

## 1    **A    Datasets**

2    We conduct our experiments on ImageNet, Stanford Dogs, Oxford 102 Flowers, and CIFAR-10/100.  
3    These datasets do not contain any personally identifiable information. They are publicly available for  
4    research use.

## 5    **B    Experimental Setup Details**

6    We specify all model architectures and the training details in Section B.1, Section B.2, and Section B.3.

### 7    **B.1    Large-scale classification: ImageNet**

8    We use ResNet-10, ResNet-18, and ResNet-26 on large-scale classification. The architectures are  
9    shown in Table 1. We use SGD optimizer with batch size 256. The initial learning rate, the momentum,  
10    and the weight decay are set to 0.1, 0.9, and  $10^{-4}$ , respectively. The training takes 100 epochs and the  
11    learning rate is scheduled to decay by a factor of 0.1 at epochs 30, 60, and 90. We augment the data  
12    via random crop and resize it to 224x224. We use "torchvision.transforms.RandomResizedCrop",  
13    which is implemented in Pytorch. Then the data is flipped horizontally with 50% probability. It takes  
14    3 days to train a model on a single GPU (Nvidia Tesla V100).

### 15    **B.2    Small-scale classification: CIFAR-10/100**

16    Both CIFAR10 and CIFAR100 contain 50,000 training images and 10,000 testing images. They are  
17    all  $32 \times 32$  colored images. We use ResNet-20 and ResNet-56 on small-scale image classification.  
18    The architectures are shown in Table 2. The training takes 200 epochs. We use SGD as the optimizer  
19    and the batch size is 128. The initial learning rate, the moment, and the weight decay are set to 0.1,  
20    0.9, and  $10^{-4}$ , respectively. The learning rate is scheduled to decay by a factor of 0.1 at epochs 100,  
21    150, and 180. In addition, the training images are augmented by random horizontal flips and random  
22    shifts by up to 4px to prevent overfitting. It takes about 1.5 hours to train a model on a single GPU  
23    (Nvidia Tesla V100).

### 24    **B.3    Fined-grained classification: Stanford Dogs and Oxford 102 Flowers**

25    We use ResNet-10, ResNet-18, and ResNet-26 on fine-grained classification. The architectures are  
26    shown in Table 1. We apply random crops, horizontal flips, and random gamma transform for data  
27    augmentation. We use SGD as the optimizer and the initial learning rate, the moment, and the weight  
28    decay are set to 0.1, 0.9, and  $10^{-4}$ , respectively. The training takes 200 epochs and the learning rate  
29    is scheduled to decay at epochs 100, 150, and 200 by a factor of 0.1. It takes about 1.5 hours to train  
30    a model on a single GPU (Nvidia Tesla V100).

## 31    **C    Details of the BlkSConv-based ResNet architectures**

32    We present more experiments and the architecture details of the BlkSConv-based ResNets that found  
33    by the proposed hyperparameter search algorithm (HSA) in Table 3, Table 4, Table 5, Table 6, Table 7,  
34    Table 8, Table 9, and Table 10. P\_ratio and MA\_ratio represent the parameter ratio and MAdds ratio  
35    compared to the standard model, respectively.

Table 1: ResNet architectures used for ImageNet, Stanford Dogs, and Oxford 102 Flowers. The PCA-based HSA is applied to conv3\_x, conv4\_x, conv5\_x layers.

ResNet-10 (L=1), ResNet-18 (L=2), ResNet-26 (L=3)				
Layers Names	Output Size	ResNet	Applying HSA	e.g.(ResNet-10)
conv1	$112 \times 112 \times 64$	$7 \times 7, 64$ , stride 2	No	
max pool	$56 \times 56 \times 64$	$3 \times 3$ , stride 2		
conv2_x	$56 \times 56 \times 64$	$\begin{bmatrix} 3 \times 3, & 64 \\ 3 \times 3, & 64 \end{bmatrix} \times L$	No	
conv3_x	$28 \times 28 \times 128$	$\begin{bmatrix} 3 \times 3, & 128 \\ 3 \times 3, & 128 \end{bmatrix} \times L$	Yes	conv-s5t2
conv4_x	$14 \times 14 \times 256$	$\begin{bmatrix} 3 \times 3, & 256 \\ 3 \times 3, & 256 \end{bmatrix} \times L$	Yes	conv-s5t2
conv5_x	$7 \times 7 \times 512$	$\begin{bmatrix} 3 \times 3, & 512 \\ 3 \times 3, & 512 \end{bmatrix} \times L$	Yes	conv-s1t1
average pool	$1 \times 1 \times 512$	$7 \times 7$		
fully connected	1000	$512 \times 1000$ fc		

Table 2: ResNet architectures used for CIFAR10/100. The PCA-based HSA is applied to conv4\_x layers.

