

Factorized Convolutional Neural Networks Supplemental Material

Stage	Output	Model N	Model O
1	112^2	$(7, 64)_2$	
2	56^2	3×3 max pooling , stride 2 $\left\{ \begin{array}{l} < 3, 32 > \\ < 3, 64 > \end{array} \right\} \times 10$	
3	28^2	2×2 max pooling , stride 2 $(1, 128)$ $\left\{ \begin{array}{l} < 3, 64 > \\ < 3, 128 > \end{array} \right\} \times 12$	
4	14^2	2×2 max pooling , stride 2 $(1, 256)$ $\left\{ \begin{array}{l} < 3, 128 > \\ < 3, 256 > \end{array} \right\} \times 12$	
5	7^2	2×2 max pooling , stride 2 $(1, 512)$ $\left\{ \begin{array}{l} < 3, 256 > \\ < 3, 512 > \end{array} \right\} \times 16$	
	1^2	$(1, 1024)$ 7×7 average pooling, stride 7 fully connected, 2048 fully connected, 1000 softmax	
	FLOPs	866M	1300M

Table 1: Configurations for our Model **N**, **O** with proposed SIC layers . Input image size is $3 \times 224 \times 224$. For each convolutional layer, we use numbers in brackets to represent its configuration. k denotes the kernel size. n is the number of output channels. Different types of bracket correspond to different convolutional layer. (k, n) is a typical standard convolutional layer. $< k, n >$ represents our SIC layer. The number after the curly brackets indicates the times that the layer is repeated in each stage.