Object-Centric EfficientZeroV2

Rodion R. Vakhitov
MIPT, Dolgoprudny
rodionvahitoff@yandex.ru

Leonid A. Ugadiarov
AIRI, Moscow
MIPT, Dolgoprudny
ulaelfray@gmail.com

Aleksandr I. Panov
AIRI, Moscow
MIPT, Dolgoprudny
panov.ai@mipt.ru

Abstract

In this paper, we introduce Object-Centric EfficientZeroV2, a novel reinforcement learning (RL) algorithm that leverages the power of object-level representations to model dynamic environments more effectively. Unlike traditional approaches that process the world as a single undifferentiated input, our method employs Graph Neural Networks (GNNs) to capture intricate interactions among multiple objects. These objects, which can be manipulated and interact with each other, serve as the foundation for our model's understanding of the environment. We trained the algorithm in a complex setting teeming with diverse, interactive objects, demonstrating its ability to learn and predict object dynamics with increased efficiency. Our results highlight the benefits of incorporating an object-centric bias in reinforcement learning, paving the way for more interpretable and effective world models in model-based RL.

1 Introduction

Research in cognitive neuroscience and psychology reveals that our minds inherently segment incoming sensory data into discrete, meaningful entities, specifically objects, a strategy that not only streamlines how we perceive the world, but also enhances our ability to generalize and adapt Ferraro et al. [2023a], Hawkins et al. [2017]. In contrast to conventional methods that simply encode entire scenes Hafner et al. [2019], Mnih et al. [2015], this paradigm emphasizes uncovering the intricate interactions among individual elements. Recent advances in unsupervised learning have enabled the development of object-centric representations Jiang et al. [2023] and Chakravarthy et al. [2023] allowing models to automatically discover and disentangle objects from raw sensory input. Similar methods introduce an inherent bias toward objectness, resulting in representations that are both more interpretable and generalizable. Embedding such priors into representation learning not only reflects cognitive principles, but also enhances performance across various downstream tasks.

Interest in model-based reinforcement learning (MBRL) approaches Moerland et al. [2020] has been steadily growing, driven by their ability to improve sample efficiency and generalization in reinforcement learning. Unlike model-free approaches that learn policies directly from experience, MBRL agents construct predictive models of state transitions and reward functions through environment interactions. These world models enable agents to simulate trajectories and perform multi-step planning without costly environment queries, potentially achieving superior data efficiency when the model achieves sufficient predictive accuracy. However, the effectiveness of MBRL critically depends on the quality of these learned models - particularly their ability to capture the underlying structure of complex environments. This dependence motivates our object-centric approach, as representing environments through their constituent objects and interactions may provide more accurate and generalizable dynamics modeling compared to unstructured representations Yoon et al. [2023]. Factorized world models, where environments are decomposed into objects and their interactions, provide a natural framework for MBRL Diuk et al. [2008]. Traditional monolithic world models often struggle with generalization due to their tendency to learn dense, entangled represen-

