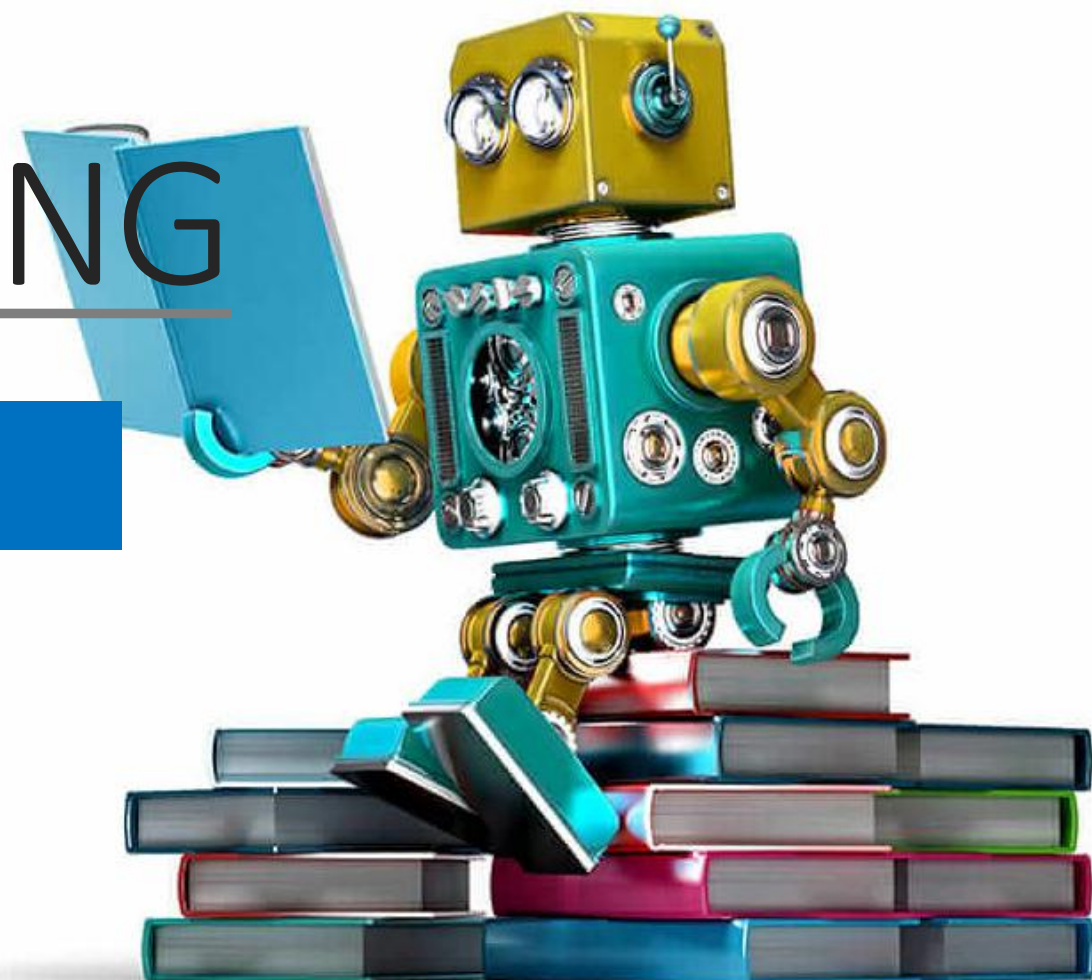


MACHINE LEARNING

LAB6 Neural Network

贾艳红 Jana

Email: jiayh@mail.sustech.edu.cn





Outline



- What can neural networks do?
- What are neural networks?
- What are deep neural networks?
- What are Convolutional neural networks?



