International Journal of Research Publication and Reviews, Vol (5), Issue (5), May (2024) Page – 10283-10287

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

BRAILLE TO TEXT AND VOICE CONVERSION USING CONVOLUTIONAL NEURAL NETWORK

Parth Bangar[1], Saurabh Gadhave[2], Aditya Lokhande[3], Tejas Kapare[4]

[1,2,3,4] Department of Computer Engineering Pune Vidyarthi Griha's College of Engineering and Technology (Savitribai Phule Pune University) Pune , 411009 , Maharashtra , India

A B S T R A C T

In an era of increasing technological advancement, accessibility remains a paramount concern. This project presents a robust and innovative solution for the visually impaired by harnessing the power of computer vision to seamlessly convert Braille into text and voice. The Braille to Text and Voice Conversion System (BTVCS) serves as a transformative tool that bridges the gap between tactile and auditory communication, empowering individuals with visual impairments to access written information independently. The core of the system lies in its ability to accurately recognize Braille characters through image processing and computer vision techniques. Utilizing state-of-the-art machine learning algorithms and deep neural networks, BTVCS interprets Braille patterns with precision, ensuring minimal error rates. The extracted Braille text is then seamlessly converted into natural language text, opening up a world of printed content to the visually impaired. Additionally, BTVCS features a robust text-to-speech (TTS) engine that transforms the converted text into high-quality, natural-sounding voice output. Users can select from a range of voices and settings to personalize their experience. This dynamic TTS capability extends beyond mere translation; it provides an immersive auditory experience that conveys tone, context, and emotion, making the content more engaging and informative.

Keywords: Accessibility, Visually Impaired, Braille, Computer Vision, Text-to-Speech (TTS), Machine Learning, Deep Neural Networks, Image Processing, Natural Language Text, Voice Output, Personalization, Convolutional Neural Network.

INTRODUCTION

Braille, a tactile writing system, has been instrumental in enabling blind individuals to read and write. However, the transition from Braille to digital text and speech remains an essential task. The advent of deep learning techniques, particularly CNNs, has opened up new possibilities for efficient and accurate Braille conversion. The Braille to Text and Voice Conversion System (BTVCS) represents a groundbreaking technological advancement with transformative potential for the visually impaired community. By harnessing cutting-edge techniques, this system seamlessly translates Braille characters into both text and spoken language, significantly enhancing accessibility, communication, and participation in society. Braille communication has a long and fascinating history. Louis Braille invented the Braille system in 1824, and it has been an essential tool for the visually impaired ever since. However, Braille has its limitations. It can be challenging to learn, and not all information is available in Braille format. The BTVCS system uses Convolutional Neural Network technology to convert Braille into text and voice. This technology has been developed over several years and is continually improving. The BTVCS system has many advantages, including increased accessibility to information, reduced costs, and improved communication. However, it also has some limitations, such as the need for a Braille display device and the high cost of the technology.

With the BTVCS system, the visually impaired can access information that was previously unavailable to them, such as online books, articles, and websites. This technology has also made it easier for the visually impaired to communicate with others, whether it be in-person or through electronic devices. Furthermore, the BTVCS system has opened up new opportunities for the visually impaired in the workplace, allowing them to participate in a broader range of jobs and industries. The main objective is to create a complete system that will facilitate the translation of Braille to standard languages like English and then to voice, using the user's earphone or speaker connected to the computer. The significance of the system lies in its ability to empower the visually impaired by providing a means of understanding and communicating in a world that is dominated by visual signs. By making technology more accessible to the visually impaired, this will certainly benefit them and may even open up new opportunities and possibilities to improve their quality of life.

LITERATURE SURVEY

Braille in the early 19th century, revolutionized communication for blind individuals. However, the transition from Braille to digital text and spoken language remained a challenge. In recent decades, advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have paved the way for efficient and accurate Braille conversion. Braille alphabet is composed of 64 dot patterns where each pattern may contain one or more individual dots. In general, one cell includes 2 vertical columns and 3 horizontal rows of dots. It is known as 2*3 dot matrix cells. But, this design is varied in different languages. The characters in the cells are formed by the patterns. And then these patterns are recognized by the Braille system as characters. The characters in Braille system are not just alphabet letters; they also include numbers, punctuations, arithmetic and other symbols. So, variety designs of Braille system are also provided based on the different demands and requirements of them. For instance, we have the computer Braille codes, mathematics Braille system and etc.

