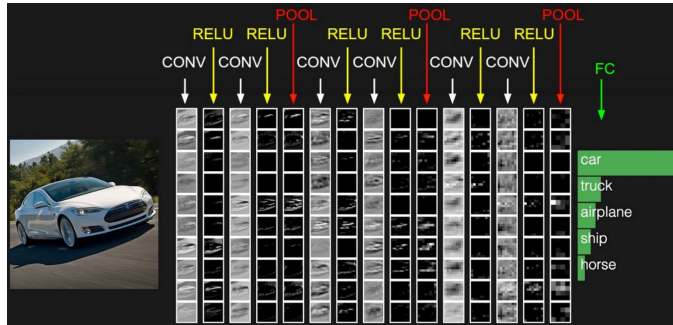


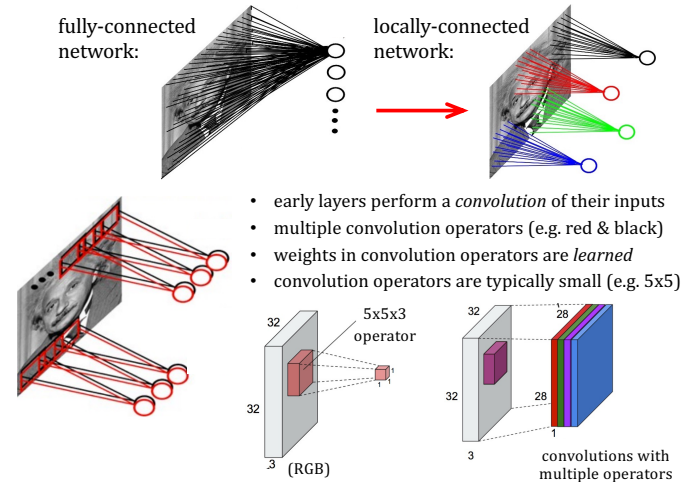
Sample stages of a CNN



CONV: "convolution" layer with weights that are learned
 RELU: "rectified linear unit" applies an activation function
 POOL: "pooling" selects maximum value in small neighborhoods
 FC: "fully-connected" neural network

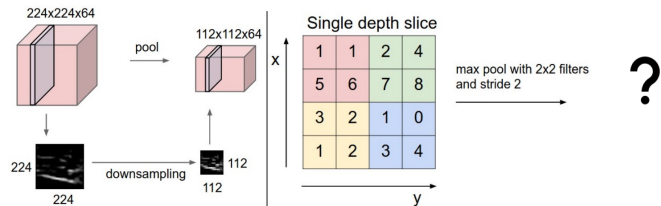
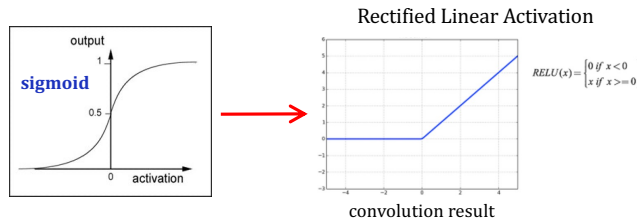
1

Convolutional layer



2

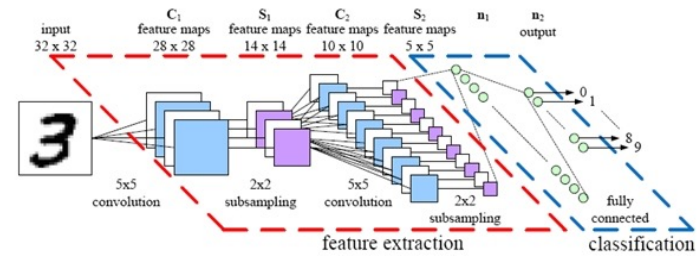
ReLU & max pooling layers



3

Adding a fully-connected neural net layer

Recognizing digits from the MNIST database with a CNN:

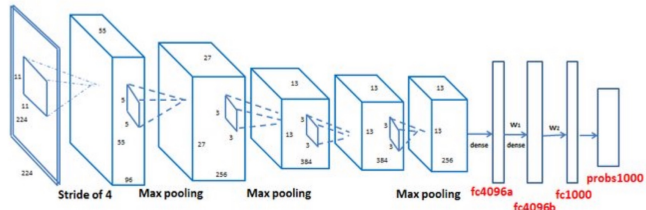


LeNet

LeCun, Bottou, Bengio, Haffner (1998)

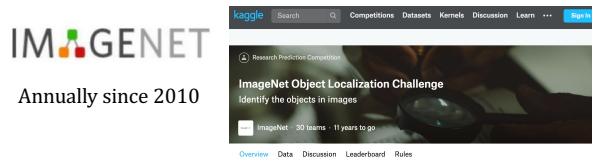
4

AlexNet, ZF Net, GoogLeNet, VGGNet, ResNet, ...

