# Machine Learning Short Laboratory Course

04/03/2022

# Previous lesson

- Unsupervised learning:
  - Clustering:
    - K-means
    - DBScan
    - Hierarchical Clustering
  - Dimensionality Reduction:
    - PCA
    - t-SNE

- Materials available on MLAdventure/ML_Short_Lab

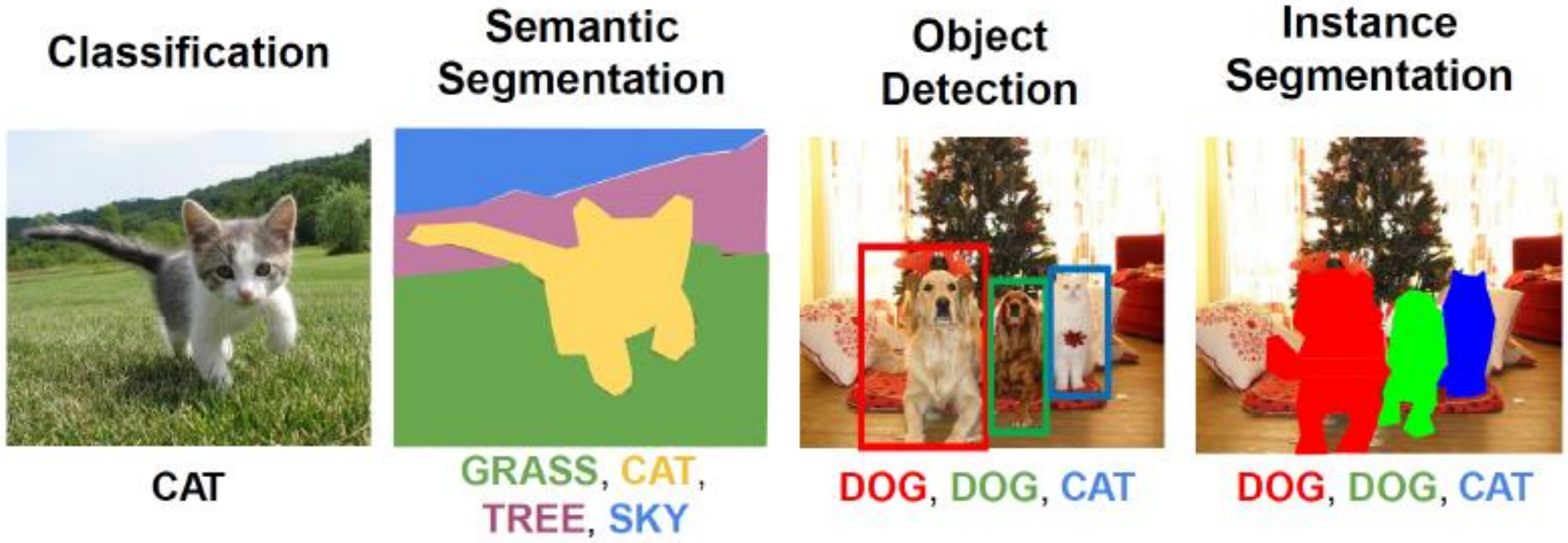
# Computer Vision

- What is Perception:
  - Science:
    - Understanding how do we see (explore computational model of human vision)
  - Engineering:
    - Build systems that perceive the world
  - Applications:
    - Medical imaging
    - Surveillance
    - Entertainment
    - Car industry

# Computer Vision

- Basic concepts:
  - Goal: localize object
  - Recognition problems:
    - Identification: recognize a specific object (es: your pen)
    - Classification: recognize a class of objects (es: any pen)
  - Segmentation:
    - Separate pixels belonging to the foreground (object) and the background
  - Localization/Detection:
    - Position of the object in the scene, pose estimate
  - Challenges:
    - Multi scale
    - Multi view
    - Multi class

# Computer Vision



# Digital Image Filtering

- Image filtering :
  - The process of applying some function to local image patches.
  - Goal :
    - Reduce noise
    - Fill-in missing values
    - Extract image features

- Linear filtering :
  - Simplest case of image filtering
  - Replace each pixel by a linear combination of its neighbours :
    - 2D discrete convolution

