CS5242 : Neural Networks and Deep Learning

Lecture 7: Convolutional Neural Networks Introduction

Semester 1 2021/22

Xavier Bresson

https://twitter.com/xbresson

Department of Computer Science
National University of Singapore (NUS)

Outline

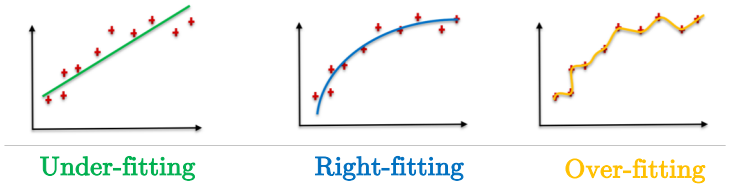
- Data structures
- Local reception fields
- Modeling of hierarchical organization
- One-layer convolutional neural network
- Paradigm shift in computer vision

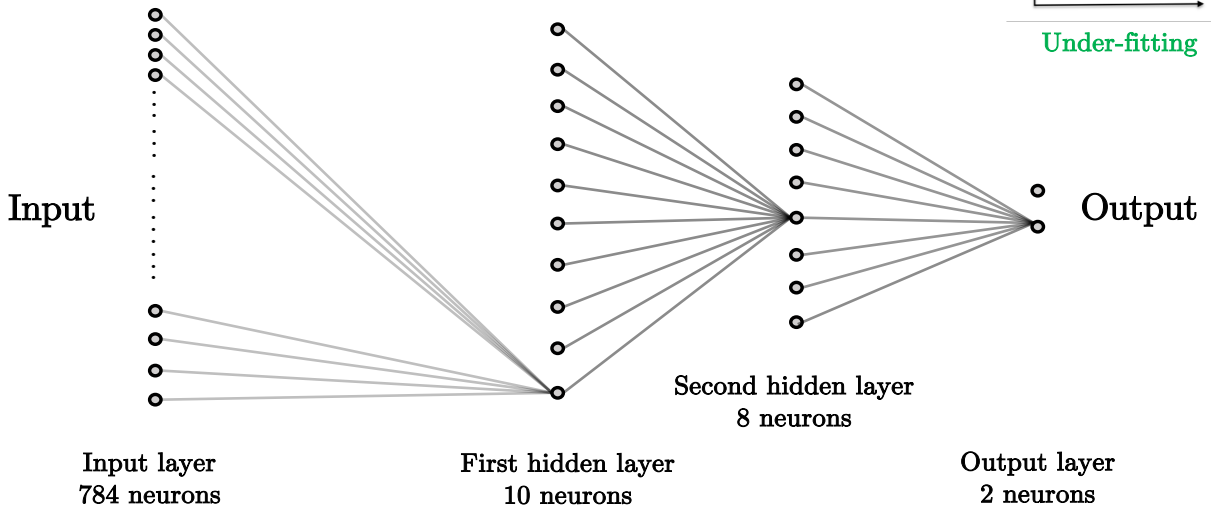
Outline

- Data structures
- Local reception fields
- Modeling of hierarchical organization
- One-layer convolutional neural network
- Paradigm shift in computer vision

Motivation

- MLP (a.k.a. Fully connected networks) :
  - MLP networks do not consider any specific structure of data, only pattern matching.
  - Theoretically, they can learn anything (universal approximation theorem), but they are practically very hard to train (too much parameters, too long).
  - They are pruned to overfitting and do not generalize well.





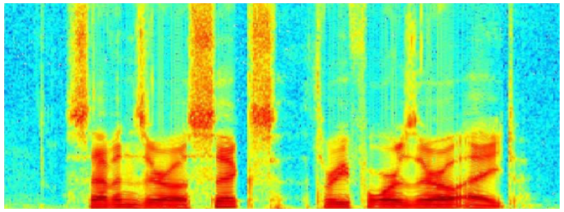
MLP/Fully connected networks

Data structures

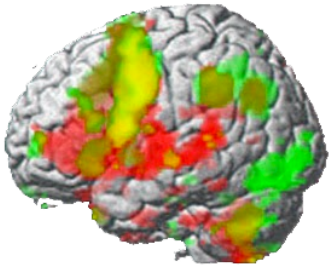
- Observe that data are not random, but exhibit special structures to find and to leverage.
- Invariance/symmetry refer to common and meaningful data structure.
- Identifying fundamental, minimal and universal data structures, that can be encoded into a large number of layers, is the best way to design neural networks.

Image

Video



Spectogram



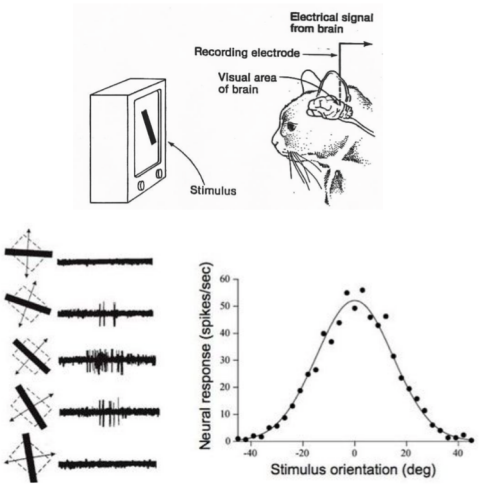
fMRI

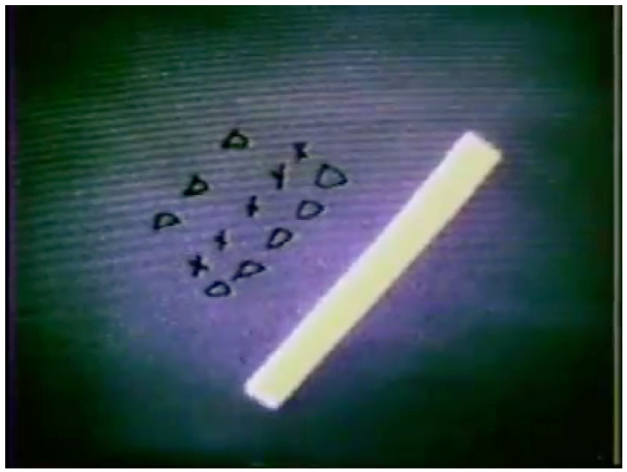
Outline

- Data structures
- Local reception fields
- Modeling of hierarchical organization
- One-layer convolutional neural network
- Paradigm shift in computer vision