The conversion of braille to text and voice is essential for providing access to literacy and information for individuals with visual impairments, with Convolutional Neural Networks (CNNs) emerging as a promising solution in recent years. CNNs, inspired by the visual cortex organization, have shown remarkable success in various image processing tasks, making them well-suited for recognizing and interpreting braille patterns accurately. By leveraging large-scale datasets and advanced architectures, researchers have developed CNN-based models capable of seamlessly converting braille symbols into textual representations and synthesizing natural-sounding speech. Despite notable advancements, challenges such as handling variations in braille formatting and improving speech synthesis robustness remain. However, the integration of CNNs into end-to-end conversion systems holds great promise for enhancing accessibility and empowering visually impaired individuals in navigating the digital world independently, highlighting the importance of continued research and innovation in this field.

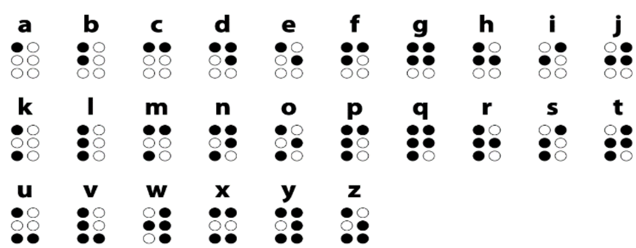

Fig. 1 – Braille Characters

METHODOLGY

Model - Mobile-Net V2

MobileNetV2 is tailor-made for the demands of mobile devices, making it an ideal solution for Braille conversion systems intended for smartphones and tablets. Its lightweight architecture, featuring an inverted residual structure and efficient depthwise convolutions, ensures minimal computational burden without compromising on representation power. This design facilitates real-time processing of Braille images, catering to on-the-go applications where responsiveness is paramount. Moreover, MobileNetV2 strikes a delicate balance between accuracy and resource usage, crucial for mobile hardware with limited computational resources, ensuring fast inference times for seamless user experiences.

By leveraging pre-trained checkpoints on ImageNet-1k and fine-tuning them with Braille data, MobileNetV2 is equipped to handle the intricacies of Braille recognition tasks effectively. Its robust feature extraction capabilities, particularly through depthwise convolutions, enable the model to adapt to varying Braille formats, accommodating differences in embossing techniques, language-specific rules, and contractions. Additionally, MobileNetV2 seamlessly integrates with text-to-speech (TTS) models, completing the user interaction loop by providing both text and voice outputs. This comprehensive approach fosters inclusivity and enhances communication for visually impaired individuals, making MobileNetV2 a standout choice for Braille to Text and Voice Conversion systems.

FLOWCHART:

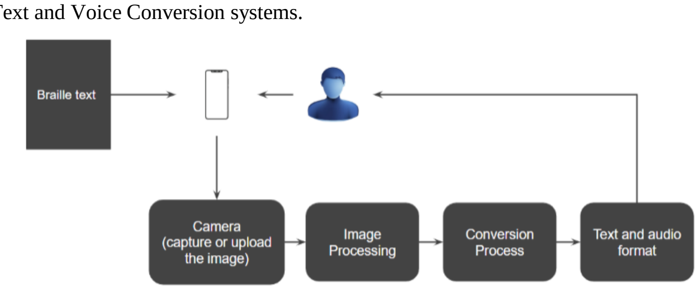

Fig.2-Flowchart of Braille to Text Conversion

3.1 Feature Extraction

In the preprocessing phase of Braille to text and speech conversion using Convolutional Neural Networks (CNNs), several essential steps are involved to enhance the quality and extract meaningful features from the input images. Initially, the raw input image undergoes image processing techniques to improve its clarity and remove any potential noise. This may involve operations such as noise reduction, contrast adjustment, and normalization to ensure consistent and optimal input for subsequent processing. Following this, image sharpening techniques are applied to enhance the edges and details of the Braille characters, making them more distinct and easier for the CNN to recognize.

