

Convolutional Neural Networks

A dark blue diagonal gradient bar that starts from the bottom left and extends towards the top right, covering the lower half of the slide.

Some Facts...

- Inspired by Hubel and Wiesel's [experiments in vision](#)
- Base on [Neocognitron](#) - 1980's
- Invented by Yann LeCun
- AlexNet's win in the 2012 ImageNet competition was a CNN architecture
- Currently the de facto image search algorithm
- Can be used in time series tasks, audio tasks like speech to text, language processing and others



Yann Lecun, co-winner of the ACM Turing Award

The Importance of Vision

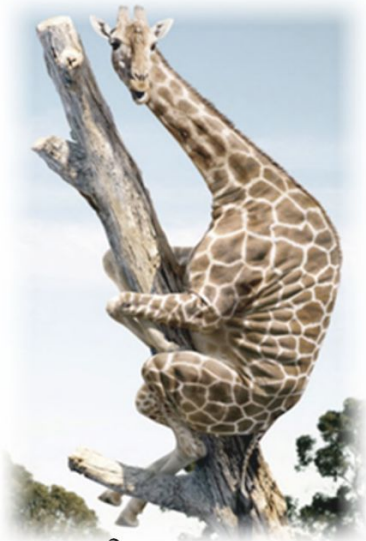
- 60% of the Brain
- Images sizes in computers and processing time
- A brain runs on 40 Watts - not much more than a lightbulb

What is Happening in this picture?

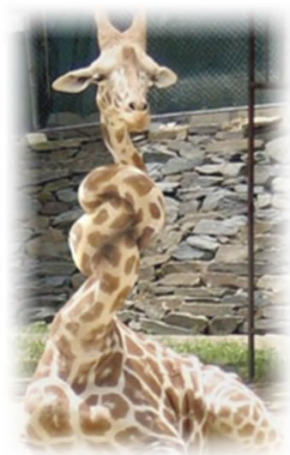
- Mirrors
- Sparse information
- 3 dimensions
- Physics
- Mental Models
- Mental Models about mental models
- Humor
- Importance of context, predicting how the scene will evolve



