NAVER CLOVA[1]
NAVER AI LAB[2]
Search Solution Inc.[3] GIST[4]

Rainbow Memory: Continual Learning with a Memory of Diverse Samples

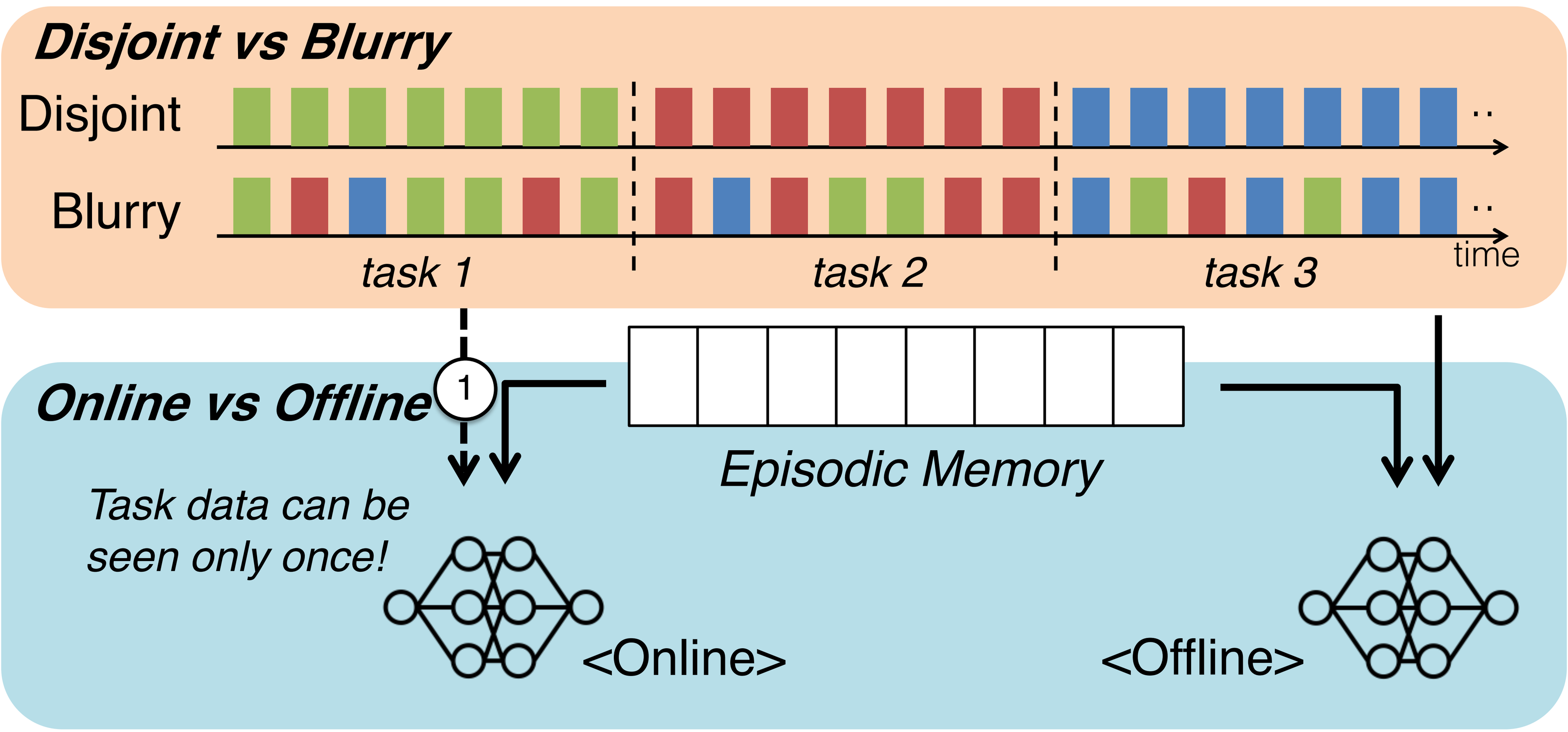
Jihwan Bang[1,2,3] Heesu Kim[1,2] YoungJoon Yoo[1,2] Jung-Woo Ha[1,2] Jonghyun Choi[4]

