

Translation / Transliteration of Vernacular Languages from Signboards

Project ID:21G378383

Review - I

Group Members

RA1711003030378 Vishnu Teja Chikkala

RA1711003030383 Rajpreet Srivastav

Supervised By:

Mr. Sunil Kumar

Assistant Professor

Department of Computer Science & Engineering
Faculty of Engineering & Technology
SRM Institute of Science & Technology



Table of Contents I

- 1 Abstract
- 2 Literature Survey
- 3 Architectural Design for Proposed System
- 4 Algorithms / Techniques to be used
- 5 References

Abstract I

India has 22 constitutionally recognized languages written in 13 different scripts. An average traveler, when travelling to a new region, might often get confused with signboards written in an unfamiliar language. It is also impossible to have every signboard in every city / town / village written in 22 different languages, as there will not be enough room to accommodate more than 2 – 3 scripts.

The goal of this project is to develop an App which translates the text written on a signboard into another language as desired by the user. The user should simply point at and click a picture of the signboard using the camera of their phone and the app will translate the text written on the signboard. This will be accomplished with the use of neural networks trained for text detection, recognition and translation tasks on collected and freely-available datasets, which will be developed using Python and frameworks such as TensorFlow. The user app will be designed in Java.



Abstract II

For the scope of this project, we will design a system which works for names (such as road names, city names, shop names, organization names, etc.) which typically are not longer than 4-5 words, and support translation for 1 or 2 languages. Further scope for the project involves including support for more languages and building models to support translation / transliteration of longer pieces of text.

- Indian community faces a “Digital Divide” due to dominance of English as mode of communication in higher education, judiciary, corporate sector and Public administration at Central level whereas the government in states work in their respective regional languages [6]
- India has 22 scheduled languages. While 99 % of the population speak one of these scheduled languages in various dialects (which number in the thousands) [1], according to Census 2011, the total percentage of English speakers is at 10 %, and that too is skewed towards the urban population. [9] Hence, there lies a need for developing NLP architectures for facilitating flow of digital content and information in and between local, national and international levels.

Literature Survey II

- The above also means that a large percentage of the literate population is either monolingual or bilingual, and across 22 languages, an accessible, easy-to-use and intuitively developed system is required, which enables intercommunication.
- While traditionally NLP has been approached with statistical methods such as Hidden Markov Machines (HMM), Support vector machine(SVM), Conditional Random Field(CRF), Naive Bayes(NB), etc, which take a large amount of tagged/annotated data (corpus) to statistically analyze and learn the language characteristics [3], the research into deep learning or 'connectionist approach' [3] with the use of trained artificial neural networks (ANNs), has gained impetus due to (i) the simplicity of the solution in rapidly prototyping and establishing practically effective systems (ii) the lower cost of annotation of the training data [7], and the fact that they attempt to

more closely emulate the learning process of biological brains, among other reasons. [3], [8], [4]

- Particularly, the collection of a uniform corpus and standard datasets for training models remains a challenge across all regional languages. The large number of morphological variations across Indic languages also contributes to this issue.[5], [10]
- Most of the Indian population accesses digital content through smartphones. In 2019, the number of smartphone users in the country passed 500 million [2], and is estimated to increase to 850 million by 2022 [11]. This, hence, also makes smartphones and smartphone apps in particular an ideal platform on which to launch NLP applications for the wider population, and directly help facilitate flow of information past language barriers.

- The basic functioning of the app is as follows:
 - User captures a photo of the signboard
 - The image is resized so as to be suitable as input to the model
 - The model takes the image as input. The text within the message is detected, extracted and transcribed to target language.
 - The text output (or error message, in case of failure to generate output within threshold confidence), is displayed on screen.

Architectural Design II

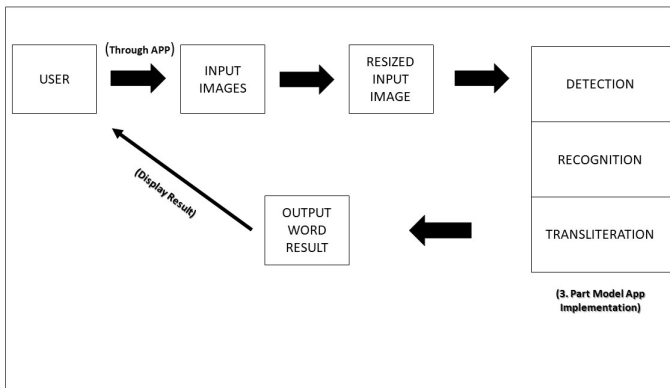


Figure: Functioning of App

- The artificial neural network (ANN) behind the core functioning of the app is made up of 3 models performing consecutive tasks. That is, the output of a preceding model will be fed as input to the succeeding model, and thus they act as one model unit.

