Transformer-based Imagination with Slot Attention

Yosuke Nishimoto[1] Takashi Matsubara[2]
[1]Osaka University [2]Hokkaido University

1 Introduction

World Models for Reinforcement Learning

- World model is the simulator of the world.
- RL agents train a world model and optimize their policies within an "imagined" environment generated by the world model.
- It remains challenging for world models to effectively replicate environments comprising multiple objects and their interactions.
- **RL agents with Transformer-based world models** (e.g., TWM [1])
  - ✘ perceive the world as a monolithic entity.
  - ✔ learn dynamics.
  - ✔ learn policies.

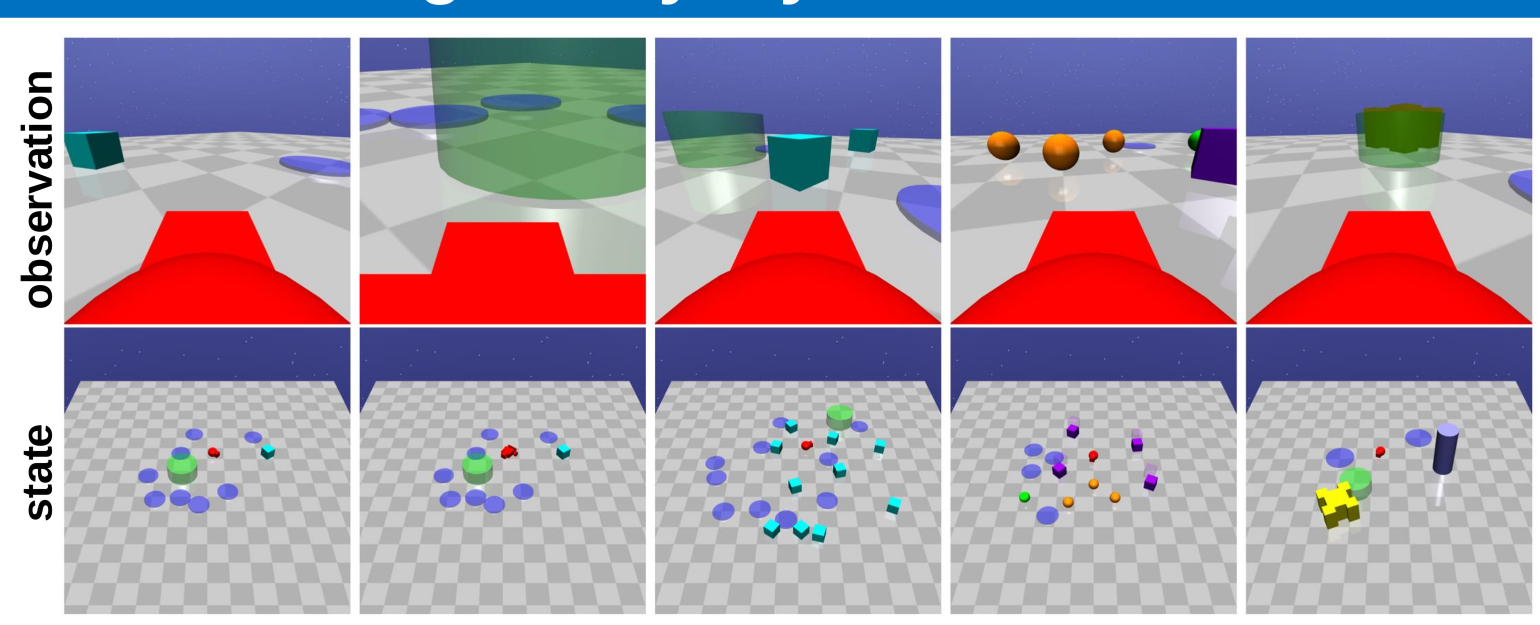
Object-Centric Representation Learning

- Object-centric representation learning is a method that accurately represents a visual scene by segmenting it into multiple entities and extracting individual representations for each one.
- **Video prediction methods** (e.g., SlotFormer [2])
  - ✔ perceive the world by decomposing it into objects.
  - ✔ learn dynamics.
  - ✘ do not learn policies.
- **RL methods** (e.g., OCRL [3], EIT [4])
  - ✔ perceive the world by decomposing it into objects.
  - ✘ do not learn dynamics.
  - ✔ learn policies.

Our Work

- **Transformer-based Imagination with Slot Attention (TISA),** an RL agent that integrates a world model, policy function, and value function, all based on Transformer architecture for object-centric representations.
  - ✔ perceives the world by decomposing it into objects.
  - ✔ learns dynamics.
  - ✔ learns policy.

