Supplementary Material of
Dynamic Residual Classifier for Class Incremental Learning

Xiuwei Chen[1], Xiaobin Chang[1,2,3*]
[1]School of Artificial Intelligence, Sun Yat-sen University, China
[2]Guangdong Key Laboratory of Big Data Analysis and Processing, Guangzhou 510006, P.R.China
[3]Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China

A. Additional Detailed Analysis Results

The average accuracy of the CIFAR100 / ImageNet100 B0 5, 20 steps settings are presented, in addition to Sec. 4.3.

Contributions of Different Components As shown in Tab. I, clear improvements are observed with our DRC.

Components			CIFAR100		ImageNet100	
MAF	DRC	LA	5 steps	20 steps	5 steps	20 steps
			65.82	64.66	70.62	66.53
✓			71.24	67.01	77.49	68.50
✓	✓		74.43	71.50	82.16	73.67
✓	✓	✓	74.87	71.75	82.22	75.21

Table I. Ablation of MAFDRC components, complement to Tab. 4.

Branch Layer Merging Tab. II shows the effectiveness of residual classifier and branch layer merging.

Methods	CIFAR100		ImageNet100	
	5 steps	20 steps	5 steps	20 steps
MAF	71.24	67.01	77.49	68.50
MAFRC	75.12	71.23	82.51	75.11
MAFDRC	74.87	71.75	82.22	75.21

Table II. The results of MAF with residual classifier and dynamic residual classifier, complement to Tab. 5.

CIL Pipelines with DRC As shown in Tab. III, DRC consistently improves all pipelines under the new settings.

Model	CIFAR100		ImageNet100	
	5 steps	20 steps	5 steps	20 steps
MAF	71.24	67.01	77.49	68.50
+DRC	74.43(↑3.19)	71.50(↑4.49)	82.16(↑4.67)	73.67(↑5.17)
MEC	69.02	67.52	71.49	67.61
+DRC	70.64(↑1.62)	69.29(↑1.77)	73.60(↑2.11)	68.40(↑0.79)
MDT	68.24	67.01	73.07	67.12
+DRC	70.08(↑1.84)	68.51(↑1.50)	73.18(↑0.11)	68.39(↑1.27)

Table III. Results of different pipelines with DRC, complement to Tab. 6 in the main text.

Data Imbalanced Methods As shown in Tab. IV, DRC is superior to its counterparts, BFT and WA, in handling the data imbalance in CIL.

Methods	CIFAR100		ImageNet100	
	5 steps	20 steps	5 steps	20 steps
MAF	71.24	67.01	77.49	68.50
MAF+BFT	71.36	67.58	78.21	68.76
MAF+WA	71.39	67.70	79.62	68.71
MAF+DRC	**74.43**	**71.50**	**82.16**	**73.67**

Table IV. Results of different imbalanced methods, complement to Tab. 7.

Impact of Memory size Tab. V shows more results under the memory size of 20 exemplars per class. Results in the memory size of 1,000 exemplars are shown in Tab. VI.

Methods	CIFAR100		ImageNet100	
	5 steps	20 steps	5 steps	20 steps
DER w/o P	**74.10**	**67.06**	**79.85**	**72.97**
FOSTER B4	**74.55**	66.65	78.64	**71.31**
FOSTER	72.34	66.06	76.62	70.45
MAFDRC	73.19	**66.91**	**81.31**	70.07

Table V. CIL results with reserving 20 exemplars per class, complement to Tab. 8.

Methods	CIFAR100			ImageNet100		
	5 steps	10 steps	20 steps	5 steps	10 steps	20 steps
DER w/o P	**74.37**	**73.07**	**72.57**	**79.67**	**76.90**	**75.54**
FOSTER B4	73.04	71.16	67.86	75.58	72.99	71.41
FOSTER	70.34	69.63	66.92	72.88	71.55	70.48
MAFDRC	**73.46**	**71.62**	**67.91**	**81.20**	**78.39**	**72.37**

Table VI. CIL results with a memory size of 1,000 exemplars.

