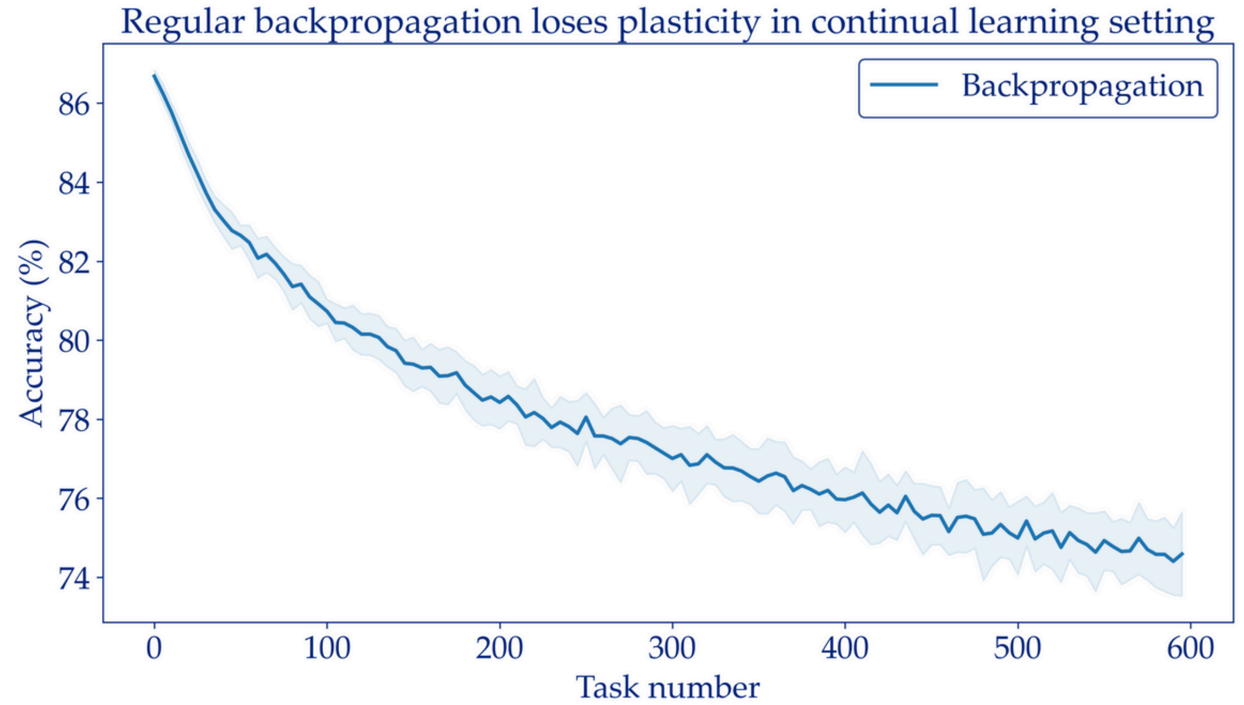
Exploring Alternatives to Full Neuron Reset for Maintaining Plasticity in Continual Backpropagation

Author: Urtė Urbonavičiūtė (uurbonaviciute@tudelft.nl), Supervisor: Laurens Engwegen, Responsible Professor: Wendelin Böhmer

Introduction

Standard deep learning models are limited in real-world scenarios as their fixed weights prevent ongoing learning. Continual learning addresses this by allowing models to adapt over time, but often suffers from reduced plasticity – a gradual decline in the ability to acquire new knowledge. Recent methods like Continual Backpropagation (CBP) aim to preserve plasticity by reinitializing neurons that are estimated to contribute least to the output. While this helps restore learning ability, full reinitialization risks discarding useful knowledge and slowing down learning.

Regular backpropagation loses plasticity in continual learning setting



This research investigates three less disruptive alternatives:

1. Injecting Gaussian noise into low-utility neurons
2. Reinitializing outgoing weights using Kaiming uniform distribution
3. Rescaling weights to restore original variance

