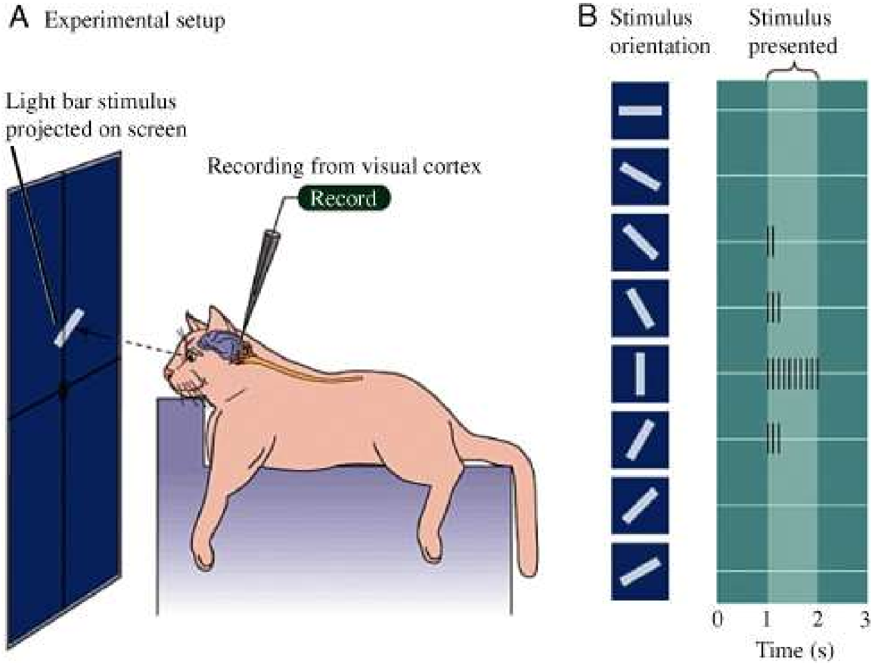
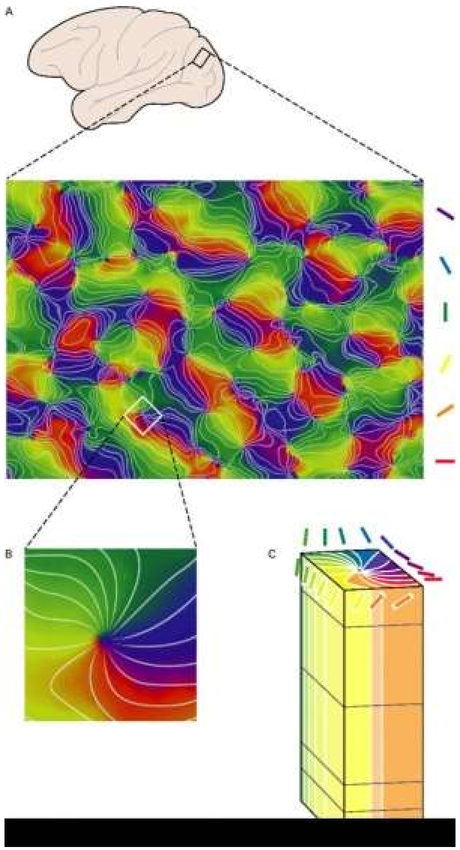
Convolutional Neural Networks

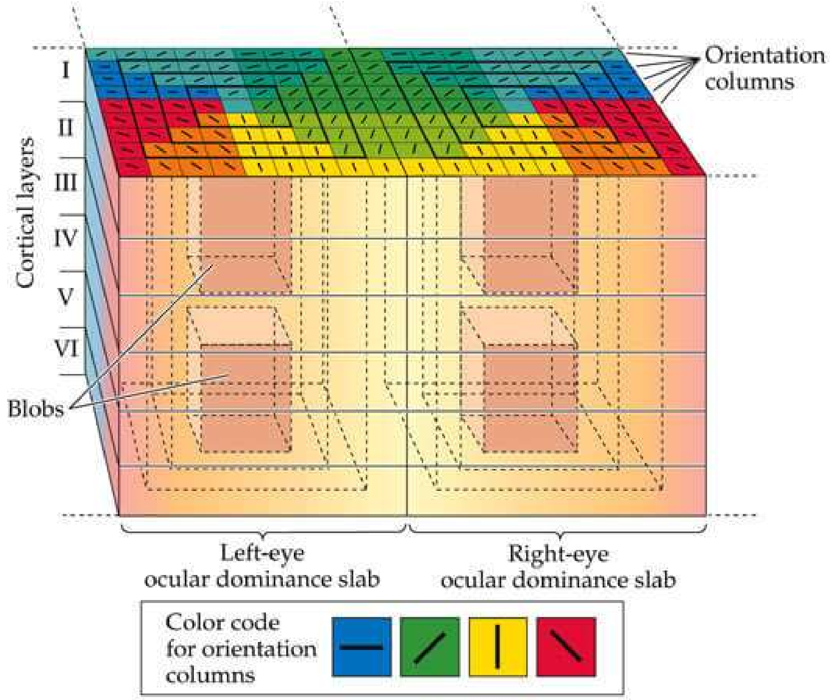
Hubel & Wiesel (1962)



Nobel Prize in Medicine and Physiology (1981) for their discoveries concerning information processing in the visual system

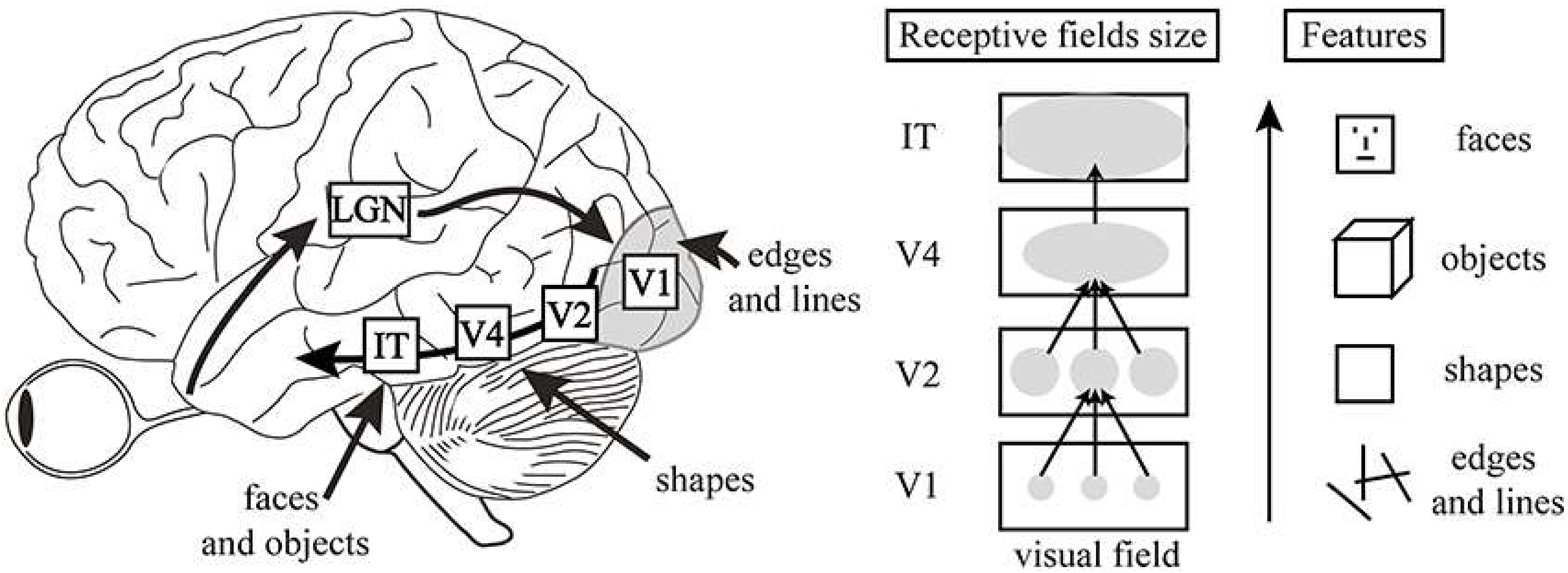
Orientation & Ocular Dominance Columns of the Visual Cortex



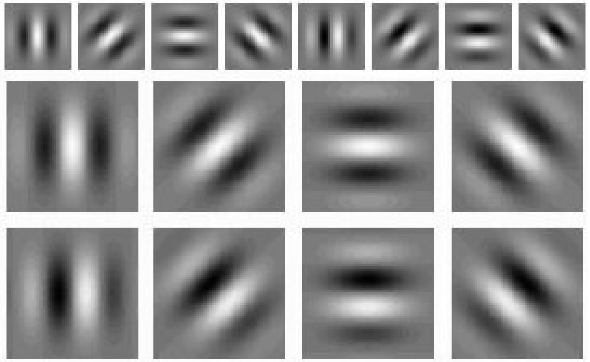


- **Topographic organization of oriented edge detectors in V1** -- neighboring regions of neurons have similar orientation tuning, as shown in this colorized map where different colors indicate orientation preference as shown in panel C. Panel B shows how a full 360 degree loop of orientations nucleate around a central point -- these are known as pinwheel structures.

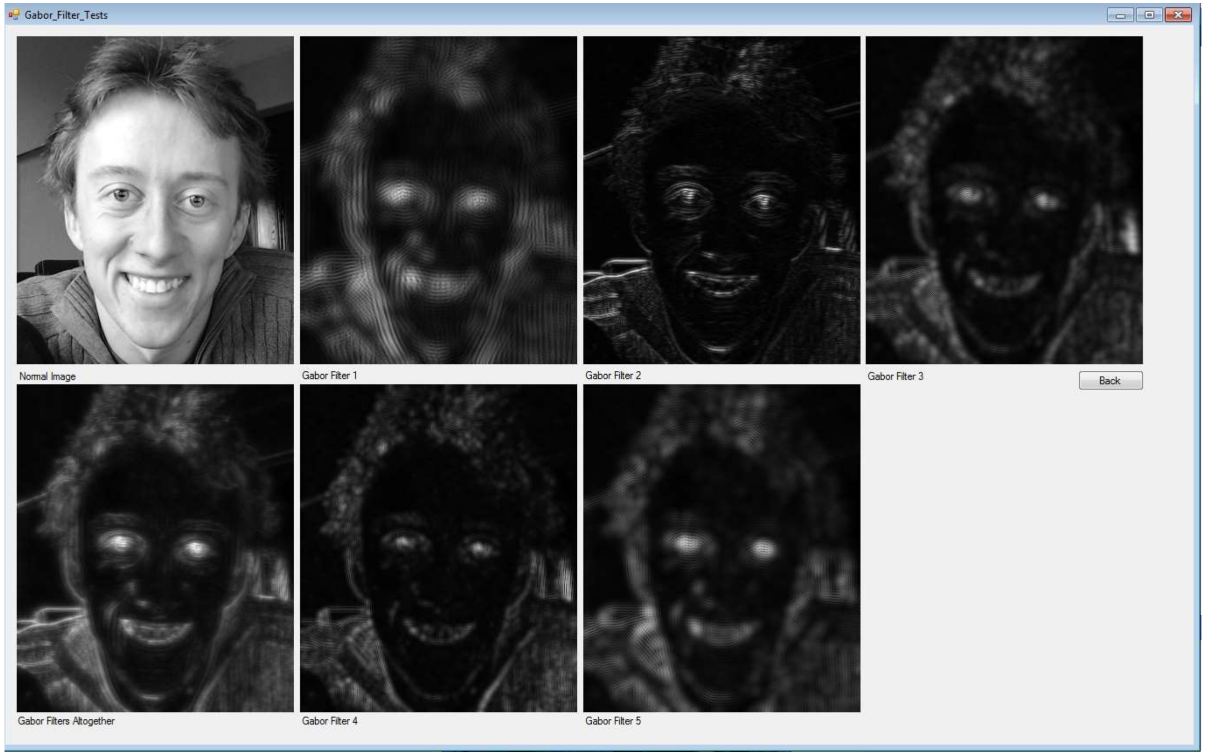
Visual Processing of The Brain



Gabor Filtering at the Visual Cortex (V1)

