ntic master
revolucionamos la comunicación

UNIVERSIDAD COMPLUTENSE DE MADRID

Deep Learning

Alberto Ezpondaburu

Analytical Index

Convolutional Neural Networks (CNN)

a. Introduction to Computer Vision
b. Introduction to CNN

2

Convolutional Neural Networks (CNN)

2.1

Introduction to Computer Vision

Computer Vision I

- Massive growth in visual data:
  - 500 hours of video are uploaded to YouTube every minute.
  - TikTok has over 1 billion monthly active users that spend 1.5 hours on average per day.
  - 100 million post and 1 billion stories are uploaded to Instagram every day.
  - There are cameras everywhere.
- Techniques and ideas can be used in other types of ordered data.
- People create and share visual data continuously.

Computer Vision II

“Building artificial systems that process, perceive and reason about visual data”

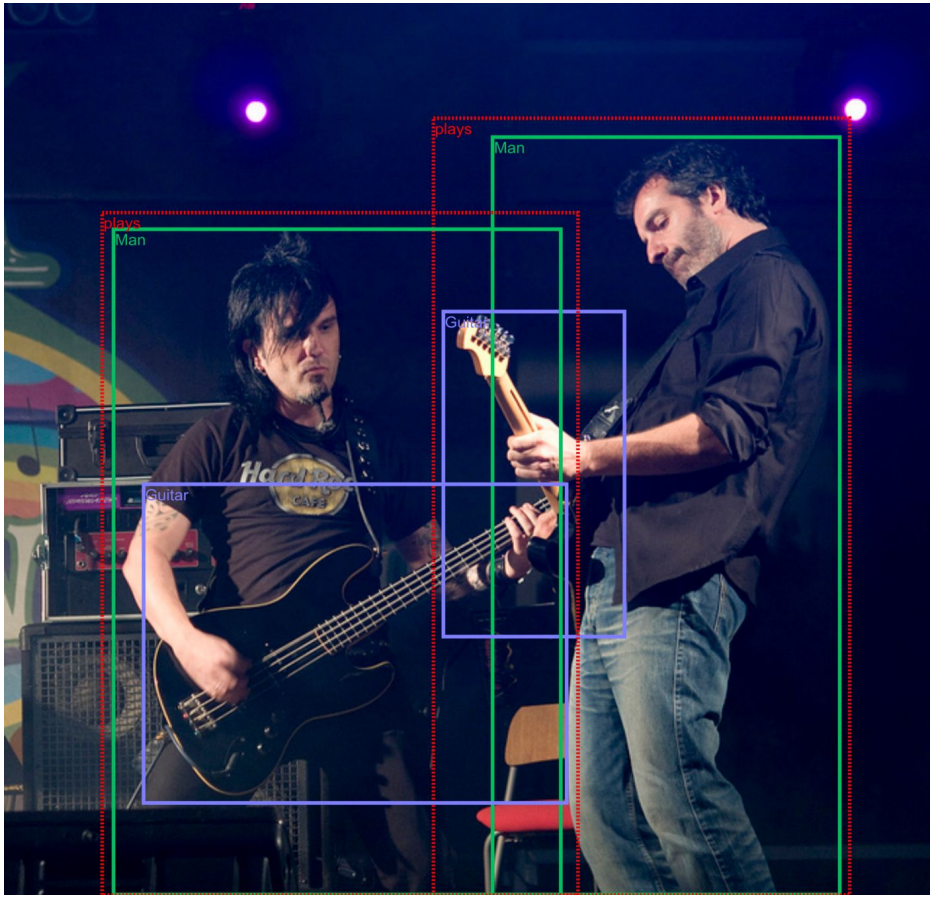
Visual Perception: 500 millions of years of evolution, involves both the evaluation of stimulus features and attention (differential processing of selected information that is relevant to behavior)

Importance of computer vision:

1. Automated analysis and understanding of visual data to enhance human capabilities.
2. Enables applications in diverse industries, including healthcare, automotive, retail, security, digital marketing…

Computer Vision: Image Classification & Location

Computer Vision: Relationship Detection



Computer Vision: Pose Estimation

Determines the position and orientation of objects or people, useful in AR, gaming, and healthcare

Computer Vision: Object Tracking

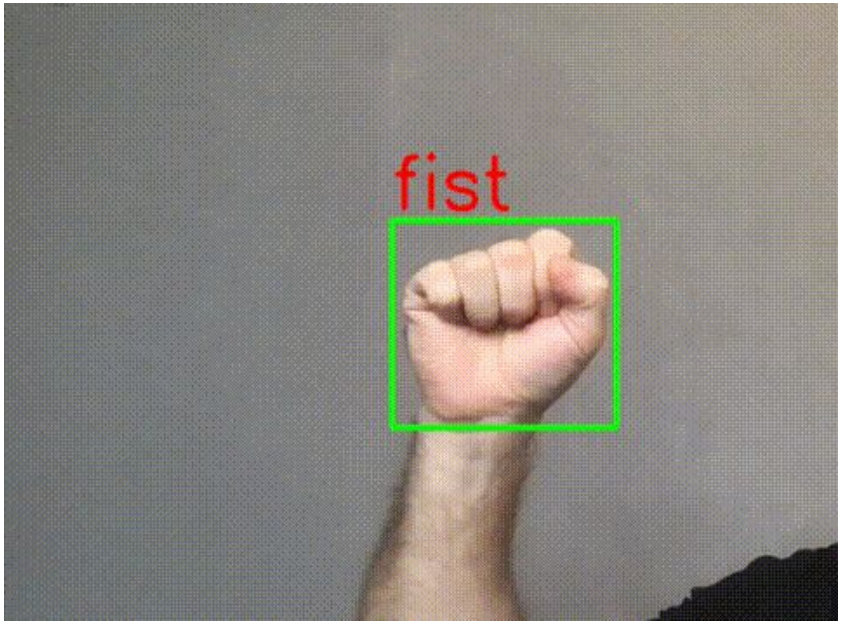
Identifies and locates objects within images or video streams, used in retail, security, and many other sectors

