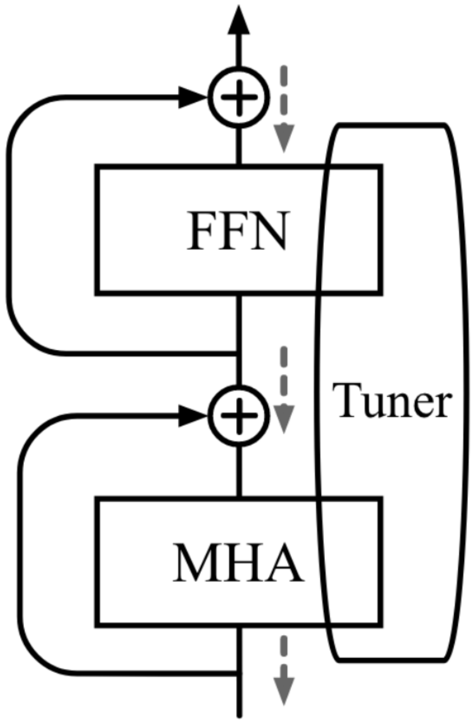
Res-Tuning: A Flexible and Efficient Tuning Paradigm via Unbinding Tuner from Backbone

Zeyinzi Jiang[1], Chaojie Mao[1], Ziyuan Huang[2], Ao Ma[1]
Yiliang Lv[1], Yujun Shen[3], Deli Zhao[1], Jingren Zhou[1]

Project page: https://res-tuning.github.io/

[1] Alibaba Group
[2] National University of Singapore
[3] Ant Group

Efficient Tuners



Existing methods are deeply embedded

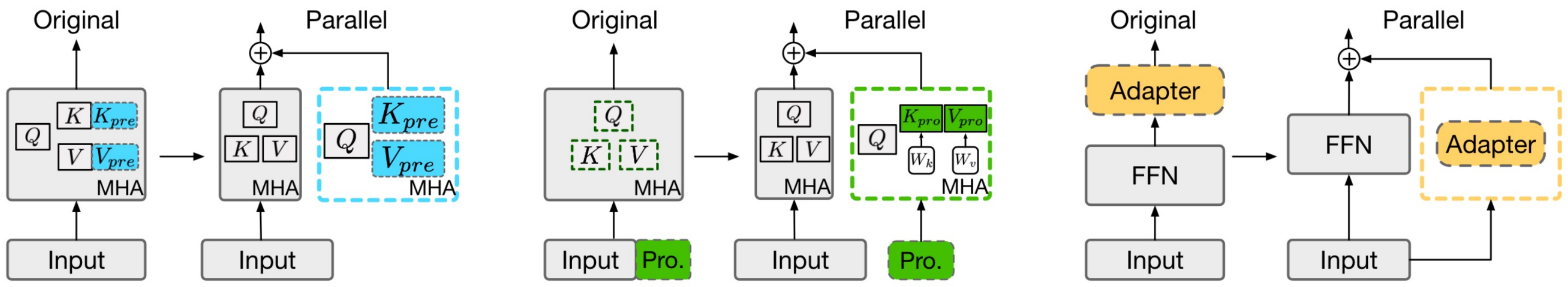
into original structures

Flexible
combination

Only parameter-efficient

Efficient
parameter and memory

Res-Tuner



(a) Prefix Tuning (b) Prompt Tuning (c) Adapter Tuning

$$\mathrm{MHA}_{\mathrm{pre}} = \mathrm{Attn}(\boldsymbol{x}\boldsymbol{W}_q, [\boldsymbol{K}_{pre}; \boldsymbol{x}\boldsymbol{W}_k], [\boldsymbol{V}_{pre}; \boldsymbol{x}\boldsymbol{W}_v])$$

$$\mathrm{MHA}_{\mathrm{pre}} = (1-\lambda)\underbrace{\mathrm{Attn}\left(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}\right)}_{\text{original attention}} + \lambda\underbrace{\mathrm{Attn}\left(\boldsymbol{Q}, \boldsymbol{K}_{pre}, \boldsymbol{V}_{pre}\right)}_{\text{prefix attention in parallel}}$$

