Method	CIFAR100		ImageNet100		ImageNet-R		CUB200		VTAB	
	Avg ↑	**Last** ↑	**Avg** ↑	**Last** ↑	**Avg** ↑	**Last** ↑	**Avg** ↑	**Last** ↑	**Avg** ↑	**Last** ↑
DualPrompt [5]	82.11	74.31	-	-	82.73	76.41	82.37	76.29	-	-
CODA-P [2]	85.19	76.4	85.93	79.02	82.06	79.5	84.77	80.39	87.5	81.2
Ours	**86.13**	**78.21**	**87.76**	79.16	85.77	79.98	**86.93**	**81.64**	**91.37**	**89.67**
AttriCLIP + Ours	78.06	67.59	87.37	**79.3**	**86.35**	**80.6**	83.71	79.01	74.84	71.12

Table 1: **Performance comparison with CODA-Prompt** (avg. over 3 runs): Similar to L2P and DualPrompt, we re-implement CODA-P using the pre-trained OpenAI CLIP's ViT-B/16 backbone and run it with memory replay (see Table 2 for results without memory replay). Best results are in **bold**.

Method	CIFAR100		ImageNet100		ImageNet-R		CUB200		VTAB	
	Avg ↑	**Last** ↑	**Avg** ↑	**Last** ↑	**Avg** ↑	**Last** ↑	**Avg** ↑	**Last** ↑	**Avg** ↑	**Last** ↑
CODA-P	74.66	63.7	79.8	72.66	76.44	72.19	74.32	70.15	70.59	66.12
AttriCLIP [4]	71.35	61.4	80.55	73.08	79.57	76.1	73.89	72.36	71.25	66.02
CoOp + Ours	74.19	63.45	81.07	72.0	81.22	75.8	**82.59**	**74.98**	**82.11**	**80.11**
AttriCLIP + Ours	**76.94**	**69.39**	**84.1**	**75.83**	**85.2**	**78.57**	77.2	73.58	72.98	68.5
AttriCLIP + Ours (w/o adapter init)	71.87	60.35	75.9	69.41	77.1	70.53	69.82	65.95	65.48	61.35
AttriCLIP + Ours (w/o distribution reg.)	64.2	53.01	64.03	52.55	62.81	53.19	58.52	51.0	58.22	56.14

Table 2: **Performance comparison without memory replay** (avg. over 3 runs). For a thorough analysis, the last two rows ablate our proposed pretrained CLIP's language-aware anti-forgetting components (Sec. 3.3, main paper): distribution regularization and adapter weight initialization. Best results are in **bold**.

Method	ImageNet100	CIFAR100 + ImageNet100
DualPrompt [5]	81.9	67.1
Continual-CLIP [3]	75.4	54.9
AttriCLIP [4]	83.3	78.3
PROOF []	81.26	82.59
Ours	83.51	83.83
CoOp + Ours	82	82.63
MaPLe + Ours	82.97	83.6
AttriCLIP + Ours	**84.14**	**84.56**

Table 3: **Performance comparison with PROOF on the Cross-Datasets Continual Learning (CDCL) setting** [4].

Method	CIFAR100		ImageNet100	
	Avg ↑	**Last** ↑	**Avg** ↑	**Last** ↑
CODA-P	52.13	49.5	52.99	48.03
AttriCLIP [4]	58.61	52.1	60.54	57.4
PROOF	55.29	50.3	56.8	54.37
AttriCLIP + Ours	**61.7**	**55.89**	**62.91**	**60.2**
AttriCLIP + Ours (w/o init.)	61.33	54.95	62.14	59.86
AttriCLIP + Ours (w/o reg.)	58.95	52.6	60.04	57.93

Table 4: **Results for computationally budgeted CL setup [1]:** we follow the "Normal" budget setup from [1] where each incremental task is allocated training iterations equivalent of 1 epoch on the first task of each dataset. Scores reported are averages over three runs. Best results are in **bold**.

References

[1] A. Prabhu, H. A. A. K. Hammoud, S.-N. Lim, B. Ghanem, P. H. Torr, and A. Bibi. From categories to classifier: Name-only continual learning by exploring the web. *arXiv preprint arXiv:2311.11293*, 2023.

[2] J. S. Smith, L. Karlinsky, V. Gutta, P. Cascante-Bonilla, D. Kim, A. Arbelle, R. Panda, R. Feris, and Z. Kira. Coda-prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11909–11919, 2023.

[3] V. G. Thengane, S. A. Khan, M. Hayat, and F. S. Khan. Clip model is an efficient continual learner. *ArXiv*, abs/2210.03114, 2022.

[4] R. Wang, X. Duan, G. Kang, J. Liu, S. Lin, S. Xu, J. Lü, and B. Zhang. Attriclip: A non-incremental learner for incremental knowledge learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3654–3663, 2023.

[5] Z. Wang, Z. Zhang, S. Ebrahimi, R. Sun, H. Zhang, C.-Y. Lee, X. Ren, G. Su, V. Perot, J. Dy, et al. Dualprompt: Complementary prompting for rehearsal-free continual learning. In *European Conference on Computer Vision*, pages 631–648. Springer, 2022.