MePo: Meta Post-Refinement for Rehearsal-Free General Continual Learning

Guanglong Sun[1], Hongwei Yan[1], Liyuan Wang[2,#], Zhiqi Kang[3], Shuang Cui[4], Hang Su[2], Jun Zhu[2], Yi Zhong[1,#]

[1] School of Life Sciences, IDG/McGovern Institute for Brain Research, Tsinghua University, Tsinghua-Peking Center for Life Sciences, Beijing, China
[2] Department of Computer Science and Technology, Institute for AI, BNRist Center, Tsinghua-Bosch Joint ML Center, THBI Lab, Tsinghua University, Beijing, China
[3] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, Grenoble, France
[4] Institute of Software Chinese Academy of Sciences, Beijing, China
[#] Corresponding author.
Email: sgl23@mails.Tsinghua.edu.cn

Core idea: meta training the pretrained model before deployment, then continually adapt with a stable reference geometry.

1. Introduction

General continual learning (GCL): real-world scenarios continual learning with **online data streams** and **blurry task boundaries**, making it increasingly difficult for AI models to rapidly capture and effectively balance successive information[1,2].

Pretrained models (PTMs) in CL: parameter-efficient tuning (PET) for representation learning, and recover old task distributions in the representation space for output alignment[3].

Biologic inspiration: (1) **meta-plasticity**[4]: biological brain retain substantial "pre-trained knowledge", positioning them in a critical state of neuro dynamics for rapid adaptation; (2) **reconstructive memory**[5]: the neural representations of incoming memories are continually encoded into and reconstructed from a shared representation space that enables real-time generalization.

2. Contribution

(1) Meta Rep. learns a GCL-ready initialization.

(2) Meta Cov. stabilizes feature geometry online.

Strength: Plug-and-play, reusable, generalizable, and no extra computational overhead.

