IMAGENET

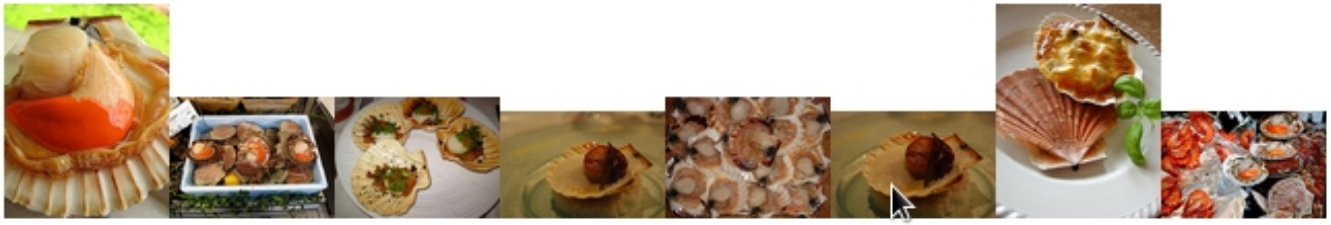
14,197,122 images, 21841 synsets indexed

Explore Download Challenges Publications CoolStuff About

Not logged in. Login | Signup

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.

Click here to learn more about ImageNet, Click here to join the ImageNet mailing list.



What do these images have in common? *Find out!*

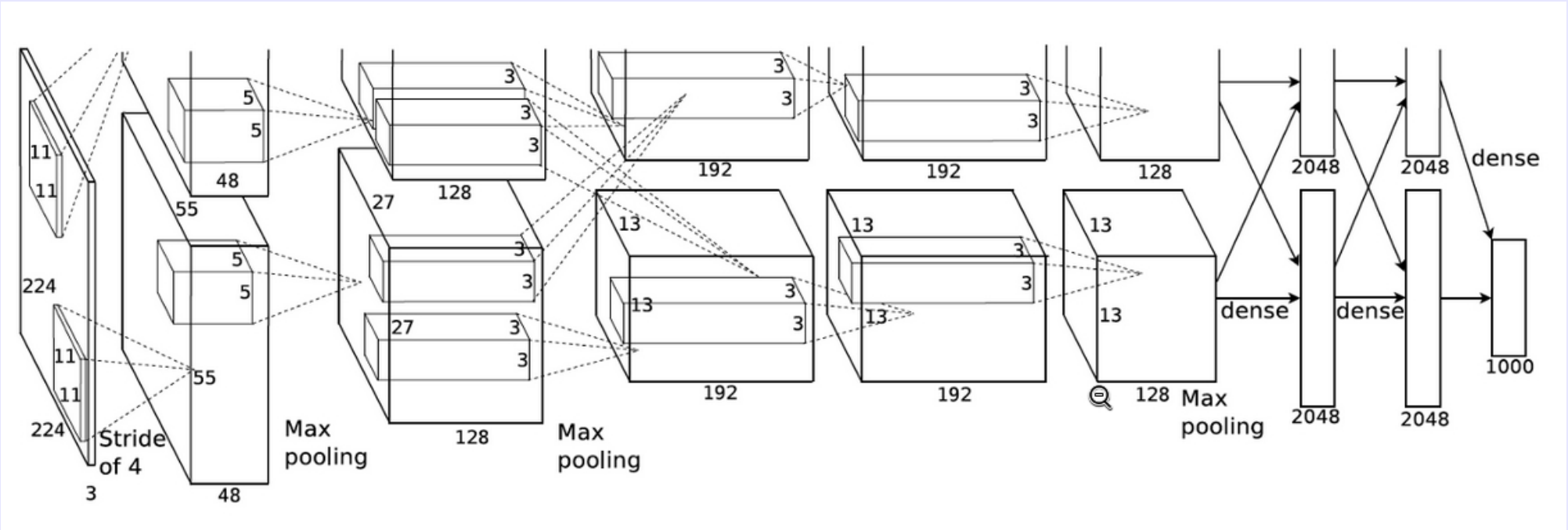
Check out the ImageNet Challenge on Kaggle!

