Lifelong Learning via Progressive Distillation and Retrospection

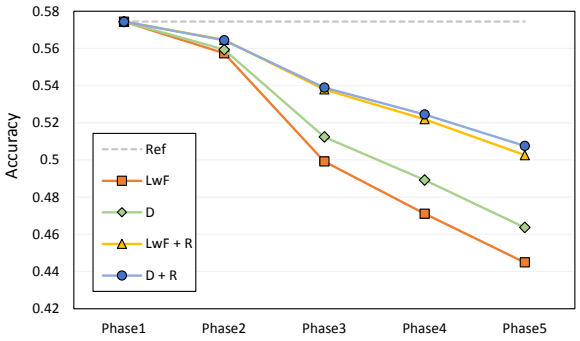
Saihui Hou[1][0000−0003−4689−2860], Xinyu Pan[2][0000−0001−9951−4842], Zilei Wang[1][0000−0003−1822−3731], Chen Change Loy[2][0000−0001−5345−1591], and Dahua Lin[2][0000−0002−8865−7896]

[1] Department of Automation, University of Science and Technology of China
[2] Department of Information Engineering, The Chinese University of Hong Kong

1 Appendix

1.1 Accuracy Degradation on ImageNet

Figure 1 displays the trends of accuracy degradation on ImageNet in another sequence of five-task scenario presented in the paper, which shows that the curve of *Distillation+Retrospection* goes down in the slowest rate.



(a) Imagenet→Flowers→Aircrafts→Scenes→Birds.

Fig. 1: Accuracy degradation on ImageNet in five-task scenario. D for *Distillation*, and R for *Retrospection*.

1.2 Comparison with *Encoder-based-LwF*

Table 1 shows the comparison with *Encoder-based-LwF* [1] in two-task scenario. ImageNet→Birds and Flowers→Scenes are adopted as the benchmarks, which are also evaluated in [1]. The performance with *Encoder-based-LwF* implemented by us basically agree with those presented in [1]. From the results in Table 1,

Table 1: Classification accuracy (%) for comparison with *Encoder-based-LwF*. The reference results are respectively given by *LwF* and *LwF+R*. *D* for *Distillation*, *R* for *Retrospection*, and *Encoder* for the approach in [1].

	Imagenet → Birds			Flowers → Birds		
	ImageNet	Birds	Average	Flowers	Scenes	Average
LwF [2]	54.49 (ref)	57.45 (ref)	55.97 (ref)	84.86 (ref)	61.87 (ref)	73.36 (ref)
LwF + Encoder [1]	54.99 (+0.50)	57.10 (-0.34)	56.05 (+0.08)	85.41 (+0.55)	62.16 (+0.30)	73.79 (+0.43)
DT (**ours**)	55.34 (+0.84)	58.21 (+0.76)	**56.77 (+0.80)**	85.36 (+0.50)	62.31 (+0.45)	73.84 (+0.48)
DT + Encoder (**ours**)	55.52 (+1.03)	57.84 (+0.40)	56.68 (+0.71)	86.16 (+1.30)	62.54 (+0.67)	**74.35 (+0.99)**

	Imagenet → Birds			Flowers → Birds		
	ImageNet	Birds	Average	Flowers	Scenes	Average
LwF+R	55.61 (ref)	57.79 (ref)	56.70 (ref)	85.31 (ref)	62.54 (ref)	73.92 (ref)
LwF+Encoder+R	55.76 (+0.15)	57.41 (-0.38)	56.59 (-0.11)	86.19 (+0.88)	62.91 (+0.37)	74.55 (+0.63)
DT+R (**ours**)	55.85 (+0.24)	59.55 (+1.76)	57.70 (+1.00)	85.73 (+0.42)	64.03 (+1.49)	74.88 (+0.96)
DT+Encoder+R (**ours**)	55.95 (+0.34)	59.55 (+1.76)	**57.75 (+1.05)**	86.39 (+1.07)	63.88 (+1.34)	**75.13 (+1.21)**

we get similar observations to those described in the paper. Specifically, the performance with *Distillation* is superior or at least comparable to that with *Encoder-based-LwF*. The two methods are orthogonal and their combination can further improve the performance. Besides, *Retrospection* is still helpful with an auto-encoder [1] incorporated for the old task.

1.3 Ablation Study on Retrospection Strategy

Here we provide the detailed results for the ablation study on *Retrospection* strategy. The experiments are conducted on ImageNet→Birds and Flowers→Birds. As shown in Table 2, the performance on the old task rises as the number of reserved images for each class increases. Besides, choosing images close the class center does not show significant superiority to random selection.

Table 2: Ablation study on *Retrospection* strategy. *Random* for random selection, and *Center* for selecting images close to the class center.

		#Images per class reserved					
		1	2	5	10	20	All
ImageNet	*Random*	54.57	55.10	55.61	55.87	56.00	56.27
	Center	54.54	54.99	55.64	55.82	56.06	56.27
Birds	*Random*	58.55	58.12	57.79	57.88	57.28	57.17
	Center	58.60	58.26	58.33	57.09	57.66	57.17
Flowers	*Random*	83.19	83.84	85.15	86.22	-	86.91
	Center	83.47	84.09	85.32	86.30	-	86.91
Birds	*Random*	57.28	57.01	56.79	57.26	-	56.47
	Center	56.94	56.50	56.60	56.17	-	56.47

References

1. Rannen Ep Triki, A., Aljundi, R., Blaschko, M., Tuytelaars, T.: Encoder based lifelong learning. In: ICCV. (2017)
2. Li, Z., Hoiem, D.: Learning without forgetting. IEEE Transactions on Pattern Analysis and Machine Intelligence (2017)