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

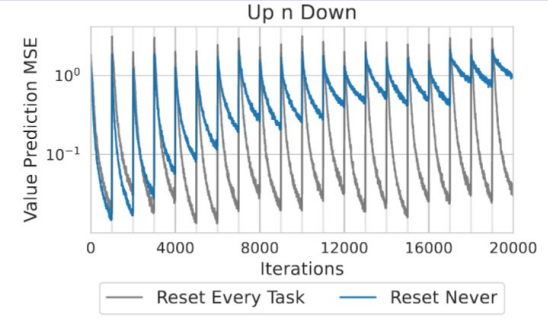
Haoyu Albert Wang, Wei P. Dai, Jiawei Zhang,
Jialun Ma, Mingyi Huang, Yuguo Yu

------- *Neural Information Processing Systems (2025)* -------

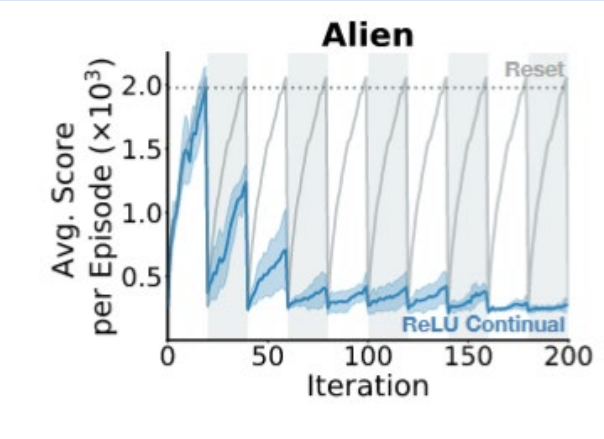
# Loss of Plasticity in Deep Continual Learning

