IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 26, Issue 6, Ser. 1 (Nov. – Dec. 2024), PP 45-52
www.iosrjournals.org

Realtime Multilingual Sign Language Translation App

Mr. Kamlesh Tripathi
Department Of Computer Science And Engineering
Chhatrapati Shivaji Maharaj University Navi Mumbai

Dr. Vikas Kumar
Department Of Computer Science And Engineering
Chhatrapati Shivaji Maharaj University Navi Mumbai

Kalpesh Patil
Department Of Computer Science And Engineering
Chhatrapati Shivaji Maharaj University Navi Mumbai

Abstract
In a globalized world, effective communication is a necessity. As we live in rapidly changing world, communication enables individuals to adapt to new circumstances, innovate, and seize opportunities. It an important aspect which focuses on building relationships facilitates the exchange of ideas, thoughts, feedback, and feelings with others. Thus, hearing people breakdown and smash those barriers encountered by the Deaf and Hard of Hearing communities. The main barrier is communication. Sign language is just like any other language used for communication purpose. Addressing this challenge, the development of a Multilingual Sign Language App is proposed. This app aims to bridge the communication gap by offering a comprehensive platform for learning and using sign languages from around the world and is designed to be a versatile and inclusive tool that caters to the unique linguistic and cultural aspects of different sign languages. The app includes features such as sign language dictionaries with video demonstrations, interactive learning modules, translation tools, customizable learning experiences, and a vibrant user community. Users can access content in various sign languages. The app promotes not only language acquisition but also cultural understanding. By addressing linguistic diversity and cultural nuances, this app stands as a vital resource for both Deaf and hearing communities, fostering a more inclusive and understanding world.

Date of Submission: 29-11-2024 Date of Acceptance: 09-12-2024

I. Introduction

In a world that thrives on communication, there are various challenges faced by Deaf and hard of hearing individual in today's era. Sign languages are the primary means of communication for an estimated 466 million deaf or hard-of-hearing people worldwide. However, the difference between sign language and spoken language causes some communication barriers between them and hearing-unimpaired people, which brings inconvenience to their daily lives. This motivates researchers to design more efficient and accurate sign language translation systems. Additionally, they face problems in expressing themselves properly and responding to the people around them. This constant struggle may cause social and mental issues. The Sign Language Translator App emerges as a beacon of inclusivity and understanding. As communication is a fundamental aspect of human interaction, transcending borders, cultures, and languages. However, for the Deaf and Hard of Hearing communities, traditional barriers to communication have persisted, limiting their ability to fully participate in a world that often takes spoken language for granted.

The Sign Language Translator App steps in as a ground-breaking solution to bridge this gap, The creation of a multilingual sign language app can be a valuable tool for promoting communication and inclusivity among people who are deaf or hard of hearing. Such an app can help users learn and communicate in different sign languages from around the world. Also ensures fostering inclusivity and fostering connections in an increasingly interconnected global society by seamlessly translating sign language gestures into multiple spoken languages, enabling real-time communication that knows no bounds.

As an individual point of view, it has a feature of video demonstrations provide high definition of signs to ensure clarity and accuracy respectively of cultural and regional variations, which allow us to get knowledge and understand of the sentiments and thought which are been proposed by sign language thus this help to improve content over time also on other hand it translates regional, cultural language gesture into sign language which decrease the gap of communication.

The Multilingual Sign Language App is more than just a language-learning tool. It's a catalyst for change that will have a profound impact on society. It will Promote equal access to education and employment opportunities for the deaf and hard of hearing, raise awareness of Deaf culture and languages, fostering a sense of belonging, Encourage the hearing community to learn sign languages, breaking down communication barriers, facilitate effective communication in healthcare settings, ensuring that individuals with hearing impairments receive quality care.

The app includes a vast collection of signs from multiple sign languages. Each sign is accompanied by ensure

II. Releated Works

When creating a Realtime Multilingual Sign Language Translation App, it's quite essential to decrease the communication barrier through empowering inclusivity through technology. Here's a concise overview of potential related works:

(A) Bridging Communication Barriers: The process of overcoming obstacles or differences in communication through technologies in order to facilitate effective communication between individuals. It involves implementing various approaches to facilitate clear and meaningful interaction, understanding, and collaboration among diverse individuals or communities. The goal is to ensure that message are accurately conveyed and understood.

