International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Sign Language Interpretation using Machine Learning

Aniruddha Jathar (B212036) Omkar Kodmur (B213043) Prathamesh Belurkar (B212007)
Shrikant Bhandalkar (B211012) Dipali Pawar (Professor)
Assistant Professor, Department of Computer Engineering, Zeal College of Engineering and Research, Pune.
BE Students, Department of Computer Engineering, Zeal College of Engineering and Research, Pune

Abstract-- Every day we see many people with disabilities like the deaf, the dumb and the blind, etc. There are around 500 million people with loss of hearing disability. Sign language is one of the communication tools for the deaf people's community and common people's community. But normal people find it hard to understand the sign language and gestures of the deaf. Many tools can be used to translate the sign language created by the disabled into a form that normal people can understand. The studies are based on various image acquisition, preprocessing, hand gesture segmentation, extraction of features, and classification methods. This paper aims to research and examine the use of LSTM for creating a application to detect sign language using a webcam. Due to the latest advancement in classification methods, we can use TensorFlow library which has various machine learning models for creating a neural network for sign language detection. This paper specializes in the classification strategies utilized in earlier Sign Language Recognition. Based on our review, HMM-based techniques were explored significantly in previous studies, which include modifications. Deep learning consisting of convolutional neural networks has become popular over the past five years.

Keywords:

Sign language, Recognition, TensorFlow, Media Pipe, Neural networks, Deep learning, LSTM.

Introduction

Sign language is a unique form of communication that relies heavily on hand movements and facial expressions to convey meaning. It is predominantly used by individuals who are deaf or hard-of-hearing to communicate with each other, but it is not commonly used by individuals who are not hearing-impaired. This

lack of familiarity with sign language can make it difficult for hearing-impaired individuals to communicate with others, which can limit their social interactions. While real-time translation with interpreters is an option, it can be expensive and not always feasible. Therefore, there is a need for an automatic translation system that can bridge the communication gap between hearing-impaired and non-hearing-impaired individuals.

In recent years, researchers have developed novel strategies for translating sign language into written or spoken language. In this project, we aim to create a program that can recognize and translate custom sign language gestures into normal text using OpenCV. Sign language recognition algorithms must accurately identify and interpret sequences of hand and facial movements, but sign language is more than just a collection of well-articulated motions.

Gesture recognition is a branch of computer science and language technology that involves translating human touch into mathematical algorithms. While gestures can arise from any movement or posture of the body, they are most expressed through facial and hand movements. Emotional recognition through facial expression and hand gestures is an active area of research, and many methods have been developed that use cameras and computer vision algorithms to translate sign language.

Sign languages, also known as sign languages, are fully functional natural languages that convey meaning through visual clues. These languages have their own grammar and lexicon and are not universal or well understood. While there are some similarities across sign languages, they evolved independently and are unique to each culture. Sign language should not be confused with body language, which is a form of nonverbal communication.

Theory

For static hand recognition, it is easy to recognize hand posture by extracting some features such as finger directions, fingertips, skin color, and hand contours. Such features are not always available and reliable due to lighting conditions and the background of the image. There are also many other non-geometric features, such as the silhouette, color, and textures, that are inadequate in recognition. Since it is not easy to define features clearly, the entire frame or transformed image is taken as the input, and features are chosen automatically and implicitly by the recognizer. This paper aims to review and evaluate the approaches used in previous studies. It also aims to recommend the best method to investigate for future research. This paper is

considering using LSTM (Long-Short-Term-Memory) algorithm available in Keras API in python for direct implementation using TensorFlow library.

Literature Survey:

In this paper [1], they used LSTM for detection purpose using TensorFlow and Keras API. They imported LSTM Model directly from Keras and used it to detect alphabets. These alphabets are used to form words which forms a sentence to communicate.

In this paper [2], Haar Cascade Classifier is used for classifying and detecting words like “Bye", "Excuse Me” etc. This model achieved the average accuracy of 92.68%. They also use speech synthesis to convert text into direct audio format.

