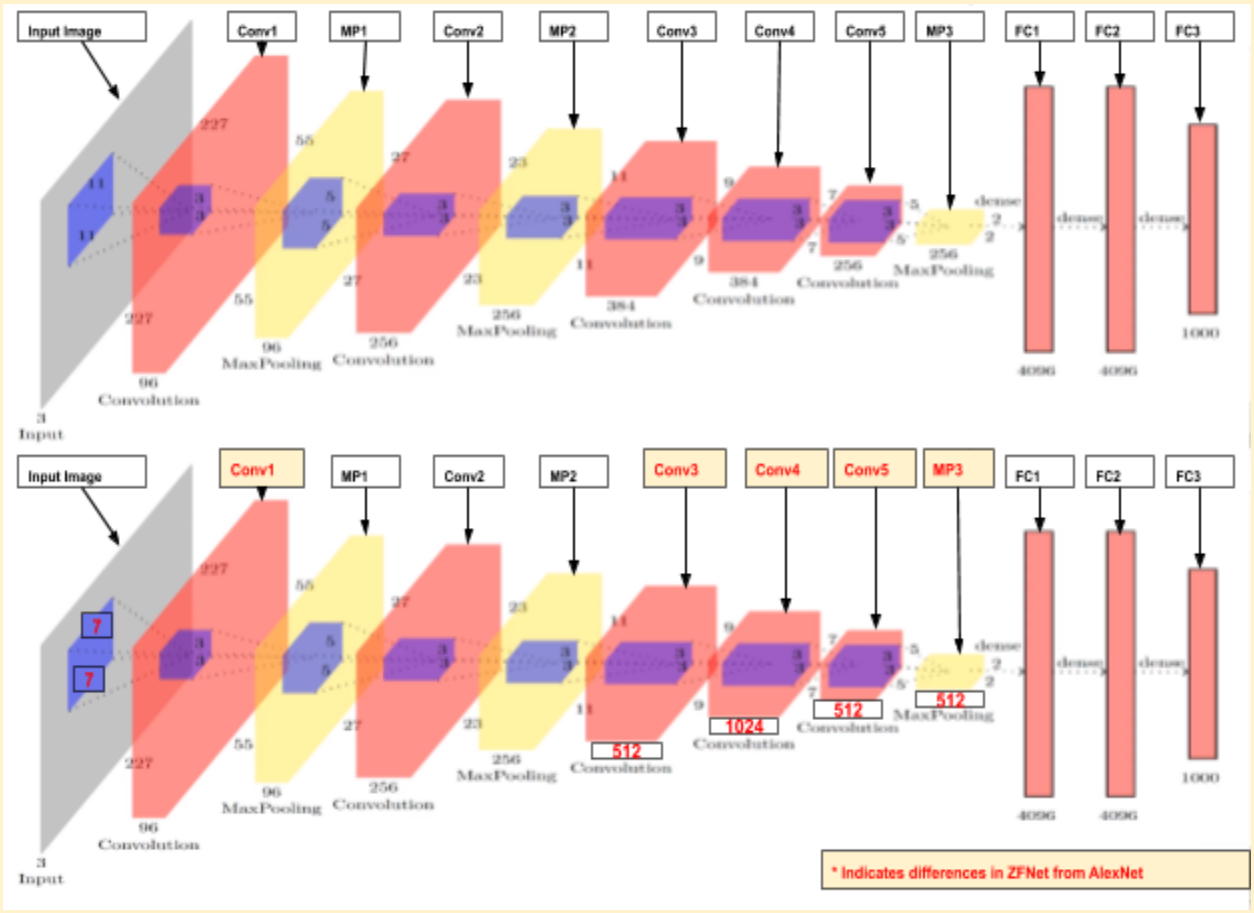
PadhAI: CNN Architectures

One Fourth Labs

ZFNet

Let's now look at the entire AlexNet

1. ZFNet is another 8-layer CNN architecture. Let's understand it better with a side-by-side comparison with AlexNet.



2. ZFNet is largely similar to AlexNet, with the exception of a few of the layers. Let us highlight those differences.
3. **Convolutional Layer 1**: Input is 227x227x3
   a. Filter Size (**F**) = 7 (7x7x3)
   b. No. of Filters (**K**) = 96
   c. Stride (**S**) = 4
   d. Padding (**P**) = 0
   e. **Parameters** = (7x7x3) x 96 = 14,112
   f. $\mathbf{W_1}$ = 55
   g. $\mathbf{H_1}$ = 55
   h. $\mathbf{D_1}$ = K = 96
   i. **ReLU** Non-linearity function is applied to every 2D area in the output volume.

4. **Convolutional Layer 3**: input is 11x11x256
   a. Filter Size (**F**) = 3 (3x3x256)
   b. No. of Filters (**K**) = 512
   c. Stride (**S**) = 1
   d. Padding (**P**) = 0
   e. **Parameters** = (3x3x256) x 512 = 1,179,648
   f. $\mathbf{W_3}$ = 9
   g. $\mathbf{H_3}$ = 9
   h. $\mathbf{D_3}$ = K = 512
   i. **ReLU** Non-linearity function is applied.
5. **Convolutional Layer 4**: input is 9x9x512
   a. Filter Size (**F**) = 3 (3x3x512)
   b. No. of Filters (**K**) = 1024
   c. Stride (**S**) = 1
   d. Padding (**P**) = 0
   e. **Parameters** = (3x3x512) x 1024 = 4,718,592
   f. $\mathbf{W_4}$ = 7
   g. $\mathbf{H_4}$ = 7
   h. $\mathbf{D_4}$ = K = 1024
   i. **ReLU** Non-linearity function is applied.
6. **Convolutional Layer 5**: input is 7x7x1024
   a. Filter Size (**F**) = 3 (3x3x1024)
   b. No. of Filters (**K**) = 512
   c. Stride (**S**) = 1
   d. Padding (**P**) = 0
   e. **Parameters** = (3x3x1024) x 512 = 4,718,592
   f. $\mathbf{W_4}$ = 5
   g. $\mathbf{H_4}$ = 5
   h. $\mathbf{D_4}$ = K = 512
   i. **ReLU** Non-linearity function is applied.
7. **Max-Pooling Layer 3**: input is 5x5x512
   a. Filter Size (**F**) = 3 (3x3x512)
   b. Stride (**S**) = 2
   c. **Parameters** = 0
   d. $\mathbf{W_{2m}}$ = 2
   e. $\mathbf{H_{2m}}$ = 2
   f. $\mathbf{D_{1m}}$ = 512
8. **Fully Connected Layer 1**: input is 2x2x512 = 2048
   a. Number of Neurons = 4096
   b. Parameters = (2x2x512) x 4096 = 8,388,608
9. The **total difference in the number of parameters** ZFNet - AlexNet = 1.45 Million

10. There are other variants of ZFNet where we use a stride of 2 in the first convolutional layer, thereby changing the subsequent layer dimensions.