Computer Vision: Self-Driving Cars

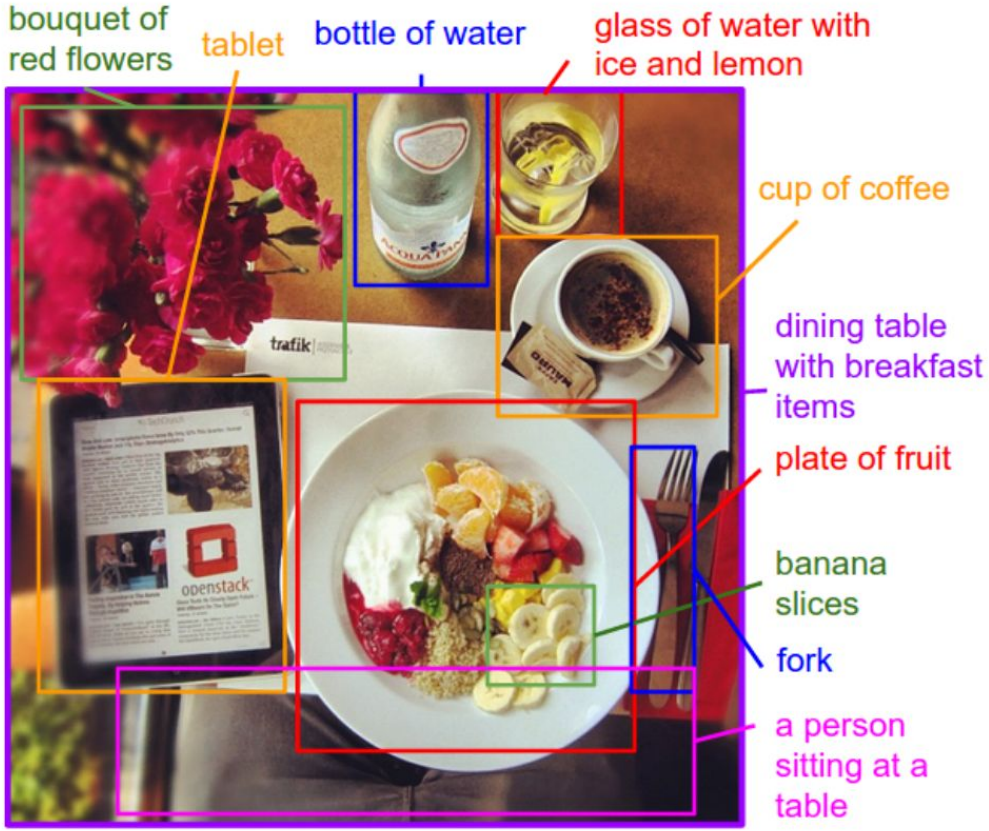
Identifies and locates objects within images or video streams, used in retail, security, and many other sectors

Computer Vision: Gesture Recognition

Interpreting human gestures to control devices or applications, used in gaming, healthcare, and smart home systems.