In this paper [3], use of TensorFlow and Keras is made to detect if the person is wearing a facemask or not. They used sequential convolutional network model to detect the presence of mask without causing overfitting.

In this paper [4], a review is made of the understanding of all the viable techniques for detection of sign language. Thus, this paper reviews various methods which can be potentially used for sign language detection. Some of the methods discussed in this paper are use of Hidden Makarov Model, use of CNN, use of Neural Networks, use of SVM.

In this paper [5], they have used Convolutional Neural Network (CNN) for hand gesture recognition. They were making use of the neural network for diving the image into smaller parts and extracting the features from the image to identify the show gesture. They achieved the accuracy of 98.74% on the testing dataset.

In this paper [6], they made a program to detect Egyptian sign language with use of the LSTM model available in Keras API from TensorFlow library. Also, they made use of CNN for extraction of spatial features and LSTM for extraction of temporal features. Both these models achieved the accuracy of 90% and 70% respectively.

In this paper [7], a thorough study about the framework "Mediapipe" is done. This paper also displays how Mediapipe builds pipelines to perform inference over arbitrary sensory data.

In this paper [8], the authors used Siamese Neural Network for hand gesture recognition of motor impaired people. Also, the hand recognition system they created is customizable . They achieved the accuracy of 80.12%.

In this paper [9], Deep Multi-layered CNN is used for hand gestures recognition in the images sent to the CNN model. This model contained 32 convolutional filters with 3 x 3 kernel, LeakyReLU activation function and 2 x 2 max pooling operations have been performed. They achieved testing accuracy of 99.89%.

In this paper [10], they created a program for sign language detection using LSTM Deep Learning Model. They used MediaPipe Holistic, OpenCV and TensorFlow for building a model which detects sign language. They achieved the accuracy of 91.1%

Methodology

Numerous approaches have been attempted to resolve the challenge of facilitating two-way communication between hearing and non-hearing individuals, but many of these strategies necessitate the use of complicated equipment or multiple cameras, rendering them inconvenient for deployment in public spaces. Moreover, these solutions can hamper accessibility due to their dependency on excessive equipment. In contrast, this project seeks to develop a more user-friendly implementation approach that can be utilized through mobile or online applications with the mere use of a camera.

To train a neural network for sign language recognition using LSTM, video sequence frames must be captured, and custom gestures must be represented within the collection of frames. Once the frames have been captured, a file must be produced that effectively describes the gesture to the network. These images and supplementary files are subsequently bifurcated into two folders for training and testing purposes. The testing phase allows for an accurate representation of how the network will recognize novel videos that it has not previously encountered.

Designing:

In this sign language recognition project, we developed a customized sign detector capable of detecting a range of unique hand signs, including letters and numbers, that can be easily expanded to include additional signs and gestures. The project was implemented using various Python modules, such as OpenCV, Mediapipe, Tensorflow, and Keras. The OpenCV module processes live video frames from a camera to detect the actions of the individual being displayed at that moment in time. Using MediaPipe Holistic, the frames are further processed to extract keypoints from the hands, torso, and face. These keypoints are then utilized by the prediction algorithm to make real-time predictions of the sign being made by the user. The predicted sign is then displayed for the user.

Prerequisites:

The prerequisites software & libraries for the sign language project are:

Python (3.10.4)

IDE (Jupyter)

Media Pipe (version 0.8.10)

NumPy (version 1.22.4)

cv2 (OpenCV) (version 4.5.5.64)

Keras (version 2.9.0)

TensorFlow (as Keras uses TensorFlow in backend and for image preprocessing) (version 2.9.1)

Dataset /Labeling:

The data collection process involves capturing the motions and signals of sign language users in a dataset. This is achieved by live streaming video from a camera and recording every frame that detects a gesture or motion within the region of interest (ROI) established. The collected data is stored in a gesture directory and each sign is represented by approximately 30 video sequences, with each sequence comprising of 30 frames of significant moments, saved as Numpy arrays. It is important to correctly identify the gesture being signed throughout the video sequence to ensure accurate data representation. Additionally, this process may require significant time and effort, but it is essential for developing a reliable sign language recognition system.

