NeuMoSync: End-to-End Neuromodulatory Control for Plasticity and Adaptability in Continual Learning

Seyed Roozbeh Razavi Rohani[1], Khashayar Khajavi[1], Wesley Chung[2], Mandana Samiei[2], and Mo Chen[1]

1 Simon Fraser University 2 McGill University, Mila

Introduction

Continual learning poses several challenges. In this work, we focus on two: loss of plasticity and slow adaptation.

i. Loss of plasticity: The capacity of the model to acquire new knowledge diminishes (Figure 1).

ii. Slow adaptation: Ineffective forward transfer limits the model's ability to quickly learn new tasks.

Biologically inspired global modulation keeps deep nets plastic and fast-adapting across 6 continual-learning benchmarks.

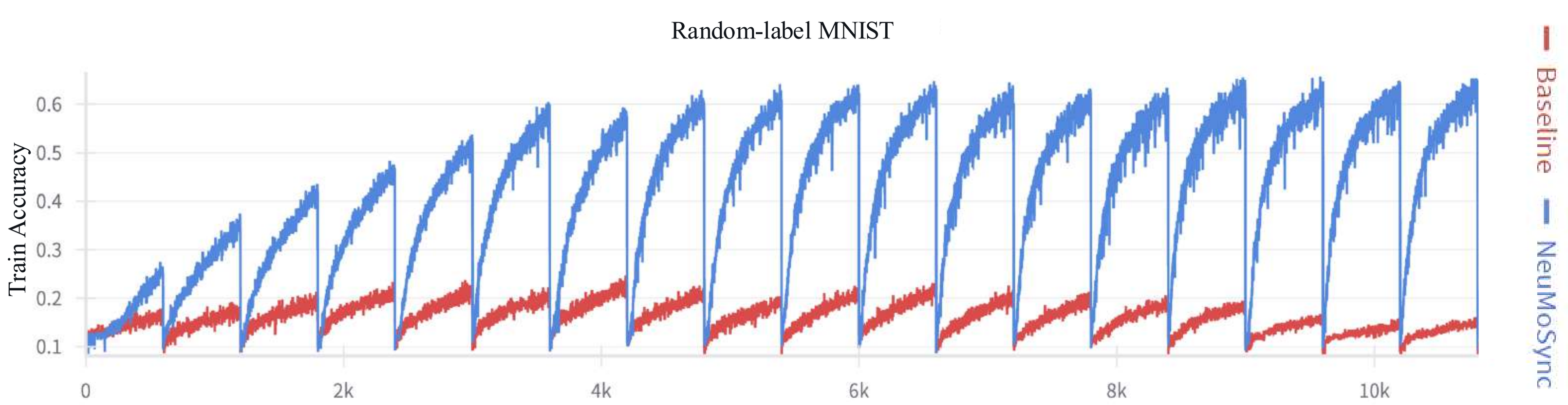

Figure 1: The baseline exhibits loss of plasticity and slow adaptation. Accuracy is reported at each training step.

Method

We aim to simulate three neuromodulator-driven mechanisms inspired by the brain: synaptic consolidation, plasticity modulation, and activity modulation.

- Each neuron is associated with a feature vector comprising positional information, activity statistics, and learnable features from the MainNetwork θ (Figure 2 A).
- We maintain a slowly updating moving average of the MainNetwork, referred to as ConsolidatedNetwork ϕ (Figure 2 A).
- The input and network state, captured by neuron features, are fed into a global modulatory network, NeuroSync, implemented using a Transformer. This module outputs four α-parameters per neuron that modulate both plasticity and activation (Figure 2 B).
- Based on these signals, the InferenceNetwork is constructed (Figure 2C), and its activation is modulated by additional generated signals (Figure 2D).