Biological reception fields

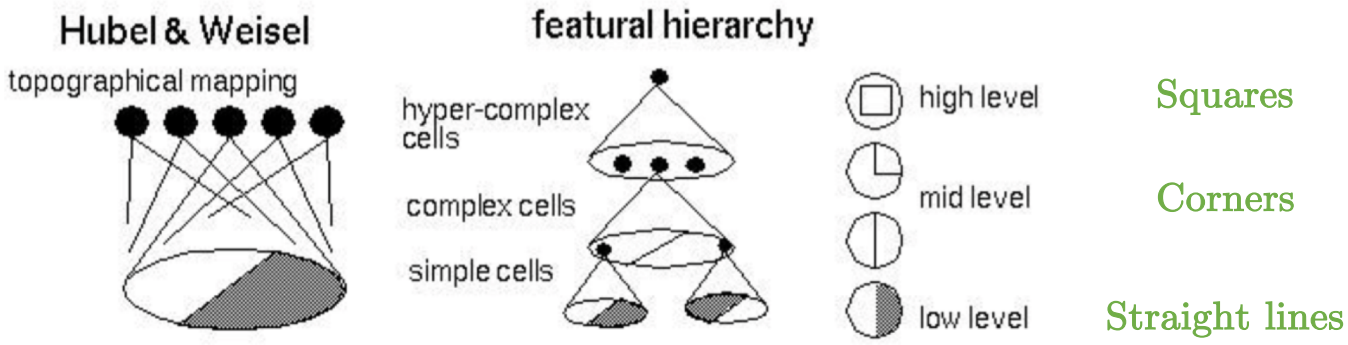
- Hubel and Wiesel 1959 :
  - Nobel Prize in Medicine for the understanding of the primary visual cortex system (first 2 layers):
  - Visual system is composed of receptive fields called V1 cells that are composed of neurons that activate depending on the local orientation.





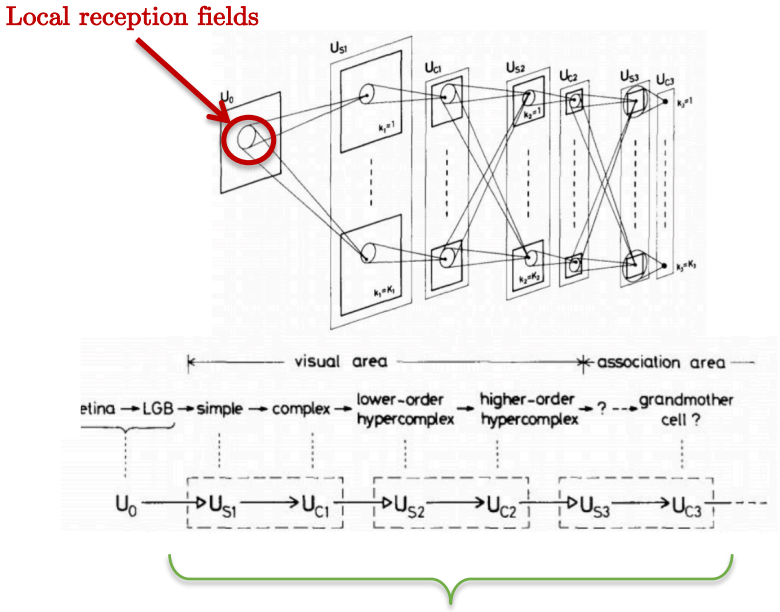
Biological reception fields

- The second layer of the visual cortex takes and composes outputs of V1 neurons.
  - The second layer neurons are called V2 cells.
  - This forms a hierarchical organization.



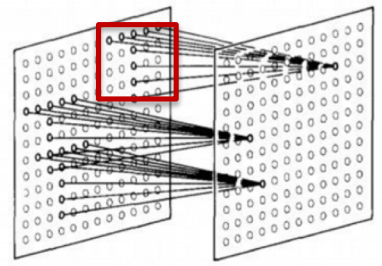
Neocognitron

- First algorithm that implements Huber-Wiesel visual recognition system [Fukushima, 1987] :
  - Introduction of concepts of local features to model biological reception fields.
  - Cascade simple and complex cells (V1 and V2 cells) to form hierarchical organization of features
  - Precursor of convolutional neural networks [LeCun, Bottou, Bengio, Haffner, 1998].



The importance of local features

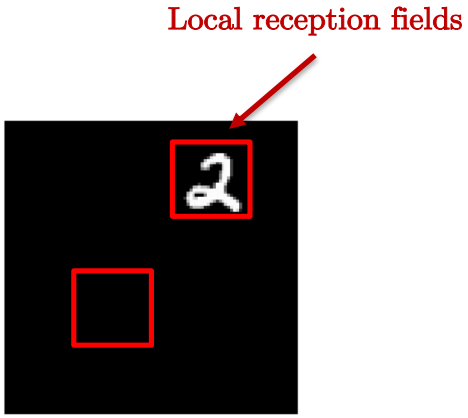
- Local features model the biological local reception fields :



