EVAN RACAH

EXPLORING "WORLD MODELS"

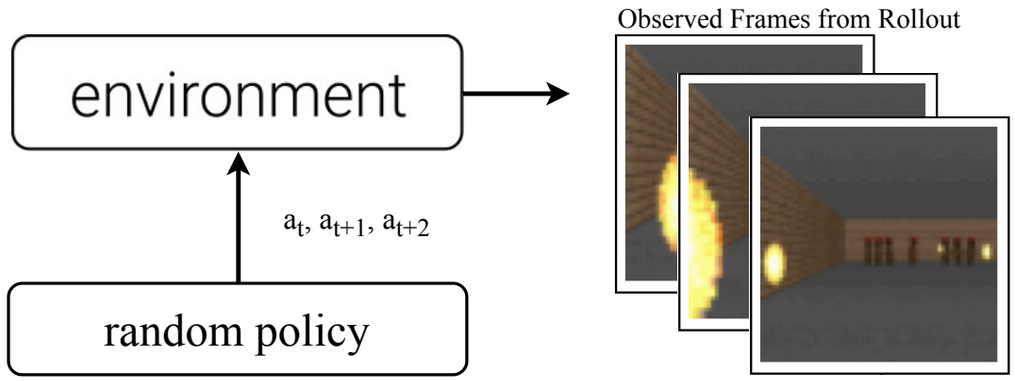
(HA ET AL., 2018)

WHY WORLD MODELS?

- Focus on the representation (controller is simple)
- Decoupling of the model components
- Action-conditioned prediction in latent space
- Evolutionary methods - interesting gradient-free approach to optimize policy
- Many avenues of exploration and extension

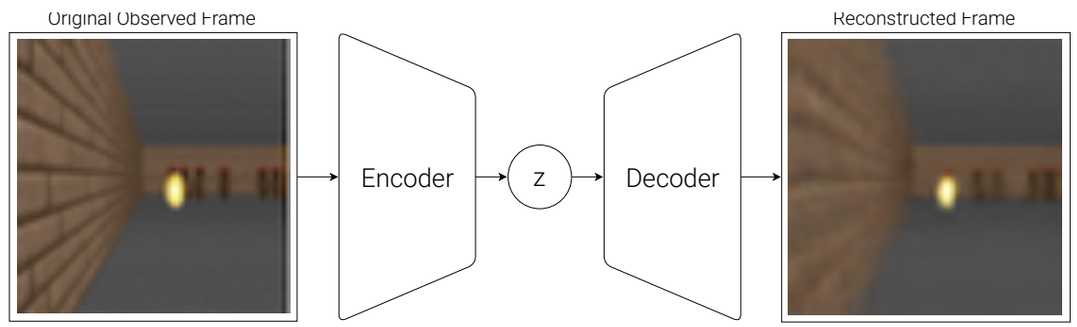
THEIR APPROACH

1



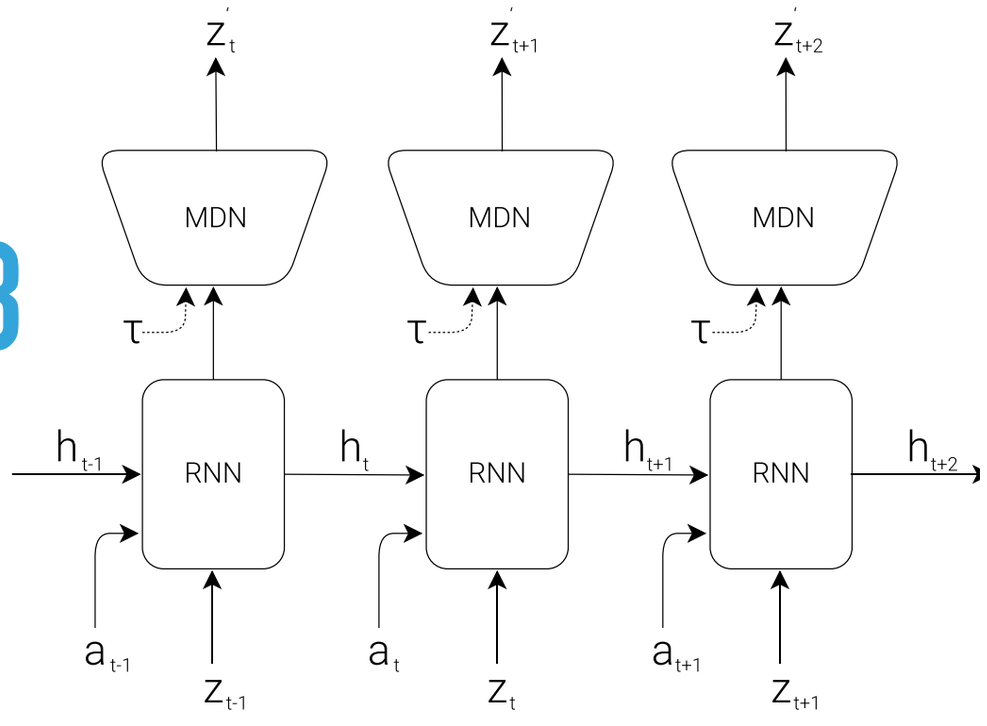
Generate actions, frames from random policy (10,000 rollouts)

