Evolutionary Planning in Latent Space

Thor V.A.N. Olesen[*,1] Dennis T.T. Nguyen[*,1] Rasmus B. Palm[1] Sebastian Risi[1,2]

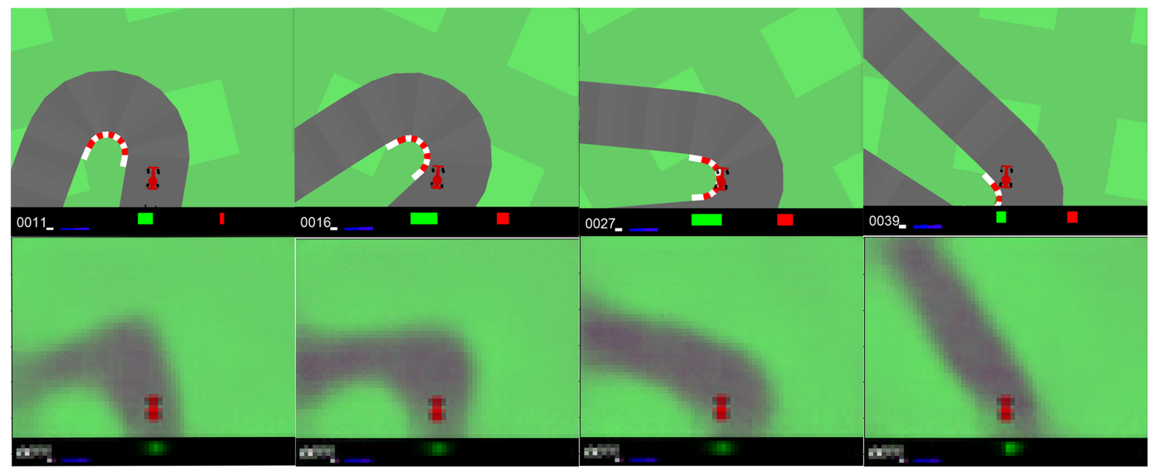
[1]IT University of Copenhagen, [2]modl.ai, [*] equal contribution, *arxiv.org/abs/2011.11293*

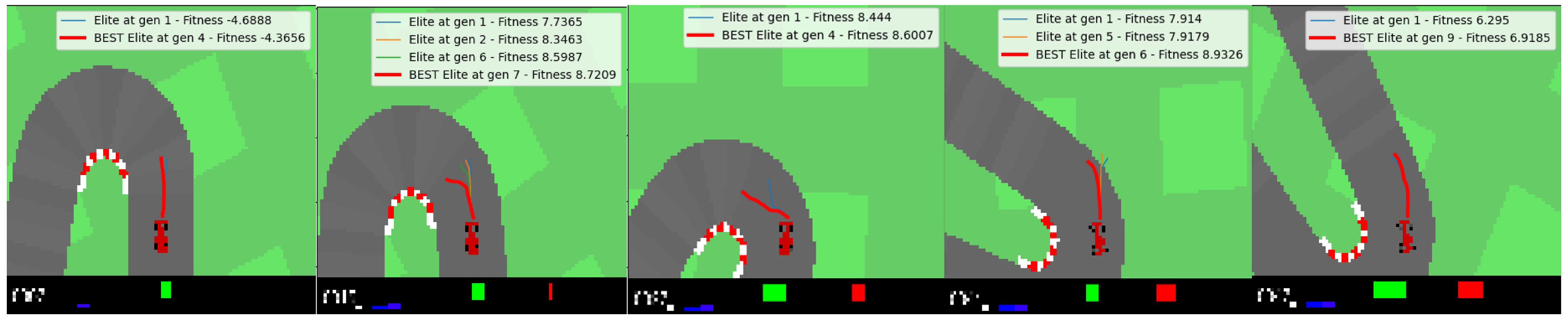
Abstract

- Planning is a powerful approach to reinforcement learning. However, it requires a world model, which is not available in many real-life problems.
- We propose to learn a world model that enables *Evolutionary Planning in Latent Space* (EPLS).
- We initialize our world model with rollouts from a random policy and iteratively refine it with rollouts from an increasingly accurate planning policy using the learned world model.
- After a few iterations, our planning agents perform well on a difficult car racing task, which demonstrates the viability of our approach.

