Dual-Enhanced Coreset Selection with Class-wise Collaboration for Online Blurry Class Incremental Learning
– Supplementary Material –

Yutian Luo[1], Shiqi Zhao[2], Haoran Wu[2], Zhiwu Lu[1,*]
[1]Gaoling School of Artificial Intelligence, Renmin University of China, Beijing, China
[2]China Unicom Research Institute, Beijing, China
{luoyutian2021, luzhiwu}@ruc.edu.cn

1. Algorithms

Algorithm 1: Class-wise Balanced Memory

Input: the minibatch stream data $\mathcal{B}_t$ in task T_t
the class-wise saved number $\{q_{c_i}\}_{i=1}^{K}$
the class-wise seen number $\{z_{c_i}\}_{i=1}^{K}$
the memory buffer $\mathcal{M}_t$ and buffer size m
the random.choice function RC

for $(x, y) \in \mathcal{B}_t$ **do**
  $z_{c_i} \leftarrow z_{c_i} + 1, c_i = y$;
  if $|\mathcal{M}_t| < m$ **then**
    $\mathcal{M}_t.append(x, y)$
  else if $q_{c_i} < \frac{m}{K}$ **then**
    $c_k \leftarrow RC(\{c_j | q_{c_j} > \frac{m}{K}, 1 \leq j \leq K\})$;
    $(\hat{x}, \hat{y}) \leftarrow RC(\{(x', y') | (x', y') \in \mathcal{M}_t, y' = c_k\})$;
    $q_{c_k} \leftarrow q_{c_k} - 1, q_{c_i} \leftarrow q_{c_i} + 1$;
    $\mathcal{M}_t.delete(\hat{x}, \hat{y})$;
    $\mathcal{M}_t.append(x, y)$;
  else
    $c_k \leftarrow y$;
    $\mathcal{M}_t \leftarrow CW\_RSV(z_{c_k}, \frac{m}{K}, (x, y), \mathcal{M}_t^{c_k})$
end

Algorithm 2: Class-wise Reservoir (CW_RSV)

Input: the class-wise seen number z_{c_i}
the stream data (x, y)
the memory buffer $\mathcal{M}_t^{c_k}$
the class-wise maximum buffer size $\frac{m}{K}$

Output: the updated buffer $\mathcal{M}_t$

j = randint(0, z_{c_i});
if $j \leq \frac{m}{K}$ **then**
  $\mathcal{M}_t^{c_k}[j] \leftarrow (x, y)$
end
return the $\mathcal{M}_t$;

*Corresponding author.

Table 1. Average ablation results of different combinations of strategies in our devised diverse score guidance (DSG).

	MNIST		CIFAR10		CIFAR100	
Methods	FAA ↑	FF ↓	FAA ↑	FF ↓	FAA ↑	FF ↓
$STG-PPP$	89.25	1.48	48.44	0.67	30.55	2.54
$STG-PPF$	89.30	1.18	52.78	0.53	30.62	**2.21**
$STG-PCP$	89.91	1.38	56.34	0.58	32.22	4.05
$STG-PCF$ (**DSG**)	**90.20**	**1.17**	**57.36**	**0.51**	**33.07**	4.40

2. Ablation on Strategies in DSG

In Section 3.4 of the main paper, we analyzed and demonstrated the effectiveness of DSG's specific design elements, including CW_RSV and the exclusive use of $\mathcal{CW}\_\mathcal{A}$. However, we still need to evaluate the effectiveness of each strategy — $STG-P$, $STG-C$, and $STG-F$ — in DSG and the improvements achieved by their combinations.

In Table 1, we present additional ablation results for various strategy combinations in DSG, tested on MNIST, CIFAR10, and CIFAR100. We use $STG-PPP$ as our baseline, which applies the sum of all class-wise scores for guiding $\mathcal{C}^P$, $\mathcal{C}^C$, and $\mathcal{C}^F$. It can be observed that: (1) Combinations $STG-PPF$ and $STG-PCP$ both improve the final performance of the model on three datasets, indicating that either $STG-F$ or $STG-C$ make a positive contribution to the critical coreset selection. (2) Compared to $STG-PCP$, the combination $STG-PPF$ makes a more obvious effect in reducing the final forgetting, while $STG-PCP$ makes a more significant improvement to the final average accuracy. (3) By combining all our designed strategies, $STG-PCF$ (*i.e.*, DSG) reaches the highest accuracy and lowest forgetting, which further proves that all these strategies work well together and consistently improve the final average accuracy and reduce the final forgetting of the model under the OBCIL. Overall, all these observations verify the effectiveness of each individual strategy and their combined implementation in DSG.

