Assignment 02 Solutions

1. Explain convolutional neural network, and how does it work ?

Ans. A Convolutional Neural Network, also known as CNN or ConvNet, is a class of neural networks that specializes in processing data that has a grid-like topology, such as an image. A digital image is a binary representation of visual data.

2. How does refactoring parts of your neural network definition favor you ?

Ans. The goal of refactoring is not to add new functionality or remove an existing one. The main goal of refactoring is to make code easier to maintain in future and to fight technical debt. We do refactor because we understand that getting design right in first time is hard and also you get the following benefits from refactoring:Code size is often reduced

Confusing code is restructured into simpler code

Both of the above benefits greatly improve maintainability which is required because requirements always keep changing.

3. What does it mean to flatten? Is it necessary to include it in the MNIST CNN? What is the reason for this ?

Ans. flatten function flattens the multi-dimensional input tensors into a single dimension, so you can model your input layer and build your neural network model, then pass those data into every single neuron of the model effectively. You can understand this easily with the fashion MNIST dataset.

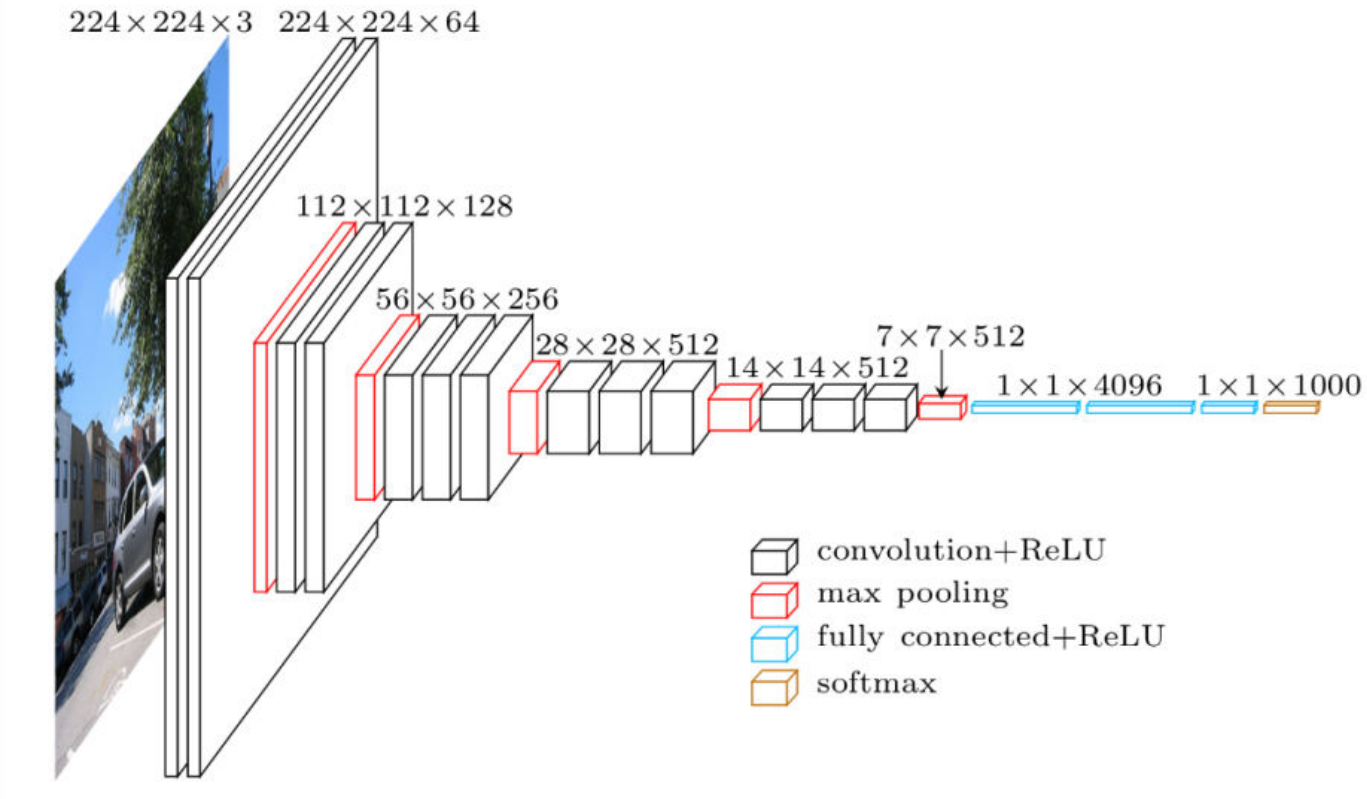
4. What exactly does NCHW stand for ?

Ans: NCHW stands for: **batch N, channels C, depth D, height H, width W**. It is a way to store multidimensional arrays / data frames / matrix into memory, which can be considered as a 1-D array.

5. Why are there 77(1168-16) multiplications in the MNIST CNN's third layer ?

Ans. Suppose you are working with MNIST dataset, you know each image in MNIST is 28 x 28 x 1(black & white image contains only 1 channel). Total number of neurons in input layer will 28 x 28 = 784, this can be manageable. What if the size of image is 1000 x 1000 which means you need 10^6 neurons in input layer. Oh! This seems a huge number of neurons are required for operation. It is computationally ineffective right. So here comes Convolutional Neural Network or

CNN. In simple word what CNN does is, it extract the feature of image and convert it into lower dimension without loosing its characteristics. In the following example you can see that initial the size of the image is 224 x 224 x 3. If you proceed without convolution then you need 224 x 224 x 3 = 100, 352 numbers of neurons in input layer but after applying convolution you input tensor dimension is reduced to 1 x 1 x 1000. It means you only need 1000 neurons in first layer of feedforward neural network.



6.Explain definition of receptive field ?

Ans. The receptive field encompasses the sensory receptors that feed into sensory neurons and thus includes specific receptors on a neuron as well as collectives of receptors that are capable of activating a neuron via synaptic connections.

7. What is the scale of an activation's receptive field after two stride-2 convolutions? What is the reason for this ?

Ans. in a neural network context, the receptive field is defined as the size of the region in the input that produces the feature. Basically, it is a measure of association of an output feature (of any layer) to the input region (patch).

8. What is the tensor representation of a color image ?

Ans: The representation of an image can take many forms. Most of the time, it refers to the way that the conveyed information, such as color, is coded digitally and how the image is stored, i.e., how is structured an image file. Several open or patented standards were proposed to create, manipulate, store and exchange digital images. They describe the format of image files, the algorithms of image encoding such as compression as well as the format of additional information often called metadata.

TENSORS IMAGES
Tensors can be understood as nested lists of objects of the previous order all with the same size. For example, an order three tensor can be thought of as a list of matrices all of which have the same number of rows and columns. These matrices are tensors of order two and since they have all the same number of rows and columns, the tensor of order three is actually like a cuboid of numbers and we can find numbers by going along any of the three-axis. Each number is identified by the row, the column, and the depth at which it's stored. We can formalize this idea in the concept of shape.

9. How does a color input interact with a convolution ?

Ans. The CNN only has the data to learn if color is a decisive factor for recognizing an object or not. If you only present it with red 'A's, it will learn that red is a decisive factor for recognizing the 'A'. By presenting it with a large number of different 'A's that are colored differently.