CSC415: Introduction to Reinforcement Learning

Lecture 8: Representation Learning for RL

Dr. Amey Pore

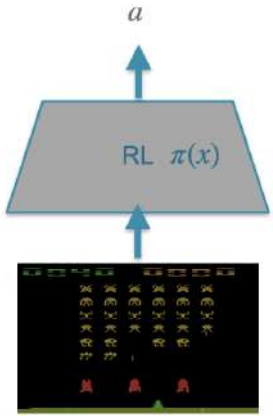
Winter 2026

March 4, 2026

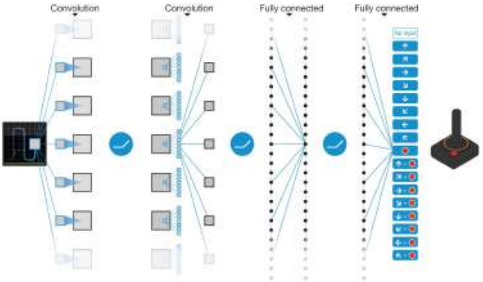
Lecture Outline

1. **End-to-End Reinforcement Learning**
2. What are Good Representations?
3. Implicit Regularisation: Data Augmentation
4. Course Logistics
5. Explicit Regularisation of Representations
6. Conclusions

End-to-End Reinforcement Learning



- Learn mapping from observations to actions



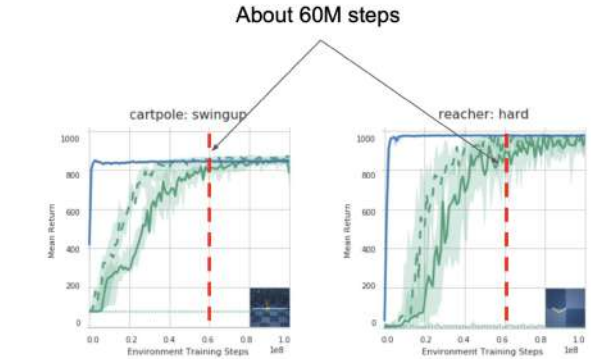
[Human-level control through deep reinforcement learning, Mnih et al, Nature 2015]

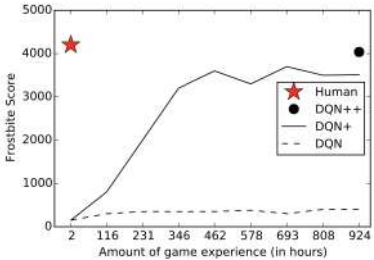
Catch

1. Inefficiency

- Millions of transitions (sample inefficient)

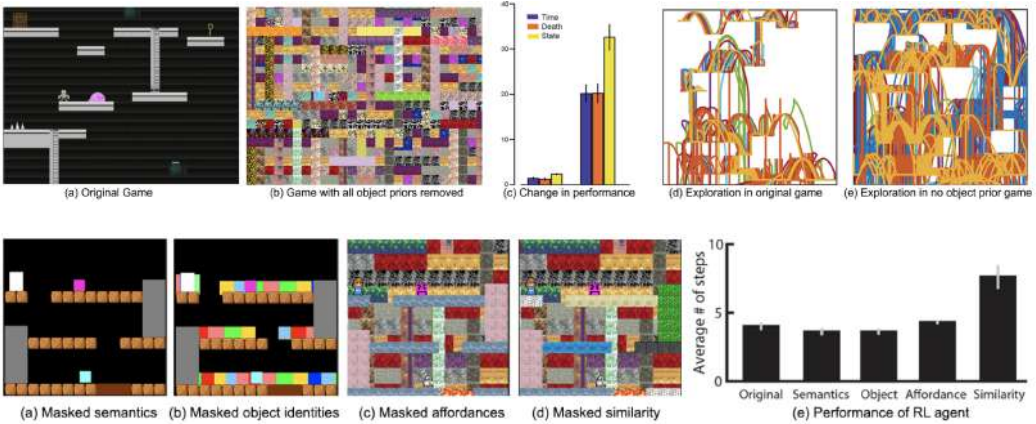
→ **Good representations can accelerate** learning from images





[DeepMind Control Suite, Tassa et al., 2018
Building Machines That Learn and Think Like People, Lake et al., Behavioral and Brain Sciences 2016]

Humans come with prior knowledge



[Investigating Human Priors for Playing Video Games, Dubey et al. ICML 2018]

Catch

1. Inefficiency

- millions of transitions (sample inefficient)

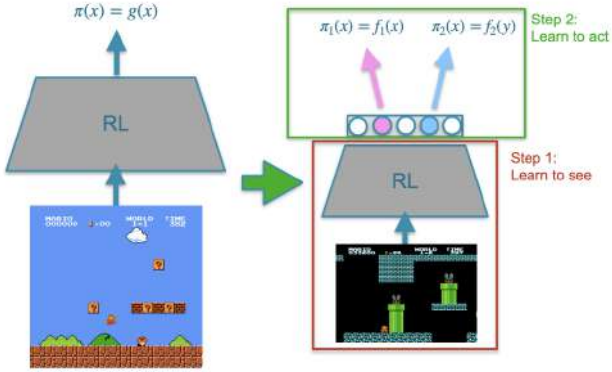
→ Good representations can accelerate learning from images

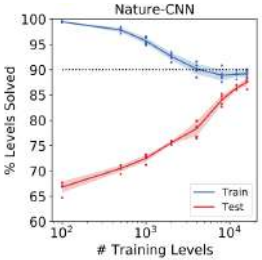
2. Generalisation

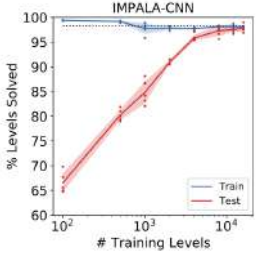
- Works really well in single task setting

→ 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)

→ 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

3. Requires lots of supervision

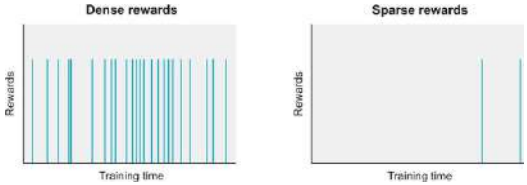
- Dense reward function
- Effective exploration is challenging in many RL tasks

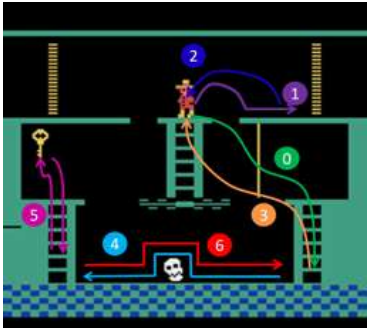
→ Instead of only learning from reward signals, we can also learn from unsupervised collected data.