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

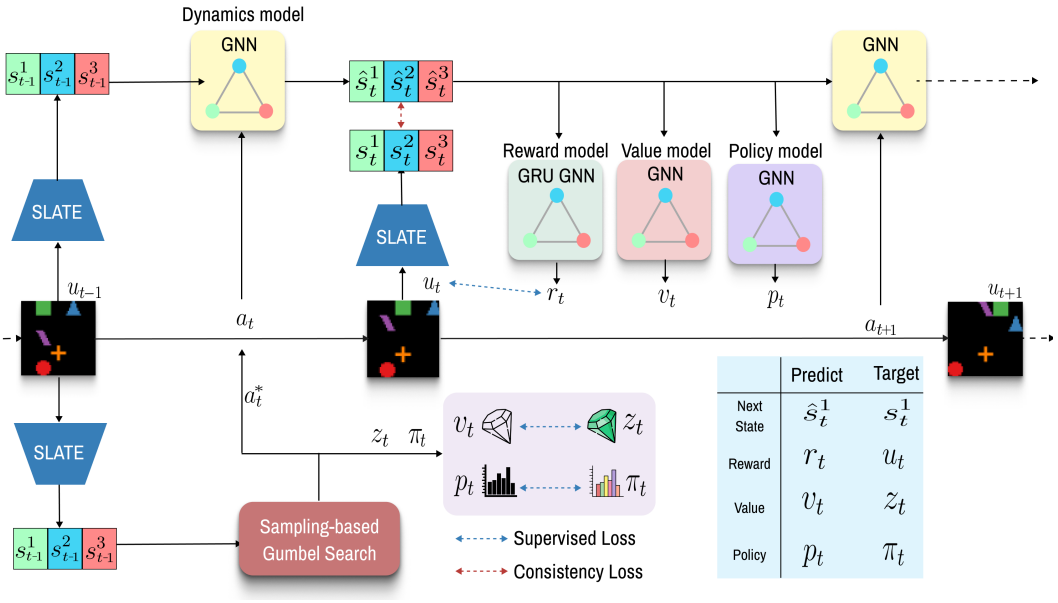

Figure 1: How OCEZ-V2 trains its model. The SLATE extracts object slots from observations. Separate GNN models predict next-state slots, rewards, and policy/value outputs. Gumbel search generates target policy π_t and value z_t.

tations. In contrast, object-centric models promote compositionality, allowing learned dynamics to be transferred across different tasks and environments. Incorporating object-centric world models into model-based RL not only improves efficiency, but also aligns with the principles of human cognition.

This object-centric perspective is particularly advantageous in robotics and MBRL, where structured representations of the world can significantly enhance both perception and decision-making. In robotics, an agent must efficiently parse its environment, distinguishing objects and their properties to facilitate interaction. Object-centric representations enable robots to generalize across different contexts, improving their ability to manipulate novel objects, adapt to unseen scenarios, and reason about the consequences of their actions Devin et al. [2018]. The ability to reason about individual entities and their interactions mirrors how humans infer causal relationships and predict outcomes. By embedding such structured representations into robotic systems, we can enhance their ability to perform complex tasks with greater autonomy and robustness. This structured decomposition can be particularly beneficial in real-world applications, such as robotic manipulation, autonomous driving, and embodied AI, where agents must operate in complex, dynamic environments.

This paper introduces OCEfficientZeroV2, an object-centric reinforcement learning algorithm that integrates a structured world model with online policy learning. The world model utilizes a pre-trained, frozen object-centric encoder based on the SLATE Singh et al. [2022a] architecture, which employs Slot Attention Locatello et al. [2020] to extract sets of object representations from high-dimensional visual inputs. A GNN aggregates information across these object representations, similar to C-SWMs Kipf et al. [2020], to predict environmental rewards, state values, dynamics and policy. Evaluations across various object-centric tasks demonstrate that this structured approach outperforms both monolithic and existing object-centric methods, highlighting the advantages of integrating object-centric representations.

2 Related Work

Object-Centric Representation Learning In recent years, object-centric representation learning methods have gained significant traction, aiming to extract and track individual entities from raw visual input without supervision. A central component in many of these approaches is Slot Attention, which employs a normalized cross-attention mechanism to iteratively bind slots to different regions

in the input. Building on this foundation, SAVi Kipf et al. [2022] and SAVi++ Elsayed et al. [2022] introduce temporal consistency by predicting optical flow and depth, respectively. SLATE Singh et al. [2022a] combines Slot Attention with a VQ-VAE and Transformer decoder for image reconstruction, while STEVE Singh et al. [2022b] scales this architecture to video data. More recently, DINOSAUR Seitzer et al. [2023] leverages slot-level features as training targets, shifting towards a distillation-based learning paradigm that enhances abstraction and robustness.

