Text Recognition, Object Detection and Language Translation App

Chaitra Naik[1], Amruta Khot[1], Arti Jha[1], Sejal D'mello[2]

[1]Dept. of Information Technology, Atharva College of Engineering, Maharashtra, India
[2]Assitant Professor, Dept. of Information Technology, Atharva College of Engineering, Maharashtra, India

---***---

Abstract - *This study presents a comprehensive review of OCR (optical character recognition), Translation, and Object Detection Research from a single image. With the fast advancement of deep learning, more powerful tools that can learn semantic, high-level, and deeper features have been proposed to solve the issues that plague traditional systems. The rise of high-powered desktop computer has aided OCR reading technology by permitting the creation of more sophisticated recognition software that can read a range of common printed typefaces and handwritten texts. However, implementing an OCR that works in all feasible scenarios and produces extremely accurate results remains a difficult process. Object detection is also the difficult problem of detecting various items in photographs. Object identification using deep learning is a popular use of the technology, which is distinguished by its superior feature learning and representation capabilities when compared to standard object detection approaches. The major focus of this review paper is on text recognition, object detection, and translation from an image-based input application employing OCR and the YOLO technique.*

***Key Words*: Text recognition, Optical character recognition, Object detection, Language translation, YOLO**

1.INTRODUCTION

With the advent of numerous photography gadgets and powerful mobile camera characteristics, all papers have become electronic in nature, such as pdf files and jpg files. Text recognition has risen in popularity in recent years as it has expanded into a wide range of applications, from scanning papers – bank statements, receipts, handwritten documents, coupons, and so on – to reading street signs in autonomous cars. Language obstacles may be overcome all across the world. For example, if a person is travelling to Paris and is unfamiliar with the French language, the text recognition function of the app may be used to detect and translate text seen on a picture.

Object detection has received a lot of academic attention in recent years because of its tight association with video analysis and picture interpretation. Object detection is a sophisticated computer vision technology that identifies and labels items in photos, videos, and even live video. However, there are other issues with photographs captured in the actual world, such as noise, blurring, and spinning jitter. Object detection suffers as a result of these issues. Issues with photographs recorded in the actual world include the camera's instability, which causes the acquired image to be blurry. To address these challenges, object identification algorithms are trained with a large number of annotated images before being used on fresh data. It's as easy as inputting input images and getting a completely marked-up output graphic. Object detection characteristics may be utilized to interpret traffic signs.

2. PROPOSED SOLUTION

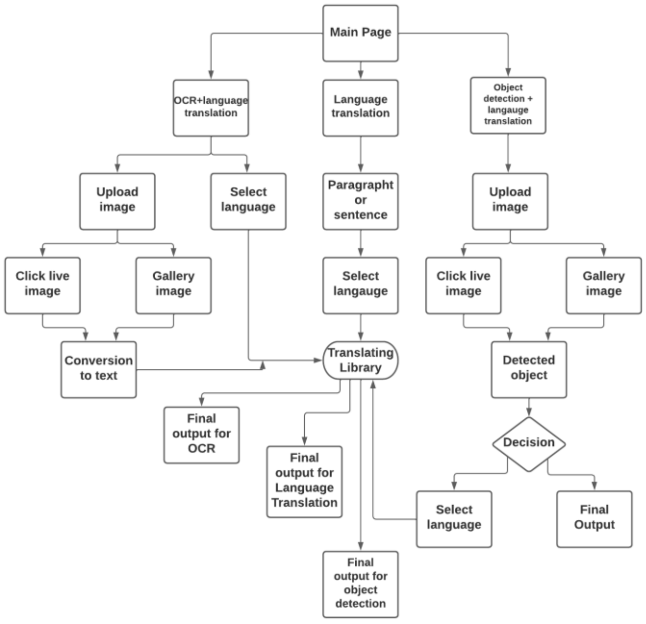


Figure-1: Block Diagram

Many applications based on OCR, language translation, and object identification have been seen. However, the majority of applications do not provide all of these functionalities. All of these characteristics have been included into this system. On the app's main screen, the user will be presented with three alternatives. Text recognition, object detection, and language translation are the three possibilities. Any essential option can be selected by the user. There are two alternative possibilities for text recognition. The user may either choose a picture from the gallery or click on a live image. After the user has provided input, the user must pick a language for translation and click submit.