Training:

To train the gesture recognition system, the machine learning platform TensorFlow is utilized in Python. To enable transfer learning, it is necessary to transform the datasets and label files into a format that TensorFlow can understand. The data is then converted into ".nyp"(numpy) files and saved in the folders containing the training and testing data. To initiate the training of the neural network via transfer learning, the object detection models are downloaded, and the configuration files are modified to correspond to the number of classes in the dataset. Achieving high accuracy requires at least 2000 training steps. To anticipate actions on the screen, particularly in the case of sign language gestures, a Long Short-Term Memory (LSTM) model is created by combining TensorFlow and Keras.

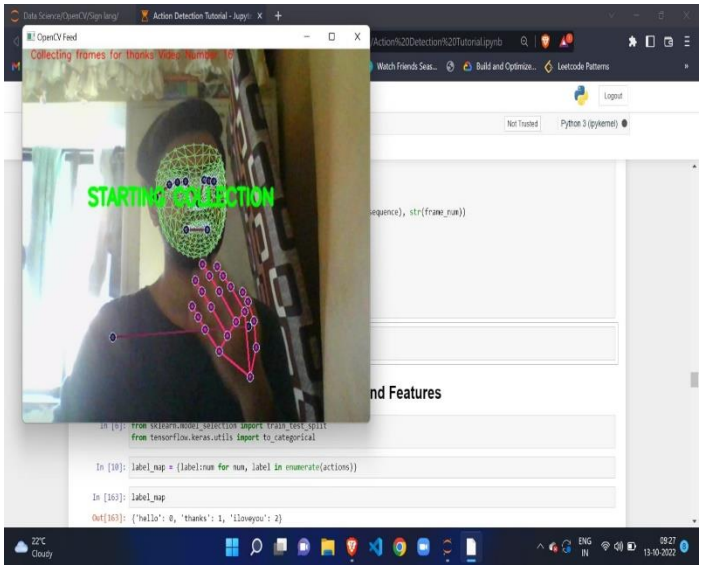

Fig 1: Model training for "Thanks"

Creating Website:

We created a website using Stream lit for hosting our machine learning project. Stream lit simplifies the process of creating interactive web applications by providing a simple and intuitive interface for building user interfaces, integrating data, and running machine learning models. Stream lit allows the user to create easily customizable and easy to share web application.

Now this website can be used from anywhere and by anyone. Also, we added a feature to train customizable gestures for the user. Also, the user can add new words if they want to.

Discussion and results

The primary objective of this project was to make use of deep neural networks and Media Pipe Holistic to determine sign language gestures using forearm, hand, and finger kinematics models. The Media Pipe LSTM model, with data augmentation, attained the highest precision of more than 80 percent on the test sets. This cutting-edge sign language recognition system is equipped to detect, comprehend and display hand gestures and generate coordinators in real-time. It is a remarkable technological advancement that will greatly benefit the deaf and community.

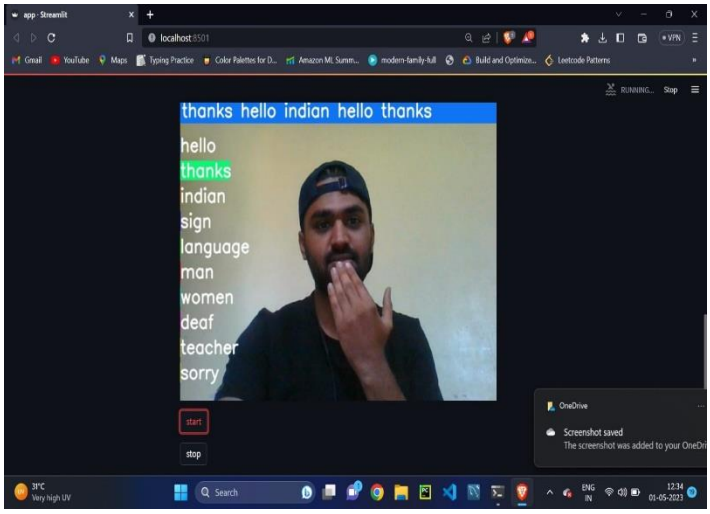

Fig 2: Sign language detection for “thanks”

