A new direction for continual learning:
ask not only *where* to go, also *how* to get there

Gido van de Ven

E-mail: gido.vandeven@kuleuven.be

Website: https://gmvandeven.github.io

26 March 2025

Overview

Introduction

- ➔ What is continual learning?
- ➔ Why is continual learning important?
- ➔ Why is continual learning difficult?

Main part

- ➔ Current approach to continual learning
- ➔ Problem with current approach: the stability gap
- ➔ The road forward: a new direction for continual learning

Overview

Introduction

➔ **What is continual learning?**

➔ Why is continual learning important?

➔ Why is continual learning difficult?

Main part

➔ Current approach to continual learning

➔ Problem with current approach: the stability gap

➔ The road forward: a new direction for continual learning

What is continual learning?

- In *standard machine learning*, a learning algorithm has access to all training data at the same time

- In *continual learning*, two key differences are:
  - the training data arrive incrementally
  - the distribution from which the training data are sampled changes over time

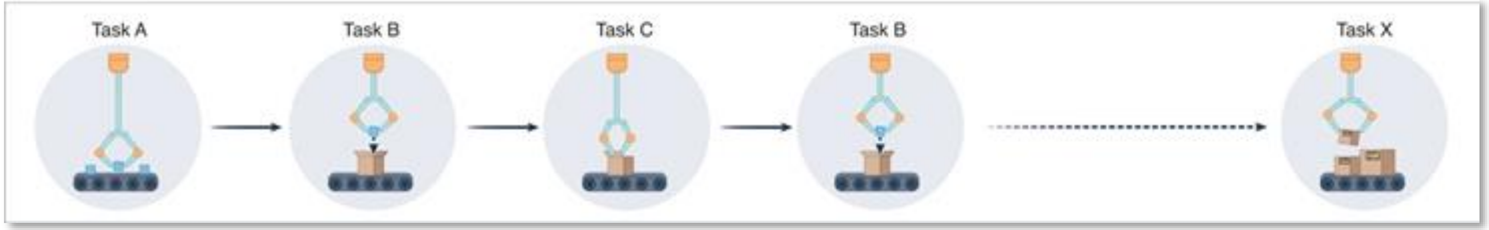


Figure source: Kudithipudi et al. (2022, Nature Machine Intelligence)

What is continual learning?

- In *standard machine learning*, a learning algorithm has access to all training data at the same time

- In *continual learning*, two key differences are:
  - the training data arrive incrementally
  - the distribution from which the training data are sampled changes over time

Continual learning is a **key aspect of intelligence**, but **very challenging** for current AI algorithms

Overview

Introduction

➔ What is continual learning?

➔ **Why is continual learning important?**

➔ Why is continual learning difficult?

Main part

➔ Current approach to continual learning

➔ Problem with current approach: the stability gap

➔ The road forward: a new direction for continual learning

Motivating example: self-driving cars

Cartoon created with generative AI

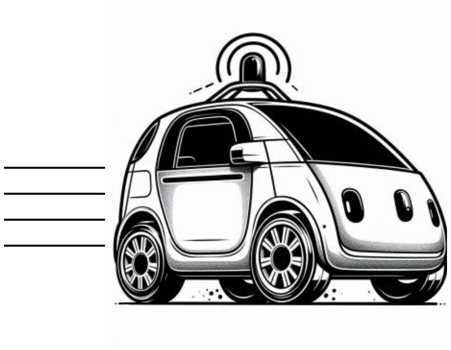
Motivating example: self-driving cars

A LOT of training data

Cartoon created with generative AI

Motivating example: self-driving cars

A LOT of training data

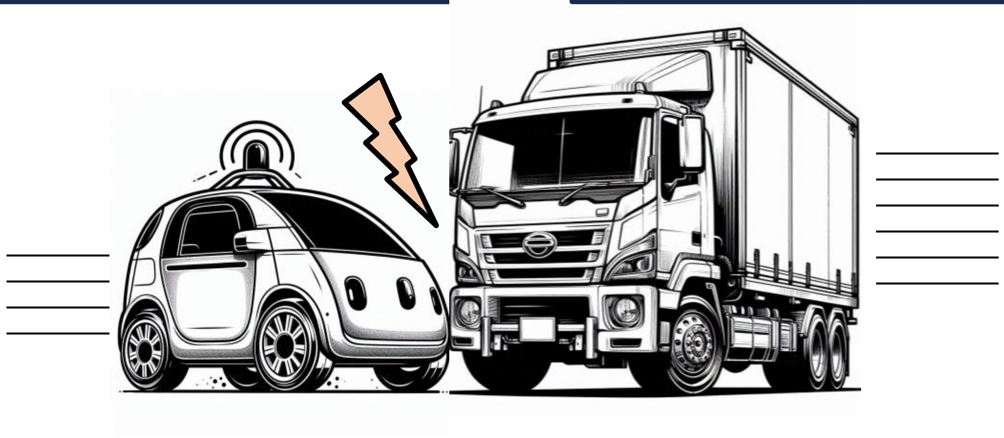


Cartoon created with generative AI

Motivating example: self-driving cars

A LOT of training data

More data of
white trucks



Cartoon created with generative AI

Motivating example: self-driving cars

A LOT of training data

More data of
white trucks

Cartoon created with generative AI

Motivating example: self-driving cars

A LOT of training data

More data of
white trucks

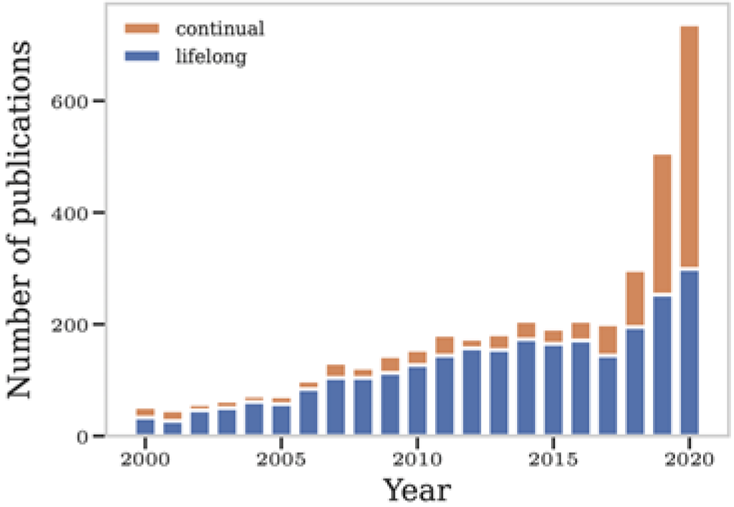
→ **Inefficient!**

Cartoon created with generative AI

Promising solution: continual learning

- Add capabilities to existing models in much more efficient manners
- Hot topic in deep learning

Figure source: Mundt et al. (2022, ICLR)



Overview

Introduction

→ What is continual learning?

→ Why is continual learning important?

→ **Why is continual learning difficult?**

Main part

→ Current approach to continual learning

→ Problem with current approach: the stability gap

→ The road forward: a new direction for continual learning

Catastrophic forgetting (McCloskey and Cohen, 1989; Ratcliff, 1990)

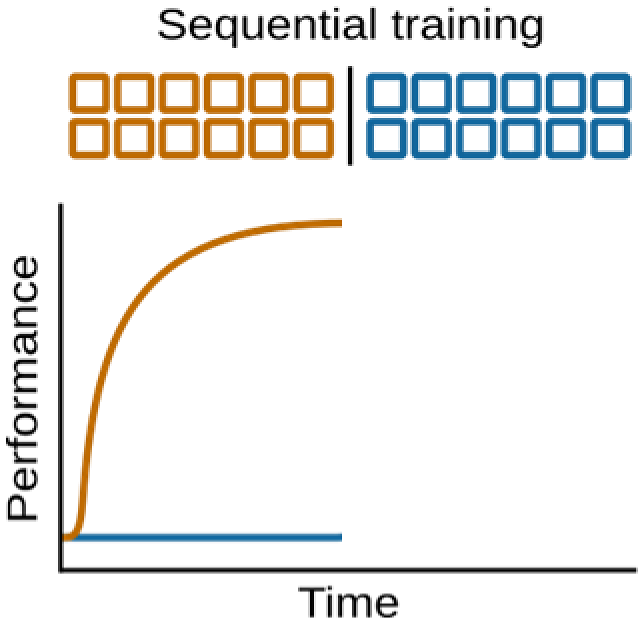


Figure source: van de Ven, Soures et al. (2024, arXiv)

Catastrophic forgetting (McCloskey and Cohen, 1989; Ratcliff, 1990)