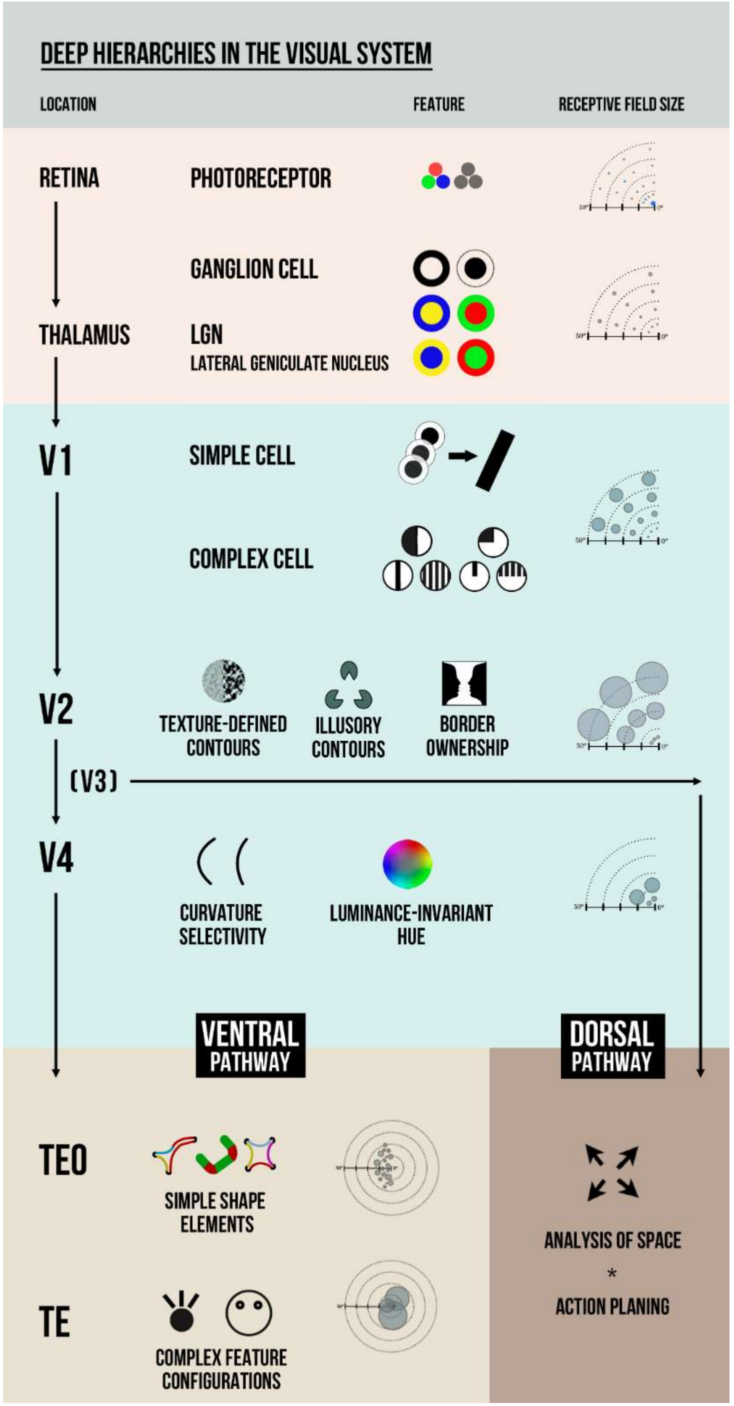


- Various Gabor Filters

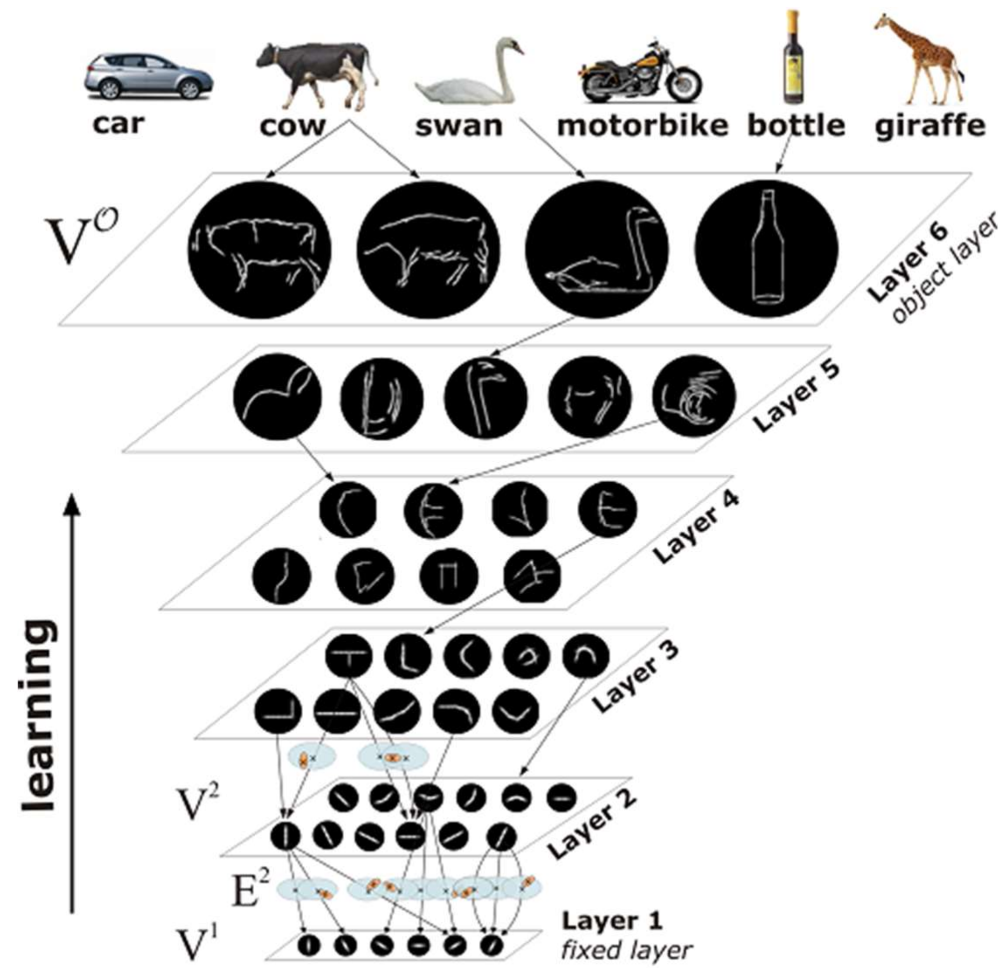


- Original and filtered images

Feature maps


DEEP HIERARCHIES IN THE VISUAL SYSTEM
LOCATION
FEATURE
RECEPTIVE FIELD SIZE
RETINA
PHOTORECEPTOR
GANGLION CELL
THALAMUS
LGN
LATERAL GENICULATE NUCLEUS
V1
SIMPLE CELL
COMPLEX CELL
V2
TEXTURE-DEFINED CONTOURS
ILLUSORY CONTOURS
BORDER OWNERSHIP
(V3)
V4
CURVATURE SELECTIVITY
LUMINANCE-INVARIANT HUE
VENTRAL PATHWAY
DORSAL PATHWAY
TEO
SIMPLE SHAPE ELEMENTS
TE
COMPLEX FEATURE CONFIGURATIONS
ANALYSIS OF SPACE
*
ACTION PLANING

