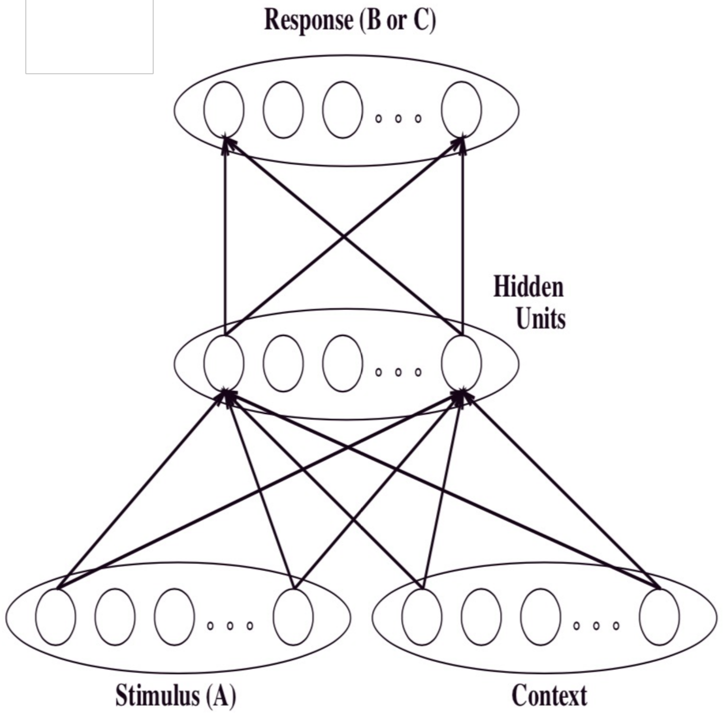
Brain-Inspired Continual Learning

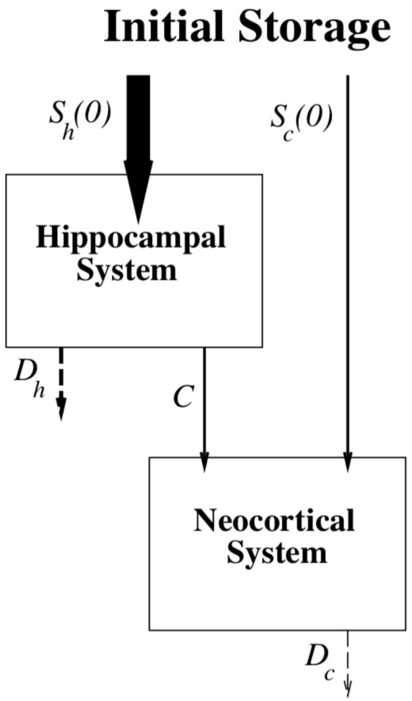
Liyuan Wang

School of Life Sciences
Department of Computer Science and Technology
Tsinghua University

Continual / Incremental / Lifelong Learning

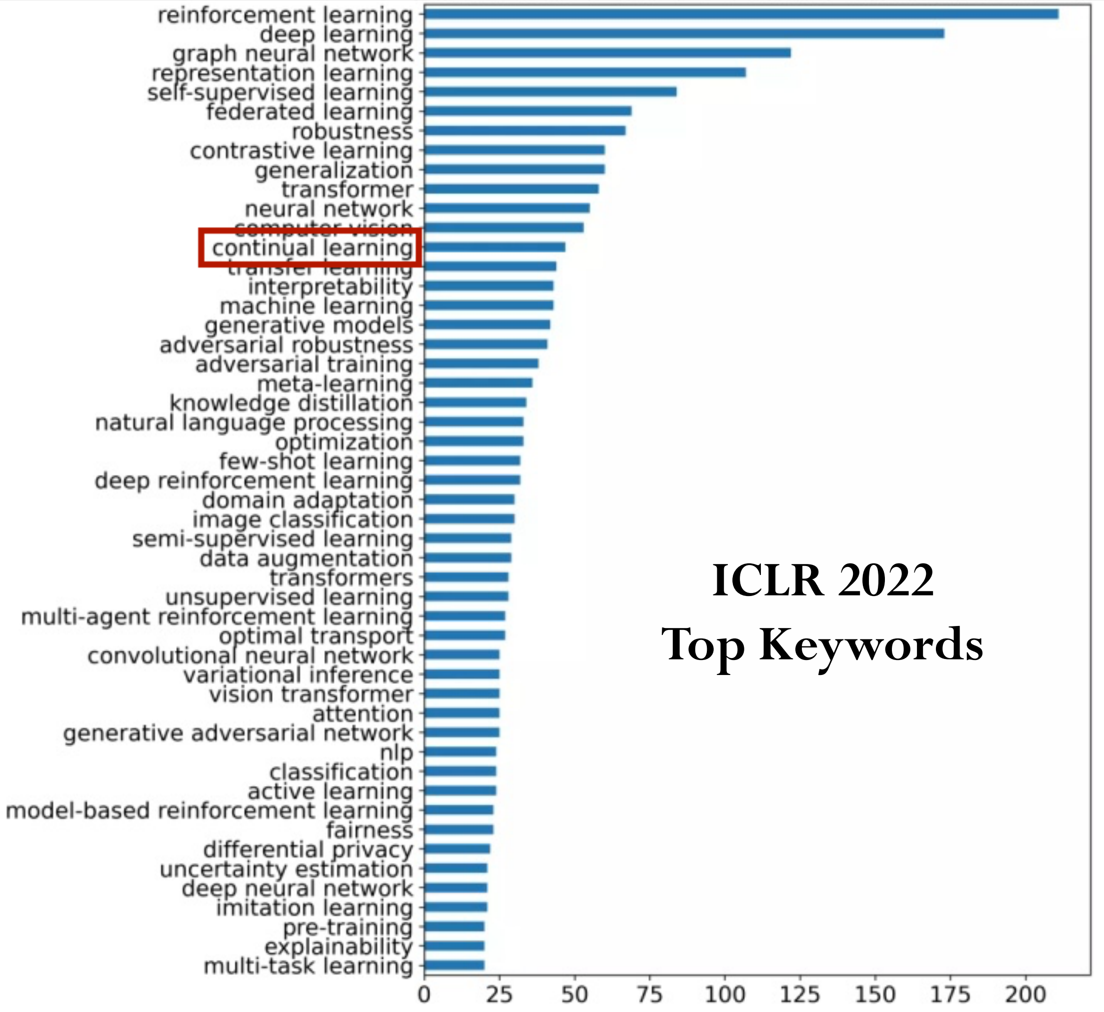
- New Task / Class, New Instance, New Domain
- Catastrophic Forgetting
- Stability-Plasticity Trade-off





