

History Compression via Language Models in Reinforcement Learning



Fabian Paischer, Thomas Adler, Vihang Patil, Angela Bitto-Nemling,
Markus Holzleitner, Sebastian Lehner, Hamid Eghbal-zadeh, Sepp Hochreiter

Partial Observability in Reinforcement Learning

- Memory mechanism to approximate MDP
 - LSTM [1]
 - Transformer [2]
- Transformer for on-policy RL [3]
 - requires plenty of interaction steps
 - prone to overfitting on scarce data

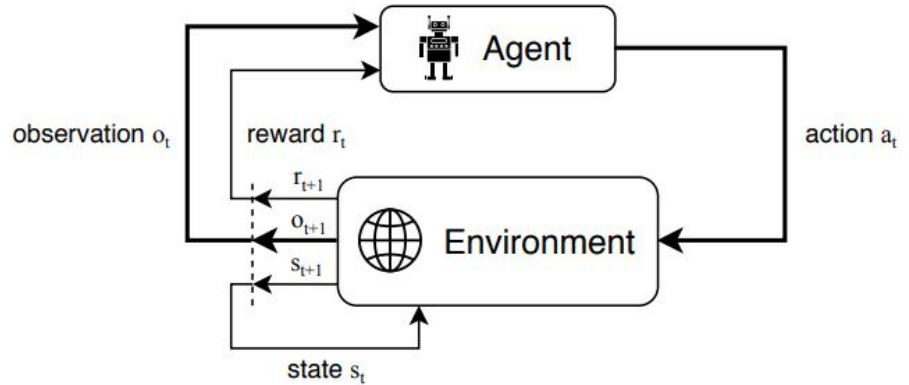
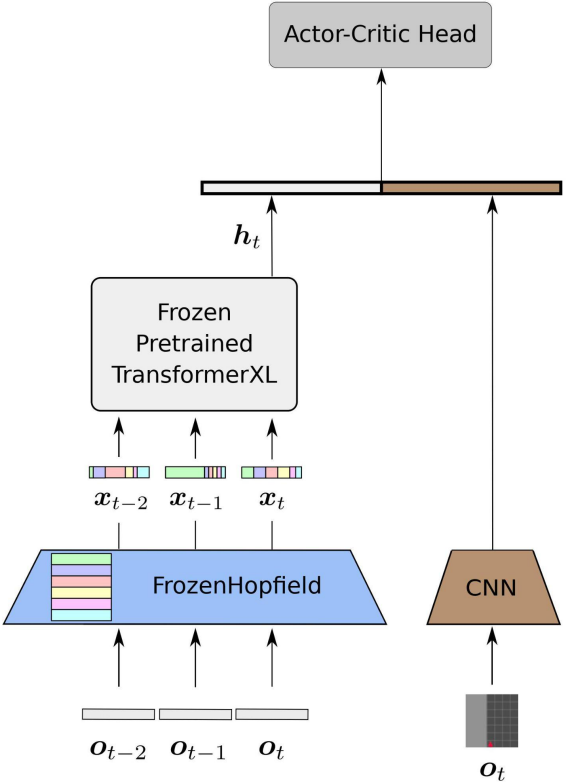
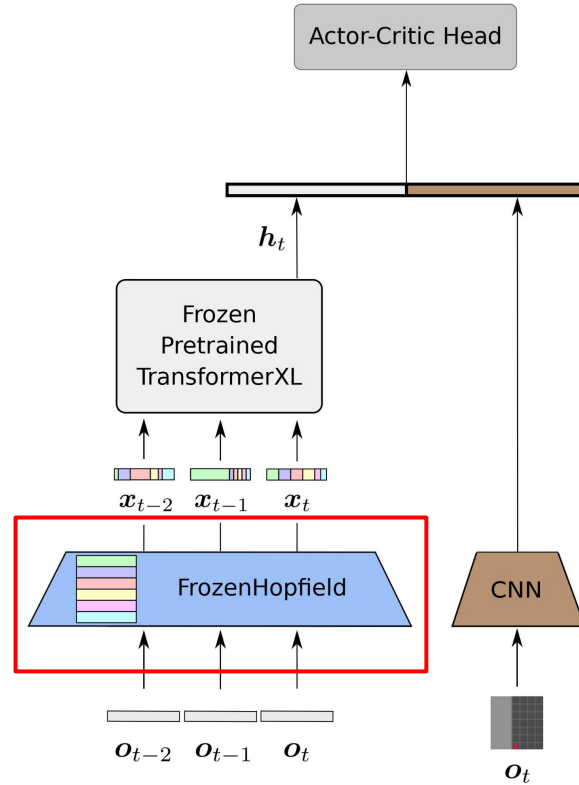