Describe a Giraffe

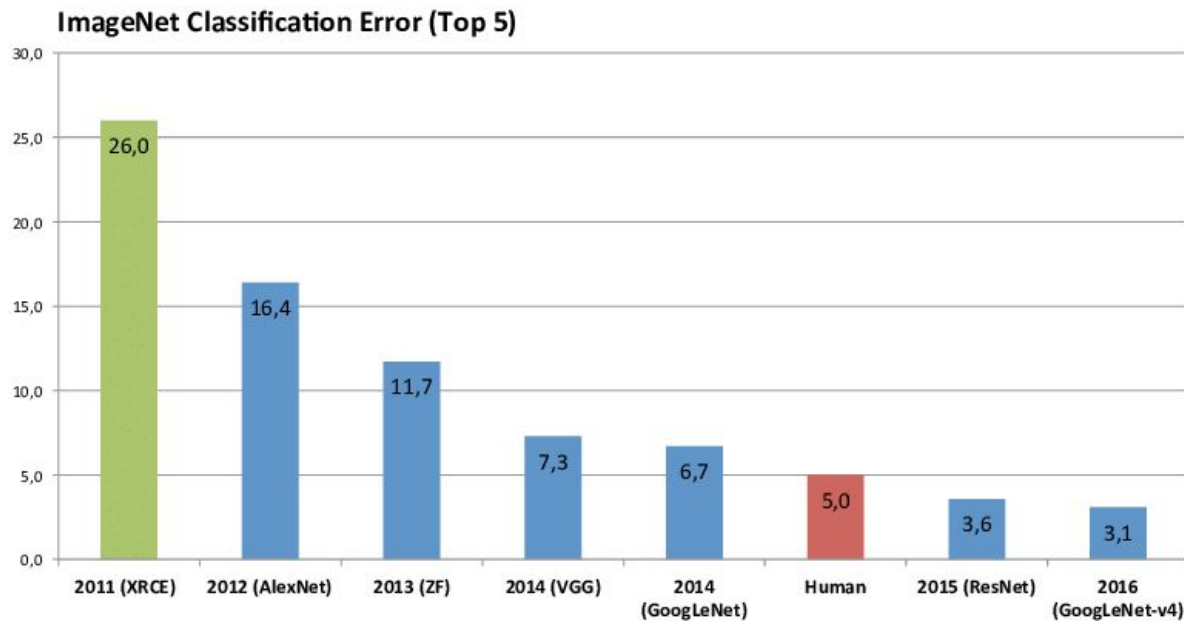


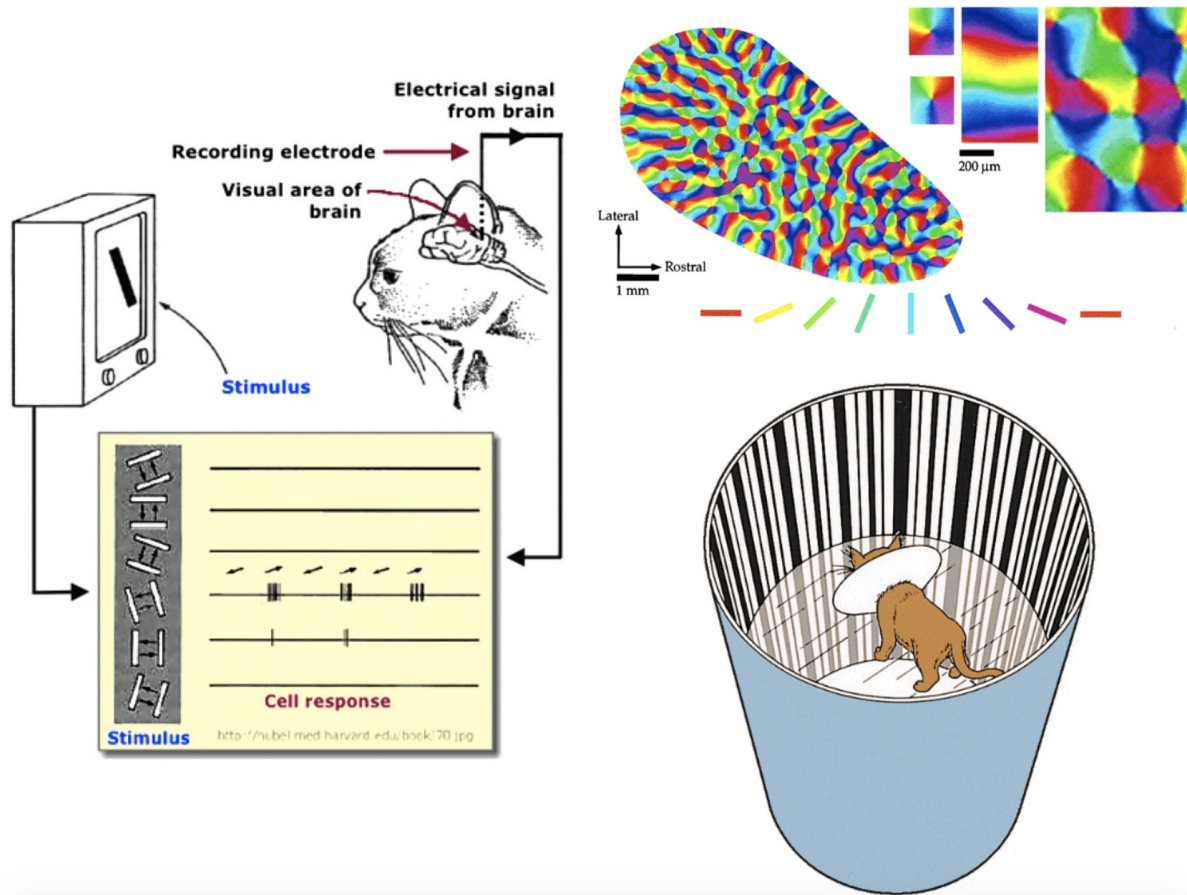
2



[Thomas Serre 2012]

