Text Detection -

The Text Detection process is mainly derived from the object detection process which mainly can be divided into two types.

1. Region Based Detectors
2. Single Shot Detectors

1. Region-based Detectors -
    a. They find all the regions that have the presence of objects (Bounding Boxes are drawn over the objects)
    b. They then pass these onto a classifier (CNN) which then labels these objects into different classes.
    c. Classes detected with accuracy less than 60% are usually deemed as noise but in this case it can also be considered as a lack of training data
        i. Speed - Slow
        ii. Accuracy - More accurate
        iii. Object Localization - A separate step is involved
2. Single Shot Detectors -
    a. It detects the bounding box and the class at the same time as it is mainly used for real-time applications
        i. Speed - Faster compared to Region-based detectors
        ii. Accuracy - less accurate
        iii. Object Localization - uses predefined presets for bounding boxes.

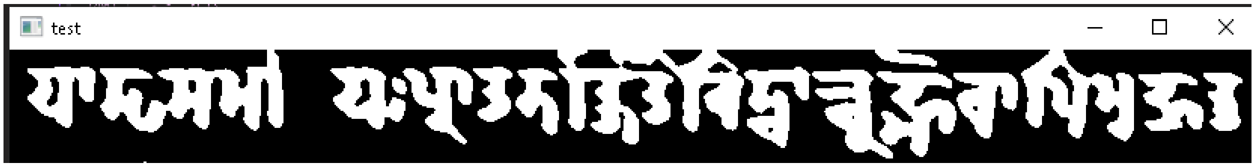
In this case we decided to go with Region-based detectors as we aimed for a higher accuracy with a smaller dataset.

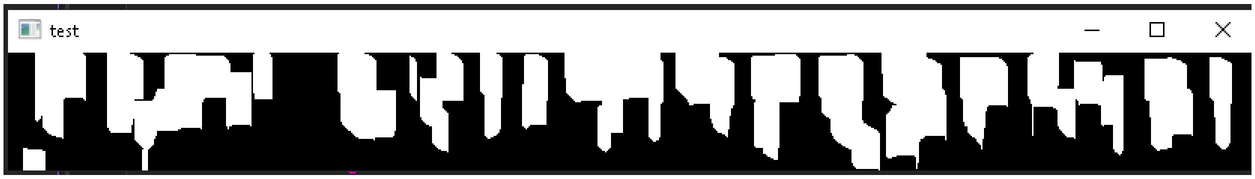
YOLO being one of the examples of SSD requires an extensive dataset and is usually used in conjunction with tesseract.

Deep learning models like Resnet-50 and VGG-16 generalize images with lesser amounts of data and they perform accurately on external data.

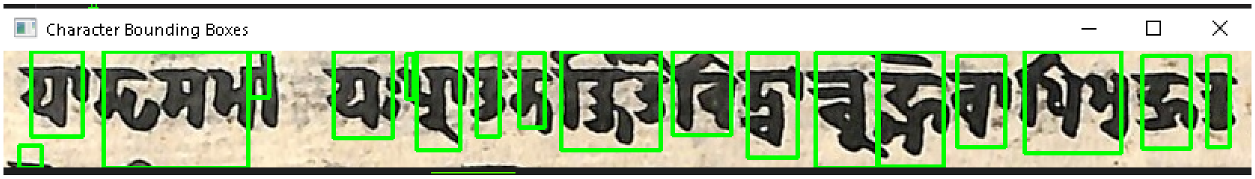
Treating matras / vowel marks as a separate character

Though it may seem practical to treat the anusvara and visarga as a separate character and then adding them to the main word , it’s a bit hard to accomplish that.





It can be seen in the process of dilation , the pixels which are in close proximity merge with each other and it’s hard to treat them as separate entities. This sometimes results in bad segmentation of bounding boxes.

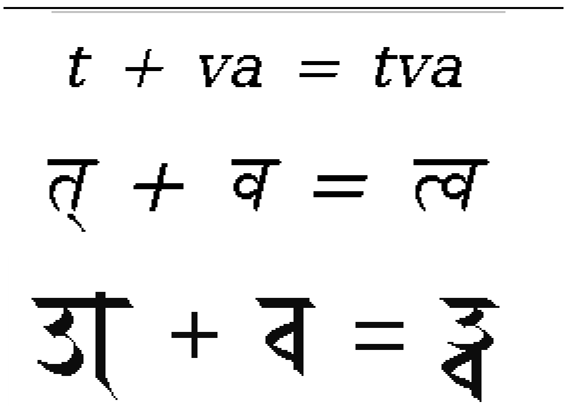


The classifier can take care of prediction of most of these labels and we also plan to implement a transformer in the pipeline which will correct the word errors using a dictionary.

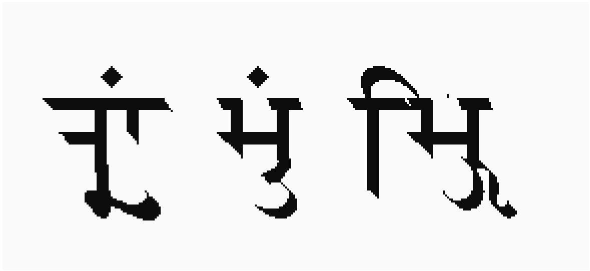
Hence minor outliers like the anusavara and visarga can be permitted.

Complex Characters

Treating complex characters as two separate is possible if they join in a horizontal manner.



From my observations , most of the compound characters are combined in a vertical fashion.



Annotations for compound characters which conjoin horizontally can surely be considered as an outlier and avoided. As mentioned before , the transformer will act as a spell checker and fix the words.