3. Method Overview

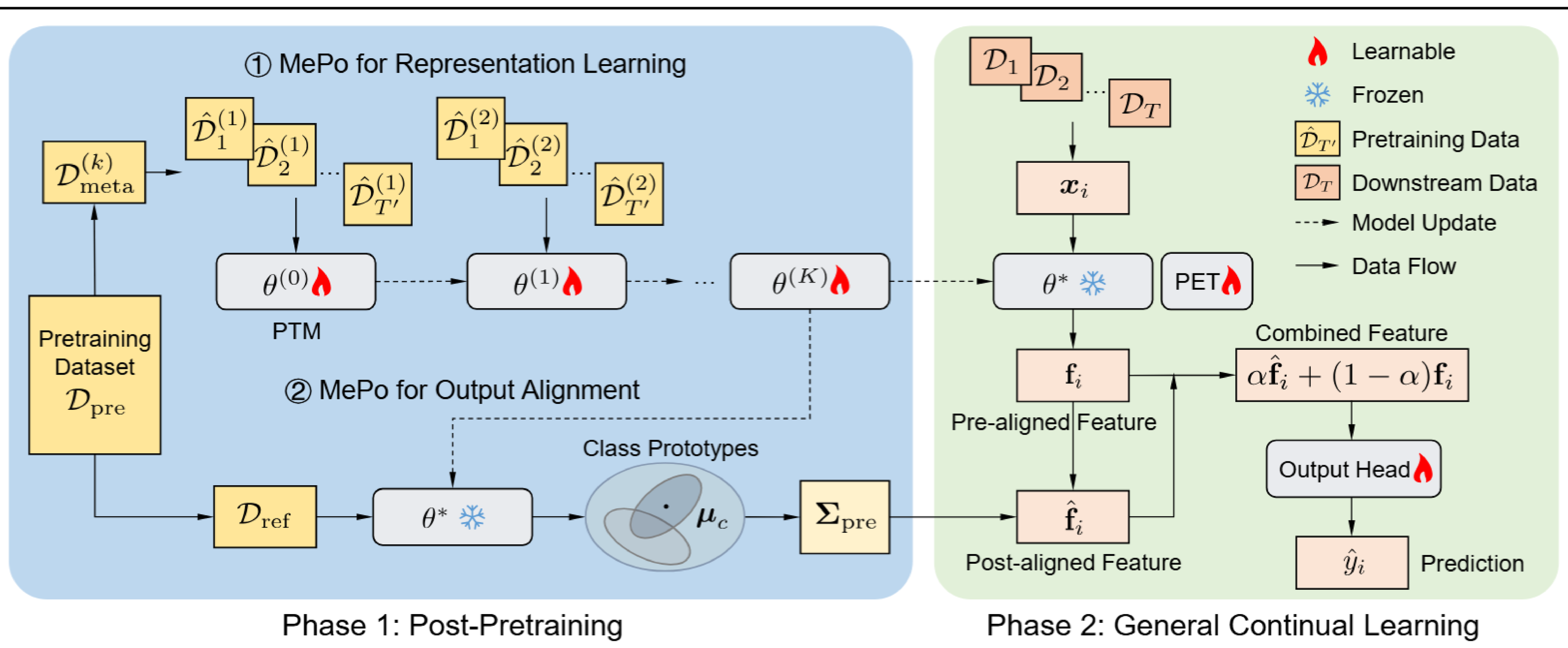

Fig. 1 The proposed MePo framework.

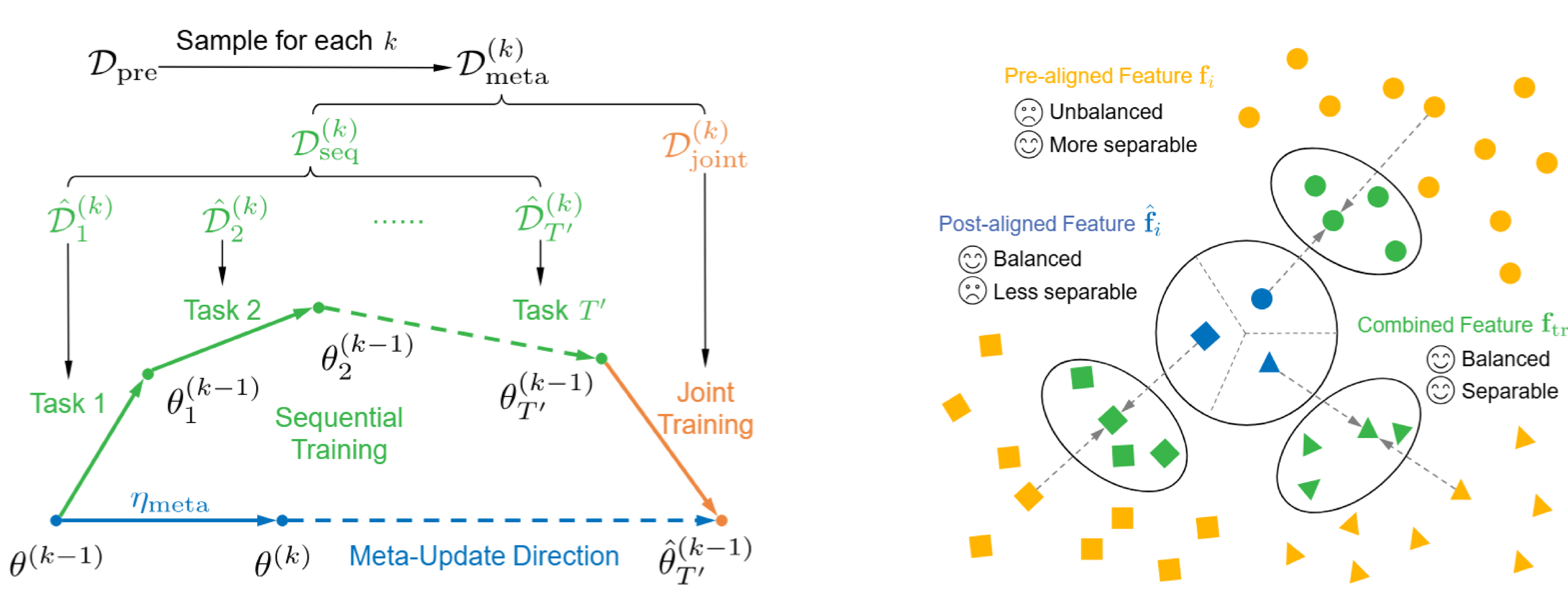

Fig. 2 MePo representation learning (left) and feature alignment (right).

