AI and image recognition

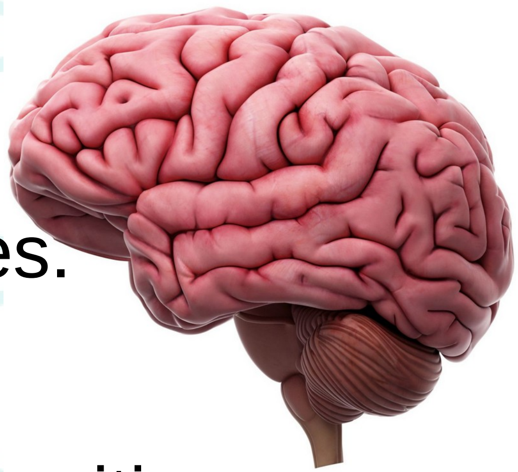
Dr. Charles Kind

Bristol University

May 2023

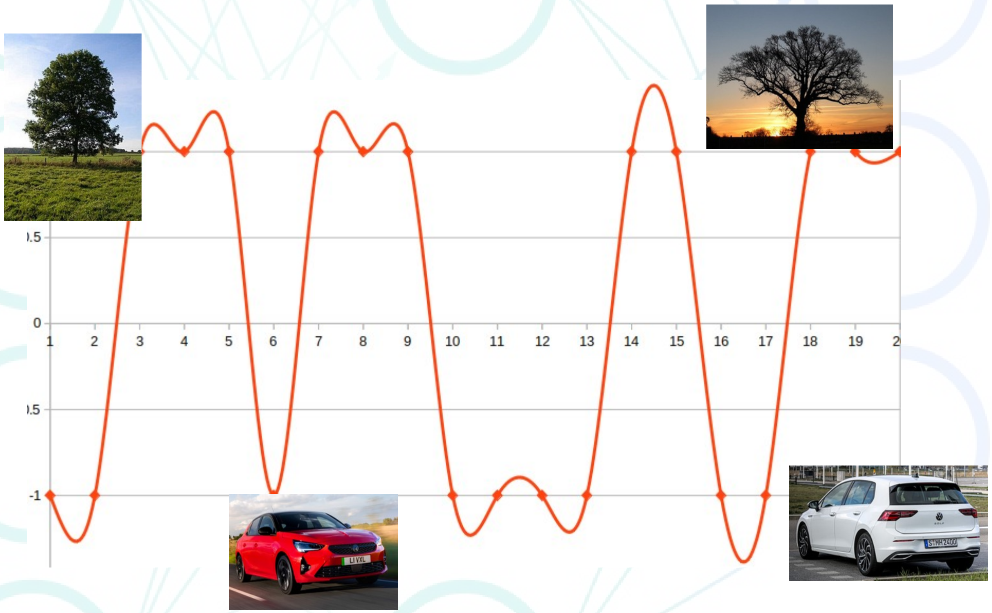
Summary of learning objectives

- Why do we want to recognise images?
- Why do standard neural networks fail at these tasks?
- What property of images makes them hard to process?
- Image and object recognition.
- How the human brain processes images.
- Simplification of images.
- Neural network solutions to image recognition.



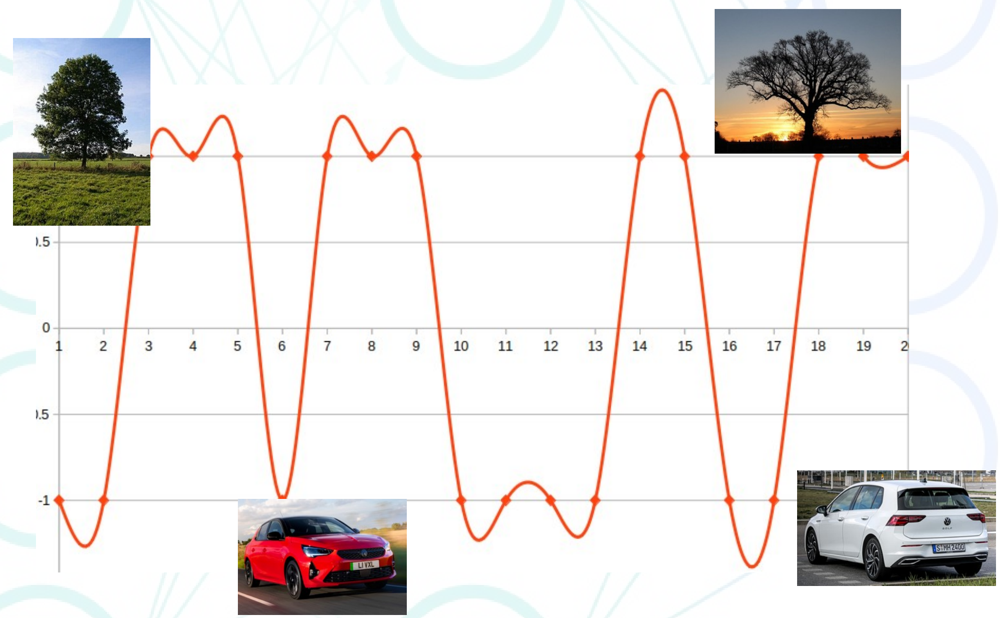
Why neural networks failed to recognise images