$$f[m,n] = I \otimes g = \sum_{k,l} I[m-k, n-l] \cdot g[k,l],$$

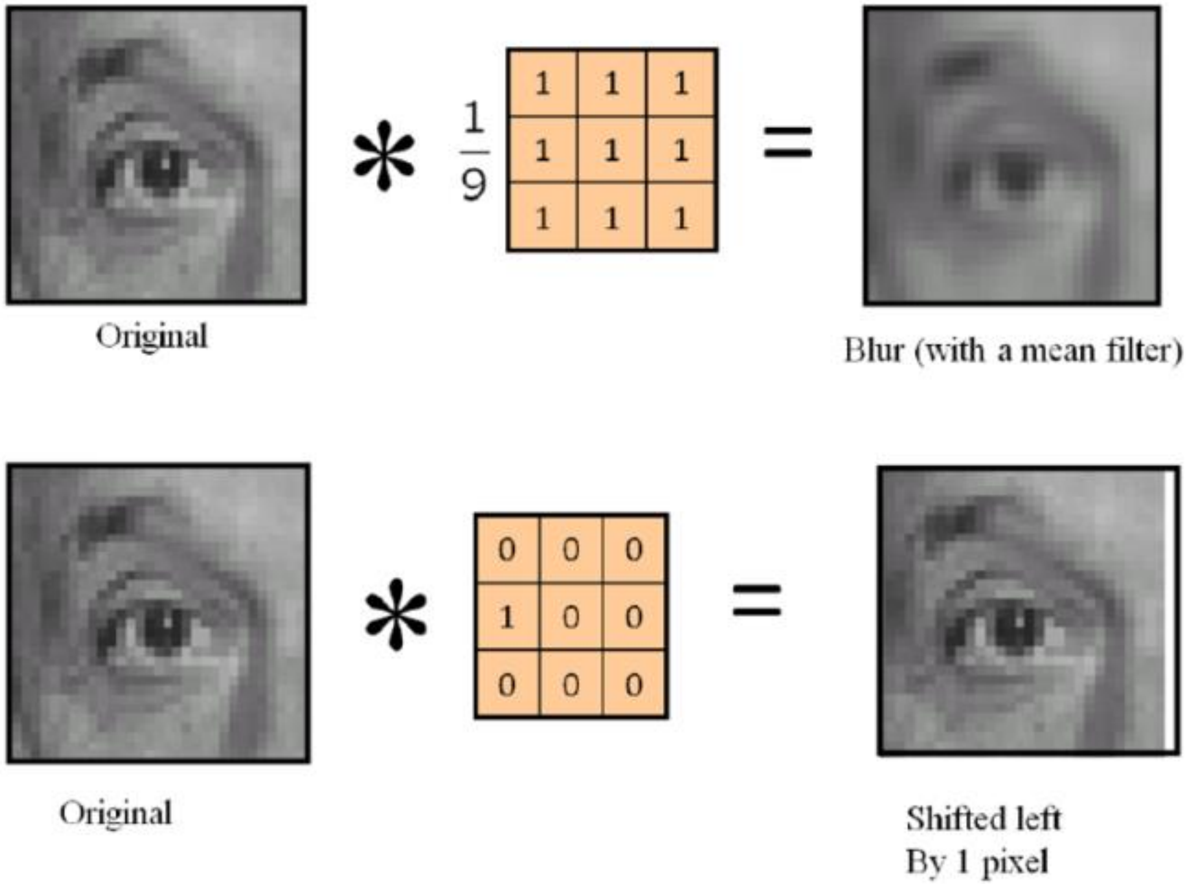
# Digital Image Filtering

- Analogously, if we have a 1D-lter and a 2D image, the convolution operation is:

$$f[m,n] = I \otimes g = \sum_k I[m-k,n] \cdot g[k].$$

- Examples:
  - With a filter [0; 0; 0; 0; 1; 0; 0; 0; 0] the image stays the same
  - With a filter [0; 0; 0; 0; 0; 0; 0; 1; 0] the image is shifted on the left of 3 pixels
  - With a filter [0; 0; 0; 1; 1; 1; 0; 0; 0] the image is blurred

