SilentSync-AI

Pranav Khaire[1], Sanket Mahajan[2], Pushkar Pawar[3], and Sahil Pawar[4]

[1]Prof. Chaitali Shewale
[2]Dr. Suruchi Dedgaonkar

Vishwakarma Institute of Information Technology (VIIT)
Department of Information Technology
Savitribai Phule Pune University

December 22, 2023

Abstract

Sign language, a natural language developed for communication with the deaf and mute, plays a crucial role. It serves as an essential component in non-interpreter-mediated interpersonal interactions between individuals with normal hearing and those who are deaf or mute.

This project's main objective is to develop a system that allows people with speech and hearing impairments to interact with everyday people. Various sign languages exist worldwide. However, the primary focus of this system is Indian Sign Language (ISL), which is headed toward standardization. The system will exclusively focus on hand movements, a vital component of the body for conveying ideas, messages, and thoughts between individuals who are deaf or mute.

The suggested approach will interpret Indian Sign Language gestures into Marathi audio, identifying words and sentences. The procedure is broken down into three stages: data collection, training, and testing. Initially, it will recognize Indian Sign Language gestures. Ultimately, the algorithm interprets the motion and turns it into audio in Marathi.

Keywords

Sign language interpretation, ISL communication, Marathi audio conversion, gesture recognition technology, inclusive communication tech.

1 Introduction

For those who have trouble hearing or speaking, sign language is their main form of communication. Deaf or mute individuals encounter daily challenges where body language, hand gestures, and facial expressions play crucial roles in sign language communication. Not all individuals can communicate through sign language, and various sign languages, including Indian Sign Language, exist. Language signs depend on regional dialects.

Finding interpreters in day-to-day life is increasingly difficult, and deaf and mute individuals require interpreters to communicate with the general population [1]. A system designed to facilitate communication would greatly benefit those who are deaf or mute. Our goal is to create a system that simplifies communication. Sign language involves making gestures with hands in coordination with head or other body parts, along with specific facial expressions.

In this project, we initially created a dataset and then utilized technologies such as Mediapipe, CNN, and other Python modules to train and test our dataset, converting hand gestures to English speech. Finally, we use Google APIs to convert English speech into Marathi.

2 Literature Survey

The field of artificial intelligence (AI) is one that is developing quickly and has shown greater productivity in a variety of applications. The focus of this literature survey is on AI's development in recognizing hand gestures, particularly in the context of a chatbot designed to encourage human-machine dialogue.

2.1 Indian Sign Language in Communication

This section addresses the definition, applications, and varieties of sign language, with a specific focus on Indian Sign Language (ISL). It explores how sign language functions in communication, both in-person and in front of the camera. The section delves into database organization and functionality within the context of sign language, providing insights into various sign languages.

2.2 Hand Gesture for Communication

This paper examines the interactive and communicative aspects of hand gestures, distinguishing between static and dynamic gestures. Tracking methods, including image tracking, coordinate tracking, and label tracking, are discussed to achieve precise gesture recognition. The section also explores applications of hand gestures in communication.

2.3 Artificial Intelligence in Sign Language

Focusing on the advancements in AI applications for sign language, this paper highlights its potential to enhance communication between individuals with hearing impairments and the general population. It covers the broad spectrum of sign language technology, including translation and representation, emphasizing the crucial role of AI in facilitating verbal communication for the deaf or hard of hearing.

2.4 Implementation of Hand Gesture Recognition

In this paper, the author proposed a system that translates hand gestures into appropriate text messages to facilitate communication between deaf and mute people using Indian Sign Language (ISL) with ordinary individuals. The primary objective is to create an algorithm that can instantly translate dynamic gestures into text. This addresses the need for people who are deaf or hard of hearing to communicate verbally with various groups.

2.5 Machine Learning and Deep Learning in Hand Gesture

This paper presents a CNN-based method for computer vision-based hand gesture recognition and classification. The model is trained using the Hand Gesture dataset, where each gesture is mapped to a label. Hand gestures are recognized in this way.

3 Proposed Methodology

3.1 Dataset Generation

Creating a comprehensive database of sign language gestures is essential for effectively comparing photos captured during communication with the system. The following steps outline the process we undertook to construct our dataset.

