AttriCLIP: A Non-Incremental Learner for Incremental Knowledge Learning

Supplementary Material

1. Classes of ImageNet100

We choose 100 classes from ImageNet ISLVRC2012 [1] to build ImageNet100 in the paper. The class names of ImageNet100 are:

'American robin', 'Gila monster', 'eastern hog-nosed snake', 'garter snake', 'green mamba', 'European garden spider', 'lorikeet', 'goose', 'rock crab', 'fiddler crab', 'American lobster', 'little blue heron', 'American coot', 'Chihuahua', 'Shih Tzu', 'Papillon', 'toy terrier', 'Treeing Walker Coonhound', 'English foxhound', 'borzoi', 'Saluki', 'American Staffordshire Terrier', 'Chesapeake Bay Retriever', 'Vizsla', 'Kuvasz', 'Komondor', 'Rottweiler', 'Dobermann', 'Boxer', 'Great Dane', 'Standard Poodle', 'Mexican hairless dog (xoloitzcuintli)', 'coyote', 'African wild dog', 'red fox','tabby cat', 'meerkat', 'dung beetle', 'stick insect', 'leafhopper', 'hare', 'wild boar', 'gibbon', 'langur', 'ambulance', 'baluster handrail', 'bassinet', 'boathouse', 'poke bonnet', 'bottle cap', 'car wheel', 'bell or wind chime', 'movie theater', 'cocktail shaker', 'computer keyboard', 'Dutch oven', 'football helmet', 'gas mask or respirator', 'hard disk drive', 'harmonica', 'honeycomb', 'clothes iron', 'jeans', 'lampshade', 'laptop computer', 'milk can', 'mixing bowl', 'modem', 'moped', 'graduation cap', 'mousetrap', 'obelisk', 'park bench', 'pedestal', 'pickup truck', 'pirate ship', 'purse', 'fishing casting reel', 'rocking chair', 'rotisserie', 'safety pin', 'sarong', 'balaclava ski mask', 'slide rule', 'stretcher', 'front curtain', 'throne', 'tile roof', 'tripod', 'hot tub', 'vacuum cleaner', 'window screen', 'airplane wing', 'cabbage', 'cauliflower', 'pineapple', 'carbonara', 'chocolate syrup', 'gyromitra', 'stinkhorn mushroom'.

2. More Detailed Results of Experiments

2.1. More Detailed Results on Crosss-Datasets Continual Learning

We also test L2P [5] and S-liPrompts [3] on the proposed Cross-Datasets Continual Learning (CDCL) benchmark in this supplementary material. DualPrompt [4] is an updated version of L2P, so we compare with DualPrompt directly, rather than L2P, in the paper. S-liPrompts is a CLIP-based method of domain incremental learning. We treat each dataset as a domain, and test S-liPrompts on the CDCL benchmark.

Table 1. Accuracy of different methods on CIFAR100. The models are either trained from scratch on CIFAR100 (CIFAR100), or fine-tuned on CIFAR100 after being continually trained from scratch on ImageNet100 (CIFAR100-I2C).

Method	Memory	CIFAR100	CIFAR100-I2C	FT
iCaRL-1	2000	49.5	49.7	+0.2
iCaRL-2	2000	49.1	46.5	-2.6
CoOp-1	1000	67.6	61.1	-6.5
CoOp-2	1000	67.6	59.0	-8.6
ARI-1	2000	80.9	74.5	-6.4
ARI-2	2000	79.7	59.9	-19.8
Continual-CLIP	0	66.7	66.7	0
S-liPrompts	0	58.9	53.3	-5.6
L2P-1	0	83.8	78.9	-4.9
L2P-2	0	80.7	72.4	-8.3
DualPrompt-1	0	**86.5**	80.7	-5.8
DualPrompt-2	0	84.1	74.7	-9.4
AttriCLIP	0	81.4	**82.3**	+**0.9**

The complete experimental comparisons on CDCL are reported in Tables 1, 2, 3. AttriCLIP obtains the best results in the long-sequence, domain-shift continual learning task. Moreover, it achieves remarkable performance in generalizing to a new dataset and preventing forgetting the previous dataset.

