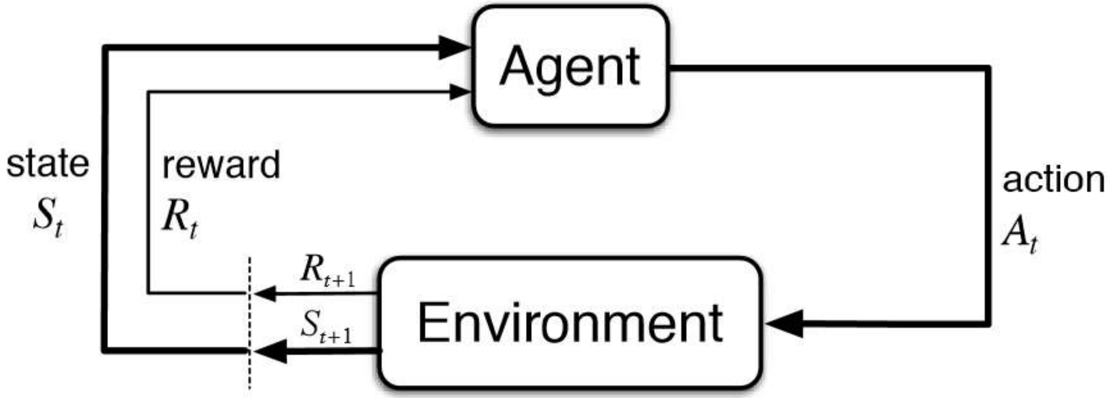
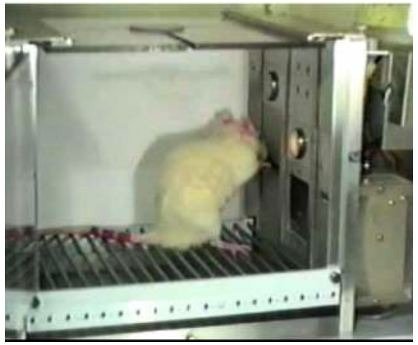
Representation learning and Reinforcement learning

Amey Pore
AI ML Club, Verona
30th May 2024

Marie Sklodowska-Curie Training Network GA No 813782

What we know: Reinforcement learning





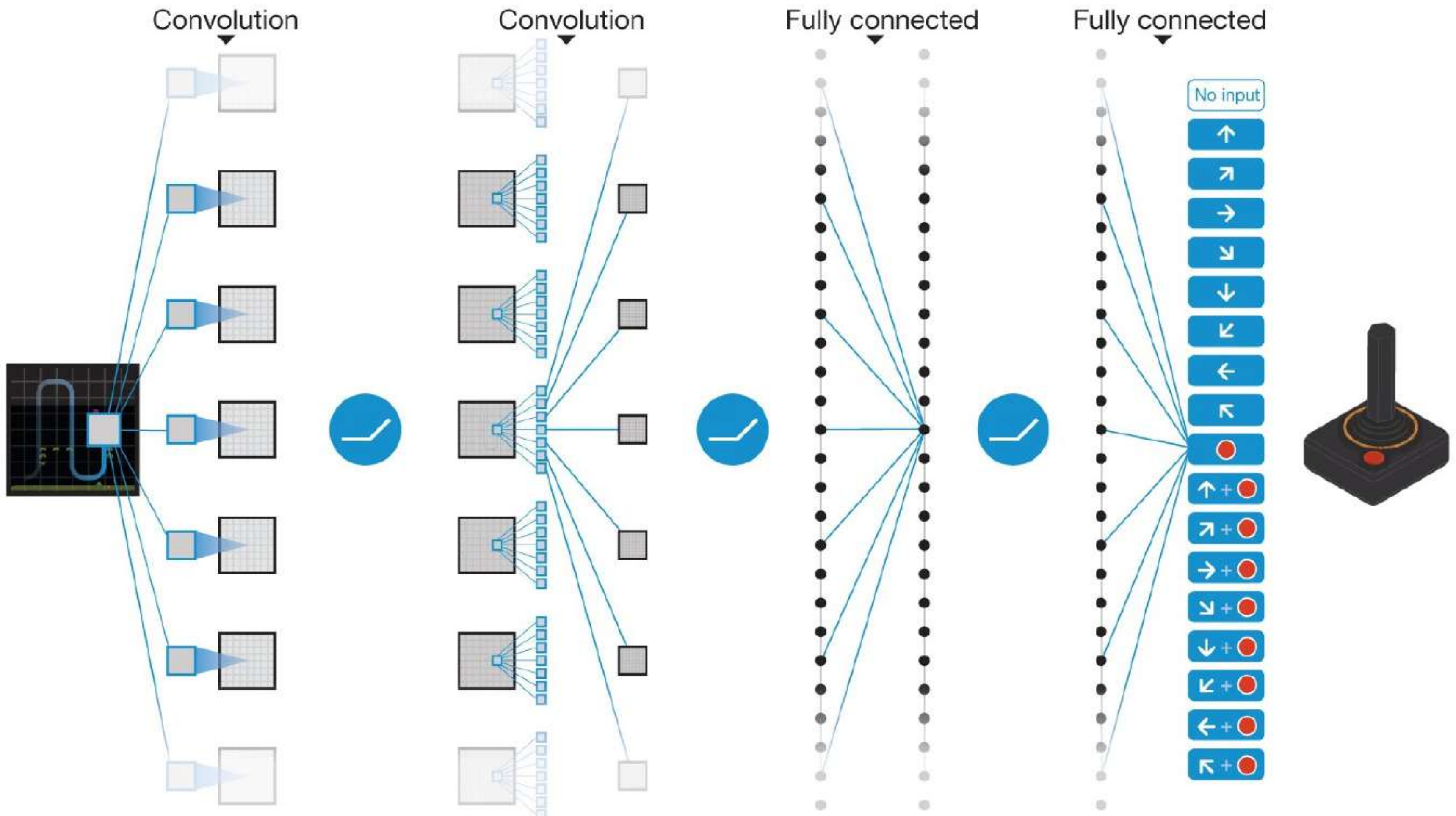
$$\max_{\theta} \quad \mathrm{E}[\sum_{t=0}^{H} R(s_t)|\pi_\theta]$$

Compared to supervised learning;
Additional challenges

- Credit assignment
- Exploration
- Stability

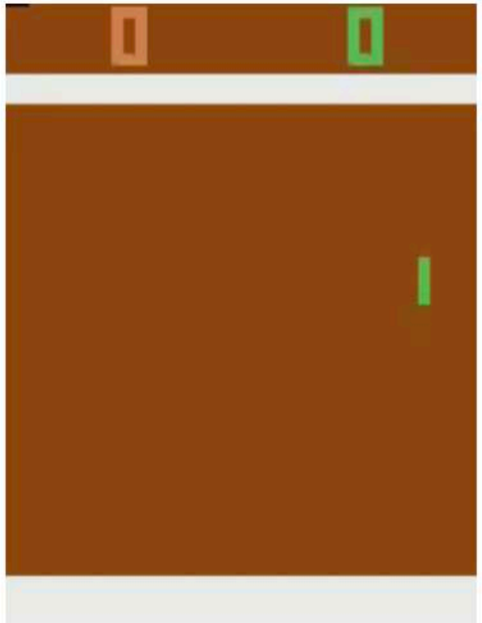
Image credit: Sutton and Barto 1998

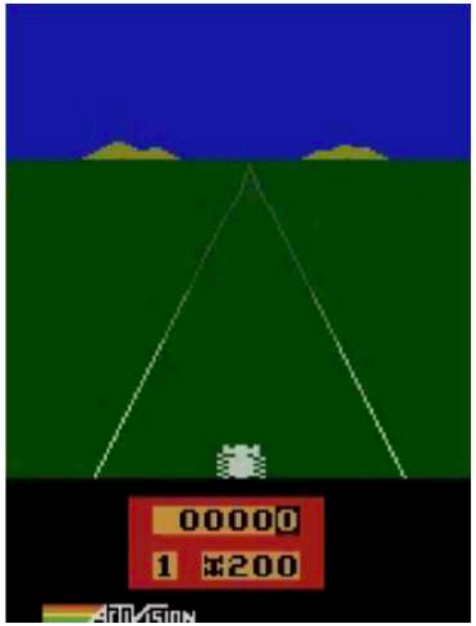
Deep RL success story: Atari

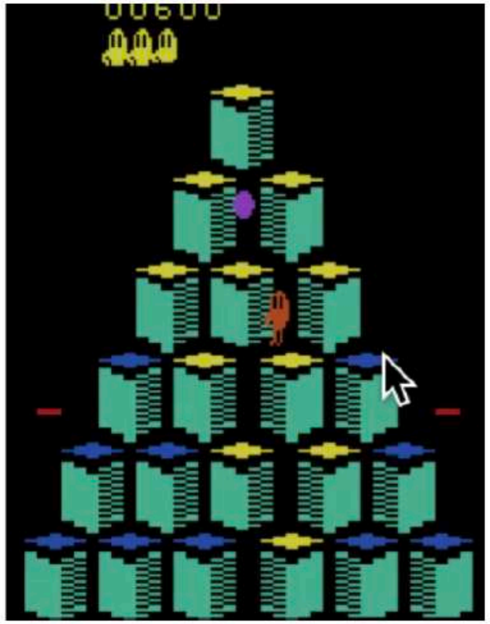


Human-level control through deep reinforcement learning, Mnih et al, Nature 2015

Deep RL success story: Atari



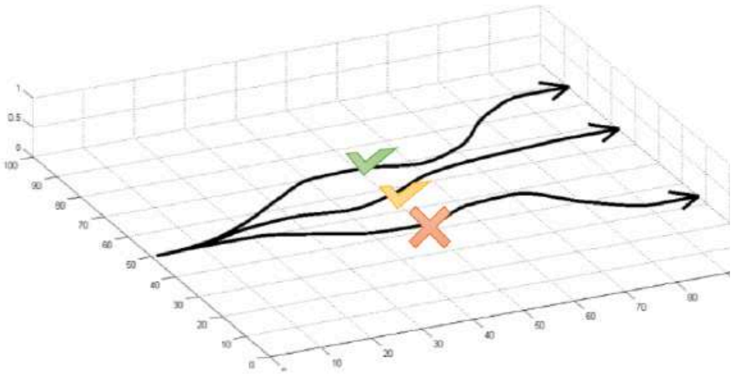


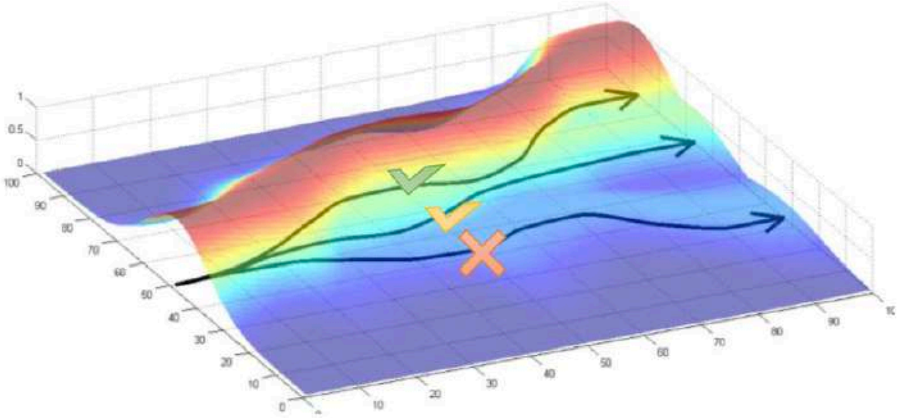


DQN Mnih et al, NIPS 2013 / Nature 2015
MCTS Guo et al, NIPS 2014; **TRPO** Schulman, Levine, Moritz, Jordan, Abbeel, ICML 2015; **A3C** Mnih et al, ICML 2016; **Dueling DQN** Wang et al ICML 2016; **Double DQN** van Hasselt et al, AAAI 2016; **Prioritized Experience Replay** Schaul et al, ICLR 2016; **Bootstrapped DQN** Osband et al, 2016; **Q-Ensembles** Chen et al, 2017; **Rainbow** Hessel et al, 2017; **Accelerated** Stooke and Abbeel, 2018; …

Deep RL success story: policy gradient

$$\text{policy gradient:} \quad \nabla_\theta J(\theta) \approx \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} \nabla_\theta \log \pi_\theta(\mathbf{a}_{i,t} | \mathbf{s}_{i,t}) \right) \left(\sum_{t=1}^{T} r(\mathbf{s}_{i,t}, \mathbf{a}_{i,t}) \right)$$





Deep RL success story: Go

AlphaGo Silver et al, Nature 2015

AlphaGoZero Silver et al, Nature 2017

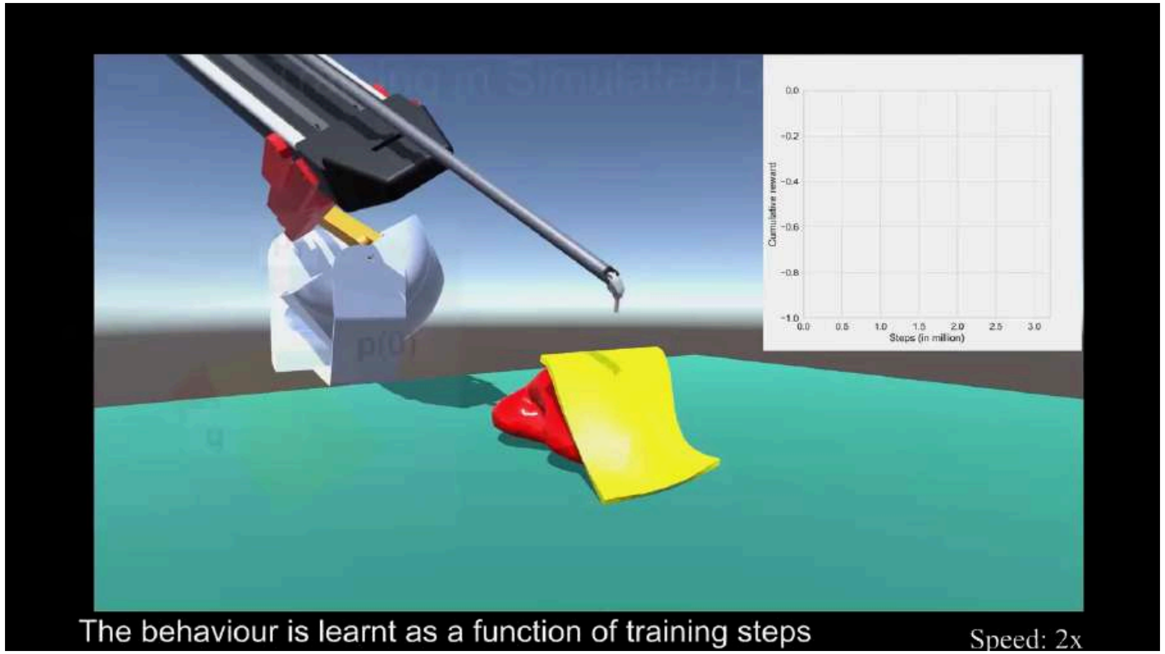
AlphaZero Silver et al, 2017

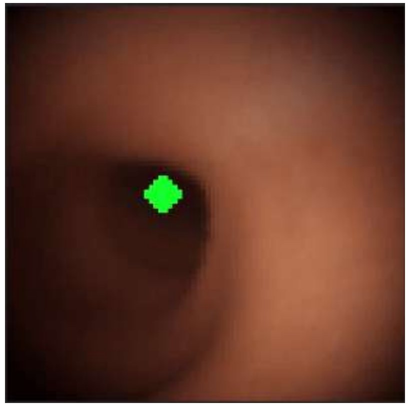
Tian et al, 2016; Maddison et al, 2014; Clark et al, 2015

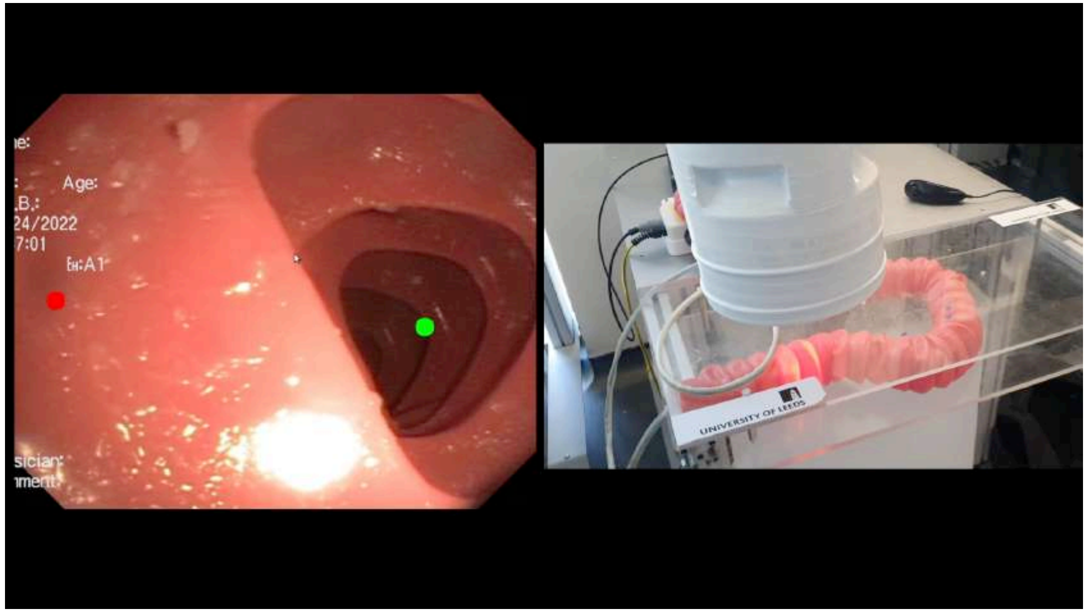
Deep RL Success: Locomotion

TRPO Schulman, Levine, Moritz, Jordan, Abbeel, 2015 + GAE Schulman, Moritz, Levine, Jordan Abbeel, 2016

Deep RL Success: robotic surgery





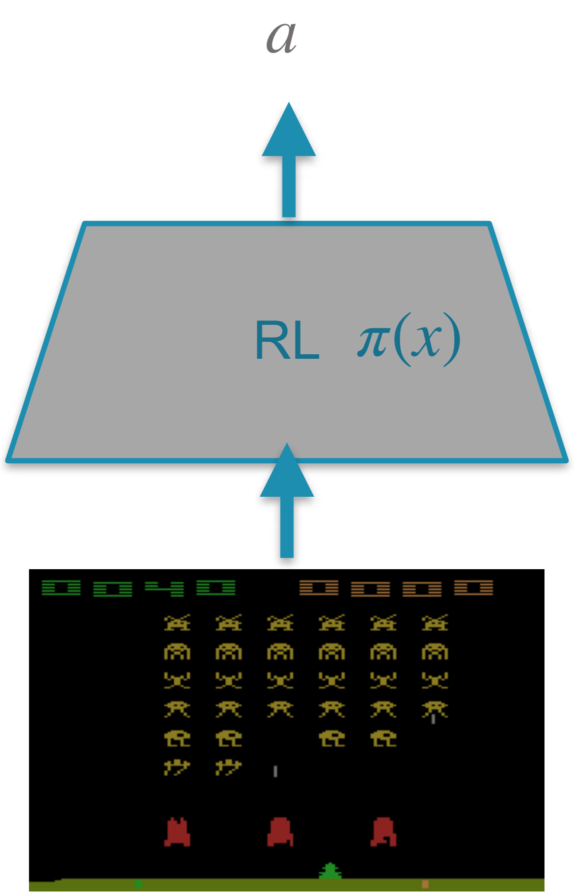


UnityFlexML *Pore et al, 2020 +* ***DVC*** *Pore et al, 2022*

In this talk

- What is representation learning in RL
- What are good representations
- How do we learn them?