Subsequently, edge detection algorithms are employed to identify and highlight the edges of the Braille dots and characters within the image. This step helps in segmenting the individual Braille characters from the background, facilitating more accurate recognition by the CNN model. Finally, feature extraction techniques are utilized to extract discriminative features from the preprocessed images, which are essential for distinguishing between different Braille characters. These features may include the shape, size, and spatial arrangement of the dots, as well as any other distinctive characteristics unique to Braille patterns. By effectively preprocessing the input images through image processing, sharpening, edge detection, and feature extraction, the CNN model can better interpret and convert Braille images into text and speech, thereby enhancing accessibility for visually impaired individuals.

Pre processing : the image is converted to the appropriate size for the model which is (28,28)

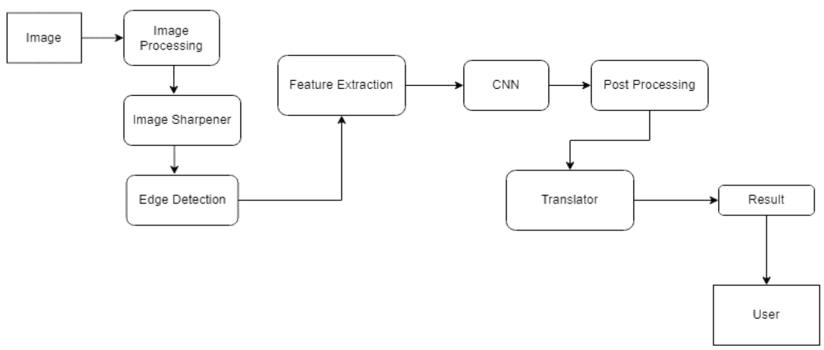

Fig.3

For Braille dot prediction, a character-based method is employed. The model is trained on a Braille character dataset, achieving 100% accuracy in predicting single characters. The image is segmented into parts containing Braille characters, each passed to the recognition model. This iterative process predicts characters individually, storing them in a string. This approach ensures accurate identification of Braille characters within the image, enabling comprehensive conversion to text or speech for visually impaired users.

3.2 Post Processing

Post processing: post processing contains three processes

1.Calculating the space between the words: In the post-processing phase, determining the spacing between words involves computing the average distance between consecutive characters recognized in the converted text. This average distance is then used to establish a threshold, typically set at 1.3 times the computed average. When analyzing the distances between characters, if the distance between two adjacent letters exceeds this threshold, it suggests the presence of a space between those characters, indicating the boundary between two words in the text. This method ensures accurate segmentation of words, contributing to the readability and coherence of the final output.

2. Translation: Translation of the recognized English text into multiple languages is facilitated through the utilization of the Google Translate API. This application programming interface (API) provided by Google allows developers to integrate automated translation capabilities into their software or systems. When applied to the converted English text, the Google Translate API automatically translates it into various languages according to the user's preferences or specified settings. This enables the system to provide accessibility and comprehension to users from diverse linguistic backgrounds by offering translations in multiple languages, thereby enhancing inclusivity and usability of the Braille conversion system.

3. Text to speech: **pyttsx3** simplifies text-to-speech conversion in Python by providing a straightforward interface to the system's speech synthesis capabilities. It abstracts the complexities of interacting with different operating systems' speech engines, enabling developers to focus

on their application logic. With **pyttsx3**, you can easily convert textual content into spoken words within your Python scripts, making it ideal for various applications, including accessibility tools, educational software, and interactive voice interfaces. The library allows customization of speech properties such as speaking rate and volume, providing flexibility in speech output. By integrating **pyttsx3**, developers can create more inclusive and interactive applications that cater to users with visual impairments or those who benefit from auditory feedback. Moreover, its open-source nature and active community support ensure ongoing development and compatibility with future Python releases, making it a reliable choice for text-to-speech conversion needs.

