A Project Report on

OCR Model For Visually Impaired People Using TensorFlow

Submitted in fulfillment of the requirements for the award
of the degree of

Bachelor of Engineering

in

Computer Engineering

by

Nidhi Munavalli(16102049)
Apurva Waingankar(16102050)
Anuja Velaskar(16102042)

Under the Guidance of

Prof.Ramya RB

Department of Computer Engineering
A.P. Shah Institute of Technology
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615
UNIVERSITY OF MUMBAI

Academic Year 2019-2020

Approval Sheet

This Project Report entitled ***"OCR Model For Visually Impaired People Using TensorFlow "*** Submitted by ***"Nidhi Munavalli"(16102049),"Apurva Waingankar "(16102050),"Anuja Velaskar"(16102042)*** is approved for the fulfillment of the requirenment for the award of the degree of ***Bachelor of Engineering*** in ***Computer*** from ***University of Mumbai***.

Prof. Ramya RB
Guide

Prof. Sachin Malave
Head Department of Computer Engineering

Place:A.P.Shah Institute of Technology, Thane
Date:

CERTIFICATE

This is to certify that the project entitled ***"'OCR Model For Visually Impaired People Using TensorFlow"*** submitted by ***"Nidhi Munavalli"(16102049),"Apurva Waingankar "(16102050),"Anuja Velaskar"(16102042)*** for the fulfillment of the requirement for award of a degree ***Bachelor of Engineering*** in ***Computer.***,to the University of Mumbai,is a bonafide work carried out during academic year 2019-2020.

Prof. Ramya RB
Guide

Prof. Sachin Malave
Head Department of Computer Engineering

Dr. Uttam D.Kolekar
Principal

External Examiner(s)

1.

2.

Place:A.P.Shah Institute of Technology, Thane
Date:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

_____

(Nidhi Munavalli (16102049))

_____

(Apurva Waingankar (16102050))

_____

(Anuja Velaskar (16102042))

Date:

Abstract

There are many cultural, governmental, commercial and educational organization that manage large number of manuscript textual information. English being one of the most widely used language, organization include English documents.Optical Character Recognition (OCR) is the process of extracting text from an image. The main purpose of an OCR is to make editable documents from existing paper documents or image files.

In this digital world, accurate identification and recognition of the text has been an important key area of image processing and document analysis. It is difficult to interpret some alphanumeric symbols which are similar in appearance. To avoid this a perfect and better guidance is needed for correct identification of characters and which can distinguish between symbols sharing similar or even identical physical characteristics. In such cases Artificial Neural Network (ANN) pays off. Since single ANN's can't get the appropriate results, we employ best Deep Learning Neural Networks emphasizing mainly on Convolution Neural Network (CNN). The model can handle different font types and font sizes. The experimental results show high level precision for detection of machine printed document images. In continuation to this the project helps visually impaired or blind people for recognising the text as it is converted to speech as well as braille.

Contents

List of Figures

List of Abbreviations

OpenCV:	Open Computer Vision.
CNN:	Convolutions Neural Network
ANN:	Artificial Neural Network.
OCR:	Optical Character Recognition
SVM:	Support Vector Machine
TAS:	Triple Adjacent Segment
MST:	Minimal Spanning Tree
CDM:	Character Deformation Model
CDF:	Character Deformation Field
CC:	Connected Components
MRF:	Markov Random Field
SLF:	Standard Lattice Format
FVT:	Feature Vector Table
KNN:	K Nearest Neighbour

Chapter 1

Introduction

Optical character recognition (OCR) is identification of optically processed characters. One of the fields of pattern recognition which contributes to recognise printed documents is Optical Character Recognition. It is a technique to process different type of documents, images, pdfs to American Standard Code for Information Interchange (ASCII) or machine editable form which can be further edited or processed. OCR is widely used as document scanners, recognition of characters and language, authentication in banks, security purposes etc. OCR can further be classified into two types: offline character recognition and online character recognition system. In online, the initial stage of character identification is not necessary as the characters are processed as it is. Offline are classified to handwritten and printed OCR. In this paper, the character recognition is done for machine printed documents or images using Convolution Neural Network. The recognised text is then converted to Braille Language which is used by the blind or visually challenged people to read and write. The best use of this technology is to have a computer read text to them out loud hence the text is also converted to Speech. Braille has six dots organised in a 3x2 matrix and therefore has 64 (2^6) symbols.

One of the fields of pattern recognition which contributes to recognise printed documents is Optical Character Recognition. They scan the printed documents to an image and recognise characters present to create a respective digital text document which can be further edited or processed. The best use of this technology is to have a computer read text to them out loud or can have it printed in braille.

OCR, or optical character recognition, is one of the earliest addressed computer vision tasks, since in some aspects it does not require deep learning. Therefore there were different OCR implementations even before the deep learning boom in 2012. This makes many people think the OCR challenge is "solved", it is no longer challenging. Another belief which comes from similar sources is that OCR does not require deep learning, or in other words, using deep learning for OCR is an over skill.Text line segmentation of a document image is considered as a critical stage towards unconstrained document recognition. Line segmentation is the first and the most critical pre-processing step for a document recognition, followed by word segmentation, word recognition and other indexing steps. Different types of documents give arise to different types of problem.

The following are the steps of OCR model

A) Input Image.
B) Pre-processing
C) Segmentation
D) Extraction
E) Training
F) Optimization
E) Convert to speech
F) Translation to Braille Language

The above steps gives the overview of the proposed system, a text document undergoes pre-processing steps, This output is given to the Segmentation process where every line following the words and letters are being Segmented by using bounding boxes.This output is given to the background cleaning step. In this stage, all the noise is removed. Then the Extraction step is done,then the image passes through the model and training of letters are implemented.The output then is optimized by using English corpus.Later on,the identified text is converted to Braille language.

1.1 Objective

- Recognize text in scanned text documents,text images, and any picture taken which is in English.The primary goal is to speed up the purpose of character recognition in document processing.As a result the system can process huge number of documents with-in-less time and hence saves the time.

- To extract and convert to speech.

1.2 Problem Definition

The problem in the project is to mainly to recognize the English text then to convert the English text into Braille Language and Convert English text to speech.Many methods have been proposed but most of them are restricted and complicated.To improve the efficiency of OCR segmentation plays a vital role The text line segmentation in the documents remains an open document analysis problem.Hence in this project,the problem is to segment,extract the English document and also convert to Braille language and convert to speech.

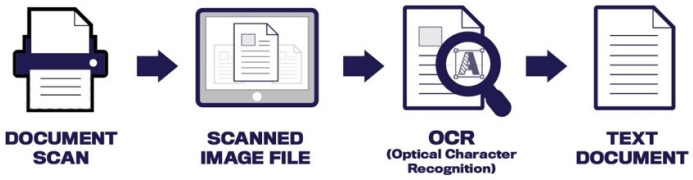


Figure 1.1: Process of OCR Model

1.3 Scope

OCR model is conversion of images of typed or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo) .OCR is a field of research in pattern recognition, artificial intelligence and computer vision.
The scope of our project Optical Character Recognition is to provide an efficient and enhanced software tool for the users to perform Document Image Analysis, document processing by reading and recognizing the characters in research, academic, governmental, business organizations and for blind people that are having large pool of documented, scanned images. Irrespective of the size of documents and the type of characters in documents, the product is recognizing them, searching them and processing them faster according to the needs of the environment.
The model takes English text scanned image as an input. This image is analysed in order to identify each letter or digit. When a character is recognised it converts it into braille language as well as in English language and also speech conversion is implemented. The output is in the form of well recognised and understandable document.

