Preserving Plasticity in Continual Learning with Adaptive Linearity Injection

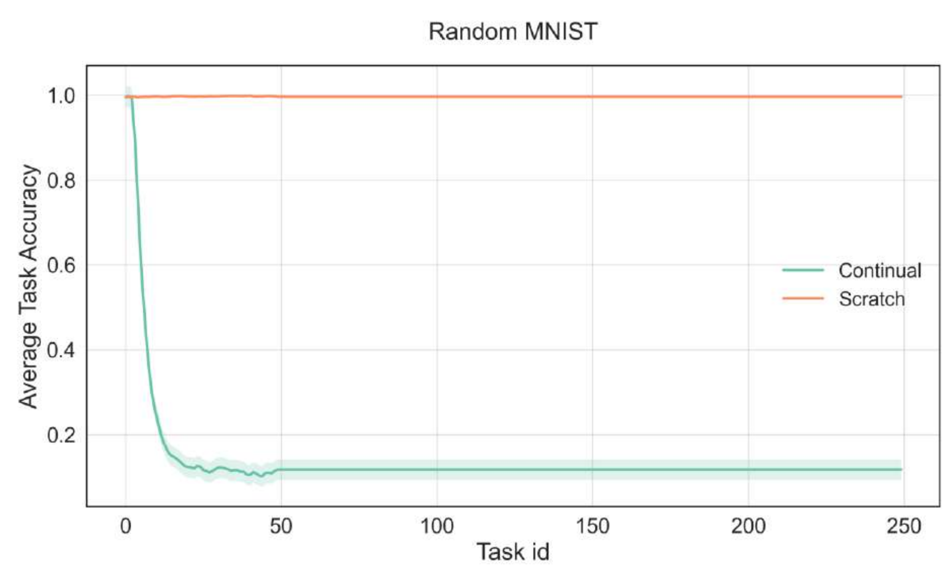
Seyed Roozbeh Razavi Rohani*[1], Khashayar Khajavi*[1], Wesley Chung[2,3], Mo Chen[1] and Sharan Vaswani[1]

1 Simon Fraser University 2 McGill University 3 Mila

Introduction

Plasticity: The capacity of a model to adjust and learn from new data

Loss of plasticity frequently arises in **Continual Learning** and **Reinforcement Learning**, where the underlying data distribution is **non-stationary**. Below is an example of this:



- We address problem of **loss of plasticity** through the lens of activation functions.
- We modify any activation function by adding an **adaptive linear term**.
- We demonstrated the effectiveness of our method on several **supervised** and **reinforcement learning** benchmarks.
- We showed that our method mitigates loss of plasticity for several well-known activation functions.

