GACL: Exemplar-Free Generalized Analytic Continual Learning

Huiping Zhuang[1*] , Yizhu Chen[1*] , Di Fang[1], Run He[1], Kai Tong[1], Hongxin Wei[2], Ziqian Zeng[1†], Cen Chen[1,3,5†]

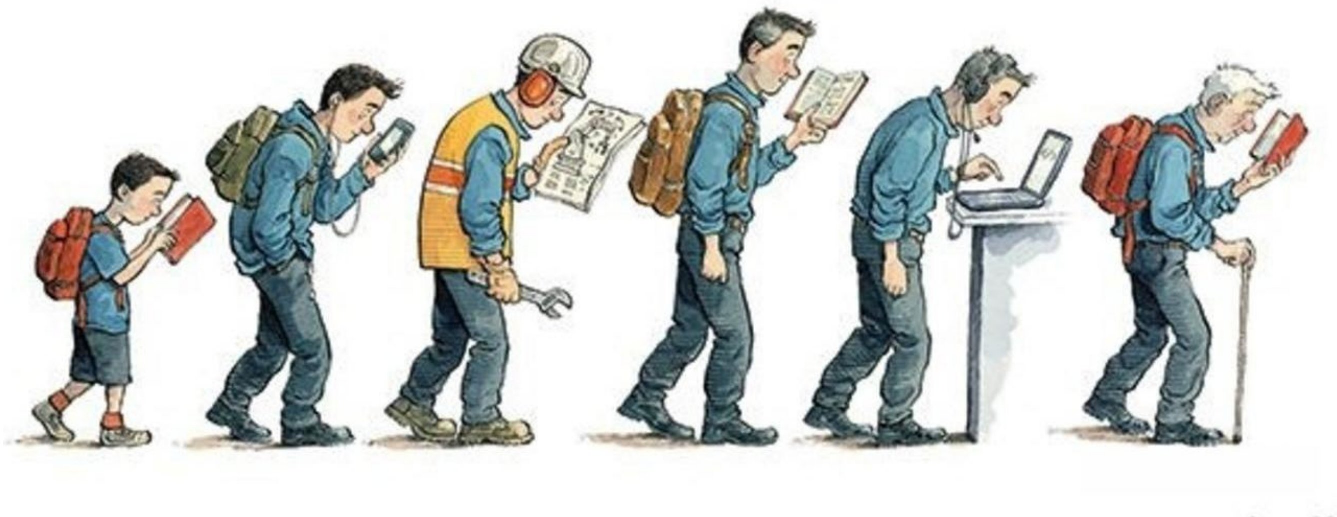
[1]South China University of Technology, China
[2]Southern University of Science and Technology, China
[3]Shenzhen Institute, Hunan University, China
[4]Pazhou Lab, China
* Equal contribution
† Corresponding authors

The 38th Conference on Neural Information Processing Systems (NeurIPS 2024)

NEURAL INFORMATION PROCESSING SYSTEMS

Continual learning, Incremental learning, Lifelong learning

Enables **continuous knowledge acquisition**, mimicking human behavior.

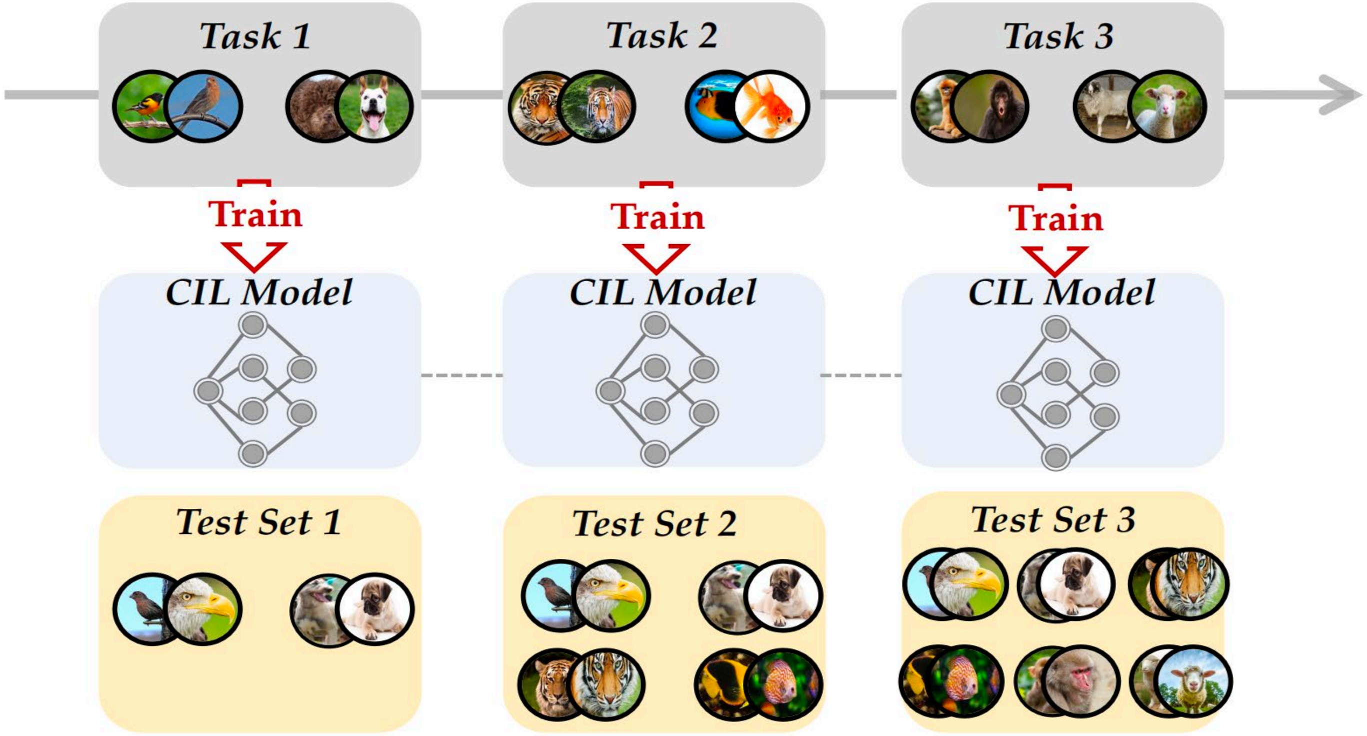


Practical Significance:

