e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

HANDWRITING TEXT RECOGNITION USING NEURAL NETWORK

Neelisetty Nikith[*1], Anand Sai M[*2], Kumaravel P[*3], V Gowthami[*4]

[*1,2,3]UG Scholar, ECE Department, ANNA University : MIT Campus, Chennai, Tamil Nadu, India.

[*4]Teaching Fellow, ECE Department, ANNA University : MIT Campus, Chennai, Tamil Nadu, India.

DOI : https://www.doi.org/10.56726/IRJMETS29660

ABSTRACT

This project introduces a way to recognize text or handwritten information covert into machine readable text in a form the computer can process, store and edit as a text file or as a part of a data entry and manipulation software. Then it converts the machine readable text to speech. People who suffer from low vision, sight and visual impairments, who are not able to see words and letters in ordinary newsprint, books and notes were the ones this proposal was meant to help. Moreover this project also can be capstone for research and development of artificially intelligent virtual assistants that require a glimpse into the human world for information such as street signs and receipts. Using a combination of Convolution neural networks and Recurrent Neural Networks the project successfully came about to recognize text and hand written information. Another deep learning network, WaveNet was employed to convert the text which is recognized from the handwriting to covert into a raw audio file. Hence using this complete system one can pass an image containing text, have it recognized by the computer and also be read out to the user.

I. INTRODUCTION

OVERVIEW

Handwriting is one thing that is unique to an individual. Handwriting Recognition is gaining importance in various fields eg: Authentication of signature in banks, forensic evidences, ZIP code addresses on letters. System will be suitable for converting handwritten documents into structural text form and recognizing handwritten names

A. IMAGE PROCESSING

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Image processing includes the following three steps Importing the image via image acquisition tools; Analyzing and manipulating the image, Output in which result can be altered image or report that is based on image analysis The two types of methods used for image processing are, analog and digital image processing. Analog image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. Digital image processing techniques help in manipulation of the digital images by using computers. The three general phases that all types of data have to undergo while using digital technique are pre-processing, enhancement, and display, information extraction.

B. IMAGE FILE FORMATS

Image file formats are standardized means of organizing and storing digital images. An image file format may store data in an uncompressed format, a compressed format (which may be lossless or lossy), or a vector format. Image files are composed of digital data in one of these formats so that the data can be rasterized for use on a computer display or printer. Rasterization converts the image data into a grid of pixels. Each pixel has a number of bits to designate its color. The PNG (Portable Network Graphics), JPEG (Joint Photographic Experts Group), and GIF (Graphics Interchange Format) formats are most often used to display images.

C. OPENCV

OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. The library has more than 2500

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. Python is a general purpose programming language started by Guido van Rossum, which became very popular in short time mainly because of its simplicity and code readability. It enables the programmer to express his ideas in fewer lines of code without reducing any readability. How OpenCV-Python works is a Python wrapper around original C++ implementation. And the support of Numpy makes the task more easier. Numpy is a highly optimized library for numerical operations. It gives a MATLAB - style syntax. All the OpenCV array structures are converted to-and-from Numpy arrays. So whatever operations you can do in Numpy, you can combine it with OpenCV, which increases number of weapons in your arsenal. Besides that, several other libraries like SciPy, Matplotlib which supports Numpy can be used with this. OpenCV-Python is an appropriate tool for fast prototyping of computer vision problems. OpenCV 3.4.0 and python 2.7.10 are the versions used

D. KERAS

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. It focuses on being user-friendly, modular, and extensible. Keras supports both convolutional networks and recurrent networks, as well as combinations of the, It runs seamlessly on CPU and GPU. New models are easy to create and existing models available on eras provide ample examples . This feature supports Keras as a tool used in advanced research

E. TENSOR FLOW

TensorFlow is a free and open-source software library for dataflow 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. TensorFlow can run on multiple CPUs and GPUs (with optional CUDA and SYCL extensions for general-purpose computing on graphics processing units). TensorFlow is available on 64-bit Linux, macOS, Windows, and mobile computing platforms including Android and iOS. Its flexible architecture allows for the easy deployment of computation across a variety of platforms such as CPUs (Central Processing Unit), GPUs (Graphical processing units), TPUs (Tensor Processing unit), and from desktops to clusters of servers to mobile and edge devices. TensorFlow computations are expressed as stateful dataflow graphs. The name TensorFlow derives from the operations that such neural networks perform on multidimensional data arrays, which are referred to as tensors

II. LITERATURE SURVEY

2.1. TTS SPEECH CODING A. Acero, "An overview of text-to-speech synthesis," 2000 IEEE Workshop on Speech Coding. Proceedings. Meeting the Challenges of the New Millennium (Cat. No.00EX421), Delavan, WI, USA, 2000, pp. 1-, doi: 10.1109/SCFT.2000.878372.

The article gives an overview of text-to-speech (TTS) technology and a description of some issues of potential interest to speech coding experts. After motivation for the use of TTS technology, it describes the general architecture of a text-to-speech system with particular emphasis on the speech synthesis component. Both formant synthesis and concatenative synthesis are presented, offering different degrees of flexibility and quality. Several well-known speech coding techniques (including LPC vocoders, waveform interpolation, harmonic coding, and layered coding) have been used in speech synthesis. It explains how they have been applied, and the advantages and limitations of those techniques when used in speech synthesis. The main goal is to increase cooperation between the speech coding community and the TTS community, and in particular to motivate the need for speech coding algorithms that meet the requirements of the next generation speech synthesis technology.

2.2. TTS FOR INDIAN ENGLISH D. Mahanta, B. Sharma, P. Sarmah and S. R. M. Prasanna, "Text to speech synthesis system in Indian English," 2016 IEEE Region 10 Conference (TENCON), Singapore, 2016, pp. 2614-2618, doi: 10.1109/TENCON.2016.7848511.

In this work, an effort has been made to modify the existing English grapheme to phoneme dictionary by implementing specific rules for one particular variety of Indian English, namely Assamese English. The proposed method of dictionary modification is applied at the front end of the Indian English TTS, developed using unit selection synthesis and statistical parametric speech synthesis frameworks. In both frameworks, significant improvement is achieved in subjective evaluation when the dictionary is adapted to Assamese English pronunciation. The word error rate decreased from 46.67% to 7.69% after incorporating the variety specific modifications to the dictionary, indicating significant perceptual improvement.

2.3. SPATIOSPECTRAL CONCENTRATION Frederik J. Simons · Dong V. Wang, " Spatiospectral concentration in the Cartesian plane", Department of Statistics and Operations Research, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, U.S.A, 2011

The paper solves analogue of slepians time frequency concentration problem in the 2D plane. The orthogonal family of strictly band limited functions functions that are optimally concentrated within a closed region of the plane. the concentration domains are circularly symmetric in both spaces, the Slepian functions are also eigen functions of a Sturm – Liouville operator, leading to special algorithms are discussed in the paper.

2.4. DIAGONAL BASED FEATURE EXTRACTION J.Pradeep , E.Srinivasan and S.Himavathi, "Diagonal based feature extraction for handwritten alphabets recognition system using neural network" , International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 1, 2010

The authors train a neural network for 26 alphabets and 570 different handwritten alphabetical characters are used for testing. The authors then implement diagonal based feature extraction method as opposed to conventional horizontal and vertical methods for feature extraction.

2.5. TTS SYNTHESIZER FOR ENGLISH P. Jayawardhana, A. Aponso, N. Krishnarajah and A. Rathnayake, "An Intelligent Approach of Text-To-Speech Synthesizers for English and Sinhala Languages," 2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT), Kahului, HI, USA, 2019, pp. 229-234, doi: 10.1109/INFOCT.2019.8711051.

