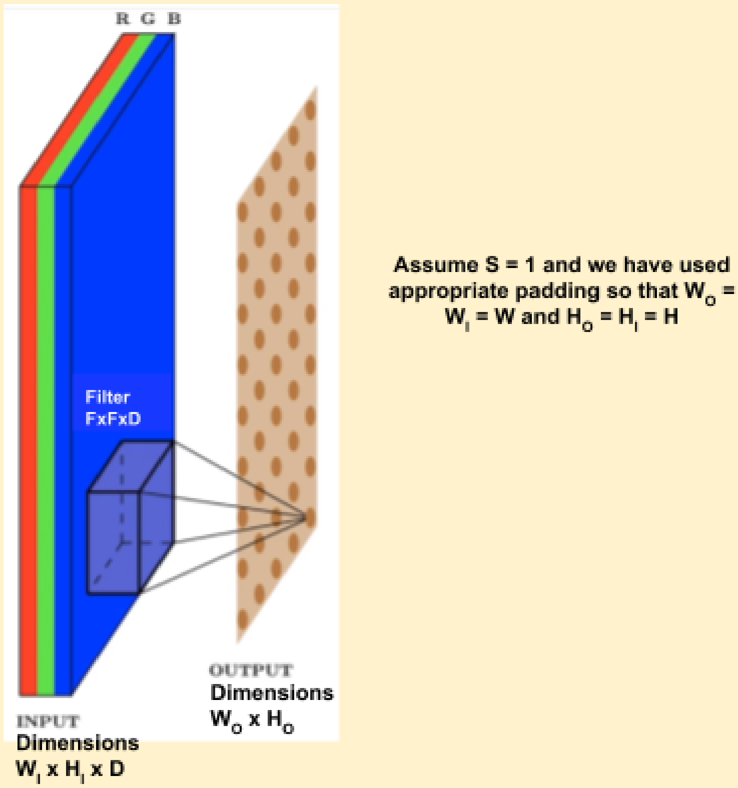
PadhAI: CNN Architectures

One Fourth Labs

Number of computations in a convolution layer

Let's see how many computations are needed in a CNN.

1. We will be looking at the GoogLeNet architecture as an improvement to the VGGNet based on the points discussed in the previous section.
2. However, before that, we must look at two key concepts in the GoogLeNet layout: **1x1 convolution** and an **interesting way to perform max-pooling**.
3. To approach these two, we first need to see how many computations are needed in one convolutional layer



   a. **Input dimensions**: $W_I \times H_I \times D_I$
   b. **Filter size**: $F \times F \times D_I$
   c. **Output dimensions**: $W_O \times H_O$
   d. Stride = 1 and appropriate padding so that $\mathbf{W_O = W_I = W}$ **and** $\mathbf{H_O = H_I = H}$
4. To calculate the number of computations:
   a. For every pixel of interest, for D layers, we perform FxFxD computations
   b. So for an output area of WxH, we perform (WxH) x (FxFxD) computations
   c. From the previous point, we can observe that the Depth of the output layer will be very large if there is a large number of filters applied on the input layer, as each filter generates a 2D area of unit depth.
   d. So if we use a **large number of filters**, the **output volume will be very deep**, subsequently **increasing the number of computations in the next layer's calculation** (Due to high D value).
   e. We can also try controlling W and H, but they can be more easily regulated using max-pooling. However, depth is directly related to the number of filters used.