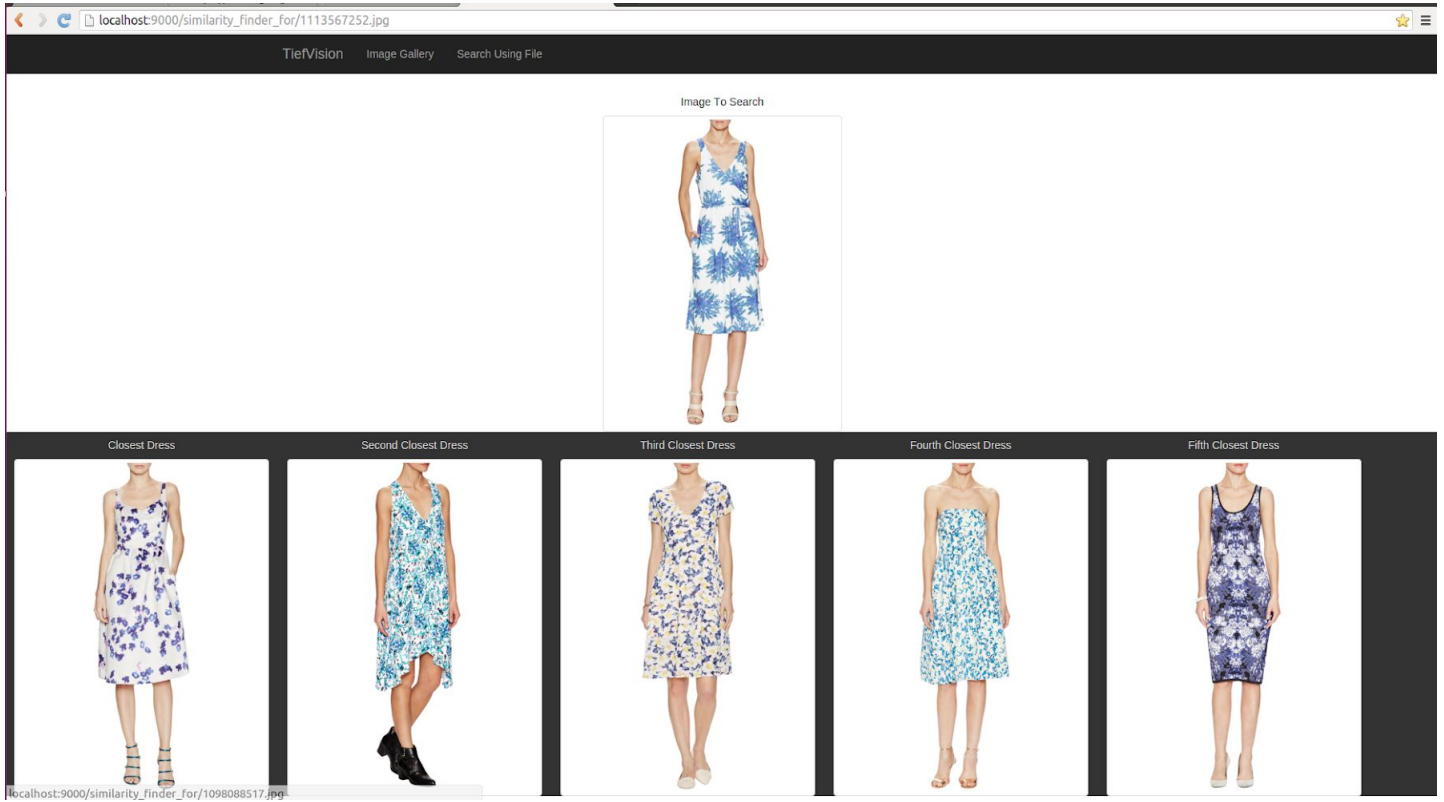


Computer Vision: Image annotation



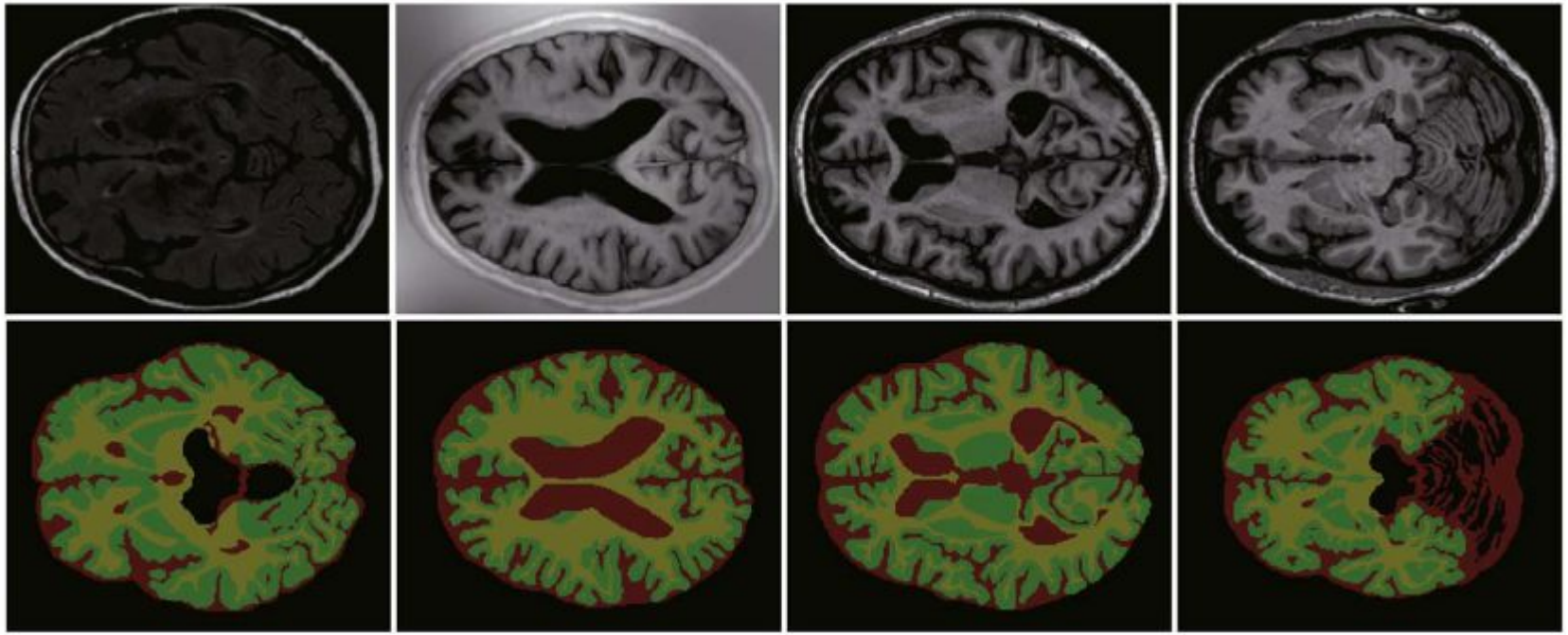
"Deep Visual-Semantic Alignments for Generating Image Descriptions", 2015.

Computer Vision: Image Similarity Search Engine



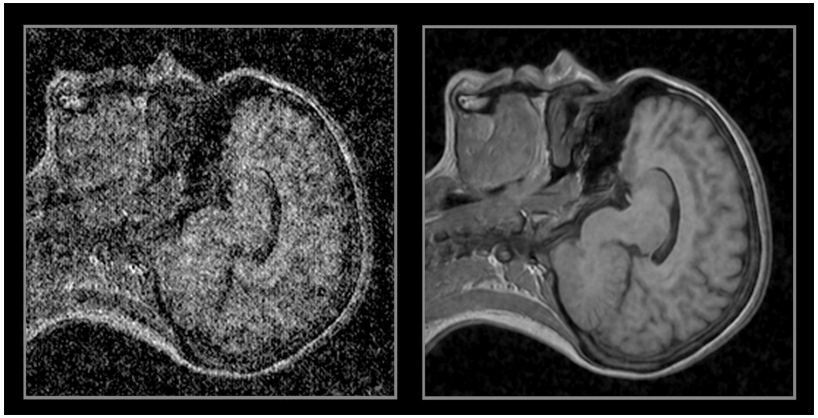
https://github.com/paucarre/tiefvision

Computer Vision: Image segmentation



Computer Vision: Image Reconstruction

Computer Vision: Image Denoising

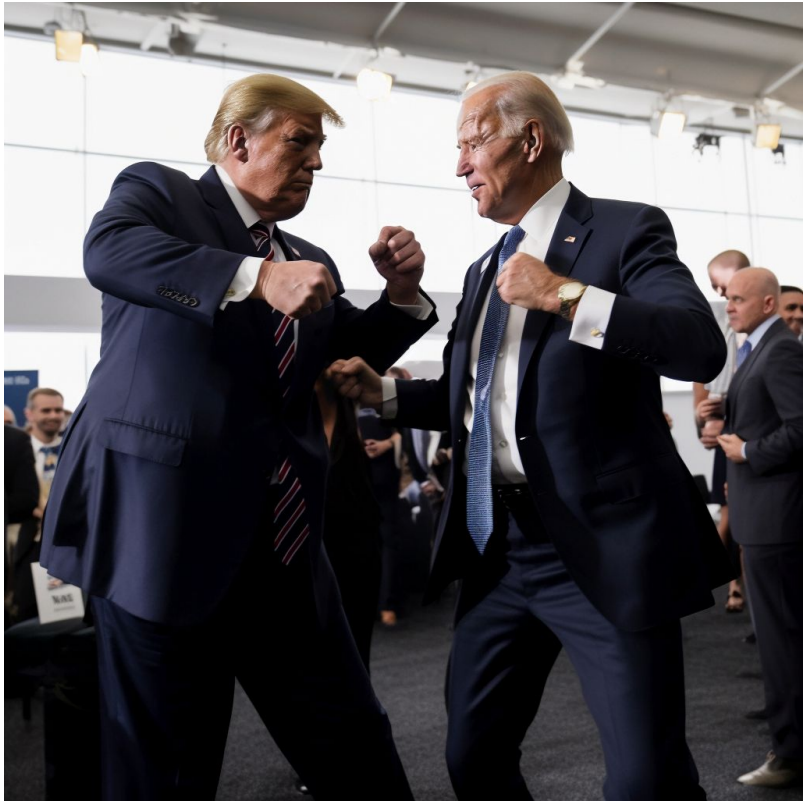


Computer Vision: Image search

```
image_search('Find me an image of a famous monument in India',

Input query: Find me an image of a famous monument in India

0.3326866328716278
```

