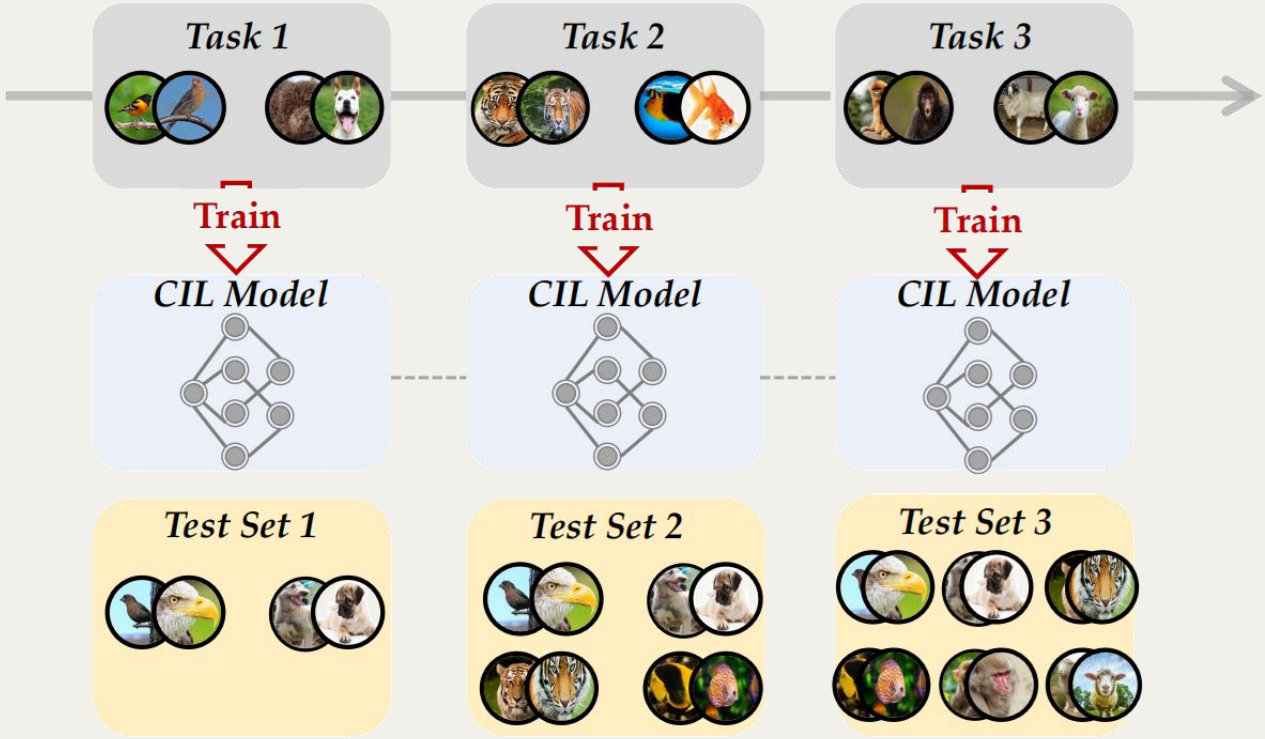
Incremental Learning

马鸣霄

2026.01.09

Incremental learning

增量学习（Incremental Learning，持续学习(Continual Learning)）



动机：随着时间的推移，更多的新数据逐渐可用，同时旧数据可能由于存储限制或隐私保护等原因逐渐不可用。

能力：不断地处理现实世界中连续的信息流，在吸收新知识的同时保留甚至整合、优化旧知识的能力。[1]

操作：通过对新的复杂多变环境下的数据进行持续学习，而不是重头训练整个模型。

[1] Parisi G I, Kemker R, Part J L, et al. Continual lifelong learning with neural networks: A review[J]. Neural networks, 2019, 113: 54-71.
[2] Zhou D W, Wang Q W, Qi Z H, et al. Class-incremental learning: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.

Incremental learning

关键点：灾难性遗忘（catastrophic forgetting）

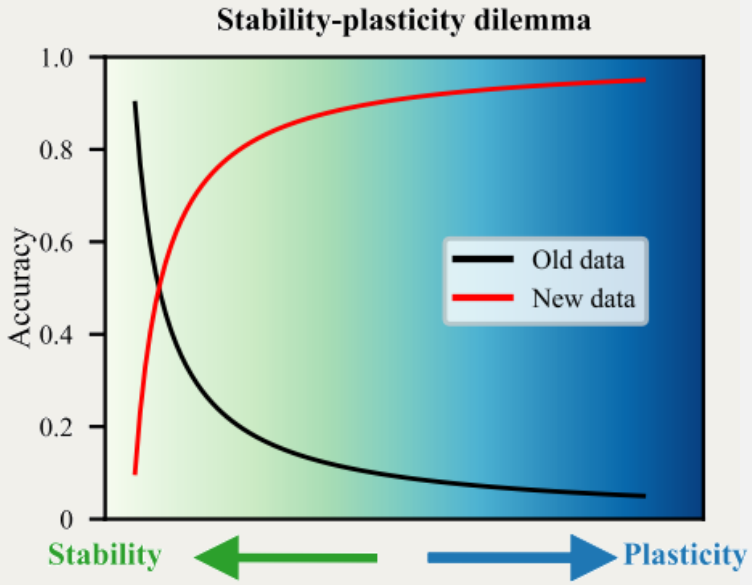
原因：原有的固定数据→连续的数据流，Stability-Plasticity dilemma

