Transfer Learning

Rishabh Samra

1 Introduction

It is the process in which parameters from one of the trained model is taken, freezed and used while training another dataset with more complex network. Here knowledge gained while solving one problem is stored and reused while solving another problem

2 Transfer Learning On Characters' Dataset

The process of transfer learning is illustrated while training characters' dataset comprising 10 different characters. Initially a Convolutional Neural Network is trained on dataset comprising of images of 3 classes (single line at different inclinations) and the parameters of the network are saved for reusing while training another network for classifying dataset of 10 classes(different alphabets)

3 Details of the Dataset

3.1 Line Dataset

It comprises of 32x32 grayscale images with 3 classes.

Training dataset comprises of 40000 images and validation dataset is having 10000 images distributed equally among the 3 classes.

Following are some sample images from the dataset:

Figure 1: Lines at different inclination

3.2 Characters' Dataset

It comprises of 32x32 grayscale images of 10 alphabets thus having 10 classes which was created after augmenting images with operations like rotation, adding noise etc.

Training Dataset comprises of 60000 images and validation dataset has 10000 images in equal proportion among 10 classes.

The dataset with 7 class classifier has 42000 training images and 7000 validation images in equal proportion among the classes. Following are some sample images from the dataset:

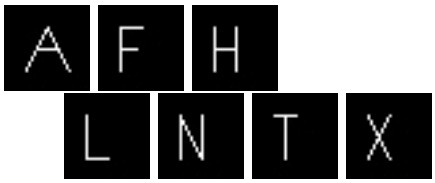


Figure 2: Characters' Dataset

4 Network and Training Details:

For training dataset of lines, 3 class classifier was trained with the 2 set of convolutional and maxpool(2x2) layers and then 2 convolutional layers without maxpool layers. First set of convolutional layer consists of 3x3 kernel with 10 filters and second set consists of 3x3 kernel with 20 filters and other two convolutional layers have 30 and 40 filters respectively. The output of the final convolutional layer is linearized and this layer consists of 160 neurons. Another fully connected layer of 84 neurons was used and finally output layer has 3 neurons to classify the four classes.

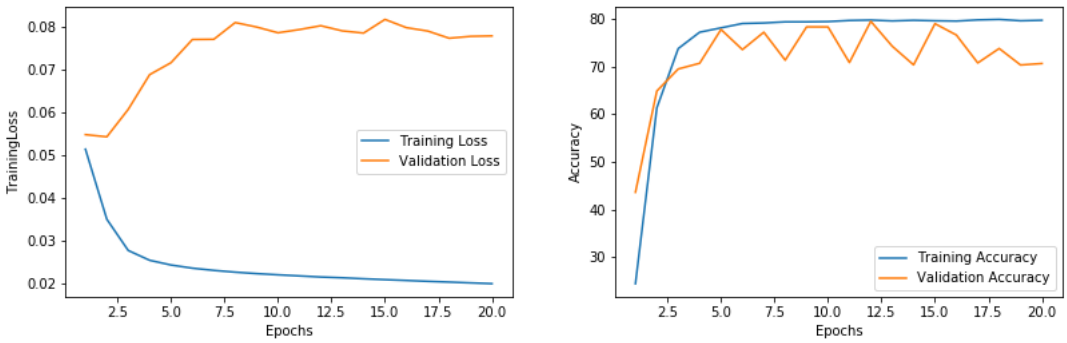
Parameters of the trained model was saved for reuse while training another network.

For characters' dataset, a 7 class classifier which is a CNN is trained. The dimensions of the convolutional and maxpool layers are same as mentioned above. Then after linearizing output of final convolutional and maxpool layer to 500, fully connected layers of 160, 84, 42 neurons were added and finally output layer consists of 7 neurons for classifying 7 classes. Dropout was implemented between convolutional layers and also l2 regularization was introduced with regularization constant value equal to 0.05. For training the network, we used learning rate of 0.0001 with Stochastic Gradient Descent Optimization with momentum constant = 0.9. Training was done for 20 epochs with batch size of 32

Then same dataset was used to train the same network but this time the parameters of initial convolutional layer was copied from the model trained on line dataset and freezed and parameters of other layers were trained and we plot training loss against number of epochs.

