

Convolutional Neural Networks

Plan of Attack

Plan of Attack

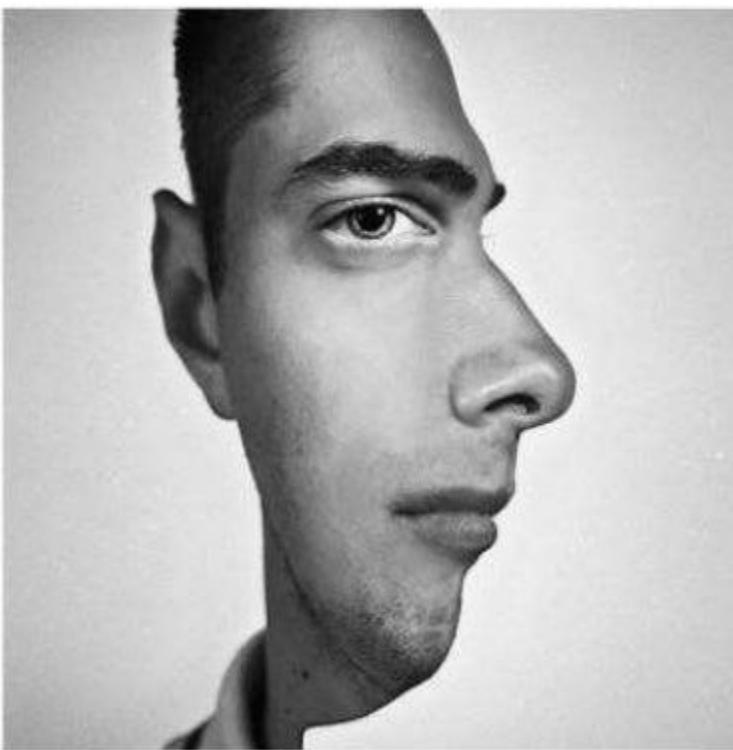
What we will learn in this section:

- What are Convolutional Neural Networks?
- Step 1 - Convolution Operation
- Step 1(b) - ReLU Layer
- Step 2 - Pooling
- Step 3 - Flattening
- Step 4 - Full Connection
- Summary

- EXTRA: Softmax & Cross-Entropy

What are convolutional neural networks?

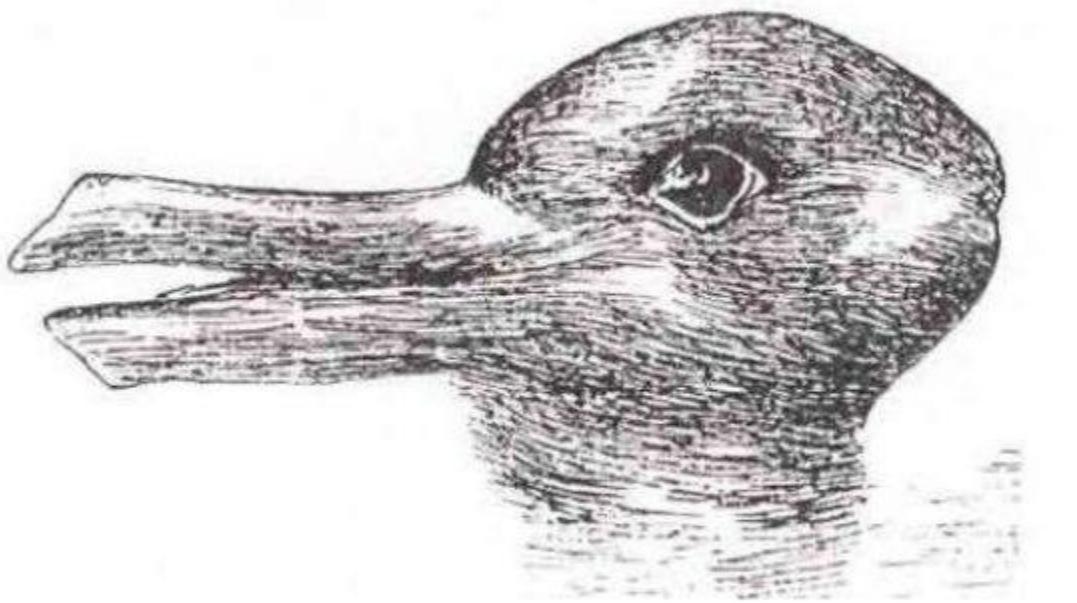
Convolutional Neural Networks



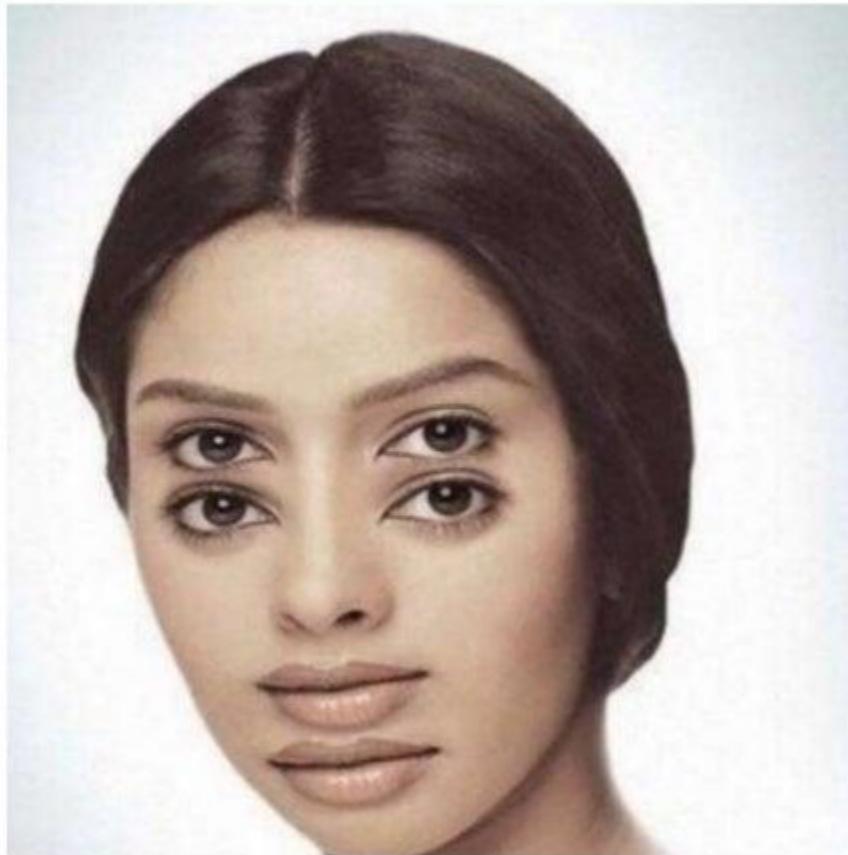
Convolutional Neural Networks



Convolutional Neural Networks



Convolutional Neural Networks



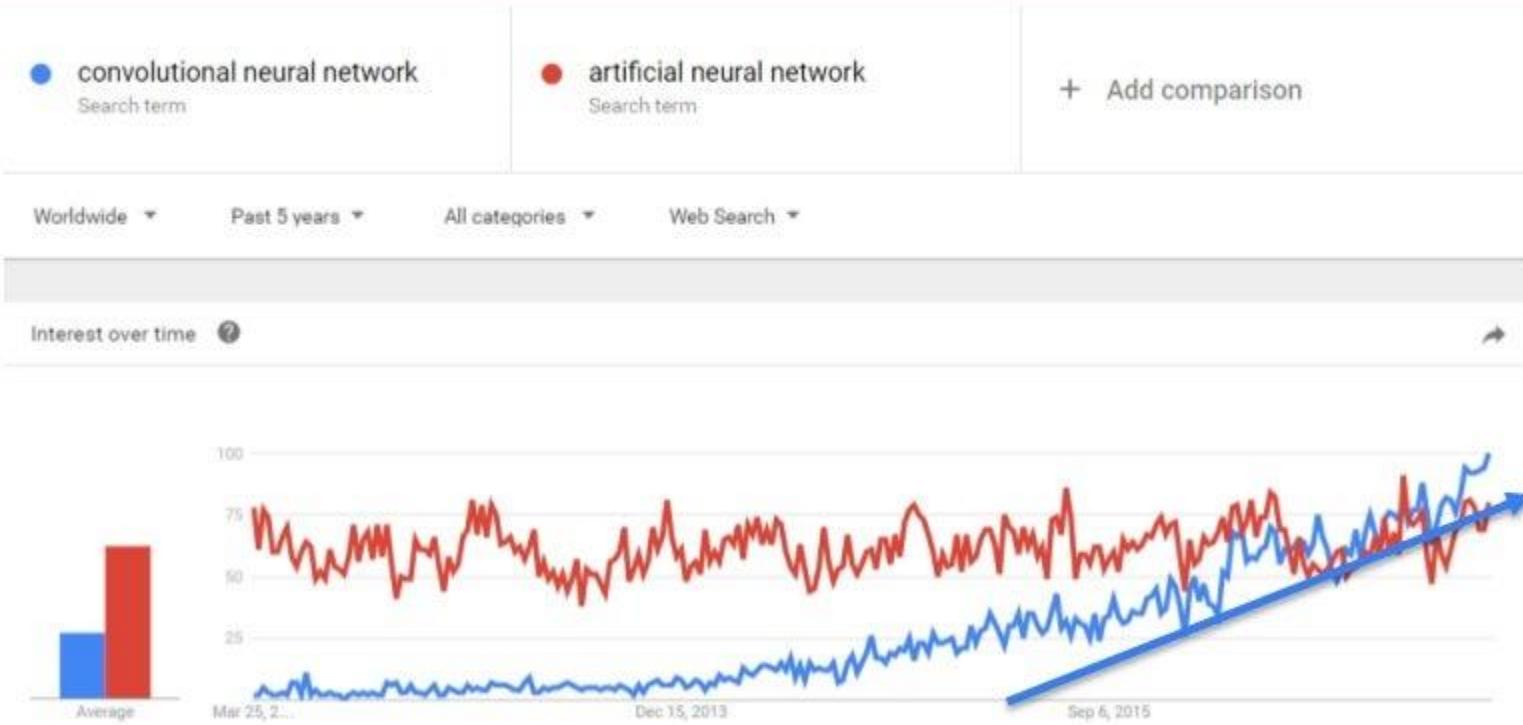
Convolutional Neural Networks

Examples from the test set
(with the network's guesses)