ResNet-20 (L=3), ResNet-56 (L=9)			
Layers Names	Output Size	ResNet	Applying HSA
conv1	$32 \times 32 \times 16$	$3 \times 3, 16$	No
conv2_x	$16 \times 16 \times 16$	$\begin{bmatrix} 3 \times 3, & 16 \\ 3 \times 3, & 16 \end{bmatrix} \times L$	No
conv3_x	$8 \times 8 \times 16$	$\begin{bmatrix} 3 \times 3, & 32 \\ 3 \times 3, & 32 \end{bmatrix} \times L$	No
conv4_x	$4 \times 4 \times 32$	$\begin{bmatrix} 3 \times 3, & 64 \\ 3 \times 3, & 64 \end{bmatrix} \times L$	Yes
average pool	$1 \times 1 \times 64$	$4 \times 4$	
fully connected	10 (or 100)	$64 \times 10$ (or 100) fc	

Table 3: BlkSConv-based ResNet-10 on ImageNet.

ResNet-10 on ImageNet						
$(\alpha_v, \alpha_c, \alpha_s, SS)$	conv3_x	conv4_x	conv5_x	Accuracy	P_ratio	MA_ratio
(0.50, 0.50, 0.50, max)	s1t1-s4t1	s1t1-s4t1	s1t1-s4t1	62.736	0.3420	0.4596
(0.50, 0.50, 0.75, max)	s1t1-s4t1	s1t1-s4t1	s1t1-s4t1	62.736	0.3420	0.4596
(0.50, 0.75, 0.50, max)	s1t1-s4t1	s1t1-s4t1	s1t1-s4t1	62.736	0.3420	0.4596
(0.50, 0.75, 0.75, max)	s1t1-s6t1	s1t1-s6t1	s1t1-s6t1	62.79	0.4936	0.6139
(0.75, 0.50, 0.50, max)	conv-s4t1	conv-s4t1	conv-s4t1	63.646	0.6366	0.6418
(0.75, 0.50, 0.75, max)	conv-s4t1	conv-s4t1	conv-s4t1	63.646	0.6366	0.6418
(0.75, 0.75, 0.50, max)	conv-s4t1	conv-s4t1	conv-s4t1	63.646	0.6366	0.6418
(0.75, 0.75, 0.75, max)	conv-s6t1	conv-s6t1	conv-s6t1	63.622	0.7882	0.7960
(0.50, 0.50, 0.50, min)	s1t1-s1t1	s1t1-s2t2	s1t1-s1t2	61.958	0.0888	0.2560
(0.50, 0.50, 0.75, min)	s1t1-s1t1	s1t1-s2t2	s1t1-s1t2	61.958	0.0888	0.2560
(0.50, 0.75, 0.50, min)	s1t1-s1t1	s1t1-s2t2	s1t1-s1t2	61.958	0.0888	0.2560
(0.50, 0.75, 0.75, min)	s1t1-s1t1	s1t1-s2t2	s1t1-s1t2	61.958	0.0888	0.2560
(0.75, 0.50, 0.50, min)	conv-s3t1	conv-s3t1	conv-s1t1	63.398	0.4459	0.5144
(0.75, 0.50, 0.75, min)	conv-s3t1	conv-s3t1	conv-s1t1	63.398	0.4459	0.5144
(0.75, 0.75, 0.50, min)	conv-s5t2	conv-s5t2	conv-s1t1	63.464	0.4423	0.6314
(0.75, 0.75, 0.75, min)	conv-s5t2	conv-s5t2	conv-s1t1	63.464	0.4423	0.6314
Standard	conv-conv	conv-conv	conv-conv	63.686	4.64M	520M

Table 4: BlkSConv-based ResNet-18 on ImageNet.

