Sanjay Talbar
Professor,
Electronics & Telecommunication Engg.,
SGGS Institute of Engineering & Technology, Nanded

Convolutional Neural Networks and Deep Learning

Sanjay Talbar
Professor, Department of E&TC,
SGGSIET, Nanded

Acknowledgements

- Researcher in the area of Machine Learning, Deep Learning and Computer Vision who have made their resources publicly available on the internet.
- Online coarse “Machine Learning” and “Deep Learning Specialization” by Andrew Ng on Coursera.
- “CS231n: Convolutional Neural Networks for Visual Recognition” by Fei Fei Li and Andrej Karpathy.
- ACM’s “Turing Award” 2018
  - Prof Geoffrey Hinton, (English-Canadian) Toronto University, Canada & Google fellow
  - Prof Yoshua Bengio (Canadian Computer Scientist) University of Montreal, Canada
  - Prof Yann LeCunn, (French-American scientist) NYU, USA
- Many others who may not be explicitly cited.

Outline

- What is Machine Learning?
- ML Applications
- Traditional ML framework and flow
- Introduction to Deep Learning
- Brief review of Neural Network
- Basics of Convolutional Neural Network
- Popular CNN architectures
- Programming Frameworks and GPUs
- Discussion

Machine Learning is…

Machine learning is the science of getting computers to learn without being explicitly programmed.

-- Andrew Ng

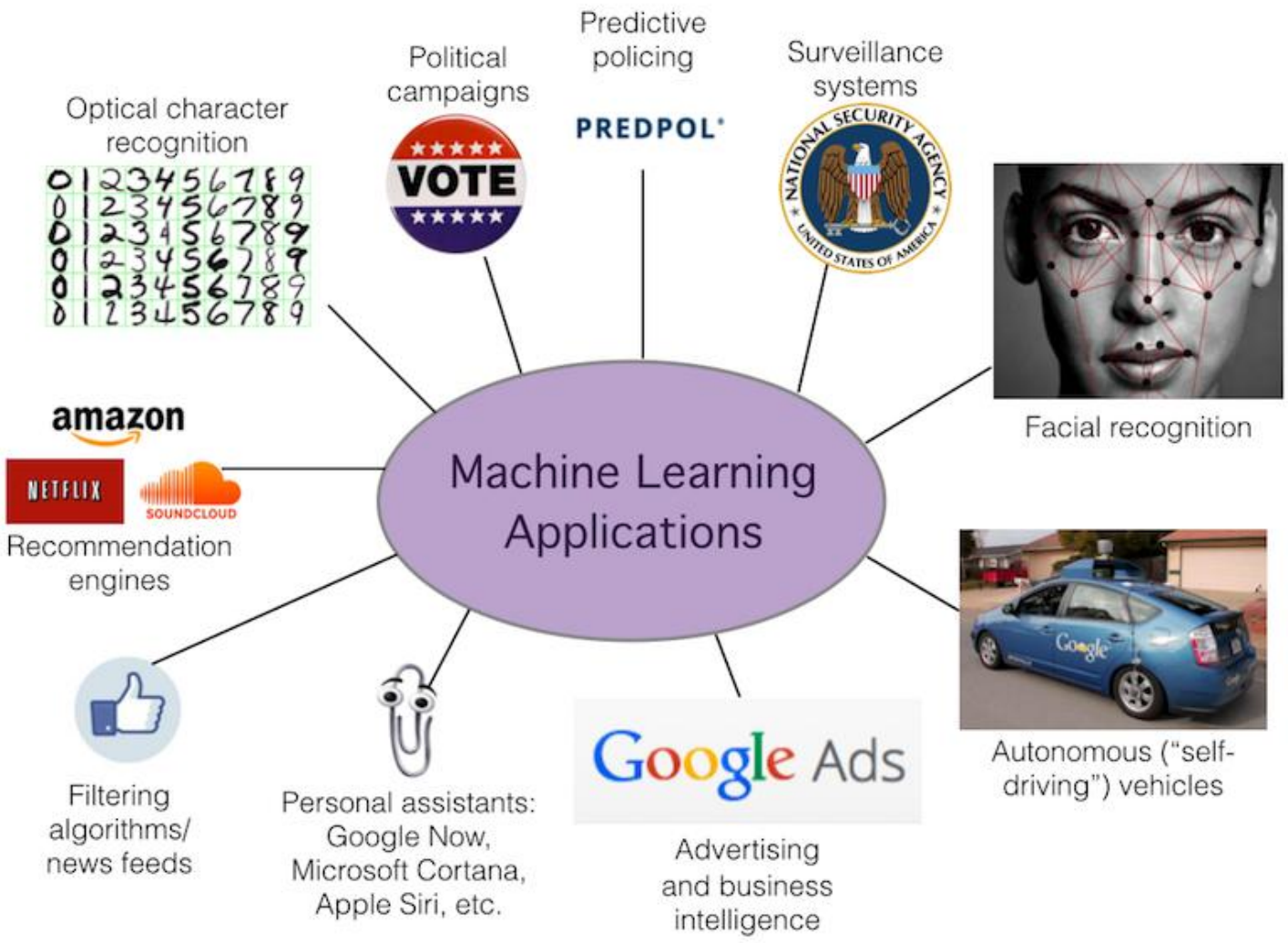
The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest.

-- Kevin P. Murphy

The field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions.

