World Models: DreamerV3, IRIS, and Beyond

Ayeen Poostforoushan   Kooshan Fattah

Sharif University of Technology

Background

Reinforcement learning is often **sample inefficient**, requiring millions of interactions with the environment. **World models** address this by learning a compact latent representation of the environment and using it to *imagine rollouts*, reducing the need for real experience. The core research question: **Can learned world models significantly improve sample efficiency while scaling to complex, diverse domains?**

Dyna-Q: The Foundation

DYNA-Q laid the groundwork for modern world models by introducing the concept of learning and planning with an internal model.

Core Ideas:

- Learn environment model from real experience
- Generate synthetic experience through model rollouts
- Update policy using both real and synthetic data

While DYNA-Q used simple tabular representations, modern world models extend these ideas to high-dimensional spaces with deep neural networks and latent representations.

World Models (2018)

- **VAE (Vision Model)**: Compresses raw images into a compact latent vector, making high-dimensional visual input tractable.
- **MDN-RNN (Memory Model)**: Predicts future latent states and captures temporal dependencies by modeling sequence dynamics.
- **Controller**: A small, simple policy network using output of the RNN. Trained with evolutionary strategies because the gradient wouldn't get passed through the roll-out samplings.

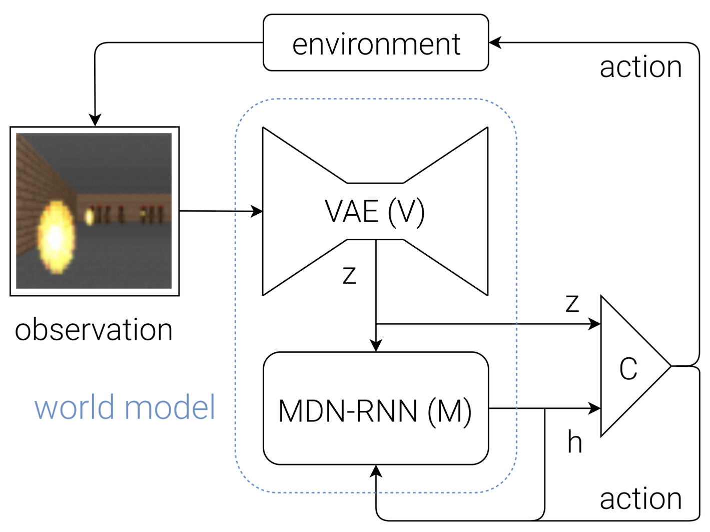