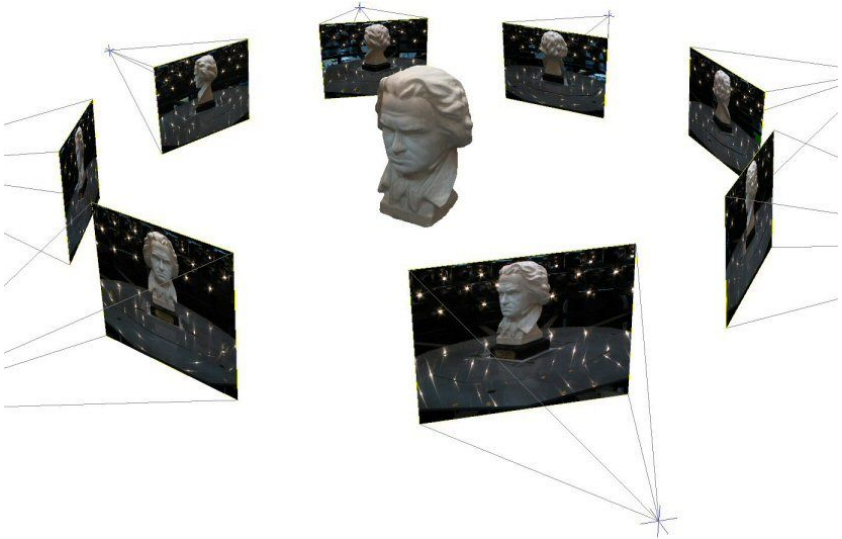
```
image_search_from_path('Two dogs playing in the snow', model, img_embeddings, img_

Input query: Two dogs playing in the snow

0.31620916724205017
```

Computer Vision: Image Generation



Computer Vision: 3D Reconstruction

Constructs 3D models from 2D images, used in virtual reality, architecture, and entertainment



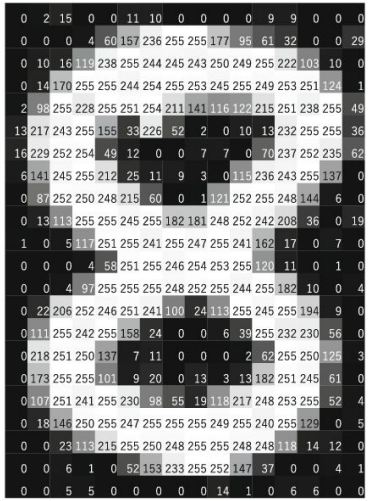
Computer Vision: Super Resolution

Enhances the resolution of images, making them clearer, which is widely used in satellite imaging, medical imaging, and surveillance

2.2

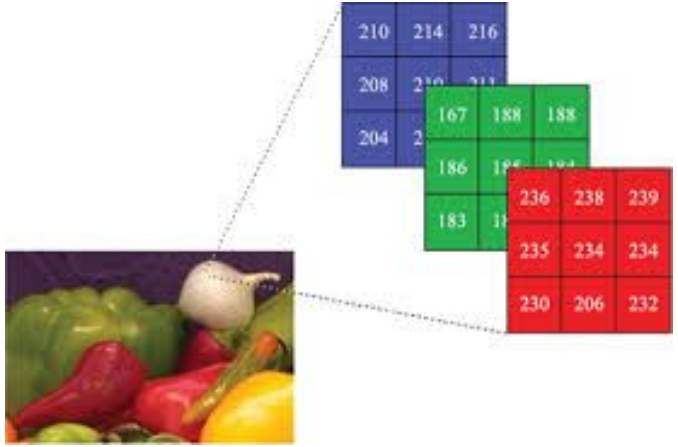
Introduction CNN:

Image Data

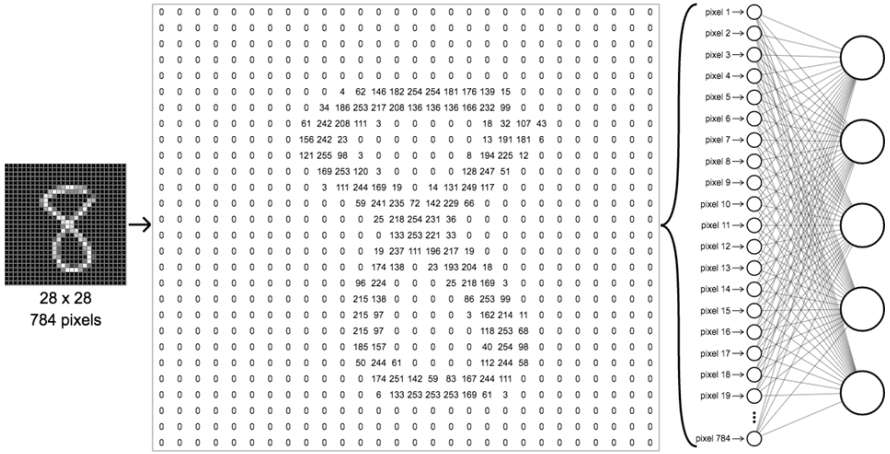


- Images are composed by pixels. 1 Megapixel can be a 1000x1000 matrix.
- Grayscale images can be seen as matrices of integers (0 -255) (black - white).

