Sign Language Detection model using Action Recognition for people with difficulty in speaking

Sandaruwan P.K.U.I.
Department of Electrical Engineering
Univeristy of Moratuwa
Sri Lanka
udithishanka.s@gmail.com

***Abstract*—Speech is still the most common form of communication, however some people have trouble hearing or speaking. For those with these difficulties, communication is a big challenge. Deep learning techniques can be used to break through communication obstacles. The model described in this work uses deep learning to identify and detect words and phrases in a person's actions. To detect signals from isolated video clips, deep learning models LSTM and GRU (feedback-based learning models) are utilized. We are able to create our own dataset, any kind of action from a mute person can be interpreted if the model is properly trained with.**

***Index Terms*—Action recognition, Deep learning, Sign language detection, Machine learning**

I. Introduction

According to the World Federation of the Deaf, there are over 300 sign languages around the world that 70 million deaf people are using them (Murray, 2018). Sign language recognition would help break down the barriers for sign language users in society. Most of the communication technologies have been developed to support spoken or written language (which excludes sign language). While communication technologies and tools such as IMO (Page-bites, 2018) and Whats App (Acton Koum, 2009) have become an important part of our life, deaf people have many problems for using these technologies. Daily communication of the deaf community with the hearing majority community can be facilitated using these technologies.

As a result, sign language, as a structural form of the hand gestures involving visual motions and signs, is used as a communication system to help the deaf and speech-impaired community for daily interaction. Sign language involves the usage of different parts of the body, such as fingers, hand, arm, head, body, and facial expression (Cheok, Omar, Jaward, 2017). [1] In sign language each gesture has a specific meaning. So, therefore complex meanings can be explain by the help of combination of various basic elements. It is basically a non-verbal language which is usually used to deaf and dumb people to communicate more effectively with each other or normal people. Sign language contains special rules and grammar's for expressing effectively [2]. We can detect several common gestures like hello, thanks, I love you, etc. in with any action that we have introduced.

The main goal of this model is to provide better accuracy by using efficient and effective technologies [3]. This model uses TensorFlow and media pipe libraries.And also os library for making and editing directories, NumPy for mathematical calculations ,matplotlib for graph plotting and cv2 for capturing the video clips (or to access the webcam).

II. System Overview

The overall structure of the proposed system is illustrated in Fig. 1. The system is capable of recognizing the action being done by the person in front of the camera. Actions are taken as small video clips by the system and interprets the meaning from the Deep learning model. Facial and hand landmarks are detected from media-pipe library. After defining the landmarks , system take data from the user. In this stage the model is being trained. First we define how many actions we take into the data set and we give the actions clips related to that expression. After data is collected, a neural network is created using Keras API which is merged with TensorFlow. A model is created with an epoch of 2000. After creating the Model with LTSM, System can now identify the actions in front of the camera according to the data ; the model is trained with.

III. User Command Evaluation

Upon collecting data sets for the Neural Network model, We can check the Landmarks of face, hands and body, whether they are detected from the system as it is the foundation of this system. Fig. 2 After configuring the landmarks we keep an array to store extracted key points. Then we create directories for data collection and we keep a separate directory for each action and we gather 30 clips of each action to train the model.

Using the cv2 library we start collecting data and store them in a .npy file and save into the relevant directory. Then we create a neural network model and compile it with

Identify applicable funding agency here. If none, delete this.

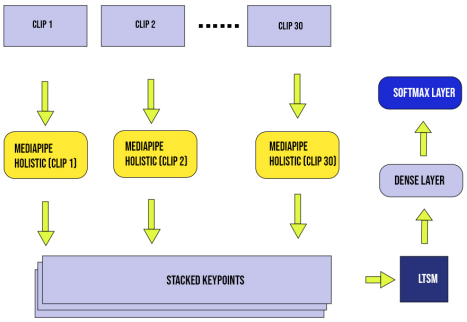

Fig. 1. Structure of the proposed system.