End-to-End Reinforcement learning



—> Learn mapping from observations to action

—> Neural Networks are functional approximations

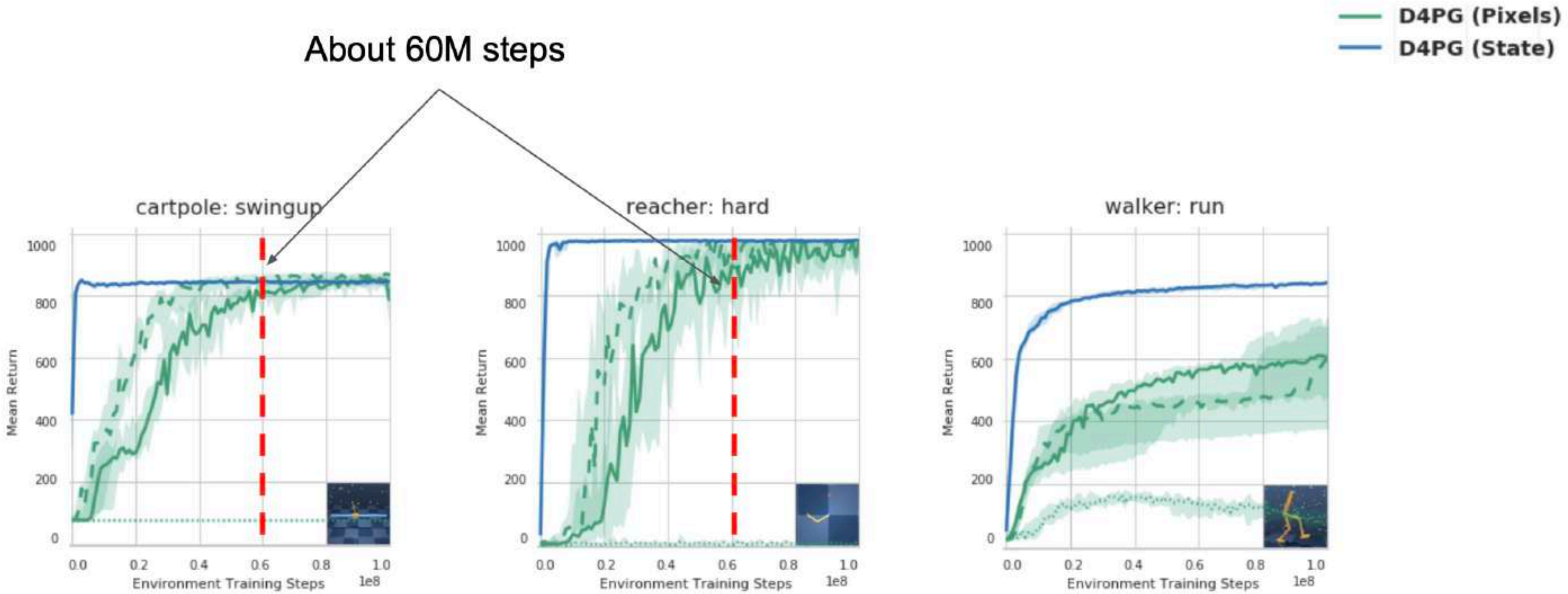
DVC *Pore et al, 2022*

Catch

1. Inefficiency

- millions of transitions (sample inefficient)

Good representations can accelerate learning from images



How do we close this gap?

Building Machines that Learn and Think like People, Lake et al, 2017, *Deepmind Control Suite, Tassa et al, 2018*

Catch

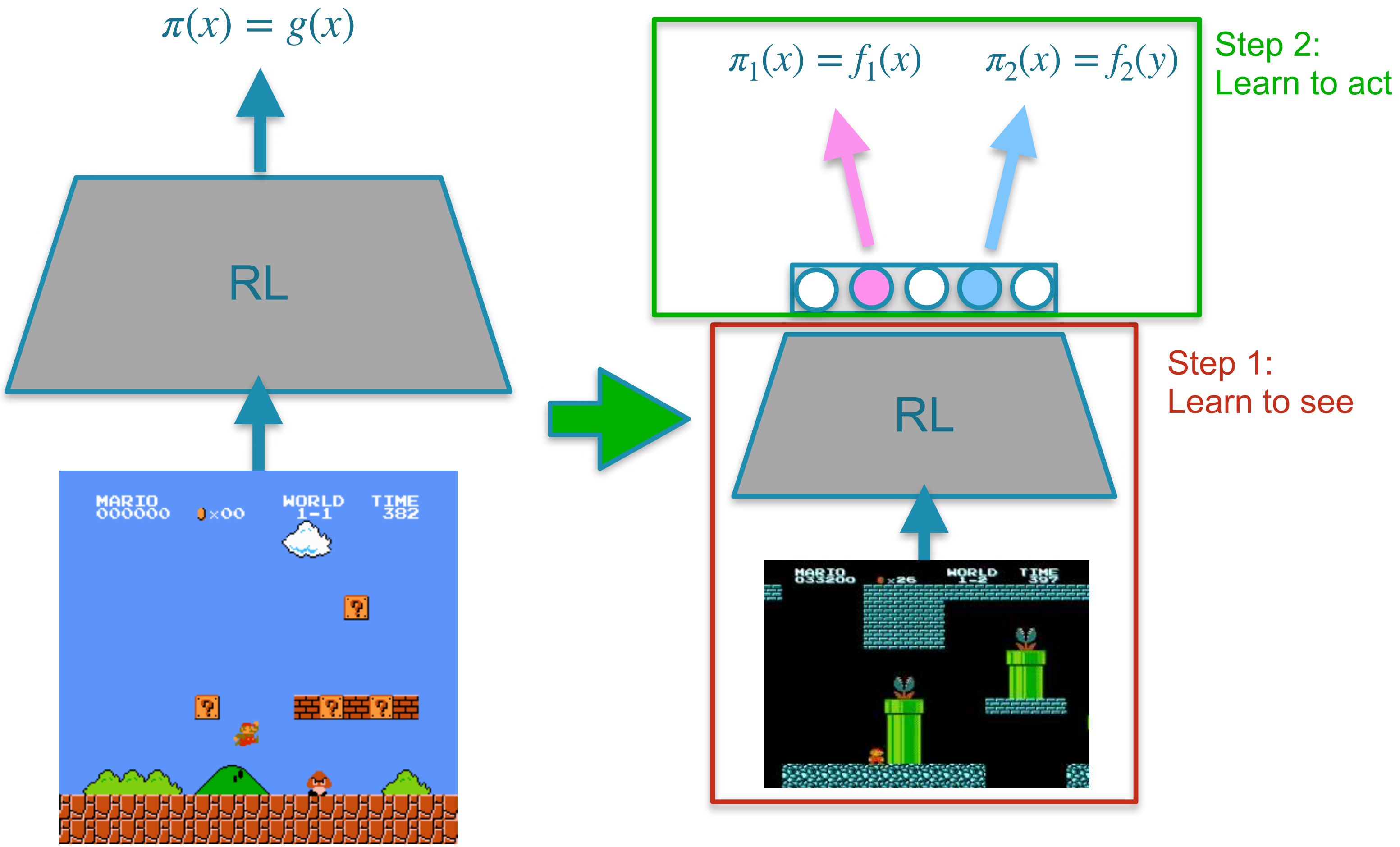
1. Inefficiency
   - millions of transitions (sample inefficient)

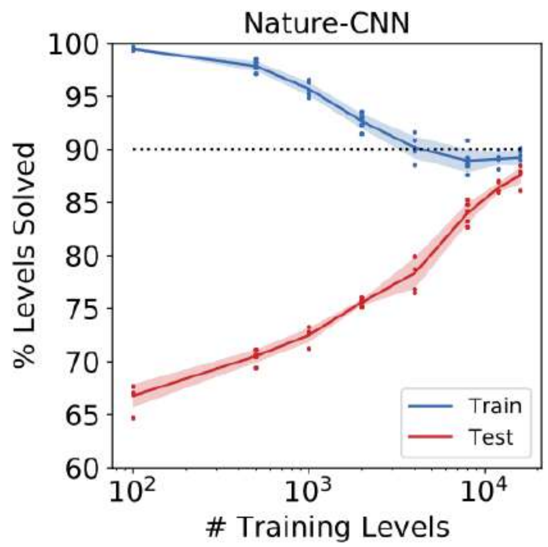
   → Good representations can accelerate learning from images

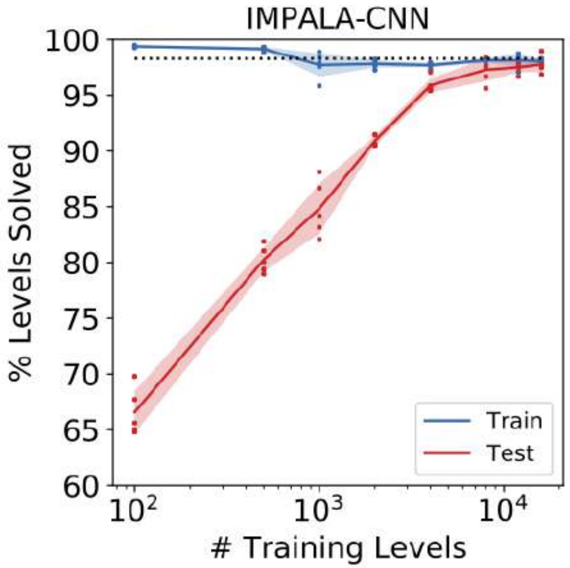
2. Generalisation
   - Works really well in single task setting

   → Good representations can generalise well across different tasks, or quickly adapt to new tasks

Generalisation







Quantifying generalisation in Reinforcement Learning, Cobbe et al, 2019

Catch

1. Inefficiency
   - millions of transitions (sample inefficient)
2. Generalisation
   - Works really well in single task setting
3. Requires lots of supervision
   - Dense reward function
   - Effective exploration is challenging in many RL tasks

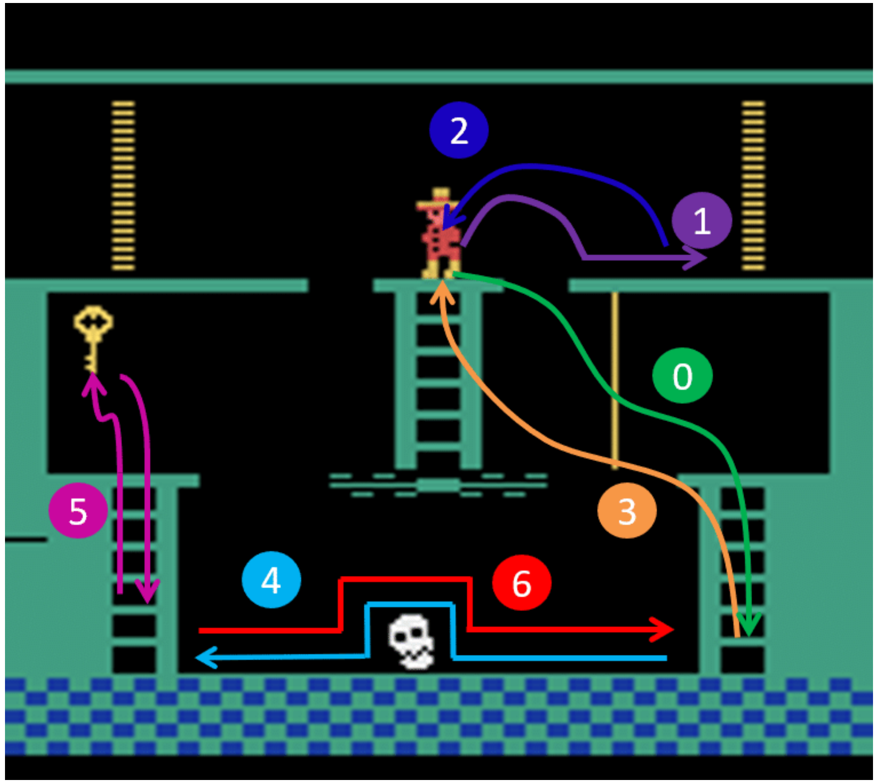
Good representations can accelerate learning from images

Good representations can generalise well across different tasks, or quickly adapt to new tasks

Instead of only learning from reward signals, we can also learn from unsupervised collected data.

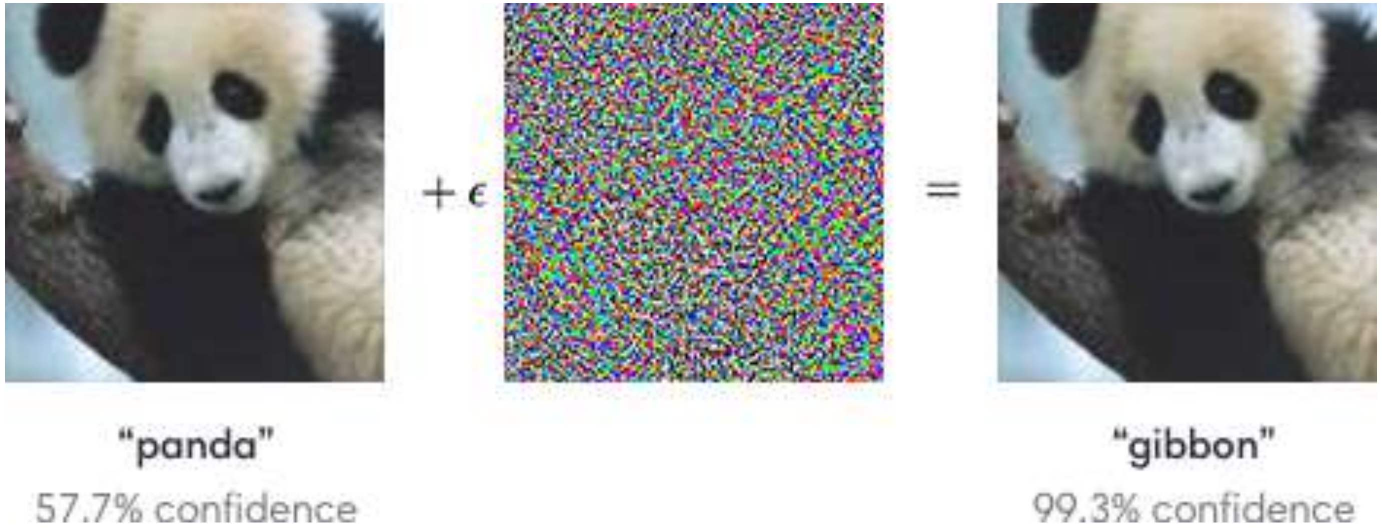
Good representations can accelerate exploration

Sparse reward and exploration



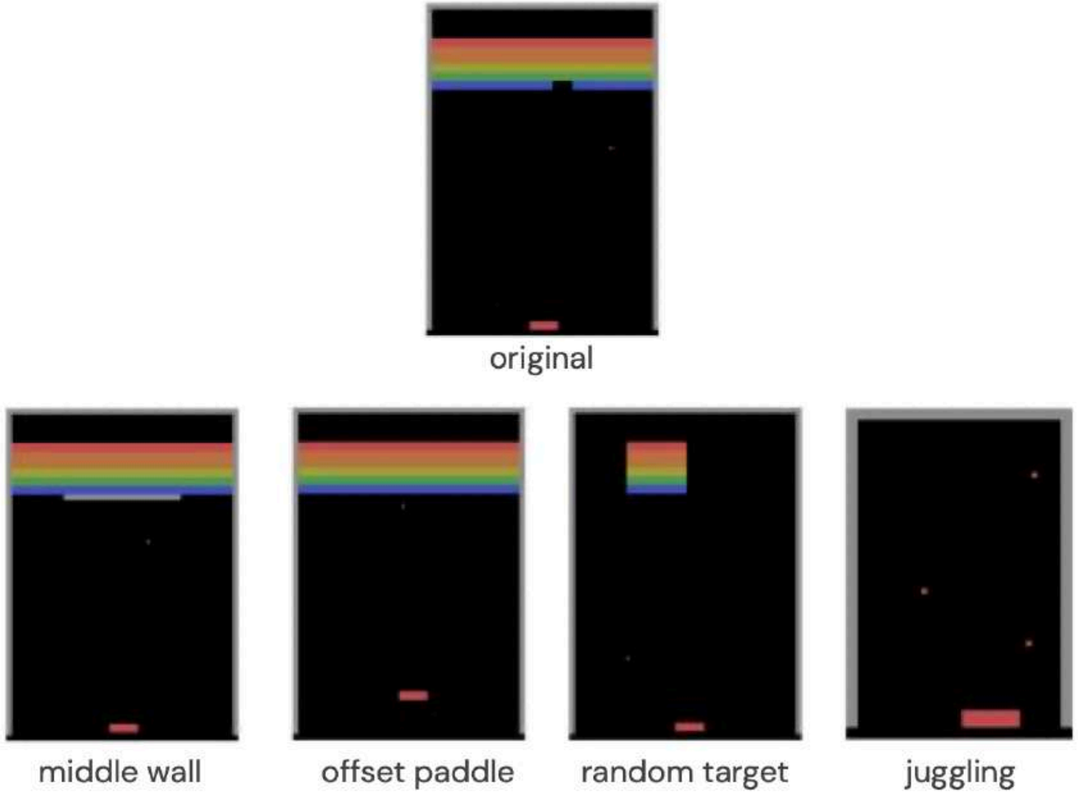
- End-to-end not preferred with sparse reward
- Need to explore novel/new states

Robustness



Explaining and harnessing adversarial examples, Goodfellow et al, ICLR 2015

Transferability

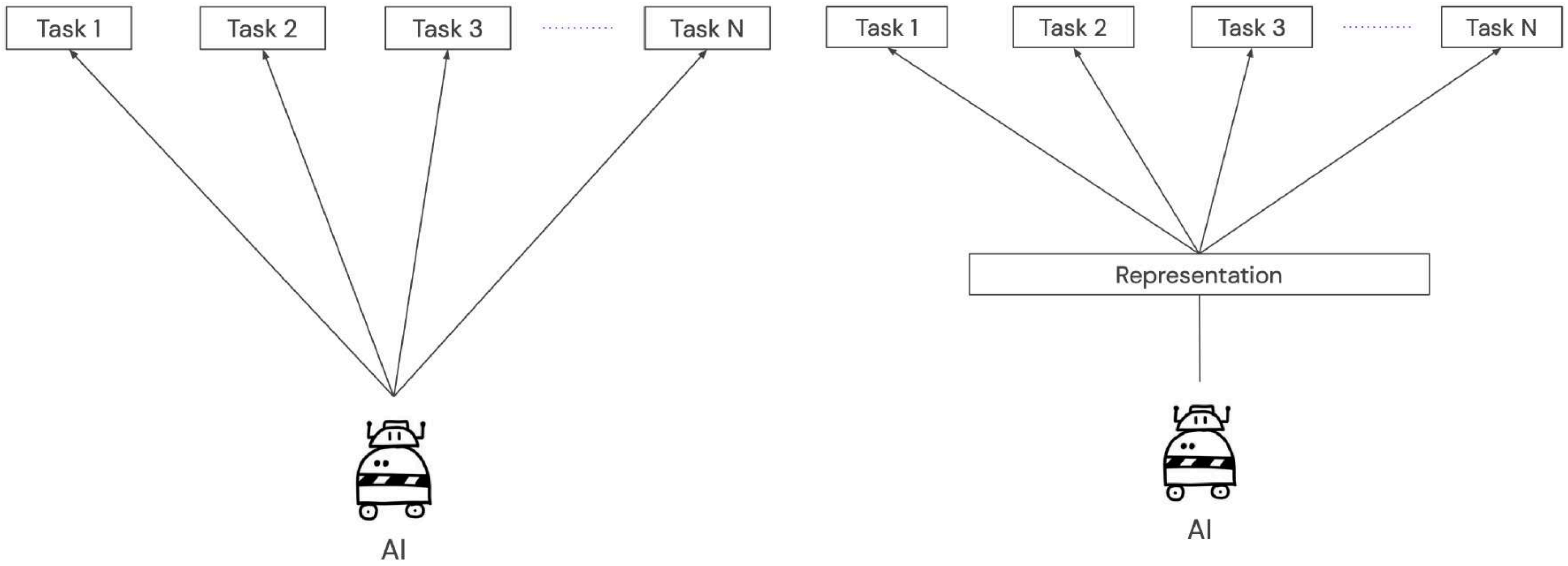


	Standard Breakout	Offset Paddle	Middle Wall	Random Target	Juggling
A3C Image Only	(36.33 ± 6.17)	0.60 ± 20.05	9.55 ± 17.44	6.83 ± 5.02	-39.35 ± 14.57

Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics, Kansky et al, ICML 2017

Desired
- General
- Robust
- Useful
- Reusable
- Flexible
- Compositional
- Interpretable

Representation learning for RL



What is a representation?