- Coloured images are usually represented as mixes of three colours: Red, Green and Blue (RGB)



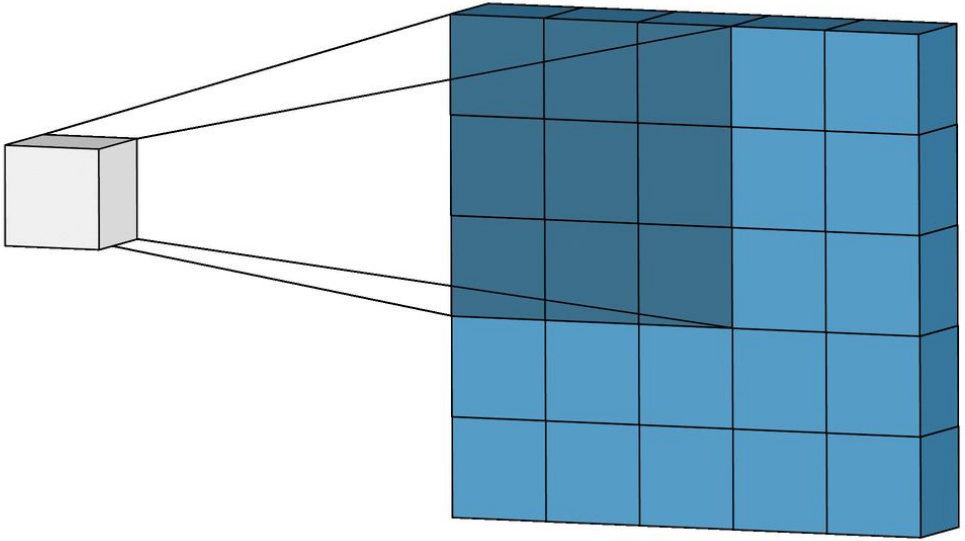
Problem with Dense Layers



- 1 Megapixel = 1000x1000 = 10^6 inputs.
- 1000 units hidden layer = 10^9 parameters!!
- No spatial information. Collapse 2D => 1D
- In most tasks, you can share the same operations in different parts of the image, In a fully-connected layer you learn pixel-by-pixel operations.

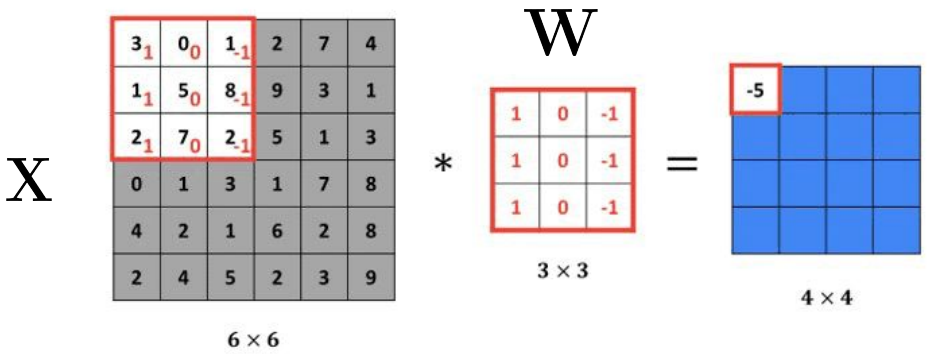
- Spotify: 44.1kHz sample rate.
- Same problems.

Convolution Operation: Intuition



- Connect input patch to a single neuron to capture spatial information
- Use sliding window to order connections.
- 3x3 filter, stride = 1 (shift by 1 pixel the sliding window)

Convolution Operation: Filtering



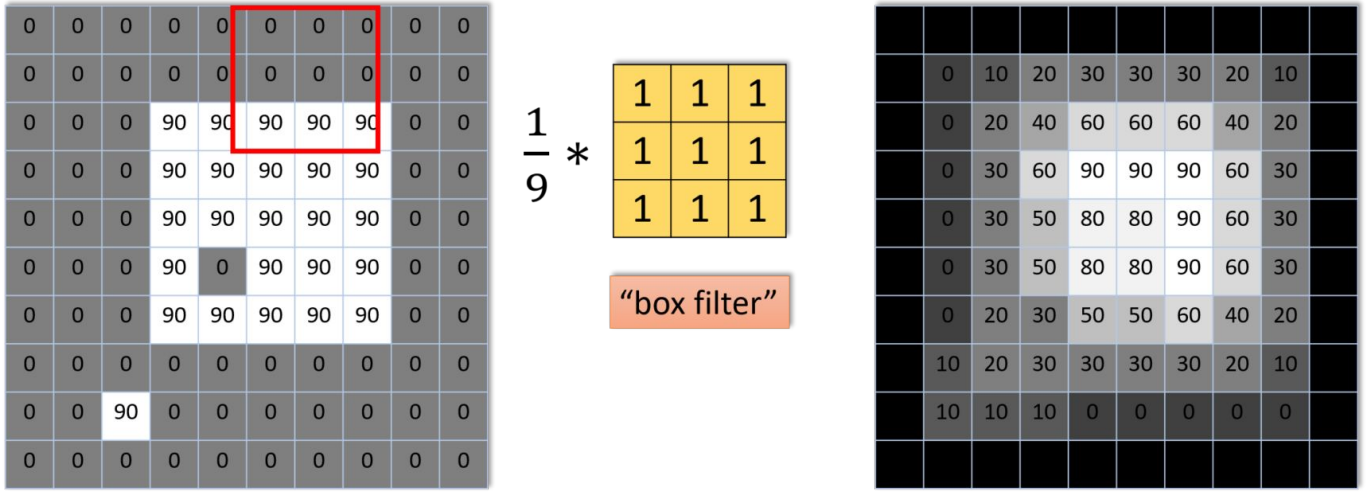
$$\mathbf{Y} = \mathbf{X} \star \mathbf{W}$$

$$y_{p,q} = \sum_{i=0}^{2} \sum_{j=0}^{2} w_{ij} x_{i+p,j+q}$$

$$3 \times 1 + 1 \times 1 + 2 \times 1 + 0 \times 0 + 5 \times 0 + 7 \times 0 + 1 \times -1 + 8 \times -1 + 2 \times -1 = -5$$

- In mathematics it is called Cross-correlation, In Deep learning convolution.
- Operator: *
- Dimensions: $(n \times n) \star (f \times f) \equiv (n - f + 1) \times (n - f + 1)$

Convolution Operation: Example



Convolution Operation: Example

10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0

6 x 6

*

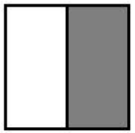
1	0	-1
1	0	-1
1	0	-1

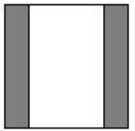
3 x 3

=

-0	30	30	0
0	30	30	0
0	30	30	0
0	30	30	0

4 x 4





Convolution Operation: Example

255	255	255	255	255	255
255	255	255	255	255	255
255	255	255	255	255	255
255	255	255	255	255	255
10	10	10	10	10	10
10	10	10	10	10	10
10	10	10	10	10	10
10	10	10	10	10	10

