iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45913

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Sign Language Recognition using Deep Learning

Smt. Sudha V Pareddy[1], Rohit C G[2], B Naveen[3], Adnan[4]
[1, 2, 3, 4]*Department of Computer Science And Engineering, Poojya Doddappa Appa College Of Engineering*

Abstract: ***Sign language is a way of communicating using hand gestures and movements, body language and facial expressions, instead of spoken words. It can also be defined as any of various formal languages employing a system of hand gestures and their placement relative to the upper body, facial expressions, body postures, and finger spelling especially for communication by and with deaf people.***
The project that is being built is to recognize the action performed by the person/user in sign language using Deep learning. Ordinary people are not well versed in sign language.
The project tries to solve this problem using deep learning that is precisely using TensorFlow.
In the project a LSTM (long-short term memory) model in deep Learning is built using TensorFlow to categories the action the user is doing. This will help the user with special needs to communicate with other people using the application we built.
By this we can bridge the gap between the especially abled people and ordinary people.
Keywords: Sign language, Deep learning, LSTM, TensorFlow

I. INTRODUCTION

Deafness has varying descriptions in artistic and medical terms. In medical terms, the meaning of deafness is hearing loss that precludes a person from understanding spoken language, an audiological condition. In this term it's written with a lower cased. In a medical terms, deafness is defined as a degree of sound loss similar that a person is unfit to understand speech, indeed in the presence of modification.

In profound deafness, indeed the loftiest intensity sounds produced by an audiometer (an instrument used to measure sound by producing pure tone sounds through a range of frequency) may not be detected. In total deafness, no sounds at each, anyhow of modification or system of product, can be heard. A mute is a person who doesn't speak, either from an incapability to speak or reluctance to speak. The term" mute" is specifically applied to a person who, due to profound natural (or beforehand) deafness, is unfit to use eloquent language and so is deaf-mute.

The problem is that there exists a communication hedge between normal people and especially- abled people as the normal person aren't clued in sign language and isn't suitable to communicate with especially- abled person. The ideal of this design is to give a communication result in- form of an operation that can fete the sign- language and give the affair in form of textbook that can be fluently understood by the normal person. can be fluently understood by the normal person. We prognosticate the sign language deep literacy that's using long term short memory algorithm this algorithm is a neural network that helps us to prognosticate the action performed by this especially- abled person dictation. In this way it decreases the communication hedge between a normal person and an especially- abled person (Deaf and Mute person). A homemade translator cannot always be present to restate the conduct of an especially- abled person and help him to overcome the difficulties faced by him in the communication with others who don't know sign- language used by the person. Our proposed system will help the deaf and hard- of- hail communicate better with members of the community. For illustration, there have been incidents where those who are deaf have had trouble communicating with first askers when in need., it's unrealistic to anticipate everyone to come completely fluent in sign language. Down the line, advancements like these in computer recognition could prop a first polled in understanding and helping those that are unfit to communicate through speech.

Another operation is to enable the deaf and hardy- of- hard equal access to videotape consultations, whether in a professional environment or while trying to communicate with their healthcare providers via telehealth. rather of using introductory converse, these advancements would allow the hearing- bloodied access to effective videotape communication.

The design being erected is an operation that can fete the stoner's conduct and restate that action to textbook and speech. The operation is doing this using deep literacy, that's we're erecting a model that will fete the conduct and orders that action and translates it to textbook and speech.

©IJRASET: All Rights are Reserved

II. LITERATURE REVIEW

In this project we are building a model that recognizes the actions that are signs in the sign language. The problem that it is trying to solve is the communication barrier between normal people and specially-abled people as the normal person is not versed in sign language and is not able to communicate with specially-abled person. We are using deep learning neural network that is LSTM (Long Short-Term Memory) to train our model. We are using Mediapipe Holistic to get the landmarks on pose, face and hands. A video is captured using camera with landmarks.

The captured data is then split in training and testing data set. LSTM model is trained using the train data set and is tested with the testing data set and the weights are arranged to improve the accuracy of the predictions. We use computer vision that is to capture the data in form of video.

