Appendix

This document contains additional experimental results and analysis.

A. ANCL implementation with TA

We follow the provided implementation of ANCL [22] based on the FACIL [32] framework. We only apply the adaptation to the main teacher network fitted on the previous data, and, in cases where we use warmup, we only apply it to the new head of the main network. The auxiliary network does not undergo the teacher adaptation process, as it is already fit for the current task, and we do not see a reason to apply the warmup to this network. We keep the same training schedule and hyperparameters as described in Section 4.1.

B. Warmup phase

As mentioned in Section 4.1, we experimented with applying the warmup phase, a technique that recently appeared in several works [7, 27, 49]. With this method, before training the full student model on the new data, we first finetune the classification head added for this new task in isolation, while the rest of the network is kept frozen. Compared to the standard practice of finetuning the full student model on the new task, starting with a warmup phase reduces the initial cross-entropy loss at the start of the new task, and avoids overwriting the knowledge from the previous tasks with large gradient updates caused by random classification head initialization.

For training the new head during the warmup phase, we use SGD optimizer and OneCycle scheduler [48] with cosine annealing, maximum learning rate of 0.1 and the number of epochs of increasing learning rate set as 40. We train the new head for 200 epochs with early stopping and freeze updates of batch normalization statistics in the model during the training.

In Appendix D, we evaluate different CL benchmarks with TA and warmup separately. We generally find that the warmup works complementary to our method when we use smaller networks such as ResNet32 and train on smaller datasets. We hypothesize that for the settings with a larger dataset and a larger network better initialization of the new head does not translate to better performance in CIL because the overall magnitude of updates required to train the model in such settings is greater and leads to overwriting the initialization over the course of the training.

C. Ablation studies of Teacher Adaptation

C.1. Teacher Adaptation with different batch sizes.

Larger batch sizes should lead to better estimation of the global statistics in the batch normalization layer, which, in theory, should lead to reduced effectiveness of TA. To investigate the impact of the batch size on our method, we train the model on CIFAR100 split into 10 tasks with different batch sizes, tuning the learning rate and λ for each batch size. We present the results in Table 6. Regardless of the batch size, TA constantly improves upon standard training, and the difference between the two approaches does not scale with batch size.

Batch size	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
32	28.94±0.29	44.24±0.27	32.57±0.92	23.89±1.19
+*TA*	**32.58±0.80**	**45.69±0.35**	**25.44±0.94**	**20.92±0.53**
64	28.92±0.57	43.89±0.27	30.48±1.11	21.70±0.39
+*TA*	**32.15±1.24**	**44.71±0.57**	**23.43±1.97**	**19.16±0.67**
128	28.27±0.44	42.52±0.76	29.59±0.92	22.26±0.31
+*TA*	**31.92±0.86**	**44.09±0.97**	**22.65±1.32**	**19.41±0.60**
256	25.13±0.46	40.44±0.11	40.06±0.38	31.39±0.98
+*TA*	**30.56±0.68**	**43.28±0.23**	**28.24±0.64**	**23.94±0.59**
512	23.01±0.89	37.44±0.34	40.57±1.20	32.03±0.69
+*TA*	**27.52±1.03**	**39.86±0.45**	**29.31±0.86**	**25.03±0.55**

Table 6. Teacher adaptation with different batch sizes. Our method improves upon the standard training regardless of the batch size.

C.2. Different CNN architectures with Teacher Adaptation.

To test the robustness of our method, in addition to experiments with ResNets, we test it with other standard convolutional neural networks. We evaluate our method with MobileNetV2 [43], MobileNetV3-small [19] and VGG11 [46]. We use CIFAR100 and ImageNet100 split into 10 equally sized tasks and train the networks as described in Section 4.1. We show the results of those experiments in the Table 7. The final results obtained with tested networks are lower than with our standard baselines, as do not tune their hyperparameters extensively, but we still notice that using TA constantly improves the results, regardless of the architecture we use.

C.3. Batch normalization statistics investigation.

We investigate the batch normalization statistics in the teacher and student model throughout CIL on CIFAR100 split into 10 equally sized tasks. To measure the divergence between the normalization statistics in the models, we compute the average Kullback-Leibler divergence (KLD) be-

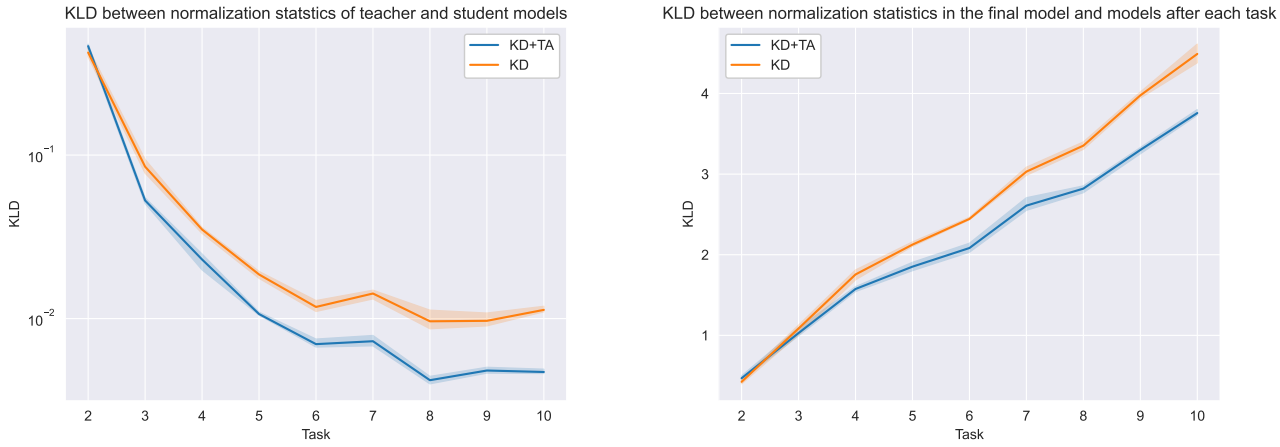

Figure 4. TA impact on batch normalization statistics. Left: KLD between the normalization statistics of the teacher and the student at the end of the training of all the tasks. Right: KLD between the normalization statistics of the model at the end of each task and the model trained on the first task.