ImageNet Winners since 2011





Visual Cortex for Mammals – Hubel and Wiesel

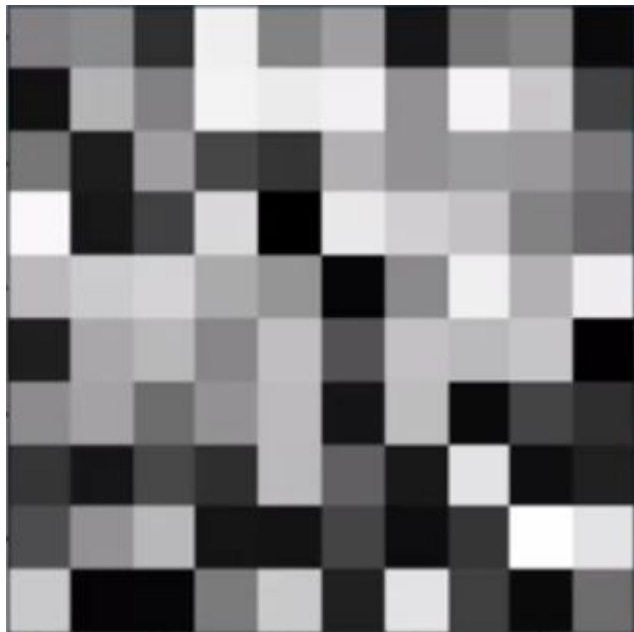
Images in Computers

Black and White



```
array([[1, 0, 1, 0, 1, 1, 1, 0, 1, 0],  
       [0, 0, 0, 0, 0, 1, 1, 1, 1, 0],  
       [0, 1, 1, 0, 0, 0, 0, 0, 1, 1],  
       [1, 1, 0, 0, 0, 0, 1, 0, 1, 0],  
       [0, 1, 1, 0, 1, 0, 0, 0, 0, 0],  
       [1, 0, 1, 0, 0, 1, 0, 0, 1, 1],  
       [1, 0, 1, 0, 1, 0, 0, 0, 1, 1],  
       [1, 1, 0, 1, 0, 0, 0, 0, 1, 1],  
       [0, 1, 1, 0, 0, 1, 0, 1, 0, 0],  
       [1, 1, 1, 1, 0, 1, 1, 0, 1, 0]])
```

Grayscale



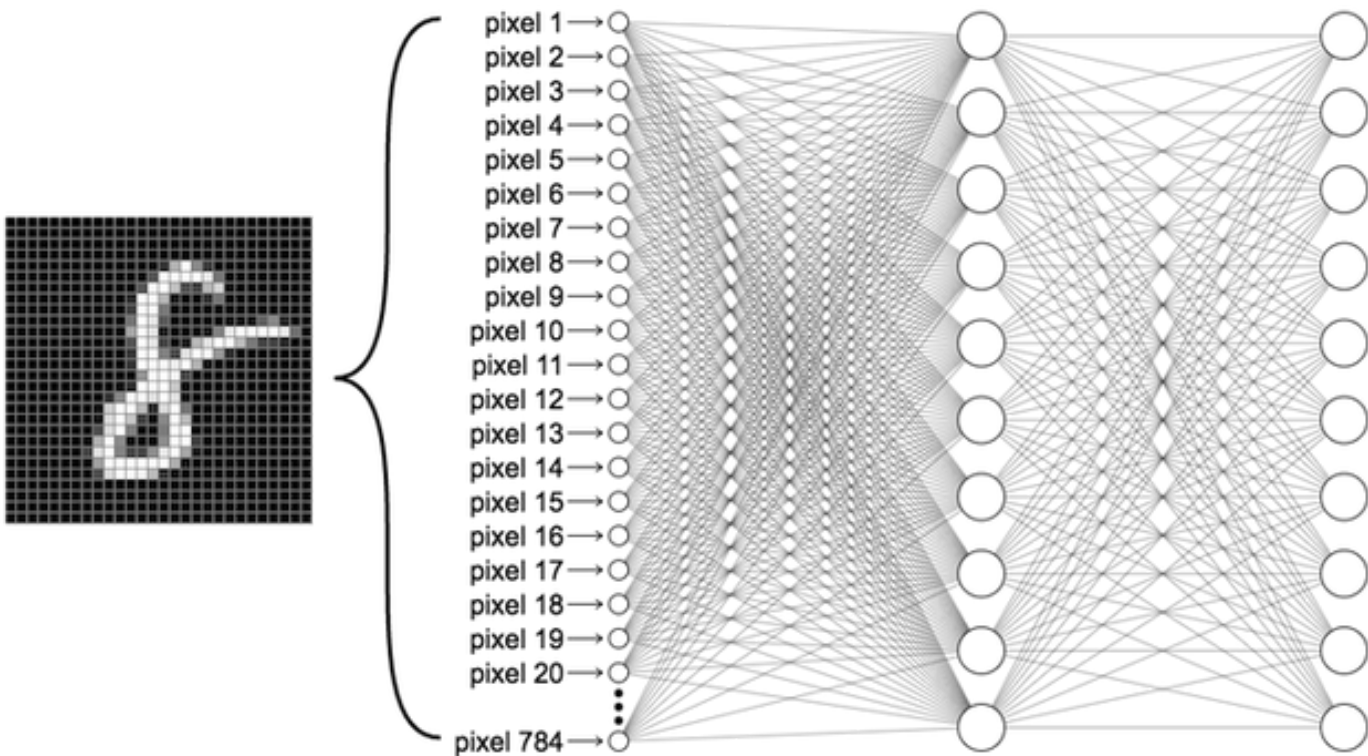
```
array([[127, 136,  48, 234, 128, 155,  24, 114, 128,  12],  
       [ 18, 177, 127, 239, 230, 236, 145, 240, 195,  65],  
       [116,  29, 155,  70,  51, 174, 144, 153, 149, 120],  
       [243,  26,  66, 213,   1, 227, 203, 191, 128, 105],  
       [183, 199, 207, 167, 146,   7, 136, 234, 175, 232],  
       [ 31, 166, 180, 133, 187,  82, 192, 182, 194,   2],  
       [139, 161, 108, 144, 183,  23, 185,  12,  67,  47],  
       [ 53,  24,  72,  46, 182,  94,  25, 221,  16,  34],  
       [ 76, 144, 181,  22,  21,  67,  16,  53, 250, 222],  
       [197,   6,   7, 122, 198,  34, 222,  64,  12, 108]])
```