5 Results:

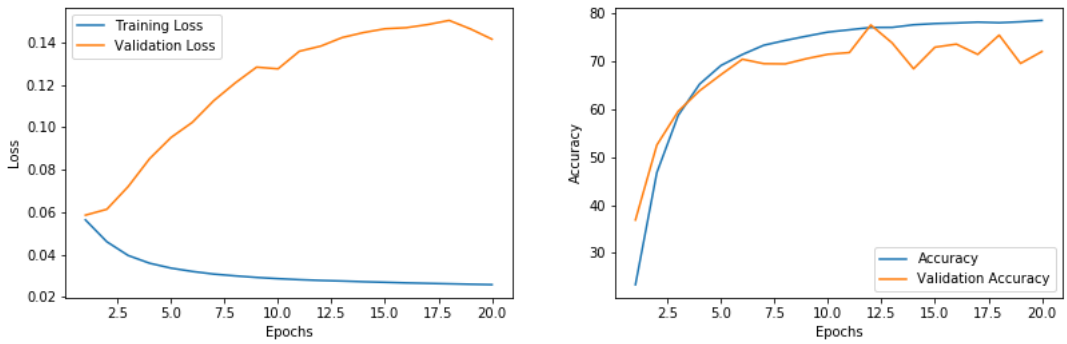
Following are the curves of Loss and Accuracy versus number of epochs for Training Characters' Dataset without Transfer Learning and with Transfer Learning respectively for 10 class classifier(A, E, F, H, I, L, M, N, T, X) and corresponding confusion matrix:



(a) Loss versus Epochs

(b) Accuracy with Epochs

Figure 3: Without Transfer Learning



(a) Loss versus Epochs

(b) Accuracy with Epochs

Figure 4: With Transfer Learning

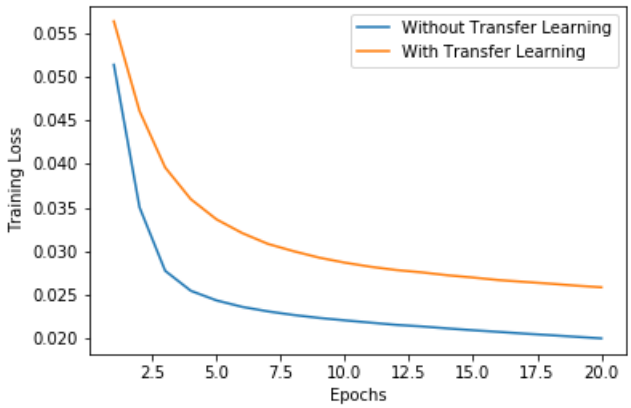


Figure 5: Training Loss Comparision

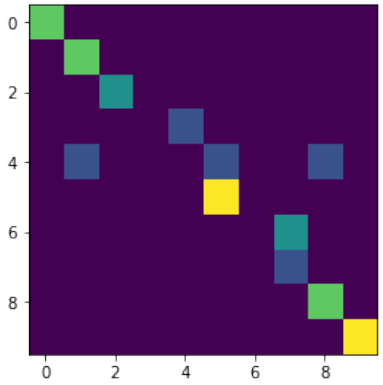
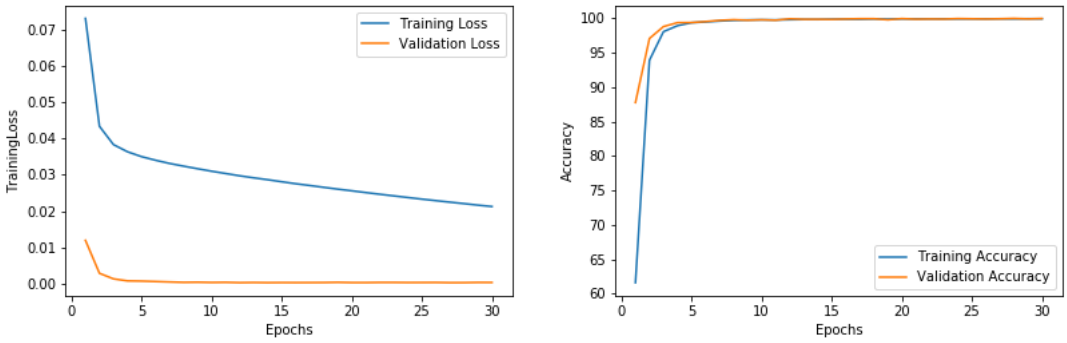


