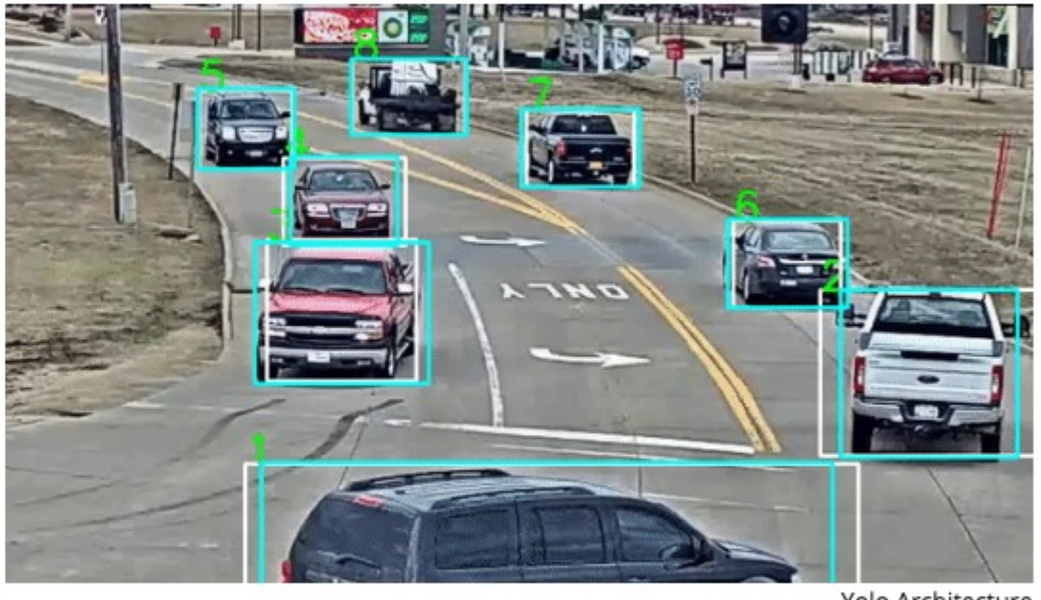
Image analysis is another task that has seen an impressive improvement over the las years because of neural networks, most specifically, Convolutional Neural Networks.

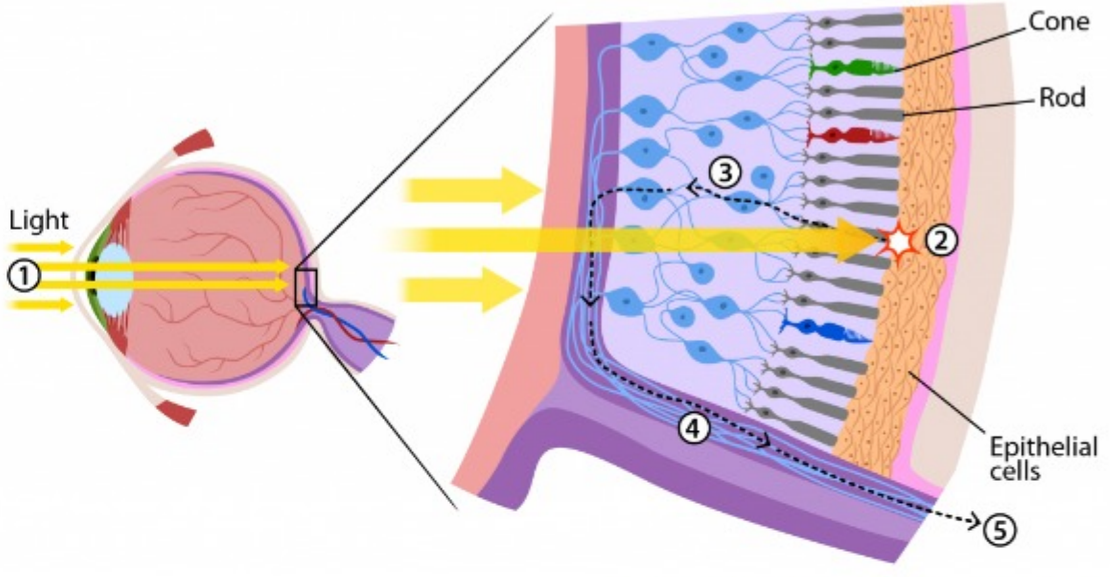
But before we take a dive into these architectures, let's see a little bit of what was image processing before NNs.



Yolo Architecture

How does the computer "see"?

As Humans, we use our eyes to receive information in the form of light that is reflected from the world around us. More concretely, we use cells inside our eyes called cones, which are stimulated by three specific frequencies in the light and our brain understands as color. In low light conditions, we use other cells called rods and are not color-specific.

