

World Model for Robot Learning

What is a World Model?

- A model which given action and the current State or Observation/Observations can predict the next state.

World Models

David Ha¹ Jürgen Schmidhuber^{2,3}

Abstract

We explore building generative neural network models of popular reinforcement learning environments. Our *world model* can be trained quickly in an unsupervised manner to learn a compressed spatial and temporal representation of the environment. By using features extracted from the world model as inputs to an agent, we can train a very compact and simple policy that can solve the required task. We can even train our agent entirely inside of its own hallucinated dream generated by its world model, and transfer this policy back into the actual environment.

An interactive version of this paper is available at <https://worldmodels.github.io>

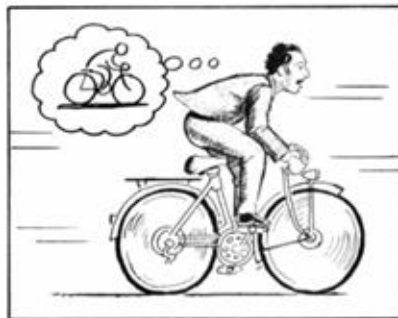
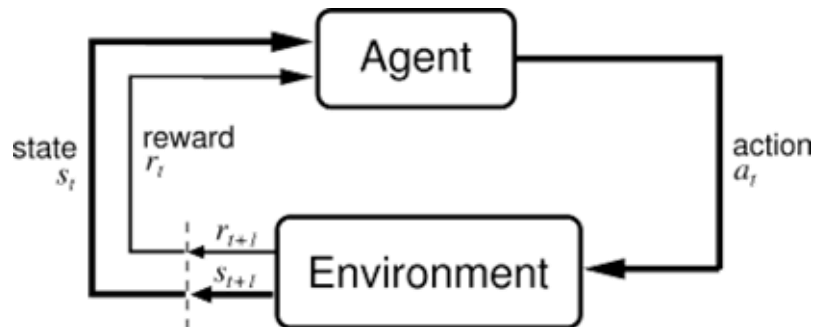
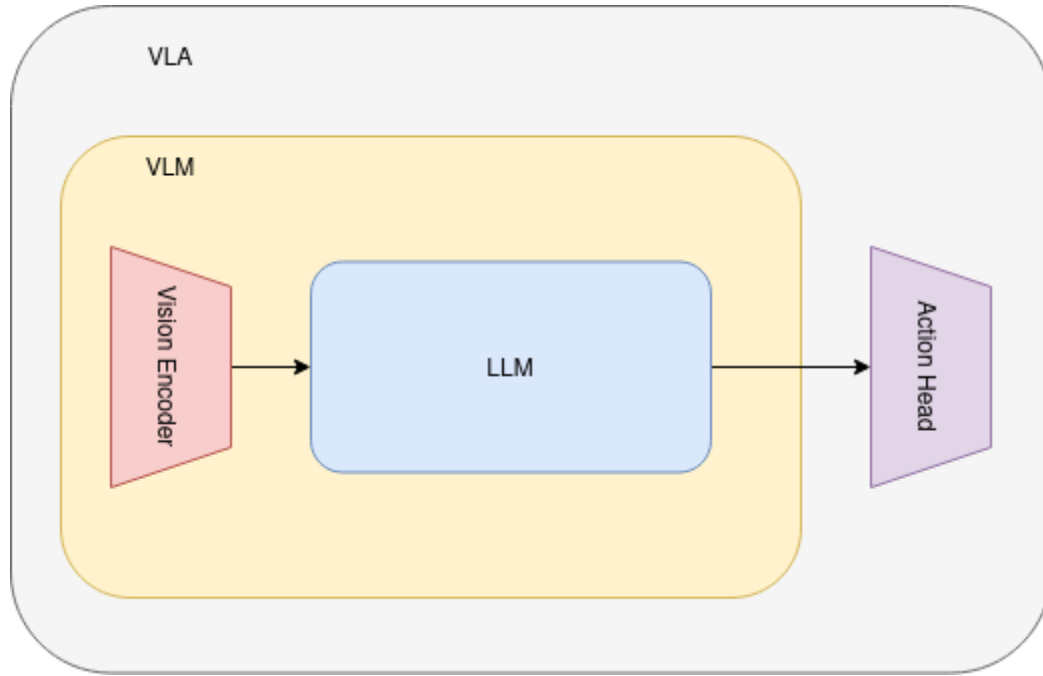


Figure 1. A World Model, from Scott McCloud's *Understanding Comics*. (McCloud, 1993; E, 2012)



What is a VLA?