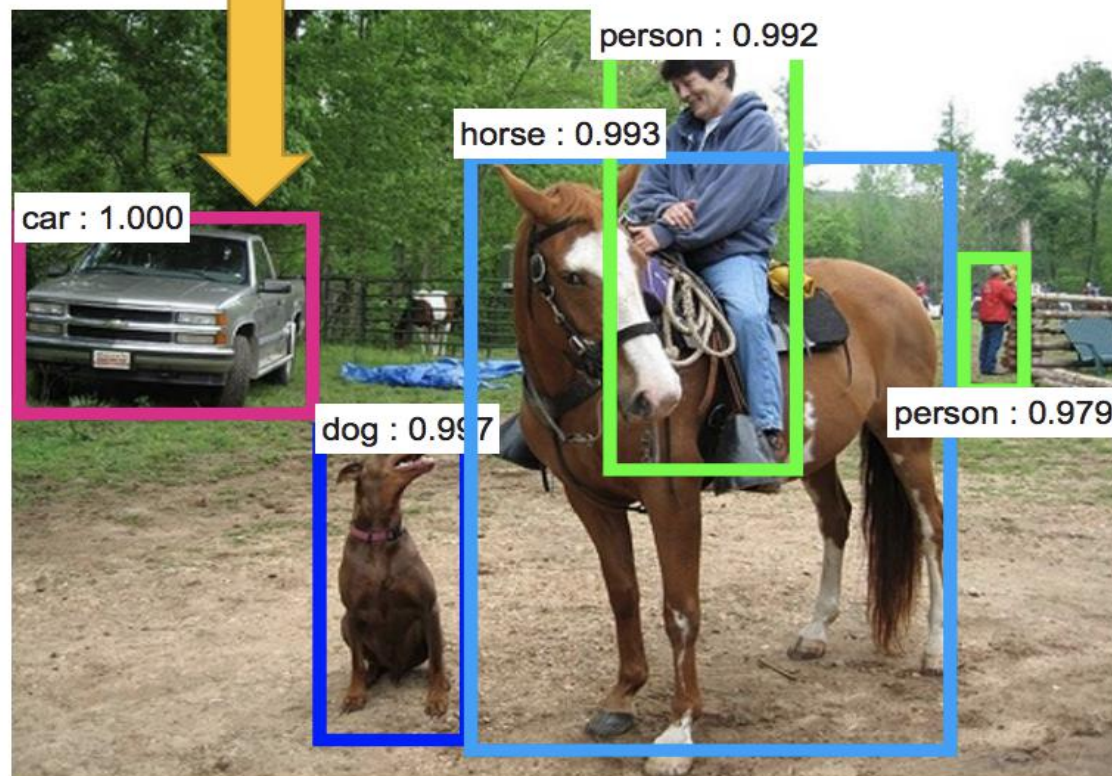
Object Detection



Object Detection = What, and Where

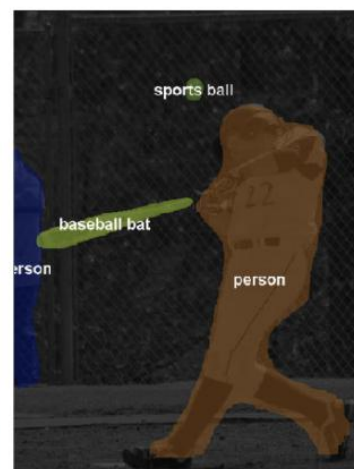
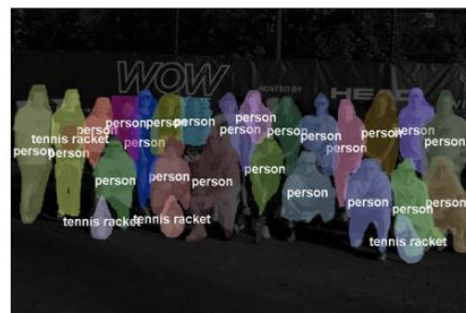
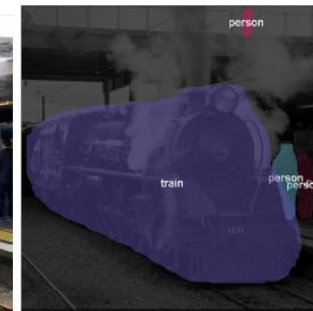
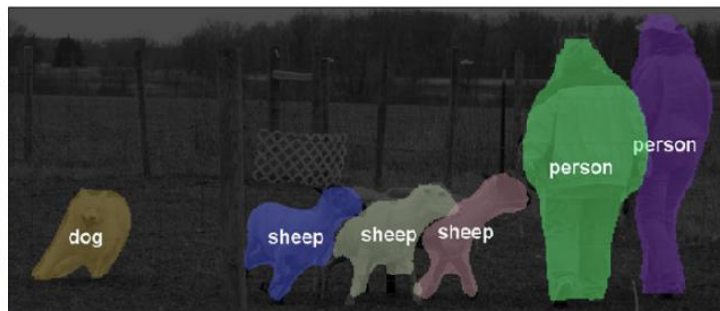
Recognition
What?

Localization
Where?



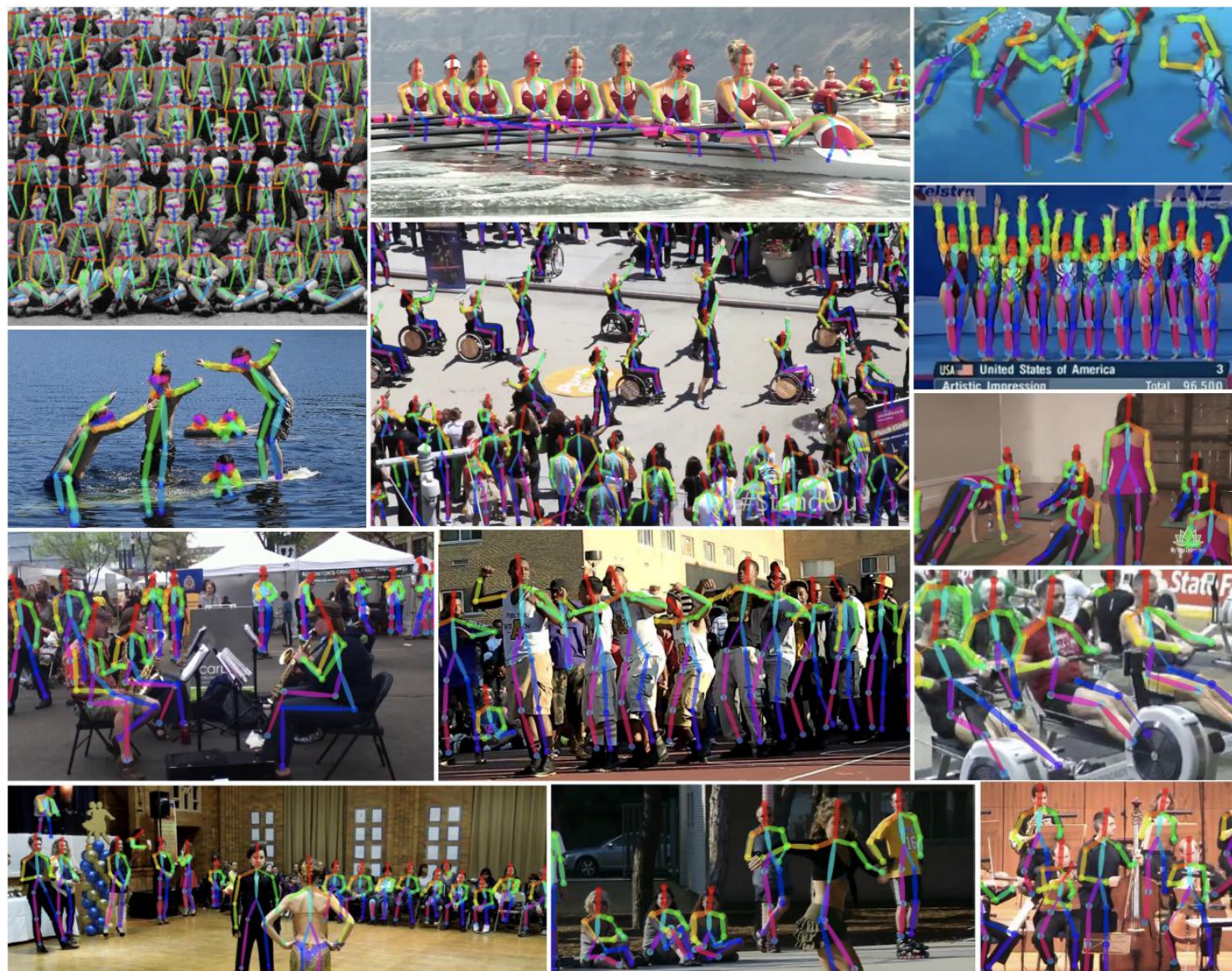


Object Segmentation



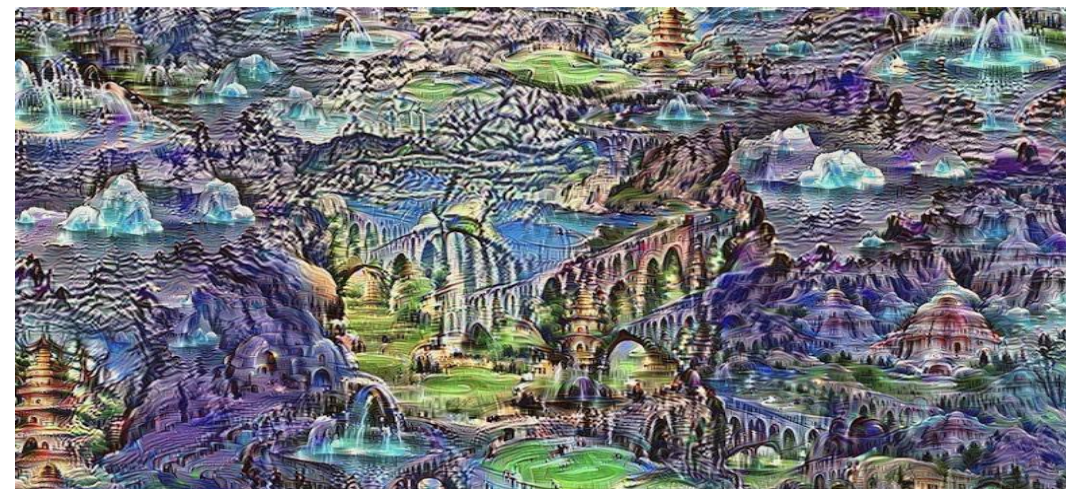


Pose Estimation





Art generation





Visual Question Answering



What color are her eyes?
What is the mustache made of?