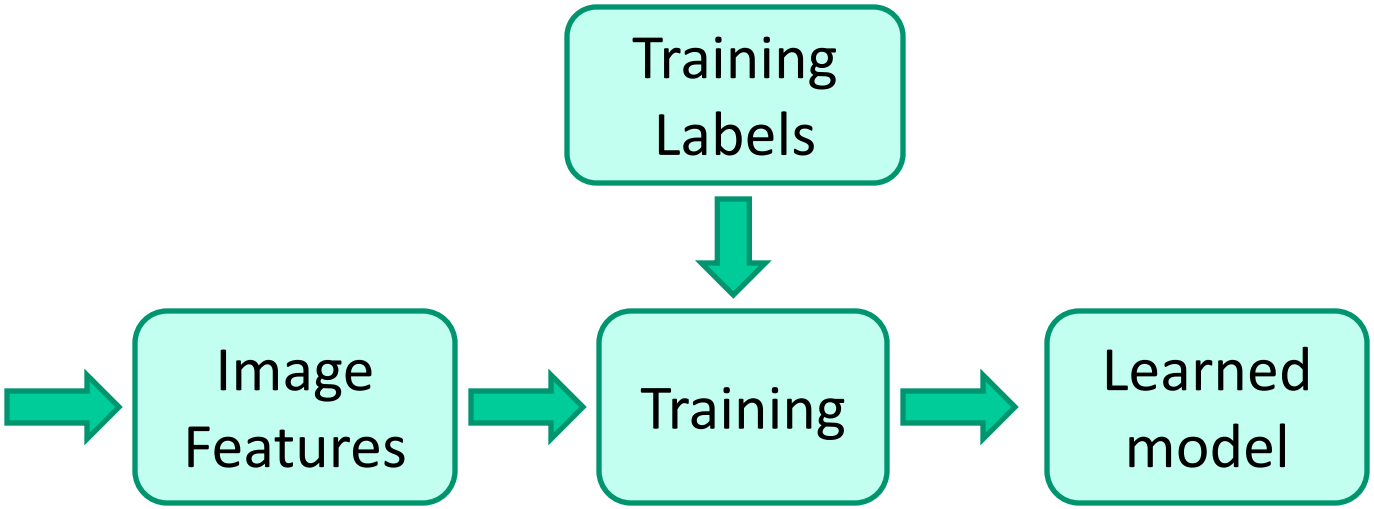
-- Christopher M. Bishop

Machine Learning Applications

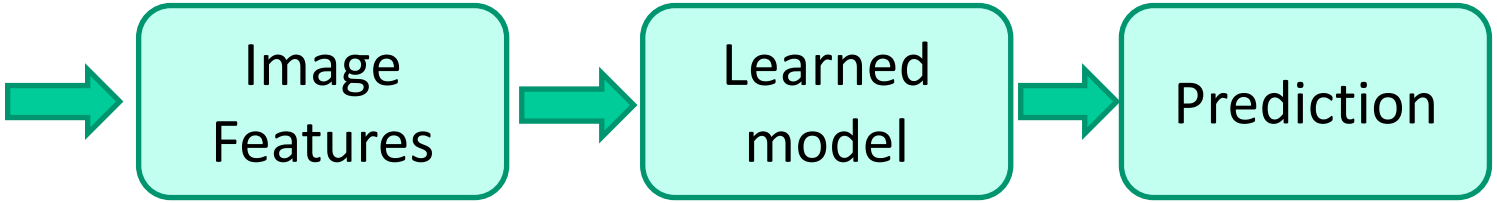


Machine Learning Flow

Training

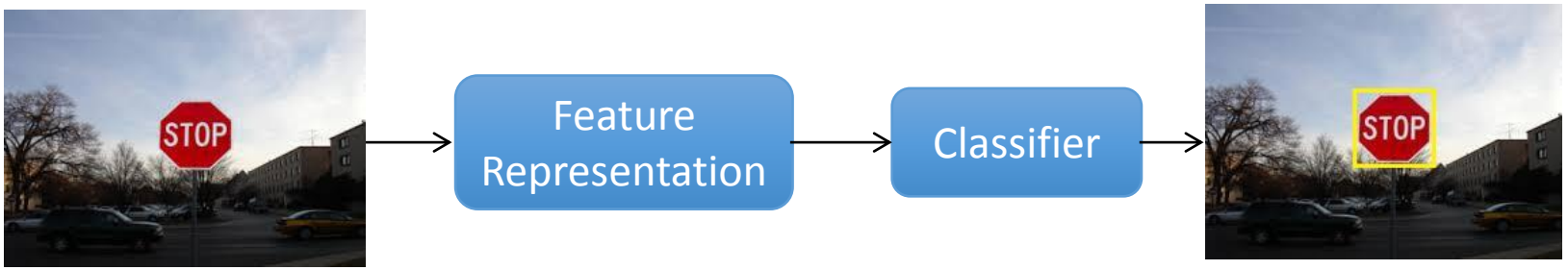


Testing



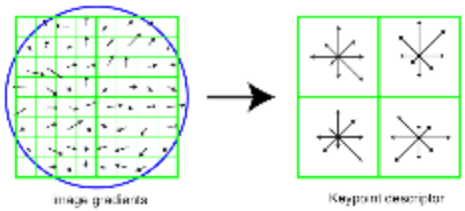
Test Image

Traditional ML

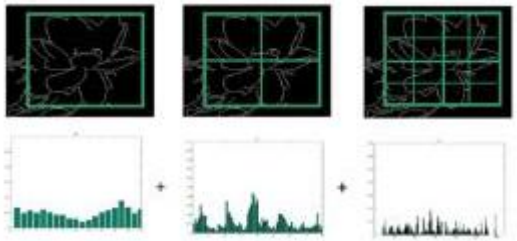


Engineered Low Level Features

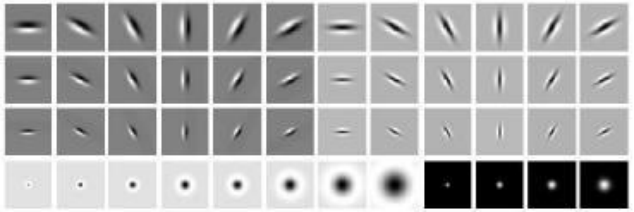
SIFT



HOG



Texton



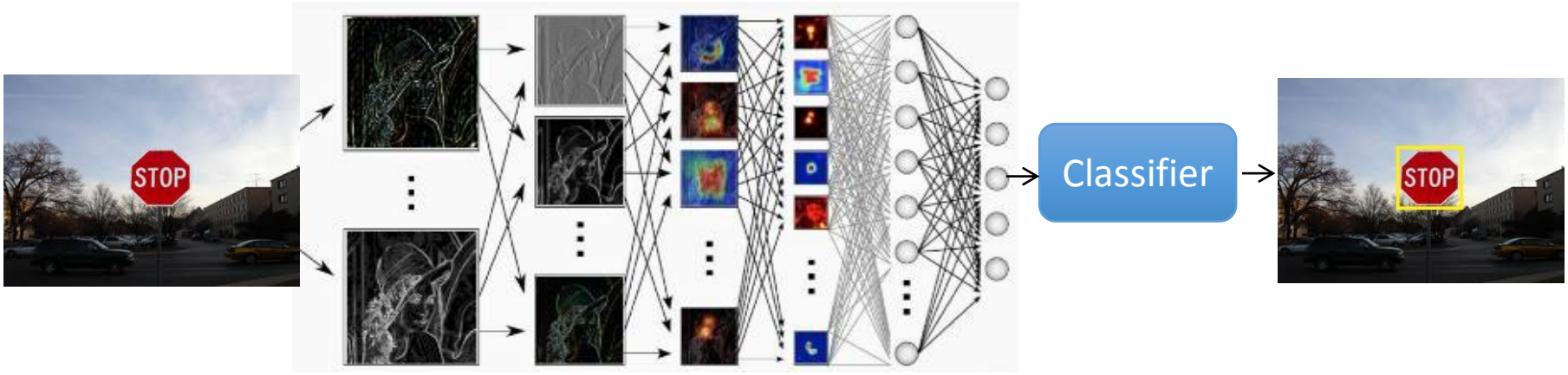
Feature Extraction Methods

- Features based on first-order/Second-order statistics (Mean, Variance, Energy, Entropy, GLCM, Law's Energy Masks)
- LBP, Ternary patterns, directional patterns
- Transform based (DWT, Gabor, Curvelet, etc)

Classical Classifiers

- Minimum distance Classifier
- Bay's Classifier
- SVM, Adaboost
- Neural Network (MLP-Backpropagation)
- Random Forest, Decision Trees etc.

