Dreaming of Many Worlds: Learning Contextual World Models aids Zero-Shot Generalization

universität freiburg

ellis UNIT FREIBURG

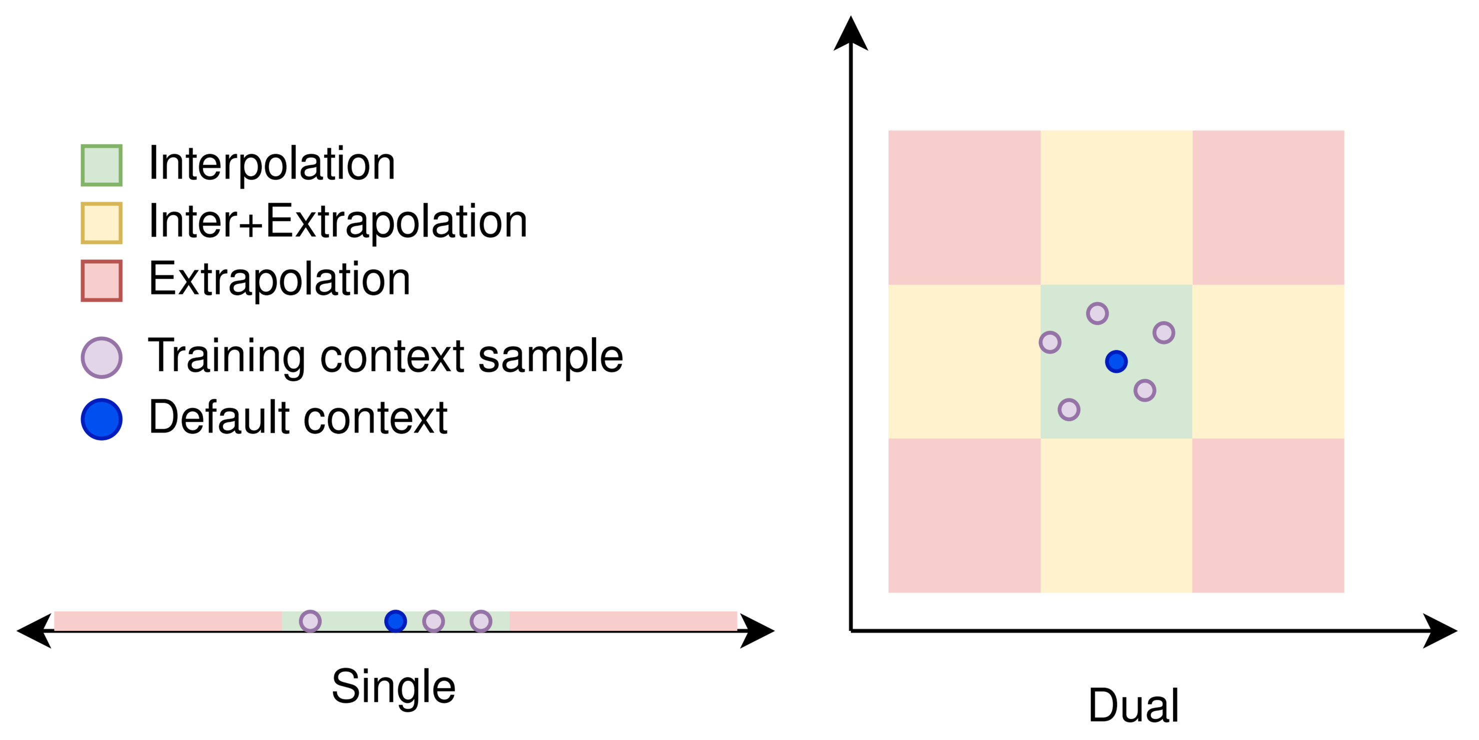
Sai Prasanna*[1], Karim Farid*[1], Raghu Rajan[1], André Biedenkapp[1]
[1]University of Freiburg, *Equal contribution

