Project Report 4

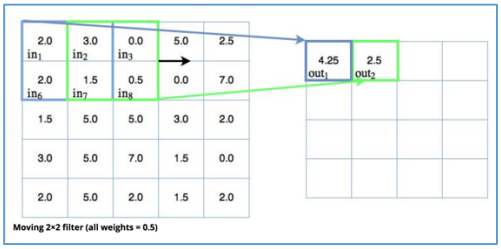
Introduction to Deep Learning

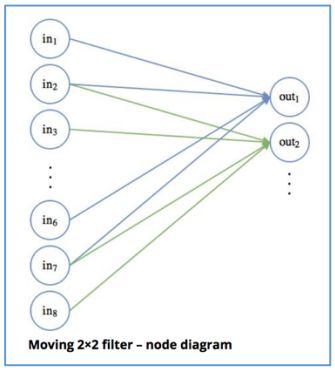
CSE-574 - Fall 2017

Deepika Chaudhary, Manpreet Dhanjal
(UB IT Name: d25, PN: 50248725), (UB IT Name: dhanjal, PN: 50248990)

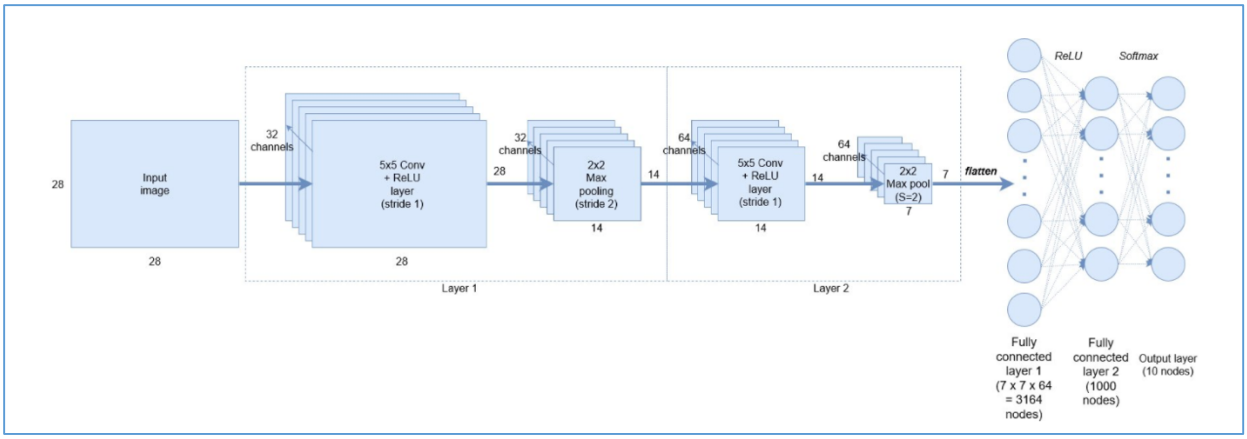
Introduction:

The assignment expects us to implement a convolution neural network to determine whether the person in a portrait image is wearing glasses or not. The idea that we implemented is of a moving filter which passes through the image in the first layer of CNN using the overlapping areas. In a convolution operation, this 5 x 5 moving filter would shuffle across each possible *x* and *y* co-ordinate combination to populate the output nodes. This operation can also be illustrated using our standard neural network node diagrams:





This reduces the number of parameters required in the network thus we will be able to converge more efficiently towards the desired feature. Here we have to determine whether the celebrity is wearing Eyeglasses or not. So this indicates that we have to classify the images into 1 subclass only. In order to proceed with the classification, we have applied 2 layers of Convolutional Neural Network and then 1 fully connected layer. Ideally the network should follow the diagram below. At each stage the image is passed through the activation function whose result is further given to the next stage and so on. At every state the image is down sampled and is divided into channels present in that layer. We are using a stride of [1,1] so that we can get the result of overlapping features. Next at every stage we perform pooling. Pooling is used in convolutional neural networks to make the detection of certain features in the input invariant to scale and orientation changes. They mainly generalize over lower level and more complex information.



Analysis and Results:

1. Loading the Celeb Dataset:

We have selected a dataset of 60K images for our experiments. The images are of high resolution so they are taking time to get loaded. The dataset of around 60k images took around 10 hours however we were not able to load the whole dataset of 2lakh images due to CPU constraints.
The 60k images dataset is further divided into 40k training, 10k validation and 10k testing data. We resized the image to the resolution of 40 x 40 and the results were obtained. We evaluated the error rate given by formula below on the validation set.

$$E = \frac{N_{\text{wrong}}}{N_{\text{V}}},$$

At image resolution of 40x40 the accuracy is:
Training Set Error: 76.4%
Validation Set Error: 75.2%
Testing Set Error: 73.86%

2. Tuning hyper Parameters:

1. **Learning Parameter:** We obtained good results by setting the learning parameter to 10^(-9). To obtain this value we ran a grid search on a bunch of learning rate and provided the input to Adam Optimizer. We were able to get the best training accuracy of around 76.4% with the learning rate of 10^(-9).

2. **Drop Out Rate:** A dropout layer is a layer of regularization which zeroes out certain activations of the layers. This helps us in generalization of our data thus preventing the chances of over fitting. We tried the grid search on different values of keep probability and we were able to obtain best results at the value of 0.7.

3. **Number of layers and number of nodes:** We have used the filter size of 5x5 for 32 channels in the first layer and the filter size of around 64 channels for the second layer. Further increasing the no of layers will increase the accuracy obtained. Each of these filters can be thought of as feature identifiers. As we go through the network and go through more convolution layers, we get activation maps that represent more and more complex features. By the end of the network, we are able to extract some high level features like eyeglasses in our case. By adding one more convolution layer, the accuracy of our classifier increased by 2% approximately.

3. **Higher Resolution Images:** High resolution images increases the accuracy as it increases the chances of feature detection more clearly. We tried the images of size 52 x 52 by downsizing the original image and better results were obtained as compared to the case with 40 x 40 resolution.

4. Increasing the size of training set: We can increase the size of training set either by adding more images or by augmentation of existing images. Both of these, are expected to increase the accuracy of the classifier.

References:

- https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/
- https://www.tensorflow.org/tutorials/
- https://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Liu_Deep_Learning_Face_ICCV_2015_paper.pdf