NEURAL INFORMATION PROCESSING SYSTEMS

INSTITUTE OF COMPUTING TECHNOLOGY,CAS

University of Chinese Academy of Sciences

Dynamical Adapter Fusion: Constructing A Global Adapter for Pre-Trained Model-based Class-Incremental Learning

Ruiqi Liu[1,2], Boyu Diao[1,2], Zijia An[1,2], Runjie Shao[1,2], Feixiang Ge[1,2], Zhulin An[1,2],

Fei Wang[1,2] Yongjun Xu[1,2]

1 Institute of Computing Technology, Chinese Academy of Sciences
2 University of Chinese Academy of Sciences

➢ Introduction

■ Class-incremental Learning

In class-incremental learning, models must learn new classes without forgetting old ones. Training a separate adapter for each task reduces interference between tasks, but creates two practical challenges.

■ Challenges

- Growing adapter storage
- Costly, error-prone adapter retrieval

中国科学院大学 University of Chinese Academy of Sciences
ICT 中科院计算所 INSTITUTE OF COMPUTING TECHNOLOGY,CAS
NEURAL INFORMATION PROCESSING SYSTEMS

➢ Related work & Challenges

■ Catastrophic Forgetting

- Regularization-based methods constrain the updates to essential parameters of the model, thereby preserving important knowledge across tasks.
- Architecture-based methods introduce task-specific components or allocate new parameters for each new task.
- Rehearsal-based methods continuously replay samples from previous tasks to ensure that knowledge is retained.

■ From Scratch to Pre-Trained Models

- Training from scratch is computationally demanding, which makes it difficult to scale under limited resources. Parameter-efficient fine-tuning reuses visual features from a frozen pre-trained backbone and updates only lightweight adapters, reducing the number of trainable parameters and the cost of learning new classes.

中国科学院大学 University of Chinese Academy of Sciences
ICT 中科院计算所 INSTITUTE OF COMPUTING TECHNOLOGY,CAS
NEURAL INFORMATION PROCESSING SYSTEMS

➢ Related work & Challenges

■ Task-Specific Adapters

A common approach freezes the pre-trained backbone and trains a lightweight adapter for each task to reduce interference between tasks.

- Growing adapter storage. Each new task adds an adapter, so storage grows with the number of tasks.
- Test-time adapter retrieval. Test inputs come without task IDs, so retrieval-based methods must select an adapter for each input. This adds latency, and selecting the wrong adapter can hurt classification accuracy.

■ Can a single adapter learn new classes and retain knowledge of earlier classes?

中国科学院大学 University of Chinese Academy of Sciences
ICT 中科院计算所 INSTITUTE OF COMPUTING TECHNOLOGY,CAS
NEURAL INFORMATION PROCESSING SYSTEMS

➢ Motivation

■ One Global Adapter for All Learned Classes

$$\theta_t^* = \underbrace{\beta_t \theta_p + \beta_t \theta_{t-1}^*}_{\text{Stability}} + \underbrace{(1 - 2\beta_t)\theta_t}_{\text{Plasticity}}$$

DAF fuses each new task adapter into a single global adapter that learns new classes while retaining prior knowledge. The same adapter serves all learned classes, eliminating test-time retrieval and keeping adapter storage constant as tasks accumulate.

中国科学院大学 University of Chinese Academy of Sciences
中科院计算所 INSTITUTE OF COMPUTING TECHNOLOGY, CAS
NEURAL INFORMATION PROCESSING SYSTEMS

