Max Planck Institute for Intelligent Systems – Autonomous Learning Group

Curious Exploration via Structured World Models Yields Zero-Shot Object Manipulation

Cansu Sancaktar, Sebastian Blaes, Georg Martius

imprs-is

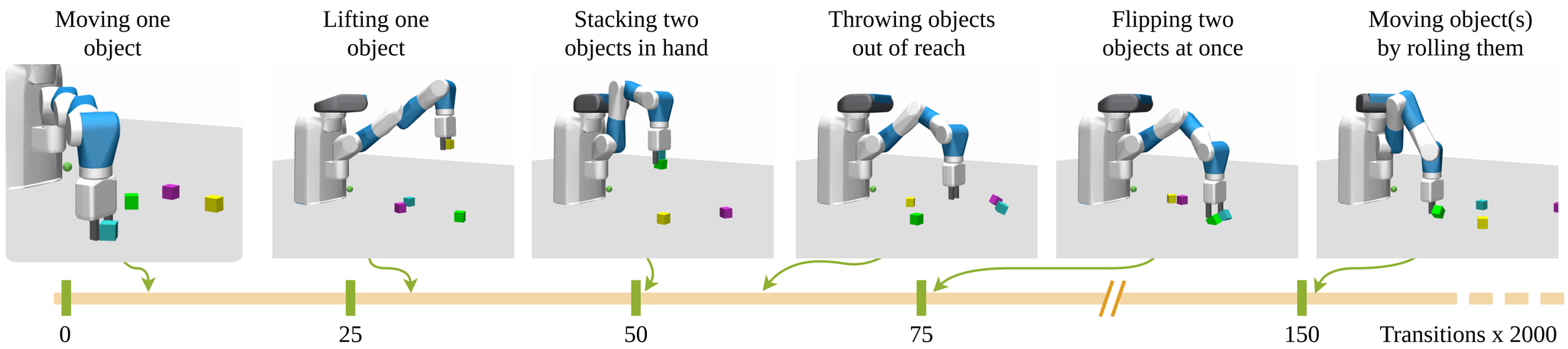
Overview

Problem:

Applying curious free play to object manipulation scenarios in a sample-efficient manner is an on-going challenge in Reinforcement Learning (RL) as the relevant information lies in the sparse agent-object interactions. A **novel stimulus** [1, 3] alone does not necessarily mean that it contains useful or generalizable information to an individual, as it is agnostic to the structure of the environment.

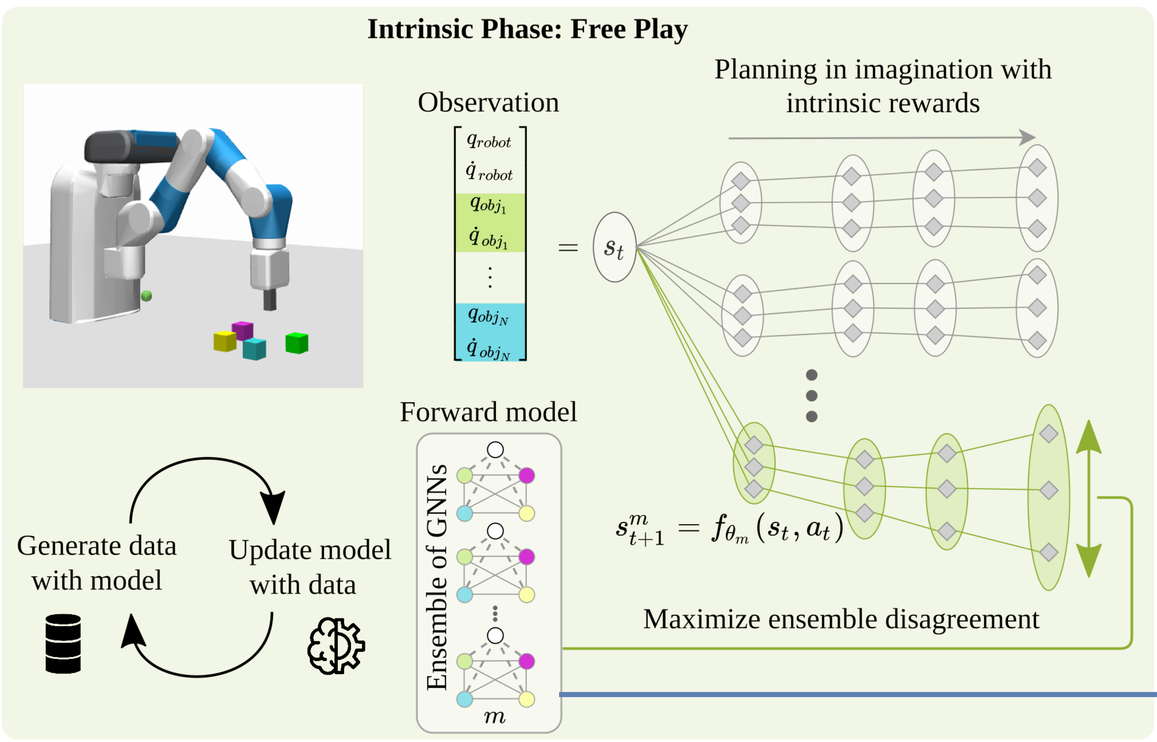
Our contributions:

