Supplementary Material for

Rainbow Memory: Continual Learning with a Memory of Diverse Samples

Jihwan Bang[1,*] Heesu Kim[2,3,*] YoungJoon Yoo[2,3] Jung-Woo Ha[2,3] Jonghyun Choi[4,†]
Search Solutions, Inc[1] NAVER CLOVA[2] NAVER AI Lab[3] GIST[4]
{jihwan.bang,heesu.kim89,youngjoon.yoo,jungwoo.ha}@navercorp.com, jhc@gist.ac.kr

1. Accuracy over the Tasks in Various CIL Setups

The evaluation set for CIL methods consists of only the seen classes. In disjoint setting, the number of seen classes increases when new tasks come, since classes of each task should be exclusive. Therefore, classes of evaluation sets increase as the task iterations proceed and the accuracy tends to decrease (see Figure 1a and 2a).

In blurry setting, on the other hand, the evaluation set comprises of entire classes as the tasks are not disjoint. Therefore, the model will see more data for each class as task iterations proceed; *e.g*., Figure 1b and 2b show the accuracy increases in later tasks in Blurry10 configuration. Interestingly, as the blurry ratio increases (*e.g*., from Blurry10 to Blurry30), the accuracy flattens for all tasks as shown in Figure 1c and 2c. We believe it is because the class frequency between minor and major classes in Blurry30 has less gap compared to Blurry10 so that the model can train well for all classes. Note that each task in the Blurry$M$ contains samples from its assigned major classes consisting of $(100 - M)\%$ and ones of minor classes consisting of remaining $M\%$.

Figure 1 and 2 show that our proposed approaches (RM w/o DA and RM) outperform other methods in the online setting, but the margin reduces or goes to negative in the offline setting as we mentioned in Section *4.2 Results* in the main paper. It is because blurry-online setting allows to see the sample of current task once, and reuse only the exemplars stored in the memory. Hence, managing diversity in the memory is more crucial compared to offline setting, and thus maximally exhibiting the efficacy of our approaches.

2. Metrics Details

We use three metrics in Section *4. Experiments* of the main paper; *Last accuracy (A)*, *Last forgetting (F)*, and *Intransigence (I)* defined in [1]. Here, we describe them in detail.

Last accuracy (A). Last accuracy reports an accuracy after entire training ends, thus it evaluates model over all classes being exposed during training.

Last forgetting (F). Forgetting measures how much the accuracy for each task is degraded (*i.e*., forgotten) compared to the best one in the training phases of previous tasks. Hence, last forgetting reports an averaged forgetting metrics over all tasks after entire training ends.

Intransigence (I). Intransigence measures the how much the accuracy for each task is achieved compared to the upper-bound, which comes from the non-CIL setting, then reports the average value for all tasks. Therefore, as model learns new knowledge, intransigence will be improved.

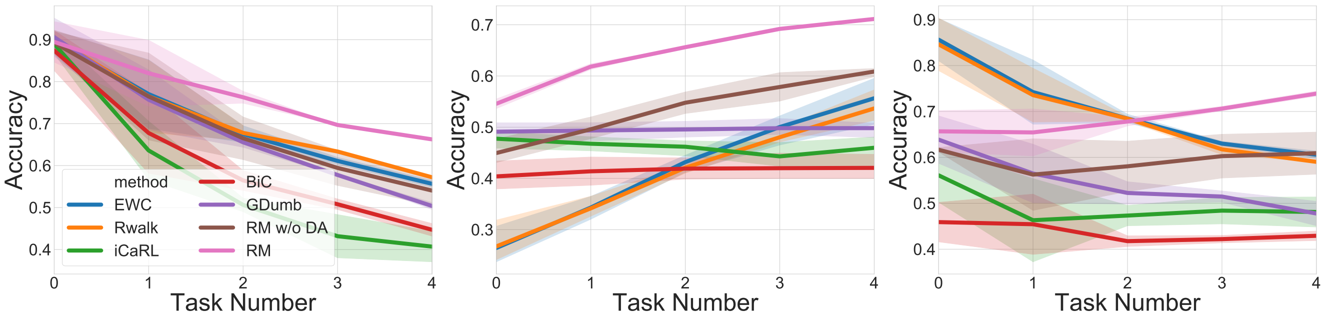
Table 1: Class splits for CIFAR10 CIL-benchmarks.

