Online Continual Learning on a Contaminated Data Stream with Blurry Task Boundaries

Jihwan Bang[1,2] Hyunseo Koh[3] Seulki Park[1,2,4] Hwanjun Song[1,2] Jung-Woo Ha[1,2] Jonghyun Choi[1,2,5]

1 NAVER CLOVA
2 NAVER AI LAB
3 GIST 4 SEOUL NATIONAL UNIVERSITY 5 YONSEI UNIVERSITY

CVPR 2022 JUNE 19-24 NEW ORLEANS • LOUISIANA

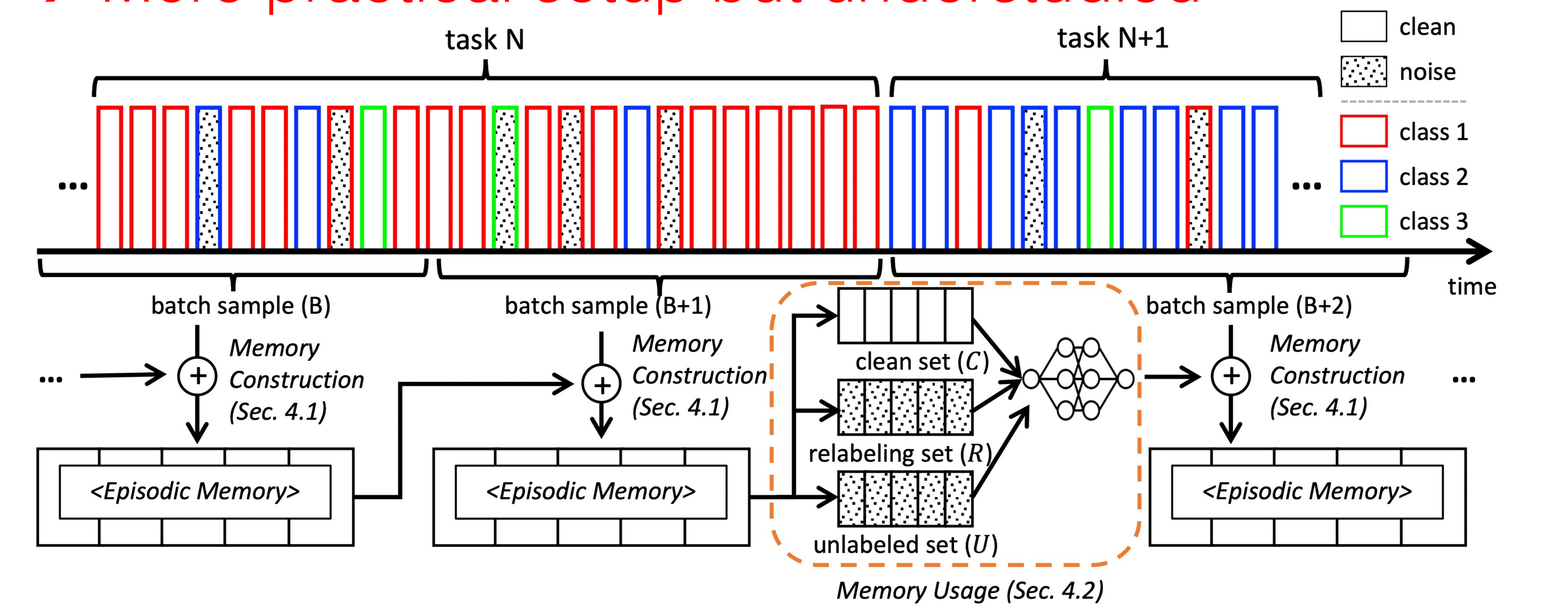
Problem Set-up

tl;dr. A new realistic continual learning set-up with noisy data and a method for it

