Aadharsh Rajagopal

21/06/2024

This documentation will cover the result of the confusion matrices generated from the FashionMNIST Dataset before and after doing some modifications to the training data.

* The aim of this is to improve the quality of the training dataset which is being used for model training
* Methodology

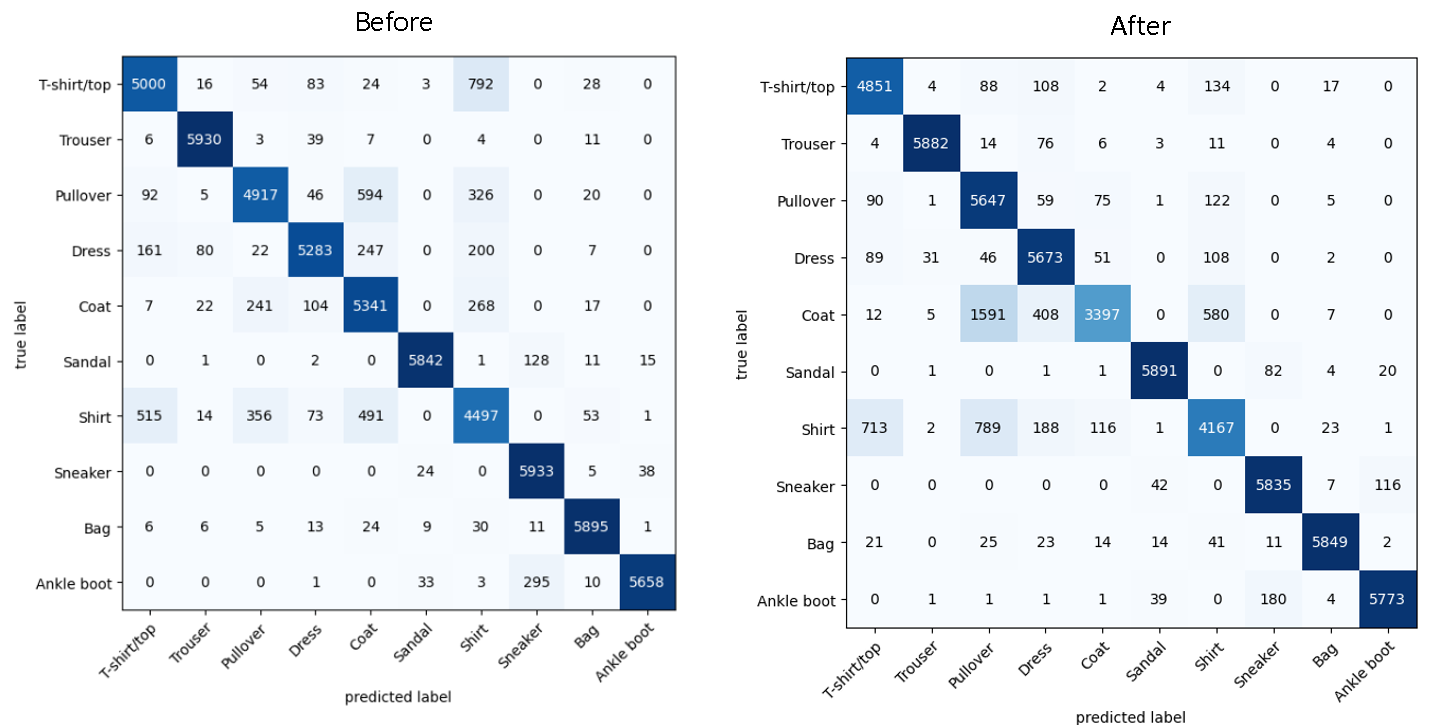
1. A baseline model is trained on all the training data until a desirable accuracy is achieved
2. A confusion matrix is constructed and inspected to see any discrepancies/inaccuracies/misclassifications can be identified
3. These misclassifications are handled in different ways (removed, swapped labels)
4. A new untrained instance of the model is trained again with this new dataset and a confusion matrix is constructed to see the difference between initial training dataset and training dataset after modifications