This paper attempts to investigate novel Text-to-Speech algorithm based on Deep voice which is an attention based, fully convolutional mechanism. The procedure of producing speech synthesis involves with learning statistical model of the human vocal production mechanism which is eligible of taking some text and vocalize that as speech. This paper would reveal the route of the attempt where there is the destination of accuracy and realism. Serenity and fluency are the most important qualities which expect from a TTS. The idea is to give an outline of discourse amalgamation in the Sinhala language, compresses and replicates about the characteristics of different blend procedures utilized. The proposed TTS synthesizing with the neural network based approach to perform phonetic-to-acoustic mapping has described by the purpose of applying for multilingual synthesizers.

2.6. ONLINE AND OFFLINE HR Plamondon, Réjean, and Sargur N. Srihari. "Online and offline handwriting recognition: a comprehensive survey." Pattern Analysis and Machine Intelligence, IEEE Transactions on 22.1 (2000): 63-84

The authors gives an overview of the nature of handwritten language and how it is transduced into electronic data. Both the online and offline cases are considered in the paper. Algorithms for pre-processing and word recognition and performance with practical systems are also mentioned.

2.7. TEXT TO SPEECH SYNTHESIZER S. Lukose and S. S. Upadhya, "Text to speech synthesizer-formant synthesis," 2017 International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, 2017, pp. 1-4, doi: 10.1109/ICNTE.2017.7947945.

In the paper, different methods of text to speech synthesizer techniques are discussed to produce intelligible and natural output and a vowel synthesizer using cascade formant technique is implemented. A text to speech output is based on generating corresponding sound output when the text is inputted. Wide range of applications use text to speech technique in medicals, telecommunications fields, etc. The Various speech synthesis methods that have been used for text to speech output for obtaining intelligible and natural output are Concatenative, Formant, Articulatory, Hidden Markov model (HMM).

2.8. A SCALABLE HTR SYSTEM R. Reeve Ingle, Yasuhisa Fujii, Thomas Deselaers, Jonathan Baccash, Ashok C. Popat, "A Scalable Handwritten Text Recognition System". Google Research Mountain View, CA 94043, 2019

The authors develop an offline Handwriting recognition system for line recognition, discussing a problem based with large scale multilingual OCR system. describe our image data generation pipeline and study how online data can be used to build HTR models. Secondly, we propose a line recognition model based on neural networks with recurrent connections. The model achieves a comparable accuracy with LSTM - based models while allowing for better parallelism in training and inference.

2.9. DATASET FOR ENGLISH TR Saad Bin Ahmed ; Saeeda Naz ; Muhammad Imran Razzak, "A Novel Dataset for English-Arabic Scene Text Recognition (EASTR)-42K and Its Evaluation Using Invariant Feature Extraction on Detected Extremal Regions", IEEE Access (Volume: 7), pp: 19801 - 19820, 13 February 2019

The authors, employ text segmentation and recognition task for English-Arabic scene text recognition. his paper presents a novel technique by using adapted maximally stable extremal region (MSER) technique and extracts scale-invariant features from MSER detected region. To select discriminant and comprehensive features, the size of invariant features is restricted and considered those specific features which exist in the extremal region. The adapted MDLSTM network is presented to tackle the complexities of cursive scene text

2.10. WAVENET FOR LOSSLESS SPEECH CODING T. Yoshimura, K. Hashimoto, K. Oura, Y. Nankaku and K. Tokuda, "WaveNet- Based Zero-Delay Lossless Speech Coding," 2018 IEEE Spoken Language Technology Workshop (SLT), Athens, Greece, 2018, pp. 153-158, doi: 10.1109/SLT.2018.8639598.

This paper presents a WaveNet-based zero-delay lossless speech coding technique for high-quality communications. The WaveNet generative model, which is a state-of-the-art model for neural-network-based speech waveform synthesis, is used in both the encoder and decoder. In the encoder, discrete speech signals are losslessly compressed using sample-by-sample entropy coding. The decoder fully reconstructs the original speech signals from the compressed signals without algorithmic delay. Experimental results show that the proposed coding technique can transmit speech audio waveforms with 50% their original bit rate and the WaveNet- based speech coder remains effective for unknown speakers.

III. SCOPE AND PROPOSED MODEL

To build a system to transform a two-dimensional image of text, that could contain machine printed or handwritten text from its image representation into machine- readable text and to then use that to convert the text to speech, for which the project started with by training a network for making predictions of the the text from any given image. The proposed engine has a pipeline-based architecture consisting of the following sequential steps: Pre-processing providing a binary threshold, determining the connected components and connections between them, character recognition and character aggregation to form words, lines, paragraphs and finally solving the problem of detecting small capitals. Before coming up with the pipeline, it was important to understand and study the history and different aspects of this technology, which gives an overview of the nature of handwritten language and how it is transduced into electronic data. From there we begin our project by concentrating on offline text recognition, which takes into account typed and handwritten text. Text recognition (HTR) system implemented with TensorFlow (TF) and trained on the IAM off-line HTR dataset using a neural network (NN) model recognizes the text contained in the images of segmented words.

Deep learning extracts features with a deep neural networks and classify itself. Compared to traditional algorithms it performance increase with amount of data. Python 3, TensorFlow 1.3, Numpy and OpenCV installed. Neural network for our task. It consists of convolutional neural network (CNN) layers, recurrent neural network (RNN) layers and a final Connectionist temporal classification (CTC) layer.

Besides the two decoders shipped with TF, it is possible to use word beam search decoding. Using this decoder, words are constrained to those contained in a dictionary, but arbitrary non-word character strings (numbers, punctuation marks) can still be recognized. The following illustration shows a sample for which word beam search is able to recognize the correct text, while the other decoders fail.

After this we obtain machine readable text from the image, the next part of the project was to get speech information from the text. For this we employed a DNN (Deep Neural Network) based approach. WaveNet is a deep neural network for generating raw audio, was used for this purpose. The network is able to generate

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

relatively realistic-sounding human-like voices by directly modelling wave forms using a neural network method trained with recordings of real speech.

A user can now, take a picture containing text or handwritten data, which is then passed through the recognition engine to get machine understandable text, which is then passed through the text to speech synthesizer to get audio information of the handwritten text. Using this engine a visually disabled and challenged people like those who are blind or have any other vision impairments can would be able to gain access to text and handwritten documents, which would have they would be unable to access prior. The engine can also be used for further research and development of artificially intelligent assistants by enabling them to become more smarter in accessing data of the human world such as street signs or bills.Multichannel

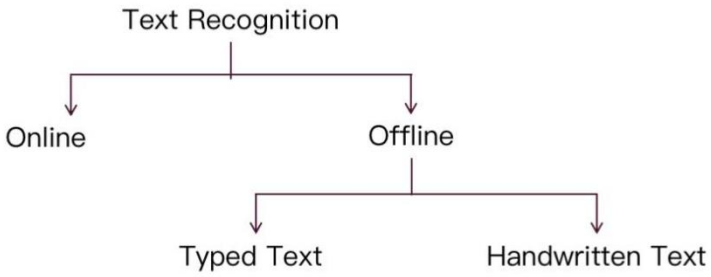
IV. SYSTEM ARCHITECTURE

4.1. INTRODUCTION

The aim is to build a system to transform a two-dimensional image of text, that could contain machine printed or handwritten text from its image representation into machine-readable text. This process generally consists of several sub-processes to perform as accurately as possible.

