Sign Language Detection using LSTM Deep Learning Model (Action Recognition with Python)

Sammon Babu
Master of Computer Applications
Amal Jyothi College of Engineering
Kanjirapally, India
sammonbabu2022b@mca.ajce.in

Grace Joseph
Master of Computer Applications
Amal Jyothi College of Engineering
Kanjirapally, India
gracejoseph@amaljyothi.ac.in

***Abstract* : This project was created intending to use computer vision to be able to recognize the Sign Language in real time with high accuracy. The reason for such a project is to help diminish the gap between those who can hear well and those hard of hearing or even deaf. This can be overcome by creating a dataset of images that correspond to the alphabets, digits or signs applied to deep neural networks. These images are labeled according to the letter being signed. They are processed through a neural network using transfer learning to help the machine "learn" what is being signed after already having been taught on larger datasets of many more images and classifications.**

Keywords : Sign language, Long Short Term Memory Neural Network , MediaPipe Holistic Keypoints, OpenCV

I. INTRODUCTION

Sign language is a form of communication that relies heavily on hand kinematics and facial emotions. Hearing-impaired persons use it frequently to communicate with one another, but it is rarely utilized by non-hearing-impaired people. As a result, they exclusively deal directly with hearing-impaired people, drastically limiting social interactions. Real-time translation with interpreters is an option, albeit it is not always possible and can be quite costly. As a result, an automatic translation system would be quite beneficial. In this subject, many novel strategies have lately been created. We will create a program to translate sign language into OpenCV in this project. It describes a method for recognizing and translating Custom Sign Language into normal text.

A. Motivation and Background

The goal of Sign Language Recognition is to create algorithms and methods for accurately identifying and interpreting the sequences of symbols presented. Many SLR approaches tackle the problem as if it were a Gesture Recognition problem (GR). As a result, research has concentrated on discovering positive qualities and differentiation approaches in order to accurately categorize a particular signal from a group of probable indicators. Sign language, on the other hand, is more than just a collection of well-articulated motions.

B. What is Gesture recognition?

Gesture recognition is a branch of computer science and language technology that includes translating a person's touch using mathematical algorithms. Computer vision is a sub-discipline of computer science. Gestures can arise from any movement or posture of the body, but they are most seen on the face or in the hands. Emotional recognition through face and hand touch detection is now a hot topic in the area. Users can control or interact with devices without touching them by employing simple touch. Many methods have been developed that use cameras and computer vision algorithms to translate the signal language.

C. Sign Language

Sign languages [also known as sign language] are languages in which meaning is conveyed by visual clues. Sign languages and non-sign language objects are both expressed in sign language. Sign languages are fully functional natural languages with their own grammar and lexicon. Although there are some remarkable parallels across sign languages, they are not universal or well understood. Both spoken and signed communications are considered natural forms of language by linguists, implying that they both evolved into a hazy ageing process that lasted longer and evolved over time without conscious planning. Body language, a kind of nonverbal communication, should not be confused with sign language.

Fig 1: Skin Masked Images of different English Alphabets

II. LITERATURE SURVEY

The paper [1], they proposed a novel deep learning-based pipeline architecture using the capabilities of SSD, 2DCNN, 3DCNN, and LSTM for efficient automatic hand sign language recognition from RGB videos. They used

DOI: 10.5281/zenodo.6905809
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

novel hand skeleton features representation in the model for projecting them to three surfaces to feed them to the 3DCNNs to get more rich features. Furthermore, they applied 3DCNNs on pixel level and heat map features to obtain discriminative features.

In the paper [2], One of the challenges in computer vision models, especially sign language, is real-time recognition. In this work, they present a simple yet low-complex and efficient model, comprising single shot detector, 2D convolutional neural network, singular value decomposition (SVD), and long short-term memory, to real-time isolated hand sign language recognition (IHSLR) from RGB video. They employ the SVD method as an efficient, compact, and discriminative feature extractor from the estimated 3D hand key points coordinators.

The paper [3], a deep CNN architecture consisting of 5 layers has been proposed to detect and classify sign languages from hand gesture images. The proposed methodology uses both static (0 - 9 and A - Z) and dynamic (alone, afraid, anger etc.) gestures in the training, validation, and blind testing phase to make the system more robust.

The paper [4] presents a novel approach in the domain of human action recognition. This approach is based on the analysis of video content and extraction features. The Motion features are presented by human motion tracking using GMM and KF methods. Then other features are based on all visual characteristic of each frame on video sequence using Recurrent Neural Networks model with Gated Recurrent Unit. The main advantages of this novel approach are the analysis and the extraction of all features in each time and in each frame of video.

The study [5], compared various conventional machine learning and deep learning models to classify American sign language. Moreover, a robust user-independent k-fold cross-validation and test phase were provided. This contrast previous work, where the validation and/or the test phase were not user-independent, or lack of information was provided.