McCloskey et al., 1989; McClelland et al., 1995

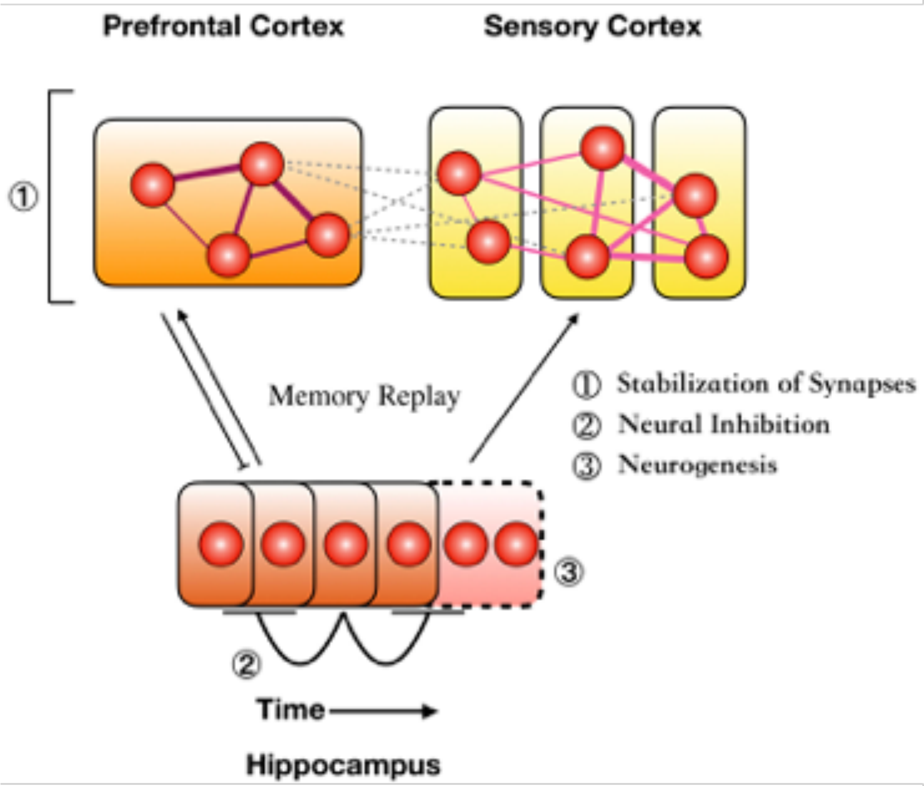
Continual Learning is Getting Hotter and Hotter



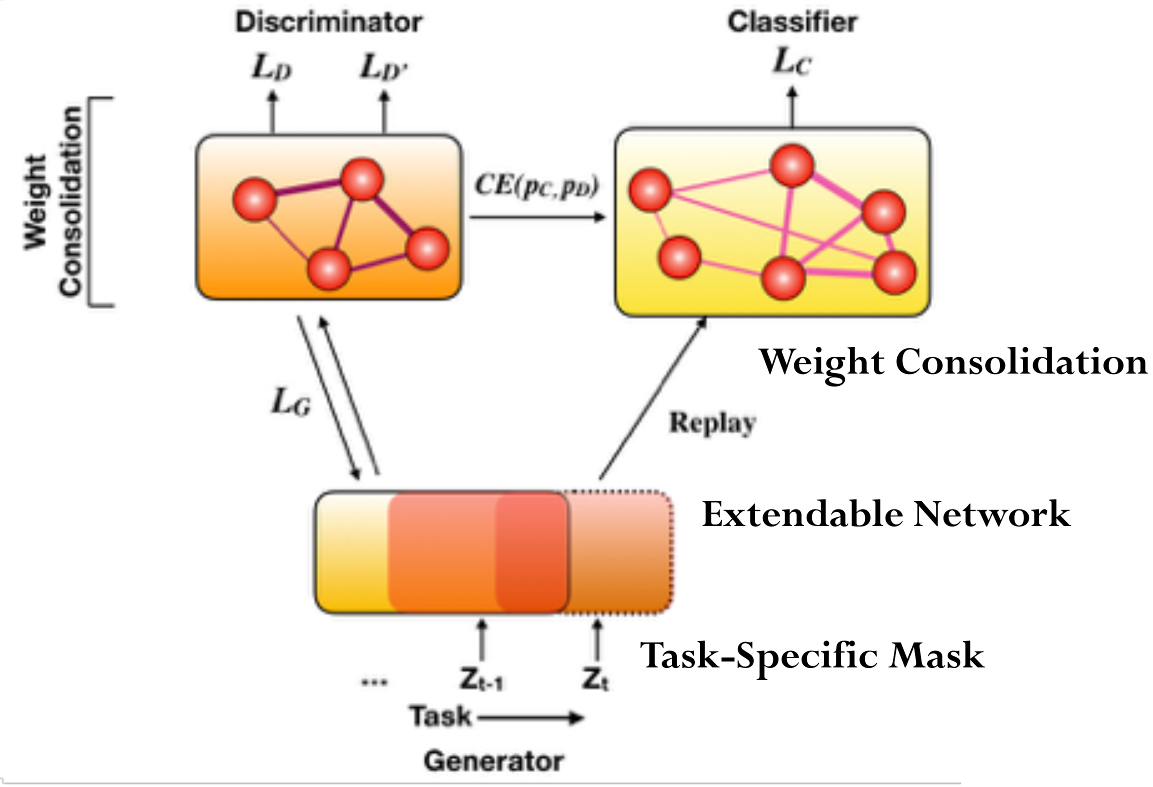
(Brain-Inspired) Continual Learning Approaches

- **Regularization-Based Methods**
  - Selectively Penalize Parameter Changes, Fast-Slow Weights
  - Synaptic Consolidation, Synaptic Plasticity
- **Replay-Based Methods**
  - Old / Generated Data, Old / Generated Feature
  - Biological Memory Replay, Complementary Learning System
- **Architecture-Based Methods**
  - Parameter Isolation, Sub-modules / Sub-networks
  - Modularization, Neural Inhibition, Engram Ensemble

Triple Memory Networks: A Brain-Inspired Framework



The Brain Memory System



Triple Memory Networks

Liyuan Wang, Bo Lei, Qian Li, Hang Su, Jun Zhu, Yi Zhong. TNNLS, 2021.

