Low-Rank Rescaled Vision Transformer Fine-Tuning: A Residual Design Approach

Supplementary Material

A. Detailed dataset statistic

We describe the details of visual adaptation classification tasks we used in Table 1 (FGVC) and Table 2 (VTAB-1k), including the class number and the train/val/test sets. we employ the split following VPT [13].

B. Detailed configuration

Table 3 summarizes the detailed configurations we used for experiments. As mentioned in Section 4, we utilize grid search to select hyper-parameters such as learning rate, weight decay, batch size, and dropout rate, using the validation set of each task. AugReg [25] provides a robust initialization for the pre-training model with varying data augmentation and regularization. Despite the need for small initializations in many PEFT methods, RLRR maintains consistent performance under different initializations, as shown in Table 10.

C. Parameter size analysis

To showcase the parameter-efficiency of our RLRR method, we compare its parameter size with other popular lightweight adaptation methods (Table 4), including Adapter [9], VPT [13], LoRA [10], SSF [18] and ARC [5]. Adapter [9] uses two linear projections to construct a bottleneck structure for each layer, resulting in the introduction of $2 \cdot D \cdot D' \cdot L$ learnable parameters, where D' denotes the size of hidden dimension and L denotes the number of layers. Furthermore, due to the presence of non-linear activations in Adapter, this structure does not allow for re-parameterization, which leads to additional computational overhead in the inference. VPT [13] incorporates m prompts into input space, leading to an increase of $m \cdot D$ parameters for VPT-Shallow and $m \cdot D \cdot L$ for VPT-Deep. In contrast to Adapter, both LoRA [10] and SSF [18] employ linear adaptation methods without incorporating non-linear functions. This design 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 o operations, resulting in an addition of $2 \cdot o \cdot D^* \cdot L$ extra parameters. D^* denotes the dimension of weight matrix, where $D^* = 4 \cdot D$ in up-projection of FFN and $D^* = D$ in other cases. ARC offers additional parameter compression by sharing symmetric projection matrices across different layers. This approach introduces $D \cdot D'$ parameters for MHA and FFN. The low-dimensional re-scaling coefficients and bias terms result in a total of $(D' + D) \cdot L$ additional parameters. The proposed RLRR introduces dual-sided scaling tuning resulting in $3 \cdot o \cdot D^* \cdot L$ trainable parameters.

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

Table 5, 6 and 7 respectively display the comprehensive results of the comparison conducted in Section 4 among ViT-Large, ViT-Huge, and Swin-Base models.

E. Expanded experiments with self-supervised pre-training

In addition to the models pre-trained with supervision, we also conduct experiments with self-supervised pre-training approaches: MAE [8] and Moco V3 [3]. Specifically, We utilize MAE and Moco V3 self-supervised pre-trained ViT-B as the backbone and evaluate the performance of our RLRR on VTAB-1k. The results of MAE and Moco V3 self-supervised models are presented in Table 8 and Table 9, respectively. Based on these results, it is noted that our proposed RLRR continues to exhibit competitive performance on two self-supervised ViT models.

F. Flexibility of RLRR

LoRA, as a universal fine-tuning paradigm, has achieved remarkable performance across multiple tasks due to its flexibility. In this section, we will elaborate on how our proposed RLRR maintains the flexibility comparable to LoRA while achieving superior performance. LoRA adjusts the trainable parameter count by altering the sampling dimensions of the bottleneck structure. Similarly, RLRR can achieve the same adjustment. Initially, we remove the $\mathbf{W}$ in fine-tuning items $\triangle\mathbf{W} = \vec{\boldsymbol{s}}_{\text{left}} \odot \mathbf{W} \odot \vec{\boldsymbol{s}}_{\text{right}}^{\top}$, defining this baseline as $\mathbf{X}(\mathbf{W} + \mathbf{S}_{\text{left}}\mathbf{S}_{\text{right}}) + \vec{\boldsymbol{b}}^{\top} + \vec{\boldsymbol{f}}^{\top}$ to simulate the LoRA rank = 1 scenario.

After this, we can also introduce variations in the RLRR variant by modifying the dimension r of the parameter scaling in the expression $\mathbf{X}(\mathbf{W} + (\mathbf{S}_{\text{left}}\mathbf{S}_{\text{right}}) \odot \mathbf{W}) + \vec{\boldsymbol{b}}^{\top} + \vec{\boldsymbol{f}}^{\top}$, where $\mathbf{S}_{\text{left}} \in \mathbb{R}^{d \times r}$ and $\mathbf{S}_{\text{right}} \in \mathbb{R}^{r \times d}$. Through this modification, we can derive adaptation matrices with varying ranks to demonstrate the flexibility of adjustments similar

Table 1. Dataset statistics for FGVC. "*" denotes the train/val split of datasets following the dataset setting in VPT [13].