➢ Method

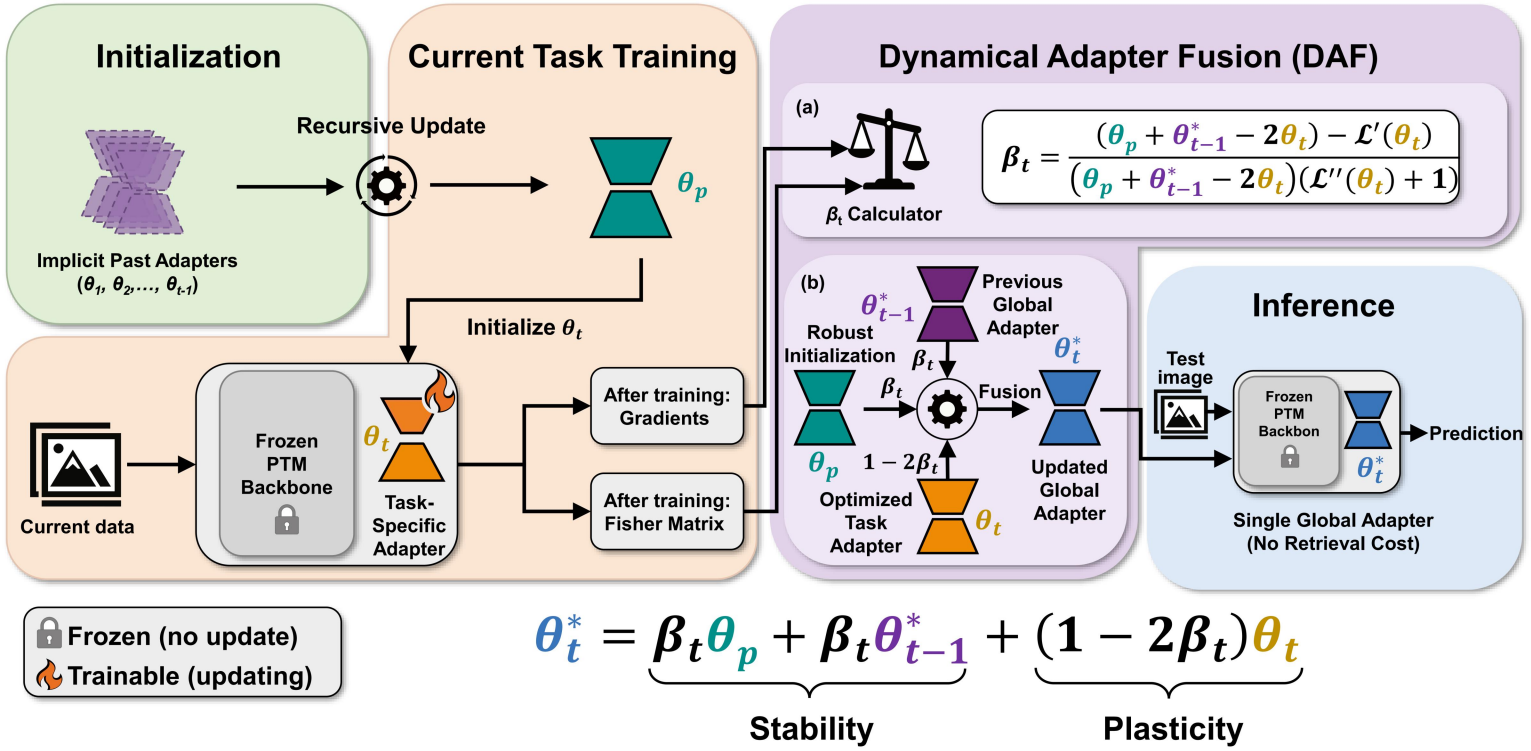
■ Dynamical Adapter Fusion (DAF)

● Robust Initialization

- We initialize each new task adapter with a running average of previously trained adapters, transferring shared knowledge without storing individual past adapters.

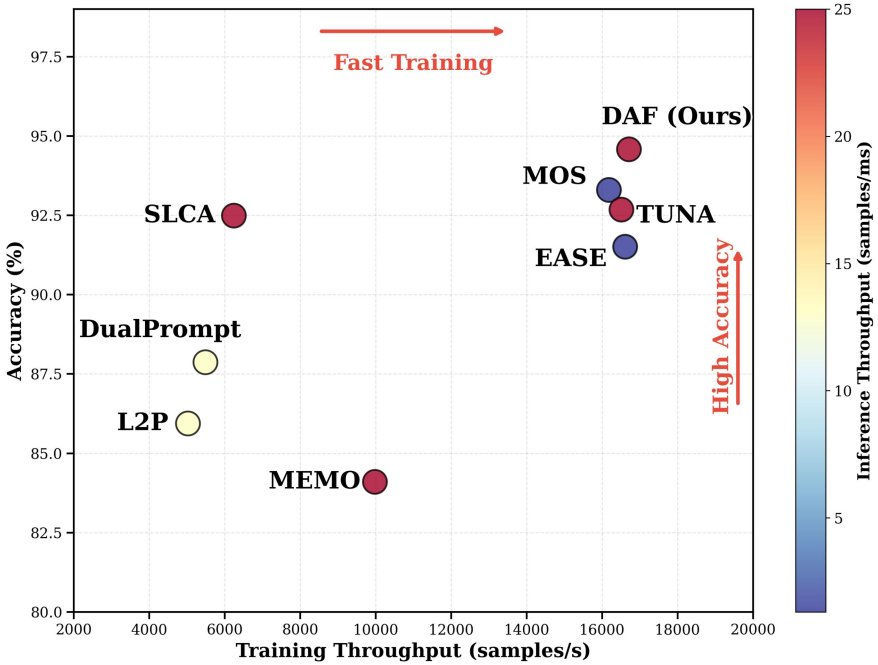
● Dynamical Fusion

- After each task, we use PAC-Bayes–guided, parameter-wise weights to fuse the initialization, previous global adapter, and newly trained task adapter.

