Convolutional Neural Networks

Giving eyes to the machines!

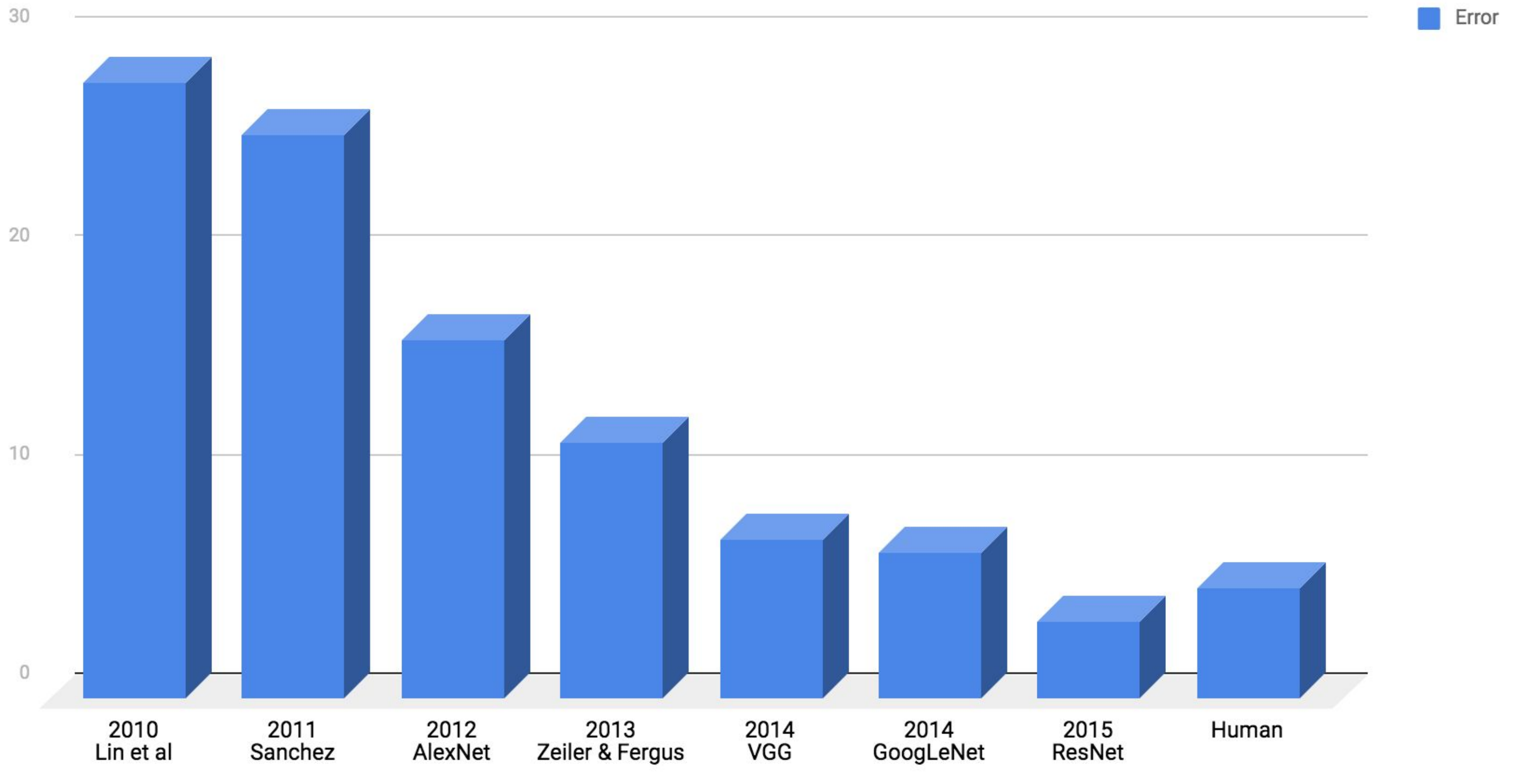
Varun Kohli

[Varun Kohli]

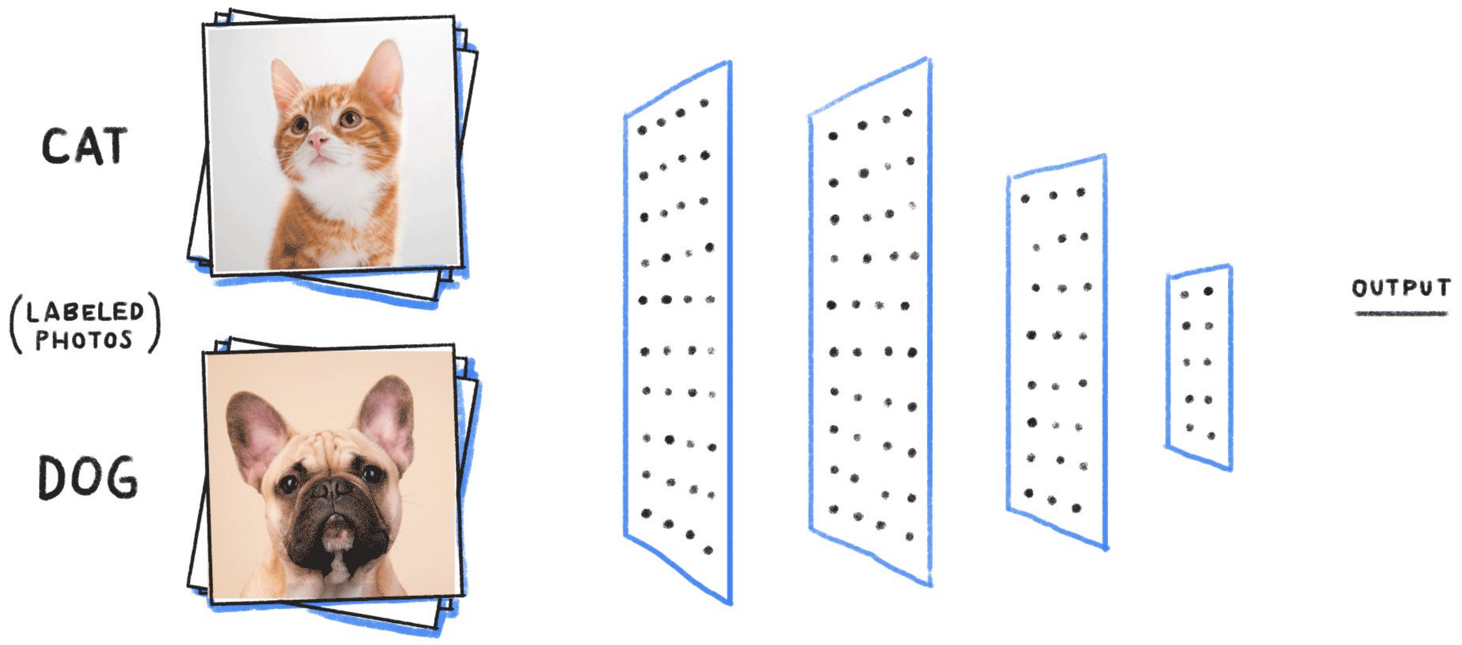
[Lead Strategist]
[Google]

Convolutional Neural Networks

The Evolution!



