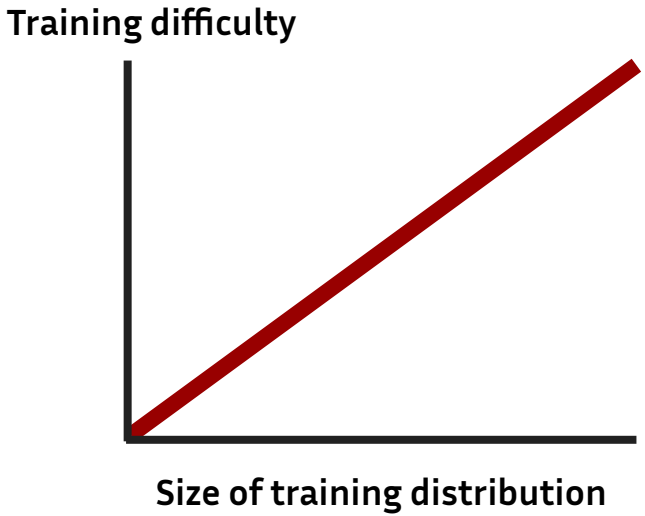
Generalization across different environments is **hard**



Generalization across different environments is **hard**

We can **randomize** elements that we expect to vary at test-time

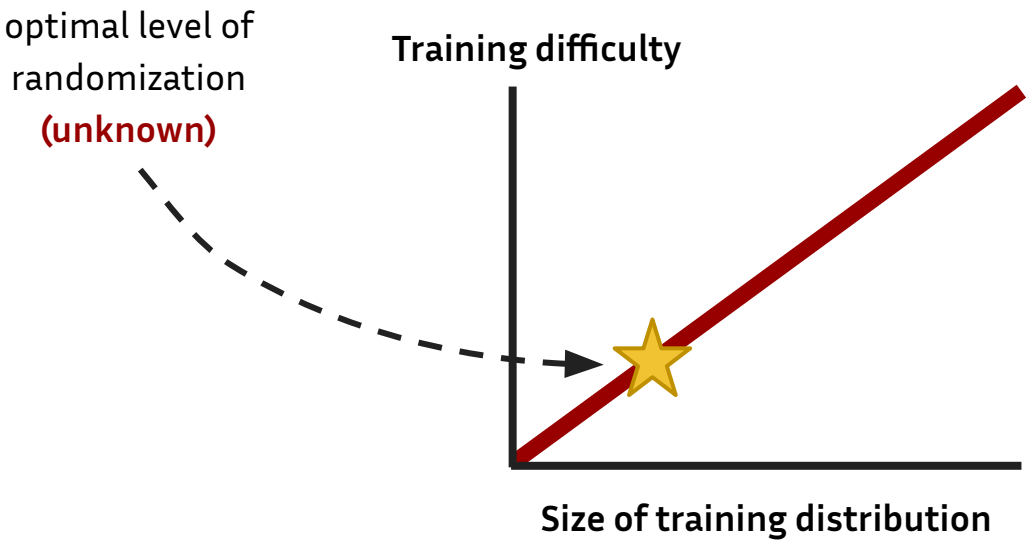
Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World (2017). Tobin et al.



Factors of variation

- Color
- Texture
- Camera pose
- Mass
- Friction
- etc...

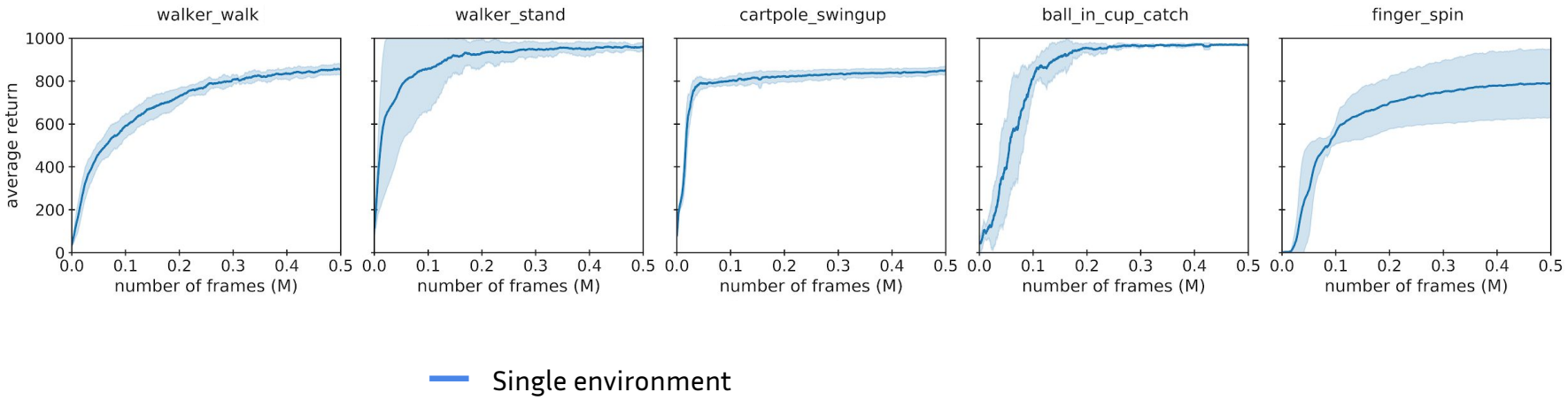
Larger training distribution = **harder** problem to solve