稳定性（memory stability）

防止新输入对已有知识的显著干扰

↓

希望参数不变



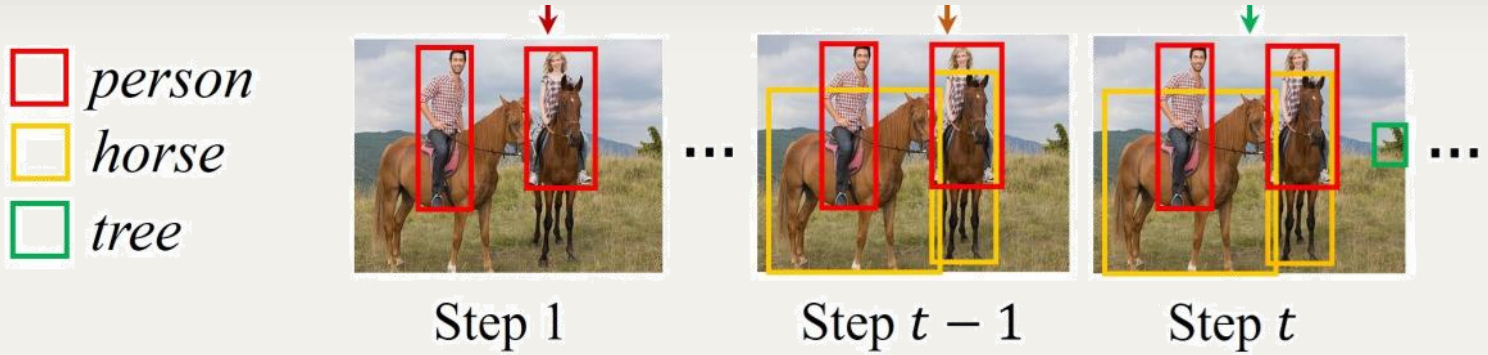
可塑性（learning plasticity）

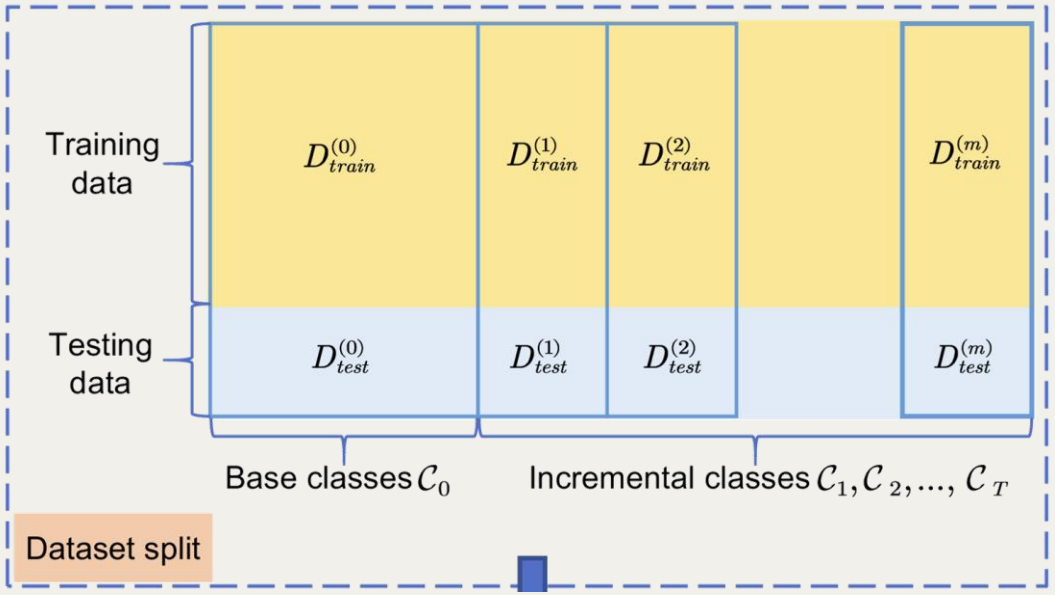
从新数据中整合新知识和提炼已有知识的能力

↓

参数可变

Incremental learning





Definition 1. Class-Incremental Learning *aims to learn from an evolutive stream with new classes. Assume there is a sequence of B training tasks*[1] *$\{\mathcal{D}^1, \mathcal{D}^2, \cdots, \mathcal{D}^B\}$ without overlapping classes, where $\mathcal{D}^b = \{(\mathbf{x}_i^b, y_i^b)\}_{i=1}^{n_b}$ is the b-th incremental step with n_b training instances. $\mathbf{x}_i^b \in \mathbb{R}^D$ is an instance of class $y_i^b \in Y_b$, Y_b is the label space of task b, where $Y_b \cap Y_{b'} = \varnothing$ for $b \neq b'$. We can only access data from $\mathcal{D}^b$ when training task b. The ultimate goal of CIL is to continually build a classification model for all classes. In other words, the model should not only acquire the knowledge from the current task $\mathcal{D}^b$ but also preserve the knowledge from former tasks. After each task, the trained model is evaluated over all seen classes $\mathcal{Y}_b = Y_1 \cup \cdots Y_b$. Formally, CIL aims to fit a*

[1] Zhou D W, Wang Q W, Qi Z H, et al. Class-incremental learning: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.

Gated Integration of Low-Rank Adaptation for Continual Learning of Large Language Models

Yan-Shuo Liang, Jia-Rui Chen and Wu-Jun Li*
National Key Laboratory for Novel Software Technology,
School of Computer Science, Nanjing University, P. R. China
{liangys,jiaruichen2005}@smail.nju.edu.cn, liwujun@nju.edu.cn

Task & Setting: Continual Learning，LLM（T5、Llama-2、Llama-3），Low-Rank Adaptation（LoRA）

Datasets（ NLP tasks ）:

① SuperNI benchmark: dialogue generation, information extraction, question answering, summarization, and sentiment analysis.

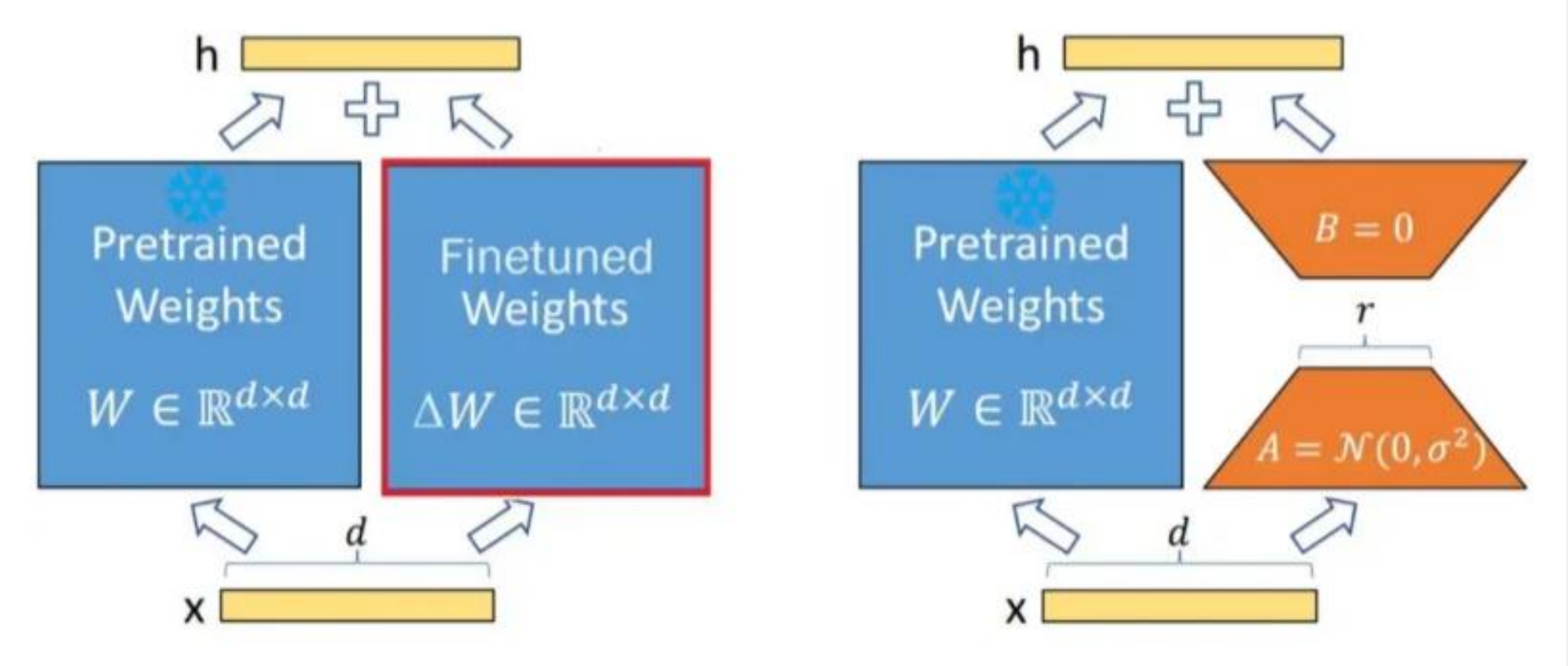
Order: 3 tasks are selected from each type, resulting in 15 tasks.

②Long Sequence benchmark.

Order: 15 diverse classification tasks.

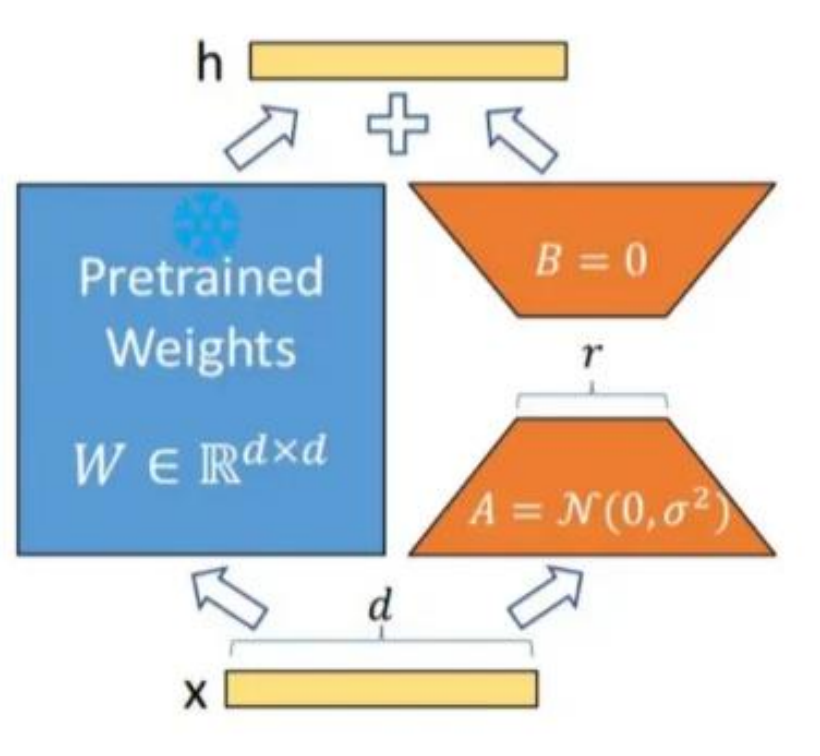
Background: Low-Rank Adaptation（LoRA）

“全量微调”的参数量为d*d；LoRA微调的参数量为2r*d（ r<<d ）



Hu E J, Shen Y, Wallis P, et al. Lora: Low-rank adaptation of large language models[J]. ICLR, 2022, 1(2): 3.

