An Introduction to Convolutional Neural Networks

Keiron O'Shea 1 and Ryan Nash 2 1
Department of Computer Science, Aberystwyth University, Ceredigion, SY23 3DB
keo7@aber.ac.uk
2 School of Computing and Communications, Lancaster University, Lancashire, LA1 4YW
nashrd@live.lancs.ac.uk

Meghna Thakur

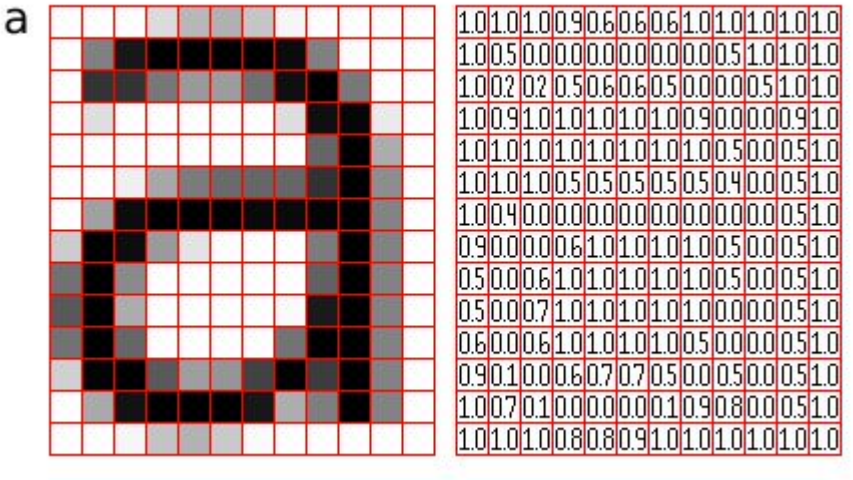
MOTIVATION

To develop a neural network with a grid-like topology, such as an image, while reducing the parameters required to set up the model.

IDEA

A digital image is a binary representation of visual data. It contains a series of pixels arranged in a grid-like fashion that contains pixel values to denote how bright and what color each pixel should be.

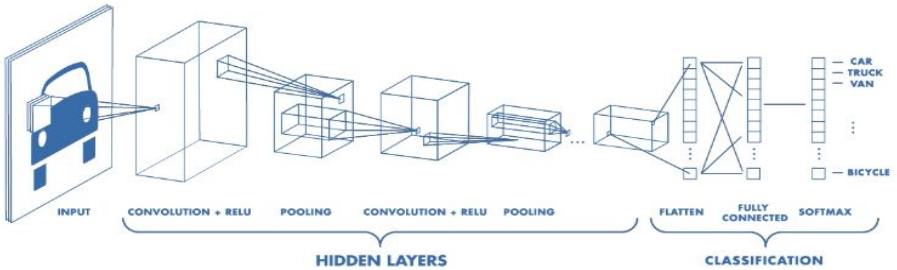
This data is feed into the CNN, after tutoring it with pre-labelled images to recognise similar images.



METHODOLOGY

CNNs are comprised of three types of layers:

1. Convolutional Layer
2. Pooling Layer
3. Fully-connected Layer

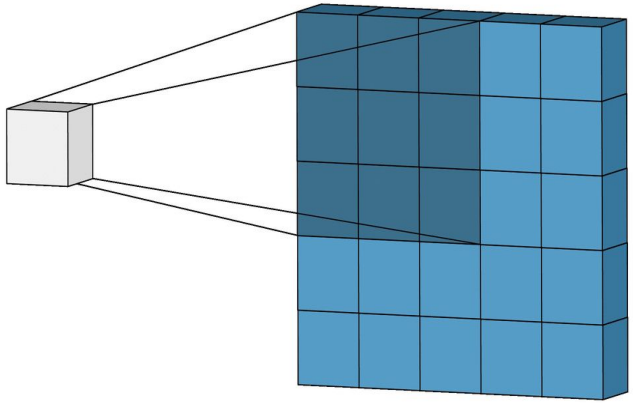


A simplified CNN architecture is formed when these layers are stacked together. The layers are arranged in such a way so that they detect simpler patterns first (lines, curves, etc.) and more complex patterns (faces, objects, etc.) further along.

