

# PILOT: a pre-trained model-based continual learning toolbox

Hai-Long SUN<sup>1,2</sup>, Da-Wei ZHOU<sup>1,2\*</sup>, De-Chuan ZHAN<sup>1,2</sup> & Han-Jia YE<sup>1,2\*</sup><sup>1</sup>School of Artificial Intelligence, Nanjing University, Nanjing 210023, China<sup>2</sup>National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China

Received 9 April 2024/Revised 12 July 2024/Accepted 17 January 2025/Published online 4 March 2025

**Citation** Sun H-L, Zhou D-W, Zhan D-C, et al. PILOT: a pre-trained model-based continual learning toolbox. *Sci China Inf Sci*, 2025, 68(4): 147101, <https://doi.org/10.1007/s11432-024-4276-4>

The rapid advancements in deep learning have resulted in significant achievements across various fields. However, our ever-changing world often presents training data in a streaming format from an open environment. For example, while ChatGPT demonstrates exceptional inference capabilities, it struggles to provide users with the most up-to-date information. This challenge arises from the high costs associated with retraining a GPT model on new data daily. Therefore, the ability to continually update the model is critically important. Continual learning has been proposed as a solution to this challenge, allowing models to learn from streaming data. A major concern in continual learning is catastrophic forgetting, where models forget previously learned information when acquiring new knowledge. Many methods have been developed to address this issue and enable models to learn from new data without forgetting former knowledge. In this article, we focus on the class-incremental learning (CIL) setting, which is a common scenario in continual learning.

Traditional approaches assume that models are “trained from scratch.” However, with the rapid evolution of pre-training techniques, pre-trained models (PTMs) have become widely used for downstream tasks. These PTMs are typically trained on extensive corpora or massive image datasets, resulting in robust generalizability. Consequently, research in CIL is shifting from training models from scratch to leveraging the power of PTMs. According to a recent survey [1], methods based on PTMs exhibit significantly superior performance compared to traditional methods relying on random initialization. This raises an important question: Is there still a need to study traditional CIL? To address this inquiry, we not only reproduce state-of-the-art methods in PTM-based CIL but also modify several traditional methods to be compatible with PTMs. This enables a fair comparison between PTM-based methods and traditional methods.

We open-source the pre-trained model-based continual learning toolbox (PILOT) for the machine learning community. It includes several traditional CIL approaches modified by PTMs and offers state-of-the-art algorithms to advance PTM-based CIL research. The source code

of PILOT is available at <https://github.com/sun-hailong/LAMDA-PILOT>.

**Compared with other toolkits.** Since the current machine learning community lacks a toolbox that includes numerous PTM-based methods, there is an urgent need to develop a dedicated PTM-based toolbox. This toolbox will facilitate cutting-edge research and allow for fair comparisons between traditional methods and PTM-based methods using the same backbone. We primarily compare PILOT and other toolkits in the following three aspects.

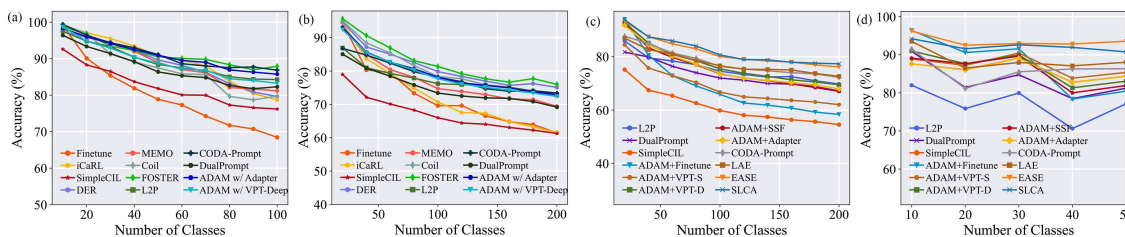
**Incorporation of PTMs.** PILOT not only encompasses traditional CIL methods but also extends support for the latest PTM-based CIL approaches. In contrast, other toolkits have mainly focused on conventional CIL methods and have not explored the integration of PTMs.

**Network architecture and parameter tuning.** By transitioning from the typical ResNet backbone to using PTMs, we design a unique parameter setting and tuning approach. While traditional toolkits can potentially be extended to accommodate PTMs, they are primarily designed with convolutional neural network (CNN) networks. Hence, the parameters and hyper-parameters suited for CNN might not be optimal for PTMs.

**Benchmarks and datasets.** We provide benchmarks and datasets specifically curated for scenarios involving PTMs. These dedicated resources can play a pivotal role in obtaining accurate performance metrics and evaluations tailored to PTM-based CIL.

**Implemented algorithms.** In PILOT, we implement 15 typical algorithms for CIL, including traditional methods modified by PTMs and PTM-based methods. Below, we list the latest PTM-based methods. Finetune involves continually training a pre-trained model on new tasks. It updates all parameters and is vulnerable to severe catastrophic forgetting. SimpleCIL [2] constructs classifiers continually by extracting prototype features using PTMs, without the need for additional training on the downstream task. L2P [3] incorporates visual prompt tuning into CIL using a pre-trained vision transformer and establishes a prompt pool for selecting the instance-specific prompts. DualPrompt [4] proposes

\* Corresponding author (email: [zhoudw@lamda.nju.edu.cn](mailto:zhoudw@lamda.nju.edu.cn), [yehj@lamda.nju.edu.cn](mailto:yehj@lamda.nju.edu.cn))



**Figure 1** (Color online) Reproduced incremental accuracy on CIFAR100, ImageNet-R, and VTAB. (a) and (b) utilize the ViT-B/16-IN1K backbone, while (c) and (d) employ the ViT-B/16-IN21K backbone. (a) CIFAR100, 10 stages; (b) ImageNet-R, 10 stages; (c) ImageNet-R, 10 stages; (d) VTAB, 5 stages.