1.4 Technology Stack

- TensorFlow-TensorFlow is a free and open-source software library for data flow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks.It is used for training the model.
- Open CV- Open source computer vision is a library of programming functions mainly aimed at real-time computer vision. In this project it is used for image processing.
- Python 3.7- Python is an interpreted, high-level, general-purpose programming language.It is the programming language used in this project with its basic libraries.
- Azure Jupyter notebook- It is a web-based interactive computational environment for creating Jupyter notebook documents.The source code of OCR model will be executed in this.

- Numpy-is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.
- Tkinter-Tkinter is the Python interface to the Tk GUI toolkit shipped with Python.
- SciPy-SciPy is a free and open-source Python library used for scientific computing and technical computing. SciPy contains modules for optimization, linear algebra, integration, interpolation, special functions, FFT, signal and image processing, ODE solvers and other tasks common in science and engineering.
- Matplotlib-Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy.
- Scikit-learn-Scikit-learn is a free software machine learning library for the Python programming language.It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.
- Pygame-Pygame is a cross-platform set of Python modules designed for writing video games. It includes computer graphics and sound libraries designed to be used with the Python programming language

1.5 Benefits For Environment And Society

- Paperless revolution- The project stores the documents in soft copy and hence the paper work is reduced.
- Environment friendly- Due to reduction in use of paper, deforestation is also reduced.
- Retyping-It reduces the work of retyping the text as it can be directly scanned and converted to document.
- Speedy digital searches -By converting scanned text into a word processing file, OCR lets you search through documents using keywords or phrases.
- Saving space - OCR can scan the documents from the paper document.This means that data can be stored in electronic format in servers eradicating the need of maintenance of paper files.
- Document saved as audio file - People with learning disability who has difficult reading large amount of text or who are visually impaired can benefit from this audio file offering them an eaiser option.

1.6 Application

- The project recognises scanned image and converts it into English text document. The document would be precise and more accurate.
- Assistive technology for blind and visually impaired users- The project converts the scanned image into braille language.This is helpful for visually impaired people.They can easily understand and interpret it.

Chapter 2

Literature Review

[1] In this paper, the Segmentation of the text is done by calculating the information energy of the energy pixel in the scanned image. The low information pixel is present in blank spaces in between the words in the documents. Hence, based on energy pixels content of the information energy map is generated from which the text lines location is found. In the event that the gaps between the characters are bigger than the word gaps, at that point, they have utilized vertical segmentation and k-means clustering. Using diagonal based feature extraction the segmented character feature is extracted, for classification and recognition feed-forward artificial neural network is utilized. Their method can be used in the future to increase the recognition, accuracy of recognition is 92% it tends to be additionally extemporized to expand the exactness.

[2] Xujun Peng, used an approach to categorized three sorts of text i.e. noise, handwritten text,and machine-printed text, from an annotated machine-printed document by Markov Random Field (MRF). Each record is displayed as a random field with numerous patches which are pieces of the image. Patches size less than or greater than a particular threshold are detected as noise and are eliminated. During training, the initial labelling is carried out by the G-means clustering method and the centers are used as hidden models in the MRF. Compared to a single classifier this method shows good performance of classification.

[3] This paper uses the Zernike moments technique for handwritten character recognition. Using SVM and KNN classification for a compound character it has given better result.0.37% rec rate is obtained when compared to other handwritten recognition software.

[4] For Feature extraction geometry-based technique is described many line types are extracted that forms a character. The geometric feature like local and global are used for feature extraction. A universe of discourse is selected of different windows that are divided on which feature is done. Toning followed by character traversal is done for the image. With the help of feature, the vector is trained and the neural network is used for testing.

[5] Here, the images are acquired only in .pbm format. Pre-processing steps include binarization and normalization. Feature extraction is done by creating a feature vector table (FVT). A total of 17 features are extracted and FVT contains 238 feature values. Training and classification are done by the HMM tool kit (HTK). For data presentation, utilizing HTK's standard lattice format (SLF) a word network is characterized. A file grammar.txt

is extracted containing all grammar, hmmlist.txt contains all character and for sorted list of character file named dictionary.txt is created which for training recognition of character using recognition tool HVite, for increasing accuracy, feature selection method is used, which are used for dimensionality reduction which removes unwanted feature and improves accuracy.

[6] Pre-processing is carried on a word. Where thinning separation is carried on words and later distance criteria are used. In post-processing process, graph theory is used for dealing with overlapped words. The segmented words are validated using SVM classifies, Tabacco-800 dB, and proprietary dB. The character recognition is calculated on Dongre and Mankar, CPAR, Chars74k database, and CVLSD with a proprietary database. For feature extractions, Fourier transform, wavelet transform, Gabor T, Scale Invariant feature T, Hartley transform Gaber filter are used.

[7] Optical Chinese character recognition for low-quality images are dealt with the character deformation model (CDM).Two phases of character matching are, a)Estimate deformation between images and ref images, b)Fine-tuning to reduce the distance between character. A character is represented by a bitmap of a four-level intensity. For detailed character matching, a vector map is used. Utilization of Image editing rule helps to find the correct bitmap template matching the image. Character deformation field (CDF) is utilized to deform one image to match well with other images. Iterative deformation method guides to match templates to an unknown character. A mutual match is used to match reliability.

[8] Josè A. Roodriguez, uses a sliding window approach which slides across the image from left to right, at each point sliding window is partitioned into cells and in every cell histogram of orientation is accumulated. Handwritten word spotting is a method to detected keywords from the handwritten document image. The fundamental commitment is another sequential feature that gets execution well past the state-of-the-art in an unbounded word spotting task. Which is influence by SHIFT key point descriptor which is a histogram of oriented gradient of an image. The hidden Markov models and Dynamic time warping are two different word spotting systems used. In hidden Markov models the training and testing are carried out in 5 folds i.e. trained on 4 folds and tested on 1 fold, while 5 random images are used as queries in DTW. The negative distance to the closest query is taken as a similar score for an input image.

[9] In the paper, the extraction of printed text and handwritten is separated by using a triple adjacent segment (TAS) and then calculate the normalized histogram of code-words for every segmented zone and utilize for training a Support Vector Machine classifier (SVM). Firstly, the zone present in the document is extracted using Voronoi segmentation. Secondly, a shape codebook is generated using canny edge detector it obtains a list of edges present in the image, then by fitting a line to each edge segment with a certain tolerance we find a list of similar line segments, then categorizing the neighboring segment to form a connected component (CC) the CC triplet form TAS basic shapes. Simple k-means is used to cluster the TAS extraction. For each zone of segmentation the descriptor is constructed which shows the rate of occurrence of each feature of TAS component.