Model-based Reinforcement Learning MBRL has demonstrated significant improvements in sample efficiency and planning capabilities by learning internal models of environment dynamics. A prominent example is MuZero Schrittwieser et al. [2020], which combines learned dynamics and Monte Carlo Tree Search (MCTS) without requiring access to the true environment model. It has achieved strong performance in a wide range of domains, including board games and Atari. Building on this, EfficientZero Ye et al. [2021] (EZ) and EfficientZeroV2 Wang et al. [2024] (EZ-V2) extend MuZeros planning framework to pixel-based environments, introducing architectural improvements to boost data efficiency in visual tasks. The Dreamer series leverages latent dynamics models trained end-to-end with actor-critic objectives. DreamerV3 Hafner et al. [2023], in particular, achieves robust performance through categorical latents and online learning from high-dimensional observations. Similarly, TD-MPC Hansen et al. [2022] and TD-MPC2 Hansen et al. [2024] integrate model-predictive control with learned dynamics, enabling trajectory optimization without explicit reconstruction losses. Additionally, STORM Zhang et al. [2023] incorporates stochasticity into a Transformer-based world model, enhancing robustness to modeling errors in visually complex environments.

Object-Centric Representations and MBRL Recent work has explored the integration of object-centric representations into MBRL to improve sample efficiency and robustness in complex environments. COBRA Watters et al. [2019] COBRA introduces a modular approach using intrinsically motivated exploration to build object-based models, enabling efficient task learning through model-based search in continuous control environments. FOCUS Ferraro et al. [2023b] introduces an object-centric world model with a novel exploration bonus, enhancing manipulation tasks by explicitly modeling object interactions. Similarly, OC-STORM Zhang et al. [2025] combines segmentation-based object features with a learned dynamics model, enabling better handling of visually complex scenes in games. SOLD Mosbach et al. [2024] advances this direction by employing Slot-Attention for unsupervised object-centric dynamics learning, outperforming state-of-the-art MBRL methods in robotic tasks requiring relational reasoning. These approaches highlight the growing recognition of structured representations as a key component for scalable and interpretable MBRL.

3 Background

3.1 Reinforcement Learning

Reinforcement Learning can be formulated as a Markov Decision Process (MDP) Bellman [1957]. An MDP in this context is formalized as a tuple $(\mathcal{S}, \mathcal{A}, T, R, \gamma)$, where $s \in \mathcal{S}$ represents states, $a \in \mathcal{A}$ denotes actions, $T\colon \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ is the transition function, $R\colon \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the reward function associated with a particular task, and γ is the discount factor. The goal in RL is to find an optimal policy π^* that maximizes the expected sum of discounted rewards over time, formally expressed as $\max_\pi \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)\right]$, where a_t is an action taken according to the policy π at state s_t.

3.2 EfficientZeroV2

Our model builds on top of the EfficientZeroV2, a model-based reinforcement learning algorithm that significantly improves upon its predecessor, EfficientZero, by extending its capabilities to continuous action spaces, enhancing planning efficiency, and introducing more accurate value estimation methods. Like EfficientZero, EZ-V2 is based on the MuZero-style framework and learns a latent model of the environment, which is used both for planning and for learning policy and value functions. In addition to architectural refinements, EZ-V2 incorporates more robust training mechanisms, including improved value estimation and temporal consistency objectives.

Table 1: Comparison of component models and implementations in EZ-V2 and OCEZ-V2

Component	Model	Implementation
	EZ-V2	
Dynamics model	$\hat{s}_{t+1} \sim p_\phi(\hat{s}_{t+1} \mid s_t, a_t)$	CNN and MLP
Encoder	$s_t \sim p_\theta(s_t \mid o_t)$	CNN and MLP
Reward predictor	$\hat{r}_t \sim p_\psi(\hat{r}_t \mid s_t)$	LSTM
Value predictor	$\hat{v}_t \sim p_\xi(\hat{v}_t \mid s_t)$	CNN and MLP
Policy model	$a_t \sim p_\omega(\hat{a}_t \mid s_t)$	CNN and MLP
	OCEZ-V2	
Dynamics model	$\hat{s}^i_{t+1} \sim p_\phi(\hat{s}^i_{t+1} \mid \hat{s}^i_t, a_t)$	GNN
Encoder	$\bar{s}_t = \text{SLATE}(o_t)$	SLATE (frozen)
Reward predictor	$\hat{r}_t \sim p_\psi(\hat{r}_t \mid \bar{s}_t)$	GRU-GNN
Value predictor	$\hat{r}_t \sim p_\xi(\hat{r}_t \mid \bar{s}_t)$	GNN and MLP
Policy model	$a_t \sim p_\omega(\hat{a}_t \mid \bar{s}_t)$	GNN and MLP

