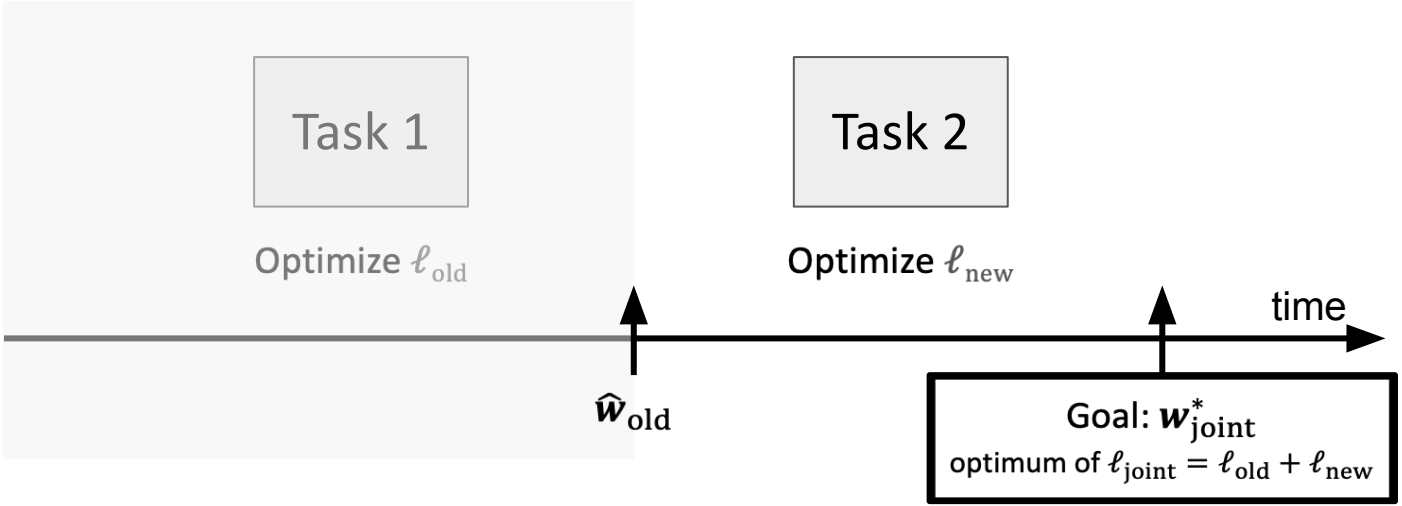
Research Project 2024/2025 Q4

The Stability Gap in Continual Learning with Deep Neural Networks

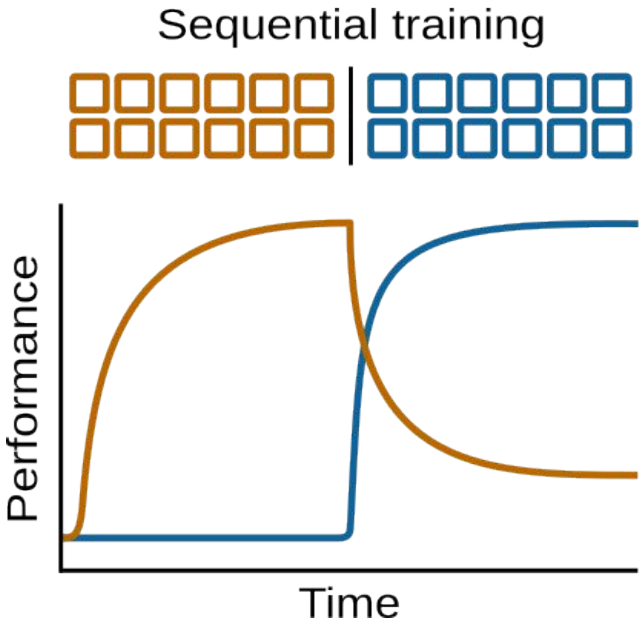
Gido van de Ven (Supervisor),
Tom Viering (Responsible Professor)

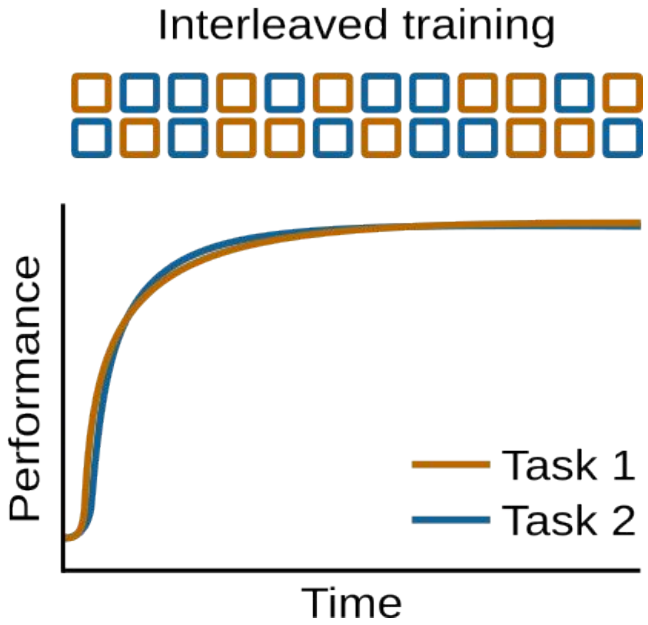
The continual learning problem

→ Optimize the parameters w of a neural network f_w for two tasks that are observed one after the other



