VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Winter Semester 2020-21

BTECH IN COMPUTER SCIENCE ENGINEERING-AI AND ML

CSE1015-MACHINE LEARNING ESSENTIALS

PROJECT REPORT

FOR

SIGNBOARD TRANSLATION FROM VERNACULAR LANGUAGES

Submitted By

Sruthi Srinivasan-19BAI1046

A.Prasad-19BAI1061

Vishnu Prasad.M-19BAI1127

To

Dr. Rajalakshmi R

DECLARATION BY THE CANDIDATE

I hereby declare that the report titled “Signboard translation from vernacular languages” submitted by me to VIT Chennai is a record of bona-fide work undertaken by me under the supervision of Dr. R. Rajalakshmi, Associate Professor, SCOPE, Vellore Institute of Technology, Chennai.

Signature of the Candidate

ACKNOWLEDGEMENT

We wish to express our sincere thanks and deep sense of gratitude to our project guide, Dr. R. Rajalakshmi, School of Computer Science and Engineering for her consistent encouragement and valuable guidance offered to us throughout the course of the project work.

We are extremely grateful to Dr. R. Jagadeesh Kannan, Dean, School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology, Chennai, for extending the facilities of the School towards our project and for his unstinting support.

We express our thanks to our Head of the Department for his support throughout the course of this project.

We also take this opportunity to thank all the faculty of the School for their support and their wisdom imparted to us throughout the courses.

We thank our parents, family, and friends for bearing with us throughout the course of our project and for the opportunity they provided us in undergoing this course in such a prestigious institution

BONAFIDE CERTIFICATE

Certified that this project report entitled “J-Title” is a bona-fide work of Sruthi Srinivasan 19BAI1046, A.Prasad 19BAI1061, Vishnu Prasad 19BAI1127 carried out the “J”-Project work under my supervision and guidance for CSE1015-Machine learning Essentials.

Dr. R. Rajalakshmi

SCOPE

TABLE OF CONTENTS

Abstract

India is a very diverse country as a result of which numerous vernacular languages are spoken around the country. For an interstate traveller it would be highly tedious to understand and decipher every language and read various sign boards written in different languages along his journey.

Traditionally, people would use several translation tools in order to overcome this language barrier. However, this would prove to be inefficient since text in different languages often cannot be translated directly. In order to facilitate translation of written text we chose to build this application which can directly translate text written in several indian vernacular languages into any other language of the user's choice merely by the input of an image of the signboard clicked by the user. Subsequently, the user can obtain the translated text as an audio message as well.

When an input image is fed, the text in the image is firstly detected, then recognised and finally extracted from the image. This type of text recognition can be done using OCR or optical character recognition. In our project, the EasyOCR software is used for the purpose of multilingual text recognition. The text can be translated into any language of the user's choice. This is carried out using the google translate module in python and finally the audio conversion is done using a module in python named pyttsx3 used for text to speech conversion.

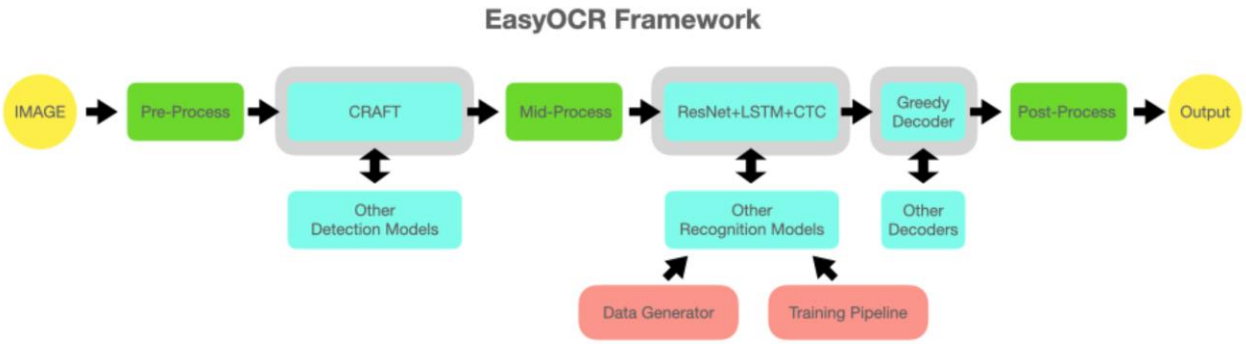
The dataset in consideration consists of images of several words in multiple languages and these are trained to form the EasyOCR module.

Introduction

Text detection is the premise and guarantee of text recognition. Multi-oriented text detection is the current research hotspot. Due to the variability in size, spatial layout, color and the arrangement direction of natural scene text, natural scene text detection is still very challenging. In order to carry out text detection and recognition we are using the EasyOCR module.