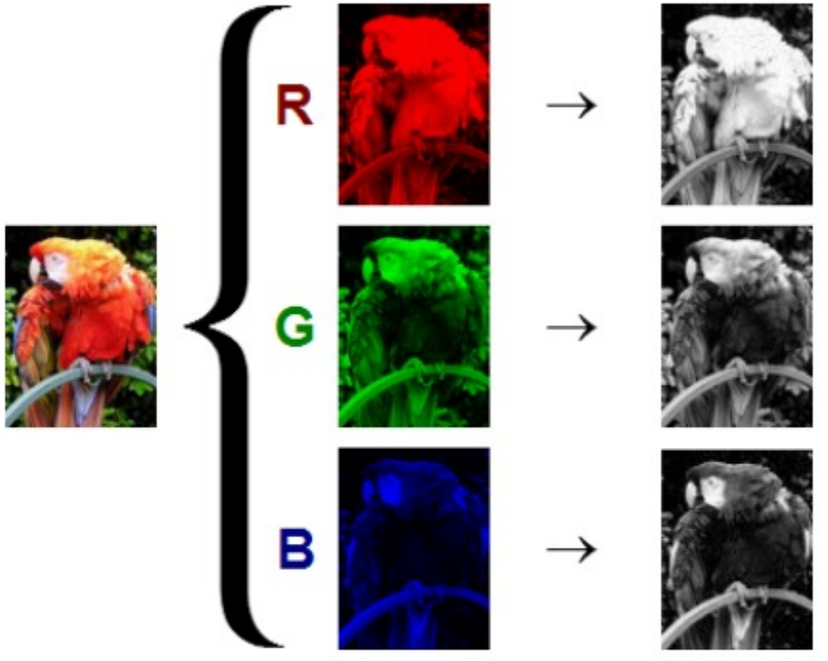


https://askabiologist.asu.edu/rods-and-cones

By comparison, normal color cameras use electronic sensors that also react to specific light frequencies and produce color images. While night vision cameras usually are sensible to the infrared part of the spectrum and do not produce colored images.

These images that are captured by the camera are stored in the computer's memory as arrays of different dimensionality depending on the image type.

2D arrays for B&W images
3D arrays for normal color images or B&W video
4D arrays for colored video



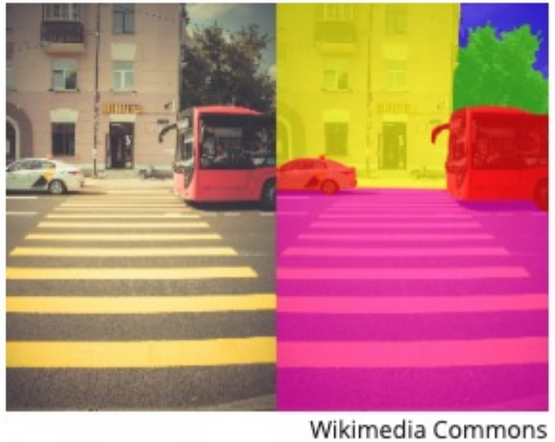
Wikimedia Commons

Usual tasks in Computer Vision

1. Transformations
    a. Change brightness/contrasts
    b. Make them sharper/blurrier
    c. Color Correction
    d. Filters
    e. Scaling
2. Analysis
    a. Find things
    b. Track things
    c. OCRs (Recognize written characters)
    d. ...

openCV

Wikimedia Commons

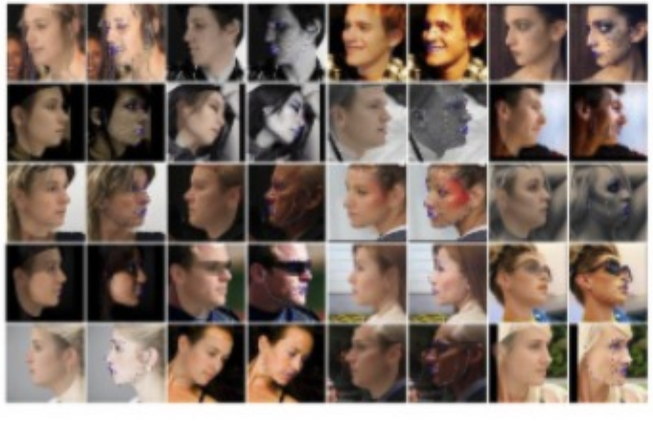


Wikimedia Commons

pyimagesearch

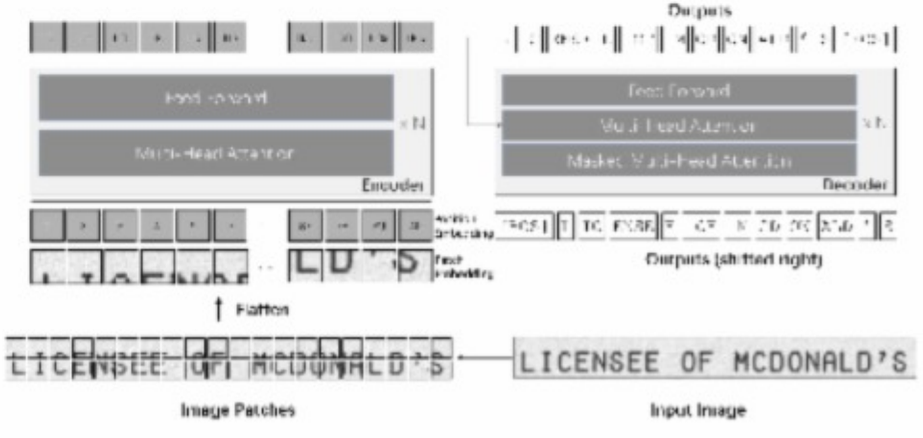
How many of these tasks can be taken by a Neural Network?