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

## RL

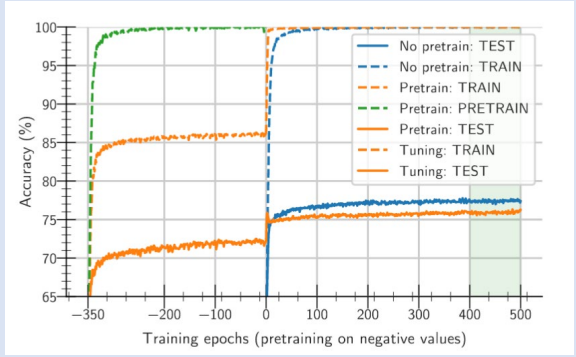



Evgenii Nikishin et al. 2023

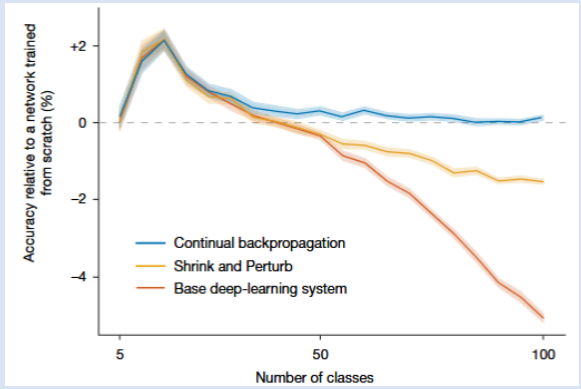



Zaheer Abbas et al. 2023

## CL

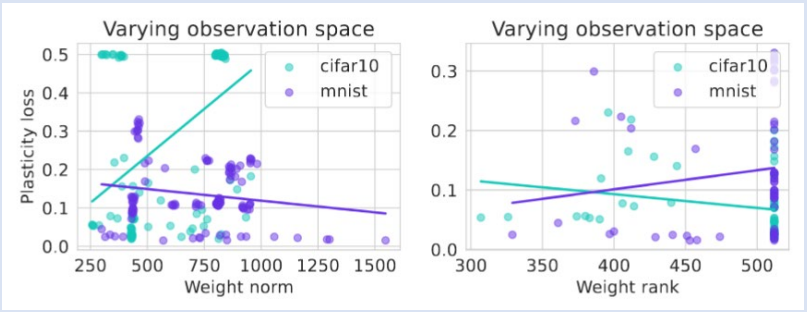


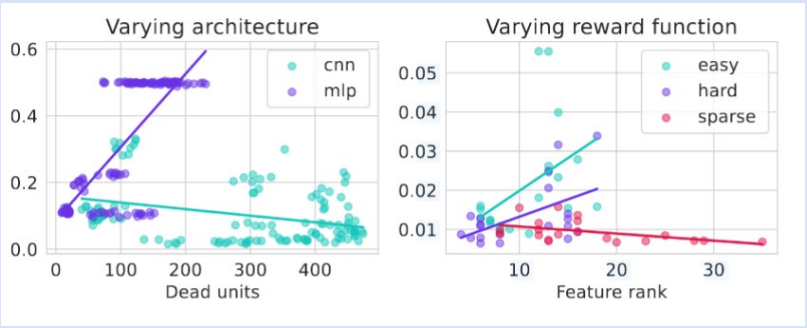

Tudor Berariu et al. 2023




Shibhansh Dohare et al. 2024

## Potential factors of LoP



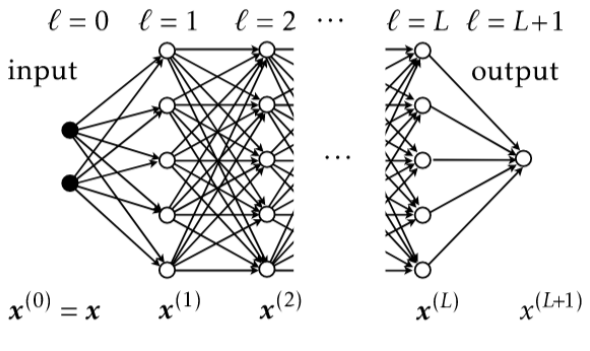

Clare Lyle et al. 2023

**Inconclusive and Indirect**

# Preliminary

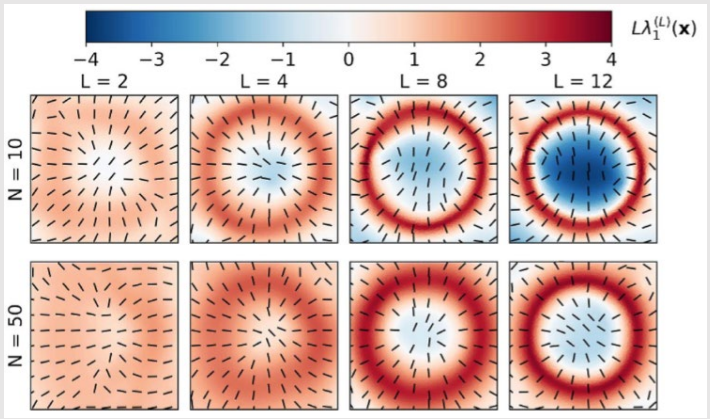
## Combining FTLE and Neural Collapse Reveals the Dynamics of Representation Space Underlying Continual Learning

### FTLE as a Framework for Understanding the Geometric Structure of Neural Network Mappings



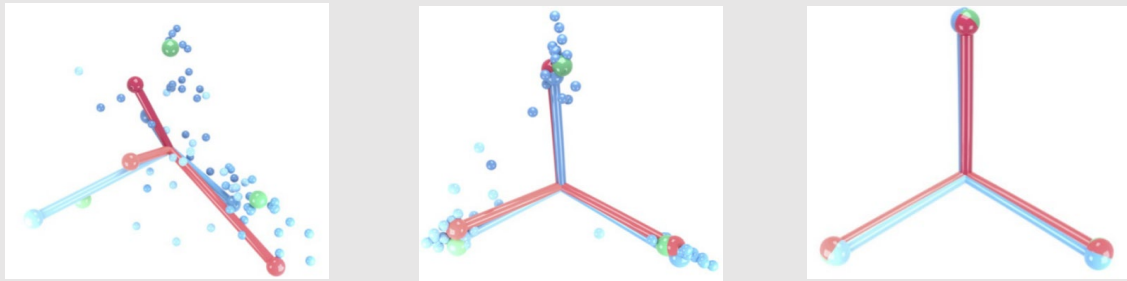

$$\delta x^{(\ell)} = \mathbb{J}_\ell \delta x$$

$$\lambda_1^{(\ell)}(x) \equiv \ell^{-1} \log \Lambda_1^{(\ell)}(x)$$




L. Storm et al. 2024

### Prevalence of neural collapse during the terminal phase of deep learning training



**(NC1) Variability collapse: $\Sigma_W \to 0$.**

**(NC2) Convergence to simplex ETF:**

$$|\|\mu_c - \mu_G\|_2 - \|\mu_{c'} - \mu_G\|_2| \to 0 \quad \forall\, c, c'$$

$$\langle \tilde{\mu}_c, \tilde{\mu}_{c'} \rangle \to \frac{C}{C-1}\delta_{c,c'} - \frac{1}{C-1} \quad \forall\, c, c'.$$

