Australian National University

Deep Learning Tutorial

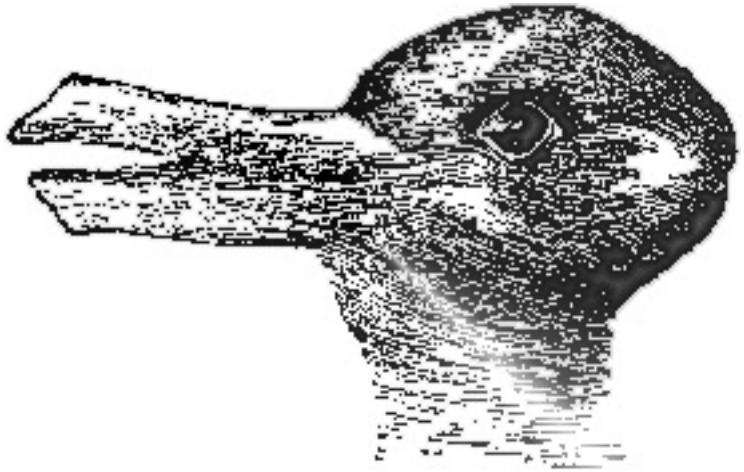
Ibrahim Radwan

What you will learn

- What are the Convolutional Neural Networks?
- Convolution operation.
- Rectifier Linear Units (ReLUs)
- Pooling
- Flattening
- Full connection
- Summary
- Softmax and Cross-entropy

Convolutional Neural Networks

How does the brain work and classify an image?

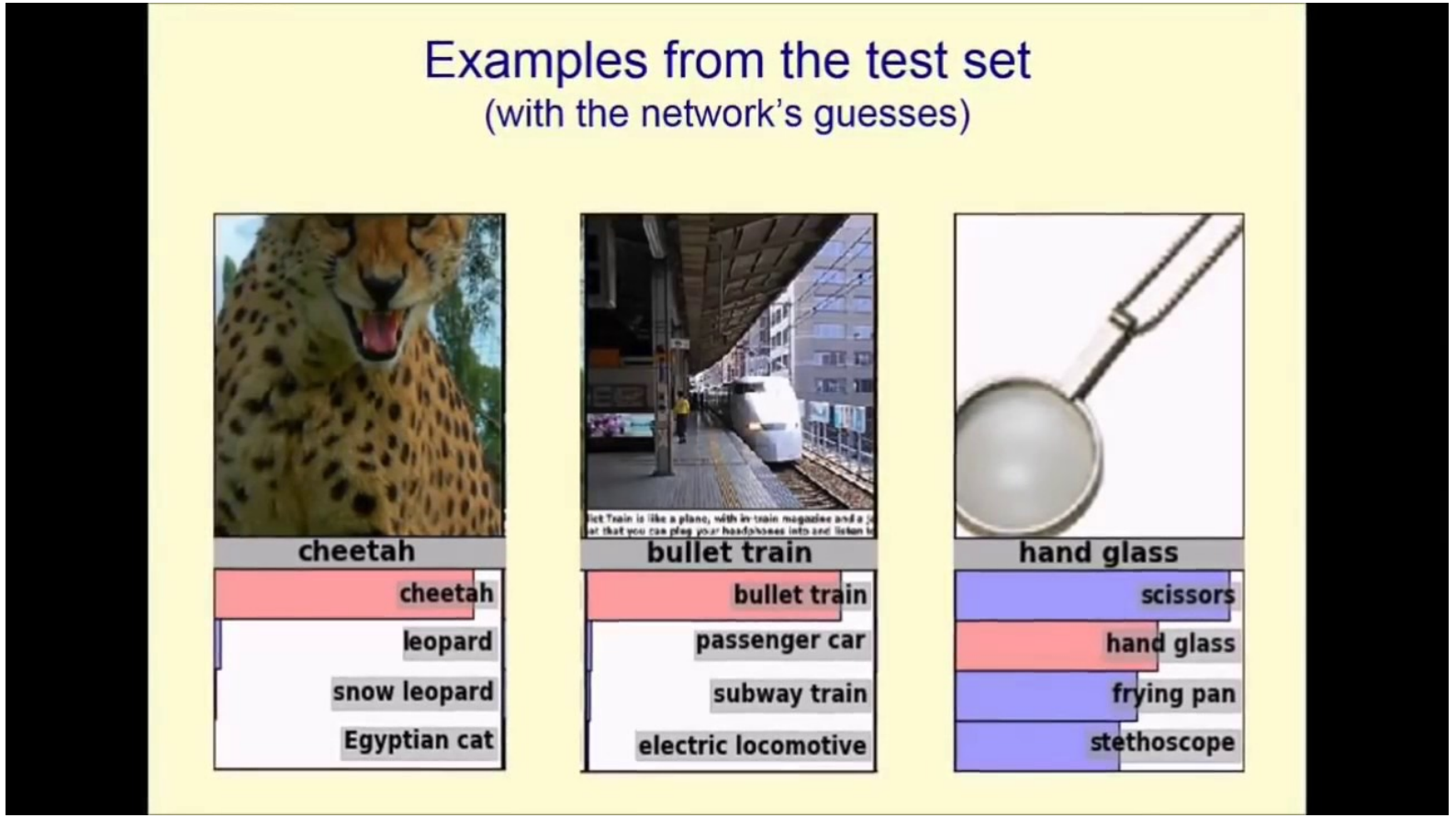


Duck or Rabbit?