Object detection is the app's second feature. The user is given two alternatives here as well. The user can either choose a picture for the gallery or click on a live image. After the object has been discovered, the user may translate the object's name into whatever language they like.

Only language translation is the last option, which requires the user to compose a paragraph or sentence and pick a language for translation.

3. METHODOLOGY

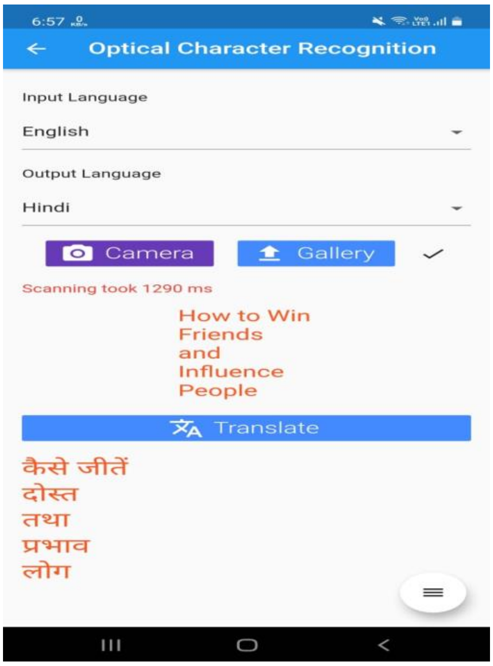
The three components of our project are OCR and language translation, object detection and language translation, and merely language translation. As a result, the user must first choose one of the aforementioned possibilities. We used OCR for text recognition, which was imported via tesseract. Tesseract employs a two-step process known as adaptive recognition. Character recognition is the first phase, and there are three sub-steps in this step as well. Image pre-processing is the initial stage. Images are preprocessed in this stage to increase the likelihood of successful recognition. Unwanted distortions are reduced, and certain visual characteristics are accentuated in this stage. The next two stages rely heavily on this phase. The actual recognition of character is the second sub step. It is based on the feature extraction idea. When the initial input is too vast to handle, just a subset of features is chosen. The features that are not picked are redundant, but the ones that are selected are critical. The performance is improved by using the smaller set of data instead of the initial huge one. The final sub step is picture post-processing. It is another high-accuracy error correcting approach. The tesseract's second step is to fill in any missing letters with letters that fit the word or phrase context. The text will be submitted to the language translation library, which will be imported using "googletrans" after it has been identified.

The YOLO method is used to detect objects in the second section of the app. The YOLO algorithm's first step is to partition the entire image into grids. There are seven vectors connected with each of the grid cells. Probability of the class, bounding box x, bounding box y, bounding box width, bounding box height, and classes are the vectors. As a result, anytime we come across an object grid cell at that moment, we check for the centroid first. Even if parts of two separate objects are present in a single grid, the centroid of whatever item is present in that grid is linked with that picture. If each grid is 4x4, for example, the real size of the grid becomes 4x4x7, where 4x4 is the grid size and each grid has 7 vectors. As a result, the train and test datasets are created. The picture is present in the train dataset, whereas vectors are present in the test dataset. We've made projections based on this. If the user desires to translate the object name into another language after it has been identified, the user must first pick a language for translation, which will be handled by the "googletrans" library.

The app's last feature is language translation, which requires the user to compose a sentence or a paragraph in any desired language and then pick a language for translation via the "googletrans" library.

4. RESULTS

Following are the screenshots of the interface and output of the proposed system.