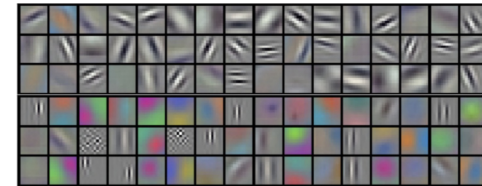


AlexNet: Krizhevsky, Sutskever, Hinton (2012)

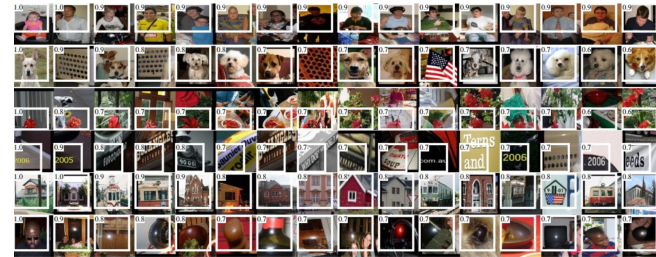
ImageNet Large Scale Visual Recognition Challenge (ILSVRC)



5



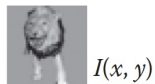
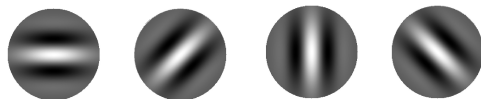
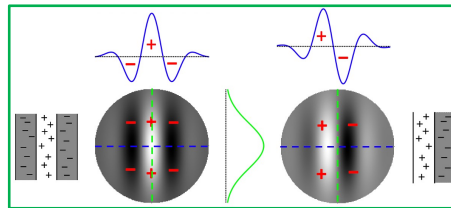
Filters generated in first convolution layer of AlexNet



Maximally activating images from some POOL5 neurons of AlexNet (Girshick et al., 2014)

6

2D Gabor functions



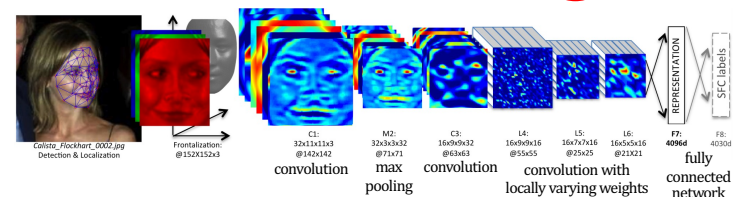
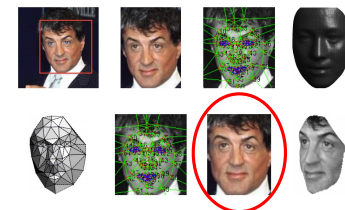
convolution operators learned in CNNs

7

Facebook's DeepFace system

Taijman et al., 2014

- detect face
- 2D align face in crop window using 6 fiducial points
- align to 3D shape model using 67 fiducial points
- use 3D model + image to generate frontal view

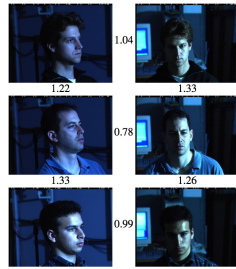
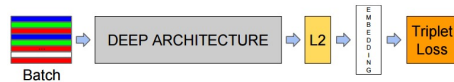


8

Google's FaceNet system

Schroff et al., 2015

FaceNet also uses a deep convolutional network

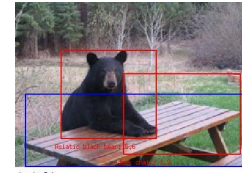


- learns mapping from images to a space where distance between images captures similarity
- training data: triplets of face thumbnails
 - two same ID, one different ID
- learning process: minimize distance between anchor & positive images (same ID), maximize distance between anchor & negative images



threshold = 1.1 classifies pairs correctly
(smaller value means more similar)

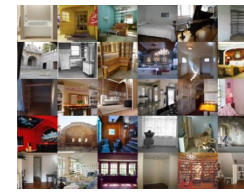
9



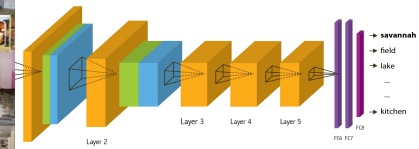
scene understanding

places.csail.mit.edu

205 scene categories
2.5 millions of images



Initialization: Trained AlexNet

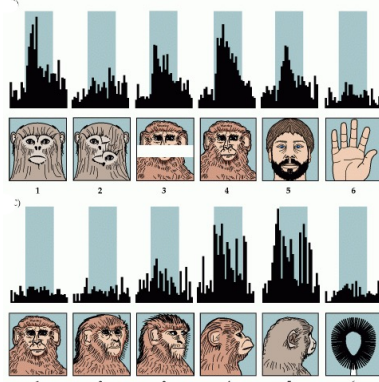


places images

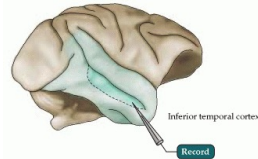
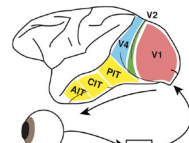
transfer learning

