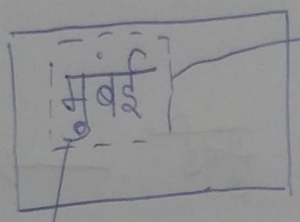
Project : App for Signboard translation of Vernacular Languages

Ideation

Part 1 : Neural Network



i) Text region detection (Object detection)

└ Data : Text in images and exact coordinates of their bounding box given

└ Model : RCNN / YOLO / FB Detectron kit

ii) Recognition of Individual characters

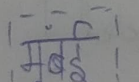
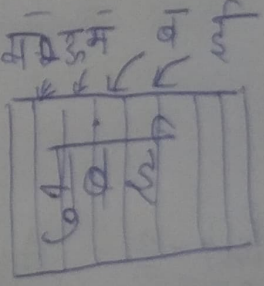
└ Data :  → मुंबई

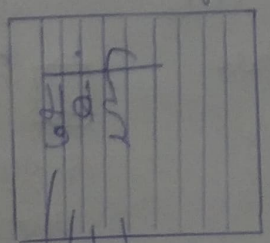
Image Cropped Section — Unicode String



└ Model : Encoder - Decoder Model

(CNN) (RNN)

iii) Recognition of individual characters from columns



Data : Text in Bounding Box

└ Feature representation using CNN (Columns)

└ Columns are a sequence. Passed through RNN. [Sequence Labelling Problem]

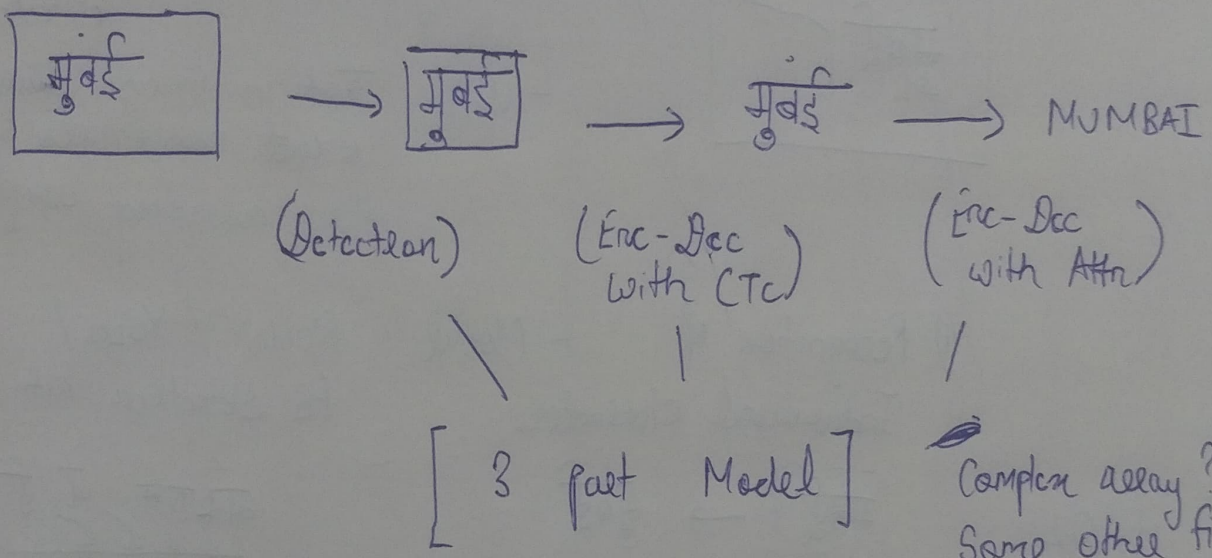
CTC loss is used ← (eliminates duplicate outputs among columns)

iii) Translation / Transliteration

Data : मुंबई ⟶ MUMBAI [redacted] : Output

Model : Encoder - Decoder model with Attention.

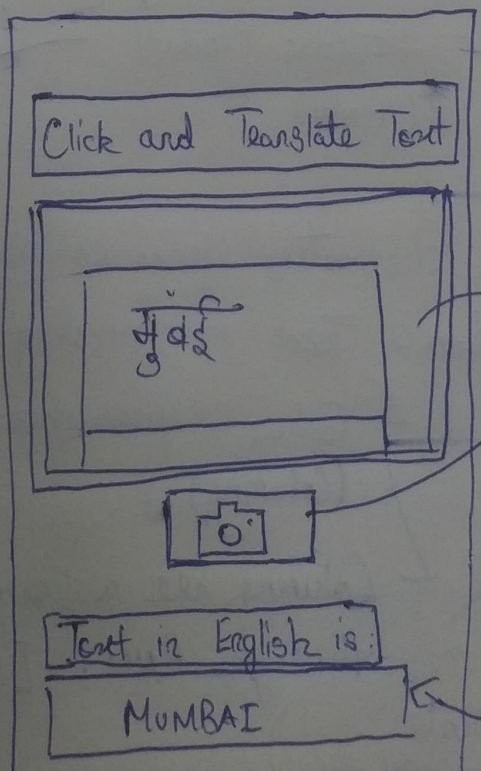
Process



Part - 2 : App Development

Main Page Example

[Coded in Java]



User Points and Clicks Picture

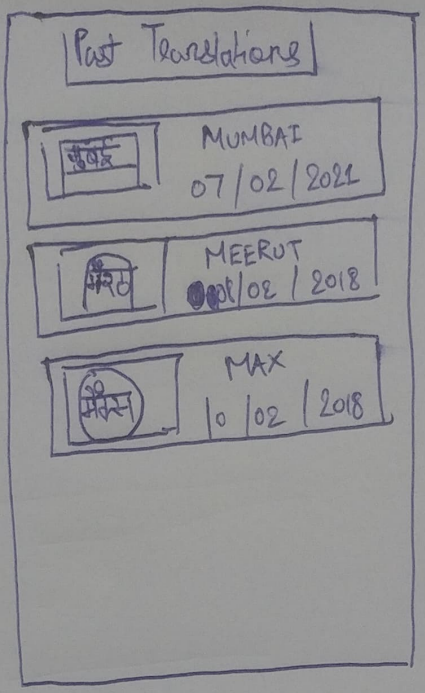
Picture is resized to fit model

Picture is given as input to Model

Output is displayed

PHOTO (Resized) → [3 part model]

Secondary Page



History of all translations with thumbnail and timestamp

[illegible] order dates from latest to [illegible] past

Research Topics

- Code model from scratch / use already available toolkits ?
- Code model in Python or Java ? How to integrate model with Java Code, if coded in Python ?
- Final form of model (complex array or another file) ? How to pass newly taken image through this model ?
- Optimizing model / app to work fast
- Augment app with dictionary / Google Search ?

How to use model in an app

Goals For February 2021

- Revise Encoder - Decoder Models, Android App Development
- Collect dataset for model training, choose model
- Develop main page UI for app