Good representations can accelerate exploration

Sparse Reward and Exploration

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





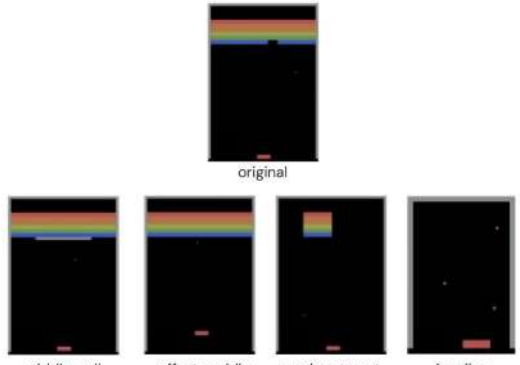
Robustness

[Explaining and Harnessing Adversarial Examples, Goodfellow et al., ICLR 2015]

Transferability

Desired properties:

- General
- Robust
- Useful
- Reusable
- Compositional
- Interpretable



	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]

Lecture Outline

1. End-to-End Reinforcement Learning
2. **What are Good Representations?**
3. Implicit Regularisation: Data Augmentation
4. Course Logistics
5. Explicit Regularisation of Representations
6. Conclusions

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

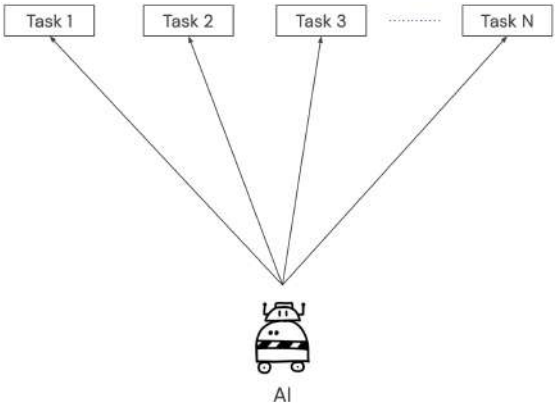
- Representational form orthogonal to the information content
- Useful abstraction to make different computations more efficient

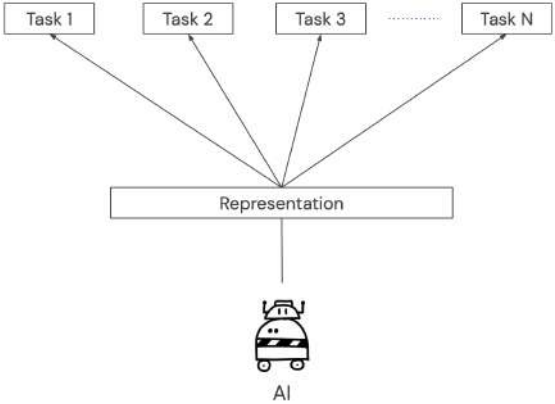
Representation Learning

- *"... 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

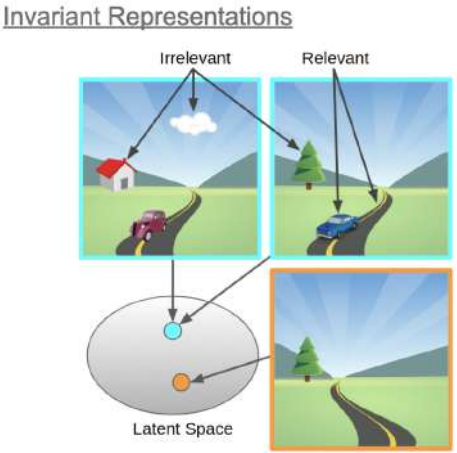
[Representation Learning: A Review and New Perspectives, Bengio et al., 2013]

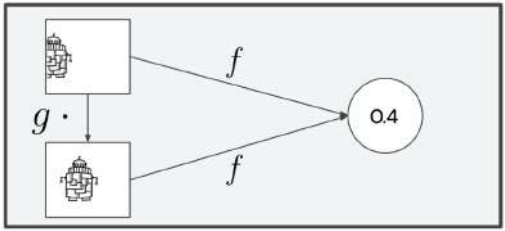
Representation learning for RL





What are good representations?





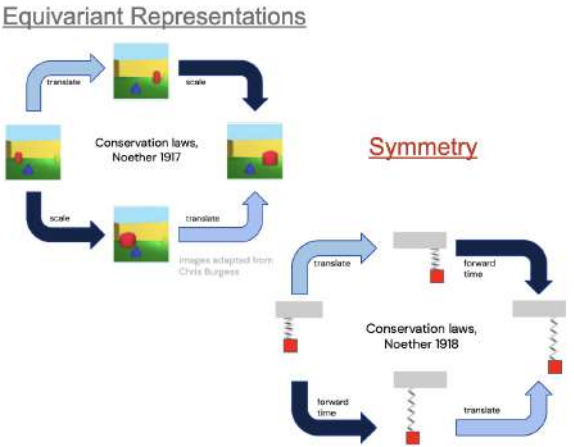
Invariance

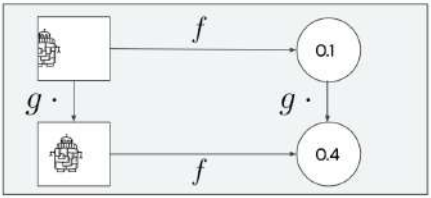
– representation remains unchanged when a certain type of transformation is applied to the input

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

[Learning Invariant Representations for Reinforcement Learning without Reconstruction, Zhang et al., ICLR 2021]

What are good representations?