This module was built by JaidedAI for the purpose of multilingual text detection and recognition. Most text detection and recognition techniques are applicable only to english. EasyOCR however is an OCR that performs text detection and recognition for several indian vernacular languages which is the requirement in this project. This is done by incorporating all the characters present in that language supported by nearly 30000 to almost 50000 words if the language is popular.

EasyOCR uses a generative adversarial network(GAN) to generate a realistic handwritten dataset to improve handwritten text recognition.



EasyOCR uses the CRAFT text detector for text detection. Text recognition consists of ResNet and LSTM with Connectionist temporal classification(CTC) for neural network output and scoring.

Related Works

1. This work[1] presents an image operator that seeks to find the value of stroke width for each image pixel, and demonstrates its use on the task of text detection in natural images. The suggested operator is local and data dependent, which makes it fast and robust enough to eliminate the need for multi-scale computation or scanning windows. Extensive testing shows that the suggested scheme outperforms the latest published algorithms. Its simplicity allows the algorithm to detect texts in many fonts and languages.

2. This work[2] proposes a simple and fast multi-oriented text detection method. This method first optimizes the regression branch by designing a diagonal adjustment factor to make the position regression more accurate, Secondly, this method adds an attention module to the model, which improves the accuracy of detecting small text regions.Then, it introduces DR Loss to solve the problem of positive and negative sample imbalance.

3. This method[3] proposes an attention mechanism which roughly identifies text regions via an automatically learned attentional map. This substantially suppresses background interference in the convolutional features, which is the key to producing accurate inference of words, particularly at extremely small sizes.

4. This method[4] proposes an end-to-end method for scene text localization and recognition. It introduces features such as a departure from a strict feed-forward pipeline that is replaced by a hypotheses-verification framework simultaneously processing multiple text line hypotheses, the use of synthetic fonts to train the algorithm eliminating the need for time-consuming acquisition and labeling of real-world training data and the use of MSERs which provides robustness to geometric and illumination conditions.

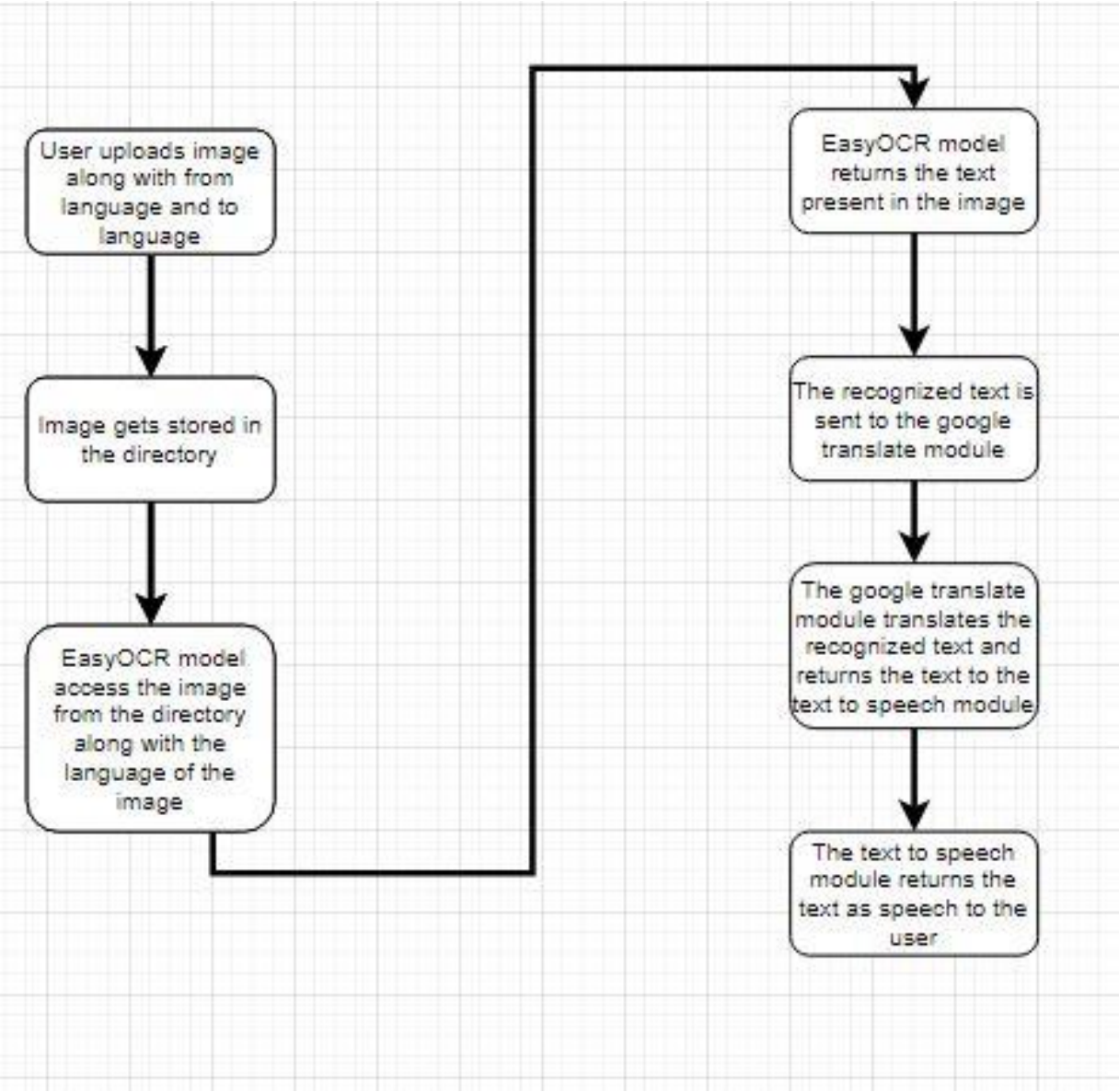
Methodologies:

The user needs to upload the image to the software from their device. After uploading the image the user needs to select the language present in the image and the language that the user desires to be translated. Once choosing the languages the user needs to press the Process button.
Once the process button is pressed the image chosen by the user is sent to the EasyOCR function.

The EasyOCR model initially converts the images to gray scale and then to binary images made up of only 2 colors (black and white). Once the images are pre-processed the image is sent to the CRAFT model(Character Region Awareness For Text detection), here is where the model will find the bounding box of where the text is present. Once the bounding box is formed, the image is cropped based on the coordinates of the bounding box.

This OCR uses the pattern recognition approach which works by identifying the character as a whole we can identify a line of text by looking for rows of white pixels with rows of black pixels in between in the same way we can identify what an individual character begins and ends next we convert the image of the character into a binary matrix or white pixels or zeros and black pixels or ones then by using the distance formula we can find the distance from the centre of the matrix to the farthest one we then create a circle of that radius and split it up into more granular sections at this point the algorithm will compare every single subsection and can send database of matrices representing characters with various fonts to find a character it statistically has the most in common with doing this for every line in every character makes it easy to bring printed media into the digital world.
Once the text from the image is recognised the image is sent to the Google Translator which translates the given text to the desired language specified by the user and this text is displayed on the screen along with the speech output using Pyttsx3.

Flowchart of the project



Experiments and results:

MARATHI	Pytesseract	EasyOCR
	द घ् वभ र र्न न." ना-) 7ठर गं । नि 11 आ, नि. डर झट[टा। वि टी:	महानगर पालिका भवन डेक्न शनिवार बाडा शिवाजीनगर
	आ न टा	पॉडर्नटीडेपो ३त३५ 5टँ [ँ2ऐ२
	का ् पा 8: 8) । ।)प. व 841	माऊल
	३... स. >. चस ळा र न ळे! क) र	साइन्स ईल मरीत लाईन्स "{ड २/६6

ENGLISH	PyTesseract	EasyOCR

	13103 [h] 2,01 .	EMERGENCY EXIT
	o @ = Oakland |[Berkeley enevrl |\ SIS\ 4 3 ’.‘ 'g o)	MEST 24) 13) Wa Oakland Berkeley LEFT EYIT
	I ‘/AIE o1 8	Plecse be mindful cf distoncing soaidl
	1L {[TR N S B =N AU b 1IN ATt M, m"?’!‘!!m e L - ‘ _:: Ll WARNING! G_ e p—S / M _ Do not proceed until "3 . ® bariers fully raised.N , No Talgating ——— // \ .	WARNING! STOP Do not proceed until barrier is raised. No 0 fully Tailgating SEGR eHL{ 8JS655["0#5

Tamil	pytesseract	Easy OCR
	துப்ப டூவண்டஈம் 30.00 (=... 10.00 0...	துப்ப & 8 வேண்டாம் 30.00 ணஸ
	_... 3060(12ாள) க _ வு) ம லவா 911011	ஞ ச2ெமம ம 30 ஸ 12 3றந^) ஏ5=ெள}ொ அனுமதி இல்லை $:9"=ெள}
	... குப்னப ... டூபஈடஈடூத மொ பாவோகௌள	குப்பை போடாதே வ0 _:{1}3^9
	தியப ம வத பயன்ப ப்ப லவ்வ அ ஆட. பஹ்ல் க வல்லவ வவததல் ப த்ய க க த ன் க ல்லை ன ணை ந் ன் ணட. ் 7 ம் ் ன ளா!	நெல்லை மன\ை ஈ4