MNIST



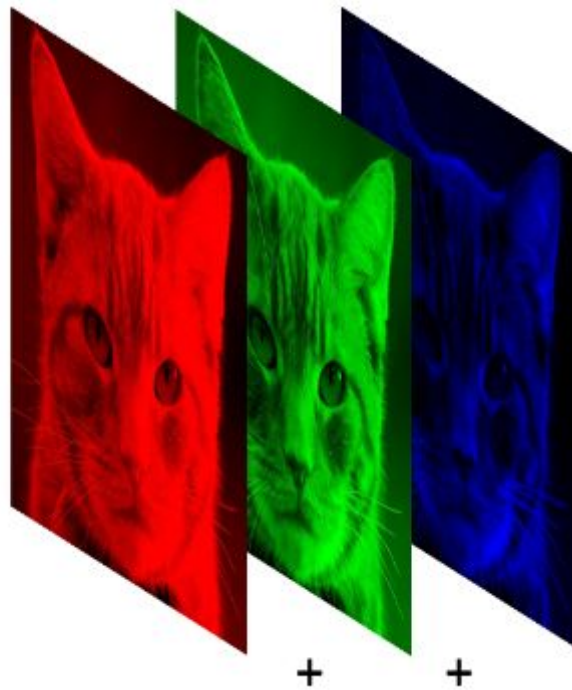
5 0 4 1

MNIST in Dense Networks



Order	Name	Example	Shape
0	Scalar	3	No Shape
1	Vector	[4, 6, 0, 5, 7, 1]	(6,)
2	Matrix	[[0, 1, 2], [6, 7, 8]]	(2, 3)
3	Tensor	[[[0, 1, 2], [6, 7, 8]], [[0, 1, 2], [6, 7, 8]]]	(2,2,4)

