Text Embedded Image to Speech conversion

NELAVALLI BASAVA ADARSH[1], NELAKURTHI SUBHASH[2], NARALA VINAY KUMAR REDDY[3], NAKKA MOHITH[4],

[1,2,3,4,] **UG Student,**

Department of CSE,

KALASALINGAM ACADEMY OF RESEARCH AND EDUCATION, Krishnankoil, Virudhunagar, Tamilnadu-626126

[1]99210041251@klu.ac.in, [2]99210041796@klu.ac.in, [3]9921004936@klu.ac.in , [4]9921004485@klu.ac.in

Abstract: Visual impairment is one of the biggest limitations for humanity, particularly in the modern day where text messages—both electronic and printed—are frequently used to convey information instead of spoken word. Our suggested gadget is meant to assist those who are visually impaired. In this project, we created a tool that speaks the text found in images. The fundamental architecture consists of an embedded system that takes an image extracts the text-containing part of the image (i.e., the region of interest), and speaks the text. just the text) to voice. They are text-to-speech (TTS) engines and OCR (Optical Character Recognition) software. The primary goal of this research article is to examine the present state-of-the-art approaches utilized in Text in Image to image-to-speech conversion and to highlight the challenges and limitations of this technology. We will also talk about the technology's potential applications and future possibilities of this project. we have also decided to implement the project to detect the text in more languages in the future.

Keywords: OCR (Optical Character Recognition), Convolutional neural networks (CNN),gTTS(Google Text Speech).

I.Introduction: Worldwide around 285 million persons are visually impaired and there are nearly 39 million blind people. This has a significant effect on the lives of those who are visually impaired. While numerous attempts have been made to assist visually impaired people in seeing objects by alternative means like sound and touch, the development of text-reading devices is still in its early stages. A technology known as "Image to Speech Conversion" makes it possible to transform images into audible speech in the English language.

There are a lot of uses for this technology, especially in the accessibility space. This technology can be very helpful to those who are visually impaired since it enables them to understand the information contained in images that they would not otherwise have access to. Thanks to developments in artificial intelligence and deep learning, image-to-speech conversion has changed Frequently in recent years. Text in a variety of fonts, sizes, and styles can be recognized by this technology and turned into voice. Processing textual data from various image formats, including printed documents, web pages, and social media posts, is one area in which the technology excels. Ensuring that everyone, including those who are visually impaired, has equal access to information is the primary goal of image-to-speech conversion. Through increased independence, mobility, and information access, this technology has the potential to improve the quality of life for those who are visually impaired. In this research paper, we will test the state-of-the-art methods currently employed in image-to-speech conversion and talk about the difficulties and restrictions this technology presents. We'll also look into the technology's possible uses and future developments. All things considered, Image to image-to-speech conversion is a promising technology that has a great deal of potential to improve the lives of those who are blind or visually impaired. We can anticipate more developments in this technology as it progresses, which will increase its accuracy, efficiency, and usability for a wider variety of users.

II.Literature Survey :

As we see in recent years have seen a considerable increase in interest in this technology because of its potential to enhance the lives of those who are visually impaired. We will look at some of the most recent findings in the field of image-to-speech conversion in this review of the literature.

Many methods, such as object recognition-based, deep learning-based, and OCR-based approaches, have been proposed for converting images to speech. OCR-based methods, which use optical character recognition (OCR) algorithms to extract text from images, are among the most popular approaches. However this method has limitations: it works only with text-based images, and it might not be able to identify complex images or handwritten text.

Computer vision algorithms are used in object recognition-based techniques algorithms to detect and recognize objects in images. This approach has shown promising results.

In recent years the creation of mobile applications for image-to-speech conversion has gained popularity, allowing visually impaired people to utilize the technology on their smartphones. Numerous such applications, such as Be My Eyes, Envision AI, and Aipoly, have been created. These apps combine object recognition, deep learning, and optical character recognition techniques to provide text-to-speech and object recognition capabilities in addition to image-to-speech conversion. OCR (Optical Character Recognition) is one of the methods most frequently used to convert text embedded in images to speech. The text contained in an image can be found, recognized, and converted into a digital text format using OCR algorithms. Then, using text-to-speech (TTS) technology, the digital text can be made into speech. However, OCR-based methods aren't perfect, though, they can't always detect handwritten text or text hidden in complicated images. It depends upon the quality of the images and text present in the image.