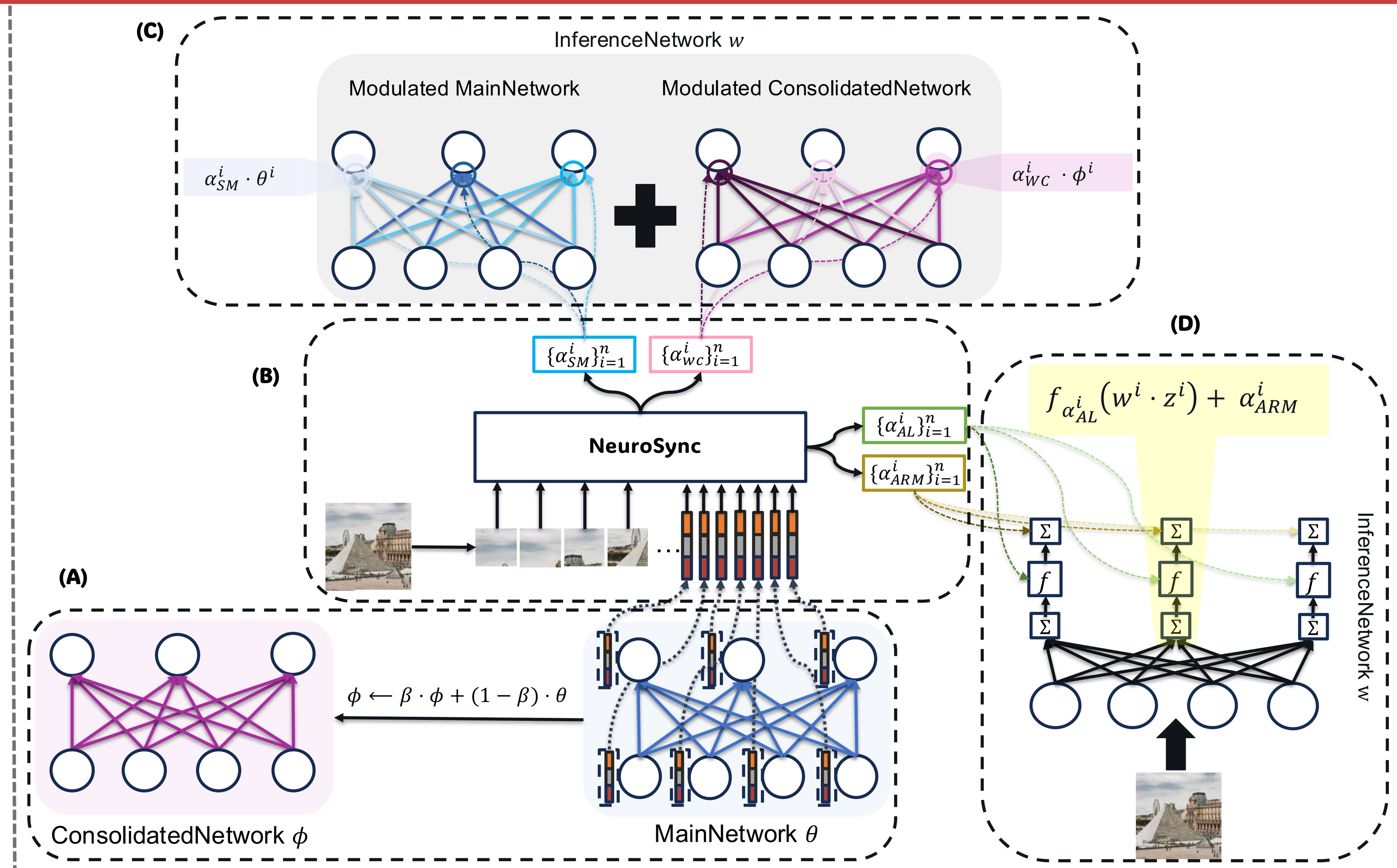

Figure 2: Overview of the architecture, illustrating key components and the flow of information.