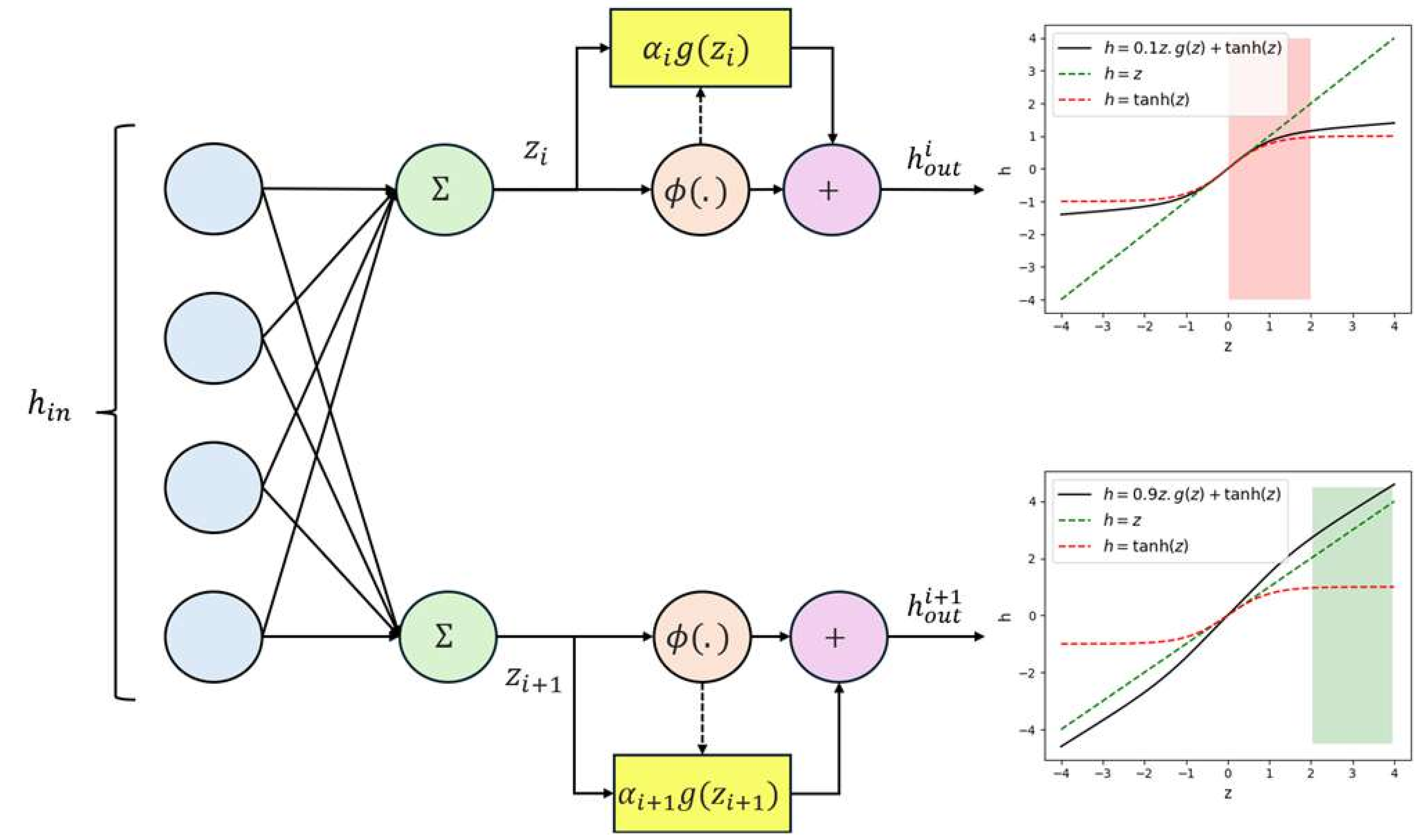
Method

Deep linear networks preserve plasticity by maintaining stable gradients and avoiding saturation but lack **expressivity** and struggle to model complex, nonlinear patterns [1].

We **dynamically** inject **linearity** into activations to preserve plasticity across tasks, enabling the network to balance the **stability** of linear models with the **expressive** power of nonlinear activations. Below is the output of neuron i:

$$\underbrace{f_i(x)}_{\text{Output of Neuron } i} = \underbrace{\phi(x)}_{\text{Preactivation}} + \underbrace{\alpha_i}_{\text{Learnable Parameter Per-neuron}} x\underbrace{[g(x)]_{sg}}_{\text{Stop Gradient}}$$