"Formal system for making explicit certain entities or types of information, together with a specification of how the system does this"

- Marr and Nishihara, 1978

XXXVII 37 0b100101

- Representational form orthogonal to the information content
- Useful abstraction to make different computations more efficient
- Not defined by a single piece of information but rather by the shape of the manifold on which the data lie within the representational space

Representation Learning

"… learning representations of the data that make it easier to extract useful information when building classifiers or other predictors" — Bengio et al. [2013]

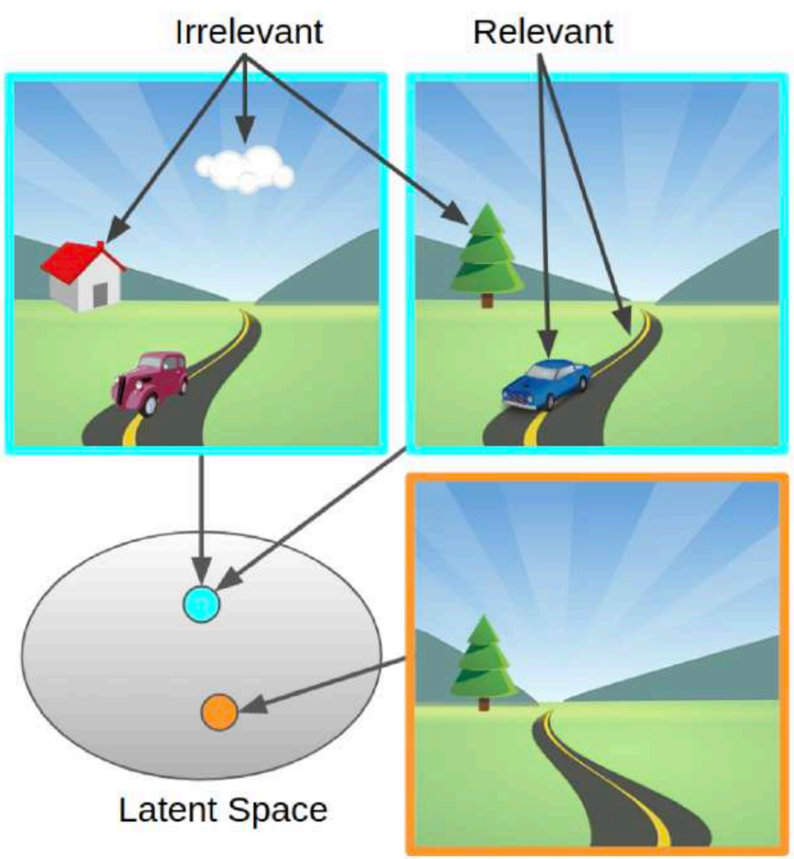
"Is a way of injecting some (hopefully useful) inductive bias in the features" — anonymous

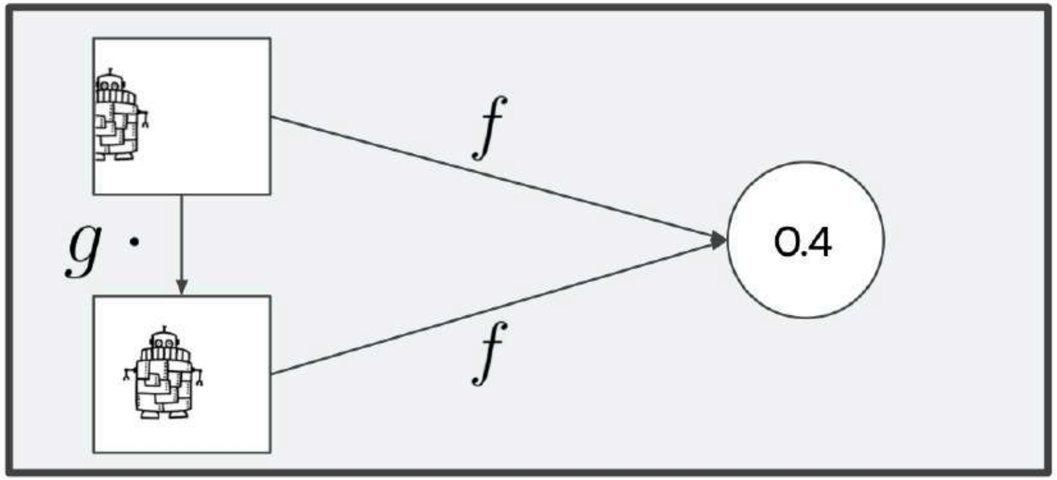
"Is a way of making Reinforcement Learning more efficient" — anonymous

Representation Learning: A review and new perspective, Bengio et al, 2013

What are good representations?

Invariant Representations





Invariance

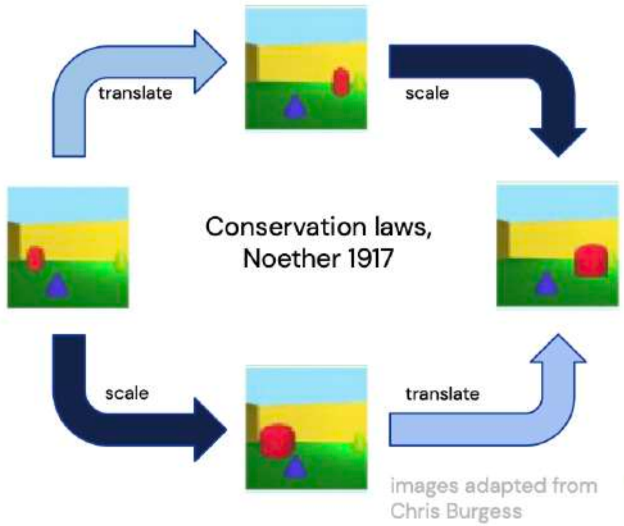
– representation remains unchanged when a certain type of transformation is applied to the input

$$f(g \cdot x) = f(x)$$

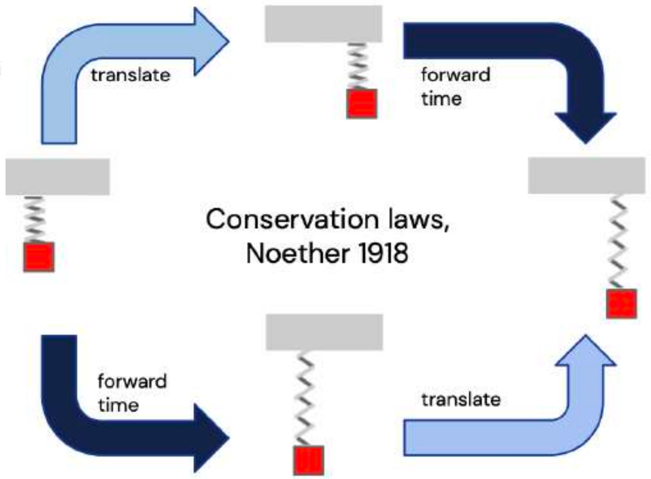
Learning Invariant Representations for Reinforcement Learning without Reconstruction, Zhang et al, 2021

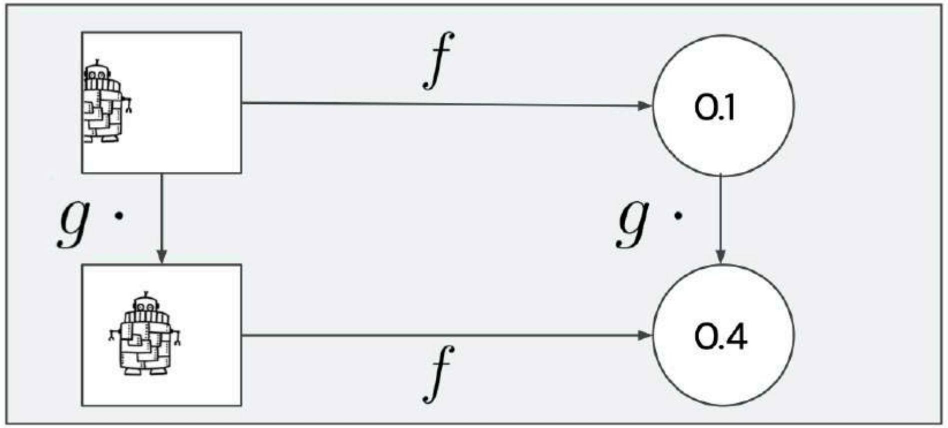
What are good representations?

Equivariant Representations



Symmetry



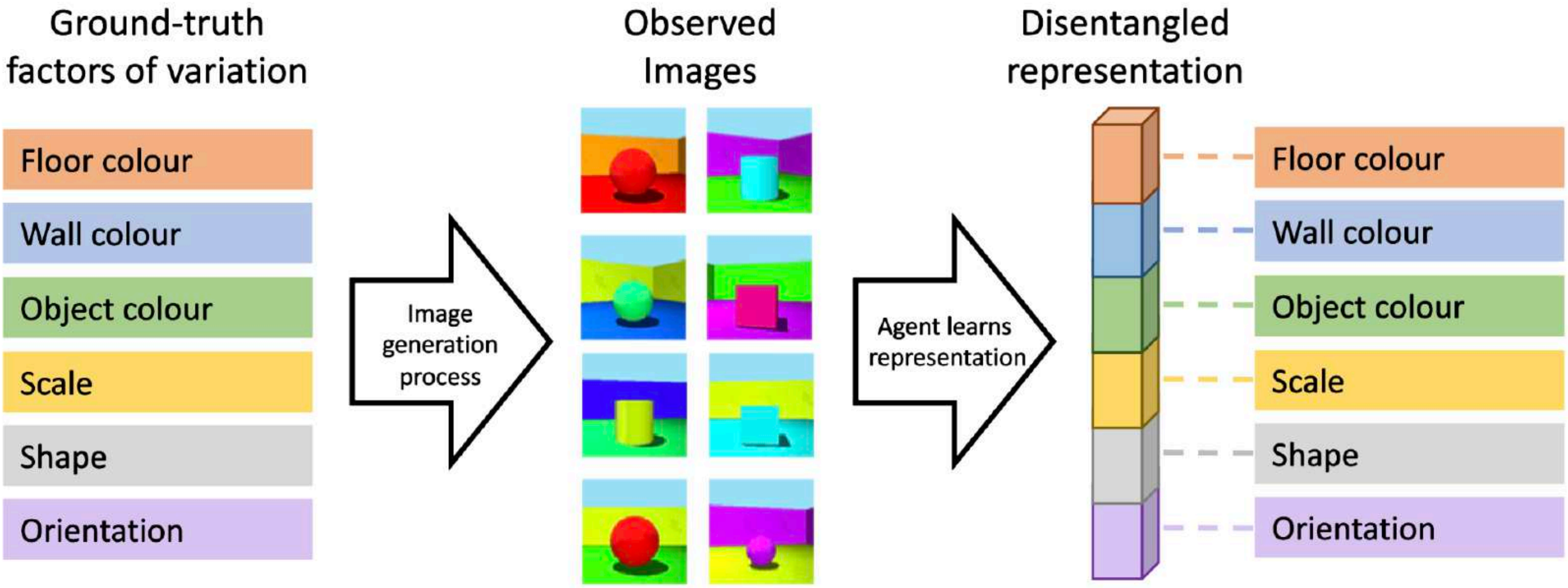


Equivariance

– representation reflects the transformation applied to the input

$$f(g \cdot x) = g \cdot f(x)$$

Disentangled representation learning

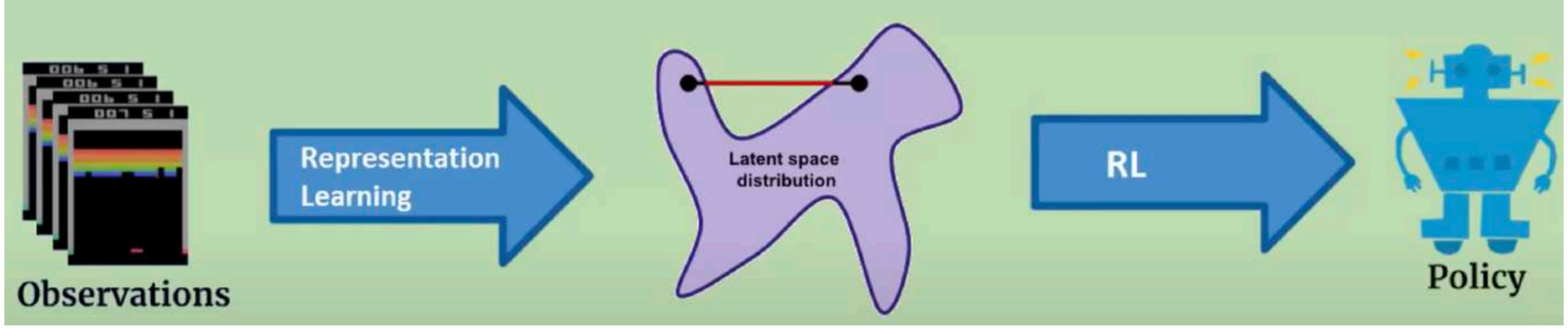


https://agents.inf.ed.ac.uk/blog/disentangled-representations-rl/

Towards a Definition of Disentangled Representations, Higgins et al., 2018

Representation learning for RL

We assume the learner has access to a representation space $\mathscr{F}$

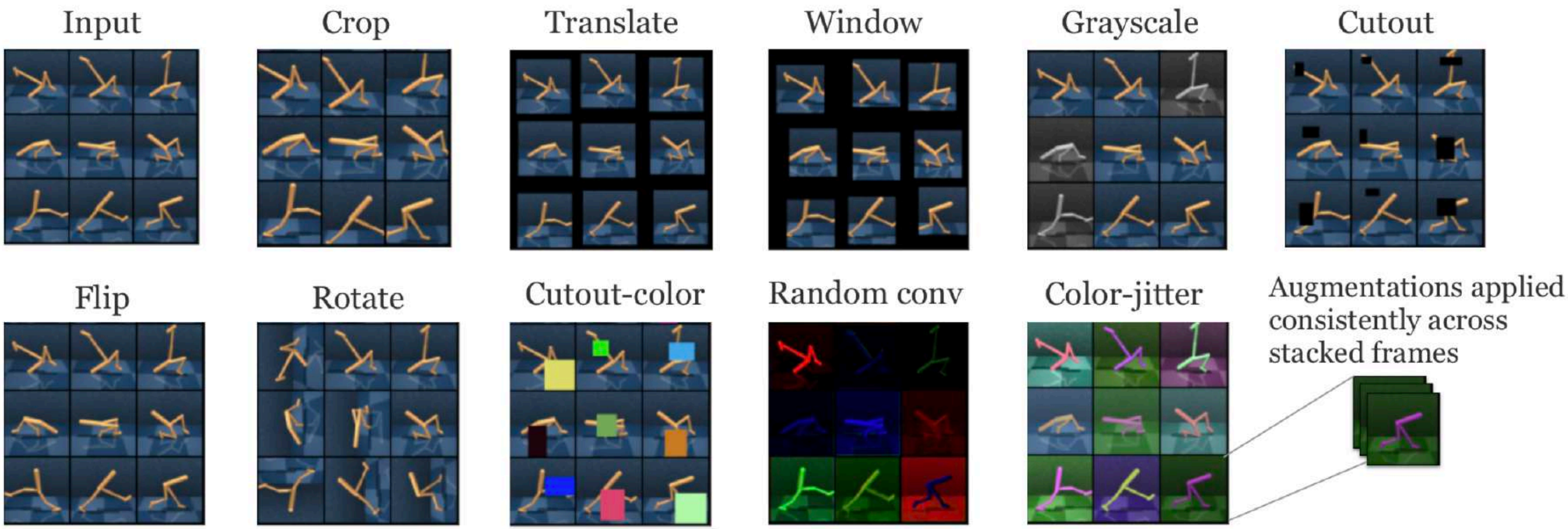


$$\forall h \in [H] \exists f \in \mathcal{F} \; s.t. \; Q_h^{\star}(s,a) = f(s,a), \; \forall s, a$$

Input: Representation space $\mathcal{F}$
$\mathcal{D}_1 = \emptyset$
for $k = 1, \ldots$ **do**
    ❶ Learn representation $f_k \in \mathcal{F}$
    ❷ Compute (explorative) policy π_k using representation f_k
    Execute policy π_k and add experience to $\mathcal{D}_{k+1}$

Implicit regularisation of the representations

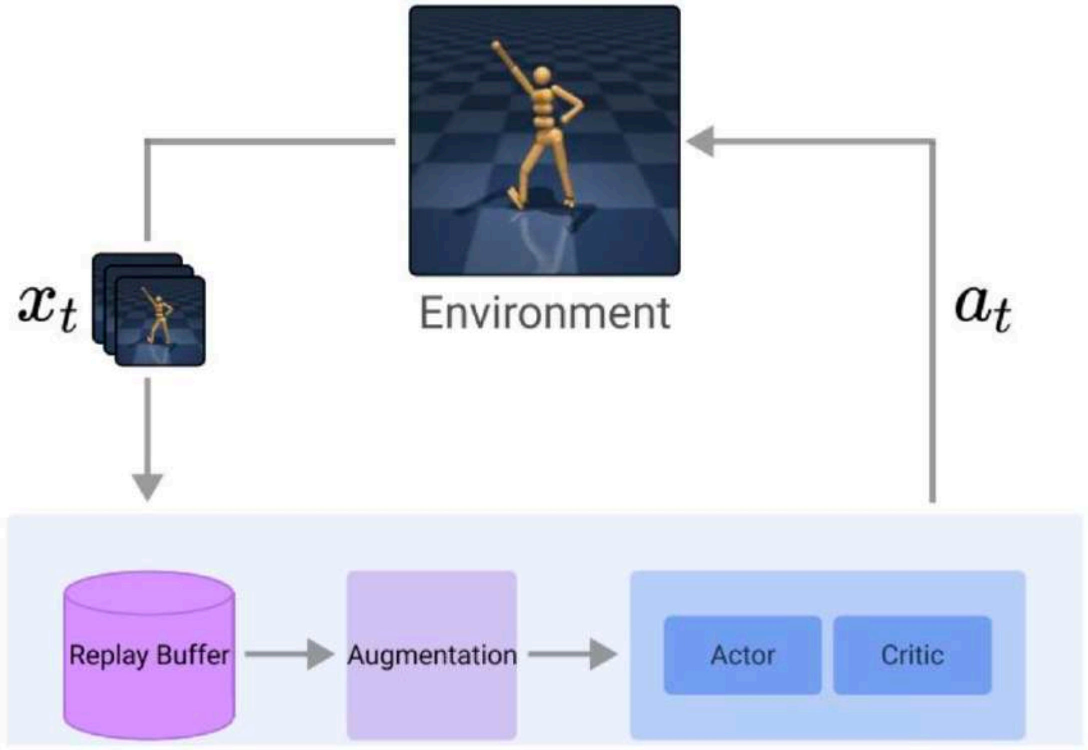
Data Augmentation



Reinforcement Learning with Augmented Data, Laskin et al., NeurIPS 2020

Data Augmentation for RL

Surprisingly data augmentation has been adopted only recently



Issues

- Unclear what are RL-driven data augmentation, in particular in state-based control

