Efficient Adaptation of Large Vision Transformer via Adapter Re-Composing

Supplementary Materials

In the supplementary materials involving our work, we demonstrate detailed dataset settings, supplemental insights and analysis, extra experimental details, supplemental experiments, and broader impacts, including:

- A **Detailed descriptions for datasets and implementation**

- B **Insights of architecture design**

- C **Parameter size analysis**

- D **Experimental details on larger-scale and hierarchical ViT backbones**

- E **Experimental details on ablation studies**

- F **Expanded experiments with self-supervised pre-training**

- G **Broader impacts**

Due to the limitation that the file "Supplementary Materials.zip" larger than 100MB cannot be uploaded on OpenReview, the supplementary materials only upload the code for the project. Please refer to the anonymous link `https://drive.google.com/file/d/1ZblHbYF1JrOuOGeTLI4uII6GHt3CV3I2/view` to obtain the complete code, datasets, and models.

13

A Detailed descriptions for datasets and implementation

We describe the details of visual adaptation classification tasks in Table 6 (FGVC) and 7 (VTAB-1k), including the class number and the train/val/test sets.

Table 6: Dataset statistics for FGVC. "*" denotes the train/val split of datasets following the dataset setting of VPT models [6].

Dataset	Description	Classes	Train size	Val size	Test size
CUB-200-2011 [31]	Fine-grained bird species recognition	200	5,394*	600*	5,794
NABirds [32]	Fine-grained bird species recognition	555	21,536*	2,393*	24,633
Oxford Flowers [33]	Fine-grained flower species recognition	102	1,020	1,020	6,149
Stanford Dogs [34]	Fine-grained dog species recognition	120	10,800*	1,200*	8,580
Stanford Cars [35]	Fine-grained car classificatio	196	7,329*	815*	8,041

Table 7: Dataset statistics for VTAB-1k [36].

Dataset	Description	Classes	Train size	Val size	Test size
CIFAR-100		100			10,000
Caltech101		102			6,084
DTD		47			1,880
Flowers102	Natural	102	800/1,000	200	6,149
Pets		37			3,669
SVHN		10			26,032
Sun397		397			21,750
Patch Camelyon		2			32,768
EuroSAT		10			5,400
Resisc45	Specialized	45	800/1,000	200	6,300
Retinopathy		5			42,670
Clevr/count		8			15,000
Clevr/distance		6			15,000
DMLab		6			22,735
KITTI/distance		4			711
dSprites/location	Structured	16	800/1,000	200	73,728
dSprites/orientation		16			73,728
SmallNORB/azimuth		18			12,150
SmallNORB/elevation		9			12,150

Table 8 summarizes the detailed configurations we used for experiments. As mentioned in Section 4.1, we utilize grid search to select hyper-parameters such as learning rate, weight decay, batch size, and adapter dropout, using the validation set of each task. Note that we also apply dropout to the middle features produced by our ARC method, which we term as "adapter dropout". Specifically, during the ARC process, we randomly drop partial features before up-projection.

B Insights of architecture design

Similar to Fig. 3, we present more visualization results of singular value distribution of adaptation matrices $\mathbf{W}_{\text{full}} \in \mathbb{R}^{D \times D}$ learned without the bottleneck operation. As shown in Fig. 4, the singular value distribution of adaptation matrices learned on *DTD* downstream task exhibits a power-law

14

Table 8: The implementation details of configurations such as optimizer and hyper-parameters. We select the best hyper-parameters for each download task via using grid search.

Optimizer	AdamW
Learning Rate	{0.2, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0001}
Weight Decay	{0.05, 0.01, 0.005, 0.001, 0}
Batch Size	{256, 128, 32}
Adapter Dropout	{0.8, 0.5, 0.1, 0}
Learning Rate Schedule	Cosine Decay
Training Epochs	100
Warmup Epochs	10

distribution across various layers in the downstream tasks. This finding provides further support for our research motivation.

(a) Singular value distribution of MHA adapter.