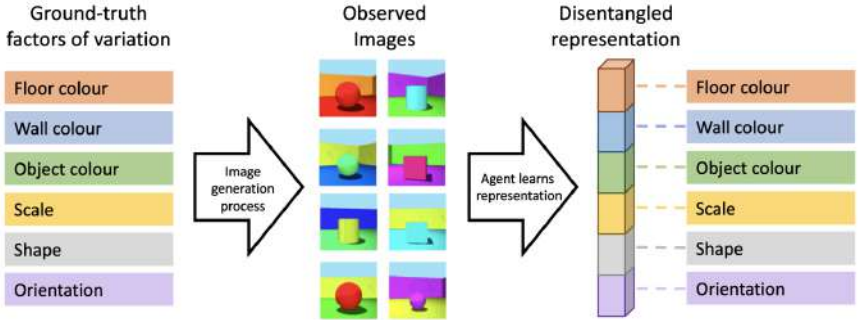


Equivariance

– representation reflects the transformation applied to the input

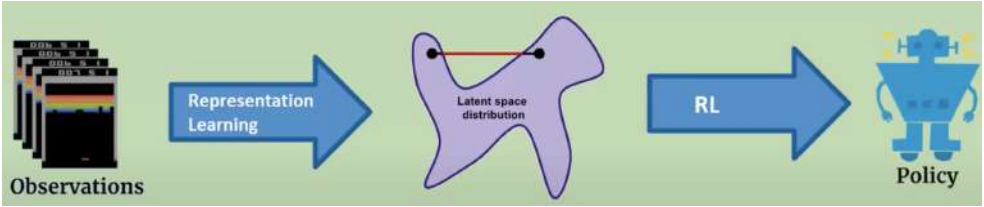
$$f(g \cdot x) = g \cdot f(x)$$

Disentangled Representation Learning



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

Representation Learning for RL



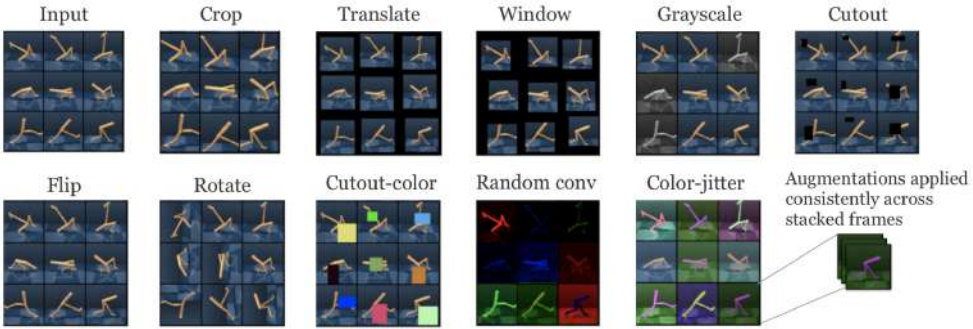
We assume the learner has access to a representation space $\mathcal{F}$

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 $\mathcal{D}_{k+1}$

Lecture Outline

Implicit Regularisation of the Representations: Data Augmentation



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

Data Augmentation for RL

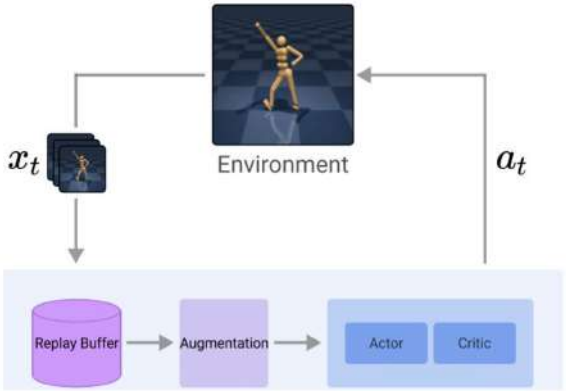
Surprisingly, data augmentation has been adopted only recently

Issues

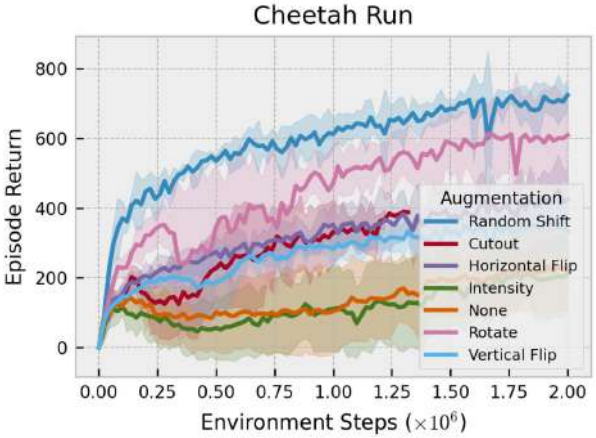
- Unclear what are RL-driven data augmentation, in particular in state-based control

Workaround

- Use standard techniques for images



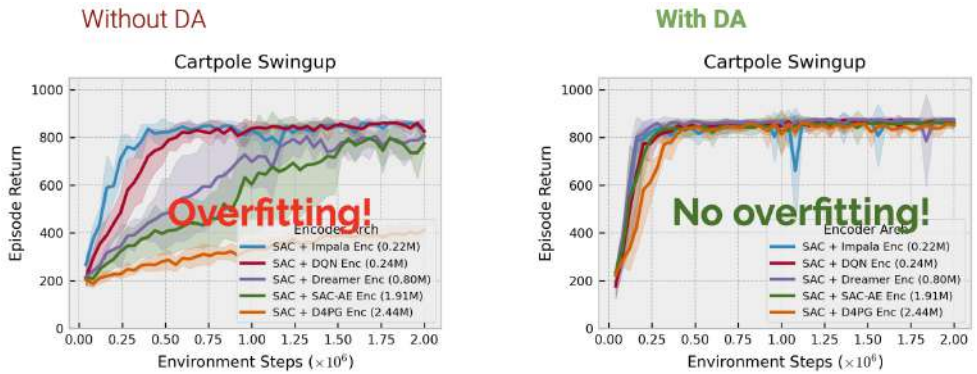
Data Augmentation for RL: Not All Augmentations Work



- Not all standard CV data augmentations can be used in RL.
- Some recent works in automatic way of selecting augmentation.