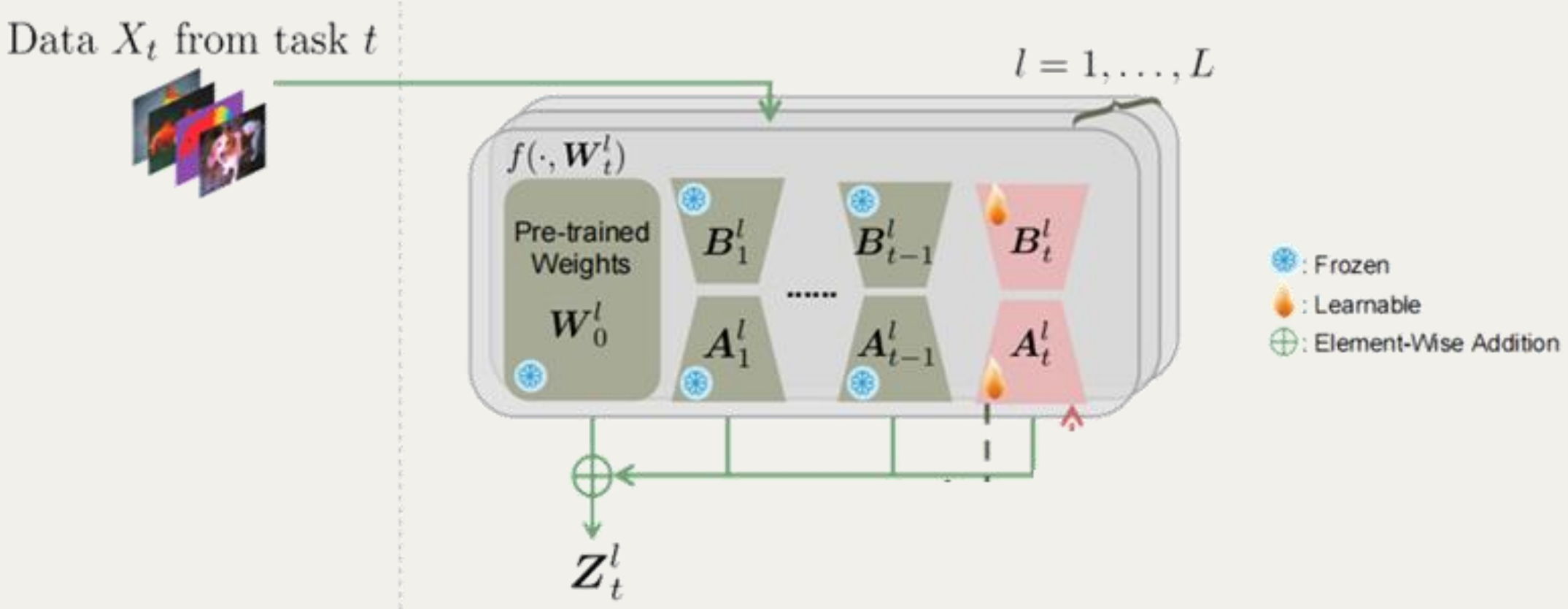
Background: LoRA——用更少的参数去近似建模一个复杂变换



- 保持原始参数矩阵 W 不变
- 在旁边引入一个由两个较小矩阵: $A \in \mathbb{R}^{r\times d}$ 和 $B \in \mathbb{R}^{d\times r}$
- 将它们的乘积 $\Delta W = B \cdot A$ 作为对 W 的一个“附加调整项”
- 整体的权重变为: $\hat{W} = W + \Delta W$
- 在训练过程中，仅更新 A 和 B 的参数，W 保持冻结状态。

Hu E J, Shen Y, Wallis P, et al. Lora: Low-rank adaptation of large language models[J]. ICLR, 2022, 1(2): 3.

Motivation: they force the new and old LoRA branches to influence old tasks equally, potentially leading to forgetting.



Liu X, Chang X. LoRA Subtraction for Drift-Resistant Space in Exemplar-Free Continual Learning[C]//Proceedings of the Computer Vision and Pattern Recognition Conference. 2025: 15308-15318.

Method:

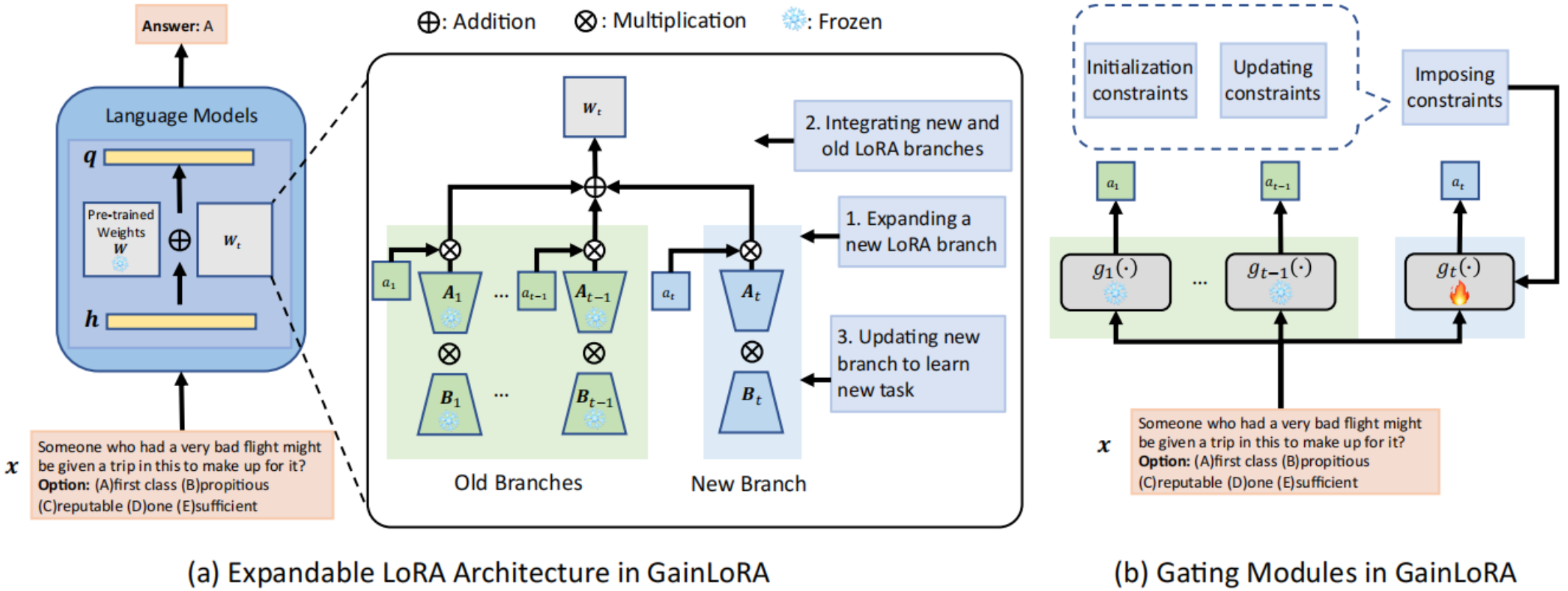
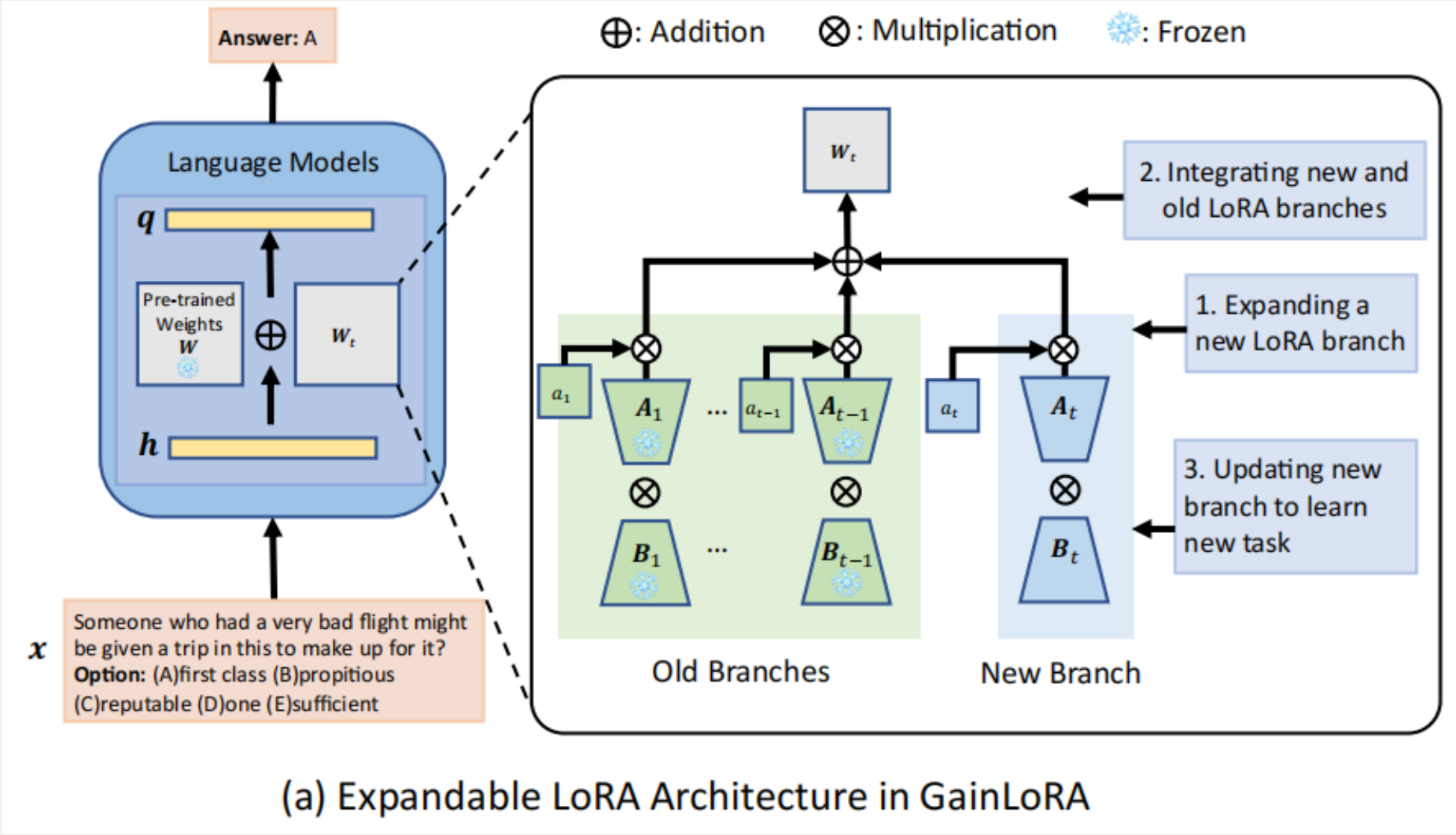