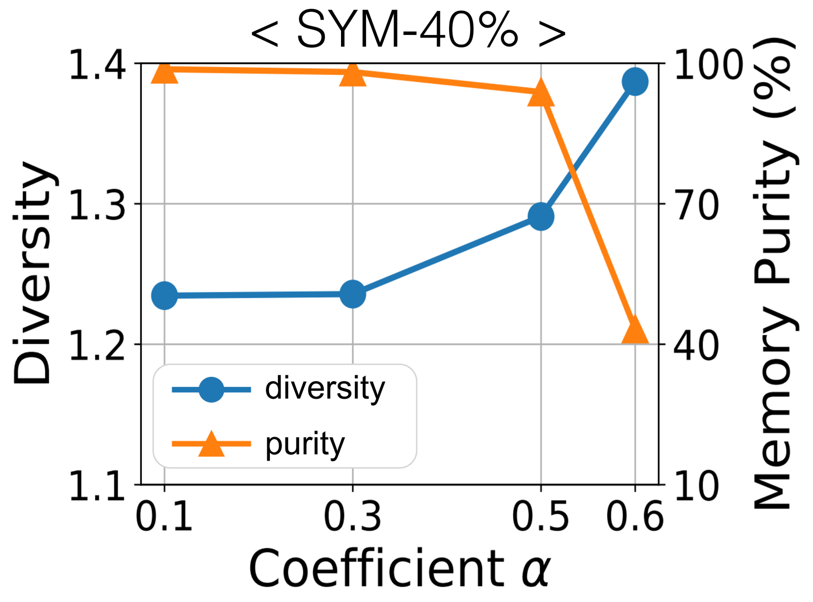
A Novel CL Set-up with Noisy Data

Online, Noisy, Blurry CL Task Set-up. Continuously changing data distribution with noisy labels in online

→ More practical setup but understudied



Trade-off between Purity and Diversity



< Last Accuracy on various setups >

	CIFAR-10 (K=500)				
	Sym.			Asym.	
α	20	40	60	20	40
0.1	57.4	55.1	51.6	58.9	47.4
0.3	58.1	55.3	**52.6**	58.4	**49.0**
0.5	**60.5**	**57.0**	47.1	**60.4**	46.7
0.6	29.5	30.3	24.4	36.0	33.7
1.0	26.1	19.4	19.3	33.0	28.1

α : coefficient for balancing diversity and purity while sampling

- Trade-off of purity and diversity empirically
  - Clean examples (small losses) less affect to model than the others
  - Large loss examples can be noisy with high probability
- Need to find the optimal α adaptively on various noisy setups for maximizing the accuracy

Method

Memory Construction: Purity vs. Diversity

Balancing purity and diversity is crucial in sampling

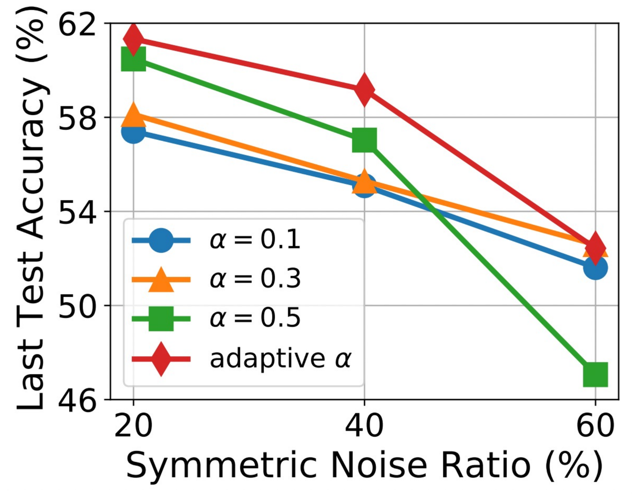
Drop the example that has the largest $S(x_i, \tilde{y}_i)$