ResNet-18 on ImageNet						
$(\alpha_v, \alpha_c, \alpha_s, SS)$	conv3_x	conv4_x	conv5_x	Accuracy	P_ratio	MA_ratio
(0.50, 0.50, 0.50, max)	s1t1-s4t1 s4t1-s4t1	s1t1-s4t1 s4t1-s4t1	s1t1-s4t1 s4t1-s4t1	69.922	0.4065	0.4614
(0.50, 0.50, 0.75, max)	s1t1-s4t1 s4t1-s4t1	s1t1-s4t1 s4t1-s4t1	s1t1-s4t1 s4t1-s4t1	69.922	0.4065	0.4614
(0.50, 0.75, 0.50, max)	s1t1-s4t1 s4t1-s4t1	s1t1-s4t1 s4t1-s4t1	s1t1-s4t1 s4t1-s4t1	69.922	0.4065	0.4614
(0.50, 0.75, 0.75, max)	s1t1-s6t1 s6t1-s6t1	s1t1-s6t1 s6t1-s6t1	s1t1-s6t1 s6t1-s6t1	69.782	0.6014	0.6597
(0.50, 0.50, 0.50, min)	s1t1-s1t1 s2t2-s3t2	s1t1-s1t1 s3t2-s3t2	s1t1-s3t2 s1t1-s1t2	67.572	0.1264	0.2700
(0.50, 0.50, 0.75, min)	s1t1-s1t1 s2t2-s3t2	s1t1-s1t1 s3t2-s3t2	s1t1-s3t2 s1t1-s1t2	67.572	0.1264	0.2700
(0.50, 0.75, 0.50, min)	s1t1-s1t1 s2t2-s3t2	s1t1-s1t1 s4t4-s4t4	s1t1-s3t2 s1t1-s1t2	67.540	0.1246	0.2986
(0.50, 0.75, 0.75, min)	s1t1-s1t1 s2t2-s3t2	s1t1-s1t1 s4t4-s4t4	s1t1-s3t2 s1t1-s1t2	67.540	0.1246	0.2986
Standard	conv-conv conv-conv	conv-conv conv-conv	conv-conv conv-conv	70.728	10.8M	1213.8M

Table 5: BlkSConv-based ResNet-26 on ImageNet.

ResNet-26 on ImageNet						
$(\alpha_v, \alpha_c, \alpha_s, SS)$	conv3_x	conv4_x	conv5_x	Accuracy	P_ratio	MA_ratio
(0.50, 0.50, 0.50, max)	s1t1-s4t1	s1t1-s4t1	s1t1-s4t1	72.038	0.4241	0.4618
	s4t1-s4t1	s4t1-s4t1	s4t1-s4t1			
	s4t1-s4t1	s4t1-s4t1	s4t1-s4t1			
(0.50, 0.50, 0.75, max)	s1t1-s4t1	s1t1-s4t1	s1t1-s4t1	72.038	0.4241	0.4618
	s4t1-s4t1	s4t1-s4t1	s4t1-s4t1			
	s4t1-s4t1	s4t1-s4t1	s4t1-s4t1			
(0.50, 0.75, 0.50, max)	s1t1-s4t1	s1t1-s4t1	s1t1-s4t1	72.038	0.4241	0.4618
	s4t1-s4t1	s4t1-s4t1	s4t1-s4t1			
	s4t1-s4t1	s4t1-s4t1	s4t1-s4t1			
(0.50, 0.75, 0.75, max)	s1t1-s6t1	s1t1-s6t1	s1t1-s6t1	72.326	0.6308	0.6722
	s6t1-s6t1	s6t1-s6t1	s6t1-s6t1			
	s6t1-s6t1	s6t1-s6t1	s6t1-s6t1			
(0.50, 0.50, 0.50, min)	s1t1-s4t1	s1t1-s1t1	s1t1-s3t2	69.970	0.1250	0.2668
	s1t1-s3t2	s1t1-s3t2	s1t1-s1t1			
	s3t2-s3t2	s3t2-s3t2	s3t4-s1t2			
(0.50, 0.50, 0.75, min)	s1t1-s4t1	s1t1-s1t1	s1t1-s3t2	69.970	0.1250	0.2668
	s1t1-s3t2	s1t1-s3t2	s1t1-s1t1			
	s3t2-s3t2	s3t2-s3t2	s3t4-s1t2			
(0.50, 0.75, 0.50, min)	s1t1-s4t1	s1t1-s1t1	s1t1-s5t4	69.922	0.1243	0.2910
	s1t1-s3t2	s1t1-s4t4	s1t1-s1t1			
	s3t2-s3t2	s3t2-s3t2	s3t4-s1t2			
(0.50, 0.75, 0.75, min)	s1t1-s4t1	s1t1-s1t1	s1t1-s5t4	69.922	0.1243	0.2910
	s1t1-s3t2	s1t1-s4t4	s1t1-s1t1			
	s3t2-s3t2	s3t2-s3t2	s3t4-s1t2			
Standard	conv-conv conv-conv conv-conv	conv-conv conv-conv conv-conv	conv-conv conv-conv conv-conv	72.604	17.03M	1907.4M

Table 6: BlkSConv-based ResNet-20 on CIFAR10.

ResNet-20 on CIFAR 10						
$(\alpha_v, \alpha_c, \alpha_s, SS)$	conv4_1	conv4_2	conv4_3	Accuracy	P_ratio	MA_ratio
(0.50, 0.50, 0.50, max)	s1t1-s3t1	s3t1-s3t1	s3t1-s3t1	91.652	0.3586	0.3889
(0.50, 0.50, 0.75, max)	s1t1-s3t1	s3t1-s3t1	s3t1-s3t1	91.652	0.3586	0.3889
(0.50, 0.75, 0.50, max)	s1t1-s5t2	s5t2-s5t2	s5t2-s5t2	91.886	0.4075	0.6903
(0.50, 0.75, 0.75, max)	s1t1-s5t1	s5t1-s5t1	s5t1-s5t1	91.836	0.5890	0.6193
(0.50, 0.50, 0.50, min)	s1t1-s3t1	s3t1-s2t1	s2t1-s1t1	91.574	0.2664	0.2967
(0.50, 0.50, 0.75, min)	s1t1-s3t1	s3t1-s2t1	s2t1-s1t1	91.574	0.2644	0.2967
(0.50, 0.75, 0.50, min)	s1t1-s4t2	s4t2-s2t1	s2t1-s1t1	91.612	0.2544	0.3655
(0.50, 0.75, 0.75, min)	s1t1-s4t2	s4t2-s2t1	s2t1-s1t1	91.612	0.2544	0.3655
Standard	conv-conv	conv-conv	conv-conv	92.006	202752	12.97M

Table 7: BlkSConv-based ResNet-20 on CIFAR100.

ResNet-20 on CIFAR 100						
$(\alpha_v, \alpha_c, \alpha_s, SS)$	conv4_1	conv4_2	conv4_3	Accuracy	P_ratio	MA_ratio
(0.50, 0.50, 0.50, max)	s1t1-s3t1	s3t1-s3t1	s3t1-s3t1	67.078	0.3586	0.3889
(0.50, 0.50, 0.75, max)	s1t1-s3t1	s3t1-s3t1	s3t1-s3t1	67.078	0.3586	0.3889
(0.50, 0.75, 0.50, max)	s1t1-s5t2	s5t2-s5t2	s5t2-s5t2	67.212	0.4075	0.6903
(0.50, 0.75, 0.75, max)	s1t1-s5t1	s5t1-s5t1	s5t1-s5t1	67.552	0.5890	0.6193
(0.50, 0.50, 0.50, min)	s1t1-s2t1	s2t1-s2t1	s2t1-s1t1	66.554	0.2203	0.2506
(0.50, 0.50, 0.75, min)	s1t1-s2t1	s2t1-s2t1	s2t1-s1t1	66.554	0.2203	0.2506
(0.50, 0.75, 0.50, min)	s1t1-s2t1	s2t1-s2t1	s2t1-s1t1	66.554	0.2203	0.2506
(0.50, 0.75, 0.75, min)	s1t1-s2t1	s2t1-s2t1	s2t1-s1t1	66.554	0.2203	0.2506
Standard	conv-conv	conv-conv	conv-conv	67.994	202752	12.97M

Table 8: BlkSConv-based ResNet-56 on CIFAR10.

ResNet-56 on CIFAR 10						
$(\alpha_v, \alpha_c, \alpha_s, SS)$	conv4_1 conv4_2 conv4_3	conv4_4 conv4_5 conv4_6	conv4_7 conv4_8 conv4_9	Accuracy	P_ratio	MA_ratio
(0.50, 0.50, 0.50, max)	s1t1-s3t1 s3t1-s3t1 s3t1-s3t1	s3t1-s3t1 s3t1-s3t1 s3t1-s3t1	s3t1-s3t1 s3t1-s3t1 s3t1-s3t1	93.372	0.3734	0.3829
(0.50, 0.50, 0.75, max)	s1t1-s3t1 s3t1-s3t1 s3t1-s3t1	s3t1-s3t1 s3t1-s3t1 s3t1-s3t1	s3t1-s3t1 s3t1-s3t1 s3t1-s3t1	93.372	0.3734	0.3829
(0.50, 0.75, 0.50, max)	s1t1-s5t2 s5t2-s5t2 s5t2-s5t2	s5t2-s5t2 s5t2-s5t2 s5t2-s5t2	s5t2-s5t2 s5t2-s5t2 s5t2-s5t2	93.332	0.4257	0.7051
(0.50, 0.75, 0.75, max)	s1t1-s5t1 s5t1-s5t1 s5t2-s5t2	s5t2-s5t2 s5t2-s5t2 s5t2-s5t2	s5t2-s5t2 s5t2-s5t2 s5t2-s5t2	93.338	0.6196	0.6292
(0.50, 0.50, 0.50, min)	s1t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t1 s1t1-s1t2	93.324	0.2335	0.2462
(0.50, 0.50, 0.75, min)	s1t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t1 s1t1-s1t2	93.324	0.2335	0.2462
(0.50, 0.75, 0.50, min)	s1t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t1 s1t1-s1t2	93.324	0.2335	0.2462
(0.50, 0.75, 0.75, min)	s1t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t1 s1t1-s1t2	93.324	0.2335	0.2462
Standard	conv-conv conv-conv conv-conv	conv-conv conv-conv conv-conv	conv-conv conv-conv conv-conv	93.218	645120	41.28M

