ARTIFICIAL NEUROGENESIS FOR ADAPTIVE CONTINUAL LEARNING

Karthik Raghunathan(1), Christian Metzner(1), Laura Kriener(1) and Melika Payvand(1)

(1) Institute of Neuroinformatics, UZH & ETH Zurich

Continual Learning (CL) requires a model to acquire new tasks without forgetting previously learned ones, a problem known as Catastrophic Forgetting (CF). Well-established regularization-based methods, such as Elastic Weight Consolidation (EWC) mitigate CF by anchoring parameters that are important for previously learned tasks [1]. This works well when successive tasks share features, as the same protected weights could support multiple tasks. However, future tasks may be largely orthogonal. In that case, each new task tends to require a different subset of parameters, such that the accumulated regularization progressively constrains a larger fraction of the network and reduces its plasticity, impairing further learning [2]. Moreover, most state-of-the-art CL methods assume a fixed network architecture throughout training, even though the required capacity depends on the unknown number and structure of future tasks. This creates an oracle architecture problem: a network sized for only a few tasks may become overly constrained as tasks accumulate, whereas a network sized for many more tasks may be unnecessarily over-provisioned if the number of future tasks is overestimated.

Biological brains face an analogous challenge and address it, in part, through adult neurogenesis. In the hippocampal dentate gyrus, activity-dependent generation of new neurons is thought to support the separation of overlapping memories and the balance between stability and plasticity [3]. Inspired by this mechanism, we present **GroHess**, a task-agnostic continual learning framework that starts from a compact network and *grows only when needed*. GroHess complements regularisation with targeted neurogenesis, ensuring that each layer retains a set of fresh, unregularised neurons that remain available for learning. To decide when, where, and how many neurons to add, GroHess monitors the representational capacity and effective plasticity of each layer during training.

We evaluate GroHess across a spectrum of task similarity on MNIST-based benchmarks, namely Split-MNIST and Rotated-MNIST for shared-feature setting and Permuted MNIST for orthogonal feature setting. Across all settings, GroHess achieves competitive accuracy relative to oracle-provisioned static baselines while using fewer parameters as shown in fig. 1. Moreover, the final architectures are interpretable, i.e. orthogonal tasks predominantly expands the feature extractor layer (first hidden layer), whereas shared-feature tasks predominantly expand the feature combining layer (second hidden layer). These results suggest that principled neurogenesis can decouple continual learning performance from prior assumptions about the number and structure of future tasks and offer a biologically motivated route toward adaptive lifelong learning systems.

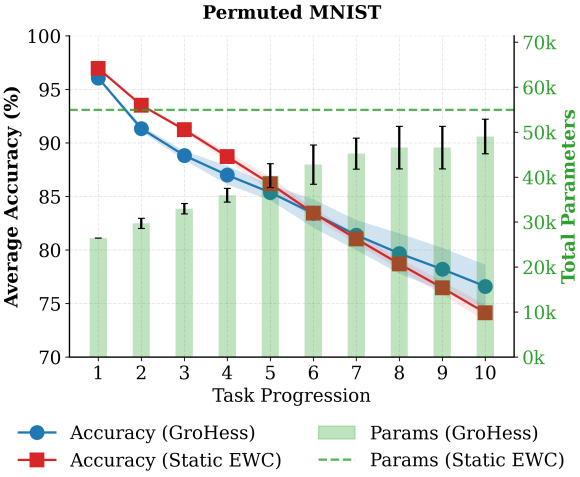

Figure 1: Average test accuracy and total parameter count across ten sequential Permuted MNIST tasks. GroHess starts from a compact network and grows on-demand, closely tracking the accuracy of an oracle-provisioned static EWC baseline while using substantially fewer parameters throughout early training.

[1] J. Kirkpatrick et al., PNAS, 114, 13, 2017.

[2] J. Schwarz et al., PMLR, 4528-4537, 2018.

[3] B. Aimone et al., Neuron, 61, 2, 187-202, 2009.