What are Convolutional Neural Networks?



Building blocks of Convolutional Neural Networks

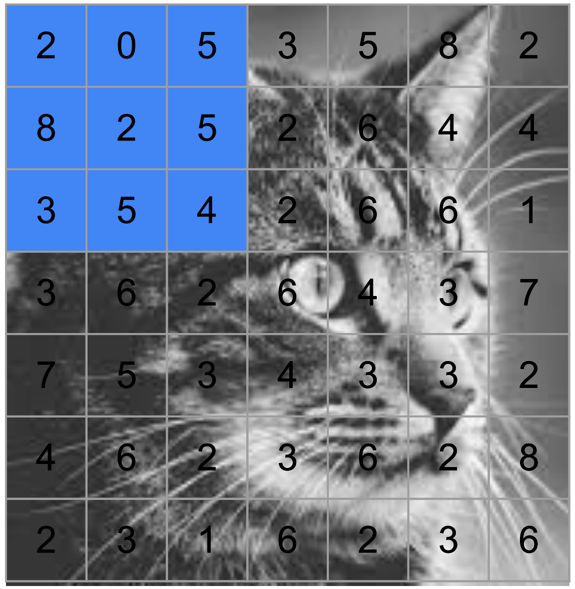
Layers:

- Convolutional Layers
- Pooling Layers
- Fully Connected Layers

Hyperparameters:

- Kernel / Filters
- Stride
- Padding

Convolutional Layers



2	0	5	3	5	8	2
8	2	5	2	6	4	4
3	5	4	2	6	6	1
3	6	2	6	4	3	7
7	5	3	4	3	3	2
4	6	2	3	6	2	8
2	3	1	6	2	3	6

7 x 7

*

2	1	3
1	2	5
2	4	2

Filter size = 3 x 3

=

90				

2 x 2 + 0 x 1 + 5 x 3 + 8 x 1 + 2 x 2 + 5 x 5 + 3 x 2 + 5 x 4 + 4 x 2 = 90

Convolutional Layers

hop 1 column
hop 1 column

$$\begin{array}{|c|c|c|c|c|c|c|}\hline 2 & 0 & 5 & 3 & 5 & 8 & 2 \\ \hline 8 & 2 & 5 & 2 & 6 & 4 & 4 \\ \hline 3 & 5 & 4 & 2 & 6 & 6 & 1 \\ \hline 3 & 6 & 2 & 6 & 4 & 3 & 7 \\ \hline 7 & 5 & 3 & 4 & 3 & 3 & 2 \\ \hline 4 & 6 & 2 & 3 & 6 & 2 & 8 \\ \hline 2 & 3 & 1 & 6 & 2 & 3 & 6 \\ \hline \end{array} * \begin{array}{|c|c|c|}\hline 2 & 1 & 3 \\ \hline 1 & 2 & 5 \\ \hline 2 & 4 & 2 \\ \hline \end{array} = \begin{array}{|c|c|c|c|c|}\hline 90 & 66 & 95 & 109 & 96 \\ \hline 100 & 70 & 104 & 100 & 85 \\ \hline 88 & 90 & 90 & 83 & 88 \\ \hline 86 & 89 & 76 & 84 & 87 \\ \hline 72 & 72 & 87 & 71 & 93 \\ \hline \end{array}$$

7 x 7

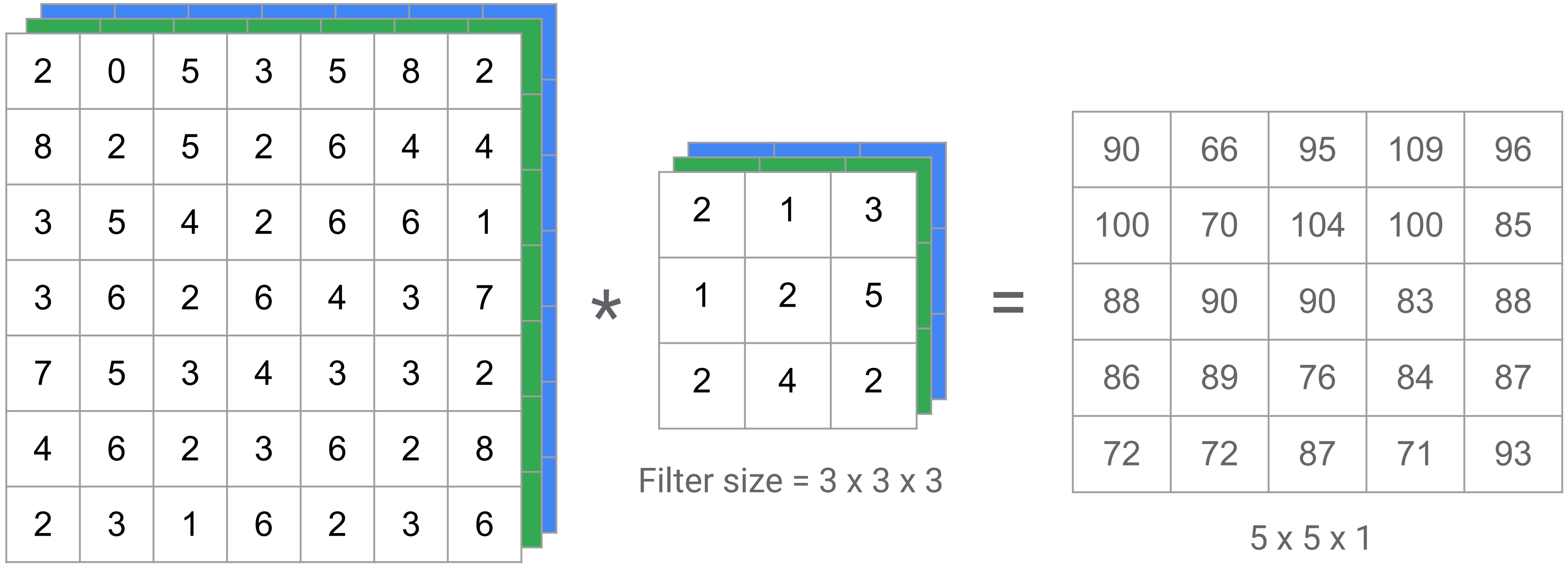
Filter size = 3 x 3

5 x 5

Size of output = [n-f+1] x [n-f+1]

