USE OF AN ARTIFICIAL HIPPOCAMPAL ALGORITHM FOR ONE-SHOT AND CONTINUAL LEARNING

Gideon Kowadlo, Abdelrahman Ahmed, David Rawlinson

Cerenaut Research (gideon/abdel/dave@cerenaut.ai)

Motivation

Conventional ML models are trained on large i.i.d. datasets. Additional training on new classes or incremental concept drift can cause catastrophic forgetting of existing representations. Sample efficient (e.g. one-shot) and continual or lifelong learning remain challenging [5].

A standard model for learning in the mammalian brain is CLS (Complementary Learning System) [3]. In CLS, Neocortex and Hippocampal Formation (HF) comprise two complementary learning systems with bidirectional connections. Neocortex learns gradually structured representations of the environment and HF learns specifics quickly (one-shot). HF slowly consolidates memories into Neocortex without interference with existing memories (continual learning), through interleaved replay of stored memory patterns. Replay occurs when the animal is in a passive state, and it also occurs as a response to external cues when in an active state. In this model, the HF constitutes short-term memory (STM) and Neocortex constitutes long-term memory (LTM). The Hippocampal model is a promising avenue to improve conventional ML.

Contributions

In our previous work, we introduced an Artificial Hippocampal Algorithm (AHA) [2] based on CLS. We now show that AHA can be used to complement a conventional, slowly-learned ML model (LTM) in a way that is analogous to the interaction of HF and Neocortex, to provide one-shot and continual learning capabilities. Here we present preliminary results.

Related Work

Replay methods are often used in continual learning. In [7], Hippocampal replay was inspiration for a deep generative model (GAN) that mimics past data to interleave training of the 'task solving' model, applied to static images. The authors of [6] used growing recurrent networks for a Hippocampal-inspired episodic module to learn instances that are replayed to a semantic module, applied to video.

Model

The model is shown in Fig. 1. AHA rapidly encodes samples with sparse representations that do not overlap or interfere with each other. This is achieved with a pattern separation and pattern retrieval pathway, giving it the ability to separate highly similar inputs as well as generalise (see [2]). AHA operates like an autoassociative memory. It uses an input cue to recall and 'reconstruct' a corresponding memory. AHA itself learns without externally provided labels. In this work, AHA is modified to memorise and recall both input samples and labels, to be used to train the LTM with supervised learning.

The LTM is pre-trained to learn visual features common to the dataset. After pre-training, when presented with new samples (including previously unseen classes), AHA receives input from LTM and learns combinations of these features in one-shot (i.e. from a single exposure). Subsequently, these classes will be recognised by AHA prompting replay to LTM. The replayed and current encodings of LTM are interpolated, which we define as 'Enhanced Inference'. AHA can also be used in a passive replay mode, independent of external input. Patterns of stored memories are replayed to LTM in a randomly interleaved fashion, in an 'internal' training phase. Memories are consolidated to long term memory, LTM, and the short term memories can be overwritten. LTM is able to learn new classes from only one exposure.

LTM learns incrementally. An unsupervised single layer sparse convolutional autoencoder learns features and a single layer softmax is used as a classifier. The classifier is trained with supervised learning in a pre-training phase, and during consolidation.

