International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 05 | May 2023 www.irjet.net p-ISSN: 2395-0072

Audio computing Image to Text Synthesizer - A Cutting-Edge Content Generator Application

Abhishek Venkata Shiva Siripalli[1], Nikhil Shinde[2], Prof. Lovenish Sharma[3]

[1]*Student, School of Engineering, Ajeenkya DY Patil University, Pune, Maharashtra, India*
[2]*Student, School of Engineering, Ajeenkya DY Patil University, Pune, Maharashtra, India*
[3]*Professor, Ajeenkya DY Patil University, Pune, Maharashtra, India*

Abstract - *In the anti-establishment world, there is a first-rate extent in the utilization of digital technological know-how to be aware of how and a vary of methods are on hand for a character to catch images. Such images may additionally comprise necessary textual information that the customer may additionally desire to edit or store digitally. This can be completed the utilization of Optical Character Recognition with the help of Tesseract OCR Engine. OCR is a branch of artificial Genius that is used in features to apprehend textual content material from scanned documents or images. The recognized textual content material can moreover be changed to audio sketch to aid visually impaired human beings hear the data that they wish to understand and additionally to the illiterate. So, truly at the existing day purposes convert image to textual content, picture to handwritten notes and later provide its audio contents is the use of Optical Character Recognition (OCR) tool.*

Now, we additionally introduced new attribute like image to text, textual content material to speech, and we can convert the textual content material to any language as per individual requirement, it will be increased available and accustomed way to do. All the journal, have reply in addition we're alongside with translator that can be google translated bundle deal for our project. In this we will be exploring wonderful bundle and mission will comprise web page the region customer can add photograph and in the returned of at the backend it will process enter and ship lower back aspect in form of API. This utility can be used for character focus from scanned archives so that information can be digitalized. Also, the data can be converted to audio form to aid visually impaired people obtain the records easily. In this, we can prolong the utility to that is can apprehend greater languages, one of a form fonts. Various accents can moreover be delivered for audio files in the upcoming future.

***Key Words*: OCR (Optical character recognition), translator, Hand written notes, Tesseract, Text-to-Speech (TTS), Tesseract, OCR Engine.**

1.INTRODUCTION

Audio computing Text and Image Synthesizer makes it doable to extract textual content material from pictures to automate the processing of texts on images, videos, and scanned documents. In this, we show up at how to manner an image to textual content material with React and Tesseract.js(OCR), pre-process images, and deal with the obstacles of Tesseract (OCR) and later provide an output in audio structure which can be downloaded and saved for the future preference. Text is without problems on hand in many belongings in the structure of documents, newspapers, faxes, printed information, handwritten notes, etc. Many people sincerely scan the report to preserve the records in the computers. When a document is scanned with a scanner, it is saved in the shape of images. But these photographs are no longer editable and it is very hard to find out what the man or woman requires as they will have to go via the entire image, inspecting each line and phrase to determine if it is relevant to their need. Images moreover take up more residence than phrase archives on the computer. It is fundamental to be in a role to maintain this records in such a way so that it will end up less difficult to search and edit the data. There is a growing demand for features that can apprehend characters from scanned archives or captured photographs and make them editable and besides troubles reachable[1].

As analyzing is of excessive magnitude in the day with the aid of day hobbies (text being current in all locations from newspapers, commercial enterprise products, sign-boards, digital shows etc.) of mankind, visually impaired human beings face a lot of difficulties. Our software assists the visually impaired by way of the usage of reading out the textual content to them and additionally to the illiterate[2].

This utility can be useful in many methods they are as follows;

1.1 Digitalizing Documents

An OCR application can convert printed or handwritten archives into digital text format, making it less difficult to store, edit, and share the information

1.2 Saves Time

Rather than manually typing out textual content material from a document, an OCR application can unexpectedly and exactly extract the text, saving time and reducing the hazard of errors. It additionally offers an output of an audio file that can be downloaded and pay attention when in your free time.

