The Dual Nature of Plasticity Loss in Deep Continual Learning: Dissection and Mitigation

Haoyu Albert Wang[1,2,†] , Wei P. Dai[1,4, †,‡], Jiawei Zhang[1,3], Jialun Ma[1], Mingyi Huang[1,2], Yuguo Yu[1,2,3,4,‡]

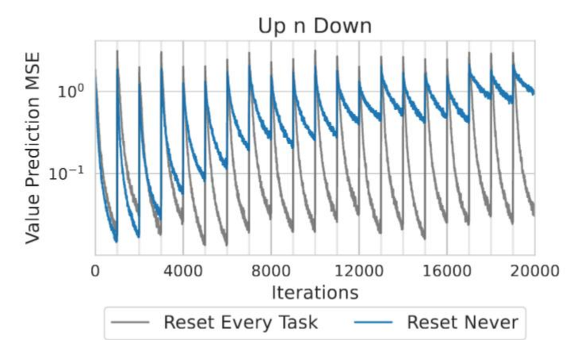
1.Research Institut[1,2,3,4]e of Intelligent Complex Systems, Fudan University.2. Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University.3. State Key Laboratory of Brain Function and Disorders and MOE Frontiers Center for BrainScience, Institutes of Brain Science, Fudan University.4. Shanghai Artificial Intelligence Laboratory.

† These authors contributed equally to this work.‡ Corresponding author.

Introduction

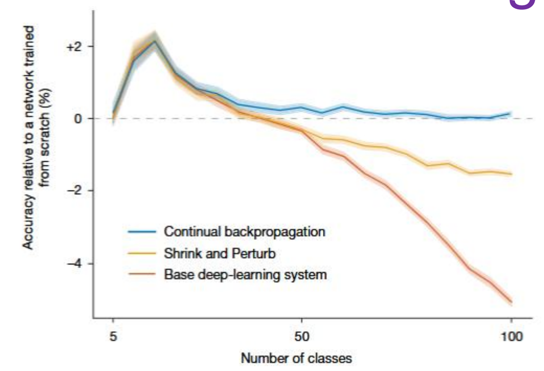
Loss of Plasticity in Deep Continual Learning

Reinforcement Learning



Nikishin et. al., 2023

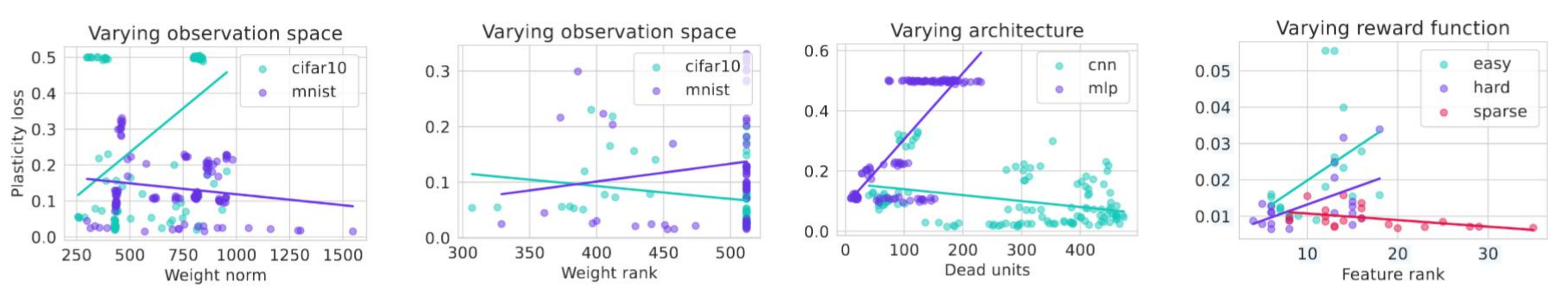
Continual Learning



Dohare et. al., 2024

***Loss of plasticity** is a widely observed phenomenon in both continual learning and RL. It refers to the degradation of performance on new tasks, which eventually prevents the system from learning continuously.*

Potential factors of LoP are Inconclusive and Indirect

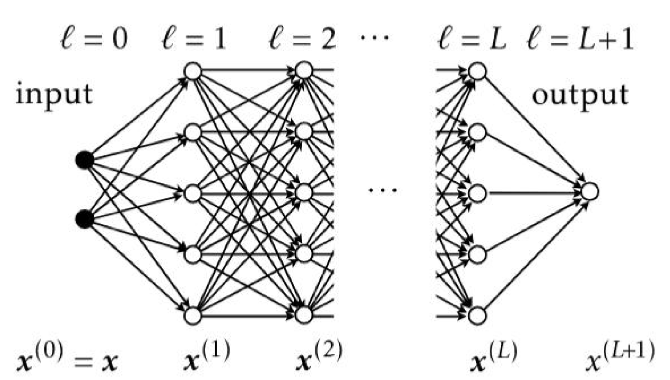


Lyle et. al., 2023

A more comprehensive understanding of LoP is in need.

