Grenze International Journal of Engineering and Technology, Jan Issue

Sign Language Interpreter using Convolutional Neural Network

Ranjana.Battur[1], Arundhati Nelli[2], Pavan Kunchur[3], Sadhana Bangarashetti[4] and Nelson Lobo[5]
[1-3]Asst Prfoessor, Department of Computer Science and Engineering,KLS,Gogte Institute of Technology,Belagavi
EmailID: rbbattur@git.edu, pnkunchur@git.edu
[4]Professor, Department of Information Science and Engineering, Basaweshwar Engineering College, Bagalkot
Email: sadhanapb@gmail.com
[5]Department of Computer Science and Engineering, KLS, Gogte Institute of Technology, Belagavi
Email: 2gi18cs422@students.git.edu

***Abstract*—People with hearing and discourse debilitation cannot bring using discourse. They depend upon conversation through gestures for correspondence. communique through gestures is utilized by every and every person whose discourse hindered, however, they enjoy troubles speaking with unsigned people. So individuals with discourse disabilities want a gesture-based totally communique translator. This makes their casual and formal correspondence difficult. The proposed framework endeavors to make an ongoing interpretation of hand signals into comparable English content and later on convert it to sound. The framework takes hand signals as contribution thru video and deciphers text this is justifiable to the unsigned. there is the utilization of Convolutional Neural community and Recurrent Neural network classification for the identity and order of hand signals. through sending this framework, the correspondence hollow among the underwriter and the non-endorser. therefore, a mechanized mediator can be a stable and reasonably priced other choice. A framework that may make an interpretation of gesture-based totally verbal exchange into simple content or sound aides continuously correspondence.**

***Index Terms*— Neural network , gestures , pre-processing.**

I. Introduction

The motive of this model is to format a system that is useful for people who have listening to and speech troubles and in favored people who use quite simple and effective techniques sign language. This model may be used for changing sign language to textual content and voice. A gesture popularity system is used for sign language detection and an audio API for voice conversion. The version captures the symptoms and symptoms and dictates at the display as writing. It additionally captures in movement and shows the sign language this means that on the show as motioned textual content and audio. we have got were given carried out a model to ease the communique hole some of the impaired guys and women (deaf and dumb) and an ordinary guy or woman. The version takes enter within side the shape of video. The video captured or fed is then converted into frames which help us within side the characteristic extraction method and map the sign that turns into expressed to the actual closely associated sign from the dataset. as quickly as characteristic mapping is performed, a related text token is generated, that is fed to the Audio API to translate the sign interpreted to audio format. accordingly, supporting

Grenze ID: 01.GIJET.8.1.512_1

in the translation of sign language to a comprehensible format. The version may be able to spotting a particular subset of a sign language dialect. there are numerous dialects of sign languages and frequently many dialects for the same ethnic language. or even, in that case, each dialect can embody many thousand symptoms and symptoms and symptoms. using appropriate algorithms to recognize and extract the characters from the signal captured in the video. The recognized gestures will need to be translated into both text and/or speech. due to the fact a laptop system is concerned, it's miles low priced to expect that there is a show wherein text may be displayed. ideally, textual content to speech is used for more practical communique. Speech-to-textual content popularity could be used to permit bidirectional conversation.

