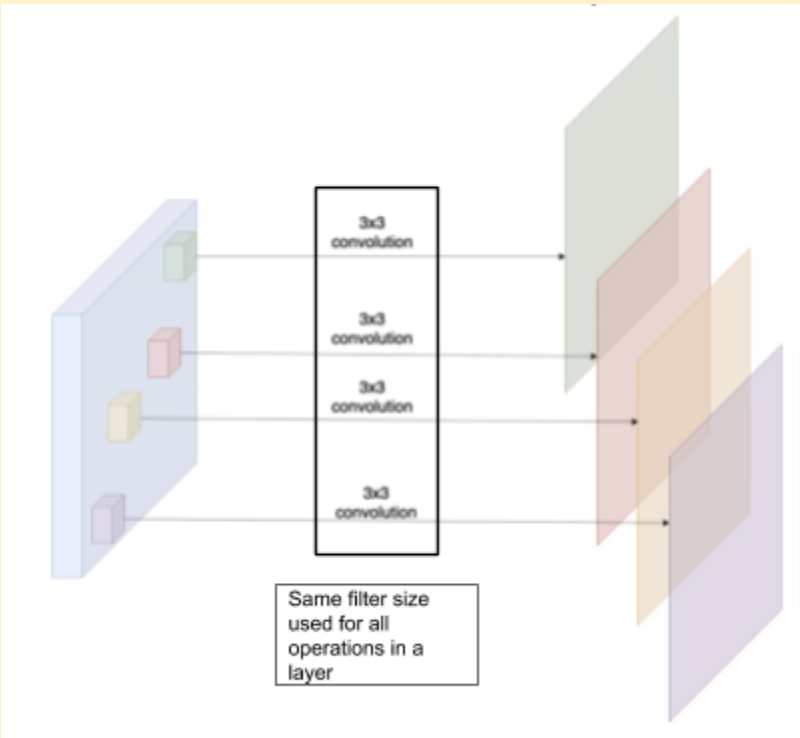
PadhAI: CNN Architectures

One Fourth Labs

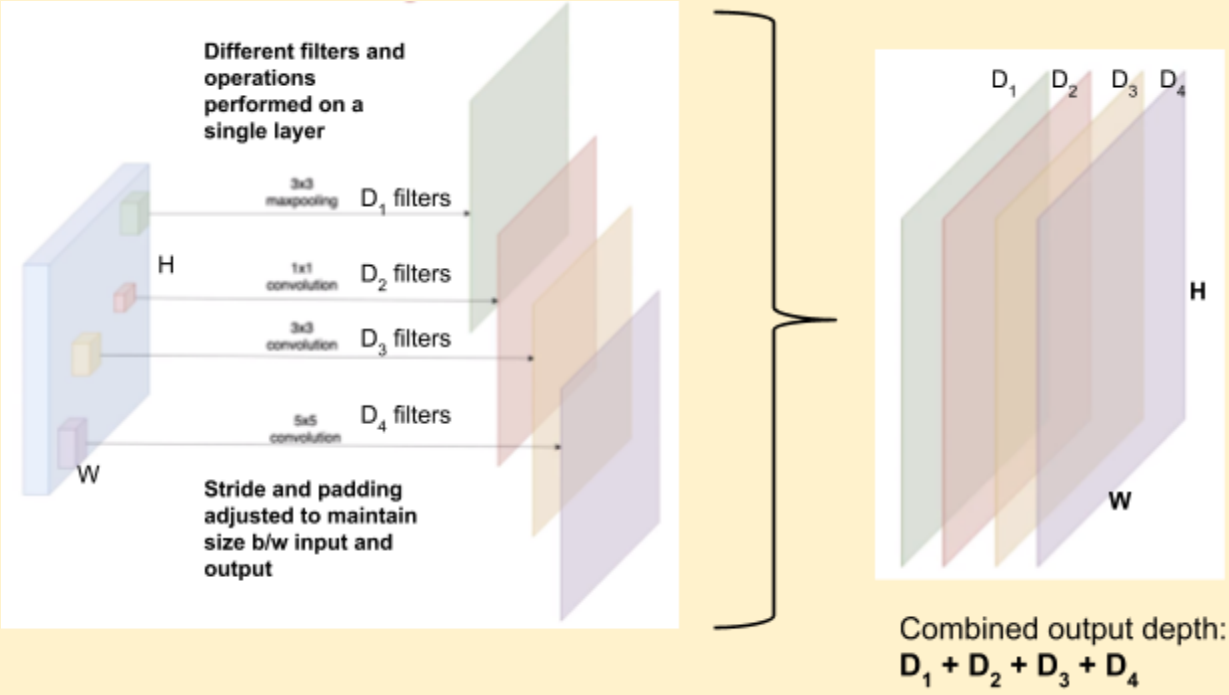
The Intuition behind GoogLeNet

What is the intuition behind GoogLeNet?

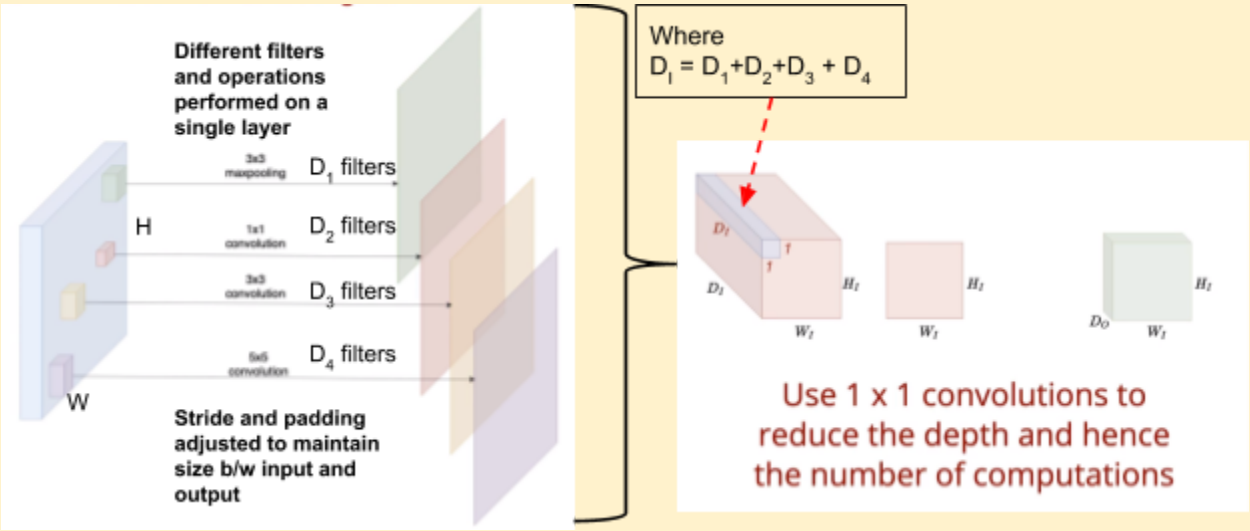
1. One point to note in the architectures used thus far, is that we must always make a choice of a particular filter size for any given layer. For eg, in VGGNet, all filters used were 3x3
2. Another point is the interspersing of Max-pooling layers between convolutional layers. How do we decide the arrangement to follow?



3. In GoogLeNet, the choice was eliminated, instead we are able to apply all our operations whatever combination of filter size and max-pooling/non that we'd like.



4. In GoogLeNet, we can apply multiple filters of varying size to perform either convolutional or max-pooling operations
5. Through experimentation with the older architectures, we have found that when there are multiple convolutional operations in a layer, we needn't use larger filter sizes.
6. Therefore, 5x5 filters are usually the upper limit of size.
7. One constraint is that for each operation in a layer, appropriate padding and stride must be taken so as to preserve the width and height between the input and the output
8. The Number of filter for each operation can vary, (D_0 to D_3 etc)
9. In the output volume, the total depth is the combined depth of the individual volumes from each of the operations. $D = D_0 + D_1 + D_2 + D_3$
10. The problem with combining the depths is that it has the potential to become very large, thus drastically increasing the number of computations.
11. We can mitigate this problem by performing 1x1 convolutions



12. By performing 1x1 convolutions, we can reduce the depth of the output volume, thereby reducing the number of computations to be performed in the subsequent layers.