DAYANANDA SAGAR COLLEGE OF ENGINEERING

(An Autonomous Institute affiliated to VTU, Belagavi, Approved by AICTE & ISO 9001:2008 Certified)

Accredited by National Assessment & Accreditation Council (NAAC) with 'A' grade,

Shavige Malleshwara Hills, Kumaraswamy Layout, Bengaluru-111

"Dynamic Gesture Recognition"

PROJECT REFERENCE NO: 46S_BE_0078

Submitted by

PRAJNA HARISH (1DS19CS732)
[prajnah03@gmail.com ; 87626447698]
MOWNIKA REDDY K A (1DS19CS728)
[mownikareddy2001@gmail.com ; 9606232453]
SAMBHRAMA K (1DS19CS743)
[sambhramapatel@gmail.com ; 7349242409]
VINAYAK BHAT(1DS19CS757)
[vinayaknbhat10@gmail.com ; 9901125242]

OF

Eight Semester B.E (CSE)

Under the guidance of

DR. DEEPAK G
Associate Professor
Dept. of CSE
DSCE, Bangalore
[deepak-cs@dayanandasagar.edu ; 9886166985]
&
DR. HARISH KUMAR N
Assistant Professor
Dept. of CSE
DSCE, Bangalore

KEYWORDS

Gesture recognition; Deep learning techniques; Human-computer interaction; Keyboard, mouse, touch screen; Sign language recognition; Hand gestures; Medium for recognizing and interpreting motions; Communication gap; Complexity and diversity of gestures; Machine learning techniques; Real-time performance; Customization and adaptability; Evaluation and benchmarking; Data collection; Input pre-processing; Feature learning; YOLO implementation; Feature fusion and classification; Results and conclusion; High accuracy; Real-time performance; Sign language recognition

INTRODUCTION

This project's major goal is to use advanced deep learning techniques to recognize human gestures for human-computer interaction. There are only a few ways to interact with computers: using a keyboard, mouse, or touch screen, for example. When it comes to adapting more adaptable computer technology, each of these devices has its own set of constraints. One of the most important methods for creating user-friendly interfaces is gesture recognition. Gestures can come from any part of the body, although they most usually come from the face or the hand. Users may interact with devices without touching them thanks to gesture detection. Most deaf people communicate with non-impaired people in their daily lives via sign language created by bodily activities such as hand gestures, body mobility, eyes, and facial expressions. It has, however, become a barrier for deaf people who want to integrate into society. As a result, having a medium that can recognize and interpret motions into words that ordinary people can comprehend is critical, as the information conveyed by hand gestures is always primary in sign language. We suggest a sign language to speech functionality for the blind in addition to the deaf. A hand gesture recognition system for Sign Language Recognition (SLR) is required to bridge the communication gap.

In the process of recognizing hand gestures, the complexity and diversity of gestures will extremely impact on the recognition rate and reliability. The existence of machine learning techniques can be effectively exploited in the task of improving the rate of hand gesture recognition.

OBJECTIVES

1. **Create a powerful gesture recognition system:** The main goal is to build a system that can correctly identify and decipher a variety of dynamic motions made by users. Different hand forms, movement rates, and differences in illumination should all be supported by the system.

2. **Real-time performance:** The system must be built to recognise gestures in real-time, enabling smooth user-system communication. For effective and low-latency processing, this goal entails optimising the algorithms and methods employed.

3. **Customization and adaptability:** The system should be flexible enough to accommodate various users and situations. It should be able to accurately recognise objects in a variety of situations while also learning and adapting to the preferences of each user.

4. **Robustness to noise and variability**: The objective is to develop a system that can handle noise and variability commonly encountered in real-world scenarios. This includes dealing with occlusions, background distractions, and variations in hand movements or appearance.

5. **Evaluation and benchmarking:** The project should establish a comprehensive evaluation methodology to assess the performance of the gesture recognition system. This involves collecting a diverse dataset, defining evaluation metrics, and comparing the system's performance with existing approaches or benchmarks.

6. **Integration and application**: The ultimate objective is to integrate the gesture recognition system into practical applications or use cases. This may involve integrating it into virtual reality (VR) or augmented reality (AR) environments, gaming consoles, smart home devices, or other interactive systems.

