Convolutional Neural Network

Group 16:

1. Chintan Panchamia
2. Akshat Shah
3. Kevin Desai
4. Kunal Bhandari
5. Prutha Khandeparker
6. Karthik Palaniappan

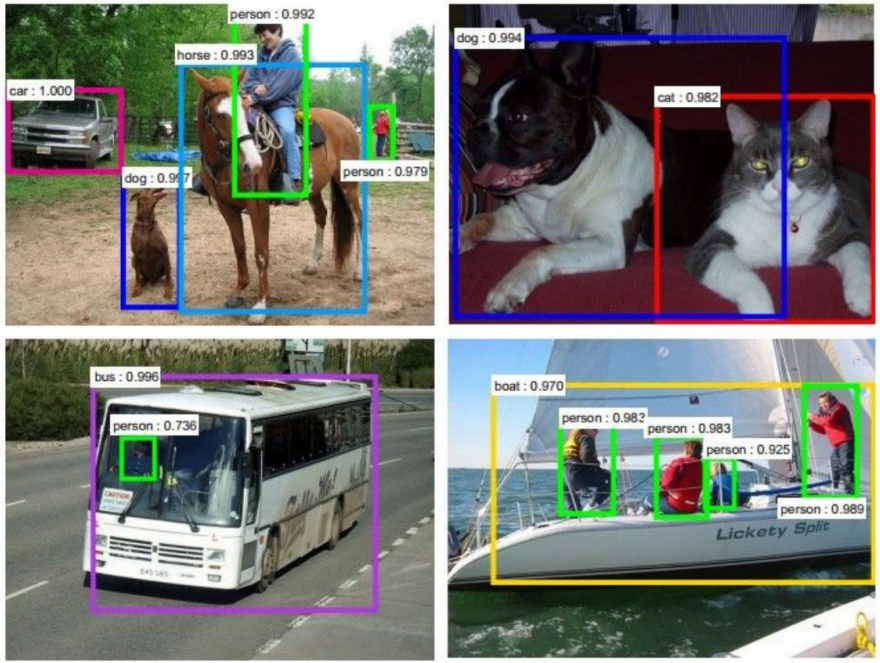
What is CNN?

Classification

Retrieval

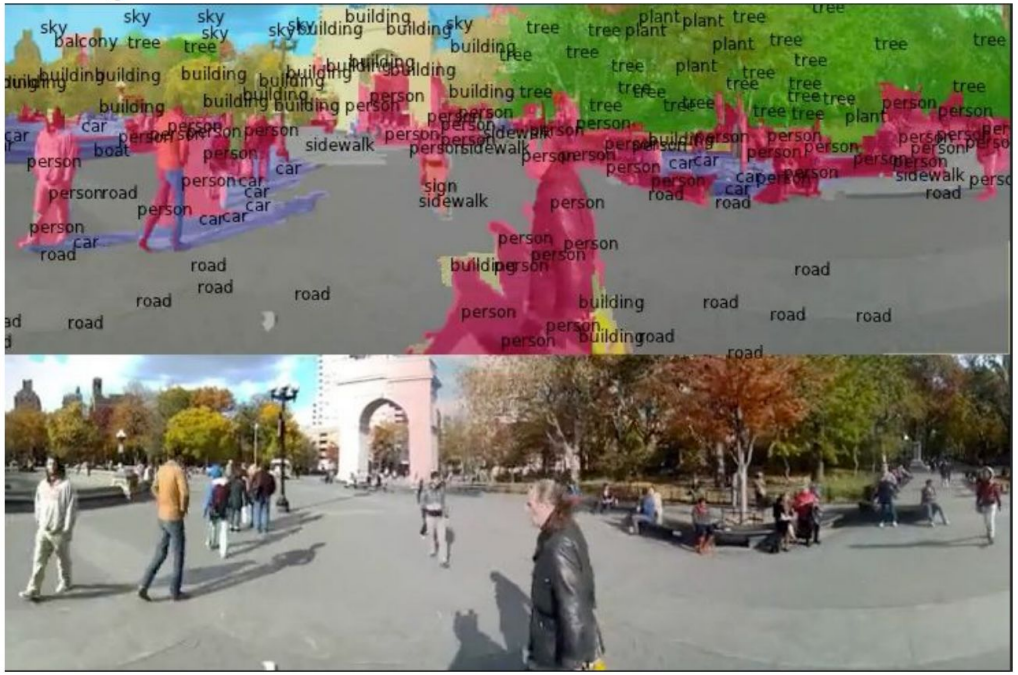
[Krizhevsky 2012]

