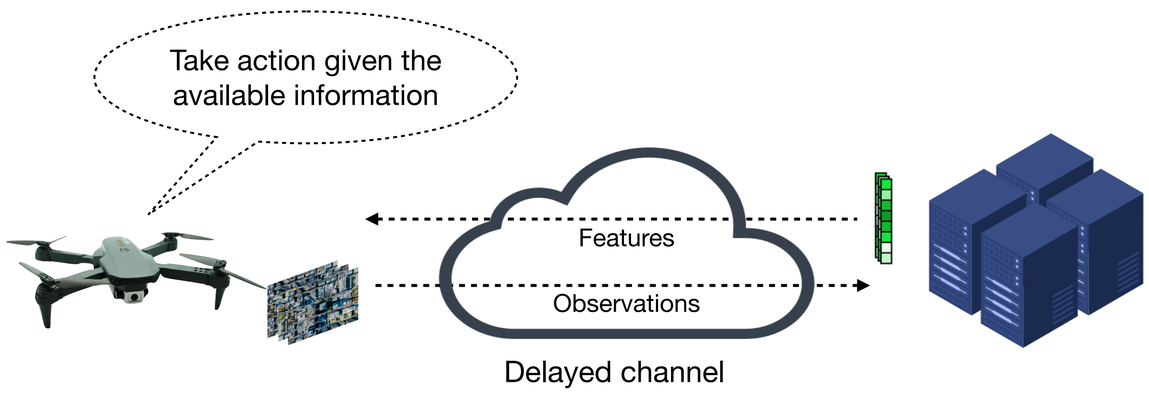
Reinforcement Learning from Delayed Observations via World Models

Armin Karamzade, Kyungmin Kim, Montek Kalsi, Roy Fox

University of California, Irvine

Motivation

Agent receives observation o_{t-d} at time t, and should take a_t



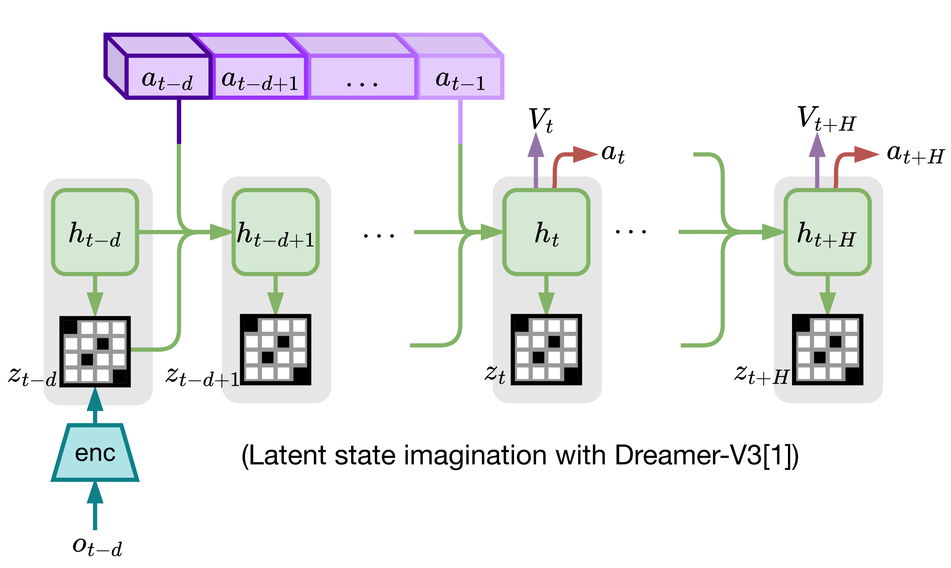
Methods

Observation: World model converts a delayed POMDP to a delayed MDP in the latent space

Two methods for addressing observation delays via WMs:

1. Latent state imagination

Using imagination to estimate m_t: $a_t \sim \pi(\cdot \mid \hat{m}_t)$ with $\hat{m}_t$ imagined from m_{t-d} and $a_{t-d}, \ldots, a_{t-1}$