CIFAR100				
Network	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
MobilenetV2	12.47±0.69	25.71±1.05	49.86±0.61	41.23±1.25
+*TA*	**20.62±1.01**	**31.91±0.57**	**26.66±1.34**	**22.31±0.56**
MobilenetV3-s	16.59±0.09	29.86±0.06	38.89±0.42	30.99±0.12
+*TA*	**21.22±0.91**	**32.24±0.60**	**20.98±0.13**	**16.84±1.19**
ResNet32	28.27±0.44	42.52±0.76	29.59±0.92	22.26±0.31
+*TA*	**31.92±0.86**	**44.09±0.97**	**22.65±1.32**	**19.41±0.60**
VGG11	17.10±0.12	34.34±0.39	62.36±0.62	52.99±0.97
+*TA*	**24.87±0.35**	**42.50±0.15**	**43.96±0.62**	**32.92±0.51**
ImageNet100				
Network	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
ResNet18	40.99±0.45	54.82±0.56	31.99±0.87	25.90±0.65
+*TA*	**42.69±0.53**	**55.84±0.51**	**25.26±0.23**	**20.43±0.33**
VGG11	24.35±0.32	44.72±0.50	61.62±0.10	46.73±0.31
+*TA*	**30.67±0.31**	**48.81±0.29**	**50.04±0.76**	**37.60±0.56**

Table 7. Teacher Adaptation with different network architectures on CIFAR100 and ImageNet100 split into 10 equally sized tasks.

tween the distributions of normalization statistics in every batch normalization layer. We show the results of this analysis in the Figure 4. Specifically, we measure KLD between the final teacher and student model at the end of each task, as well as the difference between the final student model trained for each task and the model initially learned after the first task. Applying TA leads to reduced KLD between the statistics in both teacher and student, and also student and initial model. This proves that our method leads to more stable representations throughout the training.

C.4. Hyperparameters for alternative adaptation methods ablation

The best results reported in Table 5 were obtained with teacher learning rate of 1e-7. For variants with teacher pertaining, the best results were obtained with training for 5 epochs. We keep λ fixed at 10 for all methods reported in the Table. While all variants that allow adaptation of batch norm statistics improve upon the baseline, the values of best hyperparameters are generally small, suggesting that the main source of improvement in all the methods comes from changes in normalization. This is further supported by the fact that all improvements vanish when normalization statistics are fixed.

C.5. Task recency bias with TA

We also conduct additional analysis of our method of Teacher Adaptation (TA) to understand the mechanism with which it improves upon the standard knowledge distillation. At Figure 5, we analyze task confusion matrices of standard knowledge distillation (LwF) and its extension with TA. We find that applying TA results in a model that is better at distinguishing between the tasks, and generally exhibits lower recency bias.

We hypothesize that the lower KD loss that we observe when using TA results in smaller updates to the model, so the difference between the magnitudes of logits learned for different tasks is smaller. Therefore, Teacher Adaptation helps to alleviate the recency bias in class-incremental learning.

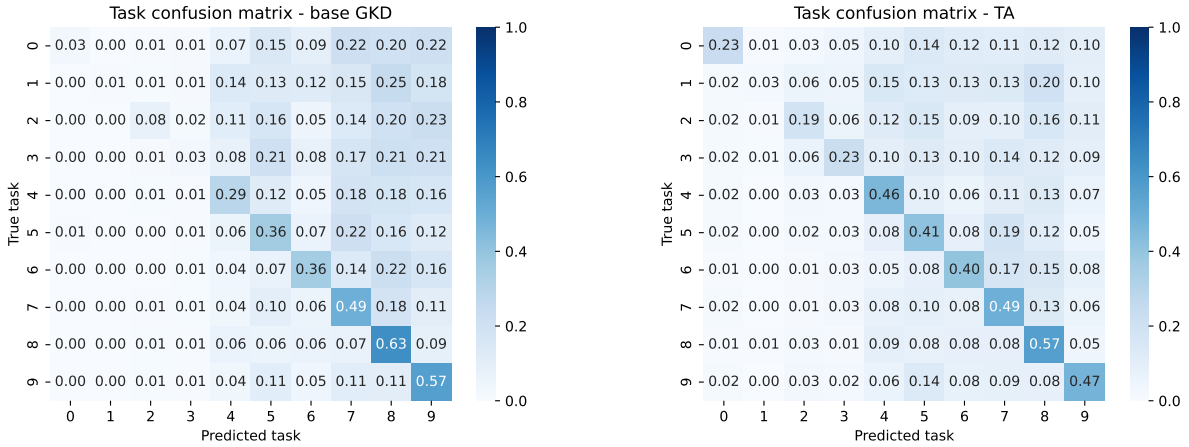

Figure 5. Task confusion matrix after learning all ten tasks on CIFAR100/10 for (left) base GKD and (right) GKD+TA. We see that TA leads to a model that is better at distinguishing between the first tasks and exhibits lower recency bias.

D. Additional experiments for different CL benchmarks

We provide the results of additional experiments conducted with our method and warmup. To shorten the notation, we denote the total number of tasks, including the first pretraining task (if present), with **T**, and the number of classes in the first task by **S**. For example, for CIFAR100 **T10S10** is a setting composed of 10 tasks with 10 classes each, while **T11S50** is a setting where the first task contains 50 classes and the next 10 tasks contain 5 classes each.

D.1. Standard benchmarks

In addition to Section 4.2, we conduct more experiments on CIFAR100, TinyImageNet200 and ImageNet100, adding two settings with a smaller number of tasks. We report final accuracy and forgetting in addition to incremental accuracy and forgetting. Additionally, we include the results for the warmup phase (WU) applied in isolation, without the Teacher Adaptation.

We report the results for CIFAR100, TinyImageNet200 and ImageNet100 in Table 8, Table 9, and Table 10 respectively.

D.2. Fine-grained classification using full datasets

We conduct experiments using the same datasets as in the Section 4.3.1, but without sampling classes, so each of the 6 datasets is treated as a single task. Additionally, we add the results for experiments with a reversed order of tasks to ensure that the ordering does not affect our results. We report final and incremental accuracy and forgetting in Table 11. We observe that Teacher Adaptation (TA) improves the final accuracy, but applying warmup (WU) in this setting results in significant drops in performance regardless of the order of tasks.

D.3. Fine-grained classification with reverse order of datasets

We show results for the experiments conducted in Section 4.3.1, extended with additional metrics. Additionally, we add the results for the results with reversed order of tasks, as in Appendix D.2. Similar to the results in the main paper, the impact of our method is more visible in settings with a larger number of tasks.

D.4. DomainNet

We present the full experimental results for DomainNet in Table 3. In addition to incremental accuracy and forgetting, we also add final accuracy and forgetting to the metrics in the experiments. Additionally, we analyze the random initialization and impact of starting from pretrained checkpoint on ImageNet. For task-wise knowledge distillation (TKD) we apply λ equal to 1, for global knowledge distillation we notice that $\lambda = 10$ works better. We set the learning rate to 0.01 and use the scheduler with LR decay after 60, 120 and 180 epochs.

