Supplementary Materials for Feature Drift Oriented Distribution Reconstruction for Imbalanced Class Incremental Learning

Tingmin Li[a], Fengqiang Wan[a], Yipeng Lin[a] and Yang Yang [a,*]

[a]Nanjing University of Science and Technology

1 Performance under Extreme Class imbalance

To validate the performance of FDDR under extreme class imbalance scenarios, we conducted experiments using a total buffer size of 500 on both ImageNet100 and CIFAR100 datasets. The results on the ImageNet100 dataset are reported in Table 1, showing that FDDR achieves the best performance under the B0 10 steps setting and is comparable to DER under B0 20 steps and B50 5 steps settings. Notably, as shown in Table 3, FDDR does not introduce any additional parameters during inference. In contrast, DER requires a larger number of parameters, highlighting the efficiency of FDDR in real-world deployment scenarios. The detailed incremental accuracy across different learning phases is illustrated in Figure 1.

Table 1. Performance on ImageNet100 dataset with a limited buffer size of $\mathcal{M} = 500$.

Methods	ImageNet100 B0				ImageNet100 B50	
	10 steps		20 steps		5 steps	
	Avg	Last	Avg	Last	Avg	Last
DER w/o P	75.09	63.48	**72.13**	**58.56**	**79.87**	71.36
FOSTER B4	70.43	56.22	60.83	50.02	78.33	68.30
FOSTER	68.99	53.66	59.81	49.92	76.77	68.26
MAFDRC	76.52	65.92	70.22	57.08	68.00	58.28
FDDR	**78.89**	**69.26**	71.34	57.30	79.28	**73.30**

Table 2 summarizes the results on the CIFAR100 dataset. The proposed FDDR achieves the second-best performance, outperforming the SOTA competitor MAFDRC across all settings. The detailed accuracies for each incremental step under various settings are presented in Figure 2, which further supports this conclusion.

2 Hyperparameter Analysis

Analysis of the hyperparameter α of MAF distill loss. To investigate the impact of MAF distill loss on model performance, we varing $\alpha = 3.0, 3.5, 4.0, 4.5, 5.0, 5.5$ under the B0 10 steps setting on CIFAR100 dataset. As shown in Table 4, the model achieves the highest Last accuracy and Avg accuracy when $\alpha = 4.0$. Moreover, we observe a performance degradation when α is either increased or decreased beyond this value, suggesting that $\alpha = 4.0$ provides an optimal trade-off between stability and plasticity.

* Corresponding Author. Email: yyang@njust.edu.cn.

Table 2. Performance on CIFAR100 dataset with buffer size of $\mathcal{M} = 500$.

Methods	CIFAR100 B0				CIFAR100 B50	
	10 steps		20 steps		5 steps	
	Avg	Last	Avg	Last	Avg	Last
DER w/o P	**72.51**	**61.09**	**69.20**	**56.45**	**69.07**	**61.73**
FOSTER B4	67.10	53.69	61.37	45.88	68.02	58.88
FOSTER	65.82	49.88	60.10	45.43	64.09	55.73
MAFDRC	69.11	53.99	65.61	47.14	59.86	50.64
FDDR	70.41	56.14	66.34	48.37	68.25	60.33

Table 3. The number of parameters used during inference phase for various CL methods under B0 10 steps setting on ImageNet100 dataset.

Method	DER w/o P	FOSTER B4	FOSTER	MAFDRC	FDDR
Params	112,277,220	22,455,524	11,227,812	11,229,860	11,227,812

Analysis of hyperparamter β of GCR loss. The hyperparameter β is introduced to constrain the structure of historical distributions and further alleviate the feature drift phenomenon. As shown in Table 5, we vary β in the range from 0.0 to 0.5, and observe that the best performance is achieved when $\beta = 0.1$.

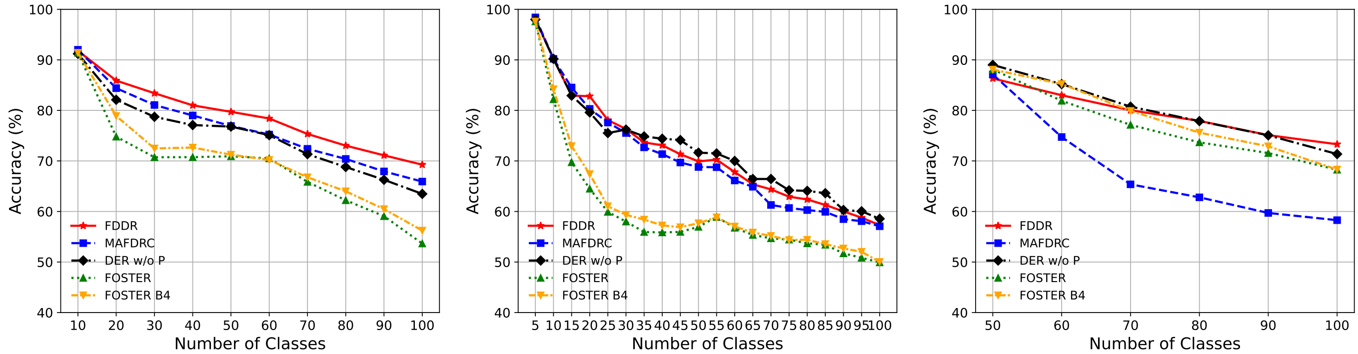
Table 4. Performance of different α on CIFAR100 B0 10 steps setting.