Paper: https://rlj.cs.umass.edu/2024/papers/Paper167.html

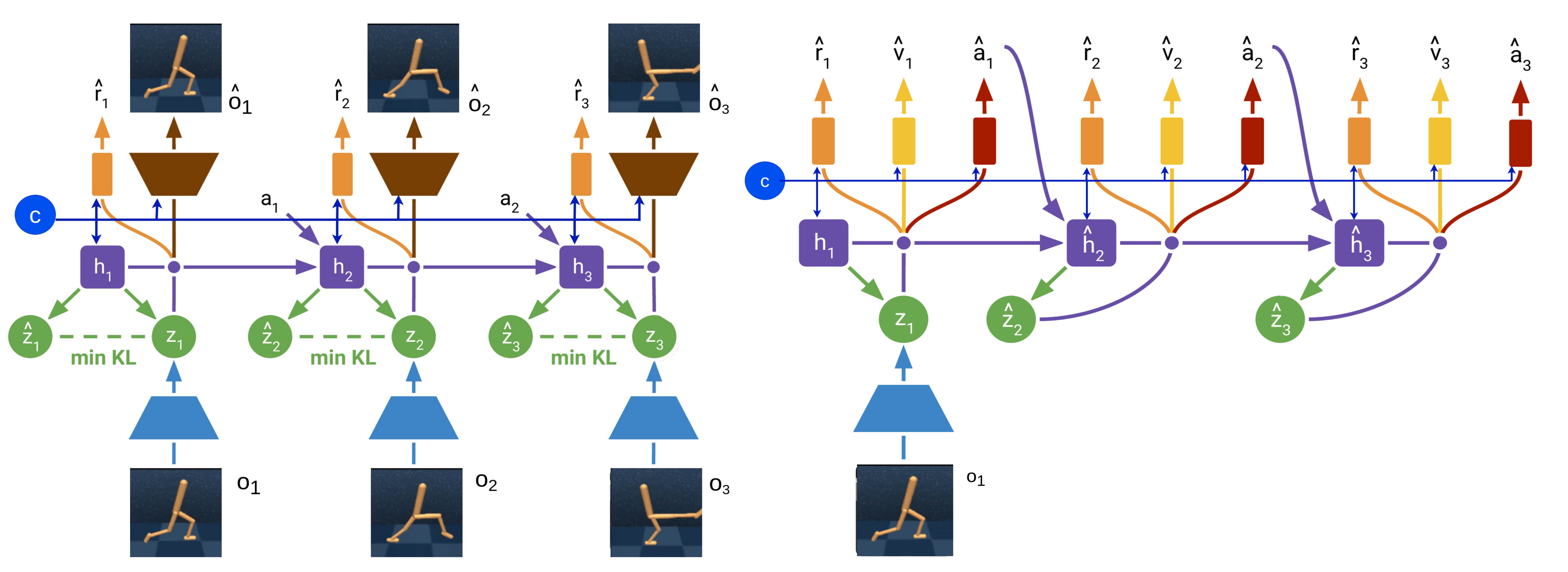
Motivation

- **Context:** parameters of a (PO)MDP that remain **unchanged** during an episode but can vary across episodes;
  - Affects the **dynamics** and **rewards.**
  - In this work, we assume the context is observable.
  - **Examples:** Height of a robot, mass of the load carried, actuator strength, etc.
- **Zero-Shot Generalization (ZSG)** to context distributions without adapting weights.
  - Explicitly conditioning a SOTA MbRL agent, DreamerV3, should improve ZSG / OOD robustness.
- **World models** could afford a promising avenue for ZSG.

