Neural Networks

MNIST Dataset

Siddik Ayyappa
Dr Chiranjeevi Yarra
11 July 2022

Accuracy Tables for MNIST dataset

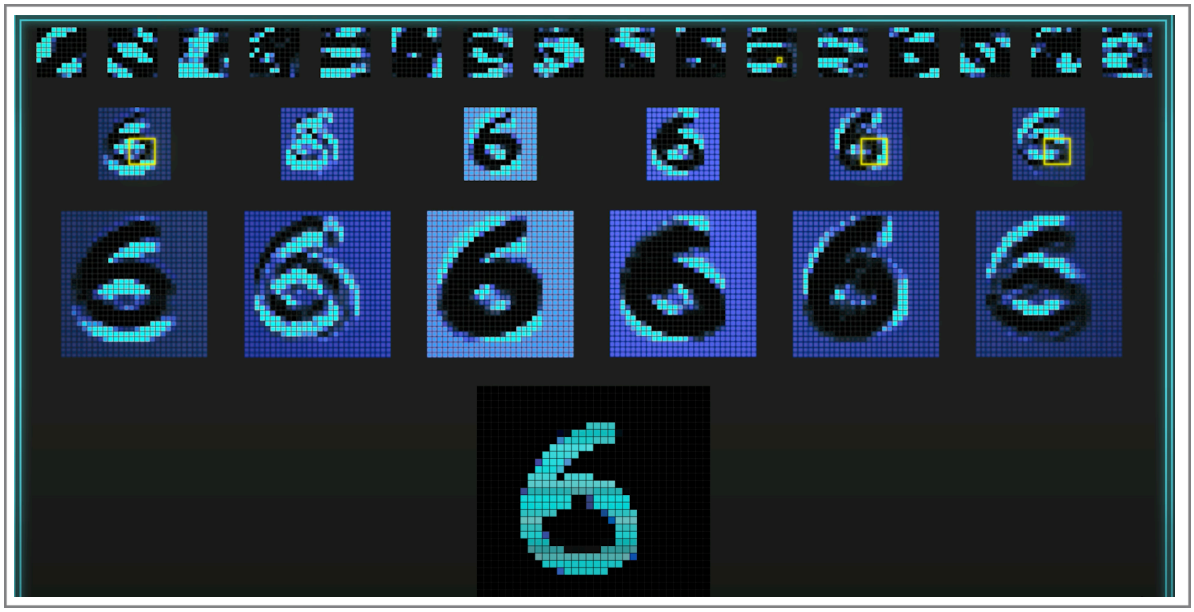
Neural network \ Set	Training Set	Test Set
DNN	95.77	94.58
CNN	98.23	98.21

DNN Attributes

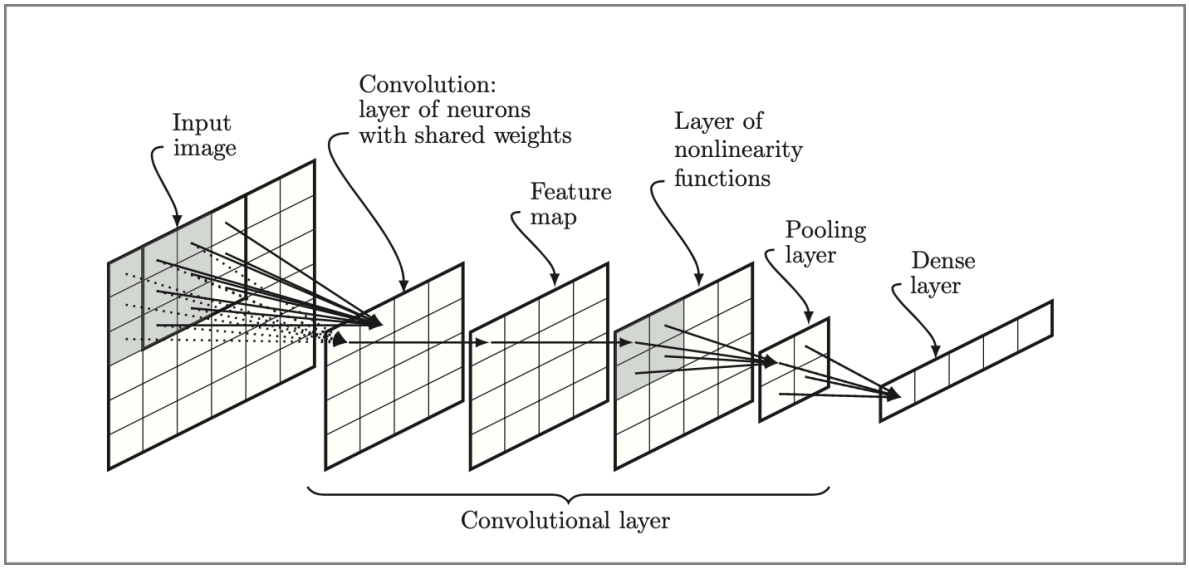
- Input Layer → 28 * 28 Inputs
- Hidden Layer(s)
  - Dense Layer → 200 Neurons; ReLU Activation;
- Output Layer → 10; SoftMax Activation
- Loss → Sparse Categorical Cross entropy