8 X 6

*

1	1	1
0	0	0
-1	-1	-1

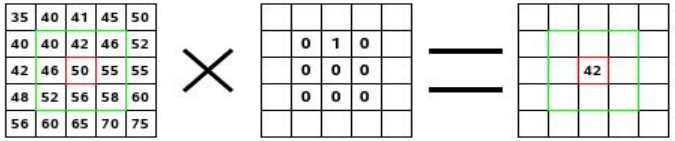
3 X 3

=

6 X 4

0	0	0	0
0	0	0	0
735	735	735	735
735	735	735	735
0	0	0	0
0	0	0	0

Convolution Operation: Example



Edge Detection

	0	1	0	
	1	-4	1	
	0	1	0	

Edge Detection (right)

	0	0	0	
	-1	1	0	
	0	0	0	

Convolution Operation: Example

Sharpen

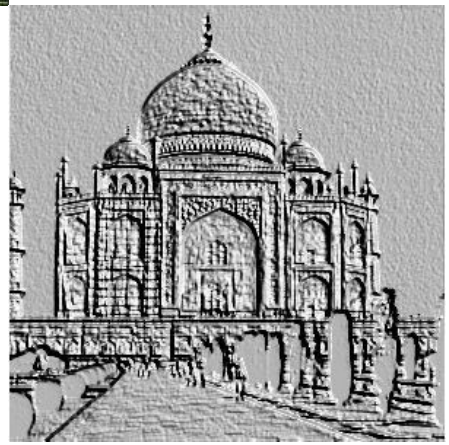
0	0	0	0	0
0	0	-1	0	0
0	-1	5	-1	0
0	0	-1	0	0
0	0	0	0	0

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

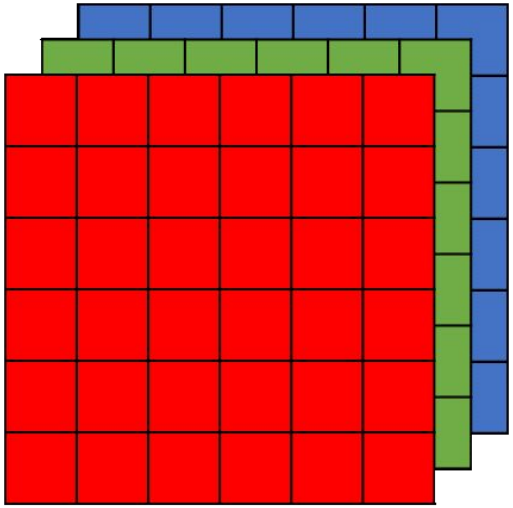
Emboss

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



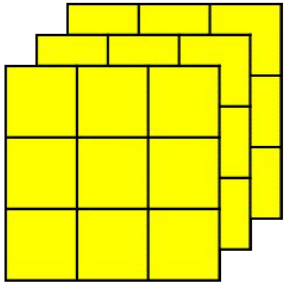
Convolution Operation: RGB

3 separate convolutions and then sum all 3



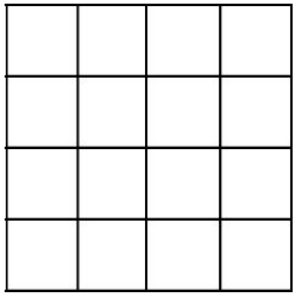
$6 \times 6 \times 3$

*



$3 \times 3 \times 3$

=



4×4

Convolution: Padding & Stride

- The image shrinks every time a convolution operation is perform

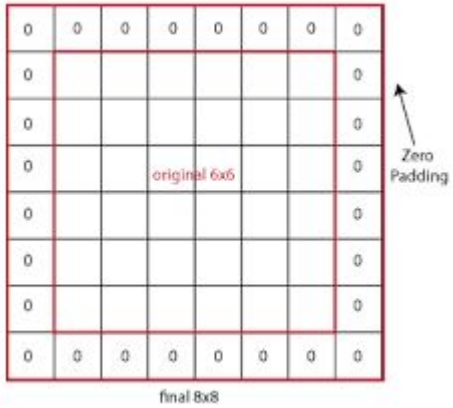
$$(n \times n) * (f \times f) = (n - f + 1) \times (n - f + 1)$$

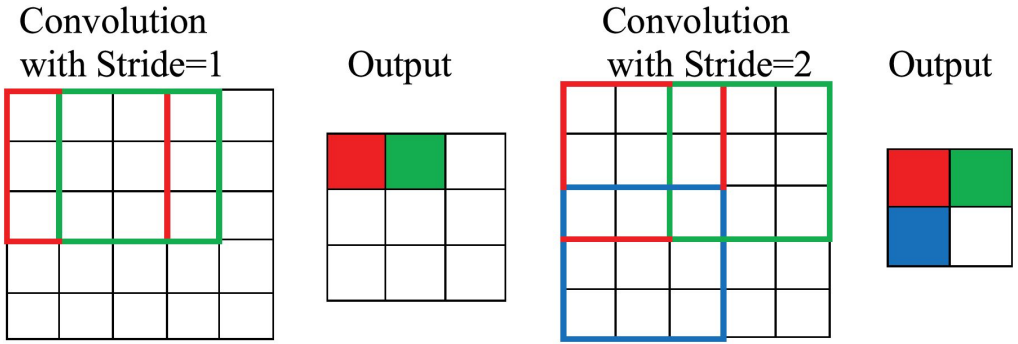
- Padding (p): amount of zero-pixels added to an image.

$$(n - f + 2p + 1) \times (n - f + 2p + 1)$$

- Types convolution:
  - **“valid”**: No padding (p =0)
  - **“same”**: Output size = Input size