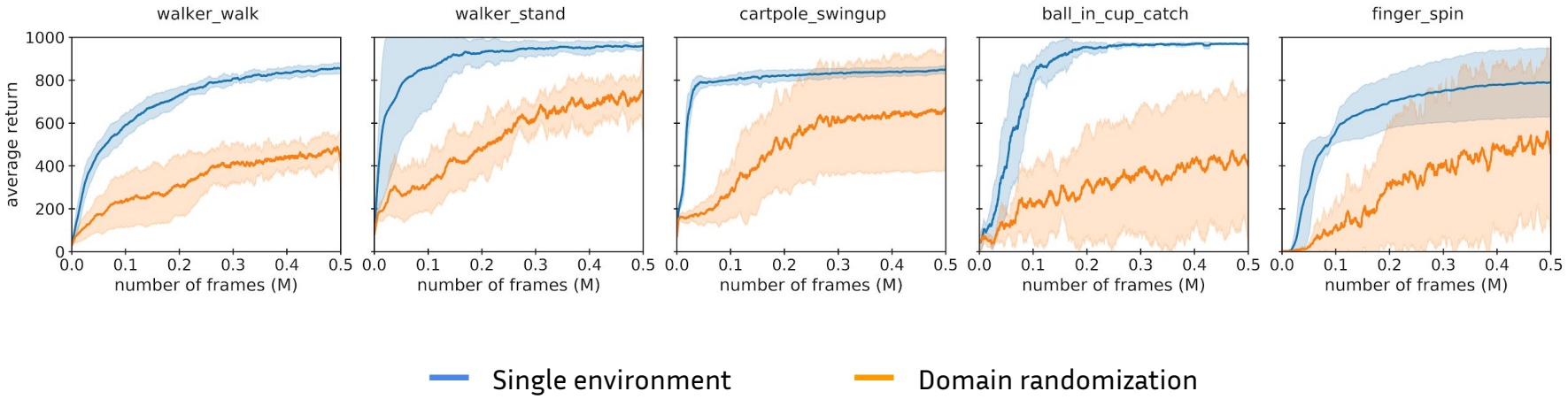


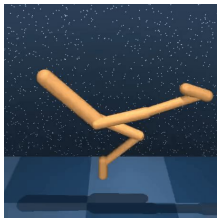
Factors of variation

- Color
- Texture
- Camera pose
- Mass
- Friction
- etc...

Larger training distribution = **harder** problem to solve


walker_walk
walker_stand
cartpole_swingup
ball_in_cup_catch
finger_spin
average return
0
200
400
600
800
1000
0.0
0.1
0.2
0.3
0.4
0.5
number of frames (M)
Single environment


walker_walk
walker_stand
cartpole_swingup
ball_in_cup_catch
finger_spin
average return
0
200
400
600
800
1000
0.0
0.1
0.2
0.3
0.4
0.5
number of frames (M)
Single environment
Domain randomization

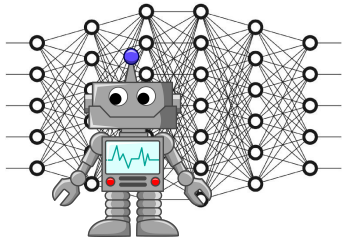


Training environment

Deploy

Test environment

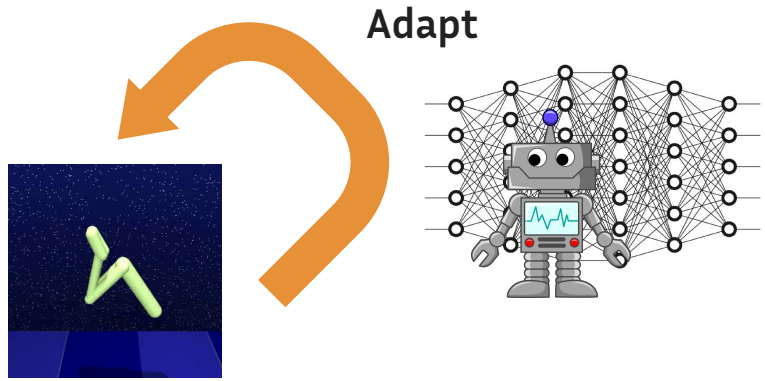
Adapt



Can we instead **adapt** to new environments **during deployment**?