Image Source: a talk by Geoffrey Hinton

Convolutional Neural Networks



Source: google trends

Convolutional Neural Networks



Yann LeCun

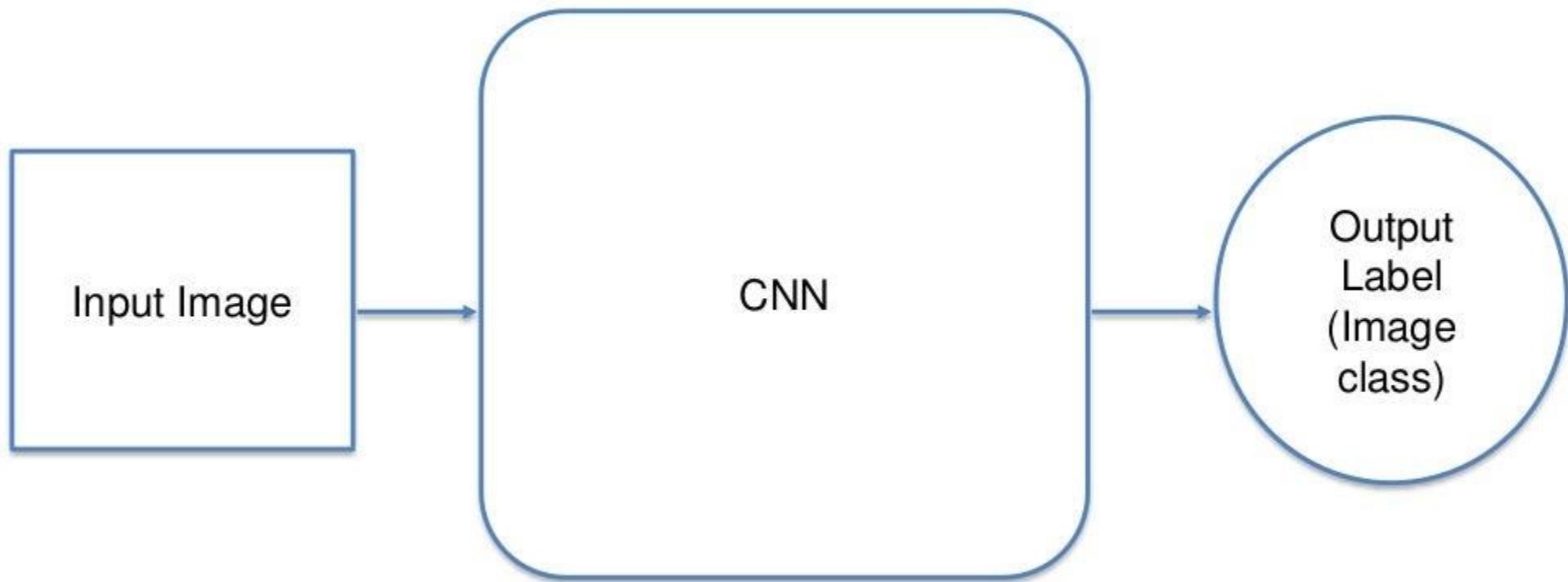
Convolutional Neural Networks

Google

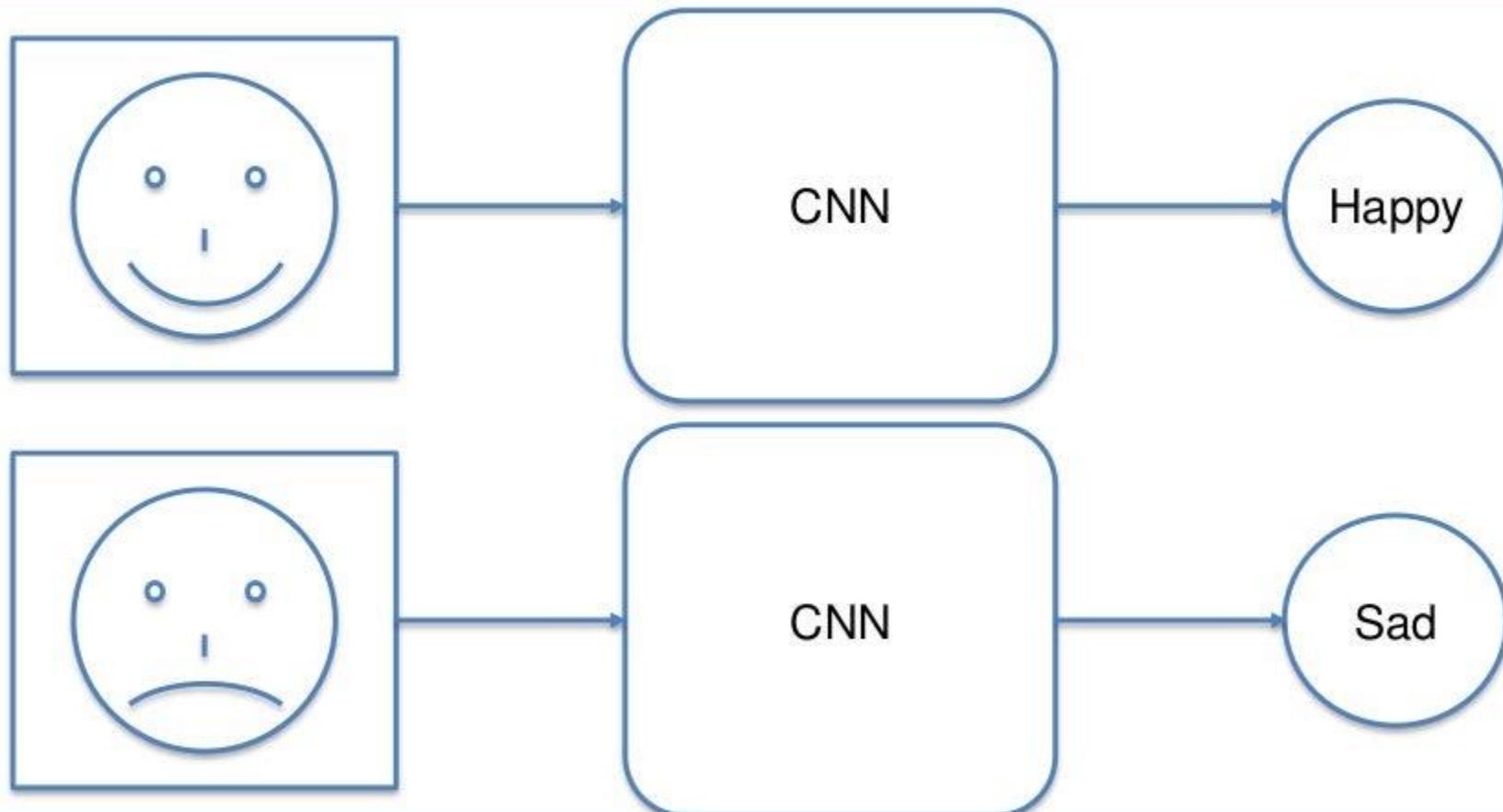
Facebook



Convolutional Neural Networks



Convolutional Neural Networks



Convolutional Neural Networks

B / W Image 2x2px

Pixel 1	Pixel 2
Pixel 3	Pixel 4

2d array

Pixel 1	Pixel 2
0 ≤ pixel value ≤ 255	0 ≤ pixel value ≤ 255
Pixel 3	Pixel 4
0 ≤ pixel value ≤ 255	0 ≤ pixel value ≤ 255

Colored Image 2x2px

Pixel 1	Pixel 2
Pixel 3	Pixel 4

3d array

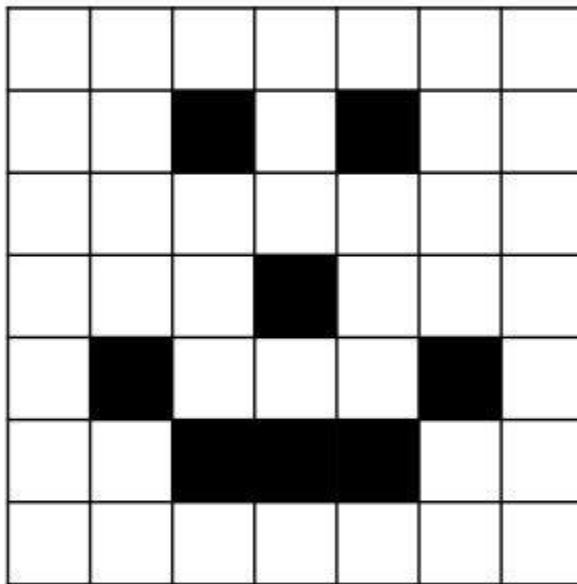
Red channel

Green channel

Pixel 1	Pixel 2
0 ≤ pixel value ≤ 255	0 ≤ pixel value ≤ 255
Pixel 3	Pixel 4
0 ≤ pixel value ≤ 255	0 ≤ pixel value ≤ 255

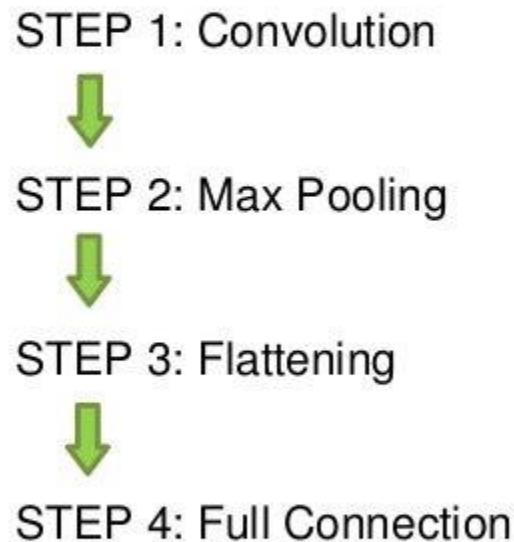
Blue channel

Convolutional Neural Networks



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

Convolutional Neural Networks



Convolutional Neural Networks

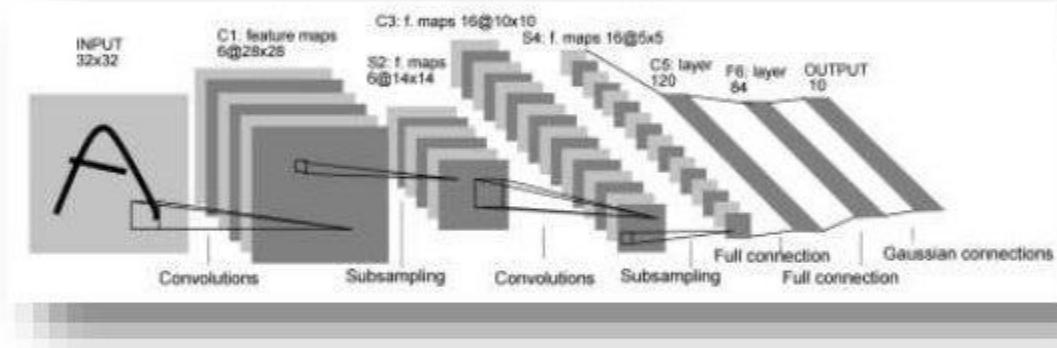
Additional Reading:

*Gradient-Based Learning
Applied to Document
Recognition*

By Yann LeCun et al. (1998)

Link:

<http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf>



Step 1 – Convolution

Step 1 - Convolution

$$(f * g)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau) g(t - \tau) d\tau$$

Step 1 - Convolution

Additional Reading:

*Introduction to Convolutional
Neural Networks*

By Jianxin Wu (2017)

$$\begin{aligned}\frac{\partial z}{\partial(\text{vec}(\mathbf{y})^T)}(F^T \otimes I) &= \left((F \otimes I) \frac{\partial z}{\partial \text{vec}(\mathbf{y})} \right)^T \\ &= \left((F \otimes I) \text{vec} \left(\frac{\partial z}{\partial Y} \right) \right)^T \\ &= \text{vec} \left(I \frac{\partial z}{\partial Y} F^T \right)^T \\ &= \text{vec} \left(\frac{\partial z}{\partial Y} F^T \right)^T,\end{aligned}$$

Link:

<http://cs.nju.edu.cn/wujx/paper/CNN.pdf>

Step 1 - 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

0	0	1
1	0	0
0	1	1

Feature Detector

Step 1 - 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

0	0	1
1	0	0
0	1	1

Feature Detector



0				

Feature Map

Step 1 - 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

0	0	1
1	0	0
0	1	1

Feature Detector



0	1			

Feature Map

Step 1 - 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

0	0	1
1	0	0
0	1	1

Feature Detector



0	1	0		

Feature Map

Step 1 - 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

0	0	1
1	0	0
0	1	1

Feature Detector



0	1	0	0	

Feature Map

Step 1 - Convolution

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



Input Image

0	0	1
1	0	0
0	1	1

Feature Detector



0	1	0	0	0

Feature Map

Step 1 - 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



0	0	1
1	0	0
0	1	1

Feature Detector



0	1	0	0	0
0				

Feature Map

Step 1 - Convolution

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



Input Image

0	0	1
1	0	0
0	1	1

Feature Detector



0	1	0	0	0
0	1			

Feature Map

Step 1 - 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



0	0	1
1	0	0
0	1	1

=

0	1	0	0	0
0	1	1		

Input Image

Feature Detector

Feature Map

Step 1 - 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

0	0	1
1	0	0
0	1	1

Feature Detector



0	1	0	0	0
0	1	1	1	

Feature Map

Step 1 - 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

0	0	1
1	0	0
0	1	1

Feature Detector



0	1	0	0	0
0	1	1	1	0

Feature Map

Step 1 - Convolution

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



Input Image

0	0	1
1	0	0
0	1	1

Feature Detector



0	1	0	0	0
0	1	1	1	0
1				

Feature Map

Step 1 - Convolution

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



Input Image

0	0	1
1	0	0
0	1	1

Feature Detector



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

Feature Map

Step 1 - Convolution

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



Input Image

0	0	1
1	0	0
0	1	1

Feature Detector



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

Feature Map

Step 1 - 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



0	0	1
1	0	0
0	1	1

=

0	1	0	0	0
0	1	1	1	0
1	0	1	2	

Input Image

Feature Detector

Feature Map

Step 1 - Convolution

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



Input Image

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

Feature Map

Step 1 - 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

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				

Feature Map

Step 1 - 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

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			

Feature Map

Step 1 - 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

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		

Feature Map

Step 1 - 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



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	

Input Image

Feature Detector

Feature Map

Step 1 - Convolution

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



Input Image

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

Feature Map

Step 1 - 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

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				

Feature Map

Step 1 - 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

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			

Feature Map

Step 1 - Convolution

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



Input Image

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		

Feature Map

Step 1 - 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



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	

Input Image

Feature Detector

Feature Map

Step 1 - Convolution

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



Input Image

0	0	1
1	0	0
0	1	1

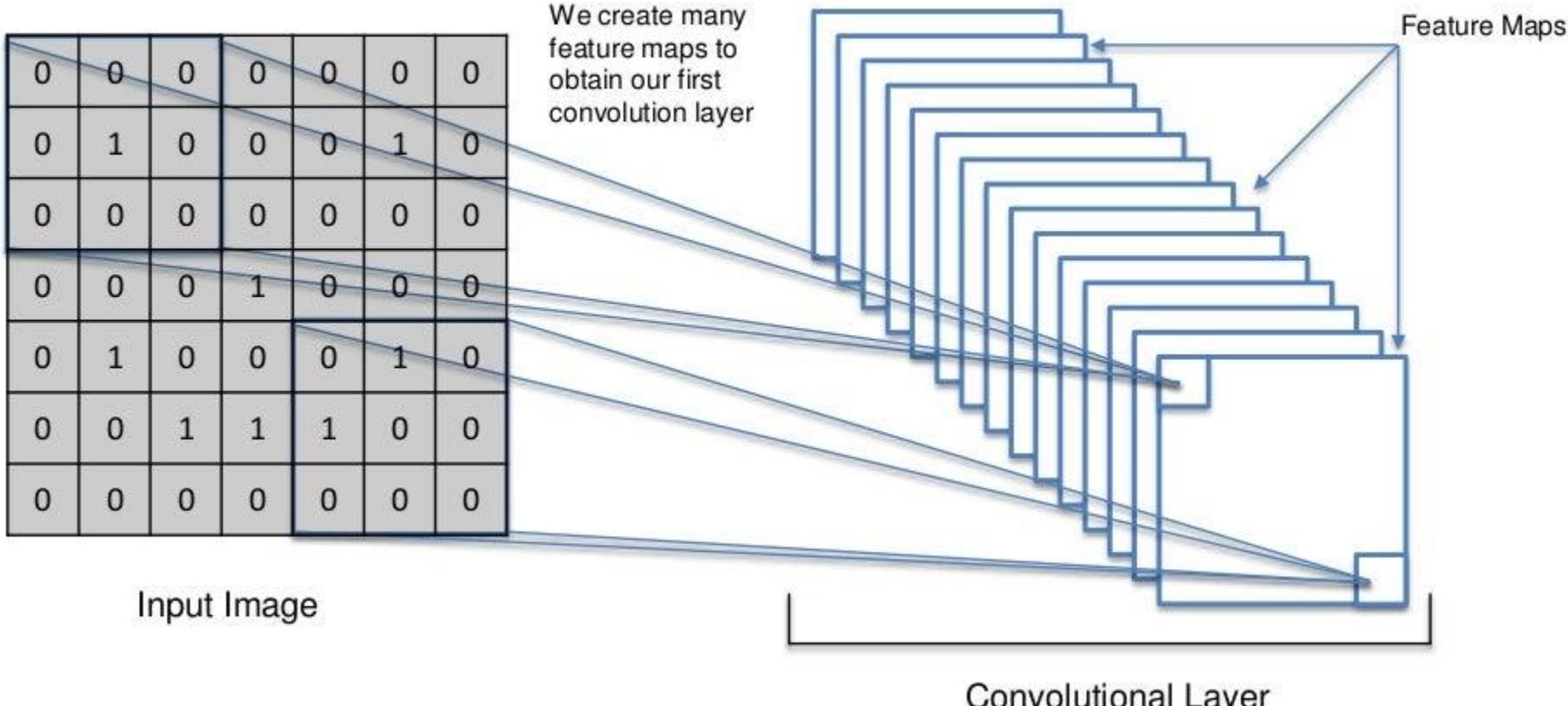
Feature Detector



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

Step 1 - Convolution



Step 1 - Convolution

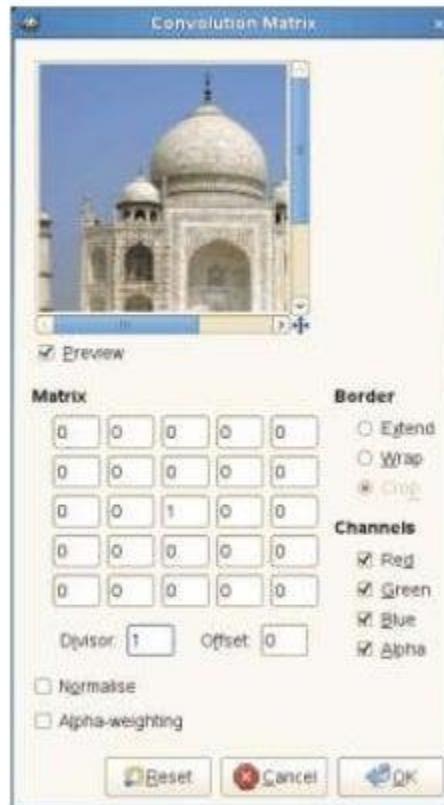


Image Source: docs.gimp.org/en/plug-in-convmatrix.html

Step 1 - Convolution

Sharpen:

0	0	0	0	0
0	0	-1	0	0
0	-1	5	-1	0
0	0	-1	0	0
0	0	0	0	0

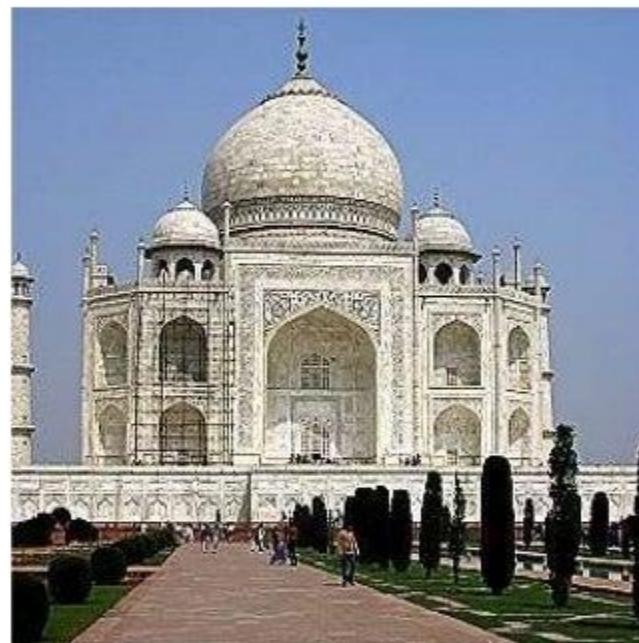


Image Source: docs.gimp.org/en/plug-in-convmatrix.html

Step 1 - Convolution

Blur:

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



Image Source: docs.gimp.org/en/plug-in-convmatrix.html

Step 1 - Convolution

Edge Enhance:

$$\begin{array}{|c|c|c|} \hline & 0 & 0 & 0 \\ \hline 0 & -1 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline \end{array}$$



Image Source: docs.gimp.org/en/plug-in-convmatrix.html

Step 1 - Convolution

Edge Detect:

$$\begin{array}{|c|c|c|} \hline & 0 & 1 & 0 \\ \hline 0 & 1 & -4 & 1 \\ \hline & 0 & 1 & 0 \\ \hline \end{array}$$

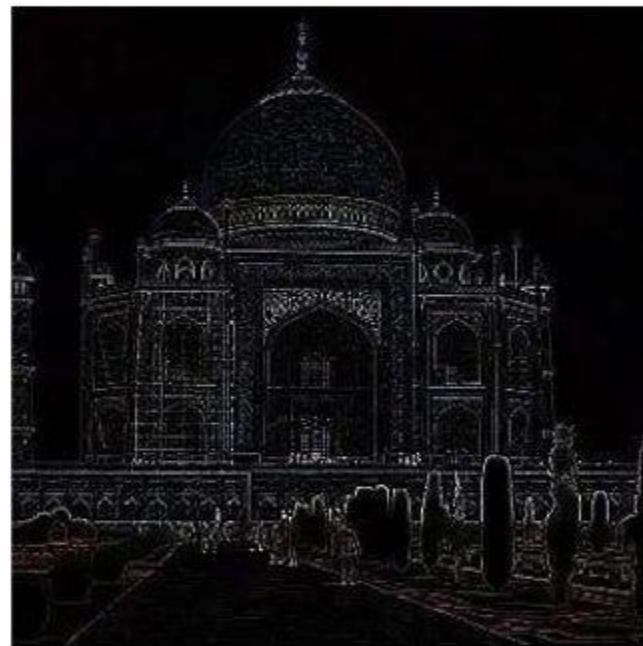


Image Source: docs.gimp.org/en/plug-in-convmatrix.html

Step 1 - Convolution

Emboss:

$$\begin{array}{c} \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & -2 & -1 & 0 \\ \hline & -1 & 1 & 1 \\ \hline & 0 & 1 & 2 \\ \hline & & & \end{array} \end{array}$$

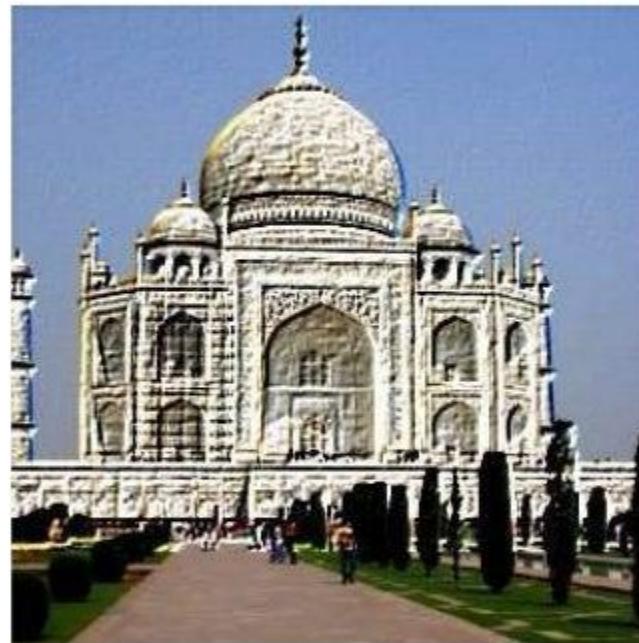


Image Source: docs.gimp.org/en/plug-in-convmatrix.html

Step 1 - Convolution

 $*$

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



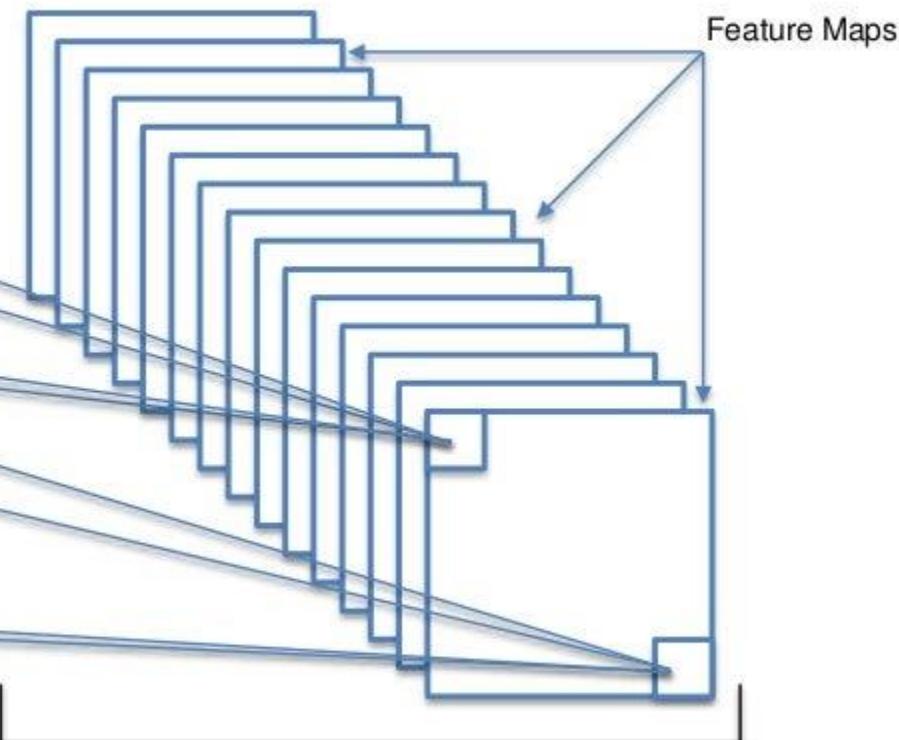
Image Source: eonardoaraujosantos.gitbooks.io

Step 1(B) – ReLU Layer

Step 1(B) – ReLU Layer

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

We create many feature maps to obtain our first convolution layer



Convolutional Layer

Step 1(B) – ReLU Layer

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

Input Image

Feature Maps

Convolutional Layer

Rectifier

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

$$\sum_{i=1}^m w_i x_i$$

y

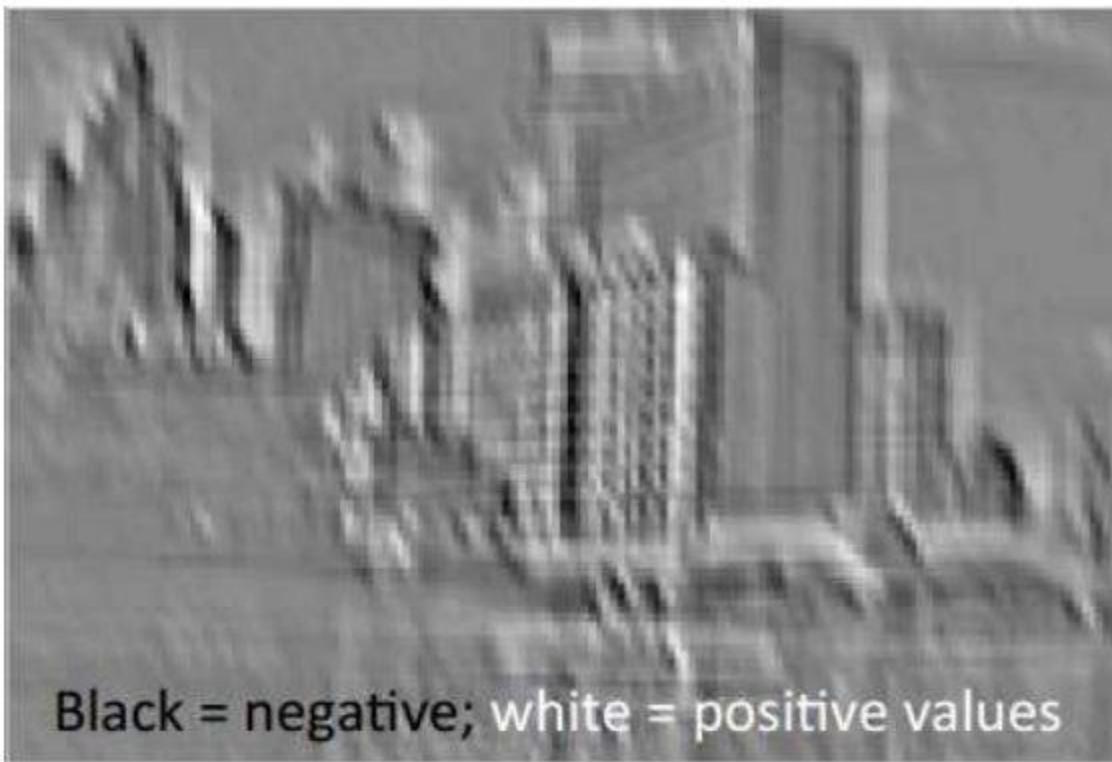
0

Step 1(B) – ReLU Layer



Image Source: http://mlss.tuebingen.mpg.de/2015/slides/fergus/Fergus_1.pdf

Step 1(B) – ReLU Layer



Black = negative; white = positive values

Image Source: http://mlss.tuebingen.mpg.de/2015/slides/fergus/Fergus_1.pdf

Step 1(B) – ReLU Layer



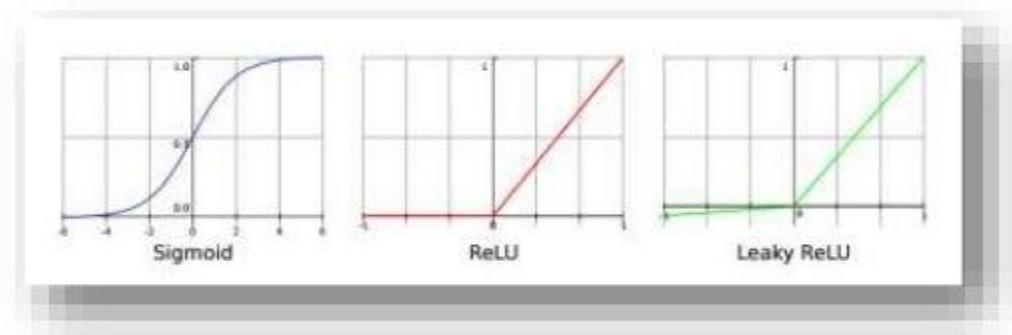
Image Source: http://mlss.tuebingen.mpg.de/2015/slides/fergus/Fergus_1.pdf

Step 1(B) – ReLU Layer

Additional Reading:

Understanding Convolutional Neural Networks with A Mathematical Model

By C.-C. Jay Kuo (2016)



Link:

<https://arxiv.org/pdf/1609.04112.pdf>

Step 1(B) – ReLU Layer

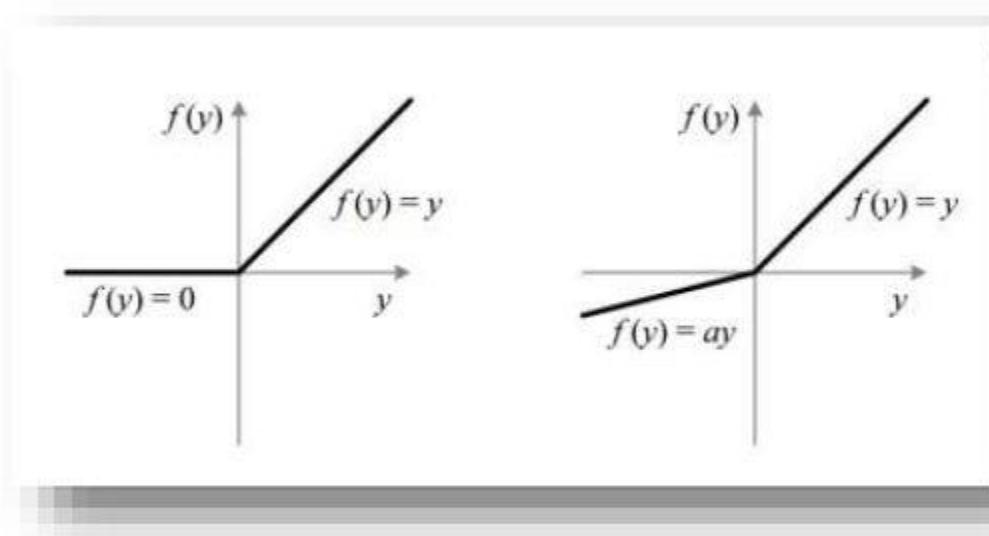
Additional Reading:

*Delving Deep into Rectifiers:
Surpassing Human-Level
Performance on ImageNet
Classification*

By Kaiming He et al. (2015)

Link:

<https://arxiv.org/pdf/1502.01852.pdf>



Step 2 – Max Pooling

Step 2 - Max Pooling



Image Source: Wikipedia

Step 2 - Max Pooling

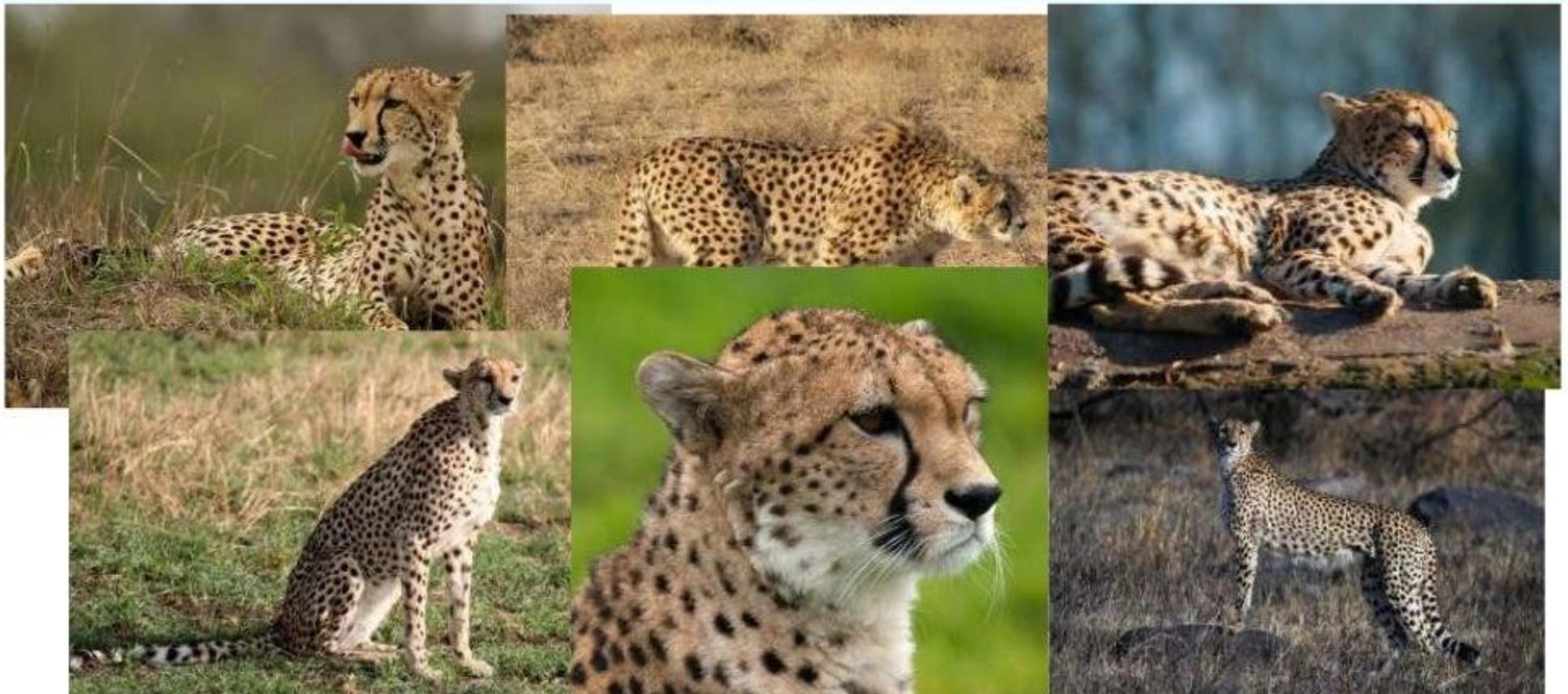


Image Source: Wikipedia

Step 2 - Max 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

Feature Map

Step 2 - Max 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

Feature Map

Max Pooling



Pooled Feature Map

Step 2 - Max 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

Feature Map

Max Pooling



1		

Pooled Feature Map

Step 2 - Max 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

Feature Map

Max Pooling



1	1	

Pooled Feature Map

Step 2 - Max 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	

Feature Map

Max Pooling



1	1	0

Pooled Feature Map

Step 2 - Max 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

Feature Map

Max Pooling



1	1	0
4		

Pooled Feature Map

Step 2 - Max 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

Feature Map

Max Pooling



1	1	0
4	2	

Pooled Feature Map

Step 2 - Max 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

Feature Map

Max Pooling



1	1	0
4	2	1

Pooled Feature Map

Step 2 - Max 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

Feature Map

Max Pooling



1	1	0
4	2	1
0		

Pooled Feature Map

Step 2 - Max 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

Feature Map

Max Pooling



1	1	0
4	2	1
0	2	

Pooled Feature Map

Step 2 - Max 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

Feature Map

Max Pooling



1	1	0
4	2	1
0	2	1

Pooled Feature Map

Step 2 - Max Pooling

Additional Reading:

Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition

By Dominik Scherer et al. (2010)

Link:

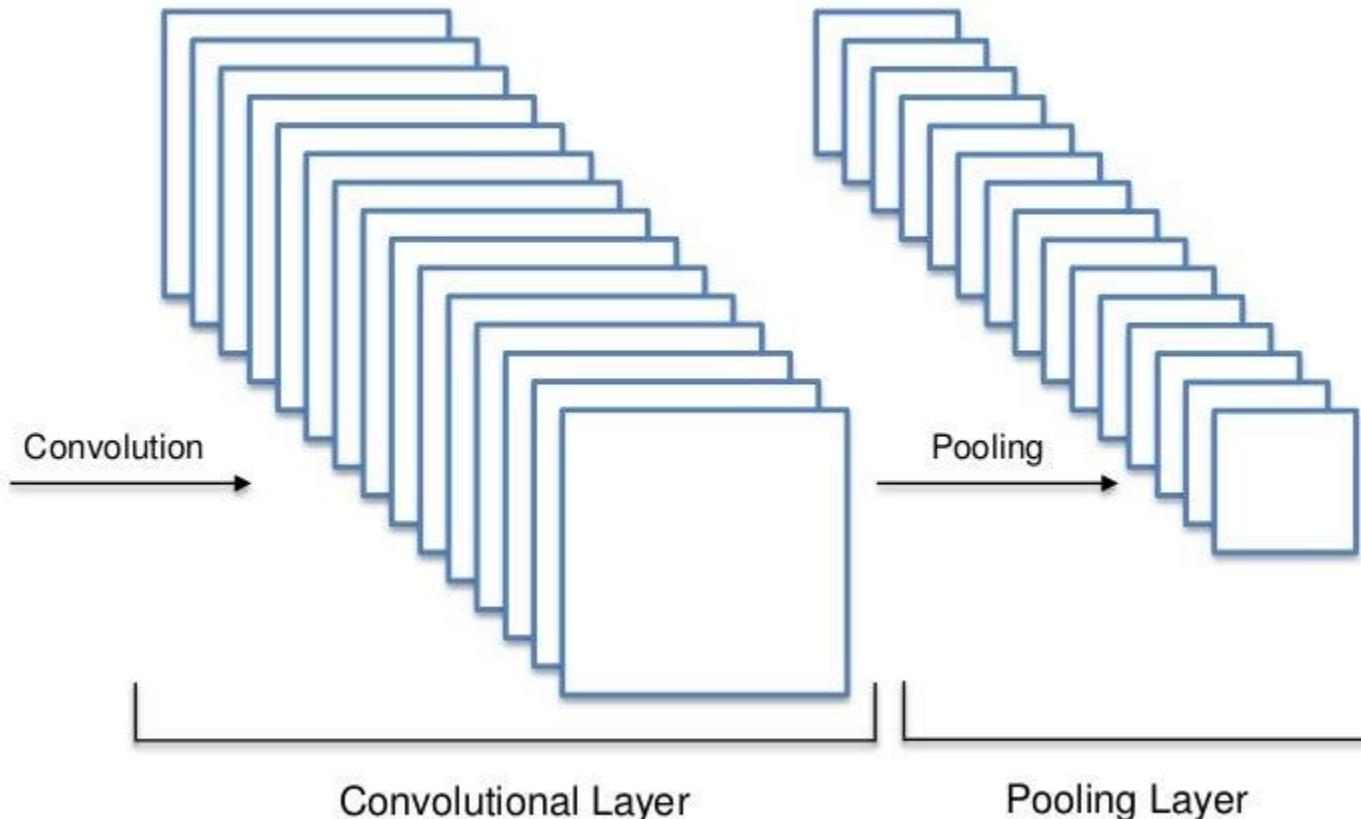
http://ais.uni-bonn.de/papers/icann2010_maxpool.pdf