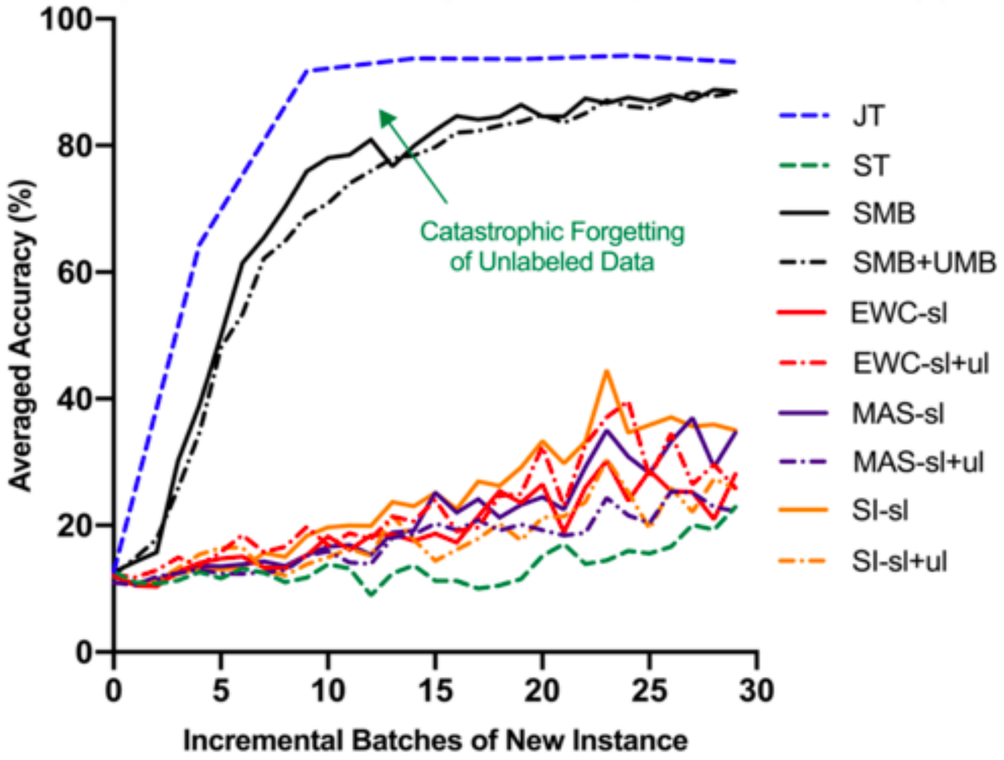
Experimental Results

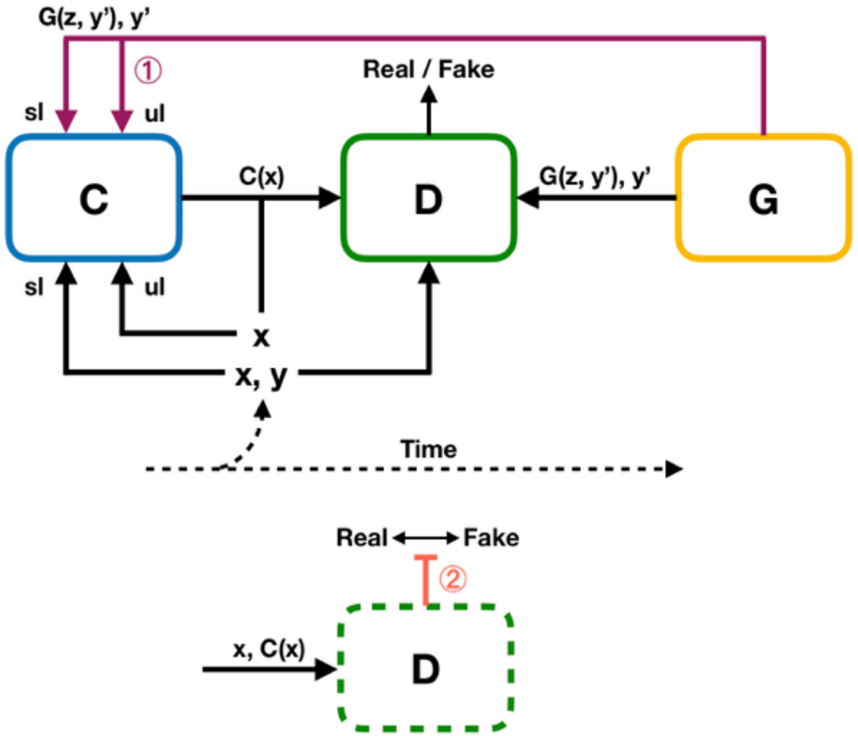
- Without accessing to the old data, Triple Memory Networks (TMNs) achieve the state-of-the-art performance in supervised class-incremental learning.

	Methods	MNIST A_5	MNIST A_{10}	SVHN A_5	SVHN A_{10}	CIFAR-10 A_5	CIFAR-10 A_{10}	ImageNet-50 A_{30}	ImageNet-50 A_{50}
	Joint Training	99.87	99.24	92.99	88.72	83.40	77.82	57.35	49.88
+ Training Data	EWC-S [13]	79.36	60.83	38.65	25.36	37.39	21.13	-	-
	SI-S [14]	78.40	60.18	37.21	23.86	36.96	20.16	-	-
	RWalk-S [25]	82.08	62.84	39.25	26.63	35.75	22.27	-	-
	MAS-S [15]	80.40	67.66	37.57	25.11	44.38	19.56	-	-
	iCarl [18]	-	-	-	-	57.30	43.69	29.38	**28.98**
	DGMw-S [22]	-	-	-	-	-	-	36.87	18.84
- Training Data	EWC-M [41]	70.62	77.03	39.84	33.02	-	-	-	-
	DGR [3]	90.39	85.40	61.29	47.28	-	-	-	-
	MeRGAN [21]	98.19	**97.00**	80.90	66.78	-	-	-	-
	DGMw [22]	98.75	96.46	83.93	74.38	72.45	56.21	32.14	17.82
	TMNs (ours)	**98.80**	96.72	**87.12**	**77.08**	**72.72**	**61.24**	**38.23**	28.08

ORDisCo: Semi-supervised Continual Learning

- The incremental data are typically partially-labeled in realistic scenarios.
- Representative methods lack the ability to exploit the incremental unlabeled data.

