Convolutional Neural Networks

A convolutional neural network will likely be the first neural network that you approach when wanting to work with images. The network is able to look at an image, identify unique features to each image and make a prediction based on the content. CNNs are made of what are known as Neurons and have learnable weights and biases.

Unlike other neural networks, the neurons in a convolutional neural network are organised into 3 dimensions; width, height and depth. An image can be broken down into 3 dimensions, red, green and blue. Each colour represents an array storing numbers which correspond to the colour value of each pixel in the image.

Image Data Augmentations

Image augmentation will allow us to increase our total amount of collected data substantially by altering each image a number of times with the parameters that we set. Doing so has significant benefits to our final trained model. Augmenting combats overfitting and helps the model to generalise better, thus increasing the accuracy of predictions and overall performance.

The Layers

It's important that we understand what the layers are in a CNN as we will either be building ours from the ground up or using a pre-trained model. There is no point in choosing a model that is pre-defined if we don't understand what it is useful for.

Input

This has been covered above, but I'll cover it again here. The input of a CNN will consist of 3 dimensions, for example, an image may have the dimensions 32x32 which is really 32x32x3. The 3 represents the depth of the image, the 3 arrays, each of which represent the 3 primary colours that make up the colour image file. If we were to look at one of the arrays it would be a 1D array with a total of 32x32 values.

Convolution

The role of the convolution layer is to extract features from our images. A filter of a much smaller size than the images (E.g. 5x5) will scan over the image similar to how you would imagine a flashlight to look against a surface like a wall. It scans a group of pixels, 5x5 in this case, produce an activation map and finds unique features like edges or colour patterns. Eventually these features that are being recognised could become wheels or faces.

Relu Activation

ReLU (Rectified Linear Unit) steps are normally introduced after convolution operations in CNN's. ReLU is applied per pixel and replaces any negative values with a 0 value; this is due to wanting to introduce non-linearity. The purpose of introducing non-linearity is due to the fact that most real world data is not linear.

Pooling

The pooling stage of the network reduces dimensionality of the feature maps while retaining the important information. This is to reduce the number of parameters and computation in the network, while also reducing the risk of overfitting.

Fully Connected Layer

The fully connected layer implies that every neuron in the previous layer is connected to every neuron in the next layer.