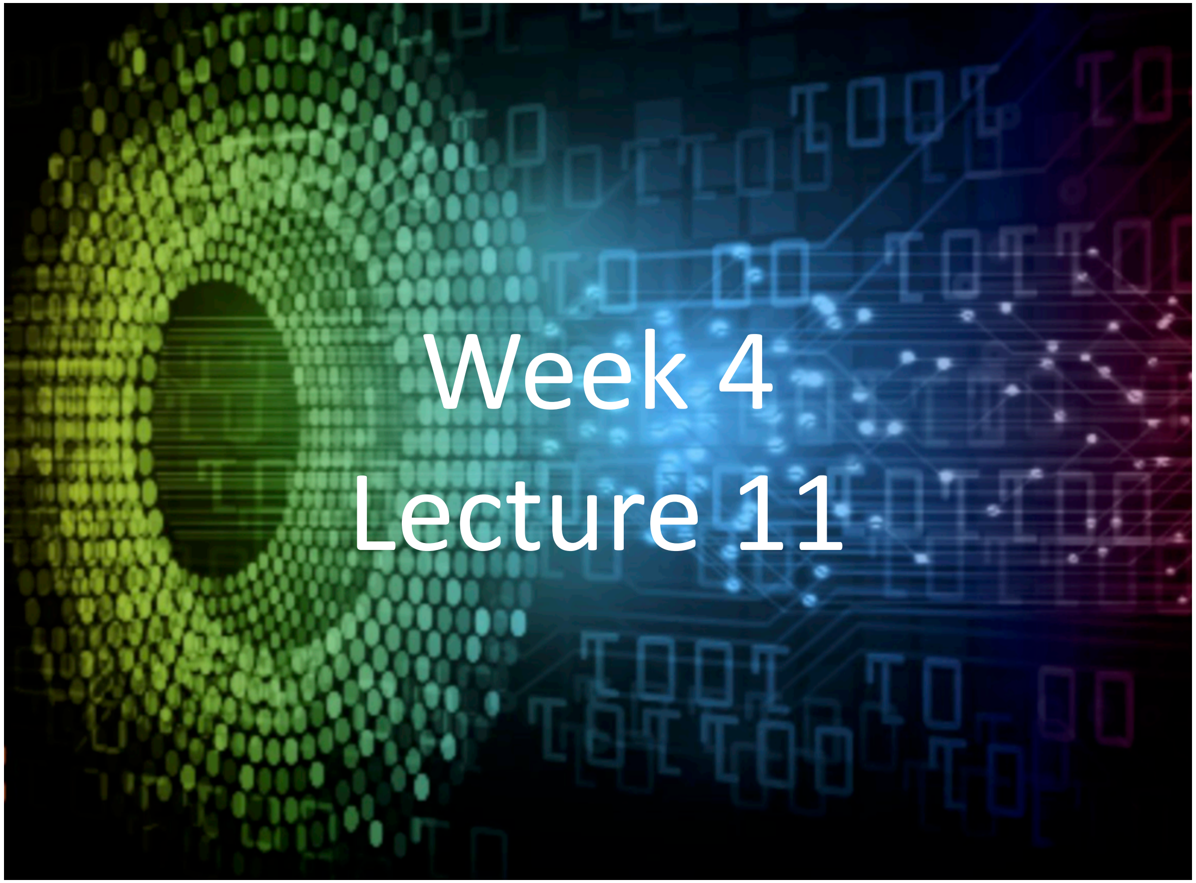
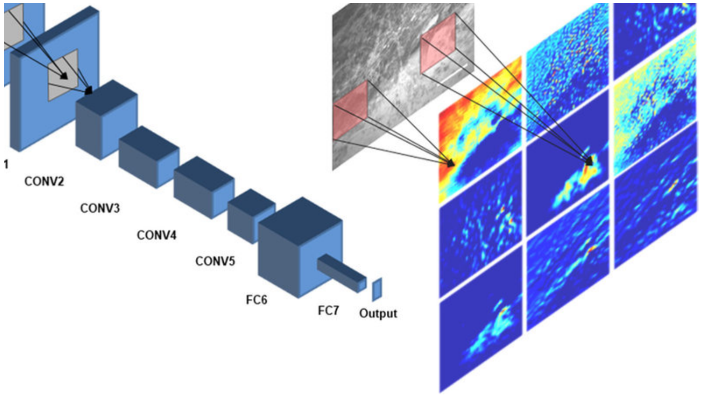
Introduction to Deep Learning Applications and Theory