- Standard neural networks of whatever depth are essentially curve, or surface, fitting functions.
- Imagine trying to plot a graph of two different types of images, say trees and cars.
- We can train our classical neural network on an arbitrarily large dataset and surely then it could recognise cars and trees? NO!
- Why not?



Why neural networks failed to recognise images

- Standard neural networks of whatever depth are essentially curve, or surface, fitting functions.
- Imagine trying to plot a graph of two different types of images, say trees and cars.
- We can train our classical neural network on an arbitrarily large dataset and surely then it could recognise cars and trees? NO!
- Why not?



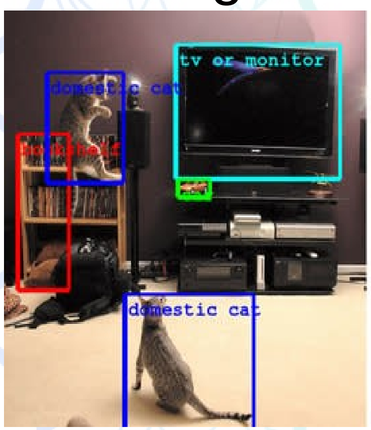
- We cannot describe an image of this level of complexity with a single number. A pixel alone uses three numbers.

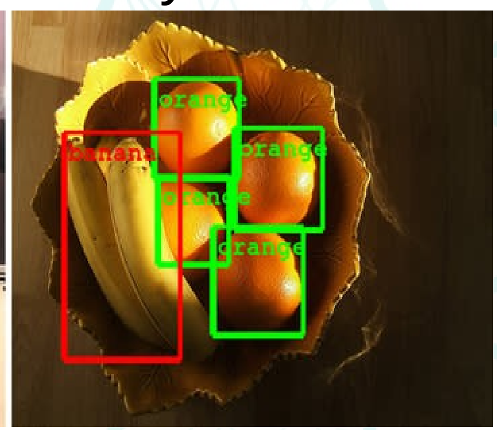
What are images composed of?

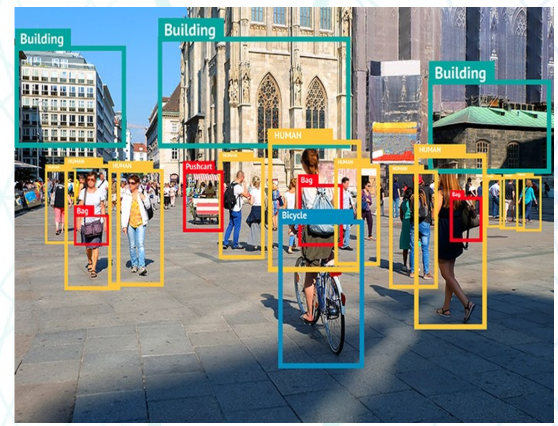
- A colour picture is composed of a number of pixels, each of which have at least three floating point values. Let us call them (R,G,B).
- Consider a 32x32 image. It has 1024 pixels and therefore 3072 numbers per image.
- Imagine you want to fill the problem space with a reasonable number of data points for learning. Let us say 3 for each dimension (which is not very reasonable).
- Why do this? Each dimension reduces the relative density of local points.
- You will therefore need to label around 3^3072 images for training.
- To give you some perspective it is estimated that there are 3^81 atoms in the universe.
- Oh … well … this is a large problem space we want to fit our network to!