1.3 Accessibility

For visually impaired individuals, an OCR application can convert textual content material from a photo into a design that can be learned about aloud via way of text-to-speech convertor application.

1.4 Language translation

For people who are illiterate, OCR application can translate textual content material from one language to another, making it much less tough for human beings to understand and speak with others who talk unique languages.

1.5 Data Extraction

An OCR application can be used to extract precise information from documents, such as names, dates, and addresses, making it much less challenging to analyze and put together the data.

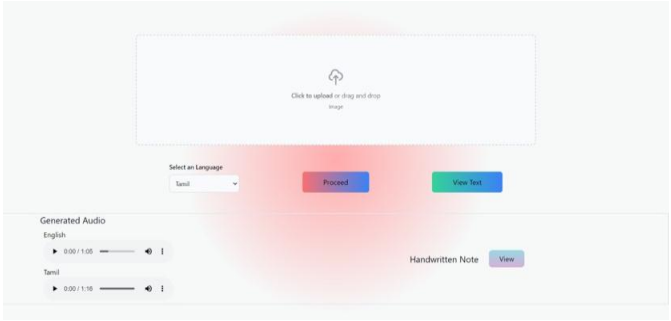

Fig -1: Image to Text Audio Convertor application

2. DETAIL DESCRIPTION OF TECHNOLOGY USED

2.1 Tesseract

Tesseract is an optical character recognition (OCR) engine developed by Google. Its primary purpose is to recognize text embedded in images and convert it into machine-readable text format. Tesseract's proficiency at quickly and accurately identifying written text in a variety of languages, such as English, French, Spanish, German, and many more, is well-known. Applications for Tesseract include data extraction, document management, and machine translation. It is simple to use and implement because it can be integrated into several computer languages, including Python, Java, and C++. Tesseract can recognise text from scanned documents and supports a number of image formats, including JPG, PNG, and TIFF [7].

2.2 Next.js

The well-known open-source web framework Next.js, which is built on React, aids programmers in creating server-side rendered (SSR) and statically generated web apps. Additionally, Next.js has a number of built-in capabilities that improve performance and shorten development times, such as automatic code splitting, hot module replacement, and optimised image loading. It also has a sizable and vibrant community and a variety of plugins and libraries that may be used to increase its capability.

2.3 Flask

A well-liked open-source Python web framework called Flask enables programmers to create web apps quickly and effortlessly. Since Flask is a micro-framework, it is small and doesn't need any specialised libraries or tools to operate. Developers can select the tools and libraries they want to utilise, making it flexible and simple to use. Flask offers a wide range of plugins and extensions, making it simple to add functionality to the application[5].

2.4 Firebase (Cloud Storage)

Google offers developers the Firebase Cloud Storage service, a cloud-based storage solution that enables them to store and serve user-generated material including photographs, videos, and audio files. Built on Google Cloud Storage, a dependable and scalable object storage service, is Firebase Cloud Storage. It is simple to use and integrate Firebase Cloud Storage into online and mobile applications. It offers a straightforward API that enables developers to handle metadata and access control as well as upload and download files. In order to protect user data, Firebase Cloud Storage additionally offers built-in security measures like encryption at rest and in transit.

2.5 googletrans

Using Google Translate to translate text is made simple with the help of the googletrans Python package. Text across different languages is translated using the Google Translate API. Python application developers may rapidly and efficiently translate text across languages using googletrans. More than a hundred languages are supported, including widely used ones like English, Spanish, French, German, and Chinese as well as uncommon ones like Afrikaans, Bengali, and Icelandic. The simplicity and use of googletrans are two of its main advantages. It offers a straightforward API that enables programmers to translate text using very little code. The library takes care of the rest after developers specify the source and target languages.