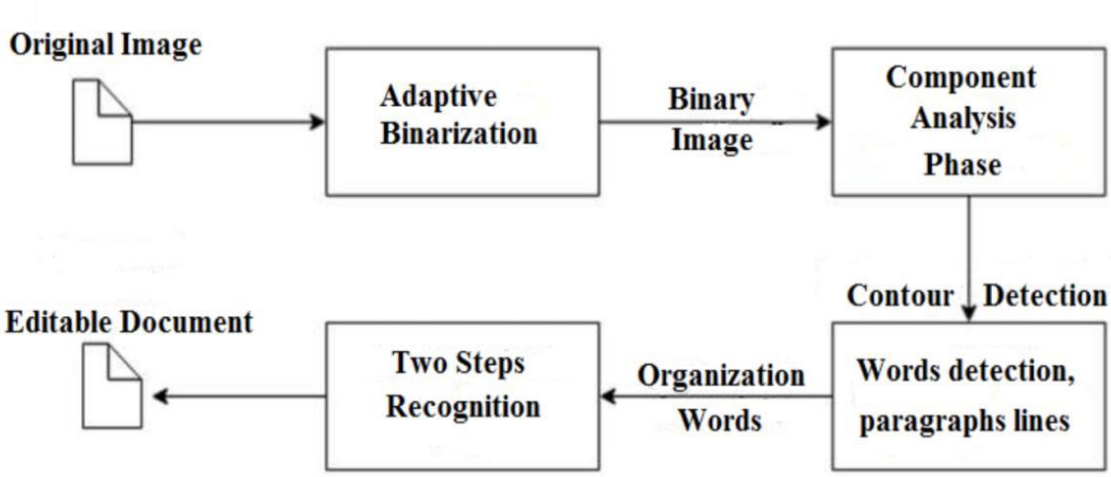
The sub-processes are:

- Pre-processing of the Image
- Text Localization
- Character Segmentation
- Character Recognition
- Post Processing



4.2. TYPED TEXT RECOGNITION

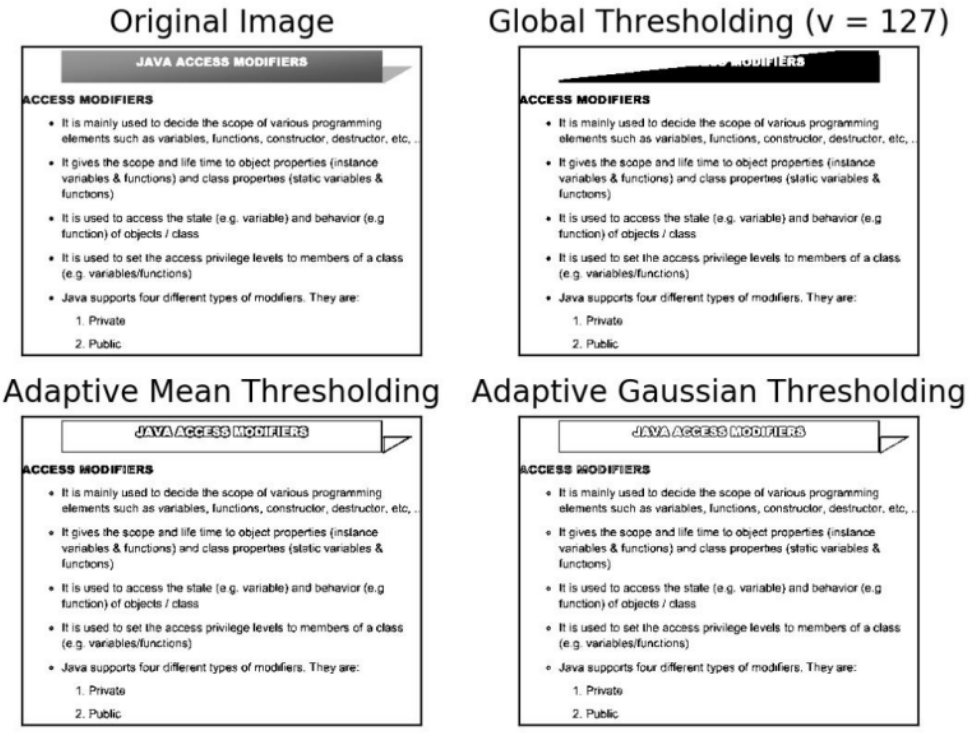
The Character Recognition Engine (CRE) has a pipeline-based architecture presented in Figure 4.1. It consists of the following sequential steps: preprocessing providing a binary threshold, determining the connected components and connections between them (also storing them in objects called blobs), character recognition and character aggregation to form words, lines, paragraphs and finally solving the problem of detecting small capitals.



e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

4.2.1. BINARIZATION

Digital images are composed of pixels, however the range of colors that can be displayed depends on the number of bits used to represent each such pixel (BPP or Bits-Per-Pixel). For example, a binary image having a BPP of one and a single component means that the image will be represented using only two colors: black and white, one color for each possible value of the binary representation. The thresholding operation is the processing stage that takes as input an image having a different representation and converts it to a black and white image and based on Architecture of the CRE determining a computer threshold, hence the name. The thresholding step is essential for an character recognition because the analyzed picture becomes easier to process and the background noise is largely reduced being lower than the threshold limit and thus removed.



Adaptive Thresholding with gaussian weighted average is done using the function cv2.adaptiveThreshold. The signature of the function is - cv2.adaptive Threshold (src, maxValue, adaptive Method, threshold Type, blockSize, C[, dst]) **->** dst

4.2.2. IMAGE SEGMENTATION

Thresholding simplifies the input image transforming it into black and white, it cannot identify the elements of an image. Segmentation is the process of identifying the objects in an image based on certain properties like pixel color, intensity, texture. Typically, segmentation creates a mask image consisting of input pixels belonging to a zone of interest (an object image) of a certain color and/or property. Because a normal image that would be processed through a character recognition engine can contain text, graphics and complex layouts, the goal of segmentation is the identification of the areas of interest as well as the evaluation of their type. This image segmentation step is particularly useful in the detection of lines or other layout separators. Any method of segmentation must meet the following criteria: the recombination of all regions or segments, must reconstruct the original image (i.e. segmentation must be complete) and the regions must be disjointed to avoid duplication and different from each other in the sense that each pixel region only groups based on fixed conditions. Segmentation may be performed using multiple approaches like histogram analysis region growing, watershed or even a voting-based segmentation. The different approaches used in the project

