



Solutions to Catastrophic Forgetting

Praveen K. Pilly
HRL Laboratories, LLC



Specialized Functions: Lifelong Learning Machines (L2M)



Background

What is Catastrophic Forgetting?

- Most deep neural networks (DNN) are first trained and then deployed.
- If unexpected events occur, the DNN will fail unless it is retrained upon encountering new events.
- But the new training could degrade state of the DNN corresponding to the earlier training (catastrophic forgetting).
- New methods have been devised that enables training on new inputs while preserving the training from past inputs.

Overview

How can we overcome Catastrophic Forgetting?

Structural Plasticity

Learn the importance of synapses in solving different tasks and make the important synapses less plastic while learning new tasks

Memory Augmentation

Hold a memory of data from previously learned tasks and interleave old data and new data when learning a new task

Architectural Change

Change the architecture of the network to learn new tasks, which is related to epigenetic neurogenesis

Functional Constraints

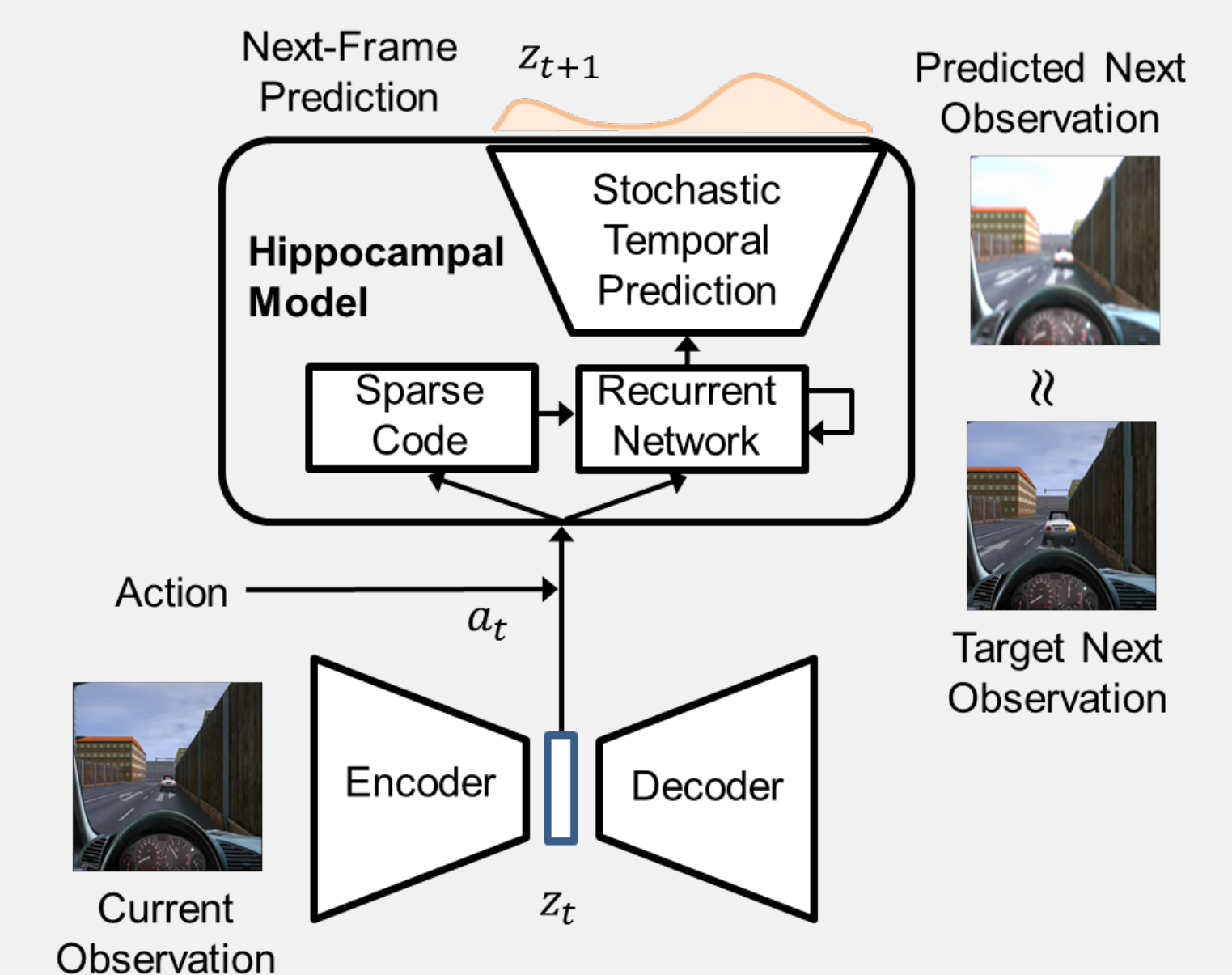
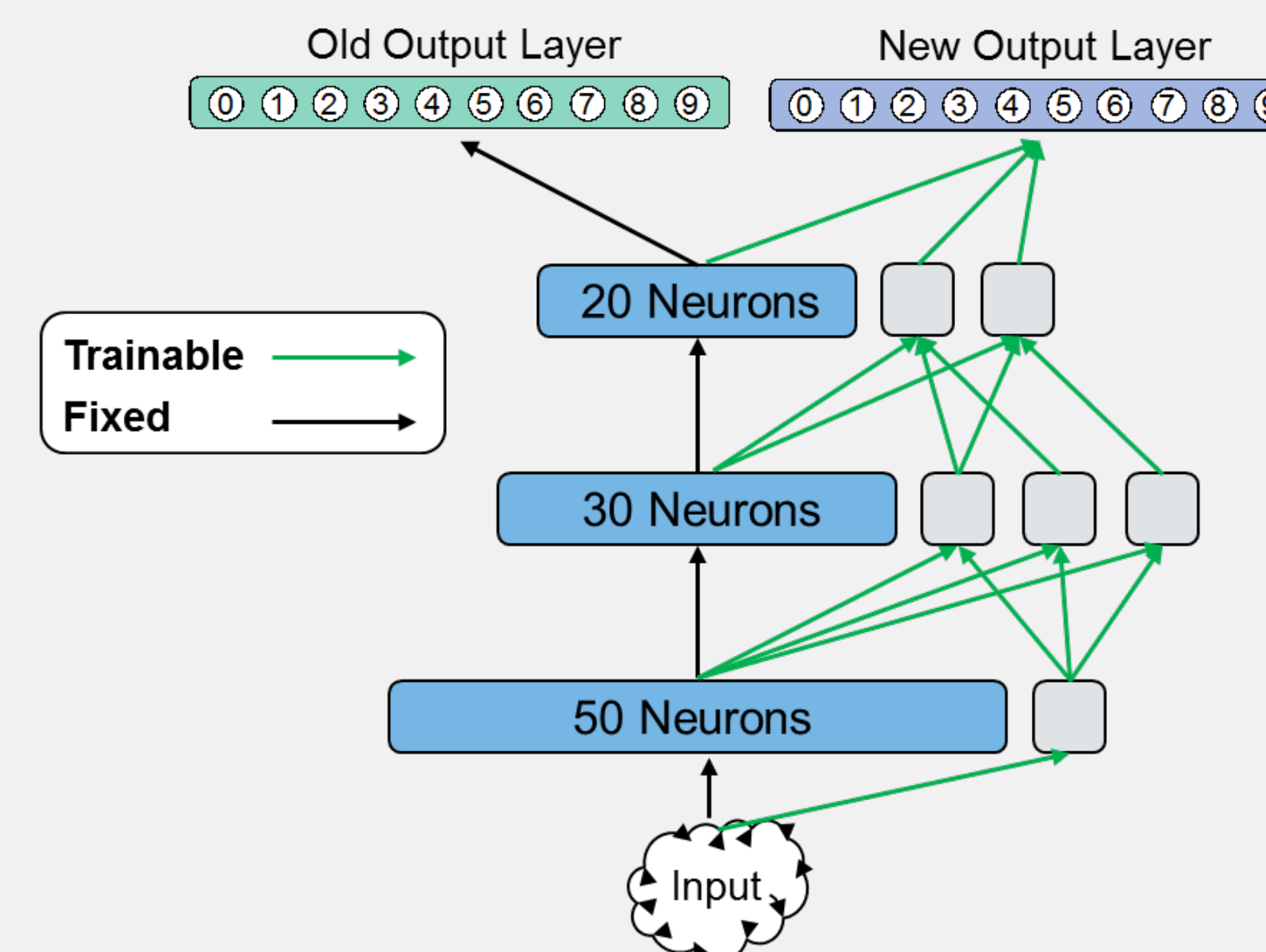
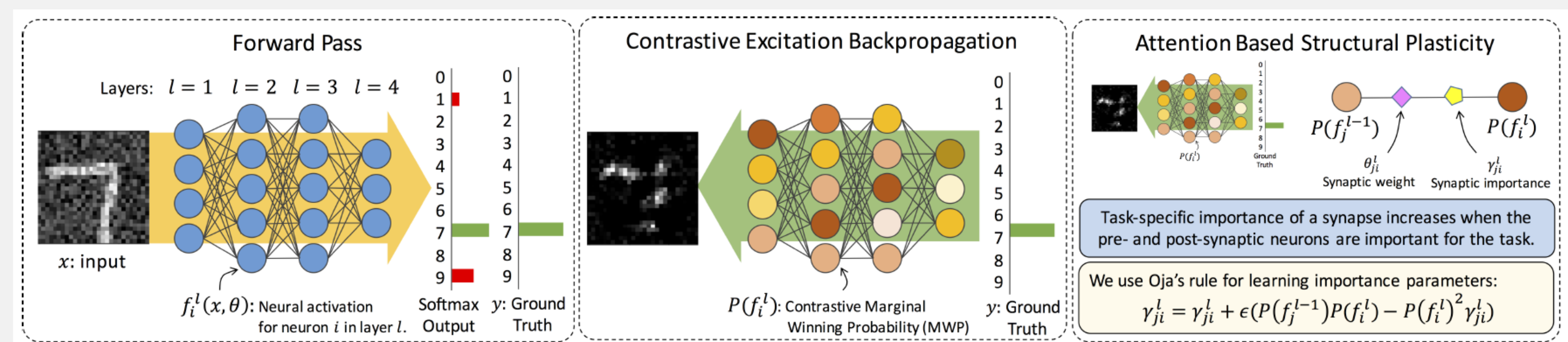
Regularize the objective function to penalize the changes in the input-output function of the neural network