	T5S20				T6S50			
	$Acc_{Final}\uparrow$	$Acc_{Inc}\uparrow$	$Forg_{Final}\downarrow$	$Forg_{Inc}\downarrow$	$Acc_{Final}\uparrow$	$Acc_{Inc}\uparrow$	$Forg_{Final}\downarrow$	$Forg_{Inc}\downarrow$
GKD	37.63±0.52	48.80±0.36	23.13±2.28	19.30±2.37	40.74±0.72	51.93±1.24	25.87±0.75	18.90±0.29
+*TA*	40.84±0.23	50.08±0.26	**16.76±1.68**	**16.66±1.25**	43.18±1.66	52.22±1.28	**18.73±1.61**	**14.09±0.66**
+*WU*	38.82±0.06	49.69±0.26	22.08±0.61	19.34±1.24	43.19±0.95	**53.42±0.50**	24.54±2.17	18.94±2.69
+*TA+WU*	**41.50±0.60**	**50.70±0.27**	16.85±0.74	17.16±1.24	**44.66±0.54**	53.40±0.73	**18.73±1.73**	14.53±1.56
MKD	34.41±0.16	47.42±0.32	23.74±1.63	23.51±1.75	39.33±0.72	50.28±1.12	23.54±0.62	18.19±0.31
+*TA*	37.03±0.82	48.63±0.17	**14.64±0.22**	**17.63±0.73**	39.45±1.29	49.42±1.38	**13.97±0.89**	**11.02±0.28**
+*WU*	35.37±0.19	47.96±0.15	24.63±0.81	24.35±0.44	40.89±1.52	**51.46±0.68**	24.45±2.59	19.53±2.86
+*TA+WU*	**38.26±0.24**	**49.58±0.06**	15.18±0.41	17.96±0.65	**41.42±0.78**	51.04±1.17	15.00±1.07	12.03±1.32
TKD	38.33±0.70	49.56±0.48	24.78±2.58	25.04±2.55	41.19±0.42	52.07±1.35	17.31±1.20	15.02±0.39
+*TA*	41.12±0.35	50.87±0.15	18.12±1.55	**21.09±1.39**	41.36±0.89	51.88±0.80	**12.84±1.25**	**11.42±0.64**
+*WU*	40.03±0.66	50.50±0.17	22.79±0.73	24.19±1.61	**43.16±0.26**	**53.51±0.39**	18.40±1.73	16.03±2.77
+*TA+WU*	**42.07±0.67**	**51.47±0.33**	**17.77±1.37**	21.21±1.34	43.08±0.35	53.12±0.60	13.64±0.89	12.13±1.37
ANCL	37.13±0.33	48.60±0.27	32.41±0.26	31.71±0.81	43.41±0.62	53.40±0.49	23.54±1.29	18.08±0.40
+*TA*	41.04±0.22	50.36±0.43	24.07±0.81	26.18±0.51	44.64±1.65	53.33±0.66	18.43±3.12	14.44±1.29
+*WU*	38.94±0.57	49.26±0.61	29.84±0.88	30.98±1.98	44.74±0.30	**54.25±0.29**	22.42±1.75	17.49±1.56
+*TA+WU*	**42.08±0.53**	**50.59±0.21**	**22.55±1.15**	**26.09±1.11**	**45.64±0.87**	54.05±0.42	**18.12±1.58**	**14.28±1.06**
	T10S10				T11S50			
GKD	28.27±0.44	42.52±0.76	29.59±0.92	22.26±0.31	30.79±1.62	41.69±1.18	26.84±2.12	18.09±0.88
+*TA*	31.92±0.86	44.09±0.97	22.65±1.32	**19.41±0.60**	33.20±0.76	44.05±1.12	**18.90±0.19**	**12.97±0.43**
+*WU*	31.11±0.58	43.95±0.65	25.26±0.93	21.95±0.74	33.17±0.86	44.09±1.26	23.85±1.29	17.25±0.74
+*TA+WU*	**33.76±0.78**	**45.25±1.02**	**20.87±0.13**	19.87±0.34	**34.72±0.77**	**46.27±1.09**	19.13±1.77	13.98±0.98
MKD	25.41±1.07	39.36±0.70	44.04±0.98	42.74±0.52	31.31±1.22	41.04±0.93	22.56±0.94	15.37±0.33
+*TA*	29.97±1.44	43.55±0.96	32.37±2.17	32.61±0.80	30.90±0.69	41.67±1.35	**14.76±0.25**	**10.51±0.36**
+*WU*	27.32±1.28	40.78±0.95	37.89±0.22	39.03±0.53	31.63±1.16	42.45±0.78	22.57±1.57	16.98±1.23
+*TA+WU*	**31.94±0.23**	**44.85±0.80**	**26.83±1.01**	**30.08±0.57**	**32.79±0.55**	**44.19±1.17**	16.44±0.63	12.56±0.53
TKD	30.05±0.81	43.74±0.84	24.53±0.23	23.65±0.79	28.38±1.46	40.44±1.40	15.68±0.84	12.20±0.46
+*TA*	31.80±0.67	45.29±1.02	18.59±0.90	**19.42±0.85**	28.50±0.39	41.68±1.03	**11.58±0.38**	**9.29±0.75**
+*WU*	31.23±0.94	44.59±0.72	23.21±0.35	23.73±0.32	**30.88±1.16**	42.45±1.31	17.94±0.39	14.30±0.99
+*TA+WU*	**33.14±1.03**	**46.21±0.86**	**17.55±0.66**	20.45±0.57	30.60±0.37	**44.22±1.08**	14.65±0.49	11.79±0.57
ANCL	29.77±1.10	43.15±0.49	36.37±1.48	32.78±1.52	29.77±1.10	43.15±0.49	36.37±1.48	32.78±1.52
+*TA*	33.47±0.17	45.67±0.14	**28.04±1.61**	**26.71±1.71**	33.47±0.17	45.67±0.14	**28.04±1.61**	**26.71±1.71**
+*WU*	31.41±0.54	44.28±0.08	34.70±0.25	32.06±1.07	31.41±0.54	44.28±0.08	34.70±0.25	32.06±1.07
+*TA+WU*	**34.17±0.30**	**46.73±0.20**	28.67±0.63	26.86±1.17	**34.17±0.30**	**46.73±0.20**	28.67±0.63	26.86±1.17
	T20S5				T26S50			
GKD	15.59±0.32	31.89±0.45	43.28±0.56	34.68±1.87	10.10±0.71	17.64±0.93	15.29±0.27	9.67±0.26
+*TA*	19.55±0.24	35.99±0.79	30.38±2.08	**23.32±1.79**	11.99±0.66	19.37±1.73	**9.05±0.63**	**8.31±0.68**
+*WU*	19.22±0.80	34.67±0.55	37.05±1.03	31.14±1.41	10.41±1.01	21.87±0.16	17.92±0.48	11.02±0.65
+*TA+WU*	**21.52±0.62**	**37.11±0.64**	**30.34±0.78**	24.87±1.04	**18.11±0.52**	**26.15±0.94**	10.11±0.97	8.73±0.85
MKD	17.61±0.51	32.89±0.42	33.36±0.80	32.01±1.36	11.55±0.86	19.14±1.36	14.04±0.71	8.76±0.35
+*TA*	19.07±0.65	35.36±0.85	**21.27±1.08**	**20.96±1.31**	13.91±0.80	20.99±1.53	**9.52±0.26**	**8.20±0.98**
+*WU*	18.96±0.82	34.30±0.94	30.45±0.70	29.75±1.56	11.48±0.36	22.42±0.76	17.42±1.70	11.51±1.19
+*TA+WU*	**20.52±0.85**	**36.79±0.70**	21.88±0.81	21.84±0.33	**18.15±0.70**	**26.10±0.75**	11.64±1.04	9.17±0.20
TKD	19.39±0.41	34.58±0.34	22.06±0.46	21.13±1.17	7.88±0.08	14.64±0.33	7.96±0.47	**6.02±0.54**
+*TA*	18.30±0.50	34.62±0.92	**15.22±1.25**	**14.72±1.28**	9.05±0.64	16.66±1.66	**7.17±0.53**	6.88±0.36
+*WU*	**20.77±0.59**	35.72±0.09	21.14±0.88	20.93±1.62	9.55±0.53	18.01±0.59	12.63±1.04	9.90±0.99
+*TA+WU*	20.24±0.97	**36.26±0.71**	17.98±1.00	17.01±0.89	**13.34±0.73**	**22.00±0.97**	10.80±1.03	9.36±0.68
ANCL	19.21±0.75	34.32±0.41	39.97±1.81	36.74±1.38	11.04±0.40	21.84±1.33	18.31±1.64	11.79±0.51
+*TA*	20.63±1.04	37.52±0.78	**31.07±1.00**	**26.77±0.49**	18.96±0.91	26.47±0.41	**12.69±1.49**	**10.33±0.57**
+*WU*	21.53±0.38	36.06±0.73	37.34±0.53	35.34±1.52	11.81±0.24	23.81±0.88	20.05±0.95	13.18±0.34
+*TA+WU*	**22.54±0.33**	**38.48±0.94**	31.87±0.92	27.99±1.20	**20.79±0.28**	**29.67±1.13**	13.84±0.45	10.98±0.56

