

# PyCIL: a Python toolbox for class-incremental learning

Da-Wei ZHOU<sup>†</sup>, Fu-Yun WANG<sup>†</sup>, Han-Jia YE<sup>\*</sup> & De-Chuan ZHAN*State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China*

Received 7 April 2022/Accepted 6 September 2022/Published online 19 April 2023

**Citation** Zhou D-W, Wang F-Y, Ye H-J, et al. PyCIL: a Python toolbox for class-incremental learning. *Sci China Inf Sci*, 2023, 66(9): 197101, <https://doi.org/10.1007/s11432-022-3600-y>

With the rapid development of deep learning, current deep models can learn a fixed number of classes with high performance. However, in our ever-changing world, data often come from the open environment, which is with stream format or available temporarily due to privacy issues. As a result, the classification model should learn new classes incrementally instead of restarting the training process. A straightforward approach is to finetune the model with the incoming new data, while it suffers catastrophic forgetting phenomena: due to the absence of previous data, the prediction on former classes drastically drops. Class-incremental learning (CIL) aims to extend the acquired knowledge with only new classes. For example, when training a robot in the open-world, it meets new objects as time goes by, and in the electronic commerce platform, new types of products appear daily. We give an example to demonstrate the setting of CIL. In the first task, the model needs to classify birds and dogs. After that, the model is incrementally updated with two new classes, i.e., tigers and fish, and it needs to classify among two old classes (birds and dogs) and two new classes (tigers and fish). Similarly, new classes like monkeys and sheep will emerge in the next task, requiring the model to incorporate them incrementally. New categories arrive progressively, and the model needs to classify more classes without forgetting the former ones.

With the growing interest of the machine learning community in class-incremental learning, it is essential to provide a simple and efficient toolbox with several class-incremental learning algorithms. We choose to conduct its development in the Python programming language for its wide use in the machine learning community. Its high-level interactive nature makes it an appealing tool for both academic and industrial software developments, and several popular machine learning libraries and deep learning open-source frameworks are built upon it.

The Python class-incremental learning (PyCIL) library takes advantage of Python to make class-incremental learning accessible to the machine learning community. It contains implementations of several founding studies of CIL and provides current state-of-the-art algorithms that can be

used to conduct novel fundamental research. As PyCIL is designed to be user-focused and friendly, we have kept our toolbox easy to use and accessible with convention consistencies and syntax over all the available functions. Moreover, our toolbox depends only on standard open-source libraries, and it is usable under many operating systems such as Linux, MacOSX, or Windows. The source code of PyCIL is available at <https://github.com/G-U-N/PyCIL>.

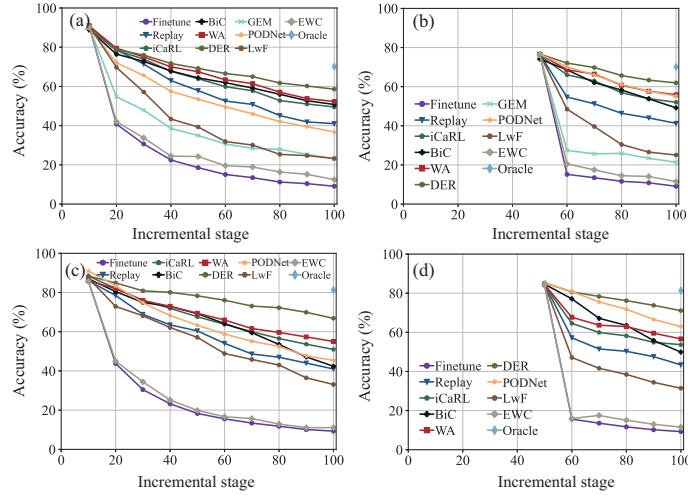
**Definition 1** (Class-incremental learning). Class-incremental learning was proposed to learn a stream of data incrementally from different classes. Assume there are a sequence of  $B$  training tasks  $\{\mathcal{D}^1, \mathcal{D}^2, \dots, \mathcal{D}^B\}$  without overlapping classes, where  $\mathcal{D}^b = \{(\mathbf{x}_i^b, y_i^b)\}_{i=1}^{n_b}$  is the  $b$ -th incremental step with  $n_b$  instances. Besides,  $\mathbf{x}_i^b \in \mathbb{R}^D$  is a training instance of class  $y_i \in Y_b$ ,  $Y_b$  is the label space of task  $b$ , where  $Y_b \cap Y_{b'} = \emptyset$  for  $b \neq b'$ . During the training process of task  $b$ , we can only access data from  $\mathcal{D}^b$ . The aim of CIL at each step is not only to acquire the knowledge from the current task  $\mathcal{D}^b$ , but also to preserve the knowledge from former tasks. After each task, the trained model is evaluated over all seen classes  $\mathcal{Y}_b = Y_1 \cup \dots \cup Y_b$ .

**Definition 2** (Exemplar set). In the  $b$ -th stage, typical CIL methods update the model with only the current dataset  $\mathcal{D}^b$ , which suffers severe catastrophic forgetting. As a result, current CIL methods propose to maintain an extra exemplar set  $\mathcal{E} = \{(\mathbf{x}_j, y_j)\}_{j=1}^M$ .  $\mathcal{E}$  helps to reserve a limited amount of instances for the classes seen before, and revisiting them can help the model overcome catastrophic forgetting. The exemplars are selected with the herding algorithm to make them more representative.

