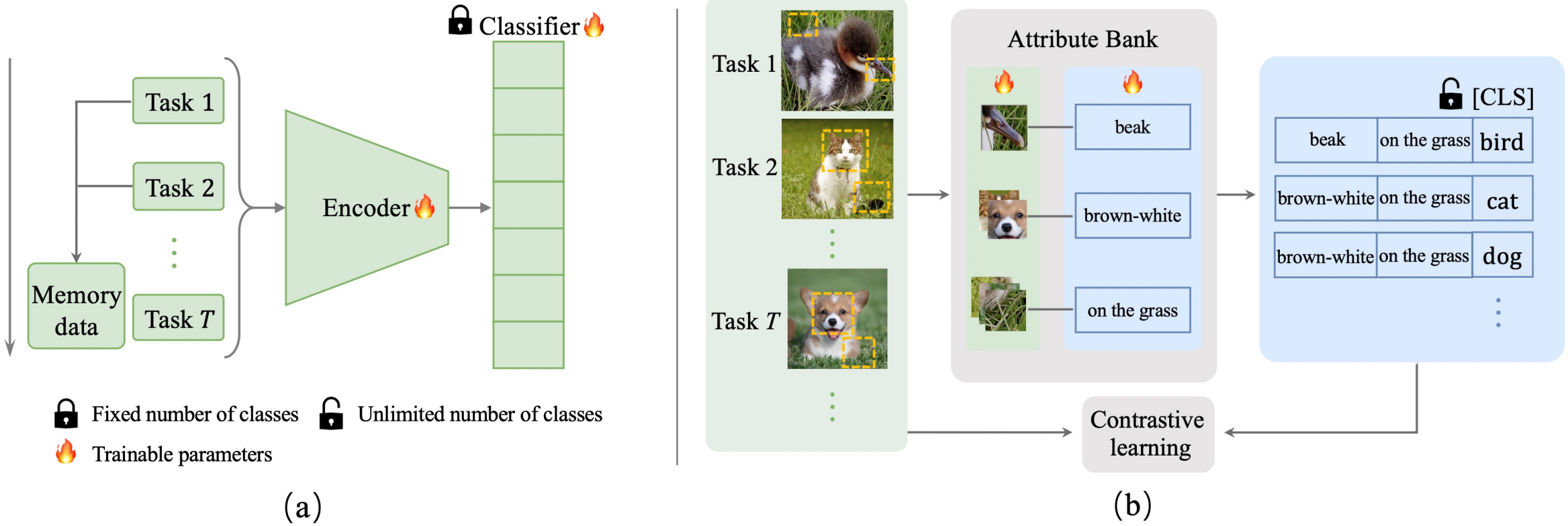
AttriCLIP: A Non-Incremental Learner for Incremental Knowledge Learning