All the main components of EZ-V2 and OCEZ-V2 are presented in Table 1. Among these components, o_t represents the current observation, s_t is the current latent state. The representation function learns a compact state representation of the input o_t. The dynamic function predicts the next state $\hat{s}_{t+1}$ based on the s_t current state and the action taken a_t. The policy model outputs the current policy p_t, the value predictor provides the value estimation $\hat{v}_t$ at the current state and reward predictor estimates $\hat{r}_t$.

Action Planning with Sampling-Based Gumbel Search. In contrast to traditional MuZero or EfficientZero, which rely on MCTS, EZ-V2 uses a novel sampling-based Gumbel Search, like in Danihelka et al. [2022] that allows effective planning in both discrete and continuous action spaces. It is the most critical and impactful innovation in EZ-V2.

The action candidates are sampled using the Gumbel-Top-k trick, which allows sampling without replacement from a categorical distribution. For each action a, a Gumbel noise $g(a)$ is added to its logit:

$$A_k = \arg \max_{a \notin A_1, \ldots, A_{k-1}} \left(g(a) + \text{logits}(a) \right), \quad g(a) \sim \text{Gumbel}(0) \tag{1}$$

Among the sampled actions, the search algorithm performs iterative elimination (Sequential Halving) to select the action a^*_S with the highest estimated Q-value. This ensures that, on average, planning improves the policy's performance relative to the current distribution.

Also EZ-V2 includes several refinements that improve sample efficiency and stability:

- A Gaussian policy head enables effective sampling in continuous action spaces;
- Search-Based Value Estimation provides more accurate value targets by leveraging simulated rollouts during planning;
- Temporal consistency loss encourages stable latent dynamics across predicted future states;
- A unified loss function supervises reward, value, policy, and state predictions.

All components are trained jointly using experience replay and model-generated planning targets.

4 Object-Centric EfficientZeroV2

4.1 Objects Extractor

As an objects extractor we use a SLATE, a recent unsupervised object-centric model. The SLATE incorporates a dVAE Van Den Oord et al. [2017] for internal feature extraction, a GPT-like transformer Ramesh et al. [2021] for decoding, and a slot-attention module to group features associated with the same object. The SLATE takes an image-based observation o_t as input and extracts object

representations as an unordered set of low-dimensional vectors $\bar{s}_t = \{s_t^1, \cdots, s_t^K\}$, where K is a non-learned parameter of the SLATE which equals to the maximum possible number of objects in the image. The SLATE can not guarantee the fixed order of object representations due to the stochastic nature of the slot-attention module. To enforce the order of object representation during the episode, we preinitialize the slots of the SLATE model for the next step $t+1$ with the current object representations $\bar{s}_t$. We pretrain the SLATE on the dataset of observations collected in the environment by a uniform random policy and freeze it. ctor model, the actor, and the critic follow the architecture described above.

4.2 Dynamics Model

Figure 1 illustrates the architectures of OCEZ-V2.

$$y_h^i = \texttt{node}_h(s_t^i, a_t, \sum_{j \neq i} \texttt{edge}_h(s_t^i, s_t^j, a_t)). \tag{2}$$

$$\bar{s}_{t+1} = \mathrm{GNN}_h(\bar{s}_t, a_t). \tag{3}$$

This procedure allows us to model pairwise interactions between objects while being invariant to the order in which objects are presented.

4.3 Policy model, Reward Predictor and Value Predictor

$$\overline{y}_{r/v/p} = \mathrm{GNN}_{r/v/p}(\bar{s}_t). \tag{4}$$

$$r_t/v_t/p_t = \mathrm{MLP}_{\mathrm{r/v/p}}(y^1_{\mathrm{r/v/p}} \oplus \cdots \oplus y^K_{\mathrm{r/v/p}}) \tag{5}$$

5 Environments

The performance of the OCEZ-V2 algorithm was tested in the 2D compositional environment Shapes2D Kipf et al. [2020] on the Navigation and PushingNoAgent task.



