International Journal of Computer Applications Technology and Research
Volume 13–Issue 08, 34 – 38, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1308.1004

Two-Way Tamil to Indian Sign Language Translator Using Mediapipe and Natural Language Processing

Kalaimathi S
Dept. of Electronics and Communication Engineering, Puducherry Technological University, Puducherry, India

Mangaiyarkarasi M
Dept. of Electronics and Communication Engineering, Puducherry Technological University, Puducherry, India

Moginder E
Dept. of Electronics and Communication Engineering, Puducherry Technological University, Puducherry, India

Rihana M
Dept. of Electronics and Communication Engineering, Puducherry Technological University, Puducherry, India

Thachayani M
Dept. of Electronics and Communication Engineering, Puducherry Technological University, Puducherry, India

Abstract: Sign-language is the primary mode of communication for the hearing- or speech-impaired people. With the technological advancements in machine-learning and computational technologies the models are developed for the translating sign-language to other natural language. In this paper, a two-way translation model based on CNN-LSTM is presented for the translation of Indian sign language to one of the India's regional language, Tamil. The proposed model is trained and tested for a subset of commonly used words and resulted in an accuracy of 92.9% and a precision of 93.7%. This two-way ISL communication system aims to revolutionize the way ISL is taught and learned.

Keywords: Indian sign language; Assistive technology; Sign-language; Natural language processing; Gesture recognition

1. INTRODUCTION

Sign language is a form of communication using hand motions, facial expressions, and body postures, enhances communication for those with hearing impairments or speech disorders. Various countries have their sign languages, including ASL, BSL, ISL, Arabic Sign Language, and others [1], [2]. Each sign language has its characteristics; for example, ISL requires both hands, while ASL can be communicated with one hand.

Indian Sign Language (ISL) is a visual-gestural language used by the hearing-impaired community in India which allows them to express themselves, share ideas, and connect with others. It is a rich and complex language with its own syntax, grammar and vocabulary [1]. ISL is a vital means of communication for these people and the main challenge faced by ISL users is the lack of resources for learning and practicing the language. Traditional methods of teaching ISL often rely on in-person instruction, which may not be easily accessible to all individuals. To address this issue, researchers have explored the use of technology to develop digital tools for learning ISL [3], [4]. This paper, presents a two-way translation of the sign language to one of the India's regional language, Tamil. By harnessing the power of CNNs and LSTMs, this two-way ISL communication system has the potential to revolutionize the way ISL is taught and learned.

2. LITERATURE SURVEY

There are several basic techniques used to develop sign language translation systems. An extensive review of the literature in sign language translation is reported in [4]. Few other relevant literature follows. A mobile application for speech to ISL translation based on Google API with 91% accuracy of prediction is reported in [5]. In this work, animations are created from recorded videos for various words and stored as database. The speech is converted into text by Google API and an ISL parser is used to separate the keywords. Then the corresponding animation for the keywords are displayed. English speech to ISL sign translation is done using the natural language processing tokenization and lemmatization and the corresponding videos for the keywords are played from the video database using Youtube IFrame API [6]. An online application is developed to translate sentences in English or Hindi to ISL gesture animations [7]. The testing with the data base of more than 300 common sentences resulted in a

similarity score or accuracy of 95%. A hand gesture recognition of English alphabets in ASL using LSTM which utilizes Google-Mediapipe for getting the landmark from the custom video is reported in [8]. The accuracy for alphabet detection is reported to be 99%. Bora et. al., reported Assamese sign language recognition system based on feed-forward neural network which uses Mediapipe for landmark detection and reported an accuracy of 99% in detecting the nine alphabets [9]. Another speech to ISL based sign-language translation system using Google API is reported in [10]. In this work the speech is first converted to text and the words are mapped using LSTM, Bi-LSTM and Google API based classifiers. The corresponding video animations from the data base are displayed and the system reported accuracy of 45%, 65%, and 95% respectively. A custom CNN-RNN model for converting sign-gestures into English text with an accuracy of 89.99% is reported in [11]. In this work gestures are recognized using the CNN model and RNN-LSTM is used for semantic verification of the sentences.