Figure 1: Reinforcement Learning under Partial Observability [4]

Motivation

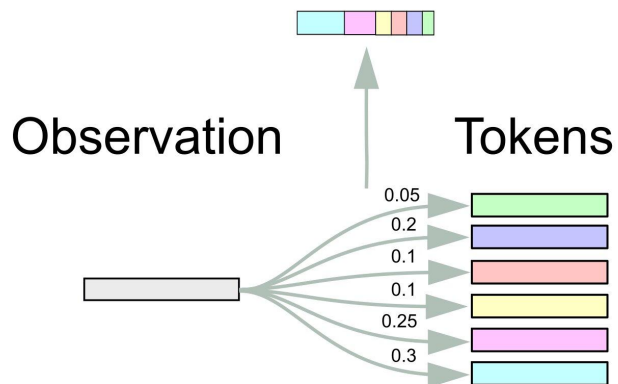
- Improve sample efficiency via abstract history compression
 - Language was optimized to provide high-level abstractions [5]
 - Abstraction facilitates forming and applying concepts and analogies [6]
 - Leveraged by humans to summarize and communicate past events to one another
- Pretrained language Transformers (PLTs) transfer well across modalities [7]
- Leverage language abstraction via PLTs in RL
 - freeze PLT to avoid instabilities during training/finetuning [8]