Detection



[Faster R-CNN: Ren, He, Girshick, Sun 2015]

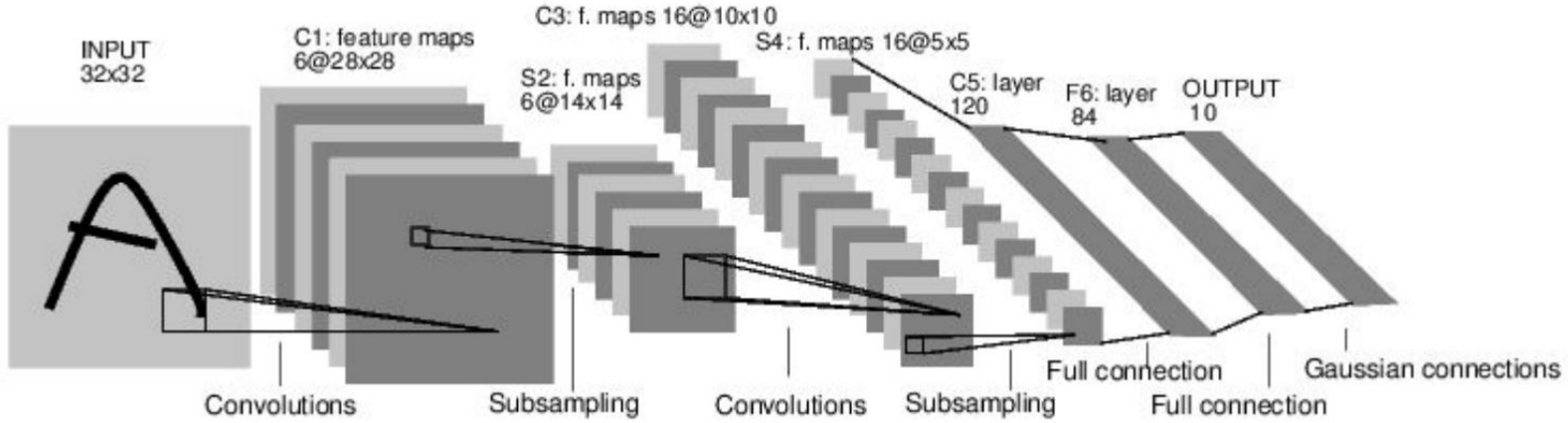
Segmentation



[Farabet et al., 2012]

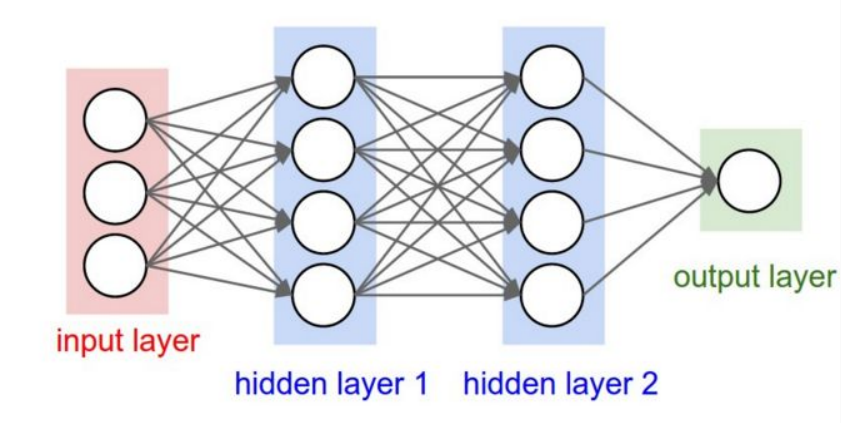

[Taigman et al. 2014]
Calista_Flockhart_0002.jpg
Detection & Localization
Frontalization:
@152X152x3
C1:
32x11x11x3
@142x142
M2:
32x3x3x32
@71x71
C3:
16x9x9x32
@63x63
L4:
16x9x9x16
@55x55
L5:
16x7x7x16
@25x25
L6:
16x5x5x16
@21X21
F7:
4096d
REPRESENTATION
F8:
4030d
SFC labels