- No need for retraining
- Adapt used models to new tasks

Traditional Class Incremental Learning

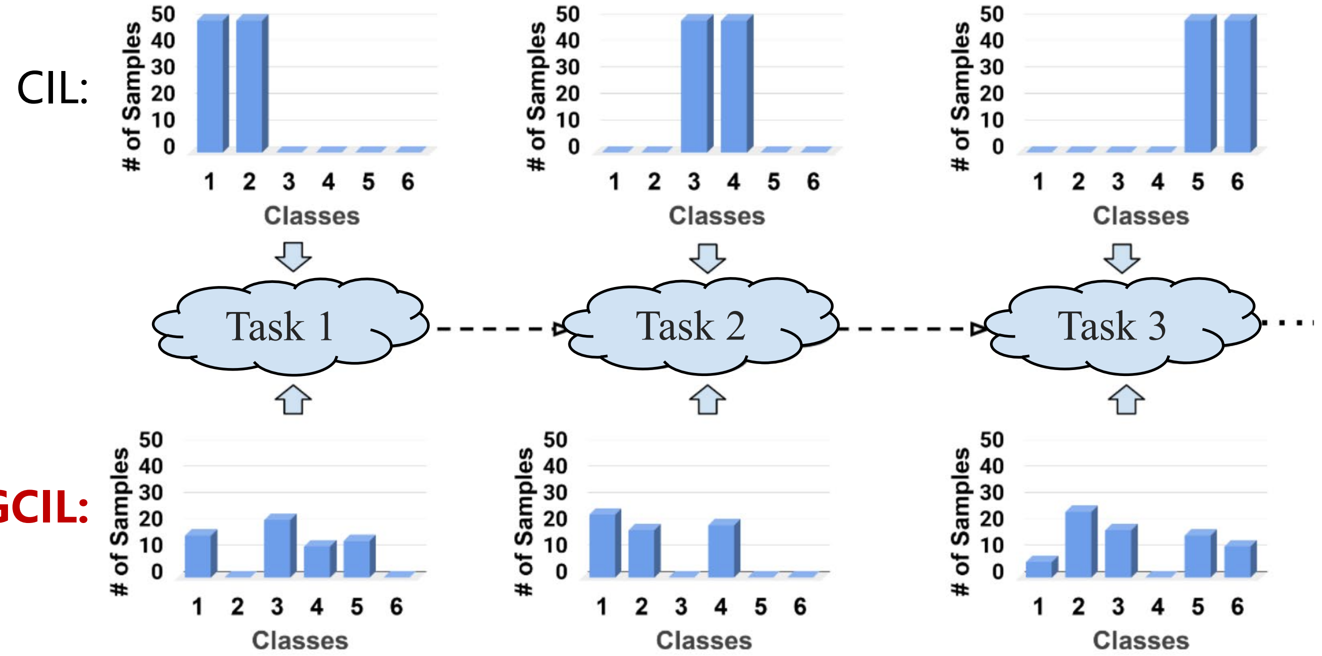
Class Incremental Learning (CIL) aims to continually build a holistic classifier among **all seen classes** with **non-overlapping classes** arriving sequentially.

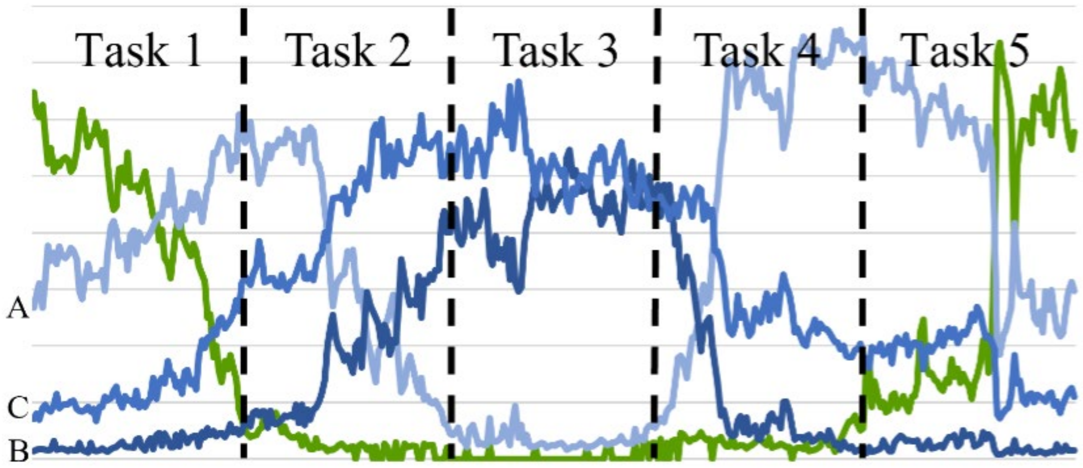


Zhou, Da-Wei, et al. "Class-incremental learning: A survey." IEEE Transactions on Pattern Analysis and Machine Intelligence (2024).

Generalized Continual Learning

Generalized Class Incremental Learning (GCIL) simulates **real-world** incremental learning, as **distributions of data category and sample size could be uneven** between tasks.



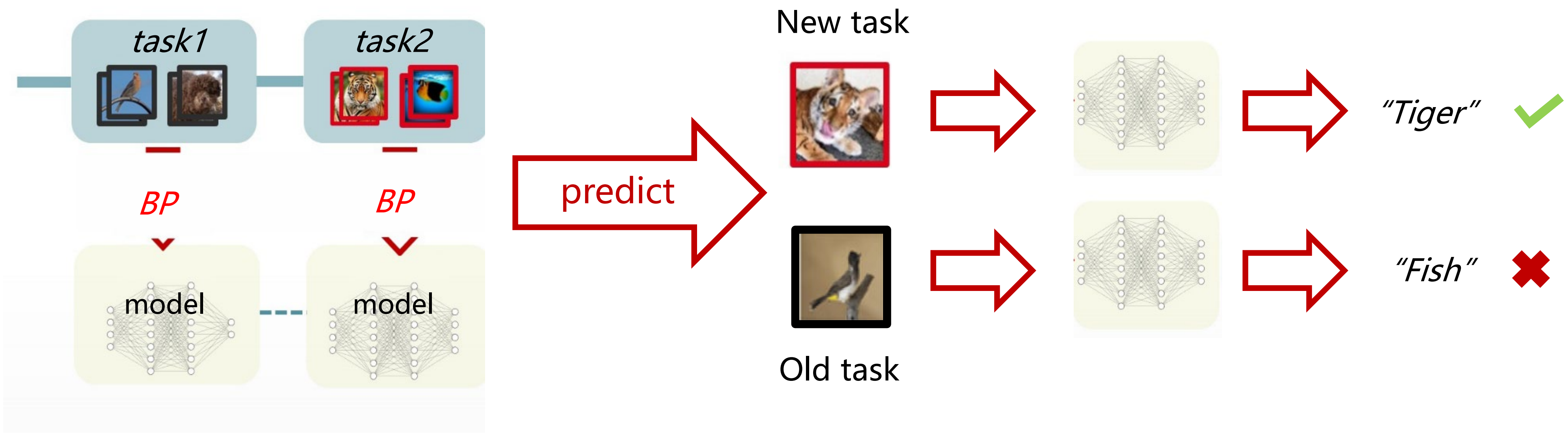


Real-World Data Example: Normalized Search History from Dec 28, 2021, to Dec 28, 2022 (NAVER Search Trend API)

Mi, Fei, et al. "Generalized class incremental learning." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020.
Moon, Jun-Yeong, et al. "Online class incremental learning on stochastic blurry task boundary via mask and visual prompt tuning." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

