Dyn-O: Building Structured World Models with Object-Centric Representations

Zizhao Wang, Kaixin Wang, Li Zhao, Peter Stone, Jiang Bian





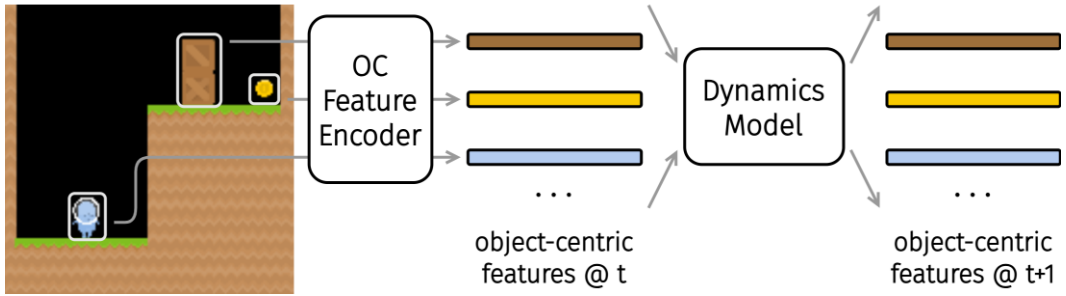
Microsoft Research | LARG Learning Agents Research Group The University of Texas at Austin | Sony AI

Problem Setup

World Models

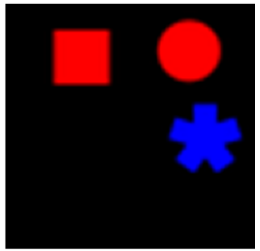
- Simulate environment dynamics in the latent space.
- Most prior methods use **monolithic** representation.
  - Considers the entire scene as a whole and ignores its internal structure.
  - For example, most interactions are object centered.

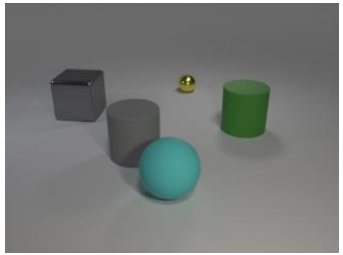
→ We propose to learn object-centric world models.



Wang, Z., K. Wang, L. Zhao, P. Stone, J. Bian. “Dyn-O”, NeurIPS 2025.

Related Work & Motivation





Prior object-centric world models are usually evaluated in simple domains.

- basic geometries
- linear motion

We want to enhance object-centric world model's applicability to more challenging environments, such as Procgen.

- rich, diverse visuals
- complex, nonlinear dynamics

Wang, Z., K. Wang, L. Zhao, P. Stone, J. Bian. "Dyn-O", NeurIPS 2025.

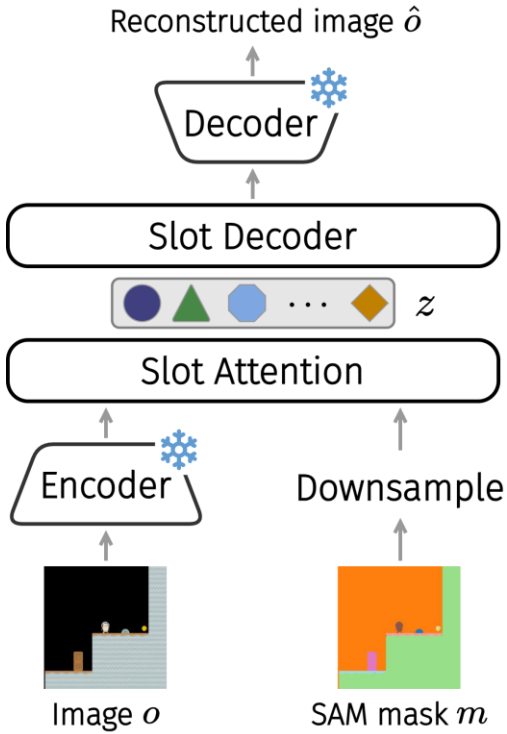
Method – object-centric representation learning

Dyn-O consists of two stages:

- object-centric representation learning
- dynamics learning

During object-centric representation learning, we

- learn on top of pretrained tokenizer
  - utilize high-quality features
  - avoid learning from scratch with raw pixels
- guide object-slot binding with segmentation masks from SAM2