Zero-Shot Generalization



Contextual Recurrent State Space Model (cRSSM)



Figures are adapted from Dreamer v2 (Hafner et al. 2020)

Experiments

Incorporating context

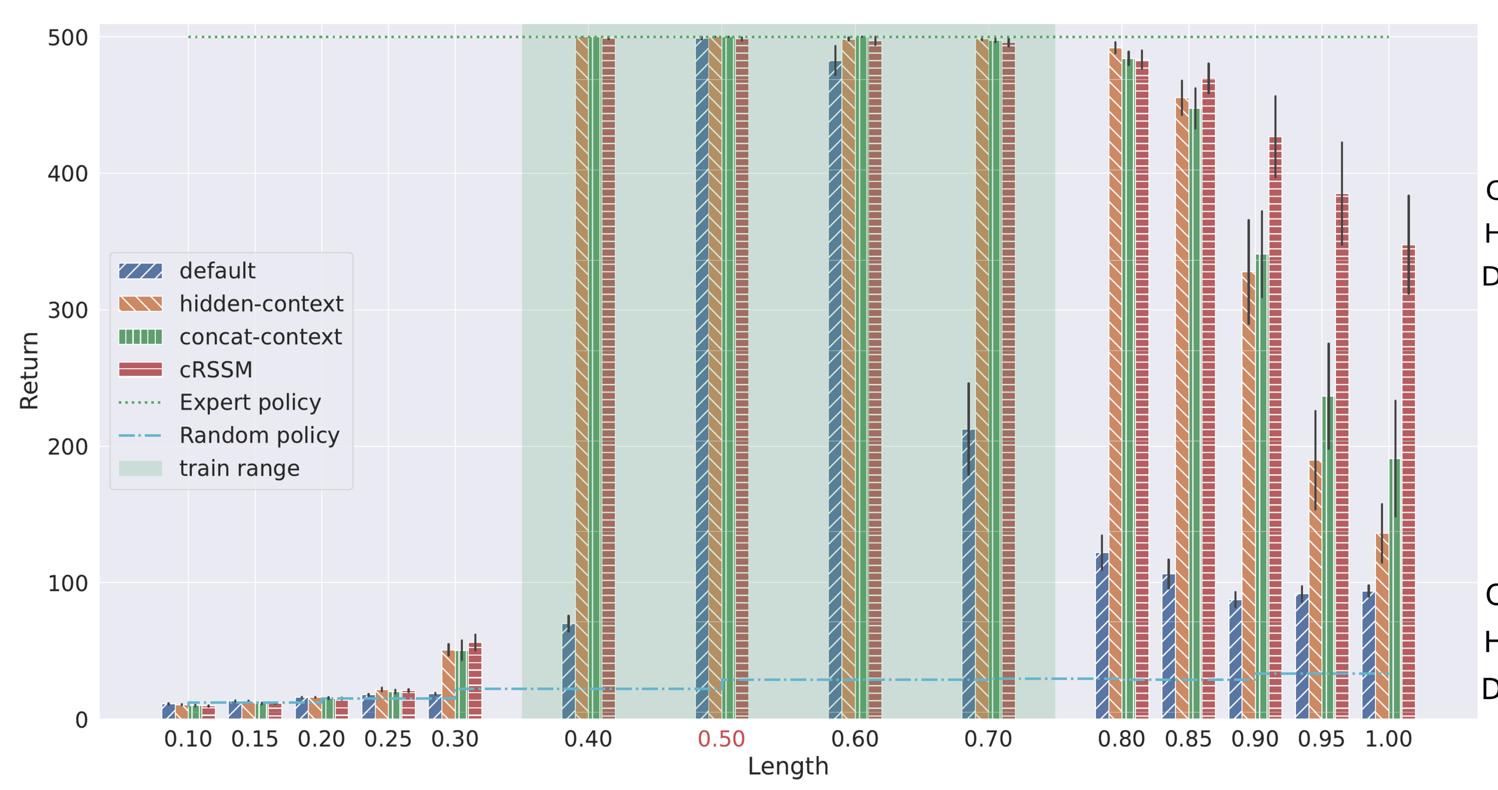
- **Default**: train on only the "default" context
- **Hidden Context**: train on multiple contexts without context being observable
- **Concat Context**: context as an observation
- **cRSSM:** appropriate conditioning of the world model and policy with context