(B) Real-time Sign Language Detection via Mobile Camera: Incorporates computer vision technology to uses a camera to capture images of a person signing and it analyzes each frame to detect and classify sign language gestures as they are performed then processes the image frames to detect and recognize hand gestures in real-time through a mobile camera. Thus, models learn to associate hand movements and configurations with corresponding signs in the sign language

(C) A Gesture translated into thousand of words: Sign language provides a means of communication for individuals who are deaf or hard of hearing, ensuring their inclusion in society and facilitating communication with both deaf and hearing individuals. These hand gestures are capable of conveying complex meanings which can be expressed through thousand of words.

(D) Real-time communication and Real-time understanding: This refers to any form of communication that occurs instantaneously where participants can communicate without any noticeable delay and play critical roles in enhancing efficiency, productivity, and user experience. Real-time communication and understanding is a key focus.

(E) Translation in Multilingual Language: This app promotes huge collection of signs form multiple sign language which ensure clarity and accessibility across different linguistic communities and facilitates communication as well as comprehension in diverse cultural also linguistic contexts, promoting inclusivity and understanding across language barriers.

III. Methodology And Architecture

Overview

Sign language is a crucial mode of communication for millions of Deaf and Hard of Hearing individuals around the world. However, the barrier between the Deaf and the hearing communities often remains insurmountable due to the lack of understanding and accessibility to sign language. To address this challenge, we propose the development of a cutting-edge real-time multilingual Sign Language Translator application. This innovative application aims to bridge the communication gap by providing a comprehensive set of features, including sign language gesture learning lessons, text-to-sign language translation, real-time sign language detection via mobile camera, and video calls with translation capability.

Sign language is one of the most important tools for speech impaired people to communicate with others. As sign language has an extensive vocabulary and complex expressions, it requires a lot of time and effort to learn, but it is not easy for anyone to master. It is more challenging to automatically perform Sign Language Recognition (SLR) because of an extensive vocabulary and complex expressions from a machine perspective. We propose using a deep learning model as a translator that can convert sign language to text. The entire model has been converted into an application that involves the user's action being captured by his/her camera and the data will be sent to the rest API in base64 format.

The American Sign Language (ASL) is one of the most popular languages used by people who are deaf to communicate with each other. With its natural syntax, which shares the same etymological roots as spoken languages but has a different grammar, ASL may express the outcome of bodily acts. It becomes difficult for people not familiar with ASL to communicate with people who are hearing impaired. ASL [17] is a full, natural language with English-like syntax that shares many of the same linguistic characteristics as spoken languages. Hand and face gestures are used to convey meaning in ASL.

The conventional methods [5], [6] utilizing I3D network [7] have become state-of-the-art methods for WSLR. Although I3D was mainly proposed for video-based action recognition [7], it has found success in other video recognition tasks, including SLR, due to its spatiotemporal representation capability. It is because I3D algorithm uses appearance information of the upper body of the signers to Each stream is trained separately, and the recognition scores extracted from each recognize sign language words and treats various appearance information equally. It extracts global features by observing the whole image instead of capturing only the image's local regions.

Using neural networks, we can isolate regions of interest to better distinguish between characters and gestures. Most pre-trained networks by MATLAB perform vision tasks which require a convolutional layer, which are highly effective in image classification [3]. Using a convolutional layer, we can extract features frame by frame of our image/video data set. Our feature selection will require a temporal convolutional network to encompass spatial and local features. Our CNN network will also comprise of a max pooling layer, the reason is to down-sample our inputs, this will reduce the dimensionality of our input features allowing assumptions within sub-regions [37, 4]. In mathematics the idea is to take the filtered input from the Convolutional layer, which normally decodes everything in matrices. These matrices need to be reduced in scale for our network and thus the pooling layer operates on these matrices, the operations performed are normally Max or Average, thus dubbed max-pooling or just pooling. Our application's learning sign language lesson feature offers a structured and interactive educational platform, providing users with comprehensive sign language lessons that include video demonstrations and textual explanations. Users can monitor their progress, engage in practice exercises, and connect with a supportive community of learners. This feature is de- signed to empower users to effectively acquire sign language skills, ultimately fostering inclusively and facilitating improved communication between the Deaf and Hard of Hearing com- munities and the hearing population.