V1 layer V2 layer

Modeling

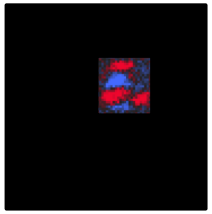
$\Rightarrow$



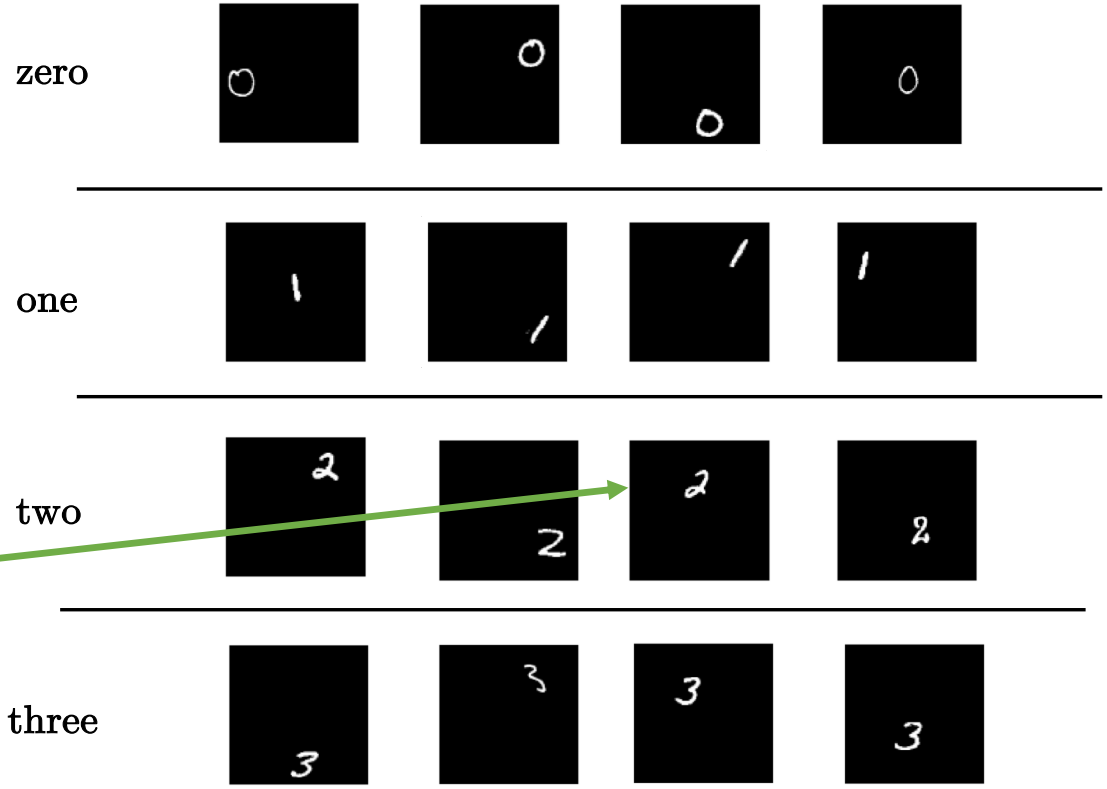
- Why local features are important for image recognition?

The importance of local features

- What if images are not well centered?
  - Suppose the handwritten digits are not in the center of the image.
  - It is not possible to find a single template for all the two's !



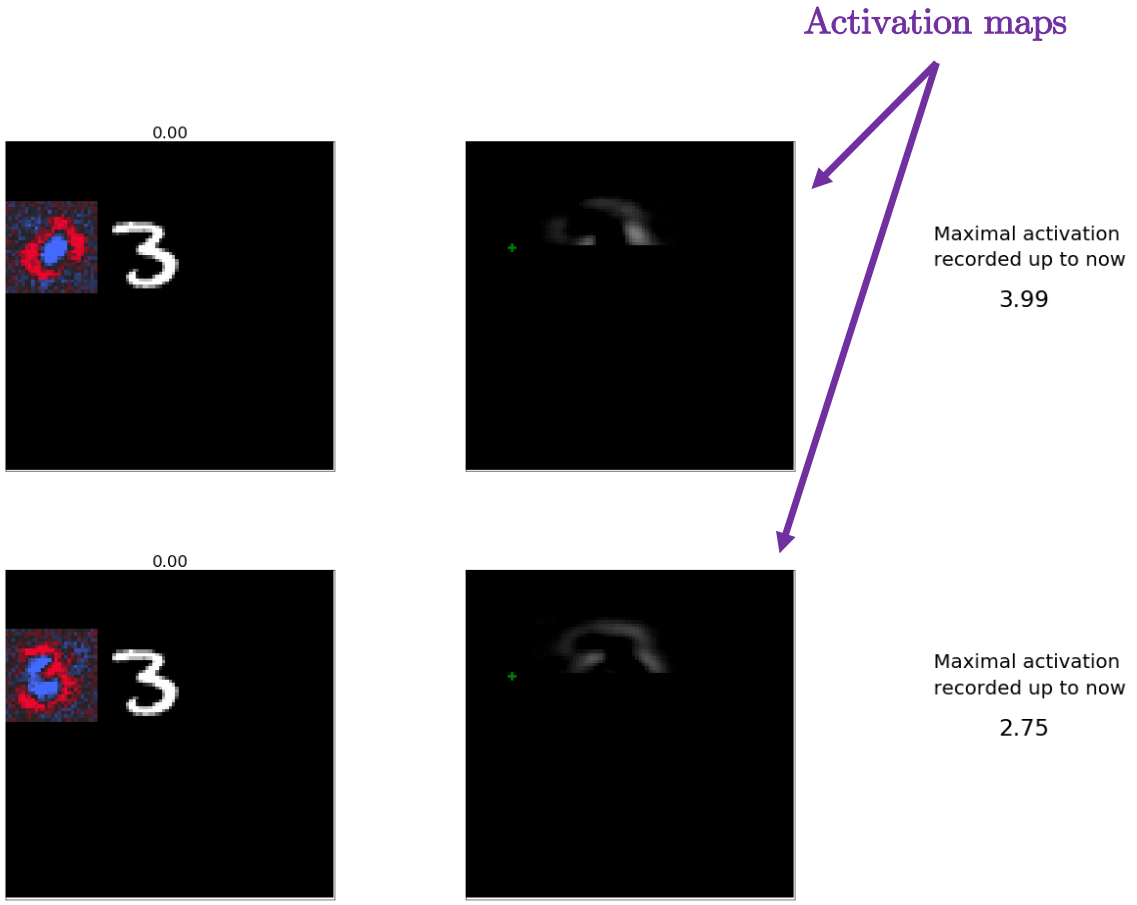
This template will only be useful for this image.



  - It is intractable to design a pattern for all locations in the image.

Detecting local features

- Sliding templates :
  - Define small templates and scan them across the image domain.
  - This operation is known as convolution.

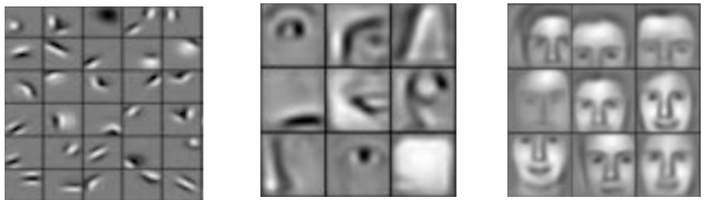


Outline

- Data structures
- Local reception fields
- Modeling of hierarchical organization
- One-layer convolutional neural network
- Paradigm shift in computer vision

Hierarchical features

- Local feature patterns can be composed to form abstract complex patterns :