ImageNet classification (ILSVRC)

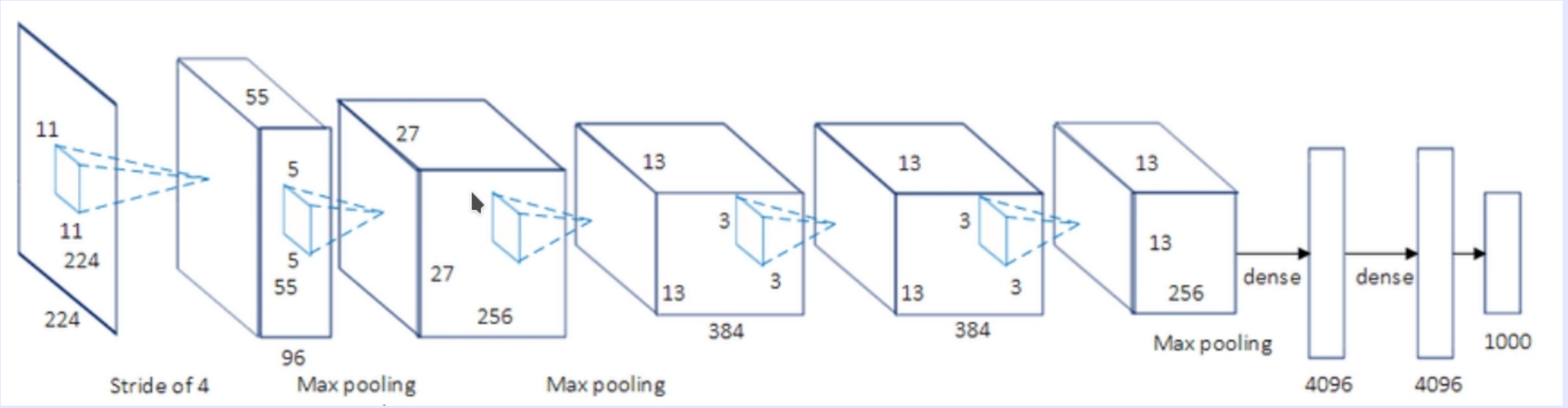
Large Scale Visual Recognition Challenge (2012)

- 1000 objects
- 1.2 million training images
- 100 000 test images

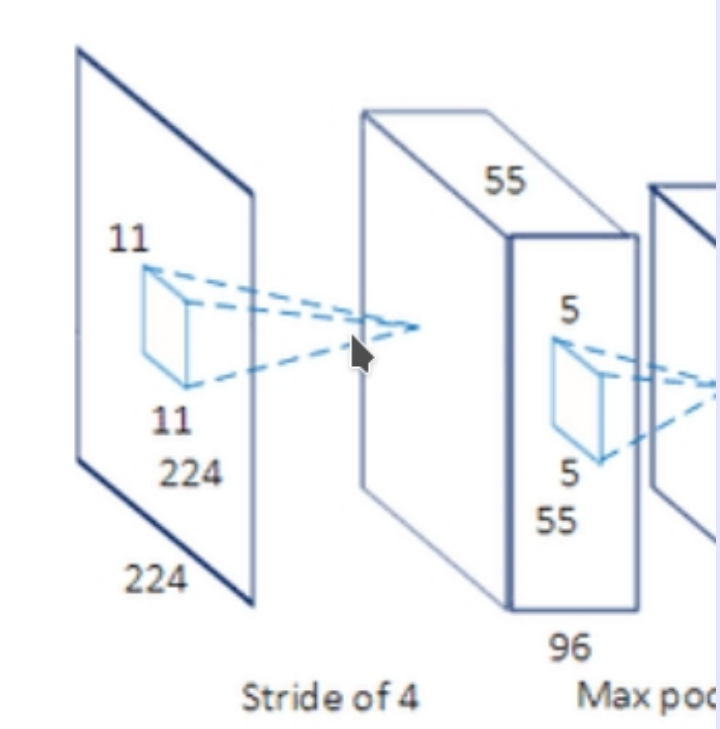
AlexNet


11
11
224
224
3
Stride
of 4
5
5
48
55
5
5
48
55
Max
pooling
3
3
3
128
27
3
3
3
128
27
Max
pooling
3
3
192
3
3
192
13
13
3
3
192
13
13
3
3
192
13
13
128
Max
pooling
dense
2048
2048
dense
dense
2048
2048
1000