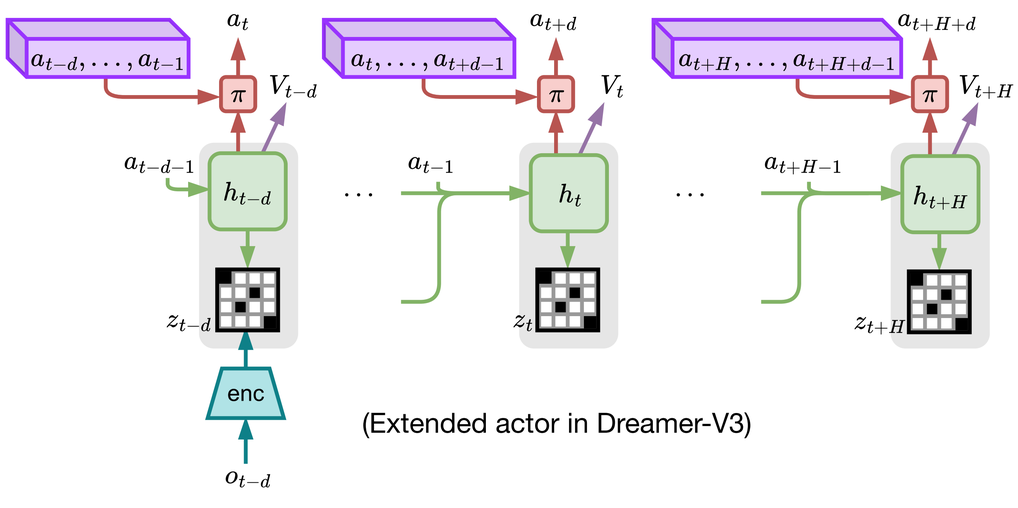


(Latent state imagination with Dreamer-V3[1])

Applying latent state imagination only in the test time refers to **Agnostic**

2. Extended actor

Conditioning the policy explicitly on the extended state while keeping the critic unchanged: $a_t \sim \pi(\cdot \mid m_{t-d}, a_{t-d}, \ldots, a_{t-1})$



(Extended actor in Dreamer-V3)

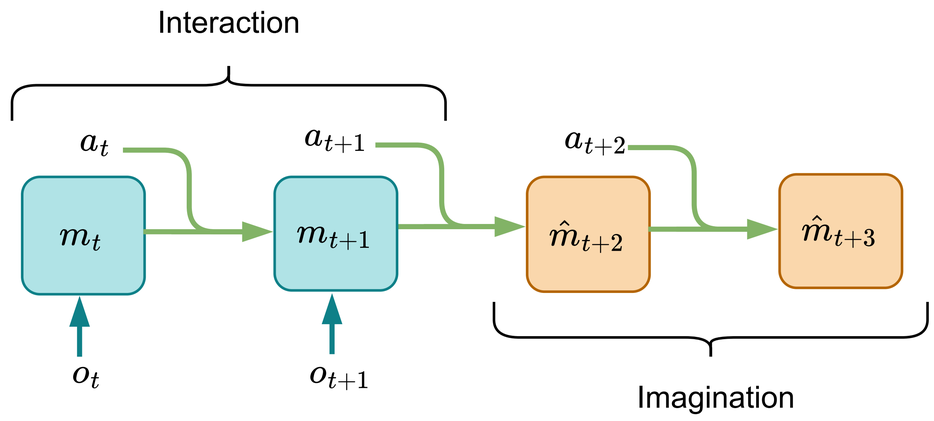
Another variant without the actions buffer is called **Memoryless**

