Neural network pretraining and its influence on continual learning

Piotr Gręda[1] Kamil Deja[1,2] Tomasz Trzciński [1,2,3]

[1]Warsaw University of Technology [2]IDEAS NCBR [3]Tooploox

#TLDR

- We train models in Continual Learning setup
- We check whether starting from pretrained model can boost a neural network's performance in new tasks and reduce catastrophic forgetting
- We inspect pretrianing methods to determine wich one is the most profitable
- We demonstrate differences on how supervision in pretraining affects the performance of the continually trained model

Continual Learning (CL)

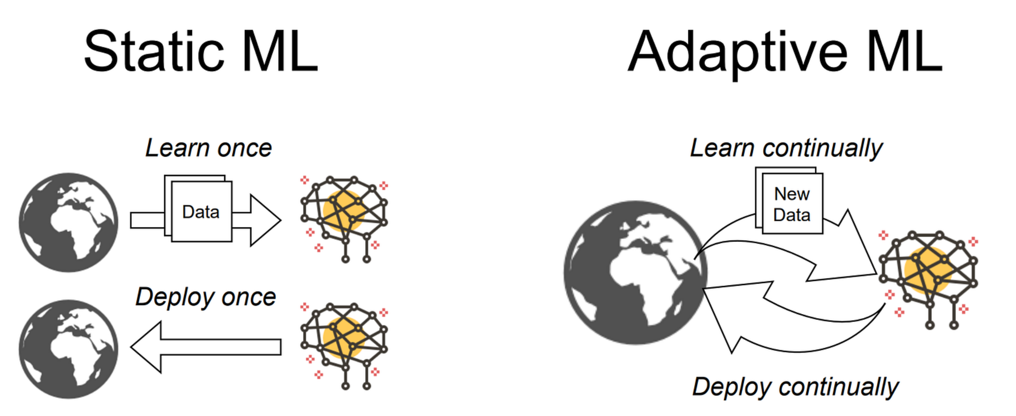

Figure 1. Difference between static and adaptive Machine Learning. Visualisation from: Towards Adaptive AI with Continual Learning

- Current dataset is not the only one - new tasks can be added during training
- Sequential model training while adding new classes without forgetting previously obtained knowledge

Self-supervised Learning (SSL)

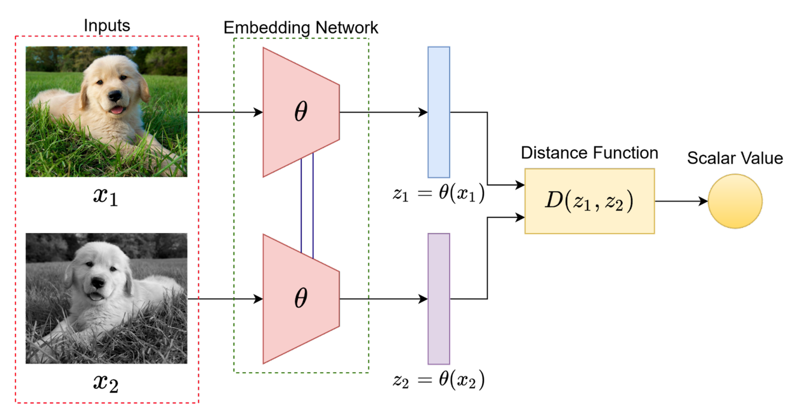

Figure 2. Example of Self Supervised Learning method. Visualisation from: Alessio Lerede - A world without "labels"

- Focuses more on data itself - human made labels are not needed
- Trains model to recognize patterns, similarities, relations between images
- Can be turned into supervised models by adding classification head

Continual Learning with replay

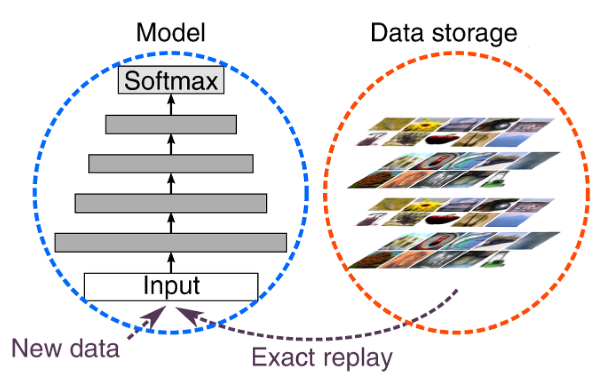

Figure 3. Overview of Replay way of training using buffer. Visualisation from: Ven et al., Brain-inspired replay for continual learning with artificial neural networks

- Replay method uses buffer that stores a history of past experiences sampled randomly
- Size of a buffer can differ, the bigger the buffer the higher the accuracy on previously learnt tasks
- The new images mix with the old ones taken from the buffer, creating a batch.