**(NC3) Convergence to self-duality:**

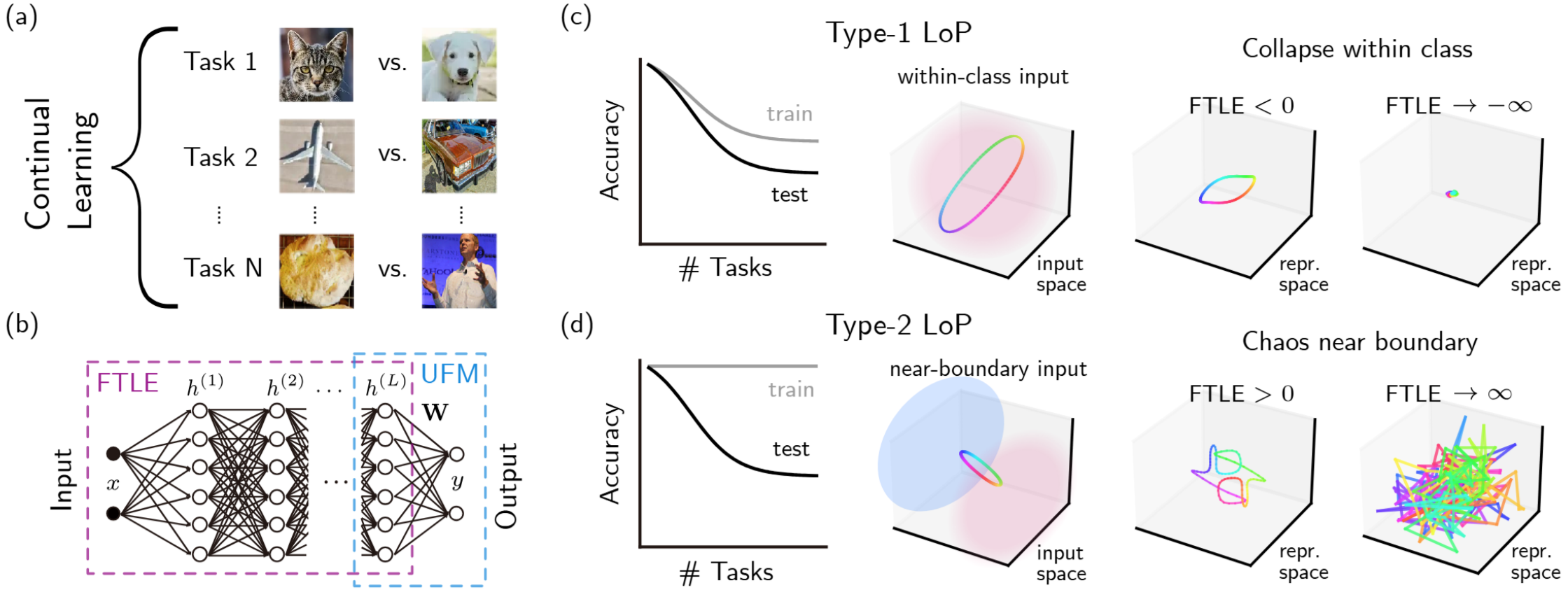
$$\left\| \frac{W^\top}{\|W\|_F} - \frac{\dot{M}}{\|\dot{M}\|_F} \right\|_F \to 0.$$