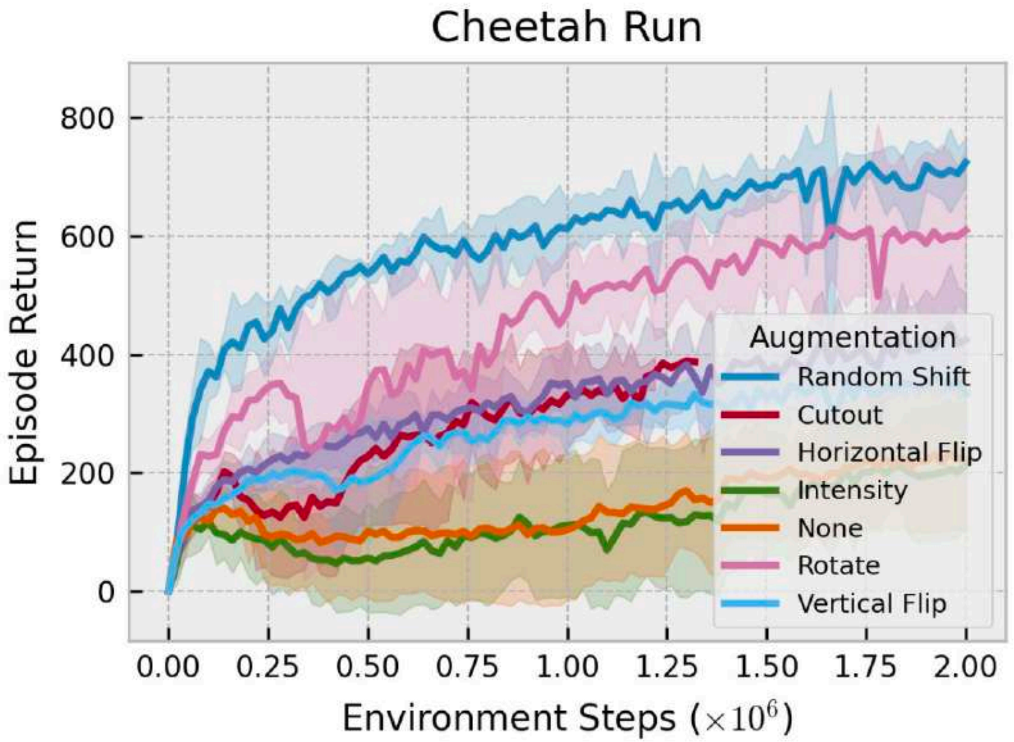
Workaround

- Use standard techniques for images

Image Augmentation Is All You Need: Regularizing Deep Reinforcement Learning from Pixels, Yarats et al., ICLR 2021
Mastering Visual Continuous Control: Improved Data-Augmented Reinforcement Learning, Yarats et al., ICLR 2021

Data Augmentation for RL

Not all standard CV data augmentations can be used in RL

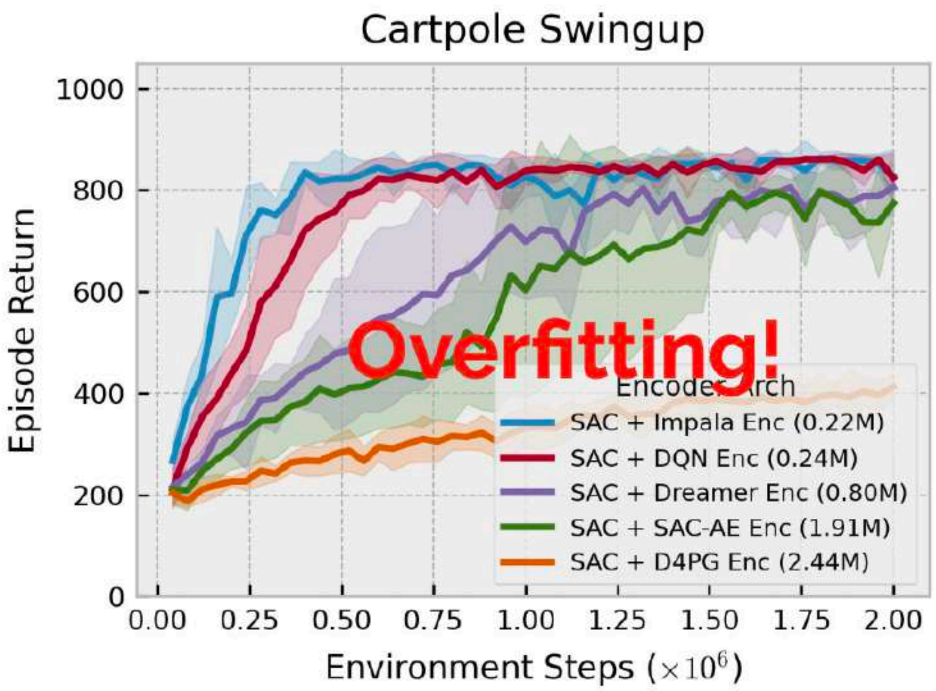


Some recent works in automatic way of selecting augmentation.

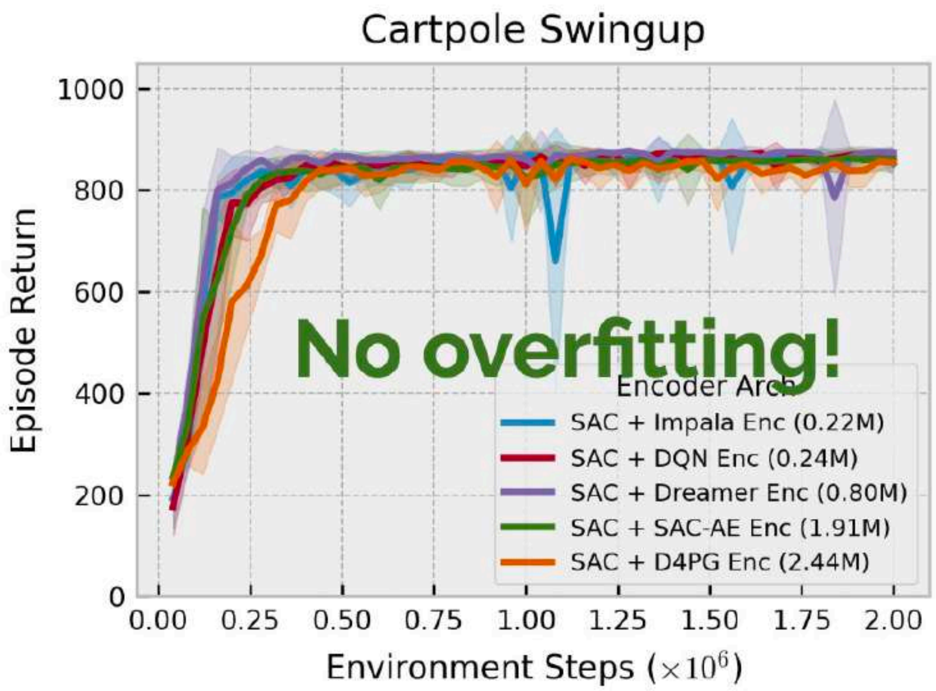
Image Augmentation Is All You Need: Regularizing Deep Reinforcement Learning from Pixels, Yarats et al., ICLR 2021
Automatic Data Augmentation for Generalisation in Reinforcement Learning, Raileanu et al., NeurIPS 2021

Data Augmentation prevents overfitting

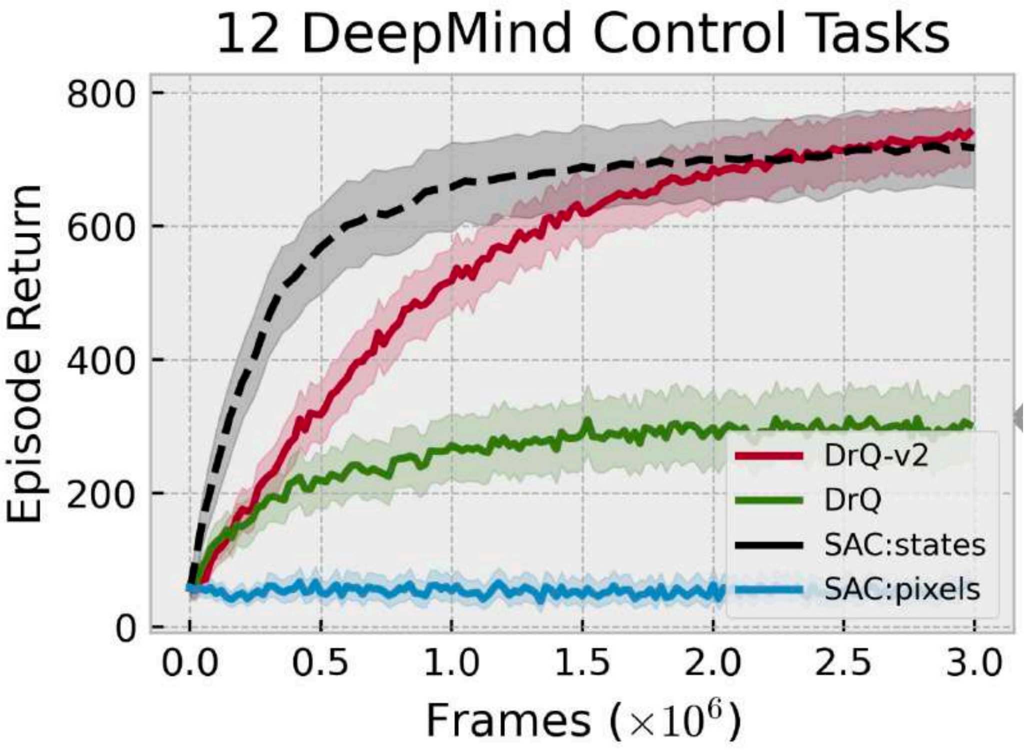
Without DA

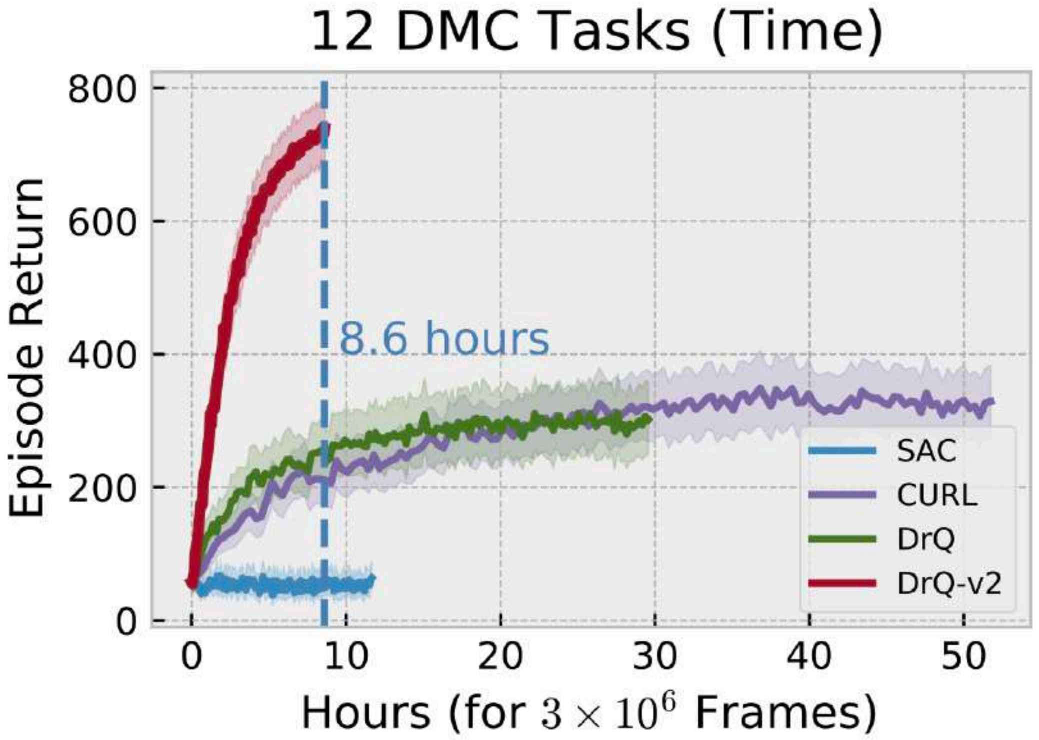


With DA



Data Augmentation works

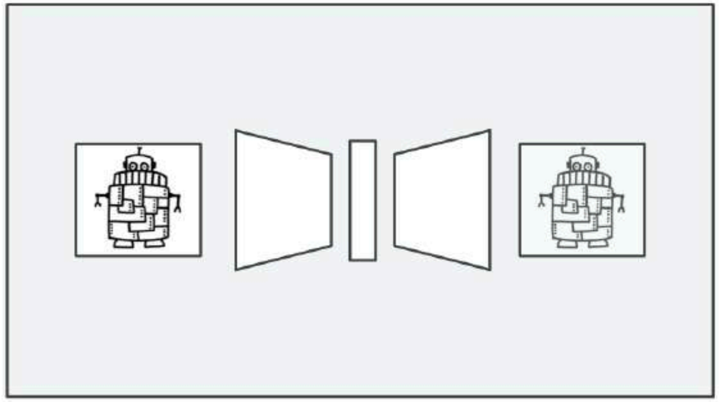


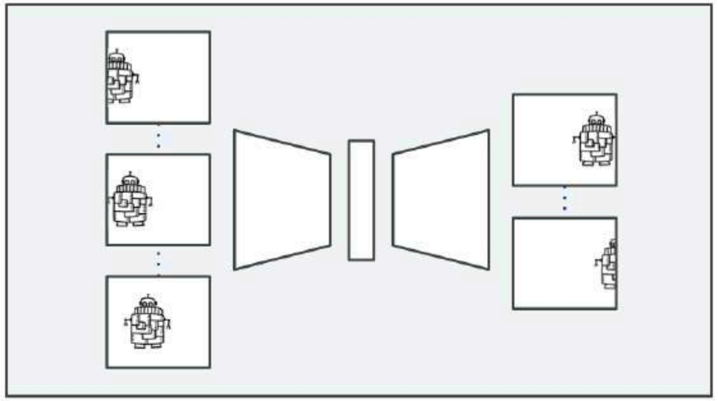


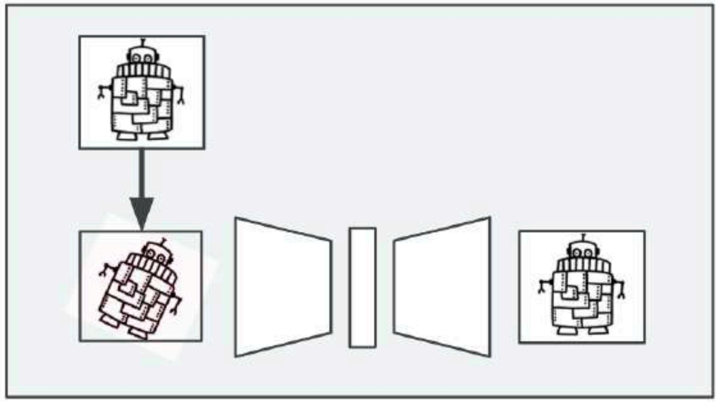
Time for Break!!

ATLAS

Explicit regularisation of representations







Generative modeling

Learn the data distribution using generative modeling, often through reconstructions.

Contrastive losses

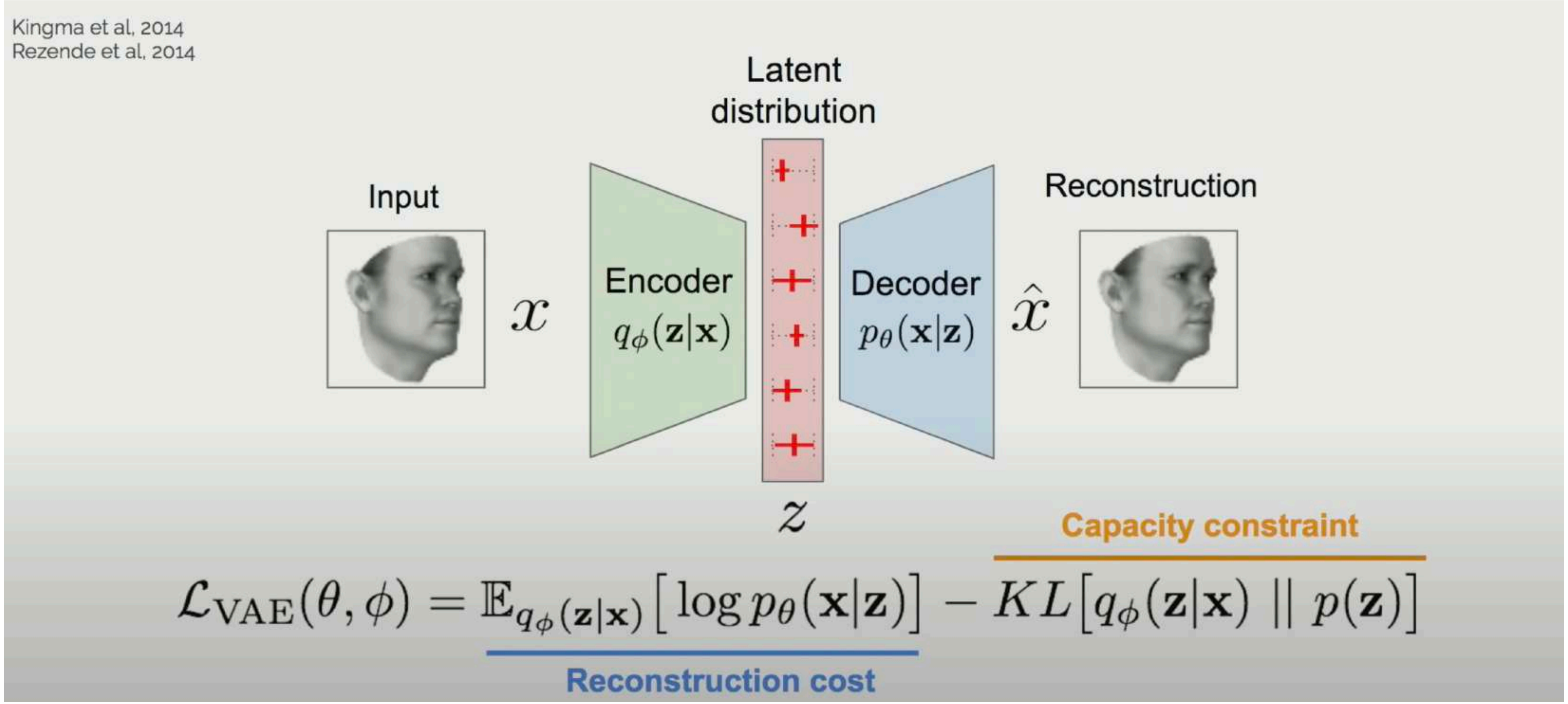
Use classification losses to learn representations that preserve temporal or spatial data consistency.