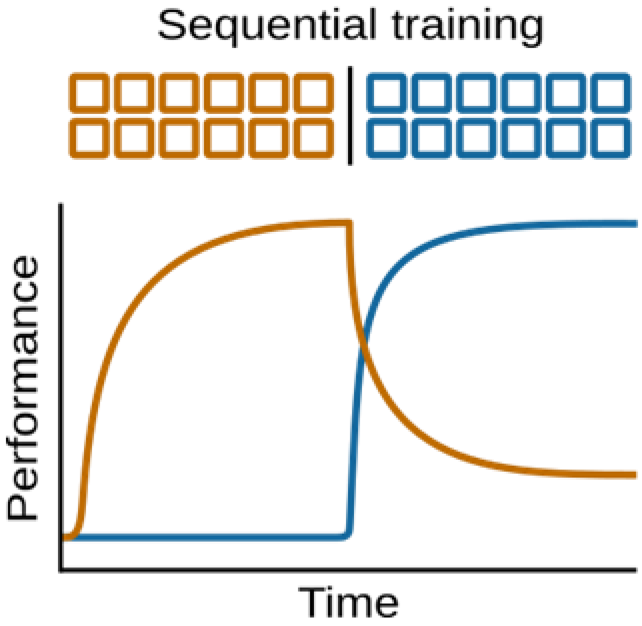


Figure source: van de Ven, Soures et al. (2024, arXiv)

Catastrophic forgetting (McCloskey and Cohen, 1989; Ratcliff, 1990)

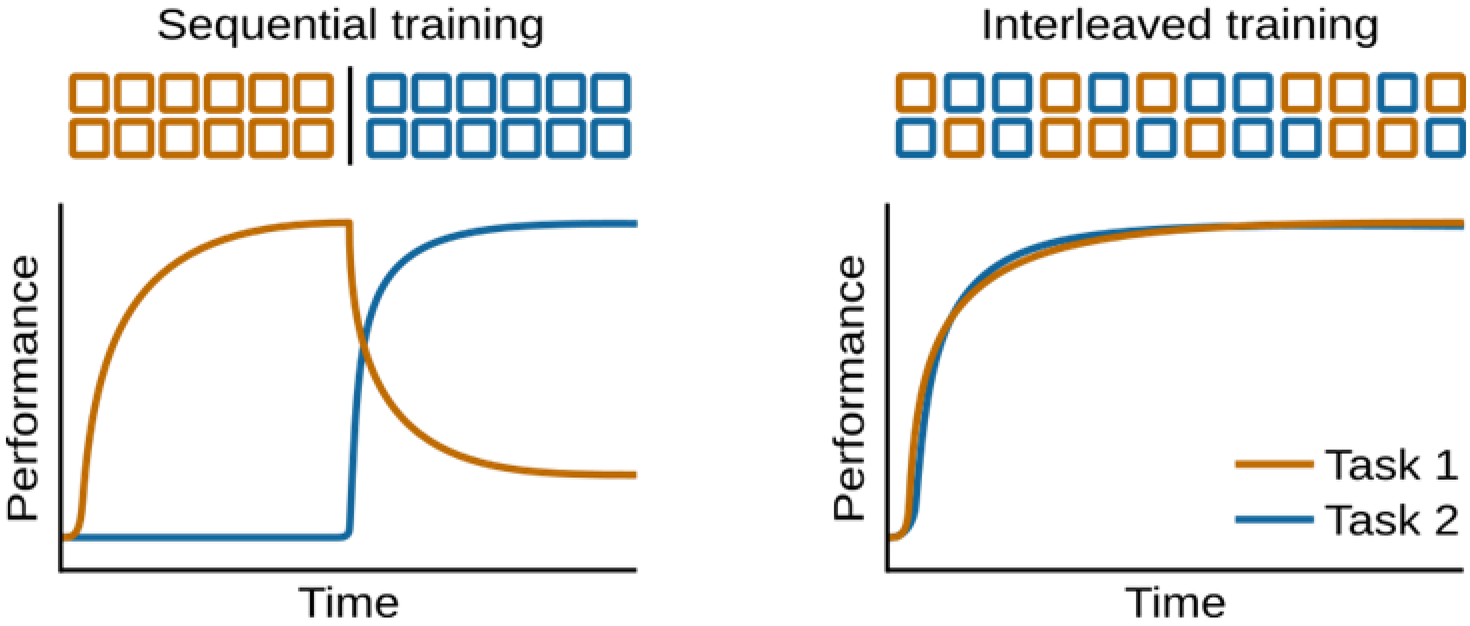
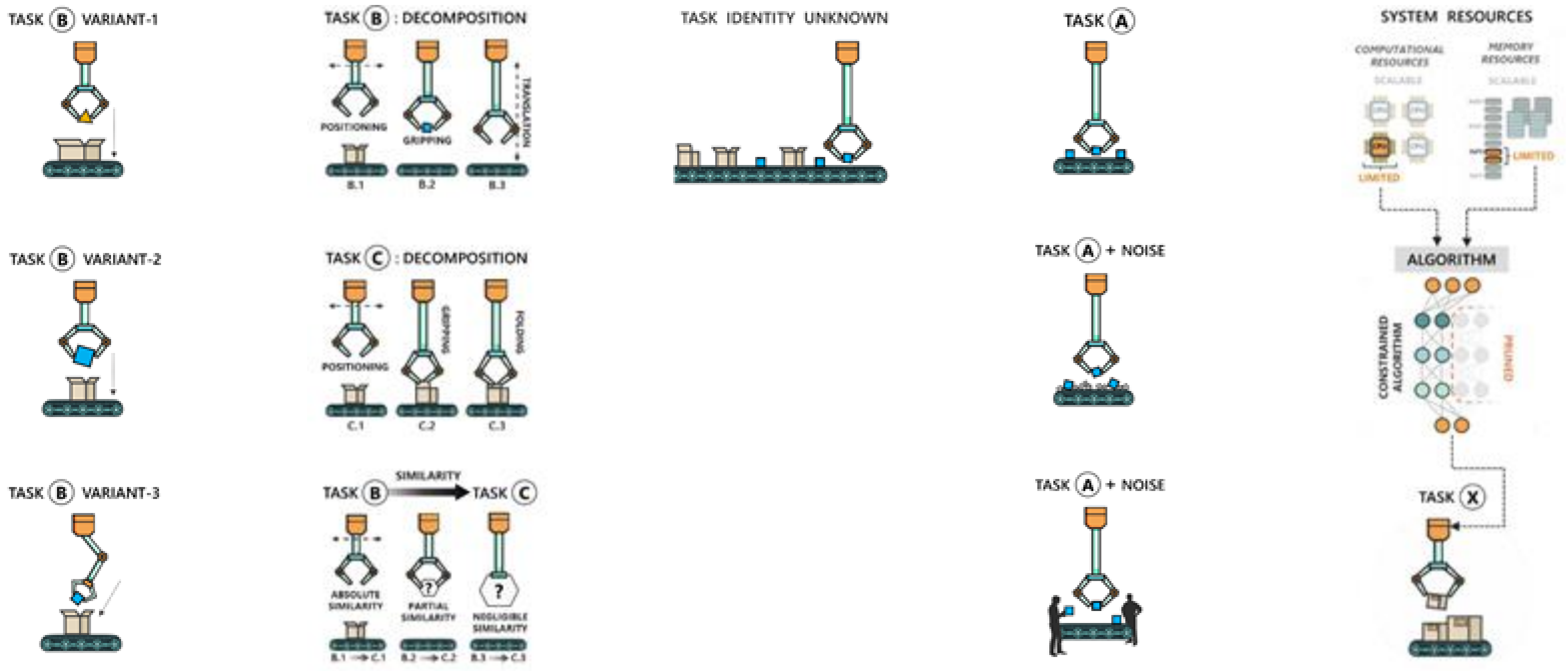


Figure source: van de Ven, Soures et al. (2024, arXiv)

Other features important for continual learning

Rapid adaptation	Exploiting task similarity	Task agnosticity	Noise tolerance	Resource efficiency and sustainability



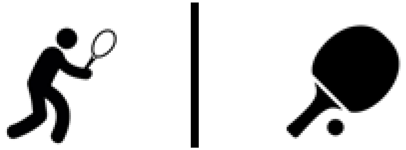
Kudithipudi et al. (2022, *Nature Machine Intelligence*)

Three types of incremental learning

- Task-incremental learning *(Task-IL)*

  "Incrementally learn a set of clearly distinguishable tasks"

Important challenge: achieve positive transfer between tasks



- Domain-incremental learning *(Domain-IL)*

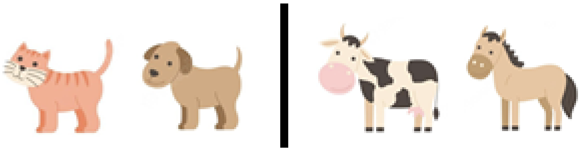
  "Learn the same type of problem in different contexts"

Important challenge: alleviate catastrophic forgetting / identify the domain

- Class-incremental learning *(Class-IL)*

  "Incrementally learn a growing number of classes"

Important challenge: learn to discriminate between objects not observed together



Images designed by Freepik

van de Ven & Tolias (2018, *NeurIPS workshop*) – van de Ven et al. (2022, *Nature Machine Intelligence*)

Summary of the introduction

➔ Continual learning is the problem of incrementally learning from a non-stationary stream of data

➔ Successful continual learning can enable updating trained models in a much more efficient manner

➔ Catastrophic forgetting is an important challenge, but not the only one

Next: how to address the continual learning problem

Overview

Introduction

➔ What is continual learning?