Step 2 - Max Pooling

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



Convolutional Layer

Pooling Layer

Example

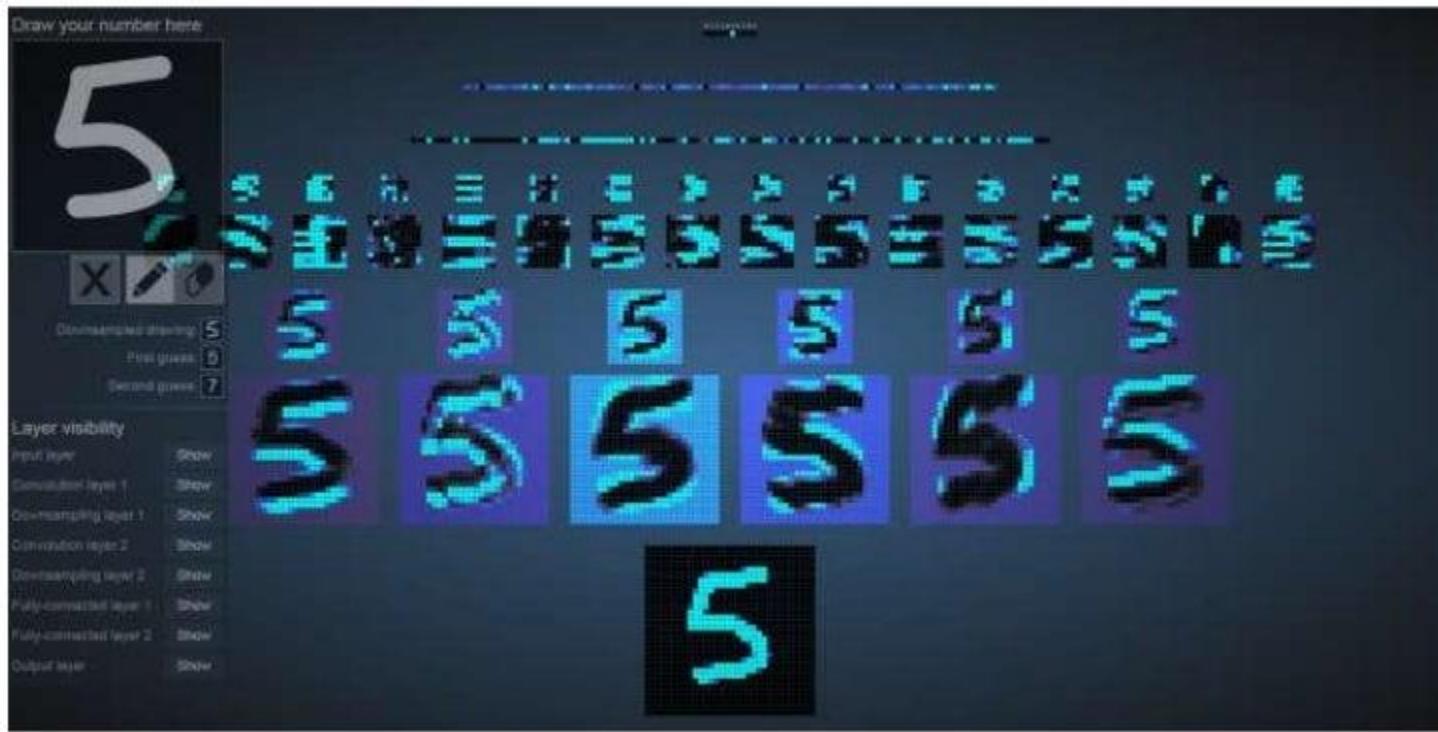


Image Source: scs.ryerson.ca/~aharley/vis/conv/flat.html

Step 3 – Flattening

Step 3 - Flattening

1	1	0
4	2	1
0	2	1

Pooled Feature Map

Step 3 - Flattening

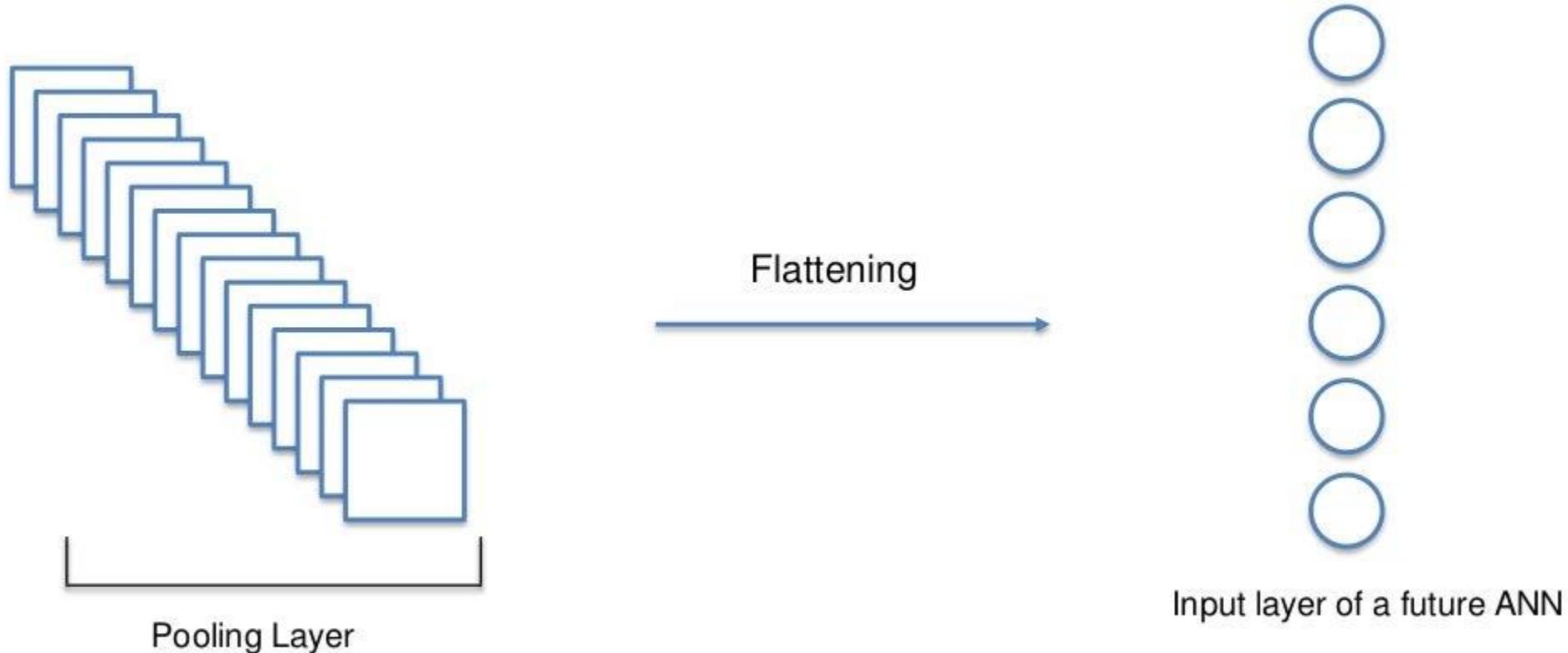
1	1	0
4	2	1
0	2	1

Pooled Feature Map

Flattening

1
1
0
4
2
1
0
2
1

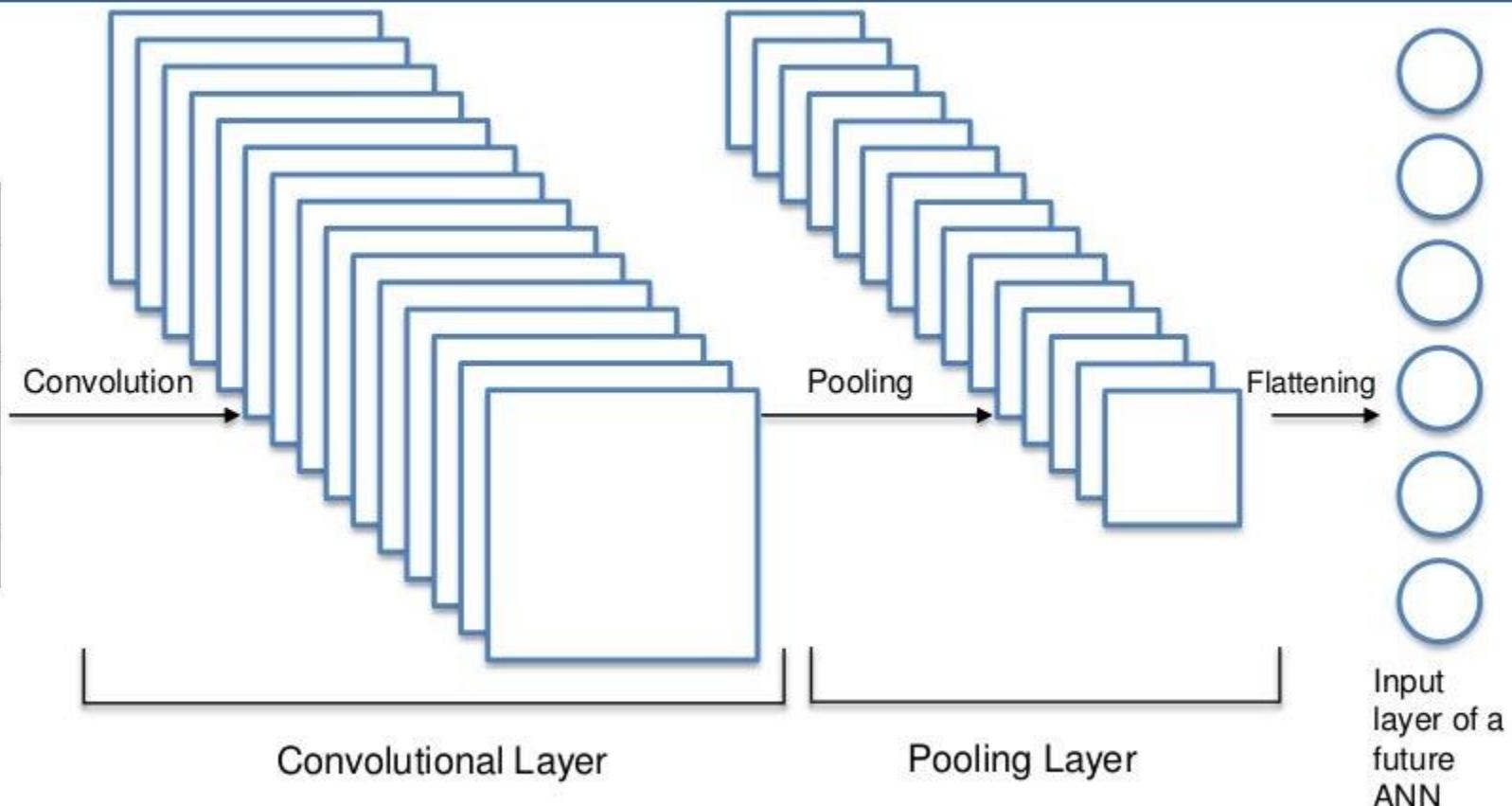
Step 3 - Flattening



Step 3 - Flattening

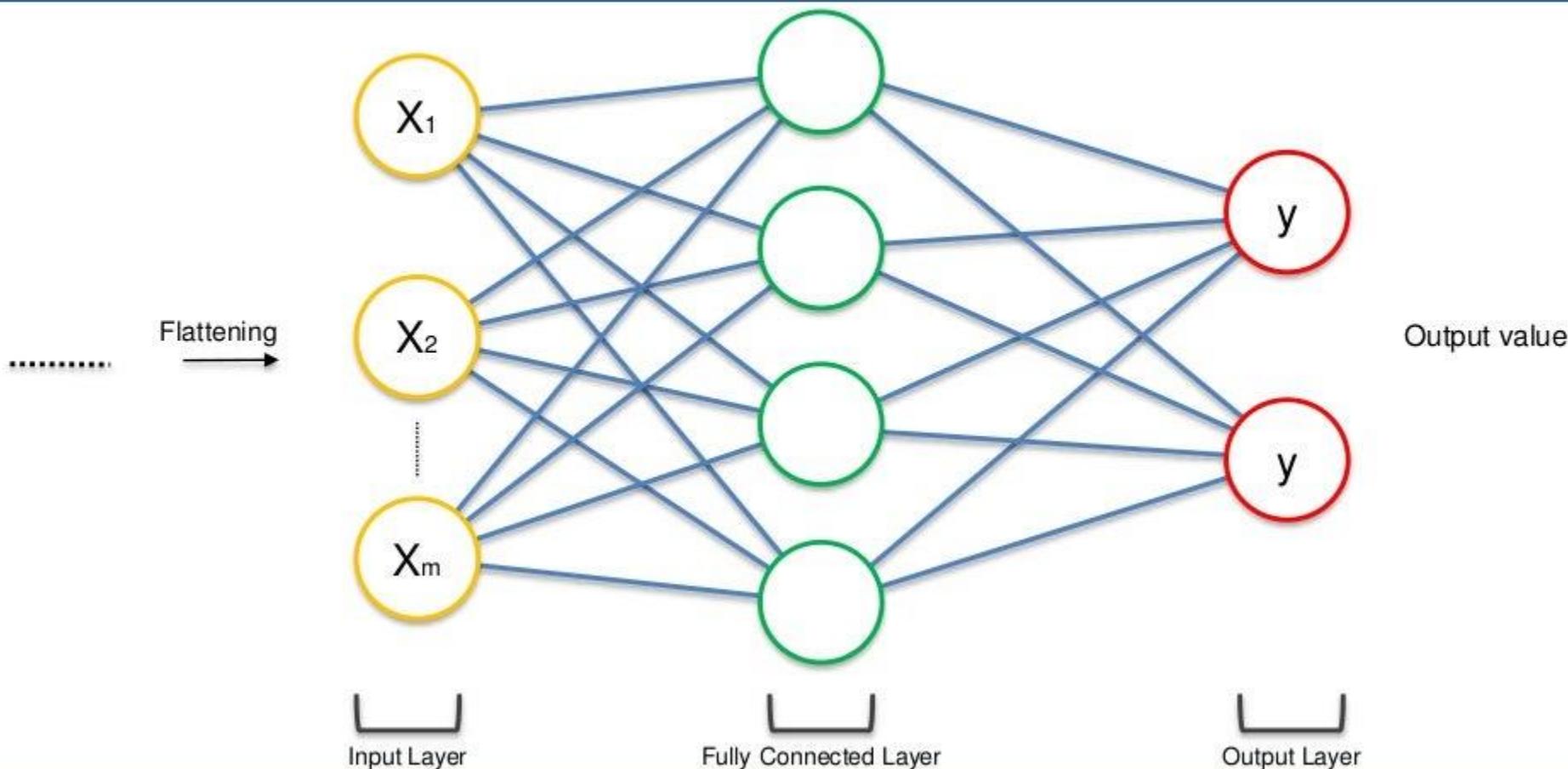
0	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	1	0	0	0	1	0	0
0	0	1	1	1	0	0	0
0	0	0	0	0	0	0	0

