Continual Pre-training via Sparse Updates

Research internship 2026 – M2 – 5 to 6 months

Keywords

parameter-efficient fine-tuning, continual learning, multimodal models, image classification

Context

Continual learning is a long-standing challenge in artificial intelligence, aiming to train models to retain and refine their learning, thereby enabling them to perform a wider, more complex range of tasks through growing experience. However, neural networks are usually faced with the phenomenon of *catastrophic forgetting*, namely a performance drop on past data when updated using novel data [McCloskey and Cohen, 1989].

Today, so-called *foundation models*, trained on enormous datasets, have come close to or even surpassed human-level performance for a wide range of perceptual and generative tasks, from image classification to text summarization [Achiam et al., 2023, Radford et al., 2021]. Large pre-training boosts transferability and generally improves continual learning performance [Feillet et al., 2025]. However, foundation models have not eliminated the need for continual learning algorithms.

First, general-purpose pre-trained models often require an additional adaptation step to achieve optimal performance in specific domains (e.g., industrial images, medical text, etc.). Adapting these models without compromising their generalizability is a challenging task, as it risks overfitting to the target training corpus. Second, despite most of the computational effort being allocated to the pre-training phase, computational efficiency remains an issue, particularly when adapting large-scale models. Third, it would be more efficient to cumulatively enhance a pre-trained model to perform new tasks, rather than developing a specific, adapted model for each task. This is the purpose of *continual pre-training* [Roth et al., 2024, Cossu et al., 2024]. Hence, continual learning is a rising topic in the era of foundation models.

Fostered by the works on pre-training of large language models, parameter-efficient fine-tuning (PEFT) methods such as Adapters [Houlsby et al., 2019] and LoRA [Hu et al., 2021] are now widely used across various deep learning architectures and applications. PEFT is also considered in the context of continual learning; for example, [He et al., 2025] proposes an adaptation of LoRA for a sequence of tasks. Alternatively, [Lin et al., 2025] designs a sparse memory layer to tackle efficient fine-tuning on question-answering corpora. Overall, continual learning via parsimonious updates is a promising line of research to address a sequence of learning tasks efficiently while preserving previously learned capabilities to tackle future tasks.

In this context, this internship will focus on sparse update methods to tackle a continual pre-training problem. Depending on the candidate's profile, we will focus either on *image classification* using vision and language-vision models (e.g., CLIP) or on a *textual question-answering* task.

Objectives of the internship

The aim of the internship is to review, implement, benchmark, and improve state-of-the-art continual learning methods that rely on parsimonious updates.

We propose to focus on (i) continual low-rank adaptation [He et al., 2025, Wistuba et al., 2023] and (ii) sparse update methods [Yildirim et al., 2024, Lin et al., 2025, Wang et al., 2022a] (i.e. sparse gradient updates).

Other baseline methods for continual pre-training include classic full fine-tuning, prototype-based methods [Ostapenko et al., 2022, McDonnell et al., 2023] and prompt-based methods [Wang et al., 2022b, Wei et al., 2022]. Multimodal models also offer competitive zero-shot learning capabilities [Radford et al., 2021, Liu et al., 2025, Kojima et al., 2022]. Finally, we will also compare continual pre-training against test-time adaptation methods [Eddine Marouf et al., 2023]

For the continual pre-training task, we propose to rely on the framework of [Roth et al., 2024] for image classification, or to build on [Lin et al., 2025] for question answering.

The main steps of the internship can be summarized as follows:

- **Literature review**: getting familiar with the continual learning framework, identifying the main methods of the state-of-the-art
- **Implementation**: ensure compatibility of evaluation set-ups of existing implementations, implement missing methods.
- **Benchmark**: define experiment settings and run experiments to compare the selected algorithms on the task of continual pre-training.
- **Improvements**: based on the knowledge and experience gathered through the previous steps of the internship, propose improvements for existing algorithms.

Applicant Profile

Education: You are pursuing an engineering degree (M1/M2) or a Master's program specializing in a scientific field related to machine learning, computer science, applied statistics or signal processing. You have followed theoretical courses in machine learning and gained practical experience in deep learning related projects.

Technical Skills:

- Proficiency in Python and familiarity with deep learning libraries such as TensorFlow/Keras or PyTorch.
- Experience in computer vision and/or NLP.

Other Skills:

- Fluent in French and/or English (spoken and written).
- Strong analytical abilities. Ability to communicate about your work.
- Ability to work autonomously and take the initiative.

Working at LISN

The Interdisciplinary Laboratory of Numérical Sciences (Laboratoire Interdisciplinaire des Sciences du Numérique – LISN) is a joint research unit of CNRS, Université Paris-Saclay, INRIA, and CentraleSupélec. With over 400 members, it leads multidisciplinary research at the crossroads of artificial intelligence, physics, and the humanities. This internship will take place in the Department of Language Sciences and Technologies, which studies fundamental questions relating to linguistic systems and develops statistical learning models adapted to natural language processing.