Challenges:

- No data prior to deployment
- Potentially no reward signal



Test environment

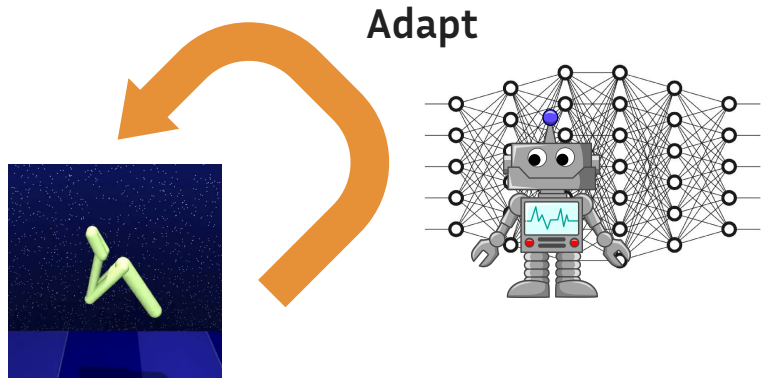
Can we instead **adapt** to new environments **during deployment**?

Challenges:

- No data prior to deployment
- Potentially no reward signal

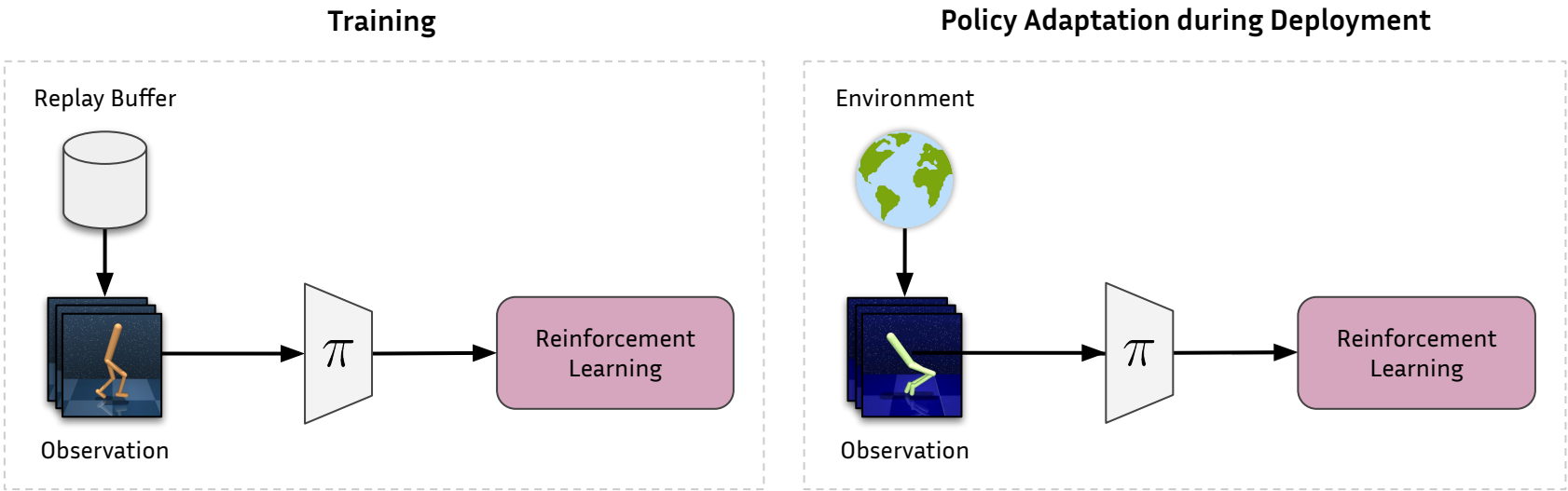
Solution:

- **Online** adaptation
- **Self-supervised** training signal

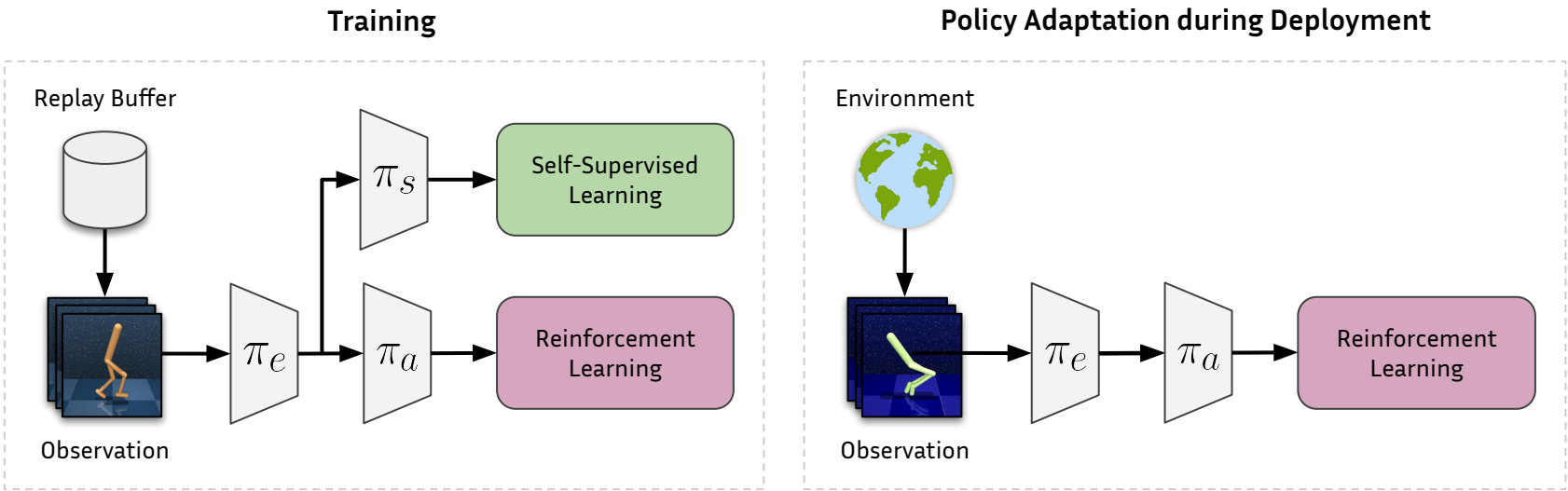


Can we instead **adapt** to new environments **during deployment**?