Layer 1 Layer 2 Layer 3 Layer 4

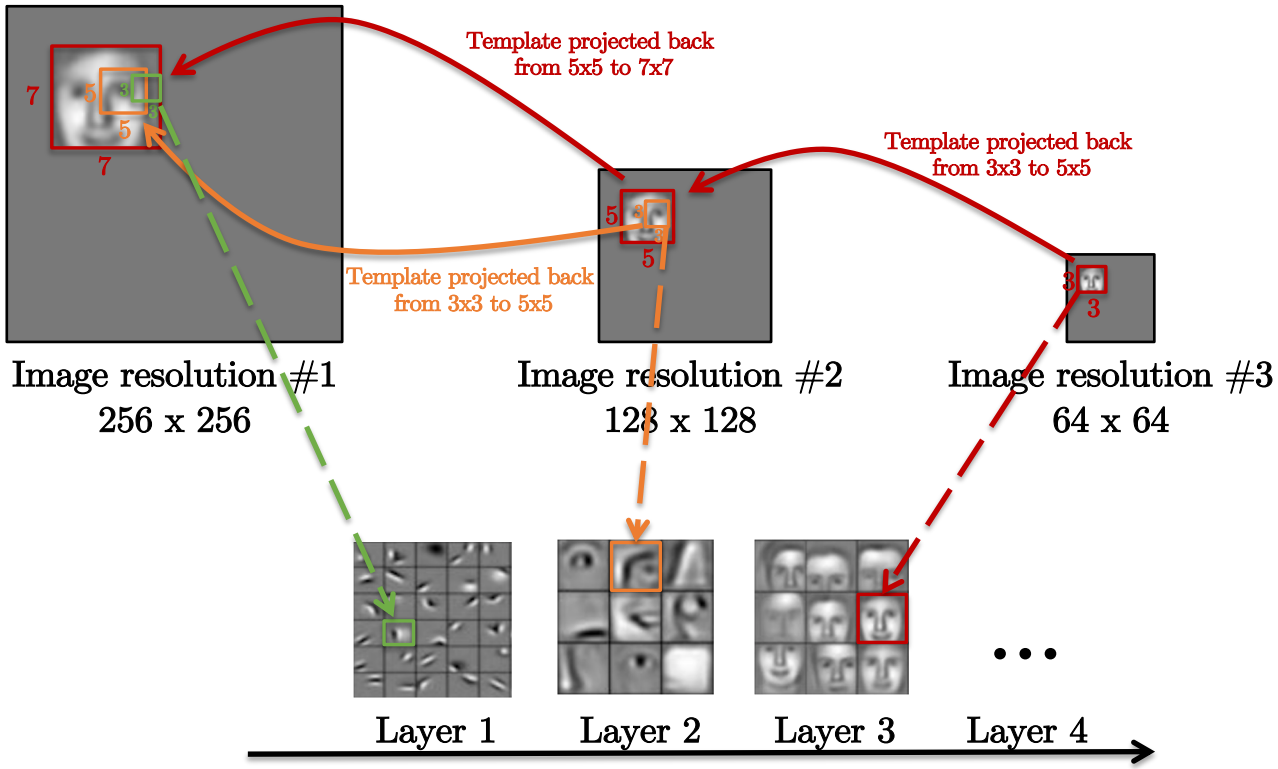
Deep/hierarchical features

(simple to abstract concepts)

- Image data is compositional :
  - It is formed from hierarchical local stationary patterns.

Hierarchical features

- How to extract hierarchical patterns (low-level to high-level features)?
  - Hierarchical patterns are multi-scale features.
    - They are captured by 3x3 templates at each image resolution :



We will only learn 3x3 templates !

At Image resolution #1 : 3 3

At Image resolution #2 : 3 3

At Image resolution #3 : 3 3

Hierarchical features

- How to change image resolution?
  - Down-sampling and
  - Pooling (for examples max pooling, average pooling)

3	3	0	1	1	1
7	2	1	1	5	2
7	9	1	0	5	6
5	6	1	0	3	2
9	7	0	0	0	0
4	7	0	0	0	1

Down-sampling
+ 2x2 max pooling

7	1	5
9	1	6
9	0	1

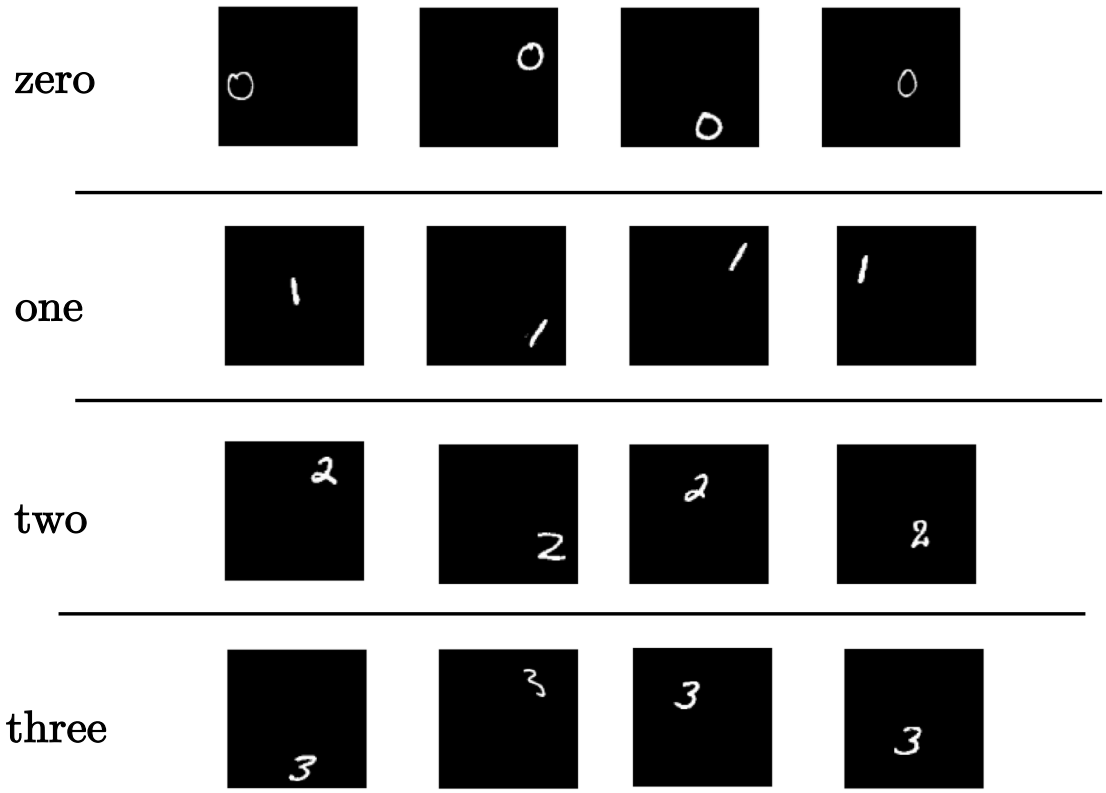
Outline

- Data structures
- Local reception fields
- Modeling of hierarchical organization
- **One-layer convolutional neural network**
- Paradigm shift in computer vision