METHODOLOGY

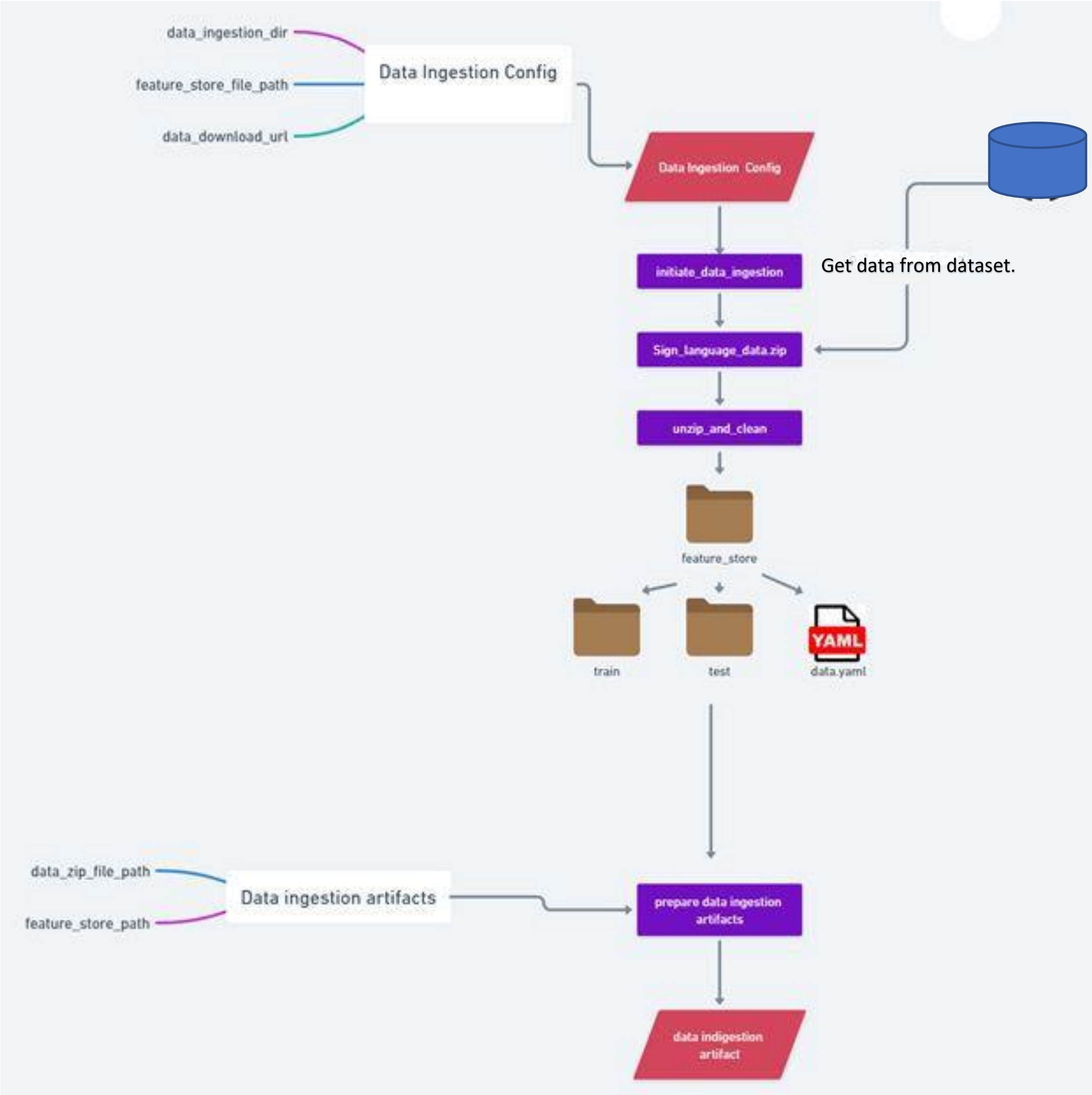


FIG 1: Final Sample output (voice is dispatched in the background automatically

➢ Workflow of Gesture Recognition

- Extracting Key point Values using Media pipe
- Collect Key point Values for Training and Testing
- Pre-process Data and Create Labels and Features using labelimg
- Build and Train YOLO V5 Neural Network.
- Make Predictions & Save Weights
- Detecting Gestures

➢ Data Collection

Over the course of two recording sessions, we recorded the dataset using LogiTech C270 HD Webcam for a few commonly used signs. There were no restrictions on the setting for the recording, the participants' clothes, or the lighting arrangement. The dataset may reveal significant variations in illumination, participant dress, position, scale, and gesturing speed.