Face Alignment Across Large Poses: A 3D Solution

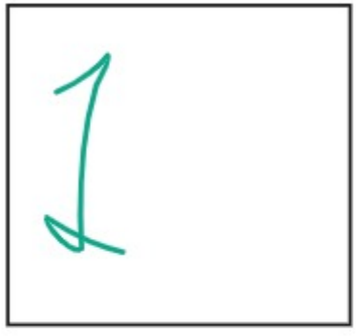


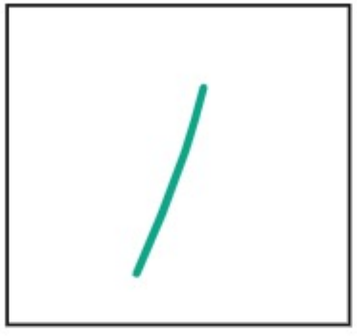
Semantic White Balance: Semantic Color Constancy Using Convolutional Neural Network

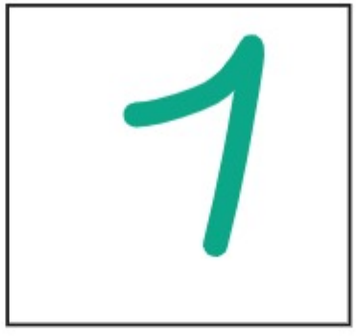
(a) Input image (b) Gray world (c) Adobe Photoshop (d) Semantic mask (e) Ours



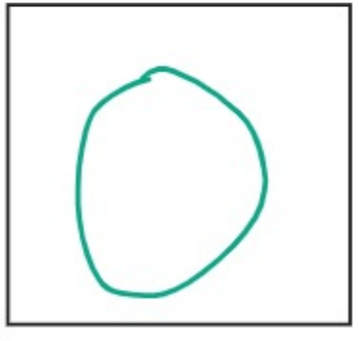
Let's go back to a known problem, hand-written digits

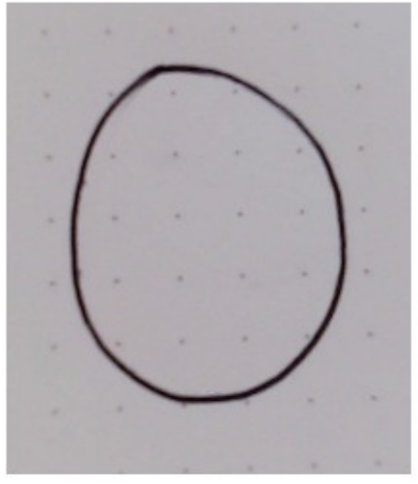


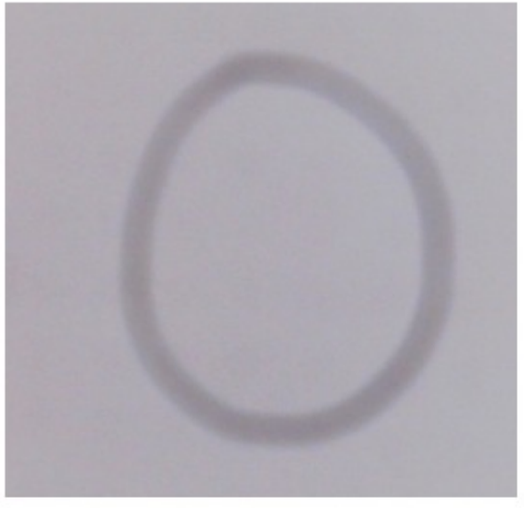


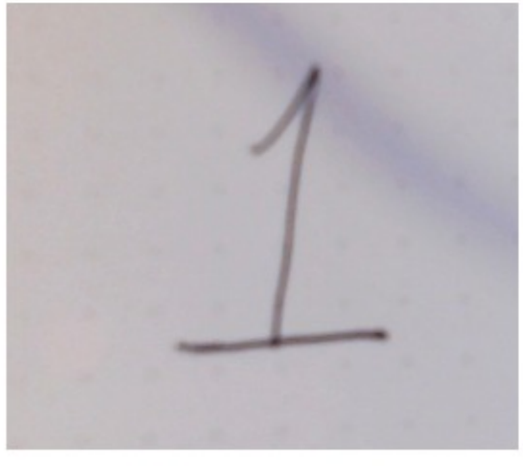


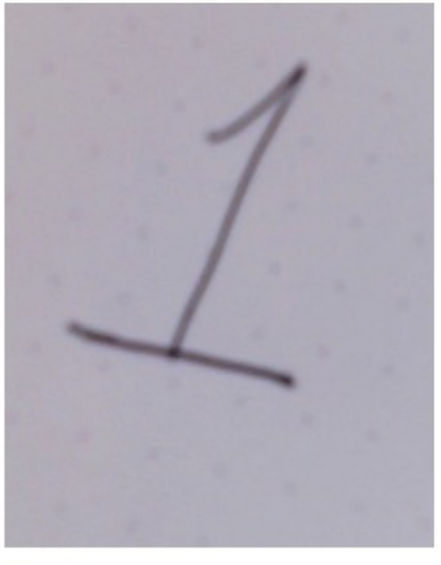
How many ways to write a number can be? What sources of variability do you think there can be, aside from just the fact that each person has their own way to write?