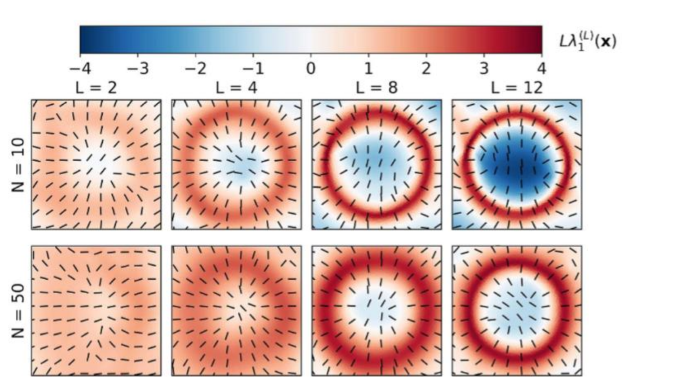
Preliminary

A Framework Combining FTLE and UFM

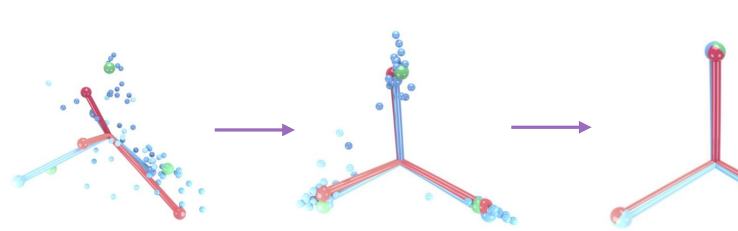


Finite Time Lyapunov Exponent:

$$\delta x^{(\ell)} = \mathbb{J}_\ell \delta x \qquad \lambda_1^{(\ell)}(x) \equiv \ell^{-1} \log \Lambda_1^{(\ell)}(x)$$



L. Storm et. al., 2024



Vardan Papyan et al. 2020

(NC1) Variability collapse: $\Sigma_W \to 0$.
(NC2) Convergence to simplex ETF:

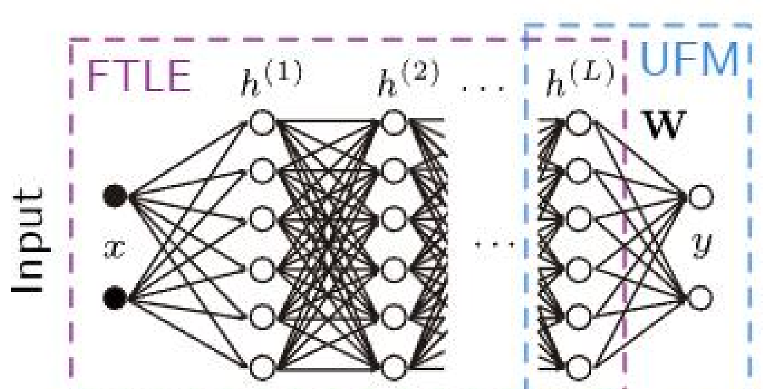
$$\big|\|\mu_c - \mu_G\|_2 - \|\mu_{c'} - \mu_G\|_2\big| \to 0 \quad \forall\, c, c'$$
$$\langle \tilde{\mu}_c, \tilde{\mu}_{c'} \rangle \to \frac{C}{C-1}\delta_{c,c'} - \frac{1}{C-1} \quad \forall\, c, c'.$$

(NC3) Convergence to self-duality:

$$\left\| \frac{W^\top}{\|W\|_F} - \frac{\dot{M}}{\|\dot{M}\|_F} \right\|_F \to 0.$$

(NC4) Simplification to NCC:

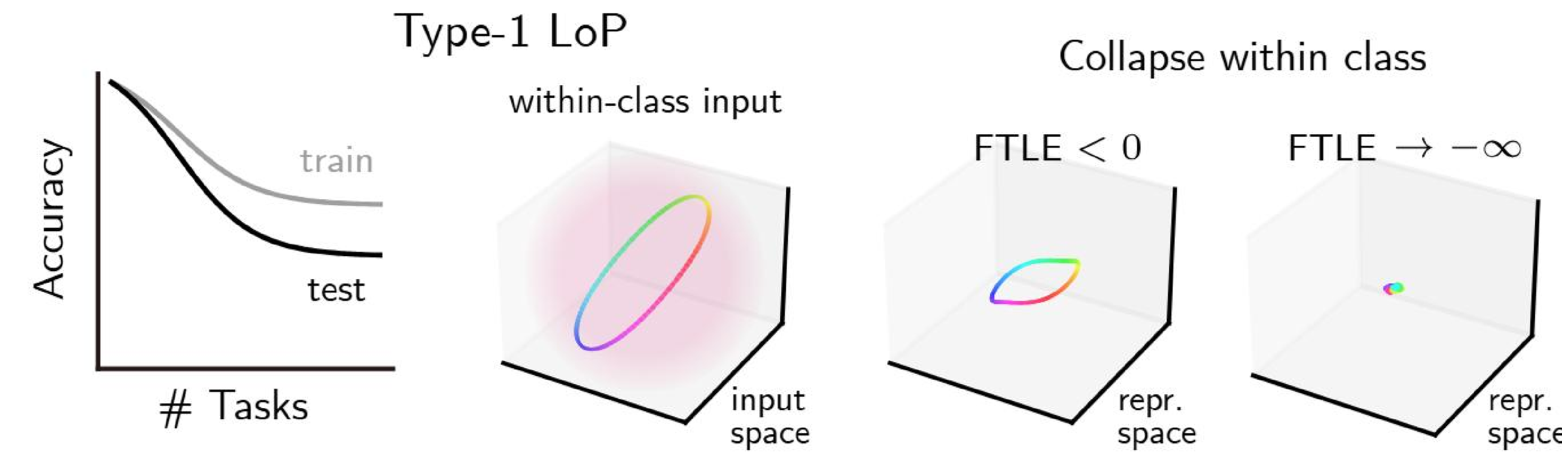
$$\arg\max_{c'} \langle w_{c'}, h\rangle + b_{c'} \to \arg\min_{c'} \|h - \mu_{c'}\|_2,$$



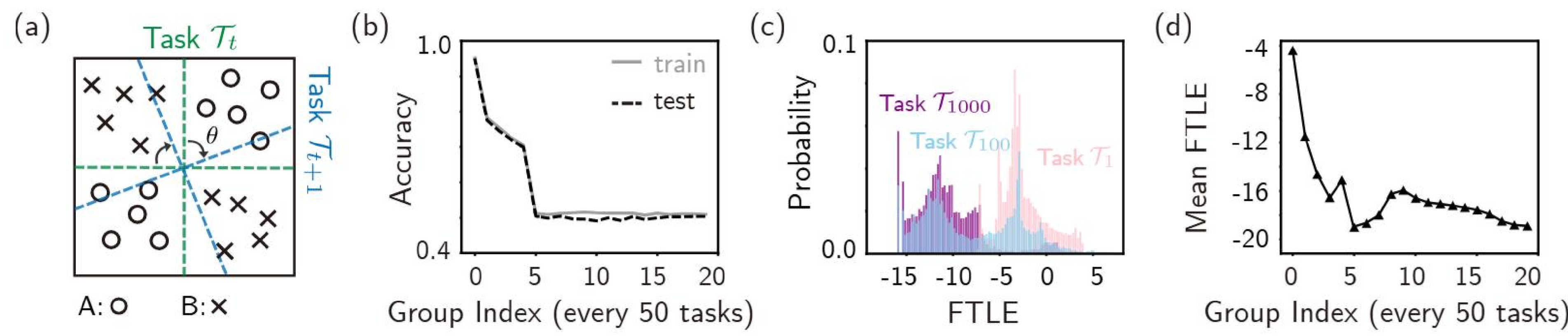
FTLE quantifies how the mapping properties of neural networks from the input layer to the representation layer evolve during training, while **UFM** offers analytical tractability for optimization in the representation space.

