Dual-consistency Model Inversion for Non-exemplar Class Incremental Learning (Supplementary Materials)

Zihuan Qiu[1] Yi Xu[2] Fanman Meng[1] Hongliang Li[1] Linfeng Xu[1] Qingbo Wu[1]

[1]University of Electronic Science and Technology of China [2]Dalian University of Technology

{zihuanqiu@std.,fmmeng@,hlli@,lfxu@,qbwu@}uestc.edu.cn yxu@dlut.edu.cn

In this supplementary material, we provide additional explanations §A and results §B that cannot fit into the main paper due to the page limit. Finally, we discuss the limitations of our method in §C.

A. Additional Explanations

A.1. The Details of Benchmark Datasets

To achieve a comprehensive study, we conduct extensive experiments in the main paper, including datasets CIFAR-100 [7], Tiny-ImageNet [8], and ImageNet-Subset [1], and ImageNet-Full [1].

CIFAR-100 is a well-known image classification dataset that contains 32×32 images for 100 classes. The training set contains 50,000 images with 500 images per class, and the validation set contains 10,000 images with 100 per class. ImageNet-Full is a large-scale classification dataset with 1,000 classes, that contains about 1.2 million images for training and 50,000 images for validation. Tiny-ImageNet is a subset of 200 classes from ImageNet, with image size rescaled to 64×64. The training set contains 100,000 images with 500 per class. The validation and evaluation set both contain 10,000 images with 50 per class. ImageNet-subset is a 100-class subset from ImageNet, where each class contains 1,300 training images and 50 validation images.

A.2. The Details of Evaluation Metrics

Following most previous works [10, 11], we report average incremental accuracy A_N and average forgetting F_N as the main metric. The A_N reflects the average performance of the model on all tasks, while F_N measures the ability to resist catastrophic forgetting. A desired CIL learner needs to simultaneously achieve a high A_N and a low F_N.

Average incremental accuracy is computed as the average result of the accuracy a_i of all phases (initial and incremental):

$$A_N = \frac{1}{N+1}\sum_{i=0}^{N} a_i. \tag{S.1}$$

Average forgetting is defined as:

$$F_N = \frac{1}{N}\sum_{i=0}^{N-1} f_N^i, \tag{S.2}$$

where $f_N^i = \max_{t\in i,\dots,N-1}(a_{t,i} - a_{N,i})$ and $a_{m,n}$ is the accuracy of task n after training task m. f_N^i reflects the accuracy drop of task i between the peak accuracy $a_{t,i}$ and the accuracy $a_{N,i}$ of last phase.

A.3. The Details of Training

We utilize RandomResizedCrop, RandomHorizontalFlip, and ColorJitter for data augmentation, similar to [10]. During the initial task training phase, we optimize all model parameters. For incremental updates, only the parameters of the last stage are updated to mitigate the risk of forgetting previously learned knowledge. In training the generator and discriminator, we employ the WGAN framework with gradient penalty [4].

A.4. Proof for Main Paper Eq.2

Let $\phi_{\mathcal{D}_o}$ and $\phi_{\hat{\mathcal{D}}_o}$ be the density functions of $\mathcal{D}_o$ and $\hat{\mathcal{D}}_o$ respectively. Then, we have:

$$\begin{aligned}
&\epsilon_{\mathcal{D}_n}(h_n,f_n)+\epsilon_{\mathcal{D}_o}(h_n,f_o)\\
&=\epsilon_{\mathcal{D}_n}(h_n,f_n)+\epsilon_{\mathcal{D}_o}(h_o,f_o)-\epsilon_{\mathcal{D}_o}(h_o,f_o)+\epsilon_{\mathcal{D}_o}(h_n,f_o)\\
&\le\epsilon_{\mathcal{D}_n}(h_n,f_n)+\epsilon_{\mathcal{D}_o}(h_o,f_o)+|\epsilon_{\mathcal{D}_o}(h_n,f_o)-\epsilon_{\mathcal{D}_o}(h_o,f_o)|\\
&\le\epsilon_{\mathcal{D}_n}(h_n,f_n)+\epsilon_{\mathcal{D}_o}(h_o,f_o)+\mathrm{E}_{\mathbf{x}\sim\mathcal{D}_o}[|h_n(\mathbf{x})-h_o(\mathbf{x})|]\\
&=\epsilon_{\mathcal{D}_n}(h_n,f_n)+\epsilon_{\mathcal{D}_o}(h_o,f_o)+\epsilon_{\hat{\mathcal{D}}_o}(h_n,h_o)-\epsilon_{\hat{\mathcal{D}}_o}(h_n,h_o)\\
&\qquad+\epsilon_{\mathcal{D}_o}(h_n,h_o)\\
&\le\epsilon_{\mathcal{D}_n}(h_n,f_n)+\epsilon_{\mathcal{D}_o}(h_o,f_o)+\epsilon_{\hat{\mathcal{D}}_o}(h_n,h_o)\\
&\qquad+|\epsilon_{\mathcal{D}_o}(h_n,h_o)-\epsilon_{\hat{\mathcal{D}}_o}(h_n,h_o)|\\
&\le\epsilon_{\mathcal{D}_n}(h_n,f_n)+\epsilon_{\mathcal{D}_o}(h_o,f_o)+\epsilon_{\hat{\mathcal{D}}_o}(h_n,h_o)\\
&\qquad+\int|\phi_{\mathcal{D}_o}(\mathbf{x})-\phi_{\hat{\mathcal{D}}_o}(\mathbf{x})||h_n(\mathbf{x})-h_o(\mathbf{x})|d\mathbf{x}\\
&\le\epsilon_{\mathcal{D}_n}(h_n,f_n)+\epsilon_{\mathcal{D}_o}(h_o,f_o)+\epsilon_{\hat{\mathcal{D}}_o}(h_n,h_o)+d_1(\hat{\mathcal{D}}_o,\mathcal{D}_o)
\end{aligned}$$