Environments & Context

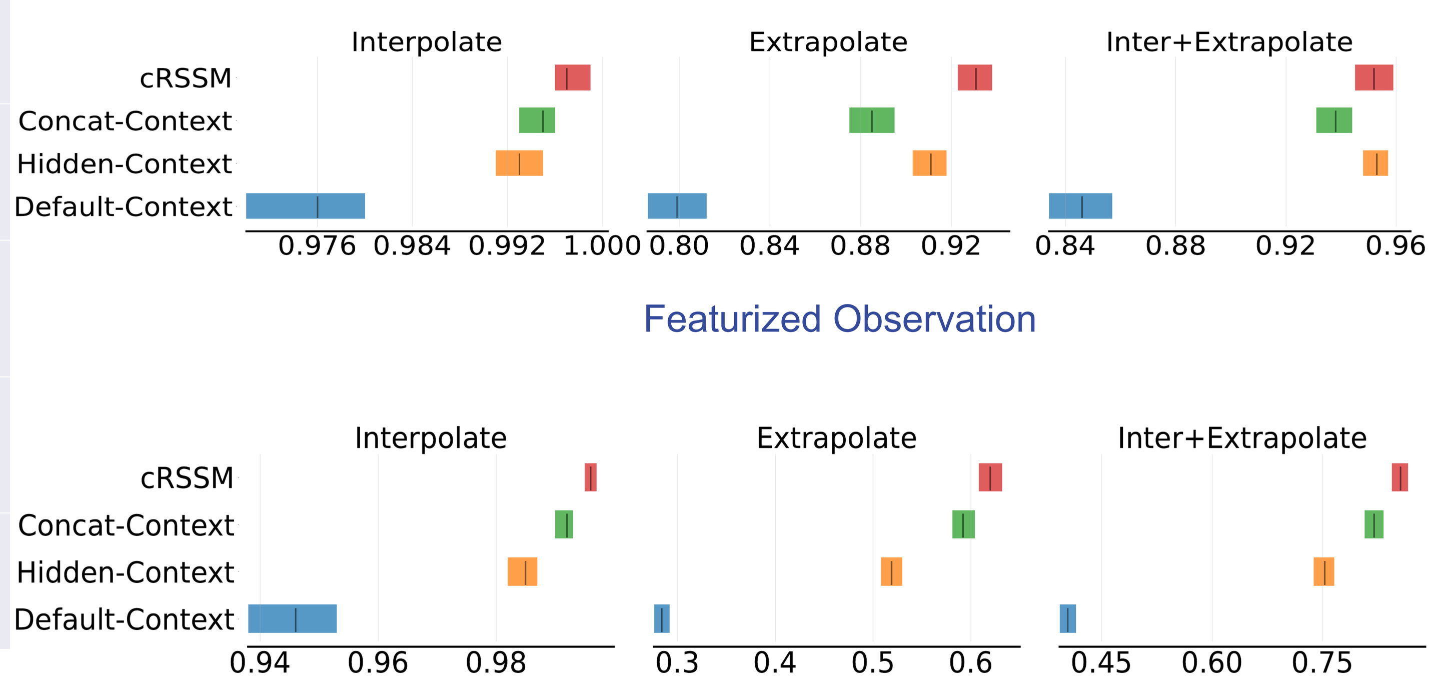
- **Cartpole** — Discrete Control Task
  - Gravity, Length
- **Walker Walk** — Continuous Control Task
  - Gravity, Actuator Strength

