

INCLUSIVE PLATFORM FOR THE VISUALLY HEARING-IMPAIRED STUDENTS

Dr. Mohanapriya N
Department of
Computer Science and
Engineering
Vivekanandha College
of Engineering for
Women Namakkal,
India

Ms. Sinduja S
Department of
Computer Science and
Engineering
Vivekanandha College
of Engineering for
Women Namakkal,
India

Ms. Misha D
Department of
Computer Science and
Engineering
Vivekanandha College
of Engineering for
Women Namakkal,
India

Ms. Neha Sharma
Department of
Computer Science and
Engineering
Vivekanandha College
of Engineering for
Women Namakkal,
India

Ms. Sowmiya C M
Department of
Computer Science and
Engineering
Vivekanandha College
of Engineering for
Women Namakkal,
India

Abstract- Sign language is the primary mode of communication for speech and hearing-impaired individuals. However, understanding sign language is often challenging for those unfamiliar with it. This project aims to bridge the communication gap between sign language users and non-signers through an automated real-time sign language interpretation system.

The proposed system utilizes Python programming with OpenCV for video processing and deep learning algorithms like YOLOv5 and Convolutional Neural Networks (CNN) for gesture recognition. The system captures real-time video input using a webcam, detects hand gestures, classifies them, and converts them into readable text instantly on the screen.

This system can recognize various hand gestures and interpret alphabets or words based on the sign shown. The accurate and real-time translation of sign language will make communication easier and more effective for deaf and mute individuals, promoting inclusivity in daily life. Future enhancements may include voice output, support for multiple sign languages, and mobile-based applications.

I. INTRODUCTION

Sign language is a visual language used by individuals with speech and hearing impairments to communicate with others. It plays a vital role in conveying thoughts, ideas, and information through specific hand gestures and movements. However, many people are unaware of sign language gestures, which leads to communication barriers between sign language users and non-signers.

In today's world, technology can be utilized to eliminate this gap by developing systems that can interpret sign language and convert it into understandable text or speech

In the existing scenario, deaf and mute people often face difficulties while communicating with normal people, as their language is not easily understood by everyone. The traditional method involves either learning sign language or having an interpreter in between, which is not always possible or convenient. This increases the dependency on others, reduces confidence, and makes communication less effective. Hence, an automated system that recognizes sign language gestures in real-time and converts them into readable text is essential for better communication.

To overcome these limitations, this project proposes a Real-Time Sign Language Recognition System using Python, OpenCV, YOLOv5, and CNN. The system captures real-time video through a webcam, detects hand gestures, and converts them into text output instantly. YOLOv5 helps in detecting hand regions, while CNN classifies the gestures accurately.

This system bridges the communication gap between hearing and speech-impaired people and others. It ensures easier and faster communication and can be enhanced further with features like voice output, multi-language support, and mobile application development.

II. RELATED WORKS

Sign language plays a crucial role in the communication of speech and hearing-impaired individuals. Several researchers have contributed to the development of sign language recognition systems using various machine learning and deep learning techniques.

In the earlier stage, sign language recognition was implemented using simple image processing techniques like edge detection and skin color segmentation. However, these methods faced many challenges in real-time applications.

Many researchers have proposed systems that use contour detection and background subtraction techniques to identify hand gestures. But these techniques were affected by the changing background and lighting conditions.

To overcome these limitations, machine learning-based algorithms like Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) were used for classifying hand gestures based on extracted features from images. Some researchers have implemented the use of Haar Cascade Classifier for detecting hand gestures in static images. But this method was not suitable for dynamic hand movements in real-time scenarios.

Later, Convolutional Neural Network (CNN) models were introduced in the field of sign language recognition to overcome the drawbacks of manual feature extraction methods. CNN has the ability to automatically learn important features from the dataset during the training process without any manual effort. This deep learning approach enabled the system to extract spatial features like hand shape, position, and movement patterns more effectively from the input images or video frames. As a result, CNN-based gesture recognition systems provided better accuracy and improved performance in identifying complex hand gestures when compared to traditional image processing techniques. This advancement helped researchers to develop more accurate and efficient models for recognizing dynamic hand gestures used in sign language communication.