2.6 gTTS

Using Google's Text-to-Speech API, developers can translate text into spoken language using the gTTS (Google Text-to-Speech) Python package. It offers a simple user interface that makes it possible to create audio files from text in a range of languages. Python programmers may rapidly and simply generate spoken language files from written text with gTTS. It is capable of speaking a broad variety of dialects and languages, including widely used ones like English, Spanish,

French, German, and Chinese as well as less widely used ones like Bengali, Gujarati, and Swahili. The simplicity and usability of gTTS are two of its main advantages. It offers a straightforward API that enables programmers to create text-to-speech conversions using very little code[4].

2.7 PyWhatKit

Python's pywhatkit package deal gives a simple interface for turning text into handwritten notes. It creates images of handwritten notes that appear like handwriting by using the Pillow library. It is easy for developers to use because it is constructed on pinnacle of some of the most widespread Python modules, together with PyAutoGUI, Pillow, and Paperclip. The simplicity and use of PyWhatKit are two of its primary advantages. It affords a simple API that enables programmers to complete difficult duties with a little quantity of code. Developers can use pywhatkit to, for instance, send emails with attachments, convert textual content to handwriting, or even take screenshots.

3. METHODOLOGY

OCR (optical character recognition) is a technology that converts embedded texts from images into a text. Using, Tesseract-OCR library it can extract text from images and that text can be saved in cloud that is Firebase cloud storage. It can take one text input and convert that text into another language text by selecting the language from dropdown using the googletrans API provided by google. And we can also the text in Firebase storage. In the next part, the application is converting the text into audio by using gTTS that is Google text to speech, save it in Firebase and display it on UI(user interface). And then, later it can convert text into handwritten notes by using PyWhatKit library and display it on application.

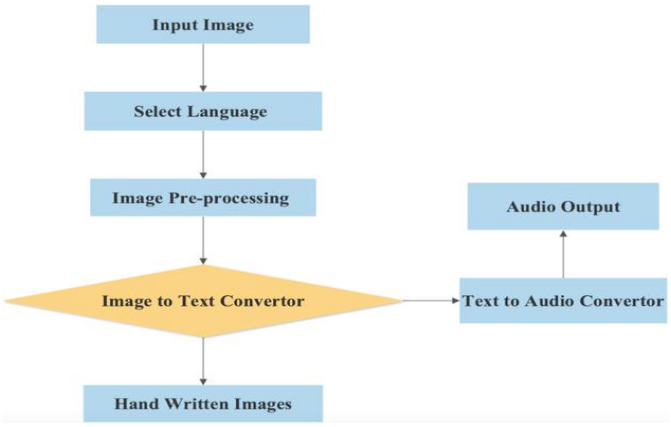

Fig -2: Working Flowchart

3.1 Image Input

To convert and image to text we required data as it contains images in various format. The very first step is to upload the image for pre-processing to extract text content.

3.2 Select Language

To obtain different languages from text we have to select languages to convert the text in various languages obtained from the image.

3.3 Image Pre-Processing

This step consists of shade to grey scale conversion, part detection, noise removal, warping and cropping and thresholding. The photograph is transformed to grey scale as many OpenCV features require the input parameter as a grey scale image. This permits us to become aware of and extract solely that location which carries textual content and eliminates the undesirable background. In the end, Thresholding is accomplished so that the picture appears like a scanned document. This is carried out to permit the OCR to effectively convert the photograph to text.

3.4 Image to Text Convertor

In the given figure(fig.3) suggests the go with the flow of Text-To-Speech. The first block is the photograph pre-processing modules and the OCR. It converts the pre-processed image, which is in .png/jpg/jpeg form, to a .txt file. We are the use of the Tesseract OCR.

