Computers perceive the world as numbers

Pixels can be used as features

AI | BUSINESS SCHOOL

- **A pixel** in a grayscale image is <u>a single value</u>, while in color images it is an array of <u>3 values.</u>

- Each value can usually be scaled between **0** (no color) and **1** (full color).

But there is a problem with using pixels as features

AI BUSINESS SCHOOL

Disadvantages

The model cannot identify which pixels are closer to each other and which ones are far.

Solutions

We need features that consider neighboring pixels that give the spatial information

A method to extract spatial information from the images: convolution

AI | BUSINESS SCHOOL

What is convolution?

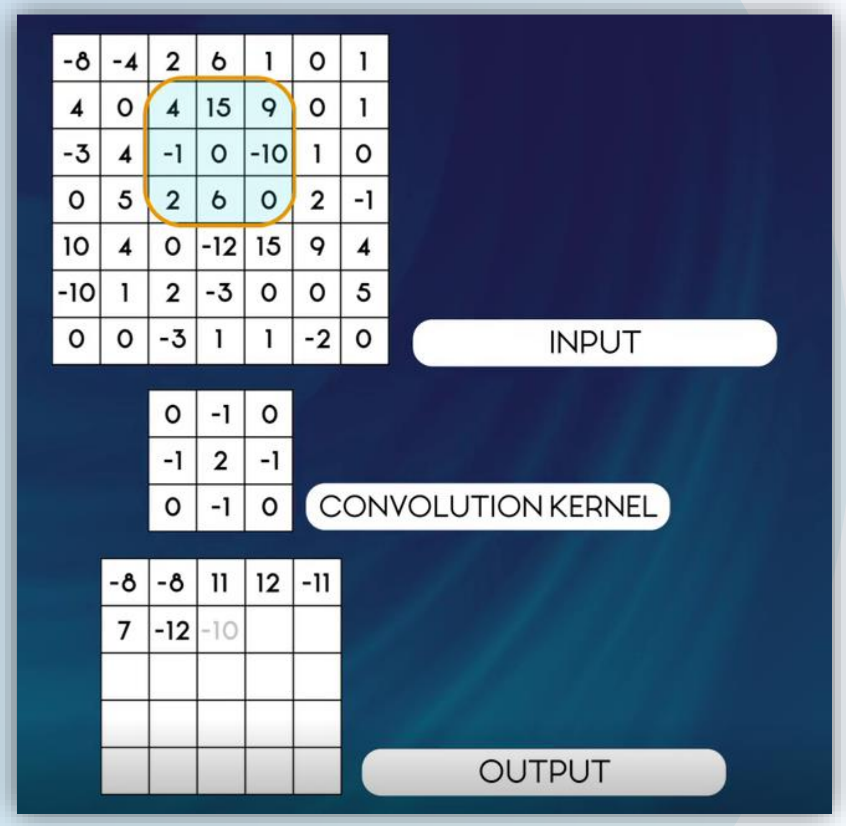
It is a simple mathematical operation where a weighted sum is calculated of each element and its neighbors. This helps **the computer to understand the spatial context of the data.**

How does convolution work?

AI | BUSINESS SCHOOL

In convolution, there is a predefined kernel...

- The kernel slides over the input image and performs matrix multiplication element by element
- The results of each area where the kernel slides over is being written in the output grid



***A kernel = A filter used to extract features from images**

Using multiple kernels to compute feature maps

AI BUSINESS SCHOOL

- In cases where one kernel is <u>not enough</u> to extract the features, multiple kernels can be used to **compute feature maps.** And these feature maps can be fed into the neural network and solve problems.

Practical example: reducing the unwanted noise from an image

AI | BUSINESS SCHOOL

- The mean filter **averages** the content of each window of the input image and outputs a **smoother** version of it without any sudden jumps.

  Now the salt and pepper noise in the input is removed by taking the local averages.

A crucial component in deep learning architectures: Convolutional Neural Networks (CNNs)

AI BUSINESS SCHOOL

What do CNNs do?

Convolutional Neural Networks improve the performance of classification models. They also serve as the basis for more complicated tasks and can be used to solve some common computer vision problems such as **semantic segmantation**, **object detection**, **instance segmentation**.