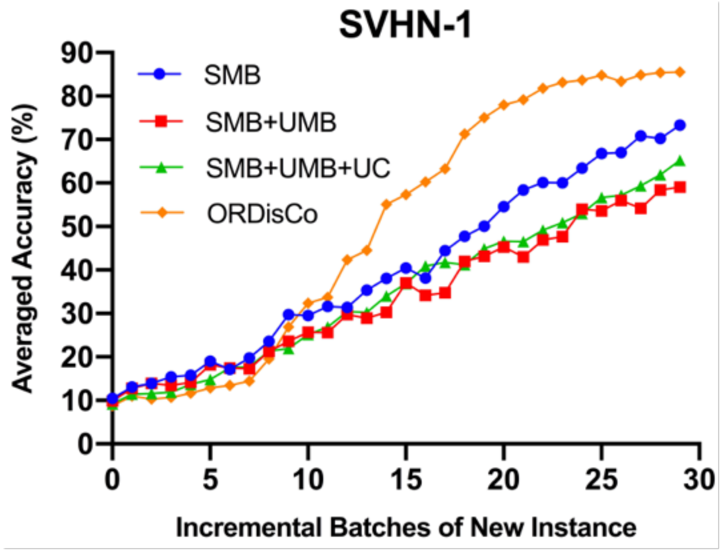


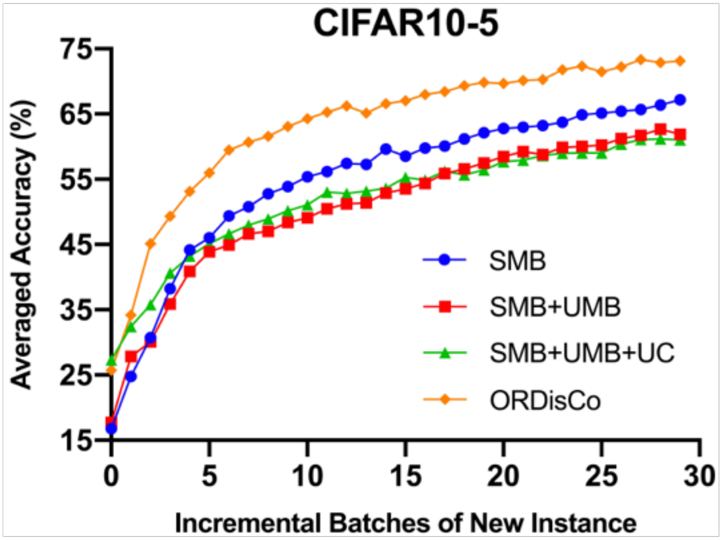


Liyuan Wang, Kuo Yang, Chongxuan Li, Lanqing Hong, Zhenguo Li, Jun Zhu. CVPR 2021.

Experimental Results

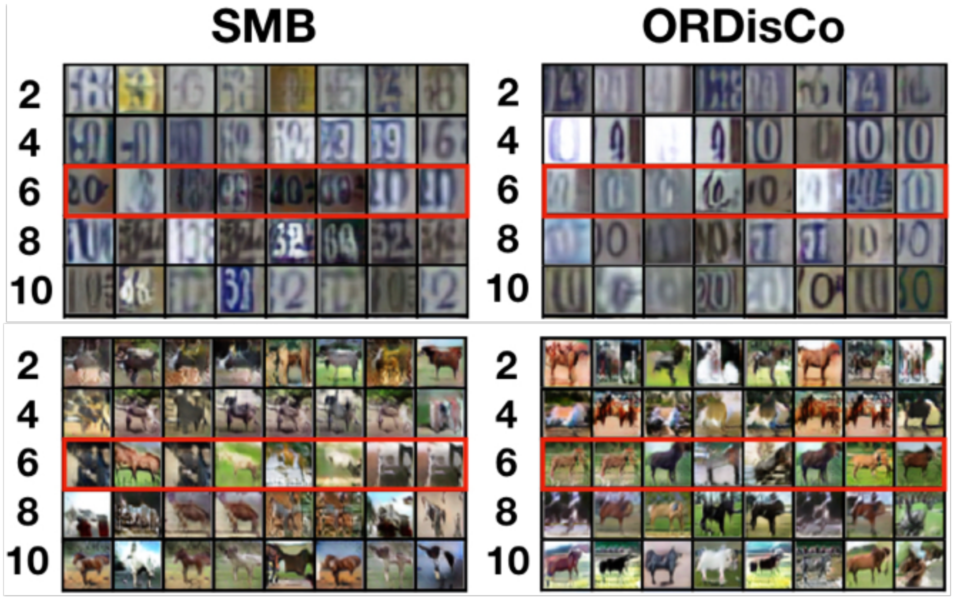
Classification



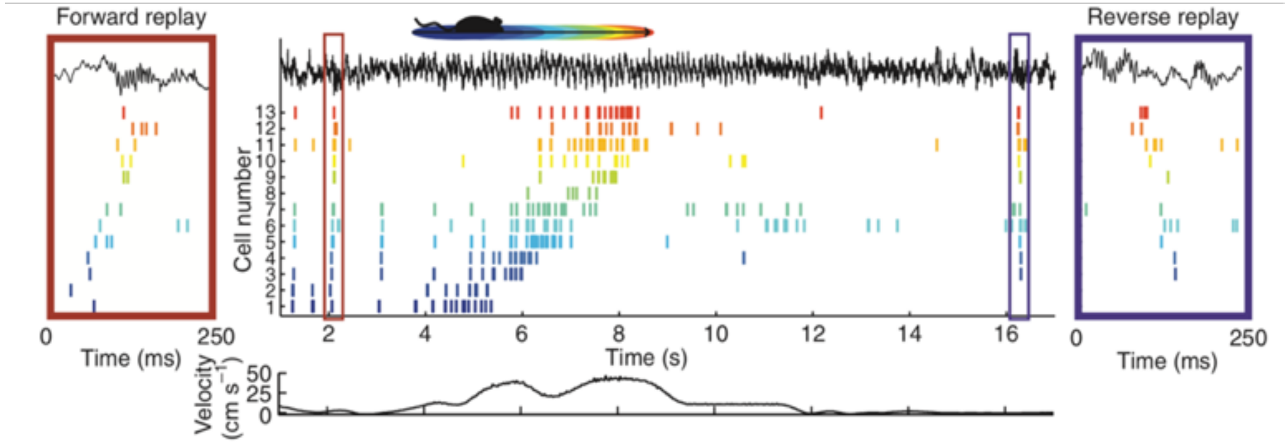


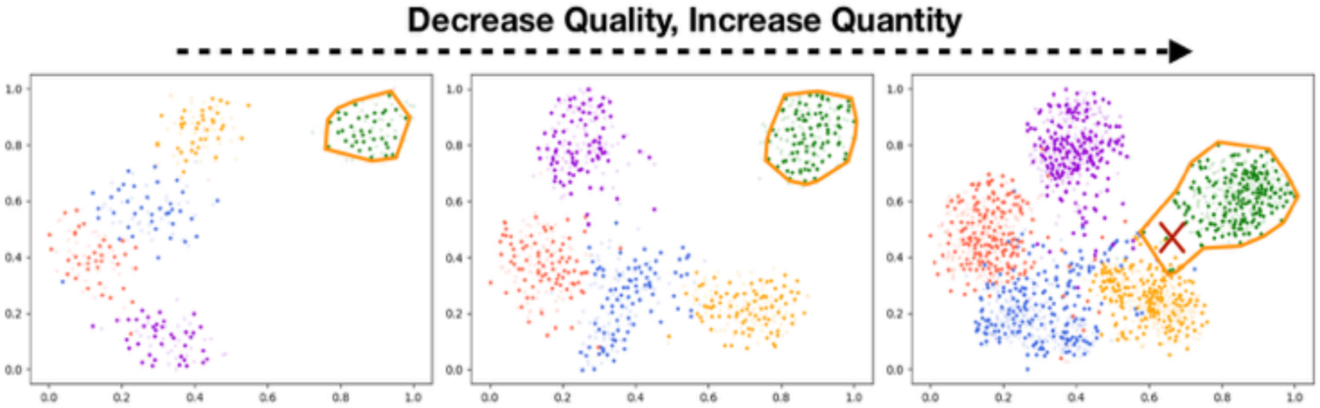
SMB: Replay of Supervised Memory Buffer
UMB: Replay of Unsupervised Memory Buffer