The answer is the features and the structure of the image.

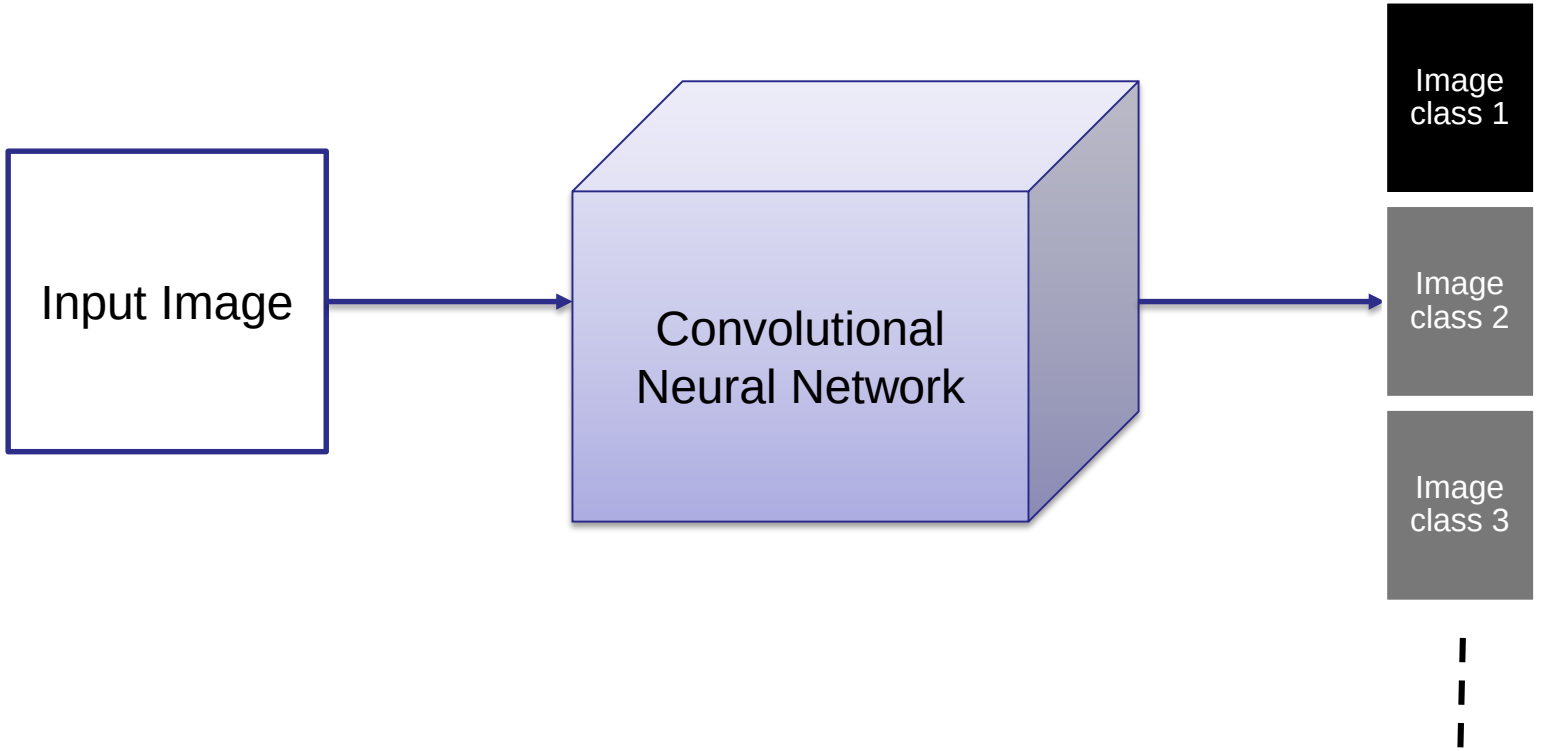
Picture from Wikipedia

Convolutional Neural Networks



Picture from Geoffrey Hinton *presentation on YouTube*

Convolutional Neural Networks



What does the computer see?

B / W Image

P1	P2
P3	P4

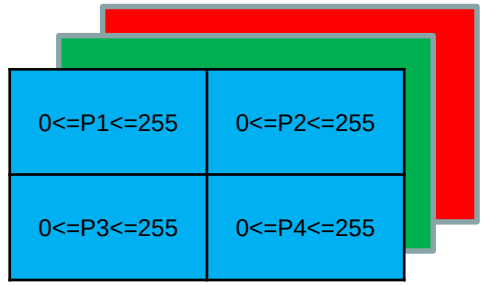
2D array

0<=P1<=255	0<=P2<=255
0<=P3<=255	0<=P4<=255

Coloured Image

P1	P2
P3	P4

3D array



Convolution

$$(f * g)(t) \int_{-\infty}^{\infty} f(\tau) g(t - \tau) dt$$

Convolution

0	0	0	0	0	0	0
0	1	0	0	0	1	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	1	0	0	0	1	0
0	0	1	1	1	0	0
0	0	0	0	0	0	0

