e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

A REVIEW ON THE SIGN SPEAK WITH REAL TIME GESTURES

Dr. A.D. Bhange[*1], Harsh Huke[*2], Divya Motghare[*3], Digeshwar Parshuramkar[*4], Himani Hatwar[*5], Gaurav Shrinath[*6], Yash Wasalwar[*7]

[*1]Professor, Computer Science And Engineering, K.D.K. College Of Engineering, RTMNU, Nagpur, Maharashtra, India.

[*2,3,4,5,6,7]Computer Science And Engineering, K.D.K. College Of Engineering, RTMNU, Nagpur, Maharashtra, India.

ABSTRACT

The Sign Speak project aims to bridge the communication gap between individuals who use sign language and those who do not understand it by translating real-time sign language into speech. Utilizing the YOLO algorithm, the system detects and recognizes hand gestures corresponding to the alphabets A to Z and predefined phrases such as "Hi," "Hello," "Okay," "I love you," and "Bye.", etc. Once detected, the gestures are converted into text and then into coherent sentences. The system encompasses three modules for different input sources: images, videos, and live camera feeds. Python libraries are used to facilitate the automated generation of sentences from detected signs. This project offers an inclusive and accessible solution, enabling individuals who are unable to speak to communicate their thoughts effectively to a wider audience.

Keywords: Utilizing The YOLO Algorithm, Python Libraries (Q Rious).

I. INTRODUCTION

Communication is an essential part of human interaction, but it can be challenging for those who use sign language to communicate with people who do not understand it. The Sign Speak project aims to solve this problem by creating a system that can translate sign language into spoken words in real time. This way, people who are unable to speak can easily convey their thoughts to others.

Sign Speak uses a powerful object detection model called the YOLO algorithm to recognize hand gestures. These gestures represent the letters of the English alphabet (A to Z) and common phrases like "Hi," "Hello," "Okay," "I love you," and "Bye.", etc. The system then converts these gestures into text and combines them to form sentences. Finally, the text is turned into speech, allowing the user's message to be heard aloud.

The project includes three modules that can work with different types of input: images, videos, and live camera feeds. This flexibility makes Sign Speak suitable for various situations, whether it's a recorded video or a real-time conversation. Additionally, the system uses Python libraries to help create complete sentences from the recognized signs.

In summary, Sign Speak is designed to give a voice to those who use sign language, helping them communicate more effectively with people who don't know how to sign. This project takes an important step toward a more inclusive world where everyone can express themselves clearly.

II. METHODOLOGY

Data Collection and Preparation:

Gather a comprehensive dataset of sign language gestures representing the English ABC(A to Z) and predefined expressions like" Hi,"" Hello,"" Okay,"" I love you," and" Fine." Annotate the dataset with markers to indicate the matching letter or expression for each gesture. Perform data addition ways to increase the diversity of the dataset, icing the model's robustness to different lighting conditions, backgrounds, and hand exposures.

Model Training:

Model Training use the YOLO(You Only Look formerly) algorithm for object discovery and train the model on the set dataset. Split the dataset into training, confirmation, and test sets to estimate the model's performance. Optimize the model by tuning hyperparameters and using ways similar as early stopping and literacy rate scheduling to help overfitting.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

Development of Module:

Gesture Recognition Module apply the trained YOLO model in a module that can reuse inputs from images, videos, and live camera feeds. This module will descry and fete hand gestures in real time.

Text Generation Module:

Develop a module that converts the recognized gestures into text. Utilize natural language processing techniques to assemble detected gestures into coherent sentences.

Speech Synthesis Module: Integrate a text-to-speech engine that converts the generated text into spoken words, providing an audible output of the translated sign language.

User Interface Module: Design and implement a user-friendly interface that allows users to select the input type (image, video, or camera), view recognized gestures and corresponding text, and listen to the synthesized speech.

Integration and Testing:

Integrate all modules into a cohesive system, ensuring smooth data flow and interaction between modules.

Conduct thorough testing of the system to validate its performance and accuracy in translating sign language gestures into speech.

Perform user testing with individuals who use sign language to gather feedback and make necessary improvements.

