Introduction to Computer Vision

Discussion questions

1. What are convolutional neural networks and how are they different from ANNs?
2. How do filters/kernels work for feature detection in CNNs?
3. How does pooling, padding and stride operations work in CNNs?

Convolutional Neural Networks

CNNs are a special type of neural network designed to work with image data. CNNs use convolutional layers - hidden layers which perform convolution operations. They have some different characteristics to ANNs:

1. Unlike ANNs, CNNs capture the spatial structure of the image
2. CNNs follow the concept of parameter sharing i.e. one filter is applied over the whole image, because of which they are much more computationally efficient.

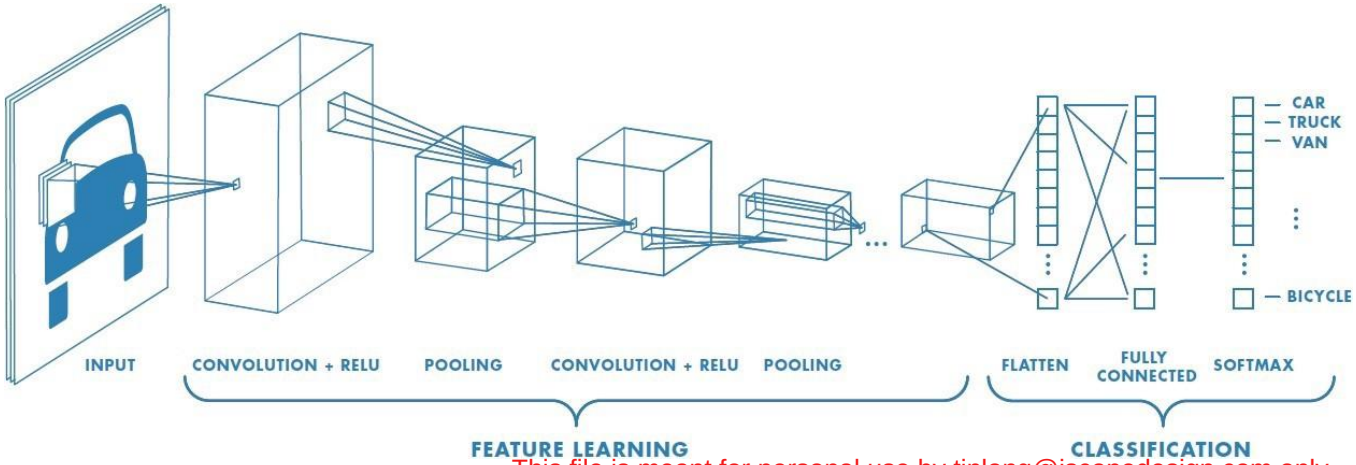


Image Source

The first part in this architecture is the convolutional layer followed by the pooling layer and the second part is the fully connected layer. This whole architecture is called a convolutional neural network.

The convolutional layer - Filter/Kernel

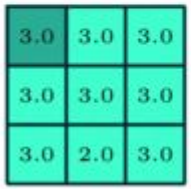
- A convolution operation uses a small array of numbers called a filter/kernel on the input image.
- Each filter is designed to identify a specific feature in the input space of the image, such as horizontal edges, vertical edges etc.
- A CNN is able to successfully capture the spatial and temporal dependencies in an image through the application of relevant filters.
- The role of the CNN is to reduce the images into a form which is easier to process, without losing features which are important for getting a good prediction.

- This image shows how the convolution operation works in a CNN
- It uses a 3x3 filter on a 5x5 image
- The resulted feature is a 3x3 image which is the convolved feature

Image Source

Pooling layer in CNNs

- After a convolution operation, we usually perform pooling to reduce the dimensions of the feature map
- It enables us to reduce the number of parameters, which both reduces the training time and the overfitting
- Pooling layers downsample each feature map independently, reducing the height and width, but keeping the depth same.



There are two types of pooling - Max and Average

- Max pooling just takes the maximum value whereas average pooling takes the average value in the pooling window
- Contrary to the convolution operation, pooling has no parameters.
- In these gifs, we can see how both max and average pooling work

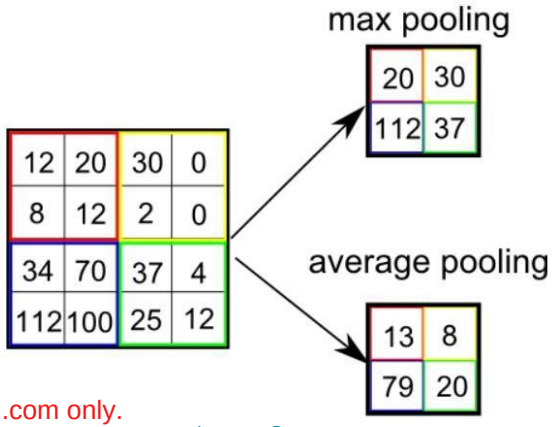
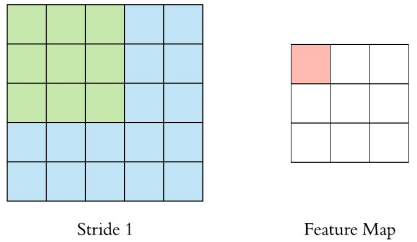
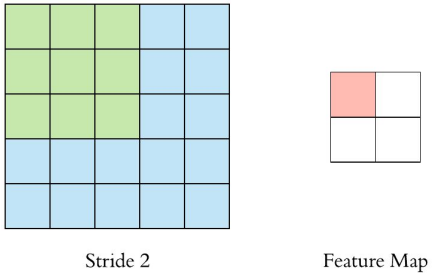


Image Source

Padding & Stride in CNNs

- **Stride** specifies how much we move the filter at each step. By default the value of the stride is 1 and is represented by the first figure
- We can also increase the value of stride if we want less overlap between the filters. It also makes the resulting feature map smaller since we are skipping over some locations.
- The second figure demonstrates the stride 2





We see that after using stride, the size of the feature map is smaller than the input. If we want to maintain the same dimensions, we can use **padding** to surround the input with zeros.

- The grey area around the input in the third figure is the padding.
- We either pad with zeros or the values on the edge, to match the dimensions of the feature map with the input.
- Padding is commonly used in CNNs to preserve the size of the feature maps, otherwise they would shrink at each layer, which is not desirable.

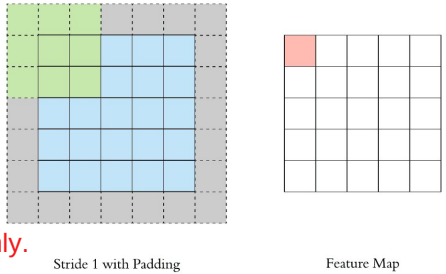


Image Source

Happy Learning !