(b) Singular value distribution of FFN adapter.
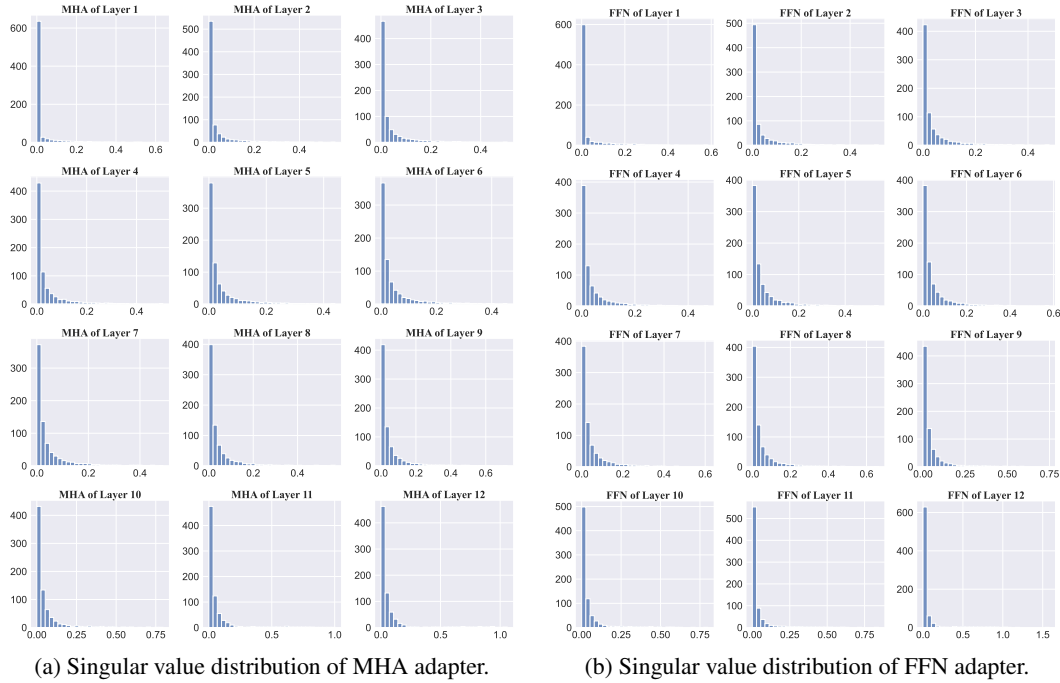
Figure 4: Singular value distribution of adaptation matrices without the bottleneck structure. Two adaptation matrices of both MHA and FFN blocks are fine-tuned on the *DTD* downstream task. The X-axis represents the singular values, while the Y-axis represents the count of singular values within specific ranges.

C Parameter size analysis

To showcase the parameter-efficiency of our ARC method, we compare its parameter size with other popular lightweight adaptation methods (Table 9), including Adapter [7], VPT [6], LoRA [24], and SSF [9]. Adapter [7] adds two linear projections to each encoder layer during fine-tuning, resulting in the introduction of $2 \cdot D \cdot D' \cdot L$ learnable parameters, where D' denotes the hidden dimensionality of the linear projections. Furthermore, due to the presence of non-linear activations in Adapter, the additional parameters contribute to supernumerary overhead during the inference phase. VPT [6] incorporates m prompts into input space, leading to an increase of $m \cdot D$ parameters for VPT-Shallow and $m \cdot D \cdot L$ parameters for VPT-Deep. In contrast to Adapter, both LoRA [24] and SSF [9] employ linear adaptation methods without incorporating non-linear functions. This design choice allows them to leverage re-parameterization benefits, thereby mitigating additional computations during inference. Specifically, the adaptation matrix of LoRA, which consists of a down-projection and an up-projection, introduces $2 \cdot w \cdot D \cdot D' \cdot L$ learnable parameters, where w denotes the number of attention matrices undergoing adaptation. SSF inserts linear scaling and shifting coefficients after

15

16

o operations, resulting in an addition of $2 \cdot o \cdot D \cdot L$ extra parameters. The proposed ARC method offers additional parameter compression by sharing symmetric projection matrices across different layers. This approach introduces only $D \cdot D'$ parameters. Additionally, we learn low-dimensional re-scaling coefficients and bias terms for each layer, resulting in a total of $(D' + D) \cdot L$ additional parameters. Overall, the number of parameters in our default ARC is $2 \cdot ((D \cdot D') + (D' + D) \cdot L)$.

Table 9: Comparison of the additional parameter size in both fine-tuning and inference stages with other lightweight adaptation methods.

Method Stage	Adapter [7]	VPT-Shallow [6]	VPT-Deep [6]	LoRA [24]	SSF [9]	ARC
Fine-Tuning	$2 \cdot D \cdot D' \cdot L$	$m \cdot D$	$m \cdot D \cdot L$	$2 \cdot w \cdot D \cdot D' \cdot L$	$2 \cdot o \cdot D \cdot L$	$2 \cdot (D \cdot D' + (D' + D) \cdot L)$
Inference	$2 \cdot D \cdot D' \cdot L$	$m \cdot D$	$m \cdot D \cdot L$	0	0	0