Table 9: BlkSConv-based ResNet-56 on CIFAR100.

ResNet-56 on CIFAR 100						
$(\alpha_v, \alpha_c, \alpha_s, SS)$	conv4_1 conv4_2 conv4_3	conv4_4 conv4_5 conv4_6	conv4_7 conv4_8 conv4_9	Accuracy	P_ratio	MA_ratio
$(0.50, 0.50, 0.50, \max)$	s1t1-s3t1 s3t1-s3t1 s3t1-s3t1	s3t1-s3t1 s3t1-s3t1 s3t1-s3t1	s3t1-s3t1 s3t1-s3t1 s3t1-s3t1	70.636	0.3734	0.3829
$(0.50, 0.50, 0.75, \max)$	s1t1-s3t1 s3t1-s3t1 s3t1-s3t1	s3t1-s3t1 s3t1-s3t1 s3t1-s3t1	s3t1-s3t1 s3t1-s3t1 s3t1-s3t1	70.636	0.3734	0.3829
$(0.50, 0.75, 0.50, \max)$	s1t1-s5t2 s5t2-s5t2 s5t2-s5t2	s5t2-s5t2 s5t2-s5t2 s5t2-s5t2	s5t2-s5t2 s5t2-s5t2 s5t2-s5t2	70.790	0.4257	0.7051
$(0.50, 0.75, 0.75, \max)$	s1t1-s5t1 s5t1-s5t1 s5t1-s5t1	s5t1-s5t1 s5t1-s5t1 s5t1-s5t1	s5t1-s5t1 s5t1-s5t1 s5t1-s5t1	70.668	0.6196	0.6292
$(0.50, 0.50, 0.50, \min)$	s1t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t2 s2t2-s1t2	69.994	0.2316	0.2570
$(0.50, 0.50, 0.75, \min)$	s1t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t2 s2t2-s1t2	69.994	0.2316	0.2570
$(0.50, 0.75, 0.50, \min)$	s1t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t2 s2t2-s1t2	69.994	0.2316	0.2570
$(0.50, 0.75, 0.75, \min)$	s1t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t1 s2t1-s2t1	s2t1-s2t1 s2t1-s2t2 s2t2-s1t2	69.994	0.2316	0.2570
Standard	conv-conv conv-conv conv-conv	conv-conv conv-conv conv-conv	conv-conv conv-conv conv-conv	70.998	645120	41.28M

Table 10: BlkSConv-based ResNet-18 on Stanford Dogs.

ResNet-18 on Stanford Dogs						
$(\alpha_v, \alpha_c, \alpha_s, SS)$	conv3_x	conv4_x	conv5_x	Accuracy	P_ratio	MA_ratio
$(0.5, 0.5, 0.5, \max)$	conv-s4t1 s4t1-s4t1	conv-s4t1 s4t1-s4t1	conv-s4t1 s4t1-s4t1	53.005	0.5327	0.5394
$(0.5, 0.75, 0.75, \max)$	conv-s6t1 s6t1-s6t1	conv-s6t1 s6t1-s6t1	conv-s6t1 s6t1-s6t1	53.359	0.7277	0.7377
$(0.5, 0.5, 0.5, \min)$	conv-s3t1 s2t1-s3t2	conv-s2t1 s2t1-s2t1	conv-s3t1 s3t2-s3t4	53.159	0.3273	0.4006
$(0.5, 0.75, 0.75, \min)$	conv-s4t2 s2t1-s3t2	conv-s2t1 s2t1-s2t1	conv-s5t2 s5t4-s3t4	53.615	0.3171	0.4611
Standard	conv-conv conv-conv	conv-conv conv-conv	conv-conv conv-conv	52.436	10.8M	1213.8M