# Semantic HELM: A Human-Readable Memory for RL

ELLIS Unit Linz, Institute for Machine Learning
Fabian Paischer, Thomas Adler, Markus Hofmarcher, Sepp Hochreiter

JOHANNES KEPLER
UNIVERSITY LINZ

# Partial Observability in Reinforcement Learning

- Most real-world problems are partially observable
  - Agent never observes true state of environment
- Approximate environment state via memory mechanism
  - LSTM [1]
  - Transformer [2]
- State-of-the-art systems lack interpretability [3,4]
  - It is not comprehensible for a human what pieces of information entered the memory

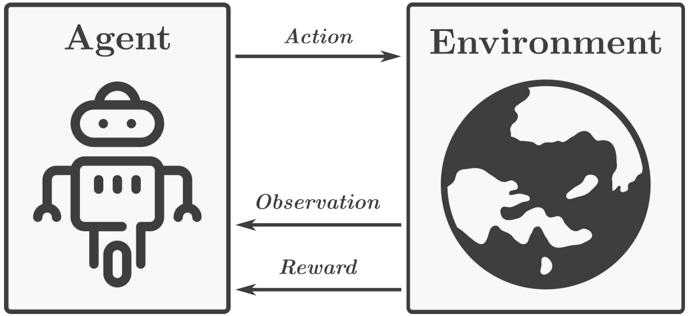



Figure 1: Reinforcement Learning under partial observability.

JOHANNES KEPLER UNIVERSITY LINZ

[1] Long Short-Term Memory, S. Hochreiter and J. Schmidhuber, Neural Computation, 1997
[2] Attention is All you Need, A. Vaswani et al., Advances in Neural Information Processing Systems, 2017
[3] Dota 2 with Large Scale Deep Reinforcement Learning, Berner et al., 2019
[4] Grandmaster level in Starcraft II using multi-agent Reinforcement Learning, Vinyals et al., Nature Communications, 2019

# How can we make our agents more interpretable?

- By using human language to compress past observations
  - Language was optimized to provide high-level abstractions [1]
  - Humans memorize abstract concepts rather than every single detail [2]

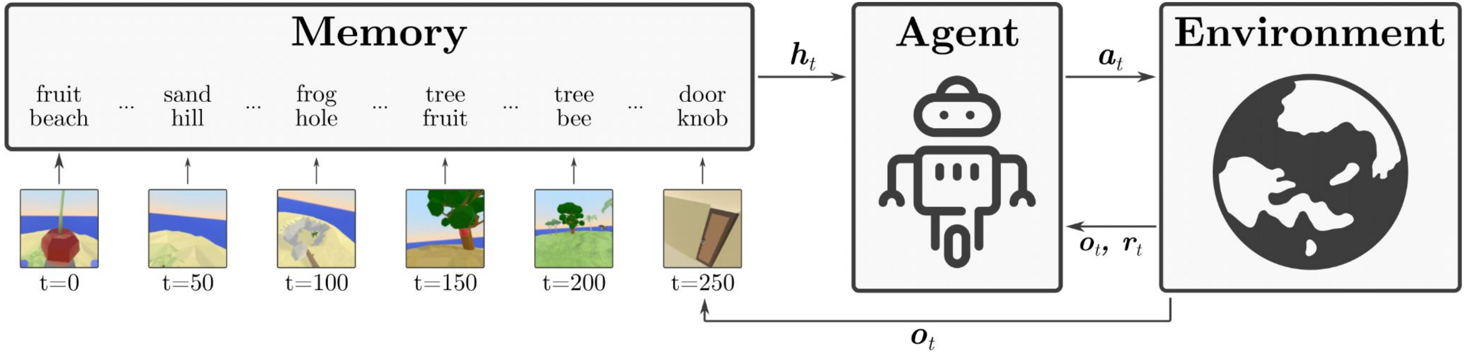
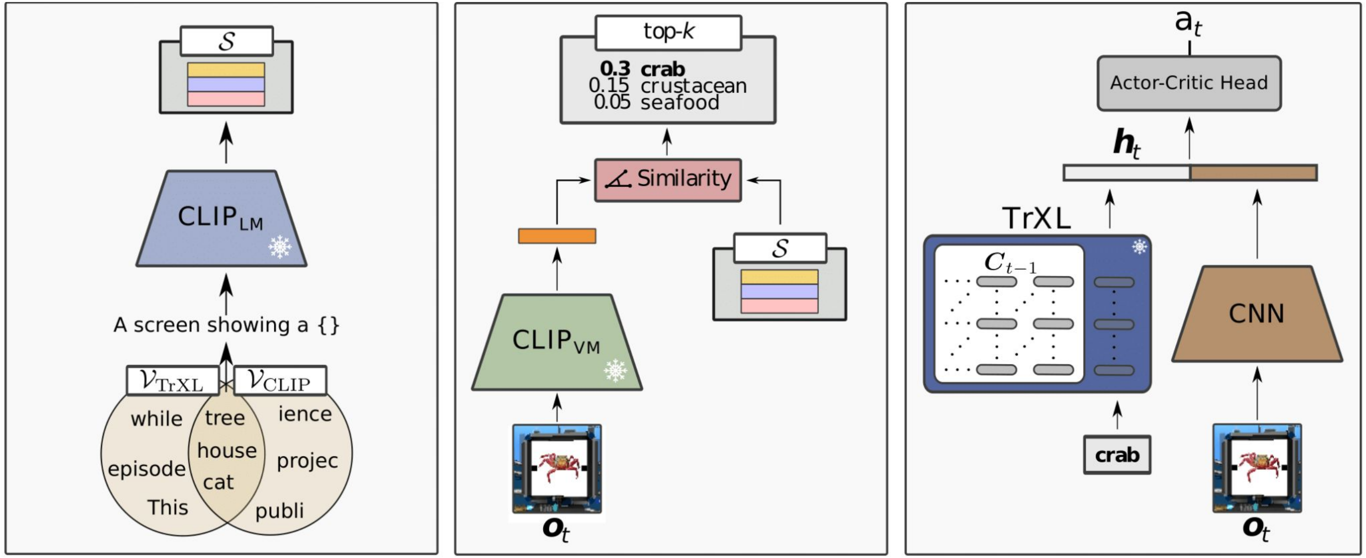



