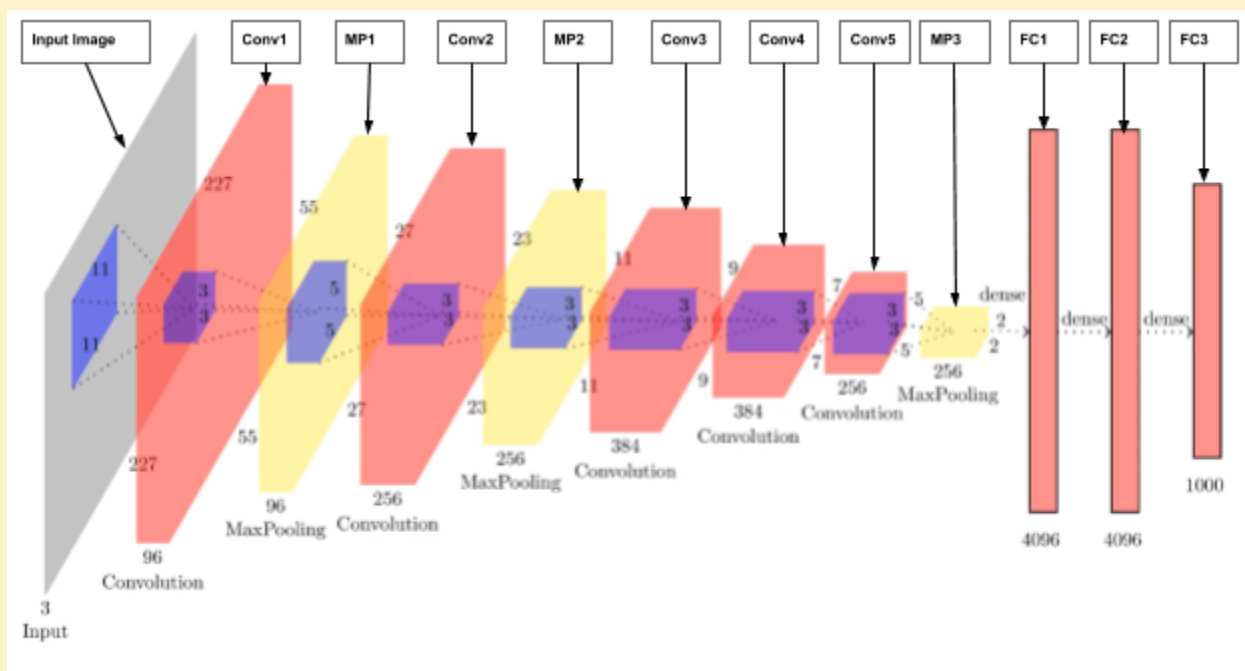


Understanding all the layers of AlexNet

Let's now look at the entire AlexNet

- Before moving into the AlexNet architecture, let us understand why we have decided to use a Convolutional layer instead of a Fully-connected layer
 - In the convolutional layer, the number of parameters is $(11 \times 11 \times 3) \times 96 = 34,848$
 - However, in a FC layer, the number of parameters would be
 - Input: $(227 \times 227 \times 3) \times$ Output: $(55 \times 55 \times 96) = 4.49 \times 10^{10}$
 - Thus, because of sparse-connectivity and weight sharing, we are able to achieve a similar degree of complexity with a significantly smaller number of parameters with a Convolutional layer.
- Now, let us **break down the entire AlexNet architecture**



- Let's look at each of the layers in depth
- Input Layer:** 227x227x3 (colour images of 227x227 Width x Height)
 - $W_{in} = 227$
 - $H_{in} = 227$
 - $D_{in} = 3$
- Convolutional Layer 1:** Input is 227x227x3
 - Filter Size (F) = 11 (11x11x3)
 - No. of Filters (K) = 96
 - Stride (S) = 4
 - Padding (P) = 0
 - Parameters** = $(11 \times 11 \times 3) \times 96 = 34,848$
 - $W_1 = 55$
 - $H_1 = 55$
 - $D_1 = K = 96$
 - ReLU** Non-linearity function is applied to every 2D area in the output volume.
- Max-Pooling Layer 1:** Input is 55x55x96
 - Filter Size (F) = 3 (i.e. 3x3x96)
 - Stride (S) = 4