We also compare the parameter size with lightweight adaptation methods on backbones of different scales, as shown in Fig. 5. Our ARCs demonstrate parameter efficiency across various model sizes, comparable to VPT-Shallow [6]. However, the unique advantage of our approach lies in its ability to effectively balance lower overheads and maintain competitive performance. Furthermore, the parameter count of our ARC remains stable even as the model scale increases, showcasing the scalability of our method with minimal additional resource consumption.
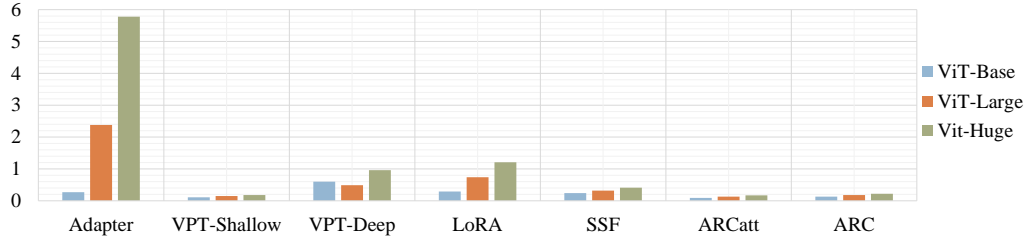
Figure 5: The parameter size comparison of lightweight adaptation methods on ViT Backbones of Different Scales. The X-axis represents different adaptation methods, while the Y-axis represents the parameter size in Million (M).
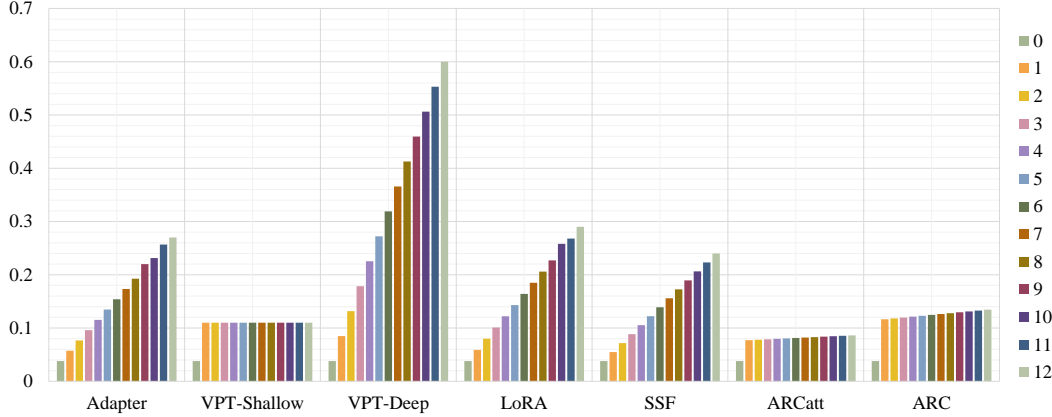
Figure 6: The parameter size comparison with lightweight adaptation methods with a different number of inserted layers. The X-axis represents different adaptation methods, while the Y-axis represents the parameter size in Million (M).

Thanks to our adaptation parameter sharing strategy, the ARC method avoids a linear increase in the number of learnable parameters as the number of layers grows. We employ ViT-B as the backbone and integrate adapters into different layers. As shown in Fig.6, in contrast to other adaptation methods, both our ARCs and VPT-Shallow[6] effectively manage parameter growth as the number of inserted layers increases, but only our methods achieve promising performance without significant cost escalation. This highlights the scalability and effectiveness advantages of our ARCs.

D   Experimental details on larger-scale and hierarchical ViT backbones

Table 10, 11 and 12 respectively display the comprehensive results of the comparison conducted in Section 4.2 among ViT-Large, ViT-Huge, and Swin-Base models.

Table 10: This table is extended from Table 3a in Section 4.2 and describes the detailed experimental results of the performance comparison on VTAB-1k using ViT-Large pre-trained on ImageNet-21k as the backbone.

