International Journal of Innovative Research in Engineering
Volume 4, Issue 3(May-June 2023), PP: 613-616.
www.theijire.com ***ISSN No: 2582-8746***

Sign Language to Speech Translation

Salman Bagwan[1], Juned Tamboli[2], Mubasir Siddiqui[3], Shrikant Shinde[4]
[1,2,3,4]Computer Science, Savitribai Phule Pune University, Maharashtra, India

How to cite this paper:
Salman Bagwan[1], Juned Tamboli[2], Mubasir Siddiqui[3], Shrikant Shinde[4], "Sign Language to Speech Translation" IJIRE-V4I03-613-616.

https://www.doi.org/10.59256/ijire.20230403127

Abstract: *Sign language is a way of communicating using hand gestures, movements, and facial expressions, instead of spoken words. It is the medium of communication used by people who are deaf or have hearing impairments to exchange information between their own community and with normal people. In order to bridge the communication gap between people with hearing and speaking disabilities and people who do not use sign language, a lot of research work using machine learning algorithms has been done. Hence, Sign language translator came into picture. Sign Language Translators are generally used to interpret signs and gestures from deaf and hard hearing people and convert them into text.*

Key Word: *Sign language, Hand Gestures, Helping Impaired People, Machine Learning, Deep Learning.*

I.INTRODUCTION

Sign language is a visual form of communication that uses body movements and facial expressions to convey meaning between people. For communication with their social environment, Deaf people only use sign language, a non-verbal language. It is based on visual cues sent by the hands, eyes, and face. In sign language, the gestures or symbols are arranged linguistically. It is a complex fusion of nonverbal cues like body language, facial expressions, and hand gestures that convey ideas or thoughts without using words.

There are many spoken languages in the world, and they are all distinctive from one another in some ways. Similar to spoken languages, there are numerous sign languages that use various hand gestures and visual cues. Some of them include Indian Sign Language (ISL), British Sign Language (BSL), French Sign Language (LSF), Pakistani Sign Language (PSL), American Sign Language (ASL), and British Sign Language (BSL). The framework for sign language translation enables deaf, dumb, and speech-impaired people to interact with hearing people using sign language. This results in the elimination of the middleman, who typically serves as a translator. Text to Sign Conversion and Sign to Text Conversion are examples of conversion modules.

Motivation:

Sign language is used by people who have trouble speaking to interact with others. The traditional approach to gesture recognition involves tracking hand gestures with a camera-based system. Comparatively speaking, the camera-based system is less user-friendly because it would be cumbersome to transport. Additionally, using it in crowded spaces would not be practical because it would pick up multiple gestures from various people who are in its field of view. Separating such offensive gestures is time-consuming and not a practical solution. Basically, we will use the device's built-in camera to capture images, perform vision analysis, operate the operating system, and output speech using the built-in audio device.

Need of the project

- To help dumb and deaf people to communication with people.
- Help normal people to understand what dumb people are trying to say.

Outcomes

- This model takes live image as an input and detects the gesture.
- Once the gesture is detected it gives output in both text as well as audio format along with the accuracy in percentage.

II.RELATED WORK

1. Sign Language Translator

The author of the paper [1] has outlined all the methods that can be used to create a standalone Sign Language translator that is installed

on a Raspberry Pi and can convert dynamic fingerspells, predict the words that go with them, and construct sentences using the Hand-mesh model that is included in the Mediapipe framework. The generated sentence will also be rendered as audio. Additionally, we have used the face-mesh model, which is also found in MediaPipe, to recognise emotions. We have also included a technique that can recognise text embedded in images on boards, flyers, and other surfaces. Using Google's text-to-speech API, we were able to successfully translate this recognised text into the chosen regional language.

2. Sign Language Interpretation

In the paper [2], research was initiated with a number of speeches to text experiments to gauge deaf people's communication abilities and better comprehend their day-to-day issues. The project's main objective was to create a deaf person's communication aid that could be incorporated into a mobile phone. In order to communicate with deaf users, this system displayed a face that was only partially animated. They have many applications and are very helpful.

3. Sign Language Machine Translation

[3] The literature review that is part of this paper is split into two sections: the traditional SLT and the neural SLT, which has recently taken over the research scene in a similar way that neural architectures have in the Natural Language Processing (NLP) sector. Transformer layers in particular, along with encoder-decoder neural architectures, have become the industry standard for handling this task. They also provide the opportunity to develop multilingual systems, though this is uncommon for SLs. Additionally, datasets are extremely scarce, and this is especially true for SLT since it is very expensive to annotate SL videos with spoken language text translations. Additionally, this makes it difficult for neural models to learn.