*Implemented algorithms.* In PyCIL, we implemented 11 typical algorithms for class-incremental learning. They are listed as follows. Finetune: The baseline method simply updates parameters on new tasks and suffers from severe catastrophic forgetting. Replay: The baseline method updates parameters on new tasks with instances from the new dataset and exemplar set. EWC [1]: Use the Fisher information matrix to weigh the importance of each parameter and regularize them to overcome forgetting. LwF [2]: Use knowl-

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

<sup>†</sup> Zhou D-W and Wang F-Y have the same contribution to this work.



**Figure 1** (Color online) Reproduced incremental accuracy on CIFAR100 and ImageNet100. (a) CIFAR100, 10 stages; (b) CIFAR100, B50, 5 stages; (c) ImageNet100, 10 stages; (d) ImageNet100, B50, 5 stages.

edge distillation to align the output probability between old and new models. iCaRL [3]: Based on LwF, it introduces an exemplar set for rehearsal and uses the nearest center mean classifier. GEM [4]: Use exemplars as the regularization of gradient updating. BiC [5]: Train an extra adaptation layer based on iCaRL, which adjusts the logits on new classes. WA [6]: Normalize the classifier weight after each learning session based on iCaRL. PODNet [7]: Introduce pooled outputs distillation to constrain the network. DER [8]: A two-stage learning approach utilizes a dynamically expandable representation for more effective incremental concept modeling. Coil [9]: Build bi-directional knowledge transfer in the incremental learning process with optimal transport.

Dependencies: PyCIL relies on open-source libraries such as NumPy and SciPy for linear algebra and optimization problems. The network structure is designed with PyTorch.

Basic usage: PyCIL provides implementations of the above 11 methods. As for the benchmark dataset setting in class-incremental learning, we provide the environment of CIFAR100 and ImageNet100/1000. When using PyCIL, users can edit the global parameters and algorithm-specific hyper-parameter, and then run the main function. The aforementioned global parameters include the following. Memory-Size: The total exemplar number in the incremental learning process. Init-Cls: The number of classes in the first incremental stage. Increment: The number of classes in each incremental stage  $b, b > 1$ . Convnet-type: The backbone network for the incremental model. Seed: The random seed for shuffling the class order, which is set to 1993 by default.

*Evaluation.* The common performance measure for CIL is the test accuracy after every stage, denoted by  $\mathcal{A}_b$ , where  $b$  is the stage index. Similarly, the averaged accuracy across all stages is also a common measure, i.e.,  $\bar{\mathcal{A}} = \frac{1}{B} \sum_{b=1}^B \mathcal{A}_b$ . As a preliminary step for research in the machine learning field, we have tested the incremental performance (Top-1 accuracy) along the incremental stages, and the results are shown in Figure 1. We use the benchmark datasets, i.e., CIFAR100 and ImageNet100, and divide the 100 classes into several incremental stages. Since some parameters are not reported in the original paper, we search for a good parameter set in our re-implementation. Most reproduced algorithms have the same or even better performance than the

results reported in the original paper.

*Conclusion.* We have presented PyCIL, a class-incremental learning toolbox written in Python. It contains implementations of a number of founding studies of CIL, but also provides current state-of-the-art algorithms that can be used to conduct novel fundamental research. Code consistency makes it an easy tool for research purposes, teaching, and industrial applications.

**Acknowledgements** This work was supported by National Key R&D Program of China (Grant No. 2020AAA0109401), National Natural Science Foundation of China (Grant Nos. 61773198, 61921006, 62006112), NSFC-NRF Joint Research Project (Grant No. 61861146001), Collaborative Innovation Center of Novel Software Technology and Industrialization, and Natural Science Foundation of Jiangsu Province (Grant No. BK20200313).

## References

- Kirkpatrick J, Pascanu R, Rabinowitz N, et al. Overcoming catastrophic forgetting in neural networks. *Proc Natl Acad Sci USA*, 2017, 114: 3521–3526
- Li Z Z, Hoiem D. Learning without forgetting. *IEEE Trans Pattern Anal Mach Intell*, 2018, 40: 2935–2947
- Rebuffi S A, Kolesnikov A, Sperl G, et al. iCaRL: incremental classifier and representation learning. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 2001–2010
- Lopez-Paz D, Ranzato M. Gradient episodic memory for continual learning. In: *Proceedings of International Conference on Neural Information Processing Systems*, 2017. 6467–6476
- Wu Y, Chen Y, Wang L, et al. Large scale incremental learning. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 374–382
- Zhao B, Xiao X, Gan G, et al. Maintaining discrimination and fairness in class incremental learning. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 13208–13217
- Douillard A, Cord M, Ollion C, et al. PODNet: pooled outputs distillation for small-tasks incremental learning. In: *Proceedings of European Conference on Computer Vision*, 2020. 86–102
- Yan S, Xie J, He X. DER: dynamically expandable representation for class incremental learning. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 3014–3023
- Zhou D W, Ye H J, Zhan D C. Co-transport for class-incremental learning. In: *Proceedings of ACM International Conference on Multimedia*, 2021. 1645–1654