Unleashing the Power of Task-Specific Directions in Parameter Efficient Fine-Tuning

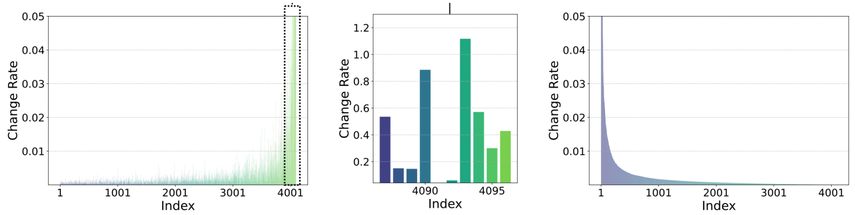
Chongjie Si[1], Zhiyi Shi[2], Shifan Zhang[1], Xiaokang Yang[1], Hanspeter Pfister[2], Wei Shen[1]
[1]Shanghai Jiao Tong University [2]Harvard University

ICLR

Observation

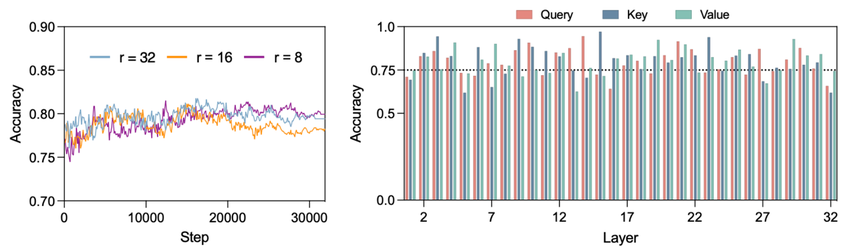
Task: Parameter Efficient Fine-Tuning (PEFT), minimal trainable parameters, comparable or even superior performance.

Question: What are task specific directions?



For a specific task and a pre-trained weight matrix W, considering the optimal weights for this task as W., the task-specific directions of this task on W are W's core singular directions whose singular values exhibit significantly higher change rates through the alteration from W to W*.*

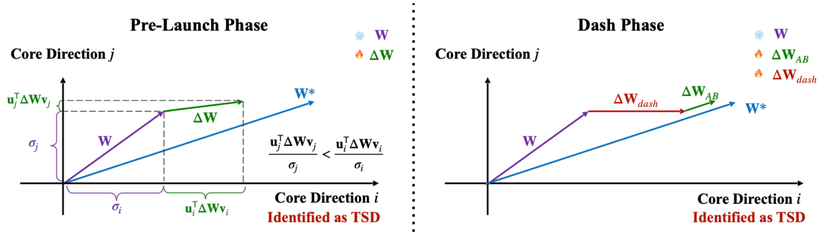
Challenge: How to obtain task specific directions?



Irrespective of the rank setting in LoRA, the training step, or the specific layer within the model, LoRA consistently captures the information of the task-specific directions.

LoRA-Dash

Framework:



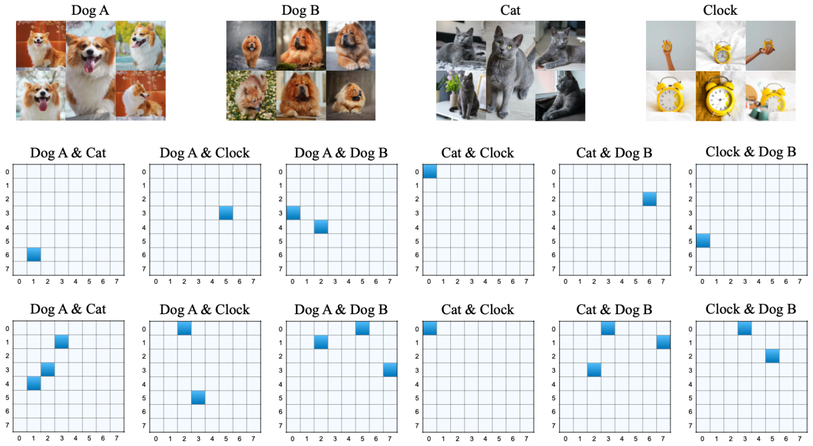
Pre-launch Phase:

- Train LoRA for several training steps, then calculate the change rates for W.
- Identify the top k singular directions of W with the highest change rates.
- These directions are denoted as Task-specific Directions.