B. Additional Results

B.1. Detailed Values of the Curves

For a fair comparison with subsequent work, we provide the detailed values of all accuracy curves in Fig. 5 of the main paper. The results are listed in Tab. S.3, S.4, S.5.

B.2. Results on Modified 32-layer ResNet

In tab. S.2, we report average incremental accuracy A_N and last phase accuracy A_L on CIFAR-100 using a modified 32-layer ResNet [5]. The means and standard deviations are reported of three runs. In Fig. S.2, we present a comprehensive comparison between DCMI and two preceding inversion methods, ABD [9] and RDFCIL [3]. The results show that our method achieves comparable performance with the previous methods under 5 phases, while notably surpassing them under 10 and 25 phases. This highlights the superiority of our method, particularly in scenarios involving long sequence increments.

B.3. Sensitive Analysis of Hyper-parameter

We perform a sensitivity analysis on the hyper-parameter λ in Eq.13 of the main paper. As illustrated in Fig. S.1, the results of 5 and 10 phases are less sensitive to λ than 20 phases. Optimal performance is obtained when selecting $\lambda = 0.5$.

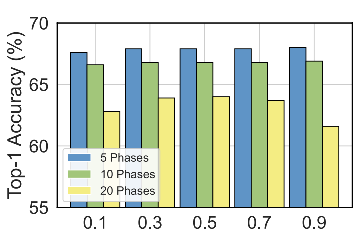

Figure S.1. Sensitive analysis of λ on CIFAR-100.

B.4. Time complexity

We evaluate the training time by comparing our method with non-generative (PASS [10]) and generative (ABD [9]) approaches with the same training epochs, as shown in Tab. S.1. ABD experiences prolonged training times due to its generator requiring larger epochs, and PASS is relatively inefficient due to LabelAug in each incremental task. Our method demonstrates efficiency in comparison.

Method	CIFAR-100			Tiny-ImageNet		
	P=5	*P=10*	*P=20*	*P=5*	*P=10*	*P=20*
PASS	774	389	236	5978	3018	1535
ABD	750	599	542	3462	2433	1809
Ours	542	342	272	3095	1653	894

Table S.1. Comparison of the training time (s) during each phase.

C. Limitations

Similar to other generative methods, DCMI requires time to train the generator. Meanwhile, replaying generated data inevitably increases computational costs. Additionally, the training of the generator relies on new class data, which imposes certain requirements on the amount of new class data, making it difficult to apply with the few shot increments. A potential solution is to involve domain-consistent extra data. Finally, the current methodology may not seamlessly extend to other critical tasks, such as segmentation and detection.

References

[1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, K. Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. 1

[2] Arthur Douillard, Matthieu Cord, Charles Ollion, Thomas Robert, and Eduardo Valle. Podnet: Pooled outputs distillation for small-tasks incremental learning. In *European Conference on Computer Vision*, 2020. 3

[3] Qiankun Gao, Chen Zhao, Bernard Ghanem, and Jian Zhang. R-dfcil: Relation-guided representation learning for data-free class incremental learning. In *European Conference on Computer Vision*, pages 423–439. Springer, 2022. 2, 3

[4] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30, 2017. 1

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[6] Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and Dahua Lin. Learning a unified classifier incrementally via rebalancing. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 831–839, 2019. 3

[7] Alex Krizhevsky. Learning multiple layers of features from tiny images. 2009. 1

[8] Ya Le and Xuan S. Yang. Tiny imagenet visual recognition challenge. 2015. 1

[9] James Smith, Yen-Chang Hsu, John C. Balloch, Yilin Shen, Hongxia Jin, and Zsolt Kira. Always be dreaming: A new approach for data-free class-incremental learning. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9354–9364, 2021. 2, 3

[10] Fei Zhu, Xu-Yao Zhang, Chuan Wang, Fei Yin, and Cheng-Lin Liu. Prototype augmentation and self-supervision for incremental learning. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5867–5876, 2021. 1, 2

[11] Kai Zhu, Kecheng Zheng, Ruili Feng, Deli Zhao, Yang Cao, and Zheng-Jun Zha. Self-organizing pathway expansion for non-exemplar class-incremental learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19204–19213, 2023. 1