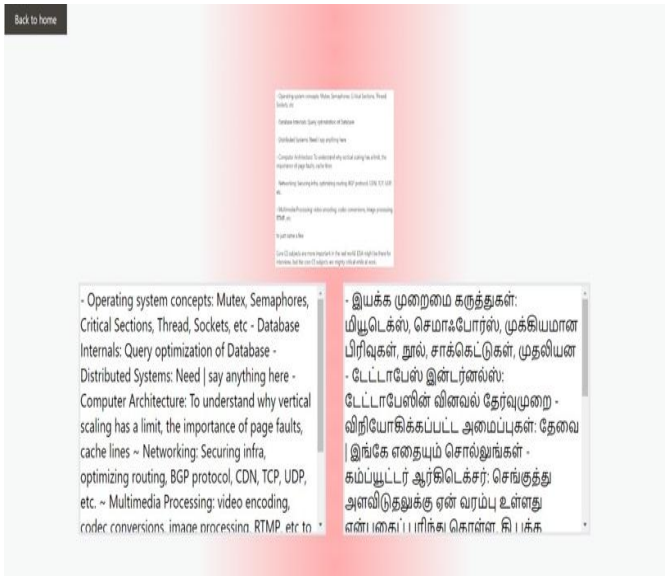

Fig -3: Conversion of English to Various Languages

3.5 Text to Audio-Convertor

In the below (fig 4)it converts the .txt file to an audio output. Here, the textual content is transformed to speech the use of a speech synthesizer. This Audio file can be generated in various other languages.

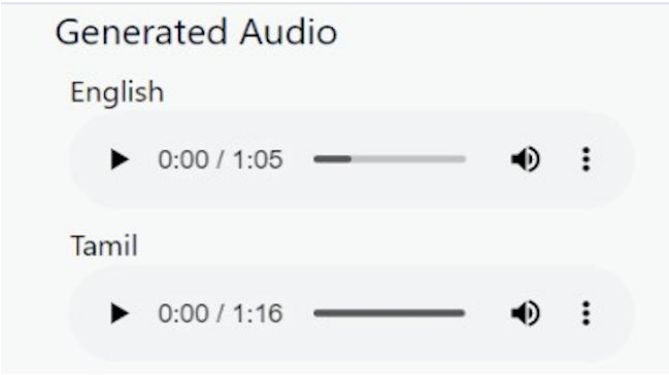

Fig -4: Text to Audio Generator

3.6 Image to Handwritten Notes

This application can also convert the image text into the handwritten format with the help of PyWhatKit.

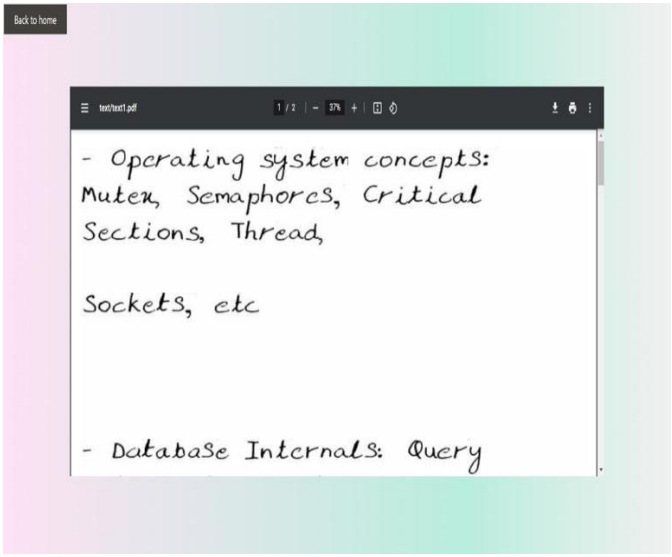

Fig -5: Conversion of Text to Handwritten notes

4. RESULTS

Text extraction from photos is really useful in many true world applications. The data that is saved in textual content is huge and there is favour to store this statistics in such a manner that can be searched except problem every time required. Elimination of the use of paper is one of the steps to improvement in the course of a world of electronics. Also, records that can be changed to audio shape is a way to ease the lives of visually impaired people. Likewise, the textual content material identified can be translated into a variety of languages and can be processed in the chosen language into speech or document. The knowledgeable statistics is created for all on hand fonts and handwritten texts in English so that the OCR will be capable to convert any textual content reachable in the photo into text. The computing device moreover acknowledges textual content in one-of-a-kind patterns or fonts and techniques it to be reachable for pre-mentioned elements such as conversion to speech or document and moreover helps translation.