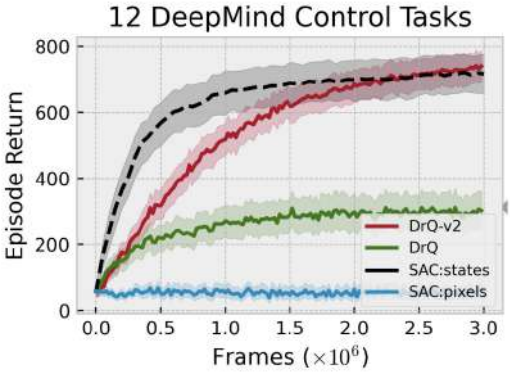
[Image Augmentation Is All You Need: Regularizing Deep RL from Pixels, Yarats et al., ICLR 2021
Automatic Data Augmentation for Generalisation in RL, Raileanu et al., NeurIPS 2021]

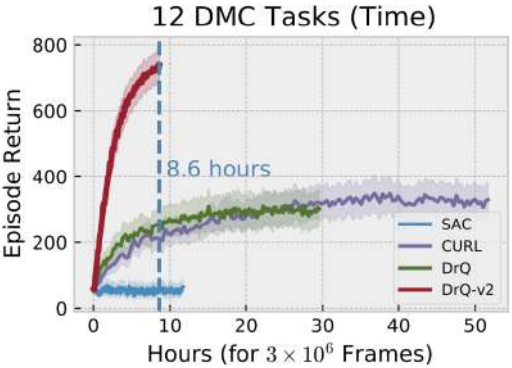
Data Augmentation Prevents Overfitting



[Image Augmentation Is All You Need: Regularizing Deep RL from Pixels, Yarats et al., ICLR 2021]

Data Augmentation Works





[Mastering Visual Continuous Control: Improved Data-Augmented RL, Yarats et al., ICLR 2021]

Lecture Outline

Course Logistics

Please vote for the quiz.

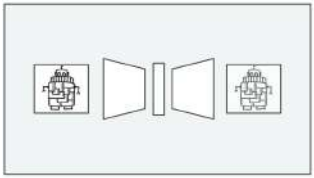
Reasons:

1. You get more time for the project submission. Can be extended by 2 weeks from March 23 to April 7th.
2. Quiz will not be mostly conceptual.
3. Will help you revise from lecture 4-8.
4. 2 more labs remaining.
5. A1 grading should be out next week.

Break

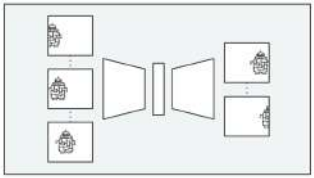
Lecture Outline

Explicit Regularisation of Representations



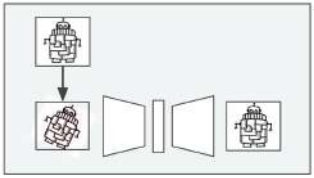
Generative modeling

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.

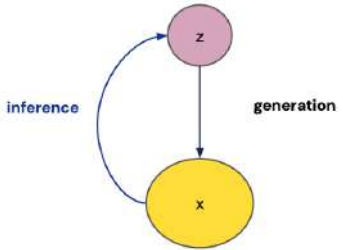


Self-supervision

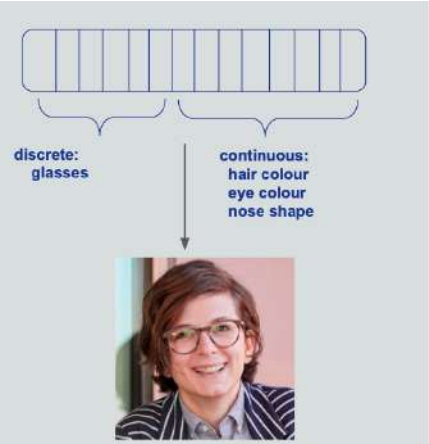
Exploit knowledge of data to design learning tasks which lead to useful representations.

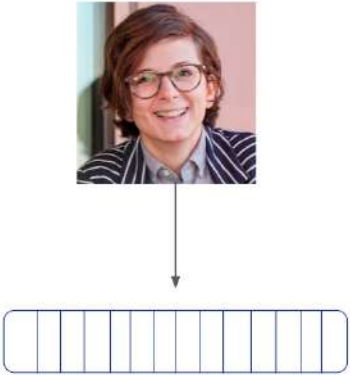
Latent variable models

- Model the data generating process as a mapping from a low dimensional unknown (latent) space to the data distribution
- Inference: Find $p(\mathbf{z}|\mathbf{x})$
- Intuition: Find the underlying factors which generated the data (with uncertainty estimates)
- Finding $p(\mathbf{z}|\mathbf{x})$ is often intractable, and we thus have to resort to approximations