Our proposed Video Call Sign Language Translator feature is a ground-breaking addition to our application, enabling real-time communication between Deaf and Hard of Hearing individuals and the hearing population through video calls. Using cutting-edge computer vision technology, this feature accurately detects and interprets sign language gestures as they occur during video calls, ensuring a smooth and natural conversation flow. The system offers two translation options: text and voice output, catering to diverse user preferences. Users can engage in rich, real-time conversations with confidence, as the application ensures that sign language nuances are conveyed accurately. This trans formative feature aims to bridge the communication gap, enabling sign language users to participate fully in video conversations and expanding their opportunities for interaction and engagement, both locally and globally. Text-to- Speech (TTS) refers to the ability of computers to read text aloud. A TTS Engine converts written text to a phonemic representation, and then converts the phonemic representation to waveforms that can be output as sound. Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech synthesizer, and can be implemented in software or hardware.

Design Approach

For the making of project, we have used AGILE Model. Agile refers to something that is quick or adaptable and prioritize delivering value quickly and are willing to adjust plans and priorities as needed. A software development approach based on iterative development is referred to as an "agile process model." Agile approaches divide projects into smaller iterations or sections and avoid long-term planning. The scope and requirements of the project are defined at the start of the development phase. As the scope was well defined before the building of the project, we decided to go with Agile model. Various different phases of Agile Model are:

- Requirement Gathering
- Design
- Develop
- Test
- Deploy
- Review

Requirement Gathering

In this phase, the requirements must be defined. Business opportunities and plan the time and effort needed to build the project must be explained. Based on this information, technical and economic feasibility can be evaluated.

Design

When the project has been identified the, work with stakeholders to define requirements. Need to use the user flow diagram or the high-level UML diagram to show the work of new features and show how it will apply to your existing system.

Develop

When the team defines the requirements, the work begins. Designers and developers start work- ing on their project, which aims to deploy a working product. The product will undergo various stages of improvement, so it includes simple, minimal functionality.

Test

In this phase, the Quality Assurance team examines the product's performance and looks for the bug.

Deploy

In this phase, the team issues a product for the user'swork environment.

Review

After releasing the product, the last step is feedback. In this, the team receives feedback about the product and works through the feedback.

Working

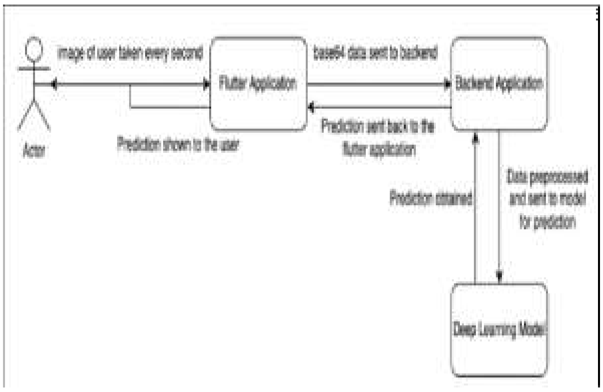


image of user taken every second
Flutter Application
base64 data sent to backend
Backend Application
Actor
Prediction shown to the user
Prediction sent back to the flutter application
Prediction obtained
Data preprocessed and sent to model for prediction
Deep Learning Model

Figure.3.1 Level 1 Use case Diagram

In the fig.3.1 a use case diagram for a real-time speech- to-sign language translation app.

The diagram shows the followinginteractions:

- A user takes a video Stream of Sign Language Person.
- The Flutter application sends the base64 data to the backend application.
- The backendapplication preprocessesthe data and sends it to the deep learning model for prediction.
- The deep learning model predicts the sign language corresponding to the user'sspeech
- The backend application sends the prediction back to the Flutter application.
- The Flutter application shows the prediction to the user.

This use case diagram illustrates how the app can be used to bridge the communication gap between deaf/mute people and the hearing population. It also highlights the importance of real- time translation, as this allows deaf/mute people to communicate effectivelyin any situation.