2



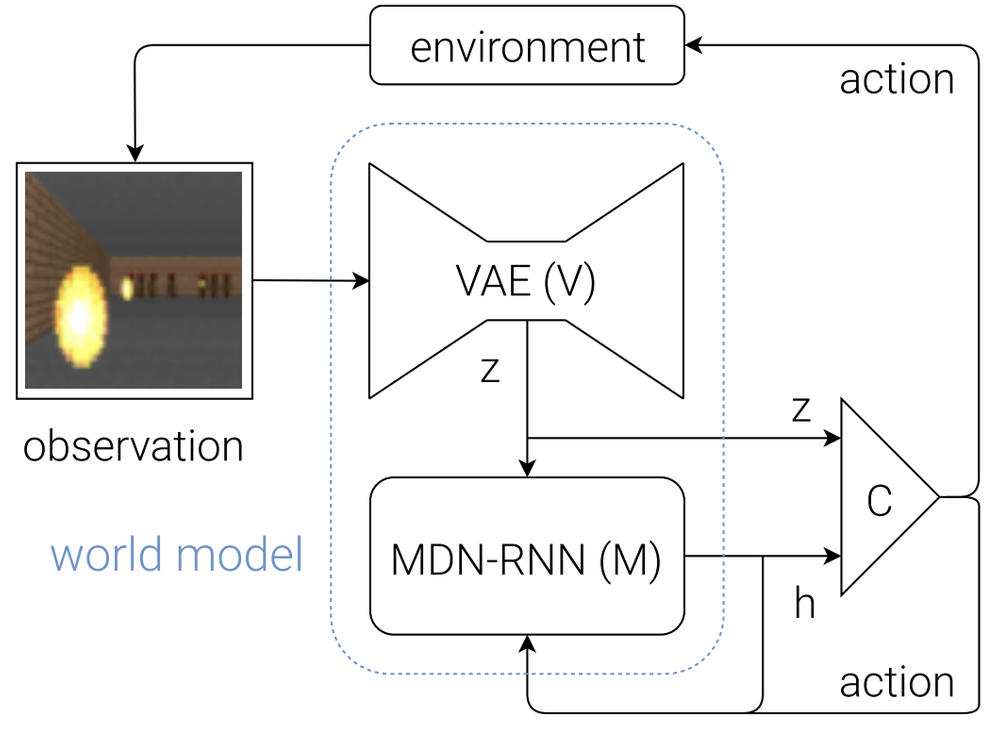
Train VAE to encode these frames into z

3



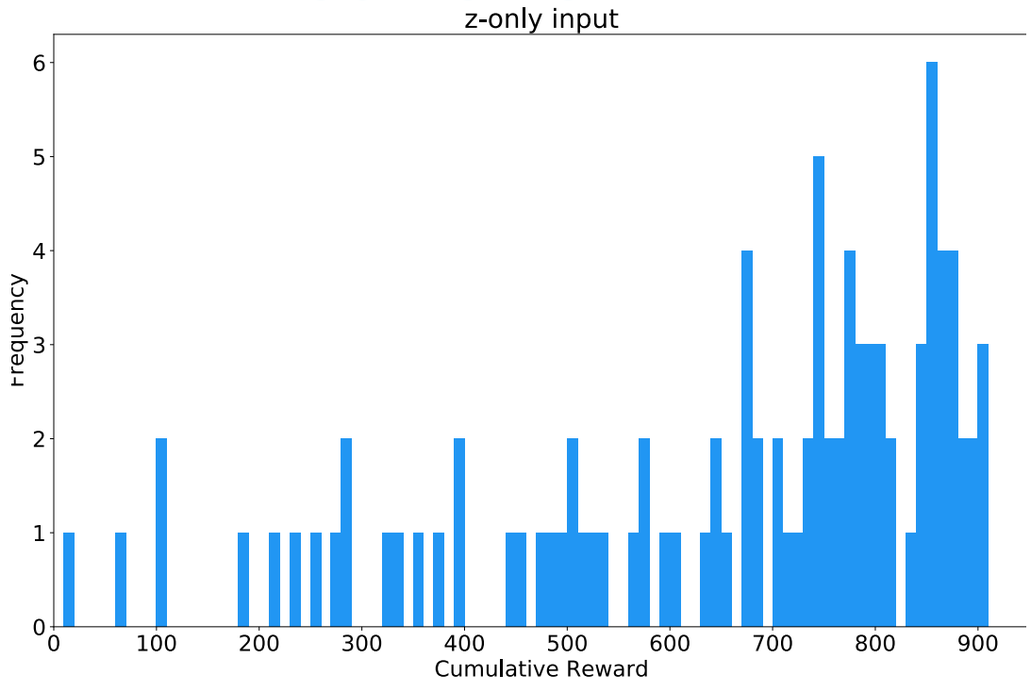
Learn to predict future z, given previous z and