ECE 596 / AMATH 563

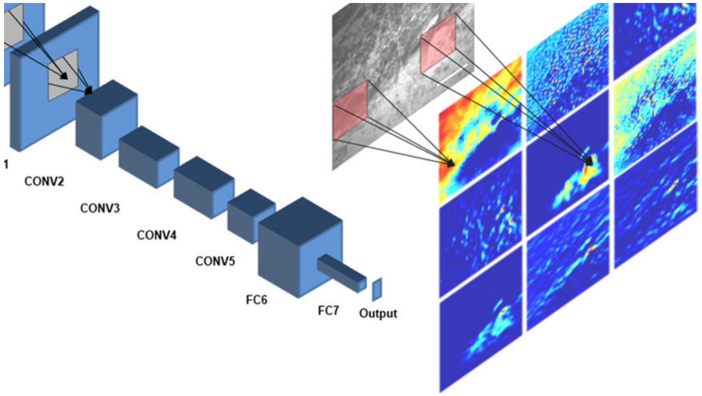
Previous Lecture: Convolutional Layer

1. The need for CNN
2. Convolutional Components Definition
    1. 2D Convolution
    2. Volume Convolution
    3. Padding, Stride
3. Complexity

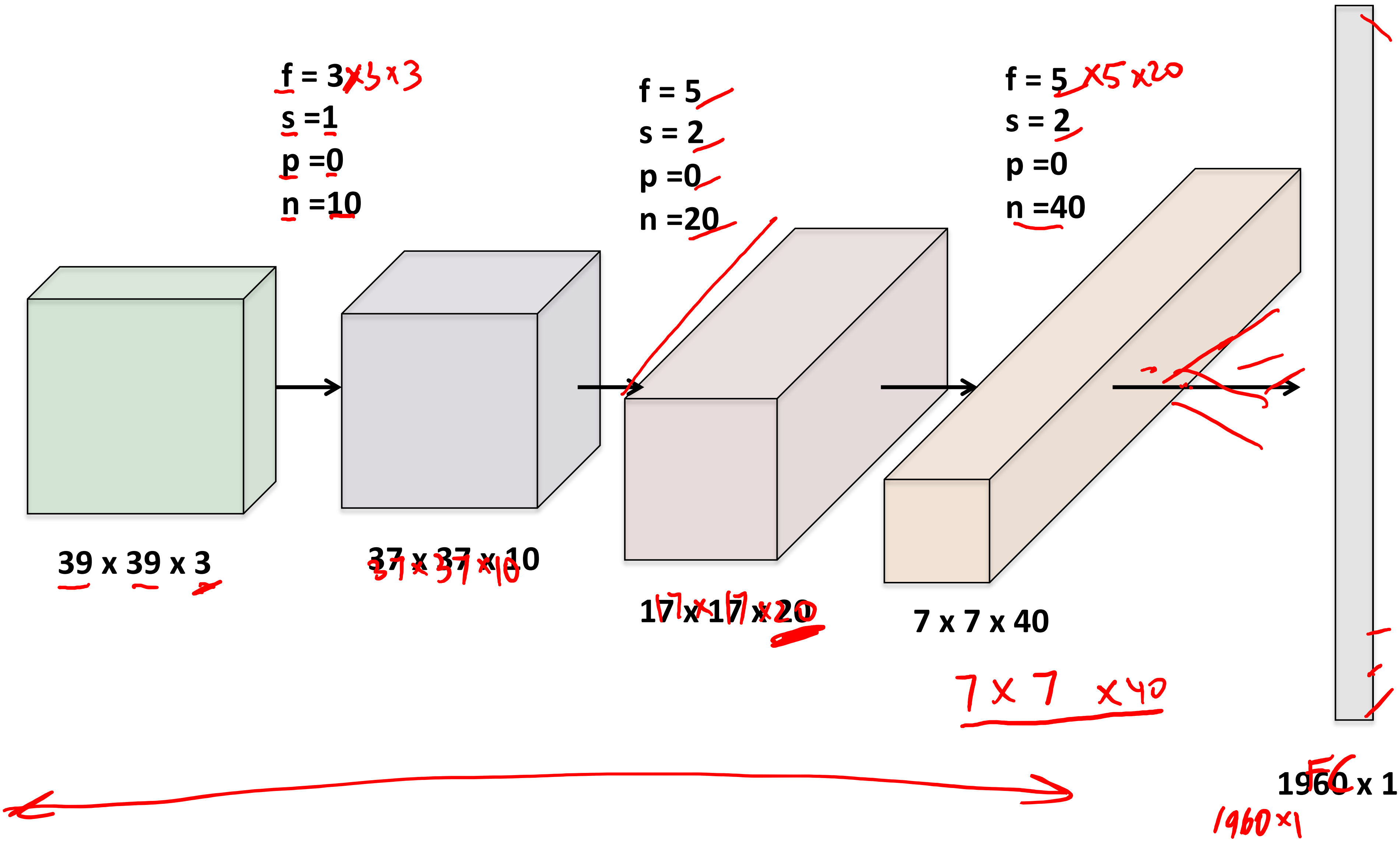


This Lecture: CNN

- **Pool Layer**
- **Combining all together**
- **Fundamental Networks (Case Studies)**
  - LeNet 5
  - AlexNet 7
  - VGG 16