The two normalized histograms of handwritten and printed is linked to obtain single feature vector. we utilize a variation of SVM called v-SVM to train a two-class classifier for handwritten and machine printed zones. The advantage of this technique is it is powerful to size and noise of the zone.

[10] Fei Yin utilizes a methodology an approach of text-line segmentation algorithm based on Minimal Spanning Tree clustering (MST) along with distance spanning learning. This is a bottom-up technique. With the assistance of the distance metric a tree structure is formed with the documented image. Supervised learning designs the distance metric. Every text-line can be seen as a cluster of connected components (CC) or stroke pixels. The algorithm takes those CC and changes over it into text-lines since it is productive for detecting clusters with irregular boundaries. This bottom-up technique can separate curved, marginally covering text lines and multi-skewed lines. The algorithm is liberated from the supervised learning and artificial parameter of distance metric improves the precision of text line identification basically.

[11]This paper proposes techniques for skew correction and text line extraction for extracted text lines using a cost function which considers the spaces between the skew of each line and text line. The issue is expressed as an energy minimization problem, so minimum cost function is yield as a set of text lines .it is also an efficient technique for the baseline correction, using the skating window the lower baseline is normalized to the horizontal line. The methodology includes input image, segmentation, background cleaning, and skew correction. The cost function helps to reduce errors and increases the accuracy of the project.

[12] Sunanda Dixit proposes an approach for segmentation of line for handwritten text documents that are in the Kannada Language. The algorithm finds the components, bounding box, and coefficient of variance. By connecting the centroids within the bounding box and coefficient of variance the segmentation of the line is performed. This procedure has been tested on Kannada script and it results in high degree accuracy and performance. The process consists of three main steps first, componentization which classifies image to connected component and identifying the bounding box second, filter process eliminates the components with very small area finally, line construction process which computes the coefficient of variance and line segmentation.

[13] The main objective of this paper is to present essential information about skew correction, zone segmentation, and character segmentation. The paper presents a better technique for text line segmentation two major techniques used are sliding window and adaptive histogram adaptation equalization. After pre-processing the image is passed to histogram equalization where the text character of the document image is enhanced for higher accuracy this text line is segmented by utilizing the sliding window operation.

[14] In this paper, segmentation and recognition of handwritten and scanned documents are implemented. It defines a novel technique for unconstrained text line segmentation of handwritten documents in the Kannada script. First, the scanned document is taken then all the pre-processing methods are implemented on the image then, connect components are

found then destructor from the image are removed then the image is normalized using standard error then component syllable is found and they are grouped, then the line segregation using weighted bucket algorithm is implemented then text line extraction is done as a final step. This approach has gained higher accuracy.

Chapter 3

Methodology

3.1 Pre-Processing

Image preprocessing involves the following steps:
• Image Acquisition.
• Noise Removal.
• Normalization.
• Skew Detection and Correction.

In Image Acquisition the saved image is retrieved from the computer.The images of format jpeg, jpg, png are supported. The images are then converted to grayscale image. These pictures are taken from digital camera or a scanner. Gaussian blurring technique is used for noise removal in the images. Instead of a normalized box filter, a Gaussian kernel is used for convolution. The kernel dimension and standard deviation in both directions can be determined independently. Otsu technique is used which is an adaptive thresholding way of binarization in image processing. It finds the optimal threshold value by going through all the possible threshold values(0-255).In tilt detection, the text pixels of the image are detected and a minimum area enclosing the pixels are created and the tilt angle is calculated. Necessary rotation is performed to straighten the image.

3.1.1 Image Acquisition

The input image is a retrieve image saved in a remote location in the computer.Image formats supported are:JPG, PNG.

3.1.2 Noise Removal

Gaussian noise from the image is removed by Gaussian Blurring. The Gaussian smoothing is a convolution operator which is majorly used for 'blurring' the images so that detail and noise are removed. Kernels with Special properties are used. By Gaussian's standard deviation the smoothing degree is determined. Because of the frequency response, Gaussian is justified in its use.

3.1.3 Normalization

Otsu technique is used for binarization which is an adaptive thresholding way. The optimal value of the threshold is found by checking the various values from 0 to 255 and the measure of spread is calculated for pixel levels. It is used in applications of computer vision.

3.1.4 Skew detection and correction

The text pixels of the image are detected and the minimum area enclosing the pixels is created. Then the skew angle is calculated. Necessary rotation is performed to correct the skew.

3.2 Segmentation

Text image segmentation can be achieved at three levels. As we move at different levels of text segmentation hierarchy, we obtain specifically finer details. Segmentation at any of these levels directly depends on the nature of the application. More the details required for the image, the more is the level of segmentation.

3.2.1 Line Segmentation

For segmentation of line, the image is projected horizontally to get a projection profile which is the histogram of pixels having zero on each row of the image. The threshold value is used to differentiate between two lines. In horizontal projection, the words in the same line have same threshold values and this is considered for line detection.
After changing over the coloured image to a binary image we have just white pixel indicates the appearance of foreground pixel and background pixel implies nonappearance of foreground pixel. In this technique, we count the number of foreground pixels along all rows of the image, and the resultant histogram array exhibit is of the size, equivalent to the number of rows in the image. Higher peaks suggest that number of Foreground pixels along that row are high and number of Background pixels along that row are low, Similarly Lower peaks suggest that the number of Foreground pixels along that row is low and a number of Background pixels along that line are high. The objective of horizontal projection is to determine the co-ordinates i.e. starting and end point of each row, which can be divided in to lines. Horizontal projection profile of the given binary image is:

$$ph(j) = \sum_{i=1}^{m} f(i,j) \tag{3.1}$$

Equation represents the sum of black pixel perpendicular to y-axis.

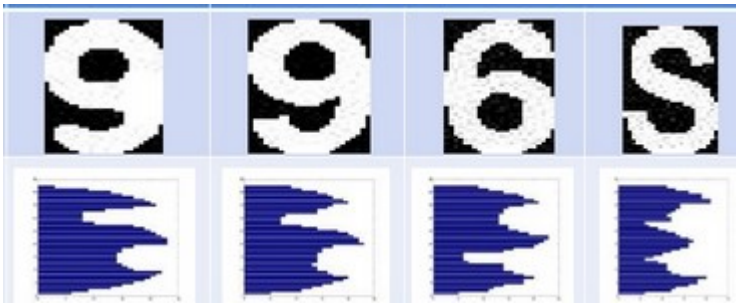


Figure 3.1: Horizontal Projection

3.2.2 Word Segmentation

The lines segmented from document image are further segmented into individual words. Vertical projections of the image is calculated and is used for finding the spacing between the words. Vertical projection profile is the histogram of pixels having zero on each column of the image. The white spaces differentiate the two words. The projection in the space of two words have value zero. This difference in threshold values helps in segmentation of words.
In the previous step we segment the whole image into lines, in this step input is the segmented single line. Here the main objective is to segment the lines into word. Using the idea of pervious step, but changing the projection from horizontal to vertical projection i.e. sum taken along the columns. When vertical projection is implemented columns that have text pixels have higher peak value and column which have spaces will have lower peak values.