Input Image

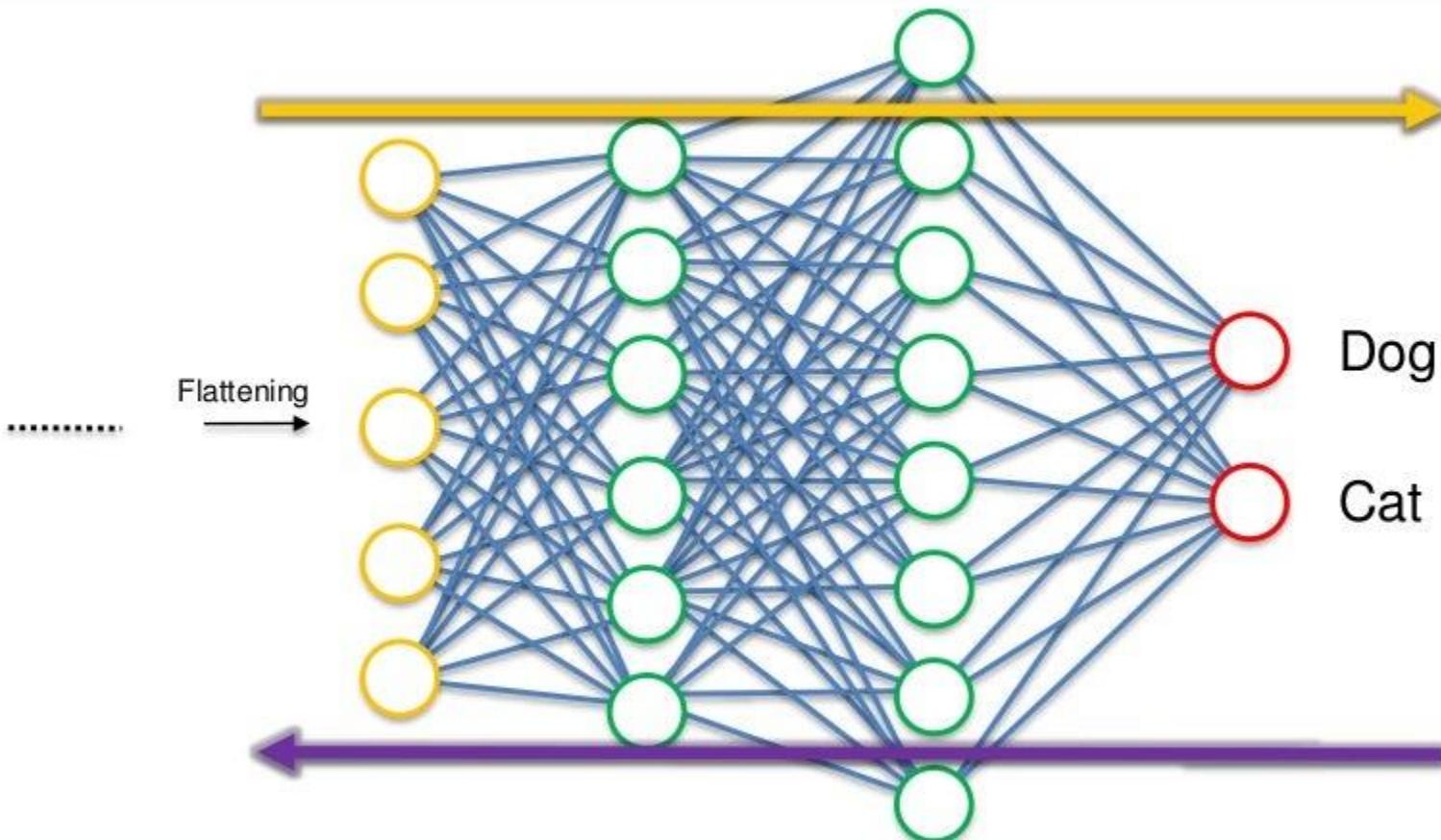


Step 4 – Full Connection

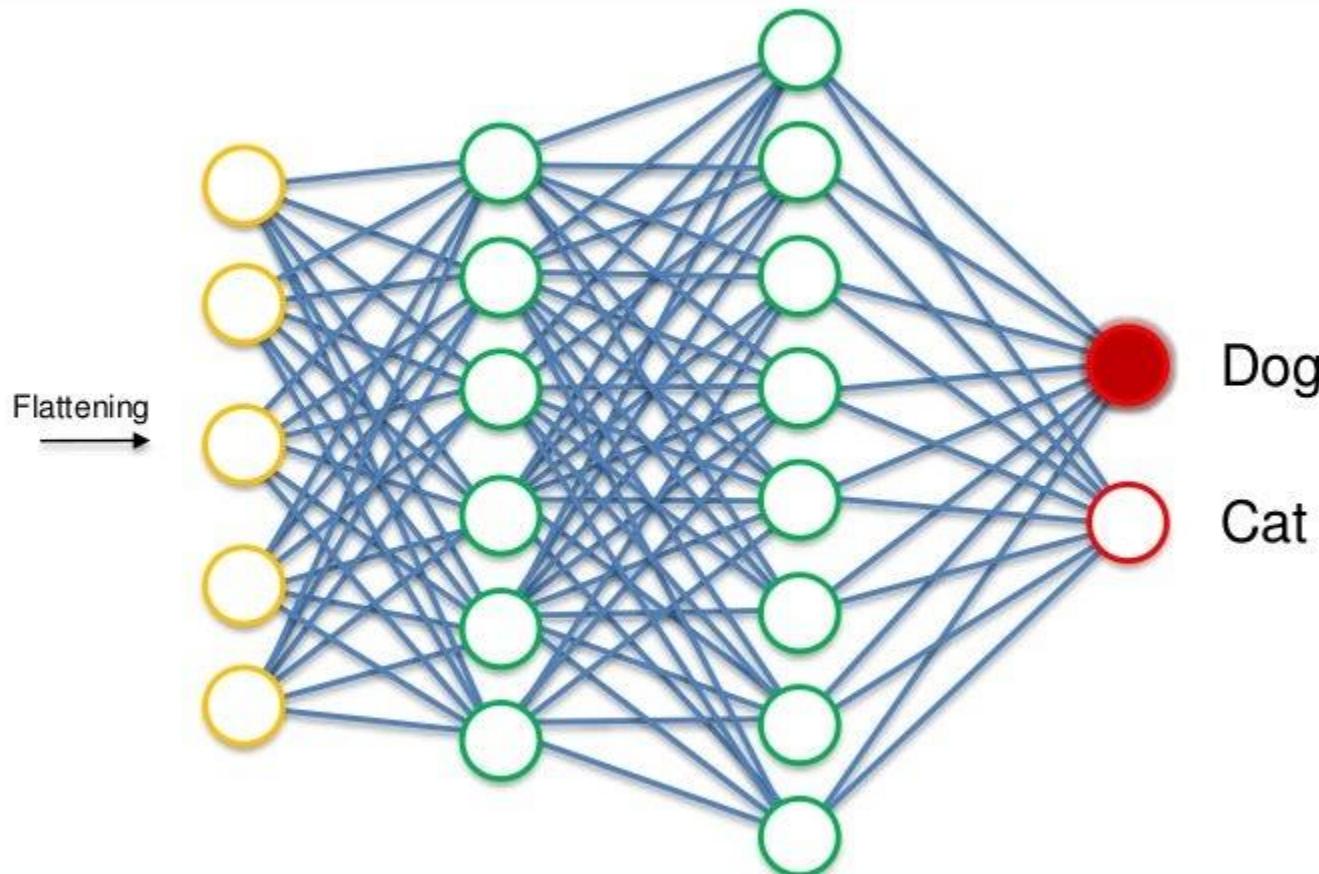
Step 4 - Full Connection



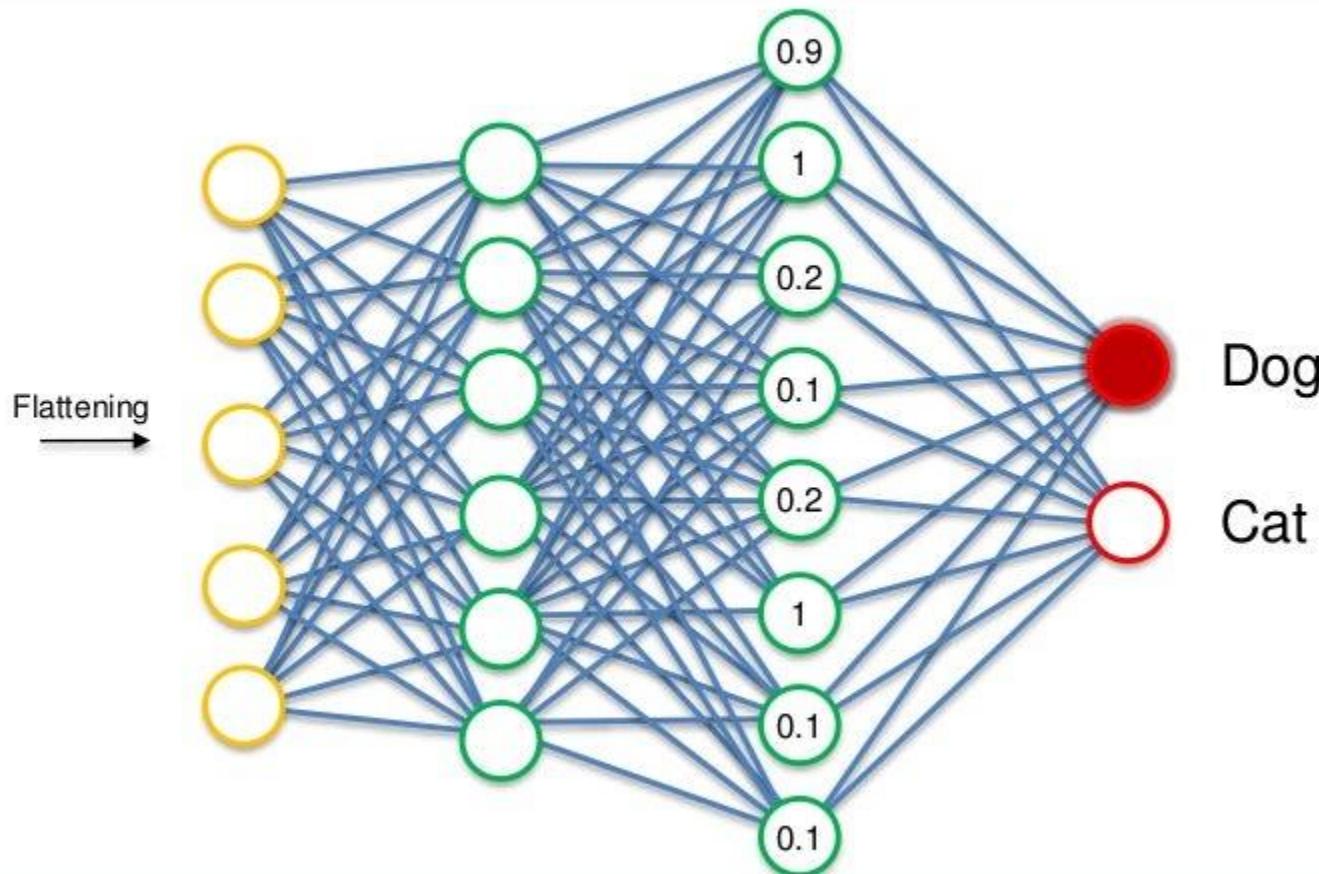
Step 4 - Full Connection



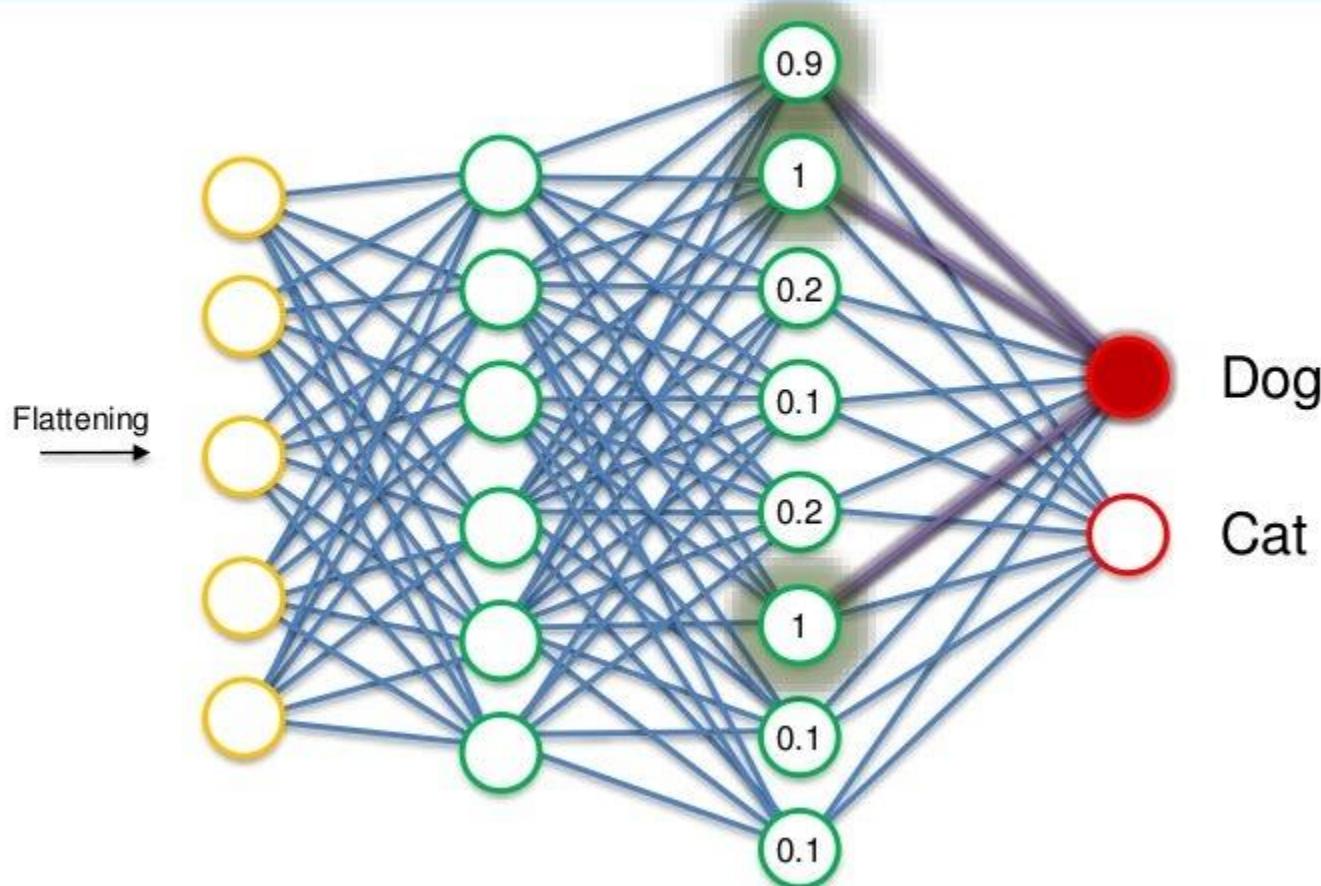
Step 4 - Full Connection



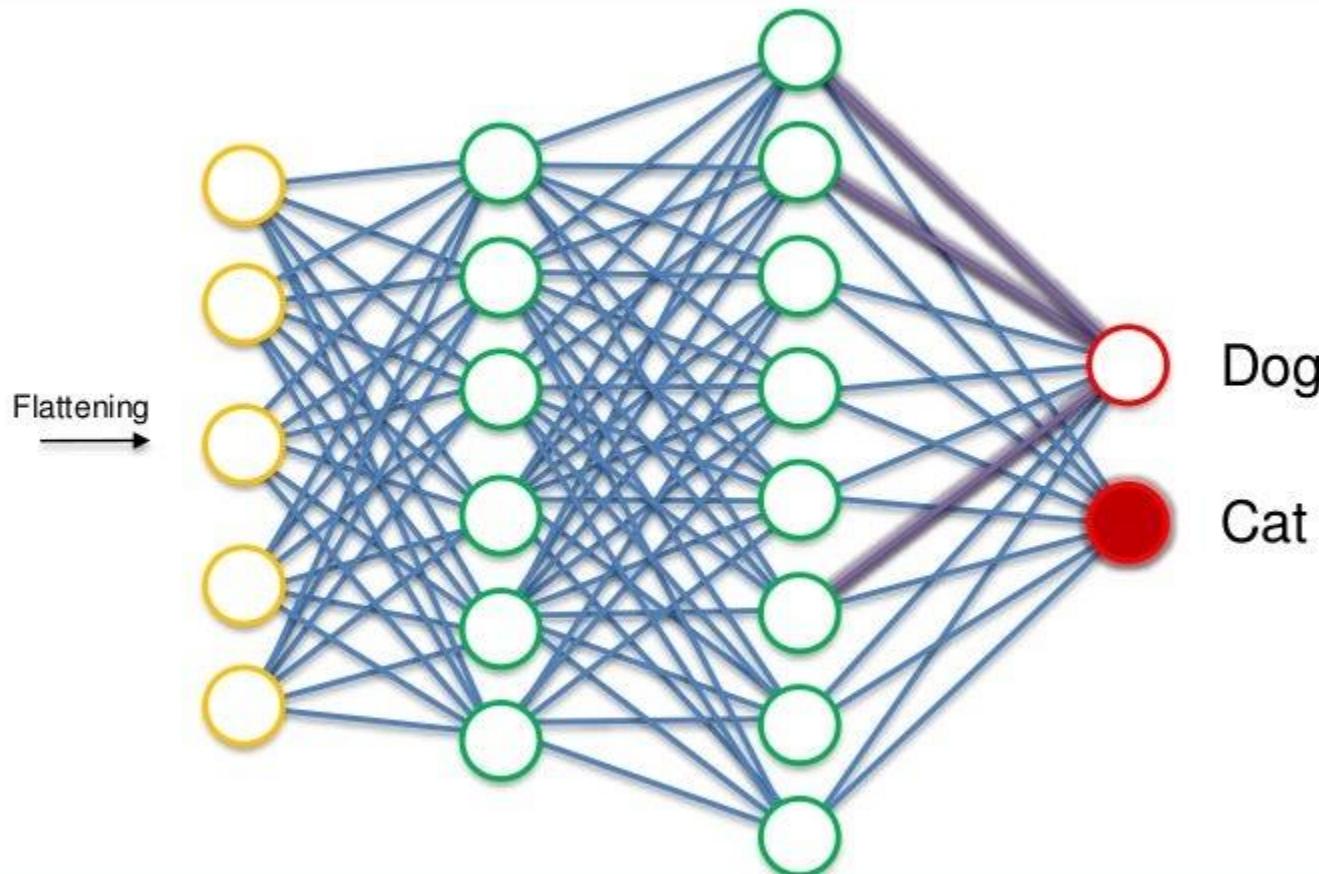
Step 4 - Full Connection



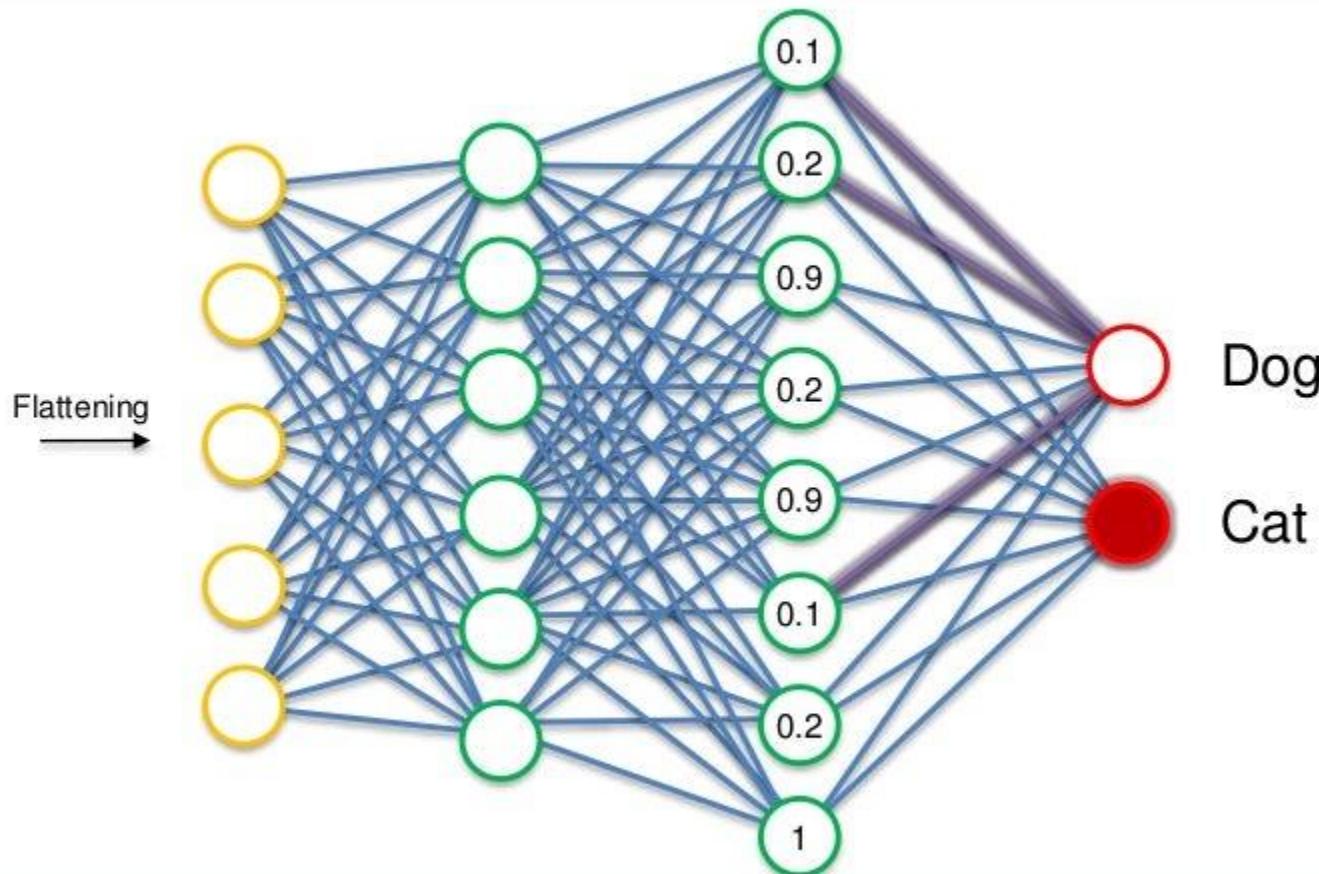
Step 4 - Full Connection