5 x 2 + 3 x 1 + 5 x 3 + 5 x 1 + Code demonstration + 2 x 4 + 6 x 2 = 95

Convolutional Layers with multiple channels



2	0	5	3	5	8	2
8	2	5	2	6	4	4
3	5	4	2	6	6	1
3	6	2	6	4	3	7
7	5	3	4	3	3	2
4	6	2	3	6	2	8
2	3	1	6	2	3	6

7 x 7 x 3

*

2	1	3
1	2	5
2	4	2

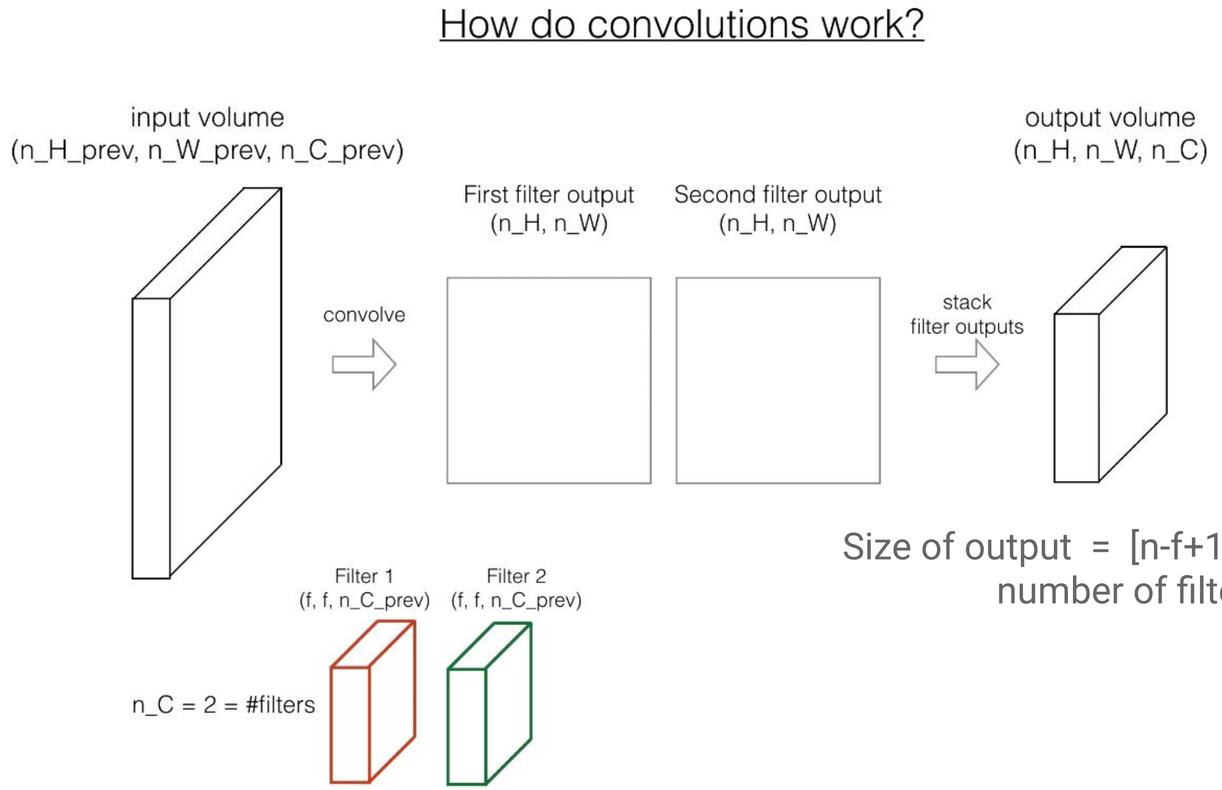
Filter size = 3 x 3 x 3

=

90	66	95	109	96
100	70	104	100	85
88	90	90	83	88
86	89	76	84	87
72	72	87	71	93

5 x 5 x 1

Convolutional Layer with multiple filters



Size of output = [n-f+1] x [n-f+1] x number of filters

Convolutional Layers

256

256

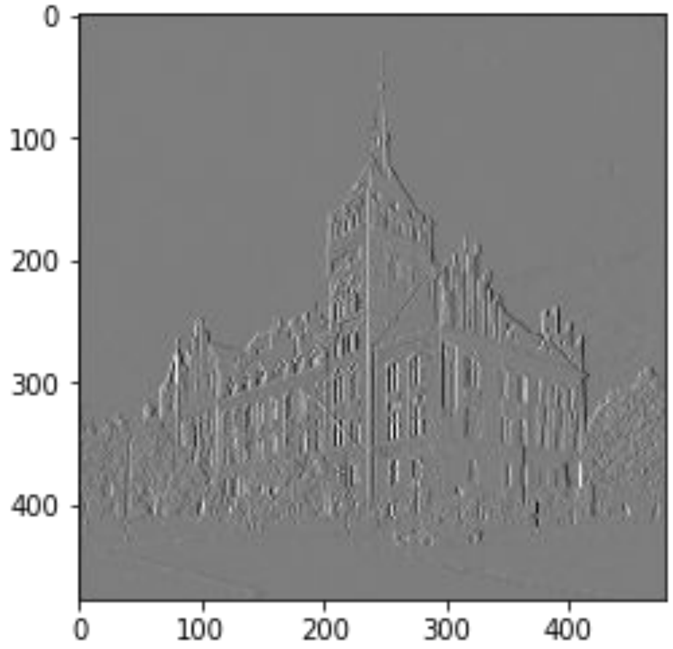
Silesian University of Technology, Faculty of Chemistry