Conditional Generation



Memory Replay with Compression





$$\mathcal{P}_q(M_q^c|D) = \frac{\det(L_{M_q^c}(D;q,\theta))}{\sum_{|M|=N_q^{mb}} \det(L_M(D;q,\theta))},$$

(1) Maximize $\mathcal{P}_q(M_q^*|D)$ $\quad$ $\mathcal{P}_q(M_q^c|D) \leq \mathcal{P}_q(M_q^*|D)$

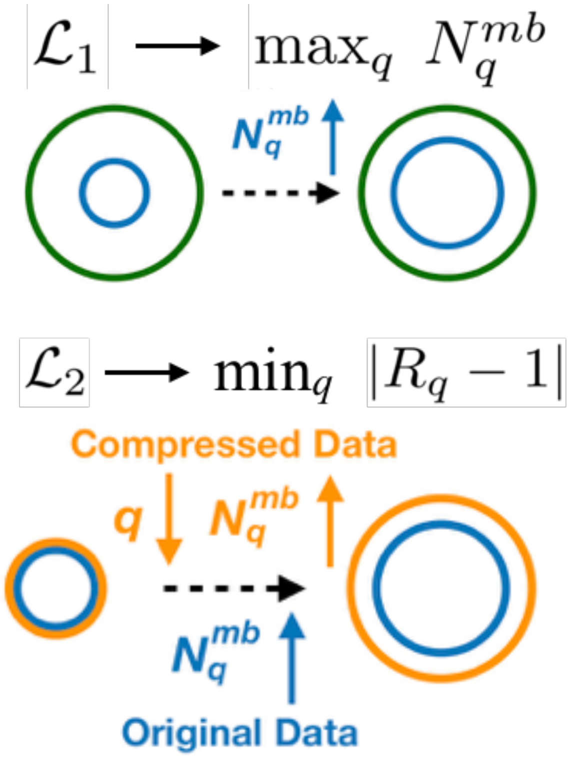
(2) Constrain that $\mathcal{P}_q(M_q^c|D)$ **is consistent with** $\mathcal{P}_q(M_q^*|D)$

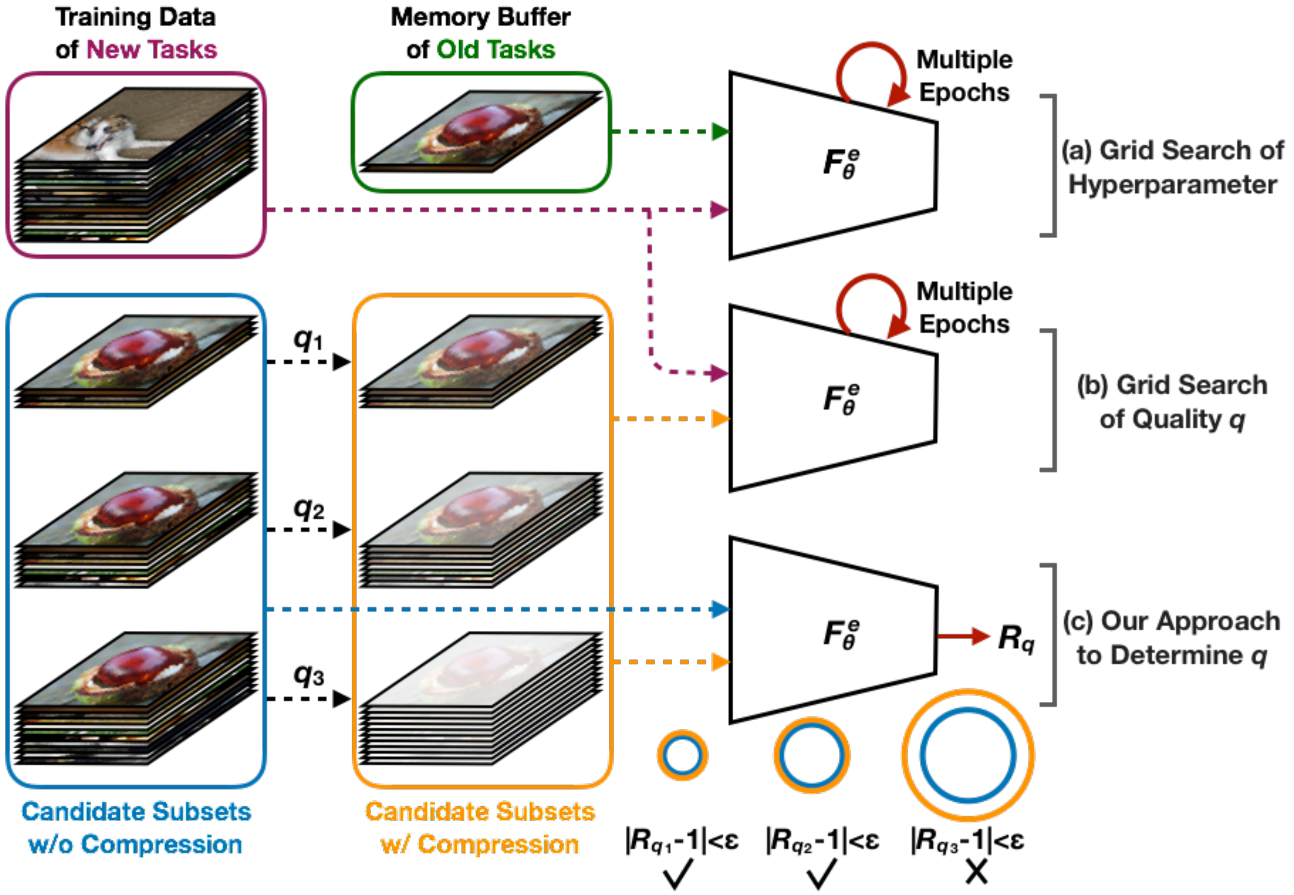
$$\mathcal{L}_2(q) = \left|\frac{\mathcal{P}_q(M_q^c|D)}{\mathcal{P}_q(M_q^*|D)} - 1\right| = \left|\frac{\det(M_q^{c\top} M_q^c)}{\det(M_q^{*\top} M_q^*)} Z_q - 1\right|$$
$$= \left|\left(\frac{\mathrm{Vol}_q^c}{\mathrm{Vol}_q^*}\right)^2 Z_q - 1\right| = |R_q^2 Z_q - 1|$$

Liyuan Wang, Xingxing Zhang, Kuo Yang, Longhui Yu, Chongxuan Li, Lanqing Hong, Shifeng Zhang, Zhenguo Li, Yi Zhong, Jun Zhu. ICLR 2022.