Workflow

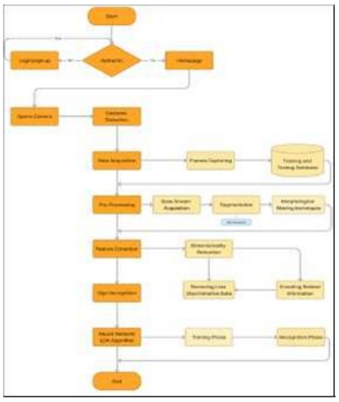


Figure.3.3 Flowchart

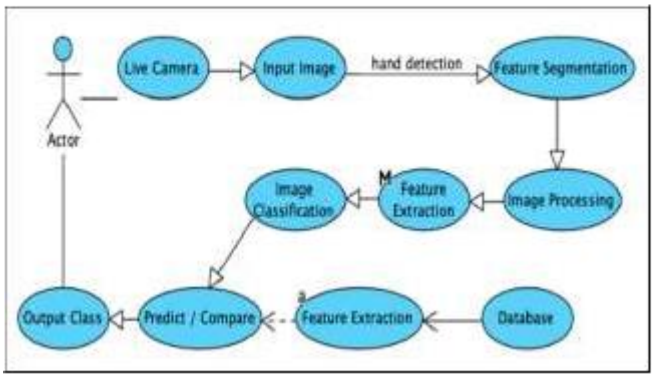


Figure.3.2 Level 2 Use case Diagram

The use case diagram also shows the following benefits of the app:
•It is accessible to deaf/mute people, as they can simply take a picture of themselves to communicate.
•It is easy to use, as the Flutter application takes care of all of the underlying processing.
•It is accurate, as the deep learning model has been trained on a large dataset of speech and sign language data.

Overall, the use case diagram demonstrates how the real- time speech-to-sign language translation app can be used to improve the quality of life for deaf/mute people and to make the world a more inclusive place.

Video Processing and Feature Extraction:

OpenCV is used for capturing video frames from a webcam (cv2.VideoCapture). Frames are resized and normalized to be feed into the Inception I3D model. Frames are stored in a buffer for further processing. The Inception I3D model is used for extracting features from video frames.

Sign Language Recognition:

The Inception I3D model is loaded and modified for the specific task of sign language recognition (InceptionI3d). The model is fine-tuned on the ASL2000 dataset for sign language recognition. The run_on_tensor function processes video frames through the I3D model and performs inference. The results are post-processed to obtain the predicted sign language word.

Natural Language Processing:

The keytotext library is used for converting sign language words into text. A pipeline for the KeytoText model is initialized (pipeline("k2t- new")).The nlp function converts recognized sign language words into sentences using natural language processing techniques. NGram models are used for suggesting the next word in the sentence.

Data Loading and Pre-processing:

The project involves loading preprocessed data, including NGram models (nlp_data_processed and nlp_gram counts).

Model Loading and Deployment:

The project loads pre-trained models for sign language recognition and translation. The Inception I3D model is loaded with its weights and deployed for real-time processing.

Visualization:

The processed video frames, along with recognized sentences, are displayed in real-time using OpenCV.
Additional Components:
Environment variables are loaded from a .env file. Matplotlib is used for plotting.

Configuration and User Interaction:

The project allows configuring various parameters through command-line arguments (argparse).The user can specify the mode (RGB or flow), save model paths, and root directories.

Algorithm
Inception I3D Model (Sign Language Recognition):

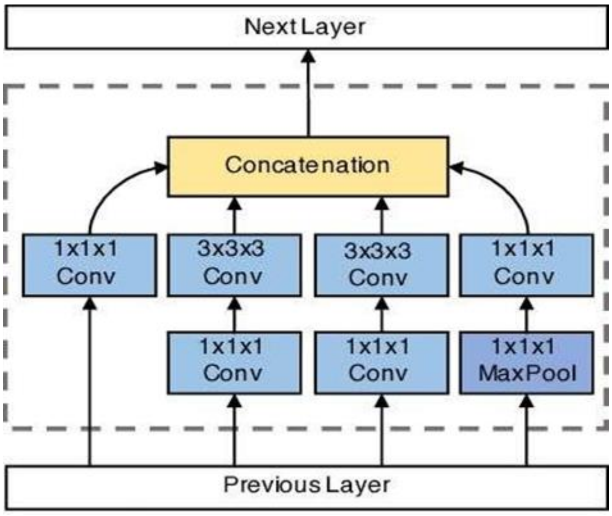
The Inception I3D model is a 3D convolutional neural network (CNN) designed for action recognition in videos. It is often pre-trained on large-scale video datasets (e.g., Kinetics) and fine-tuned on specific tasks or datasets. Fine-tuning involves training the model on the target dataset, adjusting the weights to specialize in recognizing signs from the ASL2000 dataset in this case. Common optimization algorithms for training CNNs include Stochastic Gradient Descent (SGD) with momentum or Adam.