We propose **CEE-US**: **C**urious **E**xploration using **E**pistemic **U**ncertainty via **S**tructured Models that achieves sample-efficient and interaction-rich exploration in multi-object manipulation environments. We use Graph Neural Networks (GNN) as world model which has object-based and relational inductive biases.



Emergent behaviors observed during free play with CEE-US.

Intrinsic Phase of CEE-US



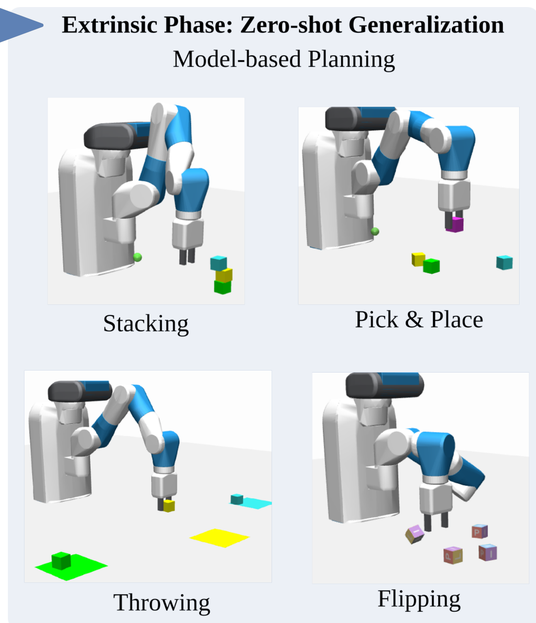
- Train **ensemble of M GNNs** $\{f_{\theta_m} \mid m=1,\dots,M\}$.
- Intrinsic reward signal as **epistemic uncertainty** of the model:

$$R_I(s_t, a_t) = \mathrm{tr}\Big(\mathrm{Cov}(\{\hat{s}^m_{t+1} = f_{\theta_m}(s_t, a_t) \mid m = 1, \dots, M\})\Big).$$

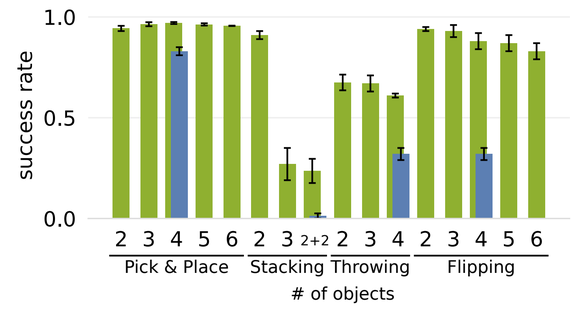
- Active exploration with **online model predictive control** using the impoved CEM (iCEM)[2], utilizing multi-step novelty with longer planning horizons.

Extrinsic Phase of CEE-US

After the intrinsic free play, we use the learned GNNs to solve downstream tasks with model-based planning zero-shot without any additional training.



Thanks to the **combinatorial generalization of GNNs**, we can solve tasks with more or less objects than seen during free play.



