Convolutional Neural Network:

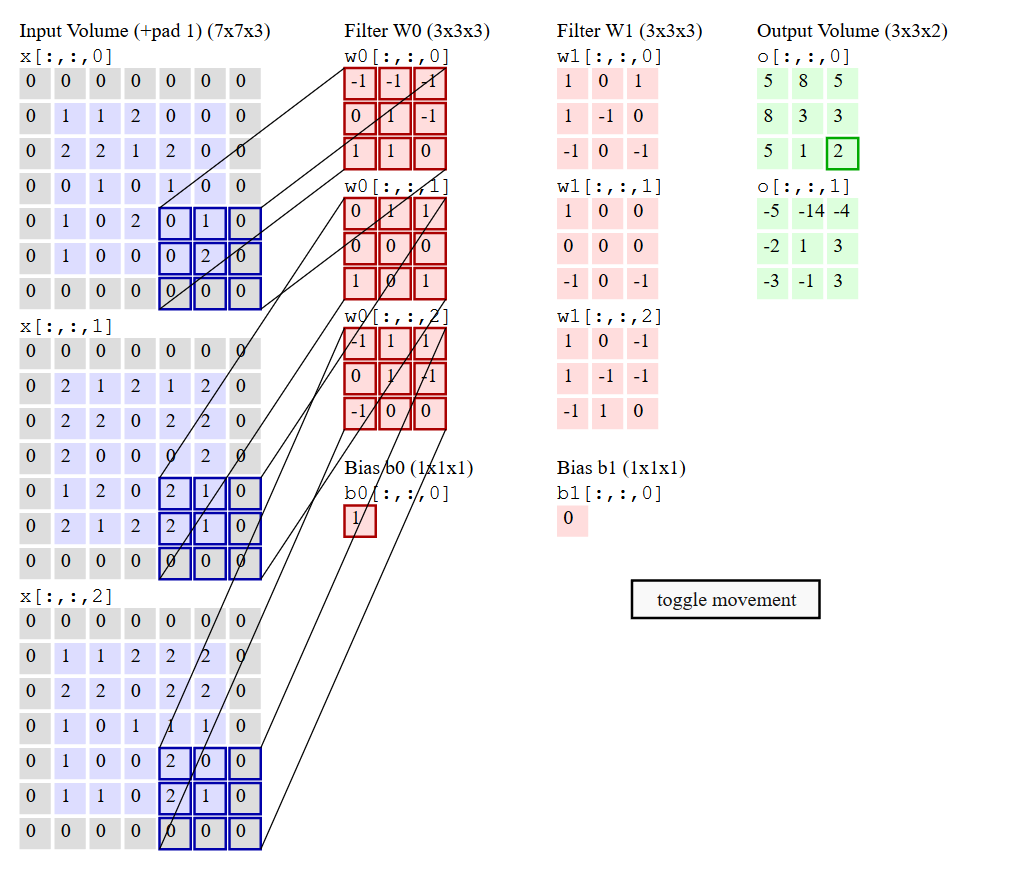
CNN have time and again proved that best way to solve any computer vision related problem is through CNN. CNN are being used different types of computer vision problems like anomaly detection, segmentation, classification, etc.

For classification part of the computer vision, the most important part of CNN are

1. Convolutional Layers.

In Convolutional layers, learnable parameters consist of set of learnable filters. Every filter is spatially matrix, which extends through the full depth of input volume. In pytorch, the typical filter size for a first Convolutional layer for RGB picture might have size 5x5x3. During forward pass, we slide each filter across width and length of input volume, and find the dot product between filter and input at any position at time. During this process, the activation map is generated, which gives us responses of that filter at every spatial position. Now in earlier layers, the filters will activate the edges, colour pallets, etc, but in more deep layers the Convolutional layers will activate more complicated features like honeycomb, wheel like patterns. Each of the layers will learn more distinct features and will generate more activation maps; we will stack these layers along the depth dimensions and produce output volume.

The filters that give us activation maps are nothing but weights and biases that are needed to be learned by the model. During the process of training, these various features are learned.



1. Pooling Layers.

Pooling layers are important to just fetch the important parameters only, this job is done by various means in this regard. Sometimes the average of the feature maps, sometimes the maximum element is taken. This process is important to lessen the amount of parameters in the CNN model, As the more number of parameters can cause overfitting.

The most common form of pooling layer is filter size of 2x2 with stride 2. Every pool operation would take 4 numbers. As in the case of pooling layers, there are no extra parameters are added.



1. Fully connected layers.

This part of CNN is as similar as a simple neural network. All the neurons are connected with each other, and significance of their connection is computed while training. But before giving the output of previous layers to the Fully connected layers, we need to convert the 2d output of previous layers into 1d vector, this process is called flattening.

In the Final layer of FC layers Sigmoid or softmax is generally used for the activation. As well as in final layers, there are generally only n numbers of neurons as there are n numbers of classes in the dataset.