Most of the existing systems consider English and one-way translation models. Few work reported on Tamil language are based on transliteration of English only. Preserving regional sign languages, like Indian Sign Language (ISL), is crucial for cultural identity. This paper presents a real-time two-way translation system for Indian Sign Language using Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, enabling seamless communication between the hearing-impaired, non-verbal community and others by accurately interpreting and translating ISL gestures into Tamil text and vice-versa.

Highlights of the proposed system are:

- Accuracy: The use of CNN and LSTM results in more accurate and contextually relevant translations, allowing for a more natural and understandable communication experience.
- Scalability: The model's design allows for scalability, making it possible to extend its capabilities to support additional languages and improve its performance over time.
- Bi-directional communication: The system facilitates both text-to-sign language and sign language-to-text translations, enabling a comprehensive and effective two-way communication process.
- Regional language –Tamil to ISL translation: Tamil sentences are translated to ISL signs and vice-versa.

3. METHODOLOGY

The proposed translator system consists of two main modules, (i) Sign-to-Text Translator and (ii) Speech/ Text to Sign Translator. Sign to Text translator uses a combination of CNN and LSTM networks to process video input of sign language gestures and output the corresponding Tamil text. Speech/Text to Sign translator takes Tamil text or speech as input and generates corresponding sign language gestures, which are output as video sequences.

Figure. 1 Block diagram of proposed Sign-Text translation system

Figure 1 shows the key steps in the Sign to Text translation process. It starts with capturing the video of the signer. The video would be divided into a number of frames of raw image sequence. This image sequence will then be processed to initially identify the boundaries. This will be useful to separate the different body parts being captured by the camera into two major subparts - head and hands. The head subpart will be further categorized into pose and movements as well as facial expressions. Postures and gestures will be extracted from the movement of the hands. The processed data is converted into array for faster and less complex computation purpose. All of the data will then be matched against the ISL Dataset which would then be used for classification purposes. The proposed work employs two key deep learning techniques: Convolutional Neural Networks (CNN) for image processing and Recurrent Neural Networks with Long Short-Term Memory (RNN -LSTM) for sentence formation. The CNN model is trained on a data set containing images of sign language gestures corresponding to different words in the input language. When the individuals express themselves through sign language, the system captures the gestures and it is recognized by CNN, RNN -LSTM comes into play for sentence formation. RNN -LSTM ensures coherence and fluency in converting the input into textual form. Frames were extracted from the input sign video and the gesture information is derived from these frames in terms of holistic landmarks. This step focuses on capturing not only hand positions and gestures, but also include full body movements and facial expressions. The dataset generated is analysed using the CNN model. Adam optimizer is used for training the model.

Figure 2 shows the internal architecture of the proposed system. It consists of two CNN layers followed by max pooling layers for features extraction, three LSTM stacks and three dense layers for sentence prediction. The output shape and parameter count for each layer in your model are shown in the Fig. The first convolutional layer consists of 24,800 parameters which represent the weights and biases of the convolutional layer. Sequence length and number of filters are 23 and 32 respectively. This is followed by a max-pooling layer. The max-pooling operation reduce the sequence length from 23 to 11 while keeping the number of filters unchanged. Max-pooling layers do not have trainable parameters, so the parameter count is 0. Second convolutional layer, convolves with 64 filters over the input data, resulting in an output shape of 9,64. There are 6,208 parameters for this layer. The max-pooling layer at the second level reduces the sequence length further, resulting in an output shape of 4, 64.

International Journal of Computer Applications Technology and Research
Volume 13–Issue 08, 34 – 38, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1308.1004