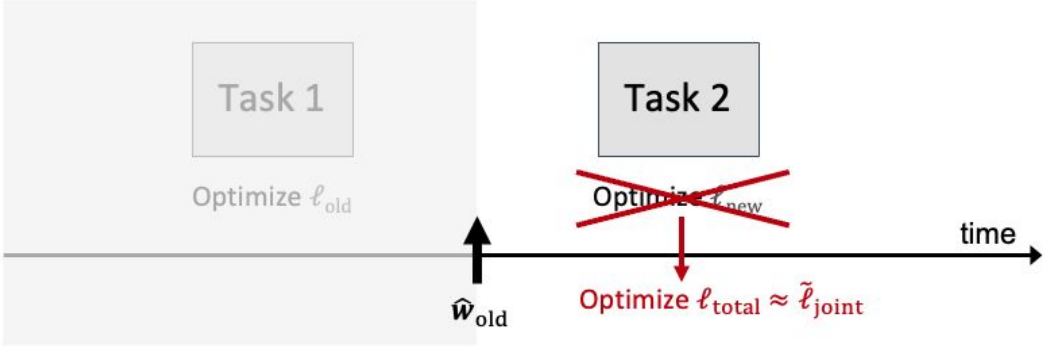
Catastrophic forgetting





Source: van de Ven, Soures et al. (2024, arXiv)

Current approach to continual learning: make changes to the loss



Replay

$$\ell_{\text{total}} = \ell_{\text{new}} + \ell_{\text{replay}} \approx \tilde{\ell}_{\text{old}}$$

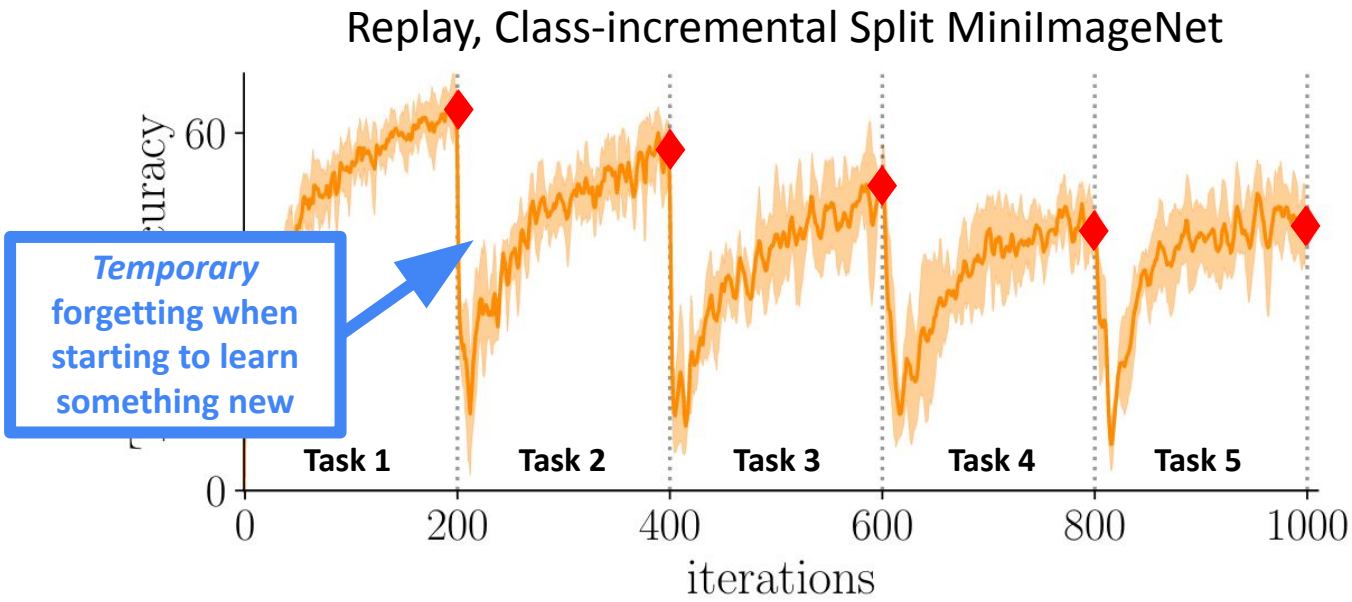
Parameter regularization

$$\ell_{\text{total}} = \ell_{\text{new}} + \|w - \widehat{w}_{\text{old}}\|_{\Sigma} \approx \tilde{\ell}_{\text{old}}$$