Is this person expecting company?
What is just under the tree?



How many slices of pizza are there?
Is this a vegetarian pizza?



Does it appear to be rainy?
Does this person have 20/20 vision?



Image	Q: Who is behind the batter?	Q: What adorns the tops of the post?	Q: How many cameras are in the photo?
Multiple Choices	A: Catcher.	A: Gulls.	A: One.
	A: Umpire.	A: An eagle.	A: Two.
w/ Image	A: Fans.	A: A crown.	A: Three.
	A: Ball girl.	A: A pretty sign.	A: Four.

Image	H: Catcher. ✓	H: Gulls. ✓	H: Three. ✗
w/ Image	M: Umpire. ✗	M: Gulls. ✓	M: One. ✓
	H: Catcher. ✓	H: Gulls. ✓	H: One. ✓
w/ Image	M: Catcher. ✓	M: A crown. ✗	M: One. ✓



Image	Q: Why is there rope?	Q: What kind of stuffed animal is shown?	Q: What animal is being petted?
Multiple Choices	A: To tie up the boats.	A: Teddy Bear.	A: A sheep.
	A: To tie up horses.	A: Monkey.	A: Goat.
w/ Image	A: To hang people.	A: Tiger.	A: Alpaca.
	A: To hit tether balls.	A: Bunny rabbit.	A: Pig.

Image	H: To hit tether balls. ✗	H: Monkey. ✗	H: A sheep. ✓
w/ Image	M: To hang people. ✗	M: Teddy Bear. ✓	M: A sheep. ✓
	H: To tie up the boats. ✓	H: Teddy Bear. ✓	H: Goat. ✗
w/ Image	M: To hang people. ✗	M: Teddy Bear. ✓	M: A sheep. ✓



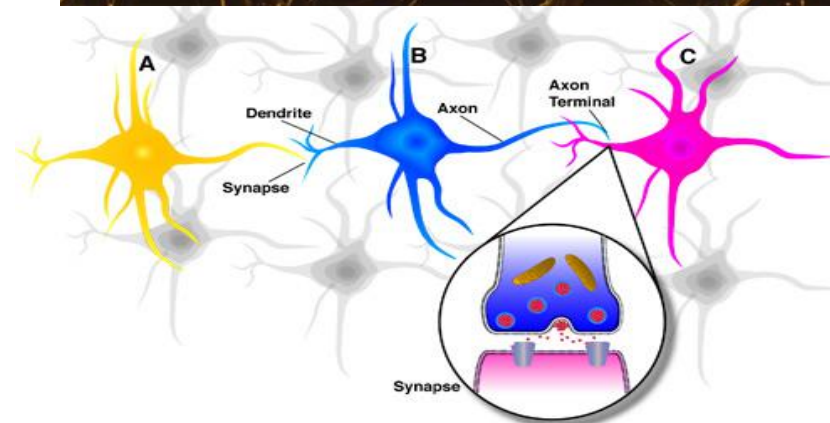
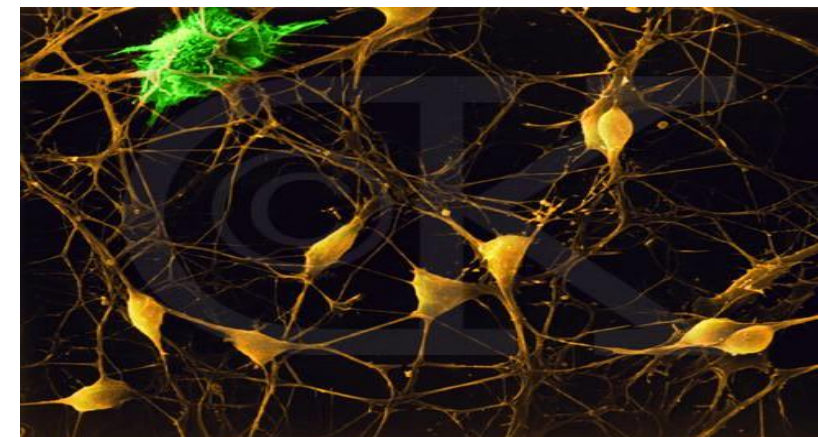
What are neural networks?



➤ Historical background

● Although heterogeneous, at a low level the brain is composed of neurons.

- A neuron receives input from other neurons (generally thousands) from its synapses
- Inputs are approximately summed
- When the input exceeds a threshold the neuron sends an electrical spike that travels that travels from the body, down the axon, to the next neuron(s)



- 10 billion neurons in human cortex
- 60 trillion synapses
- In first two years from birth ~1 million synapses / sec. formed



What are neural networks?

➤ The Neuron

The neuron that forms the basis of all Neural Network is an imitation of what was observed within the human brain.

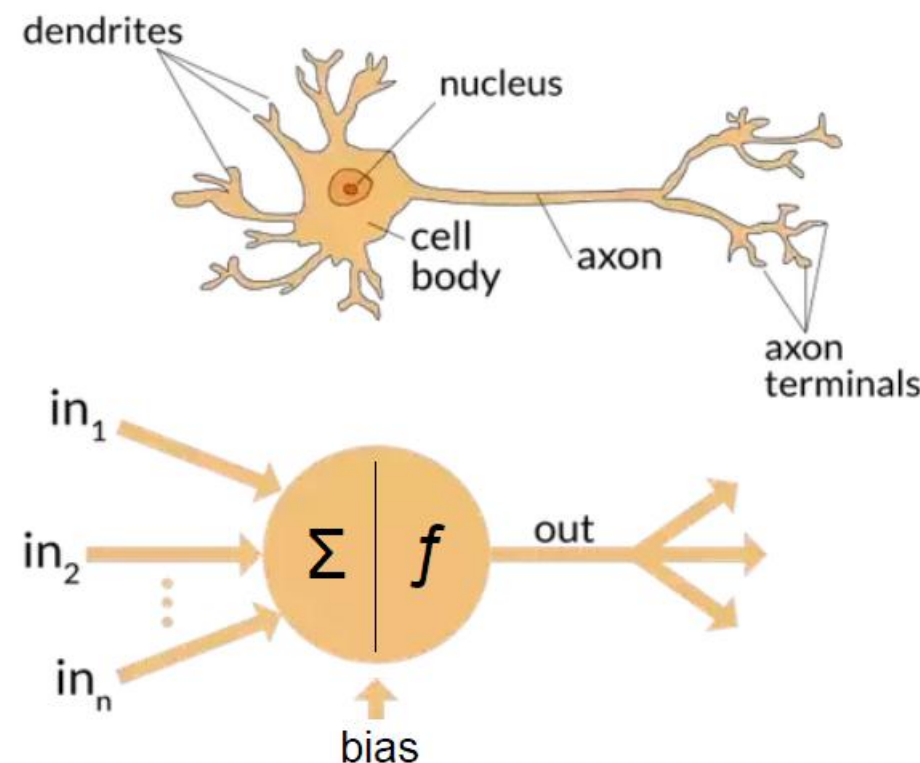
● The neuron is the basic information processing unit of a NN. It consists of:

- A set of synapses or connecting links, each link characterized by a weight: W_1, W_2, \dots, W_m
- An adder function (linear combiner) which computes the weighted sum of the inputs:

$$u = \sum_{j=1}^m w_j x_j$$

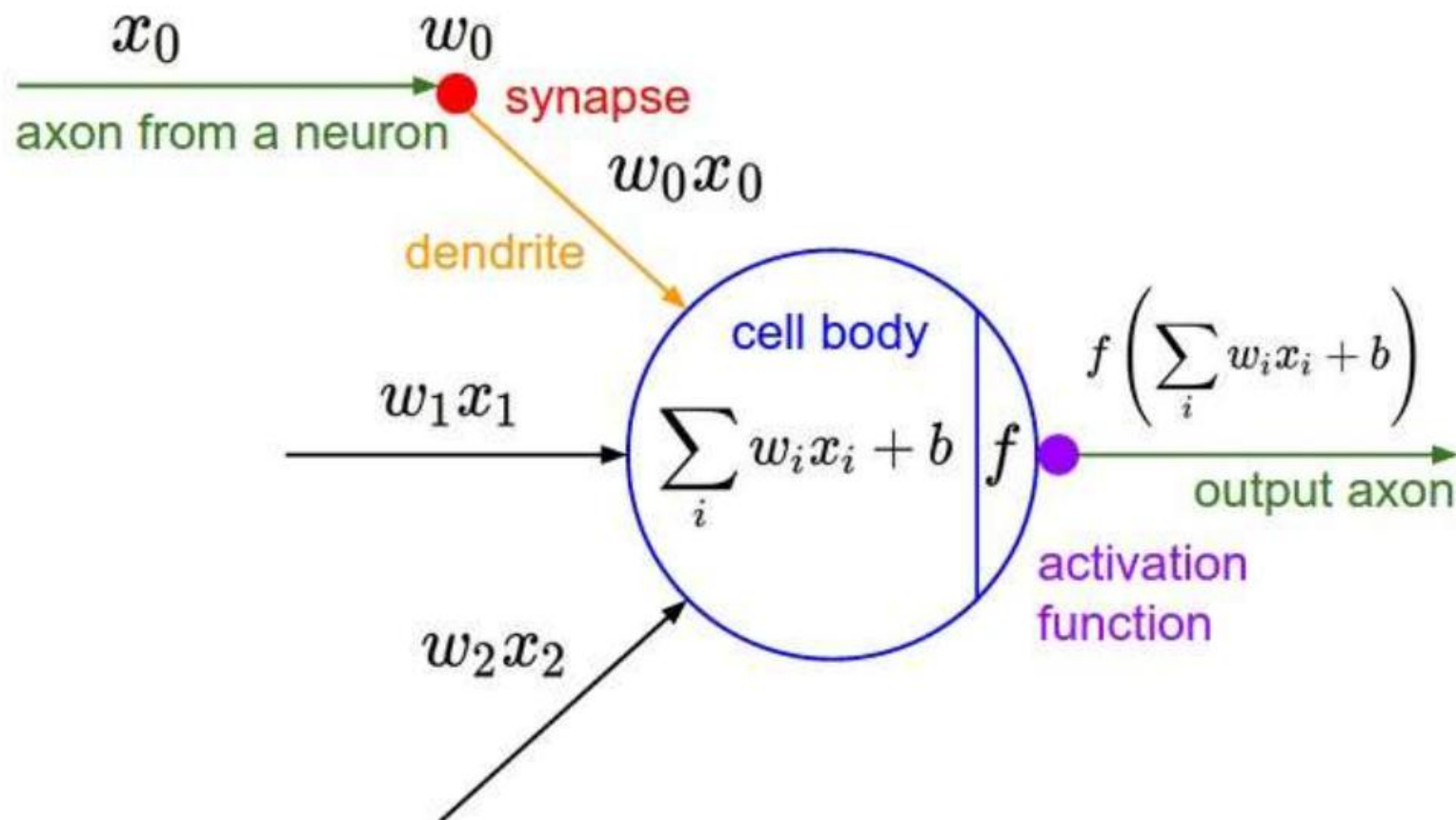
- Activation function (squashing function) for limiting the amplitude of the output of the neuron.

