Learning Latent Dynamic Robust Representations for World Models

Ruixiang Sun* [1] Hongyu Zang* [1], Xin Li† [1], Riashat Islam [2]

[1]Beijing Institute of Technology, China

[2]DreamFold AI, Canada

July 2024

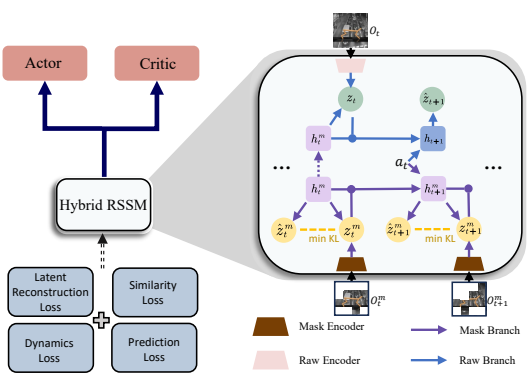
* Equal contribution
† Corresponding author

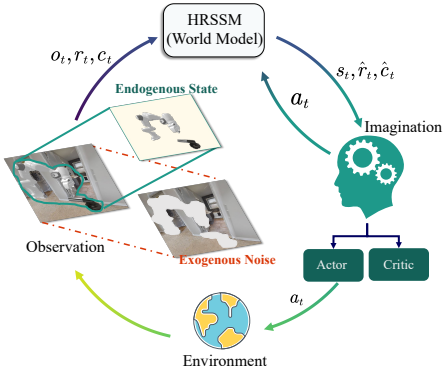
Background

How to learn sufficiently expressive state representation for a world model without the reliance of the pixel-level reconstruction?

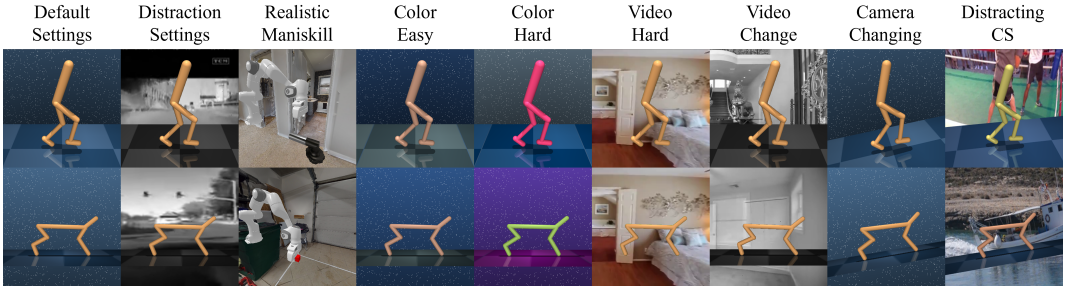
- effectively capturing task-relevant endogenous dynamics information
- be robust and compact enough to filter exogenous task irrelevant details

Hybrid RSSM


Actor
Critic
Hybrid RSSM
Latent Reconstruction Loss
Similarity Loss
Dynamics Loss
Prediction Loss
o_t
z_t
$\hat{z}_{t+1}$
h_t^m
h_{t+1}
a_t
h_t^m
h_{t+1}^m
z_t^m
$\hat{z}_t^m$
min KL
$\hat{z}_{t+1}^m$
z_{t+1}^m
min KL
o_t^m
o_{t+1}^m
Mask Encoder
Raw Encoder
Mask Branch
Raw Branch

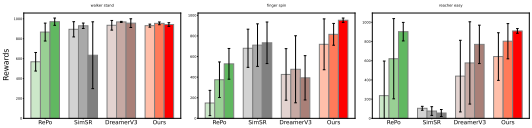

HRSSM
(World Model)
o_t, r_t, c_t
$s_t, \hat{r}_t, \hat{c}_t$
a_t
Endogenous State
Imagination
Observation
Exogenous Noise
Actor
Critic
a_t
Environment

Experimental Setup

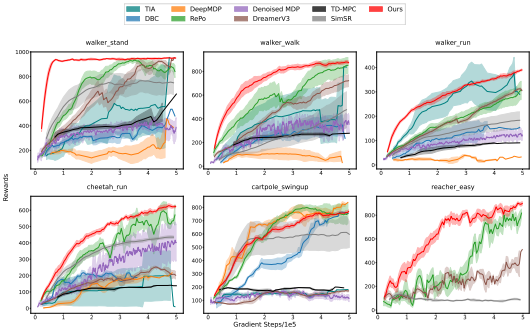


Experimental Results

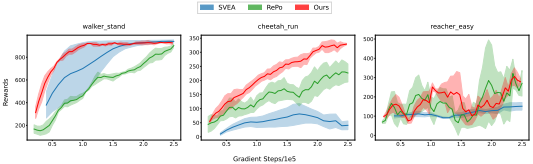
- Default Settings



- Distraction Settings



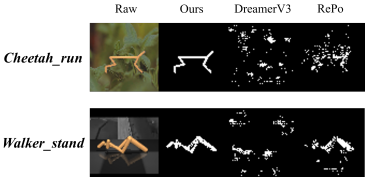
- Distracting CS



- Realistic Maniskill

Task	DreamerV3	TIA	Denoised MDP	RePo	Ours
Lift Cube	167±45	**274±173**	155±71	83±22	**284±85**
Turn Faucet	138±83	47±23	71±27	92±88	**278±97**

- Interpretability Visualization



Check our paper for . . .

- Detailed description of our method
- Theoretical guarantees
- More empirical results

paper

code