10

Face selective cells in IT cortex



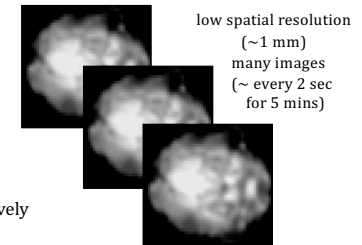
Desimone et al., 1984



Face selective cells in IT, from single cell recordings

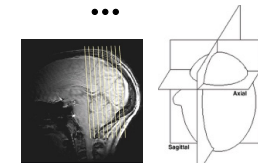
11

functional Magnetic Resonance Imaging (fMRI)

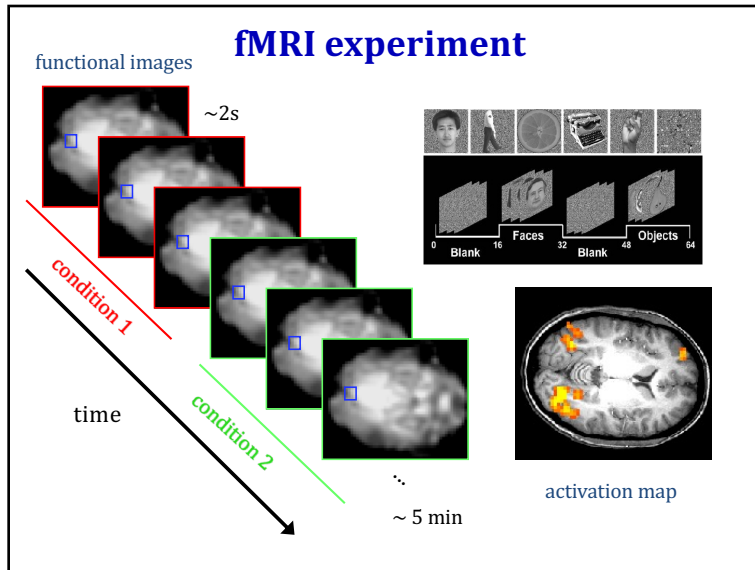


low spatial resolution
(~1 mm)
many images
(~ every 2 sec
for 5 mins)

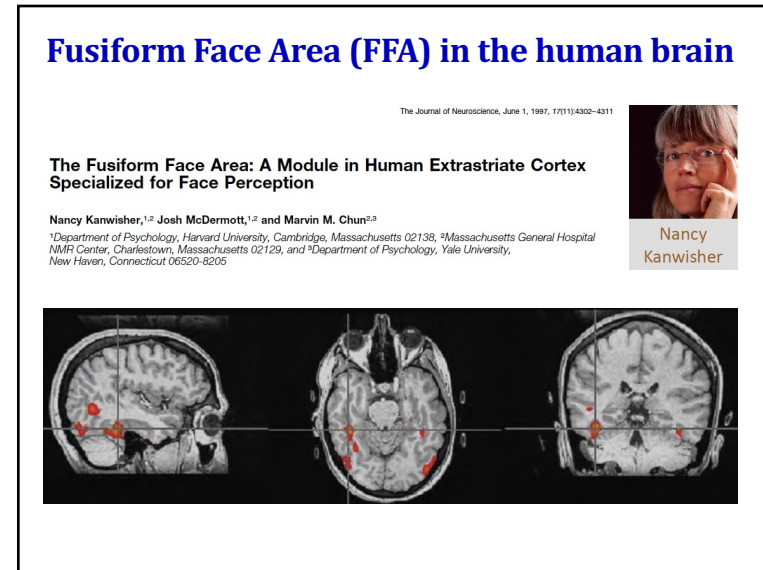
- best spatial resolution available for measuring neural activity noninvasively in the whole human brain
- increased neural activity
 - increased local blood flow
 - change in oxygenation of hemoglobin
 - increase in MRI signal
- Blood Oxygenation Level Dependent (BOLD) signal is an indirect measure of neural activity
- raw data: ~30,000 3D "voxels"
(each voxel: hundreds of thousands of neurons)



12



13



14