International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 762-764 www.ijaem.net ISSN: 2395-5252

Visibly Audible - Sign Language with Motion Detection Using Sequences and Conversion to Speech

Rashmi R, S K Vinay, Subramanya Kamble, Prof. Mushtaq Ahmed DM

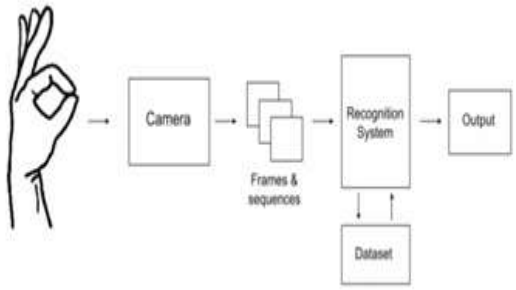
Students, Computer Science and Engineering, AMC Engineering College Bengaluru, India.
Professor, Computer Science and Engineering, AMC Engineering College Bengaluru, India.

Submitted: 05-07-2022 Revised: 15-07-2022 Accepted: 18-07-2022

ABSTRACT-Communication is one of the most important constraint in the modern world, and without interaction, there is a minimal amount of work that occurs. With rapid development in technology, most of the services are being transitioned to the online medium, including communication, which makes the work easy, efficient, and saves time. Our research aims at providing equal opportunity to everyone out there who communicates using gestures to express their thoughts comfortably. We make this happen using an advanced technology under machine learning.
Keywords-Dynamic Hand Gestures, Mediapipe, Machine Learning, OpenCV

I. INTRODUCTION

For individuals who struggle with speaking or hearing, sign language is a successful means of communicating. They organized the gestures or symbols in sign language linguistically. Hand gestures are based on hand signs performed by the user during the interaction, and we convert these gestures into speech and text. To achieve this, we use the concept of deep learning to recognize real-time hand gestures and understand the patterns. The aim is to capture and recognize the sequence of actions which makes up a gesture - generate a meaningful phrase associated with it.



Gesture recognition is a type of perceptual computing that enables computers to record and interpret human gestures. A computer's capacity to comprehend gestures is the general definition of gesture recognition. In its broadest sense, the word gesture can describe any non-verbal communication that aims to convey a particular message. A gesture is any physical movement, big or small, that a camera can recognise in gesture recognition. It includes the form of anything, from a head nod to various actions performed together.
The ability to extract meaningful gesture components from repeatedly performed movements is crucial for gesture recognition. To determine the end point of a gesture, a specific action is considered.

II. RELATED WORK

1. TensorFlow

TensorFlow is an open-source tool for machine learning. It is comprehensive and contains a flexible ecosystem. It helps developers build and train models easily. Although it may be applied on many other tasks, deep neural network training and inference are its main areas of study. An intuitive high-level API like Keras makes for immediate model iteration and easy debugging. Easy to train and deploy models in the cloud or on a device no A simple architecture to take new ideas from concept to code.

2. OpenCV – Python

OpenCV (Open Source Computer Vision Library) is a software library for computer vision and machine learning is In order to speed up the incorporation of artificial intelligence into commercial goods, OpenCV was created to offer a standard infrastructure for computer vision applications.The collection contains more than 2500 optimised algorithms, including

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 762-764 www.ijaem.net ISSN: 2395-5252

several both established and cutting-edge computer vision and machine learning methods.These algorithms can be used to identify landscapes, detect related images in an image database, erase red eyes from flash photos, track eye movements, and create overlay markers.

3. MediaPipe

MediaPipe offers cross-platform, customisable ML solutions for live and streaming media. It is a framework used to create machine learning pipelines for processing time-series data, such as audio and video. The desktop/server, Android, iOS, and embedded devices like the Raspberry Pi and Jetson Nano all support this cross-platform framework.The MediaPipe perception pipeline is called a Graph. We input a stream of photographs, and the output includes hand-rendered landmarks on the image if the desired action is to recognise hands. MediaPipe offers ready-to-use yet customizable Python solutions as a prebuilt Python package.

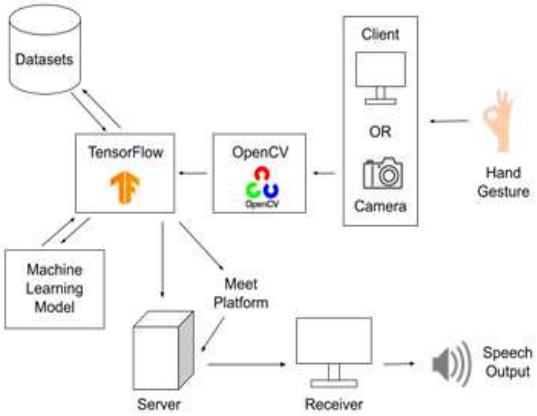
4. LSTM

Long short-term memory (LSTM) is an artificial neural network used in the fields of artificial intelligence and deep learning.LSTM features feedback connections as opposed to typical feedforward neural networks. Such a recurrent neural network may analyse complete data sequences in addition to single data points (such as photos) (such as speech or video).A cell, an input gate, an output gate, and a forget gate make up a typical LSTM unit. The three gates control the flow of information into and out of the cell, and the cell remembers values across arbitrary time intervals.

III. IMPLEMENTATION

This study develops a technology that can assist speech-impaired persons with hand gesture retrieval. We built the proposal using OpenCV, TensorFlow, and Machine Learning. The implementation is done with Jupyter Notebook, as it eases the development. Dependencies to implement are installed and imported.