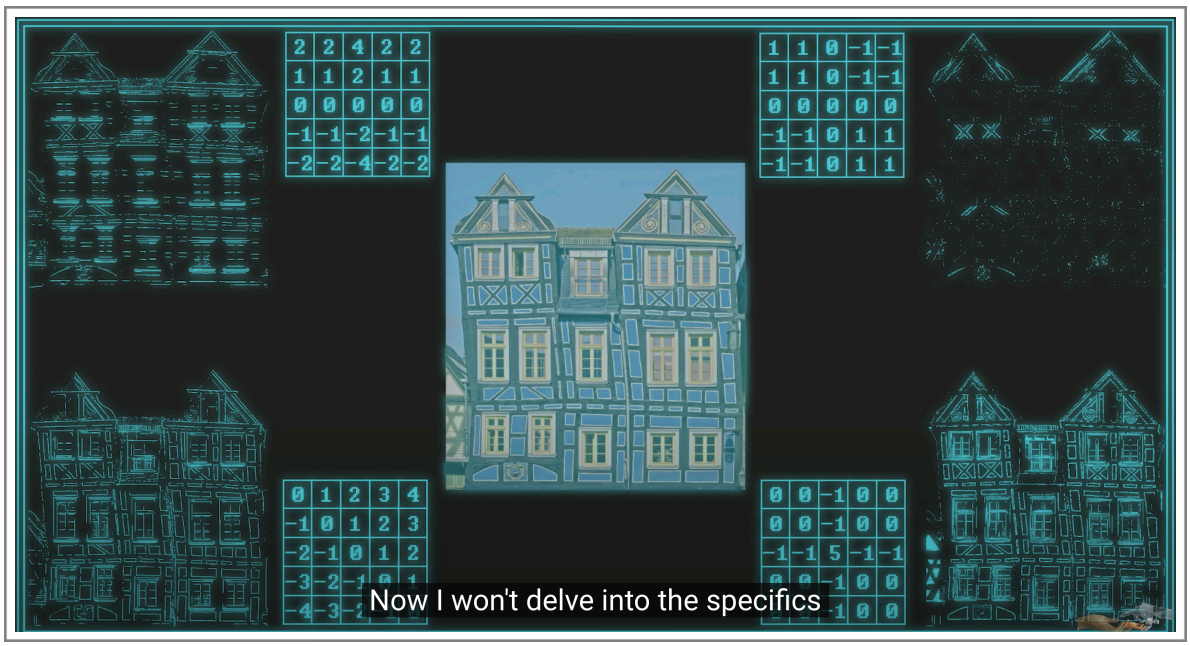
CNN Attributes

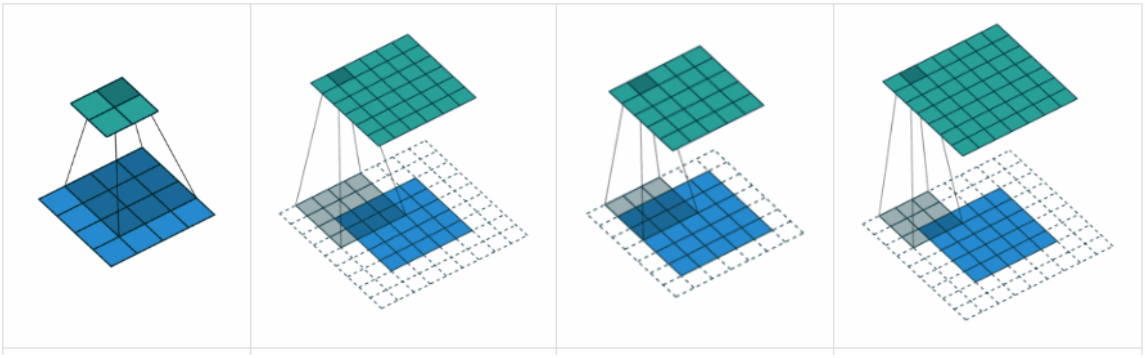
- Input Layer → 28*28
- Hidden Layers
  - Conv2D → 32; kernel_size=5; ReLU;
  - MaxPooling → Padding = “same”
  - Conv2D → 64; kernel_size = 5; ReLU;
  - MaxPooling → Padding = “same”
  - Flatten (To make the layers flat to fit into the shape of Dense layers)
  - Dense → 1024; ReLU;
  - Dropout Layer → Fraction → 0.2
- Output Layer → 10; Sigmoid