CVPR VIRTUAL JUNE 19-25

Problem Set-up

tl;dr. A better method in a realistic continual learning set-up

A Realistic Continual Learning (CL) Set-up



Previous Work

Method	CIL Settings				Knowledge Preserv.	
	Online	Disjoint	Blurry	Memory	Regularize	Distill
EWC[a]	X	O	X	X	O	X
iCaRL[b]	X	O	X	O	X	O
Rwalk[c]	X	O	X	O	O	X
GSS[d]	O	X	O	O	X	X
BiC[e]	X	O	X	O	X	O
Gdumb[f]	O	O	O	O	X	X
RM (ours)	O	O	O	O	X	X

→ CL understudied on blurry-online setting

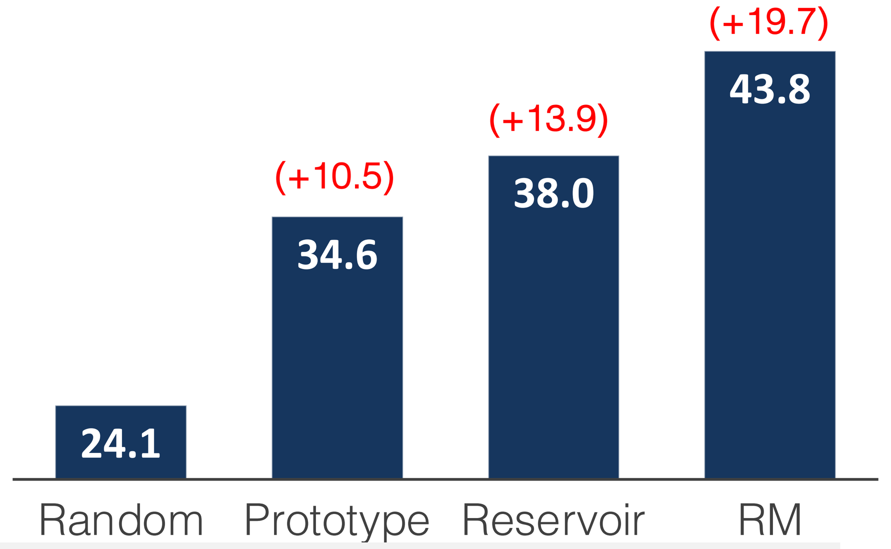
References

[a] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017
[b] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, et al. iCaRL: Incremental classifier and representation learning. In *CVPR*, 2017.
[c] Arslan Chaudhry, Puneet K. Dokania, et al. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In *ECCV*, 2018.
[d] Rahaf Aljundi, Min Lin, et al. Gradient based sample selection for online continual learning. In *NeurIPS*, pages 11816–11825, 2019
[e] Yue Wu, Yan-Jia Chen, et al. Large scale incremental learning. In *CVPR*, 2019.
[f] Ameya Prabhu, P. Torr, et al. GDumb: A simple approach that questions our progress in continual learning. In *ECCV*, 2020.

Method

Increase visual diversity in memory

1. Calculate data uncertainty
2. For each class, select data from the lowest uncertainty to the highest with uniform interval
3. Train the model with data augmentation