CNN Example



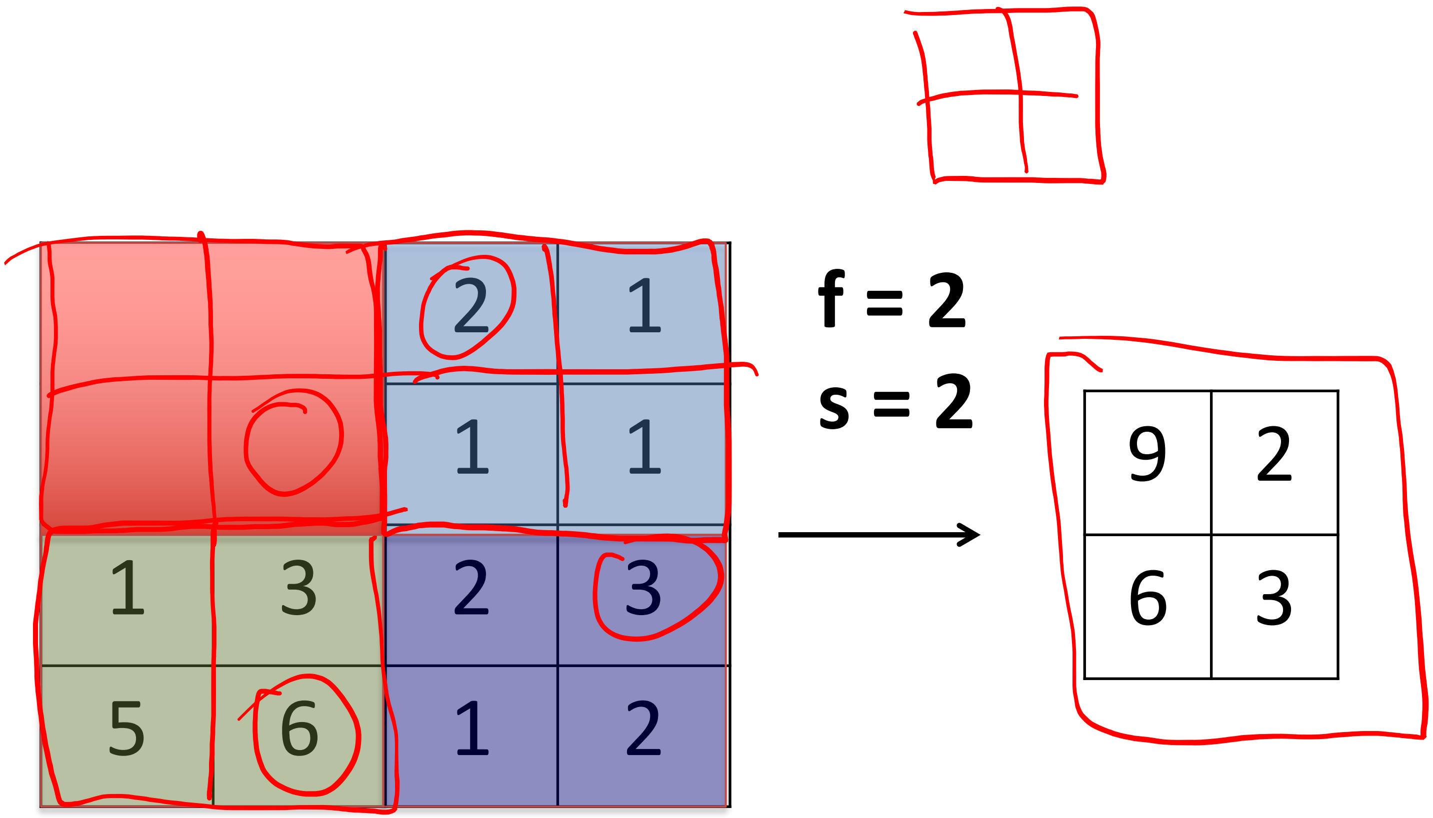
f = 3 x3x3
s =1
p =0
n =10
f = 5
s = 2
p =0
n =20
f = 5 x5x20
s = 2
p =0
n =40
39 x 39 x 3
37 x 37 x 10
37x37x10
17 x 17 x 20
17x17x20
7 x 7 x 40
7x7 x40
1960 x 1
FC
1960x1

Typical Layers

- **Convolutional Layer (CONV)**
- Pooling Layer (POOL)
- **Fully Connected (FC)**
- Normalization (NORM)

Pooling Layers (Max Pooling)