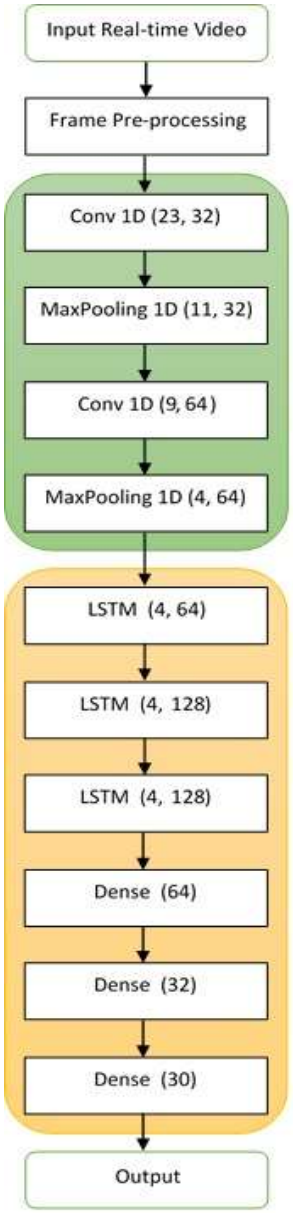

Figure. 2 Internal Framework of the Proposed Model

The LSTM layer processes the sequence data and outputs sequences of hidden states. There are 33,024 and 98,816 parameters associated with the first and second LSTM layers respectively. The third LSTM layer aggregates the information across the sequence and produces a final output for each sample. There are 49,408 parameters associated with this LSTM layer. The first dense layer has 64 neurons and applies a rectified linear activation function to its inputs. There are 4,160 parameters associated with this dense layer. The second dense layer has 32 neurons and 2,080 parameters. The final dense layer produces the output of the model with 30 neurons corresponding to the number of classes in the problem. There are 990 parameters associated with this dense layer.

Figure. 3 Block diagram of proposed Sign-Text translation system

Figure 3 shows the Speech/Text to Sign translation process. The system accepts either speech or text as input. In case of speech input captured by microphone, speech recognition library is utilized to convert speech into text. The input or converted text is parsed and converted into key parts or words using tokenisation and lemmatization from Natural Language Processing (NLP) library. This sequence of words is mapped with the labelled sign – video data base and eventually corresponding gesture videos are rendered as output.

4. IMPLEMENTATION AND RESULTS

Dataset is created by capturing sign – videos for a set of thirty commonly used Tamil words, sixty clips are generated per one word acted by five different signers and each clip is of twenty-five frames size. For each selected word, the code runs in loop to collect 60 different videos. The dataset of holistic gesture information is generated after converting the videos into frames and this part is done using media pipe and open-CV. The data is split into two sets - 70%(42 videos) for training and 30% (18 videos) for testing for each sign. The performance of the model is evaluated using accuracy, precision and recall metrics.

4.1 Text to Sign Translation

During testing, a live video is captured using the help of a camera and fed to the pre-trained translator system. The system reads the first 25 frames whenever the left or right hand movement is detected. The classification task results in the prediction of the words and if this probability of prediction is over 0.8, the predicted word is displayed. Comparison with the previous prediction is done to avoid the repetition. The predicted word is then fed to the sentence array. The array is repeatedly tested for sentence prediction using the sentence predictor function, when the time threshold and the correct sentence is found the sentence displayed at the terminal.

Figure. 4 Output window

Figure 4 shows the screenshot of the IDE and output window during a sample test run. The markers of holistic gesture

detection and the predicted sentence at the output window. can be observed from the figure. This model is tested with various sentence forming words and the result is promising with an accuracy of 92.9 %.

4.2 Text to Sign Translation

Our System outlines a GUI application using the tkinter library in Python, designed for translating speech or text into Indian Sign Language (ISL) videos. This application, titled "Speech/Text to SIGN," features a main window with various interactive elements to facilitate user interaction and display ISL videos corresponding to recognized speech or entered text.

Figure. 5 Sample Input

At the core of the application is a main window configured to a size of 800x600 pixels, containing a canvas where ISL videos are displayed. The status of the application, such as "Ready," "Listening...," or any error messages, is shown at the bottom of the window in a status label, which provides real-time feedback to the user. This feedback mechanism is crucial for accessibility and usability, ensuring users are constantly informed of the application's state.

For input, the application uses speech recognition powered by the speech recognition library, which captures audio through the microphone. Users can start and stop this audio capture using "Start Listening" and "Stop Listening" buttons, respectively. An additional feature allows users to directly input text through a dialog box, catering to those who may prefer typing or have environments unsuitable for speech recognition.

