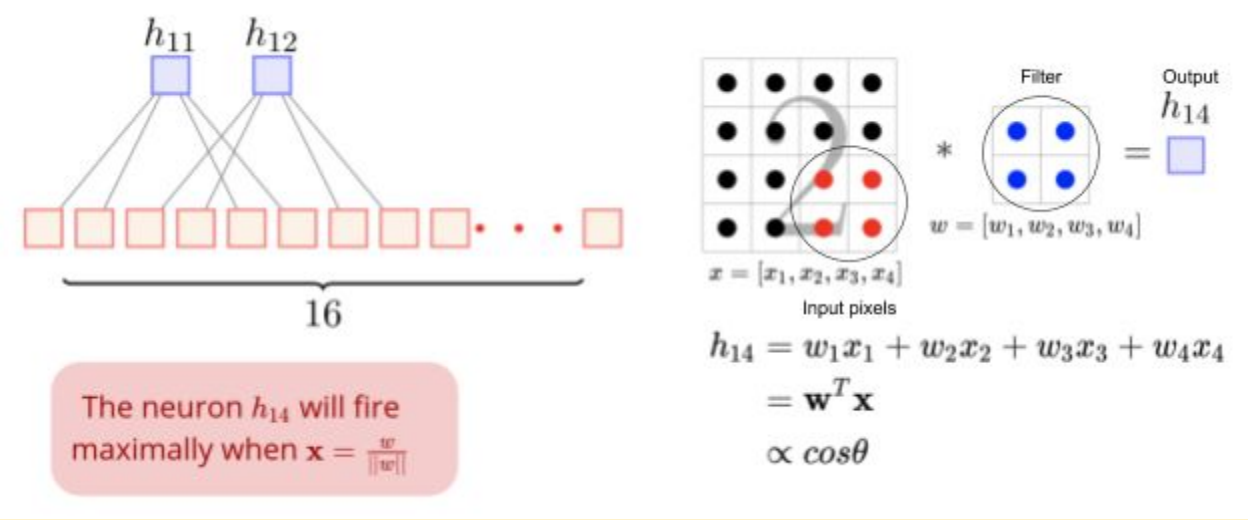
PadhAI: Visualising CNNs

One Fourth Labs

Visualising filters

What does a filter capture?

1. We have dealt with filters in all our CNN models so far. Now the question is, what exactly does a filter capture?
2. Let's look at the working of a 2x2 filter on a 4x4 input image



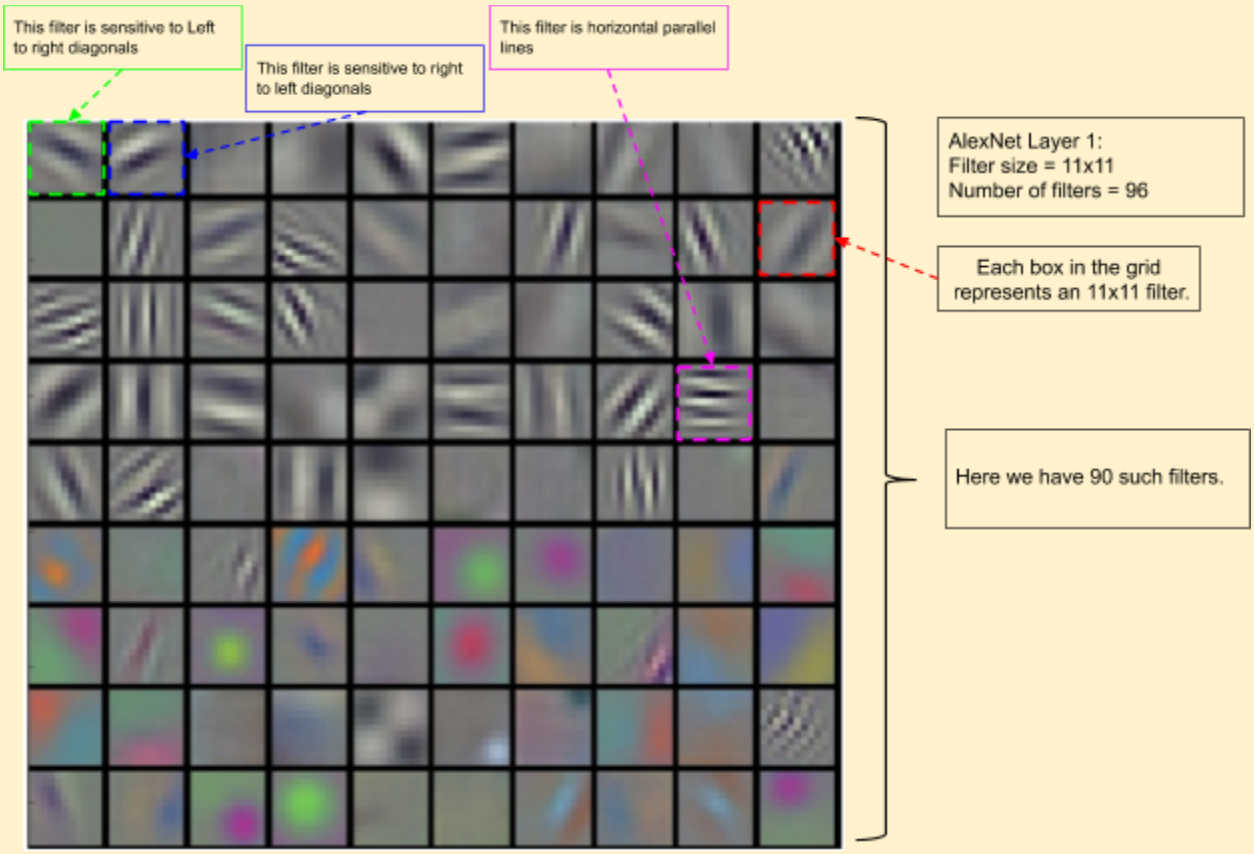
a. Here, the input image is 4x4 while the fiter is 2x2
b. The red input pixel vector $x = [x_1, x_2, x_3, x_4]$
c. The weight vector $w = [h_1, h_2, h_3, h_4]$
d. By convolving the input pixels with the filter, we get the output
e. Output $h_{14} = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4$
f. $h_{14} = w^Tx$ (This is the same as the dot product between the two vectors)
g. $h_{14} \propto cos(\theta)$ $[where\ \theta\ is\ the\ angle\ between\ the\ two\ vectors]$ $cos(\theta) = \frac{w^Tx}{||w||\ ||x||}$
h. Now for certain inputs, we want the filter to fire (give a high value).
i. Now, h_{14} will be high when $cos(\theta)$ is high (cos(θ) = 1), i.e. when θ is 0. This implies the two vectors w and x are in the same direction.
j. So, we can say that an input vector which aligns with a filter vector yields maximum output.
k. The neuron h_{14} will fire maximally when $x = \frac{w}{||w||}$ (x is a unit vector in the direction of w)
l. Thus, when we **slide the 2x2 filter w** across the 4x4 input region, whenever we **reach a 2x2 region x** that looks exactly like the filter, we get a high output. For all other regions which do not align with the filter, the output is low.

PadhAI: Visualising CNNs

One Fourth Labs

---

3. Now, let us visualize the filters in AlexNet



   a. The above image shows us how different patterns are identified by different filters.