Method	__Natural__ CIFAR-100	Caltech101	DTD	Flowers102	Pets	SVNH	Sun397	Mean	__Specialized__ Camelyon	EuroSAT	Resisc45	Retinopathy	Mean	__Structured__ Clevr-Count	Clevr-Dist	DMLab	KITTI-Dist	dSpr-Loc	dSpr-Ori	sNORB-Azim	sNORB-Ele	Mean	Mean Total	Params.(M)
Full fine-tuning	68.6	84.3	58.6	96.3	86.5	87.5	41.4	74.7	82.6	95.9	82.4	74.2	83.8	55.4	55.0	42.2	74.2	56.8	43.0	28.5	29.7	48.1	65.4	303.4
Linear probing	72.2	86.4	63.6	97.4	85.8	38.1	52.5	70.9	76.9	87.3	66.6	45.4	69.1	28.2	28.0	34.7	54.0	10.6	14.2	14.6	21.9	25.8	51.5	0.05
Adapter [7]	75.3	84.2	54.5	97.4	84.3	31.3	52.9	68.6	75.8	85.1	63.4	69.5	73.5	35.4	34.1	30.8	47.1	30.4	23.4	10.8	19.8	29.0	52.9	2.38
Bias [37]	71.0	82.4	51.3	96.3	83.2	59.5	49.9	70.5	72.9	87.9	63.1	71.3	73.8	51.2	50.7	33.5	54.8	65.9	37.3	13.7	22.2	41.2	58.9	0.32
VPT-Shallow [6]	80.6	88.2	67.1	98.0	85.9	78.4	53.0	78.7	79.7	93.5	73.4	73.1	79.9	41.5	52.5	32.3	64.2	48.3	35.3	21.6	28.8	40.6	62.9	0.15
VPT-Deep [6]	84.1	88.9	70.8	98.8	90.0	89.0	55.9	82.5	82.5	96.6	82.6	73.9	83.9	63.7	60.7	46.1	75.7	83.7	47.4	18.9	36.9	54.1	70.8	0.49
LoRA [24]	75.8	89.8	73.6	99.1	90.8	83.2	57.5	81.4	86.0	95.0	83.4	75.5	85.0	78.1	60.5	46.7	81.6	76.7	51.3	28.0	35.4	57.3	72.0	0.74
ARC_att	75.6	89.9	72.2	99.0	90.4	89.0	57.5	81.9	86.1	95.0	85.4	76.0	85.6	75.0	60.1	48.0	80.9	77.0	51.3	27.2	35.6	56.9	72.2	0.13
ARC	76.2	89.6	73.4	99.1	90.3	90.9	56.5	82.3	85.0	95.7	85.9	75.8	85.6	78.6	62.1	46.7	76.7	75.9	53.0	30.2	35.2	57.3	72.5	0.18

Table 11: This table is extended from Table 3b in Section 4.2 and describes the detailed experimental results of the performance comparison on VTAB-1k using ViT-Huge pre-trained on ImageNet-21k as the backbone.

Method	__Natural__ CIFAR-100	Caltech101	DTD	Flowers102	Pets	SVNH	Sun397	Mean	__Specialized__ Camelyon	EuroSAT	Resisc45	Retinopathy	Mean	__Structured__ Clevr-Count	Clevr-Dist	DMLab	KITTI-Dist	dSpr-Loc	dSpr-Ori	sNORB-Azim	sNORB-Ele	Mean	Mean Total	Params.(M)
Full fine-tuning	58.7	86.5	55.0	96.5	79.7	87.5	32.5	70.9	83.1	95.5	81.9	73.8	83.6	47.6	53.9	37.8	69.9	53.8	48.6	30.2	25.8	46.0	63.1	630.9
Linear probing	64.3	83.6	65.2	96.2	83.5	39.8	43.0	67.9	78.0	90.5	73.9	73.4	79.0	25.6	24.5	34.8	59.0	9.5	15.6	17.4	22.8	26.1	52.7	0.06
Adapter [7]	69.4	84.4	62.7	97.2	84.2	33.6	45.3	68.1	77.3	86.6	70.8	71.1	76.4	28.6	27.5	29.2	55.2	10.0	15.2	11.9	18.6	24.5	51.5	5.78
Bias [37]	65.7	84.3	59.9	96.6	80.6	60.1	44.9	70.3	79.7	92.8	71.5	71.6	78.9	52.3	50.4	31.2	57.7	65.9	39.7	16.7	20.2	41.7	60.1	0.52
VPT-Shallow [6]	70.6	84.7	64.8	96.4	85.1	75.6	46.2	74.8	79.9	93.7	77.7	73.6	81.2	40.3	60.9	34.9	63.3	61.3	38.9	19.8	24.9	43.0	62.8	0.18
VPT-Deep [6]	76.9	87.2	66.8	97.5	84.8	85.5	46.5	77.9	81.6	96.3	82.5	72.8	83.3	50.4	61.2	43.9	76.6	79.5	50.1	24.7	31.5	52.2	68.2	0.96
LoRA [24]	63.0	89.4	68.1	98.0	87.0	85.2	48.7	77.1	82.2	94.3	83.1	74.2	83.5	68.6	65.0	44.8	76.4	70.8	48.8	30.4	38.3	55.4	69.3	1.21
ARC_att	65.5	89.1	69.9	98.0	87.5	89.1	48.8	78.3	83.4	94.5	84.5	74.4	84.2	73.2	66.6	45.6	76.2	78.3	51.2	32.1	37.6	57.6	70.8	0.17
ARC	67.6	90.2	69.5	98.4	87.9	90.8	49.6	79.1	84.5	94.9	85.1	74.6	84.8	75.2	66.7	46.2	76.4	44.2	51.1	32.2	37.7	53.7	69.6	0.22

Table 12: This table is extended from Table 4 in Section 4.2 and describes the detailed experimental results of the performance comparison on VTAB-1k using Swin-Base pre-trained on ImageNet-21k as the backbone.