Memory Replay with Compression

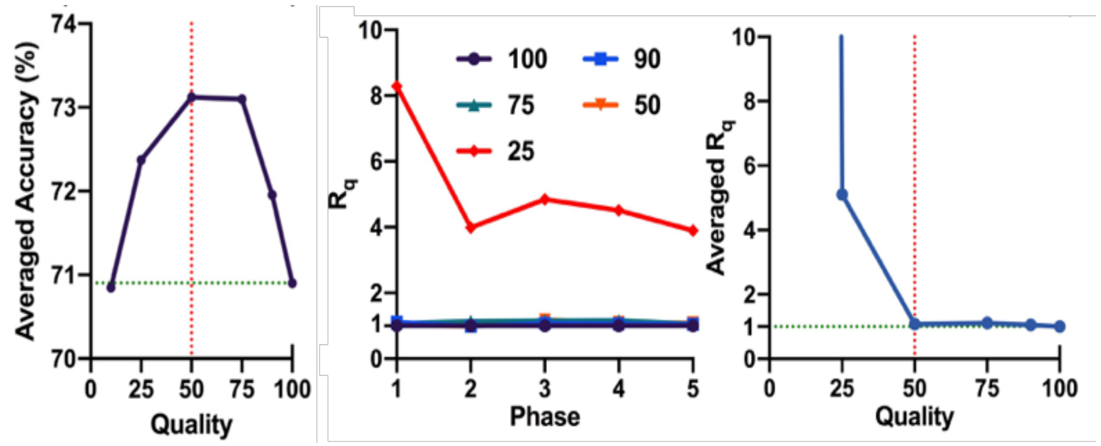
- Given a limited storage space, our method can efficiently determine a proper compression quality for incoming data, without repetitive training.

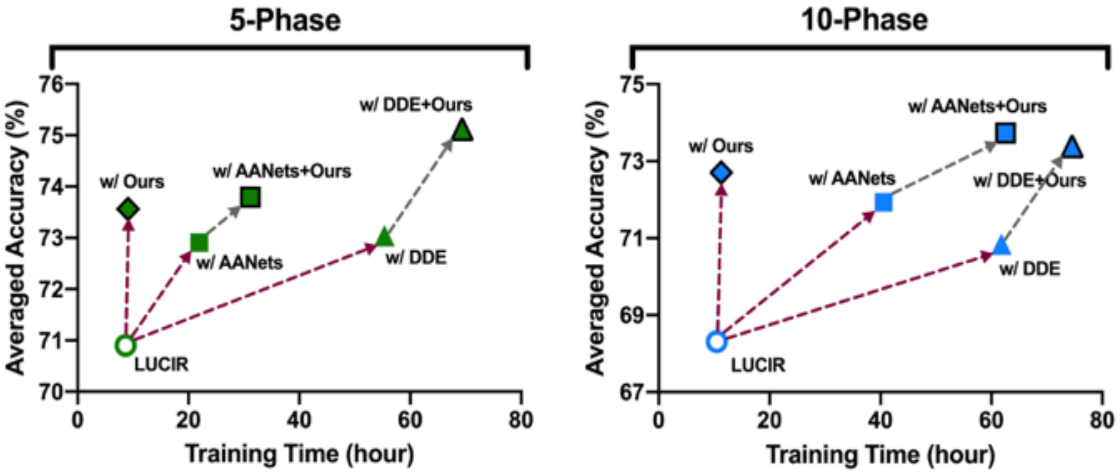




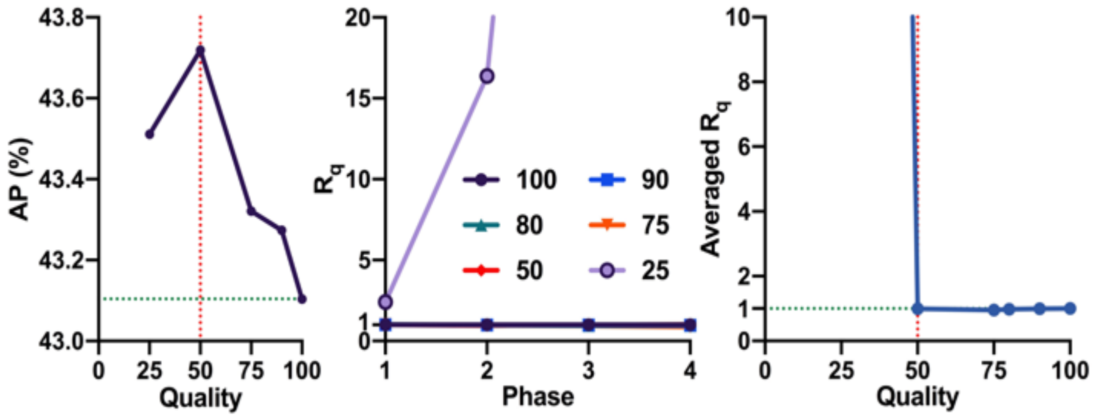
Experimental Results

Large-Scale Class-Incremental Learning





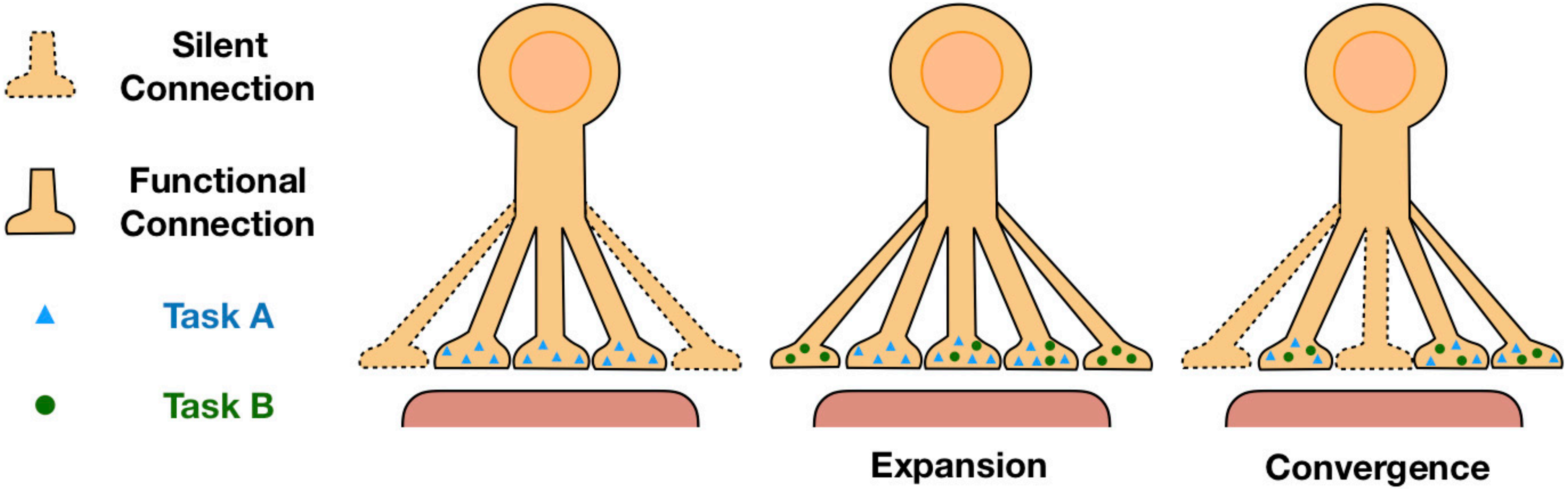
Object Detection for Autonomous Driving



	Method	AP	AP_{50}	AP_{75}
Pseudo Labeling	FT	40.36	63.83	43.82
	MR	40.75 / +0.39	65.11 / +1.28	43.53 / -0.29
	Ours	**41.50 / +1.14**	**65.36 / +1.53**	**44.95 /+1.13**
Unbiased Teacher	FT	42.88	66.70	45.99
	MR	43.10 / +0.22	66.88 / +0.18	46.62 / +0.63
	Ours	**43.72 / +0.84**	**67.80 / +1.10**	**47.36 / +1.37**