3. More Visualization

In this section, we provide more visualization to verify that different prompts do reflect different image attributes. In Fig. 1, the diversity of the learned prompts is verified on more images. In Fig. 2, we visualize the contents of more prompts (*e.g.*, $\mathbf{P}_1$, $\mathbf{P}_5$ and $\mathbf{P}_7$) on more images. The newly tested $\mathbf{P}_7$ suggests that it focuses more on the "hat" content for these images. The combinations of $\mathbf{P}_1$, $\mathbf{P}_2$, ..., $\mathbf{P}_N$ can represent different contents of images after training.

Table 2. Accuracy of different methods on ImageNet100. The models are either trained from scratch on ImageNet100 (ImageNet100), or fine-tuned on CIFAR100 after being continually trained from scratch on ImageNet100 (ImageNet100-I2C).

Method	Memory	ImageNet100	ImageNet100-I2C	BT
iCaRL-1	2000	59.5	34.5	-15.2
iCaRL-2	2000	58.7	50.9	-7.8
CoOp-1	1000	79.3	57.6	-21.7
CoOp-2	1000	79.3	75.9	-3.4
ARI-1	2000	79.3	51.2	-28.1
ARI-2	2000	77.9	61.8	-16.1
Continual-CLIP	0	75.4	75.4	0
S-liPrompts	0	61.1	40.9	-20.2
L2P-1	0	82.7	59.6	-23.1
L2P-2	0	81.1	74.8	-6.3
DualPrompt-1	0	**85.4**	63.6	-21.8
DualPrompt-2	0	81.9	77.8	-4.1
AttriCLIP	0	83.3	**90.3**	**+7.0**

Table 3. Comparison among different methods on ImageNet100 + CIFAR100 where each model is continually trained on ImageNet100 and CIFAR100 in sequence.

Method	Memory	CIFAR100+ ImageNet100
iCaRL-1	2000	30.7
iCaRL-2	2000	37.6
CoOp-1	1000	46.6
CoOp-2	1000	55.4
ARI-1	2000	32.5
ARI-2	2000	57.3
Continual-CLIP	0	54.9
S-liPrompts	0	39.6
L2P-1	0	34.9
L2P-2	0	64.6
DualPrompt-1	0	35.4
DualPrompt-2	0	67.1
AttriCLIP	0	**78.3**

4. Limitation and Ethic Impact

Although AttriCLIP is a non-incremental learner, the size of the attribute bank is related to the length of the task sequence. For long sequence tasks, a larger attribute bank is required to cover the necessary attributes in the tasks for learning. In the future, we will test AttriCLIP on more datasets and on longer-sequence CDCL benchmarks. This work does not have obvious ethic impacts.

Figure 1. More Visualization of the top three prompts of the same image using Grad-CAM [2].

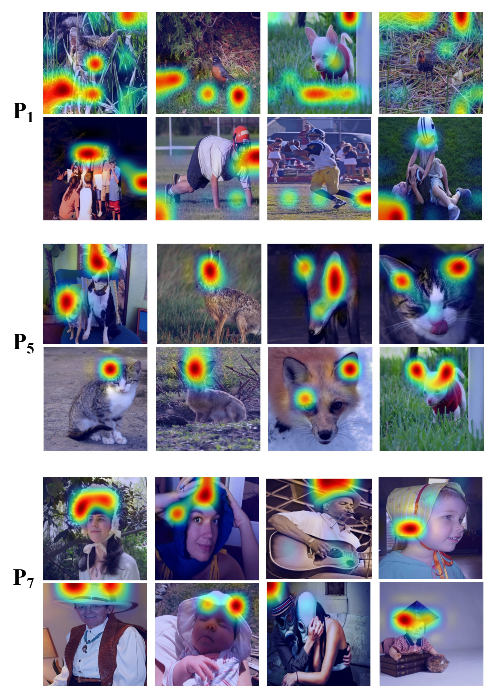

Figure 2. More Visualization of the same prompt on different images using Grad-CAM [2].

References

[1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009. 1

[2] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *ICCV*, 2017. 2

[3] Yabin Wang, Zhiwu Huang, and Xiaopeng Hong. S-prompts learning with pre-trained transformers: An occam's razor for domain incremental learning. In *arXiv preprint*, 2022. 1

[4] Zifeng Wang, Zizhao Zhang, Sayna Ebrahimi, Ruoxi Sun, Han Zhang, Chen-Yu Lee, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, et al. Dualprompt: Complementary prompting for rehearsal-free continual learning. *arXiv preprint*, 2022. 1

[5] Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Learning to prompt for continual learning. In *CVPR*, 2022. 1