Image Analysis and Computer Vision
Spring 2021
(CS 898BA)

# Final Exam Report

Submitted To
Dr. Ajita Rattani
(Course Instructor)

Submitted By
Suresh Pokharel
WSUID: X254W356
Department of Electrical Engineering and Computer Science
Wichita State University

27th April, 2021

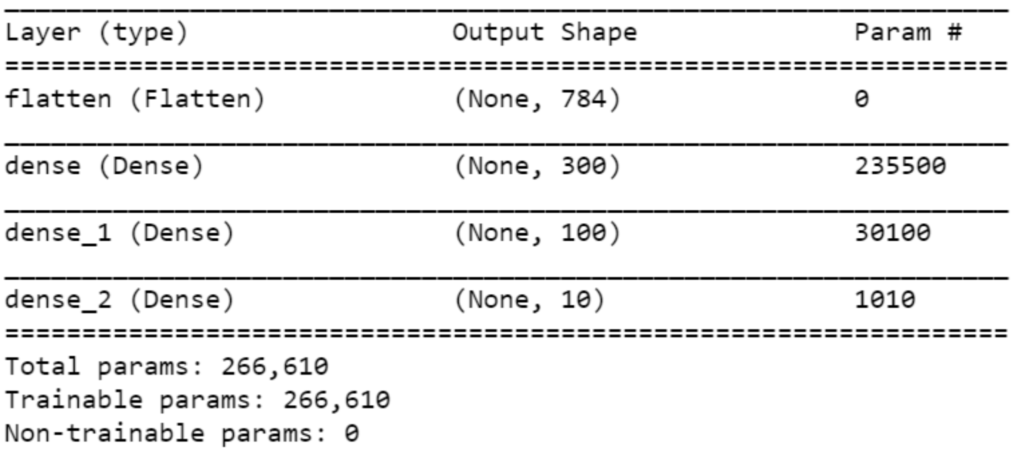
**Problem:1**

Modify the neural network model in code (CNN_basic.ipynb) provided in the package to include a stack of convolutional layers. Compare the performance of the existing neural network model in the code with your modified Convolutional Neural Network (CNN) version. Justify the observation noted.

*Solution:*

**Description of given model:**

The model summary of the given model is given in Figure 1.1. The given model is a sequential model consisting of a flatten layer and 3 dense layers.



```
_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
flatten (Flatten)            (None, 784)               0         
_________________________________________________________________
dense (Dense)                (None, 300)               235500    
_________________________________________________________________
dense_1 (Dense)              (None, 100)               30100     
_________________________________________________________________
dense_2 (Dense)              (None, 10)                1010      
=================================================================
Total params: 266,610
Trainable params: 266,610
Non-trainable params: 0
```