Deep Learning



Hierarchical Representation

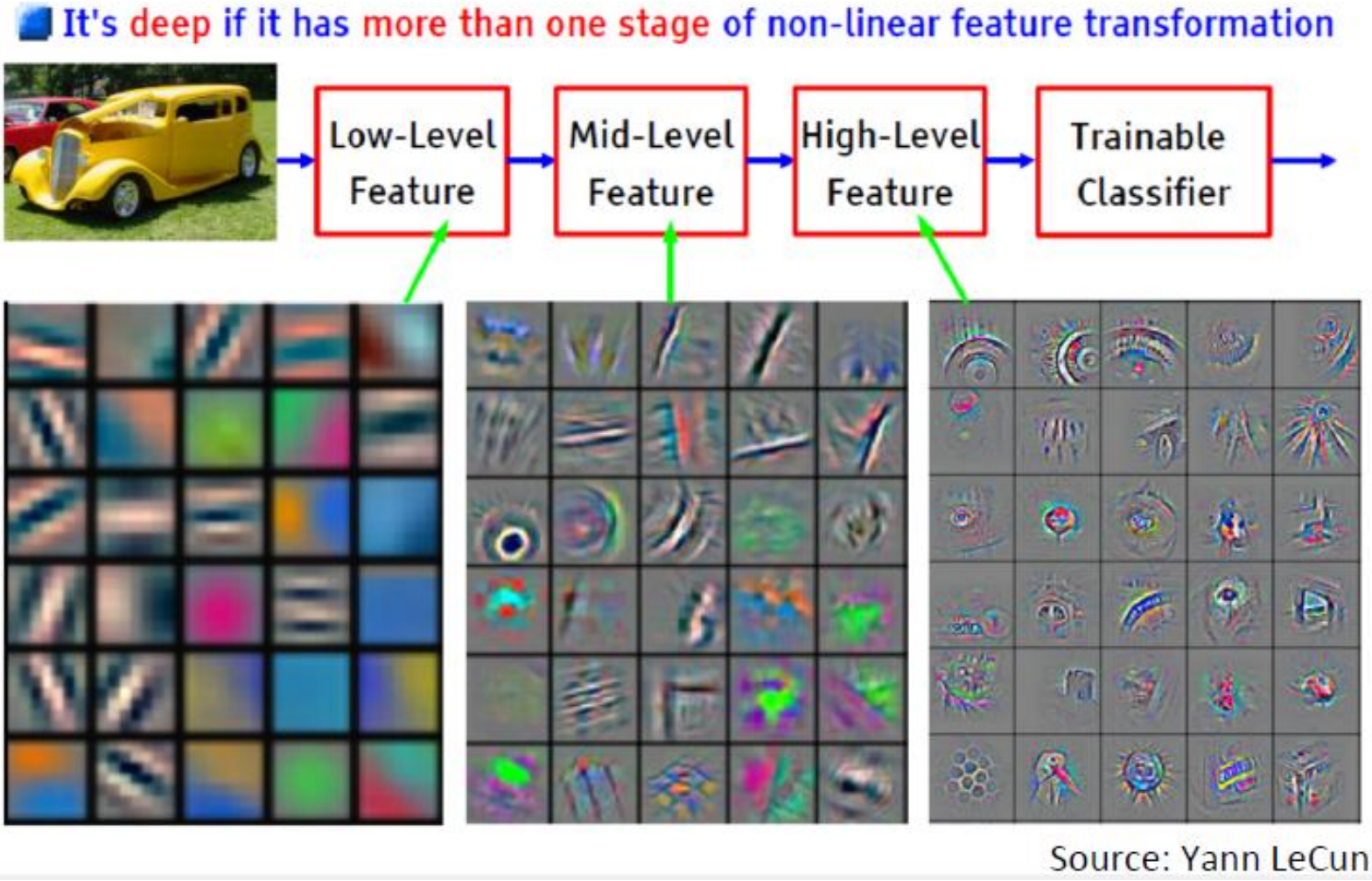
Features in DNN

(a) Low level features

(b) Mid level features

(c)High level features

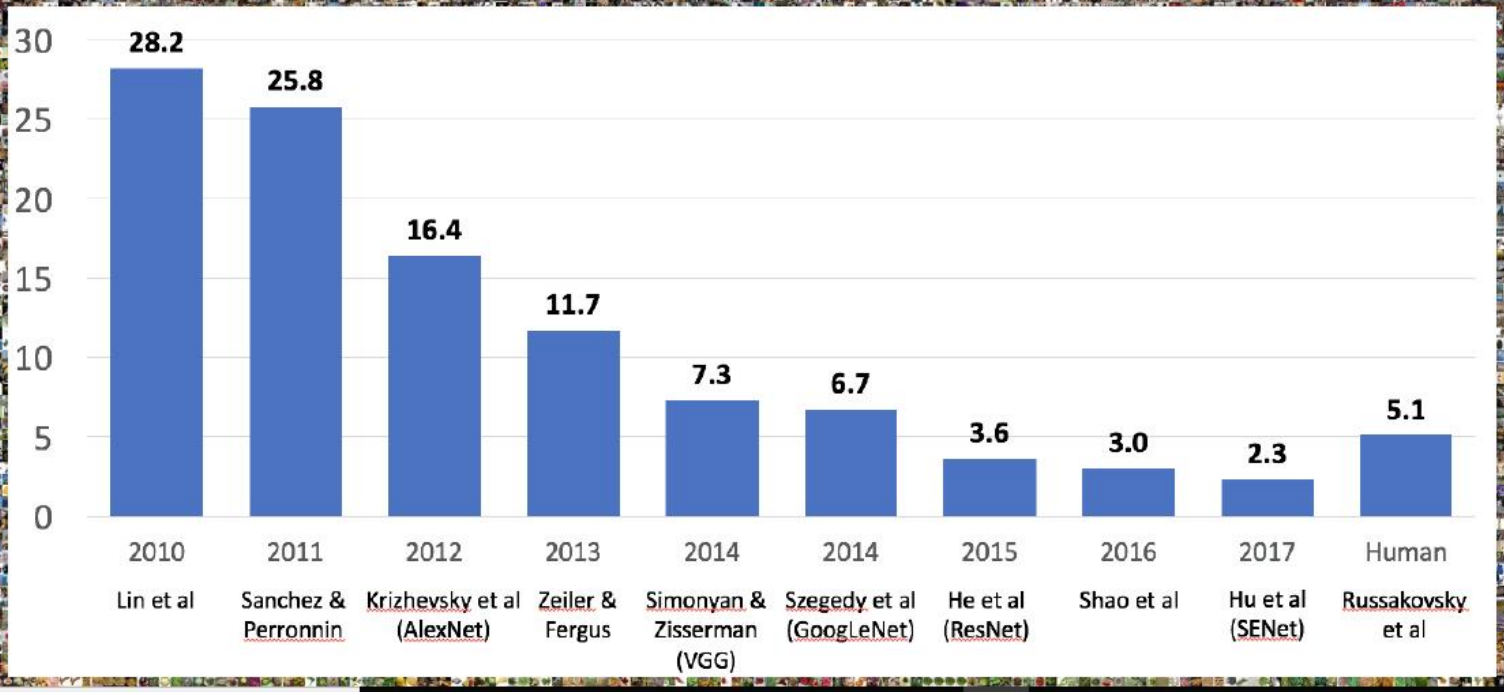
Deep Learning



Deep Learning in Visual Computing: The Turning Point

IMAGENET **Large Scale Visual Recognition Challenge**

The Image Classification Challenge:
1,000 object classes
1,431,167 images

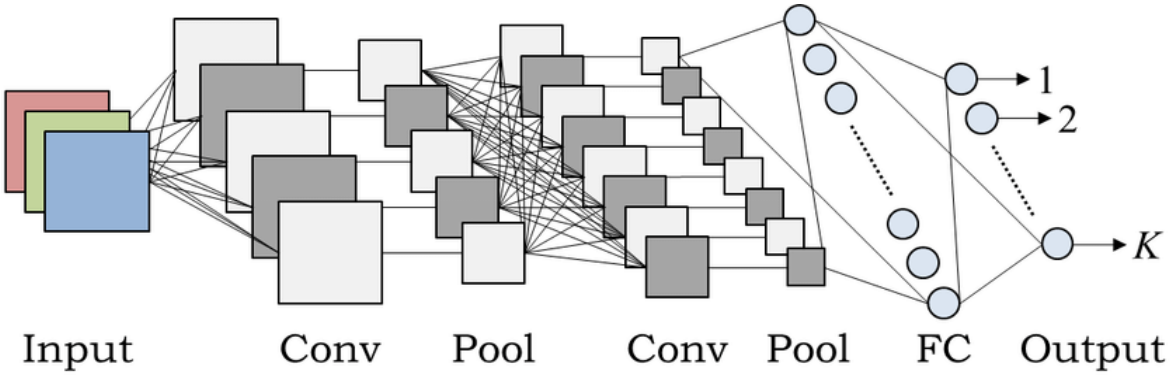


Ingredients for Deep Learning

Data

Computation

Algorithms

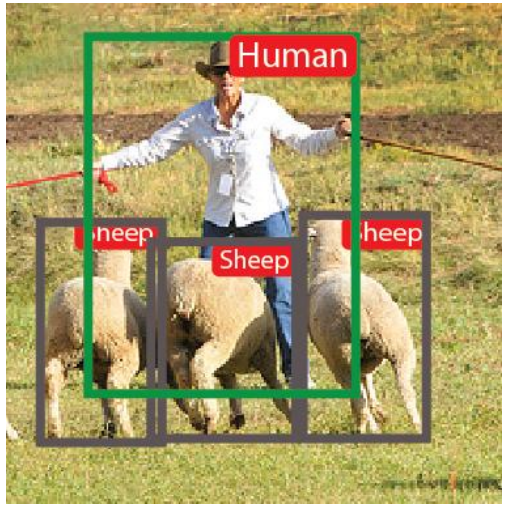


Deep Learning Applications

Image Classification