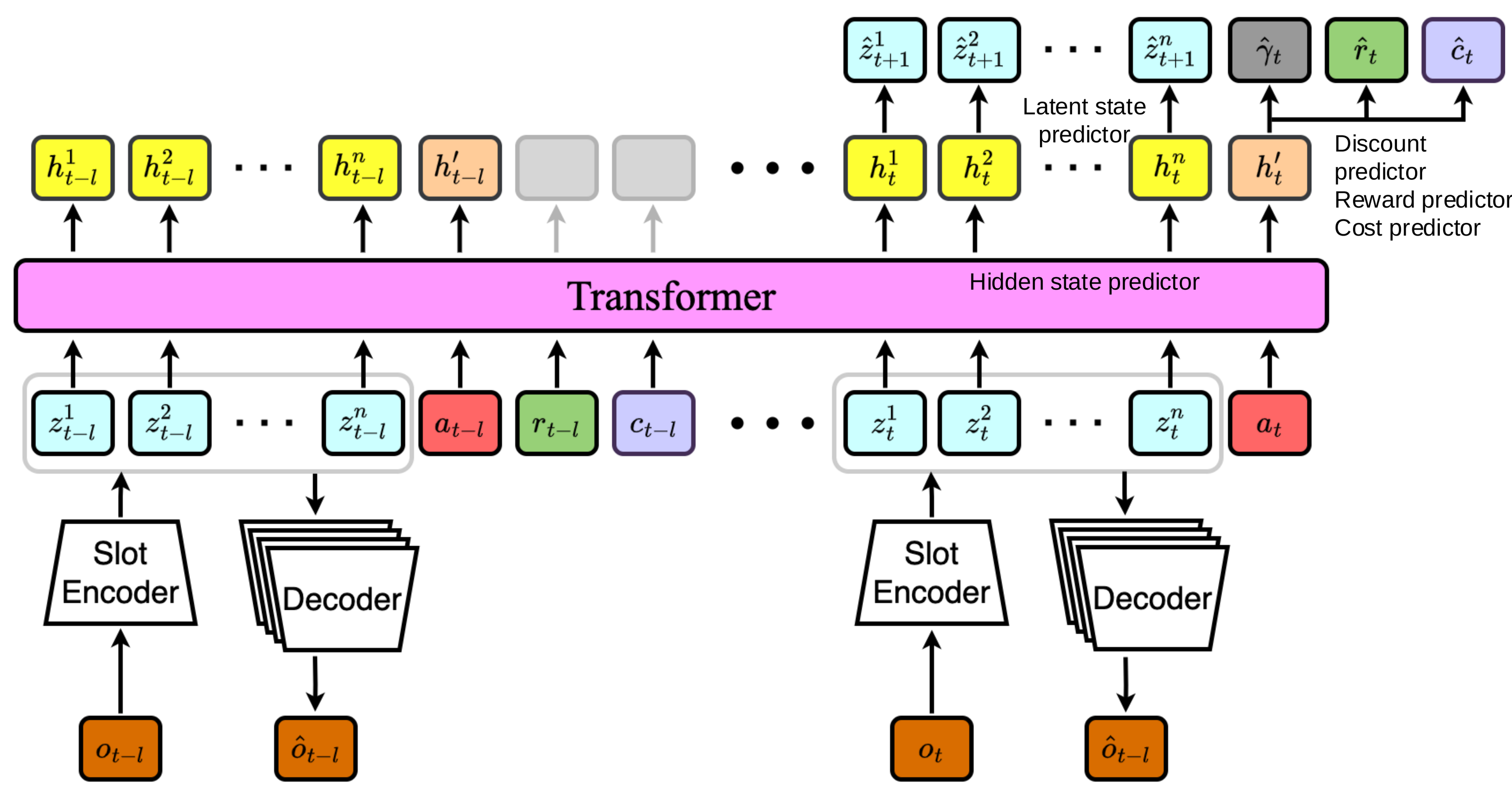
2 Introducing Safety-Gym Benchmark



PointGoal1 CarGoal1 PointGoal2 PointButton1 PointPush1

- The objective for the agent is to navigate to green goal areas while avoiding collisions with surrounding objects.
- The agent receives a reward upon reaching goals and incurs a cost penalty in instances of collision with surrounding objects.
- In PointPush1 task, the yellow box should be pushed into the goal.

3 TISA's World Model



Slot-based AutoEncoder Model

Slot Encoder: $(z^1, \dots, z^n)_t \sim p_\phi((z^1, \dots, z^n)_t | o_t)$,
Decoder: $\hat{o}_t \sim p_\phi(\hat{o}_t | (z^1, \dots, z^n)_t)$.