Self-supervision

Exploit knowledge of data to design learning tasks which lead to useful representations.

Generative Modeling

Variational Autoencoders (VAE)

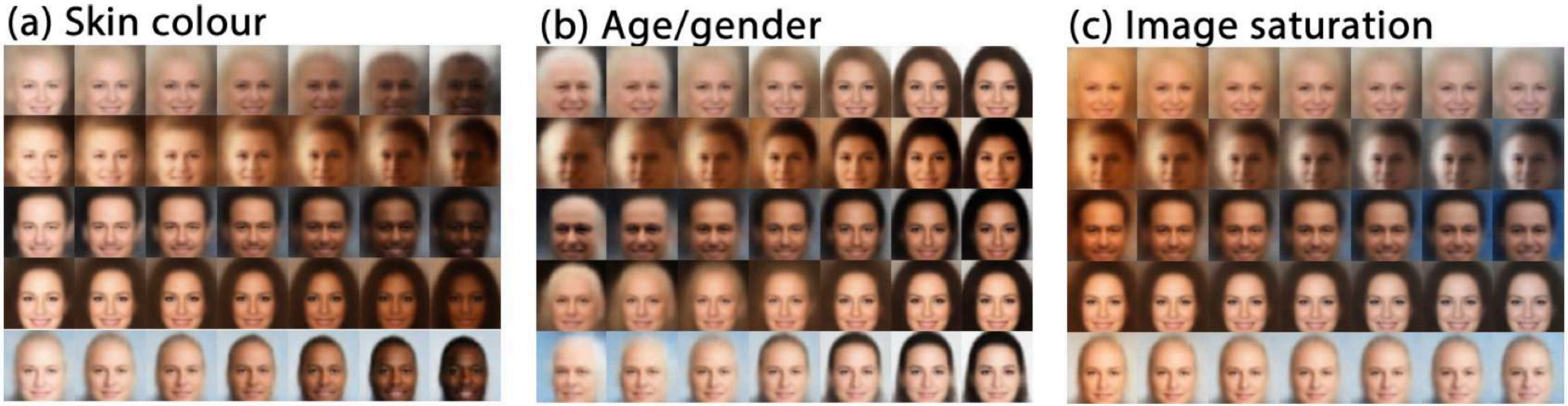


Auto-Encoding variational Bayes, Kingma et al., ICLR 2017

Beta-VAE

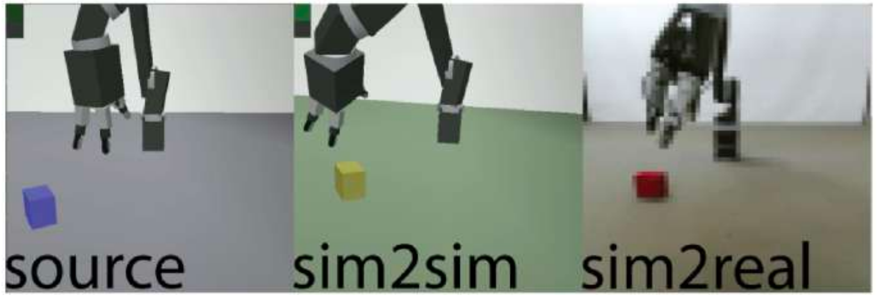
Change the weight of the KL term to encourage disentangled representations

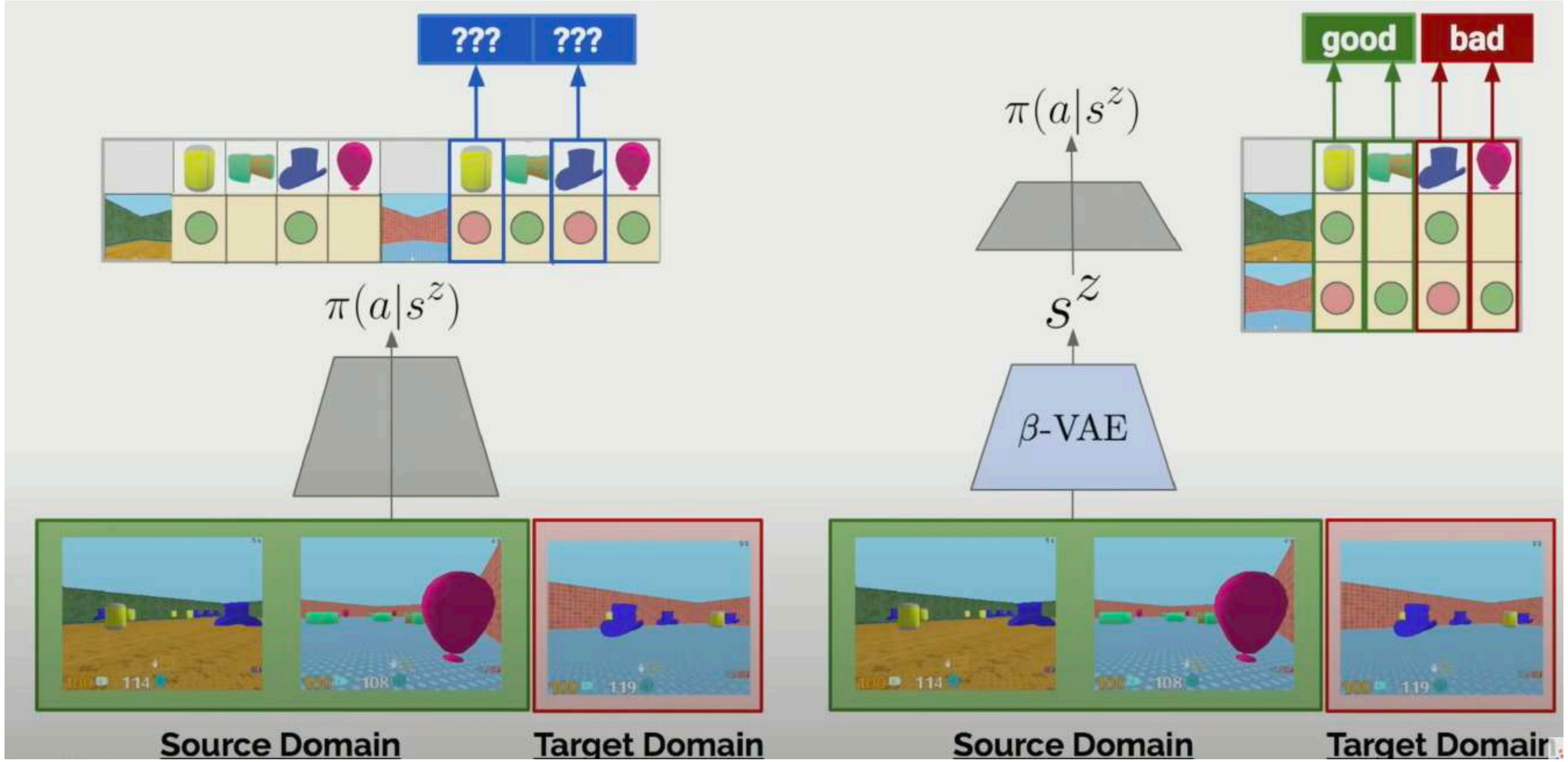
$$\mathbb{E}_{q_\eta(\mathbf{z}|\mathbf{x})} \log p_\theta(\mathbf{x}|\mathbf{z}) - \beta \mathrm{KL}(q_\eta(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))$$



beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework, Higgins et al., ICLR 2017

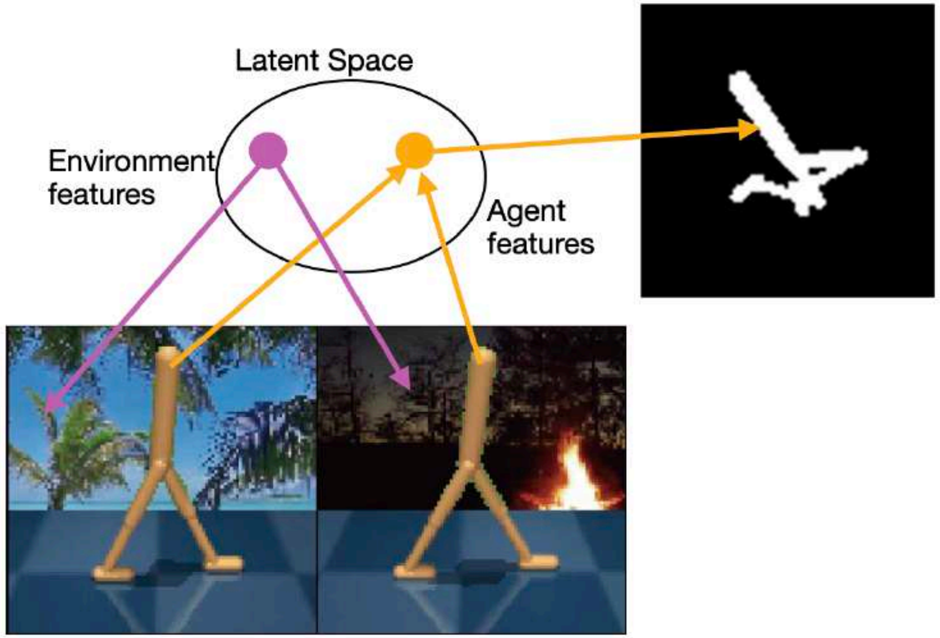
Beta-VAE in RL

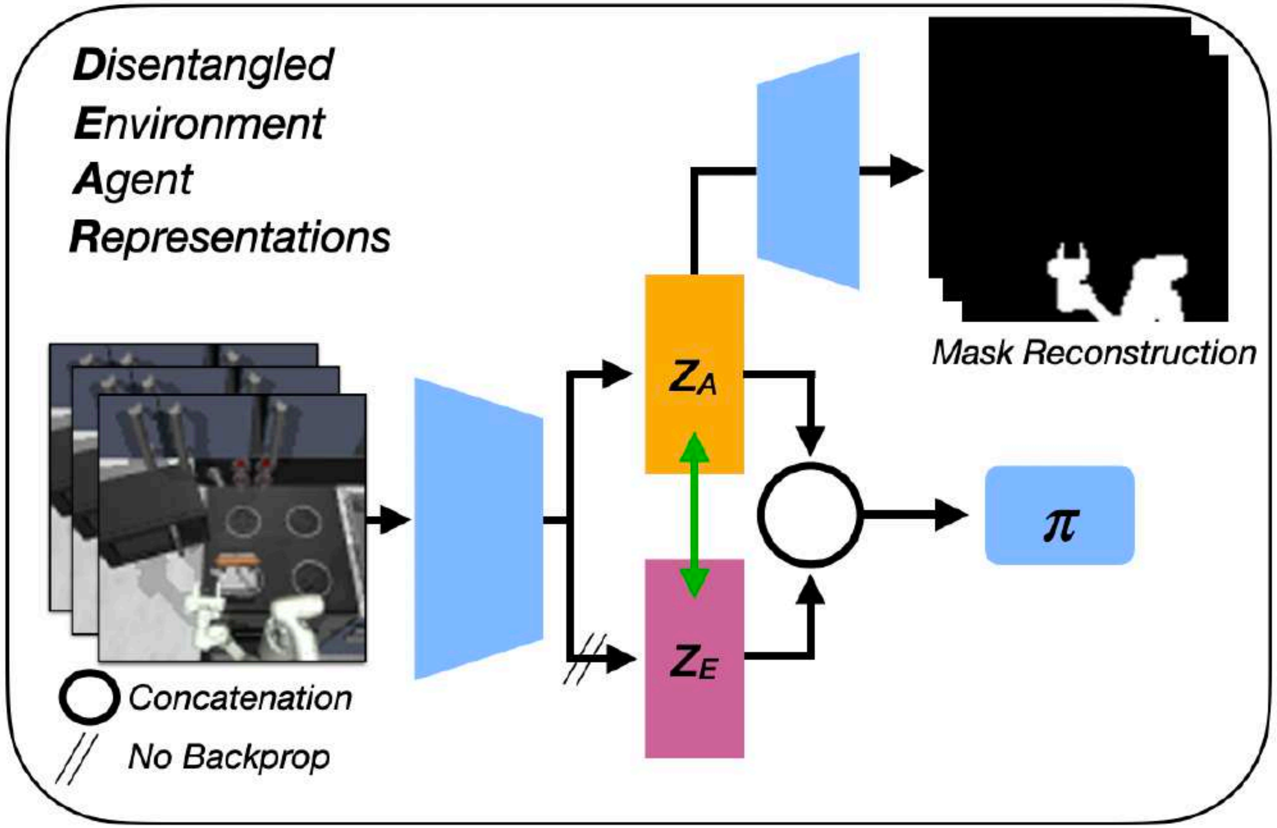




DARLA: Improving Zero-Shot Transfer in Reinforcement Learning, Higgins et al., ICML 2017

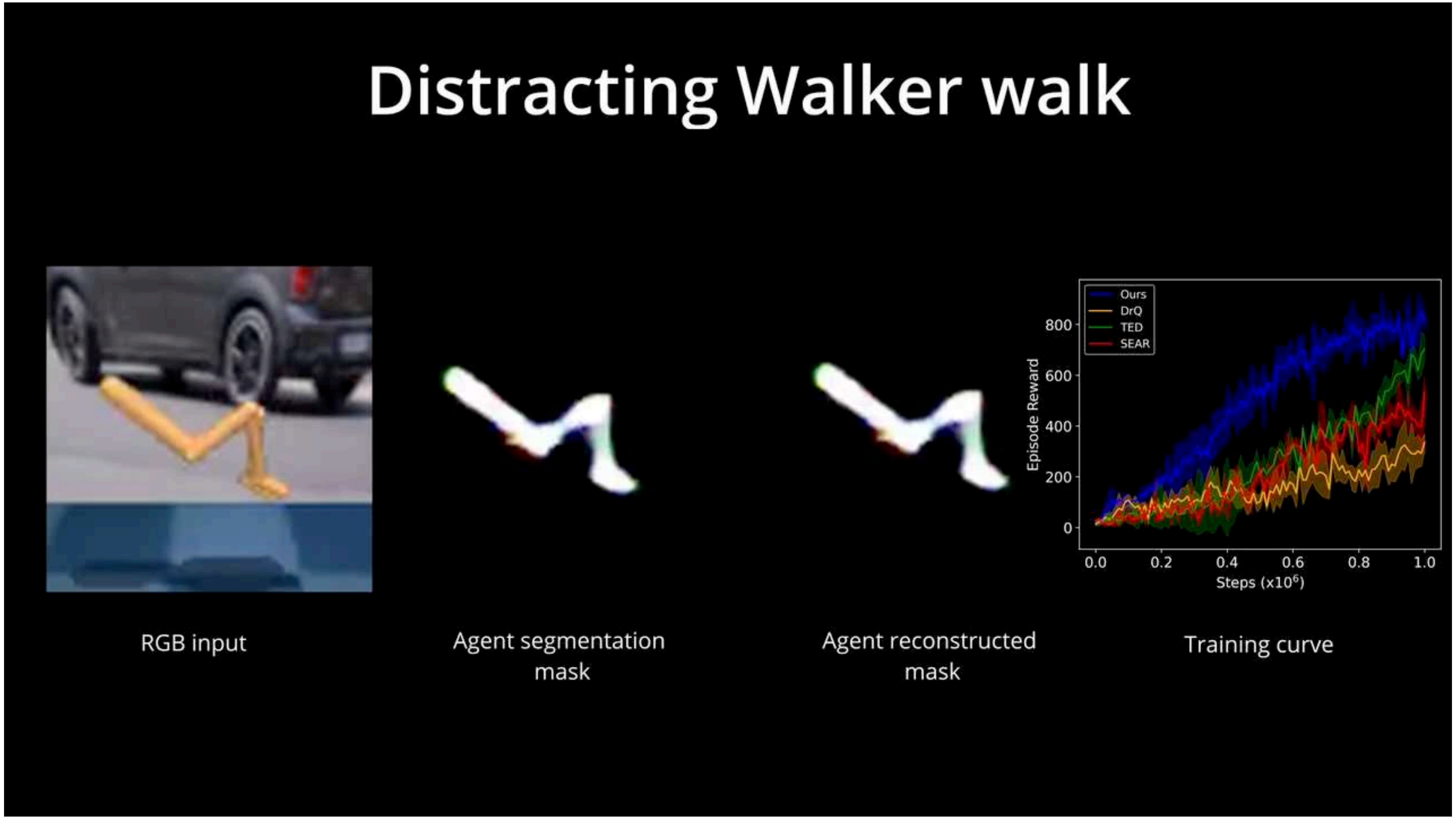
DEAR: Disentangled Env and Agent representations



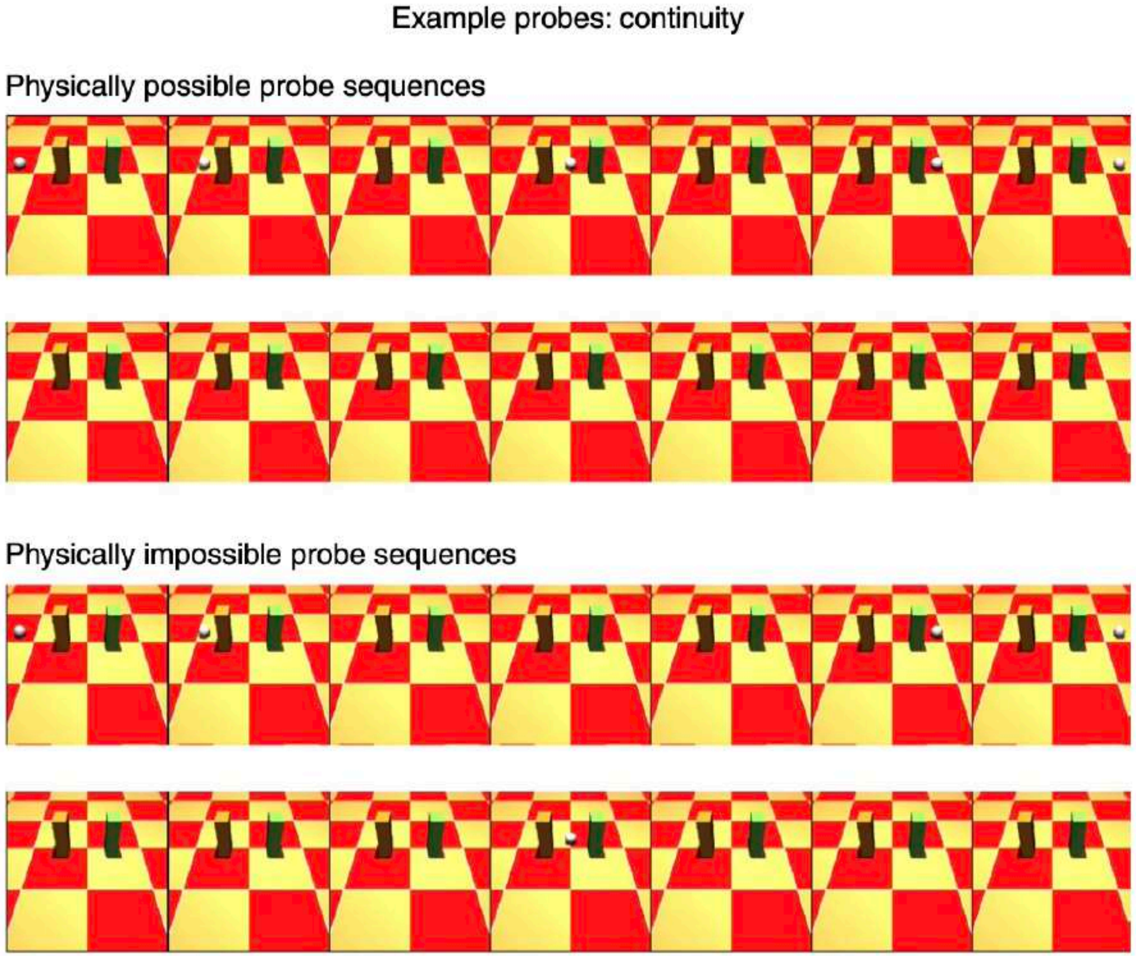


DEAR: Disentangled Environment and Agent Representations for Reinforcement Learning without Reconstruction , Pore et al., 2024

