Prototype Reminiscence and Augmented Asymmetric Knowledge Aggregation for Non-Exemplar Class-Incremental Learning

Wuxuan Shi[1] , Mang Ye[1,2*]   * Indicates Corresponding Author

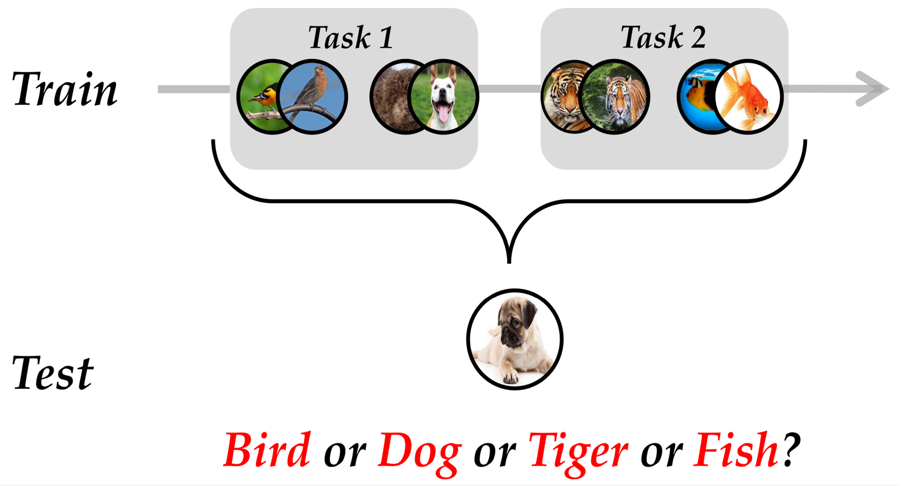
[1]School of Computer Science, Wuhan University   [2]Hubei Luojia Laboratory

ICCV23 PARIS

Project Page With Codes

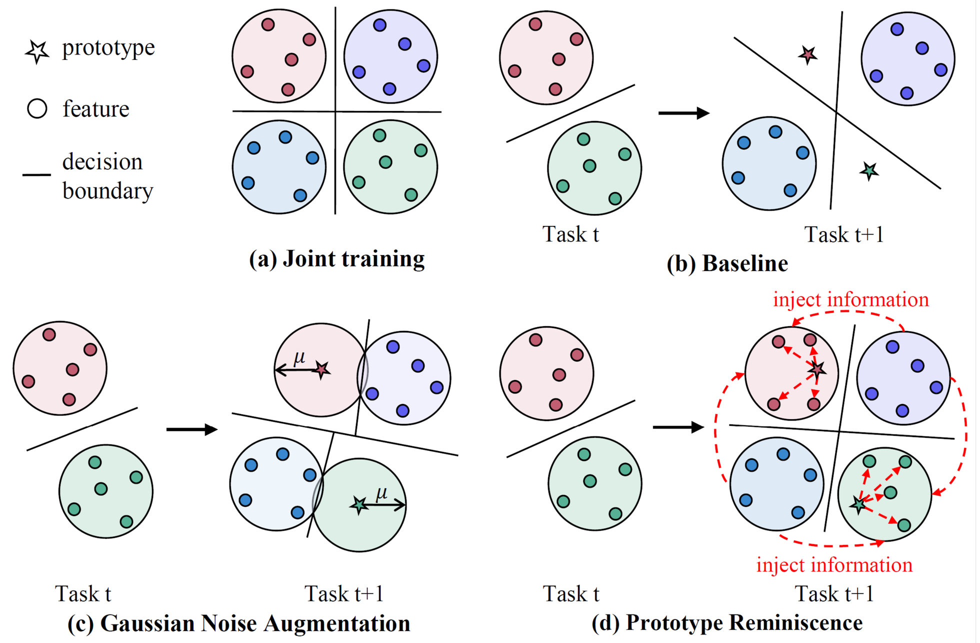
Introduction:

- Class-incremental learning (CIL) aims to continually learn new classes without forgetting the old ones, and non-exemplar class-incremental learning (NECIL) accomplishes this goal without access to the data of previously learned classes [1].