Method	__Natural__ CIFAR-100	Caltech101	DTD	Flowers102	Pets	SVNH	Sun397	Mean	__Specialized__ Camelyon	EuroSAT	Resisc45	Retinopathy	Mean	__Structured__ Clevr-Count	Clevr-Dist	DMLab	KITTI-Dist	dSpr-Loc	dSpr-Ori	sNORB-Azim	sNORB-Ele	Mean	Mean Total	Params.(M)
Full fine-tuning	72.2	88.0	71.4	98.3	89.5	89.4	45.1	79.1	86.6	96.9	87.7	73.6	86.2	75.7	59.8	54.6	78.6	79.4	53.6	34.6	40.9	59.7	72.4	86.9
Linear probing	61.4	90.2	74.8	95.5	90.2	46.9	55.8	73.5	81.5	90.1	82.1	69.4	80.8	39.1	35.9	40.1	65.0	20.3	26.0	14.3	27.6	33.5	58.2	0.05
MLP-4 [6]	54.9	87.4	71.4	99.5	89.1	39.7	52.5	70.6	80.5	90.9	76.8	74.4	80.7	60.9	38.8	40.2	66.5	9.4	21.1	14.5	28.8	31.2	57.7	4.04
Partial [6]	60.3	88.9	72.6	98.7	89.3	50.5	51.5	73.1	82.8	91.7	80.1	72.3	81.7	34.3	35.5	43.2	77.1	15.8	26.2	19.1	28.4	35.0	58.9	12.65
Bias [37]	73.1	86.8	65.7	97.7	87.5	56.4	52.3	74.2	80.4	91.6	76.1	72.5	80.1	47.3	48.5	34.7	66.3	57.6	36.2	17.2	31.6	42.4	62.1	0.25
VPT-Shallow [6]	78.0	91.3	77.2	99.4	90.4	68.4	54.3	79.9	80.1	93.9	83.0	72.7	82.5	40.8	43.9	34.1	63.2	28.4	44.5	21.5	26.3	37.8	62.9	0.05
VPT-Deep [6]	79.6	90.8	78.0	99.5	91.4	46.5	51.7	76.8	84.9	96.2	85.0	72.0	84.5	67.6	59.4	50.1	74.1	74.4	50.6	25.7	25.7	53.4	67.7	0.22
ARC_att	67.2	89.7	74.7	99.5	89.7	88.5	52.7	80.3	88.1	95.9	85.7	77.2	86.7	76.5	58.5	52.1	82.8	89.4	56.4	27.5	35.1	59.8	73.0	0.16
ARC	62.5	90.0	71.9	99.2	87.8	90.7	51.1	79.0	89.1	95.8	84.5	77.0	86.6	75.4	57.4	53.4	83.1	91.7	55.2	31.6	31.8	59.9	72.6	0.27

E   Experimental details on ablation studies

Table 13, 14, 15 and 16 display the complete results of the ablation studies in Section 4.3.

17

Table 13: This table is extended from Table 5a in Section 4.3 and describes the detailed experimental content of the performance comparison among different bottleneck dimensionality.

Dimension	CIFAR-100	Caltech101	DTD	Flowers102	Pets	SVNH	Sun397	Mean	Camelyon	EuroSAT	Resisc45	Retinopathy	Mean	Clevr-Count	Clevr-Dist	DMLab	KITTI-Dist	dSpr-Loc	dSpr-Ori	sNORB-Azim	sNORB-Ele	Mean	Mean Total	Params.(M)
				Natural							Specialized							Structured							
10	**72.2**	88.4	71.2	98.7	**91.1**	89.4	54.7	80.9	84.7	95.6	86.0	**75.8**	85.6	80.1	65.9	48.8	80.5	75.5	48.3	30.2	38.6	58.5	72.4	0.07
50	**72.2**	90.1	72.7	99.0	91.0	91.9	54.4	81.6	84.9	95.7	86.7	75.8	85.8	80.7	67.1	48.7	81.6	79.2	51.0	31.4	39.9	60.0	73.4	0.13
100	71.3	90.0	73.0	99.0	90.7	91.8	55.1	81.6	85.1	96.3	86.1	75.4	85.7	80.8	67.2	49.0	79.3	74.8	50.1	34.0	39.1	59.3	73.1	0.21
200	70.5	89.3	72.9	99.1	89.8	91.9	54.9	81.2	84.9	95.3	84.0	75.7	85.0	80.0	67.8	48.9	76.8	50.8	51.3	34.4	39.1	56.1	71.4	0.36

Table 14: This table is extended from Table 5b in Section 4.3 and describes the detailed experimental content of the performance comparison among different adapter positioning.

