

A Introduction to Embedded Implementation of Convolutional Neural Networks

Table 1. Alex-Net Layers, parameters, output and learnable sizes.

no.	name	type	parameters	activations (layer output)	learnable (weights/biases)
0	data	Image Input	-	227 x 227 x 3	-
1	conv1	Convolution	filters: 96 size: 11x 11 x 3 padding: [0 0 0 0] stride: [4 4]	55 x 55 x 96	Weights: 11 x 11 x 3 x 96 Bias: 1 x 1 x 96
	ReLU1	Rectified Linear Unit	-		-
	pool1	Max-Pooling	size: 3 x 3 padding: [0 0 0 0] stride: [2 2]	27 x 27 x 96	-
2	conv2	<i>Grouped</i> Convolution	filters: 256 size: 5 x 5 x 48 (2 groups) padding: [2 2 2 2] stride: [1 1]	27 x 27 x 256	Weights: 5 x 5 x 48 x 256 Bias: 1 x 1 x 256
	ReLU2	Rectified Linear Unit	-		-
	pool2	Max-Pooling	size: 3 x 3 padding: [0 0 0 0] stride: [2 2]	13 x 13 x 256	-
3	conv3	Convolution	filters: 384 size: 3 x 3 x 256 padding: [1 1 1 1] stride: [1 1]	13 x 13 x 384	Weights: 5 x 5 x 256 x 384 Bias: 1 x 1 x 384
	ReLU3	Rectified Linear Unit	-		-
4	conv4	<i>Grouped</i> Convolution	filters: 384 size: 3 x 3 x 192 (2 groups) padding: [1 1 1 1] stride: [1 1]	13 x 13 x 384	Weights: 3 x 3 x 192 x 384 Bias: 1 x 1 x 256
	ReLU4	Rectified Linear Unit	-		-
5	conv5	<i>Grouped</i> Convolution	filters: 256 size: 3 x 3 x 192 (2 groups) padding: [1 1 1 1] stride: [1 1]	13 x 13 x 256	Weights: 3 x 3 x 192 x 256 Bias: 1 x 1 x 256
	ReLU5	Rectified Linear Unit	-		-
	pool5	Max-Pooling	size: 3 x 3 padding: [0 0 0 0] stride: [2 2]	6 x 6 x 256	-
6	fc6	Fully Connected	nodes: 4096	1 x 1 x 4096	Weights: 4096 x 9216 Bias: 4096 x 1
	ReLU6	Rectified Linear Unit	-		-
7	fc7	Fully Connected	nodes: 4096	1 x 1 x 4096	Weights: 4096 x 4096 Bias: 4096 x 1
	ReLU7	Rectified Linear Unit	-		-
8	fc8	Fully Connected	nodes: 1000	1 x 1 x 1000	Weights: 1000 x 4096 Bias: 1000 x 1
9	prob	Soft-max	nodes: 1000	1 x 1 x 1000	-
	output	Classification Output		1000 categories	-

was first proposed in AlexNet. Various studies show that grouping can dramatically reduce resource use while helping to preserve accuracy [11,12]. In addition to simple convolution