Learning to match Natural Scenes using Biologically Plausible Deep Reinforcement Learning

Lieke J. Ceton[1]*, Sander M. Bohté[2], Pieter R. Roelfsema[1]

[1]Dept. of Vision & Cognition, Netherlands Institute for Neuroscience, KNAW, Amsterdam
[2]Machine Learning Group, Centrum Wiskunde & Informatica, Amsterdam
*l.ceton@nin.knaw.nl

NETHERLANDS INSTITUTE FOR NEUROSCIENCE
Master the mind

CWI
Centrum Wiskunde & Informatica

Comparing sensory and memory representations is crucial for intelligent behaviour. Contemporary models perform comparison using a hard-wired mechanism[1-3]

We use recent work on self-organizing comparators[4], to demonstrate that a layer with anti-Hebbian plasticity can be learned embedded in a deep reinforcement learning (DRL) agent[5] to enable comparisons.

Two learning phases

1. **Motor babbling (MB)**

   The agent randomly explores its environment[6]

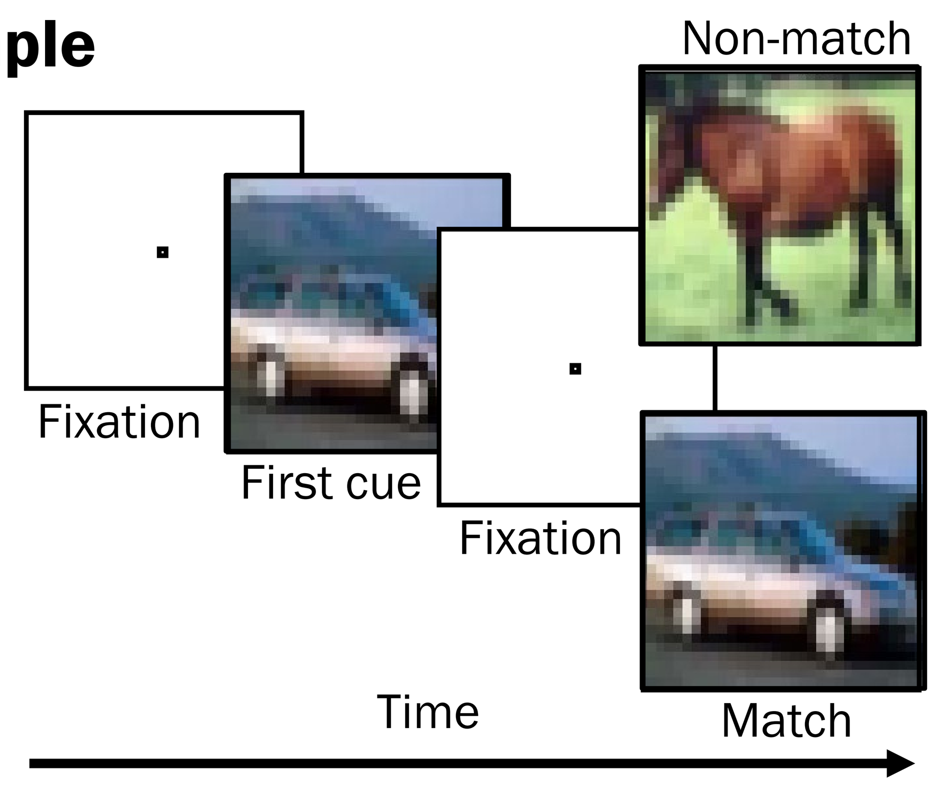
   An unsupervised anti-Hebbian learning rule trains the weights that connect I and S to the matching layer.

