Dream to Control
Learning Behaviors by Latent Imagination

Danijar Hafner$^{12}$, Timothy Lillicrap$^3$, Jimmy Ba$^2$, Mohammad Norouzi$^1$

$^1$Google Brain $^2$University of Toronto $^3$DeepMind

1 We introduce Dreamer

We present Dreamer, an RL agent that solves visual control tasks purely by latent imagination inside of a world model.

To efficiently learn behaviors, we backpropagate multi-step value gradients through imagined latent transitions.

Dreamer learns off-policy and is applicable to both continuous and discrete actions and early episode ends.

Dreamer exceeds top model-free agents in final performance, data-efficiency, and wall-clock time.

2 Agent Overview

Algorithm 1: Dreamer

1. Initialize dataset with a few random episodes.
2. While not converged do
   1. Draw sub sequences from dataset at random.
   2. Update world model using representation learning.
   3. Use action model to imagine trajectories from all latent states.
   4. Update value model for Bellman consistency on imagined rewards.
   5. Update action model to maximize imagined values.
3. Collect episode to grow the dataset using the action model.

3 Learning Latent Dynamics

World models are an explicit way to summarize an agent’s experience and represent its knowledge about the world. When inputs are large, predicting forward in a compact latent space lets us imagine thousands of trajectories in parallel.

Dreamer is compatible with any representation objective, e.g. reconstruction, contrastive estimation, or reward prediction.

4 Learning Behaviors in Imagination

Dreamer learns to predict actions and values from the compact model states.

Start imagined trajectory from each latent state of the current data batch.

Value model optimizes Bellman consistency for current action model.

Action model maximizes values by backpropagating through latent transitions.

Reparam. gradients for latent states and actions, ST for discrete actions.

5 Comparison to Existing Methods

20 visual control tasks posing challenges incl. contact dynamics, sparse rewards, many degrees of freedom, and 3D environments.

Dreamer outperforms the previous best D4PG agent in final performance, sample-efficiency, and wall-clock time. We use the same hyper parameters across all tasks.

6 Ablation Studies

Robustness to the imagination horizon

Effect of value model and action model

Choice of representation learning method

Contact: mail@danijar.com  Twitter: @danijarh

Project website with videos: danijar.com/dreamer