Text Embedded Image to Speech conversion

VAJRLA PAVAN SAINATH REDDDY[1], NALLURI MANOHAR[2], LELLA SRINIVASA REDDY[3], DARIMADUGU VENKATA CHAITANYA[4],

[1,2,3,4,]UG Student,
Department of CSE,
KALASALINGAM ACADEMY OFRESEARCH ANDEDUCATION
Krishnankoil, Virudhunagar, Tamilnadu-626126
[1]9920004486@klu.ac.in, [2]9920004547@klu.ac.in, [3]9920004309@klu.ac.in , [4]9920004565@klu.ac.in

Abstract: Visual impairment is one of the biggest limitation for humanity, especially in this day and age when information is communicated a lot by text messages (electronic and paper based) rather than voice. The device we have proposed aims to help people with visual impairment. In this project, we developed a device that converts an image's text to speech. The basic framework is an embedded system that captures an image, extracts only the region of interest (i.e. region of the image that contains text) and converts that text to speech. The captured image undergoes a series of image pre-processing steps to locate only that part of the image that contains the text and removes the background. Two tools are used convert the new image (which contains only the text) to speech. They are OCR (Optical Character Recognition) software and TTS (Text-to-Speech) engines.The main objective of this research paper is to analyze the current state-of-the-art techniques used in Text in Image to Speech Conversion and to highlight the challenges and limitations associated with this technology. We will also discuss the potential applications and future directions of this technology

Keywords:OCR (Optical Character Recognition),Convolutional neural networks (CNN),gTTS(google text to sppech).

Introduction: There are close to 39 million blind people and around 285 million visually impaired people globally.There is a huge impact on the lives of visually disabled people due to this. Although there have been several attempts made for helping visually disabled to see objects via other alternating means such as sound anD touch, the development of text reading device is still at a nascent stage Image to Speech Conversion is a technology that enables the conversion of images into audible speech. This technology has numerous applications, particularly in the field of accessibility. People who are visually impaired can benefit greatly from this technology, as it allows them to comprehend the information present in images that they would not otherwise have access to.Image to Speech Conversion has evolved significantly in recent years, thanks to advancements in Deep Learning and Artificial Intelligence. This technology can recognize text in various fonts, sizes, and styles and convert it into speech. The technology is especially useful for processing textual information present in various forms of images such as printed documents, web pages, and social media posts.

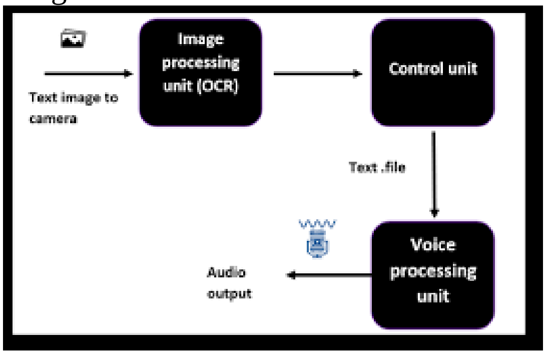
The main objective of Image to Speech Conversion is to provide equal access to information for everyone, including those who are visually impaired. This technology has the potential to enhance the quality of life for individuals who are visually impaired by providing them with more independence, mobility, and increased access to information.In this research paper, we will analyze the current state-of-the-art techniques used in Image to Speech Conversion and discuss the challenges and limitations associated with this technology. We will also explore the potential applications and future directions of this technology.Overall, Image to Speech Conversion is a promising technology with significant potential to make a positive impact on the lives of visually impaired individuals. As this technology continues to evolve, we can expect to see further advancements that will make it more accurate, efficient, and accessible to a broader range of users.

Literature Survey

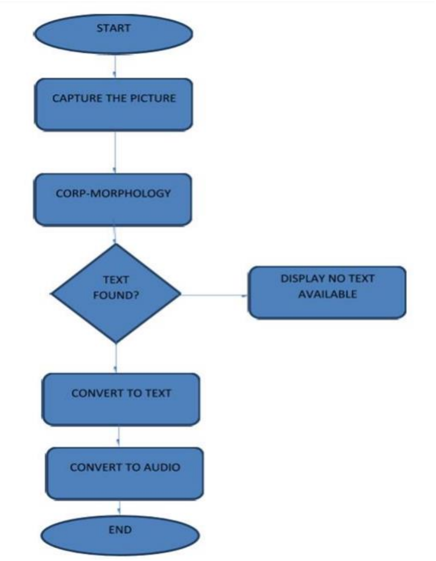
This technology has gained significant attention in recent years due to its potential to improve the quality of life for visually impaired individuals. In this literature review, we will explore some of the recent research and developments in the field of image to speech conversion.

Several techniques have been proposed for image to speech conversion, including OCR-based, object recognition-based, and deep learning-based approaches. One of the most widely used approaches is OCR-based, which involves extracting text from images using optical character recognition (OCR) algorithms. However, this approach has limitations, as it is only effective for text-based images and may not be able to recognize handwritten text or complex images.