Although CNN provided higher accuracy in recognizing hand gestures, it faced significant challenges when applied to real-time gesture detection scenarios. The major issue was the time consumption involved in processing every video frame individually, which reduced the overall speed and efficiency of the system during continuous video streaming. To overcome this drawback and to enhance real-time performance, researchers started implementing advanced object detection algorithms like SSD (Single Shot Detector) and Faster R-CNN. These algorithms were designed to detect multiple objects, such as hand gestures, from

continuous video frames with better speed and accuracy. They allowed the system to identify and locate gestures directly within the video stream rather than processing each frame separately. Despite these improvements, SSD and Faster R-CNN still required higher computational resources, which limited their usage in devices with low processing capabilities.

However, these models required high computational resources for their execution, which made them unsuitable for real-time performance on low-end devices or systems with limited hardware capabilities. The need for high processing power and memory limited their practical usage, especially in portable or mobile applications.

To overcome these limitations and to achieve faster and more accurate gesture detection, recent research works have shifted their focus towards the YOLO (You Only Look Once) family of algorithms. YOLO models are highly popular in the field of real-time object detection due to their excellent balance between speed and accuracy.

Among these, YOLOv5 has gained significant attention from researchers for its ability to detect objects like hand gestures quickly and effectively even in dynamic and complex environments. The efficiency of YOLOv5 in handling real-time video streams made it an ideal choice for gesture recognition systems designed for sign language interpretation. YOLOv5, which is the latest and most advanced version of the YOLO (You Only Look Once) family, has gained widespread attention from many researchers in recent years due to its remarkable ability to detect objects quickly and accurately in real-time environments. In the field of sign language recognition, detecting hand gestures with precision and speed is highly important for smooth communication. YOLOv5 has proven to be highly efficient in this aspect, as it can detect multiple hand gestures from continuous video input without any delay. Several research studies have explored the idea of combining YOLOv5 for accurate hand detection with Convolutional Neural Networks (CNN) for effective gesture classification.

This combination allows the system to not only detect the hand position in real-time but also classify the detected gesture into specific alphabets, words, or numbers used in sign language. This hybrid approach ensures better accuracy,

This hybrid approach ensures better accuracy, speed, and performance in developing a reliable sign language recognition system suitable for real-world applications.

Researchers have also explored the integration of Natural Language Processing (NLP) techniques into sign language recognition systems to enhance the communication experience. NLP techniques help in converting the sequence of recognized gestures into meaningful sentences, allowing the conversation to be more natural and understandable for non-sign language users.

In addition to this, several studies have proposed the incorporation of voice output features into the system. This enables the system to not only display the recognized text on the screen but also convert the text into speech, providing an audio output that makes the communication more interactive and user-friendly. Such advancements are highly beneficial for real-world scenarios where quick and effective communication is essential.

Moreover, current research in the field of sign language interpretation is primarily focused on developing systems that are accurate, real-time, and easy to use. By utilizing advanced deep learning algorithms like YOLOv5 for gesture detection and CNN for gesture classification, researchers aim to build practical solutions that can assist speech and hearing-impaired individuals in their daily life communication without any dependency on third persons. Many studies are now moving towards building mobile-friendly and lightweight models to run efficiently on smartphones for real-time sign language interpretation. These improvements are paving the way for smart, affordable, and portable solutions for sign language communication in everyday life.

III. EXISTING SYSTEM

In the existing system, sign language communication mainly happens with the help of human interpreters or by teaching sign language to non-signers, which is not always practical in real-time situations. Deaf and mute individuals mostly depend on an interpreter for communicating their thoughts and ideas to others. This dependency often leads to difficulties in situations where an interpreter is unavailable. Some traditional systems were developed using basic image processing techniques like contour detection, edge detection,

and background subtraction methods to identify hand gestures. These methods worked efficiently only in controlled environments with a fixed background and good lighting conditions. However, their performance drastically reduced in dynamic real-time scenarios.

Existing systems used machine learning algorithms like SVM and KNN for hand gesture classification, but they required manual feature extraction which made the process difficult and less accurate. Later, CNN was used for better accuracy, but it struggled with real-time video processing due to slow frame-by-frame detection. Moreover, these systems failed to work properly in dynamic backgrounds, poor lighting, and with fast hand movements. Most systems could recognize only static gestures, making them unsuitable for real-time sign language communication.