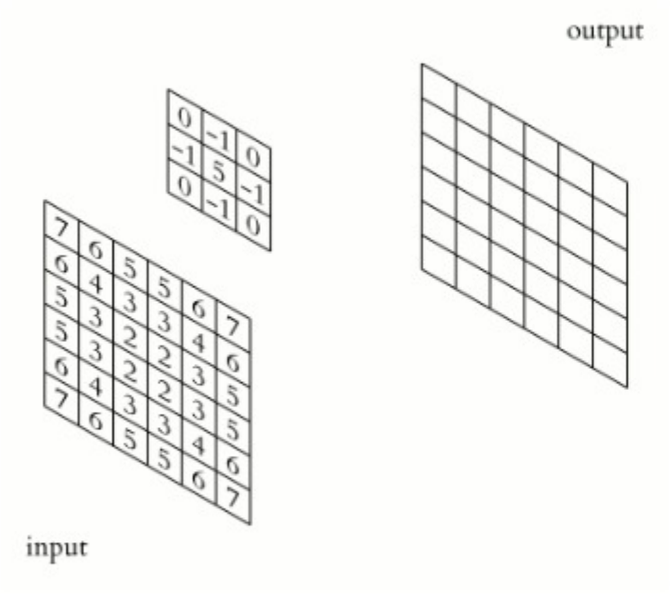


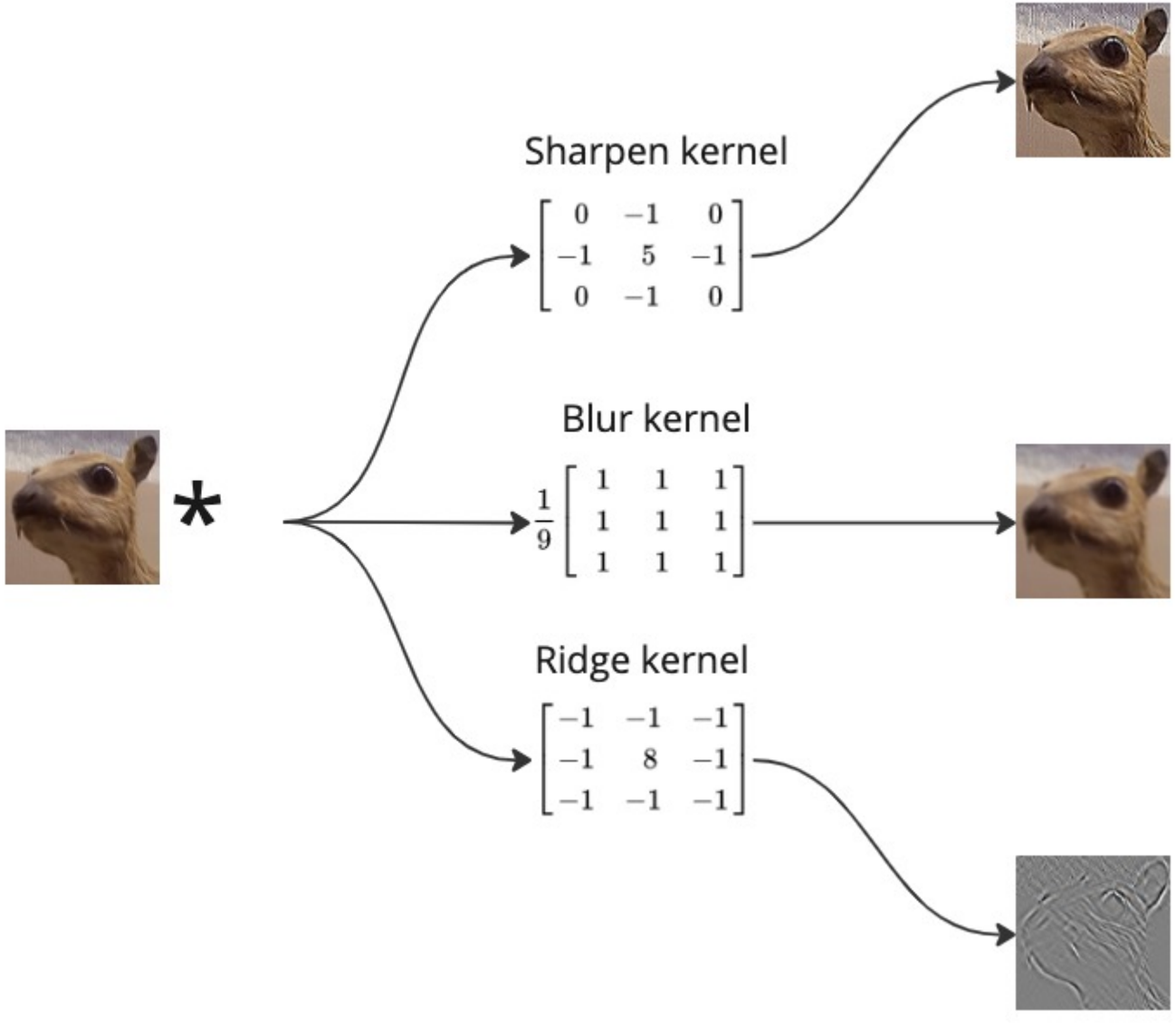


There are infinite ways to take photos of the same thing, this is one of the biggest shortcomings of traditional computer vision approaches, it is nearly impossible to manually write code for every possible situation...