The result is output in form of text and is displayed on the screen. Normal person can understand the text and gets to know what the person using sign language. For this project we referred Deep Learning book by Ian Goodfellow.[1] This book was used to get a better understanding about Machine learning especially the RNN (Recurrent Neural Network) and LSTM algorithm. We also used the techniques involved in the project: Indian Sign Language Using Holistic Pose Detection [2] by Aditya Kanodia1, Prince Singh, Durgi Rajesh, G Malathi to understand the working of OpenCV and use of Mediapipe Holistic to get or extract the key features that is the data used in as an input to the model.

Introduction to TensorFlow by Oliver Dürr [3] helped us to understand the working of TensorFlow and how it helps us in building the LSTM model and importing other dependencies.

We referred many models that had common goal but utilized different techniques [2] [4] [5] [6]

Currently there are models that work on CNN but the issue with them is that they need huge amount of data compared to LSTM that is nearly 10 times when comparing the parameters needed to train our model. The algorithm that we are using that is LSTM is faster than the CNN model as it uses less data when compared to the latter.

III. PROPOSED METHOD

This work uses a class of Artificial Neural Network knows as Recurrent Neural Network for creating a model and training it for classifying the sign-actions made by any user and produce the text output for it. To be specific it uses an advanced version of the RNN called as LSTM (Long Short-Term Memory) to create the model. It has a significant advantage over the original RNN as it suffered from vanishing gradient problem.

The LSTM doesn't suffer from the vanishing gradient problem as it forgets the irrelevant data and stores only the important data. Many computer-vision application tend to use CNN (Convolution Neural Network) but it requires too many training samples as compared to LSTM. Using LSTM, the results achieved were of comparatively high to that of the model built using CNN. Only 10 percent of the samples were required to build the model as compared to the model built using CNN. It proved to be a lot faster as it used less data.

The application was also fast to recognise the actions. Mediapipe holistic is used to add landmarks. The Landmarks are drawn on the face, hand and the body of the person in front of the camera, these landmarks represent key points these key points are extracted using computer vision.

The key points give us the exact location of the user's hands from the camera and the spatial representation of the gesture made by the user the representation of key points is in the terms of X, Y and Z coordinates.

A. Train Deep Neural Network with LSTM for Sequences

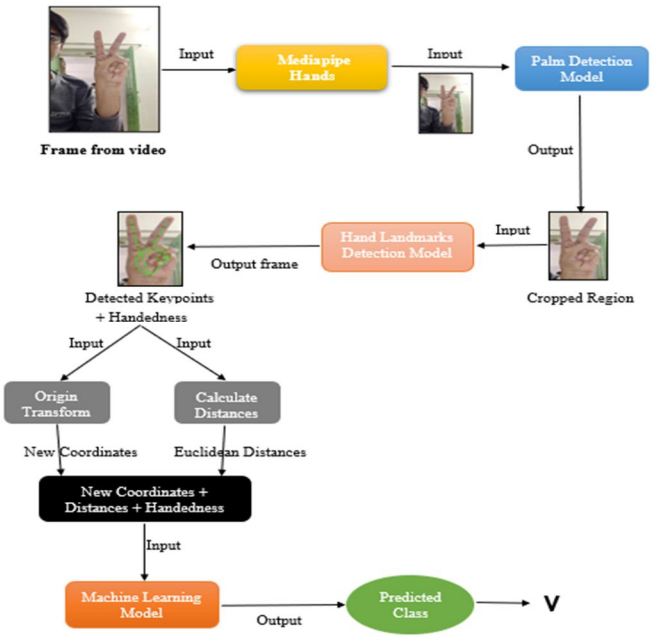
The user performs various actions and captures these actions using KeyPoint that are drawn on the hands of the user. This way the data set for training the neural network is produced. The same action is performed 30 times and 30 different folders are created on the system, the data in these folders will be used for training the deep neural network using LSTM. The model undergoes multiple epochs and the process is stopped when the accuracy is at its peak and starts to decline afterwards.

