Prototypes as Anchors: Tackling *Unseen Noise* for online continual learning

Shao-Yuan Li [a,b,c], Yu-Xiang Zheng [a], Sheng-Jun Huang [a], Songcan Chen [a], Kangkan Wang [d,*]

[a] *MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, China*

[b] *State Key Lab. for Novel Software Technology, Nanjing University, Nanjing, 211106, PR China*

[c] *Joint Laboratory of Spatial Intelligent Perception and Large Model Application, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, PR China*

[d] *School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, PR China*

Background

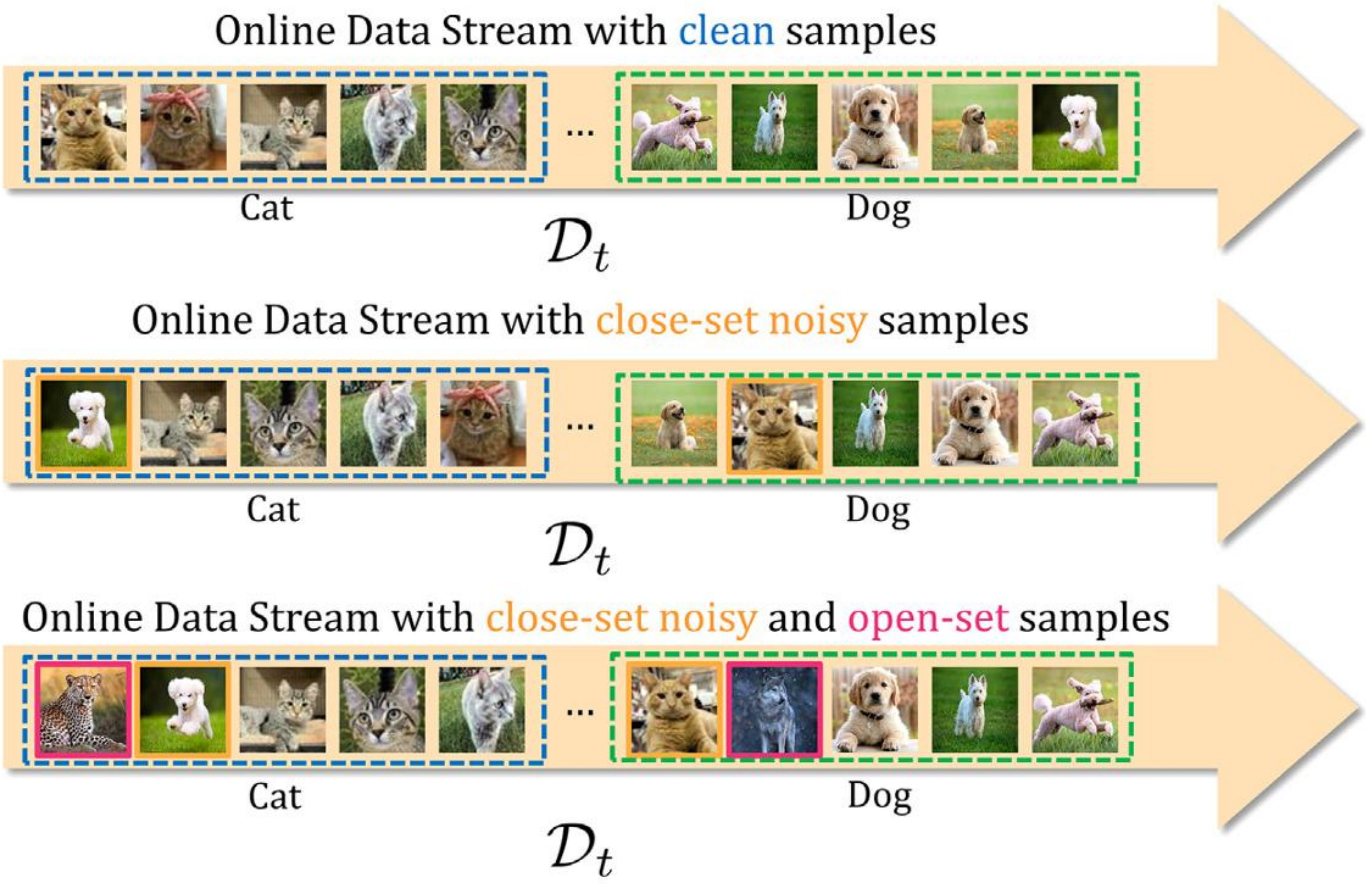


Fig. 1. Comparison of three online CIL scenarios: (Top) Each task consists of correctly labeled samples. (Mid) Each task contains*closed-set* noisy samples. (Bottom) Each task contains both *closed-set* and *open-set* noisy samples.

PAA

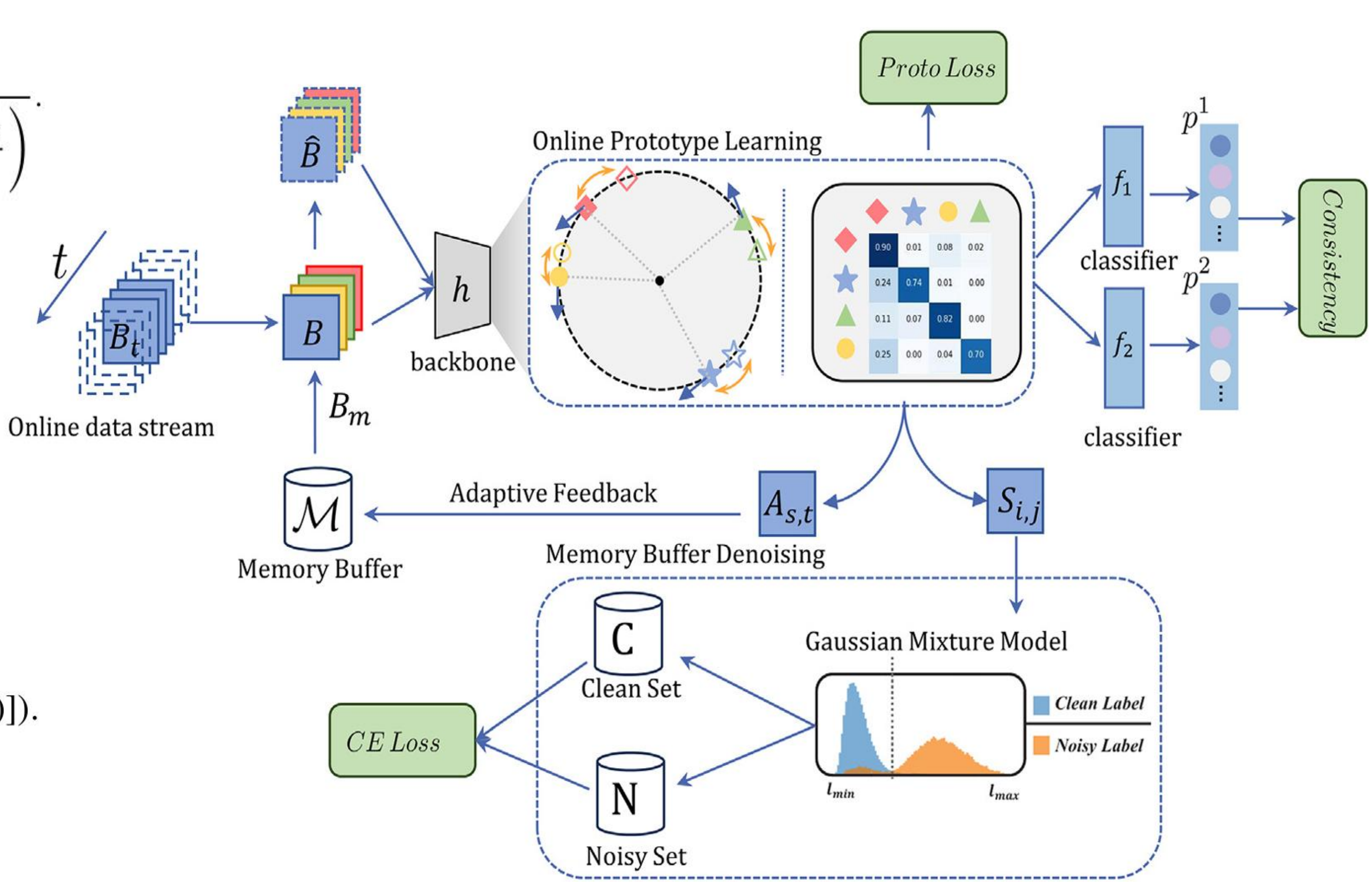
$$\mathcal{L}_{\text{NEW}} = \frac{-1}{|\mathcal{P}|}\sum_{j=1}^{|\mathcal{P}|}\log\frac{\exp\left(\frac{p_j^{\mathrm{T}}\hat{p}_j}{\tau_1}\right)}{\sum_k \exp\left(\frac{p_j^{\mathrm{T}}\hat{p}_k}{\tau_1}\right)+\sum_{k\neq j}\exp\left(\frac{p_j^{\mathrm{T}}p_k}{\tau_1}\right)}.$$