Hierarchical Visual Representation

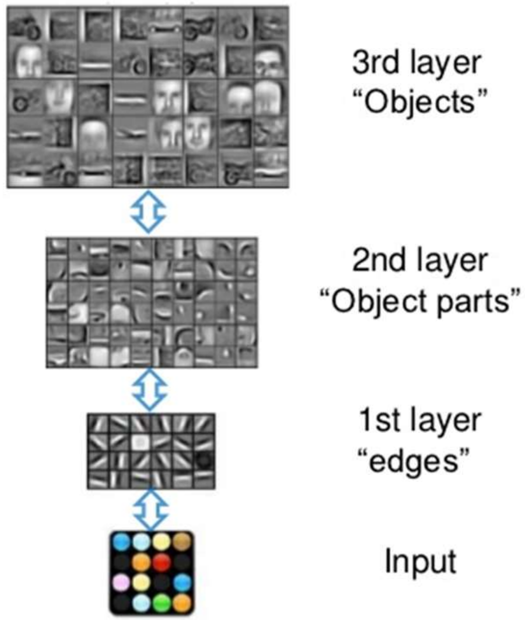


Hierarchical Processing

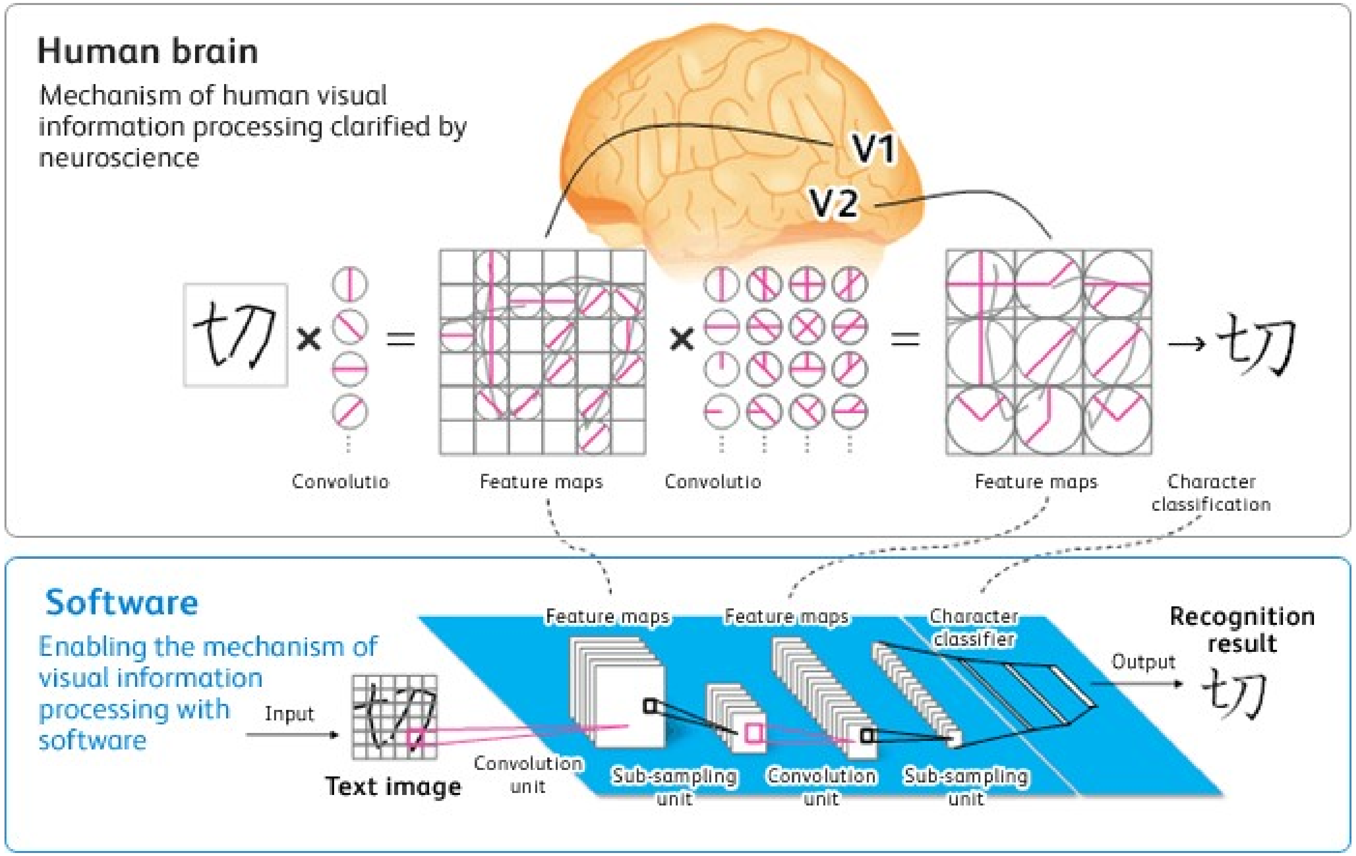
Deep Learning = Multiple Levels of Feature Representation

Learning Feature Hierarchy

- Deep Learning
  - Deep architectures can be representationally efficient.
  - Natural progression from low level to high level structures.
  - Can share the lower-level representations for multiple tasks.



Brain vs. Deep Learning

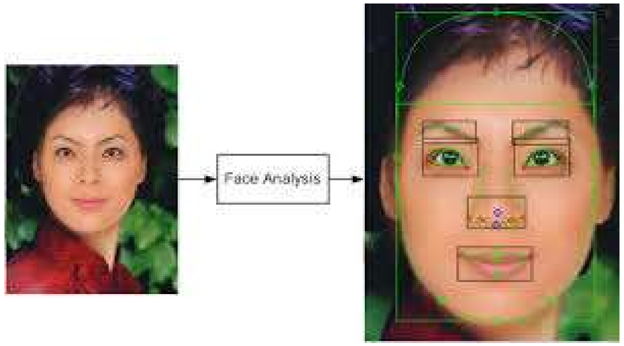


Convolutional Neural Network (CNN)

Deep Learning = Multiple Levels of Feature Representation

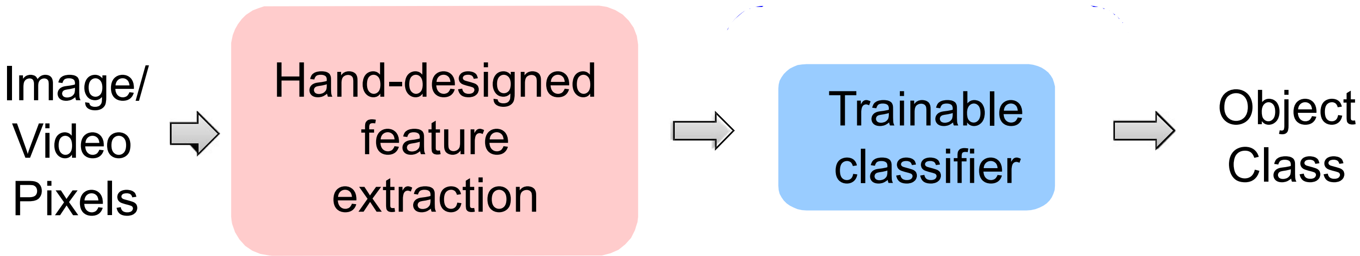
Traditional Machine Learning Solution: Feature Extraction & Classifier

- **Supervised learning**
  - Features are **not** learned, but **extracted** by humans



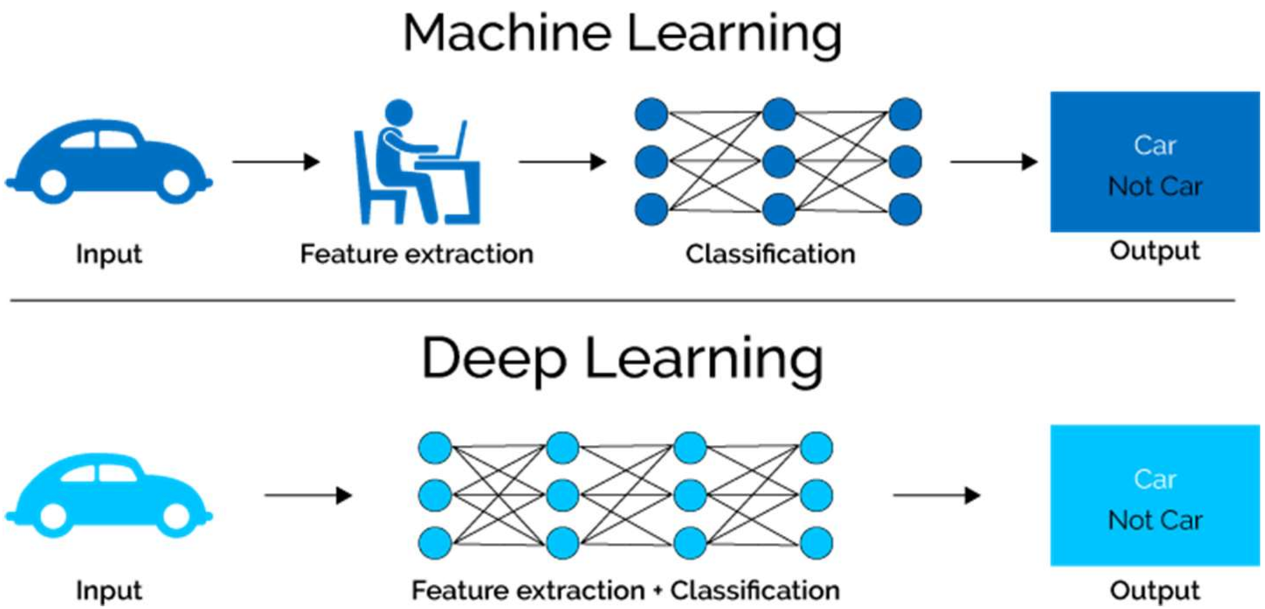
Facial Feature Extraction

  - Trainable classifier (i.e., shallow networks)



Feature Extraction via Convolution Filter

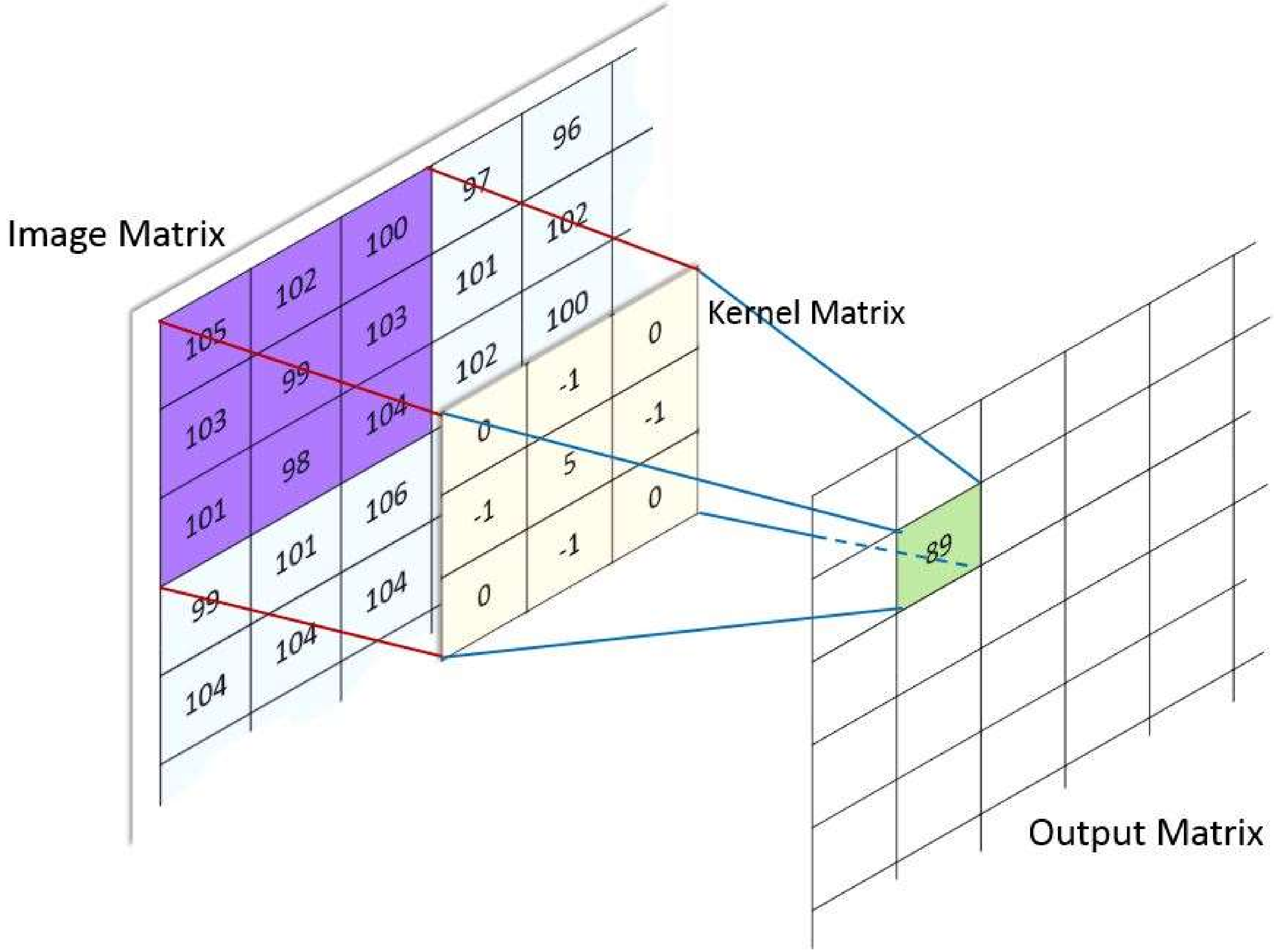
Paradigm Shift in Machine Learning



Convolution

- **Convolution** is a simple mathematical operation which is fundamental to many common **image** processing operators. **Convolution** provides a way of `multiplying together' two arrays of numbers, generally of different sizes, but of the same dimensionality, to produce a third array of numbers of the same dimensionality.