Input image

X

0	0	1
1	0	0
0	1	1

Feature detector

=

0	1	0	0	0
0	1	1	1	0
1	0	1	2	1
1	4	2	1	0
0	0	1	2	1

Feature map

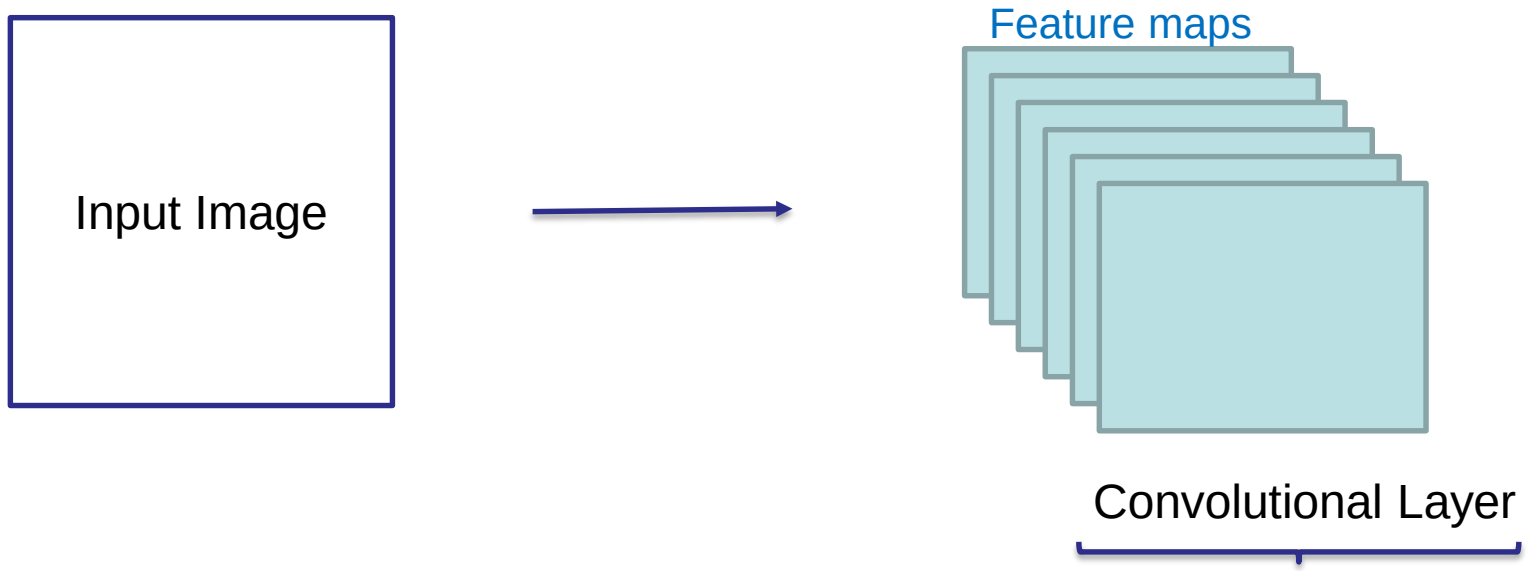
Stride = 1
Kernel size = 3x3

Convolution

- Size of the image is reduced.
- Are we losing information? Yes, but we concentrate more on the important parts.
- With many kernels/filters, we could encode many features from the images.
- The weights of the kernels are initialized randomly.

Recap

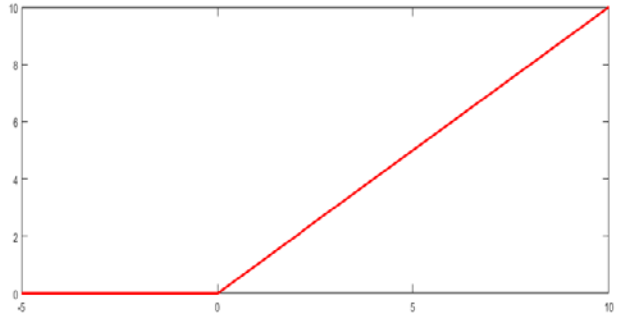
- Convolutional layers are extracting the features of the objects in the image while it keeps the spatial relationships between the pixels



Rectifier Linear Units (ReLUs)

Input Image

Feature maps



Convolutional Layer

$$\phi(x) = \max(x, 0)$$

This is to motivate the model to encode the non-linearity structure of the inputs

Pooling

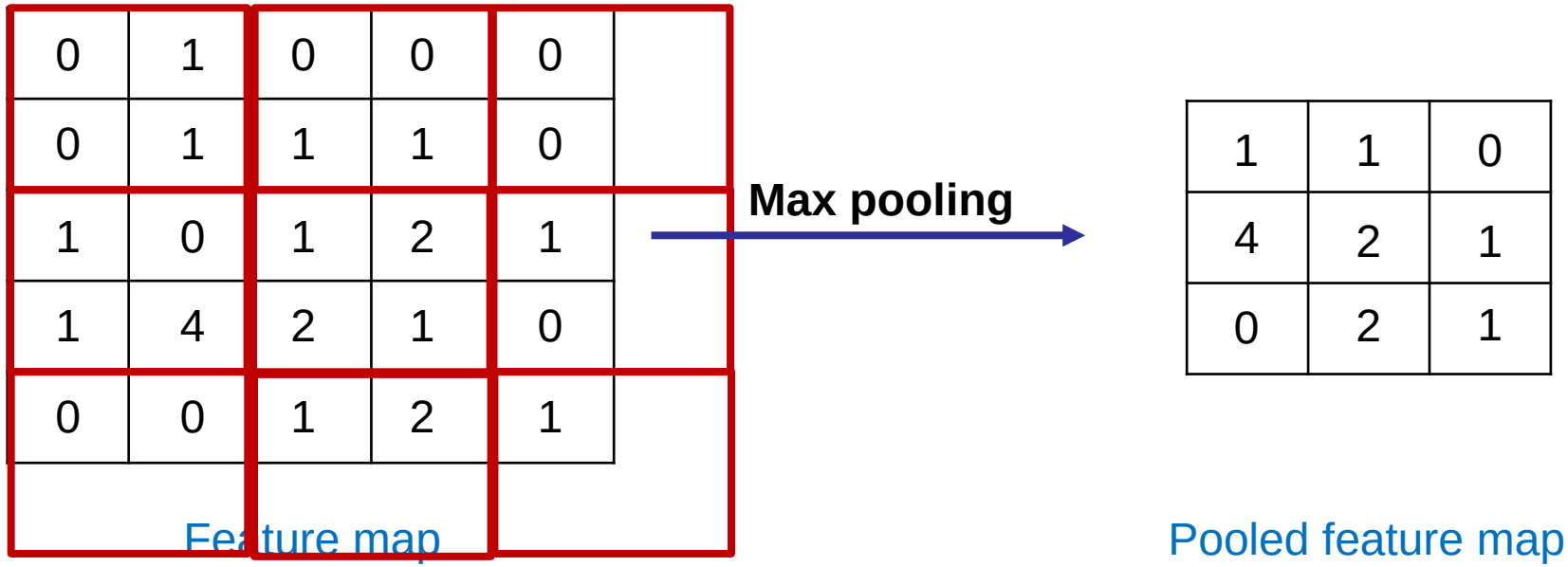
Different poses
Different rotation
Different lighting conditions
etc.

CNN should be spatial invariant

Pooling

- Max-pooling
- Average-pooling
- Sum-pooling

Pooling


0 1 0 0 0
0 1 1 1 0
1 0 1 2 1
1 4 2 1 0
0 0 1 2 1
Max pooling
1 1 0
4 2 1
0 2 1
Feature map
Pooled feature map

Pooling

- If there is a rotation, for example, where 4 in the box to left, it still get the important feature (4)
- More over, we reduce the size of feature map
- We reduce the parameter size to , for example, 75% removed. This reduce over-fitting

Pooling

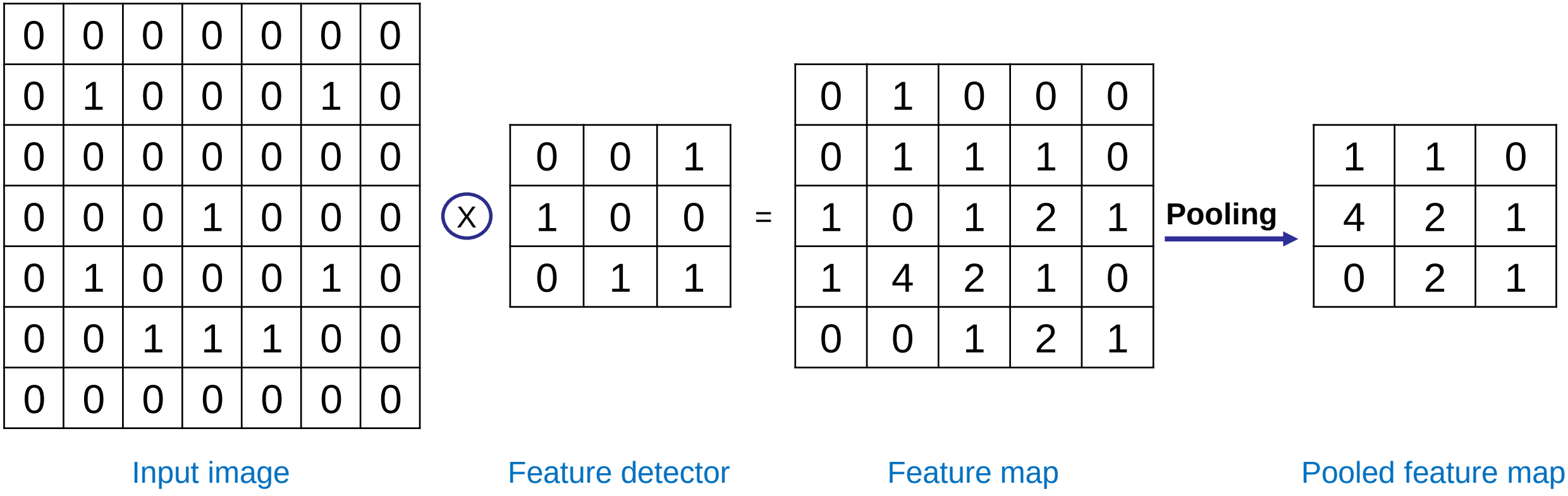
- Mean Pooling/ Average Pooling
  - Sub-sampling is just taking the average

Additional reading

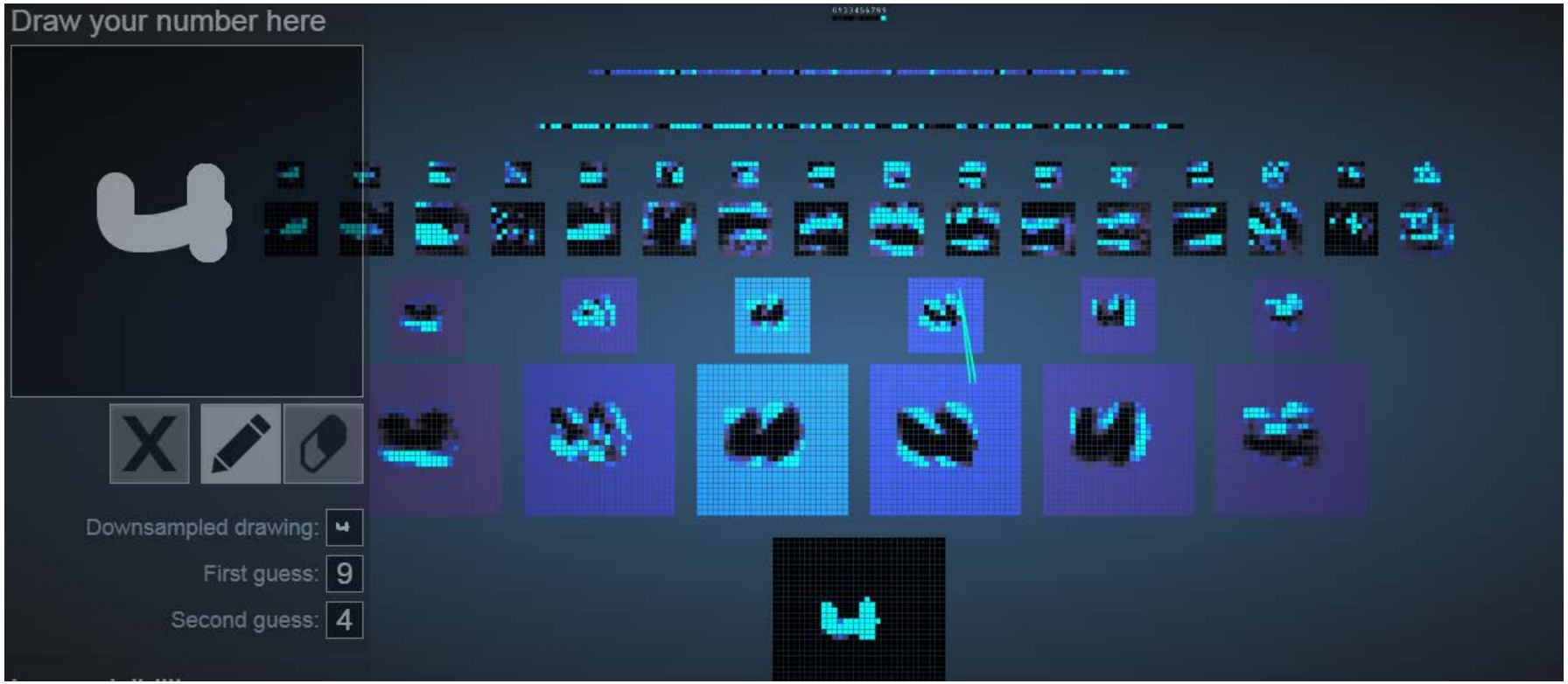
- Evaluation of Pooling Operations in Convolution Architectures for Object for Object Recognition.

  By Dominik Scherer *et al* (2010)

Recap


0 0 0 0 0 0 0
0 1 0 0 0 1 0
0 0 0 0 0 0 0
0 0 0 1 0 0 0
0 1 0 0 0 1 0
0 0 1 1 1 0 0
0 0 0 0 0 0 0
X
0 0 1
1 0 0
0 1 1
=
0 1 0 0 0
0 1 1 1 0
1 0 1 2 1
1 4 2 1 0
0 0 1 2 1
Pooling
1 1 0
4 2 1
0 2 1
Input image
Feature detector
Feature map
Pooled feature map

