

## Supplementary Material

### A. Training Curves on ImageNet

Figure 1 compares the top-1 error rate of ResNets and ScaleNets on both ImageNet training and validation dataset. It has been shown that ScaleNets achieve much lower error rates than their counterpart ResNets during all training phase.

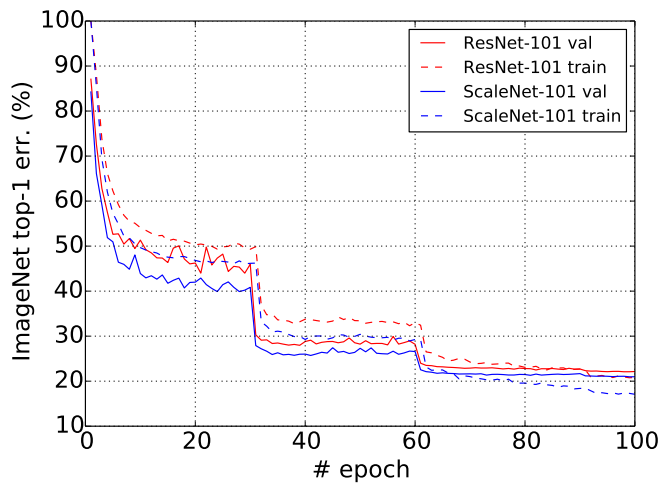
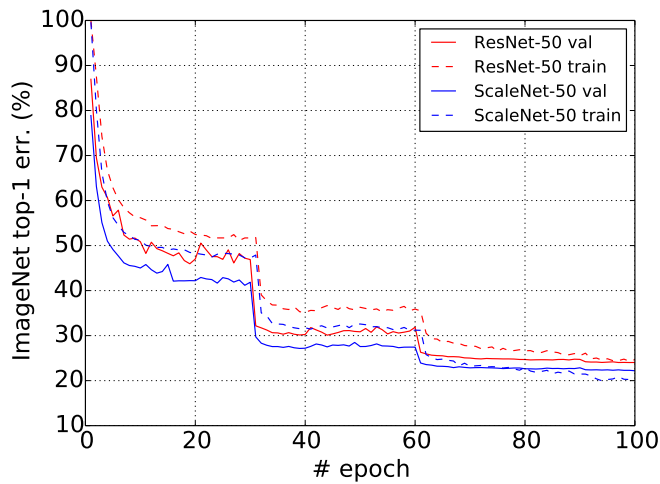


Figure 1: Comparison of the training curves. The left compares scale ScaleNet-50 and ResNet-50 on ImageNet training set and validation set, while the right compares ScaleNet-101 and ResNet-101.

### B. Allocated Neuron Numbers of ScaleNets

Table 1 lists the detailed allocated neuron numbers of  $3 \times 3$  conv in SA block on ImageNet and CIFAR-100.

block index	ScaleNets for CIFAR-100			ScaleNets for ImageNet			
	38 layers	56 layers	101 layers	50 layers (light)	50 layers	101 layers	152 layers
1	6,2,14	14,6,3	10,6,7	30,8,10,16	62,9,5,12	61,11,7,7	39,27,10,14
2	15,6,1	10,10,3	5,5,13	30,9,9,16	55,27,5,1	56,23,4,3	45,32,8,5
3	16,5,1	9,3,11	8,4,11	30,27,7,0	59,26,0,3	59,24,3,0	46,36,8,0
4	16,6,0	11,3,9	12,8,3	59,55,13,1	125,41,6,3	123,41,1,6	55,26,35,63
5	30,12,1	14,9,0	11,6,6	59,43,8,18	90,39,9,37	126,38,1,6	89,44,19,27
6	24,15,4	15,8,0	9,9,5	59,57,12,0	106,56,4,9	127,41,3,0	93,62,14,10
7	26,16,1	28,15,3	9,11,3	59,59,9,1	116,56,3,0	127,41,3,0	110,43,12,14
8	27,13,3	26,13,7	13,8,2	117,65,71,3	223,71,55,0	220,86,35,0	109,54,13,3
9	47,17,22	29,15,2	12,9,2	107,16,33,100	196,104,44,5	186,64,55,36	119,59,0,1
10	30,38,18	26,16,4	15,8,0	111,49,62,34	195,98,52,4	156,25,53,107	106,70,3,0
11	26,52,8	30,14,2	16,7,0	106,61,61,28	155,128,66,0	191,44,52,54	114,65,0,0
12	23,51,12	12,25,9	30,14,2	99,71,59,27	134,129,86,0	181,53,83,24	224,102,31,1
13		50,19,22	23,14,9	76,50,67,63	120,127,98,4	221,82,34,4	163,49,68,78
14		52,34,5	28,14,4	141,182,189,0	237,354,106,0	177,62,90,12	115,65,73,105
15		25,47,19	23,15,8	83,9,185,235	172,435,90,0	130,75,102,34	143,107,71,37
16		24,59,8	26,17,3	77,16,184,235	138,462,97,0	206,71,55,9	144,97,92,25
17		17,57,17	26,16,4			203,83,53,2	195,87,60,16
18		2,58,31	21,22,3			207,73,54,7	198,77,62,21
19			24,22,0			245,84,12,0	168,139,43,8
20			19,18,9			221,103,17,0	80,80,71,127
21			28,18,0			221,100,20,0	138,88,93,39
22			22,20,4			158,99,84,0	107,93,65,93
23			53,15,24			220,106,15,0	230,103,25,0
24			30,43,19			173,92,73,3	182,132,40,4
25			61,27,4			135,122,84,0	179,114,53,12
26			52,35,5			109,71,132,29	220,76,53,9
27			42,45,5			147,94,93,7	227,118,13,0
28			43,45,4			191,108,42,0	196,110,51,1
29			43,47,2			127,95,113,6	232,118,8,0
30			6,50,36			203,117,21,0	224,114,20,0
31			4,51,37			282,377,23,0	214,100,43,1
32			22,57,13			279,388,15,0	139,114,97,8
33			9,60,23			84,442,155,1	198,113,47,0
34							151,87,115,5
35							171,103,83,1
36							172,104,72,10
37							205,88,65,0
38							170,122,64,2
39							170,98,81,9
40							223,101,32,2
41							192,114,52,0
42							112,99,134,13
43							109,116,130,3
44							110,90,118,40
45							194,115,49,0
46							178,135,45,0
47							209,135,14,0
48							341,368,6,0
49							363,348,4,0
50							311,398,6,0

Table 1: Learned neuron numbers in each SA block in ScaleNets on CIFAR-100 and ImageNet. These numbers indicates the output channel numbers for scale 1, 2, 3, and 4 (e.g.  $3 \times 3 \text{ conv}_{[C_1, C_2, C_3, C_4]}$  in Table ??). Note that ScaleNets on CIFAR-100 have only three scales.