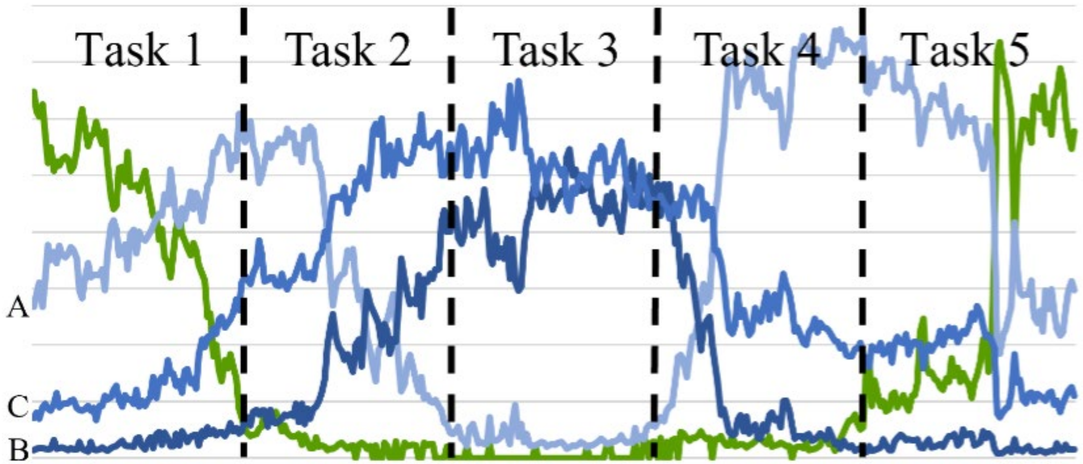
Challenges — Catastrophic Forgetting

- Model Learns in **multiple stages** and different tasks.
- New model does well in **new tasks.**
- Performance **decreases for the previous** tasks.

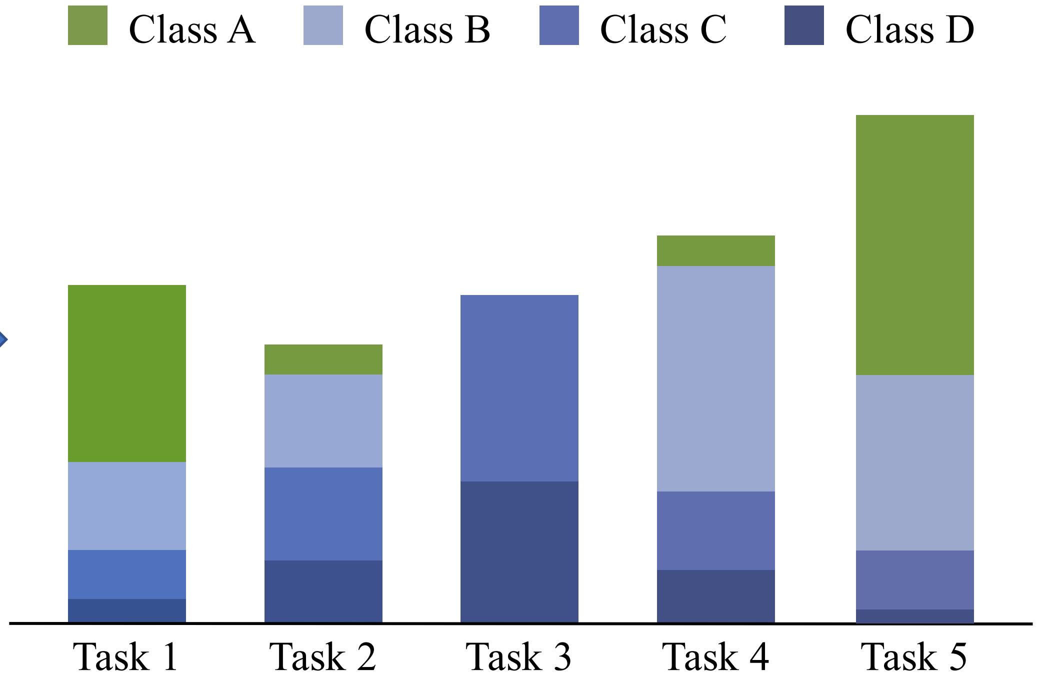


Class Imbalance

- Uneven distribution of data category and sample → **Class imbalance**
- Neglect of minority samples → **Undermining performance**



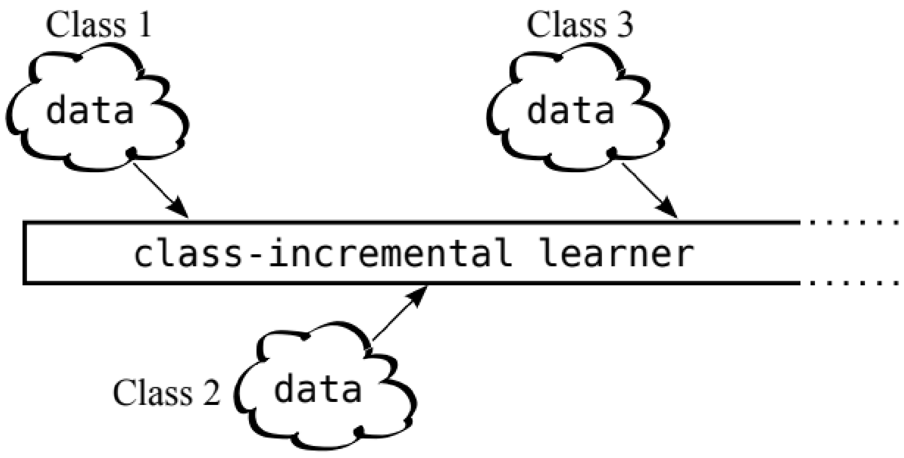
Real-World Data Example: Normalized Search History from Dec 28, 2021, to Dec 28, 2022 (NAVER Search Trend API)



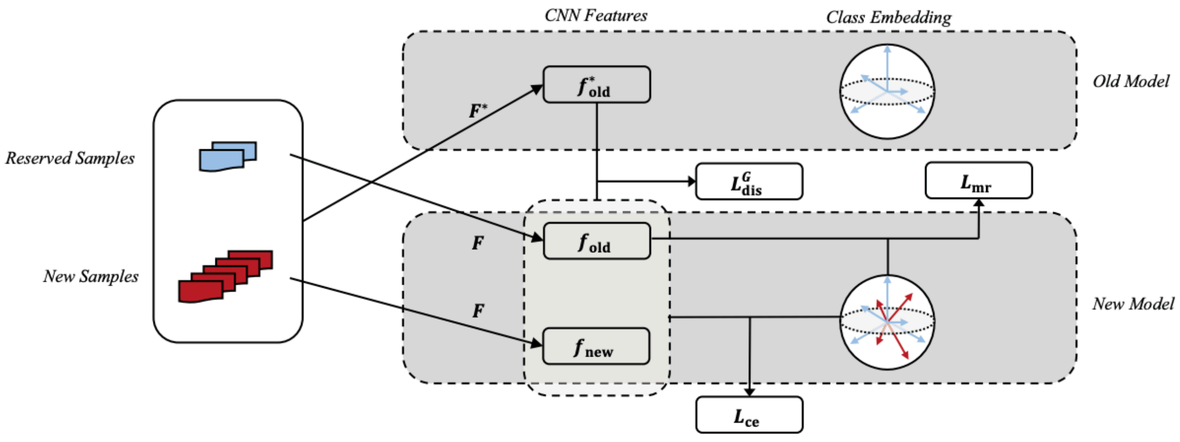
Moon, Jun-Yeong, et al. "Online class incremental learning on stochastic blurry task boundary via mask and visual prompt tuning." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

Replay-based CIL

- Replay-based CIL uses historical exemplars to replay.
- Invade the data privacy
- Hard to cross domain



iCaRL



LUCIR

Rebuffi, Sylvestre-Alvise, et al. "icarl: Incremental classifier and representation learning." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2017.
Hou, Saihui, et al. "Learning a unified classifier incrementally via rebalancing." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

Exemplar-free CIL

- Exemplar-free CIL (EFCIL) learns without storing any historical exemplars.

