

Cyclophobic Reinforcement Learning

Stefan Sylvius Wagner Martinez, Peter Arndt, Stefan Harmeling

BACKGROUND

Systematic Exploration and Feature Extraction

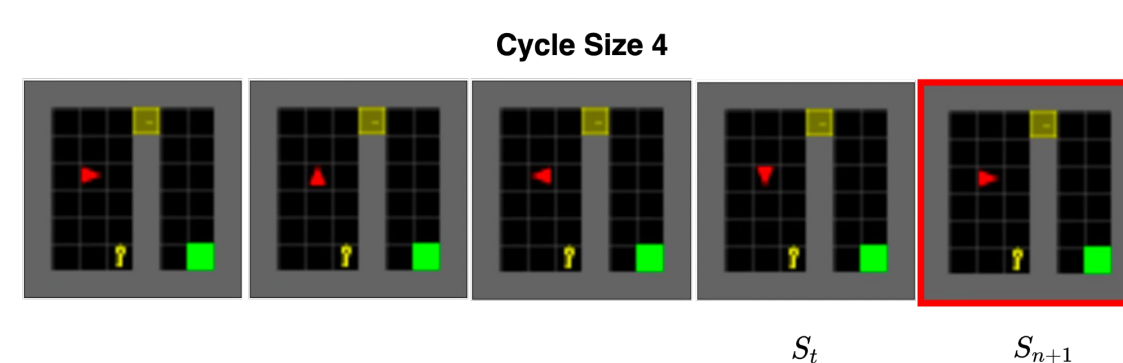
- How can we explore systematically in reinforcement learning?
- Exploration should characterize environment in a salient way
- Factorize environments in terms of salient objects.

