



Beyond Forgetting with Continual Pre-Training

continual representation learning with pre-trained models



The importance of pre-trained models

Learned representations (foundation models)

- can be kept fixed
- allow zero-shot adaptation / transfer learning
- can be powered by prompt tuning & in-context learning

What if we *need* to update the model with new information?

Continual fine-tuning

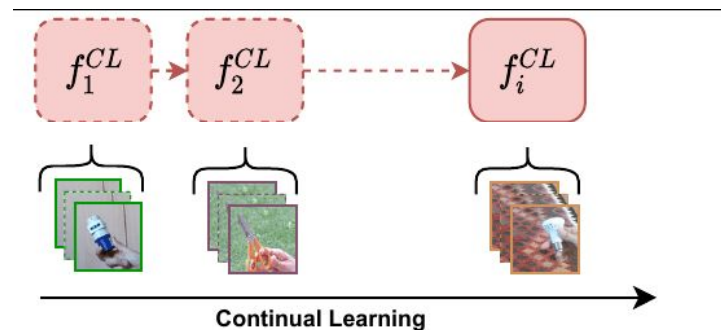
Just continue training!

- + Always ready to tackle any task
- Need to mitigate forgetting

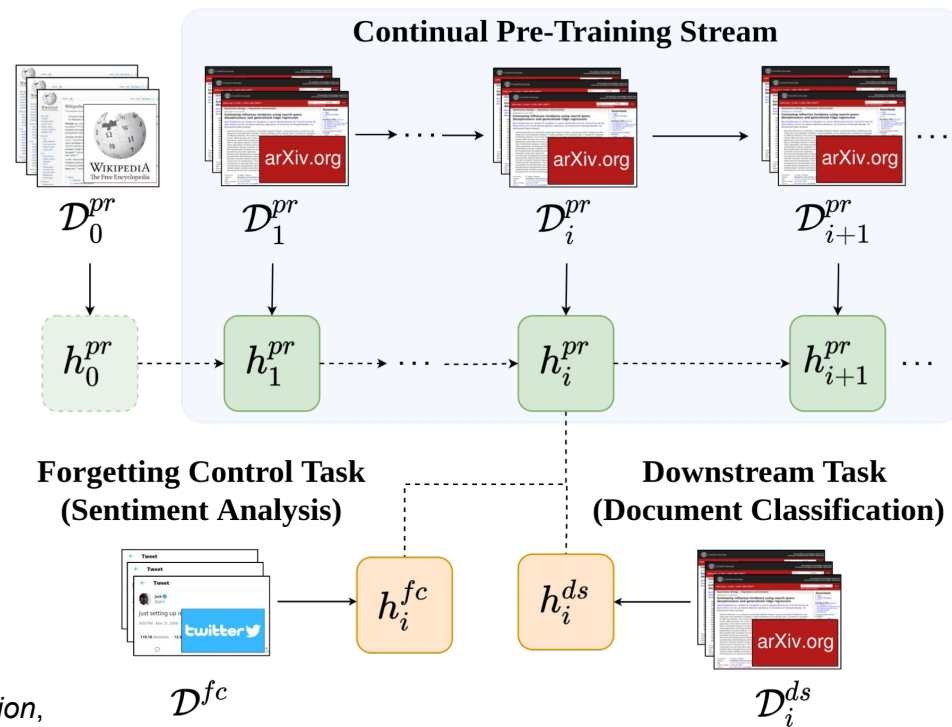
Can we do better?

Can we exploit the *core* properties of a pre-trained model?

Continual *representation* learning



Continual Pre-Training Scenario



A. Cossu, T. Tuytelaars, A. Carta,
L. Passaro, V. Lomonaco, and D. Bacciu,
Continual Pre-Training Mitigates Forgetting in Language and Vision,
arXiv, 2022.



No forgetting? Well...

Self-supervised pre-training seems to be crucial. More than scale.

- fine-tuning required to address a given task
- + fine-tuning can be very quick
- + “stronger” features for downstream tasks (compared to continual fine-tuning)

Practical scenario allowing to go *beyond* forgetting



Model personalization

- Pre-train on cloud, fine-tune on edge
- Privacy preserving
- Exploit large unlabeled corpora but also user-specific feedback

Flexibility, hybrid approaches (pre-training fine-tuning continuum)
choose when to pre-train, when to fine-tune