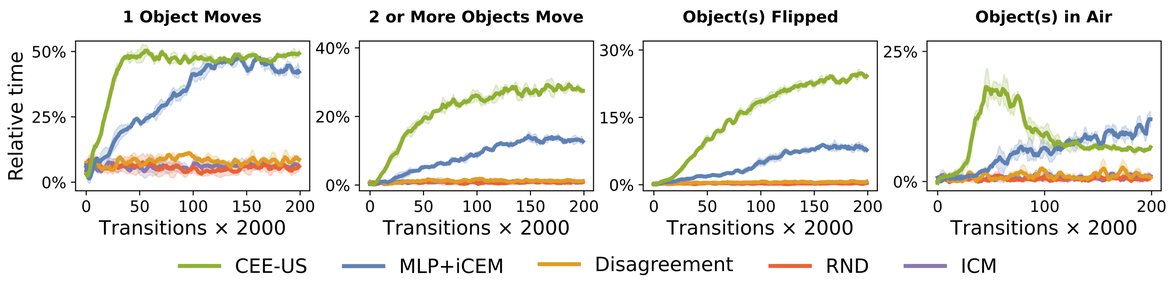
Zero-shot performance on downstream tasks for CEE-US and MLP + iCEM.

Offline Learning of Downstream Tasks

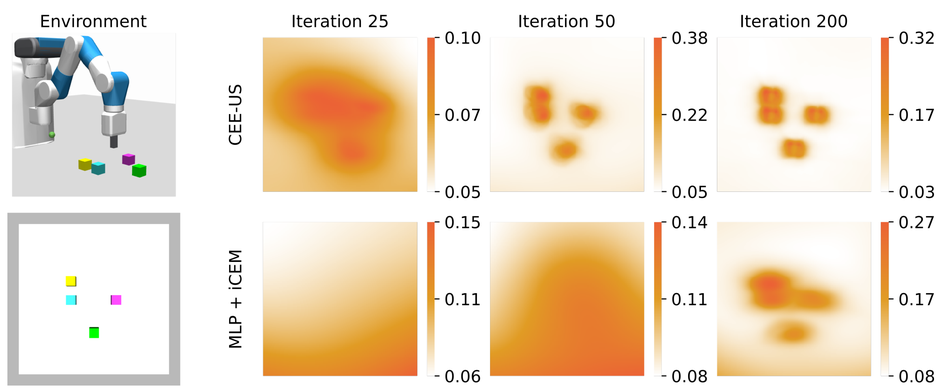
We investigate the quality of the data collected by the different methods in the free-play phase for solving downstream tasks via offline RL, where we extract a policy by repurposing the free-play data.

Task	Disagreement	RND	ICM	MLP + iCEM	**CEE-US**
Reach	0.09 ± 0.01	0.19 ± 0.05	0.2 ± 0.03	0.65 ± 0.09	$\mathbf{0.94 \pm 0.04}$
Pick & Place 1 obj.	0.07 ± 0.0	0.07 ± 0.0	0.07 ± 0.01	0.18 ± 0.06	$\mathbf{0.43 \pm 0.07}$

Interaction-Richness of Free Play



Epistemic Uncertainty Heatmaps



Conclusion

- CEE-US combines learning of GNNs as structured world models with curiosity-driven, planning-based exploration.
- We achieve **sample-efficient and interaction-rich exploration in compositional multi-object environments.**
- In the extrinsic phase: use GNNs learned during free play to **solve downstream tasks zero-shot** with model-based planning.

Paper
arXiv: 2206.11403

Website
cee-us.github.io

Autonomous Learning
al.is.mpg.de

We are hiring Post-docs!
Deep RL
Machine Learning
Robotics
Model-based learning

References

[1] Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the 34th International Conference on Machine Learning*, 2017.

[2] Cristina Pinneri, Shambhuraj Sawant, Sebastian Blaes, Jan Achterhold, Joerg Stueckler, Michal Rolinek, and Georg Martius. Sample-efficient cross-entropy method for real-time planning. In *Proceedings of the 5th Conference on Robot Learning*, 2021.

[3] Ramanan Sekar, Oleh Rybkin, Kostas Daniilidis, Pieter Abbeel, Danijar Hafner, and Deepak Pathak. Planning to explore via self-supervised world models. In *Proceedings of the 37th International Conference on Machine Learning*, 2020.