Generation vs Inference





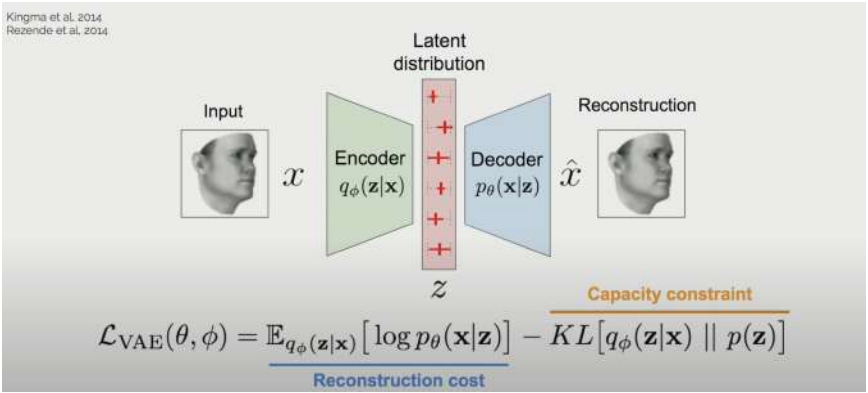
Variational Autoencoders

- **Maximum likelihood:**
  $\mathbb{E}_{p^*(\mathbf{x})}[\log p_\theta(\mathbf{x})]$
- **Latent variable model:**
  $\log p_\theta(\mathbf{x}) = \log \int p_\theta(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}$
- **Lower bound on maximum likelihood objective (ELBO):**
  $\log p_\theta(\mathbf{x}) \geq \underbrace{\mathbb{E}_{q_\eta(\mathbf{z}|\mathbf{x})} \log p_\theta(\mathbf{x}|\mathbf{z})}_{\text{reconstruct}} - \underbrace{\mathrm{KL}(q_\eta(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))}_{\text{stay close to prior}}$
- **Approximate posterior:**
  $q_\eta(\mathbf{z}|\mathbf{x})$

Role of the prior

- The KL term regularises the approximate posterior to the prior
- Use the prior to specify properties we would like the posterior to have, such as disentanglement

Variational Autoencoders (VAE)

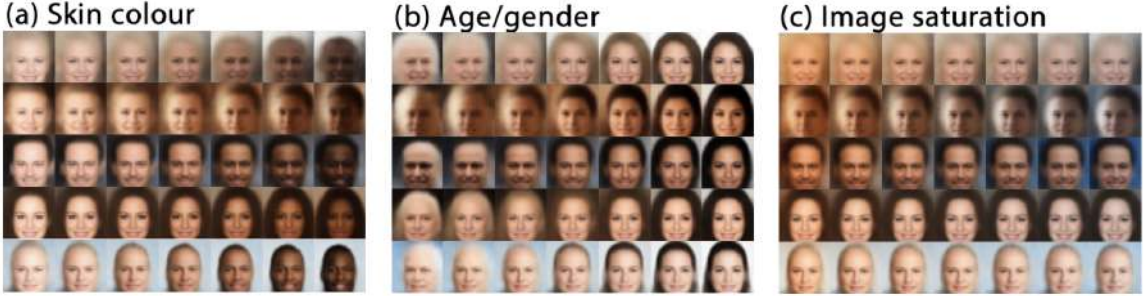


[Auto-Encoding Variational Bayes, Kingma et al., ICLR 2014]

Beta-VAE

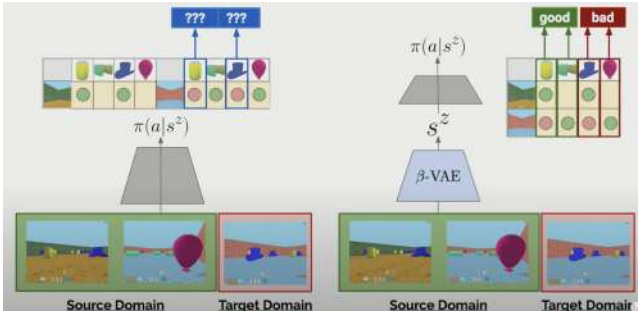
$$\mathbb{E}_{q_\eta(\mathbf{z}|\mathbf{x})} \log p_\theta(\mathbf{x}|\mathbf{z}) - \beta \mathrm{KL}(q_\eta(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))$$

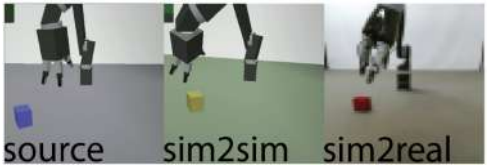
Change the weight of the KL term to encourage disentangled representations



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

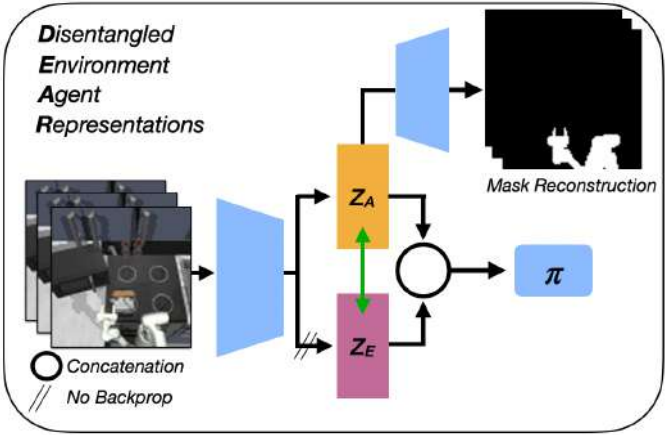
Beta-VAE in RL: DARLA

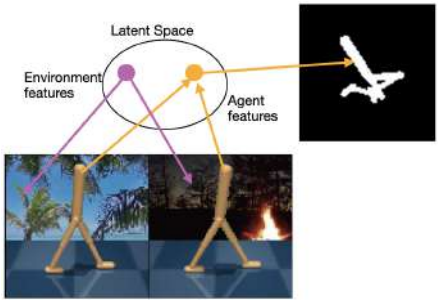




[DARLA: Improving Zero-Shot Transfer in Reinforcement Learning, Higgins et al., ICML 2017]

DEAR: Disentangled Environment and Agent Representations

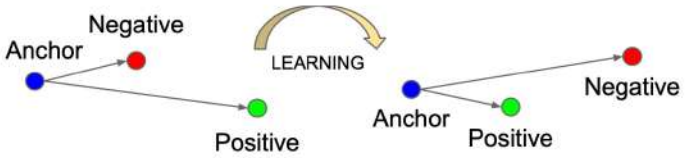




[DEAR: Disentangled Environment and Agent Representations for Reinforcement Learning without Reconstruction, Pore et al., 2024]

Contrastive Learning

Contrastive Learning

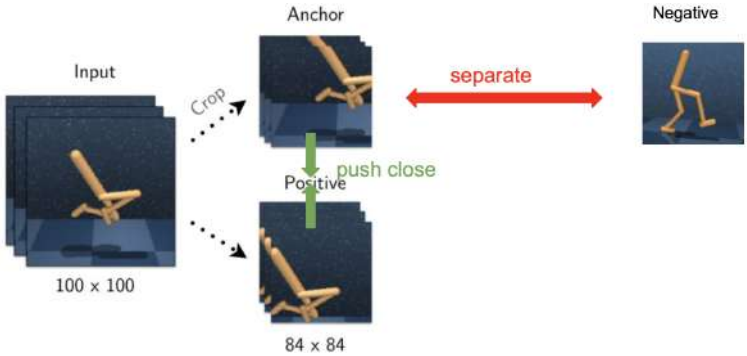


1. For an anchor x, we are given a positive sample x^+ and a negative sample x^-
2. The learning objective is to:
   - Minimize the distance between the anchor and positive
   - Maximize the distance between the anchor and negative

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, Schroff et al., 2015]

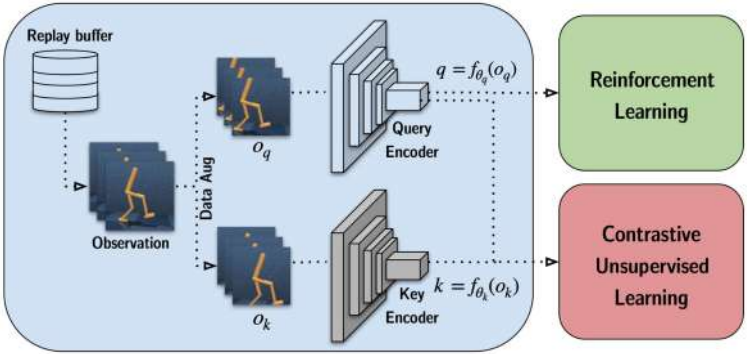
Contrastive Learning in RL



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]

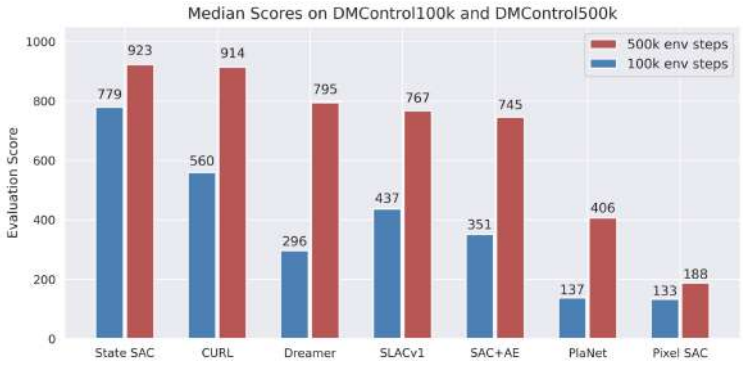
Contrastive Learning in RL



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

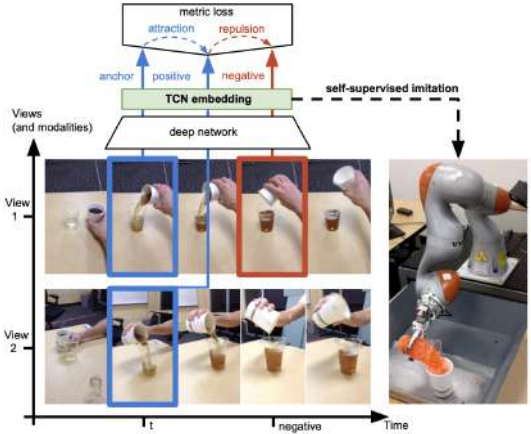
[CURL: Contrastive Unsupervised Representations for Reinforcement Learning, Srinivas et al., 2020]

Contrastive Learning in RL: Results



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

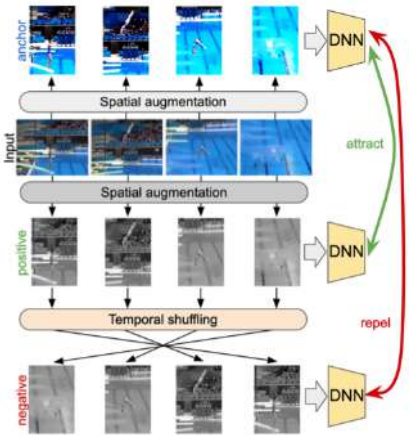
Temporal Representation Learning: Time Contrastive Networks



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: link

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

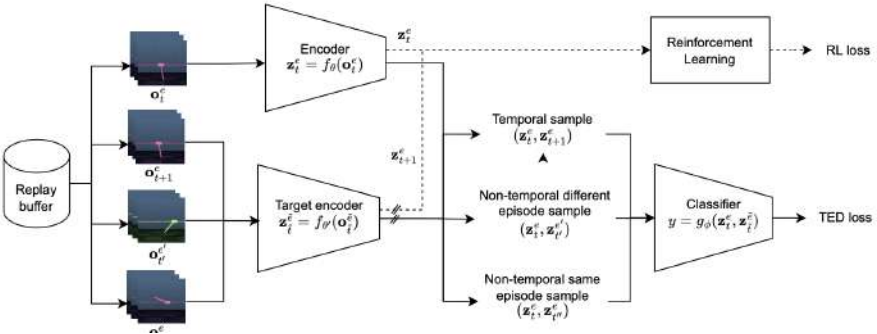
Temporal Representation Learning: Shuffling



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

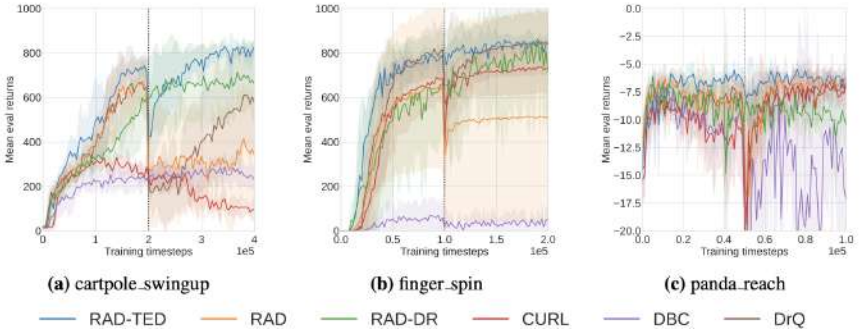
[SCVRL: Shuffled Contrastive Video Representation Learning, Dorkenwald et al., CVPR 2023]

Temporal Representation Learning: Disentanglement



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

Temporal Learning: Results



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

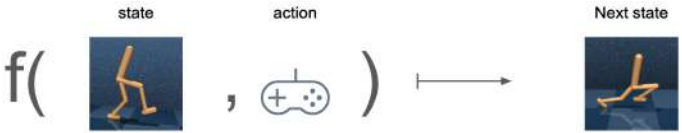
Self-Supervision

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

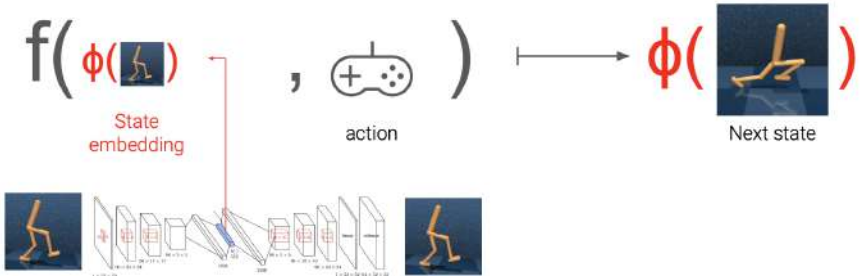
Direct pixel to pixel prediction may be too complicated, better to use a latent representation



[Incentivizing Exploration In Reinforcement Learning With Deep Predictive Models, Stadie et al., 2015]

Forward Dynamics Modelling: Latent Modelling

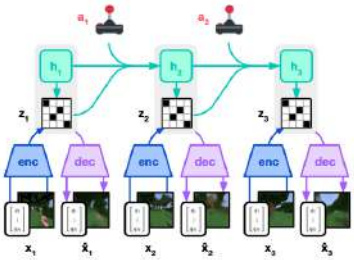
A common approach is to extract the latent representation via an auto encoder



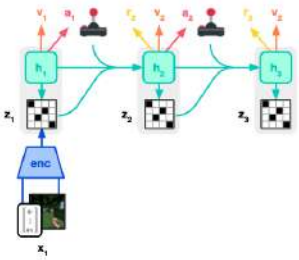
[A Study of Count Based Exploration for Deep Reinforcement Learning, Tang et al., NeurIPS 2017]

Example: DREAMER

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



(a) World Model Learning

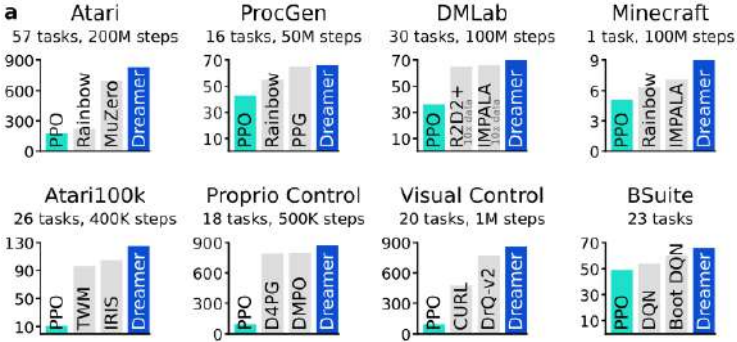


(b) Actor Critic Learning

- 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

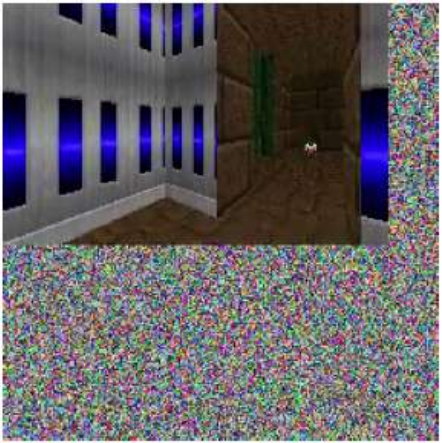
Example: DREAMER Results

Generate imagined trajectories using dynamics model



Is Everything Relevant?

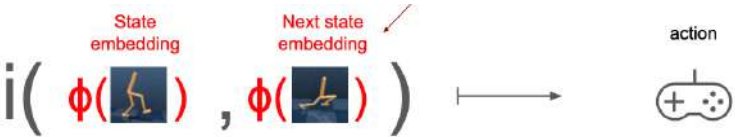
- Forward models have to concentrate on each individual pixel to be able to reconstruct the image
- For controllability, we may need to predict only changes that depend on agent's actions, ignore the rest



[Burda et al., *Large-Scale Study of Curiosity-Driven Learning*, ICLR 2019.]

Inverse Dynamics Modeling

Intuition: Inverse model I should be robust to uncontrollable components

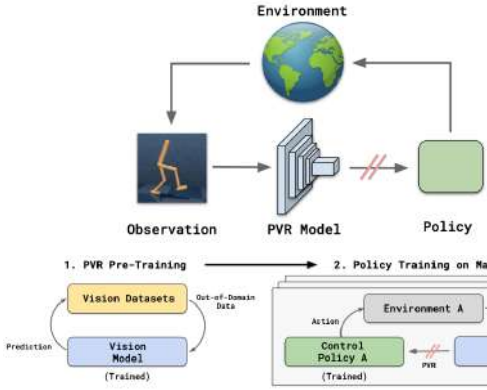


Decoupling RL and Representation Learning

Pre-trained vision models for control

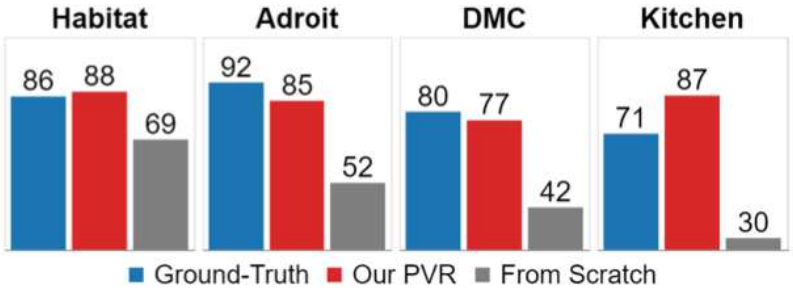
Phase 1: The perception module is detached from the policy. Trained once on out-of-domain data (e.g. ImageNet) and frozen

Phase 2: Policy training. Control policies are trained on the deployment env reusing the frozen perception module



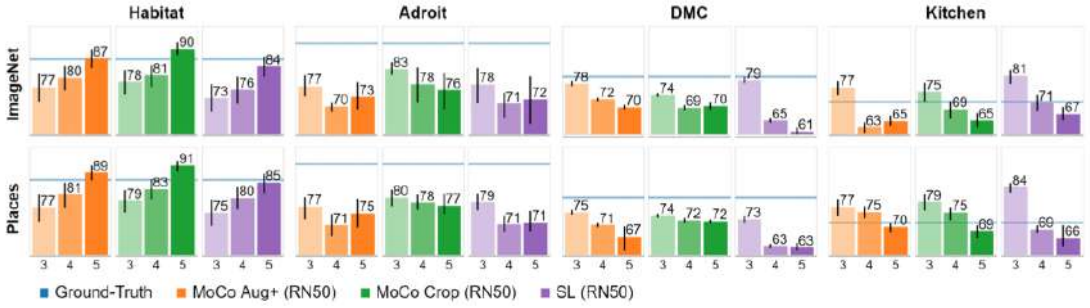
[The (Un)surprising Effectiveness of Pre-Trained Models for Control, Parisi et al., ICML 2022]

Decoupling RL and Representation Learning: Results



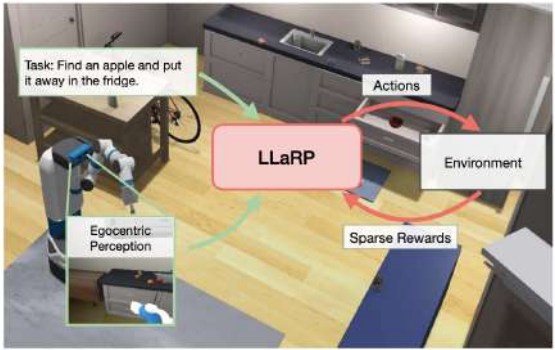
[The (Un)surprising Effectiveness of Pre-Trained Models for Control, Parisi et al., ICML 2022]

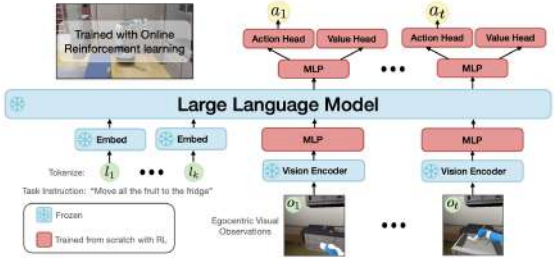
Different Layers Encode Different Invariants



- 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

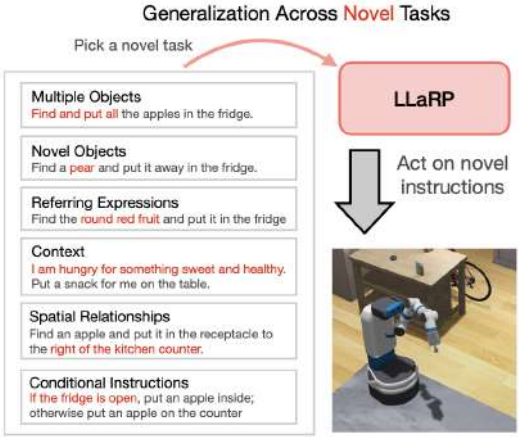




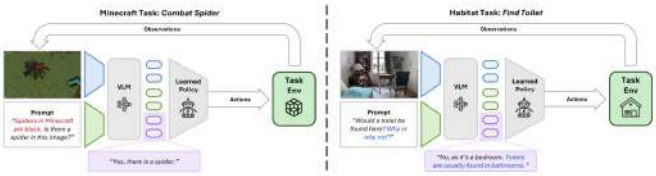
Not MDP

[Large Language Models as Generalizable Policies for Embodied Tasks, Toshev et al., ICLR 2023]

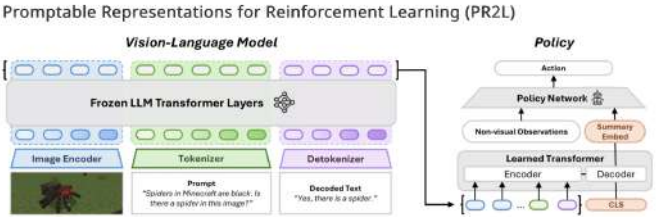
LLMs as Policy: Text + Image → Policy



Promptable Representations for Reinforcement Learning (PR2L)



Example instantiations of PR2L for Minecraft and Habitat, showing how the approach can use auxiliary text and chain-of-thought prompting.



[Promptable Representations for Reinforcement Learning (PR2L), arXiv. pr2l.github.io]

Lecture Outline

1. End-to-End Reinforcement Learning
2. What are Good Representations?
3. Implicit Regularisation: Data Augmentation
4. Course Logistics
5. Explicit Regularisation of Representations
6. **Conclusions**

Conclusions

- Representation learning in RL is a vast topic
  - We cover only a few aspects
- Pre-trained representations are popular nowadays
- Using language as a common input/representation

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!