Figure 3.2: Vertical Projection

3.2.3 Character Segmentation

The characters are segmented from the segmented words. The word images are initially grayscaled and threshold value is calculated and applied on the image for binarization.The contours are later extracted from the image for further segmentation. The contours are the connected black pixels in each character which are determined by threshold values. These contours are connected to form a single character. Two different words are recognized by the white spaces between contours. For contour detection the first pixel needs to be identified. After this all the pixels are identified following the path. All the contours of same characters are detected to form a bounding box around the character. A bounding box is created around the character according to the scope of those contours. The characters are detected using the bounding box. Hence separate character images are created based on the contours.

After segmentation of each row into words the next step is to segment the words to character, so here technique of contour based extraction is used. From each word contours are detected, from these contours all the boundary points are detected and bounding box along the bounding points are created. For further use of these character images they are saved in 32X32 size with additional padding.

3.3 Extraction

The bounding boxes around the characters are useful for extraction process. Each bounding box contains one character to be extracted. Only the boundary pixels of each bounding box is considered. The entire scanning of image takes place and bounding boxes are found. These boundaries defines the characters present in them. The single bounding box is represented as one image. They are resized to 32 x 32 with added padding and are extracted. These extracted character and words are further used to train.

3.4 Training

Convolution Neural Network comes under Artificial Neural Network which is abundantly used for processing and recognition of images. As it gives good results it is used in Natural Language Processing, speech recognition, recommendation system etc. Every images's distinctive properties are extracted for machine learning basis. Since the usage of parameters are less its more efficient compared to other deep learning models. It is difficult to interpret some alphanumeric symbols which are similar in appearance.To avoid this a perfect and better guidance is needed for correct identification of characters and which can distinguish between symbols sharing similar or even identical physical characteristics. Hence convolution neural network is used to train the model to increase the accuracy of detection.

Convolution is the first layer used in the model and it consists of convolution kernels in which every neuron act as a kernel. Here, respective fields of the image are created which means image division into small slices which helps in feature extraction. Convolution operation is expressed as:

$$f_l^k(p,q) = \sum_c \sum_{x,y} i_c(x,y) * e_l^k(u,v) \tag{3.2}$$

where, i_c (x, y) is an element of the input image tensor I_C, which is an element wise multiplied by e_l^k (u,v) index of the k^{th} convolutional kernel k_l of the l^{th} layer. There are different types of convolution operations based on size or type of the filters used, type of padding, and convolution direction. The weights can be shared hence many sets of features of the image are extracted by sliding kernel therefore CNN is parameter efficient. Strided convolution is applied to reduce the spatial dimensions. In the network architecture, the starting layers learn less convolutional filters and the deeper ones will learn more filters. The previous layers are given as an input to the succeeding layers.

To initialize weights MSRA normal distribution algorithm is used. Activation function is a decision function by which intricate pattern learning is possible. Therefore, an appropriate function should be selected to increase the process of learning. In this paper, ReLU activation function is applied along with batch normalization and dropout. ReLU (Rectified Linear Unit) is used for non-linear operations. As it introduces non-linearity it is used in CNN, the model learns fast and performs better as ReLU does not have vanishing gradient problem. Mathematically, its expressed as:

$$f(x) = max(0,x) \tag{3.3}$$

To resolve problems in the internal covariance shift of feature-maps Batch Normalization is used. The flow of gradient is smoothed by this and it also acts as a regulating factor that will further improve the network's generalization. Batch normalization is used as it stabilizes the training and tunes hyperparameters easily. To avoid overfitting and to regularize the network dropout is used, it generalizes by random skips in the units. Current layer neurons with probability disconnects randomly from the succeeding layers for the network to rely on the connections. Flattening layer prepares vectors to be fed into the fully connected layer. A fully connected layer is added at the end. It's a global operation as input is taken from the previous stages of feature extraction and to produce their output global analysis is done. One fully connected layer of 512 nodes has been added. Eventually, in the network as an activation function "softmax" is used, prediction values are the output of this layer. Softmax is a logistic classification function which is used for multiclass classification. Mathematically, its expressed as:

$$\sigma(z) = \frac{e^{(z_j)}}{\sum_{k=1}^{k} e^{(z_k)}} \quad for \;\; j = 1,2,...,k \tag{3.4}$$

Training the network is initiated when the model.fit_generator is called which accepts batch of data, performs backpropagation and the weights are updated in the model, this process

continues until the desired number of epochs are reached. Data augmentation is performed which is a form of regularization, enabling the model to generalize better. Adam is used as an optimizer. L2 normalization is used which reduces overfitting and helps in generalization. The first layer has 7 x 7 filters and the remaining layers have 3 x 3 filters. The dataset has 36,576 images which include 26 alphabets and numbers from 0-9. Each has approximately 1000 images in it. The data are split into 29260 and 7316 for training and testing respectively. The number of epochs used to train is 50. The output is the recognized English text. The accuracy obtained is 80.85% and value accuracy obtained is 90.34%.
Further, the trained model may show spelling mistakes in the output or while performing segmentation overlapping of letters may lead to bad contour values resulting in errors during recognition. Hence, to avoid this Optimization is necessary English Dictionary is used for this purpose. Each recognized word is sent to the correction function. It returns the original word present in the dictionary. If the words are unknown, it returns the words from one or two edit distance away. It estimates the probability of each word. The word having the highest probability is selected. It helps in improving the accuracy of the output.

3.5 Optimization

Overlapping of letters during segmentation leads to bad contour values which further produces errors during recognition. Optimization of text is carried out to correct the spelling mistakes if any. Dictionary is used for this purpose. Each recognized word is sent to correction function. It returns the original word present in the dictionary. If the words are unknown, it returns the words from one or two edit distance away. It estimates the probability of each word. The word having highest probability is selected. It helps in improving the accuracy of the output.