# Digital Image Filtering



# Digital Image Filtering : Filtering to Reduce Noise

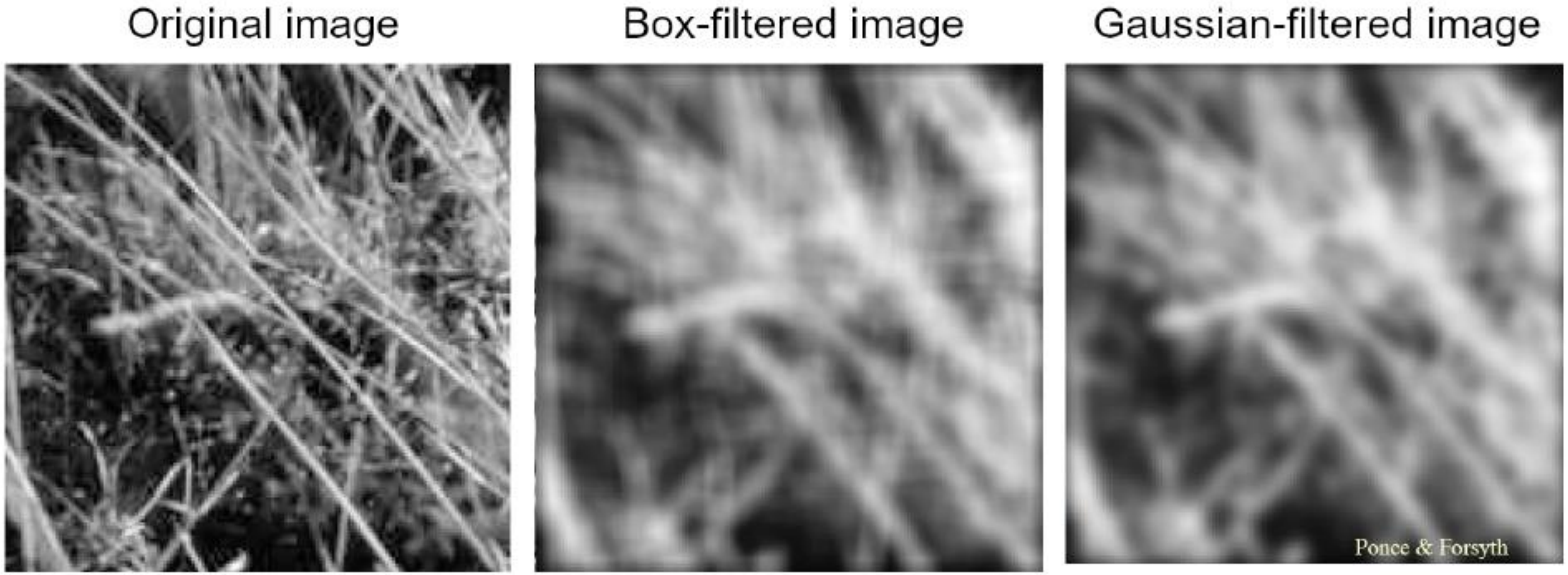
- Filtering can be applied to reduce noise :
    - light fluctuations
    - sensor noise
    - quantization effects
    - extraneous objects (complex noise)

- To do this, we assume that the neighbourhood of a pixel contains information about its intensity

# Digital Image Filtering : Filtering to Reduce Noise

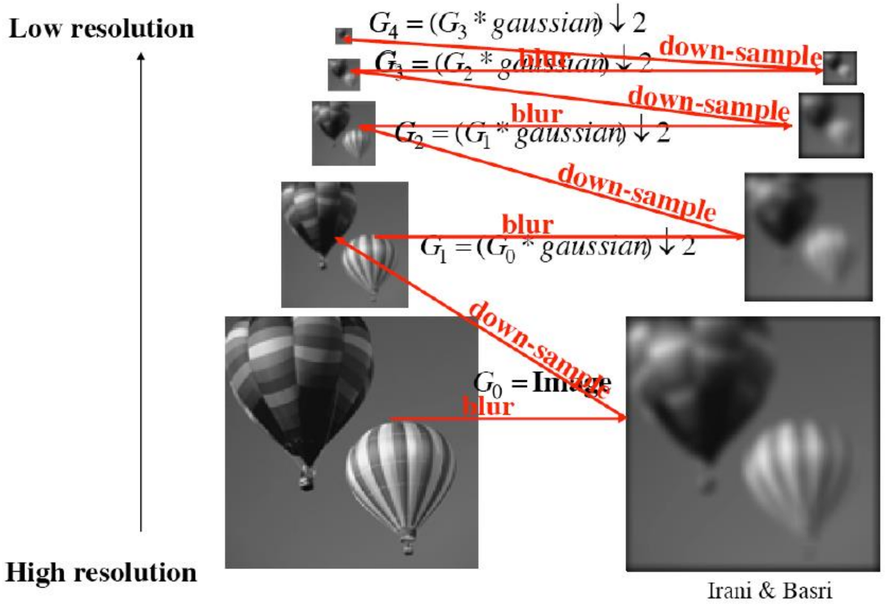
- Average Filter:
  - Replaces each pixel with an average of its neighbourhood
  - If all weights are equal, it is called a box filter (and it is separable)
- Gaussian Filter :
  - Smarter average filter
  - It's rotationally symmetric
  - It weights nearby pixels more than distant ones (probabilistic inference)
  - The “amount” of smoothing depends on the parameter of the Gaussian
  - It is separable