Location	CIFAR-100	Caltech101	DTD	Flowers102	Pets	SVNH	Sun397	Mean	Camelyon	EuroSAT	Resisc45	Retinopathy	Mean	Clevr-Count	Clevr-Dist	DMLab	KITTI-Dist	dSpr-Loc	dSpr-Ori	sNORB-Azim	sNORB-Ele	Mean	Mean Total	Params.(M)
				Natural							Specialized							Structured							
Before MHA	70.1	**90.5**	70.5	98.8	90.8	88.6	53.6	80.4	84.6	95.5	86.6	75.5	85.6	79.0	65.6	48.6	81.3	75.1	48.7	29.1	39.6	58.4	72.2	0.08
After MHA	67.0	88.9	69.8	98.8	90.8	82.2	52.3	78.5	84.1	94.6	85.1	74.8	84.8	77.4	60.1	44.3	77.1	61.2	45.7	23.0	35.6	53.0	69.1	0.08
Before FFN	70.8	89.4	71.0	99.0	89.9	86.9	53.9	80.1	85.5	94.7	84.9	75.6	85.2	77.3	63.6	46.5	77.5	70.3	48.4	27.6	37.3	56.0	71.1	0.08
After FFN	66.7	88.2	69.6	98.6	90.2	82.5	52.9	78.4	83.6	94.8	85.3	75.5	84.8	77.9	63.1	44.1	76.7	57.9	47.0	22.6	33.9	52.9	69.0	0.08
Before MHA & FFN	72.2	90.1	72.7	99.0	91.0	91.9	54.4	81.6	84.9	95.7	86.7	75.8	85.8	80.7	67.1	48.7	81.6	79.2	51.0	31.4	39.9	60.0	73.4	0.13
After MHA & FFN	70.5	89.9	71.3	99.0	91.4	86.9	53.5	80.4	84.7	94.9	86.4	76.0	85.5	80.3	62.8	46.8	80.9	66.9	49.6	28.4	36.4	56.5	71.4	0.13

Table 15: This table is extended from Table 5c in Section 4.3 and describes the detailed experimental content of the performance comparison among different parameter sharing strategy.

Strategy	CIFAR-100	Caltech101	DTD	Flowers102	Pets	SVNH	Sun397	Mean	Camelyon	EuroSAT	Resisc45	Retinopathy	Mean	Clevr-Count	Clevr-Dist	DMLab	KITTI-Dist	dSpr-Loc	dSpr-Ori	sNORB-Azim	sNORB-Ele	Mean	Mean Total	Params.(M)
				Natural							Specialized							Structured							
non-intra + non-inter	70.1	**91.1**	71.5	**99.2**	90.6	91.9	54.6	81.3	84.8	95.5	86.4	75.4	85.5	81.1	66.1	50.1	78.6	80.3	51.5	35.8	40.6	60.5	73.4	0.98
intra + inter*	72.9	89.8	72.1	98.8	91.0	90.7	54.6	81.4	85.8	95.5	86.3	75.6	85.8	80.3	66.5	48.8	79.6	77.0	50.7	30.9	39.0	59.1	72.9	0.10
intra + inter	72.2	90.1	72.7	99.0	91.0	91.9	54.4	81.6	84.9	95.7	86.7	75.8	85.8	80.7	67.1	48.7	81.6	79.2	51.0	31.4	39.9	60.0	73.4	0.13
non-intra + inter	72.9	89.5	72.9	98.8	90.6	90.2	55.8	81.5	86.2	95.5	86.2	75.9	86.0	81.1	67.1	48.3	81.0	78.5	50.6	31.5	41.9	60.0	73.4	0.21

Table 16: This table is extended from Table 5d in Section 4.3 and describes the detailed experimental content of the performance comparison among different adapter insertion.