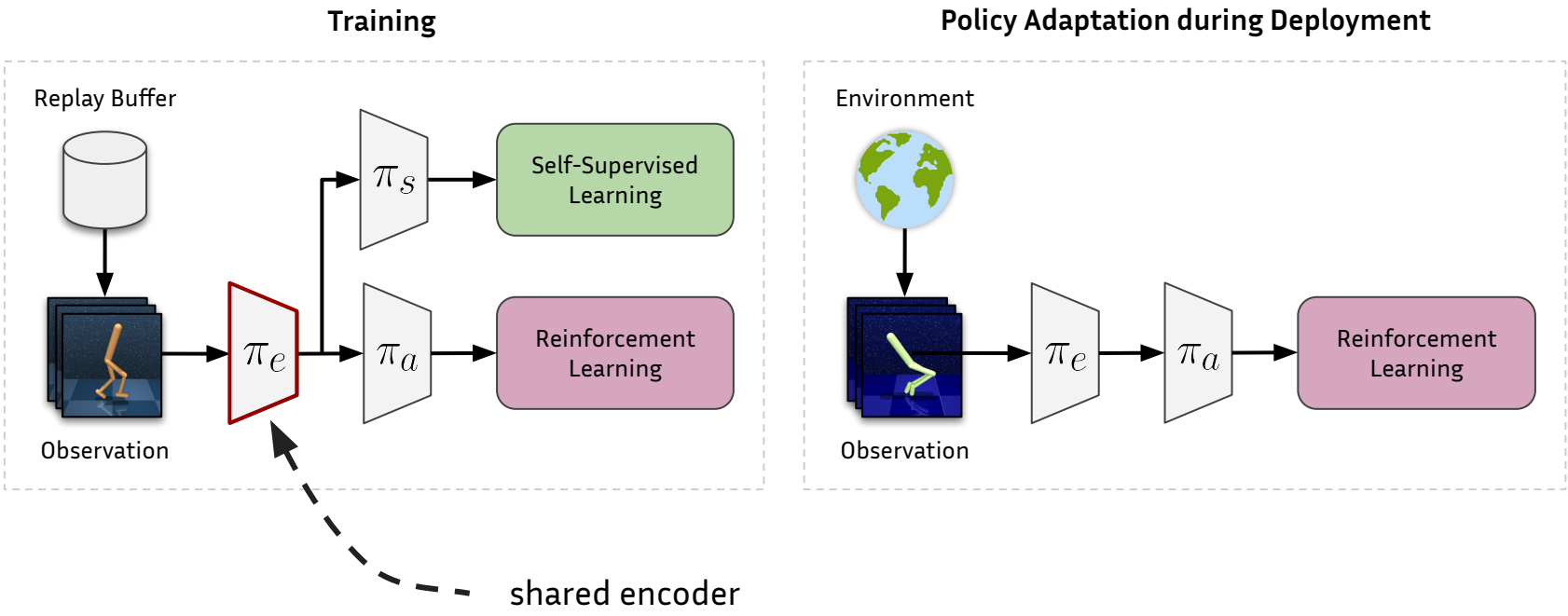
Algorithmic overview



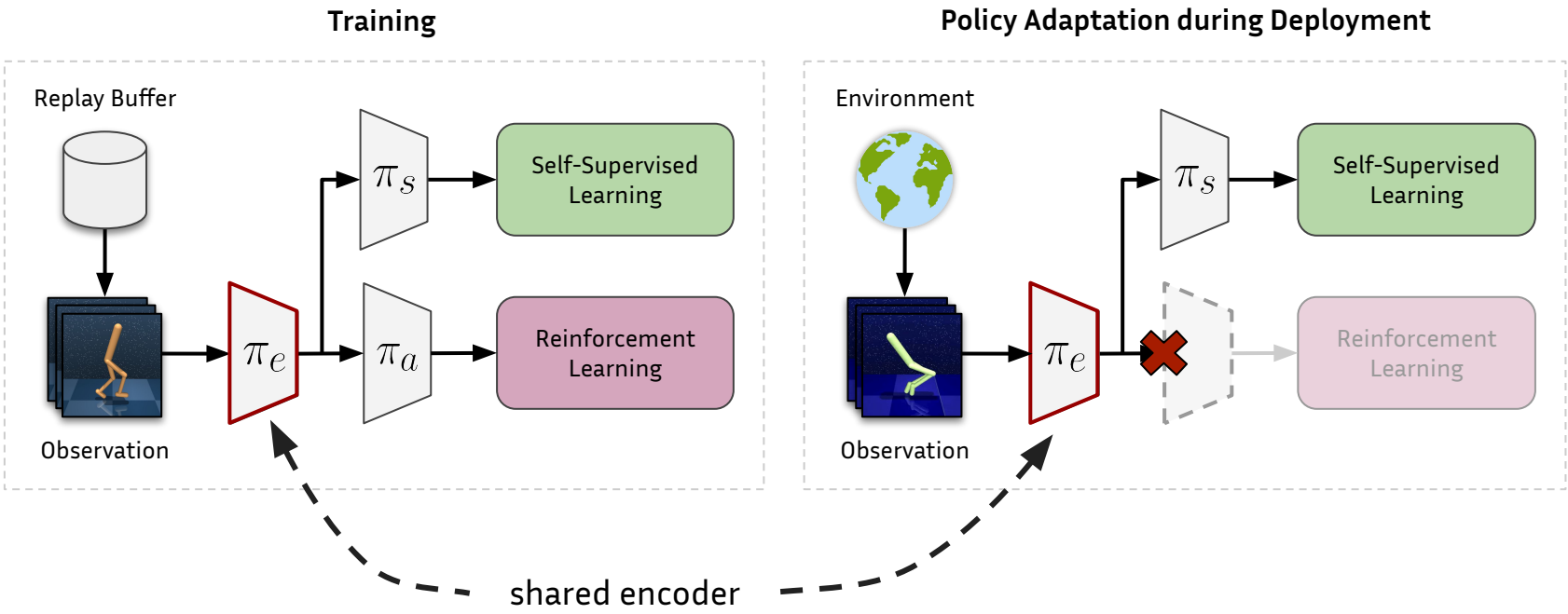
Algorithmic overview



Algorithmic overview

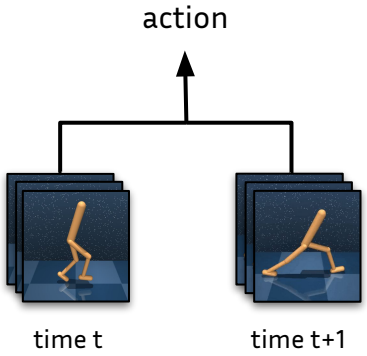


Algorithmic overview

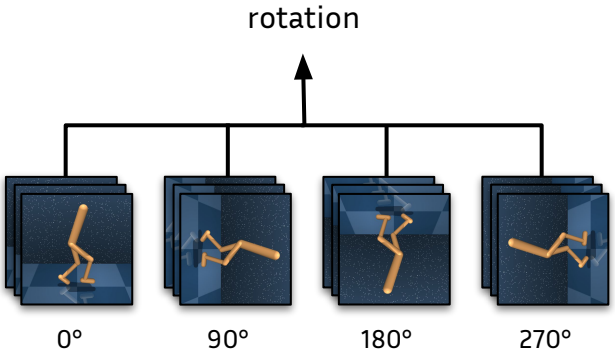


Algorithmic overview

Algorithmic overview



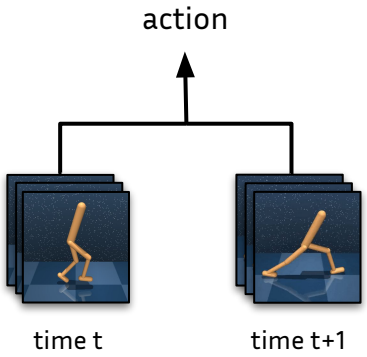
Inverse dynamics model



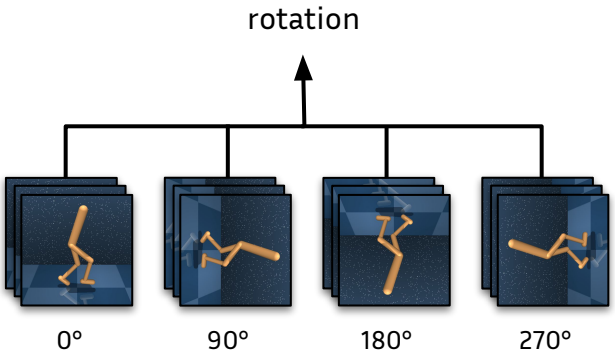
Rotation prediction

Unsupervised Representation Learning by Predicting Image Rotations (2018). Gidaris et al.

Algorithmic overview



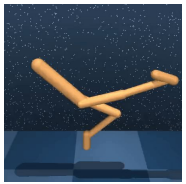
Inverse dynamics model



Rotation prediction

Choice of self-supervised task is **task-dependent**

Unsupervised Representation Learning by Predicting Image Rotations (2018). Gidaris et al.