Example



www.scs.ryerson.ca/~aharley/vis/conv/flat.html

Flattening

1	1	0
4	2	1
0	2	1

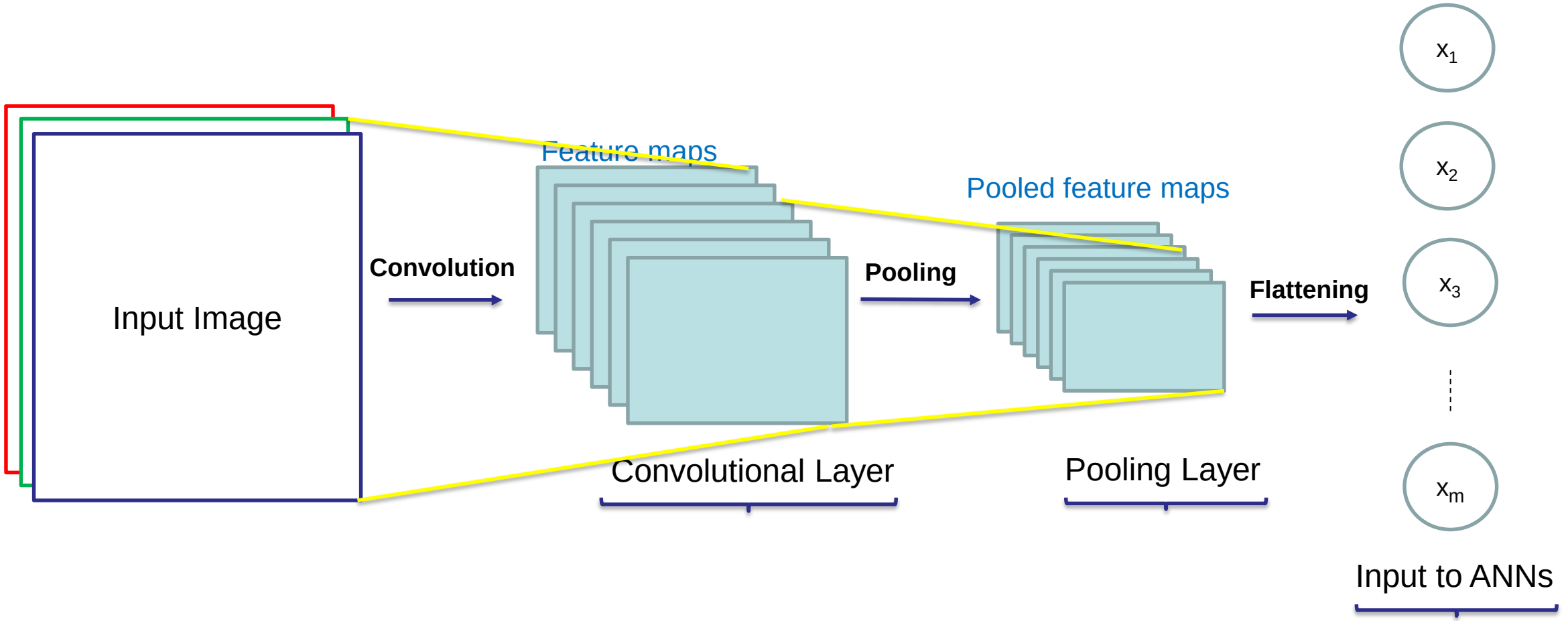
Pooled feature map

Flattening

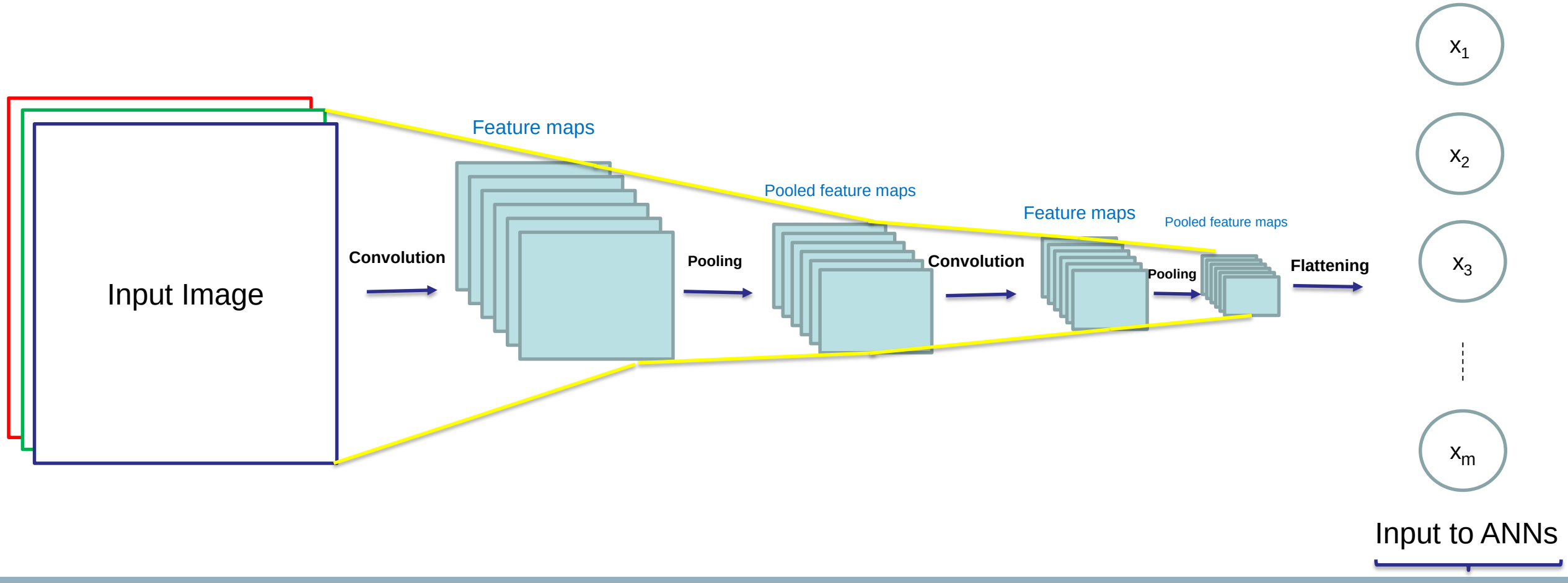
1
1
0
4
2