Functional regularization

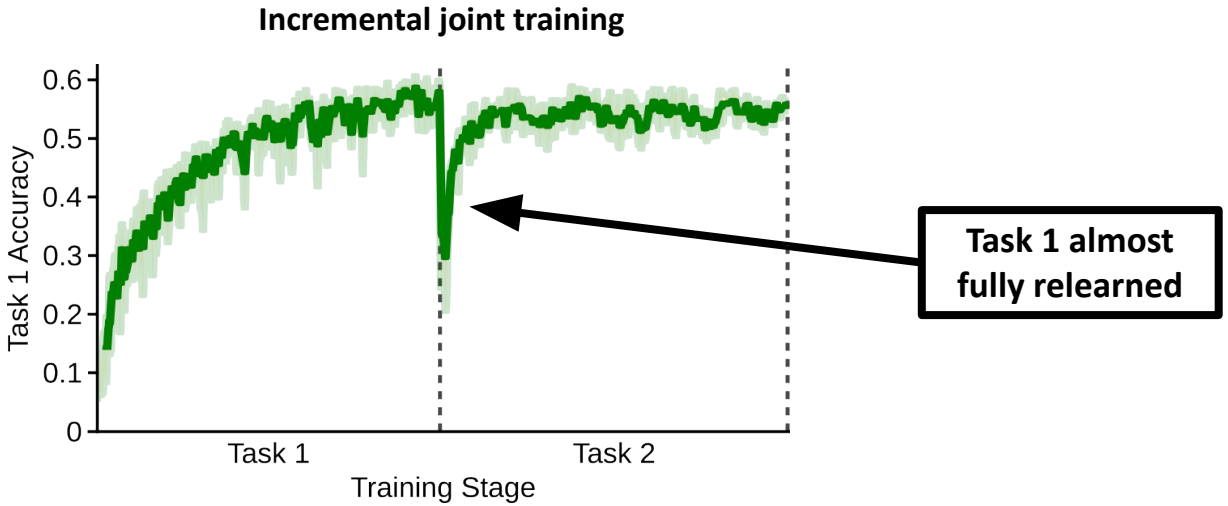
$$\ell_{\text{total}} = \ell_{\text{new}} + \langle f_w, f_{\widehat{w}_{\text{old}}} \rangle_{\mathcal{A}} \approx \tilde{\ell}_{\text{old}}$$

Does replay prevent forgetting?



Source: De Lange et al. (2023, ICLR)

The stability gap occurs even with “perfect” replay!



- Problematic for safety-critical applications
- Seems highly inefficient

Source: Hess et al. (2023, PMLR)

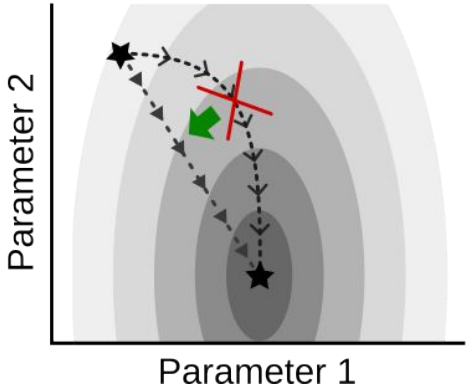
Continual learning needs a new perspective: **optimization**

- To overcome the stability gap, changes must be made not only to *which* loss function is optimized, but also to *how* it is optimized

- Standard optimization routines for deep learning have been developed for the stationary setting

- No guarantees in continual setting, yet widely used

- Fundamental difference between both settings:
  - Stationary → start from random initialization
  - Continual → start from partial solution



Research Questions

(1) How does learning rate affect the stability gap?
(2) How does momentum and/or optimizer type affect the stability gap?
(3) How does mini-batch size affect the stability gap?

(4) To what extent can the method Layerwise Proximal Replay **[Yoo et al., 2024 ICML]** reduce the stability gap? (code available)
(5) To what extent can using second-order optimization methods reduce the stability gap? (hint available)

(6) **[BONUS]** Can you propose and test your own method to reduce the stability gap?

Practicalities

- You will learn to implement continual learning experiments with deep neural networks in PyTorch. (Prior experience with PyTorch not needed.)

- Code that can be used as starting point is available: https://github.com/GMvandeVen/continual-learning

- Relatively simple, publicly available datasets (MNIST, CIFAR) will be used