- Reduces the size of the feature map
- Speeds up computation



Max Pooling Example

1	3	2	1	3
2	9	1	1	5
1	3	2	3	2
8	3	5	1	0
5	6	1	2	9

5x 5 x n_c

f = 3

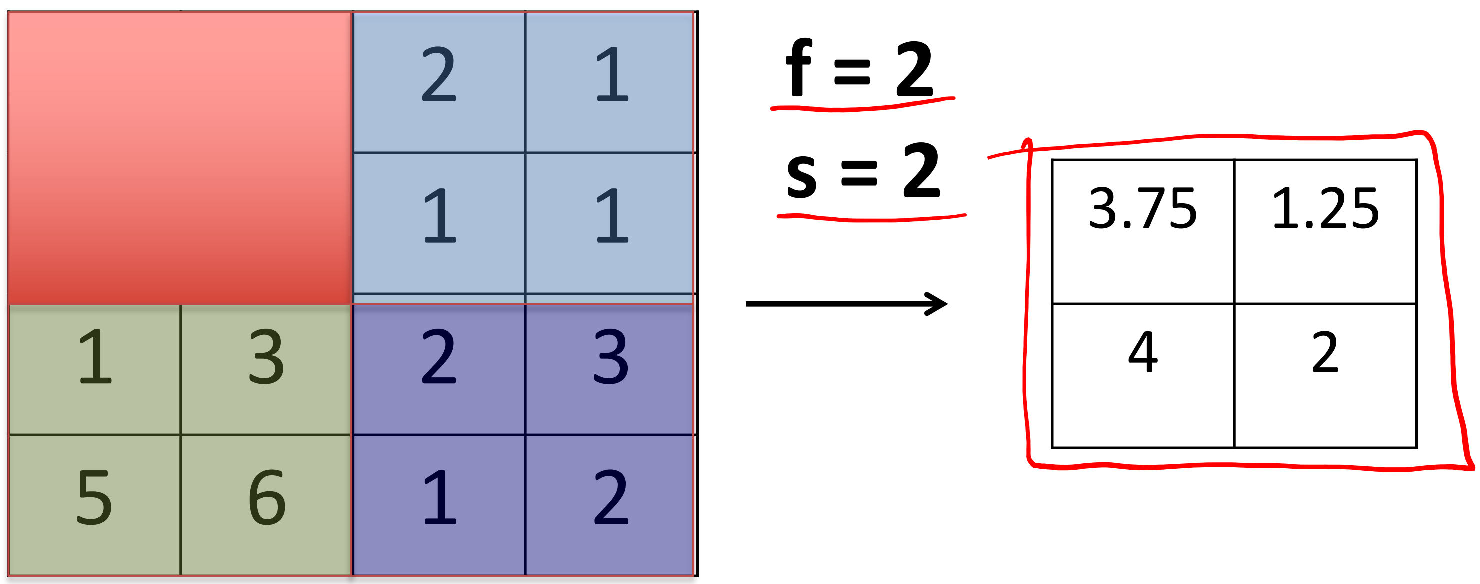
s = 1

9	9	5
9	9	5
8	6	9

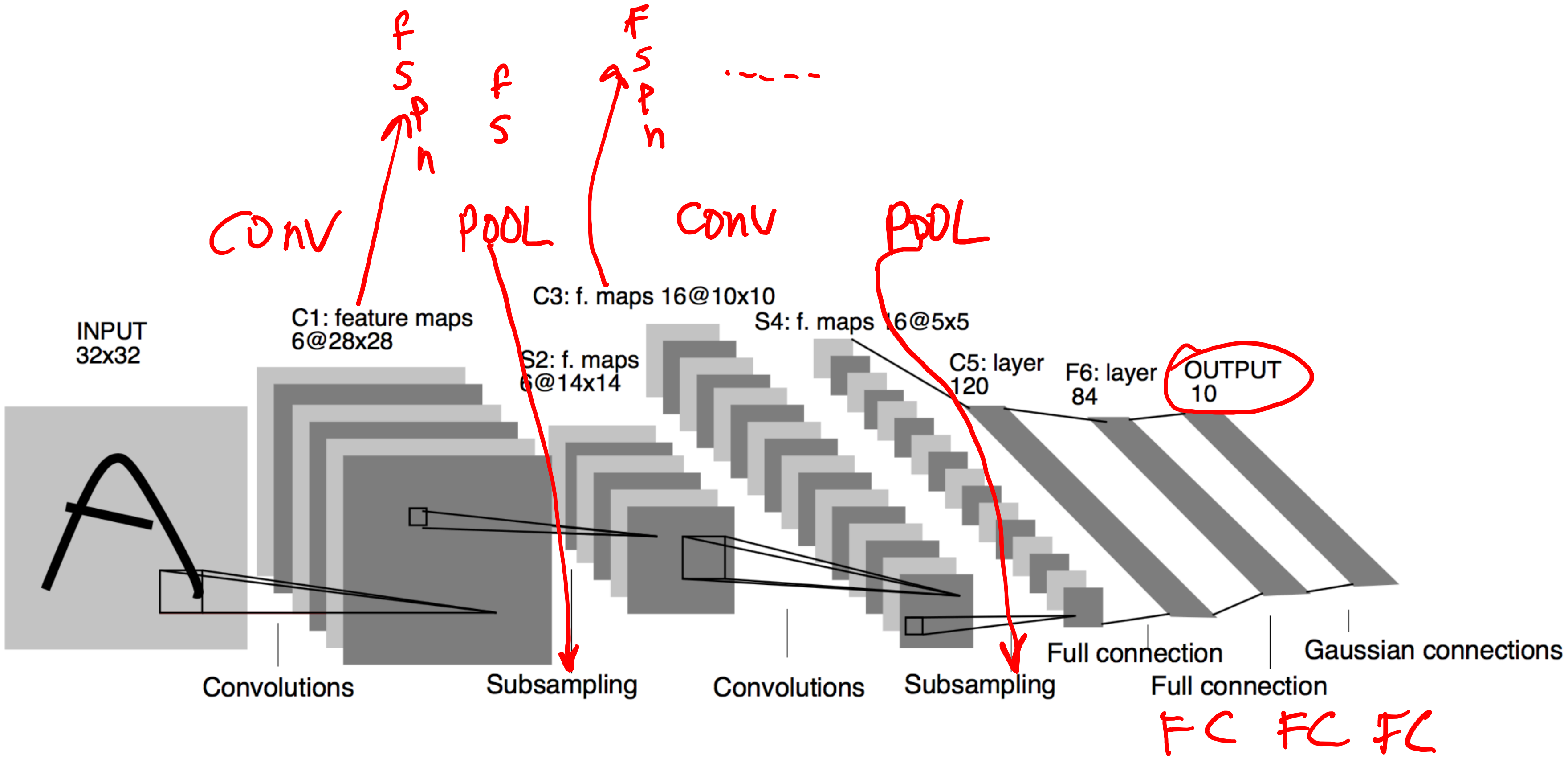
3 x 3 x n_c

Other Types of Pooling

- Max -> Average (also L2 norm)