1
0
2
1

Input to ANNs

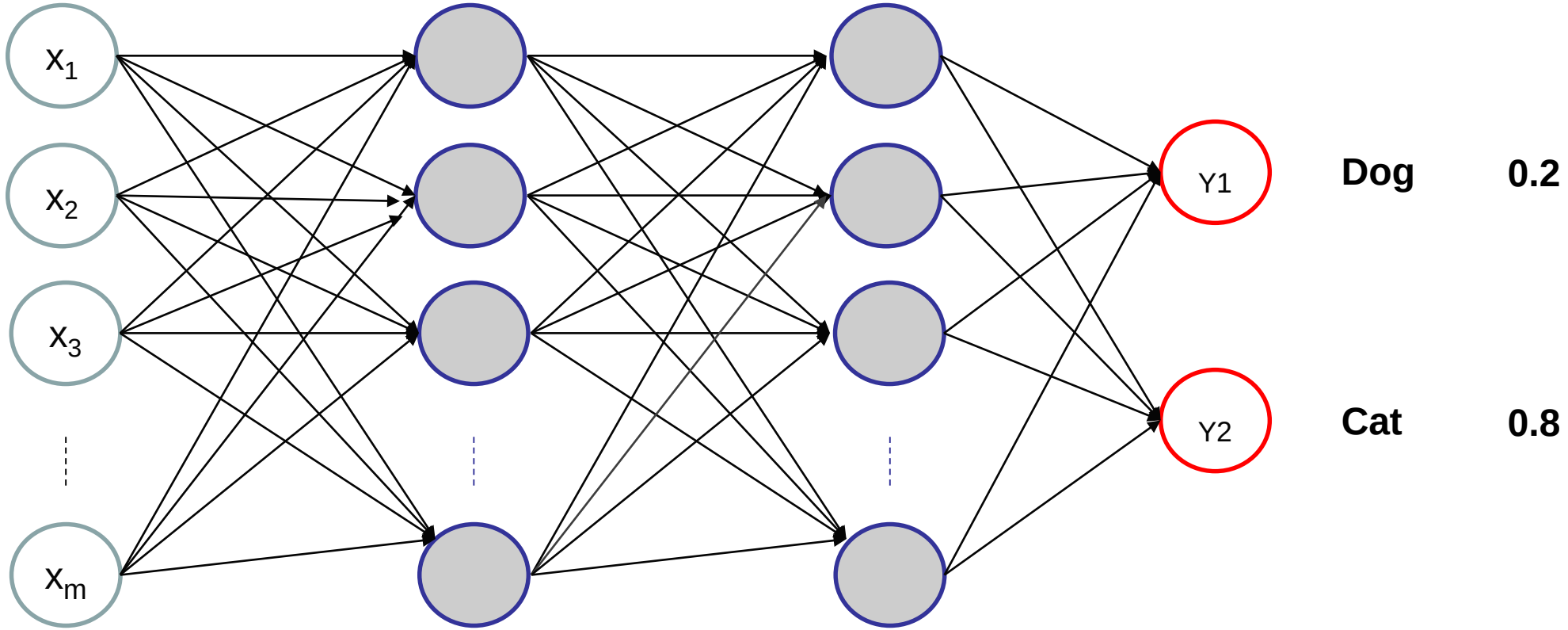
Recap


Input Image
Convolution
Feature maps
Pooling
Pooled feature maps
Flattening
x_1
x_2
x_3
x_m
Convolutional Layer
Pooling Layer
Input to ANNs