Code

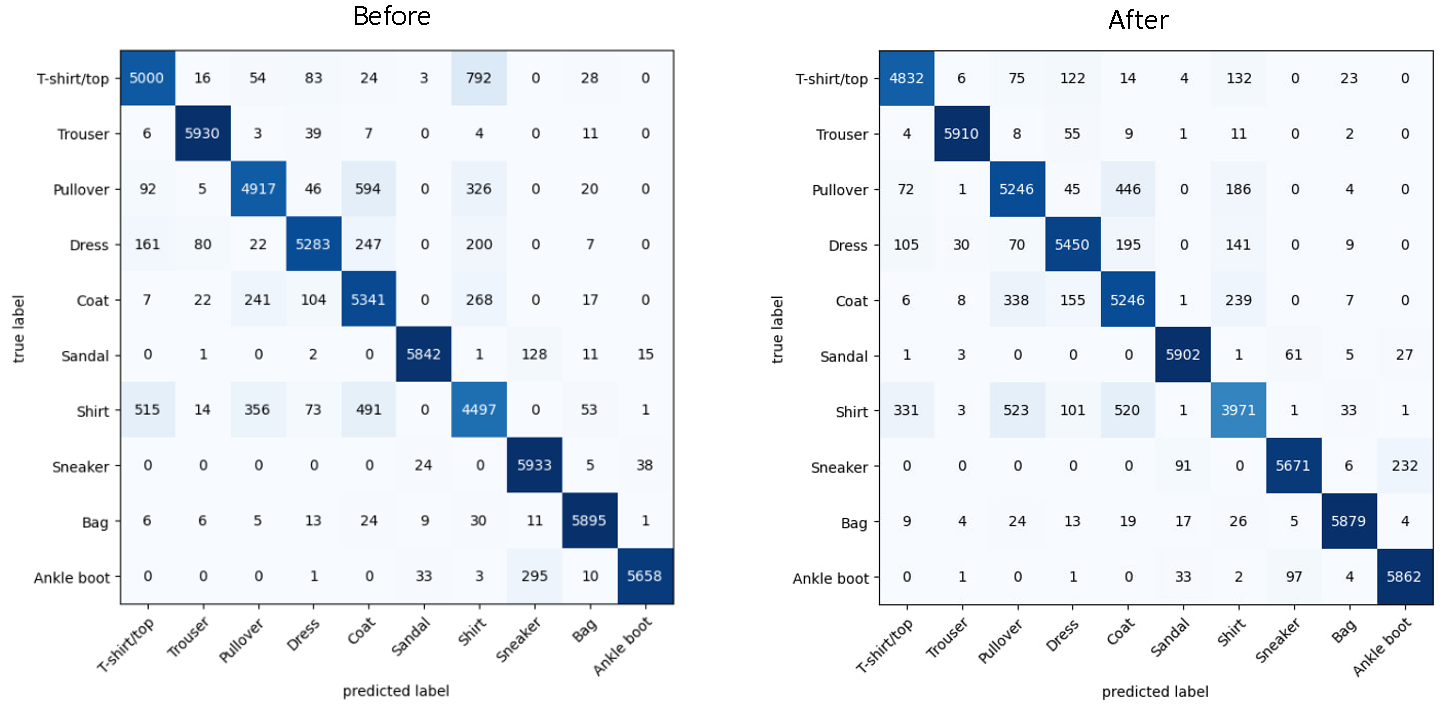
* The code used for these experiments mostly remains the same. For the swapping the labels, it is done by modifying the code in new folderdataset class to load in the new training dataset (FolderDataset2), and for removing data, we can just remove the folder that we moved the misclassified images into using shutil.rmtee() before loading the data again.
* The code involves an external file to be present, ‘model.pth’ which are the weights for the first model after I have trained it with the original dataset. This increases the reproducibility of the experiments, at least for the initial confusion matrix.
  + The file can be found in the folder in the github repository along with the code
* <https://github.com/Aadharsh1/ML-Deep-Learning/blob/main/FashionMNIST_CONFUSION_MATRIX/Fashion_mnist_19_6.ipynb>