III. METHODS AND MATERIALS

Other approaches to solving the problem of two-way communication between hearing and non-hearing people have been tried. The majority of ideas incorporate gear that is difficult to transport or many cameras. These designs can be beneficial, but they are not suitable for usage in many public areas by those who are hard of hearing. Furthermore, by needing excessive gear, these designs impede accessibility. This project develops a more user-friendly implementation approach that may be used with mobile or online apps. The only piece of gear required for this project is a camera. To train a neural network, the frames from the video sequence must be captured. Each of the custom gestures must be represented in the collection of frames that are taken. After the frames have been captured, a file must be produced that can describe the gesture to the network. These photos and supplementary files are divided into two folders: training and testing. The testing division accurately depicts how the network will recognise new videos that it has never seen before.

A. Overall Design

In this sign language recognition project, we create a sign detector that detects bespoke signs and can be easily extended to include a wide range of additional signs and hand motions, including the alphabet and numerals.The OpenCV, Mediapipe, Tensorflow, and Keras Python modules were used to create this project. The OpenCV feed examines the frames of live video from a camera to detect the action of a person who is being displayed at that moment in time. To extract keypoints from our hands, torso, and face, the video frames are processed with Media Pipe Holistic. The relevant points will then be passed to the prediction algorithm, which will begin the prediction. The technology then anticipates the hand sign that is being made in real time. The expected sign will also be displayed.

B. Prerequisites

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

- Python (3.10.4)
- IDE (Jupyter)
- Mediapipe (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)

C. Dataset /Labeling

The motions and signals are represented in a dataset. A live stream from the video camera will be available, and every frame that identifies a gesture or motion in the ROI (region of interest) established will be recorded in a directory (here gesture directory). Each sign is represented by around 30 video sequences in the collection. Each video sequence has 30 frames of several important moments, which are stored as Numpy arrays. The gesture being signed must be identified throughout the video sequence.

D. Training

Training is done in python using the Machine Learning platform called TensorFlow. To apply transfer learning, the datasets and label files must be transformed into a format that TensorFlow can process. tfrecord files are created from the folders containing the training and testing data. To begin training a network using transfer learning, the models developed for object detection must be downloaded. To reflect the number of classes in the dataset, the configuration files must be tweaked significantly. To achieve great accuracy, 2000 training steps are required. Tensorflow and Keras combine to create an LSTM model that can anticipate action on the screen, in this example a Sign Language gesture.

DOI: 10.5281/zenodo.6905809
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

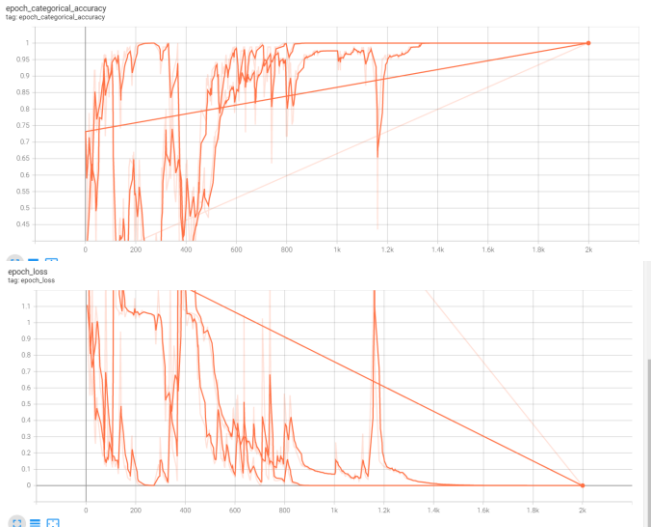

Fig 2: The accuracy and loss presentation in the training and validation phases

IV. DISCUSSION AND RESULTS

The goal of this project was to use deep neural networks and Mediapipe Holistic to anticipate signals using forearm, hand, and finger kinematics models. With 91.1 percent accuracy on the test sets, the Mediapipe LSTM with data augmentation achieved the best results. This sign language detector will be able to recognize and detect hand and produce coordinators as well as understand signs. The signage will all be updated in real time.