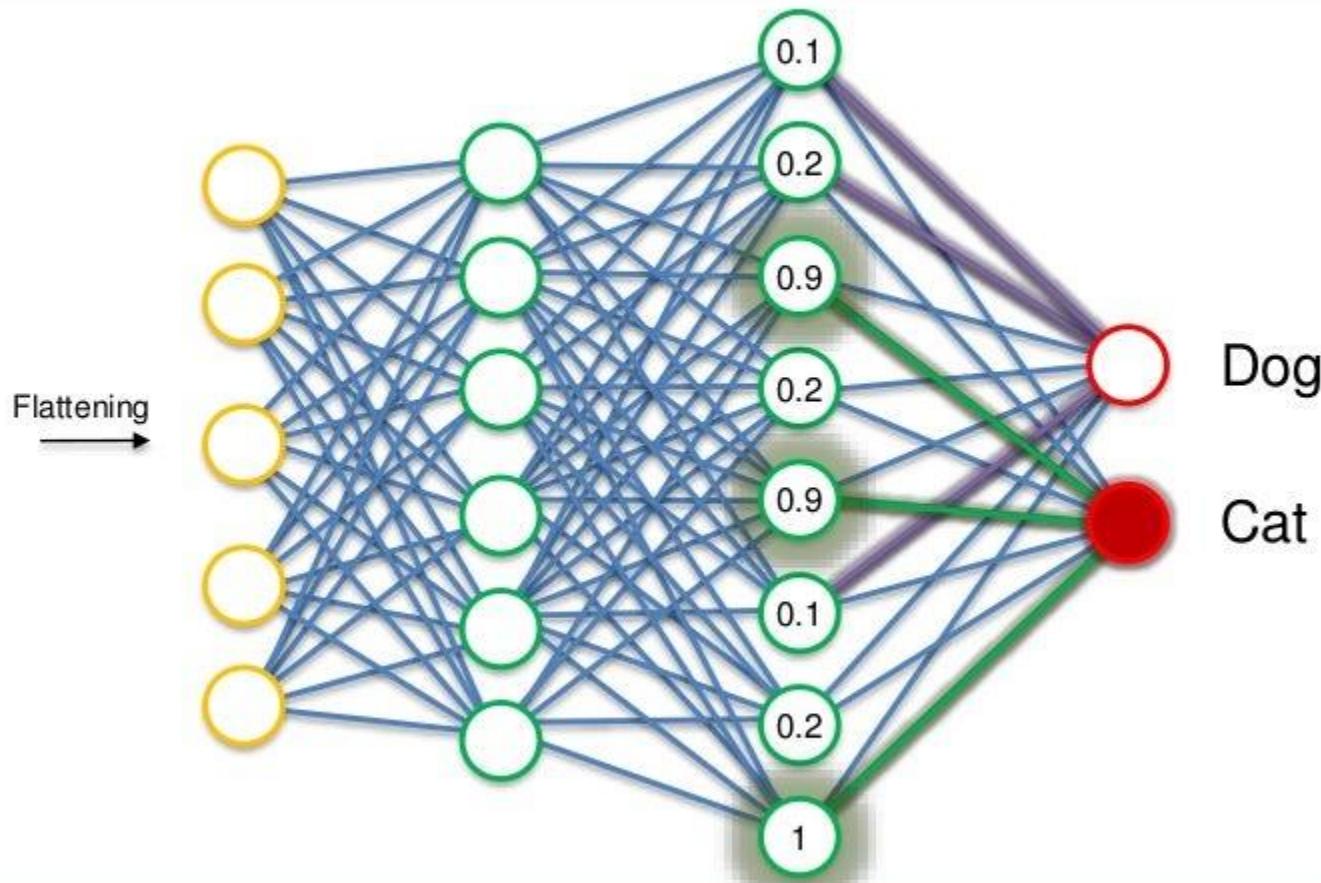
Step 4 - Full Connection



Step 4 - Full Connection



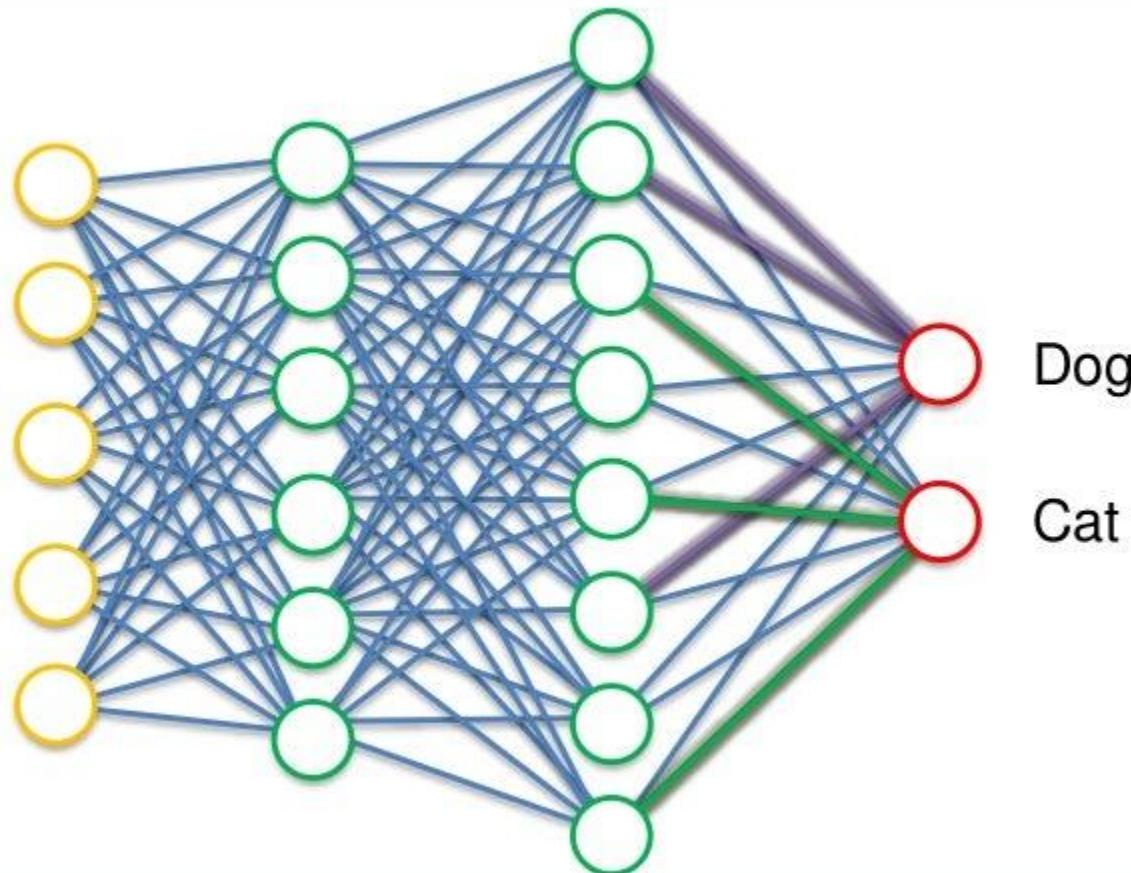
Step 4 - Full Connection



Step 4 - Full Connection



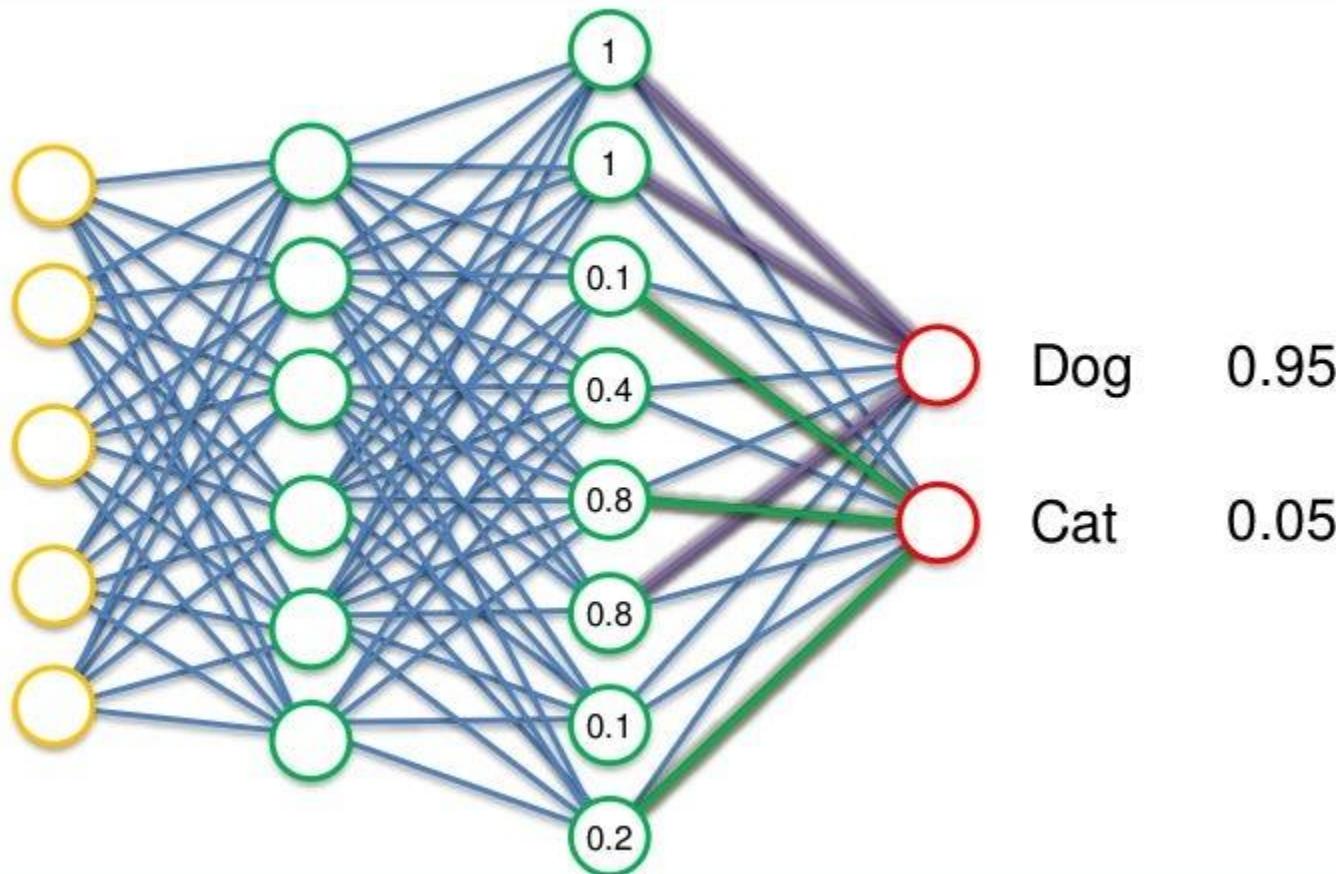
Flattening
→



Step 4 - Full Connection



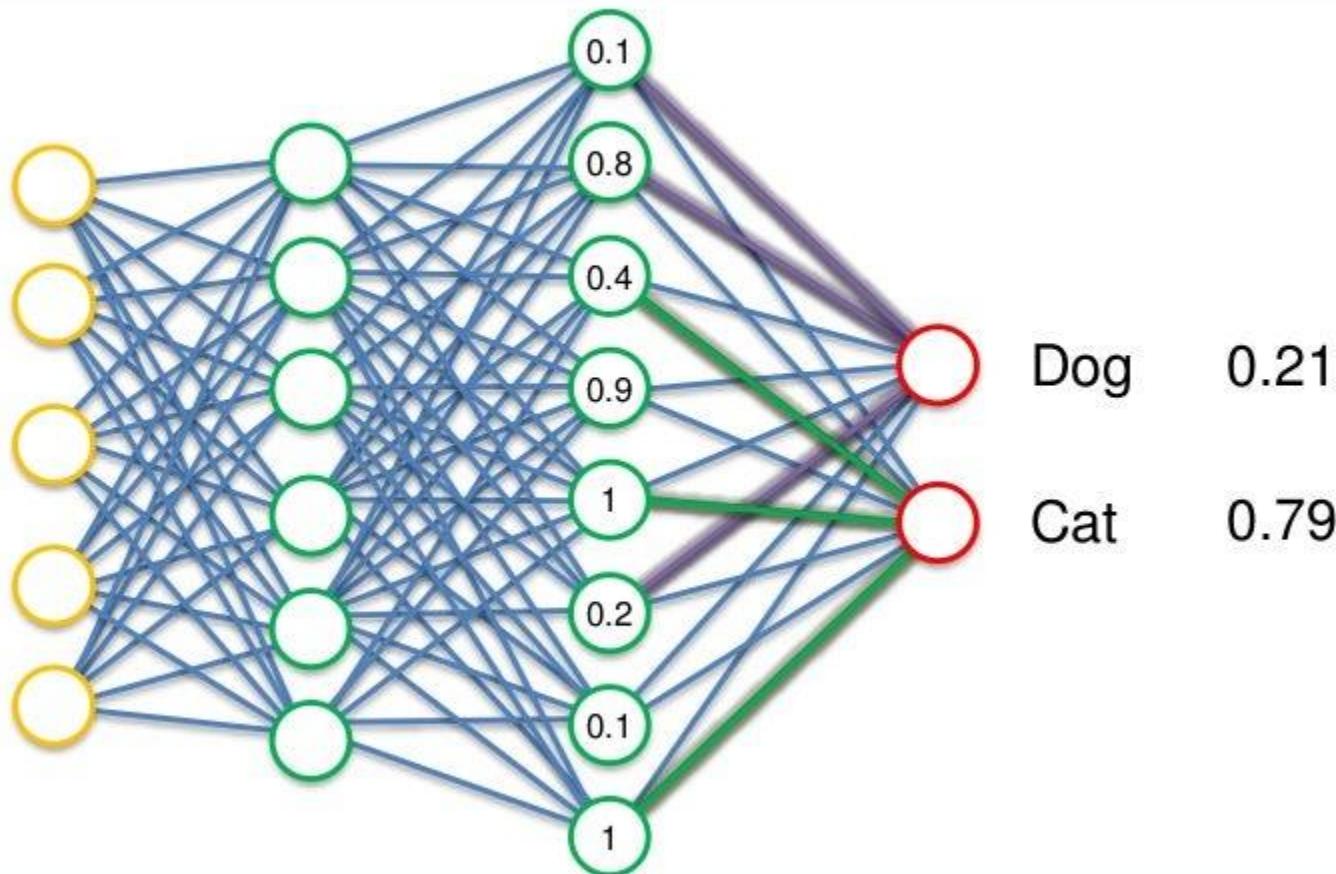
Flattening
→



Step 4 - Full Connection



Flattening
→



Step 4 - Full Connection

Examples from the test set
(with the network's guesses)