Pretraining of Neural Networks

- People learn since early age, their wealth of knowledge is vast and therefore they use formerly acquired insight when addressing new issue. We try to use similar methodology to reach better performance in Continual Learning setup.
- Human brain can learn new tasks in a sequential fashion and it differentiates it from neural networks - every new task causes a significant decrease in accuracy of previously learnt ones - this phenomenon is called "catastrophic forgetting".
- We aim to minimize catastrophic forgetting and maximize effectiveness of CL models by using different types of pretraining.

Continual Learning with pre-training

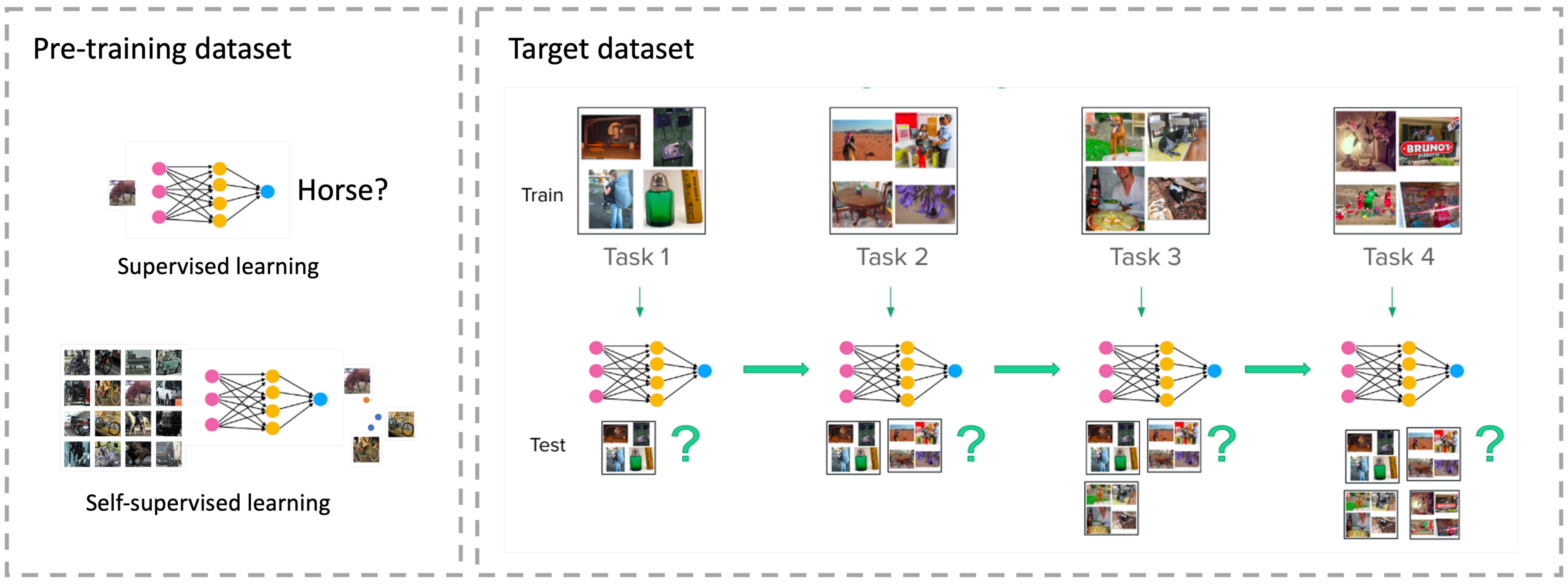

Figure 4. We pre-train models in different ways on source dataset, while evaluating them in continual-learning setup on the target one. Visualisation based on: La-MAML: Look-ahead Meta-Learning for Continual Learning by Gunshi Gupta, Liam Paull