**(NC4) Simplification to NCC:**

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

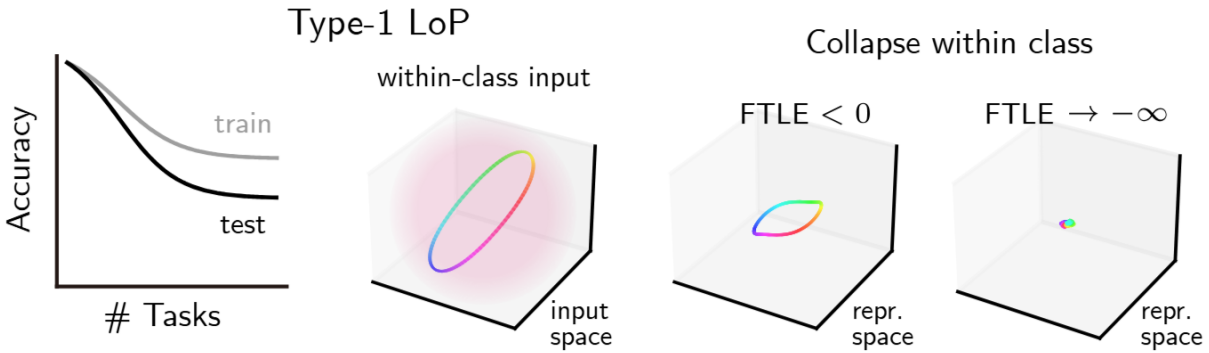
Vardan Papyan et al. 2020

# Overview




- **Propose** an FTLE + UFM framework to analyze representation-space dynamics in continual learning.
- **Identify and validate two different types of LoP, both theoretically and experimentally.**
- **Develop** a generalized Mixup method that effectively mitigates both types of LoP.

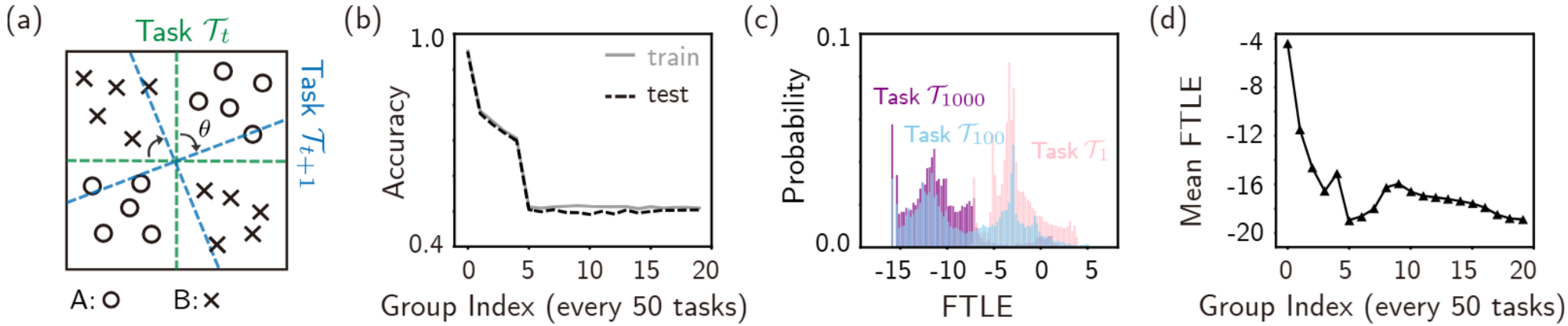
# Type-1 LoP: the Collapse of Representation Space




$$\log \beta = \lambda^{k,k}_{\text{trained}} - \lambda^{k,k}_{0} < 0 \qquad \lambda^{k,k}_{\text{trained},T} \to -\infty \quad \text{as} \quad \prod_{t=1}^{T} \beta_t \to 0$$

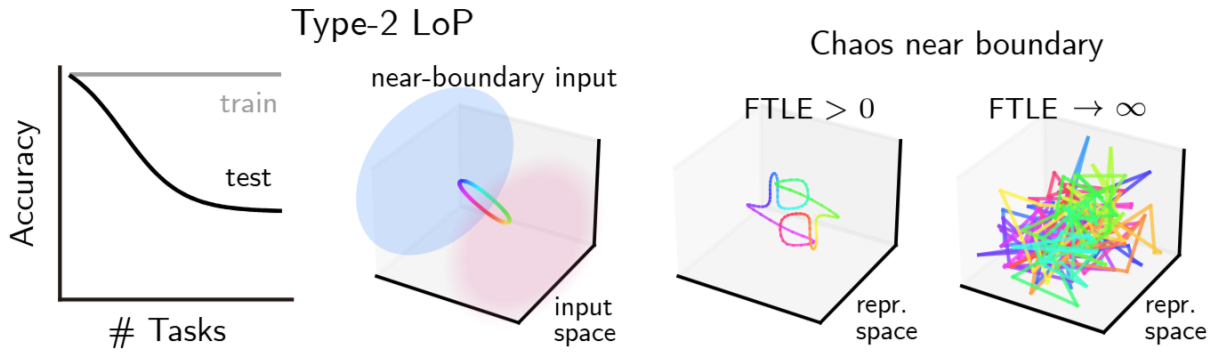
**Type-1 LoP:**

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




According to our theory, **Type-1 LoP** is more likely to occur in **low-dimensional representation spaces**. Thus we verified this using a **toy neural network** with a low-dimensional representation layer.

