DeepMind

Towards mental time travel: A hierarchical memory for RL

Andrew K. Lampinen, Stephanie C. Y. Chan, Andrea Banino, Felix Hill

Motivation

- Human learning and generalization depends on detailed memory of the past.
- Often referred to as ability to perform "mental time-travel"—to relive a past experience.
- Can allow us to e.g. combine memories of routes we've taken through a city to plan a new one.
- Or to learn something one shot and recall it for a long time.
- By contrast, RL agents have very limited memory.

Want RL agents which can recall the past in detail, in order to adapt to the challenges of the present.

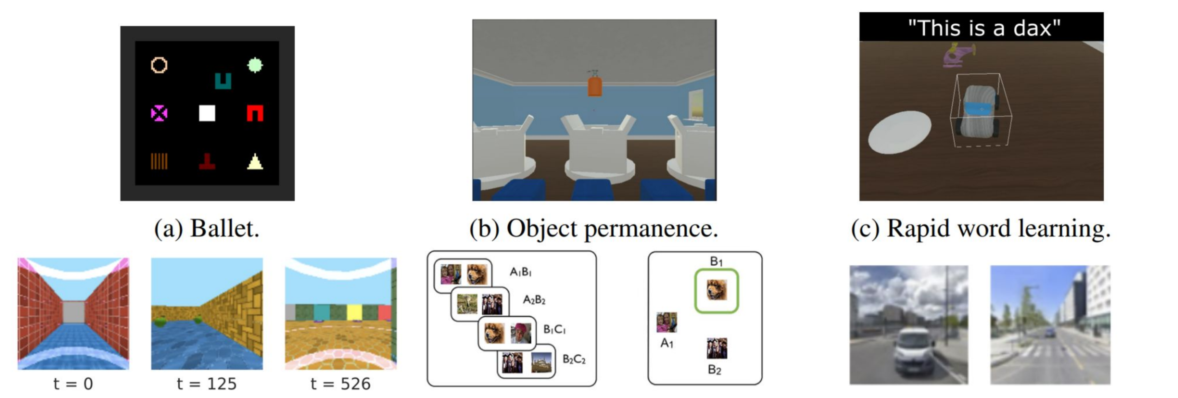
Transformers in RL

- LSTMs are the standard RL agent memory, but can struggle with long-term recall.
- Transformers can improve memory in RL [1].
- But even in supervised tasks, transformers fail to achieve detailed recall over moderately long windows [2].
- And this problem is likely exacerbated by the sparse learning signals in RL.

Goal is to augment transformers with version of mental time travel:
- **Detailed recall of specific past events.**
- **Without interference from everything else.**

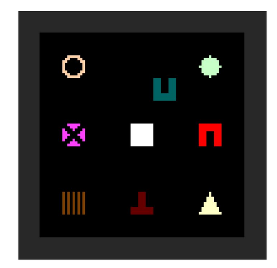
Experiments

- Evaluated on diverse domains & challenges.
- Compared to (gated) TrXL [1], LSTM and prior work on existing tasks.