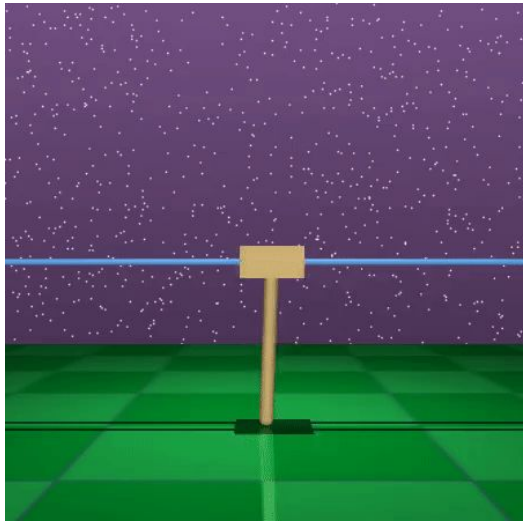


Training environment

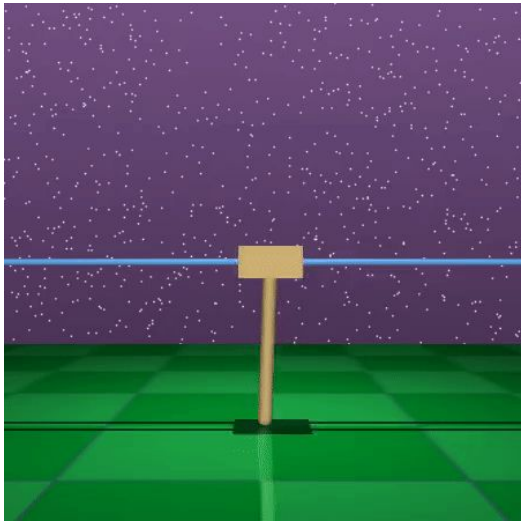
Deploy

We evaluate generalization to **100+** environments

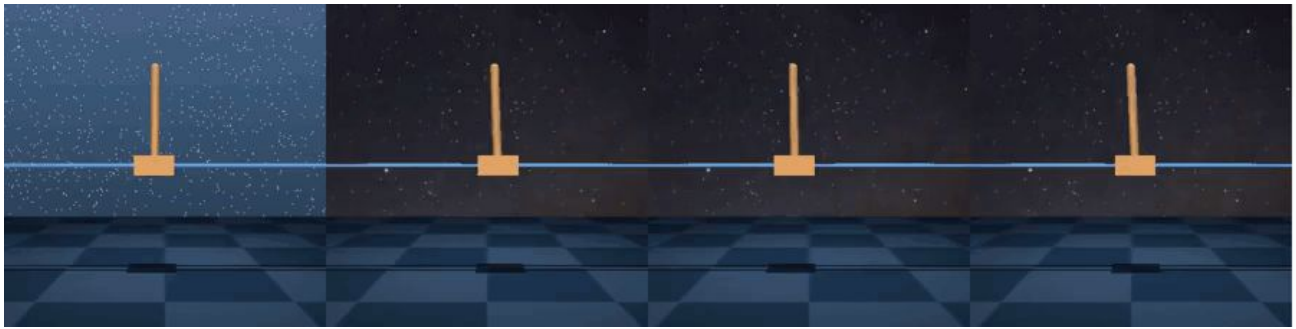
Training environment

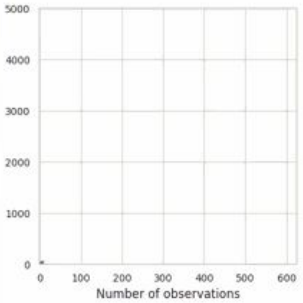


Direct transfer



PAD (ours)





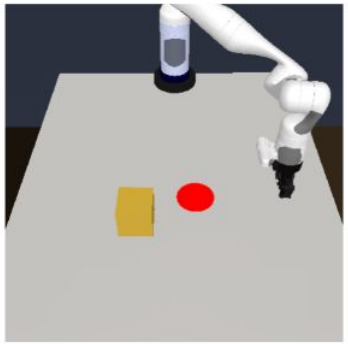
Training environment | Direct transfer | CURL | **PAD (ours)** | Cumulative reward

CURL: Contrastive Unsupervised Representations for Reinforcement Learning (2020). Srinivas et al.

Training environment | Direct transfer | CURL | **PAD (ours)** | Cumulative reward

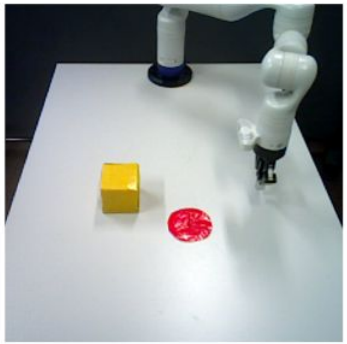
CURL: Contrastive Unsupervised Representations for Reinforcement Learning (2020). Srinivas et al.