The website additionally acknowledges handwritten textual content and strategies it to be on hand for pre-mentioned points such as conversion to speech or file and also helps translation. The method of the extraction of the textual content material can be converted into audio its be accuracy of the extraction is extra study to any other technique its be very speedy to carried out and use to the android utility it can be used. The sound excellent of the use of TTS its be good.

This application lets in its user's to understand textual content from pics and convert it into document and speech. The textual content material can be of a quantity of languages and it can moreover be translated to a range of languages. The quintessential function of the system is its potential to convert written textual content material into handwritten notes which can later be transformed to any one-of-a-kind language or into audio file. The conversion of massive volume of images into textual content will make it less difficult for translation and can be used to convert to audio file as nicely as in the structure of handwritten notes as show in the fig. 5

4.1 Performance

The precision-recall curve and F1 score are used to visualise the precision-recall curve and determine the model's performance. A dataset of 100 photos containing ground truth text and associated OCR output from Tesseract is used to evaluate the model.

To evaluate the performance of OCR and calculate the F1 score.

- True positive (TP): The OCR result agrees with the source text.
- False positive (FP): The OCR output differs from the text used as the basis for comparison.
- False negative (FN): The OCR failed to recognise the ground truth text.

The model recognised the text properly in 85 of the images (TP), wrongly in 5 of the images (FP), and not at all in 10 images (FN) as shown in the fig. 6

- Precision = TP / (TP + FP) = 85 / (85 + 15) = 0.9444
- Recall = TP / (TP + FN) = 85 / (85 + 10) = 0.8974
- F1 score = 2 * (precision * recall) / (precision + recall) = 2 * (0.9444 * 0.8974) / (0.9444 + 0.8974) = 0.9189

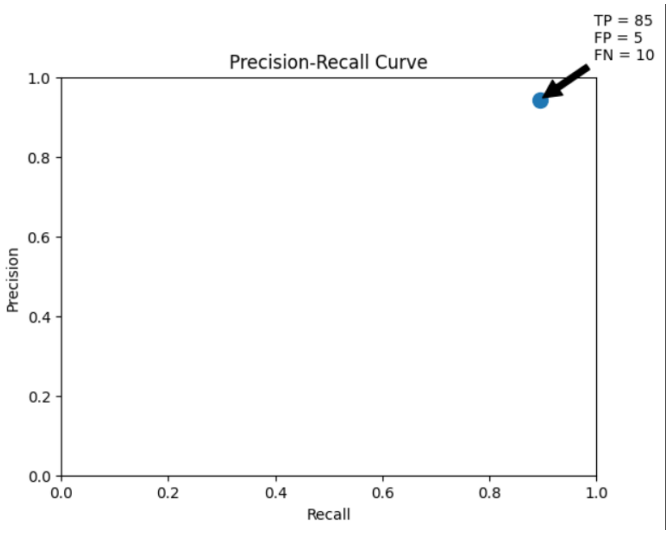

Fig -6: Precision-Recall curve

A confusion matrix is a method for assessing how well a categorization model is working. It displays how many true positives, false positives, false negatives, and true negatives a model correctly predicted for a certain collection of data as show in the fig. 7