Regularization-based
- Prevent large drift of important weights

Prototype-based
- A prototype is selected for each category

Hard to resist forgetting
Hard to cross domain

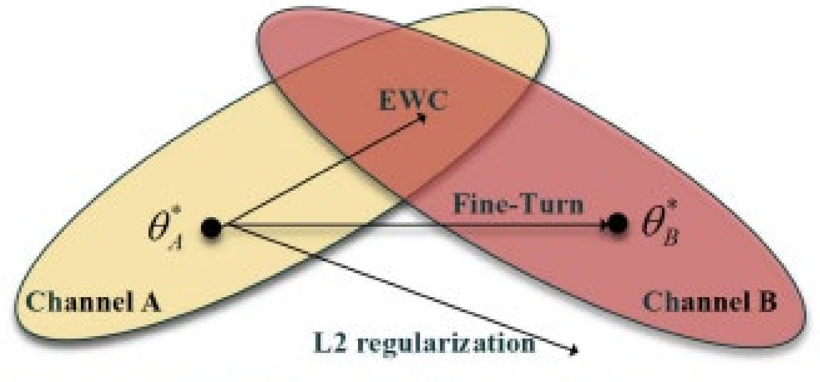
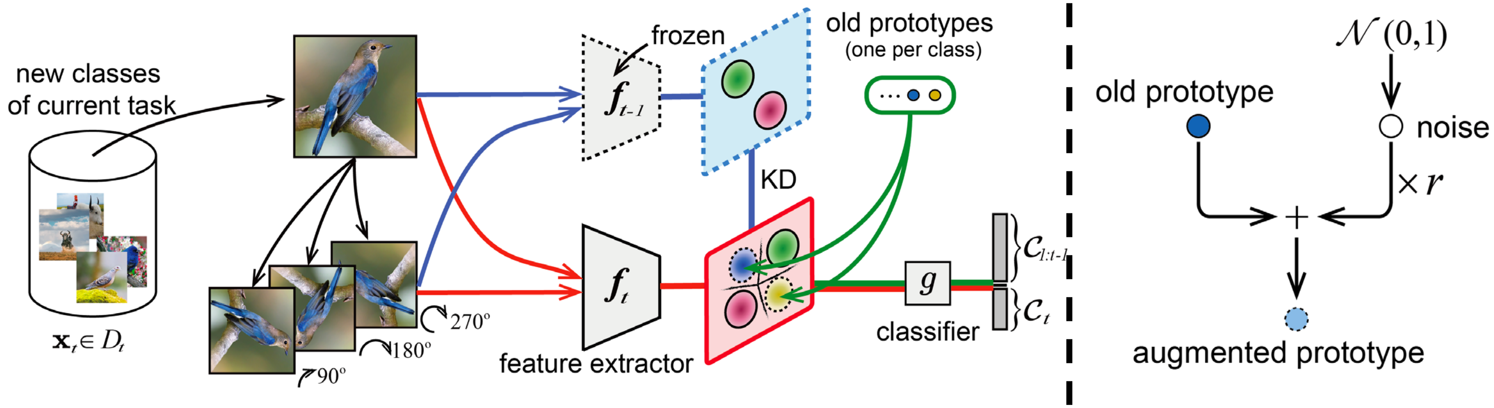


Fig. 2. The trajectory of parameter changes in parameter space

Elastic Weight Consolidation

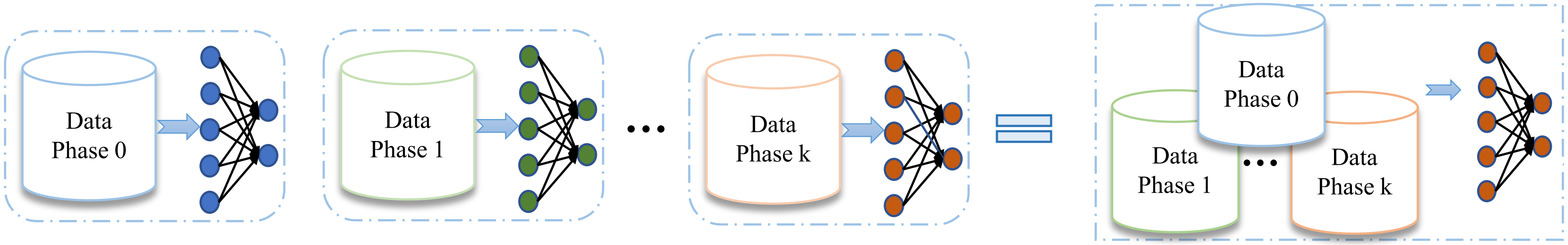


Prototype Augmentation and Self-Supervision for Incremental Learning

Liu, Liyang, et al. "Incdet: In defense of elastic weight consolidation for incremental object detection." IEEE transactions on neural networks and learning systems 32.6 (2020): 2306-2319.
Zhu, Fei, et al. "Prototype augmentation and self-supervision for incremental learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

Analytic Continual Learning

Analytic Continual Learning (ACL) makes the **equivalence** between CIL and joint learning in **linear classifiers**.



Weight-invariance

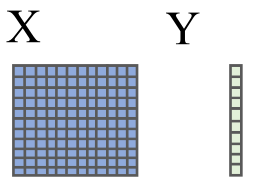
To make the equivalence

However, existing ACL are designed for traditional CIL scenarios.