cheetah

cheetah

leopard

snow leopard

Egyptian cat



bullet train

bullet train

passenger car

subway train

electric locomotive



hand glass

scissors

hand glass

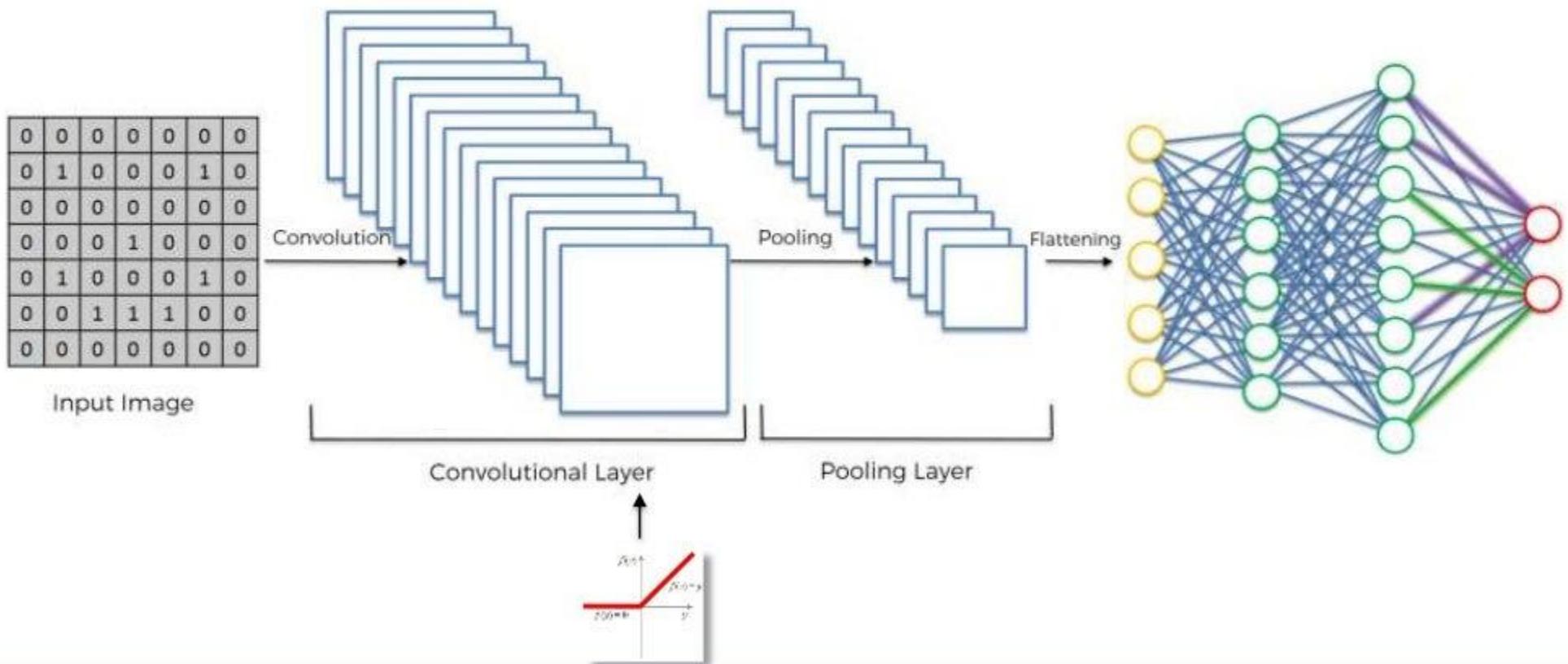
frying pan

stethoscope

Image Source: a talk by Geoffrey Hinton

Summary

Summary



Summary

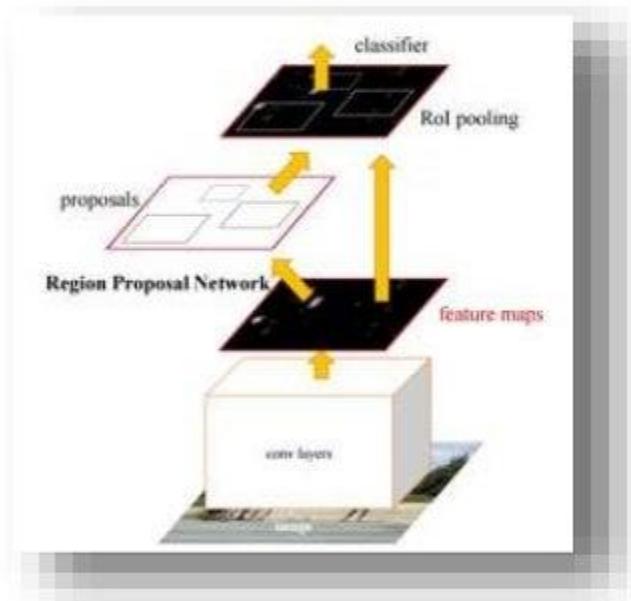
Additional Reading:

*The 9 Deep Learning Papers
You Need To Know About
(Understanding CNNs Part 3)*

Adit Deshpande (2016)

Link:

<https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html>

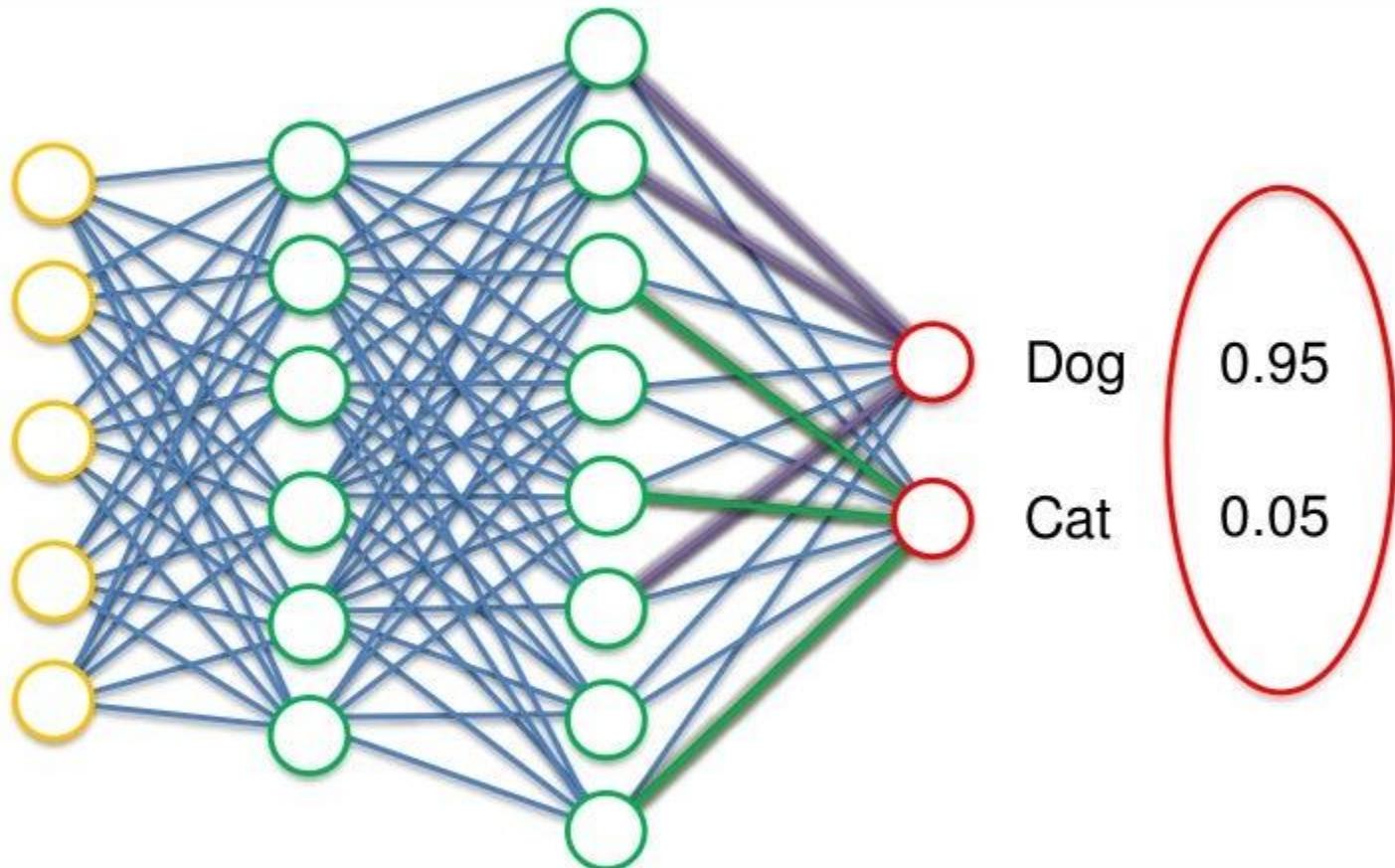


Softmax & Cross-Entropy

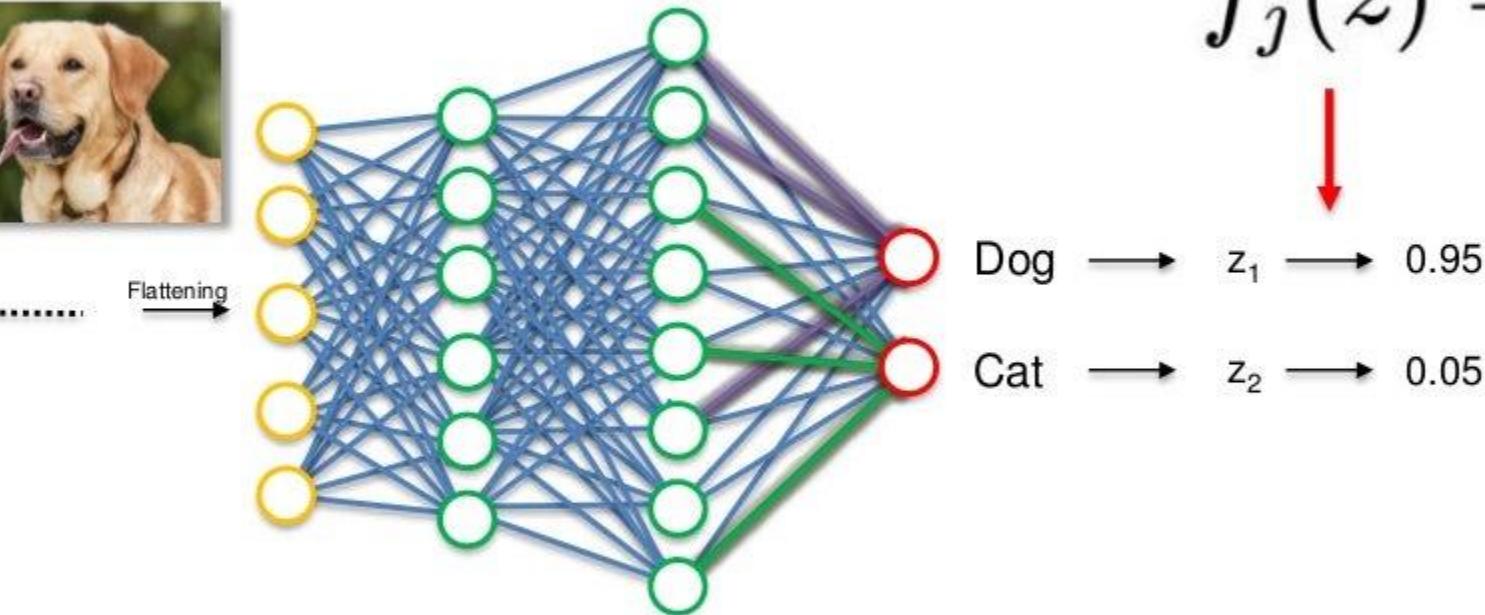
Softmax & Cross-Entropy



Flattening
→



Softmax & Cross-Entropy



$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}}$$

Dog $\longrightarrow z_1 \longrightarrow 0.95$

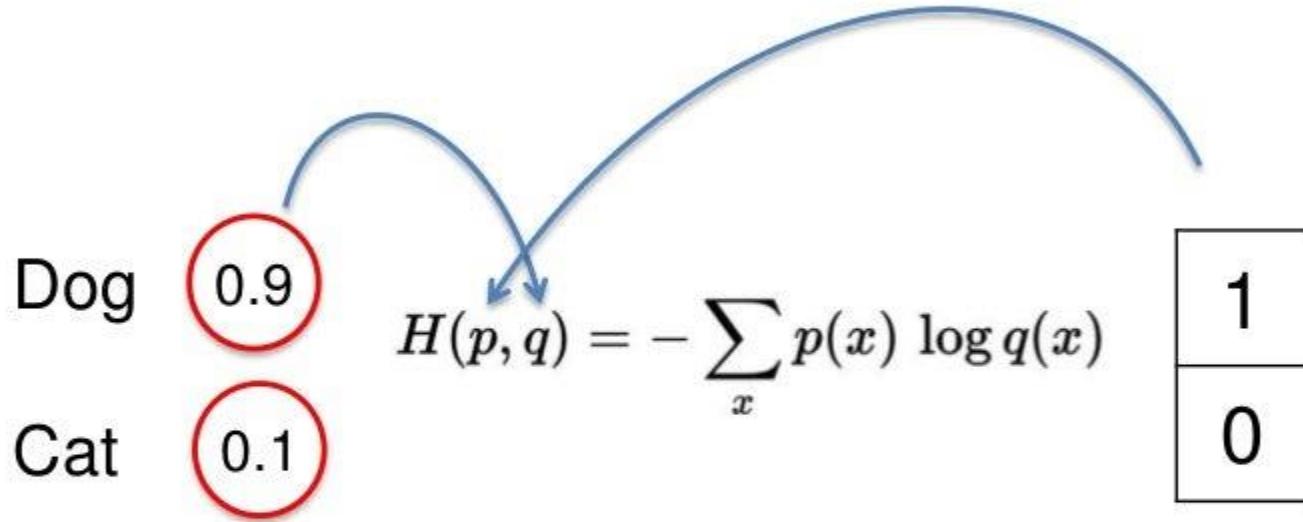
Cat $\longrightarrow z_2 \longrightarrow 0.05$

Softmax & Cross-Entropy

$$L_i = -\log \left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \right)$$

$$H(p, q) = - \sum_x p(x) \log q(x)$$

Softmax & Cross-Entropy



1
0

Softmax & Cross-Entropy

NN1 NN2



Dog	1
Cat	0



Dog	0
Cat	1



Dog	1
Cat	0



Softmax & Cross-Entropy

NN1

Row	Dog^	Cat^	Dog	Cat
#1	0.9	0.1	1	0
#2	0.1	0.9	0	1
#3	0.4	0.6	1	0

NN2

Row	Dog^	Cat^	Dog	Cat
#1	0.6	0.4	1	0
#2	0.3	0.7	0	1
#3	0.1	0.9	1	0

Classification Error

$$1/3 = 0.33$$

$$1/3 = 0.33$$

Mean Squared Error

$$0.25$$

$$0.71$$

Cross-Entropy

$$0.38$$

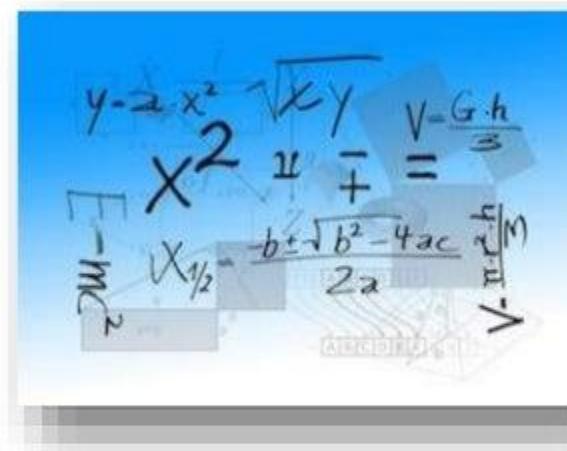
$$1.06$$

Softmax & Cross-Entropy

Additional Reading:

A Friendly Introduction to Cross-Entropy Loss

By Rob DiPietro (2016)



Link:

<https://rdipietro.github.io/friendly-intro-to-cross-entropy-loss/>

Softmax & Cross-Entropy

Additional Reading:

*How to implement a neural network
Intermezzo 2*

By Peter Roelants (2016)

$$\begin{aligned}\frac{\partial \xi}{\partial z_i} &= - \sum_{j=1}^C \frac{\partial t_j \log(y_j)}{\partial z_i} = - \sum_{j=1}^C t_j \frac{\partial \log(y_j)}{\partial z_i} = - \sum_{j=1}^C t_j \frac{1}{y_j} \frac{\partial y_j}{\partial z_i} \\ &= - \frac{t_i}{y_i} \frac{\partial y_i}{\partial z_i} - \sum_{j \neq i}^C \frac{t_j}{y_j} \frac{\partial y_j}{\partial z_i} = - \frac{t_i}{y_i} y_i (1 - y_i) - \sum_{j \neq i}^C \frac{t_j}{y_j} (-y_j y_i) \\ &= -t_i + t_i y_i + \sum_{j \neq i}^C t_j y_i = -t_i + \sum_{j=1}^C t_j y_i = -t_i + y_i \sum_{j=1}^C t_j \\ &= y_i - t_i\end{aligned}$$

Link:

http://peterroelants.github.io/posts/neural_networkImplementation_intermezzo02/