Layer Form	CIFAR-100	Caltech101	DTD	Flowers102	Pets	SVNH	Sun397	Mean	Camelyon	EuroSAT	Resisc45	Retinopathy	Mean	Clevr-Count	Clevr-Dist	DMLab	KITTI-Dist	dSpr-Loc	dSpr-Ori	sNORB-Azim	sNORB-Ele	Mean	Mean Total	Params.(M)
				Natural							Specialized							Structured							
1∼6 & sequential	69.0	88.1	70.2	98.4	90.0	89.8	52.3	79.7	84.1	94.5	85.4	75.5	84.9	80.2	67.3	46.4	78.8	74.4	48.1	29.1	37.6	57.7	71.5	0.126
7∼12 & sequential	57.9	88.2	68.4	98.3	89.2	70.3	52.1	74.9	82.2	94.3	84.3	76.4	84.3	77.0	58.1	45.4	75.3	74.0	42.7	21.0	34.9	53.6	67.9	0.126
1∼12 & sequential	72.2	90.1	72.7	99.0	91.0	91.9	54.4	81.6	84.9	95.7	86.7	75.8	85.8	80.7	67.1	48.7	81.6	79.2	51.0	31.4	39.9	60.0	73.4	0.133
1∼12 & parallel	70.7	90.9	71.5	98.9	91.1	86.1	53.8	80.4	83.5	95.1	85.6	75.4	84.9	76.6	64.1	45.9	76.9	62.0	46.0	25.3	37.2	54.3	70.4	0.133

F Expanded experiments with self-supervised pre-training

In addition to the models pre-trained with supervised objectives in Section 4, we also conduct experiments with self-supervised pre-training approaches: MAE [2] and Moco V3 [23]. Specifically, We utilize MAE [2] and Moco V3 [23] self-supervised pre-trained ViT-B as the backbone and evaluate the performance of our ARC on VTAB-1k. The results of MAE and Moco V3 self-supervised models are presented in Table 17 and Table 18, respectively. We observe that our ARC still exhibits competitive performance on two self-supervised ViTs. In addition, our ARC method outperforms other adaptation methods: Adapter[7] and LoRA [24] on the majority of downstream tasks. Surprisingly, the ARC_{att} with smaller learnable parameters even surpasses the ARC across different self-supervised pre-trained models. A possible explanation could be that ARC_{att} contains fewer parameters, which allows it to effectively prevent overfitting.

18

Table 17: Performance comparison on VTAB-1k using MAE self-supervised pre-trained ViT-Base as backbone.

	Natural								Specialized					Structured									Mean Total	Params(M)
Dataset Method	CIFAR-100	Caltech101	DTD	Flowers102	Pets	SVNH	Sun397	Mean	Camelyon	EuroSAT	Resisc45	Retinopathy	Mean	Clevr-Count	Clevr-Dist	DMLab	KITTI-Dist	dSpr-Loc	dSpr-Ori	sNORB-Azim	sNORB-Ele	Mean		
Full fine tuning	24.6	84.2	56.9	72.7	74.4	86.6	15.8	59.3	81.8	**94.0**	72.3	70.6	79.7	67.0	59.8	45.2	75.3	72.5	47.5	30.2	33.0	53.8	61.3	85.80
Linear	8.7	41.5	20.6	19.2	11.3	22.3	8.6	18.9	76.5	68.6	16.6	53.2	53.7	33.6	32.5	23.0	51.1	13.0	9.9	8.5	17.9	23.7	28.2	0.04
Bias [37]	22.4	82.6	49.7	66.2	67.7	69.0	24.3	54.6	78.7	91.4	60.0	72.6	75.7	65.9	51.0	35.0	69.1	70.8	37.6	21.5	30.7	47.7	56.1	0.14
Adapter [7]	**35.1**	85.0	56.5	66.6	71.3	45.0	**24.8**	54.9	76.9	87.1	63.5	73.3	75.2	43.8	49.5	31.2	61.7	59.3	23.3	13.6	29.6	39.0	52.5	0.76
VPT-Shallow [6]	21.9	76.2	54.7	58.0	41.3	16.1	15.1	40.0	74.0	69.5	58.9	72.7	68.8	40.3	44.7	27.9	60.5	11.8	11.0	12.4	16.3	28.1	41.2	0.04
VPT-Deep [6]	8.2	55.2	58.0	39.3	45.2	19.4	21.9	35.3	77.9	91.0	45.4	73.6	72.0	39.0	40.9	30.6	53.9	21.0	12.1	11.0	14.9	27.9	39.9	0.06
LoRA [24]	31.8	88.4	59.9	81.7	85.3	90.3	23.7	65.9	84.2	92.5	76.2	75.4	82.1	85.9	64.1	49.4	82.8	83.9	51.8	34.6	41.3	61.7	67.5	0.30
ARC_att	34.8	89.3	62.0	85.9	84.4	91.1	24.8	**67.4**	85.8	93.5	81.3	75.6	**84.1**	84.0	63.5	51.2	83.0	89.1	54.0	34.2	43.0	**62.7**	**69.0**	0.09
ARC	31.3	89.3	61.2	85.9	83.1	91.6	24.4	66.7	86.0	94.0	80.4	74.8	83.8	85.8	64.6	50.5	82.8	82.8	53.5	36.3	39.7	62.0	68.3	0.13

Table 18: Performance comparison on VTAB-1k using Moco V3 self-supervised pre-trained ViT-Base as backbone.