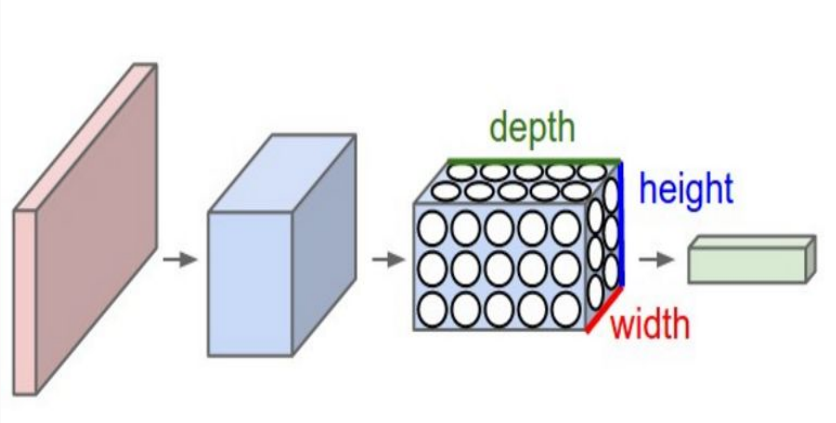
What shall you learn?



[LeNet-5, LeCun 1980]

Regular NN v/s Convolutional NN





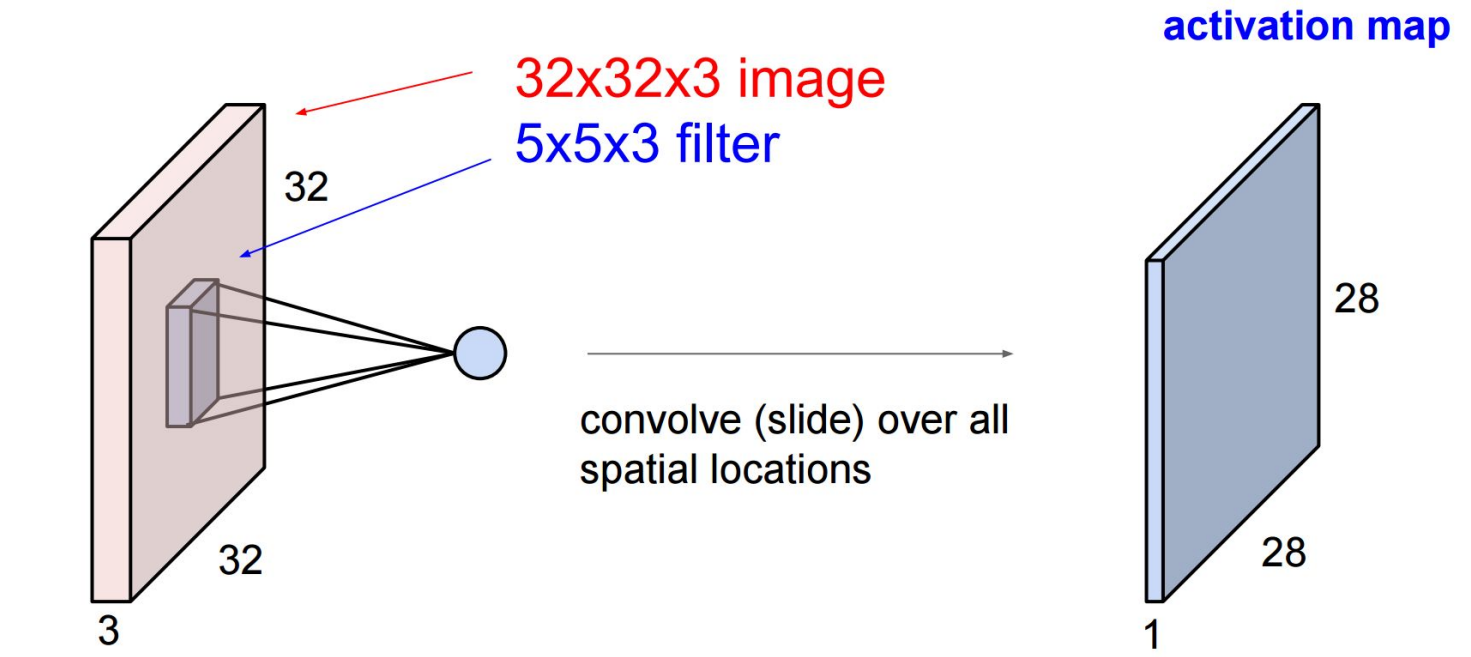
Regular Neural Nets don't work well with full images due to the high dimensions of images.