Image and object recognition

- If we are classifying just two distinct classes, trees and cars, we would need to relatively densely populate their respective areas of the entire space of 32x32 images in order for curve fitting to work.
- What if there is a long “distance” between the classes? Then my curve fitting will be very inaccurate.







- What if instead I want to recognise different objects within the image? Then I roughly increase the complexity by ^n, where n is the number of different objects. I am also now comparing sub-images of varying sizes.
- What if I want to do this with my phones camera images? It is 8700×5800 or ~50 megapixels. That is a much larger dimensional space to traverse.

What can we do to solve this problem?

- We can sometimes recognise images that have been pixelated, or reduced in dimension. Greyscale reduces the number of colour numbers to one.
- Perhaps we can reduce the dimension of the images. Why could this fail to work and in what sort of image recognition tasks?

- We could look at small, say 3x3, subsets of the entire image. But what if no sub-images contain recognisable features that can tell us about the whole image?
- How do we solve this problem ourselves? We can look at a “real” image and, within reason, identify everything in it that we have seen before. What are we doing that our DNN cannot?
- We will go back to our 3x3 subsets and see if we can find inspiration there.

Some points about how humans process images

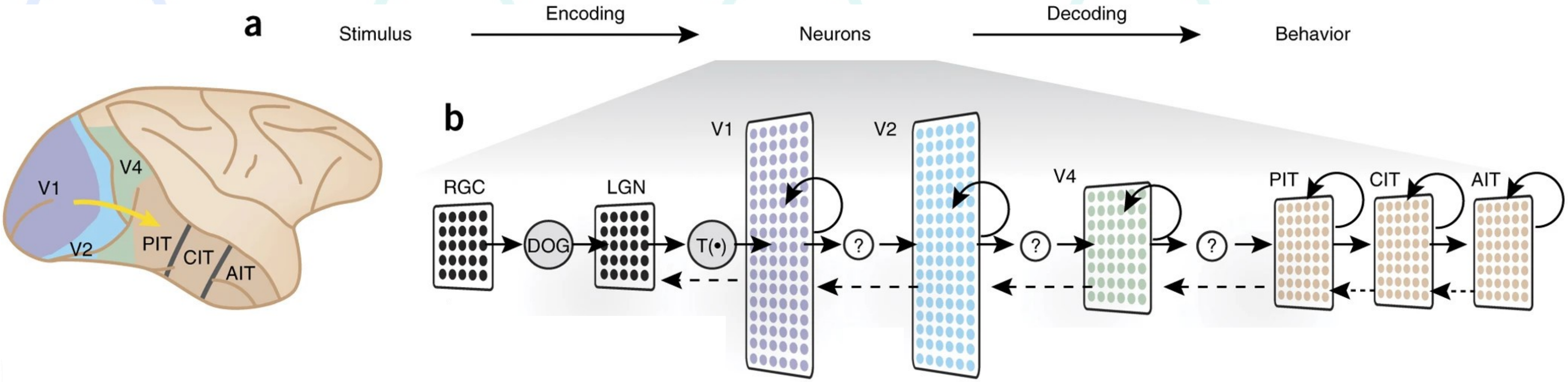
- It estimated that up to 1/3 of the human brain is devoted to parsing visual data.
- Object recognition is the ability to assign labels (e.g., nouns) to particular objects.
- Your eye feeds directly to specialist cells for colour/hue/saturation recognition as well as edges and patterns.
- There is clearly a deeply layered feed forward approach using many separate groups of neurons.
- Your cortex manages identity preserving transformations.
- Your brain stores “a copy” of the image.
- See DiCarlo et. Al, “How Does the Brain Solve Visual Object Recognition?”, 2012.

Convolutional neural networks

- We can train a neural network to analyse say 3x3 patches of an image and output a single pixel, thereby reducing the dimensionality.
- We shift our 3x3 lens by one pixel and go again. We repeat this for the entire image. Therefore for a 32x32 image we would end up with in a 30x30 image representation. (The image can be padded to remain the same dimensions.)
- What can we do now?
- Lets do it again with a new neural network.
- If we keep doing this we will end up with a final “image” of at most 3x3 pixels (without padding).
- The neural network we use only requires 3*n^2 neurons to analyse an image (where n is the side of the pixel box we are using), ie 27 for 3x3 analysis and 75 for 5x5. The times three if for the RGB.
- This is very easy to train.