Dataset	Description	Classes	Train size	Val size	Test size
CUB-200-2011 [29]	Fine-grained bird species recognition	200	5,394*	600*	5,794
NABirds [27]	Fine-grained bird species recognition	555	21,536*	2,393*	24,633
Oxford Flowers [22]	Fine-grained flower species recognition	102	1,020	1,020	6,149
Stanford Dogs [15]	Fine-grained dog species recognition	120	10,800*	1,200*	8,580
Stanford Cars [7]	Fine-grained car classificatio	196	7,329*	815*	8,041

Table 2. Dataset statistics for VTAB-1k [32].

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

Table 3. 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}
Dropout Rate	{0, 0.1, 0.3, 0.5, 0.7}
Batch Size	{256, 128, 32}
Learning Rate Schedule	Cosine Decay
Training Epochs	100
Warmup Epochs	10

Table 4. Comparison of the additional parameter size in both fine-tuning and inference stages with other lightweight adaptation methods.

Method / Stage	Adapter [9]	VPT-Shallow [13]	VPT-Deep [13]	LoRA [10]	SSF [18]	ARC [5]	RLRR
Fine-Tuning	$2 \cdot D \cdot D' \cdot L$	$m \cdot D$	$m \cdot D \cdot L$	$2 \cdot w \cdot D \cdot D' \cdot L$	$2 \cdot o \cdot D^* \cdot L$	$2 \cdot (D \cdot D' + (D' + D) \cdot L)$	$3 \cdot o \cdot D^* \cdot L$
Inference	$2 \cdot D \cdot D' \cdot L$	$m \cdot D$	$m \cdot D \cdot L$	0	0	0	0

Table 5. This table is extended from Table 4 in Section 4 and describes the detailed experimental results of the performance comparison on VTAB-1k using ViT-Large pre-trained on ImageNet-21k as the backbone.

Datasets / Methods	Natural								Specialized					Structed										
	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)
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 [9]	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 [31]	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 [13]	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 [13]	**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 [10]	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 [5]	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
SSF [18]	73.5	91.3	70.0	99.3	91.3	90.6	57.5	81.9	85.9	94.9	85.5	74.4	85.2	80.6	60.0	53.3	80.0	77.6	54.0	**31.8**	35.0	59.0	73.0	0.60
RLRR	79.3	**92.0**	**74.6**	**99.5**	**92.1**	89.6	**60.1**	**83.9**	**87.3**	95.3	**87.3**	75.7	**86.4**	**82.7**	**62.1**	**54.6**	80.6	**87.1**	**54.7**	31.3	**41.9**	**61.9**	**75.2**	0.82

Table 6. This table is extended from Table 4 in Section 4 and describes the detailed experimental results of the performance comparison on VTAB-1k using ViT-Huge pre-trained on ImageNet-21k as the backbone.

Datasets / Methods	Natural								Specialized					Structed										
	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)
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	630.90
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 [9]	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 [31]	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 [13]	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 [13]	**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 [10]	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 [5]	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
SSF [18]	66.6	**91.2**	69.0	98.4	**88.1**	88.9	50.7	79.0	85.0	94.1	79.3	73.9	83.1	73.9	61.2	47.9	76.2	82.8	51.9	25.5	33.7	56.6	70.4	0.97
RLRR	70.3	89.8	**69.7**	**98.6**	87.8	88.5	**51.3**	**79.4**	**86.0**	95.0	84.9	**74.6**	**85.1**	73.8	60.1	**49.6**	**78.6**	**83.6**	**52.4**	32.0	**41.8**	**59.0**	**72.0**	1.33

Table 7. This table is extended from Table 5 in Section 4 and describes the detailed experimental results of the performance comparison on VTAB-1k using Swin-Base pre-trained on ImageNet-21k as the backbone.

Datasets / Methods	Natural								Specialized					Structed										
	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)
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 [13]	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 [13]	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 [31]	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 [13]	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 [13]	**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 [5]	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
RLRR	66.1	90.6	75.5	99.3	**92.1**	**90.9**	54.7	**81.3**	87.1	95.9	87.1	76.5	**86.7**	66.0	57.8	**55.3**	**84.1**	91.1	**55.2**	28.6	34.0	59.0	**73.0**	0.41

to LoRA. The results of above RLRR variants are shown in Table 11, which validate our statement.