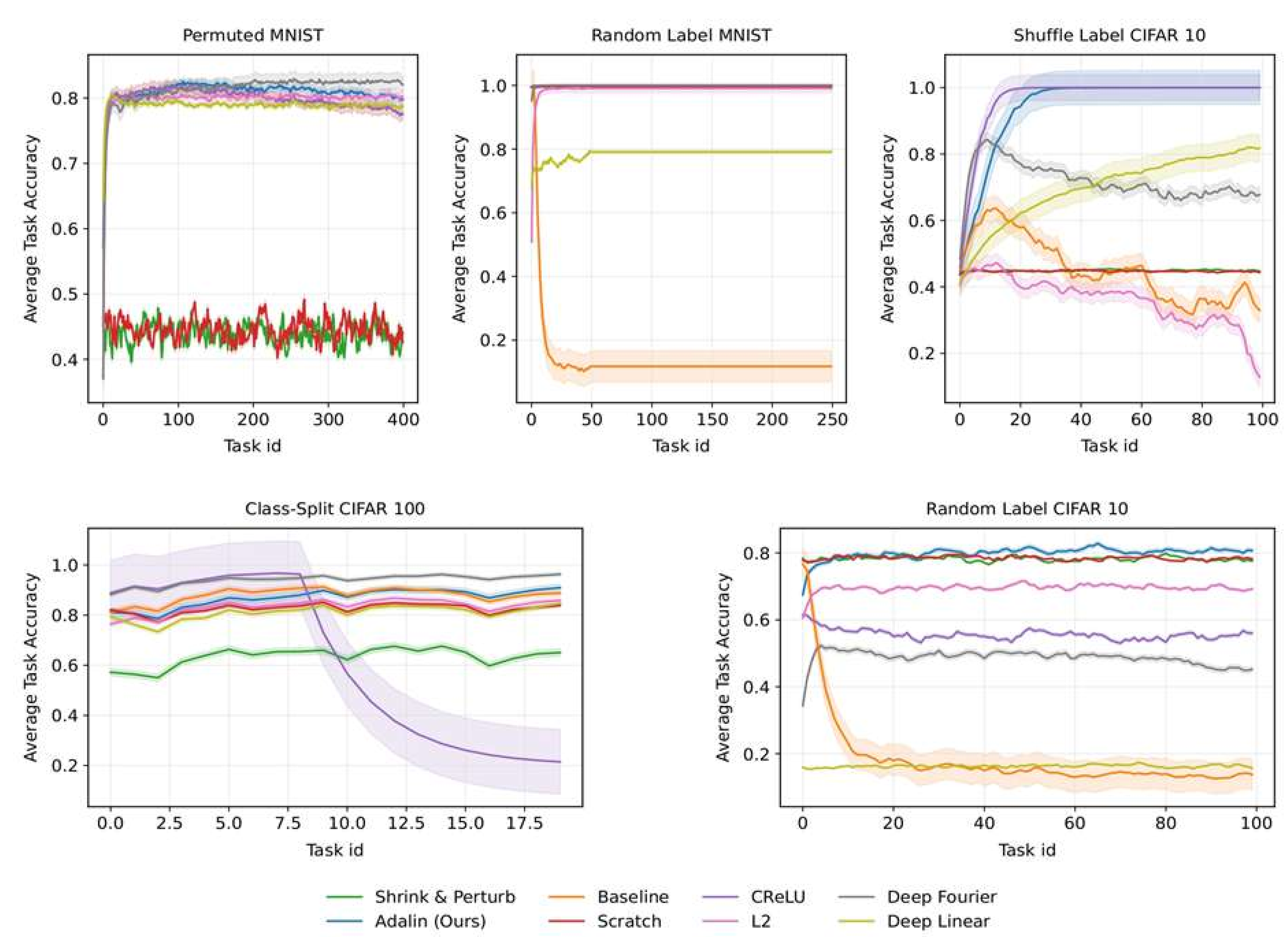
Each neuron learns a linearity injection parameter $\boldsymbol{\alpha_i}$ that adjusts how much linearity is added based on how saturated the neuron is.