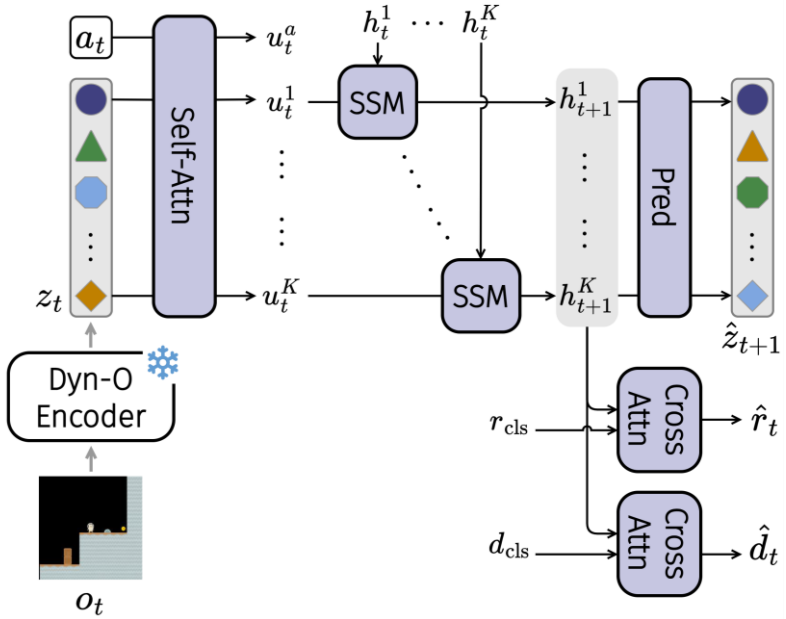


Wang, Z., K. Wang, L. Zhao, P. Stone, J. Bian. “Dyn-O”, NeurIPS 2025.

Method – dynamics learning

During dynamics learning, we

- adopt state-space models (SSMs) as the backbone
  - good at long-range temporal dependencies

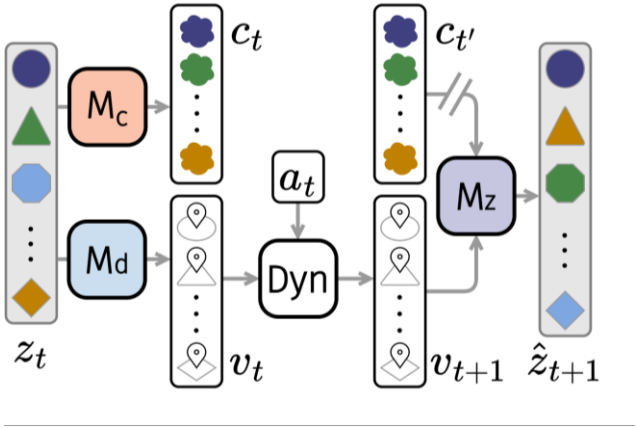


Wang, Z., K. Wang, L. Zhao, P. Stone, J. Bian. “Dyn-O”, NeurIPS 2025.

Method – dynamics learning

During dynamics learning, we

- adopt state-space models (SSMs) as the backbone
  - good at long-range temporal dependencies
- disentangle static and dynamic features
  - synthesize novel scenarios for agent learning



Wang, Z., K. Wang, L. Zhao, P. Stone, J. Bian. “Dyn-O”, NeurIPS 2025.

Experiments

Procgen

Wang, Z., K. Wang, L. Zhao, P. Stone, J. Bian. “Dyn-O”, NeurIPS 2025.