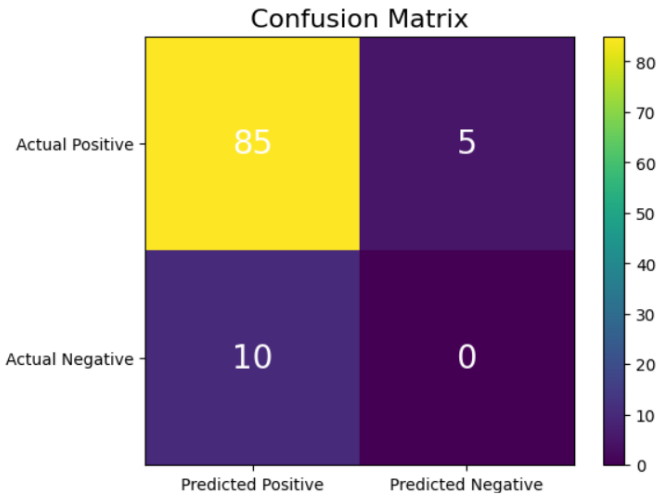

Fig -7: Confusion matrix

5. FUTURE SCOPE

In the future, we can prolong the application via including more languages, exceptional fonts and improve handwritten notes. Various accents can moreover be added for audio data. Initially, we take only one image at a time as an enter in the future we can add a multiple number of images for pre-processing. Not only images we can take any type of videos and break down into frames and that image obtained from the video can additionally be processed. This will help in making subtitles.

6. CONCLUSIONS

Our application helps the users to understand textual content of wide number of languages from images and convert them into Text and later to speech. It also consists of the characteristic of translation of textual content into a variety of languages. Many famous written works can be translated into a number languages for them to attain special people. This approach can take a look at textual content material from a range of sources, and even generate synthesized speech by using audio. It also converts textual content into the form of hand written notes to understand easily as shown in fig (5). It is more convenient to use, it is highly secure, can be used anywhere and accurate.

REFERENCES

[1] Nisha Pawar, Zainab Shaikh, Poonam Shinde, Prof. Y.P. Warke, "Image to Text Conversion Using Tesseract," International Research Journal of Engineering and Technology (IRJET) Feb 02,2019

[2] Asha G. Hagargund, Sharsha Vanria Thota, Mitadru Bera, Eram Fatima Shaik, "Image to Speech Conversion for Visually Impaired," International Journal of Latest Research in Engineering and Technology (IJLRET), ISSN: 2454-5031, Volume 03, June 06 2017

[3] Nivetha.S, Kameshwari.S, "Image to Text and Speech Converter," International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056; p-ISSN: 2395-0072, Volume 07, Issue Nov 11 2020

[4] Arjun Pratap, Kunal Wavhule, Viraj Patil, Vaibhav Narawade, "OCR-WRITTEN TEXT TO AUDIO CONVERTER" IJARIIE, ISSN(O)-2395-4396, 2022

[5] Umatia, S., Varma, A., Syed, A., Tiwari, K., & Shah, F. (2022, November 30). Text Recognition from Images. International Journal for Research in Applied Science and Engineering Technology, 10(11), 1003–1009. https://doi.org/10.22214/ijraset.2022.47498

[6] Kumar Garai, Sayan, Ojaswita Paul, Upayan Dey, Sayan Ghoshal, Neepa Biswas, and Sandip Mondal. "A Novel Method for Image to Text Extraction Using Tesseract-OCR." American Journal of Electronics & Communication 3, no. 2 (2022): 8-11

[7] Lestari, Ikha Novie Tri, and Dadang Iskandar Mulyana. "Implementation of OCR (Optical Character Recognition) Using Tesseract in Detecting Character in Quotes Text Images." Journal of Applied Engineering and Technological Science (JAETS) 4, no. 1 (2022): 58-63

[8] Patil, Shruti, Vijayakumar Varadarajan, Supriya Mahadevkar, Rohan Athawade, Lakhan Maheshwari,

Shrushti Kumbhare, Yash Garg, Deepak Dharrao, Pooja Kamat, and Ketan Kotecha. 2022. "Enhancing Optical Character Recognition on Images with Mixed Text Using Semantic Segmentation" Journal of Sensor and Actuator Networks 11, no. 4: 63. https://doi.org/10.3390/jsan11040063

[9] Karthikeyan G, Bharanidharan G, Jeevanandham D, and Balaji B G. 2022. "Text Recognition Images Using OCR". International Journal of Progressive Research in Science and Engineering 3 (05):57-60.