Model Illustrated

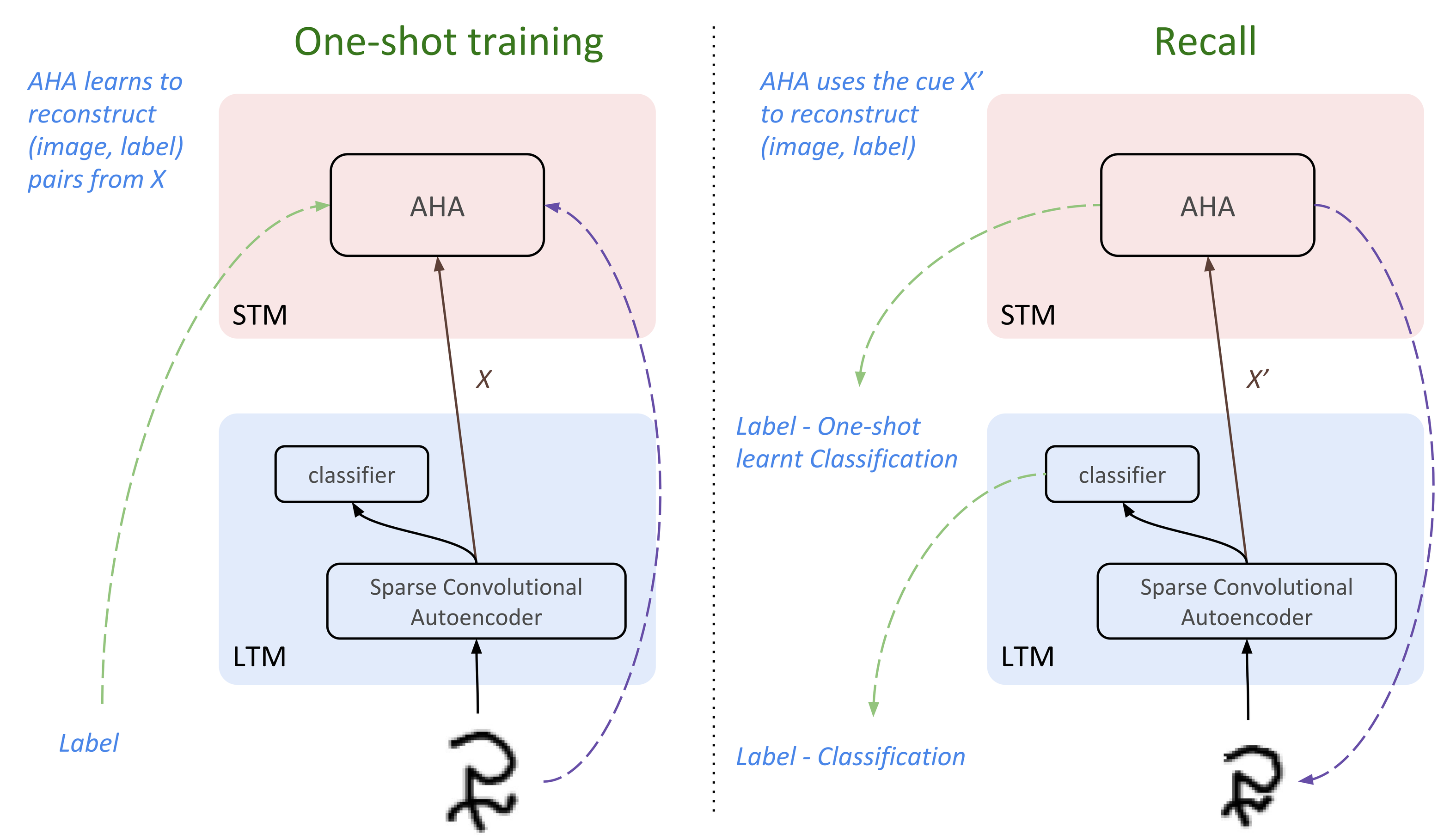

Fig. 1: **AHA+LTM in operation (Evaluation phase, see Fig. 2):** During training, AHA learns to reconstruct Label and Image in one-shot. While in STM, a different character image cues recall of the learnt Image and Label and the labels are interpolated for 'Enhanced Inference'. Offline replay consolidates memory into LTM, also improving long term performance.

Experimental Method

The overall task is classification of Omniglot characters, Fig. 3. The process is illustrated in a flowchart, Fig. 2.

