International Journal of Trend in Research and Development, Volume 8(3), ISSN: 2394-9333
www.ijtrd.com

Things Translator an Android Application using Machine Learning

[1]Manasi Gosavi, [2]Aditi Hemane, [3]Priyanka Chaurasia and [4]Vinita Bhandiwad,
[1,2,3]Student, [4]Assistant Professor,
[1,2,3,4]Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India

Abstract: Learning a new language takes time and dedication. The reasons to learn a new language can be practical, some aspirational, some intellectual and others sentimental. Words, gestures and tone are utilized in union to portray a broad spectrum of emotion. The unique and diverse methods human beings can use to communicate through written and spoken language is a large part of what allows us to harness our innate ability to form lasting bonds with one another. When you travel to countries that you don't speak their language of, Google Translate has become an invaluable tool. In most cases, you will speak to the app or type in the word to get the translation. But, Things Translator is an application that allows its users to point their camera at any object and reacquire the object's name and description in any language they desire. One can use the application to translate a text from an image which is implemented as the text extraction module with the help of OCR engine. With its voice generation application one can learn different accents and correct pronunciation of a newly learned sentences and words. The system is implemented with the help of Machine learning algorithms in association with TensorFlow Lite models for object detection and Google API for translation purposes.

Keywords: *Machine Learning, Google Translate API, Object Detection, Translation.*

I. INTRODUCTION

Language impacts the daily lives of members of any race, creed, and region of the world. Language helps express our feelings, desires, and queries to the world around us. Words, gestures and tone are utilized in union to portray a broad spectrum of emotion. The unique and diverse methods human beings can use to communicate through written and spoken language is a large part of what allows us to harness our innate ability to form lasting bonds with one another.

When on an international trip if one wishes to wander around places, to understand the local culture and lifestyle of locals, to read signs & instructions, enjoy local delicacies, board a train, shop from local spots etc, knowing the local language is necessary.

We cannot learn all the languages of the places that we want to travel to and hence at such instances 'Thing Translator' comes handy. The outcome is predicted by a Machine Learning and OCR engine models, trained to identify the objects and texts from an image. The model also consists of a Google API which will translate the desired words.

II. LITERATURE SURVEY

For Object Detection:

[1] An approach for Object Detection in Android device

Object Detection is a study of Computer Vision and Image Processing that detects the instances of semantic objects belonging to specified classes in the dataset. In this study, we'll learn how objects can be detected using Deep learning technologies on android devices. One of the best models to detect objects in android device are Tensorflow lite, Yolov4, OpenCV. Using thresholding techniques, the color images are first converted into binary images. Morphological opening and closing filters are being used in sequence for object detection. Using Contour based learning techniques, objects are detected with bounding boxes. Detected objects are stored in array. The confidence scores and location of the objects in the image can also be specified for further analysis.

[4] COCO Dataset Stuff Segmentation Challenge

Image Classification is a process of classifying objects in an image into multiple classes/categories. Image segmentation is process of dividing a digital image into multiple pixels called as super pixels. COCO is a huge dataset with a total of 164k images in which around 118k images in the training dataset and around 5k images in the validation dataset, 20K images in test-dev , 20K images in test-challenge. It consists of 80 classes. The object detection model is trained to detect objects belonging to coco dataset semantic classes. Semantic classes can be well-defined stuff such as a vehicle, an individual or things like huge backdrop areas such as an area covered with grass. stuff classes are crucial, as they explain aspects of an image such as type of scene and thing classes represent things, objects, there are various factors to it such as location, physical attributes, type of the material and properties of the scene.

For extraction of text from images using OCR:

[5] Identification of Optimal Optical Character Recognition (OCR) Engine for Proposed System

The paper suggests that a large number of research efforts have been put forward that attempts to transform a document image to format understandable for machine so that it can recognize the text or the information from the image. OCR i.e. Optical Character Recognition provides a solution for this. OCR is software that converts printed text and images into digitized form such that it can be manipulated by machine. The image of the scanned document goes through various stages like preprocessing, segmentation, feature extraction, etc. in order to retrieve the information from the image. OCR is also popular among the Android applications. Tesseract is one of the most widely used open source library for implementing OCR in Android application.

[7] Extracting table data from images using optical character recognition text

In this paper, a method is proposed in which the tabular data contents of hard-copy documents is extracted from the text and character positions which are obtained from an OCR tool and transferred to digital forms. The performance of the method is measured by the number of detected rows and columns and presented with the results of other commercial products. The conversion of image-based documents into digital and

processible forms can be accomplished quite successfully with optical character recognition (OCR) tools.

For Google APIS + Text Extraction using OCR:

[10] Android Live Text Recognition and Translation Application using Tesseract.