Table 8. Additional results for CIFAR100

	T5S40				T6S100			
	$Acc_{Final}\uparrow$	$Acc_{Inc}\uparrow$	$Forg_{Final}\downarrow$	$Forg_{Inc}\downarrow$	$Acc_{Final}\uparrow$	$Acc_{Inc}\uparrow$	$Forg_{Final}\downarrow$	$Forg_{Inc}\downarrow$
GKD	27.51±0.81	37.53±1.28	18.22±0.55	14.58±0.56	32.92±1.40	38.74±2.77	13.77±2.12	9.60±2.04
+*TA*	28.36±1.13	37.60±1.49	15.55±1.79	12.63±0.39	33.79±0.89	38.56±1.73	10.94±0.71	**8.28±0.73**
+*WU*	28.31±0.66	37.74±1.00	17.32±0.54	15.04±0.93	33.80±0.79	39.33±1.84	13.04±2.02	9.44±1.29
+*TA*+*WU*	**29.36±0.81**	**38.21±1.10**	**14.05±1.38**	**12.50±0.60**	**34.68±1.06**	**39.45±1.73**	**10.40±0.73**	8.36±0.78
MKD	24.09±0.94	34.93±1.39	11.53±0.17	10.22±1.34	29.65±0.77	35.48±2.06	9.27±1.86	6.66±1.30
+*TA*	24.17±0.77	34.67±1.58	6.57±0.26	**6.51±0.67**	28.28±0.79	34.07±1.63	**4.31±0.47**	**4.14±0.60**
+*WU*	**24.55±0.73**	**35.28±1.08**	11.10±0.07	10.79±1.45	**30.40±1.05**	**36.43±2.08**	9.79±2.19	7.10±1.53
+*TA*+*WU*	24.33±0.46	35.05±1.37	**6.28±0.70**	6.92±1.04	28.53±1.58	34.81±1.95	4.49±0.34	4.42±0.63
TKD	28.18±0.92	37.43±0.86	19.76±1.11	20.37±0.78	34.17±0.72	40.11±1.59	12.33±1.05	11.59±1.22
+*TA*	28.95±0.93	38.07±1.03	16.36±1.91	17.23±0.56	34.42±0.61	40.20±1.31	10.13±1.03	10.15±1.03
+*WU*	29.07±0.54	37.73±0.71	17.35±1.16	19.68±0.80	34.22±0.79	40.44±1.63	12.09±1.68	11.48±1.58
+*TA*+*WU*	**29.57±0.46**	**38.29±0.74**	**14.69±1.34**	**16.93±0.86**	**34.63±1.07**	**40.64±1.38**	**9.55±0.60**	**10.01±1.07**
LwF(A)	28.33±0.36	37.63±0.83	21.80±1.09	20.57±0.85	34.48±0.78	40.85±1.49	19.51±0.94	15.69±1.43
+*TA*	29.37±0.69	38.31±1.06	19.34±1.20	**18.17±0.74**	36.01±1.02	41.22±1.46	**15.83±0.35**	**13.75±1.12**
+*WU*	28.59±0.63	37.69±0.73	21.70±0.96	20.87±1.02	34.64±0.38	41.11±1.38	20.23±1.55	15.70±1.97
+*TA*+*WU*	**29.98±0.86**	**38.45±0.84**	**18.66±1.61**	18.35±0.98	**36.20±0.73**	**41.54±1.21**	16.64±1.02	13.85±1.43
	T10S20				T11S100			
GKD	20.76±0.96	32.12±0.57	26.87±2.16	21.21±1.31	27.17±0.73	34.32±1.96	17.53±3.99	11.41±2.80
+*TA*	22.94±0.70	33.12±0.53	21.39±2.14	**17.50±1.76**	30.65±0.55	37.01±1.84	**12.85±1.60**	**10.05±0.87**
+*WU*	21.94±0.71	33.00±0.59	24.84±1.49	20.37±0.65	28.77±0.81	36.54±1.45	18.11±1.90	11.66±1.83
+*TA*+*WU*	**23.72±1.34**	**33.90±0.78**	**21.29±2.19**	17.70±1.48	**31.34±0.49**	**37.50±1.84**	13.25±2.11	10.14±1.10
MKD	19.37±0.85	31.04±1.00	17.04±0.66	16.90±0.75	26.11±0.72	33.75±2.11	16.22±2.18	10.30±1.63
+*TA*	20.74±0.88	31.78±0.83	**9.89±1.58**	**11.42±1.23**	25.59±1.25	32.23±2.28	**6.18±0.99**	**5.91±0.66**
+*WU*	19.93±0.67	31.32±0.95	16.81±1.12	16.97±0.27	26.36±0.80	**33.91±1.39**	17.12±2.33	11.17±1.71
+*TA*+*WU*	**21.42±0.75**	**32.22±0.95**	10.13±1.25	11.94±1.07	**26.79±1.56**	32.99±1.98	6.78±0.92	6.42±0.80
TKD	22.88±0.29	33.15±0.45	19.17±0.02	21.05±0.39	28.32±1.18	37.31±1.36	14.34±0.78	11.83±1.31
+*TA*	23.92±0.49	33.94±0.70	**14.91±0.98**	17.32±0.97	30.18±0.92	37.97±1.47	**7.93±0.77**	**9.02±0.65**
+*WU*	23.32±0.61	33.71±0.97	18.77±1.33	20.33±0.41	27.90±2.05	37.22±1.60	14.43±1.16	11.48±0.86
+*TA*+*WU*	**24.69±0.89**	**34.58±0.96**	15.09±0.92	**17.27±0.54**	**30.80±0.75**	**38.41±1.52**	8.55±1.06	9.16±0.62
ANCL	22.58±0.83	32.84±0.78	28.76±2.52	27.24±1.47	27.78±0.96	37.74±0.60	25.24±2.77	17.28±2.59
+*TA*	23.98±0.68	33.85±0.55	**25.23±1.31**	24.02±1.39	31.34±0.34	39.50±1.18	20.36±1.54	**14.88±1.41**
+*WU*	23.20±0.57	33.33±0.81	28.66±1.18	26.86±0.20	29.05±0.75	38.86±1.29	25.35±0.64	17.11±0.89
+*TA*+*WU*	**24.78±0.86**	**34.59±0.75**	25.24±1.47	**23.64±0.76**	**32.36±0.60**	**40.10±1.39**	**20.31±1.73**	15.12±1.01
	T20S10				T26S100			
GKD	13.83±0.25	25.75±0.65	33.70±0.54	27.55±1.81	14.33±1.42	23.30±1.59	23.27±2.88	12.63±2.21
+*TA*	16.15±0.83	27.16±0.87	27.52±0.51	22.62±2.03	19.70±0.79	26.88±1.62	**14.57±1.00**	**11.08±0.47**
+*WU*	15.20±0.46	26.98±1.05	30.78±1.84	25.57±1.97	16.10±1.03	25.59±1.76	21.45±1.77	12.66±1.65
+*TA*+*WU*	**17.65±0.16**	**27.85±0.90**	**25.60±0.40**	**21.65±1.43**	**21.86±1.34**	**28.82±1.95**	15.56±0.77	11.44±0.61
MKD	13.81±0.53	25.22±0.88	24.61±1.91	25.80±2.46	15.99±1.72	23.42±1.95	18.15±3.02	10.41±1.97
+*TA*	16.66±0.59	27.05±1.11	14.46±0.75	17.55±2.36	17.65±0.32	23.29±1.61	**8.57±1.70**	**7.61±1.18**
+*WU*	13.97±0.62	25.76±0.73	24.33±2.37	24.76±2.29	16.74±1.44	24.42±1.72	19.95±2.21	12.29±2.42
+*TA*+*WU*	**17.31±0.48**	**27.39±1.33**	**13.66±0.39**	**16.74±1.52**	**20.29±1.27**	**25.75±1.74**	9.48±1.04	8.35±0.82
TKD	16.13±0.45	27.29±0.76	21.34±2.13	23.20±2.46	14.77±1.50	23.94±2.14	14.54±2.43	10.20±1.09
+*TA*	18.34±0.25	28.51±0.96	14.67±0.26	17.38±2.44	18.20±0.34	26.56±1.62	**8.11±1.01**	**8.05±0.36**
+*WU*	16.16±0.29	28.02±0.63	21.28±1.94	22.07±1.48	15.71±1.15	24.74±2.16	16.75±1.81	12.14±2.04
+*TA*+*WU*	**18.99±0.50**	**28.71±1.06**	**14.58±0.94**	**17.37±1.78**	**20.07±0.87**	**28.10±1.97**	9.74±0.90	9.56±0.65
ANCL	15.75±0.24	26.98±0.60	35.24±1.66	32.45±1.18	15.42±1.49	27.95±1.98	29.72±3.14	20.91±0.92
+*TA*	17.91±0.33	28.42±0.48	30.13±1.18	27.26±0.98	21.04±1.79	31.60±1.89	23.59±1.71	18.10±0.55
+*WU*	16.40±0.39	27.54±0.44	34.51±1.06	31.76±0.91	16.47±1.58	28.80±1.59	29.62±2.49	20.74±1.10
+*TA*+*WU*	**19.02±0.34**	**29.18±0.71**	**28.78±0.40**	**26.50±1.34**	**22.32±1.25**	**32.60±1.45**	**23.21±1.01**	**17.94±0.39**

