RobustMerge: Parameter-Efficient Model Merging for MLLMs with Direction Robustness

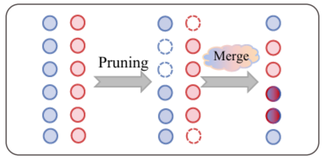
Spotlight

Fanhu Zeng[1], Haiyang Guo[1], Fei Zhu[2*], Li Shen[3], Hao Tang[4*]

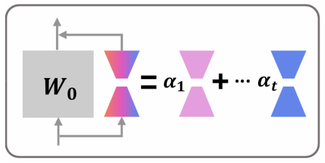
[1]Institute of Automation, Chinese Academy of Sciences [2]HKISI-CAS [3]Sun Yat-sen University [4]Peking University

Background

Two types of merging paradigm

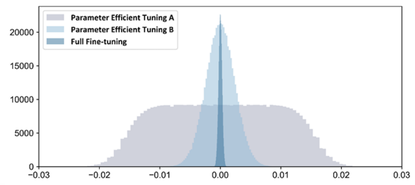


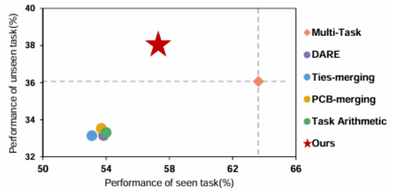
Full fine-tune (FFT) merging



Parameter-efficient (PEFT) merging

Existing problems

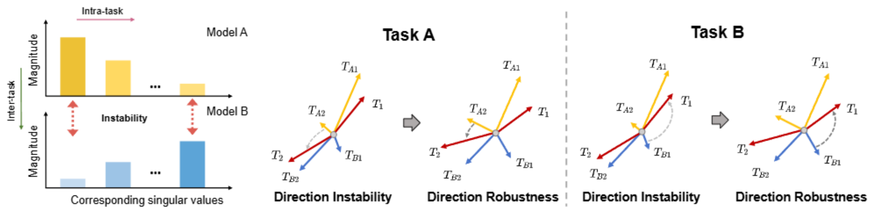




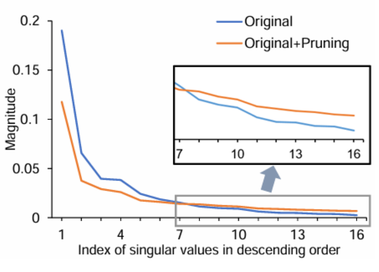
Distinct difference between two paradigms **degrades** traditional FFT methods on PEFT merging

Motivation

New perspective: Direction robustness



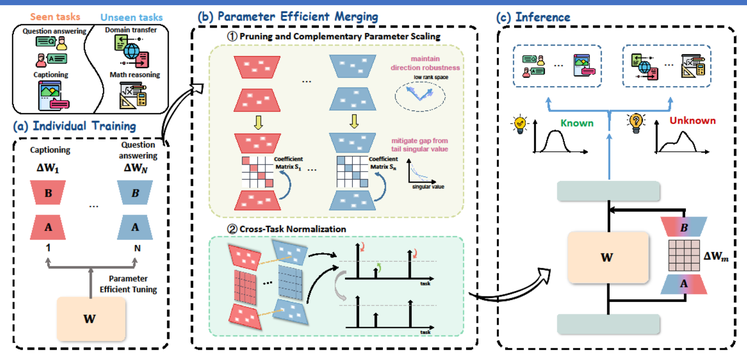
Merging PEFT modules can be formulated into **merging singular vectors** in low-rank space



(1) **Large gap** exists between head and tail singular values, causing **direction instability**.

(2) **Protecting tail values** is more important, and adaptively scale singular values helps **direction robustness**.

Method



Insight

1. Changing larger parameters are more likely to **alter directions** in low-rank space
2. **Imbalanced data size** leads to overfitting and underfitting for different tasks

1. Pruning and complementary parameter scaling

$$\widetilde{\mathbf{A}} = \mathcal{M}_{\mathrm{A}}(k) \odot \mathbf{A}, \quad \widetilde{\mathbf{B}} = \mathcal{M}_{\mathrm{B}}(k) \odot \mathbf{B},$$

$$\mathbf{S}^i = \frac{\sum_{j=1}^{d_i} \mathrm{abs}(\mathbf{A}_{[i,j]})}{\sum_{j=1}^{d_i} \mathrm{abs}(\mathcal{M}_{\mathrm{A}[i,j]} \odot \mathbf{A}_{[i,j]})}, \quad i = 1, \cdots, r.$$