$*$

1	0	-1
1	0	-1
1	0	-1

Filter size = 3 x 3

$=$



Code demonstration

Padding

0	0	0	0	0	0	0	0	0
0	2	0	5	3	5	8	2	0
0	8	2	5	2	6	4	4	0
0	3	5	4	2	6	6	1	0
0	3	6	2	6	4	3	7	0
0	7	5	3	4	3	3	2	0
0	4	6	2	3	6	2	8	0
0	2	3	1	6	2	3	6	0
0	0	0	0	0	0	0	0	0

9 x 9

padding = 1

*

2	1	3
1	2	5
2	4	2

Filter size = 3 x 3

=

40	61	53	66	89	67	36
50	90	66	95	109	96	46
69	100	70	104	100	85	54
92	88	90	90	83	88	44
88	86	89	76	84	87	56
74	72	72	87	71	93	56
41	33	58	48	43	76	27

7 x 7

Size of output = [n+2p-f+1] x [n+2p-f+1]

Code demonstration

Stride

hop 2 columns

hop 2 rows

$$\begin{bmatrix} 2 & 0 & 5 & 3 & 5 & 8 & 2 \\ 8 & 2 & 5 & 2 & 6 & 4 & 4 \\ 3 & 5 & 4 & 2 & 6 & 6 & 1 \\ 3 & 6 & 2 & 6 & 4 & 3 & 7 \\ 7 & 5 & 3 & 4 & 3 & 3 & 2 \\ 4 & 6 & 2 & 3 & 6 & 2 & 8 \\ 2 & 3 & 1 & 6 & 2 & 3 & 6 \end{bmatrix} * \begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 5 \\ 2 & 4 & 2 \end{bmatrix} = \begin{bmatrix} 90 & 95 & 96 \\ 88 & 90 & 88 \\ 72 & 87 & 93 \end{bmatrix}$$

7 x 7

Filter size = 3 x 3

Stride = 2

3 x 3

Size of output = $\lfloor (n+2p-f)/s + 1 \rfloor \times \lfloor (n+2p-f)/s + 1 \rfloor$

3 x 2 + 5 x 1 + 4 x 3 + 3 x 1 + Code demonstration + 5 x 4 + 3 x 2 = 88

Pooling Layers

Max Pooling Average Pooling

2	0	5	3	5	8	2
8	2	5	2	6	4	4
3	5	4	2	6	6	1
3	6	2	6	4	3	7
7	5	3	4	3	3	2
4	6	2	3	6	2	8
2	3	1	6	2	3	6

=

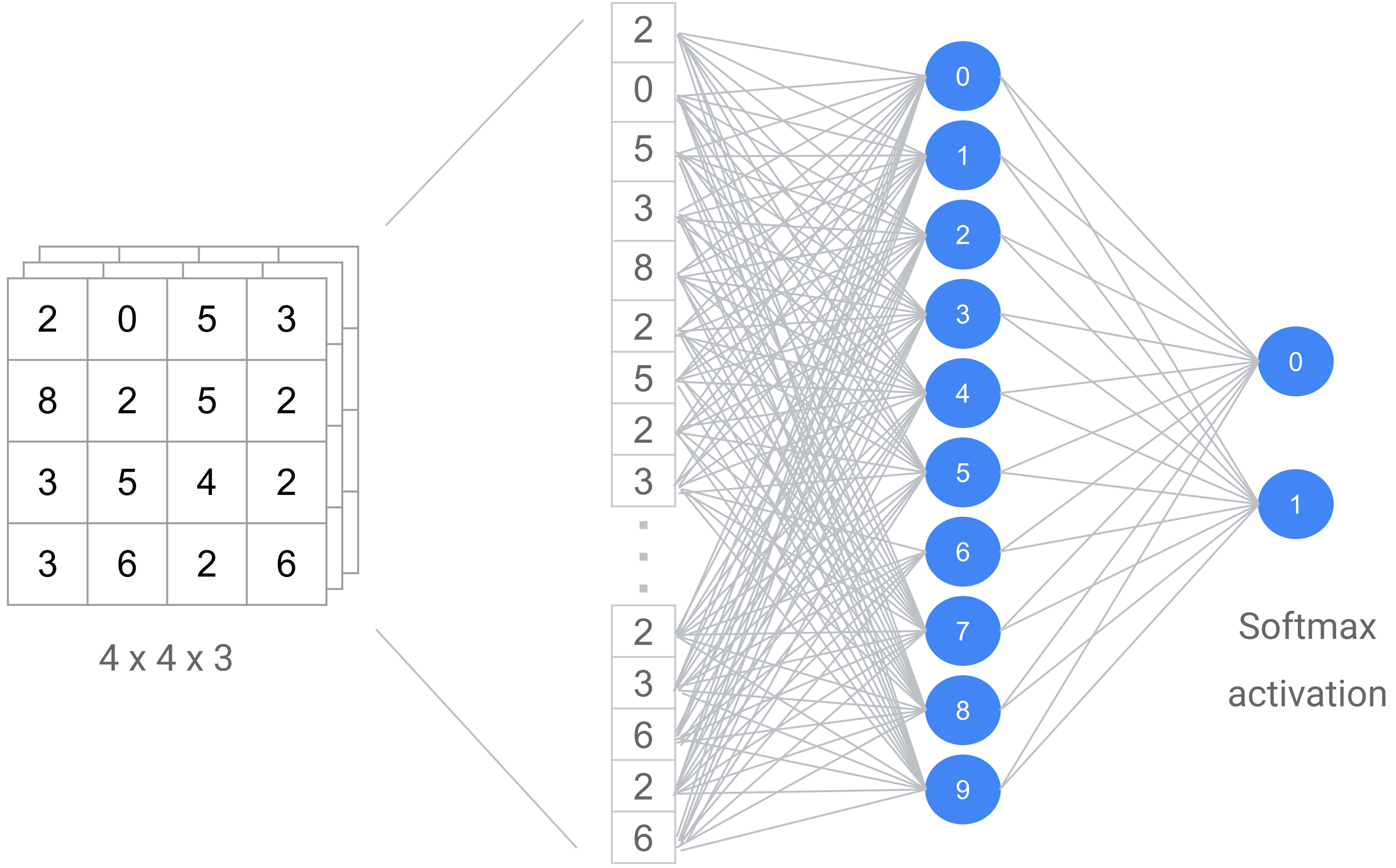
Filter size = 3 x 3

Stride = 2

3.7	4.2	4.6
4.2	3.7	3.8
3.6	3.3	3.8

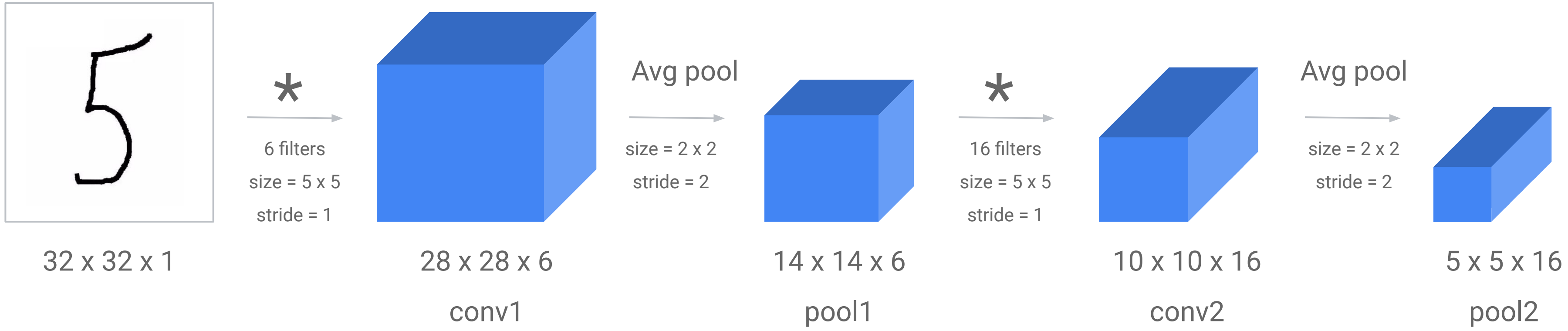
Code demonstration

Fully Connected Layers



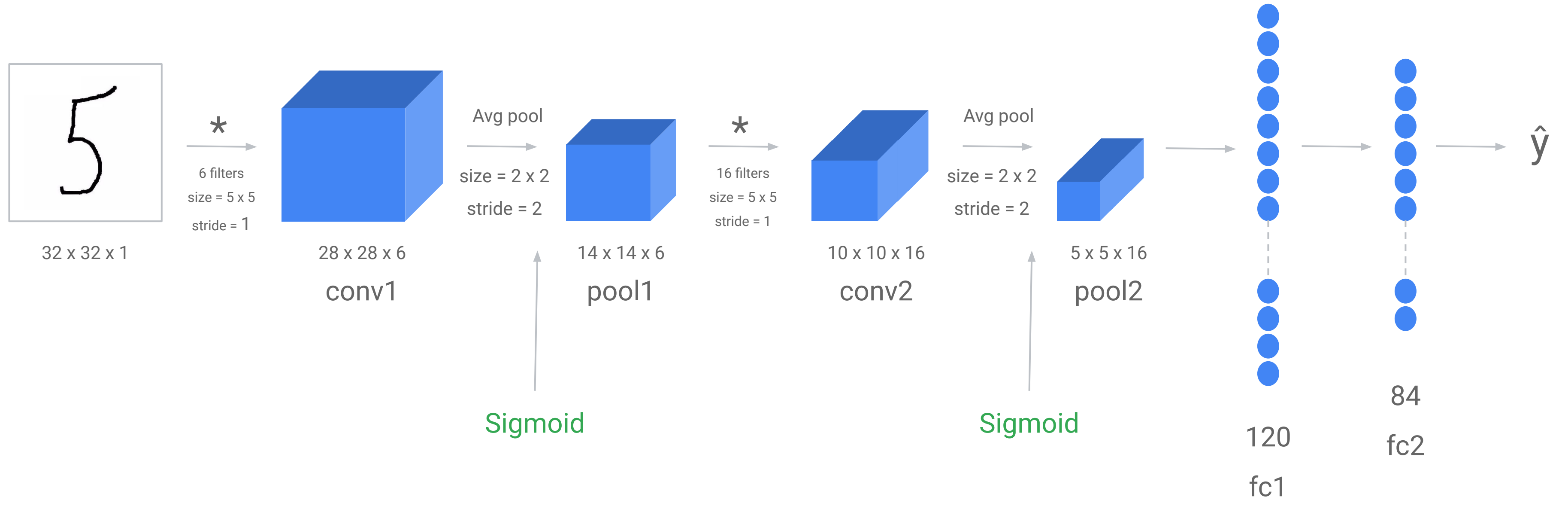
LeNet

LeNet - 5



[LeCun et al., 1998. Gradient-based learning applied to document recognition]