Table 9. Additional results for TinyImageNet200.

	T5S20				T6S50			
	$Acc_{Final}\uparrow$	$Acc_{Inc}\uparrow$	$Forg_{Final}\downarrow$	$Forg_{Inc}\downarrow$	$Acc_{Final}\uparrow$	$Acc_{Inc}\uparrow$	$Forg_{Final}\downarrow$	$Forg_{Inc}\downarrow$
GKD	51.06±0.59	61.95±0.50	27.57±0.79	23.16±0.88	57.30±1.40	66.36±0.75	19.19±0.53	12.92±0.59
+*TA*	**52.29±0.28**	**62.89±0.32**	**23.40±0.31**	**18.94±0.43**	57.86±0.68	66.20±0.40	**16.81±1.00**	**12.70±0.36**
+*WU*	51.04±0.53	60.24±0.23	26.26±0.63	27.22±0.29	**59.52±0.27**	**67.63±0.39**	18.26±0.76	14.45±0.47
+*TA+WU*	51.27±0.18	61.01±0.12	23.98±0.36	23.51±0.29	58.40±0.86	66.86±0.31	18.87±0.75	15.29±0.69
MKD	50.17±0.25	61.48±0.35	23.21±0.78	21.01±0.35	54.70±1.02	65.82±0.24	27.21±0.70	18.25±0.32
+*TA*	50.92±0.42	62.18±0.45	**16.00±0.61**	**14.71±0.35**	57.26±0.83	66.03±0.32	**19.00±1.04**	**14.34±0.11**
+*WU*	50.45±0.60	61.45±0.47	23.20±0.50	21.42±1.21	56.78±0.61	66.36±0.31	25.14±0.24	19.54±0.67
+*TA+WU*	**51.59±0.24**	**62.37±0.59**	16.66±0.72	15.56±1.62	**57.82±0.37**	**66.60±0.14**	19.86±0.60	16.32±0.62
TKD	**53.73±0.25**	62.91±0.44	20.77±0.58	20.30±0.80	58.90±1.02	67.49±0.42	17.38±0.80	14.15±0.13
+*TA*	53.29±0.04	**63.04±0.30**	**18.28±0.72**	**17.26±0.89**	58.80±0.48	67.00±0.23	**15.07±0.66**	**13.48±0.23**
+*WU*	52.14±0.16	60.83±0.24	22.90±0.70	25.45±0.55	**59.85±0.95**	**67.77±0.46**	18.40±0.63	16.35±0.64
+*TA+WU*	51.69±0.19	61.09±0.28	21.68±0.15	23.15±0.59	58.45±0.41	66.94±0.26	18.31±0.51	16.40±0.60
ANCL	54.50±0.28	63.96±0.32	20.26±0.36	17.56±0.40	58.24±1.28	66.98±0.70	16.34±0.36	11.40±0.48
+*TA*	54.39±0.64	63.92±0.42	18.04±0.33	15.29±0.37	59.16±0.63	66.92±0.56	15.03±0.91	11.91±0.10
+*WU*	**54.84±0.20**	**64.25±0.37**	18.71±0.18	16.55±0.51	**60.14±0.66**	**67.78±0.41**	**14.04±0.27**	**10.56±0.42**
+*TA+WU*	54.39±0.60	64.08±0.40	**17.81±0.53**	**15.00±0.59**	59.72±0.56	67.00±0.20	14.62±0.32	12.12±0.51
	T10S10				T11S50			
GKD	40.33±0.37	54.62±0.52	32.72±0.09	25.95±0.11	43.34±0.85	57.94±0.90	26.04±2.37	**14.47±0.83**
+*TA*	**43.17±1.06**	**55.82±0.61**	25.10±1.03	**20.52±0.24**	48.80±0.21	57.18±0.45	**17.56±0.53**	17.24±0.39
+*WU*	41.64±0.32	53.12±0.61	28.09±0.62	27.10±0.55	47.81±1.13	**60.93±0.31**	23.24±1.20	14.88±0.68
+*TA+WU*	42.37±1.01	54.23±0.99	**24.94±0.76**	23.31±0.65	**50.79±0.11**	59.56±0.21	19.53±0.13	18.56±0.28
MKD	40.35±0.43	54.01±0.01	30.94±1.01	28.19±0.61	43.01±0.94	56.18±0.90	26.99±0.88	14.94±0.17
+*TA*	42.35±0.50	**56.02±0.20**	**19.37±1.18**	**18.60±0.76**	43.80±0.49	52.05±0.24	**13.67±1.09**	**14.23±0.46**
+*WU*	40.49±0.52	53.79±0.54	29.89±1.41	27.98±0.27	43.61±1.63	**57.32±1.60**	27.71±1.83	16.09±3.04
+*TA+WU*	**42.94±0.88**	55.39±0.25	20.62±0.53	21.11±0.38	**45.33±3.15**	54.05±3.32	15.21±2.99	15.60±2.54
TKD	43.19±0.16	55.70±0.49	24.84±0.35	23.55±0.35	40.56±1.30	54.72±0.86	15.39±1.01	**10.16±0.34**
+*TA*	**43.93±0.72**	**56.23±0.70**	**17.89±0.14**	**18.09±0.26**	42.83±0.61	53.85±0.39	**10.13±0.20**	13.15±0.16
