Class-Incremental Exemplar Compression for Class-Incremental Learning

SMU SINGAPORE MANAGEMENT UNIVERSITY · mpi max planck institut informatik

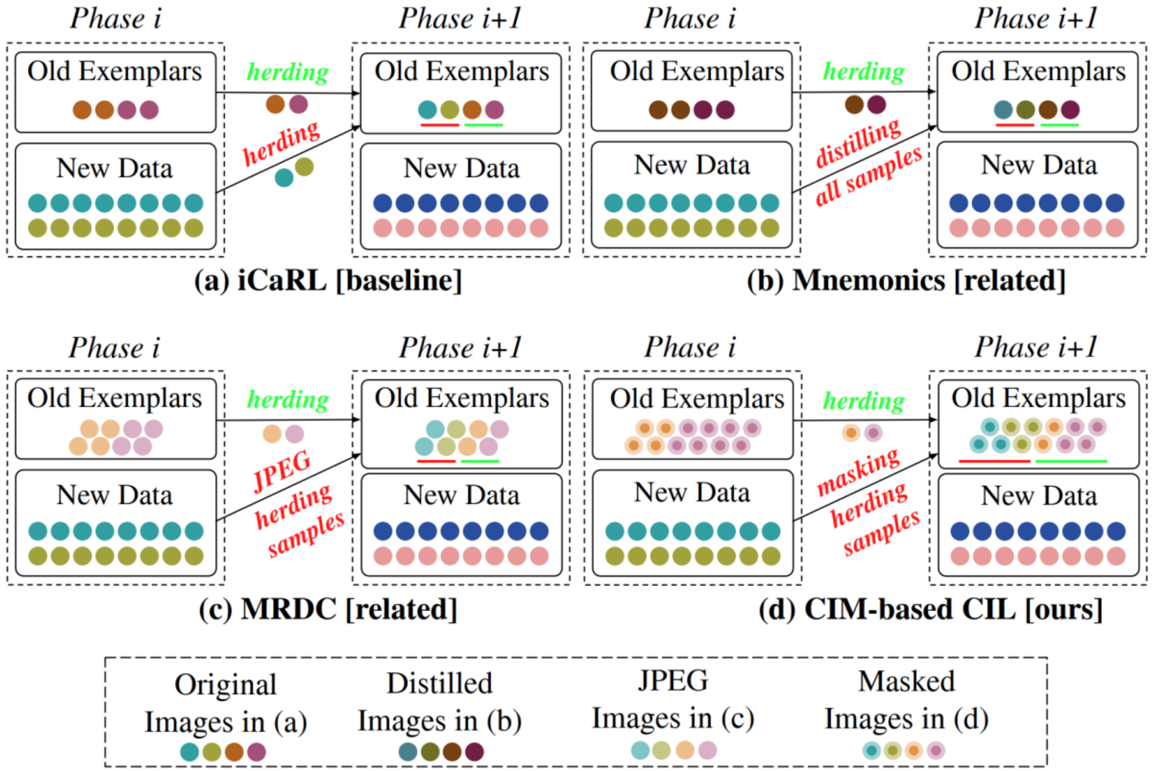
Zilin Luo, Yaoyao Liu, Bernt Schiele, Qianru Sun

Paper · Code

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Exemplar-based Class-Incremental Learning

- **Basic Task**
  - Different classes arrive in different phases.
  - Data in former phases is not available in later phases.
  - An extra memory with a limited budget is provided.
- **Challenge**
  - Catastrophic forgetting due to few-shot old-class exemplars.



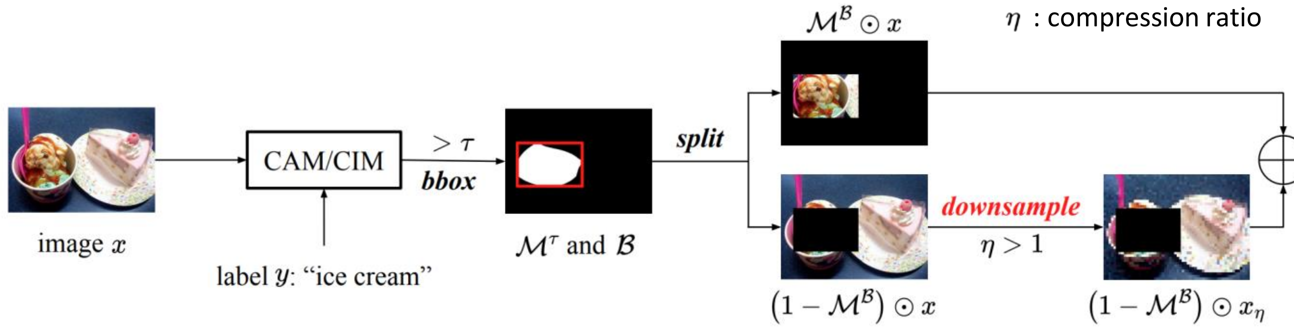
- **Contribution**
  - A pixel-selective compression framework for Exemplar-based CIL.
  - An optimization pipeline that associates the CIM model with the CIL model to produce adaptive compression masks.