Computing Uncertainty

Uncertainty U = Data Difficulty

$$p(y=c\mid x)=\int p(y=c\mid \tilde{x}_t)\,p(\tilde{x}_t\mid x)\,d\tilde{x}_t$$

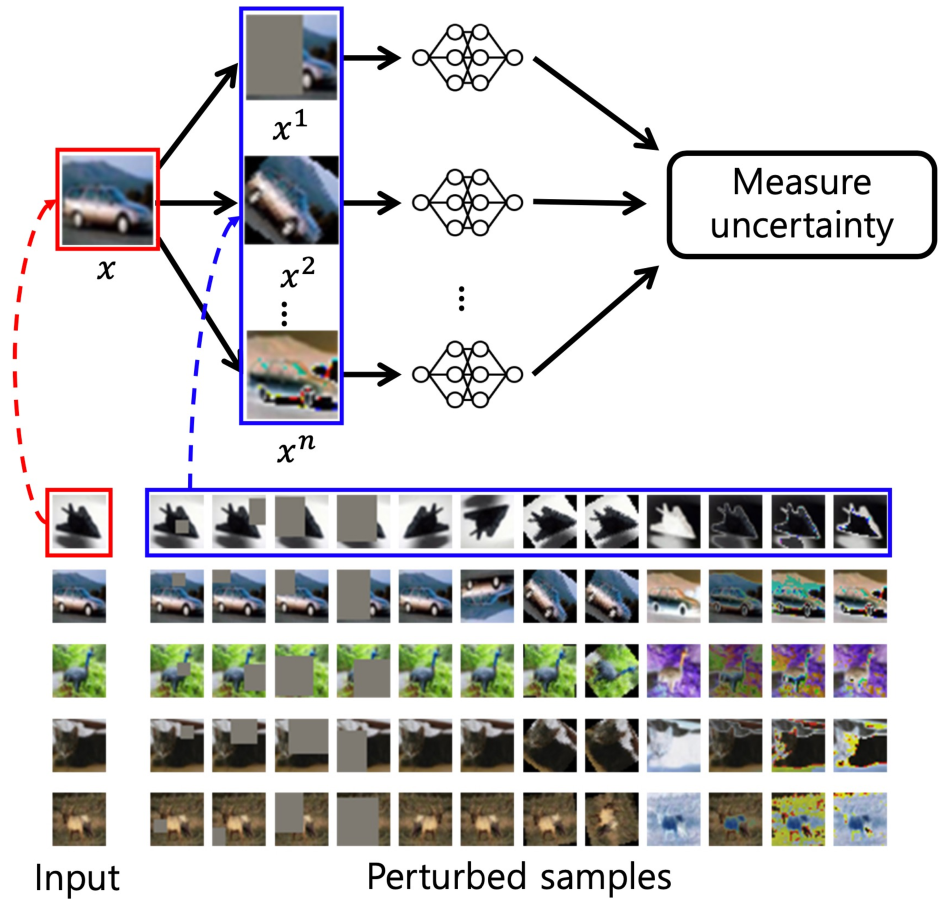
$$\approx \frac{1}{A}\sum_{t=1}^{A} p(y=c\mid \tilde{x}_t)$$

$$\rightarrow U(x)=1-\frac{1}{T}\max_{c} S_c$$

Certain data (uncert. ↓) is robust for perturbed images

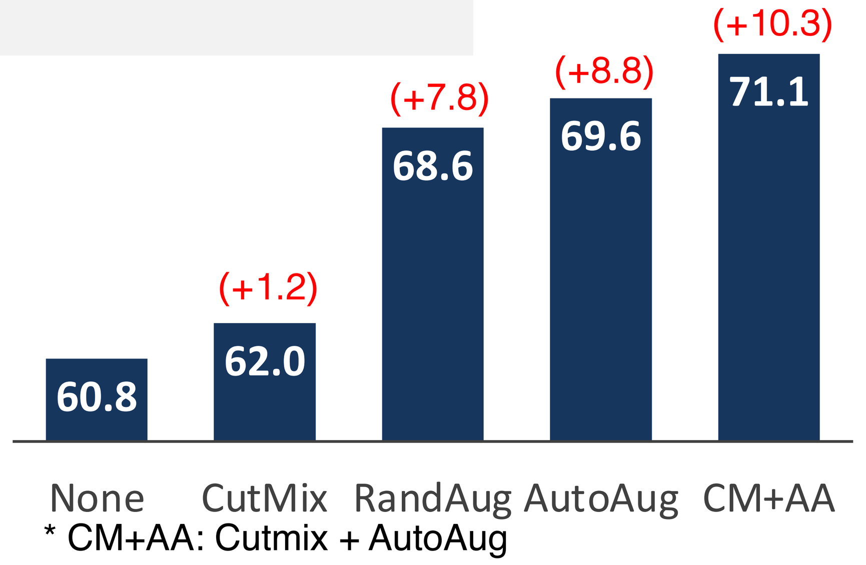
$$S_c=\sum_{t=1}^{T}\mathbb{1}_c \arg\max_{\hat{c}} p(y=\hat{c}\mid x_t)$$

T: Number of perturbed samples



Ablations

- Memory Selection
  +19.7% accuracy improvement by RM compared to Random.
- Data Augmentation (DA)
  +10.3% accuracy improvement by CM+AA compared to not using DA