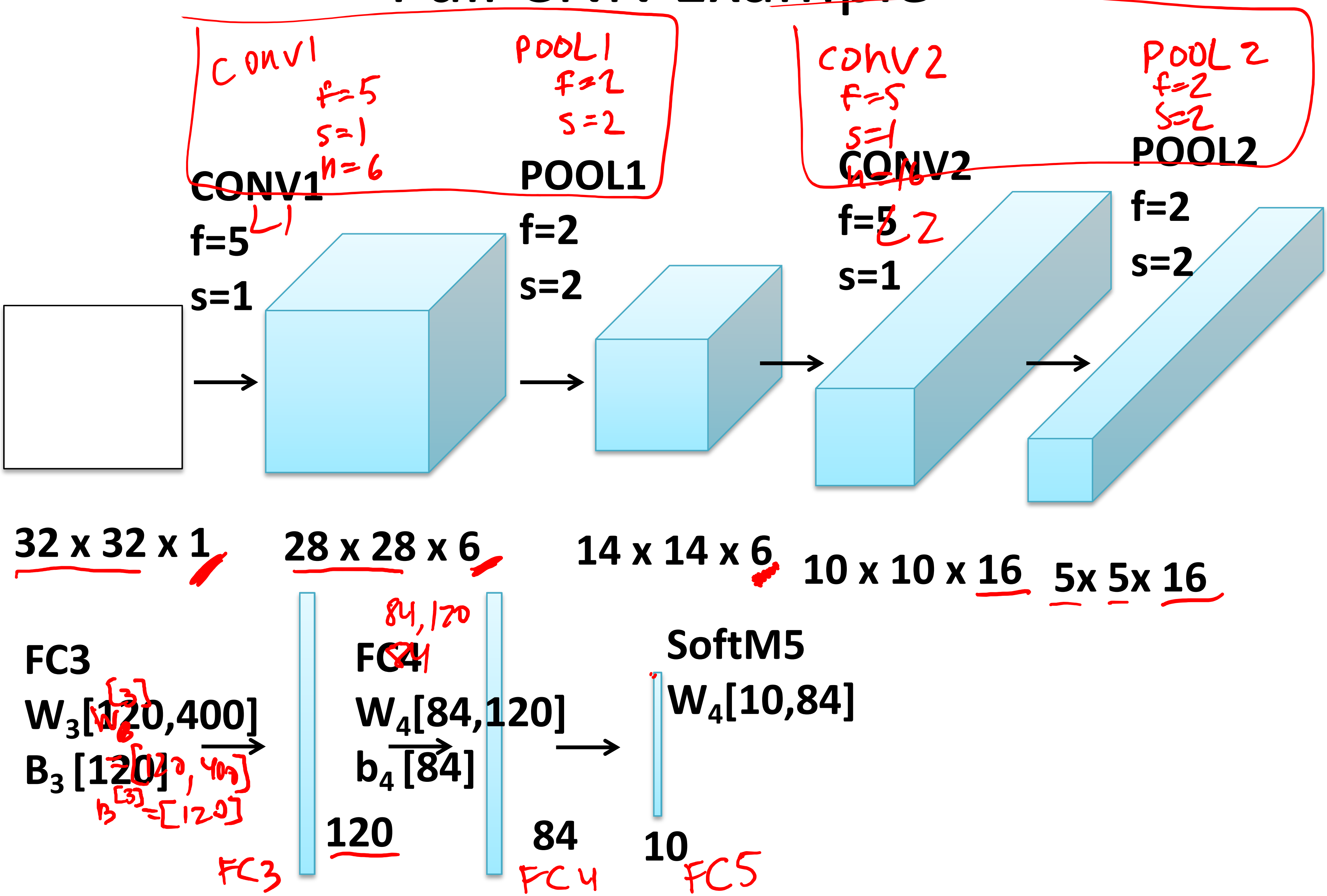


LeNet 5



f
s
p
n
f
s
F
s
p
n
CONV
POOL
Conv
POOL
C3: f. maps 16@10x10
S4: f. maps 16@5x5
INPUT
32x32
C1: feature maps
6@28x28
S2: f. maps
6@14x14
C5: layer
120
F6: layer
84
OUTPUT
10
Convolutions
Subsampling
Convolutions
Subsampling
Full connection
Full connection
Gaussian connections
FC FC FC