Figure 2: We add a semantic and human-readable memory to an agent to tackle partially observable RL tasks. Visual observations are mapped to the language domain via a CLIP retrieval. The memory component, a pretrained language encoder, operates on text only and compresses a history of tokens into a vector. The agent takes an action based on the current observation and the compressed history.

JOHANNES KEPLER UNIVERSITY LINZ


[1] The evolution of language, Nowak et al., Proceedings of the National Academy of Sciences, 1999
[2] The persistence and Transience of Memory, Richards and Frankland, Neuron, 2017

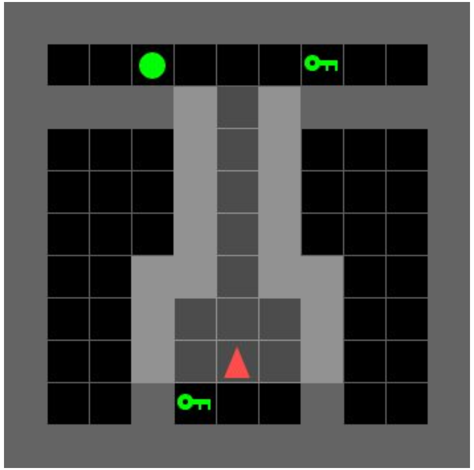
# Semantic HELM




**(a)** **(b)** **(c)**

Figure 3: Architecture of SHELM.
**(a)** We compile a semantic database by encoding prompt-augmented tokens from the overlapping vocabularies of CLIP and the TrXL.
**(b)** Given an observation, we retrieve the top-k embeddings and select their corresponding text tokens.
**(c)** These tokens are passed to the TrXL which represents the memory module of SHELM.

# MiniGrid-Memory



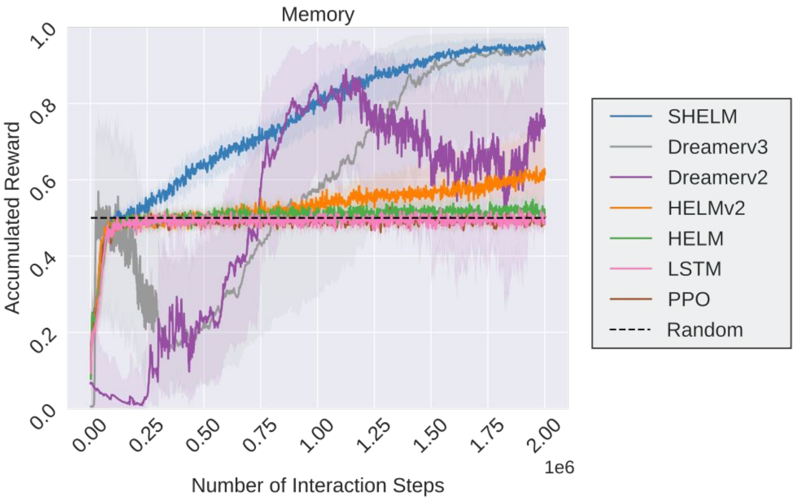
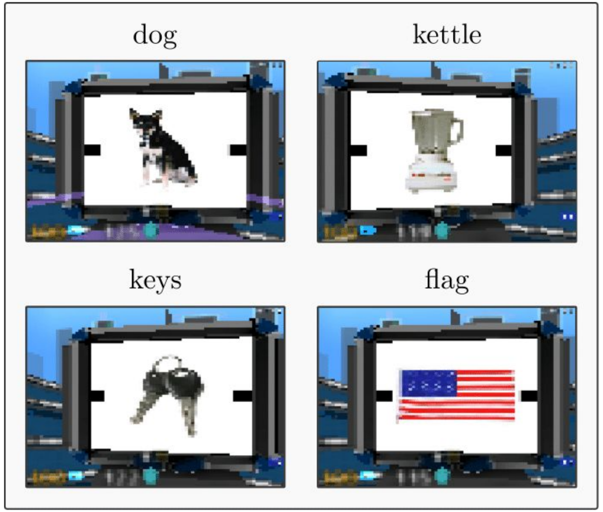