the action that caused it from random policy

4

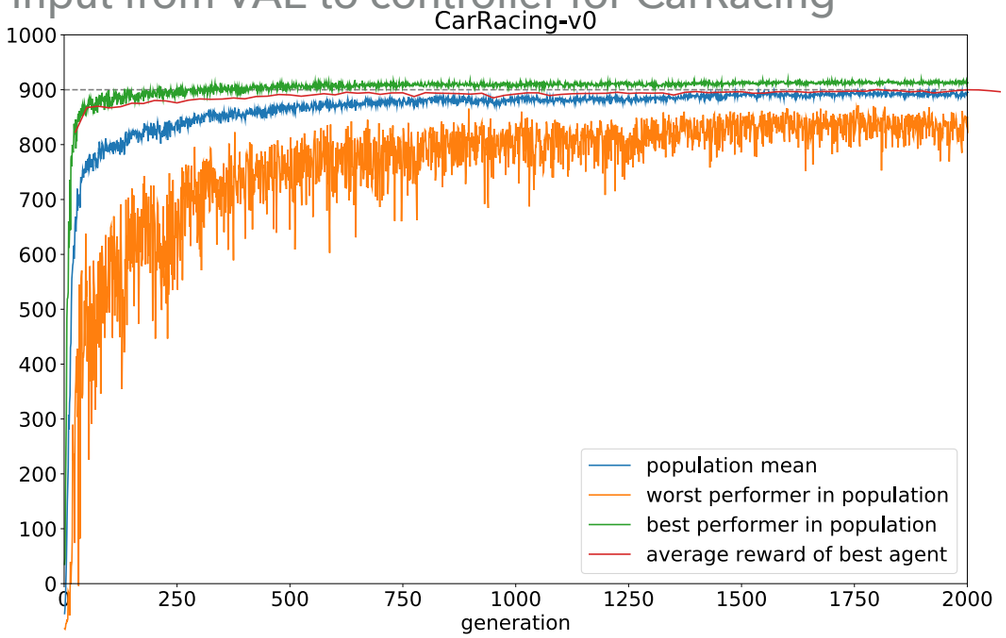


Learn best action using z from VAE and h from RNN as input using evolutionary algorithms

