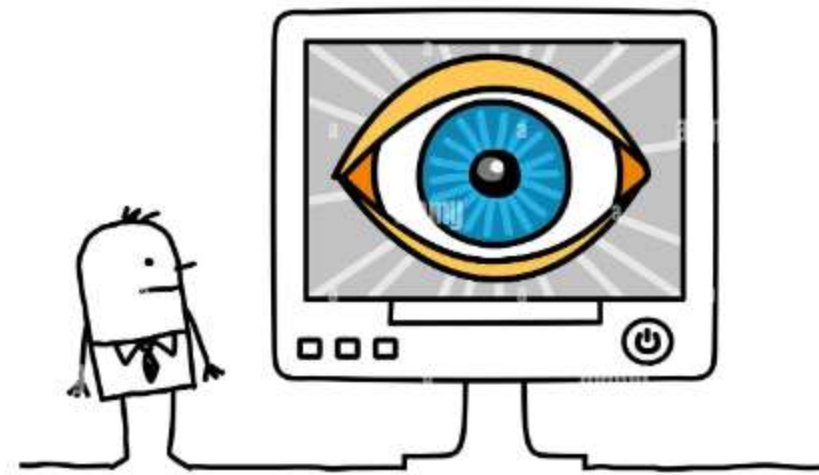


COMPUTER VISION



Vision and perception are very natural to humans.

But is it same to computers?

NO !!



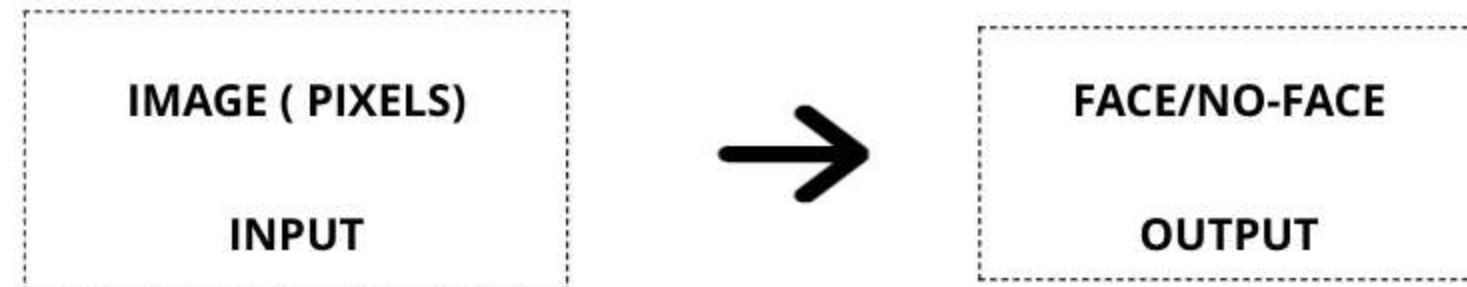
1970's in MIT

Which is more important : Data or algorithm?

The essence of computer vision is not solely about developing complex algorithms
but about enabling a system to relate its current visual input to its past experiences
and knowledge.

WHY??

EX : FACE DETECTION SYSTEM



The idea of face can't be mathematically conveyed to a computer !!

How does CV does it then?

Pattern recognition

Feature detection

Can you now think of some applications of CV ?