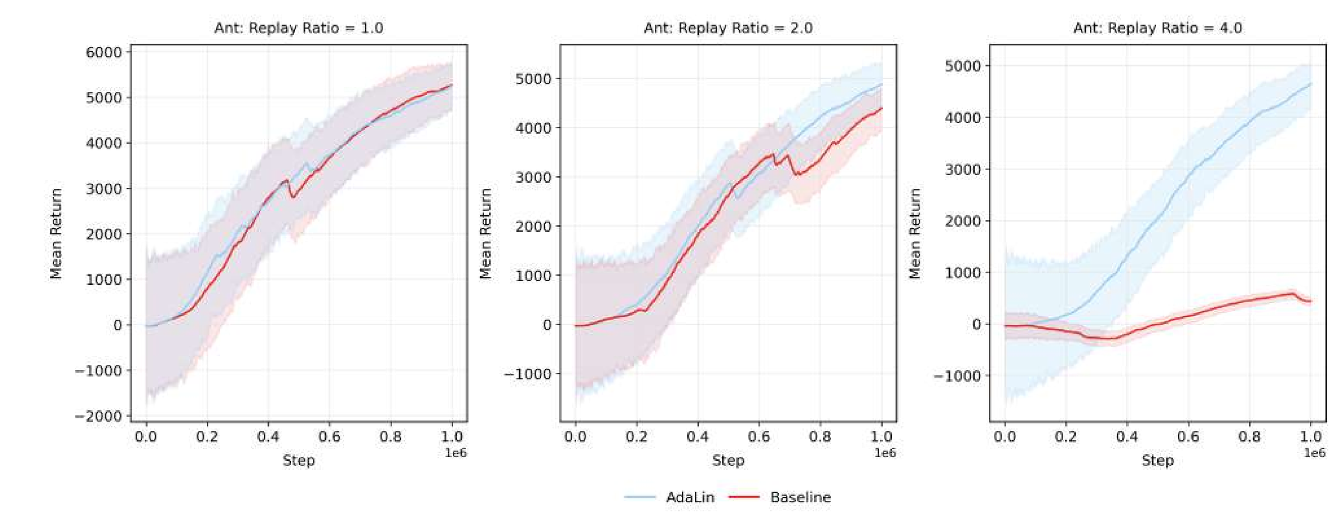
Where $g(x) = \cos\left(\frac{\pi}{2}.\frac{\phi'(x)}{L}\right)$ is a **gating function** that controls how much linearity is injected based on the **saturation** of $\phi(x)$.

Experiments

We tested our method on Class-incremental (Class-split CIFAR100), Domain-incremental (Permuted MNIST), Memorization (Random Label MNIST/CIFAR10), and Concept Drift (Shuffle CIFAR10) benchmarks. AdaLin consistently mitigates the loss of plasticity.

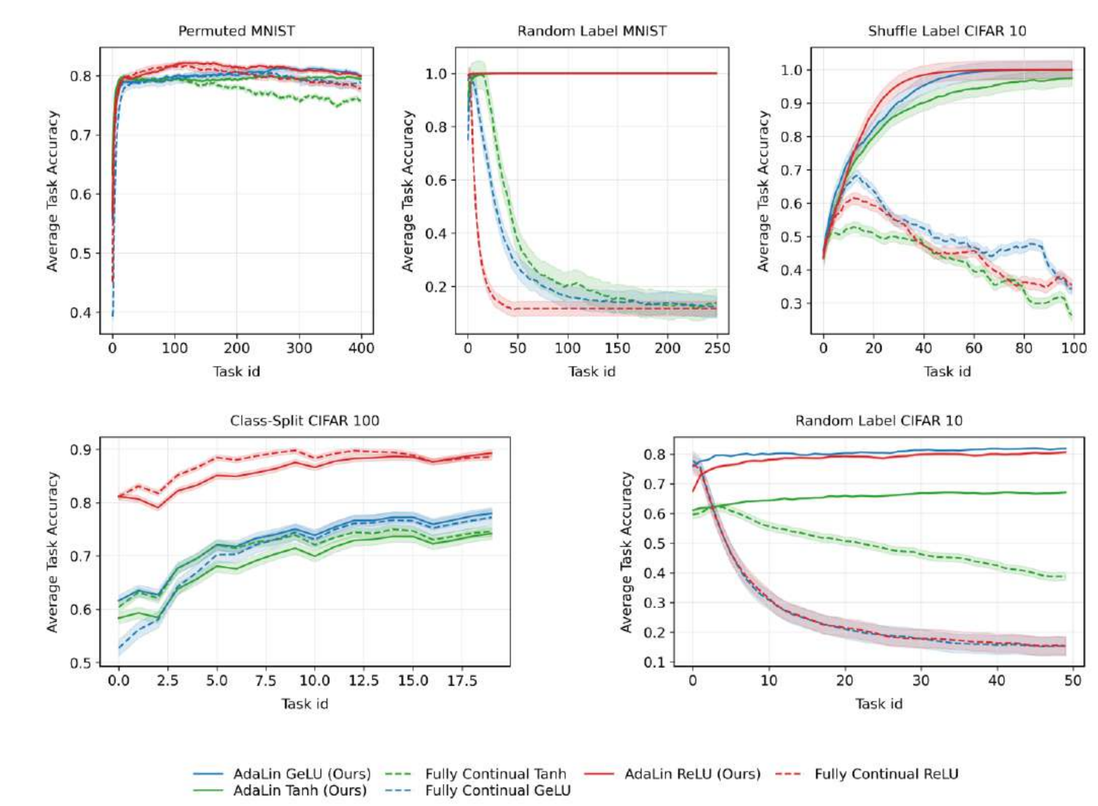


We tested our method on several MuJoCo environments under different replay ratio regimes. The results show AdaLin helps in RL with larger replay ratios.