From this paper we understand about an application that can be used to take pictures of text in the paper, screen, signboards, or anywhere and the application will translate the text. The goal is to be able to understand any text in any language and translate it in real-time to any other language of preference. Its implemented using Android as it is a widely used mobile OS around the globe. Also uses many open-source libraries such as Tesseract Leptonica for image computation, google-API-translate to translate the recognized text. It can recognize text from printed papers or screens and translate them into any language of preference. The proposed work has a scope to be used by visually impaired to recognize the text from anywhere and read the translated text to them.

[13] Development of an Automated Low-cost Book Scanner and Translator

A low-priced book scanner and translator using Raspberry Pi is the idea put forward by the mentioned paper. The system is able to read a book, translate parts of it to English if necessary and scan the whole book in the process to create a digitized soft copy of the book. A printed book can be converted to an audio book or PDF. For the conversion process, it interfaces an external webcam with the Raspberry Pi. The pictures taken from a book are processed and converted to text with the Tesseract OCR. Any part, that is not in English, is translated using the Google translator API, then the text is converted to audio file with eSpeak TTS engine.

III. PROPOSED SYSTEM

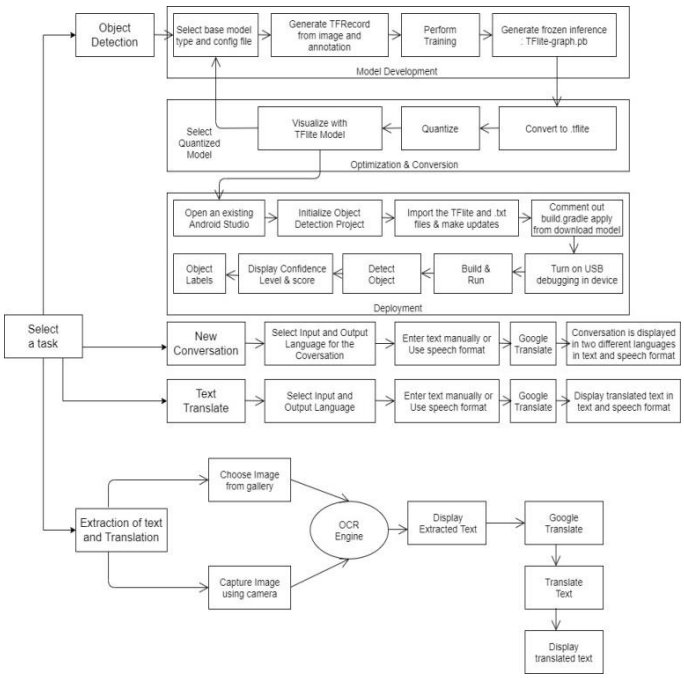

Fig.1. System Design

Things Translator is all in one android application which proves to be feasible for its users in multiple ways. The application comprises of various modules such as object detection with confident score, text extraction and translation of sentences identified from a stored or captured image, text translation in various languages and a bilingual conversation or chat system. Along with this it includes a text to speech module to understand the appropriate pronunciations and accents.

Fig.1. System Design - shows the detailed architecture of the system. As we can see, all the modules are expanded here. Based on the various parameters, the application consists of following modules, one for identifying the object in an image which will take into consideration factors such as size, shape, color, dimensions etc. of an object and provide with its identity and description in any language chosen by the user. The second module will have a captured or a stored image as its input and will return the extracted text from an image translated into a language of user's choice. The third module takes manually or verbally entered text as an input and converts into desired language as an output. The speech conversion of text will help end users with correct pronunciations and accents of words or sentences.

IV. METHODOLOGY

The Things Translator application went through different phases such as analysis, design, development, testing and integration and deployment over the span of a year. Using Machine Learning algorithms for training the dataset to classify and identify the objects, OCR engine to extract texts from images, using Google Translate API for translation, creating a bilingual conversation module and integrating various modules with the application created with the help of Android Studio. The system also stores the login credentials of the users with the help of Firebase database.

A. For Object Detection

An object detection model is trained to detect the presence and location of multiple classes of objects. For example, a model might be trained with images that contain various pieces of fruit, along with a label that specifies the class of fruit they represent (e.g. an apple, a banana, or a strawberry), and data specifying where each object appears in the image.

When an image is subsequently provided to the model, it will output a list of the objects it detects, the location of a bounding box that contains each object, and a score that indicates the confidence that detection was correct.

Confidence Score

The score is a number between 0 and 1 that indicates confidence that the object was genuinely detected. The closer the number is to 1, the more confident the model is.

Location

For each detected object, the model will return an array of four numbers representing a bounding rectangle that surrounds its position. For the starter model provided, the numbers are ordered as follows:

[top, left, bottom, right]

The top value represents the distance of the rectangle's top edge from the top of the image, in pixels. The left value represents the left edge's distance from the left of the input image. The other values represent the bottom and right edges in a similar manner.