Main CV Tasks



Object detection



Object segmentation



Object landmark detection



Object recognition

Object identification



Object verification

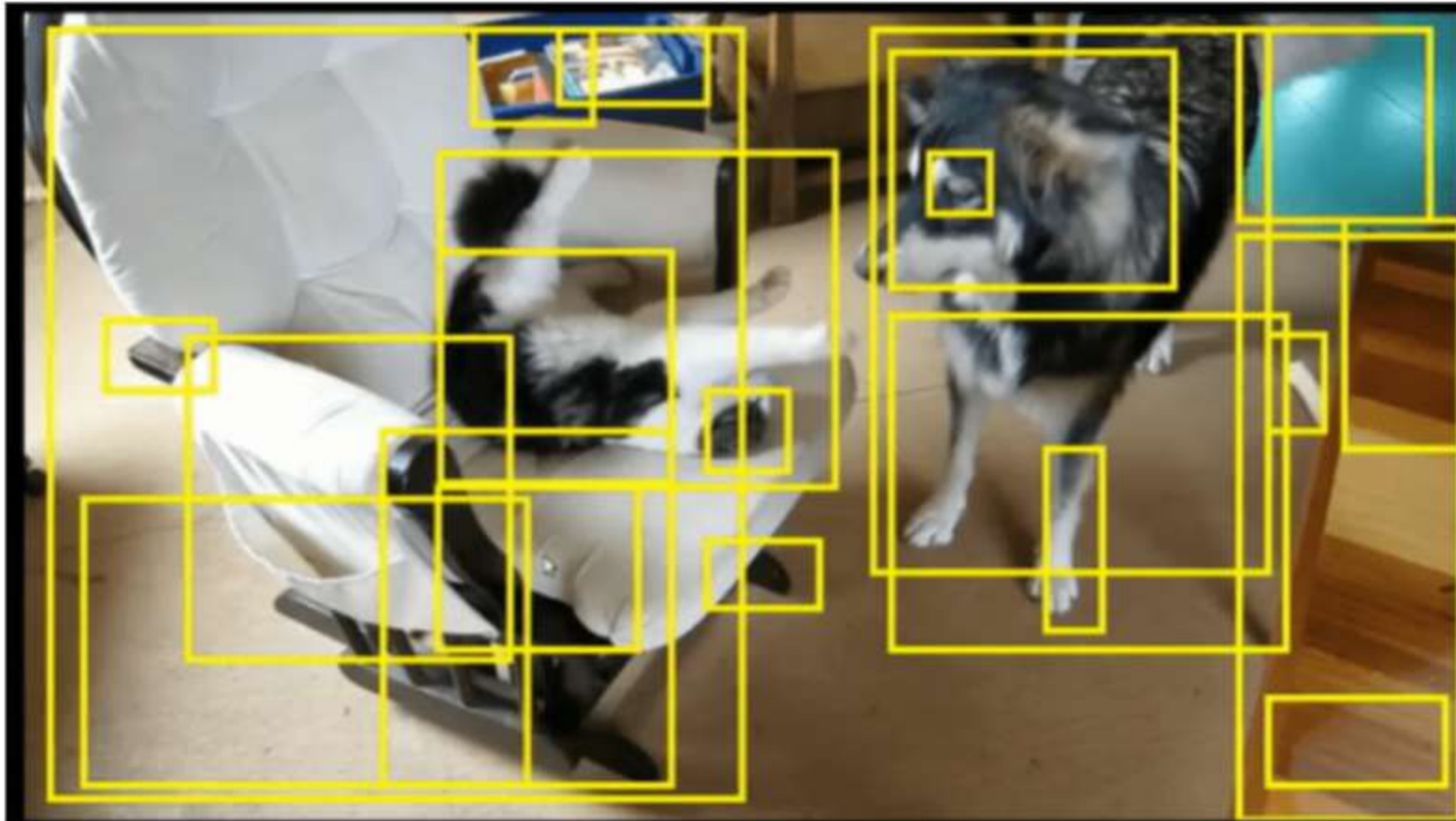
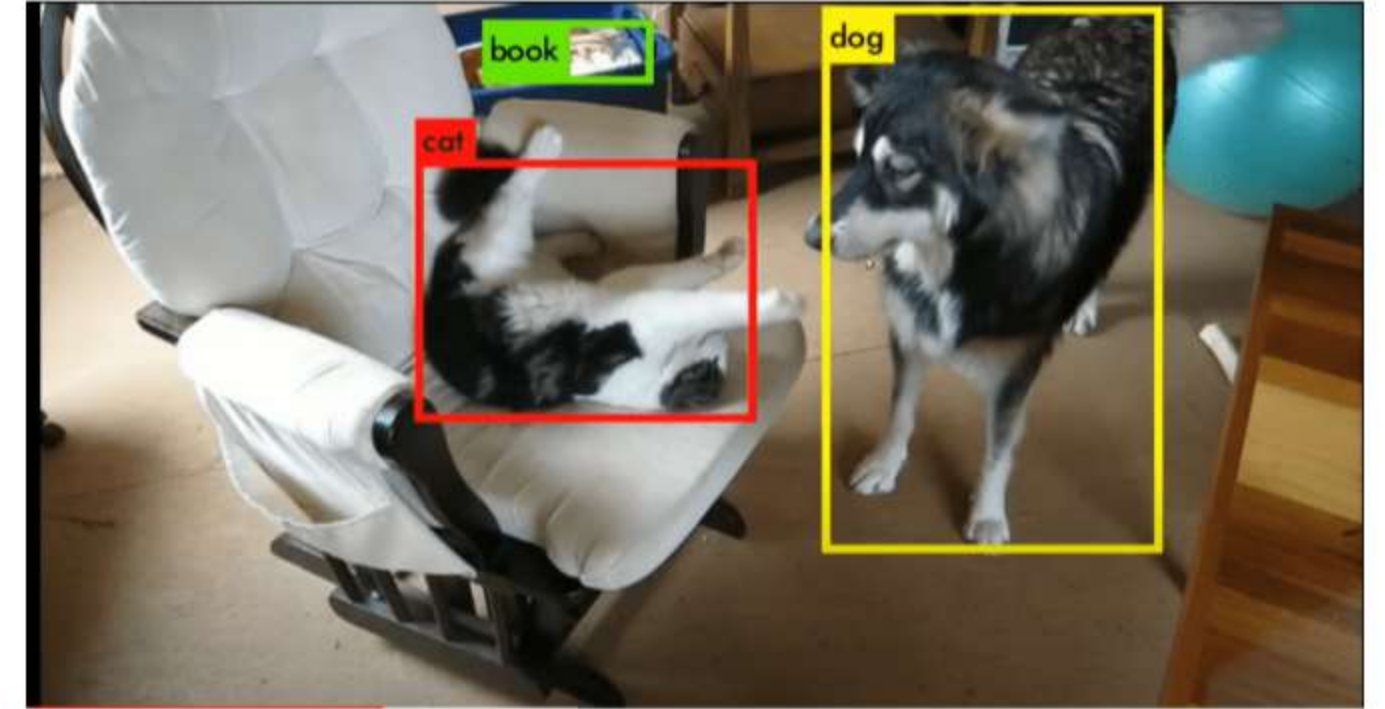