II. METHODOLOGY

The entire methodology is divided into 4 steps namely Video Input, Pre-Processing, Model Input and Output. **Video input:** Hand Gestures are given as input in to the system in the video format. Then the video would be converted into individual frames. After that pre- processing would be carried out on the frames.
Pre-processing: In pre-processing the frames are converted from RGB to grayscale and the noise in the image is reduced and the size is reduced. The size is reduced that the computational power required will be less. The skin tone values are selected to isolate the hands form the background. Then it is converted into the grayscale image to obtain the features of the hand. **Model input:** The frames are then separated as training and test sets and then the training set is fed to the model for training. Next the testing frames are fed into the trained neural network. At last the neural network would classify the hand gesture in the frame depending on the most appropriate feature that as mapped from the dataset. Output: The mapped sign from the dataset is then sent out as a text token, this text token can be read by the user. As the final output the generated text is converted into audio format. The data flow design is shown in figure 1. The various steps followed are-
Capture Video: Capture video from the webcam of a laptop or from a mobile phone. Divide video into frames: In this phase we split the video captured from the camera into a set of frames so that each frame can be processed by the system individually. Image Processing: This phase deals with noise reduction and scaling of high resolution images. Feature Extraction: It is the process of using Mediapipe Holistic for Live perception of simultaneous human pose, face landmarks, and hand tacking in real-time. Classification: This phase deals with classification of data using Neural Networks. Output Text: The classified data is processed to get output text. TTS API: Uses speech synthesizer to convert text data into audio. Audio Output: The output of TTS API.

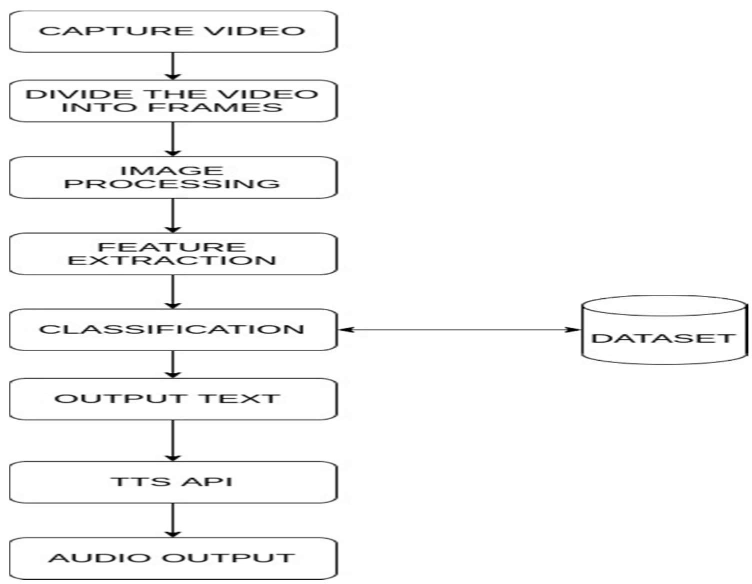

Figure 1: Data Flow Design

A standard sequence classification architecture is used. The model takes as input a series of feature vectors, constructed from a flat vector representation of the pose concatenated with the 2D angle and length of every limb, using the pose-format library. These representations are subjected to a 20% dropout, normalized using 1D batch normalization, and are projected to a lower dimension hidden state vector (512 dimensions). This is then

