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.

![Training Curves](image)

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.

![Trainig Curves](image)
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. 3×3 conv [C1, C2, C3, C4] in Table ??). Note that ScaleNets on CIFAR-100 have only three scales.