Our dataset was meticulously crafted utilizing the Open Computer Vision Library (OpenCV). To facilitate training, we initially captured rough photographs of each hand gesture in Indian Sign Language (ISL). Subsequently, we recorded every frame points generated by our computer's webcam and MediaPipe, as shown in Fig. 2.

Figure 1: 21 Hand Points Captured

The dataset is specifically tailored to include demonstrations of the gestures intended for identification by the system. Each gesture is associated with corresponding labels for enhanced training accuracy.

The system collects key points on the hand to accurately interpret sign language gestures. These key points include 21 points of each hand as shown in Fig. **??** and stored in Data.csv Fig. 3 for training purpose.

0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

Figure 2: 21 Hand Points Captured

3.3 Recognition

To extract relevant features from the hand gesture photos, a CNN architecture is created using TensorFlow. This list of features could include edges, forms, textures, and other visual characteristics that are crucial for distinguishing one gesture from another. The CNN model is then trained using the preprocessed data set, and it modifies its internal parameters to identify patterns and characteristics in the movements.

A list of landmarks is sent to the model.predict() function, which produces an array with prediction classes for each landmark.

1	0	0	0.1875	0.017857	0.383929	-0.05357	0.535714	-0.10714	0.660714	-0.14286	0.428571	-0.24107	0.651786	-0.26786	0.830357	-0.24107	0.955357	-0.20536	0.419643
1	0	0	0.171171	0.045045	0.342342	0.009009	0.477477	0.009009	0.594595	0.009009	0.414414	-0.20721	0.648649	-0.15315	0.81982	-0.07207	0.927928	0.018018	0.441441
1	0	0	0.155963	0.009174	0.33945	-0.01835	0.477064	-0.02752	0.605505	-0.02752	0.422018	-0.20183	0.642202	-0.13761	0.798165	-0.05505	0.899083	0.027523	0.449541
1	0	0	0.134454	0	0.294118	-0.05882	0.411765	-0.06723	0.521008	-0.07563	0.411765	-0.2437	0.655462	-0.21008	0.789916	-0.10924	0.857143	-0.0084	0.436975
1	0	0	0.134921	0.015873	0.293651	-0.03968	0.428571	-0.05556	0.539683	-0.06349	0.365079	-0.27778	0.626984	-0.24603	0.801587	-0.15873	0.912698	-0.07143	0.388889
1	0	0	0.151515	-0.06818	0.325758	-0.14394	0.469697	-0.16667	0.598485	-0.17424	0.348485	-0.31818	0.606061	-0.34848	0.780303	-0.30303	0.893939	-0.22727	0.356061
1	0	0	0.18705	-0.05036	0.359712	-0.11511	0.503597	-0.13669	0.633094	-0.14388	0.381295	-0.32374	0.654676	-0.34532	0.827338	-0.29496	0.942446	-0.22302	0.381295
2	0	0	-0.34409	-0.17204	-0.5914	-0.33333	-0.77419	-0.46237	-0.91398	-0.58065	-0.36559	-0.73118	-0.55914	-0.87097	-0.72043	-0.7957	-0.80645	-0.69892	-0.25806
2	0	0	-0.36782	0.034483	-0.6092	0.091954	-0.81609	0.16092	-1	0.149425	-0.58621	-0.49425	-0.86207	-0.51724	-0.77011	-0.34483	-0.62069	-0.31034	-0.49425
2	0	0	-0.36	0.093333	-0.66667	0.106667	-0.86667	0.08	-1	0.04	-0.62667	-0.4	-0.86667	-0.54667	-0.8	-0.37333	-0.65333	-0.26667	-0.53333
2	0	0	-0.34667	0	-0.58667	-0.13333	-0.69333	-0.30667	-0.77333	-0.46667	-0.54667	-0.52	-0.8	-0.58667	-0.70667	-0.38667	-0.54667	-0.28	-0.46667
2	0	0	-0.36508	0.063492	-0.63492	-0.04762	-0.74603	-0.2381	-0.77778	-0.42857	-0.63492	-0.30159	-0.87302	-0.44444	-0.7619	-0.33333	-0.5873	-0.26984	-0.57143
2	0	0	-0.38095	0.079365	-0.68254	0	-0.85714	-0.15873	-1	-0.33333	-0.68254	-0.33333	-0.93651	-0.44444	-0.8254	-0.25397	-0.63492	-0.14286	-0.60317
2	0	0	-0.2963	0.074074	-0.62963	0.018519	-0.85185	-0.11111	-1	-0.22222	-0.7037	-0.37037	-0.96296	-0.44444	-0.90741	-0.25926	-0.75926	-0.12963	-0.61111

Figure 3: data.csv

These points are gathered with the aid of the MediaPipe library and are essential for accurately identifying and deciphering hand gestures used in sign language communication.