Adapting to the real world



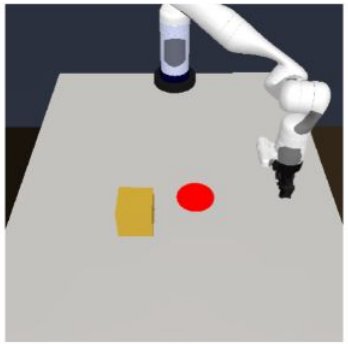
Simulation

Deploy



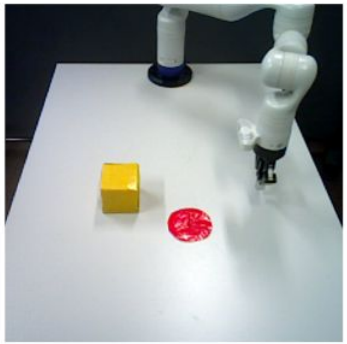
Default transfer

Adapting to the real world

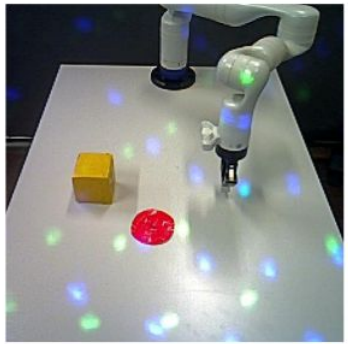


Simulation

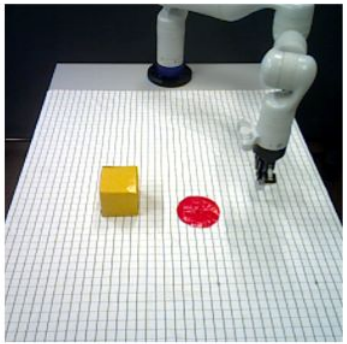
Deploy



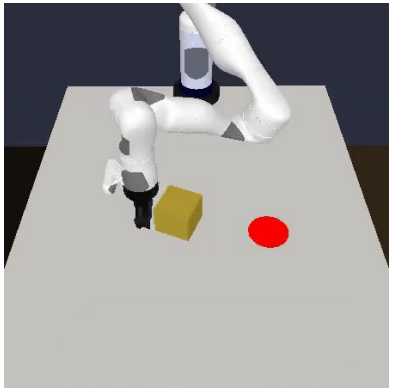
Default transfer



Disco lights



Tablecloth

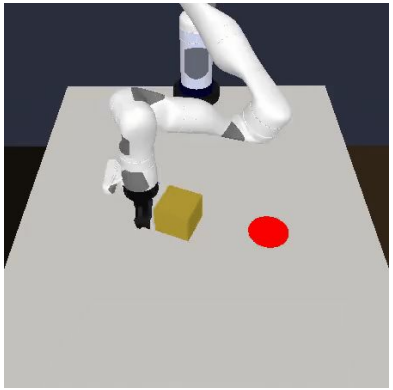


Training environment

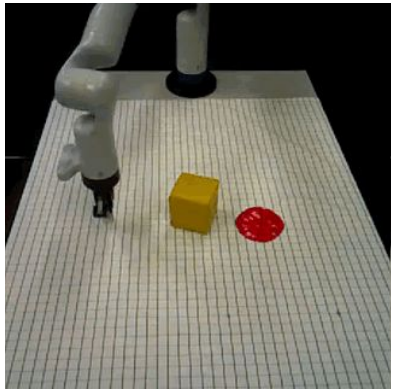
Direct transfer

PAD (ours)

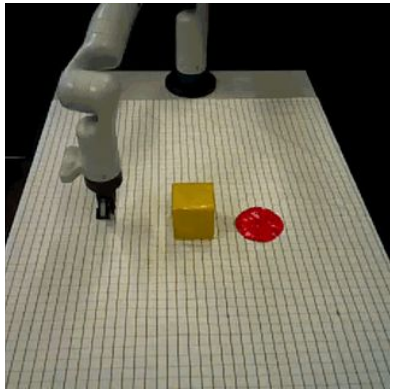
Our method improves generalization in **sim2real**



Training environment

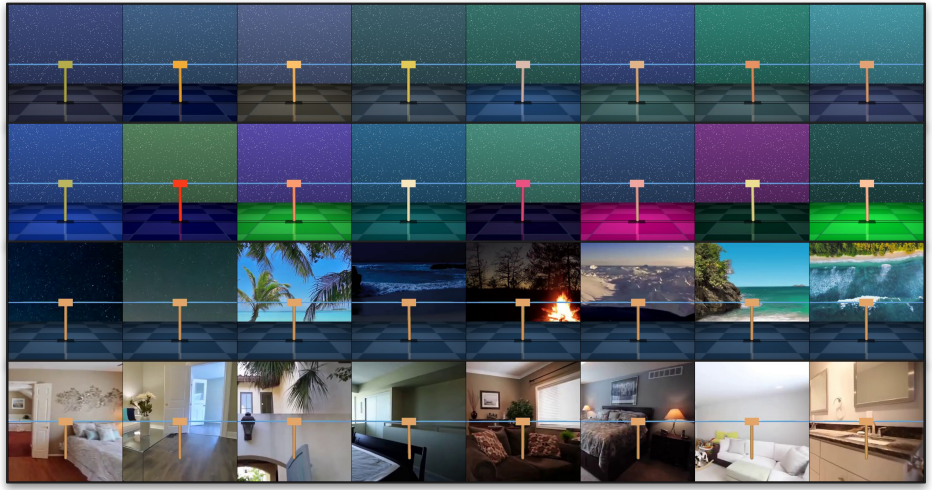


Direct transfer



PAD (ours)

Our method improves generalization in **sim2real**



For more information:

https://nicklashansen.github.io/PAD

Self-Supervised Policy Adaptation during Deployment

Nicklas Hansen, Rishabh Jangir, Yu Sun, Guillem Alenyà,
Pieter Abbeel, Alexei A. Efros, Lerrel Pinto, Xiaolong Wang

ICLR 2021

UC San Diego

DTU

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Berkeley
UNIVERSITY OF CALIFORNIA

Institut de Robòtica
i Informàtica Industrial

NYU