AFEC: Active Forgetting of Negative Transfer

- If the old knowledge conflicts with the new task learning, then precisely remembering the old knowledge will further aggravate the interference.
- Biological neural networks can **actively forget** the conflicting information, through regulating the learning-triggered synaptic expansion and synaptic convergence.



Liyuan Wang, Mingtian Zhang, Zhongfan Jia, Qian Li, Chenglong Bao, Kaisheng Ma, Jun Zhu, Yi Zhong. NeurIPS, 2021

AFEC: Active Forgetting of Negative Transfer

- We introduce a forgetting factor β and replace the posterior that absorbs all the information of the old tasks by a weighted product distribution:

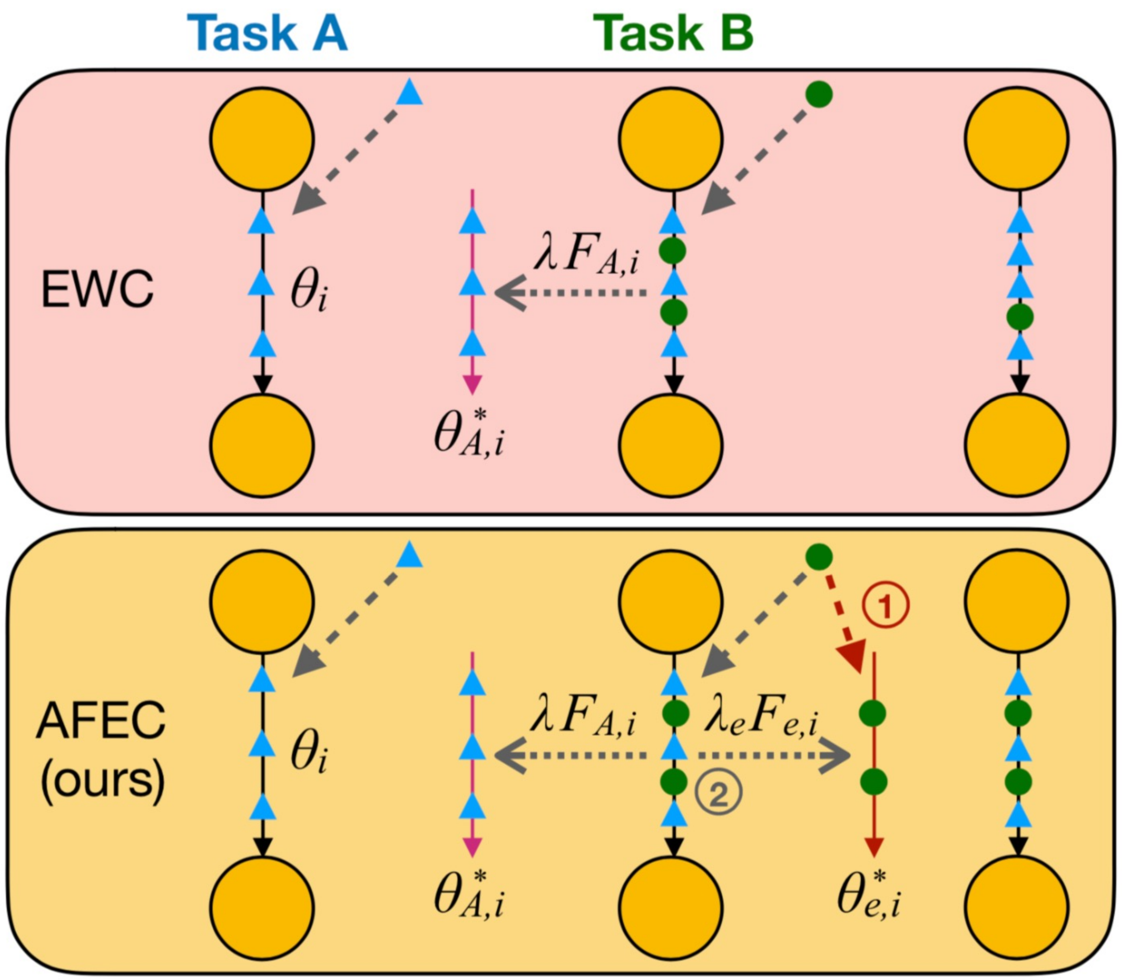
$$p(\theta|D_A^{train}) = \frac{p(D_A^{train}|\theta)p(\theta)}{p(D_A^{train})}$$

Replace

$$p_m(\theta|D_A^{train}, \beta) = \frac{p(\theta|D_A^{train})^{(1-\beta)}p(\theta)^{\beta}}{Z}$$

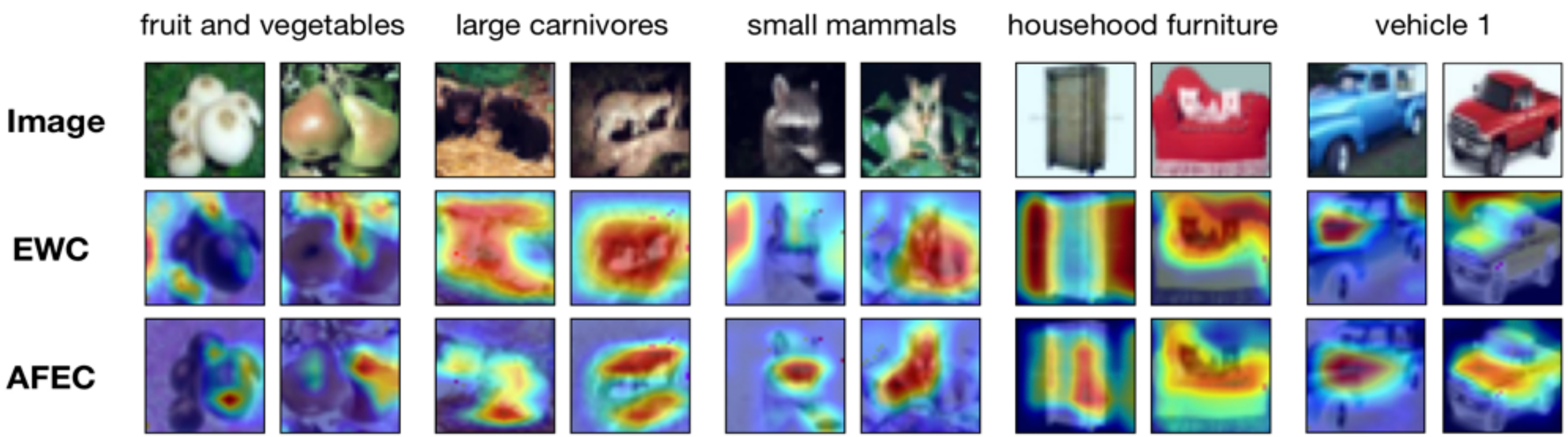
- The optimal forgetting factor can maximize the learning of each new task:

$$\beta^* = \arg\max_{\beta} p(D_B^{train}|D_A^{train}, \beta)$$
$$= \arg\max_{\beta} \int p(D_B^{train}|\theta)p_m(\theta|D_A^{train}, \beta)d\theta$$