Bengali	Pytesseract	Easy OCR
	ক সা দূর	করোনাভাইরাস সামাজিক দূরত্ব
	নু ৬.গগনা.*১// 4... / .ন্ ১৬৮৬২ নে নর	গগনা /%,)2]'%<]5^][8
	৮6145517106 51. 5.2...জজ যা সা -.সে [নে ৬০ ০ আন আচে শারযার নীট[সস আর ক মত তি ।মে মি তে	@চচঙঞঋ| চ 31: !@ [চেশায়ার ঠট
	বর্ণপরিচখুঁ	বর্ণপরিচয়

Hindi	Pytesseract	EasyOCR
	न ह ही की ४ कै कल्व्यी । ड - ज क्र 1 ः « हि, दी ! ष्थ्ि ७ है को [0 जी ह	रूकिये
	कर दिस का, 2, 2 “सु क की ९»... भ्शट । ल हि जि । है हे न के 2८ अर हे. की न ४4 अं, लि... । बाल हि के कि धाा 00९0९ पर. ७ उख न जा ॥॥||||| त लय घद , [जिला जा 0 है फककन्नननत - उब। हे) 4 ही: रे न । भा के हि दि जिस	नरसिंहपुर क़"$।ढःलणर
	'फोटो सींचना मनाहै। शिण 0५९१४ 15 शश०ाेाए)	फोटो खींचना मना हे। क़=07कक4शख } $ शब0=।8च7६२
	। बटचलर) _ ·	बैट्री चार्जर रूम

Telugu	Pytesseract	EasyOCR
	క గ "గా న స వక క్ న ? 1 | తా	గు టుపల్లి
	క ద.	గాతమి ఎక్స్రెప్రెస్
	శే గ్ల 04| |	దిసురరుపరని 7స0 గాేడే @`>క]-ట^]
	0121212211 0210003 1-ఎ పన ము య... | చం...	5డ 'ప`ఎ:" ద-చ] 05" ఫలకొనుమా రైల్వే స్టేషన్ రఇఇన] #శఇ ~^ఎలగ<^ న>ల. $+శ+చంన 29థ

Kannada	Pytesseract	EasyOCR
	ಜ. ಶಿ | ಎ ಬ ಟಕ 138.2	% " |(ಫ " ೀ"6೦ ಚ ಚ^) 4 '(1 51 ಐ೦೦ಟಲ [ಥ.ಪ] ೨:ಳಷ[ಶ [ತಚ:^ಱ "( ) ಥಾ್ಷಗತ್ [ಸಸೆೃಪ್
	ತ ದ್ದ. ''ಷತಾಷಾಗ ಷ್ಟ)	ಎತ ಉಆಯ ೨"ದ 19 ಳ19]]
	ಸ್ ಸ್ರ 1 ಱ 11.1.1... ಹ ಕ (1! ಇ ಬೂ ಱ ಱ 8.126	ಕರ್ನಾಟಕ ಉಚ್ಯ ನ್ಯಾಯಾಲಯ }3ೂ೦ (0=ಳ1 05 ^4ಳೈೃ^[4^^ ~6452.53
	. ' ನಂಜನಗೂಡು 3 ಅ ಱ.1 8)876ರ	ನಂಜನಗೂಡು]ಸಮ್ೂ೦ೊತ 3ಋಊಇ/9-] ಇ.ಇಖ 5[ಚ- 06-.೦

Results Discussion:

Accuracy calculation:

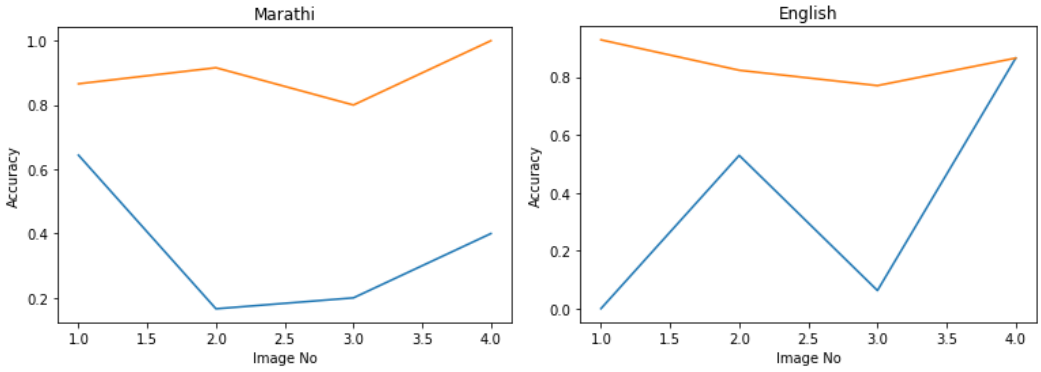
Single image accuracy=Correctly Predicted Characters/Total no of actual characters