➢ Input Pre-processing

Input pre-processing refers to the steps taken to prepare raw data before feeding it into a machine learning model. It involves various techniques to transform and manipulate the data to make it suitable for the model. Some common pre-processing steps include cleaning the data by handling missing values and outliers, scaling the features to a similar range, encoding categorical features, transforming features, reducing dimensionality, and addressing imbalanced data. These steps ensure that the data is in a format that the model can understand and learn from effectively, leading to better model performance and accurate predictions.

➢ Feature Learning

Feature learning enables models to learn directly from raw data, eliminating the need for manual feature engineering and potentially uncovering hidden patterns that may not be apparent to human experts.

Overall, feature learning plays a crucial role in machine learning and has revolutionized various domains, including computer vision, natural language processing, and speech recognition, by enabling models to automatically learn powerful and relevant features directly from the data.

➢ YOLO Implementation

Image Pre-processing: Before feeding an image into YOLO, it's usually necessary to preprocess the image to bring it into a suitable format. This typically involves resizing the image to match the input size expected by the YOLO model (e.g., 320x416 pixels) and normalizing the pixel values.

Input Encoding: YOLO takes input in the form of tensors or arrays. Therefore, the preprocessed image needs to be converted into a tensor/array that the model can process. This step usually involves converting the image to a multi-dimensional array and reordering the dimensions to match the input format required by the YOLO model (e.g., [batch_size, height, width, channels]).

Anchor Box Assignment: YOLO uses anchor boxes of different sizes and aspect ratios to detect objects of various shapes. During input processing, anchor boxes are usually assigned to the ground truth objects based on their similarity in terms of intersection over union (IoU). This assignment helps the model learn to predict the corresponding anchor box during training.

Label Encoding: Object detection involves assigning class labels to the detected objects. Each object in the training data needs to be encoded with its corresponding class label. This step often involves mapping the object labels to numerical values or one-hot encoding, depending on the requirements of the YOLO model.

Data Augmentation: Data augmentation techniques, such as random cropping, rotation, flipping, or color jittering, can be applied during input processing to increase the diversity of training data and improve the model's generalization ability.

➢ Feature Fusion and Classification

Feature fusion involves combining or integrating features from multiple sources or modalities to create a unified representation that captures complementary information. This process aims to leverage the strengths of each modality and enhance the overall performance of the model. Feature fusion can be performed at different levels, such as early fusion (combining raw features), mid-level fusion (combining intermediate representations), or late fusion (combining predictions or decision scores).

Classification is the task of assigning labels or categories to input data based on its features or representations. After feature fusion, the unified representation is used as input to a classification model or algorithm. The goal is to learn a decision boundary or mapping that can accurately classify new, unseen instances.

RESULTS AND CONCLUSION

The model has undergone successful training for a range of specific signs. We anticipate expanding its capabilities to encompass additional signs shortly. It is noteworthy to mention that the achieved results are contingent upon the signer involved, as individual characteristics can impact the system's performance. However, it is gratifying to report that our current implementation has demonstrated commendable accuracy levels, accompanied by remarkably low latency.