	Seed 1	Seed 2	Seed 3
Task 1	truck, automobile	airplane, dog	ship, airplane
Task 2	frog, airplane	ship, cat	dog, truck
Task 3	cat, bird	horse, truck	automobile, frog
Task 4	dog, horse	bird, frog	horse, cat
Task 5	deer, ship	automobile, deer	bird, deer

3. Class Distribution over Tasks

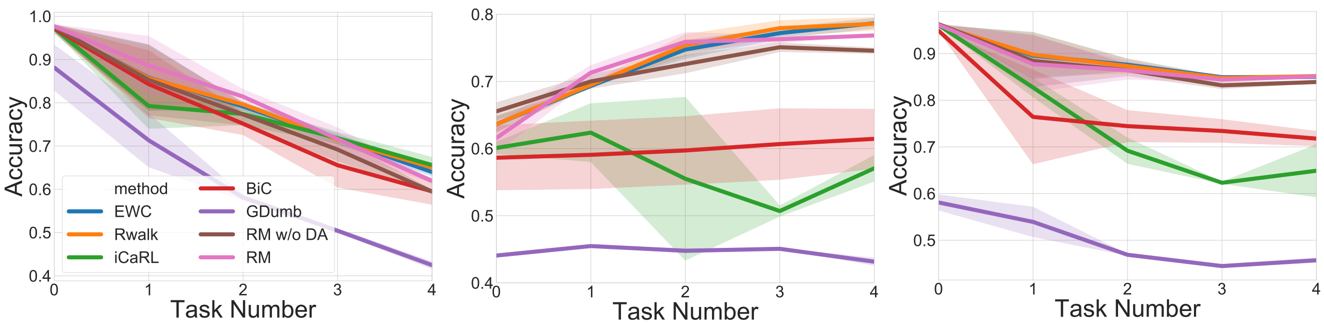
As we mentioned in Section *4.1 Experimental Setup* of the main paper, classes of CIFAR10 and CIFAR100 were randomly split into five tasks (2 and 20 classes per task, respectively), and classes of ImageNet were split into ten tasks to generate CIL-benchmark. Moreover, we iterated every experiments three times with different class splits from three different random seeds except for ImageNet. Here, we summarize the class splits of CIFAR10 CIL-benchmarks used for our experiments in Table 1. We will release the splits and other configuration along with the code in our github repo: https://github.com/clovaai/rainbow-memory.

* indicates equal contribution. † indicates corresponding author.



(a) CIFAR10-Disjoint-Online (b) CIFAR10-Blurry10-Online (c) CIFAR10-Blurry30-Online

Figure 1: Illustration of accuracy changes as tasks are being learned in (a) CIFAR10-Disjoint-Online, (b) CIFAR10-Blurry10-Online, (c) CIFAR10-Blurry30-Online settings.



(a) CIFAR10-Disjoint-Offline (b) CIFAR10-Blurry10-Offline (c) CIFAR10-Blurry30-Offline

Figure 2: Illustration of accuracy changes as tasks are being learned in (a) CIFAR10-Disjoint-Offline, (b) CIFAR10-Blurry10-Offline, (c) CIFAR10-Blurry30-Offline settings.

Table 2: Comparison of last accuracy (A5 (↑), %) over methods with data augmentations in CIFAR10-Blurry10-Online on $K = 1000$.

Methods	None	CutMix	RandAug	AutoAug	CutMix +AutoAug
EWC	68.6±0.9	70.5±0.6	73.0±0.5	75.1±2.2	75.2±0.0
Rwalk	68.2±1.8	69.7±1.0	73.5±0.1	76.0±4.0	76.2±0.4
iCaRL	53.6±2.8	56.1±2.6	57.7±0.7	62.5±6.1	63.8±1.1
GDumb	59.1±0.3	64.2±1.2	67.5±1.3	67.6±2.2	70.3±0.6
BiC	47.8±3.0	47.8±3.0	45.3±7.7	45.6±5.8	48.5±5.0
RM (Ours)	**70.9±1.5**	**74.7±0.7**	**76.4±0.4**	**77.5±0.7**	**78.0±0.5**

References

[1] A. Chaudhry, P. K. Dokania, T. Ajanthan, and P. H. S. Torr. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In *ECCV*, 2018. 1