COMP 3106

Optical Character Recognition Project Proposal

Jigar M Dhameliya – 101177665
Rhythm Manish Shah - 101177565
Siddharth Reddy Busreddy – 101211997

Background information:

Optical Character Recognition (OCR) is a method for efficiently digitizing and processing huge amounts of text-based data by turning printed or handwritten text into machine-encoded text. The automation of data entry, document management, and text analysis across numerous industries has been greatly aided by OCR technology. OCR has also been really helpful to us in our effort to transform ancient literature and scripture into text data. The incorporation of AI provides an edge in overcoming constraints like difficult fonts, languages, handwriting, and low-quality documents.

Objectives:

The creation of a Transfer Learning model that significantly enhances text recognition for a range of fonts, sizes, and styles is the major goal of our OCR project. We will train our model on the MNIST-extended dataset and then apply what we have learned to the characters from Professor Matthew Holden's COMP3106 lecture notes to finetune the model. We want this project to benefit persons who are differently abled by enabling them to utilize it to transform handwritten notes into text-based data.
We also aim to create software that is simple to use and straightforward so that both technical and non-technical users can utilize the OCR system. Users need to have no trouble uploading single or multiple photos and getting precise text results for them.

Methods:

Our primary model will be trained utilizing the transfer model approach. Transfer learning is a key OCR technique. Pre-trained models can be tailored for certain OCR tasks, especially if they were trained on big datasets. In order to learn the fundamental characteristics of the characters, we will train the model using a sizable dataset (MNIST-extended). Then we will use hyperparameter optimization and fine-tuning methods to validate our data on the dataset from Professor Matthew Holden's lecture notes.

Dataset:

The dataset we've selected for our project is the result of team discussions and conversations with Professor Matthew Holden. For our project, we selected the MNIST-Extended dataset for a number of reasons; a Dataset of handwritten characters that aligns with our objective to recognize handwritten characters. Fairly large dataset as it has 100K+ handwritten letter data points and close to 280K digit data points. Aligns with the general rule of thumb to have more than 1000 training data points for each class. Data set with true labels which make validation easy and is Easily and publicly available. The dataset can be downloaded from Kaggle using this link.

Validation:

The MNIST-Extended has the added advantage of having true labels, which we can compare to the outcomes of our model to determine accuracy. Therefore, after we have trained our model on the training dataset, we will test it on the test dataset, compare the results with the true label, and record the accuracy. We will retrain our dataset to increase accuracy with each testing dataset iteration until we achieve an OCR model accuracy of

80%. When training our model using the dataset generated from the notes, we will employ the same techniques to perfect and optimize it.

Novelty:

OCR models and methods have been developed by other computer scientists, but what sets our project apart is the way we'll apply what we've learned from the COMP3106 course to assist students and teachers in digitizing their own handwritten notes for simple storage and usage. This is especially crucial for students with disabilities, as they frequently struggle to understand handwritten notes. This project intends to provide them with immediate assistance and create a welcoming learning atmosphere for everyone. We got the idea for this project by experiencing this issue firsthand in some of our courses.

Weekly Schedule:

Our weekly schedule is listed below:

Week	Planning
10/16 to 10/22	Research on how to use Transfer learning and its implementation in the project.
10/23 to 10/29	Reading Break (Mid term preparation)
10/30 to 11/05	Coding and implementation of the transfer learning model
11/06 to 11/12	Training and Testing datasets on our training model to check the accuracy. With this we will also use this week to modify the code and make any necessary changes in our implementation
11/13 to 11/19	Research and implementation of ways to use the notes as a dataset.
11/20 to 11/26	Fine-tuning model, and final testing on new datapoints
11/27 to 12/03	Report writing and submission
12/04 to 12/08	Spare week to rely on in case of any unforeseen circumstances in the weeks above

Availability:

We all three have pretty much the same courses and availability. Most of it is summarized below

Day	Time	Place
Monday	10:00 - 12:00 PM	In Person at university
Tuesday	3:00 - 5:00PM	In Person at university
Wednesday	4:00 - 6:00PM	In Person at university
Thursday	Not available	NA
Friday	4:00 - 6:00PM	In Person at university
Saturday	Varying availability	Online
Sunday	Varying availability	Online

GPU needs:

We conducted research on the length of time required for model training and testing as none of the three team members are experienced in the field of AI and machine learning. Since we have other courses going on at the same time and we learned that we would need to keep our computers running for extended periods of time for the model to train, it would be helpful to have access to a GPU to speed up the process. When creating the dataset from professors' notes, which may demand a lot of image processing and GPU power, we may also use the GPU for image processing.