Full CNN Example



Benefit of Convolutions (Pros)

- **Parameter Sharing**
  - Filter can be useful in different parts of the input (image)

- **Sparsity of Connections**
  - In each layer each output value depends only on small number of inputs (local)
  - Translation invariance

Challenges of Convolutions (Cons?)

- **Computational Complexity**
  - Convolutions are expensive $O(N^2n^4)$

input dim

filter/channel dim

- **Deeper Structure Needed**
  - In each layer each output value depends only on small number of inputs (local)

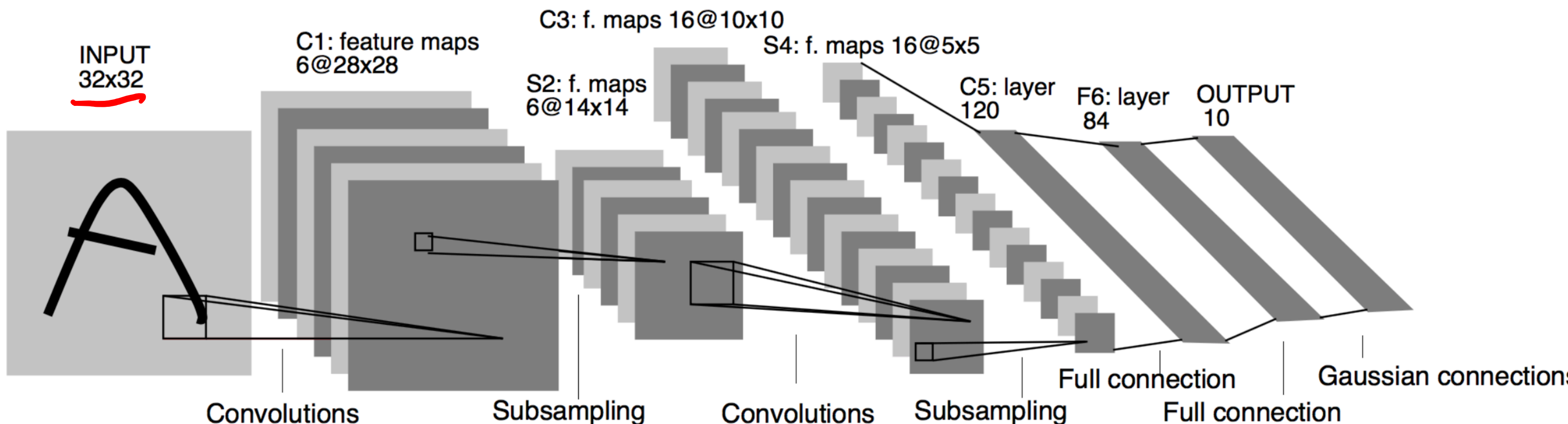
LeNet 5

Yann LeCun

Leon Bottou

Yoshua Bengio

Patrick Haffner



Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998.

Parameters

- CONV Layers ~= 1,500
- POOL Layers = 0
- FC Layers ~= 60,000
- Total ~= **61,500**