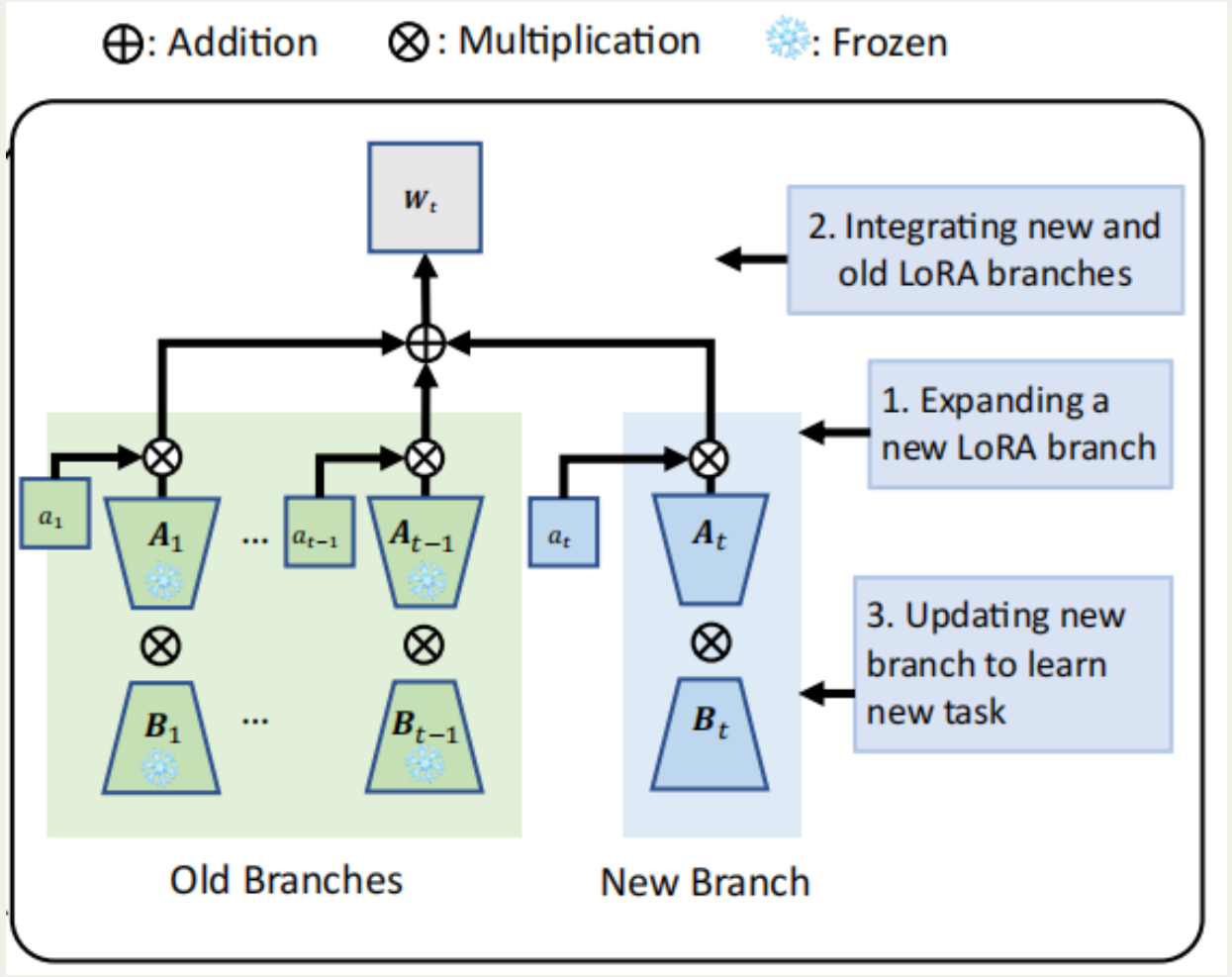
Figure 1: (a) shows the expandable LoRA architecture of our GainLoRA for learning the t-th new task. (b) shows that for each task $\mathcal{T}_i$, GainLoRA uses an independent gating module $g_i(\cdot)$ to generate integration coefficient a_i.

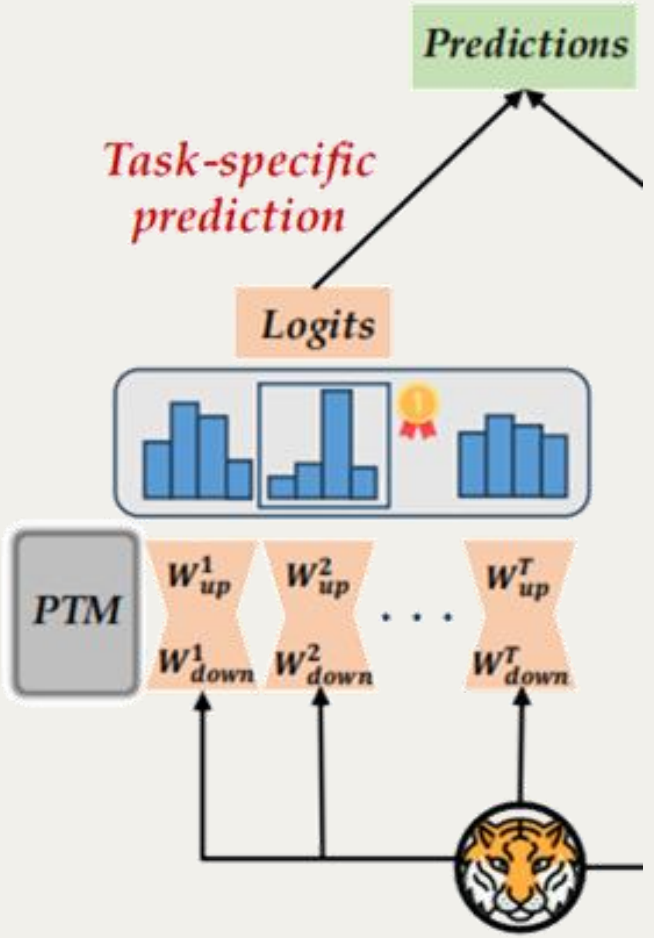
Method:



(a) Expandable LoRA Architecture in GainLoRA

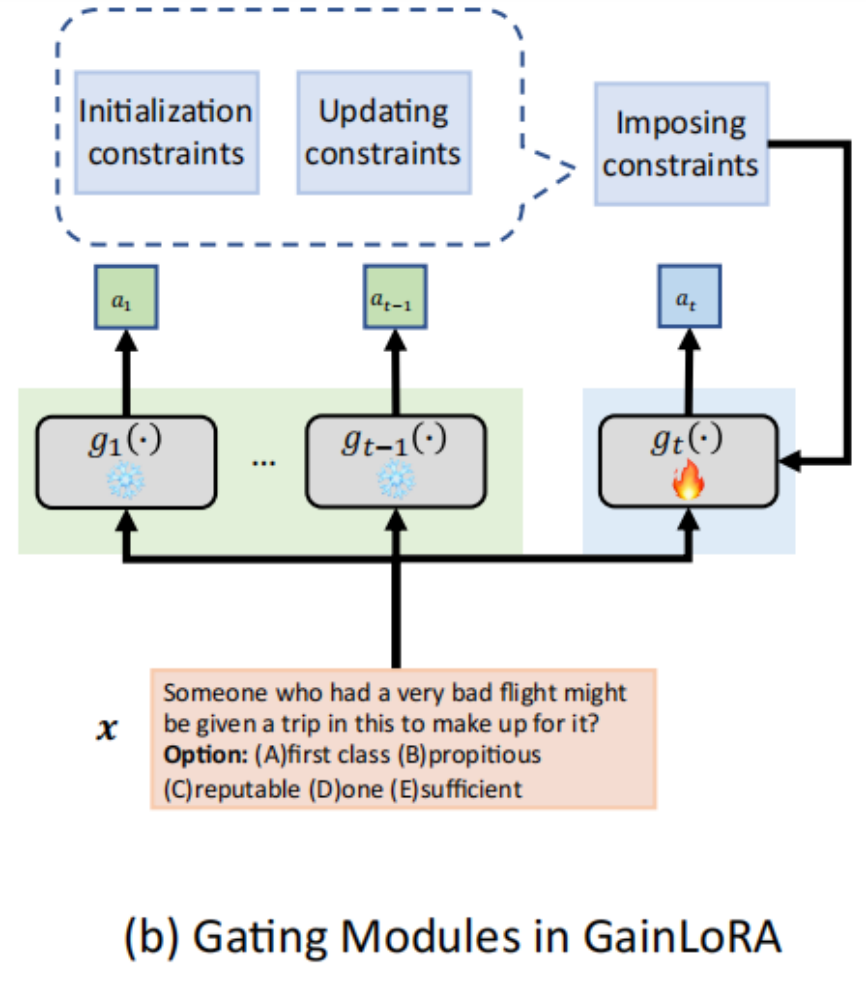
Method:





Wang Y, Zhou D W, Ye H J. Integrating Task-Specific and Universal Adapters for Pre-Trained Model-based Class-Incremental Learning[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2025: 806-816.

Method:

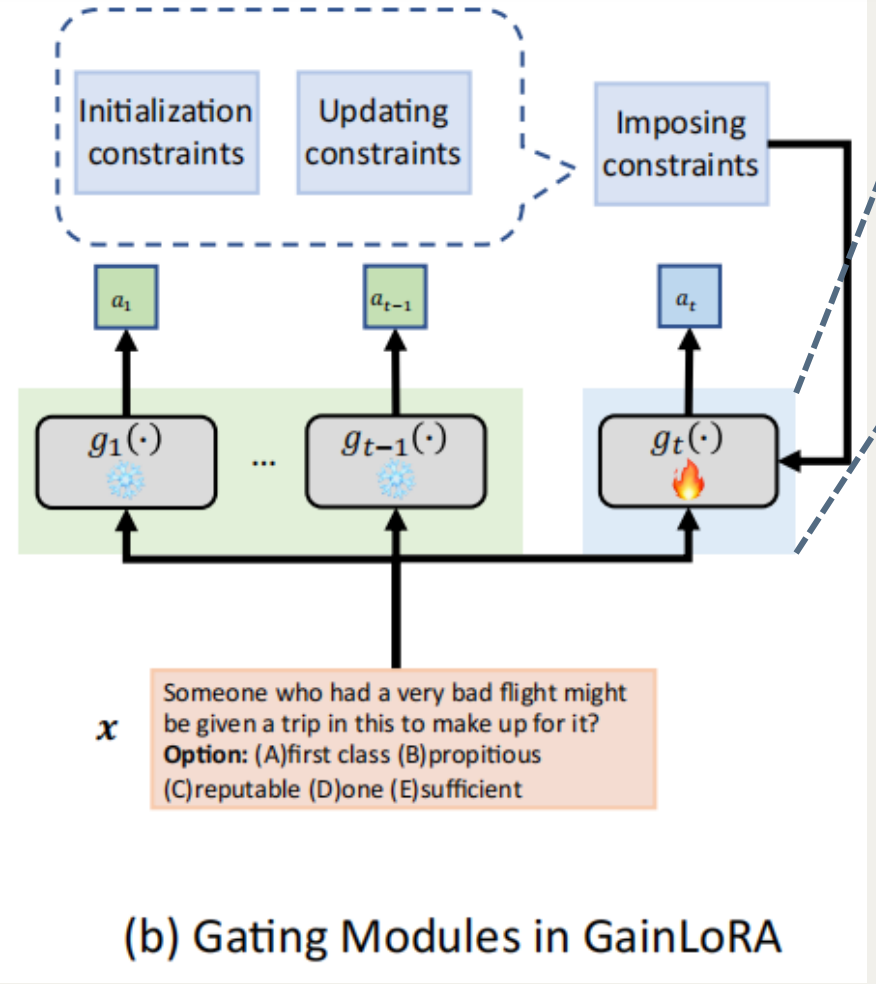


(b) Gating Modules in GainLoRA

学习第 t 个新任务时，新 LoRA 分支（第 t 个分支）的整合系数 $a_t = g_t(x)$ 需满足：

- 对所有旧任务（$T_1, T_2, ..., T_{t-1}$）的输入 x：$a_t \approx 0$（最小化新分支对旧任务的影响）
- 对新任务（T_t）的输入 x：a_t 无强制约束（保证新任务能正常学习）

Method:



(b) Gating Modules in GainLoRA

$$\begin{aligned} \boldsymbol{p}_{i,0} &= \boldsymbol{p}_0 = \text{Pool}(\text{Token}(\boldsymbol{x})), \\ \boldsymbol{p}_{i,l} &= \sigma(\boldsymbol{G}_{i,l}\boldsymbol{p}_{i,l-1}),\ l \in \{1, 2, ..., L\}, \\ g_i(\boldsymbol{x}) &= f(\boldsymbol{G}_{i,L+1}\boldsymbol{p}_{i,L}). \qquad f(\cdot) : \mathbb{R} \to [0, 1]. \end{aligned}$$

$$\mathcal{M}_{t,l} = \text{span}\{\boldsymbol{p}_{t,l-1} |\ \boldsymbol{p}_{t,l-1} \text{ is defined in } (3), (\boldsymbol{x}, \boldsymbol{y}) \in \cup_{i=1}^{t-1}\mathcal{D}_i\}.$$

the initialization of $\boldsymbol{G}_{t,l}$ $(1 \le l \le L)$

$$\text{Init}(\boldsymbol{G}_{t,l}) \leftarrow \boldsymbol{G}_{t-1,l}.$$

$$\text{Init}(\boldsymbol{G}_{t,L+1}) \perp \mathcal{M}_{t,L+1},\ f(0) = 0$$

$$\Delta\boldsymbol{G}_{t,l} \perp \mathcal{M}_{t,l} \quad \text{for} \quad 1 \le l \le L+1$$

Experiments:

Table 1: Results on different task sequences with T5-large model. Results of methods with * are copied from existing paper [75].

