Image Classification with CNNs

Élèves :
Achraf Akhlifa
Mohamed El Hasnaoui
Walid Tarazi Younes Janah

Enseignant :
Adil Ahidar

31 décembre 2020

1 Deep computer vision

Vision is one of the most important senses that humans possess. Sighted people rely on vision every day from things like navigation, manipulation of objects, how to pick up an object to how to recognize complex human emotions and behaviors.
In this project we will discover how deep learning can build powerful computer vision systems capable of solving complex tasks.

1.1 How computers process images

A 'pixel' (short for ‘picture element’) is a tiny square of colour. Lots of these pixels together can form a digital image. Each pixel has a specific number and this number tells the computer what colour the pixel should be. The process of digitisation takes an image and turns it into a set of pixels.

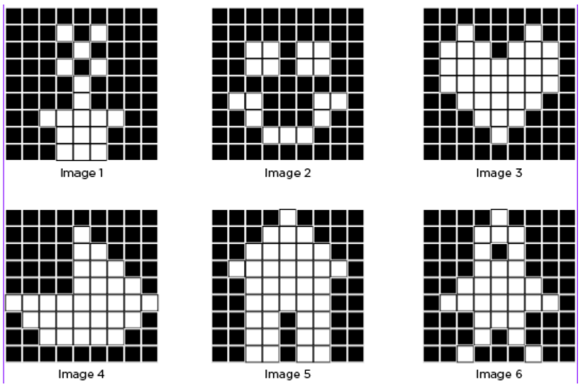


FIGURE 1 – Images are numbers

1.2 Computer Vision Related Problems Using Traditional Machine Learning

In automated visual inspection systems offer manufacturers the ability to monitor and respond in real time to production problems, reduce costs and improve quality. Most visual inspection systems nowadays consist of some form of hardware for image capture, and an integrated or discrete device equipped with specialized image processing software. At the heart of this program is a computer vision algorithm which encompasses the array of numbers representing the product picture, performs some mathematical operations on those numbers, and calculates a final result. For example, the computer vision algorithm can determine that a whole product is defective, detect the type and position of a defect on a product, test for the existence of some sub-component, or calculate the overall finish quality. Driver-less cars also features high artificial vision as one of their key inputs.
This computer vision algorithm is divided into two stages in traditional machine vision systems. A series of mathematical operations are performed on the image's raw pixel values in the first step, which is usually called feature extraction. For example, when searching for defects in a product image, the feature extraction step may consist of sliding a small window over the entire image, and computing the contrast for the pixels within the window for each window position-the difference between the brightest and darkest

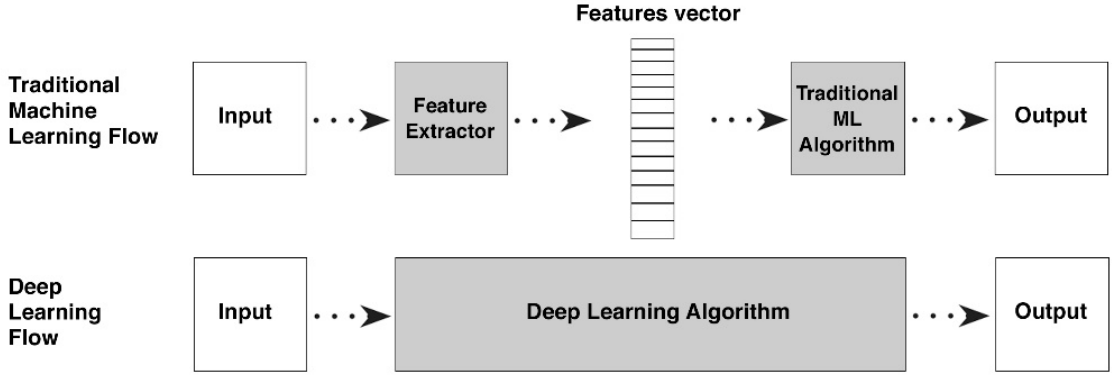


FIGURE 2 – Traditional Machine Vision Uses A Two Step Process To Process Images

pixels. This feature may be useful in making a final determination, since higher-contrast windows may be more likely to contain defects.