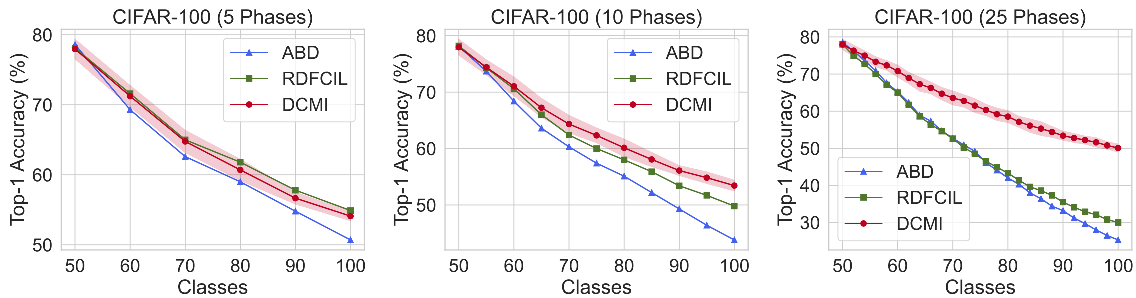

Figure S.2. Accuracy for each phase on CIFAR-100 using a modified 32-layer ResNet.

Method	*P=5*		*P=10*		*P=20*	
	$A_N(\uparrow)$	$A_L(\uparrow)$	$A_N(\uparrow)$	$A_L(\uparrow)$	$A_N(\uparrow)$	$A_L(\uparrow)$
UCIR [6]	$65.6_{\pm 1.0}$	$55.7_{\pm 0.9}$	$63.5_{\pm 1.1}$	$53.2_{\pm 0.7}$	$60.3_{\pm 1.1}$	$50.1_{\pm 0.4}$
PODNet [2]	$66.8_{\pm 1.3}$	$56.2_{\pm 1.0}$	$63.9_{\pm 1.1}$	$52.5_{\pm 0.6}$	$61.6_{\pm 1.0}$	$49.1_{\pm 0.3}$
ABD [9]	$62.4_{\pm 1.2}$	$50.6_{\pm 1.1}$	$59.0_{\pm 1.9}$	$43.7_{\pm 2.4}$	$48.9_{\pm 1.9}$	$25.3_{\pm 1.1}$
R-DFCIL [3]	$64.8_{\pm 1.6}$	$54.8_{\pm 0.8}$	$61.7_{\pm 1.2}$	$49.7_{\pm 0.6}$	$50.0_{\pm 0.8}$	$30.0_{\pm 0.6}$
DCMI	$64.2_{\pm 1.2}$	$54.1_{\pm 0.6}$	$63.6_{\pm 1.3}$	$53.4_{\pm 0.9}$	$62.0_{\pm 1.5}$	$50.1_{\pm 1.0}$

Table S.2. Average accuracy on CIFAR-100 using a modified 32-layer ResNet. Results from [3].

Dataset	*Phase*					
	0	1	2	3	4	5
CIFAR-100	79.5	72.6	68.8	65.4	62.0	59.3
Tiny-ImageNet	65.7	58.8	55.8	52.5	49.3	46.5
ImageNet-Subset	84.9	76.8	71.4	67.5	62.8	59.2

Table S.3. Detailed results (%) of classification accuracy under 5 phases.

Dataset	*Phase*										
	0	1	2	3	4	5	6	7	8	9	10
CIFAR-100	79.5	75.7	72.2	69.5	67.8	66.1	64.3	62.8	60.5	59.1	57.6
Tiny-ImageNet	65.7	61.5	58.6	56.1	55.5	53.6	51.7	50.2	48.1	46.6	45.2
ImageNet-Subset	84.9	81.0	76.6	73.6	70.7	69.4	67.0	64.4	62.3	60.8	58.9
ImageNet-Full	76.4	72.2	68.7	65.9	63.0	60.6	58.4	56.4	54.5	53.0	51.5

Table S.4. Detailed results (%) of classification accuracy under 10 phases.

Dataset	*Phase*									
	0	1	2	3	4	5	6	7	8	9
CIFAR-100	81.1	77.8	76.2	73.8	72.3	70.2	68.6	66.7	65.1	63.8
Tiny-ImageNet	65.7	63.0	61.4	60.0	58.4	57.4	55.6	55.0	54.4	53.1
ImageNet-Subset	86.2	83.1	78.7	76.3	73.8	72.2	70.1	67.7	66.1	65.6

Dataset	*Phase*										
	10	11	12	13	14	15	16	17	18	19	20
CIFAR-100	63.0	61.6	60.9	59.5	58.2	57.2	55.8	54.5	53.6	52.6	51.9
Tiny-ImageNet	52.3	51.2	50.3	49.2	48.2	47.2	46.0	45.1	43.9	42.9	42.1
ImageNet-Subset	63.5	63.1	62.2	61.1	59.2	57.7	56.2	55.0	53.7	53.0	51.6

Table S.5. Detailed results (%) of classification accuracy under 20 phases.