CS590: Socially Cognizant Robotics
Deliverable 1

Group Members:
Rohit Bellam, rsb204
Karim Smires, ks1686

February 9th, 2026

Selected Paper

Title: *Dream to Control: Learning Behaviors by Latent Imagination*
Authors: Danijar Hafner, Timothy Lillicrap, Jimmy Ba, Mohammad Norouzi
Venue: ICLR 2020
Paper: https://arxiv.org/pdf/1912.01603
Code: https://github.com/danijar/dreamer

Summary

Dreamer is a model-based reinforcement learning agent that learns long-horizon behaviors from high-dimensional image observations purely by *latent imagination*. The problem is formulated as a POMDP, and the architecture consists of three interleaved components: a latent dynamics model, a behavior learning module, and an environment interaction loop. The dynamics model includes a representation model that encodes observations and actions into compact latent states, a transition model that predicts future latent states without generating images (enabling parallel imagination of thousands of trajectories), and a reward model that predicts rewards from latent states. Behavior learning uses an actor-critic approach: the action model outputs tanh-transformed Gaussian actions with reparameterized sampling, while the value model estimates expected rewards beyond the imagination horizon using V_λ returns—an exponentially-weighted average of multi-step estimates balancing bias and variance. The key novelty is using stochastic backpropagation to propagate analytic gradients of state values back through the learned dynamics, rather than relying on derivative-free optimization. The world model is trained via image reconstruction using a Recurrent State Space Model (RSSM) optimized with a variational lower bound; ablations show pixel reconstruction outperforms both contrastive estimation and reward-only prediction.

Evaluated on 20 visual control tasks from the DeepMind Control Suite—spanning contact dynamics, sparse rewards, and 3D environments such as Quadruped, Hopper, and Walker—Dreamer achieves an average score of 823 after 5×10^6 steps, surpassing D4PG (786 after 10^8 steps) and PlaNet (332), while training in roughly 3 hours per million steps versus 11 for PlaNet and 24 for D4PG. Ablations confirm that removing the value model degrades long-horizon performance and that Dreamer is robust across imagination horizons. The paper also demonstrates applicability to discrete actions and early termination on Atari and DeepMind Lab. Overall, Dreamer bridges model-based data efficiency with model-free asymptotic performance, setting a new standard for visual RL.