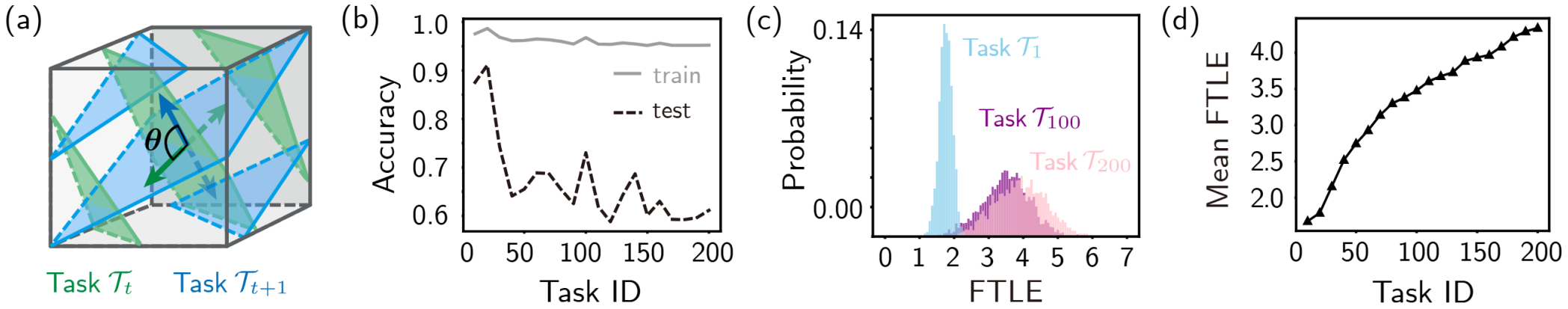
# Type-2 LoP : Over-stretched Boundaries and Chaotic Behaviors




$$\log \alpha = \lambda^{k,k'}_{\text{trained}} - \lambda^{k,k'}_{0} > 0 \quad \lambda^{k,k'}_{\text{trained},T} \to \infty \quad \text{as} \quad \prod_{t=1}^{T} \alpha_t \to \infty$$