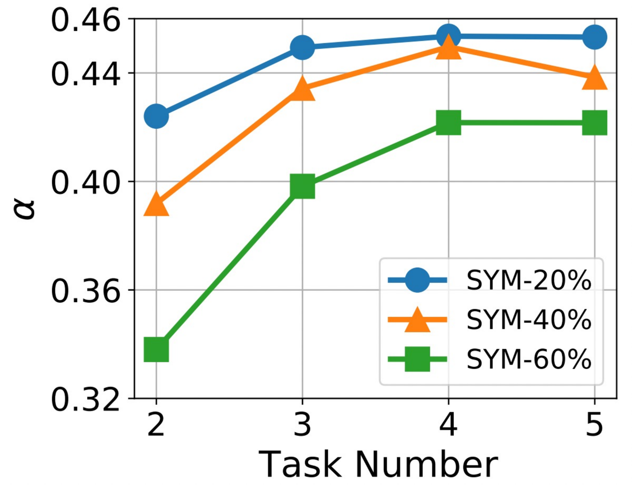
$$S(x_i, \tilde{y}_i) = (1-\alpha_k)\overbrace{\ell(x_i, \tilde{y}_i)}^{\text{purity}}$$

$$+\,\alpha_k \overbrace{\frac{1}{|\mathcal{M}[\tilde{y}_i]|}\sum_{\hat{x}_i \in \mathcal{M}[\tilde{y}_i]} \cos\big(f_{rel}(x_i;\tilde{y}_i), f_{rel}(\hat{x}_i;\tilde{y}_i)\big)}^{\text{diversity}}$$

Adaptive coefficient strategy

$$\alpha_k = 0.5 \cdot \min\left(1/\ell(\mathbf{X}_k, \tilde{\mathbf{Y}}_k), 1\right)$$





Memory Usage: Split data into 3 groups; C, R, U

M contains noisy data → it needs to be splitted into three groups to apply different training methods

- C: correct labels
- R: incorrect labels, correct prediction
- U: incorrect labels, incorrect prediction

Re-labeling on R

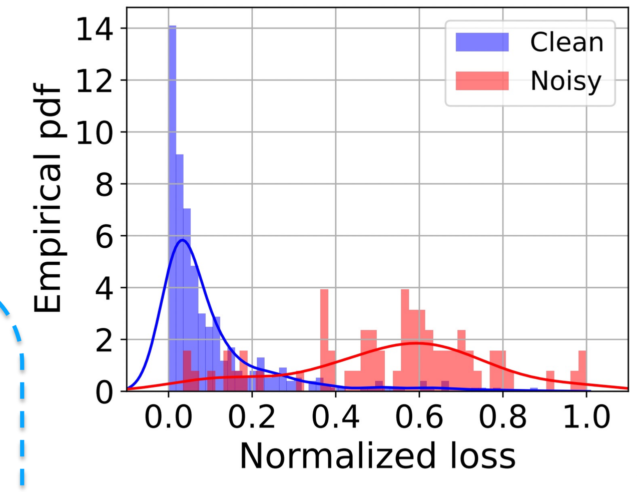
$$\hat{y}_i = p_G(u|\mathrm{U}(x_i))\cdot p_m(x_i) + \big(1.0 - p_G(u|\mathrm{U}(x_i))\big)\cdot \tilde{y}_i$$

Consistency regularization on U

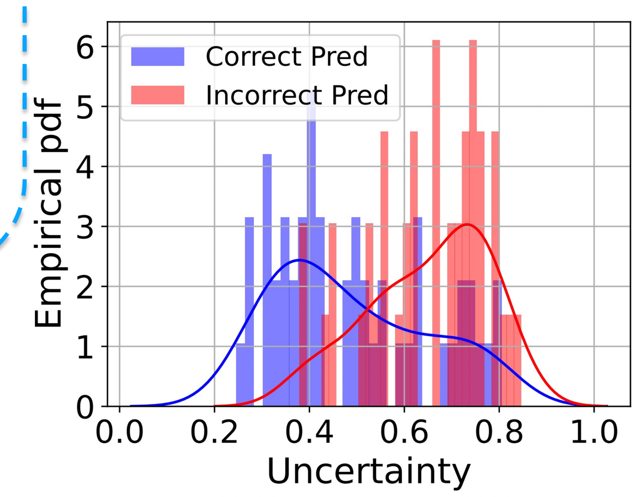
$$\ell_{reg} = \frac{1}{|\mathcal{U}|}\sum_{x_i \in \mathcal{U}} \left\|p_m(s(x_i);\Theta) - p_m(w(x_i);\Theta)\right\|_2$$

$$\ell(\mathcal{M}) = \ell_{cls}(\mathcal{C}\cup\mathcal{R}) + \eta \cdot \ell_{reg}(\mathcal{U})$$