4. Sign Language and Gesture Recognition System

This paper [4] discusses various algorithms and methods that can be applied to identify hand gestures and sign language used by various deaf and dumb people. A more proficient and intuitive tool for human-computer interaction is the hand gesture recognition system. Applications range from sign language interpretation to virtual prototyping to medical education. Sign Language is one of the means of communication for the deaf and dumb people. The aforementioned analysis shows that the field of hand gesture recognition has advanced significantly thanks to vision-based hand gesture recognition. C, C++, and Java are the programming languages used to implement the gesture recognition system. to streamline the process, particularly for image processing operations are needed, MATLAB with image processing toolbox is used.

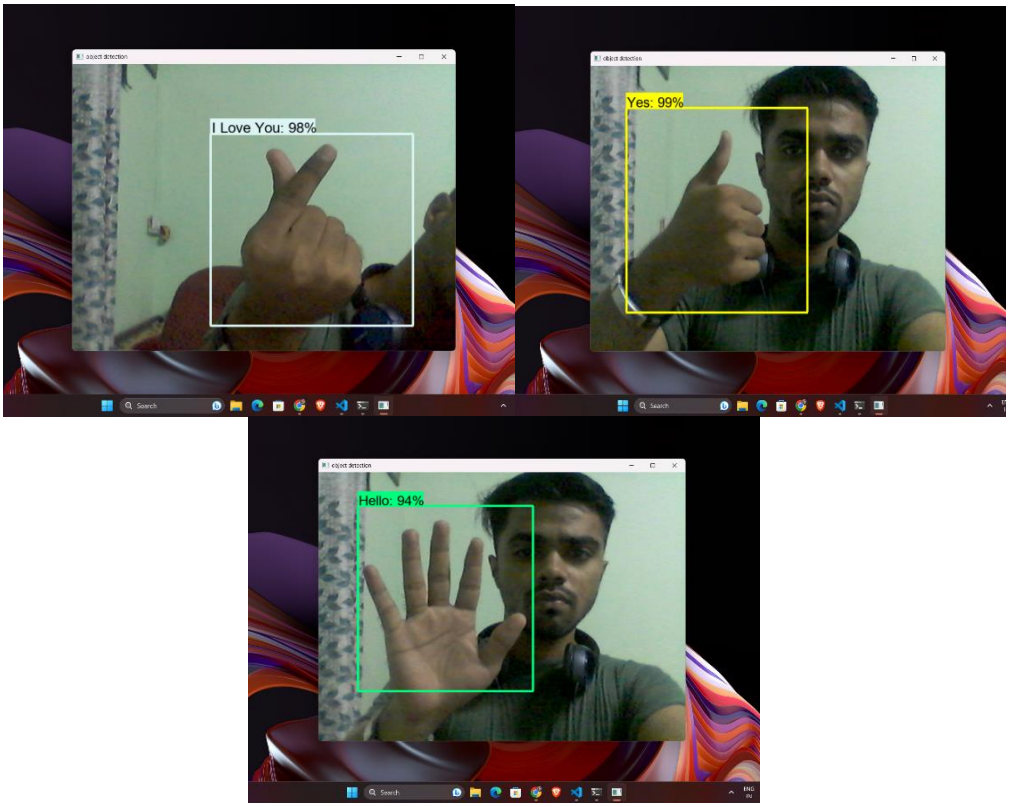
5. Sign Language Translation Systems for Hearing/Speech Impaired People

[5]The author has examined different methods and tools used for sign language translation as well as the kinds of datasets used. The majority of the systems that have already been developed are in American or British Sign Language, but the survey's primary focus was on Indian Sign Language. Systems for translation in languages like Chinese Sign Language and Russian Sign Language have also been noted. The majority of translation systems only translated one way, either from sign language to text-to-speech or from speech to sign language generation. A full two-way Indian Sign Language sign language system, however, was not seen. Existing systems are either one-way translators or domain-specific. They don't really help many people who are generally deaf or hard of hearing. They want a system that can provide them with a variety of features, such as real-time translation between sign languages. Additionally, the system should be general rather than domain-specific. This would enable them to more effectively and profoundly express their thoughts and feelings. As a result, a two-way Indian Sign Language translation system is required to completely close the communication gap between hearing- and speech-impaired individuals and other people, who could then fit into the majority of the proposed classification.