2. **Reward-driven learning**

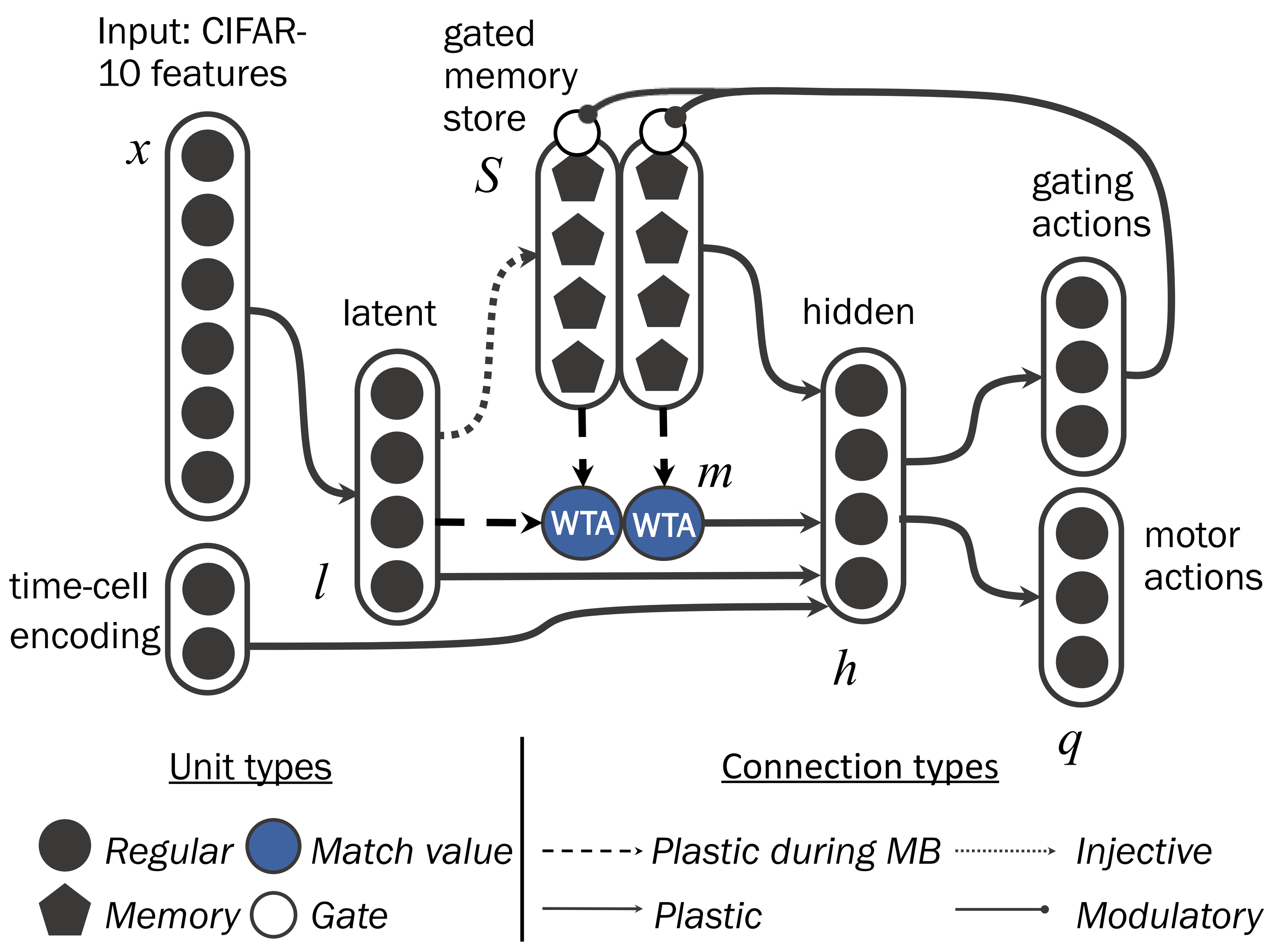
   The biologically plausible learning rule BrainProp[7] trains the network to perform matching tasks in a reinforcement learning environment.