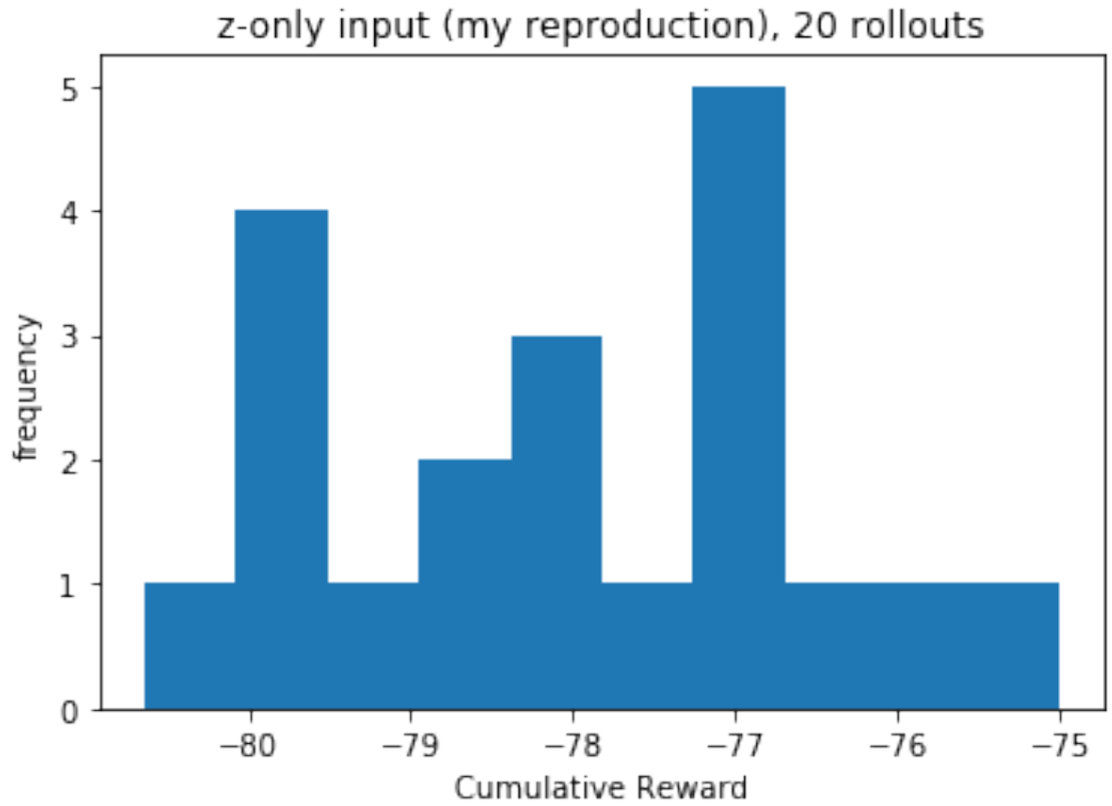
THEIR RESULTS VS. MINE



Their reward distribution for 100 rollouts with only input from VAE to controller for CarRacing