$$\mathcal{L}_{\text{OLD}} = \frac{-1}{|B_m|}\sum_{i=1}^{|B_m|}\log\frac{\exp\left(\hat{\boldsymbol{z}}_i\cdot\boldsymbol{p}^m_{\hat{y}_i}/\tau_2\right)}{\sum_{c=1}^{C_{old}}\exp\left(\hat{\boldsymbol{z}}_i\cdot\boldsymbol{p}^m_c/\tau_2\right)}.$$

$$\mathcal{L}_{\text{CON}}(\boldsymbol{x}) = \left|f_1(h(\boldsymbol{x})) - f_2(h(\boldsymbol{x}))\right|,$$
$$\mathcal{L}_{\text{CON}} = \sum_{\boldsymbol{x}\in B_t\cup B_m}\mathcal{L}_{\text{CON}}(\boldsymbol{x}).$$

$$\mathcal{L}_{\text{CE}} = \sum_{(\boldsymbol{x},\hat{\boldsymbol{y}})\in\mathcal{M}}\text{CE}(\hat{\boldsymbol{y}}, 0.5 * [f_1(h(\boldsymbol{x})) + f_2(h(\boldsymbol{x}))]).$$

$$\mathcal{L} = \mathcal{L}_{\text{NEW}} + \mathcal{L}_{\text{OLD}} + \mathcal{L}_{\text{CON}} + \mathcal{L}_{\text{CE}}.$$



Memory buffer denoising

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

$$S_{i,j} \propto \exp(-\|\boldsymbol{p_j} - g(h(\boldsymbol{x_i}))\|_2^2)$$

$$p_G\left(g \mid S_{i,j}\right) = \frac{\pi_g \cdot \mathcal{N}\left(S_{i,j} \mid \mu_g, \sigma_g^2\right)}{\sum_{l=1}^{C_M} \pi_l \cdot \mathcal{N}\left(S_{i,j} \mid \mu_l, \sigma_l^2\right)}$$

$$\mathbf{C} := \left\{(\boldsymbol{x}, \tilde{\boldsymbol{y}}) \in \mathcal{M} : p_G\left(g = \tilde{\boldsymbol{y}} \mid S_{i,j}\right) \geq \lambda_1\right\}$$

$$\mathbf{N} := \left\{(\boldsymbol{x}, \tilde{\boldsymbol{y}}) \in \mathcal{M} : p_G\left(g = \tilde{\boldsymbol{y}} \mid S_{i,j}\right) < \lambda_1\right\}$$

$$\mathbf{N_c} := \left\{(\boldsymbol{x}, \tilde{\boldsymbol{y}}) \in \mathbf{N} : \mathbf{1}\left\{-s_k(\boldsymbol{x}) < \lambda_2\right\}\right\}$$

Experiment

南京航空航天大學 NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Table 1
Last test accuracy on CIFAR100N and CIFAR80N under various noise scenarios. The size of the memory buffer both are 2000. Bold value represents the best method.