Classify an image based on the dominant object inside it.
Datasets: MNIST, CIFAR, ImageNet

Object Detection/Localization



Localize and classify all objects appearing in the image.
Datasets: PASCAL, COCO

Deep Learning Applications

Semantic Segmentation

Label each pixel of an image by the object class that it belongs to, such as human, sheep, and grass in the example.

Datasets: PASCAL, COCO

Instance Segmentation

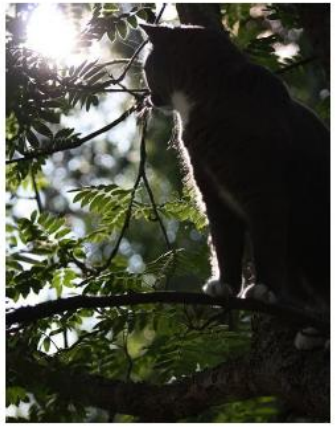
Label each pixel of an image by the object class and object instance that it belongs to.

Datasets: PASCAL, COCO

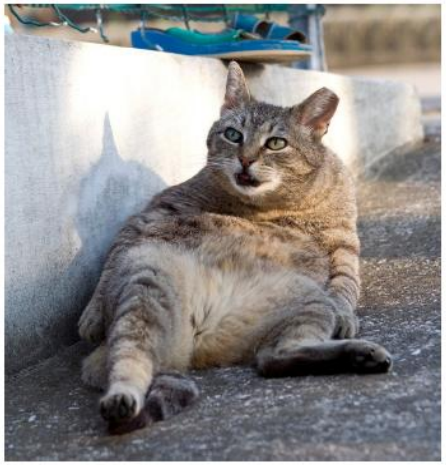
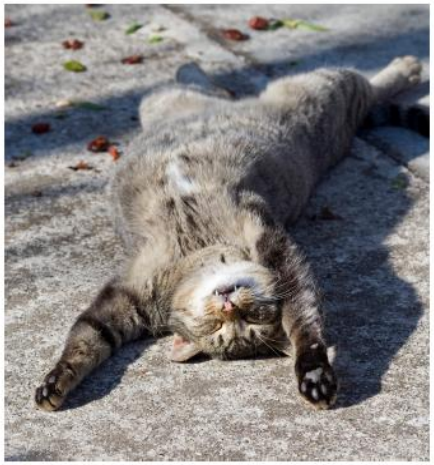
Challenges