1.5k

60k

Classic Networks

- LeNet 5 –
  – Handwritten digits recognition

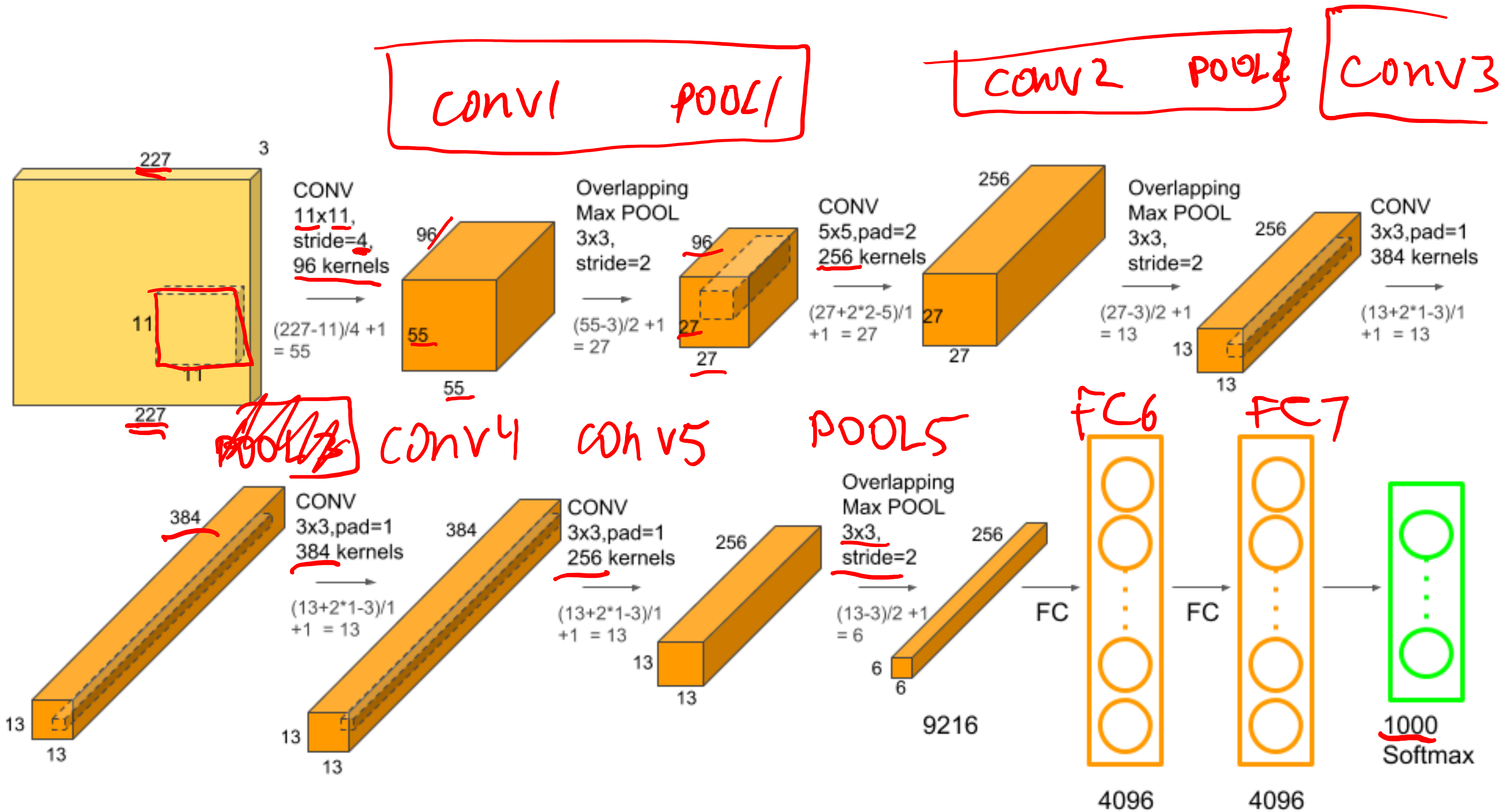
32×32×1 input

- AlexNet (7)
  – Object Classification

227×227×3 input

- VGG (16)
  - Recognition / Classification

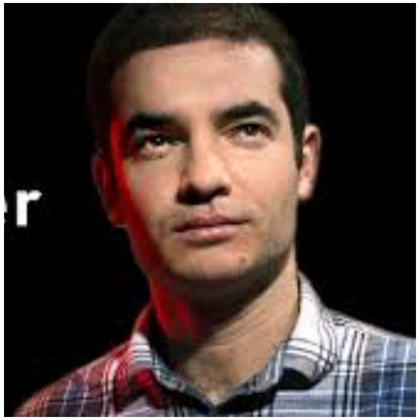
AlexNet (7)



conv1
pool1
conv2
pool2
conv3
227
3
227
11
11
CONV
11x11,
stride=4,
96 kernels
(227-11)/4 +1
= 55
96
55
55
Overlapping
Max POOL
3x3,
stride=2
(55-3)/2 +1
= 27
96
27
27
CONV
5x5,pad=2
256 kernels
(27+2*2-5)/1
+1 = 27
256
27
27
Overlapping
Max POOL
3x3,
stride=2
(27-3)/2 +1
= 13
256
13
13
CONV
3x3,pad=1
384 kernels
(13+2*1-3)/1
+1 = 13
pool3
conv4
conv5
pool5
FC6
FC7
384
13
13
CONV
3x3,pad=1
384 kernels
(13+2*1-3)/1
+1 = 13
384
13
13
CONV
3x3,pad=1
256 kernels
(13+2*1-3)/1
+1 = 13
256
13
13
Overlapping
Max POOL
3x3,
stride=2
(13-3)/2 +1
= 6
256
6
6
9216
FC
4096
FC
4096
1000
Softmax

AlexNet

Alex Krizhevsky



Ilya Sutskever

Jeoffrey Hinton

Krizhevsky et al., Imagenet classification with deep convolutional neural networks, 2012

14 years

Parameters

Layer Name	Tensor Size	Weights	Biases	Parameters
Input Image	227x227x3	0	0	0
Conv-1	55x55x96	34,848	96	34,944
MaxPool-1	27x27x96	0	0	0
Conv-2	27x27x256	614,400	256	614,656
MaxPool-2	13x13x256	0	0	0
Conv-3	13x13x384	884,736	384	885,120
Conv-4	13x13x384	1,327,104	384	1,327,488
Conv-5	13x13x256	884,736	256	884,992
MaxPool-3	6x6x256	0	0	0
FC-1	4096×1	37,748,736	4,096	37,752,832
FC-2	4096×1	16,777,216	4,096	16,781,312
FC-3	1000×1	4,096,000	1,000	4,097,000
Output	1000×1	0	0	0
Total				**62,378,344**