Research Question

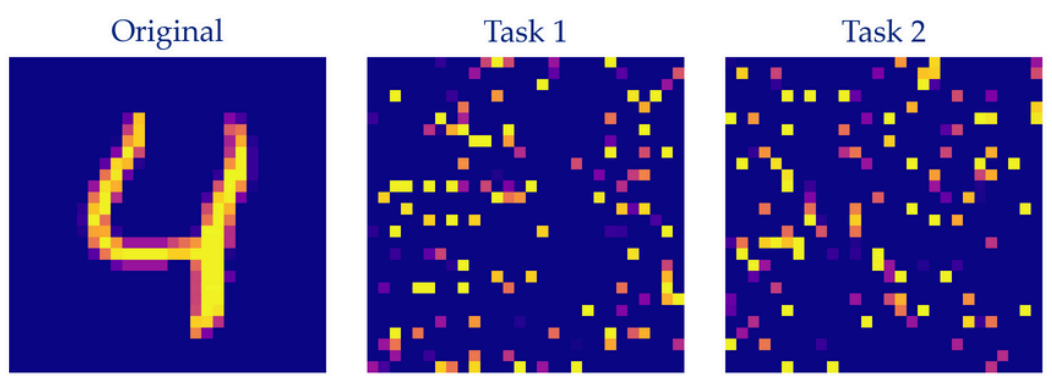
What is the effect of alternative methods to neuron reinitialization on maintaining plasticity in continual learning?

References:
Shibhansh Dohare, J. Fernando Hernandez-Garcia, Qingfeng Lan, Parash Rahman, A. Ruapm Mahmood, and Richard S. Sutton. Loss of plasticity in deep continual learning. Nature, 632:768—774, 2024.
Jordan T. Ash and Ryan P. Adams. On warm-starting neural network training, 2020.
Lukas Niehaus, Ulf Krumnack, and Gunther Heidemann. Weight rescaling: Applying initialization strategies during training, 06 2024.

Benchmark

Permuted MNIST: A sequence of 600 image classification tasks, each applying a different fixed random pixel permutation. It tests adaptation to abruptly changing input representations across tasks.

Digit '4' permutations in Permuted MNIST



Methodology

All three methods are applied to low-utility neurons, which are detected based on their contribution to learning using a utility metric. Instead of completely resetting these neurons as in CBP, each method modifies them more subtly.

1. **Noise Injection**

Shrinks incoming weights, outgoing weights, and biases by $\lambda = 0.2$ and perturbs them with Gaussian noise.

$$\boldsymbol{\theta}_i^t = \boldsymbol{\lambda} \cdot \boldsymbol{\theta}_i^{t-1} + \boldsymbol{p}_t, \quad \text{where } \boldsymbol{p}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}^2)$$

2. **Kaiming Reinitialization**

Reinitializes outgoing weights from the Kaiming uniform distribution, rather than setting them to zero.

3. **Weight Rescaling**

Rescales weights to match the original variance from initialization:

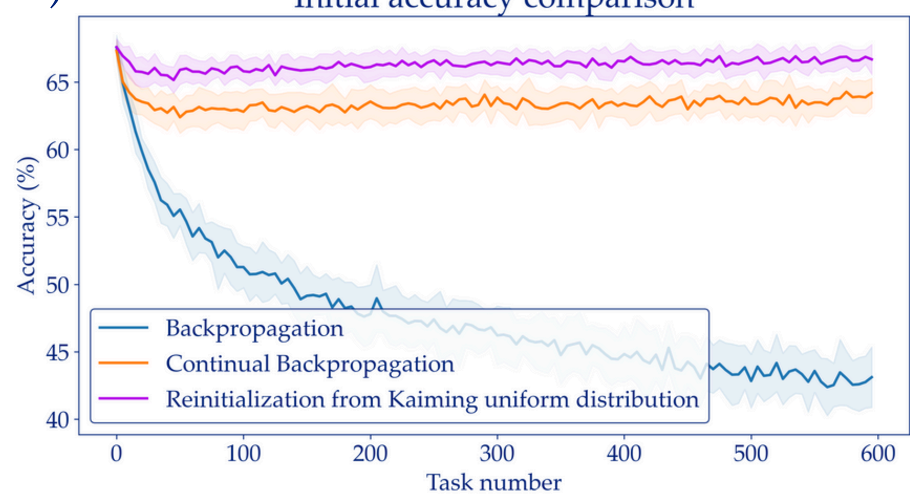
$$\boldsymbol{w}^{(\ell)} \leftarrow \left(\frac{\boldsymbol{w}^{(\ell)} - \mu(\boldsymbol{w}^{(\ell)})}{\sigma(\boldsymbol{w}^{(\ell)})} \right) \cdot \boldsymbol{\sigma}_{\text{init}} + \mu(\boldsymbol{w}^{(\ell)})$$

