AlexNet.

Equation: $N = \frac{(W - F + 2P)}{S + 1}$

F: filter_size
P: Padding
S: stride

Cov_No	Input size	K_No	K_size	padding	stride	Output size
Cov1	(3, 224, 224)	96	11	(1, 2)	4	(96, 55, 55)
Max1	(96, 55, 55)	96	3	(0, 0)	2	(96, 27, 27)
Cov2	(96, 27, 27)	256	5	(2, 2)	1	(256, 27, 27)
Max2	(256, 27, 27)	256	3	(0, 0)	2	(256, 13, 13)
Cov3	(256, 13, 13)	384	3	(1, 1)	1	(386, 13, 13)
Cov4	(384, 13, 13)	384	3	(1, 1)	1	(384, 13, 13)
Cov5	(384, 13, 13)	256	3	(1, 1)	1	(256, 13, 13)
Max3	(256, 13, 13)	256	3	(0, 0)	2	(256, 6, 6)
FC1			2048			
FC2			2048			
FC3			1000			

Features of AlexNet.

1. ReLU Nonlinearity.

$f(x) = \max(0, x)$

ReLU is faster than $\tanh(x)$.

2. Training on Multiple GPU.

3. Local Response Normalization (LRN)

Equation: $b^i_{xy} = a^i_{xy} / \left(k + \alpha \sum_{j=\max(0, i-n/2)}^{\min(N-1, i+n/2)} (a^j_{xy})^2\right)^\beta$

a^i_{xy} is the ith active neuron computed by applying kernel i at position (x, y)

N is the total number of kernels in the layer.

LRN inspired by Real Neuron, its a competition form.

~~X~~ However, in 2015 (Very deep Convolutional Networks for Large Scale Image Recognition), LRN was been proved invalid with memory wasting.