3.2 Labelling and Tracking

Hand gesture recognition is a multi-step process that uses a range of tools and libraries, such as OpenCV, MediaPipe, Scikit (scikitlearn), TensorFlow, Matplotlib, Pandas, and Convolutional Neural Networks (CNNs).

First, a collection of images or videos depicting hand gestures points is assembled into a dataset, and each action points is assigned the proper label. Next, the data is preprocessed by resizing, standardizing, and augmenting hand gestures in order to enhance the model's learning capability. MediaPipe and OpenCV combine to preprocess, track, and extract patterns and features from hand movements in real time.

After that, the images is fed into a model for recognising sign language, which is then used by CNN to interpret and convert the sign language gestures into text or speech.

[2.0656423e-28 1.9525415e-24 9.9780710e-01 9.7559216e-85 1.66172453e-66 1.0834080e-88 1.1111732e-56 4.564236e-46 6.6467859e-15 4.5745162e-01] is the output look like.

The index of the maximum value from the database is returned by Np.argmax(). Once the index has been obtained, all that needs to be done is select the English and Marathi translations from the database, and the text will quickly appear on the screen.We translate from English to Marathi using Google Api.Hand motions can be converted from text to Marathi voice using the Google Text to voice Text API.

3.4 Displaying

Each translation is shown on the website that we design. We employed Django based web application using HTML, and CSS, three contemporary web technologies. Additionally, a user-friendly user interface design that makes it simple for users to input hand gestures. Our recognition system can be easily connected to the website to deliver real-time results. Our approach ensures that hand gestures on the website we created will be successfully recognized.

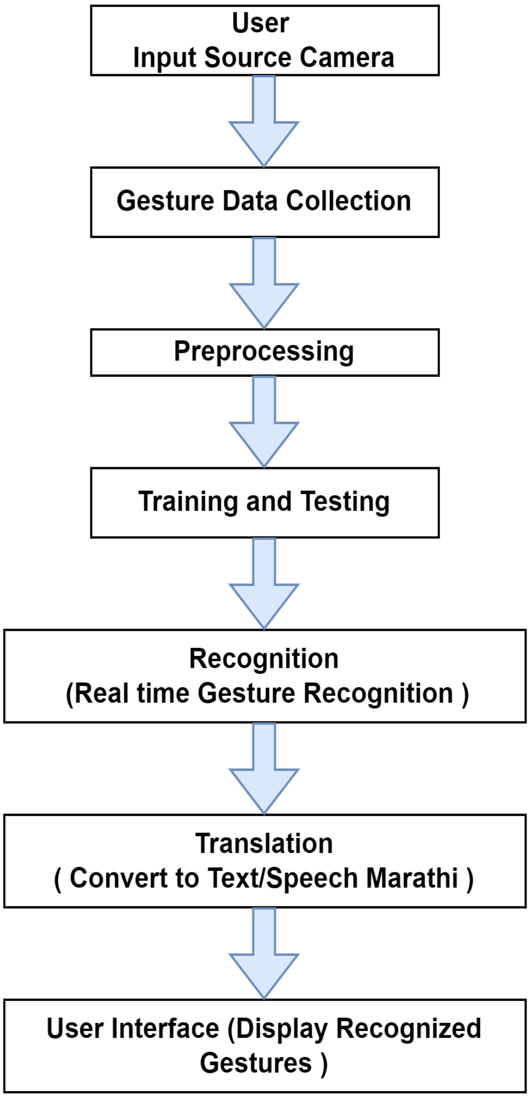

Figure 4: Data Flow Diagram

4 Results

Using computer vision, Silent Sync-AI recognizes hand gestures in real-time, translates them into regional languages like Marathi and English, and then speaks the gestures in Marathi. For this project, we used HTML, CSS, JavaScript, and Django to create a web application that allows dumb people to vocalize their actions.