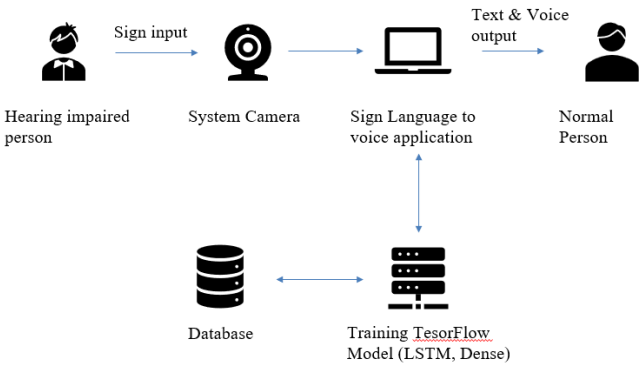
III.PROPOSED WORK

1. Setup Paths
2. Create Label Map
3. Create TF records
4. Download TF Models Pretrained Models from Tensorflow Model Zoo
5. Copy Model Config to Training Folder
6. Update Config For Transfer Learning
7. Train the model
8. Load Train Model From Checkpoint
9. Detect in Real-Time and convert text to speech

Outputs:



IV.SYSTEM ARCHITECTURE



The workflow for sign to speech model works in following manner –

1. Create a images dataset for the labels.
2. Create a train and test dataset.
3. Write code to start the training of the object-detection model.
4. Train the object-detection model using the dataset.
5. Once the object-detection model is trained, it can identify the hand gesture.
6. The output is displayed live as the camera frame is running.
7. It displays accuracy of the detected gesture as well as an audio reading the text.

V.ADVANTAGES

1. Helps communication with dumb and deaf people easy.
2. Easy and simple to use.
3. Simple to understand UI

VI.LIMITATIONS

4. Accuracy of model learning algorithms.
5. Assigning weight for extracted text.

VII.FUTURE SCOPE

The proposed sign language translation system used to translate sign language gestures can be further extended to recognize gestures facial expressions. Instead of displaying word labels it will be more appropriate to display sentences as

more appropriate translation of language. This also increases readability. The scope of different sign languages can be increased. More training data can be added to detect the words with more accuracy.

VIII. CONCLUSION

This paper addresses and overcomes the problem of communication with deaf and dumb people. Given an input image, the goal is to extract hand landmark key points and create a data set and give an output in text and speech form. We can say that the accuracy of the individual components of our design is good, however, with a huge scope for improvement.

References

[1] *Sujay R, Somashekar M, Aruna Rao B P. "Literature Survey: Sign Language Translator" In IJIRT, Volume 8, Issue 12, May 2022*

[2] *Sandra E. F. de Avila, Antonio da Luz Jr., Arnaldo de A. Araujo, and Matthieu Cord§ † Computer Science Department — Federal University of Minas Gerais. "VSUMM: An Approach for Automatic Video Summarization and Quantitative Evaluation" In IEEE Conference, 2016*

[3] *Mrs.Poonam S.Jadhava, Prof. Dipti S. Jadhavb * a Department of Information Technology, Ramrao Adik Institute Of Technology, Navi Mumbai,400706, India. "Video Summarization using Higher Order Color Moments (VSUHCM)"*

[4] *Ting Yao, Tao Mei, and Yong Rui Microsoft Research, Beijing, China. "Highlight Detection with Pairwise Deep Ranking for First-Person Video Summarization "*

[5] *Wei Zhang Faculty of Computer Guangdong University of Technology Guangzhou City, Guangdong Province, China "State Transition-Based for Cooperative Shot Boundary Detection"*

[6] *Zuzana Cernekov´a, Ioannis Pitas ˇ , Senior Member, IEEE, and Christophoros Nikou, Member, IEEE "] Information Theory-Based Shot Cut/Fade Detection and Video Summarization"*

[7] *Shrikant A. Shinde, Sakshi Bondre, Priyanka Parmar, Khushboo Patil, Pratiksha Yewalkar, "An Social Distance Violation Detector System", International Journal of Innovative research in Engineering, Vol 3 , Issue 3, PP 191-195, May-June 20222.*

[8] *Shrikant A. Shinde, Bhavesh Patil, Mrunali Ghate, Poonam Shinare, Ajay Patil, "Skin Burn and Skin Cancer Detection using Image Processing ", International ResearchJournal of Engineering and Technology, Vol 9, Issue 6, June 2022.*

[9] *Shrikant A Shinde, Shailaja N Uke, "Transmission Policies for Data Aggregation using Cooperate Node in Wireless Sensor Networks" IJSR Vol 6, Issue 1, Jan 2017*

[10] *Shrikant A Shinde, Abhilasha V. Biradar, "Genetic Algorithm for Privacy Protected Personalized Web Search" JETIR, Vol 7, Issue 4, Apr 2020*