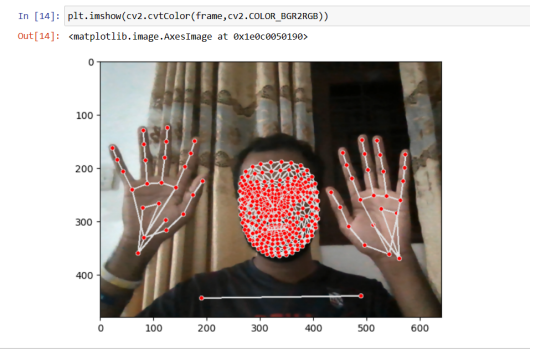

Fig. 2. Landmarks of face,hands,body

optimizer='Adam', loss function='categorical crossentropy'. After that we train the model with 600 epochs. Figure below shows the state after the training of the model.

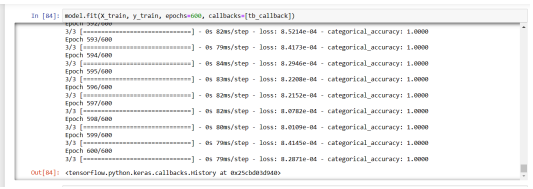

Fig. 3. After the neural network is created.

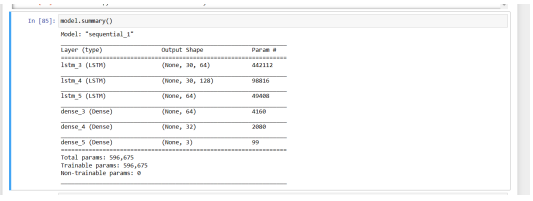

Fig. 4. Model summary

Now we can predict the actions done by any user in front of the camera and interprets what they are trying to say. So,now we can test this model by assuming a word beforehand and we can ask the model to predict the word we assumed, if we get same answers for almost everyone then we can say our model is working. Following is a block of code that has been used to test this accuracy.Fig. 5. Here 'res[1]', 'res[4]' are the assumed values. 'ytest[1]', 'ytest[4]' are the predicted values from the model. As you can see both 'res[1]' and 'ytest[1]' are equal also the same for 'res[4]' and 'ytest[4]'. We can evaluate the accuracy score for each model we create, for this model have gained an accuracy of 1.0 which is the highest that one can achieve so we can say we have obtained a good Deep learning model.Fig. 6. yhat is the predicted value. ytrue is the true value.

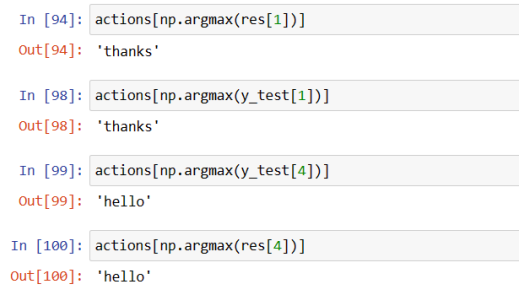

Fig. 5. Model testing

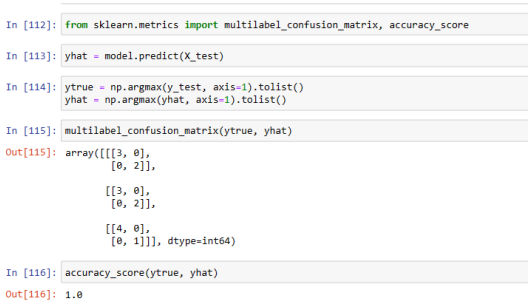

Fig. 6. Model Accuracy Score

Results and Discussion

For this model I have included three actions which are 'hello', 'thanks', 'I love you'. I have given 30 sets of data for each actions.From Fig. 7,Fig. 8, Fig. 9, you can see for each actions the model understands the meaning of the action and the interpreted meaning of the action is put on the blue label of capture. I further improved it by giving a threshold and with the probability. As you can see in the figures I have put the three actions in the left side and with our actions model gives a probability of the interpreted action and it implements a progress bar on the relevant words. Also the blue bar of the model can be used as a sentence generator. It only shows the action interpretation if its not equal to the previous one. So that way the mute people can generate a sentence with this model.