Dash Phase:

- Model the changes of task-specific directions:

$$\mathbf{W} + \Delta\mathbf{W}_{all} = \mathbf{W} + \Delta\mathbf{W}_{AB} + \Delta\mathbf{W}_{dash} = \mathbf{W} + \mathbf{AB} + \sum_{i=1}^{s} \Delta\sigma_i \bar{\mathbf{u}}_i \bar{\mathbf{v}}_i^{\mathsf{T}}$$



Experiment

Numerical Results

Model	Method	Params(%)	BoolQ	PIQA	SIQA	HellaS.	WinoG.	ARC-e	ARC-c	OBQA	Avg.
ChatGPT	-	-	73.1	85.4	68.5	78.5	66.1	89.8	79.9	74.8	77.0
LLaMA-7B	Fully FT	100	69.9	84.2	78.9	92.3	83.3	86.6	72.8	83.4	81.4
	LoRA$_{r=4}$	0.10	2.3	46.1	18.3	19.7	55.2	65.4	51.9	57.0	39.5
	LoRA-Dash	0.10	65.2	79.9	78.3	82.8	77.1	78.6	65.4	78.4	**75.7**
	LoRA$_{r=8}$	0.21	31.3	57.0	44.0	11.8	43.3	45.7	39.2	53.8	40.7
	LoRA-Dash	0.21	69.8	81.1	77.3	85.1	81.1	77.2	64.1	79.6	**76.9**
	LoRA$_{r=16}$	0.42	69.9	77.8	75.1	72.1	55.8	77.1	62.2	78.0	70.9
	LoRA-Dash	0.42	66.9	80.2	77.8	78.8	79.2	78.0	61.9	77.4	**75.0**
	LoRA$_{r=32}$	0.83	68.9	80.7	77.4	78.1	78.8	77.8	61.3	74.8	74.7
	LoRA-Dash	0.83	69.9	82.8	78.6	84.9	81.6	82.3	66.5	80.8	**78.4**
	LoRA$_{r=64}$	1.66	66.7	79.1	75.7	17.6	78.8	73.3	59.6	75.2	65.8
	LoRA-Dash	1.66	69.6	79.5	76.0	82.8	75.8	81.5	64.7	81.0	**76.4**
LLaMA2-7B	Fully FT	100	72.2	84.9	80.9	93.1	84.7	87.5	74.2	85.1	82.8
	LoRA$_{r=16}$	0.41	71.7	81.6	79.5	89.5	81.9	82.9	67.9	79.6	79.3
	LoRA-Dash	0.41	70.9	82.2	80.5	90.2	80.1	83.5	68.9	80.8	**79.6**
	LoRA$_{r=32}$	0.82	69.8	79.9	79.5	83.6	82.6	79.8	64.7	81.0	77.6
	LoRA-Dash	0.82	71.0	75.7	79.3	91.1	78.6	84.2	69.8	78.8	**78.6**
LLaMA3-8B	Fully FT	100	75.3	89.9	81.5	95.8	87.6	91.6	79.3	87.4	86.1
	LoRA$_{r=16}$	0.35	72.3	86.7	79.3	93.5	84.8	87.7	75.7	82.8	82.8
	LoRA-Dash	0.35	74.8	88.0	80.6	95.2	85.6	89.0	77.4	84.8	**84.4**
	LoRA$_{r=32}$	0.70	70.8	85.2	79.9	91.7	84.3	84.2	71.2	79.0	80.8
	LoRA-Dash	0.70	75.3	88.5	80.2	95.7	86.8	90.7	80.2	85.6	**85.4**

Visual Results