KeytoText Model (Natural Language Processing):

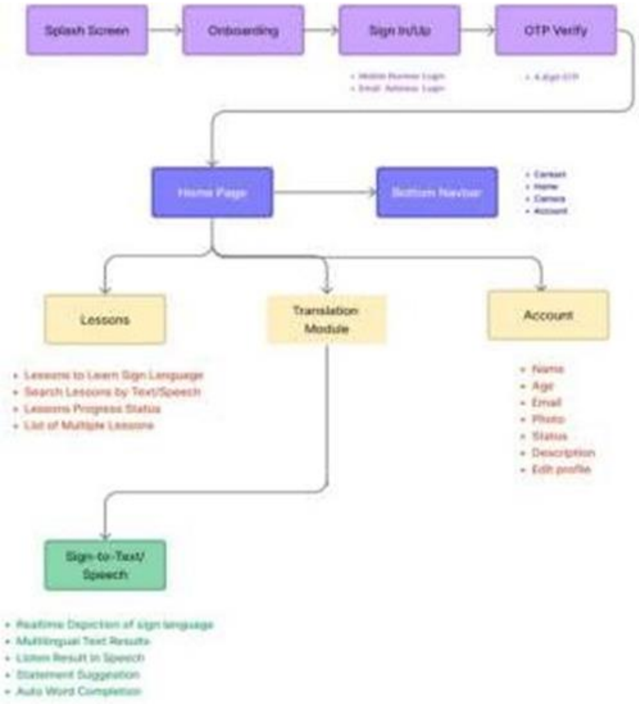
The KeytoText model is initialized with pre-trained weightsIt might involve training on a large corpus of text data to learn language patterns and relationships.

NGram Model (Natural Language Processing):

NGram models are statistical language models that predict the probability of the next word based on the previous n-1 words. Training an NGram model typically involves counting the occurrences of n-grams in a training corpus to estimate probabilities. Smoothing techniques might be applied to handle unseen n-grams.



Architecture/Framework:



User Login

The login feature in your application plays a pivotal role in offering users a personalized and interactive experience centered around real-time sign language gesture translation and the learning of Indian Sign Language.

The application caters to two main user groups. Firstly, general users seeking real-time sign language interpretation and education. These users can explore the core functionality of the application, which involves interpreting sign language gestures in real-time and providing corresponding meanings in both Indian regional languages and English. Additionally, the application offers a comprehensive learning experience for Indian Sign Language through modules tailored to various proficiency levels. Secondly, users preferring a simplified login process can sign in with their Google ID, ensuring a seamless and efficient on boarding experience.

Upon successful login, general users gain access to the application's primary features. This includes real-time sign language interpretation, where the application employs a sophisticated gesture recognition system to decipher gestures and deliver accurate meanings in multiple languages. Furthermore, users can engage in structured learning modules designed to teach Indian Sign Language. Progress tracking, achievements, and interactive lessons enhance the learning journey. The application allows customization of settings to tailor the experience to individual preferences, covering aspects like language choices, gesture recognition sensitivity, and notification. For users opting for Google ID login, the authentication process is seamlessly integrated, eliminating the need for manual registration. Once logged in, users can enjoy the benefits of account synchronization, where information from their Google account enhances the overall experience. This may include preferences, history, or other relevant data that can be synchronized with the application, contributing to a more personalized and efficient interaction.

Real-time Translation

The real-time translation feature utilizing the device's camera represents the core functional- ity of your application, providing users with an immersive and instantaneous experience in interpreting sign language gestures. Upon activating the camera for real-time translation, the application leverages advanced computer vision and machine learning algorithms to detect and interpret sign language gestures. The camera captures live video input, and the system processes each frame to identify and analyze specific hand movements, facial expressions, and body language associated with sign language.

The application incorporates a robust gesture recognition system that has been trained on a diverse dataset of sign language gestures. This system utilizes deep learning techniques to recognize patterns and correlations between different gestures and their corresponding mean- ings. The model is continuously refined through machine learning updates, ensuring accuracy and adaptability to a wide range of sign language variations. Once a sign language gesture is successfully identified, the application translates it into meaningful words or phrases. The translation process is designed to provide outputs in both Indian regional languages and English, catering to a diverse user base. The translated text is then displayed on the user interface in real- time, offering immediate comprehension of the communicate message. To ensure accessibility in various scenarios, the real-time translation feature may include an offline mode. The application could store essential gesture recognition models locally, allowing users to continue using the translation functionality even in areas with limited or no internet connectivity. Continuous updates and improvements to the gesture recognition and translation models are crucial for staying at the forefront of accuracy and reliability. Regularly incorporating user feedback, expanding the dataset for training, and refining the algorithms based on real-world usage contribute to the application's ongoing enhancement.