➔ Why is continual learning important?

➔ Why is continual learning difficult?

Main part

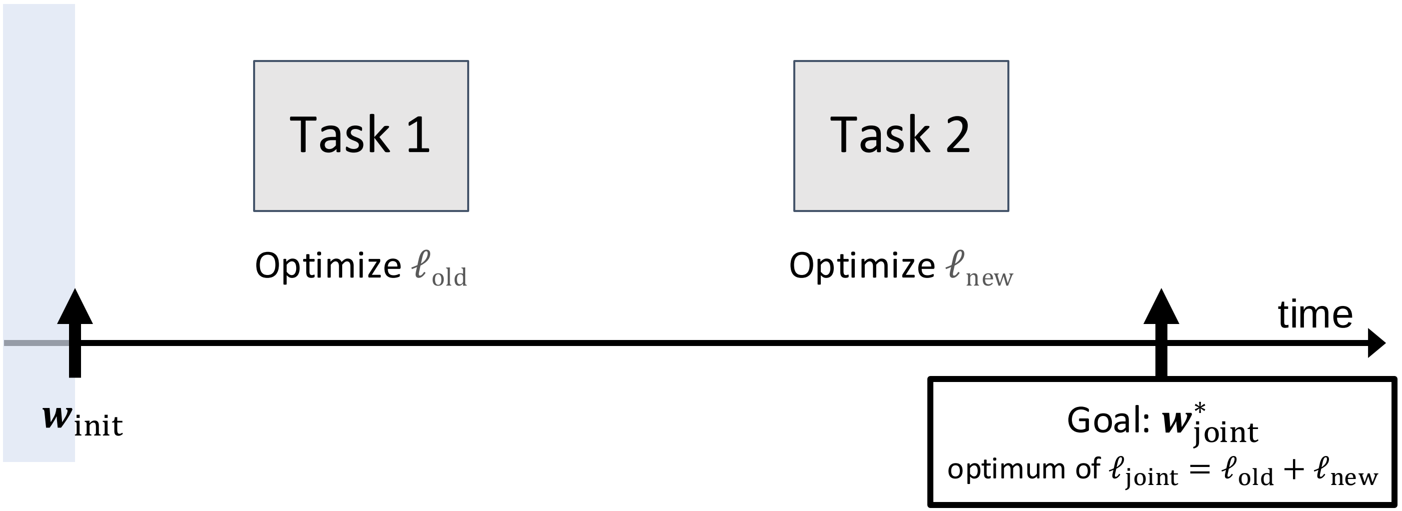
➔ **Current approach to continual learning**

➔ Problem with current approach: the stability gap

➔ The road forward: a new direction for continual learning

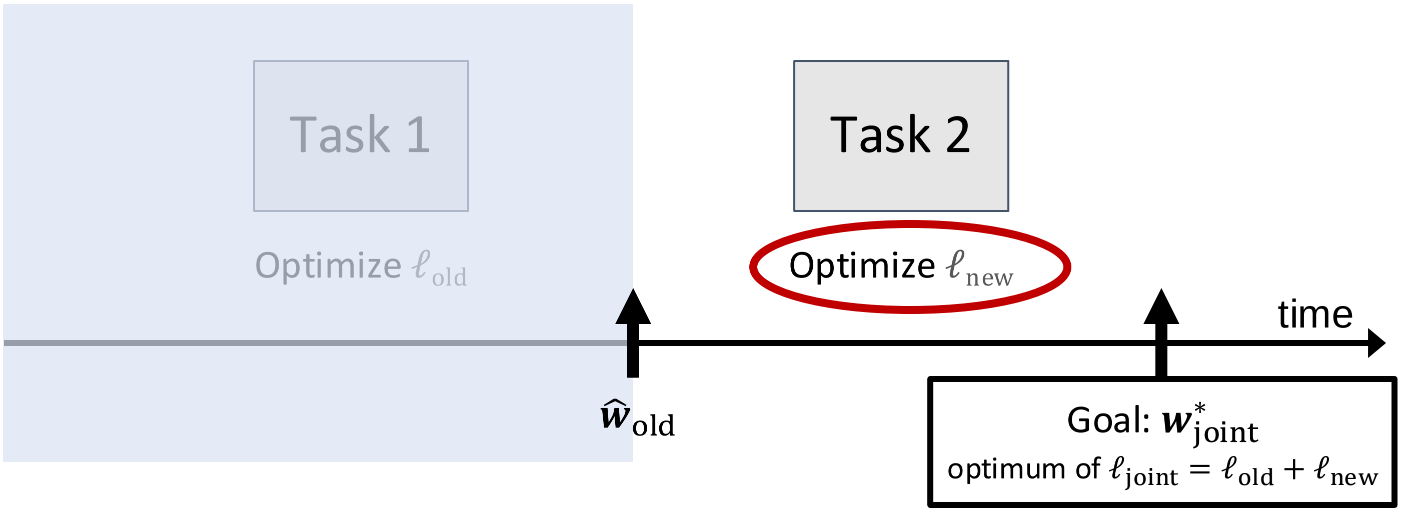
The core continual learning problem

→ Optimize parameters $\boldsymbol{w}$ of model $f_{\boldsymbol{w}}$ for two tasks observed one after the other



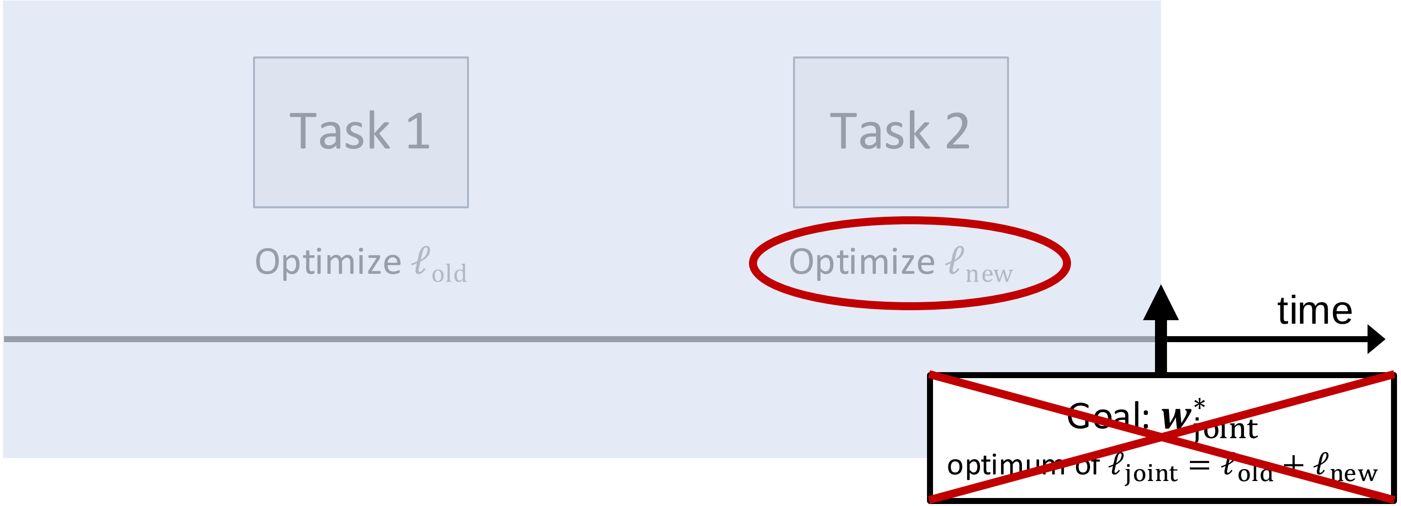
The core continual learning problem

→ Optimize parameters $\boldsymbol{w}$ of model $f_{\boldsymbol{w}}$ for two tasks observed one after the other



