Deep Learning – Lab Project Report : Phase 1

Manvendra Singh – 200968146
Udit Amin - 200968132

Problem Statement :

Given a set of grayscale document images, the task is to classify each image into one of the 16 classes or document types. The training dataset consists of 16000 images with 1000 images belonging to each class. Example images are provided below for some classes. The dataset was collected from the RVL-CDIP dataset [1].

Objectives :

- Replicate the original DocFormer and LayoutLM models
- Achieve an accuracy threshold of atleast 75%
- Real life implementation and deployment into a cloud service using Docker

Metrics :

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix
- ROC AUC
- Log-loss

Meta-data :

The RVL-CDIP dataset consists of scanned document images belonging to 16 classes such as letter, form, email, resume, memo, etc. The dataset has 320,000 training, 40,000 validation and 40,000 test images. The images are characterized by low quality, noise, and low resolution, typically 100 dpi.

The data used for this problem statement is based on the above mentioned dataset.

The dataset at hand has document images, categorized into train and test directories. These directories contain image (.tiff) files. There is also a file for labels which has id and it’s corresponding label for mapping.

The dataset contain 16 classes of documents, which are :

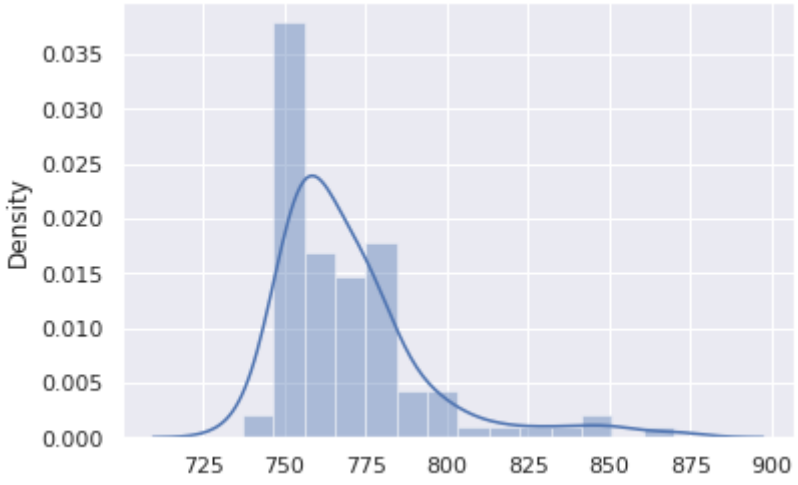
Deep Learning – Lab Project Report : Phase 1

1 : memo
2 : form
3 : email
4 : handwritten
5 : advertisement
6 : scientific report
7 : scientific publication
8 : specification
9 : file folder
10 : news article
11 : budget
12 : invoice
13 : presentation
14 : questionaire
15 : resume
16 : letter

Reference Paper

Exploratory Analysis :

The dataset is clean as it's been provided on kaggle. There isn't much analysis apart from the images themselves.



The above distplot provides the class distributions for the different classes. There is a class imbalance which can have an impact on the performance of the classifier.

While outputting the different images from each class shows that some classes like files and folder doesn't have that well defined images, the reason could be the way the images were processed.

Deep Learning – Lab Project Report : Phase 1

Pre-processing Pipeline :

For the baseline model, we are using CNN models. So for that we will be using the required tasks like applying convolutinal kernels for extracting feature map, and experiment with it and the pooling operations. Further we will be moving along with experimentations with different models like transformer networks and OCRs.