# Supplementary material for "FeTrIL: Feature Translation for Exemplar-Free Class-Incremental Learning"

Grégoire Petit[1,2], Adrian Popescu[1], Hugo Schindler[1], David Picard[2], Bertrand Delezoide[3]
[1]Université Paris-Saclay, CEA, LIST, F-91120, Palaiseau, France
[2]LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France
[3]Amanda, 34 Avenue Des Champs Elysées, F-75008, Paris, France
{gregoire.petit, adrian.popescu}@cea.fr,hugo-schindler@orange.fr
david.picard@enpc.fr,bertrand.delezoide@amanda.com

## 1. Introduction

In this supplementary material, we provide:

- details about datasets used in experiments;
- comparison with other recent methods which were designed for memory-free class-incremental learning (MFCIL);
- implementation details for all tested approaches;
- extended results on the negative examples ratios;
- supplementary results for the use of a ratio between positives and negatives;
- working FeTrIL code for CIFAR-100 [6] with $T = 10$ states.

## 2. Datasets details

| Dataset | #Train | #Test | $\mu$(Train) | $\sigma$(Train) |
|---|---|---|---|---|
| CIFAR-100 [6] | 50,000 | 10,000 | 500.0 | 0.0 |
| TinyImageNet [7] | 100,000 | 10,000 | 500.0 | 0.0 |
| ImageNet-Subset [9] | 128856 | 5,000 | 1288.56 | 44.85 |
| ILSVRC [9] | 1,231,167 | 50,000 | 1231.2 | 70.2 |

Table 1. Summary of datasets. $\mu$ is the mean number of train images per class and $\sigma$ is the standard deviation

The datasets used in evaluation are designed for visual classification tasks. Their main statistics are in Table 1. Since the actual test subsets are not provided by the organizers of the ImageNet LSVRC competition, we follow common practice in incremental learning [8, 3, 5] and use the original validation subsets for the test phase.

## 3. Implementation details

When implementations of compared methods were available, we first tested them using the protocol and datasets from the original paper to make sure that we reproduced their results. We then used the authors' optimal parameters to test these methods in our evaluation setting. Note that for sake of fairness, all baselines were run using both training and validation sets (from Table 1). A ResNet-18 model [4] and an $SGD$ optimizer with $momentum = 0.9$ are used for all methods. We explicitly list the learning parameters of each method hereafter:

1. **Training the initial model:**

   This training regime is needed to obtain the initial model for each method, and also Joint training which can be considered the upper bound method where all classes are learned with all their data at once. We used the parameters provided by the authors as follows.

   Joint and the first models of FT and SIW are training using the parameters from [2]. Each model is learned for 120 epochs using $batch\ size = 256$ and $weight\ decay = 0.0001$. The $lr$ is set to 0.1 and is divided by 10 when the error plateaus for 10 epochs.

   The $lr$ is set to its initial value decayed by 10 every 30 epochs. The $lr$ is constrained to do not decrease beneath 0.001.

   For LUCIR, he first model is trained in the same manner than subsequent models (detailed below), following the original protocol from [5].

2. **Training the incremental models:**

   Here, we describe the hyper-parameters used to train the methods which were retrained in Table 1 of the main paper.

   - LUCIR [5] - all models are trained for 90 epochs using $lr = 0.1$, $batch\ size = 128$ and $weight\ decay = 0.0001$. The $lr$ is divided by

| CIL Method | CUB200 | | Flower102 | |
|---|---|---|---|---|
| | $T = 5$ | $T = 10$ | $T = 5$ | $T = 10$ |
| SDC[11] *(CVPR'20)* | 70.0 | 65.8 | 86.8 | 80.4 |
| FeTrIL[1] | **71.6** | **71.0** | **90.4** | **89.7** |

Table 2. Comparison of SDC [11] with FeTrIL[1] using the evaluation protocol for two supplemenary datasets used in [11]. **Best results in bold.**

| CIL Method | ImageNet50 | ImageNet100 |
|---|---|---|
| | $T = 5$ | $T = 20$ |
| ABD[10] *(ICCV'21)* | 71.5 | 12.1 |
| FeTrIL[1] | **89.0** | **39.0** |

Table 3. Comparison of ABD [10] with FeTrIL using the authors' evaluation protocol. ImageNet50 includes 50 classes and 5 states of 10 classes. ImageNet100 includes 100 classes, with 20 states of 5 classes each. Note that [10] uses top-5 accuracy for ImageNet50 and top-1 for ImageNet100 and we present the same numbers. **Best results in bold.**

10 at epochs 30 and 60. The method-specific parameters are the same as those from the original paper [5] and can also be found once we release the codes and configuration files.

- DeeSIL [1] - the initial model is the same one used for FeTrIL. The training of linear classifier is also done using the same parameters.