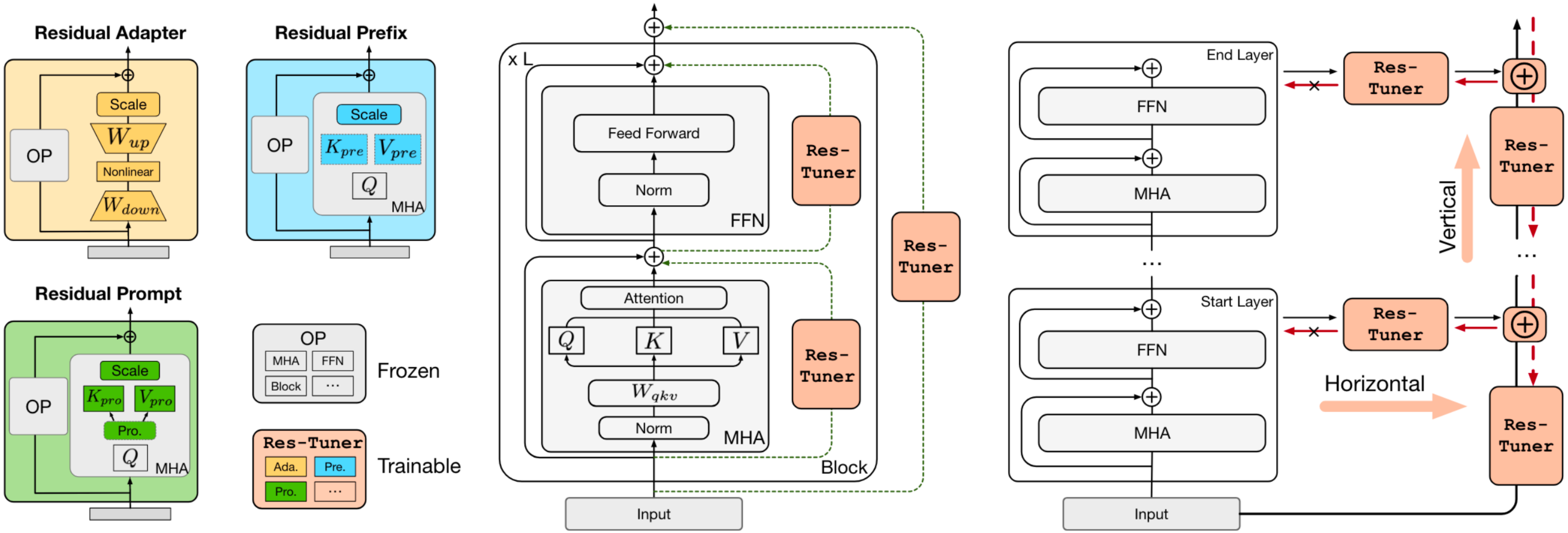
$$\mathrm{MHA}_{\mathrm{pro}} = \mathrm{Attn}\left([\boldsymbol{x}; \boldsymbol{x}_{pro}]\boldsymbol{W}_q, [\boldsymbol{x}; \boldsymbol{x}_{pro}]\boldsymbol{W}_k, [\boldsymbol{x}; \boldsymbol{x}_{pro}]\boldsymbol{W}_v\right)$$

$$\mathrm{MHA}_{\mathrm{pro}} = [(1-\lambda)\underbrace{\mathrm{Attn}\left(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}\right)}_{\text{original attention}} + \lambda\underbrace{\mathrm{Attn}\left(\boldsymbol{Q}, \boldsymbol{K}_{pro}, \boldsymbol{V}_{pro}\right)}_{\text{prompt attention in parallel}}; \boldsymbol{D}]$$

$$\mathrm{FFN}_{\mathrm{adapter}} = \underbrace{\mathrm{FFN}(\boldsymbol{x})}_{\text{original module}} + \underbrace{\phi(\mathrm{FFN}(\boldsymbol{x})\boldsymbol{W}_{down})\boldsymbol{W}_{up}}_{\text{adapter module in parallel}}$$

$$\boxed{\boldsymbol{x}' = \mathrm{OP}(\boldsymbol{x}) + \texttt{Res-Tuner}(\boldsymbol{x})}$$

Res-Tuning framework



Flexible	Parameter-Efficient	Memory & Parameter-Efficient
Res-Tuner : unbinds tuners from backbone	*Res-Tuning* : unified formulation	*Res-Tuning-Bypass* : backpropagation only on Bypass