$$y = \varphi(u + b)$$



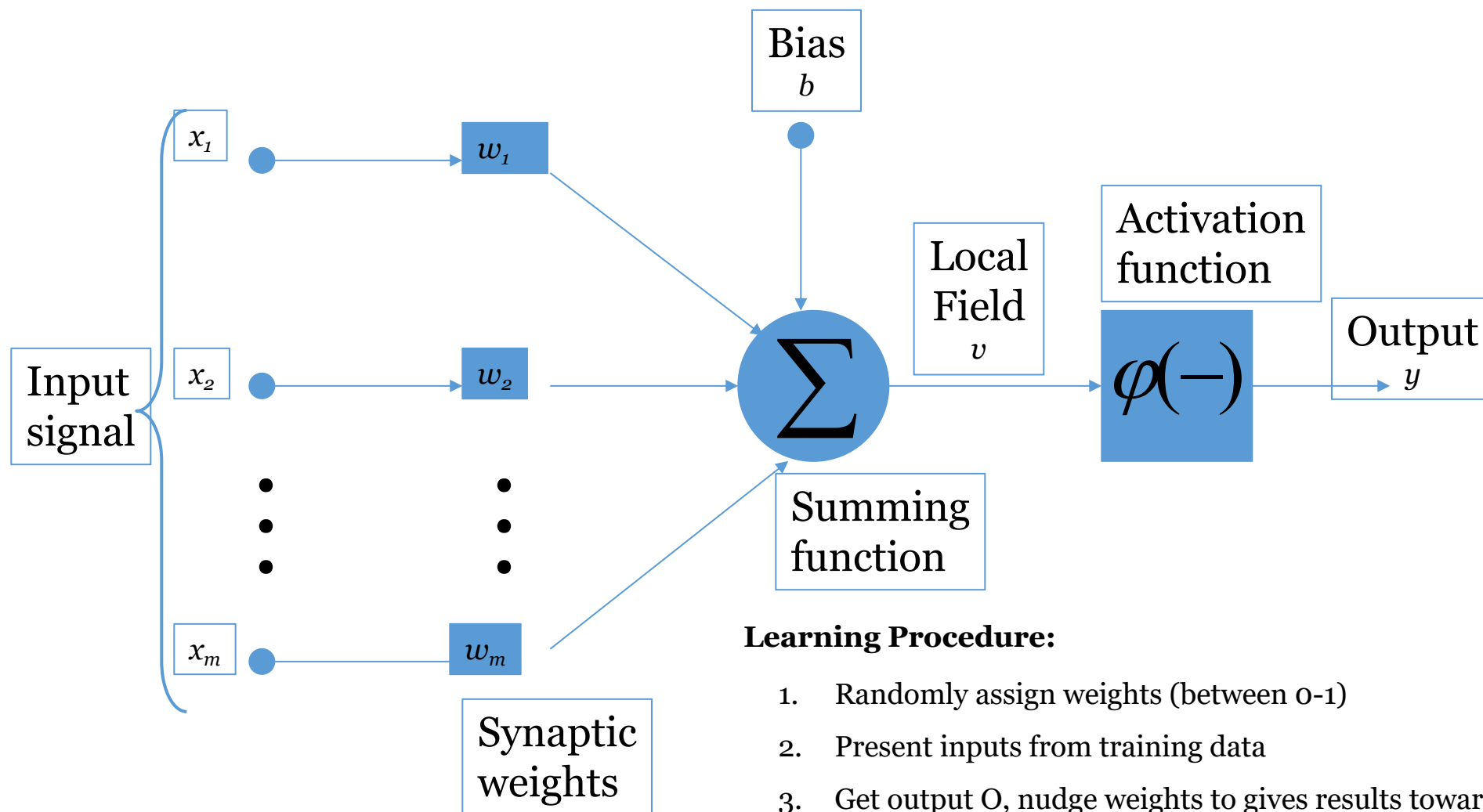


Modeling the Neuron





Modeling the Neuron



Learning Procedure:

1. Randomly assign weights (between 0-1)
2. Present inputs from training data
3. Get output O , nudge weights to gives results toward our desired output T
4. Repeat; stop when no errors, or enough epochs completed



What is DNN ?

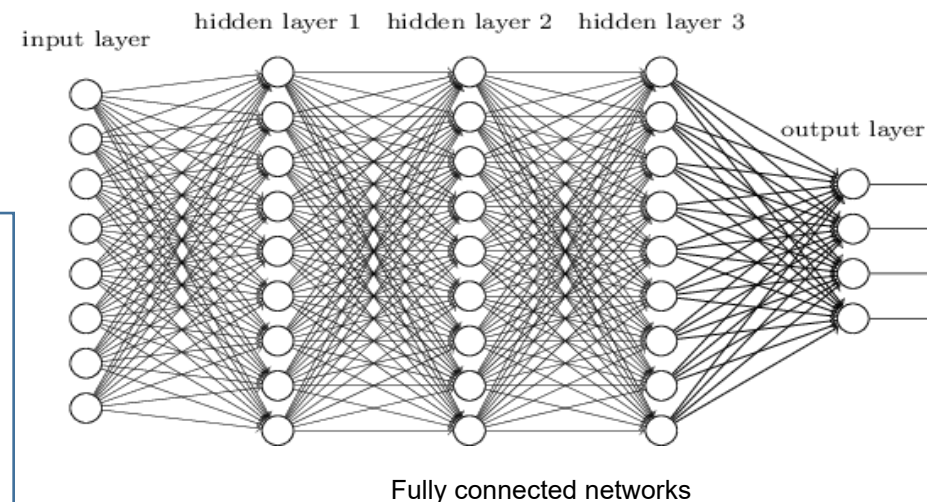


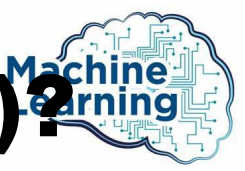
So, 1. **what exactly is deep learning ?**

And, 2. **why is it generally better** than other methods on image, speech and certain other types of data?

The short answers

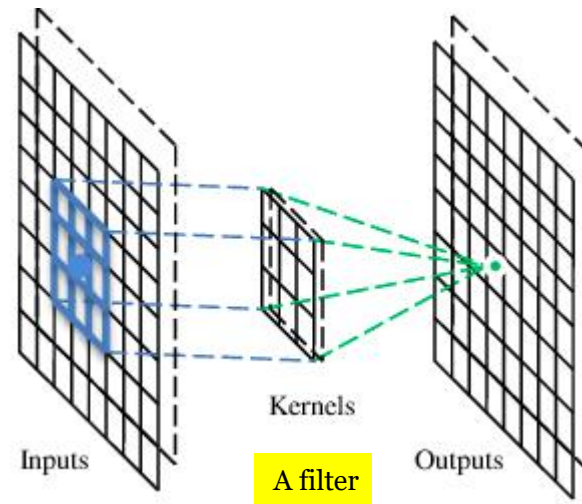
- **‘Deep Learning’ means using a neural network with several layers of nodes between input and output**
- **The series of layers between input & output do feature identification and processing in a series of stages, just as our brains seem to.**





What is a Convolutional neural network(CNN)?

A CNN is a neural network with some convolutional layers (and some other layers). A convolutional layer has a number of filters that does convolutional operation.