Figure 1. World Model Schematic

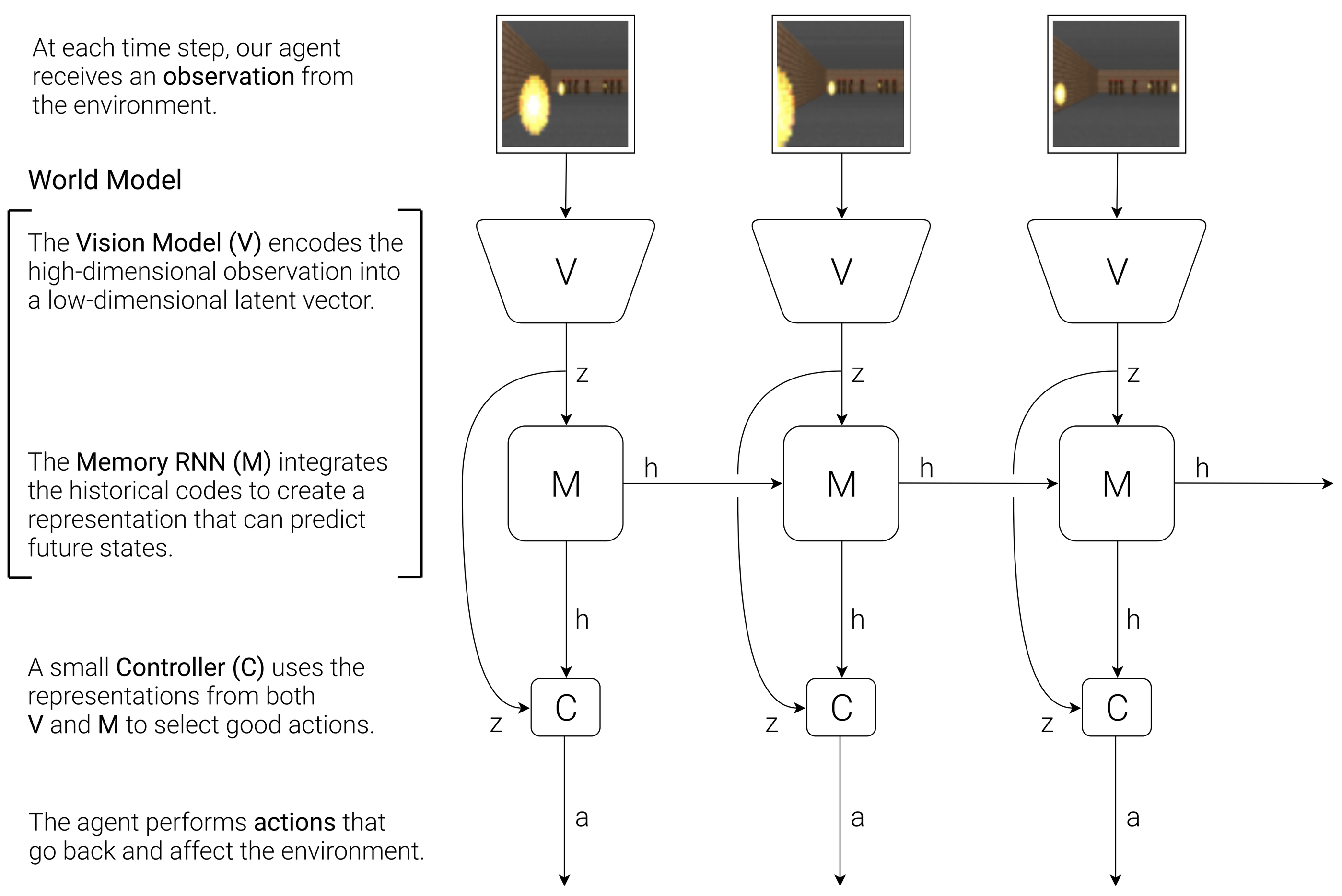
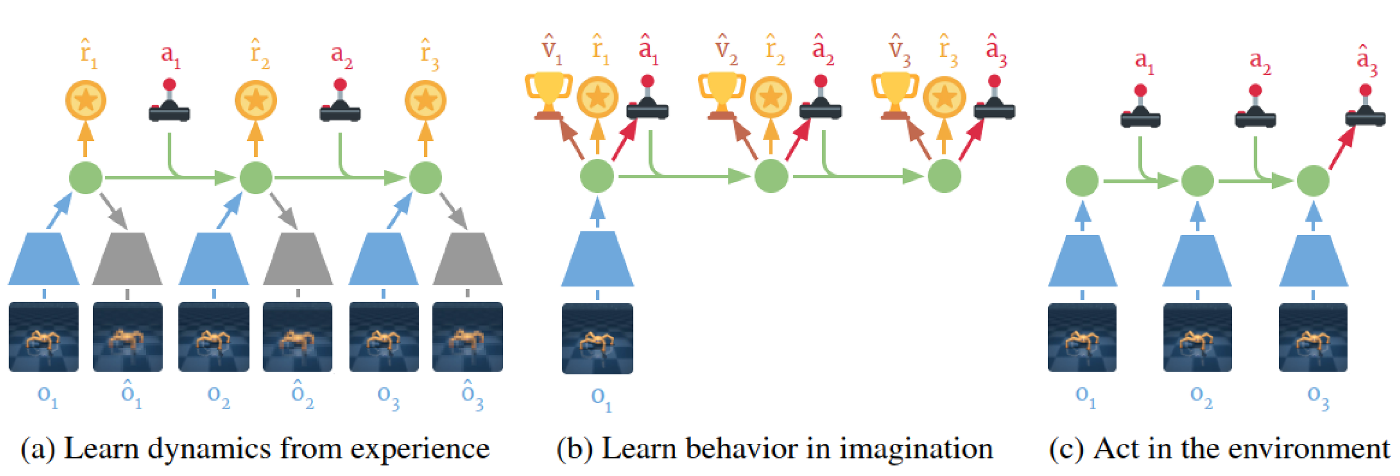


Figure 2. Original World Model Overview

Dreamer (2020)

Dreamer advanced World Models by integrating actor and critic training directly in the latent space (which is stochastic because of the stochastic latent with hidden state). The re-parametrization trick replaced evolutionary search with gradient-based updates, allowing efficient imagination-based planning and higher sample efficiency.



DreamerV3: Mastering Diverse Domains

- **Categorical Latents:** DreamerV3 represents the stochastic state z_t using multiple categorical variables with the Gumbel-Softmax reparameterization, yielding a discrete but differentiable latent space.
- **Symlog Transformation:** Rewards and values are normalized with the log function to have compressed and stable rewards.
- **Free Bits:** The KL loss for training the encoder and dynamics model is lower-bounded by a threshold to prevent posterior collapse and ensure informative latents.

IRIS: Transformers as World Models

- **Discrete Autoencoder:** Observations are compressed into sequences of discrete *tokens*, providing a compact symbolic representation of the environment.
- **Transformer World Model:** An autoregressive Transformer predicts future tokens, enabling accurate imagination of long-horizon dynamics in token space.
- **Imagination-Based Learning:** Policies are trained on imagined token rollouts, achieving strong sample efficiency on benchmarks such as Atari-100k.