two kinds of prompts based on L2P, i.e., general and expert prompts. CODA-Prompt [5] improves the prompt selection process with an attention mechanism. APER [2], based on SimpleCIL, employs parameter-efficient fine-tuning to acquire an adapted model. Subsequently, it concatenates the adapted model with the original one to obtain augmented features for constructing a prototype-based classifier. RanPAC [6] injects a frozen random projection layer with nonlinear activation to capture interactions between features with expanded dimensionality. SLCA [7] improves the classification layer by modeling the class-wise distributions and aligning the classification layers in a post-hoc fashion. LAE [8] defines the online and offline learning protocol, where the online model is updated with cross-entropy loss, aiming to acquire new knowledge in new tasks. EASE [9] designs an expandable subspace ensemble method for PTM-based CIL.

*Supported datasets.* Due to the overlap in data between ImageNet-based benchmarks and the pre-trained dataset, ImageNet is not an appropriate choice for assessing PTM-based CIL methods, we provide some novel benchmarks for CIL which (1) are entirely distinct from the ImageNet dataset, (2) present a significant domain gap from ImageNet, thereby challenging the PTM’s ability to generalize, and (3) encompass large-scale datasets from various domains to establish a cross-domain class-incremental benchmark. On the other hand, since pre-trained models may possess extensive knowledge of upstream tasks, we evaluate performance on CIFAR100, CUB200, ImageNet-R, ImageNet-A, ObjectNet, OmniBenchmark, and VTAB. These datasets represent typical CIL benchmarks and include out-of-distribution datasets that exhibit a significant domain gap with ImageNet (i.e., the pre-trained dataset). Specifically, there are 50 classes in VTAB, 100 classes in CIFAR100, 200 classes in CUB, ImageNet-R, ImageNet-A, and ObjectNet, and 300 classes in OmniBenchmark.

*Evaluation methodology.* In CIL, a widely used performance metric is the test accuracy at each incremental stage, defined as  $\mathcal{A}_b$ , where  $b$  represents the stage index. Another important metric is the average accuracy across all stages, given by  $\bar{\mathcal{A}} = \frac{1}{B} \sum_{b=1}^B \mathcal{A}_b$ . In this work, we evaluate the incremental performance (Top-1 accuracy) at each stage, with results shown in Figure 1. We utilize datasets such as CIFAR100, ImageNet-R, ObjectNet, and VTAB, dividing all classes into several incremental stages. Due to some missing parameters in some studies (e.g., L2P), we have optimized a suitable parameter set for these methods. It is encouraging to observe that most re-implemented algorithms either match or exceed the performance benchmarks of the original publication. Moreover, we find that although traditional

methods use PTM backbones and preserve some samples for replay, their performance is generally lower than PTM-based methods. This highlights the importance of leveraging pre-training techniques to design efficient CIL methods.

*Conclusion.* We have introduced PILOT, a pre-trained model-based continual learning toolbox. It includes a collection of reproduced PTM-based CIL methods and provides state-of-the-art algorithms for advanced research. PILOT aims to facilitate innovative research and development in the field of continual learning. In the future, we will continue to update our toolbox, expanding it to include more algorithms and datasets, and applying it to a wider range of settings.

**Acknowledgements** This work was partially supported by National Key R&D Program of China (Grant No. 2022ZD0114805), National Natural Science Foundation of China (Grant Nos. 62476123, 62376118, 62006112, 62250069, 61921006), Fundamental Research Funds for the Central Universities (Grant Nos. 2024300373, 14380021), Key Program of Jiangsu Science Foundation (Grant No. BK20243012), CCF-Tencent Rhino-Bird Open Research Fund (Grant No. RAGR20240101), and AI & AI for Science Project of Nanjing University, Collaborative Innovation Center of Novel Software Technology and Industrialization.

**Supporting information** The supporting information is available online at [info.scichina.com](http://info.scichina.com) and [link.springer.com](http://link.springer.com). The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

## References

- Zhou D W, Sun H L, Ning J, et al. Continual learning with pre-trained models: a survey. In: *Proceedings of IJCAI*, 2024. 8363–8371
- Zhou D W, Cai Z W, Ye H J, et al. Revisiting class-incremental learning with pre-trained models: generalizability and adaptivity are all you need. *Int J Comput Vis*, 2024. doi: 10.1007/s11263-024-02218-0
- Wang Z, Zhang Z, Lee C Y, et al. Learning to prompt for continual learning. In: *Proceedings of CVPR*, 2022. 139–149
- Wang Z, Zhang Z, Ebrahimi S, et al. DualPrompt: complementary prompting for rehearsal-free continual learning. In: *Proceedings of ECCV*, 2022. 631–648
- Smith J S, Karlinsky L, Gutta V, et al. CODA-Prompt: continual decomposed attention-based prompting for rehearsal-free continual learning. In: *Proceedings of CVPR*, 2023. 11909–11919
- McDonnell M D, Gong D, Parvaneh A, et al. RanPAC: random projections and pre-trained models for continual learning. In: *Proceedings of Advances in Neural Information Processing Systems*, 2024
- Zhang G, Wang L, Kang G, et al. SLCA: slow learner with classifier alignment for continual learning on a pre-trained model. In: *Proceedings of ICCV*, 2023. 19148–19158
- Gao Q, Zhao C, Sun Y, et al. A unified continual learning framework with general parameter-efficient tuning. In: *Proceedings of ICCV*, 2023. 11483–11493
- Zhou D W, Sun H L, Ye H J, et al. Expandable subspace ensemble for pre-trained model-based class-incremental learning. In: *Proceedings of CVPR*, 2024. 23554–23564