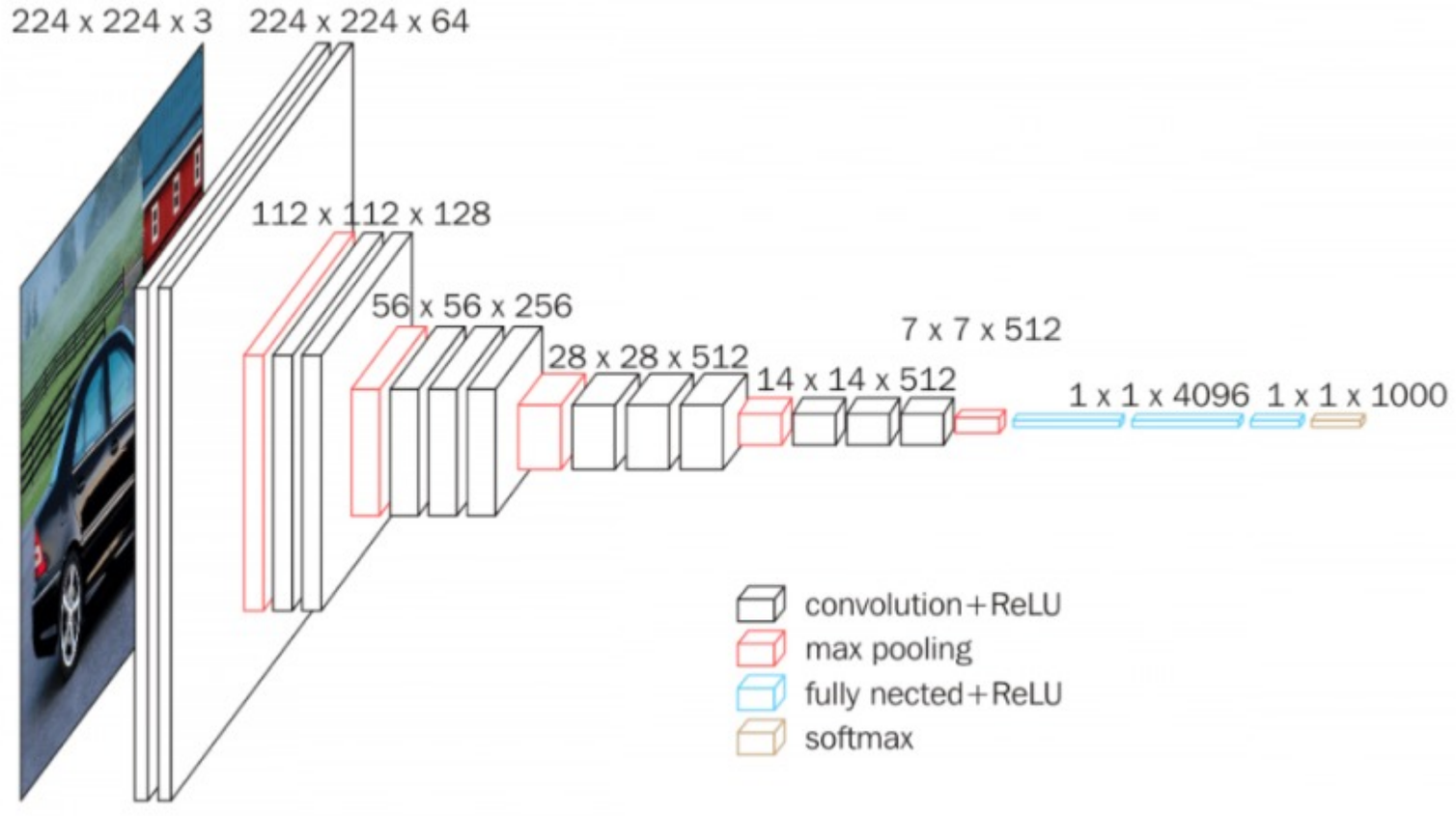
So... How can we use neural networks to take on computer vision tasks? How can we create a neural network that differentiates not only hand-written digits, but also cats from dogs, or even a person smiling or frowning?

The answer is ✨ Convolutions✨





So, what is the logic of the CNNs?