Dissection

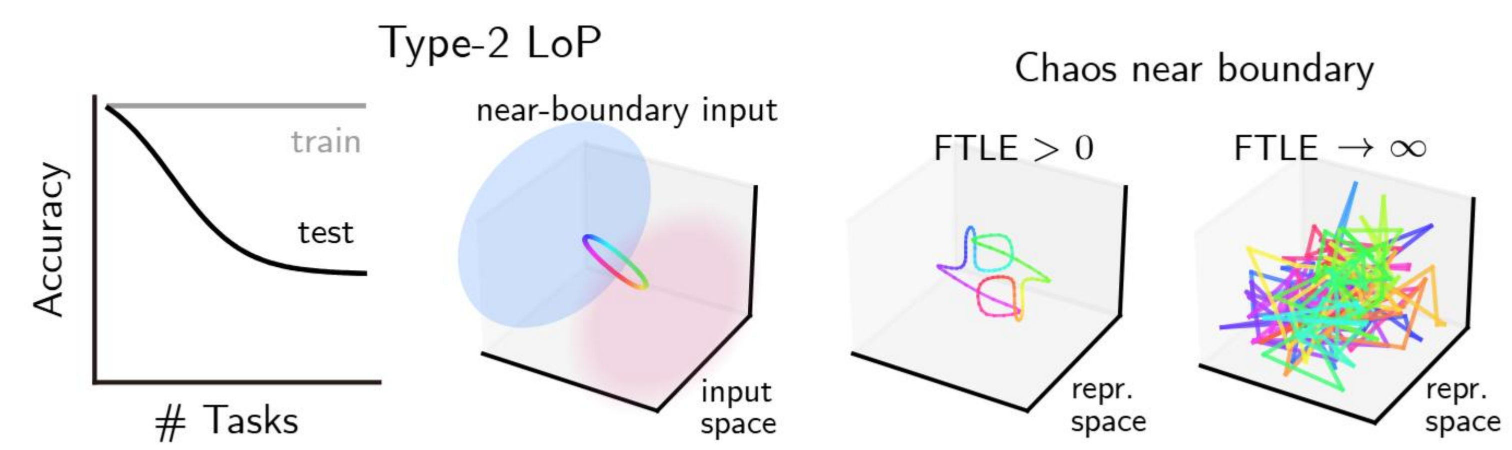
- We identify the existence of **two subtypes of LoP**. On the level of task performance, they only **differ in training accuracy**.
- We unveil that the causes of the two LoP subtypes are exactly opposite: **collapse of representation** vs. **chaotic behavior**.



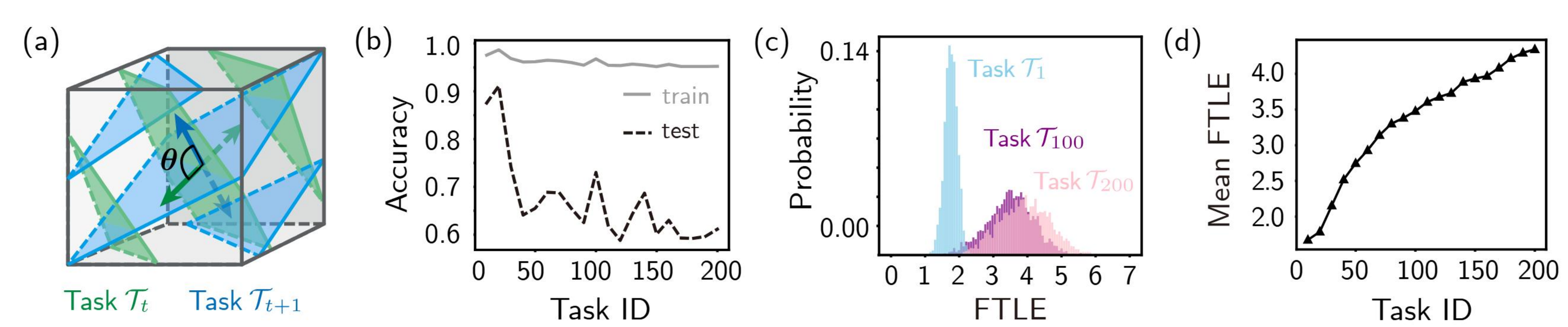
Type-1 LoP: the Collapse of Representation Space



***Type-1 LoP:** Learning causes within-class regions to collapse progressively. These collapsed areas accumulate during continual learning. Type-1 LoP occurs when representations of a new task approach these collapsed regions, characterized by **highly negative FTLEs.** Both training and test accuracies drop sharply, indicating a loss of capacity in learning.*