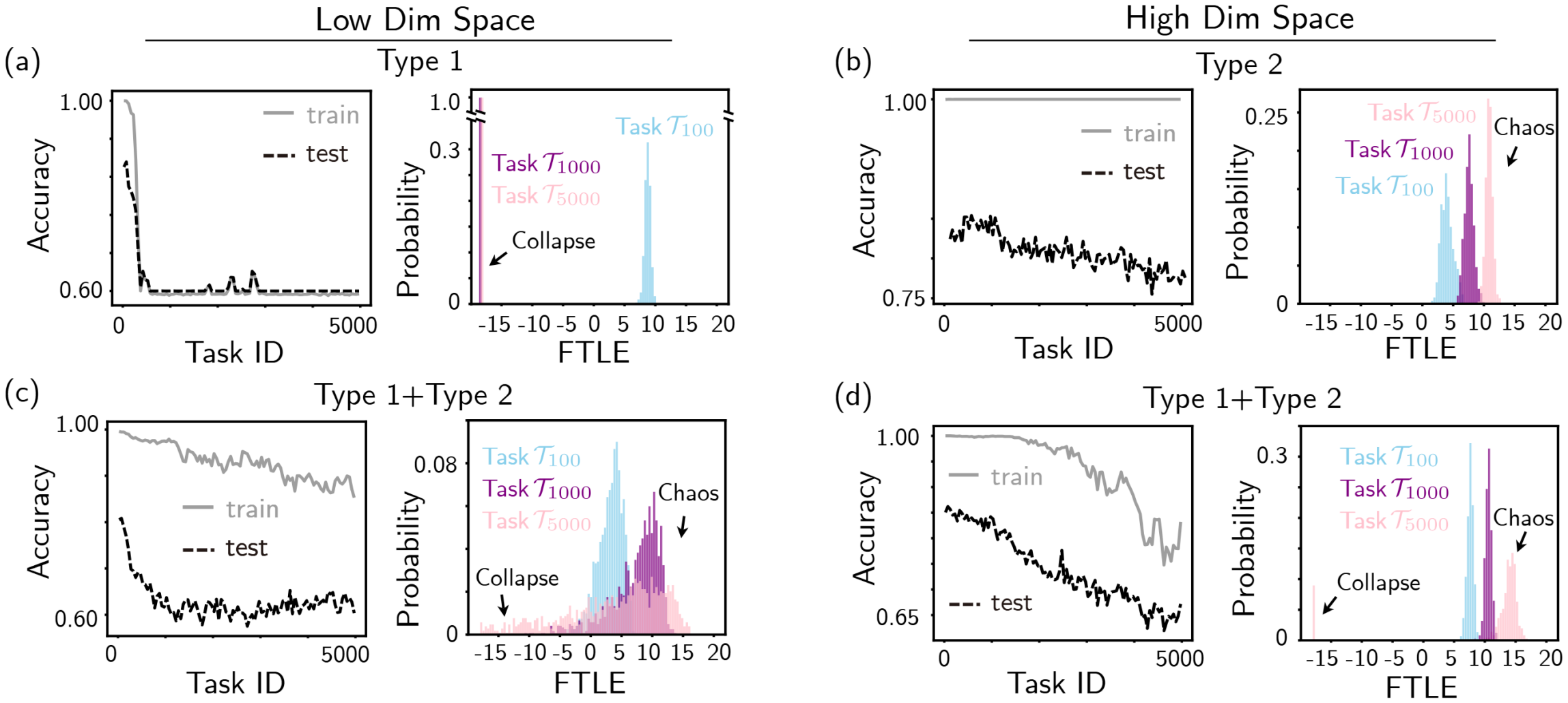
**Type-2 LoP:**

- Learning progressively causes **inter-class regions to expand**, and this expansion accumulates 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 generalization capacity**.



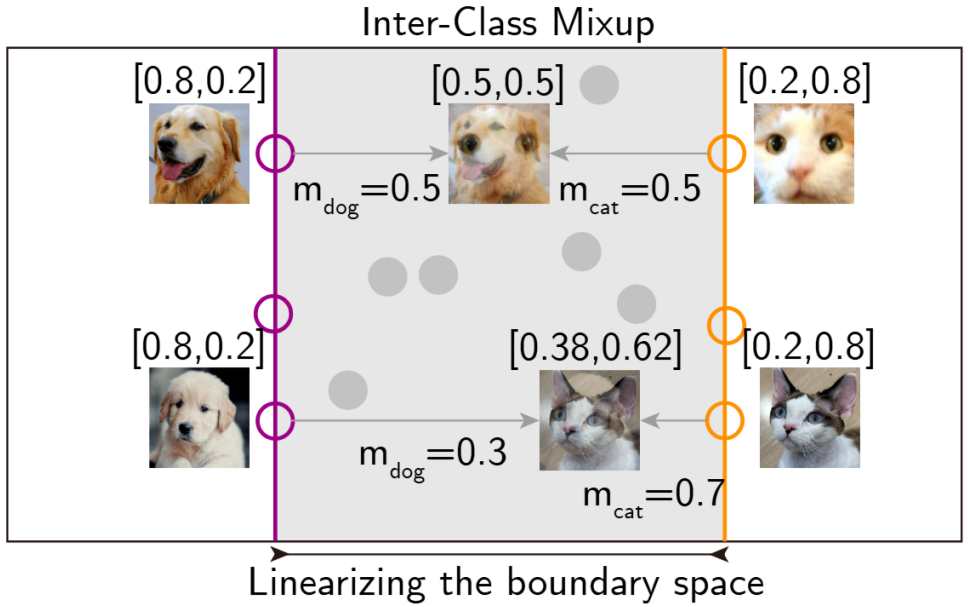

According to our theory, **Type-2 LoP** is more likely to occur in **high-dimensional representation spaces**. Thus we verified this using a **toy neural network** with a high-dimensional representation layer.