Figure 2: Examples of observations and attention maps produced by the SLATE model in the Navigation 5x5 task.

5.1 Shapes2d

Navigation Task Shapes2D is a grid-based environment where objects are represented as simple two-dimensional shapes. The environment operates on a four-connected grid, allowing movement in four directions. One of the objects, shaped like a cross, serves as a stationary target and does not move. All other objects in the environment are movable and controlled by the agent. During each step, the agent can move one of the controlled objects to an adjacent, unoccupied cell. The goal is to move the objects in such a way that they collide with the stationary target. When a movable object successfully collides with the target, it is removed from the grid and the agent receives a reward of $+1$. If an attempted move results in a collision with either another movable object or the boundaries of the grid, the move is canceled and a penalty of -0.1 is applied. Additionally, each time step incurs a small penalty of -0.01 regardless of the outcome. The episode ends when only the target object remains on the grid. The environment is used in two configurations: a 5×5 grid with five total objects, and a 10×10 grid with eight total objects. In the 5×5 setup, the agent controls four movable objects and has access to 16 discrete actions. In the 10×10 version, seven movable objects are controlled, with a corresponding action space of 28 discrete actions.

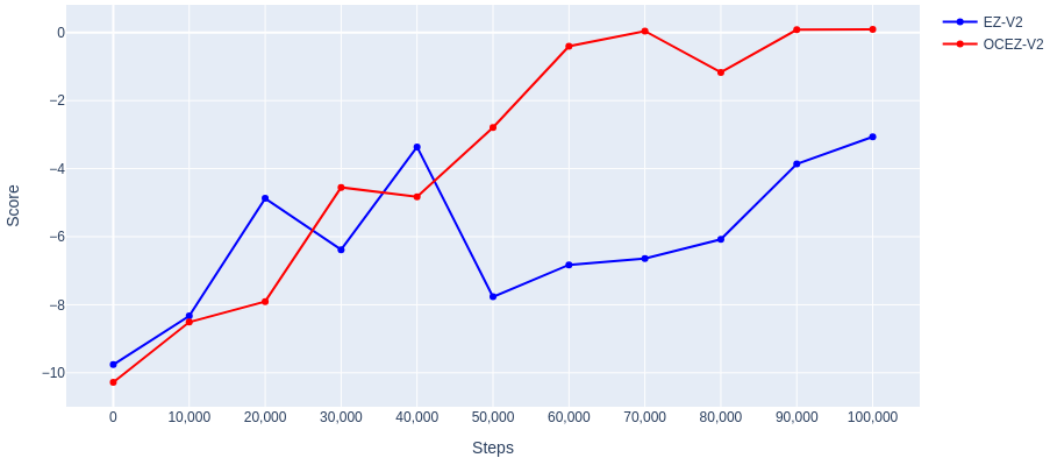

Figure 3: Return averaged over 100 episodes and three seeds for EZ-V2 and OCEZ-V2 on Shapes2D Navigation Task

PushingNoAgent Task In the this task the agent controls all movable objects similarly to the Navigation task, but with modified collision dynamics where collisions between two movable objects cause both objects to move in the direction of motion. The agent's primary objective is to push another movable object into the target while controlling the current object. When successfully pushing an object into the target, that object disappears and the agent receives a +1 reward. The agent incurs a -0.1 penalty when the currently controlled object collides with either the target object or field boundaries, and the same penalty applies when pushing a movable object into the boundaries. Each timestep carries a -0.01 penalty. The episode terminates when only the target object and one movable object remain in the environment. The experimental setup uses a 5×5 grid with five total objects. While the action space mirrors the Navigation 5×5 task with 16 possible actions, the physics are more complex. Developing effective pushing strategies becomes essential for the agent, as this represents the sole method to obtain positive rewards in this challenging environment.