The core continual learning problem

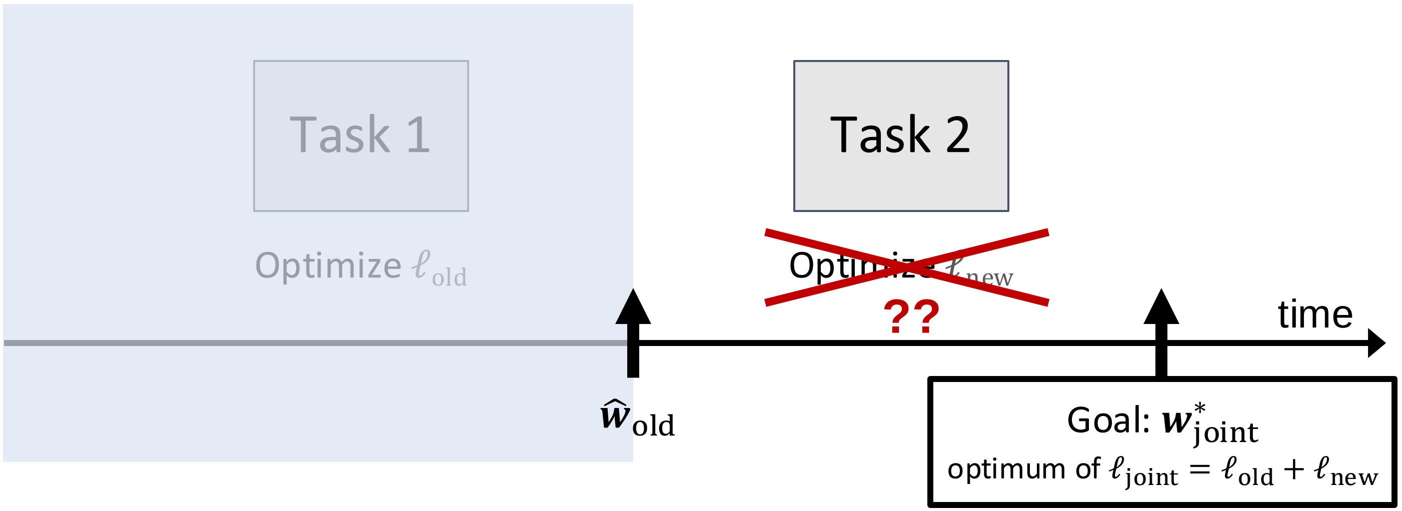
→ Optimize parameters $\boldsymbol{w}$ of model $f_{\boldsymbol{w}}$ for two tasks observed one after the other



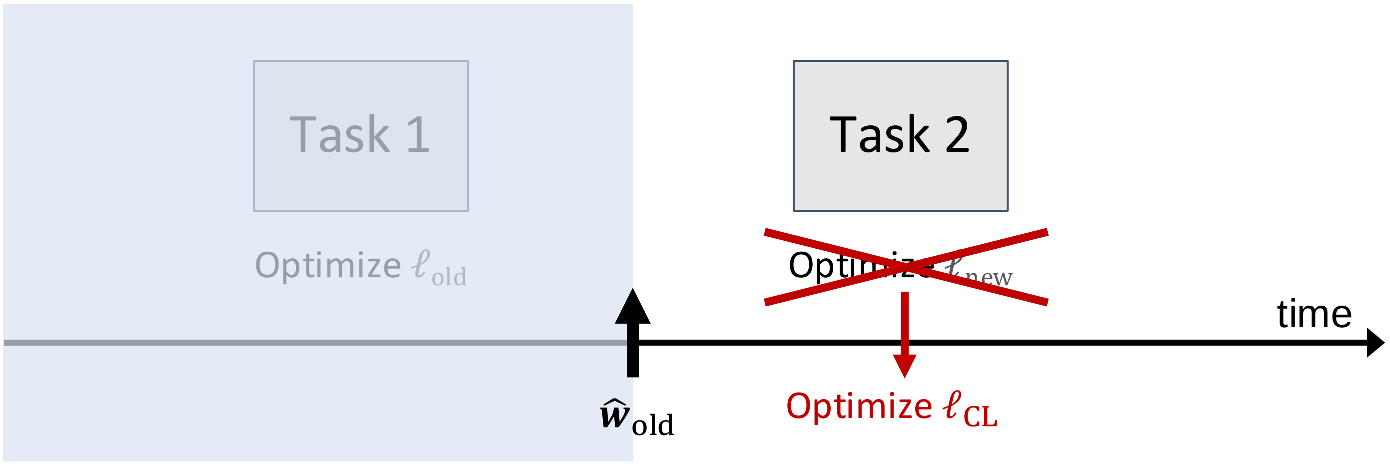
Catastrophic forgetting

The core continual learning problem

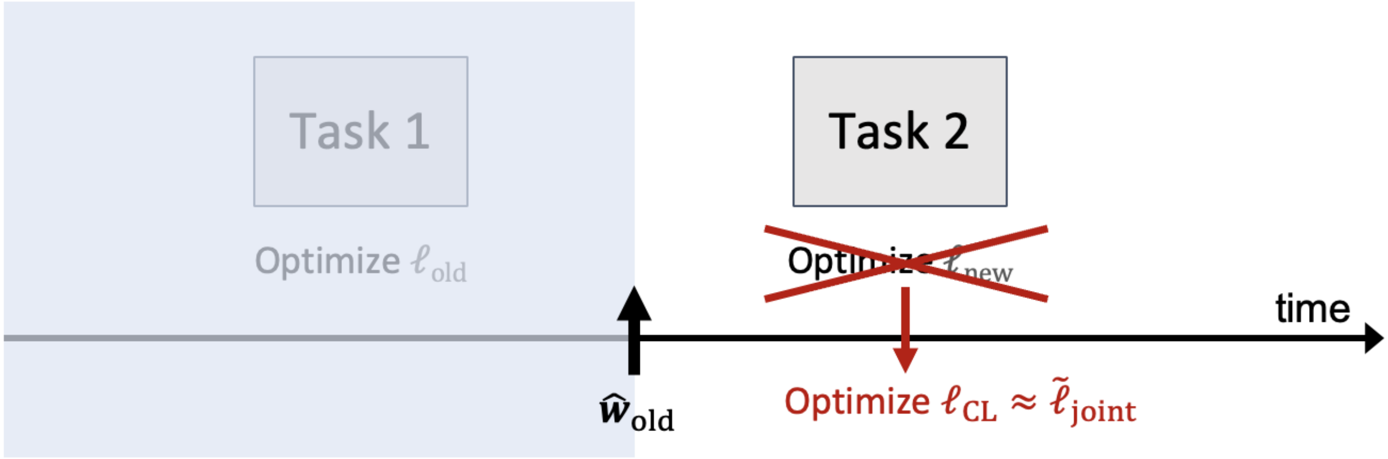
→ Optimize parameters $\boldsymbol{w}$ of model $f_{\boldsymbol{w}}$ for two tasks observed one after the other



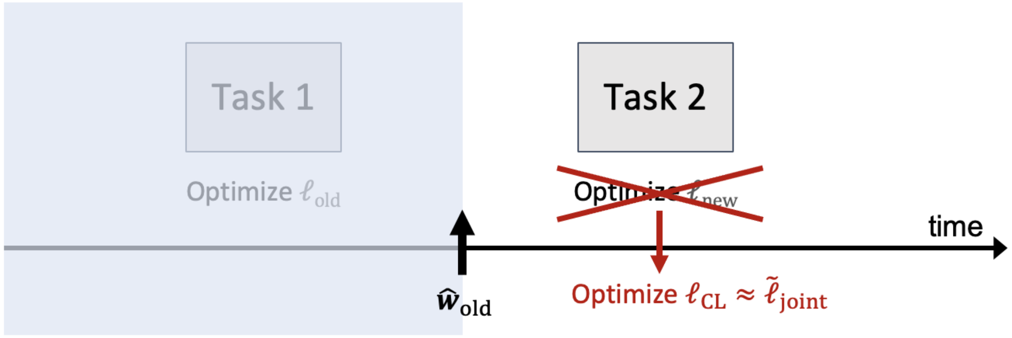
Current approach to continual learning: make changes to the loss



Current approach to continual learning: make changes to the loss



Current approach to continual learning: make changes to the loss



Replay

$$\ell_{\text{CL}} = \ell_{\text{new}} + \underbrace{\ell_{\text{replay}}}_{\approx\, \tilde{\ell}_{\text{old}}}$$

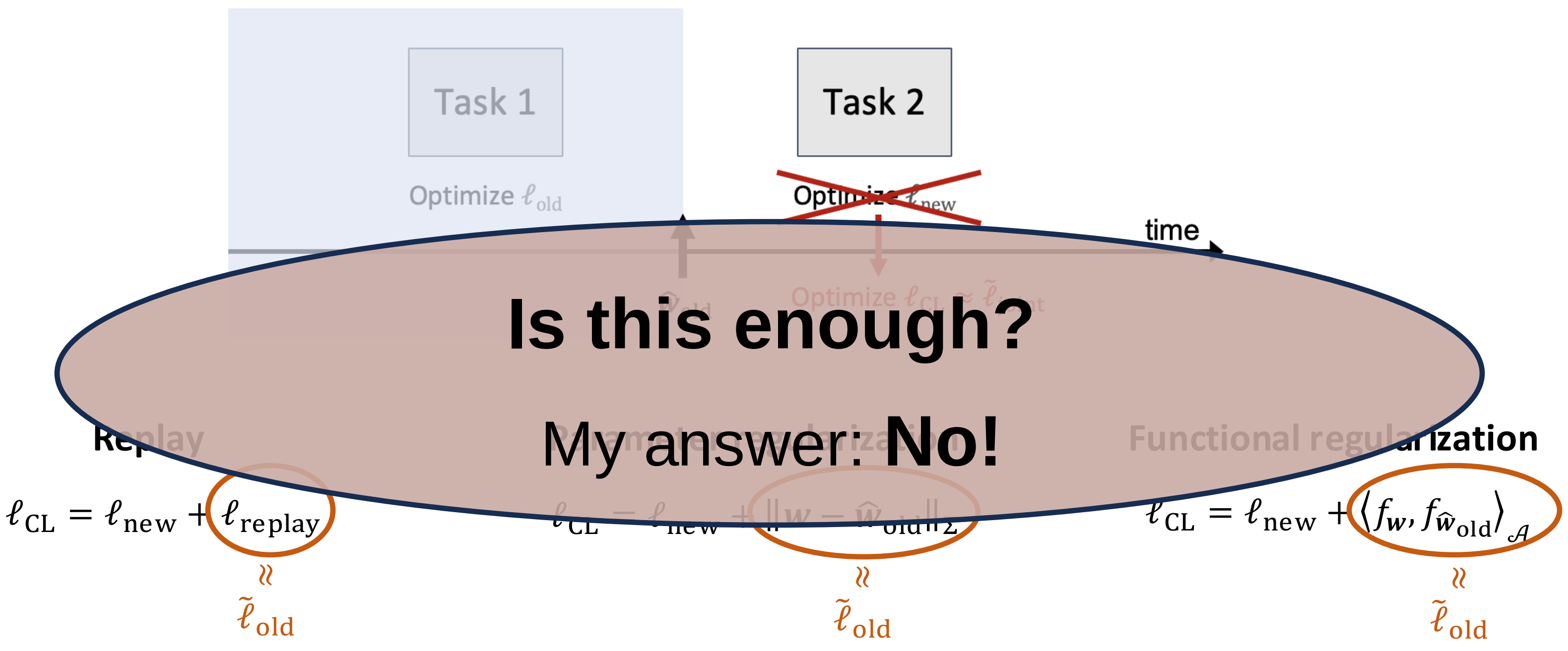
Parameter regularization

$$\ell_{\text{CL}} = \ell_{\text{new}} + \underbrace{\|\boldsymbol{w} - \widehat{\boldsymbol{w}}_{\text{old}}\|_{\Sigma}}_{\approx\, \tilde{\ell}_{\text{old}}}$$

Functional regularization

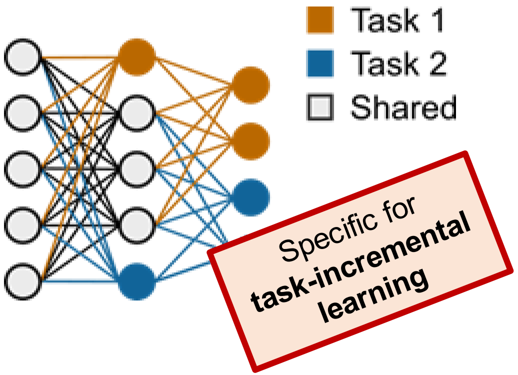
$$\ell_{\text{CL}} = \ell_{\text{new}} + \underbrace{\langle f_{\boldsymbol{w}}, f_{\widehat{\boldsymbol{w}}_{\text{old}}} \rangle_{\mathcal{A}}}_{\approx\, \tilde{\ell}_{\text{old}}}$$

Current approach to continual learning: make changes to the loss

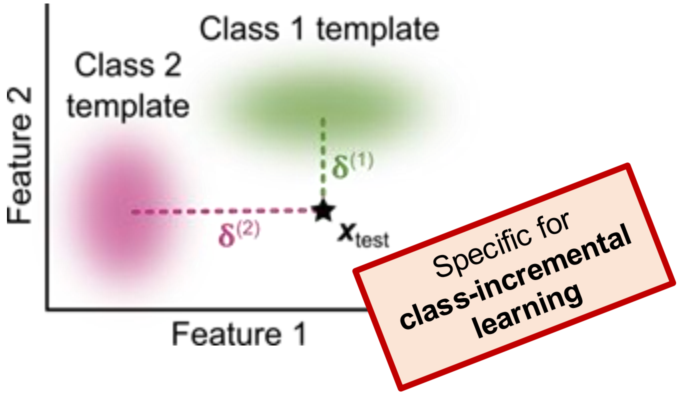


Other approaches to continual learning

Context-specific components



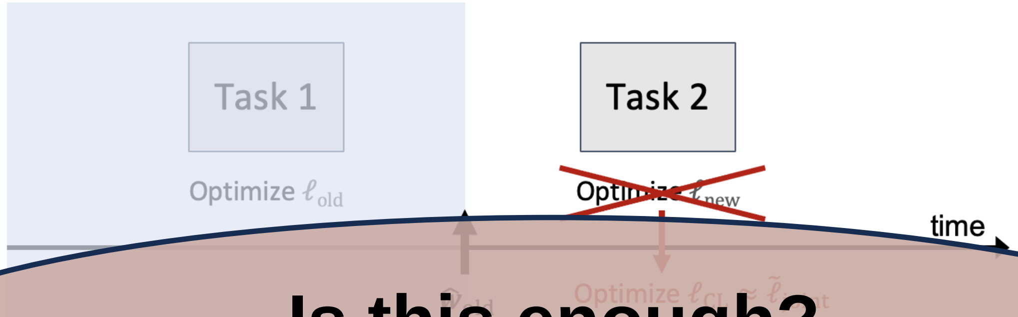
Template-based classification



These specialized approaches do not address the “core continual learning problem” (i.e., optimizing a shared set of parameters in a continual manner)

These approaches are useful and important, but not what I mean here

Current approach to continual learning: make changes to the loss



Is this enough?

My answer: **No!**

Replay

$$\ell_{\text{CL}} = \ell_{\text{new}} + \ell_{\text{replay}}$$

$$\approx \tilde{\ell}_{\text{old}}$$

Parameter regularization

$$\ell_{\text{CL}} = \ell_{\text{new}} + \|w - \widehat{w}_{\text{old}}\|_{\Sigma}$$

$$\approx \tilde{\ell}_{\text{old}}$$

Functional regularization

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

$$\approx \tilde{\ell}_{\text{old}}$$

Overview

Introduction

→ What is continual learning?

→ Why is continual learning important?

→ Why is continual learning difficult?

Main part

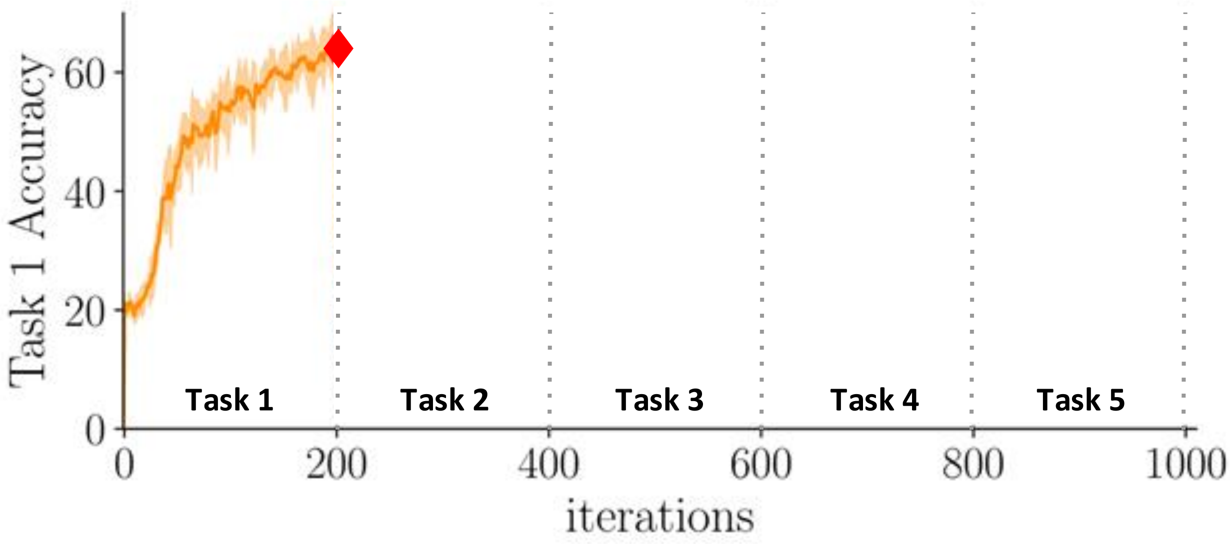
→ Current approach to continual learning

→ **Problem with current approach: the stability gap**

→ The road forward: a new direction for continual learning

Does replay prevent forgetting?