< GMM: $M \rightarrow C,\ N$ >



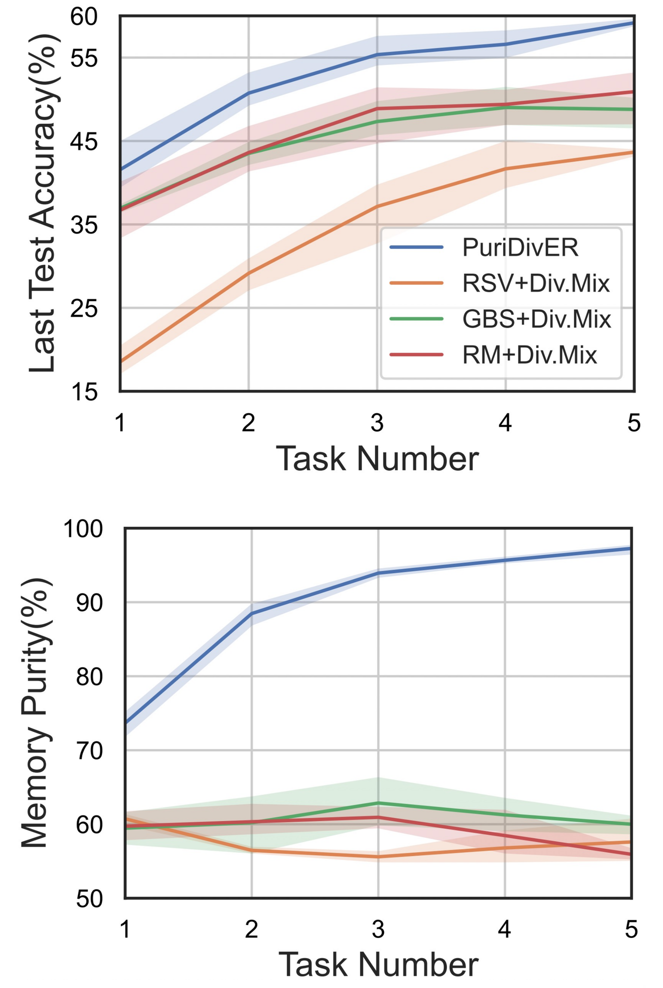
< GMM: $N \rightarrow R,\ U$ >



Experiments

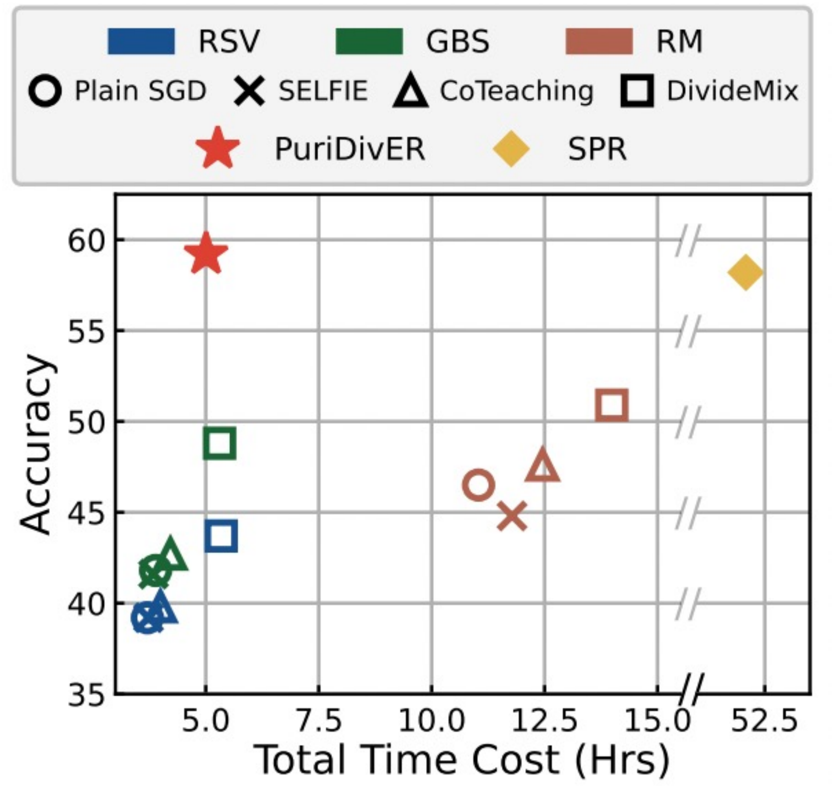
Last Accuracy on CIFAR-10 with Various Noise

Methods	CIFAR-10				
	Sym.			Asym.	
	20	40	60	20	40
RSV [29]	54.5 ± 2.1	39.2 ± 0.9	28.7 ± 0.4	53.6 ± 1.6	40.0 ± 1.2
+ SELFIE [32]	54.5 ± 1.7	39.2 ± 1.1	28.8 ± 2.9	51.8 ± 4.1	40.4 ± 1.4
+ Co-teaching [14]	56.1 ± 2.3	39.8 ± 4.4	30.5 ± 3.6	53.5 ± 2.9	38.7 ± 1.3
+ DivideMix [22]	56.1 ± 1.3	43.7 ± 0.3	35.1 ± 1.9	56.1 ± 0.6	38.9 ± 2.5
GBS [27]	54.8 ± 1.2	41.8 ± 0.9	27.3 ± 2.1	54.2 ± 1.7	40.4 ± 1.1
+ SELFIE [32]	55.4 ± 0.8	41.6 ± 0.3	27.8 ± 1.8	51.3 ± 4.7	40.7 ± 0.9
+ Co-teaching [14]	55.1 ± 0.9	42.7 ± 0.9	31.4 ± 3.6	54.0 ± 0.9	39.5 ± 1.3
+ DivideMix [22]	57.8 ± 1.9	48.8 ± 1.9	34.3 ± 1.3	57.4 ± 0.6	44.6 ± 5.5