REGION GROWING BASED SEGMENTATION

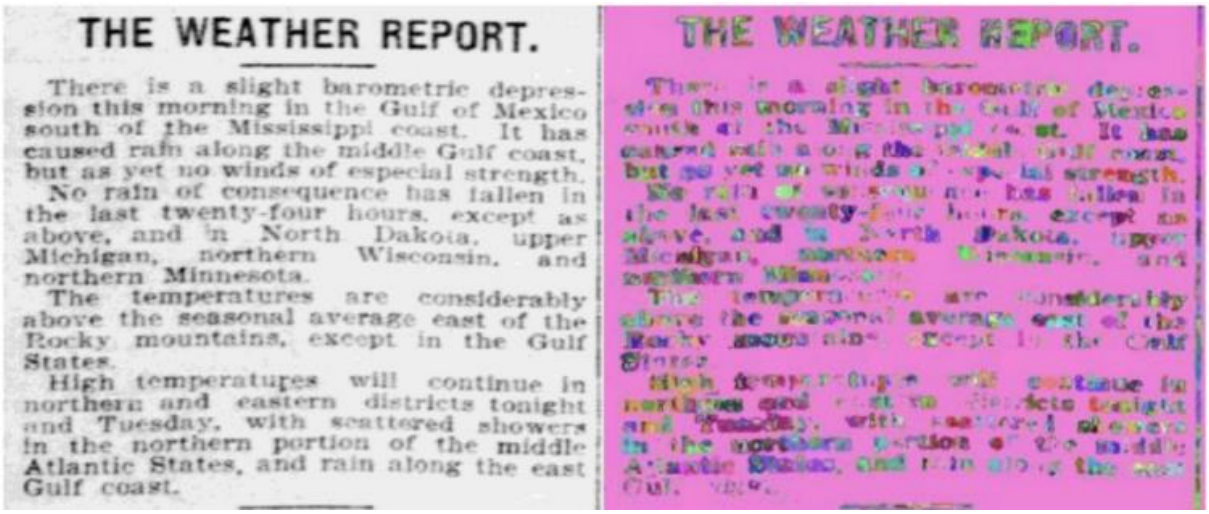
GRAPH BASED SEGMENTATION

WATERSHED BASED SEGMENTATION

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

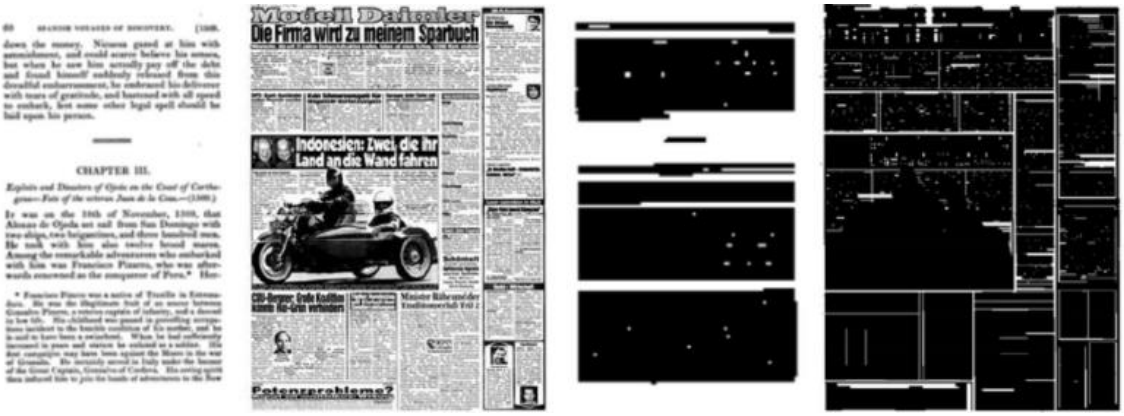
VOTING BASED SEGMENTATION

DOCUMENT IMAGE LAYOUT ANALYSIS (DILA)



The Original Image And Output Of DILA Algorithm

Through a process called dilation, the elements of the resulting image are thinned, making the background image bolder and sometimes causing darker elements to be separated. Erosion is the reverse process by which elements of the resulting image contours are thicker and the background image is thinned and sometimes brighter elements are combined.



Different Stage Outputs Of DILA Algorithm

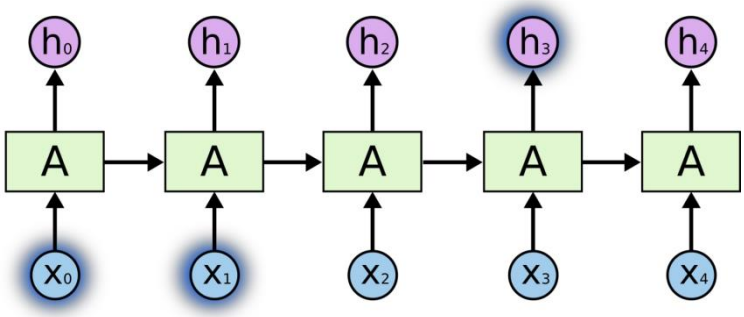
4.3 LONG SHORT-TERM MEMORY (LSTM)

Long short-term memory is an artificial recurrent neural network architecture used in the field of deep learning. Unlike standard feed - forward neural networks, LSTM has feedback connections. It can not only process single data points, but also entire sequences of data

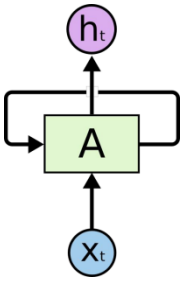
RECURRENT NEURAL NETWORKS

A recurrent neural network (RNN) is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic behavior. Derived from feed - forward neural networks, RNNs can use their internal state (memory) to process variable length sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com



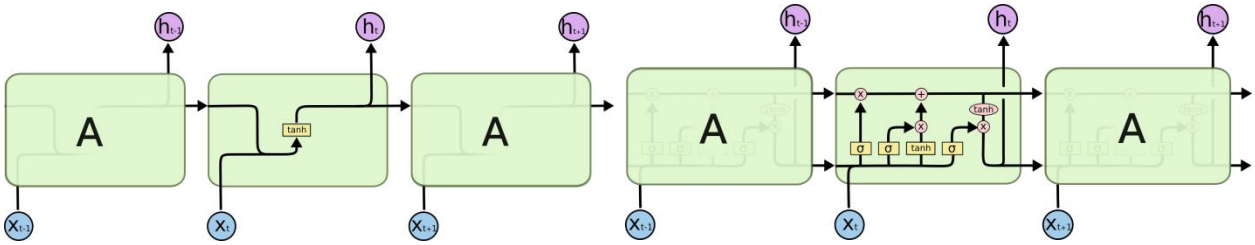
A Typical RNN Network



A section of Neural Network

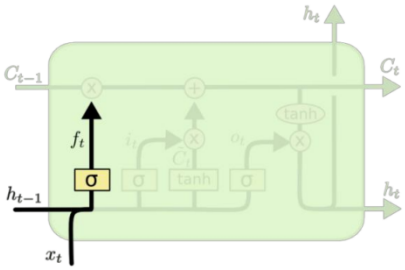
LSTM NETWORKS

Long Short Term Memory networks abbreviated to LSTMs. LSTMs are special kind of RNN, capable of learning long-term dependencies. They work tremendously on a large variety of problems. LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn. All recurrent neural networks have the form of a chain of repeating modules of neural network. In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer.

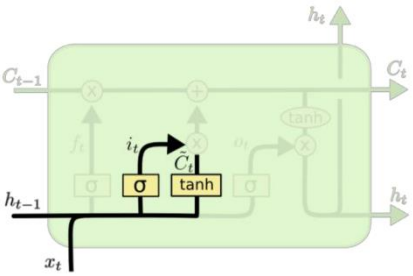


A Standard RNN Network A Standard LSTM Network

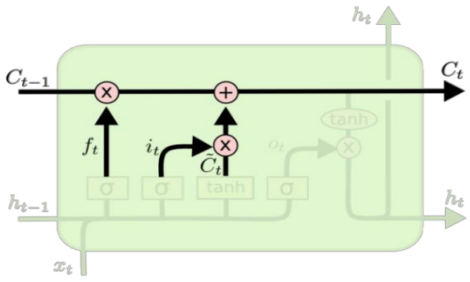
PHASED APPROACH OF LSTM



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

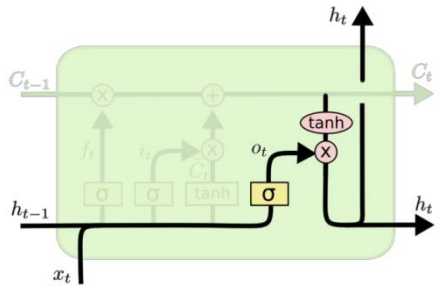


$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com



$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t]\ +\ b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

