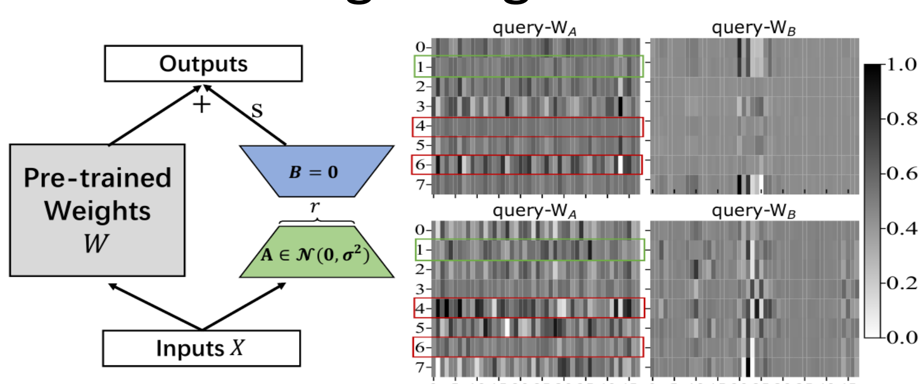
Low-Rank Spectral Adapter for Parameter-Efficient Fine-tuning of Transformer

Kun ZHANG, Guangyi LV, Yubin HUANGFU, Minglong XUE, Richang HONG, Jianping FAN, Xin LI, Si WEI

Frontiers of Computer Science, DOI: 10.1007/s11704-025-50771-9

Problems & Ideas

- Problems of hyper-parameter sensitivity and training instability:
  – LoRA module relies on a pre-defined intrinsic rank r and an scaling s, selecting the best values is resource-intensive and experience-dependent.
  – LoRA modules in different layers are usually initialized randomly for better generalization, which will cause the gradient changes of different layers to be overly dramatic.
- Ideas: Inspired by spectral theorem, we can first use spectral decomposition methods (e.g., SVD) to extract the pre-trained knowledge as orthonormal bases and subsequently recombine them for better fine-tuning and generalization.



(a)LoRA module for a single component of Pre-trained LLMs

(b) LoRA gradients from different layers at different stage

Low-Rank Adapter (LoRA) framework and the visualization of LoRA gradients from different layers at different training stages (i.e., the 5,000th step and 9,000th step).

Main Contributions

- Contributions:
  - a novel SpecAdapt that views the model fine-tuning process as a re-combination of the spectral bases without requiring direct training of a small number of parameters, allowing for effective prevention of overfitting or underfitting ;
  - A soft pruning technique: Singular-guided Weight Decay enabling the sharing of a unified hyper-parameter setting across different tasks, and a Gradient Normalization module to help to better propagate gradients across different layers.
  - Extensive experiments on multiple transformer-based large models over three different benchmarks are conducted to prove the superiority and efficiency of SpecAdapt in various aspects.