Discriminative tasks

Transfer Learning

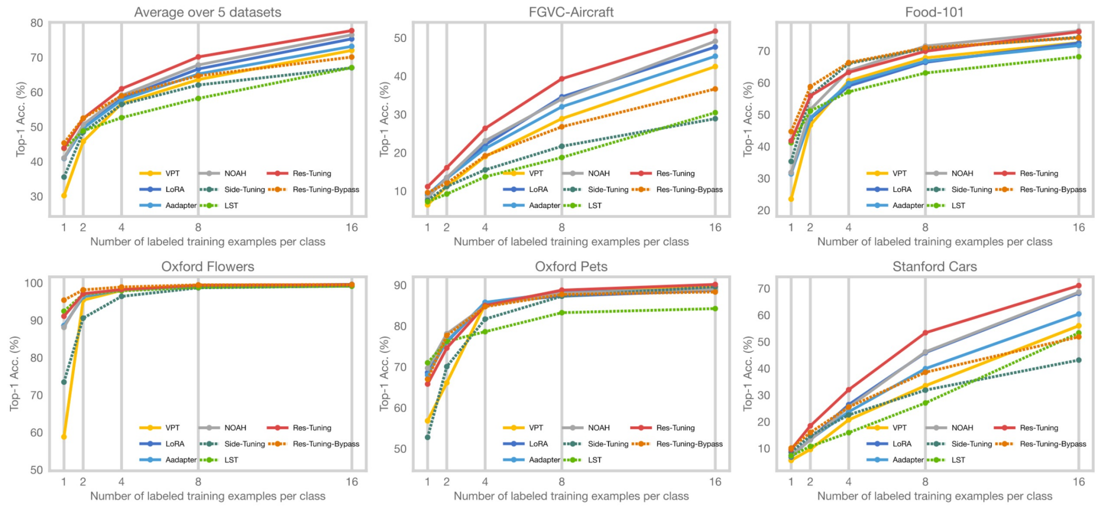
	Natural							Specialized				Structured											
	CIFAR-100	Caltech101	DTD	Flowers102	Pets	SVHN	Sun397	Camelyon	EuroSAT	Resisc45	Retinopathy	Clevr-Count	Clevr-Dist	DMLab	KITTI-Dist	dSpr-Loc	dSpr-Ori	sNORB-Azim	sNORB-Elev	Group Mean	All Mean	Param. (M)	Mem. (GB)
Traditional methods																							
Full	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	68.96	65.57	85.84	9.40
Linear	63.4	85.0	63.2	97.0	86.3	36.6	51.0	78.5	87.5	68.6	74.0	34.3	30.6	33.2	55.4	12.5	20.0	9.6	19.2	57.64	52.94	0.04	3.09
Parameter-efficient tuning methods																							
Adapter [24]	74.2	85.7	62.7	97.8	87.2	36.4	50.7	76.9	89.2	73.5	71.6	45.2	41.8	31.1	56.4	30.4	24.6	13.2	22.0	60.52	56.35	1.82	6.53
LoRA [26]	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	74.60	72.30	0.29	6.88
VPT-Deep [27]	**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	71.96	69.43	0.60	8.13
SSF [41]	69.0	92.6	**75.1**	**99.4**	91.8	**90.2**	52.9	**87.4**	95.9	**87.4**	75.5	75.9	62.3	**53.3**	80.6	77.3	54.9	29.5	37.9	75.69	73.10	0.24	7.47
NOAH [79]	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**	75.48	73.25	0.42	7.27
Res-Tuning	75.2	**92.7**	71.9	99.3	**91.9**	86.7	**58.5**	86.7	95.6	85.0	74.6	80.2	63.6	50.6	80.2	**85.4**	**55.7**	31.9	42.0	**76.32**	**74.10**	0.55	8.95
Memory-efficient tuning methods																							
Side-Tuning [78]	60.7	60.8	53.6	95.5	66.7	34.9	35.3	58.5	87.7	65.2	61.0	27.6	22.6	31.3	51.7	8.2	14.4	9.8	21.8	49.91	45.65	9.59	3.48
LST† [65]	58.0	87.1	66.2	99.1	89.7	63.2	52.6	81.9	92.2	78.5	69.4	68.6	56.1	38.8	**73.4**	72.9	30.5	16.6	31.0	67.56	64.52	0.89	5.13
Res-Tuning-Bypass	**64.5**	**88.8**	**73.2**	**99.4**	**90.6**	**63.5**	**57.2**	**85.5**	**95.2**	**82.4**	**75.2**	**70.4**	**61.0**	**40.2**	66.8	**79.2**	**52.6**	**26.0**	**49.3**	**72.32**	**69.51**	0.42	4.73

VTAB-1K Benchmark

Method	Acc.	Param. (M)	Mem.
Full	89.12	85.9 (100%)	9.02G
Linear	85.95	0.07 (0.08%)	2.72G
Parameter-efficient tuning methods			
MAM-Adapter† [19]	91.70	10.08 (11.72%)	9.57G
AdaptFormer [7]	91.86	1.26 (1.46%)	6.32G
`Res-Tuning`	**93.25**	0.48 (0.55%)	6.85G
Memory-efficient tuning methods			
Side-Tuning [82]	87.16	9.62 (11.18%)	3.48G
LST† [68]	88.72	0.93 (1.08%)	5.26G
`Res-Tuning-Bypass`	**89.33**	0.46 (0.53%)	4.72G