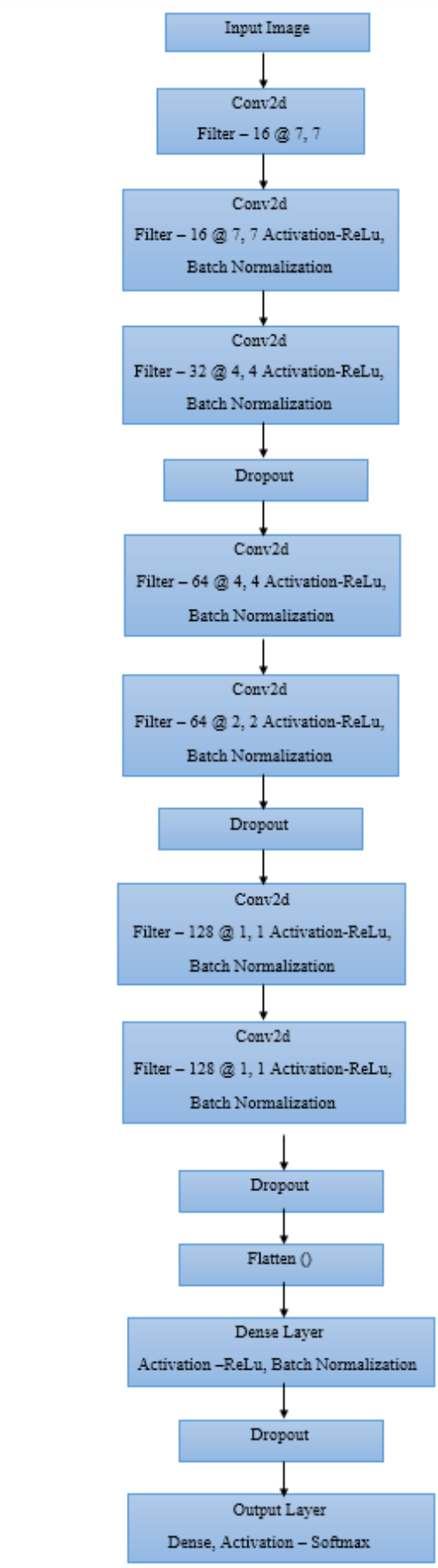


Figure 3.3: Internal Block of CNN

3.6 Conversion to Speech

This phase is specifically done for the visually impaired people by converting the words from the document to speech. The file of identified letters is broken down into lines and the lines are further broken down into words. These words are then converted to speech. Conversion to speech is done using Google Text to Speech API. The converted speech is stored in .mp3 format. Google Text to Speech API supports many languages like English, Dutch, Bengali, Hindi, French and many more in both male and female voices. This project makes use of text to speech in the English language with a voice of a female. Speed of the speech can also be adjusted as slow or fast according to user's needs.

3.7 Translation to Braille

Translation of English characters to braille is carried out for visually impaired people. They use dots to feel and read the text. The documents in Braille can be printed by special printers known as braille embosser which renders text as tactile braille cells.
A translation table is created containing both English and braille alphabets. In this table the English alphabets are considered as old values and braille alphabets are considered as new values. When a English character is obtained an input, it is converted to braille character with reference to the translation table. For translation of numbers [Script=Braille] " is appended in the beginning of braille characters taken from the translation table to convert them to braille form.

Chapter 4

Project Design

4.1 Proposed System

The block diagram of the proposed method is shown in Fig.1. The proposed system in this paper is to recognize English text using Deep Convolution Neural Network (CNN). It is a special type of Neural Network which has given a great output in Image Processing and Computer Vision. Applications of CNN are Image Classification, Image Segmentation, Video Processing, Natural Language Processing, Speech Recognition, Object Detection, etc. The parameters used are less hence it provides high efficiency when compared to other models. The Deep Convolution Neural Network uses many feature extraction stages that learn representations automatically from the data. The Deep CNN is a multi-layered hierarchical structure that has an ability to extract features of low, mid and high levels. A Mixture of lower and mid-level features are high-level features or abstract features. In tasks like recognition as there are enormous image categories, CNN has given a better performance than conventional vision-based models.

The proposed system has six phases, the scanned input image is sent to the pre-processing stage which has three main steps, the output of which is given to the segmentation which includes three main steps. Phase three is the extraction of the segmented images followed by the fourth phase Training and Optimization. In this phase Convolution Neural Network is used to train the model and gives recognized English text as an output. Optimization using English Dictionary is done to increase the accuracy of the model. Braille Conversion is the fifth phase in which the recognized English text is converted to Braille-the language of visually impaired or blind people. The sixth phase is the conversion of the recognized text to speech to help the blind people.

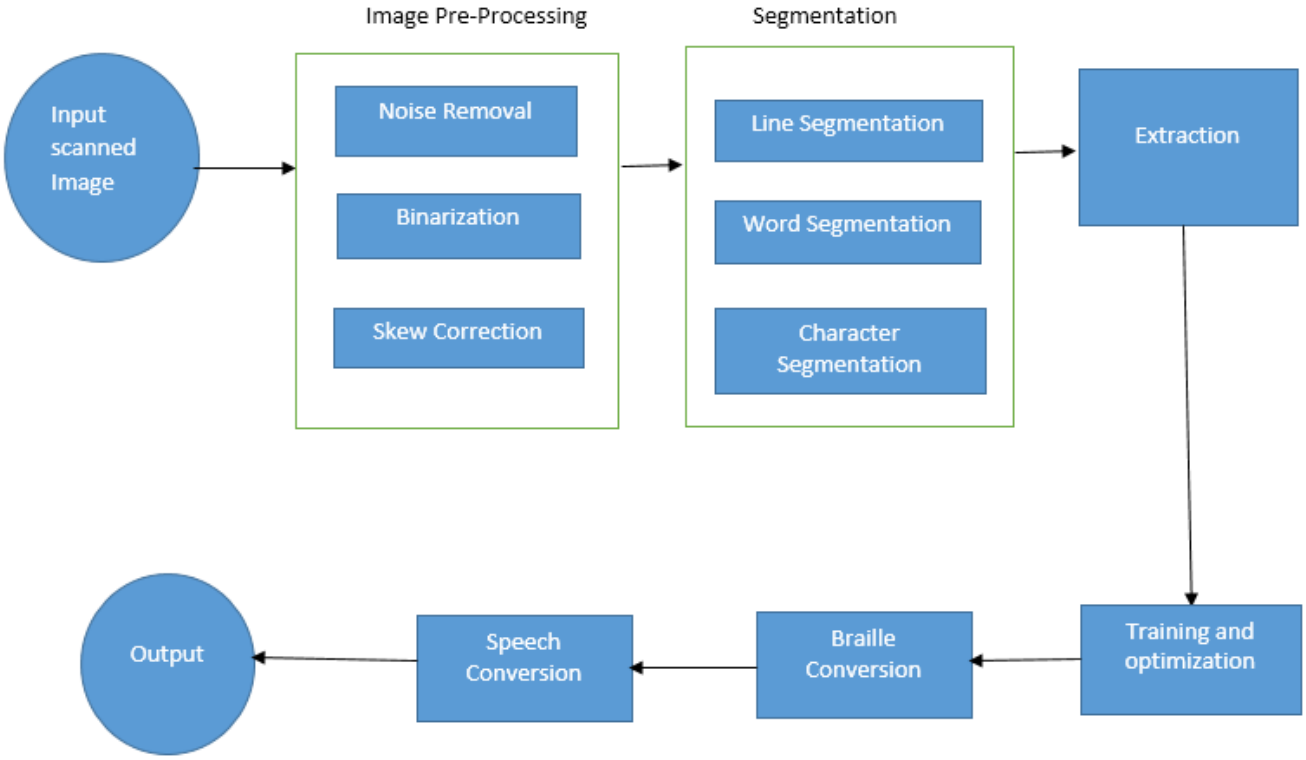


Figure 4.1: Flow Diagram of the Proposed System

4.2 Design

The Modules in the OCR Model are:
1) Input Image
2) Pre-processing
3) Segmentation
4) Extraction
5) Training
6) Optimization
7) Convert to speech
8) Translation to Braille Language

The input image is provided to the model and necessary pre-processing steps are done,which includes Re scaling the image ,increasing or decreasing the contrast of the image,increasing and decreasing the brightness,grey scaling the image,binarization of the image,noise removal,erosion and dilation,finding contours.By this step,we get a clear image all the background noise is cleared here.The output of this step is given as input to the Segmentation step where there are various levels.Firstly in a given paragraph,line segmentation is done.It includes horizontal scanning of the image, pixel-row by pixel-row from left to right and top to bottom.Next,word segmentation is done.It includes vertical scanning of the image, pixel-row by pixel-row from left to right and top to bottom.Lastly,character segmentation is done,according to the intensity the bounding box is created to each letter in the word.Then

extracting the letters and saving them where every letter is extracted.Training in which data set is given and then it recognises by matching the letters and if they are matched they are printed.Next,by using google conversion API,we convert the English letters into speech.