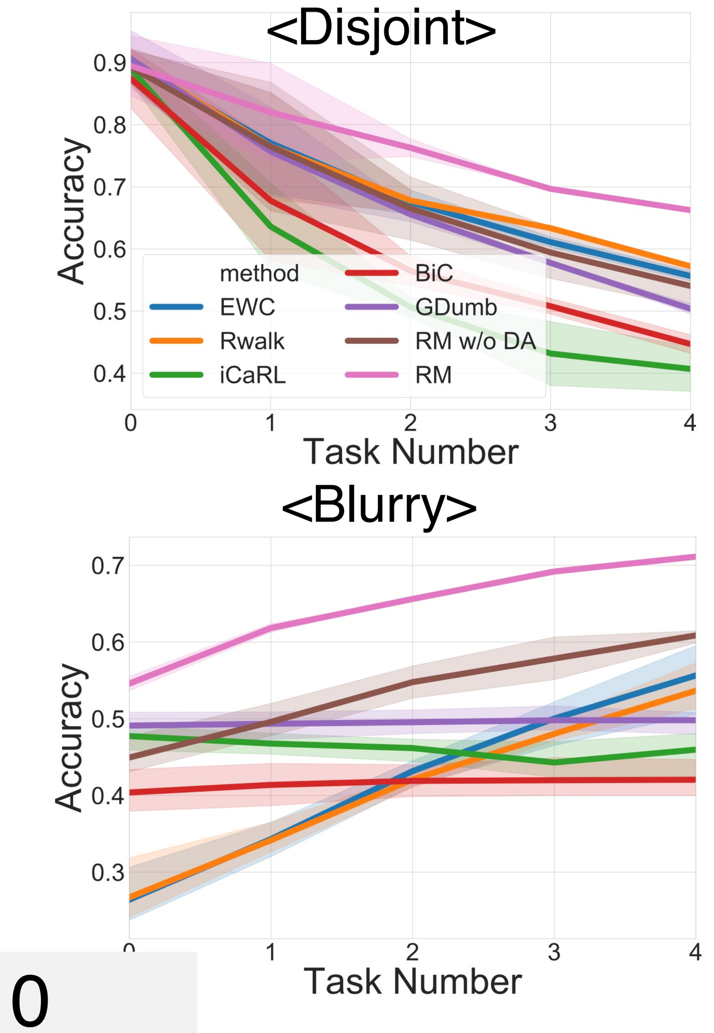


* CM+AA: Cutmix + AutoAug

Experiments

Last Accuracy After Training All Tasks

Methods	MNIST	CIFAR100	ImageNet
EWC	90.98 ± 0.61	26.95 ± 0.36	39.54
Rwalk	90.69 ± 0.62	32.31 ± 0.78	35.26
iCaRL	78.09 ± 0.60	17.39 ± 1.04	17.52
GDumb	88.51 ± 0.52	27.19 ± 0.65	21.52
BiC	77.75 ± 1.27	13.01 ± 0.24	37.20*
RM w/o DA	**92.65 ± 0.33**	34.09 ± 1.41	37.96
RM	91.80 ± 0.69	**41.35 ± 0.95**	**50.11**



Last Accuracy on Various CL Settings on CIFAR10

Methods	Blurry0 (=Disjoint)		Blurry10		Blurry30	
	Online	Offline	Online	Offline	Online	Offline
EWC	55.66 ± 1.18	64.00 ± 1.34	55.65 ± 4.60	**78.67 ± 1.06**	60.57 ± 1.15	85.00 ±0.42
Rwalk	55.91 ± 1.85	65.04 ± 0.11	53.66 ± 3.18	78.59 ± 1.37	59.03 ± 0.05	**85.18 ± 0.57**
iCaRL	40.70 ± 5.13	**65.61 ± 2.57**	45.98 ± 3.04	57.07 ± 2.74	48.11 ± 4.63	64.90 ± 7.95
GDumb	50.37 ± 1.17	50.37 ± 1.17	46.70 ± 1.53	46.70 ± 1.53	47.78 ± 3.77	47.78 ± 3.77
BiC	44.70 ± 2.12	59.53 ± 4.30	42.06 ± 2.41	61.45 ± 6.25	42.92 ± 1.47	71.93 ± 2.45
RM w/o DA	54.05 ± 4.94	59.47 ± 0.61	60.87 ± 0.88	74.58 ± 0.60	60.92 ± 6.48	83.91 ± 0.40
RM	**66.25 ± 0.21**	61.91 ± 0.63	**71.13 ± 0.18**	76.86 ± 0.04	**73.90 ± 0.80**	85.10 ± 0.16

Summary

- Tackle blurry setting on continual learning, and analyze various setups
- Propose a memory management scheme using data uncertainty
- Show DA improves significant performance in online blurry CL

Code, Data, Set-ups →