RGB

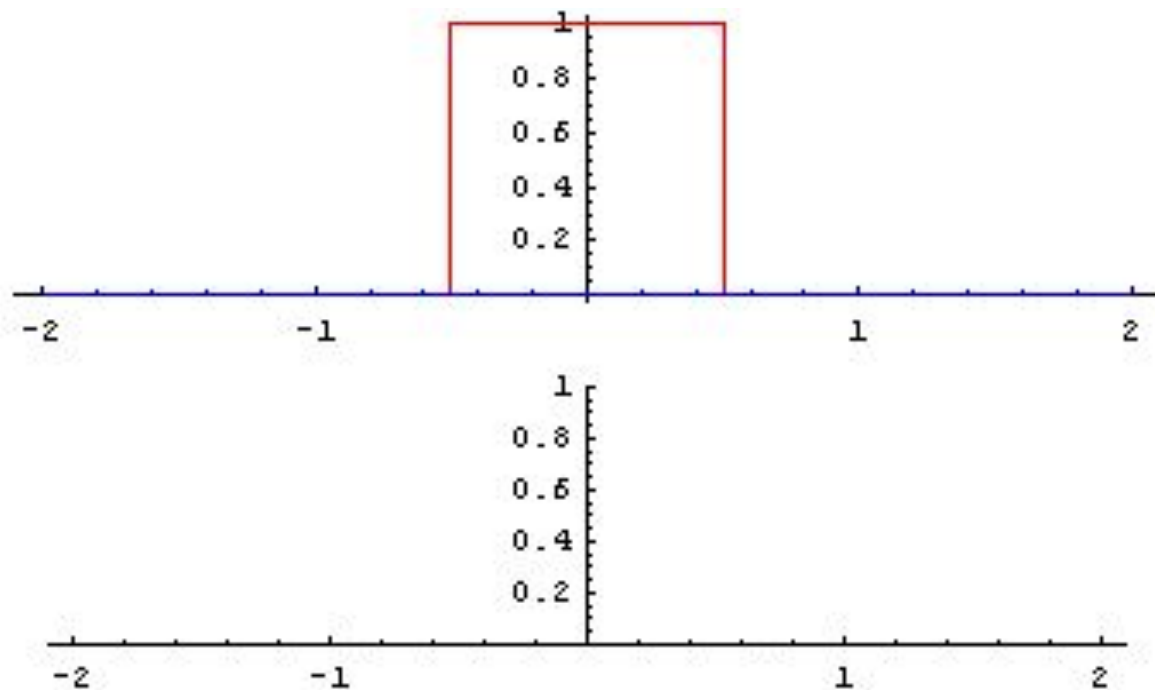


So what is a Convolution?

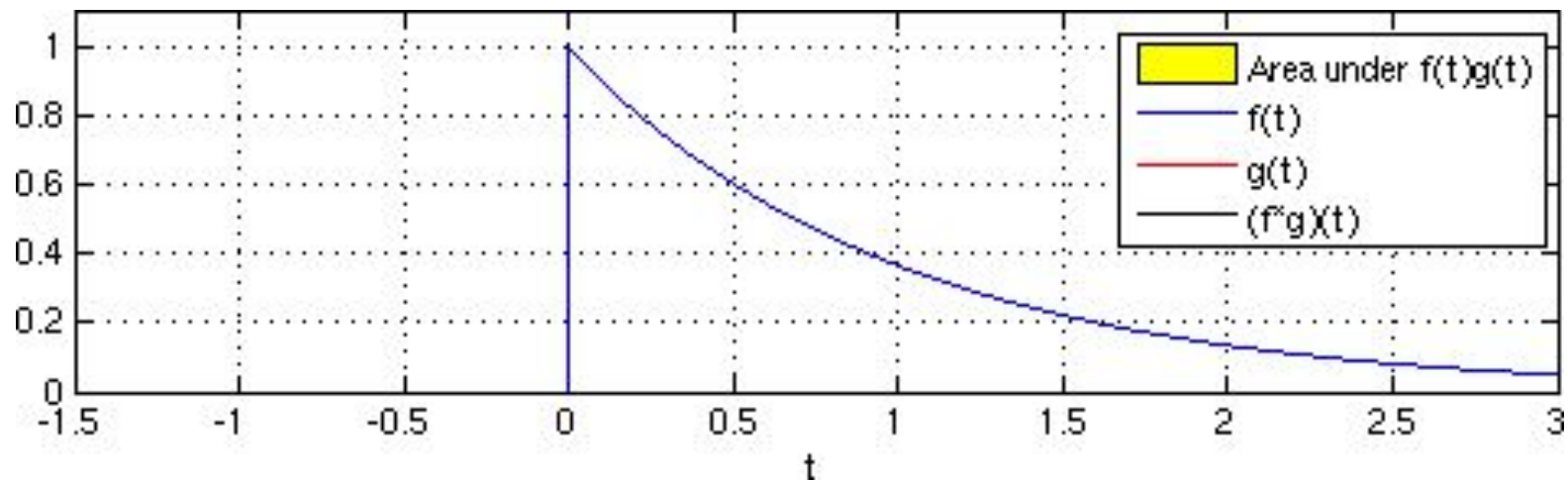
1D Convolution



1D Convolution



1D Convolution



Code...

2D Convolution

Reference Image

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

Kernel or Filter

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

Result

0.77	-0.11	0.11	0.33	0.55	-0.11	0.33
-0.11	1.00	-0.11	0.33	-0.11	0.11	-0.11
0.11	-0.11	1.00	-0.33	0.11	-0.11	0.55
0.33	0.33	-0.33	0.55	-0.33	0.33	0.33
0.55	-0.11	0.11	-0.33	1.00	-0.11	0.11
-0.11	0.11	-0.11	0.33	-0.11	1.00	-0.11
0.33	-0.11	0.55	0.33	0.11	-0.11	0.77

Code...