	Natural								Specialized					Structured									Mean Total	Params(M)
Dataset Method	CIFAR-100	Caltech101	DTD	Flowers102	Pets	SVNH	Sun397	Mean	Camelyon	EuroSAT	Resisc45	Retinopathy	Mean	Clevr-Count	Clevr-Dist	DMLab	KITTI-Dist	dSpr-Loc	dSpr-Ori	sNORB-Azim	sNORB-Ele	Mean		
Full fine tuning	57.6	91.0	64.6	91.5	79.9	89.8	29.1	72.0	85.1	**96.4**	83.1	74.3	84.7	55.1	56.9	44.7	77.9	63.8	49.0	31.5	36.9	52.0	66.2	85.69
Linear	62.9	85.1	68.8	87.0	85.8	41.8	**40.9**	67.5	80.3	93.6	77.9	72.6	81.1	42.3	34.8	36.4	59.2	10.1	22.7	12.6	24.7	30.3	54.7	0.04
Bias [37]	65.5	89.2	62.9	88.9	80.5	82.7	40.5	72.9	80.9	95.2	77.7	70.8	81.1	71.4	59.4	39.8	77.4	70.2	49.0	17.5	42.8	53.4	66.4	0.14
Adapter [7]	**73.0**	88.2	69.3	90.7	87.4	69.9	**40.9**	74.2	82.4	93.4	80.5	74.3	82.7	55.6	56.1	39.1	73.9	60.5	40.2	19.0	37.1	47.7	64.8	0.98
VPT-Shallow [6]	68.3	86.8	**69.7**	90.0	59.7	56.9	39.9	67.3	81.7	94.7	78.9	73.8	82.3	34.3	56.8	40.6	49.1	40.4	31.8	13.1	34.4	37.6	57.9	0.05
VPT-Deep [6]	70.1	88.3	65.9	88.4	85.6	57.8	35.7	70.3	83.1	93.9	81.2	74.0	83.0	48.5	55.8	37.2	64.6	52.3	26.5	19.4	34.8	42.4	61.2	0.05
LoRA [24]	58.8	90.8	66.0	91.8	88.1	87.6	40.6	74.8	86.4	95.3	83.4	75.5	85.1	83.0	64.6	51.3	81.9	83.2	47.5	32.4	47.3	61.4	71.3	0.30
ARC_att	59.3	90.9	67.7	**93.6**	89.2	90.5	40.3	75.9	87.1	94.8	85.4	75.5	85.7	84.0	64.9	52.5	83.1	88.2	53.4	33.0	46.2	63.2	72.6	0.09
ARC	60.0	**91.3**	67.9	92.8	89.3	91.4	40.9	**76.2**	87.5	95.6	86.1	75.6	**86.2**	83.0	64.2	50.2	80.6	85.0	53.0	34.6	47.4	62.3	72.4	0.13

G Broader impacts

Efficient usability. Unlike previous approaches, our method incorporates a parameter sharing scheme across different layers of the model, resulting in a significant reduction in the number of parameters that need to be fine-tuned. This approach allows us to maintain competitive performance while achieving parameter efficiency. By maximizing the utilization of large-scale pre-trained models, our ARC methods offer enhanced usability and practicality in various applications.

Environmental-friendly consumption. In addition to the reduction in computational overheads, another significant benefit of our method is the positive impact on carbon emissions reduction and environmental protection. By optimizing the computational efficiency of the model, we minimize the energy consumption required during the training and deployment of the model. This reduction in energy consumption leads to a decrease in carbon emissions, contributing to environmental sustainability. Our method not only delivers improved performance and efficiency but also aligns with the larger goal of mitigating the environmental impact of AI technologies.

Ethical Considerations. Our model focuses on utilizing the representation and generalization capacity obtained from large-scale pre-trained datasets and models. However, it is crucial to acknowledge that if the pre-training datasets contain bias or illegal information, there is a risk of inheriting such issues into our model.

In order to address this concern, it becomes imperative to explore research directions that aim to identify and prevent privacy leakage and correct model bias. This involves developing robust mechanisms to detect and mitigate bias in training data, as well as implementing privacy-preserving techniques to safeguard sensitive information.

19