Where Purely Reactive VLAs Might Fail?

- VLA's are VLM's which output actions and as a result they are good at semantic knowledge or command following. This doesn't necessarily mean they would be good in the physical world at least not without feeding them tons of data!
- They don't explicitly learn how the world evolves and don't anticipate change in the physical sense, at least not explicitly.
- For a VLA to be able to make an accurate implicit world model of the environment, you would need to collect a lot of teleoperation data.

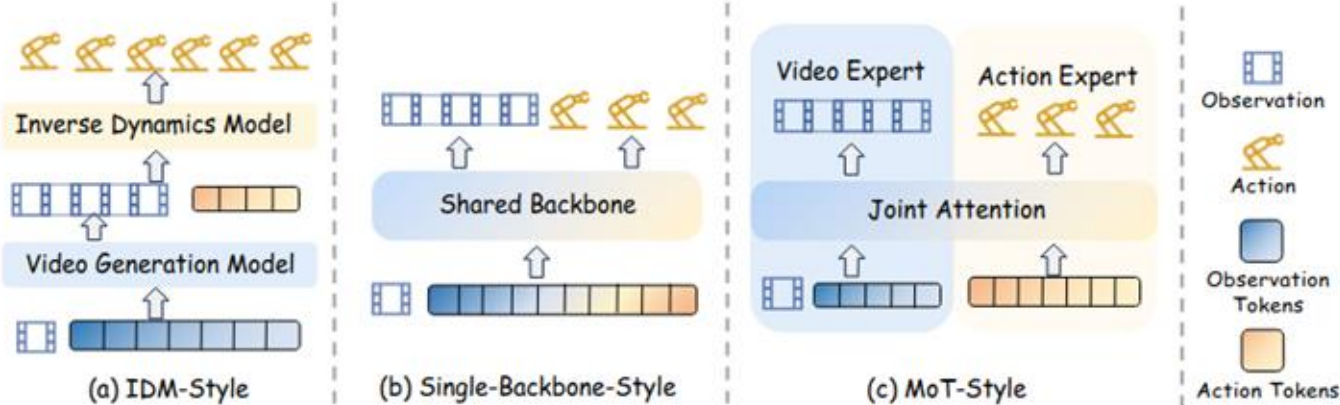
Why a World Model?

- Introducing the 'Explicit' predictive structure for anticipating how the world may evolve under the agent's behavior.
 - Can Capture Environmental Dynamics.
 - Enable Reasoning about future states in a physical context before acting.
- How can it help:
 - Foresight = Anticipate future states or action consequences before execution
 - Imagination-driven planning = compare and select candidate behavior
 - Data Amplification = Synthesize data

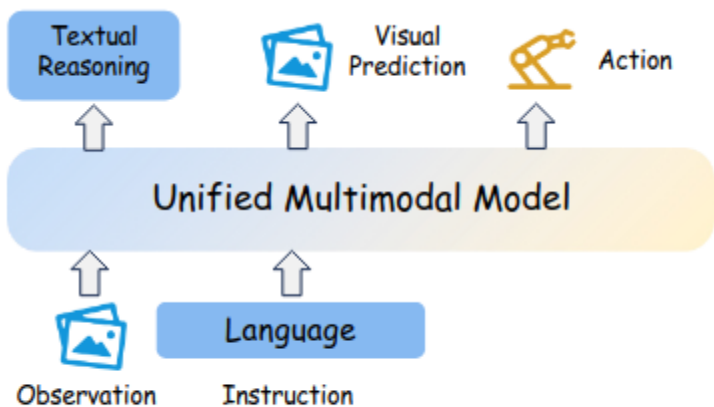
How are World Models used:

- 1) For the policy model (used as the brain of the robot)
- 2) As simulation
- 3) For synthesizing data