The image produced by each kernel is called a feature map

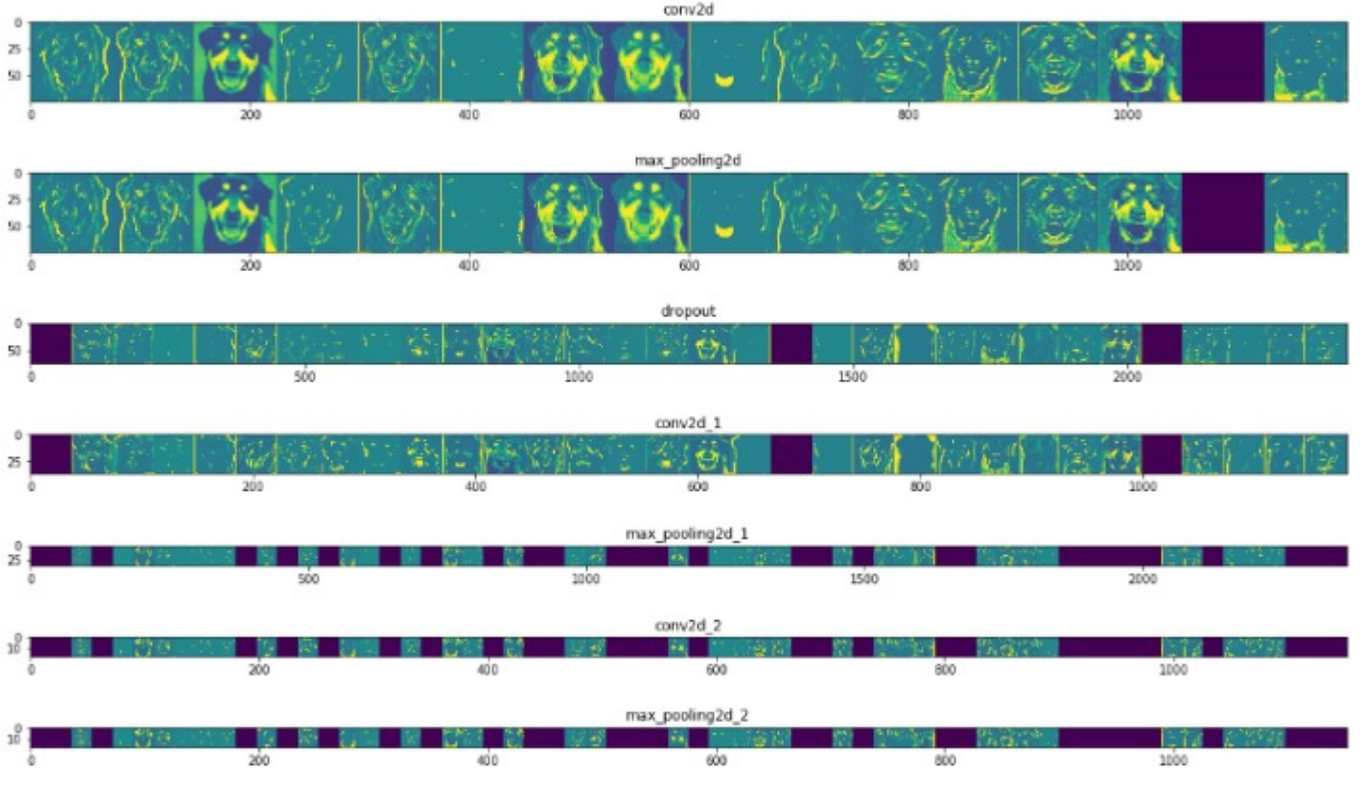
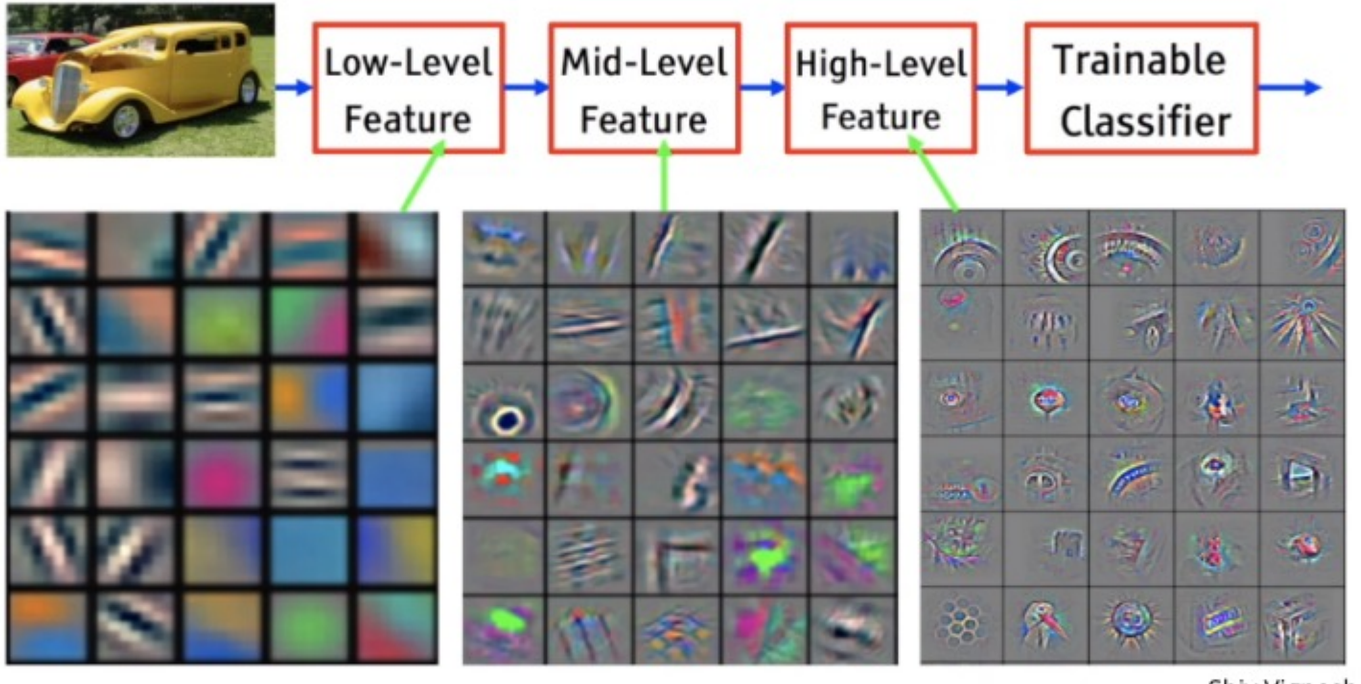