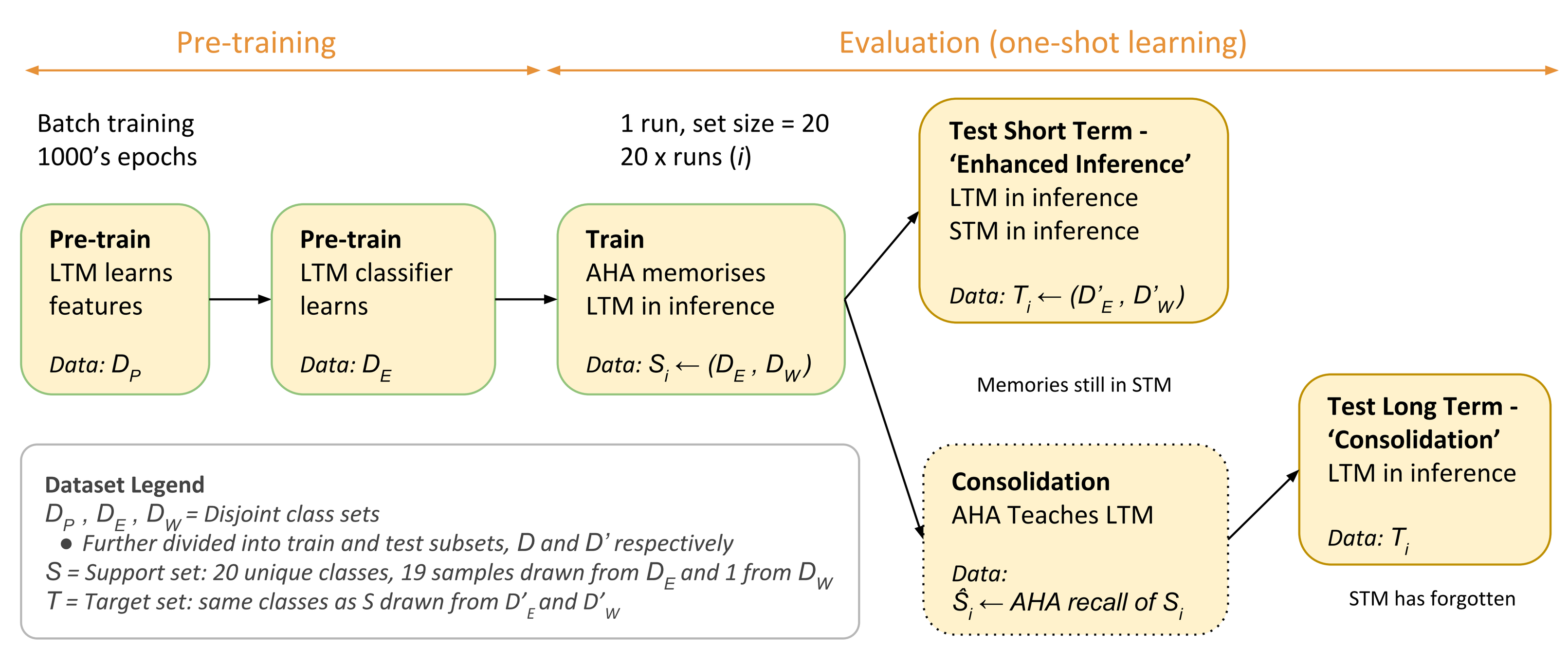

Fig. 2: **Experimental Method Flowchart**

The consolidation stage is achieved as follows. AHA repeatedly recalls (Image, Label) pairs from internally generated random cues (uniformly sampled). The resulting outputs are fed back into the system to retrieve a crisper reconstruction and more accurate label classification (referred to as big-loop recurrence [3]). A threshold is used to filter out remaining faint images. The 'random recall' procedure is repeated for 25 steps per run, and the resulting (Image, Label) pairs are stored in a replay buffer for consolidation. The size of the replay buffer was chosen to make it highly probable to recall all of the memorised samples, and was optimised empirically. We randomly sample from the replay buffer (with a bias towards the unseen sample) and allow the LTM to train on these samples for 160 steps. The replay buffer sampling mechanism ensures that there is at least one unseen sample (based on the label, as predicted by AHA).