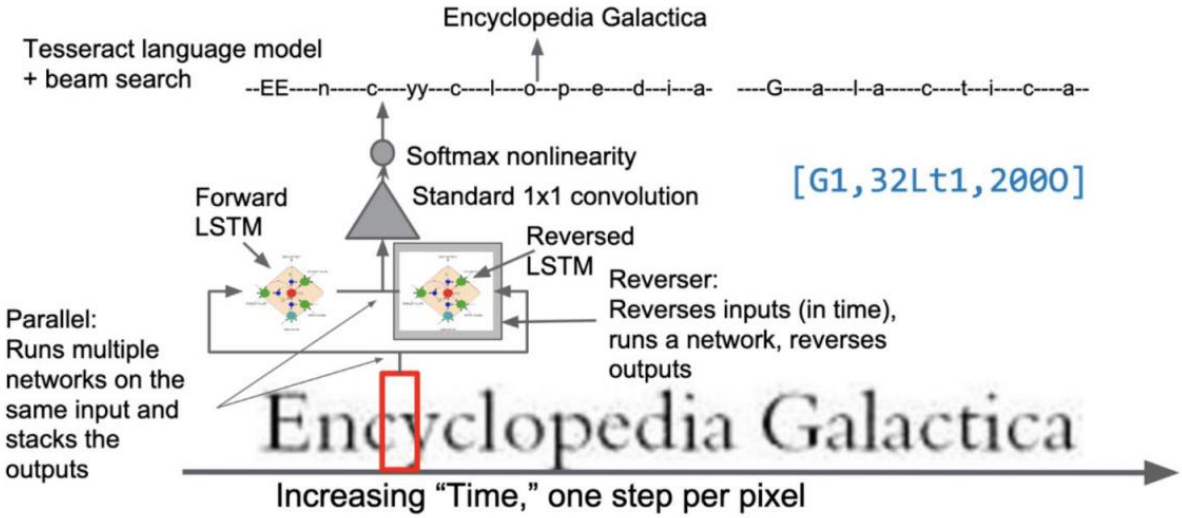
The engine is dependant on the multistage process where we can differentiate steps:

- Word finding
- Line finding
- Character classification

Word finding was done by organizing text lines into blobs, and the lines and regions are analyzed for fixed pitch or proportional text. Text lines are broken into words differently according to the kind of character spacing. Recognition then proceeds as a two-pass process. In the first pass, an attempt is made to recognize each word in turn.

Each word that is satisfactory is passed to an adaptive classifier as training data. The adaptive classifier then gets a chance to more accurately recognize text lower down the page.

Modernization of the Tesseract tool was an effort on code cleaning and adding a new LSTM model. The input image is processed in boxes (rectangle) line by line feeding into the LSTM model and giving output. In the figure below we can visualize how it works



4.3. HANDWRITTEN TEXT RECOGNITION

Handwriting is the writing down with a writing instrument (such as a pen or pencil) in the hand. Handwriting includes both printing and cursive styles and is separate from formal calligraphy or typeface. The handwriting of each person is unique and the handwriting can be used to verify the author of a document. Each human has a unique style of handwriting, irrespective of his regular day handwriting or personal signature. Even identical twins who share appearance and genetics do not have the same handwriting. The place where a human grows up and 33the first language the human learns, the humans style of handwriting is created with the different distribution of force and ways of shaping words. Because handwriting is relatively stable, a change in the handwriting can be indicative of the nervousness or intoxication of the author. A sample of a person writing can be used to determine/authenticate a document, if the document author is denying. The sample of accused writing can be used to compare to that of a written document to determine and authenticate the written document's author, if the writing styles is 100% match, it is likely that accused written both documents.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

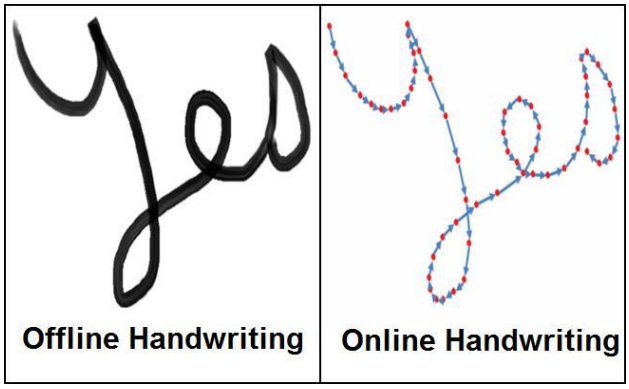
Handwriting is historically considered the widest taught motor skill. Handwriting is also one of the first, and often the only motor skill that children will learn at elementary school. It takes years of practice and maturing before a person has mastered the adult handwriting skill. Handwriting is not considered only as a movement that leaves a visible trace of ink on paper (product) but it can also be considered as a movement (process).

Characteristics of handwriting include:

- Specific shape of letters, For example, The roundness or sharpness of a letter
- Space in between the letters
- The slope of the sentence
- The rhythmic repetition of the elements or Arrhythmia
- The pressure to the paper
- The average size of letters
- The thickness of letters

CHALLENGES IN HTR

Huge variability and ambiguity of strokes from person to person. Handwriting style of an individual person also varies time to time and is inconsistent. Poor quality of the source document/image due to degradation over time. Text in printed documents sit in a straight line whereas humans need not write a line of text in a straight line on white paper. Cursive handwriting makes separation and recognition of characters challenging. Text in handwriting can have variable rotation to the right which is in contrast to printed text where all the text sits up straight. Collecting a good labelled dataset to learn is not cheap compared to synthetic data Methods Handwriting text Recognition methods can be broadly classified into two types



METRICS

There are two metrics for evaluating any text recognition module

1. Character Error Rate :- It is computed as the Levenshtein distance which is the sum of the character substitutions (Sc), insertions (Ic) and deletions (Dc) that are needed to transform one string into the other, divided by the total number of characters in the groundtruth (Nc)

$$CER = \frac{S_c + I_c + D_c}{N_c}$$

2. Word Error Rate :- It is computed as the sum of the word substitutions (Sw), insertions (Iw) and deletions (Dw) that are required to transform one string into the other, divided by the total number of words in the groundtruth (Nw)

$$WER = \frac{S_w + I_w + D_w}{N_w}$$

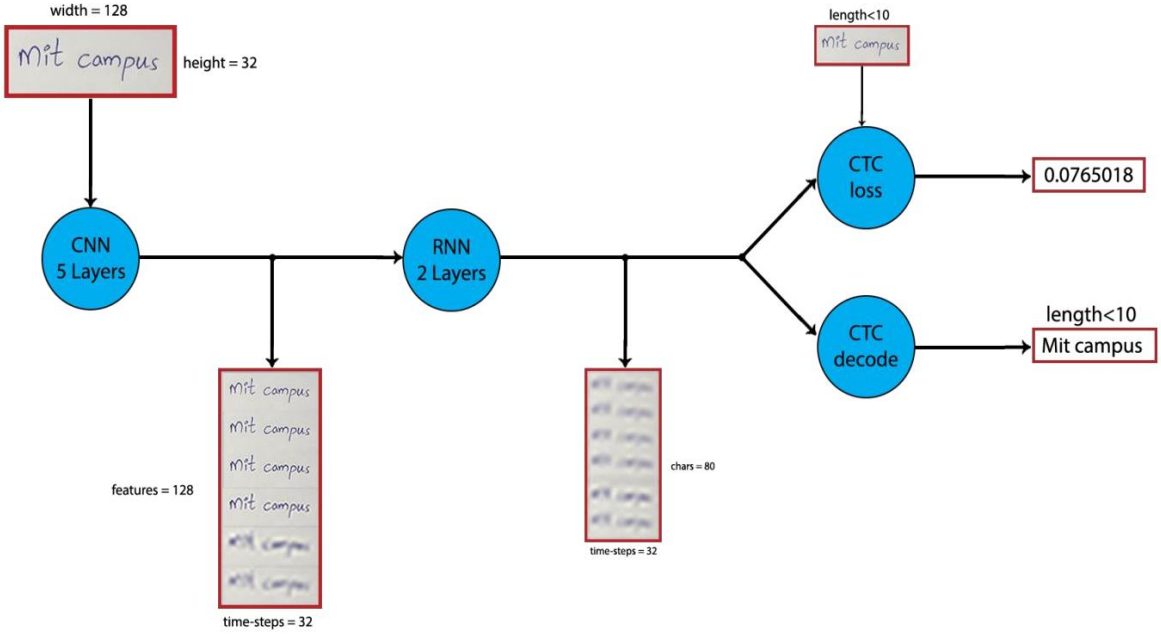
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

V. SYSTEM IMPLEMENTATION

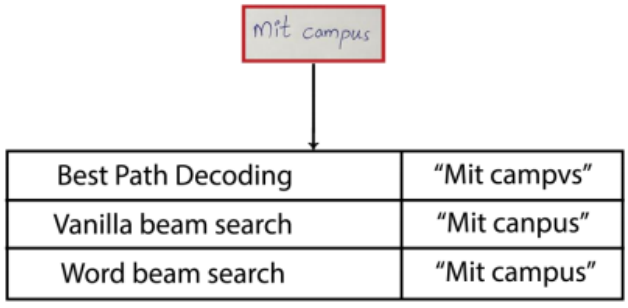
INTRODUCTION

Handwritten Text Recognition (HTR) system implemented with TensorFlow (TF) and trained on the IAM off-line HTR dataset. This Neural Network (NN) model recognizes the text contained in the images of segmented words as shown in the illustration below. As these word-images are smaller than images of complete text- lines, the NN can be kept small and training on the CPU is feasible. 3/4 of the words from the validation-set are correctly recognized and the character error rate is around 10%.

Expected outcome of HTR module



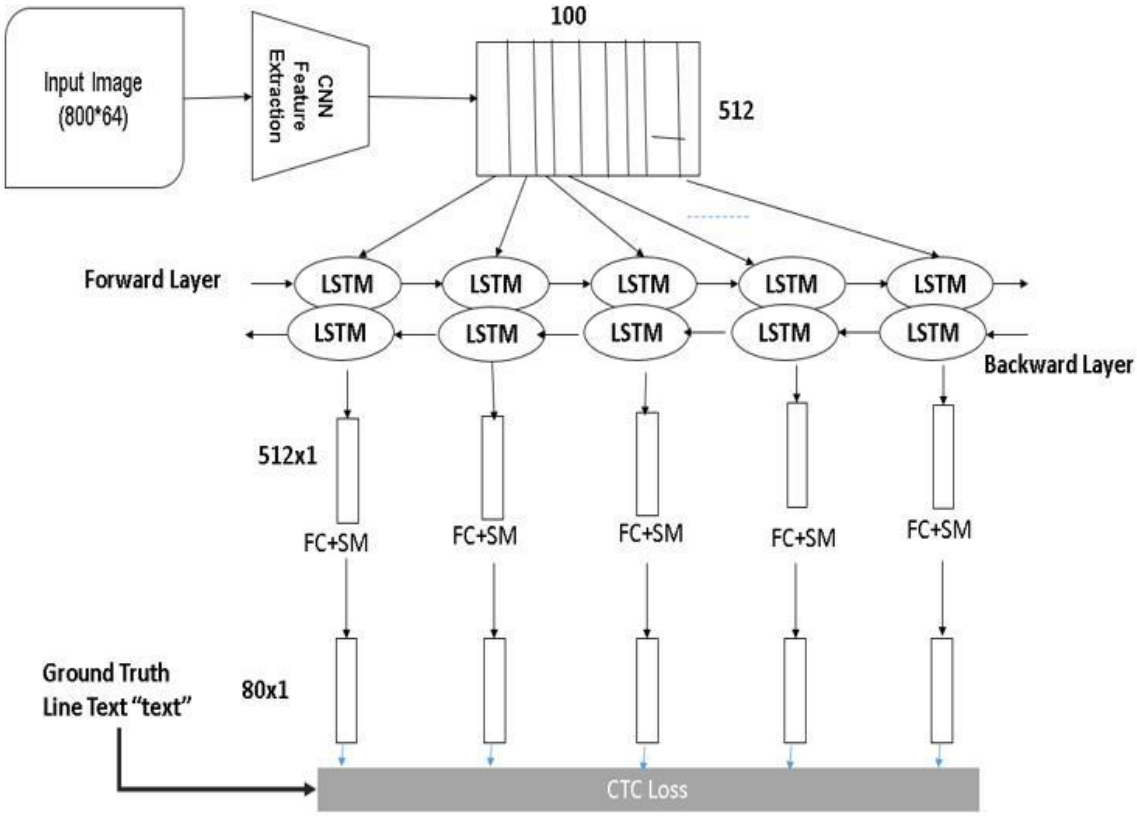
Overview of HTR system



Best Path Decoding	"Mit campvs"
Vanilla beam search	"Mit canpus"
Word beam search	"Mit campus"

Outputs of Different CTC Decoding Techniques

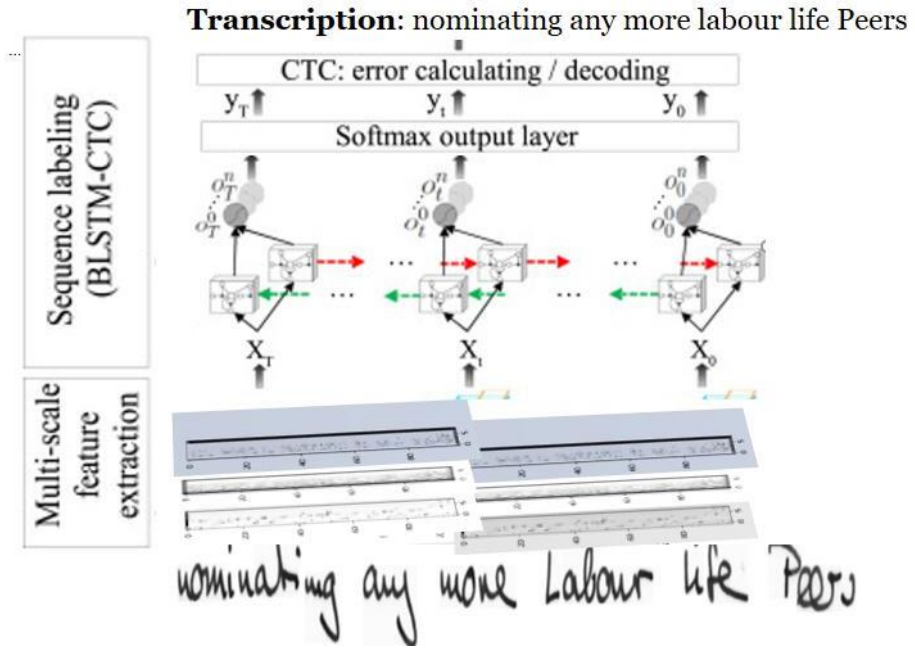
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com



Work flow in HTR system

Project consists of Three steps:

- Data is extracted form Multi-scale feature Extraction block and sent to Convolutional Neural Network 7 Layers
- Data from CNN is sent to Sequence Labeling (BLSTM-CTC) and then transfered to Recurrent Neural Network (2 layers of LSTM) with CTC
- From CTC the text is converted into Transcription and the Decoding the output of the RNN (CTC decode)