α	3.0	3.5	4.0	4.5	5.0	5.5
Avg(%)	74.12	74.26	**74.40**	73.95	73.82	73.43
Last(%)	61.78	61.90	**62.72**	62.20	62.24	61.94

Analysis of mixing hyperparameter η. To evaluate the impact of the hyperparameter η on model performance, we conduct an ablation study by varying η from 0 to 0.5 with a step size of 0.1 under the B0 10 steps setting on the CIFAR100 dataset. As shown in Table 6, the performance remains relatively stable across the tested range, demonstrating the robustness of the proposed method with respect to η. Moreover, incorporating the evolved historical prototypes consistently improves performance, indicating that the proposed JPS strategy facilitates more accurate similarity estimation, thereby contributing to enhanced overall performance.

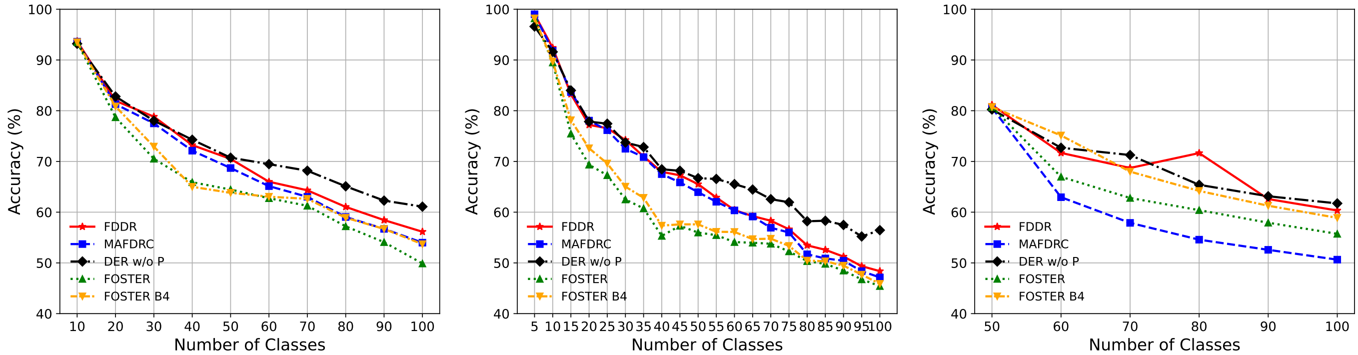
3 Visualization of Confusion Matrix

To facilitate a comprehensive comparison, we visualize the confusion matrices of various methods at the final incremental step un-



(a) ImageNet100 B0 10 Steps (b) ImageNet100 B0 20 Steps (c) ImageNet100 B50 5 Steps

Figure 1. Classification accuracy of each incremental task on ImageNet100 dataset.



(a) CIFAR100 B0 10 Steps (b) CIFAR100 B0 20 Steps (c) CIFAR100 B50 5 Steps

Figure 2. Classification accuracy of each incremental task on CIFAR100 dataset.

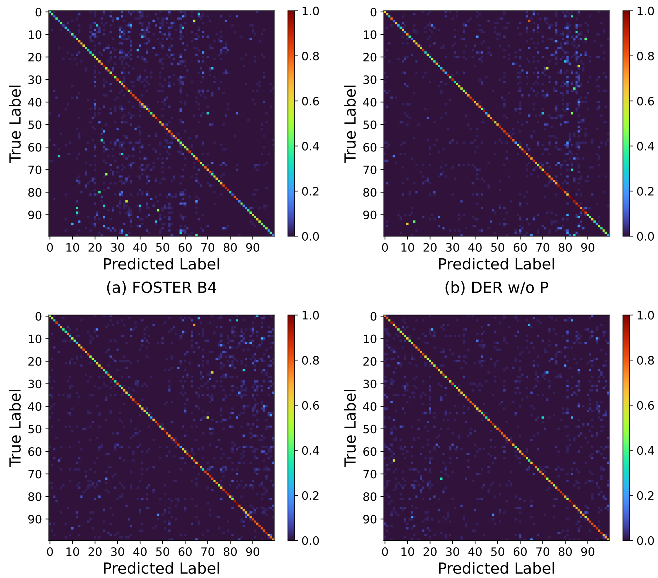
Table 5. Ablation study on the hyperparameter β of GCR loss. We conduct experiments on the CIFAR100 B0 10 steps setting.

β	0.0	0.1	0.2	0.3	0.4	0.5
Avg(%)	74.08	**74.40**	74.26	74.01	74.18	73.88
Last(%)	62.51	**62.72**	62.11	62.01	62.46	61.98

Table 6. Comparison of Avg accuracy and Last accuracy on CIFAR100 B0 10 steps setting with the η from 0.0 to 0.5.

η	0.0	0.1	0.2	0.3	0.4	0.5
Avg(%)	74.05	74.30	**74.40**	74.27	74.26	74.18
Last(%)	61.86	**62.73**	62.72	62.55	62.51	62.11

der the ImageNet100 B0 10 steps setting, as illustrated in Figure 3. In these matrices, warm colors correspond to higher prediction frequencies, whereas cool colors indicate lower ones. It can be clearly observed that FOSTER B4, DER, and MAFDRC exhibit pronounced brightness across many columns, indicating the severity of misclassifications. Moreover, DER and MAFDRC show a strong tendency to misclassify samples into new classes, which reflects a significant classification bias between historical and new classes, and highlights the catastrophic forgetting of historical classes. In contrast, the confusion matrix of our proposed method, FDDR, shows a more uniform distribution of colors across all classes, indicating its effectiveness in mitigating forgetting and maintaining a balanced classification performance between old and new classes.



(a) FOSTER B4 (b) DER w/o P (c) MAFDRC (d) FDDR

Figure 3. Confusion matrices of different methods at the last stage under ImageNet100 B0 10 steps setting.