AlexNet single GPU equivalent

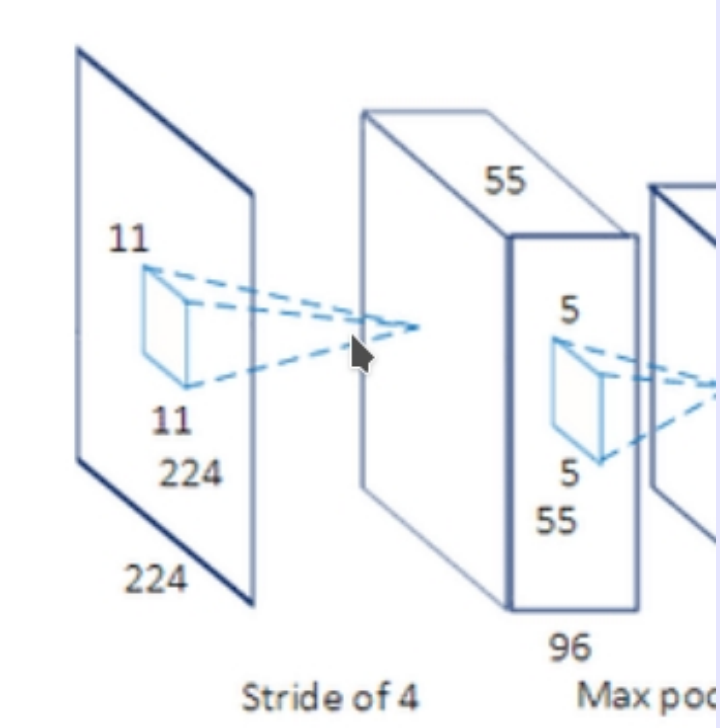


First convolutional layer



- Images: 227x227x3
- Filter size: 11x11
- Stride: 4
- Conv layer output: 55x55x**96**

First convolutional layer

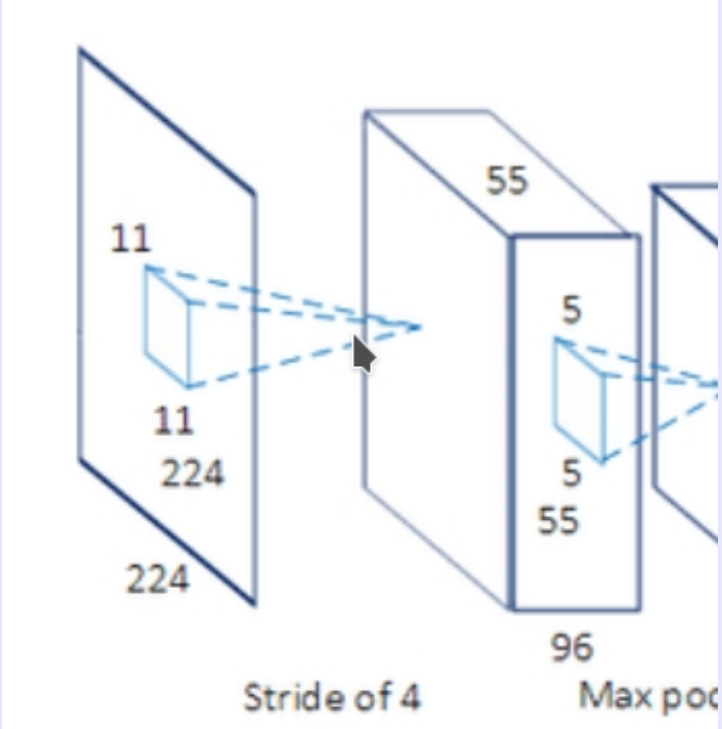


- Images: 227x227x3
- Filter size: 11x11
- Stride: 4
- Conv layer output: 55x55x**96**

Why this output size?

(227-11)/4 + 1 = 55

First convolutional layer



- Images: 227x227x3
- Filter size: 11x11
- Stride: 4
- Conv layer output: 55x55x**96**