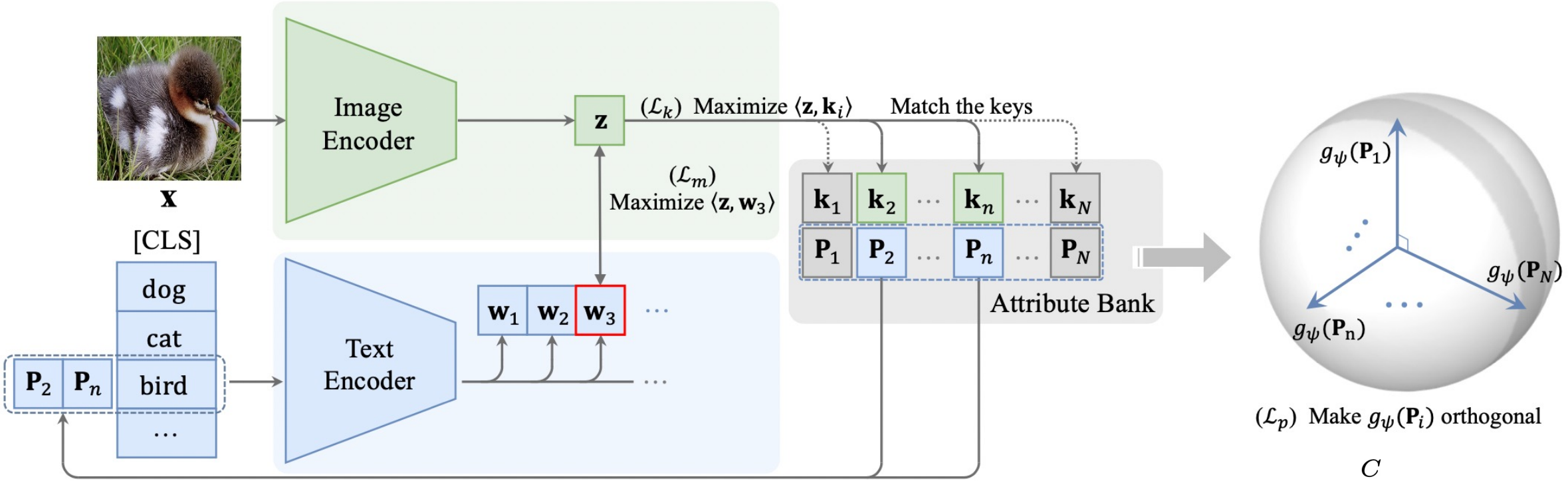
Wang Runqi, Duan Xiaoyue, Kang Guoliang, Liu Jianzhuang,
Lin Shaohui, Xu Songcen, Lv Jinhu, Zhang Baochang

CVPR2023-5850 TUE-AM-348

Motivation


Classifier
Task 1
Task 2
Task T
Memory data
Encoder
Fixed number of classes
Unlimited number of classes
Trainable parameters
(a)
Task 1
Task 2
Task T
Attribute Bank
beak
brown-white
on the grass
[CLS]
beak
on the grass
bird
brown-white
on the grass
cat
brown-white
on the grass
dog
Contrastive learning
(b)

Method



$$\{\mathcal{K}, \mathcal{P}\} \triangleq \{(\mathbf{k}_1, \mathbf{P}_1), \ldots, (\mathbf{k}_N, \mathbf{P}_N)\}$$

$$\mathcal{P}_j = \{\mathbf{P}_{j_i}\}_{i=1}^{C}$$

$$\mathcal{L}_k = \sum_{i=1}^{C} \gamma(\mathbf{z}_j, \mathbf{k}_{j_i})$$

$$\mathbf{z}_j = f_\theta(\mathbf{x}_j)$$

$$\mathbf{t}_k(\mathcal{P}_j) = \text{concat}(\mathbf{P}_{j_1}; \ldots; \mathbf{P}_{j_C}; [\text{CLS}]_k)$$

$$\mathcal{L}_p = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} |\langle g_\psi(\mathbf{P}_i), g_\psi(\mathbf{P}_j) \rangle|$$

$$\mathcal{K}_j = \text{Top-}C^{min}\{\gamma(\mathbf{z}_j, \mathbf{k}_{j_i})\}_{i=1}^{N}$$

$$p(y_i|\mathbf{x}_j) = \frac{e^{\langle \mathbf{z}, g_\psi(\mathbf{t}_{y_i}(\mathcal{P}_j)) \rangle / \tau}}{\sum_{k=1}^{K} e^{\langle \mathbf{z}, g_\psi(\mathbf{t}_k(\mathcal{P}_j)) \rangle / \tau}}$$

$$\mathcal{L} = \mathcal{L}_m + \lambda_k \mathcal{L}_k + \lambda_p \mathcal{L}_p$$

Experiments

Table 1. Average accuracy [14] of different continual learning methods on CIFAR100 [9]. The accuracy of Task $t, t \in \{1, 2, \ldots, 10\}$ reported here is the test accuracy averaged over all the previous tasks (*i.e.*, Tasks $1, 2, \ldots, t$).