Plasticity Preservation

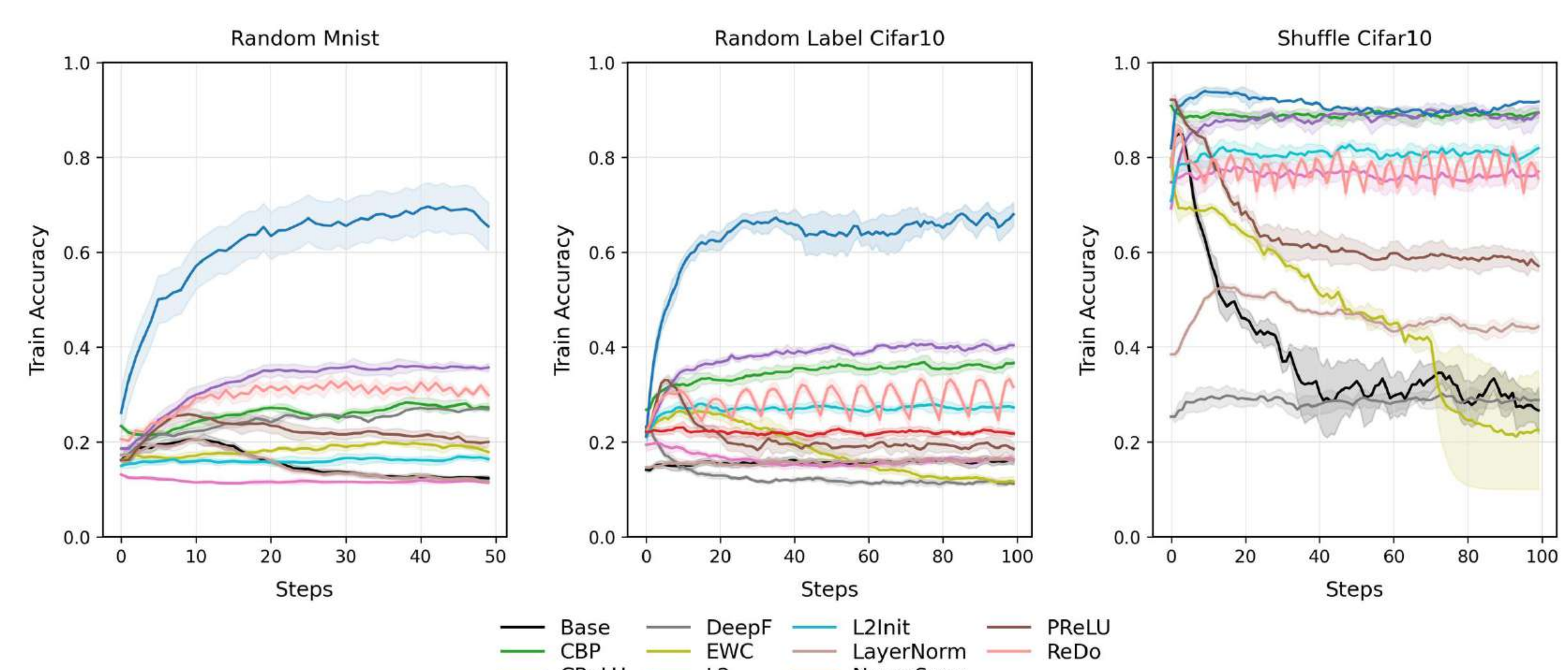

Figure 3: Performance comparison of our method against baselines on three challenging plasticity preservation tasks. Final accuracy for each task is reported.