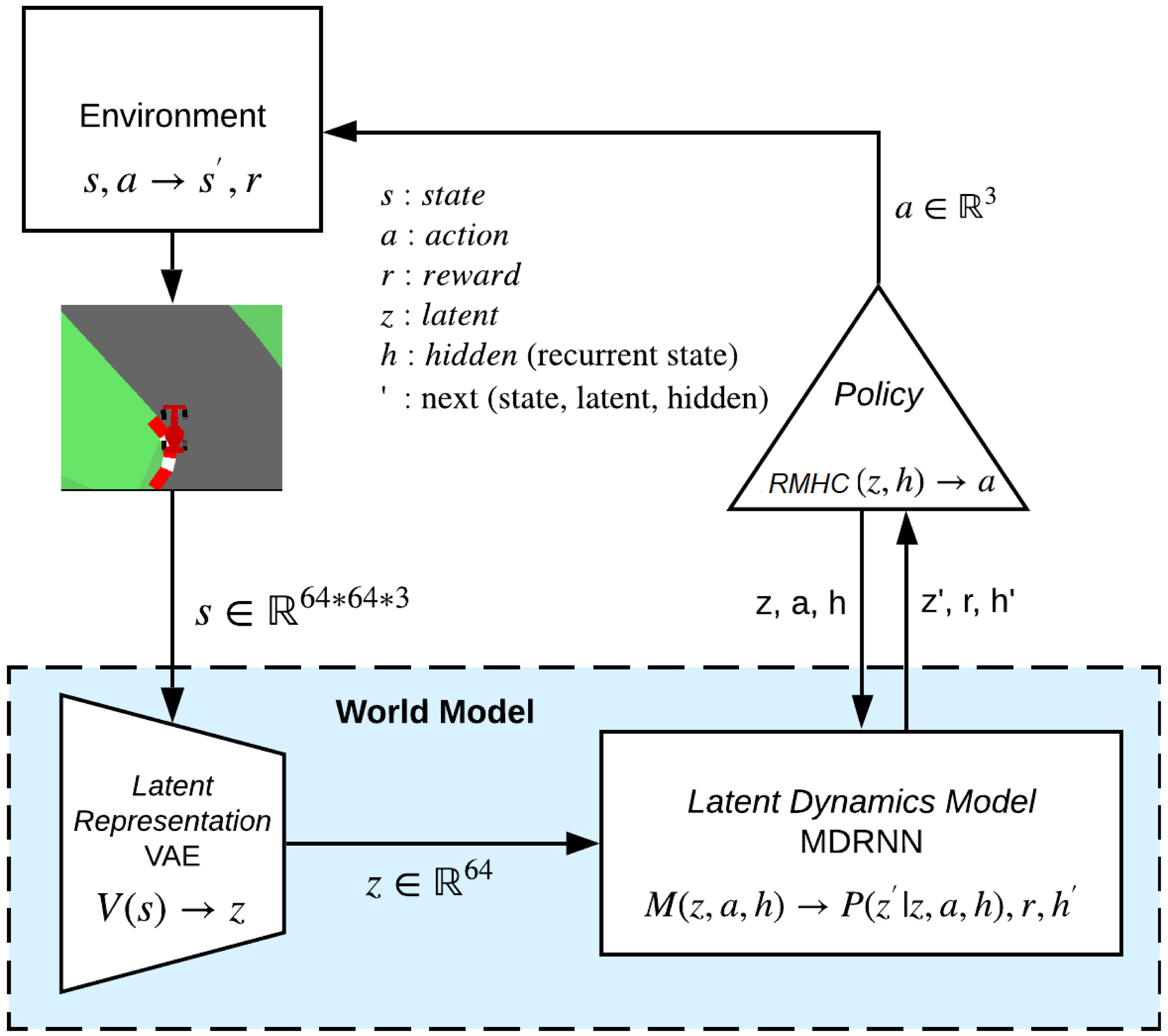
Approach

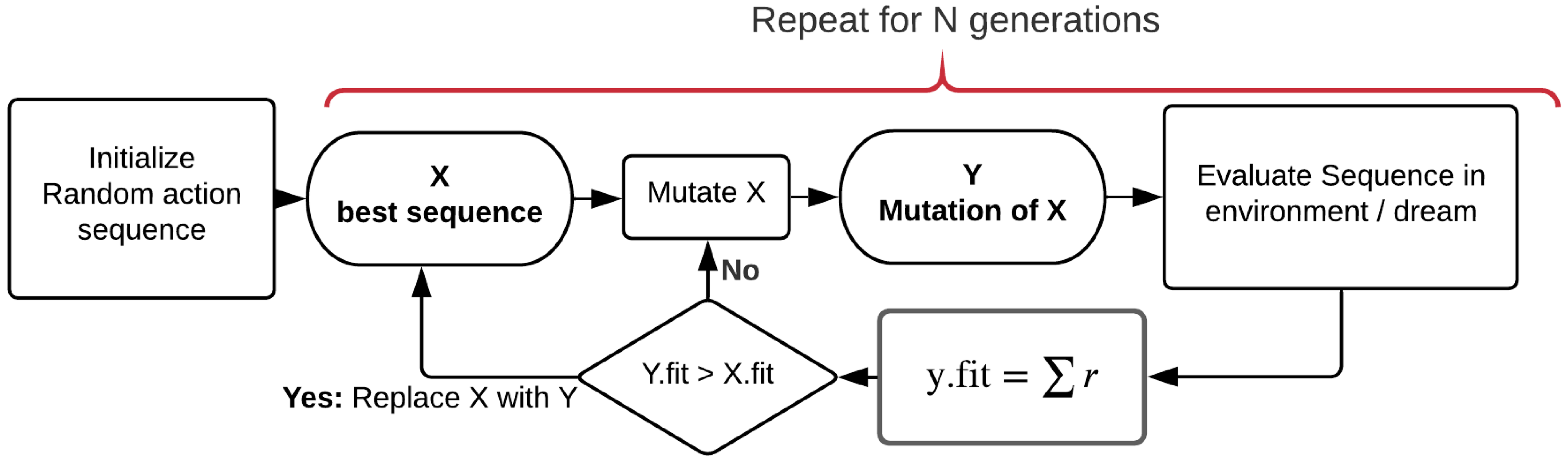
- A vision model (VAE) encodes frames, s, into low dimensional state vectors: $V(s) \to z$.
- A dynamics model (LSTM) predicts future state vectors z', rewards r' and termination τ' given the current state z and action a.
- At each time step an evolutionary planning algorithm (RMHC) evolves an optimal action sequence by evaluating it in the dynamics model.