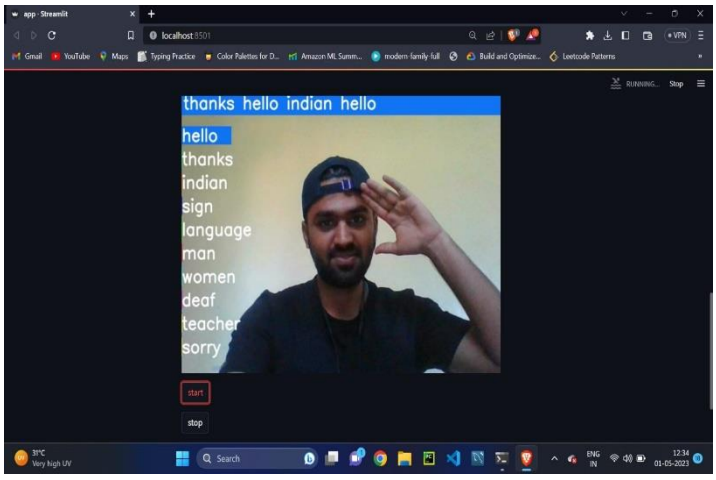

Fig 3: Sign language detection for “hello”

Conclusion

The main goal of this research was to make a Machine Learning Algorithm Model such that it will use TensorFlow, Keras, OpenCV, MediaPipe for detection and interpretation of sign language.

Future improvements in this model and its training process will also help the people with problems like motor impairment to customize their own gestures as they feel comfortable with. This website can be accessed from all around the world if hardware requirements are met.

References

[1] Sundar B, Bagyammal T
American Sign Language Recognition for Alphabets Using Media Pipe and LSTM
4th International Conference on Innovative Data Communication Technology and Application

[2] Kanchan Dabre, Surekha Dholay
Machine Learning Model for Sign Language Interpretation using Webcam Images
2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)

[3] Arjya Das, Mohammad Wasif Ansari, Rohini Basak
Covid-19 Face Mask Detection Using TensorFlow, Keras and OpenCV
2020 IEEE 17th India Council International Conference (INDICON)

[4] Mohammed Safeel, Tejas Sukumar, Shashank K S, Arman M D, Shashidhar R, Puneeth S B
Sign Language Recognition Techniques- A Review
2020 IEEE International Conference for Innovation in Technology (INOCON) Bengaluru, India. Nov 6-8, 2020

[5] Felix Zhan
Hand Gesture Recognition with Convolution Neural Networks
2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science

[6] Ahmed Adel Gomaa Elhagry, Rawan Gla Elrayes
Egyptian Sign Language Recognition Using CNN and LSTM

[7] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg and Matthias Grundmann.

Media Pipe: A Framework for Building Perception Pipelines

[8] Dr. P.Muralidhar, Amartya Saha, Prashanth Sateesh
Customizable Dynamic Hand Gesture Recognition System for Motor Impaired people using Siamese neural network

[9] Rajarshi Bhadra, Subhajit Kar
Sign Language Detection from Hand Gesture Images using Deep Multi-layered Convolution Neural Network
2021 IEEE Second International Conference on Control, Measurement and Instrumentation (CMI), India

[10] Sammon Babu, Grace Joseph
Sign Language Detection using LSTM Deep Learning Model (Action Recognition with Python)
Proceedings of the National Conference on Emerging Computer Applications (NCECA)-2022