Method	Order 1		Order 2		Order 3		Order 4	
	AP↑	FT↓	AP↑	FT↓	AP↑	FT↓	AP↑	FT↓
LFPT5* [42]	39.03	10.87	29.70	20.72	66.62	14.57	67.40	13.20
EPI* [65]	-	-	-	-	75.19	0.77	75.10	2.44
MIGU+FT [11]	-	-	-	-	71.30	11.39	69.05	14.06
EWC [24]	15.32	26.78	18.19	30.28	43.24	23.66	46.25	32.90
TaSL [15]	27.51	18.53	28.05	17.39	71.37	6.20	73.11	6.52
KIFLoRA [14]	28.33	16.44	30.31	16.27	72.19	3.10	73.72	4.75
SeqLoRA	7.30	47.60	7.03	47.97	49.46	27.60	33.81	45.53
IncLoRA [61]	12.33	41.93	16.65	36.56	61.19	13.63	62.46	15.92
C-LoRA [50]	22.69	24.25	32.81	11.60	66.83	8.64	61.86	14.18
O-LoRA [61]	26.37	19.15	32.83	11.99	70.98	3.69	71.21	4.03
GainLoRA (O-LoRA)	**47.84**	**2.26**	**46.84**	2.91	73.37	3.02	76.01	2.49
InfLoRA [32]	39.78	7.64	39.57	8.93	75.15	4.19	75.79	3.47
GainLoRA (InfLoRA)	46.21	2.40	46.44	**2.61**	**78.01**	**0.77**	**77.54**	**1.25**

Experiments:

Table 3: The overall results on different task sequences with Llama-2-7B, Llama-2-13B and Llama-3-8B.

Models	Methods	Order 1		Order 2	
		AP↑	FT↓	AP↑	FT↓
Llama-2-7B	O-LoRA [61]	39.37	15.84	37.55	20.23
	GainLoRA (O-LoRA)	51.10	4.96	**51.14**	5.57
	InfLoRA [32]	42.93	11.23	39.94	15.00
	GainLoRA (InfLoRA)	**51.27**	**2.84**	50.17	**4.71**
Llama-2-13B	O-LoRA [61]	43.92	14.15	40.05	19.53
	GainLoRA (O-LoRA)	52.47	4.78	51.68	5.86
	InfLoRA [32]	43.64	14.85	45.74	10.61
	GainLoRA (InfLoRA)	**53.64**	**2.87**	**52.46**	**4.90**
Llama-3-8B	O-LoRA [61]	42.49	8.85	38.67	19.28
	GainLoRA (O-LoRA)	**53.39**	3.56	51.69	6.20
	InfLoRA [32]	43.27	6.02	48.77	5.88
	GainLoRA (InfLoRA)	52.18	**1.40**	**52.48**	**4.21**

Experiments:

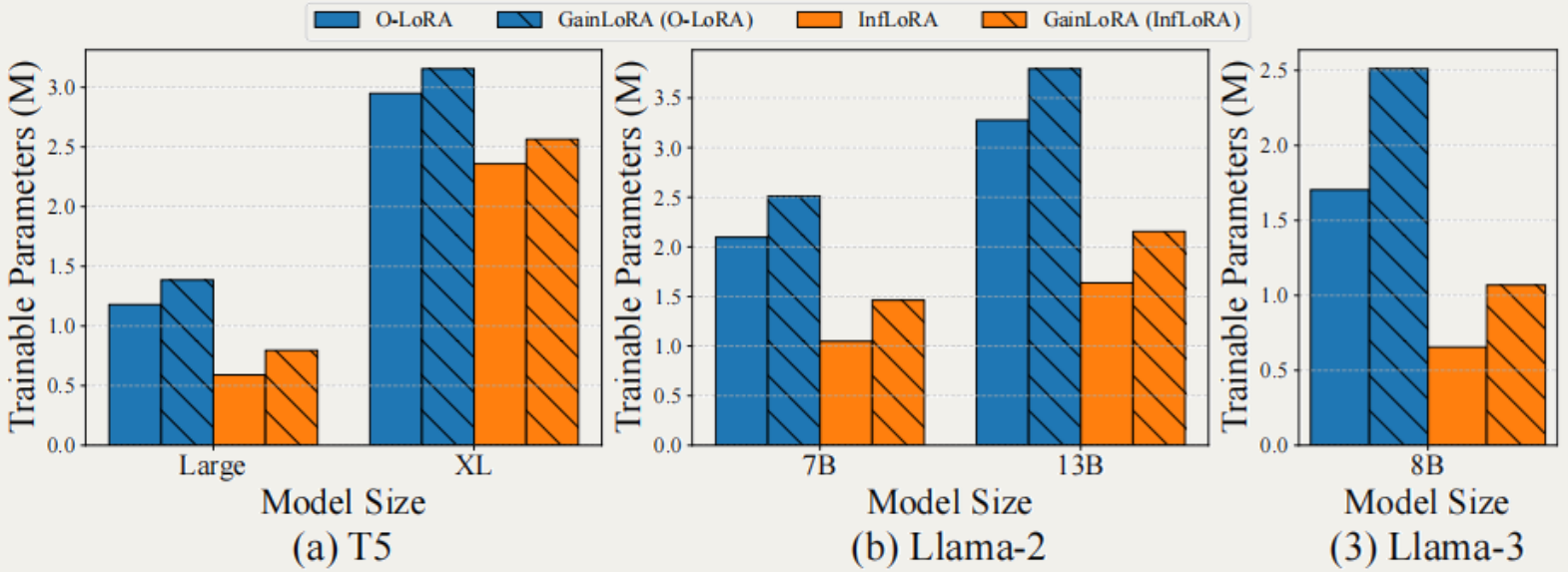


Figure 3: (a), (b) and (c) show the number of trainable parameters for different CL methods and model backbones under task sequences Order 1 and Order 2.

Continuous Subspace Optimization for Continual Learning

Quan Cheng[1,2], **Yuanyu Wan**[3,4], **Lingyu Wu**[1,2], **Chenping Hou**[5], **Lijun Zhang**[1,2,*]
[1]National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
[2]School of Artificial Intelligence, Nanjing University, Nanjing, China
[3]School of Software Technology, Zhejiang University, Ningbo, China
[4]Hangzhou High-Tech Zone (Binjiang) Institute of Blockchain and Data Security, Hangzhou, China
[5]College of Science, National University of Defense Technology, Changsha, China
{chengq, wuly, zhanglj}@lamda.nju.edu.cn
wanyy@zju.edu.cn, hcpnudt@hotmail.com

Continuous Subspace Optimization for Continual Learning

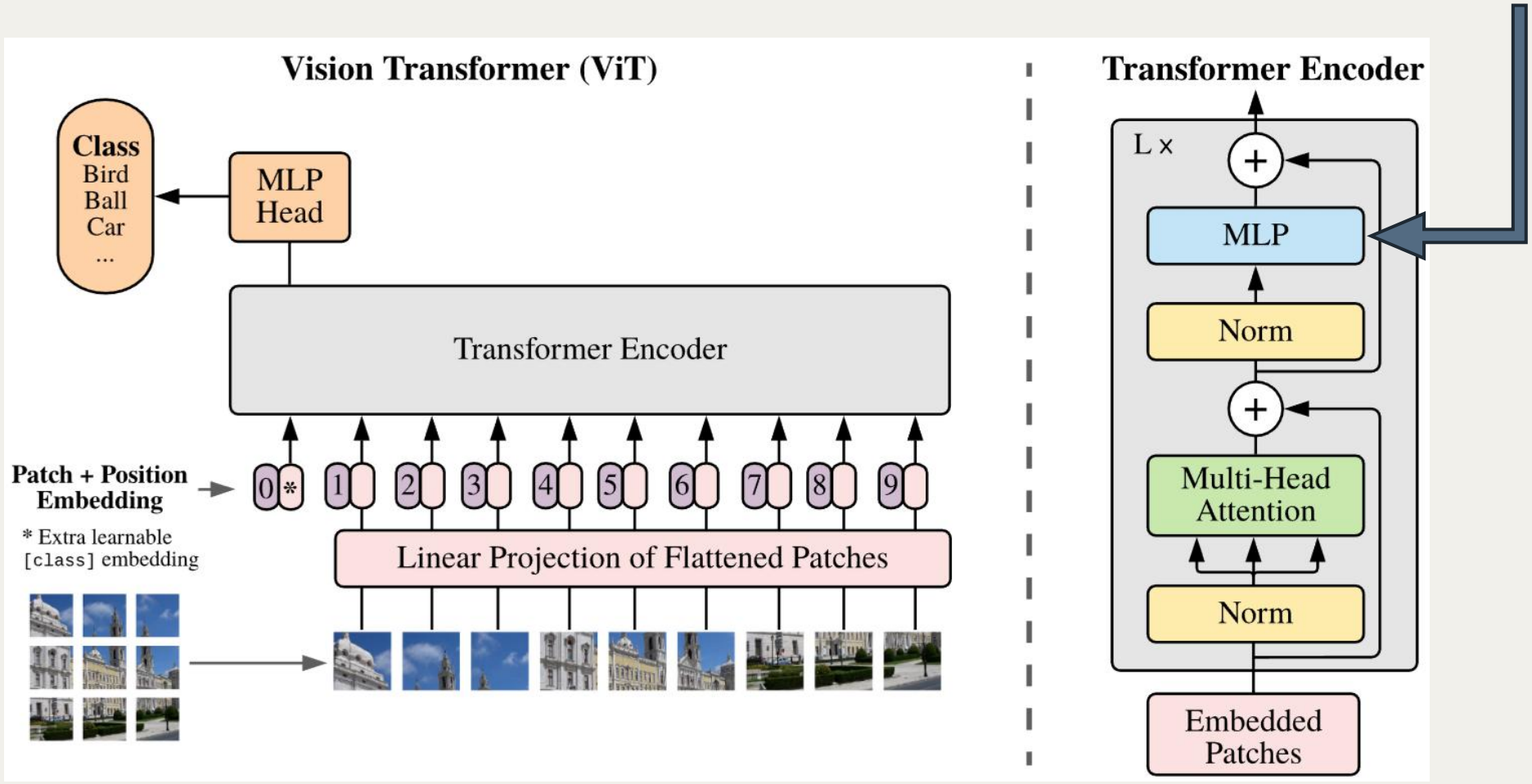
Task & Setting: Class Incremental Learning，Pre-Trained Model (ViT-B/16)