**Experiments**

1. Removing the images which were labelled as T-shirts but misclassified as shirts.

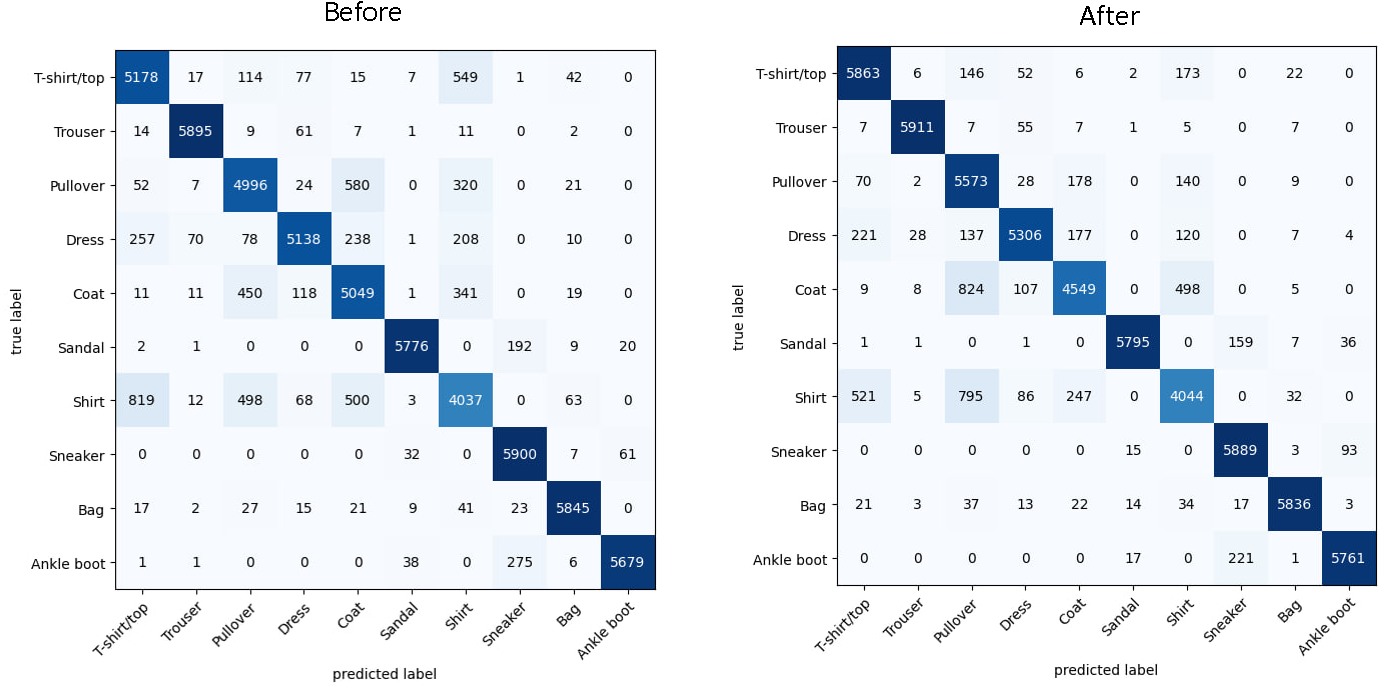


1. Removing both the folders where T-shirts were misclassified as shirts and shirts were misclassified asT-shirts.



1. Swapping the labels of the misclassified tshirts as shirts to shirts and the misclassified shirts as T-shirts to T-shirts and retraining the model.

* The labels of 549 T-shirts were reassigned to shirts
* The labels of 819 shirts were reassigned to T-shirts



1. Only reassigning the labels of one group of misclassified images

* The labels of 885 shirts which were misclassified as T-shirts were reassigned to T-shirts.