4.3 Activity Diagram

An activity diagram is a behavioral diagram i.e. it depicts the behavior of a system.In this case it is showing the activities needed to get from an input condition in the system to an end condition or output.

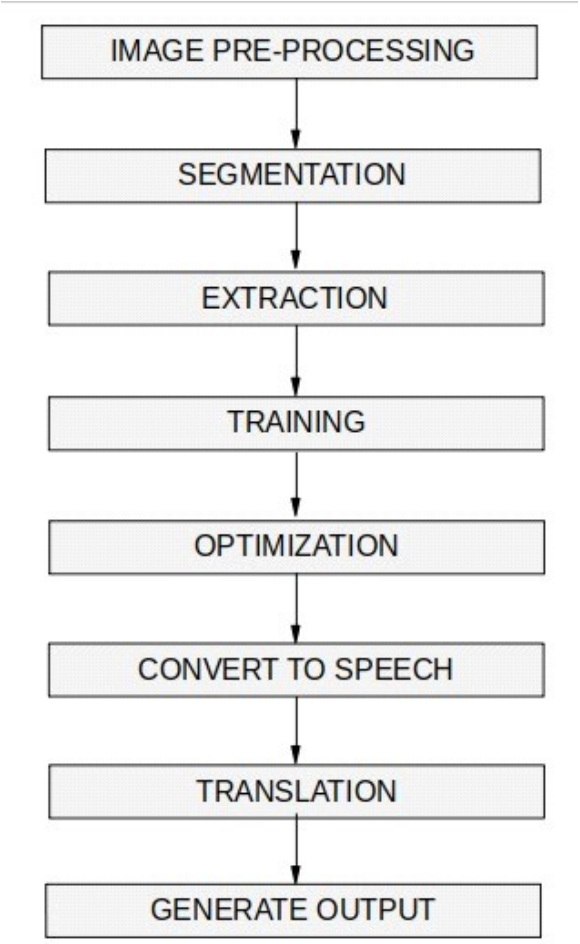


Figure 4.2: Acitivity Diagram

4.4 Use case Diagram

4.4.1 Description Of Use Case

Use cases used in this project are input image,pre-processing,segmentation, extraction. Text document undergoes segmentation whose output is given to background cleansing in this stage all the noise is removed based on the area for the accurate detection of the text line.The text line is detected and segmented,each detected line is indicated by bounding box

.It includes re scaling, increases in brightness, contrast, grey scaling, binarization.The system eliminates small text fragment in the background cleansing stage. After pre-processing step it finds all the connected documents then grouping and text line extraction and then conversion to braille language. An input image is taken, for image pre-processing-resize it accordingly and then increase or decrease the contrast and brightness of the image as per required.Text image segmentation can be achieved at three levels. It includes horizontal scanning of the image, pixel-row by pixel-row from left to right and top to bottom.It includes vertical scanning of the image, pixel-row by pixel-row from left to right and top to bottom.Character segmentation is the final level for text based image segmentation. It is similar to in operations as word segmentation.Then extracting the letters and saving them where every letter is extracted.Training in which data set is given and then it recognises by matching the letters and if they are matched they are printed.Next,by using google translate API,we translate the English letters into Braille Language. In this step the output of the previous step is taken and with the help of google conversion API conversion of English words to speech.

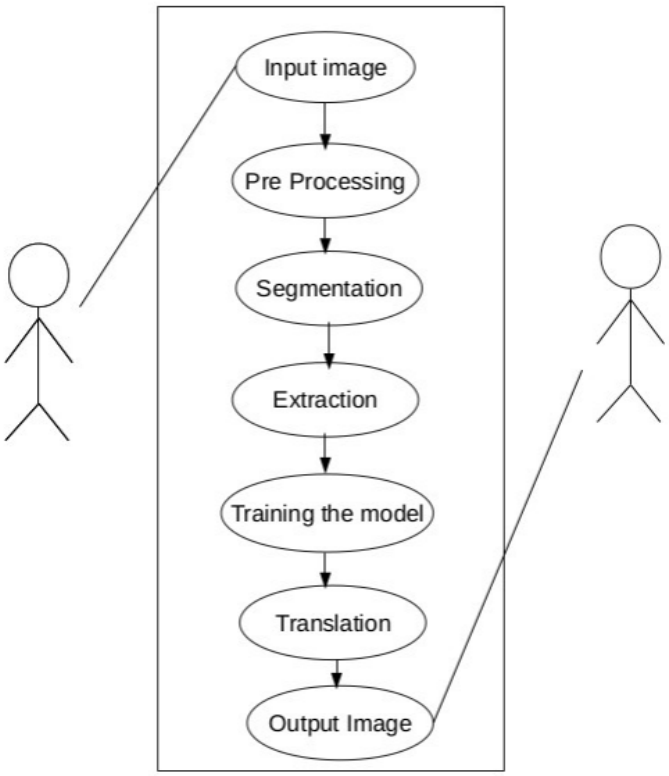


Figure 4.3: Use Case Diagram

Chapter 5

Result

A scanned image is the input to the model which undergoes various pre-processing steps making it suitable and more optimised for the model. The unwanted noise of the image is removed and skew correction of the image is done. The image is then segmented into lines followed by word and character segmentation. The segmented characters are extracted and then fed to the model for feature extraction process. The convolutional neural network has been trained with 26 English alphabets and numbers in the range 0 to 9. The model recognises the characters according to features recognised from the dataset. The characters from the documents are identified as English characters. The recognised words are converted to English speech and stored in mp3 format. These English words are also converted to Braille language for visually impaired people