Results – object-centric representation

(a) Oracle (b) Dyn-O (ours) (c) SOLV

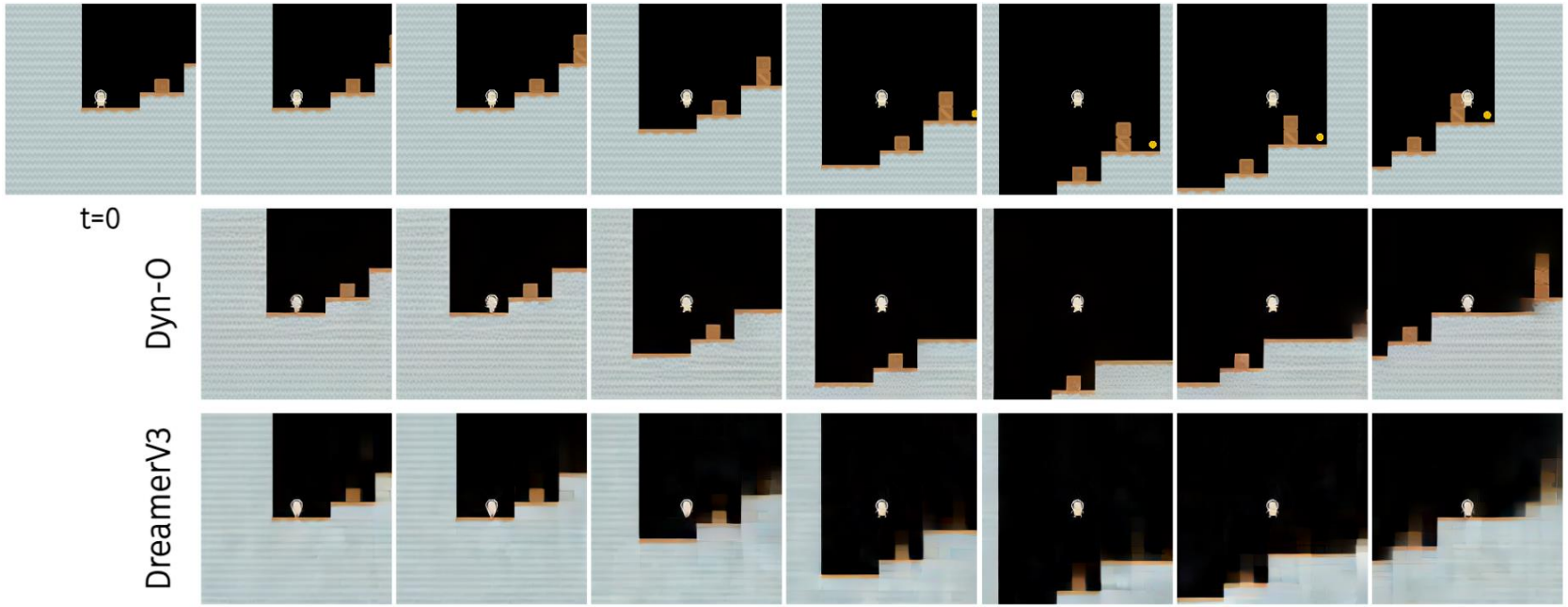
Method	FR-ARI (↑)
Oracle	0.96
Dyn-O (ours)	0.80
SOLV	0.54

(d) slot-object binding accuracy

Dyn-O learns more accurate object binding than the baseline, by distilling from SAM2.

Wang, Z., K. Wang, L. Zhao, P. Stone, J. Bian. “Dyn-O”, NeurIPS 2025.

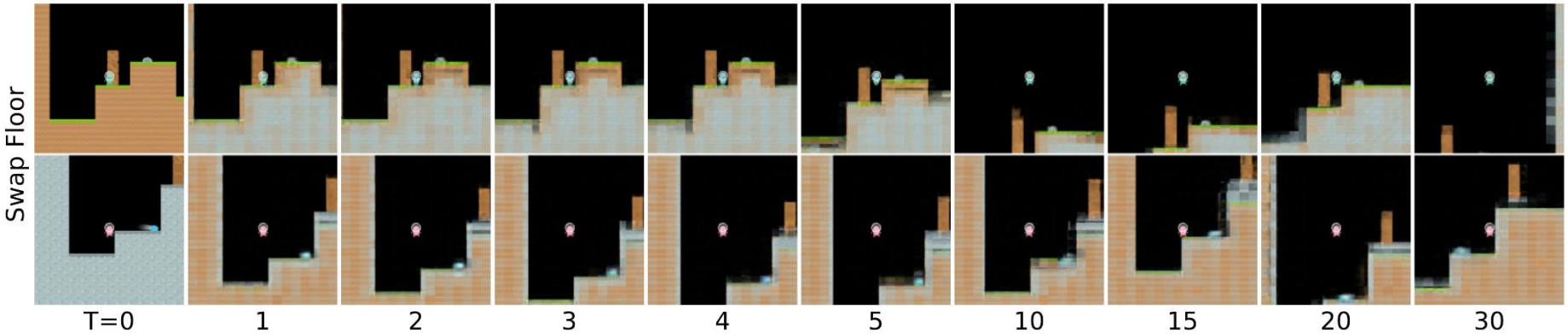
Results – dynamics model



Dyn-O generates more accurate rollouts than baselines (see quantitative results in the paper).

Wang, Z., K. Wang, L. Zhao, P. Stone, J. Bian. “Dyn-O”, NeurIPS 2025.

Results – static-dynamic disentanglement



Dyn-O synthesizes novel experiences for agent learning by shuffling static features.

Wang, Z., K. Wang, L. Zhao, P. Stone, J. Bian. “Dyn-O”, NeurIPS 2025.

Thank you!

Dyn-O: Building Structured World Models with Object-Centric Representations

Zizhao Wang, Kaixin Wang, Li Zhao, Peter Stone, Jiang Bian





Microsoft Research | LARG Learning Agents Research Group, The University of Texas at Austin | Sony AI