			Natural							Specialized				Structured								
Method	# params(M)	extra-inference time	Cifar100	Caltech101	DTD	Flower102	Pets	SVHN	Sun397	Camelyon	EuroSAT	Resisc45	Retinopathy	Clevr-Count	Clevr-Dist	DMLab	KITTI-Dist	dSpr-Loc	dSpr-Ori	sNORB-Azim	sNORB-Ele	Average
Traditional Fine-Tuning																						
Full	85.8	–	68.9	87.7	64.3	97.2	86.9	87.4	38.8	79.7	95.7	84.2	73.9	56.3	58.6	41.7	65.5	57.5	46.7	25.7	29.1	68.9
Linear	0	–	64.4	85.0	63.2	97.0	86.3	36.6	51.0	78.5	87.5	68.5	74.0	34.3	30.6	33.2	55.4	12.5	20.0	9.6	19.2	57.6
PEFT methods																						
BitFit	0.103	–	72.8	87.0	59.2	97.5	85.3	59.9	51.4	78.7	91.6	72.9	69.8	61.5	55.6	32.4	55.9	66.6	40.0	15.7	25.1	65.2
VPT-Shallow	0.063	✓	77.7	86.9	62.6	97.5	87.3	74.5	51.2	78.2	92.0	75.6	72.9	50.5	58.6	40.5	67.1	68.7	36.1	20.2	34.1	67.8
VPT-Deep	0.531	✓	**78.8**	90.8	65.8	98.0	88.3	78.1	49.6	81.8	96.1	83.4	68.4	68.5	60.0	46.5	72.8	73.6	47.9	**32.9**	37.8	72.0
Adapter	0.157	✓	69.2	90.1	68.0	98.8	89.9	82.8	54.3	84.0	94.9	81.9	75.5	80.9	65.3	48.6	78.3	74.8	48.5	29.9	41.6	73.9
AdaptFormer	0.157	✓	70.8	91.2	70.5	99.1	90.9	86.6	54.8	83.0	95.8	84.4	**76.3**	81.9	64.3	49.3	80.3	76.3	45.7	31.7	41.1	74.7
LoRA	0.295	–	67.1	**91.4**	69.4	98.8	90.4	85.3	54.0	**84.9**	95.3	84.4	73.6	**82.9**	**69.2**	49.8	78.5	75.7	47.1	31.0	**44.0**	74.5
FacT-TT_4	0.008	–	69.4	88.5	70.6	98.8	90.0	83.3	53.7	83.9	95.1	81.5	75.4	78.2	69.0	47.7	79.0	75.2	42.7	27.2	38.7	73.5
FacT-TK_8	0.014	–	70.3	88.7	69.8	99.0	90.4	84.2	53.5	82.8	95.6	82.8	75.7	81.1	68.0	48.0	80.5	74.6	44.0	29.2	41.1	74.0
SpecAdapt	0.370	–	73.7	90.2	**73.1**	**99.4**	**91.6**	**89.1**	**55.4**	84.6	**96.4**	**86.9**	72.6	80.0	65.4	**52.7**	**82.1**	**81.4**	**49.4**	31.2	36.6	**75.6**
Searching-based methods																						
NOAH*	0.361	✓	69.6	92.7	70.2	99.1	90.4	86.1	53.7	84.4	95.4	83.9	75.8	82.8	68.9	49.9	81.7	81.8	48.3	32.8	44.2	75.5
FacT-$TT_{\leq 16}$†	0.037	–	71.3	89.6	70.7	98.9	91.0	87.8	54.6	85.2	95.5	83.4	75.7	82.0	69.0	49.8	80.0	79.2	48.4	34.2	41.4	75.3
FacT-$TK_{\leq 32}$†	0.069	–	70.6	90.6	70.8	99.1	90.7	88.6	54.1	84.8	96.2	84.5	75.7	82.6	68.2	49.8	80.7	80.8	47.4	33.2	43.0	75.6
RepAdapter†	0.22	–	72.4	91.6	71.0	99.2	91.4	90.7	55.1	85.3	95.9	84.6	75.9	82.3	68.0	50.4	79.9	80.4	49.2	38.6	41.0	76.1
$GLoRA_{4/8}$*	0.86	–	76.4	92.9	74.6	99.6	92.5	91.5	57.8	87.3	96.8	88.0	76.0	83.1	67.3	54.5	86.2	83.8	52.9	37.0	41.4	78.0

† The results of these methods are obtained by searching hyperparameters on the val set of VTAB-1K benchmark.
* These methods are based on neural architecture search (NAS).

Backbone	method	MNLI	QQP	QNLI	SST-2	STS-B	MRPC	RTE	CoLA	AVG
$BERT_{large}$	Finetuning	85.3	**73.5**	91.1	**94.2**	**87.5**	**89.1**	70.4	61.2	81.54
	Adapter	84.6	71.9	90.9	93.7	87.2	88.2	69.1	57.5	80.39
	$LoRA_{r=8}$	85.4	72.6	91.4	93.8	86.3	88.6	**75.1**	60.3	81.69
	DyLoRA	84.8	72.3	92.4	93.5	85.5	88.7	73.2	60.2	81.33
	AdaLoRA	85.3	**73.5**	**92.6**	94.0	85.7	88.4	73.2	61.8	81.81
	SpecAdapt	**85.8**	72.3	91.7	93.5	86.4	88.5	74.8	**62.7**	**81.96**
$T5_{base}$	Finetuning	85.7	91.1	92.0	92.5	88.8	90.2	75.4	54.9	83.83
	Adapter	86.3	90.5	93.2	93.0	89.9	90.2	70.3	61.5	84.36
	$LoRA_{r=8}$	85.8	89.2	93.1	93.5	90.4	89.9	76.3	62.8	85.13
	DyLoRA	86.5	89.8	92.6	94.5	90.8	**91.1**	74.2	62.8	85.26
	AdaLoRA	**87.1**	**91.5**	93.9	**94.7**	**91.1**	88.6	75.4	61.0	85.41
	SpecAdapt	86.1	90.7	93.3	93.8	91.0	90.4	**76.5**	**66.4**	**86.03**

Experimental results on image benchmarks and Language benchmark. Left: results with ViT-B/16 on VTAB-1K benchmark; Right: performance on GLUE Benchmark.