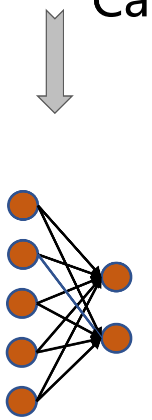
Analytic learning (AL): Use **Least Square** (LS) solutions to train the network

Block-wise recursive Moore-Penrose inverse (BRMP): convert LS to **block-wise calculation**

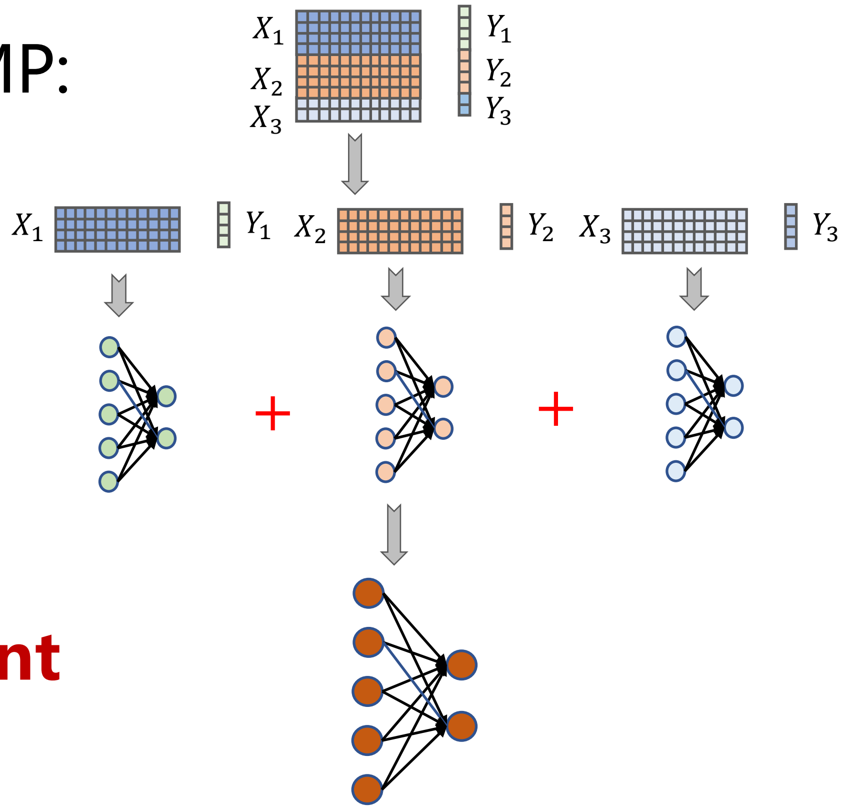
Train with LS:



Calculate **optimal weight** via LS



BRMP:



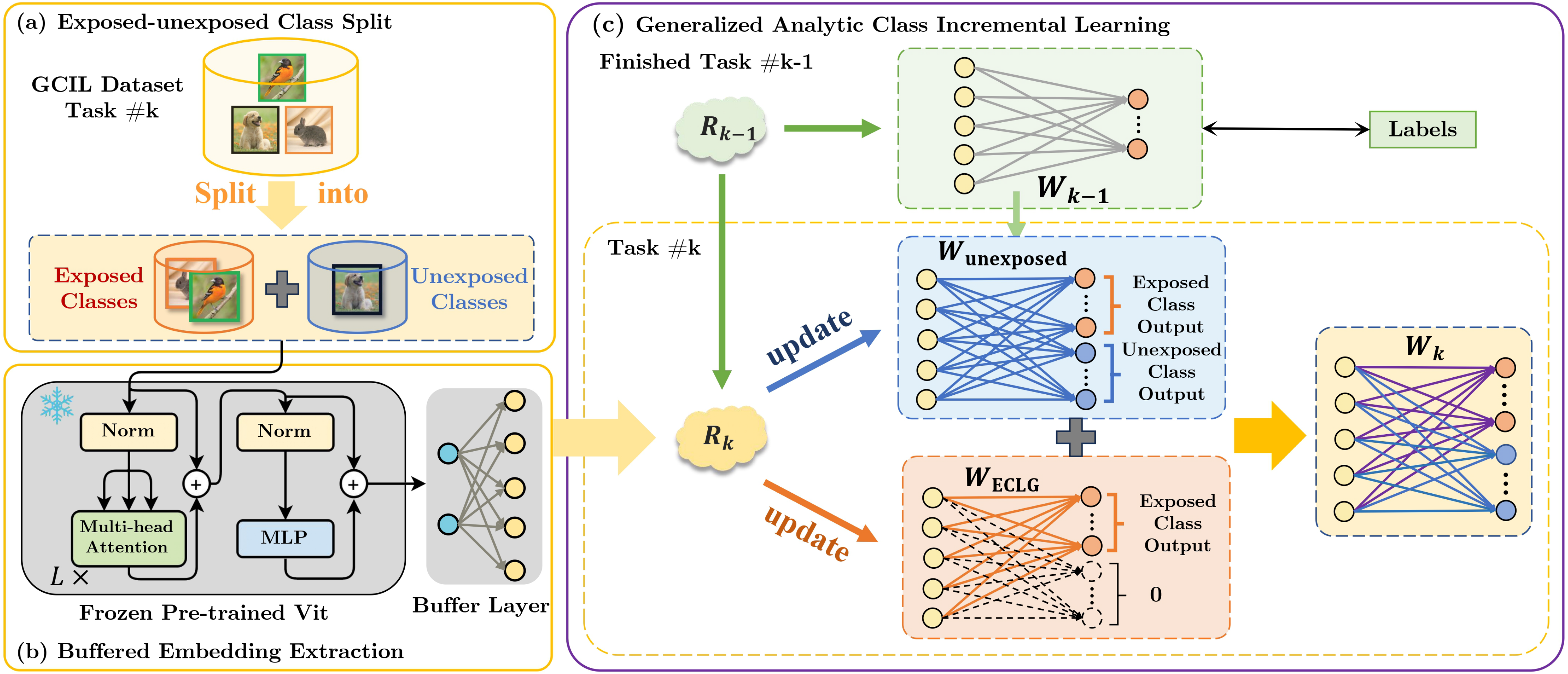
Weight-invariant

Proposed Method

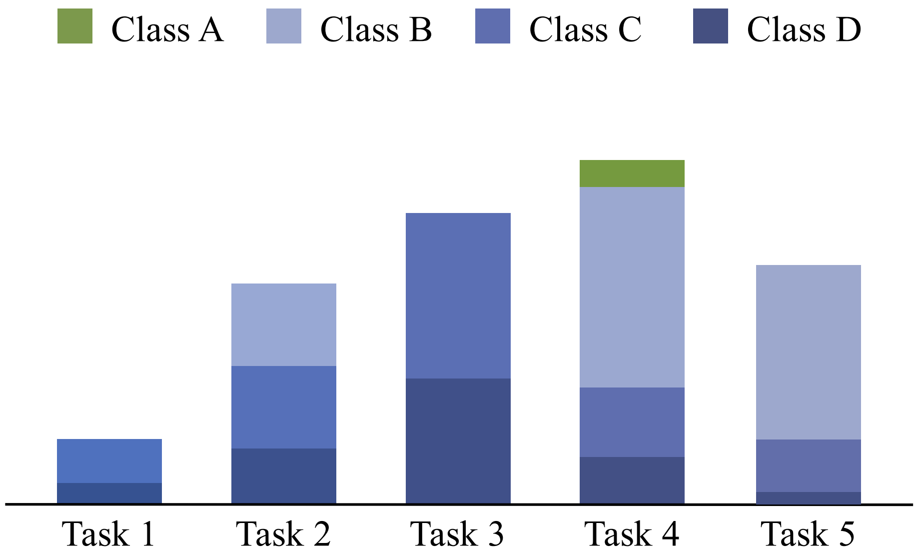
Structure of the GACL

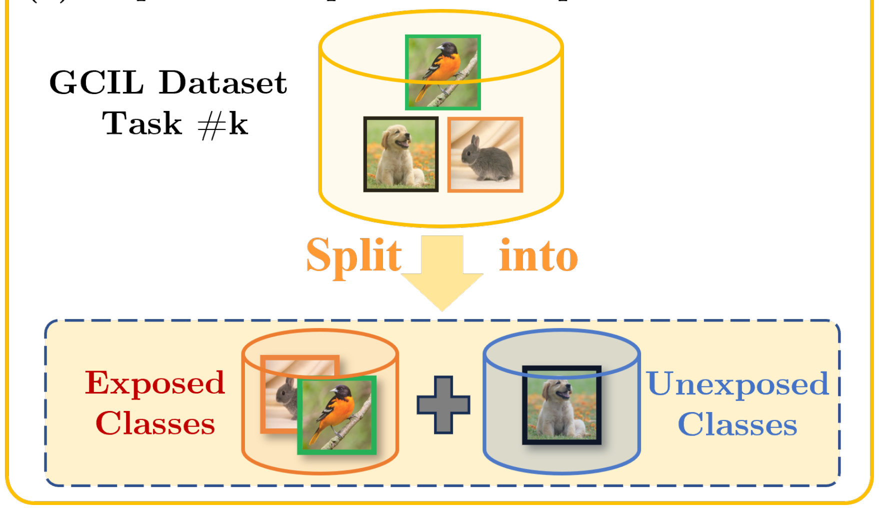
NEURAL INFORMATION PROCESSING SYSTEMS

- Address the uneven data distributions of GCIL
- Extend ACL to GCIL tasks



Exposed-unexposed Class Split





(a) Exposed-unexposed Class Split

- **Exposed classes** in task k: classes that **have appeared** in previous tasks 1 to k − 1
- **Unexposed classes** in task k: classes that **make their initial appearance**

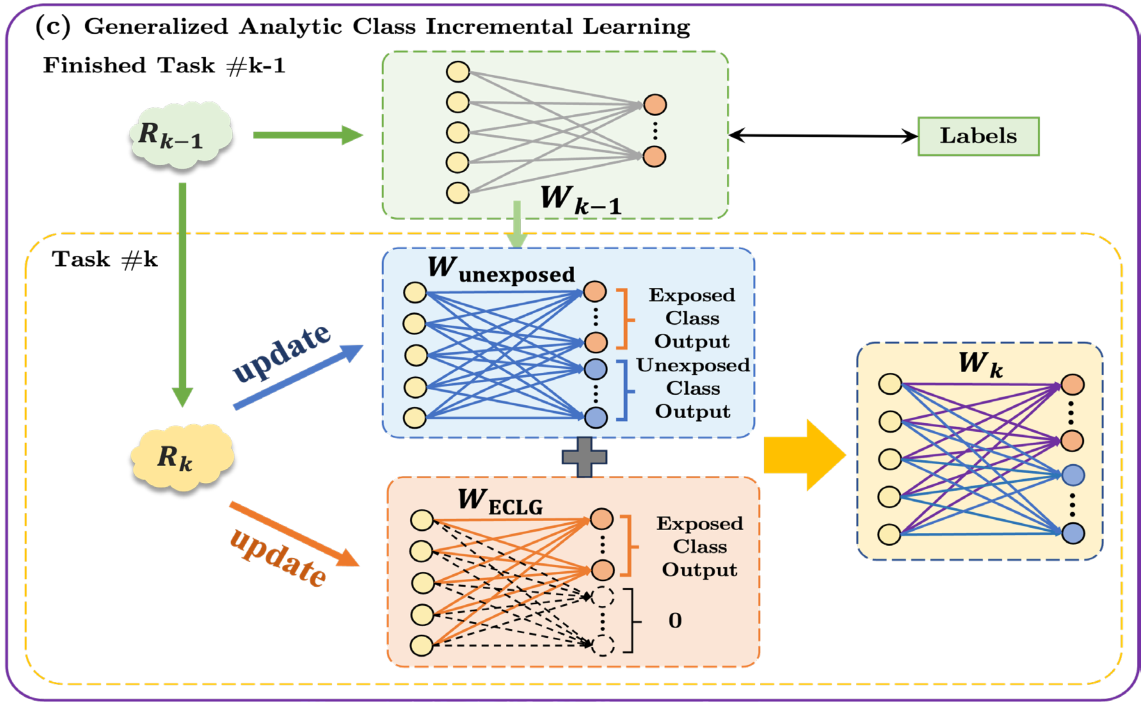
e.g., in task 2, exposed classes: C & D, unexposed class: B

Learning problem in task k

$$\mathbf{Y}_k^{\text{train}} = \left[\boxed{\bar{\mathbf{Y}}_k^{\text{train}}}\ \boxed{\tilde{\mathbf{Y}}_k^{\text{train}}}\right]$$

Unexposed Class Label **Exposed Class Label**

Generalized Analytic Class Incremental Learning



Learning problem in task k

$$X^{\text{total}}_{1:k} = \begin{bmatrix} X^{\text{total}}_{1:k-1} \\ X^{(\text{B})}_k \end{bmatrix} \qquad Y^{\text{total}}_{1:k} = \begin{bmatrix} Y^{\text{total}}_{1:k-1} & \mathbf{0} \\ \bar{Y}^{\text{train}}_k & \tilde{Y}^{\text{train}}_k \end{bmatrix}$$

$$\underset{W^{(k)}_{\text{FCN}}}{\operatorname{argmin}} \left\| \left[\boxed{\begin{matrix} Y^{\text{total}}_{1:k-1} \\ \bar{Y}^{\text{train}}_k \end{matrix}} \ \boxed{\begin{matrix} \mathbf{0} \\ \tilde{Y}^{\text{train}}_k \end{matrix}} \right] - \begin{bmatrix} X^{\text{total}}_{1:k-1} \\ X^{(\text{B})}_k \end{bmatrix} W^{(k)}_{\text{FCN}} \right\|^2_{\text{F}} + \gamma \left\| W^{(k)}_{\text{FCN}} \right\|^2_{\text{F}}$$