Omniglot Dataset

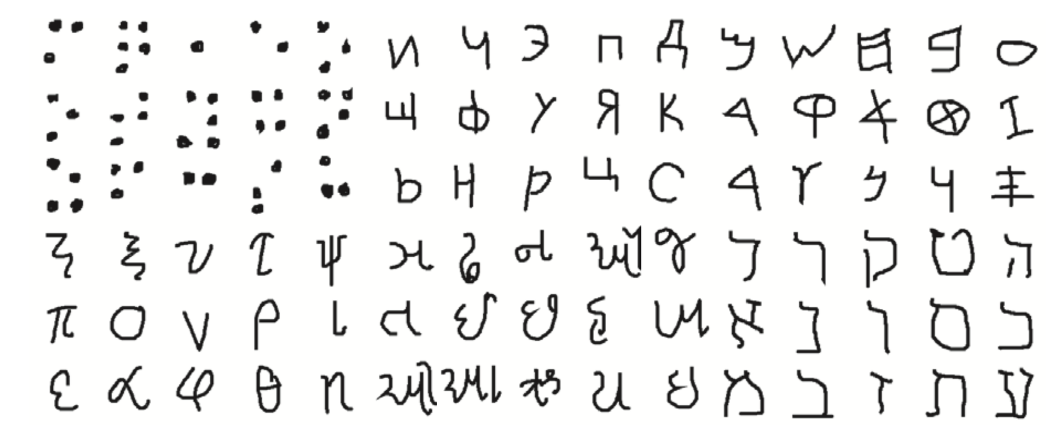

Fig. 3: **Omniglot:** Handwritten characters from a range of alphabets of different styles. Modified from [4]

Results

Classification using the slowly learned representation of LTM achieves accuracy of 97.5% overall, 45% on the one-shot classes. Enhanced Inference lowers accuracy to 80% overall, but boosts it from 45% to 60% on one-shot classes. After consolidation the accuracy of LTM is increased to 98% overall and 60% on the one-shot classes.

Model	Accuracy (per run)	Accuracy (per one-shot learnt class)
LTM	97.5%	45.0%
LTM+AHA - Enhanced Inference	80%	60%
LTM+AHA - Consolidation	98%	60%

Tab. 1: **Classification accuracy for baseline (LTM) and LTM+AHA:**. AHA dramatically improved one-shot learning performance and was able to consolidate that learning into Long Term Memory (LTM)

Conclusions and Future Work

This work shows how an artificial hippocampal algorithm can be used in a practical system to improve a standard supervised learning model, enabling it to learn new classes after only one exposure. We are currently running similar experiments with the continual fewshot learning framework [1] comparing performance to several algorithms on Omniglot and Slimagenet.

References

[1] Antreas Antoniou et al. "Defining Benchmarks for Continual Few-Shot Learning". In: *arXiv preprint arXiv:2004.11967* (2020).

[2] Gideon Kowadlo, Abdelrahman Ahmed, and David Rawlinson. "Unsupervised One-shot Learning of Both Specific Instances and Generalised Classes with a Hippocampal Architecture". In: *Australasian Joint Conference on Artificial Intelligence*. 2020.

[3] Dharshan Kumaran, Demis Hassabis, and James L. McClelland. "What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated". In: *Trends in Cognitive Sciences* 20.7 (July 2016), pp. 512–534.

[4] Brenden M. Lake, Ruslan Salakhutdinov, and Joshua B. Tenenbaum. "Human-level concept learning through probabilistic program induction". In: *Science* 350.6266 (2015), pp. 1332–1338.

[5] Brenden M. Lake et al. "Building machines that learn and think like people". In: *Behavioral and Brain Sciences* 40.2012 (2017), pp. 1–58. arXiv: 1604.00289.

[6] German I Parisi et al. "Lifelong learning of spatiotemporal representations with dual-memory recurrent self-organization". In: *Frontiers in neurorobotics* 12 (2018), p. 78.

[7] Hanul Shin et al. "Continual learning with deep generative replay". In: *Advances in Neural Information Processing Systems*. 2017, pp. 2990–2999.