



Image Scraping and Classification Project

Submitted by:

Manish

Kumar

ACKNOWLEDGMENT

I would like to thank Keshav Bansal Sir, Sapna Verma ma'am who helped me in this project by clearing my doubts in doubt clearing sessions and providing me with learning material that proved to be helpful for my project. Also, DataTrained live sessions and youtube channels like Krish Naik and CodeBasics helped me understand concepts that were alien to me.

INTRODUCTION

Images are one of the major sources of data in the field of data science and AI. This field is making appropriate use of information that can be gathered through images by examining its features and details.

The idea behind this project is to build a deep learning-based Image Classification model on images that will be scraped from e-commerce portals. This is done to make the model more and more robust.

Analytical Problem Framing

- **Data Sources and their formats**

The data was collected from amazon.in from three different sections of clothing. These sections consisted of jeans, trousers and sarees. The data was in image format (jpeg) and was collected with the help of browser automation with selenium and downloaded with the help of request and shuttle library.

- **Data Preprocessing Done**

All the downloaded images were preprocessed with the help of keras image_dataset_from_directory, they were reshaped into 256*256 pixels and divided into different batches of size 16. We also implemented some data augmentations like flipping images vertically or horizontally or rotating images by certain degrees with the help of sequential.

- **Hardware and Software Requirements and Tools Used**

All of the work in this project was done on Jupyter notebook. We used pandas and NumPy for working on data and using all the basic mathematical functions on it. All the data preprocessing, model building and other tasks were done with the help of tensorflow and keras.

Model/s Development and Evaluation

- **Testing of Identified Approaches (Algorithms)**

We used convolutional neural networks (CNN) for building our deep learning model and our model consisted of 17 layers.

- **Run and Evaluate selected models**

Initially I trained the model for 50 epochs and it gave accuracy around 86, then I increased the epochs to 100 and the accuracy also increased to 92 and finally I settled with 1000 epochs which gave us a pretty good accuracy score.

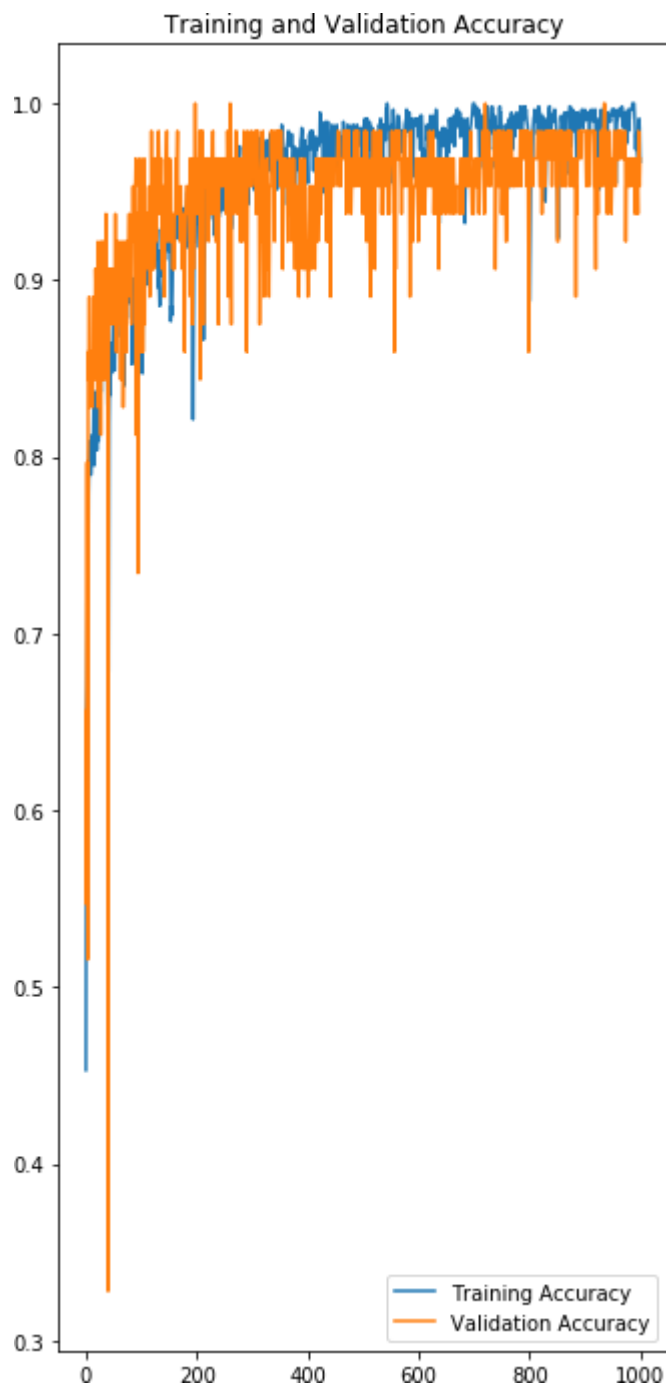
- **Key Metrics for success in solving the problem under consideration**

For evaluation of our models, we used accuracy scores.