Convolution



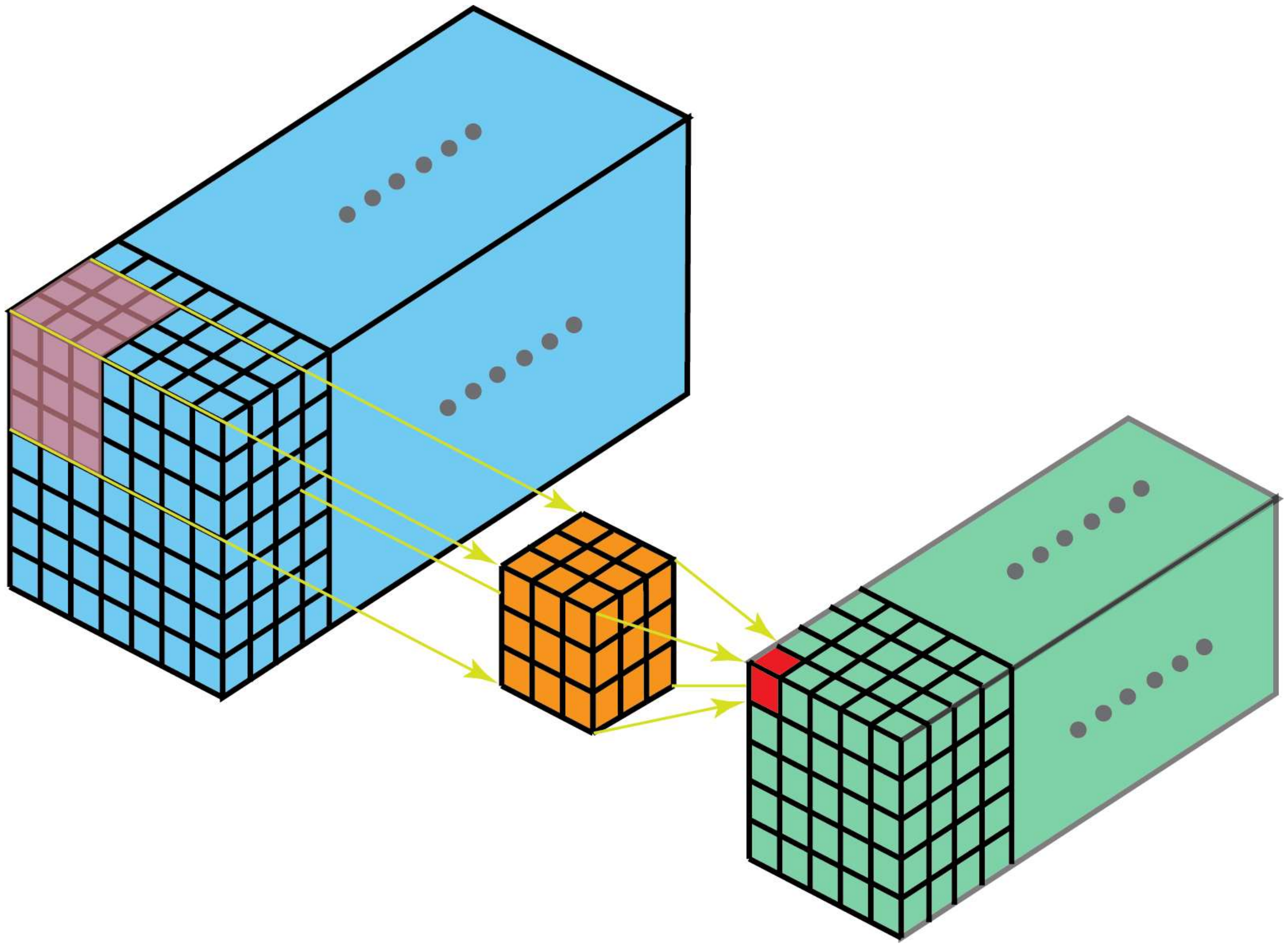
Convolution

1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	32	33	34	35
36	37	38	39	40	41	42
43	44	45	46	47	48	49

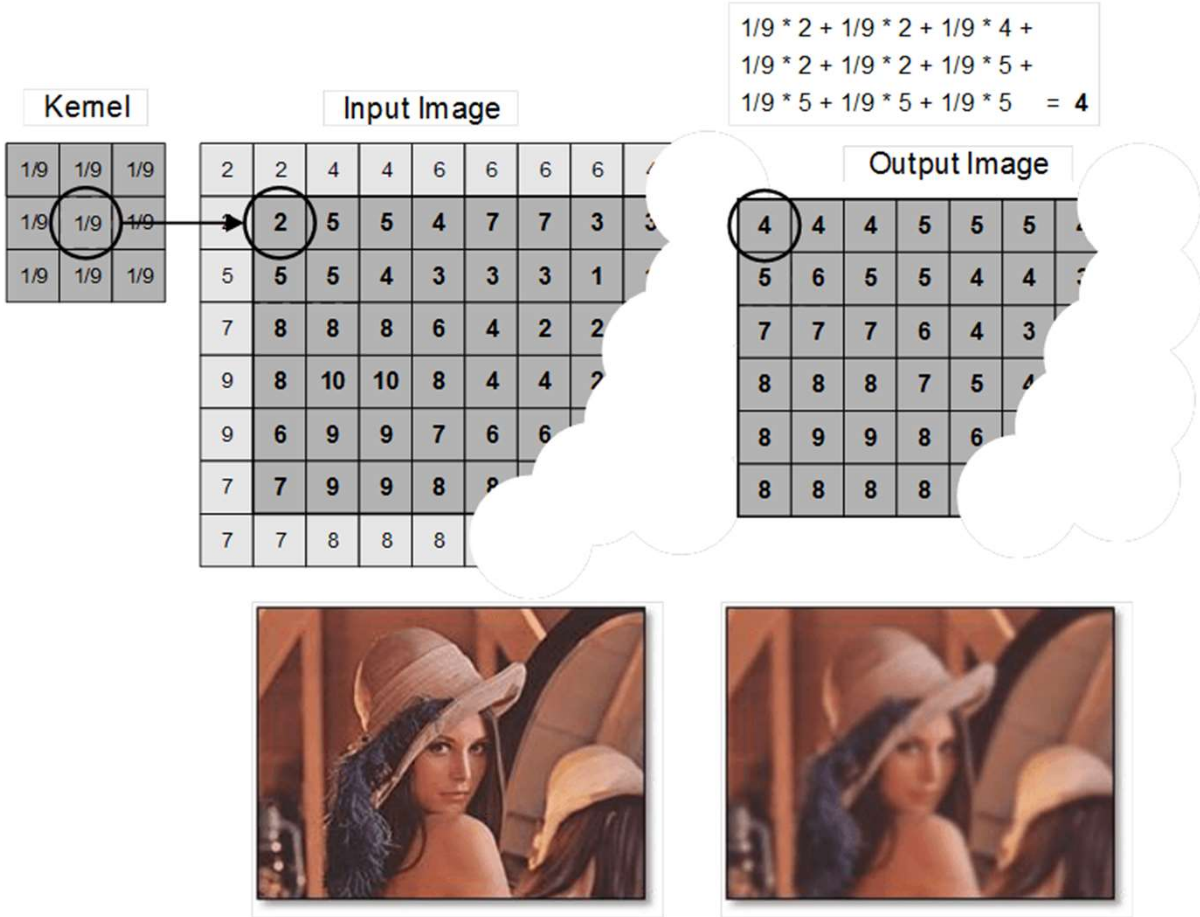
0.1	0.2	0.3
0.4	0.5	0.6
0.7	0.8	0.9

$$
\begin{aligned}
&= 0.1 \times 10 + 0.2 \times 11 + 0.3 \times 12 \\
&+ 0.4 \times 17 + 0.5 \times 18 + 0.6 \times 19 \\
&+ 0.7 \times 24 + 0.8 \times 25 + 0.9 \times 26 \\
&= 94.2
\end{aligned}
$$

3D Convolution



Convolution Edge Handling



Convolution Edge Handling

0	0	0	0	0	0
0	105	102	100	97	96
0	103	99	103	101	102
0	101	98	104	102	100
0	99	101	106	104	99
0	104	104	104	100	98