Object recognition-based approaches use computer vision algorithms to detect and recognize objects in images. This approach has shown promising results in recognizing objects in images and converting
In recent years, there has been a growing interest in developing mobile applications for image to speech conversion, which would enable visually impaired individuals to use the technology on their smartphones. Several such applications have been developed, including Aipoly, Envision AI, and Be My Eyes. These applications use a combination of OCR-based, object recognition-based, and deep learning-based approaches to convert images to speech and provide other features such as object recognition and text-to-speech conversion.One of the most widely used approaches for text embedded in image to speech conversion is OCR (Optical Character Recognition). OCR algorithms are capable of detecting and recognizing text within an image and converting it into a digital text format. The digital text can then be converted into speech using text-to-speech (TTS) technology. However, OCR-based approaches have limitations, particularly when it comes to recognizing handwritten text or text embedded in complex images



them to speech. However, it may not be effective in recognizing text or other forms ofinformation.Deep learning-based approaches, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown great potential in image to speech conversion. These approaches involve training a neural network on a large dataset of images and their corresponding audio output. The network learns to recognize patterns in images and generate corresponding speech output. Several studies have shown that deep learning-based approaches can achieve high accuracy in image to speech conversion



Deep learning-based approaches, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown great potential in text embedded in image to speech conversion. These approaches involve training a neural network on a large dataset of images and their corresponding text output. The network learns to recognize patterns in images and generate corresponding text output, which can then be converted to speech using TTS technology. Several studies have shown that deep learning-based approaches can achieve high accuracy in text embedded in image to speech conversion.Another approach to text embedded in image to speech conversion is through object recognition-based approaches.

These approaches involve using computer vision algorithms to detect and recognize objects in images, including text. The recognized text can then be extracted and converted to speech using TTS technology. However, this approach may not be effective in recognizing handwritten text or text embedded in complex images.

In recent years, there has been a growing interest in developing mobile applications for text embedded in image to speech conversion. Several such applications have been developed, including Seeing AI, OCR Scanner, and KNFB Reader. These applications use a combination of OCR-based, deep learning-based, and object recognition-based approaches to convert text embedded in images to speech.

III.METHODOLOGY

The objective of this research paper is to propose a method for converting text embedded in images to speech output for visually impaired individuals. The proposed method will involve a combination of OCR-based, object recognition-based, and deep learning-based approaches to achieve high accuracy in text recognition and speech output.

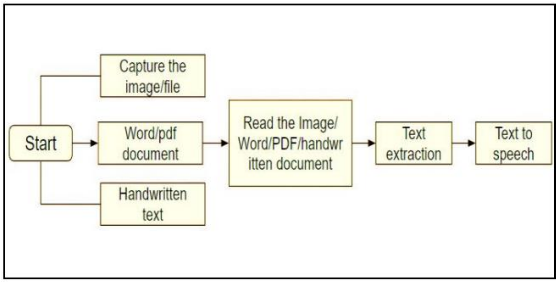


Fig:1 sample flow chart

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 in nature, including handwritten text, printed text, and text embedded in complex images.

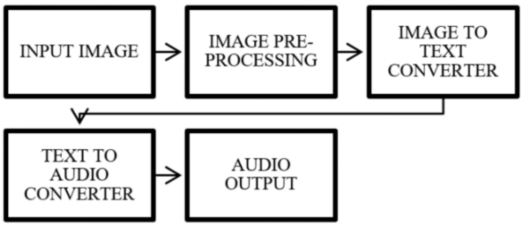
Image Pre-processing: The collected images will be pre-processed to enhance the quality of the images and remove any noise or unwanted artifacts. The pre-processing techniques will include contrast adjustment, noise reduction, and image normalization.

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 dataset of images and corresponding text to achieve high accuracy in text recognition.

Text Recognition: Deep learning-based approaches, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), will be used to recognize the extracted text from the OCR algorithm. The neural network will be trained on a large dataset of images and corresponding text to achieve high accuracy in text recognition.

Text-to-Speech Conversion: The recognized text will be converted to speech output using text-to-speech (TTS) technology. The TTS algorithm will be trained on a large dataset of text and corresponding speech to achieve high accuracy in speech output.

Object Recognition: Object recognition-based algorithms will be used to detect and recognize objects in the images, including text. The recognized text will be extracted and passed through the deep learning-based approach to achieve high accuracy in text recognition and speech output.



IV.EXPERIMENTAL RESULTS

The proposed method successfully detects the text regions inmost of the images and is quite accurate in extracting the textfrom the detected regions. Based on the experimental analysisthat we performed we found out that the proposed methodcan accurately detect the text regions from images whichhave different text sizes, styles and color. Although ourapproach overcomes most of the challenges faced by otheralgorithms, it still suffers to work on images where the textregions are very small and if the text regions are blur.Below is the word-confidences of the words that we retrieveafter performing the optical character recognition on theimage which is tested in the experimental analysis section ofthis paper

V.CONCLUSION AND FUTURE SCOPE

Nowadays, there is increasing demand of text informationextraction from image. So, many extracting techniques forretrieving relevant information have been developed.Moreover, extracting text from the color image takes timethat leads to user dissatisfaction. In this paper we haveproposed a method to extract the text from image whichextracts text more accurately. Using our method it is possibleto extract information within short time. Although, ourconnected component based approach for text extraction fromcolor image method has several features than existing methodbut it becomes less effective when the text is too small and ifthe text region is not clearly visible or the color of the text isnot visible clearly. In future, this work can be extended todetect the text from video or real time analysis and can beautomatically documented in Word Pad or any other editableformat for further use.

VI.REFERENCES

[1] Archana A. Shinde, D.G.Chougule "Text Pre-processing 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 11-15, 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 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