Address : LISN – Site Belvédère
Campus Universitaire, bât.507 - Rue du Belvédère F- 91405 - Orsay 91400 - ORSAY

Contact

- Eva Feillet, Université Paris-Saclay, LISN.
  eva [point] feillet [at] universite-paris-saclay [point] fr
- Sahar Ghannay, Université Paris-Saclay, LISN, sahar [point]ghannay [at] lisn [point] fr

References

[Achiam et al., 2023] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

[Cossu et al., 2024] Cossu, A., Carta, A., Passaro, L., Lomonaco, V., Tuytelaars, T., and Bacciu, D. (2024). Continual pre-training mitigates forgetting in language and vision. *Neural Networks*, page 106492.

[Eddine Marouf et al., 2023] Eddine Marouf, I., Roy, S., Tartaglione, E., and Lathuilière, S. (2023). Rethinking class-incremental learning in the era of large pre-trained models via test-time adaptation. *arXiv e-prints*, pages arXiv–2310.

[Feillet et al., 2025] Feillet, E., Popescu, A., and Hudelot, C. (2025). A reality check on pre-training for exemplar-free class-incremental learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7614–7625.

[He et al., 2025] He, J., Duan, Z., and Zhu, F. (2025). Cl-lora: Continual low-rank adaptation for rehearsal-free class-incremental learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 30534–30544.

[Houlsby et al., 2019] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., and Gelly, S. (2019). Parameter-efficient transfer learning for nlp. In *International conference on machine learning*, pages 2790–2799. PMLR.

[Hu et al., 2021] Hu, E. J., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al. (2021). Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

[Kojima et al., 2022] Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., and Iwasawa, Y. (2022). Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213.

[Lin et al., 2025] Lin, J., Zettlemoyer, L., Ghosh, G., Yih, W.-T., Markosyan, A., Berges, V.-P., and Oğuz, B. (2025). Continual learning via sparse memory finetuning. *arXiv preprint arXiv:2510.15103*.

[Liu et al., 2025] Liu, W., Zhu, F., Wei, L., and Tian, Q. (2025). C-clip: Multimodal continual learning for vision-language model. In *The Thirteenth International Conference on Learning Representations*.

[McCloskey and Cohen, 1989] McCloskey, M. and Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. *The Psychology of Learning and Motivation*, 24:104–169.

[McDonnell et al., 2023] McDonnell, M. D., Gong, D., Parvaneh, A., Abbasnejad, E., and Van den Hengel, A. (2023). Ranpac: Random projections and pre-trained models for continual learning. *Advances in Neural Information Processing Systems*, 36:12022–12053.

[Ostapenko et al., 2022] Ostapenko, O., Lesort, T., Rodriguez, P., Arefin, M. R., Douillard, A., Rish, I., and Charlin, L. (2022). Continual learning with foundation models: An empirical study of latent replay. In *Conference on lifelong learning agents*, pages 60–91. PMLR.

[Radford et al., 2021] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR.

[Roth et al., 2024] Roth, K., Udandarao, V., Dziadzio, S., Prabhu, A., Cherti, M., Vinyals, O., Hénaff, O., Albanie, S., Bethge, M., and Akata, Z. (2024). A practitioner's guide to continual multimodal pretraining. *arXiv preprint arXiv:2408.14471*.

[Wang et al., 2022a] Wang, Z., Zhan, Z., Gong, Y., Yuan, G., Niu, W., Jian, T., Ren, B., Ioannidis, S., Wang, Y., and Dy, J. (2022a). Sparcl: Sparse continual learning on the edge. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A., editors, *Advances in Neural Information Processing Systems*, volume 35, pages 20366–20380. Curran Associates, Inc.

[Wang et al., 2022b] Wang, Z., Zhang, Z., Lee, C.-Y., Zhang, H., Sun, R., Ren, X., Su, G., Perot, V., Dy, J., and Pfister, T. (2022b). Learning to prompt for continual learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 139–149.

[Wei et al., 2022] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E. H., Le, Q. V., and Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA. Curran Associates Inc.

[Wistuba et al., 2023] Wistuba, M., Balles, L., Zappella, G., et al. (2023). Continual learning with low rank adaptation. In *NeurIPS 2023 Workshop on Distribution Shifts: New Frontiers with Foundation Models*.

[Yildirim et al., 2024] Yildirim, M. O., Gok, E. C., Sokar, G., Mocanu, D. C., and Vanschoren, J. (2024). Continual learning with dynamic sparse training: Exploring algorithms for effective model updates. In Chi, Y., Dziugaite, G. K., Qu, Q., Wang, A. W., and Zhu, Z., editors, *Conference on Parsimony and Learning*, volume 234 of *Proceedings of Machine Learning Research*, pages 94–107. PMLR.