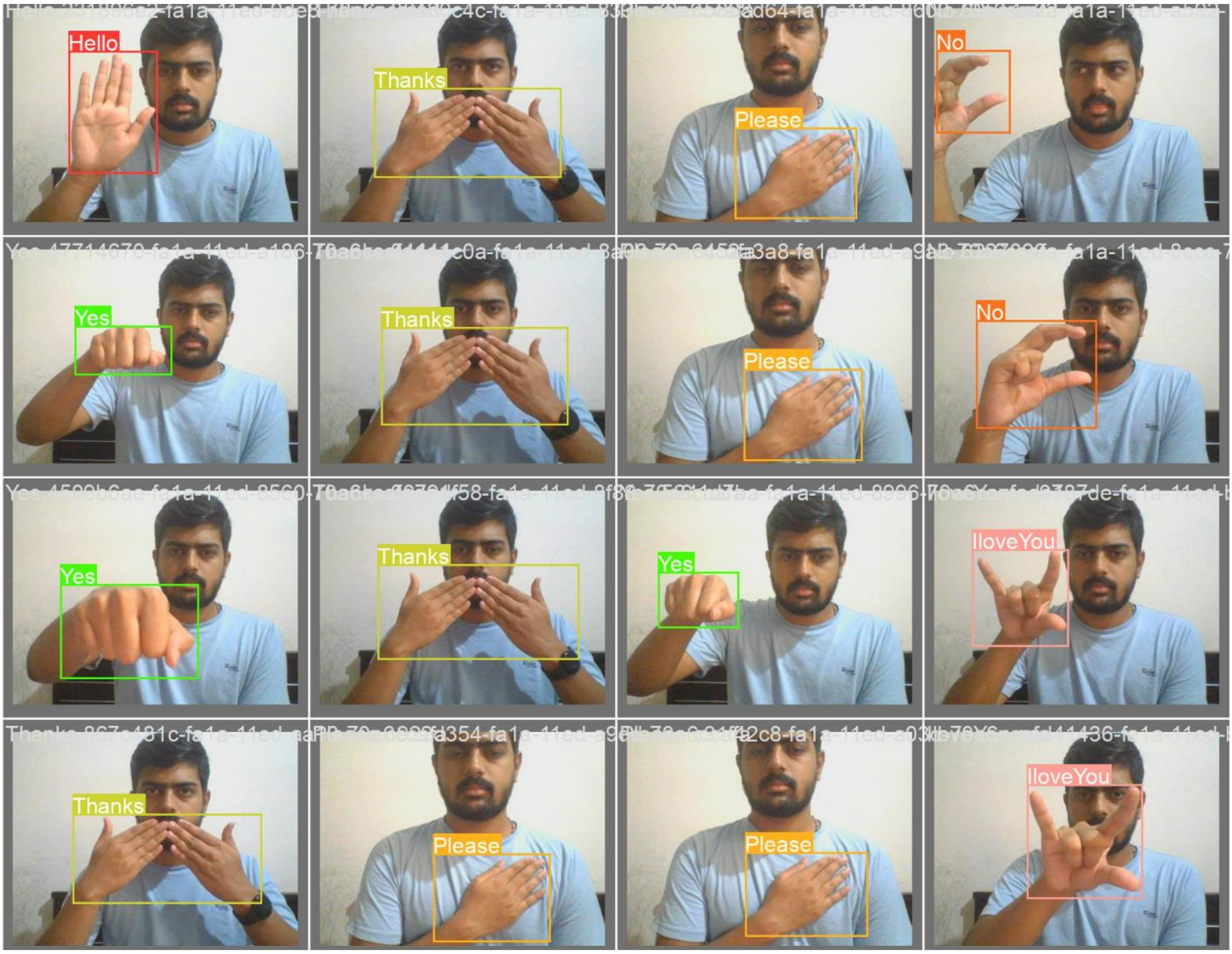