SSL pretraining for online and offline CL

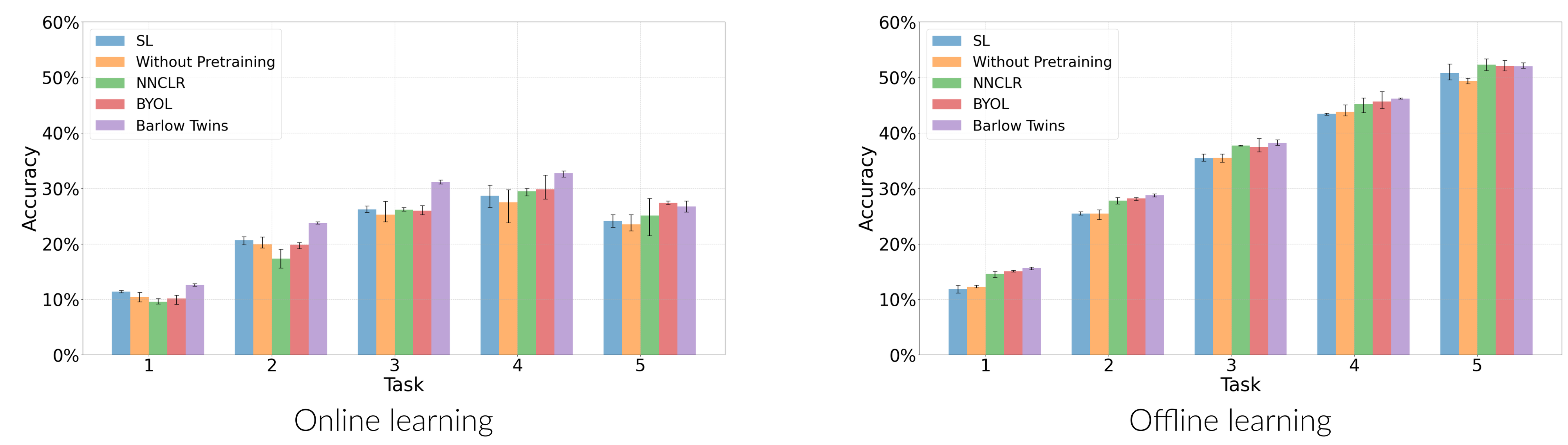

Figure 5. Comparison of different SSL pretraining methods on CIFAR10->CIFAR100 datasets

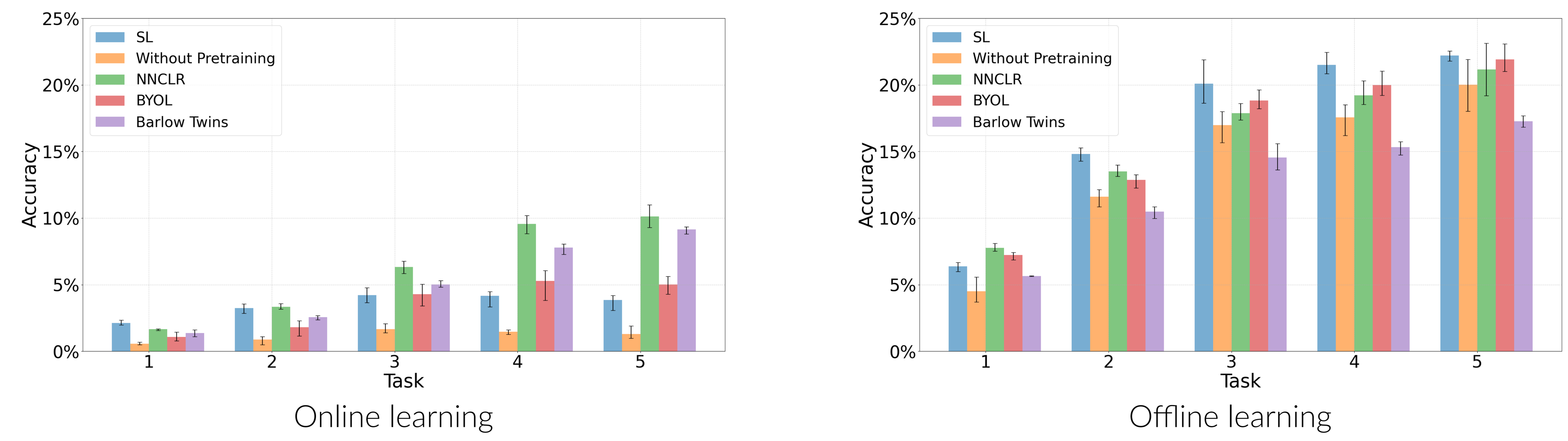

Figure 6. Comparison of different SSL pretraining methods on ImageNet100->CUB 200 datasets

- The decrease in accuracy is attributed to the shift towards a more challenging evaluation dataset.
- The effect of pre-training is more visible in online setting when model has access to the new data samples only once