The first step is to instantiate mediapipe holistic and drawing libraries to extract key points from the frames and sequences of images provided. We customize the draw_styled_landmarks function to show the extracted vital points appealingly. OpenCV-Python is used to interact with the device's camera that captures videos. Mediapipe holistic is customized to use a minimum detection confidence rate of 0.5 and minimum tracking confidence of 0.5. The draw_styled_landmarks function defines the key points and stores the image for every frame captured. The function extract_keypoints exports the generated results into a NumPy array with left-hand, right-hand, face, and pose attributes. We gather 30 sequences of every action comprising 30 frames in each sequence. This is stored in folders for every separate sequence under the parent folder name representing the action.



Considering all these variations, we need to perform some pre-processing of data and create labels and features. The processed labeled data is split into two separate sections. One is for training the model, and the other is used as testing data by a library called train_test_split from sklearn. The function train_test_split is called by passing the ready data, and it returns four sections of data which are x_train, x_test, y_train, and y_test, respectively. Then a deep learning neural network model is built using the LSTM layers. We used the Keras module in the TensorFlow package to build the model. TensorBoard is an exciting and handy feature of TensorFlow that provides a great visualization of how the model is getting trained. We use 3 LSTM layers and 3 Dense layers to create the Machine Learning model, all of which has activation attribute set to "relu". The model is compiled with attributes such as "Adam" as an optimizer, "categorical_cassentropy" for loss, and the metrics having "catergorical_accuracy". Then the model is trained for approximately 2000 epochs, resulting in an accuracy of roughly between 0.95 (95%) and 1.0 (100%). The model summary states the total params and the trainable params, which should both be equal to indicate a successful process. We now save the model, and it is time to test it.

Before directly testing, it is a good practice to evaluate using the confusion matrix and accuracy. The confusion matrix is a matrix used to determine the performance of the classification models for a given set of test data. It can only be determined if the true values for test data are known. The matrix can be easily understood, but the related terminologies may be confusing. We use multi-label_confusion_matrix from sklearn.metrics module. The accuracy returns

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 762-764 www.ijaem.net ISSN: 2395-5252

the one that we found earlier while training the model. Now, the model is tested in real-time.

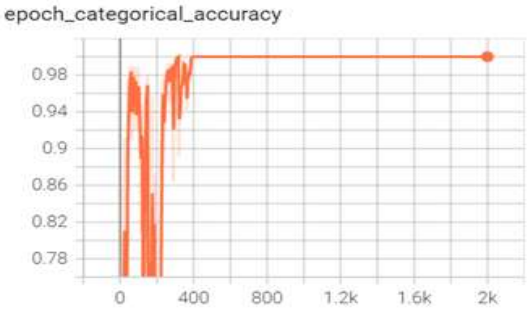
At the camera's start, a person starts making gestures and tries to communicate using sign language. Every frame is captured, and the model tries to fit the sequences and predict the associated action. For every action predicted, a list is appended withthe action's label. We use the pyttsx3 module to convert every phrase into the text to make it look like the person is continuously talking. The computer's processing speed tries to avoid the generated overhead. In this way, we provide virtual speech to a speech-impaired person to communicate with society.

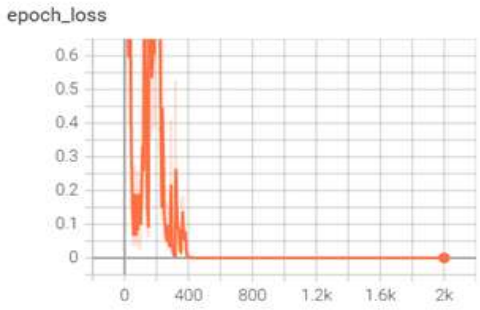
IV. ACCURACY AND PREDICTION

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 30, 64)	48896
lstm_1 (LSTM)	(None, 30, 128)	98816
lstm_2 (LSTM)	(None, 64)	49408
dense (Dense)	(None, 64)	4160
dense_1 (Dense)	(None, 32)	2080
dense_2 (Dense)	(None, 10)	330

Total params: 203,690
Trainable params: 203,690
Non-trainable params: 0





V. CONCLUSION

The interface solution for mute people's practical adaptation is bound by its simplicity and applicability in real world circumstances. We have employed hand gesture to speech and text translation in this approach to allow the reduction of hardware components as a simple and practical technique to establish human-computer interaction. Overall, the approach tries to help people in need while maintaining societal significance. People can speak with no hindrance to one another. The user friendly design ensures that anybody may use it without difficulty or complication. The application is low cost and does not require the use of expensive technology.

REFERENCES

[1]. K. Manikandan, Ayush Patidar*, Pallav Walia*, Aneek Barman Roy*, "Hand Gesture Detection and Conversion to Speech and Text", Department of Information Technology, SRM Institute of Science and Technology.

[2]. Daeha Lee*, Hosub Yoon, Jaehong Kim, "Continuous gesture recognition by using gesture spotting", Intelligent Cognitive Technology Research Department, ETRI DAEJEON, Korea, Oct 2016.

[3]. Iker Vazquez Lopez, "Hand Gesture Recognition for Sign Language Transcription", Master of Science in Computer Science, Boise State University, May 2017.

[4]. Xie Han, Jameel Ahmed Khan, Prof. Hyunchul Shin, "Hand Gesture Recognition Using Deep Learning", Department of Electronics and Communication Engineering, Hanyang University, Sangnok-gu, Korea. 2017.

[5]. Archana S. Ghotkar, Rucha Khatal, Sanjana Khupase, Surbhi Asati, Mithila Hadap, "Hand Gesture Recognition for Indian Sign Language", Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India. Jan 2012.