PLAN: Proactive Low-Rank Allocation for Continual Learning

Xiequn Wang, Zhan Zhuang, Yu Zhang

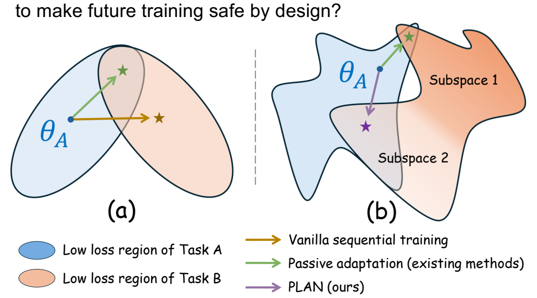
香港城市大學 City University of Hong Kong

Introduction

- Continual learning with large pre-trained models must add new tasks without forgetting; existing LoRA-style methods rely on passive orthogonality and fail to *plan* for future interference.
- **PLAN—Proactive Low-Rank Allocation** pre-assigns task-specific orthogonal subspaces and trains them with a perturbation-aware **min–max** objective that anticipates worst-case future updates.
- Preserves past knowledge while enabling efficient adaptation with minimal extra parameters—a simple, scalable recipe for rehearsal-free continual learning.

Why Continual Learning Still Forgets

- Large pre-trained models must learn new tasks without revisiting past data. Most LoRA-style approaches enforce *passive* orthogonality, which reduces interference but doesn't **plan** for future updates. We ask: can we proactively allocate subspaces to make future training safe by design?



- Setting: rehearsal-free CL; no access to past data.
- Base: low-rank adapters (LoRA family).
- Goal: maintain past performance while learning Task_t efficiently.

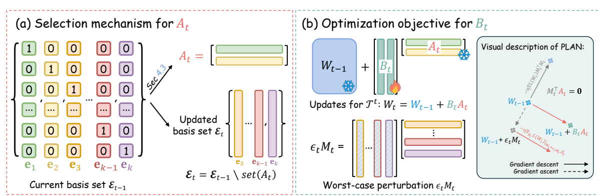
Method

$$\min_{B_t} \mathcal{L}_{\text{new}}(\theta + A_t B_t) + \lambda \max_{|\xi| \le \epsilon} I(\theta + A_t B_t + \xi, \text{ past })$$

- Train new task in its own subspace A_t, while guarding against worst-case future perturbations ξ that could harm past tasks, I is Interference risk on past tasks; either projection magnitude onto past subspaces.

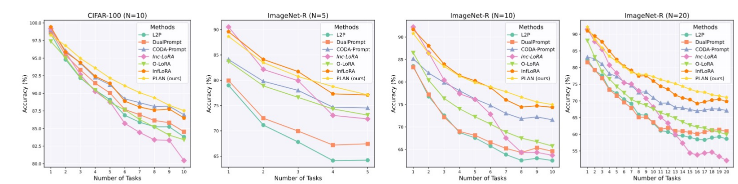
Algorithm 1 PLAN Method for Continual Learning

Input: a pre-trained ViT model f_θ, number of tasks T, training set $\{\{x_i^t, y_i^t\}_{i=1}^{n_t}\}_{t=1}^{T}$, number of training epochs E, predefined LoRA basis set $\mathcal{E}_0$.
Output: The learned LoRA parameters $\{A_t, B_t\}_{t=1}^{T}$.
for t in 1, ..., T **do**
  Construct A_t through Eqs. (12) and (13);
  $\mathcal{E}_t \leftarrow \mathcal{E}_{t-1} \setminus set(A_t)$;
  for e = 1, ..., E **do**
    Sample batch $\mathcal{B} = \{(\boldsymbol{x}_1^t, y_1^t), ...(\boldsymbol{x}_b^t, y_b^t)\}$;
    $\mathbf{g} \leftarrow \nabla_{W_t} L_{\mathcal{B}}(W_t) M_t^\top$;
    Compute $\hat{\epsilon}_t(W_t)$ with $\mathbf{g}$ according to Eq. (7);
    $\mathbf{g}^{\text{PLAN}} \leftarrow \nabla_{B_t} L_{\mathcal{B}}(W_t)|_{W_{t-1}+A_t B_t + \hat{\epsilon}_t(W_t) M_t}$;
    Update B_t with $\mathbf{g}^{\text{PLAN}}$ through gradient descent;
  end for
end for



Experiments

Method	*ImageNet-R* ($N = 5$)		*ImageNet-R* ($N = 10$)		*ImageNet-R* ($N = 20$)	
	Acc ↑	AAA ↑	Acc ↑	AAA ↑	Acc ↑	AAA ↑
L2P	64.20 (±0.30)	69.25 (±0.63)	62.52 (±0.41)	68.69 (±0.35)	58.63 (±0.52)	65.67 (±0.33)
Dual-Prompt	67.43 (±1.13)	71.40 (±0.85)	64.59 (±1.24)	69.59 (±0.72)	60.89 (±0.62)	66.20 (±0.51)
CODA-Prompt	74.52 (±4.25)	78.21 (±2.73)	71.58 (±0.26)	76.47 (±0.28)	67.10 (±0.46)	72.38 (±0.42)
Inc-LoRA	72.36 (±0.57)	79.60 (±0.27)	63.69 (±0.84)	74.54 (±0.35)	52.12 (±0.72)	67.73 (±0.45)
O-LoRA	73.12 (±6.09)	77.33 (±3.67)	65.74 (±0.81)	72.89 (±0.87)	59.94 (±0.82)	68.92 (±0.69)
InfLoRA	77.09 (±0.33)	**81.96** (±0.28)	74.37 (±0.54)	80.37 (±0.62)	69.83 (±0.65)	76.83 (±0.54)
PLAN (ours)	**77.79** (±0.24)	81.93 (±0.63)	**75.25** (±0.42)	**80.41** (±0.56)	**71.06** (±0.42)	**77.93** (±0.56)



Method	*ImageNet-R* ($N = 5$)		*ImageNet-R* ($N = 10$)		*ImageNet-R* ($N = 20$)	
	Acc ↑	AAA ↑	Acc ↑	AAA ↑	Acc ↑	AAA ↑
Inc-LoRA	72.36 (±0.57)	79.60 (±0.27)	63.69 (±0.84)	74.54 (±0.35)	52.12 (±0.72)	67.73 (±0.45)
w/o A_t selection	75.97 (±0.77)	80.69 (±0.49)	72.14 (±0.51)	78.56 (±0.56)	68.35 (±0.53)	76.33 (±1.12)
w/o perturbation	76.66 (±0.76)	80.38 (±0.51)	74.97 (±0.40)	79.57 (±0.30)	70.65 (±0.68)	76.42 (±0.76)
PLAN (ours)	**77.79** (±0.24)	**81.93** (±0.63)	**75.25** (±0.42)	**80.41** (±0.56)	**71.06** (±0.42)	**77.93** (±0.56)

Conclusion

- **Proactive subspace planning:** We introduce **PLAN (Proactive Low-Rank Allocation)**, which pre-assigns **orthogonal** low-rank subspaces to incoming tasks, explicitly planning to avoid future interference under a fixed adapter budget.