Reoccurring Setting

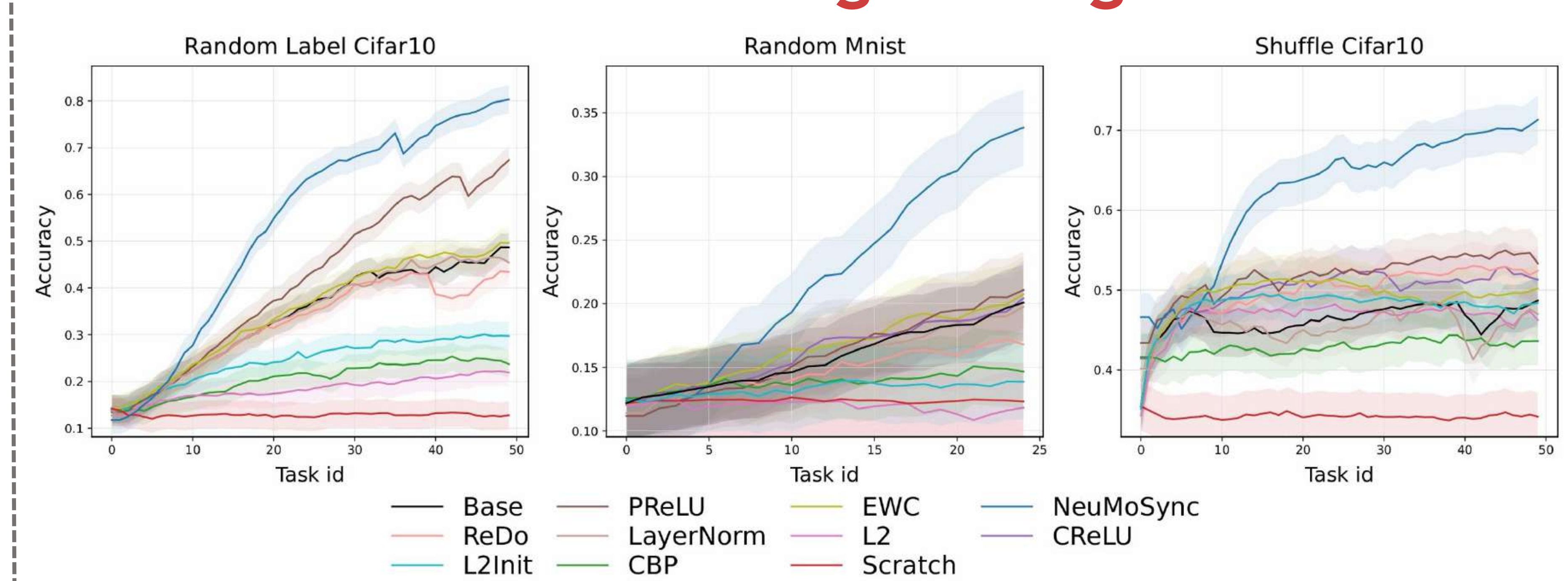

Figure 4: In the recurring setting, a fixed task is repeated every two tasks. The final accuracy on the recurring task is reported.

Metrics

$$\mathrm{LCA}_F = \frac{1}{\beta+1}\sum_{b=0}^{\beta}\left(\frac{1}{T-1}\sum_{k=1}^{T-1} A_{b,k}^{k-1}\right) \quad , \quad \mathrm{LCA}_B = \frac{1}{\beta+1}\sum_{b=0}^{\beta}\left(\frac{1}{T-1}\sum_{k=0}^{T-2} A_{b,k}^{T-1}\right).$$

$$\mathrm{FKT}_b = \frac{1}{T-1}\sum_{k=1}^{T-1}(A_{b,k}^{k-1} - A_{b,k}^{0}), \quad \mathrm{BKT}_b = \frac{1}{T-1}\sum_{k=0}^{T-2}(A_{b,k}^{T-1} - A_{b,k}^{0})$$

$$\mathrm{LCA}_B = \mathrm{LCA}_0 + \frac{1}{\beta+1}\sum_{b=0}^{\beta}\mathrm{BKT}_b, \quad \mathrm{LCA}_F = \mathrm{LCA}_0 + \frac{1}{\beta+1}\sum_{b=0}^{\beta}\mathrm{FKT}_b.$$

Adaptation Speed & Knowledge Transfer

Table 1: NeuMoSync outperforms other methods in adaptation speed, enabled by more effective knowledge transfer.