Method	Memory	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task9	Task10
LwF	0	89.3	70.1	54.3	45.8	39.8	36.1	31.7	28.9	24.4	23.9
iCaRL	2000	88.7	78.1	72.4	67.2	63.7	60.2	56.4	54.4	51.9	49.5
iTAML	2000	89.2	89.0	87.3	86.2	84.3	82.1	80.7	79.1	78.4	77.8
ARI	2000	88.6	86.9	85.8	84.6	83.1	81.8	81.6	81.0	80.2	80.9
CoOp	1000	95.8	90.7	85.2	83.4	80.8	75.8	74.7	71.7	71.3	67.6
Continual-CLIP	0	96.7	92.2	86.0	80.4	77.5	75.8	73.0	71.4	69.8	66.7
AttriCLIP	0	**97.8**	**93.7**	**91.0**	**87.5**	**84.7**	**82.5**	**82.3**	**81.9**	**81.7**	**81.4**
Upper-bound	-	-	-	-	-	-	-	-	-	-	**86.3**

Table 2. Average accuracy [14] of different continual learning methods on ImageNet100 [4]. The accuracy of Task $t, t \in \{1, 2, \ldots, 10\}$ reported here is the test accuracy averaged over all the previous tasks (*i.e.*, Tasks $1, 2, \ldots, t$).

Method	Memory	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task9	Task10
iCaRL	2000	82.1	80.6	75.5	70.1	68.1	65.8	62.5	61.3	60.7	59.5
DER	2000	81.7	80.6	76.0	72.1	74.4	71.8	70.5	68.3	67.3	66.7
ARI	2000	87.6	85.4	83.1	82.6	80.4	80.8	80.5	80.1	79.6	79.3
CoOp	1000	89.2	83.2	76.7	79.8	79.9	82.34	79.7	80.1	80.3	79.3
Continual-CLIP	0	93.3	87.6	83.1	81.7	80.5	80.2	79.3	78.5	76.9	75.4
AttriCLIP	0	**95.4**	**89.4**	**84.5**	**86.7**	**84.4**	**86.6**	**85.9**	**85.6**	**86.9**	**83.3**
Upper-bound	-	-	-	-	-	-	-	-	-	-	**91.4**

Experiments

Table 4. 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
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**

Experiments

Table 6. Comparison of different loss functions for $\mathcal{L}_k$ on CIFAR100.

Loss function	Triplet loss	Cosine loss	MSE loss
Average acc.	80.22	**81.38**	80.81

Table 7. Average acc. of different loss weights λ_k on CIFAR100.

λ_k	0.1	0.3	0.5	0.7	0.9
Average acc.	80.28	80.30	81.11	**81.38**	80.86

Table 8. Average acc. of different loss weights λ_p on CIFAR100.

λ_p	0	0.1	0.3	0.5	0.7	0.9
Average acc.	78.10	79.23	**81.38**	81.28	81.17	80.88

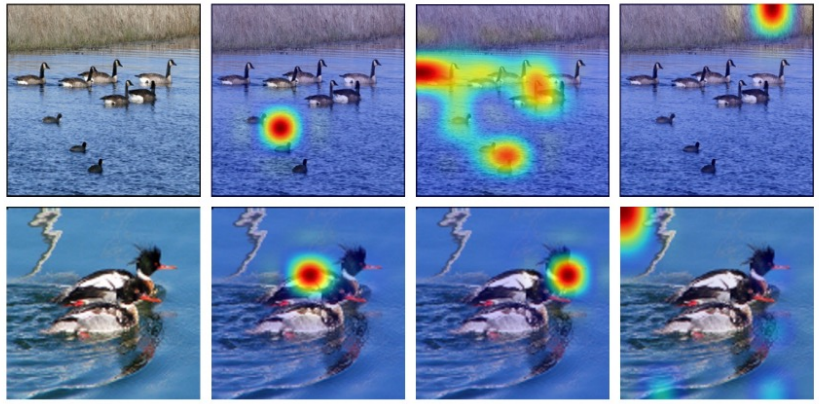


Figure 4. Visualization of the selected prompts of the same image using Grad-CAM [23].

Figure 5. Visualization of the same prompts on different images using Grad-CAM [23].

AttriCLIP: A Non-Incremental Learner for Incremental Knowledge Learning

Thanks for listening

CVPR2023