6 Experiments

6.1 Shapes2d

For experiments in the Shapes2D environment, we use SLATE as an object-centric extractor. We utilize a single SLATE model for the Navigation5x5 task, as it provides a consistent observation space. Appendix A provides detailed information regarding the hyperparameters of the SLATE model. The SLATE model extracts high-quality object-centric representations in the Shapes2D environment, achieving a foreground ARI score of 95% for the Navigation5x5 task. The high quality of the encoder allows us to evaluate the impact of the object-oriented decomposition of the world model on the performance of the MBRL algorithm. Our experiments demonstrate that OCEZ-V2 performs better than EZV2, highlighting the benefits of object-centric representations in model-based reinforcement learning

7 Limitations

Despite the advances in modern unsupervised object-centric learning, most state-of-the-art approaches still face significant challenges when it comes to decomposing complex, realistic scenes into distinct objects. We consider this limitation a major barrier to deploying our algorithm in arbitrary, real-world environments. Additionally, since we represent the environments state as a fully connected graph of object states, which is processed by graph neural networks, the computational complexity of Object-Centric EfficientZeroV2 becomes quadratic with respect to the number of slots produced by the object-centric extractor. Both of these factors can significantly slow down training

in practical, object-rich scenariosparticularly in real-world robotics settings, where efficiency and scalability are crucial for learning from interaction.

8 Conclusion

In this work, we introduced Object-Centric EfficientZeroV2, a model-based reinforcement learning algorithm that integrates object-centric representations with a structured world model built upon a pretrained SLATE encoder and Graph Neural Networks. Through experiments in a multi-object 2D environment (Shapes2D), we demonstrated that OCEZ-V2 effectively learns policies and outperforms both monolithic and existing object-centric baselines in tasks requiring relational reasoning. Our results highlight the advantages of incorporating inductive biases for objectness and structured interaction modeling into world models. In particular, we show that GNN-based dynamics and policy models can leverage the compositional nature of object representations to improve learning efficiency and generalization. As future work, we plan to explore the joint training of the object-centric encoder and world model components, as well as the application of OCEZ-V2 in more realistic and high-dimensional environments where object discovery remains a major challenge.

References

Stefano Ferraro, Toon Van de Maele, Tim Verbelen, and Bart Dhoedt. Symmetry and complexity in object-centric deep active inference models. *Interface Focus*, 13(3):20220077, 2023a.

Jeff Hawkins, Subutai Ahmad, and Yuwei Cui. A theory of how columns in the neocortex enable learning the structure of the world. *Frontiers in neural circuits*, 11:295079, 2017.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *International conference on machine learning*, pages 2555–2565. PMLR, 2019.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

Jindong Jiang, Fei Deng, Gautam Singh, and Sungjin Ahn. Object-centric slot diffusion. *Advances in Neural Information Processing Systems*, 36:8563–8601, 2023.

Ayush Chakravarthy, Trang Nguyen, Anirudh Goyal, Yoshua Bengio, and Michael C Mozer. Spotlight attention: Robust object-centric learning with a spatial locality prior. *arXiv preprint arXiv:2305.19550*, 2023.

Thomas M. Moerland, Joost Broekens, and Catholijn M. Jonker. Model-based reinforcement learning: A survey. *CoRR*, abs/2006.16712, 2020. URL `https://arxiv.org/abs/2006.16712`.

Jaesik Yoon, Yi-Fu Wu, Heechul Bae, and Sungjin Ahn. An investigation into pre-training object-centric representations for reinforcement learning. *arXiv preprint arXiv:2302.04419*, 2023.

Carlos Diuk, Andre Cohen, and Michael L Littman. An object-oriented representation for efficient reinforcement learning. In *Proceedings of the 25th international conference on Machine learning*, pages 240–247, 2008.

Coline Devin, Pieter Abbeel, Trevor Darrell, and Sergey Levine. Deep object-centric representations for generalizable robot learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7111–7118. IEEE, 2018.

Gautam Singh, Fei Deng, and Sungjin Ahn. Illiterate DALL-e learns to compose. In *International Conference on Learning Representations*, 2022a. URL `https://openreview.net/forum?id=h0OYV0We3oh`.

Francesco Locatello, Dirk Weissenborn, Thomas Unterthiner, Aravindh Mahendran, Georg Heigold, Jakob Uszkoreit, Alexey Dosovitskiy, and Thomas Kipf. Object-centric learning with slot attention. *Advances in neural information processing systems*, 33:11525–11538, 2020.