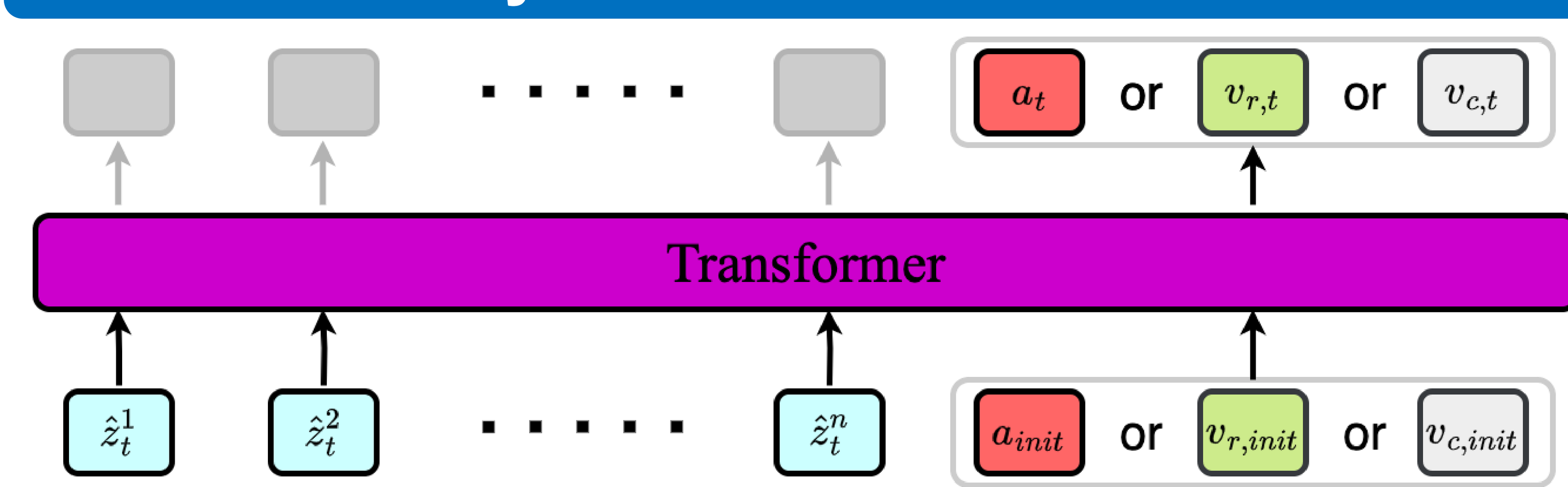
- Using Slot Attention [5], the slot encoder extracts object-centric latent states $z_t^1, \dots, z_t^n$ from an observation o_t.

Transformer-based Dynamics Model

Hidden state predictor: $(h^1, \dots, h^n)_t, h' = f_\psi((z^1, \dots, z^n)_{t-l:t}, a_{t-l:t}, r_{t-l:t-1}, c_{t-l:t-1})$,
Latent state predictor: $\hat{z}_{t+1}^k \sim p_\psi(\hat{z}_{t+1}^k | h_t^k)$, for $k = 1, \dots n$,
Reward predictor: $\hat{r}_t \sim p_\psi(\hat{r}_t | h'_t)$,
Cost predictor: $\hat{c}_t \sim p_\psi(\hat{c}_t | h'_t)$,
Discount predictor: $\hat{\gamma}_t \sim p_\psi(\hat{\gamma}_t | h'_t)$,

- The Transformer-based dynamics model predicts the future latent states $\hat{z}_{t+1}^1, \dots, \hat{z}_{t+1}^n$, along with the reward $\hat{r}_t$, cost $\hat{c}_t$ and discount factor $\hat{\gamma}_t$ for the current step, from latent states $(z^1, \dots, z^n)_{t-l:t}$ and actions $a_{t-l:t}$ from steps $t-l$ to t, as well as the past rewards $r_{t-l:t-1}$ and costs $c_{t-l:t-1}$ from steps $t-l$ to $t-1$.

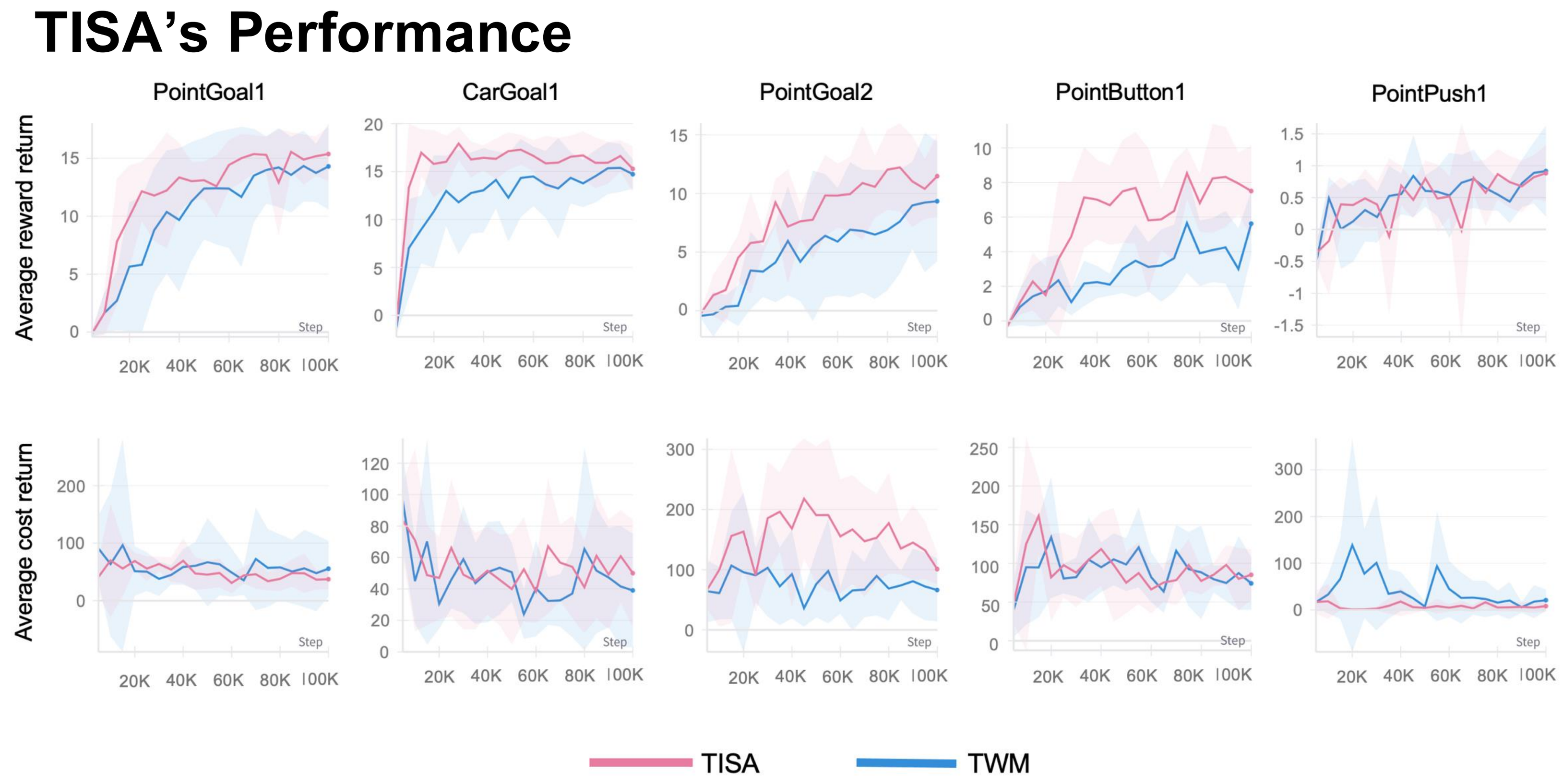
4 TISA's Policy and Value Functions



- We adopted the actor-critic method utilizing the Augmented Lagrangian to build a safe policy [6] .
- The policy function $\pi_\theta(a_t | \hat{z}_t^1, \dots, \hat{z}_t^n)$, the reward function $v_{\zeta_r}(\hat{z}_t^1, \dots, \hat{z}_t^n)$ and the cost value function $v_{\zeta_c}(\hat{z}_t^1, \dots, \hat{z}_t^n)$ are implemented using Transformers.
- The initial values a_{init}, $v_{r,init}$, $v_{c,init}$ are sampled from learnable normal distributions.