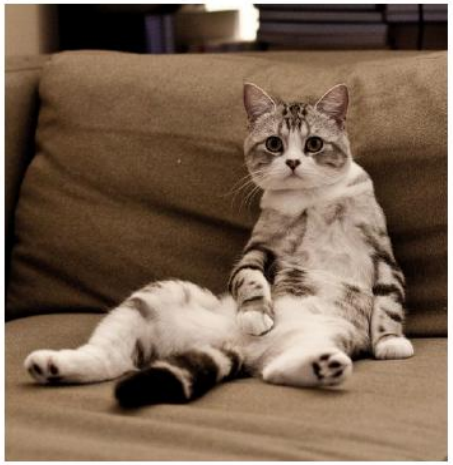
Illumination



Deformation



Challenges

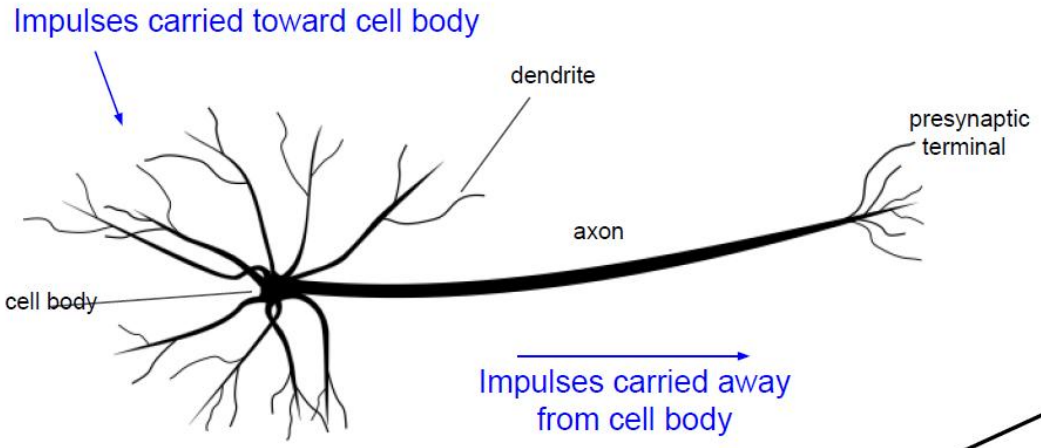
Occlusion

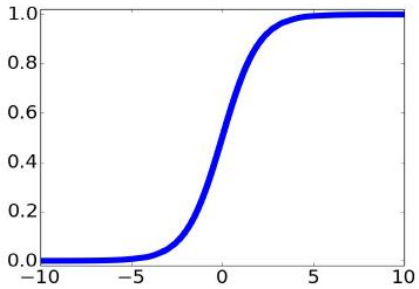


Clutter


Intra class Variation

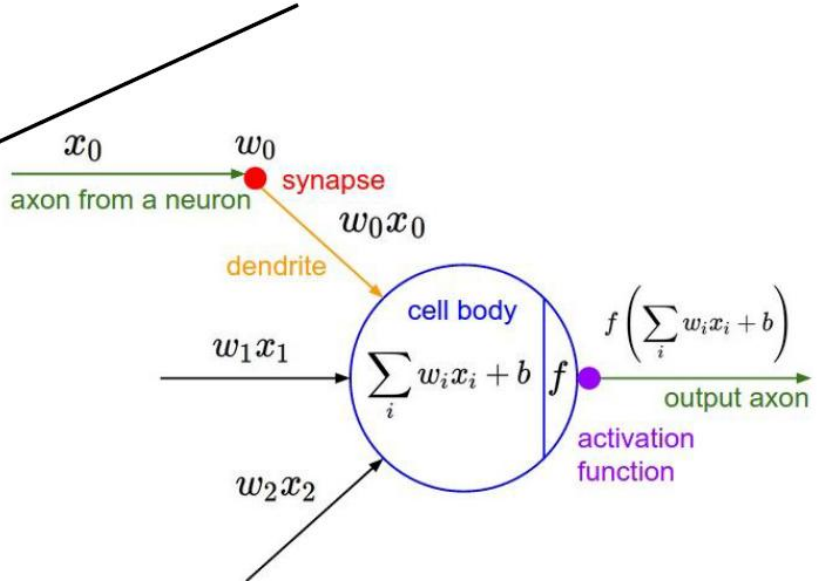
Analogy of Neural network



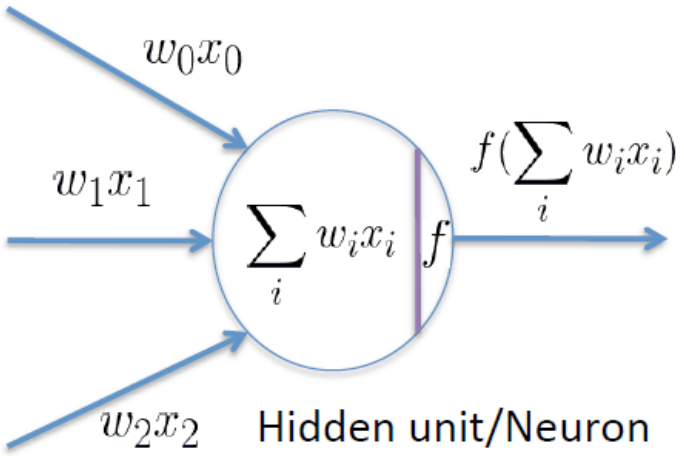


sigmoid activation function

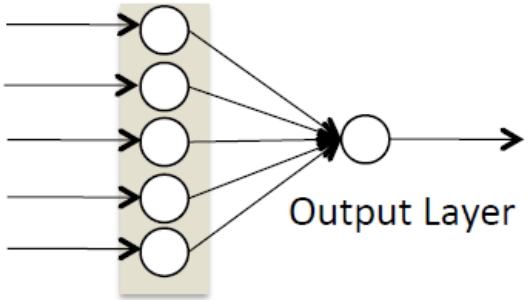
$$\frac{1}{1+e^{-x}}$$



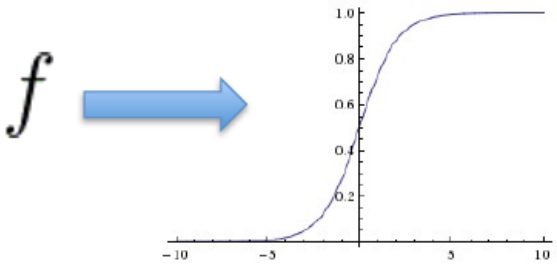
Brief Review Neural network



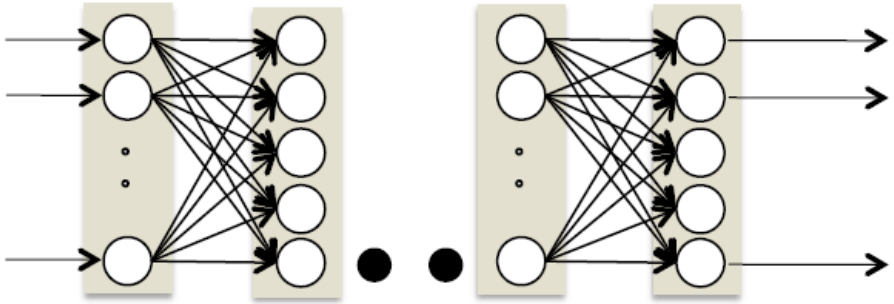
Hidden unit/Neuron



Perceptron



E.g. Sigmoid Activation Function



Multi Layer Perceptron

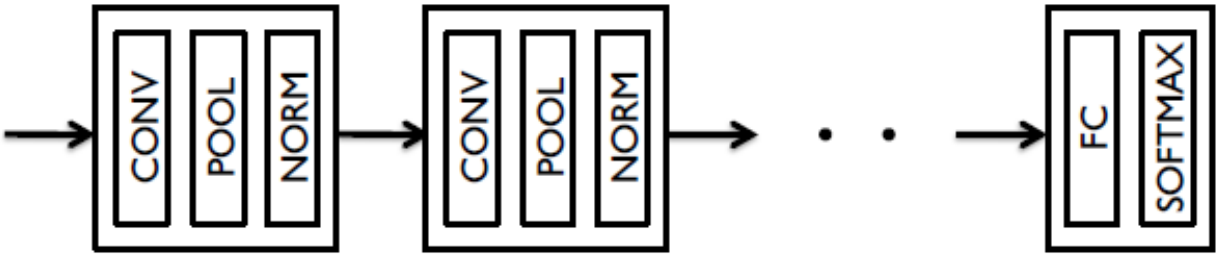
Handwritten Digit Recognition: CNN Approach

- Convolution is a simple mathematical operation between image matrix and filter matrix in which one is multiplied with the other elementwise and sum of all these multiplications is calculated.
- Convolution provide better feature extraction
- CNN requires less number of parameters and hence save lot of computation compared to ANNs
- Due to less number of required parameters, memory requirement and training time is reduced drastically