4. Main Results

PTM	Method	CIFAR-100 $A_{AUC}(\uparrow)$	CIFAR-100 $A_{Last}(\uparrow)$	ImageNet-R $A_{AUC}(\uparrow)$	ImageNet-R $A_{Last}(\uparrow)$	CUB-200 $A_{AUC}(\uparrow)$	CUB-200 $A_{Last}(\uparrow)$
Sup-21K	Seq FT	$19.71_{\pm 3.39}$	$10.42_{\pm 4.92}$	$7.51_{\pm 3.94}$	$2.29_{\pm 0.85}$	$3.47_{\pm 0.41}$	$1.49_{\pm 0.42}$
	Linear Probe	$49.69_{\pm 6.09}$	$23.07_{\pm 7.33}$	$29.24_{\pm 1.26}$	$16.87_{\pm 3.14}$	$28.96_{\pm 2.46}$	$17.33_{\pm 3.08}$
	Seq FT (SL) [38]	$64.90_{\pm 7.18}$	$62.06_{\pm 1.89}$	$47.20_{\pm 1.47}$	$39.60_{\pm 2.43}$	$56.16_{\pm 4.32}$	$56.50_{\pm 3.08}$
	CODA-P [25]	$78.81_{\pm 3.38}$	$80.30_{\pm 1.58}$	$50.11_{\pm 2.14}$	$46.17_{\pm 2.00}$	$64.96_{\pm 3.30}$	$59.28_{\pm 3.14}$
	Deep L2P [33]	$78.12_{\pm 0.61}$	$77.73_{\pm 1.09}$	$42.39_{\pm 0.23}$	$38.16_{\pm 1.37}$	$60.95_{\pm 1.22}$	$56.31_{\pm 2.53}$
	w/ MePo (Ours)	$\mathbf{83.63}_{\pm 0.61}$	$83.98_{\pm 0.29}$	$\mathbf{58.71}_{\pm 1.28}$	$\mathbf{55.13}_{\pm 1.16}$	$64.92_{\pm 1.47}$	$63.30_{\pm 1.52}$
	DualPrompt [32]	$66.36_{\pm 4.42}$	$58.09_{\pm 4.40}$	$38.63_{\pm 2.19}$	$30.71_{\pm 0.82}$	$55.73_{\pm 2.77}$	$47.08_{\pm 4.94}$
	w/ MePo (Ours)	$71.37_{\pm 4.07}$	$66.48_{\pm 2.82}$	$44.65_{\pm 2.09}$	$36.76_{\pm 1.21}$	$58.36_{\pm 2.59}$	$52.16_{\pm 3.74}$
	MVP [19]	$68.13_{\pm 4.34}$	$60.56_{\pm 2.57}$	$41.50_{\pm 1.15}$	$34.14_{\pm 3.95}$	$56.78_{\pm 2.88}$	$50.25_{\pm 3.53}$
	w/ MePo (Ours)	$72.18_{\pm 4.50}$	$68.45_{\pm 1.59}$	46.35±1.31	38.21±3.66	58.73±3.31	52.22±2.80
	MISA [12]	$80.35_{\pm 2.39}$	$80.75_{\pm 1.24}$	$51.52_{\pm 2.09}$	$45.08_{\pm 1.43}$	$65.40_{\pm 3.01}$	$60.20_{\pm 1.82}$
	w/ MePo (Ours)	$82.30_{\pm 2.83}$	$\mathbf{83.99}_{\pm 1.35}$	$54.86_{\pm 2.20}$	$49.18_{\pm 1.38}$	$\mathbf{68.13}_{\pm 3.17}$	$\mathbf{64.75}_{\pm 1.00}$
Sup-21/1K	Deep L2P [33]	$69.15_{\pm 1.66}$	$68.57_{\pm 1.38}$	$42.74_{\pm 0.83}$	$39.22_{\pm 2.14}$	$39.20_{\pm 1.69}$	$46.76_{\pm 1.87}$
	w/ MePo (Ours)	$\mathbf{78.75}_{\pm 1.18}$	$\mathbf{77.52}_{\pm 1.03}$	$62.71_{\pm 1.09}$	$58.91_{\pm 0.08}$	$48.36_{\pm 1.88}$	$50.88_{\pm 2.85}$
	DualPrompt [32]	$64.84_{\pm 2.62}$	$67.22_{\pm 8.54}$	$49.52_{\pm 2.92}$	$47.14_{\pm 3.39}$	$43.96_{\pm 2.00}$	$41.20_{\pm 7.61}$
	w/ MePo (Ours)	$67.18_{\pm 4.48}$	$57.95_{\pm 3.69}$	$54.75_{\pm 1.66}$	$44.75_{\pm 0.74}$	$47.06_{\pm 3.19}$	$38.24_{\pm 9.29}$
	MVP [19]	$65.26_{\pm 3.87}$	$53.66_{\pm 5.61}$	$51.26_{\pm 1.47}$	$41.41_{\pm 4.81}$	$45.12_{\pm 3.08}$	$37.95_{\pm 9.32}$
	w/ MePo (Ours)	$70.25_{\pm 4.23}$	$62.05_{\pm 2.39}$	$61.28_{\pm 1.21}$	$50.82_{\pm 3.70}$	$49.72_{\pm 3.53}$	$42.81_{\pm 6.74}$
	MISA [12]	$62.91_{\pm 7.96}$	$67.99_{\pm 7.41}$	$50.87_{\pm 1.69}$	$47.75_{\pm 2.87}$	$42.76_{\pm 2.33}$	$44.05_{\pm 1.94}$
	w/ MePo (Ours)	$78.01_{\pm 3.09}$	$76.73_{\pm 1.06}$	$\mathbf{64.23}_{\pm 1.30}$	$\mathbf{58.20}_{\pm 0.51}$	$\mathbf{55.31}_{\pm 4.52}$	$\mathbf{56.58}_{\pm 2.33}$