Now Let's discuss the implementation of this module in a practical way. We have to design both Front end and back end APIs and if we are going to make this available for users. We can use this on social media as an extension; which helps

the mute people to communicate through social media as they are helpless against the modern communication. We can have standard data set with sign languages and also we can define some actions in sign language and make them available in the software.Also we can let the user to add their own data set of data which makes it convenient for them. By having this software the normal people can understands what the mute people are saying.

Fig. 7. Action Recognition - Hello

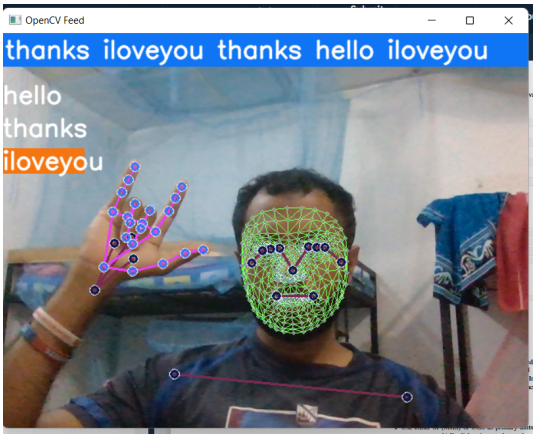

Fig. 8. Action Recognition - I love you

CONCLUSION

This paper proposes a model of sign language recognition using LTSM using Deep learning neural networks. It is able to determine the actions of a mute user and give the output as a word or sentences.This model has room for improvement and more actions can be added to the data set. This if the actions are bit similar, the accuracy can go down. So, when choosing the actions for the words, user need to be mindful about that.

ACKNOWLEDGEMENT

Internet was used to gather the information. Author participated for the data set generation.

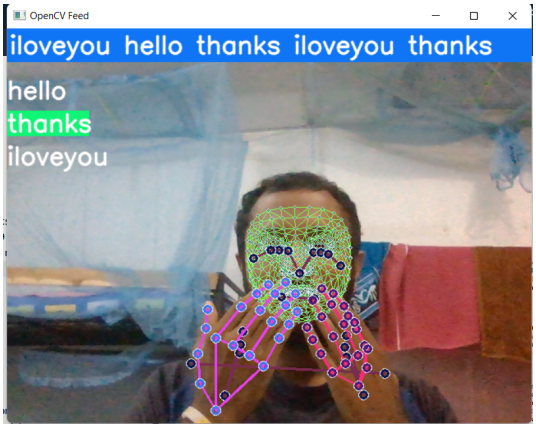

Fig. 9. Action Recognition - Thanks

REFERENCES

[1] Sign Language Recognition: A Deep Survey Author: Razieh Rastgoo,Kourosh Kiani,Sergio Escalera Publication: Expert Systems with Applications Publisher: Elsevier Date: February 2021

[2] A. Deep, A. Litoriya, A. Ingole, V. Asare, S. M. Bhole and S. Pathak, "Realtime Sign Language Detection and Recognition," 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), Ravet, India, 2022, pp. 1-4, doi: 10.1109/ASIANCON55314.2022.9908995.

[3] Kaur, H., Mirza, A., Alankar, B., Chauhan, R. (2022). Sign Language Detection Using Tensorflow Object Detection. In: Mohanty, M.N., Das, S., Ray, M., Patra, B. (eds) Meta Heuristic Techniques in Software Engineering and Its Applications. METASOFT 2022. Artificial Intelligence-Enhanced Software and Systems Engineering, vol 1. Springer, Cham. https://doi.org/10.1007/978-3-031-11713-820