Background

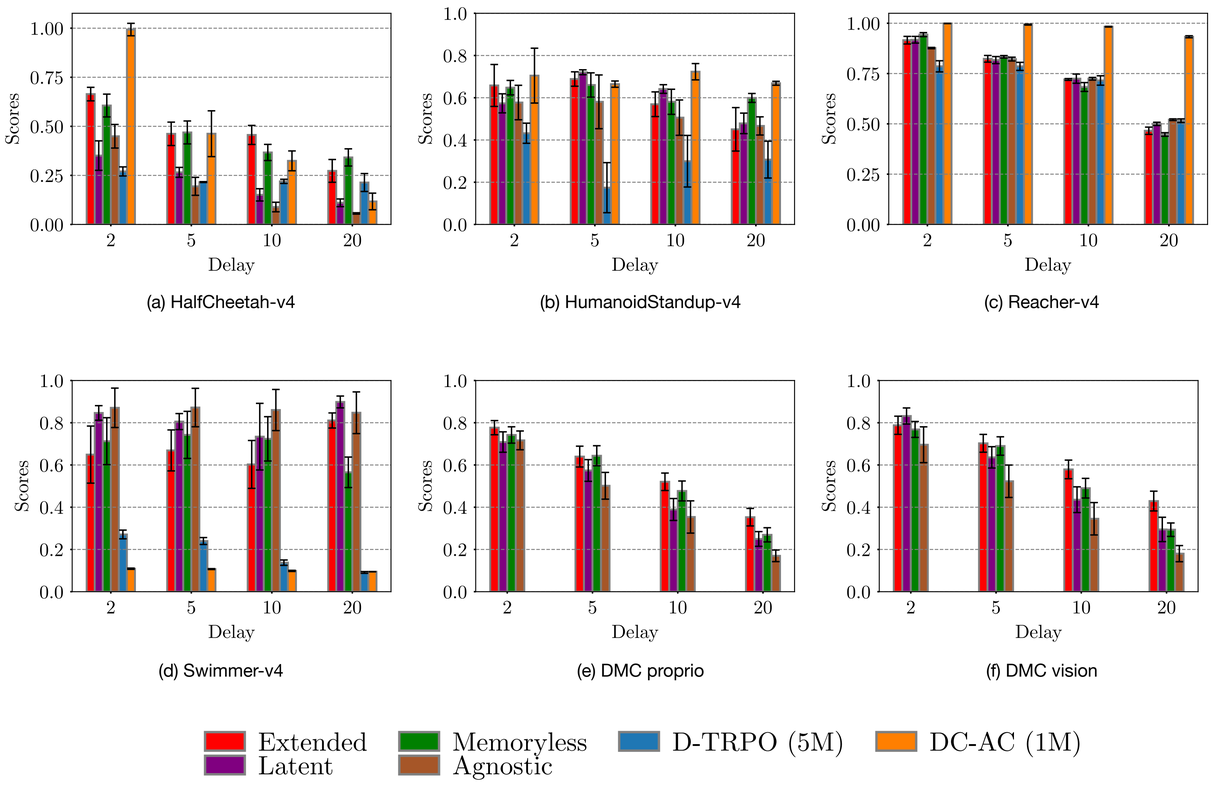
Delayed MDP: $\langle \mathscr{M}, \mathscr{D} \rangle$ where $\mathscr{M} = \langle S, A, \mathscr{T}, \gamma \rangle$ is MDP and $\mathscr{D}$ is observation delays distribution

DMDP = MDP with extended states $x_t = (s_{t-d}, a_{t-d}, \ldots, a_{t-1})$

Two modes of world models:



Experimental Results



- Ours is comparable with DC/AC[2] and outperforms D-TRPO[3]
- Latent and Memoryless models are effective for shorter delays, while the Extended model is better for longer delays but with added architectural complexity
- Extended improves up to **2.5x** over Agnostic

References

[1] Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. arXiv, 2023.
[2] Pierre Liotet, Erick Venneri, and Marcello Restelli. Learning a belief representation for delayed reinforcement learning. IJCNN, 2021.
[3] Yann Bouteiller, Simon Ramstedt, Giovanni Beltrame, Christopher Pal, and Jonathan Binas. Rein- forcement learning with random delays. ICLR, 2020.

Check out the full paper for more!