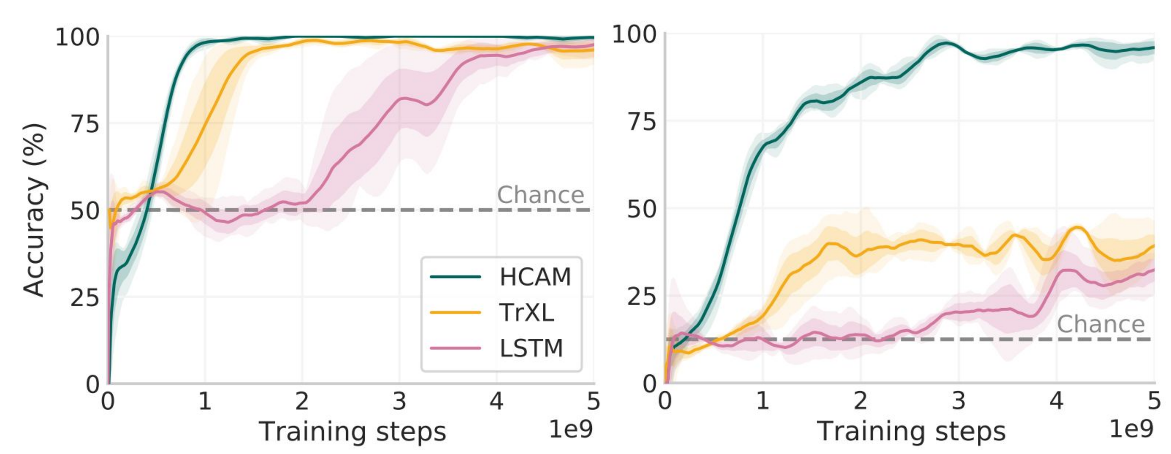


(a) Ballet. (b) Object permanence. (c) Rapid word learning. (d) Visual match. (e) Paired associative inference. (f) One-shot StreetLearn navigation.

Recalling a Ballet



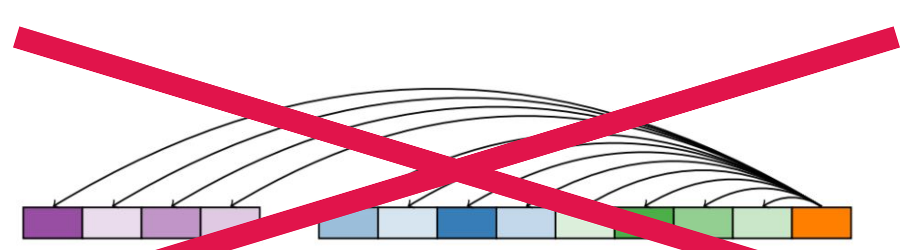
- Agent watches sequence of dances.
- At end, prompted with dance name, rewarded for navigating to correct dancer.
- Necessary to recall spatio-temporal sequences rather than just single timepoints.
- HCAM performs best, especially on long tasks.



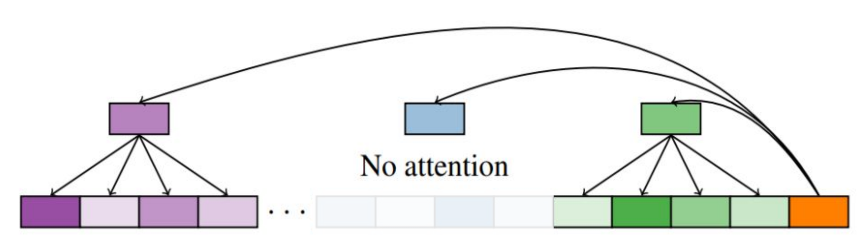
(a) 2 dances. (c) 8 dances, long delays.

Proposal: Hierarchical Chunk Attention

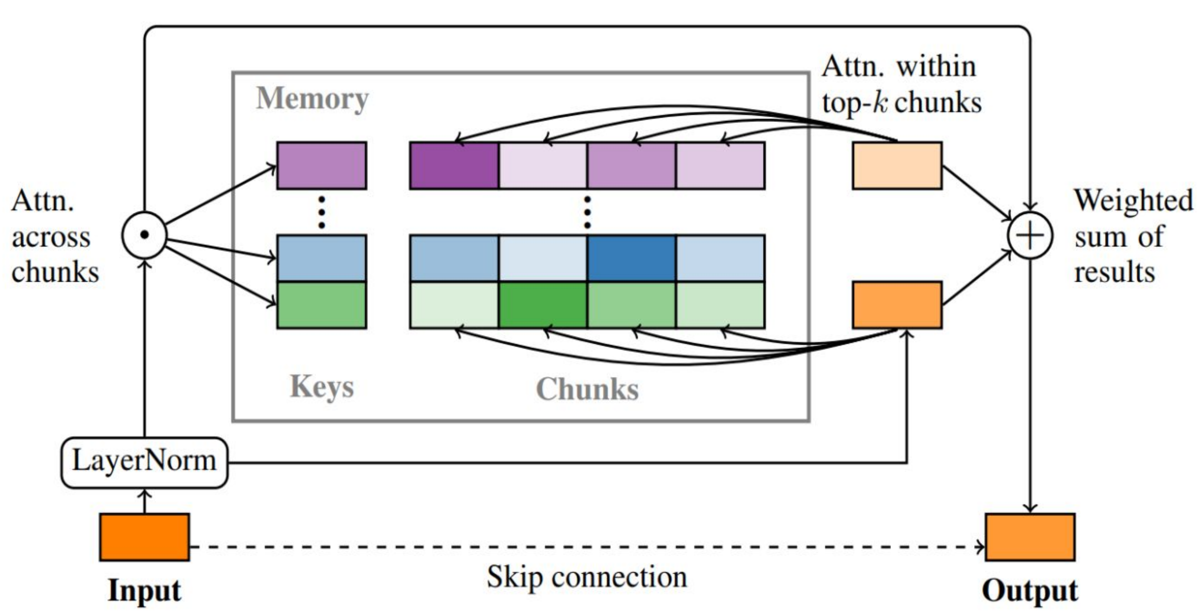
- Rather than attend over large sequences of past events, attend hierarchically.
- First, attend to summaries of past chunks, to see if they are relevant.
- Then, attend in detail **only** within the top-k most relevant chunks.
- This allows detailed attention to relevant events:
  - Regardless of how far back in memory.
  - Without interference from everything else.