IV. PROPOSED SYSTEM

The proposed system aims to overcome the limitations of existing systems by developing a real-time sign language recognition system using Python, OpenCV, YOLOv5, and Convolutional Neural Networks (CNN). This system is designed to detect and recognize hand gestures quickly and accurately from real-time video input and convert them into readable text.

In this system, YOLOv5 is used for detecting hand gestures effectively from the video frames. YOLOv5 helps to detect the hand region even in dynamic environments and varying lighting conditions. After detecting the hand, CNN is used to classify the gestures into corresponding alphabets or words of sign language.

The main advantage of the proposed system is its ability to detect and recognize hand gestures in real-time with high speed and accuracy. It works efficiently in different environments without being affected by background or lighting conditions. The system is user-friendly and eliminates manual feature extraction. It can also be enhanced in future with voice output, multiple language support, and mobile application for easier communication between speech and hearing-impaired people and normal users.

This system does not require any wearable devices or sensors, making it simple, cost-effective, and easy to use. It only needs a webcam and computer to detect both static and dynamic gestures accurately in real-time without delay.

In future, the system can be improved by adding voice output, multiple sign language support, and mobile application development, helping to reduce the communication gap between speech and hearing-impaired people and normal users for easier daily interaction.

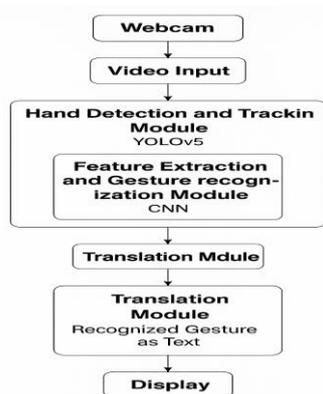


Figure 1 - Block Diagram

V. METHODOLOGY

Hand detection and tracking

The Hand Detection and Tracking Module is responsible for detecting the hand and tracking its movements from the real-time video input. A webcam captures the video of the user performing sign language gestures. The captured video is processed using OpenCV to extract each frame. YOLOv5, a powerful object detection model, detects the hand accurately and draws bounding boxes around it. This module tracks the hand continuously even if the user moves or changes position. It focuses only on the hand region while ignoring the background to ensure better accuracy. It helps the system to perform smoothly in real-time. This module plays an important role in detecting hand gestures effectively for further processing. It provides the base for the next modules to recognize and translate gestures easily.

Once the hand is successfully detected, the next important step is tracking the movement of the hand when the user performs gestures. This module continuously monitors and follows the hand region even if the user moves their hand across different positions or angles. The tracking process ensures that the system does not lose focus on the hand and can accurately recognize both static and dynamic gestures. This module plays a major role in filtering out unnecessary background objects and noise, focusing only on the required hand region. The main advantage of this module is that it eliminates the need for wearable devices like gloves or sensors, allowing the system to detect gestures naturally with only a webcam, making the overall system cost-effective, easy to use, and highly efficient in real-time environments.

Feature Extraction and Gesture Recognition

The Feature Extraction and Gesture Recognition Module plays a vital role in identifying and classifying the detected hand gestures into their respective sign language alphabets, numbers, or words. After the hand is detected and tracked by the first module, this module receives the cropped hand region for further analysis. The main focus here is to extract important hand features like shape, finger positions, and the overall pattern of the gesture. This process is handled by Convolutional Neural Networks (CNN), a powerful deep learning model that automatically learns and extracts significant features from the input hand images without the need for manual effort. CNN processes the input images through multiple layers, focusing on the important gesture patterns required for accurate recognition.

It extracts the important features like hand shape, finger positions, and orientation from the detected hand using CNN. These features help the system to identify what gesture the user is performing. CNN is trained with a large sign language dataset, so it easily compares the extracted features with stored patterns. This helps in recognizing both static and dynamic gestures shown by the user. The extracted features improve the accuracy of the system even when gestures vary in size, angle, or distance. After the gesture is classified into its respective alphabet, number, or word, the result is sent to the next module for further processing and output display.

Translation

The Translation Module plays a vital role in converting the recognized gesture into readable text. After detecting and classifying the gesture using CNN, this module receives the output and maps it to the corresponding alphabet, number, or word from the sign language dataset. It ensures that the identified gesture is translated clearly and accurately, helping normal users easily understand the message shown by speech and hearing-impaired people.