B. Perform Real Time Detection using OpenCV

The data used to create a model that can predict the sign that the user is performing. The user performs an action in front of the camera while OpenCV feed is active. The open CV uses the LSTM model that we trained to predict the action being performed in front of the camera by the user and produces a plain text as an output for the action that has been performed. The plain text represents the action that was performed.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

In this way we are able to detect the sign performed in front of the camera in real time using open CV and LSTM



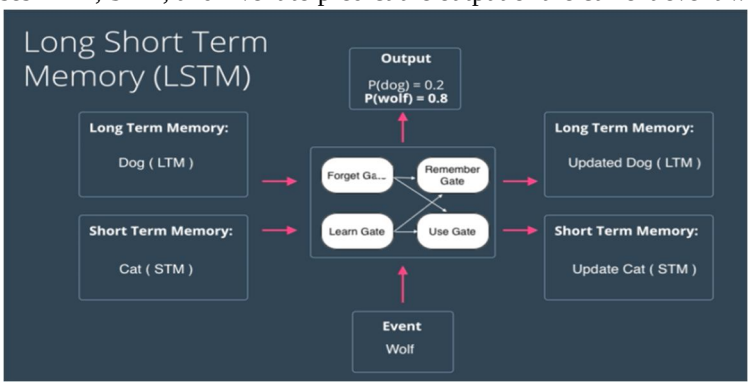
Sign language Recognition using Deep learning

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can process not only single data points (such as images), but also entire sequences of data (such as speech or video). For example, LSTM is applicable to tasks such as unsegmented, connected handwriting recognition, speech recognition and anomaly detection in network traffic or IDSs (intrusion detection systems). A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.

C. The Architecture of LSTM

LSTMs deal with both Long-Term Memory (LTM) and Short-Term Memory (STM) and for making the calculations simple and effective it uses the concept of gates.

1) *Forget Gate:* LTM goes to forget gate and it forgets information that is not useful.
2) *Learn Gate:* Event (current input) and STM are combined together so that necessary information that we have recently learned from STM can be applied to the current input.
3) *Remember Gate:* LTM information that we haven't forget and STM and Event are combined together in Remember gate which works as updated LTM.
4) *Use Gate:* This gate also uses LTM, STM, and Event to predict the output of the current event which works as an updated STM.

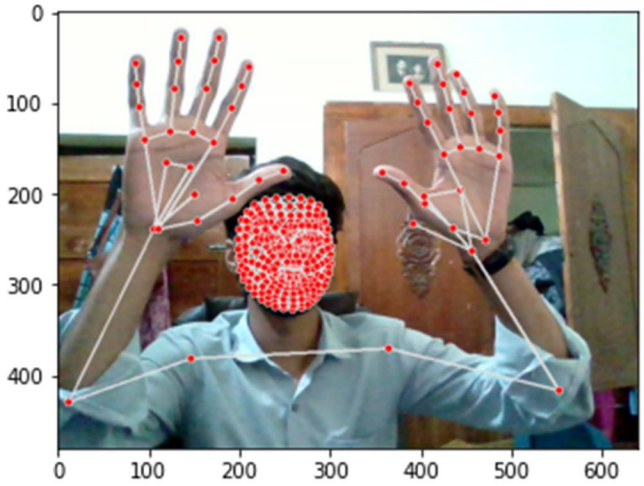


ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

IV. IMPLEMENTATION

A. Collecting the Data for Creating the data Samples

The implementation for the proposed system is done in jupyter notebook. The language used is python 3.9 for the proposed system. The keras library from the TensorFlow library is used to import the LSTM model required for training. The OpenCV is used to capture the actions for the training and testing. Mediapipe Holistic is a pipeline used to create the landmarks that serves as the keypoints. The landmarks from the user's hands are captured and saved in the file. This process repeated for 30 times for each action that is to be included.