Object classification



OBJECT DETECTION

This involves not only identifying objects within an image but also locating them by drawing bounding boxes around each object.

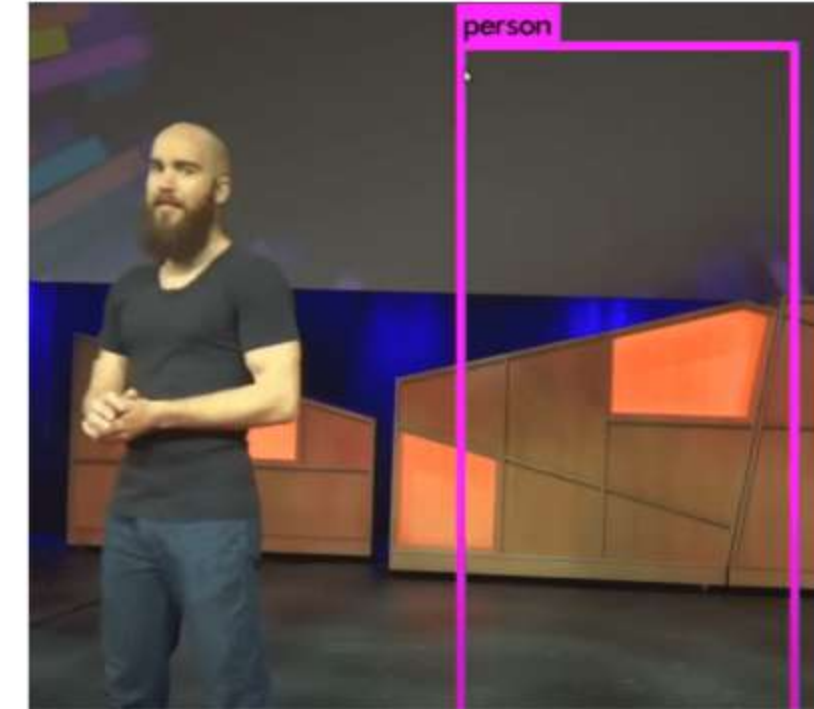


Split the image into a bunch of regions and run a classifier on each of these regions, then high score for a classifier means, detection of object. But this would require to run the classifier thousands of time over a same image

BUT WHAT'S THE PROBLEM WITH THAT ?

SPEED !

We have transitioned from processing static images to processing videos, in which we have continuous motion of the objects.