Method	Backward Knowledge Transfer			Forward Knowledge Transfer			Learning Speed			Average
	$\mathrm{BKT}_{\beta/2}$	BKT_{β}	BKT_{mean}	$\mathrm{FKT}_{\beta/2}$	FKT_{β}	FKT_{mean}	LCA_0	LCA_F	LCA_B	Final Accuracy
Random Label CIFAR-10 (Memorization)										
CBP	0.00 (0.19)	2.00 (0.53)	0.77 (0.41)	2.00 (0.72)	3.00 (1.24)	1.60 (1.18)	**12.13 (1.39)**	14.23 (0.59)	12.30 (0.92)	38.21 (0.24)
CReLU	3.00 (0.53)	5.00 (0.72)	3.33 (1.01)	2.00 (0.53)	4.00 (0.70)	2.13 (1.15)	11.87 (1.04)	14.42 (0.72)	15.20 (1.09)	35.03 (0.25)
ReDo	0.00 (0.21)	0.04 (0.01)	0.43 (1.21)	0.00 (0.05)	0.00 (0.21)	0.00 (0.20)	11.87 (1.46)	12.19 (0.84)	11.33 (0.59)	27.14 (0.87)
L2Init + EWC	1.00 (0.05)	2.00 (1.35)	1.50 (1.10)	0.00 (0.21)	1.00 (0.36)	0.17 (0.80)	11.81 (1.47)	12.58 (0.88)	13.47 (1.05)	25.14 (0.32)
NeuMoSync	**7.00 (1.33)**	**13.00 (1.12)**	**8.10 (1.36)**	**5.00 (1.08)**	**8.00 (1.20)**	**4.77 (0.55)**	10.68 (0.73)	**16.06 (0.79)**	**18.90 (0.58)**	**54.74 (1.96)**
Random Label MNIST (Memorization)										
CBP	0.00 (0.15)	1.00 (0.05)	0.50 (0.42)	1.00 (0.32)	1.00 (0.18)	0.70 (0.56)	10.65 (0.71)	11.68 (0.77)	10.73 (1.44)	36.41 (1.25)
CReLU	2.00 (1.15)	3.00 (1.11)	1.80 (0.67)	2.00 (1.23)	3.00 (0.66)	1.47 (0.28)	10.74 (1.49)	12.65 (1.14)	12.53 (1.06)	32.47 (0.81)
ReDo	0.00 (0.03)	0.00 (0.26)	0.33 (1.16)	0.00 (0.13)	0.00 (0.05)	0.00 (0.47)	**10.74 (0.77)**	10.97 (0.71)	10.00 (1.44)	27.24 (0.74)
L2Init + EWC	1.00 (0.44)	1.00 (0.16)	0.83 (0.98)	0.00 (0.05)	0.00 (0.13)	0.00 (0.06)	10.03 (0.76)	10.55 (0.75)	11.10 (1.06)	17.49 (0.12)
NeuMoSync	**3.00 (0.76)**	**7.00 (1.08)**	**8.53 (1.40)**	**4.00 (0.90)**	**6.00 (0.72)**	**7.17 (1.50)**	10.13 (1.01)	**18.81 (0.59)**	**19.87 (0.55)**	**58.67 (1.42)**
Shuffle CIFAR-10 (Concept Drift)										
CBP	0.00 (0.16)	0.00 (0.19)	0.36 (1.19)	0.00 (0.26)	0.00 (0.10)	0.00 (0.05)	**37.17 (1.50)**	37.72 (1.03)	38.67 (1.47)	90.14 (0.47)
CReLU	14.00 (1.36)	**17.00 (0.51)**	12.10 (1.22)	6.00 (1.18)	12.00 (1.04)	8.22 (0.77)	31.12 (1.14)	**39.93 (0.61)**	43.29 (0.93)	89.41 (0.34)
ReDo	12.00 (0.95)	13.00 (1.45)	10.44 (1.38)	2.00 (0.76)	3.00 (1.00)	1.44 (0.13)	32.17 (1.41)	33.88 (1.37)	42.71 (0.80)	77.36 (0.57)
L2Init + EWC	5.00 (1.14)	7.00 (1.11)	4.53 (0.65)	3.00 (1.26)	4.00 (1.04)	2.92 (1.28)	32.05 (1.03)	35.46 (0.50)	36.75 (0.82)	69.65 (0.24)