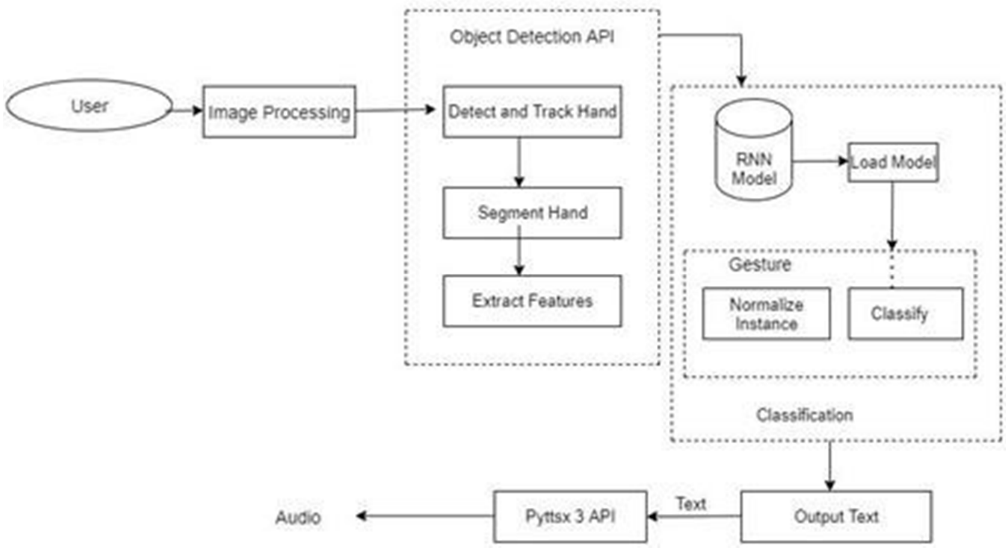

Figure 2. Software Architecture

passed to a two layer BiLSTM with hidden dimension 256, followed by a max-pooling operation to obtain a single representation vector per video. Finally, the output is fed to a linear layer and softmax to give probabilities for each class. The model is trained using cross-entropy loss with the Adam optimizer (with default parameters) and a batch size of 512 on a single GPU. No data augmentation or frame dropout is applied at training time, except for horizontal frame flip to account for left-handed signers in the dataset. The Software architecture is shown in figure 2.

III. LITERATURE SURVEY

To make a communication bridge, a highly accurate, cost effective and an independent glove was designed for deaf/mute people to enable them to communicate. The glove translates the sign language gestures into speech according to the American Sign Language Standard [1]. The glove contained flex and contact sensors to detect the movements of the fingers and bending of the palm. In addition an accelerometer was built in the glove to measure the acceleration produced by the changing positions of the hand. Principal Component Analysis (PCA) was used to train the glove into recognizing various gestures, and later classify the gestures into alphabets in real time. The glove then established a Bluetooth link [2] with an Android phone, which was used to display the received letters and words and convert the text into speech [1], At least 9 analog voltage channels for flex sensors and at least 10 digital channels for contact sensors were required. To enhance the performance and accuracy of the system, signal conditioning for accelerometer readings was needed. Baud rate was set at 9600 and sampling rate at 500 samples per second. The device used was: NI PCI-6250. SCB-68A I/O block was used for interfacing signals to plug in DAQ devices with 68 pin connectors [3]. The cable used was NI SHC68-68-EPM which was specially designed to work with NI X Series and M Series data acquisition (DAQ) devices, and included separate digital and analog sections, individually shielded twisted pairs for analog inputs, individually shielded analog outputs, and twisted pairs for critical digital I/O. The NI PXI 8330 module was used to connect an external controller (our desktop computer) to a PXI chassis. It was paired with a PCI 8330 in desktop. computer with a MXI cable running between the two. It contained a MXI-3 Multi-System Extension Interface for PCI [4]. C-05 (Bluetooth to Serial Port Module) was used because of its ease of availability in Islamabad and its low cost. Service Discovery Protocol (SDP) was used to allow devices to discover what services each other support, and what parameters to use to connect them. Each service was identified by a universally unique identifier. The default setting of serial port of HC05 was: baud 38400 and our serial communication in the Arduino was happening at a baud rate of 9600, thus baud rate of Bluetooth module of SignSpeak was needed to be matched. To change it, a command was entered through the AT mode on the module. The module was set as Master mode [5].

The connection steps of Arduino and Bluetooth module remained the same. For receiving the data on an android device, an application was developed for this purpose which used Bluetooth API. The application was written in

Java using Eclipse, by modifying the Bluetooth Chat sample code from Android. Two APIs Bluetooth API (for scanning of other Bluetooth devices and paired Bluetooth connections) and Text to Speech API (for conversion of text into speech featuring more than 30+ languages and implementation of different features such as voice, speed and pitch) were used [6].