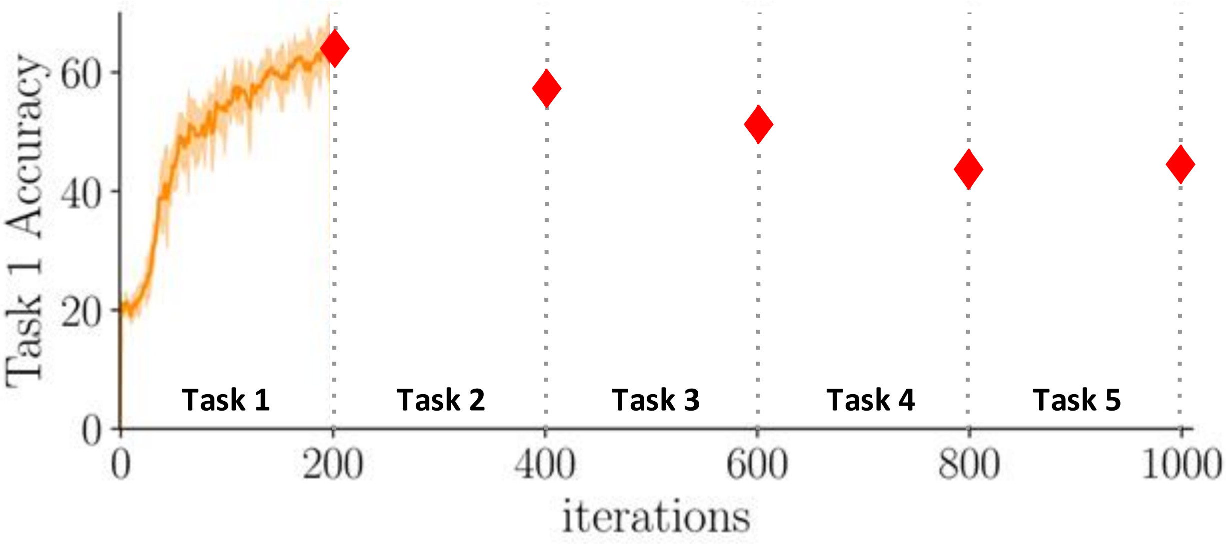
Replay, Class-incremental Split MiniImageNet



De Lange et al. (2023) "Continual evaluation for lifelong learning: Identifying the stability gap" *ICLR, spotlight*

Does replay prevent forgetting?

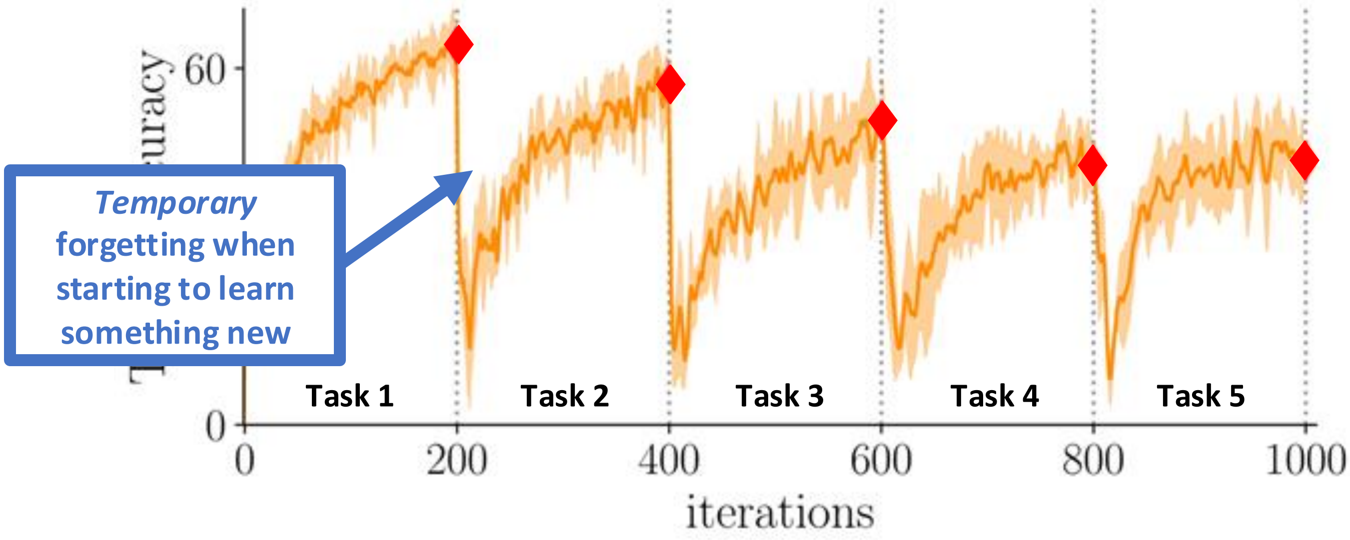
Replay, Class-incremental Split MiniImageNet



De Lange et al. (2023) "Continual evaluation for lifelong learning: Identifying the stability gap" *ICLR, spotlight*

Does replay prevent forgetting?

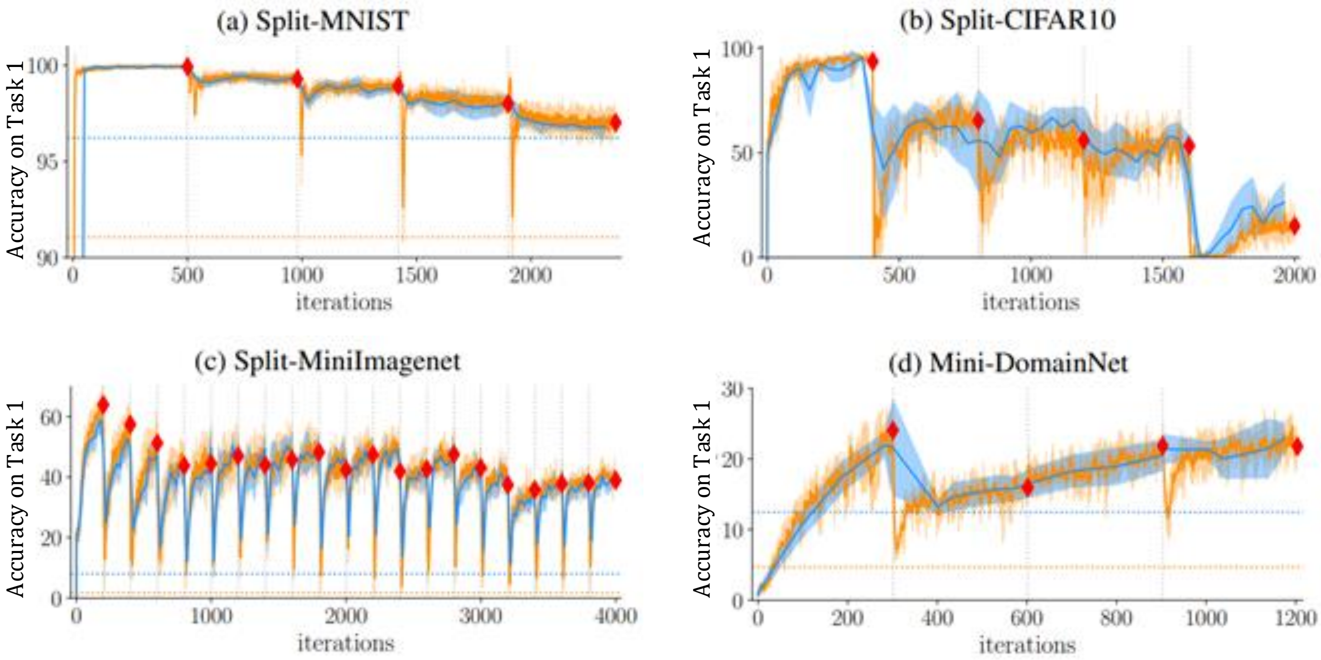
Replay, Class-incremental Split MiniImageNet



De Lange et al. (2023) "Continual evaluation for lifelong learning: Identifying the stability gap" *ICLR, spotlight*

The stability gap is consistently observed

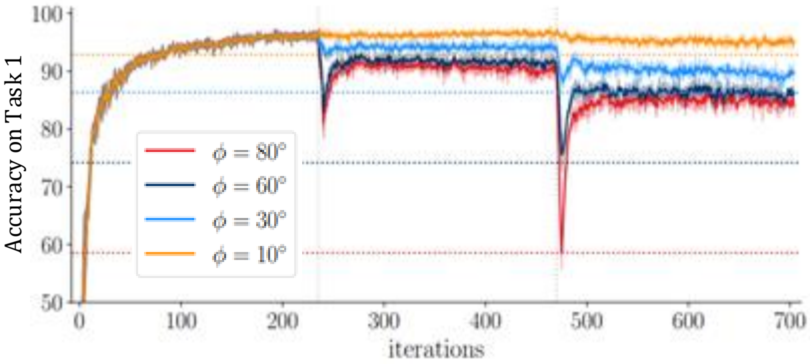
Replay, Class-incremental on …



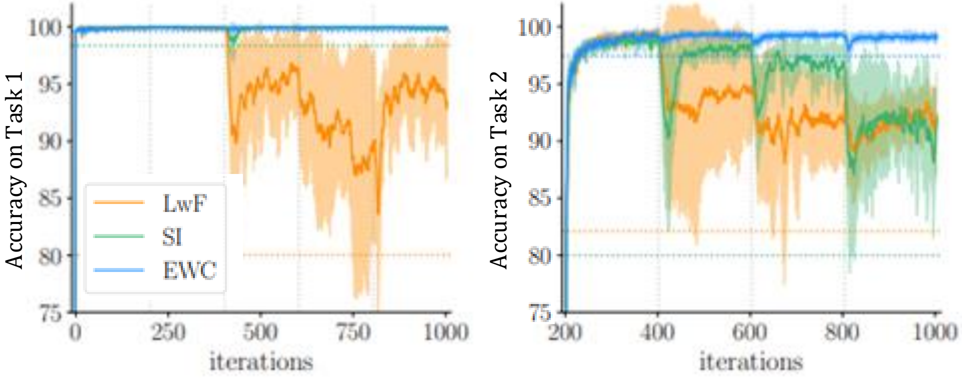
De Lange et al. (2023) “Continual evaluation for lifelong learning: Identifying the stability gap” *ICLR, spotlight*

… also in other settings or with other methods

Replay, Domain-incremental Rotated MNIST



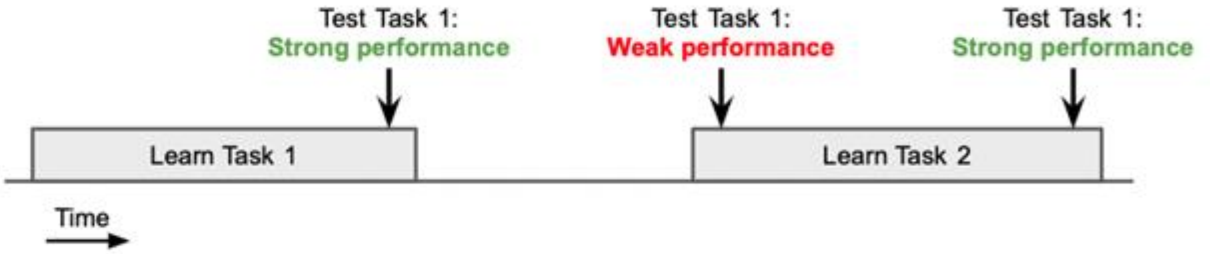
Regularization, Task-incremental Split MNIST



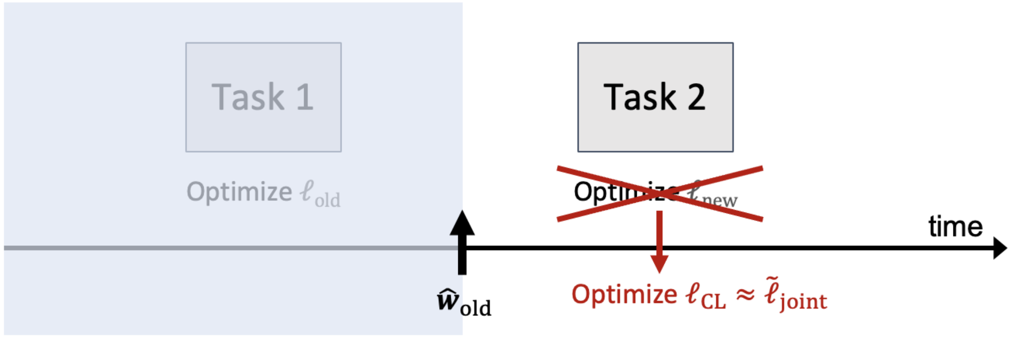
De Lange et al. (2023) [“Continual evaluation for lifelong learning: Identifying the stability gap”] *ICLR, spotlight*

Why should we care?

- Problematic for safety-critical applications
  - Worst-case performance might be important
  - Could be exploited by adversarial agent with control over training stream
- Seems highly inefficient
  - Preventing forgetting seems more efficient than having to re-learn
- Scientifically interesting
  - Do humans suffer from transient forgetting upon learning something new?



Can the "current approach to continual learning" avoid the stability gap?



Replay

$$\ell_{\text{CL}} = \ell_{\text{new}} + \underbrace{\ell_{\text{replay}}}_{\approx\, \tilde{\ell}_{\text{old}}}$$

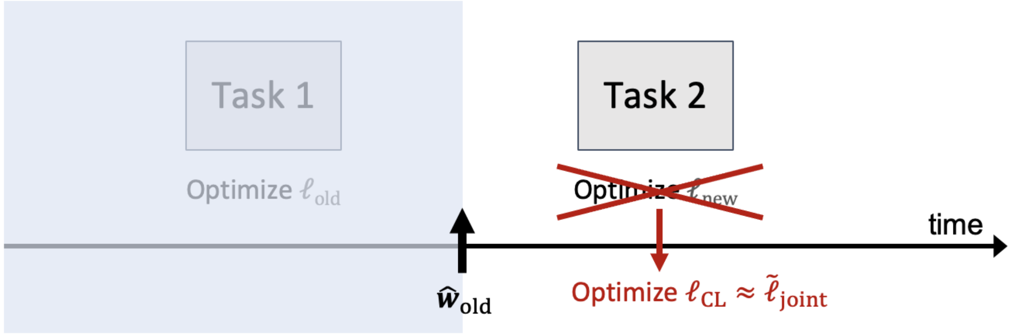
Parameter regularization

$$\ell_{\text{CL}} = \ell_{\text{new}} + \underbrace{\|\boldsymbol{w} - \widehat{\boldsymbol{w}}_{\text{old}}\|_{\Sigma}}_{\approx\, \tilde{\ell}_{\text{old}}}$$

Functional regularization

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

Can the "current approach to continual learning" avoid the stability gap?



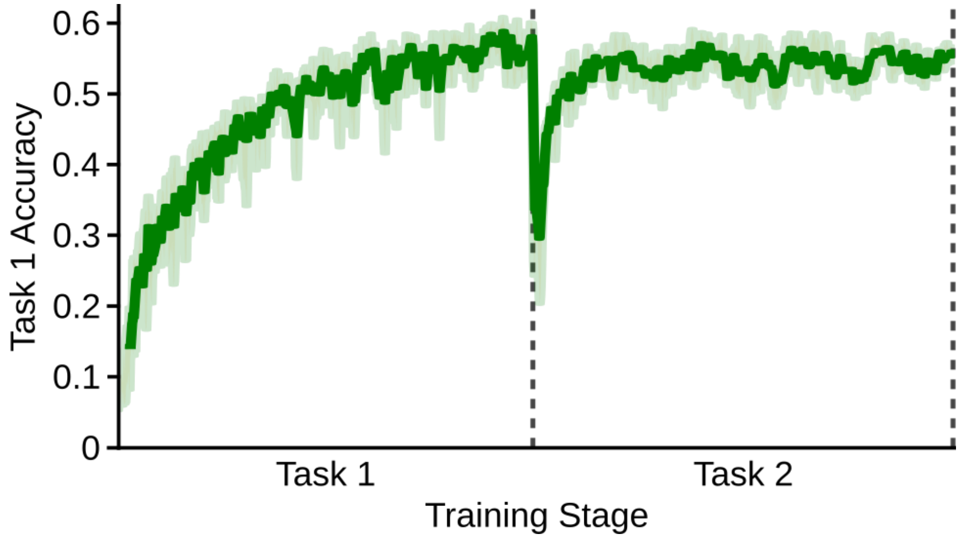
Upper bound of the current approach to continual learning