$$p = \frac{f - 1}{2}$$





- Stride: (s) number of rows and columns traversed per slide

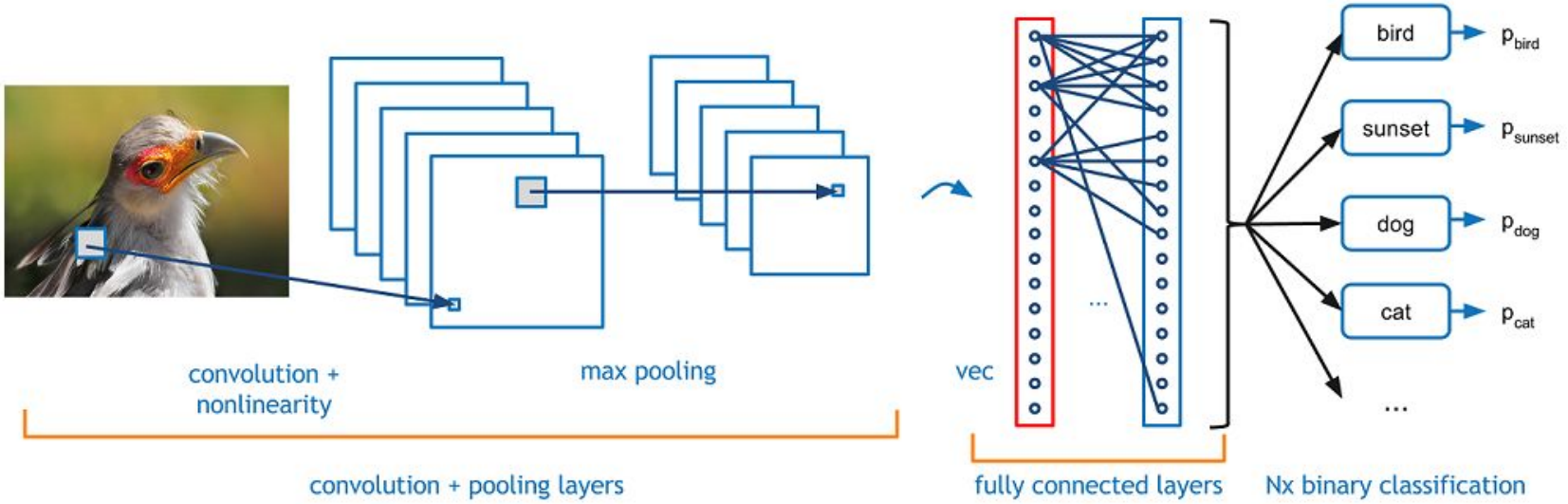
$$\left\lfloor \frac{n - f + 2p}{s} + 1 \right\rfloor$$

CNNs Advantages

- **Sparse Connections:** we need to store fewer parameters, which both reduces the memory requirements of the model and improves its statistical efficiency.

- **Shareable Weights:** use the same parameter for more than one function in a model. Rather than learning a separate set of parameters for every location, we learn only one set. **Equivariance to translation**, if the input changes, the output changes in the same way.

CNN Layers



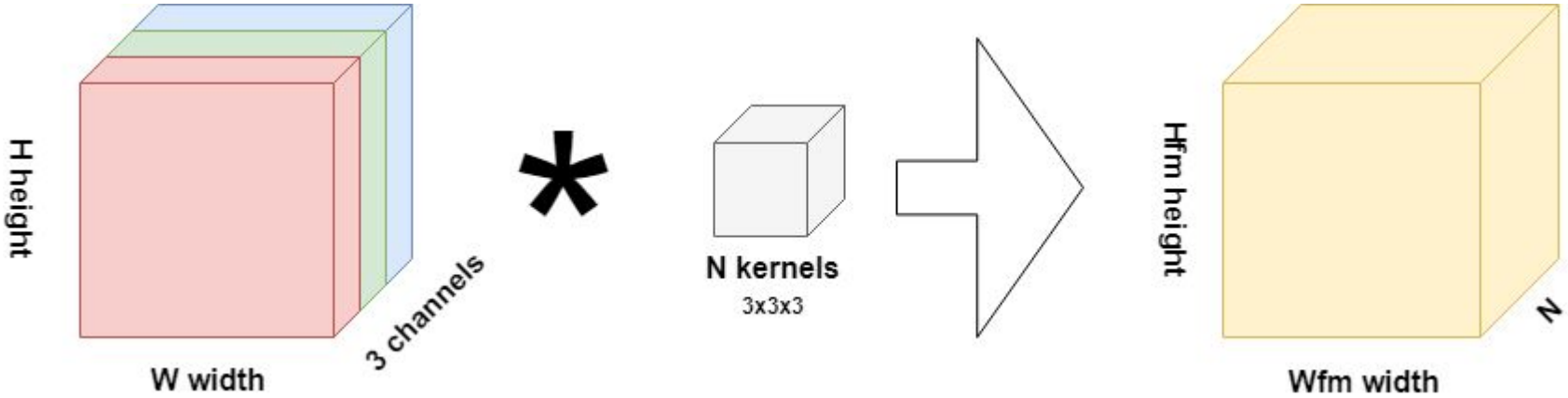
- Convolution + non-linearity (ReLu)
- Pooling: Downsampling
- Fully-connected

```
tf.keras.layers.Conv2D
```

```
tf.keras.layers.MaxPool2D
```

```
tf.keras.layers.Dense
```

CNN Layers: Convolution

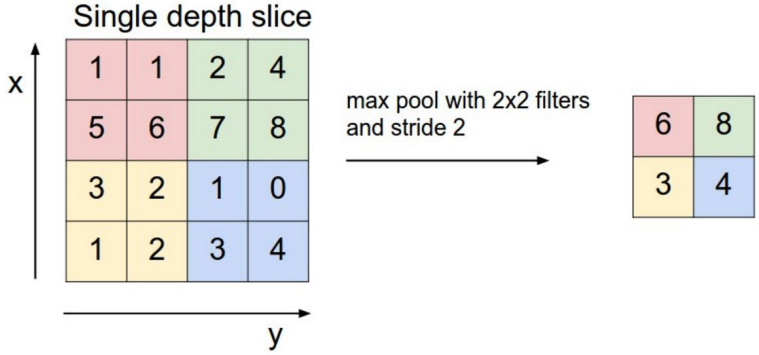


Layer dimensions: (h x w x N) Relu(convolution + b) for every filter (N different filters)

```
tf.keras.layers.Conv2D(
    filters=N,
    kernel_size=(h,w),
    strides=(1, 1),
    padding="valid"
)
```