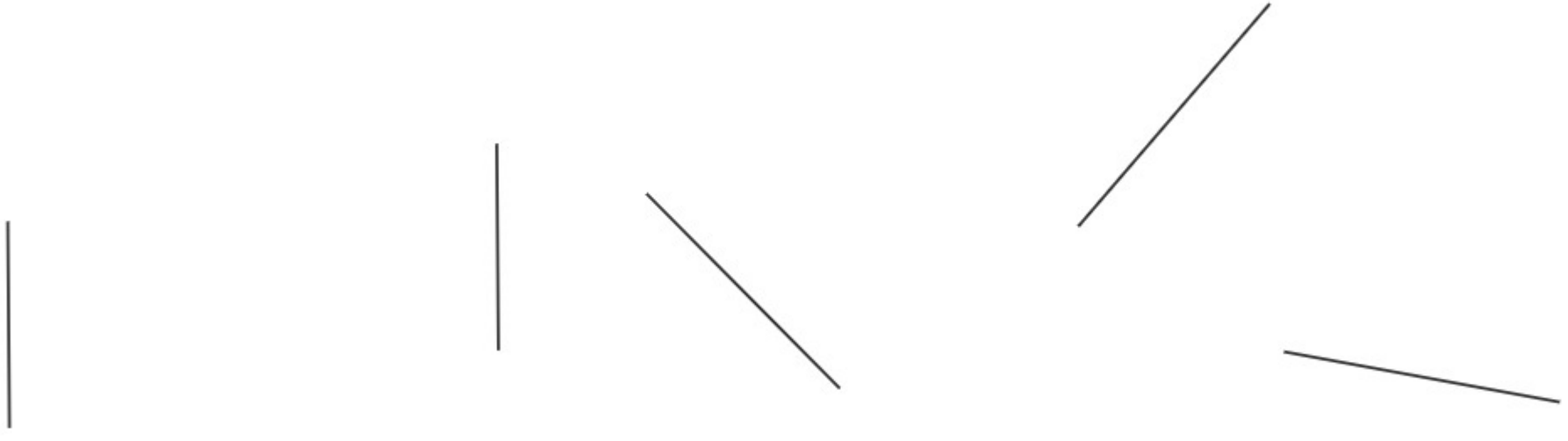
Image after first layer

Image after last layer

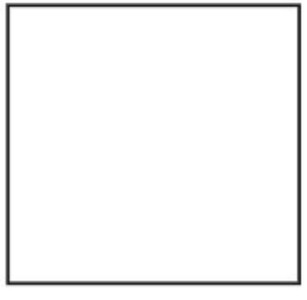


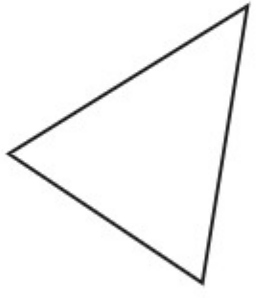
Each convolutional layer that we add is trained to find more complex patterns...

So, first layers find straight lines, color blops, curves in one or other direction.



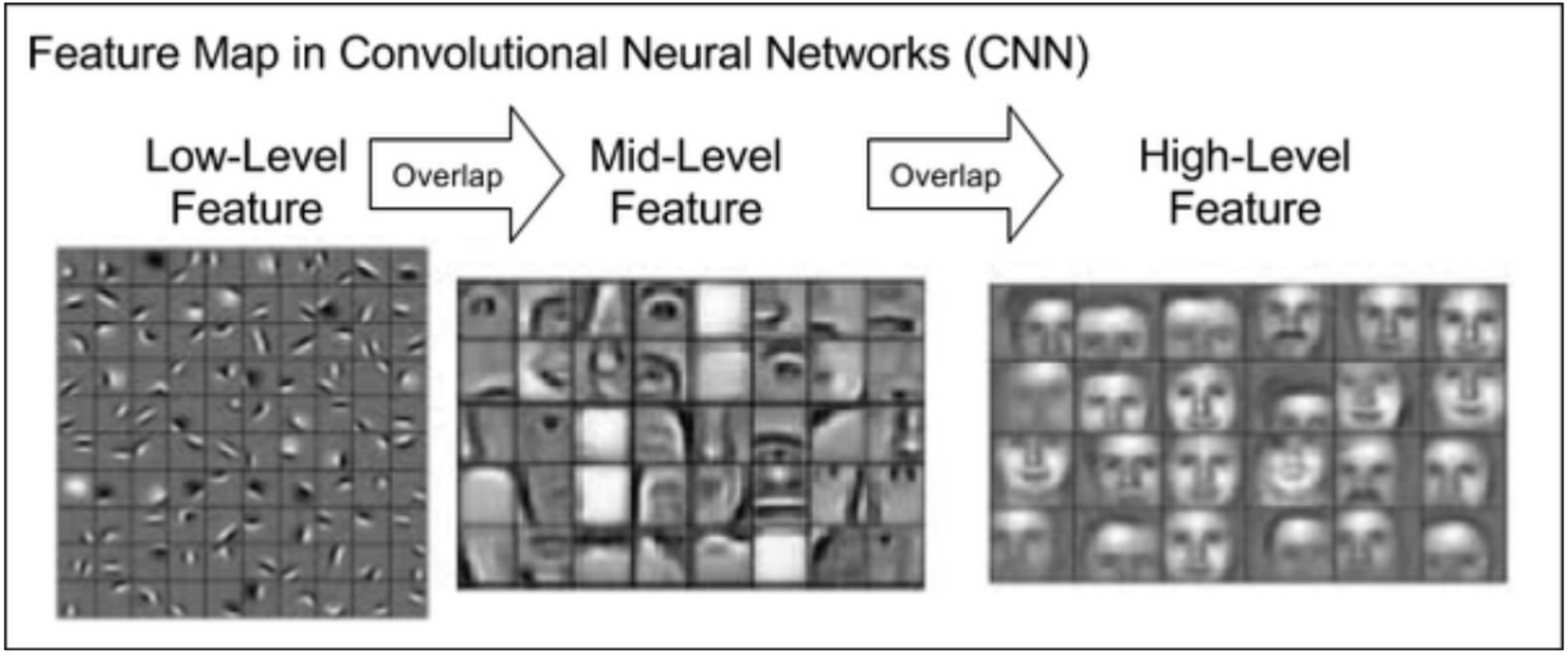
Then, the middle layers pick up these features: two vertical lines and two horizontal lines? Hmmm, that could be a square. And these two oblique lines in addition to a horizontl one looks like a triangle!