Image Matrix

Kernel Matrix

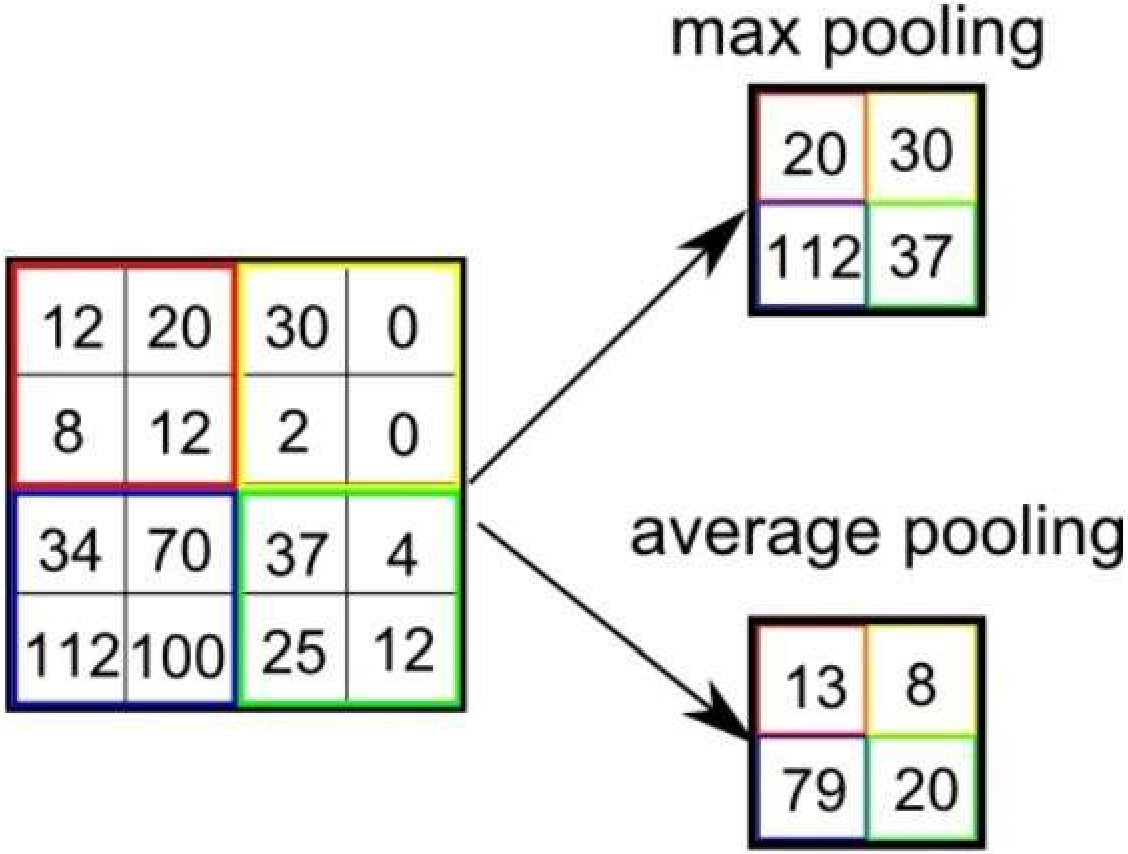
0	-1	0
-1	5	-1
0	-1	0

320				
210	89	111		

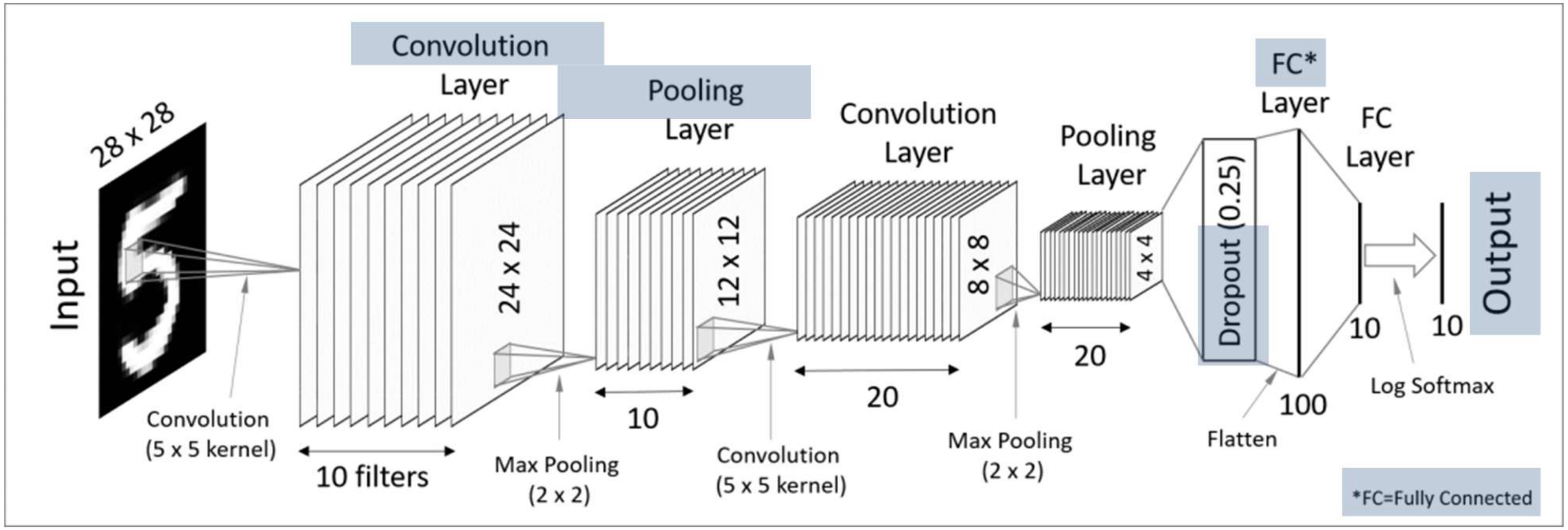
Output Matrix

$$0*0+0*-1+0*0$$
$$+0*-1+105*5+102*-1$$
$$+0*0+103*-1+99*0 \quad = 320$$

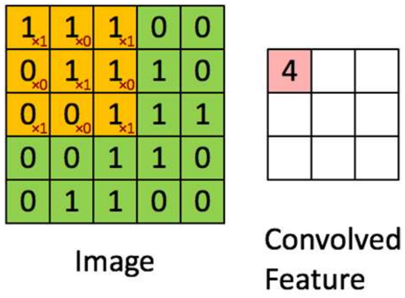
Pooling Operation



Convolutional Neural Networks: Basics

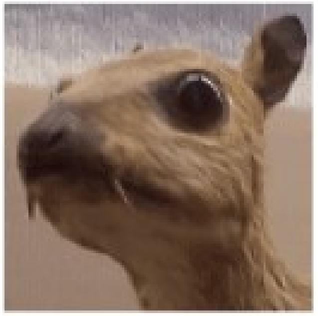


Convolution



Note image size reduction!

Input image



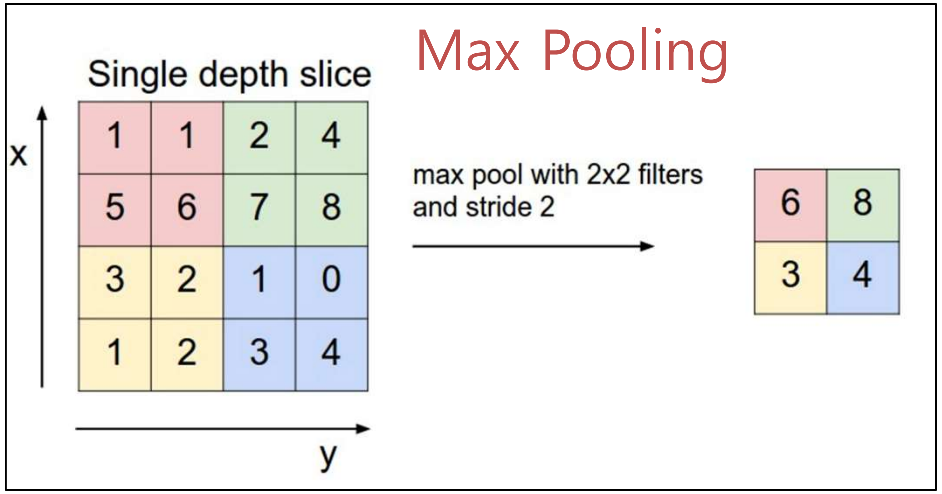
Convolution Kernel

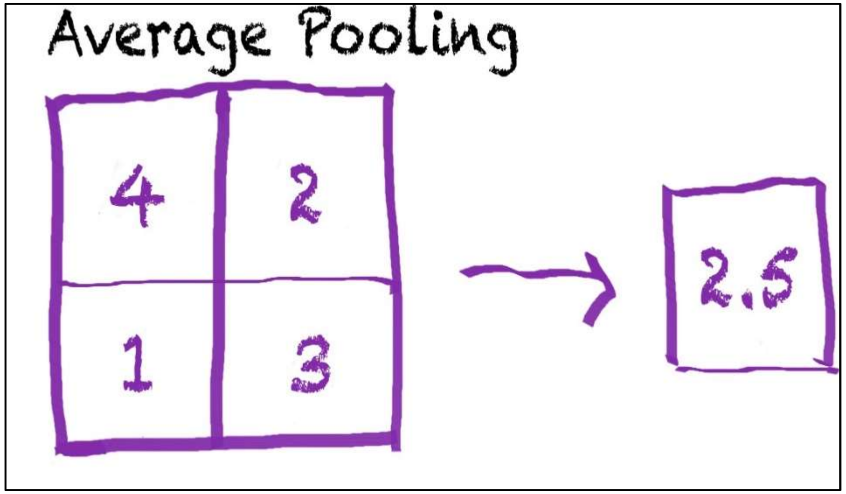
$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

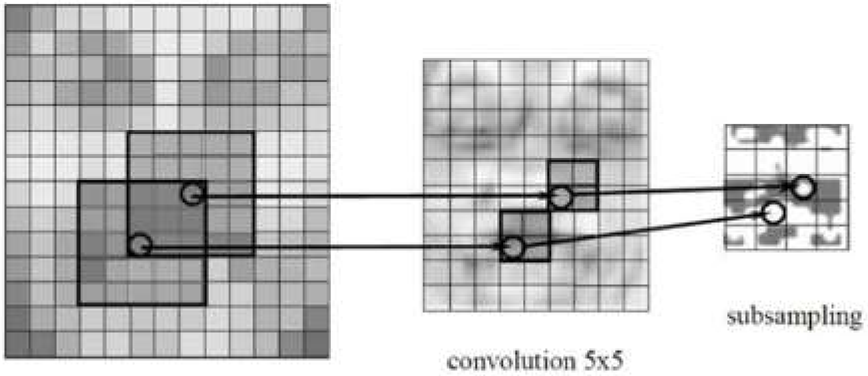
Feature map

Weights of Deep Network
(to be learned or estimated)