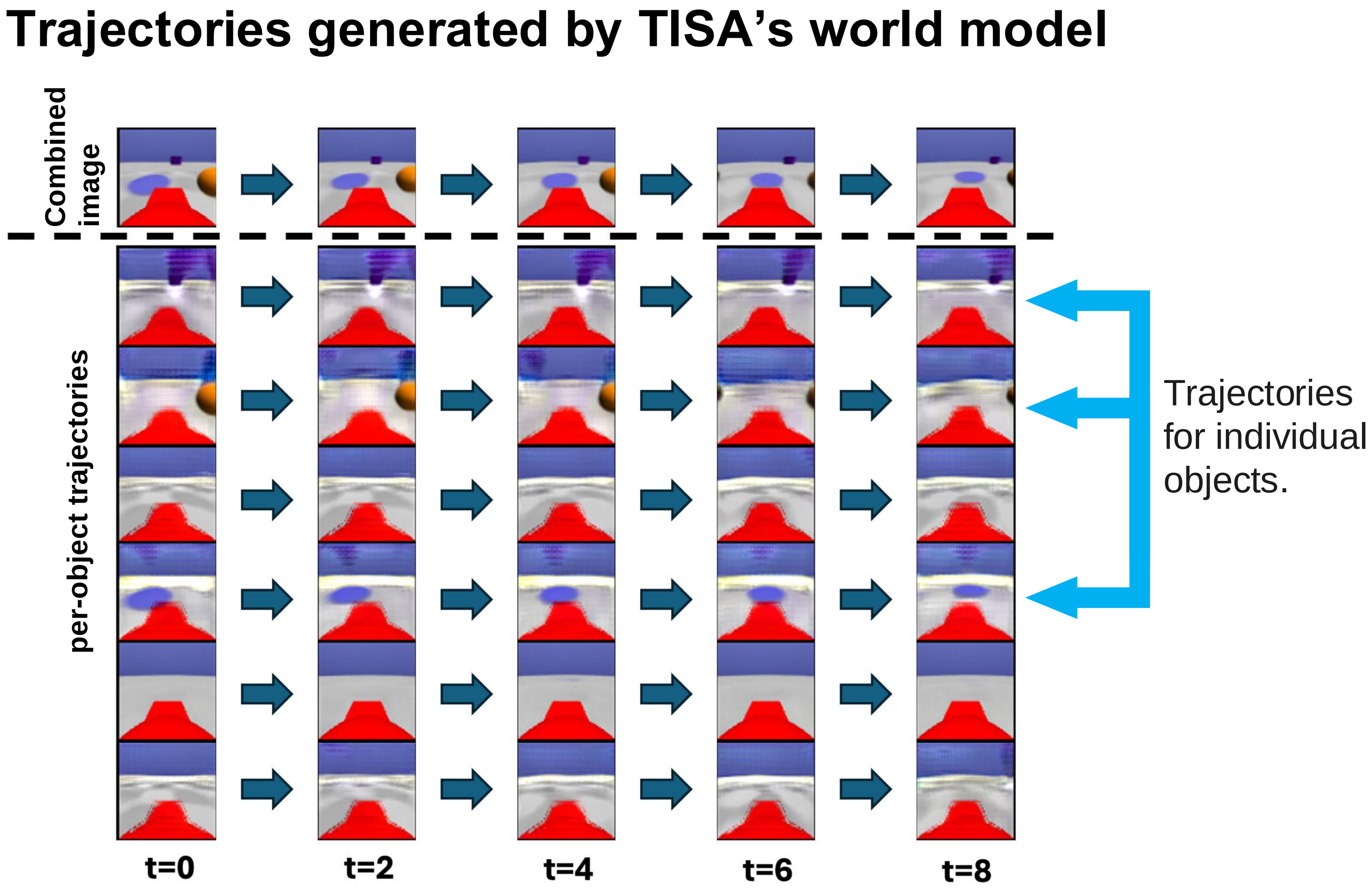
5 Experiments

TISA's Performance



- In three out of the five environments– PointGoal1, CarGoal1 and PointButton1–TISA has achieved the better rewards with the costs at the same levels as TWM.

Trajectories generated by TISA's world model



References

[1] Jan Robine, et al. Transformer-based world models are happy with 100k interactions. In International Conference on Learning Representations, 2023.
[2] Ziyi Wu, et al. Slotformer: Unsupervised visual dynamics simulation with object-centric models. In International Conference on Learning Representations, 166 2023.
[3] Jaesik Yoon, et al. An investigation into pre-training object-centric representations for reinforcement learning. In International Conference on Machine Learning, 2023.
[4] Dan Haramati, et al. Entity-centric reinforcement learning for object manipulation from pixels. In International Conference on Learning Representations, 2024.
[5] Francesco Locatello, et al. Object-centric learning with slot attention. In Advances in Neural Information Processing Systems, 2020.
[6] Yarden As, et al. Constrained policy optimization via bayesian world models. In International Conference on Learning Representations, 2022.