My understanding of Convolutional Neural Network

We humans see the world through our eyes, the 'sensory' organ. We see and learn to differentiate. Someone captions the name of the object, faces or scenes for us. The leaves of a tree are green, the color of sky is blue, etc.

We differentiate objects, faces or scenes based on their physical boundaries (technically their edges). This process happens so quick that before captioning what is seen, we cannot undo seeing it.

Artificial Intelligence is basically the intelligent machines making appropriate decisions by themselves without human supervision or intervention.

To transfer the visual prowess of humans to machines, the Convolutional Neural Networks (CNN) can be used so that they can have human like vision and operate on what they see.

But why Convolutional Neural Networks only?

> Because the way Convolutional Neural Networks work is much similar to how the human visual cortex learns to see. The real-life image in human brain is created from observing smaller features of the objects first and then forming the big picture. This hierarchical learning is similarly managed in CNNs too.

The high-level application of CNNs is image processing, image analysis, image recognition and image generation. (Including elements of image that is objects, faces, moving images etc.)

The digital image is arrangement of pixels in rows and columns. This digital image is converted to 2D array during preprocessing.

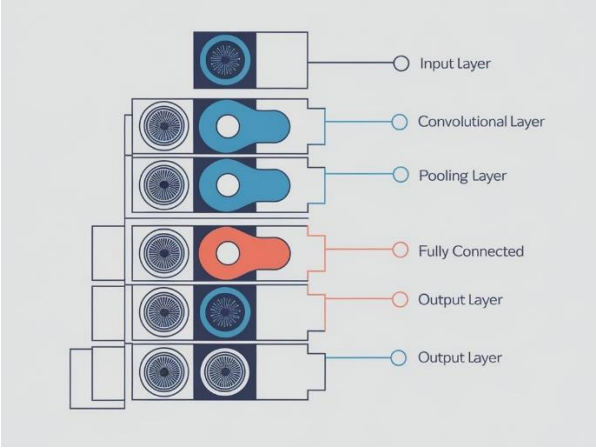
To learn various features of the original image, the CNNs use filters which are also arrays (or matrices).

To start learning about the input image, the filters will start element wise multiplication from the top left corner of the image and progress until the whole image is covered. This is called filter sliding.

The final output after applying filters is the feature maps. Feature maps can be simply called as a version of the input image which holds data of only a particular feature of the image. For any input image fed to CNN, there will be 'n' feature maps for 'n' filters.

This is how the CNNs learn multiple features of the image layer by layer.

The feature maps of previous layer are inputs for next convolutional layer.



Source: Image generated using Ideogram