Task = Delayed Match-to-Sample

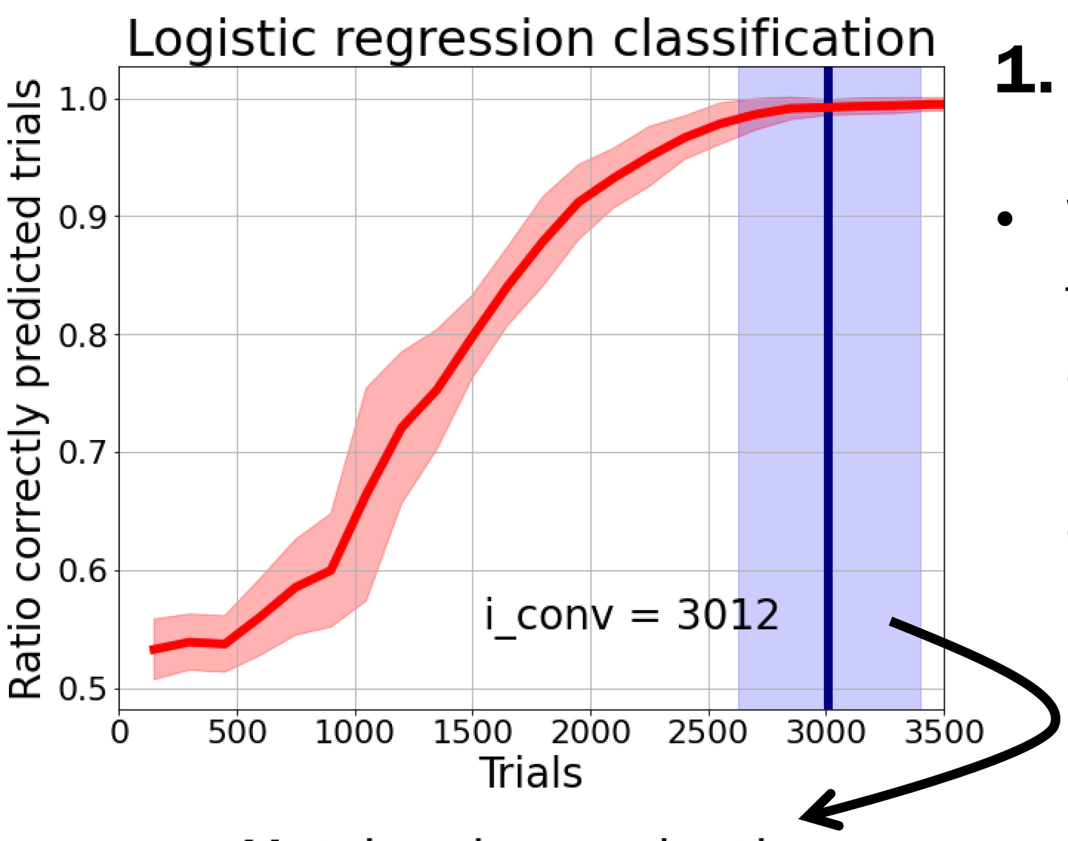
Subjects are shown a natural scene, hold it in memory during a delay and then match it to a second shown image, giving one response for a match and another one for mismatch.



Agent = Deep Reinforcement Learning architecture with learned matching layers and gated memory

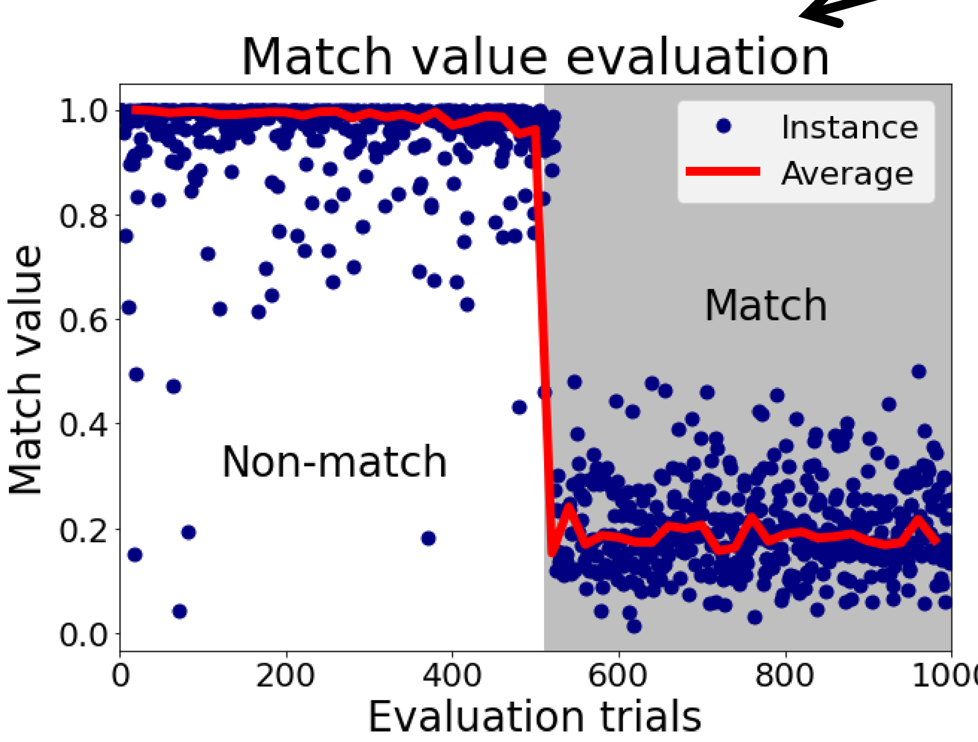


- A deep SARSA RL agent with input layer x, hidden layers (l, h) and output layer q that gates the persistently active working memory stores (S)
- Each cell in l is coupled to a memory cell in S
- Lateral inhibition is modelled in m with a winner-takes-all (WTA) mechanism, producing a scalar match value at each timestep[5]