+*WU*	42.71±1.29	53.51±0.56	24.84±0.96	26.84±0.28	45.44±0.98	**58.94±0.45**	23.85±1.16	18.09±0.76
+*TA+WU*	43.05±0.55	54.40±0.64	21.32±0.73	22.50±0.12	**49.48±0.92**	58.86±0.38	17.65±1.12	19.07±0.75
ANCL	44.01±0.19	55.81±0.41	29.56±0.51	27.13±0.50	46.94±0.67	60.97±0.62	26.30±0.84	15.72±0.14
+*TA*	45.75±0.90	**57.41±0.22**	23.71±1.08	**21.54±0.50**	50.35±0.45	59.22±0.26	18.07±1.14	17.34±0.53
+*WU*	45.08±0.08	56.13±0.40	27.27±0.34	26.13±0.20	50.18±0.87	**63.15±0.12**	23.10±0.73	**13.29±0.35**
+*TA+WU*	**46.37±0.36**	57.20±0.32	**22.79±0.74**	21.70±0.29	**52.70±0.25**	61.18±0.34	**17.93±0.82**	16.92±0.26
	T20S5				T26S50			
GKD	24.14±0.91	42.82±0.58	45.53±0.88	35.39±0.88	14.82±0.85	21.91±0.06	17.99±0.53	9.29±0.69
+*TA*	31.89±1.63	**45.88±0.79**	**27.74±1.03**	**23.25±0.62**	17.16±0.84	22.31±0.64	**12.71±1.11**	**11.28±0.98**
+*WU*	28.27±1.19	43.92±0.38	37.93±1.38	32.37±0.66	17.85±0.27	30.15±0.56	23.05±1.17	15.86±0.91
+*TA+WU*	**32.63±0.69**	45.75±0.61	29.33±0.37	25.37±0.04	**24.90±0.24**	**34.19±0.28**	20.82±1.10	18.97±1.23
MKD	26.27±1.28	43.39±0.66	40.22±1.04	34.25±0.81	15.83±1.18	26.07±0.29	29.85±0.95	16.00±0.06
+*TA*	31.90±0.34	**46.18±0.54**	**18.92±0.14**	**19.14±0.79**	16.12±0.25	22.25±0.13	**11.65±0.66**	**11.94±1.23**
+*WU*	27.50±1.10	43.73±0.43	38.01±1.61	33.47±0.76	16.15±0.27	**28.42±1.65**	32.35±3.78	18.39±4.16
+*TA+WU*	**33.55±1.15**	46.12±0.64	22.06±1.28	21.56±0.80	**19.85±6.72**	25.79±6.22	13.35±3.06	13.72±3.01
TKD	29.39±0.47	44.75±0.28	35.96±0.95	32.16±0.14	10.99±0.22	19.32±0.23	13.55±0.88	**9.67±0.61**
+*TA*	32.71±0.27	**46.45±0.42**	**19.12±1.02**	**19.55±0.30**	13.90±0.52	22.55±0.83	**7.24±0.71**	9.96±0.28
+*WU*	30.79±0.60	45.29±0.38	33.50±0.33	31.09±0.94	15.48±1.06	27.81±0.55	29.99±0.49	23.07±0.97
+*TA+WU*	**33.73±0.68**	**46.45±0.27**	23.74±0.63	22.98±0.53	**22.59±0.66**	**32.46±0.52**	21.63±1.15	20.65±0.99
ANCL	28.83±0.71	44.94±0.63	41.59±1.13	34.24±1.14	17.92±0.24	29.19±0.76	28.50±0.64	16.12±0.27
+*TA*	33.92±0.68	47.49±0.35	**25.89±0.82**	**22.61±0.40**	23.52±0.73	29.64±0.50	**15.06±0.78**	**14.67±0.84**
+*WU*	31.66±0.44	46.28±0.56	37.07±0.77	32.07±1.20	19.79±0.47	32.25±0.15	28.02±0.46	15.11±0.46
+*TA+WU*	**35.79±0.39**	**47.81±0.30**	27.14±0.10	24.93±0.89	**27.49±1.10**	**33.31±0.58**	15.23±0.58	14.68±0.35

Table 10. Additional results for ImageNet100

	Basic order				Reverse order			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
GKD	26.71±0.53	**46.74±0.94**	46.57±0.89	**37.47±1.07**	26.61±1.07	54.88±0.62	52.71±1.01	39.34±0.41
+TA	**31.16±0.75**	46.38±0.74	44.09±0.68	37.84±0.98	**33.61±0.40**	**54.97±0.35**	41.89±0.64	**36.63±0.58**
+WU	23.96±1.53	36.37±0.95	**34.78±1.80**	46.72±1.22	23.60±1.33	43.02±0.56	39.25±1.36	51.69±0.76
+TA+WU	24.74±0.95	37.74±1.03	40.44±0.87	44.85±0.74	27.06±0.44	44.68±0.39	**37.26±0.77**	47.10±0.61

Table 11. Additional results for training on full fine-grained datasets in standard and reversed order.