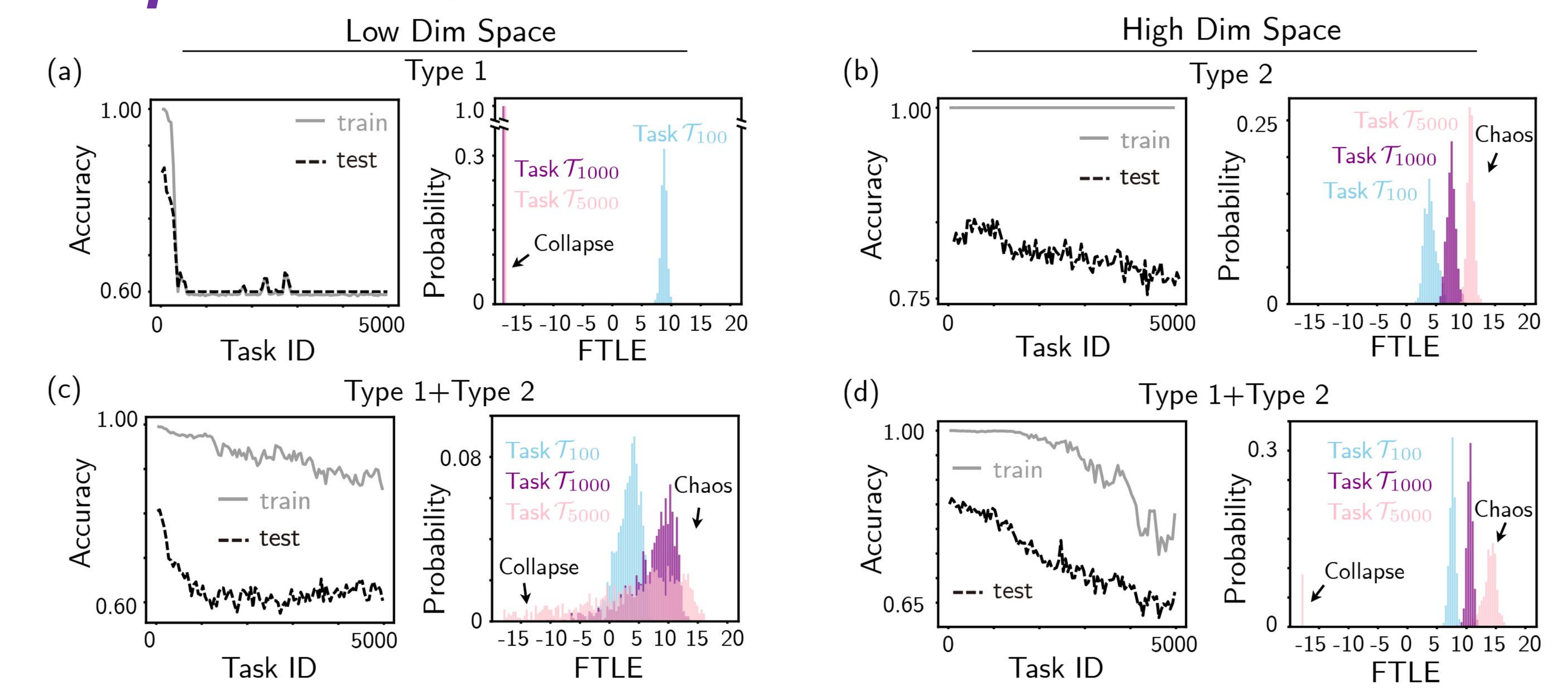


Type-2 LoP : Over-stretched Boundaries and Chaotic Behaviors



***Type-2 LoP:** Learning causes inter-class regions to expand progressively. These expansions accumulate during continual learning. Type-2 LoP occurs when representations of a new task approach these overly stretched regions in representation space, characterized by highly positive FTLEs. Training accuracy remains high due to the chaotic and over-expressive representation space, while test accuracy degrades, indicating a loss of capacity to generalize.*