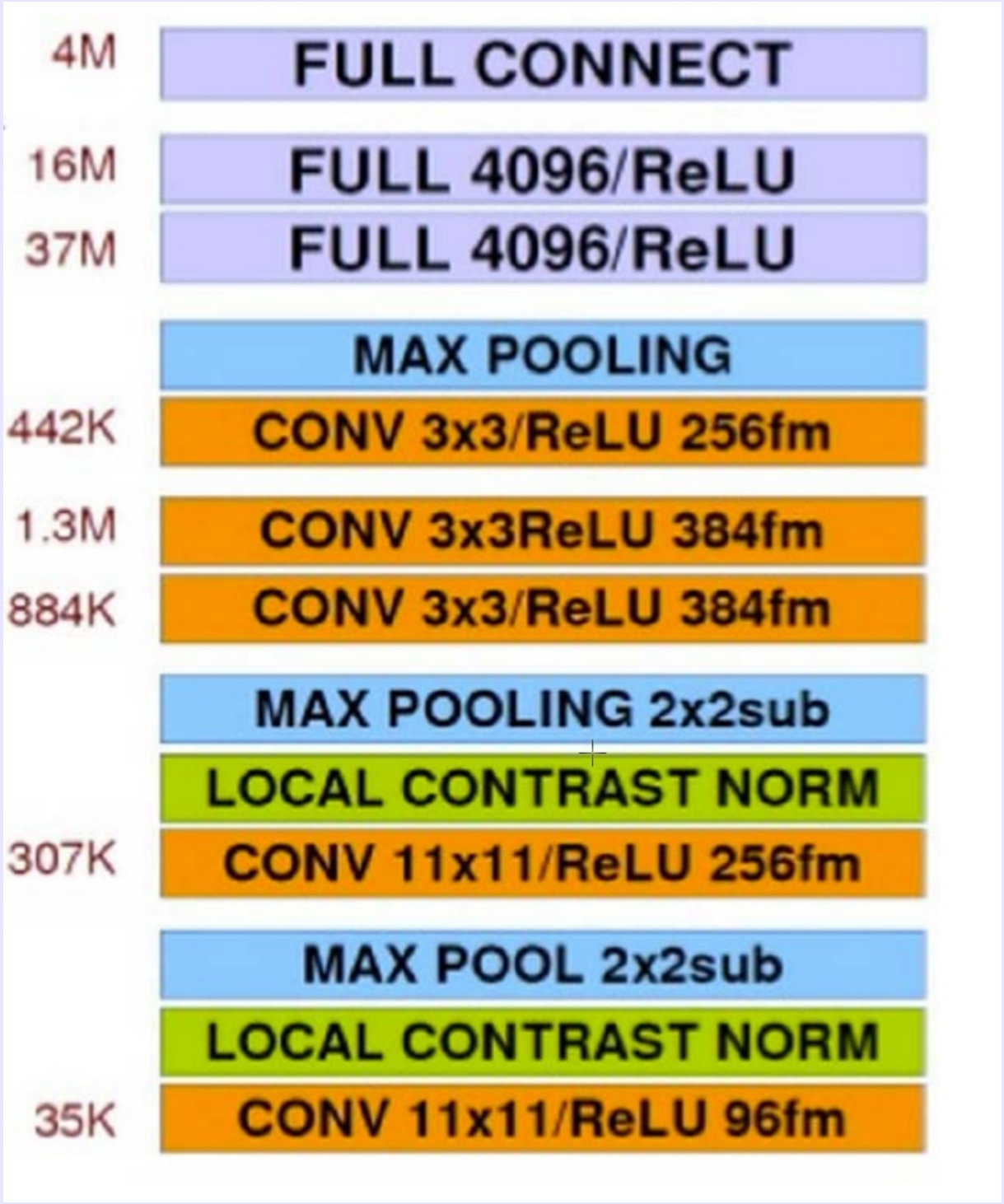
How many weights?