One-layer ConvNet

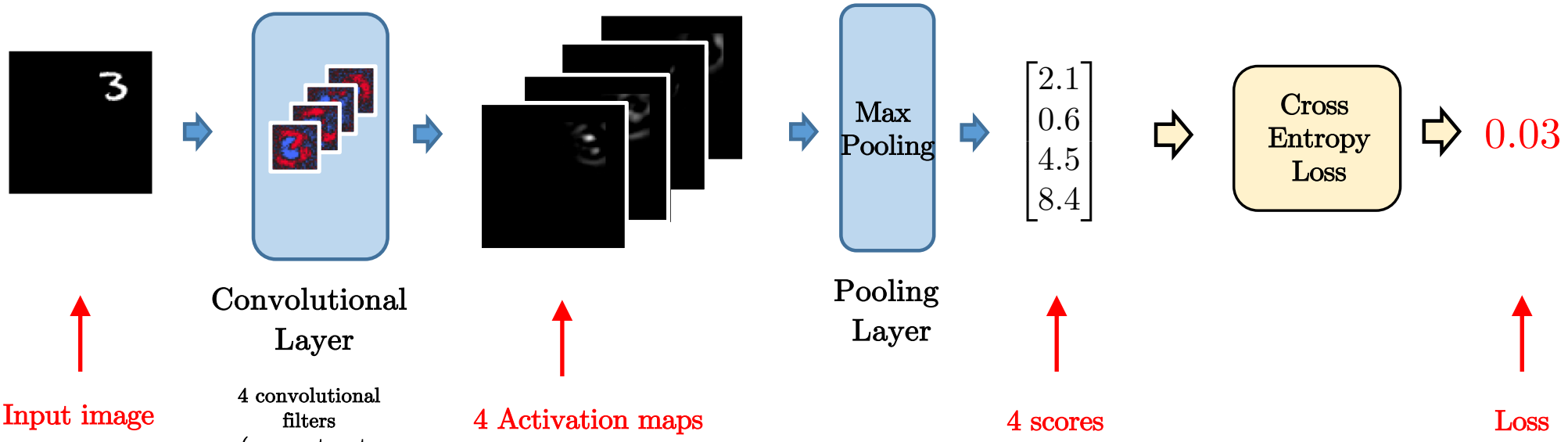
- Assume there are only four classes :

  (0,1,2,3)



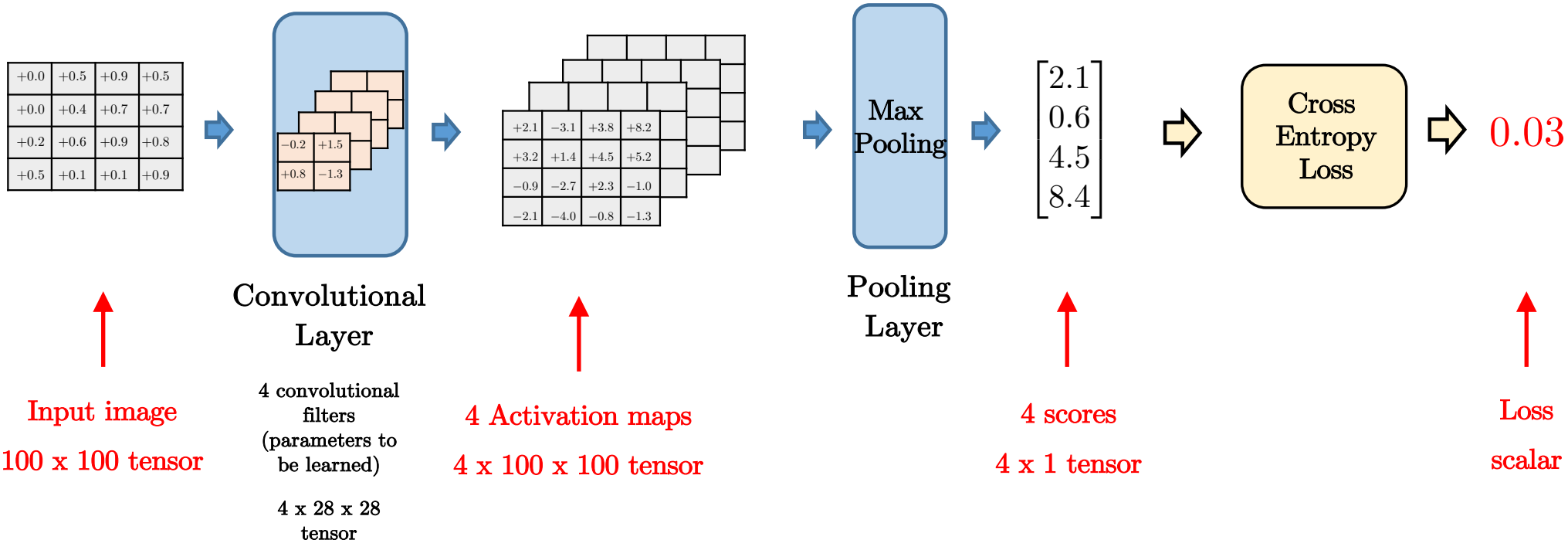
One-layer ConvNet

- Architecture : 1 CL + MP + Loss



One-layer ConvNet

- Tensor representation :



Convolution

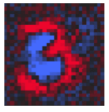
- Exact formula for the convolution layer:
  - Each filter generates an activation map.

$$(0)(-1) + (3)(1) + (2)(2) + (0)(1) = 7$$

$$\begin{bmatrix} 0 & 3 & 0 & 1 \\ 2 & 0 & 1 & 2 \\ 3 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 \end{bmatrix} \star \begin{bmatrix} -1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 7 & -2 & 5 \\ 4 & 1 & 3 \\ -2 & 4 & 6 \end{bmatrix}$$