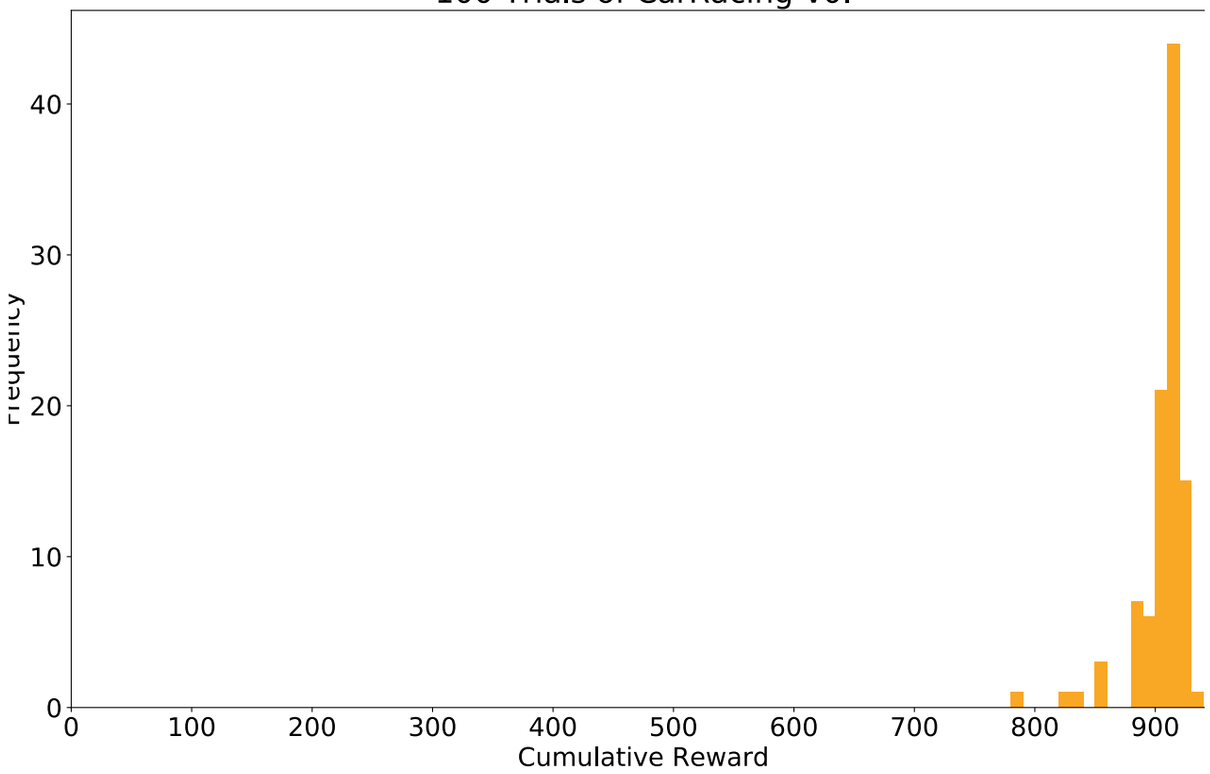


Performance over time during evolutionary search of full model

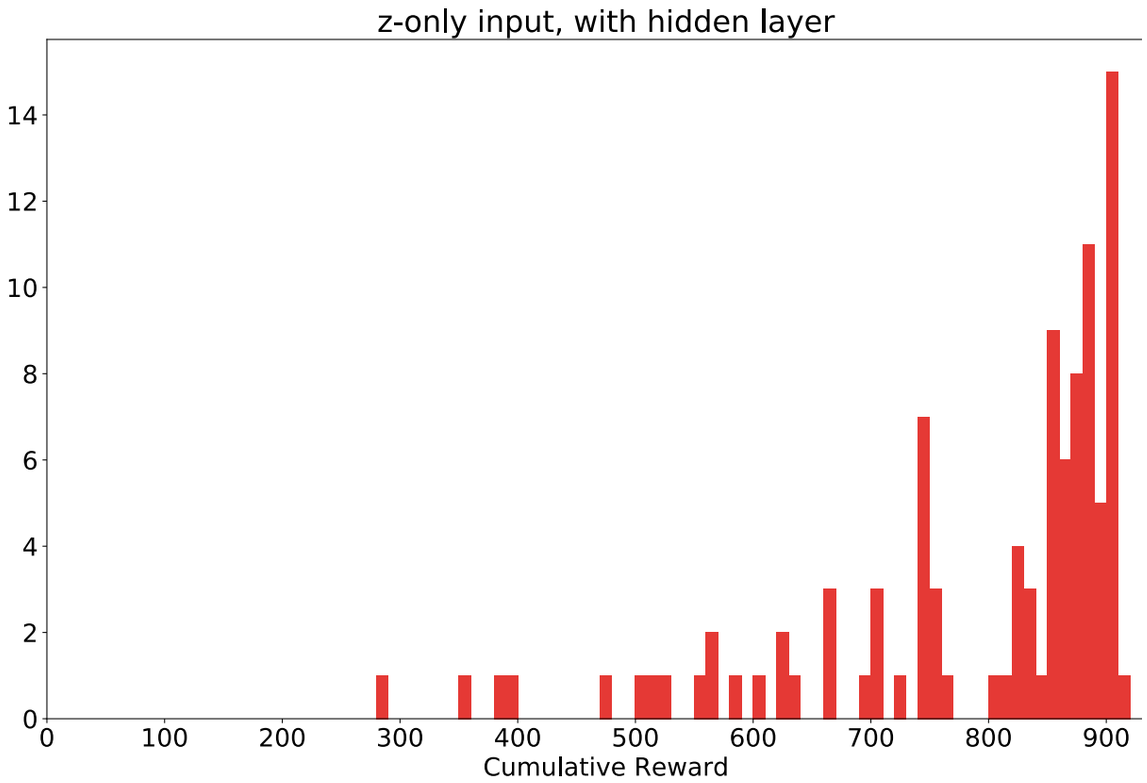


RESULTS/DISCUSSION

MORE OF THEIR RESULTS



100 rollouts with z from VAE and h from LSTM



Just z from VAE, but added complexity to controller

HOW DID IT GO?

- Doing 10,000 rollouts serially can be slow
- Evolutionary methods can be slow (especially when they rely on rollouts to evaluate)
- What more needs to be done to reproduce
  - Run evolutionary search longer to match their performance
    - Parallelize it
  - Add in features from LSTM
- Future Questions to Address
  - Policy gradient (just for controller and end-to-end)
  - Why can’t VAE see actions too?
  - Does it work for Atari?