# Digital Image Filtering: Gaussian Filter



# Multi Scale Image Representation

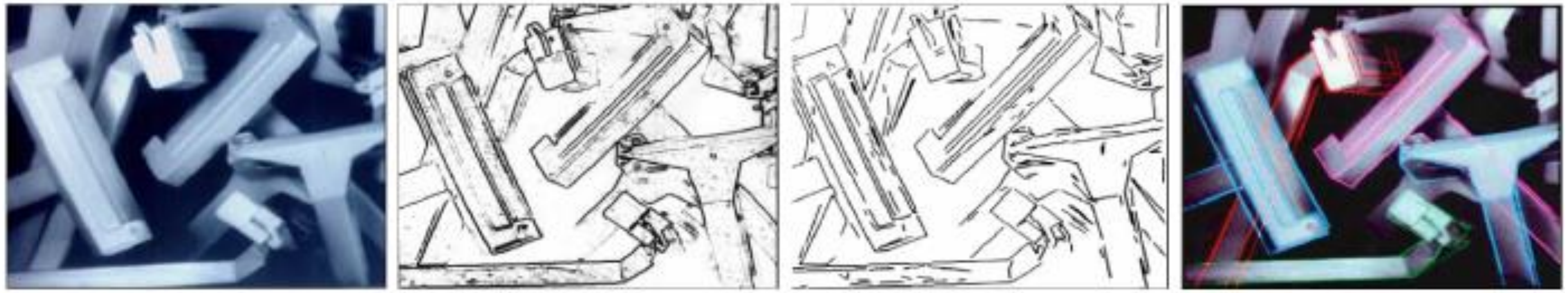
- To classify objects, a possibility is to compare the image with a set of objects' templates, used as kernels
    - Problem : computationally expensive, different sizes image
    - Solution : Gaussian Pyramid




- Blurring: fast -> small linear filter
- Down-sampling: fast -> remove a certain percentage of the pixels in the image

# Edge Detection: Old Style

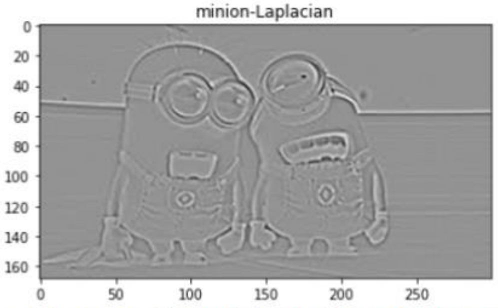
- The common approach in the 80s was to match models to edges and lines, so it was needed to reliably extract lines and edges
- Goals of edge detection:
  - Good detection: filter responds to edge, not to noise
  - Good localization: detected edge near true edge
  - Single response: one per edge