In the post-processing phase, we employ a method to locate spaces within the text by analyzing the distances between letters. This approach hinges on measuring the gap between consecutive characters, determined by calculating the disparity between their respective bounding box centers. By computing the distance between the bounding box centers of each character pair, we discern the spaces between words. Essentially, this involves determining the separation between the first and second characters' bounding box centers. This method effectively delineates word boundaries, facilitating accurate text segmentation and subsequent processing for tasks such as natural language understanding or synthesis.
x1, y1, x2, y2 are the coordinates of the bounding box.

center_x = (x1 + x2) / 2

center_y = (y1 + y2) / 2

Following the calculation of distances between adjacent characters, we utilize a thresholding technique. Here, if the disparity between distances exceeds a predetermined threshold, it signifies a space. This threshold acts as a delimiter, allowing us to identify and insert spaces into the sentence appropriately. By comparing the differences between consecutive distances against this threshold, we effectively discern word boundaries and augment the sentence accordingly. This method ensures accurate space allocation within the text, enabling seamless conversion from Braille to natural language text, thereby enhancing the accessibility and usability of the system for individuals with visual impairments.

x - previous of x > threshold: space at i position

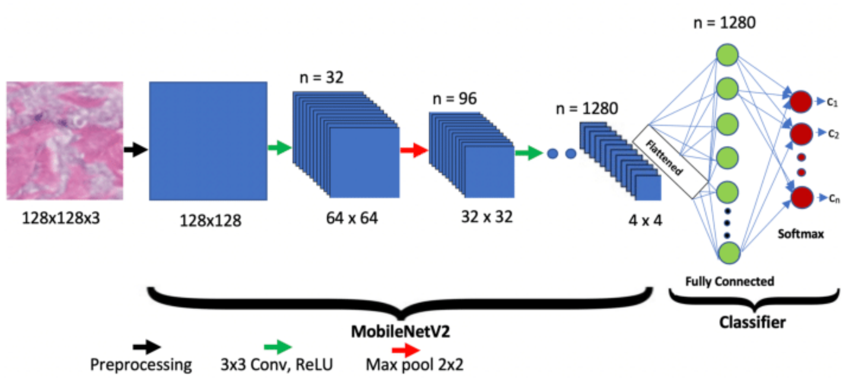

Fig.4 Convolutional Neural Network

Conclusion

The project on Braille to text and speech conversion using Convolutional Neural Networks (CNN) represents a significant advancement in enhancing accessibility and communication for visually impaired individuals. By seamlessly converting Braille characters into natural language text and high-quality voice output, this system bridges the gap between tactile and auditory communication, empowering the visually impaired to access written information independently. Through the utilization of state-of-the-art machine learning algorithms and deep neural networks, the system ensures precise recognition of Braille patterns with minimal error rates, offering a transformative tool for the visually impaired community. The developed system not only facilitates the translation of Braille into standard languages like English but also provides a dynamic text-to-speech engine that conveys tone, context, and emotion, enhancing the overall auditory experience. By making technology more accessible and inclusive, this project opens up new opportunities and possibilities to improve the quality of life for visually impaired individuals, emphasizing the importance of continued research and innovation in this field to further enhance accessibility and empowerment.

REFERENCES

1. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H, arXiv:1704.04861, 2017.

2. MobileNetV2: Inverted Residuals and Linear Bottlenecks, Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. arXiv preprint. arXiv:1801.04381, 2018.

3. Speed/accuracy trade-offs for modern convolutional object detectors, Huang J, Rathod V, Sun C, Zhu M, Korattikara A, Fathi A, Fischer I, Wojna Z, Song Y, Guadarrama S, Murphy K, CVPR 2017.

4. Scientific Reports (nature.com)

5. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385, 2015.

6. Mathivani ‘Braille Language converts for visually impaired people’ International Journal of Intellecutal Advancement and Research in Engineering Computations, India, vol-06, Issue-02.

7. Article:99 June Volumen 2020 [Online], Available (https://www.gjstx-e.com)

8. A survey of text detection and recognition algorithms based on deep learning technology Xiao-Feng Wang, Zhi-Huang He , Kai Wang , Yi Fan Wang, Le Zou , Zhi-Ze Wu 2023 Version of Record 25 August 2023.