Input image

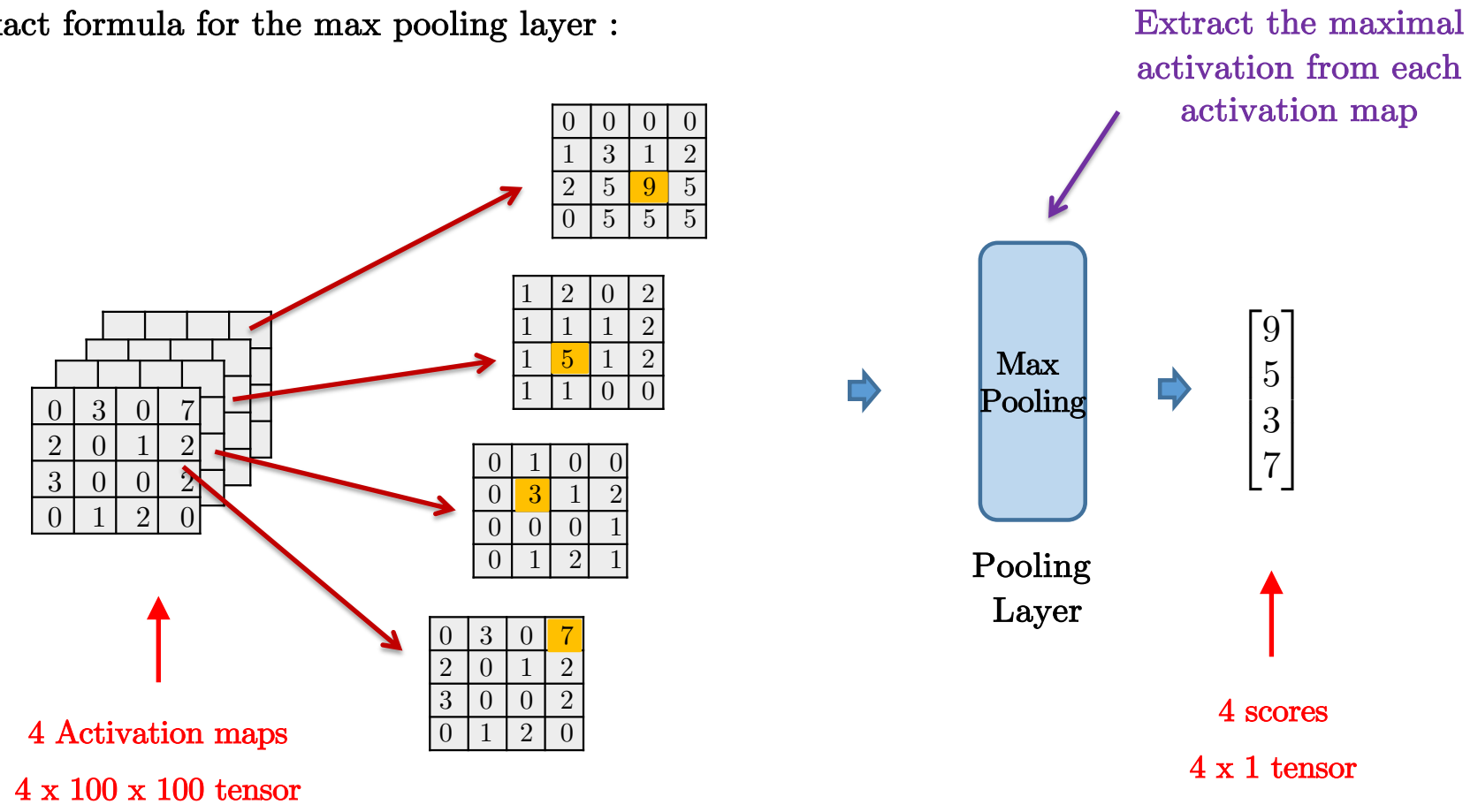
One filter



One activation map

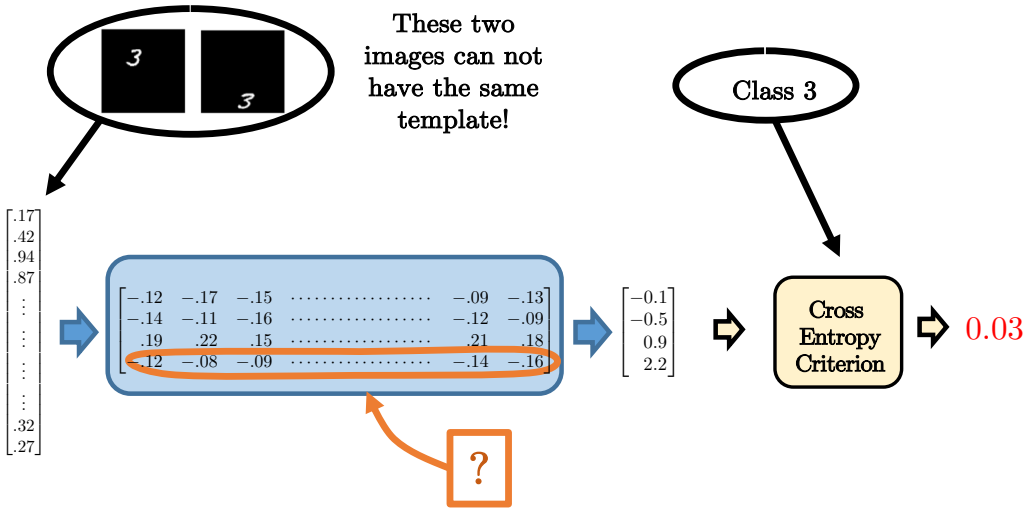
One-layer ConvNet

- Exact formula for the max pooling layer :

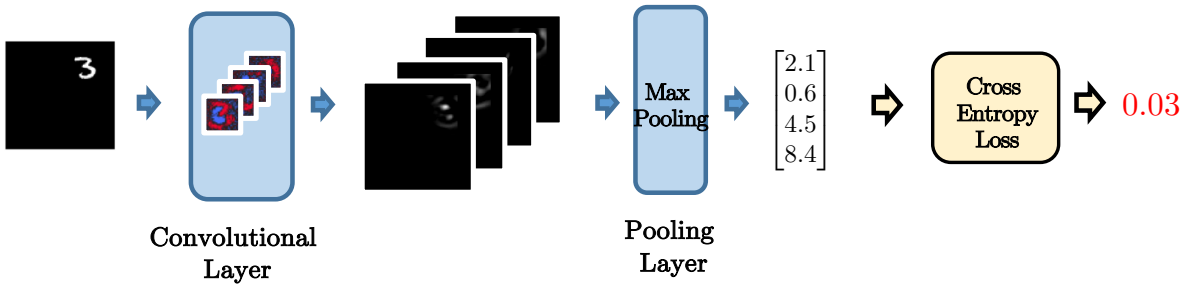


MLP vs ConvNets

- A regular one-layer neural net can not find one template that matches all the threes:



- But a one-layer convnet can find one template that can do it with the convolution layer:



Outline

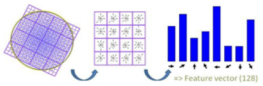
- Data structures
- Local reception fields
- Modeling of hierarchical organization
- One-layer convolutional neural network
- **Paradigm shift in computer vision**

From handcrafted to learned features

- Handcrafted approaches:

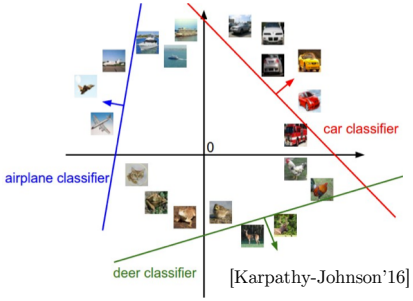
Input image

Handcrafted Features



SIFT, Lowe, 1999
(most cited paper in CV)

SVM classifier



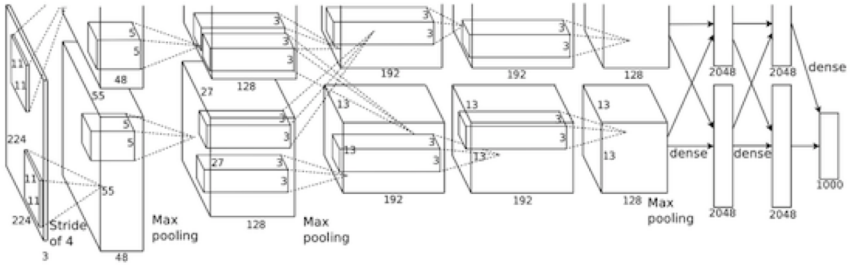
[Karpathy-Johnson'16]

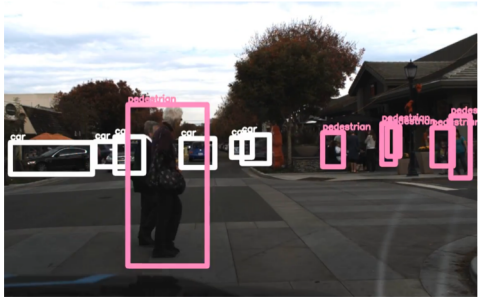
Before deep learning

- ConvNet approaches:
  - AlexNet made a breakthrough in 2012 in image recognition.

Input image

Learned features + Classifier





Deep learning

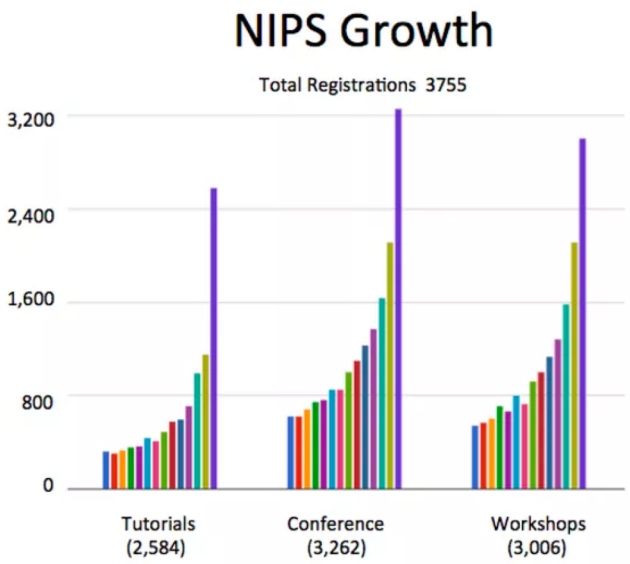
ConvNets

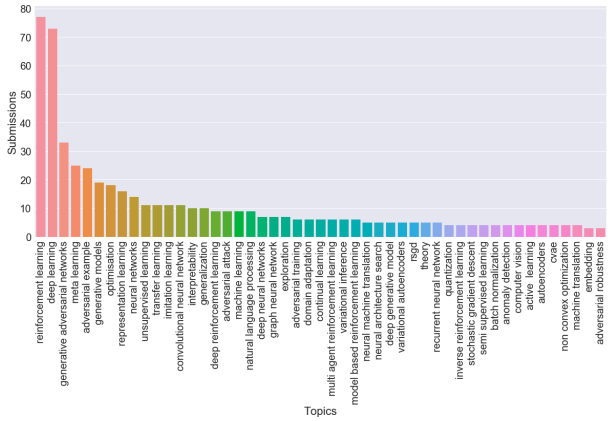
- Deep ConvNets:
  - Stack multiple convolutional+pooling layers
    - LeNet5, AlexNet, VGG, Inception, ResNet, etc

- ConvNets are extremely efficient at extracting compositional statistic patterns in large-scale and high-dimensional image datasets.

- Almost the entire ML, CV, NLP communities have moved to deep learning techniques and industry is investing considerably in deep learning.

- ICLR'19 topics

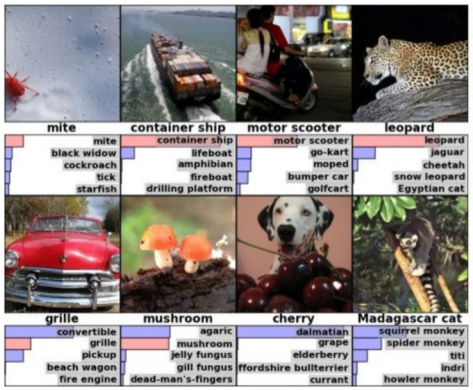




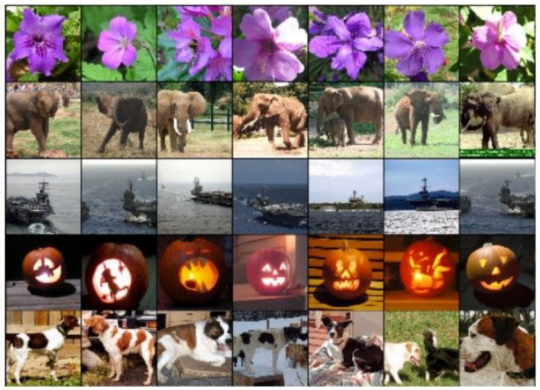
ConvNet applications

- ConvNets are today ubiquitous in computer vision !