Architectural Design IV

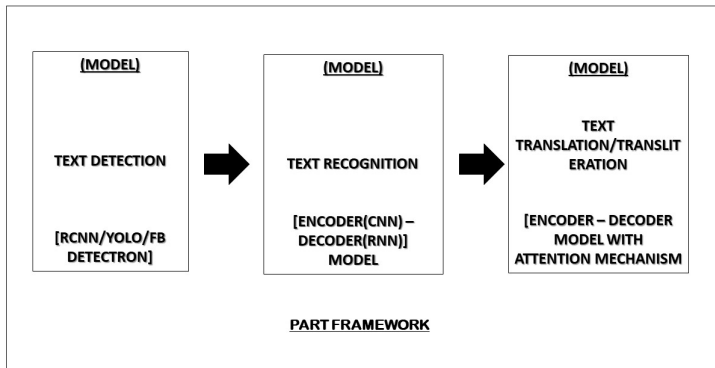


Figure: 3-part model

Techniques to be used I

- The 3 ANN models will be trained and tested on separate datasets curated to the task. The code for all the models will be written in Python using libraries such as PyTorch and Tensorflow, on Google Colab / Jupyter (.ipynb) notebooks, and using MLFlow for hyperparameter tuning. The models individually will perform the following tasks:
 - The first task is **Text Detection**, a subset of object detection, and will be performed by a Region-based Convolutional Neural Network (R-CNN). A classic R-CNN uses Support Vector Machines(SVMs) in its object-classifying stage, but for more efficient training and prediction, these will be replaced by a feedforward network. The possibility of using a You Only Look Once (YOLO) model will be explored.





Techniques to be used II

- The second task is **Text Recognition**, which is a sequence labelling task. This will be achieved by an Encoder- Decoder model. The encoder model will be a Convolutional Neural Network (CNN) which encodes input picture of text into series of feature representations, while the decoder will be a Recurrent Neural Network (RNN) which extracts text characters from the feature representations of the image. Connectionist Temporal Classification (CTC) loss will be used
- The third task will be **Text Translation / Transliteration**, that is, converting script from source language to target language and is a sequence-to-sequence generation task, which will again be modelled by an Encoder - Decoder model, this time with both encoder and decoder being RNNs. The encoder will convert the input to an intermediate form. The decoder will then generate output in target script from this intermediate form. Attention mechanism will be implemented into the model.

Techniques to be used III

- These models will be then be deployed as a single unit onto the mobile app, either in pruned and compressed form or using Tensorflow Lite / Mobile.
- The app itself will be for Android platforms, and will be developed in Java.

References I

-  More than 19,500 mother tongues spoken in india: Census, Jul 2018.
-  Smartphone users in india crossed 500 million in 2019, states report, Jan 2020.
-  N. P. Desai and V. K. Dabhi.
Taxonomic survey of hindi language nlp systems.
arXiv preprint arXiv:2102.00214, 2021.
-  T. Deselaers, S. Hasan, O. Bender, and H. Ney.
A deep learning approach to machine transliteration.
In Proceedings of the Fourth Workshop on Statistical Machine Translation, pages 233–241, 2009.

References II



A. Kunchukuttan, D. Kakwani, S. Golla, A. Bhattacharyya, M. M. Khapra, P. Kumar, et al.

Ai4bharat-indicnlp corpus: Monolingual corpora and word embeddings for indic languages.

arXiv preprint arXiv:2005.00085, 2020.



C. Kurian and K. Kannan Balakrishnan.

Natural language processing in india prospects and challanges.

In Proceedings of the International Conference on "Recent Trends in Computational Science, 2008.



J. Philip, V. P. Namboodiri, and C. Jawahar.

A baseline neural machine translation system for indian languages.

arXiv preprint arXiv:1907.12437, 2019.

References III



M. Rosca and T. Breuel.

Sequence-to-sequence neural network models for transliteration.
arXiv preprint arXiv:1610.09565, 2016.



R. S.

In india, who speaks in english, and where?, May 2019.



N. Singh.

Nlp for indian languages.
2020.



www.ETTelecom.com.

India to have 820 million smartphone users by 2022 - et telecom, Jul 2020.

Thank You