Reduces the amount of parameters in the network.

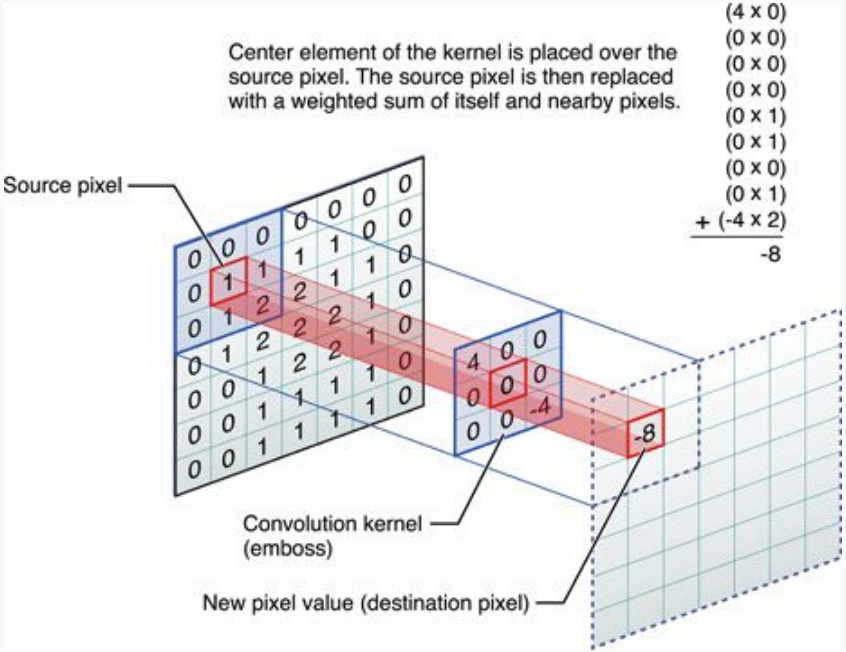
Layers of CNN

1. Input - contains the pixels values of the image.
2. CONV - computes the dot product between the weights and the regions they are connected to, for the neurons that are connected to local regions.
3. RELU - applies an elementwise activation function.
4. POOL - performs downsampling along the spatial dimensions of the image (width, height)
5. Fully Connected (FC) - connects all the neurons and gives a class score.