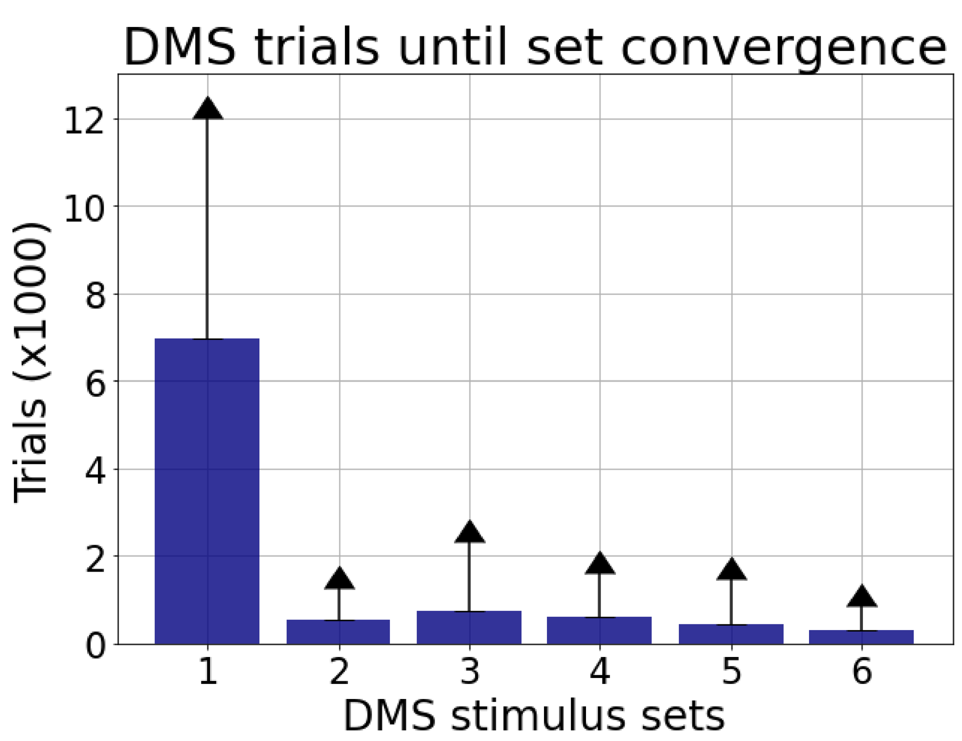


1. Matching performance

- Within ~3000 trials (n=50) the match layers self-organise to correctly predict match/non-match >99% of the last 250 trials



- Anti-Hebbian learning causes low match values, and high non-match values[5]
- Random sampling and the maintenance of half of the inputs in S leads to a general matching function



2. DMS task performance

- Input stimuli are activity vectors of a fully connected layer in a CNN with 80% training acc. on CIFAR-10 images[8]
- Learning is step-wise, switching stimulus sets when 85 of the last 100 trials are performed correctly (Example in red, dashed line is set switch)
- Task is fully learned in ~8000 trials, and speeds up after the first level (average over 17/25 converged agents)

A biologically plausible matching mechanism in a DRL agent

- Plasticity during early developmental stages allows for the fundamental capacity to match
- Later phases can use a reinforcement learning scheme to learn advanced memory-guided behaviours that rely on these matching circuits
- The deep architecture is learned fully locally and can match high-dimensional natural scenes

References: [1] Miller, E. K., Erickson, C. A., & Desimone, R. (1996). *Journal of Neuroscience, 16*(16), 5154–5167. [2] Engel, T. A., & Wang, X. J. (2011). *Journal of Neuroscience*, 31(19), 6982–6996. [3] Zagoruyko, S., & Komodakis, N. (2015). *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 4353-4361. [4] Ludueña, G. A., & Gros, C. (2013). *Neural Computation*, 25(4), 1006–1028. [5] Kruijne, W., Bohte, S. M., Roelfsema, P. R., & Olivers, C. (2021). *Neural computation*, 33(1), 1–40. [6] Piaget, J. (1952). The origins of intelligence in children (Second edi). International Universities Press. [7] Pozzi, I., Bohté, S. M., & Roelfsema, P. R. (2020). *Advances in Neural Information Processing Systems 33* (NeurIPS 2020).[8] Caleb Woy, Kaggle.

Human Brain Project