This module provides real-time conversion without any delay, ensuring smooth communication. It stores gesture data for fast access and helps in reducing errors during translation. The output generated is instantly passed to the User Interface Module for display. It also makes the communication process more user-friendly and efficient.

In addition, this module helps in reducing the communication gap between normal users and sign language users. It ensures accuracy and clarity in gesture-to-text conversion. In future, this module can be enhanced to form complete sentences, support multiple languages, and provide voice output to improve communication even better.

User Interface

The User Interface (UI) Module is responsible for displaying the recognized gesture output in a clear and readable text format. After detecting, recognizing, and translating the gesture in the previous modules, this module presents the output directly on the screen for easy communication. It provides a simple and user-friendly platform where users can view the converted text instantly, making the system more interactive and convenient for everyone.

This module ensures that the output is displayed in real-time without any delay. It updates the screen immediately for every new gesture detected, maintaining smooth and continuous communication. The design of this module mainly focuses on clarity, simplicity, and providing easy access to the output for both

technical and non-technical users.

In future, this module can be enhanced by adding features like voice output, text-to-speech conversion, and customizable display options. This will improve user experience and make the system more flexible for communication between speech and hearing-impaired people and normal users.

VI. CONCLUSION

Communication is an essential part of human life, and it becomes even more important when it comes to bridging the gap between speech and hearing-impaired people and normal users. In this modern world, technology plays a vital role in solving real-time problems, and one such innovative solution is the *Real-Time Sign Language Recognition System*. This project focuses on developing a smart, user-friendly platform for converting hand gestures into readable English text, enabling easy communication between sign language users and those who do not understand sign language.

The proposed system is implemented using advanced technologies like Python programming, OpenCV for image processing, YOLOv5 for accurate hand detection, and Convolutional Neural Network (CNN) for feature extraction and gesture recognition. This system captures real-time video input using a webcam, detects and tracks the hand region from every video frame, extracts necessary features like hand shape, finger positions, and movement, and classifies the gestures into corresponding English alphabets, numbers, or words.

The recognized gesture is then converted into readable text and displayed on the screen for easy understanding. One of the main advantages of this system is its ability to perform real-time gesture detection without using any wearable devices like gloves or sensors. This makes the system more cost-effective, convenient, and comfortable for users. The system works efficiently in different lighting conditions, dynamic backgrounds, and for different users with varying hand positions. It is highly accurate, fast, and suitable for recognizing both static and dynamic gestures.

The application of this project can be extended to various fields like educational institutions, hospitals, government offices, customer care centers, and public places, where communication barriers need to be eliminated for specially-abled people. The system

helps them to interact independently without depending on any human interpreter, promoting inclusivity and equal participation in society.

Furthermore, the future scope of this project is very wide and promising. The system can be enhanced by integrating voice output for the recognized English text, making the communication even more interactive and user-friendly. Sentence formation techniques can also be incorporated using NLP (Natural Language Processing) so that multiple gestures can be converted into a meaningful complete sentence rather than just words or alphabets. This will help in improving the communication flow and reduce the effort of showing individual signs for each word. Another future enhancement is the development of a mobile application that allows users to use this system anytime, anywhere, increasing its portability and usability.

Additional features like changing text size, font style, colour customization, voice assistance, and storing recognized text for future reference can also be added to improve user experience. With the advancement of artificial intelligence and deep learning techniques, this project can be further upgraded to provide multi-mode communication support in real-world scenarios.

Overall, this *Real-Time Sign Language Recognition System* is an efficient, reliable, and innovative solution that bridges the communication gap between speech and hearing-impaired individuals and normal users. It promotes inclusivity, equal participation, independence, and a better standard of living for specially-abled people in the society.

Thus, this project stands as a valuable contribution towards improving communication between speech and hearing-impaired people and normal users. It ensures real-time gesture recognition with accuracy, speed, and user-friendly interaction. With further advancements, this system can bring a positive change in the lives of specially-abled individuals, making communication easier, faster, and more accessible in daily life.

