

CHAPTER 35

Convolutional Neural Networks (CNN)

Convolutional neural networks (CNN) are a specific type of neural network systems that are particularly suited for computer vision problems such as image recognition. In such tasks, the dataset is represented as a 2-D grid of pixels. See Figure 35-1.

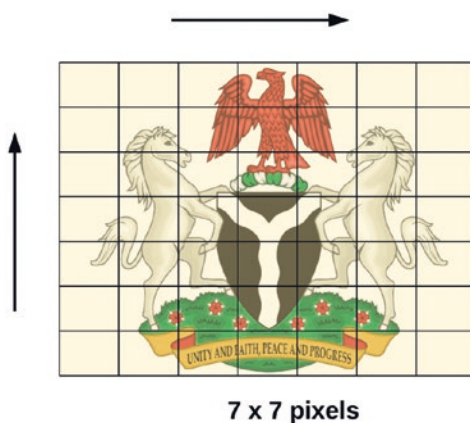


Figure 35-1. 2-D representation of an image

An image is depicted in the computer as a matrix of pixel intensity values ranging from 0 to 255. A grayscale (or black and white) image consists of a single channel with 0 representing the black areas and 255 the white regions with the values in between for various shades of gray.

For example, the image in Figure 35-2 is a 10 x 10 grayscale image with its matrix representation.

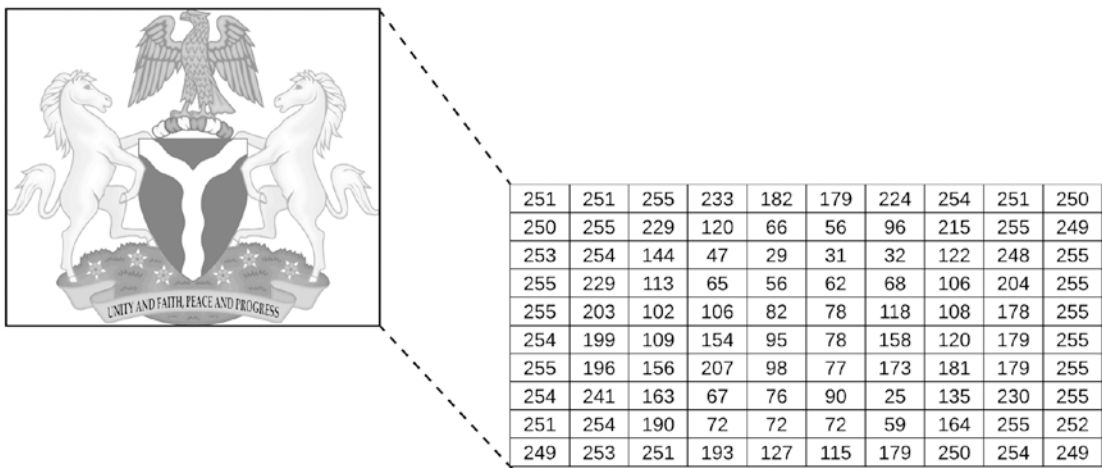


Figure 35-2. Grayscale image with matrix representation

On the other hand, a colored image consists of three channels, red, green, and blue, with each channel also containing pixel intensity values from 0 to 255. A colored image has a matrix shape of [height x width x channel]. In Figure 35-3, we have an image of shape [10 x 10 x 3] indicating a 10 x 10 matrix with three channels.

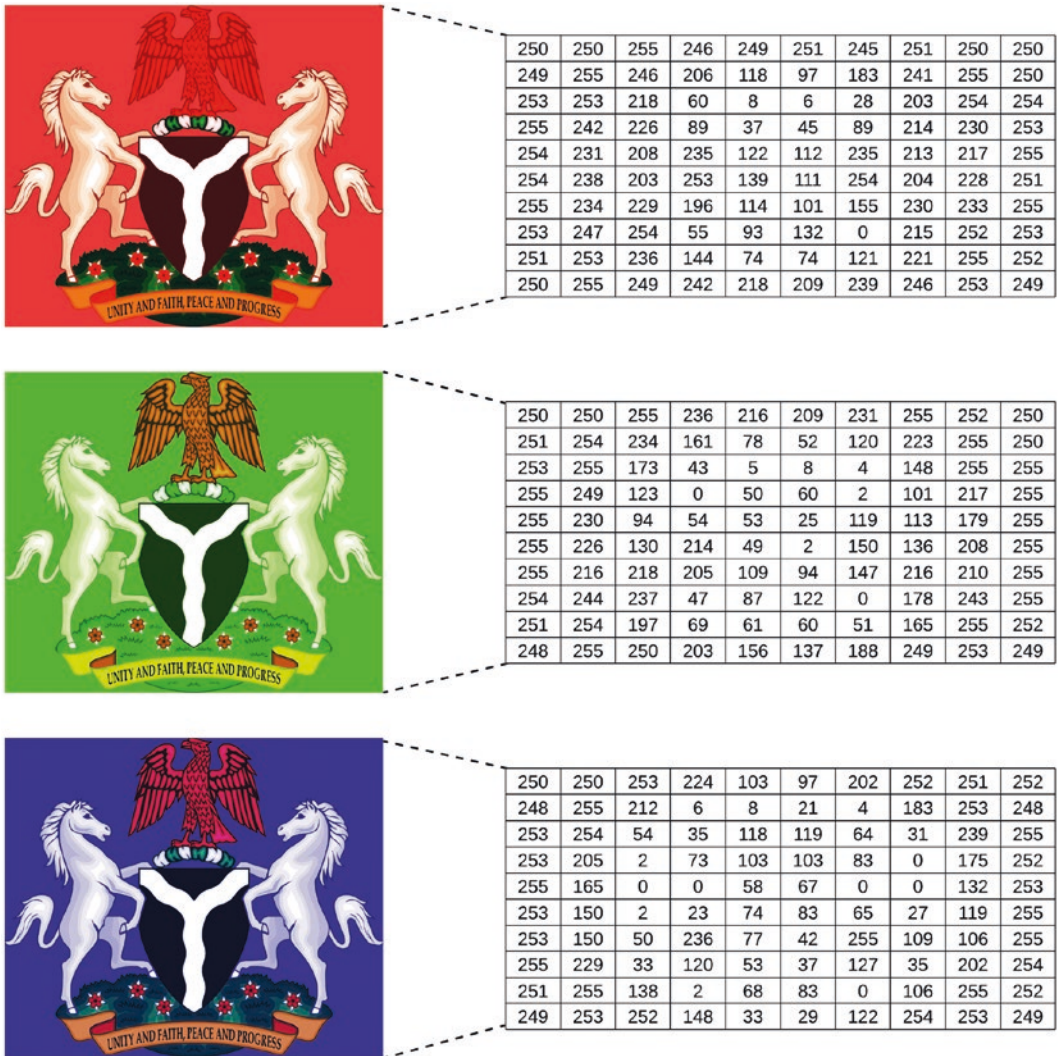


Figure 35-3. Colored image with matrix representation

Local Receptive Fields of the Visual Cortex

The core concept of convolutional neural networks is built on understanding the local receptive fields found in the neurons of the visual cortex – the part of the brain responsible for processing visual information.

A local receptive field is an area on the neuron that excites or activates that neuron to fire information to other neurons. When viewing an image, the neurons in the visual cortex react to a small or limited area of the overall image due to the presence of a small local receptive field.