Unexposed Class Label

Exposed Class Label

- **Solution of Least Square** = **Solution of Recursive Least Square**

$$\hat{W}^{(k)}_{\text{FCN}} = (X^{\text{total}\top}_{1:k} X^{\text{total}}_{1:k} + \gamma I)^{-1} X^{\text{total}\top}_{1:k} Y^{\text{total}}_{1:k}$$

$$R_k = (X^{\text{total}\top}_{1:k} X^{\text{total}}_{1:k} + \gamma I)^{-1}$$

$$\hat{W}^{(k)}_{\text{FCN}} = \left[\hat{W}^{(k-1)}_{\text{FCN}} - R_k X^{(\text{B})\top}_k X^{(\text{B})}_k \hat{W}^{(k-1)}_{\text{FCN}} + R_k X^{(\text{B})\top}_k \bar{Y}^{\text{train}}_k \quad R_k X^{(\text{B})\top}_k \tilde{Y}^{\text{train}}_k \right]$$

$$= \hat{W}^{(k)}_{\text{unexposed}} + \hat{W}^{(k)}_{\text{ECLG}}$$

$$\boxed{\hat{W}^{(k)}_{\text{unexposed}} = \left[\hat{W}^{(k-1)}_{\text{FCN}} - R_k X^{(\text{B})\top}_k X^{(\text{B})}_k \hat{W}^{(k-1)}_{\text{FCN}} \quad R_k X^{(\text{B})\top}_k \tilde{Y}^{\text{train}}_k \right]}$$

$$\boxed{\hat{W}^{(k)}_{\text{ECLG}} = \left[R_k X^{(\text{B})\top}_k \bar{Y}^{\text{train}}_k \quad \mathbf{0} \right]}$$

ECLG Module

$$R_k = R_{k-1} - R_{k-1} X^{(\text{B})\top}_k (I + X^{(\text{B})}_k R_{k-1} X^{(\text{B})\top}_k)^{-1} X^{(\text{B})}_k R_{k-1}$$

Experiments | Compare with State-of-the-arts

NEURAL INFORMATION PROCESSING SYSTEMS

- **Outperform existing** EFCIL methods by a considerable margin
- **Better than** most SOTA replay-based methods

Mem Size	Method	EFCIL	CIFAR-100 (%)			ImageNet-R (%)			Tiny-ImageNet (%)		
			$\mathcal{A}_{\text{AUC}}$	$\mathcal{A}_{\text{Avg}}$	$\mathcal{A}_{\text{Last}}$	$\mathcal{A}_{\text{AUC}}$	$\mathcal{A}_{\text{Avg}}$	$\mathcal{A}_{\text{Last}}$	$\mathcal{A}_{\text{AUC}}$	$\mathcal{A}_{\text{Avg}}$	$\mathcal{A}_{\text{Last}}$
2000	EWC++ [16]	✗	$53.31_{\pm1.70}$	$50.95_{\pm1.50}$	$52.55_{\pm0.71}$	$36.31_{\pm0.72}$	$39.87_{\pm1.35}$	$29.52_{\pm0.43}$	$52.43_{\pm0.52}$	$54.61_{\pm1.54}$	$37.67_{\pm0.77}$
	ER [35]	✗	$56.17_{\pm1.84}$	$53.80_{\pm1.46}$	$55.60_{\pm0.69}$	$39.31_{\pm0.70}$	$43.03_{\pm1.19}$	$32.09_{\pm0.44}$	$55.69_{\pm0.47}$	$57.87_{\pm1.42}$	$41.10_{\pm0.57}$
	RM [32]	✗	$53.22_{\pm1.82}$	$52.99_{\pm1.69}$	$55.25_{\pm0.61}$	$32.34_{\pm1.88}$	$36.46_{\pm2.23}$	$25.26_{\pm1.08}$	$49.28_{\pm0.43}$	$57.74_{\pm1.57}$	$41.79_{\pm0.34}$
	MVP-R [30]	✗	$60.62_{\pm1.03}$	$57.58_{\pm0.56}$	$64.30_{\pm0.29}$	$47.16_{\pm1.00}$	$50.36_{\pm0.90}$	$42.05_{\pm0.15}$	$61.15_{\pm0.86}$	$62.41_{\pm0.50}$	$51.12_{\pm0.67}$
500	EWC++ [16]	✗	$48.31_{\pm1.81}$	$44.56_{\pm0.96}$	$40.52_{\pm0.83}$	$32.81_{\pm0.76}$	$35.54_{\pm1.69}$	$23.43_{\pm0.61}$	$45.30_{\pm0.61}$	$46.34_{\pm2.05}$	$27.05_{\pm1.35}$
	ER [35]	✗	$51.59_{\pm1.94}$	$48.03_{\pm0.80}$	$44.09_{\pm0.80}$	$35.96_{\pm0.72}$	$39.01_{\pm1.54}$	$26.14_{\pm0.44}$	$48.95_{\pm0.58}$	$50.44_{\pm1.71}$	$29.97_{\pm0.75}$
	RM [32]	✗	$41.07_{\pm1.30}$	$38.10_{\pm0.59}$	$32.66_{\pm0.34}$	$22.45_{\pm0.62}$	$22.08_{\pm1.78}$	$9.61_{\pm0.13}$	$36.66_{\pm0.40}$	$38.83_{\pm2.33}$	$18.23_{\pm0.22}$
	MVP-R [30]	✗	$56.20_{\pm1.47}$	$53.61_{\pm0.04}$	$55.35_{\pm0.43}$	$43.28_{\pm1.41}$	$45.74_{\pm0.97}$	$35.60_{\pm1.18}$	$55.28_{\pm1.42}$	$55.45_{\pm1.02}$	$40.12_{\pm0.40}$
0	LwF [14]	✓	$40.71_{\pm2.13}$	$38.49_{\pm0.56}$	$27.03_{\pm2.92}$	$29.41_{\pm0.83}$	$31.95_{\pm1.86}$	$19.67_{\pm1.27}$	$39.88_{\pm0.90}$	$41.35_{\pm2.59}$	$24.93_{\pm2.01}$
	L2P [36]	✓	$42.68_{\pm2.70}$	$39.89_{\pm0.45}$	$28.59_{\pm3.34}$	$30.21_{\pm0.91}$	$32.21_{\pm1.73}$	$18.01_{\pm3.07}$	$41.67_{\pm1.17}$	$42.53_{\pm2.52}$	$24.78_{\pm2.31}$
	DualPrompt [33]	✓	$41.34_{\pm2.59}$	$38.59_{\pm0.68}$	$22.74_{\pm3.40}$	$30.44_{\pm0.88}$	$32.54_{\pm1.84}$	$16.07_{\pm3.20}$	$39.16_{\pm1.13}$	$39.81_{\pm3.03}$	$20.42_{\pm3.37}$
	MVP [30]	✓	$45.07_{\pm2.43}$	$44.93_{\pm0.54}$	$39.94_{\pm0.47}$	$35.77_{\pm2.55}$	$35.58_{\pm1.20}$	$22.06_{\pm5.01}$	$46.43_{\pm3.07}$	$45.41_{\pm1.09}$	$28.21_{\pm2.89}$
	SLDA [37]	✓	$53.00_{\pm3.85}$	$50.09_{\pm2.77}$	$61.79_{\pm3.81}$	$33.11_{\pm3.17}$	$33.78_{\pm1.76}$	$39.02_{\pm1.30}$	$46.43_{\pm3.07}$	$45.41_{\pm4.43}$	$53.13_{\pm2.29}$
	GACL (ours)	✓	$\mathbf{57.99}_{\pm2.46}$	$\mathbf{56.24}_{\pm3.12}$	$\mathbf{70.31}_{\pm0.06}$	$\mathbf{41.68}_{\pm0.78}$	$\mathbf{47.30}_{\pm0.84}$	$\mathbf{42.22}_{\pm0.10}$	$\mathbf{63.14}_{\pm0.66}$	$\mathbf{69.32}_{\pm0.87}$	$\mathbf{62.68}_{\pm0.08}$