Methods	CIFAR100N					CIFAR80N				
	Sym.			Asym.		Sym.			Asym.	
	20	40	60	20	40	20	40	60	20	40
MIR	15.1 ± 1.0	7.7 ± 0.3	5.1 ± 0.4	11.3 ± 1.5	7.2 ± 0.6	11.6 ± 0.5	4.2 ± 1.6	1.4 ± 0.3	7.7 ± 0.6	4.4 ± 1.1
+ JoCoR	15.7 ± 1.6	7.9 ± 0.4	5.5 ± 2.1	13.7 ± 0.8	7.7 ± 0.9	12.9 ± 0.4	5.8 ± 1.2	2.6 ± 0.3	10.6 ± 1.4	3.7 ± 0.6
+ PNP	15.5 ± 0.7	8.2 ± 1.3	5.4 ± 1.7	12.9 ± 0.3	7.9 ± 1.2	13.1 ± 0.8	4.7 ± 0.9	2.3 ± 0.5	10.9 ± 0.5	4.8 ± 0.7
GDumb	15.8 ± 1.7	8.1 ± 0.5	5.6 ± 1.1	13.9 ± 1.3	9.0 ± 0.9	13.0 ± 0.2	5.0 ± 1.7	2.5 ± 0.8	10.7 ± 0.9	5.8 ± 0.5
+ JoCoR	16.1 ± 0.3	8.9 ± 0.4	6.1 ± 2.6	15.0 ± 0.2	9.5 ± 1.3	12.8 ± 0.7	5.9 ± 1.3	3.6 ± 2.6	11.3 ± 0.2	6.9 ± 1.3
+ PNP	16.5 ± 0.2	9.3 ± 0.7	6.4 ± 0.4	14.7 ± 0.8	10.1 ± 2.4	12.7 ± 0.8	7.0 ± 1.1	2.6 ± 0.4	11.9 ± 0.7	7.4 ± 1.1
OnPro	24.7 ± 0.1	22.8 ± 1.4	16.5 ± 0.3	22.4 ± 0.7	19.2 ± 0.2	22.5 ± 1.3	20.1 ± 0.9	14.4 ± 0.4	19.7 ± 1.0	16.2 ± 1.2
+ JoCoR	23.6 ± 1.4	23.1 ± 0.9	17.8 ± 1.2	23.1 ± 1.5	20.8 ± 1.4	21.1 ± 0.8	19.6 ± 1.0	15.4 ± 0.7	20.8 ± 1.3	17.8 ± 0.9
+ PNP	24.1 ± 0.9	20.6 ± 1.3	16.4 ± 0.2	22.7 ± 1.0	19.4 ± 0.6	21.6 ± 0.9	17.3 ± 1.1	12.8 ± 0.2	19.1 ± 0.7	17.1 ± 1.1
PCR	23.8 ± 0.1	21.7 ± 0.4	17.1 ± 0.7	22.3 ± 1.9	18.2 ± 0.8	21.6 ± 1.3	18.9 ± 0.8	14.6 ± 0.7	18.9 ± 1.1	14.5 ± 0.8
+ JoCoR	24.3 ± 0.5	22.6 ± 0.8	18.9 ± 1.5	22.7 ± 0.4	19.6 ± 0.5	22.1 ± 0.6	19.6 ± 1.0	16.3 ± 0.8	19.2 ± 1.2	16.3 ± 0.5
+ PNP	23.1 ± 0.4	21.6 ± 0.6	18.7 ± 1.1	22.1 ± 0.1	18.9 ± 0.2	19.6 ± 1.0	18.5 ± 0.7	15.6 ± 0.7	19.0 ± 0.6	15.8 ± 1.0
SPR	21.5 ± 0.1	21.1 ± 0.3	18.1 ± 1.0	20.5 ± 0.6	19.8 ± 2.2	18.3 ± 0.2	18.7 ± 0.3	16.1 ± 1.1	18.4 ± 0.8	17.7 ± 2.0
CNLL	38.7 ± 0.6	32.1 ± 0.4	26.2 ± 1.2	39.0 ± 0.9	32.6 ± 1.1	30.1 ± 0.5	28.8 ± 0.4	22.5 ± 0.7	34.1 ± 0.9	26.5 ± 1.1
PuriDivER	32.1 ± 0.4	30.6 ± 0.3	24.7 ± 0.9	31.4 ± 0.3	22.5 ± 0.2	28.9 ± 0.4	24.0 ± 0.3	21.5 ± 0.8	28.1 ± 0.5	20.4 ± 0.6
PAA(ResNet)	40.2 ± 1.6	35.4 ± 0.5	28.3 ± 1.4	37.8 ± 1.9	34.7 ± 1.7	36.6 ± 0.6	32.8 ± 0.4	29.7 ± 0.1	35.4 ± 0.7	31.9 ± 0.2
PAA+	**41.4 ± 1.1**	**37.8 ± 0.9**	**30.2 ± 2.1**	**39.6 ± 2.4**	**36.1 ± 1.2**	**38.2 ± 1.0**	**34.3 ± 0.7**	**31.5 ± 0.3**	**37.9 ± 1.5**	**34.1 ± 0.5**