Figure 5: Results on the MiniGrid-Memory task.
**Left:** The MiniGrid-Memory task [1].
**Right:** Mean IQM and 95% bootstrapped CIs across 30 seeds on MiniGrid-Memory environment for different memory-based agents.

JOHANNES KEPLER UNIVERSITY LINZ

[1] Minimalistic gridworld environment for openai gym, Chevalier-Boisvert et al., 2018

# Psychlab - Continuous Recognition



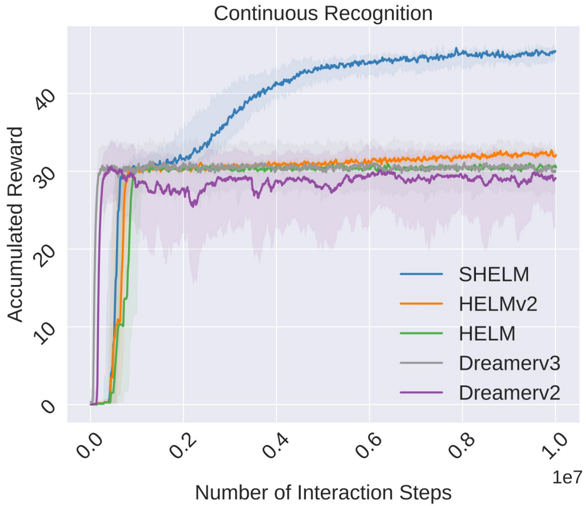



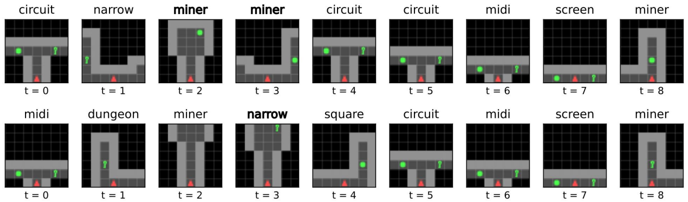

Figure 9: Results on the continuous recognition task of Psychlab.
**Left:** Sample observations and associated language tokens that were stored in the memory for SHELM on Psychlab.
**Right:** Mean IQM [1] and 95% bootstrapped CIs across 5 seeds over the Psychlab continuous recognition task for different memory-based agents.

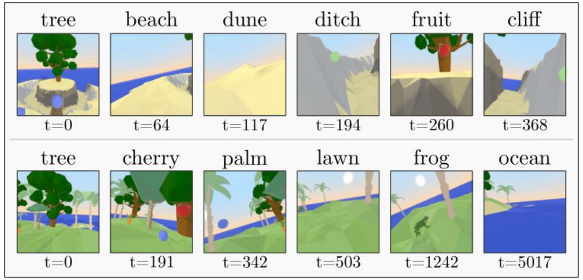
JOHANNES KEPLER UNIVERSITY LINZ

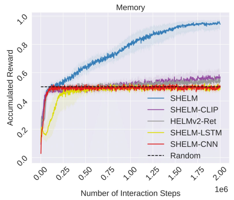
[1] Psychlab: A Psychology Laboratory for Deep Reinforcement Learning Agents, Leibo et al, 2018

# Conclusions

- SHELM adds interpretability to memory mechanism
- Semantics are not always important as long as vision encoder can discriminate between objects
- SHELM excels in environments that heavily rely on memory
- Partial observability does not necessarily imply memory dependency




SCAN ME

JOHANNES KEPLER UNIVERSITY LINZ