Motivation：不想引入LoRA，认为$\Delta W = B \times A$中每一个列向量，都是A矩阵列向量的线性组合（列空间），这是ΔW的范围；A的更新本质是“微调这r个列向量的数值”，微调列空间的“维度”和“核心方向”有限。

Continuous Subspace Optimization for Continual Learning

Task & Setting: Class Incremental Learning， Pre-Trained Model (ViT-B/16)

CoSO: only optimize the output projection layers in multi-head attention module.



Continuous Subspace Optimization for Continual Learning

Method:

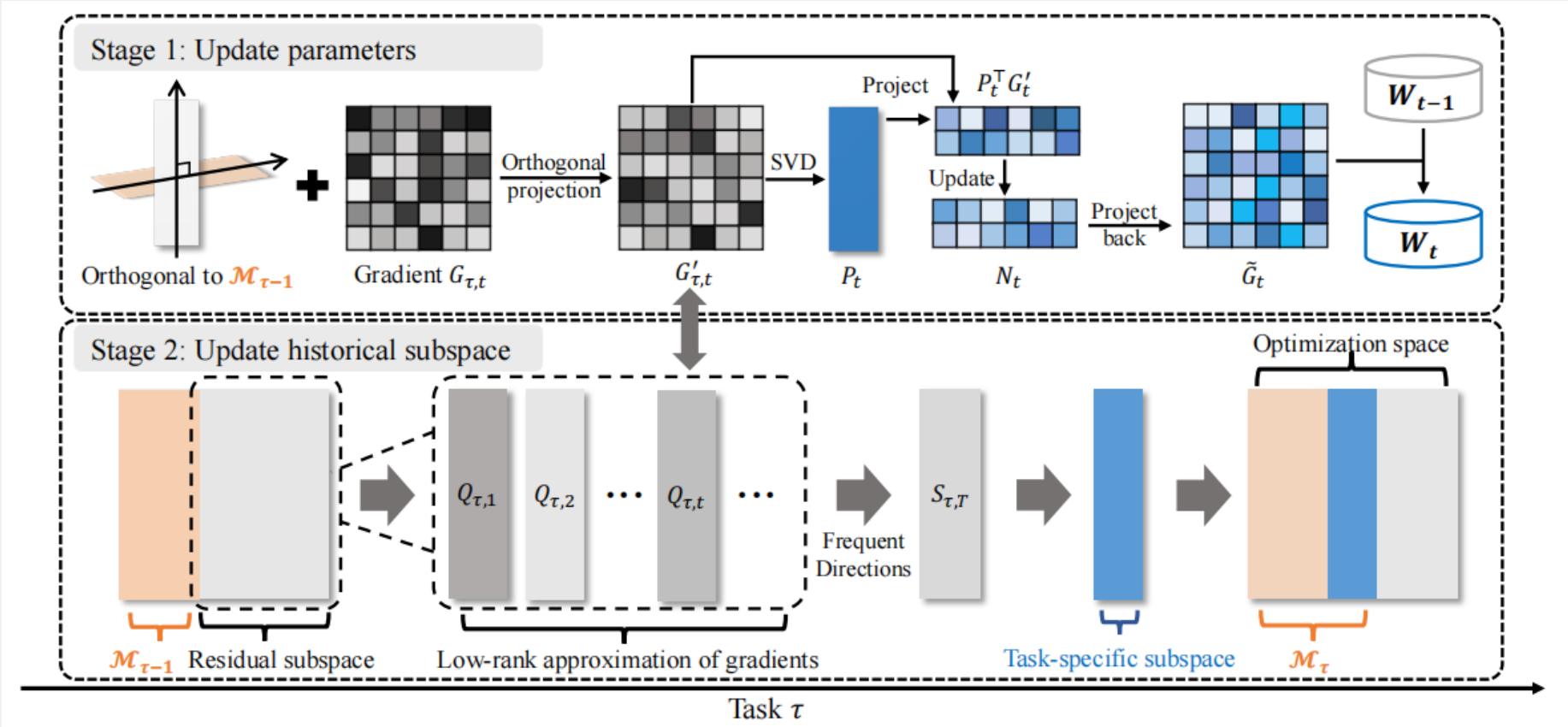


Figure 1: CoSO optimizes the parameters in continual low-rank subspaces, enhancing the learning capacity of models. To mitigate forgetting, the optimization subspaces of the current task are set to be orthogonal to the historical task subspace. While learning a task, CoSO consolidates the low-rank approximation matrices $\{Q_{\tau,t}\}_{t=1}^{T}$ into a task-specific component $S_{\tau,T}$ through Frequent Directions. The dedicated component is then used to update the historical task subspace spanned by $\mathcal{M}_{\tau-1}$.

Continuous Subspace Optimization for Continual Learning

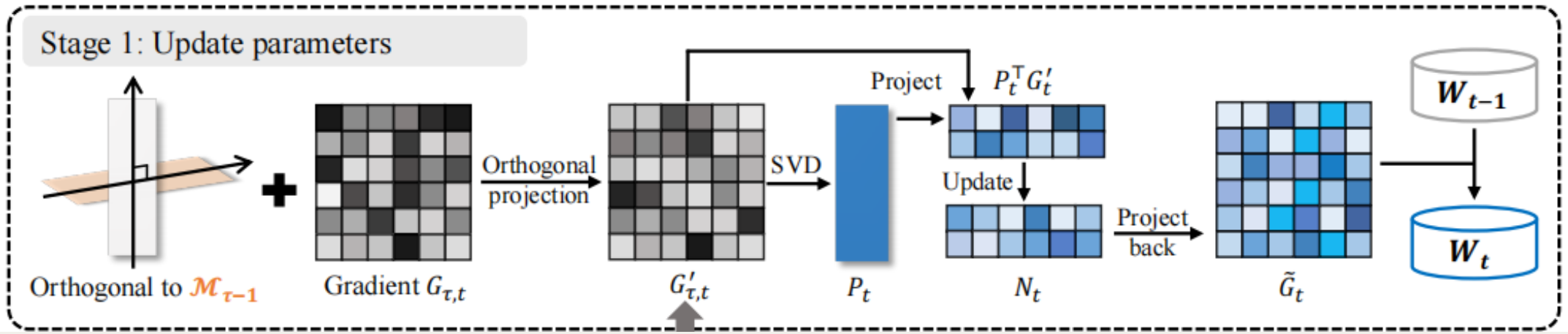
Method:

$\mathcal{M}_\tau$是“参数更新方向的集合”（正交基矩阵），维度随任务增多而扩大（每新增一个任务，拼接其核心方向），训练时的“约束工具”。

Continuous Subspace Optimization for Continual Learning

Method: 提取低维核心梯度 → 低维优化 → 映射回高维更新梯度

减少计算开销!