METHODOLOGY

CONVOLUTIONAL LAYER

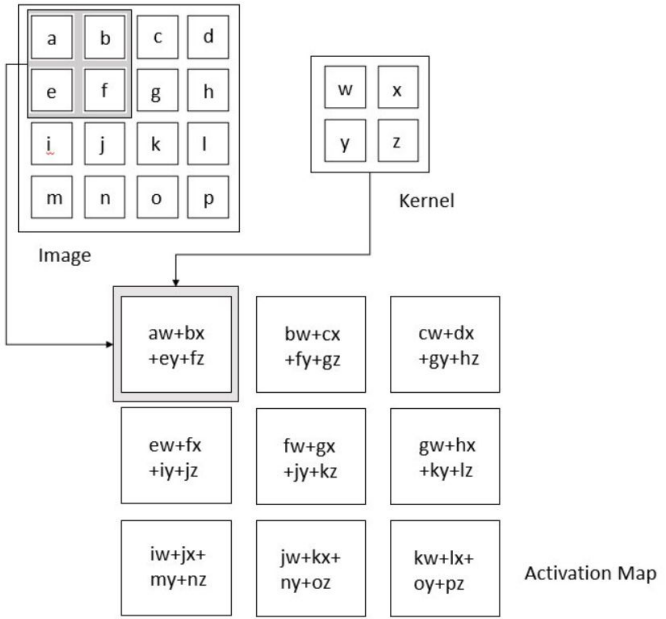
This layer performs a dot product between two matrices, where one matrix is the set of learnable parameters otherwise known as a kernel, and the other matrix is the restricted portion of the receptive field.



METHODOLOGY

The kernel is spatially smaller than an image but is more in-depth. This means that, if the image is composed of three (RGB) channels, the kernel height and width will be spatially small, but the depth extends up to all three channels.

Kernel slides across the height and width of the image-producing the image representation of that receptive region. This produces a two-dimensional representation of the image known as an activation map.

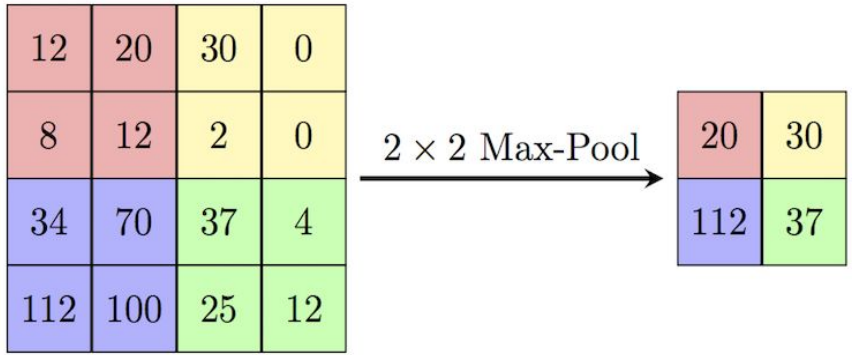


METHODOLOGY

POOLING LAYER

Pooling layers aim to gradually reduce the dimensionality of the representation, and thus further reduce the number of parameters and the computational complexity of the model.

The pooling layer operates over each activation map in the input, and scales its dimensionality using the “MAX” function. In most CNNs, these come in the form of max-pooling layers with kernels of a dimensionality of 2×2 applied with a stride of 2 along the dimensions of the input. This scales the activation map down to 25% of the original size - whilst maintaining the depth volume to its standard size.

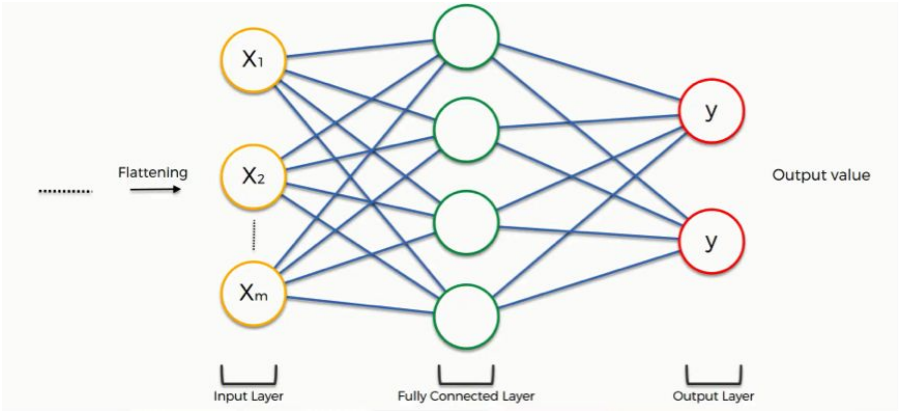


METHODOLOGY

FULLY-CONNECTED LAYERS

Neurons in this layer have full connectivity with all neurons in the preceding and succeeding layer.

The fully-connected layer helps to map the representation between the input and the output.



DISCUSSION POINTS

CNNs are extremely powerful machine learning algorithms, however they can be horrendously resource-heavy. What operation could be performed to reduce the size of the memory per image?

THANK YOU