Lessons

The learning modules within your application provide users with a structured and comprehensive approach to mastering Indian Sign Language. This feature is designed to cater to both beginners and those seeking to enhance their proficiency in sign language. The learning mod- ules are organized into a well-defined structure, covering a range of sign language gestures from basic to advanced levels. Each module is thoughtfully curated to introduce users to the fundamental elements of Indian Sign Language progressively. The progression ensures that users build a solid foundation before moving on to more complex gestures and expressions. To enhance the learning experience, the application includes a robust progress tracking system. Users can monitor their advancement through the modules, track completed lessons, and receive feedback on their performance. Gamification elements, such as achievements and rewards, may be incorporated to motivate users and make the learning process enjoyable. Recognizing the diversity of users, the learning modules offer support for multiple languages, including Indian regional languages and English. Users can choose their preferred language, allowing them to grasp the nuances of sign language in a language they are most comfortable with.

IV. Experimental Results

In this section, we unveil the empirical outcomes of our research efforts, providing a comprehensive analysis of the proposed system conducted in accordance with the methodologies outlined earlier. The presented results offer crucial insights into the setup and the execution of our approach.

System Information

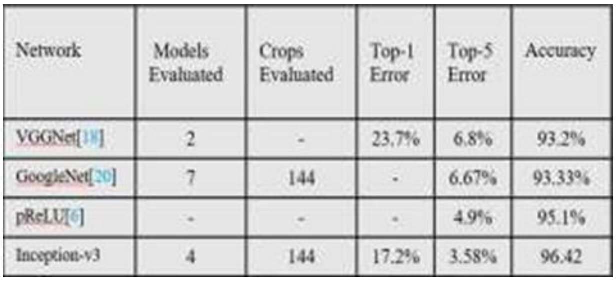
1. Hardware Details

- GPU : GTX 1650 Super or Mobile Graphics
- Processor : AMD Ryzen 3300h or Any Mobile processor upto 1.5GHz
- RAM : 8GB
- OS : Browser or Android 10+

2. Software Details

- Pytorch
- Python
- Cv2
- HTML, CSS, Javascript
- Flutter
- MongoDB and Firebase (B)Accuracy

Accuracy



Network	Models Evaluated	Crops Evaluated	Top-1 Error	Top-5 Error	Accuracy
VGGNet[18]	2	-	23.7%	6.8%	93.2%
GoogleNet[20]	7	144	-	6.67%	93.33%
pReLU[6]	-	-	-	4.9%	95.1%
Inception-v3	4	144	17.2%	3.58%	96.42

- VGGNet : Visual Geometry Group Network, is a convolutional neural network architecture designed for image recognition and classification tasks.
- GoogleNet : It also known as Inception-v1, is a convolutional neural network architecture, designed for image classification tasks.
- pReLU : Parametric Rectified Linear Unit is an activation function commonly used in neural networks. PReLU introduces learnable parameters to the rectified linear unit (ReLU) activation function, allowing it to adaptively learn the slope of the negative part of the activation.
- Inception-v3 : The Inception V3 is a deep learning model based on CNN, which is used for image classification. It is used on the coarsest (8 × 8) grids to promote high dimensional representations. It is designed for image classification, object detection, and other computer vision tasks.

V. Conclusion And Future Work

In conclusion, our real-time multilingual Sign Language Translator application stands as a trans formative solution for bridging the gap between the Deaf and Hard of Hearing communities and the hearing world. With its extensive set of features, including sign language learning lessons, real- time sign language detection, text-to-sign language translation, voice output, and video call interpretation, our application promises to empower individuals with the tools they need for effective and inclusive communication. By breaking down communication barriers and offering accessible learning opportunities, we aim to promote greater understanding and connectivity between these communities, fostering a more inclusiveand equitable society.

Looking ahead, in Continuously improving the accuracy and recognition capabilities of sign language gestures is essential. Incorporating machine learning and AI techniques can lead to more precise and adaptable sign language interpretation. Adding more interactive lessons which will expanding the range of lessons to cover advanced sign language topics, specific dialects, or regional variations, catering to a broader user base.

References

[1] Shubham Thakar, Samveg Shah, Bhavya Shah, And Anant V. Nimkar, ”Sign Language To Text Conversion In Real Time Using Transfer Learning:” In IEEE Access, Vol. 10, Pp. 69436-69451, 2022.

[2] M. Madhisaran And Partha Pratim Roy, ”A Comprehensive Review Of Sign Language Recognition: Different Types, Modalities, And Datasets” In Arxiv Preprint Arxiv:2204.03328, April 2022.