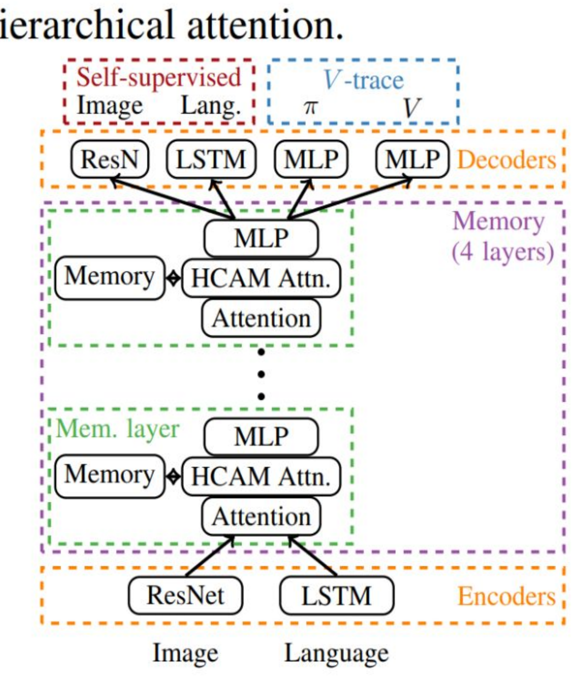
(a) Standard attention.



(b) Hierarchical attention.



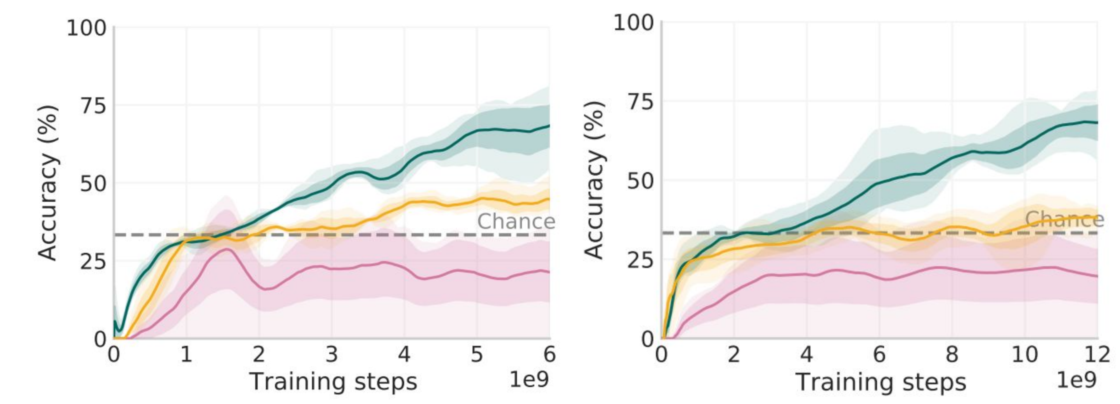
(c) HCAM attention block.