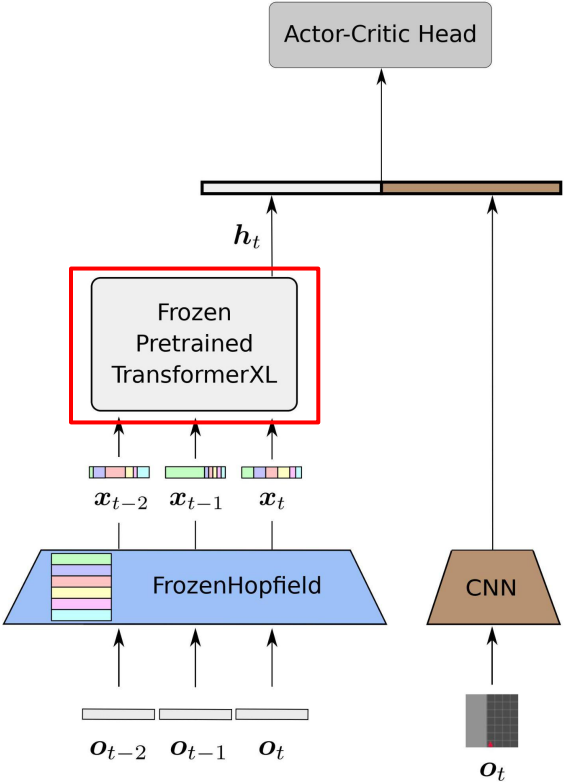


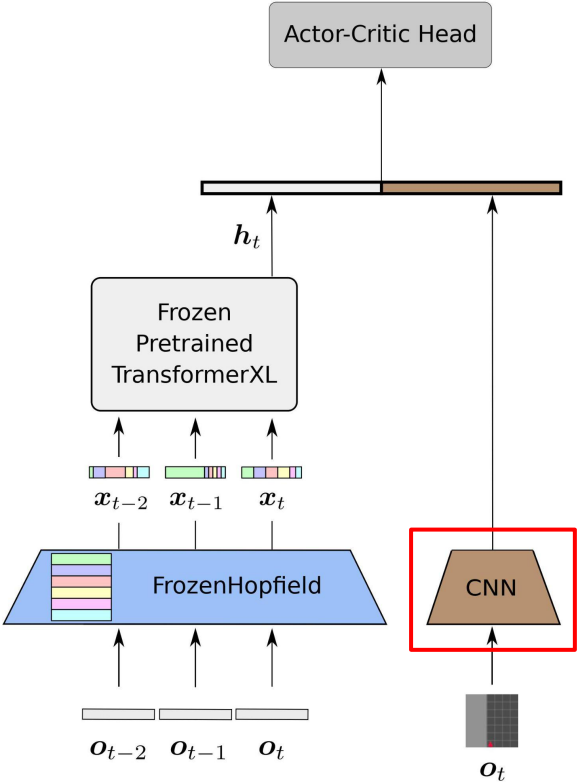


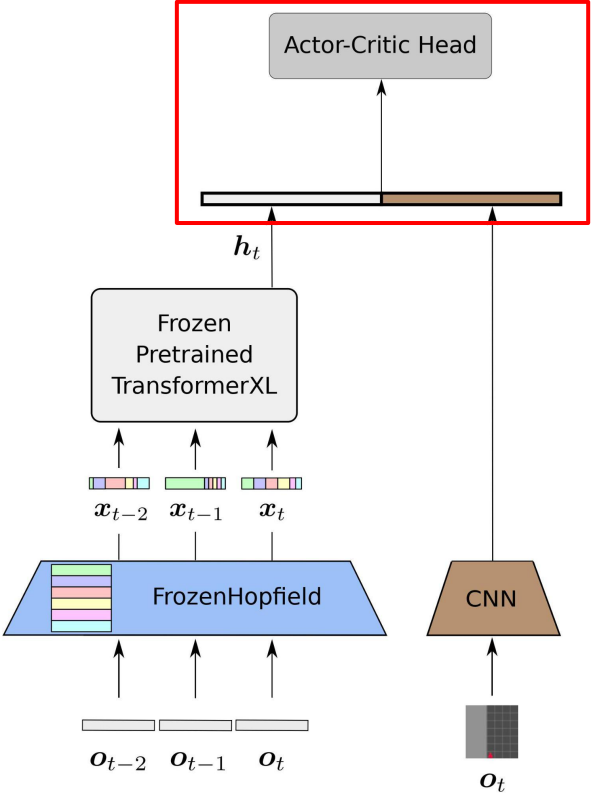
FrozenHopfield

- Translate observations to language domain
 - Leverage Johnson-Lindenstrauss lemma [9] to project observations
 - Store pretrained token embeddings of PLT in Modern Hopfield Network [10]
 - Retrieve a combination of tokens using projected observations as query









Experiment Design

- Baselines
 - Recurrent agent based on small Impala architecture (Impala-PPO [11])
 - Markovian Baseline (CNN-PPO)
- Gridsearch for hyperparameter optimization
 - Training via PPO [12]
 - Evaluation via interquartile mean (IQM, [13]) and Wilcoxon to test for statistical significance