Approach

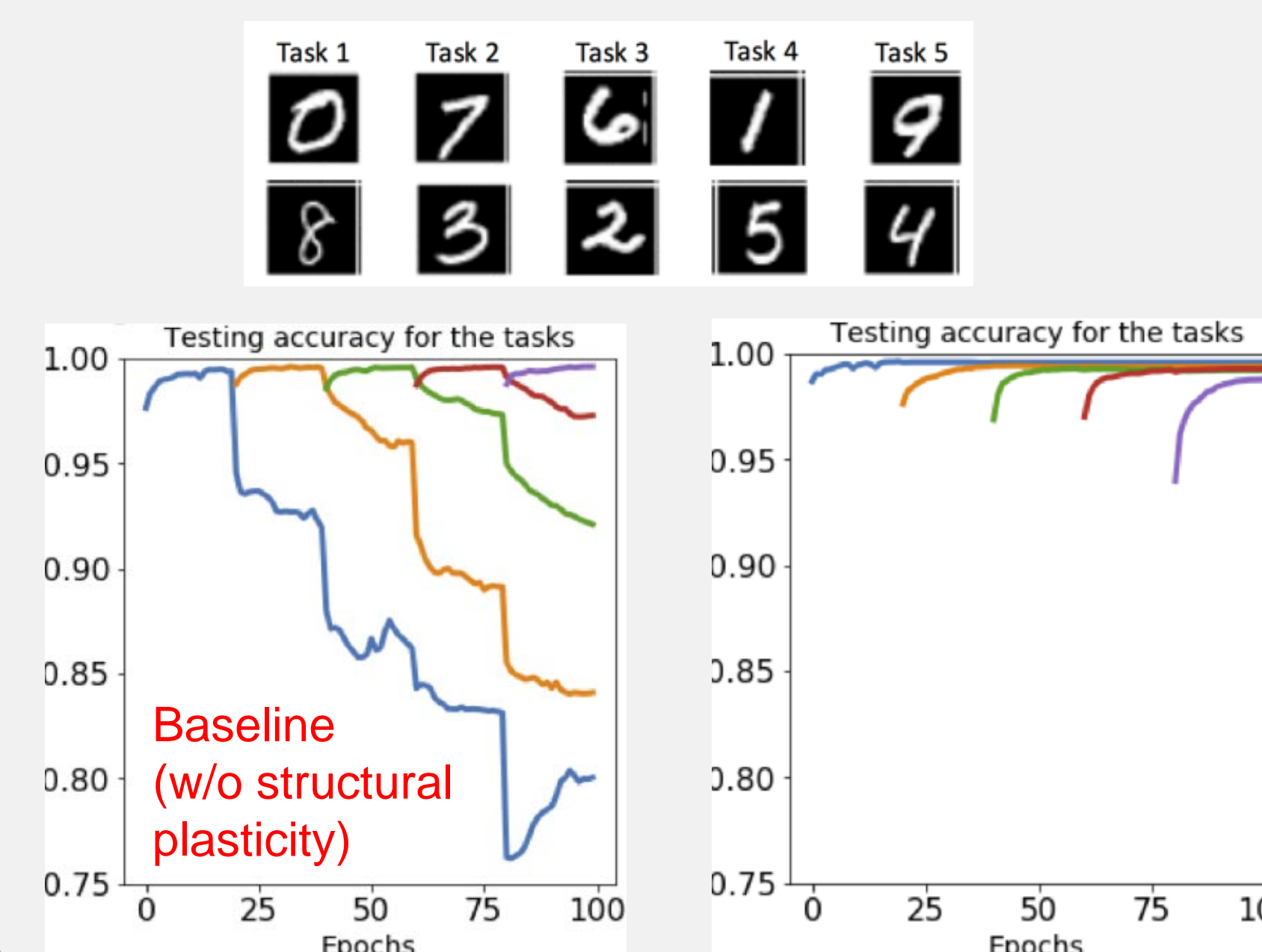
Online: Determine “synaptic importance” for the DNN weights