Convolutional Layer

The input is an order 4 tensor

(N, H, W, C)

N: Number of Images

H: Height of Images

W: Width of Images

C: Number of color channels



(60000, 28, 28, 1)

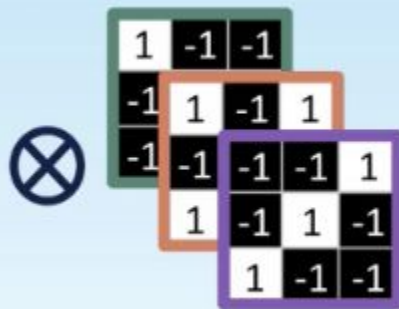
Convolutional Layer

Layer Input



(-, 9, 9, 1)

Layer Kernels



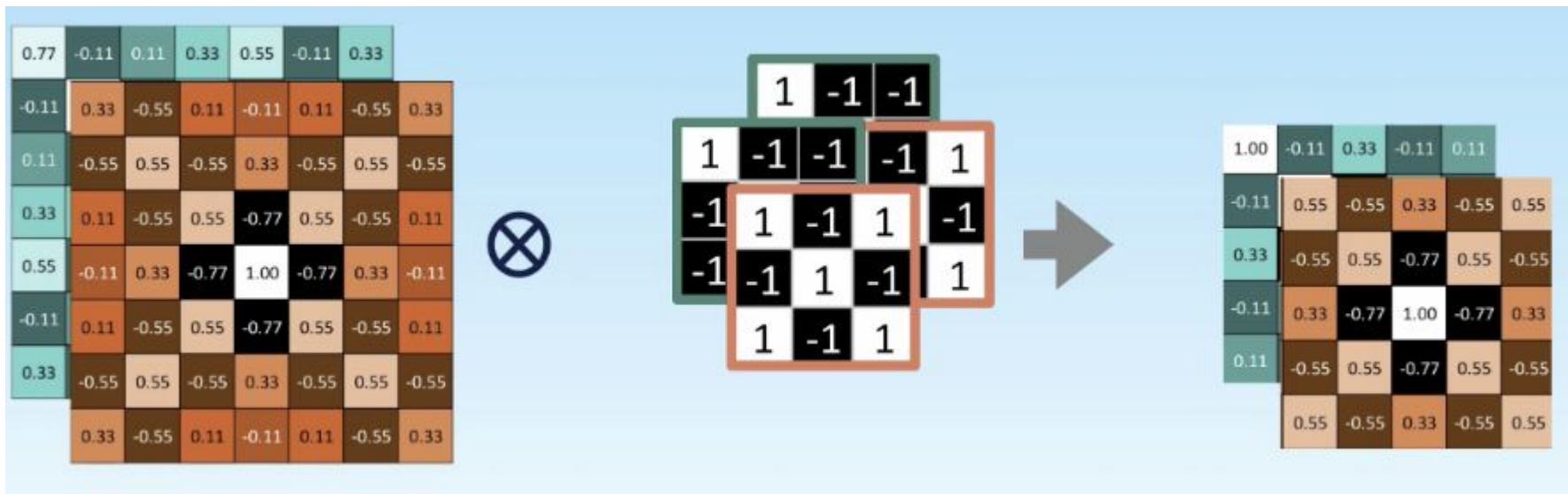
(3, 3, 1, 2)

Layer Output



(-, 7, 7, 2)

Next Layer



(3, 3, 1, 2)

Code...

Strides: (2, 2)