Pretraining methods

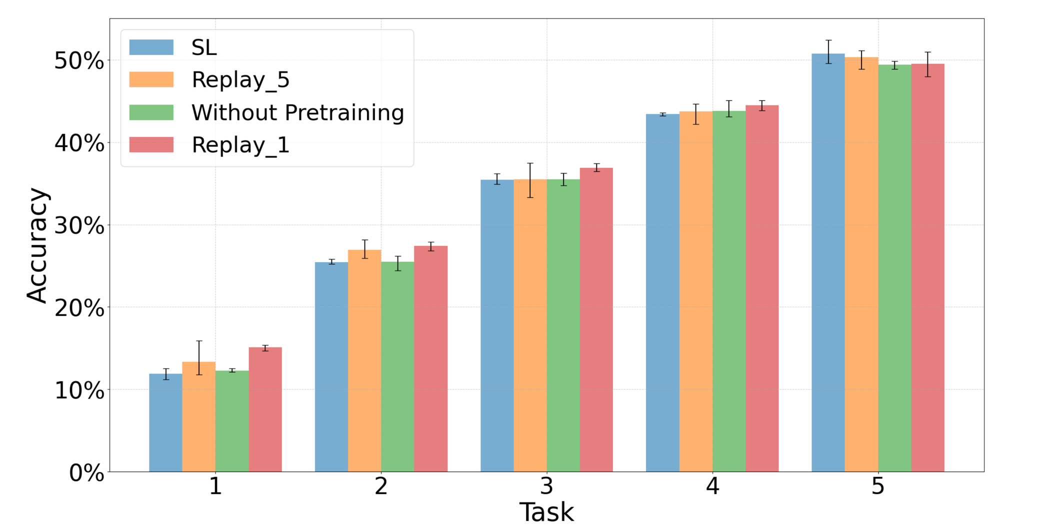

Figure 7. All models were pretrained using written methods on SplitCIFAR10 and later evaluated on SplitCIFAR100 using Continual Learning Replay. Replay_5 means that pretraining on the dataset took place 5 times in a cyclical manner. Replay_1 implies that the pretraining took place only once without any repetiton

- Pretraining the model improves effectiveness in a CL setup around the margin of statistical error
- When applied to more complex dataset differences between them are still negligible

Freezing the backbone

We evaluate the performance of a frozen pre-trained model without fine-tuning on new-target dataset

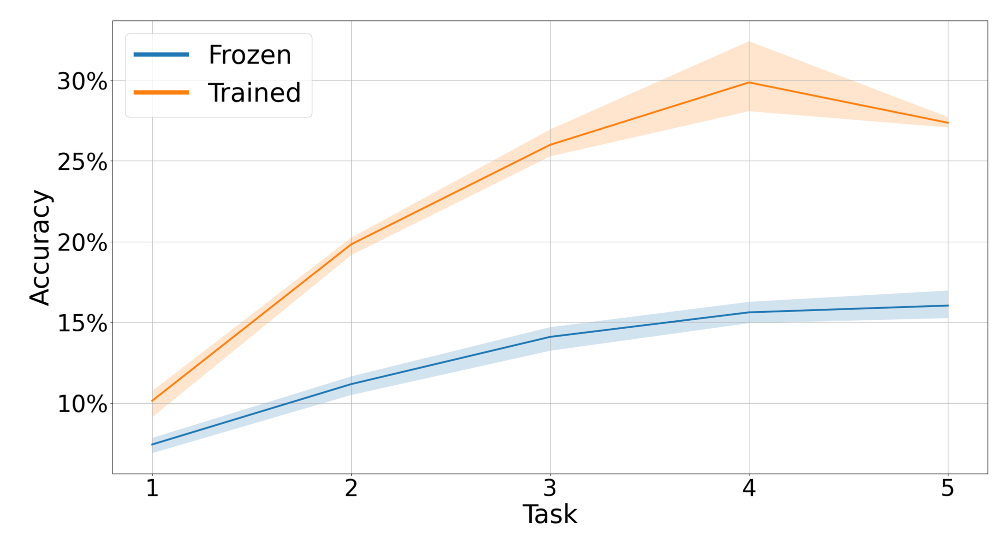

Figure 8. Both models were pretrained on 50 epochs using BYOL

- Freezing model's backbone reduces performance
- This aligns with expectations - CL exhibits high dynamism, and freezing the model does not favor its fine-tuning.

Take-Away Points

- Pretraining has instrumental effect on online continual learning - especially SSL methods make CL models perform better
- For offline continual learning pre-training does not significantly influence the final performance, pretraining on some SSL methods can even decreace final accuracy
- In both scenarios, fine-tuning on target dataset is essential to adapt the model to new data