3M

57M

60M vs 60k

AlexNet Specifics

- Much bigger than LeNet (**60M** parameters)
- ReLU
- Multiple GPUs
- Local Response Normalization (LRN)

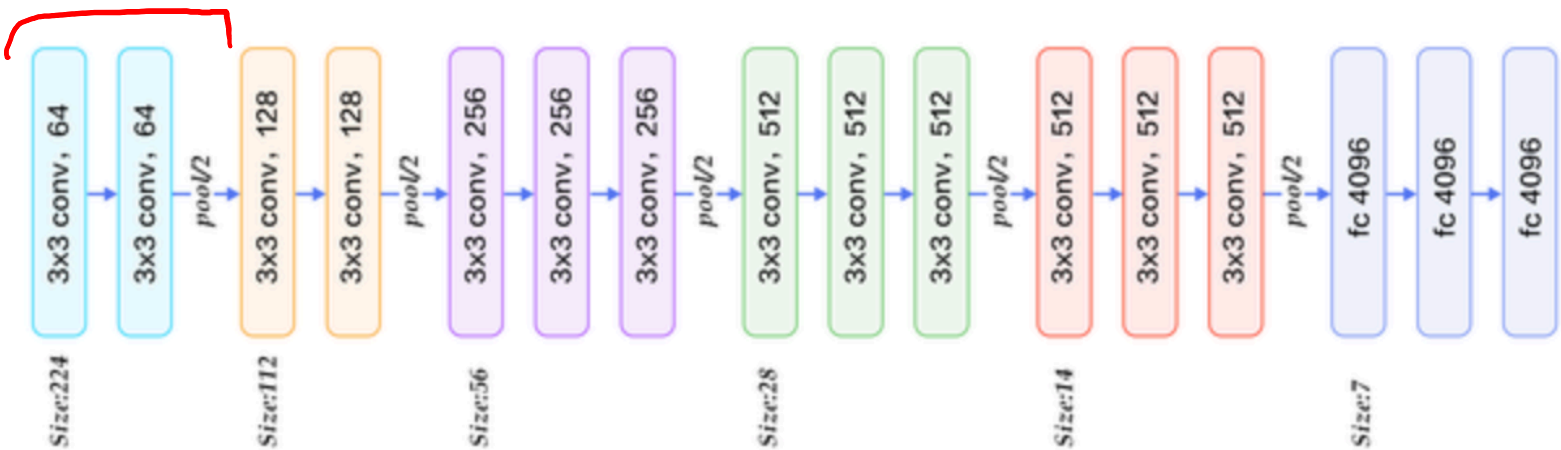
VGG - 16 / 19

CONV: f=3, s=1, same

POOL: f=2, s=2

Order: $\underbrace{CCP}_{2}$ $\underbrace{CCP}_{2}$ $\underbrace{CCCP}_{3}$ $\underbrace{CCCP}_{3}$ $\underbrace{CCCP}_{3}$ FFS

Nf: 2^6 2^7 2^8 2^9 2^9



VGG Specifics

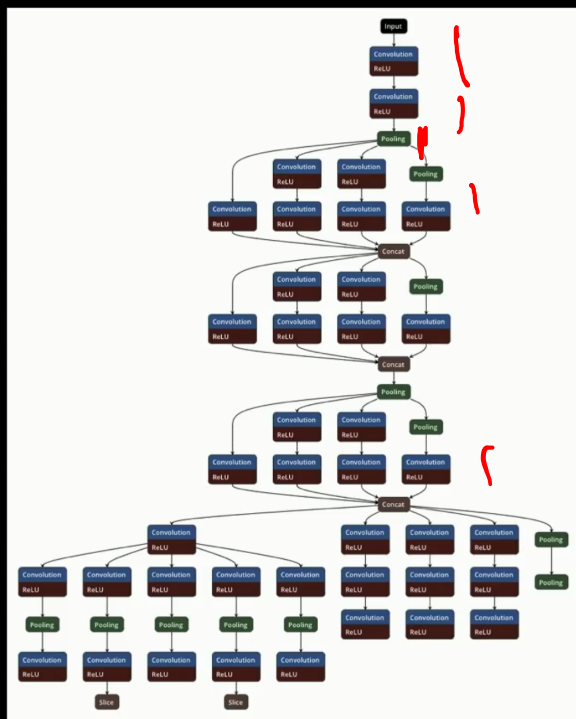
- From Visual Geometry Group (2015)
- Organized Structure
- More Layers – Deeper
- Parameters: **138M**

	L5	A7	V16
Layers	5	7	16
Params	60k	60M	~140M
Input	1k	50k	50k

Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. 2014.

Goal

Tesla NN



Operation	MOPS	%
Convolution	34275	98.1
Deconvolution	576	1.6
ReLU	123	0.1
Pooling	13	0.2

99.7% of operations are multiply add