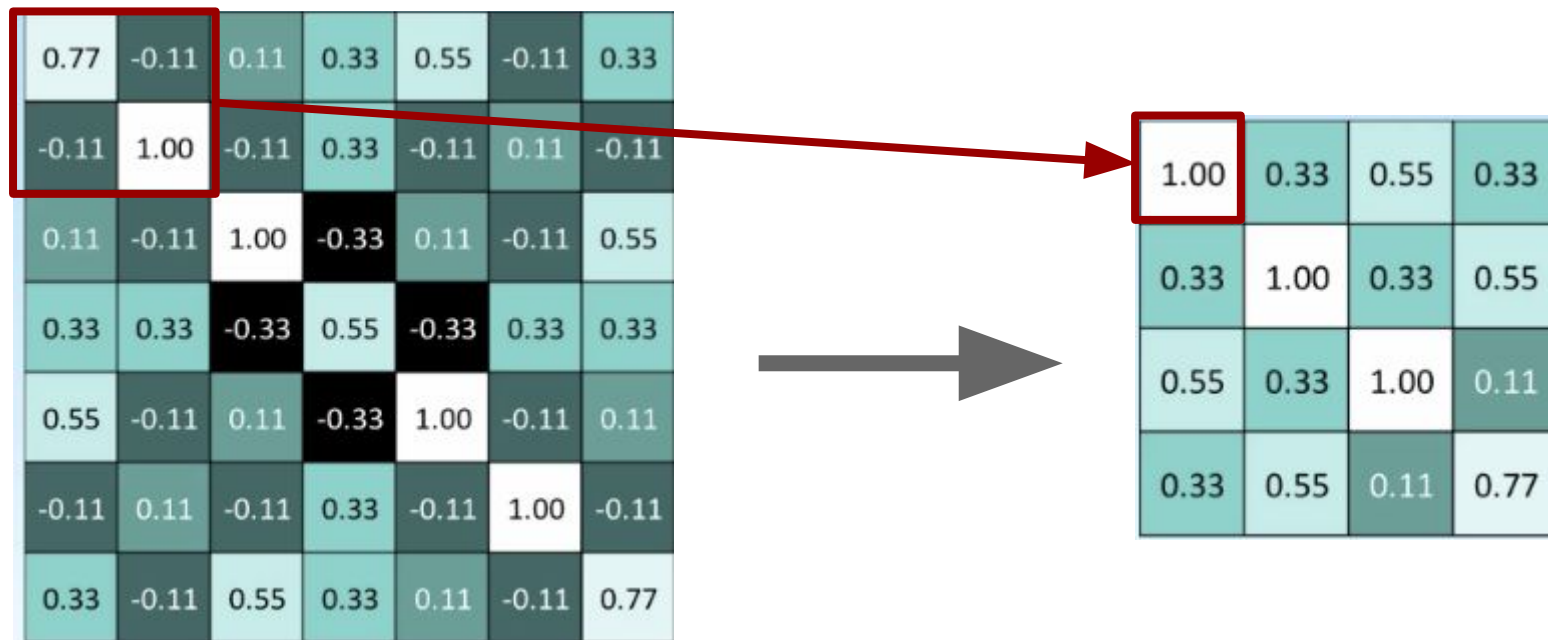
[illegible][illegible][illegible]

Padding Layer

- **Valid**
- **Same**

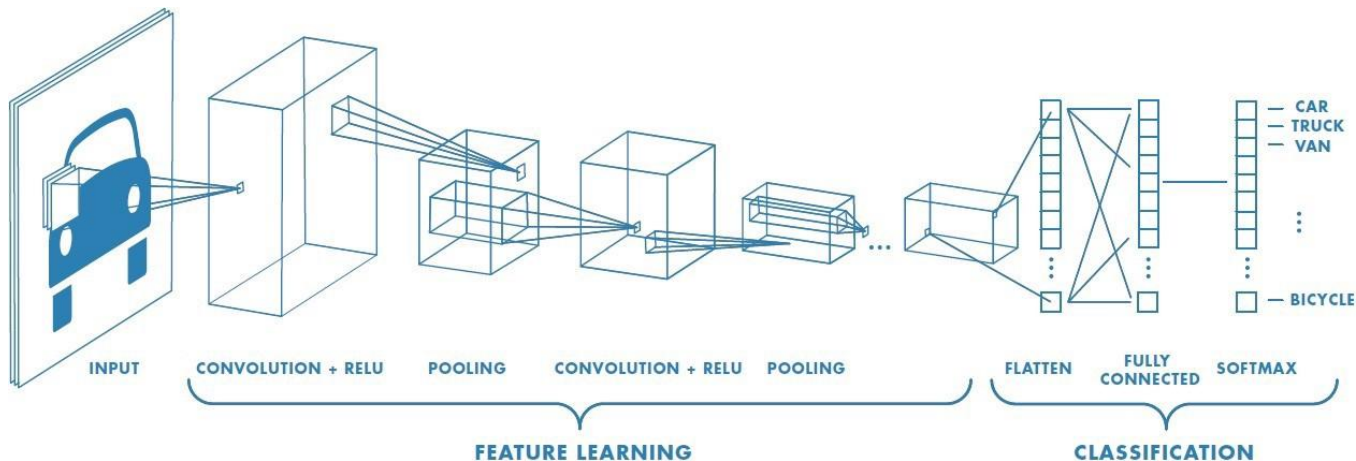
[illegible]

Pooling Layer



Code...