Input Language
English

Output Language
Hindi

Scanning took 1290 ms

How to Win
Friends
and
Influence
People

कैसे जीतें
दोस्त
तथा
प्रभाव
लोग

Figure-2: OCR + language translation

The operation of Optical Character Recognition is depicted in Figure 2. The user must first choose the input and output languages, after which the picture for character recognition must be added. Once the picture has been added, it will recognize the text using OCR and translate it into the output language using the "googletrans" library, before displaying the recognized text.

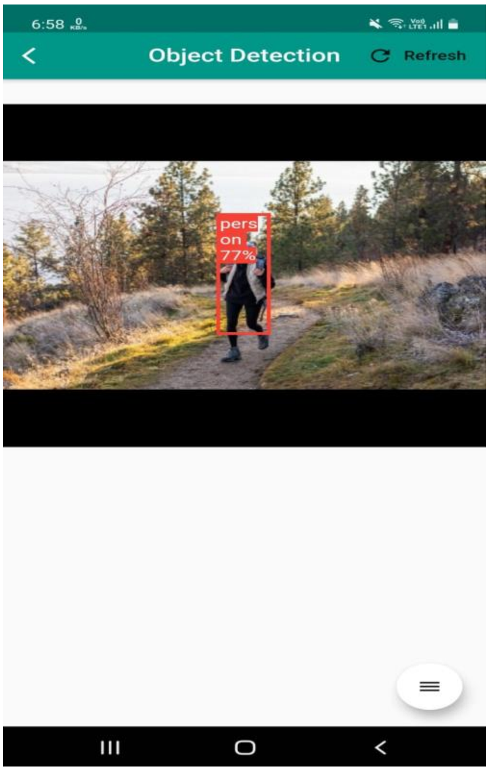
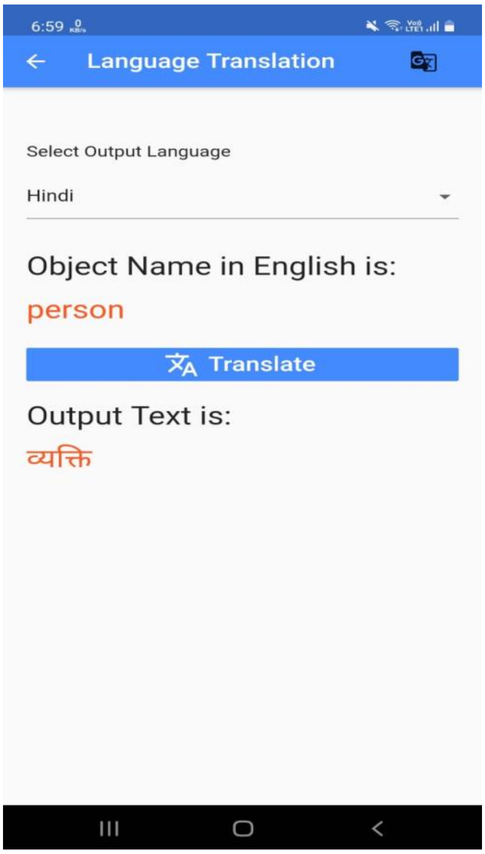


Figure-3: Object detection

Object detection is depicted in the above figure. We utilized the COCO dataset, which comprises roughly 328k photos, for object detection. YOLOv4 has been utilized. When compared to YOLOv3, YOLOv4 has a higher mAP (mean average precision). The major motivation for employing this approach was a 10% increase in mean average accuracy (mAP). YOLOv4 has an average precision of 38 to 44, while YOLOv3 has an average precision of 31 to 33.