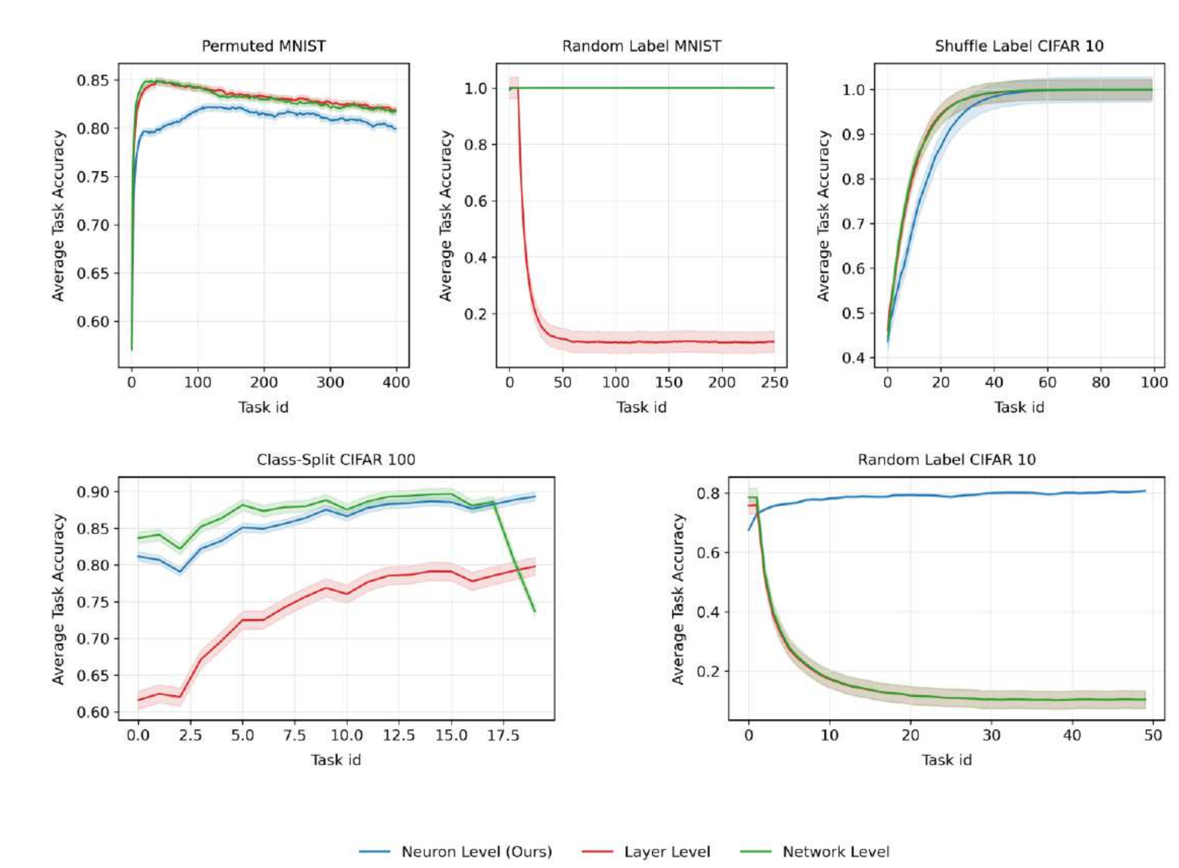


Ablation Study

AdaLin can be applied to different activation functions:



The learnable parameter should be per-neuron:



[1] Lewandowski et al. "Plastic learning with deep fourier features." ICLR 2025.