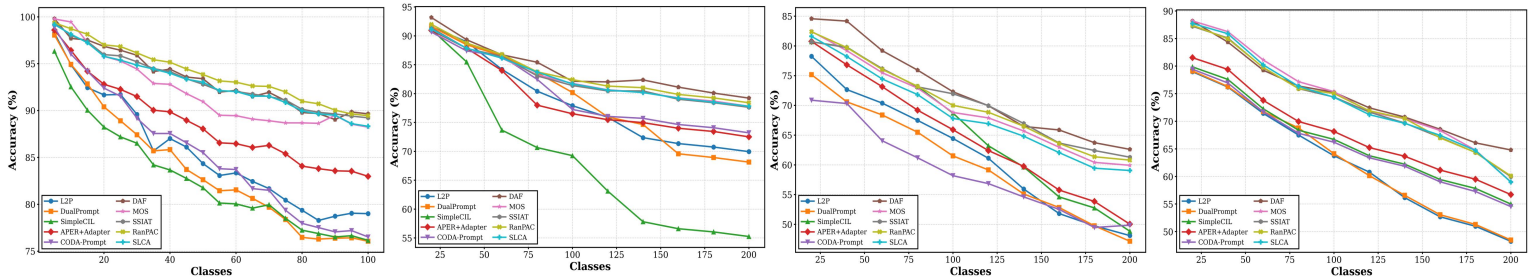


➢ Experiments

DAF achieves the highest average and final accuracy across four benchmarks, with high training and inference throughput.



Method	CIFAR-100		ImageNet-R		ImageNet-A		ObjectNet	
	$\bar{\mathcal{A}}$ (%)	$\mathcal{A}_T$ (%)	$\bar{\mathcal{A}}$ (%)	$\mathcal{A}_T$ (%)	$\bar{\mathcal{A}}$ (%)	$\mathcal{A}_T$ (%)	$\bar{\mathcal{A}}$ (%)	$\mathcal{A}_T$ (%)
Finetune	38.90	20.17	32.32	22.78	24.28	14.51	19.14	8.73
L2P [43]	85.94	79.93	75.46	69.77	49.39	41.71	63.78	52.19
DualPrompt [41]	87.87	81.15	73.10	67.18	53.71	41.67	59.27	49.33
CODA-Prompt [32]	89.11	81.96	77.97	72.27	53.54	42.73	66.07	53.29
SLCA [46]	92.49	88.55	81.17	77.00	68.66	58.74	72.55	61.30
RanPAC [24]	94.00	90.62	82.98	77.94	69.32	61.82	72.76	62.02
SSIAT [36]	93.52	90.07	83.20	78.85	70.83	62.23	73.65	62.45
SimpleCIL [50]	87.57	81.26	61.26	54.55	59.77	48.91	65.45	53.59
APER + Adapter [50]	90.65	85.15	75.82	67.95	60.47	49.37	67.18	55.24
EASE [49]	91.51	85.80	81.74	76.17	65.34	55.04	70.84	57.86
TUNA† [39]	92.68	86.19	81.24	72.92	56.14	28.64	72.61	58.52
MOS [34]	93.30	89.25	82.96	77.93	67.08	56.22	74.69	63.62
DAF (Ours)	**94.58**	**91.15**	**84.01**	**79.63**	**72.06**	**62.54**	**76.11**	**66.37**