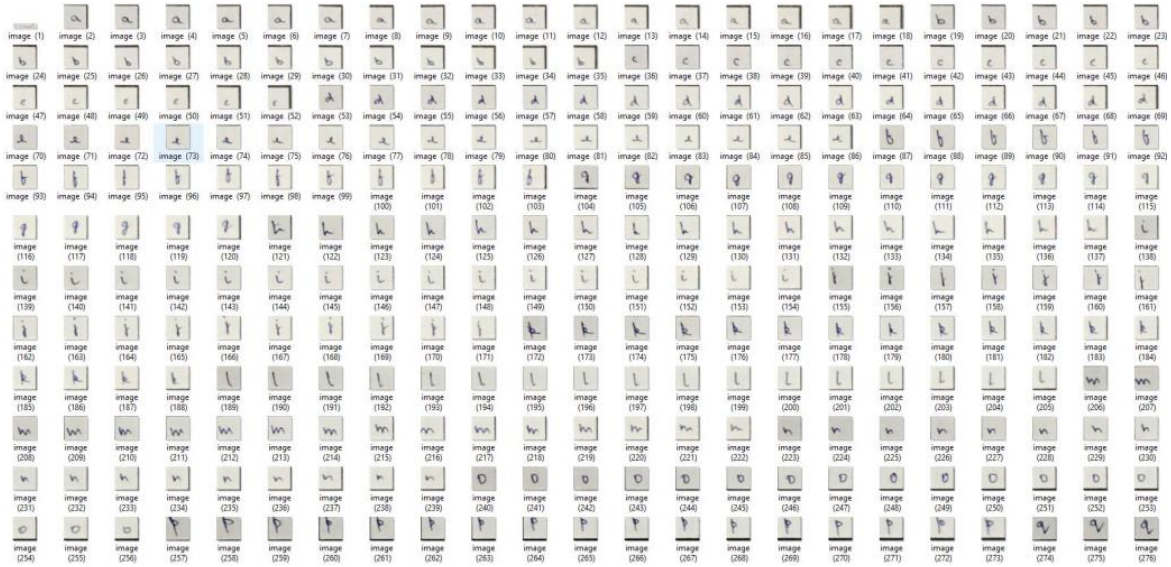


Schematic of HTR system

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

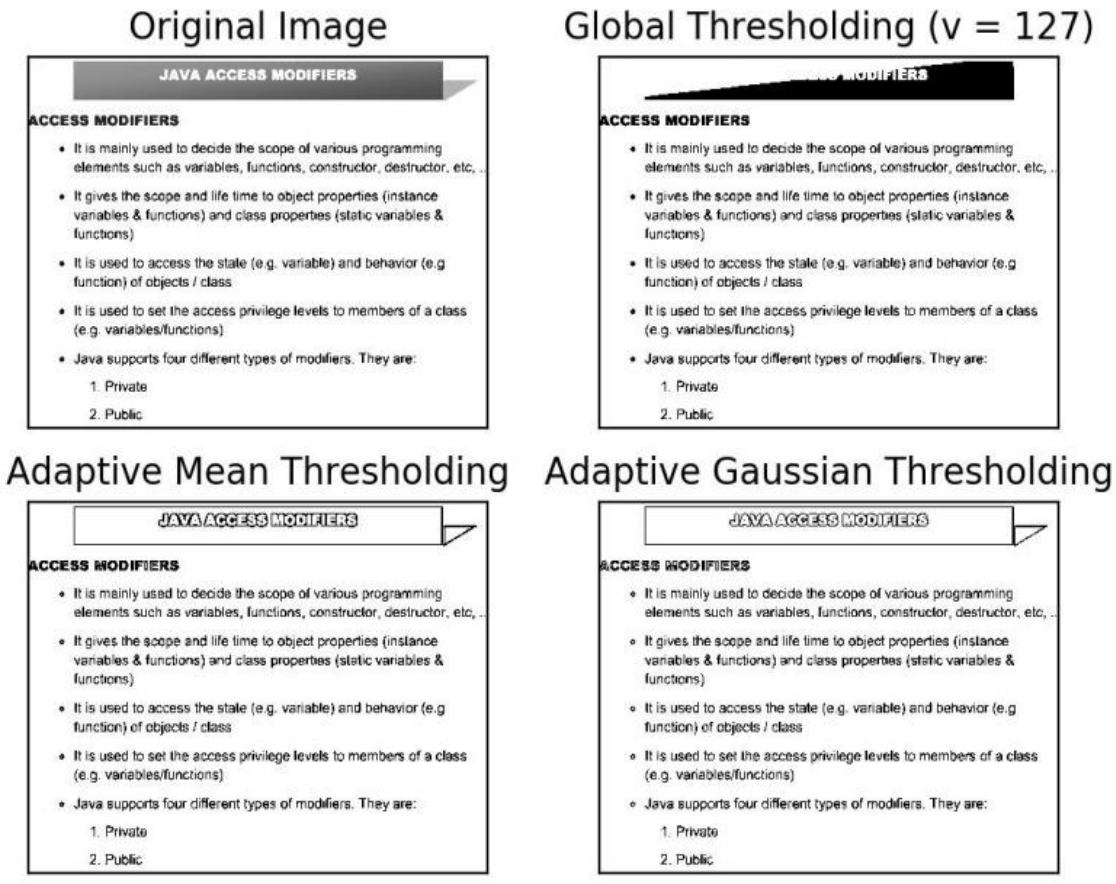
VI. RESULT AND OUTPUT

1. DATASET



Sample Of Dataset Used For Neural Net Training

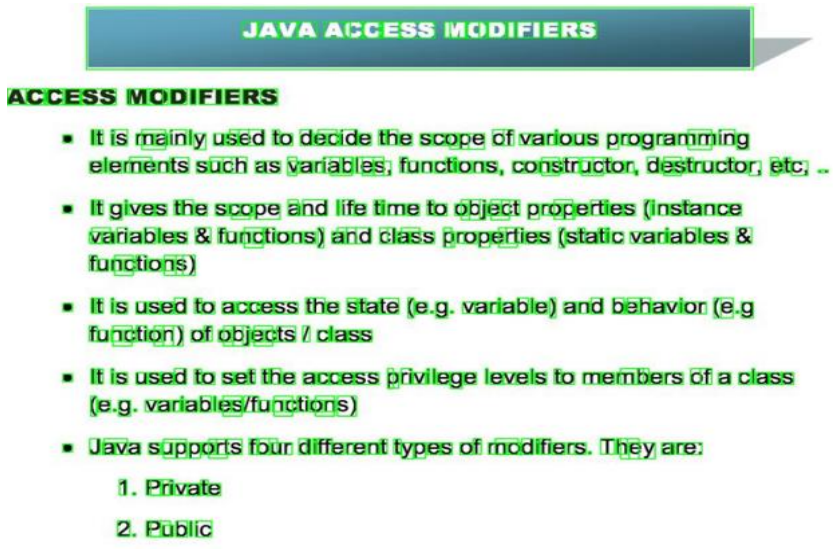
2. BINARISATION



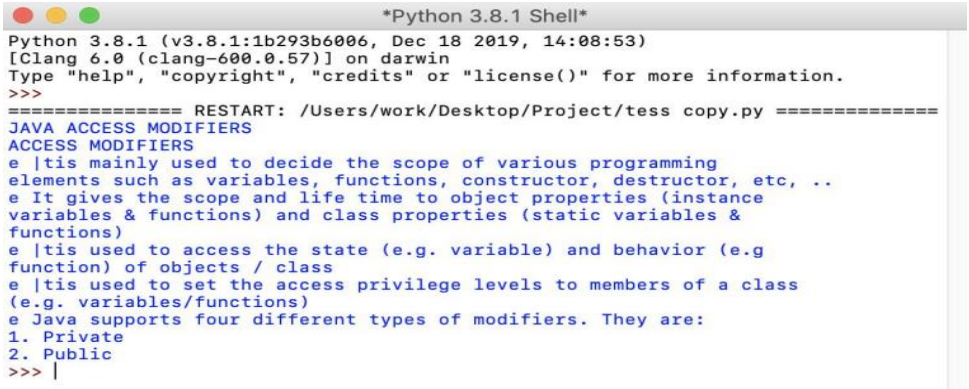
Output Of Different Binarisation Techniques

