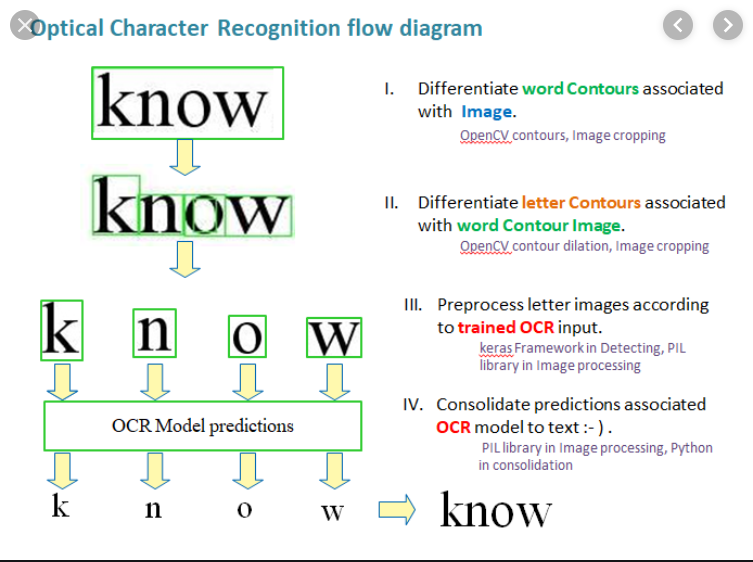
**Document Text Recognition**

**Project Idea:**

Optical Character Recognition has taken a big leap after the advent of deep learning algorithms. In terms of accuracy it almost equals to human beings intelligence.

In this project we will be identifying the text present in the given document/PDG by using deep learning algorithms.

The main idea of project is to build an end to end application that accepts an image/document/PDF and returns the text present inside the image content.



The entire idea of document text recognition is split into two sub tasks. One task would be text detection and the next one is to identify the text inside the detected regions.

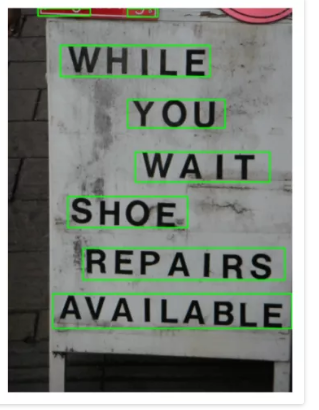
The entire document text recognition can be summarized by using above image. Where we start with detecting the regions containing the image. The regions containing the images will be highlighted with some rectangular boxes as shown in the figure above.

The process then continues to extract the text from the detected text segments.

The entire process of text detection and recognition is implemented in a Pipeline.

**Text Detection:**

This is the process of localizing all the text present in the image in the form of text segments.



This task can be accomplished by using various deep learning algorithms. Few of them are

EAST: Efficient and Accurate Scene Text Detector.

It is a deep learning algorithm which is capable of learning text regions in a given image and return precise text segment during the prediction time.

CTPN: Connectionist Text Proposal Network

This algorithm is capable of extracting text lines in a given natural image.

Textboxes++:

Uniqueness of this algorithm is that it precisely predicts the oriented text segments.

**Text Recognition Algorithms:**

1. **Convolution Recurrent Neural Networks with CTC:**

The text segments/regions containing the text are made to pass through few convolution layer. The feature maps thus obtained are flattened and then passed through the RNN’s(Recurrent Neural Networks).

RNN’s are one type of neural networks specifically suited for decoding sequential data. Since text is sequential data the flattened feature maps also called feature sequences are passed as input to a Bidirectional LSTM RNN.  
RNN’s are trained by using the flattened sequences using CTC as loss function.

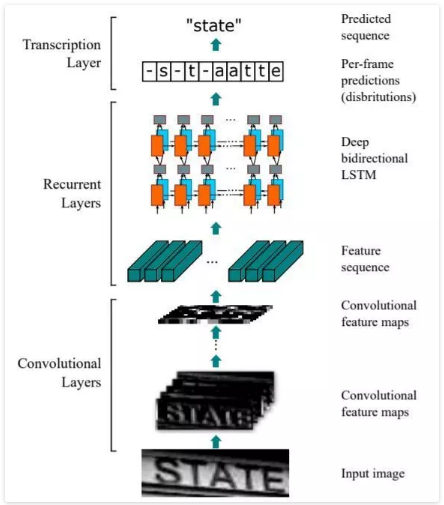
**CTC-Connectionist Temporal Classification**

A CTC works as a transcription layer in a neural network. The output from the RNN is passed to this CTC which will create precise text from the predictions of each time step of RNN.

**Why CTC?**

Normally supervised learning algorithms can have a problem with alignment in case of sequence generation. Let say if we do not use the CTC, we need to label the dataset character by character. Let say we are having images of width 150 pixels, and we are taking 150 time-steps as input sequence to the RNN, we need to label each time step in the training dataset. CTC takes care of all this. You only need to provide the text present in the image as the label to an image.

Look into below image for a detailed view.

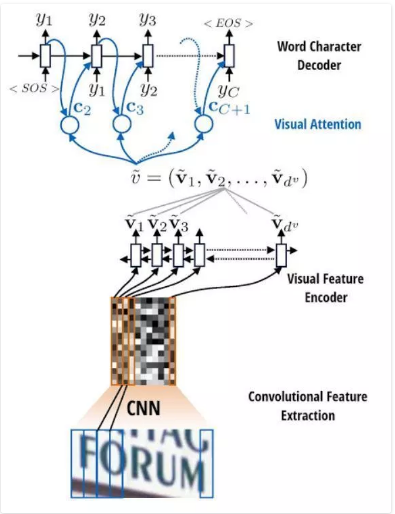


1. **Attention-OCR**

Text inside the text segments are obtained by application of attention mechanism which is one kind of Recurrent Neural Networks. Attention models are generally used to solve neural machine translation problems involving text sentences. Attention models tries to predict the output text by paying attention to few important words in the given input hence the name attention.

Since text recognition in OCR is also sequence based problem attention mechanism can be applied for accurate text predictions.

The below diagram explains the entire pipeline of Attention based OCR



**Dataset used for Capstone Project**

Brno Mobile OCR Dataset:

Brno Mobile OCR dataset is used for text recognition task. The dataset is collection of images captured by handheld devices like mobile cameras.

The dataset is different from existing datasets because of its large collection of images captured by using mobile cameras.

Images that are captured by mobile cameras having non uniform lighting, image blur, strong noise, built in denoising, sharpening.

The main aim of using this dataset is to train a model which will be capable of text recognition on images of various qualities and even on low resolution images.

More details of dataset is available on Data Collection section of capstone project.