Because the website is designed to be user-friendly, it will be easier to record real-time video and images and translate them into Marathi and English. The pre-trained TensorFlow model for hand gestures that accurately recognizes the various actions is the foundation of this project.

In the evaluation of our trained model, we achieved a commendable accuracy of 83 percent in recognizing and interpreting hand gestures for sign language communication.

We integrated voice synthesis using the Google Text-to-Speech (gTTS) and pygame libraries to provide a comprehensive user experience. This improves accessibility by giving the communication process an auditory component.

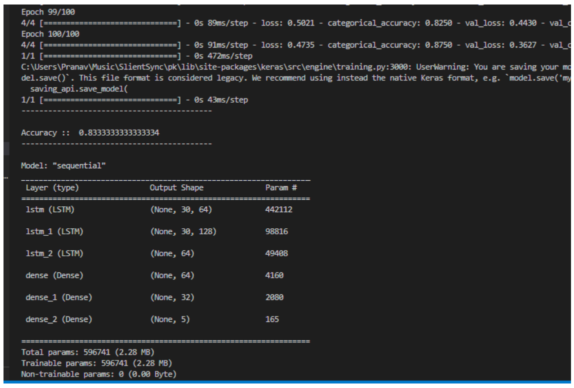

Figure 5: Accuracy

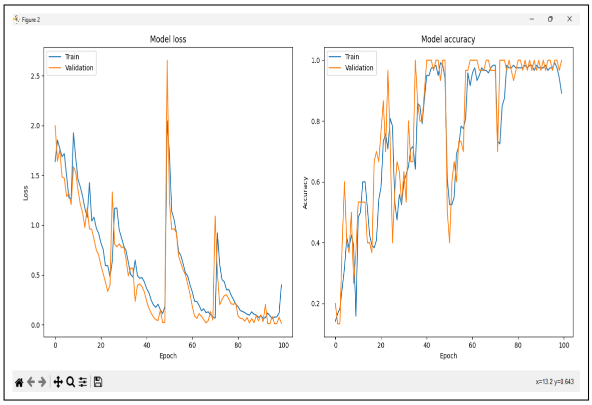

Figure 6: Accuracy Graph

In Fig.6 graph between loss-epoch and Accuracy-Epoch represents Model Loss and Model accuracy graph respectively. It is used to predict loss and accuracy.