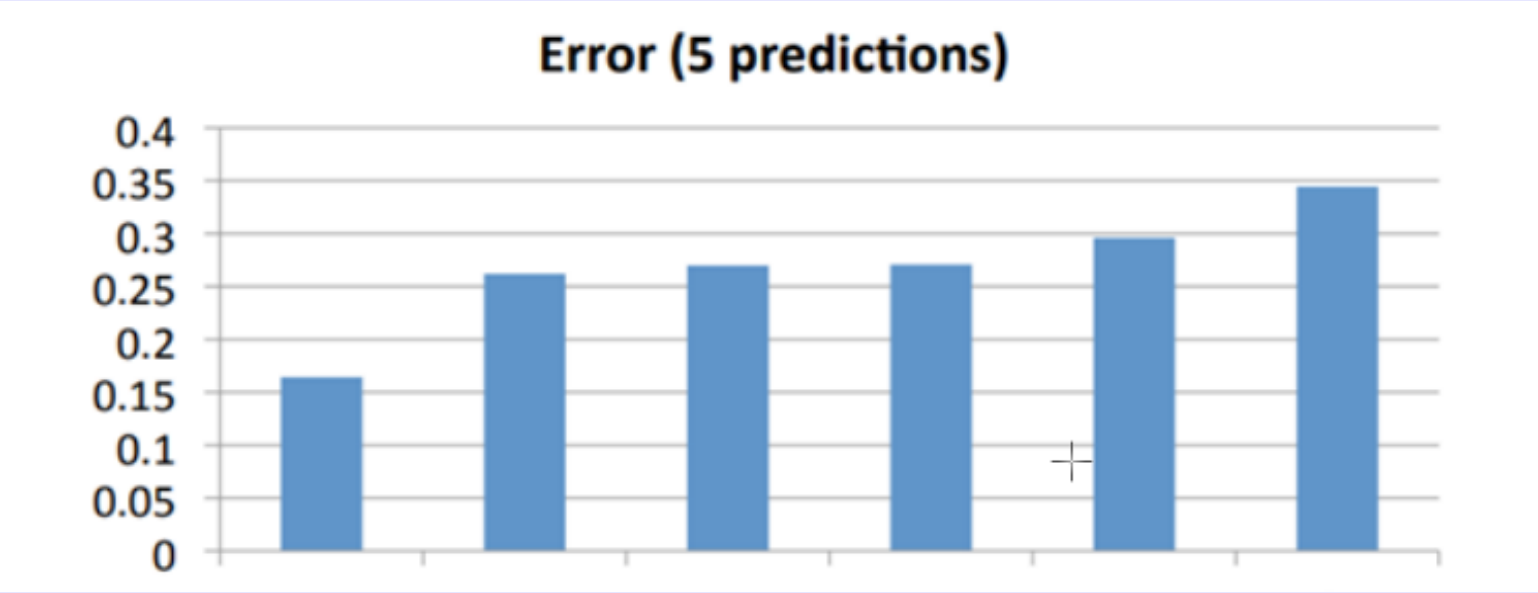
(11*11*3+1) weights per filter
96 filters

(11*11*3+1)*96 = 34944 weights

Size / Operation	Filter	Depth	Stride	Padding	Number of Parameters
3* 227 * 227					
Conv1 + Relu	11 * 11	96	4		(11*11*3 + 1) * 96=34944
96 * 55 * 55					
Max Pooling	3 * 3		2		
96 * 27 * 27					
Norm					
Conv2 + Relu	5 * 5	256	1	2	(5 * 5 * 96 + 1) * 256=614656
256 * 27 * 27					
Max Pooling	3 * 3		2		
256 * 13 * 13					
Norm					
Conv3 + Relu	3 * 3	384	1	1	(3 * 3 * 256 + 1) * 384=885120
384 * 13 * 13					
Conv4 + Relu	3 * 3	384	1	1	(3 * 3 * 384 + 1) * 384=1327488
384 * 13 * 13					
Conv5 + Relu	3 * 3	256	1	1	(3 * 3 * 384 + 1) * 256=884992
256 * 13 * 13					
Max Pooling	3 * 3		2		
256 * 6 * 6					
Dropout (rate 0.5)					
FC6 + Relu					256 * 6 * 6 * 4096=37748736
4096					
Dropout (rate 0.5)					
FC7 + Relu					4096 * 4096=16777216
4096					
FC8 + Relu					4096 * 1000=4096000
1000 classes					
Overall					62369152=62.3 million
Conv VS FC					Conv:3.7million

- Stochastic gradient descent
- ReLU
- Dropout
- Local Response Normalization
  (Today replaced by batch normalization)


4M
FULL CONNECT
16M
FULL 4096/ReLU
37M
FULL 4096/ReLU
MAX POOLING
442K
CONV 3x3/ReLU 256fm
1.3M
CONV 3x3ReLU 384fm
884K
CONV 3x3/ReLU 384fm
MAX POOLING 2x2sub
LOCAL CONTRAST NORM
307K
CONV 11x11/ReLU 256fm
MAX POOL 2x2sub
LOCAL CONTRAST NORM
35K
CONV 11x11/ReLU 96fm



AlexNet

The 96 filters from the first conv. Layer.

(11x11x3)