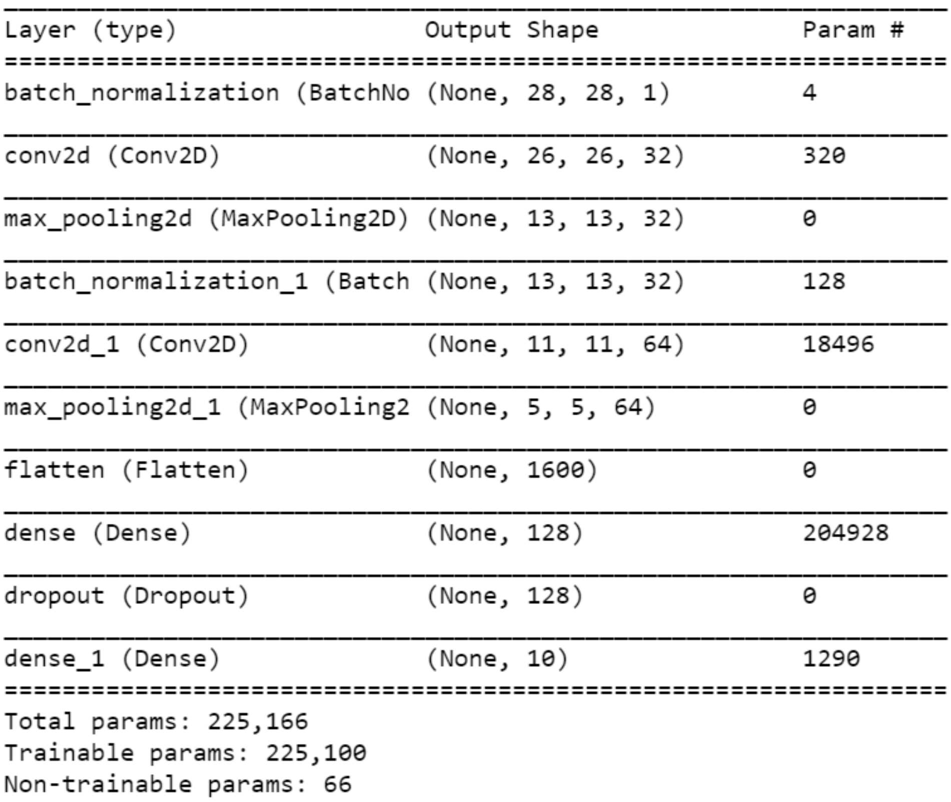
Figure 1.1: Model architecture of given model

The flatten layer basically converts a n-dimensional array into a 1D array to create a single long feature vector.. For example, for an input of shape (c , h , w), flatten converts to vector output of shape (c* h* w). In the next, The dense layer is a fully connected layer where all the neurons in a layer are connected to the neurons of the next layer.

The result after training the model upto 30 epochs on the fashion mnist dataset and obtained 85% accuracy on independent test sets.

**Description of modified model**

The summary of the modified model is shown in figure 1.2. The input is passed through a Batch Normalization layer. A stack of two convolution layers followed by max pooling of (2,2) pixels is used in the modified model. The kernel size of each convolution filter is taken (3,3). The number of filters used in the first and second convolution layer are 32 and 64 respectively. After stacking convolution layers, the result is flatten into a 1-dimensional array. Then, Dense layer of 128 neurons is followed by a dropout of 10% for regularization purposes. In the final output layer, 10 output nodes are used to give the probability of 10 different classes ranging from 0-9. The predicted class will be the maximum probability found among the 10 nodes.



```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
batch_normalization (BatchNo (None, 28, 28, 1)         4
_________________________________________________________________
conv2d (Conv2D)              (None, 26, 26, 32)        320
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 13, 13, 32)        0
_________________________________________________________________
batch_normalization_1 (Batch (None, 13, 13, 32)        128
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 11, 11, 64)        18496
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 5, 5, 64)          0
_________________________________________________________________
flatten (Flatten)            (None, 1600)              0
_________________________________________________________________
dense (Dense)                (None, 128)               204928
_________________________________________________________________
dropout (Dropout)            (None, 128)               0
_________________________________________________________________
dense_1 (Dense)              (None, 10)                1290
=================================================================
Total params: 225,166
Trainable params: 225,100
Non-trainable params: 66
```

Figure 1.2: Modified model architecture after adding convolution layers

Early stopping technique is used to prevent the model from overfitting. Also, it helps to reduce the training time by evaluating the performance. In this work, loss metric was monitored by using keras callback to stop the training where the loss observed is minimum. Other details on the modified model are as follows:

**Loss function**: sparse_categorical_crossentropy

**Optimizer**: adam

**Steps involved to train the model**

After developing the model, following steps are carried out to train and evaluate the model:

a) Import necessary libraries
b) Load Fashion-Mnist data set and split in train(60,000 images) and test(10,000 images) sets.
c) Split 10,000 train data for validation data.
d) Divide all pixels by 255 to normalize them in between [0,1].
e) Reshape image (28,28) to (28,28,1) to make it fit for Conv2D (2D convolution) operation.
f) Train the model(shown in figure 1.2) on train and validation data
g) Evaluate the model using independent test data
h) Compare with the result of model-1 given in the question.