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

3. PRINTED TEXT PREDICTION

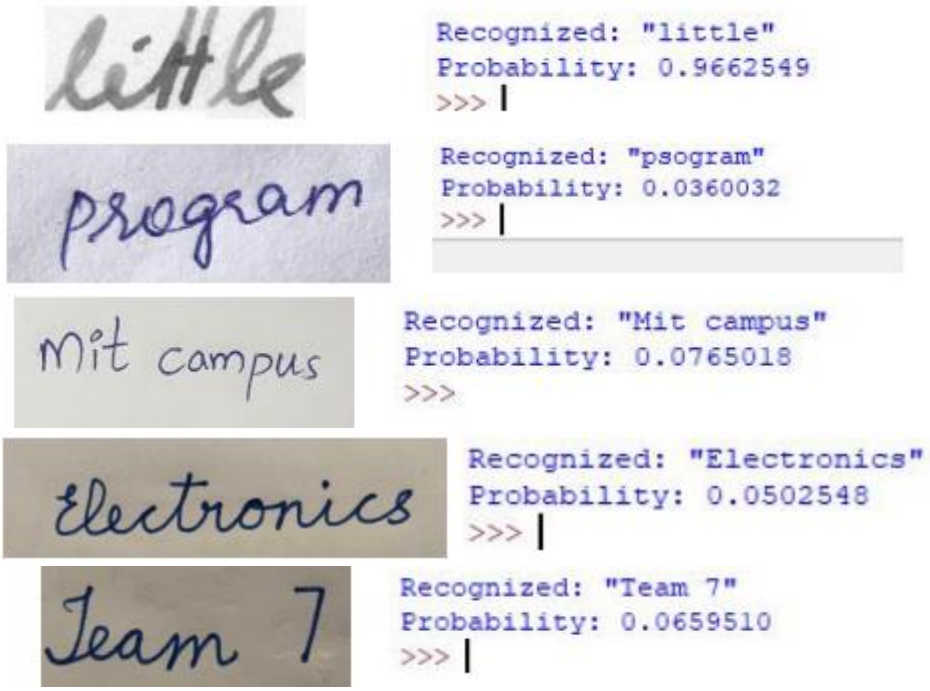


Output Form Printed Text Recognizer



Output Form Printed Text Recognizer

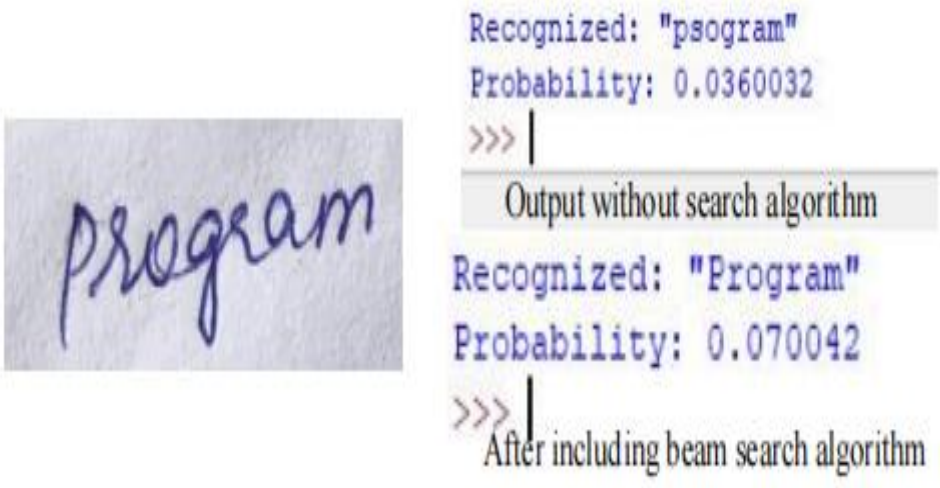
4. HANDWRITTENTEXT PREDICTION



Input Images, Output Of The Handwritten Text Predictor

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:09/September-2022 Impact Factor- 6.752 www.irjmets.com

5. SEARCH ALGORITHMS



VII. CONCLUSION

The dataset was collected and we implemented a robust text recognition and speech synthesizer system capable of converting handwritten and text data in an image to speech format. These technologies can be used separately for text recognition purposes and/or for text to speech conversion purposes, and can be combined too. The text recognition system we developed a deep learning engine consisting of multidimensional recurrent neural network and convolutional neural network for intelligent character recognition. For the text to speech synthesizer, WaveNet, a deep neural network for generating raw audio waveforms was used. A user can now, take a picture containing text or handwritten data, which is then passed through the recognition engine to get machine understandable text, which is then passed through the text to speech synthesizer to get audio information of the handwritten text. Using this engine a visually disabled and challenged people like those who are blind or have any other vision impairments can would be able to gain access to text and handwritten documents, which would have they would be unable to access prior. The engine can also be used for further research and development of artificially intelligent assistants by enabling them to become more smarter in accessing data of the human world such as street signs or bills.

VIII. REFERENCES

[1] A. Acero, "An overview of text-to-speech synthesis," 2000 IEEE Workshop on Speech Coding. Proceedings. Meeting the Challenges of the New Millennium (Cat. No.00EX421), Delavan, WI, USA, 2000, pp. 1-, doi: 10.1109/SCFT.2000.878372.

[2] D. Mahanta, B. Sharma, P. Sarmah and S. R. M. Prasanna, "Text to speech synthesis system in Indian English," 2016 IEEE Region 10 Conference (TENCON), Singapore, 2016, pp. 2614-2618, doi: 10.1109/TENCON.2016.7848511.

[3] Frederik J. Simons · Dong V. Wang, " Spatiospectral concentration in the Cartesian plane", Department of Statistics and Operations Research, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, U.S.A, 2011

[4] J.Pradeep , E.Srinivasan and S.Himavathi, "Diagonal based feature extraction for handwritten alphabets recognition system using neural network" , International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 1, 2010

[5] P. Jayawardhana, A. Aponso, N. Krishnarajah and A. Rathnayake, "An Intelligent Approach of Text-To-Speech Synthesizers for English and Sinhala Languages," 2019 IEEE 2nd International Conference on Information and58 Computer Technologies (ICICT), Kahului, HI, USA, 2019, pp. 229-234, doi: 10.1109/INFOCT.2019.8711051.

[6] S. Lukose and S. S. Upadhya, "Text to speech synthesizer-formant synthesis," 2017 International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, 2017, pp. 1-4, doi: 10.1109/ICNTE.2017.7947945.

[7] Plamondon, Réjean, and Sargur N. Srihari. "Online and offline handwriting recognition: a comprehensive survey." Pattern Analysis and Machine Intelligence, IEEE Transactions on 22.1 (2000): 63-84

[8] R. Reeve Ingle, Yasuhisa Fujii, Thomas Deselaers, Jonathan Baccash, Ashok C. Popat, "A Scalable Handwritten Text Recognition System". Google Research Mountain View, CA 94043, 2019

[9] Saad Bin Ahmed ; Saeeda Naz ; Muhammad Imran Razzak, "A Novel Dataset for English-Arabic Scene Text Recognition (EASTR)-42K and Its Evaluation Using Invariant Feature Extraction on Detected Extremal Regions",

[10] IEEE Access (Volume: 7), pp: 19801 - 19820, 13 February 2019

[11] T. Yoshimura, K. Hashimoto, K. Oura, Y. Nankaku and K. Tokuda, "WaveNet-Based Zero-Delay Lossless Speech Coding," 2018 IEEE Spoken Language Technology Workshop (SLT), Athens, Greece, 2018, pp. 153-158, doi: 10.1109/SLT.2018.8639598