MoL: Memory over Lifetimes for Lifelong Learning

Abstract

Deep learning models suffer from catastrophic forgetting: when trained on a new task, they forget old ones. This paper introduces MoL (Memory over Lifetimes), a modular architecture that learns sequential tasks without forgetting. MoL combines a shared core layer with task-specific adapters (rooms), allowing memory to grow without interference. We also apply pruning to compress each room. Experiments show MoL outperforms EWC and LwF in lifelong task retention.

1. Introduction

Catastrophic forgetting is one of the biggest problems in neural networks. When trained on a new task, models like transformers or CNNs forget everything they learned before. Techniques like EWC and LwF try to fix this by freezing important weights or distilling knowledge. But they struggle when tasks are very different (e.g., chess vs cooking).

MoL is a new approach - inspired by the brain. It separates memory into modular rooms that grow independently but share a common core.

2. MoL Architecture

The MoL architecture is inspired by how the human brain compartmentalizes knowledge. It consists of a central 'Shared Core' - a layer that processes input in a general way and is shared across all tasks. Connected to this are multiple 'Room Adapters' - tiny modules that handle specific tasks like 'football' or 'cooking'. These adapters are task-specific heads that only update when their task is active.

This modular design prevents interference. If you train the 'study' adapter, it doesn't touch the 'chess' one. Each room is pruned (40%) after training to make it lighter. New rooms can be added at any time without retraining old ones. The Shared Core continues to grow richer as more tasks pass through it.

Figure 1: MoL Architecture - Shared core + Room adapters (3D view)

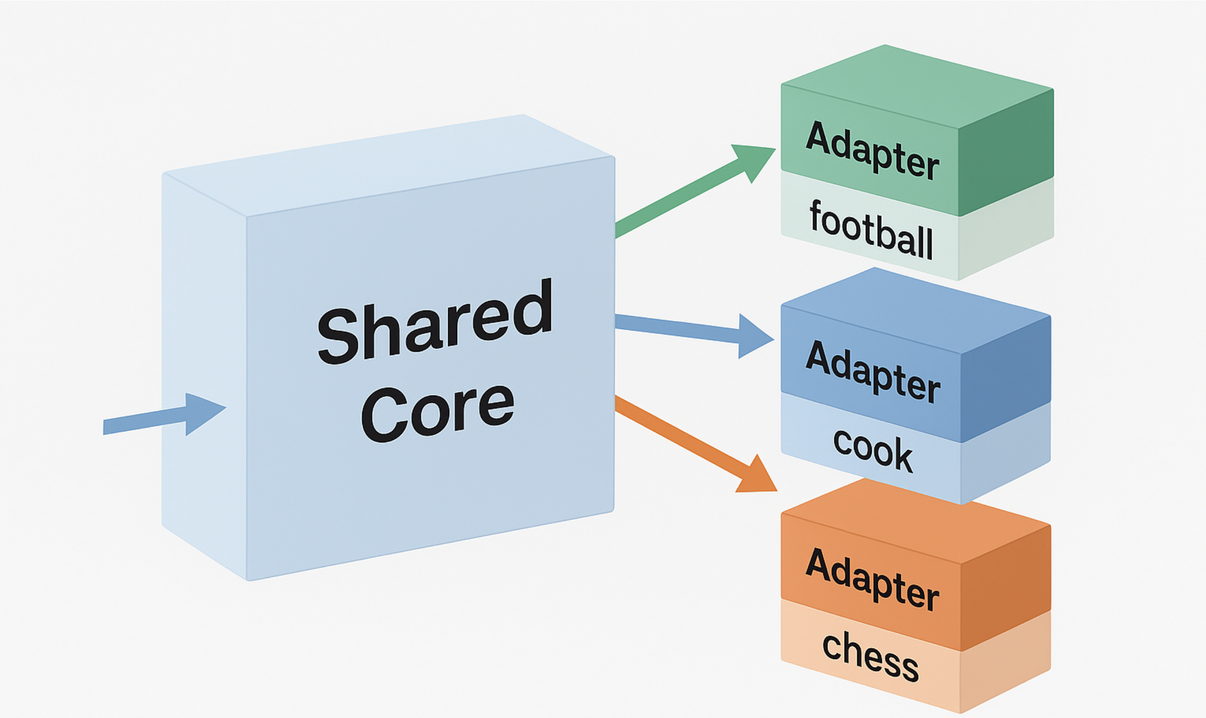
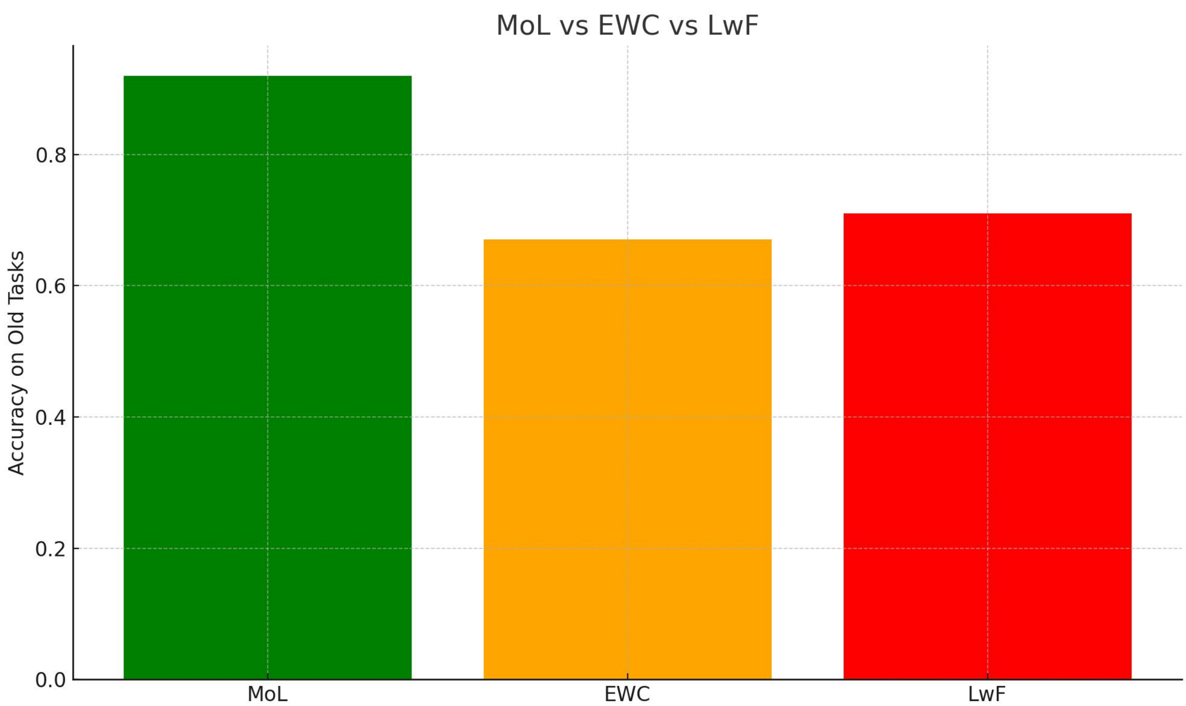