- $\mathcal{M}_{\tau-1}$：历史任务子空间（由之前所有任务的关键更新方向构成，是正交基矩阵）；
- $G_{\tau,t}$：当前任务 τ 第 t 步的原始梯度；
- $G'_{\tau,t}$：正交投影后的梯度（去除了与历史子空间重叠的分量，避免干扰旧知识）；
- $P_{\tau,t}$：梯度低秩投影矩阵（由 $G'_{\tau,t}$ 的 SVD 得到，捕捉当前梯度的关键方向）；

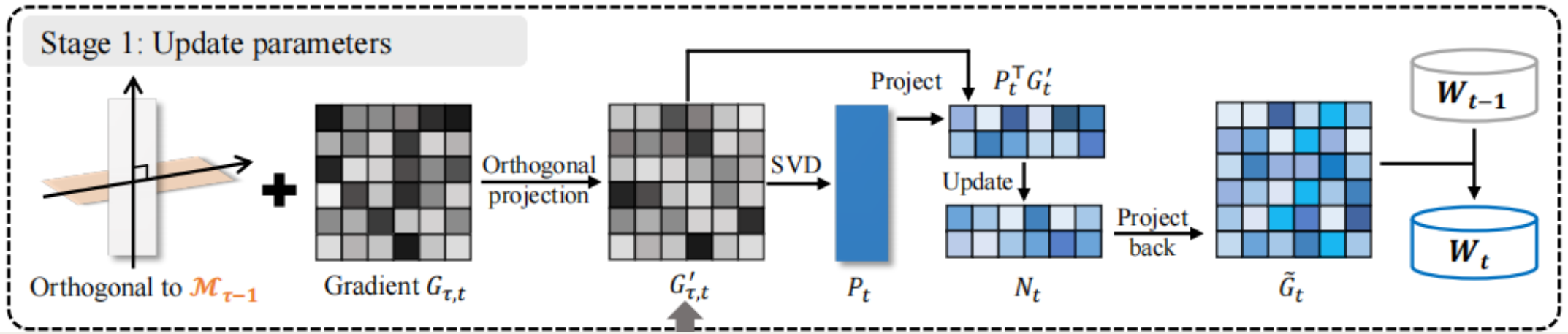
$$U\Sigma V^\top = \mathrm{SVD}_{r_1}(G'_{\tau,t})$$

$$P_{\tau,t} = U[:, :r_1],$$

Continuous Subspace Optimization for Continual Learning

Method: 提取低维核心梯度 → 低维优化 → 映射回高维更新梯度

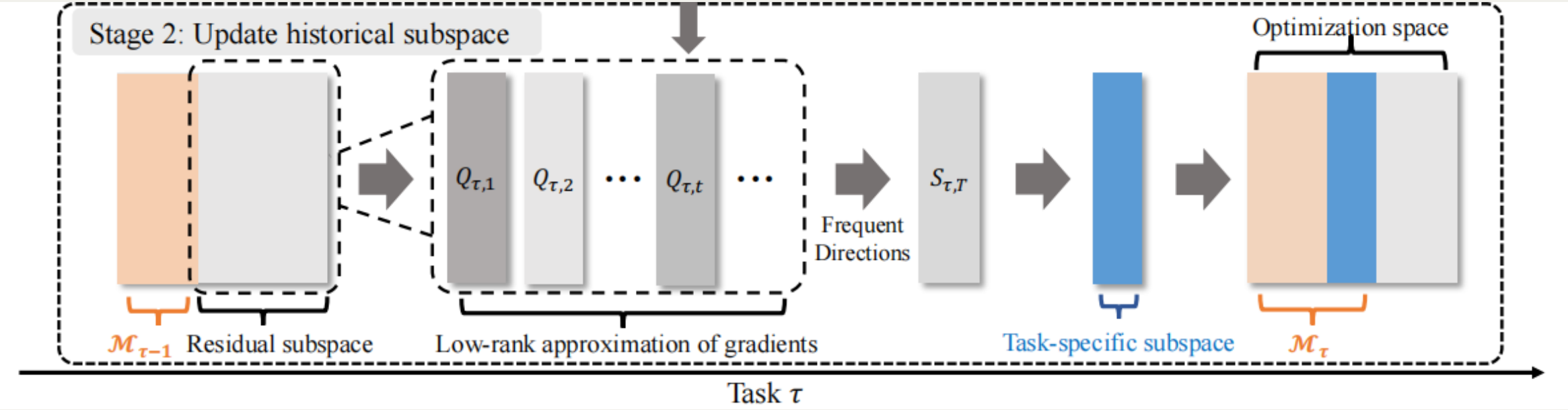
减少计算开销!



Hyperparameter	CIFAR100	ImageNet-R	DomainNet
Projection rank (r_1)	15	50	70

Continuous Subspace Optimization for Continual Learning

Method:



$$U\Sigma V^{\top} = \mathrm{SVD}_{r_2}(G'_{\tau,t})$$

$$Q_{\tau,t} = U\Sigma.$$

Hyperparameter	CIFAR100	ImageNet-R	DomainNet
Projection rank (r_1)	15	50	70
Frequent directions rank (r_2)	100	120	160

Continuous Subspace Optimization for Continual Learning

Experiments:

Table 1: Results (%) on ImageNet-R with varying numbers of tasks (5, 10 and 20). All reported results with mean and standard deviation are computed over 3 independent runs.

Method	ImageNet-R (5 Tasks)		ImageNet-R (10 Tasks)		ImageNet-R (20 Tasks)	
	ACC_5	$\overline{ACC}_5$	ACC_{10}	$\overline{ACC}_{10}$	ACC_{20}	$\overline{ACC}_{20}$
L2P	$65.03_{\pm 0.03}$	$69.97_{\pm 0.15}$	$62.87_{\pm 0.72}$	$68.90_{\pm 0.58}$	$58.64_{\pm 0.34}$	$65.57_{\pm 0.35}$
DualPrompt	$68.24_{\pm 0.23}$	$71.82_{\pm 0.39}$	$65.30_{\pm 0.52}$	$69.62_{\pm 0.29}$	$60.47_{\pm 0.54}$	$65.91_{\pm 0.52}$
CODA-P	$73.65_{\pm 0.15}$	$77.88_{\pm 0.30}$	$72.10_{\pm 0.29}$	$76.90_{\pm 0.41}$	$67.16_{\pm 0.11}$	$72.34_{\pm 0.44}$
InfLoRA	$77.53_{\pm 0.30}$	$82.24_{\pm 0.11}$	$74.43_{\pm 0.31}$	$80.50_{\pm 0.06}$	$70.30_{\pm 0.14}$	$77.04_{\pm 0.06}$
SD-LoRA	$79.15_{\pm 0.20}$	$83.01_{\pm 0.42}$	$77.34_{\pm 0.35}$	$82.04_{\pm 0.24}$	$75.26_{\pm 0.37}$	$80.22_{\pm 0.72}$
VPT-NSP^2	$79.72_{\pm 0.19}$	$84.33_{\pm 0.29}$	$77.87_{\pm 0.10}$	$83.09_{\pm 0.26}$	$75.42_{\pm 0.27}$	$81.32_{\pm 0.21}$
CoSO	$\mathbf{82.10}_{\pm 0.13}$	$\mathbf{86.38}_{\pm 0.07}$	$\mathbf{81.10}_{\pm 0.39}$	$\mathbf{85.56}_{\pm 0.13}$	$\mathbf{78.19}_{\pm 0.28}$	$\mathbf{83.69}_{\pm 0.12}$

Continuous Subspace Optimization for Continual Learning

Experiments:

Table 2: Results (%) on CIFAR100 (10 Tasks) and DomainNet (5 Tasks). All reported results with mean and standard deviation are computed over 3 independent runs.

Method	CIFAR100 (10 Tasks)		DomainNet (5 Tasks)	
	ACC_{10}	$\overline{ACC}_{10}$	ACC_5	$\overline{ACC}_5$
L2P	$82.64_{\pm 0.26}$	$87.90_{\pm 0.19}$	$70.03_{\pm 0.09}$	$75.65_{\pm 0.06}$
DualPrompt	$84.68_{\pm 0.22}$	$90.12_{\pm 0.05}$	$72.25_{\pm 0.05}$	$77.84_{\pm 0.02}$
CODA-P	$86.60_{\pm 0.37}$	$91.46_{\pm 0.20}$	$73.16_{\pm 0.07}$	$78.75_{\pm 0.04}$
InfLoRA	$86.85_{\pm 0.08}$	$91.45_{\pm 0.16}$	$73.09_{\pm 0.11}$	$79.21_{\pm 0.08}$
SD-LoRA	$87.30_{\pm 0.45}$	$91.81_{\pm 0.27}$	$73.20_{\pm 0.12}$	$79.03_{\pm 0.04}$
VPT-NSP^2	$88.09_{\pm 0.12}$	$92.48_{\pm 0.11}$	$72.52_{\pm 0.13}$	$78.68_{\pm 0.06}$
CoSO	$\mathbf{88.77}_{\pm 0.16}$	$\mathbf{92.99}_{\pm 0.23}$	$\mathbf{74.27}_{\pm 0.07}$	$\mathbf{80.05}_{\pm 0.04}$

Incremental Learning

马鸣霄

2026.01.09