Dynamic Integration of Task-Specific Adapters for Class Incremental Learning

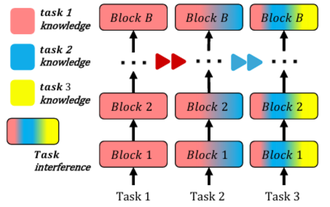
Jiashuo Li[1], Shaokun Wang[1†], Bo Qian[1], Yuhang He[2], Xing Wei[1], Qiang Wang[1], Yihong Gong[1,2†]
[1] School of Software Engineering, Xi'an Jiaotong University
[2] College of Artificial Intelligence, Xi'an Jiaotong University

CVPR Nashville JUNE 11-15, 2025

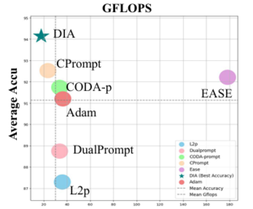
Introduction

1. Compositional Deficiency

a) Shared parameter space leads to task interference.
b) High computation costs characterize current PET-based methods.



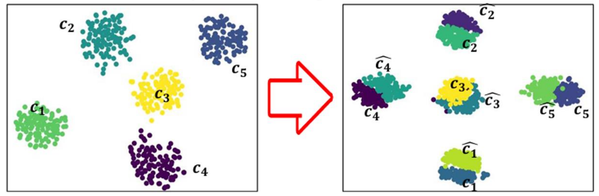
(a) Task interference



(b) GFLOPS versus Accuracy

2. Compositional Deficiency

c) Gaussian-based feature reconstruction may have a significant deviation from the actual feature distribution. c_i indicates the actual feature distribution, $\hat{c}_i$ indicates the pseudo feature distribution.

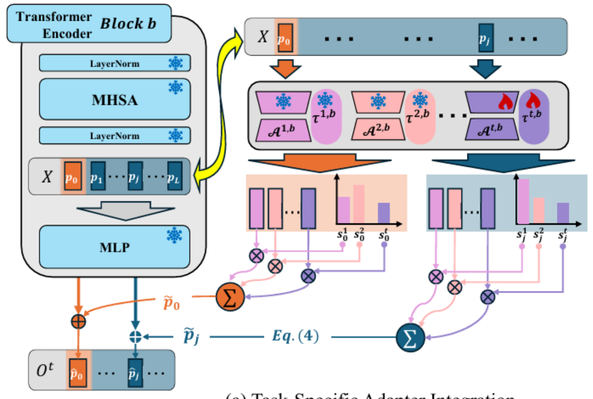


(c) Distribution Shift: Task 1 → Task t

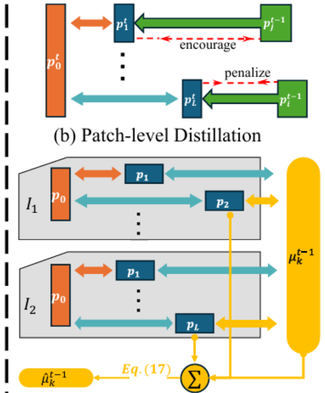
Methodology

We introduce the Dynamic Integration of task-specific Adapters (DIA) framework, which comprises two components: Task Specific Adapter Integration (TSAI) and Patch-Level Model Alignment.

(a) Task-Specific Adapter Integration. For incremental task t, we learn a task adapter $A_{t,b}$ and a task signature vector $\tau_{t,b} \in R_d$ at each transformer block b. Each token is routed to the relevant task adapters through the signature vectors, processed independently, and combined into an integrated, task-informed output.



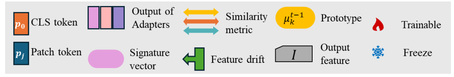
(a) Task-Specific Adapter Integration



(b) Patch-level Distillation

(c) Patch-level Feature Reconstruction

(a) Patch-level Distillation. We promote feature drift in patch tokens that contribute to new task learning while penalizing those that do not, thus regulating the feature shift associated with old tasks.



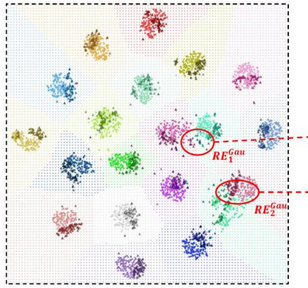
(a) Patch-level Feature Reconstruction. We identify patch tokens that are related to old class knowledge and integrate them with the old class prototype μ_{t-1}^k to reconstruct old class feature $\hat{\mu}_{t-1}^k$ aligned with new tasks.

Experiments

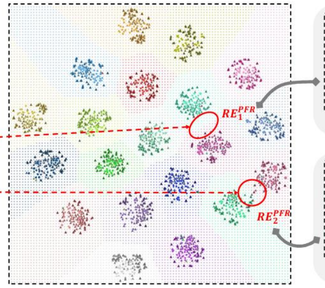
We conduct experiments on
1. Typical benchmarks Cifar-100,CUB 200
2. Challenging Datasets: ImageNet-R, ImageNet-A.