DEAR: Disentangled Env and Agent representations



Learning Intuitive Physics



1. We don’t need to know Conservation of laws of motion to predict the motion of objects
2. Violation of expectation

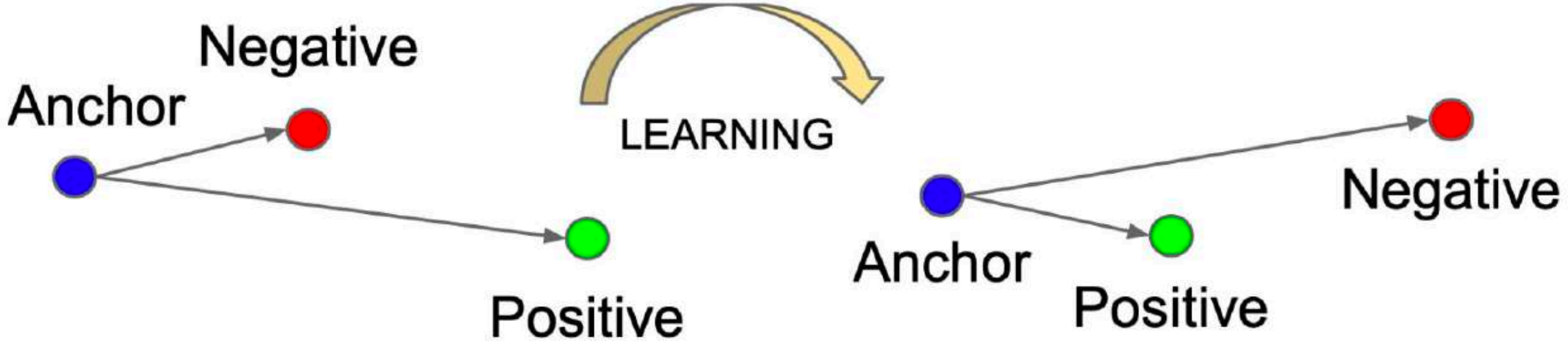
Intuitive physics learning in a deep-learning model inspired by developmental psychology, Piloto et al., Nature Human Behaviour, 2022

Contrastive Learning

Explicit regularisation of the representations

Contrastive learning

1. For an anchor **x**, we are given a positive sample **x+** and a negative sample **x-**
2. The learning objective is to
   - Minimize the distance between the anchor and positive
   - And maximise the distance between the anchor and negative



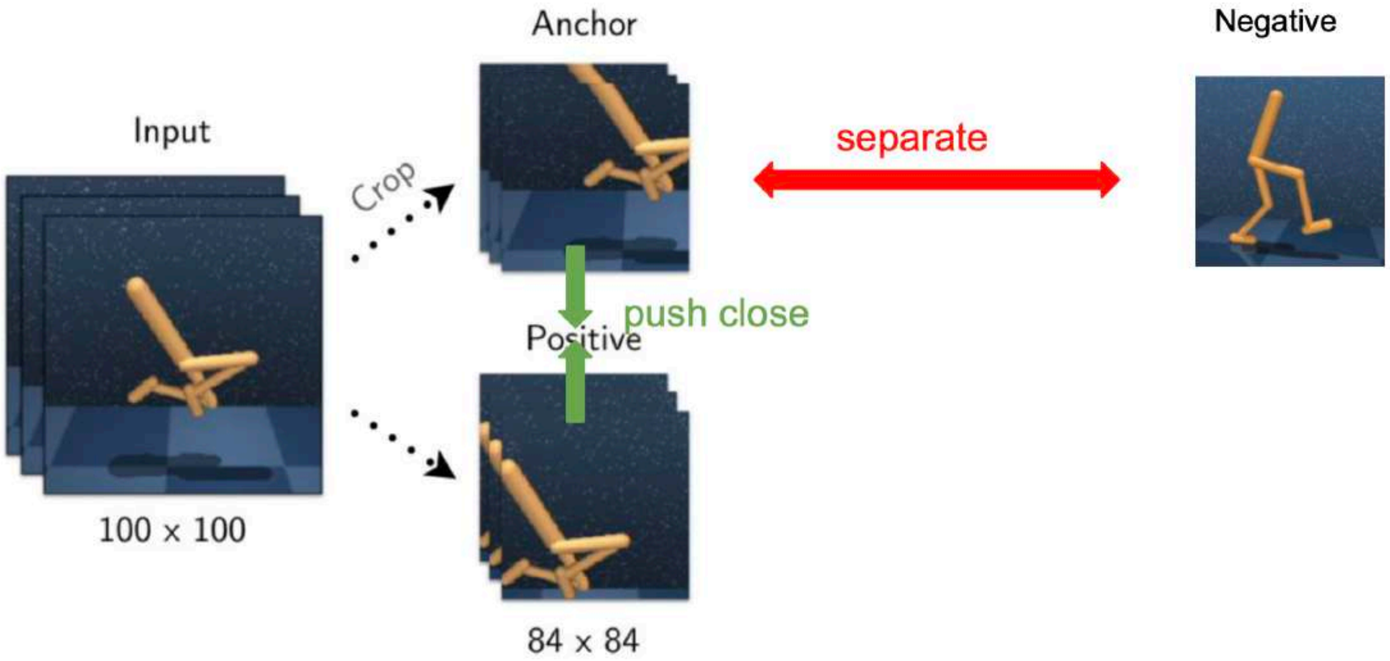
Idea

Learn features that are common between data classes and features that set apart a data class from another.

FaceNet: A Unified Embedding for Face Recognition and Clustering , Shroff et al., 2015

Contrastive learning in RL

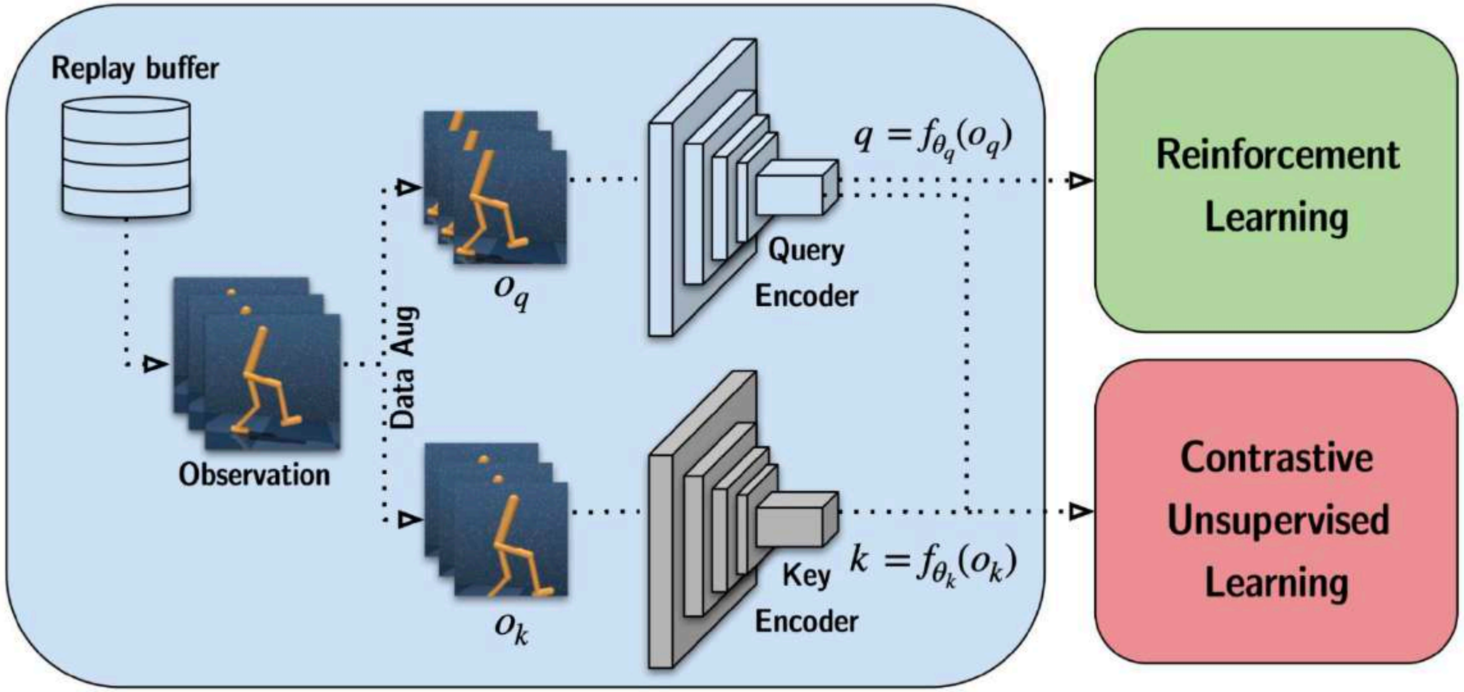
1. Anchor and positive observations are two different augmentations of the same image
2. Negative observations come from other images



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

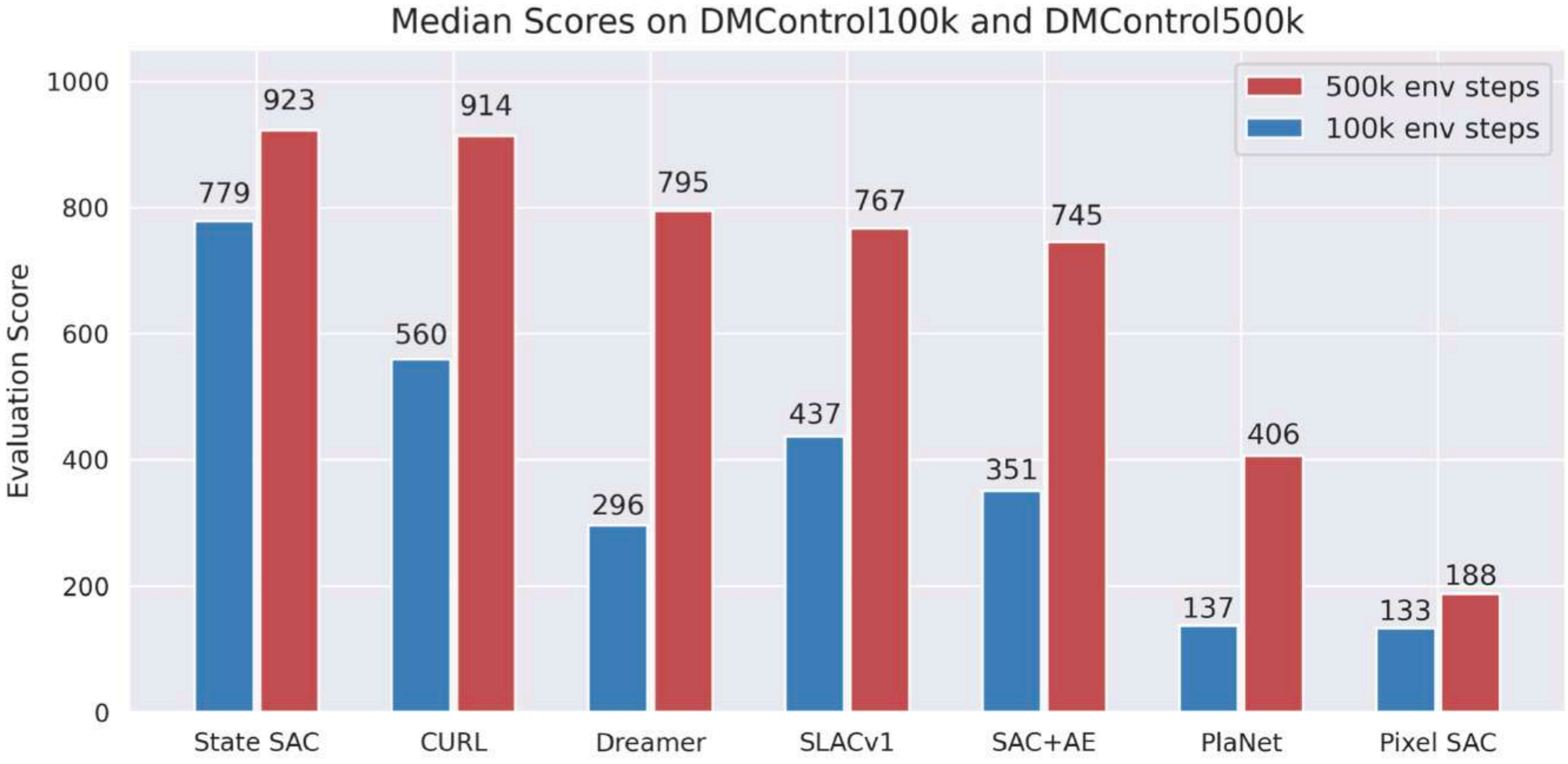
Contrastive learning in RL

1. During the gradient update step, only the query encoder is updated
2. The key encoder weights are the moving average (EMA) of the query weights



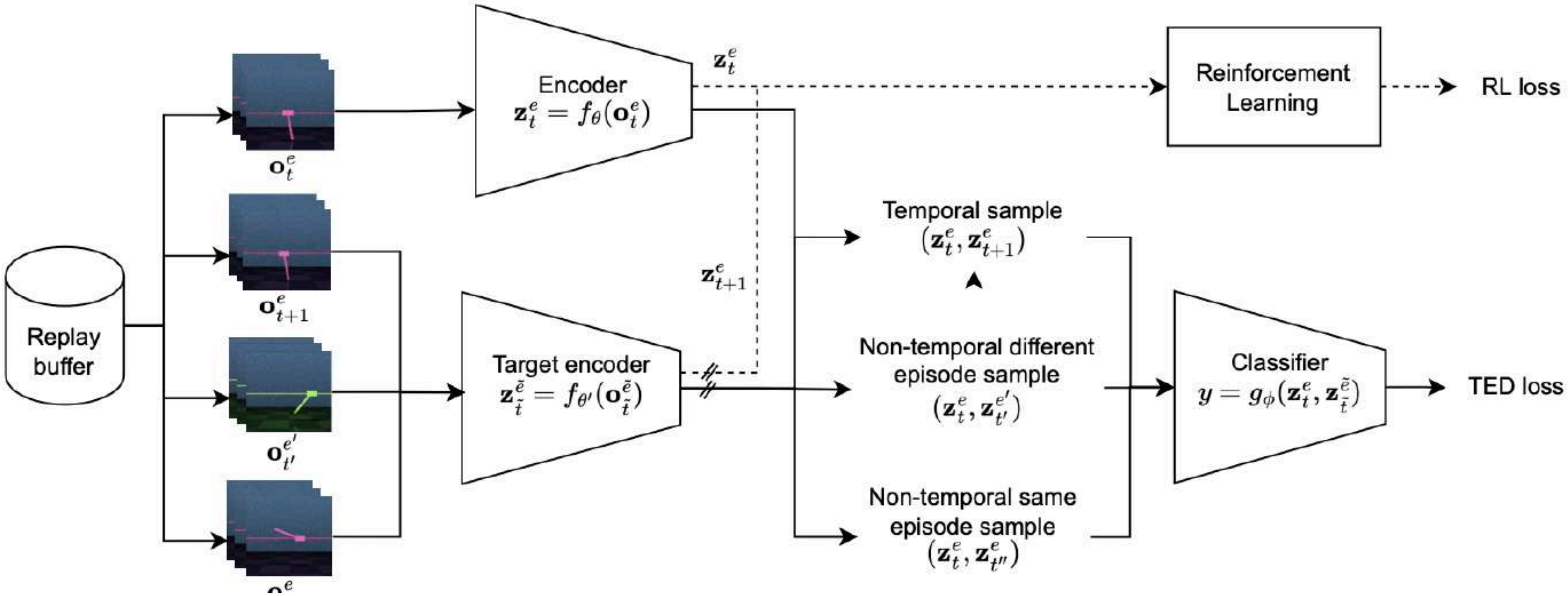
CURL: Contrastive Unsupervised Representations for Reinforcement Learning , Srinivas et al., 2020

Contrastive learning in RL



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

Temporal learning

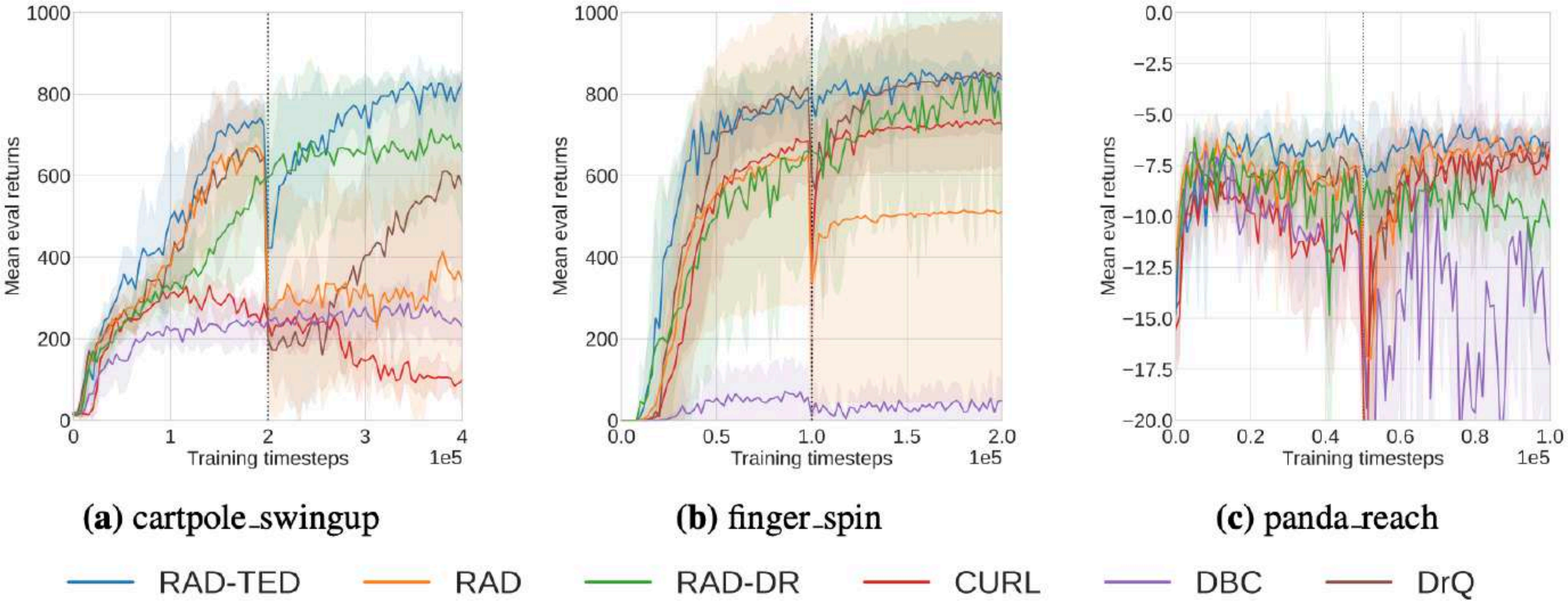


Intuitive Physics (IP):

- Ability to understand and predict physical interactions
- Closer to understanding physical concepts:
  - Object permanence
  - Gravity
  - Momentum

Temporal Disentanglement of Representations for Improved Generalisation in Reinforcement Learning., Dunion et al., ICLR 2023