B. Training the LSTM Model

The LSTM network is imported from the Keras library which comes under TensorFlow. The data which is collected and stored in the folder is fed to the model. The model is then run for many epochs and the accuracy is seen using TensorBoard, when the accuracy reaches its max value and begins to fall the execution is stopped and the model is saved. The model is used to recognise the user's action and produce an output in the form of plain text.

C. Testing the Model

The model is saved and stored in the local machine. The application is deployed and the user can now begin to provide arbitrary input in the form of action (sign in sign-language), the input is fed to the model and the prediction is made. The predicted action is shown to the user in form of text.

V. CONCLUSION

The work successfully covers the commonly used gestures and interprets them into a sentence at high speed and accuracy. Recognition of the gestures does not get affected by the lighting of the environment, color or size of the person. This Application requires less data when compared to the applications that were built on the CNN algorithm. It is also faster to train as it takes less data as the input. It also performs faster detections when compared to a CNN model. And also got good accuracy score for the validation data. We are including as many words as possible in the near future. Model training gets complex as the number of different words increases. As this work can bridge a gap between normal people and disabled people. Hence, our future enhancements or work would primarily focus on two things.

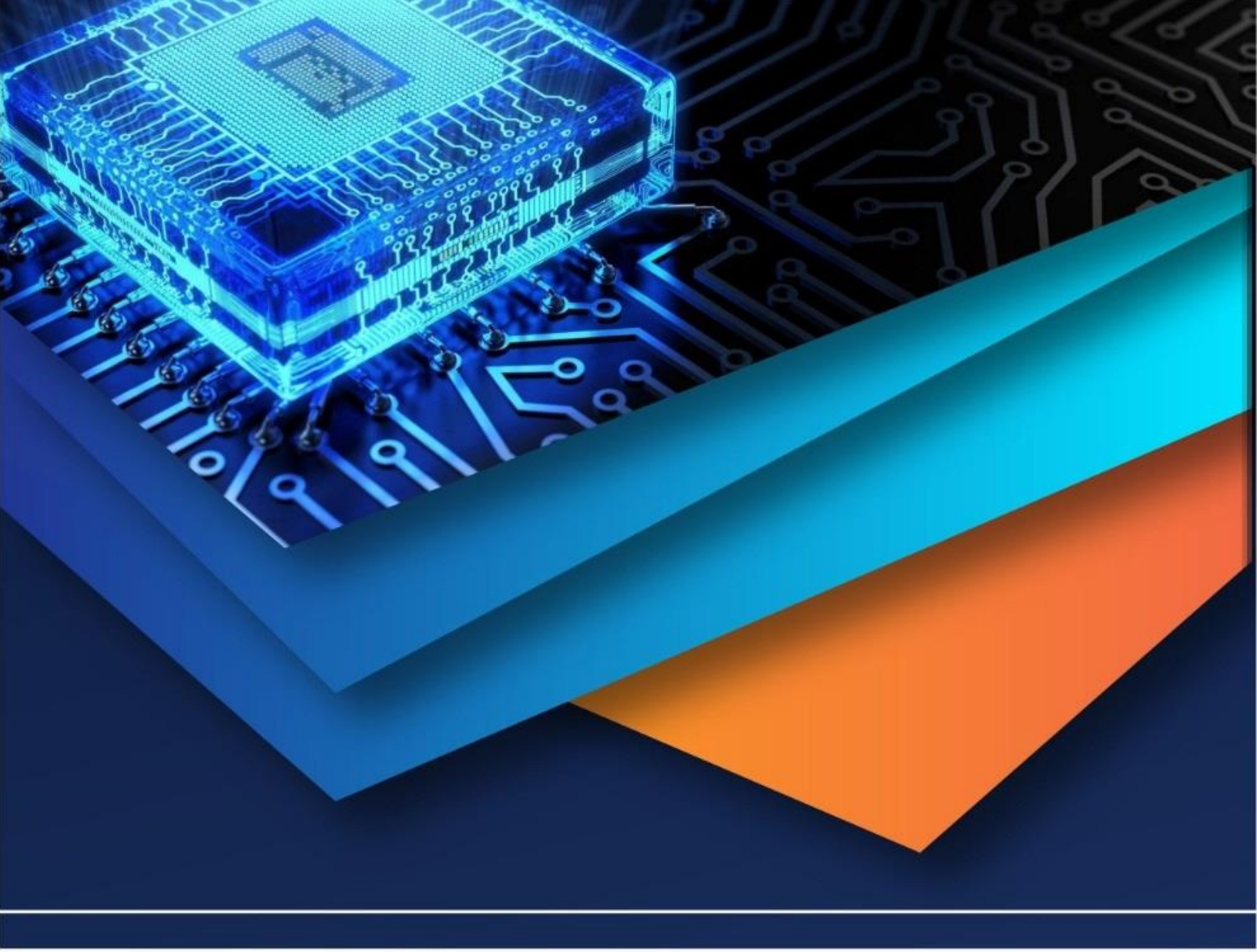
REFERENCES

[1] S. Nikam and A. G. Ambekar, "Sign language recognition using image-based hand gesture recognition techniques," 2016 Online International Conference on Green Engineering and Technologies (IC-GET), 2016, pp. 1-5, doi: 10.1109/GET.2016.7916786.
[2] S. Suresh, H. T. P. Mithun and M. H. Supriya, "Sign Language Recognition System Using Deep Neural Network," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019, pp. 614-618, doi: 10.1109/ICACCS.2019.8728411.