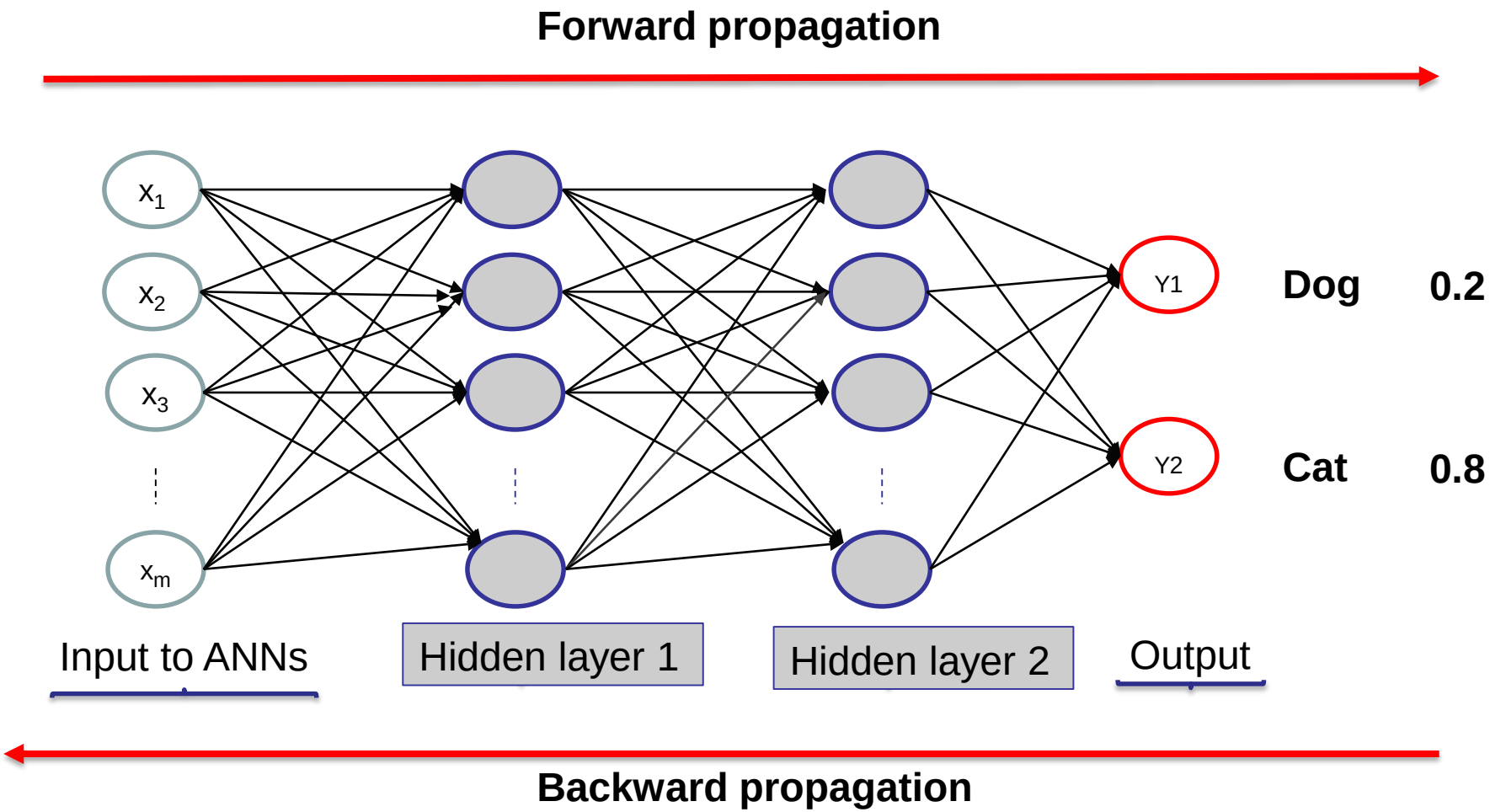
Recap


Input Image
Convolution
Feature maps
Pooling
Pooled feature maps
Convolution
Feature maps
Pooling
Pooled feature maps
Flattening
x_1
x_2
x_3
x_m
Input to ANNs

Full Connection



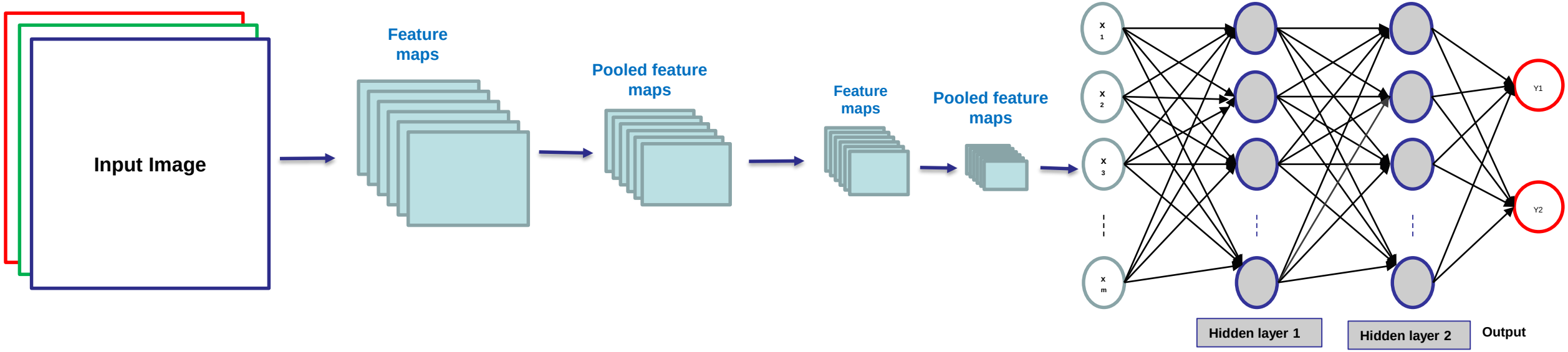
Training


Forward propagation
x_1
x_2
x_3
x_m
Y1
Y2
Dog 0.2
Cat 0.8
Input to ANNs
Hidden layer 1
Hidden layer 2
Output
Backward propagation

Training

- Two type of parameters are adjusted:
  - Fully connected layers' weights.
  - Values of the kernels/Filters.
- Please Note that in CNNs, the cost function is called "Loss-function".

Summary


Input Image
Feature maps
Pooled feature maps
Feature maps
Pooled feature maps
x 1
x 2
x 3
x m
Y1
Y2
Hidden layer 1
Hidden layer 2
Output

Additional reading

- Great blog by Adit Dishpande(2016)
  - "The 9 deep learning papers you need to know, (understanding CNNs, part3)".

Drop-out

- The main idea is to allow the network to ignore some units and their weights in either the visible or hidden layers for seek of generality.
- Drop-out reduces overfitting the model on the training data.
- 20% drop-out means 1 node from 5 will not be updated during the training.

Momentum

- Momentum is a technique that accelerates the gradient descent direction toward the minimum.

Let us Practice!