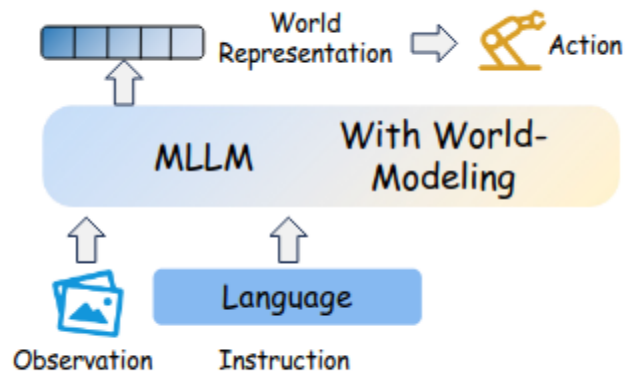
For policy using video generator models



For policy using Internalized World Models



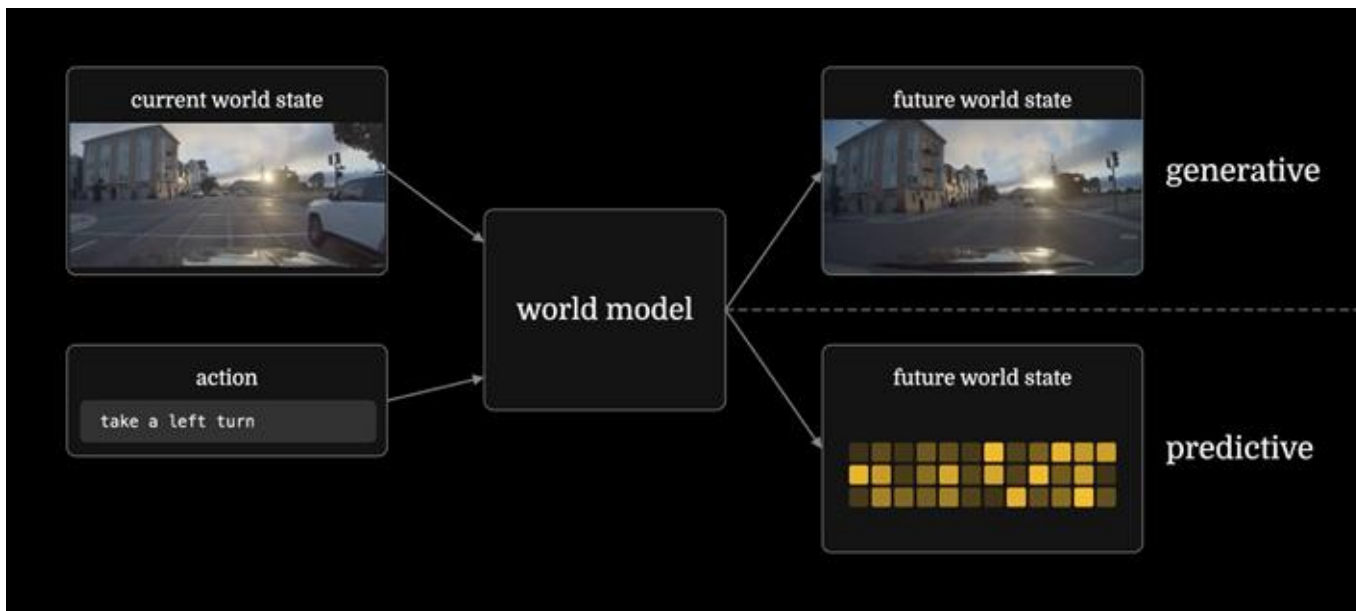
(a) Unified Vision-Language-Action Model