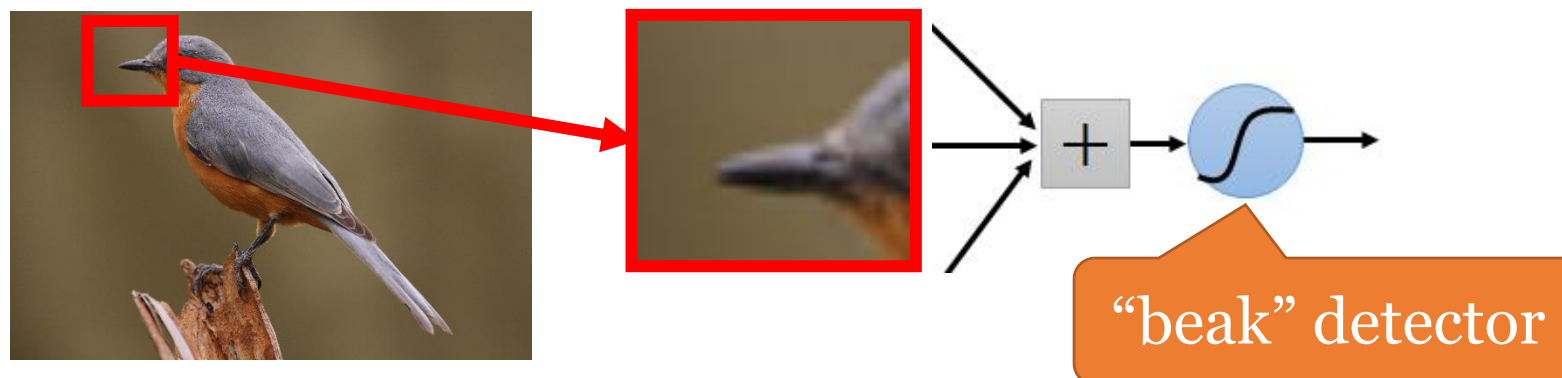




What are Convolutional neural networks

- Some patterns are much smaller than the whole image

Can represent a small region with fewer parameters

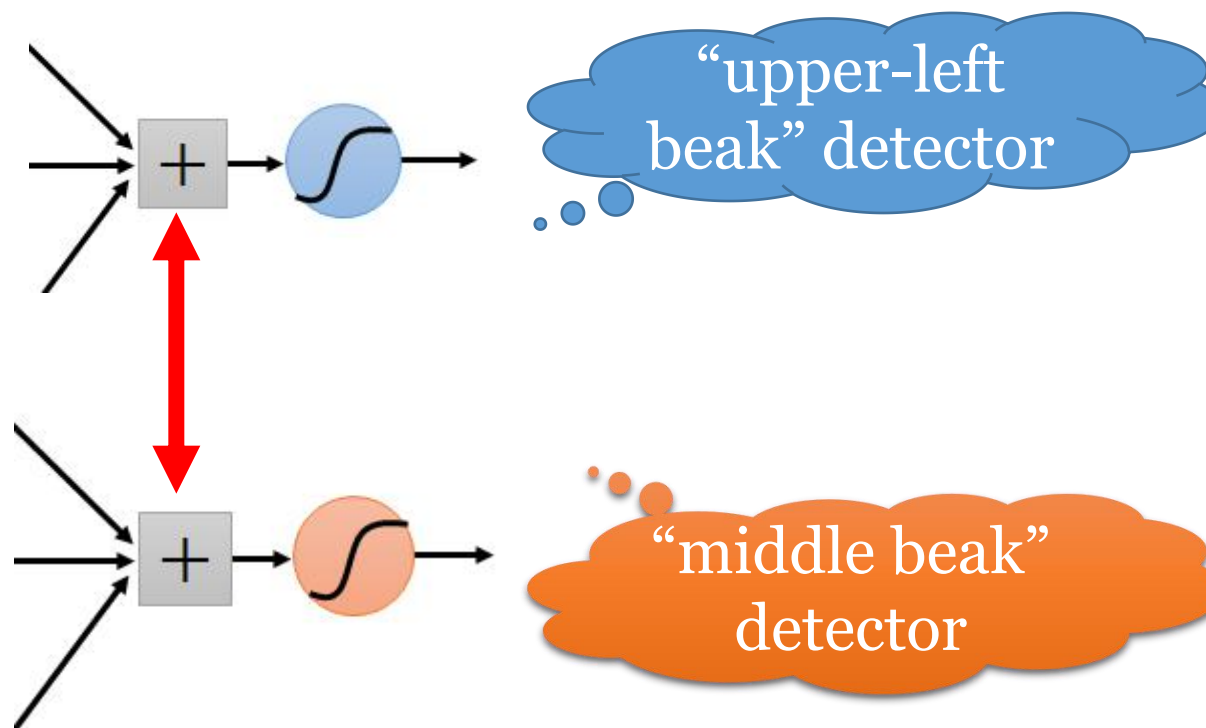
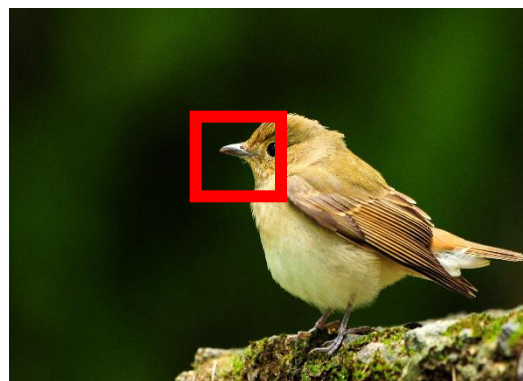




What are Convolutional neural networks



Same pattern appears in different places:
They can be compressed!
What about training a lot of such “small” detectors
and each detector must “move around”.





Convolution



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

6 x 6 image

These are the network parameters to be learned.

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

-1	1	-1
-1	1	-1
-1	1	-1

Filter 2

⋮ ⋮

Each filter detects a small pattern (3 x 3).



Convolution



stride=1

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

6 x 6 image

Dot
product



3

-1

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1



Convolution



If stride=2

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

6 x 6 image

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

3

-3



Convolution



stride=1

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

6 x 6 image

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

3	-1	-3	-1
-3	1	0	-3
-3	-3	0	1
3	-2	-2	-1



Convolution



stride=1

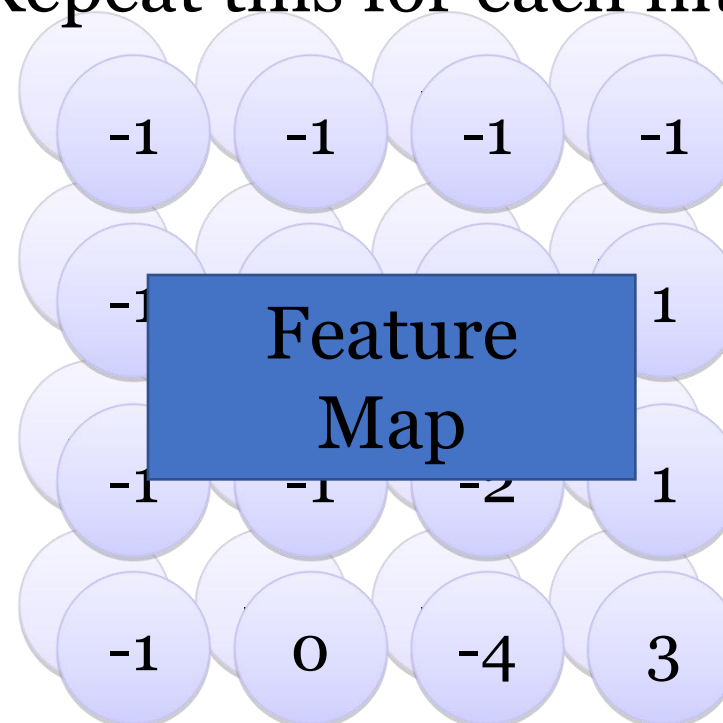
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

-1	1	-1
-1	1	-1
-1	1	-1

Filter 2

Repeat this for each filter



Two 4 x 4 images
Forming 2 x 4 x 4 matrix



Summary. To summarize, the Conv Layer:

- Accepts a volume of size $W_1 \times H_1 \times D_1$
- Requires four hyperparameters:
 - Number of filters K ,
 - their spatial extent F ,
 - the stride S ,
 - the amount of zero padding P .
- Produces a volume of size $W_2 \times H_2 \times D_2$ where:
 - $W_2 = (W_1 - F + 2P)/S + 1$
 - $H_2 = (H_1 - F + 2P)/S + 1$ (i.e. width and height are computed equally by symmetry)
 - $D_2 = K$
- With parameter sharing, it introduces $F \cdot F \cdot D_1$ weights per filter, for a total of $(F \cdot F \cdot D_1) \cdot K$ weights and K biases.
- In the output volume, the d -th depth slice (of size $W_2 \times H_2$) is the result of performing a valid convolution of the d -th filter over the input volume with a stride of S , and then offset by d -th bias.