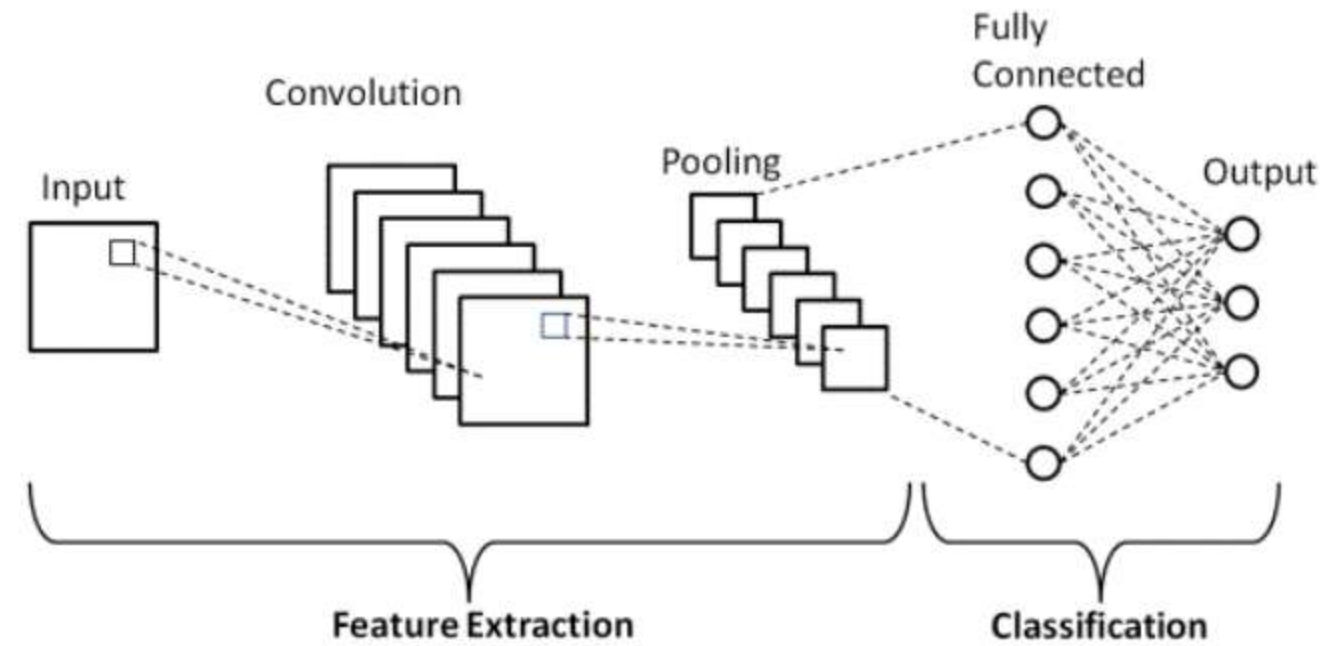


A slower detection in video would lead to something like this

The YOLOv5 logo, consisting of the text "YOLOv5" in white, bold, sans-serif font, centered on a solid black rectangular background.

<https://www.youtube.com/watch?v=VOC3huqHrss>

CONVOLUTION NEURAL NETWORK



Convolutional Layer: This layer uses filters (kernels) to detect features.

Pooling Layer: This layer reduces the size of the feature maps from the convolutional layer, making the network more efficient.