(b) Latent World Modeling Policy

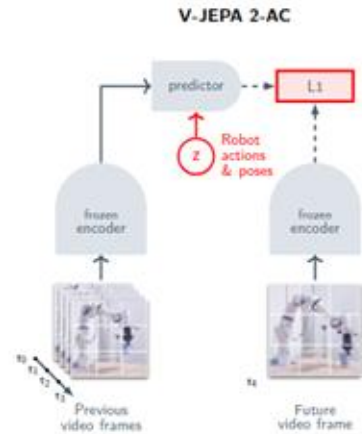
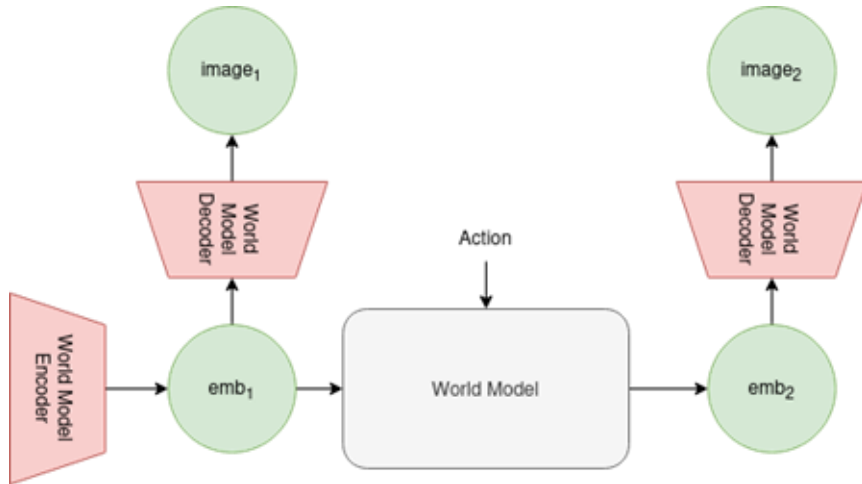
Loss Function Design

- Pure picture construction (thinking in pixel space) can make the model spend compute on parts of the scene which are not valuable.