## 4. Comparison to other recent MFCIL methods

In Table 2, we compare FeTrIL SDC [11] using the evaluation protocol and datasets from [11]. Half of the datasets are assigned to the initial state and the rest of classes are split evenly among the remaining states. Following [11], the training of the initial FeTrILmodel for CUB200 and Flower102 datasets is initialized with a pretrained ILSVRC model. We do the same here to facilitate comparison with the original paper. The results from Table 2 indicate that FeTrIL [1] is clearly better than SDC [11] in all tested configurations. The better stability of FeTrIL results with the increase of the number of CIL states observed in the main submission is also confirmed for CUB200, Flower102, the three medium-scale datasets used in [11].

In Table 3, we present results obtained with FeTrIL and Always Be Dreaming (ABD) [10] a recent method which combines distillation and image inversion to address MF-CIL. The comparison is done for two ILSVRC [9] subsets which include 50 and 100 classes, respectively. We thank the authors of [10] for providing the lists of classes for these two subsets in a personal communication. FeTrIL outperforms ABD by a large margin in both configurations. This result is explained by difficulty of deploying image inversion in an efficient manner for visually complex images, such as those included in ImageNet.

## 5. Effect of a positives-to-negatives ratio

In addition to Figure 4 of the main submission, we present in Figure 1 the behavior of FeTrIL[1] when we approximate the negative example pool with different ratios $r$ for CIFAR-100 and TinyImageNet. These results confirm the observations made in the main paper since the accuracy trends observed when using a positives-to-negatives ratio is similar for all three datasets. This highlights the possibility to accelerate the training of FeTrIL with very limited accuracy loss.

Generally speaking, no EFCIL method can ensure a class separability comparable to that provided by standard learning with all images of all classes available simultaneously. The objective is to find a good balance between the stability and the plasticity of EFCIL representations. The experiments from the main paper show that, while imperfect, the combination of features and of pseudo-features used in FeTrIL[1] provides better performance compared to methods which update the model using variants of knowledge distillation and more complicated class prototypes.

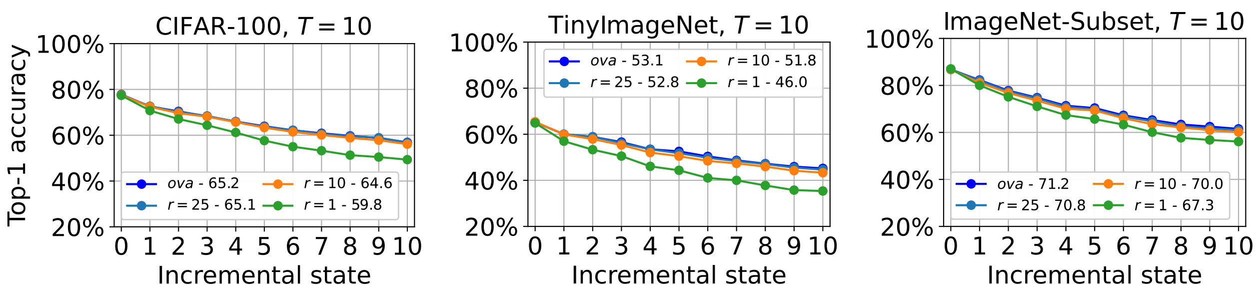


Figure 1. Top-1 incremental accuracy of FeTrIL[1] for approximate training of the classification layer with different ratios for negative sampling. $ova$ denotes a classical one-vs-all training procedure.

## References


[1] Eden Belouadah and Adrian Popescu. Deesil: Deep-shallow incremental learning. *TaskCV Workshop @ ECCV 2018.*, 2018.

[2] Eden Belouadah, Adrian Popescu, and Ioannis Kanellos. Initial classifier weights replay for memoryless class incremental learning. In *British Machine Vision Conference (BMVC)*, 2020.

[3] Francisco M. Castro, Manuel J. Marín-Jiménez, Nicolás Guil, Cordelia Schmid, and Karteek Alahari. End-to-end incremental learning. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XII*, pages 241–257, 2018.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Conference on Computer Vision and Pattern Recognition*, CVPR, 2016.

[5] Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and Dahua Lin. Learning a unified classifier incrementally via rebalancing. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 831–839, 2019.

[6] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

[7] Ya Le and Xuan Yang. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7):3, 2015.

[8] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H. Lampert. icarl: Incremental classifier and representation learning. In *Conference on Computer Vision and Pattern Recognition*, CVPR, 2017.

[9] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, and Fei-Fei Li. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[10] James Smith, Yen-Chang Hsu, Jonathan Balloch, Yilin Shen, Hongxia Jin, and Zsolt Kira. Always be dreaming: A new approach for data-free class-incremental learning. *arXiv preprint arXiv:2106.09701*, 2021.

[11] Lu Yu, Bartlomiej Twardowski, Xialei Liu, Luis Herranz, Kai Wang, Yongmei Cheng, Shangling Jui, and Joost van de Weijer. Semantic drift compensation for class-incremental learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 6980–6989. IEEE, 2020.