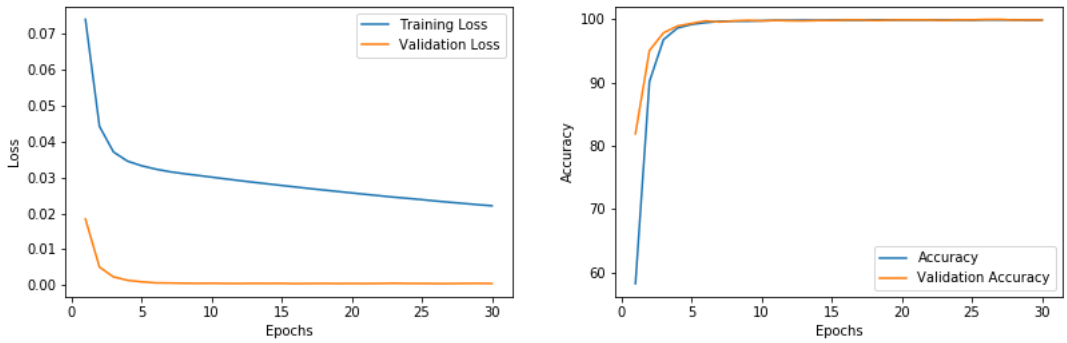
Figure 6: Confusion Matrix

Following are the curves of Loss and Accuracy versus number of epochs for Training Characters' Dataset without Transfer Learning and with Transfer Learning respectively for 7 class classifier(A, F, H, L, N, T, X) and corresponding confusion matrix:



(a) Loss versus Epochs (b) Accuracy with Epochs

Figure 7: Without Transfer Learning



(a) Loss versus Epochs (b) Accuracy with Epochs

Figure 8: With Transfer Learning

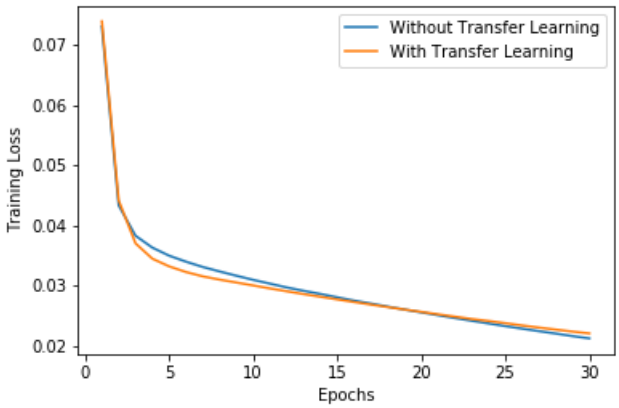


Figure 9: Training Loss Comparision

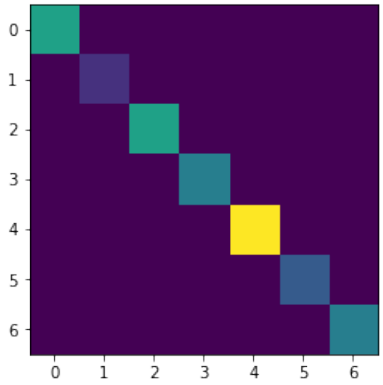


Figure 10: Confusion Matrix

6 Conclusion:

From the confusion matrix for 10 class classifier we can infer that in some cases H is predicted as I and I sometimes predicted as E, T, L and M predicted as N which decreases the accuracy so by removing 3 classes(E,I,M), we repeat the experiment and trained 7 class classifier which increases the accuracy significantly. Also while comparing the training loss for 7 class classifier, we can see that initially rate of decrease of training loss with respect to time is more while doing transfer learning.