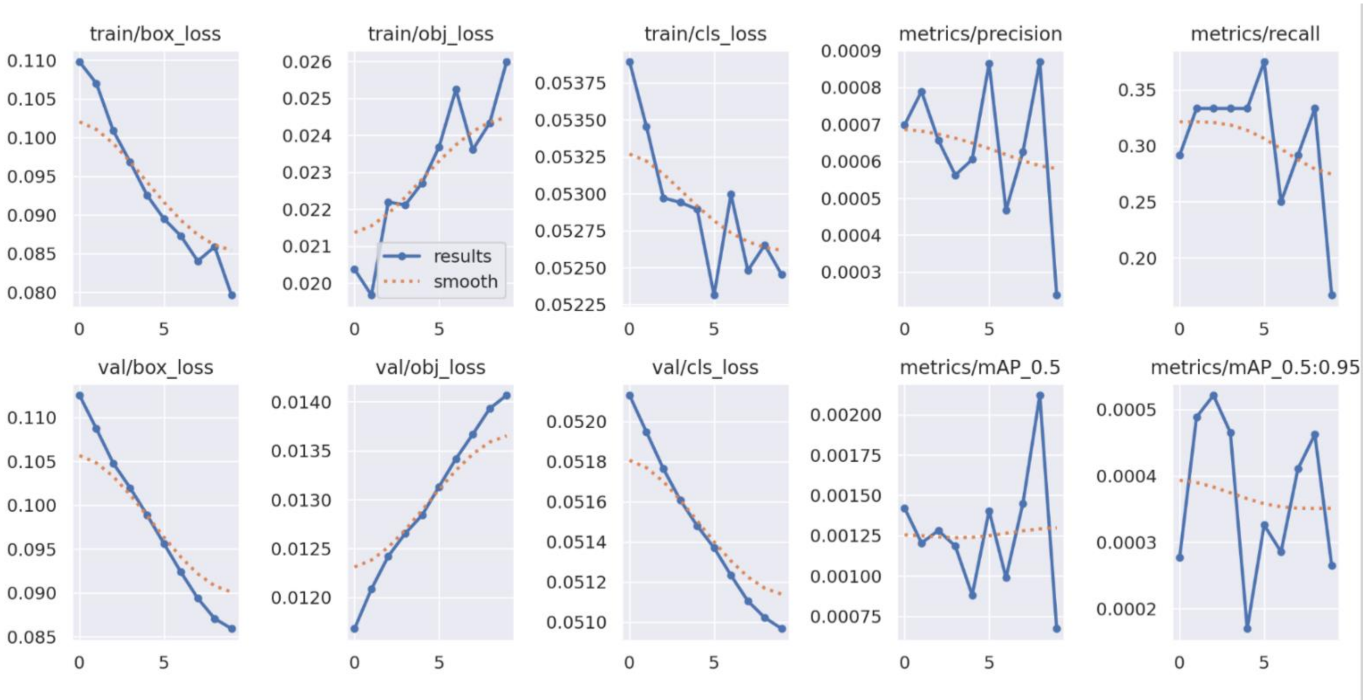