Evaluation Metrics

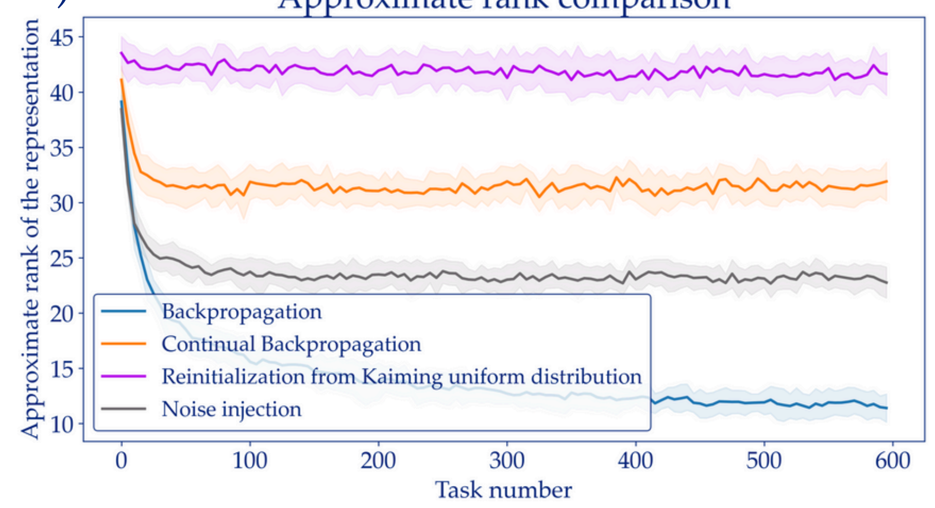
- **Overall Accuracy** – the model's average prediction accuracy across all samples within the task, indicating overall performance.
- **Initial Task Accuracy** – average accuracy of the first 10% of samples in each task, indicating how quickly the model learns.
- **Approximate Rank** – a proxy for the representational capacity and plasticity of each layer.

Results

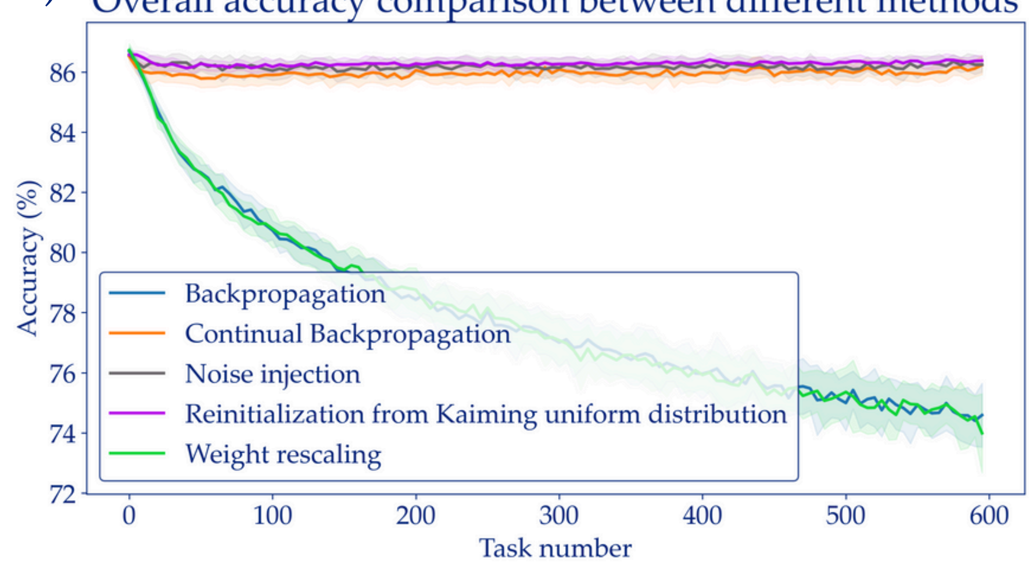
a)



b)



c)



Figures a and b show that Kaiming Reinitialization significantly improves initial accuracy and approximate rank compared to standard CBP.

Figure c presents final overall accuracy for all methods.

Conclusions and Future Work

- Kaiming reinitialization was the most effective method. It improved plasticity and outperformed standard CBP in all metrics. It also enabled faster learning.
- Noise injection performed comparably to CBP, achieving similarly high accuracy while retaining parts of the original weights.
- Weight rescaling showed no significant improvement, likely because it is not disruptive enough to help neurons relearn new patterns.
- In the future, methods could be tested on more complex continual learning datasets to assess scalability and generalization.