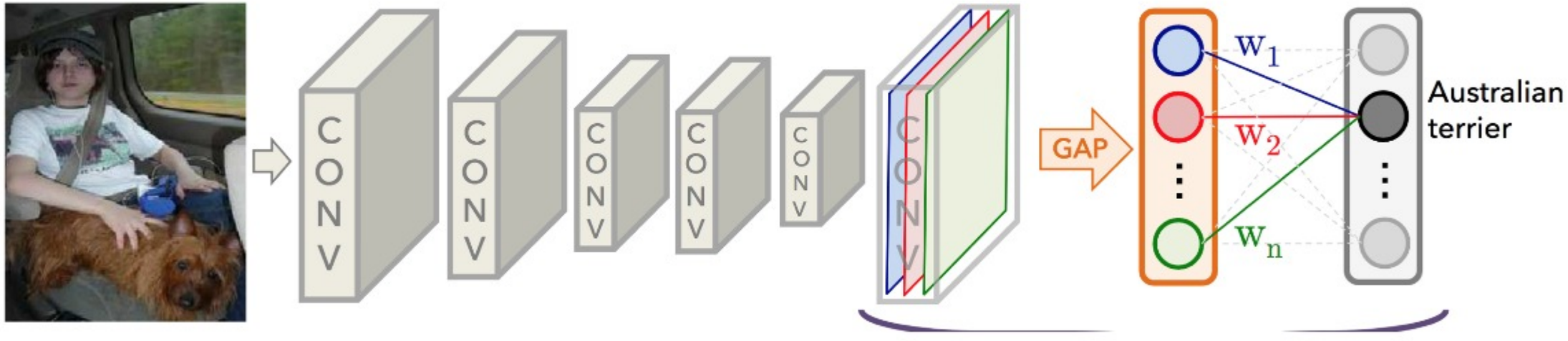


Finally, the last layers, could say something like: Well... that is a square with a triangle on top so... It must be a house!

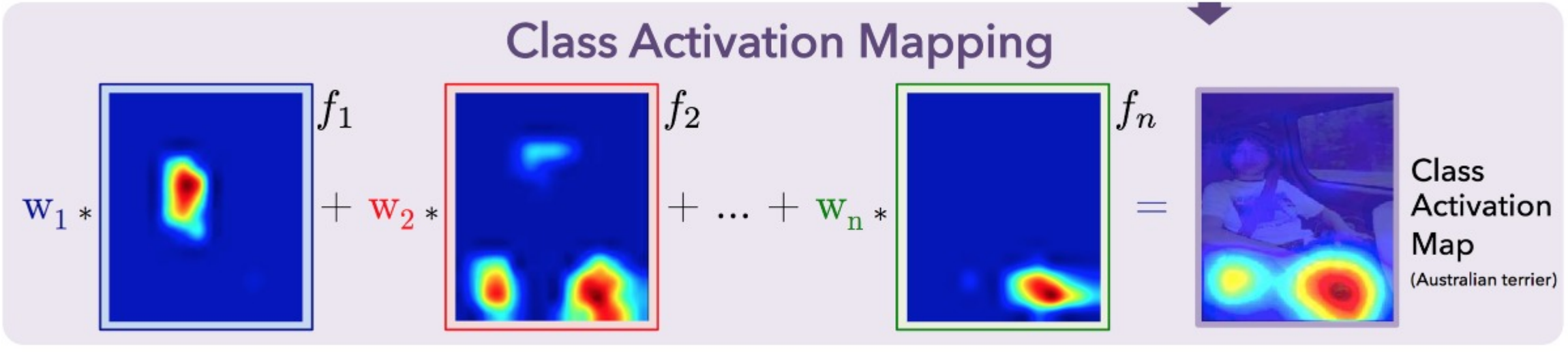
So in the same way, a human face is made up by eyes, noses, etc...



And if you want to classify lots of classes?...



And if you want to know which pixels are from the dog, and which from the kid?



You see which pixels were responsible for the highest activation of each class and then create a mask for them