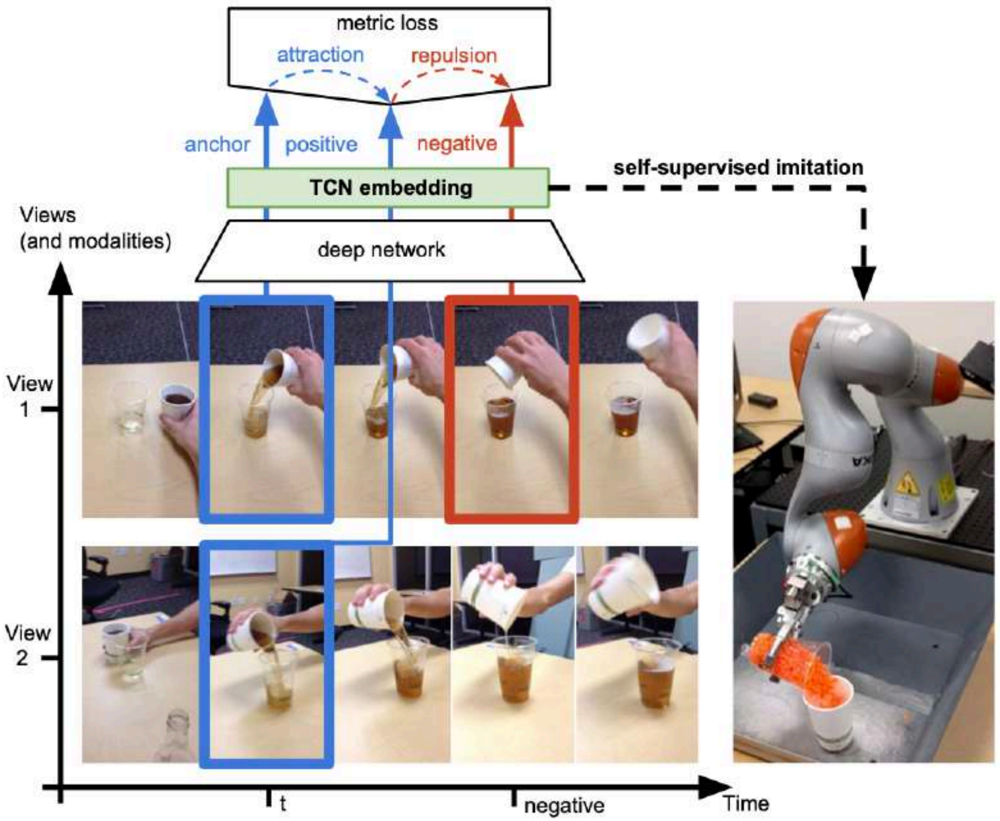
Temporal learning



Temporal Disentanglement of Representations for Improved Generalisation in Reinforcement Learning., Dunion et al., ICLR 2023

Temporal learning

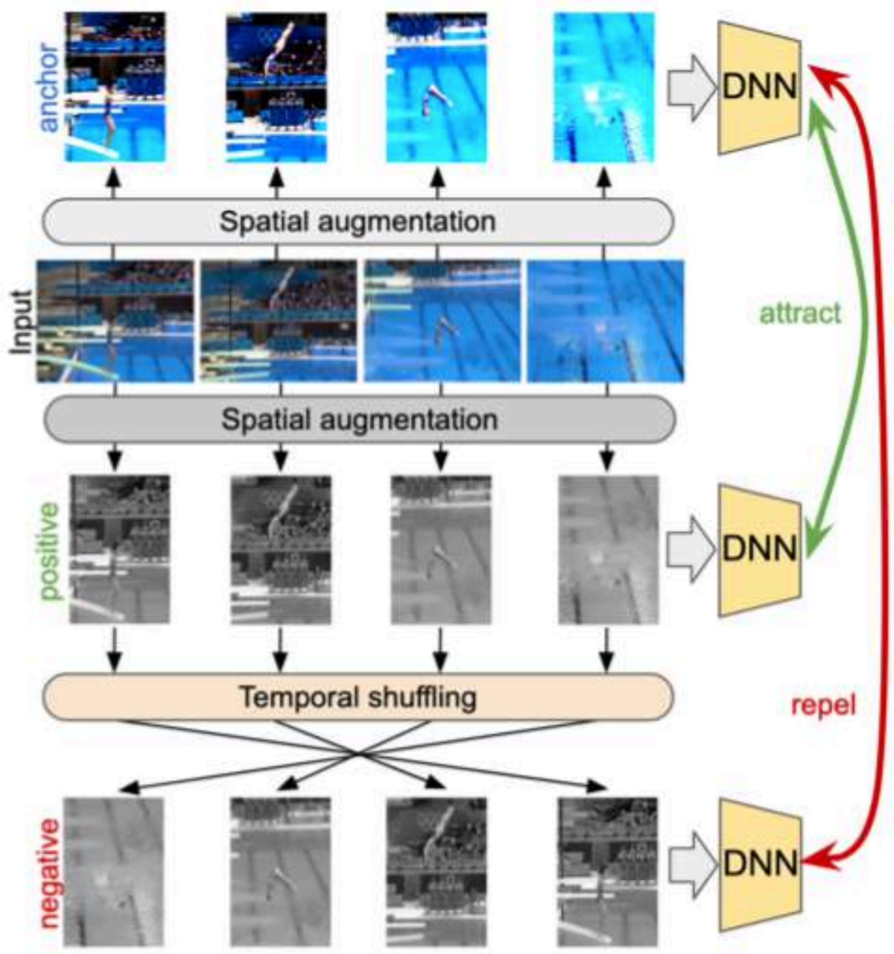
Time contrastive networks



1. Extract features that are invariant to the camera angle and the manipulated objects
2. Reward function based on the distance between the TCN embeddings of human demo and the camera images recorded with robot camera.
3. Video: https://www.youtube.com/watch?v=b1UTUQpxPSY

Time Contrastive Networks: Self Supervised Learning from Video, Levine et al. ,NeurIPS 2017

Temporal learning: shuffling



1. Learns representations that are temporally different
2. Could help RL: Not applied yet

SCVRL: Shuffled Contrastive Video Representation Learning, Dorkenwald et al. ,CVPR 2023

Self-Supervision

World modelling

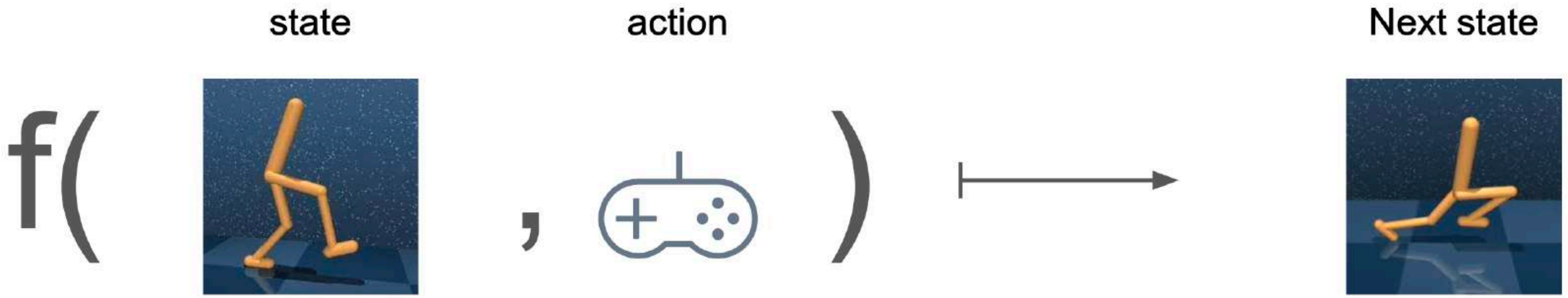
1.Forward

- Predict next state and possibly reward

2.**Inverse**

- Predict the action that generated the transition from s to s’

Forward dynamics modelling

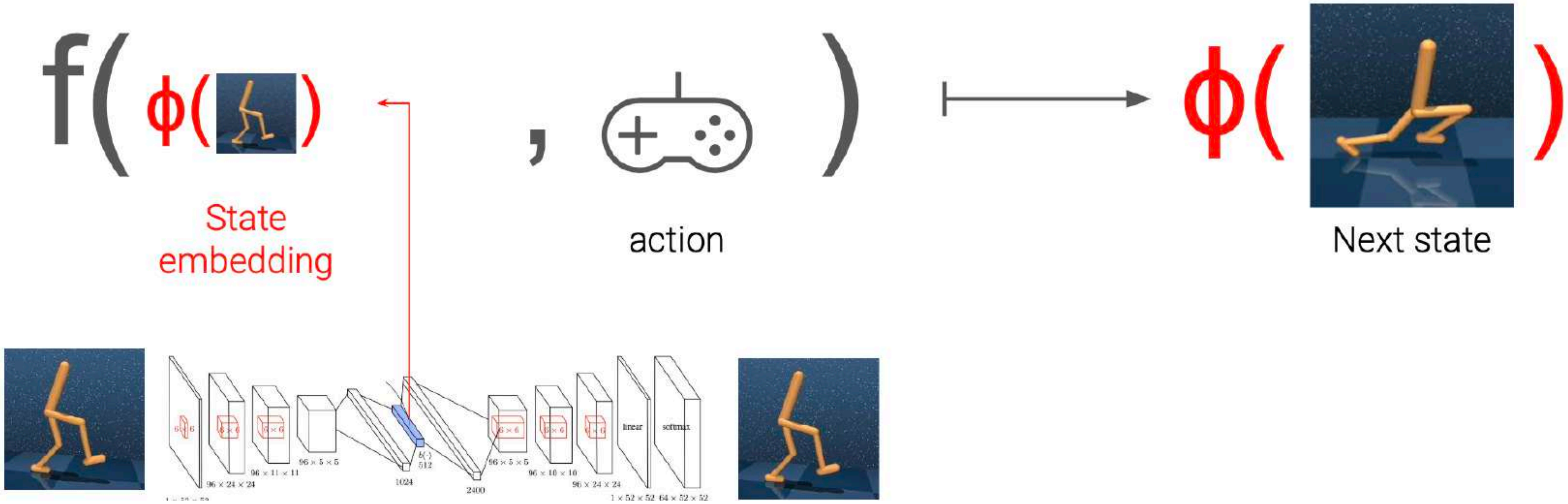


Direct pixel to pixel prediction may be too complicated, better to use a latent representation

Incentivizing Exploration In Reinforcement Learning With Deep Predictive Models, Dorkenwald et al., CORR 2015

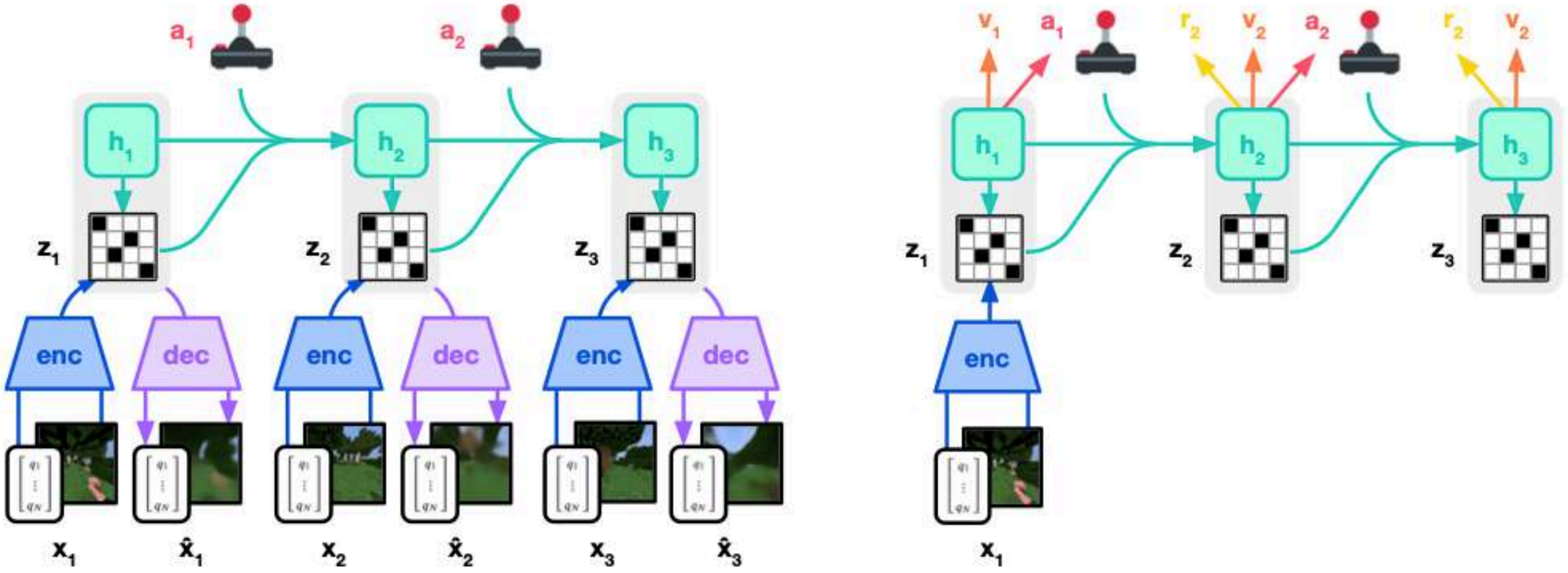
Forward dynamics modelling: latent modelling

A common approach is to extract the latent representation via an auto encoder



A study of count based exploration for deep reinforcement learning, Tang et al., NeurIPS 2017

Example: DREAMER

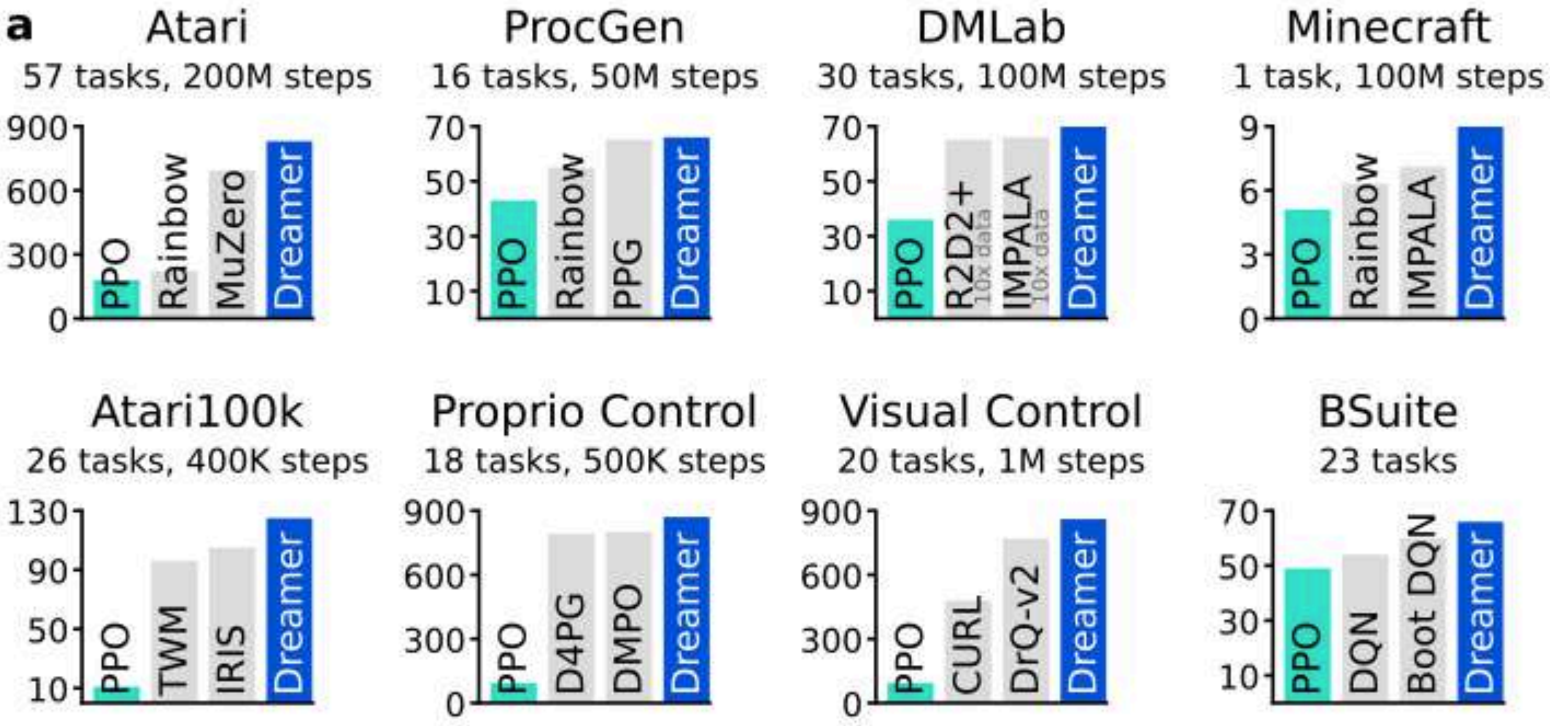


(a) World Model Learning (b) Actor Critic Learning

1. Learn latent space dynamics model
2. Multi-step prediction
3. Planning in latent space

Learning Latent Dynamics for Planning from Pixels, Hafner et al., ICML 2019
Dream to Control: Learning Behaviours by Latent Imagination, Hafner et al., ICLR 2020
Mastering Atari with Discrete World Models, Hafner et al., ICLR 2021
Mastering Diverse Domains through World Models, Hafner 2024

Example: DREAMER



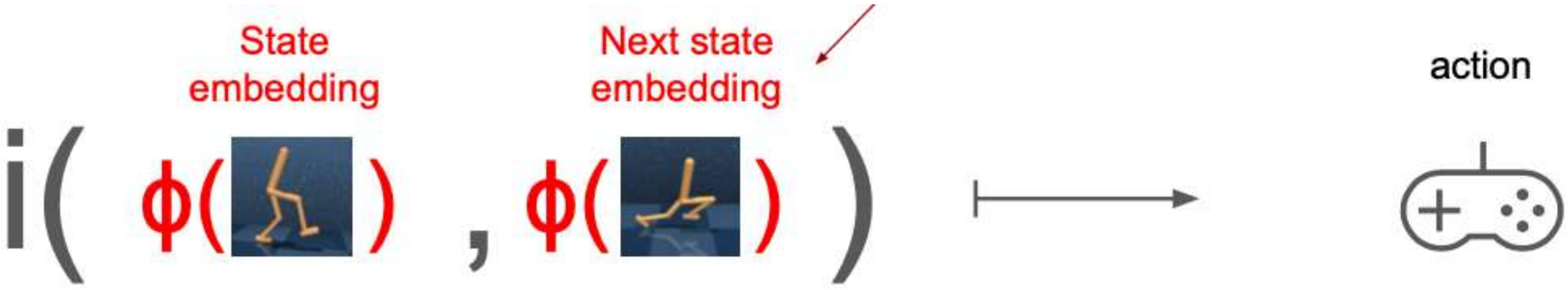
Generate imagined trajectories using dynamics model

Is everything relevant?