Convolutional neural networks

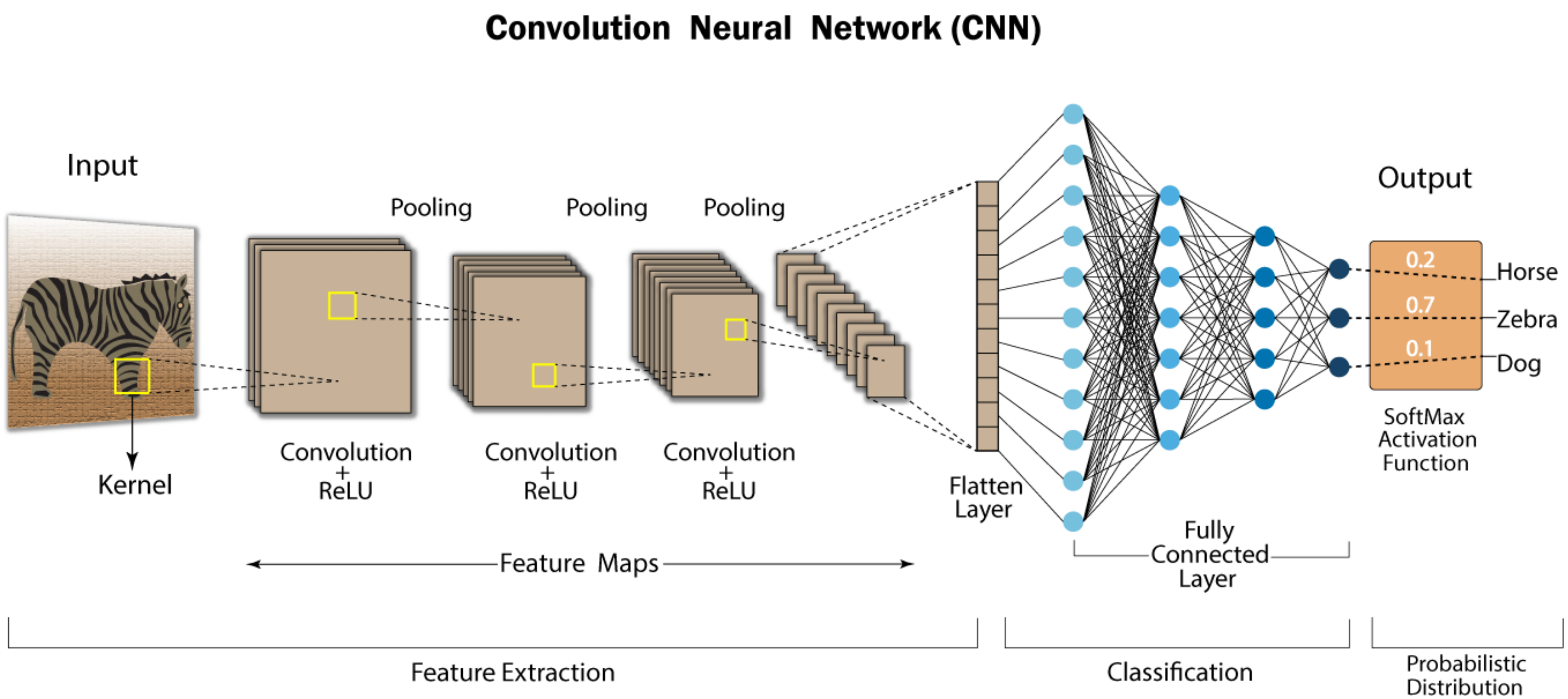
- How does the brain do it?