## Comparison of Given and Modified Model

The comparison between the results obtained from the given model and the model after stacking convolution layers (Modified Model) are presented in the table below.

| **Metric** | **Given Model** | **Modified Model** |
|---|---|---|
| No. of Epochs | 30 | 8 |
| Training Loss | 0.2235 | 0.1432 |
| Training Accuracy | 0.9194 | 0.9450 |
| Validation Loss | 0.2995 | 0.2697 |
| Validation Accuracy | 0.8878 | 0.9122 |
| **Test Accuracy** | **0.8517** | **0.9068** |

## Conclusion and Discussion

It can be clearly observed that the modified model has outperformed the results of the given model. The use of early stopping has drastically reduced (30 to 8) the number of epochs. The use of the dropout layer has prevented the model from overfitting. The overall test accuracy tested on the independent test set is improved by more than 5% after adding the stack of two convolution layers. More convolution layers, regularization, data augmentation etc. concepts could be added to the experiment to observe if the performance can be improved further.

**Problem:2**

Summarize the paper titled “Classification of Fashion Article Images using “Convolutional Neural Networks” (https://ieeexplore.ieee.org/document/8313740) in one or two pages. Point out the limitations of this paper if applicable.

*Solution:*

The article “Classification of Fashion Article Images using Convolutional Neural Networks” has proposed three different architectures for image classification. The authors have used the Fashion-MNIST dataset as a benchmark dataset to compare the performance of their model with existing state of art network architectures. The Fashion-MNIST dataset consists of 10 classes of fashion clothes with 60,000 training and 10,000 testing image samples. As compared to similar research researches, this paper has additionally employed batch normalization and residual skip connections on their different convolutional neural network architectures. Batch normalization is basically a technique used in deep learning works that makes the training process faster and more stable by re-scaling the input features to the next layer. On other parts, the authors have employed residual skip connections to avoid the vanishing gradient problem,and to mitigate the accuracy saturation problem. In this paper, the authors have used skip connections by adding the previous input and current value of convoluted output to get the final output.

The first model is a most common CNN architecture that consists of two 2D convolution layers followed by a MaxPooling layer. Then the network is followed by Flatten, Dense, Dropout of 50% and Dense layers.  The convolutional layer was composed by using 32 filters of window size 3x3, and max pooling of 2x2 pixels were used. This model has gained accuracy of 91.17% on the independent test set.

Similarly, the second model, the authors have employed Batch Normalization to the input layer before feeding to the convolution layer. Then each convolution layer is followed by a Max Pooling layer of 2x2 pixels. After max pooling layer, Flatten, Dense, Dropout of 50% and Dense layers are used like in the first model. This model has achieved 92.22% of test accuracy which is more than 1% improvement  after addition of batch normalization.

In the third model architecture, they have proposed the addition of residual skip connections to the second model.  They have named the network “CNN2 + BatchNorm + Skip model” as they have two convolution layers. This model has shown performance improvement over the first and second models. The accuracy of 92.54% was observed on the independent test set by introducing skip-residual connections with convolution layers.

The authors have compared their results with the results from the research that has used Support Vector Machines(SVM) on the same data. The comparison has shown remarkable improvement (more than 3%) by the proposed model with residual skip connections. The authors have also used other performance metrics other than Accuracy like F1 Score, Precision, and Recall. The proposed model has obtained better results with respect to all the performance evaluation metrics.

The potential shortcoming of this paper is that the authors have not used image augmentation techniques while training the image. The use of simple image augmentation techniques like geometric transformation(rotation, translation, scaling etc.) increases the generalization capacity of the model and that may help to improve the classification accuracy in this project. In addition, the use of regularization techniques like L1,L2 regularization, more number of convolution layers may help to improve the performance.

***