CIFAR-100

Few-Shot Learning



Fine-Grained Visual Recognition (FGVC) Datasets

Domain Generalization

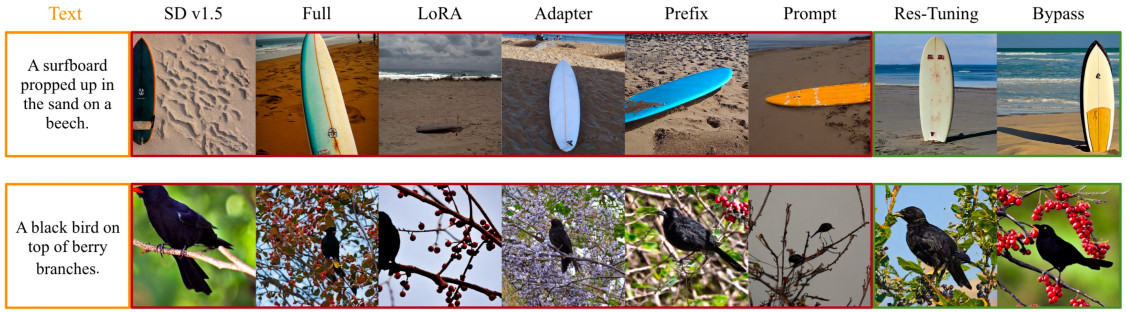
	Source	Target				
	ImageNet	IN-V2	IN-Sketch	IN-A	IN-R	Mean
Parameter-efficient tuning methods						
Adapter [25]	70.5	59.1	16.4	5.5	22.1	25.8
VPT [28]	70.5	58.0	18.3	4.6	23.2	26.0
LoRA [27]	70.8	59.3	20.0	6.9	23.3	27.4
NOAH [83]	71.5	66.1	24.8	11.9	28.5	32.8
Res-Tuning	**78.04**	**66.58**	**29.23**	**13.15**	**29.01**	**34.50**
Memory-efficient tuning methods						
Side-Tuning [82]	74.57	62.52	23.55	10.37	25.06	30.38
LST [68]	70.00	57.04	14.39	7.21	17.02	23.92
Res-Tuning-Bypass	**77.30**	**65.23**	**27.39**	**10.66**	**26.45**	**32.43**

ImageNet and variants

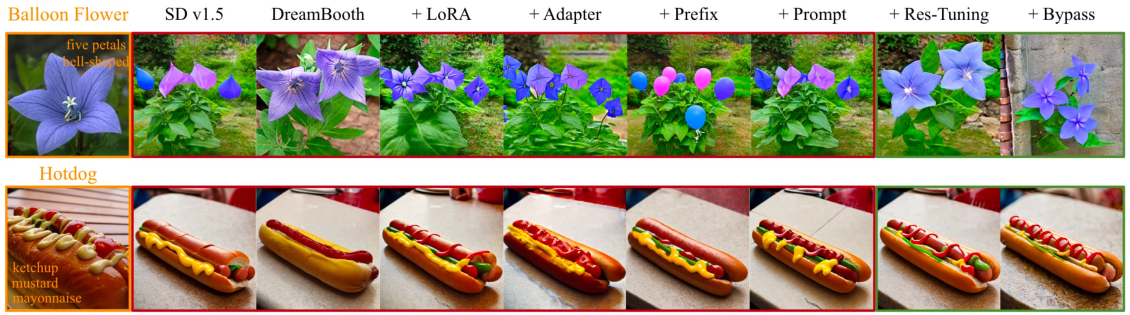
Generative task

Method	FID	Param. (M)	Mem. (GB)	Train (Hour/Epoch)
SD v1.5	15.48	-	-	-
+ Full	14.85	862 (100%)	72.77	1.98
+ LoRA	14.50	9.96 (1.15%)	61.03	1.42
+ Adapter	14.73	2.51 (0.29%)	54.30	1.30
+ Prefix	15.36	4.99 (0.58%)	64.91	2.20
+ Prompt	14.90	**1.25** (0.14%)	63.70	2.17
+ Res-Tuning	**13.96**	2.54 (0.29%)	54.49	1.38
+ Res-Tuning Bypass	14.89	3.76 (0.44%)	**21.35**	**0.82**

Performance and Efficiency Comparison on COCO2017 Dataset



Qualitative Results on COCO2017 Validation Set



Qualitative Results on Fine-grained Dataset

Thank you !