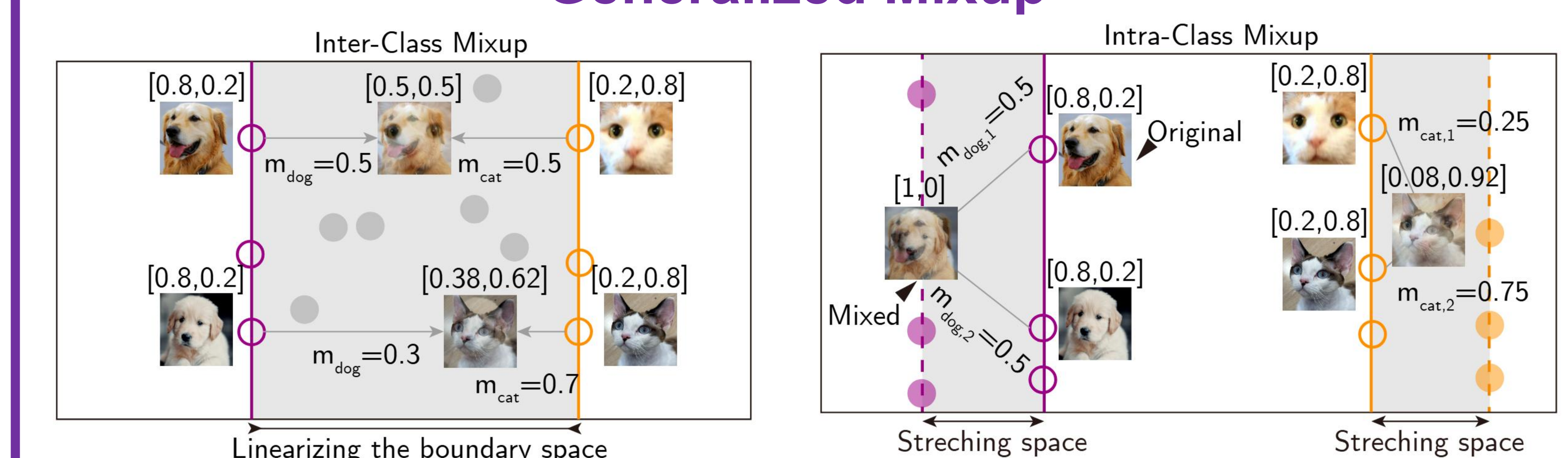
Experiments



- **Both types of LoP exist in real datasets** and they can impair task performance simultaneously.

Mitigation

Generalized Mixup



$$x^m = mx^K + (1-m)x^{K'}, \qquad y^m = my^K + (1-m)y^{K'}$$

$$y^m_K = y^K + \frac{M}{2} - M|0.5-m| \qquad y^m_{K'} = y^{K'} - \frac{M}{2} + M|0.5-m|$$

- **Classical Mixup** (Inter-Class) suppresses Type-2 LoP by **linearizing the decision boundaries.**
- **Generalized Mixup** (Intra-Class) mitigates Type-1 LoP by **expanding the intra-class representation.**

Verified in Continual Imagenet Benchmark

Table 1: SmallConv Acc. on Continual ImageNet

Task (×1000)	0-1	1-2	2-3	3-4	4-5
No Intv.	0.817	0.805	0.562	0.500	0.500
Retrained	0.853	0.845	0.845	0.840	0.840
L2 init	0.804	0.796	0.786	0.785	0.788
Layernorm	0.753	0.760	0.759	0.751	0.751
CBP	0.834	0.847	0.846	0.847	0.857
G-mixup[ours]	**0.866**	**0.881**	**0.885**	**0.880**	**0.879**

Table 2: ConvNet Acc. on Continual ImageNet

Task (×1000)	0-1	1-2	2-3	3-4	4-5
No Intv.	0.794	0.778	0.604	0.537	0.500
Retrained	0.857	0.851	0.850	0.849	0.846
L2 init	0.814	0.805	0.800	0.803	0.807
Layernorm	0.782	0.768	0.752	0.749	0.755
CBP	0.848	0.867	0.864	0.863	0.878
G-mixup[ours]	**0.875**	**0.896**	**0.899**	**0.894**	**0.896**

G-Mixup consistently outperforms other methods across tasks and settings, demonstrating superior capacity to preserve plasticity over long task sequences.

Acknowledgment

This work is supported by Science and Technology Innovation 2030 – Brain Science and Brain-Inspired Intelligence Project (2021ZD0201301), the National Natural Science Foundation of China (12201125, 9257020), Shanghai Municipal Science and Technology Committeeof Shanghai outstanding academic leaders plan (21XD1400400); Shanghai Institute for Mathematics and Interdisciplinary Sciences (SIMIS-ID-2025-NC). The computations in this research were performed using the CFFF platform of Fudan University.

haoyuwang18@fudan.edu.cn, weidai@fudan.edu.cn, yuyuguo@fudan.edu.cn