Experiments — Ablation Study on the ECLG Module

- The ECLG module captures knowledge from exposed-class labels
- Demonstrate competence with ECLG module

r_B	Dataset	With ECLG			Without ECLG		
		$\mathcal{A}_{\text{AUC}}(\%)$	$\mathcal{A}_{\text{Avg}}(\%)$	$\mathcal{A}_{\text{Last}}(\%)$	$\mathcal{A}_{\text{AUC}}(\%)$	$\mathcal{A}_{\text{Avg}}(\%)$	$\mathcal{A}_{\text{Last}}(\%)$
10%	CIFAR-100	**57.99**$_{\pm 2.46}$	**56.24**$_{\pm 3.12}$	**70.31**$_{\pm 0.06}$	45.68$_{\pm 7.74}$	42.04$_{\pm 4.52}$	47.30$_{\pm 2.61}$
	ImageNet-R	**41.68**$_{\pm 0.78}$	**47.30**$_{\pm 0.84}$	**42.22**$_{\pm 0.10}$	40.29$_{\pm 2.23}$	46.95$_{\pm 1.15}$	41.67$_{\pm 0.36}$
	Tiny-ImageNet	**63.14**$_{\pm 0.66}$	**69.32**$_{\pm 0.87}$	**62.68**$_{\pm 0.08}$	60.21$_{\pm 1.86}$	65.80$_{\pm 1.20}$	60.13$_{\pm 0.37}$
30%	CIFAR-100	**57.33**$_{\pm 1.03}$	**58.74**$_{\pm 1.59}$	**69.90**$_{\pm 0.01}$	42.53$_{\pm 1.97}$	42.26$_{\pm 1.75}$	45.49$_{\pm 1.17}$
	ImageNet-R	**42.19**$_{\pm 0.44}$	**47.82**$_{\pm 1.11}$	**42.90**$_{\pm 0.08}$	42.01$_{\pm 0.26}$	46.95$_{\pm 1.15}$	41.67$_{\pm 0.56}$
	Tiny-ImageNet	**60.73**$_{\pm 1.15}$	**67.31**$_{\pm 1.14}$	**59.73**$_{\pm 2.55}$	60.63$_{\pm 1.86}$	57.03$_{\pm 1.98}$	60.13$_{\pm 0.55}$
50%	CIFAR-100	**56.74**$_{\pm 1.14}$	**58.29**$_{\pm 1.95}$	**70.02**$_{\pm 0.05}$	40.91$_{\pm 3.57}$	47.25$_{\pm 2.64}$	58.61$_{\pm 2.62}$
	ImageNet-R	**41.33**$_{\pm 1.46}$	**46.42**$_{\pm 2.30}$	**42.92**$_{\pm 0.17}$	40.44$_{\pm 3.14}$	42.50$_{\pm 3.43}$	39.05$_{\pm 1.65}$
	Tiny-ImageNet	**60.96**$_{\pm 1.83}$	**66.28**$_{\pm 2.69}$	**62.24**$_{\pm 0.10}$	60.32$_{\pm 4.20}$	60.70$_{\pm 4.30}$	56.97$_{\pm 1.89}$

Experiments — Robustness Analysis in Si-Blurry Setting

NEURAL INFORMATION PROCESSING SYSTEMS

- Investigate with various disjoint class ratio and the blurry sample ratio
- Demonstrate the robustness of the GACL

r_D	Method	$A_{AUC}(\%)$	$A_{Avg}(\%)$	$A_{Last}(\%)$
0%	SLDA [37]	**55.51**$_{\pm 1.93}$	**53.94**$_{\pm 0.92}$	67.45$_{\pm 0.26}$
	MVP-R [30]	53.49$_{\pm 1.40}$	50.73$_{\pm 0.37}$	60.54$_{\pm 2.03}$
	GACL (ours)	49.96$_{\pm 0.61}$	50.56$_{\pm 0.49}$	**69.94**$_{\pm 0.09}$
50%	SLDA [37]	53.00$_{\pm 3.85}$	50.09$_{\pm 2.77}$	61.79$_{\pm 3.81}$
	MVP-R [30]	56.20$_{\pm 1.47}$	53.61$_{\pm 0.04}$	55.35$_{\pm 0.43}$
	GACL (ours)	**57.99**$_{\pm 2.46}$	**56.24**$_{\pm 3.12}$	**70.31**$_{\pm 0.06}$
100%	SLDA [37]	65.46$_{\pm 4.79}$	67.29$_{\pm 5.28}$	63.56$_{\pm 2.68}$
	MVP-R [30]	68.43$_{\pm 0.28}$	68.04$_{\pm 1.48}$	53.14$_{\pm 0.72}$
	GACL (ours)	**70.72**$_{\pm 0.32}$	**77.57**$_{\pm 1.02}$	**69.97**$_{\pm 0.03}$

The performance at different disjoint class ratio with blurry sample ratio = 10% on CIFAR-100.

r_B	Method	$A_{AUC}(\%)$	$A_{Avg}(\%)$	$A_{Last}(\%)$
10%	SLDA [37]	53.00$_{\pm 3.85}$	50.09$_{\pm 2.77}$	61.79$_{\pm 3.81}$
	MVP-R [30]	56.20$_{\pm 1.47}$	53.61$_{\pm 0.04}$	55.35$_{\pm 0.43}$
	GACL (ours)	**57.99**$_{\pm 2.46}$	**56.24**$_{\pm 3.12}$	**70.31**$_{\pm 0.06}$
30%	SLDA [37]	54.55$_{\pm 4.66}$	54.06$_{\pm 2.41}$	63.04$_{\pm 2.56}$
	MVP-R [30]	59.65$_{\pm 2.04}$	58.31$_{\pm 1.52}$	58.16$_{\pm 1.38}$
	GACL (ours)	**57.33**$_{\pm 1.03}$	**58.74**$_{\pm 1.59}$	**69.90**$_{\pm 0.01}$
50%	SLDA [37]	53.81$_{\pm 3.43}$	52.93$_{\pm 2.36}$	63.45$_{\pm 2.72}$
	MVP-R [30]	59.10$_{\pm 1.98}$	57.34$_{\pm 1.96}$	54.81$_{\pm 0.21}$
	GACL (ours)	**56.74**$_{\pm 1.14}$	**58.29**$_{\pm 1.95}$	**70.02**$_{\pm 0.05}$

The performance at different blurry sample ratio with disjoint class ratio = 50% on CIFAR-100.

NEURAL INFORMATION PROCESSING SYSTEMS

Thanks for Watching!

- Our paper is available at:

  https://arxiv.org/abs/2403.15706

- Our codes are available at:

  https://github.com/CHEN-YIZHU/GACL