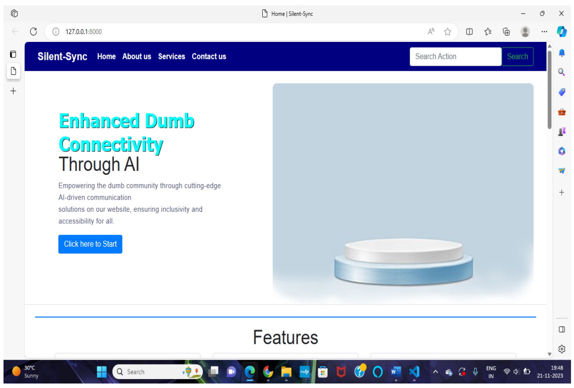

Figure 7: Home Page

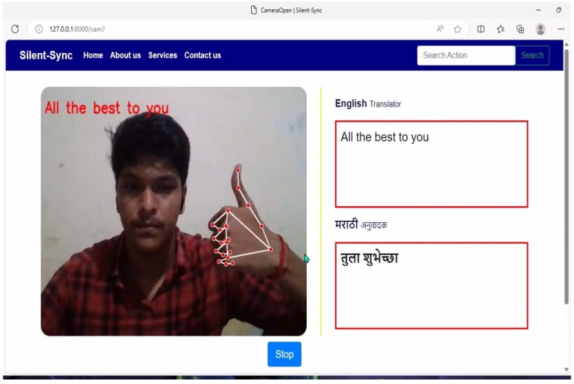

Figure 8: All the best

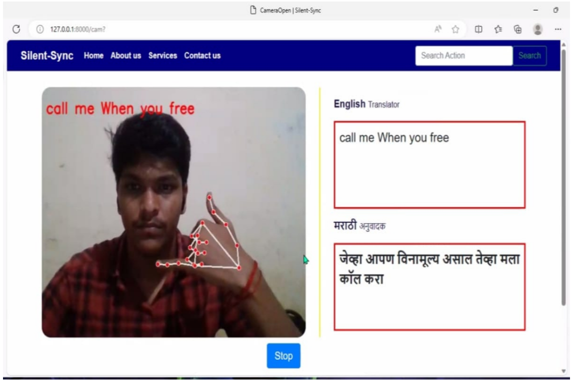

Figure 10: Call me when you free

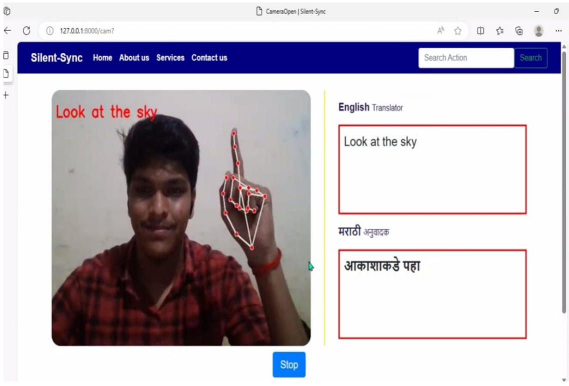

Figure 9: Look at the sky

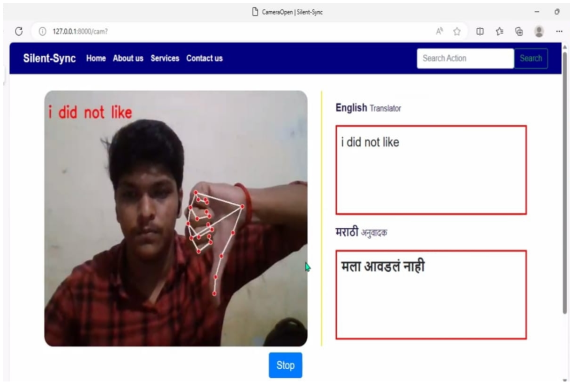

Figure 11: I did not like

5 Conclusion

The demands placed on hand gesture recognition in real-world applications are high due to its need for efficiency, accuracy, and robustness. In this system, hand landmarks are extracted using Mediapipe , after which we tested and trained our model. To translate English text to Marathi and then back again into Marathi audio, we used the Google API.

The actual implementation of the system has demonstrated the effectiveness of the design in a practical setting. Python is used in the development of this system.

In the preexisting system, only 10 actions can be recognized, and translations are available only for English and Japanese. However, in our SilentSync AI, we have expanded the repertoire to include more than 10 actions and introduced support for regional language Marathi.

6 Future Work

In future we are going to expand language support to include more regional languages, enabling a broader audience to communicate using sign language , Allow users to define and customize their own gestures , Develop a mobile application version of the system for on-the-go accessibility, making it easier for users to communicates in various situations.

7 References

[1] Ashok Sahoo, Gouri Mishra and Kiran Ravulakollu, "SIGN LANGUAGE RECOGNITION: STATE OF THE ART", 2014

[2]Rafiqul Zaman Khan, Noor Adnan Ibraheem, "HAND GESTURE RECOGNITION: A LITER-ATURE REVIEW", 201

[3]Ilias Papastratis, Christos Chatzikonstantinou, Dimitrios Konstantinidis, Kosmas Dimitropoulos,and Petros Dara, "Artificial Intelligence Technologies for Sign Language", 2021

[4]Reddygari Rani, R Rumana, R. Prema, "A Review Paper on Sign Language Recognition for Dumb", 2021

[5] R.S. Sabeenian, S. Bharathwaj, M. Mohamed Aadhil, "Sign Language Recognition Using Deep Learning and Computer Vision", 2020

[6] Statistical Analysis of Design Aspects of Various Based Deep Learning Models for sign Detection.International Journal of Computational Intelligence Systems (2023) 16:126

[7] Driver drowsiness detection using ANN image processing,IOP Conf. Series: Materials Science and Engineering .