2. Cross-task normalization $\widetilde{\mathbf{S}}_n^i = \mathbf{S}_n^i \Big/ \sum_{\cdot}^{N} \mathbf{S}_n^i, n = 1, \cdots, N.$

3. Merge parameter-efficient modules $\Delta\widetilde{\mathbf{W}}_n = \widetilde{\mathbf{B}}_n \cdot \widetilde{\mathbf{S}}_n \cdot \widetilde{\mathbf{A}}_n, n = 1, \cdots, N,$

Experiment

	Seen Tasks									Unseen Tasks				
Method	**SciQA**	**Image**	**VQA**	**REC**	**OCR**	**Viz**	**Flickr**	**IconQA**	**Avg**	**AVQA**	**Image-R**	**S2W**	**TabMWP**	**Avg**
Individual	83.74	96.02	67.58	43.40	65.50	64.80	57.29	75.54	69.23	-	-	-	-	-
Zero-Shot	61.73	40.87	62.88	36.10	41.16	41.03	49.07	14.09	43.37	51.62	28.27	5.98	15.01	25.22
Multi-Task	76.90	74.08	67.05	35.98	65.37	66.67	56.09	66.87	63.62	76.33	41.39	8.34	18.20	36.06
Task Arithmetic	71.94	57.49	67.06	38.90	62.87	44.80	49.20	39.21	53.93	74.78	37.37	7.52	13.57	33.31
DARE	71.59	57.25	66.26	39.38	62.56	44.93	49.13	39.59	53.84	73.75	37.67	7.56	13.62	33.15
Ties-merging	71.49	55.88	66.73	39.67	**65.12**	44.35	47.06	34.46	53.09	73.43	38.44	7.47	13.23	33.14
PCB-merging	71.10	57.82	**67.59**	38.22	64.35	44.58	48.90	37.01	53.70	74.57	36.28	7.84	15.44	33.53
DIR-Merging	**73.43**	**65.54**	67.20	**44.80**	62.97	**46.61**	**52.80**	**45.90**	**57.33** (+3.4)	**79.30**	**45.79**	**9.23**	**17.62**	**37.99** (+4.5)

Method	POPE	MME	MMBench
Zero-Shot	86.4	1476.9	66.1
Traditional MTL	86.9	1433.5	62.9
Task Arithmetic	87.0	1465.2	67.3
DARE	86.4	1475.7	67.4
Ties-merging	86,7	1489.4	66.6
PCB-merging	86.6	1490.7	66.3
RobustMerge	**87.2**	**1494.9**	**68.1**

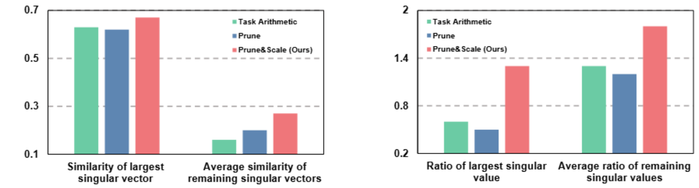
(1) Promote **seen tasks** and generalize well to **unseen tasks**

(2) Enhance ability on **general multimodal benchmarks**

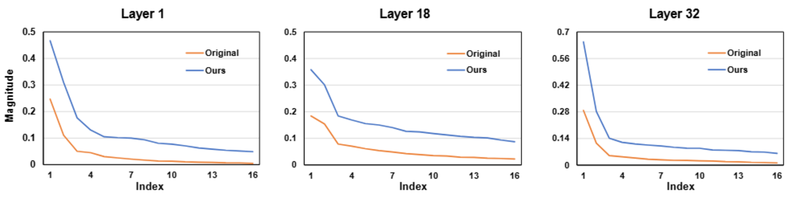
Further Analysis

Prune&Scale	Norm	SciQA	Image	VQA	REC	OCR	Viz	Flickr	IconQA	Avg
		71.94	57.49	67.06	38.90	62.87	44.80	49.20	39.21	53.93
✓		73.03	64.18	**67.50**	43.12	58.19	46.36	52.24	44.54	56.14 (+2.21)
✓	✓	**73.43**	**65.54**	67.20	**44.80**	**62.97**	**46.61**	**52.80**	**45.90**	**57.33** (+3.40)

Effectiveness of each component



Quantitative evaluation of direction robustness



Distribution of singular value in different layers

Conclusion

- We focus on PEFT model merging, highlighting the necessity of high-performance parameter-efficient merging algorithms.
- We analyze from the perspective of direction robustness of singular values in low-rank space and propose an effective training-free merging algorithm to maintain direction.

Dataset Model Code