Convolutional Neural Networks... Finally



Comparison: FC vs CNN

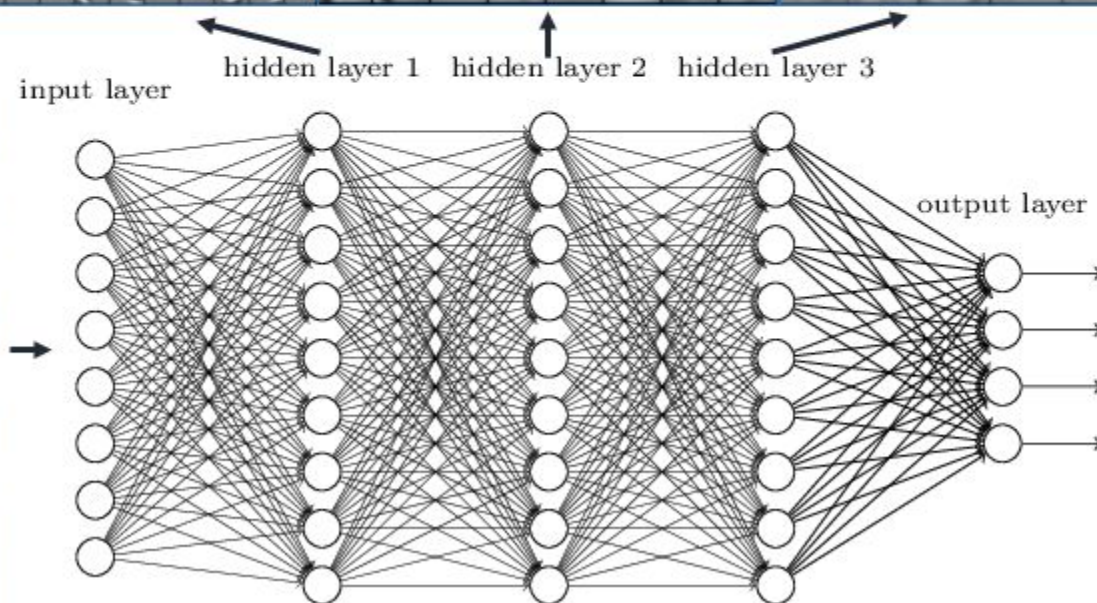
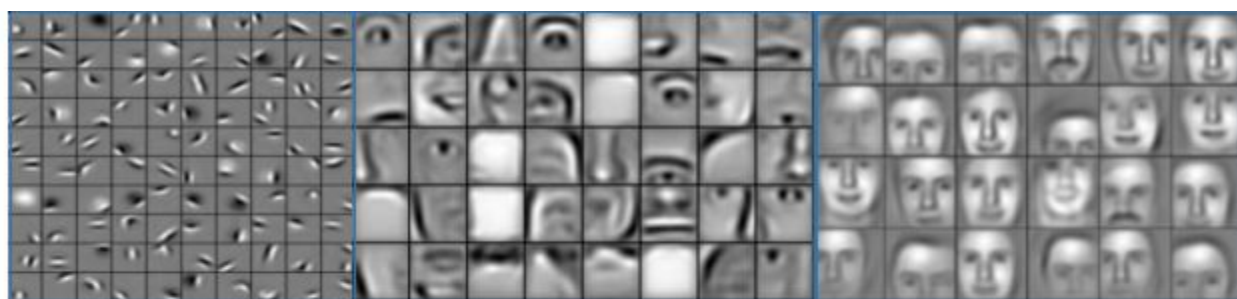
Fully Connected

- Input data: image $10 \times 10 \Rightarrow 100$ numbers
- First Layer: 32 nodes $\Rightarrow 100$ weights
- Total $\Rightarrow 3200$ weights

Convolutional

- Input data: image $10 \times 10 \Rightarrow 100$ numbers
- First Layer: 32 conv, kernels $3 \times 3 \Rightarrow 9$ weights each
- Total $\Rightarrow 32 \times 9 = 288$ weights

Deep neural networks learn hierarchical feature representations



Feature Hierarchy

Code...

Famous Architectures

Names:

- LetNet-5
- AlexNet
- GoogLeNet
- VGGNet
- ResNet
- Xception
- SENet