Select Output Language

Hindi

Object Name in English is:

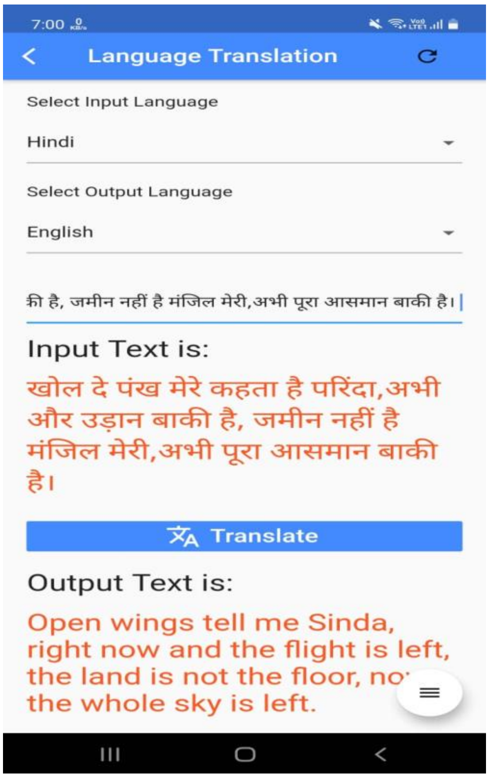
person

Output Text is:

व्यक्ति

Figure-4: Translation of detected object

The translated text of object detection is shown in the above figure. The user is given the option of translating the text that was found throughout the detection process. The user must first pick a language and then click the Translate button. The translation activity is started in the backend and utilizes the “googletrans” library to translate the content.



Select Input Language
Hindi

Select Output Language
English

फी है, जमीन नहीं है मंजिल मेरी,अभी पूरा आसमान बाकी है।

Input Text is:
खोल दे पंख मेरे कहता है परिंदा,अभी और उड़ान बाकी है, जमीन नहीं है मंजिल मेरी,अभी पूरा आसमान बाकी है।

Output Text is:
Open wings tell me Sinda, right now and the flight is left, the land is not the floor, no the whole sky is left.

Figure-5: Only language translation

Figure 5 depicts how Language Translation works. The user must first choose the input and output languages before adding the text to be translated. When text is entered, the “googletrans” library is used to translate it. Finally, the translated text is shown.

4. ADVANTAGES

1. The developed application is User-friendly.
2. The application includes text recognition, language translation, and object detection, so the user may get all of these functions in one application rather than having to install separate apps for each feature.
3. This application reduces the language barrier.

5. LIMITATIONS

When using the Text Recognition tool, the user must provide the input language for the text that is contained in the image.

6. CONCLUSION

For both characteristics, the created program can perform text recognition, object identification, and language translation into a chosen language with high accuracy. This application may be improved to handle the issue of translating pdfs and other documents from one language to another.

7. FUTURE SCOPE

1. This project may be improved by converting detected text to editable text, allowing the user to amend the text that was identified from an image before translating it.
2. The identified text may be turned into voice in a variety of languages.
3. The project may be improved to deal with real-time data instead of labels from a prepared dataset for object detection.

REFERENCES

[1] Thakare, Sahil, et al. "Document Segmentation and Language Translation Using Tesseract-OCR." 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS). IEEE, 2018.

[2] Li, Gaohe, Xinhao Li, and Bo Xu. "Numerical Simulation Technology Study on Automatic Translation of Foreign Language Images Based on Tesseract-ORC." 2019 International Conference on Robots & Intelligent System (ICRIS). IEEE, 2019.

[3] Liu, Chengji, et al. "Object detection based on YOLO network." 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC). IEEE, 2018.

[4] K. Elissa, Hiral Modi, M.C.parikh, “A Review On Optical Character Recognition Techniques”, International Journal of Computer Application,2017.

[5] Huang, Rachel, Jonathan Pedoeem, and Cuixian Chen. "YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers." 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018.

[6] Memon, Jamshed, et al. "Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR)." IEEE Access 8 (2020): 142642-142668.J.

[7] Du, Juan. "Understanding of object detection based on CNN family and YOLO." Journal of Physics: Conference Series. Vol. 1004. No. 1. IOP Publishing, 2018.

[8] Tao, Jing, et al. "An object detection system based on YOLO in traffic scene." 2017 6th International Conference on Computer Science and Network Technology (ICCSNT). IEEE, 2017.

[9] Ahmad, Tanvir, et al. "Object detection through modified YOLO neural network." Scientific Programming 2020 (2020).