2 Convolutional Neural Networks

In this section, we deal with a particular neural network's architecture used to solve image classification tasks.
Convolutional neural networks are based on the convolution operation (described earlier). The goal is to learn features directly from data and to use these learned feature maps for classification of these images. Generally, the network proceeds in two major steps :

Feature extraction in this step, we apply filters (using convolution) to generate feature maps. Then, we deal with non-linear data and introduce complexity into the learning process by applying a non-linear function on the output of the feature maps. Finally, we use pooling to reduce dimensionality and to achieve spatial invariance.

Classification the previous steps result in a vector representing the different features of the input, the output is then fed to a dense artificial neural networks, which outputs the final classification probabilities (using the softmax activation function).

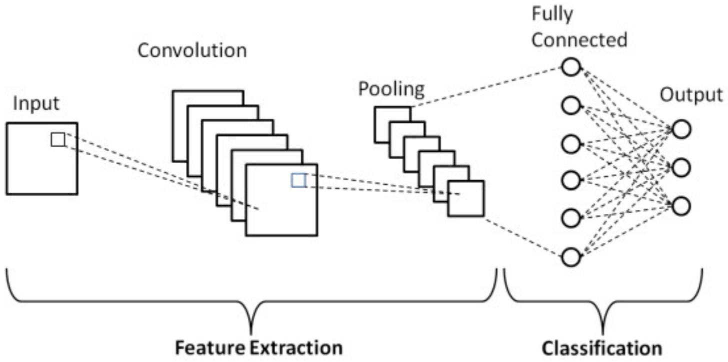


FIGURE 3 – A convolutional neural network architecture

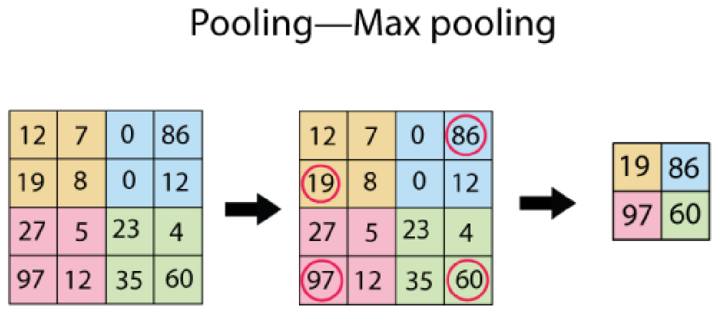
2.1 Convolutional layers and local non-linearity

In each convolutional layer, there are several neurons. Each neuron makes a filtered version of its input. It does so by performing the following operations : it takes a patch from the input, it performs a convolution operation with its weight matrix (also called kernel or filter), it adds a bias term, then it applies the non-linear activation function. Thus, the output of each neuron is a new modified version of its input. One important thing to note is that each neuron only takes a patch of the input as input ; this is called local connectivity. Also, the neurons share weights ; we use the same kernel and bias for all the patches.

A convolutional layer can perform more than what we have just described. We can associate multiple filters with one convolutional layer and the output will not be a new version of the input, but multiple modified versions. We can interpret this process, for the first layer of the network, the following way : a filter extracts one specific feature of the image, and a convolutional layer constructs multiple modified versions of the same image, each highlighting a feature of interest.

2.2 Pooling step

Output feature maps, that result from the convolution layers, have one major problem : they are sensitive to the location of the features in the input. Pooling is an approach we take to make feature maps more robust to changes in the position of the feature in the image. It downsizes the input and summarize the presence of features in patches of the feature map. Two common pooling methods are average pooling and max pooling that summarize the average presence of a feature and the most activated presence of a feature respectively.



12	7	0	86
19	8	0	12
27	5	23	4
97	12	35	60

12	7	0	86
19	8	0	12
27	5	23	4
97	12	35	60

19	86
97	60

Figure 4 – Max-pooling example