IV. Results and Discussions

Initially the input image is collected from the various sources

Figure 1: Hand and Facial Image

Figure 2: Face Image

Figure 3: Hand and Face Image with different posture

First we Capture a set of frames from the inbuilt camera. These frames are then preprocessed for noise reduction and then downscaled for faster processing. Then we make use of MediaPipe Holistic to mark key points in order to track hand and facial movements. MediaPipe is a framework for building multimodal (eg. video, audio, any time series data), cross-platform applied ML pipelines. With MediaPipe, a perception pipeline can be built as a graph of modular components, including, for instance, inference models (e.g., TensorFlow, TFLite) and media processing functions. The key point data is then passed to the neural network for training and testing. The same key point data can be used for recognition of Sign Language once the model has been trained and tested which is shown in Figure 1,2,3.
The Sign Language Interpreter model has been successfully implemented to translate the signs received from the impaired persons to understandable form (Text and Audio). The model accuracy is evaluated dynamically in the capture window by varying colors for good accuracy which is shown in Figure3.The accuracy is high. The model Loss is also shown in Figure 4.

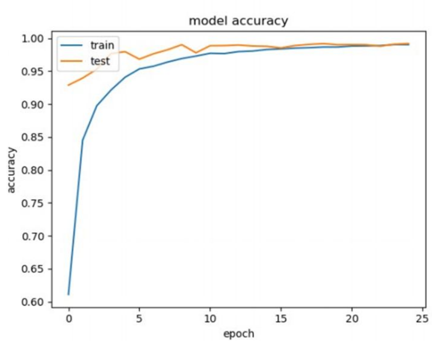

Figure 3. Results showing model accuracy

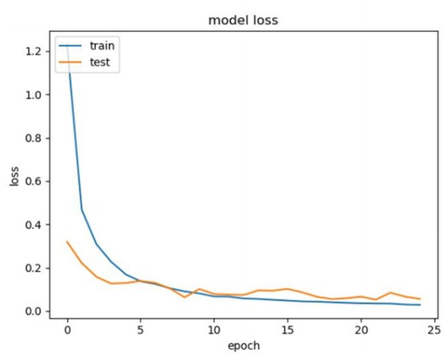

Figure 4. Results showing model loss

V. Conclusion

We have explained the implementation of the system of translating Indian Sign Language into English. We discussed the importance of ISL translators for communicating with the deaf and dumb. In this system, the hand image frame is captured using a webcam. These frames are refined to get enhanced features. Then feature extraction and classification algorithms are used to translate sign language into English text. This translation is converted from text to speech using the Speech API. The system is implemented using the above algorithms to get the final output. The proposed model is evaluated by a dataset containing 40 - 100 symbols from two different people.

For future coverage, this model can be trained with different datasets of different sign languages around the world. Similarly, the model can be integrated into other chat applications for communication purposes. It can be updated to teach others different sign languages that display two-way functionality.

REFERENCES

[1] Hai, P. T., Thinh, H. C., Van Phuc, B., & Kha, H. H. (2018). Automatic feature extraction for Vietnamese sign language recognition using a support vector machine. 2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom).

[2] Purva C. Badhe, Vaishali Kulkarni, “Indian Sign Language Translator using Gesture Recognition Algorithm”, 2015 IEEE International Conference on Computer Graphics, Graphics and Information Security (CGVIS).

[3] Md Azher Uddin, Shayhan Ameen Chowdhury, “Hand Sign language recognition for Bangla Alphabet using Support Vector Machines”, 2016 International Conference on Innovations in Science Engineering and Technology (ICISET).

[4] Muhammad Aminur Rahaman, Mahmood Jasim, Md. Haider Ali and Md. Hasanuzzaman, “Real- Time Computer Vision based Bengali Sign Language Recognition", 2014 17th International Conference on Computer and Information Technology (ICCIT).

[5] Rahat Yasir, Riasat Azim Khan, “Two Handed Hand Gesture Recognition for Bangla Sign Language using LDA and ANN”, The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014).

[6] P.Gajalaxmi, T, Sree Sharmila, “Sign Language Recognition for Invariant features based on multiclass Support Vector Machine with BeamECOC Optimization”, IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017).

[7] Sign Language Interpreter using Image Processing and Machine Learning by Omkar Vedak, Prasad Zavre, Abhijeet Todkar, Manoj Patil International Research Journal of Engineering and Technology (IRJET)