Pooling: Subsampling

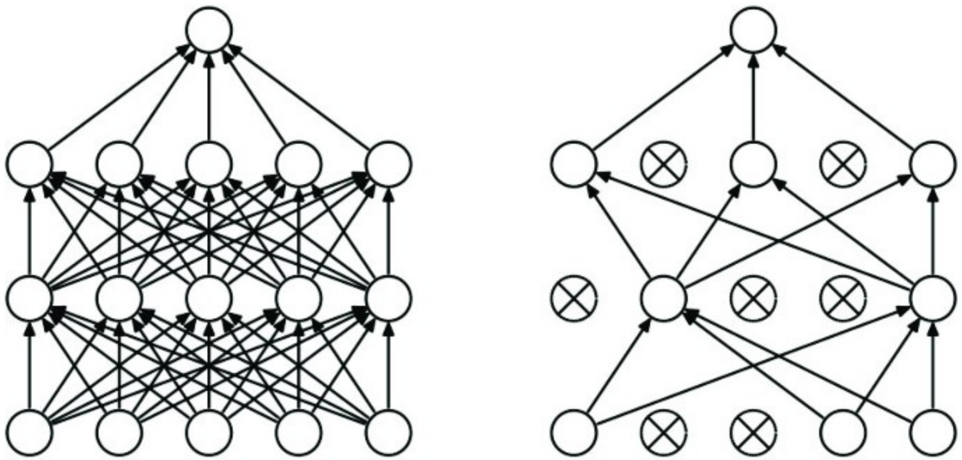






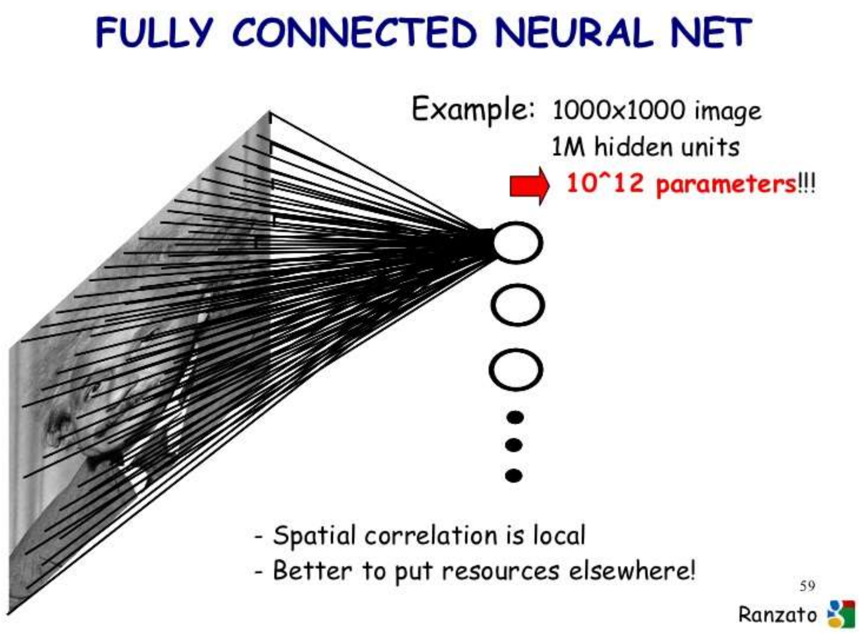
Dropout

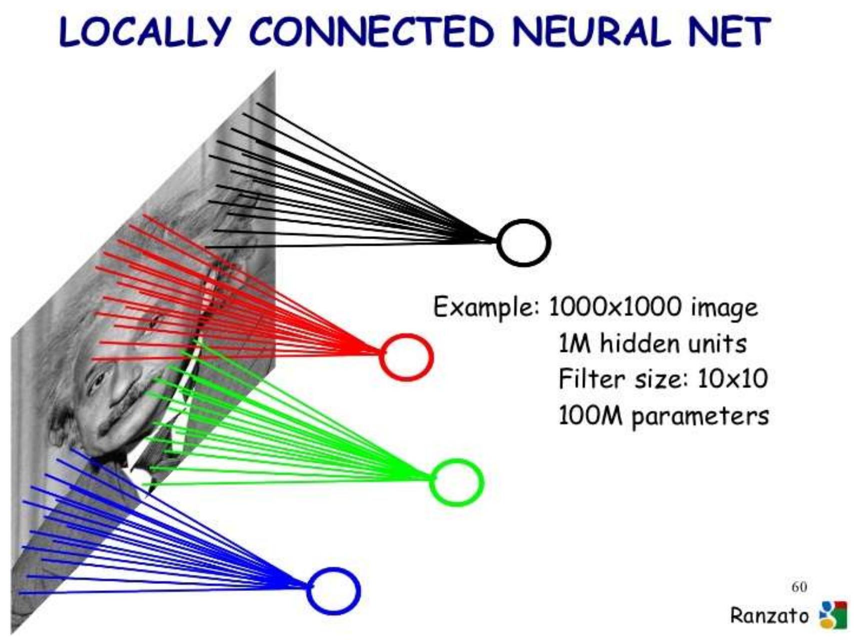
Dropout: A Simple Way to Prevent Neural Networks from Overfitting [Srivastava et al. 2014]



(a) Standard Neural Network (b) Neural Net with Dropout

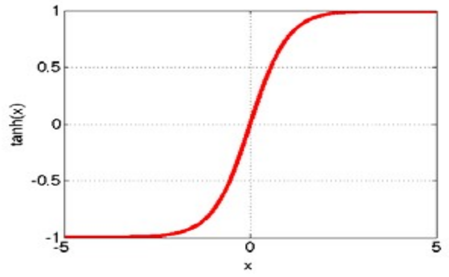
Full Connections (FC)

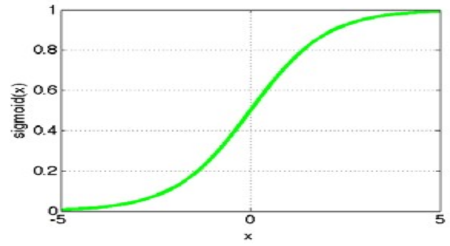


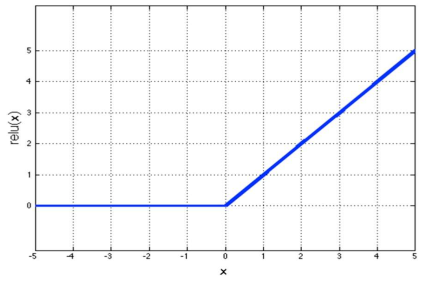


Non-Linearity (Activation Functions)

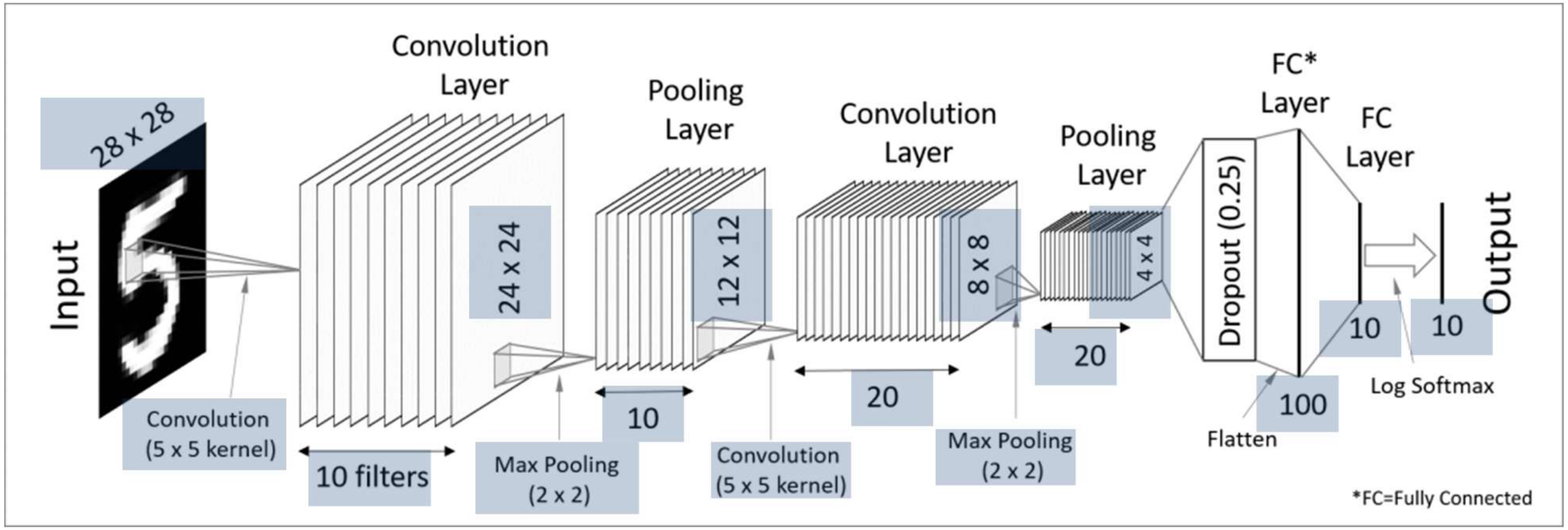
- Activation per neural element
- Options:
  - Tanh
  - Sigmoid: 1/(1+exp(-x))
  - **Rectified linear unit (ReLU)**
    - Simplifies backpropagation
    - Makes learning faster
    - Avoids saturation issues
    - Preferred option





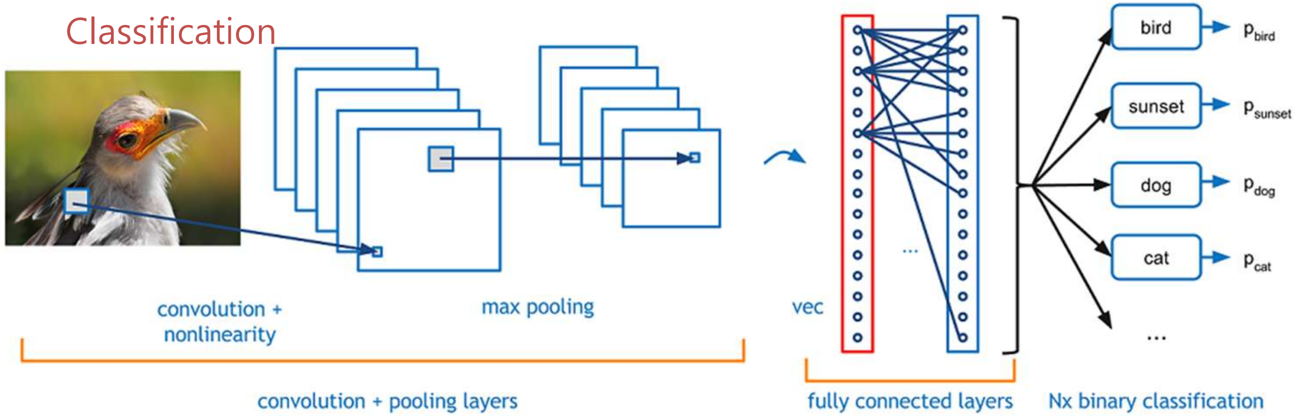


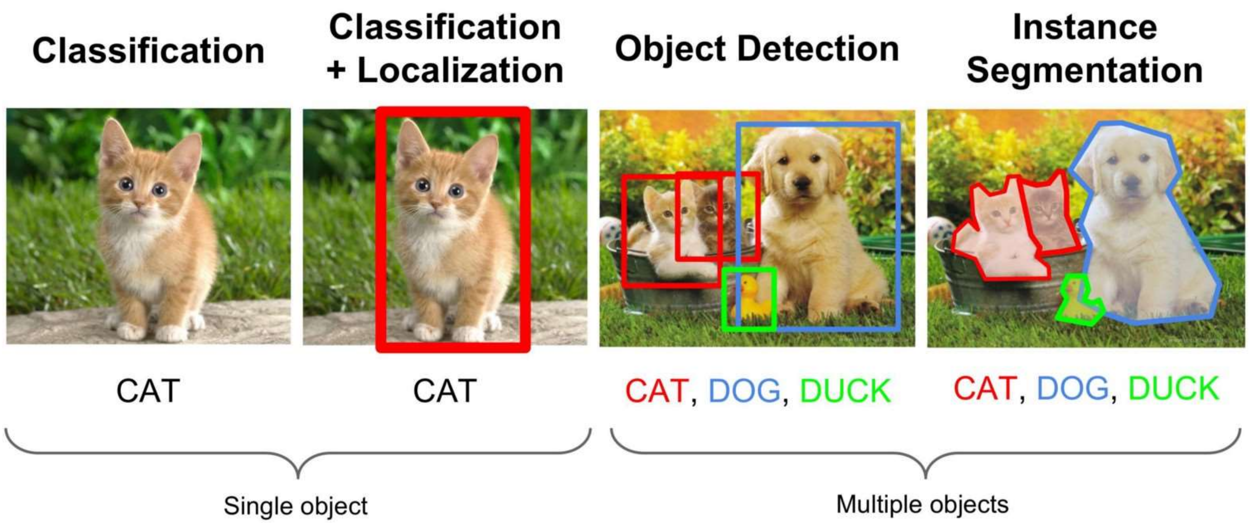
Convolutional Neural Networks: How to Read the Network



What CNN Does?