G. Transferability Analysis

We extend the RLRR variant to CNN by concatenating CNN kernels. As shown in Fig. 3, by concatenating the convolutional kernels, we transform original convolutional kernel parameters in CNN to a two-dimensional parameter matrix $\mathbf{W}'$, allowing RLRR to seamlessly migrate to CNNs. Based on this, we supplement the experiments on CIFAR-100, as shown in Table 12, which demonstrates the transferability of our RLRR on other deep learning model. We will explore applying our approach in future work under the field of NLP. The outcomes of this variant are presented in Table 12, underscoring the versatility of our design.

Table 8. Performance comparison on VTAB-1k using MAE self-supervised pre-trained ViT-Base as backbone.

Methods \ Datasets	Natural								Specialized					Structed										
	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)
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 [31]	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 [9]	**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 [13]	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 [13]	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 [10]	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 [5]	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
RLRR	33.6	88.9	**62.2**	**87.3**	**86.7**	89.1	**25.7**	**67.6**	**86.0**	93.4	**81.3**	75.1	**84.0**	77.0	**65.5**	**53.4**	**84.7**	78.5	**54.5**	**37.2**	**43.1**	61.7	**68.6**	0.33

Table 9. Performance comparison on VTAB-1k using Moco V3 self-supervised pre-trained ViT-Base as backbone.

Methods \ Datasets	Natural								Specialized					Structed										
	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)
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 [31]	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 [9]	**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 [13]	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 [13]	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 [10]	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 [5]	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
RLRR	61.8	**91.7**	68.6	91.6	**89.5**	**91.5**	**41.7**	**76.6**	**87.9**	96.0	85.4	75.4	**86.2**	79.3	**64.6**	**51.5**	81.4	77.5	50.4	**35.6**	45.9	62.1	**73.1**	0.33

Table 10. The impacts of initialization.

Initialization	Natural (7)	Specialized (4)	Structed (8)
normal	**82.9**	85.4	61.4
zero	82.4	85.1	60.8
constant	81.6	84.2	60.9
uniform	82.5	85.6	61.4
RLRR	82.7	**85.8**	**61.8**

Table 11. Ablation study on VTAB-1k to compare with baseline.

Method	Natural (7)	Specialized (4)	Structed (8)	Params
w/o W	81.3	85.5	57.5	0.33
w/ W	**82.7**	**85.8**	**61.8**	0.33
LoRA (r=16)	80.4	85.2	61.0	0.63

Table 12. Performance comparison of RLRR extended to CNNs.

Methods	ResNet-18		ResNet-50	
	CIFAR-100	Params	CIFAR-100	Params
Full fine-tuning	**79.7**	11.23	80.7	23.71
Linear probing	62.1	0.05	66.8	0.21
RLRR(r=1)	75.0	0.08	79.0	0.27
RLRR(r=10)	78.9	0.29	**82.4**	0.85

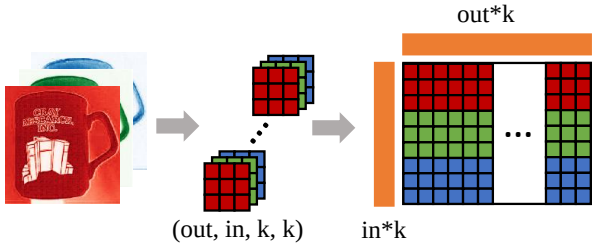

Figure 3. Illustration of the RLRR method's extension to CNN.

H. Combination of multiple RLRRs

RLRR can be likened to a LoRA with rank = 1. Consequently, operations like element-wise combination and arithmetic applied to LoRAs, as demonstrated in LCM-LoRA[21], LoRAHub [11], and Composing PEMs[33], are also applicable to PLRR. The various combinations of RLRRs within their respective frameworks can be expressed in the form of $\hat{\mathbf{S}}_{\text{left}}\hat{\mathbf{S}}_{\text{right}} = \sum_i^N w_i \mathbf{S}_{\text{left}}^i \sum_i^N w_i \mathbf{S}_{\text{right}}^i$ with $\hat{\vec{\boldsymbol{f}}}^\top = \sum_i^N w_i \vec{\boldsymbol{f}}_i^\top$, and $\hat{\mathbf{S}}_{\text{left}}\hat{\mathbf{S}}_{\text{right}} = \sum_i^N \mathbf{S}_{\text{left}}^i \mathbf{S}_{\text{right}}^i$ with $\hat{\vec{\boldsymbol{f}}}^\top = \sum_i^N w_i \vec{\boldsymbol{f}}_i^\top$.