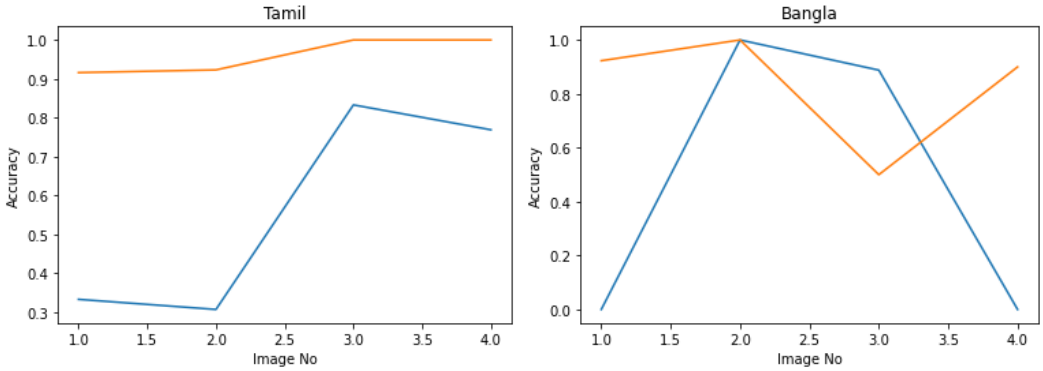
Accuracy of the OCR for a language is = sum of all single image accuracies/total images

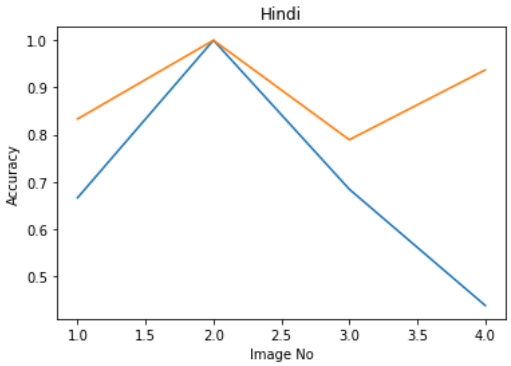
EasyOCR vs Pytesseract:

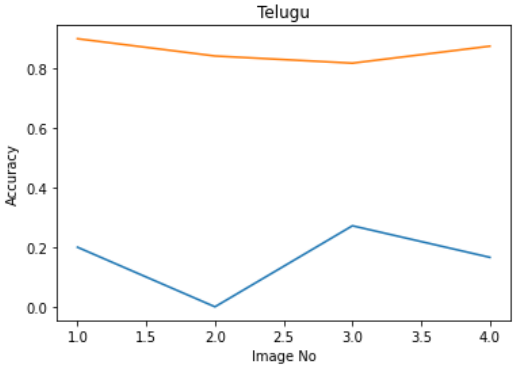
Language	Pytesseract	Average score	EasyOCR	Average score
Marathi	0.644, 0.166, 0.2, 0.4	0.352	0.866, 0.916, 0.8, 1.0	0.895
English	0.0,0.529,0.0625, 0.865	0.364	0.928,0.823, 0.770, 0.865	0.846
Tamil	0.333, 0.307, 0.833, 0.769	0.56	0.916, 0.923, 1.0, 1.0	0.959
Bengali	0.0, 1.0, 0.888, 0.0	0.478	0.923,1.0, 0.5, 0.9	0.830
Hindi	0.666, 1.0, 0.684, 0.4375	0.696	0.833, 1.0, 0.789, 0.937	0.889
Telugu	0.2,0.0,0.272, 0.166	0.159	0.9, 0.842, 0.818, 0.875	0.858
Kannada	0.0714, 0.230, 0.260, 1.0	0.39	0.642, 0.153, 0.8261, 1.0	0.655

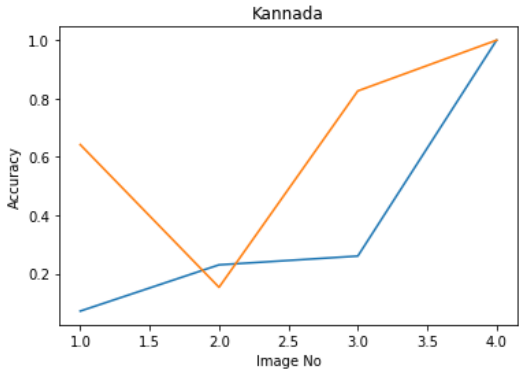
Hence, it was observed that EasyOCR has better accuracy than Pytesseract for all languages listed above.