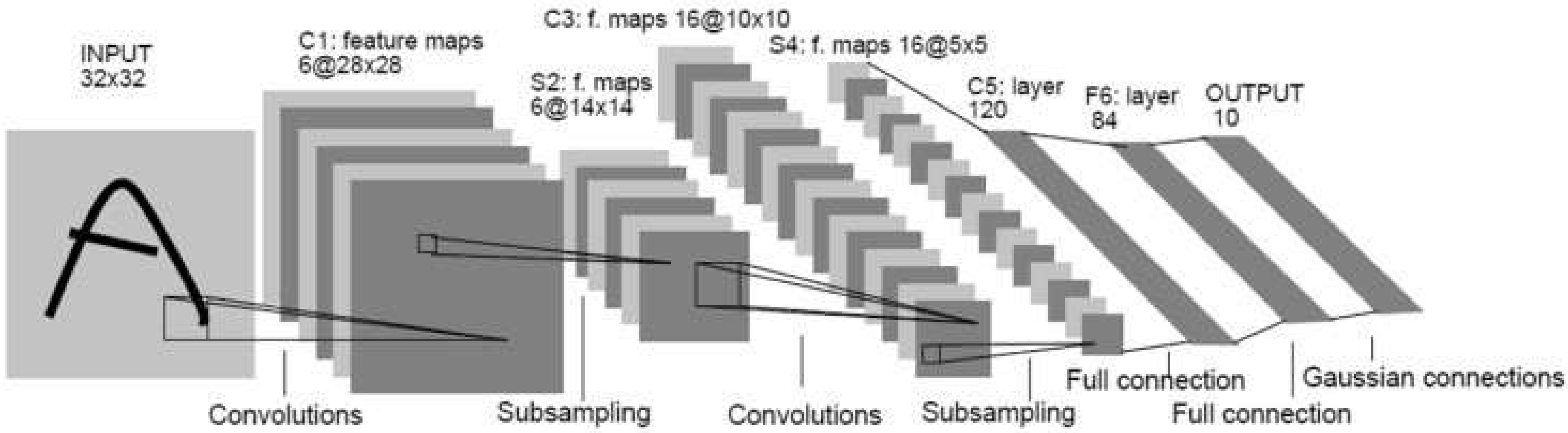
- Classification
- Detection
- Segmentation
- …





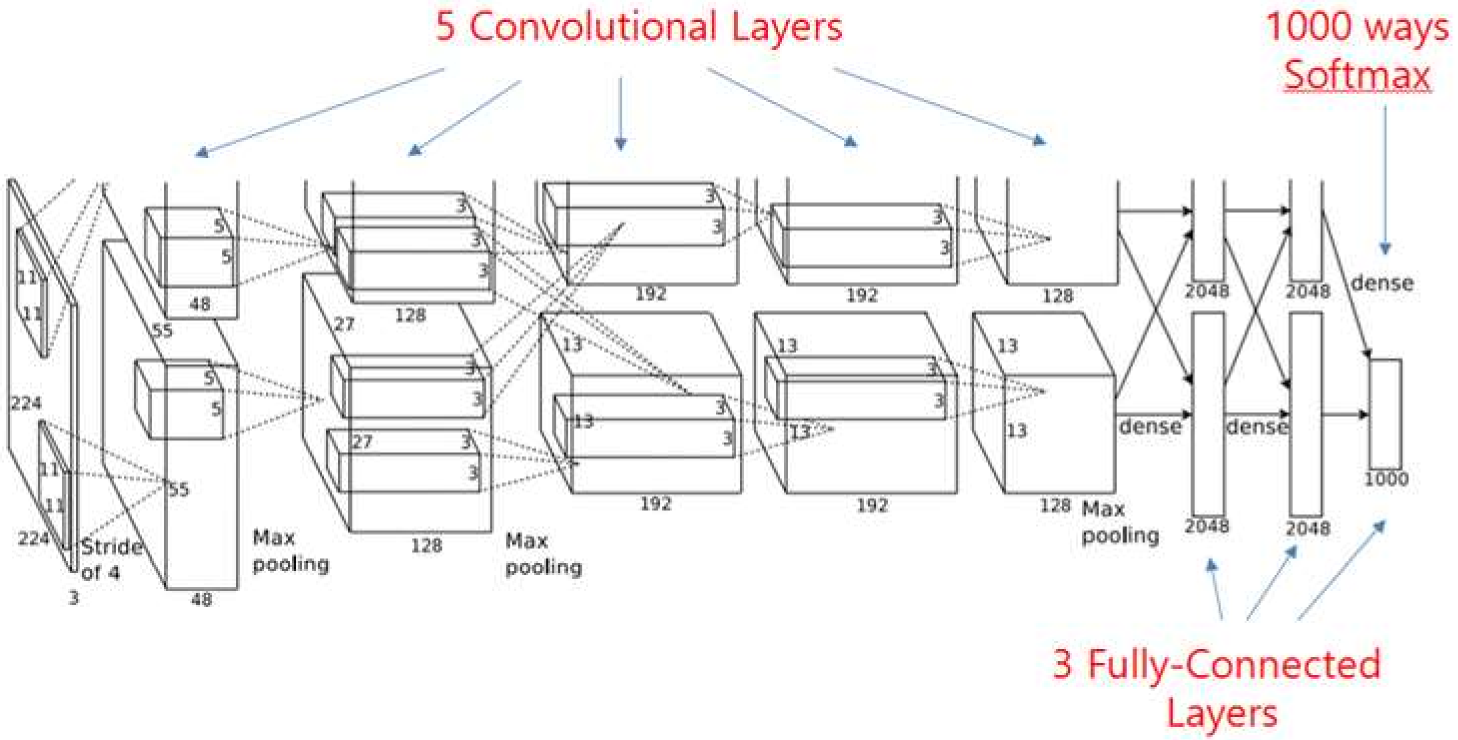
LeNet (1998)

- Developed by Yann LeCun
- First CNN Structure



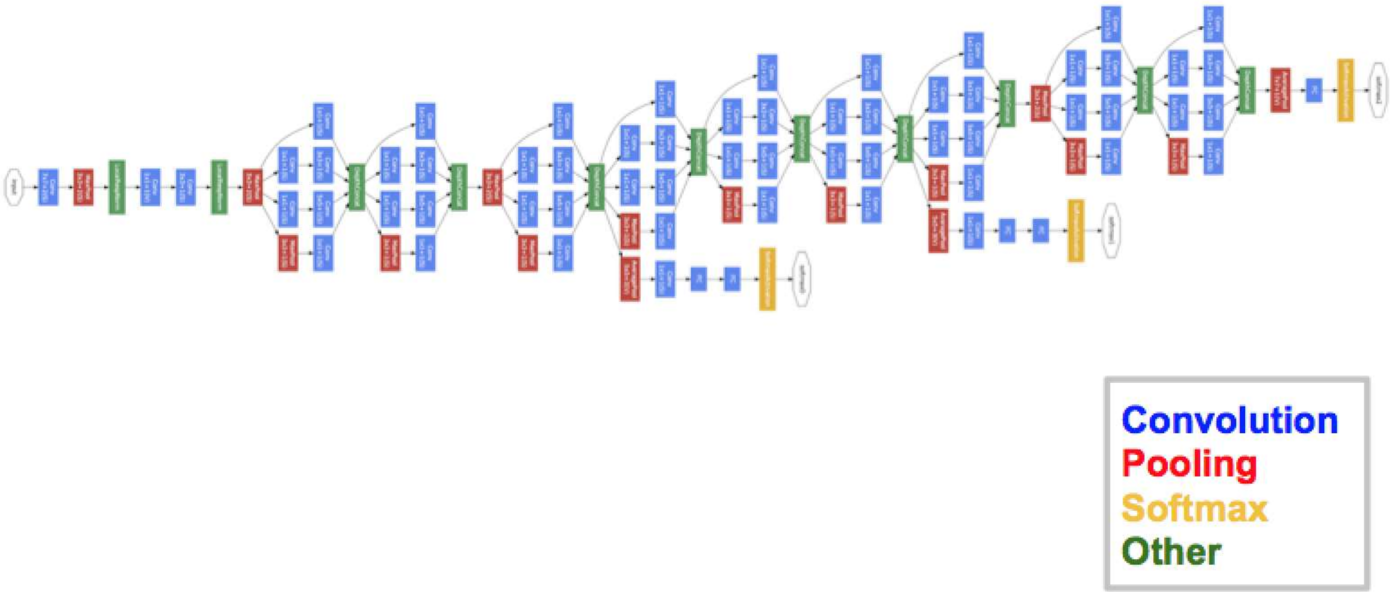
AlexNet (2012)

- Winner of ILSVRC 2012 Competition
- Recognition of ImageNet Database



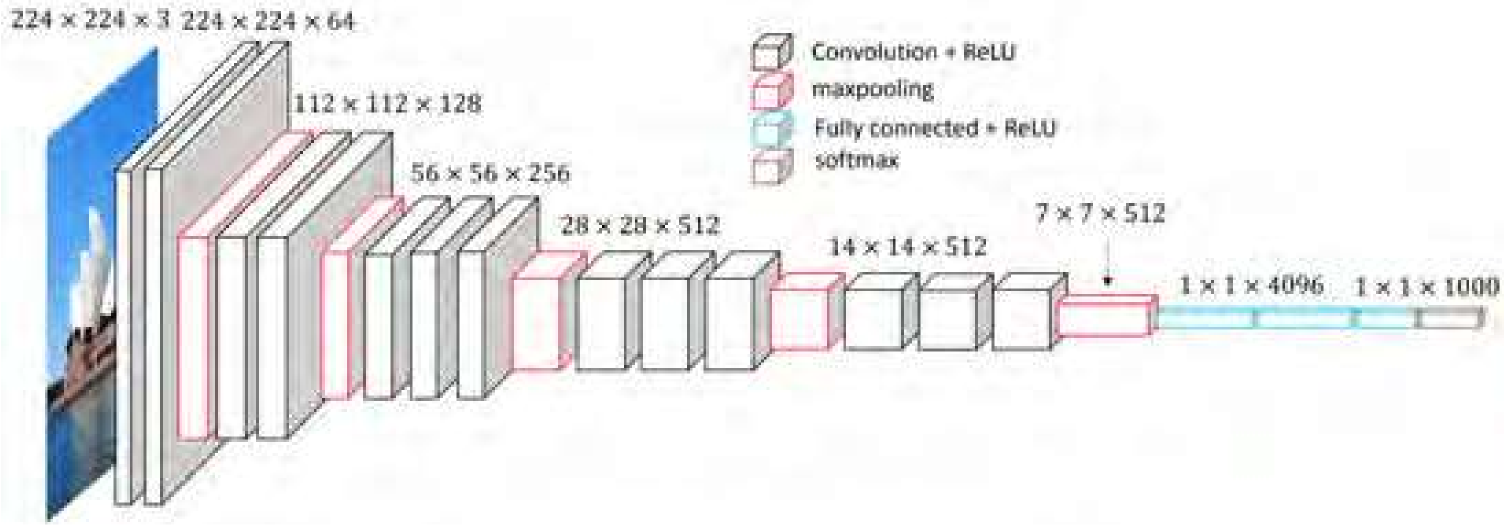
GoogLeNet (2014)