	6 tasks, 20 classes each, base order				6 tasks, 20 classes each, reverse order			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
GKD	38.12±2.00	56.03±3.75	45.18±3.15	**30.23±0.72**	37.10±3.40	**67.30±1.99**	55.82±4.80	34.60±3.77
+TA	**43.11±0.64**	**57.24±1.79**	39.58±0.80	33.12±1.76	**46.97±3.94**	67.19±4.81	41.20±4.55	**33.94±7.78**
+WU	31.56±6.58	44.06±5.43	42.48±1.08	49.24±2.65	36.70±3.30	54.88±1.53	42.92±4.58	51.30±3.39
+TA+WU	42.85±2.34	52.96±1.20	**35.97±2.75**	38.77±2.45	44.05±1.94	58.92±2.47	**38.27±2.75**	45.16±4.57
MKD	41.32±0.81	**60.12±1.59**	40.30±3.98	**26.42±1.60**	43.96±3.89	**70.65±1.67**	42.83±5.11	**25.52±3.43**
+TA	39.93±1.95	55.77±2.23	36.22±1.07	31.43±2.28	**46.94±4.42**	67.15±4.48	**37.01±6.67**	31.07±8.38
+WU	34.65±2.72	48.01±1.97	44.86±0.61	47.07±2.40	35.49±5.51	56.83±2.79	46.14±6.67	48.15±5.95
+TA+WU	**44.91±2.51**	56.41±1.04	**35.53±2.19**	34.47±0.71	44.62±2.85	61.25±3.28	38.31±3.91	41.48±6.24
TKD	38.70±3.61	56.69±3.36	44.99±1.04	33.86±1.36	40.23±4.35	66.84±2.79	50.63±5.56	35.77±4.85
+TA	42.34±0.77	**57.99±1.88**	39.97±0.52	**33.79±1.80**	**46.86±5.55**	**67.18±5.28**	39.90±6.90	**33.67±8.84**
+WU	32.71±5.27	45.64±5.00	44.43±2.13	49.06±3.58	37.53±2.65	56.12±1.34	44.06±4.80	50.70±3.59
+TA+WU	**43.04±1.30**	53.68±1.56	**37.14±1.52**	39.73±2.92	44.52±2.14	59.37±2.39	**38.32±2.53**	45.65±4.61
	12 tasks, 10 classes each, base order				12 tasks, 10 classes each, reverse order			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
GKD	23.66±2.75	44.03±1.53	58.93±1.75	42.70±2.32	26.84±4.11	51.68±2.03	59.80±5.93	47.19±2.43
+TA	**36.46±0.94**	**50.80±1.27**	43.53±1.96	37.58±3.01	41.15±5.50	**59.51±4.40**	42.61±6.36	**36.14±6.20**
+WU	11.02±1.11	29.93±2.68	50.32±3.34	54.06±0.60	19.55±0.93	39.99±0.23	46.40±0.80	54.24±0.27
+TA+WU	36.00±0.89	49.46±2.09	**35.49±1.90**	**35.81±2.02**	**44.79±4.28**	52.88±5.06	**26.81±5.14**	40.33±3.03
MKD	25.35±4.36	45.19±2.91	55.04±2.71	43.81±1.56	31.81±5.29	53.63±2.31	52.38±6.99	43.86±3.05
+TA	**35.10±1.92**	**51.16±1.84**	37.96±1.84	**34.89±2.60**	41.95±5.36	**60.11±3.87**	36.16±6.75	**33.18±5.64**
+WU	15.61±6.67	32.93±3.65	49.31±7.17	54.64±1.52	26.19±8.31	42.21±1.31	44.49±6.28	54.20±1.67
+TA+WU	34.65±4.00	50.40±3.57	**36.79±1.04**	35.74±3.45	**45.91±4.64**	54.05±4.55	**27.74±4.92**	40.08±3.32
TKD	24.69±3.39	46.28±1.42	56.57±2.87	43.38±0.82	31.50±3.26	54.13±2.24	52.51±5.19	43.47±2.21
+TA	35.06±1.42	**51.66±1.63**	39.72±2.08	35.43±2.75	41.40±4.78	**59.22±3.67**	37.10±5.97	**34.27±5.88**
+WU	15.04±1.20	33.32±1.74	50.10±1.89	53.59±1.52	26.10±4.04	42.34±1.31	43.50±3.29	52.66±0.59
+TA+WU	**35.81±1.27**	50.12±2.31	**34.73±0.52**	**34.64±1.99**	**43.91±5.45**	52.38±5.54	**26.77±4.76**	40.35±3.60
	24 tasks, 5 classes each, base order				24 tasks, 5 classes each, reverse order			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
GKD	9.41±2.78	30.14±4.00	60.16±4.76	51.95±2.28	15.33±2.59	33.65±0.76	64.43±3.19	57.18±2.13
+TA	26.92±2.31	42.86±1.82	44.80±3.25	39.01±1.55	29.52±2.33	**48.34±1.51**	48.32±2.43	41.58±3.30
+WU	7.24±1.03	24.91±1.44	43.78±2.98	51.18±1.15	6.15±1.91	25.45±2.08	39.88±3.29	46.36±1.88
+TA+WU	**27.30±1.97**	**45.28±2.25**	**34.75±4.58**	**33.67±2.48**	**34.14±9.24**	42.65±1.21	**19.92±11.23**	**32.38±8.24**
MKD	11.57±2.70	32.61±0.55	63.70±2.58	55.91±1.77	15.53±2.96	34.85±1.40	62.28±3.64	56.57±2.11
+TA	**27.88±1.58**	45.49±1.82	37.65±1.16	34.48±0.89	**31.82±1.51**	**49.47±1.16**	38.89±1.18	35.98±2.07
+WU	7.52±4.24	24.62±1.56	44.85±6.75	52.59±2.59	4.57±3.36	21.36±4.30	40.82±4.75	51.73±0.78
+TA+WU	26.71±2.29	**45.80±2.48**	**33.52±4.66**	**32.48±2.15**	30.63±6.78	41.09±4.04	**20.99±4.34**	**31.64±3.87**
TKD	9.99±0.47	32.17±2.71	60.29±3.39	51.72±1.95	16.28±2.59	34.68±1.62	59.14±4.12	54.83±1.69
+TA	**26.24±3.62**	**43.95±2.60**	36.28±2.17	33.28±1.86	**30.81±1.61**	**48.95±0.95**	38.10±1.53	35.45±1.59
+WU	3.79±1.21	22.99±1.33	36.45±3.12	42.30±4.03	8.62±4.87	24.91±6.55	39.22±7.44	49.71±2.49
+TA+WU	20.24±4.08	37.22±6.11	**23.39±2.94**	**24.53±3.01**	25.21±3.57	47.46±5.21	**28.40±4.04**	**26.15±2.93**