CNN Layers: Pooling

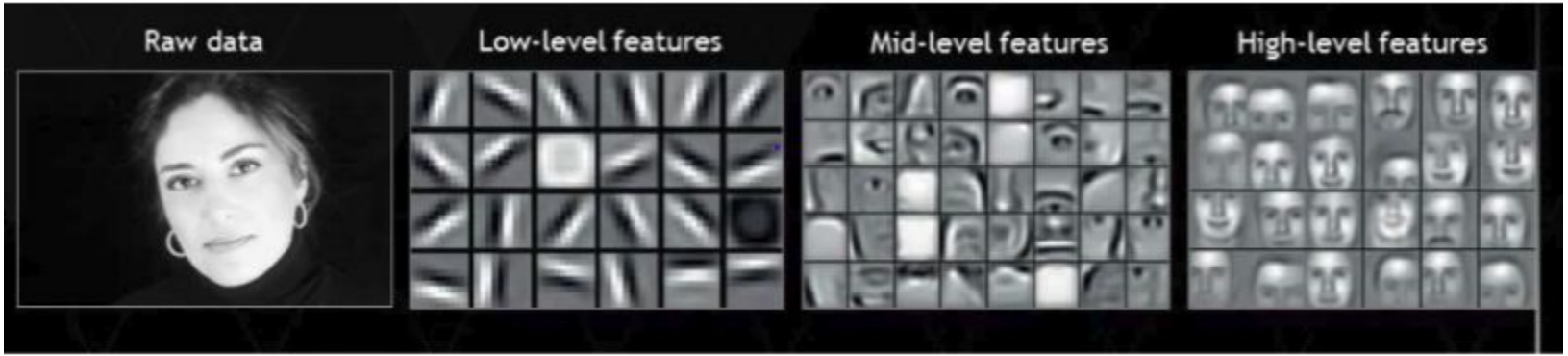
- **Dimensionality reduction:** representations turn smaller and more manageable.
- Reduces the number of parameters contributing to a smaller complexity.
- **Spatial invariance.**
- MaxPooling and average pooling are the most common.
- Backprop is not affected. **No parameters to tune**.




```
tf.keras.layers.MaxPooling2D(
    pool_size=(2, 2), strides=(1,1)
)
```

```
tf.keras.layers.AveragePooling2D(
    pool_size=(2, 2), strides=(1,1)
)
```

CNN Features

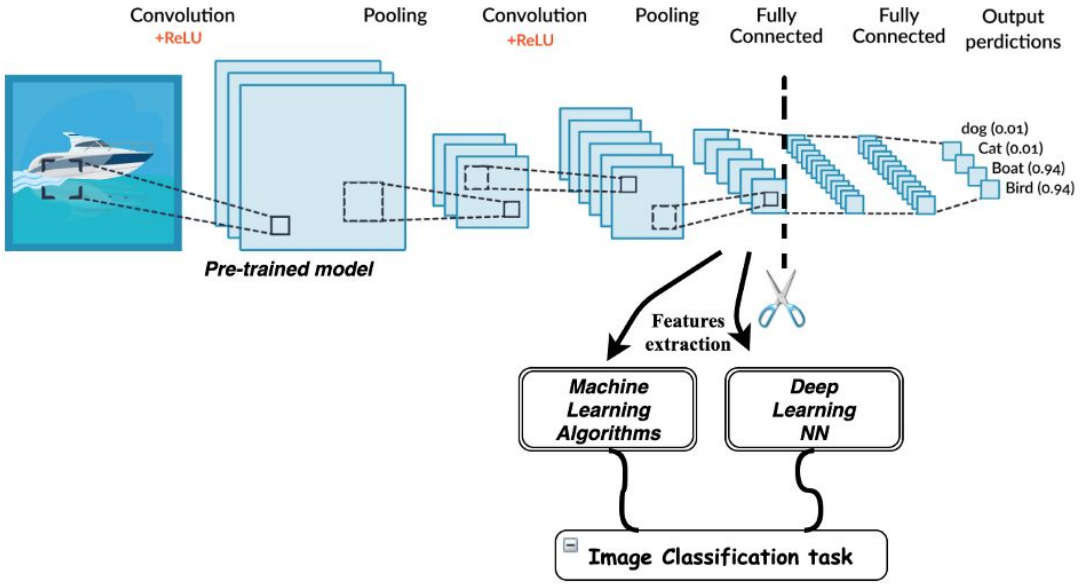


Conv 1 Conv 2 Conv 3

CNN: Data Augmentation

- Flip
- Rotation
- Scale
- Translation
- Crop
- Gaussian Noise
- ...

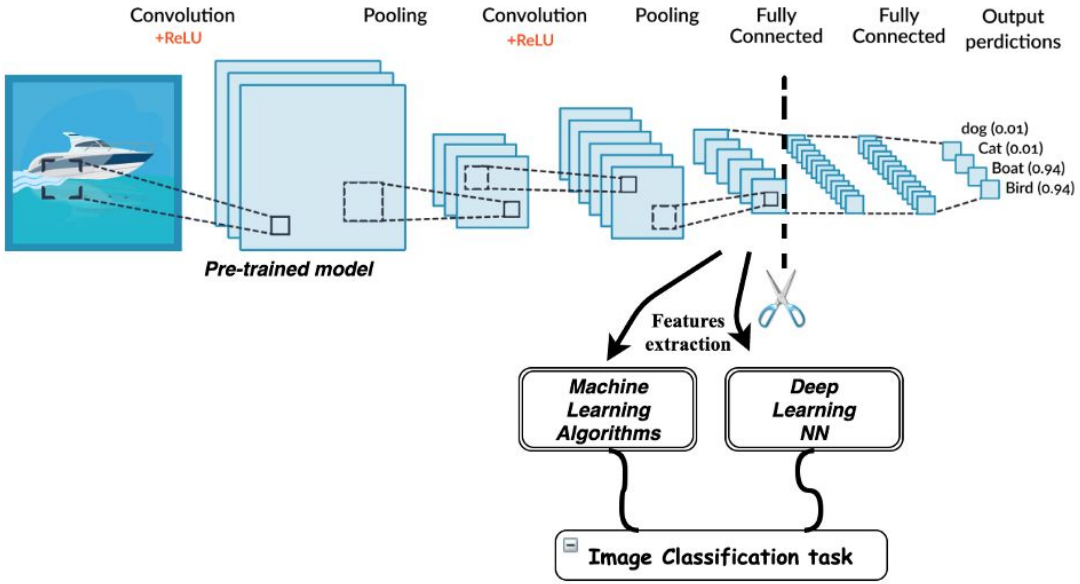
CNN: Transfer learning



Instead of training a Deep Neural Net from scratch for your task:

- Take a network trained on a different domain for a different task
- Adapt it for your domain and your task

CNN: Fine-tuning



We not only update the CNN architecture but also re-train it to learn new object classes.

CNN: Introduction & Classical Architectures

ntic master
revolucionamos la comunicación

UNIVERSIDAD COMPLUTENSE DE MADRID