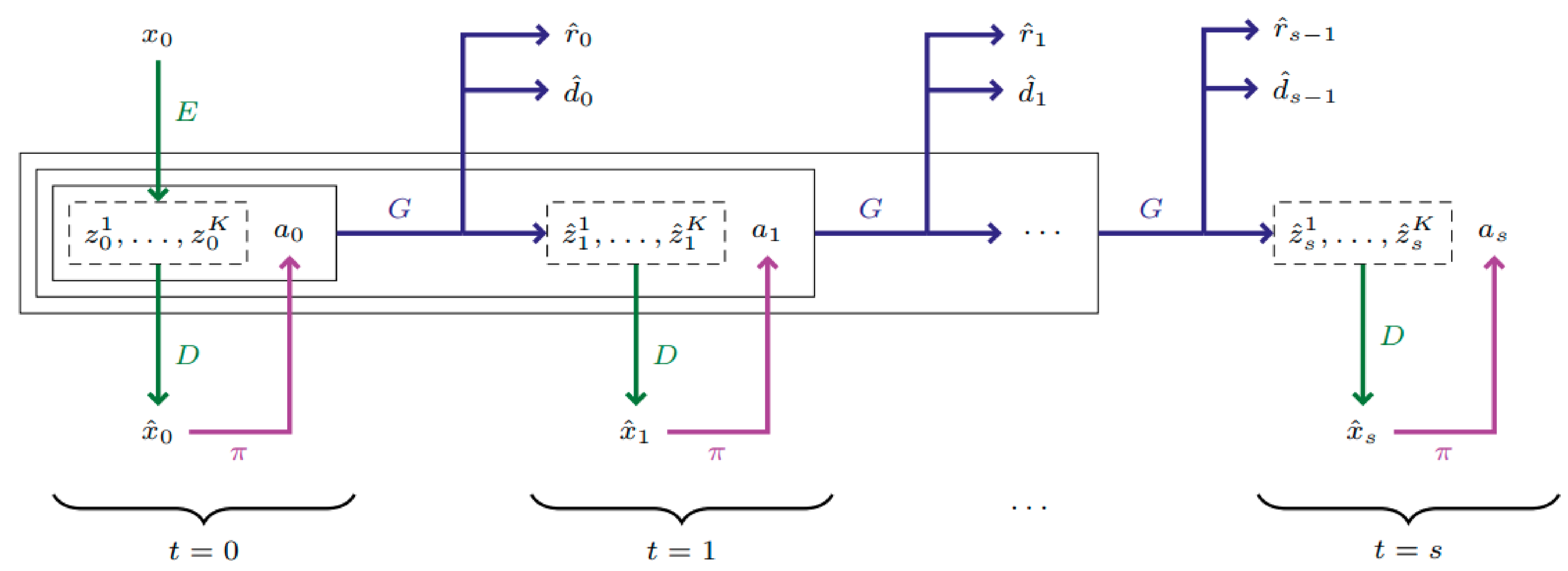


Figure 3. IRIS World Model Architecture

Performance Comparison

Method	Sample Efficiency	Latent Representation	Policy Optimization
Dyna-Q	Low	Tabular states	Q-learning with planning
World Models	Low	Continuous (VAE)	Evolution Strategies
DreamerV1	Medium	Gaussian latents	Actor–critic
DreamerV3	High	Categorical latents	Actor–critic
IRIS	Very High	Discrete tokens	Policy GD on token rollouts

Table 1. Comparison of world-model approaches across efficiency, latent representation, and policy optimization.

Key Takeaways and Future Directions

- **Latents matter:** Gaussian, categorical, or token representations affect the imagination.
- **Architecture:** RNNs use short-term memory; Transformers excel at long-horizon modeling.
- **Policy Learning:** Gradient-based updates through imagined trajectories boost sample efficiency.
- **Generality:** Modern models can work across diverse tasks without per-task tuning.

Future Directions: Multimodal inputs as tokens, long-horizon planning, and implementing new exploration frameworks in sparse-reward environments.

References

- Ha, David, and Jürgen Schmidhuber. *World Models*. NeurIPS, 2018.
- Hafner, Danijar, et al. *Learning Latent Dynamics for Planning from Pixels*. ICML, 2019.
- Hafner, Danijar, et al. *Dream to Control: Learning Behaviors by Latent Imagination*. ICLR, 2020.
- Hafner, Danijar, et al. *Mastering Diverse Domains through World Models*. arXiv, 2023.
- Mitchell, Eric, et al. *IRIS: Improving Sample Efficiency via Iterative Rollouts with Informative Signals*. ICML, 2025.