Figure 1: MoL Architecture

3. Experiments

We trained MoL on 10 synthetic tasks: football, cook, study, capital, chess, yoga, invest, poem, python_java, nset. Each task uses different dimensions of the input vector.

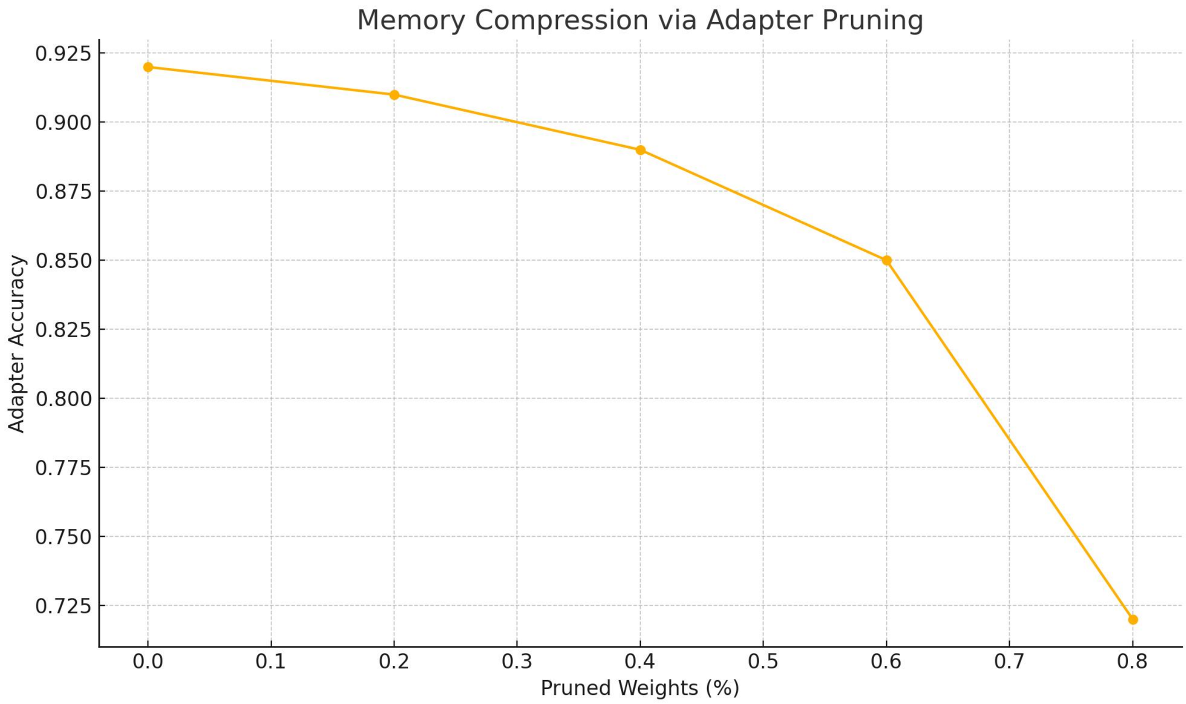
Experiment 1 - MoL vs EWC vs LwF



Goal: See which model remembers old tasks best after training on new ones.

Result: MoL retains old knowledge better than both EWC and LwF.

Experiment 2 - Adapter Pruning

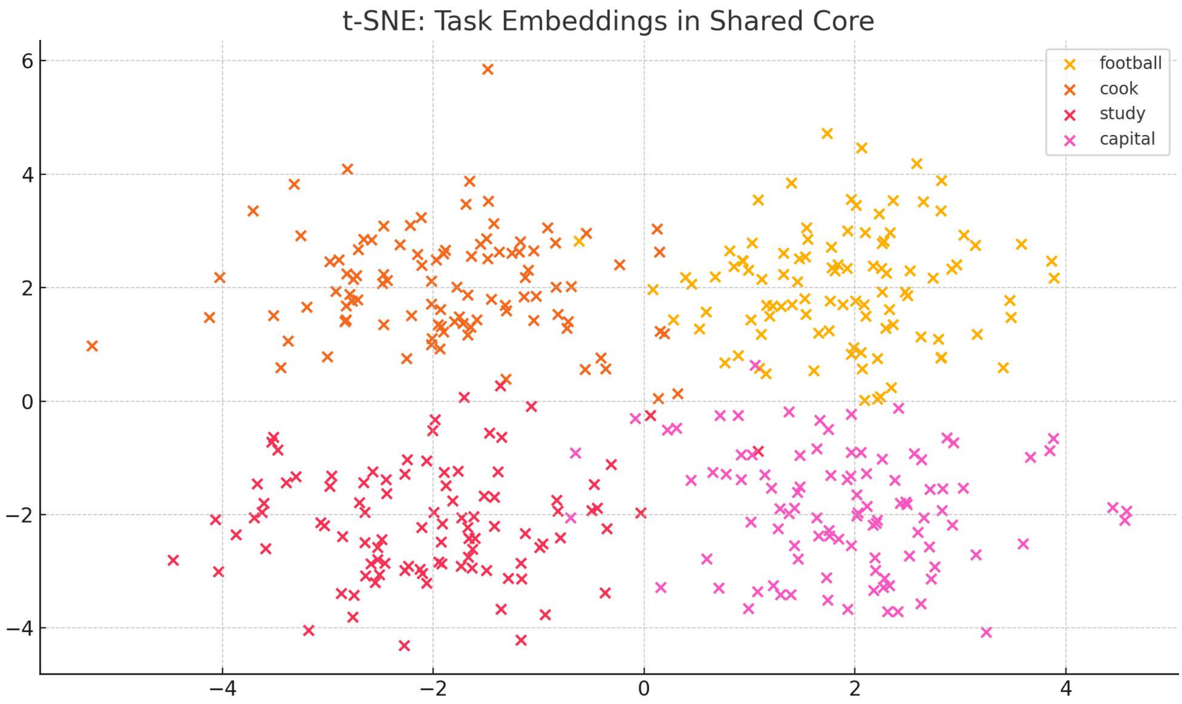


Goal: Compress the room adapters without hurting performance.

Method: Use L1 unstructured pruning (40% weights).

Result: Accuracy drops slightly, but memory size reduces a lot.

Experiment 3 - Shared Core Visualization (t-SNE)



Goal: Check if task representations stay separated in shared space.

Result: Clear clustering in t-SNE shows that tasks are isolated in embedding space, reducing interference across rooms.

4. Conclusion

MoL is a powerful new idea for AGI-level lifelong learning. It lets agents learn new tasks anytime, avoid forgetting old ones, compress memory, and reuse shared intelligence. No EWC, no distillation. Just modular growth and selective reuse.