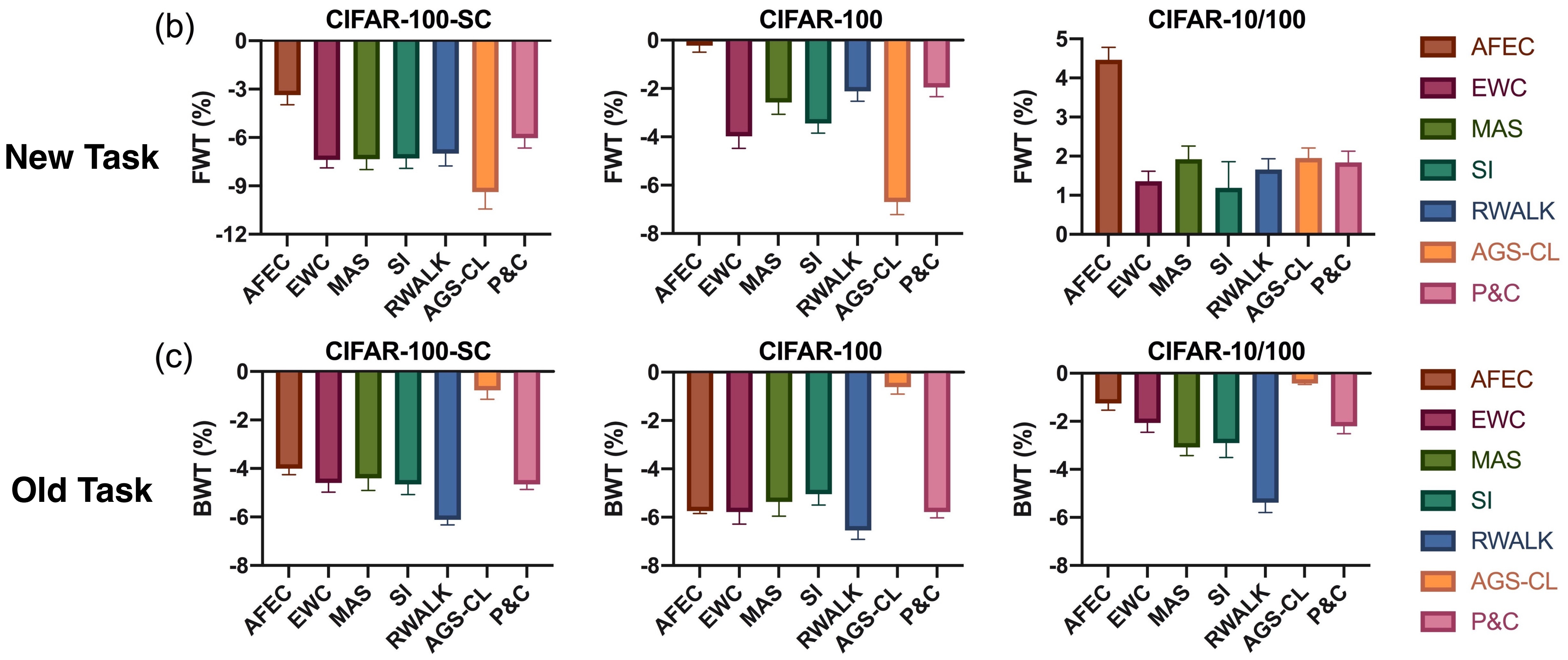


Experimental Results

Methods	CIFAR100-SC A_{10}	CIFAR100-SC A_{20}	CIFAR100 A_{10}	CIFAR100 A_{20}	CIFAR10/100 A_{2}	CIFAR10/100 A_{2+20}	CUB-200 w/ PT A_{5}	CUB-200 w/ PT A_{10}	CUB-200 w/o PT A_{5}	CUB-200 w/o PT A_{10}	ImageNet-100 A_{5}	ImageNet-100 A_{10}
EWC [13]	52.25	51.74	68.72	69.18	85.07	77.75	81.37	80.92	32.90	42.29	76.12	73.82
* AFEC (ours)	**56.28**	**55.24**	**72.36**	**72.29**	**86.87**	**81.25**	**83.65**	**82.04**	**34.36**	43.05	**77.64**	75.46
MAS [1]	52.76	52.18	67.60	69.41	84.97	77.39	79.98	79.67	31.68	42.56	75.48	74.72
w/ AFEC (ours)	55.26	54.89	69.57	71.20	86.21	80.01	82.77	81.31	34.08	42.93	75.64	**75.66**
SI [36]	52.20	51.97	68.72	69.21	85.00	76.69	80.14	80.21	33.08	42.03	73.52	72.97
w/ AFEC (ours)	55.25	53.90	69.34	70.13	85.71	78.49	83.06	81.88	34.04	**43.20**	75.72	74.14
RWALK [2]	50.51	49.62	66.02	66.90	85.59	73.64	80.81	80.58	32.56	41.94	73.24	73.22
w/ AFEC (ours)	52.62	51.76	68.50	69.12	86.12	77.16	83.24	81.95	33.35	42.95	74.64	73.86



Experimental Results



Summary

- Continual learning is complex, but all roads lead to Rome;
- Successful biological strategies can provide inspirations for and evolve with computational models;
- Order is the appearance, compatibility is the goal;
- Look to the stars (general theoretical insights) and keep feet on the ground (realistic challenges).

Acknowledgement

- School of Life Sciences in Tsinghua:

  Prof. Yi Zhong, Prof. Qian Li

- Dept. of Comp. Sci. & Tech. in Tsinghua:

  Prof. Jun Zhu, Dr. Chongxuan Li (now at Renmin U), Dr. Xingxing Zhang

- Brain-inspire AI Project:

  Prof. Kaisheng Ma, Prof. Chenglong Bao, Zhongfan Jia

- Huawei Noah's Ark Lab:

  Dr. Lanqing Hong, Dr. Zhenguo Li, Kuo Yang, Mingtian Zhang, Longhui Yu

Thank You!