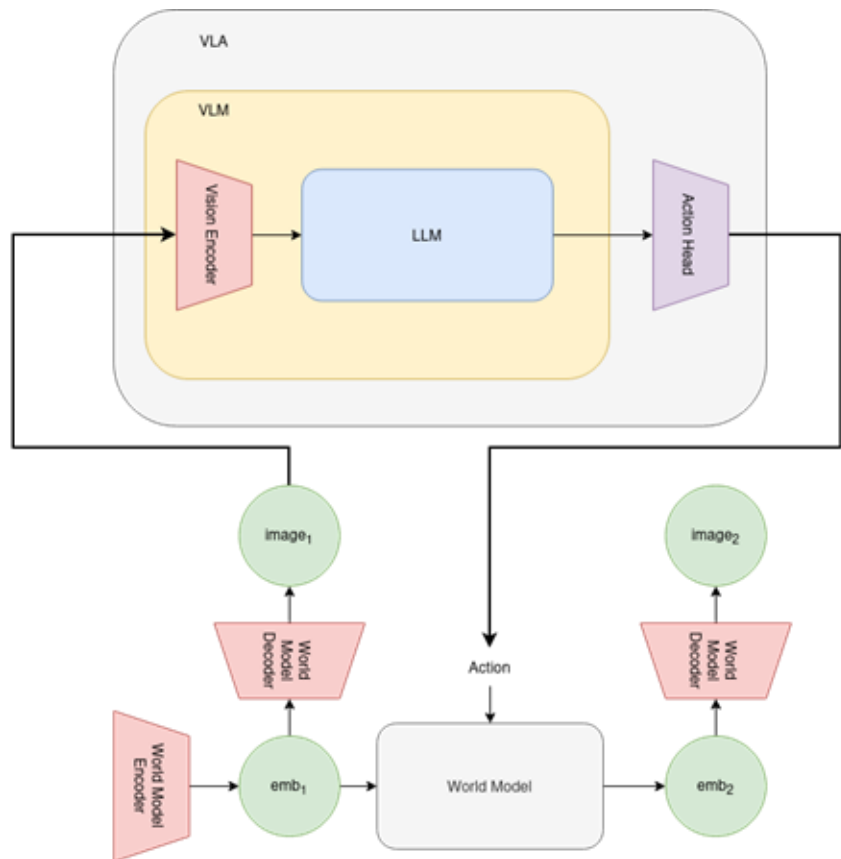


V-JEPA 2-AC

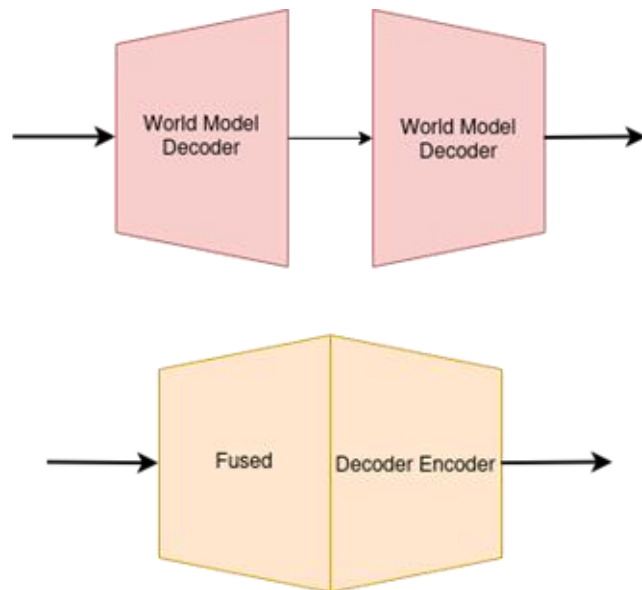
- In JEPA papers they use an MPC solver and the mean squared loss of the embedding of the goal state (embedding of the goal image) to guide the solver.
- This approach might be rather naive and in a way brute-force like.
- The model trying to go from the starting embedding to the final one doesn't have any prior knowledge and works only with the given gradients.



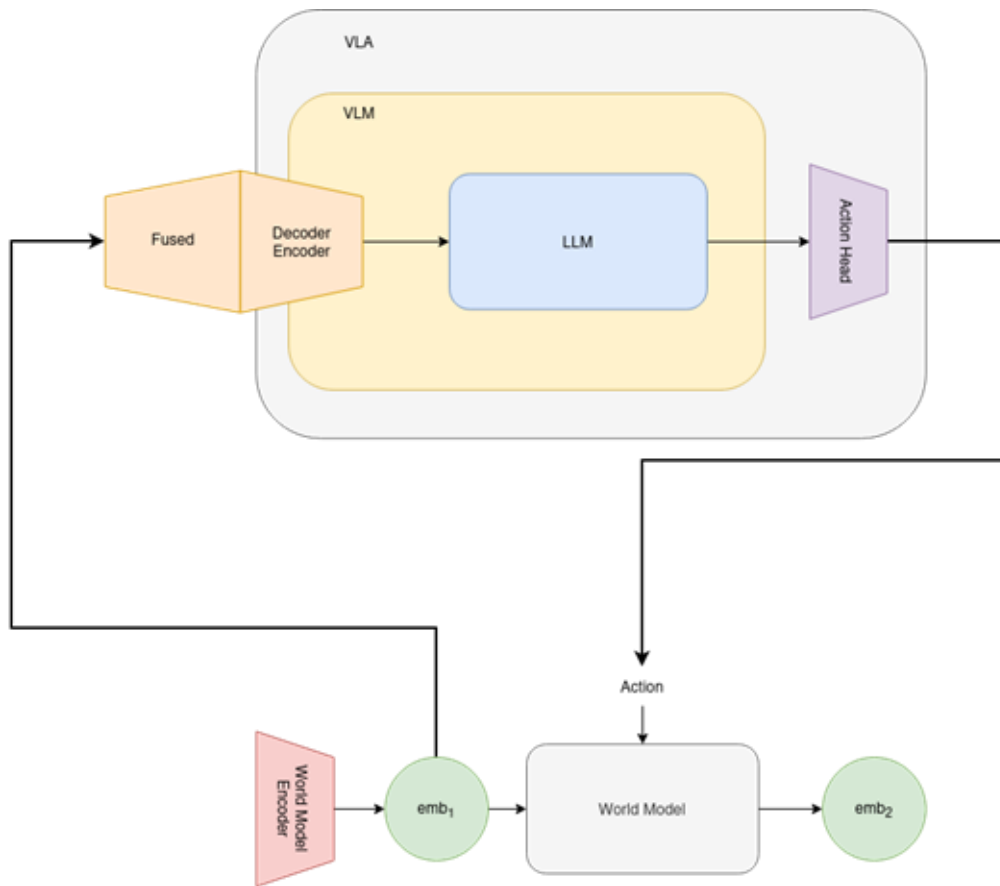
Some Ideas/Thinking Out Loud!



We can optimize this further by training a NN to fuse the two modules which would consume less parameters.



Some Ideas/Thinking Out Loud!



Steps:

- 1) Fine-tune VLA on embodiment (NicoL).
- 2) Fine-tune WM on embodiment (NicoL).
- 3) Feed starting state and goal state + the correct prompt.
- 4) RL training using the World Model as a simulation for the VLA to go to the goal state and using the mean squared error of the distance between goal embedding and current embedding as dense reward.