Common settings:

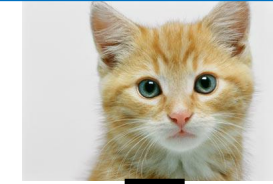
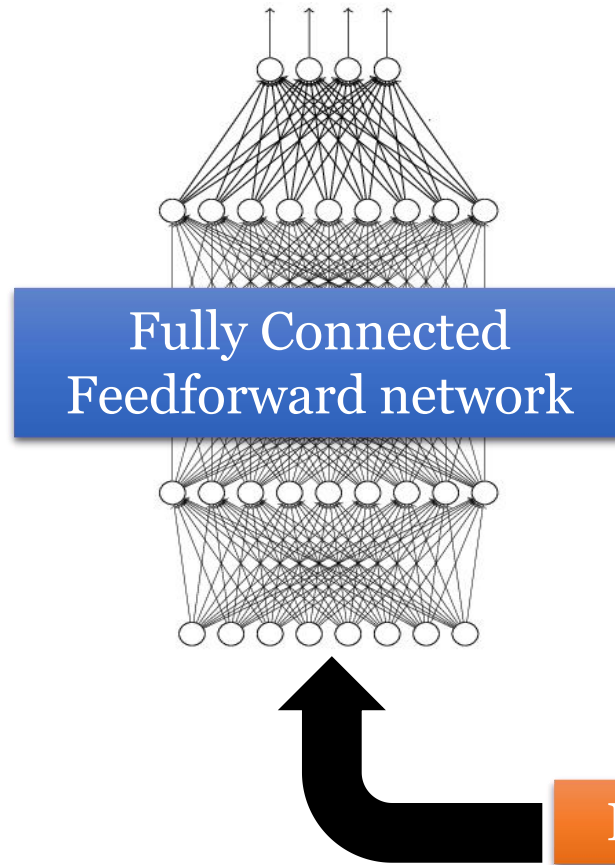
K = (powers of 2, e.g. 32, 64, 128, 512)

- $F = 3, S = 1, P = 1$
- $F = 5, S = 1, P = 2$
- $F = 5, S = 2, P = ?$ (whatever fits)
- $F = 1, S = 1, P = 0$



The whole CNN

cat dog



Convolution

Max Pooling

Convolution

Max Pooling

Flattened

Can repeat many times



Max Pooling

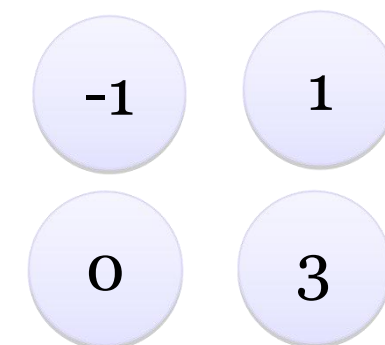
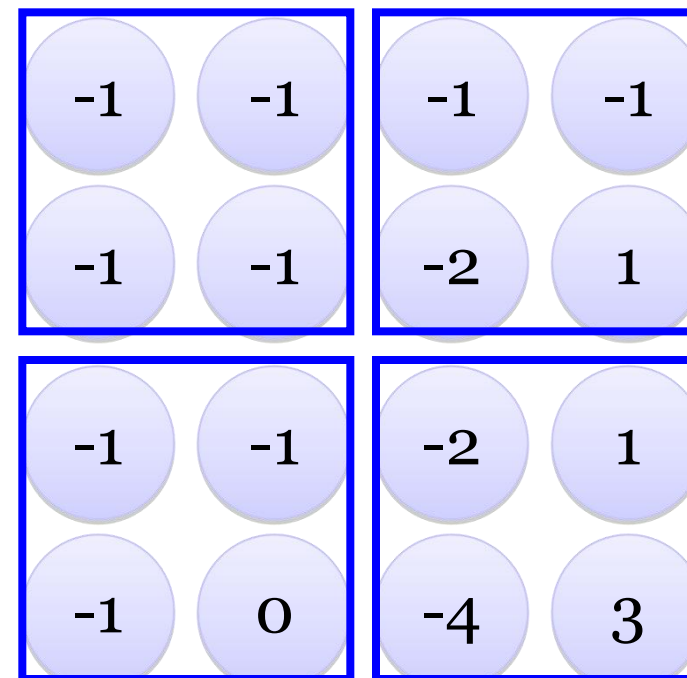
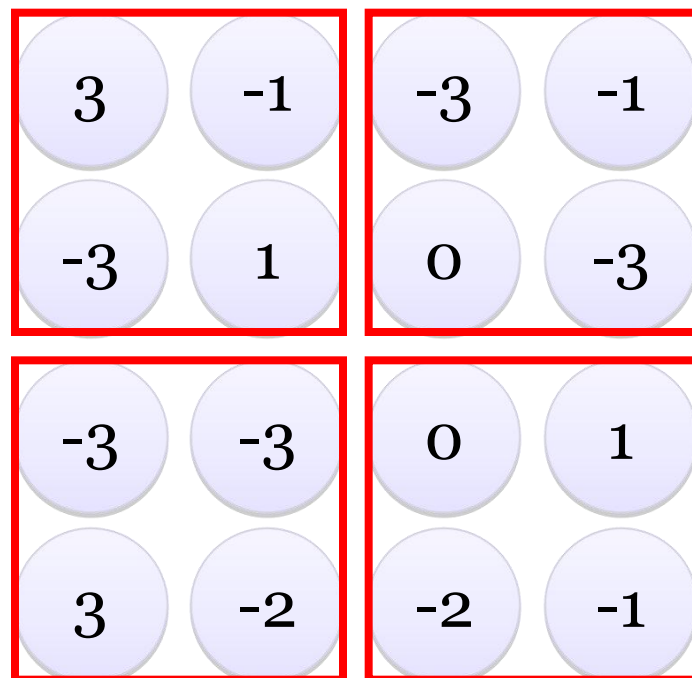
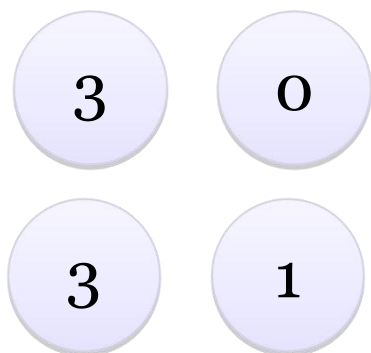