- The ventral visual pathway is the most comprehensively studied sensory cascade.
- It consists of a series of connected cortical brain areas (macaque brain shown). PIT, posterior inferior temporal cortex; CIT, central; AIT, anterior; RGC, retinal ganglion cell; LGN, lateral geniculate nucleus. DoG, difference of Gaussians model; T(•), transformation.
- Two foundational empirical observations about cortical sensory systems are that they consist of a series of anatomically distinguishable but connected areas and that the initial wave of neural activity during the first 100 ms after a stimulus change unfolds as a cascade along that series of areas.
- From Yamins et. al, 2016, ‘Using goal-driven deep learning models to understand sensory cortex’, Nature Neuroscience

Convolutional neural networks

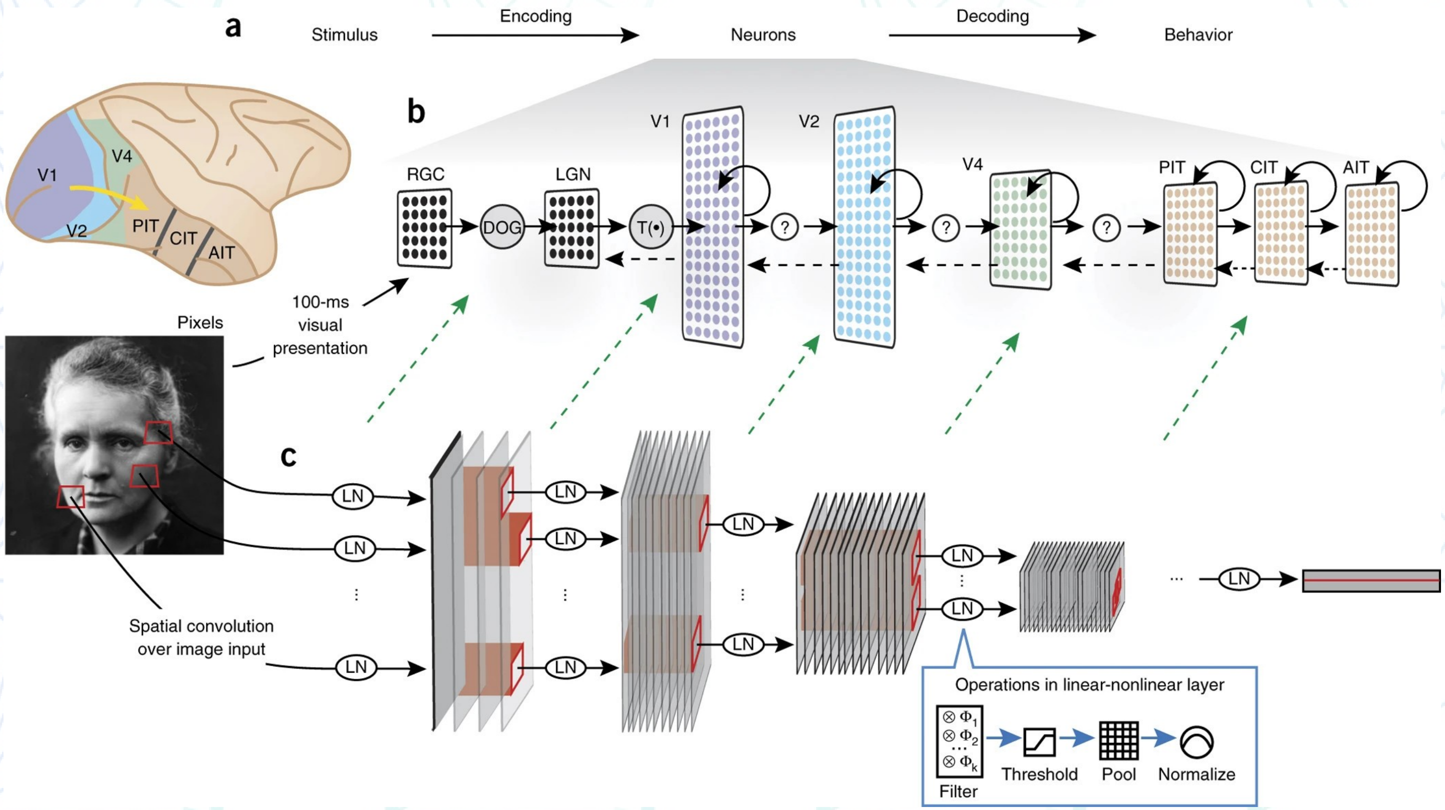
- How do computers do it?