University of Chinese Academy of Sciences

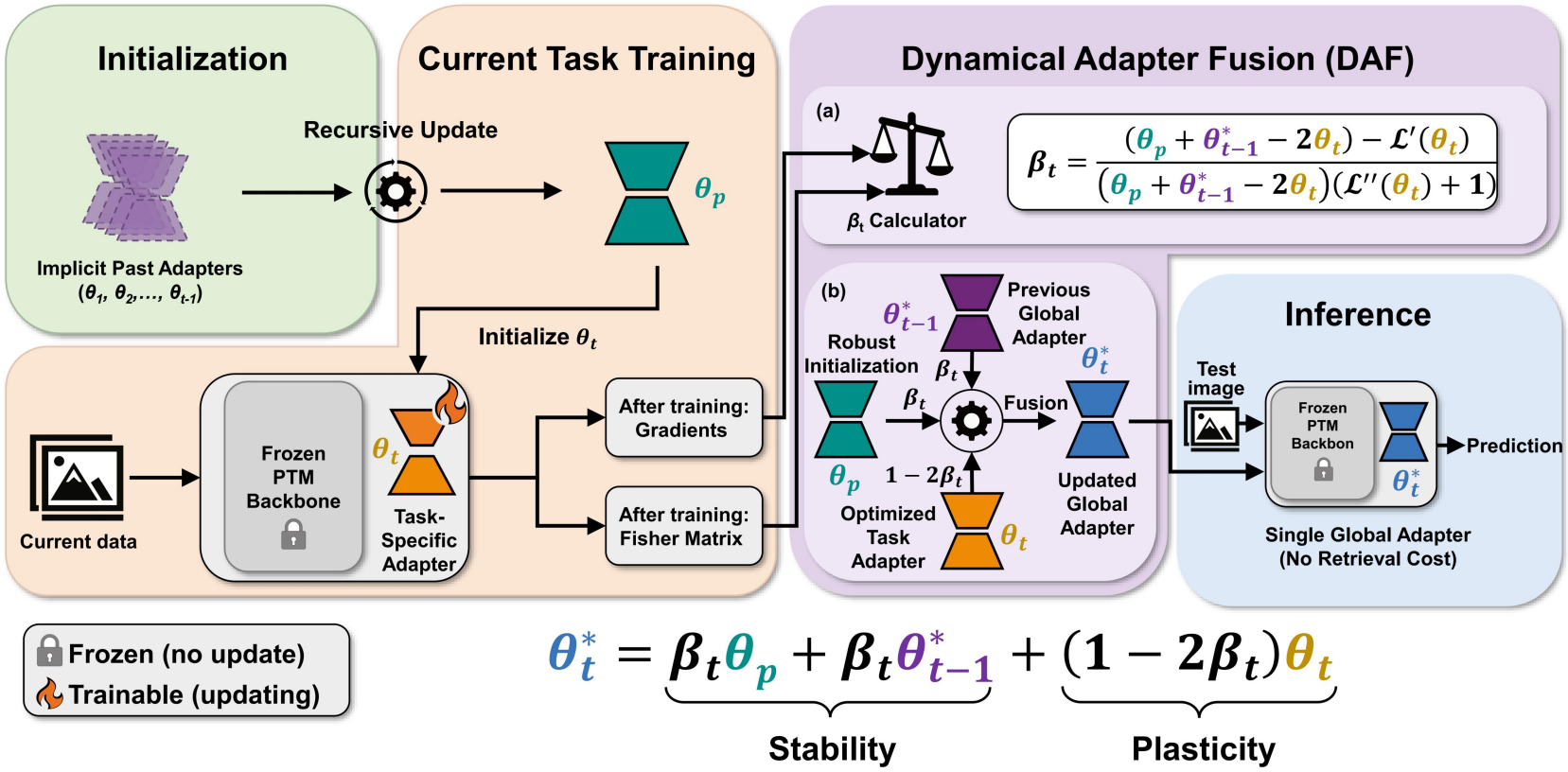
INSTITUTE OF COMPUTING TECHNOLOGY,CAS

NEURAL INFORMATION PROCESSING SYSTEMS

➢ Conclusion

■ Dynamical Adapter Fusion

- Dynamical Adapter Fusion uses robust initialization and PAC-Bayes–guided fusion to learn new classes while retaining prior knowledge in a single global adapter. DAF achieves the highest average and final accuracy across four benchmarks while keeping adapter storage constant and eliminating test-time adapter retrieval.



NEURAL INFORMATION PROCESSING SYSTEMS

中科院计算所 INSTITUTE OF COMPUTING TECHNOLOGY,CAS

中国科学院大学 University of Chinese Academy of Sciences

Dynamical Adapter Fusion: Constructing A Global Adapter for Pre-Trained Model-based Class-Incremental Learning

Thank You!

For further details, feel free to get in touch with us

liuruiqi23@mails.ucas.ac.cn

https://github.com/EMLS-ICTCAS/DAF

https://arxiv.org/abs/2601.21341