By Paul Naidoo, EOLAS

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a specialized type of (typically "deep") neural network designed for processing structured grid data, such as images. They excel in tasks like image recognition, object detection, and segmentation.

Their defining characteristic is the employment of "convolutional layers" where a kernel, which is a trained parameter in the model, performs transformations on grid data; isolating features therein.
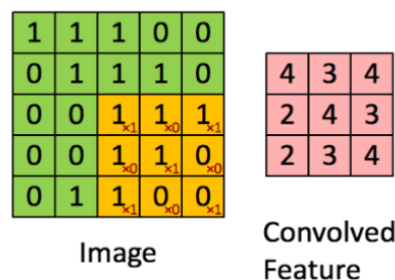

Figure 1: Illustration of a convolution kernel in action across a matrix of values. (Animation available at source)

This is depicted in Figure 1, where a kernel (values in red in the yellow window) has passed across the green matrix, resulting in the convolved matrix (pink).

Structure and Function of CNNs

A typical CNN structure can be found in Figure 2, with explanations given below.

- After convolution, activation functions like ReLU (Rectified Linear Unit) introduce non-linearity, enabling the network to learn complex patterns. These are often referred to as "feature maps".
- "Pooling layers" then down-sample the feature maps, reducing their spatial dimensions while retaining significant features. This helps in making the network more efficient and robust to variations in the input.
- After several convolutional and pooling layers, the output is flattened into a one-dimensional vector which is passed through fully connected layers (akin to a Deep Neural Network), which integrate the learned features to make final predictions.
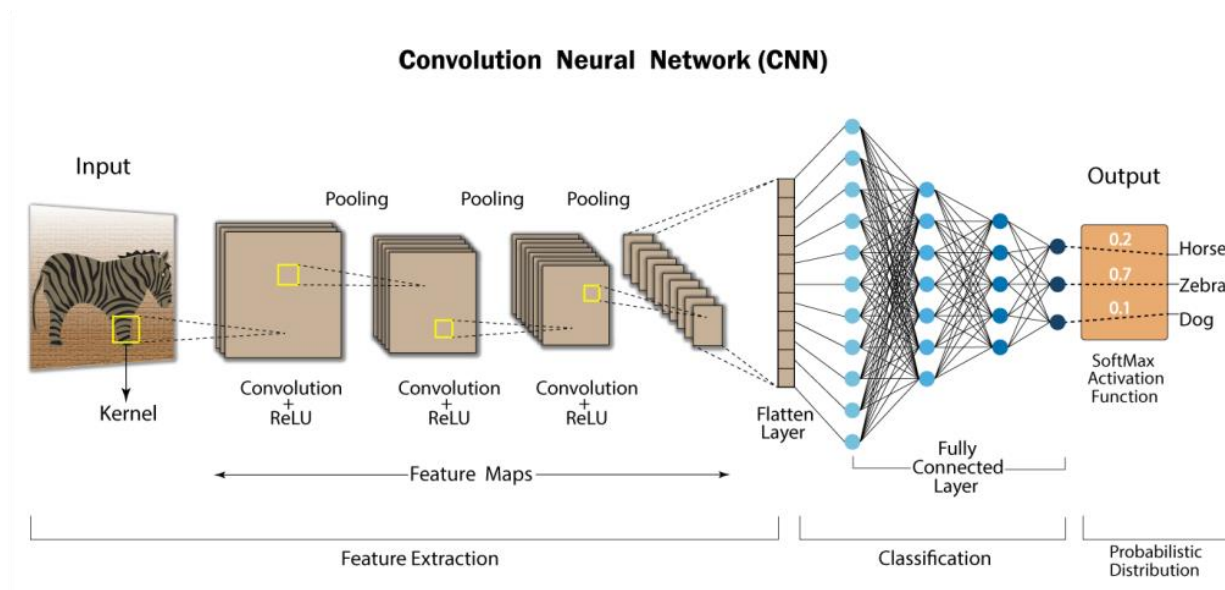
Figure 2: Illustration of a Convolutional Neural Network performing image classification (source)

Advantages of CNNs

- **Spatial Hierarchy**: CNNs effectively capture spatial hierarchies in data, making them ideal for image and video processing.

- **Parameter Sharing**: Filters are shared across the input data, reducing the number of parameters and improving computational efficiency.

- **Translation Invariance**: Pooling operations help achieve translation invariance, making the network robust to shifts and distortions in the input.

Limitations

- **Computationally Intensive**: Training CNNs can be resource-intensive, especially with large datasets and complex architectures.

- **Data Requirements**: CNNs typically require large amounts of labelled data to achieve high performance.