# Edge Detection: Old Style

- Edge detection methods are:
  - Laplacian Operator

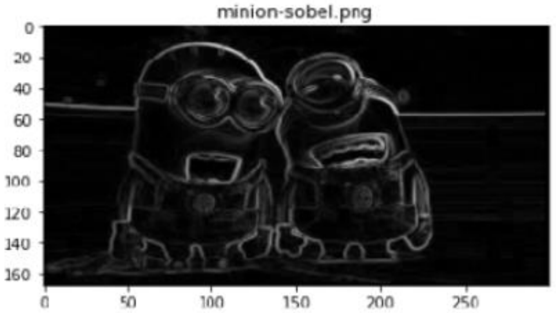
original minion image




minion image after applying Laplacian edge detection

  - Sobel Edge Detection

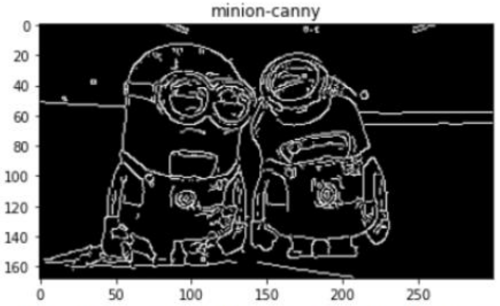
original minion image




minion image after applying sobel edge detector

  - Canny Edge Detection

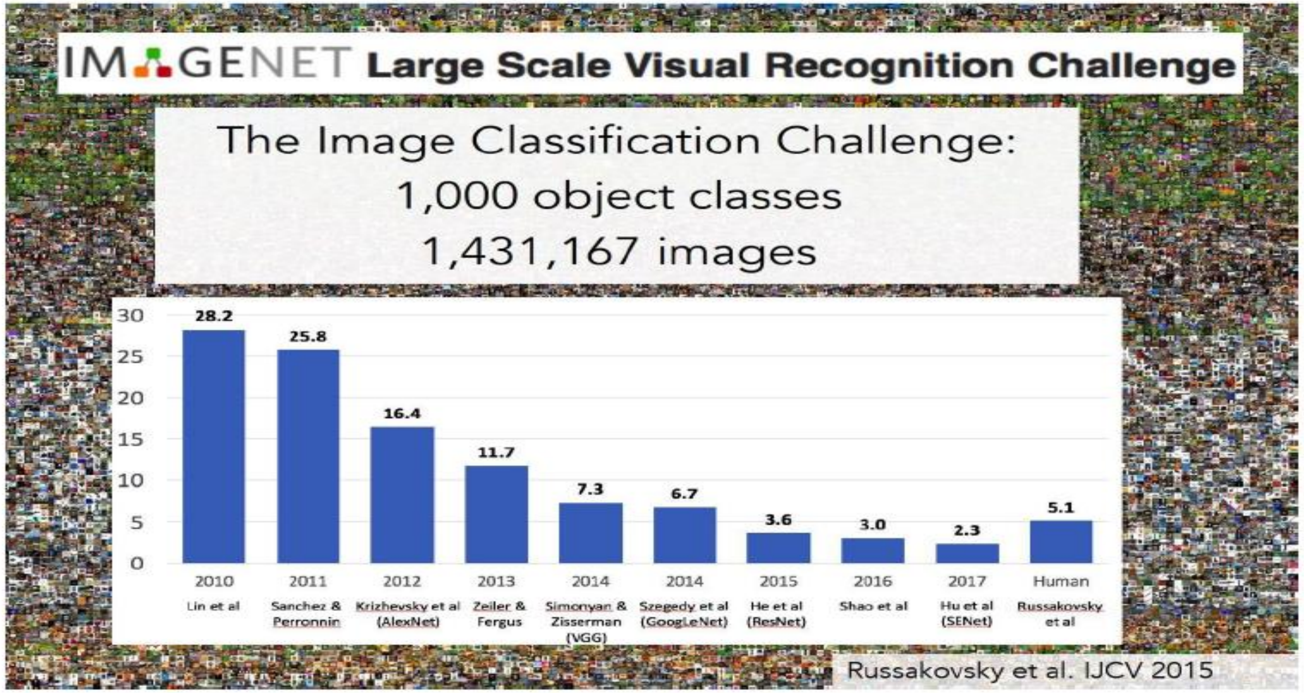
original minion image




minion image after applying canny edge detection

# Computer Vision: Today

- CNN proved to be very effective in image classification tasks
- This model is used also for semantic segmentation, detection, interpretation of pose and actions, even within videos (spatial and temporal streams).



# Computer Vision: CNN

- Charcteristics of CNN :
  - Feed-forward:
    - convolve input
    - pooling
  - Supervised
  - Train convolutional filter by back-propagating classification error

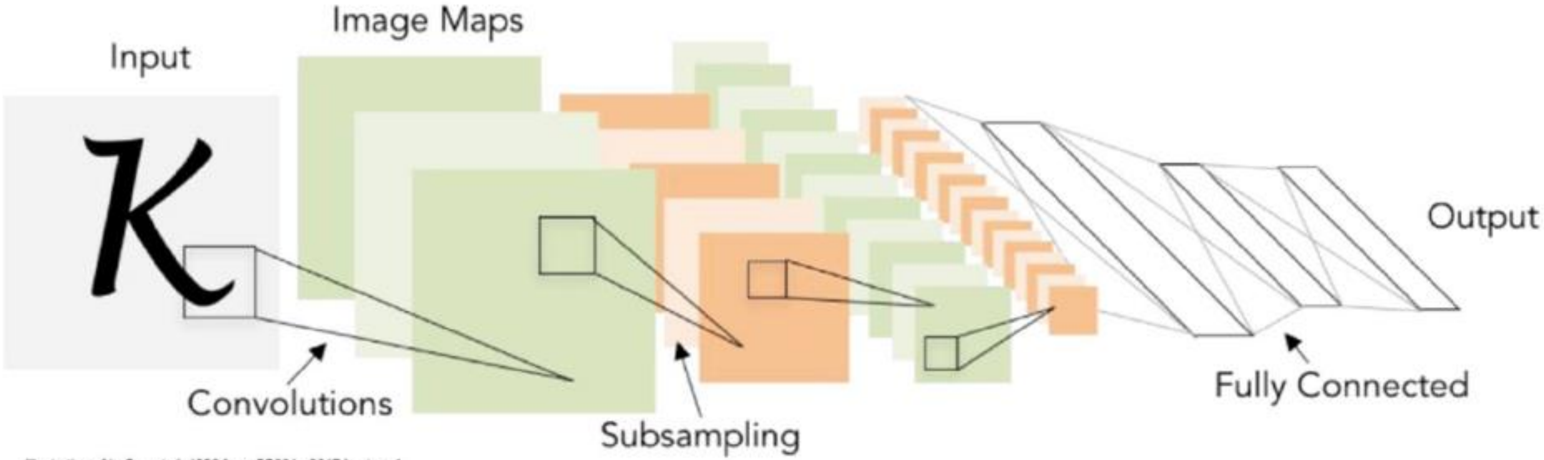
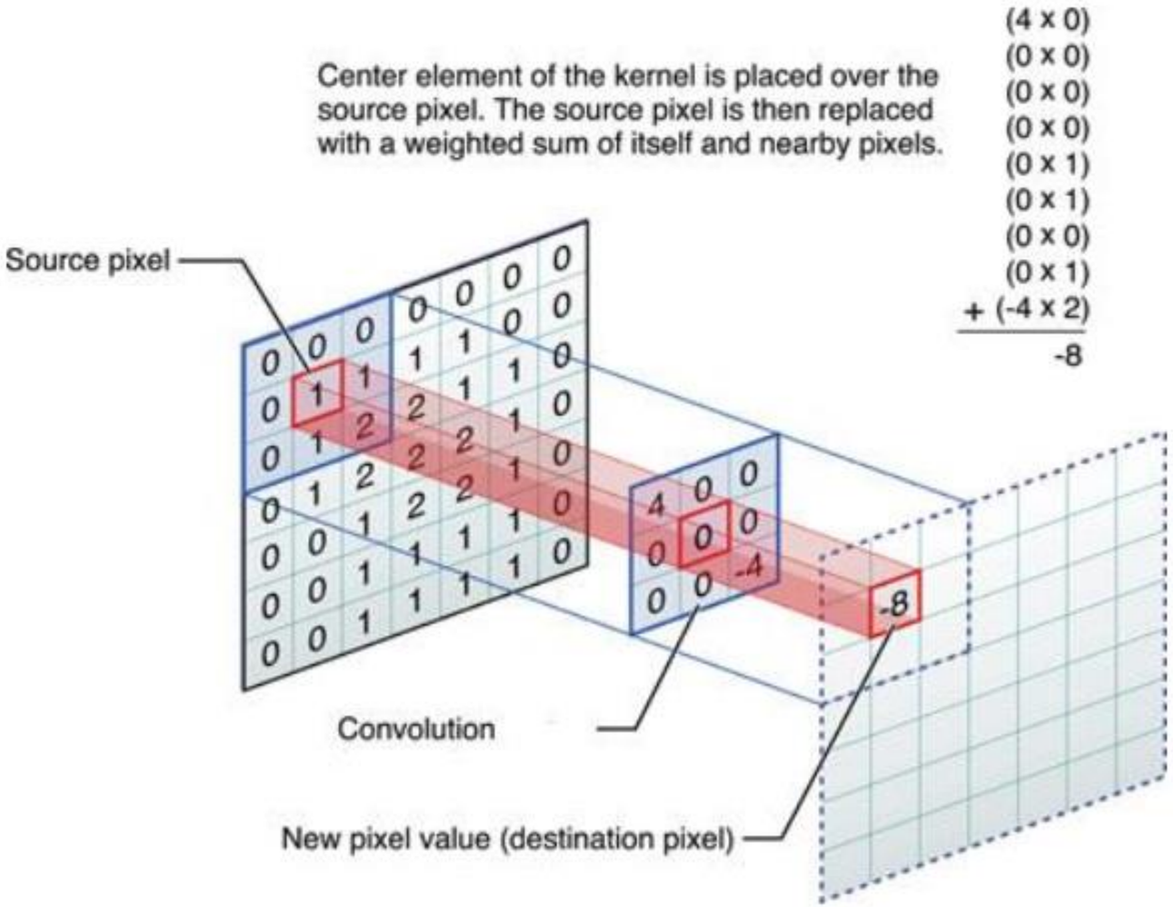



Illustration of LeCun et al. 1998 from CS231n 2017 Lecture 1

# Computer Vision: CNN

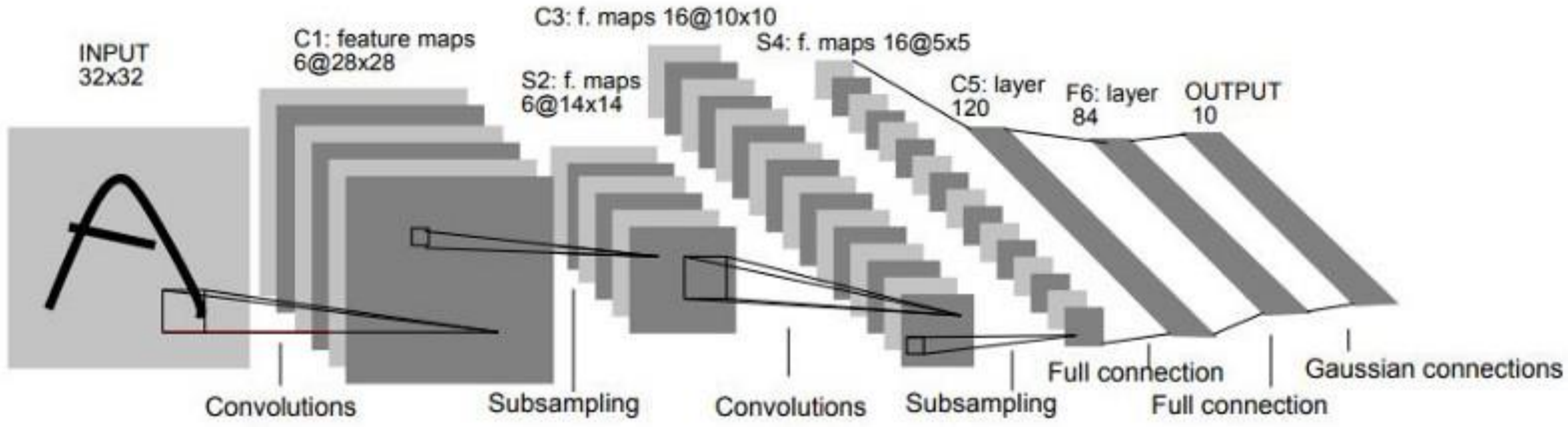
- Convolution:
  - Weight sharing
  - Image translation
  - Output is Feature maps