-(pytesseract - Blue)
-(EasyOCR - orange)

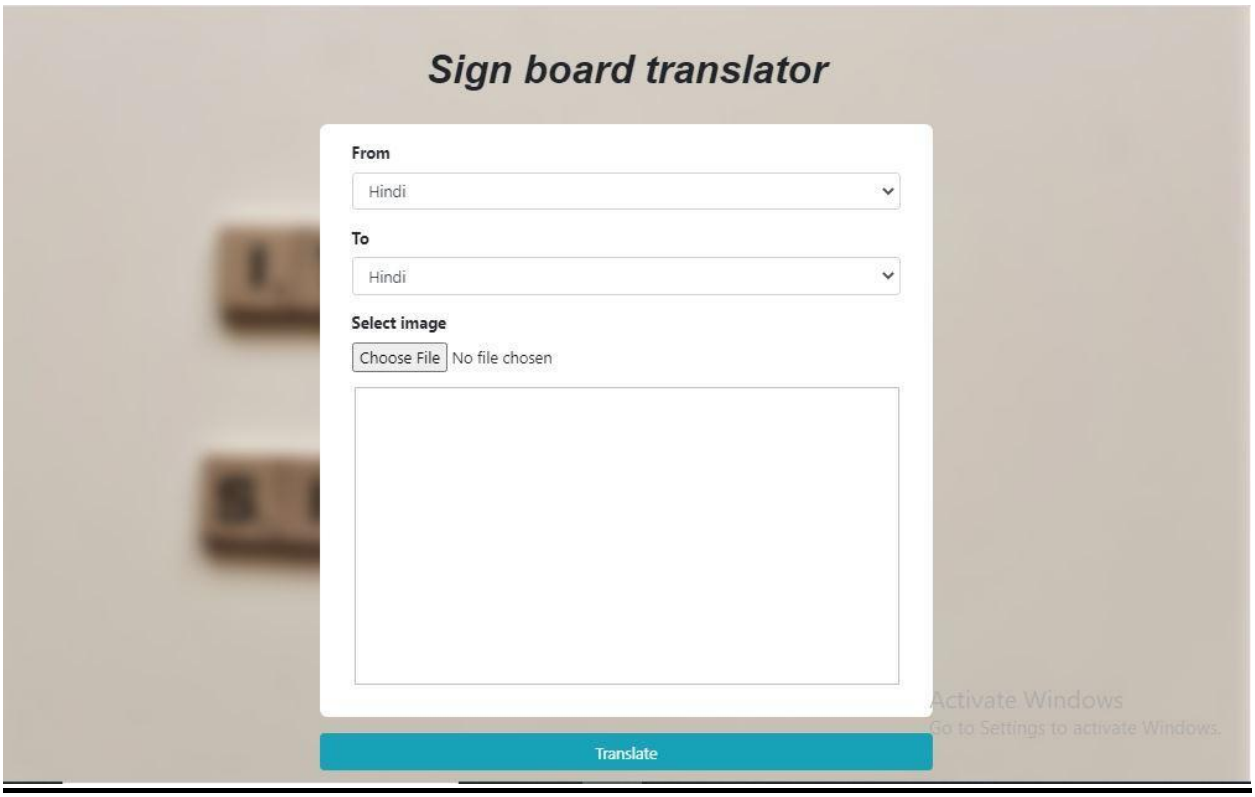
Comparative study:

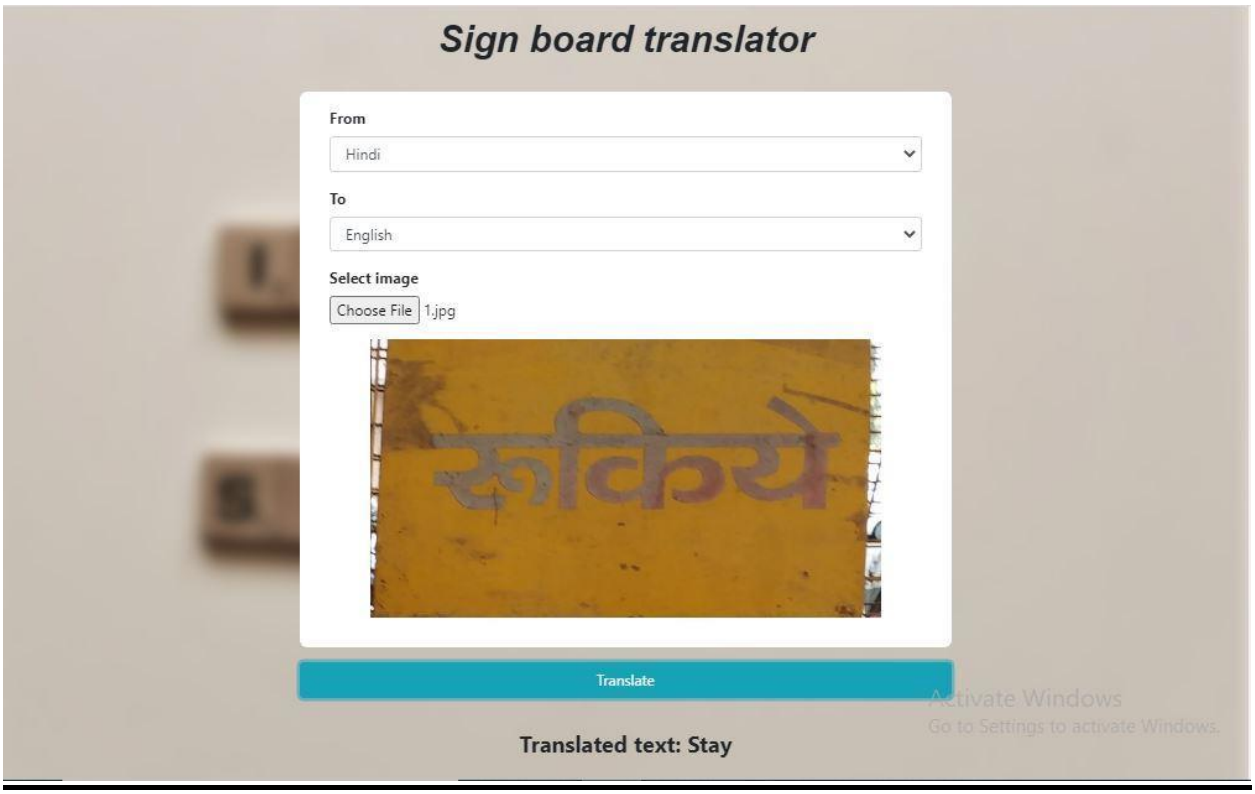
Technique	Dataset Used	Performance	Drawbacks
Stroke Width Transform[1]	ICDAR 2003	0.66 f score	The grouping of letters can be improved by considering the directions of the recovered strokes. This may allow the detection of curved text lines as well.
Text localisation and detection using MSER, hypothesis verification and synthetic fonts[4]	ICDAR 2003 CHARS74K	0.57 f score (ICDAR) 71.6 % matched (CHARS74K)	individual letters not being detected as MSERs in the projections used
Convolutional Scene Text Detection With Text Attention[3]	ICDAR 2013, ICDAR 2015 and MSRATD-500	0.84 f score (ICADR 2015) 0.88 f score (MSRA-TD500) 0.88 f score (ICADR 2013)	Has issues with large texts and curved text
Single Shot Text Detector with Regional Attention[2]	ICDAR 2013 ICDAR 2015 COCO text dataset	0.88 f score (ICDAR 2013) 0.77 f score (ICDAR 2015) 0.37 f score (COCO text dataset)	None
Benchmarking Scene Text Recognition[5]	IIIT-ILST dataset Synthetic Scene Text Dataset	42.9-57.2-73.4 WRR 75.6-86.2-92.8 CRR (Hindi-Telugu-Malayalam) (IIIT-ILST)	None