NeuMoSync	**14.90 (0.52)**	16.00 (1.43)	**13.24 (1.38)**	**4.00 (1.33)**	**12.37 (0.81)**	**8.92 (0.56)**	32.02 (1.38)	**40.17 (1.45)**	**45.90 (0.59)**	**92.84 (0.36)**
Permuted MNIST (Domain Incremental)										
CBP	17.00 (0.99)	14.00 (0.57)	13.71 (1.26)	**14.00 (1.27)**	18.00 (0.63)	17.14 (0.98)	20.50 (1.05)	38.75 (0.77)	35.29 (1.37)	69.21 (0.27)
CReLU	23.00 (0.92)	20.00 (0.71)	18.86 (1.04)	4.00 (1.23)	7.00 (0.70)	5.57 (0.81)	**33.00 (1.50)**	39.38 (1.15)	53.00 (0.94)	69.38 (0.31)
ReDo	**25.00 (1.02)**	26.00 (0.62)	20.71 (0.72)	13.00 (0.84)	14.00 (1.09)	12.43 (0.73)	12.12 (0.72)	25.50 (0.57)	36.29 (1.13)	35.71 (0.38)
L2Init + EWC	24.00 (0.73)	24.00 (1.41)	19.86 (1.36)	12.00 (0.57)	13.00 (0.74)	12.00 (1.17)	12.12 (0.71)	25.00 (0.63)	35.57 (1.44)	34.24 (0.42)
NeuMoSync	22.00 (1.07)	**28.00 (0.97)**	**24.57 (1.28)**	12.00 (1.31)	**19.00 (0.69)**	**17.71 (0.60)**	30.00 (0.93)	**47.13 (0.92)**	**55.71 (0.97)**	**74.36 (1.36)**
Class Split CIFAR-100 (Class Incremental)										
CBP	**12.00 (1.23)**	10.00 (1.17)	**13.86** (1.48)	**8.00 (0.60)**	5.00 (0.90)	7.52 (0.84)	51.09 (1.36)	59.08 (0.75)	**63.57 (0.69)**	72.36 (0.67)
ReDo	2.00 (1.50)	2.00 (1.34)	2.46 (1.47)	0.00 (0.19)	0.00 (0.27)	0.00 (0.25)	**56.20 (0.99)**	58.19 (0.71)	56.06 (0.90)	72.51 (0.75)
L2Init + EWC	3.00 (0.56)	3.00 (0.88)	3.94 (1.49)	2.00 (0.77)	3.00 (1.28)	2.10 (0.96)	55.77 (0.92)	57.52 (1.46)	57.67 (1.50)	68.24 (0.31)
NeuMoSync	9.00 (1.06)	**19.00 (1.22)**	11.95 (0.65)	7.00 (0.80)	**17.00 (1.47)**	**8.19 (1.08)**	51.69 (1.04)	**60.77 (1.25)**	62.43 (0.56)	**78.68 (0.91)**
Class Split T-ImageNet (Class Incremental)										
CBP	4.00 (1.08)	3.00 (1.00)	5.27 (1.35)	2.00 (0.66)	1.00 (0.48)	1.88 (0.58)	70.30 (0.69)	**72.37 (1.10)**	**76.88 (1.18)**	83.21 (0.57)
CReLU	1.00 (0.35)	5.00 (0.18)	3.00 (1.34)	**3.00 (0.75)**	**4.00 (1.09)**	2.54 (1.12)	69.07 (0.92)	71.74 (1.08)	72.00 (0.78)	84.27 (1.02)
ReDo	5.00 (1.43)	4.00 (0.70)	6.88 (1.22)	2.00 (0.74)	2.00 (0.90)	2.19 (1.17)	69.33 (0.80)	70.52 (0.82)	74.69 (1.25)	83.26 (1.32)
L2Init + EWC	1.00 (0.04)	1.00 (0.23)	3.54 (1.50)	0.00 (0.14)	0.00 (0.11)	0.08 (0.32)	**72.48 (0.77)**	71.56 (1.43)	75.77 (1.38)	82.76 (1.27)
NeuMoSync	**7.00 (1.38)**	**6.00 (0.87)**	**8.23 (0.66)**	2.00 (1.33)	3.57 (1.20)	**3.27 (1.11)**	67.96 (1.49)	70.44 (1.15)	**76.08** (0.51)	**86.37 (1.06)**