# Verifying LoP in Real Datasets



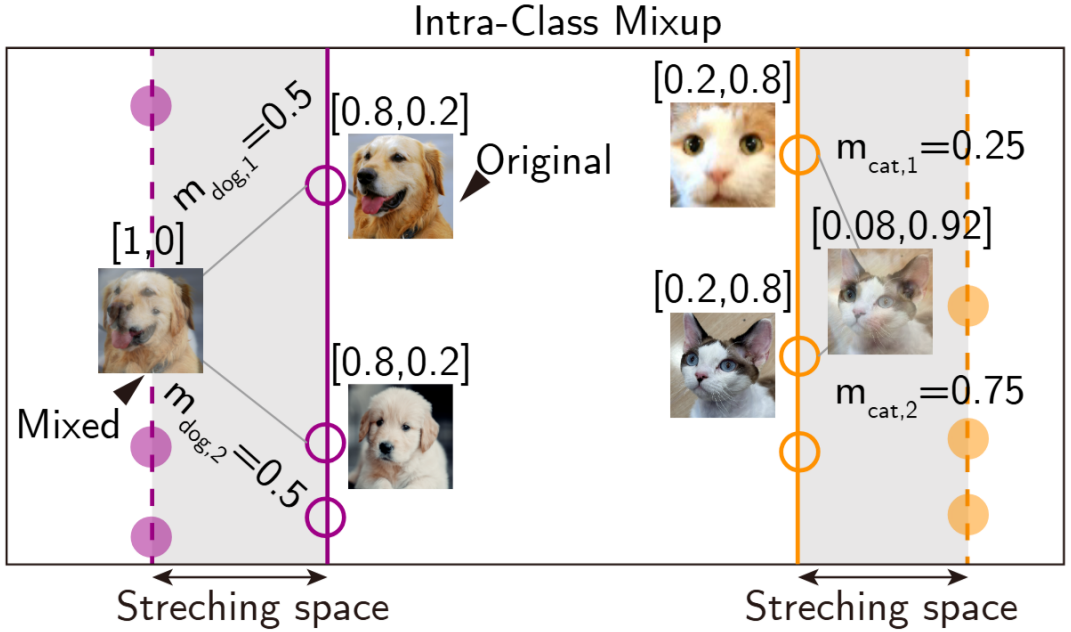

We validate our theoretical prediction on real datasets, confirming the existence of **two different types of LoP** and their **mixed forms**.

# Generalized Mixup: A Simple Method to Mitigate LoP




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

**Inter-class Mixup**, akin to the classical Mixup method, **mitigates Type-2 LoP** by alleviating over-expansion in representation space




$$x^m = mx_i^K + (1-m)x_j^K$$

$$y_K^m = y^K + \frac{M}{2} - M|0.5-m|, \quad y_{K'}^m = y^{K'} - \frac{M}{2} + M|0.5-m|$$

**Intra-class Mixup,** extending classical Mixup by combining samples **within the same class** using different **labels**, mitigates **Type-1 LoP** by preventing **within-class space collapse**

## **Generalized Mixup** achieves the **best performance across various continual learning benchmarks.**

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** |

Table 3: 0.25x Resnet-18 Acc. on CIFAR100

| Task (×4) | 0-1 | 1-2 | 2-3 | 3-4 | 4-5 |
|---|---|---|---|---|---|
| No Intv. | 0.927 | 0.812 | 0.754 | 0.708 | 0.661 |
| Retrained | 0.926 | 0.800 | 0.745 | 0.702 | 0.666 |
| L2 init | 0.925 | 0.788 | 0.724 | 0.682 | 0.649 |
| Layernorm | 0.851 | 0.755 | 0.723 | 0.674 | 0.641 |
| CBP | 0.923 | 0.812 | 0.760 | 0.713 | 0.678 |
| G-mixup[ours] | **0.932** | **0.825** | **0.772** | **0.732** | **0.697** |

Table 4: Resnet-18 Acc. on CIFAR100

| Task (×4) | 0-1 | 1-2 | 2-3 | 3-4 | 4-5 |
|---|---|---|---|---|---|
| No Intv. | 0.907 | 0.847 | 0.812 | 0.778 | 0.743 |
| Retrained | 0.901 | 0.842 | 0.815 | 0.788 | 0.767 |
| L2 init | 0.922 | 0.829 | 0.785 | 0.752 | 0.723 |
| Layernorm | 0.847 | 0.782 | 0.763 | 0.728 | 0.697 |
| CBP | 0.907 | 0.847 | 0.816 | 0.788 | **0.768** |
| G-mixup[ours] | **0.928** | **0.864** | **0.832** | **0.800** | **0.768** |