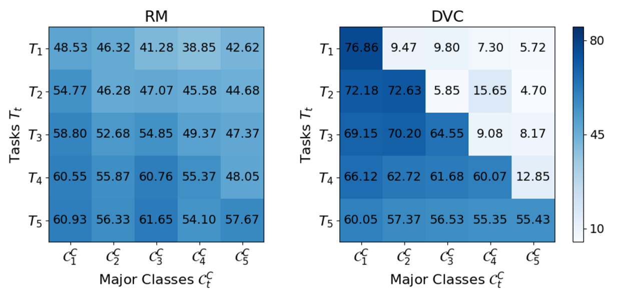

Figure 1. Average accuracy of the model in task-wise major classes with RM and DVC on CIFAR10.

Table 2. Average ablation results of our method and other competitors with different buffer sizes on MNIST.

	$\|M\| = 200$		$\|M\| = 500$		$\|M\| = 1000$	
Methods	FAA ↑	FF ↓	FAA ↑	FF ↓	FAA ↑	FF ↓
Gdumb [7]	84.61	1.67	88.66	2.31	91.56	0.56
ER [4]	86.71	6.16	88.06	9.20	91.91	5.23
MIR [1]	86.83	4.02	88.76	6.30	91.92	4.50
OCS [8]	86.42	5.90	89.12	5.90	91.83	5.12
ER-ACE [3]	86.28	5.92	89.25	6.20	91.76	5.43
DVC [6]	86.68	6.41	88.90	6.93	91.32	5.76
RM† [2]	87.52	3.52	90.24	1.02	91.82	0.95
DECO	**88.45**	**-0.15**	**90.89**	**0.72**	**92.55**	**0.48**

3. Extra Experiment Results

3.1. Additional Task-wise Results

In Figure 5 of the main paper, we present the task-wise results of the model using our DECO and MIR methods to clarify the reasons behind the significant differences in final forgetting and to highlight the superiority of our DECO in final task-wise average accuracy. For additional comparisons, we provide the task-wise results of the second-best method RM (with balanced memory) and the third-best method DVC (without balanced memory) in Figure 1.

Comparing the results in Figure 5 and Figure 1, it can be observed that: (1) Although DVC effectively reduces final forgetting and achieves higher final task-wise average accuracy than MIR, it still lags behind the final task-wise results of RM and DECO. This indicates that a balanced memory is the key to minimizing the final forgetting under the OBCIL setting. (2) Although RM also enables continual learning in all classes like our DECO, DECO reaches both higher initial task-wise accuracy and higher final task-wise accuracy, which again demonstrates that our DECO is indeed superior to any other competitor method. Overall, all these results prove that our method is more effective than other competitors under the OBCIL setting.

3.2. Additional Ablation on Buffer Size

In Table 3 of the main paper, we present the average ablation results related to buffer size for all methods on CIFAR10. For a comprehensive analysis, we also include the average ablation results on MNIST and CIFAR100 in Table 2 and Table 3, respectively. It is obvious that our DECO consistently reaches the highest final average accuracy and also keeps the lowest final forgetting in most scenarios on both datasets, regardless of buffer size variations. These results validate the good generalization ability of our DECO across various coreset sizes.

Table 3. Average ablation results of our method and other competitors with different buffer sizes on CIFAR100.

	$\|M\| = 500$		$\|M\| = 1000$		$\|M\| = 2000$	
Methods	FAA ↑	FF ↓	FAA ↑	FF ↓	FAA ↑	FF ↓
Gdumb [7]	11.13	0.67	14.84	2.61	26.58	7.17
ER [4]	15.38	12.75	21.83	13.95	32.08	12.55
MIR [1]	16.02	13.11	22.53	14.83	33.06	13.58
OCS [8]	15.53	12.88	22.32	13.87	32.48	12.13
ER-ACE [3]	15.93	13.12	22.40	13.98	32.83	11.03
DVC [6]	16.04	14.80	22.42	14.04	32.98	13.97
RM† [2]	16.33	**0.52**	22.70	-0.27	32.89	3.59
DECO	**16.84**	1.07	**23.32**	**-0.35**	**33.93**	**3.35**

Table 4. Average ablation results of our method and other competitors combined with augmentation strategy RandAug.