Basic CNN Architecture

Convolutional Neural Network

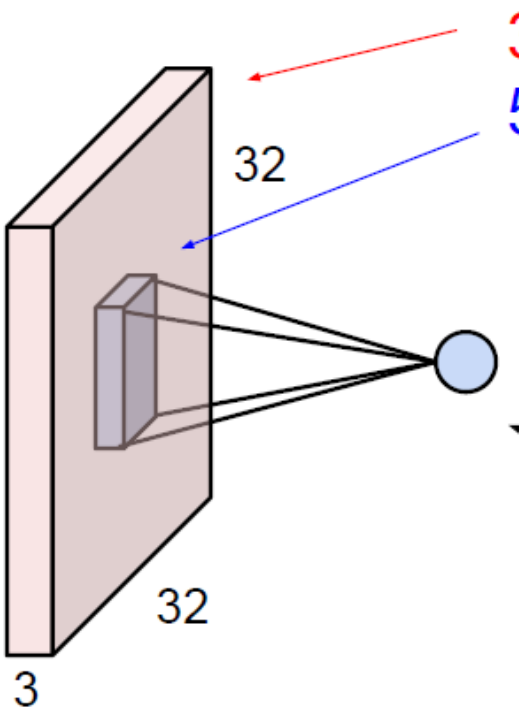
A convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied for analysing visual imagery.



- Convolutional Layer
- Activation Function
- Pooling
- Normalization
- Fully Connected/ Dense
- Classification Layer/ Softmax

Convolution Layer

CNN's make use of filters (also known as kernels), to detect what features, such as edges, are present throughout an image. A filter is just a matrix of values, called weights, that are trained to detect specific features.



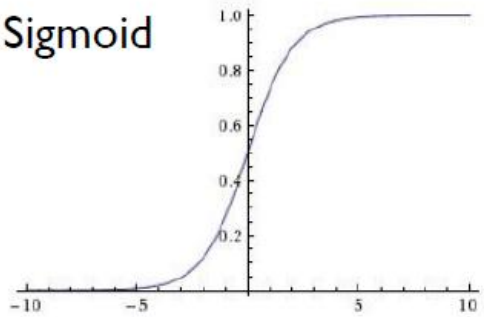
1 number:
the result of taking a dot product between the filter and a small 5x5x3 chunk of the image
(i.e. 5*5*3 = 75-dimensional dot product + bias)

$$w^T x + b$$

Activation Functions

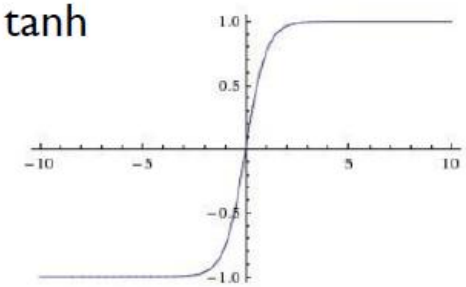
The purpose of the activation function is to introduce non-linearity into our network.

Sigmoid



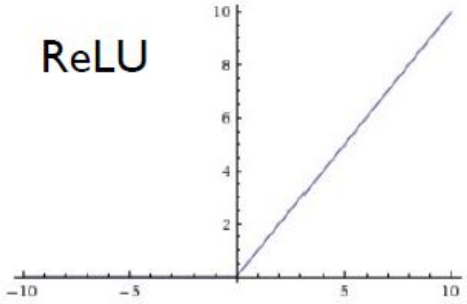
$$y = \frac{1}{1 + e^{-x}}$$

tanh



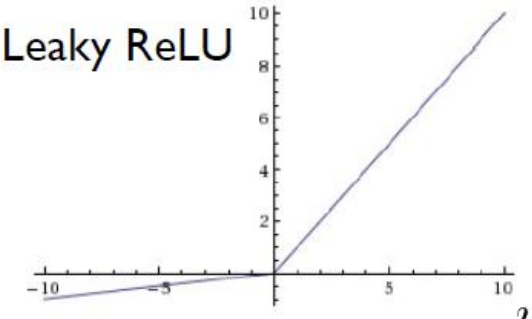
$$y = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

ReLU



$$y = max(0, x)$$

Leaky ReLU



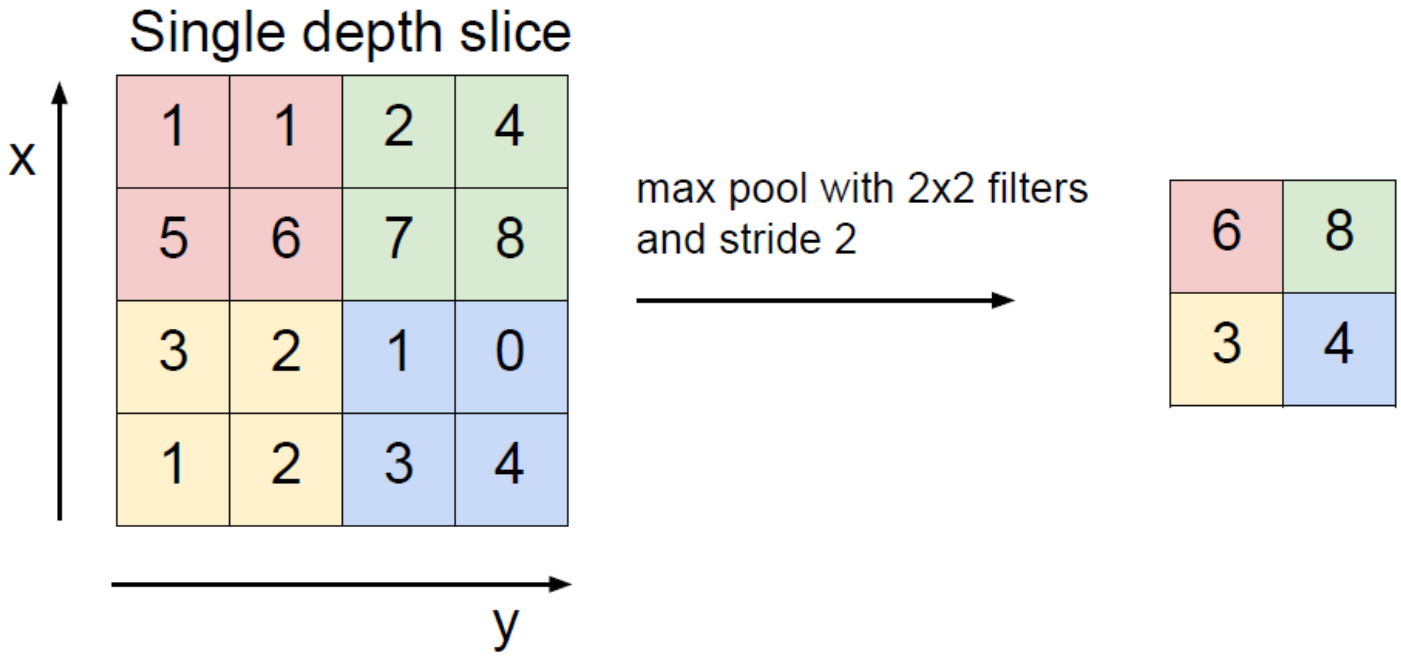
$$y = \begin{cases} x & \text{if } x < 0 \\ 0.01x & \text{if } otherwise \end{cases}$$