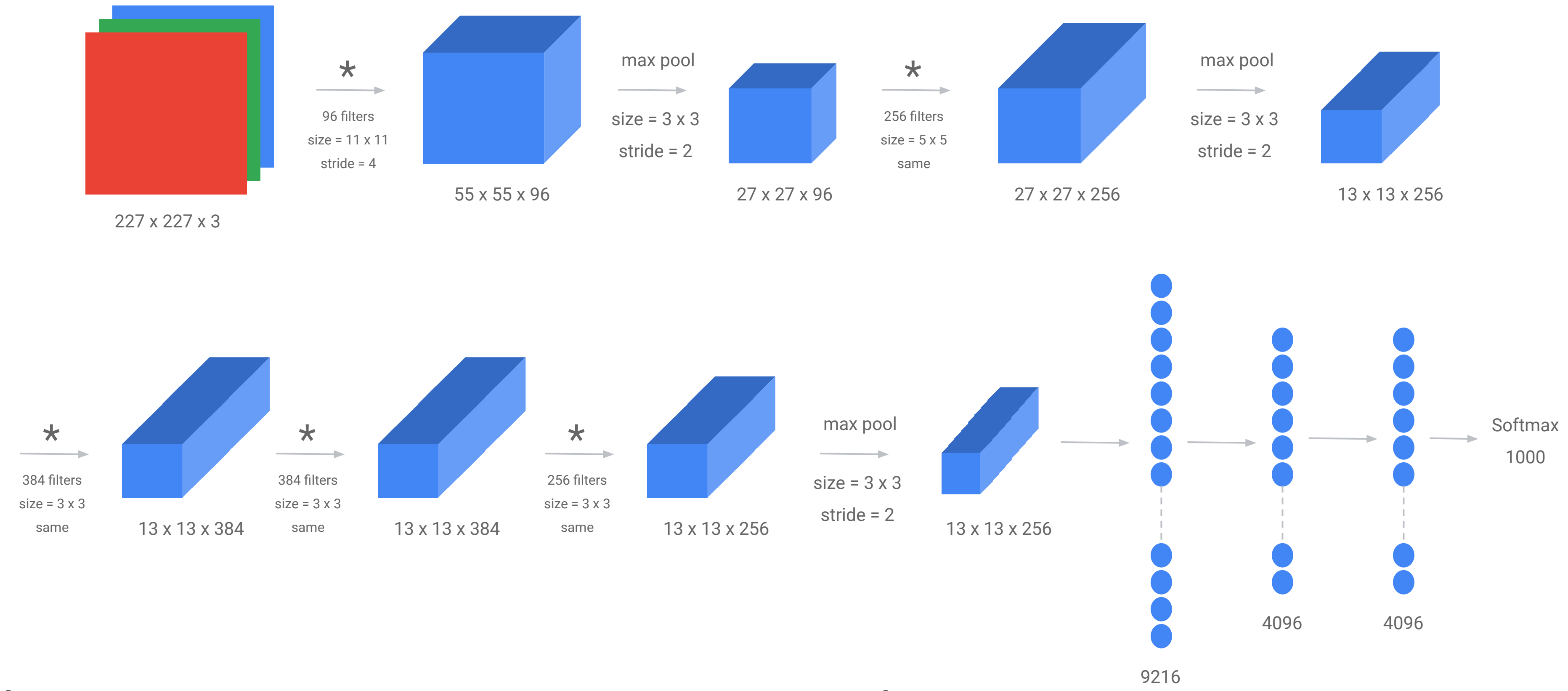
LeNet - 5



[LeCun et al., 1998. Gradient-based learning applied to document recognition]

AlexNet

AlexNet



[Krizhevsky et al., 2012. ImageNet classification with deep convolutional neural networks]

Network in Network

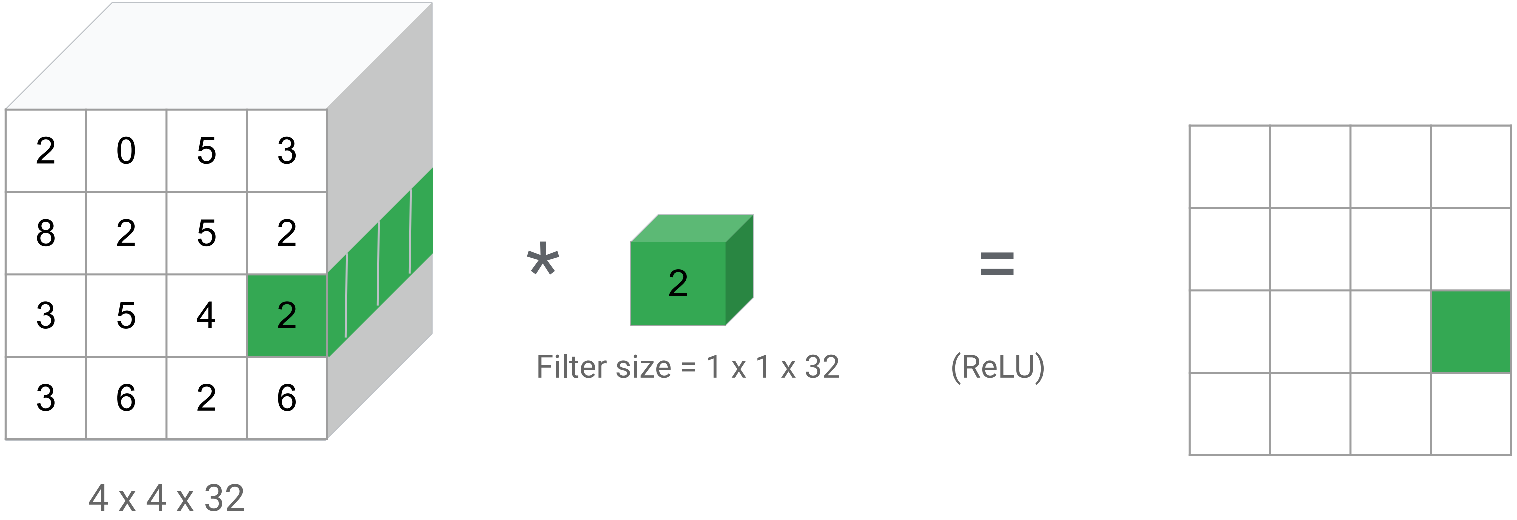
Network In Network

$$\begin{array}{|c|c|c|c|}\hline 2 & 0 & 5 & 3 \\ \hline 8 & 2 & 5 & 2 \\ \hline 3 & 5 & 4 & 2 \\ \hline 3 & 6 & 2 & 6 \\ \hline \end{array} * \boxed{2} = \begin{array}{|c|c|c|c|}\hline 4 & 0 & 10 & 6 \\ \hline 16 & 4 & 10 & 4 \\ \hline 6 & 10 & 8 & 4 \\ \hline 6 & 12 & 4 & 12 \\ \hline \end{array}$$