Input → 6 Filter → Pooling → 16 Filter







Different Kinds of Padding

Keywords and Points

1. Filters → Number of Kernels
2. Pooling → Used to downsample; Keep only the important Parts
3. Repeating the conv2d and maxpooling again would combine the features we have extracted in the previous times.
4. These Features are put into a feed forward NN, for classification.
5. In Normal Feed forward: 28*28 → 784 pixels as input
6. In CNN → 1024 pixels are the input for the feedforward.
7. Hence the better accuracy.
8. Sequence of Operation: For a general CNN

   Input → Convolving → Feature Map → Activation Function → Non Linearity → Pooling → Pooling Layer → Feed Forward Network (Classifier) → Output

Questions

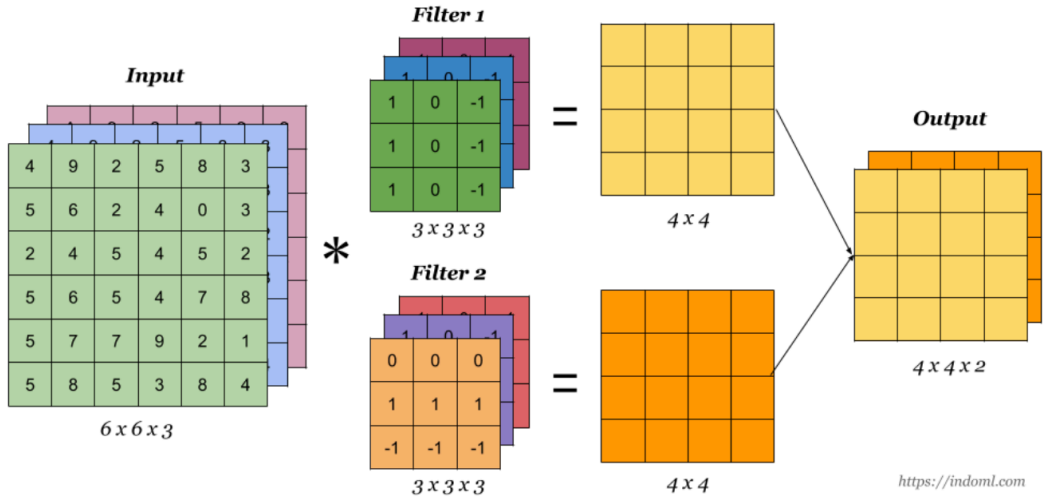
1. What is the difference between kernel and filter?

   Kernel is the matrix which is convolved over a single channel of the input tensor.

   Filter is the collection of all kernels, which are convolved on different channels of the input tensor.

2. What happens in the case where we have multiple channels of the input?

   The output is then flattened and then fed into a classical feed-forward



   network. In this case we have 32 inputs. (4*4*2).

3. What is a dropout function?

   In the layers which we think that overfitting might happen, we can place the dropout layer. What a dropout layer does is that it sets some weights to 0 randomly for each iteration. For the layers which we think overfitting might happen at, we have to place a dropout layer after it, with a smaller dropout value. *Higher the overfitting, lower the dropout value* (keep_prob). Used for very large data, especially in Computer vision, to avoid overfitting. The problem with this is that the cost function would not be well defined, so no debugging. If we needed to do debugging, we need to remove the dropouts then debug, and place the dropouts back again.

4. Do Conv1D and Conv3D exist?

   Yes, in both tensor flow and pytorch.

5. Explain about backpropogation and Cross entropy loss function. → Done.

6. Does each channel have a different kernel for a filter?

   Yes, it does. Think of the multiple channels as a single 3D image, then each filter, would just have a single 3D kernel, and hence makes sense.

7.