- Successive layers of convolution activated by ReLU
- Pooling is the averaging (downsampling) of layers to desensitise the network to feature location
- Once the features are extracted a fully connected network classifies the image

Convolutional neural networks

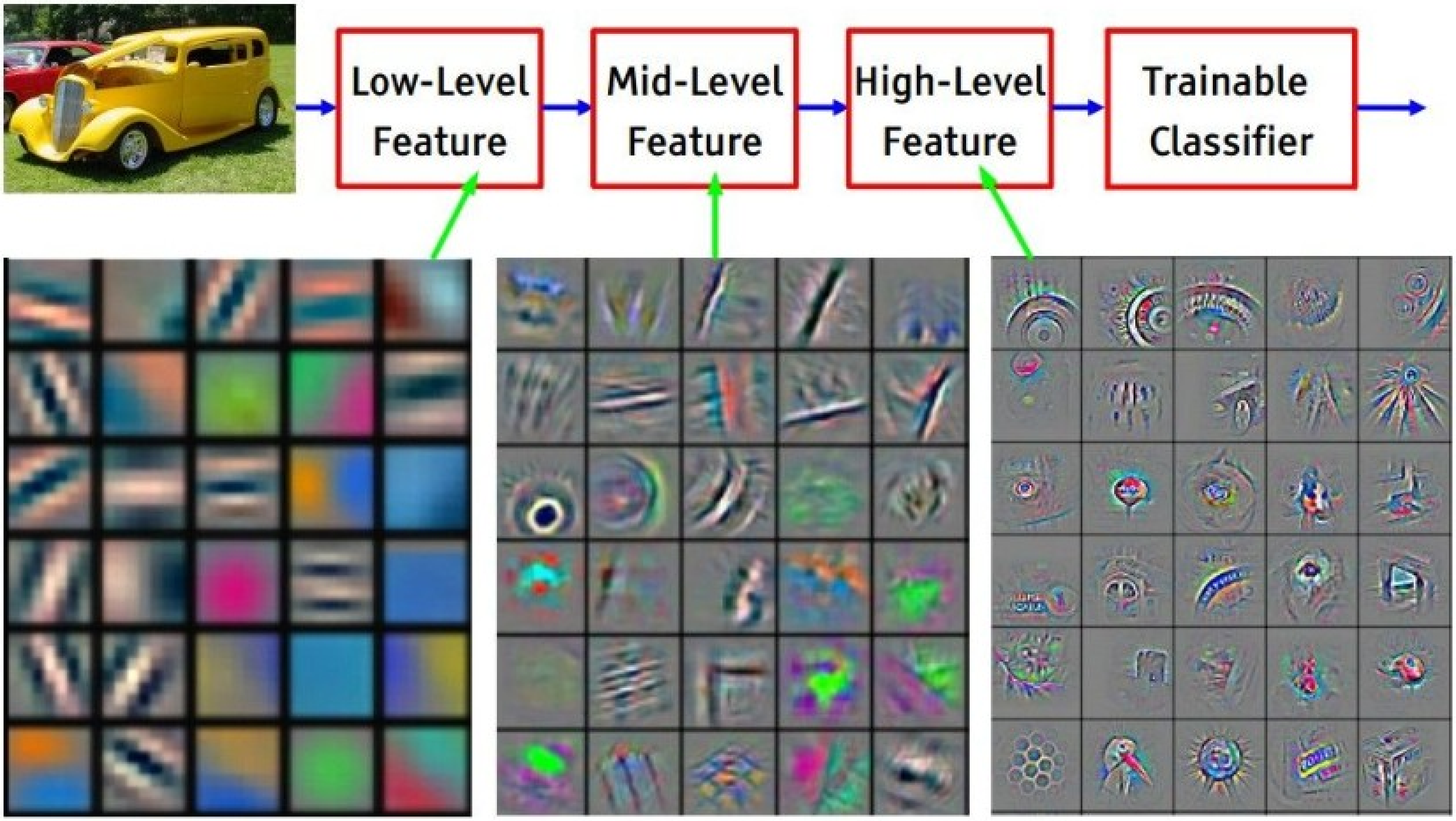
- We learn about our visual processing from designing neural networks to perform image recognition.



- We are inspired to design our neural networks using what we know about how our brains process images.

Convolutional neural networks

- If we visualise the outputs of the convolutional layers what do we find?



- We have indeed designed neural networks whose hidden layers extract recognisable features!