- Forward models have to concentrate on each individual pixel to be able to reconstruct the image

- For controllability, we may need to predict only changes that depend on agent's actions, ignore the rest

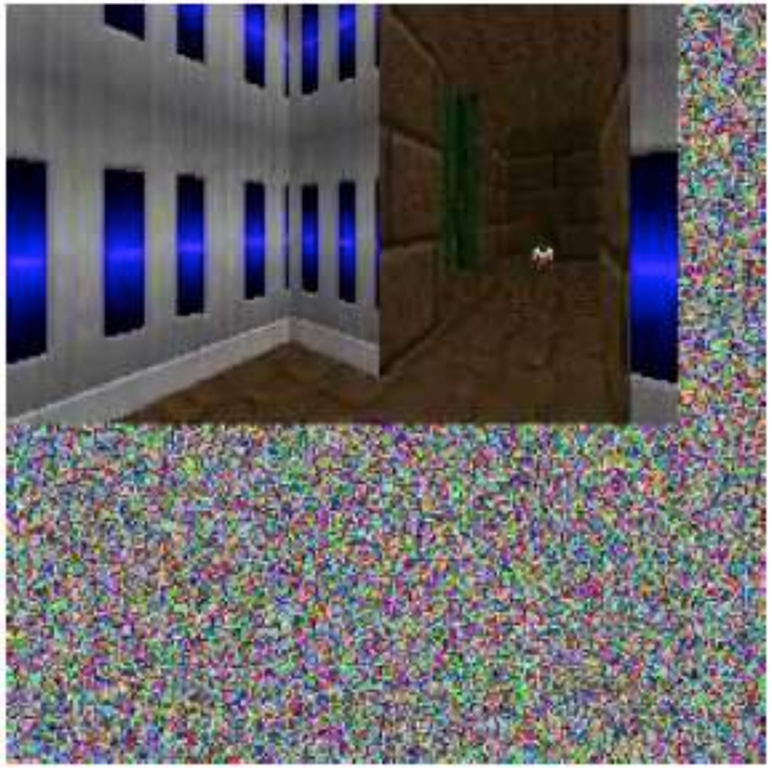
Inverse dynamics modeling

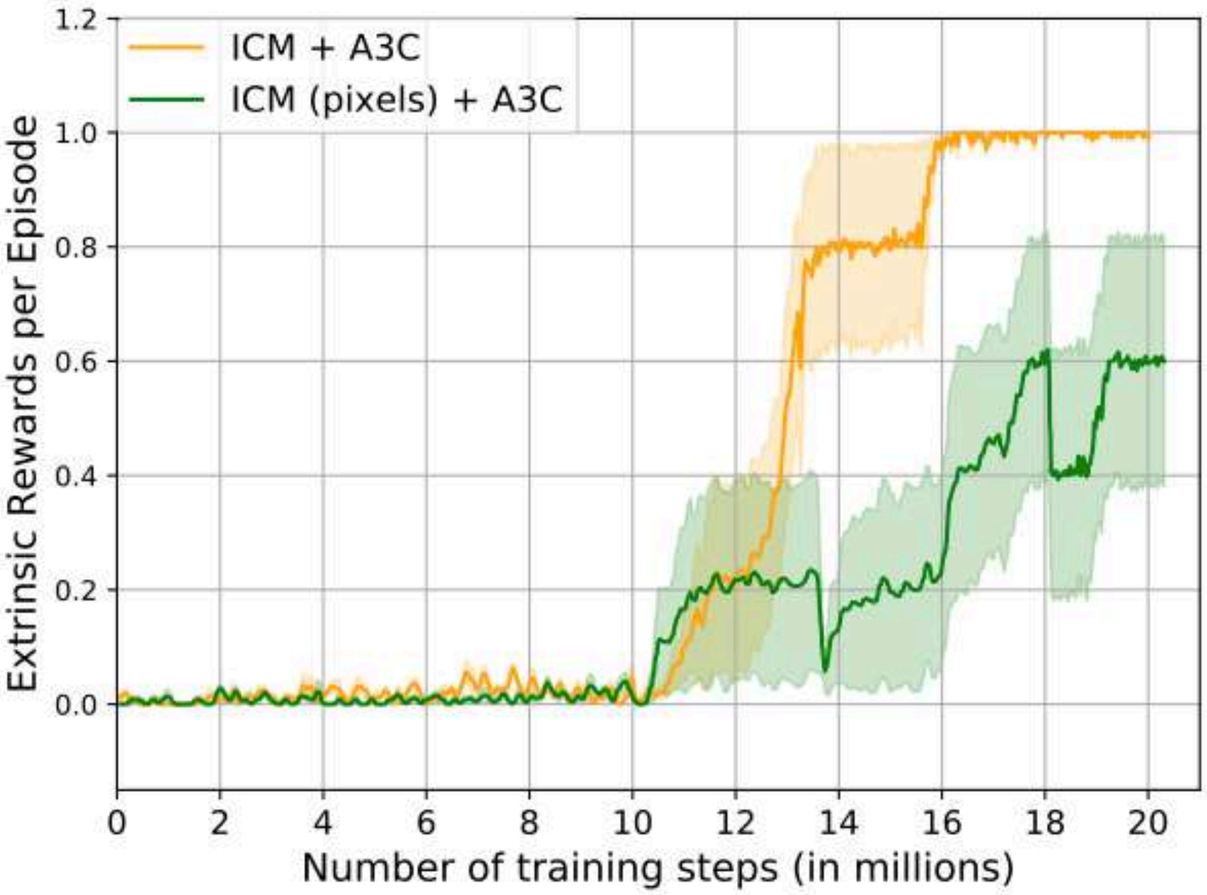


Intuition

Inverse model I should be robust to uncontrollable components

Inverse dynamics modeling





Curiosity-driven Exploration by Self-Supervised Prediction, Pathak et al., ICML 2017

Decoupling RL and Representation Learning

Pre-trained vision models for control

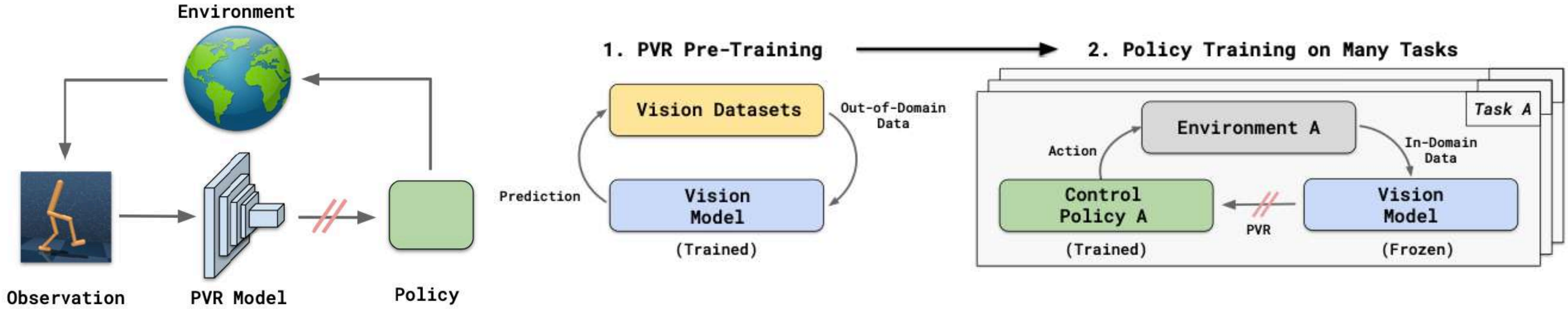
Phase 1: The perception module is detached from the policy

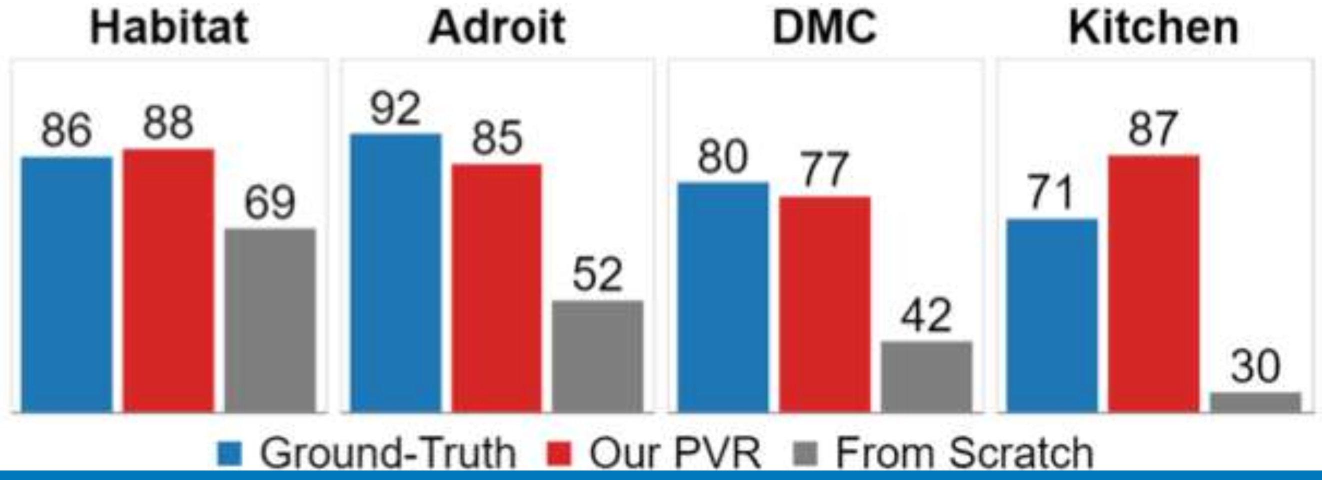
Trained once on out-of-domain data (eg: ImageNet) and frozen

Phase 2: policy training

Control policy are trained on the deployment env reusing the frozen perception module

Decoupling RL and Representation Learning

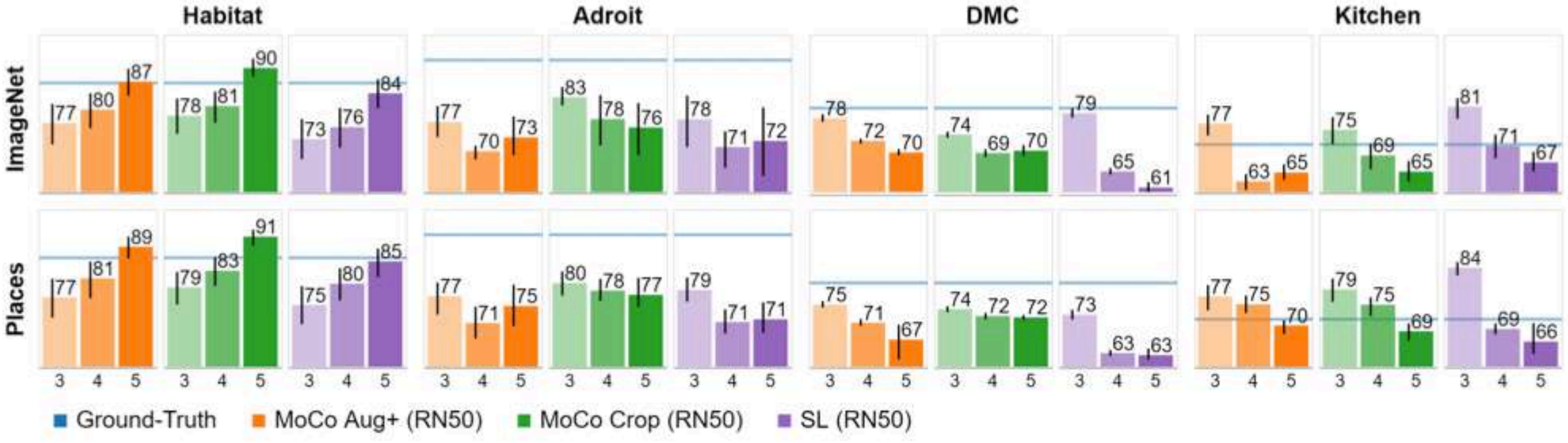




The (Un)surprising Effectiveness of Pre-Trained Models for Control, Paris et al., ICML 2022

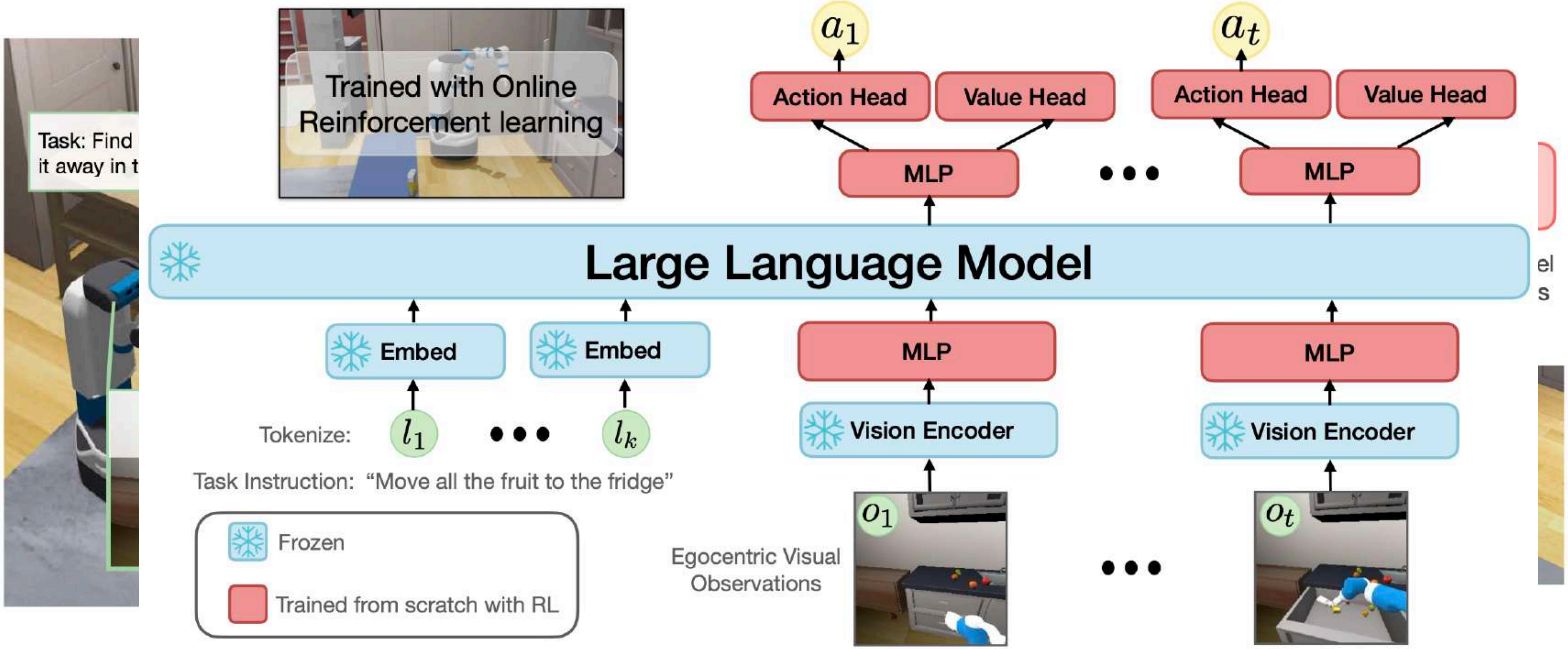
Different layers encode different invariants

- Later layer features are better for high-level semantic tasks (Habitat ImageNav)
- Early layer features are better for fine grained control tasks (manipulation in MuJoCo)



LLMs as policy: Text + Image —> Policy

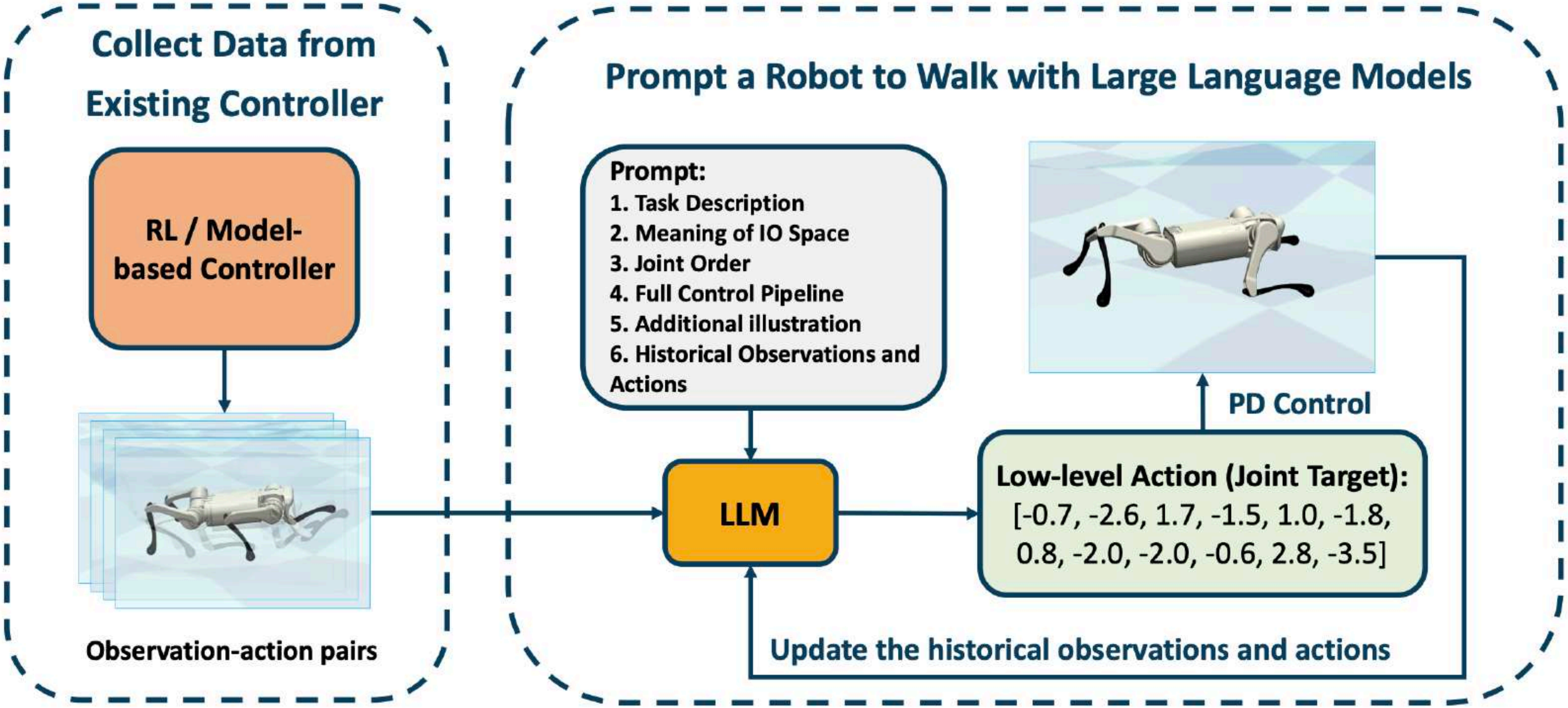
Not MDP



Large Language Models as Generalizable Policies for Embodied Tasks, Toshev et al., arXiv, Oct 2023.

Multi-modal LLMs: Text + Image —> Policy

Not MDP



Wang et al. "Prompt a Robot to Walk with Large Language Models" arXiv Nov 2023.
Yang et al. "Octopus: Embodied Vision-Language Programmer From Environmental Feedback" arXiv Oct 2023.

Conclusions

- Representation learning in RL is a vast topic
  - We cover only a few aspects
- Pre-trained representations are popular nowadays
  - Still lot of open questions: What to pre-train and how
  - Using language as a common input/representation

Common Sense

Implementations

- DrQ-V2: Mastering Visual Continuous Control: Improved Data-Augmented Reinforcement Learning
  - https://github.com/facebookresearch/drqv2
- CURL: Contrastive Unsupervised Reinforcement Learning
  - https://github.com/MishaLaskin/curl
- DEAR: Disentangled Environment and Agent Representations
  - https://github.com/Ameyapores/DEAR

Thank you!

This work was supported by the ATLAS project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813782.

groups @AtlasItn www.atlas-itn.eu