Table 12. Additional results for the fine-grained classification datasets, extended by the experiments with reversed order of the datasets.

	6 tasks 50 classes with $\lambda = 10$							
	Trained from scratch				Pre-trained on Imagenet			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
GKD	3.43±0.45	18.63±0.27	**31.05±0.42**	**23.27±0.26**	27.69±0.09	43.27±0.10	**41.61±0.71**	**36.83±0.88**
+*TA*	**6.73±0.82**	**19.55±0.42**	36.26±0.55	27.22±0.21	**30.12±0.23**	**43.52±0.17**	43.38±0.39	38.33±0.41
+*TA*+*WU*	5.83±0.70	18.68±0.36	33.00±0.59	31.12±0.38	28.83±0.26	39.00±0.12	42.64±0.56	49.44±0.37
ANCL	7.33±0.91	19.58±0.46	**33.28±0.18**	**25.63±0.20**	25.58±0.62	**42.90±0.84**	45.01±1.19	37.28±1.86
+*TA*	8.85±0.25	20.34±0.40	39.70±0.38	30.73±0.44	26.95±0.07	42.67±0.51	41.62±0.52	38.56±1.24
+*TA*+*WU*	**9.65±0.40**	**20.39±0.09**	37.57±0.12	32.83±0.66	**29.15±0.98**	42.44±0.80	**34.23±0.85**	**37.09±1.33**
	12 tasks 25 classes with $\lambda = 10$							
	Trained from scratch				Pre-trained on Imagenet			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
GKD	2.48±0.41	14.45±0.25	29.01±0.63	**29.04±0.37**	17.69±0.84	35.98±0.96	50.12±1.00	43.00±1.66
+*TA*	4.88±0.46	**16.25±0.46**	36.47±1.06	33.03±0.19	**28.02±0.68**	**38.89±0.52**	40.87±0.64	**41.10±0.42**
+*TA*+*WU*	**6.37±1.39**	16.22±0.45	**27.45±0.70**	32.63±0.11	27.97±0.38	34.24±0.42	**30.86±0.33**	44.17±0.37
ANCL	3.97±0.50	14.82±0.41	**32.80±0.61**	**33.46±0.40**	14.13±1.35	33.34±0.55	55.72±0.16	49.05±0.65
+*TA*	6.54±0.51	17.19±0.06	39.77±0.76	37.40±0.38	23.42±1.03	35.81±0.18	40.86±0.85	43.84±0.23
+*TA*+*WU*	**8.06±0.47**	**18.12±0.07**	36.07±0.13	36.71±0.05	**27.83±1.36**	**36.73±1.78**	**31.00±0.39**	**39.29±0.49**
	6 tasks 50 classes with $\lambda = 1$							
	Trained from scratch				Pre-trained on Imagenet			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
GKD	5.52±2.11	18.80±0.75	**39.48±1.12**	**32.74±0.60**	19.93±0.47	36.56±0.12	58.83±0.96	56.75±0.89
+*TA*	**11.42±0.30**	**20.84±0.24**	42.92±0.49	34.98±0.19	**25.50±0.86**	**39.01±0.60**	**56.75±0.90**	**54.99±0.93**
TKD	4.87±0.97	18.94±0.36	**32.95±0.78**	**26.60±0.29**	27.44±0.15	42.18±0.41	42.47±0.65	**41.23±1.05**
+*TA*	**6.76±1.03**	**19.67±0.19**	37.43±0.57	29.91±0.09	28.96±0.32	**42.51±0.43**	43.91±0.16	42.00±0.59
+*TA*+*WU*	5.88±0.82	18.65±0.27	33.50±0.96	32.33±0.49	**29.55±0.30**	40.62±0.07	**42.08±0.55**	46.86±0.39
MKD	3.24±1.10	18.74±0.52	**26.83±0.83**	**19.10±0.40**	**30.41±0.47**	**45.70±0.30**	32.96±0.29	**27.48±0.60**
+*TA*	3.99±0.38	18.04±0.16	29.16±0.29	22.70±0.25	27.47±0.22	42.91±0.04	**31.05±0.31**	29.43±0.04
+*TA*+*WU*	**6.26±0.02**	**19.18±0.13**	31.36±1.04	27.81±0.49	30.10±1.07	42.20±0.65	34.50±0.41	38.01±0.40
	12 tasks 25 classes with $\lambda = 1$							
	Trained from scratch				Pre-trained on Imagenet			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
GKD	2.88±0.25	14.38±0.82	**38.89±3.42**	**39.78±2.23**	9.62±0.56	27.99±0.40	64.74±0.21	59.93±1.13
+*TA*	**6.80±0.28**	**17.02±0.19**	47.37±0.82	43.44±0.16	**19.00±0.83**	**33.90±0.62**	**58.20±0.71**	**54.97±0.71**
TKD	3.39±0.43	12.28±0.99	41.22±0.28	33.32±0.33	22.66±1.59	37.56±1.03	41.65±1.73	41.21±1.97
+*TA*	5.10±0.47	**16.84±0.42**	34.03±0.61	33.46±0.41	**29.39±0.37**	**38.85±0.31**	34.51±0.54	**39.91±0.39**
+*TA*+*WU*	**6.27±0.52**	16.40±0.56	**26.51±0.57**	**32.87±0.34**	29.26±2.05	36.81±0.97	**30.00±1.48**	41.98±0.95
MKD	2.32±0.39	13.45±0.53	23.92±1.02	**27.23±0.98**	24.55±0.20	**39.14±0.21**	38.09±1.02	36.53±1.00
+*TA*	3.60±0.51	15.30±0.35	27.32±0.37	28.71±0.26	27.49±0.94	37.84±0.35	27.99±1.16	**34.09±0.39**
+*TA*+*WU*	**5.46±1.20**	**15.89±0.96**	**22.23±1.98**	30.62±0.48	**30.34±1.32**	37.79±1.11	**24.81±1.05**	37.00±1.84

Table 13. Additional results for DomainNet.