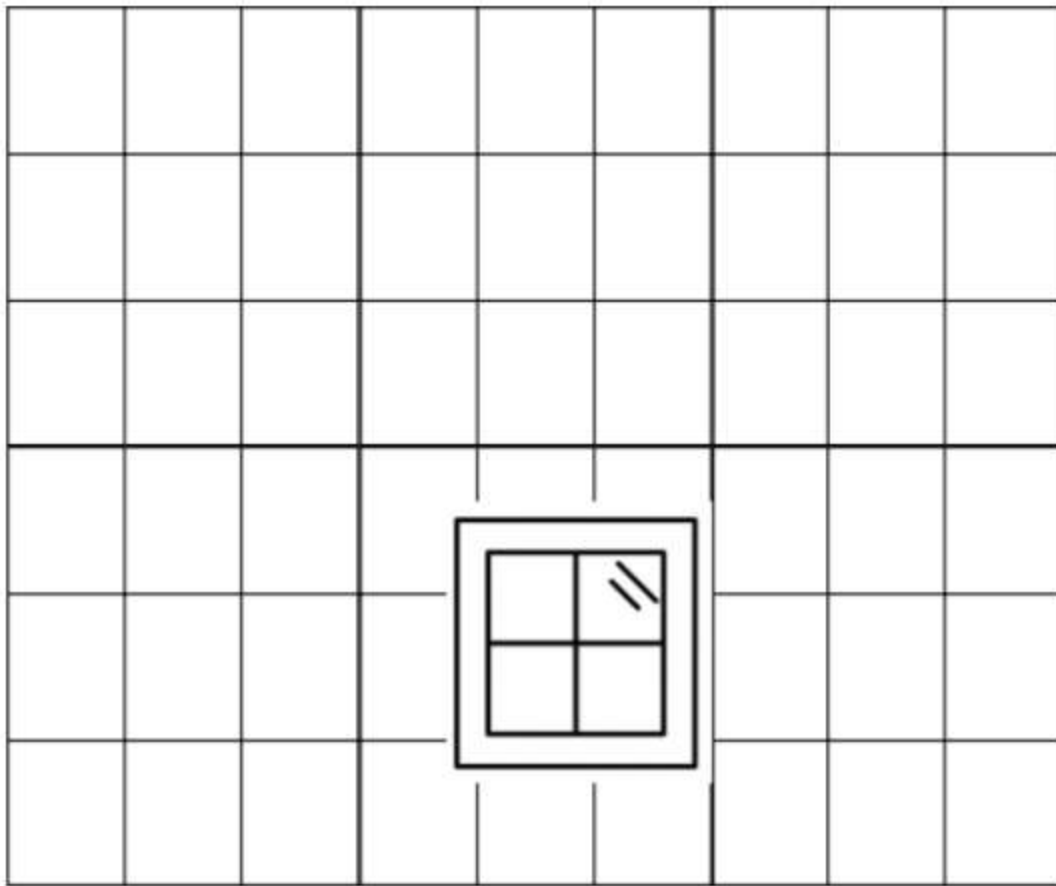
Fully Connected Layer: This layer takes the high-level features produced by previous layers and translates them into a final output, like classifying the image.



Filters (kernel) : Performs pattern recognition

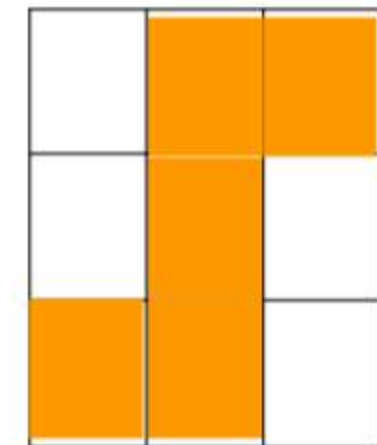
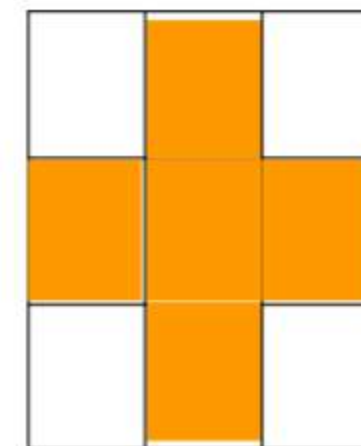
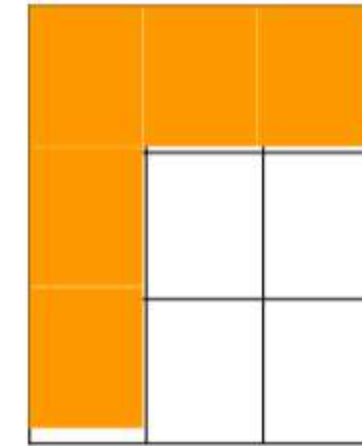
Let's take a particular part of an image, for example we