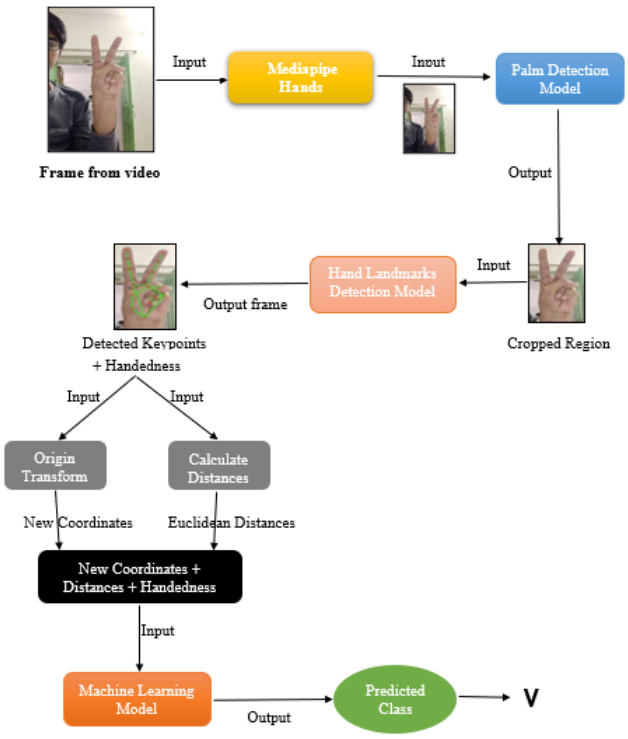

Fig 3: Workflow of the project in real-time

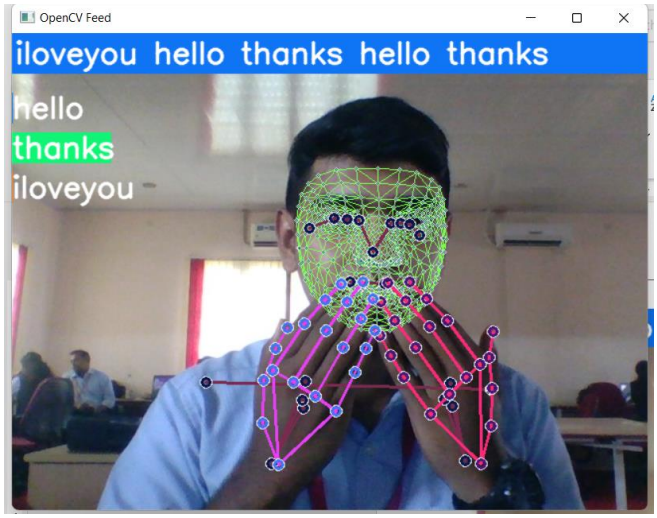

Fig 4.1 Detecting Sign Language for Thanks.

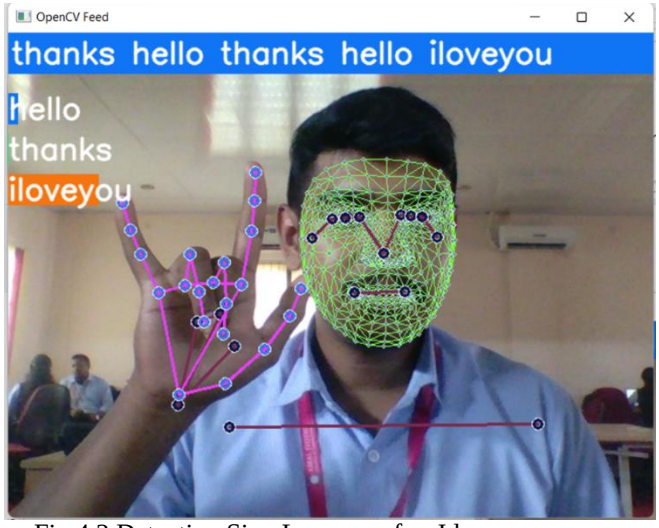

Fig 4.2 Detecting Sign Language for I love you

V. CONCLUSION

The purpose of this research was to build on the initial team's idea so that people can be detected and translated in real time utilising SL.

A future study will potentially take it a step further and develop a mobile application that can classify complete word symbols using facial emotions and relative hand movements from the face, which will be accessible for download on Android and Apple platforms.

REFERENCES

[1] Razieh Rastgooa, Kourosh Kiania, Sergio Escalerab,, "Hand sign language recognition using multi-view hand skeleton",

Electrical and Computer Engineering Department, Semnan University, Semnan 3513119111, Iran

Department of Mathematics and Informatics, Universitat de Barcelona and Computer Vision Center, Barcelona, Spain

DOI: 10.5281/zenodo.6905809
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

Received 18 September 2019, Revised 26 December 2019, Accepted 21 February 2020, Available online 22 February 2020, Version of Record 11 March 2020.

https://doi.org/10.1016/j.eswa.2020.113336

[2] Rastgoo, R., Kiani, K. & Escalera, S. "Real-time isolated hand sign language recognition using deep networks and SVD." J Ambient Intell Human Comput 13, 591–611 (2022). https://doi.org/10.1007/s12652-021-02920-8

[3] R. Bhadra and S. 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 ,2021,pp.196-200,doi:10.1109/CMI50323.2021.9362897.

https://ieeexplore.ieee.org/document/9362897

[4] "Convolutional and recurrent neural network for human activity recognition: Application on American sign language"

Hernandez V, Suzuki T, Venture G (2020) Convolutional and recurrent neural network for human activity recognition: Application on American sign language. PLOS ONE 15(2): e0228869.

https://doi.org/10.1371/journal.pone.0228869

[5] Neziha Jaouedia, Noureddine Boujnahb, Med Salim Bouhlelc, "A new hybrid deep learning model for human action recognition.",

National Engineering School, Avenue Omar Ibn El Khattab Zrig Eddakhlania, Gabes 6072, Tunisia,TSSG-Waterford Institute of Technology, West Campus Carriganore Waterford, X91 P0H, Ireland,SETIT Higher Institute of Biotechnology, Soukra Road km 4 BP 261, Sfax 3000, Tunisia

Received 9 January 2019, Revised 4 September 2019, Accepted 8 September 2019, Available online 9 September 2019, Version of Record 11 May 2020.

https://doi.org/10.1016/j.jksuci.2019.09.004

DOI: 10.5281/zenodo.6905809
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam