Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Florian Vahl

08.12.2022

Reinforcement learning with Dreamer / Dreaming V1 / V2

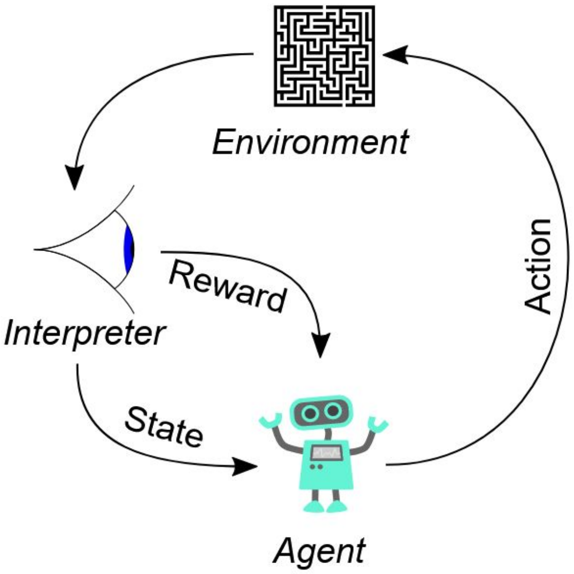
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Outline

- Fundamentals
  - Markov Decision Process
  - Reinforcement Learning
- Approaches
  - Dreamer V1
  - Dreamer V2
  - Dreaming V1
  - Dreaming V2
- Results
- Discussion

Fundamentals

Markov Decision Process



Wikimedia Commons, CC-0

Markov Decision Process

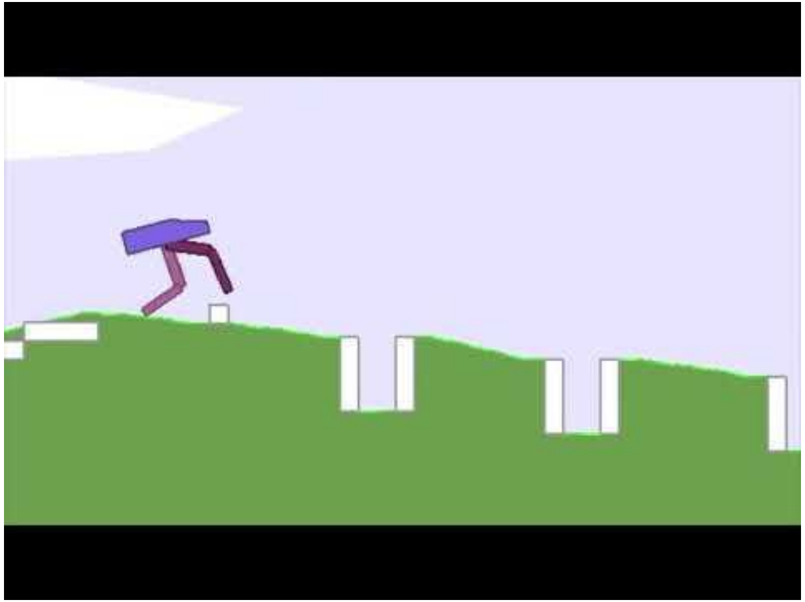
An MDP is a tuple $\langle S, A, P, R \rangle$ where:

- S is a set of states.
- A is a set of actions.
- $P : S \times A \times S \to [0, 1]$ is the state transition probability function.
- $R : S \times A \to R$ is the reward function.

Reinforcement Learning

- Type of machine learning that involves agents learning based on a reward
- No ground truth labels
- Optimizes policy to maximize the reward signal
- Applied in many domains including robotic control and game playing
- Typically formulated using the Markov Decision Process

Reinforcement Learning - Environments

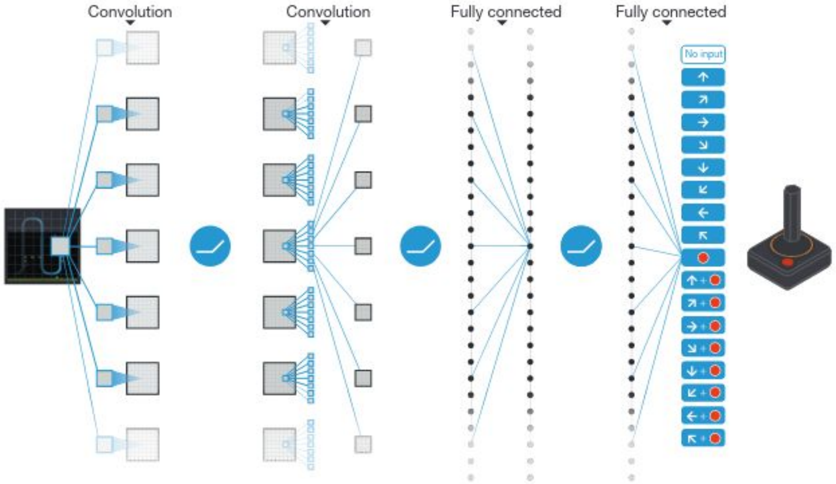


https://www.youtube.com/watch?v=SyonPJc8hMw

Reinforcement Learning - Environments

Mastering Atari with Discrete World Models (2022)

Reinforcement Learning - Deep Q-Learning (Example)



Human-level control through deep reinforcement learning

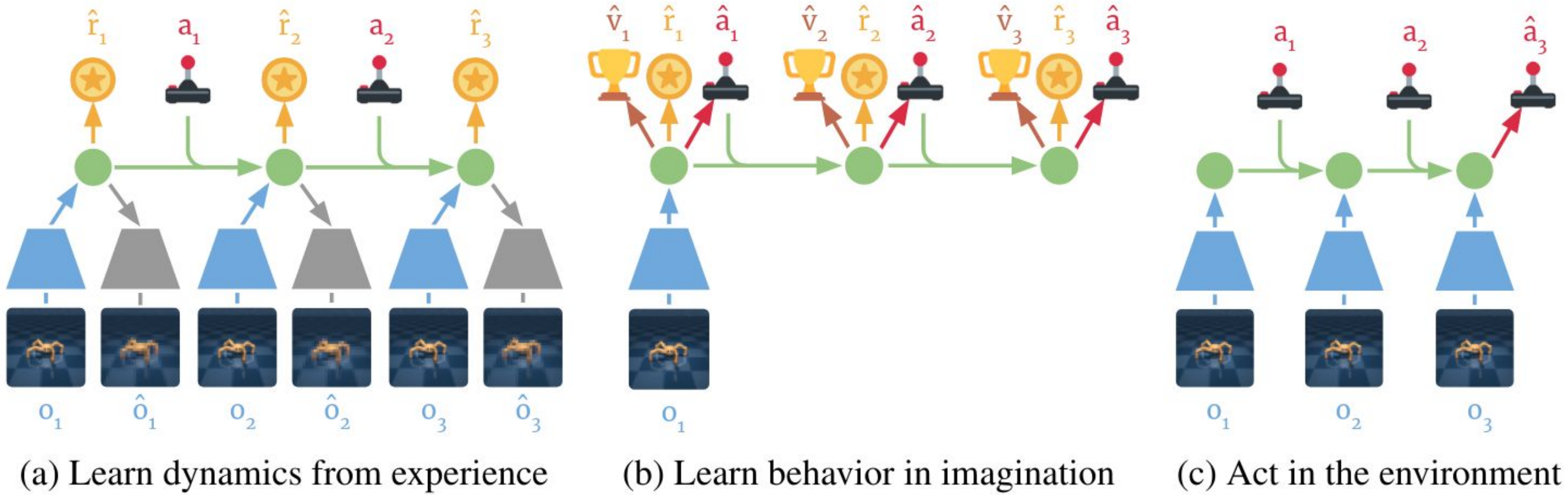
Approaches

Dreamer V1

- Paper: “Dream to Control: Learning Behaviors by Latent Imagination” by Hafner et al. from Google Brain at ICLR 2020
- Model based reinforcement learning algorithm
- Learns world model
- Solves long-horizon tasks purely from image base observations
- Imagines novel trajectories in the world models latent space
- Policy is trained using the imagined trajectories

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Dreamer V1 – Architecture



Dream to Control: Learning Behaviors by Latent Imagination (2020)

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

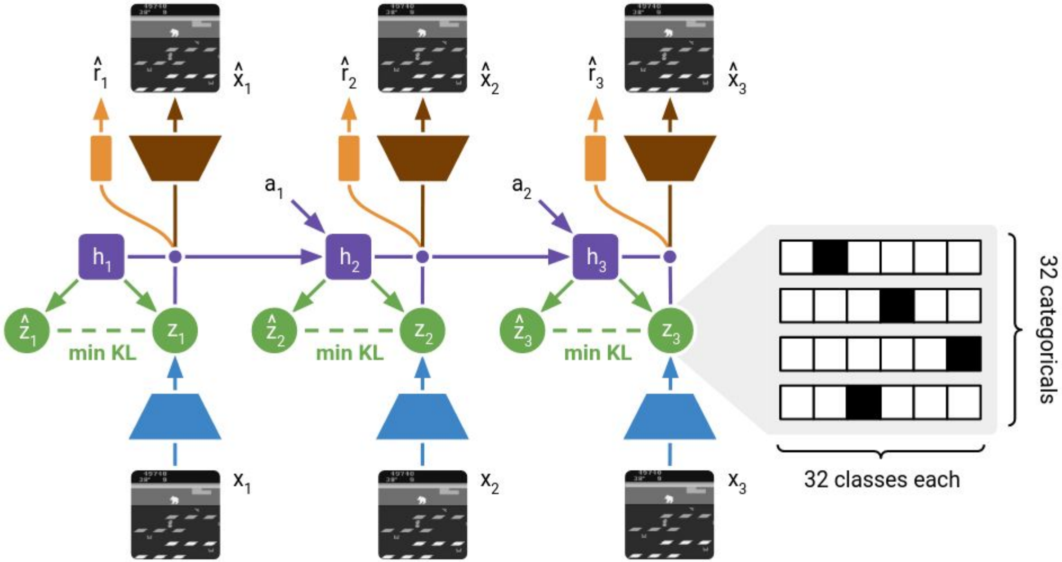
Dreamer V1 – World Model

- Representation model $p_\theta(s_t \mid s_{t-1}, a_{t-1}, o_t)$
- Observation model $q_\theta(o_t \mid s_t)$
- Reward model $q_\theta(r_t \mid s_t)$
- Transition model $q_\theta(s_t \mid s_{t-1}, a_{t-1}).$
  - Predicts next latent state
  - Outputs tanh-transformed Gaussian distribution
- Frozen while learning behaviors

Dreamer V2

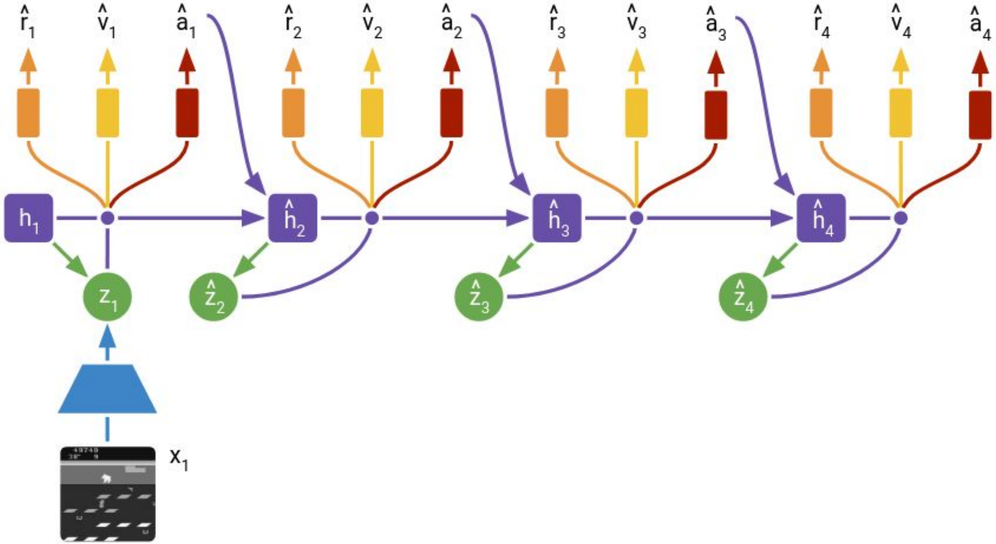
- Paper: “Mastering Atari with Discrete World Models” by Hafner et al. from Google Brain at ICLR 2021
- Similar to Dreamer V1
- Transition model uses categorical variables instead of Gaussians
- Performed well in image based Atari tasks

Dreamer V2 – Architecture



Mastering Atari with Discrete World Models (2022)

Dreamer V2 – Latent Imagination

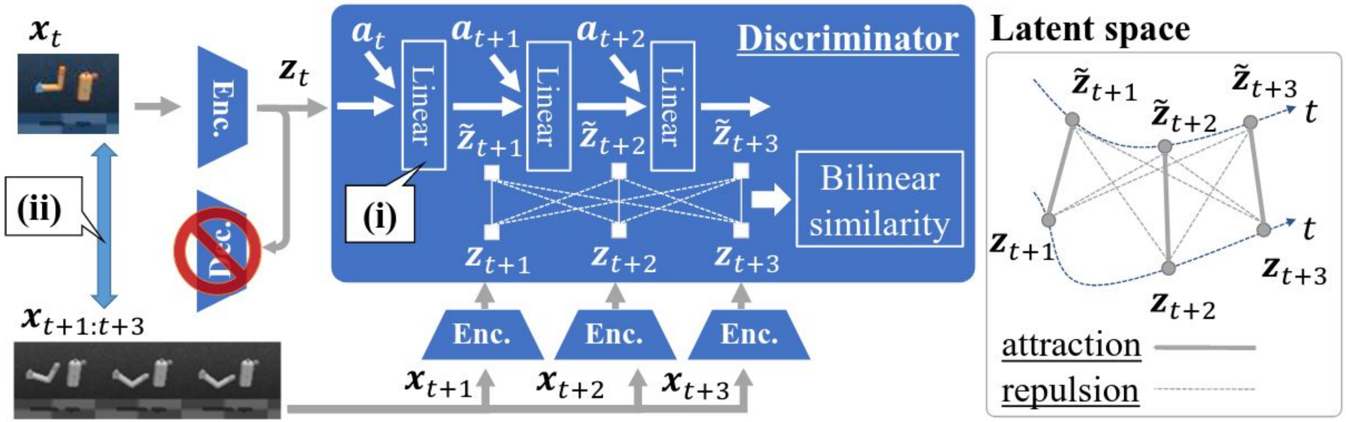


Mastering Atari with Discrete World Models (2022)

Dreaming V1

- Paper: “Dreaming: Model-based Reinforcement Learning by Latent Imagination without Reconstruction” by Okada et al. from Panasonic Corporation in ICRA 2021
- Similar to Dreamer V1
- World model is trained via contrastive learning
- No generative decoder / observation model
- Tries to avoid object vanishing

Dreaming V1 – Architecture

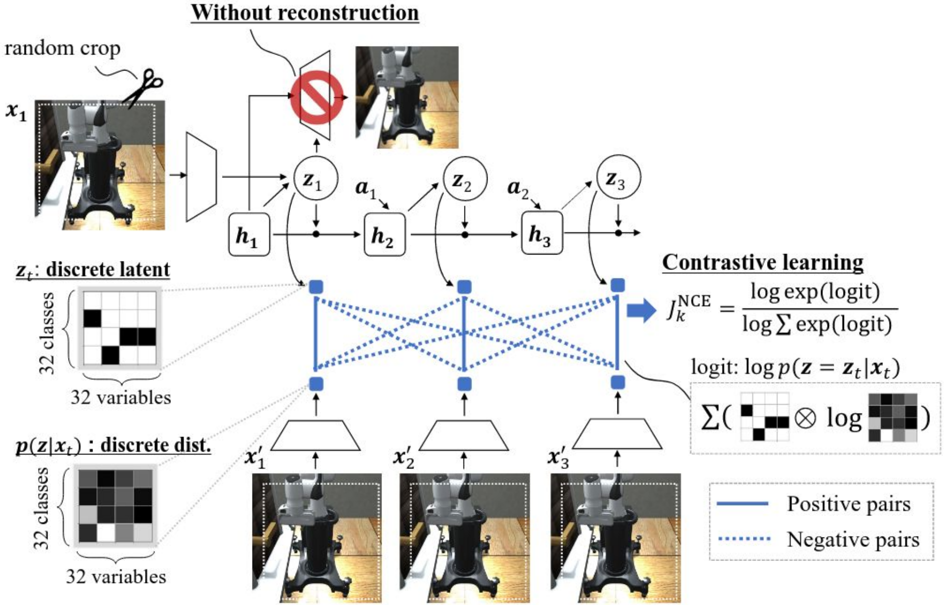


Dreaming: Model-based Reinforcement Learning by Latent Imagination without Reconstruction (2022)

Dreaming V2

- Paper: “DreamingV2: Reinforcement Learning with Discrete World Models without Reconstruction” by Okada et al. from Panasonic Corporation on arXiv (2022)
- Applies the concept concepts introduced in Dreamer V2 to Dreaming V1
- Also decoder free

Dreaming V2 – Architecture



DreamingV2: Reinforcement Learning with Discrete World Models without Reconstruction (2022)

Results

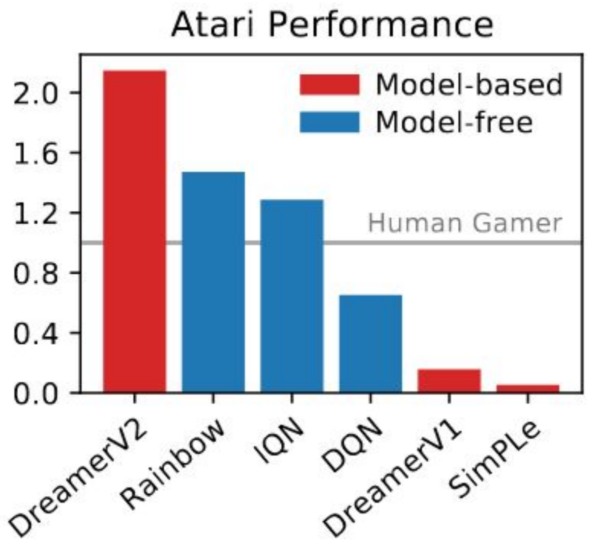
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Results

	DreamingV2 (ours)	DreamerV2 [6]	Dreaming [7]	Dreamer [5]
Discrete latent	✓	✓		
Reconstruction free	✓		✓	
(A) 3D Rotot-arm tasks				
UR5-reach	**776**±194	704±222	752±1178	701±223
Reach-duplo	**199**±43	149±62	145±61	5±11
Lift	**327**±150	165±126	174±107	134±46
Door	**383**±143	190±126	319±173	154±32
PegInHole	**436**±26	376±59	353±50	354±47
(B) 2D Manipulation tasks				
Reacher-hard	598±447	175±340	**743**±346	247±392
Finger-turn-hard	484±434	600±417	**858**±210	533±426
Reacher-easy	924±210	923±215	**947**±100	658±429
Finger-turn-easy	434±469	498±469	**842**±286	665±430
(C) Difficult pole-swingup tasks				
Acrobot-swingup	**470**±129	309±131	359±111	382±147
Cartpole-two-poles	**308**±55	248±103	273±53	256±65
(D) 3D locomotion robot tasks				
Quadruped-walk	**492**±127	350±89	379±189	242±120
Quadruped-run	**385**±91	352±68	339±128	269±114
(E) 2D locomotion tasks				
Cheetah-run	768±24	**811**±75	542±132	776±120
Walker-walk	857±115	**951**±28	518±76	906±70

