

Selected Topics in Visual Recognition and Deep Learning

Andrés Ponce

September 17, 2020

Computer vision deals with connecting a camera to a computer and trying to process an image data. The idea is also to simulate human vision using some algorithms and deep learning.

An **image** is a collection of (x, y) points which contain some intensity value. In a black and white image, the intensity value might be just how dark or light it is at the point (x, y) , where we start counting from the top left.

If we want to detect some pattern in the image, we would have to perform a **convolution**.¹ So here, one might have a linear combination of the surrounding pixels. Then we use a **filter** to somehow transform the images. An example filter might be

¹ A **convolution** is a function that describes how two *functions* influence each other.

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

If we pass this filter through an image, what will the result be? If we just take the total sum, we are going to take the difference between the pixels on the left and right side. This difference might be useful in detecting edges and such.

If we mess around with the filter structure, we can either apply a **blur** effect or a sharpening effect.

Some of the difficulties in deep learning occur from some artifacts which obfuscate the image. For example, **illuminations**, deformations, and occlusions.

Conventional Approach to Object Recognition

In the traditional approach to object recognition, there are several steps:

1. Image Collection
2. Feature Extraction
3. Classifier Training
4. Trained Classifier

The main image characteristic is **features**, which are just characteristics of the images that are of interest to us. Maybe the first layers learn some low-level, coarser features, and higher layers learn finer features of the image.