Quantitative Evaluation

Cartpole (Pixel) - Length



Aggregated Normalized IQM



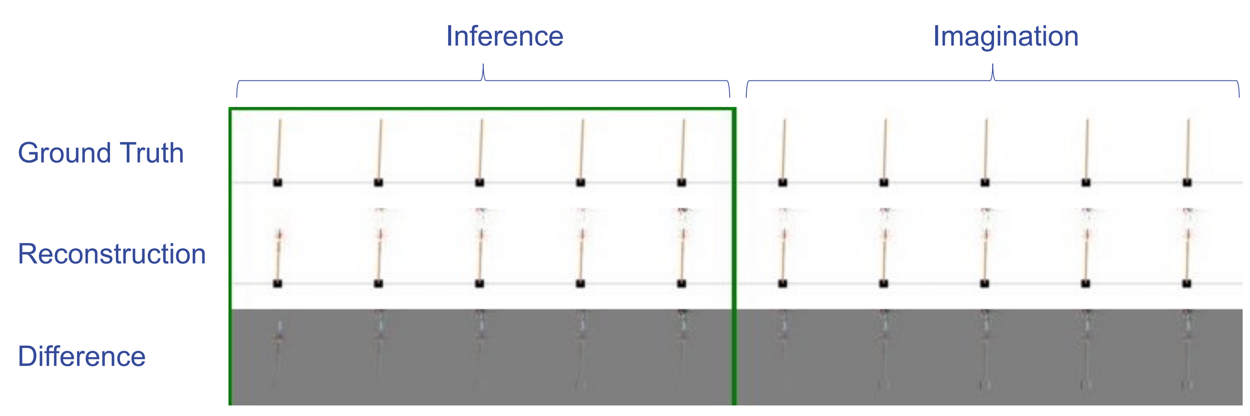
Pixel Observation

Future Work

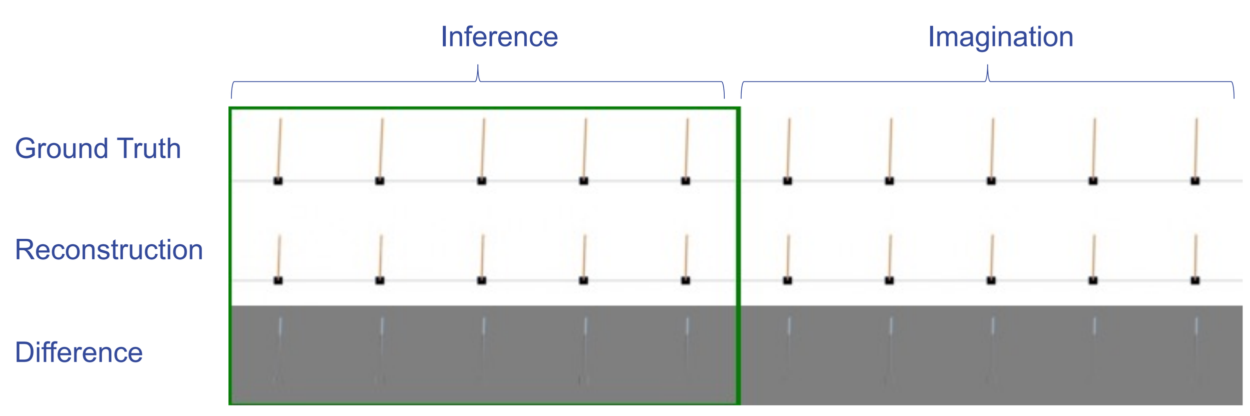
- Train in imagination in extrapolated contexts
- Inferring unobservable context and conditioning
- Standardized benchmarks for generalization with more appropriately designed environments