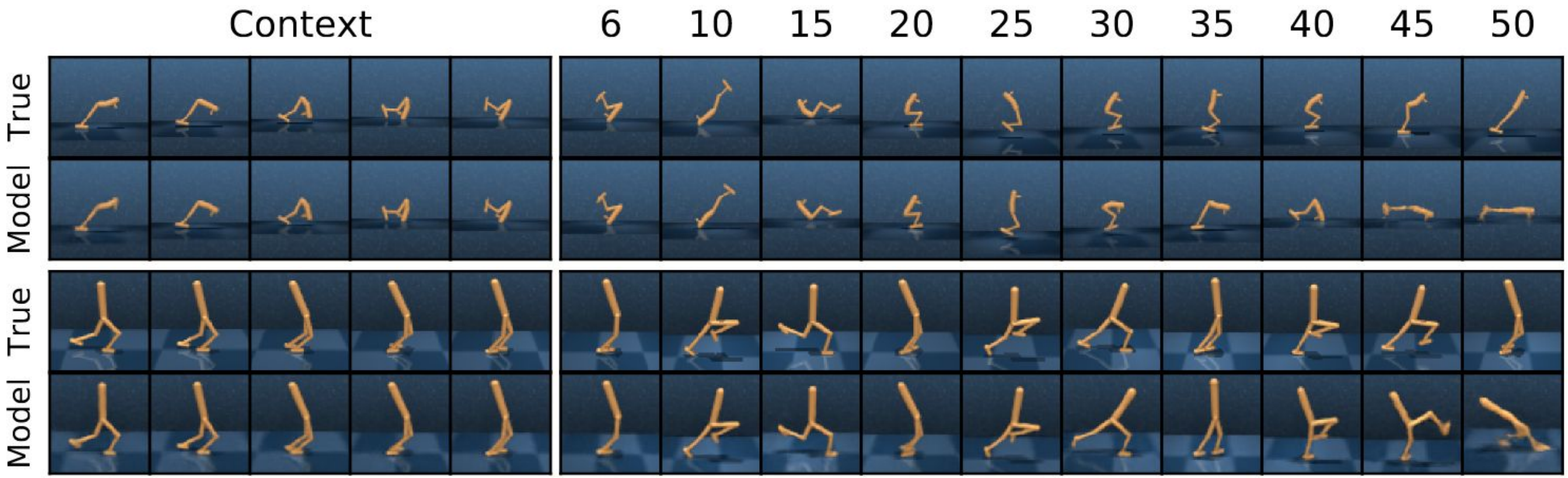
DreamingV2: Reinforcement Learning with Discrete World Models without Reconstruction (2022)

Results



Mastering Atari with Discrete World Models (2022)

Results



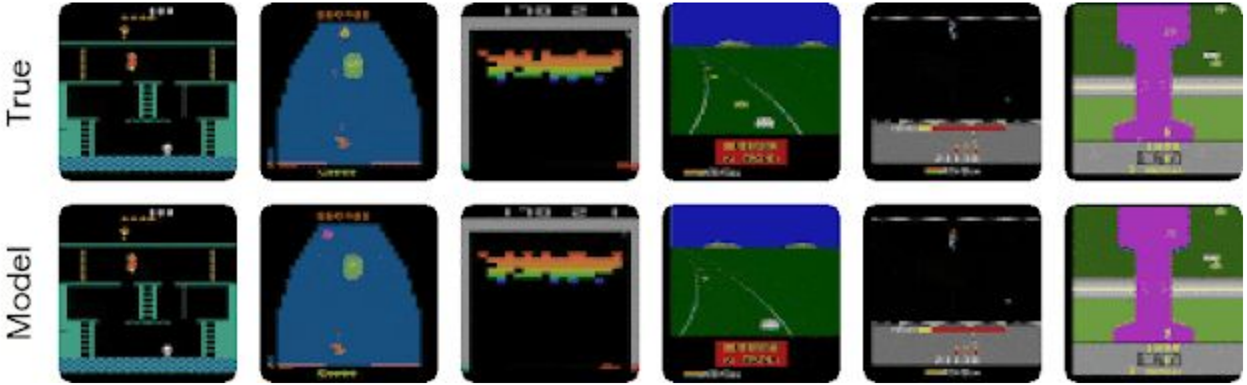
Dream to Control: Learning Behaviors by Latent Imagination (2020)

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Results – Video

Mastering Atari with Discrete World Models (2022)

Results – Video



Mastering Atari with Discrete World Models (2022)

Results – Real World Reinforcement Leaning

DayDreamer: World Models for Physical Robot Learning (2022)

Discussion

- Sampling efficiency can be improved using latent imagination
- Object vanishing is still a problem even in Dreaming
- Atari performance only for “single GPU” setups
- Categorical stochastic representations improve performance
- Good fit for situations where exploration is costly
  (e.g. real-world reinforcement learning)

Questions?