RM [5]	57.1 ± 0.1	46.5 ± 2.6	33.5 ± 3.0	58.3 ± 2.6	46.2 ± 1.9
+ SELFIE [32]	56.8 ± 1.3	44.8 ± 0.6	31.8 ± 4.4	57.9 ± 1.5	46.9 ± 1.9
+ Co-teaching [14]	57.5 ± 1.8	47.6 ± 0.7	35.1 ± 2.0	58.5 ± 1.5	45.9 ± 2.0
+ DivideMix [22]	61.3 ± 0.8	50.9 ± 3.3	34.9 ± 3.1	60.6 ± 1.7	46.4 ± 5.1
PuriDivER (ours)	**61.3** ± 2.1	**59.2** ± 0.3	**52.4** ± 2.0	**61.6** ± 1.6	**47.1** ± 3.2



Ablation Studies

Robust Learning		CIFAR-10 (K=500)				
		SYM			ASYM	
Re-label	Consistency	20	40	60	20	40
		61.8	55.4	46.9	60.6	46.4
✓		57.7	55.4	44.1	61.0	46.6
	✓	48.6	46.2	31.3	52.1	36.7
✓	✓	61.3	**59.2**	**52.4**	**61.6**	**47.1**



Summary

- Propose the first online blurry CL set-up with noisy data
- Propose memory update/usage methods considering diversity and purity
- Propose adaptive balancing coefficient to cover various environments

Code