[3] S. Gupta, R. Thakur, V. Maheshwari and N. Pulgam, "Sign Language Converter Using Hand Gestures," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. 251-256, doi: 10.1109/ICISS49785.2020.9315964.
[4] Jayaprakash, Rekha & Majumder, Somajyoti. (2011). Hand Gesture Recognition for Sign Language: A New Hybrid Approach. 1.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

[5] Starner T, Weaver J, Pentland A (1998) Real-time American sign language recognition using desk and wearable computer-based video. IEEE Trans Pattern Anal Mach Intell 20:1371–1375
[6] Moryossef, Amit & Tsochantaridis, Ioannis & Aharoni, Roee & Ebling, Sarah & Narayanan, Srini. (2020). Real-Time Sign Language Detection using Human Pose Estimation.
[7] Pose-based Sign Language Recognition using GCN and BERT Anirudh Tunga* Purdue University atunga@purdue.edu Sai Vidyaranya Nuthalapati* vidyaranya.ns@gmail.com Juan Wachs Purdue University
[8] Z. Yao and X. Song, "Vehicle Pose Detection and Application Based on Grille Net," 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE), 2019, pp. 789-793, doi: 10.1109/EITCE47263.2019.9094787.
[9] J. Su, X. Huang and M. Wang, "Pose detection of partly covered target in micro-vision system," Proceedings of the 10th World Congress on Intelligent Control and Automation, 2012, pp. 4721-4725, doi: 10.1109/WCICA.2012.6359373.
[10] Siddhartha Pratim Das, Anjan Kumar Talukdar, Kandarpa Kumar Sarma, Sign Language Recognition Using Facial Expression, Procedia Computer Science, Volume 58, 2015, Pages 210-216,
[11] Das, Siddhartha & Talukdar, Anjan & Sarma, Kandarpa. (2015). Sign Language Recognition Using Facial Expression. Procedia Computer Science.58. 10.1016/j.procs.2015.08.056.
[12] Vahdani∗ · Matt Huenerfauth · Yingli Tian
[13] Sign Language Recognition Helen Cooper, Brian Holt and Richard Bowden
[14] Aditya Kanodia, Prince Singh, Durgi Rajesh, G Malathi: INDIAN SIGN LANGUAGE USING HOLISTIC POSE DETECTION.
[15] Agarwal A, Thakur MK (2013) Sign language recognition using Microsoft Kinect. In: IEEE Sixth International Conference on Contemporary Computing (IC3), pp 181–185.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)