maxout

$$max(w_1^T x + b_1, w_2^T x + b_2)$$

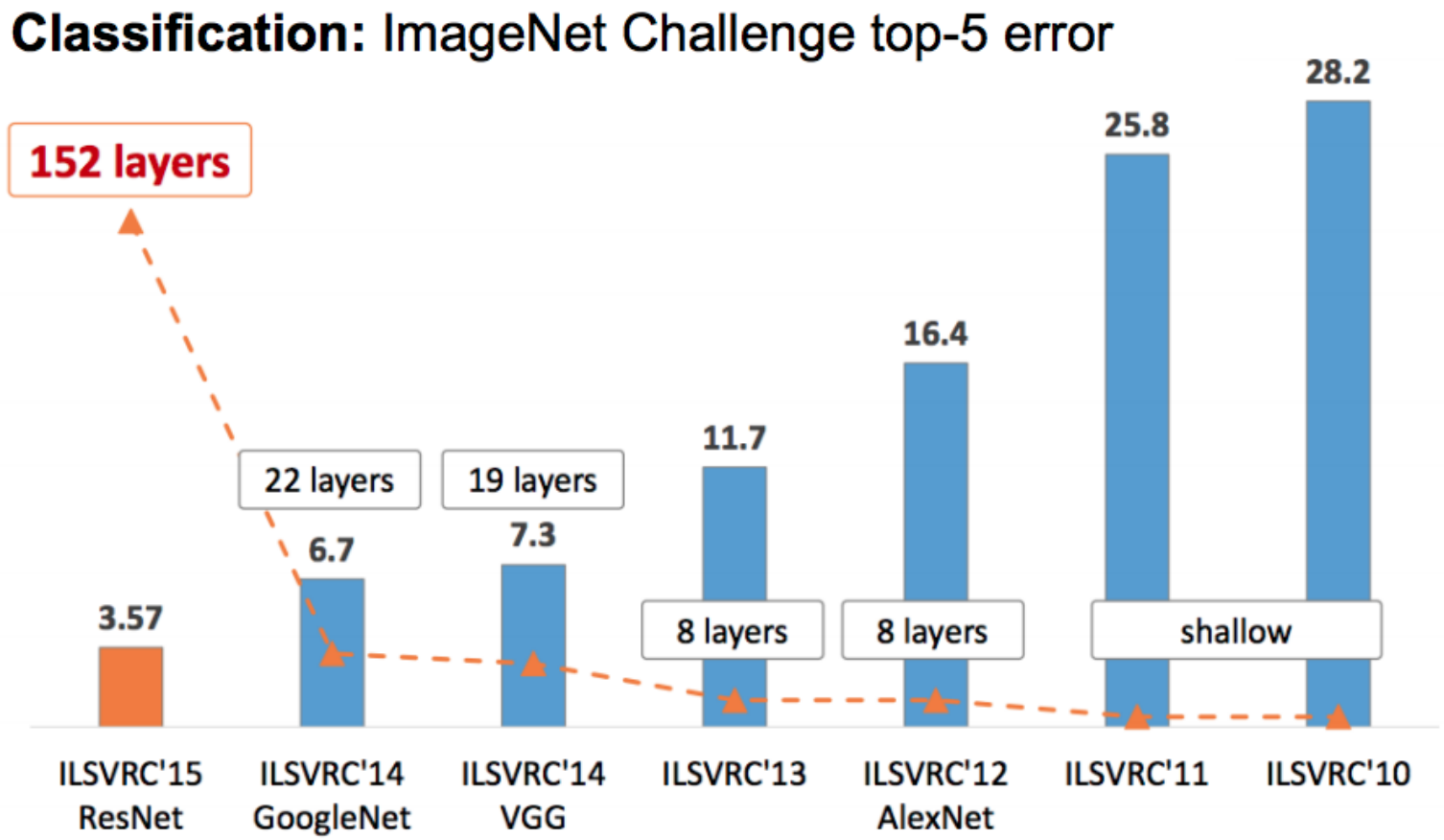
Max Pooling

- To speed up the training process and reduce the amount of memory consumed by the network, we try to reduce the redundancy present in the input feature.
- There are a couple of ways we can downsample an image, but the most common one is **max pooling**.