# LeNet (1990)

- Task: handwritten and machine-printed character recognition

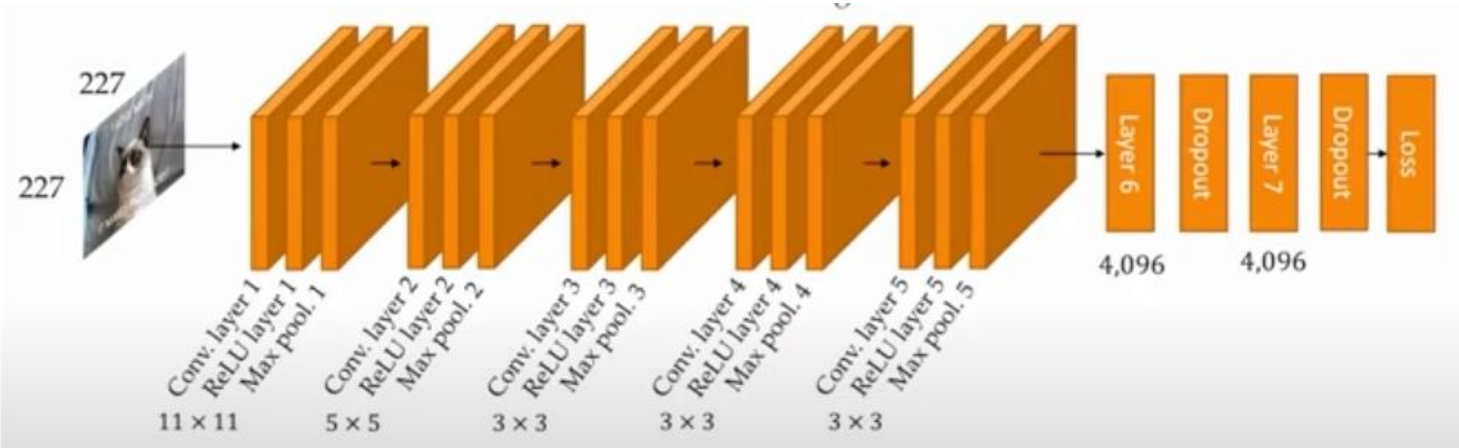
Architecture :
two sets of convolutional and average pooling layers, followed by a flattening convolutional layer, then two fully-connected layers and finally a softmax classifier



# AlexNet (2012)

- Won the Imagenet large-scale visual recognition challenge
- Accuracy of 84.7%

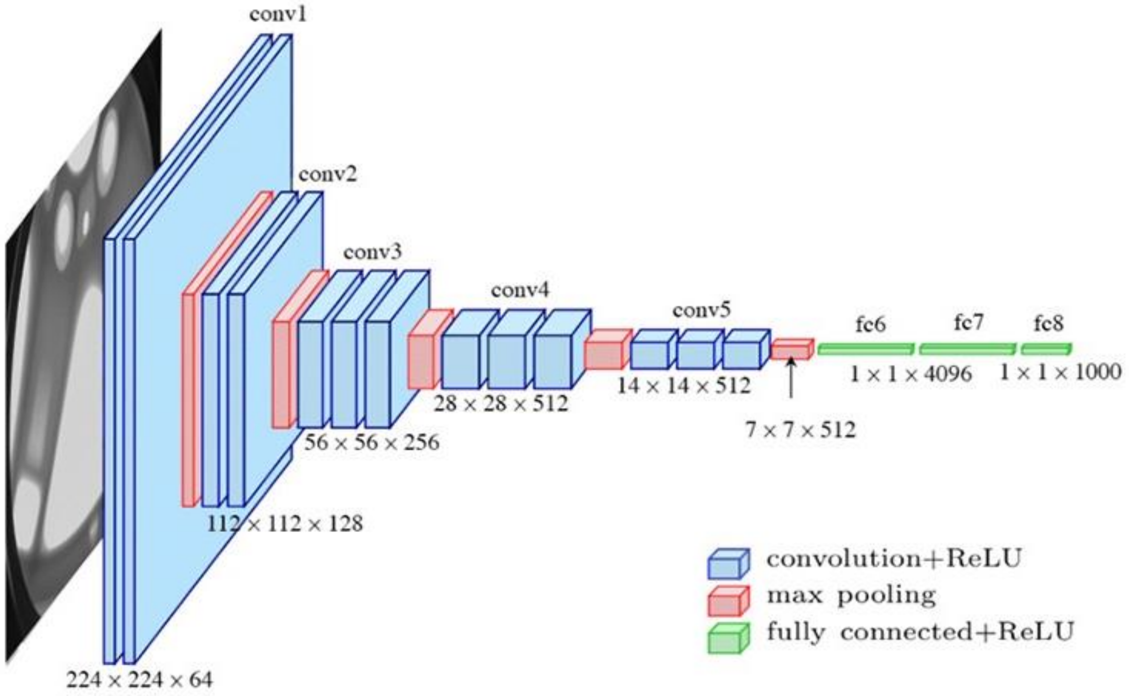
Architecture:
the model consists of five layers with a combination of max pooling followed by 3 fully connected layers with Relu activation in each of these layers except the output layer. There is also the dropout layers, that prevented the model from overfitting