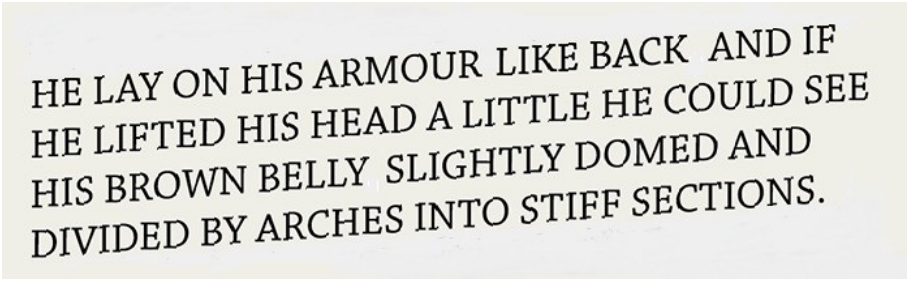


Figure 5.1: Input Image

Figure 5.2: Binarized Image

Figure 5.3: Tilt Detection

Figure 5.4: Tilt Correction

Figure 5.5: Line Segmentation

Figure 5.6: Word Segmentation

Figure 5.7: Character Segmentation

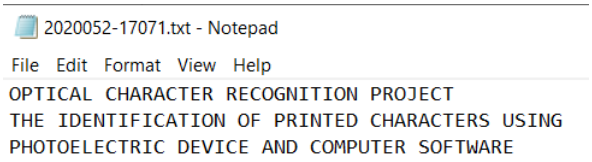


Figure 5.8: Character Recognition

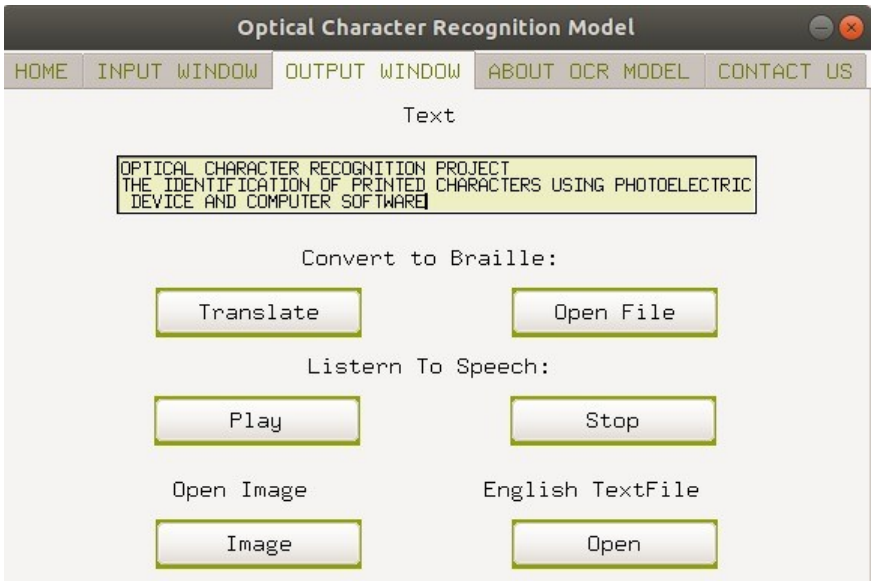


Figure 5.9: Conversion to Speech

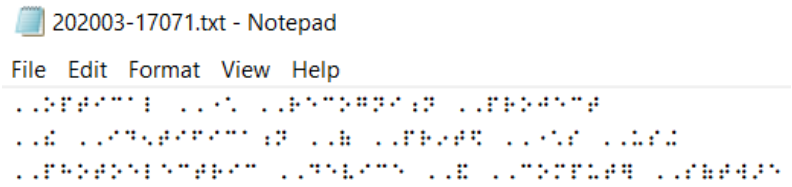


Figure 5.10: Translation to Braille

Chapter 6

Conclusions and Future Scope

This work presents a scheme recognising English characters from an image. Sometimes the images may be tilted, hence according to the region of interest the tilt correction is done. Images may contain unwanted noise in it which is removed by Gaussian blurring technique and is normalised but Otsu's normalisation technique. These image pre-processing steps helps increase the accuracy of the further steps. Segmentation of lines are done by taking horizontal projections whereas segmentation of words are done by taking vertical projections of the image. Further, character segmentation is done with the help of contours. These characters are extracted and are recognised by the trained convolutional neural network. The accuracy of line segmentation, word segmentation and character segmentation are 95%, 94% and 90% respectievely. The accuracy of the model is 80.25% and value accuracy is 90.34%.With large enough dataset more accuracy is obtained in minimal time. People who suffer from low vision, sight and visual impairment are not able to see words and letters in ordinary newsprint, books, and magazines clearly. Braille which is a tactile reading and writing system for the visually impaired people or the bind who cannot access printed materials. Hence, the text is converted to Braille which later on can be printed by braille embosser and can be used by them. Speech conversion is done for the people who are blind so that they can listen to the audio. It has less error rate and less processing time.The Future scope of this project can be using multiple languages.

Bibliography

[1] Sunanda Dixit Kanchan Keisham, "Recognition of Handwritten English Text Using Text Minimisation", Proceedings of Third International Conference INDIA 2016, vol.3, pp.607-614

[2] Xujun Peng, Srirangaraj Setlur, Venu Govindaraju, Ramachandrula Sitaram, Kiran Bhuvanagiri, "Markov Random Field Based Text Identification from Annotated Machine Printed Documents", 10th International Conference on Document Analysis and Recognition, ICDAR 2009,Barcelona, Spain, pp.431-435

[3] Karbhari V. Kale, Prapti D. Deshmukh, Shriniwas V. Chavan, Majharoddin M. Kazi, Yogesh S. Rode, "Zernike Moment Feature Extraction for Handwritten Devanagari (Marathi) Compound Character Recognition", International Journal of Advanced Research in Artificial Intelligence, 2014, vol 3,No.1,pp. 68-76

[4] Dinesh Dileep, Renu Ramesh, "A Feature Extraction Technique Based On Character Geometry for Character Recognition", arXiv February 2012.

[5] Gayathri P, Sonal Ayyappan, "Off-Line Handwritten Character Recognition Using Hidden Markov Model", IEEE, 2014, pp.518-523

[6] Parul Sahare, Sanjay B. Dhok, "Multilingual Character Segmentation and Recognition Schemes for Indian Document Images", IEEE ACCESS, 2017

[7] Tzren-Ru Chou, Fu Chang, "Optical Chinese Character Recognition for Low-Quality Document Images", IEEE, 1997, pp. 608-611

[8] Josè A. Roodriguez, Florent Perronnin, "Local Gradient Histogram Feature For Word Spotting In Unconstrained Handwritten Documents", ICFHR 2008 (International Conference on Frontiers in Handwriting Recognition), Montréal, Canada, 19-21 August, 2008

[9] Jaywant Kumar, Rohit Prasad, Huaigu Cao, Wael Abd-Almageed, David Doermann, Premkumar Natarajan, "Shape Codebook based Handwritten and Machine Printed Text Zone Extraction", SPIE Electronic Imaging, 2011, San Francisco Airport, California, United States.

[10] Fei Yin, Cheng-Lin Liu, "Handwritten Text Line Segmentation by Clustering with Distance Metric Learning", International Conference on Frontiers in Handwriting Recognition (ICFHR'08)

[11] R Sanjeev Kunte and R D Sudhaker Samuel, "A simple and efficient optical character recognition system for basic symbols in printed Kannada text", Sadhana, Vol. 32, Part 5, October 2007, pp. 521–533.