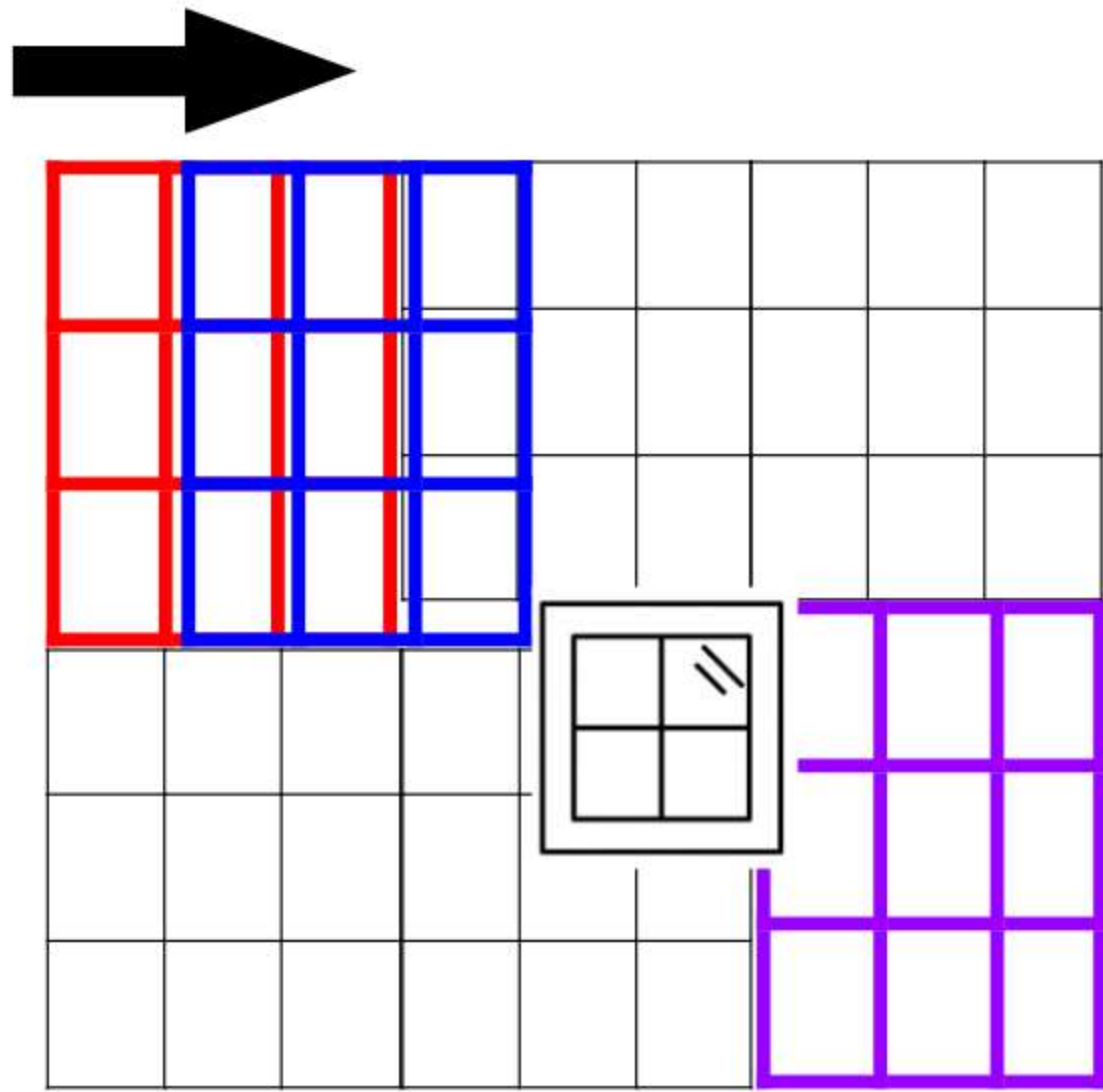
take the window (pixels)



A different window could also have curved surface. The beauty of CNN is, it recognizes both of

A filter is a 3 X 3 block and within that we define what is the pattern we are looking for.