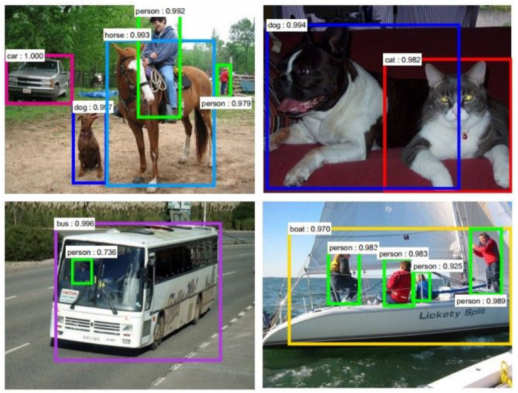
Classification



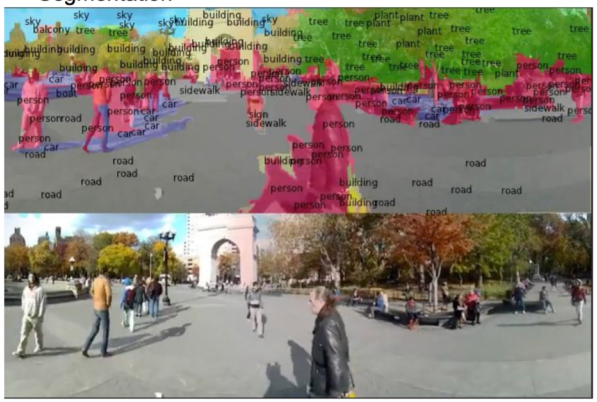
Retrieval



Detection

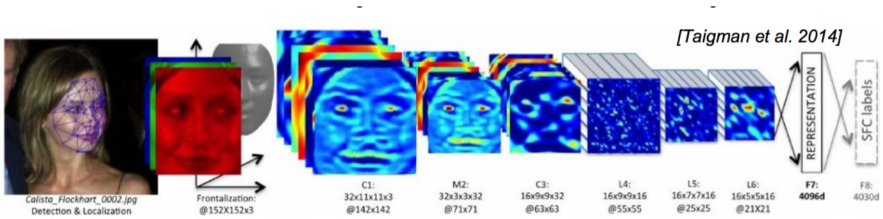


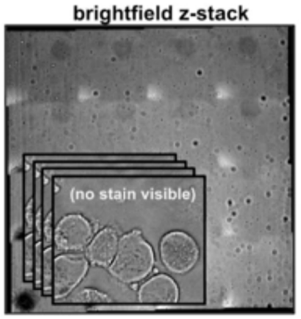
Segmentation

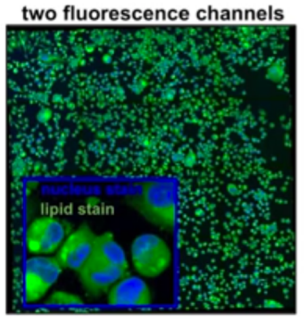


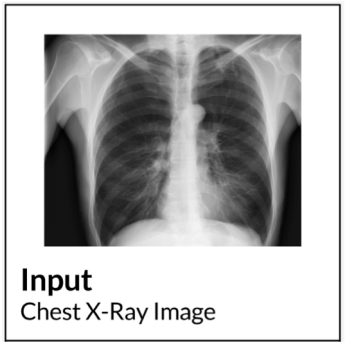
ConvNet applications

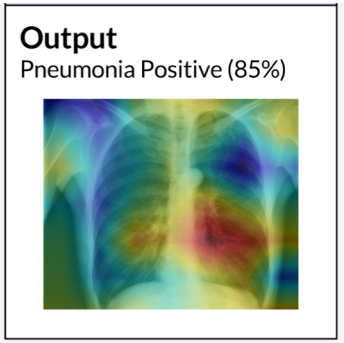
- ConvNets are today ubiquitous in computer vision !

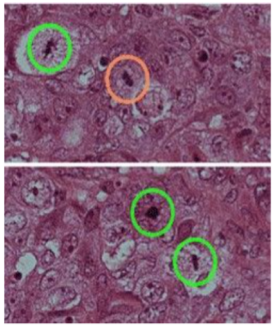


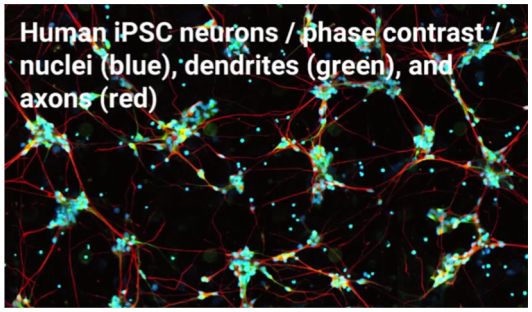








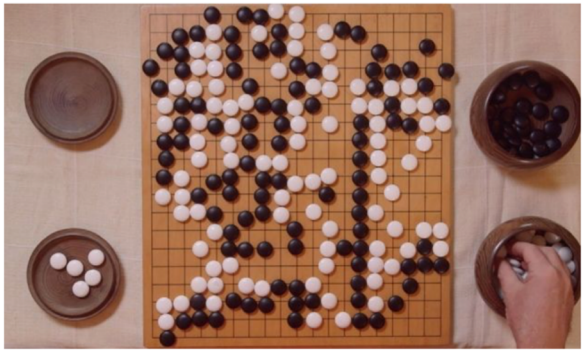


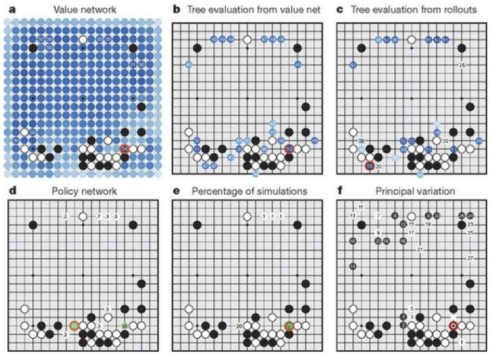


ConvNet applications

- ConvNets are today ubiquitous in computer vision !

self-driving cars





Questions?