[12] Sunanda Dixit, Suresh Hosahalli Narayan and Mahesh Bellur, “kannada Text Line Extraction Based On Energy Minimization and Skew Correction “, IEEE International Advance Computing Conference, 2014, pp. 62-67

[13] Sunanda Dixit, Suresh Hosahalli Narayan ,”Segmentation of Kannada Handwritten Text Line through Computation of Variance” , IJCSIS ,Vol. 12,No. 2, February 2014 , pp. 56-60

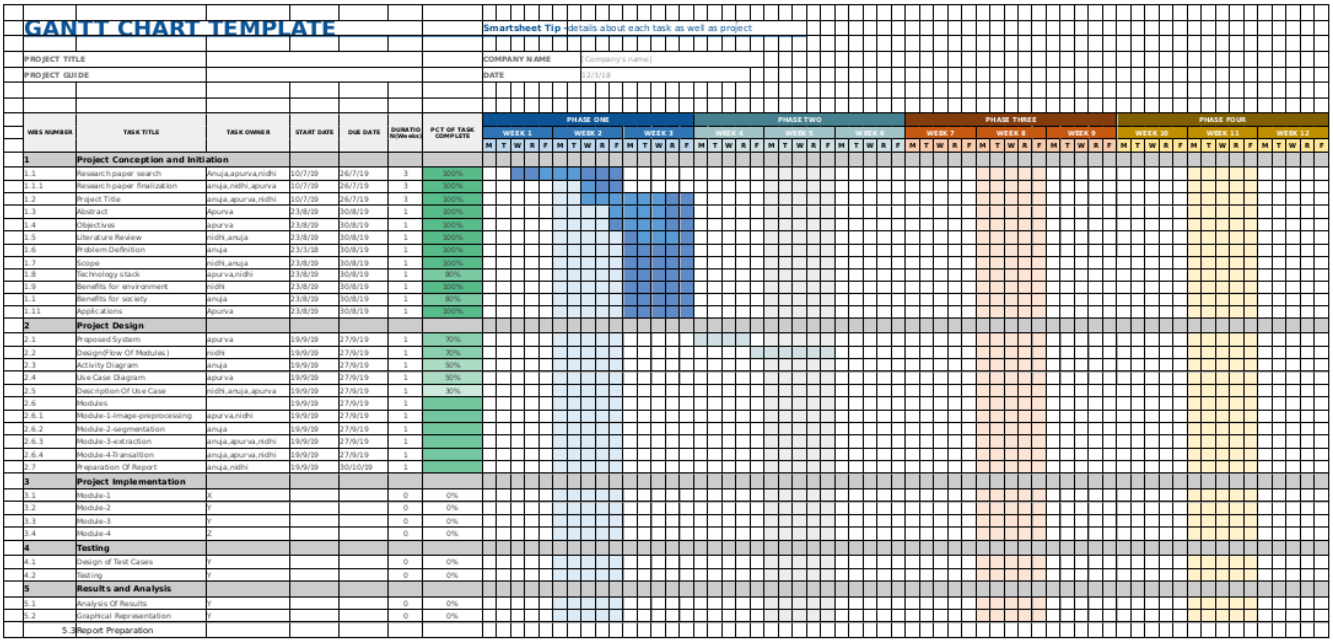
[14] Sunanda Dxit, Dr.H.N.Suresh , “South Indian Tamil Language Handwritten Document Text Line Segmentation Technique with Aid of Sliding Window and Skewing Operations”, Journal of Theoretical and Applied Information Technology, Vol.58 No.2, December 2013, pp. 430-439

[15] Sunanda Dixit, S.Ranjitha, Dr. H.N. suresh, “Segmentation of Handwritten Kannada Text Document Through Computation of standard Error and Weighted Bucket Algorithm”, Vol.3 No.2, June 2016, pp. 55-62

Chapter 7

Annexure A

7.1 Gantt chart



GANTT CHART TEMPLATE

Smartsheet Tip - details about each task as well as project

PROJECT TITLE

PROJECT GUIDE

COMPANY NAME [Company's name]

DATE 12/3/18

WBS NUMBER	TASK TITLE	TASK OWNER	START DATE	DUE DATE	DURATION (WEEKDAYS)	PCT OF TASK COMPLETE
1	**Project Conception and Initiation**					
1.1	Research paper search	Anuja,apurva,nidhi	10/7/19	26/7/19	3	100%
1.1.1	Research paper finalization	anuja,nidhi,apurva	10/7/19	26/7/19	3	100%
1.2	Project Title	anuja,apurva,nidhi	10/7/19	26/7/19	3	100%
1.3	Abstract	Apurva	23/8/19	30/8/19	1	100%
1.4	Objectives	apurva	23/8/19	30/8/19	1	100%
1.5	Literature Review	nidhi,anuja	23/8/19	30/8/19	1	100%
1.6	Problem Definition	anuja	23/3/18	30/8/19	1	100%
1.7	Scope	nidhi,anuja	23/8/19	30/8/19	1	100%
1.8	Technology stack	apurva,nidhi	23/8/19	30/8/19	1	80%
1.9	Benefits for environment	nidhi	23/8/19	30/8/19	1	100%
1.1	Benefits for society	anuja	23/8/19	30/8/19	1	80%
1.11	Applications	Apurva	23/8/19	30/8/19	1	100%
2	**Project Design**					
2.1	Proposed System	apurva	19/9/19	27/9/19	1	70%
2.2	Design(Flow Of Modules.)	nidhi	19/9/19	27/9/19	1	70%
2.3	Activity Diagram	anuja	19/9/19	27/9/19	1	50%
2.4	Use Case Diagram	apurva	19/9/19	27/9/19	1	50%
2.5	Description Of Use Case	nidhi,anuja,apurva	19/9/19	27/9/19	1	30%
2.6	Modules		19/9/19	27/9/19	1	
2.6.1	Module-1-image-preprocessing	apurva,nidhi	19/9/19	27/9/19	1	
2.6.2	Module-2-segmentation	anuja	19/9/19	27/9/19	1	
2.6.3	Module-3-extraction	anuja,apurva,nidhi	19/9/19	27/9/19	1	
2.6.4	Module-4Transaltion	anuja,apurva,nidhi	19/9/19	27/9/19	1	
2.7	Preparation Of Report	anuja,nidhi	19/9/19	30/10/19	1	
3	**Project Implementation**					
3.1	Module-1	X			0	0%
3.2	Module-2	Y			0	0%
3.3	Module-3	Y			0	0%
3.4	Module-4	Z			0	0%
4	**Testing**					
4.1	Design of Test Cases	Y			0	0%
4.2	Testing	Y			0	0%
5	**Results and Analysis**					
5.1	Analysis Of Results	Y			0	0%
5.2	Graphical Representation	Y			0	0%
5.3	Report Preparation					

PHASE ONE — WEEK 1, WEEK 2, WEEK 3

PHASE TWO — WEEK 4, WEEK 5, WEEK 6

PHASE THREE — WEEK 7, WEEK 8, WEEK 9

PHASE FOUR — WEEK 10, WEEK 11, WEEK 12

M T W R F

Figure 7.1: Gantt Chart

Acknowledgement

We have great pleasure in presenting the report on **OCR Model For Visually Impaired People Using TensorFlow.** We take this opportunity to express our sincere thanks towards our guide **Prof. Ramya RB** Department of Computer Engineering, APSIT thane for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards his constant encouragement, support and guidance through the development of project.

We thank **Prof.Sachin Malave** Head of Department,Computer, APSIT for his encouragement during progress meeting and providing guidelines to write this report.

We thank **Prof.Amol Kalugade** BE project co-ordinator, Department of Computer, APSIT for being encouraging throughout the course and for guidance.

We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

Student Name1:Nidhi Munavalli
Student ID1:16102049

Student Name2:Apurva Waingankar
Student ID2:16102050

Student Name3:Anuja Velaskar
Student ID3:16102042