Tensor Flow Lite

Things Translator for object detection makes use of TensorFlow Lite on a mobile device. We have used a pretrained MobileNet SSD quantized model. It is trained on the COCO dataset which can detect about 90 classes including banana, scissors, laptop, remote, vase etc. The main issue faced

while executing Machine learning models like object detection on mobile is that the models are often too big to run. Tensorflow Lite solves this problem by providing a lightweight solution to run machine learning models on mobile. You can see below yourself how nicely the app works and detects multiple objects quickly.

B. For Extracting Text from images

This module recognizes the text that is captured by a mobile phone camera and displays back the recognized text on to the screen. To develop this module we have used the Optical Character Recognition, OCR engine, and have implemented it in our own open source Android application.

In this module a user will be allowed to either capture image using phone camera or extract image from phone gallery to extract the text from the image. This module is mainly related to image processing to recognize characters in an image.

Considering text element, image can be classified as:

(i) document image and (ii) scene text image.

OCR algorithm deals with these images. While capturing image we can also pinch only the required text. OCR algorithm takes the input and recognizes the text present on the image displays back the recognized text on to the screen. Hence, Extraction is performed using Google's open source optical character recognition (OCR) engine 'Tesseract' and translated to a target language using Google's translation.

OCR engine – Tesseract

OCR technology allows the conversion of image which is scanned of printed character into text or any other information that user want using android mobile. OCR technology uses three phases first is Scanning of documents as optical images. Next is Recognition which involves converting those images to character streams representing letters of recognized words and the final element used to accessing or storing the text which are already converted. Converted text is nothing but the extracted text. When, the user begins by capturing an image using mobile camera containing text .To convert extracted text into any language text synthesizer is used. OCR aims to allow the user to extract the text from the image.

C. For Translation into different languages

We used the API's in this project like Google translator for translating languages and text to speech engine for converting the text retrieved into the audio. Google APIs is a set of application programming interfaces (APIs) developed by Google which allow communication with Google Services and their integration to other services. Google translator is used in this project for translating the present language into any other language that user wants. Google Translate can translate multiple forms of text and media, including text, speech, images, sites, or real-time video, from one language to another.

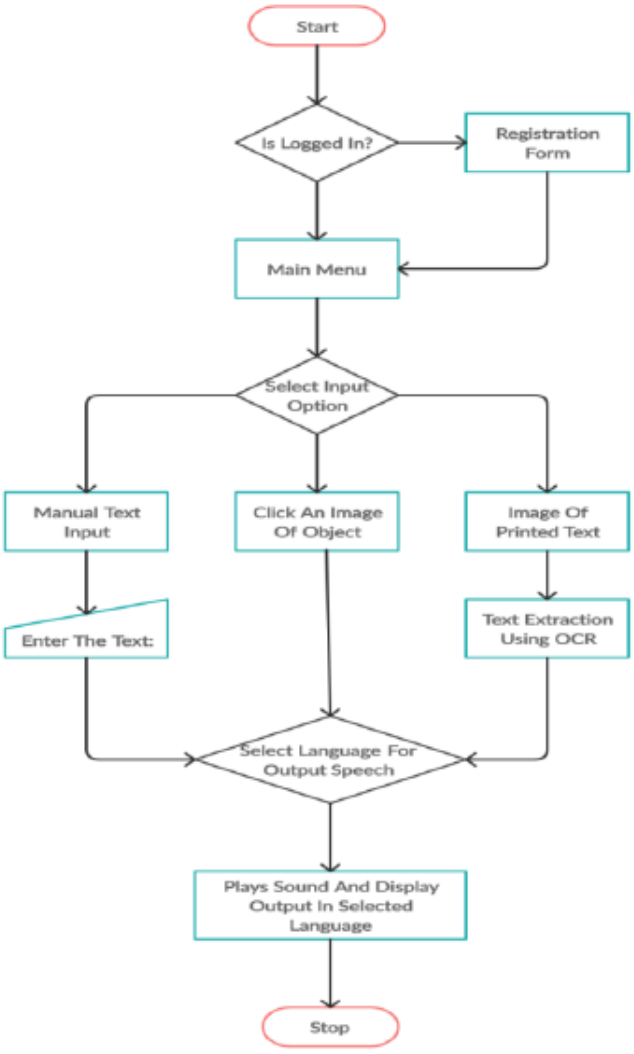

Fig.2.Flowchart

Fig.2. entails the basic flow of the entire methodology and explains each of the individual components of our approach.