Thomas Kipf, Elise van der Pol, and Max Welling. Contrastive learning of structured world models. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=H1gax6VtDB.

Thomas Kipf, Gamaleldin Fathy Elsayed, Aravindh Mahendran, Austin Stone, Sara Sabour, Georg Heigold, Rico Jonschkowski, Alexey Dosovitskiy, and Klaus Greff. Conditional object-centric learning from video. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=aD7uesX1GF_.

Gamaleldin Elsayed, Aravindh Mahendran, Sjoerd Van Steenkiste, Klaus Greff, Michael C Mozer, and Thomas Kipf. Savi++: Towards end-to-end object-centric learning from real-world videos. *Advances in Neural Information Processing Systems*, 35:28940–28954, 2022.

Gautam Singh, Yi-Fu Wu, and Sungjin Ahn. Simple unsupervised object-centric learning for complex and naturalistic videos. *Advances in Neural Information Processing Systems*, 35:18181–18196, 2022b.

Maximilian Seitzer, Max Horn, Andrii Zadaianchuk, Dominik Zietlow, Tianjun Xiao, Carl-Johann Simon-Gabriel, Tong He, Zheng Zhang, Bernhard Schölkopf, Thomas Brox, and Francesco Locatello. Bridging the gap to real-world object-centric learning. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=b9tUk-f_aG.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, 2020.

Weirui Ye, Shaohuai Liu, Thanard Kurutach, Pieter Abbeel, and Yang Gao. Mastering atari games with limited data. *Advances in neural information processing systems*, 34:25476–25488, 2021.

Shengjie Wang, Shaohuai Liu, Weirui Ye, Jiacheng You, and Yang Gao. Efficientzero v2: Mastering discrete and continuous control with limited data. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=LHGMXcr6zx.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023.

N Hansen, X Wang, and H Su. Temporal difference learning for model predictive control. In *International Conference on Machine Learning, PMLR*, 2022.

Nicklas Hansen, Hao Su, and Xiaolong Wang. TD-MPC2: Scalable, robust world models for continuous control. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=Oxh5CstDJU.

Weipu Zhang, Gang Wang, Jian Sun, Yetian Yuan, and Gao Huang. Storm: Efficient stochastic transformer based world models for reinforcement learning. *Advances in Neural Information Processing Systems*, 36:27147–27166, 2023.

Nicholas Watters, Loic Matthey, Matko Bosnjak, Christopher P Burgess, and Alexander Lerchner. Cobra: Data-efficient model-based rl through unsupervised object discovery and curiosity-driven exploration. *arXiv preprint arXiv:1905.09275*, 2019.

Stefano Ferraro, Pietro Mazzaglia, Tim Verbelen, and Bart Dhoedt. FOCUS: Object-centric world models for robotic manipulation. In *Intrinsically-Motivated and Open-Ended Learning Workshop @NeurIPS2023*, 2023b. URL https://openreview.net/forum?id=RoQbZRv1zw.

Weipu Zhang, Adam Jelley, Trevor McInroe, and Amos Storkey. Objects matter: object-centric world models improve reinforcement learning in visually complex environments. *arXiv preprint arXiv:2501.16443*, 2025.

Malte Mosbach, Jan Niklas Ewertz, Angel Villar-Corrales, and Sven Behnke. Sold: Reinforcement learning with slot object-centric latent dynamics. *arXiv preprint arXiv:2410.08822*, 2024.

Richard Bellman. A markovian decision process. *Journal of mathematics and mechanics*, pages 679–684, 1957.

Ivo Danihelka, Arthur Guez, Julian Schrittwieser, and David Silver. Policy improvement by planning with gumbel. In *International Conference on Learning Representations*, 2022. URL `https://openreview.net/forum?id=bERaNdoegnO`.

Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017.

Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In *International conference on machine learning*, pages 8821–8831. Pmlr, 2021.

A Implementation Details

B SLATE

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (12 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: **[TODO]**

Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so No is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).

- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.

- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: **[TODO]**

    Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.