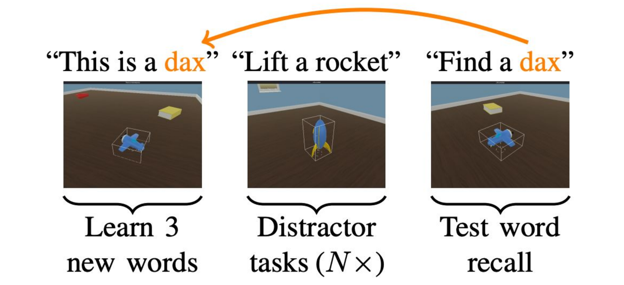
(d) HCAM in an agent.

Rapid word learning and longer term recall

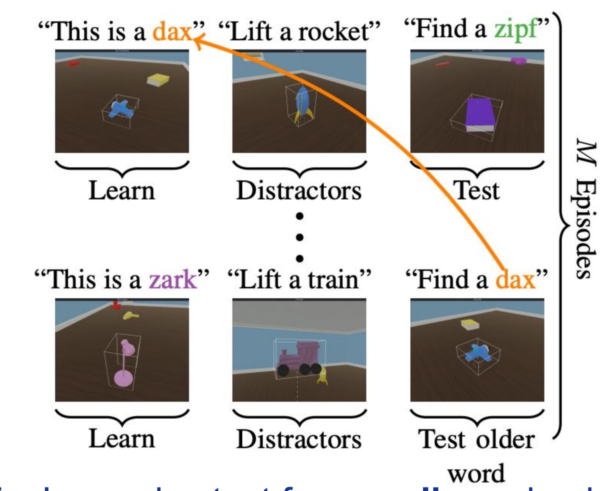
- Adapt rapid word-learning tasks [4].
- Add distractor tasks in training episodes.
- Evaluation: **zero-shot** ability to recall a word from several episodes earlier.
- HCAM performs well on these challenging tests, despite **no** multi-episode training.



(f) Eval. across 2 eps., 5 dist. each. (b) Eval. 4 eps., 1 dist.



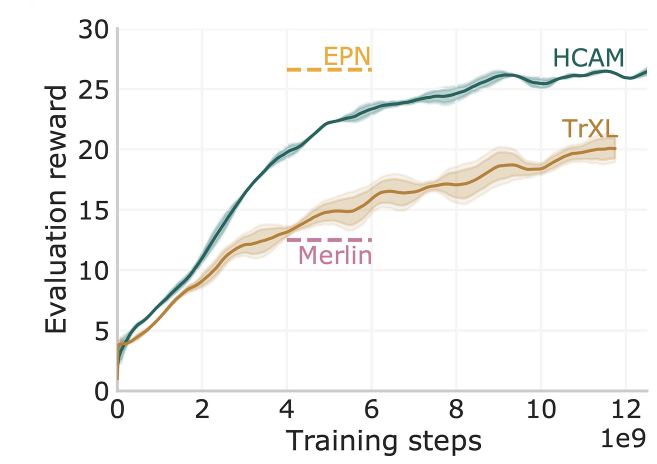
Train: single learn-dist.-test episode.



Eval: surprise test from **earlier** episode.

Navigating a new city: One-Shot Street Learn [3]

- Must rapidly learn to navigate new neighborhood from StreetView images.
- Prior work [3] had proposed complex, planning-specific memory architecture.
- HCAM matches its **near-optimal planning** with more generic architecture (though takes longer to learn fully).
- Other strong generic baselines do not.



Conclusions

- HCAM is an effective RL memory that allows "mental time travel."
- Outperforms baselines at varied tasks, e.g. recalling spatio-temporal events or newly-learned words, or navigating a new city.
- Also better in several other domains, more efficient, etc. (see paper).

Potentially broadly useful ideas:
- Hierarchical attention: early level selects for further attention.
- Key-value memories with structured values (not just vectors).
- Detailed memory allows meta-learning in activations to have longer-term consequences for behavior.

References

[1] Emilio Parisotto, et al. Stabilizing transformers for reinforcement learning. CML 2020.
[2] Yi Tay, et al. Long range arena : A benchmark for efficient transformers. ICLR 2021.
[3] Samuel Ritter, et al. Rapid task-solving in novel environments. ICLR 2021.
[4] Felix Hill, et al. Grounded language learning fast and slow. ICLR 2021.