- **Visualizations**

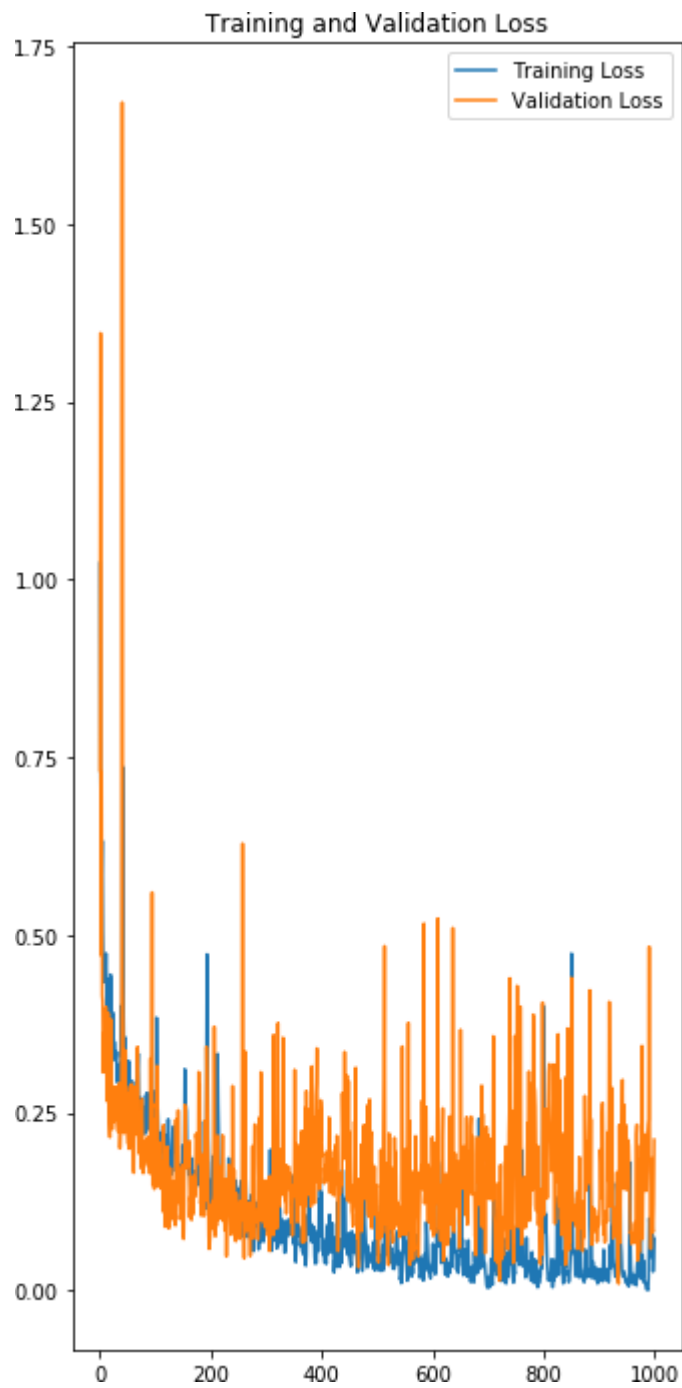
We used subplots to see the comparison between training and validation accuracy and also to see the comparison between training and validation loss.

```
plt.figure(figsize = (12, 12))
plt.subplot(1, 2, 1)
plt.plot(range(EPOCHS), acc, label = 'Training Accuracy')
plt.plot(range(EPOCHS), val_acc, label='Validation Accuracy')
plt.legend(loc='lower right')
plt.title('Training and Validation Accuracy')
```



```
plt.figure(figsize = (12, 12))

plt.subplot(1, 2, 1)
plt.plot(range(EPOCHS), loss, label = 'Training Loss')
plt.plot(range(EPOCHS), val_loss, label='Validation Loss')
plt.legend(loc='upper right')
plt.title('Training and Validation Loss')
plt.show()
```



- **Interpretation of the Results**

As we can see from the two graphs our training and validation accuracies showed a steep increment in early stages but later it slowed and was almost flat after 200 epochs. Similar was the case with the loss graph as we can see our losses decrease till around 200 epochs and then the graph straightened.

CONCLUSION

- **Key Findings and Conclusions of the Study**

In this project we build a CNN model for detection of clothing items from the image dataset.

While collecting the data from the websites with the help of an automated browser using selenium, we learned how to download image data with the help of request and shuttle library.

We created a model which could help in identifying different clothing items with very high accuracy.

- **Learning Outcomes of the Study in respect of Data Science**

While working on this project I learned about CNN, tensorflow and keras library and how we can create an object detection model with the help of these libraries and techniques with such high accuracy.