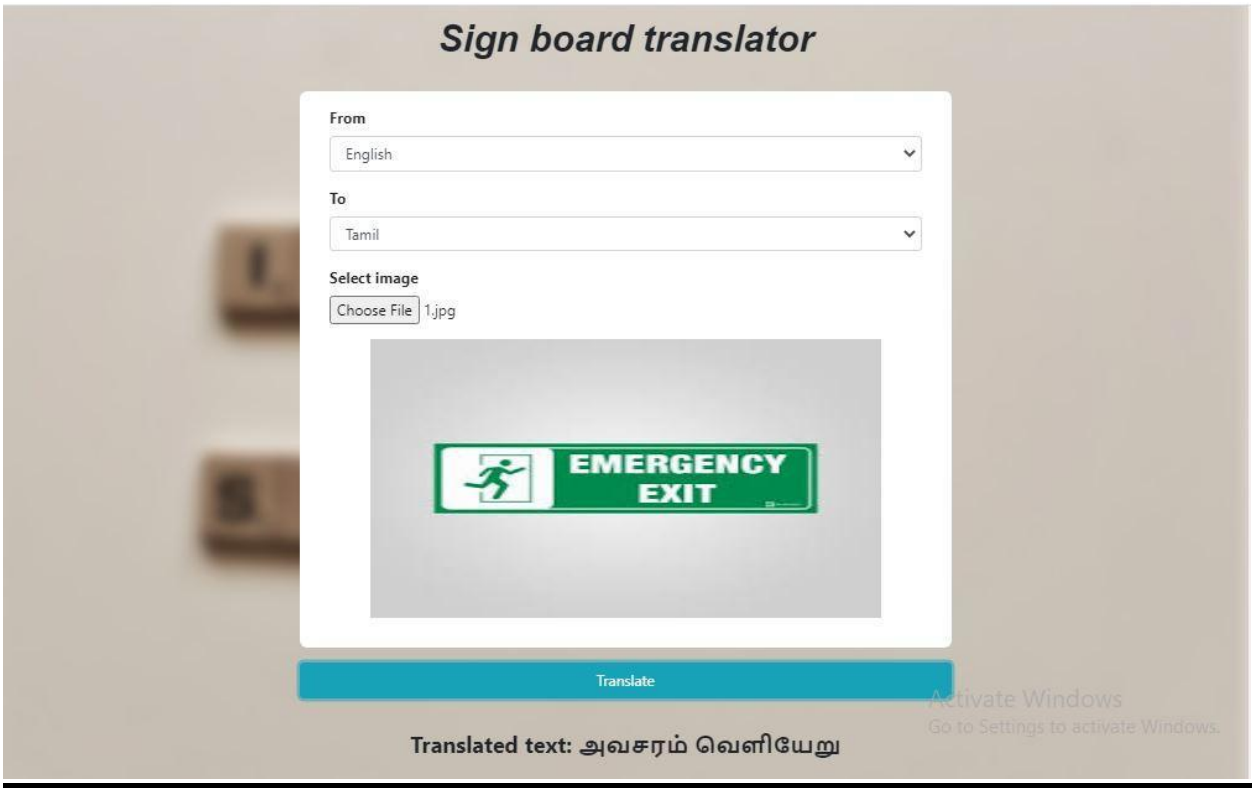
Tools Used:

- Flask- As the server for the website
- HTML,CSS and Javascript- For website design
- EasyOCR-For text detection and extraction
- Google translate module- For text translation
- Pyttsx3 - To convert text into audio

Output Screenshots:




Sign board translator
From
Hindi
To
English
Select image
Choose File
1.jpg
रुकिये
Translate
Translated text: Stay


Sign board translator
From
English
To
Tamil
Select image
Choose File
1.jpg
EMERGENCY
EXIT
Translate
Translated text: அவசரம் வெளியேறு

Conclusion:

In conclusion, a simple and user-friendly website was built in order to extract and translate the text written in vernacular languages, into a language of the user's choice.The users can upload a clear image of the signboard to be translated, into the website and subsequently click the submit button. Once the user clicks the submit button, the text on the image is detected and extracted using a model known as EasyOCR. This extracted text is then fed into the google translate module which then converts the text into any language of the user's choice.

In the course of making this project another technique to extract text on images, known as Pytesseract was also explored. The results obtained as a result of this technique were compared with those obtained for EasyOCR. It was then observed that EasyOCR is a better and more accurate method than Pytesseract for text detection and extraction from images. Finally, graphs for each language were plotted depicting the accuracy obtained for the images tested on EasyOCR.

Future Enhancements:

The software in its current form exists as a website. One major future enhancement would be to develop this software as a mobile application which would be far more convenient and practical to use on the go. Another goal would be to incorporate a technology that can convert text in any language into audio output.

References:

1. Epshtein, B., Ofek, E., & Wexler, Y. (2010, June). Detecting text in natural scenes with stroke width transform. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2963-2970). IEEE.

2. He, P., Huang, W., He, T., Zhu, Q., Qiao, Y., & Li, X. (2017). Single shot text detector with regional attention. In *Proceedings of the IEEE international conference on computer vision* (pp. 3047-3055).

3. Cao, Y., Ma, S., & Pan, H. (2020). FDTA: Fully Convolutional Scene Text Detection With Text Attention. *IEEE Access*, *8*, 155441-155449.

4. Neumann, L., & Matas, J. (2010, November). A method for text localization and recognition in real-world images. In *Asian conference on computer vision* (pp. 770-783). Springer, Berlin, Heidelberg.

5. Mathew, M., Jain, M., & Jawahar, C. V. (2017, November). Benchmarking scene text recognition in devanagari, telugu and malayalam. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (Vol. 7, pp. 42-46). IEEE.

6. Ye, Q., & Doermann, D. (2014). Text detection and recognition in imagery: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *37*(7), 1480-1500.

7. https://ai4bharat.org/articles/sign-board

8. https://github.com/JaidedAI/EasyOCR

9. https://pypi.org/project/pyttsx3/