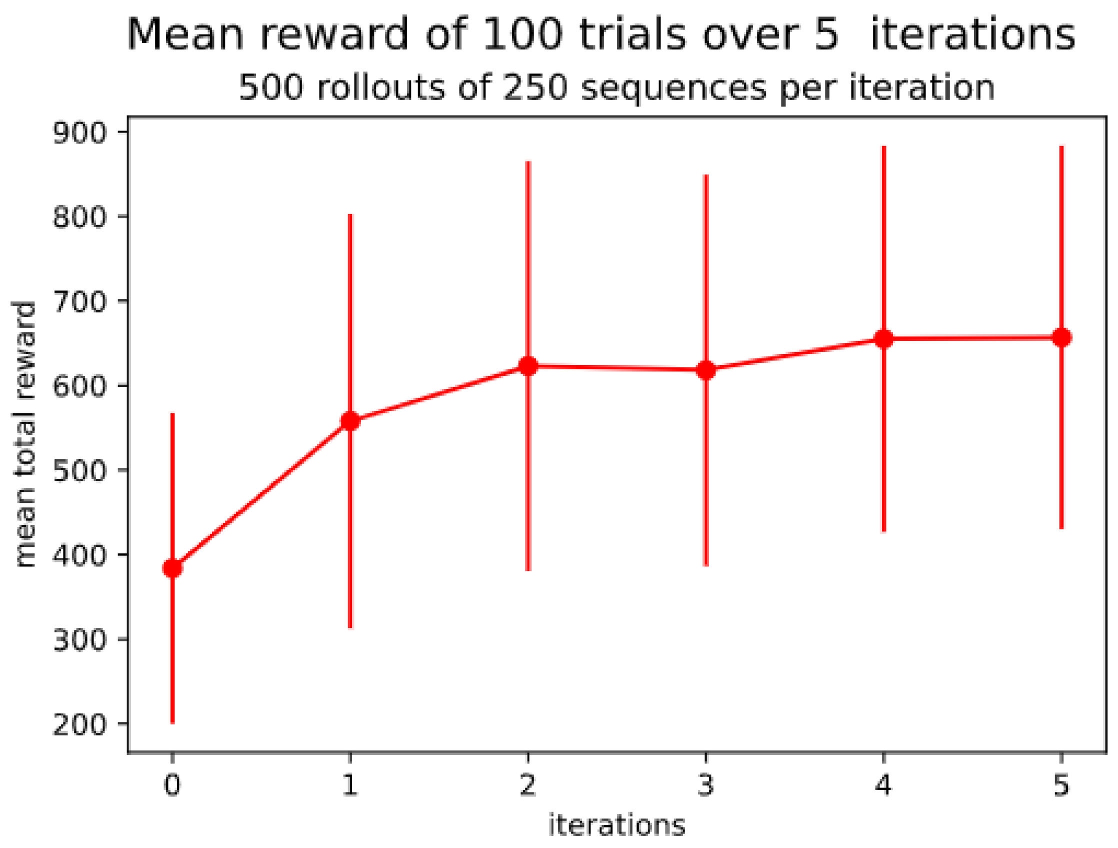


Evolutionary Planning in Latent Space





Results



- Given a world model trained on 5000 expert policy + 5000 random policy rollouts planning performs well (765 ± 102).
- Iteratively generating rollouts from the (initially random) policy and learning a better world model using the rollouts quickly and significantly improves performance, even after a single iteration.
- Using a planning horizon of 15 actions, and evolving the plan for 15 generations is sufficient for planning.

Discussion

- The agent struggles with right turns, which are rare in the data.
- Better modelling and handling of uncertainty in states and rewards will likely improve planning.
- Compare with planning by gradient descent, since dynamics model is differentiable.