	CIFAR10		**CIFAR100**		**ImageNet**	
Methods	FAA ↑	FF ↓	FAA ↑	FF ↓	FAA ↑	FF ↓
Gdumb [7]	54.13	1.68	27.60	7.69	27.86	3.98
ER [4]	64.71	11.59	36.01	16.75	42.56	14.93
MIR [1]	65.40	11.71	37.23	16.80	45.03	12.86
OCS [8]	66.23	11.31	37.33	15.31	45.74	12.53
ER-ACE [3]	65.73	11.66	37.46	15.37	46.89	11.10
DVC [6]	67.47	14.75	37.16	17.86	46.32	12.56
RM† [2]	68.03	-0.37	38.21	4.39	47.68	1.92
DECO	**69.43**	**-0.60**	**38.93**	**3.96**	**50.67**	**1.30**

3.3. Additional Ablation on Augmentation Effects

As data augmentation strategies are often used to enhance training in many rehearsal-based methods, we examine the impact of two types of data augmentation on all competitors as well as on our DECO. According to the results shown in Table 4 and Table 5, both RandAug [5] and "CutMix+AutoAug" [2] improve the performance of all methods. Notably, DECO not only achieves the best performance but also surpasses other methods by a wider margin. Overall, these findings demonstrate the superiority of DECO over other existing methods in generalization ability when combined with various data augmentation strategies.

Acknowledgements This work was supported by National Natural Science Foundation of China (62376274).

References

[1] Rahaf Aljundi, Lucas Caccia, Eugene Belilovsky, Massimo Caccia, Min Lin, Laurent Charlin, and Tinne Tuytelaars. On-

Table 5. Average ablation results of our method and other competitors combined with augmentation strategy CutMix+AutoAug.

	CIFAR10		CIFAR100		ImageNet	
Methods	FAA ↑	FF ↓	FAA ↑	FF ↓	FAA ↑	FF ↓
Gdumb [7]	58.02	3.35	32.02	7.76	31.30	3.72
ER [4]	68.07	10.08	38.52	15.62	46.77	14.16
MIR [1]	68.68	10.21	39.34	15.24	48.82	11.92
OCS [8]	68.93	10.78	39.07	14.89	48.99	12.37
ER-ACE [3]	69.17	10.51	39.44	15.79	49.51	10.80
DVC [6]	69.94	13.00	38.90	14.71	49.03	11.34
RM† [2]	71.10	1.39	41.04	4.62	51.32	1.89
DECO	**73.78**	**-0.05**	**41.93**	**4.00**	**55.34**	**1.62**

line continual learning with maximally interfered retrieval. *ArXiv*, abs/1908.04742, 2019. 2, 3

[2] Jihwan Bang, Heesu Kim, YoungJoon Yoo, Jung-Woo Ha, and Jonghyun Choi. Rainbow memory: Continual learning with a memory of diverse samples. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8218–8227, 2021. 2, 3

[3] Lucas Caccia, Rahaf Aljundi, Nader Asadi, Tinne Tuytelaars, Joelle Pineau, and Eugene Belilovsky. New insights on reducing abrupt representation change in online continual learning. *arXiv preprint arXiv:2104.05025*, 2021. 2, 3

[4] Arslan Chaudhry, Marcus Rohrbach, Mohamed Elhoseiny, Thalaiyasingam Ajanthan, Puneet K Dokania, Philip HS Torr, and Marc'Aurelio Ranzato. On tiny episodic memories in continual learning. *arXiv preprint arXiv:1902.10486*, 2019. 2, 3

[5] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *IEEE/CVF conference on Computer Vision and Pattern Recognition workshops*, pages 702–703, 2020. 2

[6] Yanan Gu, Xu Yang, Kun Wei, and Cheng Deng. Not just selection, but exploration: Online class-incremental continual learning via dual view consistency. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7442–7451, 2022. 2, 3

[7] Ameya Prabhu, Philip HS Torr, and Puneet K Dokania. Gdumb: A simple approach that questions our progress in continual learning. In *European Conference on Computer Vision*, pages 524–540. Springer, 2020. 2, 3

[8] Jaehong Yoon, Divyam Madaan, Eunho Yang, and Sung Ju Hwang. Online coreset selection for rehearsal-based continual learning. *arXiv preprint arXiv:2106.01085*, 2021. 2, 3