Table 2: NeuMoSync shows a substantial performance advantage over MAML in memorization tasks.

Method	Permuted MNIST		Random Label MNIST		Random Label CIFAR-10		Shuffle CIFAR-10		Class Split CIFAR-100		Class Split T-ImageNet	
	LCA_F	Acc	LCA_F	Acc	LCA_F	Acc	LCA_F	Acc	LCA_F	Acc	LCA_F	Acc
MAML	0.24	0.60	0.12	0.20	0.12	0.21	0.41	**0.92**	0.47	0.75	**0.64**	**0.87**
NeuMoSync	**0.26**	**0.74**	**0.25**	**0.58**	**0.21**	**0.54**	**0.51**	**0.92**	**0.57**	**0.78**	0.62	0.86

Ablation Study

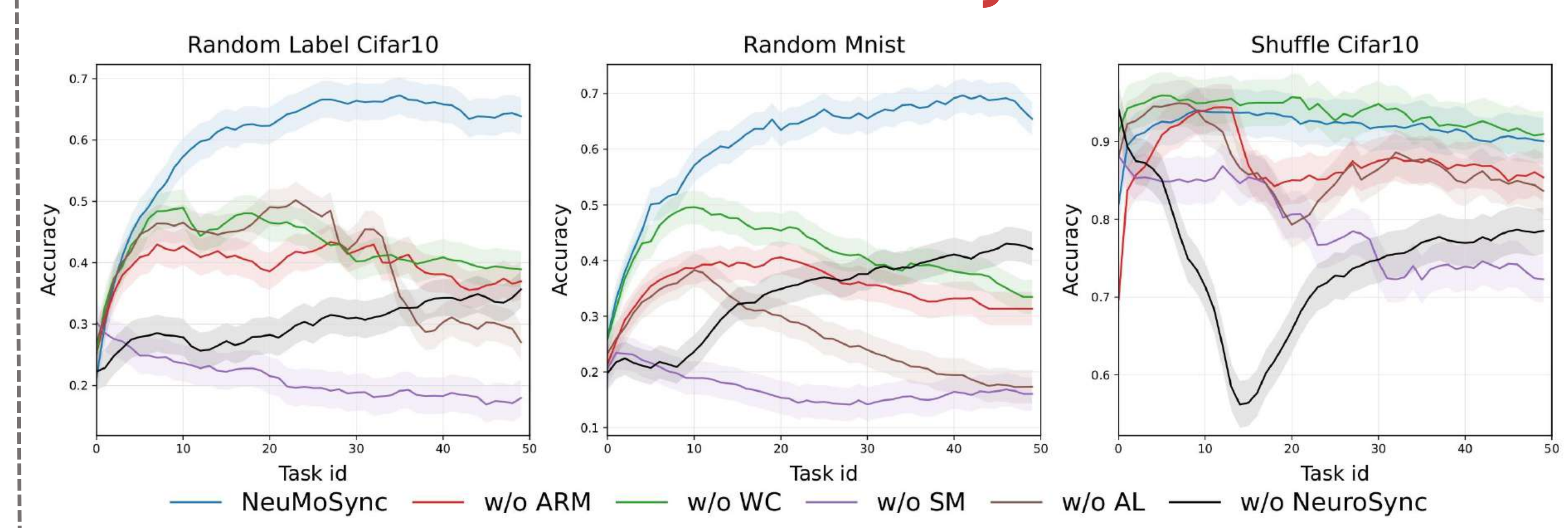

Figure 5: Removing any of the mechanisms or the NeuroSync module significantly impairs the plasticity preservation of our method.