Qualitative Evaluation

1. Extrapolation to unseen longer pole length

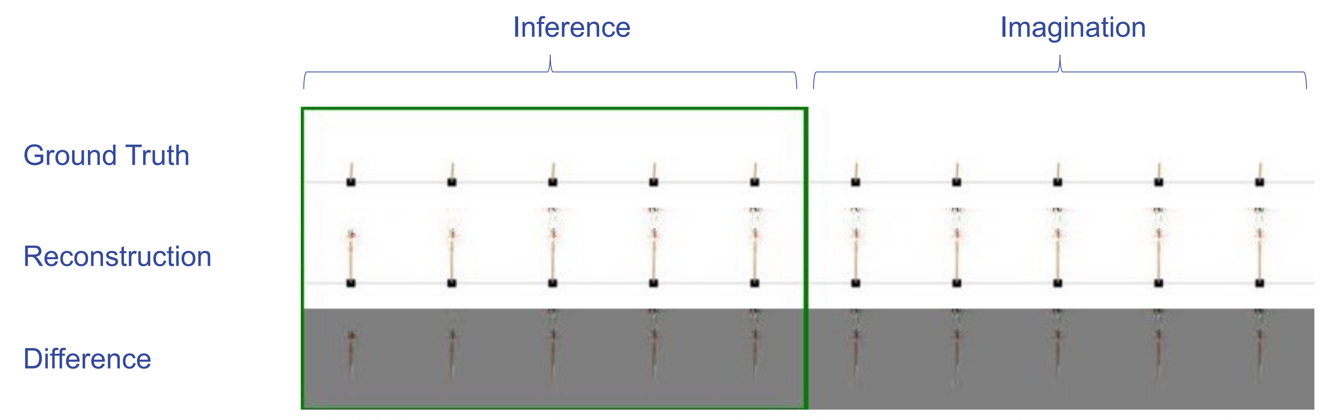


cRSSM – Additional pixels over the longest pole seen during training

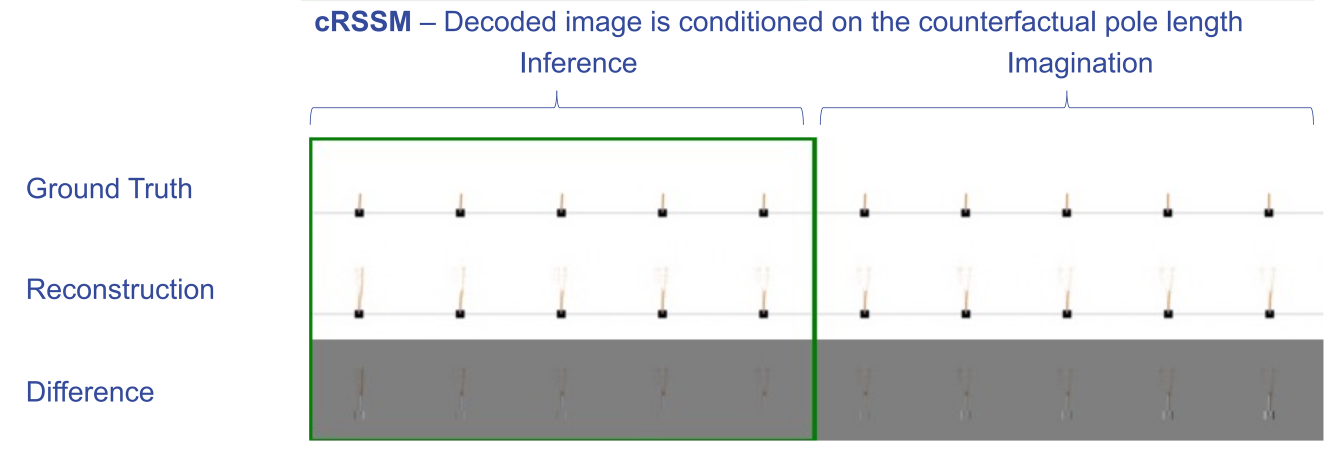


concat-context – Decoded pole image is the longest pole seen during training

2. Counterfactual conditioning on unseen longer pole length



cRSSM – Decoded image is conditioned on the counterfactual pole length



concat-context – decoded pole image is the longest pole seen during training

Paper

Code: github.com/sai-prasanna/dreaming_of_many_worlds