Figure.6 Output window for text to sign translation

Figure 5 shows the screenshot of the text input window for text to sign translation. The application also includes a multithreading approach to handle speech recognition processes, ensuring the GUI remains responsive and does not freeze during audio processing. When speech is recognized or text is entered, the application splits the input into words and searches for corresponding ISL videos in a predefined directory. Each video is played on the canvas, resized and converted from BGR to RGB format for compatibility with tkinter and PIL for display. Figure 6 shows the screenshot of the output window over the main window during the sample text to sign translation.

4.3 Performance

The performance of the model is computed and visualized using Scikit-Learn toolkit. The metrics used are accuracy, precision and recall. Accuracy measures the proportion of correct predictions which is computed as the ratio of correct predictions out of total number of predictions. Precision is the ratio between true positives and total positive predictions including true and false positives. Recall score measures the ability to detect positives and computed as the ratio between true positives and sum of true positives and false negatives. This model is tested with various sentence forming words and the result is reliable with an accuracy of 92.9%. High degree of precision is confirmed by the precision and recall score of 0.937 and 0.927 respectively.

5. CONCLUSION

By creating a more accurate and efficient system for translating sign language into text or speech, using CNN-LSTM model, this work aims to bridge the communication gap between the hearing- or speech-impaired community and the rest of society. A comprehensive approach utilizing various technologies such as Tkinter for GUI, OpenCV for video processing, speech recognition library for audio input, and threading for performance management to create a multimedia-based two-way translator for ISL is demonstrated. This also makes the application a promising resource for enhancing communication through ISL, providing a practical and interactive way for users to learn and practice sign language effectively. Beyond its technical aspects, it has the potential to transform lives, promoting understanding and empathy in the society.

6. ACKNOWLEDGMENTS

Our sincere thanks to the Govt. Special School for Differently Abled Children, Pillaichavady, Puducherry for their valuable feedback and support.

7. REFERENCES

[1] Indian Sign Language [Online]. Available: https://indiansignlanguage.org/.

[2] V. Ackroyd, B.J.D Wright. 2018. Working with British Sign Language (BSL) interpreters: lessons from child and adolescent mental health services in the U.K. Journal of Communication in Healthcare.

[3] U. Farooq, M. S. Mohd Rahim, N. Sabir, A. Hussain, and A. Abid. 2021. Advances in machine translation for sign language: approaches, limitations, and challenges, Neural Comput. Appl., 33, 14357– 14399.

[4] B. Joksimoski et al., 2022. Technological Solutions for Sign Language Recognition: A Scoping Review of Research Trends, Challenges, and Opportunities. IEEE Access, 10, 40979-40998.

[5] P. Sonawane, K. Shah, P. Patel, S. Shah, and J. Shah. 2021. Speech to Indian Sign Language (ISL) Translation System.

International Conference on Computing, Communication, and Intelligent Systems (ICCCIS).

[6] H. Monga,J, Bhutani,M. Ahuja, N. Maid, and H. Pande. 2021. Speech to Indian Sign Language Translator, Recent Trends in Intensive Computing.

[7] Sugandhi, P. Kumar and S. Kaur. 2021. Indian Sign Language Generation System, in Computer, 54:3, 37-46.

[8] Sundar, B., & Bagyammal, T. 2022. American Sign Language Recognition for Alphabets Using MediaPipe and LSTM. Procedia Computer Science, 215:642–6512.

[9] Bora, J., Dehingia, S., Boruah, A., Chetia, A. A., & Gogoi, D. 2023. Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning. Procedia Computer Science. 218, 1384–1393.

[10] Bandi Rupendra Reddy, Daka Chandra Rup, Mathi Rohith, Meena Belwal. 2023. Indian Sign Language Generation from Live Audio or Text for Tamil. 9th International Conference on Advanced Computing and Communication Systems.

[11] Jayanthi P, Ponsy R K Sathia Bhama & B Madhubalasri. 2023. Sign Language Recognition using Deep CNN with Normalised Keyframe Extraction and Prediction using LSTM, Journal of Scientific & Industrial Research, 82.