# VGG (2014)

- Visual Geometry Group, won Imagenet Challenge with 97.2% accuracy
- Represent a family of models

VGG16 Architecture:
It has two contiguous blocks of two convolution layers followed by a max-pooling, then it has three contiguous blocks of three convolution layers followed by max-pooling, and at last, there are three dense layers.
 The last three convolution layers have different depths in different architectures.



# ResNet (2015)

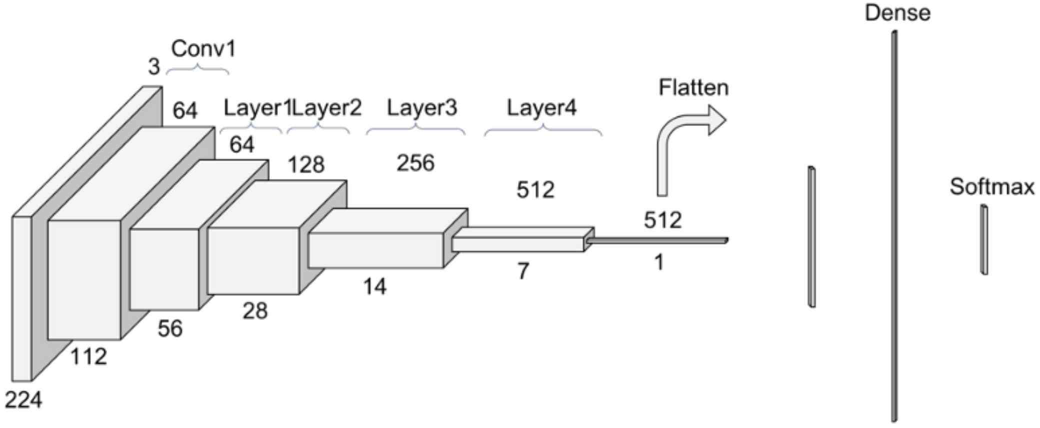
- The first truly Deep Network, going deeper than 1000 layers
- Smashed Imagenet, with a ~3% error

Architecture :

The architecture is almost similar to the previous networks. The difference is that ResNet uses Batch Normalization at its core to mitigate the problem of covariate shift.

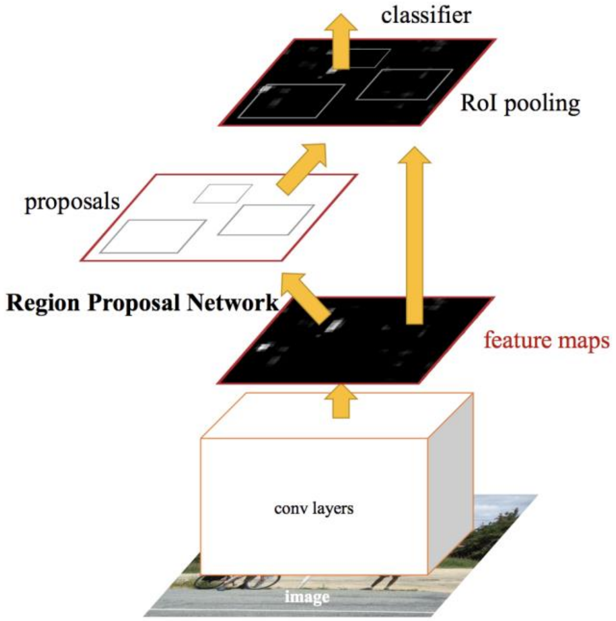
It makes use of the Identity Connection, which helps to protect the network from vanishing gradient problem.

It uses bottleneck residual block design to increase the performance of the network



# Computer Vision: Today in Object Detection

- Faster R-CNN (2015)




- YOLO (2015)