Appendix

A. Evaluation metrics. The average incremental accuracy at task k is defined as $A_k = \frac{1}{k}\sum_{j=1}^{k} a_{k,j}$, where $a_{k,j} \in [0, 1]$ be the accuracy of the j-th task ($j \leq k$) after training the network sequentially for k tasks [3]. Overall average incremental accuracy Acc_{Inc} is the mean value from all tasks. We also report *average forgetting* as defined in [8], while the $Forg_{Inc}$ is similarly the mean value from all tasks. We provide results with additional metrics such as final accuracy Acc_{Final} and final forgetting $Forg_{Final}$ in the Appendix.

B. Alternative methods of teacher adaptation. We study alternative methods of adapting the teacher model and try pertaining (P) and continuously training (CT) the teacher model. For pertaining, we train the teacher on new data in isolation for a few epochs, while during continuous training we update the teacher alongside the main model using the same batches of new data. We train either the full teacher model (FM) or only its batch normalization layers (BN). Finally, we repeat all the experiments with fixed batch normalization statistics ($fix\ BN$). We present results in Table 3. Alternative solutions perform within the standard deviation of TA, but the values of the hyperparmeters for those models are small (learning rate 10^{-7}, 5 epochs of pertaining), indicating that the teacher the crucial change in the model is batch normalization statistics.

Method	Acc_{Final}	Acc_{Inc}	$Forg_{Final}$	$Forg_{Inc}$
Base	27.53±0.15	42.22±0.38	31.28±1.64	23.11±1.58
P-FM	31.54±0.67	43.46±0.72	24.18±1.17	20.80±1.51
+*fix BN*	28.02±0.60	42.33±0.53	29.91±1.27	22.66±0.95
P-BN	31.16±0.54	43.64±0.77	24.44±0.96	20.13±0.75
+*fix BN*	27.62±0.48	42.12±0.38	29.95±1.64	22.50±0.95
T-FM	31.37±0.94	43.38±0.77	24.34±1.37	20.93±1.58
+*fix BN*	28.17±0.49	42.29±0.42	29.79±1.02	22.55±0.67
T-BN	31.35±0.63	43.69±0.76	24.29±0.61	20.23±0.59
+*fix BN*	27.33±0.50	42.09±0.45	30.20±1.73	22.50±0.85
TA	**32.15±0.12**	**44.31±0.26**	**23.55±0.51**	**19.85±0.93**

Table 3: Ablation study of different ways to adapt the teacher model. Our method achieves the best results while requiring no additional hyperparameters. All experiments were conducted on CIFAR100 split into 10 tasks.

C. Task-recency bias reduction with Teacher Adaptation. 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 4, we analyze task confusion matrices of standard KD (GKD) 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, TA helps to alleviate the recency bias in CIL.

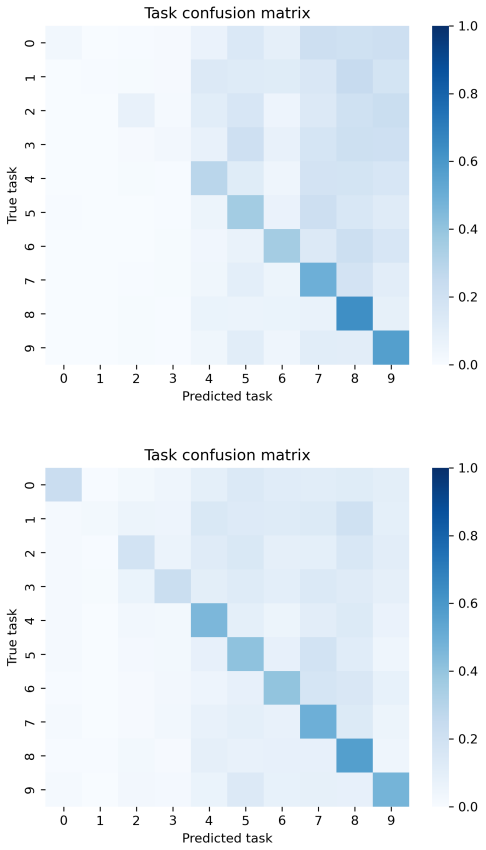

Figure 4: Task confusion matrix after learning all ten tasks on CIFAR100/10 for (upper) base GKD and (lower) GKD+TA. We see that TA leads to a model that is better at distinguishing between tasks and shows lower recency bias.

D. Additional results for standard benchmarks. In addition to results in Section 4, we conduct more experiments on CIFAR100 and ImageNet100, adding two settings with a smaller number of tasks. We report final accuracy and forgetting in addition to incremental accuracy and forgetting. We report the results for CIFAR100 in Table 4, and for ImageNet100 in Table 5.

	Equal split				First task with 50 classes			
	5 tasks				6 tasks			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
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**
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**
	10 tasks				11 tasks			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
GKD	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**
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**
	20 tasks				26 tasks			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
GKD	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**
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

Table 4: Additional results for CIFAR100.

	Equal split				First task with 50 classes			
	5 tasks				6 tasks			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
GKD	51.06±0.59	61.95±0.50	27.57±0.79	23.16±0.88	**55.71±0.82**	**63.40±0.41**	13.26±0.87	7.83±0.40
+*TA*	**52.29±0.28**	**62.89±0.32**	**23.40±0.31**	**18.94±0.43**	55.18±0.84	62.94±0.17	**13.67±0.81**	**10.25±0.38**
TKD	**53.73±0.25**	62.91±0.44	20.77±0.58	20.30±0.80	**57.17±0.45**	**66.17±0.24**	**11.38±0.39**	**8.92±0.20**
+*TA*	53.29±0.04	**63.04±0.30**	**18.28±0.72**	**17.26±0.89**	56.58±0.88	65.53±0.23	11.66±0.87	10.75±0.42
	10 tasks				11 tasks			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
GKD	40.33±0.37	54.62±0.52	32.72±0.09	25.95±0.11	41.56±1.44	**52.67±0.93**	18.94±1.21	9.92±0.83
+*TA*	**43.17±1.06**	**55.82±0.61**	**25.10±1.03**	**20.52±0.24**	**44.60±0.24**	51.44±0.51	**13.80±1.09**	**14.55±0.76**
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
	20 tasks				26 tasks			
	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$	$Acc_{Final} \uparrow$	$Acc_{Inc} \uparrow$	$Forg_{Final} \downarrow$	$Forg_{Inc} \downarrow$
GKD	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
TKD	28.90±0.42	**45.71±0.37**	29.87±0.97	25.85±0.26	10.99±0.22	19.32±0.23	13.55±0.88	9.67±0.61
+*TA*	**30.13±0.34**	45.14±0.78	**15.42±0.89**	**15.62±0.51**	**13.90±0.52**	**22.55±0.83**	**7.24±0.71**	**9.96±0.28**

Table 5: Additional results for ImageNet100.