1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

-1	1	-1
-1	1	-1
-1	1	-1

Filter 2



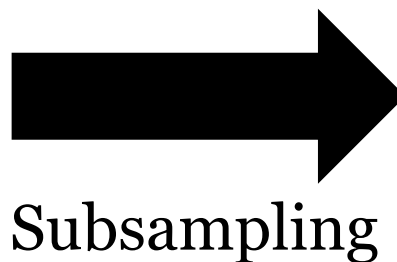


Why Pooling



- Subsampling pixels will not change the object

bird



bird



We can subsample the pixels to make image smaller



fewer parameters to characterize the image



Max Pooling



- Accepts a volume of size $W_1 \times H_1 \times D_1$
- Requires three hyperparameters:
 - their spatial extent F ,
 - the stride S ,
- Produces a volume of size $W_2 \times H_2 \times D_2$ where:
 - $W_2 = (W_1 - F)/S + 1$
 - $H_2 = (H_1 - F)/S + 1$
 - $D_2 = D_1$
- Introduces zero parameters since it computes a fixed function of the input
- Note that it is not common to use zero-padding for Pooling layers

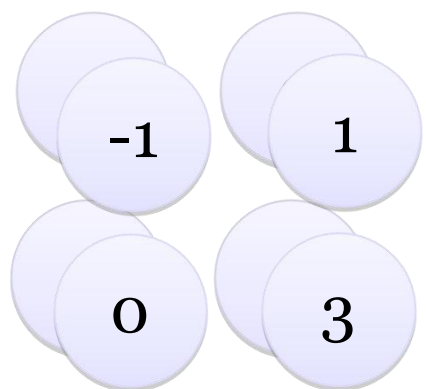
Common settings:

$$F = 2, S = 2$$

$$F = 3, S = 2$$



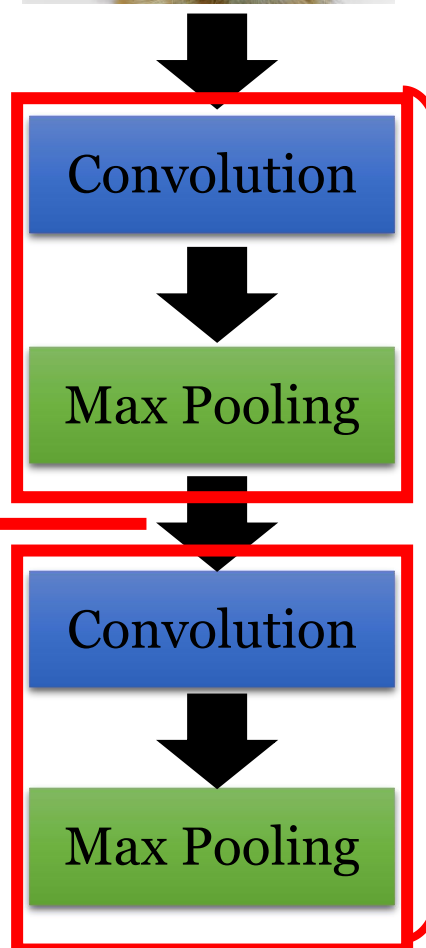
The whole CNN



A new image

Smaller than the original image

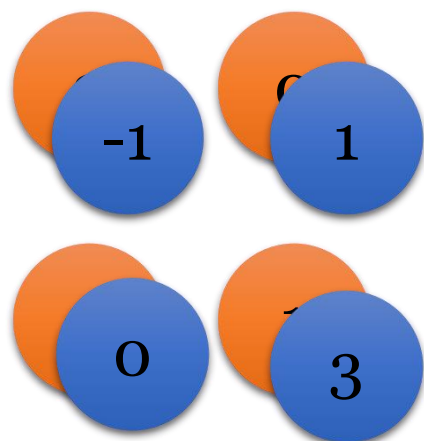
The number of channels is the number of filters



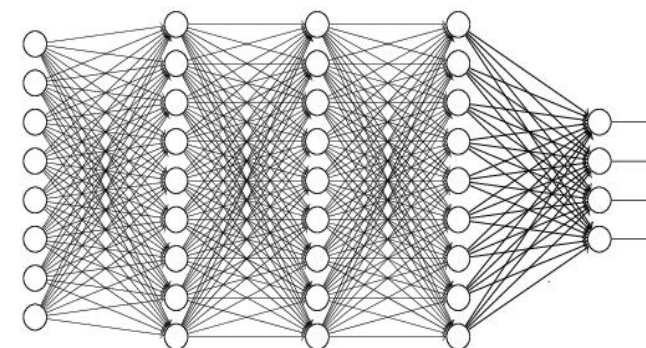
Can repeat many times



Flattening

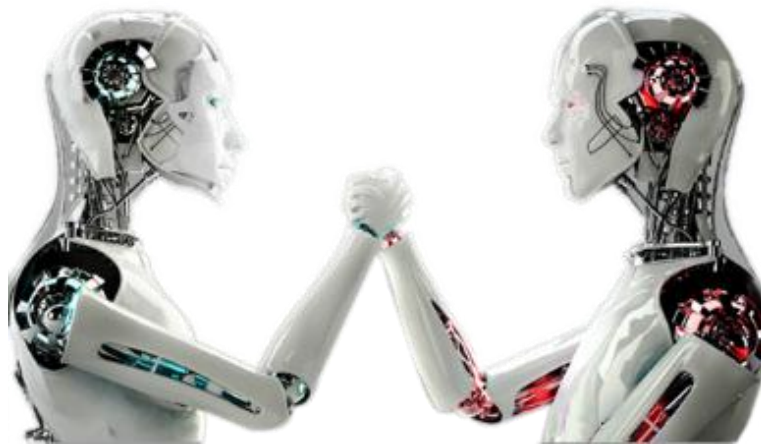


Flattened



Fully Connected
Feedforward network

Lab Task





Lab Task



1. Complete the exercises and questions in the Lab06_NN_guide.pdf
2. Submit two result files with the same content to bb. The extensions of these two files are **ipynb** and **pdf**, respectively.

Lab1: 周三 上午3-4节 荔园6栋408机房

Lab2: 周三 下午7-8节 荔园6栋406机房

Lab3: 周二下午5-6节 荔园6栋409机房



Lab References

Pytorch tutorials

 <https://pytorch.org/tutorials/>

Thanks

贾艳红 Jana

Email: jiayh@mail.sustech.edu.cn