5. Computational overheads

Phase 1: meta post-training

Training	Image Size	Epochs	Batch Size	Total Steps	Images Processed
Sup-21K (Pretraining) (Dosovitskiy et al., 2020a)	224×224	90	4096	∼310k	∼1.3B
Sup-21K (MePo Post-Refinement)	224×224	50	256	∼12.8k	∼2M

Meta epoch	1	2	3	4	5	Average
Time (mins)	14.43	14.20	14.13	14.00	13.96	14.14

Phase 2: general continual learning

Method	+Param.	+Ratio	Time	Accuracy
MVP [19]	639k	0.74%	5.34s	41.50
w/ MePo	1215k	1.41%	5.34s	46.35
MISA [12]	637k	0.74%	4.84s	51.52
w/ MePo	1213k	1.41%	4.84s	54.86

6. Analysis: MePo obtains more separable and sparse representations.

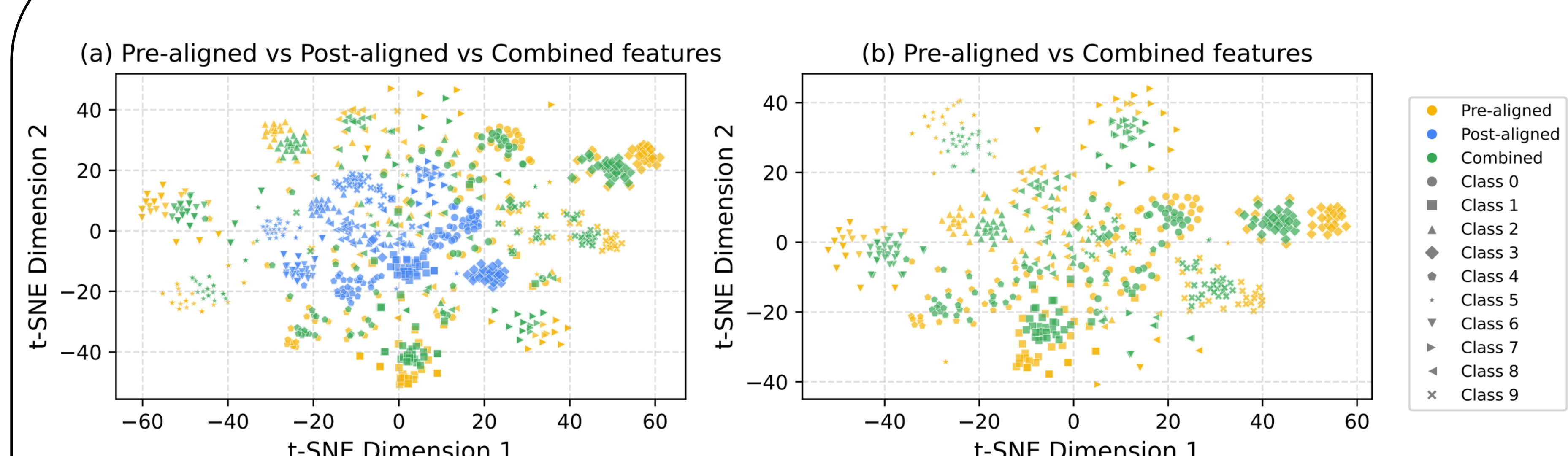

Fig. 4 Visualization of pre-aligned, post-aligned, and combined features with t-SNE. Here we take the setup of MISA w/ MePo, ImageNet-R, and Sup-21/1K as an example.

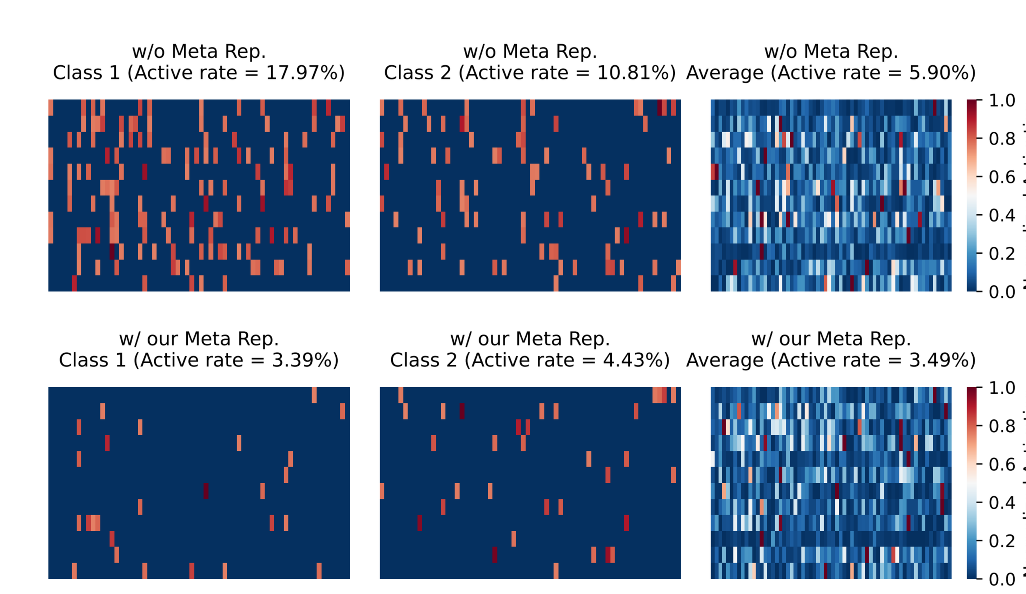

Fig. 5 Visualization of class-wise prototypes on ImageNet-R under Sup-21K.