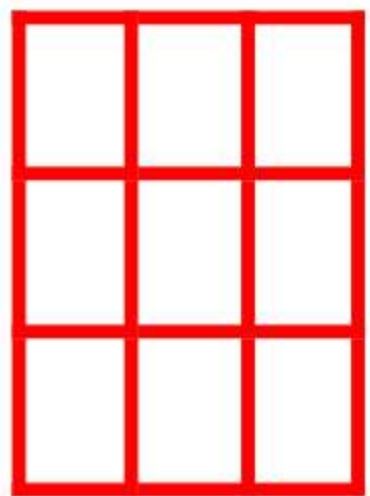




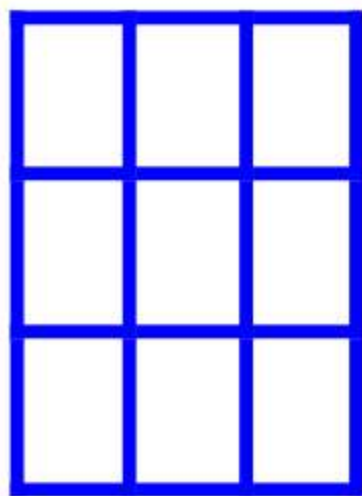
We take any one of the filters and compare them with a 3×3 subset block of the image and analyze how close the pattern in the image is similar to the filter. We get a numeric score of how similar the patterns are

Similarly we slide over through every 3×3 subset block in the image

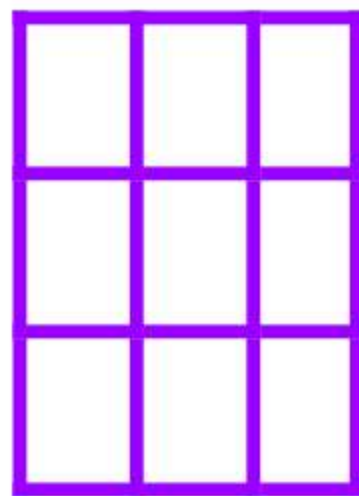
We do the same for other filters too. we take the numeric array from all these filters and combine them together in a process called pooling, we will have a better idea of what is represented in the group of pixels



Step 1



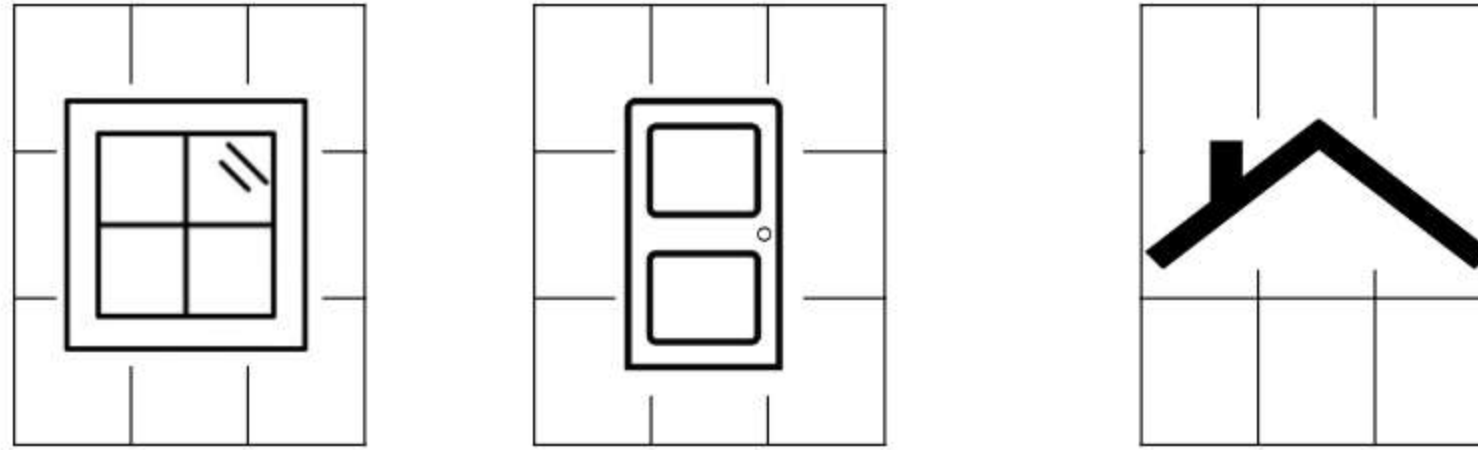
Step 2



Step n

<https://commons.wikimedia.org/wiki/File:CNN-filter-animation-1.gif>

As we go deeper into neural network, the filters becomes more abstract. The second layer may perform like basic object detection such as presence of window or a door or a roof



As we deeper, the filters perform even more abstract tasks like being able to determine if we are looking at a house or an apartment or a skyscraper. So the application of filer keeps increasing as we go through the network



IT ALL SEEMS EASY, DOESN'T IT ?

SO, IS THIS ALL CV IS ?

NO

THANK YOU :)