DIA achieve improved performance on benchmark datasets while maintaining an optimal balance between computational complexity and accuracy

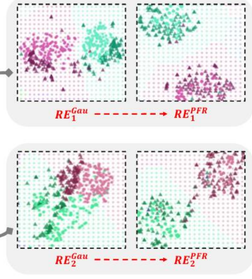
Method	Params	Flops	ImageNet-R		ImageNet-A		CUB-200		Cifar-100	
			$\mathcal{A}^{10}\uparrow$	$\bar{\mathcal{A}}^{10}\uparrow$	$\mathcal{A}^{10}\uparrow$	$\bar{\mathcal{A}}^{10}\uparrow$	$\mathcal{A}^{10}\uparrow$	$\bar{\mathcal{A}}^{10}\uparrow$	$\mathcal{A}^{10}\uparrow$	$\bar{\mathcal{A}}^{10}\uparrow$
FT	86M	17.58B	20.93	40.35	6.03	16.57	22.05	45.67	22.17	41.83
Adam-Ft	86M	33.72B	54.33	61.11	48.52	59.79	86.09	90.97	81.29	87.15
SLCA	86M	17.58B	77.42	82.17	60.63	70.04	84.71	90.94	**91.26**	94.09
Adam-Prompt-shallow	0.04M	36.28B	65.79	72.97	29.29	39.14	85.28	90.89	85.04	89.49
Adam-Prompt-deep	0.28M	36.28B	72.30	78.75	53.46	64.75	86.6	91.42	83.43	89.47
L2P	0.04M	35.85B	72.34	77.36	44.04	51.24	79.62	85.76	82.84	87.31
DualPrompt	0.13M	33.72B	69.10	74.28	53.19	61.47	81.10	88.23	83.44	88.76
CODA-Prompt	0.38M	33.72B	73.30	78.47	52.08	63.92	77.23	81.90	87.13	91.75
ConvPrompt	0.17M	17.98B	77.86	81.55	—	—	82.44	85.59	88.10	92.39
CPrompt	0.25M	23.62B	77.15	82.92	55.23	65.42	80.35	87.66	88.82	92.53
C-ADA	0.06M	17.62B	73.76	79.57	54.10	65.43	76.13	85.74	88.25	91.85
LAE	0.19M	35.24B	72.39	79.07	47.18	58.15	80.97	87.22	85.33	89.96
Adam-Adapter	1.19M	36.47B	65.29	72.42	48.81	58.84	85.84	91.33	87.29	91.21
EASE	1.19M	177.11B	76.17	81.73	55.04	65.34	84.65	90.51	87.76	92.35
InfLoRA	0.19M	21.79B	76.78	81.42	52.16	61.27	80.78	89.21	88.31	92.57
DIA (Ours)	0.17M	17.91B	**79.03**	**85.61**	**61.69**	**71.58**	**86.73**	**93.21**	90.80	**94.29**



(a) FR by Gaussian Sampling



(b) FR by PFR



(c) Optimize decision boundaries

The decision boundaries formed by these pseudo features are also visualized through the background color.

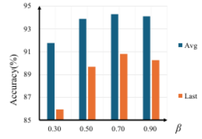
- **Dark-colored triangles** represent actual features
- **Light-colored circles** denote pseudo features generated by feature reconstruction (FR) methods.

Distillation Loss	ImageNet-R		Cifar-100	
	$\mathcal{A}^{10}\uparrow$	$\bar{\mathcal{A}}^{10}\uparrow$	$\mathcal{A}^{10}\uparrow$	$\bar{\mathcal{A}}^{10}\uparrow$
$\mathcal{L}_{pdl}$ w/ α_{eu}	76.38	81.73	87.42	92.91
$\mathcal{L}_{pdl}$ w/ α_{cos}	78.28	85.16	89.67	93.88
$\mathcal{L}_{pdl}$ w/ $\alpha_{\angle}$ (ours)	79.03	85.61	90.80	94.29
$\mathcal{L}_{fd}$	76.67	83.82	88.31	92.97
$\mathcal{L}_{pdl}$ w/ cls	77.01	83.87	88.87	93.02
$\mathcal{L}_{pdl}$ w/o cls (ours)	79.03	85.61	90.80	94.29

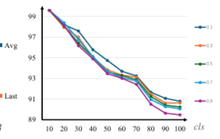
Ablation Study on Similarity Metrics in L_{pdl} and other distillation methods.

Ablation	ImageNet-R		Cifar100	
	$\mathcal{A}^{10}$	$\bar{\mathcal{A}}^{10}$	$\mathcal{A}^{10}$	$\bar{\mathcal{A}}^{10}$
DIA w SDC	75.13	83.63	89.13	93.46
DIA w LDC	76.43	84.08	88.66	93.66
DIA w PFR	79.03	85.61	90.80	94.29

Performance comparison with different feature reconstruction (FR) methods.



(a) Ablation on hyperparameter β



(b) Ablation on hyperparameter λ

Ablation experiments on hyper-parameters β and λ conducted on the CIFAR-100.