III. MODELING AND ANALYSIS

Gesture Recognition Module: This module employs the YOLO algorithm to detect and identify hand gestures representing the letters A to Z and predefined phrases such as "Hi," "Hello," "Okay," "I love you," and "Fine." It processes the input from images, videos, and live camera feeds, accurately recognizing and classifying the gestures in real-time.

Text Generation Module: Once gestures are recognized, this module converts them into text. It assembles the detected signs into meaningful words and sentences using Python libraries, ensuring that the generated text accurately reflects the intended message.

Speech Synthesis Module: The text generated by the previous module is then converted into spoken words by this module. It uses text-to-speech technology to produce clear and natural-sounding audio output, allowing the translated sign language to be heard by others.

User Interface Module: This module provides a user-friendly interface for interacting with the system. It allows users to choose the type of input (image, video, or camera), displays the recognized gestures and corresponding text, and plays the synthesized speech. The interface is designed to be intuitive and accessible, making it easy for users of all skill levels to operate the system.

Video Module

The Video Module is designed to process pre-recorded video files containing sign language gestures. It extracts frames from the video and applies the YOLO algorithm to detect and recognize hand gestures in each frame. The module then sends the recognized gestures to the Text Generation Module, which converts them into text. This allows users to translate sign language from any recorded video into spoken words, making it accessible to those who do not understand sign language.

Camera Module

The Camera Module facilitates real-time sign language detection and translation by using a live camera feed. It continuously captures frames from the camera and processes them using the YOLO algorithm to detect and identify hand gestures. This real-time recognition allows the system to generate and synthesize speech immediately as the gestures are performed, enabling seamless communication between individuals who use sign language and those who do not.

Image Module

The Image Module processes individual images containing hand gestures. Users can upload or capture still images of sign language gestures, which the module then analyzes using the YOLO algorithm to identify the gestures. Once identified, the gestures are converted into text by the Text Generation Module and then into

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

speech by the Speech Synthesis Module. This feature is useful for translating static images of sign language into a format that is understandable to a wider audience.

IV. CONCLUSION

The "Sign Speak with Real-Time Gestures" project successfully addresses the communication barrier between the hearing-impaired community and the general public by providing an innovative solution for translating sign language gestures into spoken language. Through the integration of computer vision, machine learning, and real-time processing, the system is able to accurately recognize and translate sign language gestures into text or voice output.

Key outcomes and contributions of the project include:

Enhanced Accessibility: The system allows users to communicate seamlessly with individuals who do not know sign language, improving social integration and accessibility for the hearing-impaired.

Real-Time Processing: By utilizing real-time gesture recognition, the project enables instantaneous communication, which is crucial for effective and natural conversations.

User-Friendly Interface: The system's design prioritizes ease of use, with intuitive interactions and minimal hardware requirements, ensuring that users can comfortably engage with the technology.

Scalability and Adaptability: The platform can be further developed to support multiple sign languages, making it adaptable for a global audience and expanding its usability in diverse linguistic contexts.

In conclusion, the "Sign Speak with Real-Time Gestures" project represents a significant step toward bridging communication gaps, empowering individuals with hearing impairments to communicate freely and effectively with society. With continued refinement, it holds the potential to become a transformative tool in promoting inclusivity and accessibility in both personal and professional environments.

V. REFERENCES

[1] Faisal, M.A.A., Abir, F.F., Ahmed, M.U. et al. Exploiting domain transformation and deep learning for hand gesture recognition using a low-cost dataglove. Sci Rep 12, 21446 (2022).

[2] Zhang, Y. et al. Static and dynamic human arm/hand gesture capturing and recognition via multiinformation fusion of flexible strain sensors. IEEE Sens. J. **20**, 6450–6459 (2020).

[3] Adithya, V. & Rajesh, R. A deep convolutional neural network approach for static hand gesture recognition. Proc.Comput.Sci. 171,23532361. https://doi.org/10.1016/j.procs.2020.04.255 (2020).

[4] Hill, J. C., Lillo-Martin, D. C. & Wood, S. K. Sign Languages: Structures and Contexts 1st edn, Ch. 1 (Routledge, 2019).4:50 PM11/14/2024