VII. OUTPUT

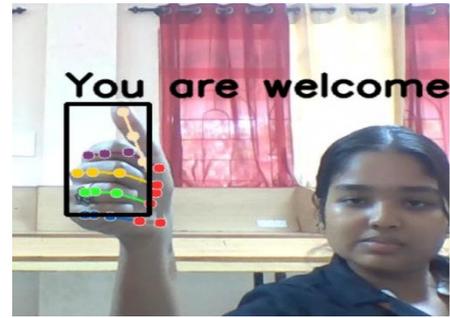


Figure 2 - sample output 1

Predicted character : You are welcome

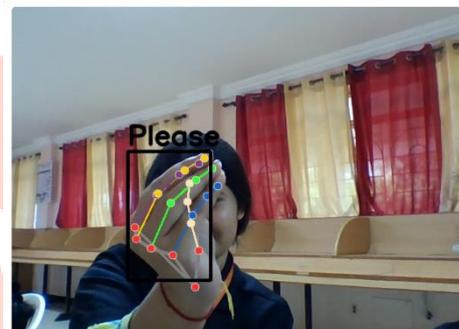


Figure 3 - sample output 2

Predicted character : Please

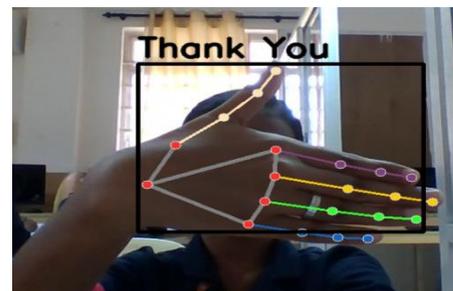


Figure 4 - sample output 3

Predicted character : Thank You

REFERENCES

- [1] C. Keskin, F. Kirac, Y. E. Kara, and L. Akarun, "Real time hand tracking and gesture recognition using the sum of anisotropic Gaussians model," *IEEE International Conference on Computer Vision*, pp. 1–8, 2011.
- [2] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 3, pp. 311–324, 2007.
- [3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.
- [4] L. Pigou, S. Dieleman, P. Kindermans, and B. Schrauwen, "Sign language recognition using convolutional neural networks," *European Conference on Computer Vision (ECCV) Workshops*, pp. 572–578, 2014.
- [5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *International Conference on Learning Representations (ICLR)*, 2015.
- [6] A. Bhuyan, S. Dutta, and M. Nasipuri, "Continuous sign language recognition using CNN and BiLSTM," *International Journal of Computer Applications*, vol. 179, no. 44, pp. 25–30, 2018.
- [7] S. Molchanov et al., "Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4207–4215, 2016.
- [8] T. Shanableh and M. Al-Rushdan, "Real-time hand gesture recognition for sign language using a CNN-LSTM model," *IEEE Access*, vol. 9, pp. 138165–138175, 2021.
- [9] S. Chattopadhyay, P. Mitra, and S. Dutta, "Real-time Indian sign language recognition using CNN and transfer learning," *Procedia Computer Science*, vol. 167, pp. 2141–2148, 2020.
- [10] A. Farhadi, R. Collobert, and P. Dollár, "YOLOv5: Advancements in real-time object detection," *arXiv preprint arXiv:2104.02767*, 2021.
- [11] S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1–54, 2015.
- [12] P. Molchanov et al., "Hand gesture recognition with 3D convolutional neural networks," *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1–7, 2015.
- [13] S. S. Rautaray and A. Agrawal, "Real-time hand gesture recognition using skin color detection and feature extraction," *International Journal of Computer Applications*, vol. 80, no. 16, pp. 1–7, 2013.
- [14] M. Van den Bergh and L. Van Gool, "Combining RGB and ToF cameras for real-time 3D hand gesture interaction," *IEEE Workshop on Applications of Computer Vision (WACV)*, pp. 66–72, 2011.
- [15] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 25, pp. 1097–1105, 2012.
- [16] G. Liang and S. Zhang, "Multimodal gesture recognition using deep learning for human-computer interaction," *Neurocomputing*, vol. 356, pp. 188–197, 2019.
- [17] K. P. Khandait, R. C. Thool, and P. D. Khandait, "Automatic sign language recognition using gesture segmentation," *International Journal of Engineering Research and Applications*, vol. 2, no. 3, pp. 105–110, 2012.