Transfer Learning

- Use extracted features which factorize environment for transfer learning

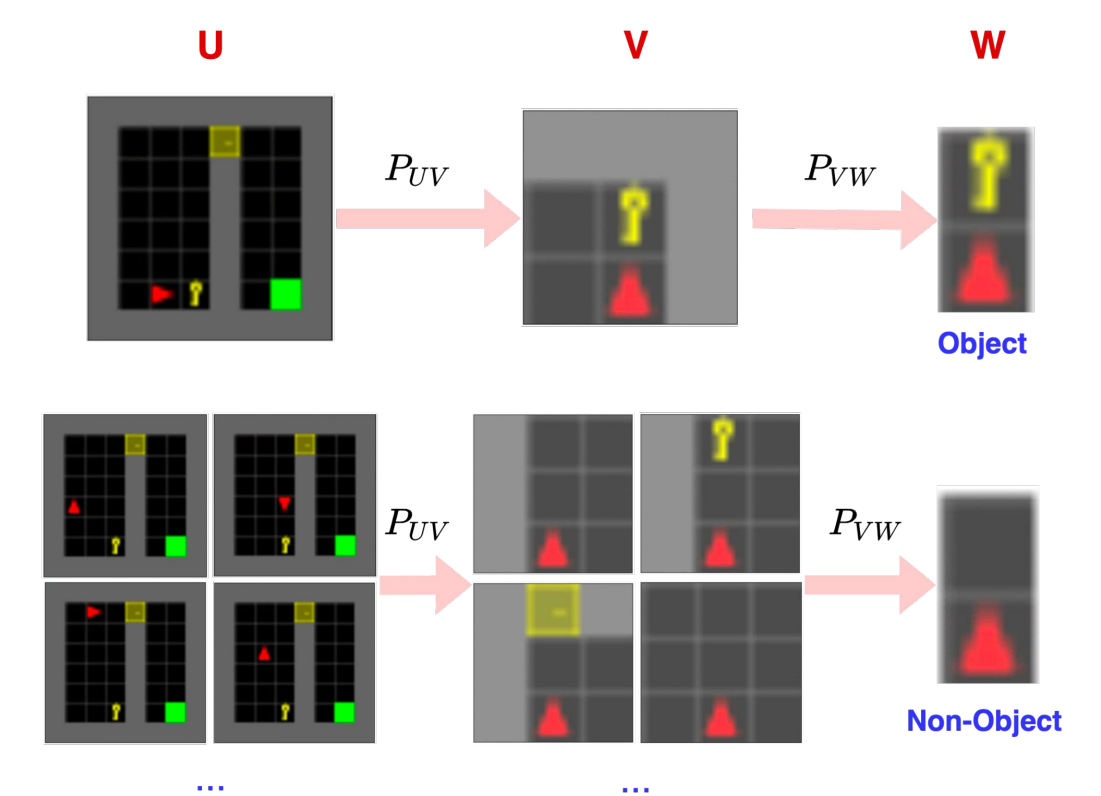
Cyclophobic Intrinsic Reward

Cycle Avoidment



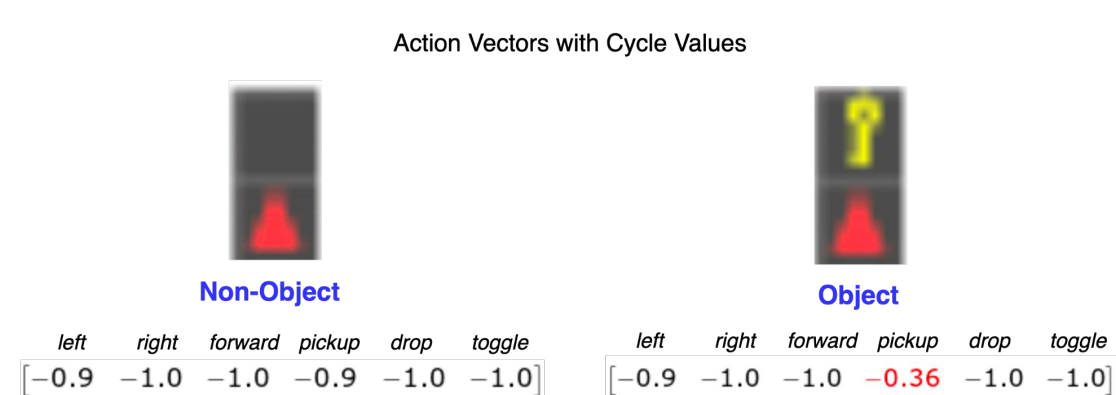
- A repeated observation is a cycle.
- We can determine how long the cycle is and how often it occurs.
- Systematic: First smaller cycles are avoided, then larger cycles.

Hierarchical Observations



- Salient objects induce different structure over multiple views.
- Learn different cycle-values over different views to characterize environment.

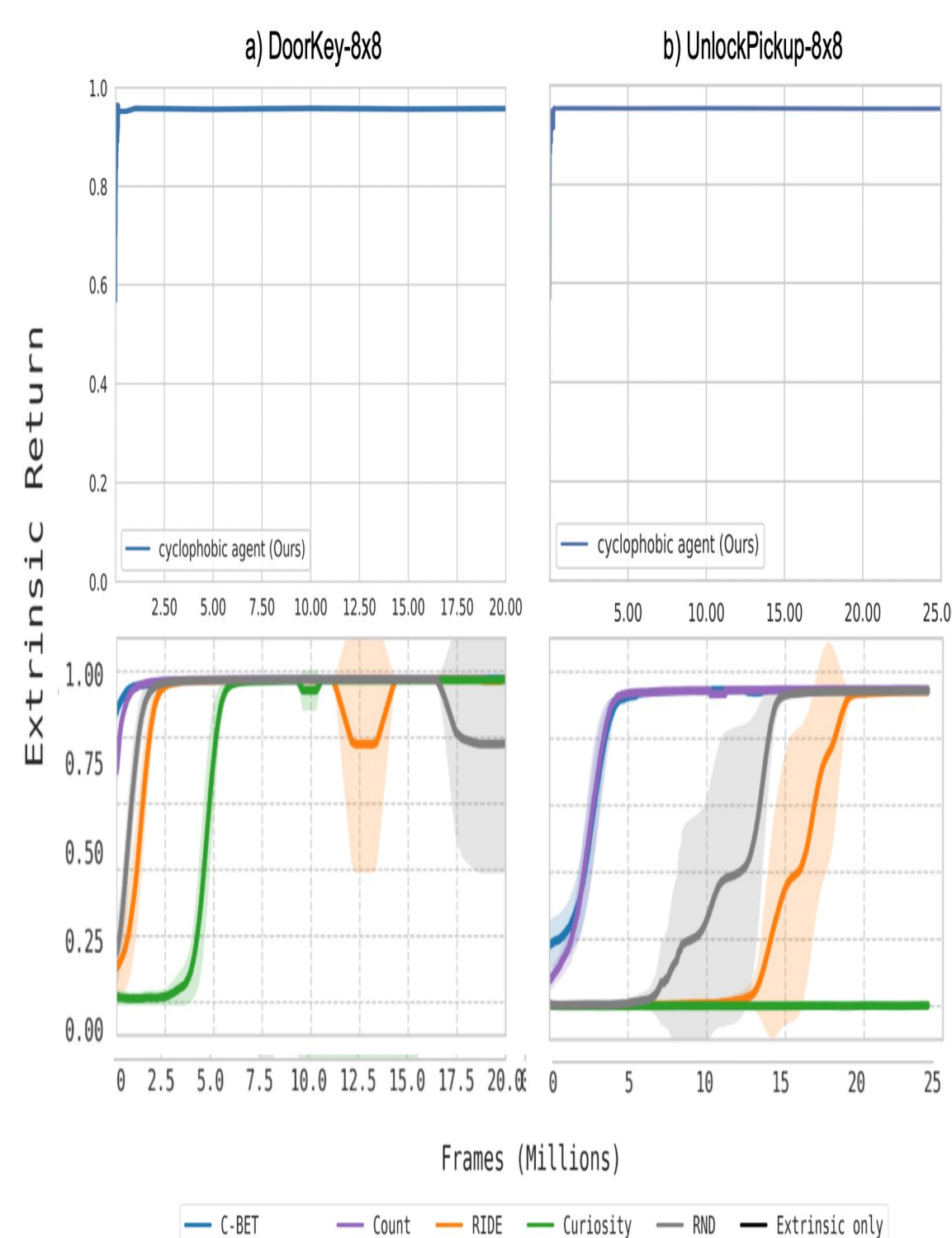
Finding salient objects



Using Cyclophobic Intrinsic Reward and Hierarchical Observations

- Some observations in the smallest view will have salient actions.
- Use these salient events for transfer learning.

Preliminary Results



CONCLUSIONS

So far...

- Very efficient training: 10x times more efficient than black-box end to end approaches.
- Learned features are transparent and interpretable.

Future work...

- Increase experiment variety
- Use learned objects to perform symbolic AI i.e. neuro-symbolic AI.