Examples of the most common computer vision problems

AI | BUSINESS SCHOOL

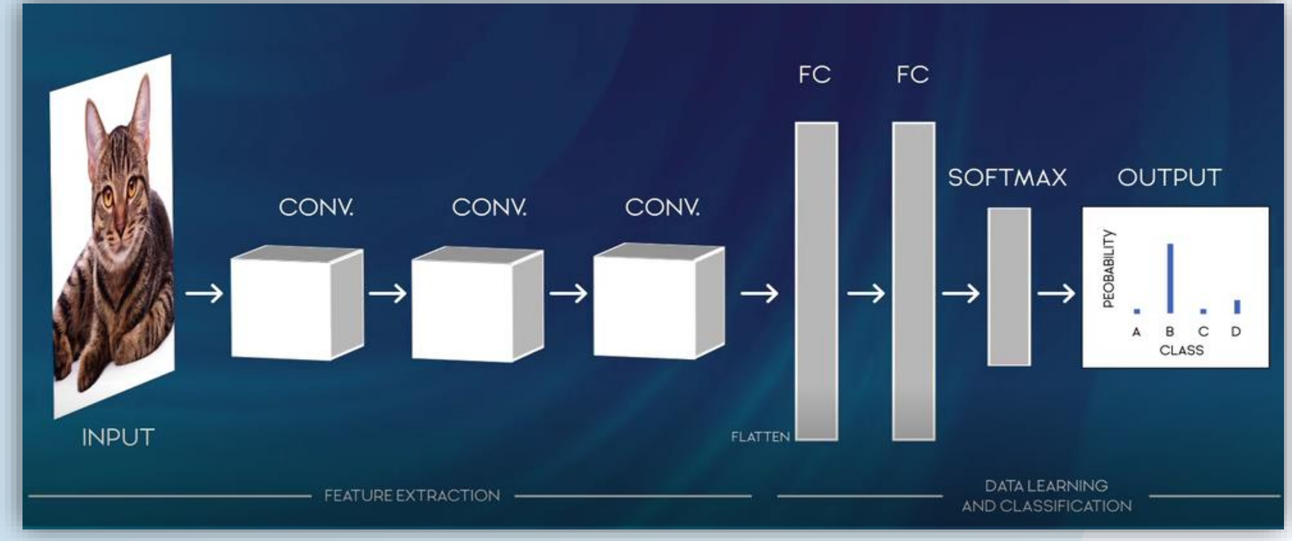
Semantic Segmentation
Every pixel in the image is classified

Object Detection
The model detects different objects in the image and draws boxes around them

Instance Segmentation
The model detects each object separately and segments it

How do these complex DL architectures work?

- Convolutional Neural Networks allow machines to learn the features by **automating the spatial information extraction process**.
- CNNs combine the information from pixel-level information and **create complex features**.



What could go wrong with convolutional operation?

AI BUSINESS SCHOOL

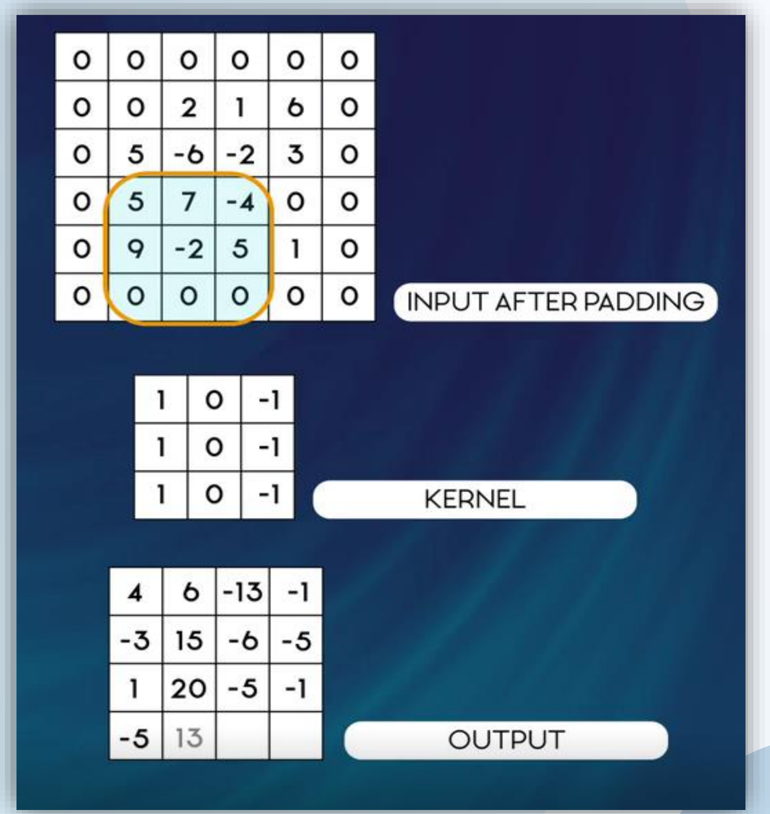
When convolution is applied to an image, the kernel is fit to the upper left corner and slides right at each step. In this process, **the borders do not have their pixels in the output** and therefore the dimension is reduced **which can cause information loss**...

Padding is used to prevent information loss at the borders

What is padding?

- It is the process of concatenating some numbers to the borders of the input frame, either padding by **zeros** or any other **constant**, or by **mirroring the edge**.

- Padding helps to <u>retain the shape of the input</u> when extracting the features and allows taking edges and corners of the input into account.



Pooling is used to reduce the size of the convolved features

AI | BUSINESS SCHOOL

After padding, some features extracted from these convolutional layers may be redundant or useless or even add noise to our model. Therefore, it is important to use only the useful or meaningful features.

👍 This will increase the performance of the model and reduce the computational cost by reducing the parameters. And this is where the pooling layer helps.

Common types of pooling

There are two most common types of pooling

Max pooling	Average pooling
The maximum value of each block is the layer's output	Each block's average value is the layer's output

How to transfer information from a trained model to an untrained model

AI | BUSINESS SCHOOL

By using existing models that have a well-trained feature extractor…

… we can train only the classification part of the model according to the given problem by using **the same optimized weights of pre-trained models**. And there is **no need to adjust them again.**

This process is called **“transfer learning”** and it helps a lot when training a CNN model with less data.