Overall Compression Framework

τ : mask threshold

η : compression ratio

$$m_{\tilde{x}} = \frac{H_{\mathcal{B}}W_{\mathcal{B}}}{HW} + \frac{1}{\eta}\left(1 - \frac{H_{\mathcal{B}}W_{\mathcal{B}}}{HW}\right) = 1 - \left(1 - \frac{1}{\eta}\right)\cdot\left(1 - \frac{H_{\mathcal{B}}W_{\mathcal{B}}}{HW}\right)$$



Global Optimization Pipeline

(θ, ω): CIL model $\quad \phi$: CIM model

- **Task-level Optimization** $\quad (\theta_i, \omega_i) \leftarrow (\theta_i, \omega_i) - \lambda \nabla_{(\theta,\omega)} \mathcal{L}_{\text{CIL}}(\tilde{\mathcal{E}}_{0:i-1} \cup \mathcal{D}_i; \theta_i, \omega_i)$
- **Mask-level Optimization**

$$\begin{cases} \text{inner step: } \theta_i^+ \leftarrow \theta_i - \beta_1 \nabla_\theta \mathcal{L}_{\text{CIL}}(\tilde{\mathcal{E}}_{0:i-1} \cup \tilde{\mathcal{D}}_i(\phi_i); \theta_i, \omega_i) \\ \text{outer step: } \phi_i \leftarrow \phi_i - \beta_2 \nabla_\phi [\underbrace{\mathcal{L}_{\text{CE}}(\mathcal{D}_i; \theta_i^+, \omega_i)}_{①} + \underbrace{\mu \mathcal{R}(\phi_i)}_{②} + \underbrace{\mu' \mathcal{L}_{\text{CE}}(\tilde{\mathcal{E}}_{0:i-1} \cup \mathcal{D}_i; \theta_i, \phi_i, \omega_i)}_{③}] \end{cases}$$

① validation loss
② memory constraint
③ mask diversity loss

Experimental Results

- **Comparing with SOTA in multiple settings and datasets**

Method	Learning from Scratch (LFS)						Learning from Half (LFH)					
	Food-101			ImageNet-100			Food-101			ImageNet-100		
	N=5	10	20	5	10	20	5	10	25	5	10	25
DER [46]	73.88	70.76	64.39	78.50	76.12	73.79	78.13	73.45	-	79.08	77.73	-
DER *w/* ours	75.63	73.09	69.17	79.63	77.57	**75.36**	79.25	75.76	-	80.30	**79.05**	-
FOSTER [42]	75.03	72.72	66.73	79.93†	76.55†	74.49	79.08	75.07	68.08	80.07†	77.54	72.40*
FOSTER *w/* ours	**76.44**	**74.85**	**70.20**	**80.58**	**77.94**	75.23	**79.76**	**76.86**	**70.50**	**80.93**	78.66	**75.74**

#phases

Memory Budget	Method	N=5		N=10	
		Avg.	Last	Avg.	Last
$M=20k$	FOSTER [42]	69.21	64.88	68.34	60.14
	FOSTER *w/* ours	**69.93**	**66.05**	**69.53**	**62.07**
$M=5k$	FOSTER	57.19	49.42	54.72	44.96
	FOSTER *w/* ours	**61.37**	**54.46**	**59.48**	**50.83**

- The results on ImageNet-1000 (LFS) show that our method brings larger performance improvement under a stricter memory budget.

- **Ablation Studies**

Ablation Method	Food-101		ImageNet-100	
	N=10	20	10	20
1 Baseline	72.72	66.73	76.55	72.37
2 Artifact Aug.	71.38	66.03	75.63	71.45
3 Full Comp.	73.03	67.38	76.92	73.26
4 Random Acti.	73.10	67.54	76.88	73.54
5 Center Acti.	73.29	67.88	76.78	73.82
6 Class Acti.	73.76	68.65	77.21	74.67
7 Phase-wise τ	73.83	69.17	77.06	74.78
8 Joint Train	73.44	69.01	77.34	74.59
9 BOP (ours)	74.85	**70.20**	**77.94**	75.23
10 LastBlock Only	74.55	69.87	77.72	74.86
11 Fg Compressed	**75.02**	70.13	77.87	**75.46**

- Line 1-2: SOTA baselines.
- Line 3-6: different activation methods.
- Line 7-9: different optimization strategies.
- Line 10-11: two variants.

- **Performances on Different-Size Objects**

Metric	Small	Middle	Large
Mean of #Exemplars	39.40	38.30	34.77
Last Acc. (%, baseline)	66.13	68.40	69.93
Last Acc. (%, ours)	70.00	71.10	72.26
Improvement (%)	+3.87	+3.65	+2.33

- The baseline has 20 exemplars for each individual class.
- Our method achieves larger performance boost for small objects.

- "DER: Dynamically Expandable Representation for Class Incremental Learning." In CVPR 2021.
- "FOSTER: Feature Boosting and Compression for Class-Incremental Learning." In ECCV 2022.