One Fourth Labs

- c. **Parameters** = 0 (no parameters in max pooling)
- d. $W_{1m} = 27$
- e. $H_{1m} = 27$
- f. $D_{1m} = 96$
- 7. **Convolutional Layer 2:** Input is 27x27x96
 - a. Filter Size (**F**) = 5 (5x5x96)
 - b. No. of Filters (**K**) = 256
 - c. Stride (**S**) = 1
 - d. Padding (**P**) = 0
 - e. **Parameters** = (5x5x96) x 256 = 614,400
 - f. $W_2 = 23$
 - g. $H_2 = 23$
 - h. $D_2 = K = 256$
 - i. **ReLU** Non-linearity function is applied.
- 8. **Max-Pooling Layer 2:** input is 23x23x256
 - a. Filter Size (**F**) = 3 (3x3x256)
 - b. Stride (**S**) = 3
 - c. **Parameters** = 0
 - d. $W_{2m} = 11$
 - e. $H_{2m} = 11$
 - f. $D_{2m} = 256$
- 9. **Convolutional Layer 3:** input is 11x11x256
 - a. Filter Size (**F**) = 3 (3x3x256)
 - b. No. of Filters (**K**) = 384
 - c. Stride (**S**) = 1
 - d. Padding (**P**) = 0
 - e. **Parameters** = (3x3x256) x 384 = 884,736
 - f. $W_3 = 9$
 - g. $H_3 = 9$
 - h. $D_3 = K = 384$
 - i. **ReLU** Non-linearity function is applied.
- 10. **Convolutional Layer 4:** input is 9x9x384
 - a. Filter Size (**F**) = 3 (3x3x384)
 - b. No. of Filters (**K**) = 384
 - c. Stride (**S**) = 1
 - d. Padding (**P**) = 0
 - e. **Parameters** = (3x3x384) x 384 = 1,327,104
 - f. $W_4 = 7$
 - g. $H_4 = 7$
 - h. $D_4 = K = 384$
 - i. **ReLU** Non-linearity function is applied.
- 11. **Convolutional Layer 5:** input is 7x7x384
 - a. Filter Size (**F**) = 3 (3x3x384)
 - b. No. of Filters (**K**) = 256
 - c. Stride (**S**) = 1
 - d. Padding (**P**) = 0
 - e. **Parameters** = (3x3x384) x 256 = 884,736
 - f. $W_5 = 5$
 - g. $H_5 = 5$

One Fourth Labs

- h. $D_5 = K = 256$
- i. **ReLU** Non-linearity function is applied.
- 12. **Max-Pooling Layer 3:** input is $5 \times 5 \times 256$
 - a. Filter Size (**F**) = 3 ($3 \times 3 \times 256$)
 - b. Stride (**S**) = 2
 - c. **Parameters** = 0
 - d. $W_{5m} = 2$
 - e. $H_{5m} = 2$
 - f. $D_{5m} = 256$
- 13. **Fully Connected Layer 1:** input is $2 \times 2 \times 256 = 1024$
 - a. Number of Neurons = 4096
 - b. Parameters = $(2 \times 2 \times 256) \times 4096 = 4,194,304$
- 14. **Fully Connected Layer 2:** input = 4096
 - a. Number of Neurons = 4096
 - b. Parameters = $4096 \times 4096 = 16,777,216$
- 15. **Fully Connected Layer 3:** input = 4096
 - a. Number of Neurons/output-classes = 1000
 - b. Parameters = $4096 \times 1000 = 4,096,000$
- 16. Totally, there are around 27.55 Million parameters, out of which roughly 25 Million parameters were in the last 3 Fully-connected layers.
- 17. When counting the total number of layers, we do not include the max-pooling layers as they do not carry weights. Thus, we say that **AlexNet has 8 layers**