Gridworld Experiments

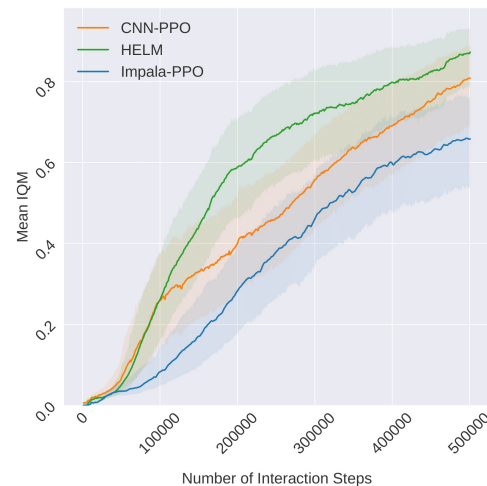
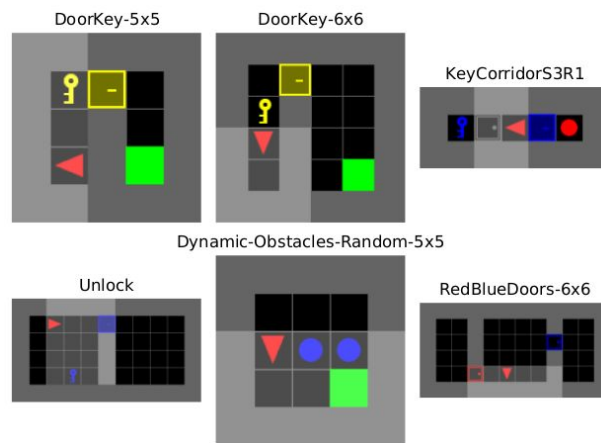


Figure 8: MiniGrid [14] toytasks (left), mean IQM across all environments over 30 seeds (right), HELM significantly outperforms all competitors.

Procgen

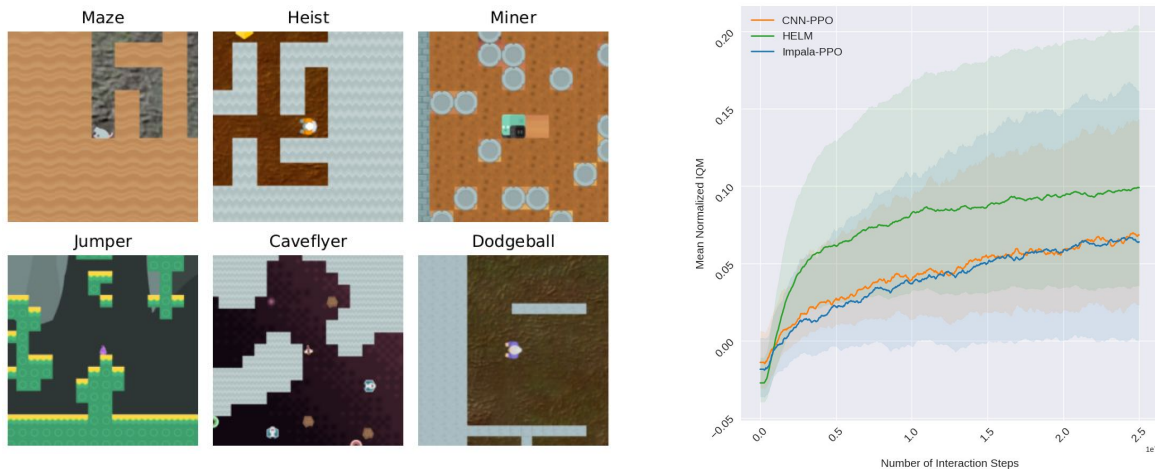


Figure 8: Procgen [15] environments (left), mean normalized IQM across all environments over 10 seeds (right), HELM significantly outperforms Impala-PPO.

Conclusions

- Language is well-suited for constructing abstract representations of the past
- FrozenHopfield allows mapping into the language domain without any training
- HELM outperforms LSTM on Minigrad and Procgen environments

 : <https://twitter.com/PaischerFabian>

 : <https://arxiv.org/abs/2205.12258>

 : <https://ml-jku.github.io/blog/2022/helm/>

 : <https://github.com/ml-jku/helm>