Impacts of tro To further mitigate classification bias, our method adopts adjusted loss, logit adjustment (LA) [35], as mentioned in Sec. 4.3. Specifically, an adjusting vector γ_t

Methods	ImageNet100 B0						ImageNet100 B50				ImageNet1000	
	5 steps		10 steps		20 steps		5 steps		10 steps		10 steps	
	Avg	Last	Avg	Last	Avg	Last	Avg	Last	Avg	Last	Avg	Last
iCaRL [36]	71.25	60.02	65.82	50.86	61.07	44.66	58.90	49.34	48.59	41.70	53.32	33.96
BiC [42]	71.24	60.98	64.93	45.98	56.18	32.32	61.65	44.98	53.71	37.38	-	-
WA [48]	73.89	64.06	68.00	54.48	61.96	46.02	62.78	54.24	52.84	45.70	-	-
PODNet [15]	72.53	58.68	62.85	44.84	54.88	36.86	75.33	66.58	72.91	62.90	-	-
DER w/o p [43]	**77.57**	**71.28**	**75.49**	**66.34**	**72.87**	**64.92**	**77.36**	**70.82**	**75.59**	**68.94**	**66.74**	**57.84**
FOSTER B4 [41]	75.81	69.66	71.37	62.58	66.42	**54.18**	77.15	70.06	73.75	63.98	64.15	45.03
FOSTER [41]	74.90	68.54	71.07	63.14	66.48	55.34	76.89	70.92	73.86	64.50	63.14	44.90
MAFDRC	**78.48**	**71.28**	**74.31**	**63.48**	**66.84**	52.16	**78.00**	**70.94**	**75.51**	**68.04**	**67.40**	**57.26**

Table VII. ImageNet Results without AutoAugment [6].

Methods	CIFAR100 B0						CIFAR100 B50			
	5 steps		10 steps		20 steps		5 steps		10 steps	
	Avg	Last	Avg	Last	Avg	Last	Avg	Last	Avg	Last
iCaRL [36]	66.28	53.45	64.69	48.69	64.06	47.22	52.04	43.96	43.25	37.70
BiC [42]	65.03	54.19	62.21	47.69	61.16	40.91	58.36	44.40	55.71	42.93
WA [48]	66.07	55.51	65.66	51.25	65.05	48.63	60.86	51.99	55.84	48.12
PODNet [15]	63.69	50.18	55.98	38.04	48.68	29.10	64.69	55.38	**63.11**	52.76
DER w/o P [43]	**71.35**	**63.55**	**71.02**	**60.18**	**69.75**	**57.22**	**68.09**	**61.90**	**66.36**	**59.70**
FOSTER B4 [41]	67.95	56.17	64.20	49.86	60.80	45.67	**65.47**	**57.27**	61.81	**53.17**
FOSTER [41]	65.77	54.94	62.70	49.14	59.94	45.04	64.48	55.65	61.26	51.71
MAFDRC	**70.12**	**59.05**	**67.72**	**54.24**	**65.84**	**50.76**	64.53	56.33	60.95	52.49

Table VIII. CIFAR100 Results without AutoAugment [6].

is used to adjust the original classifier logit $\bar{\ell}_t$,

$$\eta_t = \bar{\ell}_t + \gamma_t, \tag{i}$$

and the adjusted one, η_t, can be used to compute losses. γ_t is computed as,

$$\gamma_t^i = \log\left[\left(\frac{m_i}{m_1 + m_2 + \cdots + m_t}\right)^{tro}\right], \tag{ii}$$

where m_i is the sample number of the $i^{th}, i = \{1, ..., t\}$ task in new data D_t or memory buffer M_t. tro is a hyper-parameter of LA and is fixed at 1.2. As shown in Tab. IX, our method is not sensitive to different tros.

tro	1.0	1.2	1.4	1.6	1.8	2.0
Avg	71.04	71.75	71.99	71.75	71.76	71.00

Table IX. Impacts of different tros on model performance. CIFAR100 B0 20 steps results are reported.