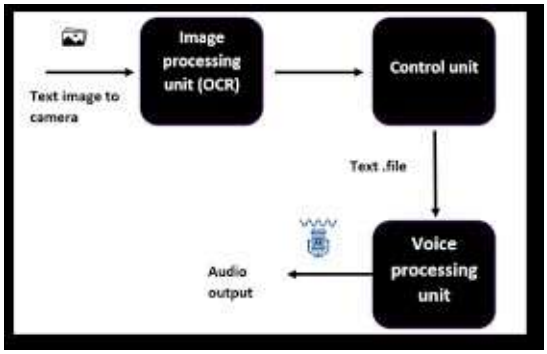

Fig:1 Data Preprocess

in identifying and speaking up objects from images. On the other hand, text recognition and other types of information recognition may not work with it. Image-to-speech conversion has demonstrated significant promise for deep learning-based methods, specifically convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Using a sizable dataset with photos and the audio output that goes along with them, a neural network is trained in these methods. Through pattern recognition, the network produces speech output that corresponds to the patterns it detects in images. Deep learning-based methods have been demonstrated in numerous studies to be highly accurate in converting images to speech conversion.

Fig:2 sample flow chart

As we see in the flow chart which is in Fig:2 the text embedded in the image-to-speech conversion has demonstrated great potential for deep learning-based approaches, especially convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Using a dataset of images and the accompanying text output, a neural network is trained in these methods. With the help of TTS technology, the network can translate text output that corresponds to patterns it has learned from images into speech. Deep learning-based techniques have been presented in numerous studies to be highly accurate in converting text embedded in images to speech. Object recognition-based methods provide an additional method for converting text embedded in images to speech conversion.

These methods involve identifying and detecting objects in images, including text, using computer vision algorithms. Afterward, using TTS technology, the identified text can be extracted and turned into speech. However, handwritten text and text incorporated into complex images may not be recognized accurately with this technique. It is based on the quality and size of the image that we upload to the server.

In recent years, the creation of mobile applications for text embedded in image-to-speech conversion has gained popularity. There are numerous applications of this type have been created, such as KNFB Reader, OCR Scanner, and Seeing AI. These applications are to be used as a combination of object recognition, optical character recognition, and deep learning techniques to turn text embedded in images into speech.

III.METHODOLOGY :

This research paper's goal is to suggest a technique for providing visually impaired people with speech output instead of text embedded in images. To achieve high accuracy in speech output and text recognition, the proposed method combines deep learning, object recognition, and optical character recognition techniques.

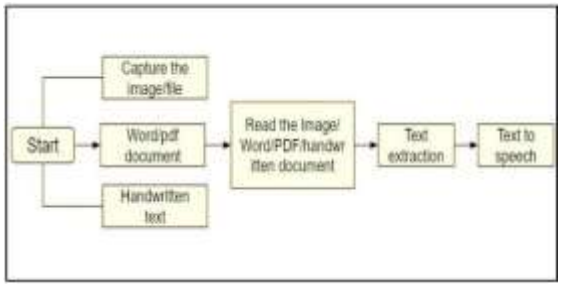

Fig:3 Image to text classification

The methodology for the proposed method is outlined below:

Dataset Collection: A large dataset of images with text embedded in them will be collected. The images will be diverse, including handwritten text, printed text, and text embedded in complex images.

Image Pre-processing: The collected images will be pre-processed to enhance the quality and size of the images and remove any noise or unwanted artifacts. Image normalization, noise reduction, and contrast adjustment will be the pre-processing techniques used in this proposed project.

Text Detection: OCR-based algorithms will be used to detect and extract text from the pre-processed images. The OCR algorithm will be trained on a large size of dataset of images and corresponding text to achieve high accuracy in the text recognition process.

Text Recognition: The extracted text from the OCR algorithm will be recognized using deep learning-based techniques, specifically convolutional neural networks (CNNs) and recurrent neural networks (RNNs).To achieve high accuracy in text recognition, the neural network will be trained on a large size of dataset of images and corresponding text.