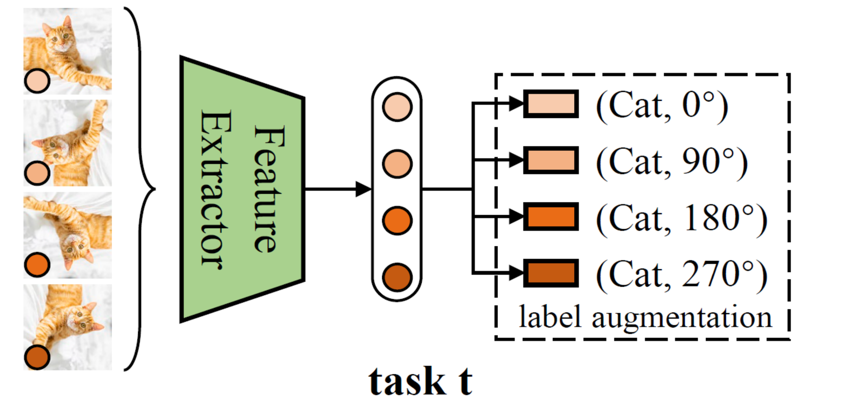


Main Idea:

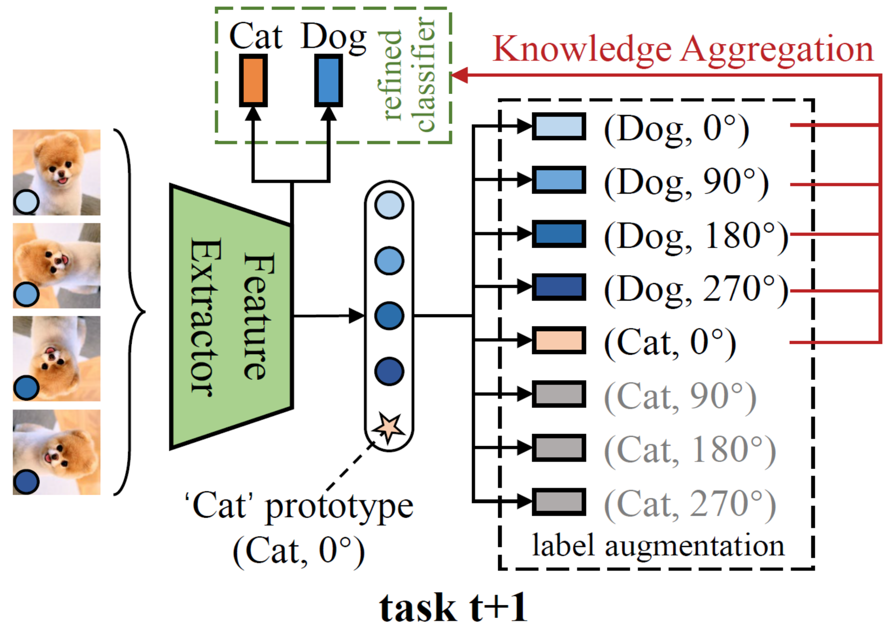
- **For catastrophic forgetting**
  - ☐ We propose a **prototype reminiscence** mechanism to track the evolution of the old class representations by injecting new knowledge from the updating network while reshaping the feature distribution.



- **For the plasticity of model**
  - ☐ We introduce **self-supervised label augmentation** to learn generalizable and transferable representations.



  - ☐ We propose a **asymmetric knowledge aggregation** scheme to selectively aggregate the valuable knowledge in the augmented classifier—valid weights of past tasks are extracted, and the information captured on the current task is sufficiently exploited.



Proposed Method:

- **Prototype Reminiscence (PR)**
  - ☐ Compute and memorize one prototype in the deep feature space for each class:

$$P_{t,k_{new}} = \frac{1}{N_{t,c_{new}}} \sum_{n=1}^{N_{t,c_{new}}} \mathcal{F}\left(\mathbf{X}_{t,c_{new}};\theta_t\right)$$

  - ☐ Then we perform a random bidirectional interpolation operation between prototypes P^n_{old} and current features F^m_{new} :

$$F^n_{old} = \begin{cases} (1-\lambda)\left(P^n_{old}\right) + \lambda\left(F^m_{new}\right), p_e < 0.5 \\ (1+\lambda)\left(P^n_{old}\right) - \lambda\left(F^m_{new}\right), otherwise \end{cases}$$

- **Self-supervised Label Augmentation (SLA)**
  - ☐ Augment the new classes with rotation as self-supervision:

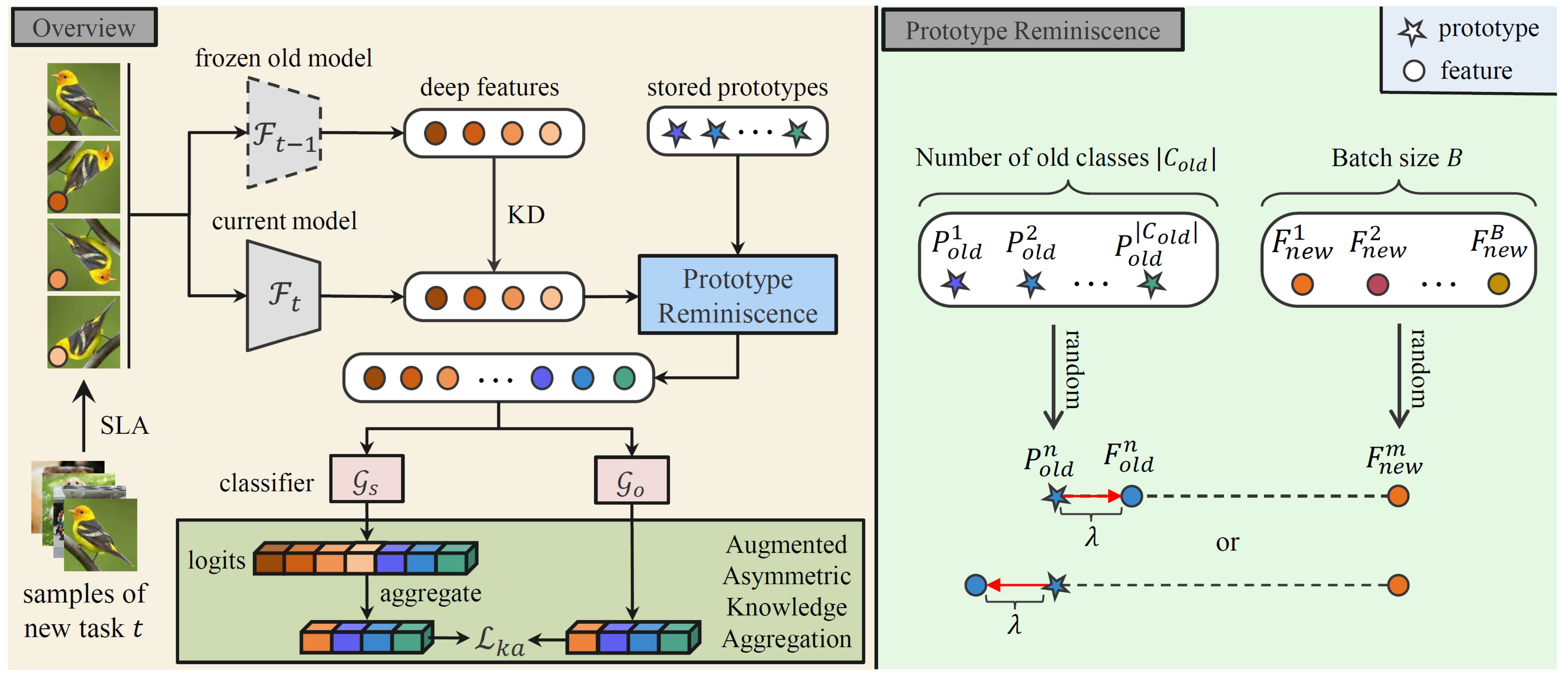
$$\tilde{x}^i_t = \{x^{i,j}_t\}_{j=0}^3 = \{rotate(x^i_t, j\times 90°)\}_{j=0}^3 \qquad \tilde{y}^i_t = \{y^i_t \times 4 + j\}_{j=0}^3$$

- **Asymmetric Knowledge Aggregation (AKA)**
  - ☐ Aggregate conditional probabilities of the augmented classifier $\mathcal{G}_s$:

$$\mathcal{P}_{agg}(c|\tilde{F}_{new}) = \frac{\exp\left(\frac{1}{4}\sum_{j=0}^{3} w^T_{c,j}F^j_{new}\right)}{\sum_{k=1}^{K}\exp\left(\frac{1}{4}\sum_{j=0}^{3} w^T_{k,j}F^j_{new}\right)}$$

  - ☐ Transfer the aggregated knowledge to another refined classifier $\mathcal{G}_o$:

$$\begin{aligned}\mathcal{L}_{ka} &= KLD(\mathcal{P}_{agg}(\cdot|F)\|\mathcal{G}_O(F;\phi)) \\ &= \mathcal{P}_{agg}(\cdot|F)\log\frac{\mathcal{P}_{agg}(\cdot|F)}{\mathcal{G}_O(F;\phi)}\end{aligned}$$



Experimental Results:

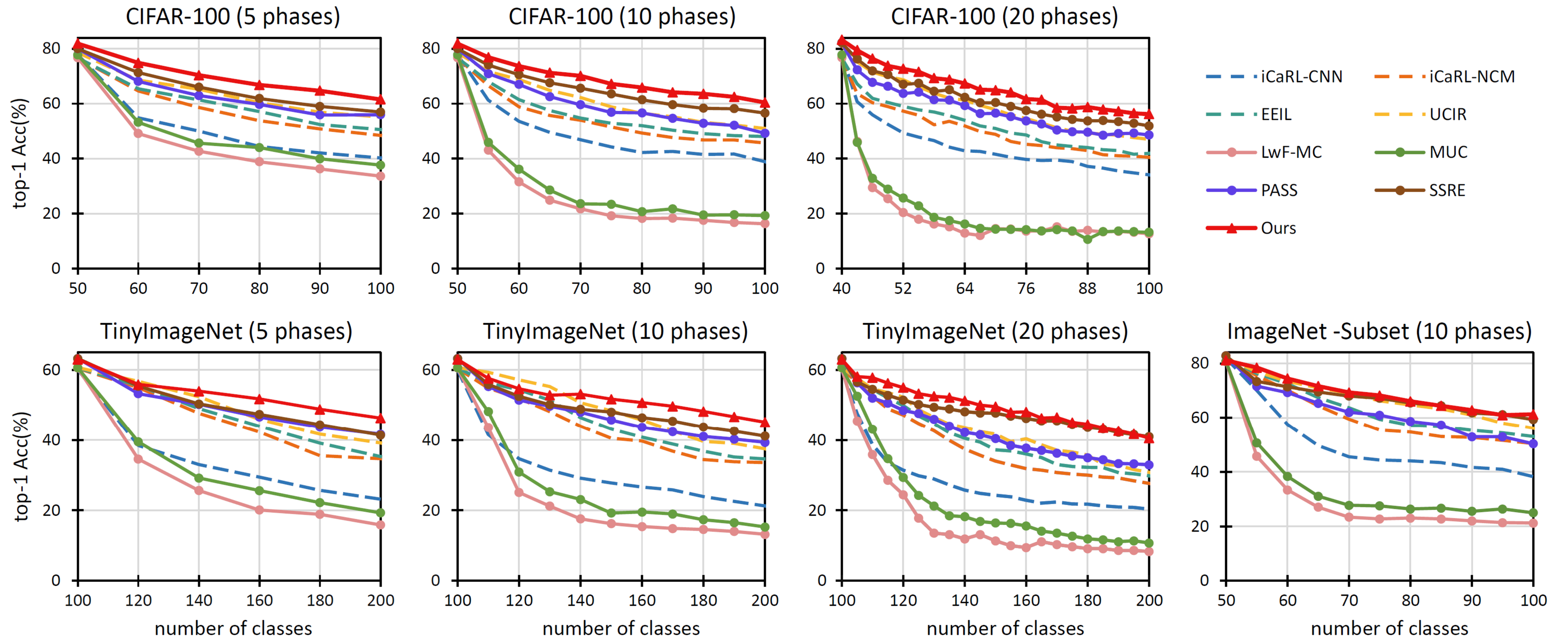
- Comparison of the average incremental accuracy with the SOTA method

Methods	CIFAR-100						ImageNet-Subset
	P=5	P=10	P=20	P=5	P=10	P=20	P=10
LwF_MC	45.93	27.43	20.07	29.12	23.10	17.43	31.18
PASS	63.47	61.84	58.09	49.55	47.29	42.07	61.80
SSRE	65.88	65.04	61.70	50.39	48.93	48.17	67.69
Ours	**70.02**	**68.86**	**65.86**	**53.32**	**52.61**	**49.83**	**68.98**

- Ablation study of our method

Components			CIFAR-100			Tiny-ImageNet		
PR	SLA	AKA	P=5	P=10	P=20	P=5	P=10	P=20
			56.27	51.02	43.98	37.93	32.44	23.98
√			66.21	63.80	57.31	45.85	44.04	35.93
	√		61.27	59.59	55.14	46.65	43.88	38.13
√	√		68.48	67.56	65.03	52.15	51.67	49.30
√	√	√	**70.02**	**68.86**	**65.86**	**53.32**	**52.61**	**49.83**

- Complete curves of the classification accuracy



Reference:

[1] Da-Wei Zhou, Qi-Wei Wang, Zhi-Hong Qi, Han-Jia Ye, De-Chuan Zhan and Ziwei Liu. Deep class-incremental learning: A survey. arXiv preprint arXiv:2302.

Acknowledgement:

This work is supported by the Key Research and Development Program of Hubei Province (2021BAA187), National Natural Science Foundation of China under Grant (62176188), Zhejiang lab (NO.2022NF0AB01), the Special Fund of Hubei Luojia Laboratory (220100015) and CAAI-Huawei MindSpore Open Fund.