The basic flow is as follows:

- When a user visits the 'Things Translator' android application, s/he should be logged in to continue the further process and if not then just simply register for further course of action.
- After being authenticated by correct username and password, the user will be redirected to main menu where they can perform various activities.
- The user gets to choose between available options of entering the text manually, clicking an image of any object or text.
- The input text or the text extracted from an image is then transformed into any language as desired by the end user.
- Whereas when an image of any object is captured, the system will respond with its identity and description in any desired language.
- The output can be of two formats, text and voice. Depending on the format selected by the user the output will be displayed.

CONCLUSION

Things Translator is designed to ease the travel experience of wanderers or feed the appetite of inquisitives of learning a new language or be a handy application for the visually impaired individuals helping them through their daily routine our

system, with the assistance of Machine Learning algorithm, OCR engine and TensorFlow Lite the application will detect the objects and also extract texts from images. With many similar applications that are already in existence, Things Translator is an all in one application which will assist the end users to translate a text or identify objects and provide their name as well as description into desired language along with helping our users to explore and learn new languages with correct pronunciation and accent.

Acknowledgment

We wish to express our sincere gratitude to Prof. Vinita Bhandiwad, Assistant Professor of Information Technology Department, for her guidance and encouragement in carrying out this project.

We also thank faculty of our department for their help and support during the completion of our project. We sincerely thank the Principal of Vidyalankar Institute of Technology for providing us the opportunity to carry out this project.

References

[1] G. Savitha, P. S. Venugopal, Sarojadevi and N. Chiplunkar, "An Approach for Object Detection in Android Device," 2014 Fifth International Conference on Signal and Image Processing, 2014, pp. 9-14, doi: 10.1109/ICSIP.2014.6.

[2] Z. Domozi, D. Stojcsics, A. Benhamida, M. Kozlovszky and A. Molnar, "Real time object detection for aerial search and rescue missions for missing persons," 2020 IEEE 15th International Conference of System of Systems Engineering (SoSE), 2020, pp. 000519-000524, doi: 10.1109/SoSE50414.2020.9130475.

[3] S. Kanimozhi, G. Gayathri and T. Mala, "Multiple Real-time object identification using Single shot Multi-Box detection," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-5, doi: 10.1109/ICCIDS.2019.8862041.

[4] Puri, "COCO Dataset Stuff Segmentation Challenge," 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), 2019, pp. 1-5, doi: 10.1109/ICCUBEA47591.2019.9129255.

[5] S. Sonth and J. S. Kallimani, "OCR based facilitator for the visually challenged," 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), 2017, pp. 1-7, doi: 10.1109/ICEECCOT.2017.8284628.

[6] M. G. Marne, P. R. Futane, S. B. Kolekar, A. D. Lakhadive and S. K. Marathe, "Identification of Optimal Optical Character Recognition (OCR) Engine for Proposed System," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-4, doi: 10.1109/ICCUBEA.2018.8697487.

[7] M. Y. Akpinar, E. Emeklıgıl and S. Arslan, "Extracting table data from images using optical character recognition text," 2018 26th Signal Processing and Communications Applications Conference (SIU), 2018, pp. 1-4, doi: 10.1109/SIU.2018.8404746.

[8] B. Tejas, D. Omkar, D. Rutuja, K. Prajakta and P. Bhakti, "Number plate recognition and document verification using feature extraction OCR algorithm," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 2017, pp. 1317-1320, doi: 10.1109/ICCONS.2017.8250683.

[9] V. Kumar, P. Kaware, P. Singh, R. Sonkusare and S. Kumar, "Extraction of information from bill receipts using optical character recognition," 2020 International Conference on Smart Electronics and Communication (ICOSEC), 2020, pp. 72-77, doi: 10.1109/ICOSEC49089.2020.9215246.

[10] S. Revathy and S. Nath, "Android Live Text Recognition and Translation Application using Tesseract," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 1259-1263, doi: 10.1109/ICICCS48265.2020.9120973.

[11] Nursetyo and D. R. I. M. Setiadi, "LatAksLate: Javanese Script Translator based on Indonesian Speech Recognition using Sphinx-4 and Google API," 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2018, pp. 17-22, doi: 10.1109/ISRITI.2018.8864274.

[12] H. Rithika and B. N. Santhoshi, "Image text to speech conversion in the desired language by translating with Raspberry Pi," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2016, pp. 1-4, doi: 10.1109/ICCIC.2016.7919526.

[13] S. Nawshin, S. K. Das, N. Hossain, J. D. Mela, A. Shafin Mohammad Mahdee Jameel and S. Islam, "Development of an Automated Low-cost Book Scanner and Translator," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019, pp. 1-5, doi: 10.1109/ICASERT.2019.8934635.

[14] Rahmani, "Adapting google translate for English-Persian cross-lingual information retrieval in medical domain," 2017 Artificial Intelligence and Signal Processing Conference (AISP), 2017, pp. 43-46, doi: 10.1109/AISP.2017.8324104.