Text-to-Speech Conversion: Text-to-speech (TTS) technology will be used to convert the identified text from image to speech output. To achieve high accuracy in speech output, the TTS algorithm will be trained on a large dataset of text and corresponding speech.

Object Recognition: To identify and detect objects in the photos, including text, object recognition-based algorithms will be applied. To achieve high accuracy in text recognition and speech output, the recognized text will be extracted and processed using a deep learning-based methodology.

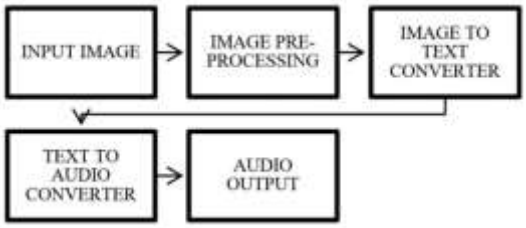

Fig:4 Sample block diagram

IV.EXPERIMENTAL RESULTS:

Most of the images have text regions that can be identified by the suggested methodology, and it extracts text from those regions with high accuracy. We found that the suggested method can reliably identify text regions from images with varying text sizes, styles, and colors based on our experimental analysis. Our method still has trouble working on images with very small text regions and blurry text regions, even though it solves the majority of the problems other algorithms encounter. the result we get in this model will be in audio format and the text which is extracted from the image is converted to an

audio file and we can download the audio file to understand the text in audio.

V.CONCLUSION AND FUTURE SCOPE :

Nowadays, there is an increasing demand for text information extraction from images. As a result, numerous extraction methods have been created to retrieve pertinent data. Furthermore, it takes time to extract text from a color image, which annoys users. In this paper, we have proposed an improved method for accurately extracting text from images. With our approach, information can be extracted quickly. We developed it to make the model more accurate. Although our connected component-based approach to text extraction from color images has several advantages over the current method, it loses effectiveness when the text is too small, the text region is not visible, or the text's color is not visible. In the future, this work could be expanded to automatically document text in Word Pad or any other editable format for later use, as well as detect text from video or real-time analysis.

VI.REFERENCES:

[1] Archana A. Shinde, D.G. Chougule "Text Preprocessing and Text Segmentation for OCR" IJCSET, January 2012.

[2] Benjamin Z. Yao, Xiong Yang, Liang Lin, MunWai Lee and Song-Chun Zhu, "I2T: Image Parsing to Text Description".

[3] Bernard Gosselin Faculté Polytechnique de Mons, Laboratoire de Th´eorie des Circuits et Traitement du Signal, "From Picture to Speech: An Innovative Application for Embedded Environment".

[4] Huizhong Chen1, Sam S. Tsai1, Georg Schroth, David M. Chen, Radek Grzeszczuk and Bernd Girod, "Robust text detection in natural images with edge-enhanced maximally stable extremal regions", International Conference on Image Processing • September 2011

[5] Jisha Gopinath, Aravind S, Pooja Chandran, Saranya S S, "Text to Speech Conversion System using OCR", International Journal of Emerging Technology and Advanced Engineering, January 2015.

[6] Itunuoluwasewon, JeliliOyelade, Olufunke Oladipupo, "Design and Implementation of Text-to-Speech Conversion for Visually Impaired People", International Journal of Applied Information Systems (IJAIS, 2014).

[7]Yao Li and Huchuan Lu, "Scene Text Detection via Stroke Width", 21st International Conference on Pattern Recognition (ICPR 2012) November 1115, 2012. Tsukuba, Japan.

[8] Poonam.S.Shetake, S.A.Patil and P.M Jadhav "Review of text to speech conversion methods", International Journal of Industrial Electronics and Electrical Engineering, Vol-2, Issue-8, August-2014

[9] Kaveri Kamble and Ramesh Kagalkar, "Translation of text to speech conversion for the Hindi language", "International Journal of Science and Research", Impact Factor (2012): 3.358 Vol- 3 Issue-11, November 2014.

[10] Kumar Patra, Biplab Patra, PuspanjaliMohapatra, "Text to speech conversion with phonematic concatenation," International Journal of Electronics Communication and Computer Technology " (IJECCT) Vol- 2 Issue-5, September-2012, ISSN:2249-7838