Malayalam Text Recognition System

Aarya R. Shankar   Amrith M   Anand R   Sarathchandran S

Department of Computer Science

College of Engineering, Trivandrum

2015-2019

Table of Contents:-

Dataset Updates

- Received a non-uniformly distributed dataset from SPACE
- Cleaned up the dataset.
- Changed all images to 32X32 pixels
- Converted all Images to Grayscale

Model Trials

- Created a Test Model with the following features:
- 2 Conv-Conv-Pool Layers (Intuition from VGG16)
- 2 Fully Connected Layers
- Output layer with Softmax Log Probabilities(133 Classes)
- Tested on a Sample dataset with 40% accuracy (BAD)

Further Updates

- Getting More Data
- Use of augmentation and ensembles
- Optimizing the model and hyperparameters

Estimated Schedule of Completion

- ▶ Oct 25 - Final Data Setup in standard MNIST format
- ▶ Nov 10 - Optimizing the model (Layer,Parameters etc)
- ▶ Nov 20 - Completing the OCR and using trained weights for prediction. Implementing this to a system.

Conclusion

More than half of the works are done. Little tweak on the dataset and the model is left for the project to end.