Experiment

Table 2
Last test accuracy on ImageNet100N under various noise scenarios. The size of the memory buffer is 2000. Bold value represents the best method.

Methods	ImageNet100N				
	Sym.			Asym.	
	20	40	60	20	40
MIR	7.8 ± 2.3	4.1 ± 0.8	2.8 ± 1.2	6.0 ± 1.0	3.9 ± 1.5
+ JoCoR	8.2 ± 0.9	4.4 ± 0.6	3.1 ± 1.0	7.3 ± 0.7	4.2 ± 1.1
+ PNP	8.0 ± 1.1	4.5 ± 0.9	3.0 ± 1.3	6.8 ± 0.5	4.4 ± 0.8
GDumb	8.3 ± 1.3	4.6 ± 0.7	3.1 ± 1.4	7.4 ± 1.2	4.9 ± 0.6
+ JoCoR	8.5 ± 0.5	5.1 ± 0.8	3.5 ± 1.6	8.2 ± 0.4	5.4 ± 0.9
+ PNP	8.8 ± 0.4	5.4 ± 0.9	3.7 ± 0.6	7.9 ± 0.9	5.6 ± 1.3
OnPro	13.1 ± 0.3	12.4 ± 0.9	9.4 ± 0.5	12.0 ± 0.5	10.3 ± 0.4
+ JoCoR	12.4 ± 0.8	12.5 ± 0.6	10.2 ± 0.8	12.4 ± 0.9	11.2 ± 0.9
+ PNP	12.7 ± 0.6	11.2 ± 0.8	9.5 ± 0.4	12.0 ± 0.7	10.6 ± 0.5
PCR	12.5 ± 0.2	11.6 ± 0.6	10.0 ± 0.8	11.8 ± 1.2	10.2 ± 0.7
+ JoCoR	12.9 ± 0.7	12.0 ± 0.9	10.8 ± 1.3	12.0 ± 0.6	10.4 ± 0.6
+ PNP	12.3 ± 0.5	11.5 ± 0.7	10.7 ± 0.9	11.8 ± 0.3	10.8 ± 0.5
SPR	11.3 ± 0.3	11.2 ± 0.5	9.6 ± 0.9	10.9 ± 0.8	10.7 ± 1.7
CNLL	21.2 ± 0.8	17.3 ± 0.6	13.8 ± 1.4	**21.3** ± **1.1**	17.5 ± 1.3
PuriDivER	17.0 ± 0.6	16.2 ± 0.5	13.1 ± 1.0	16.8 ± 0.6	12.1 ± 0.5
PAA	22.1 ± 1.4	19.0 ± 0.7	15.0 ± 1.3	20.8 ± 1.6	18.7 ± 1.5
PAA+	**22.7** ± **1.0**	**20.4** ± **0.9**	**16.0** ± **1.7**	20.3 ± 2.0	**19.4** ± **1.1**

Thanks