B. Additional Main Results

In the main text, different CIL methods are trained with AutoAugment [6]. Their results are reported in Tab. 1 and Tab. 2. In this section, we reproduce the corresponding results without such data augmentation, as shown in Tab. VII and Tab. VIII, respectively. The proposed MAFDRC still achieves the same level of performance as the state-of-the-art methods, e.g., FOSTER [41] and DER [43]. In particular, under the CIL setting of ImageNet1000 10 steps, our method achieves similar results compared to DER w/o P but with a much smaller model size.

Vision Transformer (ViT) Backbone The DyTox ViT encoder replaces ResNet as the feature extractor, resulting in MAF(ViT) and MAFDRC(ViT). As shown in Tab. X, DRC clearly boosts baseline results and performs similarly to DyTox.

Methods	CIFAR100 B0					
	5 steps		10 steps		20 steps	
	Avg	Last	Avg	Last	Avg	Last
MAF(ViT)	63.83	44.84	58.58	36.69	57.11	32.39
DyTox†	71.78	61.31	69.63	55.65	67.00	50.85
MAFDRC(ViT)	72.48	62.78	69.49	56.13	66.73	48.65

Table X. CIL results with ViT-based feature extractor. † indicates the DyTox results are reproduced under the same setting (e.g., data augmentation and trained on a single GPU) with others.

Averaged Results of Three Runs As shown in Tab. XI and Tab. XII, the proposed method achieves the SOTA-level performance and is statistically better than its counterparts.

Methods	5 steps		10 steps		20 steps	
	Avg	Last	Avg	Last	Avg	Last
DER w/o P	$81.64_{\pm 0.62}$	$75.23_{\pm 0.69}$	$79.54_{\pm 1.10}$	$70.55_{\pm 0.52}$	$78.12_{\pm 0.11}$	$70.91_{\pm 0.60}$
FOSTER B4	$80.31_{\pm 0.67}$	$73.28_{\pm 0.61}$	$77.10_{\pm 0.51}$	$67.93_{\pm 0.87}$	$74.39_{\pm 0.21}$	$62.65_{\pm 0.66}$
FOSTER	$79.35_{\pm 0.94}$	$71.87_{\pm 0.48}$	$76.51_{\pm 0.28}$	$67.09_{\pm 0.53}$	$74.20_{\pm 0.30}$	$62.78_{\pm 0.37}$
MAFDRC	$82.31_{\pm 0.33}$	$76.05_{\pm 0.30}$	$79.78_{\pm 0.28}$	$70.56_{\pm 0.39}$	$75.63_{\pm 0.25}$	$64.14_{\pm 0.71}$

Table XI. Three runs results (averaged accuracy ± standard error) on ImageNet100 B0. Red indicates the best performance and blue indicates the second best results.

Methods	5 steps		10 steps		20 steps	
	Avg	Last	Avg	Last	Avg	Last
DER w/o P	$75.19_{\pm 0.61}$	$68.78_{\pm 0.34}$	$74.84_{\pm 0.81}$	$65.74_{\pm 0.55}$	$73.85_{\pm 0.64}$	$62.92_{\pm 0.73}$
FOSTER B4	$73.96_{\pm 0.77}$	$64.78_{\pm 0.71}$	$72.91_{\pm 0.74}$	$61.54_{\pm 0.90}$	$70.40_{\pm 0.57}$	$56.87_{\pm 0.66}$
FOSTER	$71.85_{\pm 1.01}$	$62.58_{\pm 0.99}$	$71.58_{\pm 0.62}$	$60.39_{\pm 0.28}$	$69.41_{\pm 0.55}$	$56.09_{\pm 0.93}$
MAFDRC	$74.81_{\pm 0.05}$	$66.16_{\pm 0.26}$	$73.85_{\pm 0.16}$	$61.98_{\pm 0.25}$	$71.93_{\pm 0.34}$	$57.43_{\pm 0.34}$

Table XII. Three runs results (averaged accuracy ± standard error) on CIFAR100 B0. Red indicates the best performance and blue indicates the second best results.