- Winner of ImageNet 2014



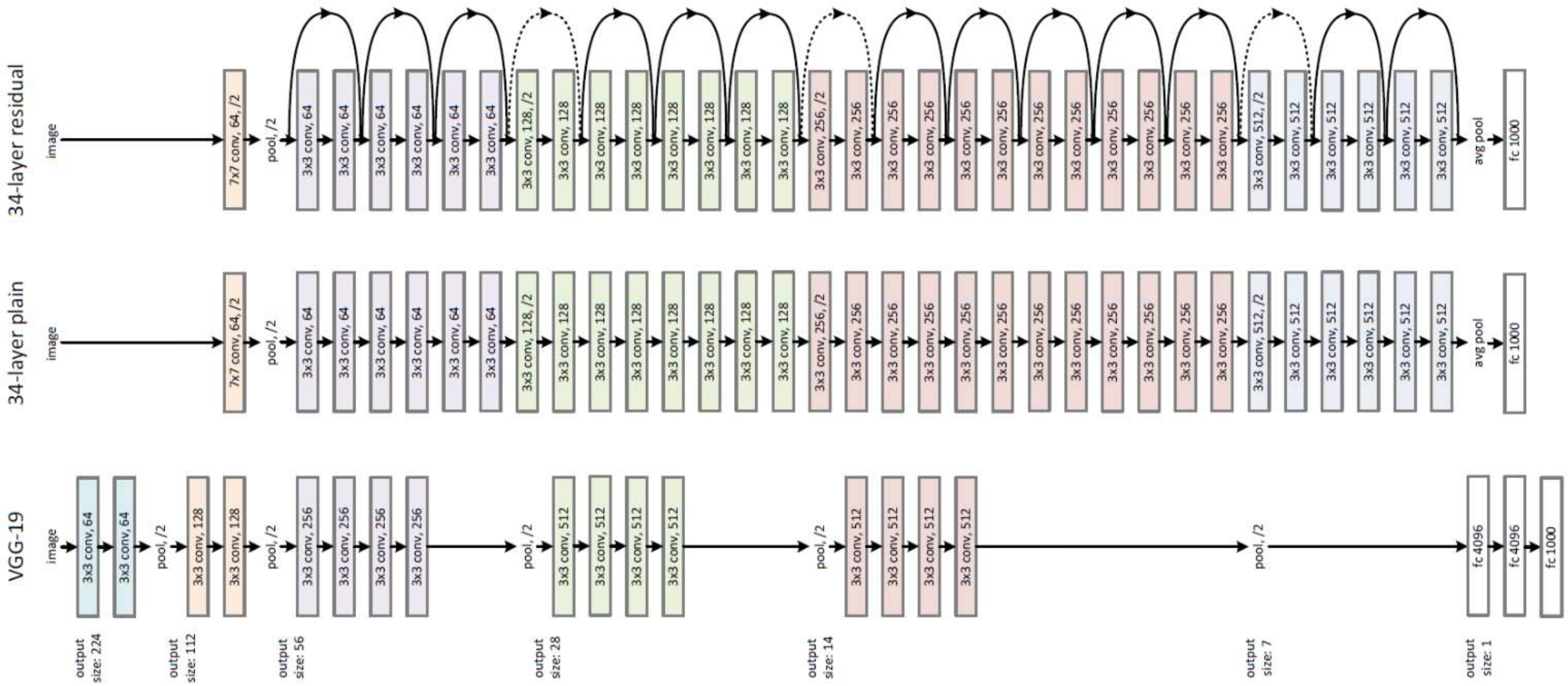
VGGNet (2014)

- 2^{nd} Winner of ImageNet 2014



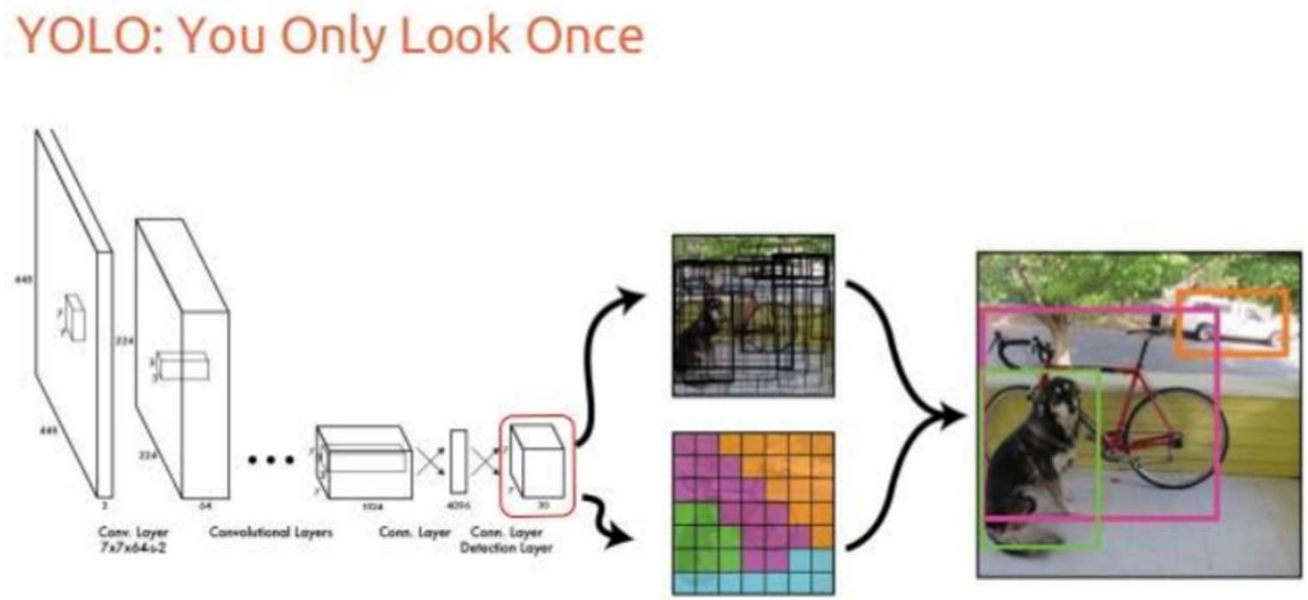
ResNet-50 (2015)

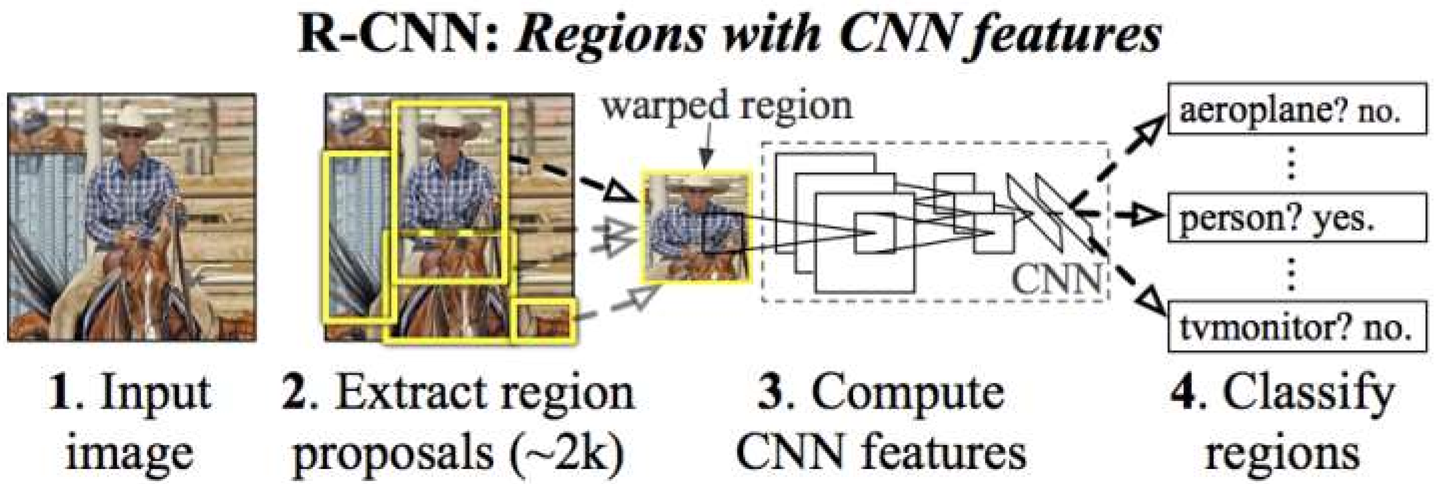
- Winner ILSVRC 2015



34-layer ResNet with Skip / Shortcut Connection (Top), 34-layer Plain Network (Middle), 19-layer VGG-19 (Bottom)

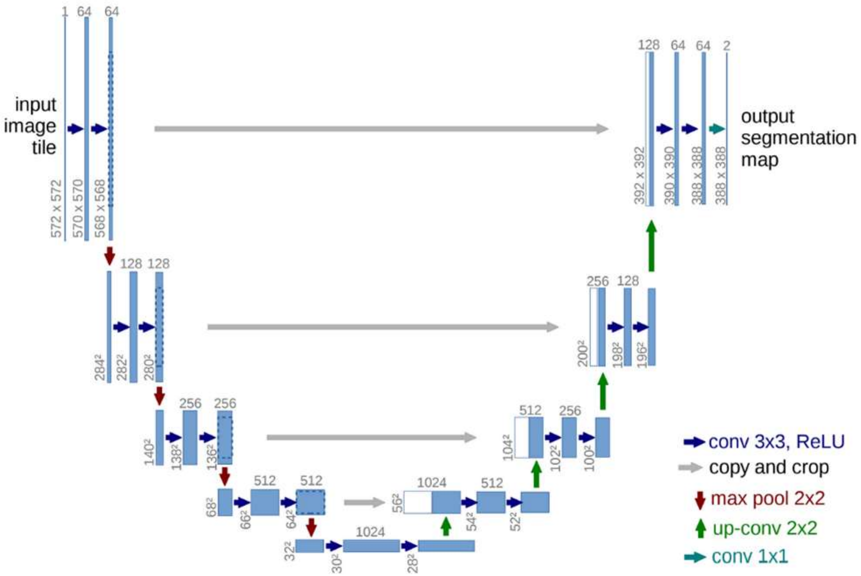
YOLO vs. Fast R-CNN

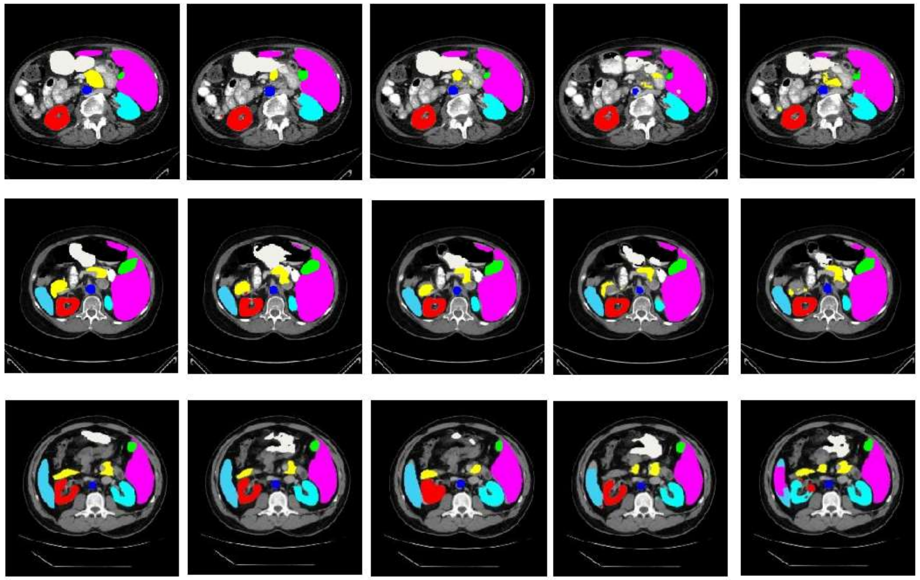




Segmentation

- U-Net





Ground Truth | Dilated-Unet | Swin-Unet | HiFormer | Unet