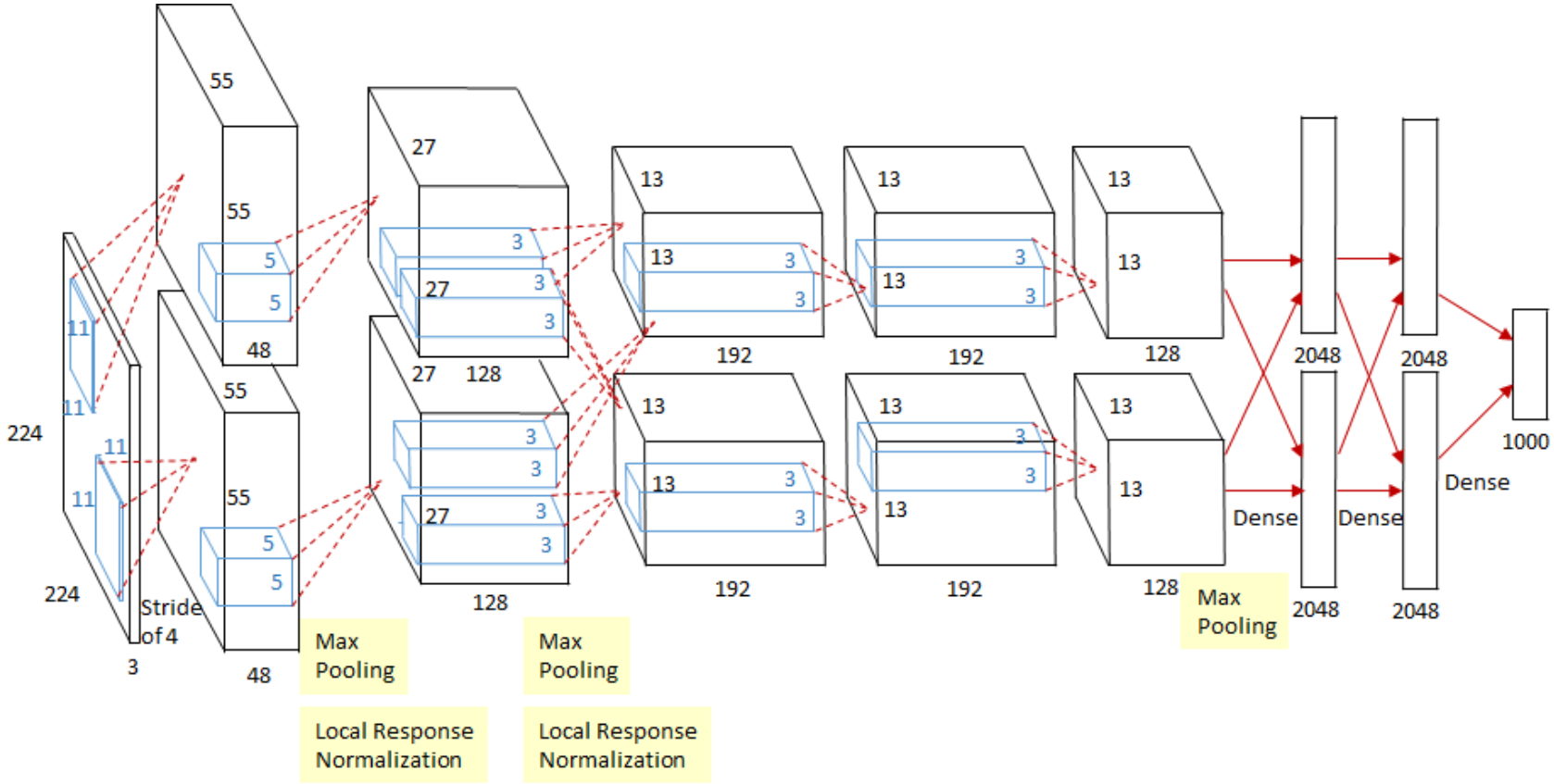


Various CNN Architectures

The ImageNet project runs an annual software contest, the **ImageNet Large Scale Visual Recognition Challenge (ILSVRC)**, where algorithms compete to correctly classify and detect objects and scenes.



Case Study: AlexNet

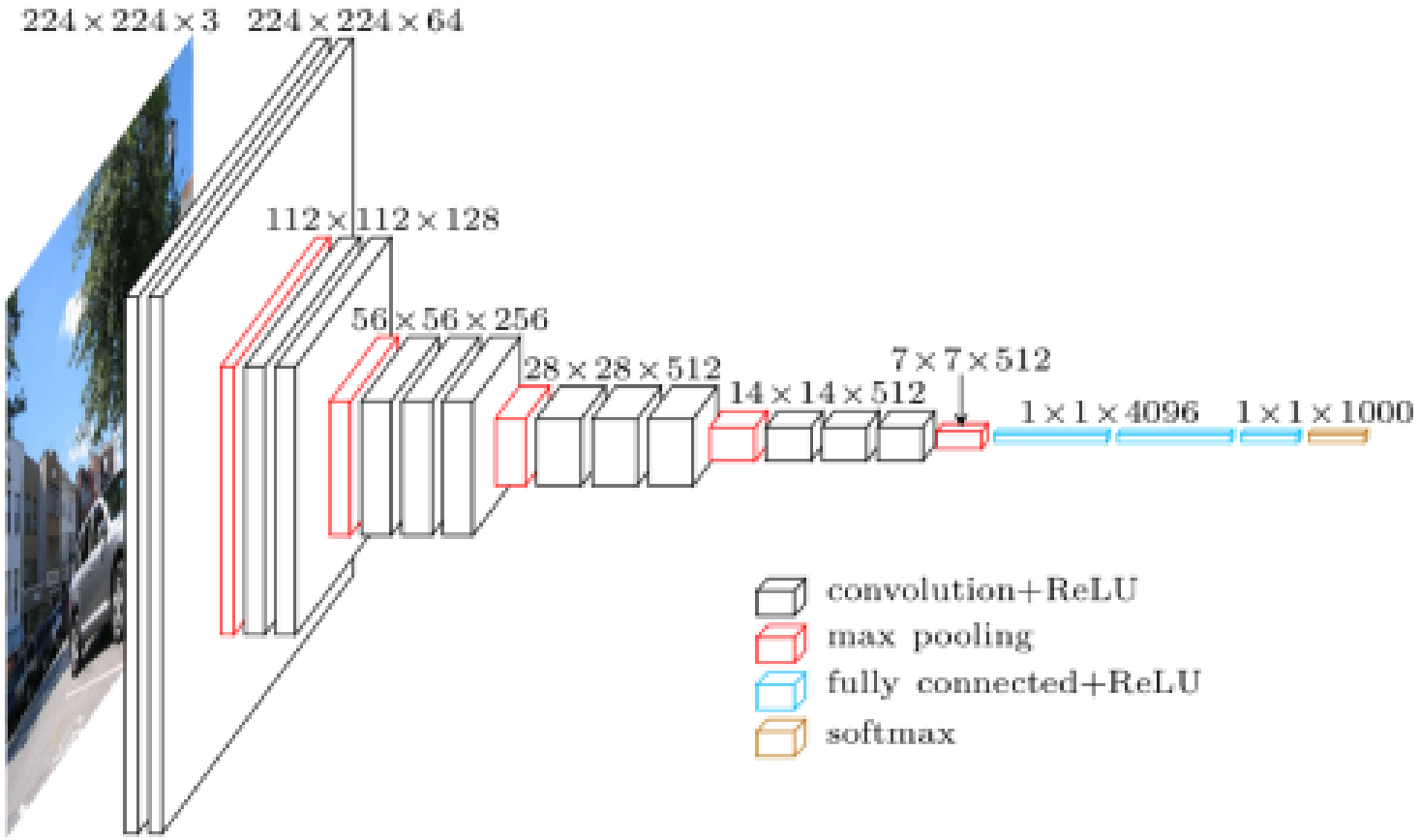


A. Krizhevsky, I. Sutskever, and G. E. Hinton, "***Imagenet classification with deep convolutional neural networks,***" in Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS) - Volume 1, NIPS 2012. USA: Curran Associates Inc., 2012, pp. 1097–1105

AlexNet Highlights

- AlexNet won the ImageNet competition in 2012 by a large margin. It was the biggest network at the time and achieved state-of-the art performance in object recognition.
- The network was split into two halves, each trained simultaneously on two different GPUs.
- It used Relu activation function instead of Tanh to add non-linearity. It accelerates the speed by 6 times at the same accuracy.
- AlexNet used 0.5 dropout during training as regularization technique to avoid overfitting.
- It also used the technique of data augmentation.
- The network has 62.3 million parameters, and needs 1.1 billion computation units in a forward pass.

VGGNet



K. Simonyan and A. Zisserman, "***Very deep convolutional networks for large-scale image recognition***," International Conference on Learning Representations (ICLR), vol. abs/1409.1556, 2015.

VGGNet Highlights

- This architecture is proposed by Visual Geometry Group of University of Oxford.
- This network is characterized by its simplicity. Uses only 3x3 convolution and 2x2 pooling layers throughout the whole network.
- Two fully-connected layers, each with 4,096 nodes are then followed by a softmax classifier.
- Several variants of VGGNet are proposed in the paper, but VGG16 performs the best.
- The weight configuration of the VGGNet is publicly available and has been used in many other applications and challenges as a baseline feature extractor.
- However, the major drawback of VGGNet is total number of parameters which is 138 million. It is a bit challenging to handle.

Summary

- Deep Convolutional Networks
    - Conv, Norm, Pool, FC Layers
    - Training by Backpropagation
- Many Specific enhancements
    - Nonlinearity (ReLu), Dropout, SGD
- Lots of data, Lots of Computations
- Anatomy and Physiology of AlexnNet
    - Architecture, Parameters


Thank
you