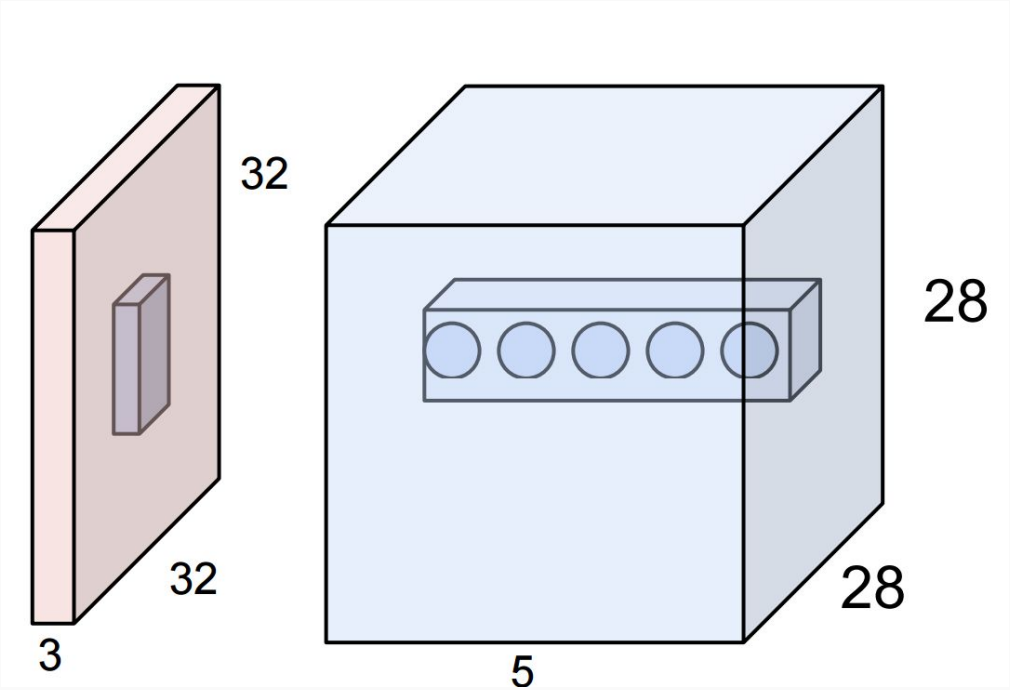
Convolutional Layer



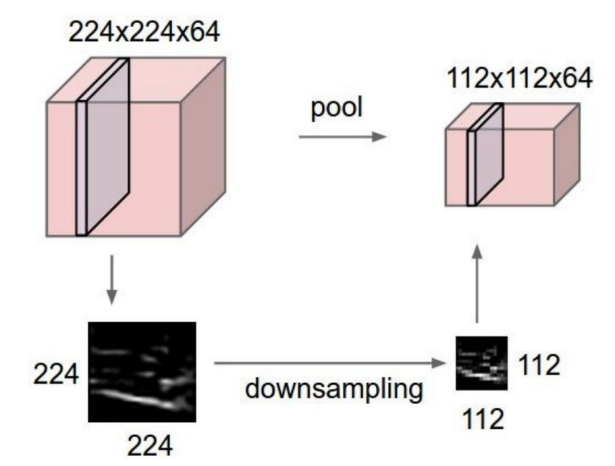
Convolutional Layer

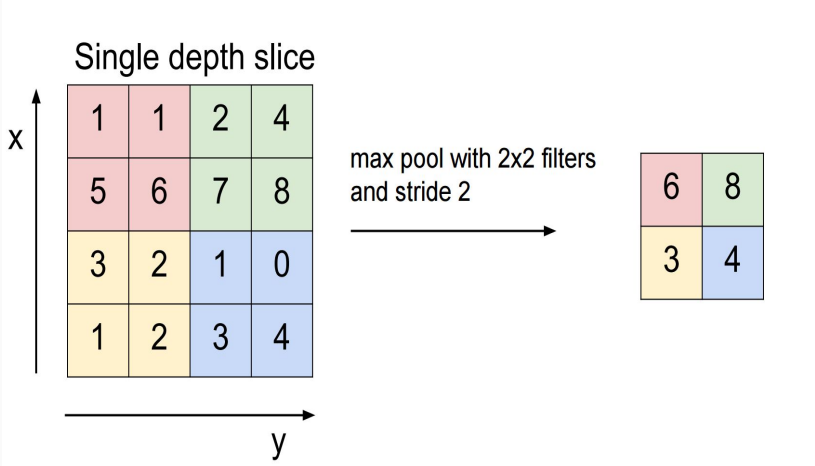


Convolutional Layer

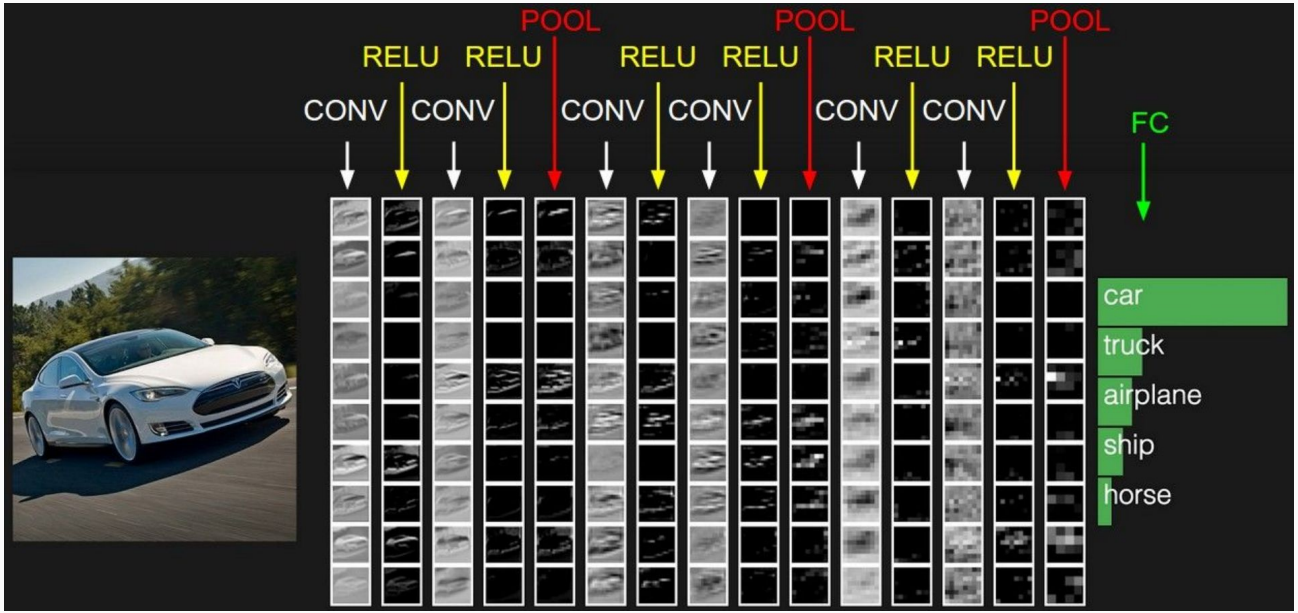


Pooling

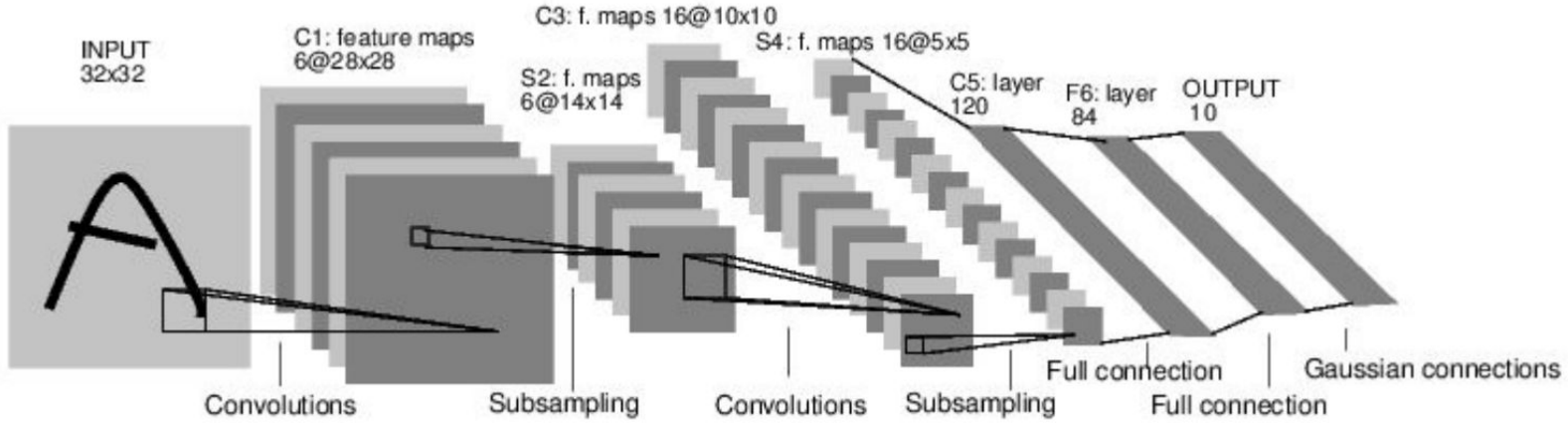

224x224x64
pool
112x112x64
224
224
downsampling
112
112


Single depth slice
x
y
1 1 2 4
5 6 7 8
3 2 1 0
1 2 3 4
max pool with 2x2 filters
and stride 2
6 8
3 4

Putting everything together

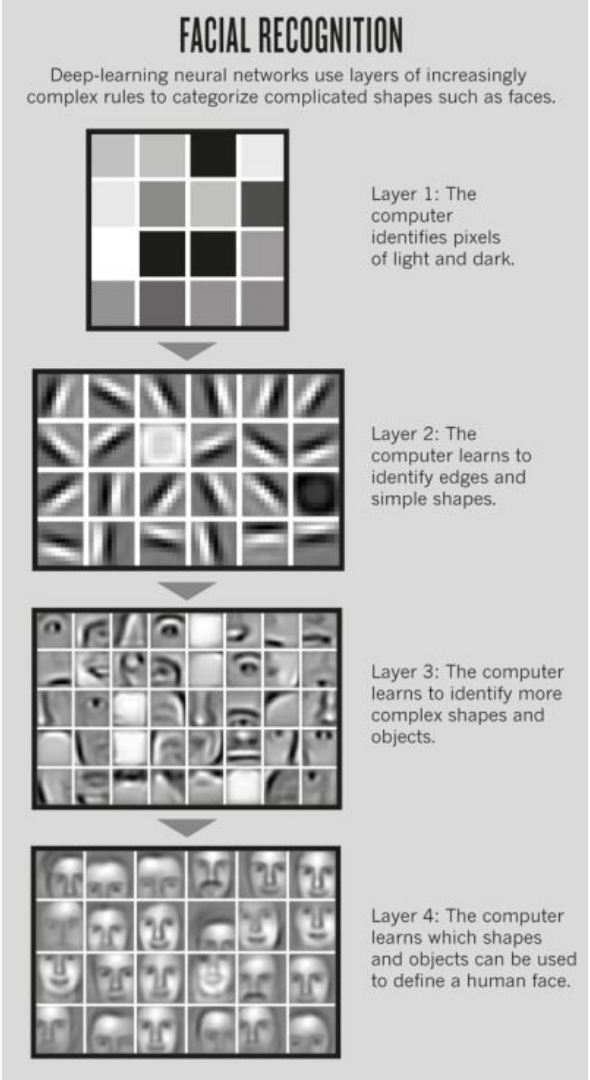


What you learned!



[LeNet-5, LeCun 1980]

Example


FACIAL RECOGNITION
Deep-learning neural networks use layers of increasingly complex rules to categorize complicated shapes such as faces.
Layer 1: The computer identifies pixels of light and dark.
Layer 2: The computer learns to identify edges and simple shapes.
Layer 3: The computer learns to identify more complex shapes and objects.
Layer 4: The computer learns which shapes and objects can be used to define a human face.

References

- Fei-Fei Li, Andrej Karpathy and Justin Johnson - http://cs231n.github.io/convolutional-networks/
- http://deeplearning.net/tutorial/lenet.html
- http://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf

Thanks!

Describes without errors	Describes with minor errors	Somewhat related to the image
		
A person riding a motorcycle on a dirt road.	Two dogs play in the grass.	A skateboarder does a trick on a ramp.
		
A group of young people playing a game of frisbee.	Two hockey players are fighting over the puck.	A little girl in a pink hat is blowing bubbles.
		
A herd of elephants walking across a dry grass field.	A close up of a cat laying on a couch.	A red motorcycle parked on the side of the road.