Incremental joint training

$$\ell_{\text{CL}} = \ell_{\text{new}} + \ell_{\text{old}} = \ell_{\text{joint}}$$

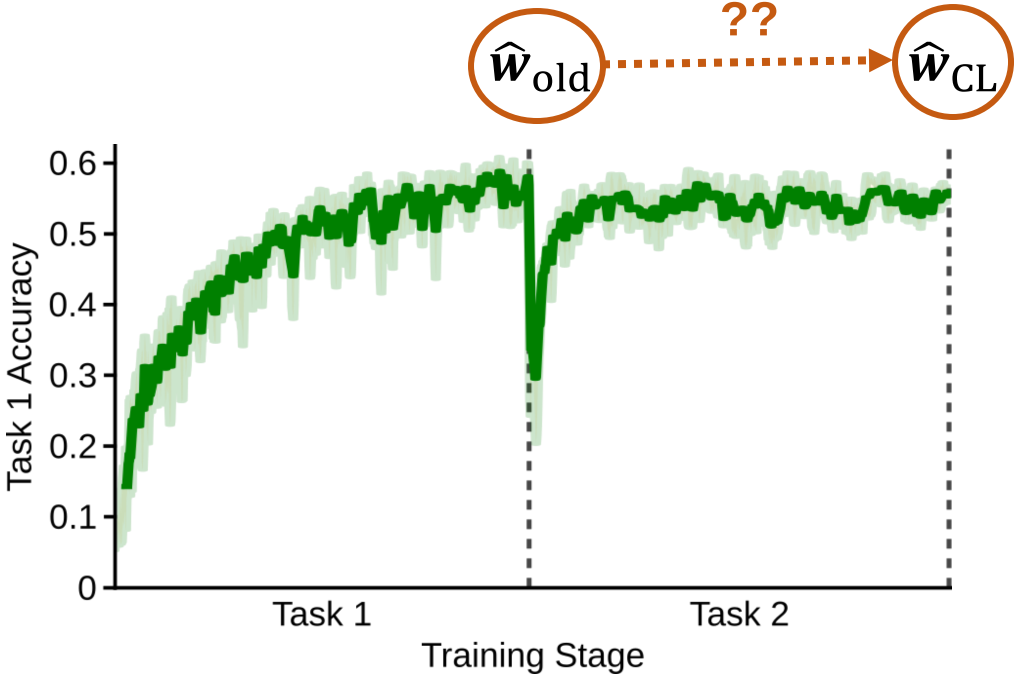
Can the “current approach to continual learning” avoid the stability gap? → No!

Incremental joint training – Domain CIFAR-100

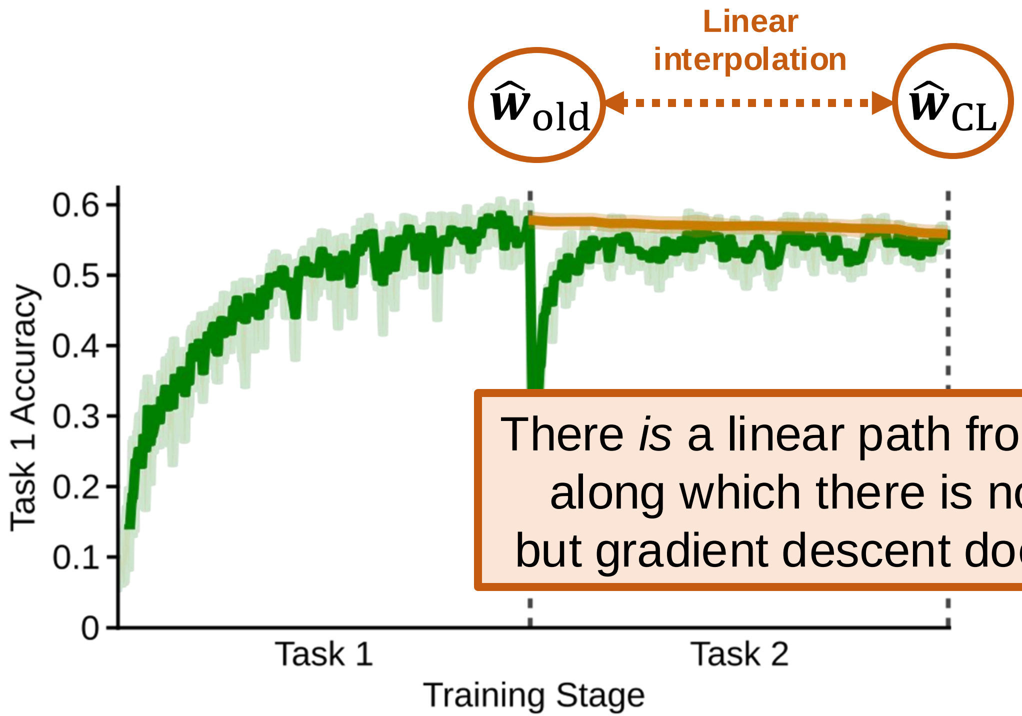


Hess et al. (2023) “Two complementary perspectives to continual learning” *1st ContinualAI Unconference, PMLR*

Can the stability gap be avoided at all?



Can the stability gap be avoided at all? → Yes!



There *is* a linear path from $\hat{\boldsymbol{w}}_{\mathrm{old}}$ to $\hat{\boldsymbol{w}}_{\mathrm{CL}}$ along which there is no forgetting, but gradient descent does not take it!

Overview

Introduction

➔ What is continual learning?

➔ Why is continual learning important?

➔ Why is continual learning difficult?

Main part

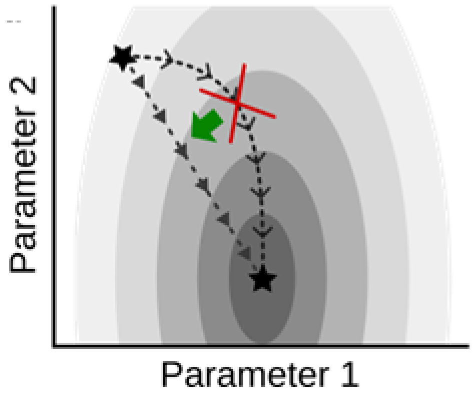
➔ Current approach to continual learning

➔ Problem with current approach: the stability gap

➔ **The road forward: a new direction for continual learning**

Continual learning needs a new direction

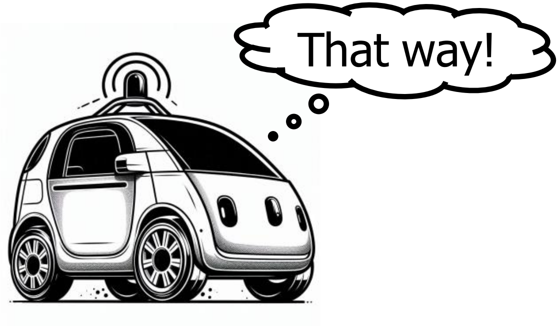
- To overcome the stability gap, changes must be made to *how* the loss is optimized
- Standard optimization routines for deep learning have been developed for the stationary setting
- No guarantees in continual setting, yet widely used
- Fundamental difference between both settings:
    - Stationary → start from random initialization
    - Continual → start from partial solution

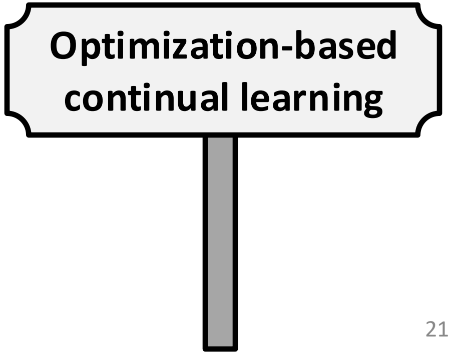


How to improve optimization for continual learning?

→ An exciting open question!

Techniques from neuromorphic computing? Any ideas??





Cartoon created with generative AI

Funding acknowledgements

The work and research towards preparing this presentation has been supported by an IBRO-ISN Research Fellowship, by the *Lifelong Learning Machines* (L2M) program of the Defense Advanced Research Projects Agency (DARPA) via contract HR0011-18-2-0025, by funding from the European Union under Horizon 2020 (ERC project *KeepOnLearning*, grant 101021347) and Horizon Europe (Marie Skłodowska-Curie fellowship, grant 101067759), and by a senior postdoctoral fellowship from the Research Foundation – Flanders (FWO, grant 1266823N).