FIG 2: Ground Truth of Trained data



GRAPH 1: YOLOv5 Performance Graph

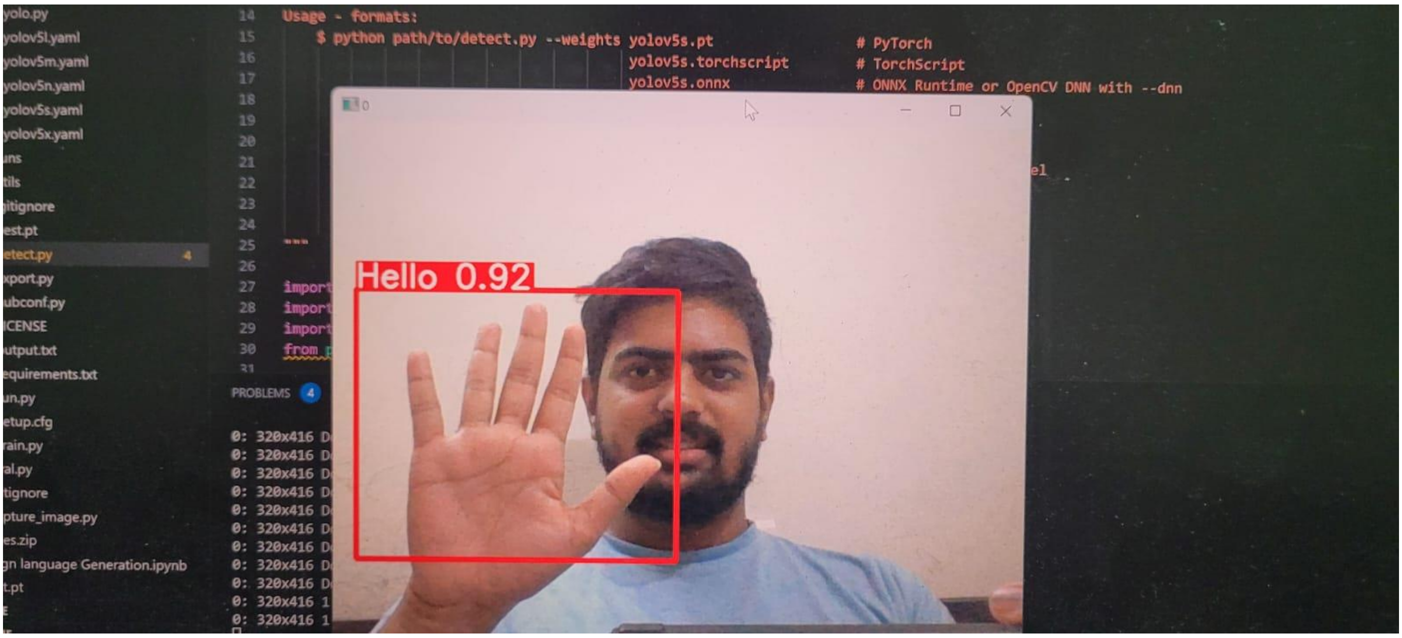


FIG 3: Final Sample output (voice is dispatched in the background automatically)

1. High Accuracy: Dynamic gesture recognition systems have achieved high accuracy rates in recognizing and classifying the trained dynamic gestures. With the advancements in deep learning techniques by using YOLOv5, R-CNN, the accuracy of our dynamic gesture recognition model has significantly improved.

2. Real-Time Performance: Dynamic gesture recognition systems have demonstrated the capability to process and recognize gestures in real time, allowing for immediate response and interaction. Our model exhibits minimal latency.

3. Sign Language Recognition: Dynamic gesture recognition has shown promise in recognizing and interpreting sign language gestures. By accurately recognizing and translating sign language gestures into text or spoken language, it can facilitate communication between hearing-impaired individuals and non-sign language users.

In conclusion, gesture recognition technology has made significant advancements in recent years, enabling various applications in fields such as human-computer interaction, virtual reality, robotics, and gaming. Gesture recognition systems can interpret and understand human gestures, allowing users to interact with devices and interfaces in a more intuitive and natural manner.

INNOVATION CARRIED OUT IN THE PROJECT

Our project stands out with two unique features: **gesture customization** and **automated sign language to voice conversion.**

The project focuses on dynamic hand gestures and sign language translation. Users have the freedom to customize their own gestures, associating them with specific commands or phrases. This customization feature adds a personalized touch and adaptability to individual preferences.

The system employs advanced algorithms and models to accurately recognize and interpret hand gestures in real-time. It automatically converts sign language into spoken words, bridging the communication gap for non-sign language users as well as the visually impaired people. The conversion process is seamless, providing instant and accurate translation without the need for manual intervention.

Our system can be integrated with various devices, including smartphones, wearables, and smart home assistants as a future enhancement. This integration allows users to interact with their devices using hand gestures, creating a natural and immersive experience. Ongoing refinement of the gesture recognition models ensures high accuracy and robustness.

As already mentioned, accessibility enhancements include text-to-speech capabilities for users with visual impairments. Our project seeks to empower individuals and promote inclusivity by enabling effective communication between sign language users and the wider community.

SCOPE FOR FUTURE WORK

1. **Real-time translation**: Expand the capabilities of your system to provide real-time translation of sign language into different spoken languages. This would allow users to communicate with people who speak different languages without the need for an interpreter.

2. **Gesture recognition accuracy:** Continuously work on improving the accuracy and robustness of your gesture recognition system. This can involve fine-tuning your training models, incorporating advanced algorithms, or exploring new sensor technologies to capture hand gestures more effectively.

3. **User-friendly interface:** Develop a user-friendly interface for the system, making it accessible to a wide range of users. We can consider designing an intuitive and visually appealing application that is easy to navigate and provides clear instructions for users to follow.

4. **Integration with other devices**: Explore the integration of the system with other devices or platforms, such as smartphones, smart home assistants, or wearable devices. It can also be integrated to Zoom calls, Google meets. This integration can enable users to interact with their devices using hand gestures, providing a more natural and immersive user experience.

5. **Gesture-based learning:** Implement a gesture-based learning component the system. This feature can enable users to learn sign language through interactive tutorials or lessons, providing them with a means to acquire sign language skills at their own pace.