Offline: Selectively add new neurons to the DNN

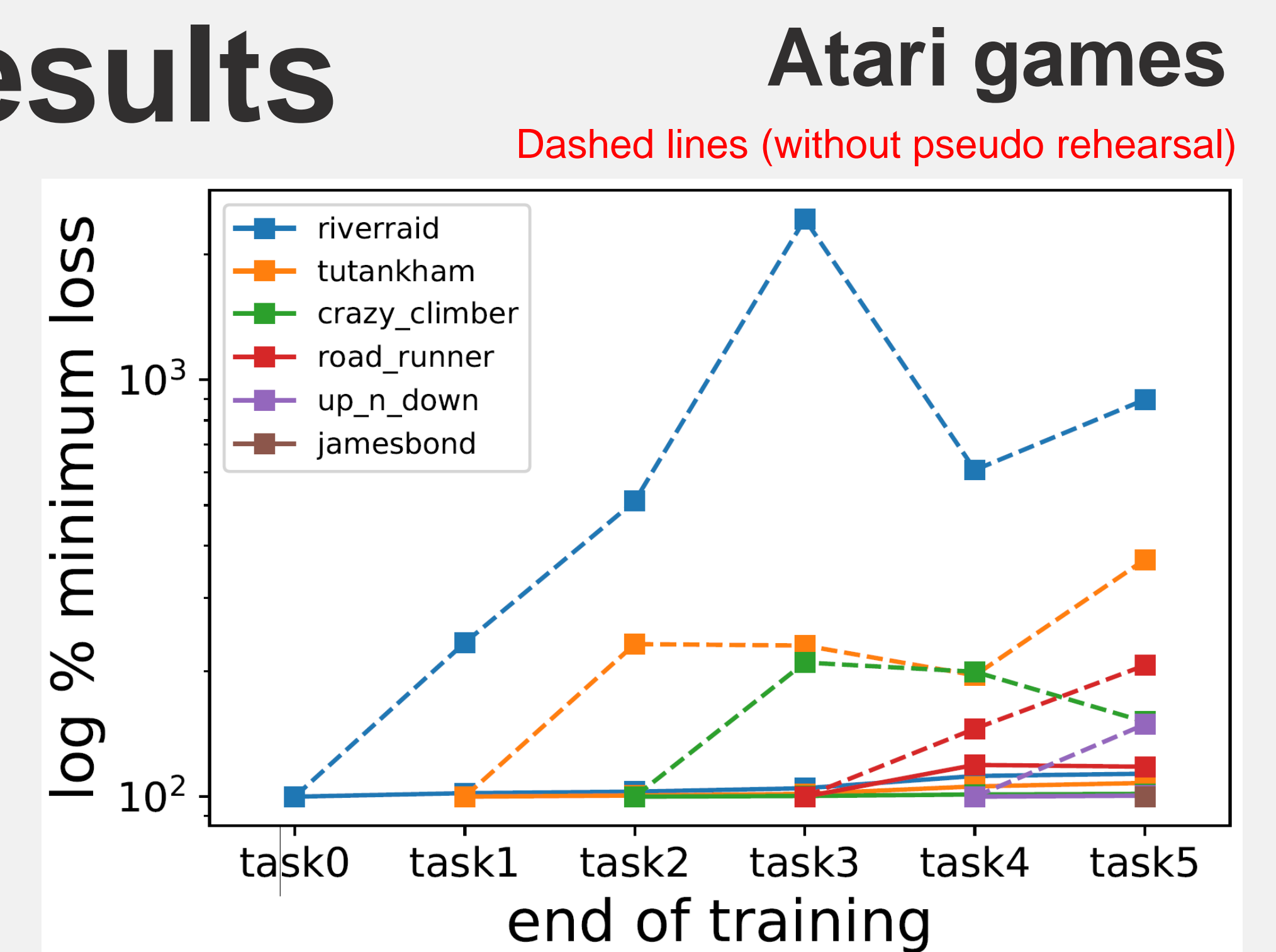
Offline: Perform pseudo-rehearsal for Deep Q networks



Sequential MNIST tasks



Results



Contact: Praveen K. Pilly, HRL, pkpilly@hrl.com

Distribution Statement A – Approved for Public Release, Distribution Unlimited