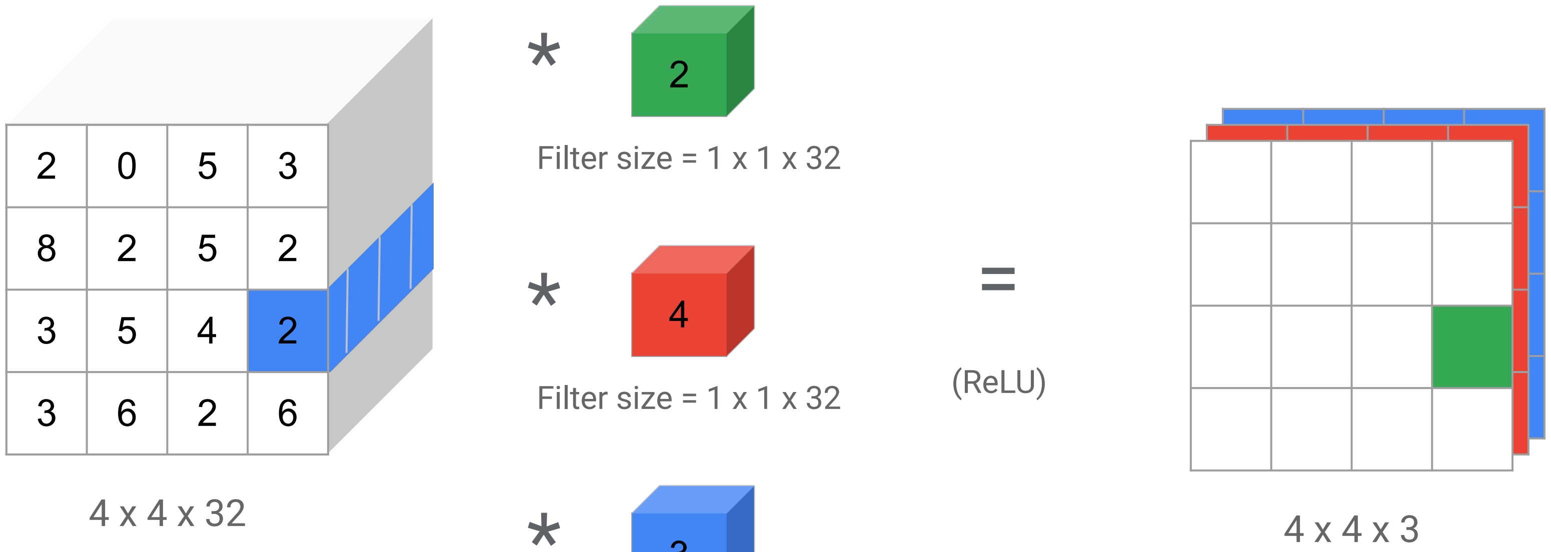
4 x 4 x 1

Filter size = 1 x 1

Network In Network

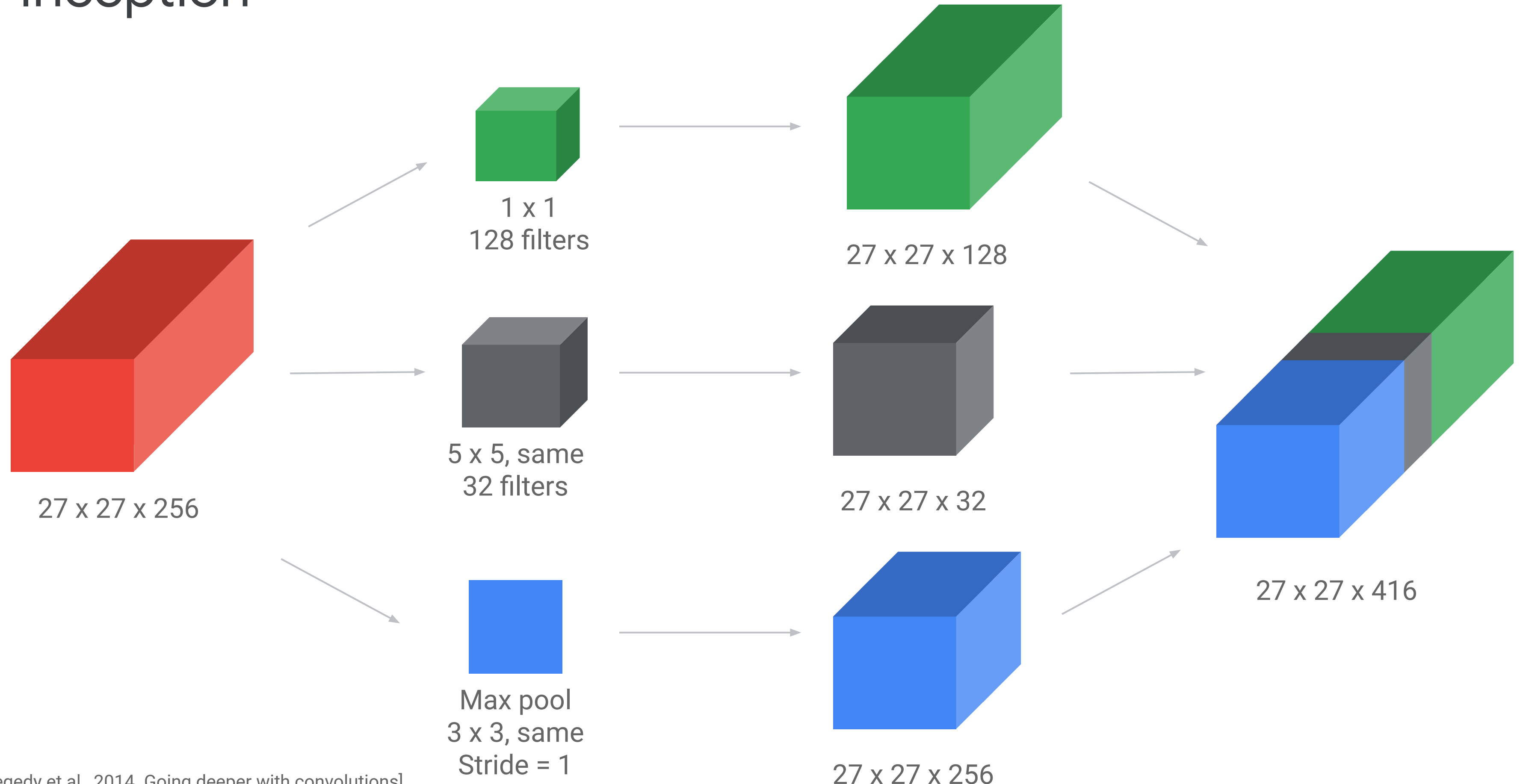


Network In Network



Inception

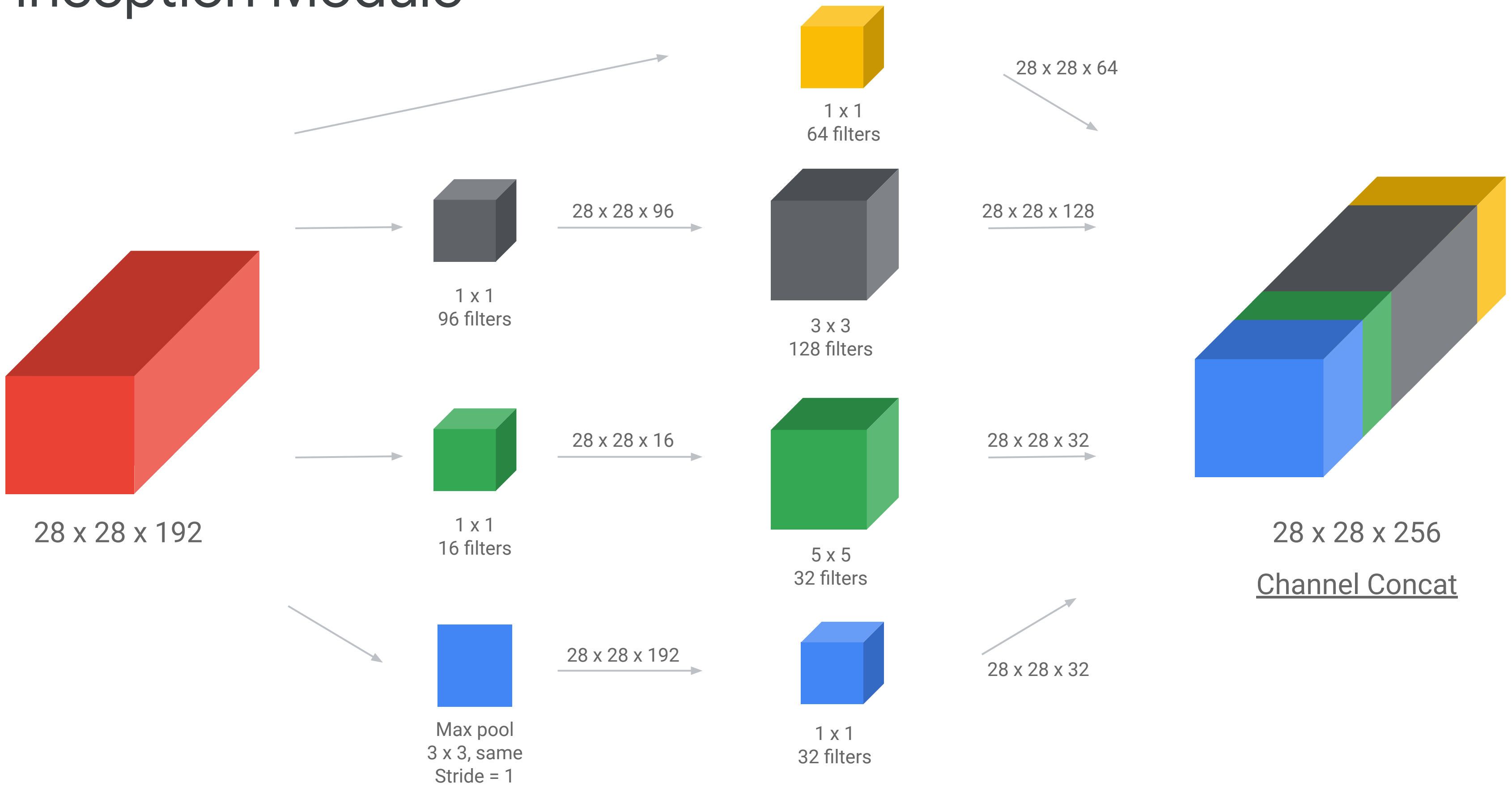
Inception



[Szegedy et al., 2014. Going deeper with convolutions]

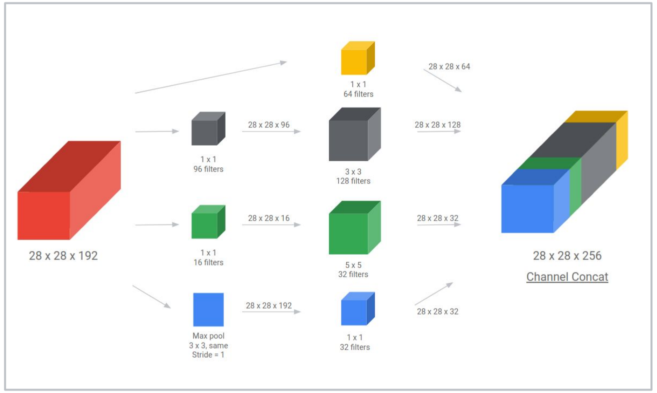
GoogLeNet

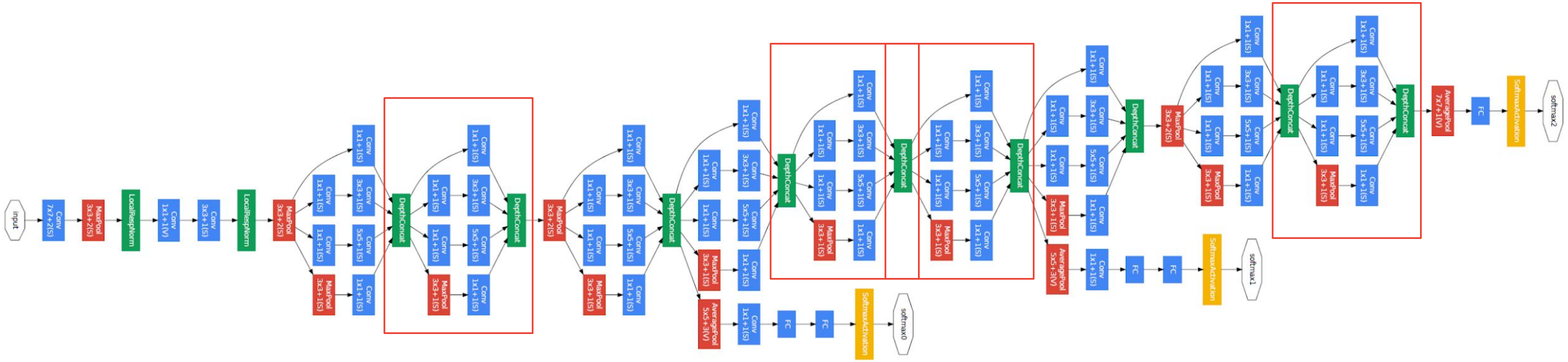
Inception Module



[Szegedy et al., 2014. Going deeper with convolutions]

GoogLeNet


28 x 28 x 192
1 x 1
64 filters
28 x 28 x 64
1 x 1
96 filters
28 x 28 x 96
3 x 3
128 filters
28 x 28 x 128
1 x 1
16 filters
28 x 28 x 16
5 x 5
32 filters
28 x 28 x 32
Max pool
3 x 3, same
Stride = 1
28 x 28 x 192
1 x 1
32 filters
28 x 28 x 32
28 x 28 x 256
Channel Concat


input
Conv 7x7+2(S)
MaxPool 3x3+2(S)
LocalRespNorm
Conv 1x1+1(V)
Conv 3x3+1(S)
LocalRespNorm
MaxPool 3x3+2(S)
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
MaxPool 3x3+1(S)
DepthConcat
AveragePool 5x5+3(V)
FC
SoftmaxActivation
softmax0
softmax1
AveragePool 7x7+1(V)
softmax2

Inception

Demo :
Detecting Airplanes in images

Questions!

See you next time!